凸优化和单调变分不等式的收缩算法
— 统一框架与应用

前言目录与简要说明

南京大学数学系　　何炳生

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

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 ✠

This lecture series study the algorithms for the following problems

- Convex Optimization. $\min\{f(x)|x \in \mathcal{X}\}$, f convex function, $\mathcal{X}$ convex set.
- Variational Inequality. $u \in \Omega, \quad (u'-u)^T F(u) \geq 0, \quad \forall u' \in \Omega$
- $\min$-$\max$ problem $\min\limits_{x\in\mathcal{X}} \max\limits_{y\in\mathcal{Y}} \{\mathcal{L}(x,y) = \theta_1(x) - y^T Ax - \theta_2(y)\}$
- Constrained Convex Optimization $\min\{\theta(x)|Ax = b\,(\text{or} \geq b),\ x \in \mathcal{X}\}$
- Separable COP $\min\{\theta_1(x) + \theta_2(y)|Ax + By = b, x \in \mathcal{X}, y \in \mathcal{Y}\}$
- Multi-blocks separable convex optimization problems
  $\min\{\theta_1(x)+\theta_2(y)+\theta_3(z)|Ax+By+Cz = b, x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\}$

变分不等式和邻近点算法是我们的两大法宝.

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.

这个系列讲义主要包括以下四个方面的研究工作：1, 变分不等式的投影收缩算法; 2, 约束凸优化问题变分不等式框架下的邻近点算法; 3, 可分离凸优化问题的乘子交替方向法 (ADMM); 4, 多块可分离凸优化问题的 ADMM-类分裂收缩算法. 做出这些工作, 需要用到的只是

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

这些算法, 已经被工程界采用解决了不少问题, 也得到了不同学科的一些著名学者引用. 常有一些年轻学者, 来信告知在读我的文章. 为让读者较快地掌握我们的基本想法与套路, 我把这些工作整理总结, 分四个主题写了 20 个讲义, 每个主题五讲. 表述或有不当, 但都是些无需高深数学就能阅读理解的篇章.

♣ 第一部分是关于投影收缩算法方面的工作. 设 $\Omega \subset \Re^n$ 是一个闭凸集, F 是从 $\Re^n$ 到自身的一个算子, 变分不等式问题 VI(Ω, F), 是求 u^*, 使得

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

投影收缩算法的基本操作是通过投影

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

提供预测点 $\tilde{u}^k$ 并由此产生距离函数的下降方向, 校正产生的新迭代点进一步向解集 Ω^* 靠近. 对投影收缩算法的研究, 我们始终追求简明统一的原则. 一个统一框架, 二个孪生方向, 三个基本不等式. 不同方向, 竟可用相同步长, 自感美妙又神奇. 投影收缩算法的名称取自在 Springer 出版社1975 年出版的 Blum 和 Oettli 的德文专著, 两位作者都是苏黎世大学的博士, Oettli 生前曾是 Math. Programming 的编委. 出于对历史的敬畏和对前人工作的尊重, 根据算法的投影收缩特征, 我们借用了 Blum 和 Oettli 书中的算法名称.

这一部分的最后两讲研究算法的收敛速率, 初学者可以跳过这些内容.

投影收缩算法得到了包括 UC Berkeley 计算机系 M. I. Jordan 教授 (Jordan 教授是机器学习界最有影响的美国科学院和工程院院士, 2018 年世界数学家大会一小时邀请被告人) 在内的学者在机器学习中应用. 我们的一些相当细致的计算法则被他们借鉴, 相关的定理被写进他们论文的附录. 国内, 除了中科院岩土所的科技工作者将投影收缩算法成功用于许多岩土工程问题的求解以外, 这类方法也被成功应用到机器人的运动规划和实时控制中.

♣ 第二部分是关于变分不等式框架下的邻近点 (PPA) 算法. 我们对线性约束的凸优化问题

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

引进 Lagrange 乘子 λ, 将求 Lagrange 函数

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

在 $\Omega = \mathcal{X} \times \Lambda$ 的鞍点 (x^*, λ^*). 求 Lagrange 函数的鞍点等价于求变分不等式

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

的解, 其中

$$u = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \qquad F(u) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix}, \qquad \Omega = \mathcal{X} \times \Lambda.$$

然后在变分不等式框架下构造邻近点算法. 这种做法在处理大型结构性优化问题时的优越性, 得到图像学界越来越多的学者认可. 我们 2012 年在 SIAM J. Imaging Science 的文章, 初稿就被一些欧美学者在他们即将发表的论文中引用. A. Chambolle（Chambolle 是在图像学领域很有影响的学者）和 T. Pock 的文章中就说到, He and Yuan 的 PPA 形式, 极大地简化了收敛性分析 (which greatly simplified the convergence analysis), 称之为一个 elegant interpretation. 此外, 一些美国名校都有青年学者关注我们的工作, 采用我们提出的变分不等式框架处理问题. S. Becker (E. Candes 的学生) 在他的一篇论文中的第一句话

就说 Recent works such as [HY12] have proposed a very simple yet powerful technique for analyzing optimization methods, 指的是我们 2012 年发表在SIAM J. Imaging Science 上的这类工作.

♣ 第三部分的五个讲义是关于交替方向法 (ADMM) . ADMM 处理的是具有可分离结构的等式约束凸优化问题

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

引进 Lagrange 乘子 λ, 问题的 Lagrange 函数是定义在 $\Omega=\mathcal{X}\times\mathcal{Y}\times\Re^m$ 上的

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

求 Lagrange 函数的鞍点 $w^*=(x^*,y^*,\lambda^*)$ 就等价于求变分不等式的解

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

其中 $\theta(u)=\theta_1(x)+\theta_2(y)$,

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

在收缩算法框架下研究交替方向法是我们从 1997 年就开始的一个主要研究课题, 尽管当时我们着眼的是管理科学中的变分不等式. ADMM 最近十多年来成为信息科学、机器学习中一个非常有用的热门工具. 我们在 2000 年文章中关于交替方向法中参数选择的调比准则, 被 Stanford 大学 S. Boyd 教授 (Boyd 教授是美国工程院院士, 2006 年世界数学家大会邀请报告人, 2017 年当选为中国工程院外籍院士) 在 2010 年的一篇综述文章中称为一个简单而有效的公式 (A simple scheme that often works well), 对我们的分析依据也作了简要介绍. 他们近年开发的凸优化求解器 SnapVX 的说明文章中也注明参考了我们的调比法则. 2012 年以来, 我们发表了 PPA 意义的 ADMM 和对称形式的 ADMM. 关于交替方向法收敛速率的文章, 分别发表在 SIAM Numer. Anal. 和 Numer. Mathematik 等有较大影响的数学期刊上.

♣ 第四部分是关于多块可分离凸优化的分裂收缩算法. 交替方向法处理的是含两块可分离结构的问题, 对于超过两块的凸优化问题, 以三块的问题

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

为例, 采用直接推广的交替方向法虽然效果还行, 但是理论上不能保证收敛. 为求解多块可分离结构的凸优化问题, 我们提出了一些修正的 ADMM 类方法. 这些修正的方法都对问题不做任何（例如函数强凸, 罚因子限制方面）的要

求, 这相当于延用了 ADMM 方法的所有好性质. 有关方法分别发表在 SIAM Optim. 和 IMA Numer. Anal. 等刊物. 有的方法已经被 UCLA 的 S. Osher 教授（Osher 是美国工程院院士, 两次世界数学家大会邀请报告人, 还是 2014 年世界数学家大会高斯奖得主）和他的合作者用在非负矩阵分解和降维问题上. 他们的文章花整整一页的篇幅介绍如何将我们的方法用来求解他们的具体问题, 指出 (The method proposed by He, Tao and Yuan is appropriate for this application). 最后一讲说明, 这些算法都属于同一个算法框架.

最优化是一门接地气的应用学科。 统一框架, 指导我们针对问题设计算法, 也帮助我们简化算法的收敛性证明. 理论有了保证, 还要计算验证.

尺有所短, 寸有所长. 没有一个方法是对所有的问题都是最好的求解方案. 自己做些数值实验, 多一点计算经验, 才能给初学者一些有价值的参考意见.

♣ 从研究出发点开始, 介绍收缩算法的基本原理。分若干篇章, 对问题类型和主要方法作了梳理。谈问题, 说算法, 讲应用, 给基本算例, 也附简单程序。一条主线, 一个模式。读懂一个算法, 理解后面的篇章就不要再费多少力气。

One algorithm framework should be flexible enough to solve many problems !

凸优化和单调变分不等式的收缩算法 — 目录

第一部分: 单调变分不等式的求解方法

第一讲	变分不等式是应用数学中许多问题的统一表述模式	1
第二讲	三个基本不等式和变分不等式的投影收缩算法	21
第三讲	单调变分不等式收缩算法中的两对孪生方法	45
第四讲	* 线性单调变分不等式投影收缩算法的收敛速率	69
第五讲	* 非线性单调变分不等式投影收缩算法的收敛速率	87

第二部分: 凸优化问题 $\{\min\theta(x)|Ax=b, x\in X\}$ 的求解方法

第六讲	为线性约束凸优化问题定制的 PPA 算法及其应用	107
第七讲	线性约束凸优化问题基于松弛 PPA 的收缩算法	129

第八讲 基于增广 Lagrange 乘子法的 PPA 收缩算法 151

第九讲 基于LVI -PC 方法求解复合凸优化的收缩算法 169

第十讲 基于梯度投影求解凸优化的收缩算法和下降算法 189

第三部分: 凸优化问题 $\min\left\{\theta_1(x)+\theta_2(y)\,\middle|\,\begin{array}{l}Ax+By=b\\ x\in X, y\in Y\end{array}\right\}$ 的交替方向法

第十一讲 结构型优化的交替方向法(ADMM) 211

第十二讲 线性化的交替方向收缩算法 (Linearized ADMM) 229

第十三讲 定制 PPA 意义下的交替方向法及其线性化方法 247

第十四讲 自变量 x-y 地位平等的对称型化交替方向法 269

第十五讲 在统一框架下研究交替方向法的收敛速率 289

第四部分: 多个可分离算子凸优化问题带简单校正的分裂方法

第十六讲 三块可分离凸优化问题的平行分裂 ALM 乘子法 313

第十七讲 三块可分离凸优化问题的略有改动的交替方向法 331

第十八讲 多块可分离凸优化带高斯回代的交替方向算法 355

第十九讲 多块可分离凸优化问题部分平行正则化的交替方向法 371

第二十讲 变分不等式意义下凸优化分裂收缩算法的统一框架 387

可只读不带 $*$ 的篇章. 分解降低难度, 整合把握方向, 是该系列讲义的主要思想.
懂些变分不等式的基本概念, 对凸优化收缩算法的设计和收敛性分析很有帮助.

懂些变分不等式的基本概念, 对凸优化收缩算法的设计和收敛性分析很有帮助.

好的优化方法, 应该是容易被工程师们掌握, 自己用来解决问题的方法.
为相关学科所用, 恰是我们从事优化方法研究的一贯追求.
授人以鱼, 不如授人以渔. 有需求的读者, 慢慢读下去, 应该能够无师自通.

凸优化和单调变分不等式的收缩算法

第一讲：变分不等式是应用数学中许多问题的统一表述模式

Variational inequality is a uniform approach
for many problems in applied mathematics

南京大学数学系　　何炳生
hebma@nju.edu.cn

设 $\Omega \subset \Re^n$ 是一个闭凸集, F 是从 $\Re^n$ 到自身的一个算子, 变分不等式问题 VI(Ω, F), 是求 x^*, 使得

$$x^* \in \Omega, \quad (x - x^*)^T F(x^*) \geq 0, \quad \forall x \in \Omega. \tag{0.1}$$

一个变分不等式称为单调的, 是指 VI(Ω, F) 中的 F 是 Ω (或 $\Re^n$) 上的一个单调算子 (monotone operator), 即 F 满足

$$(x - y)^T (F(x) - F(y)) \geq 0, \quad \forall\, x,\, y \in \Omega \,(\text{resp. } \Re^n).$$

变分不等式是许多应用数学问题的一种统一表述模式. 管理科学与统计计算中存在大量凸优化问题. 信号处理, 图像恢复, 矩阵完整化, 机器学习等信息技术领域中也有许多问题可以归结为(或松弛成)一个凸优化问题. 凸优化的一阶必要性条件就是一个单调变分不等式. 在变分不等式的框架下研究凸优化的求解方法, 就像微积分中用导数求一元函数的极值, 常常会更加方便.

除了通常的最优化问题以外, 互补问题是约束为正卦限的变分不等式. 经济活动中的空间价格平衡, 保护资源– 保障供给中的调控手段, 用经济手段解决交通疏导等问题, 都可以用变分不等式（或其特殊形式互补问题）来描述. 这一讲叙述以下一些常见问题与变分不等式的关系.

- 凸优化与单调变分不等式
- 商品流通、保护资源、保障供给中的变分不等式
- 交通疏导中的变分不等式
- 广义线性规划问题的线性变分不等式
- 最短距离和问题的线性变分不等式
- 极小化最大特征值之变分不等式

1 凸优化与单调变分不等式

连续优化方法中一些代表性的数学模型

- 简单约束优化问题 $\min\{f(x) \mid x \in \Omega\}$.
- 线性约束优化问题 $\min\{\theta(x) \mid Ax = b,\ x \in \mathcal{X}\}$.
- 结构型优化问题 $\min\{\theta_1(x) + \theta_2(y) | Ax + By = b,\ x \in \mathcal{X},\ y \in \mathcal{Y}\}$.
- 非线性互补问题 $x \geq 0,\ \ F(x) \geq 0,\ \ x^T F(x) = 0$.
- 变分不等式 $x \in \Omega,\quad (x' - x)^T F(x) \geq 0,\ \ \forall x' \in \Omega$.

1.1 可微凸优化等价于一个特殊的单调变分不等式

设 Ω 是 $\Re^n$ 中的非空闭凸集, 我们先来讨论可微凸优化问题

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

的最优性条件. 类似于瞎子爬山原理, 我们有

- 如果某一点 x^* 是问题 (1.1) 的最优点, 它必须属于 Ω,
- 并且从这点出发的所有可行方向都不是下降方向.

我们用 $\nabla f(x)$ 表示 $f(x)$ 的梯度, 并记

- $Sd(x) = \{s \in \Re^n \mid s^T \nabla f(x) < 0\}$, 为点 x 处的下降方向集;
- $Sf(x) = \{s \in \Re^n \mid s = x' - x,\ x' \in \Omega\}$, 为点 x 处的可行方向集.

$$x^* \text{ 是最优解} \iff x^* \in \Omega \quad \text{且} \quad Sf(x^*) \cap Sd(x^*) = \emptyset$$

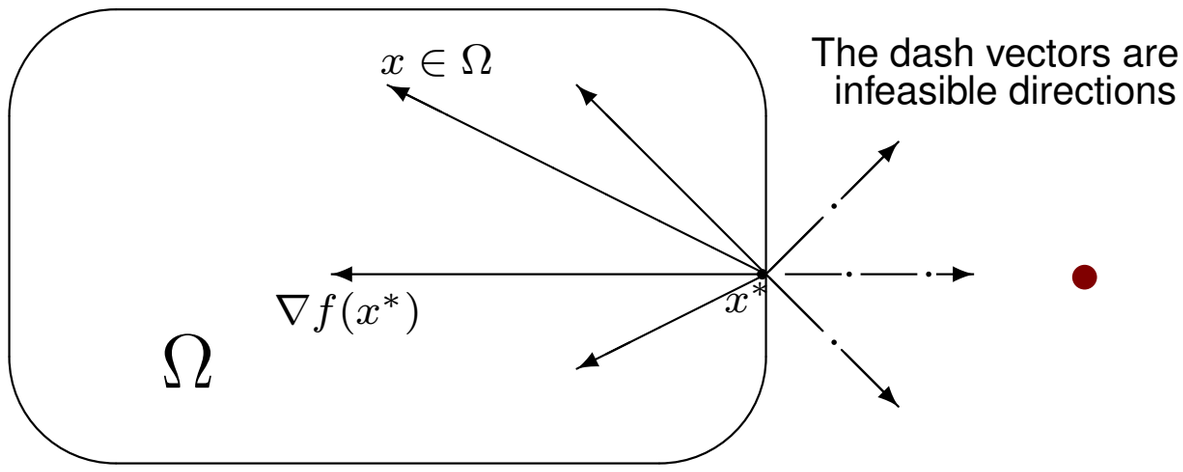


Fig. 1.1 Differential Convex Optimization and VI

上式可以写成： $x^* \in \Omega, \quad (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall x \in \Omega.$

将 $\nabla f(x)$ 写成 $F(x)$, 可微凸优化问题 (1.1) 就归结为求

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

以上形式就是一个变分不等式. 注意到, 在一般变分不等式问题 VI(Ω, F), 我们并不要求 (0.1) 中的算子 F 是某个多元函数的梯度.

1.2 线性约束的凸优化问题

我们考虑一般的线性约束凸优化问题

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

其中 $\theta : \Re^n \to \Re$ 是凸函数, $\mathcal{X} \subset \Re^n$ 是闭凸集, $A \in \Re^{m \times n}$, $b \in \Re^m$.

设 λ 是 Lagrange 乘子, 问题 (1.3) 的 Lagrange 函数

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

定义在 $\mathcal{X} \times \Re^m$ 上. Lagrange 函数的鞍点 $(x^*, \lambda^*) \in \mathcal{X} \times \Re^m$ 满足

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

为了简便, 我们假设 $\theta(x)$ 是可微的, 并且记 $\nabla\theta(x) = f(x)$, 鞍点问题等价于

$$\left\{ \begin{array}{lll} x^* \in \mathcal{X}, & (x - x^*)^T (f(x^*) - A^T \lambda^*) \geq 0, & \forall\, x \in \mathcal{X}, \\ \lambda^* \in \Re^m, & (\lambda - \lambda^*)^T (Ax^* - b) \geq 0, & \forall\, \lambda \in \Re^m. \end{array} \right. \tag{1.4}$$

上述最优性条件的第二部分也实际上就是 $Ax^* = b$. 如果记

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

那么, (1.4) 可以写成变分不等式

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

当 (1.3) 中的等式约束 $Ax = b$ 改成不等式约束 $Ax \geq b$ 时, 相应的变分不等式中只要将 $\Omega = \mathcal{X} \times \Re^m$ 改成 $\Omega = \mathcal{X} \times \Re^m_+$.

如果凸函数 $\theta(x)$ 是非光滑的, 记 $F_a(u) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix}$, 那么, 求 Lagrange 函数的鞍点就等价于求混合变分不等式 (有时就简称为变分不等式)

$$\boxed{u^* \in \Omega, \quad \theta(x) - \theta(x^*) + (u - u^*)^T F_a(u^*) \geq 0, \quad \forall\, u \in \Omega}$$

的解 u^*. 记 $F_a(u) = Mu + q$, 则 M 是斜对称矩阵, $F_a(u)$ 是仿射单调算子.

1.3 可分离结构的线性约束凸优化问题

可分离结构的线性约束凸优化问题是指

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b,\ x \in \mathcal{X},\ y \in \mathcal{Y}\} \tag{1.5}$$

其中 $\theta_1(x)$ 和 $\theta_2(y)$ 是可微凸函数. 设 λ 是 Lagrange 乘子, 上述问题的 Lagrange 函数

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

定义在 $\mathcal{X} \times \mathcal{Y} \times \Re^m$ 上, 设 $f(x) = \nabla\theta_1(x),\ g(y) = \nabla\theta_2(y)$. 这个可分离结构凸优化问题的一阶最优性条件是

$$(x^*, y^*, \lambda^*) \in \Omega, \quad \left\{ \begin{array}{l} (x - x^*)^T(f(x^*) - A^T\lambda^*) \geq 0, \\ (y - y^*)^T(g(y^*) - B^T\lambda^*) \geq 0, \\ (\lambda - \lambda^*)^T(Ax^* + By^* - b) \geq 0, \end{array} \right. \quad \forall\, (x, y, \lambda) \in \Omega, \tag{1.6}$$

其中

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

利用

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

最优性条件 (1.6) 可以写成

$$\boxed{w^* \in \Omega, \quad (w-w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega.}$$

当 $\theta_1(x), \theta_2(y)$ 是不可微凸函数时, 记 $u=(x,y)$ 为 w 的部分分量, 并记 $\theta(u)=\theta_1(x)+\theta_2(y)$, 问题 (1.5) 等价于混合变分不等式

$$\boxed{w^* \in \Omega, \ \big(\theta(u)-\theta(u^*)\big)+(w-w^*)^T F_a(w^*) \ge 0, \ \forall\, w \in \Omega.}$$

我们在后面的章节中展示, 在变分不等式框架下研究凸优化求解方法, 无论在算法设计和收敛性证明方面, 会变得相当简单和方便.

1.4 非线性互补问题是一类特殊的变分不等式

凸优化问题 (1.1) 中, 如果 $\Omega=\Re^n_+$ (实 n-维空间中的非负卦限) 并且 $f(x)$ 是可微的, 根据 §1.1 中的分析, x^* 是最优解的充要条件是

$$x^* \ge 0, \quad (x-x^*)^T \nabla f(x^*) \ge 0, \quad \forall\, x \ge 0.$$

设 F 是 $\Re^n \to \Re^n$ 的一个算子. 非负卦限 $\Re^n_+$ 上变分不等式的数学形式是

$$\boxed{\mathsf{VI}(\Re^n_+, F): \quad x^* \ge 0, \quad (x-x^*)^\top F(x^*) \ge 0, \quad \forall x \ge 0.}$$

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

$$\boxed{\textbf{(NCP)} \qquad x^* \ge 0, \quad F(x^*) \ge 0, \quad {x^*}^\top F(x^*) = 0. \qquad (1.7)}$$

事实上, **NCP** 是 $\Omega=\Re^n_+$ 的一类变分不等式. 因此, 在 $\Omega=\Re^n_+$ 时, 可微凸优化问题 (1.1) 与一个互补问题等价.

证明 为帮助阅读理解, 我们给出 **NCP** 与 **VI**$(\Re^n_+, F)$ 等价的具体证明.

如果 x^* 是 **NCP** 的解, 那么有 $x^* \geq 0$ 和 $F(x^*) \geq 0$.

对于任意的 $x \geq 0$ 有 $x^\top F(x^*) \geq 0$.

又因 ${x^*}^\top F(x^*) = 0$, 得

$$\begin{aligned}(x - x^*)^\top F(x^*) \\ = \quad x^\top F(x^*) \\ \geq \quad 0.\end{aligned}$$

所以 x^* 是 **VI**(R^n_+, F) 的一个解.

反过来, 如果 x^* 是 **VI**(R^n_+, F) 的一个解, 则 $x^* \geq 0$.

将 $x = 2x^*$ 和 $x = 0$ 代入 $(x - x^*)^\top F(x^*) \geq 0$, 我们得到 $\pm {x^*}^\top F(x^*) \geq 0$. 因此 ${x^*}^\top F(x^*) = 0$.

要证明 x^* 是 **NCP** 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 $(x - x^*)^\top F(x^*) = F_j(x^*) < 0$, 这与 x^* 是 **VI**(R^n_+, F) 的解矛盾.

1.5 最小一模问题与等价的变分不等式

最小一模问题 (Least absolute deviations) 的数学形式是

$$\min \|Ax - b\|_1, \tag{1.8}$$

其中 $A \in \Re^{m \times n}$, $b \in \Re^m$. 与最小二乘相比, 它提供了一种更稳健的数据拟合模型. 与最小二乘不同, 它是一个非光滑凸优化问题. 这类问题被广泛应用于统计学和经济研究, 一些最新的文献都有提及. 例如:

- T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer, second edition, 2009. §10.6.
- J. M. Wooldridge. Introductory Econometrics: A Modern Approach. South Western College Publications, fourth edition, 2009. §9.6.

我们用 e 表示每个分量都是 1 的 m-维向量. 容易验证

$$\|d\|_1 = \max\{y^T d \,|\, y \in B_\infty\}, \qquad B_\infty = \{y \in \Re^m \mid -e \leq y \leq e\}.$$

因此, 问题 (1.8) 就是一个形式为

$$\min_{x\in\Re^n}\ \max_{y\in B_\infty}\quad y^T(Ax-b).$$

的 min-max 问题. 它的等价形式是下面的线性变分不等式:

$$x^*\in\Re^n,\ y^*\in B_\infty,\quad \begin{cases}(x-x^*)^T(A^Ty^*)\ge 0, & \forall x\in\Re^n,\\ (y-y^*)^T(-Ax^*+b)\ge 0, & \forall y\in B_\infty.\end{cases}$$

可以写成更简单紧凑的形式

$$u^*\in\Omega,\quad (u-u^*)^T(Mu^*+q)\ge 0,\quad \forall u\in\Omega$$

其中

$$u=\begin{pmatrix}x\\ y\end{pmatrix},\quad M=\begin{pmatrix}0 & A^T\\ -A & 0\end{pmatrix},\quad q=\begin{pmatrix}0\\ b\end{pmatrix},\quad \Omega=\Re^n\times B_\infty.$$

只对结构型凸优化一阶算法感兴趣的读者, 这一讲的后继部分只需浏览一下.

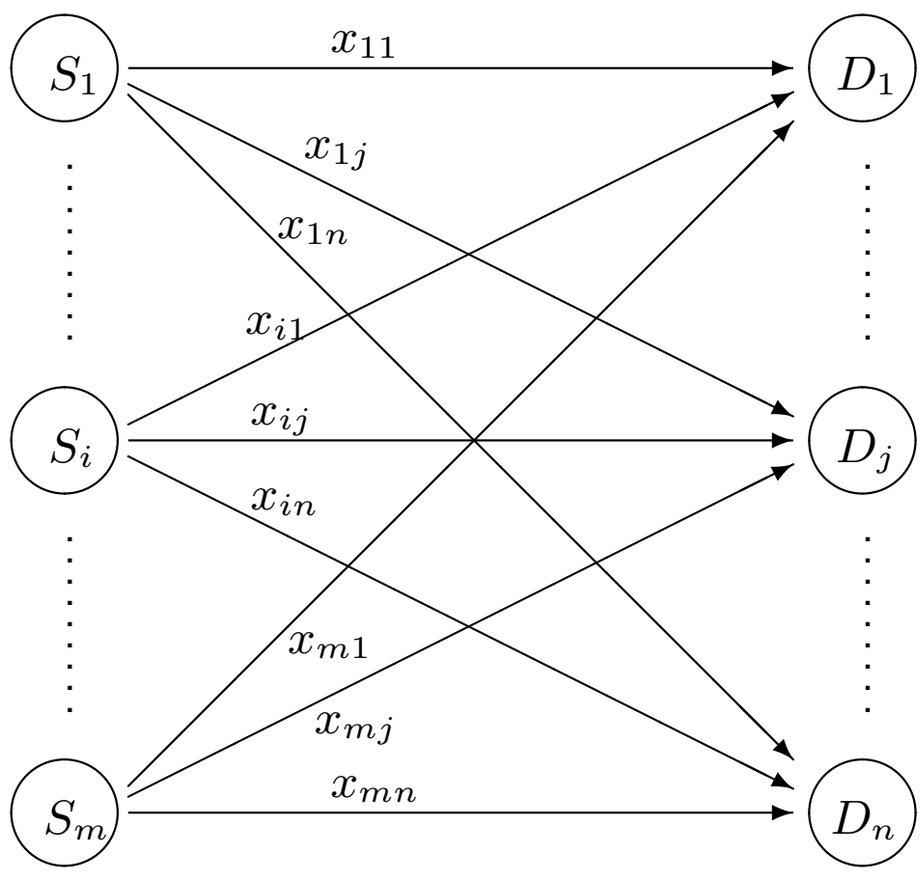
2 保护资源、保障供给中的互补问题

我们以下面的例子来说明一个经济平衡模型:

假设某种商品（例如煤）由 m 个资源地生产和 n 个需求地消费. 它由经营者们从资源地采购运到需求地销售. 经营者会根据贪婪原理找到他们的最优经营方案. 记

$s_i=\sum_{j=1}^n x_{ij}$

$d_j=\sum_{i=1}^m x_{ij}$



一些记号

S_i: 该种商品的第 i 个资源地;

D_j: 该种商品的第 j 个需求地;

x_{ij}: 从 S_i 到 D_j 的交易量;

s_i: 经营者们在资源地 S_i 的采购总量, $s_i = \sum_{j=1}^{n} x_{ij}$;

d_j: 经营者在需求地 D_j 的销售总量, $d_j = \sum_{i=1}^{m} x_{ij}$;

h_i^s: 经营者在资源地 S_i 处的采购价, 是 S_i 处采购量的函数;

h_j^d: 经营者在需求地 D_j 处的销售价, 是 D_j 处到货量的函数;

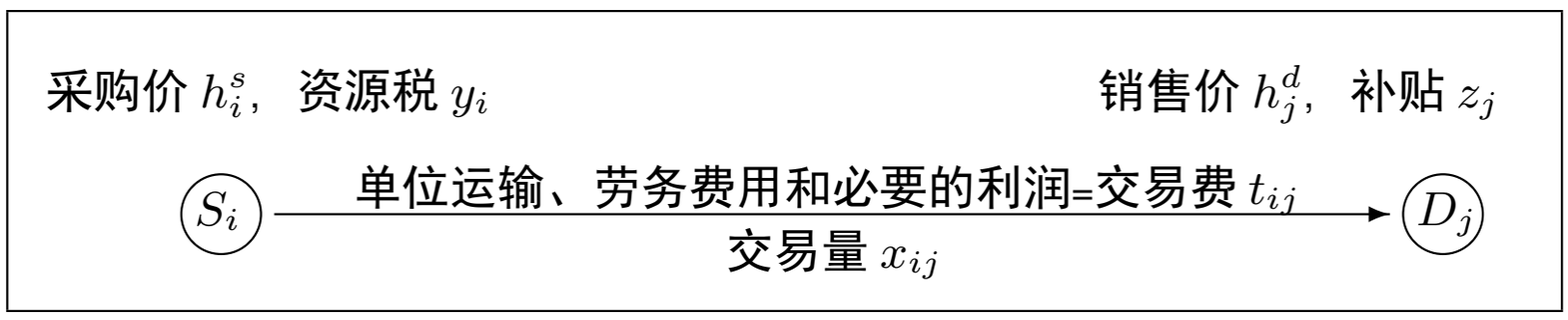
t_{ij}: 从 S_i 到 D_j 的交易费用 (包括运输费用等交易成本);

y_i: 政府为避免资源过度开采而在资源地 S_i 向经营者征收的资源税;

z_j: 政府为保障供给而在需求地 D_j 给经营者的经营补贴.

2.1 经营者追求利益最大化之互补问题

用下图表示第 i 个资源地 S_i 和第 j 个需求地 D_j 之间的采购— 销售关系.



如果$(h_i^s + y_i + t_{ij}) \geq (h_j^d + z_j)$, 没有人愿意做亏本买卖, 所以 $x_{ij} = 0$;

反之，根据贪婪原理，经营者会尽可能增大经营量 x_{ij},（通常这会导致采购价的上涨和销售价的下降）直到

$$(h_i^s + y_i + t_{ij}) = (h_j^d + z_j).$$

用数学语言描述就是:

$$h_i^s + y_i + t_{ij} \begin{cases} \geq h_j^d + z_j, & \text{如果} x_{ij} = 0, \\ = h_j^d + z_j, & \text{如果} x_{ij} > 0\,. \end{cases}$$

- 对给定的资源税率 $y_1, y_2, \cdots, y_m$ 和补贴标准 $z_1, z_2, \cdots, z_n$ (政策),

  经营者们会根据“贪婪原理”找到他们的最优经营方案 x_{ij} (对策).

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix},$$

  如果记

$$F(X) = \begin{pmatrix} F_{11} & F_{12} & \dots & F_{1n} \\ F_{21} & F_{22} & \dots & F_{2n} \\ \vdots & \vdots & & \vdots \\ F_{m1} & F_{m2} & \dots & F_{mn} \end{pmatrix},$$

  其中

$$F_{ij}(X) = \{h_i^s(s_i) + y_i + t_{ij}\} - \{h_j^d(d_j) + z_j\},$$

$$s_i = \sum_{j=1}^{n} x_{ij}, \qquad d_j = \sum_{i=1}^{m} x_{ij}.$$

  那么, 由经营者的最优经营方案形成的非负 $m \times n$ 矩阵 X 和 $F(X)$ 中, 同样下标的元素中最多只能有一个大于零. 换句话说, 经营者的最优经营方案是互补问题

$$\boxed{X \geq 0, \qquad F(X) \geq 0, \qquad \textbf{Trace}(X^T F(X)) = 0}$$

  的解. 上式中的 **Trace**$(\cdot)$ 表示矩阵的“迹”—— 矩阵对角元的和.

- 我们能观测到的是问题解中相应的

$$\boxed{s_i(y,z) = \textstyle\sum_{j=1}^{n} x_{ij}(y,z) \quad \text{and} \quad d_j(y,z) = \sum_{i=1}^{m} x_{ij}(y,z)}$$

  它们是由政策向量 (y, z) 决定的.

2.2 政府部门制定最优政策之互补问题

政府部门的职责—保护资源和保障供给

- 坚持可持续发展及环境保护，防止资源被过度开采，$s_i \le s_i^{\max}$;
- 保障基本供给，从而保证社会稳定和民生，$d_j \ge d_j^{\min}$.

手段(政策)：

- 在资源过度开采的产地征收资源税，$y_1, \cdots, y_m$；
- 供应紧张的需求地给经营者提供补贴，$z_1, \cdots, z_n$.

经营者们会根据政策给出对策. 政府的任务是给出‘最好’的政策.

政府部门需要的最优政策是一个“黑箱”条件下的互补问题:

- 保护了资源, 同时又让经济尽可能繁荣;

$$y \ge 0, \qquad s^{\max} - s(y,z) \ge 0, \qquad y^T(s^{\max} - s(y,z)) = 0,$$

- 保障了供给, 同时又尽可能节约财政支出.

$$z \ge 0, \qquad d(y,z) - d^{\min} \ge 0, \qquad z^T(d(y,z) - d^{\min}) = 0.$$

职能部门的互补问题的紧凑形式是：

$$u \ge 0, \qquad F(u) \ge 0, \qquad u^T F(u) = 0,$$

其中

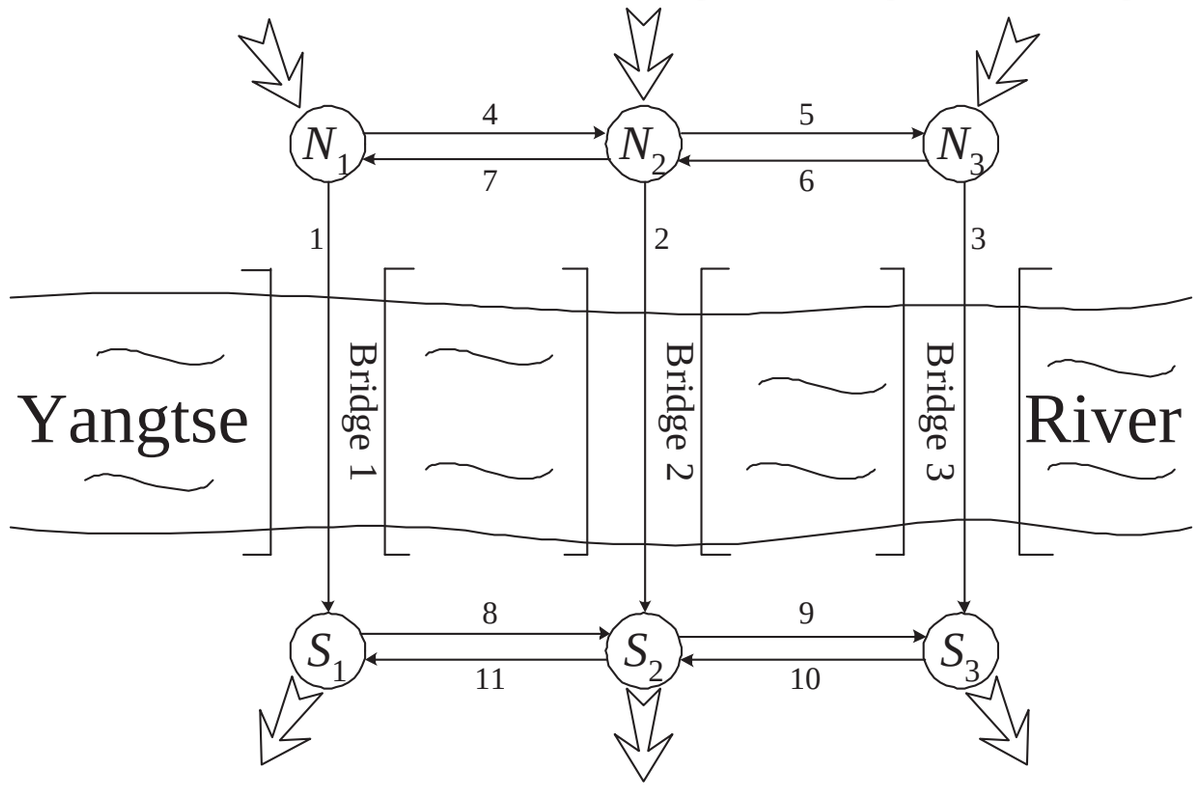
$$u = \begin{pmatrix} y \\ z \end{pmatrix}, \qquad F(u) = \begin{pmatrix} s^{\max} - s(u) \\ d(u) - d^{\min} \end{pmatrix},$$

$s(u)$ 和 $d(u)$ 都是 u 的函数.

黑箱问题, 是指 $F(u)$ 是 u 的函数, 有确定的关系, 但没有函数表达式.

3 交通疏导问题之互补问题

设某跨江城市有三座长江大桥, 分别为 Bridge-1, Bridge-2 和 Bridge-3.



不失一般性, 我们可以将 N_1, N_2, N_3 看作由北向南的车辆在江北的出发地, 把 S_1, S_2, S_3 看作它们在江南的集散地.

我们只对管理部门的问题感兴趣, 假设只收过桥费. 管理部门想制定一个适当的收费标准合理控制桥上流量.

驾驶员基于 Wardrop 原理的最优出行方案 — 最小费用路径

对给定的大桥收费 $y = (y_1, y_2, y_3)$, 驾驶员会找到他们的最优出行方案.

管理部门要求通过大桥合理收费控制桥上的流量

- $0 \le y \in \Re^3$: 桥上的收费向量;
- $f(y) \in \Re^3$: 桥上的流量, 它是收费 y 的函数;
- $0 < b \in \Re^3$: 管理部门希望控制的桥上的流量上界.

管理部门要求解的数学问题是

$$y \ge 0, \qquad F(y) = b - f(y) \ge 0, \qquad y^T F(y) = 0.$$

同样, 流量 $f(y)$ 确是收费 y 的函数, 但没有表达式. 只能对给定的自变量, 观测相应的函数值, 而这种观测, 往往代价不菲. 我们从事投影收缩算法研究, 着眼点是要得到效率高一些的、只用函数值和少用函数值的方法.

4 广义线性规划及鞍点问题之变分不等式

4.1 广义线性规划及之变分不等式

线性规划的标准形式是 $\min\{c^T x \mid Ax = b, x \geq 0\}$, 其中 $A \in \Re^{m\times n}, b \in \Re^m$, $c \in \Re^n$. 在实际经济问题中, 向量 b 一般表示需求量, c 表示价格. 我们允许 b 和 c 都在一定范围之内变动, 因此考虑更一般的问题

$$\min\{\max_{\eta\in C} \eta^T x \mid Ax \in B,\ x \in D\} \tag{4.1}$$

其中 $C, D \subset \Re^n$, $B \subset \Re^m$ 是闭凸集. 我们称这样的问题为广义线性规划. 引进辅助变量 y 和 Lagrange 乘子 λ, 得到广义线性规划 (4.1) 的 Lagrange 函数

$$L(x, y, \lambda, \eta) = \eta^T x - \lambda^T (Ax - y),$$

它定义在 $(D \times B) \times (\Re^m \times C)$ 上. 广义线性规划等价与一个 min-max 问题

$$L_{\lambda\in\Re^m, \eta\in C} L(x^*, y^*, \lambda, \eta) \leq L(x^*, y^*, \lambda^*, \eta^*) \leq L_{x\in D,\ y\in B}(x, y, \lambda^*, \eta^*).$$

设 $(x^*, y^*, \lambda^*, \eta^*)$ 是上述 min-max 问题的解, 则有

$$\begin{cases} x^* \in D, & (x - x^*)^T(-A^T\lambda^* + \eta^*) \geq 0, & \forall\, x \in D \\ y^* \in B, & (y - y^*)^T(\lambda^*) \geq 0, & \forall\, y \in B \\ \lambda^* \in \Re^m, & (\lambda - \lambda^*)^T(Ax^* - y^*) \geq 0, & \forall\, \lambda \in \Re^m \\ \eta^* \in C, & (\eta - \eta^*)^T(-x^*) \geq 0, & \forall\, \eta \in C. \end{cases}$$

更紧凑的形式可以写成

$$w^* \in \Omega, \quad (w - w^*)^T(Mw^* + q) \geq 0, \quad \forall w \in \Omega$$

其中

$$w = \begin{pmatrix} x \\ y \\ \lambda \\ \eta \end{pmatrix}, \quad M = \begin{pmatrix} 0 & 0 & -A^T & I \\ 0 & 0 & I & 0 \\ A & -I & 0 & 0 \\ -I & 0 & 0 & 0 \end{pmatrix}, \quad q = 0$$

和

$$\Omega = D \times B \times \Re^m \times C.$$

一类常常被考虑的分裂可行问题 (Split Feasibility Problem)

$$\text{Find} \quad x \in D \quad \text{such that} \quad Ax \in B,$$

更是 (4.1) 问题中的一个特例. 将其转换成

$$\text{Find} \quad x \in D,\ y \in B, \quad \text{such that} \quad Ax - y = 0.$$

上述问题相当于一个目标函数为零的约束优化问题. 设 λ 为线性约束 $Ax - y = 0$ 的 Lagrange 乘子, 问题就等价与以下的线性变分不等式:

$$\boxed{u^* \in \Omega, \quad (u - u^*)^T M u^* \geq 0, \quad \forall u \in \Omega}$$

其中

$$u = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad M = \begin{pmatrix} 0 & 0 & -A^T \\ 0 & 0 & I \\ A & -I & 0 \end{pmatrix},$$

和

$$\Omega = D \times B \times \Re^m.$$

4.2 一般鞍点问题之变分不等式

用全变差极小处理图像去模糊 [1], 经离散化以后, 问题的数学模型是

$$\boxed{\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \ \Phi(x, y) := \theta_1(x) - y^T A x - \theta_2(y) \qquad (4.2)}$$

其中$\mathcal{X} \subset \Re^n$, $\mathcal{Y} \subset \Re^m$ 是闭凸集, $A \in \Re^{m \times n}$. $\theta_1(x) : \Re^n \to \Re$, $\theta_2(y) : \Re^m \to \Re$ 为凸函数. 如果 $(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}$ 是问题 (4.2) 的解, 则有

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

换句话说, (x^*, y^*) 是 $\Phi(x, y)$ 在 $\mathcal{X} \times \mathcal{Y}$ 上的鞍点. 因此, 问题 (4.2) 能够转换成等价的变分不等式: 求 $(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}$, 使得

$$\begin{pmatrix} x - x^* \\ y - y^* \end{pmatrix}^T \begin{pmatrix} f(x^*) - A^T y^* \\ g(y^*) + A x^* \end{pmatrix} \geq 0, \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}, \qquad (4.3)$$

其中 $f(x) \in \partial\theta_1(x), g(y) \in \partial\theta_2(y)$. 用记号

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

问题 (4.3) 就是变分不等式

$$u^* \in \Omega, \qquad (u - u^*)^T F(u^*) \geq 0, \quad \forall\, u \in \Omega. \tag{4.4}$$

由于 $\theta_1(x)$ 和 $\theta_2(y)$ 是凸函数, 容易验证, 这里的 $F(u)$ 是单调的. 关心这类问题求解方法的可以参阅第六讲与第七讲. 当 $\theta_1(x), \theta_2(y)$ 是非光滑凸函数时, 问题 (4.2) 也可以表述成混合变分不等式问题:

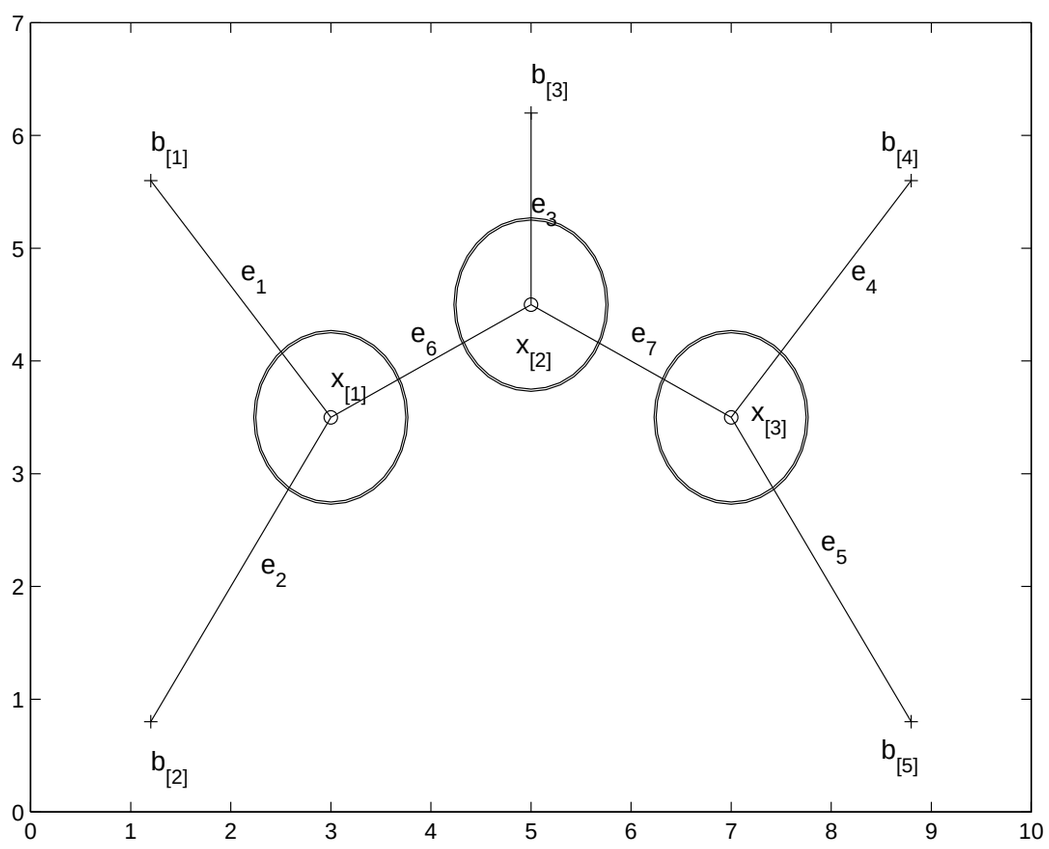
$$\theta(u) - \theta(u^*) + (u - u^*)^T M u^* \geq 0, \quad \forall\, u \in \Omega, \tag{4.5}$$

其中 $\theta(u) = \theta_1(x) + \theta_2(y)$, $M = \begin{pmatrix} 0 & -A^T \\ A & 0 \end{pmatrix}$ 是反对称矩阵. 换句话说, 混合变分不等式(4.5) 是单调的.

5 最短距离和问题之变分不等式

有些典型的非光滑凸优化问题可以化成结构相当简单的变分不等式. 这一节我们以最短距离和问题为例加以说明.

假设 $b_{[1]}, \ldots, b_{[5]}$ 是确定的村镇. 现在要用一个如右图的网络把它们连接起来. $x_{[1]}, x_{[2]}, x_{[3]}$ 是待选的连接点, 它们必须分别在划定的范围 X_1, X_2, X_3 内. 如何确定 $x_{[1]}, x_{[2]}, x_{[3]}$ 的位置, 使得网络线路长度最短.



我们在 p-模意义下求上述网络的最短距离. 该问题的数学模型是

$$\min_{x_{[j]}\in X_j}\left\{\begin{array}{lll} \|x_{[1]}-b_{[1]}\|_p & +\|x_{[1]}-b_{[2]}\|_p & +\|x_{[2]}-b_{[3]}\|_p \\ & +\|x_{[3]}-b_{[4]}\|_p & +\|x_{[3]}-b_{[5]}\|_p \\ & +\|x_{[1]}-x_{[2]}\|_p & +\|x_{[2]}-x_{[3]}\|_p \end{array}\right\}. \tag{5.1}$$

主要对 $p=1,2,\infty$ 感兴趣. 注意到这里是距离和问题, 不是距离的平方和问题. 此类问题是一个非光滑凸优化问题.

5.1 欧氏模下的最短距离和问题

将问题转化为 min-max 问题. 注意到对任意的 $d\in\Re^2$, 有

$$\|d\|_2=\max_{\xi\in B_2}\xi^T d, \tag{5.2}$$

其中

$$B_2=\{\xi\in\Re^2\mid\|\xi\|_2\le 1\}.$$

利用 (5.2), 上述欧氏模意义下的最短距离和问题可以化为 min-max 问题

$$\min_{x_{[i]}\in X_i}\ \max_{z_{[j]}\in B_2}\left\{\begin{array}{r} z_{[1]}^T(x_{[1]}-b_{[1]})+z_{[2]}^T(x_{[1]}-b_{[2]})+\ z_{[3]}^T(x_{[2]}-b_{[3]}) \\ +\ z_{[4]}^T(x_{[3]}-b_{[4]})+\ z_{[5]}^T(x_{[3]}-b_{[5]}) \\ +\ z_{[6]}^T(x_{[1]}-x_{[2]})+\ z_{[7]}^T(x_{[2]}-x_{[3]}) \end{array}\right\}$$

它的紧凑形式为

$$\boxed{\min_{x\in\mathcal{X}}\ \max_{z\in\mathcal{B}_2}\ \ z^T(Ax-b)}$$

其中

$$\mathcal{X}=X_1\times X_2\times X_3\ ,\qquad \mathcal{B}_2=B_2\times B_2\times\cdots\times B_2.$$

$$x=\begin{pmatrix} x_{[1]}\\ x_{[2]}\\ x_{[3]}\end{pmatrix},\quad z=\begin{pmatrix} z_{[1]}\\ z_{[2]}\\ \vdots\\ z_{[7]}\end{pmatrix}. \tag{5.3}$$

分块矩阵 A 和向量 b 的结构分别是

$$A = \begin{pmatrix} I_2 & 0 & 0 \\ I_2 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_2 \\ 0 & 0 & I_2 \\ I_2 & -I_2 & 0 \\ 0 & I_2 & -I_2 \end{pmatrix}, \qquad b = \begin{pmatrix} b_{[1]} \\ b_{[2]} \\ b_{[3]} \\ b_{[4]} \\ b_{[5]} \\ 0 \\ 0 \end{pmatrix}.$$

在这个 min-max 问题中, 每个 $x_{[i]}$ 对应一个点, 而每个 $z_{[j]}$ 则对应一条边.

设 $(x^*, z^*) \in \mathcal{X} \times \mathcal{B}_2$ 是 min-max 问题的解, 则对所有的 $x \in \mathcal{X}$ 和 $z \in \mathcal{B}$, 有

$$\boxed{z^T(Ax^* - b) \le z^{*T}(Ax^* - b) \le z^{*T}(Ax - b)}$$

它的等价形式是下面的线性变分不等式:

$$\boxed{x^* \in \mathcal{X},\ z^* \in \mathcal{B}_2, \quad \begin{cases} (x - x^*)^T(A^T z^*) \ge 0, & \forall x \in \mathcal{X}, \\ (z - z^*)^T(-Ax^* + b) \ge 0, & \forall z \in \mathcal{B}_2. \end{cases}}$$

可以写成更简单紧凑的形式

$$\boxed{u^* \in \Omega, \quad (u - u^*)^T(Mu^* + q) \ge 0, \quad \forall u \in \Omega}$$

其中有

$$u = \begin{pmatrix} x \\ z \end{pmatrix}, \quad M = \begin{pmatrix} 0 & A^T \\ -A & 0 \end{pmatrix}, \quad q = \begin{pmatrix} 0 \\ b \end{pmatrix}$$

和

$$\Omega = \mathcal{X} \times \mathcal{B}_2.$$

5.2 l_1-模下的最短距离和问题

由于对任意的 $d \in \Re^2$, 有

$$\|d\|_1 = \max_{\xi \in B_\infty} \xi^T d,$$

其中

$$B_\infty = \{\xi \in R^2 \mid \|\xi\|_\infty \le 1\}.$$

与欧氏模下的最短距离和问题一样, 问题 (5.1) 在 l_1-模意义下也可以表示成一个 min-max 问题

$$\min_{x_{[i]} \in X_i} \max_{z_{[j]} \in B_\infty} \left\{ \begin{array}{r} z_{[1]}^T(x_{[1]} - b_{[1]}) + z_{[2]}^T(x_{[1]} - b_{[2]}) + \ z_{[3]}^T(x_{[2]} - b_{[3]}) \\ + \ z_{[4]}^T(x_{[3]} - b_{[4]}) + \ z_{[5]}^T(x_{[3]} - b_{[5]}) \\ + \ z_{[6]}^T(x_{[1]} - x_{[2]}) + \ z_{[7]}^T(x_{[2]} - x_{[3]}) \end{array} \right\}$$

它的紧凑形式为

$$\boxed{\min_{x \in \mathcal{X}} \max_{z \in \mathcal{B}_\infty} \quad z^T(Ax - b)}$$

与欧氏模下的距离和问题一样, 它等价于变分不等式

$$\boxed{u^* \in \Omega, \quad (u - u^*)^T(Mu^* + q) \ge 0, \quad \forall u \in \Omega}$$

矩阵 M 和向量 q 都不变, 所不同的只是这时集合

$$\Omega = \mathcal{X} \times \mathcal{B}_\infty, \qquad \mathcal{B}_\infty = B_\infty \times B_\infty \times \cdots \times B_\infty.$$

5.3 l_∞-模下的最短距离和问题

由于对任意的 $d \in \Re^2$, 有

$$\|d\|_\infty = \max_{\xi \in B_1} \xi^T d,$$

其中

$$B_1 = \{\xi \in R^2 \mid \|\xi\|_1 \le 1\}.$$

与欧氏模下的最短距离和问题一样, 问题 (5.1) 在 l_∞-模意义下也可以表示成

一个 min-max 问题

$$\min_{x_{[i]}\in X_i}\ \max_{z_{[j]}\in B_1}\left\{\begin{array}{r} z_{[1]}^T(x_{[1]}-b_{[1]})+z_{[2]}^T(x_{[1]}-b_{[2]})+\ z_{[3]}^T(x_{[2]}-b_{[3]}) \\ +\ z_{[4]}^T(x_{[3]}-b_{[4]})+\ z_{[5]}^T(x_{[3]}-b_{[5]}) \\ +\ z_{[6]}^T(x_{[1]}-x_{[2]})+\ z_{[7]}^T(x_{[2]}-x_{[3]}) \end{array}\right\}$$

并化成等价的变分不等式

$$\boxed{u^*\in\Omega,\quad (u-u^*)^T(Mu^*+q)\ge 0,\quad \forall u\in\Omega}$$

矩阵 M 和向量 q 都不变, 所不同的只是这时集合

$$\Omega=\mathcal{X}\times\mathcal{B}_1,\qquad \mathcal{B}_1=B_1\times B_1\times\cdots\times B_1.$$

将处理欧氏模问题时的 $\mathcal{B}_2$ 换成了$\mathcal{B}_1$. 我们会在第三讲具体介绍这类最短距离和问题的求解方法.

6 极小化最大特征值之变分不等式

设 $A_i, i=1,\ldots,m$ 和 B 是给定的 $n\times n$ 对称矩阵, $c\in\Re^m$. 考虑问题

$$\boxed{\min_{x\in\mathcal{X}}\{\max\lambda_{\max}[\mathcal{A}(x)+B]+c^Tx\}\qquad (6.1)}$$

其中 $\mathcal{X}\subset\Re^m$,

$$\mathcal{A}(x)=\textstyle\sum_{i=1}^m x_iA_i. \qquad (6.2)$$

这是经典半定规划中的对偶问题 [10]. 对给定的 $n\times n$ 矩阵 $A=(a_{ij})$, 矩阵 A 的对角元的和称为矩阵 A 的迹, 记为 $\mathsf{Tr}(A)$, 即有

$$\mathsf{Tr}(A)=\textstyle\sum_{j=1}^n a_{jj}.$$

熟知, 矩阵的特征值的和等于矩阵的迹. 对由 (6.2) 定义的 $\mathcal{A}(x)$ 和给定的 $n\times n$ 矩阵 Y, 我们用

$$\mathcal{A}^*(Y)=\begin{pmatrix}\mathsf{Tr}(A_1Y)\\ \vdots\\ \mathsf{Tr}(A_mY)\end{pmatrix}$$

定义 $\mathcal{A}^*(Y) \in \Re^m$. 在一些应用问题中, 集合 $\mathcal{X}$ 往往是个单纯形, 即 $\mathcal{X} = \{x \in \Re^m \mid \sum_{i=1}^m = 1,\ x \geq 0\}$. 注意到相似变换不改变矩阵特征值这一事实, 可以得到

$$\lambda_{\max}(A) = \max\{\mathsf{Tr}(YA) \mid \mathsf{Tr}(Y) = 1,\ Y \succeq 0\}.$$

所以, 问题 (6.1) 等价于下面的 $\min - \max$ 问题:

$$\min_{x \in \mathcal{X}} \max_{Y \succeq 0} \{\sum_{i=1}^m \big(x_i \mathsf{Tr}(YA_i)\big) + \mathsf{Tr}(YB) + c^T x \mid \mathsf{Tr}(Y) = 1\}. \tag{6.3}$$

设 (x^*, Y^*) 是问题 (6.3) 的解, 那么有 $\lambda^* \in \Re$, 使得

$$\Phi_{y \in \mathcal{S}_+^n, \lambda \in \Re}(x^*, Y, \lambda) \leq \Phi(x^*, Y^*, \lambda^*) \leq \Phi_{x \in \mathcal{X}}(x, Y^*, \lambda^*),$$

其中

$$\Phi(x, Y, \lambda) = \sum_{i=1}^m \big(x_i \mathsf{Tr}(YA_i)\big) + \mathsf{Tr}(YB) + c^T x + \lambda(\mathsf{Tr}(Y) - 1),$$

是问题 (6.3) 的 Lagrange 函数. 问题就转化为求 函数 $\Phi(x, Y, \lambda)$ 在 $\mathcal{X} \times (\mathcal{S}_+^n \times \Re)$ 上的鞍点 (x^*, Y^*, λ^*). 因此, 问题 (6.1) 可以转换成下面的变分

不等式: 求 $(x^*, Y^*, \lambda^*) \in \mathcal{X} \times \mathcal{S}_+^n \times \Re$, 使得

$$\begin{cases} x^* \in \mathcal{X}, & (x - x^*)^T\{\mathcal{A}^*(Y^*) + c\} \geq 0, & \forall x \in \mathcal{X}, \\ Y^* \succeq 0, & \langle Y - Y^*, -\mathcal{A}(x^*) - \lambda^* I - B\rangle \geq 0, & \forall Y \succeq 0, \\ \lambda^* \in \Re, & (\lambda - \lambda^*)(\mathsf{Tr}(Y^*) - 1) \geq 0, & \forall \lambda \in \Re. \end{cases}$$

上述变分不等式的紧致形式就是

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

其中

$$u = \begin{pmatrix} x \\ Y \\ \lambda \end{pmatrix}, \quad F(u) = \begin{pmatrix} \mathcal{A}^*(Y) + c \\ -\mathcal{A}(x) - \lambda I - B \\ \mathsf{Tr}(Y) - 1 \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{S}_+^n \times \Re.$$

变分不等式是描述数学模型的强有力的统一工具

当然, 这并不是说, 上面提及的问题都要化成变分不等式去求解. 从后面的篇章能清楚看到, 变分不等式的框架有利于设计算法和分析算法的收敛性质.

岩土力学中的许多工程问题可以归结为不同类型的变分不等式

岩土工程中的力学问题, 归结成的微分方程, 离散化以后得到的非线性变分不等式, 往往具备我们所说的单调性条件 [16]. 中科院武汉岩土力学研究所的科技工作者, 用 [4, 5] 中的方法解决了长期困扰岩土力学界的一些问题 [11, 12].

References

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

[2] P.T. Harker, J.S.Pang, Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications, *Mathematical Programming*, **48**, 161-220, 1990.

[3] 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.

[4] B.S. He, A class of projection and contraction methods for monotone variational inequalities, Applied Mathematics and Optimization, **35**, 69-76, 1997.

[5] B.S He and L-Z Liao, Improvements of some projection methods for monotone nonlinear variational inequalities, Journal of Optimization Theory and Applications, **112**, 111-128, 2002

[6] B.S. He, X.M. Yuan and J.J.Z. Zhang, Comparison of two kinds of prediction-correction methods for monotone variational inequalities, Computational Optimization and Applications, **27**, 247-267, 2004

[7] B. S. He and M.-H. Xu, A general framework of contraction methods for monotone variational inequalities, Pacific J. Optimization, **4**, 195-212, 2008.

[8] B.S. He, W. Xu , H. Yang, and X.M. Yuan, Solving resource protection and supply guarantee problems in economic equilibrium, Netw. Spat. Econ., 11,. 127-138, 2011.

[9] J. Stoer and C. Witzgall, *Convexity and Optimization in Finite Dimension I*. Springer-Verlag, 1970.

[10] L. Vandenberghe and S. Boyd, Semidefinite programming, SIAM Review, **38**, 49-95, 1995.

[11] Hong Zheng, Feng Liu and Xiuli Du, Complementarity problem arising from static growth of multiple cracks and MLS-based numerical manifold method, Computer Methods in Applied Mechanics and Engineering, 295 (2015) 150-171.

[12] Hong Zheng, Peng Zhang and Xiuli Du, Dual form of discontinuous deformation analysis, Computer Methods in Applied Mechanics and Engineering, 305 (2016) 196-216.

[13] 何炳生, 论求解单调变分不等式的一些投影收缩算法, 〖计算数学〗, **18**, 54-60, 1996.

[14] 何炳生，邵虎，徐明华, 大桥流量分配调控中的隐式互补问题. 中国科技论文在线优秀论文集, 第三辑, 1-10, 2006.

[15] 何炳生, 徐薇, 杨海, 袁晓明, 经济平衡中的一类保护资源何保障供给问题, 中国科技论文在线优秀论文集, 第三辑, 11一19, 2006.

[16] 郑宏, 葛修润, 节理岩体的变分不等式模型, 中国科学(B辑) 25 卷, 第9期，986–993，1995.

凸优化和单调变分不等式的收缩算法

第二讲：三个基本不等式和变分不等式的投影收缩算法

Three fundamental inequalities and the projection and contraction methods for variational inequalities

南京大学数学系　何炳生
hebma@nju.edu.cn

The context of this lecture is mainly based on the publicatios [4, 6]

1 Basic properties of Projection Mapping and Variational Inequality

Let $\Omega \subset \Re^n$ be a closed convex set, F be a mapping from $\Re^n$ to itself . We investigate the solution methods for monotone variational inequality

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

We say the variational inequalit VI(Ω, F) is monotone, when its operator F is monotone. In other words, it satisfies

$$(u-v)^T(F(u)-F(v)) \geq 0, \quad \forall\, u,\, v \in \Re^n (\text{or } \Omega).$$

Let M be $n \times n$ matrix, $q \in \Re^n$. $F(u) = Mu + q$ is an affine operator. In this note, we say M is positive semidefinite, when

$$u^T M u \geq 0, \quad \forall u \in \Re^n.$$

Here, M is not necessarily symmetric. In other words, we say M is positive semidefinite, when the symmetric matrix $M^T + M$ is positive semidefinite.

Especially, when M is skew-symmetric, *i. e.*, $M^T = -M$, then $u^T M u \equiv 0$. In this case, the affine operator $F(u) = Mu + q$ is monotone.

Let $\Omega \subset \Re^n$ be a convex closed set and f be a convex function on Ω. Assume that f is differentiable on a open set that contains Ω. Then f is convex if and only if

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

This assertion can be found in

- R. Fletcher, Practical Methods of Optimization, Second Edition, §9.4. pp. 214–215, John Wiley & Sons, 1987.

Exchange the positions of x and y in (1.2), we have

$$f(x) - f(y) \geq \nabla f(y)^T (x - y), \quad \forall x, y \in \Omega. \tag{1.3}$$

Adding (1.2) and (1.3), we get

$$(y - x)^T (\nabla f(y) - \nabla f(x)) \geq 0, \quad \forall x, y \in \Omega. \tag{1.4}$$

Thus, the gradient operator of the differentiable convex function is monotone.

1.1 Basic properties of the projection mapping

We use $P_\Omega(\cdot)$ to denote the projection on Ω, *i. e.* ,

$$P_\Omega(v) = \arg\min\{\|u - v\| \mid u \in \Omega\}.$$

An equivalent expression is $P_\Omega(v) = \arg\min\{\frac{1}{2}\|u - v\|^2 \mid u \in \Omega\}$. When $\Omega = \Re^n_+$ (the nonnegative orthant in $\Re^n$), each element of $P_\Omega(v)$ is given by

$$(P_\Omega(v))_j = \begin{cases} v_j, & \text{if} \quad v_j \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

When $\Omega = B(c, r) = \{c + ru | \|u\| \leq 1\}$, a ball in $\Re^n$ with radius r centered on c, then

$$P_\Omega(v) = \begin{cases} \frac{r(v-c)}{\|v-c\|} + c, & \text{if} \quad \|v - c\| \geq r; \\ v, & \text{otherwise.} \end{cases}$$

The unit ball r in l_∞ and l_1 norm centered on the origin are denoted by

$$B_\infty = \{u \in \Re^n | \, \|u\|_\infty \leq 1\} \quad \text{and} \quad B_1 = \{u \in \Re^n | \, \|u\|_1 \leq 1\},$$

respectively. The projection on B_∞ and B_1 are shown in the following figures:

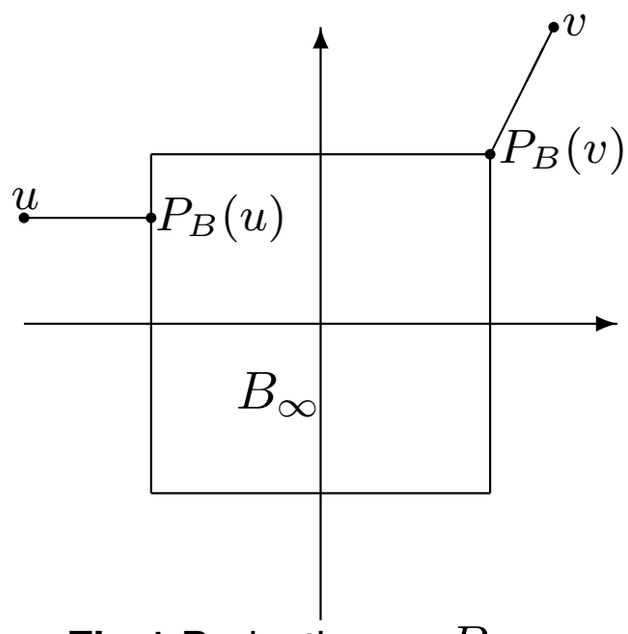


Fig.1 Projection on B_∞

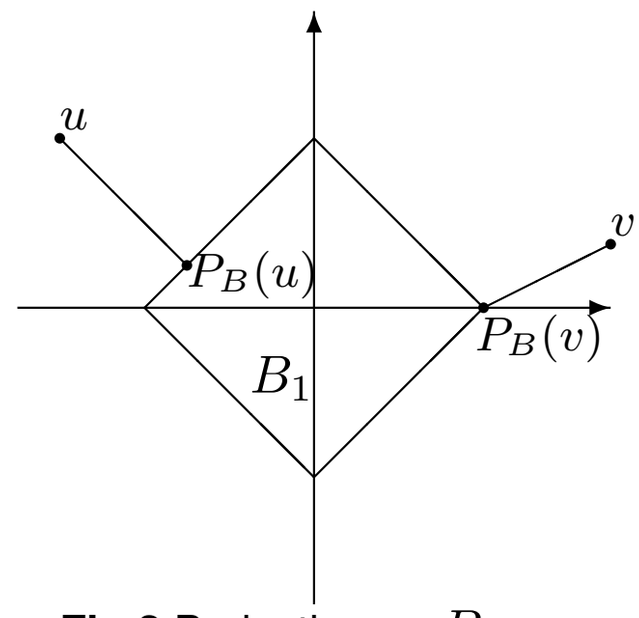


Fig.2 Projection on B_1

The following lemma illustrates an important property of the projection mapping.

Lemma 1.1 *Let $\Omega \subset \Re^n$ be a closed convex set and $P_\Omega(\cdot)$ be the projection on Ω. It holds that*

$$(v - P_\Omega(v))^T (u - P_\Omega(v)) \leq 0 \qquad \forall v \in \Re^n, \forall u \in \Omega. \tag{1.5}$$

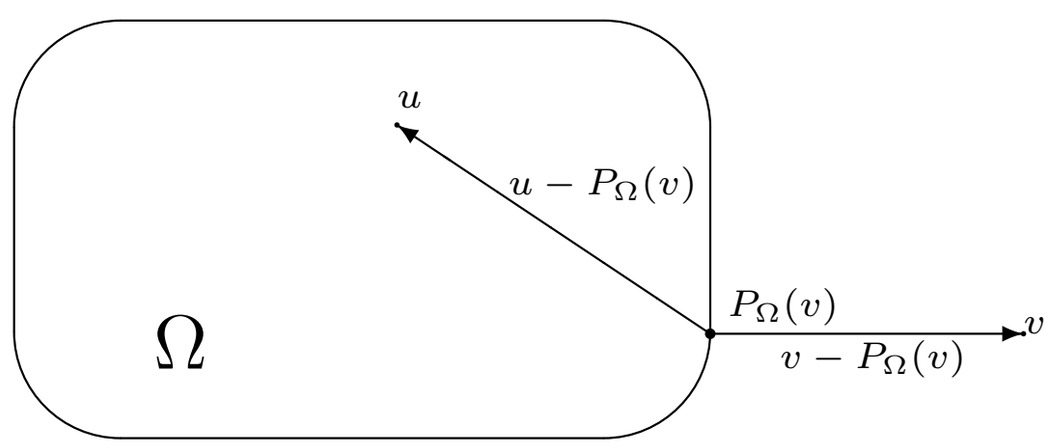


Fig.3 Geometric interpretation of the inequality (1.5)

Proof. First, since $P_\Omega(v) = \arg\min\{\|u - v\| \mid u \in \Omega\}$, we have

$$\|v - P_\Omega(v)\| \leq \|v - w\|, \quad \forall w \in \Omega. \tag{1.6}$$

Because $P_\Omega(v) \in \Omega$ and $\Omega \subset \Re^n$ is closed and convex, then for any $u \in \Omega$ and $\theta \in (0, 1)$, it holds that

$$w := \theta u + (1 - \theta) P_\Omega(v) = P_\Omega(v) + \theta(u - P_\Omega(v)) \in \Omega.$$

For this w, by using (1.6), it follows that

$$\|v - P_\Omega(v)\|^2 \leq \|v - P_\Omega(v) - \theta(u - P_\Omega(v))\|^2.$$

Expanding the last inequality, for any $u \in \Omega$ and $\theta \in (0,1)$, we have

$$[v - P_\Omega(v)]^T[u - P_\Omega(v)] \le \frac{\theta}{2}\|u - P_\Omega(v)\|^2.$$

Let $\theta \to 0_+$, we get the assertion (1.5). $\square$

In the analysis of the projection and contraction methods, the inequality (1.5) is important and useful. We call it as the **Tool inequality of the projection operator**. By using (1.5), it is easy to prove the following lemma.

Lemma 1.2 *Let $\Omega \subset \Re^n$ be a closed convex set, we have*

$$\|P_\Omega(v) - P_\Omega(u)\| \le \|v - u\|, \qquad \forall u, v \in \Re^n. \tag{1.7}$$

$$\|P_\Omega(v) - u\| \le \|v - u\|, \qquad \forall v \in \Re^n, u \in \Omega. \tag{1.8}$$

$$\|P_\Omega(v) - u\|^2 \le \|v - u\|^2 - \|v - P_\Omega(v)\|^2, \ \forall v \in \Re^n, u \in \Omega. \tag{1.9}$$

We leave the proofs to the reader.

1.2 The equivalent projection equation of the variational inequality

Assume that the solution set of (1.1), denoted by Ω^*, is nonempty. The solution set of monotone variational inequality is convex, its proof is involved in Theorem 2.3.5 [1]. We use u^* to denote any fixed point in Ω^*. For any scalar $\beta > 0$, the following statement is true.

$$\boxed{u \in \Omega^* \quad \Leftrightarrow \quad u = P_\Omega[u - \beta F(u)].}$$

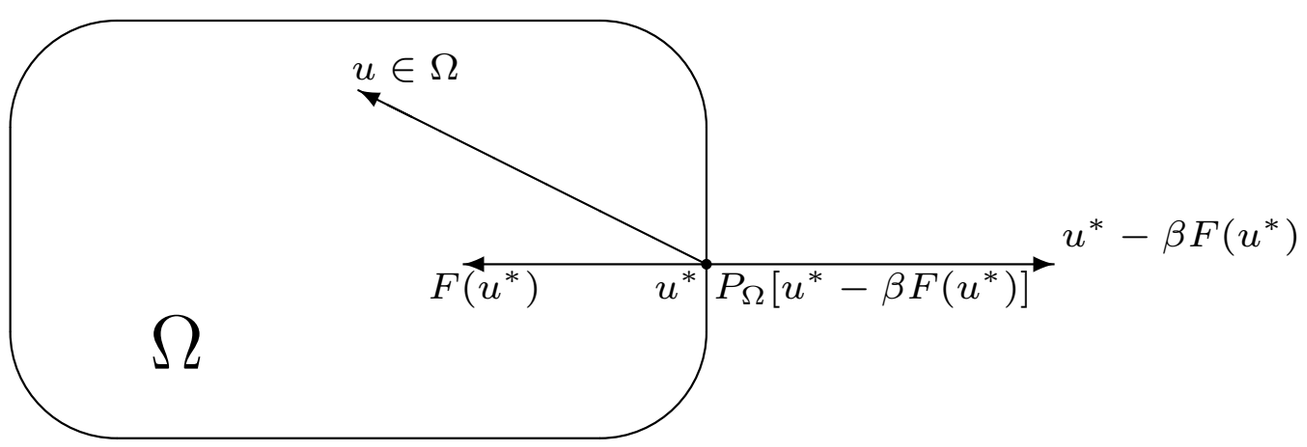


Fig.4 Geometric interpretation of $u^* \in \Omega^* \Leftrightarrow u^* = P_\Omega[u^* - \beta F(u^*)]$

In other words, solving the variational inequality (1.1) is equivalent to finding a zero point of $e(u, \beta)$,

$$e(u, \beta) := u - P_\Omega[u - \beta F(u)] \tag{1.10}$$

The proof will be given in Theorem 1.1. Thus, for a given $\beta > 0$, $\|e(u, \beta)\|$ can be viewed as the error which measure how much u fails to be a solution point.

Theorem 1.1 *For given $\beta > 0$, u^* is a solution point od VI(Ω, F) if and only if $e(u^*, \beta) = 0$.*

Proof. " $\Rightarrow$ " Let u^* be a solution point of VI(Ω, F), we show $e(u^*, \beta) = 0$. Since $u^* \in \Omega$. By using the tool inequality (1.5), we get

$$\big(v - P_\Omega(v)\big)^T \big(u^* - P_\Omega(v)\big) \le 0, \quad \forall v \in \Re^n.$$

Setting $v = u^* - \beta F(u^*)$ in the above inequality and using the notation of $e(u, \beta)$, it follows from the last inequality that $(e(u^*, \beta) - \beta F(u^*))^T e(u^*, \beta) \le 0$, and thus

$$\|e(u^*, \beta)\|^2 \le \beta e(u^*, \beta)^T F(u^*). \tag{1.11}$$

On the other hand, because $P_\Omega[u^* - \beta F(u^*)] \in \Omega$ and u^* is a solution of the VI, it

follows from (1.1) that

$$\{P_\Omega[u^* - \beta F(u^*)] - u^*\}^T F(u^*) \ge 0,$$

and thus

$$e(u^*, \beta)^T F(u^*) \le 0. \tag{1.12}$$

From (1.11) and (1.12), we get $e(u^*, \beta) = 0$.

" $\Leftarrow$ ". If $e(u^*, \beta) = 0$, we show $u^* \in \Omega^*$. Taking $v = u^* - \beta F(u^*)$ in (1.5) and using the notation $e(u^*, \beta)$, we have

$$\{e(u^*, \beta) - \beta F(u^*)\}^T \{u - P_\Omega[u^* - \beta F(u^*)]\} \le 0, \quad \forall u \in \Omega. \tag{1.13}$$

Since $e(u^*, \beta) = 0$, $u^* = P_\Omega(\cdot) \in \Omega$ and $P_\Omega[u^* - \beta F(u^*)] = u^*$. Substituting them in (1.13), we get

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

Thus u^* is a solution of VI(Ω, F). The proof is complete. $\square$

The following theorem tells us, for any fixed u, $\|e(u, \beta)\|$ is a non-decreasing function of β, and $\{\|e(u, \beta)\|/\beta\}$ is a non-increasing function of β. The following proof can be found in [14]. The tool inequality (1.5) plays the key role.

Theorem 1.2 *For any given $u \in \Re^n$ and $\tilde{\beta} \geq \beta > 0$, we have*

$$\|e(u, \tilde{\beta})\| \geq \|e(u, \beta)\| \tag{1.14}$$

and

$$\frac{\|e(u, \tilde{\beta})\|}{\tilde{\beta}} \leq \frac{\|e(u, \beta)\|}{\beta}. \tag{1.15}$$

Proof. Let $t = \|e(x, \tilde{\beta})\| / \|e(x, \beta)\|$, the assertions of this theorem is equivalent to

$$1 \leq t \leq \tilde{\beta}/\beta.$$

In other words, t is the solution of the quadratic inequality

$$(t-1)(t - \tilde{\beta}/\beta) \leq 0. \tag{1.16}$$

First, since $P_\Omega(w) \in \Omega$, according to the tool inequality (1.5), we have

$$(v - P_\Omega(v))^T (P_\Omega(v) - P_\Omega(w)) \geq 0, \quad \forall\, v \in \Re^n. \tag{1.17}$$

Setting $v := u - \beta F(u)$ and $w := P_\Omega[u - \tilde{\beta} F(u)]$ in (1.17), using the definition of $e(u, \beta)$, it follows that $(v - P_\Omega(v)) = e(u, \beta) - \beta F(u)$ and

$$P_\Omega(v) - P_\Omega(w) = P_\Omega[u - \beta F(u)] - P_\Omega[u - \tilde{\beta} F(u)] = e(u, \tilde{\beta}) - e(u, \beta).$$

Substituting it in (1.17),

$$\{e(u, \beta) - \beta F(u)\}^T \{e(u, \tilde{\beta}) - e(u, \beta)\} \geq 0. \tag{1.18}$$

Change the position of β and $\tilde{\beta}$ in (1.18), we get

$$\{e(u, \tilde{\beta}) - \tilde{\beta} F(u)\}^T \{e(u, \beta) - e(u, \tilde{\beta})\} \geq 0. \tag{1.19}$$

Multiplying (1.18) and (1.19) by $\tilde{\beta}$ and β, respectively, and then adding them up, we have

$$\{\tilde{\beta} e(u, \beta) - \beta e(u, \tilde{\beta})\}^T \{e(u, \tilde{\beta}) - e(u, \beta)\} \geq 0,$$

and thus

$$\beta \|e(x, \tilde{\beta})\|^2 - (\beta + \tilde{\beta}) e(x, \beta)^T e(x, \tilde{\beta}) + \tilde{\beta} \|e(x, \beta)\|^2 \leq 0. \tag{1.20}$$

Using Cauchy-Schwarz inequality to (1.20),

$$\beta \|e(x, \tilde{\beta})\|^2 - (\beta + \tilde{\beta}) \|e(x, \beta)\| \cdot \|e(x, \tilde{\beta})\| + \tilde{\beta} \|e(x, \beta)\|^2 \leq 0. \tag{1.21}$$

Dividing (1.21) by $\beta \|e(x, \beta)\|^2$, and using $t = \|e(x, \tilde{\beta})\| / \|e(x, \beta)\|$,

$$t^2 - \Big(1 + \frac{\tilde{\beta}}{\beta}\Big) t + \frac{\tilde{\beta}}{\beta} \leq 0.$$

Thus, the inequality (1.16) is true, and the proof of this theorem is complete. $\square$

Although Theorem 1.1 indicates that, for any $\beta > 0$, u is a solution of (1.1) if and only if $e(u, \beta) = 0$. Theorem 1.2 tell us, if we use $\|e(u, \beta)\|$ as the error measure, the constant parameter $\beta > 0$ should not be too large or too small. Generally, it should be considered in combination with the physical significance of the considered problem.

2 Three fundamental inequalities and Projection and Contraction Methods

Three fundamental inequalities Let u^* be any given solution of the monotone variational inequality (1.1). Because $\tilde{u} = P_\Omega[u - \beta F(u)] \in \Omega$, according to the definition of VI, we have

$$\text{(FI1)} \quad (\tilde{u} - u^*)^T \beta F(u^*) \geq 0.$$

Setting $v = u - \beta F(u)$ in (1.5), since $\tilde{u} = P_\Omega[u - \beta F(u)] = P_\Omega[v]$ and $u^* \in \Omega$, according to the tool inequality (1.5), we have

$$\text{(FI2)} \quad (\tilde{u} - u^*)^T \big([u - \beta F(u)] - \tilde{u}\big) \geq 0.$$

The variational inequalities considered in this Lecture Series is monotone, according to the monotonicity of the operator F, we have

$$\text{(FI3)} \quad (\tilde{u} - u^*)^T (\beta F(\tilde{u}) - \beta F(u^*)) \geq 0.$$

The search directions of some projection contraction algorithms are derived from these basic (but fundamental) inequalities.

The basic framework of the projection and contraction methods

Projection and Contraction Methods is a kind of contraction algorithms based on projection. For given $\beta > 0$ and he current point u^k, we get a predictor $\tilde{u}^k$ by making the projection $\tilde{u}^k = P_\Omega[u^k - \beta F(u^k)]$. According to Theorem 1.1,

u^k is a solution of (1.1) if and only if $u^k = \tilde{u}^k$.

Error measure function A nonnegative function $\varphi(u^k, \tilde{u}^k)$ is called the error

measure function of VI(Ω, F) (1.1), if there is a $\delta > 0$, such that

$$\varphi(u^k, \tilde{u}^k) \geq \delta \|u^k - \tilde{u}^k\|^2 \quad \text{and} \quad \varphi(u^k, \tilde{u}^k) = 0 \Leftrightarrow u^k = \tilde{u}^k. \tag{2.1a}$$

Profitable direction A vector $d(u^k, \tilde{u}^k)$ $\big(\|d(u^k, \tilde{u}^k)\| = \mathcal{O}\|u^k - \tilde{u}^k\|\big)$ is called a profitable direction associated with $\varphi(u^k, \tilde{u}^k)$, if

$$(u^k - u^*)^T d(u^k, \tilde{u}^k) \geq \varphi(u^k, \tilde{u}^k), \ \forall\, u^* \in \Omega^*. \tag{2.1b}$$

- Although u^* is unknown, for any fixed u^*, $(u^k - u^*)$ is the gradient of the distance function $\frac{1}{2}\|u - u^*\|^2$ at u^k. Thus, we call $\varphi(u, \tilde{u})$ and $d(u, \tilde{u})$ "the error measure function" and "the profitable direction", respectively.
- The basic idea of the projection and contraction algorithm is to construct a direction $d(u^k, \tilde{u}^k)$, such that (2.1b) is satisfied for any $u^* \in \Omega^*$. The definitions (2.1) indicate that $-d(u^k, \tilde{u}^k)$ is a descent direction of the distance function $\frac{1}{2}\|u - u^*\|^2$ at the current point u^k.
- In the contraction methods, it is required that the sequence $\{\|u^k - u^*\|^2\}$ is strictly monotone decreasing, where $\{u^k\}$ is the sequence generated by the algorithm and u^* is any fixed solution point.

Using the conditions (2.1a) and (2.1b), we give the following general algorithm.

The general algorithm When the condition (2.1b) is satisfied, let

$$u^{k+1}(\alpha) = u^k - \alpha d(u^k, \tilde{u}^k) \tag{2.2}$$

be the step-size α dependent new iterate. Now, we consider how to maximize the "profit" of the square of the distance, namely

$$\vartheta_k(\alpha) = \|u^k - u^*\|^2 - \|u^{k+1}(\alpha) - u^*\|^2. \tag{2.3}$$

Notice that

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

For any given u^*, the last equation indicates that $\vartheta_k(\alpha)$ is a quadratic function of α. Because u^* is unknown, we can not directly maximize $\vartheta_k(\alpha)$. However, using (2.1b), we have

$$\vartheta_k(\alpha) \geq 2\alpha\varphi(u^k, \tilde{u}^k) - \alpha^2 \|d(u^k, \tilde{u}^k)\|^2. \tag{2.4}$$

We define the right hand side of the last inequality as $q_k(\alpha)$, so, we get

$$q_k(\alpha) = 2\alpha\varphi_k(u^k, \tilde{u}^k) - \alpha^2\|d(u^k, \tilde{u}^k)\|^2, \tag{2.5}$$

which is a lower bound function of $\vartheta_k(\alpha)$. The quadratic function $q_k(\alpha)$ reaches its maximum at α_k^*,

$$\alpha_k^* = \frac{\varphi(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2}. \tag{2.6}$$

If we take

$$u^{k+1} = u^k - \alpha_k^* d(u^k, \tilde{u}^k), \tag{2.7}$$

it follows from (2.5) and (2.6) that

$$q_k(\alpha_k^*) = \alpha_k^*\varphi(u^k, \tilde{u}^k).$$

The new iterate generates by (2.7) is not necessarily in Ω, however, it satisfies

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \alpha_k^*\varphi(u^k, \tilde{u}^k).$$

The original intention of the contraction algorithms is to maximize the quadratic

function $\vartheta_k(\alpha)$ (see (2.3)) in each iteration, because it contains the unknown u^*, we have to maximize its lower bound function $q_k(\alpha)$.

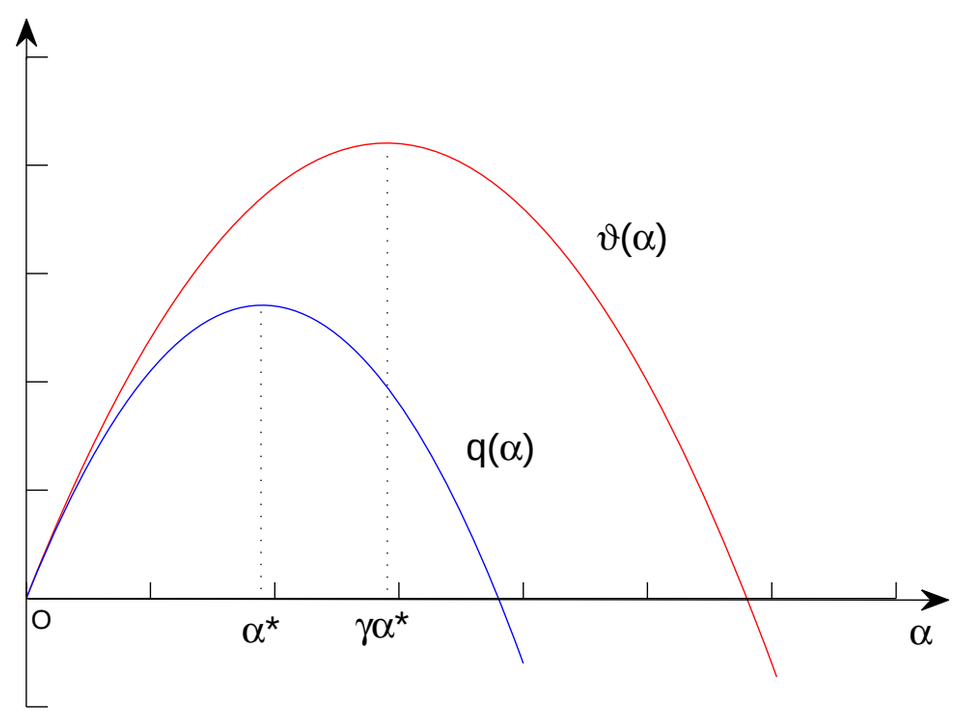


Fig.5 Interpretation of $\gamma \in [1, 2)$

Thus, in the practical computation, we take a relaxed factor $\gamma \in [1, 2)$, and set

$$u^{k+1} = u^k - \gamma\alpha_k^* d(u^k, \tilde{u}^k). \tag{2.8}$$

The reason for taking $\gamma \in [1, 2)$ in the step size is depicted in Fig. 5.

From (2.5) and (2.6), we get

$$\begin{aligned} q_k(\gamma\alpha_k^*) &= 2\gamma\alpha_k^*\varphi(u^k, \tilde{u}^k) - \gamma^2(\alpha_k^*)^2\|d(u^k, \tilde{u}^k)\|^2 \\ &= \gamma(2-\gamma)\alpha_k^*\varphi(u^k, \tilde{u}^k). \end{aligned}$$

Thus, the new iterate u^{k+1} updated by (2.8) satisfies

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \gamma(2-\gamma)\alpha_k^*\varphi(u^k, \tilde{u}^k). \tag{2.9}$$

By using (2.6) and (2.8), we have

$$\alpha_k^*\varphi(u^k, \tilde{u}^k) = \|\alpha_k^* d(u^k, \tilde{u}^k)\|^2 = \frac{1}{\gamma^2}\|u^k - u^{k+1}\|^2.$$

Substituting it in (2.9), we get the following inequality:

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \frac{2-\gamma}{\gamma}\|u^k - u^{k+1}\|^2.$$

Primary algorithm Replacing $\varphi(u^k, \tilde{u}^k) \ge \delta\|u^k - \tilde{u}^k\|^2$ (see (2.1a)) by

$$\varphi(u^k, \tilde{u}^k) \ge \frac{1}{2}\big(\|d(u^k, \tilde{u}^k)\|^2 + \tau\|u^k - \tilde{u}^k\|^2\big), \quad (\tau > 0). \tag{2.10}$$

When the conditions (2.1b) and (2.10) are satisfied, we use the simple form

$$u^{k+1} = u^k - d(u^k, \tilde{u}^k), \tag{2.11}$$

to generate the new iterate. Since it uses the unit step size, we call (2.11) as **Primary Algorithm**. By a manipulation, we get

$$\begin{aligned} \|u^{k+1} - u^*\|^2 &= \|(u^k - u^*) - d(u^k, \tilde{u}^k)\|^2 \\ &= \|u^k - u^*\|^2 - 2(u^k - u^*)^T d(u^k, \tilde{u}^k) + \|d(u^k, \tilde{u}^k)\|^2 \\ \text{(use (2.1b))} \quad &\le \|u^k - u^*\|^2 - (2\varphi(u^k, \tilde{u}^k) - \|d(u^k, \tilde{u}^k)\|^2). \end{aligned}$$

Since (2.10) is satisfied, the new iterate u^{k+1} satisfies

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \tau\|u^k - \tilde{u}^k\|^2. \tag{2.12}$$

The inequalities (2.9) and (2.12) tell us that the sequence $\{u^k\}$ is bounded, they

are key inequalities for the convergence. By using (2.1a) and Theorem 1.1, it is easy to prove the following theorem.

Theorem 2.1 *Assume that Ω^*, the solution set of the problem VI(Ω, F), is nonempty, then the sequence generated by the projection and contraction methods converges to some $u^* \in \Omega^*$.*

It should be noticed that usually we advocate to use the general algorithm to determine the new iterate by calculating the step size. Although the primary algorithm does not need to calculate the step size, according to our experience, the general algorithm converges faster than the primal one.

- The projection and contraction methods can also be regarded as a prediction-correction methods.
- The vector $\tilde{u}^k$, which is obtained by the projection, can be viewed as a predictor. It provides us the error measure function and a profitable direction.
- The updating procedure, (2.8) or (2.11), which offered u^{k+1}, can be viewed as the correction pcocess.
- Whether the projection and contraction method or prediction-correction method, their first letter is P and C, so it is called PC Methods for short.

The differentiable convex optimization problem $\min\{f(x) \mid x \in \Omega\}$ is equivalent to the variational inequality

$$x \in \Omega, \qquad (x' - x)^T \nabla f(x) \geq 0, \quad \forall x' \in \Omega. \tag{2.13}$$

If $\nabla f(x)$ is differentiable, its Hessian matrix, $\nabla^2 f(x)$ is symmetric. Especially, when $f(x)$ is a quadratic convex function, its Hessian matrix, is symmetric and positive semidefinite.

When we consider the general nonlinear monotone variational inequality VI(Ω, F), it is only required that

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

The Jacobian of F, $\nabla F(u)$, if it exist, it is not necessarily to be symmetric.

In linear variational inequality, the mapping $F(u) = Mu + q$ is affine. A linear variational inequality means that $M + M^T$ is positive semidefinite, but M is not necessarily symmetric.

Differentiable convex optimization problem is a kind of variational inequality (2.13) with special properties. We will give some better contraction algorithms for such convex optimization problems in Lecture 10.

3 PC Methods for LVI based on FI1+ FI2

In the linear variational inequality VI(Ω, F), $F(u) = Mu + q$ is an affine operator. Notice that $\tilde{u} = P_\Omega[u - \beta F(u)]$. Adding FI1 and FI2,

$$\boxed{\begin{cases} (\tilde{u} - u^*)^T \beta F(u^*) \geq 0. & \text{(FI1)} \\ (\tilde{u} - u^*)^T \big([u - \beta F(u)] - \tilde{u}\big) \geq 0. & \text{(FI2)} \end{cases}}$$

and using $F(u) = Mu + q$, we get

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

Consequently, it follows that

$$(u - u^*)^T (I + \beta M^T)(u - \tilde{u}) \geq \|u - \tilde{u}\|^2 + \beta (u - u^*)^T M (u - u^*).$$

Since $(u - u^*)^T M (u - u^*) = \frac{1}{2}(u - u^*)^T (M^T + M)(u - u^*) \geq 0$, we have

$$(u - u^*)^T (I + \beta M^T)(u - \tilde{u}) \geq \|u - \tilde{u}\|^2, \quad \forall u \in \Re^n\ u^* \in \Omega^*. \tag{3.1}$$

Let

$$\varphi(u, \tilde{u}) = \|u - \tilde{u}\|^2 \tag{3.2}$$

and

$$d(u, \tilde{u}) = (I + \beta M^T)(u - \tilde{u}). \tag{3.3}$$

The $\varphi(u, \tilde{u})$ defined by (3.2) with the $d(u, \tilde{u})$ defined by (3.3), satisfy the conditions (2.1a) and (2.1b). Especially, in (2.1a), the positive parameter $\delta = 1$.

If we use the general algorithm, the step size α_k^* is determined (2.6) and the new iterate is given by (2.8). In details, for given u^k and $\beta > 0$, let

$$\tilde{u}^k = P_\Omega[u^k - \beta(Mu^k + q)]$$

to produce a predictor. The new iterate is given by

$$u^{k+1} = u^k - \gamma\alpha_k^*(I + \beta M^T)(u^k - \tilde{u}^k)$$

where

$$\alpha_k^* = \frac{\|u^k - \tilde{u}^k\|^2}{\|(I + \beta M^T)(u^k - \tilde{u}^k)\|^2}.$$

The generated sequence $\{u^k\}$ satisfies the following contractive property

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \gamma(2-\gamma)\alpha_k^*\|u^k - \tilde{u}^k\|^2.$$

In such method, usually the parameter β is sensitive for the convergence. Thus, we suggest to adjust the parameter β_k dynamically (in every 5-10 iterations), such that

$$\beta_k\|M^T(u^k - \tilde{u}^k)\| = \mathcal{O}(\|u^k - \tilde{u}^k\|).$$

♣ If the parameter β satisfies

$$\|(I + \beta M^T)(u^k - \tilde{u}^k)\|^2 \le (2-\tau)\|u^k - \tilde{u}^k\|^2, \quad \tau \in (0,1), \qquad (3.4)$$

the $\varphi(u, \tilde{u})$ defined by (3.2) with the $d(u, \tilde{u})$ defined by (3.3) satisfy

$$2\varphi(u, \tilde{u}) \ge \|d(u, \tilde{u})\|^2 + \tau\|u - \tilde{u}\|^2.$$

This inequality tells us that the condition (2.10) is satisfied, and thus we can use the primary algorithm (2.11) (with step size 1) to update the new iterate.

