变分不等式框架下结构型凸优化的分裂收缩算法

序言 与 目录

中学的数理基础　必要的社会实践
普通的大学数学　一般的优化原理

南京大学数学系　何炳生
http://maths.nju.edu.cn/~hebma

天元数学东北中心　2023 年 10 月 17 – 27 日

序　言

这个讲座系统介绍最近 20 多年来我们对一些典型凸优化问题求解方法开展的颇具特色的研究, 包括一些被国际著名学者高度认可的成果.

变分不等式这个比较陌生的术语其实就是盲人爬山判别是否已经到达了顶点的数学表达形式, 读者将会看到, 用变分不等式的观点处理凸优化问题, 就像微积分中用导数求可微凸函数的极值点, 常常会带来很大的方便.

邻近点算法 (PPA) 和增广 Lagrange 乘子法 (ALM) 是最优化中的一些经典算法. ALM 本身就是 Lagrange 乘子的 PPA 算法, 这些算法生成的序列都具有向解集越靠越近的收缩性质, 因此我们称其为收缩算法. 变分不等式(VI)和邻近点算法(PPA)是我们开展凸优化方法研究的两大法宝.

在约束凸优化问题拉格朗日函数的鞍点和变分不等式的解点等价的基础上, 我们提出的预测–校正算法统一框架. 利用这个框架, 已有算法的收敛性证明只需要（通过简单的矩阵运算）验证两个条件; 理解掌握了这个框架, 设计求解结构型可分离凸优化问题的预测-校正方法, 难度犹如线性代数课程完成课堂练习一样. 学习这个课程, 只要求读者具备普通的大学数学知识.

根据商定课时拼凑成的讲义, 有错误的地方, 希望读者不吝指正.

变分不等式框架下结构型凸优化的分裂收缩算法 — 目录

PPT 的篇幅较长, 包含所有的证明, 内容比较详细, 便于学员课后复习理解

个 人 研 究 生 涯 的 几 点 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, 应是同一道理 ✠

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 ✠

变分不等式框架下结构型 凸优化的分裂收缩算法

I. 凸优化及其在变分不等式框架下的邻近点算法

中学的数理基础	必要的社会实践
普通的大学数学	一般的优化原理

何 炳 生　　南京大学数学系

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

天元数学东北中心　2023 年 10 月 17 – 27 日

华罗庚先生普及“双法”对我们的启示

- 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。
- 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.

能够交给“群众”的方法，应该是普通大学生能够理解，掌握的方法.

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

Department 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 and Technology (SUSTech)
2013-2015 Professor, School of Management Science and Engineering, Nanjing University
1997~2013 Professor, Department of Mathematics, Nanjing University
1992~1997 Associate Professor, Department of Mathematics,Nanjing University

我会尽量维护自己的主页, 不断修正、更新自己学术的体验.

My Thinkings: 1. 关门感想 2. 说说我的主要研究兴趣 — 兼谈华罗庚推广优选法对我的影响
3. 说说我的主要研究兴趣(续) --- 我们在ADMM类方法的主要工作 4. 古稀回首
5. 两页纸给出ADMM 收敛性证明 6. 两页纸简述我职业生涯中的主要研究工作
7. 两页纸描述古稀之后的研究进展: 从好不容易凑出一个方法到并不费力构造一簇算法
8. 两页PPT讲述求解线性约束凸优化问题预测-校正的广义邻近点(Generaized PPA)算法
《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.

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
18. 2023年1月在华南师大《华人数学家论坛》的报告 — 凸优化分裂收缩算法统一框架的最新进展
17. 2022年11月在西电《最优化前沿论坛》的报告 — 从好不容易凑出一个算法到并不费劲构造一簇算法
16. 2022 年7月南理工《计算机》方向暑期班六讲摘要 和六讲的PPT I II III IV V VI
15. 利用预测-校正统一框架构造凸优化的分裂收缩算法(由预测矩阵构造校正矩阵) (ArXiv: 2204.11522)
14. 2022 年元月南师大数科院系列报告B站视频辅助材料 A B C D E F G H I J K L
13. ADMM 类分裂收缩算法的一些最新进展 统一框架下Balanced-ALM 便于向多块推广的ADMM
12. 均困平衡的增广拉格朗日乘子法 — Balanced ALM (一类新的增广拉格朗日乘子法ArXiv: 2108.08554)
11. 一类便于向求解多块问题推广并能处理不等式约束问题的交替方向法 (ArXiv:2107.01897)
10. 瞎子爬山-步步为营—凸优化算法中的变分不等式和邻近点策 略(南京大学数学系本科生论坛上的报告)
9. 被S. Becker 誉为 Very Simple yet Powerful 的 Technique — 应用及新的进展
8. 线性化ALM 线性化ADMM 以及处理三个可分离块问题中缩小有关参数至3/4提高效率的方法

我的报告的 PDF 文件, 一般都可以在我的主页上查到.

连续优化中一些代表性数学模型

1. 鞍点问题 $\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\{\Phi(x,y)=\theta_1(x)-y^TAx-\theta_2(y)\}$
2. 线性约束的凸优化问题 $\min\{\theta(x)|Ax=b\,(\text{or}\geq b),\ x\in\mathcal{X}\}$
3. 结构型凸优化 $\min\{\theta_1(x)+\theta_2(y)|Ax+By=b,\ x\in\mathcal{X},\ y\in\mathcal{Y}\}$
4. 多块可分离凸优化 $\min\{\sum_{i=1}^p\theta_i(x_i)|\sum_{i=1}^p A_ix_i=b, x_i\in\mathcal{X}_i\}$

变分不等式(VI) 是瞎子爬山的数学表达形式

邻近点算法(PPA) 是步步为营 稳扎稳打的求解方法.

变分不等式和邻近点算法是分析和设计凸优化方法的两大法宝.

分裂是指迭代中子问题都通过分拆求解. 收缩算法有别于可行方向法,
又有别于下降算法, 它的迭代点离优化问题的拉格朗日函数的鞍点越来越近.

这一讲解释上述问题都可以化为一个单调变分不等式 并介绍什么是邻近点算法

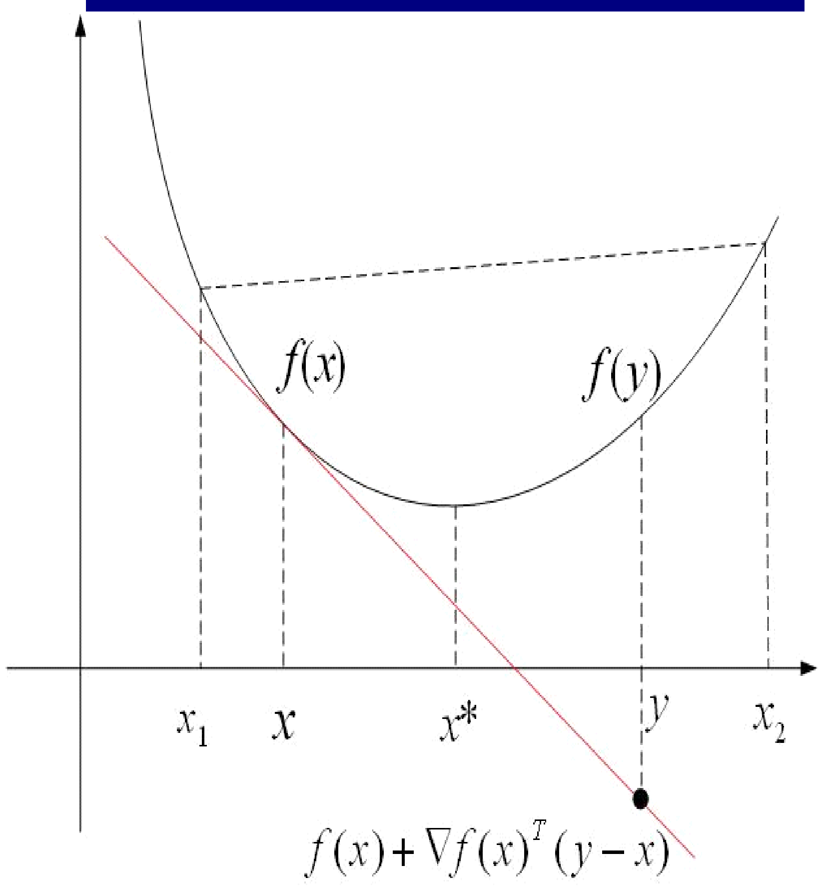
凸函数的定义和基本性质

A function $f(x)$ is convex iff

$$f((1-\theta)x+\theta y)\leq(1-\theta)f(x)+\theta f(y)$$
$$\forall\theta\in[0,1].$$

Properties of convex function

- $f\in\mathcal{C}^1$. f is convex iff
  $$f(y)-f(x)\geq\nabla f(x)^T(y-x).$$
  Thus, we have also
  $$f(x)-f(y)\geq\nabla f(y)^T(x-y).$$
- Adding above two inequalities, we get
  $$(y-x)^T(\nabla f(y)-\nabla f(x))\geq 0.$$



Convex function

- $f\in\mathcal{C}^1$, ∇f is monotone. $f\in\mathcal{C}^2$, $\nabla^2 f(x)$ is positive semi-definite.
- Any local minimum of a convex function is a global minimum.

1 Optimization problem and VI

1.1 Differential convex optimization in Form of VI

Let $\Omega \subset \Re^n$, we consider the convex minimization problem

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

What is the first-order optimal condition ?

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

Optimal condition in variational inequality form

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

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

瞎子爬山判定山顶的准则是: 所有可行方向都不再是上升方向

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

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

Substituting $\nabla f(x)$ with an operator F (from $\Re^n$ into itself), we get a classical VI.

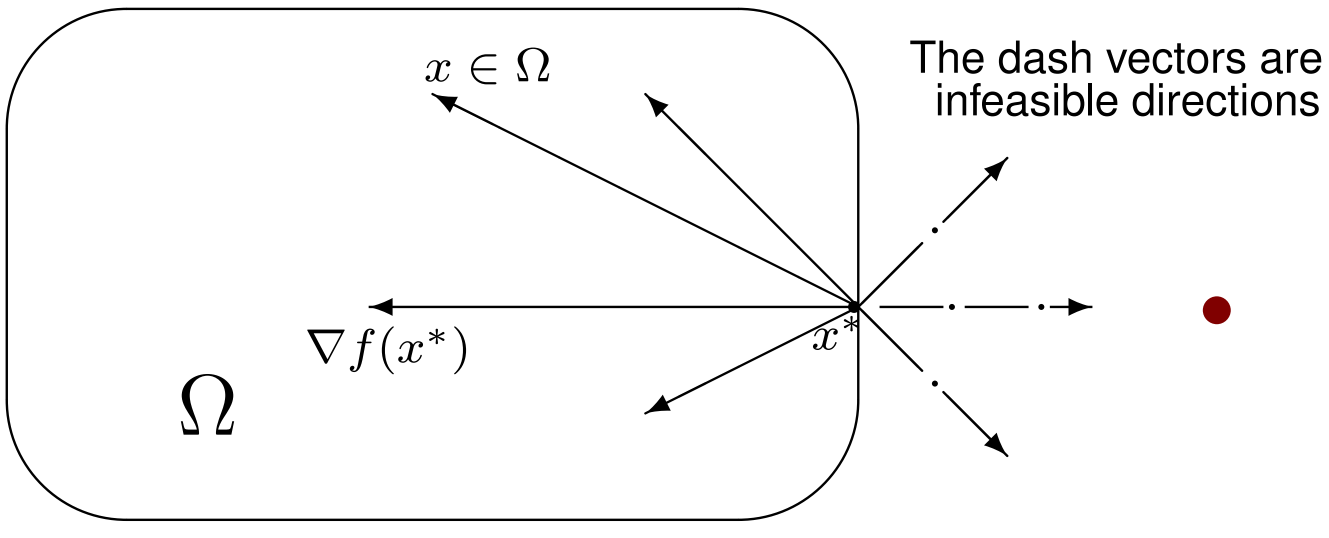


Fig. 1.1 Differential Convex Optimization and VI

Since $f(x)$ is a convex function, we have

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \text{and thus} \quad (x - y)^T (\nabla f(x) - \nabla f(y)) \geq 0.$$

We say the gradient ∇f of the convex function f is a monotone operator.

通篇我们需要用到的大学数学　主要是基于微积分学的一个引理

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

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

上面的凸优化最优性条件是最基本的, 看起来合在一起就是下面的引理:

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

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

if and only if 　　凸优化最优性条件定理

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

定理把优化问题 (1.3a) 转换成了变分不等式 (1.3b). 下面给出证明.

Proof : First, if (1.3a) is true, then for any $x \in \mathcal{X}$, we have

$$\frac{\theta(x_\alpha) - \theta(x^*)}{\alpha} + \frac{f(x_\alpha) - f(x^*)}{\alpha} \ge 0, \tag{1.4}$$

where

$$x_\alpha = (1 - \alpha)x^* + \alpha x, \quad \forall \alpha \in (0, 1].$$

Because $\theta(\cdot)$ is convex, it follows that

$$\theta(x_\alpha) \le (1 - \alpha)\theta(x^*) + \alpha\theta(x),$$

and thus

$$\theta(x) - \theta(x^*) \ge \frac{\theta(x_\alpha) - \theta(x^*)}{\alpha}, \quad \forall \alpha \in (0, 1].$$

Substituting the last inequality in the left hand side of (1.4), we have

$$\theta(x) - \theta(x^*) + \frac{f(x_\alpha) - f(x^*)}{\alpha} \ge 0, \quad \forall \alpha \in (0, 1].$$

Using $f(x_\alpha) = f(x^* + \alpha(x - x^*))$ and letting $\alpha \to 0_+$, from the above inequality we get

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

Thus (1.3b) follows from (1.3a). Conversely, since f is convex, it follow that

$$f(x_\alpha) \le (1-\alpha) f(x^*) + \alpha f(x)$$

and it can be rewritten as

$$f(x_\alpha) - f(x^*) \le \alpha(f(x) - f(x^*)).$$

Thus, we have

$$f(x) - f(x^*) \ge \frac{f(x_\alpha) - f(x^*)}{\alpha} = \frac{f(x^* + \alpha(x - x^*)) - f(x^*)}{\alpha},$$

for all $\alpha \in (0, 1]$. Letting $\alpha \to 0_+$, we get

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

Substituting it in the left hand side of (1.3b), we get

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

and (1.3a) is true. The proof is complete. □

可微约束优化问题的最优性必要条件

设 $f(x)$, $\varphi_i(x), i = 1, \ldots, m$, 都是从 $\Re^n \to \Re$ 的连续可微函数, 研究问题

$$\begin{array}{rl} \min & f(x) \\ \text{s.t} & \varphi_1(x) = 0, \\ & \vdots \\ & \varphi_m(x) = 0 \end{array}$$

相应的Lagrange函数

$$L(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i \varphi_i(x).$$

最优性必要条件是:

$$\left\{ \begin{array}{l} \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} - \begin{pmatrix} \frac{\partial \varphi_1}{\partial x_1} & \frac{\partial \varphi_2}{\partial x_1} & \cdots & \frac{\partial \varphi_m}{\partial x_1} \\ \frac{\partial \varphi_1}{\partial x_2} & \frac{\partial \varphi_2}{\partial x_2} & \cdots & \frac{\partial \varphi_m}{\partial x_2} \\ \vdots & & & \\ \frac{\partial \varphi_1}{\partial x_n} & \frac{\partial \varphi_2}{\partial x_n} & \cdots & \frac{\partial \varphi_m}{\partial x_n} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{pmatrix} = 0. \\ \varphi_i(x) = 0, \quad i = 1, \ldots, m. \end{array} \right.$$

1.2 Linear constrained convex optimization and VI

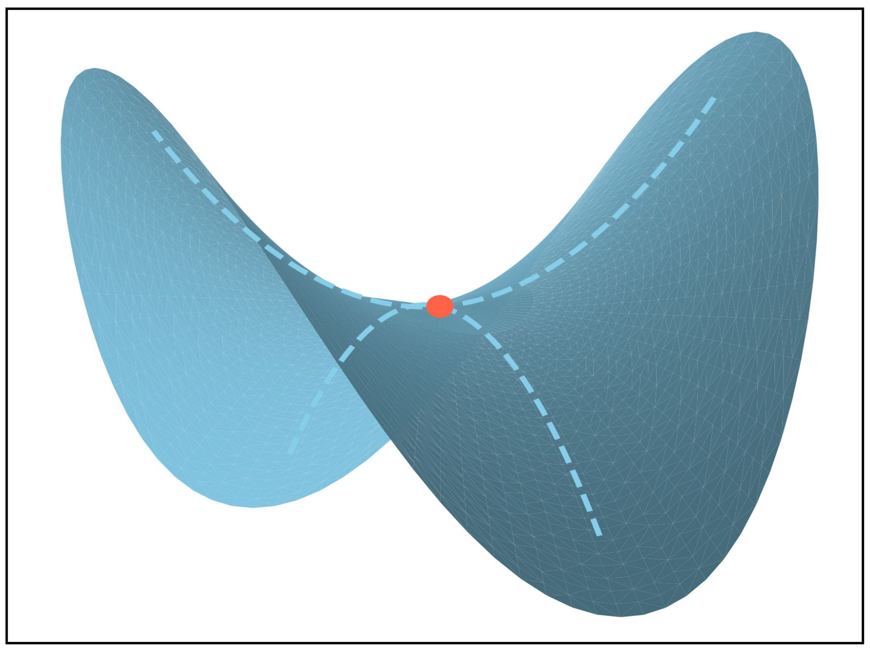
We consider the linearly constrained convex optimization problem

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

The Lagrangian function of the problem (1.5) is

$$L(u,\lambda) = \theta(u) - \lambda^T(\mathcal{A}u - b), \quad (1.6)$$

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



Example 1 of the problem (1.5): Finding the nearest correlation matrix

A positive semi-definite matrix, whose each diagonal element is equal 1, is called the correlation matrix. For given symmetric $n \times n$ matrix C, the mathematical form of finding the nearest correlation matrix X is

$$\min\{\tfrac{1}{2}\|X - C\|_F^2 \mid \text{diag}(X) = e,\ X \in S_+^n\}, \quad (1.7)$$

where S_+^n is the positive semi-definite cone and e is a n-vector whose each element is equal 1. The problem (1.7) is a concrete problem of type (1.5).

Example 2 of the problem (1.5): The matrix completion problem

Let M be a given $m \times n$ matrix, Π is the elements indices set of M,

$$\Pi \subset \{(ij) | i \in \{1,\dots,m\},\ j \in \{1,\dots,n\}\}.$$

The mathematical form of the matrix completion problem is relaxed to

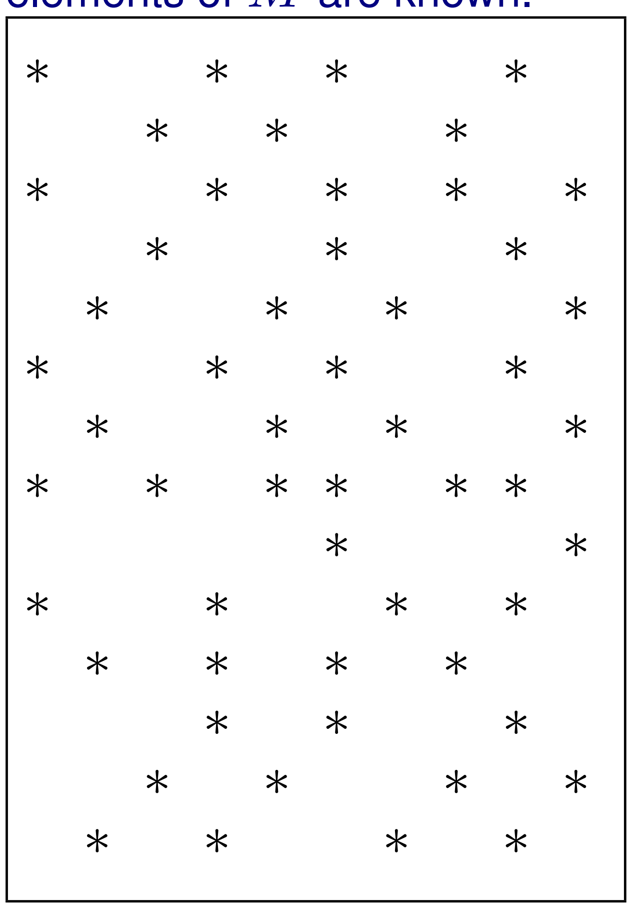
$$\min\{\|X\|_* \mid X_{ij} = M_{ij},\ (ij) \in \Pi\}, \quad (1.8)$$

where $\|\cdot\|_*$ is the nuclear norm–the sum of the singular values of a given matrix. The problem (1.8) is a convex optimization of form (1.5). The matrix A in (1.5) for the linear constraints

$$X_{ij} = M_{ij},\ (ij) \in \Pi,$$

is a projection matrix, and thus $\|A^T A\| = 1$.

M is low Rank, only some elements of M are known.



A pair of $(u^*, \lambda^*) \in \mathcal{U} \times \Re^m$ is called a saddle point of the Lagrange function (1.6), 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

$$\begin{cases} u^* \in \mathcal{U}, \quad L(u, \lambda^*) - L(u^*, \lambda^*) \ge 0, \quad \forall\, u \in \mathcal{U}, & \text{(1.9a)} \\ \lambda^* \in \Re^m, \ L(u^*, \lambda^*) - L(u^*, \lambda) \ge 0, \quad \forall\, \lambda \in \Re^m. & \text{(1.9b)} \end{cases}$$

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

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

it follows from (1.9a) that

$$u^* \in \mathcal{U}, \quad \theta(u) - \theta(u^*) + (u - u^*)^T(-\mathcal{A}^T\lambda^*) \ge 0, \quad \forall\, u \in \mathcal{U}. \quad \text{(1.10)}$$

Similarly, for (1.9b), since

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

thus we have

$$\lambda^* \in \Re^m, \quad (\lambda - \lambda^*)^T(\mathcal{A}u^* - b) \ge 0, \quad \forall\, \lambda \in \Re^m. \qquad \text{(1.11)}$$

Notice that the expression (1.11) (the inner product of the vector $(\mathcal{A}u^* - b)$ with any vector is nonnegative) is equivalent to

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

Writing (1.10) and (1.11) together, we get the following variational inequality:

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

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

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \qquad \text{(1.12a)}$$

where

$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{U} \times \Re^m. \qquad \text{(1.12b)}$$

Setting $w = (u, \lambda^*)$ and $w = (u^*, \lambda)$ in (1.12), respectively, we get (1.10) and (1.11). Because F is a affine operator and

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

The matrix is skew-symmetric, we have

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

线性约束的凸优化问题 (1.5), 转换成了混合变分不等式 (1.12).

Two block separable convex optimization

We consider the following structured separable convex optimization

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\, x \in \mathcal{X}, y \in \mathcal{Y}\}. \qquad \text{(1.13)}$$

This is a special problem of (1.5) with

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

The Lagrangian function of the problem (1.13) is

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

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

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \qquad \text{(1.14)}$$

where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y), \quad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \qquad (1.15a)$$

$$F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}, \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m. \qquad (1.15b)$$

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

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

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

可分离线性约束凸优化问题 (1.13), 转换成了变分不等式 (1.14)–(1.15).

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 (1.5) with three blocks. The Lagrangian function is

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

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

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

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

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

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

线性约束的凸优化问题, 都转换成了变分不等式. 问题归结为求一个鞍点.

2 Proximal point algorithms and its Beyond

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

$\|x\|^2 = x^T x, \quad \|x\|_H^2 = x^T H x.$

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

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

2.1 Proximal point algorithms for convex optimization

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

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

where $\theta(x)$ and $f(x)$ are convex but $\theta(x)$ is not necessary smooth, $\mathcal{X}$ is a closed convex set. For solving (2.2), the k-th iteration of the proximal point algorithm (abbreviated to PPA) [8, 10] begins with a given x^k, offers the new iterate x^{k+1} via the recursion

邻近点算法

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

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

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

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

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

Because f isconvex, $(x^{k+1} - x^*)^T \nabla f(x^{k+1}) \geq (x^{k+1} - x^*)^T \nabla f(x^*)$, it follows that

$$\begin{aligned} &\theta(x^{k+1}) - \theta(x^*) + (x^{k+1} - x^*)^T \nabla f(x^{k+1}) \\ \geq \quad &\theta(x^{k+1}) - \theta(x^*) + (x^{k+1} - x^*)^T \nabla f(x^*) \geq 0 \end{aligned}$$

and consequently,

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

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

PPA 算法的收缩性质

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

which is the nice convergence property of Proximal Point Algorithm.

The residue sequence $\{\|x^k - x^{k+1}\|\}$ is also monotonically no-increasing.

Proof. Replacing $k+1$ in (2.4) with k, we get

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

Let $x = x^{k+1}$ in the above inequality, it follows that

$$\theta(x^{k+1}) - \theta(x^k) + (x^{k+1} - x^k)^T\{\nabla f(x^k) + r(x^k - x^{k-1})\} \geq 0. \qquad (2.7)$$

Setting $x = x^k$ in (2.4), we become

$$\theta(x^k) - \theta(x^{k+1}) + (x^k - x^{k+1})^T\{\nabla f(x^{k+1}) + r(x^{k+1} - x^k)\} \geq 0. \qquad (2.8)$$

Adding (2.7) and (2.8) and using $(x^k - x^{k+1})^T[\nabla f(x^k) - \nabla f(x^{k+1})] \geq 0$, we get

$$(x^k - x^{k+1})^T\{(x^{k-1} - x^k) - (x^k - x^{k+1})\} \geq 0. \qquad (2.9)$$

Setting $a = x^{k-1} - x^k$ and $b = x^k - x^{k+1}$ in (2.9) and using (2.1), we obtain

$$\|\boldsymbol{x^k} - \boldsymbol{x^{k+1}}\|^2 \leq \|\boldsymbol{x^{k-1}} - \boldsymbol{x^k}\|^2 - \|(\boldsymbol{x^{k-1}} - \boldsymbol{x^k}) - (\boldsymbol{x^k} - \boldsymbol{x^{k+1}})\|^2. \qquad (2.10)$$

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

For the optimization problem (2.2) , namely, $\min\{\theta(x) + f(x) \mid x \in \mathcal{X}\}$,
the equivalent variational inequality form is

$$x^* \in \mathcal{X}, \;\; \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0, \;\; \forall x \in \mathcal{X}. \qquad (2.11)$$

For solving the problem (2.2), the PPA is

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

variational inequality form of the k-th iteration of the PPA (see (2.4)) is:

$$x^{k+1} \in \mathcal{X}, \;\; \theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T \nabla f(x^{k+1})$$
$$\geq (x - x^{k+1})^T r(x^k - x^{k+1}), \;\; \forall x \in \mathcal{X}. \qquad (2.12)$$

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

The solution of (2.12) is Proximal Point, it has the contraction property (2.6).

2.2 Preliminaries of PPA for Variational Inequalities

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

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{2.13}$$

PPA for VI (2.13) **in H-norm (定义)** For given w^k and $H \succ 0$, find w^{k+1},

$$w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \quad \text{邻近点算法}$$
$$\geq (w - w^{k+1})^T H(w^k - w^{k+1}), \quad \forall\, w \in \Omega, \tag{2.14}$$

w^{k+1} is called the proximal point of the k-th iteration for the problem (2.13).

(2.14) 是求解 VI (2.13) 的 PPA 算法的定义. 第二讲就会用例子说明这是容易做到的.

✠ w^{k+1} **is the solution of** (2.13) **if and only if** $w^k = w^{k+1}$ ✠

Setting $w = w^*$ in (2.14), we obtain

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

Note that (see the structure of $F(w)$ in (1.12b))

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

and consequently (by using (2.13)) we obtain

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

Thus, we have

$$(w^{k+1} - w^*)^T H(w^k - w^{k+1}) \geq 0. \tag{2.15}$$

By setting $\boldsymbol{a = w^k - w^*}$ and $\boldsymbol{b = w^{k+1} - w^*}$,
the inequality (2.15) means that $\boldsymbol{b^T H(a - b) \geq 0}$.

By using Lemma 1, we obtain

$$\boldsymbol{\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2}. \tag{2.16}$$

We get the nice convergence property of Proximal Point Algorithm.

请证明：$\|w^k - w^{k+1}\|^2 \leq \|w^{k-1} - w^k\|^2$, 即序列 $\{\|w^k - w^{k+1}\|_H\}$ 是单调不增的.

2.3 Variants of PPA for Variational Inequalities

Let v be a sub-vector of w. The k-th iteration begins with given v^k . v 核心变量

PPA for VI (2.13) in H-norm For given v^k and $H \succ 0$, find w^{k+1},

$$w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \geq (v - v^{k+1})^T H(v^k - v^{k+1}), \quad \forall\, w \in \Omega, \tag{2.17}$$

w^{k+1} is called the proximal point of the k-th iteration for the problem (2.13).

✠ w^{k+1} **is the solution of** (2.13) **if and only if** $v^k = v^{k+1}$ ✠

In this case, v is called the essential variables of w. In addition, we define

$$\mathcal{V}^* = \{v^* \text{ is a subvector of } w^* \mid w^* \in \Omega^*\}.$$

Setting $w = w^*$ in (2.17), we obtain

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

Note that (see the structure of $F(w)$ in (1.12b))

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

and consequently (by using (2.13)) we obtain

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

Thus, we have

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \geq 0. \tag{2.18}$$

By using Lemma 1, we obtain

$$\|\boldsymbol{v^{k+1} - v^*}\|_{\boldsymbol{H}}^{\mathbf{2}} \leq \|\boldsymbol{v^k - v^*}\|_{\boldsymbol{H}}^{\mathbf{2}} - \|\boldsymbol{v^k - v^{k+1}}\|_{\boldsymbol{H}}^{\mathbf{2}}. \tag{2.19}$$

We get the nice convergence property of Proximal Point Algorithm.

The residue sequence $\{\|v^k - v^{k+1}\|_H\}$ is also monotonically no-increasing.

序列 $\{\|v^k - v^{k+1}\|_H\}$ 是单调不增的. $\|v^k - v^{k+1}\|_H^2 \leq \|v^{k-1} - v^k\|_H^2$.

3 Augmented Lagrangian Method (ALM)

We consider the convex optimization, namely

$$\min\{\theta(u) \mid \mathcal{A}u = b,\ u \in \mathcal{U}\}. \tag{3.1}$$

The related variational inequality of the saddle point of the Lagrangian function is

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{3.2a}$$

where

$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{U} \times \Re^m. \tag{3.2b}$$

Augmented Lagrangian Method

The augmented Lagrangian function of the problem (3.1) is

$$\mathcal{L}_\beta(u, \lambda) = \theta(u) - \lambda^T(\mathcal{A}u - b) + \frac{\beta}{2}\|\mathcal{A}u - b\|^2,$$

The k-th iteration of the **Augmented Lagrangian Method** [7, 9] begins with a given λ^k, obtain $w^{k+1} = (u^{k+1}, \lambda^{k+1})$ via

$$\text{(ALM)} \quad \begin{cases} u^{k+1} = \arg\min\{\mathcal{L}_\beta(u, \lambda^k) \;\big|\; u \in \mathcal{U}\}, & (3.3a) \\ \lambda^{k+1} = \lambda^k - \beta(\mathcal{A}u^{k+1} - b). & (3.3b) \end{cases}$$

In (3.3), u^{k+1} is only a computational result of (3.3a) from given λ^k, it is called the intermediate variable. In order to start the k-th iteration of ALM, we need only to have λ^k and thus we call it as the essential variable.

The subproblem (3.3a) is a problem of mathematical form

$$\min\{\theta(u) + \tfrac{\beta}{2}\|\mathcal{A}u - p^k\|^2 \mid u \in \mathcal{U}\} \tag{3.4}$$

where $\beta > 0$ is a given scalar and $p^k = b + \frac{1}{\beta}\lambda^k$.

Assumption: The solution of problem (3.4) has closed-form solution or can be efficiently computed with a high precision.

Changing the constant term in the objective function does not affect the solution of the optimization problem. Thus,

$$\begin{aligned} u^{k+1} &\in \operatorname{argmin}\{\mathcal{L}_\beta(u,\lambda^k) \mid u\in\mathcal{U}\} \\ &= \operatorname{argmin}\{\theta(u) - (\lambda^k)^T\mathcal{A}u + \tfrac{\beta}{2}\|\mathcal{A}u-b\|^2 \mid u\in\mathcal{U}\} \\ &= \operatorname{argmin}\{\theta(u) + \tfrac{\beta}{2}\|(\mathcal{A}u-b) - \tfrac{1}{\beta}\lambda^k\|^2 \mid u\in\mathcal{U}\} \end{aligned}$$

According to Lemma 1, the optimal condition of (3.3a) is $u^{k+1}\in\mathcal{U}$ and

$$\theta(u) - \theta(u^{k+1}) + (u-u^{k+1})^T\{-\mathcal{A}^T\lambda^k + \beta\mathcal{A}^T(\mathcal{A}u^{k+1}-b)\} \ge 0,\ \forall u\in\mathcal{U}.$$

Because $\lambda^k - \beta(\mathcal{A}u^{k+1}-b) = \lambda^{k+1}$, the above VI can be written as

$$u^{k+1}\in\mathcal{U},\quad \theta(u) - \theta(u^{k+1}) + (u-u^{k+1})^T\{-\mathcal{A}^T\lambda^{k+1}\} \ge 0,\ \forall u\in\mathcal{U}. \quad (3.5)$$

The update form (3.3b) is

$$(\mathcal{A}u^{k+1}-b) + \frac{1}{\beta}(\lambda^{k+1}-\lambda^k) = 0.$$

and it is equivalent to

$$(\lambda-\lambda^{k+1})^T(\mathcal{A}u^{k+1}-b) \ge (\lambda-\lambda^{k+1})^T\frac{1}{\beta}(\lambda^k-\lambda^{k+1}),\quad \forall\lambda\in\Re^m. \quad (3.6)$$

Combining VI's (3.5) and (3.6), we get

$$\theta(u)-\theta(u^{k+1}) + \begin{pmatrix} u-u^{k+1} \\ \lambda-\lambda^{k+1}\end{pmatrix}^T \begin{pmatrix} -\mathcal{A}^T\lambda^{k+1} \\ \mathcal{A}u^{k+1}-b\end{pmatrix} \ge (\lambda-\lambda^{k+1})^T\frac{1}{\beta}(\lambda^k-\lambda^{k+1}),$$

for all $w=(u,\lambda)\in\Omega$. Using the notations in (3.2), we get the compact form

$$\begin{aligned} &\theta(u)-\theta(u^{k+1}) + (w-w^{k+1})^T F(w^{k+1}) \\ &\qquad \ge\ (\lambda-\lambda^{k+1})^T\frac{1}{\beta}(\lambda^k-\lambda^{k+1}),\ \forall w\in\Omega. \end{aligned} \quad (3.7)$$

This is the PPA form (2.17) in which

$$v=\lambda \quad \text{and} \quad H=\frac{1}{\beta}I_m.$$

The related contraction inequality (2.19) becomes

$$\|\lambda^{k+1}-\lambda^*\|^2_{\frac{1}{\beta}I_m} \le \|\lambda^k-\lambda^*\|^2_{\frac{1}{\beta}I_m} - \|\lambda^k-\lambda^{k+1}\|^2_{\frac{1}{\beta}I_m}$$

or

$$\|\lambda^{k+1}-\lambda^*\|^2 \le \|\lambda^k-\lambda^*\|^2 - \|\lambda^k-\lambda^{k+1}\|^2. \quad (3.8)$$

The above inequality is the key for the convergence proof of the ALM.

4 The relaxed PPA （延伸的邻近点算法）

We shall maintain our focus on the monotone variational inequality (2.13), namely,

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

The PPA form (2.17) reads as

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

Set the output of the above VI as $\tilde{w}^k$, we have

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ \geq (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \quad \forall\, w \in \Omega. \qquad (4.1)$$

Setting $w = w^*$ in (4.1), we obtain

$$(\tilde{v}^k - v^*)^T H(v^k - \tilde{v}^k) \geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k). \qquad (4.2)$$

Applying (see (1.12b)) the identity

$$(\tilde{w}^k - w^*)^T F(\tilde{w}^k) \equiv (\tilde{w}^k - w^*)^T F(w^*)$$

to (4.2), we obtain

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

Because RHS of the above inequality is , we have

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

We write it as

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

and thus

$$(v^k - v^*)^T H(v^k - \tilde{v}^k) \geq \|v^k - \tilde{v}^k\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \qquad (4.3)$$

The inequality (4.3) means that $(v^k - \tilde{v}^k)$ is the ascent direction of the unknown distance function $\frac{1}{2}\|v - v^*\|_H^2$ at the point v^k.

$$\left\langle \nabla\left(\tfrac{1}{2}\|v - v^*\|_H^2\right)\Big|_{v=v^k},\ (v^k - \tilde{v}^k)\right\rangle \ge \|v^k - \tilde{v}^k\|_H^2, \ \ \forall\, v^* \in \mathcal{V}^*.$$

The task of the algorithm is to produce a decreasing sequence $\{\|v^k - v^*\|_H^2\}$.
Set

$$v^{k+1}(\alpha) = v^k - \alpha(v^k - \tilde{v}^k) \tag{4.4}$$

which is an α dependent new iterate. It is clear we want to maximize

$$\vartheta(\alpha) = \|v^k - v^*\|_H^2 - \|v^{k+1}(\alpha) - v^*\|_H^2. \tag{4.5}$$

Note that

$$\begin{aligned} \vartheta(\alpha) &= \|v^k - v^*\|_H^2 - \|(v^k - v^*) - \alpha(v^k - \tilde{v}^k)\|_H^2 \\ &= 2\alpha(v^k - v^*)^T H(v^k - \tilde{v}^k) - \alpha^2\|v^k - \tilde{v}^k\|_H^2 \end{aligned} \tag{4.6}$$

is a quadratic function of α.

We can not directly maximize $\vartheta(\alpha)$ in (4.6) because the coefficient of the linear term $2(v^k - v^*)^T H(v^k - \tilde{v}^k)$ contains the unknown solution v^*.

Using (4.3), from (4.6) we get

$$\vartheta(\alpha) \ge 2\alpha\|v^k - \tilde{v}^k\|_H^2 - \alpha^2\|v^k - \tilde{v}^k\|_H^2 \tag{4.7}$$

Set

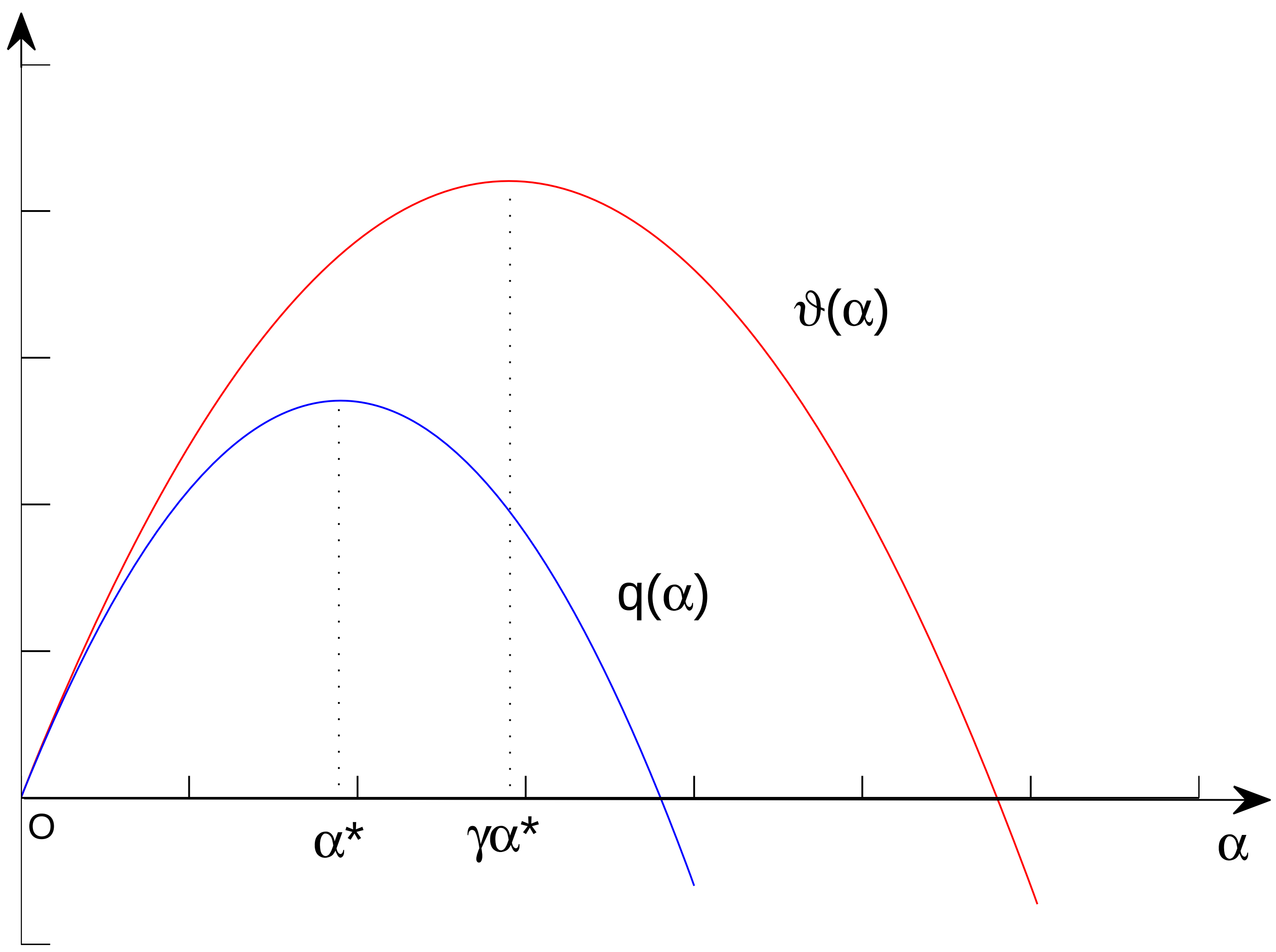
$$q(\alpha) = (2\alpha - \alpha^2)\|v^k - \tilde{v}^k\|_H^2, \tag{4.8}$$

which is a quadratic lower-bound function of $\vartheta(\alpha)$. The quadratic function $q(\alpha)$ reaches its maximum at $\alpha^* \equiv 1$.

$$v^{k+1} = v^k - \gamma(v^k - \tilde{v}^k), \quad \gamma \in (0, 2) \tag{4.9}$$

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

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \gamma(2-\gamma)\|v^k - \tilde{v}^k\|_H^2. \tag{4.10}$$



取 $\gamma \in [1,2)$ 的示意图

这一讲是预备知识. 要求读者理解（或者是先承认）优化问题拉格朗日函数的鞍点和变分不等式 (VI) 解点的等价的关系, 以及 PPA 算法的定义及收缩性质.

References

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

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

[3] R. Fletcher, Practical mathods of Optimization, Second Edition, JOHN WILEY & SONS, 1987.

[4] 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.

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

[6] B.S. He, X.M. Yuan and W.X. Zhang, A customized proximal point algorithm for convex minimization with linear constraints, Comput. Optim. Appl., 56(2013), 559-572.

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

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

[9] 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.

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

变分不等式框架下结构型
凸优化的分裂收缩算法

II 单块线性约束凸优化问题的PPA算法
和均困的增广拉格朗日乘子法

中学的数理基础	必要的社会实践
普通的大学数学	一般的优化原理

何 炳 生　　南京大学数学系

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

天元数学东北中心　2023年10月17 – 27日

1 Preliminaliers

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

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

if and only if

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

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

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

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

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

The optimal condition of the linearly constrained convex optimization

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

is characterized as a special mixed monotone variational inequality:

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

PPA with Relaxation for VI (1.3) For given v^k and $H \succ 0$, find w^{k+1},

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

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

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

2 从原始-对偶混合梯度法到按需定制的邻近点算法

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

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

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

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

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

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

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

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

where

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

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

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

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

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

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

Find the saddle point of the Lagrangian function is a special $\min - \max$ problem (2.1) whose $\theta_1(x) = \theta(x), \quad \theta_2(y) = -b^T y$ and $\mathcal{Y} = \Re^m$.

2.1 求解鞍点问题的 原始-对偶混合梯度法 PDHG [18]

For given (x^k, y^k), PDHG [18] produces a pair of (x^{k+1}, y^{k+1}). First,

$$x^{k+1} = \operatorname{argmin}\{\Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, \tag{2.4a}$$

and then we obtain y^{k+1} via

$$y^{k+1} = \operatorname{argmax}\{\Phi(x^{k+1}, y) - \frac{s}{2}\|y - y^k\|^2 \mid y \in \mathcal{Y}\}. \tag{2.4b}$$

Ignoring the constant term in the objective function, the subproblems (2.4) are reduced to

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

where

$$Q = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \qquad \text{is not symmetric.}$$

It does not be the PPA form (1.4), and we can not expect its convergence.

The following example of linear programming indicates the original PDHG (2.4) is not necessary convergent.

Consider a pair of the primal-dual linear programming:

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

We take the following example

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

where $A = [1, 1], \ b = 1, c = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and the vector $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

Note that its Lagrange function is

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

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

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

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

We use $(x_1^0, x_2^0; y^0) = (0, 0; 0)$ as the start point. For this example, the method is not convergent.

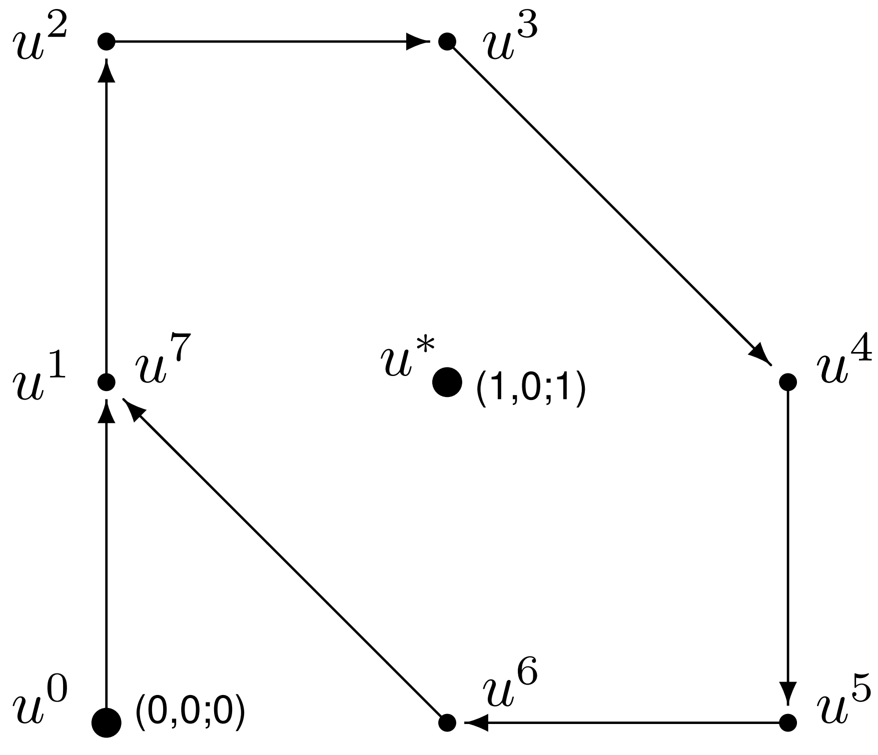
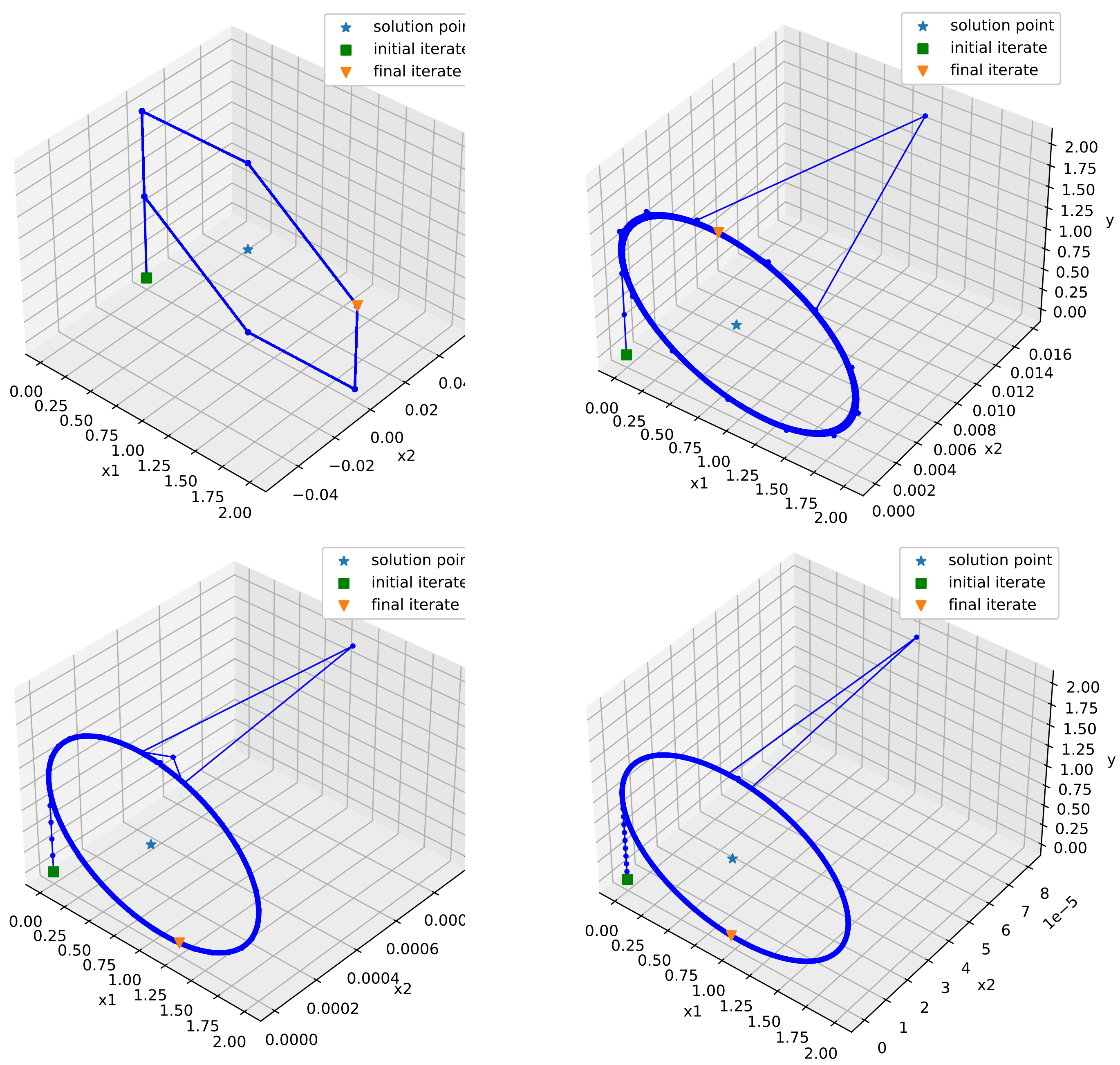


Fig. 2.1 The sequence generated by PDHG Method with $r = s = 1$

$$u^0 = (0,0;0)$$
$$u^1 = (0,0;1)$$
$$u^2 = (0,0;2)$$
$$u^3 = (1,0;2)$$
$$u^4 = (2,0;1)$$
$$u^5 = (2,0;0)$$
$$u^6 = (1,0;0)$$
$$u^7 = (0,0;1)$$
$$\boldsymbol{u^{k+6} = u^k}$$



对 $r = s = 1, 2, 5, 10$, PDHG 方法都不收敛

2.2 Customized Proximal Point Algorithm-Classical Version

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

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

then the variational inequality (2.9) will become the following desirable form:

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

For this purpose, we need only to change (2.8) in PDHG, namely,

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

to

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

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

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

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

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

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

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

The solution of (2.12b) is given by

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

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

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

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

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

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

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

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

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

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

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

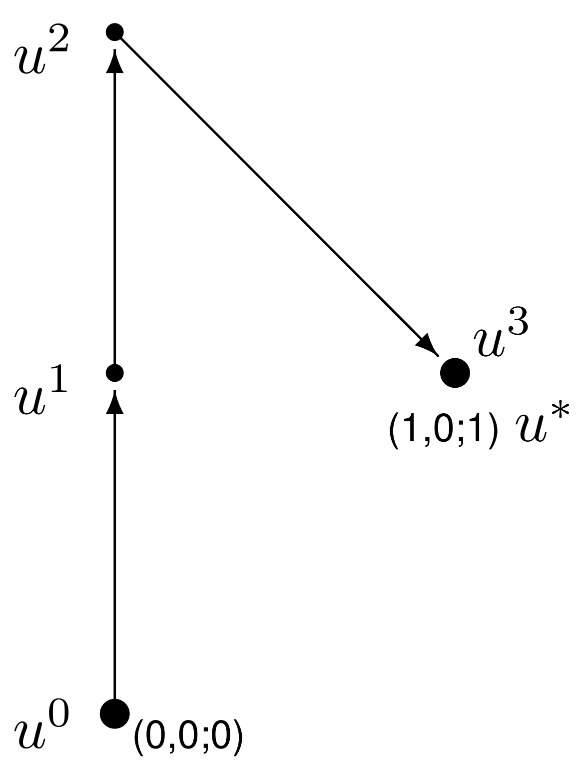
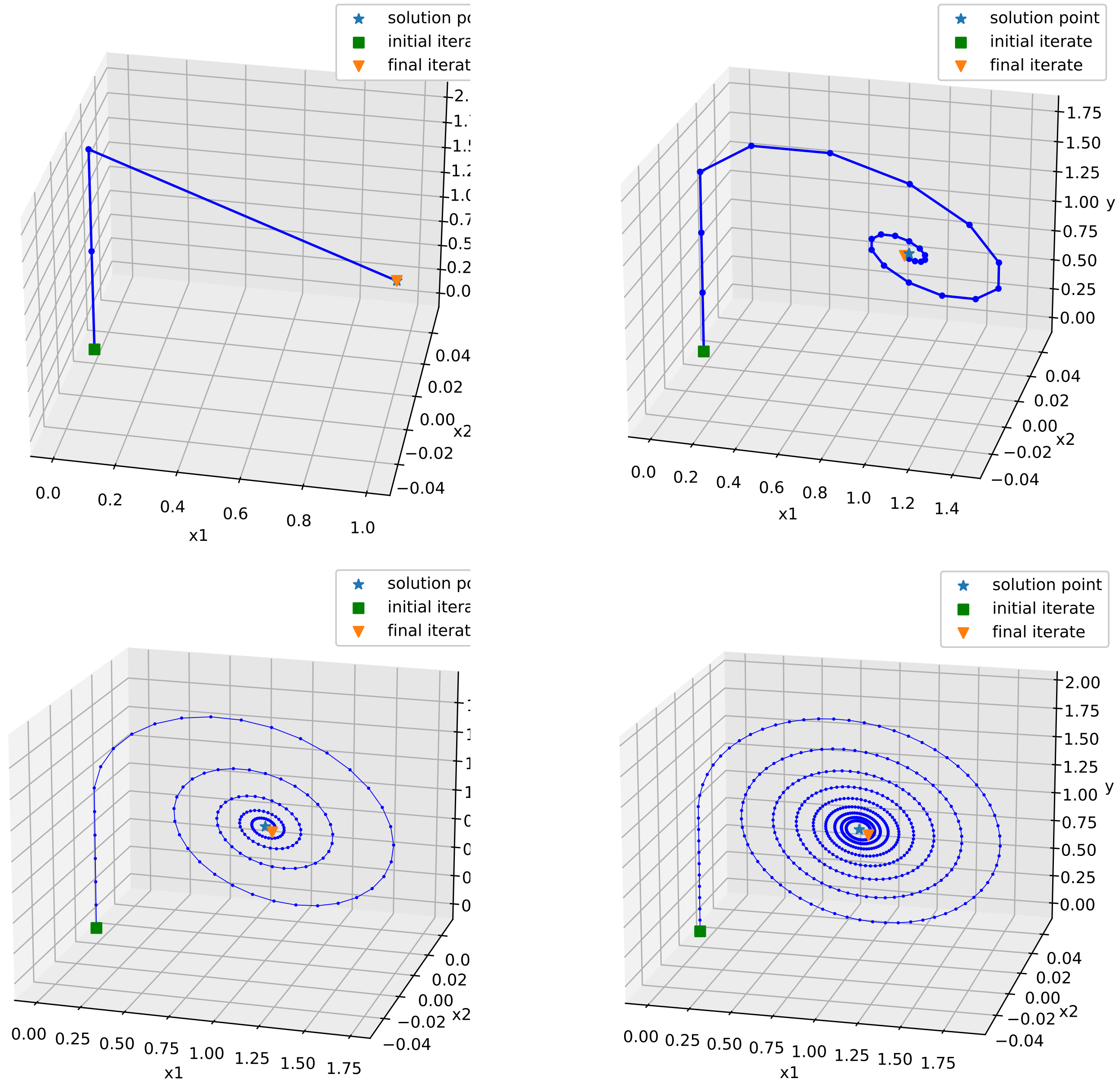


Fig. 2.2 The sequence generated by C-PPA Method with $r = s = 1$

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



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

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

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

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

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

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

where

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

Note that in the primal-dual order,

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

In the both cases, $rs > \|A^TA\|$, the matrix H is positive definite.