According to (2.12), the generated sequence $\{u^k\}$ by the primary algorithm satisfies

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \tau\|u^k - \tilde{u}^k\|^2.$$

More projection and contraction methods for LVI based on FI1+ FI2 can be found in [4] and [5]. Besides the contraction methods in Euclidean-norm, we can also build the contraction methods in G-norm, where G is a symmetric positive definite matrix. Especially, if we take $G = (I + \beta M^T)(I + \beta M)$ and consider to reduce $\|u - u^*\|_G^2$, the update form is

$$u^{k+1} = u^k - \gamma(I + \beta M)^{-1}(u^k - \tilde{u}^k), \qquad \gamma \in (0, 2). \qquad (3.5)$$

According to this formula, we have

$$\begin{aligned}\|u^{k+1}-u^*\|_G^2 &= \|(u^k-u^*)-\gamma(I+\beta M)^{-1}(u^k-\tilde{u}^k)\|_G^2 \\ &= \|u^k-u^*\|_G^2 - 2\gamma(u^k-u^*)^T(I+\beta M^T)(u^k-\tilde{u}^k) \\ &\quad + \gamma^2\|(I+\beta M)^{-1}(u^k-\tilde{u}^k)\|_G^2.\end{aligned}$$

Using (3.1) and $G=(I+\beta M^T)(I+\beta M)$, from the last inequality follows that

$$\begin{aligned}\|u^{k+1}-u^*\|_G^2 &\le \|u^k-u^*\|_G^2 - 2\gamma\|u^k-\tilde{u}^k\|^2 + \gamma^2\|u^k-\tilde{u}^k\|^2 \\ &= \|u^k-u^*\|_G^2 - \gamma(2-\gamma)\|u^k-\tilde{u}^k\|^2.\end{aligned}$$

By using this contraction method, the update form (3.5) is equivalent to solving the following system of equations:

$$(I+\beta M)(u^{k+1}-u^k)=\gamma(\tilde{u}^k-u^k).$$

Since the matrix in each iteration is in variant, we just need to do one LU decomposition for the matrix $(I+\beta M)$ in the whole iteration process.

M is a symmetric positive definite matrix

When M is symmetric positive definite, we replacing M by H, (3.1) becomes

$$(u-u^*)^T(I+\beta H)(u-\tilde{u}) \ge \|u-\tilde{u}\|^2, \quad \forall u\in\Re^n. \tag{3.6}$$

In this case, $G=I+\beta H$ is symmetric and positive definite. We consider the contraction method in G-norm. In this case, we use

$$u^{k+1}=u^k-\gamma\alpha_k^*(u^k-\tilde{u}^k), \quad \alpha_k^*=\frac{\|u^k-\tilde{u}^k\|^2}{\|u^k-\tilde{u}^k\|_G^2}, \quad \gamma\in(0,2) \tag{3.7}$$

to produce the new iterate. By using (3.6) and (3.7), we get

$$\begin{aligned}\|u^{k+1}-u^*\|_G^2 &= \|(u^k-u^*)-\gamma\alpha_k^*(u^k-\tilde{u}^k)\|_G^2 \\ &= \|u^k-u^*\|_G^2 - 2\gamma\alpha_k^*(u^k-u^*)^TG(u^k-\tilde{u}^k) + \gamma^2(\alpha_k^*)^2\|u^k-\tilde{u}^k\|_G^2 \\ &\le \|u^k-u^*\|_G^2 - 2\gamma\alpha_k^*\|u^k-\tilde{u}^k\|^2 + \gamma^2(\alpha_k^*)^2\|u^k-\tilde{u}^k\|_G^2.\end{aligned}$$

Since $\alpha_k^*\|u^k-\tilde{u}^k\|_G^2 \overset{(3.7)}{=} \|u^k-\tilde{u}^k\|^2$, it follows from the above inequality

$$\|u^{k+1}-u^*\|_G^2 \le \|u^k-u^*\|_G^2 - \gamma(2-\gamma)\alpha_k^*\|u^k-\tilde{u}^k\|^2, \quad \forall u^*\in\Omega^*$$

which is the key inequality for the proof of the convergence.

4 PC Methods for NVI based on FI1+FI2+FI3

We consider the contraction method for the general nonlinear variational inequality (1.1). Again, for given u, $\tilde{u} = P_\Omega[u - \beta F(u)]$ is obtained by the projection and it is a predictor. Adding FI1, FI2 and FI3,

$$\boxed{\begin{cases} (\tilde{u} - u^*)^T \beta F(u^*) \geq 0 & \text{(FI1)} \\ (\tilde{u} - u^*)^T \big([u - \beta F(u)] - \tilde{u}\big) \geq 0 & \text{(FI2)} \\ (\tilde{u} - u^*)^T \big(\beta F(\tilde{u}) - \beta F(u^*)\big) \geq 0 & \text{(FI3)} \end{cases}}$$

we get

$$(\tilde{u} - u^*)^T \{(u - \tilde{u}) - \beta[F(u) - F(\tilde{u})]\} \geq 0. \tag{4.1}$$

By defining

$$\varphi(u, \tilde{u}) = (u - \tilde{u})^T d(u, \tilde{u}), \tag{4.2}$$

and

$$d(u, \tilde{u}) = (u - \tilde{u}) - \beta\big(F(u) - F(\tilde{u})\big), \tag{4.3}$$

it follows from (4.1) that

$$(u - u^*)^T d(u, \tilde{u}) \geq \varphi(u, \tilde{u}), \quad \forall\, u \in \Re^n.$$

The condition (2.1b) is satisfied. Whether $\varphi(u, \tilde{u})$ is an error measure function ? Under the assumption that F is Lipschitz continuous, for a given $\nu \in (0, 1)$, we can use **Armijo** line-search strategy to get a β such that

$$(u - \tilde{u})^T \big(\beta F(u) - \beta F(\tilde{u})\big) \leq \nu \|u - \tilde{u}\|^2, \;\; \nu \in (0, 1). \tag{4.4}$$

Usually, such β also fits the request $\beta\|F(u) - F(\tilde{u})\| = \mathcal{O}(\|u - \tilde{u}\|)$. According to (4.3) and (4.4), we have

$$\begin{aligned} \varphi(u, \tilde{u}) &= (u - \tilde{u})^T d(u, \tilde{u}) \\ &= \|u - \tilde{u}\|^2 - (u - \tilde{u})^T \beta\big(F(u) - F(\tilde{u})\big) \\ &\geq (1 - \nu)\|u - \tilde{u}\|^2. \end{aligned}$$

Indeed, $\varphi(u, \tilde{u})$ is an error measure function, when (4.4) is satisfied.

The $\varphi(u,\tilde{u})$ defined by (4.2) with the $d(u,\tilde{u})$ defined by (4.3), satisfy the conditions (2.1a) and (2.1b). In (2.1a), the positive parameter $\delta = 1-\nu$.

By using the update form (2.8) with α_k^* given by (2.6), the generated sequence $\{u^k\}$ satisfies

$$\|u^{k+1}-u^*\|^2 \le \|u^k-u^*\|^2 - \gamma(2-\gamma)\alpha_k^*(1-\nu)\|u^k-\tilde{u}^k\|^2. \quad (4.5)$$

♣ Especially, when

$$\beta\|F(u)-F(\tilde{u})\| \le \nu\|u-\tilde{u}\|, \quad (4.6)$$

the direction $d(u,\tilde{u})$ defined by (4.3) and the $\varphi(u,\tilde{u})$ defined by (4.2) satisfy

$$\begin{aligned}
&2\varphi(u,\tilde{u}) - \|d(u,\tilde{u})\|^2 \\
&\quad = \quad 2(u-\tilde{u})^T d(u,\tilde{u}) - \|d(u,\tilde{u})\|^2 \\
&\quad = \quad d(u,\tilde{u})^T\big\{2(u-\tilde{u}) - d(u,\tilde{u})\big\} \\
&\quad = \quad \big\{(u-\tilde{u}) - \beta\big(F(u)-F(\tilde{u})\big)\big\}^T\big\{(u-\tilde{u}) + \beta\big(F(u)-F(\tilde{u})\big)\big\} \\
&\quad = \quad \|u-\tilde{u}\|^2 - \beta^2\|F(u)-F(\tilde{u})\|^2 \;\ge\; (1-\nu^2)\|u-\tilde{u}\|^2. \qquad (4.7)
\end{aligned}$$

This tell us that the condition (2.10) is satisfied with $\tau = 1-\nu^2$.

Thus, under the condition (4.6), we can use the primary algorithm to update the new iterate, namely,

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

According to (2.12), we have

$$\|u^{k+1}-u^*\|^2 \le \|u^k-u^*\|^2 - (1-\nu^2)\|u^k-\tilde{u}^k\|^2.$$

The earliest PC Algorithms for LVI appears in [4, 5] which is based on adding FI1 and FI2. The late PC Methods for nonlinear variational inequality [6, 8, 12] are based on FI1+FI2+FI3 .

For nonlinear variational inequality, even though the ‘strict’ condition (4.6) is satisfied, we still advocate using the computational step size for determining the next iteration point

$$u^{k+1} = u^k - \gamma\alpha_k^* d(u^k,\tilde{u}^k),$$

where

$$\alpha_k^* = \frac{(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2}.$$

Since the (4.6) is satisfied, the left hand side of (4.7) is strictly greater than 0. Thus, for any $k > 0$, we have $\alpha_k^* > 1/2$, and

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \frac{\gamma(2-\gamma)(1-\nu)}{2}\|u^k - \tilde{u}^k\|^2. \qquad (4.8)$$

The formulation of three fundamental inequalities and related methods can be seen in [6] (the first version is the preprint 94-11, Institute of mathematics, Nanjing University) and [15]. After graduating from Nanjing University, Defeng Sun also independently found the direction (4.3) and constructed the projection contraction algorithm [12]. Thank him for citing [6] in the footnote and the references of his paper [12].

The greatest truths are the simplest ! The construction of projection contraction algorithms is based on three fundamental inequalities. Its principle is simple and unified, it gives us beautiful enjoyment !

5 The extra-gradient method

The extra-gradient method [9] (abbreviated to EG method) can be applied to solve the nonlinear monotone variational inequality. Some young scholars in the universities north America use the EG method to solve the problems arising from the area of information science, such as speech recognition, optical fiber network, machine learning and others. In their doctoral dissertation, they have mentioned the relationship between the EG method and PC Algorithms. For preparing to compare their numerical efficiency in the next section, here, we will introduce the extra-gradient method. In fact, the extra gradient algorithm is a prediction-correction form of the proximal point algorithm.

The proximal point algorithm Let us first briefly review the Proximal Point Algorithm (abbreviated to PPA) VI(Ω, F) (see (1.1)). PPA is an iterative method. For given u^k and $r > 0$, the new iterate u^{k+1} is the solution of the following variational inequality:

$$u^{k+1} \in \Omega, \quad (u - u^{k+1})^T\{F(u^{k+1}) + r(u^{k+1} - u^k)\} \ge 0, \ \forall\, u \in \Omega. \qquad (5.1)$$

It is clear that u^{k+1} is a solution of (1.1) if and only if $u^{k+1} = u^k$. In the case of $u^{k+1} \ne u^k$, by setting $u = u^*$ in (5.1), we obtain

$$(u^{k+1} - u^*)^T r(u^k - u^{k+1}) \ge (u^{k+1} - u^*)^T F(u^{k+1}). \qquad (5.2)$$

Because F is monotone, we have

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

and consequently from (5.4), we obtain

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

and thus

$$(u^k - u^*)^T (u^k - u^{k+1}) \geq \|u^k - u^{k+1}\|^2.$$

By using the last inequality, we obtain

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - \|u^k - u^{k+1}\|^2. \tag{5.3}$$

The sequence $\{u^k\}$ generated by PPA has the nice convergence property, however, the subproblem (5.1) is almost difficult as the original problem (1.1). Thus, the classical PPA is not widely used in the practical application.

By using $\beta = 1/r$ in (5.1), u^{k+1} can be viewed as

$$u^{k+1} \in \Omega, \quad (u - u^{k+1})^T \{(u^{k+1} - u^k) + \beta F(u^{k+1})\} \geq 0, \ \forall\, u \in \Omega. \tag{5.4}$$

By using the equivalent representation of VI (see Theorem 1.1), it can be written as

$$u^{k+1} = P_\Omega\big[u^{k+1} - \{(u^{k+1} - u^k) + \beta F(u^{k+1})\}\big].$$

In other words,

$$u^{k+1} = P_\Omega[u^k - \beta F(u^{k+1})]. \tag{5.5}$$

The extra-gradient method It is difficult to directly get the solution of (5.5), because the both sides of this equation include the unknown u^{k+1}. Replacing u^{k+1} in the right hand side of (5.5) by u^k, we denote the output by $\tilde{u}^k$ which is produced by the projection

$$\tilde{u}^k = P_\Omega[u^k - \beta F(u^k)], \tag{5.6a}$$

We call $\tilde{u}^k$ the predictor. Then, replacing u^{k+1} in the right hand side of (5.5) by the predictor $\tilde{u}^k$, we obtain the (corrector) new iterate

$$u^{k+1} = P_\Omega[u^k - \beta F(\tilde{u}^k)] \tag{5.6b}$$

The method (5.6) is called the extra-gradient method (EG-method). Each iteration of the EG method includes two projections on Ω. In the prediction step, the parameter β should

be chosen to satisfy the following condition:

$$\beta\|F(u^k) - F(\tilde{u}^k)\| \le \nu\|u^k - \tilde{u}^k\|, \quad \nu \in (0,1), \tag{5.7}$$

which is same as in (4.6).

Convergence analysis of the extra-gradient algorithm The analysis is based on the basic property of the projection (1.5) and its consequent inequality (1.9).

♣ First, since u^{k+1} is the projection of $[u^k - \beta F(\tilde{u}^k)]$ on Ω, according to (1.9), we have

$$\|u^{k+1} - u^*\|^2 \le \|(u^k - \beta F(\tilde{u}^k)) - u^*\|^2 - \|(u^k - \beta F(\tilde{u}^k)) - u^{k+1}\|^2. \tag{5.8}$$

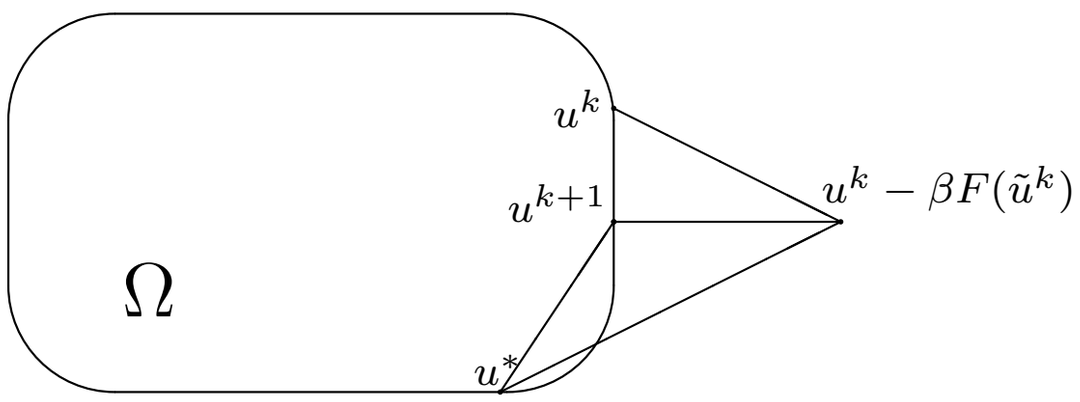


Fig.6 Geometric interpretation of inequality (5.8)

By a manipulation, it follows from (5.8) that

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \|u^k - u^{k+1}\|^2 - 2(u^{k+1} - u^*)^T \beta F(\tilde{u}^k). \tag{5.9}$$

Since $(\tilde{u}^k - u^*)^T F(\tilde{u}^k) \ge (\tilde{u}^k - u^*)^T F(u^*) \ge 0$, adding the nonnegative term $2(\tilde{u}^k - u^*)^T F(\tilde{u}^k)$ to the right hand side of (5.9), we get

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \|u^k - u^{k+1}\|^2 - 2(u^{k+1} - \tilde{u}^k)^T \beta F(\tilde{u}^k). \tag{5.10}$$

The quadratic term $\|u^k - u^{k+1}\|^2$ in the right hand side of (5.10) can be written in form

$$\|u^k - u^{k+1}\|^2 = \|u^k - \tilde{u}^k\|^2 + \|\tilde{u}^k - u^{k+1}\|^2 + 2(\tilde{u}^k - u^{k+1})^T (u^k - \tilde{u}^k).$$

Substituting it in (5.10), we get

$$\begin{aligned} \|u^{k+1} - u^*\|^2 \quad \le \quad & \|u^k - u^*\|^2 - \|u^k - \tilde{u}^k\|^2 - \|\tilde{u}^k - u^{k+1}\|^2 \\ & +2(u^{k+1} - \tilde{u}^k)^T [u^k - \beta F(\tilde{u}^k) - \tilde{u}^k]. \end{aligned} \tag{5.11}$$

♣ Now, we use the tool inequality (1.5). Setting $v = u^k - \beta F(u^k)$, then $\tilde{u}^k = P_\Omega(v)$, because $u^{k+1} \in \Omega$, according to (1.5), we have

$$2(\tilde{u}^k - u^{k+1})^T \{[u^k - \beta F(u^k)] - \tilde{u}^k\} \ge 0.$$

Adding the left term to the right hand side of (5.11), we get

$$\begin{aligned} \|u^{k+1} - u^*\|^2 \quad \le \quad & \|u^k - u^*\|^2 - \|u^k - \tilde{u}^k\|^2 - \|\tilde{u}^k - u^{k+1}\|^2 \\ & +2(u^{k+1} - \tilde{u}^k)^T \beta [F(u^k) - F(\tilde{u}^k)]. \end{aligned} \tag{5.12}$$

♣ Applying the Cauchy–Schwarz inequality to the cross term of the RHS of (5.12),

$$2(u^{k+1} - \tilde{u}^k)^T \beta [F(u^k) - F(\tilde{u}^k)] \le \|u^{k+1} - \tilde{u}^k\|^2 + \beta^2 \|F(u^k) - \beta F(\tilde{u}^k)\|^2.$$

Substituting it in (5.12), it follows that

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \|u^k - \tilde{u}^k\|^2 + \beta^2 \|F(u^k) - F(\tilde{u}^k)\|^2. \tag{5.13}$$

The contractive property of the extra-gradient method According (5.13), when the condition (5.7) is satisfied, the sequence generated by the extra-gradient method (5.6) has the following contractive property:

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - (1 - \nu^2)\|u^k - \tilde{u}^k\|^2. \tag{5.14}$$

The inequality (5.14) is the key for the proof of the convergence of the extra-gradient method. This short proof is based on the strategy in [8]. For the convergence proof of the extra-gradient algorithm, readers can also refer to [1] (Vol. II pp. 1115-1118).

6 Numerical experiments

For comparing the efficiency of the PC Algorithms and the EG Method, we test the nonlinear complementarity problem (a class of VI(Ω, F) with $\Omega = \Re^n_+$)

$$u \ge 0, \qquad F(u) \ge 0, \qquad u^T F(u) = 0.$$

In the test examples, we take

$$F(u) = D(u) + Mu + q,$$

where $Mu + q$ and $D(u)$ are the linear part and nonlinear part of $F(u)$, respectively. The linear part $Mu + q$ is generated as in [2][a], using Matlab, it produced by

```
A=(rand(n,n)-0.5)*10;  B=(rand(n,n)-0.5)*10;   B=B-B';  M=A'*A+B;
q=(rand(n,1)-0.5)*1000;        or        q=(rand(n,1)-1.0)*500;
```

In the nonlinear part $D(u)$, each element is given by $D_j(u) = d_j * \arctan(u_j)$, where d_j is a random variable between $(0, 1)$, similarly as in [13][b].

[a] In the paper by Harker and Pang [2], the matrix $M = A^T A + B + D$, where A and B are the same matrices as here, and D is a diagonal matrix with uniformly distributed random variable $d_{jj} \in (0.0, 0.3)$. In our test examples $d_{jj} \equiv 0$.

[b] In [13], the components of nonlinear mapping $D(u)$ are $D_j(u) =$ constat $* \arctan(u_j)$. Thus, $D_j(u)$ in our test example is more general.

As in [8], we have refined the EG method, and thus use the following procedure.

Refined extra-gradient method:

Step 0. Set $\beta_0 = 1$, $\nu \in (0,1)$, $u^0 \in \Omega$ and $k = 0$.

Step 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 $r_k > \nu$, $\beta_k := \frac{2}{3}\beta_k * \min\{1, \frac{1}{r_k}\}$,

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

$$r_k := \frac{\beta_k \|F(u^k) - F(\tilde{u}^k)\|}{\|u^k - \tilde{u}^k\|},$$

end(while)

$u^{k+1} = P_\Omega[u^k - \beta_k F(\tilde{u}^k)]$,

If $r_k \le \mu$ then $\beta_k := \beta_k * 1.5$, **end(if)**

Step 2. $\beta_{k+1} = \beta_k$ and $k = k+1$, go to Step 1.

When the EG method applied to solve NCP, the other people use the above program but omit the sentence $\boxed{\textbf{If } r_k \le \mu \text{ then } \beta_k := \beta_k * 1.5 \quad \textbf{end(if)}}$. Our calculation experiments shows that if this sentence is omitted, the number of iteration steps will be greatly increased, sometimes even leading to calculation failure.

The PC Algorithm (the extra computation than EG Method is indicated in a small box):

Projection and Contraction Method:

Step 0. Set $\beta_0 = 1$, $\nu \in (0,1)$, $u^0 \in \Omega$ and $k = 0$.

Step 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 $r_k > \nu$, $\beta_k := \frac{2}{3}\beta_k * \min\{1, \frac{1}{r_k}\}$,

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

$$r_k := \frac{\beta_k \|F(u^k) - F(\tilde{u}^k)\|}{\|u^k - \tilde{u}^k\|},$$

end(while)

$$\boxed{\begin{aligned} d(u^k, \tilde{u}^k) &= (u^k - \tilde{u}^k) - \beta_k[F(u^k) - F(\tilde{u}^k)], \\ \alpha_k &= \frac{(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2}, \end{aligned}}$$

$u^{k+1} = u^k - \gamma \alpha_k d(u^k, \tilde{u}^k)$,

If $r_k \le \mu$ then $\beta_k := \beta_k * 1.5$, **end(if)**

Step 2. $\beta_{k+1} = \beta_k$ and $k = k+1$, go to Step 1.

Both the EG Method and the PC Algorithm compared here can be regarded as prediction correction methods. They use the same formula

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

to produce the predictor $\tilde{u}^k$. In order to compare the efficiency, we all require that the predictor points meet (see (5.7))

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

The only difference is that EG Method uses (see (5.6b))

$$u^{k+1} = P_\Omega[u^k - \beta F(\tilde{u}^k)]$$

to update the new iterate u^{k+1}. while the PC Algorithm gives the next iterate u^{k+1} by

$$u^{k+1} = u^k - \gamma\alpha_k^* d(u^k, \tilde{u}^k), \quad \gamma \in [1,2)$$

where (see (4.3))

$$d(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k) - \beta\big(F(u^k) - F(\tilde{u}^k)\big),$$

and $\alpha_k^* = (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)/\|d(u^k, \tilde{u}^k)\|^2$. The extra work for calculating α_k^* in P-C Algorithm is small, and the correction does not need to doing projection.

In the following, we give Matlab codes for the two different algorithms

The Matlab Code of the Refined Extra-Gradient Method

```
function REG(n,M,q,d,xstart,tol,pfq)                                    %(1)
fprintf('Extragragient Method by Korpelevich   n=%4d \n',n);            %(2)
x=xstart;  Fx= d.*atan(x) + M*x + q;  stopc=norm(x-max(x-Fx,0),inf);   %(3)
beta=1;        k=0;        l=0;       tic;                             %(4)
while (stopc>tol && k<=2000)                                            %(5)
  if mod(k,pfq)==0 fprintf(' k=%4d    epsm=%9.3e  \n',k,stopc); end;   %(6)
    x0=x;   Fx0=Fx;    k=k+1;                                           %(7)
    x=max(x0-Fx0*beta,0);        Fx=d.*atan(x) + M*x + q;   l=l+1;     %(8)
       dx=x0-x;   df=(Fx0-Fx)*beta;                                     %(9)
       r=norm(df)/norm(dx);                                            %(10)
       while r>0.9     beta=0.7*beta*min(1,1/r);   l=l+1;              %(11)
         x=max(x0-Fx0*beta,0);    Fx=d.*atan(x) + M*x + q;             %(12)
         dx=x0-x;   df=(Fx0-Fx)*beta;     r=norm(df)/norm(dx);         %(13)
       end;                                                             %(14)
       x=max(x0-Fx*beta,0);                                             %(15)
       Fx= d.*atan(x) + M*x + q;     l=l+1;                             %(16)
       ex=x-max(x-Fx,0);     stopc=norm(ex,inf);                        %(17)
       if r <0.4  beta=beta*1.5;  end;                                  %(18)
end;  toc;   fprintf(' k=%4d   epsm=%9.3e     l=%4d  \n',k,stopc,l); %%%%
```

Replacing (15) in the EG code by (15a)-(15b) in the next code, we get the code for PC Algorithm.

The Matlab Code of The Projection and Contraction Method

```
function PC_G(n,M,q,d,xstart,tol,pfq)                                   %(1)
fprintf('PC Method use Direction D1 with gamma a*  n=%4d \n',n);        %(2)
x=xstart;  Fx= d.*atan(x) + M*x + q;  stopc=norm(x-max(x-Fx,0),inf);    %(3)
beta=1;        k=0;        l=0;       tic;                             %(4)
while (stopc>tol && k<=2000)                                            %(5)
  if mod(k,pfq)==0 fprintf(' k=%4d    epsm=%9.3e  \n',k,stopc); end;    %(6)
    x0=x;   Fx0=Fx;    k=k+1;                                           %(7)
    x=max(x0-Fx0*beta,0);        Fx=d.*atan(x) + M*x + q;   l=l+1;      %(8)
       dx=x0-x;   df=(Fx0-Fx)*beta;                                     %(9)
       r=norm(df)/norm(dx);                                            %(10)
       while r>0.9     beta=0.7*beta*min(1,1/r);   l=l+1;              %(11)
         x=max(x0-Fx0*beta,0);    Fx=d.*atan(x) + M*x + q;             %(12)
         dx=x0-x;   df=(Fx0-Fx)*beta;     r=norm(df)/norm(dx);         %(13)
       end;                                                            %(14)
       dxf=dx-df;    r1=dx'*dxf;    r2=dxf'*dxf;     alpha=r1/r2;     %(15a)
       x=x0- dxf*alpha*1.9;                                            %(15b)
       Fx= d.*atan(x) + M*x + q;     l=l+1;                            %(16)
       ex=x-max(x-Fx,0);     stopc=norm(ex,inf);                       %(17)
       if r <0.4  beta=beta*1.5;  end;                                 %(18)
end;  toc;   fprintf(' k=%4d   epsm=%9.3e     l=%4d  \n',k,stopc,l); %%%%
```

Instead of (15) in the EG method code, the Matlab code of PC Algorithm is (15a) and (15b).

Table 1. Numerical results for Easy Problems $q \in (-500, 500)$

	Extra-gradient Method			General PC-Method		
$n =$	No. It	No. F	CPU	No. It	No. F	CPU
500	724	1485	0.26	468	977	0.17
1000	804	1650	2.85	514	1079	1.86
2000	776	1593	10.33	407	864	5.63

Table 2. Numerical results for Hard Problems $q \in (-500, 0)$

	Extra-gradient Method			General PC-Method		
$n =$	No. It	No. F	CPU	No. It	No. F	CPU
500	1453	2983	0.53	865	1824	0.33
1000	2026	4159	7.12	1199	2553	4.38
2000	1702	3494	22.45	1025	2177	14.00

The PC method converges faster than the refined extra-gradient method.

$$\frac{\text{It. No. of Projection and Contraction Method}}{\text{It. No. of The refined extra-gradient Method}} \approx 60\%.$$

Usually, for the same problems, the PC Algorithms can save about 40% CPU time than the EG Method.

In both of the PC Algorithm and the EG Method, the computational load of each iteration is $\mathcal{O}(n^2)$. In each iteration of PC Algorithm, the computational cost for determining the step length is $\mathcal{O}(n)$, which is a small proportion in the whole calculations.

Some Ph.D Dissertations which use the EG Method to solve their problems

- Fei Sha, Large Margin Training of Acoustic Models for Speech Recognition, PhD Thesis, Computer and Information Science, University of Pennsylvania, 2007. 语音识别
- Yan Pan, A game theoretical approach to constrained OSNR optimization problems in optical network, PhD Thesis, Electrical & Computer Engineering. University of Toronto, 2009. 光纤网络
- Simon Lacoste-Julien, Discriminative Machine Learning with Structure, PhD Thesis, Computer Science. University of California, Berkeley, 2009. 机器学习
- A. G. Howard, Large Margin, Transformation Learning, PhD Thesis, Graduate School of Arts and Science. Columbia University, 2009. 机器学习

All these theses have mentioned the PC Algorithm[6]. If PC Algorithm is used instead of EG Method, the convergence speed will be greatly improved.

References

[1] F. Facchinei and J.S. Pang, Finite-dimensional variational inequality and complementarity problems, Volume I and II, Springer Series in Operations Research, 2003.

[2] Harker, P.T. Harker and J.S. Pang *A damped-Newton method for the linear complementarity problem,* Lectures in Applied Mathematics **26**, 265–284, 1990.

[3] 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.

[4] B.S. He, A new method for a class of linear variational inequalities, Math. Progr., **66**, 137-144, 1994.

[5] B.S. He, Solving a class of linear projection equations, Numerische Mathematik, **68**, 71-80, 1994.

[6] B.S. He, A class of projection and contraction methods for monotone variational inequalities, Applied Mathematics and Optimization, **35**, 69-76, 1997.

[7] B. S. He, Solving trust region problem in large scale optimization, J. Compu. Math. **18**, 1-12, 2000

[8] B.S He and L-Z Liao, Improvements of some projection methods for monotone nonlinear variational inequalities, Journal of Optimization Theory and Applications, **112**, 111-128, 2002

[9] G. M. Korpelevich. The extragradient method for finding saddle points and other problems, Ekonomika i Matematicheskie Metody, **12**, 747-756, 1976.

[10] J. Stoer and C. Witzgall, *Convexity and Optimization in Finite Dimension I*. Springer-Verlag, 1970.

[11] D.F. Sun, A new step-size skill for solving a class of nonlinear projection equations, JCM, **13**, 357-368, 1995.

[12] D.F. Sun, A class of iterative methods for solving nonlinear projection equations, JOTA **91**, 123-140, 1996.

[13] K. Taji, M. Fukushima, and T. Ibaraki, A globally convergent Newton method for solving strongly monotone variational inequalities, Math. Progr. **58**, 369-383, 1993.

[14] T. Zhu and Z. G. Yu, A simple proof for some important properties of the projection mapping. Math. Inequal. Appl. **7**, 453–456, 2004.

[15] 何炳生, 论求解单调变分不等式的一些投影收缩算法,〖计算数学〗, **18**, 54-60, 1996.

凸优化和单调变分不等式的收缩算法

第三讲：单调变分不等式投影收缩算法中的两对孪生方法

Two pairs of twin projection and contraction methods for monotone variational inequalities

南京大学数学系　何炳生
hebma@nju.edu.cn

The context of this lecture is based on the publication [5, 7, 10]

We study the solution methods for nonlinear VI. For any $\beta > 0$, the variational inequalities

$$\textbf{VI}(\Omega, F) \quad u^* \in \Omega, \quad (u-u^*)^T F(u^*) \geq 0, \quad \forall u \in \Omega \tag{0.1}$$

and

$$\textbf{VI}(\Omega, F) \quad u^* \in \Omega, \quad (u-u^*)^T \beta F(u^*) \geq 0, \quad \forall u \in \Omega \tag{0.2}$$

have the same solution set. For given u, let $e(u,\beta) := u - P_\Omega[u - \beta F(u)]$, $\|e(u,\beta)\|$ is the error function, which measures how much u failed to be a solution point.

Recent years, the projection contraction algorithm for linear variational inequality [3, 4] has been successfully applied to robot motion planning and real-time control [2, 11]. Some scholars in Geotechnical Mechanics have solved the problems that worry them for a long time [12, 13] by the PC Algorithms [6, 7] successfully .

In the last lecture, we have introduced the PC Algorithms for monotone LVI and NVI whose search directions are based on FI1+FI2 and FI1+FI2+FI3, respectively. Actually, accompanied each PC Algorithm in the last lecture, there is a twin algorithm, whose search directions are based on FI1 and FI1+FI3, respectively. It is reasonable, among the twin algorithms, the method utilizes fewer fundamental inequalities, is more efficient.

It is interesting that the same step length is used when the new iteration points are updated by the different correction procedures. From the perspective of mathematics itself, there is such a clever inner connection, which also gives us the enjoyment of the beauty of mathematics.

1 Mathematical Backgrounds

Basic property of the projection mapping We need to list a few important properties of the projection operator. Readers interested in proving these results can refer to Lecture 2 of this series of lectures.

Lemma 1.1 *Let $\Omega \subset \Re^n$ be a closed convex set, then for any $v \in \Re^n$, we have*

$$(u - P_\Omega(v))^T (P_\Omega(v) - v) \geq 0, \quad \forall u \in \Omega. \tag{1.1}$$

The equivalent projection equation By using the properties of the projection, solving the variational inequality (0.1) is equivalent to finding a zero point of $e(u, \beta)$ which is defined by

$$e(u, \beta) := u - P_\Omega[u - \beta F(u)].$$

The projection contraction algorithm for monotone variational inequality (0.2) is a prediction-correction method. The prediction is provided by projection and the contraction is realized by correction. In the k-th iteration of the PC Algorithms, for

given u^k and $\beta_k > 0$, the predictor $\tilde{u}^k$ is given by

$$\tilde{u}^k = P_\Omega[u^k - \beta_k F(u^k)]. \tag{1.2}$$

Thus, $u^k \in \Omega^*$ (or $e(u, \beta) = 0$) if and only if $u^k = \tilde{u}^k$.

Set $v = u^k - \beta_k F(u^k)$ in (1.1), because $\tilde{u}^k = P_\Omega(v)$, it follows from (1.1) that

$$\tilde{u}^k \in \Omega, \quad (u - \tilde{u}^k)^T \{\tilde{u}^k - [u^k - \beta_k F(u^k)]\} \geq 0, \quad \forall u \in \Omega. \tag{1.3}$$

Consequently, we get

$$\tilde{u}^k \in \Omega, \quad (u - \tilde{u}^k)^T \beta_k F(u^k) \geq (u - \tilde{u}^k)^T (u^k - \tilde{u}^k), \quad \forall\, u \in \Omega. \tag{1.4}$$

Three fundamental inequalities Let u^* be a solution of VI (Ω, F). Since $\tilde{u}^k \in \Omega$, according to the definition of the variational inequality (0.1), we have

$$\text{(FI1)} \qquad (\tilde{u}^k - u^*)^T \beta_k F(u^*) \geq 0. \tag{1.5}$$

Since $u^* \in \Omega$, setting the any $u \in \Omega$ in (1.3) by u^*, it follows that

$$\text{(FI2)} \qquad (\tilde{u}^k - u^*)^T \big([u^k - \tilde{u}^k] - \beta_k F(u^k)\big) \geq 0. \tag{1.6}$$

The variational inequality considered in this series is monotone, thus

$$\text{(FI3)} \qquad (\tilde{u}^k - u^*)^T \big(\beta_k F(\tilde{u}^k) - \beta_k F(u^*)\big) \geq 0. \tag{1.7}$$

We call (1.5), (1.6) and (1.7) three fundamental inequalities. Although these inequalities are simple, they are very important. The search directions of the PC Algorithms are derived from these inequalities.

2 A pair of twin PC Algorithms for LVI

When the operator $F(u)$ in (0.1) is affine, $F(u) = Mu + q$, such variational inequality is linear (abbreviated LVI):

$$u^* \in \Omega, \quad (u - u^*)^T (Mu^* + q) \geq 0, \quad \forall\, u \in \Omega.$$

We say the LVI is monotone if $M^T + M$ is positive semidefinite, it is not necessarily that M is symmetric.

2.1 The ascent directions provided by the predictor

For linear variational inequality, we use a fixed $\beta > 0$.

- **The ascent direction provided by FI1** Because $F(u) = Mu + q$, the fundamental inequality (1.5) can be written as

  $$\{(u^k - u^*) - (u^k - \tilde{u}^k)\}^T \beta \{(Mu^k + q) - M(u^k - u^*)\} \geq 0,$$

  Since $(u^k - u^*)^T M (u^k - u^*) \geq 0$, it follows from the above inequality

  $$(u^k - u^*)^T \underline{\beta[M^T(u^k - \tilde{u}^k) + (Mu^k + q)]} \geq (u^k - \tilde{u}^k)^T \beta (Mu^k + q), \tag{2.1}$$

  If $u^k \in \Omega$, using $F(u) = Mu + q$, from (1.3) we get

  $$(u^k - \tilde{u}^k)^T \beta_k (Mu^k + q) \geq \|u^k - \tilde{u}^k\|^2. \tag{2.2}$$

  Thus, if $u^k \in \Omega$, $\beta_k \big(M^T(u^k - \tilde{u}^k) + (Mu^k + q)\big)$ is a ascent direction of $\frac{1}{2}\|u - u^*\|^2$ at u^k. **(2.1)-(2.2) is true only for $u^k \in \Omega$.**

- **The ascent direction provided by FI1+FI2** Adding (1.5) and (1.6), and using $F(u) = Mu + q$, we get

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

Since $(u^k - u^*)^T M(u^k - u^*) \geq 0$, it follows from the above inequality

$$(u^k - u^*)^T \underline{(I + \beta M^T)(u^k - \tilde{u}^k)} \geq \|u^k - \tilde{u}^k\|^2. \qquad (2.3)$$

Thus, $(I + \beta M^T)(u^k - \tilde{u}^k)$ is a ascent direction of the unknown distance function of $\frac{1}{2}\|u - u^*\|^2$ at the point u^k. **(2.3) is true for any $u^k \in \Re^n$.**

A pair of the twin directions For $F(u) = Mu + q$, (1.4) can be written as

$$\tilde{u}^k \in \Omega, \quad (u - \tilde{u}^k)^T \beta(Mu^k + q) \geq (u - \tilde{u}^k)^T(u^k - \tilde{u}^k), \quad \forall\, u \in \Omega.$$

Adding $(u - \tilde{u}^k)^T \beta M^T(u^k - \tilde{u}^k)$ to the both sides of the last inequality, we get

$$\begin{aligned} \tilde{u}^k \in \Omega, \quad (u - \tilde{u}^k)^T & \underline{\beta[M^T(u^k - \tilde{u}^k) + (Mu^k + q)]} \\ & \geq (u - \tilde{u}^k)^T \underline{(I + \beta M^T)(u^k - \tilde{u}^k)}, \quad \forall u \in \Omega. \quad (2.4) \end{aligned}$$

We call (2.4) the two directions

$$\beta[M^T(u^k - \tilde{u}^k) + (Mu^k + q)] \quad \text{and} \quad (I + \beta M^T)(u^k - \tilde{u}^k) \qquad (2.5)$$

which lay on the left and right sides of (2.4), respectively, a pair of twin ascent directions for LVI. They are also the direction in (2.1) and (2.3), which derived from (FI1) and (FI1+FI2), respectively. For the notation convenience, we denote

$$g(u^k, \tilde{u}^k) = M^T(u^k - \tilde{u}^k) + (Mu^k + q). \qquad (2.6)$$

Using this expression, (2.4) can be written as

$$\begin{aligned} \tilde{u}^k \in \Omega, \qquad (u - \tilde{u}^k)^T & \underline{\beta g(u^k, \tilde{u}^k)} \\ & \geq (u - \tilde{u}^k)^T \underline{(I + \beta M^T)(u^k - \tilde{u}^k)}, \quad \forall u \in \Omega. \quad (2.7) \end{aligned}$$

Notice that for the direction $(I + \beta M^T)(u^k - \tilde{u}^k)$, there is a constant $K > 0$, such that $\|(I + \beta M^T)(u^k - \tilde{u}^k)\| \leq K\|u^k - \tilde{u}^k\|$. However, for the related direction $g(u^k, \tilde{u}^k)$, we do not have the similar statement !

2.2 Update the new iterate by the direction due to FI1+FI2

The correction uses the descent direction (the opposite of the ascent direction) of the distance function to make the new iteration point closer to the solution set. Based on the direction provided by FI1+FI2, the new iterate is updated by

$$u_{BD}^{k+1}(\alpha) = u^k - \alpha(I + \beta M^T)(u^k - \tilde{u}^k). \tag{2.8}$$

The lower index 'BD' means 'Bounded Direction'. For discussion how to determine the step length α, we denote the output of (2.8) by $u_{BD}^{k+1}(\alpha)$. Let us investigate the α-dependent reduction of the square of the distance

$$\vartheta_k^L(\alpha) := \|u^k - u^*\|^2 - \|u_{BD}^{k+1}(\alpha) - u^*\|^2. \tag{2.9}$$

According to the definition,

$$\begin{aligned} \vartheta_k^L(\alpha) &= \|u^k - u^*\|^2 - \|u^k - u^* - \alpha(I + \beta M^T)(u^k - \tilde{u}^k)\|^2 \\ &= 2\alpha(u^k - u^*)^T(I + \beta M^T)(u^k - \tilde{u}^k) \\ &\quad -\alpha^2\|(I + \beta M^T)(u^k - \tilde{u}^k)\|^2. \end{aligned} \tag{2.10}$$

For any given solution point u^*, (2.10) tell us that $\vartheta_k^L(\alpha)$ is a quadratic function of α. Since u^* is unknown, We can't directly find the maximum of $\vartheta_k^L(\alpha)$. With the help of (2.1), we have the following theorem.

Theorem 2.1 *Let $u_{BD}^{k+1}(\alpha)$ be updated by* (2.8)*. Then for $\alpha > 0$, we have*

$$\vartheta_k^L(\alpha) \geq q_k^L(\alpha), \tag{2.11}$$

where $\vartheta_k^L(\alpha)$ defined by (2.9) *and*

$$q_k^L(\alpha) = 2\alpha\|u^k - \tilde{u}^k\|^2 - \alpha^2\|(I + \beta M^T)(u^k - \tilde{u}^k)\|^2. \tag{2.12}$$

Proof. The assertion derived directly from (2.10) by using (2.1). □

Theorem 2.1 indicates that $q_k^L(\alpha)$ is a lower bound of $\vartheta_k^L(\alpha)$. The quadratic function $q_k^L(\alpha)$ reaches its maximum at

$$\alpha_k^* = \text{argmax}\{q_k^L(\alpha)\} = \frac{\|u^k - \tilde{u}^k\|^2}{\|(I + \beta M^T)(u^k - \tilde{u}^k)\|^2}. \tag{2.13}$$

From (2.13), it follows that

$$\alpha_k^* \geq \frac{1}{\|I + \beta M^T\|^2}. \tag{2.14}$$

In the practical computation, we take a relaxed factor $\gamma \in [1, 2)$ and updated the new iterate by

$$u_{BD}^{k+1} = u^k - \gamma \alpha_k^* (I + \beta M^T)(u^k - \tilde{u}^k), \tag{2.15}$$

Theorem 2.2 *Let* $u^{k+1} = u_{BD}^{k+1}$ *updated by* (2.15) *with* α_k^* *given by* (2.13)*, then*

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - \frac{\gamma(2-\gamma)}{\|I + \beta M^T\|^2} \|u^k - \tilde{u}^k\|^2. \tag{2.16}$$

Proof. According to (2.9) and (2.11), the u^{k+1} updated by (2.15) satisfies

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - q_k^L(\gamma \alpha_k^*). \tag{2.17}$$

According to the definitions of $q_k^L(\alpha)$ and α_k^* (see (2.12) and (2.13)), we get

$$q_k^L(\gamma \alpha_k^*) = \gamma(2-\gamma)\alpha_k^* \|u^k - \tilde{u}^k\|^2 \geq \frac{\gamma(2-\gamma)}{\|I + \beta M^T\|^2} \|u^k - \tilde{u}^k\|^2.$$

The last inequality follows from (2.14). The proof is complete. □

2.3 Update the new iterate by the direction due to FI1

The correction form (2.8) in §2.2 takes $(I + \beta M^T)(u^k - \tilde{u}^k)$ as the search direction. In this subsection, it is replaced by its related direction (see (2.5))

$$\beta[M^T(u^k - \tilde{u}^k) + (Mu^k + q)].$$

In §2.1, it is emphasized that (2.1)-(2.2) are true only for $u^k \in \Omega$. We use

$$u_{BLD}^{k+1}(\alpha) = P_\Omega\{u^k - \alpha\beta[M^T(u^k - \tilde{u}^k) + (Mu^k + q)]\}, \tag{2.18}$$

to update the new iterate ensured in Ω. We denote the output of (2.18) by $u_{BLD}^{k+1}(\alpha)$. The lower index 'BLD' means 'Boundless Direction', because

$g(u^k, \tilde{u}^k) \not\to 0$ as dist$(\tilde{u}^k, \Omega^*) \to 0$. For discussion how to determine the step length α, Let us investigate the α-dependent reduction of the square of the distance

$$\zeta_k^L(\alpha) = \|u^k - u^*\|^2 - \|u_{BLD}^{k+1}(\alpha) - u^*\|^2, \tag{2.19}$$

which is a function of α. We can not maximize $\zeta_k^L(\alpha)$ directly because it involves the unknown vector u^*. The following theorem indicates that for the same $\alpha > 0$, $\zeta_k^L(\alpha)$ is 'better than $\vartheta_k^L(\alpha)$ in (2.11).

Theorem 2.3 *Let $u_{BLD}^{k+1}(\alpha)$ be updated by* (2.18)*. Then for $\zeta_k^L(\alpha)$ defined in* (2.19) *with any $\alpha > 0$, we have*

$$\zeta_k^L(\alpha) \geq q_k^L(\alpha) + \|u_{BLD}^{k+1}(\alpha) - u_{BD}^{k+1}(\alpha)\|^2, \tag{2.20}$$

where $q_k^L(\alpha), u_{BD}^{k+1}(\alpha)$ and $u_{BLD}^{k+1}(\alpha)$ are given by (2.12)*,* (2.8) *and* (2.18)*, respectively.*

Proof. By using the notation $g(u^k, \tilde{u}^k)$ in (2.6), the update form (2.18) can be written as $u_{BLD}^{k+1}(\alpha) = P_\Omega[u - \alpha\beta g(u^k, \tilde{u}^k)]$. Since $u^* \in \Omega$ and the

projection operator is non-expansive, we have

$$\|u_{BLD}^{k+1}(\alpha) - u^*\|^2 \;\leq\; \|u^k - \alpha\beta g(u^k, \tilde{u}^k) - u^*\|^2.$$

However, according to the properties of projection and cosine theorem, we use more precise relations

$$\begin{aligned} \|u_{BLD}^{k+1}(\alpha) - u^*\|^2 \quad \leq \quad & \|u^k - \alpha\beta g(u^k, \tilde{u}^k) - u^*\|^2 \\ & -\|u_{BLD}^{k+1}(\alpha) - (u^k - \alpha\beta g(u^k, \tilde{u}^k))\|^2. \end{aligned} \tag{2.21}$$

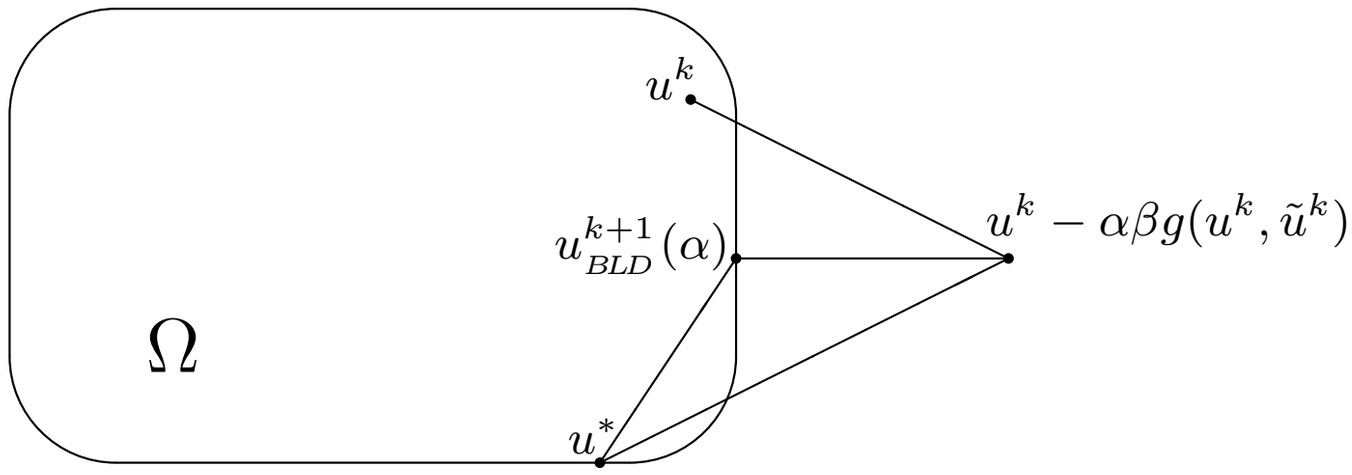


Fig.1 Geometric interpretation of the inequality (2.21)

Hence, by using $\zeta_k^L(\alpha)$ (see (2.19)) and (2.21), we have

$$\begin{aligned}
\zeta_k^L(\alpha) &\geq \|u^k - u^*\|^2 - \|(u^k - u^*) - \alpha\beta g(u^k, \tilde{u}^k)\|^2 \\
&\quad + \|(u_{BLD}^{k+1}(\alpha) - u^k) + \alpha\beta g(u^k, \tilde{u}^k)\|^2 \\
&= 2\alpha\beta(u^k - u^*)^T g(u^k, \tilde{u}^k) + 2\alpha\beta(u_{BLD}^{k+1}(\alpha) - u^k)^T g(u^k, \tilde{u}^k) \\
&\quad + \|u_{BLD}^{k+1}(\alpha) - u^k\|^2 \\
&= \|u_{BLD}^{k+1}(\alpha) - u^k\|^2 + 2\alpha(u_{BLD}^{k+1}(\alpha) - u^*)^T \beta g(u^k, \tilde{u}^k). \quad (2.22)
\end{aligned}$$

Decomposing the cross term of the right hand side of (2.22) in form

$$\begin{aligned}
&(u_{BLD}^{k+1}(\alpha) - u^*)^T \beta g(u^k, \tilde{u}^k) \\
&\quad = (u_{BLD}^{k+1}(\alpha) - \tilde{u}^k)^T \beta g(u^k, \tilde{u}^k) + (\tilde{u}^k - u^*)^T \beta g(u^k, \tilde{u}^k). \quad (2.23)
\end{aligned}$$

To the first term of the right hand side of (2.23), since $u_{BLD}^{k+1}(\alpha) \in \Omega$, by using (2.7), we have

$$(u_{BLD}^{k+1}(\alpha) - \tilde{u}^k)^T \beta g(u^k, \tilde{u}^k) \geq (u_{BLD}^{k+1}(\alpha) - \tilde{u}^k)^T (I + \beta M^T)(u^k - \tilde{u}^k).$$

In other words,

$$\begin{aligned}
(u_{BLD}^{k+1}(\alpha) - \tilde{u}^k)^T \beta g(u^k, \tilde{u}^k) &\geq (u_{BLD}^{k+1}(\alpha) - u^k)^T (I + \beta M^T)(u^k - \tilde{u}^k) \\
&\quad + (u^k - \tilde{u}^k)^T (I + \beta M^T)(u^k - \tilde{u}^k). \quad (2.24)
\end{aligned}$$

To the second term of the right hand side of (2.23), $(\tilde{u}^k - u^*)^T \beta g(u^k, \tilde{u}^k)$, we split it in form

$$(\tilde{u}^k - u^*)^T \beta g(u^k, \tilde{u}^k) = (\tilde{u}^k - u^k)^T \beta g(u^k, \tilde{u}^k) + (u^k - u^*)^T \beta g(u^k, \tilde{u}^k)$$

By using (2.1), namely,

$$(u^k - u^*)^T \beta g(u^k, \tilde{u}^k) \geq (u^k - \tilde{u}^k)^T \beta (M u^k + q),$$

we get

$$\begin{aligned}
&(\tilde{u}^k - u^*)^T \beta g(u^k, \tilde{u}^k) \\
&\quad = (\tilde{u}^k - u^k)^T \beta g(u^k, \tilde{u}^k) + (u^k - u^*)^T \beta g(u^k, \tilde{u}^k) \\
&\quad \geq (\tilde{u}^k - u^k)^T \beta g(u^k, \tilde{u}^k) + (u^k - \tilde{u}^k)^T \beta (M u^k + q).
\end{aligned}$$

Consequently, by using the notation of $g(u^k, \tilde{u}^k)$, we get

$$\begin{aligned}(\tilde{u}^k - u^*)^T \beta g(u^k, \tilde{u}^k) &\geq (\tilde{u}^k - u^k)^T \{\beta g(u^k, \tilde{u}^k) - \beta(Mu^k + q)\} \\ &= -\beta(u^k - \tilde{u}^k)^T M^T (u^k - \tilde{u}^k). \qquad (2.25)\end{aligned}$$

Adding (2.24) and (2.25), it follows that

$$(u_{BLD}^{k+1}(\alpha) - u^*)^T \beta g(u^k, \tilde{u}^k) \geq (u_{BLD}^{k+1}(\alpha) - \tilde{u}^k)^T (I + \beta M^T)(u^k - \tilde{u}^k) + \|u^k - \tilde{u}^k\|^2.$$

Substituting the above inequality in (2.22) and using the notation of $q_k^L(\alpha)$, we get

$$\begin{aligned}\zeta_k^L(\alpha) &\geq \|u_{BLD}^{k+1}(\alpha) - u^k\|^2 + 2\alpha(u_{BLD}^{k+1}(\alpha) - u^k)^T (I + \beta M^T)(u^k - \tilde{u}^k) \\ &\quad + 2\alpha\|u^k - \tilde{u}^k\|^2 \\ &= \|(u_{BLD}^{k+1}(\alpha) - u^k) + \alpha(I + \beta M^T)(u^k - \tilde{u}^k)\|^2 \\ &\quad - \alpha^2\|(I + \beta M^T)(u^k - \tilde{u}^k)\|^2 + 2\alpha\|u^k - \tilde{u}^k\|^2 \\ &= \|u_{BLD}^{k+1}(\alpha) - [u^k - \alpha(I + \beta M^T)(u^k - \tilde{u}^k)]\|^2 + q_k^L(\alpha).\end{aligned}$$

Because $[u^k - \alpha(I + \beta M^T)(u^k - \tilde{u}^k)] = u_{BD}^{k+1}(\alpha)$ (see (2.8)), the last inequality is the same as (2.20). The proof is complete. $\square$

Theorem 2.3 tells us that $q_k^L(\alpha)$ is also a lower bound of $\zeta_k^L(\alpha)$. In the practical computation, whether the correction formula

$$\text{(PC-I)} \quad u_{BD}^{k+1} = u^k - \gamma\alpha_k^*(I + \beta M^T)(u^k - \tilde{u}^k) \qquad (2.26)$$

or

$$\text{(PC-II)} \quad u_{BLD}^{k+1} = P_\Omega\big[u^k - \gamma\alpha_k^*\beta[M^T(u^k - \tilde{u}^k) + (Mu^k + q)]\big] \qquad (2.27)$$

is used to update the new iterate u^{k+1}, both the step length α_k^* is given by (2.13), which is lower bounded from 0.

An effective iterative algorithm, the step size must be lower bounded from 0.

Using the formula (2.26) to update u^{k+1}, the advantage is that the correction does not need to do an extra projection. However, in many practical problems, the cost of the projection on Ω (e. g., $\Omega = \Re_+^n$ or a 'box') is not expensive. Thus, the correction formula (2.27) is often used. The reasons are explained in paper [10].

Based on the pair of twin directions offered by (2.4), the different algorithms (2.26) and (2.27), we get the following key inequalities for the contraction:

$$\vartheta_k^L(\alpha) = \|u^k - u^*\|^2 - \|u_{BD}^{k+1}(\alpha) - u^*\|^2 \;\geq\; q_k^L(\alpha),$$

$$\begin{aligned}\zeta_k^L(\alpha) &= \|u^k - u^*\|^2 - \|u_{BLD}^{k+1}(\alpha) - u^*\|^2 \\ &\geq\; q_k^L(\alpha) + \|u_{BLD}^{k+1}(\alpha) - u_{BD}^{k+1}(\alpha)\|^2,\end{aligned}$$

where $q_k^L(\alpha)$ is defined by (2.12) and reaches its maximum at α_k^* (see (2.13)).

The PC Algorithms for solving LVI, PC-I (2.26) and PC-II (2.27), are published in [3] and [4], respectively. Both the algorithms are successfully applied to robot motion planning and real-time control by Zhang and his students [2, 11].

In [2, 11], PC-I (2.26) is called **94LVI** because the method was published in 1994 and the title of the article is 'A new method for a class of **L**inear **V**ariational **I**nequalities"; PC-II (2.27) is named as **E47** since it is described by **E**quations (**4**) to (**7**) in the original article [4]. The numerical experiments in [2, 11] verified that PC-II (2.27) is more efficient than PC-I (2.26).

3 A pair of twin PC Algorithms for NVI

For nonlinear variational inequalities (abbreviated NVI), we assume its operator F is Lipschitz continuous. In the projection for getting the predictor $\tilde{u}^k$ (1.2), the parameter β_k is chosen such that

$$\beta_k\|F(u^k) - F(\tilde{u}^k)\| \leq \nu\|u^k - \tilde{u}^k\|, \quad \nu \in (0,1). \tag{3.1}$$

In the following, for analysis convenience, we omit the index k in β_k and assume that this β satisfies the condition (3.1).

3.1 The ascent directions provided by the predictor

- **The ascent direction provided by FI1+FI3** Adding (1.5) and (1.7), it follows that

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

  From the last inequality, we get

  $$(u^k - u^*)^T\beta F(\tilde{u}^k) \geq (u^k - \tilde{u}^k)^T\beta F(\tilde{u}^k). \tag{3.2}$$

When $u^k \in \Omega$, according to (1.4), we have

$$(u^k - \tilde{u}^k)^T \beta F(u^k) \geq \|u^k - \tilde{u}^k\|^2.$$

According to the last inequality and the assumption (3.1), it follows from the Cauchy-Schwarz inequality that

$$\begin{aligned} & (u^k - \tilde{u}^k)^T \beta F(\tilde{u}^k) \\ = \quad & (u^k - \tilde{u}^k)^T \beta F(u^k) - (u^k - \tilde{u}^k)^T \beta (F(u^k) - F(\tilde{u}^k)) \\ \geq \quad & (1-\nu)\|u^k - \tilde{u}^k\|^2. \end{aligned} \tag{3.3}$$

The inequalities (3.2) and (3.3) tell us, for $u^k \in \Omega$, under the assumption (3.1), $\beta F(\tilde{u}^k)$ is an ascent direction of the unknown distance function $\frac{1}{2}\|u - u^*\|^2$ at the point u^k. **(3.2)-(3.3) is true only for $u^k \in \Omega$.**

- **The ascent direction provided by FI1+FI2+ FI3**. Adding the fundamental inequalities (1.5), (1.6) and (1.7), we get

$$\{(u^k - u^*) - (u^k - \tilde{u}^k)\}^T d(u^k, \tilde{u}^k) \geq 0,$$

where

$$d(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k) - \beta[F(u^k) - F(\tilde{u}^k)]. \tag{3.4}$$

Consequently, we have

$$(u^k - u^*)^T d(u^k, \tilde{u}^k) \geq (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \tag{3.5}$$

According to the notation $d(u^k, \tilde{u}^k)$ (3.4) and the assumption (3.1), it follows from the Cauchy-Schwarz inequality that

$$(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) \geq (1-\nu)\|u^k - \tilde{u}^k\|^2. \tag{3.6}$$

The inequalities (3.5) and (3.6) tell us, under the assumption (3.1), $d(u^k, \tilde{u}^k)$ is an ascent direction of the unknown distance function $\frac{1}{2}\|u - u^*\|^2$ at the point u^k. **(3.5)-(3.6) is true for any $u^k \in \Re^n$.**

A pair of the twin directions Notice that the inequality (1.4) is derived from the basic property of the projection. Adding the term

$$(u - \tilde{u}^k)^T \{-\beta[F(u^k) - F(\tilde{u}^k)]\},$$

to the both sides of (1.4), we get

$$\tilde{u}^k \in \Omega, \quad (u-\tilde{u}^k)^T \underline{\beta F(\tilde{u}^k)} \geq (u-\tilde{u}^k)^T \underline{d(u^k,\tilde{u}^k)}, \quad \forall u \in \Omega, \tag{3.7}$$

where $d(u^k,\tilde{u}^k)$ is given by (3.4). We call the directions

$$\beta F(\tilde{u}^k) \quad \text{and} \quad d(u^k,\tilde{u}^k) = (u^k-\tilde{u}^k) - \beta[F(u^k)-F(\tilde{u}^k)]$$

which lay on the left and right sides of (3.7), respectively, a pair of twin ascent directions for NVI. They are derived from FI1+FI3 and FI1+FI2+FI3, respectively.

3.2 Update the new iterate by the direction due to FI1+FI2+FI3

The correction uses the descent direction (the opposite of the ascent direction) of the distance function to make the new iteration point closer to the solution set. Based on the direction provided by FI1+FI2+FI3, the new iterate is updated by

$$u_{BD}^{k+1}(\alpha) = u^k - \alpha d(u^k,\tilde{u}^k). \tag{3.8}$$

where $d(u^k,\tilde{u}^k)$ is given by (3.4). The lower index 'BD' means 'Bounded

Direction' because $\|d(u^k,\tilde{u}^k)\| < 2\|u^k-\tilde{u}^k\|$. For discussion how to determine the step length α, we denote the output of (2.8) by $u_{BD}^{k+1}(\alpha)$. Let us investigate the α-dependent reduction of the square of the distance

$$\vartheta_k^N(\alpha) := \|u^k-u^*\|^2 - \|u_{BD}^{k+1}(\alpha)-u^*\|^2. \tag{3.9}$$

According to the definition,

$$\begin{aligned} \vartheta_k^N(\alpha) &= \|u^k-u^*\|^2 - \|u^k-u^*-\alpha d(u^k,\tilde{u}^k)\|^2 \\ &= 2\alpha(u^k-u^*)^T d(u^k,\tilde{u}^k) - \alpha^2\|d(u^k,\tilde{u}^k)\|^2. \end{aligned} \tag{3.10}$$

For any given solution point u^*, (3.10) tell us that $\vartheta_k^N(\alpha)$ is a quadratic function of α. Since u^* is unknown, we can't directly find the maximum of $\vartheta_k^N(\alpha)$. With the help of (3.5), we have

Theorem 3.1 *Let $u_{BD}^{k+1}(\alpha)$ be updated by* (3.8) *with $d(u^k,\tilde{u}^k)$ given by* (3.4)*. Then for $\alpha > 0$, we have*

$$\vartheta_k^N(\alpha) \geq q_k^N(\alpha), \tag{3.11}$$

where

$$q_k^N(\alpha) = 2\alpha(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \alpha^2 \|d(u^k, \tilde{u}^k)\|^2. \tag{3.12}$$

Proof. The assertion derived directly from (3.10) by using (3.5). □

Theorem 3.1 indicates that $q_k^N(\alpha)$ is a lower bound of $\vartheta_k^N(\alpha)$. The quadratic function $q_k^N(\alpha)$ reaches its maximum at

$$\alpha_k^* = \text{argmax}\{q_k^N(\alpha)\} = \frac{(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2}. \tag{3.13}$$

In the practical computation, similarly as in §2, we take a relaxed factor $\gamma \in [1, 2)$ and updated the new iterate by

$$u_{BD}^{k+1} = u^k - \gamma \alpha_k^* d(u^k, \tilde{u}^k), \tag{3.14}$$

where $d(u^k, \tilde{u}^k)$ is given by (3.5) and is α_k^* given by (3.13).

Theorem 3.2 *Let* $u^{k+1} = u_{BD}^{k+1}$ *be updated by* (3.14)*. Then for any* $\gamma \in (0, 2)$*,*

we have

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \frac{1}{2}\gamma(2-\gamma)(1-\nu)\|u^k - \tilde{u}^k\|^2. \tag{3.15}$$

Proof. According to (3.9) and (3.11), the u^{k+1} updated by (3.14) satisfies

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - q_k^N(\gamma\alpha_k^*). \tag{3.16}$$

According to the definitions $q_k^N(\alpha)$ and α_k^* (see (3.12) and (3.13)), we get

$$\begin{aligned} q_k^N(\gamma\alpha_k^*) &= 2\gamma\alpha_k^*(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \gamma^2(\alpha_k^*)^2\|d(u^k, \tilde{u}^k)\|^2 \\ &= \gamma(2-\gamma)\alpha_k^*(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \end{aligned}$$

In fact, by using (3.4) and (3.1), we have

$$\begin{aligned} & 2(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \|d(u^k, \tilde{u}^k)\|^2 \\ &= d(u^k, \tilde{u}^k)^T\{2(u^k - \tilde{u}^k) - d(u^k, \tilde{u}^k)\} \\ &= \|u^k - \tilde{u}^k\|^2 - \beta_k^2\|[F(u^k) - F(\tilde{u}^k)]\|^2 \ > \ 0. \end{aligned} \tag{3.17}$$

Thus, $\alpha_k^* > \frac{1}{2}$, and consequently, it follows that

$$q_k^N(\gamma\alpha_k^*) > \frac{1}{2}\gamma(2-\gamma)(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k).$$

Substituting it in (3.16), we get

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \frac{1}{2}\gamma(2-\gamma)(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k).$$

Using (3.6) to the term $(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)$ in the right hand side of the last inequality, we get the assertion (3.15) and the theorem is proved. □

3.3 Update the new iterate by the direction due to FI1+FI3

The correction step (3.8) in §3.3 takes $d(u^k, \tilde{u}^k)$ as the search direction. In this section, it is replaced by $\beta F(\tilde{u}^k)$ and finished the correction with an additional projection. Namely, we let

$$u_{BLD}^{k+1}(\alpha) = P_\Omega[u^k - \alpha\beta F(\tilde{u}^k)], \tag{3.18}$$

to update the α-dependent new iterate ensured in Ω. We denote the output of (3.18) by $u_{BLD}^{k+1}(\alpha)$. The lower index 'BLD' means 'Boundless Direction', because $F(\tilde{u}^k) \not\to 0$ as dist$(\tilde{u}^k, \Omega^*) \to 0$. For discussion how to determine the step length α, Let us investigate the α-dependent reduction of the square of the distance

$$\zeta_k^N(\alpha) = \|u^k - u^*\|^2 - \|u_{BLD}^{k+1}(\alpha) - u^*\|^2 \tag{3.19}$$

which is a function of α. We can not maximize $\zeta_k^N(\alpha)$ directly because it involves the unknown vector u^*. The following theorem indicates that for the same $\alpha > 0$, $\zeta_k^N(\alpha)$ is 'better than $\vartheta_k^N(\alpha)$ in (3.11)

Theorem 3.3 *Let $u_{BLD}^{k+1}(\alpha)$ be updated by* (3.18)*. Then for $\zeta_k^N(\alpha)$ defined in* (3.19) *with any $\alpha > 0$, we have*

$$\zeta_k^N(\alpha) \ge q_k^N(\alpha) + \|u_{BLD}^{k+1}(\alpha) - u_{BD}^{k+1}(\alpha)\|^2, \tag{3.20}$$

where $q_k^N(\alpha), u_{BD}^{k+1}(\alpha)$ and $u_{BLD}^{k+1}(\alpha)$ are given by (3.12)*,* (3.8) *and* (3.18)*, respectively.*

Proof. First, similarly as (2.21), because $u^{k+1}(\alpha) = P_\Omega[u - \alpha\beta F(\tilde{u}^k)]$ and $u^* \in \Omega$, we have (see Fig. 1)

$$\|u^{k+1}_{BLD}(\alpha) - u^*\|^2 \le \|u^k - \alpha\beta F(\tilde{u}^k) - u^*\|^2 - \|u^k - \alpha\beta F(\tilde{u}^k) - u^{k+1}_{BLD}(\alpha)\|^2. \tag{3.21}$$

Hence, by using $\zeta_k^N(\alpha)$ (see (3.19)) and (3.21), we have

$$\begin{aligned}
\zeta_k^N(\alpha) &\ge \|u^k - u^*\|^2 - \|(u^k - u^*) - \alpha\beta F(\tilde{u}^k)\|^2 \\
&\qquad + \|(u^k - u^{k+1}_{BLD}(\alpha)) - \alpha\beta F(\tilde{u}^k)\|^2 \\
&= 2\alpha(u^k - u^*)^T \beta F(\tilde{u}^k) + 2\alpha(u^{k+1}_{BLD}(\alpha) - u^k)^T \beta F(\tilde{u}^k) \\
&\qquad + \|u^k - u^{k+1}_{BLD}(\alpha)\|^2 \\
&= \|u^k - u^{k+1}_{BLD}(\alpha)\|^2 + 2\alpha(u^{k+1}_{BLD}(\alpha) - u^*)^T \beta F(\tilde{u}^k). \qquad (3.22)
\end{aligned}$$

Decomposing the cross term of the right hand side of (3.22) in form

$$(u^{k+1}_{BLD}(\alpha) - u^*)^T \beta F(\tilde{u}^k) = (u^{k+1}_{BLD}(\alpha) - \tilde{u}^k)^T \beta F(\tilde{u}^k) + (\tilde{u}^k - u^*)^T \beta F(\tilde{u}^k).$$

Since F is monotone and u^* is a solution of VI(Ω, F), we have

$$(\tilde{u}^k - u^*)^T \beta F(\tilde{u}^k) \ge (\tilde{u}^k - u^*)^T \beta F(u^*) \ge 0.$$

Substituting it in the right hand side of (3.22), it follows that

$$\zeta_k^N(\alpha) \ge \|u^{k+1}_{BLD}(\alpha) - u^k\|^2 + 2\alpha(u^{k+1}_{BLD}(\alpha) - \tilde{u}^k)^T \beta F(\tilde{u}^k). \tag{3.23}$$

Since $u^{k+1}_{BLD}(\alpha) \in \Omega$, replacing the any $u \in \Omega$ in (3.7) by $u^{k+1}_{BLD}(\alpha)$, we get

$$(u^{k+1}_{BLD}(\alpha) - \tilde{u}^k)^T \beta F(\tilde{u}^k) \ge (u^{k+1}_{BLD}(\alpha) - \tilde{u}^k)^T d(u^k, \tilde{u}^k).$$

Substituting it in the right hand side of (3.23),

$$\begin{aligned}
\zeta_k^N(\alpha) &\ge \|u^{k+1}_{BLD}(\alpha) - u^k\|^2 + 2\alpha(u^{k+1}_{BLD}(\alpha) - \tilde{u}^k)^T d(u^k, \tilde{u}^k) \\
&= \|u^{k+1}_{BLD}(\alpha) - u^k\|^2 + 2\alpha(u^{k+1}_{BLD}(\alpha) - u^k)^T d(u^k, \tilde{u}^k) \\
&\qquad + 2\alpha(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \qquad (3.24)
\end{aligned}$$

By using the notations $q_k^N(\alpha)$ (see (3.12)) and $u_{BD}^{k+1}(\alpha)$ (3.8), we get

$$\begin{aligned}\zeta_k^N(\alpha) &\geq \|(u_{BLD}^{k+1}(\alpha)-u^k)+\alpha d(u^k,\tilde{u}^k)\|^2-\alpha^2\|d(u^k,\tilde{u}^k)\|^2 \\ &\qquad +2\alpha(u^k-\tilde{u}^k)^T d(u^k,\tilde{u}^k) \\ &= \|u_{BLD}^{k+1}(\alpha)-(u^k-\alpha d(u^k,\tilde{u}^k))\|^2+q_k^N(\alpha) \\ &= \|u_{BLD}^{k+1}(\alpha)-u_{BD}^{k+1}(\alpha)\|^2+q_k^N(\alpha).\end{aligned}$$

This is just the assertion (3.20) and the theorem is proved. □

Theorem 3.3 tells us that $q_k^N(\alpha)$ is also the lower bound of $\zeta_k^N(\alpha)$. In practical computation, with the same predictor $\tilde{u}^k$ given by (1.2) which satisfied (3.1), the corrector u^{k+1} is updated by

$$\text{(PC Method-I)} \qquad u_I^{k+1}=u_{BD}^{k+1}(\alpha)=u^k-\gamma\alpha_k^* d(u^k,\tilde{u}^k) \tag{3.25}$$

or

$$\text{(PC Method-II)} \qquad u_{II}^{k+1}=u_{BLD}^{k+1}(\alpha)=P_\Omega[u^k-\gamma\alpha_k^*\beta_k F(\tilde{u}^k)], \tag{3.26}$$

where α_k^* is given by (3.13), which is lower bounded from $1/2$.

Based on the same predictor, if we use the formula (3.25) to update u^{k+1}, the advantage is that the correction does not need to do an extra projection. However, in many practical problems, the cost of the projection on Ω (for example, $\Omega=\Re_+^n$ or a 'box') is not expensive. Thus, the correction formula (3.26) is often used.

By using theorem 3.2, no matter which of the twin methods is applied, the generated sequence $\{u^k\}$ satisfies

$$\|u^{k+1}-u^*\|^2 \leq \|u^k-u^*\|^2-\frac{1}{2}\gamma(2-\gamma)(1-\nu)\|u^k-\tilde{u}^k\|^2.$$

Based on this inequality, we can prove the following convergence theorem.

Theorem 3.4 *Assume that the operator F in VI(Ω,F) is Lipschitz continuous and its solution set Ω^* is nonempty. Then the sequence $\{u^k\}$ generated by* (3.25) *or* (3.26) *converges to some solution point of VI(Ω,F).*

The projection contraction algorithms introduced in this section have successfully applied to solve many geotechnical engineering problems [12, 13].

4 Applications and numerical experiments

We use examples of linear and nonlinear variational inequalities to illustrate the efficiency of the twin algorithms — PC Method-I and PC Method-II.

4.1 Applied the different PC Methods for LVI

For LVI, we use the ”sum of the shortest distance” mentioned in Lecture 1 as an example. This problem is equivalent to a min-max problem whose corresponding LVI with a skew symmetric matrix. For a detailed description of this kind of problem, see §5 of Lecture 1.
The test examples taking from SIAM J. on Optimization.

- G. L. Xue and Y. Y. Ye, *An efficient algorithm for minimizing a sum of Euclidean norms with applications*, SIAM Optim. 7 (1997), 1017-1039.

Fig.1 depicters the structure of the network, where $b_{[i]}, i = 1, \ldots, 10$ are regular points whose coordinates are given. The connection between $x_{[j]}, j = 1, \ldots, 8$ and $b_{[i]}$ are also given. Fig. 2 gives the positions of $x_{[j]}, j = 1, \ldots, 8$ when the sum of the shortest distance is reached.

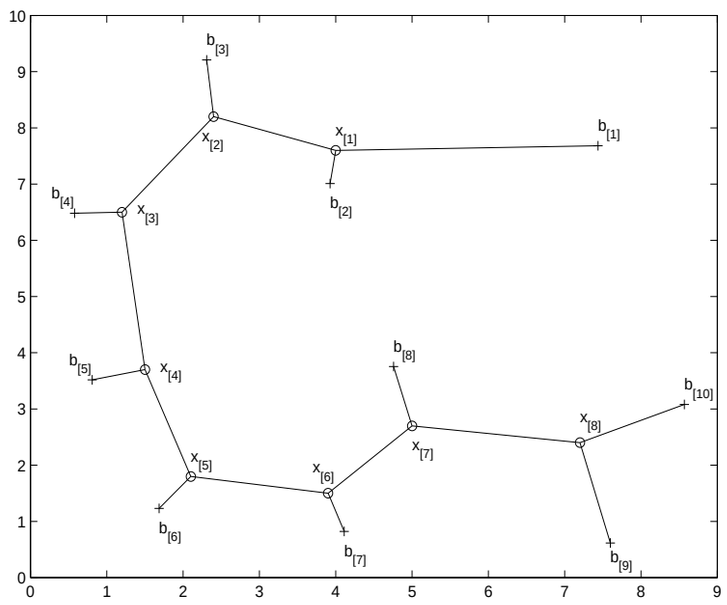


Fig. 1 Ordering of the topology

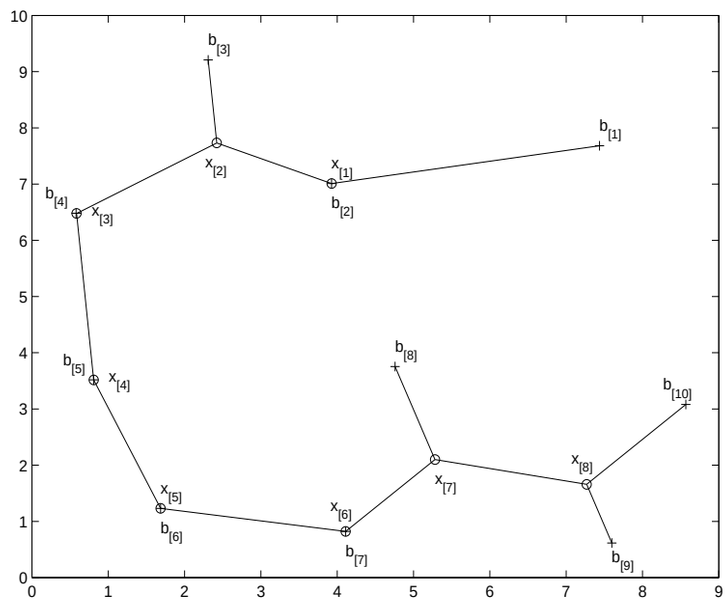


Fig. 2. Optimal solution in Euclidean-norm

The coordinates of the 10 regular points

	x-coordinate	y-coordinate		x-coordinate	y-coordinate
$b_{[1]}$	7.436490	7.683284	$b_{[6]}$	1.685912	1.231672
$b_{[2]}$	3.926097	7.008798	$b_{[7]}$	4.110855	0.821114
$b_{[3]}$	2.309469	9.208211	$b_{[8]}$	4.757506	3.753666
$b_{[4]}$	0.577367	6.480938	$b_{[9]}$	7.598152	0.615836
$b_{[5]}$	0.808314	3.519062	$b_{[10]}$	8.568129	3.079179

The update forms of using the contraction method I (3.25) and II (3.26) are

$$\text{(PC Method-I)} \qquad u^{k+1} = u^k - \gamma\alpha_k^*(I + M^T)(u^k - \tilde{u}^k),$$

and

$$\text{(PC Method-II)} \qquad u^{k+1} = P_\Omega\{u^k - \gamma\alpha_k^*[M^T(u^k - \tilde{u}^k) + (Mu^k + q)]\},$$

respectively. The numerical results are listed in the following table.

Table 1. Shortest network under l_2 norm.

PC Method-I			PC Method-II		
Iteration	$\|e(u)\|_\infty$	Total Distance	Iteration	$\|e(u)\|_\infty$	Total Distance
40	7.1e-002	25.3776304969	40	5.0e-004	25.3563526162
80	1.8e-004	25.3561050662	80	4.0e-008	25.3560677986
120	6.4e-007	25.3560678958	106	9.2e-011	25.3560677793
160	2.4e-009	25.3560677797			
183	9.5e-011	25.3560677793			
CPU-time		0.234 Sec.	CPU-time		0.125 Sec.

Here we take $\gamma = 1.8$. If $\gamma = 1$, 80% more iterations are needed for both methods.

PC Method-II for the problem in the Euclidean-norm

```
clear; % Steiner Minimum Tree * Read the coordinate of the regular points%(1)
 P1=[7.436490, 3.926097, 2.309469, 0.577367, 0.808314;                   %(2)
     7.683284, 7.008798, 9.208211, 6.480938, 3.519062];                  %(3)
 P2=[1.685912, 4.110855, 4.757506, 7.598152, 8.568129;                   %(4)
     1.231672, 0.821114, 3.753666, 0.615836, 3.079179];                  %(5)
 b=[P1,P2,zeros(2,7)]; x=zeros(2,8); z=zeros(2,17);  eps=1;   k=0; tic;  %(6)
while (eps > 10^(-10) & k<= 200)  k=k+1;   %% Beginning of an iteration  %(7)
 Ax=[x(:,1), x, x(:,8), x(:,1:7)-x(:,2:8)];   Axb=Ax-b;  %% Compute Ax-b %(8)
 ATz=z(:,2:9) + [z(:,1), -z(:,11:17)] + [z(:,11:17), z(:,10)];   % A^Tz  %(9)
 L2=0; for j=1:17  L2=L2 + norm(Axb(:,j),2); end;            % Length-2 %(10)
 if mod(k,20)==0 fprintf('k=%3d stopc=%9.1e L2=%13.10f\n',k,eps,L2);end;%(11)
 Pz=z+Axb; Dp=diag(1./max(1,sqrt(diag(Pz'*Pz)))); Pz=Pz*Dp; %P(z+(Ax-b))%(12)
 Ex = ATz;   Ez = z-Pz;  t=trace(Ex'*Ex)+ trace(Ez'*Ez);  eps=sqrt(t);  %(13)
 AEx= [Ex(:,1), Ex, Ex(:,8), Ex(:,1:7)-Ex(:,2:8)];      % Compute AEx   %(14)
 ATEz=Ez(:,2:9) + [Ez(:,1),-Ez(:,11:17)] + [Ez(:,11:17),Ez(:,10)]; %ATEz%(15)
  ta = trace(AEx'*AEx)+trace(ATEz'*ATEz);  alpha=t*1.8/(t+ta); %% Step L%(16)
  x =x-(ATz - ATEz)*alpha;                                   %% New x and z %%(17)
  z =z-(AEx - Axb)*alpha; Dz=diag(1./max(1,sqrt(diag(z'*z)))); z=z*Dz; %%(18)
 end;                                              %% End of an iteration %%(19)
 toc; fprintf(' k=%3d    eps=%9.1e   Length-2=%13.10f \n', k,eps,L2);  %%(20)
```

把上面第(18) 行改成 `z =z-(AEx + Ez)*alpha;` 就是处理同一问题的(CM-D1)程序.

Fig. 3 and 4 depict the convergence tendencies of Contraction Method–2 for the minimum sum of the distance in the Euclidean-norm with different starting points.

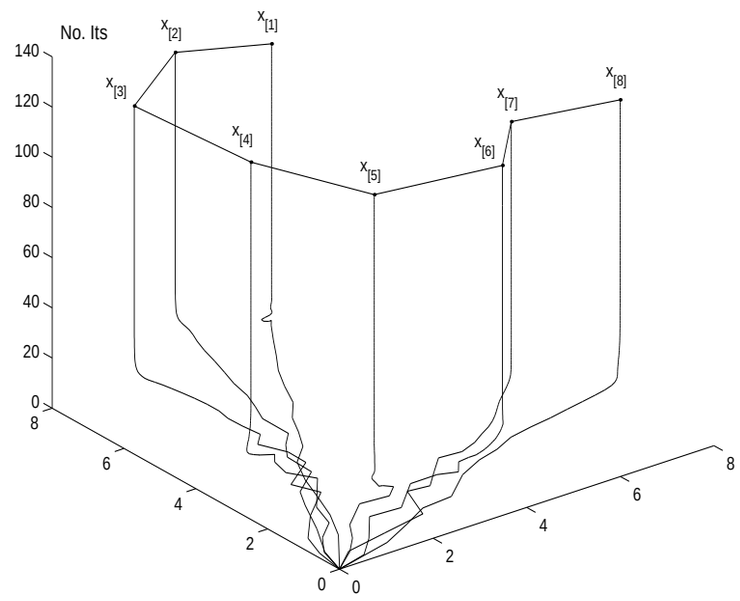


Fig. 3. Convergence tendency, $x^0 = 0$

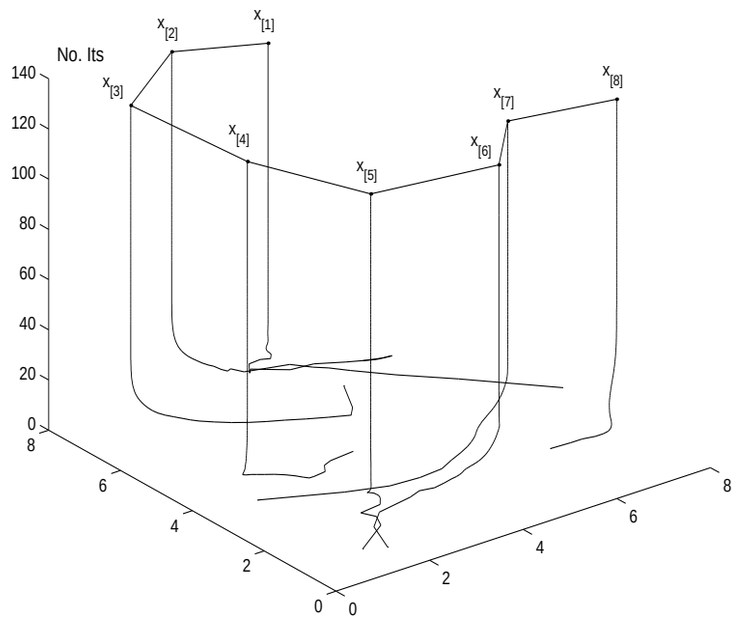


Fig. 4. Convergence tendency, x^0 random

对 l_1-模和 l_∞-模, 我们也用两种不同方法做了计算比较, 计算结果如下：

Table 2. Shortest network under l_1 norm.

PC Method-I			PC Method-II		
Iteration	$\|e(u)\|_\infty$	Total Distance	Iteration	$\|e(u)\|_\infty$	Total Distance
40	3.7e-002	28.6777786413	40	1.1e-004	28.6660178525
80	2.5e-005	28.6658649129	81	1.0e-010	28.6658580000
120	1.8e-008	28.6658580046			
149	9.4e-011	28.6658580000			
CPU-time		0.031 Sec.	CPU-time		0.016 Sec.

Table 3. Shortest network under l_∞ norm.

PC Method-I			PC Method-II		
Iteration	$\|e(u)\|_\infty$	Total Distance	Iteration	$\|e(u)\|_\infty$	Total Distance
40	9.0e-002	21.1322990353	40	2.1e-003	21.1145131146
80	4.4e-005	21.1129244226	80	4.1e-010	21.1129135002
120	2.4e-008	21.1129135060	84	7.4e-011	21.1129135000
150	9.2e-011	21.1129135000			
CPU-time		0.187 Sec.	CPU-time		0.094 Sec.

Fig. 5 and Fig. 6 depict the optimal solutions of the minimum sum of the distance in the l_1-norm and l_∞-norm, respectively.

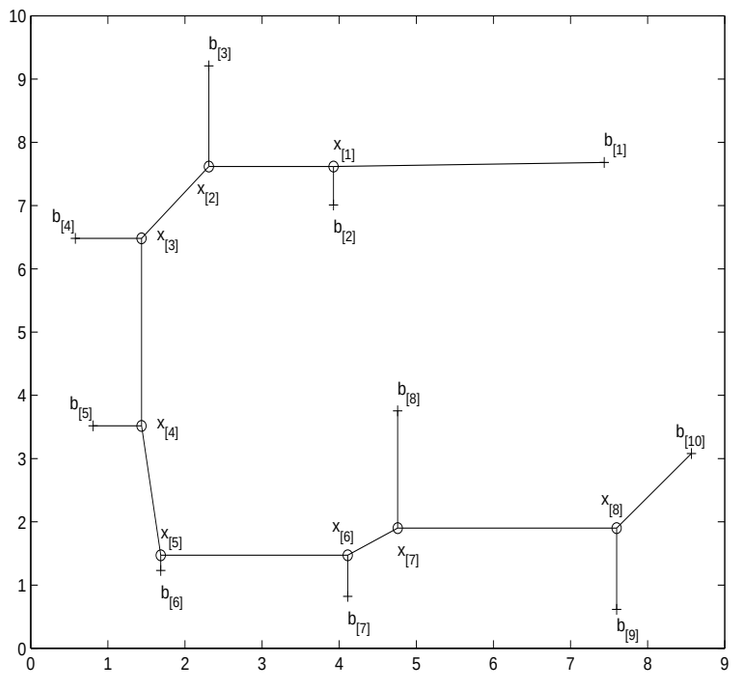


Fig. 5. Optimal solution, l_1-norm

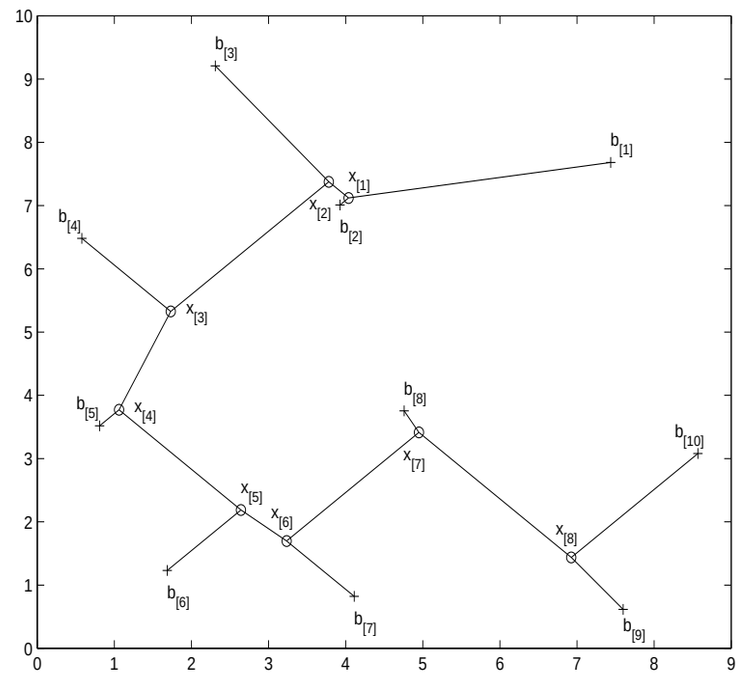


Fig. 6. Optimal solution, l_∞-norm

Fig. 7 and 8 depict the convergence tendencies of Contraction Method–2 with random starting points for the minimum sum of the distance in the in the l_1-norm and l_∞-norm, respectively.

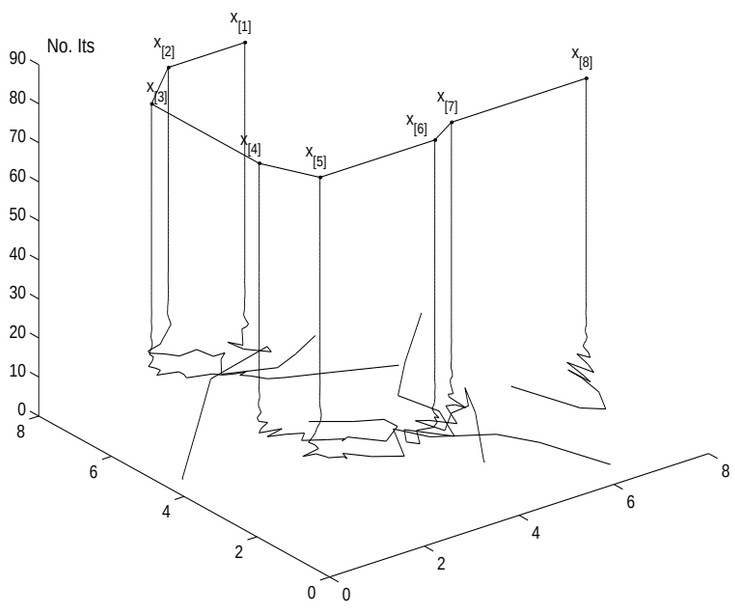


Fig. 7. Convergence tendency, l_1-norm

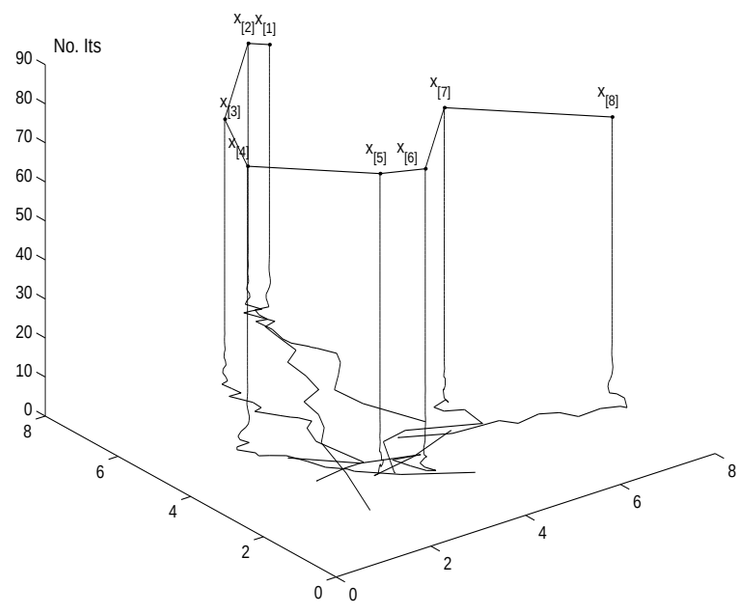


Fig. 8. Convergence tendency, l_∞-norm

4.2 Test examples of the NCP

For comparing the efficiency of PC Method-I and PC Method-II, we test the nonlinear complementarity problem (a class of VI(Ω, F) with $\Omega = \Re^n_+$)

$$u \geq 0, \qquad F(u) \geq 0, \qquad u^T F(u) = 0.$$

In the test examples, we take

$$F(u) = D(u) + Mu + q,$$

The linear part $Mu + q$ is generated by using Matlab, it produced by

```
A=(rand(n,n)-0.5)*10; B=(rand(n,n)-0.5)*10;  B=B-B'; M=A'*A+B;
q=(rand(n,1)-0.5)*1000;          or          q=(rand(n,1)-1.0)*500;
```

In the nonlinear part $D(u)$, each element is given by $D_j(u) = d_j * \arctan(u_j)$, where d_j is a random variable between $(0, 1)$.

We use the algorithms (3.25) and (3.26) in §3.3 to solve the test problems.

Notice that PC Method-I is just the PC Algorithm in Lecture 2 for NVI.

Set $\gamma\alpha_k^* \equiv 1$ in PC Method-II, it reduced to the Refined EG.

In all the tests, each element of the initial u^0 is a random variable in $(0, 10)$.

PC Method-II 的程序

PC Method-II:

Step 0. Set $\beta_0 = 1, \nu \in (0, 1), u^0 \in \Omega$ and $k = 0$.

Step 1. $\tilde{u}^k = P_\Omega[u^k - \beta_k F(u^k)]$,

$r_k := \beta_k \|F(u^k) - F(\tilde{u}^k)\| / \|u^k - \tilde{u}^k\|$,

while $r_k > \nu$, $\beta_k := \frac{2}{3}\beta_k * \min\{1, \frac{1}{r_k}\}$,

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

$r_k := \beta_k \|F(u^k) - F(\tilde{u}^k)\| / \|u^k - \tilde{u}^k\|$,

end(while)

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

$$\alpha_k = \frac{(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2},$$

$u^{k+1} = P_\Omega[u^k - \gamma\alpha_k\beta_k F(\tilde{u}^k)]$,

If $r_k \leq \mu$ then $\beta_k := \beta_k * 1.5$, **end(if)**

Step 2. $\beta_{k+1} = \beta_k$ and $k = k + 1$, go to Step 1.

From PC Method-I to PC Method-II, 只是将

$u^{k+1} = u^k - \gamma\alpha_k d(u^k, \tilde{u}^k)$ 改成了 $u^{k+1} = P_\Omega[u^k - \gamma\alpha_k\beta_k F(\tilde{u}^k)]$.

Matlab Code of Contraction Method–D2 for NCP

```
function PC_G(n,M,q,d,xstart,tol,pfq)                                   %(1)
fprintf('PC Method use Direction D1 with gamma a*  n=%4d \n',n);        %(2)
x=xstart;  Fx= d.*atan(x) + M*x + q;  stopc=norm(x-max(x-Fx,0),inf);    %(3)
beta=1;        k=0;        l=0;       tic;                             %(4)
while (stopc>tol && k<=2000)                                            %(5)
  if mod(k,pfq)==0 fprintf(' k=%4d    epsm=%9.3e  \n',k,stopc); end;    %(6)
    x0=x;   Fx0=Fx;    k=k+1;                                           %(7)
    x=max(x0-Fx0*beta,0);        Fx=d.*atan(x) + M*x + q;   l=l+1;      %(8)
       dx=x0-x;   df=(Fx0-Fx)*beta;                                     %(9)
       r=norm(df)/norm(dx);                                            %(10)
       while r>0.9     beta=0.7*beta*min(1,1/r);   l=l+1;              %(11)
         x=max(x0-Fx0*beta,0);    Fx=d.*atan(x) + M*x + q;             %(12)
         dx=x0-x;   df=(Fx0-Fx)*beta;     r=norm(df)/norm(dx);         %(13)
       end;                                                            %(14)
       dxf=dx-df;    r1=dx'*dxf;    r2=dxf'*dxf;     alpha=r1/r2;      %(15)
       x=max(x0- Fx*beta*alpha*1.9,0);                                 %(16)
       Fx= d.*atan(x) + M*x + q;     l=l+1;                            %(17)
       ex=x-max(x-Fx,0);     stopc=norm(ex,inf);                       %(18)
       if r <0.4  beta=beta*1.5;  end;                                 %(19)
end;  toc;   fprintf(' k=%4d   epsm=%9.3e     l=%4d  \n',k,stopc,l); %%%%
```


NCP 的计算结果 1 Easy Problems $q \in (-500, 500)$

	PC Method-I			PC Method-II		
$n =$	No. It	No. F	CPU	No. It	No. F	CPU
500	448	941	0.15	372	792	0.12
1000	475	995	1.37	410	852	1.17
1500	507	1064	3.17	416	887	2.64
2000	515	1080	5.53	418	892	4.55

NCP 的计算结果 2 Hard Problems $q \in (-500, 0)$

	PC Method-I			PC Method-II		
$n =$	No. It	No. F	CPU	No. It	No. F	CPU
500	908	1913	0.30	799	1704	0.27
1000	980	2068	2.87	857	1824	2.53
1500	941	1983	5.88	834	1771	5.25
2000	1112	2352	12.18	986	2105	10.87

PC Method-II converges faster than PC Method-I.

♣ 程序在附件的 Codes-03 中：运行 demo.m 输入 n 就可以, 其中也可以选择不同问题类型. PCd1.m 和 PCd2.m 分别是 PC Method-I 和 PC Method-II 的子程序.

5 更一般的孪生方法的统一框架

我们在 [8, 9] 中给出了求解变分不等式更一般的孪生收缩算法.

首先, 定义了预测点(或称检验点) . 对给定的 u^k, 根据一定法则生成的 $\tilde{u}^k \in \Omega$ 说成是一个预测点, 如果 $\boxed{u^k = \tilde{u}^k \quad \Leftrightarrow \quad u^k \in \Omega^*.}$

例如, 对给定的 u^k 和 $\beta > 0$, 由投影 $\boxed{\tilde{u}^k = P_\Omega[u^k - \beta F(u^k)]}$ 给出的 $\tilde{u}^k$ 是按确定的法则给出的, 它是一个预测点, 但这不是给出预测点的惟一方法.

统一框架. 对给定的 u^k, 设 $\tilde{u}^k \in \Omega$ 是 u^k 的一个预测点. 设有基于 $(u^k, \tilde{u}^k)$ 的一对孪生的方向 $d_1(u^k, \tilde{u}^k)$, $d_2(u^k, \tilde{u}^k)$ 和误差度量函数 $\varphi(u^k, \tilde{u}^k) \ge 0$, 它们满足以下条件:

1. 它们满足关系式

$$\tilde{u}^k \in \Omega, \quad (u - \tilde{u}^k)^T d_2(u^k, \tilde{u}^k) \ge (u - \tilde{u}^k)^T d_1(u^k, \tilde{u}^k), \quad \forall u \in \Omega. \tag{5.1a}$$

2. 存在常数 $K > 0$, 使得

$$\|d_1(u^k, \tilde{u}^k)\| \le K\|u^k - \tilde{u}^k\|. \tag{5.1b}$$

3. 对任意的 $u^* \in \Omega^*$, 有

$$(\tilde{u}^k - u^*)^T d_2(u^k, \tilde{u}^k) \ge \varphi(u^k, \tilde{u}^k) - (u^k - \tilde{u}^k)^T d_1(u^k, \tilde{u}^k), \tag{5.1c}$$

4. $\varphi(u^k, \tilde{u}^k)$ 是 VI(Ω, F) 的误差度量函数, 即存在常数 $\delta > 0$, 使得

$$\varphi(u^k, \tilde{u}^k) \ge \delta\|u^k - \tilde{u}^k\|^2 \qquad \& \qquad \varphi(u^k, \tilde{u}^k) = 0 \;\Leftrightarrow\; u^k = \tilde{u}^k. \tag{5.1d}$$

对误差度量函数 $\varphi(u^k, \tilde{u}^k)$ 而言, $d_1(u^k, \tilde{u}^k), d_2(u^k, \tilde{u}^k)$ 都是有利方向.

Lemma 5.1 如果统一框架中的条件 (5.1a) 和 (5.1c) 满足, 则有

$$(u^k - u^*)^T d_1(u^k, \tilde{u}^k) \ge \varphi(u^k, \tilde{u}^k), \quad \forall u^k \in \Re^n, u^* \in \Omega^*. \tag{5.2}$$

证明. 因为 $u^* \in \Omega$, 以 u^* 代 (5.1a) 中的 u, 就有

$$(\tilde{u}^k - u^*)^T d_1(u^k, \tilde{u}^k) \ge (\tilde{u}^k - u^*)^T d_2(u^k, \tilde{u}^k).$$

再根据条件 (5.1c), 得到

$$(\tilde{u}^k - u^*)^T d_1(u^k, \tilde{u}^k) \ge \varphi(u^k, \tilde{u}^k) - (u^k - \tilde{u}^k)^T d_1(u^k, \tilde{u}^k)$$

从上式直接得到 (5.2), 引理得证. $\square$

Lemma 5.2 如果统一框架中的条件 (5.1a) 和 (5.1c) 满足, 则有

$$(u^k - u^*)^T d_2(u^k, \tilde{u}^k) \geq \varphi(u, \tilde{u}), \quad \forall u^k \in \Omega, u^* \in \Omega^*. \tag{5.3}$$

证明. 因为 $u^k \in \Omega$, 以 u^k 代 (5.1a) 中的 u, 我们有

$$(u^k - \tilde{u}^k)^T d_2(u^k, \tilde{u}^k) \geq (u^k - \tilde{u}^k)^T d_1(u^k, \tilde{u}^k). \tag{5.4}$$

将 (5.4) 和 (5.1c) 相加, 得到

$$(u^k - u^*)^T d_2(u^k, \tilde{u}^k) \geq \varphi(u^k, \tilde{u}^k).$$

引理得证. □

根据提供的 $d_1(u^k, \tilde{u}^k)$ 和 $d_2(u^k, \tilde{u}^k)$ 一对孪生方向, 我们可以构造一对算法

(Contraction Method-I) $\quad u^{k+1} = u^k - \gamma\alpha_k^* d_1(u^k, \tilde{u}^k),$

(Contraction Method-II) $\quad u^{k+1} = P_\Omega\{u^k - \gamma\alpha_k^* d_2(u^k, \tilde{u}^k)\},$

其中 $\alpha_k^* = \varphi(u^k, \tilde{u}^k)/\|d_1(u^k, \tilde{u}^k)\|^2, \gamma \in (0, 2)$, 由(5.1b)和(5.1d), 步长是有界的.

孪生方向, 相同步长, 是 PC 方法中最优美的篇章. 证明需技巧, 使用很方便!

References

[1] E. Blum and W. Oettli, Mathematische Optimierung. Grundlagen und Verfahren. Ökonometrie und Unternehmensforschung. Berlin-Heidelberg-New York: Springer-Verlag, 1975.

[2] D. S. Guo and Y.N. Zhang, Simulation and experimental verification of weighted velocity and acceleration minimization for robotic redundancy resolution, *IEEE Transactions on Automation Science and Engineering*, 2014, 11: 1203–1217.

[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 globally linearly convergent projection and contraction method for a class of linear complementarity problems. Schwerpunktprogramm der DFG Anwendungsbezogene Optimierung und Steuerung, No. 352, 1992

[6] B.S. He, A class of projection and contraction methods for monotone variational inequalities, *Applied Mathematics and optimization*, **35**, 69–76, 1997.

[7] B.S He and L.-Z Liao, Improvements of some projection methods for monotone nonlinear variational inequalities, JOTA, **112**, 111-128, 2002

[8] B.S. He, L.Z. Liao, and X. Wang, Proximal-like contraction methods for monotone variational inequalitiesin a unified framework I: Effective quadruplet and primary methods, Comput. Optim. Appl., **51**, 649-679, 2012.

[9] 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. **51**, 681-708, 2012

[10] B.S. He, X.M. Yuan and J.J.Z. Zhang, *Comparison of two kinds of prediction-correction methods for monotone variational inequalities*, Computational Optimization and Applications, **27**, 247-267, 2004.

[11] L. Xiao and Y. N. Zhang, Acceleration-level repetitive motion planning and its experimental verification on six-link planar robot manipulator, *IEEE Transactions on Control System Technology*, 2013, 21: 906–914.

[12] 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, 295 (2015) 150-171.

[13] H. Zheng, P. Zhang and X.L. Du, Dual form of discontinuous deformation analysis, Computer Methods in Applied Mechanics and Engineering, 305 (2016) 196-216.

凸优化和单调变分不等式的收缩算法

第四讲: 线性单调变分不等式的一对孪生投影收缩算法的收敛速率

Convergence rate of the twin PC methods for monotone linear variational inequalities

南京大学数学系　何炳生
hebma@nju.edu.cn

The context of this lecture is based on the publication [2]

For solving the monotone linear variational inequalities, in 1994, we have proposed a class of projection and contraction methods [6, 7]. Before 2015, only convergence results are available to these methods. In this article, we prove the convergence rate of such projection and contraction methods.

1 Introduction

Let Ω be a closed convex subset of $\Re^n$, $M \in \Re^{n\times n}$ and $q \in \Re^n$. The linear variational inequality problem, denoted by LVI(Ω, M, q), is to find a vector $u^* \in \Omega$ such that

$$\text{LVI}(\Omega, M, q) \qquad (u-u^*)^T(Mu^*+q) \ge 0, \qquad \forall u \in \Omega. \tag{1.1}$$

Throughout this paper we assume that the matrix M is positive semi-definite (but not necessary symmetric), *i.e.*,

$$(u-v)^T M(u-v) \ge 0, \quad \forall\, u, v \in \Re^n.$$

Moreover, we assume that the solution set of LVI(Ω, M, q), denoted by Ω^*, is nonempty. The nonempty assumption of the solution set, together with the positivity assumption of the

matrix M, implies that Ω^* is closed and convex (see pp. 158 in [3]).

For any $\beta > 0$, the solution set of the following linear variational inequality

$$\text{LVI}(\Omega, M, q) \qquad (u - u^*)^T \beta(Mu^* + q) \geq 0, \qquad \forall u \in \Omega.$$

is coincides with the one of (1.1). It is well known that,

$$u^* \text{ is a solution of LVI}(\Omega, M, q) \quad \Longleftrightarrow \quad u^* = P_\Omega[u^* - \beta(Mu^* + q)], \tag{1.2}$$

where β is any positive scalar and $P_\Omega(\cdot)$ denotes the projection onto Ω with respect to the Euclidean norm, *i.e.*,

$$P_\Omega(v) = \text{argmin}\{\frac{1}{2}\|u - v\|^2 \mid u \in \Omega\}.$$

For convergence rate investigation, without loss of generality, we fix $\beta = 1$. Throughout this article, we assume that the projection on Ω in the Euclidean-norm has a closed form and it is easy to be carried out. The most important property of the projection mapping is

$$(v - P_\Omega(v))^T (u - P_\Omega(v)) \leq 0, \quad \forall\, v \in \Re^n, \forall\, u \in \Omega. \tag{1.3}$$

A pair of twin projection and contraction methods

For simpleness of the theoretical analysis, without loss the generality, we let $\beta \equiv 1$. For given $u^k \in \Re^n$, the predictor is given by

$$\tilde{u}^k = P_\Omega[u^k - (Mu^k + q)]. \tag{1.4}$$

According to (1.2), u is a solution of LVI(Ω, M, q) if and only if $u^k = \tilde{u}^k$. In [6, 7], we define $e(u^k) = u^k - \tilde{u}^k$ and use $\|e(u^k)\|$ to measure how much u^k fails to be a solution of the linear variational inequality LVI(Ω, M, q).

By using (1.4), the projection and contraction method [6] can be described as

$$\text{(PCM-I)} \qquad u^{k+1} = u^k - \gamma\alpha_k^*(I + M^T)(u^k - \tilde{u}^k), \tag{1.5}$$

where $\gamma \in (0, 2)$ and

$$\alpha_k^* = \frac{\|u^k - \tilde{u}^k\|^2}{\|(I + M^T)(u^k - \tilde{u}^k)\|^2}. \tag{1.6}$$

It was proved that (see Lecture 2 or 3 of this seriers)

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - q_k(\gamma), \tag{1.7a}$$

where

$$q_k(\gamma) = \gamma(2-\gamma)\alpha_k^* \|u^k - \tilde{u}^k\|^2. \tag{1.7b}$$

In an earlier preprint [5], it was pointed out that the method

$$\text{(PCM-II)} \qquad u^{k+1} = P_\Omega\{u^k - \gamma\alpha_k^*[(Mu^k+q) + M^T(u^k - \tilde{u}^k)]\}, \tag{1.8}$$

has the same contractive property as (1.7), the result was reported in [7] and detailed proof can be found in [8]. Note that the same step size length is used in (1.5) and (1.8) even if the search directions are different. Thus, this pair of methods are called **a pair of twin projection and contraction methods for linear variational inequalities**.

The projection and contraction methods make one (or two) projection(s) on Ω at each iteration, and the distance of the iterates to the solution set monotonically converges to zero. According to the terminology in [1], these methods belong to the class of Fejér contraction methods. Stimulated by the complexity analysis [14], this article shows the $O(1/t)$ convergence rate of the projection and contraction methods for monotone linear variational inequalities.

2 Preliminaries

Recall that Ω^* can be characterized as (see (2.3.2) in pp. 159 of [3])

$$\Omega^* = \bigcap_{u\in\Omega} \{\tilde{u} \in \Omega : (u-\tilde{u})^T(Mu+q) \ge 0\}.$$

This implies that $\tilde{u} \in \Omega$ is an approximate solution of LVI(Ω, M, q) with the accuracy ϵ if it satisfies

$$\tilde{u} \in \Omega \quad \text{and} \quad \inf_{u\in\mathcal{D}(\tilde{u})} \{(u-\tilde{u})^T(Mu+q)\} \ge -\epsilon,$$

where

$$\mathcal{D}(\tilde{u}) = \{u \in \Omega \,|\, \|u-\tilde{u}\| \le 1\}.$$

In this article, we show that, for given $\epsilon > 0$, in $O(1/\epsilon)$ iterations the pair of twin projection and contraction methods can offer a $\tilde{u}$ such that

$$\tilde{u} \in \Omega \quad \text{and} \quad \sup_{u\in\mathcal{D}(\tilde{u})} \{(\tilde{u}-u)^T(Mu+q)\} \le \epsilon. \tag{2.1}$$

Since the term $(\tilde{u}-u)^T(Mu+q)$ is essential in the approximate criterion (2.1) and the search direction in PCM-II is $[(Mu^k+q) + M^T(u^k-\tilde{u}^k)]$, now, we establish a

relation between

$$(u-\tilde{u}^k)^T(Mu+q) \quad \text{and} \quad (u-\tilde{u}^k)^T[(Mu^k+q)+M^T(u^k-\tilde{u}^k)].$$

Lemma 2.1 *For given u^k, $\tilde{u}^k$ is given by* (1.4)*. Then for any u, the inequality*

$$\begin{aligned}&(u-\tilde{u}^k)^T(Mu+q)\\ &\quad\geq\quad (u-\tilde{u}^k)^T[(Mu^k+q)+M^T(u^k-\tilde{u}^k)]-\frac{1}{2}\|u^k-\tilde{u}^k\|_D^2 \qquad (2.2)\end{aligned}$$

is always true, where $D=M^T+M$ is symmetric and positive semi-definite.

Proof. By calculating the difference and using the Cauchy-Schwarz Inequality, we get

$$\begin{aligned}&(u-\tilde{u}^k)^T(Mu+q)-(u-\tilde{u}^k)^T[(Mu^k+q)+M^T(u^k-\tilde{u}^k)]\\ &\quad=\quad (u-\tilde{u}^k)^T\{M(u-u^k)-M^T(u^k-\tilde{u}^k)\}\\ &\quad=\quad (u-\tilde{u}^k)^T\{M(u-\tilde{u}^k)-(M+M^T)(u^k-\tilde{u}^k)\}\\ &\quad=\quad \frac{1}{2}\|u-\tilde{u}^k\|_D^2-(u-\tilde{u}^k)^TD(u^k-\tilde{u}^k)\\ &\quad\geq\quad -\frac{1}{2}\|u^k-\tilde{u}^k\|_D^2.\end{aligned}$$

Consequently, we get (2.2) and the assertion is proved. □

3 Important theorems for the twin PC methods

Now, we prove the same key inequality of the twin PC Methods for the complexity analysis. The assertion (2.2) is useful for the proofs of the following theorems, even though the proof is relative simple.

Set $v=u^k-(Mu^k+q)$ in (1.3), since $\tilde{u}^k=P_\Omega[u^k-(Mu^k+q)]$, it follows from (1.3) that

$$\{[u^k-(Mu^k+q)]-\tilde{u}^k\}^T(u-\tilde{u}^k)\leq 0, \quad \forall u\in\Omega$$

and consequently,

$$(u-\tilde{u}^k)^T(Mu^k+q)\geq(u-\tilde{u}^k)^T(u^k-\tilde{u}^k), \quad \forall\, u\in\Omega.$$

Adding the term $(u-\tilde{u}^k)^TM^T(u^k-\tilde{u}^k)$ to the both sides of the above inequality, we obtain

$$\begin{aligned}&(u-\tilde{u}^k)^T\{(Mu^k+q)+M^T(u^k-\tilde{u}^k)\}\\ &\quad\geq\quad (u-\tilde{u}^k)^T(I+M^T)(u^k-\tilde{u}^k),\ \forall\, u\in\Omega. \qquad (3.1)\end{aligned}$$

Main theorem for PC-Method-I

Theorem 3.1 *For given* $u^k \in \Re^n$*, let* $\tilde{u}^k$ *be defined by* (1.4) *and the new iterate* u^{k+1} *be generated by PCM-I* (1.5) *with any* $\gamma > 0$*. Then we have*

$$\begin{aligned} &\gamma\alpha_k^*(u-\tilde{u}^k)^T(Mu+q) \\ &\quad\geq \quad \frac{1}{2}\big(\|u-u^{k+1}\|^2 - \|u-u^k\|^2\big) + \frac{1}{2}q_k(\gamma), \;\; \forall\, u \in \Omega, \end{aligned} \tag{3.2}$$

where $q_k(\gamma)$ *is defined in* (1.7b)*.*

Proof. Substituting (3.1) in the right hand side of (2.2), we get

$$\begin{aligned} &(u-\tilde{u}^k)^T(Mu+q) \\ &\quad\geq \quad (u-\tilde{u}^k)^T(I+M^T)(u^k-\tilde{u}^k) - \frac{1}{2}\|u^k-\tilde{u}^k\|_D^2. \end{aligned}$$

Thus, we have

$$\begin{aligned} &\gamma\alpha_k^*(u-\tilde{u}^k)^T(Mu+q) \\ &\quad\geq \quad (u-\tilde{u}^k)^T\gamma\alpha_k^*(I+M^T)(u^k-\tilde{u}^k) - \frac{\gamma\alpha_k^*}{2}\|u^k-\tilde{u}^k\|_D^2, \end{aligned}$$

Because $\gamma\alpha_k^*(I+M^T)(u^k-\tilde{u}^k) = u^k - u^{k+1}$ (see (1.5)), we get

$$\begin{aligned} &\gamma\alpha_k^*(u-\tilde{u}^k)^T(Mu+q) \\ &\quad\geq \quad (u-\tilde{u}^k)^T(u^k-u^{k+1}) - \frac{\gamma\alpha_k^*}{2}\|u^k-\tilde{u}^k\|_D^2. \end{aligned} \tag{3.3}$$

To the crossed term in the right hand side of (3.3), namely $(u-\tilde{u}^k)^T(u^k-u^{k+1})$, using an identity

$$(a-b)^T(c-d) = \frac{1}{2}\big(\|a-d\|^2 - \|a-c\|^2\big) + \frac{1}{2}\big(\|c-b\|^2 - \|d-b\|^2\big),$$

we obtain

$$\begin{aligned} (u-\tilde{u}^k)^T(u^k-u^{k+1}) \quad &= \quad \frac{1}{2}\big(\|u-u^{k+1}\|^2 - \|u-u^k\|^2\big) \\ &\quad + \frac{1}{2}\big(\|u^k-\tilde{u}^k\|^2 - \|u^{k+1}-\tilde{u}^k\|^2\big). \end{aligned} \tag{3.4}$$

By using $u^{k+1} = u^k - \gamma\alpha_k^*(I + M^T)(u^k - \tilde{u}^k)$ and (1.6), we get

$$\begin{aligned}
&\|u^k - \tilde{u}^k\|^2 - \|u^{k+1} - \tilde{u}^k\|^2 \\
&\quad = \quad \|u^k - \tilde{u}^k\|^2 - \|(u^k - \tilde{u}^k) - \gamma\alpha_k^*(I + M^T)(u^k - \tilde{u}^k)\|^2 \\
&\quad = \quad 2\gamma\alpha_k^*(u^k - \tilde{u}^k)^T(I + M^T)(u^k - \tilde{u}^k) \\
&\qquad\qquad -\gamma^2\alpha_k^*\big(\alpha_k^*\|(I + M^T)(u^k - \tilde{u}^k)\|^2\big) \\
&\quad = \quad 2\gamma\alpha_k^*\|u^k - \tilde{u}^k\|^2 + \gamma\alpha_k^*\|u^k - \tilde{u}^k\|_D^2 - \gamma^2\alpha_k^*\|u^k - \tilde{u}^k\|^2 \\
&\quad = \quad \gamma(2-\gamma)\alpha_k^*\|u^k - \tilde{u}^k\|^2 + \gamma\alpha_k^*\|u^k - \tilde{u}^k\|_D^2.
\end{aligned}$$

Substituting it in the right hand side of (3.4) and using the definition of $q_k(\gamma)$, we obtain

$$\begin{aligned}
(u - \tilde{u}^k)^T(u^k - u^{k+1}) \quad &= \quad \frac{1}{2}\big(\|u - u^{k+1}\|^2 - \|u - u^k\|^2\big) + \frac{1}{2}q_k(\gamma) \\
&\quad + \frac{\gamma\alpha_k^*}{2}\|u^k - \tilde{u}^k\|_D^2. \qquad (3.5)
\end{aligned}$$

Adding (3.3) and (3.5), the theorem is proved. $\square$

Main theorem for PCM-II

The both sequences $\{\tilde{u}^k\}$ and $\{u^k\}$ in the PC method II belong to Ω. In the following lemma we prove the same assertion for PC method II as in Theorem 3.1.

Theorem 3.2 *For given $u^k \in \Re^n$, let $\tilde{u}^k$ be defined by* (1.4) *and the new iterate u^{k+1} be generated by PCM-II* (1.8) *with any $\gamma > 0$. Then we have*

$$\begin{aligned}
&\gamma\alpha_k^*(u - \tilde{u}^k)^T(Mu + q) \\
&\quad \geq \quad \frac{1}{2}\big(\|u - u^{k+1}\|^2 - \|u - u^k\|^2\big) + \frac{1}{2}q_k(\gamma), \ \ \forall\, u \in \Omega, \qquad (3.6)
\end{aligned}$$

where $q_k(\gamma)$ is defined in (1.7b)*.*

Proof. By using the assertion in Lemma 2.1, we need only to show that

$$\begin{aligned}
&(u - \tilde{u}^k)^T\gamma\alpha_k^*[(Mu^k + q) + M^T(u^k - \tilde{u}^k)] \\
&\quad \geq \quad \frac{1}{2}\big(\|u - u^{k+1}\|^2 - \|u - u^k\|^2\big) + \frac{1}{2}q_k(\gamma) \\
&\qquad + \frac{\gamma\alpha_k^*}{2}\|u^k - \tilde{u}^k\|_D^2, \ \ \forall\, u \in \Omega. \qquad (3.7)
\end{aligned}$$

For investigating the term in the left hand side of (3.7), we divide it in the terms

$$(u^{k+1} - \tilde{u}^k)^T \gamma \alpha_k^* [(Mu^k + q) + M^T (u^k - \tilde{u}^k)] \tag{3.8a}$$

and

$$(u - u^{k+1})^T \gamma \alpha_k^* [(Mu^k + q) + M^T (u^k - \tilde{u}^k)]. \tag{3.8b}$$

First, we deal with the term (3.8a). Since $u^{k+1} \in \Omega$, substituting $u = u^{k+1}$ in (3.1) we get

$$\begin{aligned}
&(u^{k+1} - \tilde{u}^k)^T \gamma \alpha_k^* [(Mu^k + q) + M^T (u^k - \tilde{u}^k)] \\
\geq\ & \gamma \alpha_k^* (u^{k+1} - \tilde{u}^k)^T (I + M^T)(u^k - \tilde{u}^k) \\
=\ & \gamma \alpha_k^* (u^k - \tilde{u}^k)^T (I + M^T)(u^k - \tilde{u}^k) \\
& -\gamma \alpha_k^* (u^k - u^{k+1})^T (I + M^T)(u^k - \tilde{u}^k) \\
=\ & \gamma \alpha_k^* \|u^k - \tilde{u}^k\|^2 + \frac{\gamma \alpha_k^*}{2} \|u^k - \tilde{u}^k\|_D^2 \\
& -\gamma \alpha_k^* (u^k - u^{k+1})^T (I + M^T)(u^k - \tilde{u}^k).
\end{aligned} \tag{3.9}$$

To the crossed term of the right hand side of (3.9), using the Cauchy-Schwarz Inequality and (1.6) we get

$$\begin{aligned}
&-\gamma \alpha_k^* (u^k - u^{k+1})^T (I + M^T)(u^k - \tilde{u}^k) \\
\geq\ & -\frac{1}{2} \|u^k - u^{k+1}\|^2 - \frac{1}{2} \gamma^2 (\alpha_k^*)^2 \|(I + M^T)(u^k - \tilde{u}^k)\|^2 \\
=\ & -\frac{1}{2} \|u^k - u^{k+1}\|^2 - \frac{1}{2} \gamma^2 \alpha_k^* \|u^k - \tilde{u}^k\|^2.
\end{aligned}$$

Substituting them in the right hand side of (3.9) and using the notation of $q_k(\gamma)$, we obtain

$$\begin{aligned}
&(u^{k+1} - \tilde{u}^k)^T \gamma \alpha_k^* [(Mu^k + q) + M^T (u^k - \tilde{u}^k)] \\
\geq\ & \frac{1}{2} q_k(\gamma) + \frac{\gamma \alpha_k^*}{2} \|u^k - \tilde{u}^k\|_D^2 - \frac{1}{2} \|u^k - u^{k+1}\|^2.
\end{aligned} \tag{3.10}$$

Now, we turn to treat of the term (3.8b). The update form of PC Method II (1.8) means that u^{k+1} is the projection of $\big(u^k - \gamma \alpha_k^* [(Mu^k + q) + M^T (u^k - \tilde{u}^k)]\big)$ on Ω. Thus, it follows from (1.3) that

$$\big\{ \big(u^k - \gamma \alpha_k^* [(Mu^k + q) + M^T (u^k - \tilde{u}^k)]\big) - u^{k+1} \big\}^T \big(u - u^{k+1}\big) \leq 0, \ \ \forall\, u \in \Omega,$$

and consequently

$$\big(u-u^{k+1}\big)^T\gamma\alpha_k^*[(Mu^k+q)+M^T(u^k-\tilde{u}^k)] \ge \big(u-u^{k+1}\big)^T\big(u^k-u^{k+1}\big), \ \ \forall\, u\in\Omega.$$

For the the last inequality, by using the identity

$$b^T(b-a)=\frac{1}{2}\big(\|b\|^2-\|a\|^2\big)+\frac{1}{2}\|a-b\|^2$$

with $a=u-u^k$ and $b=u-u^{k+1}$ to the right hand side, we obtain

$$\begin{aligned}&\big(u-u^{k+1}\big)^T\gamma\alpha_k^*[(Mu^k+q)+M^T(u^k-\tilde{u}^k)]\\ &\quad\ge\ \frac{1}{2}\big(\|u-u^{k+1}\|^2-\|u-u^k\|^2\big)+\frac{1}{2}\|u^k-u^{k+1}\|^2. \qquad (3.11)\end{aligned}$$

Adding (3.10) and (3.11), we get (3.7) and the proof is complete. □

The assertion in Theorem 3.1 (see inequality (3.2)) coincides with the one in Theorem 3.2 (see inequality (3.6)). The only difference is the sequence $\{u^k\}$ in $\Re^n$ or Ω.

4 Convergence rate of the PC methods for LVI

For the different projection and contraction methods, we have the same key inequality which is shown in Theorem 3.1 (see (3.2)) and Theorem 3.2 (see (3.6)), respectively. The contraction property (1.7) of the PC methods is the consequent result of Theorem 3.1 and Theorem 3.2, respectively.

Theorem 4.1 *The sequence $\{u^k\}$ generated by PC Method-I or PC Method II satisfies*

$$\|u^{k+1}-u^*\|^2 \le \|u^k-u^*\|^2-\gamma(2-\gamma)\alpha_k^*\|u^k-\tilde{u}^k\|^2, \quad \forall\, u^*\in\Omega^*. \qquad (4.1)$$

Proof. By setting $u=u^*$ in (3.2) and (3.6), we get

$$\|u^k-u^*\|^2-\|u^{k+1}-u^*\|^2 \ge 2\gamma\alpha_k^*(\tilde{u}^k-u^*)^T(Mu^*+q)+q_k(\gamma).$$

Because $(\tilde{u}^k-u^*)^T(Mu^*+q)\ge 0$, it follows from the last inequality that

$$\|u^{k+1}-u^*\|^2 \le \|u^k-u^*\|^2-q_k(\gamma).$$

Using the notation of $q_k(\gamma)$ (see (1.7b)), the theorem is proved. □.

The following is the assertion of the convergence rate in ergodic sense.

Theorem 4.2 *For any integer $t > 0$, we have a $\tilde{u}_t \in \Omega$ which satisfies*

$$(\tilde{u}_t - u)^T F(u) \le \frac{\|I + M^T\|^2}{2\gamma(t+1)} \|u - u^0\|^2, \quad \forall u \in \Omega, \tag{4.2}$$

where

$$\tilde{u}_t = \frac{1}{\Upsilon_t} \sum_{k=0}^{t} \alpha_k^* \tilde{u}^k \qquad \textit{and} \qquad \Upsilon_t = \sum_{k=0}^{t} \alpha_k^*. \tag{4.3}$$

Proof. For the convergence rate proof, we allow $\gamma \in (0, 2]$. In this case, we still have $q_k(\gamma) \ge 0$. By using the monotonicity of F, from (3.2) and (3.6) we get

$$(u - \tilde{u}^k)^T \alpha_k^* (Mu + q) + \frac{1}{2\gamma} \|u - u^k\|^2 \ge \frac{1}{2\gamma} \|u - u^{k+1}\|^2, \quad \forall\, u \in \Omega.$$

Summing the above inequality over $k = 0, \ldots, t$, we obtain

$$\Big((\sum_{k=0}^{t} \alpha_k^*) u - \sum_{k=0}^{t} \alpha_k^* \tilde{u}^k \Big)^T (Mu + q) + \frac{1}{2\gamma} \|u - u^0\|^2 \ge 0, \quad \forall u \in \Omega.$$

Using the notations of Υ_t and $\tilde{u}_t$ in the above inequality, we derive

$$(\tilde{u}_t - u)^T (Mu + q) \le \frac{\|u - u^0\|^2}{2\gamma \Upsilon_t}, \quad \forall u \in \Omega. \tag{4.4}$$

Indeed, $\tilde{u}_t \in \Omega$ because it is a convex combination of $\tilde{u}^0, \tilde{u}^1, \ldots, \tilde{u}^t$. Because $\alpha_k^* \ge 1/\|I + M\|^2$ for all $k > 0$ (see (1.4)), it follows from (4.3) that

$$\Upsilon_t \ge \frac{t+1}{\|I + M\|^2},$$

The proof is complete. $\square$

and thus the PC methods have $O(1/t)$ convergence rate. For any substantial set $\mathcal{D} \subset \Omega$, the PC methods reach

$$(\tilde{u}_t - u)^T F(u) \le \epsilon, \quad \forall\, u \in \mathcal{D}(\tilde{u}_t), \quad \text{in at most} \quad t = \Big\lceil \frac{\|I + M\|^2 d^2}{2\gamma\epsilon} \Big\rceil$$

iterations, where $\tilde{u}_t$ is defined in (4.3) and $d = \sup\{\|u - u^0\| \mid u \in \mathcal{D}(\tilde{u})\}$. This convergence rate is in the ergodic sense, the statement (4.2) suggests us to take a larger parameter $\gamma \in (0, 2]$ in the correction steps of the PC methods.

5 PC Method for convex quadratic optimization

This section instigate the convergence rate of the PC method for the convex quadratic programming. Let Ω be a closed convex subset of $\Re^n$, $H \in \Re^{n\times n}$ is positive semidefinite and $c \in \Re^n$. Solving the convex quadratic optimization

$$\min\{\tfrac{1}{2}x^T Hx + c^T x \,|\, x \in \Omega\}$$

is equivalent to finding a vector $x^* \in \Omega$ such that

$$\text{SLVI}(\Omega, H, c) \qquad (x - x^*)^T(Hx^* + c) \geq 0, \qquad \forall x \in \Omega. \tag{5.1}$$

The problem (5.1) is called symmetric linear variational inequality problem and denoted by LVI(Ω, H, c). For given $x^k \in \Re^n$, let

$$\tilde{x}^k = P_\Omega[x^k - \beta(Hx^k + c)]. \tag{5.2}$$

For solving the monotone variational inequalities [7] are developed. For monotone linear variational inequality, SLVI(Ω, H, c), one of the method in [7] (see Method 1 therein) generates the new iterate by

$$\text{(PC Method)} \qquad x^{k+1} = x^k - \gamma\alpha_k^*(x^k - \tilde{x}^k), \tag{5.3}$$

where

$$\alpha_k^* = \frac{\|x^k - \tilde{x}^k\|^2}{\|x^k - \tilde{x}^k\|_G^2}, \qquad G = I + \beta H \qquad \text{and} \qquad \gamma \in (0, 2). \tag{5.4}$$

It was proved [7] that the sequence $\{x^k\}$ generated by PC Method satisfies

$$\|x^{k+1} - x^*\|_G^2 \leq \|x^k - x^*\|_G^2 - \gamma(2-\gamma)\alpha_k^*\|x^k - \tilde{x}^k\|^2. \tag{5.5}$$

Because a projection is needed in each iteration and the distance of the iterates to the solution set is monotonically decreasing, according to the terminology in [1], such methods are called *Projection and Contraction Methods* (see [4]).

In this section, we show that, for given $\epsilon > 0$, in $O(1/\epsilon)$ iterations the projection and contraction method can find a $\tilde{x}$ such that

$$\tilde{x} \in \Omega \quad \text{and} \quad \sup_{x\in\mathcal{D}(\tilde{x})} \left\{(\tilde{x} - x)^T(Hx + c)\right\} \leq \epsilon, \tag{5.6}$$

where

$$\mathcal{D}(\tilde{x}) = \{x \in \Omega \,|\, \|x - \tilde{x}\| \leq 1\}.$$

In this sense, we will establish the algorithmic convergence complexity for the pair of geminate projection and contraction method .

5.1 Main theorem for complexity analysis

This section shows the main theorems for the complexity analysis. Now, we prove the key inequality for the complexity analysis of the pair of geminate Algorithm.

Theorem 5.1 *For given $x^k \in \Re^n$, let $\tilde{x}^k$ be defined by* (5.2) *and the new iterate x^{k+1} be generated by PC Method* (5.3) *with any $\gamma \in (0, 2)$. Then we have*

$$\begin{aligned} &\gamma\alpha_k^*\beta(x-\tilde{x}^k)^T(Hx+c) \\ &\quad\geq \quad \frac{1}{2}\big(\|x-x^{k+1}\|_G^2 - \|x-x^k\|_G^2\big) + \frac{1}{2}q_k(\gamma), \ \ \forall\, x\in\Omega, \end{aligned} \tag{5.7}$$

where

$$q_k(\gamma) = \gamma(2-\gamma)\alpha_k^*\|x^k-\tilde{x}^k\|^2. \tag{5.8}$$

Proof. Set $v = x^k - \beta(Hx^k + c)$ in (1.3), because $\tilde{x}^k = P_\Omega[x^k - \beta(Hx^k + c)]$ and thus $\tilde{x}^k = P_\Omega(v)$, it follows that

$$(x-\tilde{x}^k)^T\{x^k - \beta(Hx^k+c) - \tilde{x}^k\} \leq 0 \quad \forall\, x\in\Omega.$$

Thus, we have

$$(x-\tilde{x}^k)^T\beta(Hx^k+c) \geq (x-\tilde{x}^k)^T(x^k-\tilde{x}^k), \quad \forall\, x\in\Omega.$$

Adding the term $(x-\tilde{x}^k)^T\beta H(x^k-\tilde{x}^k)$ to the both sides of the above inequality and using $(I+\beta H) = G$, we obtain

$$\begin{aligned} &(x-\tilde{x}^k)^T\{\beta(Hx^k+c) + \beta H(x^k-\tilde{x}^k)\} \\ &\quad\geq \quad (x-\tilde{x}^k)^T G(x^k-\tilde{x}^k), \quad \forall\, x\in\Omega. \end{aligned}$$

Then, we rewrite the above inequality in our desirable form

$$\begin{aligned} &(x-\tilde{x}^k)^T\{\beta(Hx+c) + \beta H(\tilde{x}^k-x) + 2\beta H(x^k-\tilde{x}^k)\} \\ &\quad\geq \quad (x-\tilde{x}^k)^T G(x^k-\tilde{x}^k), \quad \forall\, x\in\Omega. \end{aligned}$$

By using the Cauchy-Schwarz inequality, from the above inequality we obtain

$$\begin{aligned} &(x-\tilde{x}^k)^T\beta(Hx+c) \\ &\quad\geq \quad (x-\tilde{x}^k)^T G(x^k-\tilde{x}^k) + \beta\|x-\tilde{x}^k\|_H - 2(x-\tilde{x}^k)^T\beta H(x^k-\tilde{x}^k) \\ &\quad\geq \quad (x-\tilde{x}^k)^T G(x^k-\tilde{x}^k) - \beta\|x^k-\tilde{x}^k\|_H^2, \quad \forall\, x\in\Omega, \end{aligned}$$

and thus

$$\begin{aligned}
&\gamma\alpha_k^*(x-\tilde{x}^k)^T\beta(Hx+c)\\
&\quad\geq\quad (x-\tilde{x}^k)^T G\gamma\alpha_k^*(x^k-\tilde{x}^k)-\gamma\alpha_k^*\beta\|x^k-\tilde{x}^k\|_H^2, \quad \forall\, x\in\Omega.
\end{aligned}$$

Because $\gamma\alpha_k^*(x^k-\tilde{x}^k)=(x^k-x^{k+1})$ (see (1.5)), we get

$$\begin{aligned}
&\gamma\alpha_k^*(x-\tilde{x}^k)^T\beta(Hx+c)\\
&\quad\geq\quad (x-\tilde{x}^k)^T G(x^k-x^{k+1})-\gamma\alpha_k^*\beta\|x^k-\tilde{x}^k\|_H^2, \quad \forall\, x\in\Omega. \qquad (5.9)
\end{aligned}$$

To the crossed term in the right hand side of (5.9), namely $(x-\tilde{x}^k)^T G(x^k-x^{k+1})$, using an identity

$$(a-b)^T G(c-d)=\frac{1}{2}\big(\|a-d\|_G^2-\|a-c\|_G^2\big)+\frac{1}{2}\big(\|c-b\|_G^2-\|d-b\|_G^2\big),$$

we obtain

$$\begin{aligned}
(x-\tilde{x}^k)^T G(x^k-x^{k+1}) \quad &= \quad \frac{1}{2}\big(\|x-x^{k+1}\|_G^2-\|x-x^k\|_G^2\big)+\\
&\qquad \frac{1}{2}\big(\|x^k-\tilde{x}^k\|_G^2-\|x^{k+1}-\tilde{x}^k\|_G^2\big). \qquad (5.10)
\end{aligned}$$

By using $x^{k+1}=x^k-\gamma\alpha_k^*(x^k-\tilde{x}^k)$ and (5.4), we get

$$\begin{aligned}
&\|x^k-\tilde{x}^k\|_G^2-\|x^{k+1}-\tilde{x}^k\|_G^2\\
&\quad= \quad \|x^k-\tilde{x}^k\|_G^2-\|(x^k-\tilde{x}^k)-\gamma\alpha_k^*(x^k-\tilde{x}^k)\|_G^2\\
&\quad= \quad 2\gamma\alpha_k^*(x^k-\tilde{x}^k)^T G(x^k-\tilde{x}^k)-\gamma^2\alpha_k^*(\alpha_k^*\|x^k-\tilde{x}^k\|_G^2)\\
&\quad= \quad 2\gamma\alpha_k^*\|x^k-\tilde{x}^k\|^2+2\gamma\alpha_k^*\beta\|x^k-\tilde{x}^k\|_H^2-\gamma^2\alpha_k^*\|x^k-\tilde{x}^k\|^2\\
&\quad= \quad \gamma(2-\gamma)\alpha_k^*\|x^k-\tilde{x}^k\|^2+2\gamma\alpha_k^*\|x^k-\tilde{x}^k\|_H^2.
\end{aligned}$$

Substituting it in the right hand side of (5.10) and using the definition of $q_k(\gamma)$, we obtain

$$\begin{aligned}
(x-\tilde{x}^k)^T G(x^k-x^{k+1}) \quad &= \quad \frac{1}{2}\big(\|x-x^{k+1}\|_G^2-\|x-x^k\|_G^2\big)\\
&\qquad +\frac{1}{2}q_k(\gamma)+\gamma\alpha_k^*\beta\|x^k-\tilde{x}^k\|_H^2. \qquad (5.11)
\end{aligned}$$

Adding (5.9) and (5.11), the theorem is proved. $\square$

By setting $x=x^*$ in (5.7), we get

$$\|x^k-x^*\|_G^2-\|x^{k+1}-x^*\|_G^2\geq 2\gamma\alpha_k^*(\tilde{x}^k-x^*)^T(Hx^*+c)+q_k(\gamma).$$

Because $(\tilde{x}^k - x^*)^T(Hx^* + c) \geq 0$, it follows from the last inequality and (5.8) that

$$\|x^{k+1} - x^*\|_G^2 \leq \|x^k - x^*\|_G^2 - \gamma(2-\gamma)\alpha_k^*\|x^k - \tilde{x}^k\|^2.$$

Thus, the contraction property (5.5) of the PC-Method is a byproduct of Theorem 5.1.

5.2 Convergence rate of the PC method

This section uses Theorem 3.1 to show the convergence rate of the projection type algorithm.

Theorem 5.2 *Let the sequence $\{x^k\}$ be generated by PC Method, and $\tilde{x}^k$ be given by* (1.4)*. For any integer $t > 0$, it holds that*

$$(\tilde{x}_t - x)^T(Hx + c) \leq \frac{\|I + \beta H\|}{2\gamma\beta(t+1)}\|x - x^0\|_G^2, \quad \forall x \in \Omega, \tag{5.12}$$

where

$$\tilde{x}_t = \frac{1}{\Upsilon_t}\sum_{k=0}^{t}\alpha_k^*\tilde{x}^k \qquad \textit{and} \qquad \Upsilon_t = \sum_{k=0}^{t}\alpha_k^*. \tag{5.13}$$

Proof. For the convergence rate proof, we allow $\gamma \in (0, 2]$. In this case, we still have $q_k(\gamma) \geq 0$. From (3.2) we get

$$(x - \tilde{x}^k)^T\alpha_k^*(Hx + c) + \frac{1}{2\gamma\beta}\|x - x^k\|_G^2 \geq \frac{1}{2\gamma\beta}\|x - x^{k+1}\|_G^2, \quad \forall\, x \in \Omega.$$

Summing the above inequality over $k = 0, \ldots, t$, we obtain

$$\Big((\sum_{k=0}^{t}\alpha_k^*)x - \sum_{k=0}^{t}\alpha_k^*\tilde{x}^k\Big)^T(Hx + c) + \frac{1}{2\gamma\beta}\|x - x^0\|^2 \geq 0, \quad \forall x \in \Omega.$$

Using the notations of Υ_t and $\tilde{x}_t$ in the above inequality, we derive

$$(\tilde{x}_t - x)^T(Hx + c) \leq \frac{\|x - x^0\|^2}{2\gamma\beta\Upsilon_t}, \quad \forall x \in \Omega. \tag{5.14}$$

Indeed, $\tilde{x}_t \in \Omega$ because it is a convex combination of $\tilde{x}^0, \tilde{x}^1, \ldots, \tilde{x}^t$. Because $\alpha_k^* \geq 1/\|I + \beta H\|$ for all $k > 0$ (see (5.2)), it follows from (5.13) that

$$\Upsilon_t \geq \frac{t+1}{\|I + \beta H\|}.$$

Substituting it in (5.14), the proof is complete. $\square$

Thus the projection and contraction method have $O(1/t)$ convergence rate. For any substantial set $\mathcal{D} \subset \Omega$, the PC Method and PC Method reach

$$(\tilde{x}_t - x)^T(Hx + c) \le \epsilon, \quad \forall\, x \in \mathcal{D}(\tilde{x}_t), \quad \text{in at most} \quad t = \left\lceil \frac{\|I + \beta H\| d^2}{2\gamma\beta\epsilon} \right\rceil$$

iterations, where $\tilde{x}_t$ is defined in (5.13) and $d = \sup\{\|x - x^0\| \mid x \in \mathcal{D}(\tilde{x}_t)\}$. This convergence rate is in the ergodic sense, the statement (5.12) suggests us to take a larger parameter $\gamma \in (0, 2]$ in the correction steps of the projection and contraction method .

5.3 Convergence rate in the point manner

In this subsection, we consider a special version of PC Method which updates the new iterate by

$$x^{k+1} = x^k - (x^k - \tilde{x}^k), \tag{5.15}$$

where the positive definite matrix G satisfies the following condition:

$$G = I + \beta H \preceq 2(1 - \nu)I, \quad \nu \in (0, 1). \tag{5.16}$$

Under the condition (5.16), if we use the original PC Method (1.5), the α_k^* defined in (1.6)

satisfies

$$\alpha_k^* = \frac{\|x^k - \tilde{x}^k\|^2}{\|x^k - \tilde{x}^k\|_G^2} \ge \frac{1}{2(1 - \nu)}. \tag{5.17}$$

The update form (5.15) can be viewed in PC Method (5.3) by dynamically taking $\gamma_k \equiv 1/\alpha_k^*$. According to (5.5) and (5.17), using

$$\gamma_k \alpha_k^* = 1 \quad \text{and} \quad \|x^k - \tilde{x}^k\|^2 = \alpha_k^* \|G^{-1}(I + H)(x^k - \tilde{x}^k)\|_G^2,$$

we obtain

$$\begin{aligned} q_k(\gamma_k) &= (2 - \gamma_k)\|x^k - \tilde{x}^k\|^2 \\ &= (2 - \gamma_k)\alpha_k^* \|x^k - \tilde{x}^k\|_G^2 \ \ge\ \frac{\nu}{1 - \nu}\|x^k - \tilde{x}^k\|_G^2, \quad \forall\, x^* \in \Omega^*. \end{aligned}$$

Then, from (1.7), the above inequality and (5.15) follows that

$$\|x^{k+1} - x^*\|_G^2 \le \|x^k - x^*\|_G^2 - \frac{\nu}{1 - \nu}\|x^k - x^{k+1}\|_G^2, \quad \forall\, x^* \in \Omega^*. \tag{5.18}$$

In the following we focus on showing that the sequence $\{\|x^k - x^{k+1}\|_G\}$ is monotonically non-increasing. For this purpose, we need to prove the following lemma.

Lemma 5.1 *For given x^k, let $\tilde{x}^k$ be offered by* (1.4)*. If the algorithm uses the update form*

(5.15) *to generate the new iterate, then we have*

$$\begin{aligned}&(x^k - x^{k+1})^T G\{(x^k - \tilde{x}^k) - (x^{k+1} - \tilde{x}^{k+1})\}\\ &\quad\geq \quad \|(x^k - \tilde{x}^k) - (x^{k+1} - \tilde{x}^{k+1})\|^2 + \beta\|x^k - x^{k+1}\|_H^2. \qquad (5.19)\end{aligned}$$

Proof. First, set $y = x^k - \beta(Hx^k + c)$ in (1.3), since $\tilde{x}^k = P_\Omega[x^k - \beta(Hx^k + c)]$, it follows that

$$(x - \tilde{x}^k)^T\{\beta(Hx^k + c) - (x^k - \tilde{x}^k)\} \geq 0, \quad \forall\, x \in \Omega.$$

We rewrite it as

$$(x - \tilde{x}^k)^T\{\beta(H\tilde{x}^k + c) + 2\beta H(x^k - \tilde{x}^k) - (I + \beta H)(x^k - \tilde{x}^k)\} \geq 0, \quad \forall\, x \in \Omega,$$

and thus

$$\begin{aligned}&(x - \tilde{x}^k)^T\beta(H\tilde{x}^k + c)\\ &\quad\geq (x - \tilde{x}^k)^T G(x^k - \tilde{x}^k) - 2\beta(x - \tilde{x}^k)^T H(x^k - \tilde{x}^k),\ \forall x \in \Omega. \quad (5.20)\end{aligned}$$

Set $x = \tilde{x}^{k+1}$ in (5.20), we have

$$\begin{aligned}(\tilde{x}^{k+1} - \tilde{x}^k)^T\beta(H\tilde{x}^k + c) \quad \geq \quad & (\tilde{x}^{k+1} - \tilde{x}^k)^T G(x^k - \tilde{x}^k)\\ & -2\beta(\tilde{x}^{k+1} - \tilde{x}^k)^T H(x^k - \tilde{x}^k). \quad (5.21)\end{aligned}$$

Note that (5.20) is also true for $k := k + 1$ and thus we have

$$\begin{aligned}(x - \tilde{x}^{k+1})^T\beta(H\tilde{x}^{k+1} + c) \quad \geq \quad & (x - \tilde{x}^{k+1})^T G(x^{k+1} - \tilde{x}^{k+1})\\ & -2\beta(x - \tilde{x}^{k+1})^T H(x^{k+1} - \tilde{x}^{k+1}),\ \forall\, x \in \Omega.\end{aligned}$$

Set $x = \tilde{x}^k$ in the above inequality, we obtain

$$\begin{aligned}(\tilde{x}^k - \tilde{x}^{k+1})^T\beta(H\tilde{x}^{k+1} + c) \quad \geq \quad & (\tilde{x}^k - \tilde{x}^{k+1})^T G(x^{k+1} - \tilde{x}^{k+1})\\ & -2\beta(\tilde{x}^k - \tilde{x}^{k+1})^T H(x^{k+1} - \tilde{x}^{k+1}). \quad (5.22)\end{aligned}$$

Adding (5.21) and (5.22), we get

$$\begin{aligned}&(\tilde{x}^k - \tilde{x}^{k+1})^T G\{(x^k - \tilde{x}^k) - (x^{k+1} - \tilde{x}^{k+1})\}\\ &\quad\geq \quad \beta\|\tilde{x}^k - \tilde{x}^{k+1}\|_H^2\\ &\qquad +2\beta(\tilde{x}^k - \tilde{x}^{k+1})^T H\{(x^k - \tilde{x}^k) - (x^{k+1} - \tilde{x}^{k+1})\}. \qquad (5.23)\end{aligned}$$

Moreover, by adding the identity

$$\begin{aligned}&\{(x^k-\tilde{x}^k)-(x^{k+1}-\tilde{x}^{k+1})\}^T G\{(x^k-\tilde{x}^k)-(x^{k+1}-\tilde{x}^{k+1})\}\\ &\quad= \quad \beta\|(x^k-\tilde{x}^k)-(x^{k+1}-\tilde{x}^{k+1})\|_H^2+\|(x^k-\tilde{x}^k)-(x^{k+1}-\tilde{x}^{k+1})\|^2\end{aligned}$$

to the both sides of (5.23), we obtain

$$\begin{aligned}&(x^k-x^{k+1})^T G\{(x^k-\tilde{x}^k)-(x^{k+1}-\tilde{x}^{k+1})\}\\ &\quad\geq \quad \beta\|\tilde{x}^k-\tilde{x}^{k+1}\|_H^2+2\beta(\tilde{x}^k-\tilde{x}^{k+1})^T H\{(x^k-\tilde{x}^k)-(x^{k+1}-\tilde{x}^{k+1})\}\\ &\qquad +\beta\|(x^k-\tilde{x}^k)-(x^{k+1}-\tilde{x}^{k+1})\|_H^2+\|(x^k-\tilde{x}^k)-(x^{k+1}-\tilde{x}^{k+1})\|^2\\ &\quad= \quad \beta\|x^k-x^{k+1}\|_H^2+\|(x^k-\tilde{x}^k)-(x^{k+1}-\tilde{x}^{k+1})\|^2. \qquad (5.24)\end{aligned}$$

This is just the inequality (5.19) and the lemma is proved. □

Theorem 5.3 *For given x^k, let $\tilde{x}^k$ be offered by* (1.4)*. If the algorithm uses the update form* (5.15) *to generate the new iterate and the the condition* (5.16) *is satisfied, then for all integer $k \geq 0$, it holds that*

$$\|x^{k+1}-x^{k+2}\|_G^2 \leq \|x^k-x^{k+1}\|_G^2. \qquad (5.25)$$

Proof. Using $\tilde{x}^k = x^{k+1}$ and $\tilde{x}^{k+1} = x^{k+2}$ (see (5.15)) in (5.19), we obtain

$$\begin{aligned}&(x^k-x^{k+1})^T G\{(x^k-x^{k+1})-(x^{k+1}-x^{k+2})\}\\ &\quad\geq \quad \|(x^k-x^{k+1})-(x^{k+1}-x^{k+2})\|^2. \qquad (5.26)\end{aligned}$$

Setting $a = (x^k - x^{k+1})$ and $b = (x^{k+1} - x^{k+2})$ in the identity

$$\|a\|_G^2-\|b\|_G^2 = 2a^T G(a-b)-\|a-b\|_G^2,$$

we obtain

$$\begin{aligned}&\|x^k-x^{k+1}\|_G^2-\|x^{k+1}-x^{k+2}\|_G^2\\ &\quad= \quad 2(x^k-x^{k+1})^T G\{(x^k-x^{k+1})-(x^{k+1}-x^{k+2})\}\\ &\qquad -\|(x^k-x^{k+1})-(x^{k+1}-x^{k+2})\|_G^2.\end{aligned}$$

Inserting (5.26) into the first term of the right-hand side of the last equality, we obtain

$$\begin{aligned}&\|x^k-x^{k+1}\|_G^2-\|x^{k+1}-x^{k+2}\|_G^2\\ &\quad\geq \quad 2\|(x^k-x^{k+1})-(x^{k+1}-x^{k+2})\|^2\\ &\qquad -\|(x^k-x^{k+1})-(x^{k+1}-x^{k+2})\|_G^2. \qquad (5.27)\end{aligned}$$

Thus,

$$\begin{aligned} &\|x^k - x^{k+1}\|_G^2 - \|x^{k+1} - x^{k+2}\|_G^2 \\ &\quad \geq \quad \|(x^k - x^{k+1}) - (x^{k+1} - x^{k+2})\|_{(2I-G)}^2. \end{aligned} \tag{5.28}$$

According to (5.16), the matrix $(2I - G)$ is positive definite, and thus the right hand side of (5.28) is non-negative. The assertion (5.25) follows immediately. $\square$

With (5.18) and (5.25), we can prove the main result for the convergence rate in the non-ergodic sense.

Theorem 5.4 *For given x^k, let $\tilde{x}^k$ be offered by* (1.4)*. If the algorithm uses the update form* (5.15) *to generate the new iterate and the the condition* (5.16) *is satisfied, then for any integer $t > 0$, we have*

$$\|x^t - x^{t+1}\|_G^2 \leq \frac{1-\nu}{(t+1)\nu}\|x^0 - x^*\|_G^2. \tag{5.29}$$

Proof. First, it follows from (5.18) that

$$\frac{\nu}{1-\nu}\sum_{k=0}^{\infty}\|x^k - x^{k+1}\|_G^2 \leq \|x^0 - x^*\|_G^2, \quad \forall\, x^* \in \Omega^*. \tag{5.30}$$

According to Theorem 5.3, the sequence $\{\|x^k - x^{k+1}\|_G^2\}$ is monotonically non-increasing. Therefore, we have

$$(t+1)\|x^t - x^{t+1}\|_G^2 \leq \sum_{k=0}^{t}\|x^k - x^{k+1}\|_G^2. \tag{5.31}$$

The assertion (5.29) follows from (5.30) and (5.31) immediately. $\square$

Notice that Ω^* is convex and closed (see (2.3.2) of [3]). Let

$$d := \inf\{\|x^0 - x^*\|_G \,|\, x^* \in \Omega^*\}.$$

Then, for any given $\epsilon > 0$, Theorem 5.4 shows that the it needs at most $\lfloor d^2/c_0\epsilon \rfloor$ iterations to ensure that $\|x^k - x^{k+1}\|_G^2 \leq \epsilon$.

Recall that x^k is a solution of SLVI(Ω, H, c) if $\|x^k - x^{k+1}\|_G^2 = 0$ (see (5.15)). A worst-case $O(1/t)$ convergence rate in a non-ergodic sense is thus established.

References

[1] E. Blum and W. Oettli, *Mathematische Optimierung: Grundlagen und Verfahren. Ökonometrie und Unternehmensforschung*. Berlin-Heidelberg-New York, Springer-Verlag, 1975.

[2] C. H. Chen, X. L. Fu, B.S. He and X. M. Yuan, On the Iteration Complexity of Some Projection Methods for Monotone Linear Variational Inequalities, JOTA, 172(2017) 914-928.

[3] F. Facchinei and J.S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems, Vol. I and II*. Springer Series in Operations Research. Springer Verlag, New York, 2003.

[4] 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**, pp. 247–262, 1992.

[5] B.S. He, A globally linearly convergent projection and contraction method for a class of linear complementarity problems. Schwerpunktprogramm der Deutschen Forschungsgemeinschaft Anwen- dungsbezogene Optimierung und Steuerung, Report No. 352, 1992.

[6] B.S. He, A new method for a class of linear variational inequalities, *Mathematical Programming* **66**, pp. 137–144, 1994.

[7] B.S. He, Solving a class of linear projection equations, *Numerische Mathematik* **68**, pp. 71–80, 1994.

[8] B.S. He, A modified projection and contraction method for a class of linear complementarity problems, *Journal of Computational Mathematics* **14**, pp. 54–63, 1996.

[9] B.S. He, A class of projection and contraction methods for monotone variational inequalities, *Applied Mathematics and optimization* **35**, pp. 69–76, 1997.

[10] B.S. He and M.-H. Xu, A general framework of contraction methods for monotone variational inequalities, *Pacific J. Optimization* **4**, pp. 195-212, 2008.

[11] G.M. Korpelevich, The extragradient method for finding saddle points and other problems, *Ekonomika i Matematicheskie Metody* **12**, pp. 747-756, 1976.

[12] E.N. Khobotov, Modification of the extragradient method for solving variational inequalities and certain optimization problems, *USSR Computational Mathematics and Mathematical Physics*, Vol. 27, pp.120-127, 1987.

[13] A. Nemirovski, Prox-method with rate of convergence $O(1/t)$ for variational inequality with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems, *SIAM J. Optimization* **15**, pp. 229-251, 2005.

[14] P. Tseng, On accelerated proximal gradient methods for convex-concave optimization, Department of Mathematics, University of Washington, Seattle, WA 98195, USA, 2008.

凸优化和单调变分不等式的收缩算法

第五讲: Lipschitz 连续的单调变分不等式投影收缩算法的收敛速率

Convergence rate of the PC methods
for Lipschitz continuous monotone VIs

南京大学数学系　何炳生
hebma@nju.edu.cn

The context of this lecture is based on the publication [2]

In 2005, Nemirovski's analysis indicates that the extragradient method has the $O(1/t)$ convergence rate for variational inequalities with Lipschitz continuous monotone operators. For the same problems, in the last decades, we have developed a class of Fejér monotone projection and contraction methods. Until now, only convergence results are available to these projection and contraction methods, though the numerical experiments indicate that they always outperform the extragradient method. The reason is that the former benefits from the 'optimal' step size in the contraction sense. In this paper, we prove the convergence rate under a unified conceptual framework, which includes the projection and contraction methods as special cases and thus perfects the theory of the existing projection and contraction methods. Preliminary numerical results demonstrate that the projection and contraction methods converge twice faster than the extragradient method.

1 Introduction

Let Ω be a nonempty closed convex subset of $\Re^n$, F be a continuous mapping from $\Re^n$ to itself. The variational inequality problem, denoted by VI(Ω, F), is to find a vector

$u^* \in \Omega$ such that

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

Notice that VI(Ω, F) is invariant when F is multiplied by some positive scalar $\beta > 0$. It is well known that, for any $\beta > 0$,

$$u^* \text{ is a solution of VI}(\Omega, F) \quad \Longleftrightarrow \quad u^* = P_\Omega[u^* - \beta F(u^*)], \tag{1.2}$$

where $P_\Omega(\cdot)$ denotes the projection onto Ω with respect to the Euclidean norm, *i.e.*,

$$P_\Omega(v) = \text{argmin}\{\|u - v\| \mid u \in \Omega\}.$$

Throughout this paper we assume that the mapping F is monotone and Lipschitz continuous, *i.e.*,

$$(u - v)^T (F(u) - F(v)) \geq 0, \quad \forall\, u, v \in \Re^n,$$

and there is a constant $L > 0$ (not necessary known), such that

$$\|F(u) - F(v)\| \leq L\|u - v\|, \quad \forall\, u, v \in \Re^n.$$

Moreover, we assume that the solution set of VI(Ω, F), denoted by Ω^*, is nonempty. The nonempty assumption of the solution set, together with the monotonicity assumption of F, implies that Ω^* is closed and convex (see pp. 158 in [3]).

Among the algorithms for monotone variational inequalities, the extragradient (EG) method proposed by Korpelevich [9] is one of the attractive methods. In fact, each iteration of the extragradient method can be divided into two steps. The k-th iteration of EG method begins with a given $u^k \in \Omega$, the first step produces a vector $\tilde{u}^k$ via a projection

$$\tilde{u}^k = P_\Omega[u^k - \beta_k F(u^k)], \tag{1.3a}$$

where $\beta_k > 0$ is selected to satisfy

$$\beta_k \|F(u^k) - F(\tilde{u}^k)\| \leq \nu \|u^k - \tilde{u}^k\|, \quad \nu \in (0, 1). \tag{1.3b}$$

Since $\tilde{u}^k$ is not accepted as the new iterate, for designation convenience, we call it as a *predictor* and β_k is named the *prediction step size*. The second step (correction step) of the k-th iteration updates the new iterate u^{k+1} by

$$u^{k+1} = P_\Omega[u^k - \beta_k F(\tilde{u}^k)], \tag{1.4}$$

where β_k is called the *correction step size*. The sequence $\{u^k\}$ generated by the

extragradient method is Fejér monotone with respect to the solution set, namely,

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - (1-\nu^2)\|u^k - \tilde{u}^k\|^2. \tag{1.5}$$

For a proof of the above contraction property, the readers may consult [3] (see pp. 1115-1118 therein). Notice that, in the extragradient method, the step size of the prediction (1.3a) and that of the correction (1.4) are equal. Thus the two steps seem like 'symmetric'.

Because of its simple iterative forms, recently, the extragradient method has been applied to solve some large optimization problems in the area of information science, such as in machine learning [15], optical network [11] and speech recognition [12], etc. In addition, Nemirovski [10] and Tseng [16] proved the $O(1/t)$ convergence rate of the extragradient method. Both in the theoretical and practical aspects, the interest in the extragradient method becomes more active.

In the last decades, we devoted our effort to develop a class of projection and contraction (PC) methods for monotone variational inequalities [5, 6, 8, 13]. Similarly as in the extragradient method, each iteration of the PC methods consists of two steps. The prediction step of PC methods produces the predictor $\tilde{u}^k$ via (1.3) just as in the extragradient method. The PC methods exploit a pair of geminate directions [7, 8] offered

by the predictor, namely, they are

$$d(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k) - \beta_k(F(u^k) - F(\tilde{u}^k)) \quad \text{and} \quad \beta_k F(\tilde{u}^k). \tag{1.6}$$

Here, both the directions are ascent directions of the unknown distance function $\frac{1}{2}\|u - u^*\|^2$ at the point u^k. Based on such directions, the goal of the correction step is to generate a new iterate which is more closed to the solution set. It leads to choosing the 'optimal' step length

$$\varrho_k = \frac{(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2}, \tag{1.7}$$

and a relaxation factor $\gamma \in (0, 2)$, the second step (*correction step*) of the PC methods updates the new iterate u^{k+1} by

$$u^{k+1} = u^k - \gamma \varrho_k d(u^k, \tilde{u}^k), \tag{1.8}$$

or

$$u^{k+1} = P_\Omega[u^k - \gamma \varrho_k \beta_k F(\tilde{u}^k)]. \tag{1.9}$$

The PC methods (without line search) make one (or two) projection(s) on Ω at each iteration, and the distance of the iterates to the solution set monotonically converges to zero. According to the terminology in [1], these methods belong to the class of Fejér

contraction methods. In fact, the only difference between the extragradient method and one of the PC methods is that they use different step sizes in the correction step (see (1.4) and (1.9)). According to our numerical experiments [6, 8], the PC methods always outperform the extragradient methods.

Stimulated by the complexity statement of the extragradient method, this paper shows the $O(1/t)$ convergence rate of the projection and contraction methods for monotone VIs. Recall that Ω^* can be characterized as (see (2.3.2) in pp. 159 of [3])

$$\Omega^* = \bigcap_{u\in\Omega} \{\tilde{u} \in \Omega : (u-\tilde{u})^T F(u) \geq 0\}.$$

This implies that $\tilde{u} \in \Omega$ is an approximate solution of VI(Ω, F) with the accuracy ϵ if it satisfies

$$\tilde{u} \in \Omega \quad \text{and} \quad \inf_{u\in\Omega} \{(u-\tilde{u})^T F(u)\} \geq -\epsilon.$$

In this paper, we show that, for given $\epsilon > 0$ and $\mathcal{D} \subset \Omega$, in $O(L/\epsilon)$ iterations the projection and contraction methods can find a $\tilde{u}$ such that

$$\tilde{u} \in \Omega \quad \text{and} \quad \sup_{u\in\mathcal{D}} \{(\tilde{u}-u)^T F(u)\} \leq \epsilon. \tag{1.10}$$

As a byproduct of the complexity analysis, we find why taking a suitable relaxation factor

$\gamma \in (1, 2)$ in the correction steps (1.8) and (1.9) of the PC methods can achieve the faster convergence.

The outline of this paper is as follows. Section 2 recalls some basic concepts in the projection and contraction methods. In Section 3, we investigate the geminate descent directions of the distance function. Section 4 shows the contraction property of the PC methods. In Section 5, we carry out the complexity analysis, which results in an $O(1/t)$ convergence rate and suggests using the large relaxation factor in the correction step of the PC methods. The solution methods study for variational inequality is helpful for investigating the splitting contraction methods for separable convex optimization, some conclusion remarks are addressed in the last section.

Throughout the paper, the following notational conventions are used. We use u^* to denote a fixed but arbitrary point in the solution set Ω^*. A superscript such as in u^k refers to a specific vector and usually denotes an iteration index. For any real matrix M and vector v, we denote the transpose by M^T and v^T, respectively. The Euclidean norm will be denoted by $\|\cdot\|$.

2 Preliminaries

In this section, we summarize the basic concepts of the projection mapping and three fundamental inequalities for constructing the PC methods. Throughout this paper, we assume that the projection on Ω in the Euclidean-norm has a closed form and it is easy to be carried out. Since

$$P_\Omega(v) = \text{argmin}\{\frac{1}{2}\|u-v\|^2 \mid u \in \Omega\},$$

according to the optimal solution of the convex minimization problem, we have

$$(v - P_\Omega(v))^T(u - P_\Omega(v)) \le 0, \quad \forall\, v \in \Re^n, \forall\, u \in \Omega. \tag{2.1}$$

Consequently, for any $u \in \Omega$, it follows from (2.1) that

$$\begin{aligned} \|u-v\|^2 &= \|(u - P_\Omega(v)) - (v - P_\Omega(v))\|^2 \\ &= \|u - P_\Omega(v)\|^2 - 2(v - P_\Omega(v))^T(u - P_\Omega(v)) + \|v - P_\Omega(v)\|^2 \\ &\ge \|u - P_\Omega(v)\|^2 + \|v - P_\Omega(v)\|^2. \end{aligned}$$

Therefore, we have

$$\|u - P_\Omega(v)\|^2 \le \|u-v\|^2 - \|v - P_\Omega(v)\|^2, \quad \forall\, v \in \Re^n, \forall\, u \in \Omega. \tag{2.2}$$

For given u and $\beta > 0$, let $\tilde{u} = P_\Omega[u - \beta F(u)]$ be given via a projection. We say that $\tilde{u}$ is a test-vector of VI(Ω, F) because

$$u = \tilde{u} \quad \Leftrightarrow \quad u \in \Omega^*.$$

Since $\tilde{u} \in \Omega$, it follows from (1.1) that

$$(\textbf{FI-1}) \qquad (\tilde{u} - u^*)^T \beta F(u^*) \ge 0, \quad \forall\, u^* \in \Omega^*. \tag{2.3}$$

Setting $v = u - \beta F(u)$ and $u = u^*$ in the inequality (2.1), we obtain

$$(\textbf{FI-2}) \qquad (\tilde{u} - u^*)^T \big((u - \tilde{u}) - \beta F(u)\big) \ge 0, \quad \forall\, u^* \in \Omega^*. \tag{2.4}$$

Under the assumption that F is monotone we have

$$(\textbf{FI-3}) \qquad (\tilde{u} - u^*)^T \beta\big(F(\tilde{u}) - F(u^*)\big) \ge 0, \quad \forall\, u^* \in \Omega^*. \tag{2.5}$$

The inequalities (2.3), (2.4) and (2.5) play an important role in the projection and contraction methods. They were emphasized in [5] as *three fundamental inequalities* in the

projection and contraction methods.

3 Predictor and the ascent directions

For given u^k, the predictor $\tilde{u}^k$ in the projection and contraction methods [5, 6, 8, 13] is produced by (1.3). Because the mapping F is Lipschitz continuous (even if the constant $L > 0$ is unknown), without loss of generality, we can assume that $\inf_{k\geq 0}\{\beta_k\} \geq \beta_L > 0$ and $\beta_L = O(1/L)$. In practical computation, we can make an initial guesses of $\beta = \nu/L$ and decrease β by a constant factor and repeat the procedure whenever (1.3b) is violated.

For any but fixed $u^* \in \Omega^*$, $(u - u^*)$ is the gradient of the unknown distance function $\frac{1}{2}\|u - u^*\|^2$ in the Euclidean-norm[a] at the point u. A direction d is called an ascent direction of $\frac{1}{2}\|u - u^*\|^2$ at u if and only if the inner-product $(u - u^*)^T d > 0$.

[a] For convenience, we only consider the distance function in the Euclidean-norm. All the results in this paper are easy to extended to the contraction of the distance function in G-norm where G is a positive definite matrix.

3.1 Ascent directions by adding the fundamental inequalities

Setting $u = u^k$, $\tilde{u} = \tilde{u}^k$ and $\beta = \beta_k$ in the fundamental inequalities (2.3), (2.4) and (2.5), and adding them, we get

$$(\tilde{u}^k - u^*)^T d(u^k, \tilde{u}^k) \geq 0, \quad \forall\, u^* \in \Omega^*, \tag{3.1}$$

where

$$d(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k) - \beta_k\big(F(u^k) - F(\tilde{u}^k)\big), \tag{3.2}$$

which is the same $d(u^k, \tilde{u}^k)$ defined in (1.6). It follows from (3.1) that

$$(u^k - u^*)^T d(u^k, \tilde{u}^k) \geq (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \tag{3.3}$$

Note that, under the condition (1.3b), we have

$$\begin{aligned} & 2(u^k - \tilde{u}^k)d(u^k, \tilde{u}^k) - \|d(u^k, \tilde{u}^k)\|^2 \\ = \quad & d(u^k, \tilde{u}^k)^T\{2(u^k - \tilde{u}^k) - d(u^k, \tilde{u}^k)\} \\ = \quad & \|u^k - \tilde{u}^k\|^2 - \beta_k^2\|F(u^k) - F(\tilde{u}^k)\|^2 \\ \geq \quad & (1 - \nu^2)\|u^k - \tilde{u}^k\|^2. \end{aligned} \tag{3.4}$$

Consequently, from (3.3) and (3.4) we have

$$(u^k - u^*)^T d(u^k, \tilde{u}^k) \geq \frac{1}{2}\big(\|d(u^k, \tilde{u}^k)\|^2 + (1-\nu^2)\|u^k - \tilde{u}^k\|^2\big).$$

This means that $d(u^k, \tilde{u}^k)$ is an ascent direction of the unknown distance function $\frac{1}{2}\|u - u^*\|^2$ at the point u^k.

3.2 Geminate ascent directions

To the direction $d(u^k, \tilde{u}^k)$ defined in (3.2), there is a correlative ascent direction $\beta_k F(\tilde{u}^k)$. Use the notation of $d(u^k, \tilde{u}^k)$, the projection equation (1.3a) can be written as

$$\tilde{u}^k = P_\Omega\{\tilde{u}^k - [\beta_k F(\tilde{u}^k) - d(u^k, \tilde{u}^k)]\}. \tag{3.5a}$$

It follows that $\tilde{u}^k$ is a solution of VI(Ω, F) if and only if $d(u^k, \tilde{u}^k) = 0$. Assume that there is a constant $c > 0$ such that

$$\varrho_k = \frac{(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2} \geq c, \quad \forall k \geq 0. \tag{3.5b}$$

In this paper, we call (3.5) with $c > 0$ *the general conditions* and the forthcoming analysis is based of these conditions. For given u^k, there are different ways to construct $\tilde{u}^k$ and $d(u^k, \tilde{u}^k)$ which satisfy the conditions (3.5) (see [8] for an example). If β_k satisfies (1.3b) and $d(u^k, \tilde{u}^k)$ is given by (3.2), the general conditions (3.5) are satisfied with $c \geq \frac{1}{2}$ (see (3.4)). Note that an equivalent expression of (3.5a) is

$$\tilde{u}^k \in \Omega, \quad (u - \tilde{u}^k)^T\{\beta_k F(\tilde{u}^k) - d(u^k, \tilde{u}^k)\} \geq 0, \quad \forall\, u \in \Omega, \tag{3.6a}$$

and from (3.5b) we have

$$(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) = \varrho_k \|d(u^k, \tilde{u}^k)\|^2. \tag{3.6b}$$

In fact, $d(u^k, \tilde{u}^k)$ and $\beta_k F(\tilde{u}^k)$ in (3.5a) are a pair of geminate directions and usually denoted by $d_1(u^k, \tilde{u}^k)$ and $d_2(u^k, \tilde{u}^k)$, respectively. In this paper, we restrict $d_2(u^k, \tilde{u}^k)$ to be $F(\tilde{u}^k)$ times a positive scalar β_k. If $d(u^k, \tilde{u}^k) = u^k - \tilde{u}^k$, then $\tilde{u}^k$ in (3.6a) is the solution of the subproblem in the k-th iteration when PPA applied to solve VI(Ω, F). Hence, the projection and contraction methods considered in this paper belong to the *prox-like contraction methods*.

The following lemmas tell us that both the direction $d(u^k, \tilde{u}^k)$ (for $u^k \in \Re^n$) and $F(\tilde{u}^k)$

(for $u^k \in \Omega$) are ascent directions of the function $\frac{1}{2}\|u - u^*\|^2$ whenever u^k is not a solution point. The proof is similar to those in [7], for completeness sake of this paper, we restate the short proofs.

Lemma 3.1 *Let the general conditions* (3.5) *be satisfied. Then we have*

$$(u^k - u^*)^T d(u^k, \tilde{u}^k) \geq \varrho_k \|d(u^k, \tilde{u}^k)\|^2, \quad \forall\, u^k \in \Re^n,\ u^* \in \Omega^*. \tag{3.7}$$

Proof. Note that $u^* \in \Omega$. By setting $u = u^*$ in (3.6a) (the equivalent expression of (3.5a)), we get

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

The last inequality follows from the monotonicity of F and $(\tilde{u}^k - u^*)^T F(u^*) \geq 0$. Therefore,

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

The assertion (3.7) is followed from the above inequality and (3.6b) directly. $\square$

Lemma 3.2 *Let the general conditions* (3.5) *be satisfied. If* $u^k \in \Omega$*, then we have*

$$(u^k - u^*)^T \beta_k F(\tilde{u}^k) \geq \varrho_k \|d(u^k, \tilde{u}^k)\|^2, \quad \forall\, u^* \in \Omega^*. \tag{3.8}$$

Proof. Since $(\tilde{u}^k - u^*)^T \beta_k F(\tilde{u}^k) \geq 0$, we have

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

Note that because $u^k \in \Omega$, by setting $u = u^k$ in (3.6a), we get

$$(u^k - \tilde{u}^k)^T \beta_k F(\tilde{u}^k) \geq (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k).$$

From the above two inequalities follows that

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

The assertion (3.8) is followed from the above inequality and (3.6b) directly. $\square$

Note that (3.7) holds for $u^k \in \Re^n$ while (3.8) is hold only for $u^k \in \Omega$.

4 Corrector in the contraction sense

Based on the pair of geminate ascent directions in (3.5), namely, $d(u^k, \tilde{u}^k)$ and $\beta_k F(\tilde{u}^k)$, we use the one of the following corrector forms to update the new iterate u^{k+1}:

(Correction of PC Method-I) $$u_I^{k+1} = u^k - \gamma \varrho_k d(u^k, \tilde{u}^k), \tag{4.1a}$$

or

(Correction of PC Method-II) $$u_{II}^{k+1} = P_\Omega[u^k - \gamma \varrho_k \beta_k F(\tilde{u}^k)], \tag{4.1b}$$

where $\gamma \in (0, 2)$ and ϱ_k is defined in (3.5b). Note that the same step size length is used in (4.1a) and (4.1b) even if the search directions are different. Recall that $\tilde{u}^k$ is obtained via a projection, by using the correction form (4.1b), we have to make an additional projection on Ω in the PC methods. Replacing $\gamma \varrho_k$ in (4.1b) by 1, it reduces to the update form of the extragradient method (see (1.4)).

For any solution point $u^* \in \Omega^*$, we define

$$\vartheta_I(\gamma) = \|u^k - u^*\|^2 - \|u_I^{k+1} - u^*\|^2 \tag{4.2a}$$

and

$$\vartheta_{II}(\gamma) = \|u^k - u^*\|^2 - \|u_{II}^{k+1} - u^*\|^2, \tag{4.2b}$$

which measure the profit in the k-th iteration. The following theorem gives a lower bound of the profit function, the similar results were established in [6, 7, 8].

Theorem 4.1 *For given u^k, let the general conditions* (3.5) *be satisfied. If the corrector is updated by* (4.1a) *or* (4.1b)*, then for any $u^* \in \Omega^*$ and $\gamma > 0$, we have*

$$\vartheta_I(\gamma) \geq q(\gamma), \tag{4.3}$$

and

$$\vartheta_{II}(\gamma) \geq q(\gamma) + \|u_I^{k+1} - u_{II}^{k+1}\|^2, \tag{4.4}$$

respectively, where

$$q(\gamma) = \gamma(2 - \gamma)\varrho_k^2 \|d(u^k, \tilde{u}^k)\|^2. \tag{4.5}$$

Proof. Using the definition of $\vartheta_I(\gamma)$ and u_I^{k+1} (see (4.1a)), we have

$$\begin{aligned} \vartheta_I(\gamma) &= \|u^k - u^*\|^2 - \|u^k - u^* - \gamma \varrho_k d(u^k, \tilde{u}^k)\|^2 \\ &= 2\gamma \varrho_k (u^k - u^*)^T d(u^k, \tilde{u}^k) - \gamma^2 \varrho_k^2 \|d(u^k, \tilde{u}^k)\|^2. \end{aligned} \tag{4.6}$$

Recalling (3.7), we obtain

$$2\gamma\varrho_k(u^k - u^*)^T d(u^k, \tilde{u}^k) \geq 2\gamma\varrho_k^2 \|d(u^k, \tilde{u}^k)\|^2.$$

Substituting it in (4.6) and using the definition of $q(\gamma)$, we get $\vartheta_I(\gamma) \geq q(\gamma)$ and the first assertion is proved. Now, we turn to show the second assertion. Because

$$u_{II}^{k+1} = P_\Omega[u^k - \gamma\varrho_k\beta_k F(\tilde{u}^k)],$$

and $u^* \in \Omega$, by setting $u = u^*$ and $v = u^k - \gamma\varrho_k\beta_k F(\tilde{u}^k)$ in (2.2), we have

$$\begin{aligned} \|u^* - u_{II}^{k+1}\|^2 \quad \leq \quad & \|u^* - (u^k - \gamma\varrho_k\beta_k F(\tilde{u}^k))\|^2 \\ & -\|u^k - \gamma\varrho_k\beta_k F(\tilde{u}^k) - u_{II}^{k+1}\|^2. \end{aligned} \tag{4.7}$$

Thus,

$$\begin{aligned} \vartheta_{II}(\gamma) \quad &= \quad \|u^k - u^*\|^2 - \|u_{II}^{k+1} - u^*\|^2 \\ &\geq \quad \|u^k - u^*\|^2 - \|(u^k - u^*) - \gamma\varrho_k\beta_k F(\tilde{u}^k)\|^2 \\ &\qquad +\|(u^k - u_{II}^{k+1}) - \gamma\varrho_k\beta_k F(\tilde{u}^k)\|^2 \\ &= \quad \|u^k - u_{II}^{k+1}\|^2 + 2\gamma\varrho_k\beta_k(u_{II}^{k+1} - u^*)^T F(\tilde{u}^k) \\ &\geq \quad \|u^k - u_{II}^{k+1}\|^2 + 2\gamma\varrho_k\beta_k(u_{II}^{k+1} - \tilde{u}^k)^T F(\tilde{u}^k). \end{aligned} \tag{4.8}$$

The last inequality in (4.8) follows from $(\tilde{u}^k - u^*)^T F(\tilde{u}^k) \geq 0$. Since $u_{II}^{k+1} \in \Omega$, by setting $u = u_{II}^{k+1}$ in (3.6a), we get

$$(u_{II}^{k+1} - \tilde{u}^k)^T\{\beta_k F(\tilde{u}^k) - d(u^k, \tilde{u}^k)\} \geq 0,$$

and consequently, substituting it in the right hand side of (4.8), we obtain

$$\begin{aligned} \vartheta_{II}(\gamma) \quad &\geq \quad \|u^k - u_{II}^{k+1}\|^2 + 2\gamma\varrho_k(u_{II}^{k+1} - \tilde{u}^k)^T d(u^k, \tilde{u}^k) \\ &= \quad \|u^k - u_{II}^{k+1}\|^2 + 2\gamma\varrho_k(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) \\ &\qquad -2\gamma\varrho_k(u^k - u_{II}^{k+1})^T d(u^k, \tilde{u}^k). \end{aligned} \tag{4.9}$$

To the two crossed term in the right hand side of (4.9), we have (by using (3.6b))

$$2\gamma\varrho_k(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) = 2\gamma\varrho_k^2\|d(u^k, \tilde{u}^k)\|^2,$$

and

$$\begin{aligned} & -2\gamma\varrho_k(u^k - u_{II}^{k+1})^T d(u^k, \tilde{u}^k) \\ & \quad = \quad \|(u^k - u_{II}^{k+1}) - \gamma\varrho_k d(u^k, \tilde{u}^k)\|^2 \\ & \qquad -\|u^k - u_{II}^{k+1}\|^2 - \gamma^2\varrho_k^2\|d(u^k, \tilde{u}^k)\|^2, \end{aligned}$$

respectively. Substituting them in the right hand side of (4.9) and using

$$u^k - \gamma\varrho_k d(u^k, \tilde{u}^k) = u_I^{k+1},$$

we obtain

$$\begin{aligned} \vartheta_{II}(\gamma) &\geq \gamma(2-\gamma)\varrho_k^2\|d(u^k,\tilde{u}^k)\|^2 + \|u_I^{k+1} - u_{II}^{k+1}\|^2 \\ &= q(\gamma) + \|u_I^{k+1} - u_{II}^{k+1}\|^2, \end{aligned} \tag{4.10}$$

and the proof is complete. $\square$

Note that $q(\gamma)$ is a quadratic function of γ, it reaches its maximum at $\gamma^* = 1$. In practice, ϱ_k is the 'optimal' step size in (4.1) and γ is a relaxation factor. Because $q(\gamma)$ is a lower bound of $\vartheta_I(\gamma)$ (resp. $\vartheta_{II}(\gamma)$), the desirable new iterate is updated by (4.1) with $\gamma \in [1, 2)$.

From Theorem 4.1 we obtain

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - \gamma(2-\gamma)\varrho_k^2\|d(u^k,\tilde{u}^k)\|^2. \tag{4.11}$$

Convergence result follows from (4.11) directly. Due to the property (4.11) we call the methods which use different update forms in (4.1) PC Method-I and PC Method II,

respectively. Note that the assertion (4.11) is derived from the general conditions (3.5). For the PC methods using correction form (1.8) or (1.9), because $\varrho_k > \frac{1}{2}$, by using (3.6b) and (1.3b), it follows from (4.11) that

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - \frac{1}{2}\gamma(2-\gamma)(1-\nu)\|u^k - \tilde{u}^k\|^2. \tag{4.12}$$

5 Convergence rate of the PC methods

This section proves the convergence rate of the projection and contraction methods. Recall that the base of the complexity proof is (see (2.3.2) in pp. 159 of [3])

$$\Omega^* = \bigcap_{u\in\Omega} \{\tilde{u} \in \Omega : (u - \tilde{u})^T F(u) \geq 0\}. \tag{5.1}$$

In the sequel, for given $\epsilon > 0$ and $\mathcal{D} \subset \Omega$, we focus our attention to find a $\tilde{u}$ such that

$$\tilde{u} \in \Omega \quad \text{and} \quad \sup_{u\in\mathcal{D}} (\tilde{u} - u)^T F(u) \leq \epsilon. \tag{5.2}$$

Although the PC Method I uses the update form (4.1a) and it does not guarantee that

$\{u^k\}$ belongs to Ω, the sequence $\{\tilde{u}^k\} \subset \Omega$ in the PC methods with different corrector forms. Now, we prove the key inequality of the PC Method I for the complexity analysis.

Lemma 5.1 *For given $u^k \in \Re^n$, let the general conditions* (3.5) *be satisfied. If the new iterate u^{k+1} is updated by* (4.1a) *with any $\gamma > 0$, then we have*

$$(u - \tilde{u}^k)^T \gamma \varrho_k \beta_k F(\tilde{u}^k) + \frac{1}{2}\big(\|u - u^k\|^2 - \|u - u^{k+1}\|^2\big) \ge \frac{1}{2} q(\gamma), \ \ \forall\, u \in \Omega, \tag{5.3}$$

where $q(\gamma)$ is defined in (4.5)*.*

Proof. Because (due to (3.6a))

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

and (see (4.1a))

$$\gamma \varrho_k d(u^k, \tilde{u}^k) = u^k - u^{k+1},$$

we need only to show that

$$(u - \tilde{u}^k)^T (u^k - u^{k+1}) + \frac{1}{2}\big(\|u - u^k\|^2 - \|u - u^{k+1}\|^2\big) \ge \frac{1}{2} q(\gamma), \ \ \forall\, u \in \Omega. \tag{5.4}$$

To the crossed term in the left hand side of (5.4), namely $(u - \tilde{u}^k)^T (u^k - u^{k+1})$, using

an identity

$$(a - b)^T (c - d) = \frac{1}{2}\big(\|a - d\|^2 - \|a - c\|^2\big) + \frac{1}{2}\big(\|c - b\|^2 - \|d - b\|^2\big),$$

we obtain

$$\begin{aligned} (u - \tilde{u}^k)^T (u^k - u^{k+1}) \quad &= \quad \frac{1}{2}\big(\|u - u^{k+1}\|^2 - \|u - u^k\|^2\big) \\ &\quad + \frac{1}{2}\big(\|u^k - \tilde{u}^k\|^2 - \|u^{k+1} - \tilde{u}^k\|^2\big). \end{aligned} \tag{5.5}$$

By using $u^{k+1} = u^k - \gamma \varrho_k d(u^k, \tilde{u}^k)$ and (3.6b), we get

$$\begin{aligned} &\|u^k - \tilde{u}^k\|^2 - \|u^{k+1} - \tilde{u}^k\|^2 \\ &= \quad \|u^k - \tilde{u}^k\|^2 - \|(u^k - \tilde{u}^k) - \gamma \varrho_k d(u^k, \tilde{u}^k)\|^2 \\ &= \quad 2\gamma \varrho_k (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \gamma^2 \varrho_k^2 \|d(u^k, \tilde{u}^k)\|^2 \\ &= \quad \gamma(2 - \gamma) \varrho_k^2 \|d(u^k, \tilde{u}^k)\|^2. \end{aligned}$$

Substituting it in the right hand side of (5.5) and using the definition of $q(\gamma)$, we obtain (5.4) and the lemma is proved. $\square$

The both sequences $\{\tilde{u}^k\}$ and $\{u^k\}$ in the PC method II belong to Ω. In the following

lemma we prove the same assertion for PC method II as in Lemma 5.1.

Lemma 5.2 *For given $u^k \in \Omega$, let the general conditions* (3.5) *be satisfied. If the new iterate u^{k+1} is updated by* (4.1b) *with any $\gamma > 0$, then we have*

$$(u - \tilde{u}^k)^T \gamma \varrho_k \beta_k F(\tilde{u}^k) + \frac{1}{2}\big(\|u - u^k\|^2 - \|u - u^{k+1}\|^2\big) \ge \frac{1}{2} q(\gamma), \ \ \forall\, u \in \Omega, \tag{5.6}$$

where $q(\gamma)$ is defined in (4.5)*.*

Proof. For investigating $(u - \tilde{u}^k)^T \beta_k F(\tilde{u}^k)$, we divide it in the terms

$$(u^{k+1} - \tilde{u}^k)^T \gamma \varrho_k \beta_k F(\tilde{u}^k) \qquad \text{and} \qquad (u - u^{k+1})^T \gamma \varrho_k \beta_k F(\tilde{u}^k).$$

First, we deal with the term $(u^{k+1} - \tilde{u}^k)^T \gamma \varrho_k \beta_k F(\tilde{u}^k)$. Since $u^{k+1} \in \Omega$, substituting $u = u^{k+1}$ in (3.6a) we get

$$\begin{aligned} &(u^{k+1} - \tilde{u}^k)^T \gamma \varrho_k \beta_k F(\tilde{u}^k) \\ &\quad \ge \quad \gamma \varrho_k (u^{k+1} - \tilde{u}^k)^T d(u^k, \tilde{u}^k) \\ &\quad = \quad \gamma \varrho_k (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \gamma \varrho_k (u^k - u^{k+1})^T d(u^k, \tilde{u}^k). \end{aligned} \tag{5.7}$$

To the first crossed term of the right hand side of (5.7), using (3.6b), we have

$$\gamma \varrho_k (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) = \gamma \varrho_k^2 \|d(u^k, \tilde{u}^k)\|^2.$$

To the second crossed term of the right hand side of (5.7), using the Cauchy-Schwarz Inequality, we get

$$-\gamma \varrho_k (u^k - u^{k+1})^T d(u^k, \tilde{u}^k) \ge -\frac{1}{2}\|u^k - u^{k+1}\|^2 - \frac{1}{2}\gamma^2 \varrho_k^2 \|d(u^k, \tilde{u}^k)\|^2.$$

Substituting them in the right hand side of (5.7), we obtain

$$(u^{k+1} - \tilde{u}^k)^T \gamma \varrho_k \beta_k F(\tilde{u}^k) \ge \frac{1}{2}\gamma(2-\gamma)\varrho_k^2 \|d(u^k, \tilde{u}^k)\|^2 - \frac{1}{2}\|u^k - u^{k+1}\|^2. \tag{5.8}$$

Now, we turn to treat of the term $(u - u^{k+1})^T \gamma \varrho_k \beta_k F(\tilde{u}^k)$. Since u^{k+1} is updated by (4.1b), u^{k+1} is the projection of $\big(u^k - \gamma \varrho_k \beta_k F(\tilde{u}^k)\big)$ on Ω, it follows from (2.1) that

$$\big\{\big(u^k - \gamma \varrho_k \beta_k F(\tilde{u}^k)\big) - u^{k+1}\big\}^T \big(u - u^{k+1}\big) \le 0, \quad \forall\, u \in \Omega,$$

and consequently

$$\big(u - u^{k+1}\big)^T \gamma \varrho_k \beta_k F(\tilde{u}^k) \ge \big(u - u^{k+1}\big)^T \big(u^k - u^{k+1}\big), \quad \forall\, u \in \Omega.$$

Using the identity $a^T b = \frac{1}{2}\{\|a\|^2 - \|a-b\|^2 + \|b\|^2\}$ to the right hand side of the last

inequality, we obtain

$$\big(u-u^{k+1}\big)^T\gamma\varrho_k\beta_kF(\tilde{u}^k)\ge\frac{1}{2}\big(\|u-u^{k+1}\|^2-\|u-u^k\|^2\big)+\frac{1}{2}\|u^k-u^{k+1}\|^2. \tag{5.9}$$

Adding (5.8) and (5.9) and using the definition of $q(\gamma)$, we get (5.6) and the proof is complete. $\square$

For the different projection and contraction methods, we have the same key inequality which is shown in Lemma 5.1 and Lemma 5.2, respectively. By setting $u=u^*$ in (5.3) and (5.6), we get

$$\|u^k-u^*\|^2-\|u^{k+1}-u^*\|^2\ge 2\gamma\varrho_k\beta_k(\tilde{u}^k-u^*)^TF(\tilde{u}^k)+q(\gamma).$$

Because $(\tilde{u}^k-u^*)^TF(\tilde{u}^k)\ge(\tilde{u}^k-u^*)^TF(u^*)\ge 0$ and $q(\gamma)=\gamma(2-\gamma)\varrho_k^2\|d(u^k,\tilde{u}^k)\|^2$, it follows from the last inequality that

$$\|u^{k+1}-u^*\|^2\le\|u^k-u^*\|^2-\gamma(2-\gamma)\varrho_k^2\|d(u^k,\tilde{u}^k)\|^2.$$

This is just the form (4.11) in Section 4. In other words, the contraction property (4.11) of PC methods is the consequent result of Lemma 5.1 and Lemma 5.2, respectively.

For the convergence rate proof, we allow $\gamma\in(0,2]$. In this case, we still have $q(\gamma)\ge 0$. By using the monotonicity of F, from (5.3) and (5.6) we get

$$(u-\tilde{u}^k)^T\varrho_k\beta_kF(u)+\frac{1}{2\gamma}\|u-u^k\|^2\ge\frac{1}{2\gamma}\|u-u^{k+1}\|^2,\quad\forall\,u\in\Omega. \tag{5.10}$$

This inequality is essential for the convergence rate proofs.

Theorem 5.1 *For any integer $t>0$, we have a $\tilde{u}_t\in\Omega$ which satisfies*

$$(\tilde{u}_t-u)^TF(u)\le\frac{1}{2\gamma\Upsilon_t}\|u-u^0\|^2,\quad\forall u\in\Omega, \tag{5.11}$$

where

$$\tilde{u}_t=\frac{1}{\Upsilon_t}\sum_{k=0}^{t}\varrho_k\beta_k\tilde{u}^k\qquad\textit{and}\qquad\Upsilon_t=\sum_{k=0}^{t}\varrho_k\beta_k. \tag{5.12}$$

Proof. Summing the inequality (5.10) over $k=0,\ldots,t$, we obtain

$$\Big((\sum_{k=0}^{t}\varrho_k\beta_k)u-\sum_{k=0}^{t}\varrho_k\beta_k\tilde{u}^k\Big)^TF(u)+\frac{1}{2\gamma}\|u-u^0\|^2\ge 0,\quad\forall u\in\Omega.$$

Using the notations of Υ_t and $\tilde{u}_t$ in the above inequality, we derive

$$(\tilde{u}_t - u)^T F(u) \le \frac{\|u - u^0\|^2}{2\gamma\Upsilon_t}, \quad \forall u \in \Omega.$$

Indeed, $\tilde{u}_t \in \Omega$ because it is a convex combination of $\tilde{u}^0, \tilde{u}^1, \ldots, \tilde{u}^t$. The proof is complete. $\square$

For given u^k, the predictor $\tilde{u}^k$ is given by (1.3a) and the prediction step size β_k satisfies the condition (1.3b). Thus, the general conditions (3.5) are satisfied with $\varrho_k \ge c = \frac{1}{2}$. We choose (4.1a) (for the case that u^k is not necessary in Ω) or (4.1b) (for the case that $u^k \in \Omega$) to generate the new iterate u^{k+1}. Because $\varrho_k \ge \frac{1}{2}$, $\inf_{k\ge 0}\{\beta_k\} \ge \beta_L$ and $\beta_L = O(1/L)$, it follows from (5.12) that

$$\Upsilon_t \ge \frac{t+1}{2}\beta_L,$$

and thus the PC methods have $O(1/t)$ convergence rate. For any substantial set $\mathcal{D} \subset \Omega$, the PC methods reach

$$(\tilde{u}_t - u)^T F(u) \le \epsilon, \quad \forall\, u \in \mathcal{D}, \qquad \text{in at most} \qquad t = \left\lceil \frac{D^2}{\gamma\beta_L\,\epsilon} \right\rceil$$

iterations, where $\tilde{u}_t$ is defined in (5.12) and $D = \sup\{\|u - u^0\| \mid u \in \mathcal{D}\}$. This convergence rate is in the ergodic sense, the statement (5.11) suggests us to take a larger parameter $\gamma \in (0, 2]$ in the correction steps of the PC methods.