Remark We use CP-PPA to solve linearly constrained convex optimization.

If the equality constraints $Ax = b$ is changed to $Ax \geq b$, namely,

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

In this case, the Lagrange multiplier y should be nonnegative. $\Omega = \mathcal{X} \times \Re^m_+$.
We need only to make a slight change in the algorithms.

In the primal-dual order (2.12b), it needs to change the update dual update form

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

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

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

2.3 Simplicity recognition

Frame of VI is recognized by some Researcher in Image Science

Diagonal preconditioning for first order primal-dual algorithms in convex optimization*

Thomas Pock
Institute for Computer Graphics and Vision
Graz University of Technology
pock@icg.tugraz.at

Antonin Chambolle
CMAP & CNRS
École Polytechnique
antonin.chambolle@cmap.polytechnique.fr

- T. Pock and A. Chambolle, IEEE ICCV, 1762-1769, 2011
- A. Chambolle, T. Pock, A first-order primal-dual algorithms for convex problem with applications to imaging, J. Math. Imaging Vison, 40, 120-145, 2011.

preconditioned algorithm. In very recent work [10], it has been shown that the iterates (2) can be written in form of a proximal point algorithm [14], which greatly simplifies the convergence analysis.

From the optimality conditions of the iterates (4) and the convexity of G and F^* it follows that for any $(x, y) \in X \times Y$ the iterates x^{k+1} and y^{k+1} satisfy

$$\left\langle \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}, F\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} + M\begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} \right\rangle \geq 0, \quad (5)$$

where

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

and

$$M = \begin{bmatrix} \mathrm{T}^{-1} & -K^T \\ -\theta K & \Sigma^{-1} \end{bmatrix}. \quad (6)$$

It is easy to check, that the variational inequality (5) now takes the form of a proximal point algorithm [10, 14, 16].

作者 C-P 说到我们的 PPA 解释极大地简化了收敛性分析.

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

否则, 就像我们前面给出的例子, 方法是不一定收敛的.

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

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

[9] L. Ford and D. Fulkerson. *Flows in Networks*. Princeton University Press, Princeton, New Jersey, 1962.

[10] B. He and X. Yuan. Convergence analysis of primal-dual algorithms for total variation image restoration. Technical report, Nanjing University, China, 2010.

Later, the Reference [10] is published in SIAM J. Imaging Science [9].

Math. Program., Ser. A
DOI 10.1007/s10107-015-0957-3

CrossMark

FULL LENGTH PAPER

On the ergodic convergence rates of a first-order primal–dual algorithm

Antonin Chambolle[1] · Thomas Pock[2,3]

The paper published by Chambolle and Pock in Math. Progr. uses the VI framework

1 Introduction

In this work we revisit a first-order primal–dual algorithm which was introduced in [15, 26] and its accelerated variants which were studied in [5]. We derive new estimates for the rate of convergence. In particular, exploiting a proximal-point interpretation due to [16], we are able to give a very elementary proof of an ergodic $O(1/N)$ rate of convergence (where N is the number of iterations), which also generalizes to non-

Algorithm 1: $O(1/N)$ Non-linear primal–dual algorithm

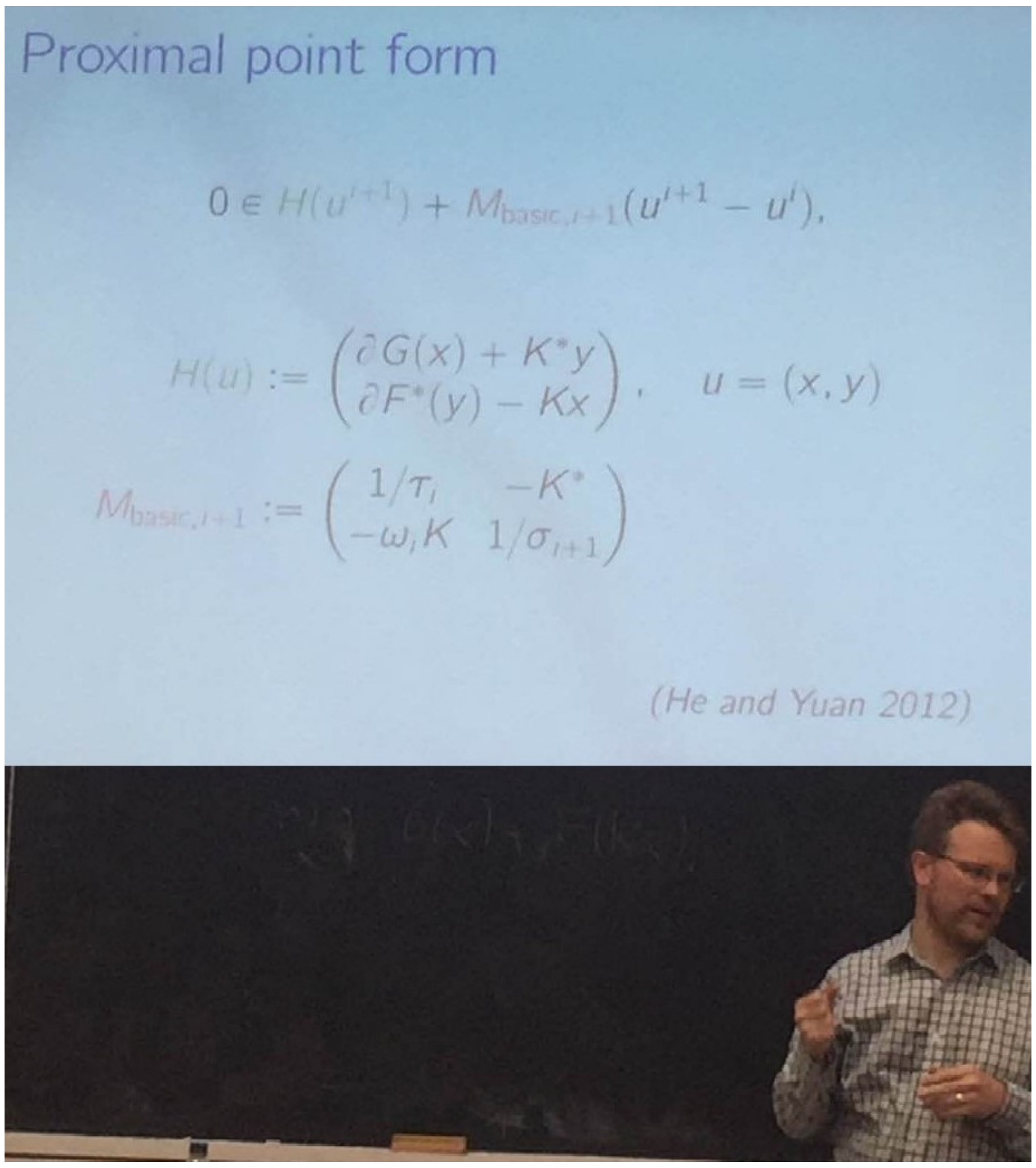
- Input: Operator norm $L := \|K\|$, Lipschitz constant L_f of ∇f, and Bregman distance functions D_x and D_y.
- Initialization: Choose $(x^0, y^0) \in \mathcal{X} \times \mathcal{Y}$, $\tau, \sigma > 0$
- Iterations: For each $n \geq 0$ let

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

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

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

B.S. He and X.M. Yuan, Convergence analysis of primal-dual algorithms for a saddle -point problem: From contraction perspective, *SIAM J. Imag. Science* **5**(2012), 119-149.



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

见到一幅幻灯片介绍我们的工作, 我的同事抢拍了一张照片发给我。

这也说明, 只有简单的思想才容易得到传播, 被人接受。

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

The Chen-Teboulle algorithm is the proximal point algorithm

Stephen Becker *

November 22, 2011; posted August 13, 2019

Abstract

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

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

1 Background

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

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

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

2.4 Relationship to Chambolle-Pock Method

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

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

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

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

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

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

当 $\tau = 1$ 并且 $rs > \|A^TA\|$, PPA 算法收敛. 当 $\tau = 0$, 方法不能保证收敛.
对 $\tau \in (0, 1)$, 收敛性没有证明, 至今还是一个 Open Problem.

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

CP 方法十年记　2020 年9 月

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

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

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

3 From ALM to Balanced ALM

We consider the generic convex minimization model with linear constraints

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

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

The Lagrangian function of (3.1) is

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

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

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

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

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

where

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

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

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

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

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

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

Then we have

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

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

3.1 Augmented Lagrangian Method

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

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

The method is implemented by

$$\begin{cases} x^{k+1} \in \arg\min\{\theta(x) - x^TA^T\lambda^k + \frac{r}{2}\|Ax-b\|^2 \mid x \in \mathcal{X}\}, & \text{(3.8a)} \\ \lambda^{k+1} = \lambda^k - r(Ax^{k+1}-b). & \text{(3.8b)} \end{cases}$$

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

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

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

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

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

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

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

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

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

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

3.2 CP-PPA method [9]

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

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

The method is implemented by

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

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

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

where

$$H = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}. \qquad \text{(3.14b)}$$

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

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

$$rs > \|A^TA\|. \qquad \text{(3.15)}$$

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

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

3.3 Balanced ALM [10]

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

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

Replaced

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

in (3.12b) by

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

The balanced ALM (3.16) is implemented by

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

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

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

where

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

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

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

$$w^{k+1}\in\Omega,\quad \theta(x)-\theta(x^{k+1}) + (w-w^{k+1})^TF(w^{k+1}) \ge (w-w^{k+1})^TH(w^k-w^{k+1}),\quad \forall\, w\in\Omega, \quad (3.20a)$$

where

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

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

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

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

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

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

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

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

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

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

Notice that the matrix H in

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

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

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

and therefore the matrix H is positive definite. □

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

4 ALM in PPA-sense

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

根据预设正定矩阵 构造 PPA 算法. 方法可以在 [19] 中查到.

The convex optimization problem,

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

is translated to the equivalent variational inequality :

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

where

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

4.1 Relaxed PPA in Primal-Dual Order

Relaxed PPA for the variational inequality (4.1) :

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

where

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

The concrete formula of (4.2) is

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

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

$$\left\{ \begin{array}{l} \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k} + \boldsymbol{(\beta A^T A + \delta I_n)}(\tilde{x}^k - x^k) + \boldsymbol{A^T}(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \\ \qquad (\underline{A\tilde{x}^k - b}) \quad + \boldsymbol{A}(\tilde{x}^k - x^k) \quad + \quad \boldsymbol{(1/\beta)}(\tilde{\lambda}^k - \lambda^k) = 0. \end{array} \right. \qquad \text{(4.3)}$$

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

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

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

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

Then, we use the form

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

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

4.2 Relaxed PPA in Dual-Primal Order

Relaxed PPA for the variational inequality (4.1) :

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

where

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

Then, we use the form

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

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

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

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

The concrete form of (4.4) is

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

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

Implementation of (4.4) is (Dual-Primal)

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

4.3 PPA in Primal-Dual Order

Relaxed PPA for the variational inequality (4.1) :

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

where

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

Then, we use the form

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

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

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

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

The concrete form of (4.6) is

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

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

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

Implementation

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

5 Different positive definite matrices H in PPA

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

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

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

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

可以根据问题的实际需要, 选择不同的正定矩阵 H

References

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

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

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

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

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

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

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

[8] B. S. He and X. M. Yuan, On the $O(1/t)$ convergence rate of the alternating direction method, *SIAM J. Numerical Analysis* **50**(2012), 700-709.

[9] B.S. He and X.M. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point

problem: From contraction perspective, *SIAM J. Imag. Science* **5**(2012), 119-149.

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

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

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

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

[14] M. J. D. Powell, A method for nonlinear constraints in minimization problems, in Optimization, R. Fletcher, ed., Academic Press, New York, NY, pp. 283-298, 1969.

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

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

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

[18] M. Zhu and T. F. Chan, An Efficient Primal-Dual Hybrid Gradient Algorithm for Total Variation Image Restoration, CAM Report 08-34, UCLA, Los Angeles, CA, 2008.

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

变分不等式框架下结构型凸优化的分裂收缩算法

III. 交替方向法 (ADMM) 和 PPA 意义下的 ADMM

中学的数理基础	必要的社会实践
普通的大学数学	一般的优化原理

何 炳 生　　南京大学数学系

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

天元数学东北中心　2023 年 10 月 17 – 27 日

1 Two blocks separable convex optimization

We consider the following separable convex optimization

$$\min\{\theta_1(x)+\theta_2(y)\,|\,Ax+By=b,\ x\in\mathcal{X},\ y\in\mathcal{Y}\} \tag{1.1}$$

Example: Best matrix approximation under some conditions

$$\min_X \{\tfrac{1}{2}\|X-C\|_F^2\,|\,X\in S_\Lambda^n\cap S_B\},$$

where

$$S_\Lambda^n=\{H\in\mathcal{S}^n\,|\,\lambda_{\min}I\preceq H\preceq\lambda_{\max}I\}$$

and

$$S_B=\{H\in\mathcal{S}^n\,|\,H_L\le H\le H_U\}.$$

It can be translated to the following equivalent problem:

$$\begin{array}{rl}\min_{X,Y} & \frac{1}{2}\|X-C\|^2+\frac{1}{2}\|Y-C\|^2\\ \text{s.t} & X-Y=0,\ X\in S_\Lambda^n,\ Y\in S_B.\end{array} \tag{1.2}$$

The problem (1.2) is a concrete problem of type (1.1).

Smooth Optimization Approach for Covariance Selection — Statistics

$$\min_X \{\mathbf{Tr}(CX) - \log(\det(X)) + \rho e^T|X|e \;\big|\; X \in S_+^n\}$$

where C is a given symmetric matrix, $e^T|X|e = \sum_{i=1}^n \sum_{j=1}^n |X_{ij}|$. Its equivalent optimization problem is

$$\begin{array}{rl} \min_{X,Y} & \mathbf{Tr}(CX) - \log(\det(X)) + \rho e^T|Y|e \\ \text{s.t} & X - Y = 0, \\ & X \in S_+^n, \; Y \in R^{n\times n}. \end{array}$$

Low rank and sparse optimization problem in statistics

$$\begin{array}{rl} \min_{X,Y} & \|X\|_* + \rho e^T|Y|e \\ \text{s.t} & X + Y = H \\ & X, \; Y \in R^{n\times n}. \end{array} \tag{1.3}$$

这些矩阵优化的数学模型本身就是一个形如 (1.1) 的结构型优化问题.

2 Mathematical Background

两大基本概念：变分不等式 和 邻近点 (PPA) 算法

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

$$x^* \in \arg\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\} \tag{2.1a}$$

if and only if

$$x^* \in \mathcal{X}, \;\; \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0, \;\; \forall x \in \mathcal{X}. \tag{2.1b}$$

2.1 Linearly constrained convex optimization and VI

The Lagrangian function of the problem (1.1) is

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

According to Lemma 1, the saddle point is a solution of the following variational inequality:

$$\begin{cases} x^* \in \mathcal{X}, & \theta_1(x) - \theta_1(x^*) + (x - x^*)^T(-A^T\lambda^*) \ge 0, \quad \forall\, x \in \mathcal{X}, \\ y^* \in \mathcal{Y}, & \theta_2(y) - \theta_2(y^*) + (y - y^*)^T(-B^T\lambda^*) \ge 0, \quad \forall\, y \in \mathcal{Y}, \\ \lambda^* \in \Re^m, & (\lambda - \lambda^*)^T(Ax^* + By^* - b) \ge 0, \quad \forall\, \lambda \in \Re^m. \end{cases}$$

Its compact form is the following variational inequality:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{2.2}$$

where

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

and

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

Note that the operator F is monotone, because

$$(w - \tilde{w})^T(F(w) - F(\tilde{w})) \ge 0, \ \text{Here } (w - \tilde{w})^T(F(w) - F(\tilde{w})) = 0. \tag{2.3}$$

2.2 Preliminaries of PPA for Variational Inequalities

The optimal condition of the problem (1.1) is characterized as a mixed monotone variational inequality:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega. \tag{2.4}$$

PPA for monotone mixed VI in H-norm

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

$$w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T \{F(w^{k+1}) + H(w^{k+1} - w^k)\} \geq 0, \ \forall\, w \in \Omega, \qquad (2.5)$$

where H is a symmetric positive definite matrix.

✠ Again, w^k is the solution of (2.4) if and only if $w^k = w^{k+1}$ ✠

Convergence Property of Proximal Point Algorithm in H-norm

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2. \qquad (2.6)$$

The sequence $\{w^k\}$ is Fejér monotone in H-norm. In customized PPA, via choosing a proper positive definite matrix H, the solution of the subproblem (2.5) has a closed form. An iterative algorithm is called the contraction method, if its generated sequence $\{w^k\}$ satisfies $\|w^{k+1} - w^*\|_H^2 < \|w^k - w^*\|_H^2$.

2.3 Augmented Lagrangian Method (ALM)

We consider the convex optimization, namely

$$\min\{\theta(u) \,|\, \mathcal{A}u = b,\, u \in \mathcal{U}\}. \qquad (2.7)$$

The related variational inequality of the saddle point of the Lagrangian function is

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \qquad (2.8a)$$

where

$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{U} \times \Re^m. \qquad (2.8b)$$

Augmented Lagrangian Method

The augmented Lagrangian function of the problem (2.7) is

$$\mathcal{L}_\beta(u, \lambda) = \theta(u) - \lambda^T(\mathcal{A}u - b) + \frac{\beta}{2}\|\mathcal{A}u - b\|^2,$$

The k-th iteration of the **Augmented Lagrangian Method** [13, 16] begins with a given λ^k, obtain $w^{k+1} = (u^{k+1}, \lambda^{k+1})$ via

$$\text{(ALM)} \quad \left\{ \begin{array}{l} u^{k+1} = \arg\min\{\mathcal{L}_\beta(u, \lambda^k) \mid u \in \mathcal{U}\}, \qquad (2.9a) \\ \lambda^{k+1} = \lambda^k - \beta(\mathcal{A}u^{k+1} - b). \qquad (2.9b) \end{array} \right.$$

In (2.9), u^{k+1} is only a computational result of (2.9a) from given λ^k, it is called the intermediate variable. In order to start the k-th iteration of ALM, we need only to have λ^k and thus we call it as the essential variable.

The subproblem (2.9a) is a problem of mathematical form

$$\min\{\theta(u) + \tfrac{\beta}{2}\|\mathcal{A}u - p^k\|^2 \mid u \in \mathcal{U}\} \qquad (2.10)$$

where $\beta > 0$ is a given scalar and $p^k = b + \frac{1}{\beta}\lambda^k$.
Assumption: The solution of problem (2.10) has closed-form solution or can be efficiently computed with a high precision.

The optimal condition of (2.9) can be written as $w^{k+1} \in \Omega = \mathcal{U} \times \Re^m$ and

$$\left\{ \begin{array}{r} \theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T\{-\mathcal{A}^T\lambda^k + \beta\mathcal{A}^T(\mathcal{A}u^{k+1} - b)\} \geq 0, \ \forall u \in \mathcal{U}, \\ (\lambda - \lambda^{k+1})^T\{(\mathcal{A}u^{k+1} - b) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} \geq 0, \ \ \forall \lambda \in \Re^m. \end{array} \right.$$

The above relations can be written as

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} u - u^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} -\mathcal{A}^T\lambda^{k+1} \\ \mathcal{A}u^{k+1} - b \end{pmatrix} \geq (\lambda - \lambda^{k+1})^T \frac{1}{\beta}(\lambda^k - \lambda^{k+1}),$$

for all $w \in \Omega$. Using the notations in (2.8), we get the compact form

$$\begin{array}{rl} \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) & \\ \geq & (\lambda - \lambda^{k+1})^T \dfrac{1}{\beta}(\lambda^k - \lambda^{k+1}), \ \forall w \in \Omega. \end{array} \qquad (2.11)$$

Setting $w = w^*$ in (2.11), we get

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

By using the monotonicity of F and the optimality of w^*, it follows that

$$\begin{aligned} & \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}) \\ = \quad & \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^*) \geq 0. \end{aligned}$$

Thus, we have

$$(\lambda^{k+1} - \lambda^*)^T(\lambda^k - \lambda^{k+1}) \geq 0. \tag{2.12}$$

By using the above inequality, we obtain

$$\begin{aligned} \|\lambda^k - \lambda^*\|^2 &= \|(\lambda^{k+1} - \lambda^*) + (\lambda^k - \lambda^{k+1})\|^2 \\ &\geq \|\lambda^{k+1} - \lambda^*\|^2 + \|\lambda^k - \lambda^{k+1}\|^2. \end{aligned}$$

It means that

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

The above inequality is the key for the convergence proof of the Augmented Lagrangian Method.

3 ADMM for two-block problems

Recall the separable convex optimization problem

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

The augmented Lagrangian function

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

Applied ALM to solve the problem (1.1), the k-th iteration begins with given λ^k,

$$\begin{cases} (x^{k+1}, y^{k+1}) = \arg\min\{\mathcal{L}_\beta(x, y, \lambda^k) \mid x \in \mathcal{X}, y \in \mathcal{Y}\}, & \text{(3.1a)} \\ \lambda^{k+1} \in \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & \text{(3.1b)} \end{cases}$$

ADMM is a relaxed ALM for the problem (1.1),the k-th iteration begins with given (y^k, λ^k),

$$\begin{cases} x^{k+1} \in \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, & \text{(3.2a)} \\ y^{k+1} \in \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \mid y \in \mathcal{Y}\}, & \text{(3.2b)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & \text{(3.2c)} \end{cases}$$

两个可分离目标函数问题的 ADMM 方法 [3, 5]

Applied ADMM to the structured COP: $(y^k, \lambda^k) \Rightarrow (y^{k+1}, \lambda^{k+1})$

First, for given (y^k, λ^k), x^{k+1} is the solution of the following problem

$$x^{k+1} \in \text{Argmin}\left\{ \begin{array}{c} \theta_1(x) - (\lambda^k)^T(Ax + By^k - b) \\ +\frac{\beta}{2}\|Ax + By^k - b\|^2 \end{array} \middle| x \in \mathcal{X} \right\} \quad (3.3a)$$

Use λ^k and the obtained x^{k+1}, y^{k+1} is the solution of the following problem

$$y^{k+1} \in \text{Argmin}\left\{ \begin{array}{c} \theta_2(y) - (\lambda^k)^T(Ax^{k+1} + By - b) \\ +\frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \end{array} \middle| y \in \mathcal{Y} \right\} \quad (3.3b)$$

$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \quad (3.3c)$$

Advantages The x and y sub-problems are separately solved one by one.

Remark Ignoring the constant term in the objective function, the sub-problems (3.22a) and (3.22b) is equivalent to

$$x^{k+1} \in \text{Argmin}\{\theta_1(x) + \tfrac{\beta}{2}\|(Ax + By^k - b) - \tfrac{1}{\beta}\lambda^k\|^2 | x \in \mathcal{X}\} \quad (3.4a)$$

and

$$y^{k+1} \in \text{Argmin}\{\theta_2(y) + \tfrac{\beta}{2}\|(Ax^{k+1} + By - b) - \tfrac{1}{\beta}\lambda^k\|^2 | y \in \mathcal{Y}\} \quad (3.4b)$$

respectively. Note that the equation (3.3c) can be written as

$$(\lambda - \lambda^{k+1})\{(Ax^{k+1} + By^{k+1} - b) + \tfrac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} \ge 0, \ \forall \lambda \in \Re^m. \quad (3.4c)$$

Notice that the sub-problems (3.4a) and (3.4b) are the type of

$$x^{k+1} \in \text{Argmin}\{\theta_1(x) + \tfrac{\beta}{2}\|Ax - p^k\|^2 | x \in \mathcal{X}\}$$

and

$$y^{k+1} \in \text{Argmin}\{\theta_2(y) + \tfrac{\beta}{2}\|By - q^k\|^2 | y \in \mathcal{Y}\},$$

respectively. (子问题求解有困难怎么处理放在后面讲)

Analysis According to Theorem 1, the solution of (3.22a) and (3.22b) satisfies

$$x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \\ \{-A^T\lambda^k + \beta A^T\big(Ax^{k+1} + By^k - b\big)\} \ge 0, \ \forall x \in \mathcal{X} \tag{3.5a}$$

and

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \\ \{-B^T\lambda^k + \beta B^T\big(Ax^{k+1} + By^{k+1} - b\big)\} \ge 0, \ \forall y \in \mathcal{Y}, \tag{3.5b}$$

respectively. Substituting $\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b)$ (see (3.3c)) in (3.5) (eliminating λ^k in (3.5)), we get

$$x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \\ \{-A^T\lambda^{k+1} + \beta A^T B(y^k - y^{k+1})\} \ge 0, \ \forall x \in \mathcal{X}, \tag{3.6a}$$

and

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^{k+1}\} \ge 0, \ \forall y \in \mathcal{Y}. \tag{3.6b}$$

The compact form of (3.6) is $u^{k+1} = (x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y}$ and

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \Bigg\{ \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \end{pmatrix} \\ + \beta \begin{pmatrix} A^T B \\ 0 \end{pmatrix} \big(y^k - y^{k+1}\big) \Bigg\} \ge 0, \tag{3.7}$$

for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

By adding and subtracting the term $\beta B^T B(y^k - y^{k+1})$, we rewrite the about variational inequality in our desirable form

$$\theta(u)-\theta(u^{k+1})+\begin{pmatrix} x-x^{k+1} \\ y-y^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \end{pmatrix} + \beta \begin{pmatrix} A^TB \\ B^TB \end{pmatrix} (y^k - y^{k+1}) + \begin{pmatrix} 0 & 0 \\ 0 & \beta B^TB \end{pmatrix} \begin{pmatrix} x^{k+1}-x^k \\ y^{k+1}-y^k \end{pmatrix} \right\} \ge 0, \ \forall (x,y) \in \mathcal{X}\times\mathcal{Y}.$$

Combining the last inequality with (3.4c), we have the following lemma.

引理 1 *Let* $w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})\in\Omega$ *be generated by* (3.3) *with given* (y^k,λ^k)*, then we have*

$$\theta(u)-\theta(u^{k+1})+\begin{pmatrix} x-x^{k+1} \\ y-y^{k+1} \\ \lambda-\lambda^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \\ Ax^{k+1}+By^{k+1}-b \end{pmatrix} + \beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k - y^{k+1}) + \begin{pmatrix} 0 & 0 \\ \beta B^TB & 0 \\ 0 & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} y^{k+1}-y^k \\ \lambda^{k+1}-\lambda^k \end{pmatrix} \right\} \ge 0, \ \forall w\in\Omega. \quad (3.8)$$

For convenience we use the notations

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

Then, we get the following lemma:

引理 2 *Let the sequence* $\{w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})\}\in\Omega$ *be generated by* (3.3)*. Then, we have*

$$(v^{k+1}-v^*)^T H(v^k-v^{k+1}) \ge (y^k-y^{k+1})^T B^T(\lambda^k-\lambda^{k+1}), \ \forall w^*\in\Omega^*, \quad (3.9)$$

where

$$H=\begin{pmatrix} \beta B^TB & 0 \\ 0 & \frac{1}{\beta}I_m \end{pmatrix}. \quad (3.10)$$

Proof. Setting $w = w^*$ in (3.8), we get

$$\begin{aligned}
&(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \\
&\geq \quad \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} \beta B(y^k - y^{k+1}) \\
&\qquad +\theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}), \ \forall\, w^* \in \Omega^*. \quad (3.11)
\end{aligned}$$

Observe the first part of the right hand side of (3.11),

$$\begin{aligned}
&\begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} \beta B(y^k - y^{k+1}) \\
&= \quad (y^k - y^{k+1})^T B^T \beta \big(A, B\big) \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix} \\
&= \quad (y^k - y^{k+1})^T B^T \beta \big(Ax^{k+1} + By^{k+1} - (Ax^* + By^*)\big) \\
&= \quad (y^k - y^{k+1}) B^T \underline{\beta \big(Ax^{k+1} + By^{k+1} - b\big)} \\
&= \quad (y^k - y^{k+1}) B^T \underline{(\lambda^k - \lambda^{k+1})}. \qquad (3.12)
\end{aligned}$$

To the second part, since $(w^{k+1} - w^*)^T F(w^{k+1}) = (w^{k+1} - w^*)^T F(w^*)$ and w^* is the optimal solution, it follows that

$$\begin{aligned}
&\theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}) \\
&\quad = \quad \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^*) \geq 0. \qquad (3.13)
\end{aligned}$$

The assertion (3.11) immediately. □

引理 3 *Let the sequence* $\{w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})\} \in \Omega$ *be generated by* (3.3). *Then, we have*

$$(y^k - y^{k+1})^T B^T (\lambda^k - \lambda^{k+1}) \geq 0. \qquad (3.14)$$

Proof. Because (3.6b) is true for the k-th iteration and the previous iteration, we have

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T \lambda^{k+1}\} \geq 0, \ \forall\, y \in \mathcal{Y}, \qquad (3.15)$$

and

$$\theta_2(y) - \theta_2(y^k) + (y - y^k)^T\{-B^T\lambda^k\} \geq 0, \ \forall\, y \in \mathcal{Y}, \qquad (3.16)$$

Setting $y = y^k$ in (3.15) and $y = y^{k+1}$ in (3.16), respectively, and then adding the two resulting inequalities, we get the assertion (3.14) immediately. □

Substituting (3.14) in (3.9), we get

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \geq 0, \quad \forall\, v^* \in \mathcal{V}^*. \qquad (3.17)$$

Using the above inequality, as in the last lecture, we have the following theorem, which is the key for the proof of the convergence of ADMM.

定理 1 *Let the sequence* $\{w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})\} \in \Omega$ *be generated by* (3.3). *Then, we have*

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \ \forall\, v^* \in \mathcal{V}^*. \qquad (3.18)$$

交替方向法收敛性证明的　再阐述

交替方向法处理的是两个可分离块的凸优化问题

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b, \ x \in \mathcal{X}, y \in \mathcal{Y}\}. \qquad (3.19)$$

将其拉格朗日函数 $L(x,y,\lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax + By - b)$ 的鞍点归结为等价的变分不等式的解点：

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \qquad (3.20a)$$

其中

$$w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}, \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m. \qquad (3.20b)$$

ADMM 的 k 步迭代从给定的核心变量 $v^k = (y^k, \lambda^k)$ 出发

$$\begin{cases} x^{k+1} = \arg\min\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\}, & (3.21a) \\ y^{k+1} = \arg\min\{\theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \mid y \in \mathcal{Y}\}, & (3.21b) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & (3.21c) \end{cases}$$

根据最优性引理 1, ADMM k-步迭代满足

$$\begin{cases} x^{k+1}\in\mathcal{X}, & \theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\{-A^T\lambda^k+\beta A^T(Ax^{k+1}+By^k-b)\}\ge 0,\ \forall x\in\mathcal{X},\\ y^{k+1}\in\mathcal{Y}, & \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\{-B^T\lambda^k+\beta B^T(Ax^{k+1}+By^{k+1}-b)\}\ge 0,\ \forall y\in\mathcal{Y},\\ \lambda^{k+1}\in\Re^m, & (\lambda-\lambda^{k+1})^T\{(Ax^{k+1}+By^{k+1}-b)+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)\}\ge 0,\ \forall\lambda\in\Re^m. \end{cases}$$

利用 $\lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}-b)$ 上面的式子可以整理改写成

$$\begin{cases} x^{k+1}\in\mathcal{X}, & \theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\{-A^T\lambda^{k+1}+\beta A^TB(y^k-y^{k+1})\}\ge 0,\ \forall x\in\mathcal{X}, & (3.22a)\\ y^{k+1}\in\mathcal{Y}, & \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\{-B^T\lambda^{k+1}\}\ge 0,\ \forall y\in\mathcal{Y}, & (3.22b)\\ \lambda^{k+1}\in\Re^m, & (\lambda-\lambda^{k+1})^T\{(Ax^{k+1}+By^{k+1}-b)+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)\}\ge 0,\ \forall\lambda\in\Re^m. & (3.22c) \end{cases}$$

在 (3.22b) 的后半部加上和为零的两项, 得到

$$\begin{cases} \underline{\theta_1(x)-\theta_1(x^{k+1})}+(x-x^{k+1})^T\{-A^T\lambda^{k+1}+\beta A^TB(y^k-y^{k+1})\}\ge 0,\\ \underline{\theta_2(y)-\theta_2(y^{k+1})}+(y-y^{k+1})^T\{-B^T\lambda^{k+1}+\underbrace{\beta B^TB(y^k-y^{k+1})+\beta B^TB(y^{k+1}-y^k)}\}\ge 0,\\ \qquad (\lambda-\lambda^{k+1})^T\{(Ax^{k+1}+By^{k+1}-b)+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)\}\ge 0. \end{cases}$$

利用变分不等式 (3.20), 进行合理整合, 得到

$$\underline{\theta(u)-\theta(u^{k+1})}+(w-w^{k+1})^TF(w^{k+1})$$
$$+\begin{pmatrix} x-x^{k+1}\\ y-y^{k+1}\end{pmatrix}^T\beta\begin{pmatrix} A^T\\ B^T\end{pmatrix}B(y^k-y^{k+1})+\begin{pmatrix} y-y^{k+1}\\ \lambda-\lambda^{k+1}\end{pmatrix}^T\begin{pmatrix}\beta B^TB & 0\\ 0 & \frac{1}{\beta}I_m\end{pmatrix}\begin{pmatrix} y^{k+1}-y^k\\ \lambda^{k+1}-\lambda^k\end{pmatrix}\ge 0.$$

将上式中那个任意的 w, 设成解点 w^* 便有

$$\underline{\theta(u^*)-\theta(u^{k+1})}+(w^*-w^{k+1})^TF(w^{k+1})$$
$$+\begin{pmatrix} x^*-x^{k+1}\\ y^*-y^{k+1}\end{pmatrix}^T\beta\begin{pmatrix} A^T\\ B^T\end{pmatrix}B(y^k-y^{k+1})+\begin{pmatrix} y^*-y^{k+1}\\ \lambda^*-\lambda^{k+1}\end{pmatrix}^T\begin{pmatrix}\beta B^TB & 0\\ 0 & \frac{1}{\beta}I_m\end{pmatrix}\begin{pmatrix} y^{k+1}-y^k\\ \lambda^{k+1}-\lambda^k\end{pmatrix}\ge 0.$$

经转换, 得到

$$\begin{pmatrix} y^{k+1}-y^*\\ \lambda^{k+1}-\lambda^*\end{pmatrix}^T\begin{pmatrix}\beta B^TB & 0\\ 0 & \frac{1}{\beta}I_m\end{pmatrix}\begin{pmatrix} y^k-y^{k+1}\\ \lambda^k-\lambda^{k+1}\end{pmatrix} \quad \text{后面记}\ v=\begin{pmatrix} y\\ \lambda\end{pmatrix},\ H=\begin{pmatrix}\beta B^TB & 0\\ 0 & \frac{1}{\beta}I_m\end{pmatrix}$$
$$\ge\begin{pmatrix} x^{k+1}-x^*\\ y^{k+1}-y^*\end{pmatrix}^T\beta\begin{pmatrix} A^T\\ B^T\end{pmatrix}B(y^k-y^{k+1})+\underbrace{[\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^TF(w^{k+1})]}. \quad (3.23)$$

假如 (3.23) 式右端非负, 证明就基本上完成了. 由于

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

(3.23) 式右端下划线部分非负. 因此从 (3.23) 式得到

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \;\geq\; \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}). \tag{3.24}$$

对 (3.24) 式的右端进行处理, 有

$$\begin{aligned} & \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}) = (y^k - y^{k+1})^T B^T \beta (A, B) \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix} \\ = \;& (y^k - y^{k+1})^T B^T \beta (Ax^{k+1} + By^{k+1} - (Ax^* + By^*)) \quad \text{利用}(Ax^* + By^* = b) \\ = \;& (y^k - y^{k+1}) B^T \beta (Ax^{k+1} + By^{k+1} - b) \\ = \;& (y^k - y^{k+1}) B^T (\lambda^k - \lambda^{k+1}). \end{aligned} \tag{3.25}$$

后面我们要证明 $(y^k - y^{k+1}) B^T (\lambda^k - \lambda^{k+1}) \geq 0.$

利用 (3.22b) 有

$$\begin{aligned} \theta_2(y) - \theta_2(y^{k+1}) &+ (y - y^{k+1})^T \{-B^T \lambda^{k+1}\} \geq 0, \ \forall\, y \in \mathcal{Y}, \\ \text{和}\ \theta_2(y) - \theta_2(y^k) &+ (y - y^k)^T \{-B^T \lambda^k\} \geq 0, \ \forall\, y \in \mathcal{Y}. \end{aligned}$$

(将任意的 y 分别设成 y^k 和 y^{k+1})

$$\begin{aligned} \theta_2(y^k) - \theta_2(y^{k+1}) &+ (y^k - y^{k+1})^T \{-B^T \lambda^{k+1}\} \geq 0. \\ \theta_2(y^{k+1}) - \theta_2(y^k) &+ (y^{k+1} - y^k)^T \{-B^T \lambda^k\} \geq 0. \end{aligned}$$

(将上面两式相加, 就有) $(y^k - y^{k+1}) B^T (\lambda^k - \lambda^{k+1}) \geq 0.$ ((3.25) 式右端非负)

证明了 (3.25) 式右端非负, 进而得到 (3.24) 式右端非负. 所以

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \geq 0. \tag{3.26}$$

Lemma 2 告诉我们:

$$b^T H(a - b) \geq 0 \quad \Rightarrow \quad \|b\|_H^2 \leq \|a\|_H^2 - \|a - b\|_H^2. \tag{3.27}$$

在 (3.27) 中置 $a = (v^k - v^*)$ 和 $b = (v^{k+1} - v^*)$, 根据 (3.26) 就得到收敛的关键不等式

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

由 $\|v^k - v^{k+1}\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2$ 得 $\sum_{k=0}^{\infty} \|v^k - v^{k+1}\|_H^2 \leq \|v^0 - v^*\|_H^2.$

How to choose the parameter β. The efficiency of ADMM is heavily dependent on the parameter β in (3.3). We discuss how to choose a suitable β in the practical computation.

Note that if $\boldsymbol{\beta A^T B(y^k - y^{k+1}) = 0}$, then it follows from (3.7)

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \end{pmatrix} \geq 0, \ \forall (x,y) \in \mathcal{X} \times \mathcal{Y}.$$

In this case, if additionally $\boldsymbol{Ax^{k+1} + By^{k+1} - b = 0}$, then we have

$$\begin{cases} \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T(-A^T\lambda^{k+1}) \geq 0, & \forall x \in \mathcal{X} \\ \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T(-B^T\lambda^{k+1}) \geq 0, & \forall y \in \mathcal{Y} \\ (\lambda - \lambda^{k+1})^T(Ax^{k+1} + By^{k+1} - b) \geq 0, & \forall \lambda \in \Re^m \end{cases}$$

and consequently $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ is a solution of the VI (2.2).

In other words, $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ is not a solution of (2.2) because

$$\beta A^T B(y^k - y^{k+1}) \neq 0 \quad \text{and/or} \quad Ax^{k+1} + By^{k+1} - b \neq 0.$$

We call

$$\boldsymbol{\|\beta A^T B(y^k - y^{k+1})\|} \quad \text{and} \quad \boldsymbol{\|Ax^{k+1} + By^{k+1} - b\|}$$

the primal-residual and the dual-residual, respectively. It seems that we should balance the primal and the dual residuals dynamically. If

$$\mu\|\beta A^T B(y^k - y^{k+1})\| < \|Ax^{k+1} + By^{k+1} - b\| \quad \text{with a} \quad \mu > 1,$$

it means that the dual residual is too large and thus we should enlarge the parameter β in the augmented Lagrangian function. Otherwise, we should reduce the parameter β. A simple scheme that often works well is (see, e.g., [9]):

$$\beta_{k+1} = \begin{cases} \beta_k * \tau, & \text{if } \mu\|\beta A^T B(y^k - y^{k+1})\| < \|Ax^{k+1} + By^{k+1} - b\|; \\ \beta_k/\tau, & \text{if } \|\beta A^T B(y^k - y^{k+1})\| > \mu\|Ax^{k+1} + By^{k+1} - b\|; \\ \beta_k, & \text{otherwise.} \end{cases}$$

where $\mu > 1, \tau > 1$ are parameters. Typical choices might be $\mu = 10$ and $\tau = 2$. 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. This self adaptive adjusting rule has been used by S. Boyd's group [1] and in their Optimization Solver [6].

4 Linearized ADMM

The augmented Lagrangian Function of the problem (1.1) is

$$\mathcal{L}_{\beta}^{[2]}(x,y,\lambda)=\theta_1(x)+\theta_2(y)-\lambda^T(Ax+By-b)+\frac{\beta}{2}\|Ax+By-b\|^2. \quad (4.1)$$

Solving the problem (1.1) by using ADMM, the k-th iteration begins with given (y^k,λ^k), it offers the new iterate (y^{k+1},λ^{k+1}) via

$$\text{(ADMM)}\quad \begin{cases} x^{k+1}=\arg\min\big\{\mathcal{L}_{\beta}^{[2]}(x,y^k,\lambda^k)\ \big|\ x\in\mathcal{X}\big\}, & (4.2a)\\ y^{k+1}=\arg\min\big\{\mathcal{L}_{\beta}^{[2]}(x^{k+1},y,\lambda^k)\ \big|\ y\in\mathcal{Y}\big\}, & (4.2b)\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}-b). & (4.2c)\end{cases}$$

In optimization problem, the solution is invariant by changing the constant term in the objective function. Thus, by using the augmented Lagrangian function,

$$\begin{aligned} y^{k+1} &= \arg\min\big\{\mathcal{L}_{\beta}^{[2]}(x^{k+1},y,\lambda^k)\ \big|\ y\in\mathcal{Y}\big\}\\ &= \arg\min\big\{\theta_2(y)-y^TB^T\lambda^k+\frac{\beta}{2}\|Ax^{k+1}+By-b\|^2\ \big|\ y\in\mathcal{Y}\big\}.\end{aligned}$$

Thus, by denoting $q^k=b-Ax^{k+1}+\frac{1}{\beta}\lambda^k$, the solution of (3.22b) is given by

$$\min\{\theta_2(y)+\frac{\beta}{2}\|By-q^k\|^2\,|\,y\in\mathcal{Y}\}. \quad (4.3)$$

In some practical applications, because of the structure of the matrix B, the subproblem (4.3) is not so easy to be solved. In this case, it is necessary to use the linearized version of the ADMM.

Notice that the Taylor expansion of the quadratic term of (4.2b), namely,

$$\begin{aligned} \frac{\beta}{2}\|Ax^{k+1}+By-b\|^2 &= \frac{\beta}{2}\|B(y-y^k)+(Ax^{k+1}+By^k-b)\|^2\\ = \ & \frac{\beta}{2}\|B(y-y^k)\|^2+\beta(y-y^k)^TB^T(Ax^{k+1}+By^k-b)\\ &+\frac{\beta}{2}\|Ax^{k+1}+By^k-b\|^2\end{aligned}$$

Changing the constant term in the objective function of (4.2b) accordingly, we have

$$
\begin{aligned}
y^{k+1} &= \arg\min\{\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^k) \mid y \in \mathcal{Y}\} \\
&= \arg\min\{\theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \mid y \in \mathcal{Y}\} \\
&= \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T \lambda^k + \beta y^T B^T (Ax^{k+1} + By^k - b) \\ + \frac{\beta}{2}\|B(y - y^k)\|^2 \end{array} \middle| \, y \in \mathcal{Y} \right\}.
\end{aligned}
$$

So-called linearized version of ADMM, we remove the unfavorable term $\dfrac{\beta}{2}\|B(y - y^k)\|^2$ in the objective function, and add the term $\dfrac{s}{2}\|y - y^k\|^2$.

Strictly speaking, it should be a "linearization" plus "regularization" method. It can also be interpreted as:

The term $\dfrac{\beta}{2}\|B(y - y^k)\|^2$ is replaced with $\dfrac{s}{2}\|y - y^k\|^2$.

In other words, it is equivalent to adding the term

$$
\frac{1}{2}\|y - y^k\|_{D_B}^2 \qquad (\text{with } D_B = sI_{n_2} - \beta B^T B) \tag{4.4}
$$

to the objective function of (4.2b), we get

$$
\begin{aligned}
y^{k+1} &= \arg\min\{\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^k) + \frac{1}{2}\|y - y^k\|_{D_B}^2 \mid y \in \mathcal{Y}\} \\
&= \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T \lambda^k + \beta y^T B^T (Ax^{k+1} + By^k - b) \\ + \frac{s}{2}\|y - y^k\|^2 \end{array} \middle| \, y \in \mathcal{Y} \right\} \\
&= \arg\min\left\{\theta_2(y) + \frac{s}{2}\left\|y - d^k\right\|^2 \mid y \in \mathcal{Y}\right\},
\end{aligned} \tag{4.5}
$$

where

$$
d^k = y^k - \frac{1}{s}B^T\left[\beta(Ax^{k+1} + By^k - b) - \lambda^k\right].
$$

By using such strategy, the sub-problems of ADMM is simplified. The linearized version of ADMM are applied successfully in scientific computing [14, 17, 18, 19]. The following analysis is based on the fact that the sub-problems (3.22a) and

$$
\min\{\theta_2(y) + \frac{s}{2}\|y - d^k\|^2 \mid y \in \mathcal{Y}\}
$$

are easy to be solved.

Linearized ADMM. For solving the problem (1.1), the k-th iteration of the linearized ADMM begins with given $v^k = (y^k, \lambda^k)$, produces the $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$

via the following procedure:

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta^{[2]}(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, & (4.6a) \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^k) + \frac{1}{2}\|y - y^k\|_{D_B}^2 \mid y \in \mathcal{Y}\}, & (4.6b) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & (4.6c) \end{cases}$$

where D_B is defined by (4.4).

First, using the optimality of the sub-problems of (4.6), we prove the following lemma as the base of convergence.

引理 4 *Let* $\{w^k\}$ *be the sequence generated by Linearized ADMM* (4.6) *for the problem* (1.1). *Then, we have*

$$\begin{aligned} w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) &+ (w - w^{k+1})^T F(w) \\ &+ \beta(x - x^{k+1})^T A^T (By^k - By^{k+1}) \\ &\geq (y - y^{k+1})^T D_B (y^k - y^{k+1}) \\ &+ \frac{1}{\beta}(\lambda - \lambda^{k+1})^T(\lambda^k - \lambda^{k+1}), \quad \forall w \in \Omega. \end{aligned} \tag{4.7}$$

Proof. For the x-subproblem in (4.6a), by using Lemma 1, we have

$$\begin{aligned} x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) & \\ + (x - x^{k+1})^T\{-A^T\lambda^k + \beta A^T(Ax^{k+1} + By^k - b)\} & \\ \geq 0, \quad \forall x \in \mathcal{X}. & \end{aligned}$$

By using the multipliers update form in (4.6), $\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b)$, the above inequality can be written as

$$\begin{aligned} x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) & \\ + (x - x^{k+1})^T\{-A^T\lambda^{k+1} + \beta A^T B(y^k - y^{k+1})\} & \\ \geq 0, \quad \forall x \in \mathcal{X}. & \end{aligned} \tag{4.8}$$

For the y-subproblem in (4.6b), by using Lemma 1, we have

$$\begin{aligned} y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) & \\ + (y - y^{k+1})^T\{-B^T\lambda^k + \beta B^T(Ax^{k+1} + By^{k+1} - b)\} & \\ + (y - y^{k+1})^T D_B(y^{k+1} - y^k) \geq 0, \ \forall y \in \mathcal{Y}. & \end{aligned}$$

Again, by using the update form $\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b)$, the above inequality can be written as

$$y^{k+1} \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T \lambda^{k+1}\} \\ \geq (y - y^{k+1})^T D_B (y^k - y^{k+1})\}, \; \forall \, y \in \mathcal{Y}. \tag{4.9}$$

Notice that the update form for the multipliers, $\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b)$, can be written as $\lambda^{k+1} \in \Re^m$ and

$$(\lambda - \lambda^{k+1})^T \{(Ax^{k+1} + By^{k+1} - b) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} \geq 0, \; \forall \, \lambda \in \Re^m. \tag{4.10}$$

Adding (4.8), (4.9) and (4.10), and using the notation in (2.2), we get

$$\begin{aligned} w^{k+1} \in \Omega, \quad \theta(u) - \theta(u^{k+1}) &+ (w - w^{k+1})^T F(w^{k+1}) \\ &+ \beta(x - x^{k+1})^T A^T (By^k - By^{k+1}) \\ &\geq (y - y^{k+1})^T D_B (y^k - y^{k+1}) \\ &\quad + \frac{1}{\beta}(\lambda - \lambda^{k+1})^T (\lambda^k - \lambda^{k+1}), \quad \forall w \in \Omega. \end{aligned} \tag{4.11}$$

For the term $(w - w^{k+1})^T F(w^{k+1})$ in the left side of (4.11), by using (2.3), we have

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

The assertion (4.7) is proved. $\Box$

This lemma is the base for the convergence analysis of the linearized ADMM.

The contractive property of the sequence $\{w^k\}$ by Linearized ADMM (4.6)

In the following we will prove, for any $w^* \in \Omega^*$, the sequence

$$\{\|v^{k+1} - v^*\|_G + \|y^k - y^{k+1}\|^2_{D_B}\}$$

is monotonically decreasing. For this purpose， we prove some lemmas.

引理 5 *Let $\{w^k\}$ be the sequence generated by Linearized ADMM* (4.6) *for the problem*

(1.1). *Then, we have*

$$w^{k+1} \in \Omega, \;\; \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w) + \beta \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}) \geq (v - v^{k+1})^T G(v^k - v^{k+1}), \;\; \forall w \in \Omega, \tag{4.12}$$

where G *is given by*

$$G = \begin{pmatrix} D_B + \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}. \tag{4.13}$$

.

Proof. Adding $(y - y^{k+1})^T \beta B^T B(y^k - y^{k+1})$ to the both sides of (4.7) in Lemma 4, and using the notation of the matrix G, we obtain (4.12) immediately and the lemma is proved. □

引理 6 *Let* $\{w^k\}$ *be the sequence generated by Linearized ADMM* (4.6) *for the problem*

(1.1). *Then, we have*

$$(v^{k+1} - v^*)^T G(v^k - v^{k+1}) \geq (\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}), \;\; \forall w^* \in \Omega^*. \tag{4.14}$$

Proof. Setting the $w \in \Omega$ in (4.12) by any $w^* \in \Omega^*$, we obtain

$$(v^{k+1} - v^*)^T G(v^k - v^{k+1}) \geq \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^*) + \beta \begin{pmatrix} x^{k+1} - x^* \\ y^{k+1} - y^* \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}). \tag{4.15}$$

According to the optimality, a part of the terms in the right hand side of the above inequality,

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

Using $Ax^* + By^* = b$ and $\lambda^k - \lambda^{k+1} = \beta(Ax^{k+1} + By^{k+1} - b)$ (see (4.6c)) to deal the last term in the right hand side of (4.15) , it follows that

$$\begin{aligned}&\beta\begin{pmatrix}x^{k+1}-x^*\\ y^{k+1}-y^*\end{pmatrix}^T\begin{pmatrix}A^T\\ B^T\end{pmatrix}B(y^k-y^{k+1})\\ &=\beta[(Ax^{k+1}-Ax^*)+(By^{k+1}-By^*)]^TB(y^k-y^{k+1})\\ &=(\lambda^k-\lambda^{k+1})^TB(y^k-y^{k+1}).\end{aligned}$$

The lemma is proved. □

引理 7 *Let* $\{w^k\}$ *be the sequence generated by Linearized ADMM* (4.6) *for the problem* (1.1). *Then, we have*

$$(\lambda^k-\lambda^{k+1})^TB(y^k-y^{k+1})\ge\frac{1}{2}\|y^k-y^{k+1}\|_{D_B}^2-\frac{1}{2}\|y^{k-1}-y^k\|_{D_B}^2. \quad (4.16)$$

Proof. First, (4.9) represents

$$\begin{aligned}&y^{k+1}\in\mathcal{Y},\ \ \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\\ &\qquad\{-B^T\lambda^{k+1}+D_B(y^{k+1}-y^k)\}\ge 0,\ \forall\, y\in\mathcal{Y}.\end{aligned} \quad (4.17)$$

Setting k in (4.17) by $k-1$, we have

$$\begin{aligned}&y^k\in\mathcal{Y},\ \ \theta_2(y)-\theta_2(y^k)+(y-y^k)^T\\ &\qquad\{-B^T\lambda^k+D_B(y^k-y^{k-1})\}\ge 0,\ \ \forall\, y\in\mathcal{Y}.\end{aligned} \quad (4.18)$$

Setting the y in (4.17) and (4.18) by y^k and y^{k+1}, respectively, and adding them, we get

$$(y^k-y^{k+1})^T\{B^T(\lambda^k-\lambda^{k+1})+D_B[(y^{k+1}-y^k)-(y^k-y^{k-1})]\}\ge 0.$$

From the above inequality we get

$$(y^k-y^{k+1})^TB^T(\lambda^k-\lambda^{k+1})\ge(y^k-y^{k+1})^TD_B[(y^k-y^{k+1})-(y^{k-1}-y^k)].$$

Using the Cauchy-Schwarz inequality for the right hand side term of the above inequality, we get (4.16) and the lemma is proved. □

By using Lemma 6 and Lemma 7, we can prove the following convergence theorem.

定理 2 *Let* $\{w^k\}$ *be the sequence generated by Linearized ADMM* (4.6) *for the problem* (1.1). *Then, we have*

$$\big(\|v^{k+1}-v^*\|_G^2+\|y^k-y^{k+1}\|_{D_B}^2\big) \le \big(\|v^k-v^*\|_G^2+\|y^{k-1}-y^k\|_{D_B}^2\big)-\|v^k-v^{k+1}\|_G^2,\ \forall w^*\in\Omega^*, \quad (4.19)$$

where G *is given by* (4.13).

Proof. From Lemma 6 and Lemma 7, it follows that

$$(v^{k+1}-v^*)^T G(v^k-v^{k+1}) \ge \frac{1}{2}\|y^k-y^{k+1}\|_{D_B}^2-\frac{1}{2}\|y^{k-1}-y^k\|_{D_B}^2, \quad \forall w^*\in\Omega^*.$$

Using the above inequality, for any $w^* \in \Omega^*$, we get

$$\begin{aligned}\|v^k-v^*\|_G^2 &= \|(v^{k+1}-v^*)+(v^k-v^{k+1})\|_G^2 \\ &\ge \|v^{k+1}-v^*\|_G^2+\|v^k-v^{k+1}\|_G^2+2(v^{k+1}-v^*)^T G(v^k-v^{k+1}) \\ &\ge \|v^{k+1}-v^*\|_G^2+\|v^k-v^{k+1}\|_G^2 \\ &\quad +\|y^k-y^{k+1}\|_{D_B}^2-\|y^{k-1}-y^k\|_{D_B}^2.\end{aligned}$$

The assertion of the Theorem 2 is proved. $\Box$

Optimal linearized ADMM – Main result in OO6228

In the subproblem of the Linearized ADMM, namely (4.6b), in order to ensure the convergence, it was required that

$$D_B = sI_{n_2} - \beta B^T B \quad \text{and} \quad s > \beta\|B^T B\|. \quad (4.20)$$

It is well known that the large parameter s will lead a slow convergence.

Recent Advance in : Bingsheng He, Feng Ma, Xiaoming Yuan: Optimally linearizing the alternating direction method of multipliers for convex programming, Comput. Optim. Appl. 75 (2020), 361-388.

We have proved: For the matrix D_B in (4.6b) with form (4.20)

- if $s > \frac{3}{4}\beta\|B^T B\|$, the method is still convergent;
- if $s < \frac{3}{4}\beta\|B^T B\|$, there is divergent example.

5 Customized PPA for Variational Inequality

We study the algorithms using the guidance of variational inequality. The optimal condition of the linearly constrained convex optimization is resulted in a variational inequality:

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

5.1 Customized PPA for VI (5.1)

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

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

where the matrix H is positive definite.

[Correction Step.] Determine a nonsingular matrix M and a scalar $\alpha > 0$, let

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

v is a part of the elements of the vector w, $\quad v = w$ is also possible.