6 Conclusions and Remarks

In a unified framework, we proved the $O(1/t)$ convergence rate of the projection and contraction methods for monotone variational inequalities. The convergence rate is the same as that for the extragradient method. In fact, our convergence rate include the extragradient method as a special case. The complexity analysis in this paper is based on the general conditions (3.5) and thus can be extended to a broaden class of similar contraction methods. Preliminary numerical results indicate that the PC methods do outperform the extragradient method.

Besides its own applications, the solution methods study for variational inequality is helpful for investigating the splitting contraction methods for separable convex optimization. Following we give a sketch for such developments.

6.1 变分不等式的投影收缩算法

设 $\Omega \subset \Re^n$ 是一个非空闭凸集, F 是 $\Re^n \to \Re^n$ 的一个映射. 考虑单调变分不等式：

$$u^* \in \Omega, \quad (u-u^*)^\top F(u^*) \ge 0, \quad \forall\, u \in \Omega. \tag{6.1}$$

我们说变分不等式单调, 是指其中的算子 F 满足 $(u-v)^\top (F(u)-F(v)) \ge 0$.

引理 1 设 $\mathcal{X} \subset \Re^n$ 是闭凸集, $\theta(x)$ 和 $f(x)$ 都是凸函数, 其中 $f(x)$ 可微. 记 x^* 是凸优化问题 $\min\{\theta(x)+f(x) \,|\, x \in \mathcal{X}\}$ 的解. 我们有

$$x^* = \arg\min\{\theta(x)+f(x) \,|\, x \in \mathcal{X}\} \tag{6.2a}$$

的充分必要条件是

$$x^* \in \mathcal{X}, \quad \theta(x)-\theta(x^*)+(x-x^*)^\top \nabla f(x^*) \ge 0, \quad \forall\, x \in \mathcal{X}. \tag{6.2b}$$

在求解 (6.1) 的投影收缩算法中, 对给定的当前点 u^k 和 $\beta_k > 0$, 我们利用投影

$$\tilde{u}^k = P_\Omega[u^k - \beta_k F(u^k)] = \text{Arg}\min\{\tfrac{1}{2}\|u-[u^k-\beta_k F(u^k)]\|^2 \,|\, u \in \Omega\} \tag{6.3}$$

生成一个预测点 $\tilde{u}^k$. 在投影 (6.3) 中, 我们假设选取的 β_k 满足

$$\beta_k \|F(u^k)-F(\tilde{u}^k)\| \le \nu \|u^k-\tilde{u}^k\|, \quad \nu \in (0,1). \tag{6.4}$$

由于 $\tilde{u}^k$ 是极小化问题 $\min\{\frac{1}{2}\|u-[u^k-\beta_k F(u^k)]\|^2 \,|\, u \in \Omega\}$ 的解, 根据引理 6.1 有

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

上式两边都加上 $(u-\tilde{u}^k)^\top d(u^k,\tilde{u}^k)$, 其中

$$d(u^k,\tilde{u}^k) = (u^k-\tilde{u}^k) - \beta_k[F(u^k)-F(\tilde{u}^k)], \tag{6.5}$$

就有我们需要的预测公式

[预测] $$\tilde{u}^k \in \Omega, \quad (u-\tilde{u}^k)^\top \beta_k F(\tilde{u}^k) \ge (u-\tilde{u}^k)^\top d(u^k,\tilde{u}^k), \quad \forall u \in \Omega. \tag{6.6}$$

将 (6.6) 中任意的 $u \in \Omega$ 选成 u^*, 就得到

$$(\tilde{u}^k-u^*)^\top d(u^k,\tilde{u}^k) \ge \beta_k(\tilde{u}^k-u^*)^\top F(\tilde{u}^k). \tag{6.7}$$

由单调性, $(\tilde{u}^k-u^*)^\top F(\tilde{u}^k) \ge (\tilde{u}^k-u^*)^\top F(u^*) \ge 0$, (6.7) 的右端非负, 随后得到

$$(u^k-u^*)^\top d(u^k,\tilde{u}^k) \ge (u^k-\tilde{u}^k)^\top d(u^k,\tilde{u}^k). \tag{6.8}$$

由 $d(u^k,\tilde{u}^k)$ 的表达式 (6.5) 和假设 (6.4), 利用 Cauchy-Schwarz 不等式便可得到

$$(u^k-\tilde{u}^k)^\top d(u^k,\tilde{u}^k) \ge (1-\nu)\|u^k-\tilde{u}^k\|^2. \tag{6.9}$$

因此，不等式 (6.8) 的右端为正. 这表示, 对任何正定矩阵 $H \in \Re^{n\times n}$, $H^{-1}d(u^k,\tilde{u}^k)$ 是未知函数 $\frac{1}{2}\|u-u^*\|_H^2$ 在 u^k 处 H-模意义下的一个上升方向. 我们用

[校正] $$u^{k+1} = u^k - \alpha H^{-1} d(u^k, \tilde{u}^k) \tag{6.10}$$

产生新的迭代点, 其中 $d(u^k, \tilde{u}^k)$ 由 (6.5) 给出. 考察与 α 相关的距离平方缩短量,

$$\vartheta_k(\alpha) := \|u^k - u^*\|_H^2 - \|u^{k+1} - u^*\|_H^2. \tag{6.11}$$

根据定义

$$\begin{aligned} \vartheta_k(\alpha) &= \|u^k - u^*\|_H^2 - \|u^k - u^* - \alpha H^{-1} d(u^k, \tilde{u}^k)\|_H^2 \\ &= 2\alpha (u^k - u^*)^\top d(u^k, \tilde{u}^k) - \alpha^2 \|H^{-1} d(u^k, \tilde{u}^k)\|_H^2 \\ \text{由(6.8)} \quad &\geq 2\alpha (u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k) - \alpha^2 \|H^{-1} d(u^k, \tilde{u}^k)\|_H^2 := q_k(\alpha) \end{aligned} \tag{6.12}$$

上式说明 $q_k(\alpha)$ 是 $\vartheta_k(\alpha)$ 的一个下界函数. 在投影收缩算法中, 我们往往考虑在欧氏模下收缩, 即 H 为单位矩阵. 在此情况下,

$$\alpha_k^* = \text{argmax}\{q_k(\alpha)\} = \frac{(u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2}. \tag{6.13}$$

由假设 (6.4), 可得 $2(u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k) > \|d(u^k, \tilde{u}^k)\|^2$, 因而 $\alpha_k^* > \frac{1}{2}$. 接着就有

$$\begin{aligned} &\|u^k - u^*\|^2 - \|u^{k+1} - u^*\|^2 \\ &\quad \geq q(\alpha_k^*) = \alpha_k^* (u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k) \geq \tfrac{1}{2}(1-\nu)\|u^k - \tilde{u}^k\|^2. \end{aligned} \tag{6.14}$$

这是证明算法收敛的关键不等式.

✠ 在投影收缩算法中, 分处不等式 (6.6) 两端的

$$\beta_k F(\tilde{u}^k) \quad \text{和} \quad d(u^k, \tilde{u}^k), \qquad \text{我们称之为一对孪生方向.}$$

奇妙的是, 采用孪生方向, 相同步长的两种不同校正方法, 具有相同的收敛性质.

在实际计算中, 我们采用校正公式

$$\text{(收缩算法–1)} \quad u_I^{k+1} = u^k - \gamma \alpha_k^* d(u^k, \tilde{u}^k) \tag{6.15}$$

或者

$$\text{(收缩算法–2)} \quad u_{II}^{k+1} = P_\Omega[u^k - \gamma \alpha_k^* \beta_k F(\tilde{u}^k)] \tag{6.16}$$

产生新的迭代点 u^{k+1}, 其中 $\gamma \in (1.5, 1.8) \subset (0, 2)$, α_k^* 都由 (6.13) 给出.

证明在《高等学校计算数学学报》38卷 1, (2016) 74-96 或我主页的报告1中可以找到.

6.2 凸优化的分裂收缩算法

我们以两个可分离算子的线性约束凸优化问题

$$\min\{\theta_1(x)+\theta_2(y) \mid Ax+By=b,\ x\in\mathcal{X}, y\in\mathcal{Y}\}. \tag{6.17}$$

为例. 这个问题的拉格朗日函数是定义在 $\mathcal{X}\times\mathcal{Y}\times\Re^m$ 上的

$$L(x,y,\lambda)=\theta_1(x)+\theta_2(y)-\lambda^\top(Ax+By-b). \tag{6.18}$$

一对 $\big((x^*,y^*),\lambda^*\big)$ 满足

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

就被称为拉格朗日函数的鞍点. 换句话说, 鞍点就是同时满足

$$\left\{\begin{array}{l} x^*=\arg\min\{L(x,y^*,\lambda^*)\,|\,x\in\mathcal{X}\},\\ y^*=\arg\min\{L(x^*,y,\lambda^*)\,|\,y\in\mathcal{Y}\},\\ \lambda^*=\arg\max\{L(x^*,y^*,\lambda)\,|\,\lambda\in\Re^m\}.\end{array}\right.$$

根据 (6.2), 上述优化子问题的最优性条件是

$$\left\{\begin{array}{lll} x^*\in\mathcal{X}, & \theta_1(x)-\theta_1(x^*)+(x-x^*)^\top(-A^\top\lambda^*)\ge 0, & \forall\, x\in\mathcal{X},\\ y^*\in\mathcal{Y}, & \theta_2(y)-\theta_2(y^*)+(y-y^*)^\top(-B^\top\lambda^*)\ge 0, & \forall\, y\in\mathcal{Y},\\ \lambda^*\in\Re^m, & (\lambda-\lambda^*)^\top(Ax^*+By^*-b)\ge 0, & \forall\,\lambda\in\Re^m.\end{array}\right. \tag{6.19}$$

换句话说, 拉格朗日函数的鞍点 $\big((x^*,y^*),\lambda^*\big)$ 可以表示成以下变分不等式的解:

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^\top F(w^*)\ge 0,\quad \forall\, w\in\Omega. \tag{6.20}$$

其中 $\theta(u)=\theta_1(x)+\theta_2(y)$,

$$u=\begin{pmatrix} x\\ y\end{pmatrix},\quad w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix},\quad F(w)=\begin{pmatrix} -A^\top\lambda\\ -B^\top\lambda\\ Ax+By-b\end{pmatrix}. \tag{6.21}$$

$\Omega=\mathcal{X}\times\mathcal{Y}\times\Re^m$. 注意到, (6.21) 中的 $F(w)$, 恰有 $(w-\tilde{w})^\top(F(w)-F(\tilde{w}))=0$, 所以也是单调的. 求解 (6.20) 这样的问题, 许多方法都属于一个预测-校正方法框架.

[预测]. 对给定的 v^k (v 是向量 w 的部分分量), 求得预测点 $\tilde{w}^k\in\Omega$, 使其满足

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^\top F(\tilde{w}^k)\ge(v-\tilde{v}^k)^\top Q(v^k-\tilde{v}^k),\ \forall w\in\Omega, \tag{6.22}$$

其中矩阵 Q 不一定对称, 但是 $Q^\top+Q$ 正定.

将 (6.22) 中任意的 $w\in\Omega$ 选成 w^*, 就得到

$$(\tilde{v}^k-v^*)^\top Q(v^k-\tilde{u}^k)\ge\theta(\tilde{u}^k)-\theta(u^*)+(\tilde{w}^k-w^*)^\top F(\tilde{w}^k). \tag{6.23}$$

由 $\theta(\tilde{u}^k)-\theta(u^*)+(\tilde{w}^k-w^*)^\top F(\tilde{w}^k)=\theta(\tilde{u}^k)-\theta(u^*)+(\tilde{w}^k-w^*)^\top F(w^*)\ge 0$. 因而 (6.23) 的右端非负, $(\tilde{v}^k-v^*)^\top Q(v^k-\tilde{u}^k)\ge 0$. 随后得到

$$(v^k-v^*)^\top Q(v^k-\tilde{v}^k)\ge(v^k-\tilde{v}^k)^\top Q(v^k-\tilde{v}^k). \tag{6.24}$$

不等式 (6.24) 和 Q 正定告诉我们, $H^{-1}Q(v^k-\tilde{v}^k)$ 是未知函数 $\frac{1}{2}\|v-v^*\|_H^2$ 在 v^k 处 H-模意义下的一个上升方向. 我们用

$$v^{k+1}=v^k-\alpha H^{-1}Q(v^k-\tilde{v}^k) \tag{6.25}$$

产生新的迭代点. 记 $M=H^{-1}Q$, 校正公式就可以写成

[校正]. 新的核心变量 v^{k+1} 由

$$v^{k+1}=v^k-\alpha M(v^k-\tilde{v}^k). \tag{6.26}$$

考察与 α 相关的距离平方缩短量,

$$\vartheta_k(\alpha):=\|v^k-v^*\|_H^2-\|v^{k+1}-v^*\|_H^2. \tag{6.27}$$

根据定义

$$\begin{aligned}\vartheta_k(\alpha) &= \|v^k-v^*\|_H^2-\|v^k-v^*-\alpha M(v^k-\tilde{v}^k)\|_H^2\\ &= 2\alpha(v^k-v^*)^\top Q(v^k-\tilde{v}^k)-\alpha^2\|M(v^k-\tilde{v}^k)\|_H^2\\ &\ge 2\alpha(v^k-\tilde{v}^k)^\top Q(v^k-\tilde{v}^k)-\alpha^2\|M(v^k-\tilde{v}^k)\|_H^2:=q_k(\alpha)\end{aligned} \tag{6.28}$$

$q_k(\alpha)$ 是 $\vartheta_k(\alpha)$ 的一个下界函数. 使 $q_k(\alpha)$ 达到极大的 α_k^* 是

$$\alpha_k^*=\text{argmax}\{q_k(\alpha)\}=\frac{(v^k-\tilde{v}^k)^\top Q(v^k-\tilde{v}^k)}{\|M(v^k-\tilde{v}^k)\|_H^2}. \tag{6.29}$$

更进一步, 如果矩阵

$$G=Q^\top+Q-M^\top HM\succ 0. \tag{6.30}$$

容易证明 $2(v^k-\tilde{v}^k)^\top Q(v^k-\tilde{v}^k)>\|M(v^k-\tilde{v}^k)\|_H^2$, 因而 $\alpha_k^*>\frac{1}{2}$. 接着就有

$$\begin{aligned}&\|v^k-v^*\|_H^2-\|v^{k+1}-v^*\|_H^2\\ &\ge\ q(\alpha_k^*)=\alpha_k^*(v^k-\tilde{v}^k)^\top Q(v^k-\tilde{v}^k)\ \ge\ \frac{1}{4}\|v^k-\tilde{v}^k\|_{(Q^T+Q)}^2.\end{aligned} \tag{6.31}$$

另外, 在条件 (6.30) 满足的情况下, 可以在 (6.26) 中取步长 $\alpha=1$, 由校正公式

$$v^{k+1}=v^k-M(v^k-\tilde{v}^k)$$

生成序列 $\{v^k\}$. 由 (6.28), $q(1)=\|v^k-\tilde{v}^k\|_G^2$. 因此, $\{v^k\}$ 有收缩性质

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

✠ 从 变分不等式的投影收缩算法 到 凸优化的分裂收缩算法, 一条主线, 一个模式! ✠

References

[1] E. Blum and W. Oettli, *Mathematische Optimierung: Grundlagen und Verfahren. Ökonometrie und Unternehmensforschung*. Berlin-Heidelberg-New York, Springer-Verlag, 1975.

[2] X.J. Cai, G.Y. Gu and B. S. He, On the O(1/t) convergence rate of the projection and contraction methods for variational inequalities with Lipschitz continuous monotone operators, *Comput. Optim. Appl.,* **57**, pp. 339-363, 2013.

[3] F. Facchinei and J.S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems, Vol. I and II*. Springer Series in Operations Research. Springer Verlag, New York, 2003.

[4] P.T. Harker and J.S. Pang, A damped-Newton method for the linear complementarity problem, *Lectures in Applied Mathematics* **26**, pp. 265–284, 1990.

[5] B.S. He, A class of projection and contraction methods for monotone variational inequalities, *Applied Mathematics and optimization* **35**, pp. 69–76, 1997.

[6] B.S. He and L.-Z Liao, Improvements of some projection methods for monotone nonlinear variational inequalities, *Journal of Optimization Theory and Applications* **112**, pp. 111-128, 2002.

[7] B.S. He and M.-H. Xu, A general framework of contraction methods for monotone variational inequalities, *Pacific J. Optimization* **4**, pp. 195-212, 2008.

[8] B.S. He, X.M. Yuan and J.J.Z. Zhang, Comparison of two kinds of prediction-correction methods for monotone variational inequalities, *Computational Optimization and Applications* **27**, pp. 247-267, 2004.

[9] G.M. Korpelevich, The extragradient method for finding saddle points and other problems, *Ekonomika i Matematicheskie Metody* **12**, pp. 747-756, 1976.

[10] A. Nemirovski, Prox-method with rate of convergence $O(1/t)$ for variational inequality with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems, *SIAM J. Optimization* **15**, pp. 229-251, 2005.

[11] Y. Pan and L. Pavel, Games with coupled propagated constraints in optical networks with multi-link topologies, *Automatica* **45**, pp. 871-880, 2009.

[12] F. Sha, Large Margin Training of Acoustic Models for Speech Recognition, PhD Thesis, Computer and Information Science, University of Pennsylvania, 2007.

[13] M.V. Solodov and P. Tseng, Modified projection-type methods for monotone variational inequalities, *SIAM J. Control and Optimization* **34**, pp. 1814–1830, 1996.

[14] K. Taji, M. Fukushima and I. Ibaraki, A globally convergent Newton mmethod for solving strongly monotone variational inequalities. *Mathematical Programming* **58**, pp. 369-383, 1993.

[15] B. Taskar, S. Lacoste-Julien and M. I. Jordan. Structured prediction via Extragradient method, In Y. Weiss and B. Schoelkopf and J. Platt (Eds.), *Advances in Neural Information Processing Systems (NIPS)* **18**, 2006.

[16] P. Tseng, On accelerated proximal gradient methods for convex-concave optimization, Department of Mathematics, University of Washington, Seattle, WA 98195, USA, 2008.

凸优化和单调变分不等式的收缩算法

第六讲：为线性约束凸优化 定制的 PPA 算法及其应用

Customized PPA for linearly constrained Optimization and its applications

南京大学数学系　何炳生
hebma@nju.edu.cn

The context of this lecture is based on the publications [6, 7]

1 Introduction

从这一讲开始的四讲, 讨论的问题是

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

其中 $\theta(x)$ 是凸函数, $A \in \Re^{m\times n}$, $b \in \Re^m$, $\mathcal{X}$ 是 $\Re^n$ 中的闭凸集.

对任意给定的 $r > 0$ 和 $a \in \Re^n$, 通篇我们假设子问题

$$\min\left\{\theta(x) + \tfrac{r}{2}\|x-a\|^2 \mid x \in \mathcal{X}\right\}$$

的求解是简单的.

目的 说明将线性约束的凸优化问题转换成混合单调变分不等式, 选择适当的正定矩阵 H, 采用 H-模下的 PPA 方法. 在上述假设条件下, 线性约束的凸优化问题就变得非常容易求解.

For the analysis in this paper, we need only the basic property which is described in the following lemma.

Lemma 1.1 *Let $\mathcal{X} \subset \Re^n$ be a closed convex set, $\theta(x)$ and $f(x)$ be convex functions and $f(x)$ is differentiable. Assume that the solution set of the minimization problem $\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\}$ is nonempty. Then,*

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

if and only if

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

Proof : First, if (1.1a) is true, then for any $x \in \mathcal{X}$, we have

$$\frac{\theta(x_\alpha) - \theta(x^*)}{\alpha} + \frac{f(x_\alpha) - f(x^*)}{\alpha} \geq 0, \tag{1.2}$$

where

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

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

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

and thus

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

Substituting the last inequality in the left hand side of (1.2), we have

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

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

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

Thus (1.1b) follows from (1.1a). Conversely, since f is convex, it follow that

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

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 (1.1b), we get

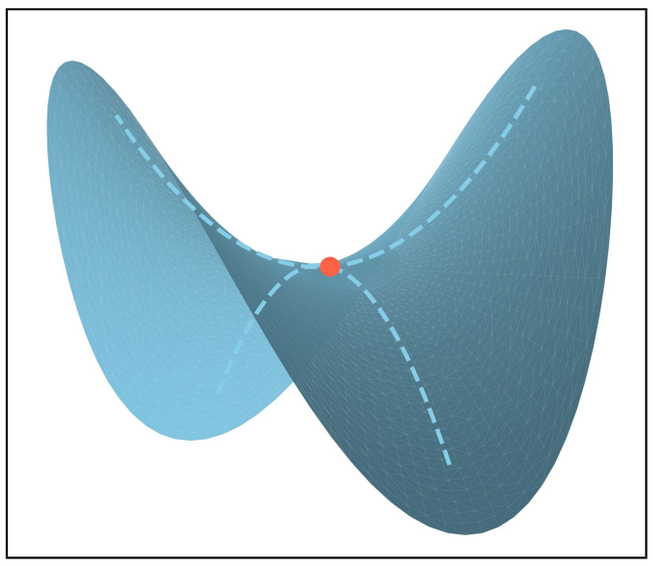
$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + f(x) - f(x^*) \ge 0, \quad \forall\, x \in \mathcal{X},$$

and (1.1a) is true. The proof is complete. $\square$

1.1 Linear constrained convex optimization and VI

We consider the linearly constrained convex optimization problem

$$\min\{\theta(x) | Ax = (\text{or } \ge)\, b, x \in \mathcal{X}\}. \quad (1.3)$$



The Lagrangian function of the problem (1.3) is

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

which is defined on $\mathcal{X} \times \Lambda$ ($\Re^m$ or $\Re^m_+$).

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

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

An equivalent expression of the saddle point is the following variational inequality:

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

The optimal condition can be characterized as a monotone variational inequality:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \qquad (1.4a)$$

where

$$w = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \Lambda. \qquad (1.4b)$$

Note that the operator F is monotone, because

$$(w - \tilde{w})^T(F(w) - F(\tilde{w})) \geq 0, \text{ Here } (w - \tilde{w})^T(F(w) - F(\tilde{w})) = 0. \qquad (1.5)$$

Example 1 of the problem (1.3): 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

$$\min\{\tfrac{1}{2}\|X - C\|_F^2 \mid \text{diag}(X) = e,\ X \in S_+^n\}, \qquad (1.6)$$

where S_+^n is the positive semi-definite cone and e is a n-vector whose each element is equal 1. The problem (1.6) is a concrete problem of type (1.3).

Example 2 of the problem (1.3): The matrix completion problem

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

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

The mathematical form of the matrix completion problem is relaxed to

$$\min\{\|X\|_* \mid X_{ij} = M_{ij},\ (ij) \in \Pi\}, \qquad (1.7)$$

where $\|\cdot\|_*$ is the nuclear norm–the sum of the singular values of a given matrix. The problem (3.5) is a convex optimization of form (1.3). The matrix A in (1.3) 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.

$$\begin{array}{cccccccccc}
* & & * & & * & & & * & & \\
 & * & & * & & & * & & & \\
* & & * & & * & & * & & & * \\
 & * & & & * & & & * & & \\
 & * & & * & & * & & & & * \\
* & & * & & * & & & * & & \\
 & * & & * & & * & & & & * \\
* & & * & & * & * & & * & * & \\
 & & & & * & & & & & * \\
* & & & * & & & * & & * & \\
 & * & & * & & * & & * & & \\
 & & & * & & * & & & * & \\
 & & * & & * & & & * & & * \\
 & * & & * & & & * & & * &
\end{array}$$

1.2 Proximal point algorithms for convex optimization

Lemma 1.2 *Let the vectors $a, b \in \Re^n$, $H \in \Re^{n\times n}$ be a positive definite matrix. If $b^T H(a-b) \geq 0$, then we have*

$$\|b\|_H^2 \leq \|a\|_H^2 - \|a-b\|_H^2. \tag{1.8}$$

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

Convex Optimization

Now, let us consider the *simple* convex optimization

$$\min\{\theta(x) + f(x) \mid x \in \mathcal{X}\}, \tag{1.9}$$

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

For solving (1.9), the k-th iteration of the proximal point algorithm (abbreviated to PPA) [9, 11] begins with a given x^k, offers the new iterate x^{k+1} via the recursion

$$x^{k+1} = \text{Argmin}\{\theta(x) + f(x) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}. \tag{1.10}$$

Since x^{k+1} is the optimal solution of (1.10), it follows from Lemma 1.1 that

$$\theta(x) - \theta(x^{k+1}) + (x - x^{k+1})^T\{\nabla f(x^{k+1}) + r(x^{k+1} - x^k)\} \geq 0, \;\forall\, x \in \mathcal{X}. \tag{1.11}$$

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

$$(x^{k+1} - x^*)^T r(x^k - x^{k+1}) \geq \theta(x^{k+1}) - \theta(x^*) + (x^{k+1} - x^*)^T \nabla f(x^{k+1}). \tag{1.12}$$

Since

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

it follows that the right hand side of (1.12) is nonnegative. And consequently,

$$(x^{k+1} - x^*)^T (x^k - x^{k+1}) \geq 0. \tag{1.13}$$

Let $a = x^k - x^*$ and $b = x^{k+1} - x^*$ and using Lemma 1.2, we obtain

$$\|\boldsymbol{x^{k+1} - x^*}\|^2 \leq \|\boldsymbol{x^k - x^*}\|^2 - \|\boldsymbol{x^k - x^{k+1}}\|^2, \tag{1.14}$$

which is the nice convergence property of Proximal Point Algorithm.

1.3 Preliminaries of PPA for Variational Inequalities

The optimal condition of the problem (1.3) is characterized as a mixed monotone variational inequality:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{1.15}$$

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

$$\begin{aligned} \tilde{w}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \\ \geq (w - \tilde{w}^k)^T H(w^k - \tilde{w}^k), \ \forall\, w \in \Omega. \end{aligned} \tag{1.16}$$

$\tilde{w}^k$ is called the proximal point of the k-th iteration for the problem (1.15).

✠ **w^k is the solution of (1.15) if and only if $w^k = \tilde{w}^k$** ✠

Setting $w = w^*$ in (1.16), we obtain

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

Note that (see the structure of $F(w)$ in (1.4b))

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

and consequently (by using (1.15)) we obtain

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

and thus

$$(\tilde{w}^k - w^*)^T H(w^k - \tilde{w}^k) \geq 0. \tag{1.17}$$

If we take $w^{k+1} = \tilde{w}^k$, and set $a = w^k - w^*$ and $b = w^{k+1} - w^*$ in the inequality (1.17), it yields that $b^T H(a - b) \geq 0$. Using Lemma1.2, we get the nice convergence property of the PPA in H-norm:

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2. \tag{1.18}$$

Extended PPA From the inequality (1.17) consequently

$$(w^k - w^*)^T H(w^k - \tilde{w}^k) \geq \|w^k - \tilde{w}^k\|_H^2. \tag{1.19}$$

Usually, we take

$$w^{k+1} = w^k - \gamma(w^k - \tilde{w}^k), \quad \gamma = 1.5 \in (0, 2) \tag{1.20}$$

as the new iterate. Then

$$\|w^{k+1} - w^*\|_H^2 \le \|w^k - w^*\|_H^2 - \gamma(2-\gamma)\|w^k - \tilde{w}^k\|_H^2. \qquad (1.21)$$

Theorem 1.1 *For given w^k, setting the solution of* (1.16) *by $\tilde{w}^k$ and let the new iterate w^{k+1} be updated by* (1.20)*. Then, the generated sequence $\{w^k\}$ converges to a solution point of* (1.15)*.*

Sketch of the proof .

- Since (1.21) is true for every solution point, the sequence $\{w^k\}$ is bounded and $\lim_{k\to\infty} \|w^k - \tilde{w}^k\|_H = 0$.
- Because $\lim_{k\to\infty} \|w^k - \tilde{w}^k\|_H = 0$ and $\{w^k\}$ is bounded, $\{\tilde{w}^k\}$ is also bounded. Moreover, $\{w^k\}$ and $\{\tilde{w}^k\}$ have the same cluster points.
- According to (1.16), every cluster point of $\{\tilde{w}^k\}$ which is also the cluster point of $\{w^k\}$, is a solution point of the VI (1.15).
- Finally, according to the contraction inequality (1.21), the generated sequence $\{w^k\}$ can only have a cluster point and the method is convergent.

2 变分不等式框架下凸优化的 PPA 算法

对等式约束 (或不等式约束) 的凸优化问题 (1.3),

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

我们用 Customized PPA 求解与它们等价的变分不等式 (1.4). 说 Customized PPA, 是因为后面的 H-模下的 PPA 方法可以看作按需定制的.

采用 Customized PPA 方法, 要求 A^TA 的模是容易估计的. 这一讲的最后一节会说明一些有应用背景的问题, 恰能满足这些要求.

2.1 Primal-dual hybrid gradient algorithm

Since the objective is to find a saddle point of the Lagrange function, a natural idea is to use the primal-dual hybrid gradient algorithm [12]. For given (x^k, λ^k), by using the primal-dual hybrid gradient algorithm,

$$x^{k+1} = \text{Argmin}\{L(x, \lambda^k) + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}, \qquad (2.1)$$

after getting x^{k+1}, we obtain

$$\lambda^{k+1} = \text{Argmax}\{L(x^{k+1}, \lambda) - \frac{s}{2}\|\lambda - \lambda^k\|^2 \mid \lambda \in \Re^m\}. \tag{2.2}$$

Combining (2.1) and (2.2), we get

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

The compact form is

$$\theta(x) - \theta(x^{k+1}) + (w - w^{k+1})^T\{F(w^{k+1}) + Q(w^{k+1} - w^k)\} \geq 0, \ \forall w \in \Omega, \tag{2.3}$$

where

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

Remark For general, min-max problem, the primal-dual hybrid gradient algorithm is not convergent. For example, we consider a pair of the primal-dual linear programming:

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

We take the following example

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

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

The optimal solutions of this pair of linear programming are $x^* = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $y^* = 1$.

Note that its Lagrange function is

$$L(x,y) = c^T x - y^T(Ax - b) \tag{2.4}$$

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

Using the Primal-dual hybrid gradient method with $r = s = 1$. For given (x^k, y^k), Zhu and Chan's procedure is

$$x^{k+1} = \text{Argmin}\{L(x, y^k) + \frac{1}{2}\|x - x^k\|^2 \mid x \ge 0\} = \max\{(x^k + A^T y^k - c), 0\},$$

and then

$$y^{k+1} = \text{Argmax}\{L(x^{k+1}, y) - \frac{1}{2}\|y - y^k\|^2\} = y^k - (Ax^{k+1} - 1).$$

In other words, the iteration formula is

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

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

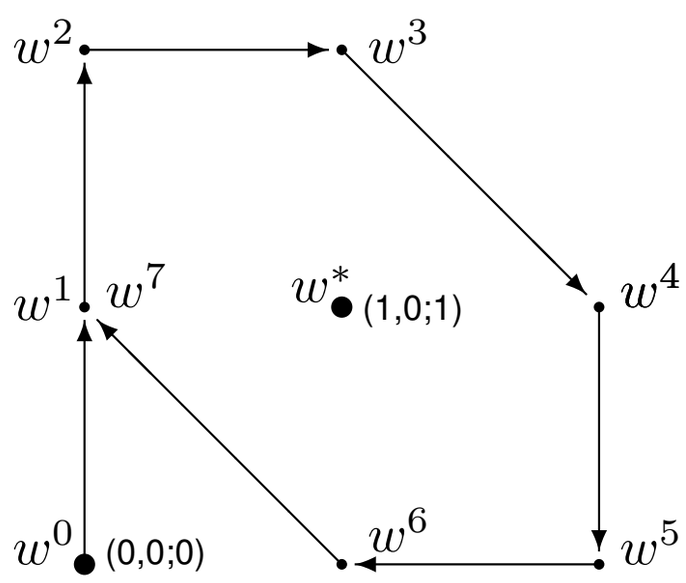


Fig. 2.1 The sequence generated by PDHG Method with $r = s = 1$

$$\begin{aligned} w^0 &= (0,0;0) \\ w^1 &= (0,0;1) \\ w^2 &= (0,0;2) \\ w^3 &= (1,0;2) \\ w^4 &= (2,0;1) \\ w^5 &= (2,0;0) \\ w^6 &= (1,0;0) \\ w^7 &= (0,0;1) \\ \boldsymbol{w^{k+6}} &\boldsymbol{= w^k} \end{aligned}$$

2.2 Customized PPA

If we can change 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 $\tilde{w}^k$ is the solution of the following variational inequality:

$$\tilde{w}^k \in \Omega, \;\; \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T \{F(\tilde{w}^k) + H(\tilde{w}^k - w^k)\} \geq 0, \; \forall w \in \Omega. \tag{2.5}$$

For this purpose, we need only change $(A\tilde{x}^k - b) + s(\tilde{\lambda}^k - \lambda^k) = 0$ to

$$(A\tilde{x}^k - b) + A(\tilde{x}^k - x^k) + s(\tilde{\lambda}^k - \lambda^k) = 0. \tag{2.6}$$

Because $\tilde{x}^k$ is known, with the given x^k and λ^k, $\tilde{\lambda}^k$ is given by

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

Thus, for given (x^k, λ^k), produce a proximal point $(\tilde{x}^k, \tilde{\lambda}^k)$ via (2.1) and (2.6) can be summarized as:

$$\tilde{x}^k = \text{argmin}\big\{L(x, \lambda^k) + \frac{r}{2}\|x - x^k\|^2 \,\big|\, x \in \mathcal{X}\big\}. \tag{2.7a}$$

$$\tilde{\lambda}^k = \text{argmax}\big\{L\big([2\tilde{x}^k - x^k], \lambda\big) - \frac{s}{2}\|\lambda - \lambda^k\|^2\big\} \tag{2.7b}$$

We call the point $(\tilde{x}^k, \tilde{\lambda}^k)$ generated by (2.7) as the predictor. The subproblem (2.7a) is

equivalent to

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

The solution of (2.7b) is given by

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

Indeed, under the assumption, the sub-problem (2.7a) is simple.

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

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

and the $\tilde{w}^k = (\tilde{x}^k, \tilde{\lambda}^k)$ generated by (2.7) is a proximal point. Based on the predictor, the new iterate is given by

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

Theorem 2.1 *The sequence* $\{w^k = (x^k, \lambda^k)\}$ *generated by the customized PPA satisfies*

$$\|w^{k+1} - w^*\|_H^2 \le \|w^k - w^*\|_H^2 - \gamma(2-\gamma)\|w^k - \tilde{w}^k\|_H^2. \tag{2.8}$$

定理 2.1 中的 (2.8) 是方法总体收敛的关键不等式. 如同 §1.3，可以由此证明算法的总体收敛性. 计算中实际困难往往在于如何选择参数 r 和 s.

我们也可以用另外的顺序产生预测点 $\tilde{u}^k = (\tilde{x}^k, \tilde{\lambda}^k)$. 对给定的 (x^k, λ^k), 求 $(\tilde{x}^k, \tilde{\lambda}^k) \in \Omega$, 使得

$$\theta(x) - \theta(\tilde{x}^k) + \begin{pmatrix} x - \tilde{x}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{\lambda}^k \\ A\tilde{x}^k - b \end{pmatrix} \right.$$
$$\left. + \begin{pmatrix} r(\tilde{x}^k - x^k) - A^T(\tilde{\lambda}^k - \lambda^k) \\ -A(\tilde{x}^k - x^k) + s(\tilde{\lambda}^k - \lambda^k) \end{pmatrix} \right\} \ge 0, \quad \forall (x, \lambda) \in \Omega. \tag{2.9}$$

这相当于求得变分不等式 (1.4) 的 PPA 子问题 (1.16) 的解, 其中对称矩阵

$$H = \begin{pmatrix} rI_n & -A^T \\ -A & sI_m \end{pmatrix}. \tag{2.10}$$

同样, 当 $rs > \|A^TA\|$ 时 G 是正定的. 注意到(2.9) 的下半部分相当于

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T\{(Ax^k - b) + s(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \quad \forall \lambda \in \Lambda,$$

它可以由

$$\tilde{\lambda}^k = P_\Lambda[\lambda^k - \frac{1}{s}(Ax^k - b)] \tag{2.11}$$

给出. 因为有了 $\tilde{\lambda}^k$, 变分不等式 (2.9) 的上半部分中未知的只有 $\tilde{x}^k$, 它可以通过求解极小化问题

$$\min\left\{\theta(x) + \frac{r}{2}\left\|x - \left[x^k + \frac{1}{r}A^T(2\tilde{\lambda}^k - \lambda^k)\right]\right\|^2 \,\middle|\, x \in \mathcal{X}\right\} \tag{2.12}$$

得到. 这是一个容易求解的问题. 生成预测点以后, 给出新迭代点 w^{k+1} 的公式是

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

3 Applications in optimization problems

3.1 Correcting the correlation matrices

在统计学中, 一个对角元均为 1 的对称半正定矩阵称为 (Correlation Matrix) 相关性矩阵. 对给定的对称矩阵 C, 求 F-模下与 C 距离最近的相关性矩阵, 其数学表达式是

$$\min\{\frac{1}{2}\|X-C\|_F^2 \mid \mathrm{diag}(X)=e,\ X\in S_+^n\}, \tag{3.1}$$

其中 e 表示每个分量都为 1 的 n-维向量, S_+^n 表示 $n\times n$ 正半定锥的集合. 问题 (3.1) 是形如 (1.3) 的等式约束凸优化问题, 其中 $\|A^TA\|=1$.

我们用 $z\in\Re^n$ 作为等式约束 $\mathrm{diag}(X)=e$ 的 Lagrange 乘子.

Dual-Primal Customized PPA 求解问题(3.1)

对给定的 (X^k,z^k), 用 (2.11)–(2.12) 产生$(\tilde{X}^k,\tilde{z}^k)$:

1. Producing $\tilde{z}^k$ by

$$\tilde{z}^k = z^k - \frac{1}{s}(\mathrm{diag}(X^k)-e).$$

2. Finding $\tilde{X}^k$ which is the solution of the following minimization problem

$$\min\{\frac{1}{2}\|X-C\|_F^2+\frac{r}{2}\|X-[X^k+\frac{1}{r}\mathrm{diag}(2\tilde{z}^k-z^k)]\|_F^2 | X\in S_+^n\}. \tag{3.2}$$

子问题 (3.2) 求解的具体做法: 化为等价问题

$$\min\{\frac{1}{2}\|X-\frac{1}{1+r}[rX^k+\mathrm{diag}(2\tilde{z}^k-z^k)+C]\|_F^2 | X\in S_+^n\}.$$

记 $A=\frac{1}{1+r}[rX^k+\mathrm{diag}(2\tilde{z}^k-z^k)+C]$, 我们只要考虑如何求解

$$\tilde{X}^k = \text{Argmin}\{\frac{1}{2}\|X-A\|_F^2 \mid X\in S_+^n\}. \tag{3.3}$$

实际上, 将对称矩阵 A 做标准特征值– 特征向量分解

$$A=V\Lambda V^T, \qquad \Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\ldots,\lambda_n). \tag{3.4}$$

其中 V 是正交矩阵. 注意到正交变换下矩阵的 Frobenius-模是不变的, 我们有

$$\|X-A\|_F=\|X-V\Lambda V^T\|_F=\|V^TXV-\Lambda\|_F.$$

要使 $\tilde{X} \succeq 0$ 并且上式右端最小, 应该有

$$V^T \tilde{X} V = \tilde{\Lambda},$$

其中

$$\tilde{\Lambda} = \mathrm{diag}(\tilde{\lambda}_1, \tilde{\lambda}_2, \ldots, \tilde{\lambda}_n), \qquad \tilde{\lambda}_j = \max\{0, \lambda_j\}.$$

最后通过

$$\tilde{X} = V \tilde{\Lambda} V^T$$

得到 $\tilde{X}$. 因此, 每次迭代的主要工作是做 (3.4) 中的特征值 (特征向量) 分解.

数值试验 为生成试验例子, 只要给定对称矩阵 C.

C=rand(n,n); C=(C'+C)-ones(n,n) + eye(n)

这样的矩阵 C 的对角元在 $(0, 2)$ 之间, 非对角元在 $(-1, 1)$ 之间。

Code 4.a. Matlab code for Creating the test examples

```
clear;  close all;
n = 1000;     tol=1e-5;     r=2.0;     s=1.01/r;     gamma=1.5;
rand('state',0);   C=rand(n,n);    C=(C'+C)-ones(n,n) + eye(n);
```


Code 4.1. Matlab code of the classical PPA

```
%%%     Classical PPA for calibrating correlation matrix          %(1)
function PPAC(n,C,r,s,tol)                                        %(2)
X=eye(n);      z=zeros(n,1);    tic;      %%  The initial iterate %(3)
stopc=1;                        k=0;                              %(4)
while (stopc>tol && k<=100)         %% Beginning of an Iteration  %(5)
 if mod(k,20)==0 fprintf(' k=%4d   epsm=%9.3e  \n',k,stopc); end; %(6)
    X0=X;    z0=z;    k=k+1;                                      %(7)
    zt=z0 - (diag(X0)-ones(n,1))/s;              EZ=z0-zt;        %(8)
    A=(X0*r + C + diag(zt*2-z0))/(1+r);                           %(9)
    [V,D]=eig(A);   D=max(0,D);  XT=(V*D)*V';     EX=X0-XT;      %(10)
    ex=max(max(abs(EX)));  ez=max(abs(EZ));   stopc=max(ex,ez);  %(11)
    X=XT;                  z=zt;                                 %(12)
end;                                        % End of an Iteration %(13)
toc;                       TB = max(abs(diag(X-eye(n))));        %(14)
fprintf(' k=%4d  epsm=%9.3e   max|X_jj - 1|=%8.5f \n',k,stopc,TB); %%
```

做 (3.4) 中的特征值 (特征向量) 分解, 在上述程序中的第 (10) 行用 Matlab 中的语句 [V,D]=eig(A) 实现的, 这是一个计算量大概 $9n^3$ 的运算.

将 Classical PPA 改成 Extended PPA, 只要将第 (12) 行改一下。

Code 4.2 Matlab Code of the Extended PPA

```
%%%     Extended PPA for calibrating correlation matrix          %(1)
function PPAE(n,C,r,s,tol,gamma)                                 %(2)
X=eye(n);      z=zeros(n,1);    tic;     %%  The initial iterate  %(3)
stopc=1;                        k=0;                             %(4)
while (stopc>tol && k<=100)          %% Beginning of an Iteration  %(5)
 if mod(k,20)==0 fprintf(' k=%4d   epsm=%9.3e  \n',k,stopc); end; %(6)
    X0=X;    z0=z;    k=k+1;                                     %(7)
    zt=z0 - (diag(X0)-ones(n,1))/s;              EZ=z0-zt;       %(8)
    A=(X0*r + C + diag(zt*2-z0))/(1+r);                          %(9)
    [V,D]=eig(A);   D=max(0,D);  XT=(V*D)*V';     EX=X0-XT;      %(10)
    ex=max(max(abs(EX)));  ez=max(abs(EZ));   stopc=max(ex,ez);  %(11)
    X=X0-EX*gamma;         z=z0-EZ*gamma;                        %(12)
end;                                          % End of an Iteration %(13)
toc;                       TB = max(abs(diag(X-eye(n))));        %(14)
fprintf(' k=%4d  epsm=%9.3e   max|X_jj - 1|=%8.5f \n',k,stopc,TB); %%
```

两个不同的方法, 程序都很简单, 用不了几行. 两个程序不同的地方仅仅是第 12 行有些差别, 取 $\gamma = 1.5$ 的方法效果却有明显的提高。

由于在这个问题中, A 是投影矩阵, $\|A^T A\| = 1$, 我们只需要选 $rs > 1$ 就能确保 $G \succ 0$. 注意到, 为了平衡 primal 和 dual 残量, 我们取 $r = 2$, $s = 1.01/r$.

特征值分解用 Matlab 中的 eig 的迭代次数和计算时间

$n \times n$ Matrix	Classical PPA		Extended PPA	
$n =$	No. It	CPU Sec.	No. It	CPU Sec.
100	31	0.35	23	0.26
200	34	2.00	25	1.48
500	39	17.28	27	13.14
800	41	72.82	29	51.97
1000	47	153.14	31	102.12
2000	62	1344.50	40	881.38

The extended PPA converges faster than the classical PPA.

$$\frac{\text{It. No. of Extended PPA}}{\text{It. No. of Classical PPA}} \approx 65\%.$$

♣ 关于相关系数矩阵校正的程序在附件的 Codes-04 的文件夹“矩阵校正”中. 只要运行 demo.m, 输入 n 就可以了. 其中的 PPAC.m 和 PPAE.m 分别是 Classical PPA 和 Extended PPA 的子程序.

确实, 用这一讲介绍的 PPA 方法求解相关矩阵矩阵校正问题, 每步迭代的主要计算工作量是对一个对称矩阵用 Matlab 中的标准子程序做 [V,D]=eig(A). 如果改用 Kim TOH 写的 mexeig 做 [V,D]=mexeig(A), 计算时间大为节省.

特征值分解使用 mexeig 的迭代次数和计算时间

$n \times n$ Matrix	Classical PPA		Extended PPA	
$n =$	No. It	CPU Sec.	No. It	CPU Sec.
100	30	0.12	23	0.10
200	33	0.54	25	0.40
500	38	7.99	26	6.25
800	38	37.44	28	27.04
1000	45	94.32	30	55.32
2000	62	723.40	38	482.18

3.2 矩阵完整化方面的应用

设 M 是一个 $m \times n$ 矩阵, Π 是矩阵元素的指标集.

$$\Pi = \{(ij) \,|\, i \in \{1, 2, \ldots, m\},\ j \in \{1, 2, \ldots, n\}\}.$$

矩阵完整化问题是 **由部分信息获取全部信息**

- 显然, 没有其他信息, 恢复一个一般的矩阵是不可能的.
- 幸运的是, 在许多情况下, 我们要恢复的矩阵是一个低秩矩阵.

恢复低秩矩阵的一般数学模型是:

$$\boxed{\min\{\text{rank}(X) \,:\, X_{ij} = M_{ij},\ (ij) \in \Pi\}.}$$

然而, 这个问题是NP-hard 的. 根据 **Candés, Recht, Tao** 最近的工作

- E. J. Candés and B. Recht, Exact Matrix Completion via Convex Optimization, 2008.
- E. J. Candés and T. Tao, The Power of Convex Relaxation: Near-Optimial Matrix Completion, 2009,

在适当(实际问题具备的)条件下, 大多数不完整信息的低秩矩阵可以通过求解松弛问题

$$\min\{\|X\|_* \mid X_{ij} = M_{ij},\ (ij) \in \Pi\} \tag{3.5}$$

得到精确恢复. 其中 $\|X\|_*$ 表示矩阵 X 的奇异值的和. 通常称为矩阵 X 的核模— Nuclear Norm. 这类问题的商业应用, 可见 [10].

问题 (3.5) 是形如 (1.3) 的等式约束凸优化问题, 其中 $\|A^T A\| = 1$. 我们将 (3.5) 的等式约束记为 $X_\Pi = M_\Pi$，并用 $Z \in \Re^{m\times n}$ 作为相应的 Lagrange 乘子.

Dual-Primal PPA 求解问题(3.5)

对给定的 (X^k, Z^k), 用 (2.11)–(2.12) 产生$(\tilde{X}^k, \tilde{Z}^k)$:

1. Producing $\tilde{Z}^k$ by
$$\tilde{Z}^k_\Pi = Z^k_\Pi - \frac{1}{s}(X^k_\Pi - M_\Pi).$$
2. Finding $\tilde{X}^k$ which is the solution of the following linear variational inequality
$$\min\ \{\|X\|_* + \frac{r}{2}\|X - [X^k + \frac{1}{r}(2\tilde{Z}^k_\Pi - Z^k_\Pi)]\|_F^2\} \tag{3.6}$$

子问题 (3.6) 求解的具体做法: 只需考虑如何求解

$$\tilde{X}^k = \text{Argmin}\{\frac{1}{r}\|X\|_* + \frac{1}{2}\|X - A\|_F^2\}. \tag{3.7}$$

我们将 A 做 SVD 分解

$$A = U\Lambda V^T,$$

并代入 (3.7), 得到

$$\frac{1}{r}\|\tilde{X}\|_* + \frac{1}{2}\|\tilde{X} - U\Lambda V^T\|_F^2 = \frac{1}{r}\|U^T\tilde{X}V\|_* + \frac{1}{2}\|U^T\tilde{X}V - \Lambda\|_F^2.$$

上面的等式是由于矩阵的奇异值及F-范数在正交变换下不变的原因. 因此, $U^T\tilde{X}V$ 应该是一个非负的对角矩阵. 设 $U^T\tilde{X}V = \tilde{\Lambda}$, 换句话说,

$$\tilde{X} = U\tilde{\Lambda}V^T. \tag{3.8}$$

对给定的非负对角矩阵 Λ,

$$\min\{\frac{1}{r}\|\tilde{\Lambda}\|_* + \frac{1}{2}\|\tilde{\Lambda} - \Lambda\|_F^2\}$$

的解对角矩阵 $\tilde{\Lambda}$, 其对角元通过

$$\tilde{\lambda}_j = \lambda_j - \min(\lambda_j, \frac{1}{r}), \tag{3.9}$$

就能得到. 代入 (3.8) 就得到 (3.7) 的解 $\tilde{X}^k$. 因此, 每次迭代的主要工作量是做一个矩阵的 SVD 分解.

如果以 λ 和 $\tilde{\lambda}$ 分别表示对角矩阵 Λ 和 $\tilde{\Lambda}$ 的对角元生成的向量, 由于 λ 是非负向量, 关系式 (3.9) 也可以写成

$$\tilde{\lambda} = \lambda - P_{B_\infty^{1/r}}[\lambda],$$

其中 $B_\infty^{1/r}$ 是无穷模下半径为 $1/r$ 的“圆”（一个立方体）. 上述运算在文献中常常称作 **Shrinkage**.

结论: 将线性约束的凸优化问题转换成单调变分不等式, 再用本文介绍的 PPA 方法求解, 每次迭代中要求解的子问题, 数值代数中都有确定的成熟的方法求解！

数值试验 数值试验例子取自[1]

一个秩为 ra 的 $n \times n$ 的自由度是 $d_{ra} := ra(2n - ra)$.

生成试验问题：

- 先用高斯同分布 (Gaussian i.i.d) 独立生成两个 $n \times ra$ 的矩阵 M_1 和 M_2, 然后令$M = M_1 M_2^T$, 则 $n \times n$ 矩阵 M 的秩为 ra.
- 随机选定 M 的 m 个元素作为已知元素, 这些元素的下标集为 Π.

计算结果：

- 矩阵完整化问题的难度与比率 m/d_{ra} 和 m/n^2 都有关系.
- 分别用 Classical PPA 和 Extended PPA ($\gamma = 1.5$) 进行计算.
- 正定矩阵 G (see (2.10)) 中的参数 $r,\ s$ 分别取 $rs = 1.01$ 和 $r = 0.005$.
- 停机准则采用相对误差 $\|X_\Pi^k - M_\Pi\|_F / \|M_\Pi\|_F \le 10^{-4}$.

我们用 $\gamma = 1.5$ 的 Extended PPA 求解, 注意到 KKT 条件 Primal 部分((2.9) 的上半部分) 的不满足量是

$$r(\tilde{X}^k - X^k) - (\tilde{Z}^k - Z^k).$$

♣ 矩阵完整化的程序在附件的 Codes-04 的文件夹"矩阵完整化"中. 只要运行 demo.m 就可以了. 要对不同情形试验, 只要在 demo.m 中用 % 做适当选择. 其中的 PPAC.m 和 PPAE.m 分别是 Classical PPA 和 Extended PPA 的子程序.

Code 4.b. Creating the test examples of Matrix Completion

```
%%  Creating the test examples of the matrix Completion problem      %(1)
clear all;  clc                                                       %(2)
maxIt=100;        tol = 1e-4;                                         %(3)
r=0.005;          s=1.01/r;        gamma=1.5;                         %(4)
  n=200;            ra = 10;         oversampling = 5;                %(5)
% n=1000;   ra=100;     oversampling = 3; %%  Iteration No. 31        %(6)
% n=1000;   ra=50;      oversampling = 4; %%  Iteration No. 36        %(7)
% n=1000;   ra=10;      oversampling = 6; %%  Iteration No. 78        %(8)
%% Generating the test problem                                        %(9)
rs = randseed;        randn('state',rs);                             %(10)
M=randn(n,ra)*randn(ra,n);         %% The matrix will be completed %(11)
df =ra*(n*2-ra);                   %% The freedom of the matrix    %(12)
mo=oversampling;                                                     %(13)
m =min(mo*df,round(.99*n*n));       %% No. of the known elements   %(14)
Omega= randsample(n^2,m);           %% Define the subset Omega     %(15)
fprintf('Matrix: n=%4d  Rank(M)=%3d  Oversampling=%2d \n',n,ra,mo);%(16)
```


Code 4.3. Extended PPA for Matrix Completion Problem

```
function PPAE(n,r,s,M,Omega,maxIt,tol,gamma)    %  Ititial Process %%(1)
X=zeros(n);     Y=zeros(n);      YT=zeros(n);                         %(2)
  nM0=norm(M(Omega),'fro');    eps=1;  VioKKT=1;  k=0;   tic;        %(3)
%% Minimum nuclear norm solution by PPA method                       %(4)
while  (eps > tol && k<= maxIt)                                      %(5)
 if mod(k,5)==0                                                      %(6)
 fprintf('It=%3d |X-M|/|M|=%9.2e VioKKT=%9.2e\n',k,eps,VioKKT); end;%(7)
  k=k+1;      X0=X;      Y0=Y;                                       %(8)
  YT(Omega)=Y0(Omega)-(X0(Omega)-M(Omega))/s;            EY=Y-YT;   %(9)
  A = X0 + (YT*2-Y0)/r;                     [U,D,V]=svd(A,0);       %(10)
  D=D-eye(n)/r;     D=max(D,0);     XT=(U*D)*V';         EX=X-XT;  %(11)
  DXM=XT(Omega)-M(Omega);           eps = norm(DXM,'fro')/nM0;     %(12)
  VioKKT = max( max(max(abs(EX)))*r, max(max(abs(EY))) );           %(13)
   if (eps <= tol)   gamma=1;   end;                                %(14)
  X = X0 - EX*gamma;                                                %(15)
  Y(Omega) = Y0(Omega) - EY(Omega)*gamma;                           %(16)
end;                                                                %(17)
 fprintf('It=%3d  |X-M|/|M|=%9.2e  Vi0KKT=%9.2e \n',k,eps,VioKKT); %(18)
   RelEr=norm((X-M),'fro')/norm(M,'fro');   toc;                    %(19)
 fprintf(' Relative error = %9.2e  Rank(X)=%3d \n',RelEr,rank(X)); %(20)
 fprintf(' Violation of KKT Condition = %9.2e  \n',VioKKT);         %(21)
```

矩阵完整化问题：用 Matlab 中标准 SVD 求解结果

Unknown $n \times n$ matrix M				Computational Results			
n	rank(ra)	m/d_{ra}	m/n^2	#iters	times(Sec.)	relative error	KKT-Violation
1000	10	6	0.12	76	841.59	9.38E-5	9.31E-6
1000	50	4	0.39	37	406.24	1.21E-4	2.11E-5
1000	100	3	0.58	31	362.58	1.50E-4	2.88E-5

♣ 用 Matlab 中的 SVD, 做一次 SVD 的花费很大, 总耗时与迭代次数成比例.

♣ 用 PROPACK [8] 中的 SVD, 快许多, 总耗时主要与问题性质有关. 我们主要对迭代次数感兴趣, 只报道用 Matlab 做 SVD 的结果, 也附上所需要的子程序.

对矩阵完整化问题, (2.5) 和 (2.9) 的下半部分误差用$\|X_\Pi^k - M_\Pi\|_F / \|M_\Pi\|_F$ 控制. 由于 (2.5) 和 (2.9) 的上半部分是

$$\theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T\{(-A^T\tilde{\lambda}^k) + [r(\tilde{x}^k - x^k) \pm A^T(\tilde{\lambda}^k - \lambda^k)]\},$$

因此上半部分误差项是 $[r(\tilde{x}^k - x^k) \mp A^T(\tilde{\lambda}^k - \lambda^k)]$. 对矩阵完整化问题, A是投影矩阵, 其模为 1, 所以我们在计算结果中也列出

$$\text{KKT-Violation} := \max\{r \max_{ij} |X_{ij}^k - \tilde{X}_{ij}^k|,\ \max_{ij} |Z_{ij}^k - \tilde{Z}_{ij}^k|\}.$$

♣ 论文 [1] 是最早发表在 SIAM J. Optimization 上的求解矩阵完整化问题的文章. 对以上试验的三个例子, [1] 中的方法达到同样精度要求的迭代次数分别是 117, 114 和 129 (See the first three examples in Table 5.1 of [1], pp. 1974), 每次迭代的主要工作量也是做一次 SVD 分解. 由于采用了不完全分解技术, [1] 中节省了 SVD 分解的时间. 我们调用的是 Matlab 中标准SVD, 虽然花费了更少的迭代次数, 但没有节省总的运行时间.

3.3 压缩传感(Compressed Sensing)问题

- 压缩传感问题是利用信号的稀疏性质, 只对其进行相对较少的测量, 然后能以较小的误差或者精确地恢复原始信号。
- CS 问题: 求满足 $Ax = b$ 的非零元个数最小的解

$$\min\{\|x\|_0 \mid Ax = b\}.$$

其中$\|x\|_0$ 表示向量 x 的非零元的个数, 它不是常规意义下的模。是一个很难求解的组合优化问题.

$$\boxed{\|x\|_1 \le \|x\|_0,\ \forall\, x \in \Omega = \{x \in R^n \mid |x_j| \le 1,\ j = 1, \ldots, n\}.}$$

Candés 和 Tao 证明了在一定条件下, 问题

$$\min\{\|x\|_0\ :\ Ax = b\} \qquad \text{与问题} \qquad \min\{\|x\|_1\ :\ Ax = b\}$$

的解在概率意义下是相同的.

极小化问题

$$\min_x \{\, \mu\|x\|_1 \mid Ax = b \,\} \qquad \text{(BP)} \tag{3.10}$$

是形如 (1.3) 的等式约束凸优化问题.

PPA 算法求解问题(3.10)

对给定的 (x^k, λ^k), 用 (2.11)–(2.12) 产生$(\tilde{x}^k, \tilde{\lambda}^k)$:

1. Producing $\tilde{\lambda}^k$ by
$$\tilde{\lambda}^k = \lambda^k - \frac{1}{s}(Ax^k - b).$$
2. Finding $\tilde{x}^k$ which is the solution of the following minimization problem
$$\min\{\, \mu\|x\|_1 + \frac{r}{2}\|x - [x^k + \frac{1}{r}A^T(2\tilde{\lambda}^k - \lambda^k)]\|^2\}. \tag{3.11}$$

子问题 (3.11) 求解的具体做法: 我们只要考虑如何求解

$$\min_{x\in\Re^n} \{\tau\|x\|_1 + \frac{1}{2}\|x - a\|^2\}. \tag{3.12}$$

设 $x^* \in \Re^n$ 是 (3.12) 的解,

- 如果 $x_j^* > 0$, 对 $\tau x + \frac{1}{2}\|x - a\|^2$ 求导, 有
$$\tau + x_j^* - a_j = 0, \quad \text{即} \quad x_j^* = a_j - \tau, \qquad (\text{这隐含了} \quad a_j > \tau).$$
- 如果 $x_j^* < 0$, 对 $-\tau x + \frac{1}{2}\|x - a\|^2$ 求导, 有
$$-\tau + x_j^* - a_j = 0, \quad \text{即} \quad x_j^* = a_j + \tau, \qquad (\text{这隐含了} \quad a_j < -\tau).$$
- 为什么 $x_j^* = 0$, 是因为 $-\tau \le a_j \le \tau$.

综上所述, 问题 (3.12) 的解由

$$x_j^* = \begin{cases} a_j - \tau, & \text{if} \quad a_j > \tau \\ a_j + \tau, & \text{if} \quad a_j < -\tau \\ 0, & \text{if} \quad -\tau \le a_j \le \tau. \end{cases}$$

给出. 这也可以写成 Shrinkage 的形式

$$x^* = a - P_{B_\infty^\tau}[a], \qquad \text{where} \qquad B_\infty^\tau = \{\xi \in \Re^n \mid -\tau e \le \xi \le \tau e\}.$$

实际 CS 问题的计算, 要加进一些 Continuation 技术, 限于篇幅, 不详细介绍.

3.4 Min-Max 问题上的应用

第一讲 §4.2 中提到用全变差极小处理图像去模糊 [2], 经离散化以后, 问题的数学模型是的 min-max 问题,

$$\min_{x\in\mathcal{X}} \max_{y\in\mathcal{Y}} \ \Phi(x,y) := \theta_1(x) + y^T Ax - \theta_2(y). \tag{3.13}$$

它可以转换成等价的变分不等式: 求 $(x^*, y^*) \in \mathcal{X}\times\mathcal{Y}$, 使得

$$\begin{pmatrix} x - x^* \\ y - y^* \end{pmatrix}^T \begin{pmatrix} f(x^*) + A^T y^* \\ g(y^*) - Ax^* \end{pmatrix} \ge 0, \quad \forall (x,y) \in \mathcal{X}\times\mathcal{Y}, \tag{3.14}$$

其中 $f(x) \in \partial\theta_1(x)$, $g(y) \in \partial\theta_2(y)$. 文献 [6] 中提到用这里的 PPA 方法去求解. 对给定的 (x^k, y^k), 求 $(\tilde{x}^k, \tilde{y}^k) \in \mathcal{X}\times\mathcal{Y}$, 对一切 $(x,y) \in \mathcal{X}\times\mathcal{Y}$, 都有

$$\begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \end{pmatrix}^T \left\{ \begin{pmatrix} f(\tilde{x}^k) + A^T\tilde{y}^k \\ g(\tilde{y}^k) - A\tilde{x}^k \end{pmatrix} + \begin{pmatrix} r(\tilde{x}^k - x^k) + A^T(\tilde{y}^k - y^k) \\ +A(\tilde{x}^k - x^k) + s(\tilde{y}^k - y^k) \end{pmatrix} \right\} \ge 0. \tag{3.15}$$

注意到 (3.15) 的下半部分中只含有未知的 $\tilde{y}^k$. 它可以通过求解 y 的子问题:

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + y^T Ax^k + \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}. \tag{3.16a}$$

得到. 然后, 为求得 (3.15) 上半部分得 $\tilde{x}^k$, 我们只要求解一个关于 x 的问题:

$$\tilde{x}^k = \text{Argmin}\big\{\theta_1(x) + \frac{r}{2}\big\|x - \big[x^k - \frac{1}{r}A^T(2\tilde{y}^k - y^k)\big]\big\|^2 \,\big|\, x \in \mathcal{X}\big\}. \tag{3.16b}$$

当然, PPA 预测点 $(\tilde{x}^k, \tilde{y}^k) \in \mathcal{X}\times\mathcal{Y}$ 也可以通过先 x, 后 y 的顺序完成.

对给定的 (x^k, y^k), 通过求解 x 的子问题

$$\tilde{x}^k = \text{Argmin}\{\theta_1(x) + x^T A^T y^k + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\}. \tag{3.17a}$$

得到 $\tilde{x}^k$, 然后求 $\tilde{y}^k$ 通过一个关于 y 的子问题:

$$\tilde{y}^k = \text{Argmin}\big\{\theta_2(y) + \frac{s}{2}\big\|y - \big[y^k + \frac{1}{s}A(2\tilde{x}^k - x^k)\big]\big\|^2 \,\big|\, y \in \mathcal{Y}\big\}. \tag{3.17b}$$

由 (3.17) 生成的 $(\tilde{x}^k, \tilde{y}^k) \in \mathcal{X} \times \mathcal{Y}$, 对一切 求 $(x, y) \in \mathcal{X} \times \mathcal{Y}$, 都有

$$\begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \end{pmatrix}^T \left\{ \begin{pmatrix} f(\tilde{x}^k) + A^T \tilde{y}^k \\ g(\tilde{y}^k) - A\tilde{x}^k \end{pmatrix} + \begin{pmatrix} r(\tilde{x}^k - x^k) - A^T(\tilde{y}^k - y^k) \\ -A(\tilde{x}^k - x^k) + s(\tilde{y}^k - y^k) \end{pmatrix} \right\} \geq 0. \tag{3.18}$$

无论是 (3.15) 还是 (3.18), 都可以写成

$$\tilde{u}^k \in \Omega, \qquad (u - \tilde{u}^k)^T \{F(\tilde{u}^k) + G(\tilde{u}^k - u^k)\} \geq 0, \quad \forall\, u \in \Omega, \tag{3.19}$$

的形式, 所不同的只是

$$G = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix} \qquad \text{和} \qquad G = \begin{pmatrix} rI_n & -A^T \\ -A & sI_m \end{pmatrix}.$$

它们都是当 $rs > \|A^T A\|$ 时候正定. 我们建议新的迭代点u^{k+1} 用迭代式

$$u^{k+1} = u^k - \gamma(u^k - \tilde{u}^k), \quad \gamma \in [1, 2),$$

生成, 一般取 $\gamma = 1.5$. 对一些工程技术上的问题, 采取 $\gamma > 1$ 的迭代满足停机准则以后, 最后加一次 $\gamma = 1$ 的迭代. 这样既提高了速度, 又满足实际问题对诸如 Low-Rank 这样的要求. 这类方法在图像处理中的应用可见论文 [6, 7].

定制的 PPA 算法具有 $O(1/k)$ 的收敛速率, 证明可以在论文 [4] 中找到.

References

[1] J. F. Cai, E. J. Candès and Z. W. Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optim. **20**, 1956-1982, 2010.

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

[3] B. S. He, X. L. Fu, Z.K. Jiang, Proximal point algorithm using a linear proximal term, JOTA **141**, 209-239, 2009.

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

[5] B. S. He and M.-H. Xu, A general framework of contraction methods for monotone variational inequalities, Pacific J. Optimization **4**, 195-212, 2008.

[6] 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**, pp. 119-149, 2012.

[7] 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**, pp. 559-572, 2013.

[8] R. M. Larsen, Propack – software for large and sparse svd calculation, http://sun.stanford.edu/rmunk/PROPACK/, 2005.

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

[10] Netflix, Netflix Prize, Available at http://www.netflixprize.com/. 2006.

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

[12] M. Zhu and T. F. Chan, An efficient primal-dual hybrid gradient algorithm for total variation image restoration, CAM Reports 08-34, UCLA, 2008.

凸优化和单调变分不等式的收缩算法

第七讲: 线性约束凸优化问题
基于松弛 PPA 的收缩算法

Relaxed PPA-based Contraction Methods for
Linearly Constrained Convex Optimization

南京大学数学系　何炳生
hebma@nju.edu.cn

The context of this lecture is based on the publications [6]

线性约束的凸优化问题

这一讲要求解的问题是

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

其中 $\theta(x)$ 是凸函数, $A \in \Re^{m\times n}$, $b \in \Re^m$, $\mathcal{X}$ 是 $\Re^n$ 中的闭凸集

对任意给定的 $r > 0$ 和 $a \in \Re^n$, 通篇我们假设子问题

$$\min\left\{\theta(x) + \tfrac{r}{2}\|x-a\|^2 \mid x \in \mathcal{X}\right\}$$

的求解是简单的.

目的 将线性约束的凸优化问题转换成单调变分不等式, 再用自调比方法, 生成类似于 PPA 的正则项. 我们并不要求正则项系数矩阵 Q 对称正定, 只要在迭代中都有 $(u^k - \tilde{u}^k)^T Q(u^k - \tilde{u}^k) \ge \tau\|u^k - \tilde{u}^k\|^2$. 基于松弛 PPA 的收缩算法是一类预测－校正方法, 产生预测点都用松弛 PPA 方法.

1 线性约束凸优化问题的 PPA 算法

1.1 线性约束凸优化问题与单调变分不等式

线性约束的凸优化问题

$$\min\{\theta(x) \mid Ax = b\,(or\ Ax \geq b)\ x \in \mathcal{X}\} \tag{1.1}$$

的 Lagrange 函数是定义在 $\mathcal{X} \times \Lambda$ 上的

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

其中

$$\Lambda = \begin{cases} \Re^m, & \text{for the equality constraints} \quad Ax = b, \\ \Re^m_+, & \text{for the inequality constraints} \quad Ax \geq b. \end{cases}$$

设 (x^*, λ^*) 是 Lagrange 函数的一个鞍点, 便有

$$L_{\lambda\in\Lambda}(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L_{x\in\mathcal{X}}(x, \lambda^*).$$

求 Lagrange 函数的一个鞍点等价于求 (x^*, λ^*) 使其满足:

$$\begin{cases} x^* \in \mathcal{X}, & \theta(x) - \theta(x^*) + (x - x^*)^T(-A^T\lambda^*) \geq 0, \quad \forall\, x \in \mathcal{X}, \\ \lambda^* \in \Lambda, & (\lambda - \lambda^*)^T(Ax^* - b) \geq 0, \quad \forall\, \lambda \in \Lambda. \end{cases} \tag{1.2}$$

换句话说, 求 Lagrange 函数的鞍点等价于求解混合变分不等式

$$\text{VI}(\Omega, F, \theta) \quad u^* \in \Omega, \quad \theta(x) - \theta(x^*) + (u - u^*)^T F(u^*) \geq 0, \quad \forall\, u \in \Omega, \tag{1.3a}$$

的解, 其中

$$u = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(u) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix} \tag{1.3b}$$

和

$$\Omega = \mathcal{X} \times \Lambda. \tag{1.3c}$$

注意到 $\theta(x)$ 是凸函数, 此外 (1.3b) 中的算子 F 是仿射单调的.

我们通过求解相应的混合单调变分不等式来求解线性约束的凸优化问题.

1.2 H-模下的 PPA 算法和松弛 PPA 的收缩算法

邻近点算法 (PPA) 是求解混合单调变分不等式的一个经典方法. 用邻近点算法求解问题 (1.3), 对给定的 $u^k \in \Omega$ 和 $r > 0$, k-次迭代是求 $\tilde{u}^k \in \Omega$, 使得

$$\theta(x) - \theta(\tilde{x}^k) + (u - \tilde{u}^k)^T \big\{F(\tilde{u}^k) + r(\tilde{u}^k - u^k)\big\} \ge 0, \quad \forall u \in \Omega. \tag{1.4}$$

我们取 $\gamma \in [1, 2)$, 用公式 $u^{k+1} = u^k - \gamma(u^k - \tilde{u}^k)$ 产生下一个迭代点. 序列 $\{u^k\}$ 满足关系式

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \gamma(2-\gamma)\|u^k - \tilde{u}^k\|^2.$$

对解集 Ω^*, 算法产生的序列 $\{u^k\}$ 是 Fejér 单调的. 这些在前一讲已经有介绍.

直接求解 PPA 子问题 (1.4) 一般来说是办不到的. 为求解与凸优化问题 (1.1) 等价的变分不等式 (1.3), 上一讲提出的 Customized PPA 方法的子问题是

$$\tilde{u}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) + (u - \tilde{u}^k)^T \{F(\tilde{u}^k) + H(\tilde{u}^k - u^k)\} \ge 0, \quad \forall u \in \Omega, \tag{1.5}$$

其中

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

注意到与讨论的优化问题等价的变分不等式 (1.2) 的具体结构 (见 (1.3b) 式), 子变分不等式 (1.5) 便是求 $(\tilde{x}^k, \tilde{\lambda}^k) \in \Omega$, 使得对一切 $(x, \lambda) \in \Omega$, 都有

$$\begin{aligned} \theta(x) - \theta(\tilde{x}^k) + \begin{pmatrix} x - \tilde{x}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \Bigg\{ & \begin{pmatrix} -A^T\tilde{\lambda}^k \\ A\tilde{x}^k - b \end{pmatrix} \\ & + \begin{pmatrix} r(\tilde{x}^k - x^k) - A^T(\tilde{\lambda}^k - \lambda^k) \\ -A(\tilde{x}^k - x^k) + s(\tilde{\lambda}^k - \lambda^k) \end{pmatrix} \Bigg\} \ge 0, \quad \forall (x, \lambda) \in \Omega. \end{aligned} \tag{1.6}$$

上式下半部分中未知的 $\tilde{x}^k$ 可以消掉, 问题 (1.6) 的一半松弛成

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T \{(Ax^k - b) + s(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \quad \forall \lambda \in \Lambda.$$

这个 $\tilde{\lambda}^k$ 可以通过 $\tilde{\lambda}^k = P_\Lambda[\lambda^k - \frac{1}{s}(Ax^k - b)]$ 直接给出. 有了 $\tilde{\lambda}^k$, 通过求解极小化问题

$$\min \Big\{\theta(x) + \frac{r}{2}\Big\|x - \Big[x^k + \frac{1}{r}A^T(2\tilde{\lambda}^k - \lambda^k)\Big]\Big\|^2 \,\Big|\, x \in \mathcal{X}\Big\} \tag{1.7}$$

得到的 $\tilde{x}^k$ 满足 (1.6) 的上半部分. 因为 (1.7) 中的 $\tilde{\lambda}^k$ 也是已知的, 极小化问题 (1.7) 的类型根据假设是容易求解的. 用公式 $u^{k+1} = u^k - \gamma(u^k - \tilde{u}^k)$ 产生

的迭代序列 $\{u^k\}$ 满足

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

说到底, 这个 H-模下的 PPA 算法, 也是对经典的 PPA 松弛后得到的收缩算法.

上述 H-模下 PPA 算法子问题 (1.6) 中的参数 r, s 要满足 $rs > \|A^TA\|$, 矩阵 H 才是正定的, 这才保证算法收敛. 有时估计 $\|A^TA\|$ 并不容易, 过大的估计相当于正则项太大, 会影响总体收敛速度. 这一讲试图解决这些问题.

1.3 松弛条件下 PPA 收缩算法

Idea of the Relaxed PPA 为求解变分不等式 (1.3), 在第 k-次迭代中, 经典的 PPA 是要求 $(\tilde{x}^k, \tilde{\lambda}^k) \in \Omega$ 使得

$$\theta(x)-\theta(\tilde{x}^k)+\begin{pmatrix} x-\tilde{x}^k \\ \lambda-\tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{\lambda}^k \\ A\tilde{x}^k-b \end{pmatrix} + \begin{pmatrix} r(\tilde{x}^k-x^k) \\ s(\tilde{\lambda}^k-\lambda^k) \end{pmatrix} \right\} \ge 0, \ (x,\lambda)\in\Omega. \tag{1.8}$$

因为上面的子变分不等式的上下两部分都含有未知的 $\tilde{x}^k$ 和 $\tilde{\lambda}^k$, 直接求解一

般办不到. 如果将上半部分中的 $\tilde{\lambda}^k$ 松弛成 λ^k, 则变成求 $(\tilde{x}^k, \tilde{\lambda}^k) \in \Omega$ 使得

$$\theta(x)-\theta(\tilde{x}^k)+\begin{pmatrix} x-\tilde{x}^k \\ \lambda-\tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\lambda^k \\ A\tilde{x}^k-b \end{pmatrix} + \begin{pmatrix} r(\tilde{x}^k-x^k) \\ s(\tilde{\lambda}^k-\lambda^k) \end{pmatrix} \right\} \ge 0, \ (x,\lambda)\in\Omega. \tag{1.9}$$

我们把子问题 (1.9) 称为松弛的 PPA (Relaxed PPA). 松弛以后的问题 (1.9) 是容易求解的. 首先 (1.9) 的上半 (Primal) 部分可以通过求解极小化问题

$$\min\left\{\theta(x)+\frac{r}{2}\|x-[x^k+\frac{1}{r}A^T\lambda^k]\|^2 \,\middle|\, x\in\mathcal{X}\right\}$$

得到. 根据假设条件, 这是一个容易求解的问题. 有了 $\tilde{x}^k$, (1.9) 的下半 (Dual) 部分中要求的只剩下 $\tilde{\lambda}^k$, 它的数学形式是

$$\tilde{\lambda}^k\in\Lambda, \quad (\lambda-\tilde{\lambda}^k)^T\{(A\tilde{x}^k-b)+s(\tilde{\lambda}^k-\lambda^k)\}\ge 0, \quad \forall\,\lambda\in\Lambda.$$

根据变分不等式和投影的关系, $\tilde{\lambda}^k$ 可以通过

$$\tilde{\lambda}^k = P_\Lambda[\lambda^k-\frac{1}{s}(A\tilde{x}^k-b)]$$

直接给出. 我们把 (1.9) 写成:

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

它的紧凑形式是

$$\tilde{u}^k \in \Omega, \ \theta(x) - \theta(\tilde{x}^k) + (u - \tilde{u}^k)^T \{F(\tilde{u}^k) + Q(\tilde{u}^k - u^k)\} \geq 0, \ \forall u \in \Omega, \quad (1.10)$$

其中矩阵

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

是非对称的. 换句话说, 这样的检验点 (或预测点) $\tilde{u}^k = (\tilde{x}^k, \tilde{\lambda}^k)$ 是通过对经典的 PPA 子问题松弛求得的. 由于 (1.10) 中除了 $F(\tilde{u}^k)$ 外就是 $Q(\tilde{u}^k - u^k)$, 也可以把由此建立的收缩算法称为基于线性临近点项的 PPA 收缩算法. 收敛证明可见参考文献[6].

我们把由 (1.10) 中产生的点 $\tilde{u}^k$ 称为预测点. 将 (1.10) 中任意的 u 设为 u^*, 就能得到

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

进而得到

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

只要 (1.10) 中的矩阵 Q 满足

$$Q^T + Q = \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix} \succ 0,$$

就可以构造收缩算法. 由上式可知, 当 $rs > \frac{1}{4}\|A^TA\|$ 时, $Q^T + Q$ 就正定. 这个条件显然比前一讲 H-模下的 PPA 算法对参数 r, s 要求 $rs > \|A^TA\|$ 来得宽松. 这是我们这一讲继续 Relaxed PPA 研究的理由.

在统一框架下的考虑 对给定的 u^k, 我们生成 $\tilde{u}^k \in \Omega$ 使得 (1.10) 成立, 其中 Q 是一个矩阵. 我们并不要求 Q 对称, 只要求存在一个 $\tau > 0$, 使得

$$(u^k - \tilde{u}^k)^T Q(u^k - \tilde{u}^k) \geq \tau\|u^k - \tilde{u}^k\|^2, \ \ \forall\, k \geq 0.$$

上述条件在 $Q^T + Q$ 正定时一定成立.

2 Primal-Dual 松弛 PPA 收缩算法

用松弛 PPA 按先 $\tilde{x}^k$ (primal) 后 $\tilde{\lambda}^k$ (dual) 的顺序生成预测点 $\tilde{u}^k = (\tilde{x}^k, \tilde{\lambda}^k)$, 然后构造的收缩算法, 称为基于Primal-Dual 松弛 PPA 的收缩算法 (Primal-dual relaxed PPA based contraction method).

2.1 Primal-Dual 生成预测点

Primal–Dual Method 生成预测点 对给定的 (x^k, λ^k) 和 $r > 0$, 通过求解

$$\min \left\{ \theta(x) + \frac{r}{2} \left\| x - \left[x^k + \frac{1}{r} A^T \lambda^k \right] \right\|^2 \,\middle|\, x \in \mathcal{X} \right\} \tag{2.1a}$$

得到 Primal 预测点 $\tilde{x}^k$. 再选取适当的 $s > 0$ 并用

$$\tilde{\lambda}^k = P_\Lambda [\lambda^k - \frac{1}{s}(A\tilde{x}^k - b)] \tag{2.1b}$$

生成 Dual 预测点 $\tilde{\lambda}^k$ (如何选取 $s > 0$ 放在后面讨论).

注意到用 (2.1a) 生成的 Primal 预测点 $\tilde{x}^k$ 满足

$$\tilde{x}^k \in \mathcal{X},\ \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T \{ r(\tilde{x}^k - x^k) - A^T \lambda^k \} \ge 0,\ \forall x \in \mathcal{X}. \tag{2.2}$$

将 (2.1b) 改写成 $\tilde{\lambda}^k = P_\Lambda \{ \tilde{\lambda}^k - [(\tilde{\lambda}^k - \lambda^k) + \frac{1}{s}(A\tilde{x}^k - b)] \}$, 则有

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T \{ (A\tilde{x}^k - b) + s(\tilde{\lambda}^k - \lambda^k) \} \ge 0, \quad \forall \lambda \in \Lambda. \tag{2.3}$$

把 (2.2) 和 (2.3) 写在一起, 有

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

它的紧凑形式是

$$\tilde{u}^k \in \Omega,\ \theta(x) - \theta(\tilde{x}^k) + (u - \tilde{u}^k)^T \{ F(\tilde{u}^k) + Q(\tilde{u}^k - u^k) \} \ge 0,\ \forall u \in \Omega, \tag{2.4}$$

其中矩阵 Q (对应于 Primal-Dual 方法生成预测点的 Q, 有时也记成 Q_{PD})

$$Q_{PD} = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix}, \tag{2.5}$$

是非对称的.

如果 (2.4) 中的矩阵 Q 是对称正定的, 这个生成预测点 $\tilde{u}^k$ 的方法就是 PPA 算法. 由于这里的 Q 是非对称矩阵, 按照 [6] 中的说法, 这里的 Relaxed PPA 是带线性临近点项的 PPA.

以 u^* 替代 (2.4) 中的 $u \in \Omega$, 可得到

$$(\tilde{u}^k - u^*)^T Q(u^k - \tilde{u}^k) \ge \theta(\tilde{x}^k) - \theta(x^*) + (\tilde{u}^k - u^*)^T F(\tilde{u}^k). \tag{2.6}$$

因为 $\tilde{u}^k \in \Omega$ 并且 u^* 是 VI(Ω, F) 的解, 根据混合变分不等式 (1.3) 的定义, 有

$$\theta(\tilde{x}^k) - \theta(x^*) + (\tilde{u}^k - u^*)^T F(u^*) \ge 0.$$

由于 F 是单调算子, 因此有

$$(\tilde{u}^k - u^*)^T F(\tilde{u}^k) \ge (\tilde{u}^k - u^*)^T F(u^*).$$

根据上面二个关系式得到 (2.6) 的右端非负. 我们定义

$$\varphi(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k)^T Q(u^k - \tilde{u}^k). \tag{2.7}$$

由 (2.6) 的右端非负得到下面的关键不等式

$$(u^k - u^*)^T Q(u^k - \tilde{u}^k) \ge \varphi(u^k, \tilde{u}^k), \quad \forall\, u^* \in \Omega^*. \tag{2.8}$$

这一节中的收缩算法, 都是基于向量 $(u^k - \tilde{u}^k)$ 的算法. 然而, 我们并不直接用 $(u^k - \tilde{u}^k)$, 而是用

$$d(u^k, \tilde{u}^k) = M(u^k - \tilde{u}^k) \tag{2.9}$$

作为寻查方向, 其中

$$M = D^{-1}Q, \tag{2.10}$$

矩阵 Q 由 (2.5) 给出, 矩阵

$$D = \begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix}. \tag{2.11}$$

根据这些定义, 我们有

$$M = \begin{pmatrix} I_n & \frac{1}{r}A^T \\ 0 & I_m \end{pmatrix},$$

是对角部分为单位阵的上三角分块矩阵. 关系式 (2.7) 和 (2.8) 可以写成等价的

$$\varphi(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k)^T DM(u^k - \tilde{u}^k). \tag{2.12}$$

和

$$\langle D(u^k - u^*), M(u^k - \tilde{u}^k)\rangle \ge \varphi(u^k, \tilde{u}^k), \quad \forall\, u^* \in \Omega^*. \tag{2.13}$$

2.2 初等的收缩算法

The Primary Contraction Methods (初等的收缩算法) 是指用确定的方向取单位步长的收缩算法.

The Primary Contraction Methods 对给定的 u^k 和由 (2.1) 生成的 $\tilde{u}^k$, 我们用

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

生成新的迭代点. 在用 (2.1) 生成的 $\tilde{u}^k$ 的过程中, 采用 Armijo 法则, 可以做到

对给定的 $r > 0$ 和 $\nu \in (0, 1)$, 选取 s 使得

$$\frac{1}{r}\|A^T(\lambda^k - \tilde{\lambda}^k)\|^2 \le \nu\big(r\|x^k - \tilde{x}^k\|^2 + s\|\lambda^k - \tilde{\lambda}^k\|^2\big). \tag{2.15}$$

特别当 s 取得使 $sr\nu \ge \|A^T A\|$ 时, 条件 (2.15) 自然成立.

我们考虑 D-模下的收缩算法. 根据迭代公式 (2.14) 有

$$\begin{aligned} &\|u^k - u^*\|_D^2 - \|u^{k+1} - u^*\|_D^2 \\ &\quad = \quad \|u^k - u^*\|_D^2 - \|(u^k - u^*) - M(u^k - \tilde{u}^k)\|_D^2 \\ &\quad = \quad 2(u^k - u^*)^T DM(u^k - \tilde{u}^k) - \|M(u^k - \tilde{u}^k)\|_D^2. \end{aligned}$$

对上式右端的 $(u^k - u^*)^T DM(u^k - \tilde{u}^k)$ 使用 (2.12)–(2.13), 我们有

$$\begin{aligned} &\|u^k - u^*\|_D^2 - \|u^{k+1} - u^*\|_D^2 \\ &\quad \ge \quad 2(u^k - \tilde{u}^k)^T DM(u^k - \tilde{u}^k) - \|M(u^k - \tilde{u}^k)\|_D^2. \end{aligned} \tag{2.16}$$

注意到 (2.16) 的右端等于

$$\begin{aligned} & 2(u^k-\tilde{u}^k)^T DM(u^k-\tilde{u}^k) - \|M(u^k-\tilde{u}^k)\|_D^2 \\ = \quad & (u^k-\tilde{u}^k)^T\big[M^TD+DM-M^TDM\big](u^k-\tilde{u}^k). \end{aligned} \tag{2.17}$$

利用矩阵关系 $DM=Q$, 有

$$M^TD+DM-M^TDM = Q^T+Q-Q^TM.$$

经过简单计算就有

$$\begin{aligned} & M^TD+DM-M^TDM = Q^T+Q-Q^TM. \\ = \quad & \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n & 0 \\ A & sI_m \end{pmatrix} \begin{pmatrix} I_n & \frac{1}{r}A^T \\ 0 & I_m \end{pmatrix} \\ = \quad & \begin{pmatrix} rI_n & 0 \\ 0 & sI_m - \frac{1}{r}AA^T \end{pmatrix}. \end{aligned}$$

因此,

$$\begin{aligned} & 2(u^k-\tilde{u}^k)^T DM(u^k-\tilde{u}^k) - \|M(u^k-\tilde{u}^k)\|_D^2 \\ = \quad & \|u^k-\tilde{u}^k\|_D^2 - \frac{1}{r}\|A^T(\lambda^k-\tilde{\lambda}^k)\|^2. \end{aligned} \tag{2.18}$$

在条件 (2.15) 满足的情况下, 从 (2.18) 式得到

$$\begin{aligned} & 2(u^k-\tilde{u}^k)^T DM(u^k-\tilde{u}^k) - \|M(u^k-\tilde{u}^k)\|_D^2 \\ \geq \quad & (1-\nu)\big(r\|x^k-\tilde{x}^k\|^2 + s\|\lambda^k-\tilde{\lambda}^k\|^2\big). \end{aligned} \tag{2.19}$$

以 (2.19) 代入 (2.16), 就有下面的定理:

Theorem 2.1 在基于 *Primal-Dual* 松弛 *PPA* 的收缩算法中, 如果产生预测点时条件 (2.15) 成立, 则由初等收缩算法 (2.14) 生成的序列 $\{u^k=(x^k,\lambda^k)\}$ 满足

$$\|u^{k+1}-u^*\|_D^2 \leq \|u^k-u^*\|_D^2 - (1-\nu)\|u^k-\tilde{u}^k\|_D^2. \tag{2.20}$$

定理 2.1 是保证初等收缩算法收敛的关键不等式.

2.3 一般的收缩算法

一般收缩算法同样用给定的方向 $M(u^k - \tilde{u}^k)$, 但通过计算步长确定下一个迭代点. 在 D-模的意义下, 使新的迭代点靠解集尽可能近一些. 在用 (2.1) 生成的 $\tilde{u}^k$ 的过程中, 采用 Armijo 法则, 可以做到

对给定的 $r > 0$, 选取 s 使得

$$\begin{aligned} &\frac{1}{s}\|A(x^k - \tilde{x}^k)\|^2 + \frac{1}{r}\|A^T(\lambda^k - \tilde{\lambda}^k)\|^2 \\ &\quad \le \quad 2\big(r\|x^k - \tilde{x}^k\|^2 + s\|\lambda^k - \tilde{\lambda}^k\|^2\big). \end{aligned} \tag{2.21}$$

特别当 s 取得使 $sr \ge \frac{1}{2}\|A^TA\|$ 时, $\|A^TA\| \le 2rs$,

$$\begin{aligned} &\frac{1}{s}\|A(x^k - \tilde{x}^k)\|^2 + \frac{1}{r}\|A^T(\lambda^k - \tilde{\lambda}^k)\|^2 \\ &\quad \le \quad \frac{1}{s}\|A^TA\| \cdot \|x^k - \tilde{x}^k\|^2 + \frac{1}{r}\|AA^T\| \cdot \|\lambda^k - \tilde{\lambda}^k\|^2, \end{aligned}$$

条件 (2.21) 自然成立.

利用 D 和 M 的表达式以及 Cauchy-Schwarz 不等式, 我们有

$$\begin{aligned} \varphi(u^k, \tilde{u}^k) &= (u^k - \tilde{u}^k)^T DM(u^k - \tilde{u}^k) \\ &= \|u^k - \tilde{u}^k\|_D^2 + (x^k - \tilde{x}^k)^T A^T(\lambda^k - \tilde{\lambda}^k) \\ &\ge r\|x^k - \tilde{x}^k\|^2 - \frac{1}{2}\|x^k - \tilde{x}^k\| \cdot \|A^T(\lambda^k - \tilde{\lambda}^k)\| \\ &\quad + s\|\lambda^k - \tilde{\lambda}^k\|^2 - \frac{1}{2}\|\lambda^k - \tilde{\lambda}^k\| \cdot \|A(x^k - \tilde{x}^k)\| \\ &\ge r\|x^k - \tilde{x}^k\|^2 - \frac{1}{4}\Big\{r\|x^k - \tilde{x}^k\|^2 + \frac{1}{r}\|A^T(\lambda^k - \tilde{\lambda}^k)\|^2\Big\} \\ &\quad + s\|\lambda^k - \tilde{\lambda}^k\|^2 - \frac{1}{4}\Big\{s\|\lambda^k - \tilde{\lambda}^k\|^2 + \frac{1}{s}\|A(x^k - \tilde{x}^k)\|^2\Big\} \\ &= \frac{3}{4}\Big\{r\|x^k - \tilde{x}^k\|^2 + s\|\lambda^k - \tilde{\lambda}^k\|^2\Big\} \\ &\quad - \frac{1}{4}\Big\{\frac{1}{s}\|A(x^k - \tilde{x}^k)\|^2 + \frac{1}{r}\|A^T(\lambda^k - \tilde{\lambda}^k)\|^2\Big\}. \end{aligned}$$

在条件 (2.21) 满足的情况下, 根据上式就有

$$\varphi(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k)^T DM(u^k - \tilde{u}^k) \ge \frac{1}{4}\|u^k - \tilde{u}^k\|_D^2. \tag{2.22}$$

The General Contraction Method I 对给定的 u^k 和由 (2.1) 生成的 $\tilde{u}^k$, 用

$$u(\alpha) = u^k - \alpha M(u^k - \tilde{u}^k) \tag{2.23}$$

产生依赖于步长 α 的迭代点. 对任意给定的 $u^* \in \Omega^*$, 我们将

$$\vartheta(\alpha) = \|u^k - u^*\|_D^2 - \|u(\alpha) - u^*\|_D^2 \tag{2.24}$$

看成是本次迭代的“进步量”, 它是步长 α 的函数. 利用 (2.23) 和 (2.24) 中 $\vartheta(\alpha)$ 的定义, 我们有

$$\begin{aligned} \vartheta(\alpha) &= \|u^k - u^*\|_D^2 - \|(u^k - u^*) - \alpha M(u^k - \tilde{u}^k)\|_D^2 \\ &= 2\alpha(u^k - u^*)^T DM(u^k - \tilde{u}^k) - \alpha^2 \|M(u^k - \tilde{u}^k)\|_D^2. \end{aligned}$$

对上式右端的 $(u^k - u^*)^T DM(u^k - \tilde{u}^k)$ 使用 (2.12) 和 (2.13), 就有

$$\vartheta(\alpha) \geq q(\alpha),$$

其中

$$q(\alpha) = 2\alpha(u^k - \tilde{u}^k)^T DM(u^k - \tilde{u}^k) - \alpha^2 \|M(u^k - \tilde{u}^k)\|_D^2. \tag{2.25}$$

同样, 注意到 (2.25) 中的 $q(\alpha)$ 是 α 的二次函数, 它在

$$\alpha_k^* = \frac{(u^k - \tilde{u}^k)^T DM(u^k - \tilde{u}^k)}{\|M(u^k - \tilde{u}^k)\|_D^2} = \frac{(u^k - \tilde{u}^k)^T Q(u^k - \tilde{u}^k)}{(M(u^k - \tilde{u}^k))^T Q(u^k - \tilde{u}^k)} \tag{2.26}$$

时取得极大值.

当条件 (2.21) 满足时, 对所有的 $k \geq 0$, 都有 $\alpha^* \geq 1/6$.

根据 (2.26), 为证明上述结论, 只要证明

$$6(u^k - \tilde{u}^k)^T DM(u^k - \tilde{u}^k) - \|M(u^k - \tilde{u}^k)\|_D^2 \geq 0. \tag{2.27}$$

由 (2.22), 我们有

$$\begin{aligned} & 6(u^k - \tilde{u}^k)^T DM(u^k - \tilde{u}^k) - \|M(u^k - \tilde{u}^k)\|_D^2 \\ \geq\ & \|u^k - \tilde{u}^k\|_D^2 + 2(u^k - \tilde{u}^k)^T DM(u^k - \tilde{u}^k) - \|M(u^k - \tilde{u}^k)\|_D^2. \end{aligned}$$

再利用 (2.18), 得到

$$\begin{aligned} & \|u^k - \tilde{u}^k\|_D^2 + 2(u^k - \tilde{u}^k)^T DM(u^k - \tilde{u}^k) - \|M(u^k - \tilde{u}^k)\|_D^2 \\ =\ & 2\|u^k - \tilde{u}^k\|_D^2 - \frac{1}{r}\|A^T(\lambda^k - \tilde{\lambda}^k)\|^2. \end{aligned}$$

当条件 (2.21) 满足时上式右端非负, (2.27) 成立, 就有 $\alpha_k^* \geq 1/6$.

我们想要极大化 $\vartheta(\alpha)$ (见 (2.24)), 由于它含有未知 u^*, 我们不得已才极大化它的下界函数 $q(\alpha)$. 在实际计算中, 我们取

$$u^{k+1} = u^k - \gamma\alpha_k^* M(u^k - \tilde{u}^k), \tag{2.28}$$

为新的迭代点, 其中 $\gamma \in [1, 2)$ 称为松弛因子.

利用 $\vartheta(\alpha) \geq q(\alpha)$, 将 (2.25) 中的 α 置换成 $\gamma\alpha^*$, 并用 (2.26), 就有

$$\begin{aligned} \|u^{k+1} - u^*\|_D^2 &\leq \|u^k - u^*\|_D^2 - q(\gamma\alpha^*) \\ &= \|u^k - u^*\|_D^2 - \gamma(2-\gamma)\alpha_k^*(u^k - \tilde{u}^k)^T DM(u^k - \tilde{u}^k). \end{aligned}$$

根据上面的不等式, 由 $\alpha_k^* \geq 1/6$ 和关系式 (2.22), 就得到下面的定理

Theorem 2.2 在基于*Primal-Dual* 松弛 *PPA* 的收缩算法中, 如果产生预测点时条件 (2.21) 成立, 则由一般收缩算法 (2.28) 生成的序列$\{u^k = (x^k, \lambda^k)\}$ 满足

$$\|u^{k+1} - u^*\|_D^2 \leq \|u^k - u^*\|_D^2 - \frac{\gamma(2-\gamma)}{24}\|u^k - \tilde{u}^k\|_D^2. \tag{2.29}$$

迭代序列 $\{u^k\}$ 是 D-模下 Fejér 单调的.