5.2 Convergence proof

We prove the following main convergence property.

定理 1 *Let* $\{v^k\}$ *be the sequence generated by* (5.2) *for the problem* (5.1) *and* $\tilde{w}^k$ *is obtained from* (5.2a)*. Then we have*

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \alpha(2-\alpha)\|v^k - \tilde{v}^k\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{5.3}$$

where $\mathcal{V}^* = \{v^* \,|\, v^* \text{is a part of } w^*, w^* \in \Omega^*\}$.

Proof. Setting $w = w^*$ in (5.2a), we get

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

By using $(\tilde{w}^k - w^*)^T F(\tilde{w}^k) = (\tilde{w}^k - w^*)^T F(w^*)$ and the optimality of w^*, we have

$$(\tilde{v}^k - v^*)^T H(v^k - \tilde{v}^k) \geq 0, \quad \forall v^* \in \mathcal{V}^*.$$

It can be written as

$$\{(v^k - v^*) - (v^k - \tilde{v}^k)\}^T H(v^k - \tilde{v}^k) \geq 0, \quad \forall v^* \in \mathcal{V}^*,$$

and thus

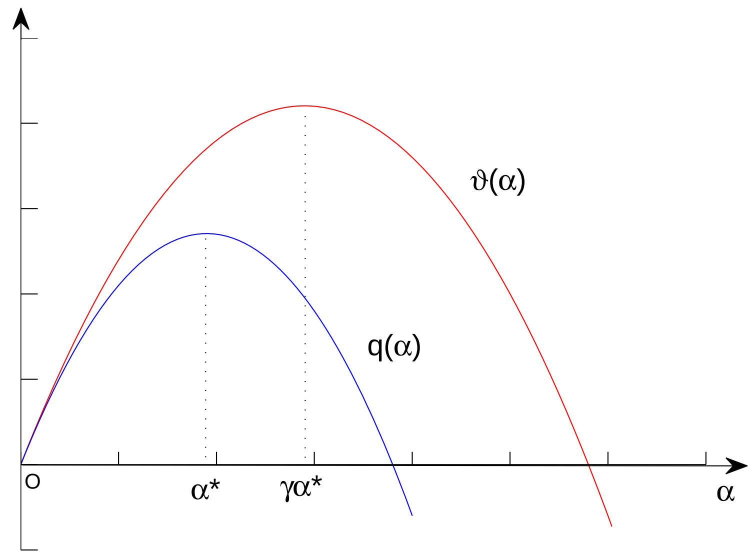
$$(v^k - v^*)^T H(v^k - \tilde{v}^k) \geq \|v^k - \tilde{v}^k\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{5.4}$$

Let

$$\vartheta(\alpha) = \|v^k - v^*\|_H^2 - \|v_\alpha^{k+1} - v^*\|_H^2.$$

It follows that

$$\begin{aligned}\vartheta(\alpha) &= \|v^k - v^*\|_H^2 - \|v_\alpha^{k+1} - v^*\|_H^2 \\ &= \|v^k - v^*\|_H^2 \\ &\quad -\|(v^k - v^*) - \alpha(v^k - \tilde{v}^k)\|_H^2 \\ &= 2\alpha(v^k - v^*)^T H(v^k - \tilde{v}^k) \\ &\quad -\alpha^2\|v^k - \tilde{v}^k\|_H^2. \end{aligned} \tag{5.5}$$



取 $\gamma \in [1,2)$ 的示意图

我们本想极大化 $\vartheta(\alpha)$, 虽然 $\vartheta(\alpha)$ 是 α 的二次函数, 但线性项系数 $2(v^k - v^*)^T H(v^k - \tilde{v}^k)$ 中含有未知的 v^*, 利用 (5.4), 得到 $\vartheta(\alpha)$ 的下界函数 $q(\alpha)$. 极大化 $q(\alpha)$, $\alpha_k^* \equiv 1$. 可以松弛延拓.

Using (5.4), we get

$$\begin{aligned}\vartheta(\alpha) &\geq 2\alpha\|v^k - \tilde{v}^k\|_H^2 - \alpha^2\|v^k - \tilde{v}^k\|_H^2 \\ &:= q(\alpha)\end{aligned} \tag{5.6}$$

The assertion (5.3) follows from (5.5) and (5.6) immediately. □

6 Applications for separable problems

6.1 ADMM in PPA-sense

根据 PPA 算法的要求 设计的右端矩阵为对称正定. 具体算法可参阅 [20]

In order to solve the separable convex optimization problem (1.1), we construct a method whose prediction-step is

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

where

$$H = \begin{pmatrix} \beta B^T B + \delta I_{n_2} & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}, \quad \text{(a small } \delta > 0). \tag{6.1b}$$

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

The concrete form of (6.1) is

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k}\} \ge 0, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \\ \qquad \{\underline{-B^T\tilde{\lambda}^k} + (\boldsymbol{\beta B^T B + \delta I_{n_2}})(\tilde{y}^k - y^k) \boldsymbol{- B^T}(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \\ (\underline{A\tilde{x}^k + B\tilde{y}^k - b}) \quad \boldsymbol{-B}(\tilde{y}^k - y^k) \quad + \quad \boldsymbol{(1/\beta)}\ (\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

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

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

In fact, the prediction can be arranged by

$$\begin{cases} \tilde{x}^k \in \text{Argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 \,|\, x \in \mathcal{X}\}, & \text{(6.2a)} \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), & \text{(6.2b)} \\ \tilde{y}^k \in \text{Argmin}\left\{ \begin{array}{c} \theta_2(y) - y^T B^T[2\tilde{\lambda}^k - \lambda^k] \\ +\frac{1}{2}\beta\|B(y - y^k)\|^2 + \frac{1}{2}\delta\|y - y^k\|^2 \end{array} \middle| \, y \in \mathcal{Y} \right\}. & \text{(6.2c)} \end{cases}$$

这个预测与经典的交替方向法 (3.3) 相当, 采用(5.2b) 校正, 会加快速度.

According to Lemma 1, the solution of (6.2a), $\tilde{x}^k$ satisfies

$$\tilde{x}^k \in \mathcal{X}, \quad \theta_1(x) - \theta_1(\tilde{x}^k) \\ \qquad + (x - \tilde{x}^k)^T\{-A^T\lambda^k + \beta A^T(A\tilde{x}^k + By^k - b)\} \ge 0, \quad \forall x \in \mathcal{X}.$$

By using (6.2b), $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b)$, the above variational inequality can be written as

$$\tilde{x}^k \in \mathcal{X}, \quad \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k}\} \ge 0, \quad \forall x \in \mathcal{X}.$$

The equation (6.2b) can written as

$$(\underline{A\tilde{x}^k + B\tilde{y}^k - b}) \boldsymbol{-B}(\tilde{y}^k - y^k) + \boldsymbol{(1/\beta)}(\tilde{\lambda}^k - \lambda^k) = 0.$$

The remainder part of the prediction (6.2c), namely,

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \\ \qquad \{\underline{-B^T\tilde{\lambda}^k} + (\boldsymbol{\beta B^T B + \delta I_{n_2}})(\tilde{y}^k - y^k) \boldsymbol{-B^T}(\tilde{\lambda}^k - \lambda^k)\} \ge 0$$

can be achieved by

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) - y^T B^T[2\tilde{\lambda}^k - \lambda^k] + \frac{1}{2}\beta\|B(y - y^k)\|^2 + \frac{1}{2}\delta\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}.$$

如果把 (6.2) 中取 $\delta=0$, 并将其输出记为 $(x^{k+1},\lambda^{k+1},y^{k+1})$, 则迭代式为

$$\left\{\begin{array}{ll} x^{k+1}\in \text{Argmin}\{\theta_1(x)-x^TA^T\lambda^k+\frac{\beta}{2}\|Ax+By^k-b\|^2\,|\,x\in\mathcal{X}\}, & \text{(6.3a)}\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^k-b), & \text{(6.3b)}\\ y^{k+1}\in \text{Argmin}\big\{\theta_2(y)-y^TB^T[2\lambda^{k+1}-\lambda^k]+\frac{\beta}{2}\|B(y-y^k)\|^2\,\big|\,y\in\mathcal{Y}\big\} & \text{(6.3c)}\end{array}\right.$$

注意在 (6.3c) 中,

$$\begin{aligned} y^{k+1}\in &\text{Argmin}\big\{\theta_2(y)-y^TB^T[2\lambda^{k+1}-\lambda^k]+\frac{\beta}{2}\|B(y-y^k)\|^2\,\big|\,y\in\mathcal{Y}\big\}\\ &=\text{Argmin}\big\{\theta_2(y)-y^TB^T\lambda^{k+1}-y^TB^T(\lambda^{k+1}-\lambda^k)+\frac{\beta}{2}\|B(y-y^k)\|^2\,\big|\,y\in\mathcal{Y}\big\}\\ &=\text{Argmin}\big\{\theta_2(y)-y^TB^T\lambda^{k+1}+\frac{\beta}{2}\|B(y-y^k)-\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)\|^2\,\big|\,y\in\mathcal{Y}\big\}\\ &=\text{Argmin}\big\{\theta_2(y)-y^TB^T\lambda^{k+1}+\frac{\beta}{2}\|B(y-y^k)+\frac{1}{\beta}(\lambda^k-\lambda^{k+1})\|^2\,\big|\,y\in\mathcal{Y}\big\}\\ &=\text{Argmin}\big\{\theta_2(y)-y^TB^T\lambda^{k+1}+\frac{\beta}{2}\|B(y-y^k)+(Ax^{k+1}+By^k-b)\|^2\,\big|\,y\in\mathcal{Y}\big\}\\ &=\text{Argmin}\big\{\theta_2(y)-y^TB^T\lambda^{k+1}+\frac{\beta}{2}\|Ax^{k+1}+By-b\|^2\,\big|\,y\in\mathcal{Y}\big\}.\end{aligned}$$

所以, (6.3) 就是

$$\left\{\begin{array}{ll} x^{k+1}\in \text{Argmin}\{\theta_1(x)-x^TA^T\lambda^k+\frac{\beta}{2}\|Ax+By^k-b\|^2\,|\,x\in\mathcal{X}\}, & \text{(6.4a)}\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^k-b), & \text{(6.4b)}\\ y^{k+1}\in \text{Argmin}\big\{\theta_2(y)-y^TB^T\lambda^{k+1}+\frac{\beta}{2}\|Ax^{k+1}+By-b\|^2\,\big|\,y\in\mathcal{Y}\big\}. & \text{(6.4c)}\end{array}\right.$$

请注意, 经典的 ADMM 是

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

所以, (6.3), 就是交换了 y,λ 顺序的交替方向法. 由于可以采用

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

通常取 $\alpha=1.5$, 收敛更快.

6.2 Linearized ADMM-Like Method

当子问题 (6.2c) 求解有困难时, 用 $\frac{s}{2}\|y-y^k\|^2$ 代替 $\frac{1}{2}\|y-y^k\|^2_{(\beta B^TB+\delta I_{n_2})}$

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

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\geq(v-\tilde{v}^k)^TH(v^k-\tilde{v}^k),\ \forall w\in\Omega, \quad (6.5)$$

where

$$H=\begin{bmatrix} sI & -B^T \\ -B & \frac{1}{\beta}I_m \end{bmatrix},\ \text{代替 (6.1) 中的}\ \begin{bmatrix} \beta B^TB+\delta I_{n_2} & -B^T \\ -B & \frac{1}{\beta}I_m \end{bmatrix}. \quad (6.6)$$

The concrete formula of (6.5) is

$$\begin{cases} \theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k}\}\geq 0, \\ \theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T \\ \qquad \{\underline{-B^T\tilde{\lambda}^k}+\ \boldsymbol{s}(\tilde{y}^k-y^k)\boldsymbol{-B^T}(\tilde{\lambda}^k-\lambda^k)\}\geq 0, \\ (\underline{A\tilde{x}^k+B\tilde{y}^k-b})\boldsymbol{-B}(\tilde{y}^k-y^k)+\boldsymbol{(1/\beta)}(\tilde{\lambda}^k-\lambda^k)=0. \end{cases} \quad (6.7)$$

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

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

How to implement the prediction? To get $\tilde{w}^k$ which satisfies (6.7), we need only use the following procedure:

$$\begin{cases} \tilde{x}^k\in \text{Argmin}\{\theta_1(x)-x^TA^T\lambda^k+\frac{1}{2}\beta\|Ax+By^k-b\|^2\,|\,x\in\mathcal{X}\}, \\ \tilde{\lambda}^k=\lambda^k-\beta(A\tilde{x}^k+By^k-b), \\ \tilde{y}^k=\text{Argmin}\{\theta_2(y)-y^TB^T[\boldsymbol{2\tilde{\lambda}^k-\lambda^k}]+\dfrac{s}{2}\|y-y^k\|^2\,|\,y\in\mathcal{Y}\}. \end{cases} \quad (6.8)$$

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

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

只有当由二次项 $\frac{1}{2}\beta\|B(y-y^k)\|^2$ 引发求解困难, 才用线性化方法.

Then, we use the form

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

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

7 Solving the primal subproblem in parallel

根据 PPA 算法的要求 设计的右端矩阵为对称正定.

Primal-Dual Order

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w - \tilde{w}^k)^T H(w^k - \tilde{w}^k), \ \forall w \in \Omega, \tag{7.1a}$$

where

$$H = \begin{pmatrix} \beta A^T A + \delta I_{n_1} & 0 & A^T \\ 0 & \beta B^T B + \delta I_{n_2} & B^T \\ A & B & \frac{2}{\beta} I_m \end{pmatrix}. \tag{7.1b}$$

The both matrices

$$\begin{pmatrix} \beta A^T A + \delta I_{n_1} & A^T \\ A & \frac{1}{\beta} I_m \end{pmatrix} \succ 0, \qquad \begin{pmatrix} \beta B^T B + \delta I_{n_2} & B^T \\ B & \frac{1}{\beta} I_m \end{pmatrix} \succ 0.$$

The concrete form of (7.1) is

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k} + \boldsymbol{(\beta A^T A + \delta I_{n_1})}(\tilde{x}^k - x^k) \boldsymbol{+ A^T}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \\ \qquad \{\underline{-B^T\tilde{\lambda}^k} + \boldsymbol{(\beta B^T B + \delta I_{n_2})}(\tilde{y}^k - y^k) \boldsymbol{+ B^T}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \\ (\underline{A\tilde{x}^k + B\tilde{y}^k - b}) \boldsymbol{+ A}(\tilde{x}^k - x^k) \boldsymbol{+ B}(\tilde{y}^k - y^k) + \boldsymbol{(2/\beta)}\ (\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

整理一下得到

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\lambda^k + (\beta A^T A + \delta I_{n_1})(\tilde{x}^k - x^k)\} \geq 0, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{-B^T\lambda^k + (\beta B^T B + \delta I_{n_2})(\tilde{y}^k - y^k)\} \geq 0, \\ [2(A\tilde{x}^k + B\tilde{y}^k - b) - (Ax^k + By^k - b)] + (2/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

In fact, the prediction can be arranged by

$$\begin{cases} \tilde{x}^k = \arg\min\left\{ \begin{array}{c} \theta_1(x) - x^T A^T \lambda^k \\ +\frac{1}{2}\beta\|A(x-x^k)\|^2 + \frac{1}{2}\delta\|x-x^k\|^2 \end{array} \middle| x \in \mathcal{X} \right\} & \text{(7.2a)} \\ \tilde{y}^k = \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T \lambda^k \\ +\frac{1}{2}\beta\|B(y-y^k)\|^2 + \frac{1}{2}\delta\|y-y^k\|^2 \end{array} \middle| y \in \mathcal{Y} \right\} & \text{(7.2b)} \\ \tilde{\lambda}^k = \lambda^k - \frac{1}{2}\beta\big[2(A\tilde{x}^k + B\tilde{y}^k - b) - (Ax^k + By^k - b)\big] & \text{(7.2c)} \end{cases}$$

$$\begin{cases} \tilde{x}^k = \arg\min\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}(x-x^k)^T(\beta A^T A + \delta I_{n_1})(x-x^k) | x \in \mathcal{X}\} \\ \tilde{y}^k = \arg\min\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}(y-y^k)^T(\beta B^T B + \delta I_{n_2})(y-y^k) | y \in \mathcal{Y}\} \\ \tilde{\lambda}^k = \lambda^k - \frac{1}{2}\beta\big[2(A\tilde{x}^k + B\tilde{y}^k - b) - (Ax^k + By^k - b)\big] \end{cases}$$

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

Dual-Primal Order

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (w - \tilde{w}^k)^T H(w^k - \tilde{w}^k), \ \forall w \in \Omega, \tag{7.3a}$$

where

$$H = \begin{pmatrix} \beta A^T A + \delta I_{n_1} & 0 & -A^T \\ 0 & \beta B^T B + \delta I_{n_2} & -B^T \\ -A & -B & \frac{2}{\beta} I_m \end{pmatrix}. \tag{7.3b}$$

The both matrices

$$H = \begin{pmatrix} \beta A^T A + \delta I_{n_1} & -A^T \\ -A & \frac{1}{\beta} I_m \end{pmatrix} \succ 0, \quad \begin{pmatrix} \beta B^T B + \delta I_{n_2} & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \succ 0.$$

The concrete form of (7.3) is

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \qquad \{\underline{-A^T\tilde{\lambda}^k} + \boldsymbol{(\beta A^T A + \delta I_{n_1})}(\tilde{x}^k - x^k)\boldsymbol{-A^T}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \\ \qquad \{\underline{-B^T\tilde{\lambda}^k} + \boldsymbol{(\beta B^T B + \delta I_{n_2})}(\tilde{y}^k - y^k)\boldsymbol{-B^T}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \\ (\underline{A\tilde{x}^k + B\tilde{y}^k - b})\boldsymbol{-A}(\tilde{x}^k - x^k)\boldsymbol{-B}(\tilde{y}^k - y^k) + \boldsymbol{(2/\beta)}\ (\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

经整理归并一下得到

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T(2\tilde{\lambda}^k - \lambda^k) \\ \qquad\qquad + (\beta A^T A + \delta I_{n_1})(\tilde{x}^k - x^k)\} \geq 0, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{-B^T(2\tilde{\lambda}^k - \lambda^k) \\ \qquad\qquad + (\beta B^T B + \delta I_{n_2})(\tilde{y}^k - y^k)\} \geq 0, \\ (Ax^k + By^k - b) + (2/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \end{cases}$$

In fact, the prediction can be arranged by

$$\begin{cases} \tilde{\lambda}^k = \lambda^k - \dfrac{1}{2}\beta(Ax^k + By^k - b), & \text{(7.4a)} \\ \tilde{x}^k \in \arg\min\left\{ \begin{array}{c} \theta_1(x) - x^T A^T[2\tilde{\lambda}^k - \lambda^k] \\ + \frac{1}{2}\beta\|A(x - x^k)\|^2 + \frac{1}{2}\delta\|x - x^k\|^2 \end{array} \middle| x \in \mathcal{X} \right\} & \text{(7.4b)} \\ \tilde{y}^k \in \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T[2\tilde{\lambda}^k - \lambda^k] \\ + \frac{1}{2}\beta\|B(y - y^k)\|^2 + \frac{1}{2}\delta\|y - y^k\|^2 \end{array} \middle| y \in \mathcal{Y} \right\}. & \text{(7.4c)} \end{cases}$$

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

我们关于 ADMM 的研究, 始于 1997 年, 第一篇 ADMM 方面的论文发表于 1998 年. 这一讲中 §4-§6 介绍的 ADMM 类方法, 可以从 [20] 中找到.

利用变分不等式 (VI) 和邻近点算法 (PPA), 更自由地设计 ADMM 类分裂收缩算法

References

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

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

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

[4] R. Glowinski, On alternating direction methods of multipliers: a historical perspective, *Modeling, Simulation and Optimization for Science and Technology*, edited by W. Fitzgibbon, Y. A. Kuznetsov, P. Neittaanmäki and O. Pironneau, Eccomas, Springer, 2014, pp. 59-82.

[5] R. Glowinski and A. Marrocco, Approximation par éléments finis d'ordre un et résolution par pénalisation-dualité d'une classe de problémes non linéaires. RAIRO Anal. Numer. R2, 41–76 (1975)

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

[7] 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.

[8] 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.

[9] 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.

[10] B. S. He and X. M. Yuan, On the $O(1/t)$ convergence rate of the Douglas-Rachford Alternating Direction Method, *SIAM J. Numerical Analysis* **50**(2012), 700-709.

[11] B.S. He and X.M. Yuan, On non-ergodic convergence rate of Douglas-Rachford alternating directions method of multipliers, Numerische Mathematik, 130 (2015), 567-577.

[12] B. S. He and X. M. Yuan, Balanced Augmented Lagrangian Method for Convex Programming, arXiv 2108.08554 [math.OC]

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

[14] Z. Lin, R. Liu and H. Li, Linearized alternating direction method with parallel splitting and adaptive penalty for separable convex programs in machine learning, Machine Learning, 99 (2015), 287-325.

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

[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] X. F. Wang and X. M. Yuan, The linearized alternating direction method of multipliers for Dantzig selector, SIAM J. Sci. Comput., 34 (2012), pp. A2792–A2811.

[18] J. F. Yang and X. M. Yuan, Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization, Math. Comp., 82 (2013), pp. 301-329.

[19] X. Q. Zhang, M. Burger, and S. Osher, A unified primal-dual algorithm framework based on Bregman iteration, J. Sci. Comput., 46 (2010), pp. 20-46.

[20] 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.

变分不等式框架下结构型
凸优化的分裂收缩算法

IV. 线性约束凸优化问题分裂收缩算法的统一框架

中学的数理基础　必要的社会实践
普通的大学数学　一般的优化原理

何 炳 生　　南京大学数学系
Homepage: maths.nju.edu.cn/~hebma

天元数学东北中心　2023 年 10 月 17 – 27 日

1 凸优化分裂收缩算法的统一框架

我们总是用变分不等式 (VI) 指导算法设计, 把线性约束的凸优化问题归结为下面的变分不等式:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega. \tag{1.1}$$

Algorithms in a unified framework

A unified Algorithmic Framework for (1.1)　统一框架由预测-校正两部分组成

[Prediction Step.] 从给定的 v^k 出发, 求得预测点 $\tilde{w}^k \in \Omega$ 使其满足

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall w \in \Omega, \tag{1.2a}$$

其中 Q 不一定对称, 但是 $Q^T + Q$ 正定.

[Correction Step.] 给一个合适的非奇异矩阵 M, 由下式确定新的迭代点

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{1.2b}$$

Q 和 M 分别叫做预测矩阵和校正矩阵

Convergence Conditions

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

$$HM = Q. \tag{1.3a}$$

In addition,

$$G = Q^T + Q - M^T HM \succ 0. \tag{1.3b}$$

其实, 只要预测 (1.2a) 中的预测矩阵 Q 满足

$$Q^T + Q \succ 0,$$

我们总可以取

$$0 \prec G \prec Q^T + Q.$$

然后记

$$D = (Q^T + Q) - G,$$

则 $D \succ 0$. 令

$$M^T HM = D.$$

由矩阵方程组解得

$$\left\{\begin{array}{r} HM = Q, \\ M^T HM = D. \end{array}\right. \quad \Leftrightarrow \quad \left\{\begin{array}{r} HM = Q, \\ Q^T M = D. \end{array}\right. \quad \Leftrightarrow \quad \left\{\begin{array}{l} H = QD^{-1}Q^T, \\ M = Q^{-T}D. \end{array}\right.$$

就得到满足收敛条件的校正矩阵 M.

实际计算中, 我们只要校正矩阵 M.

H 和 G 只是用来验证收敛条件的.

换句话说, 只要

$$Q^T + Q \succ 0.$$

我们就可以选两个正定矩阵 $D \succ 0$ 和 $G \succ 0$, 使得

$$D + G = Q^T + Q.$$

将 (1.2b) 中的校正矩阵 M 取成

$$M = Q^{-T} D$$

条件 (1.3) 自然满足.

2 预测-校正方法的例子

We consider the $\min - \max$ problem

$$\min_x \max_y \{\Phi(x,y) = \theta_1(x) - y^T Ax - \theta_2(y) \,|\, x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{2.4}$$

Let (x^*, y^*) be the solution of (2.4), then we have 根据鞍点的定义

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

上面的两个不等式可以写成等价的

$$\begin{cases} x^* \in \mathcal{X}, \quad \Phi(x, y^*) - \Phi(x^*, y^*) \ge 0, \quad \forall x \in \mathcal{X}, & (2.5a) \\ y^* \in \mathcal{Y}, \quad \Phi(x^*, y^*) - \Phi(x^*, y) \ge 0, \quad \forall y \in \mathcal{Y}. & (2.5b) \end{cases}$$

Using the notation of $\Phi(x,y)$, it can be written as 只要把 $\Phi(x,y)$ 的形式填进去

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

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

$$u \in \Omega, \quad \theta(u) - \theta(u^*) + (u - u^*)^T F(u^*) \ge 0, \ \forall u \in \Omega, \tag{2.6}$$

where 对上式中任意的 $u \in \Omega$ 分别取 $u = (x, y^*)$ 和 $u = (x^*, y)$, 就得到 $(*)$ 和 $(\diamond)$.

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

The output of Original PDHG algorithm [16] as predictor

For given (x^k, y^k), PDHG [16] produces a pair of $(\tilde{x}^k, \tilde{y}^k)$. First,

$$\tilde{x}^k = \mathrm{argmin}\{\Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\}, \tag{2.7a}$$

and then we obtain $\tilde{y}^k$ via

$$\tilde{y}^k = \mathrm{argmax}\{\Phi(\tilde{x}^k, y) - \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}. \tag{2.7b}$$

Ignoring the constant term in the objective function, the subproblems (2.7) are reduced to

$$\begin{cases} \tilde{x}^k = \operatorname{argmin}\{\theta_1(x) - x^T A^T y^k + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\}, & (2.8a) \\ \tilde{y}^k = \operatorname{argmin}\{\theta_2(y) + y^T A\tilde{x}^k + \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}. & (2.8b) \end{cases}$$

According to the basic lemma, the optimality condition of (2.8a) is $\tilde{x}^k \in \mathcal{X}$ and

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T y^k + r(\tilde{x}^k - x^k)\} \geq 0, \ \forall x \in \mathcal{X}. \quad (2.9)$$

Similarly, from (2.8b) we get $\tilde{y}^k \in \mathcal{Y}$ and

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{A\tilde{x}^k + s(\tilde{y}^k - y^k)\} \geq 0, \ \forall y \in \mathcal{Y}. \quad (2.10)$$

Combining (2.9) and (2.10), we have

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

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

$$\theta(u) - \theta(\tilde{u}^k) + (u - \tilde{u}^k)^T\{F(\tilde{u}^k) + Q(\tilde{u}^k - u^k)\} \geq 0, \ \forall u \in \Omega, \quad (2.11a)$$

where

$$Q = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix}. \quad (2.11b)$$

对于这样的预测, 我们考虑比较简单的校正

$$u^{k+1} = u^k - M(u^k - \tilde{u}^k) \quad (2.12)$$

校正. 其中 M 为单位上三角矩阵或单位下三角矩阵. 收敛性条件 (1.3)

- $H \succ 0$ and $HM = Q$.
- $G = Q^T + Q - M^T HM \succ 0$.

可以改写成等价的

(i) $H \succ 0$ and $H = QM^{-1}$.

(ii) $G = Q^T + Q - Q^T M \succ 0$.

一. 校正矩阵 M 为单位下三角矩阵

其中的 K 是待定的.

$$M = \begin{pmatrix} I_n & 0 \\ K & I_m \end{pmatrix} \quad 则 \quad M^{-1} = \begin{pmatrix} I_n & 0 \\ -K & I_m \end{pmatrix}.$$

对条件 (i), 我们在统一框架下指导下求出这个 K 的具体形式. 由于 $H = QM^{-1}$ 正定, 首先必须是对称的. 由

$$H = QM^{-1} = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \begin{pmatrix} I_n & 0 \\ -K & I_m \end{pmatrix} = \begin{pmatrix} rI_n - A^TK & A^T \\ -sK & sI_m \end{pmatrix}$$

必须对称, 推得

$$-sK = A, \qquad \Rightarrow \qquad K = -\frac{1}{s}A.$$

因此,

$$M = \begin{pmatrix} I_n & 0 \\ -\frac{1}{s}A & I_m \end{pmatrix}, \qquad H = \begin{pmatrix} rI_n + \frac{1}{s}A^TA & A^T \\ A & sI_m \end{pmatrix}.$$

对任意的 $r, s > 0$, 矩阵 H 是正定的.

对条件 (ii),

$$\begin{aligned} G &= Q^T + Q - M^T H M = Q^T + Q - Q^T M \\ &= \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n & 0 \\ A & sI_m \end{pmatrix} \begin{pmatrix} I_n & 0 \\ -\frac{1}{s}A & I_m \end{pmatrix} \\ &= \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix} = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}. \end{aligned}$$

要矩阵 G 正定, 必须有 $rs > \|A^TA\|$.

采用 PDHG 预测, 单位下三角矩阵校正, 需要 $rs > \|A^TA\|$.

二. 校正矩阵 M 为单位上三角矩阵

同样, 其中的 K 是待定的.

$$M = \begin{pmatrix} I_n & K \\ 0 & I_m \end{pmatrix} \quad 则 \quad M^{-1} = \begin{pmatrix} I_n & -K \\ 0 & I_m \end{pmatrix}.$$

对条件 (i), 我们在统一框架下指导下求出这个 K 的具体形式. 由于 $H = QM^{-1}$ 正定,

首先必须是对称的. 由

$$H = QM^{-1} = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \begin{pmatrix} I_n & -K \\ 0 & I_m \end{pmatrix} = \begin{pmatrix} rI_n & -rK + A^T \\ 0 & sI_m \end{pmatrix}$$

必须对称, 推得

$$rK = A^T, \qquad \Rightarrow \qquad K = \frac{1}{r}A^T.$$

因此,

$$M = \begin{pmatrix} I_n & \frac{1}{r}A^T \\ 0 & I_m \end{pmatrix}, \qquad H = \begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix}.$$

对任意的 $r, s > 0$, 矩阵 H 是正定的.

而对条件 (ii),

$$\begin{aligned} G &= Q^T + Q - M^T HM = Q^T + Q - Q^T M \\ &= \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n & 0 \\ A & sI_m \end{pmatrix} \begin{pmatrix} I_n & \frac{1}{r}A^T \\ 0 & I_m \end{pmatrix} \\ &= \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix} \\ &= \begin{pmatrix} rI_n & 0 \\ 0 & sI_m - \frac{1}{r}AA^T \end{pmatrix}. \end{aligned}$$

要矩阵 G 正定, 必须有 $rs > \|A^T A\|$.

采用 PDHG 预测, 单位上三角矩阵校正, 需要 $rs > \|A^T A\|$.

虽然把不能保证收敛的 PDHG 方法改造成了收敛的方法, 但是, rs 的值没有降下来.

我们的目标, 是把预测 (2.8) 中的参数 rs 想办法降下来.

对于 (2.11) 中的 Q, 我们有

$$Q^T + Q = \begin{pmatrix} 2rI & A^T \\ A & 2sI \end{pmatrix}$$

只要 $rs > \frac{1}{4}\|A^TA\|$, 矩阵 Q^T+Q 都是正定的.

当 (Q^T+Q) 正定时, 我们取

$$D = \frac{1}{2}(Q^T+Q), \qquad \text{并令} \qquad M^THM = D. \tag{2.13}$$

这样就能保证

$$G = Q^T + Q - M^THM = \frac{1}{2}(Q^T+Q) \succ 0.$$

• $H \succ 0$ and $HM = Q$. • $G = Q^T + Q - M^THM \succ 0$.	可以改写成	(i) $HM = Q$. (ii) $M^THM = D$.

$$\left\{\begin{array}{l} HM = Q, \\ M^THM = D. \end{array}\right. \Leftrightarrow \left\{\begin{array}{l} HM = Q, \\ Q^TM = D. \end{array}\right. \Leftrightarrow \left\{\begin{array}{l} H = QD^{-1}Q^T, \\ M = Q^{-T}D. \end{array}\right. \tag{2.14}$$

换句话说, 当 $(Q^T+Q) \succ 0$, 取

$$D = \begin{pmatrix} rI & \frac{1}{2}A^T \\ \frac{1}{2}A & sI \end{pmatrix}, \quad M = Q^{-T}D$$

所有收敛性条件都满足. 而

$$Q^{-T} = \begin{pmatrix} rI & 0 \\ A & sI \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{r}I & 0 \\ -\frac{1}{rs}A & \frac{1}{s}I \end{pmatrix}$$

$$\begin{aligned} M &= Q^{-T}D = \begin{pmatrix} \frac{1}{r}I & 0 \\ -\frac{1}{rs}A & \frac{1}{s}I \end{pmatrix}\begin{pmatrix} rI & \frac{1}{2}A^T \\ \frac{1}{2}A & sI \end{pmatrix} \\ &= \begin{pmatrix} I & \frac{1}{2r}A^T \\ -\frac{1}{2s}A & I - \frac{1}{2rs}AA^T \end{pmatrix} \end{aligned} \tag{2.15}$$

利用上面的校正矩阵 M

$$\left\{\begin{array}{lcl} x^{k+1} &=& \tilde{x}^k - \frac{1}{2r}A^T(y^k - \tilde{y}^k) \\ y^{k+1} &=& \tilde{y}^k + \frac{1}{2s}A[(x^k - \tilde{x}^k) + \frac{1}{r}A^T(y^k - \tilde{y}^k)]. \end{array}\right.$$

这是马峰他们 [14] 根据统一框架提出的方法. 计算效果有很大进步.

把 rs 的积降了 $\frac{3}{4}$, 有了很大进步.

3 Convergence proof in the unified framework

In this section, assuming the conditions (1.3) in the unified framework are satisfied, we prove some convergence properties.

定理 1 *Let $\{v^k\}$ be the sequence generated by a method for the problem* (1.1) *and $\tilde{w}^k$ is obtained in the k-th iteration. If v^k, v^{k+1} and $\tilde{w}^k$ satisfy the conditions in the unified framework, then we have*

$$\begin{aligned}&\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\\ &\quad\geq\ \frac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\frac{1}{2}\|v^k-\tilde{v}^k\|_G^2,\quad\forall w\in\Omega. \qquad (3.1)\end{aligned}$$

Proof. Using $Q=HM$ (see (1.3a)) and the relation (1.2b), the right hand side of (1.3a) can be written as $(v-\tilde{v}^k)^TH(v^k-v^{k+1})$ and hence

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\geq(v-\tilde{v}^k)^TH(v^k-v^{k+1}),\ \forall w\in\Omega. \qquad (3.2)$$

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

$$(a-b)^TH(c-d)=\frac{1}{2}\{\|a-d\|_H^2-\|a-c\|_H^2\}+\frac{1}{2}\{\|c-b\|_H^2-\|d-b\|_H^2\},$$

to the right hand side of (3.2) with

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

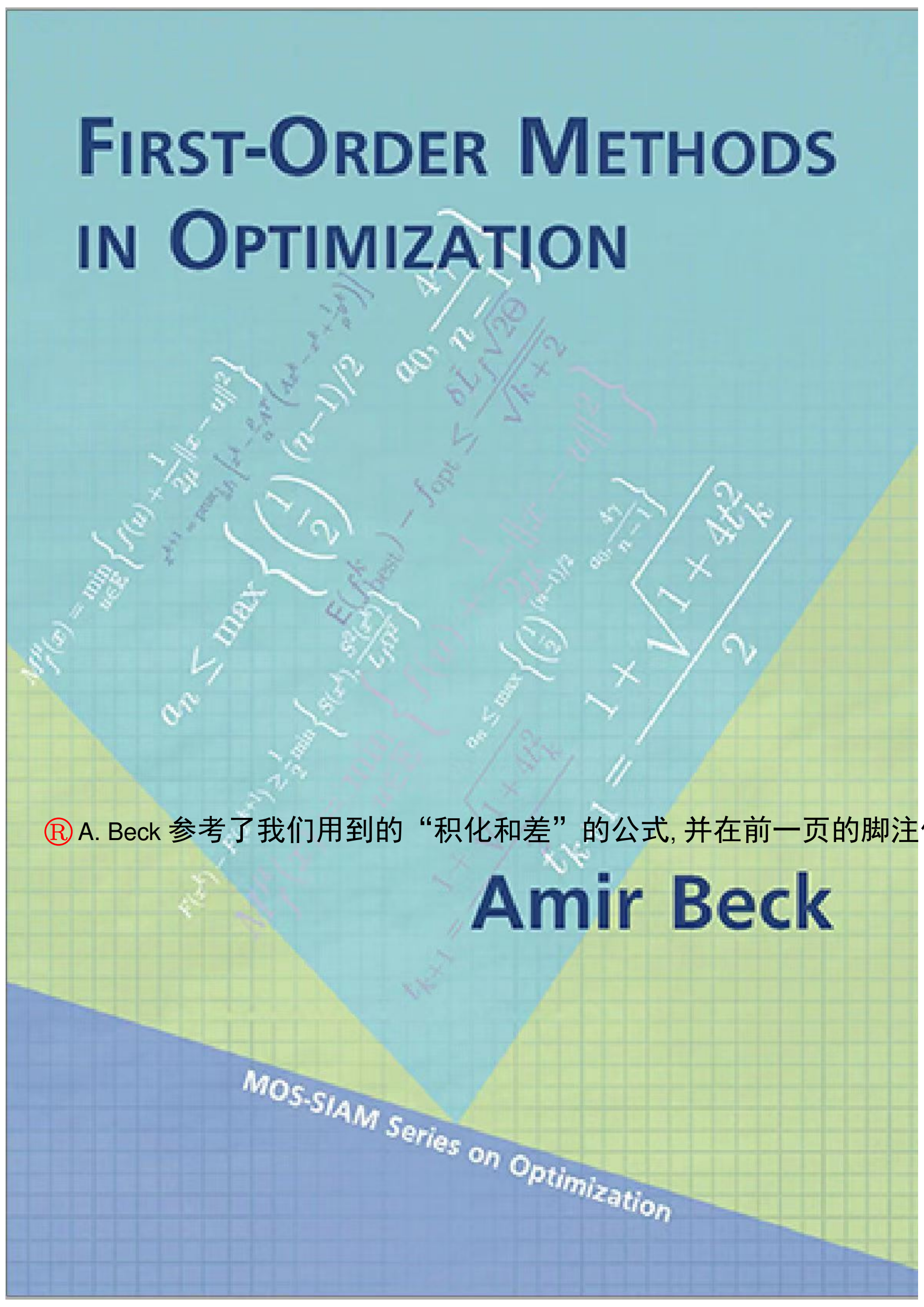
we thus obtain

$$\begin{aligned}&2(v-\tilde{v}^k)^TH(v^k-v^{k+1})\\ &\quad=\ \big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\big(\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2\big). \qquad (3.3)\end{aligned}$$

For the last term of (3.3), using $HM=Q$ and $2v^TQv=v^T(Q^T+Q)v$, we have

$$\begin{aligned}&\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2\\ &\quad=\ \|v^k-\tilde{v}^k\|_H^2-\|(v^k-\tilde{v}^k)-(v^k-v^{k+1})\|_H^2\\ &\quad\overset{(1.3a)}{=}\ \|v^k-\tilde{v}^k\|_H^2-\|(v^k-\tilde{v}^k)-M(v^k-\tilde{v}^k)\|_H^2\\ &\quad=\ 2(v^k-\tilde{v}^k)^THM(v^k-\tilde{v}^k)-(v^k-\tilde{v}^k)^TM^THM(v^k-\tilde{v}^k)\\ &\quad=\ (v^k-\tilde{v}^k)^T(Q^T+Q-M^THM)(v^k-\tilde{v}^k)\\ &\quad\overset{(1.3b)}{=}\ \|v^k-\tilde{v}^k\|_G^2. \qquad (3.4)\end{aligned}$$

Substituting (3.3), (3.4) in (3.2), the assertion of this theorem is proved. $\square$



Ⓡ A. Beck 参考了我们用到的“积化和差”的公式, 并在前一页的脚注做了说明

15.3. Convergence Analysis of AD-PMM 429

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 $\mathbf{y} \in \mathbb{R}^m$,

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

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

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

as well as

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

and

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

3.1 Convergence in a strictly contraction sense

定理 2 *Let $\{v^k\}$ be the sequence generated by a method for the problem* (1.1) *and $\tilde{w}^k$ is obtained in the k-th iteration. If v^k, v^{k+1} and $\tilde{w}^k$ satisfy the conditions in the unified framework, then we have*

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \quad (3.5)$$

Proof. Setting $w = w^*$ in (3.1), we get

$$\begin{aligned}
&\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\
&\geq \quad \|v^k - \tilde{v}^k\|_G^2 + 2\{\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k)\}. \quad (3.6)
\end{aligned}$$

By using the optimality of w^* and the monotonicity of $F(w)$, we have

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

and thus

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \geq \|v^k - \tilde{v}^k\|_G^2. \quad (3.7)$$

The assertion (3.5) follows directly. $\square$

定理 1 中的结论 (3.1)

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

是为收敛速率的证明而准备的.

否则, 我们可以通过在 (3.2)

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

中令 $w=w^*$, 得到

$$(v^k-v^{k+1})^T H(\tilde{v}^k-v^*)\geq 0. \tag{3.8}$$

将恒等式

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

用于 (3.8) 的左端, 令 $a=v^k, \ b=v^{k+1}, \ c=\tilde{v}^k$ 和 $d=v^*$, 我们得到

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

根据 (3.8) 就有

$$\|v^k-v^*\|_H^2-\|v^{k+1}-v^*\|_H^2\geq \|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2. \tag{3.9}$$

再把上式的右端化简一下,

$$\begin{aligned}&\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2\\ &\quad= \|v^k-\tilde{v}^k\|_H^2-\|(v^k-\tilde{v}^k)-(v^k-v^{k+1})\|_H^2\\ &\quad\overset{(1.2b)}{=} \|v^k-\tilde{v}^k\|_H^2-\|(v^k-\tilde{v}^k)-M(v^k-\tilde{v}^k)\|_H^2\\ &\quad= 2(v^k-\tilde{v}^k)^T HM(v^k-\tilde{v}^k)-(v^k-\tilde{v}^k)^T M^T HM(v^k-\tilde{v}^k)\\ &\quad= (v^k-\tilde{v}^k)^T(Q^T+Q-M^T HM)(v^k-\tilde{v}^k)\\ &\quad\overset{(1.3b)}{=} \|v^k-\tilde{v}^k\|_G^2.\end{aligned} \tag{3.10}$$

将 (3.10) 代入 (3.9) 就得到引理的结论. □

3.2 Convergence rate （两篇主要理论文章）

Convergence rate in an ergodic sense [10]

为了证明算法遍历意义下的迭代复杂性, 我们需要对变分不等式 (1.1) 的解集做新的刻

画. 由于 (1.1) 中的仿射算子 F 恰有

$$(w - w^*)^T F(w^*) = (w - w^*)^T F(w),$$

变分不等式问题

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

和

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

是等价的. 我们用后者定义变分不等式 (1.1) 的近似解. 对给定的 $\epsilon > 0$, 如果 $\tilde{w}$ 满足

$$\tilde{w} \in \Omega, \;\; \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \ge -\epsilon, \;\; \forall\, w \in \mathcal{D}_{(\tilde{w})}, \tag{3.11a}$$

其中

$$\mathcal{D}_{(\tilde{w})} = \{ w \in \Omega \,|\, \|w - \tilde{w}\| \le 1 \}, \tag{3.11b}$$

就叫做变分不等式 (1.1) 的 ϵ 近似解. 它可以等价地表示成

$$\tilde{w} \in \Omega, \quad \sup_{w \in \mathcal{D}_{(\tilde{w})}} \left\{ \theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w) \right\} \le \epsilon. \tag{3.12}$$

人们感兴趣的是：对给定的 $\epsilon > 0$, 经过多少次迭代得到一个 $\tilde{w} \in \Omega$, 使得 (3.12) 成立.

这就是我们要讨论的遍历意义下的收敛速率. 讨论遍历意义下的收敛性, 对 (1.3) 中的矩阵 H 和 G, 只要求它半正定.

Theorem 1 is also the base for the convergence rate proof. Using the monotonicity of F, we have

$$(w - \tilde{w}^k)^T F(w) = (w - \tilde{w}^k)^T F(\tilde{w}^k).$$

Substituting it in (3.1), we obtain

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2}\|v - v^k\|_H^2 \ge \frac{1}{2}\|v - v^{k+1}\|_H^2, \; \forall w \in \Omega. \tag{3.13}$$

Note that the above assertion is hold for $G \succeq 0$.

定理 3 *Let $\{v^k\}$ be the sequence generated by a method for the problem* (1.1) *and $\tilde{w}^k$ is obtained in the k-th iteration. Assume that v^k, v^{k+1} and $\tilde{w}^k$ satisfy the conditions in the unified framework and let $\tilde{w}_t$ be defined by*

$$\tilde{w}_t = \frac{1}{t+1} \sum_{k=0}^{t} \tilde{w}^k. \tag{3.14}$$

Then, for any integer number $t > 0$, $\tilde{w}_t \in \Omega$ and

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2(t+1)} \|v - v^0\|_H^2, \quad \forall w \in \Omega. \tag{3.15}$$

Proof. First, it holds that $\tilde{w}^k \in \Omega$ for all $k \geq 0$. Together with the convexity of Ω, (3.14) implies that $\tilde{w}_t \in \Omega$. Rewriting the inequality (3.13) in its equivalent form

$$\theta(\tilde{u}^k)-\theta(u)+(\tilde{w}^k - w)^T F(w) + \frac{1}{2}\|v - v^{k+1}\|_H^2 \leq \frac{1}{2}\|v - v^k\|_H^2, \ \forall w \in \Omega.$$

Summing the last inequality over $k = 0, 1, \ldots, t$, we obtain

$$\sum_{k=0}^{t} \theta(\tilde{u}^k)-(t+1)\theta(u)+\Big(\sum_{k=0}^{t} \tilde{w}^k-(t+1)w\Big)^T F(w) \leq \frac{1}{2}\|v-v^0\|_H^2, \quad \forall w \in \Omega.$$

Use the notation of $\tilde{w}_t$, it can be written as

$$\frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k) - \theta(u) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2\alpha(t+1)}\|v - v^0\|_H^2, \quad \forall w \in \Omega. \tag{3.16}$$

Since $\theta(u)$ is convex and

$$\tilde{u}_t = \frac{1}{t+1}\sum_{k=0}^{t}\tilde{u}^k,$$

we have

$$\theta(\tilde{u}_t) \leq \frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k).$$

Substituting it in (3.16), the assertion of this theorem follows directly. □

Recall (3.12). The conclusion (3.15) thus indicates obviously that the method is able to generate an approximate solution (i.e., $\tilde{w}_t$) with the accuracy $O(1/t)$ after t iterations. That is, in the case $G \succeq 0$, the convergence rate $O(1/t)$ of the method is established.

我们 2012 年发表在 SIAM Numerical Analysis 的论文 [10] 就是用这种方式证明了交替方向法在遍历意义下 $O(1/t)$ 的收敛速率. 这被认为我们在交替方向法方面的一个比较重要的结果, 这里只是说明, 该结果对符合统一框架收敛条件的方法都是成立的.

Convergence rate in a pointwise iteration-complexity [12]

我们 2015 年发表在 Numerische Mathematik 的论文 [12] 证明了交替方向法在点列意义下一个比较重要的结果.

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

这个性质已经被一些学者用来研发加速 ADMM. 下面证明这个结果对符合统一框架收敛条件的方法也都成立. 证明也只需要矩阵 H 和 G 半正定.

定理 4 *For solving the variational inequality* (1.1), *let* $\{w^k\}, \{\tilde{w}^k\}$ *be the sequence generated by* (1.2). *If the conditions* (1.3) *are satisfied, then we have*

$$\|v^{k+1} - v^{k+2}\|_H^2 \le \|v^k - v^{k+1}\|_H^2 - \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_G^2. \quad (3.17)$$

Proof Note that we have

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

and

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

Set the vector w in the above two inequalities by $\tilde{w}^{k+1}$ and $\tilde{w}^k$, respectively, we get

$$\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T F(\tilde{w}^k) \ge (\tilde{v}^{k+1} - \tilde{v}^k)^T Q(v^k - \tilde{v}^k)$$

and

$$\theta(\tilde{u}^k) - \theta(\tilde{u}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \ge (\tilde{v}^k - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}).$$

Adding the above two inequalities, it follows that

$$(\tilde{v}^k - \tilde{v}^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \ge 0.$$

Adding $\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\}^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\}$ to the both sides of the last inequality, we get

$$(v^k - v^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \ge \tfrac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T+Q)}^2,$$

and thus

$$(v^k - v^{k+1})^T H\{(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\} \ge \tfrac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T+Q)}^2. \quad (3.18)$$

Finally, by using $\|a\|_H^2 - \|b\|_H^2 = 2a^T H(a-b) - \|a-b\|_H^2$ and (3.18), we get

$$\begin{aligned}
&\|v^k - v^{k+1}\|_H^2 - \|v^{k+1} - v^{k+2}\|_H^2 \\
&= 2(v^k - v^{k+1})^T H\{(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\} \\
&\quad - \|(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\|_H^2 \\
&\ge \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T+Q)}^2 - \|(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\|_H^2 \\
&= \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T+Q-M^THM)}^2 \\
&= \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_G^2.
\end{aligned}$$

This is the equivalent form of (3.17) and the proof is complete. □

4 ADMM for problems with two separable blocks

This section concern the structured convex optimization problem namely,

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\ x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{4.1}$$

The Lagrangian function and the augmented Lagrange Function of (4.1) are

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

and

$$\mathcal{L}_\beta^{[2]}(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax + By - b) + \frac{\beta}{2}\|Ax + By - b\|^2, \tag{4.2}$$

respectively. Recall the model (4.1) can be explained as the VI

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{4.3a}$$

where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y), \quad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \tag{4.3b}$$

$$F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix}, \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m. \tag{4.3c}$$

Using the augmented Lagrange function, the recursion of the alternating direction method of multipliers for the structured convex optimization (4.1) can be written as

$$\begin{cases} x^{k+1} \in \text{Argmin}\{\mathcal{L}_\beta^{[2]}(x, y^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ y^{k+1} \in \text{Argmin}\{\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^k) \mid y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{cases} \tag{4.4}$$

Note that the essential variable of ADMM (4.4) is $v = (y, \lambda)$.

统一框架下的 ADMM. ADMM scheme (4.4) is also a special case which belongs to the unified algorithmic framework (1.2) and the Convergence Condition is satisfied.

In order to cast the ADMM scheme (4.4) into a special case of (1.2), let us first define the artificial vector $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ by

$$\tilde{x}^k = x^{k+1}, \quad \tilde{y}^k = y^{k+1} \quad \text{and} \quad \tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b), \tag{4.5}$$

where (x^{k+1}, y^{k+1}) is generated by the ADMM (4.4).

我们注意到 A. Beck 在他的专著 First-Order Methods in convex optimization [1], 也采用了这种转换.

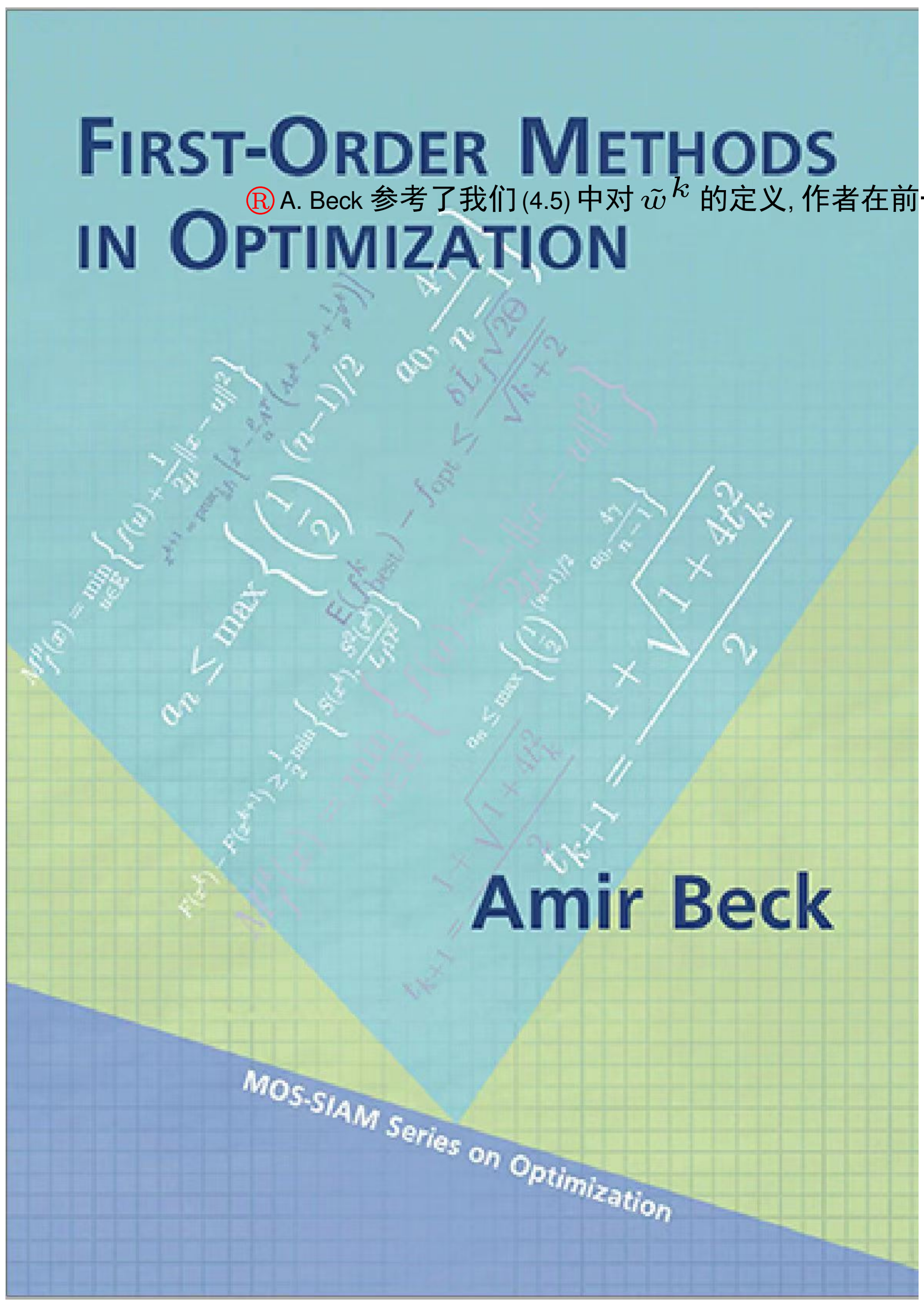
Prediction

$$\left\{\begin{array}{l} \tilde{x}^k \in \text{Argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k - b\|^2 \,|\, x \in \mathcal{X}\}, \\ \tilde{y}^k \in \text{Argmin}\{\theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2}\|A\tilde{x}^k + By - b\|^2 \,|\, y \in \mathcal{Y}\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b). \end{array}\right. \tag{4.6}$$

According to the scheme (4.4), the defined artificial vector $\tilde{w}^k$ satisfies the following VI: $\tilde{w}^k \in \Omega$,

$$\left\{\begin{array}{ll} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\tilde{\lambda}^k\} \ge 0, & \forall\, x \in \mathcal{X}, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{-B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k)\} \ge 0, & \forall\, y \in \mathcal{Y}, \\ \quad (A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + \dfrac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0. & \end{array}\right.$$

This can be written in form of (1.2a) as described in the following lemma.