The General Contraction Method II 对给定的 u^k 和由 (2.1) 生成的 $\tilde{u}^k$, 我们也可以用

$$u^{k+1} = u^k - \gamma\alpha_k^* Q^{-T} D(u^k - \tilde{u}^k) \tag{2.30}$$

生成新的迭代点, 由于矩阵

$$Q^{-T}D = \begin{pmatrix} \frac{1}{r}I_n & 0 \\ -\frac{1}{rs}A & \frac{1}{s}I_m \end{pmatrix}\begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix} = \begin{pmatrix} I_n & 0 \\ -\frac{1}{s}A & I_m \end{pmatrix}$$

的形式是简单的, 校正 (2.30) 容易实现, 步长则由

$$\alpha_k^* = (u^k - \tilde{u}^k)^T Q(u^k - \tilde{u}^k)/\|u^k - \tilde{u}^k\|_D^2, \qquad \gamma \in (0, 2)$$

给出. 由 (2.22), 可知 $\alpha_k^* \geq 1/4$. 对 $H = QD^{-1}Q^T$, 利用 (2.12) 和 (2.13), 则有

$$\begin{aligned} \|u^{k+1} - u^*\|_H^2 &= \|u^k - u^* - \gamma\alpha^* Q^{-T} D(u^k - \tilde{u}^k)\|_H^2 \\ &\leq \|u^k - u^*\|_H^2 - 2\gamma\alpha_k^*(u^k - \tilde{u}^k)^T Q(u^k - \tilde{u}^k) + \gamma^2(\alpha_k^*)^2\|u^k - \tilde{u}^k\|_D^2 \\ &= \|u^k - u^*\|_H^2 - \gamma(2-\gamma)\alpha_k^*(u^k - \tilde{u}^k)^T Q(u^k - \tilde{u}^k). \\ &\leq \|u^k - u^*\|_H^2 - \frac{\gamma(2-\gamma)}{16}\|u^k - \tilde{u}^k\|_D^2. \end{aligned}$$

由于 u^* 是任意给定的解点, 这样生成的序列 $\{u^k\}$ 在 H-模下 Fejér 单调的.

3 Dual-Primal 松弛 PPA 收缩算法

通过对经典PPA (1.8) 中下半部分中的 $\tilde{x}^k$ 松弛成 x^k 构造 Relaxed PPA. 子问题 (1.8) 就简化成求 $(\tilde{x}^k, \tilde{\lambda}^k) \in \Omega$, 使得对任意的 $u = (x, \lambda) \in \Omega$, 都有

$$\theta(x) - \theta(\tilde{x}^k) + \begin{pmatrix} x - \tilde{x}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{\lambda}^k \\ Ax^k - b \end{pmatrix} + \begin{pmatrix} r(\tilde{x}^k - x^k) \\ s(\tilde{\lambda}^k - \lambda^k) \end{pmatrix} \right\} \geq 0. \tag{3.1}$$

松弛以后的子问题 (3.1) 是容易求解的. 首先对 (3.1) 的下半 (Dual) 部分, 根据变分不等式和投影的关系, $\tilde{\lambda}^k$ 可以通过

$$\tilde{\lambda}^k = P_\Lambda[\lambda^k - \frac{1}{s}(Ax^k - b)]$$

直接给出. 有了 $\tilde{\lambda}^k$, (3.1) 的上半 (Primal) 部分中要求的只剩下 $\tilde{x}^k$, 它可以通过求解极小化问题

$$\min\left\{\theta(x) + \frac{r}{2}\|x - [x^k + \frac{1}{r}A^T\tilde{\lambda}^k]\|^2 \,\middle|\, x \in \mathcal{X}\right\}$$

得到. 根据假设条件, 这是一个容易求解的问题.

用松弛 PPA 按先 $\tilde{\lambda}^k$ (dual) 后 $\tilde{x}^k$ (primal) 的顺序生成预测点 $\tilde{u}^k = (\tilde{x}^k, \tilde{\lambda}^k)$, 然后构造的收缩算法, 我们称之为基于Dual-Primal 松弛 PPA 的收缩算法 (Dual - primal relaxed PPA based contraction method).

3.1 Dual-Primal 生成预测点

Dual–Primal Method 生成预测点 对给定的 (x^k, λ^k) 和 $s > 0$, 由

$$\tilde{\lambda}^k = P_\Lambda\{\lambda^k - \frac{1}{s}(Ax^k - b)\} \tag{3.2a}$$

生成 Dual 预测点 $\tilde{\lambda}^k$. 然后选取适当的 $r > 0$ 并求解

$$\min\left\{\theta(x) + \frac{r}{2}\|x - [x^k + \frac{1}{r}A^T\tilde{\lambda}^k]\|^2 \,\middle|\, x \in \mathcal{X}\right\} \tag{3.2b}$$

得到 Primal 预测点 $\tilde{x}^k$ (如何选取 $r > 0$ 放在后面讨论).

根据前面的分析 (见 (3.1)), 由 (3.2) 生成的预测点 $(\tilde{x}^k, \tilde{\lambda}^k)$ 满足

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

它的紧凑形式是

$$\tilde{u}^k \in \Omega, \ \theta(x) - \theta(\tilde{x}^k) + (u - \tilde{u}^k)^T \{F(\tilde{u}^k) + Q(\tilde{u}^k - u^k)\} \geq 0, \ \forall u \in \Omega, \quad (3.3)$$

其中矩阵(对应于 Dual-Primal 方法生成预测点的 Q, 有时也记成 Q_{DP})

$$Q_{DP} = \begin{pmatrix} rI_n & 0 \\ -A & sI_m \end{pmatrix}, \quad (3.4)$$

是非对称的. 类似地, 可以得到下面的关键不等式

$$(\tilde{u}^k - u^*)^T Q(u^k - \tilde{u}^k) \geq 0, \quad \forall \ u^* \in \Omega^*. \quad (3.5)$$

采用

$$M(u^k - \tilde{u}^k), \qquad (\text{其中} \quad M = D^{-1}Q) \quad (3.6)$$

作为寻查方向, 其中矩阵 H 如 (2.14) 给出. 根据这些定义, 我们有

$$M(u^k - \tilde{u}^k) = \begin{pmatrix} x^k - \tilde{x}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ -\frac{1}{s}A & 0 \end{pmatrix} \begin{pmatrix} x^k - \tilde{x}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \quad (3.7)$$

同时有

$$(u^k - u^*)^T DM(u^k - \tilde{u}^k) \geq \varphi(u^k, \tilde{u}^k), \quad \forall \ u^* \in \Omega^*, \quad (3.8)$$

其中

$$\varphi(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k)^T DM(u^k - \tilde{u}^k). \quad (3.9)$$

3.2 初等的收缩算法

这里的收缩算法是一种预测校正方法. 初等的收缩算法 (Primary Contraction Methods) 是在选取确定的方向后, 校正产生新迭代点时取单位步长的算法.

The Primary Contraction Methods 对给定的 u^k 和由 (3.2) 生成的 $\tilde{u}^k$, 我们用

$$u^{k+1} = u^k - M(u^k - \tilde{u}^k) \tag{3.10}$$

生成新的迭代点. 在用 (3.2) 生成的 $\tilde{u}^k$ 的过程中, 采用 Armijo 法则, 可以做到

对给定的$s > 0$ 和 $\nu \in (0,1)$, 选取 r 使得

$$\frac{1}{s}\|A(x^k - \tilde{x}^k)\|^2 \le \nu\big(r\|x^k - \tilde{x}^k\|^2 + s\|\lambda^k - \tilde{\lambda}^k\|^2\big). \tag{3.11}$$

特别当 s 取得使 $sr\nu \ge \|A^TA\|$ 时, 条件 (3.11) 自然成立.

与定理 2.1 类似, 对基于Dual-Primal 松弛 PPA 的收缩算法, 有如下的定理:

Theorem 3.1 在基于*Dual-Primal* 松弛 *PPA* 的收缩算法中, 如果产生预测点时条件 (3.11) 成立, 则由初等收缩算法 (3.10) 生成的序列 $\{u^k = (x^k, \lambda^k)\}$ 满足

$$\|u^{k+1} - u^*\|_D^2 \le \|u^k - u^*\|_D^2 - (1-\nu)\|u^k - \tilde{u}^k\|_D^2. \tag{3.12}$$

定理 3.1 是保证初等收缩算法收敛的关键不等式.

3.3 一般的收缩算法

一般收缩算法用同样的给定方向 $M(u^k - \tilde{u}^k)$, 通过计算步长确定下一个迭代点. 在 H-模的意义下, 使新的迭代点靠解集尽可能近一些.

对给定的 $s > 0$, 选取 r 使得

$$\begin{aligned}\frac{1}{s}\|A(x^k - \tilde{x}^k)\|^2 + \frac{1}{r}\|A^T(\lambda^k - \tilde{\lambda}^k)\|^2 \\ \le \quad 2\big(r\|x^k - \tilde{x}^k\|^2 + s\|\lambda^k - \tilde{\lambda}^k\|^2\big).\end{aligned} \tag{3.13}$$

这是能够办到的. 当 s 取得使 $sr \ge \frac{1}{2}\|A^TA\|$ 时, 条件 (3.13) 自然成立.

类似与 (2.21)–(2.22) 的分析, 在条件 (3.13) 满足的情况下, 有

$$\varphi(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k)^T DM(u^k - \tilde{u}^k) \ge \frac{1}{4}\|u^k - \tilde{u}^k\|_D^2. \tag{3.14}$$

一般的收缩算法 I 对给定的 u^k 和由 (3.2) 生成的 $\tilde{u}^k$, 我们取

$$u^{k+1} = u^k - \gamma\alpha_k^* M(u^k - \tilde{u}^k), \tag{3.15}$$

为新的迭代点, 其中

$$\alpha_k^* = (u^k - \tilde{u}^k)^T DM(u^k - \tilde{u}^k)/\|M(u^k - \tilde{u}^k)\|_D^2 \tag{3.16}$$

$\gamma \in [1, 2)$ 称为松弛因子. 同样, 在条件 (3.13) 满足时, 会有 $\alpha_k^* > 1/6$.

Theorem 3.2 在基于*Dual-Primal* 松弛 *PPA* 的收缩算法中, 如果产生预测点时条件 (3.13) 成立, 则由收缩算法 (3.15) 生成的序列 $\{u^k = (x^k, \lambda^k)\}$ 满足

$$\|u^{k+1} - u^*\|_D^2 \le \|u^k - u^*\|_D^2 - \frac{\gamma(2-\gamma)}{24}\|u^k - \tilde{u}^k\|_D^2. \tag{3.17}$$

同样, 迭代序列 $\{u^k\}$ 是 D-模下 Fejér 单调的, 上式是证明收敛的关键不等式.

一般的收缩算法 II 同样, 对给定的 u^k 和由 (3.2) 生成的 $\tilde{u}^k$, 也可以用

$$u^{k+1} = u^k - \gamma\alpha_k^* Q^{-T} D(u^k - \tilde{u}^k) \tag{3.18}$$

生成新的迭代点, 其中 $\alpha_k^* = \varphi(u^k, \tilde{u}^k)/\|u^k - \tilde{u}^k\|_D^2$, 取 $H = QD^{-1}Q^T$, 就有

$$\|u^{k+1} - u^*\|_H^2 \le \|u^k - u^*\|_H^2 - \frac{\gamma(2-\gamma)}{16}\|u^k - \tilde{u}^k\|_D^2.$$

由于 u^* 是任意给定的解点, 这样生成的序列 $\{u^k\}$ 在 H-模下 Fejér 单调.

4 数值计算

我们以第四讲中试验的相关性矩阵校正和矩阵完整化问题作为例子, 并与前一讲的 PPA 算法比较计算效果. 我们采用基于Dual-Primal 松弛 PPA 的收缩算法.

4.1 相关性矩阵校正问题

相关性矩阵校正问题的数学形式是

$$\min\{\frac{1}{2}\|X - C\|_F^2 \mid \operatorname{diag}(X) = e,\ X \in S_+^n\}, \tag{4.1}$$

在第四讲的 §4.1 中已经作了介绍. 用 $z \in \Re^n$ 作为等式约束 $\operatorname{diag}(X) = e$ 的 Lagrange 乘子. 我们用 (3.2) 生成问题 (4.1) 的预测点, 问题 (3.2b) 是求解

$$\min\{\frac{1}{2}\|X - C\|_F^2 + \frac{r}{2}\|X - [X^k + \frac{1}{r}\operatorname{diag}(\tilde{z}^k)]\|_F^2 | X \in S_+^n\}. \tag{4.2}$$

这样的工作在前一讲的 §4.1 作了介绍, 求 $(\tilde{x}^k, \tilde{\lambda}^k)$ 的主要工作是做 EIG 分解.

Relaxed PPA-Primal method 只是将前一讲 Code 4.1 (Classical PPA) 的第 (9) 行的

`A=(X0*r + C + diag(yt*2 -y0))/(1+r)` 改成 `A=(X0*r + C + diag(yt))/(1+r)`

并将 (12) 行的 `y = y0 - EY` 改成 `y = y0 - (EY - diag(EX)/s)`.

Code 4.1. Relaxed PPA – Primal method $rs = 1.01,\ s = 0.5$

```
%%%    RePPA (primal mothed) for calibrating correlation matrix   %(1)
function RePPA_MP(n,C,r,s,tol);                                   %(2)
X=eye(n);      y=zeros(n,1);    tic;      %%  The initial iterate %(3)
stopc=1;                        k=0;                              %(4)
while (stopc>tol && k<=100)          %% Beginning of an Iteration  %(5)
 if mod(k,20)==0 fprintf(' k=%4d   epsm=%9.3e  \n',k,stopc); end; %(6)
    X0=X;    y0=y;    k=k+1;                                      %(7)
    yt=y0 - (diag(X0)-ones(n,1))/s;              EY=y0-yt;        %(8)
    A=(X0*r + C + diag(yt))/(1+r);                                 %(9)
    [V,D]=eig(A);   D=max(0,D);  XT=(V*D)*V';     EX=X0-XT;       %(10)
    ex=max(max(abs(EX)));  ey=max(abs(EY));   stopc=max(ex,ey);   %(11)
    X=X0 - EX;             y=y0 - (EY - diag(EX)/s;               %(12)
end;                                        % End of an Iteration %(13)
toc;                TB = max(abs(diag(X-eye(n))));                %(14)
fprintf(' k=%4d  epsm=%9.3e   max|X_jj - 1|=%8.5f \n',k,stopc,TB); %%
```

采用 Relaxed-PPA-Prmal method 与 Classical-PPA method 计算效果完全一样.

我们下面比较 Relaxed PPA – general Method 与前一讲的 Extended PPA Method 的计算效率.

从 Relaxed PPA – Primal Method 到 Relaxed PPA – general Method, 其差别在:

- Code 4.2 与 Code 4.1 的前 11 行完全相同.
- 在 Code 4.2 中, 增加的 12–14 行是为了计算步长.
- Code 4.2 中的第 15 行给出新的迭代点, 与 Code 4.1 中给出新迭代点的第 11 行相比, 寻查方向相同, 只是步长不同. Code 4.1 中用的是单位步长. Code 4.2 中的步长要用公式 (3.16) 计算.
- 为使用 Relaxed PPA-Primal Method, 需要满足条件 $rs \geq \|A^TA\|$ (见 (3.11)), 由于 $\|A^TA\| = 1$, 我们取 $rs = 1.01$, $s = 0.5$.
- 对使用 Relaxed PPA-General Method, 只需要满足条件 $rs \geq \frac{1}{2}\|A^TA\|$ (见 (3.13)), 我们取 $rs = 0.65$, $s = 0.4$. 相对于步长为 1 的算法, 这里生成预测点时, 将参数 s 和 r 都乘上一个因子 0.8.

Code 4.2 Relaxed PPA – general method $rs = 0.65,\ s = 0.4$

```
%%%  RePPA_MG (general mothed)for calibrating correlation matrix  %(1)
function RePPA_MG(n,C,r,s,tol,gamma)                              %(2)
X=eye(n);      y=zeros(n,1);    tic;     %%  The initial iterate  %(3)
stopc=1;                       k=0;                               %(4)
while (stopc>tol && k<=100)         %% Beginning of an Iteration  %(5)
 if mod(k,20)==0 fprintf(' k=%4d   epsm=%9.3e  \n',k,stopc); end; %(6)
    X0=X;    y0=y;    k=k+1;                                      %(7)
    yt=y0 - (diag(X0)-ones(n,1))/s;              EY=y0-yt;        %(8)
    A=(X0*r + C + diag(yt))/(1+r);                                 %(9)
    [V,D]=eig(A);   D=max(0,D);  XT=(V*D)*V';     EX=X0-XT;       %(10)
    ex=max(max(abs(EX)));  ey=max(abs(EY));   stopc=max(ex,ey);  %(11)
    T1=EX(:)'*EX(:);     T2=EY(:)'*EY(:);                        %(12)
    dEX=diag(EX);        T12 =EY'*dEX;        T3=dEX'*dEX;       %(13)
    alpha=(T1*r + T2*s - T12)/(T1*r + T2*s - T12*2 + T3/s);      %(14)
    X=X0-EX*(alpha*gamma);    y=y0-(EY-dEX/s)*(alpha*gamma);     %(15)
end;                                          % End of an Iteration %(16)
toc;                  TB = max(abs(diag(X-eye(n))));             %(17)
fprintf(' k=%4d  epsm=%9.3e   max|X_jj - 1|=%8.5f \n',k,stopc,TB); %%
```


矩阵校正问题 (4.1)–使用 Matlab 中的 eig 子程序

$n \times n$ Matrix	Extended PPA		Relaxed PPA	
$n =$	No. It	CPU Sec.	No. It	CPU Sec.
100	22	0.24	22	0.24
200	25	1.42	22	1.24
500	27	11.66	22	9.62
800	29	50.47	23	39.77
1000	31	99.26	25	78.95
2000	41	883.76	33	713.36

♣ 关于相关系数矩阵校正的程序在附件的 Codes-05 的文件夹“矩阵校正”的 MaT-EIG 中. 只要运行 demo.m, 输入 n 就可以了.
其中的 PPAC.m 和 PPAG.m 分别是 Classical PPA 和 Extended PPA 的子程序.
RePPAMP.m 和 RePPAMG.m 分别是 Relaxed PPA-Primal Method 和 Relaxed PPA-General Method 的子程序.

♣ 如果改用 Kim TOH 写的 mexeig 做 $[V, D] =$ mexeig(A), 计算时间大为节省. 相应的程序在附件的 Codes-05 的文件夹"矩阵校正" 的 Mex-EIG 中.

矩阵校正问题 (4.1)–特征值分解使用 mexeig

$n \times n$ Matrix	Extended PPA		Relaxed PPA	
$n =$	No. It	CPU Sec.	No. It	CPU Sec.
100	22	0.09	22	0.09
200	25	0.37	22	0.32
500	27	3.35	22	2.67
800	29	11.56	23	9.12
1000	31	22.58	25	18.20
2000	39	209.90	33	168.79

从计算实践看, 对同样的问题, Relaxed PPA-General Method 的计算效果比 Extended PPA 要好. 这两个方法生成预测点的工作量是相同的, 虽然用 Relaxed PPA-General Method 需要计算步长, 因为计算步长是 $O(n^2)$ 的运算, Relaxed PPA-General Method 花费的时间还是有较大幅度的节省.

4.2 矩阵完整化方面的应用

矩阵完整化问题的试验例子与前一讲的 §4.2 相同, 算例来自 [1]. 我们只将 Relaxed PPA-General Method 与 Extended PPA 做比较. 这两个方法生成预测点的工作量是相同的.

♣ 矩阵完整化用 Matlab SVD 的程序在附件的 Codes-05 的文件夹 矩阵完整化–SVD-MaT 中. 只要运行 demo.m 就可以了. 要对不同情形试验, 只要在 demo.m 中用 % 做适当选择. 其中的 PPAGMaT.m 和 PPAMMaT.m 分别是 Extended PPA 和 Relaxed PPA 的子程序, 均采用 $\gamma = 1.5$.

如果用 Matlab 中的标准 SVD, 对同样规模的问题, 计算时间与迭代次数成正比. 虽然用 Relaxed PPA-General Method 需要计算步长, 因为计算步长是 $O(n^2)$ 的运算, 这在总的计算量中是微不足道的. 我们将Relaxed PPA-General Method 求解矩阵完整化问题的程序作为 Code 4.3 列在后面.

用 Relaxed PPA-General method 求解矩阵完整化问题与前一讲的 Code 4.3 (Extended PPA) 的差别是：将第 (10) 行的

`A=X0 + (YT*2 -Y0)/r` 改成 `A = X0 + (YT)/r` .

在 Code 4.3 (Extended PPA) 第 (13) 行以后加上三行计算步长 $\alpha = \gamma\alpha^*$.

并将 Code 4.3 (Extended PPA) 的 (15)-(16) 行改成了现在的 (18)-(19) 行.

矩阵完整化问题：用 Matlab 中标准SVD 求解结果

Unknown $n \times n$ matrix M				Extended PPA		Relaxed PPA	
n	rank(ra)	m/d_{ra}	m/n^2	#iters	times(Sec.)	#iters	times(Sec.)
1000	10	6	0.12	77	867.82	56	629.67
1000	50	4	0.39	37	411.28	27	302.27
1000	100	3	0.58	31	362.58	29	305.93

因此, 用 Relaxed PPA-General Method 花费的时间有较大幅度的节省.

♣ 注意到, [1] 中的方法对这三个例子的迭代次数分别是 117, 114 和 129 (See the first three examples in Table 5.1 of [1], pp. 1974). 他们的方法每次迭代的主要工作量也是做一次 SVD 分解, 由于采用了不完全分解技术, [1] 中节省了每次 SVD 分解的运行时间. 我们调用的是 Matlab 中标准SVD, 减少了迭代次数, 但没有节省总的运行时间.

Code 4.3. Relaxed PPA for Matrix Completion Problem

```
function PPAE(n,r,s,M,Omega,maxIt,tol,gamma)    %  Ititial Process %%(1)
X=zeros(n);     Y=zeros(n);      YT=zeros(n);                        %(2)
  nM0=norm(M(Omega),'fro');    eps=1;  VioKKT=1;  k=0;   tic;        %(3)
%% Minimum nuclear norm solution by PPA method                      %(4)
while  (eps > tol && k<= maxIt)                                      %(5)
 if mod(k,5)==0                                                      %(6)
 fprintf('It=%3d |X-M|/|M|=%9.2e VioKKT=%9.2e\n',k,eps,VioKKT); end;%(7)
  k=k+1;      X0=X;      Y0=Y;                                      %(8)
  YT(Omega)=Y0(Omega)-(X0(Omega)-M(Omega))/s;            EY=Y-YT;   %(9)
  A = X0 + (YT)/r;                          [U,D,V]=svd(A,0);       %(10)
  D=D-eye(n)/r;     D=max(D,0);     XT=(U*D)*V';         EX=X-XT;  %(11)
  DXM=XT(Omega)-M(Omega);           eps = norm(DXM,'fro')/nM0;      %(12)
  VioKKT = max( max(max(abs(EX)))*r, max(max(abs(EY))) );           %(13)
  T1=EX(:)'*EX(:);      T2=EY(:)'*EY(:);                            %(14)
  T12 = EY(:)'*EX(:);   EXOm = EX(Omega);  T3 = EXOm(:)'*EXOm(:);   %(15)
  alpha =(T1*r + T2*s - T12)*gamma/(T1*r + T2*s - T12*2 + T3/s);    %(16)
   if (eps <= tol)      alpha=1; end;                               %(17)
  X = X0 - EX*alpha;                                                %(18)
  Y(Omega) = Y0(Omega) -  (EY(Omega) - EXOm/s)*alpha;               %(19)
end                                                                 %(20)
 fprintf('It=%3d  |X-M|/|M|=%9.2e  Vi0KKT=%9.2e \n',k,eps,VioKKT); %(21)
   RelEr=norm((X-M),'fro')/norm(M,'fro');   toc;                    %(22)
 fprintf(' Relative error = %9.2e  Rank(X)=%3d \n',RelEr,rank(X)); %(23)
 fprintf(' Violation of KKT Condition = %9.2e  \n',VioKKT);         %(24)
```

4.3 Min-Max 问题上的应用

对第一讲 §4.2 中提到用全变差极小处理图像去模糊 [2] 的 min-max 问题,

$$\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\ \Phi(x,y) := \theta_1(x) - y^TAx - \theta_2(y). \tag{4.3}$$

它可以转换成等价的变分不等式: 求 $u^* = (x^*, y^*) \in \mathcal{X}\times\mathcal{Y}$, 使得

$$\theta(u) - \theta(u^*) + \begin{pmatrix} x - x^* \\ y - y^* \end{pmatrix}^T \begin{pmatrix} -A^Ty^* \\ Ax^* \end{pmatrix} \ge 0, \quad \forall (x,y)\in\mathcal{X}\times\mathcal{Y}, \tag{4.4}$$

其中 $\theta(u) = \theta_1(x) + \theta_2(y)$. 用这一讲介绍的 Relaxed PPA 方法去求解, 对给定的 (x^k, y^k), 可以通过

$$\tilde{x}^k = \text{Argmin}\{\theta_1(x) - x^TA^Ty^k + \frac{r}{2}\|x - x^k\|^2 \mid x\in\mathcal{X}\}. \tag{4.5a}$$

$$\tilde{y}^k = \text{Argmin}\big\{\theta_2(y) + y^TA\tilde{x}^k + \frac{s}{2}\|y - y^k\|^2 \,\big|\, y\in\mathcal{Y}\big\}. \tag{4.5b}$$

或者

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + y^TAx^k + \frac{s}{2}\|y - y^k\|^2 \mid y\in\mathcal{Y}\}. \tag{4.6a}$$

$$\tilde{x}^k = \text{Argmin}\big\{\theta_1(x) - x^TA^T\tilde{y}^k + \frac{r}{2}\|x - x^k\|^2 \,\big|\, x\in\mathcal{X}\big\}. \tag{4.6b}$$

由 (4.5) 生成的 $(\tilde{x}^k, \tilde{y}^k)\in\mathcal{X}\times\mathcal{Y}$, 对一切 求 $(x,y)\in\mathcal{X}\times\mathcal{Y}$, 都有

$$\theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^Ty^k \\ A\tilde{x}^k \end{pmatrix} + \begin{pmatrix} r(\tilde{x}^k - x^k) \\ s(\tilde{y}^k - y^k) \end{pmatrix} \right\} \ge 0. \tag{4.7}$$

由 (4.6) 生成的 $(\tilde{x}^k, \tilde{y}^k)\in\mathcal{X}\times\mathcal{Y}$, 对一切 求 $(x,y)\in\mathcal{X}\times\mathcal{Y}$, 都有

$$\theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{y}^k \\ Ax^k \end{pmatrix} + \begin{pmatrix} r(\tilde{x}^k - x^k) \\ s(\tilde{y}^k - y^k) \end{pmatrix} \right\} \ge 0. \tag{4.8}$$

无论是 (4.7) 还是 (4.8), 都可以写成

$$\tilde{u}^k\in\Omega, \qquad (u - \tilde{u}^k)^T\{F(\tilde{u}^k) + Q(\tilde{u}^k - u^k)\} \ge 0, \quad \forall\, u\in\Omega, \tag{4.9}$$

的形式, 所不同的只是

$$Q = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \quad \text{和} \quad Q = \begin{pmatrix} rI_n & 0 \\ -A & sI_m \end{pmatrix}.$$

都是当 $rs > \frac{1}{4}\|A^TA\|$ 时, 矩阵 $Q^T + Q$ 正定. 我们建议取 $rs \approx 0.7\|A^TA\|$.

对由 (2.11) 定义的矩阵 D, 新的迭代点u^{k+1} 用迭代式

$$u^{k+1} = u^k - \gamma\alpha_k^* M(u^k - \tilde{u}^k), \quad \gamma \in [1, 2),$$

生成(一般取 $\gamma = 1.5$). 采用

$$M = D^{-1}Q, \qquad \alpha_k^* = \frac{(u^k - \tilde{u}^k)^T Q(u^k - \tilde{u}^k)}{\|M(u^k - \tilde{u}^k)\|_D^2}, \tag{4.10}$$

生成的序列 $\{u^k\}$ 在 D-模下 Fejér 单调; 若采用

$$M = Q^{-T}D, \qquad \alpha_k^* = \frac{(u^k - \tilde{u}^k)^T Q(u^k - \tilde{u}^k)}{\|u^k - \tilde{u}^k\|_D^2}, \tag{4.11}$$

生成的序列 $\{u^k\}$ 则在 H-模下 Fejér 单调的($H = QD^{-1}Q^T$). 类似的结论可参阅本讲的 §2.3 和 §3.3. 用这类方法求解图像去模糊问题可以参考 [8].

References

[1] J. F. Cai, E. J. Candès and Z. W. Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optim., **20**, 1956-1982, 2010.

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

[3] B.S. He, A new method for a class of linear variational inequalities, Math. Progr., **66**, 137-144, 1994.

[4] B.S. He, A class of projection and contraction methods for monotone variational inequalities, Applied Mathematics and Optimization, **35**, 69-76, 1997.

[5] B.S. He, Inexact implicit methods for monotone general variational inequalities, Math. Program. Series A, **86**, 199-217, 1999

[6] B. S. He, X. L. Fu and Z.K. Jiang, Proximal point algorithm using a linear proximal term, JOTA **141**, 209-239, 2009.

[7] B. S. He and M.-H. Xu, A general framework of contraction methods for monotone variational inequalities, Pacific J. Optimization, **4**, 195-212, 2008.

[8] 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**, 119-149, 2012.

[9] R. M. Larsen, Propack – software for large and sparse svd calculation, http://sun.stanford.edu/rmunk/PROPACK/, 2005.

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

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

凸优化和单调变分不等式的收缩算法

第八讲: 基于增广 Lagrange 乘子法的 PPA 收缩算法

Augmented Lagrangian-based PPA contraction methods for constrained convex optimization

南京大学数学系　何炳生
hebma@nju.edu.cn

1 线性约束凸优化等价的单调变分不等式

考虑一般的线性等式约束凸优化问题

$$\min\{\theta(x) \mid Ax = b,\ x \in \mathcal{X}\}, \tag{1.1}$$

其中 $\theta(x)$ 是凸函数, $A \in \Re^{m\times n},\ b \in \Re^m$, $\mathcal{X}$ 是 $\Re^n$ 中的闭凸集. 凸优化问题 (1.1) 的 Lagrange 函数是定义在 $\mathcal{X} \times \Re^m$ 上的

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

记 $\Lambda = \Re^m$, 设 (x^*, λ^*) 是 Lagrange 函数的一个鞍点, 便有

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

求 Lagrange 函数的一个鞍点等价于求 (x^*, λ^*) 使其满足:

$$\left\{\begin{array}{lll} x^* \in \mathcal{X}, & \theta(x) - \theta(x^*) + (x - x^*)^T(-A^T\lambda^*) \ge 0, & \forall\, x \in \mathcal{X}, \\ \lambda^* \in \Lambda, & (\lambda - \lambda^*)^T(Ax^* - b) \ge 0, & \forall\, \lambda \in \Lambda, \end{array}\right. \tag{1.2}$$

通过定义

$$u = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(u) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix} \quad \text{和} \quad \Omega = \mathcal{X} \times \Lambda, \tag{1.3}$$

求解 (1.1) 相当于求解变分不等式

$$\text{VI}(\Omega, F) \quad u^* \in \Omega, \quad \theta(x) - \theta(x^*) + (u - u^*)^T F(u^*) \geq 0, \quad \forall\, u \in \Omega. \tag{1.4}$$

注意到, (1.3) 中的仿射算子 F 是单调的.

对给定的常数 $s > 0$, 定义在 $\Omega = \mathcal{X} \times \Re^m$ 上的

$$\mathcal{L}_A(x, \lambda) = \theta(x) - \lambda^T(Ax - b) + \frac{1}{2s}\|Ax - b\|^2$$

是 (1.1) 的增广 Lagrange 函数. 增广Lagrange 乘子法 [1, 8, 11] 是求等式约束优化问题的很有效的方法之一, 对此, [10] 中第 17 章有详细论述.

求上述等式约束凸优化问题的经典增广 Lagrange 乘子法 (Augmented Lagrangian Method) 的 k-步迭代, 是从给定的 λ^k 出发, 求

$$x^{k+1} = \text{Argmin}\{\mathcal{L}_A(x, \lambda^k) \,|\, x \in \mathcal{X}\},$$

然后以

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

得到新的迭代点.

设 Ω^* 是 VI(Ω, F) 的解集合. 为考虑收缩算法需要, 我们记

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

如果用经典的增广 Lagrange 方法求解问题 (1.1), 对任意给定的 $b \in \Re^m$, 我们假设子问题

$$\boxed{\min\{\theta(x) + \tfrac{1}{2s}\|Ax - b\|^2 \,|\, x \in \mathcal{X}\}}$$

是容易求解的. 否则的话, 采用基于松弛增广 Lagrange 的收缩方法. 这时, 需要假设对任意给定的 $a \in \Re^n$, 问题

$$\boxed{\min\{\theta(x) + \tfrac{r}{2}\|x - a\|^2 \,|\, x \in \mathcal{X}\}}$$

的解是容易求得的.

2 收缩意义下的增广 Lagrange 乘子法

我们称定义在 $\Omega = \mathcal{X} \times \Re^m$ 上的

$$\mathcal{L}_A(x,\lambda) = \theta(x) - \lambda^T(Ax-b) + \frac{1}{2s}\|Ax-b\|^2$$

为等式约束的问题 $\min\{\theta(x) \mid Ax = b,\ x \in \mathcal{X}\}$ 的增广 Lagrange 函数. 以下是增广 Lagrange 乘子法的框架，它从给定的 λ^k 开始.

等式约束问题的增广 Lagrange 乘子法框架 对给定的 λ^k, 先求

$$\tilde{x}^k = \text{Argmin}\{\theta(x) + \tfrac{1}{2s}\|(Ax-b) - s\lambda^k\|^2 \mid x \in \mathcal{X}\}, \tag{2.1a}$$

然后令

$$\tilde{\lambda}^k = \lambda^k - \tfrac{1}{s}(A\tilde{x}^k - b). \tag{2.1b}$$

经典的增广 Lagrange 乘子法以 $\lambda^{k+1} = \tilde{\lambda}^k$ 完成一次迭代.

我们用收缩算法的观点来分析增广 Lagrange 乘子法, 改成以

$$\lambda^{k+1} = \lambda^k - \gamma(\lambda^k - \tilde{\lambda}^k), \quad \gamma \in (0,2) \tag{2.2}$$

生成新的迭代点. 由 (2.1a) 生成的 $\tilde{x}^k \in \mathcal{X}$ 满足

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

将 (2.1b) 中的 $\tilde{\lambda}^k = \lambda^k - \frac{1}{s}(A\tilde{x}^k - b)$ 代入上式就有

$$\tilde{x}^k \in \mathcal{X}, \quad \theta(x) - \theta(\tilde{x}^k) + (x-\tilde{x}^k)^T\big(-A^T\tilde{\lambda}^k\big) \ge 0, \ \ \forall x \in \mathcal{X}. \tag{2.3}$$

把 (2.3) 和 (2.1b) 组合在一起, 就是 $\tilde{u}^k = (\tilde{x}^k, \tilde{\lambda}^k) \in \Omega$, 并且

$$\theta(x) - \theta(\tilde{x}^k) + \begin{pmatrix} x - \tilde{x}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{\lambda}^k \\ A\tilde{x}^k - b \end{pmatrix} + \begin{pmatrix} 0 \\ s(\tilde{\lambda}^k - \lambda^k) \end{pmatrix} \right\} \ge 0, \ \forall\, u \in \Omega. \tag{2.4}$$

将任意的 $u^* = (x^*, \lambda^*)$ 代入上式中的 $u \in \Omega$, 并利用 $F(u)$ 的表达式就有

$$(\tilde{\lambda}^k - \lambda^*)^T(\lambda^k - \tilde{\lambda}^k) \ge \frac{1}{s}\{(\tilde{u}^k - u^*)^T F(\tilde{u}^k) + \theta(\tilde{x}^k) - \theta(x^*)\}. \tag{2.5}$$

利用 F 的单调性和

$$\theta(\tilde{x}^k) - \theta(x^*) + (\tilde{u}^k - u^*)^T F(u^*) \geq 0,$$

推得 (2.5) 式的右端非负. 所以有

$$(\lambda^k - \lambda^*)^T(\lambda^k - \tilde{\lambda}^k) \geq \|\lambda^k - \tilde{\lambda}^k\|^2, \quad \forall\, \lambda^* \in \Lambda^*. \tag{2.6}$$

用 (2.2) 生成新的迭代点 λ^{k+1}, 在收缩意义下我们通常取 $\gamma \in [1, 2)$, 根据 (2.6) 得到

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

The sequence $\{\lambda^k\}$ (dual variable) generated by Augmented Lagrangian Method is Fejér monotone.

上述性质说明, 增广 Lagrange 乘子法是关于对偶变量 λ 的 PPA 算法. 用收缩算法的观点来考虑问题, 迭代式 (2.2) 中的 γ 可以在区间 $(0, 2)$ 中自由选取. 在实际计算中, 我们建议取 $\gamma \in [1.2, 1.8]$.

3 基于增广 Lagrange 乘子法的 PPA 算法

在实际问题中, 精确求解子问题 (2.1a) 往往是花费很大的. 我们还是假设只有形似 $\min\{\theta(x) + \frac{r}{2}\|x - a\|^2 \mid x \in \mathcal{X}\}$ 的问题是容易求解的. 如果对 (2.1a) 中的目标函数加上 Proximal 项 $\frac{r}{2}\|x - x^k\|^2$, 要处理的子问题就成为

$$\min\{\theta(x) + \frac{1}{2s}\|(Ax - b) - s\lambda^k\|^2 + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}. \tag{3.1}$$

再将 (3.1) 中的 $\frac{1}{2s}\|(Ax - b) - s\lambda^k\|^2$ 在 x^k 处做线性化近似就是

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

对 (3.1) 中的二次项线性化以后, 子问题就变成

$$\min\{\theta(x) + \big(\frac{1}{s}A^T[(Ax^k - b) - s\lambda^k]\big)^T x + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\}. \tag{3.2}$$

基于增广 Lagrange 乘子法的 PPA 算法, 只对 (3.1) 中的二次函数做线性化处理. 每步迭代从给定的 $u^k = (x^k, \lambda^k)$ 开始, 生成 $\tilde{u}^k \in \Omega$.

对二次函数线性化处理 对给定的 (x^k, λ^k), 先求

$$\tilde{x}^k = \text{Argmin}\{\theta(x) + \big(\frac{1}{s}A^T[(Ax^k - b) - s\lambda^k]\big)^T x + \frac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\} \quad (3.3a)$$

然后令

$$\tilde{\lambda}^k = \lambda^k - \frac{1}{s}(A\tilde{x}^k - b). \quad (3.3b)$$

我们按统一框架考察生成的预测点 $\tilde{u}^k$. 由 (3.3a) 生成的 $\tilde{x}^k \in \mathcal{X}$ 满足

$$\theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\lambda^k + \frac{1}{s}A^T(Ax^k - b) + r(\tilde{x}^k - x^k)\} \geq 0, \ \forall x \in \mathcal{X}.$$

将 (3.3b) 中的 $\tilde{\lambda}^k = \lambda^k - \frac{1}{s}(A\tilde{x}^k - b)$ 代入上式就有 $\tilde{x}^k \in \mathcal{X}$, 并对$x \in \mathcal{X}$, 有

$$\theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T\big(-A^T\tilde{\lambda}^k + (rI_n - \frac{1}{s}A^TA)(\tilde{x}^k - x^k)\big) \geq 0, \quad (3.4)$$

把 (3.4) 和 (3.3b) 组合在一起, 就是 $\tilde{u}^k = (\tilde{x}^k, \tilde{\lambda}^k) \in \Omega$, 并对所有的 $(x, \lambda) \in \Omega$,

均有

$$\theta(x) - \theta(\tilde{x}^k) + \begin{pmatrix} x - \tilde{x}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{\lambda}^k \\ A\tilde{x}^k - b \end{pmatrix} + \begin{pmatrix} (rI_n - \frac{1}{s}A^TA)(\tilde{x}^k - x^k) \\ s(\tilde{\lambda}^k - \lambda^k) \end{pmatrix} \right\} \geq 0. \quad (3.5)$$

我们记

$$d(u^k, \tilde{u}^k) = \begin{pmatrix} (rI_n - \frac{1}{s}A^TA)(x^k - \tilde{x}^k) \\ s(\lambda^k - \tilde{\lambda}^k) \end{pmatrix}. \quad (3.6)$$

并利用 $F(u)$ 的表达式, 就有

$$\theta(x) - \theta(\tilde{x}^k) + (u - \tilde{u}^k)^T(F(\tilde{u}^k) - d(u^k, \tilde{u}^k)) \geq 0, \ \ \forall\, u \in \Omega. \quad (3.7)$$

将任意的 $u^* = (x^*, \lambda^*)$ 代入上式中的 $(x, \lambda) \in \Omega$, 并

$$(\tilde{u}^k - u^*)^T d(u^k, \tilde{u}^k) \geq \theta(\tilde{x}^k) - \theta(x^*) + (\tilde{u}^k - u^*)^T F(\tilde{u}^k). \quad (3.8)$$

利用 F 的单调性和 $\theta(\tilde{x}^k) - \theta(x^*) + (\tilde{u}^k - u^*)^T F(u^*) \geq 0$, 上式右端非负. 进而得到

$$(u^k - u^*)^T d(u^k, \tilde{u}^k) \geq (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \quad (3.9)$$

对参数 r, s 的要求 对给定的 $s > 0$, 选取 r 使

$$rs > \|A^T A\|. \tag{3.10}$$

在上述条件下, 矩阵

$$G = \begin{pmatrix} rI_n - \frac{1}{s}A^T A & 0 \\ 0 & sI_m \end{pmatrix}, \tag{3.11}$$

是正定的. 利用 (3.6), 就有

$$d(u^k, \tilde{u}^k) = G(u^k - \tilde{u}^k). \tag{3.12}$$

不等式 (3.9) 就化成

$$(u^k - u^*)^T G(u^k - \tilde{u}^k) \geq \|u^k - \tilde{u}^k\|_G^2. \tag{3.13}$$

直接取 $u^{k+1} = \tilde{u}^k$ 为新的迭代点, 就得到 G-模下的 PPA 算法.

基于增广 Lagrange 乘子法的 PPA 算法

PPA 算法用迭代式

$$u^{k+1} = u^k - \gamma(u^k - \tilde{u}^k), \quad \gamma \in (0, 2) \tag{3.14}$$

产生新的迭代点 u^{k+1}, 我们一般取 $\gamma \in [1, 2)$. 利用 (3.13), 序列 $\{u^k\}$ 满足

$$\|u^{k+1} - u^*\|_G^2 \leq \|u^k - u^*\|_G^2 - \gamma(2-\gamma)\|u^k - \tilde{u}^k\|_G^2.$$

这是保证基于增广 Lagrange 乘子法的 PPA 算法收敛的关键不等式.

使用这一节的算法, 参数 r, s 需要满足条件 $rs > \|A^T A\|$ (见 (3.10)). 因此, 适合用来处理 $\|A^T A\|$ 容易估算的问题.

4 基于增广 Lagrange 乘子法的收缩算法

这一节的基于增广 Lagrange 乘子法的收缩算法, 同样是只对 (3.1) 中的二次函数做线性化处理. 每步迭代从给定的 $u^k = (x^k, \lambda^k)$ 开始, 还是以 (3.3) 生成 $\tilde{u}^k \in \Omega$. 由上一节的分析, 我们得到

$$(u^k - u^*)^T d(u^k, \tilde{u}^k) \geq (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k), \tag{4.1}$$

其中

$$d(u^k, \tilde{u}^k) = \begin{pmatrix} (rI_n - \frac{1}{s}A^TA)(x^k - \tilde{x}^k) \\ s(\lambda^k - \tilde{\lambda}^k) \end{pmatrix} \tag{4.2}$$

我们记

$$H = \begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix}. \tag{4.3}$$

并考虑用

$$M(u^k - \tilde{u}^k) = H^{-1}d(u^k, \tilde{u}^k) \tag{4.4}$$

作为寻查方向. 这时

$$M = \begin{pmatrix} I_n - \frac{1}{rs}A^TA & 0 \\ 0 & I_m \end{pmatrix}. \tag{4.5}$$

我们并不要求 M 正定, 因此不再要求 $rs > \|A^TA\|$, 也不能直接将 $\tilde{u}^k$ 取作新的迭代点. 换句话说, 基于增广 Lagrange 乘子法的收缩算法是一种预测－校正方法, 以 (3.3) 生成的 $\tilde{u}^k \in \Omega$ 只是一个预测点.

生成预测点时对参数 r, s 的要求 对给定的 $s > 0$, 要求 r 满足

$$\|A^TA(x^k - \tilde{x}^k)\| \le rs\nu\|x^k - \tilde{x}^k\|, \quad \nu \in (0,1). \tag{4.6}$$

如果条件 $rs > \|A^TA\|$ (见 (3.10)) 满足, 则条件 (4.6) 自然满足. 然而, 这里只要求在计算的每次迭代中验证条件 (4.6) 是否满足.

我们考虑 H 模下的收缩算法. 注意到不等式 (4.1) 可以写成

$$(u^k - u^*)^T HM(u^k - \tilde{u}^k) \ge (u^k - \tilde{u}^k)^T HM(u^k - \tilde{u}^k), \ \forall\, u^* \in \Omega^*. \tag{4.7}$$

Lemma 4.1 对给定的 $u^k = (x^k, \lambda^k)$, 设 $\tilde{u}^k \in \Omega$ 由 (3.3) 生成. 在条件 (4.6) 满足的情况下, 我们有

$$\begin{aligned} &(u^k - \tilde{u}^k)^T HM(u^k - \tilde{u}^k) \\ &\quad \ge \quad \frac{1}{2}\{\|M(u^k - \tilde{u}^k)\|_H^2 + (1-\nu^2)\|u^k - \tilde{u}^k\|_H^2\}. \end{aligned} \tag{4.8}$$

证明. 我们考察

$$2(u^k-\tilde{u}^k)^T HM(u^k-\tilde{u}^k)-\|d(u^k,\tilde{u}^k)\|_H^2.$$

由于

$$\begin{aligned}&2(u^k-\tilde{u}^k)^T HM(u^k-\tilde{u}^k)-\|M(u^k-\tilde{u}^k)\|_H^2\\ =\quad &\big(2(u^k-\tilde{u}^k)-M(u^k-\tilde{u}^k)\big)^T HM(u^k-\tilde{u}^k).\end{aligned}\tag{4.9}$$

利用 H 和 $d(u^k,\tilde{u}^k)$ 的定义(见 (4.3) 和 (4.4)), 对上式的右端进行处理, 由于

$$M(u^k-\tilde{u}^k)\quad=\quad\begin{pmatrix}(I_n-\frac{1}{rs}A^TA)(x^k-\tilde{x}^k)\\(\lambda^k-\tilde{\lambda}^k)\end{pmatrix}$$

和

$$2(u^k-\tilde{u}^k)-M(u^k-\tilde{u}^k)\quad=\quad\begin{pmatrix}(I_n+\frac{1}{rs}A^TA)(x^k-\tilde{x}^k)\\(\lambda^k-\tilde{\lambda}^k)\end{pmatrix}.$$

将这些代入 (4.9) 的右端就得到

$$\begin{aligned}&2(u^k-\tilde{u}^k)^T HM(u^k-\tilde{u}^k)-\|M(u^k-\tilde{u}^k)\|_H^2\\ =\quad &r\big(\|x^k-\tilde{x}^k\|^2-\frac{1}{r^2s^2}\|A^TA(x^k-\tilde{x}^k)\|^2\big)+s\|\lambda^k-\tilde{\lambda}^k\|^2.\end{aligned}$$

在条件 (4.6) 满足的情况下, 从上式得到

$$\begin{aligned}&2(u^k-\tilde{u}^k)^T HM(u^k-\tilde{u}^k)-\|M(u^k-\tilde{u}^k)\|_H^2\\ \geq\quad &(1-\nu^2)r\|x^k-\tilde{x}^k\|^2+s\|\lambda^k-\tilde{\lambda}^k\|^2.\end{aligned}\tag{4.10}$$

从上式马上得到 (4.8), 引理证明完毕.

我们考虑 H 模下的收缩算法. 先考虑取单位步长的**初等的收缩算法**.

初等的收缩算法 对给定的 u^k 和用 (3.3) 生成的 $\tilde{u}^k$, 用

$$u^{k+1}=u^k-M(u^k-\tilde{u}^k)\tag{4.11}$$

生成新的迭代点. 由矩阵 M 的结构 (4.5), 采用初等收缩算法时 $\lambda^{k+1}=\tilde{\lambda}^k$.

根据迭代公式 (4.11) 并使用 (4.1), 就有

$$\begin{aligned}
&\|u^k - u^*\|_H^2 - \|u^{k+1} - u^*\|_H^2 \\
&\quad = \quad \|u^k - u^*\|_H^2 - \|(u^k - u^*) - M(u^k - \tilde{u}^k)\|_H^2 \\
&\quad = \quad 2(u^k - u^*)^T HM(u^k - \tilde{u}^k) - \|M(u^k - \tilde{u}^k)\|_H^2 \\
&\quad \geq \quad 2(u^k - \tilde{u}^k)^T HM(u^k - \tilde{u}^k) - \|M(u^k - \tilde{u}^k)\|_H^2. \qquad (4.12)
\end{aligned}$$

以引理 4.1 的结论 (4.8) 代入 (4.12), 就有下面的定理:

Theorem 4.1 *Let the condition* (4.6) *be satisfied. Then the sequence* $\{u^k = (x^k, \lambda^k)\}$ *generated by the elementary contraction method satisfies*

$$\|u^{k+1} - u^*\|_H^2 \leq \|u^k - u^*\|_H^2 - (1 - \nu^2)\|u^k - \tilde{u}^k\|_H^2. \qquad (4.13)$$

定理 4.1 中的不等式 (4.13) 是保证初等收缩算法收敛的关键不等式.

再考虑通过计算步长确定下一个迭代点的**一般的收缩算法**.

一般的收缩算法 对给定的 u^k 和 (3.3) 生成的 $\tilde{u}^k$, 用

$$u(\alpha) = u^k - \alpha M(u^k - \tilde{u}^k) \qquad (4.14)$$

产生依赖于步长 α 的迭代点. 同样, 对任意给定的 $u^* \in \Omega^*$, 我们定义

$$\vartheta(\alpha) := \|u^k - u^*\|_H^2 - \|u(\alpha) - u^*\|_H^2 \qquad (4.15)$$

和

$$q(\alpha) = 2\alpha(u^k - \tilde{u}^k)^T HM(u^k - \tilde{u}^k) - \alpha^2\|M(u^k - \tilde{u}^k)\|_H^2. \qquad (4.16)$$

利用 (4.14) 和 (4.15) 中 $\vartheta(\alpha)$ 的定义以及 (4.1), 可以证明

$$\vartheta(\alpha) \geq q(\alpha) \qquad (4.17)$$

同样, 注意到 (4.16) 中的 $q(\alpha)$ 是 α 的二次函数, 它在

$$\alpha^* = \frac{(u^k - \tilde{u}^k)^T HM(u^k - \tilde{u}^k)}{\|M(u^k - \tilde{u}^k)\|_H^2} \qquad (4.18)$$

时取得极大值. 从 (4.8) 式知, 在条件 (4.6) 满足的情况下, $\alpha_k^* \geq 1/2$. 在实际计算中, 我们取

$$u^{k+1} = u^k - \gamma\alpha_k^* M(u^k - \tilde{u}^k), \qquad (4.19)$$

为新的迭代点, 其中 $\gamma \in [1,2)$ 称为松弛因子. 由 (4.15) 和 (4.17), 有

$$\begin{aligned} \|u^{k+1}-u^*\|_H^2 &\le \|u^k-u^*\|_H^2 - q(\gamma\alpha_k^*) \\ &\le \quad \|u^k-u^*\|_H^2 - \gamma(2-\gamma)\alpha_k^*(u^k-\tilde{u}^k)^T HM(u^k-\tilde{u}^k). \end{aligned} \tag{4.20}$$

根据上式, 由 (4.8) 和 $\alpha_k^* \ge 1/2$, 得到

Theorem 4.2 *The sequence* $\{u^k=(x^k,\lambda^k)\}$ *generated by the general contraction method satisfies*

$$\|u^{k+1}-u^*\|_H^2 \le \|u^k-u^*\|_H^2 - \frac{\gamma(2-\gamma)(1-\nu^2)}{4}\|u^k-\tilde{u}^k\|_H^2. \tag{4.21}$$

定理 4.2 中的不等式 (4.21) 是保证一般收缩算法收敛的关键不等式.

此外, 由 (4.18), (4.20) 和 $\alpha_k^* > 1/2$, 会有

$$\|u^{k+1}-u^*\|_H^2 \le \|u^k-u^*\|_H^2 - \frac{\gamma(2-\gamma)}{4}\|M(u^k-\tilde{u}^k)\|_H^2.$$

5 信息技术领域优化问题中的应用

我们用这一讲 §3 的方法求解第四讲中提到的算例.

5.1 相关性矩阵 (Correlation Matrix) 校正中的应用

对给定的对称矩阵 C, 求 F-模下与 C 距离最近的相关性矩阵, 其数学表达式是

$$\min\Big\{\frac{1}{2}\|X-C\|_F^2 \mid \operatorname{diag}(X)=e,\ X\in S_+^n\Big\}, \tag{5.1}$$

其中 e 表示每个分量都为 1 的 n-维向量, S_+^n 表示 $n\times n$ 正半定锥的集合. 问题 (5.1) 是形如 (1.1) 的等式约束凸优化问题, 其中 $\|A^TA\|=1$.

我们用 $z\in\Re^n$ 作为等式约束 $\operatorname{diag}(X)=e$ 的 Lagrange 乘子.

PPA 算法求解问题(5.1)

对给定的 (X^k,z^k), 用 (3.3) 产生$(\tilde{X}^k,\tilde{z}^k)$:

1. Finding $\tilde{X}^k$ which is the solution of the following minimization problem

$$\min\{\frac{1}{2}\|X-C\|_F^2+\frac{r}{2}\|X-[X^k+\frac{1}{r}\,\mathrm{diag}(z^{k+\frac{1}{2}})]\|_F^2|X\in S_+^n\},\quad (5.2)$$

where

$$z^{k+\frac{1}{2}}=z^k-\frac{1}{s}(\mathrm{diag}(X^k)-e).$$

2. Setting

$$\tilde{z}^k=z^k-\frac{1}{s}(\mathrm{diag}(\tilde{X}^k)-e).\quad (5.3)$$

子问题 (5.2) 求解的具体做法: 化为等价问题

$$\min\{\frac{1}{2}\|X-\frac{1}{1+r}[rX^k+\mathrm{diag}(z^{k+\frac{1}{2}})+C]\|_F^2|X\in S_+^n\}.$$

因此我们只要考虑如何求解

$$\tilde{X}^k=\text{Argmin}\{\frac{1}{2}\|X-A\|_F^2\mid X\in S_+^n\}.\quad (5.4)$$

实际上, 将对称矩阵 A 做标准特征值– 特征向量分解

$$A=V\Lambda V^T,\qquad \Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\ldots,\lambda_n).\quad (5.5)$$

通过

$$\tilde{X}=V\tilde{\Lambda}V^T,\qquad \tilde{\Lambda}=\mathrm{diag}(\tilde{\lambda}_1,\tilde{\lambda}_2,\ldots,\tilde{\lambda}_n),$$

就得到 $\tilde{X}$, 其中

$$\tilde{\lambda}_j=\max\{0,\lambda_j\}.$$

因此, 每次迭代的主要工作量是做 (5.5) 中的特征值 (特征向量) 分解 .

数值试验 为生成试验例子, 只要给定对称矩阵 C.

C=rand(n,n); C=(C'+C)-ones(n,n) + eye(n)

这样的矩阵 C 的对角元在 $(0,2)$ 之间, 非对角元在 $(-1,1)$ 之间。

Code 7.a. Matlab code for Creating the test examples

```
clear;  close all;
n = 1000;      tol=1e-5;     r=2.0;   s=1.05/r;    gamma=1.5;
rand('state',0);  C=rand(n,n);   C=(C'+C)-ones(n,n) + eye(n);
```

Code 7.1. Matlab code of ALM based classical PPA

```
%%%     Classical PPA for calibrating correlation matrix          %(1)
function PPAC(n,C,r,s,tol)                                        %(2)
X=eye(n);      y=zeros(n,1);    tic;      %%  The initial iterate %(3)
stopc=1;                      k=0;                                %(4)
while (stopc>tol && k<=100)         %% Beginning of an Iteration  %(5)
 if mod(k,20)==0 fprintf(' k=%4d   epsm=%9.3e  \n',k,stopc); end; %(6)
    X0=X;    y0=y;    k=k+1;                                      %(7)
    ya=y0 - (diag(X0)-ones(n,1))/s;                               %(8)
    A=(X0*r + C + diag(ya))/(1+r);                                %(9)
    [V,D]=eig(A);   D=max(0,D);  XT=(V*D)*V';     EX=X0-XT;      %(10)
    yt=y0 - (diag(XT)-ones(n,1))/s;               EY=y0-yt       %(11)
    ex=max(max(abs(EX)));  ey=max(abs(EY));   stopc=max(ex,ey);  %(12)
    X= XT;                 y=yt;                                 %(13)
end;                                         % End of an Iteration %(14)
toc;                       TB = max(abs(diag(X-eye(n))));        %(15)
fprintf(' k=%4d  epsm=%9.3e   max|X_jj - 1|=%8.5f \n',k,stopc,TB); %%
```

做 (5.5) 中的特征值 (特征向量) 分解, 在上述程序中的第 (10) 行用 Matlab 中的语句 [V,D]=eig(A) 实现的, 这是一个计算量大概 $9n^3$ 的运算.

将 Classical PPA 改成 Extended PPA, 只要将第 (13) 行改一下.

Code 7.2 Matlab code of ALM based extended PPA

```
%%%     Extended PPA for calibrating correlation matrix           %(1)
function PPAE(n,C,r,s,tol,gamma)                                  %(2)
X=eye(n);      y=zeros(n,1);    tic;      %%  The initial iterate %(3)
stopc=1;                      k=0;                                %(4)
while (stopc>tol && k<=100)         %% Beginning of an Iteration  %(5)
 if mod(k,20)==0 fprintf(' k=%4d   epsm=%9.3e  \n',k,stopc); end; %(6)
    X0=X;    y0=y;    k=k+1;                                      %(7)
    ya=y0 - (diag(X0)-ones(n,1))/s;                               %(8)
    A=(X0*r + C + diag(ya))/(1+r);                                %(9)
    [V,D]=eig(A);   D=max(0,D);  XT=(V*D)*V';     EX=X0-XT;      %(10)
    yt=y0 - (diag(XT)-ones(n,1))/s;               EY=y0-yt       %(11)
    ex=max(max(abs(EX)));  ey=max(abs(EY));   stopc=max(ex,ey);  %(12)
    X=X0 - EX*gamma;          y=y0- EY*gamma;                    %(13)
end;                                         % End of an Iteration %(14)
toc;                       TB = max(abs(diag(X-eye(n))));        %(15)
fprintf(' k=%4d  epsm=%9.3e   max|X_jj - 1|=%8.5f \n',k,stopc,TB); %%
```

两个不同的方法, 程序都很简单, 用不了几行. 两个程序不同的地方仅仅是第 (13) 行有些差别, 取 $\gamma = 1.5$ 的方法效果却有明显的提高。

矩阵校正问题 (5.1)–使用 Matlab 中的 eig 子程序

$n \times n$ Matrix	Classical PPA		Extended PPA	
$n =$	No. It	CPU Sec.	No. It	CPU Sec.
100	29	0.34	21	0.25
200	32	2.26	24	1.68
500	38	20.70	26	14.04
800	41	86.75	29	61.28
1000	47	182.82	30	117.08
2000	65	1696.50	41	1076.04

The extended PPA converges faster than the classical PPA.

$$\frac{\text{It. No. of Extended PPA}}{\text{It. No. of Classical PPA}} \approx 65\%.$$

♣ 关于相关系数矩阵校正的程序在附件的 Codes-07 的文件夹“矩阵校正”中. 只要运行 demo.m, 输入 n 就可以了. 其中的 ALM_PPAC.m 和 ALM_PPAE.m 分别是 Classical PPA 和 Extended PPA 的子程序.

确实, 用这一讲介绍的 PPA 方法求解相关矩阵矩阵校正问题, 每步迭代的主要计算工作量是对一个对称矩阵用 Matlab 中的标准子程序做 [V,D]=eig(A). 如果改用 Kim TOH 写的 mexeig 做 [V,D]=mexeig(A), 计算时间大为节省.

矩阵校正问题 (5.1)–特征值分解使用 mexeig

$n \times n$ Matrix	Classical PPA		Extended PPA	
$n =$	No. It	CPU Sec.	No. It	CPU Sec.
100	29	0.14	21	0.10
200	32	0.57	24	0.42
500	38	5.64	26	3.92
800	41	20.05	29	14.31
1000	47	41.82	30	27.10
2000	65	401.94	41	255.37

5.2 矩阵完整化方面的应用

设 M 是一个 $m \times n$ 矩阵, Π 是矩阵元素的指标集.

$$\Pi = \{(ij) \,|\, i \in \{1, 2, \ldots, m\},\ j \in \{1, 2, \ldots, n\}\}.$$

矩阵完整化问题是由部分信息获取全部信息. 在适当(实际问题具备的)条件下, 大多数不完整信息的低秩矩阵可以通过求解松弛问题

$$\min\{\|X\|_* \mid X_{ij} = M_{ij},\ (ij) \in \Pi\} \tag{5.6}$$

得到精确恢复. 其中 $\|X\|_*$ 表示矩阵 X 的奇异值的和. 通常称为矩阵 X 的核模—Nuclear Norm.

问题 (5.6) 是形如 (1.1) 的等式约束凸优化问题, 其中 $\|A^T A\| = 1$. 我们将 (5.6) 的等式约束记为 $X_\Pi = M_\Pi$，并用 $Z \in \Re^{m\times n}$ 作为相应的 Lagrange 乘子.

PPA 算法求解问题 (5.6)

对给定的 (X^k, Z^k), 用 (3.3) 产生$(\tilde{X}^k, \tilde{Z}^k)$:

1. Finding $\tilde{X}^k$ which is the solution of the following linear variational inequality

$$\min\ \{\|X\|_* + \frac{r}{2}\|X - [X^k + \frac{1}{r}Z_\Pi^{k+\frac{1}{2}}]\|_F^2\} \tag{5.7}$$

   where

$$Z_\Pi^{k+\frac{1}{2}} = Z_\Pi^k - \frac{1}{s}(X_\Pi^k - M_\Pi).$$

2. Updating $\tilde{Z}^k$ by

$$\tilde{Z}_\Pi^k = Z_\Pi^k - \frac{1}{s}(\tilde{X}_\Pi^k - M_\Pi).$$

子问题 (5.7) 求解的具体做法: 只需考虑如何求解

$$\tilde{X}^k = \text{Argmin}\{\frac{1}{r}\|X\|_* + \frac{1}{2}\|X - A\|_F^2\}. \tag{5.8}$$

我们将 A 做 SVD 分解

$$A = U\Lambda V^T,$$

并记

$$\tilde{X} = U\tilde{\Lambda}V^T. \tag{5.9}$$

注意到

$$\frac{1}{r}\|\tilde{X}\|_* + \frac{1}{2}\|\tilde{X} - A\|_F^2 = \frac{1}{r}\|\tilde{\Lambda}\|_* + \frac{1}{2}\|\tilde{\Lambda} - \Lambda\|_F^2.$$

通过

$$\tilde{\lambda}_j = \lambda_j - \min(\lambda_j, \frac{1}{r}), \tag{5.10}$$

就能得到对角矩阵 $\tilde{\Lambda}$ 的对角元 $\tilde{\lambda}_j$, 代入 (5.9) 就得到 (5.8) 的解 $\tilde{X}^k$. 同样, 每次迭代的主要工作量是做一个矩阵的 SVD 分解.

如果以 λ 和 $\tilde{\lambda}$ 分别表示对角矩阵 Λ 和 $\tilde{\Lambda}$ 的对角元生成的向量, 由于 λ 是非负向量, 关系式 (5.10) 也可以写成 Shrinkage 的形式

$$\tilde{\lambda} = \lambda - P_{B_\infty^{1/r}}[\lambda],$$

其中 $B_\infty^{1/r}$ 是无穷模下半径为 $1/r$ 的“圆”（一个立方体）.

结论: 将线性约束的凸优化问题转换成单调变分不等式, 再用这一讲 §3 中介绍的基于增广 Lagrange 乘子法的 PPA 算法求解, 每次迭代中要求解的子问题, 数值代数中都有确定的成熟的方法求解 !

数值试验 数值试验例子取自[2]

一个秩为 ra 的 $n \times n$ 的自由度是 $d_{ra} := ra(2n - ra)$.

生成试验问题:

- 先用高斯同分布 (Gaussian i.i.d) 独立生成两个 $n \times ra$ 的矩阵 M_1 和 M_2, 然后令 $M = M_1 M_2^T$, 则 $n \times n$ 矩阵 M 的秩为 ra.
- 随机选定 M 的 m 个元素作为已知元素, 这些元素的下标集为 Π.

计算结果:

- 矩阵完整化问题的难度与比率 m/d_{ra} 和 m/n^2 都有关系.
- 分别用 Classical PPA 和 Extended PPA ($\gamma = 1.5$) 进行计算.
- 正定矩阵 G (see (3.11)) 中的参数 $r,\ s$ 分别取 $rs = 1.01$ 和 $r = 0.005$.
- 停机准则采用相对误差 $\|X_\Pi^k - M_\Pi\|_F / \|M_\Pi\|_F \le 10^{-4}$.

我们用 $\gamma = 1.5$ 的Extended PPA 求解, 由于 KKT 条件 Primal 部分((3.5) 的上半

部分) 的不满足量是

$$(rI_n - \frac{1}{s}A^TA)(\tilde{x}^k - x^k).$$

对这个具体问题, $\|A^TA\| = 1$ 并且 $rs \approx 1$, 所以我们检查

$$\text{KKT-Violation} := \max\{r \max_{ij} |X_{ij}^k - \tilde{X}_{ij}^k|\}$$

并在计算结果中列出.

计算时间依赖于用什么 SVD 子程序. 我们分别列出用 Matlab 中的标准 SVD 以及 PROPACK-SVD 的不同计算效果.

♣ 矩阵完整化的程序在附件的 Codes-07 的文件夹"矩阵完整化"中. 只要运行 demo.m 就可以了. 要对不同情形试验, 只要在 demo.m 中用 % 做适当选择. 其中的 PPAC.m 和 PPAE.m 分别是 Classical PPA 和 Extended PPA 的子程序.

Code 7.b. Creating the test examples of Matrix Completion

```
%%  Creating the test examples of the matrix Completion problem     %(1)
clear all;  clc                                                     %(2)
maxIt=100;        tol = 1e-4;                                       %(3)
r=0.005;          s=1.01/r;        gamma=1.5;                       %(4)
  n=200;            ra = 10;         oversampling = 5;              %(5)
% n=1000;   ra=100;     oversampling = 3; %%  Iteration No. 31      %(6)
% n=1000;   ra=50;      oversampling = 4; %%  Iteration No. 36      %(7)
% n=1000;   ra=10;      oversampling = 6; %%  Iteration No. 78      %(8)
%% Generating the test problem                                      %(9)
rs = randseed;        randn('state',rs);                           %(10)
M=randn(n,ra)*randn(ra,n);         %% The matrix will be completed %(11)
df =ra*(n*2-ra);                   %% The freedom of the matrix    %(12)
mo=oversampling;                                                   %(13)
m =min(mo*df,round(.99*n*n));       %% No. of the known elements   %(14)
Omega= randsample(n^2,m);           %% Define the subset Omega     %(15)
fprintf('Matrix: n=%4d  Rank(M)=%3d  Oversampling=%2d \n',n,ra,mo);%(16)
```

我们只对ALM-based extended PPA 给出程序, 若用ALM-based classical PPA, 只要把下面的第(16) 行改成 `X = XT` 第(17) 行改成 `Y(Omega) = YT(Omega)`

Code 7.3. ALM-based extended PPA for Matrix Completion Problem

```
function PPAE(n,r,s,M,Omega,maxIt,tol,gamma)    %  Ititial Process %%(1)
X=zeros(n);     Y=zeros(n);      YT=zeros(n);                        %(2)
  nM0=norm(M(Omega),'fro');    eps=1;  VioKKT=1;  k=0;   tic;       %(3)
%% Minimum nuclear norm solution by PPA method                      %(4)
while  (eps > tol && k<= maxIt)                                     %(5)
 if mod(k,5)==0                                                     %(6)
 fprintf('It=%3d |X-M|/|M|=%9.2e VioKKT=%9.2e\n',k,eps,VioKKT); end;%(7)
  k=k+1;      X0=X;      Y0=Y;                                      %(8)
  Y(Omega)=Y0(Omega)-(X0(Omega)-M(Omega))/s;                         %(9)
  A = X0 + Y/r;                       [U,D,V]=svd(A,0);             %(10)
  D=D-eye(n)/r;     D=max(D,0);      XT=(U*D)*V';        EX=X0-XT;  %(11)
  DXM=XT(Omega)-M(Omega);            eps = norm(DXM,'fro')/nM0;     %(12)
  YT(Omega)=Y0(Omega)-(XT(Omega)-M(Omega))/s;           EY=Y0-YT;   %(13)
  VioKKT =  max(max(abs(EX)))*r;                                    %(14)
   if (eps <= tol)   gamma=1;   end;                                %(15)
  X = X0 - EX*gamma;                                                %(16)
  Y(Omega) = Y0(Omega) - EY(Omega)*gamma;                           %(17)
end;                                                                %(18)
 fprintf('It=%3d  |X-M|/|M|=%9.2e  Vi0KKT=%9.2e \n',k,eps,VioKKT); %(19)
   RelEr=norm((X-M),'fro')/norm(M,'fro');   toc;                    %(20)
 fprintf(' Relative error = %9.2e  Rank(X)=%3d \n',RelEr,rank(X)); %(21)
 fprintf(' Violation of KKT Condition = %9.2e  \n',VioKKT);         %(22)
```


矩阵完整化问题: 用 Matlab-SVD 求解结果

1000×1000 matrix	ALM-based Classical PPA				ALM-based Extended PPA			
(rank, $\frac{m}{d_r}$, $\frac{m}{n^2}$)	No. It.	CPU Sec.	Relative error	KKT-Violation	No. It.	CPU Sec.	Relative error	KKT-Violation
(10, 6, 0.12)	85	989.84	9.62E-5	4.34E-6	77	899.40	9.32E-5	2.79E-6
(50, 4, 0.39)	43	491.35	1.46E-4	1.85E-5	37	425.24	1.21E-4	1.43E-5
(100, 3, 0.58)	36	394.68	1.72E-4	2.78E-5	31	363.04	1.55E-4	3.26E-5

♣ 用 Matlab 中的 SVD, 做一次 SVD 的花费很大, 总耗时与迭代次数成比例.

♣ 用 Extended PPA 与 Classical PPA 相比, 几乎不加额外的负担, 效率还是提高 10% 以上. 因此, 在任何情况下, 我们都提倡用 Extended PPA.

♣ 用 PROPACK [9] 中的 SVD, 快许多, 总耗时主要与问题性质有关. 由于我们主要对迭代次数感兴趣, 我们仅仅报道用 Matlab 做 SVD 的结果, 也附上所需要的子程序. 总耗时主要与问题性质有关.

♣ 注意到, Cai, Candès and Shen [2] 的方法对这三个例子的迭代次数分别是 117, 114 和129 (See the first three examples in Table 5.1 of [2], pp. 1974). 原则上, 每次迭代的主要工作量都是一次 SVD 分解. [2] 中采用不完全分解技术, 节省了总的运行时间.

一点说明 从第六讲到第八讲, 我们主要讲一类 Customized PPA 方法, 求解

$$\min\{\theta(x) \mid Ax = b, x \in \mathcal{X}\} \tag{5.11}$$

这样一类凸优化问题. 用来计算的例子, 都有应用背景. 用 Customized PPA 方法求解, 有很不错的数值结果. 主要原因是这些问题中的线性约束的矩阵 A 的性态比较好 (注意到 $\mathrm{diag}(X) = e$ 和 $X_\Pi = M_\Pi$ 表示成 $Ax = b$ 时, A 只是一个投影矩阵), 加上这些问题中

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

这样的子问题是有成熟方法求解的.

但是, 正如 R. Fletcher 在他的著作 Practical Methods of Optimization 中说到,

Indeed there is no general agreement on the best approach and much research is still to be done.

世界上没有一个方法是对所有的问题都是最好的, 我们不能指望用这几讲介绍的方法去求解形如 (5.11) 的一般凸优化问题, 特别是人为构造的难题. 好在一些有应用背景的问题, 常常有它的特殊结构, 简单的方法有时也能凑效.

References

[1] D. P. Bertsekas. Constrained Optimization and Lagrange Multiplier Methods. Academic Press, 1982.

[2] J. F. Cai, E. J. Candès and Z. W. Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optim., **20**, 1956-1982, 2010.

[3] B.S. He, A class of projection and contraction methods for monotone variational inequalities, Applied Mathematics and Optimization, **35**, 69-76, 1997.

[4] B.S. He, Inexact implicit methods for monotone general variational inequalities, Math. Program. Series A **86**, 199-217, 1999.

[5] B. S. He, X. L. Fu and Z.K. Jiang, Proximal point algorithm using a linear proximal term, JOTA, **141**, 209-239, 2009.

[6] B.S He and L-Z Liao, Improvements of some projection methods for monotone nonlinear variational inequalities, JOTA, **112**, 111-128, 2002

[7] B. S. He and M.-H. Xu, A general framework of contraction methods for monotone variational inequalities, Pacific J. Optimization, **4**, 195-212, 2008.

[8] M. R. Hestenes, Multiplier and gradient methods. Journal of Optimization Theory and Applications,**4**, 302-320, 1969.

[9] R. M. Larsen, Propack − software for large and sparse svd calculation, http://sun.stanford.edu/rmunk/PROPACK/, 2005.

[10] J. Nocedal and S.J. Wright, Numerical Optimization, Springer Verlag, 1999.

[11] M. J. D. Powell. A method for nonlinear constraints in minimization problems. In R. Fletcher, editor, Optimization. Academic Press, 1969.

凸优化和单调变分不等式的收缩算法

第九讲: 基于线性变分不等式 PC 方法的求解复合凸优化问题的收缩算法

Contraction methods for composite convex optimization based on the PC Algorithms for LVIs

南方科技大学数学系　何炳生
hebma@nju.edu.cn

摘要. 上世纪九十年代, 我们发表了一些求解单调线性变分不等式(及凸二次优化)的投影收缩算法. 这些算法被他人用来求解一些应用问题有很好的表现. 在处理一类机器人控制问题上, 起到了其他算法不能替代的作用[3, 11, 12]. 本文介绍如何将这些方法推广, 用来求解复合凸优化问题.

1 Introduction

In the 1990s, we have published some projection and contraction algorithms for solving monotone linear variational inequalities [5, 6]. These algorithms can be applied to solve the constrained convex optimization problems

$$\min\{\frac{1}{2}x^T Hx + c^T x \,|\, x \in \mathcal{X}\} \tag{1.1}$$

and

$$\min\{\frac{1}{2}x^T Hx + c^T x \,|\, Ax = b(\text{or} \geq b), x \in \mathcal{X}\}, \tag{1.2}$$

where $H \in \Re^{n\times n}$ is a symmetric positive semi-definite matrix, $A \in \Re^{m\times n}$, $b \in \Re^m$, $c \in \Re^n$ and $\mathcal{X} \subset \Re^n$ is a closed convex set. The purpose of this

article is to develop such algorithms to solve the following composite convex optimization problems:

$$\min\{\theta(x) + \frac{1}{2}x^T Hx + c^T x \,|\, x \in \mathcal{X}\} \tag{1.3}$$

and

$$\min\{\theta(x) + \frac{1}{2}x^T Hx + c^T x \,|\, Ax = b(\text{or} \geq b), x \in \mathcal{X}\}, \tag{1.4}$$

where $\theta(x) : \Re^n \to \Re$ is a convex function (not necessarily smooth), $H,\ A,\ b,\ c$ and $\mathcal{X}$ are the same as described in (1.1) and (1.2).

Throughout this article, we assume that the solution set (1.3) and (1.4) are nonempty. In addition, we assume that for any given constant $r > 0$ and vector $a \in \Re^n$, the subproblem

$$\min\{\theta(x) + \frac{r}{2}\|x - a\|^2 \mid x \in \mathcal{X}\} \tag{1.5}$$

has a closed-form solution or can be efficiently computed with a high precision.

The analysis of this note is based on the following lemma (proof is omitted here).

Lemma 1.1 *Let $\mathcal{X} \subset \Re^n$ be a closed convex set, $\theta(x)$ and $f(x)$ be convex functions and $f(x)$ is differentiable on an open set which includes Ω. Assume that the solution set of the minimization problem $\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\}$ is nonempty. Then,*

$$x^* \in \arg\min\{\theta(x) + f(x) \,|\, x \in \mathcal{X}\} \tag{1.6a}$$

if and only if

$$x^* \in \mathcal{X}, \ \ \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \geq 0, \ \ \forall\, x \in \mathcal{X}. \tag{1.6b}$$

2 The algorithms for solving (1.3) based on P-C Algorithm for the problem (1.1)

In (1.3), since the $\theta(x)$ is convex and H is semidefinite, by using Lemma 1.1, the optimal solution of (1.3), say x^*, satisfies

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^T (Hx^* + c) \geq 0, \quad \forall\, x \in \mathcal{X}. \tag{2.1}$$

2.1 Key inequality for solving $\min\{\frac{1}{2}x^T Hx + c^T x \,|\, x \in \mathcal{X}\}$

Set $\theta(x) = 0$ in (1.3), it is reduced to the problem (1.1) whose optimal solution x^* satisfies

$$x^* \in \mathcal{X}, \quad (x - x^*)^T (Hx^* + c) \geq 0, \quad \forall\, x \in \mathcal{X}. \tag{2.2}$$

For solving (1.1) (or its equivalent (2.2)), we have proposed a class of projection and contraction algorithms [5, 6]. These algorithms are based on constructing the descent direction of the distance function $\frac{1}{2}\|x - x^*\|_G^2$, where G is some symmetric positive definite matrix.

For given $x^k \in \Re^n$ and $\beta > 0$, let

$$\tilde{x}^k = P_{\mathcal{X}}[x^k - \beta(Hx^k + c)]. \tag{2.3}$$

x^k is the optimal solution of (1.1) (or its equivalent (2.2)) if and only if $\tilde{x}^k = x^k$.

The projector $\tilde{x}^k$ is the solution of the minimization problem,

$$\tilde{x}^k = \arg\min\{\frac{1}{2}\|x - [x^k - \beta(Hx^k + c)]\|^2 \,|\, x \in \mathcal{X}\}.$$

According to Lemma 1.1, we have

$$\tilde{x}^k \in \mathcal{X}, \quad (x - \tilde{x}^k)^T \{\tilde{x}^k - [x^k - \beta(Hx^k + c)]\} \geq 0, \quad \forall x \in \mathcal{X}.$$

Set the any vector $x \in \mathcal{X}$ in the above inequality by a solution point x^*, it follows that

$$(\tilde{x}^k - x^*)^T \{(x^k - \tilde{x}^k) - \beta(Hx^k + c)\} \geq 0. \tag{2.4}$$

On the other hand, since $\tilde{x}^k \in \mathcal{X}$, it follows from (2.2) that

$$(\tilde{x}^k - x^*)^T \beta(Hx^* + c) \geq 0. \tag{2.5}$$

Adding (2.4) and (2.5), we get

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

The above inequality can be rewritten as

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

Finally, by using the semi-positiveness of H, we get

$$(x^k - x^*)^T(I + \beta H)(x^k - \tilde{x}^k) \geq \|x^k - \tilde{x}^k\|^2. \tag{2.6}$$

The above inequality is the main basis for building the projection contraction algorithms for solving (1.1) (and its equivalent variational inequality (2.2)). We hope to establish the same inequality for the problem (1.3).

2.2 Key inequality for solving $\min\{\theta(x)+\frac{1}{2}x^THx+c^Tx|x\in\mathcal{X}\}$

Our task is to solve (1.3). The purpose of this subsection is to construct the same key-inequality as (2.6). For given x^k and $\beta > 0$, we let

$$\tilde{x}^k = \arg\min\{\theta(x) + \frac{1}{2\beta}\|x - [x^k - \beta(Hx^k + c)]\|^2 \,|\, x \in \mathcal{X}\}. \tag{2.7}$$

This is an optimization problem as (1.5) and it is assumed to be solved without difficulty. According to Lemma 1.1, we have $\tilde{x}^k \in \mathcal{X}$ and

$$\theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T\frac{1}{\beta}\{\tilde{x}^k - [x^k - \beta(Hx^k + c)]\} \geq 0, \ \forall x \in \mathcal{X}.$$

We rewrite it as $\tilde{x}^k \in \mathcal{X}$ and

$$\beta\theta(x) - \beta\theta(\tilde{x}^k) + (x - \tilde{x}^k)^T\{\beta(Hx^k + c) - (x^k - \tilde{x}^k)\} \geq 0, \ \forall x \in \mathcal{X}. \tag{2.8}$$

Since $\tilde{x}^k \in \mathcal{X}$, according to the optimal condition (2.1), we have

$$\beta\theta(\tilde{x}^k) - \beta\theta(x^*) + (\tilde{x}^k - x^*)^T\beta(Hx^* + c) \geq 0. \tag{2.9}$$

Setting the any $x \in \Omega$ in (2.8) by x^* and then adding it with (2.9), we obtain

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

Rewriting the above inequality in form

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

and using the semi-positivity of H, we get

$$(x^k - x^*)^T (I + \beta H)^T (x^k - \tilde{x}^k) \geq \|x^k - \tilde{x}^k\|^2. \tag{2.10}$$

This is the same inequality as (2.6). Usually, we call the vector $\tilde{x}^k$ obtained by (2.7) the predictor in the k-th iteration of the proposed algorithm for solving the convex optimization problem (1.3).

2.3 Solving the optimization problem (1.3) by using the key inequality (2.10)

The inequality (2.10) can be written as

$$\langle (I + \beta H)(x^k - x^*), (x^k - \tilde{x}^k) \rangle \geq \|x^k - \tilde{x}^k\|^2.$$

Let $G = (I + \beta H)$, the above inequality tells us that $-(x^k - \tilde{x}^k)$ is the descent direction of the unknown distance function $\frac{1}{2}\|x - x^*\|_G^2$ at x^k.

We take

$$x^{k+1}(\alpha) = x^k - \alpha(x^k - \tilde{x}^k) \tag{2.11}$$

as the step length α dependent new iterate. In order to shorten the distance $\|x - x^*\|^2_{(I+\beta H)}$, we consider the following α-dependent benefit

$$\vartheta_k(\alpha) = \|x^k - x^*\|^2_{(I+\beta H)} - \|x^{k+1}(\alpha) - x^*\|^2_{(I+\beta H)}. \tag{2.12}$$

By using (2.10), we obtain

$$\begin{aligned} \vartheta_k(\alpha) &= \|x^k - x^*\|^2_{(I+\beta H)} - \|x^k - x^* - \alpha(x^k - \tilde{x}^k)\|^2_{(I+\beta H)} \\ &= 2\alpha(x^k - x^*)^T (I + \beta H)(x^k - \tilde{x}^k) \\ &\quad -\alpha^2 \|x^k - \tilde{x}^k\|^2_{(I+\beta H)}. \end{aligned} \tag{2.13}$$

By using the (2.10), we have the following theorem.

Theorem 2.1 *For given x^k and any $\beta > 0$, let $\tilde{x}^k$ be a predictor generated by* (2.7) *and $x^{k+1}(\alpha)$ be updated by* (2.11)*. Then for any $\alpha > 0$, we have*

$$\vartheta_k(\alpha) \geq q_k(\alpha), \tag{2.14}$$

where $\vartheta_k(\alpha)$ *is defined by* (2.12) *and*

$$q_k(\alpha) = 2\alpha\|x^k - \tilde{x}^k\|^2 - \alpha^2\|x^k - \tilde{x}^k\|^2_{(I+\beta H)}. \tag{2.15}$$

Proof. The assertion of this theorem follows directly from (2.13) and (2.10). □

Now, $q_k(\alpha)$ is a lower bound function of $\vartheta_k(\alpha)$. The quadratic function $q_k(\alpha)$ reaches its maximum at

$$\alpha_k^* = \frac{\|x^k - \tilde{x}^k\|^2}{(x^k - \tilde{x}^k)^T(I+\beta H)(x^k - \tilde{x}^k)}. \tag{2.16}$$

We use

$$x^{k+1} = x^k - \gamma\alpha_k^*(x^k - \tilde{x}^k), \quad \gamma \in (0,2) \tag{2.17}$$

to determine the new iterate x^{k+1}. By using (2.15) and (2.16), we get

$$\begin{aligned} q(\gamma\alpha_k^*) &= 2\gamma\alpha_k^*\|x^k - \tilde{x}^k\|^2 - \gamma^2\alpha_k^*\big(\alpha_k^*\|x^k - \tilde{x}^k\|^2_{(I+\beta H)}\big) \\ &= \gamma(2-\gamma)\alpha_k^*\|x^k - \tilde{x}^k\|^2. \end{aligned} \tag{2.18}$$

Thus, the profit of the k-th iteration

$$\begin{aligned} \vartheta(\gamma\alpha_k^*) &= \|x^k - x^*\|^2_{(I+\beta H)} - \|x^{k+1} - x^*\|^2_{(I+\beta H)} \\ &\geq q(\gamma\alpha_k^*) = \gamma(2-\gamma)\alpha_k^*\|x^k - \tilde{x}^k\|^2. \end{aligned} \tag{2.19}$$

Theoretically $\vartheta(\gamma\alpha_k^*) > 0$ when $\gamma \in (0,2)$, usually, we take $\gamma \in [1.2, 1.8]$.

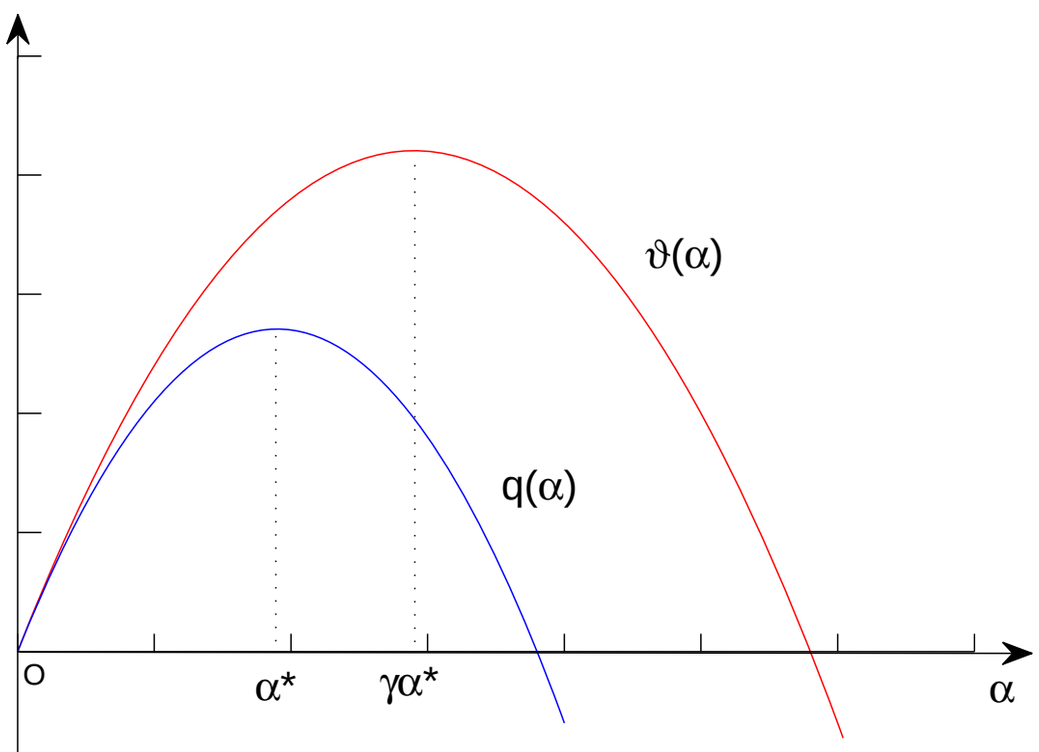


Fig. 1 It is suggested to take $\gamma \in [1.2, 1.8]$

For solving the composite convex optimization problem (1.3), we use (2.7) to generate the predictor $\tilde{x}^k$ and (2.17) to update the corrector x^{k+1}. Together with some strategy for adjusting the parameter β, we have the following algorithm.

Algorithm 2.1 Prediction-Correction method for solving the composite convex optimization problem (1.3):
Start with given x^0 and $\beta > 0$. For $k = 0, 1, \ldots$, do:

1. Prediction: $\tilde{x}^k = \arg\min\{\theta(x) + \frac{1}{2\beta}\|x - [x^k - \beta(Hx^k + c)]\|^2 \mid x \in \mathcal{X}\}$.
2. Correction: $x^{k+1} = x^k - \gamma\alpha_k^*(x^k - \tilde{x}^k)$,
   where $\alpha_k^* = \dfrac{\|x^k - \tilde{x}^k\|^2}{(x^k - \tilde{x}^k)^T(I + \beta H)(x^k - \tilde{x}^k)}$ and $\gamma \in (0, 2)$.
3. Adjust the parameter β if necessary
   $$r = \frac{\beta\|x^k - \tilde{x}^k\|_H^2}{\|x^k - \tilde{x}^k\|^2}, \qquad \beta = \begin{cases} (\beta/r) * 0.9, & \text{if } r > 1 \text{ or } r < 0.4; \\ \beta, & \text{otherwise.} \end{cases}$$
   $k := k + 1$ and go to 1.

According to our numerical experiments, the parameter β should be selected in range

$$\tfrac{2}{5}\|x^k - \tilde{x}^k\| \le \beta\|x^k - \tilde{x}^k\|_H^2 \le \|x^k - \tilde{x}^k\|.$$

We can also adjust the parameters every five or ten iterations. In practical, after some iterations, the algorithm will automatically find a suitable fixed β.

Theorem 2.2 *For solving the composite convex optimization problem* (1.3)*, the sequence* $\{x^k\}$ *generated by Algorithm 2.1 with fixed* $\beta > 0$ *satisfies*

$$\|x^{k+1} - x^*\|_{(I+\beta H)}^2 \le \|x^k - x^*\|_{(I+\beta H)}^2 - \frac{\gamma(2-\gamma)}{\|I + \beta H\|}\|x^k - \tilde{x}^k\|^2. \quad (2.20)$$

Proof. From (2.16) we have $\alpha_k^* \ge \frac{1}{\|I+\beta H\|}$. Together with (2.19), it follows the assertion (2.20) directly. $\square$

Generally speaking, whether the selection of parameters is appropriate or not, it will affect the convergence speed. Without loss of the generality, Theorem 2.2 gives us the key inequality for the convergence of Algorithm 2.1.

3 The algorithms for solving (1.4) based on P-C Algorithm for the problem (1.2)

The Lagrangian function of the linearly constrained optimization problem (1.4) is

$$L(x,\lambda) = \theta(x) + \frac{1}{2}x^T Hx + c^T x - \lambda^T(Ax - b),$$

which is defined on $\mathcal{X} \times \Lambda$ and where

$$\Lambda = \begin{cases} \Re^m, & \text{if the lineary constraints in (1.4) is } Ax = b, \\ \Re^m_+, & \text{if the lineary constraints in (1.4) is } Ax \geq b. \end{cases}$$

Let $(x^*, \lambda^*) \in \mathcal{X} \times \Lambda$ be a saddle point of the Lagrangian function, we have

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

This tells us that $(x^*, \lambda^*) \in \mathcal{X} \times \Lambda$ and

$$\begin{cases} L(x, \lambda^*) - L(x^*, \lambda^*) \geq 0, & \forall\, x \in \mathcal{X}, \\ L(x^*, \lambda^*) - L(x^*, \lambda) \geq 0, & \forall\, \lambda \in \Lambda. \end{cases}$$

Thus, finding a saddle point of the Lagrange function is equivalent to finding $(x^*, \lambda^*) \in \mathcal{X} \times \Lambda$ such that

$$\begin{cases} \theta(x) - \theta(x^*) + (x - x^*)^T(Hx^* + c - A^T\lambda^*) \geq 0, & \forall\, x \in \mathcal{X}, \\ (\lambda - \lambda^*)^T(Ax^* - b) \geq 0, & \forall\, \lambda \in \Lambda. \end{cases} \tag{3.1}$$

The above variational inequality can be written in a compact form

$$u^* \in \Omega, \quad \theta(x) - \theta(x^*) + (u - u^*)^T(Mu^* + q) \geq 0, \quad \forall\, u \in \Omega, \tag{3.2a}$$

where

$$u = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad M = \begin{pmatrix} H & -A^T \\ A & 0 \end{pmatrix}, \quad q = \begin{pmatrix} c \\ -b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \Lambda. \tag{3.2b}$$

Although the matrix M is not symmetric, however, for any u, we have $u^T M u = x^T H x \geq 0$. The variational inequality (3.2) is monotone.

3.1 Key inequality for solving the minimization problem $\min\{\frac{1}{2}x^T Hx + c^T x \mid Ax = (\text{or } \geq)\, b, x \in \mathcal{X}\}$

Set $\theta(x) = 0$ (1.4), the problem is reduced to the linear constrained quadratic programming (1.2), the related variational inequality (3.2) is reduced to

$$u^* \in \Omega, \quad (u - u^*)^T (Mu^* + q) \geq 0, \quad \forall\, u \in \Omega, \tag{3.3a}$$

where

$$u = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad M = \begin{pmatrix} H & -A^T \\ A & 0 \end{pmatrix}, \quad q = \begin{pmatrix} c \\ -b \end{pmatrix} \text{ and } \Omega = \mathcal{X} \times \Lambda. \tag{3.3b}$$

For solving (1.2) (or its equivalent (3.3)), we have proposed a projection and contraction method [5] whose search direction is established in the following way.

First, for given $u^k \in \Re^{m+n}$ and $\beta > 0$, let

$$\tilde{u}^k = P_\Omega[u^k - \beta(Mu^k + q)]. \tag{3.4}$$

Since

$$\tilde{u}^k = \arg\min\{\frac{1}{2}\|u - [u^k - \beta(Mu^k + q)]\|^2 | u \in \Omega\},$$

according to Lemma 1.1, it follows that

$$\tilde{u}^k \in \Omega, \quad (u - \tilde{u}^k)^T \{\tilde{u}^k - [u^k - \beta(Mu^k + q)]\} \geq 0, \quad \forall u \in \Omega.$$

Setting the any $u \in \Omega$ in the above inequality by u^*, we get

$$(\tilde{u}^k - u^*)^T \{(u^k - \tilde{u}^k) - \beta(Mu^k + q)\} \geq 0. \tag{3.5}$$

Because $\tilde{u}^k \in \Omega$ and $\beta > 0$, according to (3.3a), we have

$$(\tilde{u}^k - u^*)^T \beta(Mu^* + q) \geq 0. \tag{3.6}$$

Adding the above two inequalities, we get

$$(\tilde{u}^k - u^*)^T \{(u^k - \tilde{u}^k) - \beta M(u^k - u^*)\} \geq 0. \tag{3.7}$$

It can be written as

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

Consequently, using the semi-positivity of M ($v^T M v \geq 0$), it follows that

$$(u^k - u^*)^T (I + \beta M^T)(u^k - \tilde{u}^k) \geq \|u^k - \tilde{u}^k\|^2, \quad \forall u^* \in \Omega^*. \tag{3.8}$$

The above inequality is the main basis for building the projection and contraction algorithms for solving the linearly constrained convex quadratic optimization (1.2) (and its equivalent variational inequality (3.2)). We hope to establish the same inequality for the composite optimization problem (1.4).