Ⓡ A. Beck 参考了我们 (4.5) 中对 $\tilde{w}^k$ 的定义, 作者在前一页的脚注做了说明

15.3. Convergence Analysis of AD-PMM 429

We will use the following notation:

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

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

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

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

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

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 $\mathbf{x} \in \text{dom}(h_1)$, $\mathbf{z} \in \text{dom}(h_2)$, and $\mathbf{y} \in \mathbb{R}^m$,

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

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

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

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

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

引理 1 *For given* v^k, *let* w^{k+1} *be generated by* (4.4) *and* $\tilde{w}^k$ *be defined by* (4.5). *Then, we have*

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

where

$$Q = \begin{pmatrix} \beta B^T B & 0 \\ -B & \frac{1}{\beta} I \end{pmatrix}. \tag{4.7}$$

Recall the essential variable of the ADMM scheme (4.4) is (y, λ). Moreover, using the definition of $\tilde{w}^k$, the λ^{k+1} updated by (4.4) can be represented as

$$\begin{aligned} \lambda^{k+1} &= \lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k - b) \\ &= \lambda^k - \big[-\beta B(y^k - \tilde{y}^k) + \beta(A\tilde{x}^k + By^k - b)\big] \\ &= \lambda^k - \big[-\beta B(y^k - \tilde{y}^k) + (\lambda^k - \tilde{\lambda}^k)\big]. \end{aligned}$$

Therefore, the ADMM scheme (4.4) can be written as

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \tag{4.8a}$$

which corresponds to the step (1.2b) with

$$M = \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \quad \text{and} \quad \alpha = 1. \tag{4.8b}$$

验证收敛性条件. Now we check that the Convergence Condition is satisfied by the ADMM scheme (4.4). Indeed, for the matrix M in (4.8b), we have

$$M^{-1} = \begin{pmatrix} I & 0 \\ \beta B & I \end{pmatrix}.$$

Thus, by using (4.7) and (4.8b), we obtain 验证 H 半正定

$$H = QM^{-1} = \begin{pmatrix} \beta B^T B & 0 \\ -B & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} I & 0 \\ \beta B & I \end{pmatrix} = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}, \tag{4.9}$$

and consequently 验证 G 的半正定

$$\begin{aligned} G &= Q^T + Q - \alpha M^T H M = Q^T + Q - Q^T M \\ &= \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} \beta B^T B & -B^T \\ 0 & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \\ &= \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}. \end{aligned} \tag{4.10}$$

Therefore, H is symmetric and positive definite under the assumption that B is full column rank; and G is positive semi-definite. The Convergence Condition is satisfied; and thus the convergence of the ADMM scheme (4.4) is guaranteed. 见论文 [10]

收缩性和点列意义下的收敛速率. 我们将经典的 ADMM 按统一框架故意解释成预测-校正方法. 经过 (4.6) 预测以后, 再由

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k) \tag{4.11}$$

校正. 这符合统一框架的模式 (1.2). 在 (4.9) 和 (4.10) 中我们分别验证了矩阵 H 和 G 是

半正定的. 因此, 根据定理 4 就有

$$\|v^{k+1} - v^{k+2}\|_H \le \|v^k - v^{k+1}\|_H, \quad \forall\, k > 0. \tag{4.12}$$

由 (4.12), 对任意的正整数 $t > 0$,

$$\begin{aligned} \|v^t - v^{t+1}\|_H^2 &\le \frac{1}{t+1} \sum_{k=0}^{t} \|v^k - v^{k+1}\|_H^2 \\ &\le \frac{1}{t+1} \sum_{k=0}^{\infty} \|v^k - v^{k+1}\|_H^2 \\ &\overset{(4.12)}{\le} \frac{1}{t+1} \|v^0 - v^*\|_H^2. \end{aligned}$$

人们往往用 $\|v^t - v^{t+1}\|_H^2$ 的大小做停机准则的参考. 见论文 [12]

5 利用统一框架验证 ADMM 类算法的收敛性

5.1 交换顺序的交替方向法

将经典的 ADMM (4.4) 中求解 y-子问题和校正 λ 的顺序交换, 通过

$$\begin{cases} x^{k+1} & \in & \text{Argmin}\{\mathcal{L}_\beta^{[2]}(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1} + By^k - b), \\ y^{k+1} & \in & \text{Argmin}\{\mathcal{L}_\beta^{[2]}(x^{k+1}, y, \lambda^{k+1}) \,|\, y \in \mathcal{Y}\}, \end{cases} \tag{5.1}$$

得到的 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ 作为预测点, 然后

$$\begin{cases} y^{k+1} & := & y^k - \gamma(y^k - y^{k+1}), \\ \lambda^{k+1} & := & \lambda^k - \gamma(\lambda^k - \lambda^{k+1}). \end{cases} \quad \text{（松弛延拓）} \tag{5.2}$$

这里 $\gamma \in (0, 2)$. 赋值号 “:=” 表示 (5.2) 右端的 (y^{k+1}, λ^{k+1}) 是由算法的前半部分 (5.1) 产生的. (5.2) 左端才是下一步迭代开始所需要的 (y^{k+1}, λ^{k+1}). 对多数问题, 这样往往能加快收敛.

注意到, (5.1) 中核心变量还是 $v = (y, \lambda)$. 先把 (5.1) 产生的 w^{k+1} 本身定义成预测

点 $\tilde{w}^k$, 即

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k - \beta(Ax^{k+1} + By^k - b) \end{pmatrix}, \tag{5.3}$$

利用 $\mathcal{L}_\beta^{[2]}(x, y, \lambda)$ 的表达式, 交换顺序的交替方向法迭代公式 (5.1) 可以表示成

$$\begin{cases} \tilde{x}^k \in \text{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k - b\|^2 \,|\, x \in \mathcal{X}\}, & \text{(5.4a)} \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), & \text{(5.4b)} \\ \tilde{y}^k \in \text{argmin}\{\theta_2(y) - y^T B^T \tilde{\lambda}^k + \frac{\beta}{2}\|A\tilde{x}^k + By - b\|^2 \,|\, y \in \mathcal{Y}\}. & \text{(5.4c)} \end{cases}$$

我们用统一框架来证明交换顺序的交替方向法 (5.1)-(5.2) 的收敛性. 先给出由 (5.4) 求得的 $\tilde{w}^k$ 满足的形如 (1.2a) 的预测公式.

首先, 根据第一讲的定理 1, (5.4a) 的最优性条件是

$$\tilde{x}^k \in \mathcal{X}, \ \ \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\lambda^k + \beta A^T(A\tilde{x}^k + By^k - b)\} \ge 0, \ \forall x \in \mathcal{X}.$$

利用 $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b)$, 上式就是

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k}\} \ge 0, \ \ \forall x \in \mathcal{X}. \tag{5.5a}$$

类似地, 根据第一讲的定理 1, (5.4c) 的最优性条件是

$$\tilde{y}^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{-B^T\tilde{\lambda}^k + \beta B^T(A\tilde{x}^k + B\tilde{y}^k - b)\} \geq 0, \ \forall\, y \in \mathcal{Y}.$$

由于 $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b)$, 我们有

$$\begin{aligned} & -B^T\tilde{\lambda}^k + \beta B^T(A\tilde{x}^k + B\tilde{y}^k - b) \\ & \quad = \quad -B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) + \beta B^T(A\tilde{x}^k + By^k - b) \\ & \quad = \quad -B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) - B^T(\tilde{\lambda}^k - \lambda^k). \end{aligned}$$

因此, y-子问题 (5.4c) 的最优性条件是

$$\begin{aligned} \tilde{y}^k \in \mathcal{Y}, \ \ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{\underline{-B^T\tilde{\lambda}^k} + \beta B^T B(\tilde{y}^k - y^k) \\ - B^T(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \quad \forall y \in \mathcal{Y}. \end{aligned} \tag{5.5b}$$

对于 (5.4b) 中给出的 $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b)$, 可以表示成

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0,$$

也就是

$$\begin{aligned} \tilde{\lambda}^k \in \Re^m, \quad (\lambda - \tilde{\lambda}^k)^T\{\underline{(A\tilde{x}^k + B\tilde{y}^k - b)} \\ - B(\tilde{y}^k - y^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \quad \forall \lambda \in \Re^m. \end{aligned} \tag{5.5c}$$

将 (5.5a), (5.5b) 和 (5.5c) 组合在一起, 注意到下划线部分是 (4.3) 中的 $F(\tilde{w}^k)$, 我们得到下面的引理.

引理 2 求解变分不等式 (4.3), 对给定的 v^k, 由 (5.4) 提供的 $\tilde{w}^k$ 满足

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

其中

$$H = \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \tag{5.6}$$

这里的矩阵 H 是半正定的, 取 $\alpha = 1$ 的平凡校正, 就是方法 (5.1). 它的收敛效果与经典的 ADMM (4.4) 别无二致. 但是, 如果做平凡的松弛延伸, 让

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

就相当于方法 (5.1)-(5.1), 收敛速度往往会有 30% 的提高.

从理论上来讲, (5.6) 中的矩阵 H 即使在 B 列满秩的时候也是半正定的, 但这并不影响

计算和收敛性态. 当然, 我们也可以通过在子问题 (5.4c) 的目标函数中增添一项 $\frac{\delta}{2}\|y-y^k\|^2$, 就能使 相应的 H 矩阵变成

$$H=\begin{pmatrix}\beta B^TB+\delta I_{n_2} & -B^T\\ -B & \frac{1}{\beta}I_m\end{pmatrix}.$$

对任意的 $\beta,\delta>0$, 上面的 H 矩阵是正定的.

5.2 对称的交替方向法

人们习惯于用经典的乘子交替方向法 (4.4) 求解问题 (4.1). 从问题 (4.1) 本身看, 原始变量 x 和 y 是平等的, 在算法设计上平等对待 x 和 y 子问题, 也是最自然不过的考虑. 因此我们采用对称的交替方向法 [6], 它的 k 步迭代也是从给定的 (y^k,λ^k) 开始, 通过

$$\text{(S-ADMM)}\quad\left\{\begin{array}{ll}x^{k+1}\in\operatorname{argmin}\{\mathcal{L}_\beta^{[2]}(x,y^k,\lambda^k)\,|\,x\in\mathcal{X}\}, & \text{(5.7a)}\\ \lambda^{k+\frac12}=\lambda^k-\mu\beta(Ax^{k+1}+By^k-b), & \text{(5.7b)}\\ y^{k+1}\in\operatorname{argmin}\{\mathcal{L}_\beta^{[2]}(x^{k+1},y,\lambda^{k+\frac12})\,|\,y\in\mathcal{Y}\}, & \text{(5.7c)}\\ \lambda^{k+1}=\lambda^{k+\frac12}-\mu\beta(Ax^{k+1}+By^{k+1}-b). & \text{(5.7d)}\end{array}\right.$$

得到新的迭代点 $w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})$, 其中 $\mu\in(0,1)$ (通常取 $\mu=0.9$).

当 $\mu=1$ 时, 方法 (5.7) 是可以举出不收敛的反例的. 我们用统一框架证明对称型乘子交替方向法 (5.7) 的收敛性, 也是把方法 (5.7) 拆解成预测-校正两部分. 对由 (5.7) 产生的 x^{k+1} 和 y^{k+1}, 我们按如下方式定义预测点 $\tilde{w}^k$:

$$\tilde{w}^k=\begin{pmatrix}\tilde{x}^k\\ \tilde{y}^k\\ \tilde{\lambda}^k\end{pmatrix}=\begin{pmatrix}x^{k+1}\\ y^{k+1}\\ \lambda^k-\beta(Ax^{k+1}+By^k-b)\end{pmatrix}. \tag{5.8}$$

利用 $\mathcal{L}_\beta^{[2]}(x,y,\lambda)$ 的表达式, 对称型的乘子交替方向法迭代公式可以表示成等价的

$$\left\{\begin{array}{ll}\tilde{x}^k\in\operatorname{argmin}\{\theta_1(x)-x^TA^T\lambda^k+\frac{\beta}{2}\|Ax+By^k-b\|^2\,|\,x\in\mathcal{X}\}, & \text{(5.9a)}\\ \lambda^{k+\frac12}=\lambda^k-\mu\beta(A\tilde{x}^k+By^k-b), & \text{(5.9b)}\\ \tilde{y}^k\in\operatorname{argmin}\{\theta_2(y)-y^TB^T\lambda^{k+\frac12}+\frac{\beta}{2}\|A\tilde{x}^k+By-b\|^2\,|\,y\in\mathcal{Y}\}, & \text{(5.9c)}\\ \tilde{\lambda}^k=\lambda^k-\beta(A\tilde{x}^k+By^k-b). & \text{(5.9d)}\end{array}\right.$$

下面我们先找出 (5.9) 给出的 $\tilde{w}^k$ 在统一框架中形如 (1.2a) 的预测公式.

根据第一讲的定理 1, (5.9a) 的最优性条件是

$$\tilde{x}^k\in\mathcal{X},\ \ \theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T\{-A^T\lambda^k+\beta A^T(A\tilde{x}^k+By^k-b)\}\geq 0,\ \forall\, x\in\mathcal{X}.$$

利用 $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b)$, 上式就是

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k}\} \ge 0, \quad \forall\, x \in \mathcal{X}. \qquad (5.10a)$$

类似地, 根据第一讲的定理 1, (5.9c) 的最优性条件是

$$\tilde{y}^k \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{-B^T\lambda^{k+\frac{1}{2}} + \beta B^T(A\tilde{x}^k + B\tilde{y}^k - b)\} \ge 0, \; \forall\, y \in \mathcal{Y}.$$

利用 $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b)$, 我们有

$$\lambda^{k+\frac{1}{2}} = \lambda^k - \mu(\lambda^k - \tilde{\lambda}^k) = \tilde{\lambda}^k + (\mu - 1)(\tilde{\lambda}^k - \lambda^k),$$

和

$$\beta(A\tilde{x}^k + By^k - b) = -(\tilde{\lambda}^k - \lambda^k).$$

因此,

$$\begin{aligned} &-B^T\lambda^{k+\frac{1}{2}} + \beta B^T(A\tilde{y}^k + B\tilde{y}^k - b) \\ &\quad = \;\; -B^T[\tilde{\lambda}^k + (\mu - 1)(\tilde{\lambda}^k - \lambda^k)] + \beta B^T B(\tilde{y}^k - y^k) + \beta B^T(A\tilde{x}^k + By^k - b) \\ &\quad = \;\; -B^T\tilde{\lambda}^k + (1 - \mu)B^T(\tilde{\lambda}^k - \lambda^k) + \beta B^T B(\tilde{y}^k - y^k) - B^T(\tilde{\lambda}^k - \lambda^k) \\ &\quad = \;\; -B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) - \mu B^T(\tilde{\lambda}^k - \lambda^k). \end{aligned}$$

子问题 (5.9c) 的最优性条件是

$$\begin{aligned} \tilde{y}^k \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{\underline{-B^T\tilde{\lambda}^k} + \beta B^T B(\tilde{y}^k - y^k) \\ - \mu B^T(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \quad \forall y \in \mathcal{Y}. \end{aligned} \qquad (5.10b)$$

对于 (5.9d) 中定义的 $\tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k - b)$, 由于

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0,$$

可以表示成

$$\begin{aligned} \tilde{\lambda}^k \in \Re^m, \quad (\lambda - \tilde{\lambda}^k)^T\{\underline{(A\tilde{x}^k + B\tilde{y}^k - b)} \\ - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \quad \forall \lambda \in \Re^m. \end{aligned} \qquad (5.10c)$$

将 (5.10a), (5.10b) 和 (5.10c) 组合在一起并利用 (4.3) 中的记号, 我们有下面的引理.

引理 3 求解变分不等式 (4.3). 对给定的 v^k, 设 $\tilde{w}^k$ 是由 (5.9) 提供的, 则有

$$\tilde{w}^k \in \Omega, \;\; \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \qquad (5.11a)$$

其中

$$Q = \begin{pmatrix} \beta B^T B & -\mu B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \qquad (5.11b)$$

接着我们要导出形如 (1.2b) 的校正关系式. 利用 (5.8), 由 (5.7d) 给出的 λ^{k+1} 可以表示成

$$\begin{aligned}\lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - \mu\big[-\beta B(y^k - \tilde{y}^k) + \beta(A\tilde{x}^k + By^k - b)\big] \\ &= [\lambda^k - \mu(\lambda^k - \tilde{\lambda}^k)] - \mu\big[-\beta B(y^k - \tilde{y}^k) + \beta(Ax^{k+1} + By^k - b)\big] \\ &= \lambda^k - \big[-\mu\beta B(y^k - \tilde{y}^k) + 2\mu(\lambda^k - \tilde{\lambda}^k)\big]. \end{aligned} \tag{5.12}$$

跟 $y^{k+1} = \tilde{y}^k$ 结合在一起, 就有

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I & 0 \\ -\mu\beta B & 2\mu I_m \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}.$$

因此, 我们有下面的引理.

引理 4 求解变分不等式 (4.3). 对给定的 v^k, 设 w^{k+1} 由 (5.7) 提供. 那么对由 (5.8) 定义的 $\tilde{w}^k$, 我们有

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \tag{5.13a}$$

其中

$$M = \begin{pmatrix} I & 0 \\ -\mu\beta B & 2\mu I_m \end{pmatrix}. \tag{5.13b}$$

我们已经把对称的 ADMM (5.7) 拆解成 (5.11) 的预测和 (5.13) 的校正. 剩下来的事情就是根

据统一框架中的收敛性条件 (1.3) 验证算法的收敛性. 对 (5.13b) 中的矩阵 M, 算出

$$M^{-1} = \begin{pmatrix} I & 0 \\ \frac{1}{2}\beta B & \frac{1}{2\mu} I_m \end{pmatrix}.$$

由 $H = QM^{-1}$ 得到

$$H = \begin{pmatrix} \beta B^T B & -\mu B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} I & 0 \\ \frac{1}{2}\beta B & \frac{1}{2\mu} I_m \end{pmatrix} = \begin{pmatrix} (1 - \frac{1}{2}\mu)\beta B^T B & -\frac{1}{2} B^T \\ -\frac{1}{2} B & \frac{1}{2\mu\beta} I_m \end{pmatrix}.$$

因此

$$H = \frac{1}{2} \begin{pmatrix} \sqrt{\beta} B^T & 0 \\ 0 & \sqrt{\frac{1}{\beta}} I \end{pmatrix} \begin{pmatrix} (2-\mu) I & -I \\ -I & \frac{1}{\mu} I \end{pmatrix} \begin{pmatrix} \sqrt{\beta} B & 0 \\ 0 & \sqrt{\frac{1}{\beta}} I \end{pmatrix} \tag{5.14}$$

注意到

$$\begin{pmatrix} (2-\mu) & -1 \\ -1 & \frac{1}{\mu} \end{pmatrix} = \begin{cases} \succ 0, & \mu \in (0,1); \\ \succeq 0, & \mu = 1. \end{cases}$$

所以, 对所有的 $\mu \in (0,1)$, 当 B 列满秩时矩阵 H 是对称正定的.

再看矩阵$G=Q^T+Q-M^THM$. 因为$M^THM=M^TQ$, 由

$$M^TQ \;=\; \begin{pmatrix} I & -\mu\beta B^T \\ 0 & 2\mu I_m \end{pmatrix}\begin{pmatrix} \beta B^TB & -\mu B^T \\ -B & \frac{1}{\beta}I_m \end{pmatrix} = \begin{pmatrix} (1+\mu)\beta B^TB & -2\mu B^T \\ -2\mu B & 2\mu\frac{1}{\beta}I_m \end{pmatrix}.$$

得到

$$\begin{aligned} G &= (Q^T+Q)-M^THM \\ &= \begin{pmatrix} 2\beta B^TB & -(1+\mu)B^T \\ -(1+\mu)B & 2\frac{1}{\beta}I_m \end{pmatrix} - \begin{pmatrix} (1+\mu)\beta B^TB & -2\mu B^T \\ -2\mu B & 2\mu\frac{1}{\beta}I_m \end{pmatrix} \\ &= (1-\mu)\begin{pmatrix} \beta B^TB & -B^T \\ -B & \frac{2}{\beta}I_m \end{pmatrix}. \end{aligned} \tag{5.15}$$

同样, 对所有的$\mu\in(0,1)$, 当B列满秩时矩阵G正定. 所以, 根据统一框架(1.2a)-(1.2b), 方法是收敛的, 我们有下面的定理.

定理 5 求解变分不等式(4.3). 对给定的v^k, 设w^{k+1}由(5.7)提供. 我们有

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

其中$\tilde{w}^k$由(5.8)定义,

$$H=\begin{pmatrix} (1-\frac{1}{2}\mu)\beta B^TB & -\frac{1}{2}B^T \\ -\frac{1}{2}B & \frac{1}{2\mu\beta}I_m \end{pmatrix}$$

和

$$G=(1-\mu)\begin{pmatrix} \beta B^TB & -B^T \\ -B & \frac{2}{\beta}I_m \end{pmatrix}.$$

由于 $\mu\in(0,1)$, 矩阵H和G在B列满秩时都是正定的.

在矩阵B不一定列满秩的时候, 矩阵H和G半正定. 方法都具备定理3和定理4中的相关收敛速率性质.

References

[1] A. Beck, First-Order Methods in convex optimization, MOS-SIAM Series on Optimization, SIAM-Society for Industrial and Applied Mathematics, 2017.

[2] 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.

[3] 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.

[4] 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.

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

[6] 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.

[7] 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.

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

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] B.S. He and X.M. Yuan, On the optimal proximal parameter of an ADMM-like splitting method for separable convex programming. Mathematical methods in image processing and inverse problems, 139 – 163, Springer Proc. Math. Stat., 360, Springer, Singapore, 2021, and Optimization-Online 6235, 2017.

[14] Feng Ma, Yiming Bi and Bin Ga, A prediction – correction-based primal – dual hybrid gradient method for linearly constrained convex minimization, Numerical Algorithms (2019) 82:641 – 662.

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

[16] 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.

[17] 何炳生. 我和乘子交替方向法20年. 运筹学学报, 2018, 22: 1–31.

变分不等式框架下结构型凸优化的分裂收缩算法

V. 三个可分离块凸优化问题的分裂收缩方法

中学的数理基础	必要的社会实践
普通的大学数学	一般的优化原理

何 炳 生 南京大学数学系

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

天元数学东北中心 2023 年 10 月 17 – 27 日

1 Problem with three separable blocks

这一讲考虑三块可分离凸优化问题

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z) | Ax+By+Cz=b,\ x\in\mathcal{X}, y\in\mathcal{Y}, z\in\mathcal{Z}\} \tag{1.1}$$

的求解方法. 这个问题的拉格朗日函数是

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

问题 (1.1) 同样可以归结为变分不等式问题

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega, \tag{1.2a}$$

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

$$w=\begin{pmatrix} x\\ y\\ z\\ \lambda\end{pmatrix},\quad u=\begin{pmatrix} x\\ y\\ z\end{pmatrix},\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ -C^T\lambda\\ Ax+By+Cz-b\end{pmatrix}. \tag{1.2b}$$

相应的增广拉格朗日函数记为(与两个算子的符号有区别)

$$\begin{aligned}\mathcal{L}_\beta^{[3]}(x,y,z,\lambda) =& \theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^T(Ax+By+Cz-b)\\ &+\frac{\beta}{2}\|Ax+By+Cz-b\|^2.\end{aligned} \tag{1.3}$$

直接推广的 ADMM 求解三块可分离问题不保证收敛

对三个可分离块的凸优化问题, 采用直接推广的乘子交替方向法, 第 k 步迭代是从给定的 $v^k=(y^k,z^k,\lambda^k)$ 出发, 通过

$$\left\{\begin{array}{lcl} x^{k+1} & \in & \arg\min\{\mathcal{L}_\beta^{[3]}(x,y^k,z^k,\lambda^k)\mid x\in\mathcal{X}\},\\ y^{k+1} & \in & \arg\min\{\mathcal{L}_\beta^{[3]}(x^{k+1},y,z^k,\lambda^k)\mid y\in\mathcal{Y}\},\\ z^{k+1} & \in & \arg\min\{\mathcal{L}_\beta^{[3]}(x^{k+1},y^{k+1},z,\lambda^k)\mid z\in\mathcal{Z}\},\\ \lambda^{k+1} & = & \lambda^k-\beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b),\end{array}\right. \tag{1.4}$$

求得新的迭代点 $w^{k+1}=(x^{k+1},y^{k+1},z^{k+1},\lambda^{k+1})$. 当矩阵 A, B, C 中有两个是互相正交的时候, 用方法 (1.4) 求解问题 (1.1) 是收敛的 因为这种三块的可分离问题, 实际上相当于两块可分离的问题. 对一般的三块可分离问题, 是不能保证收敛的 [1].

值得继续研究的问题和猜想

譬如说, 三个可分离块的实际问题中, 线性约束矩阵

$\mathcal{A}=[A,B,C]$ 中, 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可分离块的问题, 既没有证明收敛, 也没有举出反例, 这仍然是一个有趣又特别有意义的问题！举个简单的例子来说吧:

- 经典的乘子交替方向法处理问题

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

- 将等式约束换成不等式约束, 问题就变成

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

- 再化成三个可分离块的等式约束问题就是

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

- 直接推广的乘子交替方向法 (1.4) 处理上面这种问题, 我们猜想是收敛的, 但是至今没有证明收敛性. 仍然是一个遗留的极具挑战性的问题！

在对直接推广的 ADMM (1.4) 证明不了收敛性的时候, 我们就着手对三块可分离的问题提出一些修正算法.

2 统一框架的等价表示

问题: $w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^\top F(w^*) \ge 0, \quad \forall w \in \Omega.$ (2.1)

[预测] 第 k-步迭代从给定的核心变量 v^k 开始, 求得预测点 $\tilde{w}^k$, 使得

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \ge (v - \tilde{v}^k)^\top Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \tag{2.2}$$

成立. 其中矩阵 $Q^\top + Q$ 是正定的. 左端将问题 (2.1) 的 w^* 换成了 $\tilde{w}^k$. 称 Q 为预测矩阵

[校正]. 根据预测得到的 $\tilde{v}^k$, 给出核心变量 v 的新迭代点 v^{k+1} 的公式为

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{2.3}$$

我们称 M 为校正矩阵. v 为核心变量, v 可以是 w, 也可以是 w 的部分分量

收敛性条件 对给定的预测矩阵 Q, 要求设计的校正矩阵 M 满足如下条件:

$$\exists \text{正定矩阵} \quad H \succ 0 \quad \text{使得} \quad HM = Q. \tag{2.4a}$$

此外, 能够保证

$$G = Q^\top + Q - M^\top HM \succ 0. \tag{2.4b}$$

校正 $v^{k+1} = v^k - M(v^k - \tilde{v}^k)$, 怎样给出满足收敛性条件的校正矩阵 M ?

$$\left\{\begin{array}{l} \text{预测 (2.2) 提供 } Q: Q^\top + Q \succ 0 \\ \text{收敛条件 (2.4): 选矩阵 } M \text{ 的要求:} \\ \exists H \succ 0, \text{ such that } HM = Q, \\ G = Q^\top + Q - M^\top HM \succ 0. \end{array}\right. \Longleftrightarrow \left\{\begin{array}{rcl} D \succ 0, & & G \succ 0, \\ D + G &=& Q^\top + Q, \\ M^\top HM &=& D, \\ HM &=& Q. \end{array}\right.$$

$$\Longleftrightarrow \left\{\begin{array}{rcl} D \succ 0, & & G \succ 0, \\ D + G &=& Q^\top + Q, \\ Q^\top M &=& D, \\ HM &=& Q. \end{array}\right. \Longleftrightarrow \left\{\begin{array}{rcl} D \succ 0, & & G \succ 0, \\ D + G &=& Q^\top + Q, \\ M &=& Q^{-T} D, \\ H &=& QD^{-1}Q^\top. \end{array}\right.$$

现在的做法: 有了预测矩阵 Q, 可以选定 D, 使其满足 $0 \prec D \prec Q^\top + Q$.

对给定的满足 $Q^\top + Q \succ 0$ 的预测, 从好不容易凑出一个方法, 到并不费劲构造一簇算法.

由于 $M = Q^{-T}D$, 校正 (2.3) 等价于 $\quad Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k).$ (2.5)

3 部分平行分裂的ADMM预测校正方法

这一节的方法源自2009年发表的[3], 把x当成中间变量, 迭代从$v^k=(y^k,z^k,\lambda^k)$到$v^{k+1}=(y^{k+1},z^{k+1},\lambda^{k+1})$, 只是平行处理$y$和$z$-子问题, 再更新$\lambda$. 换句话说, 把

$$\begin{cases} x^{k+1}\in\arg\min\{\theta_1(x)-x^TA^T\lambda^k+\frac{\beta}{2}\|Ax+By^k+Cz^k-b\|^2|x\in\mathcal{X}\}, \\ y^{k+1}\in\text{argmin}\{\theta_2(y)-y^TB^T\lambda^k+\frac{\beta}{2}\|Ax^{k+1}+By+Cz^k-b\|^2|y\in\mathcal{Y}\}, \\ z^{k+1}\in\text{argmin}\{\theta_3(z)-z^TC^T\lambda^k+\frac{\beta}{2}\|Ax^{k+1}+By^k+Cz-b\|^2|z\in\mathcal{Z}\}, \\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b) \end{cases} \tag{3.1}$$

生成的点$(x^{k+1},y^{k+1},z^{k+1},\lambda^{k+1})$当成预测点. 再把核心变量往回拉一点. 原因是$y$, z子问题平行处理, 包括据此更新的λ, 都太自由, 需要校正. 校正公式是

$$v^{k+1}:=v^k-\alpha(v^k-v^{k+1}),\quad \alpha\in(0,2-\sqrt{2}). \tag{3.2}$$

譬如说, 我们可以取$\alpha=0.55$. 注意到(3.2)右端的 $v^{k+1}=(y^{k+1},z^{k+1},\lambda^{k+1})$ 是由(3.1)提供的.

我们用统一框架来验证这个部分平行分裂的预测校正方法的收敛性. 先把由(3.1)生成的$(x^{k+1},y^{k+1},z^{k+1})$视为$(\tilde{x}^k,\ \tilde{y}^k,\ \tilde{z}^k)$, 并定义

$$\tilde{\lambda}^k=\lambda^k-\beta(A\tilde{x}^k+By^k+Cz^k-b). \tag{3.3}$$

这样, 预测点$(\tilde{x}^k,\tilde{y}^k,\tilde{z}^k,\tilde{\lambda}^k)$就可以看成由下式生成：

$$\begin{cases} \tilde{x}^k\in\arg\min\{\mathcal{L}_\beta^{[3]}(x,y^k,z^k,\lambda^k)\mid x\in\mathcal{X}\},, & (3.4a) \\ \tilde{y}^k\in\arg\min\{\mathcal{L}_\beta^{[3]}(\tilde{x}^k,y,z^k,\lambda^k)\mid y\in\mathcal{Y}\}, & (3.4b) \\ \tilde{z}^k\in\arg\min\{\mathcal{L}_\beta^{[3]}(\tilde{x}^k,y^k,z,\lambda^k)\mid z\in\mathcal{Z}\}, & (3.4c) \\ \tilde{\lambda}^k=\lambda^k-\beta(A\tilde{x}^k+By^k+Cz^k-b). & (3.4d) \end{cases}$$

利用增广拉格朗日函数(1.3), 子问题(3.4a)相当于

$$\tilde{x}^k=\text{argmin}\{\theta_1(x)-x^TA^T\lambda^k+\tfrac{1}{2}\beta\|Ax+By^k+Cz^k-b\|^2\mid x\in\mathcal{X}\},$$

根据最优性引理, $\tilde{x}^k\in\mathcal{X}$,

$$\theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T\{-A^T\lambda^k+\beta A^T(A\tilde{x}^k+By^k+Cz^k-b)\}\geq0,\quad\forall x\in\mathcal{X}.$$

再根据(3.4d), 就有

用(3.4d) 定义$\tilde{\lambda}^k$, 可以让(3.5a)的$-A^T\tilde{\lambda}^k$后面没有“尾巴”

$$\tilde{x}^k\in\mathcal{X},\quad\theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k}\}\geq0,\quad\forall x\in\mathcal{X}. \tag{3.5a}$$

子问题(3.4b)相当于

$$\tilde{y}^k=\text{argmin}\{\theta_2(y)-y^TB^T\lambda^k+\tfrac{1}{2}\beta\|A\tilde{x}^k+By+Cz^k-b\|^2\mid y\in\mathcal{Y}\},$$

同样根据最优性条件引理, 有 $\tilde{y}^k \in \mathcal{Y}$,

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{-B^T\lambda^k + \beta B^T(A\tilde{x}^k + B\tilde{y}^k - b)\} \geq 0, \quad \forall y \in \mathcal{Y}.$$

再根据 (3.4d), 就有

用 $\tilde{\lambda}^k$ 的定义, (3.5b) 中 $-B^T\tilde{\lambda}^k$ 后面的"尾巴"是 $\beta B^T B(\tilde{y}^k - y^k)$

$$\tilde{y}^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{\underline{-B^T\tilde{\lambda}^k} + \beta B^T B(\tilde{y}^k - y^k)\} \geq 0, \quad \forall y \in \mathcal{Y}. \tag{3.5b}$$

同理, 对子问题 (3.4c) 有

$$\tilde{z}^k \in \mathcal{Z}, \quad \theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^T\{\underline{-C^T\tilde{\lambda}^k} + \beta C^T C(\tilde{z}^k - z^k)\} \geq 0, \quad \forall z \in \mathcal{Z}. \tag{3.5c}$$

注意到 (3.4d) 可以写成

$$\underline{(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b)} - B(\tilde{y}^k - y^k) - C(\tilde{z}^k - z^k) + (1/\beta)\,(\tilde{\lambda}^k - \lambda^k) = 0. \tag{3.5d}$$

把 (3.5) 中的公式组合在一起, 可以写成统一框架中的预测形式:

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \forall w \in \Omega, \tag{3.6a}$$

其中

$$Q = \begin{pmatrix} \beta B^T B & 0 & 0 \\ 0 & \beta C^T C & 0 \\ -B & -C & \frac{1}{\beta} I \end{pmatrix}. \tag{3.6b}$$

回头来看方法 (3.1)-(3.2) 在统一框架中的校正该怎么表示. 由于

$$y^{k+1} = \tilde{y}^k, \quad z^{k+1} = \tilde{z}^k, \quad \text{和} \quad \lambda^{k+1} = \tilde{\lambda}^k + \beta B(y^k - \tilde{y}^k) + \beta C(z^k - \tilde{z}^k).$$

把 (3.4) 的输出作为预测点时, 校正公式 (3.2) 就可以表示成

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ z^k \\ \lambda^k \end{pmatrix} - \alpha \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\beta B & -\beta C & I \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ z^k - \tilde{z}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}.$$

也就是说, 利用了统一框架中 (3.6) 这样的预测表达式, 方法 (3.1)-(3.2) 的校正公式是

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \tag{3.7a}$$

其中

$$M=\alpha\begin{pmatrix} I & 0 & 0\\ 0 & I & 0\\ -\beta B & -\beta C & I\end{pmatrix}. \tag{3.7b}$$

对这样的 Q 和 M, 设

$$H=\frac{1}{\alpha}\begin{pmatrix} \beta B^TB & 0 & 0\\ 0 & \beta C^TC & 0\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix},$$

就有 $HM=Q$, 说明收敛性条件满足.

根据统一框架, 要对 (3.7b) 中的 M 找出一个 $\alpha>0$, 使得条件

$$G=(Q^T+Q)-M^THM\succ 0$$

满足. 简单的矩阵运算得到

$$Q^T+Q=\begin{pmatrix} 2\beta B^TB & 0 & -B^T\\ 0 & 2\beta C^TC & -C^T\\ -B & -C & \frac{2}{\beta}I\end{pmatrix}$$

和

$$\begin{aligned} M^TQ &= \alpha\begin{pmatrix} I & 0 & -\beta B^T\\ 0 & I & -\beta C^T\\ 0 & 0 & I\end{pmatrix}\begin{pmatrix} \beta B^TB & 0 & 0\\ 0 & \beta C^TC & 0\\ -B & -C & \frac{1}{\beta}I\end{pmatrix}\\ &= \alpha\begin{pmatrix} 2\beta B^TB & \beta B^TC & -B^T\\ \beta C^TB & 2\beta C^TC & -C^T\\ -B & -C & \frac{1}{\beta}I\end{pmatrix}. \end{aligned}$$

所以有

$$G=Q^T+Q-M^TQ=\begin{pmatrix} 2(1-\alpha)\beta B^TB & -\alpha\beta B^TC & -(1-\alpha)B^T\\ -\alpha C^TB & 2(1-\alpha)\beta C^TC & -(1-\alpha)C^T\\ -(1-\alpha)B & -(1-\alpha)C & (2-\alpha)\frac{1}{\beta}I_m\end{pmatrix}.$$

由于

$$G = \begin{pmatrix} \sqrt{\beta}B^T & 0 & 0 \\ 0 & \sqrt{\beta}C^T & 0 \\ 0 & 0 & \frac{1}{\sqrt{\beta}}I \end{pmatrix} \begin{pmatrix} 2(1-\alpha)I & -\alpha I & -(1-\alpha)I \\ -\alpha I & 2(1-\alpha)I & -(1-\alpha)I \\ -(1-\alpha)I & -(1-\alpha)I & (2-\alpha)I \end{pmatrix} \begin{pmatrix} \sqrt{\beta}B & 0 & 0 \\ 0 & \sqrt{\beta}C & 0 \\ 0 & 0 & \frac{1}{\sqrt{\beta}}I \end{pmatrix}.$$

只要验证, 对什么样的 $\alpha > 0$, 矩阵

$$\begin{pmatrix} 2(1-\alpha) & -\alpha & -(1-\alpha) \\ -\alpha & 2(1-\alpha) & -(1-\alpha) \\ -(1-\alpha) & -(1-\alpha) & (2-\alpha) \end{pmatrix} \succ 0. \tag{3.8}$$

经过计算, 对所有的 $\alpha \in (0, 2-\sqrt{2})$, (3.8) 中的矩阵正定, 收敛性条件满足.

4 带高斯回代的 ADMM 方法

Direct extension of ADMM

$$\begin{cases} x^{k+1} \in \arg\min\{\mathcal{L}_\beta^{[3]}(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ y^{k+1} \in \arg\min\{\mathcal{L}_\beta^{[3]}(x^{k+1}, y, z^k, \lambda^k) \mid y \in \mathcal{Y}\}, \\ z^{k+1} \in \arg\min\{\mathcal{L}_\beta^{[3]}(x^{k+1}, y^{k+1}, z, \lambda^k) \mid z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). \end{cases} \tag{4.1}$$

我们在 [1] 中证明, 对三个可分离块的凸优化问题, 直接推广的(4.1) 并不保证收敛.

在此之前, 我们好不容易凑成一些求解三个可分离块凸优化问题的方法 [5, 6]

直接推广的乘子交替方向法 (4.1) 对三个算子的问题不能保证收敛, 是因为它们处理有关核心变量的 y 和 z-子问题不公平. 采取补救的办法是将 (4.1) 提供的

$(y^{k+1}, z^{k+1}, \lambda^{k+1})$ 当成预测点, 校正公式为

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \\ \lambda^{k+1} \end{pmatrix} := \begin{pmatrix} y^{k} \\ z^{k} \\ \lambda^{k} \end{pmatrix} - \nu \begin{pmatrix} I & -(B^TB)^{-1}B^TC & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \\ \lambda^k - \lambda^{k+1} \end{pmatrix}. \quad (4.2)$$

其中 $\nu \in (0,1)$, 右端的 $(y^{k+1}, z^{k+1}, \lambda^{k+1})$ 是由 (4.1) 提供的. 这个方法发表在[5]. 想法是不公平, 就要做找补, 调整. 事实上, 也可以就用 (4.1) 提供的 λ^{k+1}, 只通过

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \end{pmatrix} := \begin{pmatrix} y^{k} \\ z^{k} \end{pmatrix} - \nu \begin{pmatrix} I & -(B^TB)^{-1}B^TC \\ 0 & I \end{pmatrix} \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix}. \quad (4.3)$$

校正 y 和 z (无需校正λ). 由于为下一步迭代只需要准备 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$, 我们只要做比(4.3)更简单的

$$\begin{pmatrix} By^{k+1} \\ Cz^{k+1} \end{pmatrix} := \begin{pmatrix} By^{k} \\ Cz^{k} \end{pmatrix} - \nu \begin{pmatrix} I & -I \\ 0 & I \end{pmatrix} \begin{pmatrix} By^k - By^{k+1} \\ Cz^k - Cz^{k+1} \end{pmatrix}. \quad (4.4)$$

4.1 The prediction matrix Q — Triangular Matrix

我们把直接推广 (4.1) 中的 $u^{k+1} = (x^{k+1}, y^{k+1}, z^{k+1})$ 写成 $\tilde{u}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k)$,

这样就有

$$\begin{cases} \tilde{x}^k \in \arg\min\{\mathcal{L}_\beta^{[3]}(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ \tilde{y}^k \in \arg\min\{\mathcal{L}_\beta^{[3]}(\tilde{x}^k, y, z^k, \lambda^k) \mid y \in \mathcal{Y}\}, \\ \tilde{z}^k \in \arg\min\{\mathcal{L}_\beta^{[3]}(\tilde{x}^k, \tilde{y}^k, z, \lambda^k) \mid z \in \mathcal{Z}\}. \end{cases} \quad (4.5)$$

x, y, z 子问题的形式是

$$\begin{cases} \tilde{x}^k \in \arg\min\{\theta_1(x) - x^TA^T\lambda^k + \frac{1}{2}\beta\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\}, \\ \tilde{y}^k \in \arg\min\{\theta_2(y) - y^TB^T\lambda^k + \frac{1}{2}\beta\|A\tilde{x}^k + By + Cz^k - b\|^2 \mid y \in \mathcal{Y}\}, \\ \tilde{z}^k \in \arg\min\{\theta_3(z) - z^TC^T\lambda^k + \frac{1}{2}\beta\|A\tilde{x}^k + B\tilde{y}^k + Cz - b\|^2 \mid z \in \mathcal{Z}\}. \end{cases}$$

利用优化问题和变分不等式之间等价关系的引理 1, 得到 $\tilde{u}^k \in \mathcal{U}$,

$$\left\{\begin{array}{l}\theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T\big\{-A^T\lambda^k \\ \qquad\qquad\qquad +\beta A^T\big(A\tilde{x}^k+By^k+Cz^k-b\big)\big\}\ge 0,\ \ \forall\, x\in\mathcal{X}, \\ \theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T\big\{-B^T\lambda^k \\ \qquad\qquad\qquad +\beta B^T\big(A\tilde{x}^k+B\tilde{y}^k+Cz^k-b\big)\big\}\ge 0,\ \ \forall\, y\in\mathcal{Y}, \\ \theta_3(z)-\theta_3(\tilde{z}^k)+(z-\tilde{z}^k)^T\big\{-C^T\lambda^k \\ \qquad\qquad\qquad +\beta C^T\big(A\tilde{x}^k+B\tilde{y}^k+C\tilde{z}^k-b\big)\big\}\ge 0,\ \ \forall\, z\in\mathcal{Z}.\end{array}\right. \tag{4.6}$$

定义

$$\tilde{\lambda}^k=\lambda^k-\beta\big(A\tilde{x}^k+By^k+Cz^k-b\big), \tag{4.7}$$

上式可以写成等价的等式

$$(A\tilde{x}^k+B\tilde{y}^k+C\tilde{z}^k-b)-B(\tilde{y}^k-y^k)-C(\tilde{z}^k-z^k)+\frac{1}{\beta}(\tilde{\lambda}^k-\lambda^k)=0. \tag{4.8}$$

对于给定的 $\tilde{\lambda}^k\in\Re^m$ 和 0 向量 p, 相应的关系式也可以写成

$$\tilde{\lambda}^k\in\Re^m,\quad (\lambda-\tilde{\lambda}^k)^Tp\ge 0,\quad \forall\lambda\in\Re^m.$$

将 (4.6) 和 (4.8) 加在一起, 利用变分不等式形式 (1.2), 我们得到 $\tilde{w}^k\in\Omega$,

$$\left\{\begin{array}{l}\theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T\big\{\underline{-A^T\tilde{\lambda}^k}\big\}\ge 0,\quad \forall\, x\in\mathcal{X}, \\ \theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T\big\{\underline{-B^T\tilde{\lambda}^k}+\beta B^TB(\tilde{y}^k-y^k)\big\}\ge 0,\quad \forall\, y\in\mathcal{Y}, \\ \theta_3(z)-\theta_3(\tilde{z}^k)+(z-\tilde{z}^k)^T\left\{\begin{array}{c}\underline{-C^T\tilde{\lambda}^k}+\beta C^TB(\tilde{y}^k-y^k) \\ +\beta C^TC(\tilde{z}^k-z^k)\end{array}\right\}\ge 0,\ \ \forall\, z\in\mathcal{Z}, \\ (\lambda-\tilde{\lambda}^k)^T\left\{\begin{array}{l}\underline{(A\tilde{x}^k+B\tilde{y}^k+C\tilde{z}^k-b)} \\ -B(\tilde{y}^k-y^k)-C(\tilde{z}^k-z^k)+\frac{1}{\beta}(\tilde{\lambda}^k-\lambda^k)\end{array}\right\}\ge 0,\ \ \forall\,\lambda\in\Lambda.\end{array}\right. \tag{4.9}$$

注意到 (4.9) 式中加下划线的部分恰好是 (1.2) 中定义的 $F(\tilde{w}^k)$, 合并写成 $\tilde{w}^k\in\Omega$,

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(v-\tilde{v}^k)^TQ(v^k-\tilde{v}^k),\ \forall w\in\Omega, \tag{4.10}$$

其中向量 $v=(y,z,\lambda)$ 预测矩阵

$$Q=\left(\begin{array}{ccc}\beta B^TB & 0 & 0 \\ \beta C^TB & \beta C^TC & 0 \\ -B & -C & \frac{1}{\beta}I_m\end{array}\right). \tag{4.11}$$

校正: 利用这样的预测点, 只校正 y 和 z 的公式 (4.3) (注意λ^{k+1} 和 $\tilde{\lambda}^k$ 的关系) 就可以写成

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^{k} \\ z^{k} \\ \lambda^{k} \end{pmatrix} - \begin{pmatrix} \nu I & -\nu(B^TB)^{-1}B^TC & 0 \\ 0 & \nu I & 0 \\ -\beta B & -\beta C & I \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ z^k - \tilde{z}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}.$$

也就是说, 在统一框架的校正公式中, 取

$$M = \begin{pmatrix} \nu I & -\nu(B^TB)^{-1}B^TC & 0 \\ 0 & \nu I & 0 \\ -\beta B & -\beta C & I \end{pmatrix}. \tag{4.12}$$

对于矩阵

$$H = \begin{pmatrix} \frac{1}{\nu}\beta B^TB & \frac{1}{\nu}\beta B^TC & 0 \\ \frac{1}{\nu}\beta C^TB & \frac{1}{\nu}\beta[C^TC + C^TB(B^TB)^{-1}B^TC] & 0 \\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix}, \tag{4.13}$$

可以验证$HM = Q$. 通过合同变换

$$\begin{pmatrix} I & 0 \\ -C^TB(B^TB)^{-1} & I \end{pmatrix} \begin{pmatrix} B^TB & B^TC \\ C^TB & C^TC + C^TB(B^TB)^{-1}B^TC \end{pmatrix} \begin{pmatrix} I & -(B^TB)^{-1}B^TC \\ 0 & I \end{pmatrix}$$

$$= \begin{pmatrix} B^TB & B^TC \\ 0 & C^TC \end{pmatrix} \begin{pmatrix} I & -(B^TB)^{-1}B^TC \\ 0 & I \end{pmatrix} = \begin{pmatrix} B^TB & 0 \\ 0 & C^TC \end{pmatrix},$$

得知H在B, C列满秩时正定. 此外,

$$\begin{aligned} G &= (Q^T + Q) - M^THM = (Q^T + Q) - M^TQ \\ &= \begin{pmatrix} 2\beta B^TB & \beta B^TC & -B^T \\ \beta C^TB & 2\beta C^TC & -C^T \\ -B & -C & \frac{2}{\beta}I \end{pmatrix} - \begin{pmatrix} (1+\nu)\beta B^TB & \beta B^TC & -B^T \\ \beta C^TB & (1+\nu)\beta C^TC & -C^T \\ -B & -C & \frac{1}{\beta}I \end{pmatrix} \\ &= \begin{pmatrix} (1-\nu)\beta B^TB & 0 & 0 \\ 0 & (1-\nu)\beta C^TC & 0 \\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix}. \end{aligned}$$

由于$\nu \in (0, 1)$, 当B, C列满秩时矩阵G正定. 统一框架中的收敛性条件满足.

5 Implement the correction by using (2.5)

对 (4.11) 中的预测矩阵 Q, 我们有

$$
\begin{aligned}
Q^T + Q &= \begin{pmatrix} 2\beta B^TB & \beta B^TC & -B^T \\ \beta C^TB & 2\beta C^TC & -C^T \\ -B & -C & \frac{2}{\beta}I_m \end{pmatrix} \\
&= \begin{pmatrix} B^T & 0 & 0 \\ 0 & C^T & 0 \\ 0 & 0 & I_m \end{pmatrix} \begin{pmatrix} 2\beta I_m & \beta I_m & -I_m \\ \beta I_m & 2\beta I_m & -I_m \\ -I_m & -I_m & \frac{2}{\beta}I_m \end{pmatrix} \begin{pmatrix} B & 0 & 0 \\ 0 & C & 0 \\ 0 & 0 & I_m \end{pmatrix}. \qquad (5.1)
\end{aligned}
$$

由于

$$
\begin{pmatrix} 2\beta I_m & \beta I_m & -I_m \\ \beta I_m & 2\beta I_m & -I_m \\ -I_m & -I_m & \frac{2}{\beta}I_m \end{pmatrix} = \begin{pmatrix} \beta I_m & \beta I_m & -I_m \\ \beta I_m & \beta I_m & -I_m \\ -I_m & -I_m & \frac{1}{\beta}I_m \end{pmatrix} + \begin{pmatrix} \beta I_m & 0 & 0 \\ 0 & \beta I_m & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix}
$$

是正定矩阵, 当矩阵 B 和 C 是列满秩矩阵时, 矩阵 $Q^T + Q$ 正定。

选择 $0 \prec D \prec Q^T + Q$, 可以提出自己想要的方法, 下面只是一些例子而已.

取一个比较简单的 D

对任意的 $v \in (0,1)$, 矩阵

$$
\begin{pmatrix} 2\beta I_m & \beta I_m & -I_m \\ \beta I_m & 2\beta I_m & -I_m \\ -I_m & -I_m & \frac{2}{\beta}I_m \end{pmatrix} = \begin{pmatrix} \nu\beta I_m & 0 & 0 \\ 0 & \nu\beta I_m & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} + \begin{pmatrix} (2-\nu)\beta I_m & \beta I_m & -I_m \\ \beta I_m & (2-\nu)\beta I_m & -I_m \\ -I_m & -I_m & \frac{1}{\beta}I_m \end{pmatrix}
$$

分拆成了两个正定矩阵. 因此, 可以选

$$
\begin{aligned}
D &= \begin{pmatrix} B^T & 0 & 0 \\ 0 & C^T & 0 \\ 0 & 0 & I_m \end{pmatrix} \begin{pmatrix} \nu\beta I & 0 & 0 \\ 0 & \nu\beta I & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} B & 0 & 0 \\ 0 & C & 0 \\ 0 & 0 & I_m \end{pmatrix} \\
&= \begin{pmatrix} \nu\beta B^TB & 0 & 0 \\ 0 & \nu\beta C^TC & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix}. \qquad (5.2)
\end{aligned}
$$

这时,

$$G = Q^T + Q - D = \begin{pmatrix} (2-\nu)\beta B^T B & \beta B^T C & -B^T \\ \beta C^T B & (2-\nu)\beta C^T C & -C^T \\ -B & -C & \frac{1}{\beta} I_m \end{pmatrix}. \quad (5.3)$$

Algorithms for the model (1.1)

[Prediction Step.] Obtain $(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k)$ via the direct extension of the ADMM (4.5) and define $\tilde{\lambda}^k$ by (4.7).

[Correction Step.] Get v^{k+1} by solving $Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k)$.

问题归结为如何从 $Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k)$ 求出 v^{k+1} ? 我们知道

$$Q^T = \begin{pmatrix} \beta B^T B & \beta B^T C & -B^T \\ 0 & \beta C^T C & -C^T \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix} = \begin{pmatrix} \beta B^T & 0 & 0 \\ 0 & \beta C^T & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} B & C & -\frac{1}{\beta} I_m \\ 0 & C & -\frac{1}{\beta} I_m \\ 0 & 0 & I_m \end{pmatrix},$$

和

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

对矩阵 Q^T 和 D 的分解有相同的左因子. 因此, 求解方程组

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

可以通过

$$\begin{pmatrix} B & C & -\frac{1}{\beta} I_m \\ 0 & C & -\frac{1}{\beta} I_m \\ 0 & 0 & I_m \end{pmatrix} \begin{pmatrix} y^{k+1} - y^k \\ z^{k+1} - z^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} = \begin{pmatrix} \nu B & 0 & 0 \\ 0 & \nu C & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} \tilde{y}^k - y^k \\ \tilde{z}^k - z^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix}$$

求得. 上述线性方程组等价于方程组

$$\begin{pmatrix} I & I & -\frac{1}{\beta} I \\ 0 & I & -\frac{1}{\beta} I \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} By^{k+1} - By^k \\ Cz^{k+1} - Cz^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} = \begin{pmatrix} \nu I & 0 & 0 \\ 0 & \nu I & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} B\tilde{y}^k - By^k \\ C\tilde{z}^k - Cz^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix}.$$

用回代的方法依次求得 $(\lambda^{k+1}-\lambda^k)$, $(Cz^{k+1}-Cz^k)$, $(By^{k+1}-By^k)$,

然后得到开始下一次迭代所需要的 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$.

选择 D 的一些其他方法

将 (5.2) 和 (5.3) 中的 D 和 G 互换位置, 换句话说, 取

$$D = \begin{pmatrix} (2-\nu)\beta B^T B & \beta B^T C & -B^T \\ \beta C^T B & (2-\nu)\beta C^T C & -C^T \\ -B & -C & \frac{1}{\beta} I_m \end{pmatrix}.$$

对于同样的预测, 校正可以通过

$$\begin{pmatrix} I & I & -\frac{1}{\beta} I \\ 0 & I & -\frac{1}{\beta} I \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} By^{k+1} - By^k \\ Cz^{k+1} - Cz^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} = \begin{pmatrix} (2-\nu)I & I & -I \\ I & (2-\nu)I & -I \\ -I & -I & I \end{pmatrix} \begin{pmatrix} B\tilde{y}^k - By^k \\ C\tilde{z}^k - Cz^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix}.$$

得到开始下一次迭代所需要的 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$.

选择 $D = \alpha(Q^T + Q)$, $\alpha \in (0,1)$ 的方法

这时 $D = \alpha(Q^T + Q)$ 和 $G = (1-\alpha)(Q^T + Q)$ 都是正定矩阵.

$$D = \alpha\big[Q^T + Q\big] = \alpha \begin{pmatrix} 2\beta B^T B & \beta B^T C & -B^T \\ \beta C^T B & 2\beta C^T C & -C^T \\ -B & C & \frac{2}{\beta} I_m \end{pmatrix}. \tag{5.5}$$

对于同样的预测, 校正可以通过

$$\begin{pmatrix} I & I & -\frac{1}{\beta} I \\ 0 & I & -\frac{1}{\beta} I \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} By^{k+1} - By^k \\ Cz^{k+1} - Cz^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} = \alpha \begin{pmatrix} 2I & I & -I \\ I & 2I & -I \\ -I & -I & 2I \end{pmatrix} \begin{pmatrix} B\tilde{y}^k - By^k \\ C\tilde{z}^k - Cz^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix}.$$

得到开始下一次迭代所需要的 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$.

从好不容易凑出一个方法, 到并不费力给出一簇算法.

6 部分平行并加正则项的 ADMM 方法

我们已经知道直接推广的 ADMM 求解三个可分离块的凸优化问题不能保证收敛 [1], 原因应该是对原始核心变量中 y 和 z 的子问题处理先后显得不够公平, 在 §4 采用了回代的方法. 然而, 如下的简单强制平行的方法也不能保证收敛.