3.2 Key inequality for solving the minimization problem $\min\{\theta(x) + \frac{1}{2}x^T H x + c^T x \mid Ax = (\text{or } \geq)\, b, x \in \mathcal{X}\}$

Our task is to solve the problem (1.4). This subsection will construct the same key-inequality as (3.8). For given $u^k = (x^k, \lambda^k)$ and $\beta > 0$, let

$$\tilde{x}^k = \arg\min\{\theta(x) + \frac{1}{2\beta}\|x - [x^k - \beta(Hx^k + c - A^T\lambda^k)]\|^2 \mid x \in \mathcal{X}\} \tag{3.9a}$$

and

$$\tilde{\lambda}^k = \arg\min\{\frac{1}{2}\|\lambda - [\lambda^k - \beta(Ax^k - b)] \mid \lambda \in \Lambda\}, \tag{3.9b}$$

Since (3.9a) is a convex optimization problem as (1.5) which has assumed to have closed form solution. Thus, there is no difficulty to obtain $\tilde{x}^k$ and $\tilde{\lambda}^k$ in parallel. According to Lemma 1.1, we have $\tilde{x}^k \in \mathcal{X}$ and

$$\theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T \frac{1}{\beta}\{\tilde{x}^k - [x^k - \beta(Hx^k + c - A^T\lambda^k)]\} \geq 0, \ \forall x \in \mathcal{X}.$$

Taking a solution point x^* as the any point $x \in \mathcal{X}$, it follows that

$$\beta\theta(x^*) - \beta\theta(\tilde{x}^k) + (\tilde{x}^k - x^*)^T\{(x^k - \tilde{x}^k) - \beta(Hx^k + c - A^T\lambda^k)\} \geq 0. \tag{3.10}$$

On the other hand, since $\tilde{x}^k \in \mathcal{X}$, it follows from (3.1) that

$$\beta\theta(\tilde{x}^k) - \beta\theta(x^*) + (\tilde{x}^k - x^*)^T \beta\{Hx^* + c - A^T\lambda^*\} \geq 0, \ \forall\, x^* \in \mathcal{X}^*. \tag{3.11}$$

Adding (3.10) and (3.11), we obtain

$$(\tilde{x}^k - x^*)^T\{(x^k - \tilde{x}^k) - \beta[H(x^k - x^*) - A^T(\lambda^k - \lambda^*)]\} \geq 0. \quad (3.12)$$

For (3.9b), according to Lemma 1.1, we have

$$\tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T\big\{\tilde{\lambda}^k - [\lambda^k - \beta(Ax^k - b)]\big\} \geq 0, \quad \forall\, \lambda \in \Lambda.$$

Set the fixed any point $\lambda \in \Lambda$ in the last inequality by λ^*, it follows that

$$(\tilde{\lambda}^k - \lambda^*)^T\{(\lambda^k - \tilde{\lambda}^k) - \beta(Ax^k - b)\} \geq 0. \quad (3.13)$$

On the other hand, since $\tilde{\lambda}^k \in \Lambda$, according the second part of (3.1), we have

$$(\tilde{\lambda}^k - \lambda^*)^T\beta(Ax^* - b) \geq 0. \quad (3.14)$$

Adding (3.13) and (3.14), it follows that

$$(\tilde{\lambda}^k - \lambda^*)^T\{(\lambda^k - \tilde{\lambda}^k) - \beta A(x^k - x^*)\} \geq 0. \quad (3.15)$$

Write (3.12) and (3.15) together,

$$\begin{pmatrix} \tilde{x}^k - x^* \\ \tilde{\lambda}^k - \lambda^* \end{pmatrix}^T \left\{ \begin{pmatrix} x^k - \tilde{x}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix} - \beta \begin{bmatrix} H & -A^T \\ A & 0 \end{bmatrix} \begin{pmatrix} x^k - x^* \\ \lambda^k - \lambda^* \end{pmatrix} \right\} \geq 0.$$

Using the notation in (3.3), it can be written as

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

Clearly, it is same inequality as (3.7) in §3.1. It can be written as

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

Consequently, using the semi-positivity of M ($v^T M v \geq 0$), it follows that

$$(u^k - u^*)^T(I + \beta M^T)(u^k - \tilde{u}^k) \geq \|u^k - \tilde{u}^k\|^2, \quad \forall u^* \in \Omega^*. \quad (3.16)$$

This is the same key inequality as (3.8). For solving the linearly constrained composite convex optimization problem (1.4), we call the vector $\tilde{u}^k = (\tilde{x}^k, \tilde{\lambda}^k)$ obtained from (3.9) a predictor in the k-iteration.

3.3 Solving the optimization problem (1.4) by using the key inequality (3.16)

The task of the contraction method is to generate the new iterate which is more closed to the solution set. Thus, we use

$$u^{k+1}(\alpha) = u^k - \alpha(I + \beta M^T)(u^k - \tilde{u}^k) \tag{3.17}$$

to update the new iterate which is dependent on the step length α.

In the following we discuss how to chose the step length α. For this purpose, we consider the following α-dependent benefit function

$$\vartheta_k(\alpha) := \|u^k - u^*\|^2 - \|u^{k+1}(\alpha) - u^*\|^2. \tag{3.18}$$

According to the definition,

$$\begin{aligned} \vartheta_k(\alpha) &= \|u^k - u^*\|^2 - \|u^k - u^* - \alpha(I + \beta M^T)(u^k - \tilde{u}^k)\|^2 \\ &= 2\alpha(u^k - u^*)^T(I + \beta M^T)(u^k - \tilde{u}^k) \\ &\quad -\alpha^2\|(I + \beta M^T)(u^k - \tilde{u}^k)\|^2. \end{aligned} \tag{3.19}$$

Theorem 3.1 *For given u^k and any $\beta > 0$, let $\tilde{u}^k$ be a predictor generated by* (3.9) *and $u^{k+1}(\alpha)$ be updated by* (3.17)*. Then for any $\alpha > 0$, we have*

$$\vartheta_k(\alpha) \geq q_k(\alpha), \tag{3.20}$$

where $\vartheta_k(\alpha)$ is defined by (3.18) *and*

$$q_k(\alpha) = 2\alpha\|u^k - \tilde{u}^k\|^2 - \alpha^2\|(I + \beta M^T)(u^k - \tilde{u}^k)\|^2. \tag{3.21}$$

Proof. The assertion of this theorem follows from (3.19) and (3.16) directly. □

Theorem 3.1 tells us $q_k(\alpha)$ is a lower bound function of the profit function $\vartheta_k(\alpha)$. The quadratic function $q_k(\alpha)$ (3.21) reaches its maximum at

$$\alpha_k^* = \text{argmax}\{q_k(\alpha)\} = \frac{\|u^k - \tilde{u}^k\|^2}{\|(I + \beta M^T)(u^k - \tilde{u}^k)\|^2}. \tag{3.22}$$

Notice that our intention is to maximize the quadratic profit function $\vartheta_k(\alpha)$ (see (3.19)), which includes the unknown solution u^*. As a remedy, we maximize its lower bound function $q_k(\alpha)$.

In the practical computation, we take a $\gamma \in [1, 2)$ and

$$u^{k+1} = u^k - \gamma\alpha_k^*(I + \beta M^T)(u^k - \tilde{u}^k), \tag{3.23}$$

to produce the new iterate. The reason to take $\gamma \in [1, 2)$ is illustrated in Fig. 1 of §2.3. Using (3.18) and (3.20), the new iterate u^{k+1} updated by (3.23) satisfies

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - q_k(\gamma\alpha_k^*). \tag{3.24}$$

According to the definitions of $q_k(\alpha)$ and α_k^* (see (3.21) and (3.22)), we get

$$\begin{aligned} q_k(\gamma\alpha_k^*) &= 2\gamma\alpha_k^*\|u^k - \tilde{u}^k\|^2 - \gamma^2\alpha_k^*(\alpha_k^*\|(I + \beta M^T)(u^k - \tilde{u}^k)\|^2) \\ &= 2\gamma\alpha_k^*\|u^k - \tilde{u}^k\|^2 - \gamma^2\alpha_k^*\|u^k - \tilde{u}^k\|^2 \\ &= \gamma(2 - \gamma)\alpha_k^*\|u^k - \tilde{u}^k\|^2. \end{aligned}$$

Thus, from (3.24) we get

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \gamma(2 - \gamma)\alpha_k^*\|u^k - \tilde{u}^k\|^2. \tag{3.25}$$

Algorithm 3.1 Prediction-Correction method for solving the composite convex optimization problem (1.4):

Start with given $u^0 = (x^0, \lambda^0)$ and $\beta > 0$. For $k = 0, 1, \ldots$, do:

1. Prediction:
   $\tilde{x}^k = \arg\min\{\theta(x) + \frac{1}{2\beta}\|x - [x^k - \beta(Hx^k + c - A^T\lambda^k)]\|^2 \,|\, x \in \mathcal{X}\}$
   and $\tilde{\lambda}^k = \arg\min\{\frac{1}{2}\|\lambda - [\lambda^k - \beta(Ax^k - b)] \,|\, \lambda \in \Lambda\}$.
2. Correction: $u^{k+1} = u^k - \gamma\alpha_k^*(I + \beta M^T)(u^k - \tilde{u}^k)$,
   where $\alpha_k^* = \dfrac{\|u^k - \tilde{u}^k\|^2}{\|(I + \beta M^T)(u^k - \tilde{u}^k)\|^2}$. and $\gamma \in (0, 2)$.
3. Adjust the parameter β if necessary

$$r = \frac{\|(I + \beta M^T)(u^k - \tilde{u}^k)\|}{\|u^k - \tilde{u}^k\|}, \quad \beta = \begin{cases} \beta * (2.5/r) & \text{if } r > 3 \\ \beta * (2.5/r) & \text{if } r < 2 \\ \beta & \text{otherwise.} \end{cases}$$

$k := k + 1$ and go to 1.

For solving the problem (1.4), the k-th iteration of the Algorithm3.1 start from u^k, produces the predictor $\tilde{u}^k$ by (3.9), and updated the new iterate u^{k+1} by (3.23). For this algorithm, we have the following theorem.

Theorem 3.2 *Let $\{u^k\}$ be the sequence generated by the Algorithm3.1 for the problem* (1.4)*. Then we have*

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \frac{\gamma(2-\gamma)}{\|I + \beta M^T\|^2}\|u^k - \tilde{u}^k\|^2. \qquad (3.26)$$

Proof. In fact, by using (3.22), we get

$$\alpha_k^* \ge \frac{1}{\|I + \beta M^T\|^2}.$$

Thus, it follows from (3.25) that

$$q_k(\gamma\alpha_k^*) \ge \frac{\gamma(2-\gamma)}{\|I + \beta M^T\|^2}\|u^k - \tilde{u}^k\|^2.$$

Substituting it in (3.24), the assertion (3.26) follows directly. □

Remark Since the predictor $\tilde{u}^k$ generated by (3.9) satisfies (3.16), as in [6], we can take

$$u^{k+1} = u^k - \gamma(I + \beta M)^{-1}(u^k - \tilde{u}^k) \qquad (3.27)$$

as the new iterate. For $G = (I + \beta M^T)(I + \beta M)$, by using (3.16), we get

$$\begin{aligned}
\|u^{k+1} - u^*\|_G^2 &= \|(u^k - u^*) - \gamma(I + \beta M)^{-1}(u^k - \tilde{u}^k)\|_G^2 \\
&= \|u^k - u^*\|_G^2 - 2\gamma(u^k - u^*)^T(I + \beta M)^T(u^k - \tilde{u}^k) \\
&\quad + \gamma^2\|(I + \beta M)^{-1}(u^k - \tilde{u}^k)\|_G^2 \\
&\le \|u^k - u^*\|_G^2 - \gamma(2-\gamma)\|u^k - \tilde{u}^k\|^2.
\end{aligned}$$

Thus, for solving the optimization problem (1.4), if we use (3.9) to offer the predictor and (3.27) to update the new iterate, then the sequence $\{u^k\}$ satisfies

$$\|u^{k+1} - u^*\|_G^2 \le \|u^k - u^*\|_G^2 - \gamma(2-\gamma)\|u^k - \tilde{u}^k\|^2.$$

4 Convergence rate of the Algorithm 2.1

This section studies the convergence rate of the Algorithm 2.1 for the composite convex optimization problem (1.3) (or its equivalent variational inequality (2.1)) :

$$\theta(x) - \theta(x^*) + (x - x^*)^T(Hx^* + c) \geq 0, \quad \forall x \in \mathcal{X}. \tag{4.1}$$

For given $x^k \in \Re^n$, the predictor $\tilde{x}^k$ in Algorithm 2.1 is offered by (2.7),The new iterate of the k-th iteration of Algorithm 2.1 is updated by

$$x^{k+1} = x^k - \gamma\alpha_k^*(x^k - \tilde{x}^k), \tag{4.2}$$

where

$$\alpha_k^* = \frac{\|x^k - \tilde{x}^k\|^2}{\|x^k - \tilde{x}^k\|_G^2}, \qquad G = I + \beta H \qquad \text{and} \qquad \gamma \in (0, 2). \tag{4.3}$$

It was proved [6] that the sequence $\{x^k\}$ generated by the Algorithm 2.1 satisfies

$$\|x^{k+1} - x^*\|_G^2 \leq \|x^k - x^*\|_G^2 - \gamma(2-\gamma)\alpha_k^*\|x^k - \tilde{x}^k\|^2. \tag{4.4}$$

Recall that $\mathcal{X}^*$ can be characterized as (see Theorem 2.1 in [8])

$$\mathcal{X}^* = \bigcap_{x \in \mathcal{X}} \{\tilde{x} \in \mathcal{X} : \theta(x) - \theta(\tilde{x}) + (x - \tilde{x})^T(Hx + c) \geq 0\}.$$

This implies that $\tilde{x} \in \mathcal{X}$ is an approximate solution of (4.1) with the accuracy ϵ if it satisfies

$$\tilde{x} \in \mathcal{X} \quad \text{and} \quad \inf_{x \in \mathcal{D}(\tilde{x})} \{\theta(x) - \theta(\tilde{x}) + (x - \tilde{x})^T(Hx + c)\} \geq -\epsilon,$$

where

$$\mathcal{D}(\tilde{x}) = \{x \in \mathcal{X} \mid \|x - \tilde{x}\| \leq 1\}.$$

In this section, we show that, for given $\epsilon > 0$, in $O(1/\epsilon)$ iterations the Algorithm 2.1 can find a $\tilde{x}$ such that

$$\tilde{x} \in \Omega \quad \text{and} \quad \sup_{x \in \mathcal{D}(\tilde{x})} \left\{\big(\theta(\tilde{x}) - \theta(x)\big) + (\tilde{x} - x)^T(Hx + c)\right\} \leq \epsilon, \tag{4.5}$$

where

$$\mathcal{D}(\tilde{x}) = \{x \in \Omega \mid \|x - \tilde{x}\| \leq 1\}.$$

In this sense, we will establish the algorithmic convergence complexity for the Algorithm 2.1.

4.1 Main theorem for complexity analysis

This subsection proves some main theorems for the complexity analysis. Now, we prove the key inequality for the complexity analysis of the Algorithm 2.1.

Theorem 4.1 *For given $x^k \in \Re^n$, let $\tilde{x}^k$ be generated by* (2.7)*. If he new iterate x^{k+1} is updated by* (4.2) *with any $\gamma \in (0,2)$, then we have*

$$\begin{aligned} &\gamma\alpha_k^*\beta\{\theta(x)-\theta(\tilde{x}^k)+(x-\tilde{x}^k)^T(Hx+c)\} \\ &\quad\geq \quad \frac{1}{2}\big(\|x-x^{k+1}\|_G^2-\|x-x^k\|_G^2\big)+\frac{1}{2}q_k(\gamma), \ \ \forall\, x\in\Omega, \end{aligned} \tag{4.6}$$

where

$$q_k(\gamma)=\gamma(2-\gamma)\alpha_k^*\|x^k-\tilde{x}^k\|^2. \tag{4.7}$$

Proof. Since $\tilde{x}^k$ is the solution of (2.7), we have (see (2.8))

$$\beta\theta(x)-\beta\theta(\tilde{x}^k)+(x-\tilde{x}^k)^T\{\beta(Hx^k+c)-(x^k-\tilde{x}^k)\}\geq 0, \ \forall x\in\mathcal{X}.$$

Thus, we have

$$\beta\theta(x)-\beta\theta(\tilde{x}^k)+(x-\tilde{x}^k)^T\beta(Hx^k+c)\geq(x-\tilde{x}^k)^T(x^k-\tilde{x}^k), \ \forall x\in\mathcal{X}.$$

Adding the term $(x-\tilde{x}^k)^T\beta H(x^k-\tilde{x}^k)$ to the both sides of the above inequality and using $(I+\beta H)=G$, we obtain

$$\begin{aligned} &\beta\theta(x)-\beta\theta(\tilde{x}^k)+(x-\tilde{x}^k)^T\{\beta(Hx^k+c)+\beta H(x^k-\tilde{x}^k)\} \\ &\quad\geq \quad (x-\tilde{x}^k)^TG(x^k-\tilde{x}^k), \quad \forall\, x\in\mathcal{X}. \end{aligned}$$

By using the identity

$$\beta(Hx^k+c)+\beta H(x^k-\tilde{x}^k)=\beta(Hx+c)+\beta H(\tilde{x}^k-x)+2\beta H(x^k-\tilde{x}^k)$$

Then, we rewrite the above inequality in our desirable form

$$\begin{aligned} &\beta\theta(x)-\beta\theta(\tilde{x}^k)+(x-\tilde{x}^k)^T\{\beta(Hx+c)+2\beta H(x^k-\tilde{x}^k)\} \\ &\quad\geq \quad (x-\tilde{x}^k)^TG(x^k-\tilde{x}^k)+\beta\|x-\tilde{x}^k\|_H^2, \ \forall x\in\mathcal{X}. \end{aligned}$$

By using the Cauchy-Schwarz inequality, from the above inequality we obtain

$$\begin{aligned}
&\beta\theta(x) - \beta\theta(\tilde{x}^k) + (x - \tilde{x}^k)^T \beta(Hx + c) \\
&\quad \geq \quad (x - \tilde{x}^k)^T G(x^k - \tilde{x}^k) + \beta\|x - \tilde{x}^k\|_H - 2(x - \tilde{x}^k)^T \beta H(x^k - \tilde{x}^k) \\
&\quad \geq \quad (x - \tilde{x}^k)^T G(x^k - \tilde{x}^k) - \beta\|x^k - \tilde{x}^k\|_H^2, \quad \forall\, x \in \mathcal{X},
\end{aligned}$$

and thus

$$\begin{aligned}
&\gamma\alpha_k^*\beta\{\theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T(Hx + c)\} \\
&\quad \geq \quad (x - \tilde{x}^k)^T G\gamma\alpha_k^*(x^k - \tilde{x}^k) - \gamma\alpha_k^*\beta\|x^k - \tilde{x}^k\|_H^2, \quad \forall\, x \in \mathcal{X}.
\end{aligned}$$

Because $\gamma\alpha_k^*(x^k - \tilde{x}^k) = (x^k - x^{k+1})$ (see (4.2)), thus we have

$$\begin{aligned}
&\gamma\alpha_k^*\beta\{\theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T(Hx + c)\} \\
&\quad \geq \quad (x - \tilde{x}^k)^T G(x^k - x^{k+1}) - \gamma\alpha_k^*\beta\|x^k - \tilde{x}^k\|_H^2, \ \forall\, x \in \mathcal{X}. \quad (4.8)
\end{aligned}$$

To the crossed term in the right hand side of (4.8), $(x - \tilde{x}^k)^T G(x^k - x^{k+1})$,

using an identity

$$(a-b)^T G(c-d) = \tfrac{1}{2}\big(\|a-d\|_G^2 - \|a-c\|_G^2\big) + \tfrac{1}{2}\big(\|c-b\|_G^2 - \|d-b\|_G^2\big),$$

we obtain

$$\begin{aligned}
&(x - \tilde{x}^k)^T G(x^k - x^{k+1}) \\
&\quad = \quad \tfrac{1}{2}\big(\|x - x^{k+1}\|_G^2 - \|x - x^k\|_G^2\big) + \tfrac{1}{2}\big(\|x^k - \tilde{x}^k\|_G^2 - \|x^{k+1} - \tilde{x}^k\|_G^2\big). \quad (4.9)
\end{aligned}$$

Using $x^{k+1} = x^k - \gamma\alpha_k^*(x^k - \tilde{x}^k)$ to the last part of the right hand side of (4.9), we get

$$\begin{aligned}
&\|x^k - \tilde{x}^k\|_G^2 - \|x^{k+1} - \tilde{x}^k\|_G^2 \\
&\quad = \quad \|x^k - \tilde{x}^k\|_G^2 - \|(x^k - \tilde{x}^k) - \gamma\alpha_k^*(x^k - \tilde{x}^k)\|_G^2 \\
&\quad = \quad 2\gamma\alpha_k^*\|x^k - \tilde{x}^k\|_G^2 - \gamma^2\alpha_k^*(\alpha_k^*\|x^k - \tilde{x}^k\|_G^2) \qquad \text{(see (4.3))} \\
&\quad = \quad 2\gamma\alpha_k^*\|x^k - \tilde{x}^k\|^2 + 2\gamma\alpha_k^*\beta\|x^k - \tilde{x}^k\|_H^2 - \gamma^2\alpha_k^*\|x^k - \tilde{x}^k\|^2 \\
&\quad = \quad \gamma(2-\gamma)\alpha_k^*\|x^k - \tilde{x}^k\|^2 + 2\gamma\alpha_k^*\|x^k - \tilde{x}^k\|_H^2.
\end{aligned}$$

Substituting it in the right hand side of (4.9) and using the definition of $q_k(\gamma)$, we obtain

$$\begin{aligned}&(x-\tilde{x}^k)^T G(x^k - x^{k+1})\\ &\quad = \quad \frac{1}{2}\big(\|x-x^{k+1}\|_G^2 - \|x-x^k\|_G^2\big) + \frac{1}{2}q_k(\gamma) + \gamma\alpha_k^*\beta\|x^k-\tilde{x}^k\|_H^2.\end{aligned} \tag{4.10}$$

Adding (4.8) and (4.10), the theorem is proved. $\square$

By setting $x = x^*$ in (4.6), we get

$$\begin{aligned}&\|x^k - x^*\|_G^2 - \|x^{k+1} - x^*\|_G^2\\ &\quad \geq \quad 2\gamma\alpha_k^*\beta\{(\theta(\tilde{x}^k) - \theta(x^*) + (\tilde{x}^k - x^*)^T(Hx*+c)\} + q_k(\gamma).\end{aligned}$$

Because $\theta(\tilde{x}^k) - \theta(x^*) + (\tilde{x}^k - x^*)^T(Hx^* + c) \geq 0$, it follows from the last inequality and (4.7) that

$$\|x^{k+1} - x^*\|_G^2 \leq \|x^k - x^*\|_G^2 - \gamma(2-\gamma)\alpha_k^*\|x^k - \tilde{x}^k\|^2.$$

Thus, the contraction property (4.4) is a byproduct of Theorem 4.1.

4.2 Convergence rate of the Algorithm 2.1.

This section uses Theorem 4.1 to show the convergence rate of the Algorithm 2.1.

Theorem 4.2 *Let the sequence $\{x^k\}$ be generated by the Algorithm 2.1. Then for any integer $t > 0$, it holds that*

$$\theta(\tilde{x}_t) - \theta(x) + (\tilde{x}_t - x)^T(Hx + c) \leq \frac{\|I + \beta H\|}{2\gamma\beta(t+1)}\|x - x^0\|_G^2, \ \forall x \in \mathcal{X}, \tag{4.11}$$

where

$$\tilde{x}_t = \frac{1}{\Upsilon_t}\sum_{k=0}^{t}\alpha_k^*\tilde{x}^k \qquad \textit{and} \qquad \Upsilon_t = \sum_{k=0}^{t}\alpha_k^*. \tag{4.12}$$

Proof. For the convergence rate proof, we allow $\gamma \in (0, 2]$. In this case, we still have $q_k(\gamma) \geq 0$. From (4.6) we get

$$\begin{aligned}&\alpha_k^*(\theta(x) - \theta(\tilde{x}^k)) + \alpha_k^*(x - \tilde{x}^k)^T(Hx + c)\\ &\quad \geq \quad \frac{1}{2\gamma\beta}\|x - x^{k+1}\|_G^2 - \frac{1}{2\gamma\beta}\|x - x^k\|_G^2, \ \ \forall\, x \in \mathcal{X}.\end{aligned}$$

Summing the above inequality over $k = 0, \ldots, t$, we obtain

$$\Big((\sum_{k=0}^{t} \alpha_k^*) \theta(x) - \sum_{k=0}^{t} \alpha_k^* \theta(\tilde{x}^k) \Big) + \Big((\sum_{k=0}^{t} \alpha_k^*) x - \sum_{k=0}^{t} \alpha_k^* \tilde{x}^k \Big)^T (Hx + c)$$
$$\geq \quad -\frac{1}{2\gamma\beta} \|x - x^0\|^2, \quad \forall x \in \mathcal{X}.$$

Using the notations of Υ_t and $\tilde{x}_t$ in the above inequality, we derive

$$\Big(\frac{1}{\Upsilon_t} (\sum_{k=0}^{t} \alpha_k^* \theta(\tilde{x}^k)) - \theta(x) \Big) + (\tilde{x}_t - x)^T (Hx + c) \leq \frac{\|x - x^0\|^2}{2\gamma\beta\Upsilon_t}, \quad \forall x \in \mathcal{X}. \tag{4.13}$$

Indeed, $\tilde{x}_t \in \mathcal{X}$ because it is a convex combination of $\tilde{x}^0, \tilde{x}^1, \ldots, \tilde{x}^t$. Since $\theta(x)$ is convex and

$$\tilde{x}_t = \frac{1}{\Upsilon_t} \sum_{k=0}^{t} \alpha_k^* \tilde{x}^k \quad \text{and} \quad \Upsilon_t = \sum_{k=0}^{t} \alpha_k^*.$$

we have $\theta(\tilde{x}_t) \leq \frac{1}{\Upsilon_t} \big(\sum_{k=0}^{t} \alpha_k^* \theta(\tilde{x}^k) \big)$. Thus, it follows from (4.13) that

$$\theta(\tilde{x}_t) - \theta(x) + (\tilde{x}_t - x)^T (Hx + c) \leq \frac{\|x - x^0\|^2}{2\gamma\beta\Upsilon_t}, \quad \forall x \in \mathcal{X}. \tag{4.14}$$

Because $\alpha_k^* \geq 1/\|I + \beta H\|$ for all $k > 0$ (see (4.3)), it follows from (4.12) that

$$\Upsilon_t \geq \frac{t+1}{\|I + \beta H\|}.$$

Substituting it in (4.14), the proof is complete. $\square$

Thus the Algorithm 2.1. has $O(1/t)$ convergence rate. For any substantial set $\mathcal{D} \subset \mathcal{X}$, the PC Method and PC Method reach

$$(\tilde{x}_t - x)^T (Hx + c) \leq \epsilon, \quad \forall\, x \in \mathcal{D}(\tilde{x}_t), \quad \text{in at most} \quad t = \left\lceil \frac{\|I + \beta H\| d^2}{2\gamma\beta\epsilon} \right\rceil$$

iterations, where $\tilde{x}_t$ is defined in (4.12) and $d = \sup \{ \|x - x^0\| \mid x \in \mathcal{D}(\tilde{x}_t) \}$. This convergence rate is in the ergodic sense, the statement (4.11) suggests us to take a larger parameter $\gamma \in (0, 2]$ in the correction steps of the the Algorithm 2.1.

References

[1] E. Blum and W. Oettli, *Mathematische Optimierung, Econometrics and Operations Research XX*, Springer Verlag, 1975.

[2] C. H. Chen, X. L. Fu, B.S. He and X. M. Yuan, On the Iteration Complexity of Some Projection Methods for Monotone Linear Variational Inequalities, JOTA, **172**, 914-928, 2017.

[3] D. Chen and Y. Zhang, A Hybrid Multi-Objective Scheme Applied to Redundant Robot Manipulators, IEEE Transactions on Automation Science and Engineering, **14**, 1337–1350, 2017.

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

[5] B.S. He, A new method for a class of linear variational inequalities, *Math. Progr.*, **66**, 137–144, 1994.

[6] B.S. He, Solving a class of linear projection equations, *Numerische Mathematik*, **68**, 71–80, 1994.

[7] B.S. He, A class of projection and contraction methods for monotone variational inequalities, *Appl. Math. & Optimi.*, **35**, 69–76, 1997.

[8] B.S. He and X.M. Yuan, On the $O(1/n)$ Convergence Rate of the Douglas-Rachford Alternating Direction Method, SIAM J. Numer. Anal. **50**, 700-709, 2012.

[9] A. Nemirovski, *Prox-method with rate of convergence $O(1/t)$ for variational inequality with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems*, SIAM J. Optim. 15 (2005), pp. 229-251.

[10] P. Tseng, *On accelerated proximal gradient methods for convex-concave optimization*,

manuscript, University of Washington, USA, 2008.

[11] Yunong Zhang, Senbo Fu, Zhijun Zhang, and Lin Xiao, On the LVI-Based Numerical Method (E47 Algorithm) for Solving Quadratic Programming Problems, IEEE International Conference on Automation and Logistics, 2011.

[12] Yunong Zhang, Long Jin, Robot Manipulator Redundancy Resolution, Hoboken, Wiley, 2017

[13] 何炳生, 论求解单调变分不等式的一些投影收缩算法, 计算数学, **18**, 54-60, 1996.

凸优化和单调变分不等式的收缩算法

第十讲: 基于梯度投影的凸优化 收缩算法和下降算法

Projected gradient-based contraction method
and descent method for convex optimization

南京大学数学系　何炳生
hebma@nju.edu.cn

简单约束的凸优化问题　这一讲讨论简单约束可微凸优化问题

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

的梯度算法, 其中 Ω 是$\Re^n$ 中的凸闭集, 并假设到 Ω 上的投影是容易实现的. 在第一讲中就已经提到, 简单约束可微凸优化问题等价于求变分不等式

$$\mathsf{VI}(\Omega, \nabla f) \qquad x^* \in \Omega, \qquad (x - x^*)^T \nabla f(x^*) \geq 0, \quad \forall x \in \Omega$$

的解. 这一讲的投影梯度方法, 分别是收缩算法和下降算法, 都不要用到函数值 $f(x)$, 只要对给定的 x, 能提供 $\nabla f(x)$. 收缩算法保证迭代点向解集靠近. 下降算法则隐含了目标函数值下降, 尽管目标函数值在计算过程中从不出现.

设 x^* 是变分不等式 $\mathsf{VI}(\Omega, \nabla f)$ 的解. 由于 $\tilde{x} = P_\Omega[x - \beta\nabla f(x)] \in \Omega$, 因此根据变分不等式的定义有第一个基本不等式

$$\text{(FI1)} \quad (\tilde{x} - x^*)^T \beta \nabla f(x^*) \geq 0.$$

由于 $\tilde{x}$ 是 $x - \beta\nabla f(x)$ 在 Ω 上的投影, $x^* \in \Omega$, 根据投影的基本性质, 有

$$\text{(FI2)} \quad (\tilde{x} - x^*)^T \big([x - \beta\nabla f(x)] - \tilde{x}\big) \geq 0.$$

1 凸优化的投影收缩算法

投影收缩算法由投影 $P_\Omega[x^k - \beta\nabla f(x^k)]$ 得到 $\tilde{x}^k$. $\quad x^k \in \Omega^* \;\Leftrightarrow\; x^k = \tilde{x}^k$.

误差度量函数 一个非负函数 $\varphi(x^k, \tilde{x}^k)$ 称作变分不等式 VI(Ω, F) 的误差度量函数, 如果有 $\delta > 0$, 使得

$$\varphi(x^k, \tilde{x}^k) \ge \delta\|x^k - \tilde{x}^k\|^2, \quad \text{并且} \quad \varphi(x^k, \tilde{x}^k) = 0 \;\Leftrightarrow\; x^k = \tilde{x}^k. \tag{1.1}$$

有利方向 设矩阵 G 正定, 凸优化的有利方向都是 $(x^k - \tilde{x}^k)$, 它满足

$$(x^k - x^*)^T G(x^k - \tilde{x}^k) \ge \varphi(x^k, \tilde{x}^k), \; \forall\, x^* \in \Omega^*. \tag{1.2}$$

初等收缩算法 考虑将 (1.1) 中的 $\varphi(x^k, \tilde{x}^k) \ge \delta\|x^k - \tilde{x}^k\|^2$ 改成条件

$$\varphi(x^k, \tilde{x}^k) \ge \frac{1}{2}\big(\|x^k - \tilde{x}^k\|_G^2 + \tau\|x^k - \tilde{x}^k\|^2\big), \quad (\tau > 0). \tag{1.3}$$

我们将条件 (1.2) 和 (1.3) 满足时, 用步长为 **1** 的迭代公式

$$x^{k+1} = x^k - (x^k - \tilde{x}^k) = \tilde{x}^k, \tag{1.4}$$

产生新迭代点的方法, 称为 **Primary Method** (初等方法).

1.1 凸二次优化的投影收缩算法

对凸二次优化, 我们将第一和第二个基本不等式

$$\begin{cases} (\tilde{x} - x^*)^T \beta(Hx^* + c) \ge 0 & (FI1) \\ (\tilde{x} - x^*)^T \big([x - \beta(Hx + c)] - \tilde{x}\big) \ge 0 & (FI2) \end{cases}$$

相加, 对所有的 $x \in \Re^n$, 都有

$$\{(x - x^*) - (x - \tilde{x})\}^T\{(x - \tilde{x}) - \beta H(x - x^*)\} \ge 0$$

由上式和 H 半正定,

$$(x - x^*)^T(I + \beta H)(x - \tilde{x}) \ge \|x - \tilde{x}\|^2. \tag{1.1}$$

误差度量函数就是 $\varphi(x^k, \tilde{x}^k) = \|x^k - \tilde{x}^k\|^2$. 上式说明 $(\tilde{x}^k - x^k)$ 是未知距离函数 $\frac{1}{2}\|x - x^*\|^2_{(I+\beta H)}$ 在 x^k 处的下降方向. 记 $G = I + \beta H$, 就是(1.2) 的形式

$$(x^k - x^*)^T G(x^k - \tilde{x}^k) \ge \varphi(x^k, \tilde{x}^k), \; \forall\, x^* \in \Omega^*.$$

这里考虑的投影收缩算法, 要求它产生的迭代点使得 $\|x^k - x^*\|_G^2$ 严格单调下降. 以

$$x(\alpha) = x^k - \alpha(x^k - \tilde{x}^k) \tag{1.2}$$

产生依赖于步长 α 的新迭代点. 考察与 α 相关的距离平方缩短量

$$\vartheta(\alpha) = \|x^k - x^*\|_G^2 - \|x(\alpha) - x^*\|_G^2. \tag{1.3}$$

利用 (1.1) 就有

$$\begin{aligned} \vartheta(\alpha) &= \|x^k - x^*\|_G^2 - \|x^k - x^* - \alpha(x^k - \tilde{x}^k)\|_G^2 \\ &\geq 2\alpha\|x^k - \tilde{x}^k\|^2 - \alpha^2\|x^k - \tilde{x}^k\|_G^2. \end{aligned} \tag{1.4}$$

我们得到 $\vartheta(\alpha)$ 的一个下界、二次函数 $q(\alpha)$,

$$q(\alpha) = 2\alpha\|x^k - \tilde{x}^k\|^2 - \alpha^2\|x^k - \tilde{x}^k\|_G^2. \tag{1.5}$$

使 $q(\alpha)$ 达到极大的

$$\alpha_k^* = \|x^k - \tilde{x}^k\|^2 / \|x^k - \tilde{x}^k\|_G^2. \tag{1.6}$$

用迭代式

$$x^{k+1} = x^k - \gamma\alpha_k^*(x^k - \tilde{x}^k), \quad \gamma \in (0, 2) \tag{1.7}$$

产生新的迭代点 x^{k+1}. 这样的迭代序列 $\{x^k\}$ 满足

$$\|x^{k+1} - x^*\|_{(I+\beta H)}^2 \leq \|x^k - x^*\|_{(I+\beta H)}^2 - \gamma(2-\gamma)\alpha_k^*\|x^k - \tilde{x}^k\|^2. \tag{1.8}$$

换句话说, 迭代序列 $\{x^k\}$ 在 $G = (I + \beta H)$-模下向解集收缩. 关于凸二次优化在$(I + \beta H)$-模下收缩的算法, 更详细的可参见文献 [5]. 一般说来, 如何选取适当的 β 对收敛速度影响很大.

自调比投影收缩算法 = 故意缩短了步长的最速下降法

注意到在 $G = (I + \beta H)$ 时, "最优步长"

$$\alpha_k^* = \frac{\|x^k - \tilde{x}^k\|^2}{(x^k - \tilde{x}^k)^T(I + \beta H)(x^k - \tilde{x}^k)}. \tag{1.9}$$

对给定的 $\nu \in (0, 1)$, 我们使用自调比法则选取 β 使得

$$(x^k - \tilde{x}^k)^T(\beta H)(x^k - \tilde{x}^k) \leq \nu\|x^k - \tilde{x}^k\|^2, \tag{1.10}$$

代入 (1.9) 便有

$$\alpha_k^* \geq \frac{1}{1+\nu} > \frac{1}{2}.$$

这使得我们有可能在 (1.7) 中动态地取

$$\gamma_k = 1/\alpha_k^*, \qquad \text{因此} \qquad 1 < \gamma_k \le 1 + \nu < 2.$$

迭代式 (1.7) 就成为

$$x^{k+1} = \tilde{x}^k = P_\Omega[x^k - \beta(Hx^k + c)]. \tag{1.11}$$

这样做使得有可能每步迭代只计算一次目标函数的梯度(矩阵与向量相乘).

在 (1.8) 中利用 $\gamma_k \alpha_k^* = 1$, $\gamma_k \le 1 + \nu$ 以及 $\tilde{x}^k = x^{k+1}$ 就有

$$\|x^{k+1} - x^*\|^2_{(I+\beta H)} \le \|x^k - x^*\|^2_{(I+\beta H)} - (1-\nu)\|x^k - x^{k+1}\|^2. \tag{1.12}$$

♣ 求解 $\min\{\frac{1}{2}x^T Hx + c^T x\}$ 的最速下降法是

$$x^{k+1} = x^k - \alpha_k^{SD}(Hx^k + c), \qquad \alpha_k^{SD} = \frac{\|Hx^k + c\|^2}{(Hx^k + c)^T H (Hx^k + c)}.$$

当 $\Omega = \Re^n$ 时, 根据 (1.11) 和 (1.10) 有

$$x^{k+1} = x^k - \beta(Hx^k + c) \qquad \text{和} \qquad \beta \le \nu \alpha_k^{SD}.$$

因此, [5] 中符合条件 (1.10) 的投影收缩算法(1.11) 相当于将最速下降法故意缩短了步长. 换句话说, 据此设计的求解无约束凸二次优化的算法, 恰是缩小了步长的最速下降法. 我们曾担心用这些方法求解无约束凸二次规划效果会比最速下降法还差. 事实上, 对最速下降法故意缩短步长, 收敛速率有令人难以置信的数量级提高. 读者容易用下面的例子来验证.

数值试验 试验问题中的 Hessian 矩阵是 Hilbert 矩阵, 即：

$$H = \{h_{ij}\}, \quad h_{ij} = \frac{1}{i+j-1}, \quad i = 1, \cdots, n; \quad j = 1, \cdots, n.$$

问题的规模(维数)分别从从 100 到 500. 因为 Hilbert 矩阵的条件很坏, 我们将最优解 x^* 设定为每个分量都是 1 的向量, 然后令 $c = -Hx^*$, 再用梯度类算法去求解. 试验中, 我们分别将零向量、 c 和 $-c$ 取作初始向量. 这里采用的停机准则是$\|Hx^k + c\|/\|Hx^0 + c\| \le 10^{-7}$. 表 1–3 分别列出了在不同问题规模、不同缩扩(缩减或扩张)因子和不同初始向量下的迭代次数. 其中 n 表示问题规模, r 表示缩扩因子. $r = 1$ 时, 就是最速下降法.

初步的试验结果不但解除了我们的上述担心, 同时也印证了 Dai 和 Yuan [2] 提到的将最速下降法的步长乘上一个小于 1 的因子会加快收敛的结论.

表 1. 初始向量 $x^0 = 0$, 使用不同缩扩因子 r 时的迭代次数

n=	0.1	0.3	0.5	0.7	0.8	**0.9**	0.95	0.99	**1.00**	1.20
100	2863	1346	853	627	582	**437**	565	1201	**13169**	22695
200	3283	1398	923	804	541	**669**	898	1178	**14655**	21083
300	3497	1323	856	739	720	**568**	619	1545	**17467**	24027
500	3642	1351	1023	773	667	**578**	836	2024	**17757**	22750

初始向量 $x^0 = 0$, 迭代结束时平均相对误差 $\|x^k - x^*\|/\|x^0 - x^*\| = 3.0\text{e} - 3$

表 2. 初始向量 $x^0 = c$, 使用不同缩扩因子 r 时的迭代次数

n=	0.1	0.3	0.5	0.7	0.8	**0.9**	0.95	0.99	**1.0**	1.2
100	2129	1034	544	424	302	**438**	568	919	**5527**	9667
200	1880	808	568	482	372	**339**	446	713	**6625**	11023
300	1852	1002	741	531	610	**452**	450	917	**6631**	10235
500	2059	939	568	573	379	**547**	558	874	**7739**	11269

初始向量 $x^0 = c$, 迭代结束时平均相对误差 $\|x^k - x^*\|/\|x^0 - x^*\| = 1.8\text{e} - 3$

表 3. 初始向量 $x^0 = -c$, 使用不同缩扩因子 r 时的迭代次数

n=	0.1	0.3	0.5	0.7	0.8	**0.9**	0.95	0.99	**1.0**	1.2
100	2545	1221	666	591	498	**482**	638	1581	**14442**	20380
200	2826	990	874	470	526	**455**	578	841	**15222**	18892
300	2891	1299	918	738	549	**571**	608	2552	**18762**	21208
500	3158	1769	909	678	506	**512**	678	1240	**17512**	19790

初始向量 $x^0 = -c$, 迭代结束时平均相对误差 $\|x^k - x^*\|/\|x^0 - x^*\| = 3.8\text{e} - 3$

1.2 基于 FI2 的非线性凸优化的投影收缩算法

可微凸优化问题 $\min\{f(x) \,|\, x \in \Omega\}$ 与变分不等式 VI$(\Omega, \nabla f)$ 等价. 由

$$\tilde{x} = P_\Omega[x - \beta\nabla f(x)]$$

和投影的基本性质, 可以得到

$$(\tilde{x} - x^*)^T\big([x - \beta\nabla f(x)] - \tilde{x}\big) \geq 0.$$

上面的不等式其实就是第二讲中的第二个基本不等式. 据此可以推得

$$\begin{aligned}(\tilde{x} - x^*)^T(x - \tilde{x}) &\geq (\tilde{x} - x^*)^T\beta\nabla f(x) \\ &= \quad (x - x^*)^T\beta\nabla f(x) - (x - \tilde{x})^T\beta\nabla f(x). \qquad (1.13)\end{aligned}$$

另据凸函数的性质, 有

$$(x - x^*)^T\nabla f(x) \geq f(x) - f(x^*) \geq (x - \tilde{x})^T\nabla f(\tilde{x}) + (f(\tilde{x}) - f(x^*)). \quad (1.14)$$

以 (1.14) 代入 (1.13) 的右端, 则对所有的 $x \in \Re^n$, 都有

$$(\tilde{x} - x^*)^T(x - \tilde{x}) \geq -\beta(x - \tilde{x})^T(\nabla f(x) - \nabla f(\tilde{x})).$$

由上式得

$$(x-x^*)^T(x-\tilde{x}) \geq \|x-\tilde{x}\|^2 - \beta(x-\tilde{x})^T(\nabla f(x)-\nabla f(\tilde{x})). \qquad (1.15)$$

我们定义

$$\varphi(x,\tilde{x}) = \|x-\tilde{x}\|^2 - \beta(x-\tilde{x})^T(\nabla f(x)-\nabla f(\tilde{x})). \qquad (1.16)$$

假设 ∇f 是 Lipschitz 连续的. 对于一个确定的 $\nu \in (0,1)$, 总可以采用 Armijo 技术对算子进行调比, 使得 $\beta\nabla f$ 的 Lipschitz 常数不大于 ν, 有

$$\beta\|\nabla f(x)-\nabla f(\tilde{x})\| \leq \nu\|x-\tilde{x}\|$$

式成立. 这样便有

$$(x-x^*)^T(x-\tilde{x}) \geq \varphi(x,\tilde{x}) \geq (1-\nu)\|x-\tilde{x}\|^2.$$

由 (1.16) 定义的 $\varphi(x,\tilde{x})$ 满足条件 (1.1), 其中的 $\delta = 1-\nu > 0$. 考虑欧氏模下的收缩算法, 由

$$x^{k+1} = x^k - \gamma\alpha_k^*(x^k-\tilde{x}^k),$$

产生新的迭代点 x^{k+1}, 其中

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

迭代序列 $\{x^k\}$ 就满足

$$\|x^{k+1}-x^*\|^2 \leq \|x^k-x^*\|^2 - \gamma(2-\gamma)\alpha_k^*\varphi(x^k,\tilde{x}^k).$$

由 $\varphi(x,\tilde{x}) \geq (1-\nu)\|x-\tilde{x}\|^2$ 可以进一步推得

$$\|x^{k+1}-x^*\|^2 \leq \|x^k-x^*\|^2 - \gamma(2-\gamma)(1-\nu)^2\|x^k-\tilde{x}^k\|^2.$$

有关这一节的相关论文和数值试验可见:

- 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., 51, 649-679, 2012
- 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. 51, 681-708, 2012

2 Projected Gradient Descent (PDG) method for nonlinear convex optimization

The findings on projection and contraction method for solving the quadratic programming also contribute to solving the following differentiable convex optimization problem.

Let Ω be a convex closed set in R^n. The problem concerted in this subsection is to find $x^* \in \Omega$, such that

$$(x - x^*)^T g(x^*) \geq 0, \quad \forall\, x \in \Omega, \tag{2.1}$$

where $g(x)$ is a mapping from R^n into itself. We assume that $g(x)$ is the gradient of a certain convex function, say $f(x)$, however, $f(x)$ is not provided. Only for given x, $g(x)$ is observable (sometimes with costly expenses).

In other words, (2.1) is equivalent to the following convex optimization problem

$$\min \{f(x) \mid x \in \Omega\}. \tag{2.2}$$

We call (3.2) an **oracle** convex optimization problem, because only the gradient information $g(x)$ can be used for solving (3.2). For $x^* \in \Omega^*$, we assume $f(x^*) > -\infty$.

In addition, we also assume that $g(x)$ is Lipschitz continuous, i.e., there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall\, x,\, y \in \Re^n. \tag{2.3}$$

We require that $g(x)$ is Lipschitz continuous while it does not need to know the value of L in (2.3).

The methods presented in this section do not involve the value of $f(x)$, but they can guarantee that $f(x^k)$ is strict monotonically decreasing, hence they belong to the descending methods.

2.1 Steepest descent method for convex programming

Single step projected gradient method.

Step k. ($k \geq 0$) Set

$$x^{k+1} = P_\Omega[x^k - \beta_k g(x^k)], \tag{2.4a}$$

where the step size β_k satisfies the following condition:

$$(x^k - x^{k+1})^T (g(x^k) - g(x^{k+1})) \leq \frac{\nu}{\beta_k} \|x^k - x^{k+1}\|^2. \tag{2.4b}$$

Note that the condition (2.4b) automatically holds when $\beta_k \leq \nu/L$, where L is

the Lipschitz modulus of $g(x)$. The reason is

$$\begin{aligned} &(x^k - x^{k+1})^T \beta_k (g(x^k) - g(x^{k+1})) \\ &\quad \leq \quad \|x^k - x^{k+1}\| \cdot \beta_k L \|x^k - x^{k+1}\| \;\leq\; \nu \|x^k - x^{k+1}\|^2. \end{aligned}$$

2.2 Global convergence of the proposed method

In the following, we show an important lemma by using the basic properties of the projection and convex function.

Lemma 2.1 *For given x^k, let x^{k+1} be generated by* (2.4a)*. If the step-size β_k satisfies* (2.4b)*, then we have*

$$(x - x^{k+1})^T g(x^k) \geq \tfrac{1}{\beta_k} (x - x^{k+1})^T (x^k - x^{k+1}), \quad \forall x \in \Omega, \tag{2.5}$$

and

$$\begin{aligned} &\beta_k (f(x) - f(x^{k+1})) \\ &\quad \geq \; (x - x^{k+1})^T (x^k - x^{k+1}) - \nu \|x^k - x^{k+1}\|^2, \; \forall\, x \in \Omega. \end{aligned} \tag{2.6}$$

Proof. Note that x^{k+1} is the projection of $[x^k - \beta_k g(x^k)]$ on Ω. Using the projection's property, $(x - P_\Omega(z))^T(z - P_\Omega(z)) \le 0$, $\forall x \in \Omega$, we have

$$(x - x^{k+1})^T\{[x^k - \beta_k g(x^k)] - x^{k+1}\} \le 0, \quad \forall\, x \in \Omega.$$

It follows that

$$(x - x^{k+1})^T \beta_k g(x^k) \ge (x - x^{k+1})^T (x^k - x^{k+1}), \quad \forall\, x \in \Omega, \tag{2.7}$$

and the first assertion (2.5) is proved. Using the convexity of f , we have

$$f(x) \ge f(x^k) + (x - x^k)^T g(x^k), \tag{2.8}$$

and

$$\begin{aligned} f(x^k) &\ge f(x^{k+1}) + (x^k - x^{k+1})^T g(x^{k+1}) \\ &= f(x^{k+1}) + (x^k - x^{k+1})^T g(x^k) \\ &\qquad -(x^k - x^{k+1})^T \big(g(x^k) - g(x^{k+1})\big) \\ &\ge f(x^{k+1}) + (x^k - x^{k+1})^T g(x^k) - \frac{\nu}{\beta_k}\|x^k - x^{k+1}\|^2. \end{aligned} \tag{2.9}$$

The last " $\ge$ " is due to (2.4b). From (2.8) and (2.9), we get

$$\begin{aligned} & f(x) - f(x^{k+1}) \\ &\ge f(x^k) + (x - x^k)^T g(x^k) \\ &\quad -\Big\{f(x^k) + (x^{k+1} - x^k)^T g(x^k) + \frac{\nu}{\beta_k}\|x^k - x^{k+1}\|^2\Big\} \\ &= (x - x^{k+1})^T g(x^k) - \frac{\nu}{\beta_k}\|x^k - x^{k+1}\|^2. \end{aligned} \tag{2.10}$$

Substituting (2.5) in (2.10), we obtain

$$f(x) - f(x^{k+1}) \ge \frac{1}{\beta_k}(x - x^{k+1})^T (x^k - x^{k+1}) - \frac{\nu}{\beta_k}\|x^k - x^{k+1}\|^2,$$

and the second assertion of this lemma is proved. $\square$

The following theorem shows that the projected gradient method (2.4) is a descent method whose objective function value $\{f(x^k)\}$ is monotonically decreasing.

Theorem 2.1 *Let $\{x^k\}$ be the sequence generated by the single step projected gradient method* (2.4)*. Then, we have*

$$f(x^{k+1}) \leq f(x^k) - \frac{1-\nu}{\beta_k}\|x^k - x^{k+1}\|^2, \tag{2.11}$$

and

$$\begin{aligned} \|x^{k+1} - x^*\|^2 \quad \leq \quad & \|x^k - x^*\|^2 - (1-2\nu)\|x^k - x^{k+1}\|^2 \\ & -2\beta_k(f(x^{k+1}) - f(x^*)). \end{aligned} \tag{2.12}$$

Proof. Setting $x = x^k$ in (2.6) in Lemma 2.1, we obtain the assertion (2.11) immediately. Next, setting $x = x^*$ in (2.6), we have

$$\begin{aligned} & \beta_k(f(x^*) - f(x^{k+1})) \\ & \quad \geq \quad (x^* - x^{k+1})^T(x^k - x^{k+1}) - \nu\|x^k - x^{k+1}\|^2, \end{aligned}$$

and thus

$$\begin{aligned} & (x^k - x^*)^T(x^k - x^{k+1}) \\ & \quad \geq \quad (1-\nu)\|x^k - x^{k+1}\|^2 + \beta_k(f(x^{k+1}) - f(x^*)). \end{aligned}$$

Using the above inequality, we get

$$\begin{aligned} & \|x^{k+1} - x^*\|^2 \\ & \quad = \quad \|(x^k - x^*) - (x^k - x^{k+1})\|^2 \\ & \quad = \quad \|x^k - x^*\|^2 - 2(x^k - x^*)^T(x^k - x^{k+1}) + \|x^k - x^{k+1}\|^2 \\ & \quad \leq \quad \|x^k - x^*\|^2 - 2(1-\nu)\|x^k - x^{k+1}\|^2 \\ & \qquad\qquad -2\beta_k(f(x^{k+1}) - f(x^*)) + \|x^k - x^{k+1}\|^2 \\ & \quad = \quad \|x^k - x^*\|^2 - (1-2\nu)\|x^k - x^{k+1}\|^2 \\ & \qquad\qquad -2\beta_k(f(x^{k+1}) - f(x^*)). \end{aligned}$$

This completes the proof of the assertion (2.12). □

Directly from (2.12), it follows the following corollary:

Corollary 2.1 *Let $\{x^k\}$ be the sequence generated by the single step projected gradient method* (2.4)*. If $\nu \le \frac{1}{2}$, we have $\|x^{k+1} - x^*\|^2 < \|x^k - x^*\|^2$, for any $x^* \in \Omega^*$. The generated sequence $\{x^k\}$ is in a compact set.*

2.3 Convergence rate of the proposed method

Below we show that the iteration-complexity of the single projected gradient method is $O(1/k)$. For the convenience, we assume $\beta_k \equiv \beta$.

Theorem 2.2 *Let $\{x^k\}$ be generated by the single step projected gradient method* (2.4)*. Then, we have*

$$\begin{aligned} & 2k\beta(f(x^k) - f(x^*)) \\ & \quad \le \|x^0 - x^*\|^2 - \sum_{l=0}^{k-1} \Big((1-2\nu) + 2l(1-\nu)\Big)\|x^l - x^{l+1}\|^2. \end{aligned} \tag{2.13}$$

Proof. First, it follows from (2.12) that, for any $x^* \in \Omega^*$ and all $l \ge 0$, we have

$$2\beta(f(x^*) - f(x^{l+1})) \ge \|x^{l+1} - x^*\|^2 - \|x^l - x^*\|^2 + (1-2\nu)\|x^l - x^{l+1}\|^2.$$

Summing the above inequality over $l = 0, \ldots, k-1$, we obtain

$$\begin{aligned} & 2\beta\Big(kf(x^*) - \sum_{l=0}^{k-1} f(x^{l+1})\Big) \\ & \quad \ge \quad \|x^k - x^*\|^2 - \|x^0 - x^*\|^2 + \sum_{l=0}^{k-1}(1-2\nu)\|x^l - x^{l+1}\|^2. \end{aligned} \tag{2.14}$$

It follows from (2.11) that

$$2\beta l\big(f(x^l) - f(x^{l+1})\big) \ge 2l(1-\nu)\|x^l - x^{l+1}\|^2,$$

which can be rewritten as

$$2\beta\big(lf(x^l) - (l+1)f(x^{l+1}) + f(x^{l+1})\big) \ge 2l(1-\nu)\|x^l - x^{l+1}\|^2.$$

Summing the above inequality over $l = 0, \ldots, k-1$, it follows that

$$2\beta \sum_{l=0}^{k-1} \Big(l f(x^l) - (l+1) f(x^{l+1}) + f(x^{l+1}) \Big) \ge \sum_{l=0}^{k-1} 2l(1-\nu) \|x^l - x^{l+1}\|^2,$$

which simplifies to

$$2\beta \Big(-k f(x^k) + \sum_{l=0}^{k-1} f(x^{l+1}) \Big) \ge \sum_{l=0}^{k-1} 2l(1-\nu) \|x^l - x^{l+1}\|^2. \qquad (2.15)$$

Adding (2.14) and (2.15), we get

$$\begin{aligned} & 2k\beta \big(f(x^*) - f(x^k) \big) \\ \ge \quad & -\|x^0 - x^*\|^2 + \sum_{l=0}^{k-1} \Big((1-2\nu) + 2l(1-\nu) \Big) \|x^l - x^{l+1}\|^2, \end{aligned}$$

which implies (2.13) and the theorem is proved. $\square$

From (2.13) follows directly the following theorem.

Theorem 2.3 *Let $\{x^k\}$ be generated by the single step projected gradient method. If $\nu \le \frac{1}{2}$, then we have*

$$f(x^k) - f(x^*) \le \frac{\|x^0 - x^*\|^2}{2k\beta}, \qquad (2.16)$$

and thus the iteration-complexity of this method is $O(1/k)$.

What is about for any $\nu \in (0.5, 1)$? For such ν, we define

$$p(\nu) = \operatorname{argmin}\{l \,|\, l \ge 0 \text{ is a integer}, \ (1-2\nu) + 2l(1-\nu) \ge 0\}. \qquad (2.17)$$

For any $\nu \in (0.5, 1)$, $p(\nu)$ is finite number. For example,we have

$\nu =$	0.9	0.8	0.7	$(0.5, 0.7)$
$p(\nu) =$	4	2	1	1

.

Since the term $\sum_{p(\nu)}^{k-1} \Big((1-2\nu) + 2l(1-\nu) \Big) \|x^l - x^{l+1}\|^2$ is positive, it

follows from Theorem 2.2 (see(2.13)) that

$$2k\beta(f(x^k)-f(x^*)) \le \|x^0-x^*\|^2 - \sum_{l=0}^{p(\nu)-1} \Big((1-2\nu)+2l(1-\nu)\Big)\|x^l-x^{l+1}\|^2.$$

The last inequality implies that $\lim_{k\to\infty}(f(x^k)-f(x^*))=0$.

The iteration-complexity of this method is $O(1/k)$ for any $\nu \in (0,1)$.

Theorem 2.4 *Let $\{x^k\}$ be generated by the single step projected gradient method, then we have*

$$f(x^k)-f(x^*) \le \frac{\|x^0-x^*\|^2+D}{2k\beta}, \tag{2.18}$$

where

$$D = -\sum_{l=0}^{p(\nu)-1} \big((1-2\nu)+2l(1-\nu)\big)\|x^l-x^{l+1}\|^2.$$

and $p(\nu)$ is a finite integer defined in (2.17)*.*

Self-adaptive projected gradient descent method

Self-adaptive projected gradient descent method.

Set $\beta_0 = 1$, $\mu = 0.5$, $\nu = 0.9$, $x^0 \in \Omega$ and $k = 0$. Provide $g(x^0)$.

For $k = 0, 1, \ldots$, if the stopping criterium is not satisfied, do

Step 1. $\tilde{x}^k = P_\Omega[x^k - \beta_k g(x^k)]$,

$r_k = \beta_k \|g(x^k) - g(\tilde{x}^k)\| / \|x^k - \tilde{x}^k\|$.

while $r_k > \nu$

$\boxed{\beta_k := \beta_k * 0.8/r_k,}$

$\tilde{x}^k = P_\Omega[x^k - \beta_k g(x^k)]$,

$r_k = \beta_k \|g(x^k) - g(\tilde{x}^k)\| / \|x^k - \tilde{x}^k\|$.

end(while)

$x^{k+1} = \tilde{x}^k$,

$g(x^{k+1}) = g(\tilde{x}^k)$.

$\boxed{\textbf{If} \quad r_k \le \mu \quad \text{then} \quad \beta_k := \beta_k * 1.5, \quad \textbf{end(if)}}$

Step 2. $\beta_{k+1} = \beta_k$ and $k = k+1$, go to Step 1.

Remark 2.1 *Instead of the condition* (2.4b)*, here we have*

$$\beta_k\|g(x^k) - g(x^{k+1})\| \le \nu\|x^k - x^{k+1}\|.$$

Remark 2.2 *If* $r_k \le \nu$*, we direct take* $x^{k+1} = \tilde{x}^k$*, and* $g(x^{k+1}) = g(\tilde{x}^k)$ *for the next iteration. We call the method Self-adaptive single step projected gradient method because it needs only once evaluation of the gradient* $g(x^k)$ *in each iteration when adjusting the parameter* β_k *is not necessary.*

Remark 2.3 *If* $r_k > \nu$*, we adjust the parameter* β_k *by* $\beta_k := \beta_k * 0.8/r_k$*. According to our limited numerical experiments, using the reduced* β_k*, the condition* $r_k \le \nu$ *is satisfied.*

Remark 2.4 *Too small step size* β_k *will leads to slow convergence. If* $r_k \le \mu$*, we will enlarge the trial step size* β *for the by* $\beta_k := \beta_k * 1.5$*.*

3 经济平衡问题上的一个应用

作为单步投影梯度法的应用, 我们讨论第一讲 §2 提到的保护资源、保障供给互补问题的求解方法. 读者可以从第一讲的 §2 中更好地了解问题的背景. 对所有的 $i = 1, \ldots, m,\ j = 1, \ldots, n$, 使用记号：

S_i: 该种商品的第 i 个资源地;

D_j: 该种商品的第 j 个需求地;

x_{ij}: 从 S_i 到 D_j 的交易量;

s_i: 经营者们在资源地 S_i 的总采购量, $s_i = \sum_{j=1}^n x_{ij}$;

d_j: 经营者在需求地 D_j 的总销售量, $d_j = \sum_{i=1}^m x_{ij}$;

h_i^s: 经营者在资源地 S_i 处的采购价;

h_j^d: 经营者在需求地 D_j 处的销售价;

t_{ij}: 从 S_i 到 D_j 的交易费用(包括运输费用)；

y_i: 政府为避免资源过度开采而在资源地 S_i 向经营者征收的资源税;

z_j: 政府为保障供给而在需求地 D_j 给经营者的经营补贴.

3.1 经营者追求利益最大化之互补问题

显然, 为了获利, 如果 S_i 处的采购价、资源税及从 S_i 到 D_j 的交易费的和 $(h_i^s + y_i + t_{ij})$ 不小于需求地 D_j 处的销售价与政府补贴的和 $(h_j^d + z_j)$, 经营者是不会从 S_i 采购商品运到 D_j 销售的. 反之, 根据贪婪原理, 经营者会尽可能增大经营量, 直到 $(h_i^s + y_i + t_{ij})$ 和 $(h_j^d + z_j)$ 相等. 这种关系的数学表达式是下面的平衡问题:

$$h_i^s + y_i + t_{ij} \begin{cases} \geq h_j^d + z_j, & \text{if } x_{ij} = 0, \\ = h_j^d + z_j, & \text{if } x_{ij} > 0\,. \end{cases} \tag{3.1}$$

它可以写成互补问题的形式: 对任意的 $i = 1, \ldots, m$ 和 $j = 1, \ldots, n$, 都有

$$0 \leq x_{ij} \ \perp \ \big((h_i^s + y_i + t_{ij}) - (h_j^d + z_j)\big) \geq 0. \tag{3.2}$$

通常假设 S_i 处的采购价 f_i 仅与经营者们在该地的采购量 s_i 有关; D_j 处的销售价 g_j 仅与经营者们运到该地的销售量 d_j 有关; 从 S_i 到 D_j 的交易价仅与它们之间的交易量 x_{ij}有关. 对于 $y = 0$ 和 $z = 0$ 的互补问题 (3.2), 文献中 [12] 称之为空间价格平衡问题 (Spatial Price Equilibrium Problem). 已有的求解空

间价格平衡问题的方法[?, 8, 9] 都对函数 h^s, h^d 和 t 的表达式有一定的要求. 当 a_i, b_j, c_{ij}, ξ_i, η_j, ζ_{ij} 为常数,

$$h_i^s(s_i) = \xi_i + a_i s_i, \qquad a_i \geq 0; \tag{3.3a}$$

$$h_j^d(d_j) = \eta_j - b_j d_j, \qquad b_j \geq 0; \tag{3.3b}$$

$$t_{ij}(x_{ij}) = \zeta_{ij} + c_{ij} x_{ij}, \qquad c_{ij} \geq 0 \tag{3.3c}$$

时, (3.2) 就是一个单调对称线性互补问题, 可以用求解带非负约束凸二次规划的方法求解 [3]. 在实际生活中, 这个问题是由市场根据贪婪原理自行解决的.

3.2 保护资源和保障供给的经济平衡问题

在现实生活中, 对(3.2) 中给定的 $y \geq 0$ 和 $z \geq 0$, 经济规律这只“无形的手”会让经营者们根据贪婪原理找到相应的 (x, s, d), 它是变分不等式(3.2) 的解. 换句话说, 原问题的解可由经营者们在经营活动中自行给出. 我们通常只知道函数 f, g 和 t 的一些性质而不知道它们的具体表达式, 能够观察到的是这个依赖于给定 y 和 z 的 s 与 d, 由于它是经营者根据贪婪原理给出的, 因此不可能顾及保护资源和保障供给. 假设从可持续发展的要求允许的最大资源消耗

量是$s^{\max}$, 为保障供给而必需的最小供应量是$d^{\min}$, 它们之间满足相容关系

$$\sum_{i=1}^{m} s_i^{\max} \geq \sum_{j=1}^{n} d_j^{\min}.$$

职能部门要采取经济手段来保证

$$s \leq s^{\max} \quad \text{和} \quad d \geq d^{\min}. \tag{3.4}$$

具体说来，在过度热销的资源地 S_i 通过对经营者征收资源税 y_i 保护资源，对供应不足的需求地 D_j 通过给经营者补贴 z_j 而吸引经营者增加供给. 我们的任务则是用数学方法帮助职能部门找到最优税收向量 $y^* \in \Re_+^m$ 和最优补贴向量 $z^* \in \Re_+^n$, 使得对我们给出的 (y^*, z^*) 以及经营者们由此产生的空间价格平衡问题 (3.2) 的解中的 s^* 和 d^* 满足

$$y^* \geq 0, \quad s^{\max} - s^* \geq 0, \quad {y^*}^T(s^{\max} - s^*) = 0, \tag{3.5}$$

和

$$z^* \geq 0, \quad d^* - d^{\min} \geq 0, \quad {z^*}^T(d^* - d^{\min}) = 0. \tag{3.6}$$

记

$$u = \begin{pmatrix} y \\ z \end{pmatrix} \quad \text{和} \quad F(u) = \begin{pmatrix} s^{\max} - s(u) \\ d(u) - d^{\min} \end{pmatrix}. \tag{3.7}$$

根据以上分析, 我们的任务可以归结为求解以下的隐式互补问题

$$u \geq 0, \quad F(u) \geq 0, \quad u^T F(u) = 0. \tag{3.8}$$

这里我们所说的'隐式' 是指我们不知道函数 F 的显式表达式而只能对给定的 $u \geq 0$ 观察到 $F(u)$ 的值. 本文的隐式互补问题是带有附加约束 (3.4) 的空间价格平衡问题 (3.2) 的对偶问题. 下面我们对函数 t_{ij}, h_i^s 和 h_j^d 作一定的假设后讨论隐式互补问题 (3.8) 的性质.

假设

A1. 对任意的 $i = 1, \ldots, m$ 和 $j = 1, \ldots, n$, 交易费用 $t_{ij}(x_{ij})$ 是交易量 x_{ij} 的非减函数.

A2. h_i^s 和 h_j^d 分别是 s_i 和 d_j 的一致严格增和一致严格减函数.

这样的假设应该说是合理的. 原因是由于道路拥挤, 单位交易费用(其中大部分为运输费用) 不会因交易量增加而减小, 资源地的采购价会因“采购量”

的增大(货俏)而被生产者提高, 需求地的销售价会随“到货量”的增加而降低. 由以上的假设, 隐式互补问题 (3.8) 中的向量函数 $F(u)$ 是单调和 Lipschitz 连续的.

3.3 求解隐式互补问题的直接迭代方法

对映射 F 是单调和 Lipschitz 连续的隐式互补问题 (3.8), 文献中已有的在迭代过程中只用到 $F(u)$ 的值的方法, 主要是外梯度法 (Extra-gradient Method) 和投影收缩算法 (Projection and Contraction Method). 我们分别在第二讲和第三讲中已经做了介绍. 对于这一节讨论的隐式互补问题, 实际问题也只为我们提供了对应于自变量 $u \geq 0$ 的函数值 $F(u)$ 这样的信息.

由于每调用一次函数值就等同要进行一次税收和补贴政策的调整，实际问题要求我们在求解过程中尽可能减少调用函数值的次数.

♣ 注意到预测过程中至少调用二次函数值, 分别调用 $F(u^k)$ 和 $F(\tilde{u}^k)$ 以观察 $\tilde{u}^k$ 能否被接受为预测点. 在 β_k 选取适当, 进行了一次试探 $\tilde{u}^k$ 就被接受为预测点的情况时则预测过程恰好调用了二次函数值.

3.4 初步的数值试验情况

我们感兴趣的是用 §4 中介绍的单步投影梯度法, 求解保护资源和保障供给的经济平衡问题 (3.8). 为此, 对给定的 $y \in R^m_+$ 和 $z \in R^n_+$, 经营者们要解一个原问题 (3.2). 我们对原问题 (3.2), 按照 (3.3) 中的方式定义函数 h^s, h^d 和 t. 设 $m = 20$, $n = 50$. 对 $i = 1, \ldots, m$ 和$j = 1, \ldots, n$, 取

$$
\begin{array}{lll}
a_i \in (1,2), & b_j \in (1,2), & c_{ij} \in (0.002, 0.005), \\
\xi_i \in (300, 400), & \eta_j \in (600, 700), & \zeta_{ij} \in (10, 20),
\end{array}
$$

为一定范围中的随机数. 这样, 对给定的 y 和 z, 经营者们解的原问题 (3.2) 是一个单调对称的线性互补问题, 相当于变量个数为 1000 的带非负约束的凸二次规划. 这个本该由经营者们求解的问题, 也可用 §3 中介绍的方法去求解.

为保护资源和保障供给, 分别取 $s^{\max}$ 和 $d^{\min}$ 的每个分量均为 150 和 40. 我们要求解的隐式对偶问题的变量个数是 70. 计算以 $u^0 = 0$ 为初始点. 由于求解互补问题 (3.8) 时可以将 $\|\min\{u^k, F(u^k)\}\|_\infty$ 作为误差的一种度量, 我们在下表中给出对精度 $\|\min\{u^k, F(u^k)\}\|_\infty$ 不同要求时, 不同方法所需要的迭代次数和调用函数值 $F(u)$ 的次数.

不同方法达到同样精度要求的迭代次数和调用 $F(u)$ 的次数

误差精度	外梯度方法		投影收缩算法D1		投影收缩算法D2		投影梯度法	
$\|e(u)\|_\infty$	迭代次数	调用梯度次数	迭代次数.	调用梯度次数	迭代次数	调用梯度次数	迭代次数	调用梯度次数
1	35	72	5	13	4	10	3	4
0.1	53	108	6	15	6	14	5	6
0.01	64	130	8	19	7	16	8	9
0.001	71	144	12	27	10	20	10	11

♣ 由于采用了精化的调比产生 β_k 的策略, 外梯度方法和投影收缩算法的每次迭代几乎只调用 2 次函数值. 就像在第二, 第三讲分析的, 投影收缩算法比外梯度方法收敛快. 外梯度方法的迭代次数和调用 $F(u)$ 的次数都比投影收缩算法的 6–7 倍.

♣ 在投影收缩算法中, 采用 d_2 方向的比采用 d_1 方向的快一些.

♣ 投影梯度法需要的迭代次数与投影收缩算法相当, 但调用 $F(u)$ 的次数比投影收缩算法少一半以上. 求解这类问题, 调用一次 $F(u)$, 相当于调整一次政策. 因此, 在上述算法中, 投影梯度法效率是最高的.

♣ 对变分不等式使用投影梯度法, 要求算子 F 是某个凸函数的梯度.

4 An accelerated two-steps P-G method

基于 Nesterov [10] 的思想, 采用 [1] 类似的做法, 可以构造一个只用梯度的快速算法. 这类算法在生成序列 $\{x^k\}$ 的同时, 还生成一个辅助序列 $\{y^k\}$.

A two-steps projected gradient method

Take $\beta > 0$, $x^1 \in R^n$. Set $y^1 = x^1$, $t_1 = 1$.

Step k. ($k \geq 1$) With given (x^k, y^k), let

$$x^{k+1} = P_\Omega[y^k - \beta_k g(y^k)], \tag{4.1a}$$

where the step size β_k is chosen to satisfy

$$(y^k - x^{k+1})^T (g(y^k) - g(x^{k+1})) \leq \frac{1}{2\beta_k} \|y^k - x^{k+1}\|^2. \tag{4.1b}$$

Set

$$y^{k+1} = x^{k+1} + \Big(\frac{t_k - 1}{t_{k+1}}\Big)(x^{k+1} - x^k), \tag{4.1c}$$

where

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}. \tag{4.1d}$$

The method is called two-steps projected gradient method because each iteration consists of two steps. The k-th iteration begins with (x^k, y^k), the first step (4.1a) produces x^{k+1} and the second one (4.1c) updates y^{k+1}.

It is assumed that the positive sequence $\{\beta_k\}$ is non-increasing.

We show that the proposed two-steps projected gradient method is convergent with the iteration-complexity $O(1/k^2)$. The proof is similar as those in [1].

Lemma 4.1 *Let x^{k+1} be given by* (4.1a) *and the step size condition* (4.1b) *be satisfied. Then we have*

$$2\beta_k(f(x)-f(x^{k+1})) \geq \|y^k-x^{k+1}\|^2+2(x^{k+1}-y^k)^T(y^k-x), \ \forall\, x \in \Omega. \quad (4.2)$$

Proof. By setting $x^k = y^k$ and $\nu = \frac{1}{2}$ in (2.4a) and (2.4b), we get (4.1a) and (4.1b). Therefore, substituting $x^k = y^k$ and $\nu = \frac{1}{2}$ in (2.6), we get

$$\beta_k(f(x) - f(x^{k+1})) \geq (x - x^{k+1})^T(y^k - x^{k+1}) - \frac{1}{2}\|y^k - x^{k+1}\|^2, \ \forall\, x \in \Omega.$$

The above inequality can be rewritten as (4.2) and the lemma is proved. □

To derive the iteration-complexity of the two-steps projected gradient method, we need to

prove some properties of the corresponding sequence.

Lemma 4.2 *The sequences $\{x^k\}$ and $\{y^k\}$ generated by the proposed two-steps projected gradient method satisfy*

$$2\beta_k t_k^2 v_k - 2\beta_{k+1} t_{k+1}^2 v_{k+1} \geq \|u^{k+1}\|^2 - \|u^k\|^2, \ \forall k \geq 1, \quad (4.3)$$

where $v_k := f(x^{k+1}) - f(x^*)$ and $u^k := t_k x^{k+1} - (t_k - 1)x^k - x^*$.

Proof. By using Lemma 4.1 for $k+1$, $x = x^{k+1}$ and $x = x^*$ we get

$$2\beta_{k+1}\big(f(x^{k+1})-f(x^{k+2})\big) \geq \|y^{k+1}-x^{k+2}\|^2+2(x^{k+2}-y^{k+1})^T(y^{k+1}-x^{k+1}),$$

and

$$2\beta_{k+1}\big(f(x^*) - f(x^{k+2})\big) \geq \|y^{k+1} - x^{k+2}\|^2 + 2(x^{k+2} - y^{k+1})^T(y^{k+1} - x^*).$$

Using the definition of v_k, we get

$$2\beta_{k+1}(v_k - v_{k+1}) \geq \|y^{k+1} - x^{k+2}\|^2 + 2(x^{k+1} - y^{k+1})^T(y^{k+1} - x^{k+2}), \quad (4.4)$$

and

$$-2\beta_{k+1}v_{k+1} \geq \|y^{k+1} - x^{k+2}\|^2 + 2(x^* - y^{k+1})^T(y^{k+1} - x^{k+2}). \quad (4.5)$$

To get a relation between v_k and v_{k+1}, we multiply (4.4) by $(t_{k+1}-1)$ and add it to (4.5):

$$2\beta_{k+1}\big((t_{k+1}-1)v_k - t_{k+1}v_{k+1}\big)$$
$$\geq t_{k+1}\|x^{k+2}-y^{k+1}\|^2 + 2(x^{k+2}-y^{k+1})^T\big(t_{k+1}y^{k+1}-(t_{k+1}-1)x^{k+1}-x^*\big).$$

Multiplying the last inequality by t_{k+1} and using

$$t_k^2 = t_{k+1}^2 - t_{k+1} \qquad \big(\text{and thus} \quad t_{k+1} = (1+\sqrt{1+4t_k^2}\,)/2 \quad \text{as in} \quad (4.1\text{d})\big),$$

which yields

$$\begin{aligned} 2\beta_{k+1}\big(t_k^2 v_k - t_{k+1}^2 v_{k+1}\big) & \\ \geq \quad & \|t_{k+1}(x^{k+2}-y^{k+1})\|^2 \\ & +2t_{k+1}(x^{k+2}-y^{k+1})^T\big(t_{k+1}y^{k+1}-(t_{k+1}-1)x^{k+1}-x^*\big). \end{aligned}$$

Applying the relation

$$\|a-b\|^2 + 2(a-b)^T(b-c) = \|a-c\|^2 - \|b-c\|^2$$

to the right-hand side of the last inequality with

$$a := t_{k+1}x^{k+2}, \qquad b := t_{k+1}y^{k+1}, \qquad c := (t_{k+1}-1)x^{k+1}+x^*,$$

and using the fact $2\beta_k t_k^2 v_k \geq 2\beta_{k+1}t_k^2 v_k$ (since $\{\beta_k\}$ is non-increasing), we get

$$\begin{aligned} 2\beta_k t_k^2 v_k - 2\beta_{k+1}t_{k+1}^2 v_{k+1} & \\ \geq \quad & \|t_{k+1}x^{k+2}-(t_{k+1}-1)x^{k+1}-x^*\|^2 \\ & -\|t_{k+1}y^{k+1}-(t_{k+1}-1)x^{k+1}-x^*\|^2. \end{aligned}$$

In order to write the above inequality in the form (4.3) with

$$u^k = t_k x^{k+1} - (t_k-1)x^k - x^*,$$

we need only to set

$$t_{k+1}y^{k+1}-(t_{k+1}-1)x^{k+1}-x^* = t_k x^{k+1}-(t_k-1)x^k - x^*.$$

From the last equality we obtain

$$y^{k+1} = x^{k+1} + \Big(\frac{t_k-1}{t_{k+1}}\Big)(x^{k+1}-x^k).$$

This is just the form (4.1c) in the accelerated two-steps version of the projected gradient method. $\square$

To proceed the proof of the main theorem, we need the following Lemma 4.3 and Lemma

4.4, which have also been considered in [1]. We omit their proofs as they are trivial.

Lemma 4.3 *Let $\{a_k\}$ and $\{b_k\}$ be positive sequences of reals satisfying*

$$a_k - a_{k+1} \geq b_{k+1} - b_k \quad \forall\, k \geq 1.$$

Then, $a_k \leq a_1 + b_1$ for every $k \geq 1$.

Lemma 4.4 *The positive sequence $\{t_k\}$ generated by*

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \qquad \textit{with} \qquad t_1 = 1$$

satisfies

$$t_k \geq \frac{k+1}{2}, \quad \forall\, k \geq 1.$$

Now, we are ready to show that the proposed two-steps projected gradient method is convergent with the rate $O(1/k^2)$.

Theorem 4.1 *Let $\{x^k\}$ and $\{y^k\}$ be generated by the proposed two-steps projected gradient method. Then, for any $k \geq 1$, we have*

$$f(x^k) - f(x^*) \leq \frac{2\|x^1 - x^*\|^2}{\beta_k k^2}, \quad \forall\, x^* \in \Omega^*. \tag{4.6}$$

Proof. Let us define the quantities

$$a_k := 2\beta_k t_k^2 v_k, \qquad b_k := \|u^k\|^2.$$

By using Lemma 4.2 and Lemma 4.3, we obtain

$$2\beta_k t_k^2 v_k \leq a_1 + b_1,$$

which combined with the definition v_k and $t_k \geq (k+1)/2$ (by Lemma 4.4) yields

$$f(x^{k+1}) - f(x^*) = v_k \leq \frac{2(a_1 + b_1)}{\beta_k (k+1)^2} \leq \frac{2(a_1 + b_1)}{\beta_{k+1}(k+1)^2}. \tag{4.7}$$

Since $t_1 = 1$, and using the definition of u_k given in Lemma 4.2, we have

$$a_1 = 2\beta_1 t_1^2 v_1 = 2\beta_1 v_1 = 2\beta_1\big(f(x^2) - f(x^*)\big), \qquad b_1 = \|u^1\|^2 = \|x^2 - x^*\|^2.$$

Setting $x = x^*$ and $k = 1$ in (4.2), we have

$$\begin{aligned} 2\beta_1(f(x^2) - f(x^*)) &\leq 2(y^1 - x^*)^T(y^1 - x^2) - \|y^1 - x^2\|^2 \\ &= \|y^1 - x^*\|^2 - \|x^2 - x^*\|^2. \end{aligned}$$

Therefore, we have

$$\begin{aligned} a_1 + b_1 &= 2\beta_1(f(x^2) - f(x^*)) + \|x^2 - x^*\|^2 \\ &\le \|y^1 - x^*\|^2 - \|x^2 - x^*\|^2 + \|x^2 - x^*\|^2 \\ &= \|x^1 - x^*\|^2. \end{aligned}$$

Substituting it in (4.7), the assertion is proved. $\square$

Based on Theorem 2.2, for obtaining an ε-optimal solution (denoted by $\tilde{x}$) in the sense that $f(\tilde{x}) - f(x^*) \le \varepsilon$, the number of iterations required by the proposed two-steps projected gradient method is at most $\lceil C/\sqrt{\varepsilon} - 1 \rceil$ where $C = 2\|x^1 - x^*\|^2/\beta$.

需要说明的是, 对这一讲 §5 中的问题, 我们并不提倡用这个附录中的快速方法, 原因是在 (4.1) 的 k-次迭代中, 需要至少用到两次梯度的信息, $g(y^k)$ 和 $g(x^{k+1})$. 这里的 $g(\cdot)$ 相当于 §5 中的 $F(\cdot)$. 在实际问题中, $F(\cdot)$ 的获取往往是代价不菲的. 此外, (4.1b) 中要求

$$(y^k - x^{k+1})^T(g(y^k) - g(x^{k+1})) \le \tfrac{1}{2\beta_k}\|y^k - x^{k+1}\|^2,$$

并要求 $\{\beta_k\}$ 单调不增. 这些条件远不如单步投影梯度法中相应的条件 (2.4b) 宽松. 单步投影梯度法计算实践说明限制 $\{\beta_k\}$ 单调不增会使收敛变慢许多.

References

[1] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Science, 2 (2009), pp. 183-202.

[2] Y.H. Dai and Y. Yuan, Alternate minimization gradient method. *IMA J. Numerical Analysis* **23** (2003) 377–393.

[3] R. Fletcher, *Practical Methods of Optimization*, Second Edition, John Wiley & Sons, 1987.

[4] B.S. He, *A new method for a class of linear variational inequalities*, Math. Progr., **66**, pp. 137-144, 1994.

[5] B.S. He, *Solving a class of linear projection equations*, Numerische Mathematik, **68**, pp. 71-80, 1994.

[6] B.S He and L-Z Liao, *Improvements of some projection methods for monotone nonlinear variational inequalities*, Journal of Optimization Theory and Applications, **112**, pp. 111-128, 2002

[7] G. M. Korpelevich, The Extragradient Method for Finding Saddle Points and Other Problems, *Ekonomika i Matematchskie Metody* **12** (1976) 747-756.

[8] P. Marcotte, G. Margquis and L. Zubieta, *A Newton-SOR method for spatial price equilibrium*, Transportation Science **26** (1992) 36-47.

[9] A. Nagurney, *An Algorithm for the single commodity spatial price equilibrium problem*, Reg. Sci. and Urban Econ. **16** (1986) 573-588.

[10] Y. E. Nesterov, A method for solving the convex programming problem with convergence rate $O(1/k^2)$, Dokl. Akad. Nauk SSSR, 269 (1983), pp. 543-547.

[11] Y. E. Nesterov, On an approach to the construction of optimal methods ofminimization ofsmooth convex functions, Ekonom. Mat. Metody, 24 (1988), pp. 509-517.

[12] P. A. Samuelson, *Spatial price equilibrium and linear programming*, Amer. Ecom. Rev. **42** (1952) 283-303.

Contraction Methods for Convex Optimization and Monotone Variational Inequalities – No.11

Alternating direction methods of multipliers for separable convex programming

Bingsheng He

Department of Mathematics

Nanjing University
hebma@nju.edu.cn

1 Structured constrained convex optimization

We consider the following structured constrained convex optimization problem

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

where $\theta_1(x):\Re^{n_1}\to\Re,\ \theta_2(y):\Re^{n_2}\to\Re$ are convex functions (but not necessary smooth), $A\in\Re^{m\times n_1}$, $B\in\Re^{m\times n_2}$ and $b\in\Re^m$, $\mathcal{X}\subset\Re^{n_1}$, $\mathcal{Y}\subset\Re^{n_2}$ are given closed convex sets. It is clear that the split feasibility problem:

Find a point $x\in\mathcal{X}$ such that $Ax\in\mathcal{B}$,

can be formulated as a special problem of (1.1) with $\theta_1(x)=\theta_2(y)=0$. Find (x,y) such that

$$\{Ax-y=0,\ x\in\mathcal{X},\ y\in\mathcal{B}\}. \tag{1.2}$$

Let λ be the Lagrangian multiplier for the linear constraints $Ax+By=b$ in (1.1), the Lagrangian function of this problem is

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

which is defined on $\mathcal{X} \times \mathcal{Y} \times \Re^m$. Let (x^*, y^*, λ^*) be an saddle point of the Lagrangian function, then $(x^*, y^*, \lambda^*) \in \Omega$ and it satisfies

$$\left\{ \begin{array}{l} \theta_1(x) - \theta_1(x^*) + (x - x^*)^T(-A^T\lambda^*) \geq 0, \\ \theta_2(y) - \theta_2(y^*) + (y - y^*)^T(-B^T\lambda^*) \geq 0, \quad \forall\, (x, y, \lambda) \in \Omega, \\ (\lambda - \lambda^*)^T(Ax^* + By^* - b) \geq 0, \end{array} \right. \tag{1.3}$$

where

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

By denoting

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

and

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

the first order optimal condition (1.3) can be written in a compact form such as

$$w^* \in \Omega, \ \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \ \forall\, w \in \Omega. \tag{1.4}$$

Note that the mapping F is monotone. We use Ω^* to denote the solution set of the variational inequality (1.4). For convenience we use the notations

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

Augmented Lagrangian Method to structured COP

Augmented Lagrangian Method is one of the attractive methods for nonlinear optimization as demonstrated in Chapter 17 of [21]. We try to use the Augmented Lagrangian Method to solve (1.1) and set

$$M = (A, B) \quad \text{and} \quad \mathcal{U} = \mathcal{X} \times \mathcal{Y}.$$

Now, the problem (1.1) is rewritten as

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

For given $\beta > 0$, the augmented Lagrangian function of (1.5) is

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

which defined on $\Omega = \mathcal{U} \times \Re^m$. Directly applied Augmented Lagrangian Method to the problem (1.5), the k-th iteration begins with λ^k, obtain

$$u^{k+1} = \text{Argmin}\{\mathcal{L}_A(u,\lambda^k) \,|\, u \in \mathcal{U}\}, \tag{1.6}$$

and then update the new iterate by

$$\lambda^{k+1} = \lambda^k - \beta(Mu^{k+1} - b). \tag{1.7}$$

Note that $u^{k+1} \in \mathcal{U}$ generated by (1.6) satisfies

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

By using (1.7) in the last inequality, we obtain

$$u^{k+1} \in \mathcal{U}, \quad \theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T\big(-M^T\lambda^{k+1}\big) \ge 0, \ \ \forall u \in \mathcal{U}. \tag{1.8}$$

Combining (1.8) and (1.7), we get $w^{k+1} \in \Omega$ and for any $w \in \Omega$, it holds that

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} u - u^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -M^T\lambda^{k+1} \\ Mu^{k+1} - b \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{1}{\beta}(\lambda^{k+1} - \lambda^k) \end{pmatrix} \right\} \ge 0.$$

Substituting $w = w^*$ in the above variational inequality and using the notation of $F(w)$, we get

$$\begin{aligned} &(\lambda^{k+1} - \lambda^*)^T(\lambda^k - \lambda^{k+1}) \\ &\quad \ge \quad \beta\{(w^{k+1} - w^*)^T F(w^{k+1}) + \theta(u^{k+1}) - \theta(u^*)\}. \end{aligned} \tag{1.9}$$

Using the monotonicity of F and the fact

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

we derive that the right hand side of (1.9) is non-negative. Therefore, we have

$$(\lambda^{k+1} - \lambda^*)^T(\lambda^k - \lambda^{k+1}) \ge 0, \quad \forall\, \lambda^* \in \Lambda^*. \tag{1.10}$$

It follows from (1.10) that

$$\|\lambda^k - \lambda^*\|^2 = \|(\lambda^{k+1} - \lambda^*) + (\lambda^k - \lambda^{k+1})\|^2 \ge \|\lambda^{k+1} - \lambda^*\|^2 + \|\lambda^k - \lambda^{k+1}\|^2.$$

We get the nice convergence property:

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

Summarize the Augmented Lagrangian Method to structured COP

For given λ^k, $u^{k+1} = (x^{k+1}, y^{k+1})$ is the solution of the following problem

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

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

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

Shortcoming $\quad$ The structure property is not used !

By using the augmented Lagrangian method for the structured problem (1.5), the k-th iteration is from λ^k to λ^{k+1}. The variable $u = (x, y)$ is only an intermediate variable.

2 Alternating Direction Method of Multipliers

To overcome the shortcoming the ALM for the problem (1.1), we use the alternating direction method. The main idea is splitting the subproblem (1.6) in two parts and only the x-part is the intermediate variable. Thus the iteration begins with $v^0 = (y^0, \lambda^0)$.

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

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

$$x^{k+1} = \text{Argmin} \left\{ \begin{array}{c} \theta_1(x) - (\lambda^k)^T(Ax + By^k - b) \\ + \frac{\beta}{2}\|Ax + By^k - b\|^2 \end{array} \middle| x \in \mathcal{X} \right\} \tag{2.1a}$$

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

$$y^{k+1} = \text{Argmin} \left\{ \begin{array}{c} \theta_2(y) - (\lambda^k)^T(Ax^{k+1} + By - b) \\ + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 \end{array} \middle| y \in \mathcal{Y} \right\} \tag{2.1b}$$

$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \tag{2.1c}$$

Advantages $\quad$ The sub-problems (2.1a) and (2.1b) are separately solved one by one.

Remark 2.1 *The sub-problems* (2.1a) *and* (2.1b) *is equivalent to*

$$x^{k+1} = \textit{Argmin}\left\{\theta_1(x) + \tfrac{\beta}{2}\|(Ax + By^k - b) - \tfrac{1}{\beta}\lambda^k\|^2 \big| x \in \mathcal{X}\right\} \tag{2.2a}$$

and

$$y^{k+1} = \textit{Argmin}\left\{\theta_2(y) + \tfrac{\beta}{2}\|(Ax^{k+1} + By - b) - \tfrac{1}{\beta}\lambda^k\|^2 \big| y \in \mathcal{Y}\right\} \tag{2.2b}$$

respectively. Note that the equation (2.1c) *can be written as*

$$(\lambda - \lambda^{k+1})\{(Ax^{k+1} + By^{k+1} - b) + \frac{1}{\beta}(\lambda^{k+1} - \lambda^k)\} \ge 0, \ \forall \lambda \in \Re^m. \tag{2.2c}$$

Analysis Note that the solution of (2.1a) and (2.1b) satisfies

$$\begin{aligned} x^{k+1} \in \mathcal{X}, \quad & \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \\ & \left\{-A^T\lambda^k + \beta A^T\big(Ax^{k+1} + By^k - b\big)\right\} \ge 0, \ \forall\, x \in \mathcal{X} \end{aligned} \tag{2.3a}$$

and

$$\begin{aligned} y^{k+1} \in \mathcal{Y}, \quad & \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \\ & \left\{-B^T\lambda^k + \beta B^T\big(Ax^{k+1} + By^{k+1} - b\big)\right\} \ge 0, \ \forall\, y \in \mathcal{Y}, \end{aligned} \tag{2.3b}$$

respectively. Substituting λ^{k+1} (see (2.1c)) in (2.3) (eliminating λ^k in (2.3)), we get

$$\begin{aligned} x^{k+1} \in \mathcal{X}, \quad & \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \\ & \left\{-A^T\lambda^{k+1} + \beta A^T B(y^k - y^{k+1})\right\} \ge 0, \ \forall\, x \in \mathcal{X}, \end{aligned} \tag{2.4a}$$

and

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^{k+1}\} \ge 0, \ \forall\, y \in \mathcal{Y}. \tag{2.4b}$$

For analysis convenience, we rewrite (2.4) as $u^{k+1} = (x^{k+1}, y^{k+1}) \in \mathcal{X} \times \mathcal{Y}$.

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

Combining the last inequality with (2.2c), we have $w^{k+1} \in \Omega$ and

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

$$\left. +\beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k - y^{k+1}) + \begin{pmatrix} 0 & 0 \\ \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} y^{k+1} - y^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} \right\} \geq 0, \quad (2.5)$$

for any $w \in \Omega$. The above inequality can be rewritten as $w^{k+1} \in \Omega$ and

$$\theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) + \beta \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k - y^{k+1})$$

$$\geq \begin{pmatrix} y - y^{k+1} \\ \lambda - \lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} y^k - y^{k+1} \\ \lambda^k - \lambda^{k+1} \end{pmatrix}, \ \forall\, w \in \Omega. \quad (2.6)$$

3 Convergence of ADMM

Based on the analysis in the last section, we have the following lemma.

Lemma 3.1 *Let* $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ *be generated by* (2.1) *from the given* $v^k = (y^k, \lambda^k)$. *Then, we have*

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \geq (w^{k+1} - w^*)^T \eta(y^k, y^{k+1}), \qquad (3.1)$$

where

$$\eta(y^k, y^{k+1}) = \beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k - y^{k+1}) \qquad (3.2)$$

and

$$H = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix}. \qquad (3.3)$$

Proof. Setting $w = w^*$ in (2.6), and using H and $\eta(y^k, y^{k+1})$, we get

$$\begin{aligned}
&(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \\
&\quad \geq \quad (w^{k+1} - w^*)^T \eta(y^k, y^{k+1}) \\
&\qquad\quad +\theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^{k+1}).
\end{aligned} \tag{3.4}$$

Since F is monotone, it follows that

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

The last inequality is due to $w^{k+1} \in \Omega$ and $w^* \in \Omega^*$ (see (1.4)). Substituting it in (3.4), the lemma is proved. $\square$

Lemma 3.2 *Let* $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ *be generated by* (2.1) *from the given* $v^k = (y^k, \lambda^k)$. *Then, we have*

$$(w^{k+1} - w^*)^T \eta(y^k, y^{k+1}) = (\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}), \tag{3.5}$$

and

$$(\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \geq 0. \tag{3.6}$$

Proof. By using $\eta(y^k, y^{k+1})$ (see (3.2)), $Ax^* + By^* = b$ and (2.1c), we have

$$\begin{aligned}
&(w^{k+1} - w^*)^T \eta(y^k, y^{k+1}) \\
&\quad = \quad (B(y^k - y^{k+1}))^T \beta\{(Ax^{k+1} + By^{k+1}) - (Ax^* + By^*)\} \\
&\quad = \quad (B(y^k - y^{k+1}))^T \beta(Ax^{k+1} + By^{k+1} - b) \\
&\quad = \quad (\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}).
\end{aligned}$$

Because (2.4b) is true for the k-th iteration and the previous iteration, we have

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T \lambda^{k+1}\} \geq 0, \ \forall\, y \in \mathcal{Y}, \tag{3.7}$$

and

$$\theta_2(y) - \theta_2(y^k) + (y - y^k)^T \{-B^T \lambda^k\} \geq 0, \ \forall\, y \in \mathcal{Y}, \tag{3.8}$$

Setting $y = y^k$ in (3.7) and $y = y^{k+1}$ in (3.8), respectively, and then adding the two resulting inequalities, we get

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

The assertion of this lemma is proved. $\square$

Even though H is positive semi-definite (see (3.3) when B is not full column rank), in this lecture we use $\|v-\tilde{v}\|_H$ to denote that

$$\|v-\tilde{v}\|_H^2 = (v-\tilde{v})^T H(v-\tilde{v}) = \beta\|B(y-\tilde{y})\|^2 + \frac{1}{\beta}\|\lambda-\tilde{\lambda}\|^2.$$

Lemma 3.3 *Let* $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ *be generated by* (2.1) *from the given* $v^k = (y^k, \lambda^k)$. *Then, we have*

$$(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \geq 0, \quad \forall\, v^* \in \mathcal{V}^*. \tag{3.9}$$

Proof. The assertion follows (3.1), (3.5) and (3.6) directly. □

Theorem 3.1 *Let* $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ *be generated by* (2.1) *from the given* $v^k = (y^k, \lambda^k)$. *Then, we have*

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \ \forall\, v^* \in \mathcal{V}^*. \tag{3.10}$$

Proof. By using (3.9), we have

$$\begin{aligned}
\|v^k - v^*\|_H^2 &= \|(v^{k+1} - v^*) + (v^k - v^{k+1})\|_H^2 \\
&= \|v^{k+1} - v^*\|_H^2 + 2(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \\
&\quad + \|v^k - v^{k+1}\|_H^2 \\
&\geq \|v^{k+1} - v^*\|_H^2 + \|v^k - v^{k+1}\|_H^2,
\end{aligned}$$

and thus (3.10) is proved. □

The inequality (3.10) is essential for the convergence of the alternating direction method. It tells us that the alternating direction method is a contraction method. Multiplying a factor $1/\beta$, it can be written as

$$\left\| \begin{matrix} B(y^{k+1} - y^*) \\ \frac{1}{\beta}(\lambda^{k+1} - \lambda^*) \end{matrix} \right\|^2 \leq \left\| \begin{matrix} B(y^k - y^*) \\ \frac{1}{\beta}(\lambda^k - \lambda^*) \end{matrix} \right\|^2 - \left\| \begin{matrix} B(y^k - y^{k+1}) \\ \frac{1}{\beta}(\lambda^k - \lambda^{k+1}) \end{matrix} \right\|^2, \ \forall\, v^* \in \mathcal{V}^*.$$

This result is included in Theorem 1 of [14] as a special case for fixed β and $\gamma \equiv 1$.