简单地强制 y 和 z 平等不能保证方法收敛

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

下面我们考虑强制平行, 并通过另加正则项直接解决问题

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

$$\left\{\begin{array}{l} x^{k+1}=\arg\min\left\{\mathcal{L}_{\beta}^{3}(x,y^k,z^k,\lambda^k)\mid x\in\mathcal{X}\right\}, \qquad (\tau>0 \text{ 为参数})\\ y^{k+1}=\arg\min\left\{\mathcal{L}_{\beta}^{3}(x^{k+1},y,z^k,\lambda^k)+\frac{\tau}{2}\beta\|B(y-y^k)\|^2 | y\in\mathcal{Y}\right\},\\ z^{k+1}=\arg\min\left\{\mathcal{L}_{\beta}^{3}(x^{k+1},y^k,z,\lambda^k)+\frac{\tau}{2}\beta\|C(z-z^k)\|^2 | z\in\mathcal{Z}\right\},\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b).\end{array}\right.$$

上述做法相当于

$$\left\{\begin{array}{l} x^{k+1}\in\arg\min\left\{\theta_1(x)-x^TA^T\lambda^k+\frac{\beta}{2}\|Ax+By^k+Cz^k-b\|^2\,\big|\,x\in\mathcal{X}\right\},\\ y^{k+1}\in\arg\min\left\{\begin{array}{c}\theta_2(y)-y^TB^T\lambda^k+\frac{\beta}{2}\|Ax^{k+1}+By+Cz^k-b\|^2\\ +\frac{\tau}{2}\beta\|B(y-y^k)\|^2\end{array}\Bigg|\,y\in\mathcal{Y}\right\},\\ z^{k+1}\in\arg\min\left\{\begin{array}{c}\theta_3(z)-z^TC^T\lambda^k+\frac{\beta}{2}\|Ax^{k+1}+By^k+Cz-b\|^2\\ +\frac{\tau}{2}\beta\|C(z-z^k)\|^2\end{array}\Bigg|\,z\in\mathcal{Z}\right\},\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b),\end{array}\right.$$

注意到

$$\begin{array}{rcl} y^{k+1} & \in & \arg\min\left\{\begin{array}{c}\theta_2(y)-y^TB^T\lambda^k+\frac{\beta}{2}\|Ax^{k+1}+By+Cz^k-b\|^2\\ +\frac{\tau}{2}\beta\|B(y-y^k)\|^2\end{array}\Bigg|\,y\in\mathcal{Y}\right\},\\ & = & \arg\min\left\{\begin{array}{c}\theta_2(y)+\frac{\beta}{2}\|(Ax^{k+1}+By^k+Cz^k-b)+B(y-y^k)\|^2\\ -y^TB^T\lambda^k+\frac{\tau}{2}\beta\|B(y-y^k)\|^2\end{array}\Bigg|\,y\in\mathcal{Y}\right\}\\ & = & \arg\min\left\{\begin{array}{c}\theta_2(y)-y^TB^T[\lambda^k-\beta(Ax^{k+1}+By^k+Cz^k-b)]\\ +\frac{1}{2}\beta\|B(y-y^k)\|^2+\frac{\tau}{2}\beta\|B(y-y^k)\|^2\end{array}\Bigg|\,y\in\mathcal{Y}\right\}.\end{array}$$

所以, 若令

$$\lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b),$$

这个方法就是

$$\begin{cases} x^{k+1} \in \operatorname{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 | x \in \mathcal{X}\}, \\ \quad \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b) \\ y^{k+1} \in \operatorname{argmin}\{\theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} + \frac{\mu\beta}{2}\|B(y - y^k)\|^2 \mid y \in \mathcal{Y}\}, \\ z^{k+1} \in \operatorname{argmin}\{\theta_3(z) - z^T C^T \lambda^{k+\frac{1}{2}} + \frac{\mu\beta}{2}\|C(z - z^k)\|^2 \mid z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), \end{cases} \tag{6.1}$$

其中 $\mu = \tau + 1$. 我们讨论需要多大的 μ.

把由 (6.1) 生成的

$$(x^{k+1}, y^{k+1}, z^{k+1}, \lambda^{k+\frac{1}{2}}) \qquad \text{视为预测点} \qquad (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k, \tilde{\lambda}^k), \tag{6.2}$$

这个预测公式就成为

$$\begin{cases} \tilde{x}^k = \operatorname{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\}, \\ \tilde{y}^k = \operatorname{argmin}\{\theta_2(y) - y^T B^T \tilde{\lambda}^k + \frac{\mu\beta}{2}\|B(y - y^k)\|^2 \mid y \in \mathcal{Y}\}, \\ \tilde{z}^k = \operatorname{argmin}\{\theta_3(z) - z^T C^T \tilde{\lambda}^k + \frac{\mu\beta}{2}\|C(z - z^k)\|^2 \mid z \in \mathcal{Z}\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k + Cz^k - b). \end{cases} \tag{6.3}$$

预测 (6.3) 中 x-子问题的最优性条件是

$$\tilde{x}^k = \operatorname{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \tfrac{1}{2}\beta\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\},$$

根据最优性引理, $\tilde{x}^k \in \mathcal{X}$,

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\lambda^k + \beta A^T(A\tilde{x}^k + By^k + Cz^k - b)\} \ge 0, \ \ \forall x \in \mathcal{X}.$$

再根据 $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k + Cz^k - b)$, 就有

$$\tilde{x}^k \in \mathcal{X}, \quad \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k}\} \ge 0, \ \ \forall x \in \mathcal{X}. \tag{6.4a}$$

同样根据最优性条件引理, 预测 (6.3) 中 y-子问题的最优性条件是

$$\tilde{y}^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{ \underline{-B^T \tilde{\lambda}^k} + \mu\beta B^T B(\tilde{y}^k - y^k) \} \ge 0, \quad \forall y \in \mathcal{Y}. \tag{6.4b}$$

同理, 预测 (6.3) 中 z-子问题的最优性条件是

$$\tilde{z}^k \in \mathcal{Z}, \quad \theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^T \{ \underline{-C^T \tilde{\lambda}^k} + \mu\beta C^T C(\tilde{z}^k - z^k) \} \ge 0, \quad \forall z \in \mathcal{Z}. \tag{6.4c}$$

根据 $\tilde{\lambda}^k$ 的定义, 我们有

$$\underline{(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b)} - B(\tilde{y}^k - y^k) - C(\tilde{z}^k - z^k) + (1/\beta)\,(\tilde{\lambda}^k - \lambda^k) = 0. \tag{6.4d}$$

这样, 利用最优性引理和变分不等式 (1.2) 的形式, 预测就可以写成统一框架中的形式:

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \forall w \in \Omega, \tag{6.5a}$$

其中

$$Q = \begin{pmatrix} \mu\beta B^T B & 0 & 0 \\ 0 & \mu\beta C^T C & 0 \\ -B & -C & \frac{1}{\beta} I \end{pmatrix}. \tag{6.5b}$$

由于 $\tilde{\lambda}^k = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b)$ 和

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

利用这样的预测点, 校正 y 和 z 的公式 (注意 λ^{k+1} 和 $\tilde{\lambda}^k$ 的关系) 就可以写成

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ z^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\beta B & -\beta C & I \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ z^k - \tilde{z}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}.$$

也就是说, 在统一框架的校正公式中

$$M = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\beta B & -\beta C & I \end{pmatrix}. \tag{6.6}$$

对于矩阵

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

可以验证 H 正定并有

$$\begin{aligned} HM &= \begin{pmatrix} \mu\beta B^T B & 0 & 0 \\ 0 & \mu\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\beta B & -\beta C & I \end{pmatrix} \\ &= \begin{pmatrix} \mu\beta B^T B & 0 & 0 \\ 0 & \mu\beta C^T C & 0 \\ -B & -C & \frac{1}{\beta} I \end{pmatrix} = Q \end{aligned}$$

此外,

$$\begin{aligned} G &= (Q^T + Q) - M^T H M = (Q^T + Q) - M^T Q \\ &= \begin{pmatrix} 2\mu\beta B^T B & 0 & -B^T \\ 0 & 2\mu\beta C^T C & -C^T \\ -B & -C & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} (1+\mu)\beta B^T B & \beta B^T C & -B^T \\ \beta C^T B & (1+\mu)\beta C^T C & -C^T \\ -B & -C & \frac{1}{\beta} I \end{pmatrix} \\ &= \begin{pmatrix} (\mu-1)\beta B^T B & -\beta B^T C & 0 \\ -\beta C^T B & (\mu-1)\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix}. \end{aligned}$$

由于 $\mu > 2$, 矩阵 G 正定, 收敛性条件满足. 方法的收敛性得到证明.

例如, 可以取 $\mu = 2.01$. 这类发表在 [6, 8] 的算法思想是: 让 y 和 z 各自独立, 又不准备校正, 那就预先加正则项让它们不致走得太远. [6] 中的方法被 UCLA Osher 教授的课题组成功用来求解图像降维问题 [2].

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

$$\min_{T\geq 0, V_j \in D_j, e\in E} \zeta \sum_i \max_j (T_{i,j}) + \langle R_w \sigma C_w, T\rangle$$
$$\text{such that} \quad YT - X_s = V - X_s \text{diag}(e). \quad (15)$$

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

ADMM + Parallel-Prox Splitting ALM

各自为政, 过分自由. 给它们加个适当的正则项($\tau > 1$), 方法就能保证收敛.

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, & \text{(6.7a)} \\ \begin{cases} y^{k+1} = \arg\min\{\mathcal{L}(x^{k+1}, y, z^k, \lambda^k) + \frac{\tau}{2}\|B(y - y^k)\|^2 \mid y \in \mathcal{Y}\}, \\ z^{k+1} = \arg\min\{\mathcal{L}(x^{k+1}, y^k, z, \lambda^k) + \frac{\tau}{2}\|C(z - z^k)\|^2 \mid z \in \mathcal{Z}\}, \end{cases} & \text{(6.7b)} \\ \lambda^{k+1} = \lambda^k - (Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). & \text{(6.7c)} \end{cases}$$

Notice that (6.7b) can be written as

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \end{pmatrix} = \arg\min\left\{\mathcal{L}(x^{k+1}, y, z, \lambda^k) + \frac{1}{2}\left\| \begin{matrix} y - y^k \\ z - z^k \end{matrix} \right\|^2_{D_{BC}} \left| \begin{matrix} y \in \mathcal{Y} \\ z \in \mathcal{Z} \end{matrix} \right. \right\},$$

where

$$D_{BC} = \begin{pmatrix} \tau B^T B & -B^T C \\ -C^T B & \tau C^T C \end{pmatrix}. \tag{6.8}$$

D_{BC} is positive semidefinite when $\tau \geq 1$.

However, the matrix D_{BC} is indefinite for $\tau \in (0, 1)$.

In other words, the scheme (6.7) can be rewritten as

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}(x, y^k, z^k, \lambda^k) \mid x \in \mathcal{X}\}, \\ \begin{pmatrix} y^{k+1} \\ z^{k+1} \end{pmatrix} = \arg\min\left\{\mathcal{L}(x^{k+1}, y, z, \lambda^k) + \frac{1}{2}\left\| \begin{matrix} y - y^k \\ z - z^k \end{matrix} \right\|^2_{D_{BC}} \left| \begin{matrix} y \in \mathcal{Y} \\ z \in \mathcal{Z} \end{matrix} \right. \right\}, \\ \lambda^{k+1} = \lambda^k - (Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), \end{cases}$$

The algorithm (6.7) can be rewritten in an equivalent form: $(\mu = \tau + 1 > 2)$.

$$\begin{cases} x^{k+1} = \arg\min\{\theta_1(x) + \frac{\beta}{2}\|Ax + By^k + Cz^k - b - \frac{1}{\beta}\lambda^k\|^2 \,|\, x \in \mathcal{X}\}, \\ \quad \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b) \\ y^{k+1} = \arg\min\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^T By + \frac{\mu\beta}{2}\|B(y - y^k)\|^2 \,|\, y \in \mathcal{Y}\}, \\ z^{k+1} = \arg\min\{\theta_3(z) - (\lambda^{k+\frac{1}{2}})^T Cz + \frac{\mu\beta}{2}\|C(z - z^k)\|^2 \,|\, z \in \mathcal{Z}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b), \end{cases} \tag{6.9}$$

The related publication:

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

In the above paper, in order to ensure the convergence, it **was** required

$\tau > 1$ (in (6.7)) which is equivalent to $\mu > 2$ (in (6.9)).

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.

> 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-

Thus, Osher's research group utilize the iterative formula (6.9), according to our previous paper, they set

$\mu = 2.01,$ it is only a pity larger than 2.

Large parameter μ (or τ) will lead a slow convergence.

最新进展：最优正则化因子的选择– OO6235 的结论

Bingsheng He, Xiaoming Yuan: On the optimal proximal parameter of an ADMM-like splitting method for separable convex programming. Mathematical methods in image processing and inverse problems, 139 – 163, Springer Proc. Math. Stat., 360. Springer, Singapore, 2021. Optimization Online 6235.

Our new assertion: In (6.7)

- if $\tau > 0.5$, the method is still convergent;
- if $\tau < 0.5$, there is divergent example.

Equivalently in (6.9) :

- if $\mu > 1.5$, the method is still convergent;
- if $\mu < 1.5$, there is divergent example.

For convex optimization problem (1.1) with three separable objective functions, the parameters in the equivalent methods (6.7) and (6.9) :

- **0.5** is the threshold factor of the parameter τ in (6.7) !
- **1.5** is the threshold factor of the parameter μ in (6.9) !

7 利用统一框架设计的 PPA 算法

求解变分不等式 (1.2) 的 PPA 型算法要求预测 (2.2) 中的矩阵 Q 本身是一个能写成 H 的对称正定矩阵. 这时, 我们把相应的矩阵 Q 记为 H. 这类方法中, 我们用平凡松弛的校正 (2.3) 给出 v^{k+1}, 其中 $M = \alpha I$, 实际运算中, 一般取 $\alpha \in [1.2, 1.8]$.

如果我们为求解 (1.2) 构造的预测公式中的 $\tilde{w}^k$ 满足

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T H(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \tag{7.1a}$$

其中

$$H = \begin{pmatrix} \beta B^T B + \delta I_m & 0 & -B^T \\ 0 & \beta C^T C + \delta I_m & -C^T \\ -B & -C & \frac{2}{\beta} I_m \end{pmatrix}, \tag{7.1b}$$

其中 $\beta > 0$ 和 $\delta > 0$ 都是任意给定的大于零的常数. 由于

$$H = \begin{pmatrix} \beta B^T B + \delta I_m & 0 & -B^T \\ 0 & 0 & 0 \\ -B & 0 & \frac{1}{\beta} I_m \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \beta C^T C + \delta I_m & -C^T \\ 0 & -C & \frac{1}{\beta} I_m \end{pmatrix},$$

对任意的 $\beta > 0, \delta > 0$ 和 $v = (y, z, \lambda) \neq 0$,

$$v^T H v = \Big\| \sqrt{\beta} B y - \frac{1}{\sqrt{\beta}} \lambda \Big\|^2 + \Big\| \sqrt{\beta} C z - \frac{1}{\sqrt{\beta}} \lambda \Big\|^2 + \delta(\|y\|^2 + \|z\|^2) > 0.$$

矩阵 H 是正定的. 我们用平凡松弛的校正 (2.3) 得到新的迭代点 v^{k+1}. 根据统一框架, 算法就是收敛的. 因此, 问题归结为如何实现满足 (7.1) 的预测. 用 (1.2) 中 $F(w)$ 的表达式, 把 (7.1) 的具体形式写出来就是 $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$, 使得

$$\left\{\begin{array}{ll}
\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{\underline{-A^T \tilde{\lambda}^k}\} \ge 0, \quad \forall x \in \mathcal{X}, & \text{(7.2a)} \\
\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{\underline{-B^T \tilde{\lambda}^k} + \beta B^T B(\tilde{y}^k - y^k) + \delta(\tilde{y}^k - y^k) & \\
\qquad\qquad -B^T(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \ \ \forall y \in \mathcal{Y}, & \text{(7.2b)} \\
\theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^T \{\underline{-C^T \tilde{\lambda}^k} + \beta C^T C(\tilde{z}^k - z^k) + \delta(\tilde{z}^k - z^k) & \\
\qquad\qquad -C^T(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \ \ \forall z \in \mathcal{Z}, & \text{(7.2c)} \\
\qquad \underline{(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b)} & \\
\qquad\qquad -B(\tilde{y}^k - y^k) - C(\tilde{z}^k - z^k) + (2/\beta)\ (\tilde{\lambda}^k - \lambda^k) = 0. & \text{(7.2d)}
\end{array}\right.$$

上式中, 有下划线的凑在一起, 就是 (7.1) 中的 $F(\tilde{w}^k)$. 把 (7.2) 的具体形式写出来就

是 $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$, 使得

$$\left\{\begin{array}{ll}
\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{-A^T \tilde{\lambda}^k\} \ge 0, \quad \forall x \in \mathcal{X}, & \text{(7.3a)} \\
\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{-B^T(2\tilde{\lambda}^k - \lambda^k) & \\
\qquad\qquad +\beta B^T B(\tilde{y}^k - y^k) + \delta(\tilde{y}^k - y^k)\} \ge 0, \ \ \forall y \in \mathcal{Y}, & \text{(7.3b)} \\
\theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^T C^T(2\tilde{\lambda}^k - \lambda^k) & \\
\qquad\qquad +\beta C^T C(\tilde{z}^k - z^k) + \delta(\tilde{z}^k - z^k)\} \ge 0, \ \ \forall z \in \mathcal{Z}, & \text{(7.3c)} \\
(A\tilde{x}^k + By^k + Cz^k - b) + (2/\beta)\ (\tilde{\lambda}^k - \lambda^k) = 0. & \text{(7.3d)}
\end{array}\right.$$

如果令

$$\tilde{x}^k = \mathrm{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \tfrac{1}{4}\beta\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\}, \qquad (7.4)$$

根据最优性质的定理, 问题 (7.4) 的最优性条件是 $\tilde{x}^k \in \mathcal{X}$,

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{-A^T \lambda^k + \tfrac{1}{2}\beta A^T(A\tilde{x}^k + By^k + Cz^k - b)\} \ge 0, \quad \forall x \in \mathcal{X}. \qquad (7.5)$$

再定义

$$\tilde{\lambda}^k = \lambda^k - \tfrac{1}{2}\beta(A\tilde{x}^k + By^k + Cz^k - b). \qquad (7.6)$$

将 (7.6) 代入 (7.5), 满足了 (7.3a). 注意到 (7.6) 又和 (7.3d) 等价. 这样, 有了 $\tilde{\lambda}^k$, 要得到满

足 (7.3b) 的 $\tilde{y}^k$ 和满足 (7.3c) 的 $\tilde{z}^k$, 根据最优性质的定理, 只要分别通过

$$\tilde{y}^k = \text{argmin}\left\{ \begin{array}{c} \theta_2(y) - y^T B^T [2\tilde{\lambda}^k - \lambda^k] + \\ \frac{1}{2}\beta\|B(y - y^k)\|^2 + \frac{1}{2}\delta\|y - y^k\|^2 \end{array} \,\middle|\, y \in \mathcal{Y} \right\}$$

和

$$\tilde{z}^k = \text{argmin}\left\{ \begin{array}{c} \theta_3(z) - z^T C^T [2\tilde{\lambda}^k - \lambda^k] + \\ \frac{1}{2}\beta\|C(z - z^k)\|^2 + \frac{1}{2}\delta\|z - z^k\|^2 \end{array} \,\middle|\, z \in \mathcal{Z} \right\}$$

得到. 综上所述, 按照 $x, \lambda, (y, z)$ 顺序计算:

$$\left\{ \begin{array}{ll} \tilde{x}^k \in \text{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{4}\beta\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\}, & \text{(7.7a)} \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k + Cz^k - b), & \text{(7.7b)} \\ \tilde{y}^k = \text{argmin}\left\{\theta_2(y) - y^T B^T [2\tilde{\lambda}^k - \lambda^k] + \begin{pmatrix} \frac{1}{2}\beta\|B(y - y^k)\|^2 \\ +\frac{1}{2}\delta\|y - y^k\|^2 \end{pmatrix} \,\middle|\, y \in \mathcal{Y}\right\}, & \text{(7.7c)} \\ \tilde{z}^k = \text{argmin}\left\{\theta_3(z) - z^T C^T [2\tilde{\lambda}^k - \lambda^k] + \begin{pmatrix} \frac{1}{2}\beta\|C(z - z^k)\|^2 \\ +\frac{1}{2}\delta\|z - z^k\|^2 \end{pmatrix} \,\middle|\, z \in \mathcal{Z}\right\}, & \text{(7.7d)} \end{array} \right.$$

就得到满足条件 (7.1) 的预测点. 由于预测中的矩阵对称正定, 新的迭代点可以利用预测点继续进行平凡的松弛校正得到.

References

[1] 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, Series A, 155 (2016) 57-79.

[2] 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.

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

[4] B.S. He, F. Ma and X.M. Yuan, Optimally linearizing the alternating direction method of multipliers for convex programming, Comput. Optim. Appl. 75 (2020), 361-388.

[5] 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.

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

[7] B.S. He and X.M. Yuan, On the optimal proximal parameter of an ADMM-like splitting method for separable convex programming. Mathematical methods in image processing and inverse problems, 139 – 163, Springer Proc. Math. Stat., 360, Springer, Singapore,

[8] Tao M and Yuan X M. Recovering low-rank and sparse components of matrices from incomplete and noisy observations *SIAM Journal on Optimization*, 2011, 21: 57-81.

变分不等式框架下结构型 凸优化的分裂收缩算法

VI. 多个可分离块凸优化问题的ADMM类 秩一校正 & 秩二校正方法

中学的数理基础　必要的社会实践
普通的大学数学　一般的优化原理

何 炳 生　　南京大学数学系

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

天元数学东北中心　2023年10月17 – 27日

1 p-块可分离凸优化问题的变分不等式

p-块可分离凸优化问题

$$\min\Big\{\sum_{i=1}^{p}\theta_i(x_i)\;\Big|\;\sum_{i=1}^{p}A_ix_i=b\ (\text{or}\ \ge b),\ \ x_i\in\mathcal{X}_i\Big\}. \tag{1.1}$$

The Lagrangian function is

$$L(x_1,\ldots,x_p,\lambda)=\sum_{i=1}^{p}\theta_i(x_i)-\lambda^T\Big(\sum_{i=1}^{p}A_ix_i-b\Big),$$

which is defined on $\Omega=\prod_{i=1}^{p}\mathcal{X}_i\times\Lambda$, where

$$\Lambda=\begin{cases}\Re^m, & \text{if } \sum_{i=1}^{p}A_ix_i=b,\\ \Re^m_+, & \text{if } \sum_{i=1}^{p}A_ix_i\ge b.\end{cases}$$

Let $(x_1^*, \ldots, x_p^*, \lambda^*) \in \Omega$ be a saddle point of the Lagrangian function, then

$$L_{\lambda\in\Lambda}(x_1^*, \ldots, x_p^*, \lambda) \le L(x_1^*, \ldots, x_p^*, \lambda^*) \le L_{x_i\in\mathcal{X}_i}(x_1, \ldots, x_p, \lambda^*).$$

The optimality condition of (1.1) can be written as the following VI:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{1.2a}$$

where

$$w = \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \lambda \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A_1^T\lambda \\ \vdots \\ -A_p^T\lambda \\ \sum_{i=1}^p A_i x_i - b \end{pmatrix}, \tag{1.2b}$$

and

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

Again, we denote by Ω^* the solution set of the VI (1.2).

2 从交替方向法得到的启示

Let us consider the general separable convex optimization model

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

The augmented Lagrangian function is

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

Applied the classical ADMM [1, 2] to the problem (2.1):

ADMM for (2.1) From (y^k, λ^k) to (y^{k+1}, λ^{k+1})

$$\left\{\begin{array}{lcl} x^{k+1} & \in & \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ y^{k+1} & \in & \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \,|\, y \in \mathcal{Y}\}, \\ \lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{array}\right. \tag{2.2}$$

Ignoring some constant terms in the objective functions of the corresponding subproblems, we can rewrite the ADMM (2.2) as

$$\left\{\begin{array}{rcl} x^{k+1} & \in & \text{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k - b)\|^2 \mid x \in \mathcal{X}\} \\ & = & \text{argmin}\left\{ \begin{array}{c} \theta_1(x) - x^T A^T \lambda^k \\ + \frac{\beta}{2}\|A(x - x^k) + (Ax^k + By^k - b)\|^2 \end{array} \middle| \, x \in \mathcal{X} \right\} \\ & = & \text{argmin}\left\{ \begin{array}{c} \theta_1(x) - x^T A^T [\lambda^k - \beta(Ax^k + By^k - b)] \\ + \frac{\beta}{2}\|A(x - x^k)\|^2 \end{array} \middle| \, x \in \mathcal{X} \right\} \\ y^{k+1} & \in & \text{argmin}\{\theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By - b)\|^2 \mid y \in \mathcal{Y}\} \\ & = & \text{argmin}\left\{ \begin{array}{c} \theta_2(y) - y^T B^T \lambda^k \\ + \frac{\beta}{2}\|B(y - y^k) + (Ax^{k+1} + By^k - b)\|^2 \end{array} \middle| \, y \in \mathcal{Y} \right\} \\ & = & \text{argmin}\left\{ \begin{array}{c} \theta_2(y) - y^T B^T [\lambda^k - \beta(Ax^k + By^k - b)] \\ + \frac{\beta}{2}\|A(x^{k+1} - x^k) + B(y - y^k)\|^2 \end{array} \middle| \, y \in \mathcal{Y} \right\} \\ \lambda^{k+1} & = & \lambda^k - \beta\big(Ax^{k+1} + By^{k+1} - b\big) \end{array}\right. \tag{2.3}$$

如果记

$$\lambda^{k+\frac{1}{2}} := \lambda^k - \beta(Ax^k + By^k - b). \tag{2.4a}$$

ADMM 迭代式 (2.3) 就可以写成

$$\left\{\begin{array}{rcl} x^{k+1} & \in & \text{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k - b)\|^2 \mid x \in \mathcal{X}\} \\ & = & \text{argmin}\left\{ \begin{array}{c} \theta_1(x) - x^T A^T \lambda^{k+\frac{1}{2}} \\ + \frac{\beta}{2}\|A(x - x^k)\|^2 \end{array} \middle| \, x \in \mathcal{X} \right\} \\ y^{k+1} & \in & \text{argmin}\{\theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By - b)\|^2 \mid y \in \mathcal{Y}\} \\ & = & \text{argmin}\left\{ \begin{array}{c} \theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} \\ + \frac{\beta}{2}\|A(x^{k+1} - x^k) + B(y - y^k)\|^2 \end{array} \middle| \, y \in \mathcal{Y} \right\} \\ \lambda^{k+1} & = & \lambda^k - \beta\big(Ax^{k+1} + By^{k+1} - b\big) \\ & = & \lambda^k - \beta\big(Ax^k + By^k - b\big) + \beta A(x^k - x^{k+1}) + \beta B(y^k - y^{k+1}) \end{array}\right. \tag{2.4b}$$

假如记 $\tilde{\lambda}^k = \lambda^{k+\frac{1}{2}}, \quad \tilde{x}^k = x^{k+1}, \quad \tilde{y}^k = y^{k+1}$, 则有 **预测**

$$\left\{\begin{array}{lcl} \tilde{\lambda}^k & = & \lambda^k - \beta(Ax^k + By^k - b) \\ \tilde{x}^k & \in & \mathrm{argmin}\{ \; \theta_1(x) - x^T A^T \tilde{\lambda}^k + \frac{\beta}{2}\|A(x - x^k)\|^2 \;\; \big| \, x \in \mathcal{X} \} \\ \tilde{y}^k & \in & \mathrm{argmin}\left\{ \begin{array}{l} \theta_2(y) - y^T B^T \tilde{\lambda}^k \\ + \frac{\beta}{2}\|A(\tilde{x}^k - x^k) + B(y - y^k)\|^2 \end{array} \;\Bigg| \; y \in \mathcal{Y} \right\} \end{array}\right. \tag{2.5}$$

因为 $\beta(Ax^k + By^k - b) = \lambda^k - \tilde{\lambda}^k$,

$$\begin{array}{lcl} \lambda^{k+1} & = & \lambda^k - \beta\big(Ax^k + By^k - b\big) + \beta A(x^k - x^{k+1}) + \beta B(y^k - y^{k+1}) \\ & = & \lambda^k - [(\lambda^k - \tilde{\lambda}^k) - \beta A(x^k - \tilde{x}^k) - \beta B(y^k - \tilde{y}^k) \end{array}$$

校正

$$\begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\beta A & -\beta B & I_m \end{pmatrix} \begin{pmatrix} x^k - \tilde{x}^k \\ y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}$$

2.1 ADMM with wider applications

Let us consider the general two-block separable convex optimization model

$$\min\big\{\theta_1(x) + \theta_2(y) \mid Ax + By = b \ (\text{or} \geq b), x \in \mathcal{X}, y \in \mathcal{Y}\big\}. \tag{2.6}$$

The linear constraints can be a system of linear equations or linear inequalities. We define

$$\Lambda = \left\{ \begin{array}{ll} \Re^m, & \text{if } Ax + By = b, \\ \Re^m_+, & \text{if } Ax + By \geq b, \end{array} \right.$$

and denote the projection on Λ by $P_\Lambda[\cdot]$. For such special Λ, the projection on Λ is clear !

The only difference： $\quad P_{\Re^m}(\lambda) = \lambda, \quad P_{\Re^m_+}(\lambda) = \max\{\lambda, 0\}.$

A Dual-Primal Extension of the ADMM for (2.6).

From (Ax^k, By^k, λ^k) to $(Ax^{k+1}, By^{k+1}, \lambda^{k+1})$: Find $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ via

$$\begin{cases} \tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(Ax^k + By^k - b\big)\big], \\ \tilde{x}^k \in \text{argmin}\big\{\theta_1(x) - x^TA^T\tilde{\lambda}^k + \frac{1}{2}\beta\|A(x - x^k)\|^2 \mid x \in \mathcal{X}\big\}, \\ \tilde{y}^k \in \text{argmin}\big\{\theta_2(y) - y^TB^T\tilde{\lambda}^k + \frac{1}{2}\beta\|A(\tilde{x}^k - x^k) + B(y - y^k)\|^2 \mid y \in \mathcal{Y}\big\}. \end{cases} \tag{2.7}$$

预测先做 Primal 部分, 再做 Dual 部分, 顺序也可以倒过来.

A Primal-Dual Extension of the ADMM for (2.6).

From (Ax^k, By^k, λ^k) to $(Ax^{k+1}, By^{k+1}, \lambda^{k+1})$: Find $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ via

$$\begin{cases} \tilde{x}^k \in \text{argmin}\big\{\theta_1(x) - x^TA^T\lambda^k + \frac{1}{2}\beta\|A(x - x^k)\|^2 \mid x \in \mathcal{X}\big\}, \\ \tilde{y}^k \in \text{argmin}\big\{\theta_2(y) - y^TB^T\lambda^k + \frac{1}{2}\beta\|A(\tilde{x}^k - x^k) + B(y - y^k)\|^2 \mid y \in \mathcal{Y}\big\}, \\ \tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(A\tilde{x}^k + B\tilde{y}^k - b\big)\big]. \end{cases} \tag{2.8}$$

无论是 dual-primal, 还是 primal-dual 方法, 都可以向多块问题直接推广.

多块问题 (1.2) 的 DUAL-PRIMAL 预测 Prediction

从给定的 $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ 到预测点 $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \cdots, \tilde{x}_p^k, \tilde{\lambda}^k)$:

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

$$\begin{cases} \tilde{\lambda}^k = P_\Lambda\big[\lambda^k - \beta\big(\sum_{j=1}^p A_jx_j^k - b\big)\big] \\ \tilde{x}_1^k \in \arg\min\{\theta_1(x_1) - x_1^TA_1^T\tilde{\lambda}^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\}; \\ \tilde{x}_2^k \in \arg\min\{\theta_2(x_2) - x_2^TA_2^T\tilde{\lambda}^k + \frac{\beta}{2}\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\}; \\ \qquad\vdots \\ \tilde{x}_i^k \in \arg\min_{x_i \in \mathcal{X}_i}\{\theta_i(x_i) - x_i^TA_i^T\tilde{\lambda}^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1}A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\}; \\ \qquad\vdots \\ \tilde{x}_p^k \in \text{argmin}_{x_p \in \mathcal{X}_p}\{\theta_p(x_p) - x_p^TA_p^T\tilde{\lambda}^k + \frac{\beta}{2}\|\sum_{j=1}^{p-1}A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\}. \end{cases} \tag{2.9}$$

预测先对偶再原始. 对可分离的原始变量子问题逐一按序求解.

多块问题 (1.2) 的 PRIMAL-DUAL 预测 Prediction

从给定的 $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ 到预测点 $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \cdots, \tilde{x}_p^k, \tilde{\lambda}^k)$:

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

$$\left\{\begin{array}{l} \tilde{x}_1^k \in \arg\min\{\theta_1(x_1) - x_1^T A_1^T\lambda^k + \frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\}; \\ \tilde{x}_2^k \in \arg\min\{\theta_2(x_2) - x_2^T A_2^T\lambda^k + \frac{\beta}{2}\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\}; \\ \qquad \vdots \\ \tilde{x}_i^k \in \arg\min_{x_i\in\mathcal{X}_i}\{\theta_i(x_i) - x_i^T A_i^T\lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\}; \\ \qquad \vdots \\ \tilde{x}_p^k \in \arg\min_{x_p\in\mathcal{X}_p}\{\theta_p(x_p) - x_p^T A_p^T\lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{p-1} A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\}; \\ \tilde{\lambda}^k = P_\Lambda[\lambda^k - \beta(\sum_{j=1}^p A_j\tilde{x}_j^k - b)]. \end{array}\right. \tag{2.10}$$

预测先原始再对偶. 对可分离的原始变量子问题逐一按序求解.

3 采用 Primal-Dual 预测的预测矩阵

Analysis for the P-D Prediction 我们先看 (2.10) 中 x 子问题

$$\tilde{x}_i^k \in \arg\min\{\theta_i(x_i) - x_i^T A_i^T\lambda^k + \frac{\beta}{2}\|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2 | x_i \in \mathcal{X}_i\}.$$

根据最优性引理, 最优性条件是 $\tilde{x}_i^k \in \mathcal{X}_i$ 和

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\{-A_i^T\lambda^k + \beta A_i^T(\sum_{j=1}^{i} A_j(\tilde{x}_j^k - x_j^k))\} \ge 0, \quad \forall x_i \in \mathcal{X}_i.$$

它可以改写成 $\tilde{x}_i^k \in \mathcal{X}_i$ 和对所有的 $x_i \in \mathcal{X}_i$ 都有

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\{\underline{-A_i^T\tilde{\lambda}^k} + \beta A_i^T(\sum_{j=1}^{i} A_j(\tilde{x}_j^k - x_j^k)) + A_i^T(\tilde{\lambda}^k - \lambda^k)\} \ge 0. \tag{3.1a}$$

预测的对偶部分 $\tilde{\lambda}^k = P_\Lambda[\lambda^k - \beta(\sum_{j=1}^p A_j\tilde{x}_j^k - b)]$, 等价形式

$$\tilde{\lambda}^k = \arg\min\{\|\lambda - [\lambda^k - \beta(\sum_{j=1}^p A_j\tilde{x}_j^k - b)]\|^2 \mid \lambda \in \Lambda\}.$$

最优性条件是

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T \{ \underline{(\sum_{j=1}^p A_j \tilde{x}_j^k - b)} + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \;\; \forall \lambda \in \Lambda. \tag{3.1b}$$

Summating (3.1a) and (3.1b), for the predictor $\tilde{w}^k$ generated by (2.10), we have $\tilde{w}^k \in \Omega$,

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T \underline{F(\tilde{w}^k)} \geq (w - \tilde{w}^k)^T Q (w^k - \tilde{w}^k), \;\; \forall\, w \in \Omega, \tag{3.2a}$$

where

$$Q = \begin{pmatrix} \beta A_1^T A_1 & 0 & \cdots & 0 & A_1^T \\ \beta A_2^T A_1 & \beta A_2^T A_2 & \ddots & \vdots & A_2^T \\ \vdots & & \ddots & 0 & \vdots \\ \beta A_p^T A_1 & \beta A_p^T A_2 & \cdots & \beta A_p^T A_p & A_p^T \\ 0 & 0 & \cdots & 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{3.2b}$$

3.1 变量代换下的预测矩阵

The optimization problem (1.1) has been translated to VI (1.2), namely,

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

For the easy analysis, we need to denote the following notations:

$$P = \begin{pmatrix} \sqrt{\beta} A_1 & 0 & \cdots & \cdots & 0 \\ 0 & \sqrt{\beta} A_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \sqrt{\beta} A_p & 0 \\ 0 & \cdots & \cdots & 0 & (1/\sqrt{\beta}) I_m \end{pmatrix}, \quad \xi = Pw = \begin{pmatrix} \sqrt{\beta} A_1 x_1 \\ \sqrt{\beta} A_2 x_2 \\ \vdots \\ \sqrt{\beta} A_p x_p \\ (1/\sqrt{\beta}) \lambda \end{pmatrix}. \tag{3.3}$$

Accordingly, we define

$$\Xi = \{ \xi \mid \xi = Pw, \; w \in \Omega \},$$

and

$$\Xi^* = \{ \xi^* \mid \xi^* = Pw^*, \; w^* \in \Omega^* \}.$$

Using the notation P in (3.3), for the matrix Q in (3.2b), we have

$$Q = P^T \mathcal{Q} P, \qquad \text{where} \qquad \mathcal{Q} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{3.4}$$

Thus, for the right hand side of (3.2a), we have

$$\begin{aligned} (w - \tilde{w}^k)^T Q (w^k - \tilde{w}^k) &= (w - \tilde{w}^k)^T P^T \mathcal{Q} P (w^k - \tilde{w}^k) \\ &= (\xi - \tilde{\xi}^k)^T \mathcal{Q} (\xi^k - \tilde{\xi}^k). \end{aligned}$$

Then, it follows from (3.2) that we have the following VI for the P-D prediction:

$$\begin{aligned} \tilde{w}^k \in \Omega, \ \ \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) & \\ \geq (\xi - \tilde{\xi}^k)^T \mathcal{Q} (\xi^k - \tilde{\xi}^k), \ \ \forall\, w \in \Omega. & \end{aligned} \tag{3.5}$$

where $\mathcal{Q}$ is given in (3.4).

3.2 变量代换下的算法统一框架

Prediction-Correction Framework for VI (1.2).

1. (Prediction Step) With given w^k and $\xi^k = Pw^k$, find $\tilde{w}^k \in \Omega$ such that

$$\begin{aligned} \tilde{w}^k \in \Omega, \ \ \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) & \\ \geq (\xi - \tilde{\xi}^k)^T \mathcal{Q} (\xi^k - \tilde{\xi}^k), \ \ \forall\, w \in \Omega, & \end{aligned} \tag{3.6a}$$

   with $\mathcal{Q} \in \Re^{(p+1)m \times (p+1)m}$, and the matrix $\mathcal{Q}^T + \mathcal{Q}$ is positive definite.

2. (Correction Step) With the predictor $\tilde{w}^k$ by (3.6a) and $\tilde{\xi}^k = P\tilde{w}^k$, the new iterate ξ^{k+1} is updated by

$$\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k), \tag{3.6b}$$

   where $\mathcal{M} \in \Re^{(p+1)m \times (p+1)m}$ is a non-singular matrix.

定理 1 *For the matrices* $\mathcal{Q}$ *and* $\mathcal{M}$ *in the algorithm* (3.6), *if there is a positive definite matrix* $\mathcal{H} \in \Re^{(p+1)m\times(p+1)m}$ *such that*

$$\mathcal{H}\mathcal{M} = \mathcal{Q} \tag{3.7a}$$

and

$$\mathcal{G} := \mathcal{Q}^T + \mathcal{Q} - \mathcal{M}^T\mathcal{H}\mathcal{M} \succ 0, \tag{3.7b}$$

then we have

$$\|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \le \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^k - \tilde{\xi}^k\|_{\mathcal{G}}^2, \quad \forall\, \xi^* \in \Xi^*. \tag{3.8}$$

Proof. Setting w in (3.6a) as any fixed $w^* \in \Omega^*$, and using

$$(\tilde{w}^k - w^*)^T F(\tilde{w}^k) \equiv (\tilde{w}^k - w^*)^T F(w^*),$$

we get

$$(\tilde{\xi}^k - \xi^*)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) \ge \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*), \quad \forall\, w^* \in \Omega^*.$$

The right-hand side of the last inequality is non-negative. Thus, we have

$$(\xi^k - \xi^*)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) \ge (\xi^k - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k), \quad \forall\, \xi^* \in \Xi^*. \tag{3.9}$$

Then, by simple manipulations, we obtain

$$\begin{aligned}
&\|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \\
&\overset{(3.6b)}{=} \quad \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|(\xi^k - \xi^*) - \mathcal{M}(\xi^k - \tilde{\xi}^k)\|_{\mathcal{H}}^2 \\
&\overset{(3.7a)}{=} \quad 2(\xi^k - \xi^*)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) - \|\mathcal{M}(\xi^k - \tilde{\xi}^k)\|_{\mathcal{H}}^2 \\
&\overset{(3.9)}{\ge} \quad 2(\xi^k - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) - \|\mathcal{M}(\xi^k - \tilde{\xi}^k)\|_{\mathcal{H}}^2 \\
&= \quad (\xi^k - \tilde{\xi}^k)^T[(\mathcal{Q}^T + \mathcal{Q}) - \mathcal{M}^T\mathcal{H}\mathcal{M}](\xi^k - \tilde{\xi}^k) \\
&\overset{(3.7b)}{=} \quad \|\xi^k - \tilde{\xi}^k\|_{\mathcal{G}}^2.
\end{aligned}$$

The assertion of this theorem is proved. $\Box$

We call (3.7) the convergence conditions for the algorithm framework (3.6).

The inequality (3.8) is the key for the convergence proofs, for details, see [5]

4 基于 Primal-Dual 预测的校正方法

For given $\mathcal{Q}$ which satisfies $\mathcal{Q}^T + \mathcal{Q} \succ 0$, we chose $\mathcal{D}$ and $\mathcal{G}$, such that

$$\mathcal{D} \succ 0, \qquad \mathcal{G} \succ 0, \qquad \mathcal{D} + \mathcal{G} = \mathcal{Q}^T + \mathcal{Q}.$$

Then, the correction matrix $\mathcal{M}$ in (3.6b) is given by

$$\mathcal{M} = \mathcal{Q}^{-T}\mathcal{D}.$$

选择了想要的 $0 \prec \mathcal{D}$, 构造 $\mathcal{M}$ 不再神秘！ 下面先介绍以前在 [5] 中 "凑" 出来的 $\mathcal{M}$

First, we give some correction examples which satisfy conditions (3.7) in Theorem 1.

In order to simplify the notations to be used, we define the following $p \times p$ block matrices:

$$\mathcal{L} = \begin{pmatrix} I_m & 0 & \cdots & 0 \\ I_m & I_m & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ I_m & I_m & \cdots & I_m \end{pmatrix}, \quad \mathcal{I} = \begin{pmatrix} I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{4.1}$$

We also define the $1 \times p$ block matrix

$$\mathcal{E}^T = \begin{pmatrix} I_m & I_m & \cdots & I_m \end{pmatrix}. \tag{4.2}$$

Using the notations (4.1)-(4.2), the matrix $\mathcal{Q}$ in (3.4) has the form

$$\mathcal{Q} = \begin{pmatrix} \mathcal{L} & \mathcal{E} \\ 0 & I_m \end{pmatrix} \quad \text{and} \quad \mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \mathcal{I} + \mathcal{E}\mathcal{E}^T & \mathcal{E} \\ \mathcal{E}^T & 2I_m \end{pmatrix}. \tag{4.3}$$

In order to construct a convergent algorithm, we need only to give the matrices $\mathcal{M}$ and $\mathcal{H}$ and to verify the convergence conditions (3.7)

By setting

$$\mathcal{M} = \begin{pmatrix} \nu\mathcal{L}^{-T} & 0 \\ -\nu\mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix}. \tag{4.4}$$

For the above matrices $\mathcal{Q}$ and $\mathcal{M}$, the remaining tasks is to find a positive definite matrix $\mathcal{H}$, such that the convergence conditions (3.7)are satisfied.

(4.4) 中的 $\mathcal{M}$ 是我们在 [5] 中 "凑" 出来的.

How to improvise a correction matrix $\mathcal{M}$? 因为 $\mathcal{H}\mathcal{M}=\mathcal{Q}$,

$$\mathcal{H}=\mathcal{Q}\mathcal{M}^{-1}.$$

有没有一个“块下三角矩阵” $\mathcal{M}$ 满足收敛性条件呢? 因为块下三角矩阵的逆矩阵也是块下三角矩阵, 设 $\mathcal{M}$ 的逆矩阵形式为

$$\mathcal{M}^{-1}=\begin{pmatrix} X & 0 \\ Y & I_m \end{pmatrix}.$$

$\mathcal{H}=\mathcal{Q}\mathcal{M}^{-1}$ 应该是对称矩阵

$$\mathcal{H}=\mathcal{Q}\mathcal{M}^{-1}=\begin{pmatrix} \mathcal{L} & \mathcal{E} \\ 0 & I_m \end{pmatrix}\begin{pmatrix} X & 0 \\ Y & I_m \end{pmatrix}=\begin{pmatrix} \mathcal{L}X+\mathcal{E}Y & \mathcal{E} \\ Y & I_m \end{pmatrix}. \tag{4.5}$$

因此有 $Y=\mathcal{E}^T$ 和 $X=S^{-1}\mathcal{L}^T$, S 是一个待定的正定矩阵. 所以

$$\mathcal{M}^{-1}=\begin{pmatrix} S^{-1}\mathcal{L}^T & 0 \\ \mathcal{E}^T & I_m \end{pmatrix} \quad 并有 \quad \mathcal{M}=\begin{pmatrix} \mathcal{L}^{-T}S & 0 \\ -\mathcal{E}^T\mathcal{L}^{-T}S & I_m \end{pmatrix}.$$

继续“凑”下去, 发现 $S=\nu I$ 就可以了, 我们因此也凑出了 $\mathcal{H}$.

$$\begin{aligned}\mathcal{M}^T\mathcal{H}\mathcal{M} = \mathcal{Q}^T\mathcal{M} &= \begin{pmatrix} \mathcal{L}^T & 0 \\ \mathcal{E}^T & I_m \end{pmatrix}\begin{pmatrix} \mathcal{L}^{-T}S & 0 \\ -\mathcal{E}^T\mathcal{L}^{-T}S & I_m \end{pmatrix} \\ &= \begin{pmatrix} S & 0 \\ 0 & I_m \end{pmatrix}.\end{aligned}$$

因为

$$\mathcal{Q}^T+\mathcal{Q}=\begin{pmatrix} \mathcal{L}^T+\mathcal{L} & \mathcal{E} \\ \mathcal{E}^T & I_m \end{pmatrix}=\begin{pmatrix} \mathcal{I}+\mathcal{E}\mathcal{E}^T & \mathcal{E} \\ \mathcal{E}^T & 2I_m \end{pmatrix}$$

取 $S=\nu\mathcal{I}$, 就能使 $\mathcal{Q}^T+\mathcal{Q}-\mathcal{M}^T\mathcal{H}\mathcal{M}\succ 0$.

以 $Y=\mathcal{E}^T, X=S^{-1}\mathcal{L}^T$ 和 $S=\nu\mathcal{I}$ 代入(4.5), 就有

$$\mathcal{H}=\begin{pmatrix} \mathcal{L}X+\mathcal{E}Y & \mathcal{E} \\ Y & I_m \end{pmatrix}=\begin{pmatrix} \frac{1}{\nu}\mathcal{L}\mathcal{L}^T+\mathcal{E}\mathcal{E}^T & \mathcal{E} \\ \mathcal{E}^T & I_m \end{pmatrix}.$$

引理 1 *For the matrices $\mathcal{Q}$ and $\mathcal{M}$ given by* (4.3) *and* (4.4)*, respectively, the matrix*

$$\mathcal{H} = \begin{pmatrix} \frac{1}{\nu}\mathcal{L}\mathcal{L}^T + \mathcal{E}\mathcal{E}^T & \mathcal{E} \\ \mathcal{E}^T & I_m \end{pmatrix} \quad \textit{with} \quad \nu \in (0,1) \tag{4.6}$$

is positive definite, and it satisfies $\mathcal{H}\mathcal{M} = \mathcal{Q}$.

Proof. It is easy to check the positive definiteness of $\mathcal{H}$. In addition, for the block matrix $\mathcal{Q}$ in (3.4), we have

$$\begin{aligned} \mathcal{H}\mathcal{M} &= \begin{pmatrix} \frac{1}{\nu}\mathcal{L}\mathcal{L}^T + \mathcal{E}\mathcal{E}^T & \mathcal{E} \\ \mathcal{E}^T & I_m \end{pmatrix} \begin{pmatrix} \nu\mathcal{L}^{-T} & 0 \\ -\nu\mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{L} & \mathcal{E} \\ 0 & I_m \end{pmatrix} = \mathcal{Q}. \end{aligned}$$

The assertions of this lemma are proved. $\square$

这样凑出来的 $\mathcal{M}$ 和 $\mathcal{H}$, 能否满足 $\mathcal{Q}^T + \mathcal{Q} - \mathcal{M}^T\mathcal{H}\mathcal{M} \succ 0$? 还需要检查一下.

引理 2 *Let $\mathcal{Q}$, $\mathcal{M}$ and $\mathcal{H}$ be defined in* (3.4), (4.4) *and* (4.6)*, respectively. Then the matrix*

$$\mathcal{G} := (\mathcal{Q}^T + \mathcal{Q}) - \mathcal{M}^T\mathcal{H}\mathcal{M} \tag{4.7}$$

is positive definite.

Proof. By elementary matrix multiplications, we know that

$$\mathcal{M}^T\mathcal{H}\mathcal{M} = \mathcal{Q}^T\mathcal{M} = \begin{pmatrix} \mathcal{L}^T & 0 \\ \mathcal{E}^T & I_m \end{pmatrix}\begin{pmatrix} \nu\mathcal{L}^{-T} & 0 \\ -\nu\mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix} = \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} = \mathcal{D}.$$

Then, it follows from $\mathcal{L}^T + \mathcal{L} = \mathcal{I} + \mathcal{E}\mathcal{E}^T$ (see (4.1)-(4.2)) that

$$\begin{aligned} \mathcal{G} &= (\mathcal{Q}^T + \mathcal{Q}) - \mathcal{M}^T\mathcal{H}\mathcal{M} \\ &= \begin{pmatrix} \mathcal{L}^T + \mathcal{L} & \mathcal{E} \\ \mathcal{E}^T & 2I_m \end{pmatrix} - \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} = \begin{pmatrix} (1-\nu)\mathcal{I} + \mathcal{E}\mathcal{E}^T & \mathcal{E} \\ \mathcal{E}^T & I_m \end{pmatrix}. \end{aligned}$$

Thus, the matrix $\mathcal{G}$ is positive definite for any $\nu \in (0,1)$. $\square$

Finally, correction step can be written

$$\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k). \tag{4.8}$$

Lemma 1 and Lemma 2 have verified the convergence conditions (3.7) and thus the key convergence inequality (3.8) holds. The algorithm (2.10) $\&$ (4.8) is convergent.

Recall the respective definitions $\mathcal{L}$ and $\mathcal{E}^T$ in (4.1) and (4.2). We have

$$\mathcal{L}^{-T} = \begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix}$$

and

$$\mathcal{E}^T\mathcal{L}^{-T} = \begin{pmatrix} I_m & 0 & \cdots & 0 \end{pmatrix}.$$

Thus

$$\mathcal{M} = \begin{pmatrix} \nu\mathcal{L}^{-T} & 0 \\ -\nu\mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix} = \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ -\nu I_m & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{4.9}$$

By a manipulation, we have

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^k \\ A_2x_2^k \\ \vdots \\ A_px_p^k \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 & 0 \\ 0 & \nu I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -\nu I_m \\ 0 & \cdots & 0 & \nu I_m \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \end{pmatrix}, \tag{4.10}$$

and

$$\lambda^{k+1} = \tilde{\lambda}^k + \nu\beta(A_1x_1^k - A_1\tilde{x}_1^k). \tag{4.11}$$

校正非常简单, 工作量也很小. 把校正公式分开来写就是:

$Ax_i^{k+1}, i = 1, \ldots, p$

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \end{pmatrix} - \nu \begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \end{pmatrix}, \tag{4.12}$$

λ^{k+1}

$$\begin{aligned} \lambda^{k+1} &= \lambda^k - [-\nu\beta(A_1x_1^k - A_1\tilde{x}_1^k) + (\lambda^k - \tilde{\lambda}^k)] \\ &= \tilde{\lambda}^k + \nu\beta(A_1x_1^k - A_1\tilde{x}_1^k). \end{aligned} \tag{4.13}$$

5 More Choices based on the predictions

只要 $\mathcal{Q}^{-T}$ 结构简单, 构造校正矩阵 $\mathcal{M}$ 的方法并不神秘！是非常容易的.

The matrix $\mathcal{Q}$ in (3.4) has the form

$$\mathcal{Q} = \begin{pmatrix} \mathcal{L} & \mathcal{E} \\ 0 & I_m \end{pmatrix} \quad \text{and thus} \quad \mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \mathcal{I} + \mathcal{E}\mathcal{E}^T & \mathcal{E} \\ \mathcal{E}^T & 2I_m \end{pmatrix}.$$

To further analyze the correction steps associated with the correction matrix $\mathcal{M}$, let us take a closer look at the matrix $\mathcal{Q}^{-T}$.

According to the primal-dual prediction (2.10), the matrix $\mathcal{Q}$ in (3.4), we have

$$\mathcal{Q}^{-T} = \begin{pmatrix} \mathcal{L}^T & 0 \\ \mathcal{E}^T & I_m \end{pmatrix}^{-1} = \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ -\mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix}. \tag{5.1}$$

and

$$\begin{pmatrix} \mathcal{L}^{-T} & 0 \\ -\mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix} = \begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -I_m & 0 \\ 0 & \cdots & 0 & I_m & 0 \\ -I_m & 0 & \cdots & 0 & I_m \end{pmatrix}.$$

The calculation $\mathcal{M} = \mathcal{Q}^{-T}\mathcal{D}$ is essentially very easy for different $\mathcal{D}$!

Since

$$\mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \mathcal{I} + \mathcal{E}\mathcal{E}^T & \mathcal{E} \\ \mathcal{E}^T & 2I_m \end{pmatrix},$$

it can be decomposed as

$$\mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} + \begin{pmatrix} (1-\nu)\mathcal{I} + \mathcal{E}\mathcal{E}^T & \mathcal{E} \\ \mathcal{E}^T & I_m \end{pmatrix}.$$

The both matrices in the right hand side are positive definite. If we chose

$$\mathcal{D} = \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} \quad \text{and thus} \quad \mathcal{G} = \begin{pmatrix} (1-\nu)\mathcal{I} + \mathcal{E}\mathcal{E}^T & \mathcal{E} \\ \mathcal{E}^T & I_m \end{pmatrix},$$

it is just the correction in Section $\S$4.

Conversely, we can also choose

$$\mathcal{D} = \begin{pmatrix} (1-\nu)\mathcal{I} + \mathcal{E}\mathcal{E}^T & \mathcal{E} \\ \mathcal{E}^T & I_m \end{pmatrix} \quad \text{and thus} \quad \mathcal{G} = \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix}$$

and thus get the another correction method.

There are many positive definite decompositions of $\mathcal{Q}^T + \mathcal{Q}$, for example,

$$\mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} (1-\nu)\mathcal{I} & 0 \\ 0 & (1-\nu)I_m \end{pmatrix} + \begin{pmatrix} \nu\mathcal{I} + \mathcal{E}\mathcal{E}^T & \mathcal{E} \\ \mathcal{E}^T & (1+\nu)I_m \end{pmatrix}.$$

and

$$\mathcal{Q}^T + \mathcal{Q} = \mathcal{D} + \mathcal{G} = \alpha(\mathcal{Q}^T + \mathcal{Q}) + (1-\alpha)(\mathcal{Q}^T + \mathcal{Q}), \quad \alpha \in (0,1).$$

对基于 Dual-Primal 预测的方法, 建议读者自己去构造校正矩阵 $\mathcal{M}$.