4 The extended ADMM

In the extended ADM, the k-th iteration begins with (y^k, λ^k). However, we take the solution of the classical ADM as a predictor, and denote it by $(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$.

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

$$\tilde{x}^k = \text{Argmin} \left\{ \begin{array}{c} \theta_1(x) - (\lambda^k)^T(Ax + By^k - b) \\ + \frac{\beta}{2}\|Ax + By^k - b\|^2 \end{array} \middle| x \in \mathcal{X} \right\} \tag{4.1a}$$

2. Then, use λ^k and the obtained $\tilde{x}^k$, $\tilde{y}^k$ is the solution of the following problem

$$\tilde{y}^k = \text{Argmin} \left\{ \begin{array}{c} \theta_2(y) - (\lambda^k)^T(A\tilde{x}^k + By - b) \\ + \frac{\beta}{2}\|A\tilde{x}^k + By - b\|^2 \end{array} \middle| y \in \mathcal{Y} \right\} \tag{4.1b}$$

3. Finally,

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k - b). \tag{4.1c}$$

Based on the predictor $(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$, we consider how to produce the new iterate $v^{k+1} = (y^{k+1}, \lambda^{k+1})$ and drive it more close to the set $\mathcal{V}^*$.

According to the same analysis in the last section (see (2.5)) we have $(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ and

$$\theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{\lambda}^k \\ -B^T\tilde{\lambda}^k \\ A\tilde{x}^k + B\tilde{y}^k - b \end{pmatrix} + \beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k - \tilde{y}^k) \right.$$
$$\left. + \begin{pmatrix} 0 & 0 \\ \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} \tilde{y}^k - y^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix} \right\} \geq 0, \ \forall w \in \Omega. \tag{4.2}$$

Based on the above analysis, we have the following lemma.

Lemma 4.1 *Let $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ be generated by* (4.1) *from the given $v^k = (y^k, \lambda^k)$. Then, we have*

$$(\tilde{v}^k - v^*)^T H(v^k - \tilde{v}^k) \geq (\tilde{w}^k - w^*)^T \eta(y^k, \tilde{y}^k), \tag{4.3}$$

where

$$\eta(y^k, \tilde{y}^k) = \beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k - \tilde{y}^k) \tag{4.4}$$

and H *is the same as defined in* (3.3)*.*

Proof. The prof is similar as those for Lemma 3.1 and thus omitted. □

Similarly as in (3.5), by using $\eta(y^k, \tilde{y}^k)$ (see (4.4)) and $Ax^* + By^* = b$, we have

$$(\tilde{w}^k - w^*)^T \eta(y^k, \tilde{y}^k) = (\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k). \tag{4.5}$$

Lemma 4.2 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (4.1) *from the given* $v^k = (y^k, \lambda^k)$*. Then, we have*

$$(v^k - v^*)^T H(v^k - \tilde{v}^k) \geq \varphi(v^k, \tilde{v}^k), \quad \forall\, v^* \in \mathcal{V}^*. \tag{4.6}$$

and

$$\varphi(v^k, \tilde{v}^k) = \|v^k - \tilde{v}^k\|_H^2 + (\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k). \tag{4.7}$$

Proof. It follows from (4.3) and (4.5) that

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

Assertion (4.6) follows from the last inequality and the definition of $\varphi(v^k, \tilde{v}^k)$ and the lemma is proved. □

Now, we observe the right hand side of (4.6). Note that

$$\begin{aligned} \varphi(v^k, \tilde{v}^k) &= \|v^k - \tilde{v}^k\|_H^2 + (\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k) \\ &= \frac{1}{2}\|v^k - \tilde{v}^k\|_H^2 + \frac{1}{2\beta}\|\beta B(y^k - \tilde{y}^k) + (\lambda^k - \tilde{\lambda}^k)\|^2 \\ &\geq \frac{1}{2}\|v^k - \tilde{v}^k\|_H^2. \end{aligned} \tag{4.8}$$

Ye-Yuan's Alternating Direction Method of Multipliers

Ye-Yuan's alternating direction method is a prediction-correction method. The predictor is generated by (4.1). The correction step is to update the new iterate.

Correction Using the $\tilde{v}^k$ produced by (4.1), update the new iterate v^{k+1} by

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

where

$$\alpha_k^* = \frac{\varphi(v^k, \tilde{v}^k)}{\|v^k - \tilde{v}^k\|_H^2} \tag{4.9b}$$

Usually, in comparison with the computational load for obtaining $(\tilde{x}^k, \tilde{y}^k)$ in (4.1), the calculation cost for step-size α_k^* is slight.

We obtain an essential inequality for convergence in the following theorem which was proved by Ye and Yuan in [29].

Theorem 4.1 *Let $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ be generated by (4.1) from the given $v^k = (y^k, \lambda^k)$ and the new iterate v^{k+1} be given by (4.9). Then we have*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \frac{\gamma(2-\gamma)}{4}\|v^k - \tilde{v}^k\|_H^2, \ \forall\, v^* \in \mathcal{V}^*. \tag{4.10}$$

Proof. By using (4.6) and (4.9), we obtain

$$\begin{aligned} &\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\ &= \|v^k - v^*\|_H^2 - \|(v^k - v^*) - \gamma\alpha_k^*(v^k - \tilde{v}^k)\|_H^2 \\ &\ge 2\gamma\alpha_k^*\varphi(v^k, \tilde{v}^k) - \gamma^2(\alpha_k^*)^2\|v^k - \tilde{v}^k\|_H^2 \\ &= \gamma(2-\gamma)(\alpha_k^*)^2\|v^k - \tilde{v}^k\|_H^2. \end{aligned} \tag{4.11}$$

In addition, it follows from (4.8) and (4.9b) that $\alpha_k^* \ge \frac{1}{2}$. Substituting this fact in (4.11), the theorem is proved. $\square$

Convergence Both (3.10) and (4.10) can be written as

$$\left\| \begin{matrix} B(y^{k+1} - y^*) \\ \frac{1}{\beta}(\lambda^{k+1} - \lambda^*) \end{matrix} \right\|^2 \le \left\| \begin{matrix} B(y^k - y^*) \\ \frac{1}{\beta}(\lambda^k - \lambda^*) \end{matrix} \right\|^2 - c_0 \left\| \begin{matrix} B(y^k - \tilde{y}^k) \\ \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) \end{matrix} \right\|^2, \ \forall\, v^* \in \mathcal{V}^*.$$

It leads to that

$$\lim_{k\to\infty} By^k = By^* \qquad \text{and} \qquad \lim_{k\to\infty} \lambda^k = \lambda^*.$$

5 Application and Numerical Experiments

5.1 Calibrating the correlation matrices

We consider to solve the following problem:

$$\min\{\frac{1}{2}\|X-C\|_F^2 \mid X \in S_+^n \cap S_B\}, \tag{5.1}$$

where

$$S_+^n = \{H \in R^{n\times n} \mid H^T = H,\ H \succeq 0\},$$

and

$$S_B = \{H \in R^{n\times n} \mid H^T = H,\ H_L \le H \le H_U\}.$$

H_L and H_U are given symmetric matrices.

Use the following Matlab Code to produce the matrices C, H_L and H_U

```
rand('state',0);     C=rand(n,n);  C=(C'+C)-ones(n,n) + eye(n);
%%% C is symmetric and C_{ij} is in (-1,1),  C_{jj} is in (0,2)  %%
HU=ones(n,n)*0.1;  HL=-HU; for i=1:n   HU(i,i)=1;  HL(i,i)=1; end;
```

The problem is converted to the following equivalent one:

$$\begin{array}{ll} \min & \frac{1}{2}\|X-C\|^2 + \frac{1}{2}\|Y-C\|^2 \\ \text{s.t} & X - Y = 0, \\ & X \in S_+^n,\ Y \in S_B. \end{array} \tag{5.2}$$

The basic sub-problems in alternating direction methods for the problem (5.2)

- For fixed Y^k and Z^k,

$$\tilde{X}^k = \text{Argmin}\{\frac{1}{2}\|X-C\|_F^2 - \mathbf{Tr}(Z^kX) + \frac{\beta}{2}\|X-Y^k\|_F^2 \mid X \in S_+^n\}$$

- With fixed $\tilde{X}^k$ and Z^k,

$$\tilde{Y}^k = \text{Argmin}\{\frac{1}{2}\|Y-C\|_F^2 + \mathbf{Tr}(Z^kY) + \frac{\beta}{2}\|\tilde{X}^k-Y\|_F^2 \mid Y \in S_B\}$$

$\tilde{X}^k$ can be directly obtained via

$$\tilde{X}^k = P_{S_+^n}\{\frac{1}{1+\beta}(\beta Y^k + Z^k + C)\}. \tag{5.3}$$

Note that

$$P_{S_+^n}(A) = U\Lambda^+U^T, \text{ where } \quad \Lambda^+ = \max(\Lambda, 0) \quad \text{and} \quad [U, \Lambda] = \text{eig}(A).$$

Similarly, $\tilde{Y}^k$ in is given by

$$\tilde{Y}^k = P_{S_B}\{\frac{1}{1+\beta}(\beta\tilde{X}^k - Z^k + C)\}. \tag{5.4}$$

$$S_B = \{H \mid H_L \le H \le H_U\}, \qquad P_{S_B}(A) = \min(\max(H_L, A), H_U)$$

The most time consuming calculation is $[U, \Lambda] =$ eig(A), $9n^3$

The main Matlab Code of an iteration in the Classical ADM

```
Y0=Y;                          Z0=Z;            k=k+1;
X=(Y0*beta+Z0+C)/(1+beta);
   [V,D]=eig(X);               D=max(0,D);      X=(V*D)*V';
Y=min(max((X*beta-Z0+C)/(1+beta),HL),HU);
Z=Z0-(X-Y)*beta;
```


The main Matlab Code of an iteration in Ye-Yuan's ADM

```
Y0=Y;                          Z0=Z;            k=k+1;
X=(Y0*beta+Z0+C)/(1+beta);
[V,D]=eig(X);                  D=max(0,D);      X=(V*D)*V';
Y=min(max((X*beta-Z0+C)/(1+beta),HL),HU);
%%%%%%%%%    Calculating the step size    %%%%%%%%%%%%%%
EY=Y0-Y;       EZ=(X-Y)*beta;
T1 = EY(:)'*EY(:);  T2 = EZ(:)'*EZ(:);   TA=T1*beta + T2/beta
T2 = (EY(:)'*EZ(:));
alpha=(TA-T2)*gammaY/TA;
Y=Y0-EY*alpha;                           Z=Z0-EZ*alpha;
```

Numerical results for problem (5.1)

C=rand(n,n); C=(C'+C)-ones(n,n) + eye(n)

H_U=ones(n,n)*0.2; $H_L = -H_U$; $H_U(jj) = H_L(jj)$=1.

Numerical Results for calibrating correlation matrix (Using Matlab EIG)

$n \times n$ Matrix	Classical ADM		Glowinski's ADM		Ye-Yuan's ADM	
$n =$	No. It	CPU Sec.	No. It	CPU Sec.	No. It	CPU Sec.
500	40	18.03	39	17.68	32	14.99
800	41	73.28	39	70.00	33	60.80
1000	43	141.69	42	138.30	34	114.67
1500	47	471.77	45	452.22	41	419.70
2000	55	1254.01	53	1206.94	45	1035.38

Numerical Results for calibrating correlation matrix (Using MeXEIG)

$n \times n$ Matrix	Classical ADM		Glowinski's ADM		Ye-Yuan's ADM	
$n =$	No. It	CPU Sec.	No. It	CPU Sec.	No. It	CPU Sec.
500	40	5.57	39	5.38	32	4.67
800	41	18.13	39	17.15	33	15.13
1000	43	34.75	42	34.00	34	28.50
1500	47	123.77	45	117.87	41	110.17
2000	55	306.32	53	294.75	45	255.72

$$\frac{\text{It. No. of Ye-Yuan's ADMM}}{\text{It. No. of Classical ADMM}} \approx \frac{5}{6}.$$

It seems that Ye-Yuan's Algorithm converges faster than primary ADM.

5.2 Application for Sparse Covariance Selection

For the details of applications in this subsection, please see the reference [30].

The problem

$$\min_X \{\mathbf{Tr}(\Sigma X) - \log(\det(X)) + \rho e^T |X| e \;\Big|\; X \in S^n_+\}, \tag{5.5}$$

The equivalent problem:

$$\begin{aligned} \min \quad & \mathbf{Tr}(\Sigma X) - \log(\det(X)) + \rho e^T |Y| e \\ \text{s.t} \quad & X - Y = 0, \\ & X \in S^n_+. \end{aligned} \tag{5.6}$$

For given Y^k and Z^k, get $(\tilde{X}^k, \tilde{Y}^k, \tilde{Z}^k)$ in the following procedure:

1. For fixed Y^k and Z^k, $\tilde{X}^k$ is the solution of the following problem

$$\min\{\mathbf{Tr}(\Sigma X) - \log(\det(X)) - \mathbf{Tr}(Z^k X) + \frac{\beta}{2}\|X - Y^k\|_F^2 \mid X \in S_+^n\}$$

2. Then, with fixed $(\tilde{X}^k, Z^k)$, $\tilde{Y}^k$ is a solution of

$$\min\{\rho e^T|Y|e + \mathbf{Tr}(Z^k Y) + \frac{\beta}{2}\|\tilde{X}^k - Y\|_F^2\}$$

3. Finally, update $\tilde{Z}^k$ by

$$\tilde{Z}^k = Z^k - \beta(\tilde{X}^k - \tilde{Y}^k). \tag{5.7}$$

Solving the X subproblem for getting $\tilde{X}^k$:

$$\tilde{X}^k = \text{Argmin}\{\frac{1}{2}\|X - [Y^k - \frac{1}{\beta}(\Sigma - Z^k)]\|_F^2 - \frac{1}{\beta}\log(\det(X)) \mid X \in S_+^n\}. \tag{5.8}$$

It should hold that $\tilde{X} \succ 0$ and thus $\tilde{X}$ is the solution of matrix equation

$$X - \left(Y^k - \frac{1}{\beta}(\Sigma - Z^k)\right) - \frac{1}{\beta}X^{-1} = 0. \tag{5.9}$$

By setting

$$A = Y^k - \frac{1}{\beta}(\Sigma - Z^k), \tag{5.10}$$

and using

$$[V, \Lambda] = \text{eig}(A) \tag{5.11}$$

in Matlab, we get

$$A = V\Lambda V^T, \qquad \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n).$$

In fact, the solution of matrix equation (5.9) should have the same eigenvectors as matrix A.

$$\tilde{X} = V\tilde{\Lambda}V^T, \qquad \tilde{\Lambda} = \text{diag}(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n). \tag{5.12}$$

It follows from (5.9) that

$$\tilde{\Lambda} - \Lambda - \frac{1}{\beta}\tilde{\Lambda}^{-1} = 0$$

and thus

$$\tilde{\lambda}_j = \frac{\lambda_j + \sqrt{\lambda_j^2 + (4/\beta)}}{2}, \quad j = 1, \ldots, n.$$

Indeed, $\tilde{\lambda}_j > 0$ and thus $\tilde{X} \succ 0$ (see (5.12)).

The main computational load for getting $\tilde{X}^k$ is the eigenvalues-vectors decomposition in (5.11).

Solving the Y subproblem for getting $\tilde{Y}^k$:

The first order condition for minimization problem

$$\min\{\rho e^T |Y| e + \mathbf{Tr}(Z^k Y) + \frac{\beta}{2}\|\tilde{X}^k - Y\|_F^2\}$$

is

$$0 \in \frac{\rho}{\beta}\partial(|Y|) + Y - (\tilde{X}^k - \frac{1}{\beta}Z^k).$$

In fact,

$$\tilde{Y}^k = (\tilde{X}^k - \frac{1}{\beta}Z^k) - P_{B_\infty^{\rho/\beta}}[\tilde{X}^k - \frac{1}{\beta}Z^k],$$

where $B_\infty^{\rho/\beta} = \{X \in \mathbf{R}^{n\times n} \mid -\frac{\rho}{\beta} \le X_{ij} \le \frac{\rho}{\beta}\}$. **The projection on a 'box' is very easy to be carried out !**

5.3 Split feasibility problem and Matrix completion

Applying ADMM to the reformulated split feasibility problem (1.2), the k-th iteration begins with $(y^k, \lambda^k) \in \mathcal{B} \times \Re^m$, and the new iterate is generated by the following procedure:

$$\begin{cases} x^{k+1} = \text{Argmin}\{\frac{1}{2}\|Ax - (y^k + \lambda^k/\beta)\|^2 \mid x \in \mathcal{X}\}, \\ y^{k+1} = P_{\mathcal{B}}[Ax^{k+1} - \lambda^k/\beta], \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} - y^{k+1}). \end{cases}$$

The x-subproblem is a problem of form $\min\{\frac{1}{2}\|Ax - a\|^2 \mid x \in \mathcal{X}\}$. After getting x^{k+1}, (y^{k+1}, λ^{k+1}) is obtained by a projection and an evaluation.

Matrix completion is to recover an unknown matrix from a sampling of its entries. For an $m \times n$ matrix M, Ω denotes the indices subset of the matrix

$$\Omega = \{(ij) \mid i \in \{1, 2, \ldots, m\},\ j \in \{1, 2, \ldots, n\}\}.$$

The mathematical form of the considered problem is

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

where $\|X\|_*$ is the nuclear norm of X [2]. It can convert to the following problem

$$\begin{aligned} \min_{X,Y} \quad & \|X\|_* \\ \textbf{s. t} \quad & X - Y = 0, \\ & Y_{ij} = M_{ij}, \forall\ (ij) \in \Omega. \end{aligned}$$

It belongs to the problem (1.1) and was successfully solved by the alternating direction methods [4].

Simple iterative scheme $+$ nice convergence properties
$\Rightarrow$ wide applications of ADMM in large scale optimization

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, **3**, 1-122, 2010.

[2] J. F. Cai, E. J. Candès and Z. W. Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optim. **20**, 1956-1982, 2010.

[3] R. H. Chan, J. F. Yang, and X. M. Yuan. Alternating direction method for image inpainting in wavelet domain. *SIAM J. Imag. Sci.*, 4:807–826, 2011.

[4] C. H. Chen, B.S. He and X. M. Yuan, Matrix completion via alternating direction methods, IMA Journal of Numerical Analysis, **3**, 227-245, 2012.

[5] E. Esser. Applications of Lagrangian-based alternating direction methods and connections to split Bregman. *CAM Report*, 09-31, UCLA, 2009.

[6] D. Gabay, Applications of the Method of Multipliers to Variational Inequalities, in *Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems*, M. Fortin and R. Glowinski, eds., North Holland, Amsterdam, The Netherlands, 299-331, 1983.

[7] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximations, Computers and Mathematics with Applications, **2**, 17-40, 1976.

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

[9] D. Goldfarb, S.Q. Ma, and K. Scheinberg. Fast alternating linearization methods for minimizing the sum of two convex functions. *Math. Program.*, (2012). DOI 10.1007/s10107-012-0530-2

[10] B.S. He, Inexact implicit methods for monotone general variational inequalities, Math. Program. Series A, **86**, 199-217, 1999.

[11] B.S. He, L-Z Liao, D.R. Han and H. Yang, A new inexact alternating directions method for monotone variational inequalities, Math. Progr., **92**, 103-118, 2002.

[12] B. S. He, M. Tao, and X. M. Yuan. Alternating direction method with Gaussian back substitution for separable convex programming. *SIAM J. Opt.*, 22(2): 313–340, 2012.

[13] B. S. He, M. H. Xu and X. M. Yuan, Solving large-scale least squares covariance matrix problems by alternating direction methods, SIAM J. Matrix Analy. Appli., 32: 136-152, 2011.

[14] B. S. He and H. Yang, Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities, Oper. Res. Let., **23**, 151-161, 1998.

[15] B.S. He, H. Yang and S.L. Wang, Alternating directions method with self-adaptive penalty parameters for monotone variational inequalities, JOTA **106**, 349-368, 2000.

[16] B.S. He and X.M. Yuan, On the $O(1/n)$ Convergence Rate of the Douglas-Rachford Alternating Direction Method, SIAM J. Numer. Anal. 50, 700-709, 2012.

[17] Z. Lu, T. K. Pong, and Y. Zhang. An alternating direction method for finding Dantzig selectors. Preprint, 2010.

[18] M. Ng, F. Wang, and X. M. Yuan. Inexact alternating direction methods for image recovery. *SIAM J. Sci. Comput.*, 33:1643–1668, 2011.

[19] M. Ng, P. A. Weiss, and X. M. Yuan. Solving constrained total-variation problems via alternating direction methods. *SIAM J. Sci. Comput.*, 32:2710–2736, 2010.

[20] S. Setzer. Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage. *Scale space and variational methods in computer vision*, 5567:464–476, 2009.

[21] J. Nocedal and S.J. Wright, Numerical Optimization, Springer Verlag, 1999.

[22] S. Setzer. Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage. *Scale space and variational methods in computer vision*, 5567:464–476, 2009.

[23] J. Sun and S. Zhang. A modified alternating direction method for convex quadratically constrained quadratic semidefinite programs. *European J. Oper. Res*, 207:1210–1220, 2010.

[24] M. Tao and X. M. Yuan. Recovering low-rank and sparse components of matrices from incomplete and noisy observations. *SIAM J. Opt.*, 21(1):57–81, 2011.

[25] Z. W. Wen, D. Goldfarb, and W. T. Yin. Alternating direction augmented Lagrangian methods for semidefinite programming. *Math. Program. Comput.*, 2:203–230, 2010.

[26] J. F. Yang and Y. Zhang. Alternating direction algorithms for L_1-problems in compressive sensing, *SIAM J. Sci. Comput.,* 38:250–278, 2011

[27] J. F. Yang, Y. Zhang, and W. Yin. A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data. *IEEE J. Selected Topics in Signal Processing*, 4:288–297, 2010

[28] J. F. Yang and X. M. Yuan. Sparese and low-rank matrix decomposition via alternating direction methods. *manuscript*, 2012.

[29] C. H. Ye and X. M. Yuan, A Descent Method for Structured Monotone Variational Inequalities, Optimization Methods and Software, **22**, 329-338, 2007.

[30] X. M. Yuan. Alternating direction method of multipliers for covariance selection models. *J. Sci. Comput.*, 51:261-273, 2012.

[31] X. Q. Zhang, M. Burger, and S. Osher. A unified primal-dual algorithm framework based on Bregman iteration. *J. Sci. Comput.*, 46(1):20–46, 2011.

凸优化和单调变分不等式的收缩算法

第十二讲: 线性化的交替方向收缩算法

Linearized alternating direction method
for separable convex programming

南京大学数学系　　何炳生
hebma@nju.edu.cn

1 Structured constrained convex optimization

We consider the following structured constrained convex optimization problem

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

where $\theta_1(x):\Re^{n_1}\to\Re,\ \theta_2(y):\Re^{n_2}\to\Re$ are convex functions (but not necessary smooth), $A\in\Re^{m\times n_1}$, $B\in\Re^{m\times n_2}$ and $b\in\Re^m$, $\mathcal{X}\subset\Re^{n_1}$, $\mathcal{Y}\subset\Re^{n_2}$ are given closed convex sets. We let $n=n_1+n_2$.

The task of solving the problem (1.1) is to find an $(x^*,y^*,\lambda^*)\in\Omega$, such that

$$\left\{\begin{array}{l}\theta_1(x)-\theta_1(x^*)+(x-x^*)^T(-A^T\lambda^*)\ge 0,\\ \theta_2(y)-\theta_2(y^*)+(y-y^*)^T(-B^T\lambda^*)\ge 0,\quad \forall\,(x,y,\lambda)\in\Omega,\\ (\lambda-\lambda^*)^T(Ax^*+By^*-b)\ge 0,\end{array}\right. \quad (1.2)$$

where

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

By denoting

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

and

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

the first order optimal condition (1.2) can be written in a compact form such as

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

Note that the mapping F is monotone. We use Ω^* to denote the solution set of the variational inequality (1.3). For convenience we use the notations

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

Alternating Direction Method is a simple but powerful algorithm that is well suited to distributed convex optimization [1]. This approach also has the benefit that one algorithm could be flexible enough to solve many problems.

Applied ADM 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

$$\boxed{x^{k+1} = \text{Argmin} \left\{ \theta_1(x) + \tfrac{\beta}{2}\|Ax + By^k - b - \tfrac{1}{\beta}\lambda^k\|^2 \,\middle|\, x \in \mathcal{X} \right\} \quad (1.4a)}$$

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

$$\boxed{y^{k+1} = \text{Argmin} \left\{ \theta_2(y) + \tfrac{\beta}{2}\|Ax^{k+1} + By - b - \tfrac{1}{\beta}\lambda^k\|^2 \,\middle|\, y \in \mathcal{Y} \right\} \quad (1.4b)}$$

$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \qquad (1.4c)$$

In some structured convex optimization (1.1), B is a scalar matrix. However, the solution of the subproblem (1.4a) does not have the closed form solution because of the general structure of the matrix A. In this case, we linearize the quadratic term of (1.4a)

$$\frac{\beta}{2}\|Ax + By^k - b - \tfrac{1}{\beta}\lambda^k\|^2$$

at x^k and add a proximal term $\frac{r}{2}\|x - a\|^2$ to the objective function. In other words, instead of (1.4a), we solve the following x subproblem:

$$\min\{\theta_1(x) + \beta x^T A^T(Ax^k + By^k - b - \tfrac{1}{\beta}\lambda^k) + \tfrac{r}{2}\|x - x^k\|^2 \,\big|\, x \in \mathcal{X}\}.$$

Based on linearizing the quadratic term of (1.4a), in this lecture, we construct the linearized alternating direction method. We still assume that the solution of the problem

$$\min\{\theta_1(x) + \frac{r}{2}\|x - a\|^2 \mid x \in \mathcal{X}\} \tag{1.5}$$

has a closed form.

2 Linearized Alternating Direction Method

In the Linearized ADM, x is not an intermediate variable. The k-th iteration of the Linearized ADM is from (x^k, y^k, λ^k) to $(x^{k+1}, y^{k+1}, \lambda^{k+1})$.

2.1 Linearized ADM

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

$$x^{k+1} = \text{Argmin}\left(\begin{array}{c}\left\{\theta_1(x) + \beta x^T A^T(Ax^k + By^k - b - \frac{1}{\beta}\lambda^k)\right. \\ \left. + \frac{r}{2}\|x - x^k\|^2 \;\middle|\; x \in \mathcal{X}\right\}\end{array}\right). \tag{2.1a}$$

2. Then, use λ^k and the obtained x^{k+1}, y^{k+1} is the solution of the following problem

$$y^{k+1} = \text{Argmin}\left\{\theta_2(y) + \frac{\beta}{2}\|Ax^{k+1} + By - b - \frac{1}{\beta}\lambda^k\|^2 \,\middle|\, y \in \mathcal{Y}\right\}. \tag{2.1b}$$

3. Finally,

$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \tag{2.1c}$$

Requirements on parameters β, r For given $\beta > 0$, choose r such that

$$rI_n - \beta A^T A \succeq 0. \tag{2.2}$$

Analysis of the optimal conditions of subproblems in (2.1)

Note that x^{k+1}, the solution of (2.1a), satisfies

$$x^{k+1} \in \mathcal{X}, \quad \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \Big\{ -A^T\lambda^k + \beta A^T\big(Ax^k + By^k - b\big) + r(x^{k+1} - x^k) \Big\} \ge 0, \ \forall\, x \in \mathcal{X}. \tag{2.3a}$$

Similarly, the solution of (2.1b) y^{k+1} satisfies

$$y^{k+1} \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \Big\{ -B^T\lambda^k + \beta B^T\big(Ax^{k+1} + By^{k+1} - b\big) \Big\} \ge 0, \ \forall\, y \in \mathcal{Y}. \tag{2.3b}$$

Substituting λ^{k+1} (see (2.1c)) in (2.3) (eliminating λ^k), we get $x^{k+1} \in \mathcal{X}$,

$$\theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \big\{ -A^T\lambda^{k+1} + \beta A^T B(y^k - y^{k+1}) + (rI_{n_1} - \beta A^T A)(x^{k+1} - x^k) \big\} \ge 0, \ \forall\, x \in \mathcal{X}, \tag{2.4a}$$

and

$$y^{k+1} \in \mathcal{Y}, \ \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \big\{ -B^T\lambda^{k+1} \big\} \ge 0, \ \forall\, y \in \mathcal{Y}. \tag{2.4b}$$

For analysis convenience, we rewrite (2.4) as the following equivalent form:

$$\theta(u) - \theta(u^{k+1}) + \begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \end{pmatrix} + \beta \begin{pmatrix} A^T \\ B^T \end{pmatrix} B(y^k - y^{k+1}) + \begin{pmatrix} rI_{n_1} - \beta A^T A & 0 \\ 0 & \beta B^T B \end{pmatrix} \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} \right\} \ge 0, \ \forall\, (x, y) \in \mathcal{X} \times \mathcal{Y}.$$

Combining the last inequality with (2.1c), we have

$$\theta(u)-\theta(u^{k+1})+\begin{pmatrix} x-x^{k+1} \\ y-y^{k+1} \\ \lambda-\lambda^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\lambda^{k+1} \\ -B^T\lambda^{k+1} \\ Ax^{k+1}+By^{k+1}-b \end{pmatrix} + \beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k-y^{k+1}) \right.$$

$$\left. + \begin{pmatrix} rI_{n_1}-\beta A^TA & 0 & 0 \\ 0 & \beta B^TB & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} x^{k+1}-x^k \\ y^{k+1}-y^k \\ \lambda^{k+1}-\lambda^k \end{pmatrix} \right\} \geq 0,$$

for all $(x,y,\lambda)\in\mathcal{X}\times\mathcal{Y}\times\Re^m$. The above inequality can be rewritten as

$$\theta(u)-\theta(u^{k+1})+(w-w^{k+1})^TF(w^{k+1})+\beta\begin{pmatrix} x-x^{k+1} \\ y-y^{k+1} \\ \lambda-\lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k-y^{k+1})$$

$$\geq \begin{pmatrix} x-x^{k+1} \\ y-y^{k+1} \\ \lambda-\lambda^{k+1} \end{pmatrix}^T \begin{pmatrix} rI_{n_1}-\beta A^TA & 0 & 0 \\ 0 & \beta B^TB & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} x^k-x^{k+1} \\ y^k-y^{k+1} \\ \lambda^k-\lambda^{k+1} \end{pmatrix}, \forall w\in\Omega. \quad (2.5)$$

2.2 Convergence of Linearized ADM

Based on the analysis in the last section, we have the following lemma.

Lemma 2.1 *Let* $w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})\in\Omega$ *be generated by* (2.1) *from the given* $w^k=(x^k,y^k,\lambda^k)$. *Then, we have*

$$(w^{k+1}-w^*)^TG(w^k-w^{k+1})\geq(w^{k+1}-w^*)^T\eta(y^k,y^{k+1}),\ \forall\, w^*\in\Omega^*, \quad (2.6)$$

where

$$\eta(y^k,y^{k+1})=\beta\begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k-y^{k+1}), \quad (2.7)$$

and

$$G=\begin{pmatrix} rI_{n_1}-\beta A^TA & 0 & 0 \\ 0 & \beta B^TB & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix}. \quad (2.8)$$

Proof. Setting $w = w^*$ in (2.5), and using G and $\eta(y^k, y^{k+1})$, we get

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

Since F is monotone, it follows that

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

The last inequality is due to $w^{k+1} \in \Omega$ and $w^* \in \Omega^*$ is a solution of (see (1.3)). The lemma is proved. $\square$

By using $\eta(y^k, y^{k+1})$ (see (2.7)), $Ax^* + By^* = b$ and (2.1c), we have

$$\begin{aligned}&(w^{k+1} - w^*)^T \eta(y^k, y^{k+1})\\ &\quad = \ (B(y^k - y^{k+1}))^T \beta\{(Ax^{k+1} + By^{k+1}) - (Ax^* + By^*)\}\\ &\quad = \ (B(y^k - y^{k+1}))^T \beta(Ax^{k+1} + By^{k+1} - b)\\ &\quad = \ (\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}).\end{aligned} \tag{2.9}$$

Substituting it in (2.6), we obtain

$$\begin{aligned}&(w^{k+1} - w^*)^T G(w^k - w^{k+1})\\ &\quad \geq \ (\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}), \ \forall\, w^* \in \Omega^*.\end{aligned} \tag{2.10}$$

Lemma 2.2 *Let* $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ *be generated by* (2.1) *from the given* $w^k = (x^k, y^k, \lambda^k)$. *Then, we have*

$$(\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \geq 0. \tag{2.11}$$

Proof. Since (2.4b) is true for the k-th iteration and the previous iteration, we have

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T \{-B^T \lambda^{k+1}\} \geq 0, \ \forall\, y \in \mathcal{Y}, \tag{2.12}$$

and

$$\theta_2(y) - \theta_2(y^k) + (y - y^k)^T \{-B^T \lambda^k\} \geq 0, \ \forall\, y \in \mathcal{Y}, \tag{2.13}$$

Setting $y = y^k$ in (2.12) and $y = y^{k+1}$ in (2.13), respectively, and then adding the two resulting inequalities, we get

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

The assertion of this lemma is proved. $\square$

Under the assumption (2.2), the matrix G is positive semi-definite. In addition, if B is a full column rank matrix, G is positive definite. Even if in the positive semi-definite case, we also use $\|w - \tilde{w}\|_G$ to denote

$$\|w - \tilde{w}\|_G = \sqrt{(w - \tilde{w})^T G(w - \tilde{w})}.$$

If B is a full column rank matrix, G is positive definite.

Lemma 2.3 *Let* $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \Omega$ *be generated by* (2.1) *from the given* $w^k = (x^k, y^k, \lambda^k)$. *Then, we have*

$$(w^{k+1} - w^*)^T G(w^k - w^{k+1}) \geq 0, \quad \forall\, w^* \in \Omega^*, \tag{2.14}$$

and consequently

$$\|w^{k+1} - w^*\|_G^2 \leq \|w^k - w^*\|_G^2 - \|w^k - w^{k+1}\|_G^2, \ \forall\, w^* \in \Omega^*. \tag{2.15}$$

Proof. Assertion (2.14) follows from (2.10) and (2.11) directly. By using (2.14), we have

$$\begin{aligned}
\|w^k - w^*\|_G^2 &= \|(w^{k+1} - w^*) + (w^k - w^{k+1}\|_G^2 \\
&= \|w^{k+1} - w^*\|_G^2 + 2(w^{k+1} - w^*)^T G(w^k - w^{k+1}) \\
&\quad + \|w^k - w^{k+1}\|_G^2 \\
&\geq \|w^{k+1} - w^*\|_G^2 + \|w^k - w^{k+1}\|_G^2,
\end{aligned}$$

and thus (2.15) is proved. $\square$

The inequality (2.15) is essential for the convergence of the alternating direction method. Note that G is positive semi-definite.

$$\|w^k - w^{k+1}\|_G^2 = 0 \quad \Longleftrightarrow \quad G(w^k - w^{k+1}) = 0.$$

The inequality (2.15) can be written as

$$\begin{aligned}
&\|x^{k+1}-x^*\|^2_{(rI-\beta A^TA)}+\beta\|B(y^{k+1}-y^*)\|^2+\frac{1}{\beta}\|\lambda^{k+1}-\lambda^*\|^2\\
\le\quad &\|x^{k}-x^*\|^2_{(rI-\beta A^TA)}+\beta\|B(y^{k}-y^*)\|^2+\frac{1}{\beta}\|\lambda^{k}-\lambda^*\|^2\\
&-\big(\|x^k-x^{k+1}\|^2_{(rI-\beta A^TA)}+\beta\|B(y^k-y^{k+1})\|^2+\frac{1}{\beta}\|\lambda^k-\lambda^{k+1}\|^2\big).
\end{aligned}$$

It leads to that

$$\lim_{k\to\infty} x^k = x^*, \qquad \lim_{k\to\infty} By^k = By^* \qquad \text{and} \qquad \lim_{k\to\infty}\lambda^k=\lambda^*.$$

The linearizing ADM is also known as the split inexact Uzawa method in image processing literature [9, 10].

3 Self-Adaptive ADM-based Contraction Method

In the last section, we get x^{k+1} by solving the following x-subproblem:

$$\min\{\theta_1(x)+\beta x^TA^T(Ax^k+By^k-b-\tfrac{1}{\beta}\lambda^k)+\frac{r}{2}\|x-x^k\|^2 \mid x\in\mathcal{X}\}$$

and it required that the parameter r to satisfy

$$rI_n-\beta A^TA\succeq 0 \qquad\Longleftrightarrow\qquad r>\beta\lambda_{\max}(A^TA).$$

In some practical problem, a conservative estimation of $\lambda_{\max}(A^TA)$ will leads a slow convergence. In this section, based on the linearized ADM, we consider the self-adaptive contraction methods. Each iteration of the self-adaptive contraction methods consists of two steps–prediction step and correction step. From the given w^k, the prediction step produces a test vector $\tilde{w}^k$ and the correction step offers the new iterate w^{k+1}.

3.1 Prediction

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

$$\tilde{x}^k = \text{Argmin} \left(\begin{array}{c} \Big\{ \theta_1(x) + \beta x^T A^T (Ax^k + By^k - b - \frac{1}{\beta}\lambda^k) \\ + \frac{r}{2}\|x - x^k\|^2 \quad \Big| \; x \in \mathcal{X} \Big\} \end{array} \right) \tag{3.1a}$$

2. Then, use λ^k and the obtained $\tilde{x}^k$, $\tilde{y}^k$ is the solution of the problem

$$\tilde{y}^k = \text{Argmin} \left\{ \begin{array}{c} \theta_2(y) - (\lambda^k)^T (A\tilde{x}^k + By - b) \\ + \frac{\beta}{2}\|A\tilde{x}^k + By - b\|^2 \end{array} \Bigg| \, y \in \mathcal{Y} \right\} \tag{3.1b}$$

3. Finally,

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k - b). \tag{3.1c}$$

The subproblems in (3.1) are similar as in (2.1). Instead of $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ in (2.1), we denote the output of (3.1) by $(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$.

Requirements on parameters β, r For given $\beta > 0$, choose r such that

$$\beta\|A^T A(x^k - \tilde{x}^k)\| \le \nu r \|x^k - \tilde{x}^k\|, \quad \nu \in (0,1). \tag{3.2}$$

If $rI - \beta A^T A \succ 0$, then (3.2) is satisfied. Thus, (2.2) is sufficient for (3.2).

Analysis of the optimal conditions of subproblems in (3.1)

Because we get $\tilde{w}^k$ in (3.1) via substituting w^{k+1} in (2.1) by $\tilde{w}^k$. Therefore, similar as (2.5), we get

$$\theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{\lambda}^k \\ -B^T\tilde{\lambda}^k \\ A\tilde{x}^k + B\tilde{y}^k - b \end{pmatrix} + \beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k - \tilde{y}^k) \right.$$

$$\left. + \begin{pmatrix} rI_{n_1} - \beta A^T A & 0 & 0 \\ 0 & \beta B^T B & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} \tilde{x}^k - x^k \\ \tilde{y}^k - y^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix} \right\} \ge 0, \; \forall\, w \in \Omega.$$

The last variational inequality can be rewritten as

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T \\ \left\{F(\tilde{w}^k) + \eta(y^k, \tilde{y}^k) + HM(\tilde{w}^k - w^k)\right\} \geq 0, \ \forall\, w \in \Omega, \quad (3.3)$$

where

$$\eta(y^k, \tilde{y}^k) = \beta \begin{pmatrix} A^T \\ B^T \\ 0 \end{pmatrix} B(y^k - \tilde{y}^k), \quad (3.4)$$

$$H = \begin{pmatrix} rI_{n_1} & 0 & 0 \\ 0 & \beta B^T B & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}, \quad (3.5)$$

and

$$M = \begin{pmatrix} I_{n_1} - \frac{\beta}{r} A^T A & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & 0 & I_m \end{pmatrix}. \quad (3.6)$$

Based on the above analysis, we have the following lemma.

Lemma 3.1 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (3.1) *from the given* $w^k = (x^k, y^k, \lambda^k)$. *Then, we have*

$$(\tilde{w}^k - w^*)^T HM(w^k - \tilde{w}^k) \geq (\tilde{w}^k - w^*)^T \eta(y^k, \tilde{y}^k), \ \ \forall\, w^* \in \Omega^*. \quad (3.7)$$

Proof. Setting $w = w^*$ in (3.3), we obtain

$$\begin{aligned} &(\tilde{w}^k - w^*)^T HM(w^k - \tilde{w}^k) \\ \geq \ & (\tilde{w}^k - w^*)^T \eta(y^k, \tilde{y}^k) \\ & + \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k). \end{aligned} \quad (3.8)$$

Since F is monotone and $\tilde{w}^k \in \Omega$, it follows that

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

The last inequality is due to $\tilde{w}^k \in \Omega$ and $w^* \in \Omega^*$ is a solution of (1.3). Substituting it in the right hand side of (3.8), the lemma is proved. □

In addition, because $Ax^* + By^* = b$ and $\beta(A\tilde{x}^k + B\tilde{y}^k - b) = \lambda^k - \tilde{\lambda}^k$, we have

$$\begin{aligned} & (\tilde{w}^k - w^*)^T \eta(y^k, \tilde{y}^k) \\ = \quad & (B(y^k - \tilde{y}^k))^T \beta\{(A\tilde{x}^k + B\tilde{y}^k) - (Ax^* + By^*)\} \\ = \quad & (\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k). \end{aligned} \tag{3.9}$$

Lemma 3.2 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (3.1) *from the given* $w^k = (x^k, y^k, \lambda^k)$. *Then, we have*

$$(w^k - w^*)^T HM(w^k - \tilde{w}^k) \geq \varphi(w^k, \tilde{w}^k), \quad \forall\, w^* \in \Omega^*, \tag{3.10}$$

where

$$\varphi(w^k, \tilde{w}^k) = (w^k - \tilde{w}^k)^T HM(w^k - \tilde{w}^k) + (\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k). \tag{3.11}$$

Proof. From (3.7) and (3.9) we have

$$(\tilde{w}^k - w^*)^T HM(w^k - \tilde{w}^k) \geq (\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k).$$

Assertion (3.10) follows from the last inequality and the definition of $\varphi(w^k, \tilde{w}^k)$ directly. □

3.2 The Primary Contraction Methods

The primary contraction methods use $M(w^k - \tilde{w}^k)$ as search direction and the unit step length. In other words, the new iterate is given by

$$w^{k+1} = w^k - M(w^k - \tilde{w}^k). \tag{3.12}$$

According to (3.6), it can be written as

$$\begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} \tilde{x}^k + \frac{\beta}{r} A^T A(x^k - \tilde{x}^k) \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix}. \tag{3.13}$$

In the primary contraction method, only the x-part of the corrector is different from the predictor. In the method of Section 2, we need $r \geq \beta \|A^T A\|$. By using the method in this section, we need only a r to satisfy the condition (3.2). In practical computation, we try to use the average of the eigenvalues of $\beta A^T A$.

Using (3.10), we have

$$\begin{aligned} & \|w^k - w^*\|_H^2 - \|w^{k+1} - w^*\|_H^2 \\ = \quad & \|w^k - w^*\|_H^2 - \|(w^k - w^*) - M(w^k - \tilde{w}^k)\|_H^2 \\ = \quad & 2(w^k - w^*)^T HM(w^k - \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|_H^2 \\ \geq \quad & 2\varphi(w^k, \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|_H^2. \end{aligned} \tag{3.14}$$

Because $(\tilde{y}^k, \tilde{\lambda}^k) = (y^{k+1}, \lambda^{k+1})$, the inequality (2.11) is still holds and thus

$$(\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k) \geq 0.$$

Therefore, it follows from (3.14) , (3.11) and the last inequality that

$$\begin{aligned} & \|w^k - w^*\|_H^2 - \|w^{k+1} - w^*\|_H^2 \\ \geq \quad & 2(w^k - \tilde{w}^k)^T HM(w^k - \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|_H^2. \end{aligned} \tag{3.15}$$

Lemma 3.3 *Under the condition* (3.2)*, we have*

$$\begin{aligned} & 2(w^k - \tilde{w}^k)^T HM(w^k - \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|_H^2 \\ & \geq (1 - \nu^2) r \|x^k - \tilde{x}^k\|^2 + \beta \|B(y^k - \tilde{y}^k)\|^2 + \tfrac{1}{\beta} \|\lambda^k - \tilde{\lambda}^k\|^2. \end{aligned} \tag{3.16}$$

Proof. First, we have

$$\begin{aligned} & 2(w^k - \tilde{w}^k)^T HM(w^k - \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|_H^2 \\ = \quad & (w^k - \tilde{w}^k)^T \big(M^T H + HM - M^T HM\big)(w^k - \tilde{w}^k). \end{aligned}$$

By using the structure of the matrices H and M (see (3.5) and (3.6)), we obtain

$$\begin{aligned} M^T H + HM - M^T HM &= H - (I - M^T)H(I - M) \\ &= \begin{pmatrix} rI_n & 0 & 0 \\ 0 & \beta B^T B & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix} - \begin{pmatrix} r(\frac{\beta}{r} A^T A)^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

Therefore,

$$\begin{aligned} & 2(w^k - \tilde{w}^k)^T HM(w^k - \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|_H^2 \\ = \quad & \|w^k - \tilde{w}^k\|_H^2 - r\left(\frac{\beta^2}{r^2}\right)\|A^T A(x^k - \tilde{x}^k)\|^2. \end{aligned} \tag{3.17}$$

Under the condition (3.2), we have

$$\left(\frac{\beta^2}{r^2}\right)\|A^T A(x^k - \tilde{x}^k)\|^2 \le \nu^2 \|x^k - \tilde{x}^k\|^2.$$

Substituting it in (3.17), the assertion of this lemma is proved. □

Theorem 3.1 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (3.1) *from the given*

$w^k = (x^k, y^k, \lambda^k)$ *and the new iterate* w^{k+1} *is given by* (3.12)*. The sequence* $\{w^k = (x^k, y^k, \lambda^k)\}$ *generated by the elementary contraction method satisfies*

$$\|w^{k+1} - w^*\|_H^2 \le \|w^k - w^*\|_H^2 - (1 - \nu^2)\|w^k - \tilde{w}^k\|_H^2. \tag{3.18}$$

Proof. From (3.15) and (3.16) we obtain

$$\begin{aligned} & \|w^k - w^*\|_H^2 - \|w^{k+1} - w^*\|_H^2 \\ \ge \quad & (1 - \nu^2) r\|x^k - \tilde{x}^k\|^2 + \beta\|B(y^k - \tilde{y}^k)\|^2 + \frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2 \\ \ge \quad & (1 - \nu^2)\|w^k - \tilde{w}^k\|_H^2. \end{aligned}$$

The assertion of this theorem is proved. □

Theorem 3.1 is essential for the convergence of the primary contraction method.

3.3 The general contraction method

The general contraction method For given w^k, we use

$$w(\alpha) = w^k - \alpha M(w^k - \tilde{w}^k) \tag{3.19}$$

to update the α-dependent new iterate. For any $w^* \in \Omega^*$, we define

$$\vartheta(\alpha) := \|w^k - w^*\|_H^2 - \|w(\alpha) - w^*\|_H^2 \tag{3.20}$$

and

$$q(\alpha) = 2\alpha\varphi(w^k, \tilde{w}^k) - \alpha^2\|M(w^k - \tilde{w}^k)\|_H^2, \tag{3.21}$$

where $\varphi(w^k, \tilde{w}^k)$ is defined in (3.11).

Theorem 3.2 *Let $w(\alpha)$ be defined by (3.19). For any $w^* \in \Omega^*$ and $\alpha \geq 0$, we have*

$$\vartheta(\alpha) \geq q(\alpha), \tag{3.22}$$

where $\vartheta(\alpha)$ and $q(\alpha)$ are defined in (3.20) *and* (3.21)*, respectively.*

Proof. It follows from (3.19) and (3.20) that

$$\begin{aligned} \vartheta(\alpha) &= \|w^k - w^*\|_H^2 - \|(w^k - w^*) - \alpha M(w^k - \tilde{w}^k)\|_H^2 \\ &= 2\alpha(w^k - w^*)^T HM(w^k - \tilde{w}^k) - \alpha^2\|M(w^k - \tilde{w}^k)\|_H^2. \end{aligned}$$

By using (3.10) and the definition of $q(\alpha)$, the theorem is proved. $\square$

Note that $q(\alpha)$ in (3.21) is a quadratic function of α and it reaches its maximum at

$$\alpha^* = \frac{\varphi(w^k, \tilde{w}^k)}{\|M(w^k - \tilde{w}^k)\|_H^2}. \tag{3.23}$$

In practical computation, we use

$$w^{k+1} = w^k - \gamma\alpha_k^* M(w^k - \tilde{w}^k), \tag{3.24}$$

to update the new iterate, where $\gamma \in [1, 2)$ is a relaxation factor. By using (3.20) and (3.22), we have

$$\|w^{k+1} - w^*\|_H^2 \leq \|w^k - w^*\|_H^2 - q(\gamma\alpha_k^*). \tag{3.25}$$

Note that

$$q(\gamma\alpha_k^*) = 2\gamma\alpha_k^*\varphi(w^k, \tilde{w}^k) - (\gamma\alpha_k^*)^2\|M(w^k - \tilde{w}^k)\|_H^2. \tag{3.26}$$

Using (3.23) and (3.24), we obtain

$$q(\gamma\alpha_k^*) = \gamma(2-\gamma)(\alpha_k^*)^2\|M(w^k - \tilde{w}^k)\|_H^2 = \frac{2-\gamma}{\gamma}\|w^k - w^{k+1}\|_H^2,$$

and consequently it follows from (3.25) that

$$\|w^{k+1} - w^*\|_H^2 \le \|w^k - w^*\|_H^2 - \frac{2-\gamma}{\gamma}\|w^k - w^{k+1}\|_H^2, \ \forall\, w^* \in \Omega^*. \tag{3.27}$$

On the other hand, it follows from (3.23) and (3.26) that

$$q(\gamma\alpha_k^*) = \gamma(2-\gamma)\alpha_k^*\varphi(w^k, \tilde{w}^k). \tag{3.28}$$

By using (3.11) and (3.16), we obtain

$$\begin{aligned}
&2\varphi(w^k, \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|_H^2 \\
&\quad = \quad 2(w^k - \tilde{w}^k)^T HM(w^k - \tilde{w}^k) - \|M(w^k - \tilde{w}^k)\|_H^2 \\
&\qquad\quad +2(\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k) \\
&\quad \ge (1-\nu^2)r\|x^k - \tilde{x}^k\|^2 + \beta\|B(y^k - \tilde{y}^k)\|^2 + \tfrac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2 \\
&\qquad\quad +2(\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k) \\
&\quad = \quad (1-\nu^2)r\|x^k - \tilde{x}^k\|^2 + \beta\|B(y^k - \tilde{y}^k) + \tfrac{1}{\beta}(\lambda^k - \tilde{\lambda}^k)\|^2.
\end{aligned}$$

Thus, we have $2\varphi(w^k, \tilde{w}^k) > \|M(w^k - \tilde{w}^k)\|_H^2$ and consequently

$$\alpha_k^* > \frac{1}{2}.$$

In addition, because

$$\begin{aligned}
\varphi(w^k,\tilde{w}^k) &= (w^k-\tilde{w}^k)^T HM(w^k-\tilde{w}^k) + (\lambda^k-\tilde{\lambda}^k)^T B(y^k-\tilde{y}^k) \\
&= \begin{pmatrix} x^k-\tilde{x}^k \\ y^k-\tilde{y}^k \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} rI_{n_1}-\beta A^TA & 0 & 0 \\ 0 & \beta B^TB & \frac{1}{2}B^T \\ 0 & \frac{1}{2}B & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} x^k-\tilde{x}^k \\ y^k-\tilde{y}^k \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix} \\
&\ge \|x^k-\tilde{x}^k\| \cdot \big(r\|x^k-\tilde{x}^k\| - \beta\|A^TA(x^k-\tilde{x}^k)\|\big) \\
&\quad + \frac{1}{2}\begin{pmatrix} y^k-\tilde{y}^k \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} \beta B^TB & 0 \\ 0 & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} y^k-\tilde{y}^k \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}.
\end{aligned}$$

Under the condition (3.2), it follows from the last inequality that

$$\begin{aligned}
\varphi(w^k,\tilde{w}^k) &\ge \min\big\{(1-\nu),\frac{1}{2}\big\}\|w^k-\tilde{w}^k\|_H^2 \\
&\ge \frac{1-\nu}{2}\|w^k-\tilde{w}^k\|_H^2. \qquad (3.29)
\end{aligned}$$

By using (3.25), (3.28), (3.29) and $\alpha_k^* \ge \frac{1}{2}$, we obtain the following theorem for the general contraction method.

Theorem 3.3 *The sequence* $\{w^k = (x^k, y^k, \lambda^k)\}$ *generated by the general contraction method satisfies*

$$\begin{aligned}
&\|w^{k+1}-w^*\|_H^2 \\
&\quad \le \|w^k-w^*\|_H^2 - \tfrac{\gamma(2-\gamma)(1-\nu)}{4}\|w^k-\tilde{w}^k\|_H^2, \ \forall w^* \in \Omega^*. \quad (3.30)
\end{aligned}$$

The inequality (3.30) in Theorem 3.3 is essential for the convergence of the general contraction method.

Both the inequalities (3.18) and (3.30) can be written as

$$\|w^{k+1}-w^*\|_H^2 \le \|w^k-w^*\|_H^2 - c_0\|w^k-\tilde{w}^k\|_H^2, \ \forall w^* \in \Omega^*,$$

where $c_0 > 0$ is a constant. Therefore, we have

$$\begin{aligned} & r\|x^{k+1} - x^*\|^2 + \beta\|B(y^{k+1} - y^*)\|^2 + \frac{1}{\beta}\|\lambda^{k+1} - \lambda^*\|^2 \\ \leq \quad & r\|x^k - x^*\|^2 + \beta\|B(y^k - y^*)\|^2 + \frac{1}{\beta}\|\lambda^k - \lambda^*\|^2 \\ & -c_0\big(r\|x^k - \tilde{x}^k\|^2 + \beta\|B(y^k - \tilde{y}^k)\|^2 + \frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2\big). \end{aligned}$$

It leads to that

$$\lim_{k\to\infty}\big(r\|x^k - \tilde{x}^k\|^2 + \beta\|B(y^k - \tilde{y}^k)\|^2 + \frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2\big),$$

and

$$\lim_{k\to\infty} x^k = x^*, \qquad \lim_{k\to\infty} By^k = By^* \qquad \text{and} \qquad \lim_{k\to\infty} \lambda^k = \lambda^*.$$

4 Applications in l_1-norm problems

An important l_1-norm problem in the area machine learning is the l_1 regularized linear regression, also called the lasso [8]. This involves solving

$$\min \ \tau\|x\|_1 + \tfrac{1}{2}\|Ax - b\|_2^2, \tag{4.1}$$

where $\tau > 0$ is a scalar regularization parameter that is usually chosen by cross-validation. In typical applications, there are many more features than training examples, and the goal is to find a parsimonious model for the data. The problem (1.1) can be reformulated to problem

$$\begin{aligned} \min \quad & \tau\|x\|_1 + \tfrac{1}{2}\|y\|_2^2 \\ & Ax - y = b \end{aligned} \tag{4.2}$$

which is a form of (1.1). Applied the alternating direction method (1.4) to the problem (4.2), the x-subproblem is

$$x^{k+1} = \text{Argmin } \{\tau\|x\|_1 + \tfrac{\beta}{2}\|(Ax - y^k) - \tfrac{1}{\beta}\lambda^k\|_2^2\},$$

and the solution does not has closed form. Applied the linearized alternating direction method (2.1) to the problem (4.2), the x-subproblem (2.1a) is

$$\tilde{x}^k = \text{Argmin}\{\tau\|x\|_1 + \tfrac{r}{2}\|x - [x^k + \tfrac{1}{r}\lambda^k - \tfrac{\beta}{r}A^T(Ax^k - y^k)]\|^2\}. \quad (4.3)$$

This problem is form of (1.5) and its solution has the following closed form:

$$\tilde{x}^k = a - P_{B_\infty^{\tau/r}}[a], \qquad \text{where} \qquad a = x^k + \tfrac{1}{r}\lambda^k - \tfrac{\beta}{r}A^T(Ax^k - y^k)$$

and

$$B_\infty^{\tau/r} = \{\xi \in \Re^n | -(\tau/r)e \le \xi \le (\tau/r)e\}.$$

By using the linearized alternating direction method in Section 2, for given $\beta > 0$, it needs $r > \beta\lambda_{\max}(A^TA)$. By using the self-adaptive ADM-based contraction method in Section 3, it needs r to satisfy

$$\beta\|A^TA(x^k - \tilde{x}^k)\| \le \nu r\|x^k - \tilde{x}^k\|, \quad \nu \in (0,1).$$

Because A is a generic matrix, the above condition is satisfied even if r is much less than $\beta\lambda_{\max}(A^TA)$.

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, **3**, 1-122, 2010.

[2] D. Gabay, Applications of the Method of Multipliers to Variational Inequalities, in *Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems*, M. Fortin and R. Glowinski, eds., North Holland, Amsterdam, The Netherlands, 299-331, 1983.

[3] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems viafinite element approximations, Computers and Mathematics with Applications, 2:17-40, 1976.

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

[5] B.S. He, A class of projection and contraction methods for monotone variational inequalities, Applied Math. & Optim., **35**, 69-76, 1997.

[6] B.S. He, Inexact implicit methods for monotone general variational inequalities, Math. Program. Series A, **86**, 199-217, 1999.

[7] B.S. He, L-Z Liao, D.R. Han and H. Yang, A new inexact alternating directions method for monotone variational inequalities, Math. Progr., **92**, 103-118, 2002.

[8] R. Tibshirani. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B, **58**, 267 − 288, 1996.

[9] X. Q. Zhang, M. Burger, X. Bresson and S. Osher, Bregmanized nonlocal regularization for deconvolution and sparse reconstruction, SIAM J. Imag. Sci., **3**, 253-276, 2010.

[10] X. Q. Zhang, M. Burger and S. Osher, A unified primal-dual algorithm framework based on Bregman iteration, J. Sci. Computing, **46**, 20-46, 2010.

凸优化和单调变分不等式的收缩算法

第十三讲: 定制 PPA 意义下的交替方向法及其线性化方法

Alternating direction method of multipliers in sense of customized PPA and its linearized version

南京大学数学系　何炳生
hebma@nju.edu.cn

The context of this lecture is based on the manuscript [2]

1 Structured constrained convex optimization

We consider the following structured constrained convex optimization problem

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

where $\theta_1(x):\Re^{n_1}\to\Re,\ \theta_2(y):\Re^{n_2}\to\Re$ are convex functions (but not necessary smooth), $A\in\Re^{m\times n_1}$, $B\in\Re^{m\times n_2}$ and $b\in\Re^m$, $\mathcal{X}\subset\Re^{n_1}$, $\mathcal{Y}\subset\Re^{n_2}$ are given closed convex sets.

The task of solving the problem (1.1) is to find an $(x^*,y^*,\lambda^*)\in\Omega$, such that

$$\left\{\begin{array}{l}\theta_1(x)-\theta_1(x^*)+(x-x^*)^T(-A^T\lambda^*)\ge 0,\\ \theta_2(y)-\theta_2(y^*)+(y-y^*)^T(-B^T\lambda^*)\ge 0,\quad \forall\,(x,y,\lambda)\in\Omega,\\ (\lambda-\lambda^*)^T(Ax^*+By^*-b)\ge 0,\end{array}\right. \quad (1.2)$$

where

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

By denoting

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

and

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

the first order optimal condition (1.2) can be written in a compact form such as

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

Note that the mapping F is monotone. We use Ω^* to denote the solution set of the variational inequality (1.3). For convenience we use the notations

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

Applied ADMM to the structure VI $\quad (y^k, \lambda^k) \Rightarrow (y^{k+1}, \lambda^{k+1})$

First, for given (y^k, λ^k), $\tilde{x}^k$ is the solution of the following problem

$$\tilde{x}^k = \text{Argmin} \left\{ \begin{array}{c} \theta_1(x) - (\lambda^k)^T(Ax + By^k - b) \\ + \frac{\beta}{2}\|Ax + By^k - b\|^2 \end{array} \middle| \, x \in \mathcal{X} \right\} \qquad (1.4a)$$

Use λ^k and the obtained $\tilde{x}^k$, $\tilde{y}^k$ is the solution of the following problem

$$\tilde{y}^k = \text{Argmin} \left\{ \begin{array}{c} \theta_2(y) - (\lambda^k)^T(A\tilde{x}^k + By - b) \\ + \frac{\beta}{2}\|A\tilde{x}^k + By - b\|^2 \end{array} \middle| \, y \in \mathcal{Y} \right\} \qquad (1.4b)$$

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k - b). \qquad (1.4c)$$

The sub-problems (1.4a) and (1.4b) are separately solved.

Classical Alternating Direction Method of Multipliers: $\quad v^{k+1} = \tilde{v}^k.$

2 ADMM based customized PPA

The k-th iteration of the proposed Alternating Direction Method of Multipliers in this section is also from a pair of (y^k, λ^k) to a new pair of (y^{k+1}, λ^{k+1}). In the prediction step, we generate a $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ which satisfies

$$\tilde{w}^k \in \Omega,\ \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T\{F(\tilde{w}^k)+Q(\tilde{v}^k-v^k)\} \ge 0,\ \forall\, w \in \Omega, \quad (2.1)$$

where Q is a 3×2 block matrix whose first low is zero, and the rest sub-matrix is symmetric and positive semi-definite. In details, the matrices Q and M have the following forms

$$Q = \begin{pmatrix} 0 & 0 \\ \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \quad \text{and} \quad H = \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \quad (2.2)$$

Note that the matrix H is symmetric and positive semidefinite. If we replace $Q(\tilde{v}^k - v^k)$ by $G(\tilde{w}^k - w^k)$ with a symmetric positive definite matrix G, then (2.1) becomes a sub-problem of the proximal point algorithm. Thus, the method in

this lecture is called the ADMM-based customized PPA or Alternating direction method in the sense of customized PPA.

2.1 Motivation

In the classical ADMM, x is the intermediate variable. For given (y^k, λ^k), we denote the the minimizer of the augmented Lagrangian function by $\tilde{x}^k$, *i. e.,*

$$\tilde{x}^k = \text{Argmin}\{\theta_1(x)-(\lambda^k)^T(Ax+By^k-b)+\tfrac{\beta}{2}\|Ax+By^k-b\|^2 \mid x \in \mathcal{X}\}. \quad (2.3)$$

Thus, we have $\tilde{x}^k \in \mathcal{X}$ and

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

If we write the above variational inequality as

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\tilde{\lambda}^k\} \ge 0,\ \forall\, x \in \mathcal{X}, \quad (2.4)$$

it implies that

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b). \quad (2.5)$$

According to the above definition, for any $\tilde{y}^k \in \mathcal{Y}$, we have

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + \tfrac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0. \qquad (2.6)$$

Combining (2.4) and (2.6) together, we get $(\tilde{x}^k, \tilde{\lambda}^k) \in \mathcal{X} \times \Re^m$,

$$\theta_1(x) - \theta_1(\tilde{x}^k) + \begin{pmatrix} x - \tilde{x}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{\lambda}^k \\ A\tilde{x}^k + B\tilde{y}^k - b \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ -B & \frac{1}{\beta} \end{pmatrix} \begin{pmatrix} \tilde{y}^k - y^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix} \right\} \ge 0, \quad (2.7)$$

for all $(x, \lambda) \in \mathcal{X} \times \Re^m$. In order to get $\tilde{w}^k \in \Omega$, such that

$$\begin{pmatrix} (\theta_1(x) - \theta_1(\tilde{x}^k)) + \\ (\theta_2(y) - \theta_2(\tilde{y}^k)) \end{pmatrix} + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{\lambda}^k \\ -B^T\tilde{\lambda}^k \\ A\tilde{x}^k + B\tilde{y}^k - b \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \beta B^T B & -B^T \\ -B & \frac{1}{\beta} \end{pmatrix} \begin{pmatrix} \tilde{y}^k - y^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix} \right\} \ge 0, \ \forall w \in \Omega, \quad (2.8)$$

we need only to find $\tilde{y}^k \in \mathcal{Y}$, such that

$$\tilde{y}^k \in \mathcal{Y}, \quad \big(\theta_2(y) - \theta_2(\tilde{y}^k)\big) + (y - \tilde{y}^k)^T \big\{-B^T\tilde{\lambda}^k + B^T\big(\beta B(\tilde{y}^k - y^k) - (\tilde{\lambda}^k - \lambda^k)\big)\big\} \ge 0, \ \forall\, y \in \mathcal{Y}. \quad (2.9)$$

By using (2.5), we have

$$\beta B(\tilde{y}^k - y^k) - (\tilde{\lambda}^k - \lambda^k) = \beta\big(A\tilde{x}^k + B\tilde{y}^k - b\big).$$

Thus, the variational inequality (2.9) is

$$\big(\theta_2(y) - \theta_2(\tilde{y}^k)\big) + (y - \tilde{y}^k)^T \big\{-B^T\tilde{\lambda}^k + \beta B^T\big(A\tilde{x}^k + B\tilde{y}^k - b\big)\big\} \ge 0, \ \forall\, y \in \mathcal{Y}.$$

For given $\tilde{x}^k$ and the defined $\tilde{\lambda}^k$ in (2.5), such a $\tilde{y}^k$ can be obtained via solving the following convex optimization problem:

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \tfrac{\beta}{2}\|A\tilde{x}^k + By - b - \tfrac{1}{\beta}\tilde{\lambda}^k\|^2 \,|\, y \in \mathcal{Y}\}. \qquad (2.10)$$

The above analysis guides us to construct the ADMM based customized PPA.

2.2 The proposed ADMM based customized PPA

From given $v^k = (y^k, \lambda^k)$, the prediction step produces $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$.

The prediction step:

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

$$\tilde{x}^k = \text{Argmin}\{\theta_1(x) + \tfrac{\beta}{2}\|Ax + By^k - b - \tfrac{1}{\beta}\lambda^k\|^2 \,|\, x \in \mathcal{X}\} \quad (2.11a)$$

2. Set the multipliers by

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b). \quad (2.11b)$$

3. Finally, use the obtained $\tilde{x}^k$ and $\tilde{\lambda}^k$, find $\tilde{y}^k$ by

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \tfrac{\beta}{2}\|A\tilde{x}^k + By - b - \tfrac{1}{\beta}\tilde{\lambda}^k\|^2 \,|\, y \in \mathcal{Y}\} \quad (2.11c)$$

In the ADMM view of point, we generate the predictor in the order

$$\tilde{x}^k, \quad \tilde{\lambda}^k \quad \text{and} \quad \tilde{y}^k.$$

As illustrated in the motivation, we get (2.8). This variational inequality can be written in the form of

$$\tilde{w}^k \in \Omega, \ (w - \tilde{w}^k)^T\{F(\tilde{w}^k) + Q(\tilde{v}^k - v^k)\} \ge 0, \ \forall\, w \in \Omega, \quad (2.12)$$

where Q is just the same matrix defined in (2.2). The above variational inequality is essential in the unified framework of the contraction methods.

The correction step: Update the new iterate v^{k+1} by

$$v^{k+1} = v^k - \gamma(v^k - \tilde{v}^k), \quad \gamma \in (0, 2). \quad (2.13)$$

To get the new iterate v^{k+1}, this method does not need to calculate the step size.

2.3 Convergence of the ADMM in sense of customized PPA

Based on the analysis in the last subsection, we have the following lemma.

Lemma 2.1 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (2.11) *from the given*

$v^k = (y^k, \lambda^k)$. *Then, we have*

$$(\tilde{w}^k - w^*)^T Q(v^k - \tilde{v}^k) \geq 0, \ \forall w^* \in \Omega^*, \tag{2.14}$$

where the matrix Q *is defined in* (2.2)*.*

Proof. Setting $(x, y, \lambda) = (x^*, y^*, \lambda^*)$ in (2.8), we obtain

$$(\tilde{w}^k - w^*)^T Q(v^k - \tilde{v}^k) \geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k). \tag{2.15}$$

Since F is monotone and $\tilde{w}^k \in \Omega$, it follows that

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

The last inequality is due to $\tilde{w}^k \in \Omega$ and $w^* \in \Omega^*$ (see (1.3)). Therefore, the right hand side of (2.15) is non-negative and the lemma is proved. $\square$

Lemma 2.2 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (2.11) *from the given*

$v^k = (y^k, \lambda^k)$. *Then, we have*

$$(v^k - v^*)^T H(v^k - \tilde{v}^k) \geq \|v^k - \tilde{v}^k\|_H^2, \ \forall\, v^* \in \mathcal{V}^*, \tag{2.16}$$

where H *is defined in* (2.2)*.*

Proof. Recall the matrices Q and H in (2.2). It follows from (2.14) that

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

Assertion (2.16) follows from the last inequality directly. $\square$

The matrix H is symmetric and positive semi-definite. We still use $\|v - \tilde{v}\|_H$ to denote that

$$\|v - \tilde{v}\|_H = \sqrt{(v - \tilde{v})^T H(v - \tilde{v})}.$$

If $\|v^k - \tilde{v}^k\|_H^2 = 0$, because H is symmetric and positive semi-definite, we have $H(v^k - \tilde{v}^k) = 0$. In this case, $\tilde{w}^k$ is a solution of the variational inequality (see (1.2) and (2.8)). Thus, we can take $\|v^k - \tilde{v}^k\|_H^2 \leq \epsilon$ as the stopping criterium in the iteration process.

Theorem 2.1 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (2.11) *from the given* $v^k = (y^k, \lambda^k)$ *and the new iterate* v^{k+1} *be given by* (2.13)*. Then we have*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \gamma(2-\gamma)\|v^k - \tilde{v}^k\|_H^2, \ \forall\, v^* \in \mathcal{V}^*. \quad (2.17)$$

Proof. By a simple manipulation, we obtain

$$\begin{aligned}
&\|v^{k+1} - v^*\|_H^2 \\
&\overset{(2.16)}{=} \|(v^k - v^*) - \gamma(v^k - \tilde{v}^k)\|_H^2 \\
&= \|v^k - v^*\|_H^2 - 2\gamma(v^k - v^*)^T H(v^k - \tilde{v}^k) + \gamma^2\|v^k - \tilde{v}^k\|_H^2 \\
&\overset{(2.13)}{\le} \|v^k - v^*\|_H^2 - 2\gamma\|v^k - \tilde{v}^k\|_H^2 + \gamma^2\|v^k - \tilde{v}^k\|_H^2 \\
&= \|v^k - v^*\|_H^2 - \gamma(2-\gamma)\|v^k - \tilde{v}^k\|_H^2.
\end{aligned}$$

This is true for any $v^* \in \mathcal{V}^*$ and the theorem is proved. $\square$

The inequality (2.17) is essential for the convergence of the proposed alternating direction method. The detailed convergence proof can be found in [2]. For the convergence rate of the customized PPA, the reader are refereed to [6].

2.4 Ensure the matrix H to be positive definite

In the ADMM based customized PPA (2.11), the subproblem (2.11c) can be written as

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \tfrac{\beta}{2}\|By - p^k\|^2 \,|\, y \in \mathcal{Y}\}, \quad (2.18)$$

where

$$p^k = b + \tfrac{1}{\beta}\tilde{\lambda}^k - A\tilde{x}^k.$$

If we add an additional term $\frac{\delta\beta}{2}\|B(y - y^k)\|^2$ (with any small $\delta > 0$) to the objective function of the subproblem (2.11c), we will get $\tilde{y}^k$ via

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \tfrac{\beta}{2}\|By - p^k\|^2 + \tfrac{\delta\beta}{2}\|B(y - y^k)\|^2 \,|\, y \in \mathcal{Y}\}.$$

By a manipulation, the solution point of the above subproblem is obtained via

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \tfrac{(1+\delta)\beta}{2}\|By - q^k\|^2 \,|\, y \in \mathcal{Y}\}, \quad (2.19)$$

where

$$q^k = \tfrac{1}{1+\delta}(p^k + \delta By^k).$$

In this way, the matrix Q in (2.12) will be modified to

$$Q = \begin{pmatrix} 0 & 0 \\ (1+\delta)\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix},$$

and the related matrix H in (2.2) becomes

$$\begin{aligned} H &= \begin{pmatrix} (1+\delta)\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{\beta} B^T & 0 \\ 0 & \sqrt{\frac{1}{\beta}} I_m \end{pmatrix} \begin{pmatrix} (1+\delta) I & -I \\ -I & I_m \end{pmatrix} \begin{pmatrix} \sqrt{\beta} B & 0 \\ 0 & \sqrt{\frac{1}{\beta}} I_m \end{pmatrix}. \end{aligned} \tag{2.20}$$

Thus, for any $\delta > 0$, H is positive definite when B is a full rank matrix. In other words, instead of (2.18), using (2.19) to get $\tilde{y}^k$, it will ensure the positivity of H theoretically. However, in practical computation, it works still well by using $\delta = 0$.

3 Application and Numerical Experiments

3.1 Applications to least-squares problems

We consider the following problem:

$$\min\{\frac{1}{2}\|X - C\|_F^2 \mid X \in S_+^n \cap S_B\}, \tag{3.1}$$

where

$$S_+^n = \{H \in R^{n\times n} \mid H^T = H,\ H \succeq 0\}.$$

and

$$S_B = \{H \in R^{n\times n} \mid H^T = H,\ H_L \le H \le H_U\}.$$

Use the following MATLAB Code to produce the matrices C, H_L and H_U

```
rand('state',0);  C=rand(n,n);     C=(C'+C)-ones(n,n)+eye(n);
%% C is symmetric and C_{ij} is in (-1,1), C_{jj} is in (0,2)
HU=ones(n)*0.1;  HL=-HU; for i=1:n HU(i,i)=1; HL(i,i)=1; end;
```

The problem is converted to the following equivalent one:

$$\begin{aligned} \min \quad & \tfrac{1}{2}\|X-C\|^2 + \tfrac{1}{2}\|Y-C\|^2 \\ \text{s.t} \quad & X - Y = 0, \\ & X \in S_+^n,\ Y \in S_B. \end{aligned} \tag{3.2}$$

The basic sub-problems in the ADMM based customized PPA

- For fixed Y^k and Z^k,

$$\tilde{X}^k = \text{Argmin}\{\frac{1}{2}\|X-C\|_F^2 - \mathbf{Tr}(Z^k X) + \frac{\beta}{2}\|X - Y^k\|_F^2 \mid X \in S_+^n\}$$

- Set $\tilde{Z}^k$ by

$$\tilde{Z}^k = Z^k - \beta(\tilde{X}^k - Y^k).$$

- With fixed $\tilde{X}^k$ and $\tilde{Z}^k$,

$$\tilde{Y}^k = \text{Argmin}\{\frac{1}{2}\|Y-C\|_F^2 + \mathbf{Tr}(\tilde{Z}^k Y) + \frac{\beta}{2}\|\tilde{X}^k - Y\|_F^2 \mid Y \in S_B\}$$

$\tilde{X}^k$ can be directly obtained via

$$\tilde{X}^k = P_{S_+^n}\{\frac{1}{1+\beta}(\beta Y^k + Z^k + C)\}. \tag{3.3}$$

$$P_{S_+^n}(A) = U\Lambda^+ U^T, \qquad [U,\Lambda] = \text{eig}(A), \qquad \Lambda^+ = \max(\Lambda, 0).$$

Similarly, $\tilde{Y}^k$ in is given by

$$\tilde{Y}^k = P_{S_B}\{\frac{1}{1+\beta}(\beta\tilde{X}^k - \tilde{Z}^k + C)\}. \tag{3.4}$$

$$S_B = \{H \mid H_L \le H \le H_U\}, \qquad P_{S_B}(A) = \min(\max(H_L, A), H_U)$$

The most time consuming calculation is $[U,\Lambda] =$ eig(A), $9n^3$

MATLAB Code – An iteration of the classical ADMM

```
Y0= Y;                          Z0    = Z;          k = k+1;
X = (Y0*beta+Z0+C)/(1+beta);    [V,D] = eig(X);     D = max(0,D);
X = (V*D)*V';
Y = min(max((X*beta-Z0+C)/(1+beta),HL),HU);
Z = Z0-(X-Y)*beta;
```

MATLAB Code – An iteration of the new order ADMM

```
Y0= Y;                          Z0    = Z;          k = k+1;
X = (Y0*beta+Z0+C)/(1+beta);    [V,D] = eig(X);     D = max(0,D);
X = (V*D)*V';                   Z     = Z0-(X-Y0)*beta;;
Y = min(max((X*beta-Z+C)/(1+beta),HL),HU);
```

MATLAB Code – An iteration of the extended ADMM

```
Y0= Y;                          Z0    = Z;          k = k+1;
X = (Y0*beta+Z0+C)/(1+beta);    [V,D] = eig(X);     D = max(0,D);
X = (V*D)*V';                   Z     = Z0-(X-Y0)*beta;;
Y = min(max((X*beta-Z+C)/(1+beta),HL),HU);
Y = Y0-(Y0-Y)*1.5;
Z = Z0-(Z0-Z)*1.5;
```


Numerical results for problem (3.1)

C=rand(n,n); C=(C'+C)-ones(n,n) + eye(n)

H_U=ones(n,n)/10; H_L=-ones(n,n)/10; $H_U(jj) = H_L(jj)$=1.

Table 1. Numerical results

$n \times n$ Matrix	Classical ADMM		Customized PPA		Extended C-PPA		
$n =$	No. It	CPU Sec.	No. It	CPU Sec.	No. It	CPU Sec.	β
100	46	1.39	44	1.37	28	0.94	5
200	50	3.07	50	3.05	31	4.41	10
500	48	25.50	49	24.52	32	16.50	10
800	51	110.18	50	107.29	33	72.12	10
1000	51	208.93	52	212.74	34	140.70	10
2000	55	1578.96	55	1579.68	36	1053.87	10

3.2 Applications to image restoration

The mathematical form of the image restoration problem is

$$\min \||\nabla x|\|_1 + \frac{\mu}{2}\|Kx - f\|^2, \tag{3.5}$$

where $\mu > 0$ is trade-off; K is a blur operator and f is observed image.

The equivalent problem:

$$\begin{aligned} &\min \||y|\|_1 + \frac{\mu}{2}\|Kx - f\|^2 \\ &\text{s. t. } \nabla x = y, \end{aligned} \tag{3.6}$$

This is a problem of form (1.1) where $\mathcal{X}$, $\mathcal{Y}$ are full spaces,

$$\begin{aligned} &\theta_1(x) = \frac{\mu}{2}\|Kx - f\|^2, \\ &\theta_2(y) = \||y|\|_1, \\ &A = \nabla, \quad B = -I \quad \text{and} \quad b = 0. \end{aligned}$$

The augmented Lagrangian function

$$\mathcal{L}_A(x, y, \lambda) = \||y|\|_1 + \frac{\mu}{2}\|Kx - f\|^2 - \lambda^T(\nabla x - y) + \frac{\beta}{2}\|\nabla x - y\|^2,$$

where λ is Lagrange multiplier and β is the penalty parameter.

For given (y^k, λ^k), get $(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ as follows:

1. $\tilde{x}^k$ is the solution of the following least square problem

$$\tilde{x}^k = \arg\min_x \left\{ \tfrac{\mu}{2}\|Kx - f\|^2 - (\lambda^k)^T(\nabla x - y^k) + \tfrac{\beta}{2}\|\nabla x - y^k\|^2 \right\}.$$

2. Set $\tilde{\lambda}^k$ by

$$\tilde{\lambda}^k = \lambda^k - \beta(\nabla \tilde{x}^k - y^k).$$

3. Finally, with fixed $(\tilde{x}^k, \tilde{\lambda}^k)$, $\tilde{y}^k$ are solutions of

$$\tilde{y}^k = \arg\min_y \left\{ \||y|\|_1 - (\tilde{\lambda}^k)^T(\nabla \tilde{x}^k - y) + \tfrac{\beta}{2}\|\nabla \tilde{x}^k - y\|^2 \right\}.$$

Solving the x subproblem for getting $\tilde{x}^k$:

$$(\beta\nabla^T\nabla + \mu K^T K)\tilde{x}^k = \nabla^T(\beta y^k + \lambda^k) + \mu K^T f.$$

- If ∇ and K satisfy some periodic boundary conditions, they can be factored by Fourier transform as $\nabla = \mathcal{F}^{-1}\Lambda_D\mathcal{F}$ and $K = \mathcal{F}^{-1}\Lambda_K\mathcal{F}$.
- If ∇ and K satisfy some reflective boundary conditions, they can be factored by discrete cosine transform as $\nabla = \mathcal{C}^{-1}\Lambda_D\mathcal{C}$ and $K = \mathcal{C}^{-1}\Lambda_K\mathcal{C}$.

Solving the y subproblem for getting $\tilde{y}^k$:

$$\tilde{y}^k = \text{shrink}_{\frac{1}{\beta}}\left(\nabla\tilde{x}^k - \frac{\tilde{\lambda}^k}{\beta}\right),$$

where

$$\text{shrink}_c(v) = v - \min(c, \|v\|)\frac{v}{\|v\|}.$$

Note that

$$\text{shrink}_c(v) = v - P_{B_2^c}(v) \quad \text{where} \quad B_2^c = \{v \in \Re^n | \|v\|_2 \le c\}.$$

MATLAB Code – An iteration of the classical ADMM

```
%%%%%%%%%%%%%%%%%%%%   step 1  x^{k+1}   %%%%%%%%%%%%%%%%%%%%%%%
Temp= PTx(beta*v1+lbd11) + PTy(beta*v2+lbd12) + HTx0;
un  = real(ifft2(fft2(Temp)./MDu));
%%%%%%%%%%%%%%%%%%%%   step 2  y^{k+1}   %%%%%%%%%%%%%%%%%%%%%%%
dxun= Px(un);
dyun= Py(un);
sk1 = dxun - lbd11/beta;
sk2 = dyun - lbd12/beta;
nsk = sqrt(sk1.^2 + sk2.^2); nsk(nsk==0)=1;
nsk = max(1-1./(beta*nsk),0);
vn1 = sk1.*nsk;
vn2 = sk2.*nsk;
%%%%%%%%%%%%%%%%%%%% update  \lambda     %%%%%%%%%%%%%%%%%%%%%%%
lbdn11 = lbd11 - beta*(dxun - vn1);
lbdn12 = lbd12 - beta*(dyun - vn2);
%%%%%%%%%%%%%%%%%%%% New iterative point %%%%%%%%%%%%%%%%%%%%%%%
u = un; v1 = vn1; v2 = vn2; lbd11 = lbdn11; lbd12 = lbdn12;
```

MATLAB Code – An iteration of the new order ADMM

```
%%%%%%%%%%%%%%%%%%%%   step 1  x^{k+1}  %%%%%%%%%%%%%%%%%%%%%%%
Temp= PTx(beta*v1+lbd11) + PTy(beta*v2+lbd12) + HTx0;
un  = real(ifft2(fft2(Temp)./MDu));
dxun= Px(un);
dyun= Py(un);
%%%%%%%%%%%%%%%%%%%%   update \lambda   %%%%%%%%%%%%%%%%%%%%%%%
lbdn11 = lbd11 - beta*(dxun - v1);
lbdn12 = lbd12 - beta*(dyun - v2);
%%%%%%%%%%%%%%%%%%%%   step 2  y^{k+1}  %%%%%%%%%%%%%%%%%%%%%%%
sk1 = dxun - lbdn11/beta;
sk2 = dyun - lbdn12/beta;
nsk = sqrt(sk1.^2 + sk2.^2); nsk(nsk==0)=1;
nsk = max(1-1./(beta*nsk),0);
vn1 = sk1.*nsk;
vn2 = sk2.*nsk;
%%%%%%%%%%%%%%%%%%%% New iterative point %%%%%%%%%%%%%%%%%%%%%%
u = un; v1 = vn1; v2 = vn2; lbd11 = lbdn11; lbd12 = lbdn12;
```


MATLAB Code – An iteration of the extended C-PPA

```
%%%%%%%%%%%%%%%%%%%%   step 1  x^{k+1}  %%%%%%%%%%%%%%%%%%%%%%%
Temp= PTx(beta*v1+lbd11) + PTy(beta*v2+lbd12) + HTx0;
un  = real(ifft2(fft2(Temp)./MDu));
dxun= Px(un);
dyun= Py(un);
%%%%%%%%%%%%%%%%%%%%  update \lambda    %%%%%%%%%%%%%%%%%%%%%%%
lbdn11 = lbd11 - beta*(dxun - v1);
lbdn12 = lbd12 - beta*(dyun - v2);
%%%%%%%%%%%%%%%%%%%%  step 2  y^{k+1}  %%%%%%%%%%%%%%%%%%%%%%%
sk1 = dxun - lbd11/beta;
sk2 = dyun - lbd12/beta;
nsk = sqrt(sk1.^2 + sk2.^2); nsk(nsk==0)=1;
nsk = max(1-1./(beta*nsk),0);
vn1 = sk1.*nsk;
vn2 = sk2.*nsk;
%%%%%%%%%%%%%%%%%% New iterative point %%%%%%%%%%%%%%%%%%%%%
u  = un;
v1 = v1 - gamma*(v1-vn1);
v2 = v2 - gamma*(v2-vn2);
lbd11 = lbd11 - gamma*(lbd11-lbdn11);
lbd12 = lbd12 - gamma*(lbd12-lbdn12);
```

Numerical results for image restoration

```
I = double(imread('chart.tiff'))/255; I = double(imread('house.png'))/255;
h = fspecial('disk',7); x0 = imfilter(I,h,'circular')+0.02*randn(size(I));
```

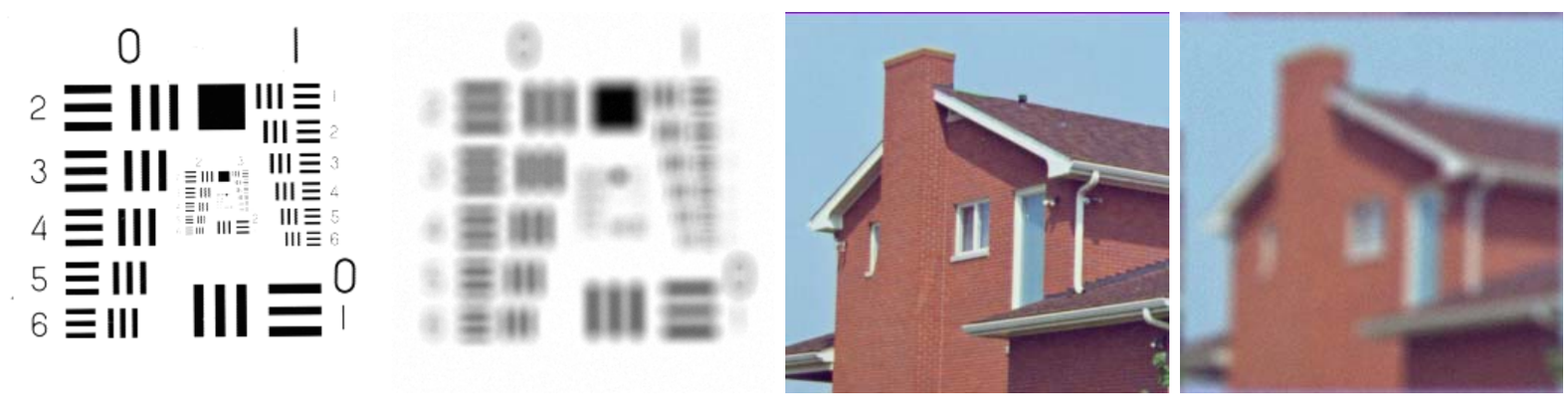


Figure 1: Original and degraded images. Left: Chart. Right: House

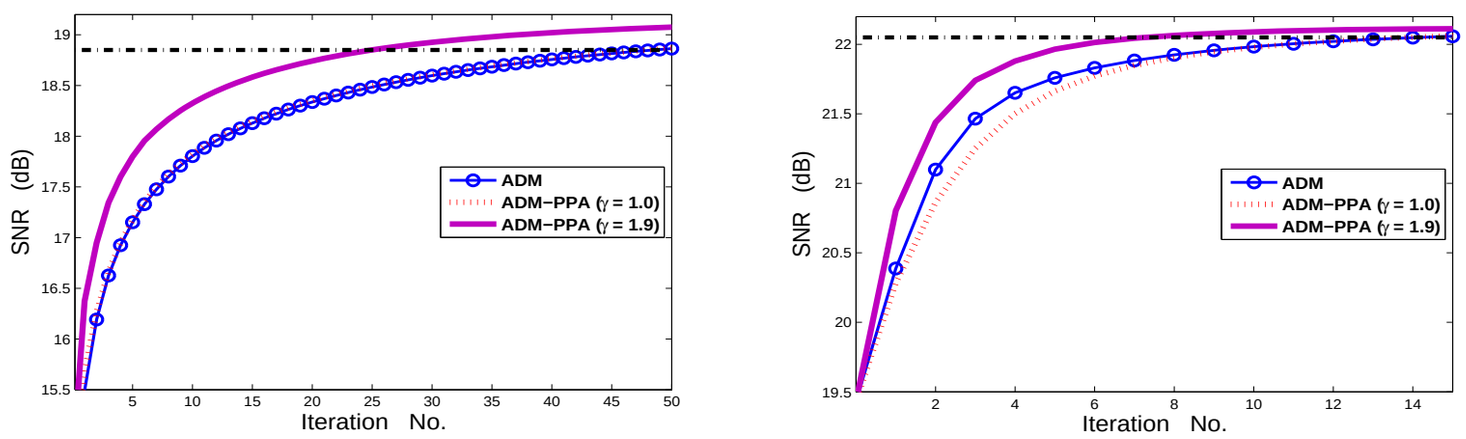


Figure 2: Performances of ADMM and two variants methods on TV-l2. Left: Chart. Right: House.

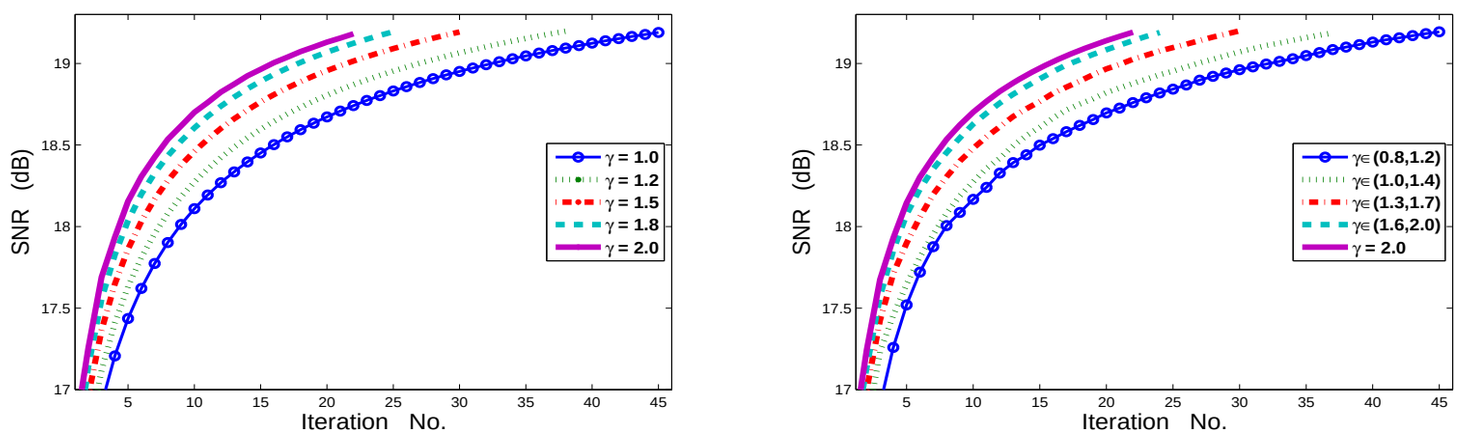


Figure 3: Performances of Algorithm 2 with different values of γ for Chart. Top: fixed γ. Bottom: random generated γ.

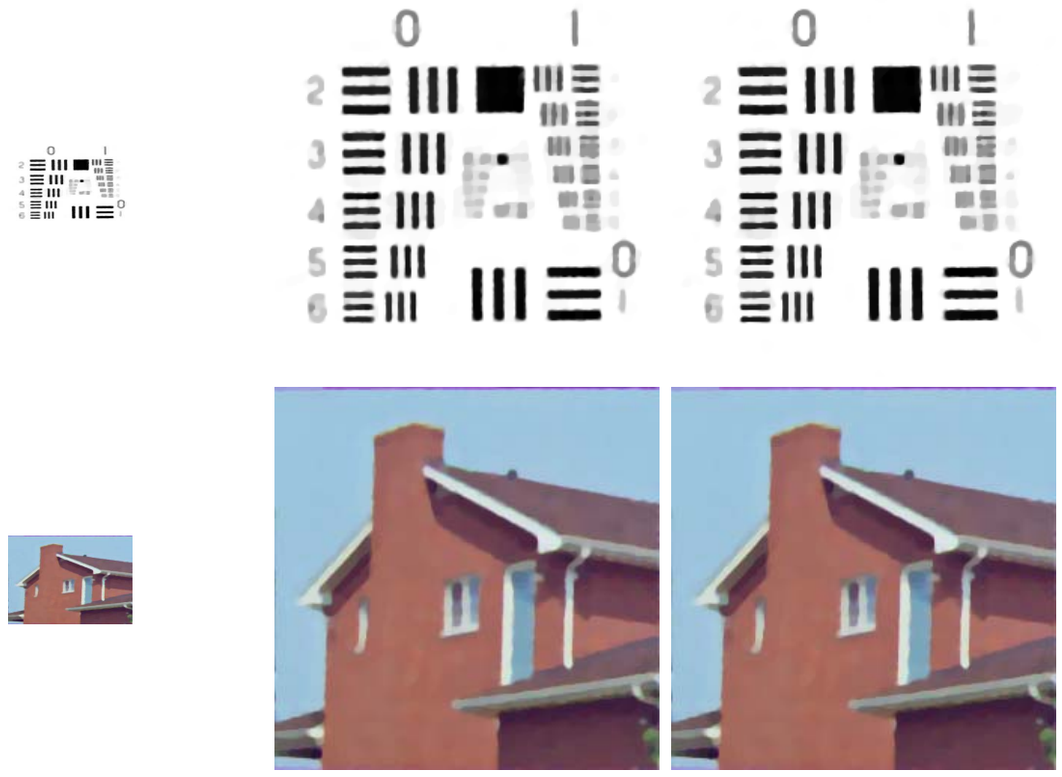


Figure 4: Restorations. From left column to right column: ADMM, new order ADMM, and the extended new order ADMM.

Table 1: Numerical comparisons of the classical ADMM (ADMM), the customized PPA and the extended customized PPA for TV-l^2 image restoration.

	Chart			House		
	ADMM	C-PPA	E-C-PPA	ADMM	C-PPA	E-C-PPA
		$\gamma = 1$	$\gamma = 1.8$		$\gamma = 1$	$\gamma = 1.8$
Iter	74	74	42	57	56	33
CPU	2.32	2.29	1.29	2.79	2.75	1.63
SNR	19.01	19.01	19.02	22.11	22.11	22.12

SNR $= 20 \log_{10} \frac{\|x\|}{\|x-I\|}$, where x is restoration and I is original image.

It seems that the new order ADMM is so good as the classical one. However, the extended-ADMM converges much faster than the other both ADMMs.

4 Linearized ADMM-based PPA Method

Note that the subproblems (2.11a) and (2.11c) in the last section are equivalent to the problems

$$\tilde{x}^k = \text{Argmin}\big\{\theta_1(x) + \tfrac{\beta}{2}\|(Ax + By^k - b) - \tfrac{1}{\beta}\lambda^k\|^2 \big| x \in \mathcal{X}\big\} \tag{4.1a}$$

and

$$\tilde{y}^k = \text{Argmin}\big\{\theta_2(y) + \tfrac{\beta}{2}\|(A\tilde{x}^k + By - b) - \tfrac{1}{\beta}\tilde{\lambda}^k\|^2 \big| y \in \mathcal{Y}\big\} \tag{4.1b}$$

respectively. In some structured optimization (1.1), the subproblem (4.1a) is easy because A is usually a scalar matrix. However, to obtain the solution of the subproblem (4.1b) is expensive in the case that B does not have a special form. In this lecture, we suppose that only the solution of the problem

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

has a closed form, and consider to linearize the quadratic function of the subproblem (4.1b) ADMM in sense of the customized PPA.