这一讲求解多块可分离线性约束的凸优化问题的方法, 仍然是变分不等式框架下的预测-校正方法. 采用 Primal-Dual 和 Dual-Primal 的 Gauss 型预测, 分别得到预测矩阵

$$Q_{PD} = \begin{pmatrix} \beta A_1^TA_1 & 0 & \cdots & 0 & A_1^T \\ \beta A_2^TA_1 & \beta A_2^TA_2 & \ddots & \vdots & A_2^T \\ \vdots & \ddots & \ddots & 0 & \vdots \\ \beta A_p^TA_1 & \cdots & \beta A_p^TA_{p-1} & \beta A_p^TA_p & A_p^T \\ 0 & 0 & \cdots & 0 & \frac{1}{\beta}I_m \end{pmatrix}$$

和

$$Q_{DP} = \begin{pmatrix} \beta A_1^TA_1 & 0 & \cdots & 0 & 0 \\ \beta A_2^TA_1 & \beta A_2^TA_2 & \ddots & \vdots & 0 \\ \vdots & \ddots & \ddots & 0 & \vdots \\ \beta A_p^TA_1 & \cdots & \beta A_p^TA_{p-1} & \beta A_p^TA_p & 0 \\ -A_1 & \cdots & -A_{p-1} & -A_p & \frac{1}{\beta}I_m \end{pmatrix}.$$

利用 (3.3) 做了替换以后, 得到特殊结构的预测矩阵 $\mathcal{Q}$, 它们分别是

$$\mathcal{Q}_{PD} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & \ddots & \ddots & 0 & \vdots \\ I_m & \cdots & I_m & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}, \quad \mathcal{Q}_{DP} = \begin{pmatrix} I_m & 0 & \cdots & 0 & 0 \\ I_m & I_m & \ddots & \vdots & 0 \\ \vdots & \ddots & \ddots & 0 & \vdots \\ I_m & \cdots & I_m & I_m & 0 \\ -I_m & \cdots & -I_m & -I_m & I_m \end{pmatrix}.$$

而 $\mathcal{Q}^T + \mathcal{Q}$ 的形式分别为

$$\begin{pmatrix} \mathcal{I} + \mathcal{E}\mathcal{E}^T & \mathcal{E} \\ \mathcal{E}^T & 2I_m \end{pmatrix} \quad \text{和} \quad \begin{pmatrix} \mathcal{I} + \mathcal{E}\mathcal{E}^T & -\mathcal{E} \\ -\mathcal{E}^T & 2I_m \end{pmatrix}.$$

其逆矩阵 $\mathcal{Q}^{-T}$ 的形式分别是

$$\mathcal{Q}_{PD}^{-T} = \begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -I_m & 0 \\ 0 & \cdots & 0 & I_m & 0 \\ -I_m & 0 & \cdots & 0 & I_m \end{pmatrix}, \quad \mathcal{Q}_{DP}^{-T} = \begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -I_m & 0 \\ 0 & \cdots & 0 & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}.$$

选取满足条件

$$\mathcal{D} + \mathcal{G} = \mathcal{Q}^T + \mathcal{Q}$$

的正定矩阵 $\mathcal{D}$ 和 $\mathcal{G}$, 策略是很多的. 然后, 令

$$\mathcal{M} = \mathcal{Q}^{-T}\mathcal{D} \quad \text{和} \quad \xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k),$$

则有

$$\|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \le \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^k - \tilde{\xi}^k\|_{\mathcal{G}}^2, \quad \forall \xi^* \in \Xi^*,$$

其中 $\mathcal{H} = \mathcal{Q}\mathcal{D}^{-1}\mathcal{Q}^T$. 由于 $\mathcal{Q}^{-T}$ 的结构相当简单, 校正是容易实现的.

6 为什么说是秩一校正

这一讲讨论的方法, 由串行预测生成的矩阵 $\mathcal{Q}_{PD}$ 和 $\mathcal{Q}_{DP}$, 都是一个容易求逆的矩阵和一个广义秩一矩阵的和. 譬如说,

$$\mathcal{Q}_{PD}^T = \mathcal{Q}_{0PD}^T \otimes I_m, \tag{6.1}$$

其中

$$\mathcal{Q}_{0PD}^T = \begin{pmatrix} 1 & 1 & \cdots & 1 & 0 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 1 & \vdots \\ 0 & \cdots & 0 & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}, \qquad \otimes \text{表示 Kronecker 积.}$$

把 $\mathcal{Q}_{0PD}^T$ 中的 1 改成 I_m, 就得到了 $\mathcal{Q}_{PD}^T$. 注意到

$$\mathcal{Q}_{0PD}^T = Q_{1PD}^T + Q_{2PD}^T,$$

其中

$$Q_{1PD}^T = \begin{pmatrix} 1 & 1 & \cdots & 1 & 0 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 1 & \vdots \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \quad Q_{2PD}^T = \begin{pmatrix} 0 & \cdots & \cdots & 0 & 0 \\ \vdots & & & \vdots & \vdots \\ \vdots & & & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & 0 \\ 1 & \cdots & \cdots & 1 & 0 \end{pmatrix}.$$

由于 Q_{1PD} 容易求逆, Q_{2PD} 是秩一矩阵

$$Q_{1PD}^{-T} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -1 & 0 \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix}, \quad Q_{2PD}^T = \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & \cdots & \cdots & 1 & 0 \end{pmatrix}.$$

利用线性代数中的秩一校正求逆公式

$$(A+uv^T)^{-1}=A^{-1}-\frac{1}{1+v^TA^{-1}u}A^{-1}uv^TA^{-1},$$

容易得到

$$\mathcal{Q}_{0PD}^{-T}=\begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -1 & 0 \\ 0 & \cdots & 0 & 1 & 0 \\ -1 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

由 (6.1) 和

$$\mathcal{Q}_{PD}^{-T}=\mathcal{Q}_{0PD}^{-T}\otimes I_m,$$

我们得到

$$\mathcal{Q}_{PD}^{-T}=\begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -I_m & 0 \\ 0 & \cdots & 0 & I_m & 0 \\ -I_m & 0 & \cdots & 0 & I_m \end{pmatrix}.$$

秩一校正是这一章用到的关键技术.

这一讲的方法主要取材于尚未正式发表的 arXiv 上的文章 [9].

7 根据统一框架设计用秩二校正的预测方法

我们仍然考虑线性约束的多块可分离凸优化问题. 其相应的变分不等式在前一讲也已经做了介绍. 采用统一框架中的方法求解变分不等式. 这些方法的第 k-步迭代从给定的 $(A_1x_1^k,\ldots,A_px_p^k,\lambda^k)$ 出发, 生成预测点 $\tilde{w}^k\in\Omega$, 满足

$$\theta(x)-\theta(\tilde{x}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(w-\tilde{w}^k)^TQ(w^k-\tilde{w}^k),\quad \forall\, w\in\Omega, \tag{7.1}$$

其中 Q^T+Q 是本质上正定的. 这一讲前面的方法采用的是串行预测, 然后用广义秩一校正. 这一节介绍的方法, 我们仍然采用的是串行预测, 然后进行广义秩二校正.

利用前一讲定义的变换, 可以把预测 (7.1) 改写成

$$\tilde{w}^k\in\Omega,\quad \theta(x)-\theta(\tilde{x}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(\xi-\tilde{\xi}^k)^T\mathcal{Q}(\xi^k-\tilde{\xi}^k),\quad \forall\, w\in\Omega, \tag{7.2}$$

其中

$$Q=P^T\mathcal{Q}P,\qquad \text{并且}\qquad \mathcal{Q}^T+\mathcal{Q}\succ 0. \tag{7.3}$$

由 (7.2) 得到

$$(\tilde{\xi}^k-\xi^*)^T\mathcal{Q}(\xi^k-\tilde{\xi}^k)\ge 0,\quad \forall\,\xi\in\Xi^*. \tag{7.4}$$

接着, 我们就可以选择正定矩阵 $\mathcal{D}$ 和 $\mathcal{G}$, 使得

$$\mathcal{D}\succ 0,\quad \mathcal{G}\succ 0,\quad \mathcal{D}+\mathcal{G}=\mathcal{Q}^T+\mathcal{Q}. \tag{7.5}$$

最后, 用

$$\xi^{k+1}=\xi^k-\mathcal{Q}^{-T}\mathcal{D}(\xi^k-\tilde{\xi}^k) \tag{7.6}$$

得到 ξ^{k+1}. 算法产生的序列 $\{\xi^k\}$ 满足收敛性质

$$\|\xi^{k+1}-\xi^*\|_{\mathcal{H}}^2\le\|\xi^k-\xi^*\|_{\mathcal{H}}^2-\|\xi^k-\tilde{\xi}^k\|_{\mathcal{G}}^2,\quad \forall\,\xi^*\in\Xi^*. \tag{7.7}$$

讨论根据统一框架构造算法, 实际上就是预先设定矩阵 $\mathcal{Q}$, 使得

1. $\mathcal{Q}^T+\mathcal{Q}\succ 0$.
2. 对 $Q=P^T\mathcal{Q}P$ 的预测矩阵 Q, 相应的预测 (7.1) 可以实施.
3. $\mathcal{Q}^{-T}$ (或者 $\mathcal{Q}^{-1}$) 的表达式简单, 使得校正 (7.6) 容易实现.

求解多块可分离凸优化问题, 预测按串行逐渐向前推进, 如果将矩阵 $\mathcal{Q}$ 写成 2×2 的分块形式, 其左上角是下三角矩阵 $\mathcal{L}$.

7.1 Primal-Dual 预测后再秩二校正的方法

设计一个可以执行 Primal-Dual 预测的矩阵

$$\mathcal{Q} = \begin{pmatrix} \mathcal{L} & \mathcal{E} \\ \mathcal{E}^T & \frac{5}{2} I_m \end{pmatrix}. \tag{7.8}$$

由于

$$\mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} \mathcal{I} + \mathcal{E}\mathcal{E}^T & 2\mathcal{E} \\ 2\mathcal{E}^T & 5I_m \end{pmatrix} = \begin{pmatrix} \mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} + \begin{pmatrix} \mathcal{E} \\ 2I_m \end{pmatrix} \big(\mathcal{E}^T, \ 2I_m\big), \tag{7.9}$$

$\mathcal{Q}^T + \mathcal{Q}$ 是单位矩阵与一个半正定矩阵的和, 所以是正定的. 注意到如果将 (7.8) 中 $\mathcal{Q}$ 矩阵左上角的 $\mathcal{L}$ 换成 $\mathcal{I}$, $\mathcal{Q}$ 就成了对称矩阵, 但对 $p \geq 3$, 这样的矩阵就不再是正定的. 利用变换前一讲的记号, 对应于 (7.8) 中的 $\mathcal{Q}$, 相应

的 $Q = P^T \mathcal{Q} P$, 所以

$$Q = \begin{pmatrix} \beta A_1^T A_1 & 0 & \cdots & 0 & A_1^T \\ \beta A_2^T A_1 & \beta A_2^T A_2 & \ddots & \vdots & A_2^T \\ \vdots & & \ddots & 0 & \vdots \\ \beta A_p^T A_1 & \beta A_p^T A_2 & \cdots & \beta A_p^T A_p & A_p^T \\ A_1 & A_2 & \cdots & A_p & \frac{5}{2\beta} I_m \end{pmatrix}. \tag{7.10}$$

要实现预测

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w - \tilde{w}^k)^T Q(w^k - \tilde{w}^k), \quad \forall w \in \Omega, \tag{7.11}$$

其中矩阵 Q 由 (7.10) 给出. 根据多块问题变分不等式的形式, 预测 (7.11) 的原始和对偶部分可以分别通过

$$\left\{\begin{array}{l} \tilde{x}_1^k \in \arg\min\Big\{\theta_1(x_1) - x_1^T A_1^T \lambda^k + \frac{1}{2}\beta\|A_1(x_1 - x_1^k)\|^2 \;\Big|\; x_1 \in \mathcal{X}_1\Big\}; \\ \qquad\qquad\vdots \\ \tilde{x}_i^k \in \arg\min\left\{\begin{array}{c} \theta_i(x_i) - x_i^T A_i^T \lambda^k + \\ \frac{1}{2}\beta\|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2 \end{array}\right| x_i \in \mathcal{X}_i\Bigg\}; \\ \qquad\qquad\vdots \\ \tilde{x}_p^k \in \arg\min\left\{\begin{array}{c} \theta_p(x_p) - x_p^T A_p^T \lambda^k + \\ \frac{1}{2}\beta\|\sum_{j=1}^{p-1} A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2 \end{array}\right| x_p \in \mathcal{X}_p\Bigg\} \end{array}\right. \tag{7.12a}$$

和

$$\tilde{\lambda}^k = P_\Lambda\Big\{\lambda^k - \frac{2}{5}\beta\Big[\big(\sum_{i=1}^p A_i\tilde{x}_i^k - b\big) + \sum_{i=1}^p A_i(\tilde{x}_i^k - x_i^k)\Big]\Big\} \tag{7.12b}$$

完成. 根据最优性定理, (7.12a) 中 x_i-子问题的最优性条件是

$$\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\big\{-A_i^T\lambda^k + A_i^T\beta\big[\sum_{j=1}^i A_j(\tilde{x}_j^k - x_j^k)\big]\big\} \ge 0, \quad \forall x_i \in \mathcal{X}_i.$$

这可以写成

$$\begin{aligned} \tilde{x}_i^k \in \mathcal{X}_i, \;\; & \theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\big\{-A_i^T\tilde{\lambda}^k\big\} \\ & \ge (x_i - \tilde{x}_i^k)^T\big\{A_i^T\beta\big[\sum_{j=1}^i A_j(x_j^k - \tilde{x}_j^k)\big] + A_i^T(\lambda^k - \tilde{\lambda}^k)\big\}, \quad \forall x_i \in \mathcal{X}_i. \end{aligned} \tag{7.13a}$$

对偶预测 (7.12b) 的最优性条件是 $\tilde{\lambda}^k \in \Lambda$,

$$(\lambda - \tilde{\lambda}^k)^T\Big\{\tilde{\lambda}^k - \Big[\lambda^k - \frac{2}{5}\beta\big[\big(\sum_{i=1}^p A_i\tilde{x}_i^k - b\big) + \sum_{i=1}^p A_i(\tilde{x}_i^k - x_i^k)\big]\Big]\Big\} \ge 0, \quad \forall\, \lambda \in \Lambda.$$

这可以改写成等价的 $\tilde{\lambda}^k \in \Lambda$,

$$(\lambda-\tilde{\lambda}^k)^T\Big\{\Big[(\sum_{i=1}^p A_i\tilde{x}_i^k-b)+\sum_{i=1}^p A_i(\tilde{x}_i^k-x_i^k)\Big]+\frac{5}{2\beta}(\tilde{\lambda}^k-\lambda^k)\Big\}\geq 0,\ \ \forall\,\lambda\in\Lambda$$

并进一步有

$$\begin{aligned}\tilde{\lambda}^k\in\Lambda,\ \ (\lambda-\tilde{\lambda}^k)^T\{\sum_{i=1}^p A_i\tilde{x}_i^k-b\}&\\ \geq(\lambda-\tilde{\lambda}^k)^T\Big\{\sum_{i=1}^p A_i(x_i^k-\tilde{x}_i^k)+\frac{5}{2\beta}(\lambda^k-\tilde{\lambda}^k)\Big\},&\ \ \forall\,\lambda\in\Lambda.\end{aligned}\tag{7.13b}$$

把 (7.13a) 和 (7.13b) 放在一起, 就是预测 (7.11), 其中的矩阵 Q 由 (7.10) 给出. 得到了满足 (7.11) 的 $\tilde{w}^k$, 也得到了相应的 $\tilde{\xi}^k=P\tilde{w}^k$.

还需要关心的是, 对 (7.8) 中的 $\mathcal{Q}$, $\mathcal{Q}^{-T}$ 的形式是否简单. 对这里的 $\mathcal{Q}$, 为了防止混淆, 我们记其为 $\mathcal{Q}_{PD}$, 有

$$\mathcal{Q}_{PD}^T=\mathcal{Q}_1^T+\mathcal{Q}_2^T,$$

其中

$$\mathcal{Q}_1^T=\begin{pmatrix}\mathcal{L}^T & 0\\ 0 & \frac{5}{2}I_m\end{pmatrix},\qquad \mathcal{Q}_2^T=\begin{pmatrix}0 & \mathcal{E}\\ \mathcal{E}^T & 0\end{pmatrix}.$$

$\mathcal{Q}_1$ 是个容易求逆的矩阵, 而

$$\mathcal{Q}_2^T=\begin{pmatrix}I_m & 0\\ \vdots & \vdots\\ I_m & 0\\ 0 & I_m\end{pmatrix}\begin{pmatrix}0 & \cdots & 0 & I_m\\ I_m & \cdots & I_m & 0\end{pmatrix}=\begin{pmatrix}\mathcal{E} & 0\\ 0 & I_m\end{pmatrix}\begin{pmatrix}0 & I_m\\ \mathcal{E}^T & 0\end{pmatrix}$$

是个广义秩二矩阵. 利用线性代数中的求逆公式

$$(A+UV)^{-1}=A^{-1}-A^{-1}U(I+VA^{-1}U)^{-1}VA^{-1},$$

经过演算可得

$$\mathcal{Q}_{PD}^{-T}=\begin{pmatrix}\mathcal{L}^{-T} & 0\\ 0 & \frac{2}{5}I_m\end{pmatrix}+\frac{2}{3}\begin{pmatrix}\mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T} & -\mathcal{L}^{-T}\mathcal{E}\\ -\mathcal{E}^T\mathcal{L}^{-T} & \frac{2}{5}I_m\end{pmatrix}.$$

上式也可以写成

$$\mathcal{Q}_{PD}^{-T} = \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ 0 & 0 \end{pmatrix} + \frac{2}{3} \begin{pmatrix} \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T} & -\mathcal{L}^{-T}\mathcal{E} \\ -\mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix}. \tag{7.14}$$

由于

$$\mathcal{L}^{-T}\mathcal{E} = \begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} I_m \\ I_m \\ \vdots \\ I_m \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ I_m \end{pmatrix}$$

和

$$\mathcal{E}^T\mathcal{L}^{-T} = \big(I_m, I_m, \ldots, I_m\big) \begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix} = \big(I_m, 0, \ldots, 0\big),$$

我们有

$$\begin{pmatrix} \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T} & -\mathcal{L}^{-T}\mathcal{E} \\ -\mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ I_m & 0 & \cdots & 0 & -I_m \\ -I_m & 0 & \cdots & 0 & I_m \end{pmatrix}.$$

所以, (7.14) 中的 $\mathcal{Q}_{PD}^{-T}$ 形式是相当简单的. 写开来就是

$$\mathcal{Q}_{PD}^{-T} = \begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -I_m & 0 \\ 0 & \cdots & 0 & I_m & 0 \\ 0 & \cdots & 0 & 0 & 0 \end{pmatrix} + \frac{2}{3} \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ I_m & 0 & \cdots & 0 & -I_m \\ -I_m & 0 & \cdots & 0 & I_m \end{pmatrix},$$

校正容易实现. 由 (7.9) 知道

$$\mathcal{Q}_{PD}^{T} + \mathcal{Q}_{PD} = \begin{pmatrix} \mathcal{I} + \mathcal{E}\mathcal{E}^T & 2\mathcal{E} \\ 2\mathcal{E}^T & 5I_m \end{pmatrix} = \begin{pmatrix} \mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} + \begin{pmatrix} \mathcal{E} \\ 2I_m \end{pmatrix} \left(\mathcal{E}^T,\ 2I_m \right).$$

选择符合条件 (7.5) 的矩阵 $\mathcal{D}$ 有许多选法. 例如, 若取

$$\mathcal{D} = \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix}, \quad \nu \in (0,1),$$

条件 (7.5) 满足. 由

$$\xi^{k+1} = \xi^k - \mathcal{Q}_{PD}^{-T}\mathcal{D}(\xi^k - \tilde{\xi}^k)$$

产生的迭代序列 $\{\xi^k\}$ 满足关键收缩不等式 (7.7). 与上式等价的校正公式是

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \\ \lambda^{k} \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}$$

$$- \frac{2}{3} \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ \nu I_m & 0 & \cdots & 0 & -\frac{1}{\beta}I_m \\ -\nu\beta I_m & 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \quad (7.15)$$

7.2 Dual-Primal 预测后再校正的方法

同样, 可以设计一个 Dual-Primal 的预测矩阵

$$\mathcal{Q}=\begin{pmatrix}\mathcal{L}^T & -\mathcal{E}\\ -\mathcal{E}^T & \frac{5}{2}I_m\end{pmatrix}. \tag{7.16}$$

其中的 $\mathcal{L}$, $\mathcal{E}$ 如前一讲给出. 由于

$$\begin{aligned}\mathcal{Q}^T+\mathcal{Q} &= \begin{pmatrix}\mathcal{I}+\mathcal{E}\mathcal{E}^T & -2\mathcal{E}\\ -2\mathcal{E}^T & 5I_m\end{pmatrix}\\ &= \begin{pmatrix}\mathcal{I} & 0\\ 0 & I_m\end{pmatrix}+\begin{pmatrix}\mathcal{E}\\ -2I_m\end{pmatrix}\big(\mathcal{E}^T,\ -2I_m\big).\end{aligned} \tag{7.17}$$

$\mathcal{Q}^T+\mathcal{Q}$ 是单位矩阵与一个半正定矩阵的和, 所以是正定的. 同样, 如果将 (7.16) 中的 $\mathcal{Q}$ 矩阵左上角的 $\mathcal{L}$ 换成 $\mathcal{I}$, $\mathcal{Q}$ 就成了对称矩阵, 但对 $p\ge 3$, 这样的矩阵就不再是正定的. 利用相应变换的记号, 对应于 (7.16) 的 $\mathcal{Q}$, 相应

的 $Q=P^T\mathcal{Q}P$, 所以

$$Q=\begin{pmatrix}\beta A_1^TA_1 & 0 & \cdots & 0 & -A_1^T\\ \beta A_2^TA_1 & \beta A_2^TA_2 & \ddots & \vdots & -A_2^T\\ \vdots & & \ddots & 0 & \vdots\\ \beta A_p^TA_1 & \beta A_p^TA_2 & \cdots & \beta A_p^TA_p & -A_p^T\\ -A_1 & -A_2 & \cdots & -A_p & \frac{5}{2\beta}I_m\end{pmatrix}, \tag{7.18}$$

要实现预测

$$\theta(x)-\theta(\tilde{x}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(w-\tilde{w}^k)^TQ(w^k-\tilde{w}^k),\quad \forall\, w\in\Omega, \tag{7.19}$$

其中矩阵 Q 由 (7.18) 给出. 根据多块问题变分不等式的形式, (7.19) 的最后一行是

$$\tilde{\lambda}^k \in \Lambda, \;\; (\lambda - \tilde{\lambda}^k)^T \Big\{ \big(\sum_{i=1}^{p} A_i \tilde{x}_i^k - b \big) \\ \geq (\lambda - \tilde{\lambda}^k)^T \Big\{ - \sum_{i=1}^{p} A_i (x_i^k - \tilde{x}_i^k) + \tfrac{5}{2\beta} (\lambda^k - \tilde{\lambda}^k) \Big\}, \;\; \forall \, \lambda \in \Lambda.$$

也就是

$$\tilde{\lambda}^k \in \Lambda, \;\; (\lambda - \tilde{\lambda}^k)^T \Big\{ \big(\sum_{i=1}^{p} A_i x_i^k - b \big) + \tfrac{5}{2\beta} (\tilde{\lambda}^k - \lambda^k) \Big\} \geq 0, \;\; \forall \, \lambda \in \Lambda.$$

上面的 $\tilde{\lambda}^k$ 可以通过

$$\tilde{\lambda}^k = P_\Lambda \big[\lambda^k - \tfrac{2}{5} \beta \big(\sum_{i=1}^{p} A_i x_i^k - b \big) \big] \tag{7.20a}$$

得到. 有了对偶变量的预测, 串行迭代的 x_i-子问题需要满足的最优性条件是

$$\tilde{x}_i^k \in \mathcal{X}_i, \;\; \theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T \big\{ - A_i^T \tilde{\lambda}^k + A_i^T \beta \big[\sum_{j=1}^{i} A_j (\tilde{x}_j^k - x_j^k) \big] \\ - A_i^T (\tilde{\lambda}^k - \lambda^k) \big\} \geq 0, \quad \forall x_i \in \mathcal{X}_i,$$

其中第一个 $-A_i^T \tilde{\lambda}^k$ 对应的是 $F(\tilde{w}^k)$ 中相应的那部分. 上式归并以后得到

$$\tilde{x}_i^k \in \mathcal{X}_i, \;\; \theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T \big\{ -A_i^T (2\tilde{\lambda}^k - \lambda^k) \\ + A_i^T \beta \big[\sum_{j=1}^{i} A_j (\tilde{x}_j^k - x_j^k) \big] \big\} \geq 0, \;\; \forall x_i \in \mathcal{X}_i.$$

根据最优性条件的定理, 它是优化问题

$$\tilde{x}_i^k \in \arg\min \big\{ \theta_i(x_i) - x_i^T A_i^T (2\tilde{\lambda}^k - \lambda^k) + \frac{\beta}{2} \| \sum_{j=1}^{i-1} A_j (\tilde{x}_j^k - x_j^k) + A_i (x_i - x_i^k) \|^2 \, \big| \, x_i \in \mathcal{X}_i \big\}$$

的最优性条件. 因此, 原始变量 x 的预测是

$$\left\{\begin{array}{l}
\tilde{x}_1^k \in \arg\min\left\{\begin{array}{c}\theta_1(x_1) - x_1^T A_1^T(2\tilde{\lambda}^k - \lambda^k) \\ +\frac{\beta}{2}\|A_1(x_1 - x_1^k)\|^2\end{array}\;\middle|\; x_1 \in \mathcal{X}_1\right\}; \\
\qquad\qquad\qquad\vdots \\
\tilde{x}_i^k \in \arg\min\left\{\begin{array}{c}\theta_i(x_i) - x_i^T A_i^T(2\tilde{\lambda}^k - \lambda^k) \\ +\frac{\beta}{2}\|\sum_{j=1}^{i-1} A_j(\tilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2\end{array}\;\middle|\; x_i \in \mathcal{X}_i\right\}; \\
\qquad\vdots \\
\tilde{x}_p^k \in \arg\min\left\{\begin{array}{c}\theta_p(x_p) - x_p^T A_p^T(2\tilde{\lambda}^k - \lambda^k) \\ +\frac{\beta}{2}\|\sum_{j=1}^{p-1} A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\end{array}\middle| x_p \in \mathcal{X}_p\right\}.
\end{array}\right. \tag{7.20b}$$

这样, 我们就得到了满足 (7.4) 的 $\tilde{w}^k$, 也得到了相应的 $\tilde{\xi}^k = P\tilde{w}^k$. 同样需要关心的是, 对 (7.16) 中的 $\mathcal{Q}$, $\mathcal{Q}^{-T}$ 的形式是否简单. 对这里的 $\mathcal{Q}$, 为了防止混淆,

我们记其为 $\mathcal{Q}_{DP}$, 又因为

$$\mathcal{Q}_{DP}{}^T = \mathcal{Q}_1^T + \mathcal{Q}_2^T,$$

其中

$$\mathcal{Q}_1^T = \begin{pmatrix} \mathcal{L}^T & 0 \\ 0 & \frac{5}{2}I_m \end{pmatrix}, \qquad \mathcal{Q}_2 = \begin{pmatrix} 0 & -\mathcal{E} \\ -\mathcal{E}^T & 0 \end{pmatrix}.$$

$\mathcal{Q}_1^T$ 是个容易求逆的矩阵, 而

$$\mathcal{Q}_2^T = \begin{pmatrix} I_m & 0 \\ \vdots & \vdots \\ I_m & 0 \\ 0 & -I_m \end{pmatrix} \begin{pmatrix} 0 & \cdots & 0 & -I_m \\ I_m & \cdots & I_m & 0 \end{pmatrix} = \begin{pmatrix} \mathcal{E} & 0 \\ 0 & -I_m \end{pmatrix} \begin{pmatrix} 0 & -I_m \\ \mathcal{E}^T & 0 \end{pmatrix}$$

是个广义秩二矩阵. 利用 $\mathcal{Q}_1^T$ 求 $\mathcal{Q}^T$ 是个秩二校正的过程. 经过简单演算可得

$$\mathcal{Q}_{DP}^{-T} = \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ 0 & 0 \end{pmatrix} + \frac{2}{3}\begin{pmatrix} \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T} & \mathcal{L}^{-T}\mathcal{E} \\ \mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix}. \tag{7.21}$$

读者将上式和 (7.14) 做比较, 就能得到 (7.21) 中的 $\mathcal{Q}^{-T}$ 的具体形式

$$\mathcal{Q}_{DP}^{-T} = \begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -I_m & 0 \\ 0 & \cdots & 0 & I_m & 0 \\ 0 & \cdots & 0 & 0 & 0 \end{pmatrix} + \frac{2}{3} \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ I_m & 0 & \cdots & 0 & I_m \\ I_m & 0 & \cdots & 0 & I_m \end{pmatrix},$$

校正容易实现. 由于

$$\mathcal{Q}_{DP}^{T} + \mathcal{Q}_{DP} = \begin{pmatrix} \mathcal{I} + \mathcal{E}\mathcal{E}^T & -2\mathcal{E} \\ -2\mathcal{E}^T & 5I_m \end{pmatrix} = \begin{pmatrix} \mathcal{I} & 0 \\ 0 & I_m \end{pmatrix} + \begin{pmatrix} \mathcal{E} \\ -2I_m \end{pmatrix} \begin{pmatrix} \mathcal{E}^T & -2I_m \end{pmatrix}.$$

同样, 若取

$$\mathcal{D} = \begin{pmatrix} \nu\mathcal{I} & 0 \\ 0 & I_m \end{pmatrix}, \quad \nu \in (0,1),$$

条件 (7.5) 满足. 由

$$\xi^{k+1} = \xi^k - \mathcal{Q}_{DP}^{-T}\mathcal{D}(\xi^k - \tilde{\xi}^k)$$

产生的迭代序列 $\{\xi^k\}$ 满足关键收缩不等式 (7.7). 与上式等价的校正公式是

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \\ \lambda^{k} \end{pmatrix} - \begin{pmatrix} \nu I_m & -\nu I_m & 0 & \cdots & 0 \\ 0 & \nu I_m & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\nu I_m & 0 \\ 0 & \cdots & 0 & \nu I_m & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}$$

$$- \frac{2}{3} \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ \nu I_m & 0 & \cdots & 0 & \frac{1}{\beta} I_m \\ \nu\beta I_m & 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \tag{7.22}$$

8 从秩一校正到秩二校正的预测方法

这一讲, 我们讨论了多块可分离问题的求解方法, 包括等式线性约束问题和不等式约束问题, 建立了求解方法的统一的框架 [5, 9].

- 秩一校正的方法, 起源于对 ADMM 的等价改写, 得到了渐进式预测, 然后进一步扩展到多块的可分离问题的求解上. 无论是 Primal-Dual 还是 Dual -Primal 预测, 生成的预测矩阵都是一个容易求逆的矩阵和一个广义秩一矩阵的和.
- §7 则是根据设定的预测矩阵, 再去构造预测方法. 其中的预测矩阵都是一个容易求逆的矩阵和一个广义秩二矩阵的和. 这个迭代预测的变分不等式可以分解, 然后通过求解相应的分裂后简单的凸优化问题去实现.
- 有了满足 $\mathcal{Q}^T + \mathcal{Q} \succ 0$ 的预测矩阵 $\mathcal{Q}$, 校正的方法是千变万化的. 只要选择

  $$0 \prec \mathcal{D} \prec \mathcal{Q}^T + \mathcal{Q},$$

  采用

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

  就能实现, 其中 ξ 和 w 的关系是由变换 (3.3) 确定的.
- 这里介绍的满足统一框架的算法, 都有 [6, 8] 中提到的类似的收敛性质.

References

[1] R. Glowinski and A. Marrocco, Approximation par éléments finis d'ordre un et résolution par pénalisation-dualité d'une classe de problémes non linéaires. RAIRO Anal. Numer. R2, 41–76 (1975)

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

[3] B. S. He, My 20 years research on alternating directions method of multipliers (in Chinese), Operations Research Trasactions, 22 (2018), 1-31.

[4] B. S. He, Using a unified framework to design the splitting and contraction methods for convex optimization (in Chinese), 高等学校计算数学学报 44 (2022), 1-35.

[5] B. S. He, S. J. Xu and X. M. Yuan, Extensions of ADMM for separable convex optimization problems with linear equality or inequality constraints, arXiv:2107.01897v2[math.OC].

[6] B. S. He and X. M. Yuan, On the $O(1/n)$ convergence rate of the alternating direction method, SIAM J. Numerical Analysis, 50 (2012), 700-709.

[7] B. S. He and X. M. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective, SIAM Journal on Imaging Science, 5 (2012) 119-149.

[8] 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.

[9] B. S. He and X. M. Yuan, On construction of splitting contraction algorithms in a prediction-correction framework for separable convex optimization, arXiv 2204.11522 [math.OC]

变分不等式框架下结构型
凸优化的分裂收缩算法

VII. 由预测决定的广义 PPA 方法

中学的数理基础	必要的社会实践
普通的大学数学	一般的优化原理

何 炳 生 南京大学数学系

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

天元数学东北中心 2023 年 10 月 17 – 27 日

1 广义邻近点算法

我们介绍了变分不等式和邻近点算法的概念. 讨论了基于合格预测构造单位步长校正方法的策略, 证明了迭代序列满足收缩不等式

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*, \tag{1.1}$$

其中 H 和 G 分别称为范数矩阵和效益矩阵. 当 $H = G$ 时, 这犹如 PPA 算法的结论. 因此, 我们把预测-校正方法的收缩不等式 (1.1) 中 $H = G$ 的方法称为广义 PPA 算法.

前面我们介绍的凸优化的分裂收缩算法基本上都在变分不等式的邻近点算法(PPA)和可分离凸优化的交替方向法(ADMM)的基础上发展起来的. 我们回顾一下这些算法的主要共有性质.

1. 变分不等式 PPA 算法的主要性质

我们对变分不等式问题

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall w \in \Omega, \tag{1.2}$$

定义了PPA算法, 设H为对称正定矩阵, H-模下的PPA算法的第k步从已知的w^k出发, 求得的新迭代点w^{k+1}使得

$$\begin{aligned} w^{k+1} \in \Omega, \quad \theta(u)-\theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \\ \geq \ (w - w^{k+1})^T H(w^k - w^{k+1}), \ \ \forall\, w \in \Omega. \end{aligned} \qquad (1.3)$$

w^{k+1}是变分不等式问题(1.2)的解的充分必要条件是(1.3)中的$w^k = w^{k+1}$. PPA算法产生的迭代序列$\{w^k\}$满足

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2, \quad \forall w^* \in \Omega^*, \qquad (1.4)$$

并有

$$\|w^k - w^{k+1}\|_H^2 \leq \|w^{k-1} - w^k\|_H^2. \qquad (1.5)$$

不等式(1.4)和(1.5)是PPA算法的两条重要而又漂亮的性质.

2. ADMM算法的主要性质

把两块可分离凸优化问题

$$\min\{\theta_1(x) + \theta_2(y) | Ax + By = b, x \in \mathcal{X}, y \in \mathcal{Y}\} \qquad (1.6)$$

转换成变分不等式(1.2), 其中

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

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

ADMM的k次迭代从给定的$v^k = (y^k, \lambda^k)$开始, 通过

$$\left\{ \begin{array}{lcl} x^{k+1} & \in & \arg\min\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\}, \\ y^{k+1} & \in & \arg\min\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\beta\|Ax^{k+1} + By - b\|^2 \mid y \in \mathcal{Y}\}, \\ \lambda^{k+1} & = & \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b) \end{array} \right. \qquad (1.7)$$

求得$w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$. 这个方法中的核心变量是$v = (y, \lambda)$.

在第三讲中我们证明了收缩性质

$$\|v^{k+1}-v^*\|_H^2 \le \|v^k-v^*\|_H^2 - \|v^k-v^{k+1}\|_H^2, \quad \forall v^* \in \mathcal{V}^*, \tag{1.8}$$

其中

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

除此之外, 我们在 [16] 中证明了 ADMM 的迭代序列 $\{v^k\}$ 具备性质

$$\|v^{k+1}-v^{k+2}\|_H^2 \le \|v^k-v^{k+1}\|_H^2. \tag{1.9}$$

不等式 (1.8) 和 (1.9) 展示了 ADMM 很好的性质. 在一些快速 ADMM 的研究 [2] 中, 都用到了 (1.9) 这条性质.

2 预测-校正的广义 PPA 算法

求解变分不等式 (1.2) 采用单位步长校正的时候, 如果预测公式

$$\tilde{w}^k \in \Omega, \;\; \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \ge (v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k), \;\; \forall\, w \in \Omega, \tag{2.1}$$

中的预测矩阵 Q 满足 $Q^T+Q \succ 0$, 若将 Q^T+Q 分拆成

$$D \succ 0, \qquad G \succ 0 \quad 和 \quad D+G = Q^T+Q, \tag{2.2}$$

再令

$$M = Q^{-T}D \qquad 和 \qquad H = QD^{-1}Q^T. \tag{2.3}$$

则由单位步长校正

$$v^{k+1} = v^k - M(v^k-\tilde{v}^k) \tag{2.4}$$

产生的新的迭代序列 $\{v^k\}$ 满足

$$\|v^{k+1}-v^*\|_H^2 \le \|v^k-v^*\|_H^2 - \|v^k-\tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.5}$$

如果我们采用一对特殊的 D 和 G, 使得

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

那么, (2.5) 就变成了

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_D^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.6}$$

对选定的 D, 根据 (2.3), 总有

$$M^T H M = D,$$

因此, (2.6) 就成了

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

再利用 $M(v^k - \tilde{v}^k) = v^k - v^{k+1}$ (见 (2.4)), 上式就了

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.7}$$

此外, 关于统一框架中所有固定步长的方法都证明了 (见本系列讲座第四讲定理4) 都证明了

$$\|v^{k+1} - v^{k+2}\|_H^2 \le \|v^k - v^{k+1}\|_H^2. \tag{2.8}$$

我们把上述分析结果写成下面的定理.

定理 1 用预测校正方法求解变分不等式 (1.2), 设预测 (2.1) 中的预测矩阵 Q 满足 $Q^T + Q \succ 0$. 若令

$$D = \tfrac{1}{2}(Q^T + Q), \qquad \text{和} \qquad M = Q^{-T} D$$

则由单位步长校正公式

$$v^{k+1} = v^k - Q^{-T} D(v^k - \tilde{v}^k) \tag{2.9}$$

产生的新的迭代点具有性质 (2.7) 和 (2.8), 其中

$$H = Q[\tfrac{1}{2}(Q^T + Q)]^{-1} Q^T.$$

求解变分不等式 (1.2), 我们把迭代序列具有性质 (2.7) 和 (2.8) 的方法, 称为广义 PPA 方法. 在实际计算中, 我们并不要求显式写出 H 的表达式,

3 变量替换下的广义 PPA 算法

仍然考虑线性约束的多块可分离凸优化问题. 这些方法的第 k-步迭代从给定

的 $(A_1x_1^k, \ldots, A_px_p^k, \lambda^k)$ 出发, 生成预测点 $\tilde{w}^k$ 满足

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (w - \tilde{w}^k)^T Q(w^k - \tilde{w}^k), \quad \forall\, w \in \Omega. \tag{3.1}$$

作为合格的预测, 其中的矩阵 $Q^T + Q$ 往往只是本质上正定的. 利用上一讲的变换, 把预测 (3.1) 改写成 $\tilde{w}^k \in \Omega$,

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (\xi - \tilde{\xi}^k)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k), \quad \forall\, w \in \Omega, \tag{3.2}$$

其中 $Q = P^T \mathcal{Q} P$,

$$\mathcal{Q}^T + \mathcal{Q} \succ 0 \tag{3.3}$$

是正定矩阵. 在 $\mathcal{Q}$ 非对称但 (3.3) 满足的时候, 必须采用必要的校正. 我们总可以选两个矩阵 $\mathcal{D}$ 和 $\mathcal{G}$, 使得

$$\mathcal{D} \succ 0, \quad \mathcal{G} \succ 0, \quad \text{和} \quad \mathcal{D} + \mathcal{G} = \mathcal{Q}^T + \mathcal{Q}. \tag{3.4}$$

根据前一讲的分析, 我们有如下的定理.

定理 2 设预测点 $\tilde{\xi}^k$ 满足条件 (3.2), 其中 $\mathcal{Q}^T + \mathcal{Q}$ 是正定矩阵. 如果由两个正定矩阵 $\mathcal{D}$ 和 $\mathcal{G}$, 使得 (3.4) 成立.

$$\mathcal{M} = \mathcal{Q}^{-T} \mathcal{D} \tag{3.5}$$

那么, 利用矩阵 (3.5) 校正

$$\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k), \tag{3.6}$$

产生的 ξ^{k+1} 满足

$$\|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \leq \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^k - \tilde{\xi}^k\|_{\mathcal{G}}^2, \quad \forall\, \xi^* \in \Xi^*, \tag{3.7}$$

其中矩阵 $\mathcal{H} = \mathcal{Q}\mathcal{D}^{-1}\mathcal{Q}^T$.

如果选

$$\mathcal{D} = \mathcal{G} = \frac{1}{2}(\mathcal{Q}^T + \mathcal{Q}) \tag{3.8}$$

那么, (3.7) 就变成了

$$\|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \leq \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^k - \tilde{\xi}^k\|_{\mathcal{D}}^2, \quad \forall \xi^* \in \Xi^*.$$

对选定的 $\mathcal{D}$, 根据 $\mathcal{D} = \mathcal{M}^T \mathcal{H} \mathcal{M}$, 并利用 (3.6), 上式就成了

$$\|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \leq \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^k - \xi^{k+1}\|_{\mathcal{H}}^2, \quad \forall \xi^* \in \Xi^*. \tag{3.9}$$

下面证明收敛性的另一条重要性质: 序列 $\{\|\xi^k - \xi^{k+1}\|_{\mathcal{H}}\}$ 是单调不增的.

定理 3 如果预测点 $\tilde{\xi}^k$ 满足条件 (3.2), 那么, 由校正 (3.6) 产生的新的迭代

点 ξ^{k+1} 满足

$$\|\xi^{k+1}-\xi^{k+2}\|_{\mathcal{H}}^2 \le \|\xi^k-\xi^{k+1}\|_{\mathcal{H}}^2. \tag{3.10}$$

证明. 首先, 我们证明迭代序列满足

$$\begin{aligned}(\xi^k-\xi^{k+1})^T\mathcal{H}[(\xi^k-\xi^{k+1})-(\xi^{k+1}-\xi^{k+2})]\\ \ge \quad \frac{1}{2}\|(\xi^k-\tilde{\xi}^k)-(\xi^{k+1}-\tilde{\xi}^{k+1})\|_{(\mathcal{Q}^T+\mathcal{Q})}^2.\end{aligned} \tag{3.11}$$

将预测 (3.2) 中的 k 改为 $k+1$, 我们有

$$\theta(x)-\theta(\tilde{x}^{k+1})+(w-\tilde{w}^{k+1})^TF(\tilde{w}^{k+1}) \ge (\xi-\tilde{\xi}^{k+1})^T\mathcal{Q}(\xi^{k+1}-\tilde{\xi}^{k+1}), \quad \forall\, w\in\Omega,$$

将上式中任意的 w 设为 $\tilde{w}^k$, 得到

$$\theta(\tilde{x}^k)-\theta(\tilde{x}^{k+1})+(\tilde{w}^k-\tilde{w}^{k+1})^TF(\tilde{w}^{k+1}) \ge (\tilde{\xi}^k-\tilde{\xi}^{k+1})^T\mathcal{Q}(\xi^{k+1}-\tilde{\xi}^{k+1}). \tag{3.12}$$

将预测 (3.2) 式中任意的 w 设为 $\tilde{w}^{k+1}$, 就有

$$\theta(\tilde{x}^{k+1})-\theta(\tilde{x}^k)+(\tilde{w}^{k+1}-\tilde{w}^k)^TF(\tilde{w}^k) \ge (\tilde{\xi}^{k+1}-\tilde{\xi}^k)^T\mathcal{Q}(\xi^k-\tilde{\xi}^k). \tag{3.13}$$

将 (3.12), (3.13) 加在一起, 利用 $(\tilde{w}^k-\tilde{w}^{k+1})^T(F(\tilde{w}^k)-F(\tilde{w}^{k+1}))\equiv 0$, 得到

$$(\tilde{\xi}^k-\tilde{\xi}^{k+1})^T\mathcal{Q}\{(\xi^k-\tilde{\xi}^k)-(\xi^{k+1}-\tilde{\xi}^{k+1})\}\ge 0.$$

对上式两边加上

$$\{(\xi^k-\tilde{\xi}^k)-(\xi^{k+1}-\tilde{\xi}^{k+1})\}^T\mathcal{Q}\{(\xi^k-\tilde{\xi}^k)-(\xi^{k+1}-\tilde{\xi}^{k+1})\}$$

并利用 $\xi^T\mathcal{Q}\xi=\frac{1}{2}\xi^T(\mathcal{Q}^T+\mathcal{Q})\xi$, 我们得到

$$\begin{aligned}(\xi^k-\xi^{k+1})^T\mathcal{Q}\{(\xi^k-\tilde{\xi}^k)-(\xi^{k+1}-\tilde{\xi}^{k+1})\}\\ \ge \quad \frac{1}{2}\|(\xi^k-\tilde{\xi}^k)-(\xi^{k+1}-\tilde{\xi}^{k+1})\|_{(\mathcal{Q}^T+\mathcal{Q})}^2.\end{aligned}$$

在上式左端利用 $\mathcal{Q}=\mathcal{H}\mathcal{M}$ 和校正公式 (3.6), 就得到 (3.11).

下面, 我们在恒等式 $\|a\|_{\mathcal{H}}^2-\|b\|_{\mathcal{H}}^2=2a^T\mathcal{H}(a-b)-\|a-b\|_{\mathcal{H}}^2$ 中置 $a=(\xi^k-\xi^{k+1})$ 和 $b=(\xi^{k+1}-\xi^{k+2})$, 得到

$$\begin{aligned}&\|\xi^k-\xi^{k+1}\|_{\mathcal{H}}^2-\|\xi^{k+1}-\xi^{k+2}\|_{\mathcal{H}}^2\\ &\quad = \quad 2(\xi^k-\xi^{k+1})^T\mathcal{H}\{(\xi^k-\xi^{k+1})-(\xi^{k+1}-\xi^{k+2})\}\\ &\qquad\qquad -\|(\xi^k-\xi^{k+1})-(\xi^{k+1}-\xi^{k+2})\|_{\mathcal{H}}^2.\end{aligned}$$

利用 (3.11) 替换上面等式右端的第一项, 得到

$$\begin{aligned}
&\|\xi^k-\xi^{k+1}\|_{\mathcal{H}}^2-\|\xi^{k+1}-\xi^{k+2}\|_{\mathcal{H}}^2 \\
&\quad\ge \|(\xi^k-\tilde{\xi}^k)-(\xi^{k+1}-\tilde{\xi}^{k+1})\|_{(\mathcal{Q}^T+\mathcal{Q})}^2 \\
&\qquad -\|(\xi^k-\xi^{k+1})-(\xi^{k+1}-\xi^{k+2})\|_{\mathcal{H}}^2. \qquad (3.14)
\end{aligned}$$

用校正公式 (3.6) 处理上式右端得到

$$\begin{aligned}
&\|(\xi^k-\tilde{\xi}^k)-(\xi^{k+1}-\tilde{\xi}^{k+1})\|_{(\mathcal{Q}^T+\mathcal{Q})}^2-\|(\xi^k-\xi^{k+1})-(\xi^{k+1}-\xi^{k+2})\|_{\mathcal{H}}^2 \\
&\quad= \quad \|(\xi^k-\tilde{\xi}^k)-(\xi^{k+1}-\tilde{\xi}^{k+1})\|_{(\mathcal{Q}^T+\mathcal{Q}-\mathcal{M}^T\mathcal{H}\mathcal{M})}^2.
\end{aligned}$$

由于 $(\mathcal{Q}^T+\mathcal{Q})-\mathcal{M}^T\mathcal{H}\mathcal{M}=\mathcal{G}\succeq 0$, (3.14) 右端非负, 定理结论得证. □

不等式 (3.9) 和 (3.10) 说明, 变量替换下的广义 PPA 算法同样具备和 PPA 算法的性质 (1.4) 和 (1.5).

在广义邻近点算法(Generalized PPA)中, 校正矩阵 $\mathcal{M}$ 是由 (3.2) 中的预测矩阵 $\mathcal{Q}$ 唯一确定的. 如果 (3.2) 中的 $\mathcal{Q}$ 是对称的, 根据相关的定义, 校正矩阵为

$$M=\tfrac{1}{2}(I+Q^{-T}Q) \qquad \text{或} \qquad \mathcal{M}=\tfrac{1}{2}\big(\mathcal{I}+\mathcal{Q}^{-T}\mathcal{Q}\big), \qquad (3.15)$$

就是单位矩阵. 我们将校正矩阵并非单位矩阵, 迭代序列又具备 (1.4)-(1.5) 这类性质的算法, 称为广义邻近点算法.

4 基于秩一校正的广义 PPA 算法

前一讲介绍的方法, 预测产生的 $\mathcal{Q}$ 矩阵是一个容易求逆的矩阵与一个广义秩一矩阵的和. 我们对这样的串型预测, 给出广义邻近点算法的校正公式.

4.1 Primal-Dual 预测的广义 PPA 算法

设预测是由前一讲中的 Primal-Dual 预测给出的, 我们得到形如 (3.2) 的变分不

等式, 其中

$$\mathcal{Q}_{PD} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{4.1}$$

利用前一讲给出的$\mathcal{L}$, $\mathcal{E}$, 那么

$$\mathcal{Q}_{PD} = \begin{pmatrix} \mathcal{L} & \mathcal{E} \\ 0 & I_m \end{pmatrix}. \tag{4.2}$$

由于

$$\mathcal{M}_{PD} = \frac{1}{2}\big(\mathcal{I}_{p+1} + \mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD}\big).$$

我们先来考察一下 $\mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD}$. 注意到

$$\mathcal{Q}_{PD}^{T} = \begin{pmatrix} \mathcal{L}^T & 0 \\ \mathcal{E}^T & I_m \end{pmatrix} \quad \text{和} \quad \mathcal{Q}_{PD}^{-T} = \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ -\mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix}.$$

所以

$$\begin{aligned} \mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD} &= \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ -\mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix}\begin{pmatrix} \mathcal{L} & \mathcal{E} \\ 0 & I_m \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{L}^{-T}\mathcal{L} & \mathcal{L}^{-T}\mathcal{E} \\ -\mathcal{E}^T\mathcal{L}^{-T}\mathcal{L} & I_m - \mathcal{E}^T\mathcal{L}^{-T}\mathcal{E} \end{pmatrix} \end{aligned} \tag{4.3}$$

分别计算 $\mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD}$ 的四块. 因为

$$\mathcal{L}^{-T} = \begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix},$$

矩阵 $\mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD}$ 左上角块,

$$
\begin{aligned}
\mathcal{L}^{-T}\mathcal{L} &= \begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} I_m & 0 & \cdots & 0 \\ I_m & I_m & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ I_m & I_m & \cdots & I_m \end{pmatrix} \\
&= \begin{pmatrix} 0 & -I_m & 0 & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -I_m \\ I_m & \cdots & I_m & I_m \end{pmatrix}. \qquad (4.4)
\end{aligned}
$$

矩阵 $\mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD}$ 右上角块,

$$
\mathcal{L}^{-T}\mathcal{E} = \begin{pmatrix} I_m & -I_m & 0 & 0 \\ 0 & I_m & \ddots & 0 \\ \vdots & \ddots & \ddots & -I_m \\ 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} I_m \\ I_m \\ \vdots \\ I_m \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ I_m \end{pmatrix}. \qquad (4.5)
$$

矩阵 $\mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD}$ 左下角块, 利用 (4.4), 得到

$$
\begin{aligned}
-\mathcal{E}^T\mathcal{L}^{-T}\mathcal{L} &= -\begin{pmatrix} I_m & I_m & \cdots & I_m \end{pmatrix} \begin{pmatrix} 0 & -I_m & 0 & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -I_m \\ I_m & I_m & \cdots & I_m \end{pmatrix} \\
&= \begin{pmatrix} -I_m & 0 & \cdots & 0 \end{pmatrix}. \qquad (4.6)
\end{aligned}
$$

矩阵 $\mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD}$ 右下角块,

$$I_m - \mathcal{E}^T\mathcal{L}^{-T}\mathcal{E} = I_m - \begin{pmatrix} I_m & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} I_m \\ I_m \\ \vdots \\ I_m \end{pmatrix} = 0. \tag{4.7}$$

组装在一起就是

$$\mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD} = \begin{pmatrix} 0 & -I_m & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & -I_m & 0 \\ I_m & \cdots & I_m & I_m & I_m \\ -I_m & \cdots & 0 & 0 & 0 \end{pmatrix}. \tag{4.8}$$

最后得到

$$\mathcal{M}_{PD} = \frac{1}{2}\big(\mathcal{I}_{p+1} + \mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD}\big) = \frac{1}{2} \begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & I_m & -I_m & 0 \\ I_m & \cdots & I_m & 2I_m & I_m \\ -I_m & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{4.9}$$

利用前一讲的变换, 采用 (4.9) 中的矩阵 $\mathcal{M}_{PD}$ 的校正 (3.6) 可以写成等价的

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \\ \lambda^{k} \end{pmatrix} - \frac{1}{2} \begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & I_m & -I_m & 0 \\ I_m & \cdots & I_m & 2I_m & \frac{1}{\beta}I_m \\ -\beta I_m & 0 & \cdots & 0 & I_m \end{pmatrix} \begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix} \tag{4.10}$$

4.2 Dual-Primal 预测的广义 PPA 算法

设预测时由前一讲中的 Dual-Primall 预测给出的. 我们得到形如 (3.2) 的变分不等式, 其中

$$\mathcal{Q}_{DP}=\begin{pmatrix} I_m & 0 & \cdots & 0 & 0\\ I_m & I_m & \ddots & \vdots & 0\\ \vdots & & \ddots & 0 & \vdots\\ I_m & I_m & \cdots & I_m & 0\\ -I_m & -I_m & \cdots & -I_m & I_m \end{pmatrix}. \tag{4.11}$$

利用记号 $\mathcal{L}$, $\mathcal{E}$, 那么

$$\mathcal{Q}_{DP}=\begin{pmatrix} \mathcal{L} & 0\\ -\mathcal{E}^T & I_m \end{pmatrix}. \tag{4.12}$$

由于

$$\mathcal{M}_{DP}=\frac{1}{2}(\mathcal{I}+\mathcal{Q}_{DP}^{-T}\mathcal{Q}_{DP}).$$

我们先来考察一下 $\mathcal{Q}_{DP}^{-T}\mathcal{Q}_{DP}$. 注意到

$$\mathcal{Q}_{DP}^{T}=\begin{pmatrix} \mathcal{L}^T & -\mathcal{E}\\ 0 & I_m \end{pmatrix} \quad 和 \quad \mathcal{Q}_{DP}^{-T}=\begin{pmatrix} \mathcal{L}^{-T} & \mathcal{L}^{-T}\mathcal{E}\\ 0 & I_m \end{pmatrix}.$$

则有

$$\mathcal{Q}_{DP}^{-T}\mathcal{Q}_{DP}=\begin{pmatrix} \mathcal{L}^{-T} & \mathcal{L}^{-T}\mathcal{E}\\ 0 & I_m \end{pmatrix}\begin{pmatrix} \mathcal{L} & 0\\ -\mathcal{E}^T & I_m \end{pmatrix}=\begin{pmatrix} \mathcal{L}^{-T}\mathcal{L}-\mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T & \mathcal{L}^{-T}\mathcal{E}\\ -\mathcal{E}^T & I_m \end{pmatrix}. \tag{4.13}$$

分别计算分块矩阵 $\mathcal{Q}_{DP}^{-T}\mathcal{Q}_{DP}$ 中的四块. 从 (4.4) 和 (4.5) 我们已经有了

$$\mathcal{L}^{-T}\mathcal{L}=\begin{pmatrix} 0 & -I_m & 0 & 0\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & -I_m\\ I_m & \cdots & I_m & I_m \end{pmatrix} \quad 和 \quad \mathcal{L}^{-T}\mathcal{E}=\begin{pmatrix} 0\\ \vdots\\ 0\\ I_m \end{pmatrix},$$

因此 (4.13) 的左上角部分

$$\mathcal{L}^{-T}\mathcal{L} - \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T = \mathcal{L}^{-T}\mathcal{L} - \begin{pmatrix} 0 \\ \vdots \\ 0 \\ I_m \end{pmatrix} \begin{pmatrix} I_m & I_m & \cdots & I_m \end{pmatrix}$$
$$= \begin{pmatrix} 0 & -I_m & 0 & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -I_m \\ 0 & \cdots & 0 & 0 \end{pmatrix}.$$

矩阵 $\mathcal{Q}_{DP}^{-T}\mathcal{Q}_{DP}$ 的右上角部分, $\mathcal{L}^{-T}\mathcal{E}$ 在 (4.5) 中已经有了交代. 所以

$$\mathcal{Q}_{DP}^{-T}\mathcal{Q}_{DP} = \begin{pmatrix} \mathcal{L}^{-T}\mathcal{L} - \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T & \mathcal{L}^{-T}\mathcal{E} \\ -\mathcal{E}^T & I_m \end{pmatrix}$$
$$= \begin{pmatrix} 0 & -I_m & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -I_m & 0 \\ 0 & \cdots & 0 & 0 & I_m \\ -I_m & \cdots & \cdots & -I_m & I_m \end{pmatrix}. \qquad (4.14)$$

最后, 我们得到

$$\mathcal{M}_{DP} = \frac{1}{2}(\mathcal{I} + \mathcal{Q}_{DP}^{-T}\mathcal{Q}_{DP}) = \frac{1}{2}\begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -I_m & 0 \\ 0 & \cdots & 0 & I_m & I_m \\ I_m & \cdots & I_m & I_m & 2I_m \end{pmatrix}. \quad (4.15)$$

由相应的变换, 采用 (4.15) 中的矩阵 $\mathcal{M}_{DP}$ 的校正 (3.6) 可以写成等价的

$$\begin{pmatrix} A_1x_1^{k+1} \\ A_2x_2^{k+1} \\ \vdots \\ A_px_p^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} A_1x_1^{k} \\ A_2x_2^{k} \\ \vdots \\ A_px_p^{k} \\ \lambda^{k} \end{pmatrix} - \frac{1}{2}\begin{pmatrix} I_m & -I_m & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -I_m & 0 \\ 0 & \cdots & 0 & I_m & \frac{1}{\beta}I_m \\ -\beta I_m & -\beta I_m & \cdots & -\beta I_m & 2I_m \end{pmatrix}\begin{pmatrix} A_1x_1^k - A_1\tilde{x}_1^k \\ A_2x_2^k - A_2\tilde{x}_2^k \\ \vdots \\ A_px_p^k - A_p\tilde{x}_p^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \quad (4.16)$$

5 基于秩二校正的广义 PPA 算法

第六讲 §7 讲述的方法, 预测产生的 $\mathcal{Q}$ 矩阵是一个容易求逆的矩阵与一个广义秩二矩阵的和. 广义 PPA 的校正完全是由预测决定的. 因此, 广义邻近点算法只需根据预测矩阵 $\mathcal{Q}$ 给出相应的校正矩阵 $\mathcal{M}$.

5.1 Primal-Dual 预测的广义 PPA 算法

设预测是由第六讲 §7.1 讲述的方法中的 Primal-Dual 预测给出的, 我们得到形如 (3.2) 的变分不等式, 其中的 $\mathcal{Q}$ 我们记为 $\mathcal{Q}_{PD}$.

$$\mathcal{Q}_{PD} = \begin{pmatrix} I_m & 0 & \cdots & 0 & I_m \\ I_m & I_m & \ddots & \vdots & I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & I_m \\ I_m & I_m & \cdots & I_m & \frac{5}{2}I_m \end{pmatrix} = \begin{pmatrix} \mathcal{L} & \mathcal{E} \\ \mathcal{E}^T & \frac{5}{2}I_m \end{pmatrix}. \quad (5.1)$$

我们要给出

$$\mathcal{M} = \frac{1}{2}\mathcal{Q}_{PD}^{-T}(\mathcal{Q}_{PD}^{T} + \mathcal{Q}_{PD}),$$

首先看给出 $\mathcal{Q}_{PD}^{-T}$ 的形式. 由第六讲的 §7.1 节的 (7.14) 式

$$\mathcal{Q}_{PD}^{-T} = \frac{2}{3}\begin{pmatrix} \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T} & -\mathcal{L}^{-T}\mathcal{E} \\ -\mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix} + \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ 0 & 0 \end{pmatrix}.$$

得到

$$\begin{aligned} \mathcal{Q}_{PD}^{-T}\mathcal{Q}_{PD} &= \left\{\frac{2}{3}\begin{pmatrix} \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T} & -\mathcal{L}^{-T}\mathcal{E} \\ -\mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix} + \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ 0 & 0 \end{pmatrix}\right\}\begin{pmatrix} \mathcal{L} & \mathcal{E} \\ \mathcal{E}^T & \frac{5}{2}I_m \end{pmatrix} \\ &= \frac{2}{3}\begin{pmatrix} \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T}\mathcal{L} - \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T & \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T}\mathcal{E} - \frac{5}{2}\mathcal{L}^{-T}\mathcal{E} \\ -\mathcal{E}^T\mathcal{L}^{-T}\mathcal{L} + \mathcal{E}^T & -\mathcal{E}^T\mathcal{L}^{-T}\mathcal{E} + \frac{5}{2}I_m \end{pmatrix} \\ &\quad + \begin{pmatrix} \mathcal{L}^{-T}\mathcal{L} & \mathcal{L}^{-T}\mathcal{E} \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

因此校正矩阵

$$\mathcal{M}_{PD} = \frac{1}{2}\mathcal{Q}_{PD}^{-T}(\mathcal{Q}_{PD}^{T} + \mathcal{Q}_{PD}) = \frac{1}{3}\mathcal{B}_{PD} + \frac{1}{2}\mathcal{C}_{PD}, \tag{5.2}$$

其中

$$\mathcal{B}_{PD} = \begin{pmatrix} \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T}\mathcal{L} - \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T & \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T}\mathcal{E} - \frac{5}{2}\mathcal{L}^{-T}\mathcal{E} \\ -\mathcal{E}^T\mathcal{L}^{-T}\mathcal{L} + \mathcal{E}^T & -\mathcal{E}^T\mathcal{L}^{-T}\mathcal{E} + \frac{5}{2}I_m \end{pmatrix} \tag{5.3}$$

和

$$\mathcal{C}_{PD} = \begin{pmatrix} \mathcal{I} + \mathcal{L}^{-T}\mathcal{L} & \mathcal{L}^{-T}\mathcal{E} \\ 0 & I_m \end{pmatrix}. \tag{5.4}$$

我们先计算矩阵 $\mathcal{B}_{PD}$ 的四块. 利用 (参见第六讲 §7.1节)

$$\mathcal{L}^{-T}\mathcal{E} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ I_m \end{pmatrix}, \quad \mathcal{E}^T\mathcal{L}^{-T} = \big(I_m, 0, \ldots, 0\big) \quad \text{和} \quad \mathcal{L} = \begin{pmatrix} I_m & 0 & \cdots & 0 \\ I_m & I_m & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ I_m & I_m & \cdots & I_m \end{pmatrix},$$

得到

$$\mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T}\mathcal{L} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \\ I_m & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} I_m & 0 & \cdots & 0 \\ I_m & I_m & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ I_m & I_m & \cdots & I_m \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \\ I_m & 0 & \cdots & 0 \end{pmatrix}$$

和

$$\mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \\ I_m & I_m & \cdots & I_m \end{pmatrix}$$

因此矩阵$\mathcal{B}_{PD}$ 的 (1,1) 块

$$\mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T}\mathcal{L} - \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \\ 0 & -I_m & \cdots & -I_m \end{pmatrix}.$$

矩阵$\mathcal{B}_{PD}$ 的 (1,2) 块

$$\mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T}\mathcal{E} - \frac{5}{2}\mathcal{L}^{-T}\mathcal{E} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ I_m \end{pmatrix} - \frac{5}{2}\begin{pmatrix} 0 \\ \vdots \\ 0 \\ I_m \end{pmatrix} = -\frac{3}{2}\begin{pmatrix} 0 \\ \vdots \\ 0 \\ I_m \end{pmatrix}.$$

矩阵$\mathcal{B}_{PD}$ 的 (2,1) 块

$$\mathcal{E}^T - \mathcal{E}^T\mathcal{L}^{-T}\mathcal{L} = \mathcal{E}^T - \big(I_m,\, 0,\, \ldots,\, 0\big)\begin{pmatrix} I_m & 0 & \cdots & 0 \\ I_m & I_m & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ I_m & I_m & \cdots & I_m \end{pmatrix}$$
$$= \big(0,\, I_m,\, \ldots,\, I_m\big).$$

矩阵$\mathcal{B}_{PD}$ 的 (2,2) 块

$$\frac{5}{2}I_m - \mathcal{E}^T\mathcal{L}^{-T}\mathcal{E} = \frac{3}{2}I_m.$$

所以

$$\mathcal{B}_{PD} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & -I_m & \cdots & -I_m & -\frac{3}{2}I_m \\ 0 & I_m & \cdots & I_m & \frac{3}{2}I_m \end{pmatrix}. \qquad (5.5)$$

利用

$$\mathcal{L}^{-T}\mathcal{L} = \begin{pmatrix} 0 & -I_m & 0 & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -I_m \\ I_m & \cdots & I_m & I_m \end{pmatrix} \quad \text{和} \quad \mathcal{L}^{-T}\mathcal{E} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ I_m \end{pmatrix},$$

得到

$$\mathcal{C}_{PD} = \begin{pmatrix} \mathcal{I} + \mathcal{L}^{-T}\mathcal{L} & \mathcal{L}^{-T}\mathcal{E} \\ 0 & I_m \end{pmatrix} = \begin{pmatrix} I_m & -I_m & 0 & 0 & 0 \\ 0 & \ddots & \ddots & 0 & \vdots \\ 0 & 0 & I_m & -I_m & 0 \\ I_m & \cdots & I_m & 2I_m & I_m \\ 0 & \cdots & \cdots & 0 & I_m \end{pmatrix}. \tag{5.6}$$

有了 (5.5) 和 (5.6), 校正矩阵

$$\begin{aligned} \mathcal{M}_{PD} &= \frac{1}{3}\mathcal{B}_{PD} + \frac{1}{2}\mathcal{C}_{PD} \\ &= \frac{1}{3}\begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & -I_m & \cdots & -I_m & -\frac{3}{2}I_m \\ 0 & I_m & \cdots & I_m & \frac{3}{2}I_m \end{pmatrix} \\ &\quad + \frac{1}{2}\begin{pmatrix} I_m & -I_m & 0 & 0 & 0 \\ 0 & \ddots & \ddots & 0 & \vdots \\ 0 & 0 & I_m & -I_m & 0 \\ I_m & \cdots & I_m & 2I_m & I_m \\ 0 & \cdots & \cdots & 0 & I_m \end{pmatrix} \end{aligned}$$

的形式是非常简单的.

5.2 Dual-Primal 预测的广义 PPA 算法

由第六讲 §7.2 讲述的方法 中的Dual-Primal 预测给出的, 我们得到形如 (3.2) 的变分不等式, 其中的 $\mathcal{Q}$ 我们记为 $\mathcal{Q}_{DP}$.

$$\mathcal{Q}_{DP} = \begin{pmatrix} I_m & 0 & \cdots & 0 & -I_m \\ I_m & I_m & \ddots & \vdots & -I_m \\ \vdots & & \ddots & 0 & \vdots \\ I_m & I_m & \cdots & I_m & -I_m \\ -I_m & -I_m & \cdots & -I_m & \frac{5}{2}I_m \end{pmatrix} = \begin{pmatrix} \mathcal{L} & -\mathcal{E} \\ -\mathcal{E}^T & \frac{5}{2}I_m \end{pmatrix}. \tag{5.7}$$

我们要给出

$$\mathcal{M} = \frac{1}{2}\mathcal{Q}_{PD}^{-T}(\mathcal{Q}_{PD}^T + \mathcal{Q}_{PD}),$$

首先看给出 $\mathcal{Q}_{DP}^{-T}$ 的形式. 由第六讲§7.2节中的(7.21) 式

$$\mathcal{Q}_{DP}^{-T} = \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ 0 & 0 \end{pmatrix} + \frac{2}{3}\begin{pmatrix} \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T} & \mathcal{L}^{-T}\mathcal{E} \\ \mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix}.$$

得到

$$\begin{aligned} \mathcal{Q}_{DP}^{-T}\mathcal{Q}_{DP} &= \left\{ \frac{2}{3}\begin{pmatrix} \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T} & \mathcal{L}^{-T}\mathcal{E} \\ \mathcal{E}^T\mathcal{L}^{-T} & I_m \end{pmatrix} + \begin{pmatrix} \mathcal{L}^{-T} & 0 \\ 0 & 0 \end{pmatrix} \right\} \begin{pmatrix} \mathcal{L} & -\mathcal{E} \\ -\mathcal{E}^T & \frac{5}{2}I_m \end{pmatrix} \\ &= \frac{2}{3}\begin{pmatrix} \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T}\mathcal{L} - \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T & -\mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T}\mathcal{E} + \frac{5}{2}\mathcal{L}^{-T}\mathcal{E} \\ \mathcal{E}^T\mathcal{L}^{-T}\mathcal{L} - \mathcal{E}^T & -\mathcal{E}^T\mathcal{L}^{-T}\mathcal{E} + \frac{5}{2}I_m \end{pmatrix} \\ &\quad + \begin{pmatrix} \mathcal{L}^{-T}\mathcal{L} & -\mathcal{L}^{-T}\mathcal{E} \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

因此校正矩阵

$$\mathcal{M}_{DP} = \frac{1}{2}\mathcal{Q}_{DP}^{-T}(\mathcal{Q}_{DP}^T + \mathcal{Q}_{DP}) = \frac{1}{3}\mathcal{B}_{DP} + \frac{1}{2}\mathcal{C}_{DP}, \tag{5.8}$$

其中

$$\mathcal{B}_{DP} = \begin{pmatrix} \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T}\mathcal{L} - \mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T & -\mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T}\mathcal{E} + \frac{5}{2}\mathcal{L}^{-T}\mathcal{E} \\ \mathcal{E}^T\mathcal{L}^{-T}\mathcal{L} - \mathcal{E}^T & -\mathcal{E}^T\mathcal{L}^{-T}\mathcal{E} + \frac{5}{2}I_m \end{pmatrix} \tag{5.9}$$

和

$$\mathcal{C}_{DP} = \begin{pmatrix} \mathcal{I} + \mathcal{L}^{-T}\mathcal{L} & -\mathcal{L}^{-T}\mathcal{E} \\ 0 & I_m \end{pmatrix}. \tag{5.10}$$

将 (5.9) 跟 (5.3) 比较, 利用 (5.5), 有

$$\mathcal{B}_{DP} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & -I_m & \cdots & -I_m & \frac{3}{2}I_m \\ 0 & -I_m & \cdots & -I_m & \frac{3}{2}I_m \end{pmatrix}. \tag{5.11}$$

将 (5.10) 跟 (5.4) 比较, 利用 (5.6), 有

$$\begin{aligned} \mathcal{C}_{DP} &= \begin{pmatrix} \mathcal{I} + \mathcal{L}^{-T}\mathcal{L} & -\mathcal{L}^{-T}\mathcal{E} \\ 0 & I_m \end{pmatrix} \\ &= \begin{pmatrix} I_m & -I_m & 0 & 0 & 0 \\ 0 & \ddots & \ddots & 0 & \vdots \\ 0 & 0 & I_m & -I_m & 0 \\ I_m & \cdots & I_m & 2I_m & -I_m \\ 0 & \cdots & \cdots & 0 & I_m \end{pmatrix}. \end{aligned} \tag{5.12}$$

有了 (5.11) 和 (5.12), 校正矩阵

$$
\begin{aligned}
\mathcal{M}_{DP} &= \frac{1}{3}\mathcal{B}_{DP} + \frac{1}{2}\mathcal{C}_{DP} \\
&= \frac{1}{3}\begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & -I_m & \cdots & -I_m & \frac{3}{2}I_m \\ 0 & -I_m & \cdots & -I_m & \frac{3}{2}I_m \end{pmatrix} \\
&\quad + \frac{1}{2}\begin{pmatrix} I_m & -I_m & 0 & 0 & 0 \\ 0 & \ddots & \ddots & 0 & \vdots \\ 0 & 0 & I_m & -I_m & 0 \\ I_m & \cdots & I_m & 2I_m & -I_m \\ 0 & \cdots & \cdots & 0 & I_m \end{pmatrix}
\end{aligned}
$$

的形式是非常简单的.

6 Conclusions

- 我的学术报告中常用的一个题目是 "构造凸优化的分裂收缩算法– 用好 VI 和 PPA 两大法宝", 是指构造变分不等式意义下的 PPA 算法, 文章首先发表在 [14]. 后来又做了一些人为地将预测矩阵设计成对称正定矩阵的方法 [1, 3], 包括我们 2021 年才提出的均困平衡的增广拉格朗日乘子法 [17]. 有时我们也称这样的方法为按需定制的 PPA - (Customized PPA).
- 对预测矩阵 Q 为非对称的预测-校正方法, 利用统一框架的套路证明收敛性, 最初出现在我和袁晓明 (Xiaoming Yuan) 2012 年 SIAM 数值分析的文章 [13] 中, 后面我们发表的一些论文 [8, 10, 11, 16], 都用这个套路证明收敛性. 把它归结为统一框架, 是在南京大学讨论班上, 那是在我 2013 年即将退休之前, 以后便常常出现在我的 "讲习班" 讲义和报告的 PPT 中.
- 第一次在正式出版物里提到这个统一框架, 是在 2016 年《高校计算数学学报》的我的中文文章 [4] 中. 2018 年我在《运筹学学报》的综述文章 "我和乘子交替方向法 20 年"[5] 中指出, 我们发表的方法都可以用这个框架非常简单地证明收敛性. 英文出版物中首次出现统一框架的是我和袁晓明 2018 年在 COAP 的文章 [15].
- 从 2018 年开始, 我在自己的报告和论文 [7] 中, 经常讲用统一框架去构造算法主要还是按收敛条件去 "凑". 如何根据确定的预测矩阵 $\boldsymbol{Q}$ 凑出满足收敛条件的校正矩阵 $\boldsymbol{M}$. 似乎给人一种难以效仿的神秘感觉.

- 2022年初我在南师大做报告时有人问过这样的问题, 后来我又在中科大和南航做线上报告, 教学相长, 得到一些新的, 整理成下面的材料与听众共享.
- 我们从预测矩阵满足 $Q^T + Q \succ 0$ 出发. 根据条件 $HM = Q$, 我们有

$$H = QM^{-1}.$$

因为 H 是正定矩阵, 必须对称. 从上式又看到, H 有个左因子 Q, 那它必须有个右因子 Q^T, 中间夹一个“待定的”正定矩阵. 我们设这个正定矩阵为 D^{-1}, 则有

$$H = QD^{-1}Q^T.$$

比较上面两式, 我们得到 $M^{-1} = D^{-1}Q^T$, 因此

$$\boldsymbol{M = Q^{-T}D}.$$

这个我们大概在10年前就知道. 当时往往考虑选择的 D 应该是个块对角矩阵.

- 至此, 我们还不知道矩阵 D 具体形式是什么. 计算一下收敛性条件中的 M^THM,

$$M^THM = \big(DQ^{-1}\big)\big(QD^{-1}Q^T\big)\big(Q^{-T}D\big) = D.$$

上式已经出现在我2018的暑期讲习班的讲义中, 没有向前再迈一步.

- 利用上式和 $G = Q^T + Q - M^THM \succ 0$, 这个待定的正定矩阵 D 只需要满足

$$\boldsymbol{0 \prec D \prec Q^T + Q} \qquad (\text{因此}, 0 \prec G = Q^T + Q - D)$$

就可以了. 明确这一条, 得益于为2022年以来在南师大(线下), 南航 和 中科大(线上)讲课, 促使我深入思考, 努力想办法讲清楚.

- 在选了满足上述条件的矩阵 D 以后, 根据确定的 Q 和 D, 找未知矩阵 H 和 M 使得

$$HM = Q \qquad \text{和} \qquad M^THM = D,$$

我们的目的就达到了.

- 这样的 M 和 H: 可以通过求解下面的矩阵方程组得到.

$$\left\{\begin{array}{rcl} HM &=& Q, \\ M^THM &=& D. \end{array}\right. \Leftrightarrow \left\{\begin{array}{rcl} HM &=& Q, \\ Q^TM &=& D. \end{array}\right. \Leftrightarrow \left\{\begin{array}{rcl} H &=& QD^{-1}Q^T, \\ M &=& Q^{-T}D. \end{array}\right. .$$

- 选择不同的满足条件的矩阵 D (这非常容易), 就有不同的校正方法. 譬如说,

$$D = \alpha[Q^T + Q], \quad \alpha \in (0,1).$$

- 前一讲的方法, 对一般线性约束凸优化问题, 采用 primal-dual 预测, 子问题的求解方式是 ADMM 类型的逐个向前. 我们需要的 $\boldsymbol{Q^{-T}}$ 形式非常简单. 是的, 它需要额外的校正. 可喜的是, 校正花费很少, 又特别容易实现!

- 我们特别推崇“预测-校正”, 尤其是那种代价很小的校正. 生机勃勃的果树, 修剪就是校正. 社会治理也是一种校正, 当然也考虑成本! 交替按序预测, 降低了问题难度; 全局整体校正, 把握了收敛方向.

- 预测-校正方法既可以用来求解等式约束的问题, 又可以用来求解不等式约束的问题. 适用从一块到任意多块的可分离问题, 算法结构和收敛性证明完全统一.
- 适用范围广的算法会不会影响效率？对经典 ADMM 擅长的两块可分离的等式约束凸优化问题, 我们用前一讲提到的带校正的交替方向法去求解, 与网上他人提供的 ADMM 代码比较, 发现这种担心是多余的.
- **Question A.** In the prediction step, how to arrange a "good" prediction matrix whose matrix $\mathcal{Q}$ satisfies
$$\mathcal{Q}^T + \mathcal{Q} \succeq \mathcal{I}.$$
- **Question B** For the given prediction matrix $\mathcal{Q}$, what are the criteria for choosing matrix $\mathcal{D}$ which satisfies
$$0 \prec \mathcal{D} \prec \mathcal{Q}^T + \mathcal{Q}.$$
- 这一讲介绍的广义 PPA 算法, 是由预测唯一确定的. 对给定的预测矩阵 Q, 取
$$D = \tfrac{1}{2}(Q^T + Q), \qquad 和 \qquad v^{k+1} = v^k - Q^{-T}D(v^k - \tilde{v}^k),$$
迭代序列 $\{v^k\}$ 满足（其中 $H = QD^{-1}Q^T$）
$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2$$
和
$$\|v^{k+1} - v^{k+2}\|_H^2 \le \|v^k - v^{k+1}\|_H^2.$$

为什么说我们在数值优化方面做出了一些颇有特色又自成系统的工作呢？

首先, 变分不等式和邻近点算法是我们的主要工具. 任何一本关于数值优化的书, 都没有专门提及变分不等式 (VI), 也不会刻意介绍邻近点算法 (PPA), 尽管线性约束的凸优化问题的增广拉格朗日乘子法 (ALM) 是乘子 λ 的 PPA 算法.

- 我们把线性约束的凸优化问题转换成一个等价的结构型单调变分不等式，然后说明什么是变分不等式的 PPA 算法, 讨论了 PPA 算法的收敛性质.
- 变分不等式的 PPA 算法迭代的每一步, 都利用其可分离结构, 分解成一些简单的"小微"变分不等式, 求解这些小微变分不等式, 又可以通过求解相应的凸优化问题实现.
- 后来我们又有了基于 VI 的预测-校正方法的统一框架, 既可以用它来验证算法的收敛性, 又可以用它"按需设计"求解可分离凸优化问题的算法, 这就是我们与众不同的逻辑.
- 我们又应该保持清醒的头脑, 即使是 ADMM, 它也是松弛了的 ALM, 是关于乘子 λ 的 PPA 算法. 同时也可以强调, 求解线性约束凸优化问题, ALM 是个有竞争力的好方法.

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.

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] T. Goldstein, B. O' Donoghue, S. Setzer and R. Baraniuk, Fast Alternating Direction Optimization Methods, SIAM J. Imaging Science, Vol. 7, No. 3, pp. 1588 – 1623, 2014.

[3] 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.

[4] B. S. He, From the projection and contraction methods for variatioonal inequality to the splittingcontraction methods fro convex optimization (in Chinese), Numerical Mathematics, 高等学校计算数学学报 38 (2016), 74-96.

[5] B. S. He, My 20 years research on alternating directions method of multipliers (in Chinese), Operations Research Trasactions, 22 (2018), 1-31.

[6] B. S. He, A uniform framework of contraction methods for convex optimization and monotone variational inequality (in Chinese). Sci Sin Math, 48 (2018), 255-272, doi: 10.1360/N012017-00034

[7] B. S. He, Using a unified framework to design the splitting and contraction methods for convex optimization (in Chinese), 高等学校计算数学学报 44 (2022), 1-35.

[8] B. S. He, H. Liu, Z. R. Wang and X. M. Yuan, A strictly Peaceman-Rachford splitting method for convex programming, SIAM J. Optim. 24 (2014), 1011-1040.

[9] B.S. He, F. Ma, S. J. Xu and X. M. Yuan, A rank-two relaxed parallel splitting version of the augmented Lagrangian method with step size in (0,2) for separable convex programming, Mathematics of Computation, 92(2023), 1633-1663.

[10] 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.

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

[12] B. S. He, S. J. Xu and X. M. Yuan, Extensions of ADMM for separable convex optimization problems with linear equality or inequality constraints, arXiv:2107.01897v2[math.OC].

[13] B. S. He and X. M. Yuan, On the $O(1/n)$ convergence rate of the alternating direction method, SIAM J. Numerical Analysis, 50 (2012), 700-709.

[14] B. S. He and X. M. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective, SIAM Journal on Imaging Science, 5 (2012) 119-149.

[15] B. S. He and X. M. Yuan, A class of ADMM-based algorithms for three-block separable convex programming, Comput. Optim. Appl. 70 (2018), 791-826.

[16] 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.

[17] B. S. He and X. M. Yuan, Balanced Augmented Lagrangian Method for Convex Programming, arXiv 2108.08554 [math.OC]

[18] B. S. He and X. M. Yuan, On construction of splitting contraction algorithms in a prediction-correction framework for separable convex optimization, arXiv 2204.11522 [math.OC]

变分不等式框架下结构型凸优化的分裂收缩算法

VIII. 投影收缩算法和可微凸优化问题的算法设计

中学的数理基础　必要的社会实践
普通的大学数学　一般的优化原理

何 炳 生　　南京大学数学系

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

天元数学东北中心　2023 年 10 月 17 – 27 日

1　引言

研究凸优化分裂收缩算法的统一框架, 跟我们长期研究经典单调变分不等式的投影收缩算法 [1, 3, 4, 5, 6, 9, 10, 11, 14] 有着密切的联系. 变分不等式是描述平衡问题的数学工具, 在管理科学和工程计算中都有广泛的应用.

设 $\Omega \subset \Re^n$ 是一个非空闭凸集, $\mathbb{F}$ 是 $\Re^n \rightarrow \Re^n$ 的一个映射. 我们考虑单调变分不等式

$$\text{VI}(\Omega, \mathbb{F}) \qquad u^* \in \Omega, \quad (u-u^*)^T \mathbb{F}(u^*) \ge 0, \quad \forall\, u \in \Omega \tag{1.1}$$

的求解. 变分不等式 (1.1) 单调, 是指其中的算子 $\mathbb{F}$ 满足

$$(u-v)^T(\mathbb{F}(u) - \mathbb{F}(v)) \ge 0, \quad \forall\, u, v \in \Re^n \text{ (或 } \Omega). \tag{1.2}$$

这里的单调算子 $\mathbb{F}$ 可以是非线性的, 有别于由凸优化转换而来的变分不等式中的仿射算子 F, 其中的系数矩阵是反对称矩阵, 具备性质

$$(u-v)^T(F(u) - F(v)) \ge 0, \quad \forall\, u, v \in \Re^n \text{ (或 } \Omega).$$

为了有所区别, 在这一讲的讨论中, 我们把由线性约束凸优化问题转换得来的变分不等式称为混合变分不等式, 形式为 (1.1) 的变分不等式称为经典变分不等式.

非线性互补问题是一类特殊的变分不等式

在第一讲中, 我们就讨论过凸优化问题

$$\min\{f(x) \mid x \in \Omega\}. \tag{1.3}$$

它的最优性条件

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

当 $\Omega = \Re^n_+$ 时, 就是一个约束集合为非负卦限的变分不等式

$$x^* \geq 0, \quad (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall x \geq 0.$$

设 F 是 $\Re^n \to \Re^n$ 的一个算子. 非负卦限 $\Re^n_+$ 上变分不等式的一般形式是

$$\text{VI}(\Re^n_+, F) \qquad x \geq 0, \quad (x' - x)^T F(x) \geq 0, \quad \forall x' \geq 0. \tag{1.4}$$

非线性互补问题是最优化理论与方法中一类很重要的问题. 它的数学形式是

$$\text{(NCP)} \qquad x \geq 0, \quad F(x) \geq 0, \quad x^T F(x) = 0. \tag{1.5}$$

事实上, NCP 是 $\Omega = \Re^n_+$ 的一类变分不等式.

定理 1 非负卦限上的变分不等式 (1.4) 和互补问题 (1.5) 是等价的.

证明 如果 x 是互补问题 (1.5) 的解, 那么有 $x \geq 0$ 和 $F(x) \geq 0$. 对于任意的 $x' \geq 0$ 有 ${x'}^T F(x) \geq 0$. 又因 $x^T F(x) = 0$, 所以有

$$x \geq 0, \quad (x' - x)^T F(x) = {x'}^T F(x) - x^T F(x) \geq 0, \quad \forall x' \geq 0.$$

所以 x 是 VI(R^n_+, F) (1.4) 的一个解.

反过来, 如果 x 是 VI(R^n_+, F) (1.4) 的一个解, 则 $x \geq 0$. 将 $x' = 2x$ 和 $x' = 0$ 代入

$$(x' - x)^T F(x) \geq 0,$$

得到 $\pm x^T F(x) \geq 0$. 因此 $x^T F(x) = 0$.

要证明 x 是互补问题 (1.5) 的解, 只剩下 $F(x) \geq 0$ 需要证明, 对此采用反证法. 如果 $F(x)$ 的某个分量 $F_j(x) < 0$, 我们取 x', 使得

$$x'_i = \begin{cases} x_i, & \text{if } i \neq j \\ x_j + 1, & \text{if } i = j \end{cases}$$

这样的 $x' \geq 0$. 但 $(x' - x)^T F(x) = F_j(x) < 0$, 这与 x 是 VI(R^n_+, F) (1.4) 的解矛盾.

因此, 在 $\Omega = \Re^n_+$ 时, 可微凸优化问题 (1.3) 与一个互补问题等价.

经典单调变分不等式的投影收缩算法也是一种预测-校正方法. 经典变分不等式框架下的投影收缩算法也为线性约束可微凸优化问题的求解提供了一些其他的途径.

这一讲的内容安排上,

- 首先介绍经典的单调变分不等式及其等价的投影方程,
- 经典变分不等式投影收缩算法和凸优化分裂收缩算法中的预测
- 经典变分不等式投影收缩算法和凸优化分裂收缩算法中的预测校正
- 投影收缩算法中的孪生方向和姊妹方法
- 投影收缩算法在求解可分离凸优化上的应用

2 投影与变分不等式的一些基本性质

投影收缩算法中的基本运算是执行向量到简单闭凸集上的投影.

2.1 投影的基本性质

用 $P_\Omega(\cdot)$ 表示欧氏范数下在凸集 Ω 上的投影, 也就是说

$$P_\Omega(v) = \text{Argmin}\{\|u - v\| \mid u \in \Omega\}.$$

如果 $\Omega = \Re^n_+$ (n-维空间的非负卦限), 那么 $P_\Omega(v)$ 的每个分量为

$$(P_\Omega(v))_j = \begin{cases} v_j, & \text{if} \quad v_j \ge 0; \\ 0, & \text{otherwise}. \end{cases}$$

如果 Ω 是 n-维空间中以 c 为球心半径为 r 的球, 那么

$$P_\Omega(v) = \begin{cases} \frac{r(v-c)}{\|v-c\|} + c, & \text{if} \quad \|v - c\| \ge r; \\ v, & \text{otherwise}. \end{cases}$$

在 l_∞ 和 l_1 模意义下的“单位球”上投影如下图所示:

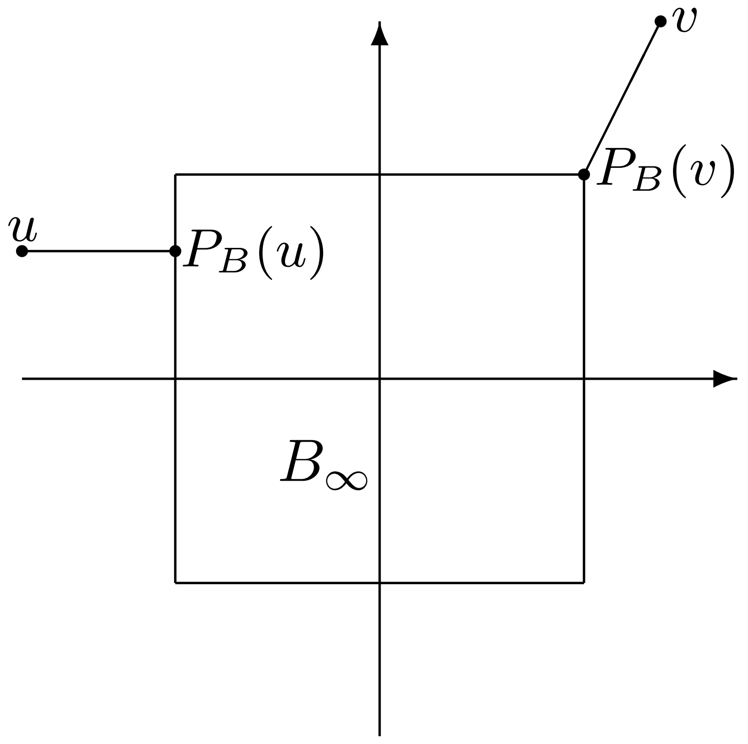


Fig. 1. Projection on B_∞

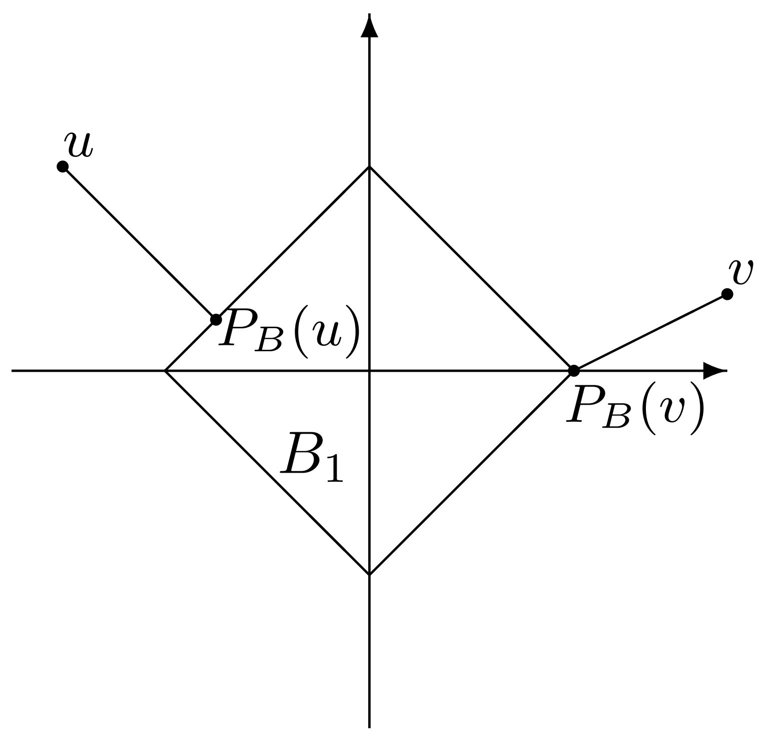


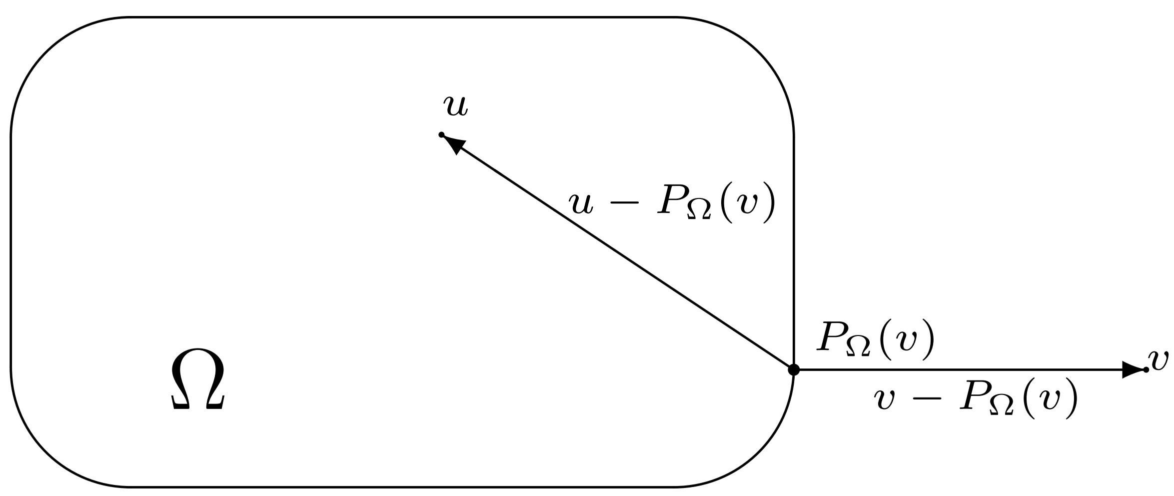
Fig. 2. Projection on B_1

通篇, 我们记

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

引理 1 设 $\Omega \subset R^n$ 是闭凸集, 则有

$$(v - P_\Omega(v))^T(u - P_\Omega(v)) \le 0 \qquad \forall v \in R^n, \forall u \in \Omega. \tag{2.1}$$



不等式(2.1) 的几何解释.

证明. 首先, 根据 $P_\Omega(v)$ 的定义, 有

$$\|v - P_\Omega(v)\| \le \|v - w\|, \quad \forall w \in \Omega. \tag{2.2}$$

注意到对任意的 $v \in \Re^n$, 都有 $P_\Omega(v) \in \Omega$, 由于 $\Omega \subset \Re^n$ 是闭凸集, 则对任意的 $u \in \Omega$ 和 $\theta \in (0, 1)$, 都有

$$w := \theta u + (1 - \theta)P_\Omega(v) = P_\Omega(v) + \theta(u - P_\Omega(v)) \in \Omega.$$

对这个 w, 利用 (2.2), 就有

$$\|v - P_\Omega(v)\|^2 \le \|v - P_\Omega(v) - \theta(u - P_\Omega(v))\|^2.$$

对上式展开, 对任意的 $u \in \Omega$ 和 $\theta \in (0,1)$, 都有

$$[v - P_\Omega(v)]^T[u - P_\Omega(v)] \le \frac{\theta}{2}\|u - P_\Omega(v)\|^2.$$

令 $\theta \to 0_+$, 引理 (2.1) 得证. $\square$

在投影收缩算法的分析中, 不等式 (2.1) 是一个非常有用的基本工具. 我们因此而称之为 投影算子的工具不等式. 由(2.1), 容易证明下面的引理.

引理 2 设 $\Omega \subset R^n$ 是闭凸集, 则有

$$\|P_\Omega(v) - P_\Omega(u)\| \le \|v - u\|, \qquad \forall u, v \in \Re^n. \tag{2.3}$$

$$\|P_\Omega(v) - u\| \le \|v - u\|, \qquad \forall v \in \Re^n, u \in \Omega. \tag{2.4}$$

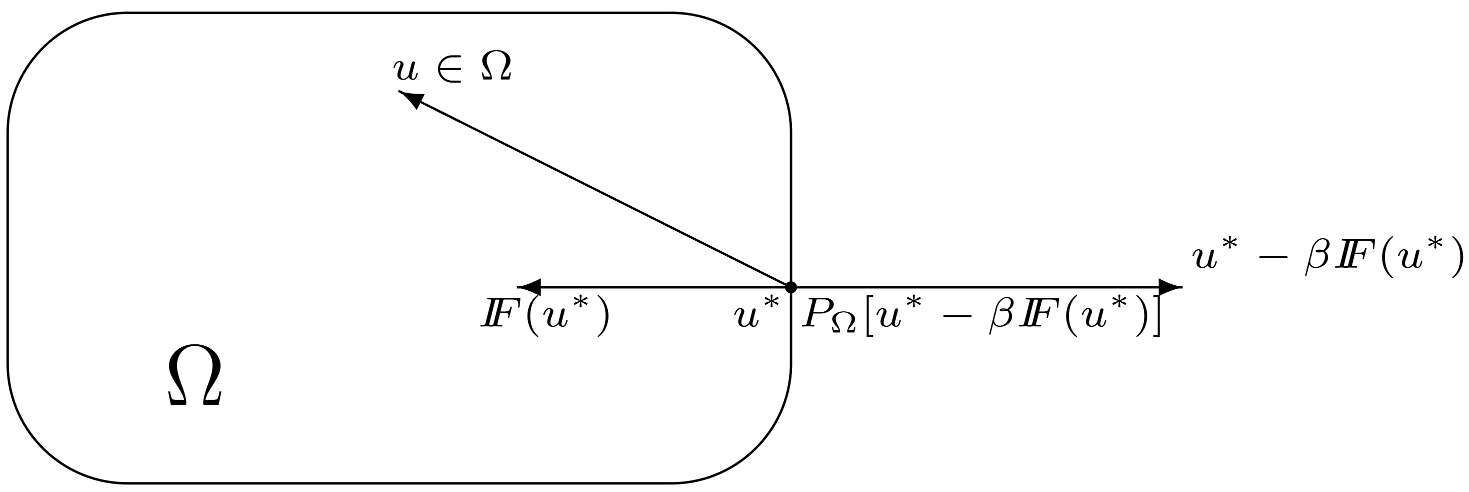
$$\|P_\Omega(v) - u\|^2 \le \|v - u\|^2 - \|v - P_\Omega(v)\|^2, \ \forall v \in \Re^n, u \in \Omega. \tag{2.5}$$

我们把这几个性质留给读者自己去证明.

2.2 变分不等式等价的投影方程

设变分不等式 (1.1) 的解集 Ω^* 非空. 我们用 u^* 表示一个确定的解点. 对任意的 $\beta > 0$, 变分不等式等价于投影方程

$$u \in \Omega^* \quad \Leftrightarrow \quad u = P_\Omega[u - \beta I\!\!F(u)].$$



u^* 是 VI(Ω, F) 的解等价于 $u^* = P_\Omega[u^* - \beta I\!\!F(u^*)]$ 的几何解释

换言之, 求解变分不等式可以归结为求

$$e(u,\beta) := u - P_\Omega[u - \beta I\!\!F(u)] \tag{2.6}$$

的一个零点 u^*, 后面我们会给出证明. 因此, 对确定的 $\beta > 0$, $\|e(u,\beta)\|$ 可以看作一种误差的度量函数. 为了方便, 我们往往把 $e(u,1)$ 记成 $e(u)$.

定理 2 设 $\beta > 0$. u^* 是 *VI*$(\Omega, I\!\!F)$ 的解当且仅当 $e(u^*,\beta) = 0$.

证明. 先证必要性. 若 u^* 是 VI(Ω, F) 的解, 则 $u^* \in \Omega$. 由于 $\Omega \subset R^n$ 是闭凸集, 利

用 (2.1) 得到

$$\big(v - P_\Omega(v)\big)^T \big(u^* - P_\Omega(v)\big) \le 0, \qquad \forall v \in R^n.$$

上式中取 $v := u^* - \beta I\!\!F(u^*)$, 则有 $\big(e(u^*,\beta) - \beta I\!\!F(u^*)\big)^T e(u^*,\beta) \le 0$, 即

$$\|e(u^*,\beta)\|^2 \le \beta e(u^*,\beta)^T I\!\!F(u^*). \tag{2.7}$$

另一方面, 由于$P_\Omega[u^* - \beta I\!\!F(u^*)] \in \Omega$, 而且 u^* 是变分不等式的解, 根据 (1.1) 可以得到

$$\{P_\Omega[u^* - \beta I\!\!F(u^*)] - u^*\}^T I\!\!F(u^*) \ge 0,$$

即

$$e(u^*,\beta)^T I\!\!F(u^*) \le 0. \tag{2.8}$$

由不等式 (2.7) 和 (2.8) 可得$e(u^*,\beta) = 0$.

再证充分性. 取 $v = u^* - \beta I\!\!F(u^*)$, 利用 (2.1) 和 $e(u^*,\beta)$ 的表达式, 有

$$\{e(u^*,\beta) - \beta I\!\!F(u^*)\}^T\{u - P_\Omega[u^* - \beta I\!\!F(u^*)]\} \le 0, \quad \forall u \in \Omega. \tag{2.9}$$

根据条件 $e(u^*,\beta) = 0$, 有 $u^* = P_\Omega(\cdot) \in \Omega$ 和 $P_\Omega[u^* - \beta I\!\!F(u^*)] = u^*$. 代入不等式 (2.9), 可以得到

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

即 u^* 是VI(Ω, F) 的解. 定理得证. □

下面的定理说明 $\|e(u,\beta)\|$ 是 β 的不减函数, 而 $\{\|e(u,\beta)\|/\beta\}$ 是 β 的不增函数. 这个简单证明只用到一元二次不等式的初等知识和工具不等式 (2.1), 它初见于 [19].

定理 3 对所有的 $u \in \Re^n$ 和 $\tilde\beta \ge \beta > 0$, 我们有

$$\|e(u,\tilde\beta)\| \ge \|e(u,\beta)\| \tag{2.10}$$

和

$$\frac{\|e(u,\tilde\beta)\|}{\tilde\beta} \le \frac{\|e(u,\beta)\|}{\beta}. \tag{2.11}$$

证明. 设 $t = \|e(x,\tilde\beta)\|/\|e(x,\beta)\|$, 定理的结论就相当于要证明

$$1 \le t \le \tfrac{\tilde\beta}{\beta}.$$

注意到它的等价表达式是 t 的一元二次不等式

$$(t-1)(t - \tfrac{\tilde\beta}{\beta}) \le 0 \tag{2.12}$$

的解. 首先, 由工具不等式 (2.1), 我们有

$$(v - P_\Omega(v))^T (P_\Omega(v) - w) \ge 0, \quad \forall w \in \Omega. \tag{2.13}$$

在 (2.13) 中令 $w := P_\Omega[u - \tilde\beta I\!\!F(u)]$ 和 $v := u - \beta I\!\!F(u)$, 利用 $e(u,\beta)$ 的定义和

$$P_\Omega[u - \beta I\!\!F(u)] - P_\Omega[u - \tilde\beta I\!\!F(u)] = e(u,\tilde\beta) - e(u,\beta),$$

我们得到

$$\{e(u,\beta)-\beta I\!F(u)\}^T\{e(u,\tilde{\beta})-e(u,\beta)\}\ge 0. \tag{2.14}$$

用相应的方法(将上式中的 β 和 $\tilde{\beta}$ 互换位置), 可得

$$\{e(u,\tilde{\beta})-\tilde{\beta} I\!F(u)\}^T\{e(u,\beta)-e(u,\tilde{\beta})\}\ge 0. \tag{2.15}$$

分别将不等式 (2.14) 和 (2.15) 乘上 $\tilde{\beta}$ 和 β, 然后再将它们相加, 我们得到

$$\{\tilde{\beta}e(u,\beta)-\beta e(u,\tilde{\beta})\}^T\{e(u,\tilde{\beta})-e(u,\beta)\}\ge 0 \tag{2.16}$$

并有

$$\beta\|e(x,\tilde{\beta})\|^2-(\beta+\tilde{\beta})e(x,\beta)^Te(x,\tilde{\beta})+\tilde{\beta}\|e(x,\beta)\|^2\le 0.$$

对上式采用 Cauchy-Schwarz 不等式, 就有

$$\beta\|e(x,\tilde{\beta})\|^2-(\beta+\tilde{\beta})\|e(x,\beta)\|\cdot\|e(x,\tilde{\beta})\|+\tilde{\beta}\|e(x,\beta)\|^2\le 0. \tag{2.17}$$

将 (2.17) 除 $\beta\|e(x,\beta)\|^2$, 并利用 t 的定义便得

$$t^2-\Big(1+\frac{\tilde{\beta}}{\beta}\Big)t+\frac{\tilde{\beta}}{\beta}\le 0.$$

因此不等式 (2.12) 成立, 定理得证. □

定理 3 说明, 若以 $\|e(u,\beta)\|$ 作为停机的误差度量, 常数 $\beta>0$ 不宜过大, 也不宜过小. 一般要结合问题的物理意义考虑.

2.3 投影收缩算法的三个基本不等式

设 u^* 是变分不等式 VI$(\Omega,I\!F)$ 的解. 记

$$\tilde{u}=P_\Omega[u-\beta I\!F(u)] \tag{2.18}$$

根据变分不等式的定义 (1.1), 有第一个基本不等式

$$\textbf{(FI1)}\qquad (\tilde{u}-u^*)^T\beta I\!F(u^*)\ge 0,\quad \forall u^*\in\Omega^*. \tag{2.19}$$

由于 $u^*\in\Omega$, 由 (2.18) 给出的 $\tilde{u}$ 是 $[u-\beta I\!F(u)]$ 在 Ω 上的投影. 在投影的基本性质不等式 (2.1) 中, 分别设 $v=u-\beta I\!F(u)$ 和任意的属于 Ω 的 $u=u^*$, 则有

$$\textbf{(FI2)}\qquad (\tilde{u}-u^*)^T\big\{[u-\beta I\!F(u)]-\tilde{u}\big\}\ge 0,\quad \forall u^*\in\Omega^*. \tag{2.20}$$

此外, 根据单调算子的性质, 有

$$\textbf{(FI3)}\qquad (\tilde{u}-u^*)^T\big\{\beta I\!F(\tilde{u})-\beta I\!F(u^*)\big\}\ge 0,\quad \forall u^*\in\Omega^*. \tag{2.21}$$

我们把 (2.19)-(2.21) 称为投影收缩算法中的三个基本不等式 [5]. 将这里的 (2.19), (2.20) 和 (2.21) 加在一起, 就得到

$$(\tilde{u}-u^*)^Td(u,\tilde{u})\ge 0,\quad \forall u^*\in\Omega^*. \tag{2.22}$$

其中

$$d(u,\tilde{u})=(u-\tilde{u})-\beta[I\!F(u)-I\!F(\tilde{u})]. \tag{2.23}$$

如果只将 (2.19) 和 (2.21) 加在一起, 得到的是

$$(\tilde{u} - u^*)^T \{\beta F(\tilde{u})\} \ge 0, \quad \forall u^* \in \Omega^*. \tag{2.24}$$

后面我们将会定义, 这里由不同组合得到的 $d(u, \tilde{u})$ 和 $\beta F(\tilde{u})$, 称为一对孪生方向.

3 投影收缩算法和分裂收缩算法中的预测

和求解混合变分不等式的分裂收缩算法一样, 求解单调经典变分不等式 (1.1) 的投影收缩算法 [5, 9, 14, 6] 是一个采用投影为预测的预测-校正方法.

3.1 求解经典变分不等式的投影收缩算法中的预测

求解经典变分不等式 (1.1) 的投影收缩算法的第 k-步迭代从给定的 u^k 开始, 通过欧氏模意义下的投影得到预测点 $\tilde{u}^k$, 具体公式是

$$\tilde{u}^k = P_\Omega[u^k - \beta_k F(u^k)], \tag{3.1}$$

其中 β_k 的选择要求满足

$$\beta_k \|F(u^k) - F(\tilde{u}^k)\| \le \nu \|u^k - \tilde{u}^k\|, \quad \nu \in (0,1). \tag{3.2}$$

定义 (合格的投影预测) 求解非线性变分不等式 (1.1) 的预测-校正方法中, 对给定的常数 $\nu \in (0,1)$, 若通过投影 (3.1) 得到的预测点 $\tilde{u}^k$ 满足条件 (3.2), 则称其为一个合格的投影预测.

在 F 为 Lipschitz 连续的条件下,(3.2) 是能够实现的. 由于 $\tilde{u}^k$ 可以表示成

$$\tilde{u}^k = \arg\min\{\tfrac{1}{2}\|u - [u^k - \beta_k F(u^k)]\|^2 \,|\, u \in \Omega\},$$

根据最优性定理, 由 (3.1) 得到的 $\tilde{u}^k$ 满足

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

进而得到

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

两边都加上 $(u - \tilde{u}^k)^T\{-\beta_k[F(u^k) - F(\tilde{u}^k)]\}$, 就有

$$\tilde{u}^k \in \Omega, \quad (u - \tilde{u}^k)^T \beta_k F(\tilde{u}^k) \ge (u - \tilde{u}^k)^T d(u^k, \tilde{u}^k), \quad \forall u \in \Omega, \tag{3.3}$$

其中的 $d(u^k, \tilde{u}^k)$ 是由 (2.23) 定义的.

不等式 (3.3) 对应于求解混合变分不等式的统一框架中的预测, 这里的 $d(u^k, \tilde{u}^k)$ 相当于其中的 $Q(v^k - \tilde{v}^k)$. 差别在于 $d(u^k, \tilde{u}^k)$ 并不是 $(u^k - \tilde{u}^k)$ 的线性函数.

将 (3.3) 中的 $u \in \Omega$ 选成任意确定的 $u^* \in \Omega^*$, 得到

$$(\tilde{u}^k - u^*)^T d(u^k, \tilde{u}^k) \geq (\tilde{u}^k - u^*)^T \beta_k F(\tilde{u}^k). \tag{3.4}$$

由 F 的单调性和 u^* 的最优性, 上式右端

$$(\tilde{u}^k - u^*)^T F(\tilde{u}^k) \geq (\tilde{u}^k - u^*)^T F(u^*) \geq 0,$$

不等式 (3.4) 的左端非负, 随后改写成

$$\{(u^k - u^*) - (u^k - \tilde{u}^k)\}^T d(u^k, \tilde{u}^k) \geq 0,$$

得到

$$(u^k - u^*)^T d(u^k, \tilde{u}^k) \geq (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \tag{3.5}$$

根据 $d(u^k, \tilde{u}^k)$ 的表达式 (2.23) 和假设 (3.2), 利用 Cauchy-Schwarz 不等式, 有

$$(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) \geq (1-\nu)\|u^k - \tilde{u}^k\|^2. \tag{3.6}$$

由上面两个不等式得知

$$(u^k - u^*)^T d(u^k, \tilde{u}^k) \geq (1-\nu)\|u^k - \tilde{u}^k\|^2. \tag{3.7}$$

定义 (上升方向) 求解变分不等式 (1.1) 的方法中, 假如存在一个常数 $\delta > 0$, 向量 $d(u^k, \tilde{u}^k)$ 满足关系式

$$(u^k - u^*)^T d(u^k, \tilde{u}^k) \geq \delta\|u^k - \tilde{u}^k\|^2, \quad \forall\, u^* \in \Omega^*, \tag{3.8}$$

则称其为距离函数 $\|u - u^*\|^2$ 在 u^k 处的上升方向.

因此, 由 (2.23) 给出的 $d(u^k, \tilde{u}^k)$ 是未知距离函数 $\|u - u^*\|^2$ 在 u^k 处欧氏模下的一个上升方向. 虽然我们并不知道解点在哪里, 但是沿着方向 $-d(u^k, \tilde{u}^k)$, 选取适当步长, 可以找到欧氏模下比 u^k 更靠近解集的 u^{k+1}.

3.2 求解混合变分不等式的分裂收缩算法中的预测

由线性约束的凸优化问题得到的混合变分不等式

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{3.9}$$

统一框架中定义的预测为

$$\tilde{w}^k \in \Omega, \ \ \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \ \forall\, w \in \Omega, \tag{3.10}$$

其中矩阵 $Q^T + Q$ 是本质上正定的. 将 (3.10) 中任意的 $w \in \Omega$ 选成 $w^* \in \Omega^*$, 我们有

$$(\tilde{v}^k - v^*)^T Q(v^k - \tilde{v}^k) \geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k). \tag{3.11}$$

由

$$(\tilde{w}^k - w^*)^T F(\tilde{w}^k) = (\tilde{w}^k - w^*)^T F(w^*)$$

和 w^* 的最优性, (3.11) 的左端非负. 随后由它得到

$$(v^k - v^*)^T Q(v^k - \tilde{v}^k) \geq (v^k - \tilde{v}^k)^T Q(v^k - \tilde{v}^k).$$

对正定矩阵 H 和 $H^{-1}Q = M$ 上式可以表示成

$$\langle H(v^k - v^*), M(v^k - \tilde{v}^k)\rangle \geq (v^k - \tilde{v}^k)^T Q(v^k - \tilde{v}^k). \tag{3.12}$$

不等式 (3.12) 告诉我们, 在条件 $Q^T + Q$ 正定的情况下, 向量 $M(v^k - \tilde{v}^k)$ 是 H-模下未知距离函数 $\|v - v^*\|_H^2$ 在 v^k 处的一个上升方向.

这两类方法的预测中, (3.3) 和 (3.4) 分别跟 (3.10) 和 (3.11) 相对应. 对应于经典变分不等式的投影收缩算法和混合变分不等式的分裂收缩算法, 不等式 (3.5) 和 (3.12) 分别提供了相应的上升方向. 它们的右端严格大于零分别由假设 (3.2) 和 $Q^T + Q \succ 0$ 得到保证.

4 投影收缩算法和分裂收缩算法中的校正

校正是利用距离函数的下降方向(上升方向的反方向), 使得新的迭代点在某种确定的模的意义下离解集比原来的点更近一些.

4.1 两类方法中采用固定步长的校正

1. 变分不等式投影收缩算法固定步长的校正 在投影收缩算法中, 我们一般考虑欧氏模下的收缩. 用关系式 (3.5), 校正通过

$$u^{k+1} = u^k - \alpha d(u^k, \tilde{u}^k) \tag{4.1}$$

产生新的迭代点, 其中 $d(u^k, \tilde{u}^k)$ 是由 (2.23) 给出的. 下面我们讨论如何选取步长 α.

投影收缩算法中, 我们将条件 (3.2) 满足时, 由单位步长校正产生新迭代点

$$u^{k+1} = u^k - d(u^k, \tilde{u}^k) \tag{4.2}$$

的方法, 称为初等方法 (Primary Method). 利用 (3.5), 由简单计算可得

$$\begin{aligned}\|u^{k+1} - u^*\|^2 &= \|(u^k - u^*) - d(u^k, \tilde{u}^k)\|^2 \\ &= \|u^k - u^*\|^2 - 2(u^k - u^*)^T d(u^k, \tilde{u}^k) + \|d(u^k, \tilde{u}^k)\|^2 \\ &\leq \|u^k - u^*\|^2 - \big[2(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \|d(u^k, \tilde{u}^k)\|^2\big]. \end{aligned} \tag{4.3}$$

利用 (2.23) 和 (3.2), 可以得到

$$\begin{aligned} & 2(u^k-\tilde{u}^k)^T d(u^k,\tilde{u}^k) - \|d(u^k,\tilde{u}^k)\|^2 \\ &\quad= \quad d(u^k,\tilde{u}^k)^T\{2(u^k-\tilde{u}^k) - d(u^k,\tilde{u}^k)\} \\ &\quad= \quad \big\{(u^k-\tilde{u}^k) - \beta_k[F(u^k)-F(\tilde{u}^k)]\big\}^T\big\{(u^k-\tilde{u}^k)+\beta_k[F(u^k)-F(\tilde{u}^k)]\big\} \\ &\quad= \quad \|u^k-\tilde{u}^k\|^2 - \beta_k^2\|F(u^k)-F(\tilde{u}^k)\|^2 \\ &\quad\geq \quad (1-\nu^2)\|u^k-\tilde{u}^k\|^2. \end{aligned} \tag{4.4}$$

代入 (4.3), 说明由 (4.2) 产生的序列 $\{u^k\}$ 满足

$$\|u^{k+1}-u^*\|^2 \leq \|u^k-u^*\|^2 - (1-\nu^2)\|u^k-\tilde{u}^k\|^2. \tag{4.5}$$

2. 凸优化分裂收缩算法固定步长的校正 在凸优化的分裂收缩算法中, 我们一般考虑 H-模下的收缩. 用关系式 (3.12), 校正通过

$$v^{k+1} = v^k - \alpha M(v^k-\tilde{v}^k) \tag{4.6}$$

产生新的迭代点, 其中 $M=H^{-1}Q$. 下面我们讨论如何选取步长 α. 由简单计算可得

$$\begin{aligned} \|v^{k+1}-v^*\|_H^2 &= \|(v^k-v^*) - \alpha M(v^k-\tilde{v}^k)\|_H^2 \\ &= \|v^k-v^*\|_H^2 - 2\alpha(v^k-v^*)^T HM(v^k-\tilde{v}^k) + \alpha^2\|M(v^k-\tilde{v}^k)\|_H^2 \\ &\leq \|v^k-v^*\|_H^2 - \alpha\big((v^k-\tilde{v}^k)^T[Q^T+Q-\alpha M^THM](v^k-\tilde{v}^k)\big). \end{aligned}$$

可以得到

$$\|v^{k+1}-v^*\|_H^2 \leq \|v^k-v^*\|_H^2 - \alpha\|v^k-\tilde{v}^k\|_G^2, \tag{4.7}$$

其中

$$G = Q^T + Q - \alpha M^T HM.$$

根据统一框架收敛性的要求, G 是正定矩阵.

4.2 两类方法中采计算步长的校正

1. 变分不等式投影收缩算法计算步长的校正 我们将 (4.1) 中的 u^{k+1} 记为 $u^{k+1}(\alpha)$, 表示新的迭代点依赖于步长 α. 考察与 α 相关的距离平方缩短量,

$$\vartheta_k(\alpha) = \|u^k-u^*\|^2 - \|u^{k+1}(\alpha)-u^*\|^2. \tag{4.8}$$

根据定义

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

对任意给定的确定解点 u^*, 上式表明 $\vartheta_k(\alpha)$ 是 α 的一个二次函数. 只是 u^* 是未知的, 我们无法直接求 $\vartheta_k(\alpha)$ 的极大. 利用 (3.5), 对任意的 $\alpha>0$, 有

$$\vartheta_k(\alpha) \geq q_k(\alpha), \tag{4.9}$$

其中

$$q_k(\alpha) = 2\alpha(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \alpha^2 \|d(u^k, \tilde{u}^k)\|^2. \qquad (4.10)$$

既然二次函数 $q_k(\alpha)$ 是 $\vartheta_k(\alpha)$ 的一个下界函数. 使 $q_k(\alpha)$ 达到极大的 α_k^* 是

$$\alpha_k^* = \text{argmax}\{q_k(\alpha)\} = \frac{(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2}. \qquad (4.11)$$

注意到这里的 α_k^* 是由 (3.5) 确定的, 分子是 (3.5) 的右端, 分母是 (4.1) 中 $d(u^k, \tilde{u}^k)$ 的欧氏长度的平方.

在实际计算中, 我们一般取一个松弛因子 $\gamma \in [1.2, 1.8]$, 令

$$u^{k+1} = u^k - \gamma \alpha_k^* d(u^k, \tilde{u}^k), \qquad (4.12)$$

根据 (4.8) 和 (4.9), 由 (4.12) 产生的 u^{k+1} 满足

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - q_k(\gamma\alpha_k^*). \qquad (4.13)$$

其中

$$\begin{aligned} q_k(\gamma\alpha_k^*) &= 2\gamma\alpha_k^*(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \gamma^2(\alpha_k^*)^2 \|d(u^k, \tilde{u}^k)\|^2 \\ &= \gamma(2-\gamma)\alpha_k^*(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \end{aligned} \qquad (4.14)$$

此外, 从 (4.4), 我们已经有 $2(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \|d(u^k, \tilde{u}^k)\|^2 > 0$, 因而根

据 (4.11) 得到 $\alpha_k^* > \frac{1}{2}$ (见(4.4)). 结合 (3.6),

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \tfrac{1}{2}\gamma(2-\gamma)(1-\nu)\|u^k - \tilde{u}^k\|^2. \qquad (4.15)$$

不等式 (4.5) 和 (4.15) 说明, 用固定步长和计算步长的算法产生的序列 $\{u^k\}$ 都是收缩和有界的. 利用这些关键不等式, 容易证明收敛定理.

虽然投影收缩算法的这些结论都在假设 (3.2) 满足时才成立, 但是当 $I\!\!F$ 是 Lipschitz 连续, 采用 Armijo 法则选取适当的 β_k 实施的预测 (3.1), 是能够使条件 (3.2) 满足的. 采用下面程序的方法我们称其为投影收缩算法-I.

求解经典单调变分不等式的投影收缩算法-I

给定 $\beta_0 = 1, \mu = 0.4, \nu = 0.9, u^0 \in \Omega.$

For $k = 0, 1, \ldots,$ 假如停机准则尚未满足, **do**

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

$$r_k := \frac{\beta_k \| F(u^k) - F(\tilde{u}^k)\|}{\|u^k - \tilde{u}^k\|},$$

while $r_k > \nu, \quad \beta_k := \frac{2}{3}\beta_k * \min\{1, \frac{1}{r_k}\},$

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

$$r_k := \frac{\beta_k \| F(u^k) - F(\tilde{u}^k)\|}{\|u^k - \tilde{u}^k\|},$$

end(while)

2). $d(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k) - \beta_k[F(u^k) - F(\tilde{u}^k)],$

$$\alpha_k^* = \frac{(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2},$$

$$u^{k+1} = u^k - \gamma \alpha_k^* d(u^k, \tilde{u}^k),$$

3). **If** $r_k \le \mu$ then $\beta_k := \beta_k * 1.5,$ **end(if)**

4). 令 $\beta_{k+1} = \beta_k$ 和 $k := k + 1,$ 开始新的一次迭代.

- 在做投影预测 (3.1) 的时候, 要求满足条件 (3.2). 这里取的 $\nu = 0.9$, 是经验值. 在投

影收缩算法-I 的 1) 中, 当 $r_k > \nu$ 时, 用

$$\beta_k := \tfrac{2}{3}\beta_k * \min\{1, \tfrac{1}{r_k}\}$$

对 β_k 做调正. 这里的两种不同情形是: 当 $r_k \in (\nu, 1]$ 之间, 就取 $\beta_k := \frac{2}{3}\beta_k$; 当 $r_k > 1$ 时, 相当于把原来的 β_k 缩小 r_k 倍再乘上 $(2/3)$. 调正参数 β_k 以后, 重做一次投影预测, 条件 (3.2) 一般能够得到满足. 这里的 $(2/3)$ 也是经验值, 实际计算中, 也可以改成 $(3/4)$.

- 在条件 (3.2) 满足的前提下, 我们同时希望

$$\frac{\beta_k \| F(u^k) - F(\tilde{u}^k)\|}{\|u^k - \tilde{u}^k\|}$$

不要太小. 根据我们计算的一些例子, 程序中的那句

If $r_k \le \mu \quad \text{then} \quad \beta_k := \beta_k * 1.5,$ **end(if)**

是不可缺少的. 就像信赖域方法 [16] 中的信赖域半径, 在迭代计算过程发现过小也需要增大, 其中的 1.5 也是经验值.

2. 凸优化分裂收缩算法计算步长的校正 分裂收缩算法的计算步长的校正方法的根据是 (3.12). 利用 $HM = Q$ 并考虑距离函数 $\|v - v^*\|_H^2$ 的下降, 新的迭代点 v^{k+1} 由

$$v^{k+1} = v^k - \gamma \alpha_k^* M(v^k - \tilde{v}^k), \tag{4.16a}$$

产生, 其中

$$\alpha_k^* = \frac{(v^k - \tilde{v}^k)^T Q (v^k - \tilde{v}^k)}{\|M(v^k - \tilde{v}^k)\|_H^2}. \tag{4.16b}$$

分子是 (3.12) 的右端, 分母 $\|M(v^k - \tilde{v}^k)\|_H^2$ 可以通过

$$\|M(v^k - \tilde{v}^k)\|_H^2 = (v^k - \tilde{v}^k)^T M^T H M (v^k - \tilde{v}^k) = (v^k - \tilde{v}^k)^T M^T Q (v^k - \tilde{v}^k)$$

得到. 单调收缩关系式满足

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \frac{\gamma(2-\gamma)}{2}\alpha_k^* \|v^k - \tilde{v}^k\|_{(Q^T+Q)}^2, \quad \forall\, v^* \in \mathcal{V}^*.$$

凸优化分裂收缩算法的统一框架源自经典单调变分不等式的投影收缩算法. 不同的是：投影收缩算法通过投影得到预测点并提供欧氏模下的下降方向, 分裂收缩算法则是通过求解一些子问题得到 H-模下的下降方向.

5 投影收缩算法中的孪生方向和姊妹方法

我们在经典单调变分不等式求解方面发表的计算步长的投影收缩算法 [5, 9], 被工程力学界的一些学者用来解决了一些长期困扰他们的岩土工程问题 [17, 18]. 与 §4.2 中介绍的投影收缩算法-I 相对应的投影收缩算法-II, 它们分别取孪生方向之一作为寻查方向

而取相同的步长, 得到一对姊妹方法. 根据我们的数值试验 [9] 和工程界的计算实践 [17, 18], 算法-II 要比算法-I 效率高一些. 这一节, 我们专门讲一下基于同一预测的一对孪生方向和姊妹方法.

定义 (孪生方向) 求解经典变分不等式的方法中, 假如 $d(u^k, \tilde{u}^k)$ 是一个上升方向, 则称分处不等式 (3.3) 两边的

$$\beta_k F(\tilde{u}^k) \quad \text{和} \quad d(u^k, \tilde{u}^k), \tag{5.1}$$

为一对孪生方向.

我们也可以从得到关系式 (2.22) 和 (2.24) 的过程中知道, 这一对孪生方向是由基本不等式的不同组合生成的.

定义 (姊妹方法) 求解经典单调变分不等式的方法中, 由同一个预测点 $\tilde{u}^k$ 提供了一对 (5.1) 中的孪生方向 $d(u^k, \tilde{u}^k)$ 和 $\beta_k F(\tilde{u}^k)$. 用相同的步长 α, 分别由

$$u^{k+1}(\alpha) = u^k - \alpha H^{-1} d(u^k, \tilde{u}^k) \tag{5.2}$$

和

$$u^{k+1}(\alpha) = \arg\min\{\|u - [u^k - \alpha\beta_k H^{-1} F(\tilde{u}^k)]\|_H^2 \,|\, u \in \Omega\} \tag{5.3}$$

给出新的收缩迭代点的方法称为一对 H-模下收缩的姊妹方法.

在 H 为单位阵的时候, (5.3) 的右端是欧氏模下向量 $[u^k - \alpha\beta_k F(\tilde{u}^k)]$ 到 Ω 上的投影.

我们先讨论欧氏模下依赖步长 α 的姊妹方法, 分别用带下标的

$$u_I^{k+1}(\alpha) = u^k - \alpha d(u^k, \tilde{u}^k) \tag{5.4}$$

和

$$u_{II}^{k+1}(\alpha) = P_\Omega[u^k - \alpha\beta_k I\!F(\tilde{u}^k)] \tag{5.5}$$

表示不同方法依赖步长 α 的新的迭代点. 对任意给定的 $u^* \in \Omega^*$, 我们用

$$\vartheta_k(\alpha) = \|u^k - u^*\|^2 - \|u_I^{k+1}(\alpha) - u^*\|^2 \tag{5.6}$$

和

$$\zeta_k(\alpha) = \|u^k - u^*\|^2 - \|u_{II}^{k+1}(\alpha) - u^*\|^2 \tag{5.7}$$

看成是本次迭代的进步量, 它是步长 α 的函数. 根据 §4.2 的分析 (见(4.8) - (4.10)),

$$\vartheta_k(\alpha) \ge q_k(\alpha) = 2\alpha(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \alpha^2\|d(u^k, \tilde{u}^k)\|^2. \tag{5.8}$$

下面的定理说明, 对同样的 α, $\zeta_k(\alpha)$ 的下界不小于 $\vartheta_k(\alpha)$ 的下界.

定理 4 设 $u_{II}^{k+1}(\alpha)$ 由 (5.5) 生成. 对任意的 $\alpha > 0$, 对由 (5.7) 定义的 $\zeta_k(\alpha)$ 有

$$\zeta_k(\alpha) \ge \|u_I^{k+1}(\alpha) - u_{II}^{k+1}(\alpha)\|^2 + q_k(\alpha), \tag{5.9}$$

其中 $q_k(\alpha)$ 由 (4.10) 给出.

证明. 首先, 因为 $u_{II}^{k+1}(\alpha) = P_\Omega[u^k - \alpha\beta_k I\!F(\tilde{u}^k)]$ 和 $u^* \in \Omega$, 根据投影的性

质 (见 (2.5)), 我们有

$$\begin{aligned}\|u_{II}^{k+1}(\alpha) - u^*\|^2 \le\ & \|u^k - \alpha\beta_k I\!F(\tilde{u}^k) - u^*\|^2 \\ & - \|u^k - \alpha\beta_k I\!F(\tilde{u}^k) - u_{II}^{k+1}(\alpha)\|^2, \quad \forall u^* \in \Omega^*\end{aligned} \tag{5.10}$$

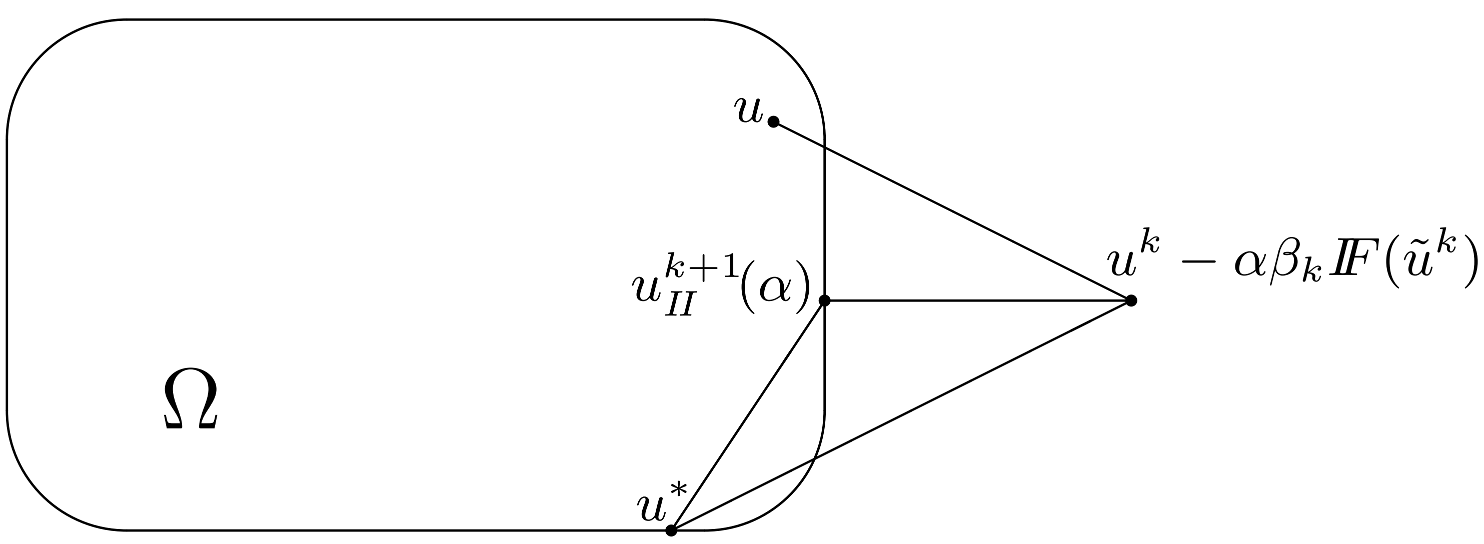


图6.1. 不等式 (5.10) 的几何解释

因此, 利用 $\zeta_k(\alpha)$ 的定义(见 (5.7)), 我们有

$$\begin{aligned}\zeta_k(\alpha) &\geq \|u^k-u^*\|^2-\|(u^k-u^*)-\alpha\beta_k\mathbb{F}(\tilde{u}^k)\|^2\\ &\quad+\|(u^k-u_{II}^{k+1}(\alpha))-\alpha\beta_k\mathbb{F}(\tilde{u}^k)\|^2\\ &= 2\alpha(u^k-u^*)^T\beta_k\mathbb{F}(\tilde{u}^k)+2\alpha(u_{II}^{k+1}(\alpha)-u^k)^T\beta_k\mathbb{F}(\tilde{u}^k)\\ &\quad+\|u^k-u_{II}^{k+1}(\alpha)\|^2\\ &= \|u^k-u_{II}^{k+1}(\alpha)\|^2+2\alpha(u_{II}^{k+1}(\alpha)-u^*)^T\beta_k\mathbb{F}(\tilde{u}^k). \qquad (5.11)\end{aligned}$$

将 (5.11) 中右端的最后一项 $(u_{II}^{k+1}(\alpha)-u^*)^T\beta_k\mathbb{F}(\tilde{u}^k)$ 分解成

$$(u_{II}^{k+1}(\alpha)-u^*)^T\beta_k\mathbb{F}(\tilde{u}^k)=(u_{II}^{k+1}(\alpha)-\tilde{u}^k)^T\beta_k\mathbb{F}(\tilde{u}^k)+(\tilde{u}^k-u^*)^T\beta_k\mathbb{F}(\tilde{u}^k), \qquad (5.12)$$

利用 $(\tilde{u}^k-u^*)^T\beta_k\mathbb{F}(\tilde{u}^k)\geq(\tilde{u}^k-u^*)^T\beta_k\mathbb{F}(u^*)\geq 0$, (5.12) 右端的最后一部分非负. 代入 (5.11) 的右端, 进一步得到

$$\zeta_k(\alpha)\geq\|u^k-u_{II}^{k+1}(\alpha)\|^2+2\alpha(u_{II}^{k+1}(\alpha)-\tilde{u}^k)^T\beta_k\mathbb{F}(\tilde{u}^k). \qquad (5.13)$$

因为 $u_{II}^{k+1}(\alpha)\in\Omega$, 用它替代 (3.3) 中的任意 $u\in\Omega$, 得到

$$(u_{II}^{k+1}(\alpha)-\tilde{u}^k)^T\beta_k\mathbb{F}(\tilde{u}^k)\geq(u_{II}^{k+1}(\alpha)-\tilde{u}^k)^Td(u^k,\tilde{u}^k). \qquad (5.14)$$

将它们代入 (5.13) 的右端, 就有

$$\zeta_k(\alpha)\geq\|u^k-u_{II}^{k+1}(\alpha)\|^2+2\alpha(u_{II}^{k+1}(\alpha)-\tilde{u}^k)^Td(u^k,\tilde{u}^k). \qquad (5.15)$$

对上式右端, 利用 $u_I^{k+1}(\alpha)$ 和 $q_k(\alpha)$ 的表达式 (见(5.4)和(4.10)), 进一步化成

$$\begin{aligned}\zeta_k(\alpha) &\geq \|u^k-u_{II}^{k+1}(\alpha)\|^2+2\alpha(u_{II}^{k+1}(\alpha)-\tilde{u}^k)^Td(u^k,\tilde{u}^k)\\ &= \|u^k-u_{II}^{k+1}(\alpha)\|^2+2\alpha(u_{II}^{k+1}(\alpha)-u^k)^Td(u^k,\tilde{u}^k)\\ &\quad+2\alpha(u^k-\tilde{u}^k)^Td(u^k,\tilde{u}^k)\\ &= \|(u^k-u_{II}^{k+1}(\alpha))-\alpha d(u^k,\tilde{u}^k)\|^2-\alpha^2\|d(u^k,\tilde{u}^k)\|^2\\ &\quad+2\alpha(u^k-\tilde{u}^k)^Td(u^k,\tilde{u}^k)\\ &= \|u_I^{k+1}(\alpha)-u_{II}^{k+1}(\alpha)\|^2+q_k(\alpha).\end{aligned}$$

这样就完成了定理结论 (5.9) 的证明. □

定理 4 说明, $q_k(\alpha)$ 也是 $\zeta_k(\alpha)$ 的下界. 对同样的 α, $\zeta_k(\alpha)$ 优于 $\vartheta_k(\alpha)$, 除非 $u_I^{k+1}(\alpha)=u_{II}^{k+1}(\alpha)$. 在实际计算中, 这对姊妹方法分别采用校正公式

$$\text{(投影收缩算法-I)}\qquad u^{k+1}=u^k-\gamma\alpha_k^*d(u^k,\tilde{u}^k) \qquad (5.16)$$

和

$$\text{(投影收缩算法-II)}\qquad u^{k+1}=P_\Omega[u^k-\gamma\alpha_k^*\beta_k\mathbb{F}(\tilde{u}^k)] \qquad (5.17)$$

产生新的迭代点, 其中的 α_k^* 都由 (4.11) 给出. 由于 $\zeta_k(\gamma\alpha_k^*) \ge q_k(\gamma\alpha_k^*)$, 它们所产生的迭代序列 $\{u^k\}$ 都满足

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \gamma(2-\gamma)\alpha_k^*(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \tag{5.18}$$

事实上, 由于 $\alpha_k^* > \frac{1}{2}$, 结合关系式 (3.6), 从 (5.18) 得到

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \tfrac{1}{2}\gamma(2-\gamma)(1-\nu)\|u^k - \tilde{u}^k\|^2. \tag{5.19}$$

这就是 §4.2 中的 (4.15). 根据上面这个不等式, 利用简单的分析术语就可以证明投影收缩算法的收敛性.

采用校正公式 (5.16) 的好处是生成 u^{k+1} 不用再做投影. 实际问题中, 到 Ω 上的投影代价往往是不高的 (例如 Ω 常常是一个正卦限或者框形), 因此常采用校正公式 (5.17). 这方面的理由我们在论文 [14] 中有更详细的说明.

求解经典单调变分不等式的投影收缩算法-II

给定 $\beta_0 = 1, \nu \in (0,1), u^0 \in \Omega.$

For $k = 0, 1, \ldots$, 假如停机准则尚未满足, **do**

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

$$r_k := \frac{\beta_k\|F(u^k) - F(\tilde{u}^k)\|}{\|u^k - \tilde{u}^k\|},$$

while $r_k > \nu, \quad \beta_k := \frac{2}{3}\beta_k * \min\{1, \frac{1}{r_k}\},$

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

$$r_k := \frac{\beta_k\|F(u^k) - F(\tilde{u}^k)\|}{\|u^k - \tilde{u}^k\|},$$

end(while)

2). $d(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k) - \beta_k[F(u^k) - F(\tilde{u}^k)],$

$$\alpha_k^* = \frac{(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2},$$

$$u^{k+1} = P_\Omega[u^k - \gamma\alpha_k^*\beta_k F(\tilde{u}^k)],$$

3). **If** $r_k \le \mu$ then $\beta_k := \beta_k * 1.5,$ **end(if)**

4). 令 $\beta_{k+1} = \beta_k$ 和 $k := k + 1,$ 开始新的一次迭代.

从投影收缩算法-I 到投影收缩算法-II, 只是将

$$u^{k+1} = u^k - \gamma\alpha_k^* d(u^k, \tilde{u}^k) \qquad \text{改成了} \qquad u^{k+1} = P_\Omega[u^k - \gamma\alpha_k^*\beta_k F(\tilde{u}^k)].$$

论文 [17, 18] 中解决岩土工程问题用的就是这一节的方法. 根据我们的数值经验 [9] 和他们的计算实践, 算法-II 的效率比算法-I 的效率高一些.

在条件 (3.2) 满足的合格预测 (3.1) 的基础上, 校正也可以用

$$u^{k+1} = P_\Omega[u^k - \beta_k F(\tilde{u}^k)]$$

实现. 这时候, 预测-校正写在一起就是

$$\begin{cases} \tilde{u}^k = P_\Omega[u^k - \beta_k F(u^k)], \\ u^{k+1} = P_\Omega[u^k - \beta_k F(\tilde{u}^k)]. \end{cases} \tag{5.20}$$

这就是人们所说的 Korpelevich 外梯度法 [15].

凸优化的分裂收缩算法的预测, 同样提供了一对孪生方向. 如何借鉴投影收缩算法中的孪生方向和姊妹方法, 我们在 [8] 中开展了一些讨论.

6 投影收缩算法在求解可分离凸优化上的应用

作为投影收缩算法在求解可分离凸优化上的应用, 我们考虑更一般两块可分离凸优化问题

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b \ (\text{或} \geq b), \ x \in \mathcal{X}, \ y \in \mathcal{Y}\}. \tag{6.1}$$

假设问题有解并且 $\theta_1(x)$ 和 $\theta_2(y)$ 分别在包含凸闭集 $\mathcal{X}$ 和 $\mathcal{Y}$ 的一个开集上可微. 问题 (6.1) 的拉格朗日函数是定义在 $\mathcal{X} \times \mathcal{Y} \times \Lambda$ 上的

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

其中

$$\Lambda = \begin{cases} \Re^m, & \text{if } Ax + By = b, \\ \Re^m_+, & \text{if } Ax + By \geq b. \end{cases}$$

拉格朗日函数的鞍点 $(x^*, y^*, \lambda^*) \in \mathcal{X} \times \mathcal{Y} \times \Lambda$ 满足不等式

$$L(x^*, y^*, \lambda) \leq L(x^*, y^*, \lambda^*) \leq L(x, y, \lambda^*), \quad \forall (x, y, \lambda) \in \mathcal{X} \times \mathcal{Y} \times \Lambda.$$

这意味着

$$\begin{cases} x^* \in \mathcal{X}, & L(x, y^*, \lambda^*) \geq L(x^*, y^*, \lambda^*) & \forall x \in \mathcal{X}, \\ y^* \in \mathcal{Y}, & L(x^*, y, \lambda^*) \geq L(x^*, y^*, \lambda^*) & \forall y \in \mathcal{Y}, \\ \lambda^* \in \Lambda, & L(x^*, y^*, \lambda^*) \geq L(x^*, y^*, \lambda) & \forall \lambda \in \Lambda. \end{cases}$$

也就是说

$$\begin{cases} x^* & \in & \arg\min\{\theta_1(x)+\theta_2(y^*)-(\lambda^*)^T(Ax+By^*-b)\,|\,x\in\mathcal{X}\}, \\ y^* & \in & \arg\min\{\theta_1(x^*)+\theta_2(y)-(\lambda^*)^T(Ax^*+By-b)\,|\,y\in\mathcal{Y}\}, \\ \lambda^* & \in & \arg\max\,\{\theta_1(x^*)+\theta_2(y^*)-\lambda^T(Ax^*+By^*-b)|\lambda\in\Lambda\}. \end{cases}$$

当 $\theta_1(x)$, $\theta_2(y)$ 可微时, 根据最优性质定理, 我们有

$$\begin{cases} x^*\in\mathcal{X}, & (x-x^*)^T(\nabla\theta_1(x^*)-A^T\lambda^*)\geq 0, & \forall\,x\in\mathcal{X}, \\ y^*\in\mathcal{Y}, & (y-y^*)^T(\nabla\theta_2(y^*)-B^T\lambda^*)\geq 0, & \forall\,y\in\mathcal{Y}, \\ \lambda^*\in\Lambda, & (\lambda-\lambda^*)^T(Ax^*+By^*-b)\geq 0, & \forall\,\lambda\in\Lambda. \end{cases}$$

用记号

$$\nabla\theta_1(x)=f(x), \qquad \nabla\theta_2(y)=g(y),$$

相应的紧凑形式的单调变分不等式就是

$$w^*\in\Omega, \quad (w-w^*)^T F(w^*)\geq 0, \quad \forall\,w\in\Omega, \tag{6.2a}$$

其中

$$w=\begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad F(w)=\begin{pmatrix} f(x)-A^T\lambda \\ g(y)-B^T\lambda \\ Ax+By-b \end{pmatrix}, \qquad \Omega=\mathcal{X}\times\mathcal{Y}\times\Lambda. \tag{6.2b}$$

假如交替方向法中子问题求解比较困难, 而 $\theta_1(x)$ 和 $\theta_2(y)$ 的梯度随手可得并且 Lipschitz 连续, 可以考虑用投影收缩算法求解可微的线性约束凸优化问题等价的变分不等式 (6.2). 对给定的 $w^k=(x^k,y^k,\lambda^k)$, 通过

$$\begin{cases} \tilde{x}^k = P_{\mathcal{X}}\{x^k-\frac{1}{r}[f(x^k)-A^T\lambda^k]\}, & (6.3a) \\ \tilde{y}^k = P_{\mathcal{Y}}\{y^k-\frac{1}{s}[g(y^k)-B^T\lambda^k]\}, & (6.3b) \\ \tilde{\lambda}^k = P_{\Lambda}\{\lambda^k-\beta(Ax^k+By^k-b)\}, & (6.3c) \end{cases}$$

得到预测点 $\tilde{w}^k=(\tilde{x}^k,\tilde{y}^k,\tilde{\lambda}^k)$. 其中的 $r,s>0$ 是适当选取的正常数使得

$$\|f(x^k)-f(\tilde{x}^k)\|\leq\nu r\|x^k-\tilde{x}^k\|, \quad \text{和} \quad \|g(y^k)-g(\tilde{y}^k)\|\leq\nu s\|y^k-\tilde{y}^k\|. \tag{6.4}$$

当 $f(x)$ 和 $g(y)$ Lipschitz 连续时, 这是可以办到的.

下面对预测 (6.3) 进行分析. 通过投影 (6.3a) 得到的 $\tilde{x}^k$ 是极小化问题

$$\min\{\|x-[x^k-\tfrac{1}{r}[f(x^k)-A^T\lambda^k]]\|^2|x\in\mathcal{X}\}$$

的解, 根据最优性条件的定理, 有

$$\tilde{x}^k \in \mathcal{X} \quad (x-\tilde{x}^k)^T\{\tilde{x}^k - [x^k - \tfrac{1}{r}[f(x^k) - A^T\lambda^k]]\} \ge 0, \quad \forall x \in \mathcal{X}.$$

这可以写成

$$\tilde{x}^k \in \mathcal{X}, \quad (x-\tilde{x}^k)^T\{f(x^k) - A^T\lambda^k + r(\tilde{x}^k - x^k)\} \ge 0, \quad \forall x \in \mathcal{X}.$$

利用 (6.2), 我们有

$$\tilde{x}^k \in \mathcal{X}, \quad (x-\tilde{x}^k)^T \left\{ \begin{array}{l} \underline{f(\tilde{x}^k) - A^T\tilde{\lambda}^k} + A^T(\tilde{\lambda}^k - \lambda^k) \\ +\big[r(\tilde{x}^k - x^k) - \big(f(\tilde{x}^k) - f(x^k)\big)\big] \end{array} \right\} \ge 0, \quad \forall x \in \mathcal{X}. \tag{6.5a}$$

通过投影 (6.3b) 得到的 $\tilde{y}^k$, 有

$$\tilde{y}^k \in \mathcal{Y}, \quad (y-\tilde{y}^k)^T \left\{ \begin{array}{l} \underline{g(\tilde{y}^k) - B^T\tilde{\lambda}^k} \ + B^T(\tilde{\lambda}^k - \lambda^k) \\ +\big[s(\tilde{y}^k - y^k) - \big(g(\tilde{y}^k) - g(y^k)\big)\big] \end{array} \right\} \ge 0, \quad \forall y \in \mathcal{Y}. \tag{6.5b}$$

根据 (6.3c) 得到的 $\tilde{\lambda}^k$, 我们有

$$\tilde{\lambda}^k \in \Re^m, \quad (\lambda-\tilde{\lambda}^k)^T \left\{ \begin{array}{l} \underline{(A\tilde{x}^k + B\tilde{y}^k - b)} + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) \\ -A(\tilde{x}^k - x^k) - B(\tilde{y}^k - y^k) \end{array} \right\} \ge 0, \quad \forall \lambda \in \Re^m. \tag{6.5c}$$

这样写, 是为了让上面三个式子中下波纹线的部分合在一起就成了 (6.2) 中的 $\mathbb{F}(\tilde{w}^k)$.

引理 3 从给定的 $w^k = (x^k, y^k, \lambda^k)$ 出发, 由 (6.3) 产生的预测点 $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ 满足

$$\tilde{w}^k \in \Omega, \quad (w-\tilde{w}^k)^T \mathbb{F}(\tilde{w}^k) \ge (w-\tilde{w}^k)^T d(w^k, \tilde{w}^k), \quad \forall w \in \Omega \tag{6.6a}$$

其中

$$d(w^k, \tilde{w}^k) = \left(\begin{array}{c} \big[r(x^k - \tilde{x}^k) - \big(f(x^k) - f(\tilde{x}^k)\big)\big] + A^T(\lambda^k - \tilde{\lambda}^k) \\ \big[s(y^k - \tilde{y}^k) - \big(g(y^k) - g(\tilde{y}^k)\big)\big] + B^T(\lambda^k - \tilde{\lambda}^k) \\ \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) - A(x^k - \tilde{x}^k) - B(y^k - \tilde{y}^k) \end{array} \right). \tag{6.6b}$$

证明 将 (6.5) 的三部分写在一起并利用 $\mathbb{F}(w)$ 的定义经整理便得引理之结论. □

下面我们证明由 (6.6b) 给出的 $d(w^k, \tilde{w}^k)$ 是上升方向. 将 (6.6) 中任意的 w 设为属于 Ω^* 的 w^*, 则有

$$(\tilde{w}^k - w^*)^T d(w^k, \tilde{w}^k) \ge (\tilde{w}^k - w^*)^T \mathbb{F}(\tilde{w}^k). \tag{6.7}$$

利用 $\mathbb{F}$ 的单调性和 w^* 的最优性,

$$(\tilde{w}^k - w^*)^T \mathbb{F}(\tilde{w}^k) \ge (\tilde{w}^k - w^*)^T \mathbb{F}(w^*) \ge 0.$$

因此

$$(w^k - w^*)^T d(w^k, \tilde{w}^k) \ge (w^k - \tilde{w}^k)^T d(w^k, \tilde{w}^k). \tag{6.8}$$

利用 $d(w^k,\tilde{w}^k)$ 的表达式 (6.6b) 和预测条件 (6.4) 得到

$$\begin{array}{rl} & (w^k-\tilde{w}^k)^T d(w^k,\tilde{w}^k) \\ = & r\|x^k-\tilde{x}^k\|^2-(x^k-\tilde{x}^k)^T(f(x^k)-f(\tilde{x}^k)) \\ & +s\|y^k-\tilde{y}^k\|^2-(y^k-\tilde{y}^k)^T(g(y^k)-g(\tilde{y}^k))+\frac{1}{\beta}\|\lambda^k-\tilde{\lambda}^k\|^2 \\ \geq & (1-\nu)\big(r\|x^k-\tilde{x}^k\|^2+s\|y^k-\tilde{y}^k\|^2\big)+\frac{1}{\beta}\|\lambda^k-\tilde{\lambda}^k\|^2. \end{array} \tag{6.9}$$

由于 $\nu\in(0,1)$, 上式右端大于 0. 这里的 (6.8) 和 (6.9) 相当于§3 中的 (3.5) 和 (3.6). 因此, $d(w^k,\tilde{w}^k)$ 是距离函数 $\|w-w^*\|^2$ 的上升方向.

由于 $d(w^k,\tilde{w}^k)$ 是距离函数 $\|w-w^*\|^2$ 的上升方向, 根据 (6.6), 分处不等式 (6.6a) 两边的

$$F(\tilde{w}^k) \qquad 和 \qquad d(w^k,\tilde{w}^k),$$

为一对孪生方向. 我们可以用 H-模下的姊妹方法进行校正. 对给定的正定矩阵 H, 校正公式-I 通过

$$\textbf{(校正公式-I)} \qquad w^{k+1}=w^k-\gamma\alpha_k^* H^{-1}d(w^k,\tilde{w}^k), \quad \gamma\in(0,2) \tag{6.10a}$$

产生新的迭代点. 其中

$$\alpha_k^*=\frac{(w^k-\tilde{w}^k)^T d(w^k,\tilde{w}^k)}{\|H^{-1}d(w^k,\tilde{w}^k)\|_H^2}. \tag{6.10b}$$

定理 5 求解变分不等式 (6.2), 由 (6.3) 预测和 (6.10) 校正产生的序列 $\{\tilde{w}^k\}$ 和 $\{w^k\}$ 满足

$$\|w^{k+1}-w^*\|_H^2\leq\|w^k-w^*\|_H^2-\gamma(2-\gamma)\alpha_k^*(w^k-\tilde{w}^k)^T d(w^k,\tilde{w}^k), \quad \forall w\in\Omega, \tag{6.11}$$

其中 $d(w^k,\tilde{w}^k)$ 由 (6.6b) 给出.

证明 先将 (6.10a) 中的 $\gamma\alpha_k^*$ 置为任意的 $\alpha>0$, 并将输出记为 $w_I^{k+1}(\alpha)$, 并记

$$\vartheta_k^H(\alpha)=\|w^k-w^*\|_H^2-\|w_I^{k+1}(\alpha)-w^*\|_H^2. \tag{6.12}$$

这样

$$\begin{array}{rcl} \vartheta_k^H(\alpha) & \stackrel{(6.10a)}{=} & \|w^k-w^*\|_H^2-\|(w^k-w^*)-\alpha H^{-1}d(w^k,\tilde{w}^k)\|_H^2 \\ & = & 2\alpha(w^k-w^*)^T d(w^k,\tilde{w}^k)-\alpha^2\|H^{-1}d(w^k,\tilde{w}^k)\|_H^2 \\ & \stackrel{(6.8)}{\geq} & 2\alpha(w^k-\tilde{w}^k)^T d(w^k,\tilde{w}^k)-\alpha^2\|H^{-1}d(w^k,\tilde{w}^k)\|_H^2 \\ & =: & q_k^H(\alpha). \end{array} \tag{6.13}$$

对二次函数 $q_k^H(\alpha)$ 求极值得到 (6.10b) 中的 α_k^*. 当 (6.13) 中的 $\alpha=\gamma\alpha_k^*$ 时,

$$\begin{array}{rcl} \vartheta_k^H(\gamma\alpha_k^*) & \geq & q_k^H(\gamma\alpha_k^*) \\ & = & 2\gamma\alpha_k^*(w^k-\tilde{w}^k)^T d(w^k,\tilde{w}^k)-\gamma^2(\alpha_k^*)^2\|H^{-1}d(w^k,\tilde{w}^k)\|_H^2 \\ & = & \gamma(2-\gamma)\alpha_k^*(w^k-\tilde{w}^k)^T d(w^k,\tilde{w}^k). \end{array}$$

证明的最后一个等式使用了 (6.10b) 中

的 $\alpha_k^*\|H^{-1}d(w^k,\tilde{w}^k)\|_H^2 = (w^k-\tilde{w}^k)^T d(w^k,\tilde{w}^k)$. 定理结论得证. □

采用 (6.10) 校正, 理论上可以取维数相配的任何正定的矩阵. 对 (6.6) 给出的 $d(w^k,\tilde{w}^k)$, 建议取

$$H = \begin{pmatrix} rI & 0 & 0 \\ 0 & sI & 0 \\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix}, \tag{6.14}$$

这时, H 是正定的数量对角矩阵,

$$H^{-1}d(w^k,\tilde{w}^k) = \begin{pmatrix} (x^k-\tilde{x}^k) - \frac{1}{r}\big[\big(f(x^k)-f(\tilde{x}^k)\big) - A^T(\lambda^k-\tilde{\lambda}^k)\big] \\ (y^k-\tilde{y}^k) - \frac{1}{s}\big[\big(g(y^k)-g(\tilde{y}^k)\big) - B^T(\lambda^k-\tilde{\lambda}^k)\big] \\ (\lambda^k-\tilde{\lambda}^k) - \beta A(x^k-\tilde{x}^k) - \beta B(y^k-\tilde{y}^k) \end{pmatrix}.$$

对应于校正公式 (6.10), 我们也可以采用姊妹方法中的另一种方法校正, 由

(校正公式-II) $$w^{k+1} = \arg\min\big\{\|w - [w^k - \gamma\alpha_k^* H^{-1}F(\tilde{w}^k)]\|_H^2 \,|\, w\in\Omega\big\} \tag{6.15}$$

产生新的迭代点 w^{k+1}, 其中 $\gamma\in(0,2)$, α_k^* 由 (6.10b) 提供, 可以得到跟校正方法 (6.10) 同样的收缩性质.

定理 6 求解变分不等式 (6.2), 由 (6.3) 预测和 (6.15) 校正产生的序列 $\{\tilde{w}^k\}$ 和 $\{w^k\}$ 满足

$$\|w^{k+1}-w^*\|_H^2 \le \|w^k-w^*\|_H^2 - \gamma(2-\gamma)\alpha_k^*(w^k-\tilde{w}^k)^T d(w^k,\tilde{w}^k), \quad \forall w\in\Omega, \tag{6.16}$$

其中 $d(w^k,\tilde{w}^k)$ 由 (6.6b) 给出.

证明. 先将 (6.15) 中的 $\gamma\alpha_k^*$ 置为任意的 $\alpha>0$ 并将输出记为 $w_{II}^{k+1}(\alpha)$. 我们考察

$$\zeta_k^H(\alpha) = \|w^k-w^*\|_H^2 - \|w_{II}^{k+1}(\alpha)-w^*\|_H^2 \tag{6.17}$$

首先, 因为 $w_{II}^{k+1}(\alpha) = \arg\min\big\{\|w-[w^k-\alpha H^{-1}F(\tilde{w}^k)]\|_H^2 \,|\, w\in\Omega\big\}$, 根据最优性质定理, 有

$$(w - w_{II}^{k+1}(\alpha))^T H\{w_{II}^{k+1}(\alpha) - w^k + \alpha H^{-1}F(\tilde{w}^k)\} \ge 0, \quad \forall w\in\Omega.$$

将上面任意的 w 替换成 w^*, 就有不等式

$$(w_{II}^{k+1}(\alpha)-w^*)^T H\{[w^k-\alpha H^{-1}F(\tilde{w}^k)] - w_{II}^{k+1}(\alpha)\} \ge 0. \tag{6.18}$$

在恒等式

$$(a-b)^T H(c-a) = \tfrac{1}{2}\big(\|c-b\|_H^2 - \|c-a\|_H^2\big) - \tfrac{1}{2}\|a-b\|_H^2$$

中置

$$a = w_{II}^{k+1}(\alpha), \qquad b = w^* \quad \text{和} \quad c = w^k - \alpha H^{-1}F(\tilde{w}^k)$$

并利用 (6.18), 则有

$$\|w^k-\alpha H^{-1}F(\tilde{w}^k) - w^*\|_H^2 \\ \qquad - \|w^k-\alpha H^{-1}F(\tilde{w}^k) - w_{II}^{k+1}(\alpha)\|_H^2 - \|w_{II}^{k+1}(\alpha)-w^*\|_H^2 \;\ge\; 0.$$

因此,

$$\begin{aligned}\|w_{II}^{k+1}(\alpha)-w^*\|_H^2 \quad &\le \quad \|(w^k-w^*)-\alpha H^{-1}F(\tilde{w}^k)\|_H^2 \\ &\quad -\|(w^k-w_{II}^{k+1}(\alpha))-\alpha H^{-1}F(\tilde{w}^k)\|_H^2.\end{aligned}$$

将此代入 (6.17), 我们有

$$\begin{aligned}\zeta_k^H(\alpha) \quad &\ge \quad \|w^k-w^*\|_H^2-\|(w^k-w^*)-\alpha H^{-1}F(\tilde{w}^k)\|_H^2 \\ &\quad +\|(w^k-w_{II}^{k+1}(\alpha))-\alpha H^{-1}F(\tilde{w}^k)\|_H^2 \\ &= \quad 2\alpha(w^k-w^*)^TF(\tilde{w}^k)+2\alpha(w_{II}^{k+1}(\alpha)-w^k)^TF(\tilde{w}^k) \\ &\quad +\|w^k-u_{II}^{k+1}(\alpha)\|_H^2 \\ &= \quad \|w^k-w_{II}^{k+1}(\alpha)\|_H^2+2\alpha(w_{II}^{k+1}(\alpha)-w^*)^TF(\tilde{w}^k).\end{aligned} \tag{6.19}$$

将 (6.19) 中右端的最后一项 $(w_{II}^{k+1}(\alpha)-w^*)^TF(\tilde{w}^k)$ 分解成

$$(w_{II}^{k+1}(\alpha)-w^*)^TF(\tilde{w}^k)=(w_{II}^{k+1}(\alpha)-\tilde{w}^k)^TF(\tilde{w}^k)+(\tilde{w}^k-w^*)^TF(\tilde{w}^k),$$

利用

$$(\tilde{w}^k-w^*)^TF(\tilde{w}^k)\ge(\tilde{w}^k-w^*)^TF(w^*)\ge 0,$$

上式右端的最后一部分非负. 代入 (6.19) 的右端, 进一步得到

$$\zeta_k^H(\alpha)\ge\|w^k-w_{II}^{k+1}(\alpha)\|_H^2+2\alpha(w_{II}^{k+1}(\alpha)-\tilde{w}^k)^TF(\tilde{w}^k). \tag{6.20}$$

因为 $w_{II}^{k+1}(\alpha)\in\Omega$, 用它替代 (6.6a) 中的任意 $w\in\Omega$, 得到

$$(w_{II}^{k+1}(\alpha)-\tilde{w}^k)^TF(\tilde{w}^k)\ge(w_{II}^{k+1}(\alpha)-\tilde{w}^k)^Td(w^k,\tilde{w}^k). \tag{6.21}$$

将它们代入 (6.20) 的右端, 就有

$$\zeta_k^H(\alpha)\ge\|w^k-w_{II}^{k+1}(\alpha)\|_H^2+2\alpha(w_{II}^{k+1}(\alpha)-\tilde{w}^k)^Td(w^k,\tilde{w}^k).$$

对上式右端, 进一步化成

$$\begin{aligned}\zeta_k^H(\alpha) \quad &\ge \quad \|w^k-w_{II}^{k+1}(\alpha)\|_H^2+2\alpha(w_{II}^{k+1}(\alpha)-\tilde{w}^k)^Td(w^k,\tilde{w}^k) \\ &= \quad \|w^k-w_{II}^{k+1}(\alpha)\|_H^2+2\alpha(w_{II}^{k+1}(\alpha)-w^k)^Td(w^k,\tilde{w}^k) \\ &\quad +2\alpha(w^k-\tilde{w}^k)^Td(w^k,\tilde{w}^k) \\ &= \quad \|(w^k-w_{II}^{k+1}(\alpha))-\alpha H^{-1}d(w^k,\tilde{w}^k)\|_H^2-\alpha^2\|H^{-1}d(w^k,\tilde{w}^k)\|_H^2 \\ &\quad +2\alpha(w^k-\tilde{w}^k)^Td(w^k,\tilde{w}^k) \\ &= \quad \|w_I^{k+1}(\alpha)-w_{II}^{k+1}(\alpha)\|_H^2-\alpha^2\|H^{-1}d(w^k,\tilde{w}^k)\|_H^2 \\ &\quad +2\alpha(w^k-\tilde{w}^k)^Td(w^k,\tilde{w}^k) \\ &\ge \quad 2\alpha(w^k-\tilde{w}^k)^Td(w^k,\tilde{w}^k)-\alpha^2\|H^{-1}d(w^k,\tilde{w}^k)\|_H^2\big).\end{aligned} \tag{6.22}$$

在上式中取 $\alpha=\gamma\alpha_k^*$ 并利用 (6.10b) 中的

$$\alpha_k^*\|H^{-1}d(w^k,\tilde{w}^k)\|_H^2=(w^k-\tilde{w}^k)^Td(w^k,\tilde{w}^k),$$

从 (6.22) 得到

$$\zeta_k^H(\gamma\alpha_k^*) \quad \geq \quad \gamma(2-\gamma)\alpha_k^*(w^k-\tilde{w}^k)^T d(w^k,\tilde{w}^k).$$

这样就完成了定理结论 (6.16)的证明. □

我们关心校正公式-II (6.15) 如何实现. 由于 (6.14) 中的 H 是与 $\Omega = \mathcal{X}\times\mathcal{Y}\times\Lambda$ 相对应的分块数量矩阵. 利用可分离结构, 我们有

$$\begin{aligned}
&\min\{\|w-[w^k-\alpha_k H^{-1}I\!\!F(\tilde{w}^k)]\|_H^2 \,|\, w\in\Omega\} \\
= &\min\left\{ \left\| \begin{array}{c} x-[x^k-\alpha_k\frac{1}{r}(f(\tilde{x}^k)-A^T\tilde{\lambda}^k)] \\ y-[y^k-\alpha_k\frac{1}{s}(g(\tilde{y}^k)-B^T\tilde{\lambda}^k)] \\ \lambda-[\lambda^k-\alpha_k\beta(A\tilde{x}^k+B\tilde{y}^k-b)] \end{array}\right\|_H^2 \;\middle|\; \begin{array}{c} x\in\mathcal{X} \\ y\in\mathcal{Y} \\ \lambda\in\Lambda \end{array}\right\} \\
= &\min\left\{ \begin{array}{c} r\|x-[x^k-\alpha_k\frac{1}{r}(f(\tilde{x}^k)-A^T\tilde{\lambda}^k)]\|^2 \,|\, x\in\mathcal{X} \\ +s\|y-[y^k-\alpha_k\frac{1}{s}(g(\tilde{y}^k)-B^T\tilde{\lambda}^k)]\|^2 \,|\, y\in\mathcal{Y} \\ +\frac{1}{\beta}\|\lambda-[\lambda^k-\alpha_k\beta(A\tilde{x}^k+B\tilde{y}^k-b)]\|^2 \,|\, \lambda\in\Lambda \end{array}\right\}.
\end{aligned}$$

因此,

$$\begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} \text{argmin}\{\|x-[x^k-\alpha_k\frac{1}{r}(f(\tilde{x}^k)-A^T\tilde{\lambda}^k)]\|^2 \,|\, x\in\mathcal{X}\} \\ \text{argmin}\{\|y-[y^k-\alpha_k\frac{1}{s}(g(\tilde{y}^k)-B^T\tilde{\lambda}^k)]\|^2 \,|\, y\in\mathcal{Y}\} \\ \text{argmin}\{\|\lambda-[\lambda^k-\alpha_k\beta(A\tilde{x}^k+B\tilde{y}^k-b)]\|^2 \,|\, \lambda\in\Lambda\} \end{pmatrix}.$$

换句话说,

$$\begin{aligned}
w^{k+1} &= \arg\min\{\|w-[w^k-\alpha_k H^{-1}I\!\!F(\tilde{w}^k)]\|_H^2 \,|\, w\in\Omega\} \\
&= \arg\min\{\|w-[w^k-\alpha_k H^{-1}I\!\!F(\tilde{w}^k)]\|^2 \,|\, w\in\Omega\} \\
&= P_\Omega[w^k-\alpha_k H^{-1}I\!\!F(\tilde{w}^k)]. \qquad (6.23)
\end{aligned}$$

这里, 我们需要再一次强调, 上式只有当 H 是形如 (6.14) 的正定分块数量矩阵时才成立. 因此, 求解变分不等式 (6.2), 由 (6.3) 预测, 通过 (6.15) 校正是由

$$\begin{cases} x^{k+1} = P_{\mathcal{X}}\{x^k - \dfrac{\alpha_k}{r}[f(\tilde{x}^k)-A^T\tilde{\lambda}^k]\}, & \text{(6.24a)} \\ y^{k+1} = P_{\mathcal{Y}}\{y^k - \dfrac{\alpha_k}{s}[g(\tilde{y}^k)-B^T\tilde{\lambda}^k]\}, & \text{(6.24b)} \\ \lambda^{k+1} = P_{\Lambda}\{\lambda^k - \alpha_k\beta(A\tilde{x}^k+B\tilde{y}^k-b)\}, & \text{(6.24c)} \end{cases}$$

实现的. 运算执行的主要是分别在 $\mathcal{X}$, $\mathcal{Y}$ 和 Λ 上的欧氏模下的投影.

References

[1] B.S. He, A projection and contraction method for a class of linear complementarity problems and its application in convex quadratic programming, Applied Mathematics and Optimization, **25**, 247–262, 1992.

[2] 何炳生, 论求解变分不等式的一些投影收缩算法, 计算数学, 1996, 18: 54–60.

[3] B.S. He, A new method for a class of linear variational inequalities, Math. Progr., **66**, 137-144, 1994.

[4] B.S. He, Solving a class of linear projection equations, Numerische Mathematik, **68**, 71-80, 1994.

[5] B.S. He, A class of projection and contraction methods for monotone variational inequalities, Applied Mathematics and Optimization, **35**, 69-76, 1997.

[6] 何炳生. 从变分不等式的投影收缩算法到凸优化的分裂收缩算法, 高等学校计算数学学报, 2016, 38: 74–96.

[7] 何炳生. 我和乘子交替方向法20年. 运筹学学报, 2018, 22: 1–31.

[8] 何炳生. 凸优化和单调变分不等式收缩算法的统一框架. 中国科学：数学, 2018, 48: 255–272.

[9] He B S and Liao L Z. Improvements of some projection methods for monotone nonlinear variational inequalities, *Journal of Optimization Theory and Applications*, 2002, 112: 111–128.

[10] B.S. He, L.Z. Liao, and X. Wang, Proximal-like contraction methods for monotone variational inequalities in a unified framework I: Effective quadruplet and primary methods, Comput. Optim. Appl., 2012, 51: 649-679.

[11] B.S. He, L.Z. Liao, and X. Wang, Proximal-like contraction methods for monotone variational inequalities in a unified framework II: General methods and numerical experiments, Comput. Optim. Appl., 2012, 51: 681-708.

[12] 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.

[13] B. S. He and X. M. Yuan, On construction of splitting contraction algorithms in a prediction-correction framework for separable convex optimization, arXiv 2204.11522 [math.OC]

[14] B.S. He, X.M. Yuan and J.J.Z. Zhang, *Comparison of two kinds of prediction-correction methods for monotone variational inequalities*, Comput. Optim. Appl., 2004, 27: 247-267.

[15] G. M. Korpelevich. The extragradient method for finding saddle points and other problems, Ekonomika i Matematicheskie Metody, 1976, 12: 747-756.

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

[17] H. Zheng, F. Liu and X.L. Du, Complementarity problem arising from static growth of multiple cracks and MLS-based numerical manifold method, Computer Methods in Applied Mechanics and Engineering, 2015, 295: 150–171.

[18] H. Zheng, P. Zhang and X.L. Du, Dual form of discontinuous deformation analysis, Computer Methods in Applied Mechanics and Engineering, 2016, 305: 196–216.

[19] T. Zhu and Z. G. Yu, A simple proof for some important properties of the projection mapping. Math. Inequal. Appl. **7**, 453–456, 2004.