4.1 Linearized alternating direction method

The prediction step:

1. First, for given (x^k, y^k, λ^k), solving the x subproblem to get $\tilde{x}^k$ by

$$\tilde{x}^k = \text{Argmin}\left\{\theta_1(x) + \tfrac{\beta}{2}\|(Ax + By^k - b) - \tfrac{1}{\beta}\lambda^k\|^2 \big| x \in \mathcal{X}\right\} \tag{4.2a}$$

2. Set the new multipliers by

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b). \tag{4.2b}$$

3. Finally, use the obtained $\tilde{x}^k$ and $\tilde{\lambda}^k$, solving the y subproblem to get $\tilde{y}^k$ by

$$\tilde{y}^k = \text{Argmin}\left(\begin{array}{c}\Big\{\theta_2(y) + \beta y^T B^T(A\tilde{x}^k + By^k - b - \tfrac{1}{\beta}\tilde{\lambda}^k) \\ +\tfrac{s}{2}\|y - y^k\|^2 \quad \Big| \, y \in \mathcal{Y}\Big\}\end{array}\right). \tag{4.2c}$$

Request on the parameter s For given $\beta > 0$, s should satisfy

$$sI - \beta B^T B \succeq 0. \tag{4.3}$$

The correction step: Update the new iterate v^{k+1} by

$$v^{k+1} = v^k - \gamma(v^k - \tilde{v}^k), \quad \gamma \in [1, 2). \tag{4.4}$$

To get the new iterate v^{k+1}, this method does not need to calculate the step size. However, it needs to estimate the max-eigenvalue of $B^T B$, *i. e.*, $\lambda_{\max}(B^T B)$.

4.2 Analysis in the PPA framework

Note that the solution of (4.2a), $\tilde{x}^k$ satisfies

$$\begin{aligned} \tilde{x}^k \in \mathcal{X}, \;\; \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \\ \{-A^T\lambda^k + \beta A^T(A\tilde{x}^k + By^k - b)\} \ge 0, \; \forall\, x \in \mathcal{X}. \end{aligned} \tag{4.5}$$

Substituting $\tilde{\lambda}^k$ (see (4.2b)) in (4.5) (eliminating λ^k), we get

$$\tilde{x}^k \in \mathcal{X}, \;\; \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \big\{-A^T\tilde{\lambda}^k\big\} \ge 0, \; \forall\, x \in \mathcal{X}. \tag{4.6}$$

The solution of (4.2c), $\tilde{y}^k$ satisfies

$$\begin{aligned} \tilde{y}^k \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \big\{-B^T\tilde{\lambda}^k \\ +\beta B^T\big(A\tilde{x}^k + By^k - b\big) + s(\tilde{y}^k - y^k)\big\} \ge 0, \; \forall\, y \in \mathcal{Y}. \end{aligned} \tag{4.7}$$

Note that $\beta\big(A\tilde{x}^k + By^k - b\big) = (\lambda^k - \tilde{\lambda}^k)$ (see (4.2b)). Substituting it in (4.7), we obtain

$$\begin{aligned} \tilde{y}^k \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \big\{-B^T\tilde{\lambda}^k \\ -B^T(\tilde{\lambda}^k - \lambda^k) + s(\tilde{y}^k - y^k)\big\} \ge 0, \; \forall\, y \in \mathcal{Y}. \end{aligned} \tag{4.8}$$

From (4.2b) we have

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0. \tag{4.9}$$

Combining the inequalities (4.6), (4.8) and (4.9), we obtain

$$\theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{\lambda}^k \\ -B^T\tilde{\lambda}^k \\ A\tilde{x}^k + B\tilde{y}^k - b \end{pmatrix} \right.$$

$$\left. + \begin{pmatrix} 0 & 0 \\ sI & -B^T \\ -B & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} \tilde{y}^k - y^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix} \right\} \geq 0, \ \forall\, w \in \Omega. \tag{4.10}$$

The last variational inequality can be written in form of $\tilde{w}^k \in \Omega$ and

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T\{F(\tilde{w}^k) + Q(\tilde{v}^k - v^k)\} \geq 0, \ \forall\, w \in \Omega, \tag{4.11}$$

where

$$Q = \begin{pmatrix} 0 & 0 \\ sI & -B^T \\ -B & \frac{1}{\beta}I_m \end{pmatrix} \tag{4.12}$$

which is essential in the framework of the PPA contraction methods.

4.3 Convergence of the Linearized ADM-based PPA Method

Based on the analysis in the last subsection, we have the following lemma.

Lemma 4.1 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (4.2) *from the given* $v^k = (y^k, \lambda^k)$. *Then, we have*

$$(\tilde{v}^k - v^*)^T G(v^k - \tilde{v}^k) \geq 0, \ \forall w^* \in \Omega^*, \tag{4.13}$$

where

$$G = \begin{pmatrix} sI & -B^T \\ -B & \frac{1}{\beta}I_m \end{pmatrix}. \tag{4.14}$$

Proof. Setting $w = w^*$ in (4.10), we get

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

Since F is monotone and $\tilde{w}^k \in \Omega$, it follows that

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

The right hand side of the last inequality is non-negative because $\tilde{w}^k \in \Omega$ and $w^* \in \Omega^*$. the assertion follows directly. $\square$

Lemma 4.2 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (4.2) *from the given* $v^k = (y^k, \lambda^k)$. *Then, we have*

$$(v^k - v^*)^T G(v^k - \tilde{v}^k) \geq \|v^k - \tilde{v}^k\|_G^2, \ \forall\, w^* \in \Omega^*, \tag{4.15}$$

where G *is defined in* (4.14).

Proof. Assertion (4.15) follows from the last inequality directly. $\square$

Since G is symmetric and positive semi-definite, we have

$$v^k - \tilde{v}^k = 0 \quad \text{or} \quad G(v^k - \tilde{v}^k) = 0,$$

whenever $\|v^k - \tilde{v}^k\|_G^2 = 0$. Therefore, it follows from (4.10) that $\tilde{w}^k$ is a solution of the variational inequality when $\|v^k - \tilde{v}^k\|_G^2 = 0$.

Theorem 4.1 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (4.2) *from the given* $v^k = (y^k, \lambda^k)$ *and the new iterate* w^{k+1} *be given by* (4.4). *Then we have*

$$\|v^{k+1} - v^*\|_G^2 \leq \|v^k - v^*\|_G^2 - \gamma(2-\gamma)\|v^k - \tilde{v}^k\|_G^2, \ \forall\, w^* \in \Omega^*. \tag{4.16}$$

Proof. By using (4.4) and (4.15), we obtain

$$\begin{aligned}
&\|v^{k+1} - v^*\|_G^2 \\
&\overset{(4.4)}{=} \quad \|(v^k - v^*) - \gamma(v^k - \tilde{v}^k)\|_G^2 \\
&\overset{(4.15)}{\leq} \quad \|v^k - v^*\|_G^2 - 2\gamma\|v^k - \tilde{v}^k\|_G^2 + \gamma^2\|v^k - \tilde{v}^k\|_G^2 \\
&= \quad \|v^k - v^*\|_G^2 - \gamma(2-\gamma)\|v^k - \tilde{v}^k\|_G^2.
\end{aligned}$$

This is true for any $w^* \in \Omega^*$ and the theorem is proved. $\square$

The inequality (4.16) is essential for the convergence of the Linearized alternating direction method. By using (4.4), the result of Theorem 4.1 can be written as

$$\|v^{k+1} - v^*\|_G^2 \leq \|v^k - v^*\|_G^2 - \frac{2-\gamma}{\gamma}\|v^k - v^{k+1}\|_G^2, \ \forall\, w^* \in \Omega^*.$$

5 Convergence rate of ADMM-CPPA Method

Lemma 5.1 *Let $\{w^k\}$ be the sequence generated by the customized PPA (4.2) with (4.4). Then, we have*

$$(\tilde{v}^k - \tilde{v}^{k+1})^T G\{(v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\} \geq 0. \tag{5.1}$$

Proof. Set $w = \tilde{w}^{k+1}$ in (4.11), we have

$$\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T\{F(\tilde{w}^k) + Q(\tilde{v}^k - v^k)\} \geq 0. \tag{5.2}$$

Note that (4.11) is also true for $k := k+1$ and thus we have

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

Set $w = \tilde{w}^k$ in the above inequality, we obtain

$$\theta(\tilde{u}^k) - \theta(\tilde{u}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T\{F(\tilde{w}^{k+1}) + Q(\tilde{v}^{k+1} - v^{k+1})\} \geq 0. \tag{5.3}$$

Adding (5.2) and (5.3) and using the monotonicity of F, we get

$$(\tilde{w}^k - \tilde{w}^{k+1})^T Q\{(v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\} \geq 0.$$

we obtain (5.1) immediately. $\square$

Lemma 5.2 *Let $\{w^k\}$ be the sequence generated by the customized PPA (4.2) with (4.4). Then, we have*

$$\begin{aligned} &(v^k - \tilde{v}^k)^T G\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \\ \geq \quad & \frac{1}{\gamma}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_G^2. \end{aligned} \tag{5.4}$$

Proof. Adding the term $\|(v^k - v^{k+1}) - (\tilde{w}^k - \tilde{w}^{k+1})\|_G^2$ to the both sides of (5.1), we obtain

$$\begin{aligned} &(v^k - v^{k+1})^T G\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \\ \geq \quad & \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_G^2. \end{aligned} \tag{5.5}$$

Substituting the term $(v^k - v^{k+1})$ in the left hand side of (5.5) by $\gamma(v^k - \tilde{v}^k)$ (see (4.4)), we obtain (5.4) and the lemma is proved. $\square$

Lemma 5.3 *Let $\{v^k\}$ be the sequence generated by the customized PPA (4.2) with (4.4). Then, we have*

$$\|v^{k+1} - \tilde{v}^{k+1}\|_G^2 \le \|v^k - \tilde{v}^k\|_G^2. \tag{5.6}$$

Proof. Setting $a = v^k - \tilde{v}^k$ and $b = v^{k+1} - \tilde{w}^{k+1}$ in the identity

$$\|a\|_G^2 - \|b\|_G^2 = 2a^T G(a-b) - \|a-b\|_G^2,$$

we obtain

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

By using (5.4) to the first term of the right hand side of the last equality, we obtain

$$\|v^k - \tilde{v}^k\|_G^2 - \|v^{k+1} - \tilde{v}^{k+1}\|_G^2 \ge \frac{2-\gamma}{\gamma}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_G^2.$$

The assertion of this lemma is proved. $\square$

Having the assertion (4.16) and Lemma 5.3, we are ready to present the $O(1/t)$ convergence rate of the customized PPA in the residue sense.

Theorem 5.1 *Let $\{w^k\}$ be the sequence generated by the customized PPA (4.2) with (4.4). Then, we have*

$$\|v^k - \tilde{v}^k\|_G^2 \le \frac{1}{(k+1)\gamma(2-\gamma)}\|v^0 - v^*\|_G^2, \quad \forall w^* \in \Omega^*. \tag{5.7}$$

Proof. First, it follows from (4.16) that

$$\gamma(2-\gamma)\sum_{t=0}^{\infty}\|v^t - \tilde{v}^t\|_G^2 \le \|v^0 - v^*\|_G^2, \quad \forall\, w^* \in \Omega^*. \tag{5.8}$$

According to Lemma 5.3, the sequence $\{\|v^t - \tilde{v}^t\|_G^2\}$ is non-increasing. Therefore, we have

$$(k+1)\|v^k - \tilde{v}^k\|_G^2 \le \sum_{i=0}^{k} \|v^i - \tilde{v}^i\|_G^2. \tag{5.9}$$

The assertion of this theorem follows from (5.8) and (5.9) directly. $\square$

The solution set of the variational inequality VI(Ω, F, θ) is convex and closed. Theorem 5.1 indicates that ADMM has $O(1/k)$ iteration convergence rate. Let

$$d = \inf\{\|v^0 - v^*\|_G \,|\, w^* \in \Omega^*\}.$$

For any given $\epsilon > 0$, in order to enforce the error $\|v^k - \tilde{v}^k\|_G^2 \le \epsilon$, according to (5.7), it needs at most $k = \lfloor d^2/\gamma(2-\gamma)\epsilon \rfloor$ iterations.

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, **3**, 1-122, 2010.

[2] X.J. Cai, G.Y. Gu, B.S. He and X.M. Yuan, A relaxed customized proximal point algorithm for separable convex programming. http://www.optimization-online.org/DB_HTML/2011/08/3141.html

[3] J. Eckstein and M. C. Ferris, Operator-splitting methods for monotone affine variational inequalities, with a parallel application to optimal control, INFORMS J. Comput., **10**, 218-235, 1998.

[4] D. Gabay, Applications of the Method of Multipliers to Variational Inequalities, in *Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems*, M. Fortin and R. Glowinski, eds., North Holland, Amsterdam, The Netherlands, 299-331, 1983.

[5] E. G. Gol'shtein and N. V. Tret'yakov, Modified Lagrangian in convex programming and their generalizations, Math. Program. Studies, **10**, 86-97, 1979.

[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 uniform approach. http://www.optimization-online.org/DB_HTML/2011/10/ 3220.html

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

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

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

凸优化和单调变分不等式的收缩算法

第十四讲: 自变量 x-y 平等的 对称型交替方向法

Symmetric ADMM

南京大学数学系　何炳生
hebma@nju.edu.cn

The context of this lecture is based on the manuscript [4]

1 Introduction

We consider the following structured constrained convex optimization problem

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

where $\theta_1(x):\Re^{n_1}\to\Re,\ \theta_2(y):\Re^{n_2}\to\Re$ are convex functions (but not necessary smooth), $A\in\Re^{m\times n_1}$, $B\in\Re^{m\times n_2}$ and $b\in\Re^m$, $\mathcal{X}\subset\Re^{n_1}$, $\mathcal{Y}\subset\Re^{n_2}$ are given closed convex sets.

First, as the work [18, 19] for analyzing the convergence of ADMM, we need a variational inequality (VI) reformulation of the model (1.1) and a characterization of its solution set. More specifically, solving (1.1) is equivalent to finding $w^*=(x^*,y^*,\lambda^*)\in\Omega$ such that

$$\text{VI}(\Omega,F,\theta):\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega, \quad (1.2a)$$

where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \qquad \theta(u) = \theta_1(x) + \theta_2(y), \qquad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix} \tag{1.2b}$$

$$F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m \tag{1.2c}$$

Since the mapping $F(w)$ defined in (1.2c) is affine with a skew-symmetric matrix, it is thus monotone. We denote by Ω^* the solution set of VI(Ω, F, θ), and consider the problem under the nonempty assumption onto Ω^*.

The augmented Lagrangian function of the problem (1.1) is defined by

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

Applied ADMM to the problem (1.1)

From given $v^k = (y^k, \lambda^k)$, the iteration produces $w^k = (x^{k+1}, y^{k+1}, \lambda^{k+1})$.

The iteration scheme:

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

$$x^{k+1} \in \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) | x \in \mathcal{X}\}. \tag{1.3a}$$

2. Use λ^k and the obtained x^{k+1}, find y^{k+1} by

$$y^{k+1} \in \text{Argmin}\left\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \middle| y \in \mathcal{Y}\right\}. \tag{1.3b}$$

3. Set the multipliers by

$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \tag{1.3c}$$

The sequence $\{v^k\}$ generated by (1.4) satisfies

$$\|v^{k+1} - v^*\|_{H_c}^2 \le \|v^k - v^*\|_{H_c}^2 - \|v^k - v^{k+1}\|_{H_c}^2, \; H_c = \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}.$$

Applied ADMM-based customized PPA to the problem (1.1)

From given $v^k = (y^k, \lambda^k)$, the prediction step produces $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$.

The prediction step:

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

$$\tilde{x}^k = \in \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) | x \in \mathcal{X}\}. \tag{1.4a}$$

2. Set the multipliers by

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b). \tag{1.4b}$$

3. Finally, use the obtained $\tilde{x}^k$ and $\tilde{\lambda}^k$, find $\tilde{y}^k$ by

$$\tilde{y}^k = \text{Argmin}\{\mathcal{L}_\beta(\tilde{x}^k, y, \tilde{\lambda}^k) | y \in \mathcal{Y}\}. \tag{1.4c}$$

The new iterate v^{k+1} is given by

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

With given v^k, the predictor $\tilde{w}^k$ is generated by an "order changed ADMM".
As analyzed in the last section, for the predictor $\tilde{w}^k \in \Omega$, we have

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T H_c(v^k - \tilde{v}^k) \geq 0, \ \forall\, w \in \Omega, \tag{1.6}$$

where

$$H_c = \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \tag{1.7}$$

Set $w = w^*$ in (1.6), we get

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

and thus

$$(v^k - v^*)^T H_c(v^k - \tilde{v}^k) \geq \|v^k - \tilde{v}^k\|_{H_c}^2$$

Using the above inequality, it is easy to show that generated sequence $\{v^k\}$ by (1.4)-(1.5) satisfies

$$\|v^{k+1} - v^*\|_{H_c}^2 \leq \|v^k - v^*\|_{H_c}^2 - \alpha(2 - \alpha)\|v^k - \tilde{v}^k\|_{H_c}^2.$$

Symmetric ADMM for the problem (1.1)

In practice, the primal variables x and y should be treated fairly. In this paper, we propose the following iterative scheme for the problem (1.1). Again, $v = (y, \lambda)$ are the essential variables and x is the intermediate variable. The k-th iteration starts with $v^k = (y^k, \lambda^k)$, produces $v^{k+1} = (y^{k+1}, \lambda^{k+1})$ via

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(Ax^{k+1} + By^k - b), \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \,|\, y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \alpha\beta(Ax^{k+1} + By^{k+1} - b), \end{cases} \tag{1.8}$$

where $\alpha \in (0, 1)$. If $\alpha = 1$, the method is the Peaceman-Rachford splitting method (PRSM) in [20, 21] to the problem (1.1).

We will prove the sequence $\{v^k\}$ is strictly contractiveto the solution set and establish a worst-case $O(1/t)$ convergence rate for the PRSM scheme (2.1).

2 Preliminaries

In this section, we summarize some useful preliminaries known in the literature and prove some simple conclusions for further analysis.

Let $\alpha = 1$, the iterative scheme of SPRSM (1.8) for (1.1) becomes

$$\begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k - b), \\ y^{k+1} = \arg\min\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \,|\, y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} + By^{k+1} - b), \end{cases} \tag{2.1}$$

where $\lambda \in \Re^m$ is the Lagrange multiplier associated with the linear constraints in (1.1) and $\beta > 0$ is a penalty parameter. As analyzed in [9], the PRSM scheme (2.1) differs from ADMM “only through the addition of the intermediate update of the multipliers ($\lambda^{k+\frac{1}{2}}$); it thus offers the same set of advantages”. The PRSM scheme (2.1), however, according to [9] again (see also [14]), “is less ‘robust’ in

that it converges under more restrictive assumptions than ADMM" and "if it does converge, then its rate of convergence is faster". We refer to [1, 12] for some numerical verification of the efficiency of PRSM.

According to Theorem 2.3.5 in [7], a very useful characterization of the solution set Ω^* of VI(Ω, F, θ) can be summarized in the following theorem. Its proof can be found in [7, 18].

Theorem 2.1 *The solution set of VI(Ω, F, θ) is closed and convex; and it can be characterized as*

$$\Omega^* = \bigcap_{w \in \Omega} \{\tilde{w} \in \Omega : \big(\theta(u) - \theta(\tilde{u})\big) + (w - \tilde{w})^T F(w) \geq 0\}. \tag{2.2}$$

Theorem 2.1 thus implies that $\tilde{w} \in \Omega$ is an approximate solution of VI(Ω, F, θ) with an accuracy of $O(1/t)$ if it satisfies

$$\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w) \leq \epsilon, \quad \forall w \in \Omega, \tag{2.3}$$

with $\epsilon = O(1/t)$. In fact, this characterization makes it possible to analyze the

convergence rate of ADMM and other splitting methods via the VI approach rather than the conventional approach based on the functional values in the literature. In the following, we shall show that either the sequence (2.1) or (1.8) provides us such an iterate satisfying (2.3) after t iterations.

As mentioned in [2] for ADMM, the variable x is an intermediate variable during the PRSM iteration since it essentially requires only (y^k, λ^k) in (2.1) or (1.8) to generate the $(k+1)$-th iterate. For this reason, we define the notations $v^k = (y^k, \lambda^k)$, $\mathcal{V} = \mathcal{Y} \times \Re^m$,

$$\mathcal{V}^* := \{v^* = (y^*, \lambda^*) \,|\, w^* = (x^*, y^*, \lambda^*) \in \Omega^*\};$$

and it suffices to analyze the convergence rate of the sequence $\{v^k\}$ to the set $\mathcal{V}^*$ in order to study the convergence rate of the sequence $\{w^k\}$ generated by (2.1) or (1.8). Note $\mathcal{V}^*$ is also closed and convex.

Finally, we define some matrices in order to present our analysis in a compact

way. Let

$$M = \begin{pmatrix} I_{n_2} & 0 \\ -\alpha\beta B & 2\alpha I_m \end{pmatrix}. \tag{2.4}$$

and

$$\widetilde{Q} = \begin{pmatrix} 0 & 0 \\ \beta B^T B & -\alpha B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \tag{2.5}$$

In (2.5), "0" is a matrix with all zero entries in appropriate dimensionality. We further define

$$Q = \begin{pmatrix} \beta B^T B & -\alpha B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \tag{2.6}$$

as the submatrix of $\widetilde{Q}$ excluding all the first zero rows. The matrices $\widetilde{Q}$ and Q are associated with the analysis for the sequences $\{w^k\}$ and $\{v^k\}$, respectively.

Last, for $\alpha \in (0, 1]$ we define a symmetric matrix

$$H = \frac{1}{2} \begin{pmatrix} (2-\alpha)\beta B^T B & -B^T \\ -B & \frac{1}{\alpha\beta} I_m \end{pmatrix}. \tag{2.7}$$

Lemma 2.1 *The matrix H defined in (2.7) is positive definite for $\alpha \in (0, 1)$ and positive semi-definite for $\alpha = 1$.*

Proof We have

$$H = \frac{1}{2} \begin{pmatrix} \sqrt{\beta} B^T & 0 \\ 0 & \sqrt{\frac{1}{\beta}} I \end{pmatrix} \begin{pmatrix} (2-\alpha) I & -I \\ -I & \frac{1}{\alpha} I \end{pmatrix} \begin{pmatrix} \sqrt{\beta} B & 0 \\ 0 & \sqrt{\frac{1}{\beta}} I \end{pmatrix}.$$

Note that the matrix

$$\begin{pmatrix} (2-\alpha) & -1 \\ -1 & \frac{1}{\alpha} \end{pmatrix}$$

is positive definite if $\alpha \in (0, 1)$ and positive semi-definite if $\alpha = 1$. Thus

assertion is thus proved. $\square$.

Lemma 2.2 *The matrices M, Q and H defined respectively in* (2.4)*,* (2.6) *and* (2.7)*) have the following relationships:*

$$HM = Q \tag{2.8}$$

and

$$G = Q^T + Q - M^T HM \succeq \frac{(1-\alpha)}{2(1+\alpha)} M^T HM. \tag{2.9}$$

Proof. Using the definition of the matrices M, Q and H, by a simple manipulation, we obtain

$$\begin{aligned} HM &= \frac{1}{2}\begin{pmatrix} (2-\alpha)\beta B^T B & -B^T \\ -B & \frac{1}{\alpha\beta} I_m \end{pmatrix}\begin{pmatrix} I & 0 \\ -\alpha\beta B & 2\alpha I_m \end{pmatrix} \\ &= \frac{1}{2}\begin{pmatrix} 2\beta B^T B & -2\alpha B^T \\ -2B & \frac{2}{\beta} I_m \end{pmatrix} = Q. \end{aligned}$$

The first assertion is proved. Consequently, we get

$$\begin{aligned} M^T HM &= M^T Q = \begin{pmatrix} I_{n_2} & -\alpha\beta B^T \\ 0 & 2\alpha I_m \end{pmatrix}\begin{pmatrix} \beta B^T B & -\alpha B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \\ &= \begin{pmatrix} (1+\alpha)\beta B^T B & -2\alpha B^T \\ -2\alpha B & \frac{2\alpha}{\beta} I_m \end{pmatrix}. \end{aligned} \tag{2.10}$$

Using (2.6) and the above equation, we have

$$G = (Q^T + Q) - M^T HM = (1-\alpha)\begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{2}{\beta} I_m \end{pmatrix}. \tag{2.11}$$

Since $\begin{pmatrix} \beta & -1 \\ -1 & \frac{2}{\beta} \end{pmatrix}$ is positive definite, we say G is essentially positive definite .

Note that

$$2(1+\alpha)\begin{pmatrix} \beta B^TB & -B^T \\ -B & \frac{2}{\beta}I_m \end{pmatrix} - M^THM = \begin{pmatrix} (1+\alpha)\beta B^TB & -2B^T \\ -2B & \frac{4+2\alpha}{\beta}I_m \end{pmatrix}. \tag{2.12}$$

Because

$$\begin{pmatrix} (1+\alpha) & -2 \\ -2 & 4+2\alpha \end{pmatrix} \succeq 0, \quad \forall\, \alpha \geq 0,$$

the right-hand side of (2.12) is positive semidefinite. Thus, it follows that

$$\begin{pmatrix} \beta B^TB & -B^T \\ -B & \frac{2}{\beta}I_m \end{pmatrix} \succeq \frac{1}{2(1+\alpha)} M^THM. \tag{2.13}$$

Substituting (2.13) in (2.11), we obtain (2.9) and the lemma is proved. □

Remark 2.1 *When $\alpha = 1$, the matrices H defined in* (2.7) *and $Q^T + Q - M^THM$ are both positive semi-definite. But, in the following analysis we still use $\|v - \tilde{v}\|_H$ and $\|v - \tilde{v}\|_{Q^T+Q-M^THM}$ to denote*

respectively

$$\|v - \tilde{v}\|_H = \big((v-\tilde{v})^T H (v - \tilde{v})\big)^{1/2}$$

and

$$\|v - \tilde{v}\|_{Q^T+Q-M^THM} = \big((v-\tilde{v})^T (Q^T + Q - M^THM)(v - \tilde{v})\big)^{1/2},$$

for $v, \tilde{v} \in \mathcal{Y} \times \Re^m$. This slight abuse of notation will alleviate the notation in our analysis greatly.

We also have the following lemma.

Lemma 2.3 *Let $\{w^k\}$ be the sequence generated by the strictly contractive PRSM (1.8), and H be defined in (2.7). Then, w^{k+1} is a solution of VI(Ω, F, θ) if $\|v^k - v^{k+1}\|_H^2 = 0$.*

Proof. See Lemma 3.2 in [19]. □

3 Contraction Analysis

In this section, we analyze the contraction property for the sequence $\{v^k\}$ generated by the PRSM scheme (2.1) or the strictly contractive PRSM scheme (1.8) with respect to the set $\mathcal{V}^*$. The convergence rate analysis for (2.1) and (1.8) to be presented is based on this analysis of contraction property. Since (2.1) can be included by the strictly contractive PRSM scheme (1.8) if we extend the value of $\alpha = 1$ and the algebra of convergence analysis for these two schemes are of the same framework, below we only present the contraction analysis for (1.8) and the analysis for (2.1) is readily obtained by taking $\alpha = 1$ in our analysis.

First, to further alleviate the notation in our analysis, we need to define an auxiliary sequence $\{\tilde{w}^k\}$ as

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k - \beta(Ax^{k+1} + By^k - b) \end{pmatrix}, \tag{3.1}$$

where (x^{k+1}, y^{k+1}) is generated by (2.1) or (1.8). Note with the notation of $\tilde{w}^k$, the strictly contractive PRSM (1.8) can be written as

$$\left\{ \begin{array}{l} \tilde{x}^k = \arg\min \left\{ \theta_1(x) - (\lambda^k)^T Ax + \frac{\beta}{2}\|Ax + By^k - b\|^2 \;\middle|\; x \in \mathcal{X} \right\}, \\ \tilde{y}^k = \arg\min \left\{ \begin{array}{c} \theta_2(y) - [\lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k)]^T By \\ +\frac{\beta}{2}\|A\tilde{x}^k + By - b\|^2 \end{array} \;\middle|\; y \in \mathcal{Y} \right\}. \end{array} \right. \tag{3.2}$$

Then, based on (1.8) and (3.1), we immediately have

$$x^{k+1} = \tilde{x}^k, \qquad y^{k+1} = \tilde{y}^k \qquad \text{and} \qquad \lambda^{k+\frac{1}{2}} = \tilde{\lambda}^k$$

and

$$\begin{aligned} \lambda^{k+1} &= \lambda^{k+1/2} - \alpha\beta(A\tilde{x}^k + B\tilde{y}^k - b) \\ &= \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k) - \alpha\big[\beta(A\tilde{x}^k + By^k - b) - \beta B(y^k - \tilde{y}^k)\big] \\ &= \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k) - \alpha\big[(\lambda^k - \tilde{\lambda}^k) - \beta B(y^k - \tilde{y}^k)\big] \\ &= \lambda^k - \big[2\alpha(\lambda^k - \tilde{\lambda}^k) - \alpha\beta B(y^k - \tilde{y}^k)\big]. \end{aligned} \tag{3.3}$$

Furthermore, together with $y^{k+1} = \tilde{y}^k$, we have the following relationship

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I_{n_2} & 0 \\ -\alpha\beta B & (1+\alpha)I_m \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix};$$

which can be rewritten into a compact form by using the notations of v^k and $\tilde{v}^k$:

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \tag{3.4}$$

where M is defined in (2.4).

Now, we start to prove some properties for the sequence $\{\tilde{w}^k\}$ defined in (3.1). Recall our primary purpose is to analyze the convergence rate for the sequences (2.1) and (1.8) based on the solution characterization (2.2), and the accuracy of an approximate solution $\tilde{w} \in \Omega$ is measured by a upper bound of the quantity of $\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)$ for all $w \in \Omega$ (see (2.3)). Hence, we are interested in estimating how accurate the point $\tilde{w}^k$ defined in (3.1) is to a solution point of VI(Ω, F, θ). The main result is proved in Theorem 3.1. But before that, we first show two lemmas. The first lemma presents a upper bound of

$\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)$ for all $w \in \Omega$ in term of a quadratic term involving the matrix Q.

Lemma 3.1 *For given $v^k \in \mathcal{Y} \times \Re^m$, let w^{k+1} be generated by the strictly contractive PRSM scheme (1.8) and $\tilde{w}^k$ be defined in* (3.1)*. Then, we have $\tilde{w}^k \in \Omega$ and*

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall\, w \in \Omega, \tag{3.5}$$

where the matrix Q is defined in (2.6)*.*

Proof. Since $x^{k+1} = \tilde{x}^k$, by deriving the first-order optimality condition of the x-minimization problem in (3.2), we have

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{A^T[\beta(A\tilde{x}^k + By^k - b) - \lambda^k]\} \geq 0, \ \forall\, x \in \mathcal{X}. \tag{3.6}$$

According to the definition (3.1), we have

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b). \tag{3.7}$$

Using (3.7), the inequality (3.6) can be written as

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T\{-A^T\tilde{\lambda}^k\} \geq 0, \quad \forall\, x \in \mathcal{X}. \tag{3.8}$$

Similarly, by deriving the first-order optimality condition of the y-minimization problem in (3.2), we get

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T B^T \left\{ \begin{array}{c} \beta(A\tilde{x}^k + B\tilde{y}^k - b) \\ -[\lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k)] \end{array} \right\} \geq 0, \ \forall\, y \in \mathcal{Y}. \tag{3.9}$$

Again, using (3.7), we have

$$\begin{aligned} & \beta(A\tilde{x}^k + B\tilde{y}^k - b) - [\lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k)] \\ = \quad & -(\tilde{\lambda}^k - \lambda^k) + \beta B(\tilde{y}^k - y^k) - [\lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k)] \\ = \quad & -\tilde{\lambda}^k + \beta B(\tilde{y}^k - y^k) - \alpha(\tilde{\lambda}^k - \lambda^k). \end{aligned}$$

Consequently, it follows from (3.9) that

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \left\{ \begin{array}{c} -B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) \\ -\alpha B^T(\tilde{\lambda}^k - \lambda^k) \end{array} \right\} \geq 0, \ \forall\, y \in \mathcal{Y}. \tag{3.10}$$

In addition, based on (3.1) we have

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0. \tag{3.11}$$

Combining (3.8), (3.10) and (3.11) together, we get $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$; and for any $w = (x, y, \lambda) \in \Omega$, it holds

$$\begin{aligned} & \theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} -A^T\tilde{\lambda}^k \\ -B\tilde{\lambda}^k \\ A\tilde{x}^k + B\tilde{y}^k - b \end{pmatrix} \\ & + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} 0 \\ \beta B^T B(\tilde{y}^k - y^k) - \alpha B^T(\tilde{\lambda}^k - \lambda^k) \\ -B(\tilde{y}^k - y^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) \end{pmatrix} \geq 0. \end{aligned}$$

The assertion (3.5) is only a compact form of the above inequality by using the

notations of Q in (2.6), w and F in (1.2b) and v. The proof is complete. $\square$

The second lemma aims at expressing the bound $(v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k)$ found in Lemma 3.1 by the difference of two quadratic terms involving two consecutive iterates of the sequence $\{v^k\}$ and a quadratic term involving v^k and the auxiliary iterate $\tilde{v}^k$. This equivalent expression is convenient for the manipulation over the whole sequence $\{v^k\}$ recursively and thus for establishing the convergence rate of $\{v^k\}$ in either ergodic or nonergodic sense.

Lemma 3.2 *Let $\{w^k\}$ be generated by the strictly contractive PRSM scheme* (1.8) *and $\{\tilde{w}^k\}$ be defined in (3.1); M, Q and H be defined in (2.4), (2.6) and (2.7), respectively. Then we have*

$$\begin{aligned}(v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k) \;=\; & \frac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)\\ & +\frac{(1-\alpha)}{4(1+\alpha)}\|M(v^k-\tilde{v}^k)\|_H^2,\ \forall v\in\mathcal{V}. \qquad (3.12)\end{aligned}$$

Proof. By using $Q=HM$ and $M(v^k-\tilde{v}^k)=(v^k-v^{k+1})$ (see (3.4)), it

follows that

$$\begin{aligned}(v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k) &= (v-\tilde{v}^k)^T HM(v^k-\tilde{v}^k)\\ &= (v-\tilde{v}^k)^T H(v^k-v^{k+1}). \qquad (3.13)\end{aligned}$$

For the vectors a,b,c,d in the same space and a matrix H with appropriate dimensionality, we have the identity

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

In this identity, we take

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

and submit it to the right-hand side of (3.13). The resulting equation is

$$\begin{aligned}(v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k) \;=\; & \frac{1}{2}\big(\|v-v^k\|_H^2-\|v-v^{k+1}\|_H^2\big)\\ & +\frac{1}{2}(\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2). \qquad (3.14)\end{aligned}$$

Now, we deal with the last term of the right-hand side of (3.14). By using (3.4) and (2.8), we get

$$\begin{aligned}
&\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\
&\quad = \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\
&\quad \overset{(3.4)}{=} \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\
&\quad = \quad 2(v^k - \tilde{v}^k)HM(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)M^T HM(v^k - \tilde{v}^k) \\
&\quad \overset{(2.8)}{=} \quad (v^k - \tilde{v}^k)^T(Q^T + Q - M^T HM)(v^k - \tilde{v}^k) \\
&\quad \overset{(2.9)}{\geq} \quad \frac{(1-\alpha)}{2(1+\alpha)}(v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\
&\quad = \quad \frac{(1-\alpha)}{2(1+\alpha)}\|M(v^k - \tilde{v}^k)\|_H^2.
\end{aligned}$$

Substituting it in (3.14), we obtain the assertion (3.12). The proof is complete. □

Now we are ready to present an inequality where a upper bound of $\theta(\tilde{u}^k) - \theta(u) + (\tilde{w}^k - w)^T F(w)$ is found for all $w \in \Omega$. This inequality is

also crucial for analyzing the contraction property and the convergence rate for the iterative sequence generated by either (2.1) or (1.8).

Theorem 3.1 *For given $v^k \in \mathcal{Y} \times \Re^m$, let w^{k+1} be generated by the strictly contractive PRSM scheme (1.8) and $\tilde{w}^k$ be defined in* (3.1)*; M and H be defined in (2.4) and (2.7), respectively. Then, we have $\tilde{w}^k \in \Omega$ and*

$$\begin{aligned}
&\theta(\tilde{u}^k) - \theta(u) + (\tilde{w}^k - w)^T F(w) \\
&\quad \leq \quad \frac{1}{2}\big(\|v - v^k\|_H^2 - \|v - v^{k+1}\|_H^2\big) \\
&\qquad\quad - \frac{1-\alpha}{4(1+\alpha)}\|M(v^k - \tilde{v}^k)\|_H^2, \ \forall w \in \Omega.
\end{aligned} \tag{3.15}$$

Proof. First, by using the structure of $F(w)$ (see (1.2c)), we have

$$(w - \tilde{w})^T(F(w) - F(\tilde{w})) = 0, \ \ \forall w, \ \tilde{w} \in \Omega.$$

Then, using the above equality and replacing the right-hand side term in (3.5) with the identity (3.12), we obtain the assertion (3.15). The proof is complete. □

The assertion (3.15) also enables us to study the contraction property of the sequence $\{v^k\}$ generated by (2.1) or (1.8). In fact, setting $w = w^*$ in (3.15) where w^* being an arbitrary solution point in Ω^*, we get

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

which implies that

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \frac{1-\alpha}{2(1+\alpha)}\|M(v^k - \tilde{v}^k)\|_H^2. \tag{3.16}$$

Recall the identity (3.4). We thus have

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \frac{1-\alpha}{2(1+\alpha)}\|v^k - v^{k+1}\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{3.17}$$

Therefore, when $\alpha = 1$, i.e., for the PRSM scheme (2.1), we have

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2, \tag{3.18}$$

which means the sequence $\{v^k\}$ generated by (2.1) is contractive, but not strictly, to the set $\mathcal{V}^*$. In fact, it is possible that the sequence $\{v^k\}$ stays away from the solution set with a constant distance, hence no convergence is guaranteed by (3.18). In [5], such an example is constructed. On the other hand, when $\alpha \in (0,1)$, the inequality (3.17) ensures a reduction of $\frac{1-\alpha}{2(1+\alpha)}\|v^k - v^{k+1}\|_H^2$ to the set $\mathcal{V}^*$ at the $(k+1)$-th iteration, i.e., the strict contraction of $\{v^k\}$ is guaranteed for the sequence generated by (1.8). We can thus expect that the strictly contractive PRSM (1.8) converges faster to $\mathcal{V}^*$ than the original PRSM (2.1). As we have mentioned, the difference of contraction between (3.17) and (3.18) is also the reason why we can establish a nonergodic convergence rate for the strictly contractive PRSM (1.8) while only an ergodic convergence rate can be established for the original PRSM (2.1).

4 Convergence rate of (2.1) in ergodic sense

In this section, we show that although the original PRSM (2.1) might not be convergent to a solution point of the model (1.1), it is still possible to find an approximate solution of VI(Ω, F, θ) with an accuracy of $O(1/t)$ based on the first t iterations of the PRSM scheme (2.1). This estimate helps us better understand the convergence property of the original PRSM (2.1).

Theorem 4.1 *Let $\{w^k\}$ be generated by* (2.1) *and $\{\tilde{w}^k\}$ be defined by (3.1). Let $\tilde{w}_t$ be defined as*

$$\tilde{w}_t = \frac{1}{t+1}\sum_{k=0}^{t}\tilde{w}^k. \tag{4.1}$$

Then, for any integer number $t > 0$, $\tilde{w}_t \in \Omega$ and

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2(t+1)}\|v - v^0\|_H^2, \quad \forall w \in \Omega, \tag{4.2}$$

where H is defined in (2.7).

Proof. First, because of (3.1), it holds that $\tilde{w}^k \in \Omega$ for all $k \ge 0$. Together with the convexity of $\mathcal{X}$ and $\mathcal{Y}$, (4.1) implies that $\tilde{w}_t \in \Omega$. Second, by taking $\alpha = 1$ in (3.15) (and using $(w - \tilde{w})^T F(\tilde{w}) = (w - \tilde{w})^T F(w)$) we have

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2}\|v - v^k\|_H^2 \ge \frac{1}{2}\|v - v^{k+1}\|_H^2, \ \forall w \in \Omega. \tag{4.3}$$

Summing the inequality (4.3) over $k = 0, 1, \ldots, t$, we obtain

$$(t+1)\theta(u) - \sum_{k=0}^{t}\theta(\tilde{u}^k) + \Big((t+1)w - \sum_{k=0}^{t}\tilde{w}^k\Big)^T F(w) + \frac{1}{2}\|v - v^0\|_H^2 \ge 0,$$

for all $w \in \Omega$. Use the notation of $\tilde{w}_t$, it can be written as

$$\frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2(t+1)}\|v - v^0\|_H^2, \quad \forall w \in \Omega. \tag{4.4}$$

Since $\theta(u)$ is convex and

$$\tilde{u}_t = \frac{1}{t+1}\sum_{k=0}^{t}\tilde{u}^k,$$

we have that

$$\theta(\tilde{u}_t) \le \frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k).$$

Substituting it in (4.4), the assertion of this theorem follows directly. $\square$

Let $v^0 = (y^0, \lambda^0)$ be the initial iterate. For a given compact set $\mathcal{D} \subset \mathcal{Y} \times \Re^m$, let $d = \sup\{\|v - v^0\|_H \,|\, v \in \mathcal{D}\}$. Then, after t iterations of the PRSM (2.1), the point $\tilde{w}_t \in \Omega$ defined in (4.1) satisfies

$$\sup_{w\in\mathcal{D}}\left\{\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w)\right\} \le \frac{d^2}{2(t+1)},$$

which means $\tilde{w}_t$ is an approximate solution of VI(Ω, F, θ) with an accuracy of $O(1/t)$ (recall (2.3)).

Remark 4.2 *In the proof of Theorem 4.1, we take* $\alpha = 1$ *in* (4.3)*. Obviously, the*

proof is still valid if we take $\alpha \in (0,1)$*. Thus, a worst-case* $O(1/t)$ *convergence rate in ergodic sense can be established easily for the strictly contractive PRSM (1.8). As we shall show in Section 5, this is less interesting because a nonergodic worst-case* $O(1/t)$ *convergence rate can be established for (1.8). We thus omit the detail.*

5 Convergence rate of (1.8) in nonergodic sense

In this section, we show that the sequence $\{v^k\}$ generated by the strictly contractive PRSM scheme (1.8) is convergent to a point in $\mathcal{V}^*$, and its worst-case convergence rate is $O(1/t)$ in nonergodice sense. Our starting point for the analysis is the inequality (3.17), and a crucial property is the monotonicity of the sequence $\{\|v^k - v^{k+1}\|_H^2\}$. That is, we will prove that

$$\|v^{k+1} - v^{k+2}\|_H^2 \le \|v^k - v^{k+1}\|_H^2, \quad \forall\, k \ge 0.$$

We first take a closer look at the assertion (3.5) in Lemma 3.1.

Lemma 5.1 *Let* $\{w^k\}$ *be the sequence generated by the strictly contractive PRSM* (1.8) *and* $\tilde{w}^k$ *be defined in (3.1); the matrix* Q *be defined in* (2.6)*. Then, we have*

$$(\tilde{v}^k - \tilde{v}^{k+1})^T Q\{(v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\} \geq 0. \tag{5.1}$$

Proof. Set $w = \tilde{w}^{k+1}$ in (3.5), we have

$$\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T F(\tilde{w}^k) \geq (\tilde{v}^{k+1} - \tilde{v}^k)^T Q(v^k - \tilde{v}^k). \tag{5.2}$$

Note that (3.5) is also true for $k := k+1$ and thus

$$\theta(u) - \theta(\tilde{u}^{k+1}) + (w - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \geq (v - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}),$$

for all $w \in \Omega$. Set $w = \tilde{w}^k$ in the above inequality, we obtain

$$\theta(\tilde{u}^k) - \theta(\tilde{u}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \geq (\tilde{v}^k - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}). \tag{5.3}$$

Adding (5.2) and (5.3) and using the monotonicity of F, we get (5.1) immediately.
□

Lemma 5.2 *Let* $\{w^k\}$ *be the sequence generated by the strictly contractive PRSM* (1.8) *and* $\tilde{w}^k$ *be defined in (3.1); the matrices* M, Q *and* H *be defined in* (2.4), (2.6) *and* (2.7)*, respectively. Then, we have*

$$\begin{aligned} &(v^k - \tilde{v}^k)^T M^T H M\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \\ \geq \quad & \frac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)}. \end{aligned} \tag{5.4}$$

Proof. Adding the equation

$$\begin{aligned} &\{(v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\}^T Q\{(v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\} \\ = \quad & \frac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)} \end{aligned}$$

to both sides of (5.1), we get

$$\begin{aligned}&(v^k - v^{k+1})^T Q\{(v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\}\\ \geq\ & \frac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)}.\end{aligned} \tag{5.5}$$

By using (see (3.4) and (2.8))

$$v^k - v^{k+1} = M(v^k - \tilde{v}^k) \qquad \text{and} \qquad Q = HM,$$

to the term $v^k - v^{k+1}$ in the left hand side of (5.5), we obtain

$$\begin{aligned}&(v^k - \tilde{v}^k)^T M^T HM\{(v^k - v^{k+1}) - (\tilde{v}^k - \tilde{v}^{k+1})\}\\ \geq\ & \frac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)}.\end{aligned}$$

and the lemma is proved. $\square$

Now, we are ready to prove the monotonicity of the sequence $\{\|v^k - v^{k+1}\|^2_H\}$.

Theorem 5.1 *Let $\{w^k\}$ be the sequence generated by the strictly contractive PRSM* (1.8) *and $\tilde{w}^k$ be defined in (3.1); the matrix H be defined in* (2.7)*. Then,*

we have

$$\|v^{k+1} - v^{k+2}\|^2_H \leq \|v^k - v^{k+1}\|^2_H. \tag{5.6}$$

Proof. Setting $a = M(v^k - \tilde{v}^k)$ and $b = M(v^{k+1} - \tilde{v}^{k+1})$ in the identity

$$\|a\|^2_H - \|b\|^2_H = 2a^T H(a-b) - \|a-b\|^2_H,$$

we obtain

$$\begin{aligned}&\|M(v^k - \tilde{v}^k)\|^2_H - \|M(v^{k+1} - \tilde{v}^{k+1})\|^2_H\\ =\ & 2(v^k - \tilde{v}^k)^T M^T HM\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\}\\ & -\|M(v^k - \tilde{v}^k) - M(v^{k+1} - \tilde{v}^{k+1})\|^2_H.\end{aligned}$$

Inserting (5.4) into the first term of the right-hand side of the last equality, we obtain

$$\begin{aligned}&\|M(v^k - \tilde{v}^k)\|^2_H - \|M(v^{k+1} - \tilde{v}^{k+1})\|^2_H\\ \geq\ & \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q-M^T HM)}.\end{aligned}$$

The assertion (5.6) follows from the above inequality, the relationship (3.4) and

(2.9) immediately. The proof is complete. □

Now, we can establish a worst-case $O(1/t)$ convergence rate in nonergodic sense for the strictly contractive PRSM scheme (1.8).

Theorem 5.2 *Let $\{w^t\}$ be the sequence generated by the strictly contractive PRSM scheme* (1.8)*. For any $v^* \in \mathcal{V}^*$, we have*

$$\|v^t - v^{t+1}\|_H^2 \le \frac{2(1+\alpha)}{(t+1)(1-\alpha)}\|v^0 - v^*\|_H^2. \tag{5.7}$$

Proof. First, it follows from (3.17) that

$$\frac{1-\alpha}{2(1+\alpha)}\sum_{k=0}^{\infty}\|v^k - v^{k+1}\|_H^2 \le \|v^0 - v^*\|_H^2, \quad \forall\, v^* \in \mathcal{V}^*. \tag{5.8}$$

According to Theorem 5.1, the sequence $\{\|v^k - v^{k+1}\|_H^2\}$ is monotonically

non-increasing. Therefore, we have

$$(t+1)\|v^t - v^{t+1}\|_H^2 \le \sum_{k=0}^{t}\|v^k - v^{k+1}\|_H^2. \tag{5.9}$$

The assertion (5.7) follows from (5.8) and (5.9) immediately. The proof is complete. □

Notice that $\mathcal{V}^*$ is convex and closed. Let $v^0 = (y^0, \lambda^0)$ be the initial iterate and $d := \inf\{\|v^0 - v^*\|_H \mid v^* \in \mathcal{V}^*\}$. Then, for any given $\epsilon > 0$, Theorem 5.2 shows that the strictly contractive PRSM scheme (1.8) needs at most $\lfloor d^2/\epsilon \rfloor$ iterations to ensure that $\|v^k - v^{k+1}\|_H^2 \le \epsilon$. It follows from Lemma 2.3 that w^{k+1} is a solution of VI(Ω, F, θ) if $\|v^k - v^{k+1}\|_H^2 = 0$. A worst-case $O(1/t)$ convergence rate in nonergodic sense for the strictly contractive PRSM scheme (1.8) is thus established in Theorem 5.2.

References

[1] D.P. Bertsekas, Constrained Optimization and Lagrange Multiplier Methods, Academic Press, 1982.
Ökonometrie und Unternehmensforschung, Springer-Verlag, Berlin-Heidelberg-New York, 1975.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, Nov. 2010.

[3] T. F. Chan and R. Glowinski, *Finite element approximation and iterative solution of a class of mildly non-linear elliptic equations*, technical report, Stanford University, 1978.

[4] X.J. Cai, G.Y. Gu, B.S. He and X.M. Yuan, A relaxed customized proximal point algorithm for separable convex programming. http://www.optimization-online.org/DB_HTML/2011/08/3141.html

[5] E. Corman and X. M. Yuan, *A generalized proximal point algorithm and its convergence rate*, manuscript, 2012.

[6] J. Douglas and H. H. Rachford, *On the numerical solution of the heat conduction problem in 2 and 3 space variables*, Trans. Amer. Math. Soc., 82 (1956), pp. 421-439.

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

[8] M. Fortin and R. Glowinski, Augmented Lagrangian Methods: Applications to the Numerical Solutions of Boundary Value Problems, Stud. Math. Appl. 15, North − Holland, Amsterdam, 1983.

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

[10] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems viafinite element approximations, Computers and Mathematics with Applications, 2:17-40, 1976.

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

[12] R. Glowinski and P. Le Tallec, Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics, SIAM Studies in Applied Mathematics, Philadelphia, PA, 1989.

[13] R. Glowinski, On alternating direction methods of multipliers: A historical perspective, 2012.

[14] R. Glowinski, T. K$\ddot{a}$rkk$\ddot{a}$inen and K. Majava, *On the convergence of operator-splitting methods*, in Numerical Methods for Scienfic computing, Variational Problems and Applications, edited by Y. Kuznetsov, P. Neittanmaki and O. Pironneau, Barcelona, 2003.

[15] E. G. Gol'shtein and N. V. Tret'yakov, *Modified Lagrangian in convex programming and their generalizations*, Math. Program. Studies, 10 (1979), pp. 86-97.

[16] B.S. He, A class of projection and contraction methods for monotone variational inequalities, Applied Mathematics and Optimization, **35**, pp. 69-76, 1997.

[17] B.S. He, "Inexact implicit methods for monotone general variational inequalities", Math. Program. Vol. 86, Series A, pp. 199-217, 1999

[18] B. S. He and X. M. Yuan, *On the O(1/n) convergence rate of Douglas-Rachford alternating direction method*, SIAM J. Numer. Anal. 50 (2012), pp. 700-709.

[19] 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.

[20] P. L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Nmuer. Anal., 16(1979), pp. 964-979.

[21] D. H. Peaceman and H. H. Rachford, *The numerical solution of parabolic elliptic differential equations*, SIAM J. Appl. Math. 3 (1955), pp. 28-41.

凸优化和单调变分不等式的收缩算法

第十五讲: 统一框架下交替方向法的收敛速率

Convergence rate study of ADMM
under a unified framework

南京大学数学系　何炳生
hebma@nju.edu.cn

The context of this lecture is based on the paper [10, 12]

In this note, we discuss the ADMM and ADM-like methods for the separable convex optimization problem in a unified framework, and study the their global convergence and convergence rate.

1 Algorithms in a unified framework

The classical separable convex optimization problem has the following mathematical form:

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

where $\theta_1(x) : \Re^{n_1} \to \Re,\ \theta_2(y) : \Re^{n_2} \to \Re$ are convex functions (but not necessary smooth), $A \in \Re^{m\times n_1}$, $B \in \Re^{m\times n_2}$ and $b \in \Re^m$, $\mathcal{X} \subset \Re^{n_1}$, $\mathcal{Y} \subset \Re^{n_2}$ are given closed convex sets. Let $n = n_1 + n_2$.

The optimal condition of the linearly constrained convex optimization is resulted in a variational inequality: The Lagrangian function of the problem (1.1) is

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

which is defined on $\Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m$. Let $(x^*, y^*, \lambda^*) \in \mathcal{X} \times \mathcal{Y} \times \Re^m$ be an saddle point which satisfies

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

These inequalities can be interpreted as

$$\left\{ \begin{array}{rll} x^* \in \mathcal{X}, & L(x, y^*, \lambda^*) - L(x^*, y^*, \lambda^*) \ge 0, & \forall x \in \mathcal{X} \\ y^* \in \mathcal{Y}, & L(x^*, y, \lambda^*) - L(x^*, y^*, \lambda^*) \ge 0, & \forall y \in \mathcal{Y} \\ \lambda^* \in \Re^m, & L(x^*, y^*, \lambda^*) - L(x^*, y^*, \lambda) \ge 0, & \forall \lambda \in \Re^m \end{array} \right. \tag{1.2}$$

The equivalent expression is the following variational inequality:

$$\left\{ \begin{array}{rll} x^* \in \mathcal{X}, & \theta_1(x) - \theta_1(x^*) + (x - x^*)^T(-A^T\lambda^*) \ge 0, & \forall x \in \mathcal{X} \\ y^* \in \mathcal{Y}, & \theta_2(y) - \theta_2(y^*) + (y - y^*)^T(-B^T\lambda^*) \ge 0, & \forall y \in \mathcal{Y} \\ \lambda^* \in \Re^m, & (\lambda - \lambda^*)^T(Ax^* + By^* - b) \ge 0, & \forall \lambda \in \Re^m \end{array} \right. \tag{1.3}$$

This first order optimal condition (1.3) can be written in a compact form such as

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{1.4a}$$

where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ Ax + By - b \end{pmatrix} \tag{1.4b}$$

$$\theta(u) = \theta_1(x) + \theta_2(y) \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m. \tag{1.4c}$$

We study the algorithms using the guidance of variational inequality. The analysis can be fund in [6] (Sections 4 and 5 therein). Notice that the augmented Lagrangian function of (1.1) is

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

The recursion of the classical alternating direction method of multipliers for the structured convex optimization (1.1) can be written as

$$\text{(ADMM)} \quad \left\{ \begin{array}{l} x^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ y^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \,|\, y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{array} \right. \tag{1.5}$$

Thus, ADMM can be viewed as a relaxed Augmented Lagrangian Method. The main advantage of ADMM is that one can solve the x and y-subproblem separately. Note that the essential variable of ADMM (1.5) is $v = (y, \lambda)$.

The linearized ADMM for (1.1) which taking $v = (y, \lambda)$ as the essential variable can be written as

$$\text{(L-ADMM)} \quad \left\{ \begin{array}{l} x^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ y^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^k) + \frac{1}{2}\|y - y^k\|_D^2 \,|\, y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{array} \right. \tag{1.6}$$

where

$$D = sI - \beta B^T B, \quad s > \beta\rho(B^T B). \tag{1.7}$$

1.1 The algorithmic framework

We study the convergence and the convergence rate of the different ADMM methods. First, we interpret these methods in a prediction-correction framework.

A Prototype Algorithm for (1.4)

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

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall w \in \Omega, \tag{1.8a}$$

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

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

$$v^{k+1} = v^k - \alpha M(v^k - \tilde{v}^k). \tag{1.8b}$$

Convergence Conditions

For the matrices Q and M, and the step size α determined in (1.8), the matrices

$$H = QM^{-1} \tag{1.9a}$$

and

$$G = Q^T + Q - \alpha M^T H M. \tag{1.9b}$$

are positive definite.

1.2 Interpreting ADMM in the framework

The ADMM scheme (1.5) (and its linearized version (1.6)) is also a special case of the prototype algorithm (1.8). Recall the model (1.1) can be explained as the VI (1.4) with the specification given in (1.4b)-(1.4c).

In order to cast the ADMM scheme (1.5) (and (1.6)) into a special case of (1.8), let us first define the artificial vector $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ by

$$\left\{\begin{array}{lcl} \tilde{x}^k & = & x^{k+1}, \\ \tilde{y}^k & = & y^{k+1}, \\ \tilde{\lambda}^k & = & \lambda^k - \beta(Ax^{k+1} + By^k - b), \end{array}\right. \tag{1.10}$$

where (x^{k+1}, y^{k+1}) is generated by the ADMM (1.5). Using these notations, we define the prediction by

$$\left\{\begin{array}{l} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \tilde{y}^k = \text{Argmin}\{\mathcal{L}_\beta(\tilde{x}^k, y, \lambda^k) \,|\, y \in \mathcal{Y}\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b). \end{array}\right. \tag{1.11}$$

According to the scheme (1.5), the defined artificial vector $\tilde{w}^k$ satisfies the following VI:

$$\left\{\begin{array}{ll} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T(-A^T\tilde{\lambda}^k) \ge 0, & \forall\, x \in \mathcal{X}, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T(-B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k)) \ge 0, & \forall\, y \in \mathcal{Y}, \\ \qquad (A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. & \end{array}\right. \tag{1.12}$$

This can be written in form of (1.8a) as described in the following lemma.

Lemma 1.1 *For given v^k, let w^{k+1} be generated by* (1.5) *and $\tilde{w}^k$ be defined by* (1.10)*. Then, we have $\tilde{w}^k \in \Omega$ and*

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

where

$$Q = \left(\begin{array}{cc} \beta B^T B & 0 \\ -B & \frac{1}{\beta} I \end{array}\right). \tag{1.13}$$

Recall the essential variable of the ADMM scheme (1.5) is (y, λ). Moreover, using the

definition of $\tilde{w}^k$, the λ^{k+1} updated by (1.5) can be represented as

$$\begin{aligned}
\lambda^{k+1} &= \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b) \\
&= \lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k - b) \\
&= \lambda^k - \big[-\beta B(y^k - \tilde{y}^k) + \beta(A\tilde{x}^k + By^k - b)\big] \\
&= \lambda^k - \big[-\beta B(y^k - \tilde{y}^k) + (\lambda^k - \tilde{\lambda}^k)\big].
\end{aligned}$$

Therefore, the output of ADMM scheme (1.5) can be written as the correction

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}. \tag{1.14a}$$

which corresponds to the step (1.8b) in the framework with

$$M = \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \quad \text{and} \quad \alpha = 1. \tag{1.14b}$$

Now we check that the Convergence Condition is satisfied by the ADMM scheme (1.5).

Indeed, for the matrix M in (1.14b), we have

$$M^{-1} = \begin{pmatrix} I & 0 \\ \beta B & I \end{pmatrix}.$$

Thus, by using (1.13) and (1.14b), we obtain $H = QM^{-1}$ and thus

$$H = \begin{pmatrix} \beta B^T B & 0 \\ -B & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} I & 0 \\ \beta B & I \end{pmatrix} = \begin{pmatrix} \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}. \tag{1.15}$$

and consequently (since $\alpha = 1$)

$$\begin{aligned}
G &= Q^T + Q - \alpha M^T H M = Q^T + Q - Q^T M \\
&= \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} \beta B^T B & -B^T \\ 0 & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \\
&= \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} 2\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}. \qquad (1.16)
\end{aligned}$$

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 (1.5) is guaranteed.

1.3 Interpreting the Linearized ADMM in the framework

For the linearized version (1.6), using the same definition about $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ as in (1.10), the predictor is given by

$$\left\{\begin{array}{l} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \tilde{y}^k = \text{Argmin}\{\mathcal{L}_\beta(\tilde{x}^k, y, \lambda^k) + \frac{1}{2}\|y - y^k\|_D^2 \,|\, y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). \end{array}\right. \tag{1.17}$$

Because $D = sI - \beta B^T B$, in comparing with (1.12), the defined artificial vector $\tilde{w}^k$ satisfies the following VI:

$$\left\{\begin{array}{ll} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T(-A^T\tilde{\lambda}^k) \ge 0, & \forall\, x \in \mathcal{X}, \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T(-B^T\tilde{\lambda}^k + s(\tilde{y}^k - y^k)) \ge 0, & \forall\, y \in \mathcal{Y}, \\ (A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. & \end{array}\right. \tag{1.18}$$

This can be written in form of (1.8a) as described in the following lemma.

Lemma 1.2 *For given v^k, let w^{k+1} be generated by* (1.6) *and $\tilde{w}^k$ be defined by* (1.10)*. Then, we have $\tilde{w}^k \in \Omega$ and*

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

where

$$Q = \begin{pmatrix} sI_m & 0 \\ -B & \frac{1}{\beta}I \end{pmatrix}. \tag{1.19}$$

In this way, the output of the linearized ADMM scheme (1.6) can be written as (1.14) with the same M and $\alpha = 1$. By using (1.19) and (1.14b), we obtain $H = QM^{-1}$ and thus

$$H = \begin{pmatrix} sI_m & 0 \\ -B & \frac{1}{\beta}I \end{pmatrix} \begin{pmatrix} I & 0 \\ \beta B & I \end{pmatrix} = \begin{pmatrix} sI_m & 0 \\ 0 & \frac{1}{\beta}I \end{pmatrix}. \tag{1.20}$$

and consequently (since $\alpha = 1$)

$$\begin{aligned} G &= Q^T + Q - M^T H M = Q^T + Q - Q^T M \\ &= \begin{pmatrix} 2sI_m & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} sI_m & -B^T \\ 0 & \frac{1}{\beta} I \end{pmatrix} \begin{pmatrix} I & 0 \\ -\beta B & I \end{pmatrix} \\ &= \begin{pmatrix} 2sI_m & -B^T \\ -B & \frac{2}{\beta} I \end{pmatrix} - \begin{pmatrix} sI_m + \beta B^T B & -B^T \\ -B & \frac{1}{\beta} I \end{pmatrix} \\ &= \begin{pmatrix} sI_m - \beta B^T B & 0 \\ 0 & \frac{1}{\beta} I \end{pmatrix}. \end{aligned} \tag{1.21}$$

Since $sI_m - \beta B^T B \succ 0$, both the matrices H (in (1.20)) and G (in (1.21)) are symmetric and positive definite. The Convergence Condition is satisfied; and thus the convergence of the ADMM scheme (1.5) is guaranteed.

2 Convergence proof in the unified framework

In this section, assuming the conditions (1.9) in the unified framework are satisfied, we prove some convergence properties.

Theorem 2.1 *Let $\{v^k\}$ be the sequence generated by a method for the problem* (1.4) *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*

$$\begin{aligned} &\alpha\big(\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k)\big) \\ &\quad \geq \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \frac{\alpha}{2}\|v^k - \tilde{v}^k\|_G^2, \quad \forall w \in \Omega. \end{aligned} \tag{2.1}$$

Proof. Using $Q = HM$ (see (1.9a)) and the relation (1.8b), the right hand side of (1.9a) can be written as $(v - \tilde{v}^k)^T \frac{1}{\alpha} H(v^k - v^{k+1})$ and hence

$$\alpha\{\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k)\} \geq (v - \tilde{v}^k)^T H(v^k - v^{k+1}), \ \forall w \in \Omega. \tag{2.2}$$

Applying the identity

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

to the right hand side of (2.2) with

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

we thus obtain

$$\begin{aligned}&(v - \tilde{v}^k)^T H(v^k - v^{k+1}) \\ &\quad = \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \frac{1}{2}\big(\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2\big). \end{aligned} \tag{2.3}$$

For the last term of (2.3), we have

$$\begin{aligned}&\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\ &\quad = \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\ &\quad \overset{(1.9a)}{=} \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - \alpha M(v^k - \tilde{v}^k)\|_H^2 \\ &\quad = \quad 2\alpha(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - \alpha^2(v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\ &\quad = \quad \alpha(v^k - \tilde{v}^k)^T(Q^T + Q - \alpha M^T HM)(v^k - \tilde{v}^k) \\ &\quad \overset{(1.9b)}{=} \quad \alpha\|v^k - \tilde{v}^k\|_G^2. \end{aligned} \tag{2.4}$$

Substituting (2.3), (2.4) in (2.2), the assertion of this theorem is proved. □

2.1 Convergence in a strictly contraction sense

Theorem 2.2 *Let $\{v^k\}$ be the sequence generated by a method for the problem* (1.4) *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*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \alpha\|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.5}$$

Proof. Setting $w = w^*$ in (2.1), we get

$$\begin{aligned}&\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\ &\quad \ge \quad \alpha\|v^k - \tilde{v}^k\|_G^2 + 2\alpha\{\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k)\}. \end{aligned} \tag{2.6}$$

By using the optimality of w^* and the monotonicity of $F(w)$, we have

$$\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k) \ge \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*) \ge 0$$

and thus

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \ge \alpha\|v^k - \tilde{v}^k\|_G^2. \tag{2.7}$$

The assertion (2.5) follows directly. □

2.2 Convergence rate in an ergodic sense

Equivalent Characterization of the Solution Set of VI

For the convergence rate analysis, we need another characterization of the solution set of VI (1.4). It can be described the following theorem and the proof can be found in [**?**] (Theorem 2.3.5) or [10] (Theorem 2.1). We include all the details for completeness.

Theorem 2.3 *The solution set of VI*(Ω, F, θ) *is convex and it can be characterized as*

$$\Omega^* = \bigcap_{w \in \Omega} \{\tilde{w} \in \Omega : \big(\theta(u) - \theta(\tilde{u})\big) + (w - \tilde{w})^T F(w) \geq 0\}. \tag{2.8}$$

Proof. Indeed, if $\tilde{w} \in \Omega^*$, we have

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

By using the monotonicity of F on Ω, this implies that

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

Thus, $\tilde{w}$ belongs to the right-hand set in (2.8). Conversely, suppose $\tilde{w}$ belongs to the latter

set of (2.8). Let $w \in \Omega$ be arbitrary. The vector

$$\bar{w} = \alpha\tilde{w} + (1 - \alpha)w$$

belongs to Ω for all $\alpha \in (0, 1)$. Thus we have

$$\theta(\bar{u}) - \theta(\tilde{u}) + (\bar{w} - \tilde{w})^T F(\bar{w}) \geq 0. \tag{2.9}$$

Because $\theta(\cdot)$ is convex, we have

$$\theta(\bar{u}) \leq \alpha\theta(\tilde{u}) + (1 - \alpha)\theta(u) \quad \Rightarrow \quad (1 - \alpha)(\theta(u) - \theta(\tilde{u})) \geq \theta(u) - \theta(\tilde{u}).$$

Substituting it in (2.9) and using $\bar{w} - \tilde{w} = (1 - \alpha)(w - \tilde{w})$, we get

$$(\theta(u) - \theta(\tilde{u})) + (w - \tilde{w})^T F(\alpha\tilde{w} + (1 - \alpha)w) \geq 0$$

for all $\alpha \in (0, 1)$. Letting $\alpha \to 1$, it yields

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

Thus $\tilde{w} \in \Omega^*$. Now, we turn to prove the convexity of Ω^*. For each fixed but arbitrary $w \in \Omega$, the set

$$\{\tilde{w} \in \Omega : \theta(\tilde{u}) + \tilde{w}^T F(w) \leq \theta(u) + w^T F(w)\}$$

and its equivalent expression

$$\{\tilde{w} \in \Omega : \ \big(\theta(u) - \theta(\tilde{u})\big) + (w - \tilde{w})^T F(w) \geq 0\}$$

is convex. Since the intersection of any number of convex sets is convex, it follows that the solution set of VI(Ω, F, θ) is convex. $\square$

In Theorem 2.3, we have proved the equivalence of

$$\tilde{w} \in \Omega, \ \ \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(\tilde{w}) \geq 0, \ \ \forall\, w \in \Omega,$$

and

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

We use the late one to define the approximate solution of VI (1.4). Namely, for given $\epsilon > 0$, $\tilde{w} \in \Omega$ is called an ϵ-approximate solution of VI(Ω, F, θ), if it satisfies

$$\tilde{w} \in \Omega, \ \ \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \geq -\epsilon, \ \ \forall\, w \in \mathcal{D}_{(\tilde{w})},$$

where

$$\mathcal{D}_{(\tilde{w})} = \{w \in \Omega \,|\, \|w - \tilde{w}\| \leq 1\}.$$

We need to show that for given $\epsilon > 0$, after t iterations, it can offer a $\tilde{w} \in \mathcal{W}$, such that

$$\tilde{w} \in \mathcal{W} \quad \text{and} \quad \sup_{w \in \mathcal{D}_{(\tilde{w})}} \big\{\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)\big\} \leq \epsilon. \tag{2.10}$$

Theorem 2.1 is also the base for the convergence rate proof. Using the monotonicity of F, we have

$$(w - \tilde{w}^k)^T F(w) \geq (w - \tilde{w}^k)^T F(\tilde{w}^k).$$

Substituting it in (2.1), we obtain

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2\alpha}\|v - v^k\|_H^2 \geq \frac{1}{2\alpha}\|v - v^{k+1}\|_H^2, \ \forall w \in \Omega. \tag{2.11}$$

Note that the above assertion is hold for $G \succeq 0$.

Theorem 2.4 *Let $\{v^k\}$ be the sequence generated by a method for the problem* (1.4) *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*

$$\tilde{w}_t = \frac{1}{t+1}\sum_{k=0}^{t} \tilde{w}^k. \tag{2.12}$$

Then, for any integer number $t > 0$, $\tilde{w}_t \in \Omega$ *and*

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2\alpha(t+1)} \|v - v^0\|_H^2, \quad \forall w \in \Omega. \quad (2.13)$$

Proof. First, it holds that $\tilde{w}^k \in \Omega$ for all $k \geq 0$. Together with the convexity of $\mathcal{X}$ and $\mathcal{Y}$, (2.12) implies that $\tilde{w}_t \in \Omega$. Summing the inequality (2.11) over $k = 0, 1, \ldots, t$, we obtain

$$(t+1)\theta(u) - \sum_{k=0}^{t} \theta(\tilde{u}^k) + \Big((t+1)w - \sum_{k=0}^{t} \tilde{w}^k\Big)^T F(w) + \frac{1}{2\alpha}\|v - v^0\|_H^2 \geq 0, \quad \forall w \in \Omega.$$

Use the notation of $\tilde{w}_t$, it can be written as

$$\frac{1}{t+1}\sum_{k=0}^{t} \theta(\tilde{u}^k) - \theta(u) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2\alpha(t+1)} \|v - v^0\|_H^2, \quad \forall w \in \Omega. \quad (2.14)$$

Since $\theta(u)$ is convex and

$$\tilde{u}_t = \frac{1}{t+1}\sum_{k=0}^{t} \tilde{u}^k,$$

we have that

$$\theta(\tilde{u}_t) \leq \frac{1}{t+1}\sum_{k=0}^{t} \theta(\tilde{u}^k).$$

Substituting it in (2.14), the assertion of this theorem follows directly. □

Recall (2.10). The conclusion (2.13) 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.

- **For the unified framework and the convergence proof, the reader can consult: 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.**
- **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.**

2.3 Convergence rate in pointwise iteration-complexity

In this subsection, we show that if the matrix G defined in (1.9b) is positive definite, a worst-case $O(1/t)$ convergence rate in a nonergodic sense can also be established for the prototype algorithm (1.8). Note in general a nonergodic convergence rate is stronger than the ergodic convergence rate.

We first need to prove the following lemma.

Lemma 2.1 *For the sequence generated by the prototype algorithm (1.8) where the Convergence Condition is satisfied, we have*

$$\begin{aligned}&(v^k-\tilde{v}^k)^T M^T H M\{(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\}\\ \geq\ & \frac{1}{2\alpha}\|(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\|^2_{(Q^T+Q)}.\end{aligned} \tag{2.15}$$

Proof. First, set $w=\tilde{w}^{k+1}$ in (1.8a), we have

$$\theta(\tilde{u}^{k+1})-\theta(\tilde{u}^k)+(\tilde{w}^{k+1}-\tilde{w}^k)^T F(\tilde{w}^k)\geq(\tilde{v}^{k+1}-\tilde{v}^k)^T Q(v^k-\tilde{v}^k). \tag{2.16}$$

Note that (1.8a) is also true for $k:=k+1$ and thus we have

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

Set $w=\tilde{w}^k$ in the above inequality, we obtain

$$\theta(\tilde{u}^k)-\theta(\tilde{u}^{k+1})+(\tilde{w}^k-\tilde{w}^{k+1})^T F(\tilde{w}^{k+1})\geq(\tilde{v}^k-\tilde{v}^{k+1})^T Q(v^{k+1}-\tilde{v}^{k+1}). \tag{2.17}$$

Combining (2.16) and (2.17) and using the monotonicity of F, we get

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

Adding the term

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

to the both sides of (2.18), and using $v^T Qv=\frac{1}{2}v^T(Q^T+Q)v$, we obtain

$$(v^k-v^{k+1})^T Q\{(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\}\geq\frac{1}{2}\|(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\|^2_{(Q^T+Q)}.$$

Substituting $(v^k-v^{k+1})=\alpha M(v^k-\tilde{v}^k)$ in the left-hand side of the last inequality and using $Q=HM$, we obtain (2.15) and the lemma is proved. □

Now, we are ready to prove (2.19), the key inequality in this section.

Theorem 2.5 *For the sequence generated by the prototype algorithm (1.8) where the Convergence Condition is satisfied, we have*

$$\|M(v^{k+1} - \tilde{v}^{k+1})\|_H \le \|M(v^k - \tilde{v}^k)\|_H, \quad \forall\, k > 0. \tag{2.19}$$

Proof. Setting $a = M(v^k - \tilde{v}^k)$ and $b = M(v^{k+1} - \tilde{v}^{k+1})$ in the identity

$$\|a\|_H^2 - \|b\|_H^2 = 2a^T H(a-b) - \|a-b\|_H^2,$$

we obtain

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

Inserting (2.15) into the first term of the right-hand side of the last equality, we obtain

$$\begin{aligned} &\|M(v^k - \tilde{v}^k)\|_H^2 - \|M(v^{k+1} - \tilde{v}^{k+1})\|_H^2 \\ \ge \quad & \frac{1}{\alpha}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T+Q)}^2 - \|M[(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})]\|_H^2 \\ = \quad & \frac{1}{\alpha}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_G^2 \ge 0, \end{aligned}$$

where the last inequality is because of the positive definiteness of the matrix $(Q^T + Q) - \alpha M^T H M \succeq 0$. The assertion (2.19) follows immediately. □

Note that it follows from $G \succ 0$ and Theorem 2.2 there is a constant $c_0 > 0$ such that

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - c_0\|M(v^k - \tilde{v}^k)\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.20}$$

Now, with (2.20) and (2.19), we can establish the worst-case $O(1/t)$ convergence rate in a nonergodic sense for the prototype algorithm (1.8).

Theorem 2.6 *Let $\{v^k\}$ and $\{\tilde{w}^k\}$ be the sequences generated by the prototype algorithm (1.8) under the Convergence Condition. For any integer $t > 0$, we have*

$$\|M(v^t - \tilde{v}^t)\|_H^2 \le \frac{1}{(t+1)c_0}\|v^0 - v^*\|_H^2. \tag{2.21}$$

Proof. First, it follows from (2.20) that

$$\sum_{k=0}^{\infty} c_0 \|M(v^k - \tilde{v}^k)\|_H^2 \le \|v^0 - v^*\|_H^2, \quad \forall\, v^* \in \mathcal{V}^*. \tag{2.22}$$

According to Theorem 2.5, the sequence $\{\|M(v^k - \tilde{v}^k)\|_H^2\}$ is monotonically non-increasing. Therefore, we have

$$(t+1)\|M(v^t - \tilde{v}^t)\|_H^2 \le \sum_{k=0}^{t} \|M(v^k - \tilde{v}^k)\|_H^2. \tag{2.23}$$

The assertion (2.21) follows from (2.22) and (2.23) immediately. □

Let $d := \inf\{\|v^0 - v^*\|_H \,|\, v^* \in \mathcal{V}^*\}$. Then, for any given $\epsilon > 0$, Theorem 2.6 shows that the it needs at most $\lfloor d^2/c_0\epsilon \rfloor$ iterations to ensure that $\|M(v^k - \tilde{v}^k)\|_H^2 \le \epsilon$. Recall that v^k is a solution of VI(Ω, F, θ) if $\|M(v^k - \tilde{v}^k)\|_H^2 = 0$ (see (1.8a) and due to $Q = HM$). A worst-case $O(1/t)$ convergence rate in pointwise iteration-complexity is thus established for the prototype algorithm (1.8).

Notice that, for a differentiable unconstrained convex optimization $\min f(x)$, it holds that

$$f(x) - f(x^*) = \nabla f(x^*)^T (x - x^*) + O(\|x - x^*\|^2) = O(\|x - x^*\|^2).$$

3 Classical ADMM in the Unified Framework

This section interprets the classical ADMM (1.5) and the linearized lADMM (1.6).

ADMM (1.5) Note that Theorem 2.4 is true for $G \succeq 0$. Thus the classical ADMM (1.5) has $O(1/t)$ convergence rate in the ergodic sense.

Since $\alpha = 1$, according to (2.5) and the form of G in (1.16), we have

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \tfrac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2, \quad \forall v^* \in \mathcal{V}^*. \tag{3.1}$$

Lemma 3.1 *For given v^k, let w^{k+1} be generated by* (1.5) *and $\tilde{w}^k$ be defined by* (1.10)*. Then, we have*

$$\tfrac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2 \ge \|v^k - v^{k+1}\|_H^2. \tag{3.2}$$

Proof. According to (1.5) and (1.10), the optimal condition of the y-subproblem is

$$\tilde{y}^k \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\{-B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k)\} \ge 0, \; \forall\, y \in \mathcal{Y}.$$

Because

$$\lambda^{k+1} = \tilde{\lambda}^k - \beta B(\tilde{y}^k - y^k) \quad \text{and} \quad \tilde{y}^k = y^{k+1},$$

it can be written as

$$y^{k+1} \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^T\{-B^T\lambda^{k+1}\} \geq 0, \; \forall\, y \in \mathcal{Y}. \quad (3.3)$$

The above inequality is hold also for the last iteration, *i. e.*, we have

$$y^k \in \mathcal{Y}, \;\; \theta_2(y) - \theta_2(y^k) + (y - y^k)^T\{-B^T\lambda^k\} \geq 0, \; \forall\, y \in \mathcal{Y}. \quad (3.4)$$

Setting $y = y^k$ in (3.3) and $y = y^{k+1}$ in (3.4), and then adding them, we get

$$(\lambda^k - \lambda^{k+1})^T B(y^k - y^{k+1}) \geq 0. \quad (3.5)$$

Using $\lambda^k - \tilde{\lambda}^k = (\lambda^k - \lambda^{k+1}) + \beta B(y^k - y^{k+1})$ and the inequality (3.5), we obtain

$$\begin{aligned} \tfrac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2 &= \tfrac{1}{\beta}\|(\lambda^k - \lambda^{k+1}) + \beta B(y^k - y^{k+1})\|^2 \\ &\geq \tfrac{1}{\beta}\|\lambda^k - \lambda^{k+1}\|^2 + \beta\|B(y^k - y^{k+1})\|^2 \\ &= \|v^k - v^{k+1}\|_H^2. \end{aligned}$$

The assertion of this lemma is proved. □

Substituting (3.2) in (3.1), we get the following nice property of the classical ADMM as in Lecture 11.

Theorem 3.1 *For given v^k, let w^{k+1} be generated by* (1.5)*. Then, we have*

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \;\; \forall v^* \in \mathcal{V}^*. \quad (3.6)$$

Notice that the sequence $\{\|v^k - v^{k+1}\|_H^2\}$ generated by the classical ADMM is monotone non-increasing. In fact, in Theorem 2.5, we have proved that (see (2.19))

$$\|M(v^k - \tilde{v}^k)\|_H \leq \|M(v^{k-1} - \tilde{v}^{k-1})\|_H, \quad \forall\, k \geq 1. \quad (3.7)$$

Because (see the correction formula (1.14)), $v^k - v^{k+1} = M(v^k - \tilde{v}^k)$, it follows from (3.7) that

$$\|v^k - v^{k+1}\|_H^2 \leq \|v^{k-1} - v^k\|_H^2. \quad (3.8)$$

On the other hand, the inequality (3.6) tell us that

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

Thus, we have

$$\begin{aligned}\|v^t - v^{t+1}\|_H^2 &\le \frac{1}{t+1}\sum_{k=0}^{t}\|v^k - v^{k+1}\|_H^2 \\ &\le \frac{1}{t+1}\sum_{k=0}^{\infty}\|v^k - v^{k+1}\|_H^2 \\ &\le \frac{1}{t+1}\|v^0 - v^*\|_H^2.\end{aligned}$$

Therefore, ADMM (1.5) has $O(1/t)$ convergence rate in pointwise iteration-complexity [12].

ADMM (1.6) According to Theorem 2.2, for the v^{k+1} generated by (1.6) and $\tilde{v}^k$ defined by (1.10), we have

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*, \tag{3.9}$$

where

$$H = \begin{pmatrix} sI_m & 0 \\ 0 & \frac{1}{\beta}I \end{pmatrix} \quad \text{and} \quad G = \begin{pmatrix} sI_m - \beta B^T B & 0 \\ 0 & \frac{1}{\beta}I \end{pmatrix}.$$

Since

$$v^k - \tilde{v}^k = M^{-1}(v^k - v^{k+1}),$$

it follows from (3.9) that

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|M^{-1}(v^k - v^{k+1})\|_G^2, \quad \forall v^* \in \mathcal{V}^*, \tag{3.10}$$

Using $M = \begin{bmatrix} I & 0 \\ -\beta B & I \end{bmatrix}$ (see (1.14b)) and defining $S = M^{-T}GM^{-1}$, we have

$$\begin{aligned} S &= \begin{bmatrix} I & \beta B \\ 0 & I \end{bmatrix}\begin{bmatrix} sI_m - \beta B^T B & 0 \\ 0 & \frac{1}{\beta}I \end{bmatrix}\begin{bmatrix} I & 0 \\ \beta B & I \end{bmatrix} \\ &= \begin{bmatrix} sI_m & B^T \\ B & \frac{1}{\beta}I \end{bmatrix}. \end{aligned}$$

Since $sI_m \succ \beta\rho(B^T B)$, the matrix S is positive definite and it follows from (3.10) that

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

4 ADMM in Sense of Customized PPA [3]

If we change the performance order of y and λ of the classical ADMM (1.5), it becomes

$$\left\{\begin{array}{l} x^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^k - b), \\ y^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+1}) \,|\, y \in \mathcal{Y}\}. \end{array}\right. \tag{4.1}$$

In this way we can get a positive semidefinite matrix Q in (1.8a). We define

$$\tilde{x}^k = x^{k+1}, \quad \tilde{y}^k = y^{k+1}, \quad \tilde{\lambda}^k = \lambda^{k+1}, \tag{4.2}$$

where $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ is the output of (4.1) and thus it can be rewritten as

$$\left\{\begin{array}{l} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), \\ \tilde{y}^k = \text{Argmin}\{\mathcal{L}_\beta(\tilde{x}^k, y, \tilde{\lambda}^k) \,|\, y \in \mathcal{Y}\}. \end{array}\right. \tag{4.3}$$

Because $\tilde{\lambda}^k = \lambda^{k+1} = \lambda^k - \beta(A\tilde{x}^k + By^k - b)$, the optimal condition of the x-subproblem of (4.3) is

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T(-A^T\tilde{\lambda}^k) \ge 0, \quad \forall\, x \in \mathcal{X}. \tag{4.4}$$

Notice that

$$\mathcal{L}_\beta(\tilde{x}^k, y, \tilde{\lambda}^k) = \theta_1(\tilde{x}^k) + \theta_2(y) - (\tilde{\lambda}^k)^T(A\tilde{x}^k + By - b) + \tfrac{\beta}{2}\|A\tilde{x}^k + By - b\|^2,$$

ignoring the constant term in the y optimization subproblem of (4.3), it turns to

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) - (\tilde{\lambda}^k)^T By + \tfrac{\beta}{2}\|A\tilde{x}^k + By - b\|^2 \,|\, y \in \mathcal{Y}\},$$

and consequently, the optimal condition is $\tilde{y}^k \in \mathcal{Y}$,

$$\theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T\big[-B^T\tilde{\lambda}^k + \beta B^T(A\tilde{x}^k + B\tilde{y}^k - b)\big] \ge 0, \ \forall y \in \mathcal{Y}.$$

For the term $[\,\cdot\,]$ in the last inequality, using $\beta(A\tilde{x}^k + By^k - b) = -(\tilde{\lambda}^k - \lambda^k)$, we have

$$\begin{array}{rl} & -B^T\tilde{\lambda}^k + \beta B^T(A\tilde{x}^k + B\tilde{y}^k - b) \\ = & -B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) + \beta B^T(A\tilde{x}^k + By^k - b) \\ = & -B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) - B^T(\tilde{\lambda}^k - \lambda^k). \end{array}$$

Finally, the optimal condition of the y-subproblem can be written as $\tilde{y}^k \in \mathcal{Y}$ and

$$\theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T\left[-B^T\tilde{\lambda}^k+\beta B^TB(\tilde{y}^k-y^k)-B^T(\tilde{\lambda}^k-\lambda^k)\right] \geq 0, \ \forall y \in \mathcal{Y}. \tag{4.5}$$

From the λ update form in (4.3) we have

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \tag{4.6}$$

Combining (4.4), (4.5) and (4.6), and using the notations of (1.4), we get following lemma.

Lemma 4.1 *For given v^k, let $\tilde{w}^k$ be generated by* (4.3)*. Then, we have*

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

where

$$Q = \begin{pmatrix} \beta B^TB & -B^T \\ -B & \frac{1}{\beta}I_m \end{pmatrix}. \tag{4.7}$$

Because Q is symmetric and positive semidefinite, according to (1.8), we can take

$$M = I \qquad \alpha \in (0,2) \qquad \text{and thus} \qquad H = Q.$$

In this way, we get the new iterate by

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

The generated sequence $\{v^k\}$ has the convergence property

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

Ensure the matrix H to be positive definite If we add an additional proximal term $\frac{\delta\beta}{2}\|B(y-y^k)\|^2$ to the y-subproblem of (4.3) with any small $\delta > 0$, it becomes

$$\begin{cases} \tilde{x}^k = \text{Argmin}\{\mathcal{L}_\beta^{(2)}(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), \\ \tilde{y}^k = \text{Argmin}\{\mathcal{L}_\beta^{(2)}(\tilde{x}^k, y, \tilde{\lambda}^k) + \frac{\delta\beta}{2}\|B(y-y^k)\|^2 \,|\, y \in \mathcal{Y}\}. \end{cases} \tag{4.8}$$

In the ADMM based customized PPA (4.3), the y-subproblem can be written as

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \tfrac{\beta}{2}\|By - p^k\|^2 \,|\, y \in \mathcal{Y}\}, \tag{4.9}$$

where

$$p^k = b + \tfrac{1}{\beta}\tilde{\lambda}^k - A\tilde{x}^k.$$

If we add an additional term $\frac{\delta\beta}{2}\|B(y-y^k)\|^2$ (with any small $\delta > 0$) to the objective function of the y-subproblem, we will get $\tilde{y}^k$ via

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \tfrac{\beta}{2}\|By - p^k\|^2 + \tfrac{\delta\beta}{2}\|B(y-y^k)\|^2 \,|\, y \in \mathcal{Y}\}.$$

By a manipulation, the solution point of the above subproblem is obtained via

$$\tilde{y}^k = \text{Argmin}\{\theta_2(y) + \tfrac{(1+\delta)\beta}{2}\|By - q^k\|^2 \,|\, y \in \mathcal{Y}\}, \tag{4.10}$$

where

$$q^k = \tfrac{1}{1+\delta}(p^k + \delta By^k).$$

In this way, the matrix Q in (4.7) will turn to

$$Q = \begin{pmatrix} (1+\delta)\beta B^T B & -B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}.$$

Take $H = Q$, for any $\delta > 0$, H is positive definite when B is a full rank matrix. In other words, instead of (4.9), using (4.10) to get $\tilde{y}^k$, it will ensure the positivity of H theoretically. However, in practical computation, it works still well by using $\delta = 0$.

5 Symmetric ADMM [7]

In the problem (1.1), x and y are a pair of fair variables. It is nature to consider a symmetric method: Update the Lagrangian Multiplier after solving each x and y-subproblem. .

We take $\mu \in (0,1)$ (usually $\mu = 0.9$), the method is described as

$$\begin{cases} x^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \mu\beta(Ax^{k+1} + By^k - b), \\ y^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \,|\, y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \mu\beta(Ax^{k+1} + By^{k+1} - b). \end{cases} \tag{5.1}$$

This method is called **Alternating direction method of multipliers with symmetric multipliers updating**, or **Symmetric Alternating Direction Method of Multipliers**.

The convergence of the proposed method is established via the unified framework.

For establishing the main result, we introduce an artificial vector $\tilde{w}^k$ by

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k - \beta(Ax^{k+1} + By^k - b) \end{pmatrix}, \tag{5.2}$$

where (x^{k+1}, y^{k+1}) is generated by the ADMM (5.1).

According to (5.2), the optimal condition of the x-subproblem of (5.1) is

$$\theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T(-A^T\tilde{\lambda}^k) \geq 0, \quad \forall\, x \in \mathcal{X}. \tag{5.3}$$

Notice that the objective function of the y-subproblem in (5.1) is

$$\begin{aligned} &\mathcal{L}_\beta(\tilde{x}^k, y, \lambda^{k+\frac{1}{2}}) \\ &\quad = \quad \theta_1(\tilde{x}^k) + \theta_2(y) - (\lambda^{k+\frac{1}{2}})^T(A\tilde{x}^k + By - b) + \tfrac{\beta}{2}\|A\tilde{x}^k + By - b\|^2. \end{aligned}$$

Ignoring the constant term in the y-subproblem, it turns to

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

Consequently, according to the optimality equivalent lemma , we have

$$\begin{aligned} &\tilde{y}^k \in \mathcal{Y}, \quad \theta_2(y) - \theta_2(\tilde{y}^k) \\ &\qquad +(y - \tilde{y}^k)^T\big\{-B^T\lambda^{k+\frac{1}{2}} + \beta B^T(A\tilde{x}^k + B\tilde{y}^k - b)\big\} \geq 0, \ \forall\, y \in \mathcal{Y}. \end{aligned}$$

Using

$$\lambda^{k+\frac{1}{2}} = \lambda^k - \mu(\lambda^k - \tilde{\lambda}^k) = \tilde{\lambda}^k + (1-\mu)(\lambda^k - \tilde{\lambda}^k),$$

and $\beta(A\tilde{x}^k + By^k - b) = (\tilde{\lambda}^k - \lambda^k)$, we get

$$\begin{aligned} &-B^T\lambda^{k+\frac{1}{2}} + \beta B^T(A\tilde{x}^k + B\tilde{y}^k - b) \\ &\quad = \quad -B^T(\tilde{\lambda}^k + (1-\mu)(\lambda^k - \tilde{\lambda}^k)) + \beta B^T(A\tilde{x}^k + B\tilde{y}^k - b) \\ &\quad = \quad -B^T(\tilde{\lambda}^k + (1-\mu)(\lambda^k - \tilde{\lambda}^k)) + \beta B^T B(\tilde{y}^k - y^k) \\ &\qquad\quad +\beta B^T(A\tilde{x}^k + By^k - b) \\ &\quad = \quad -B^T\tilde{\lambda}^k - (1-\mu)B^T(\lambda^k - \tilde{\lambda}^k) + \beta B^T B(\tilde{y}^k - y^k) \\ &\qquad\quad +B^T(\lambda^k - \tilde{\lambda}^k) \\ &\quad = \quad -B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) - \mu B^T(\tilde{\lambda}^k - \lambda^k). \end{aligned}$$

Finally, the optimal condition of the y-subproblem can be written as $\tilde{y}^k \in \mathcal{Y}$, and

$$\theta_2(y) - \theta_2(\tilde{y}^k) \; + \; (y - \tilde{y}^k)^T \big\{ -B^T \tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k) \\ -\mu B^T (\tilde{\lambda}^k - \lambda^k) \big\} \geq 0, \; \forall y \in \mathcal{Y}. \tag{5.4}$$

According to the definition of $\tilde{w}^k$ in (5.2), we have

$$(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + (1/\beta)(\tilde{\lambda}^k - \lambda^k) = 0. \tag{5.5}$$

Combining (5.3), (5.4) and (5.5), and using the notations of (1.4), we get following lemma.

Lemma 5.1 *For given v^k, let w^{k+1} be generated by* (5.1) *and $\tilde{w}^k$ be defined by* (5.2)*. Then, we have*

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

where

$$Q = \begin{pmatrix} \beta B^T B & -\mu B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix}. \tag{5.6}$$

Using this notation and by a manipulation, the update form of λ in (5.1) can be represented

as

$$\begin{aligned} \lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - \mu\big[-\beta B(y^k - \tilde{y}^k) + \beta(Ax^{k+1} + By^k - b)\big] \\ &= \lambda^k - \big[-\mu\beta B(y^k - \tilde{y}^k) + 2\mu(\lambda^k - \tilde{\lambda}^k)\big]. \end{aligned} \tag{5.7}$$

Thus, together with $y^{k+1} = \tilde{y}^k$, we have the following useful relationship

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I & 0 \\ -\mu\beta B & 2\mu I_m \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}.$$

This can be rewritten into a compact form:

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k),$$

with

$$M = \begin{pmatrix} I & 0 \\ -\mu\beta B & 2\mu I_m \end{pmatrix}. \tag{5.8}$$

These relationships greatly simplify our analysis and presentation.

In order to use the unified framework, we only need to verify the positiveness of H and G.
For the matrix M given by (5.8), we have

$$M^{-1} = \begin{pmatrix} I & 0 \\ \frac{1}{2}\beta B & \frac{1}{2\mu} I_m \end{pmatrix}.$$

For $H = QM^{-1}$, it follows that

$$H = \begin{pmatrix} \beta B^T B & -\mu B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} I & 0 \\ \frac{1}{2}\beta B & \frac{1}{2\mu} I_m \end{pmatrix} = \begin{pmatrix} (1-\frac{1}{2}\mu)\beta B^T B & -\frac{1}{2} B^T \\ -\frac{1}{2} B & \frac{1}{2\mu\beta} I_m \end{pmatrix}.$$

Thus

$$H = \frac{1}{2} \begin{pmatrix} \sqrt{\beta} B^T & 0 \\ 0 & \sqrt{\frac{1}{\beta}} I \end{pmatrix} \begin{pmatrix} (2-\mu) I & -I \\ -I & \frac{1}{\mu} I \end{pmatrix} \begin{pmatrix} \sqrt{\beta} B & 0 \\ 0 & \sqrt{\frac{1}{\beta}} I \end{pmatrix}.$$

Notice that

$$\begin{pmatrix} (2-\mu) & -1 \\ -1 & \frac{1}{\mu} \end{pmatrix} = \begin{cases} \succ 0, & \mu \in (0,1); \\ \succeq 0, & \mu = 1. \end{cases}$$

Therefore, H is positive definite for any $\mu \in (0,1)$ when B is a full column rank matrix.

It remains to check the positiveness of $G = Q^T + Q - M^T H M$. Note that

$$\begin{aligned} M^T H M &= M^T Q = \begin{pmatrix} I & -\mu\beta B^T \\ 0 & 2\mu I_m \end{pmatrix} \begin{pmatrix} \beta B^T B & -\mu B^T \\ -B & \frac{1}{\beta} I_m \end{pmatrix} \\ &= \begin{pmatrix} (1+\mu)\beta B^T B & -2\mu B^T \\ -2\mu B & \frac{2\mu}{\beta} I_m \end{pmatrix}. \end{aligned}$$

Using (5.6) and the above equation, we have

$$G = (Q^T + Q) - M^T H M = (1-\mu) \begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{2}{\beta} I_m \end{pmatrix}.$$

Thus

$$G = (1-\mu) \begin{pmatrix} \sqrt{\beta} B^T & 0 \\ 0 & \sqrt{\frac{1}{\beta}} I \end{pmatrix} \begin{pmatrix} I & -I \\ -I & 2I \end{pmatrix} \begin{pmatrix} \sqrt{\beta} B & 0 \\ 0 & \sqrt{\frac{1}{\beta}} I \end{pmatrix}.$$

Because the matrix

$$\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$$

is positive definite, for any $\mu \in (0,1)$, G is essentially positive definite (positive definite when B is a full column rank matrix). The convergence conditions (1.9) are satisfied.

Take $\mu = 0.9$, it will accelerate the convergence much. For the numerical experiments of this method, it is refereed to consult [7].

The symmetric ADMM is a special version of the unified framework (1.8) - (1.9) whose $\alpha = 1$,

$$H = \begin{pmatrix} (1-\frac{1}{2}\mu)\beta B^T B & -\frac{1}{2}B^T \\ -\frac{1}{2}B & \frac{1}{2\mu\beta}I_m \end{pmatrix} \quad \text{and} \quad G = (1-\mu)\begin{pmatrix} \beta B^T B & -B^T \\ -B & \frac{2}{\beta}I_m \end{pmatrix}.$$

Both the matrices H and G are positive definite for $\mu \in (0,1)$. According to Theorem 2.2, we have

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

Conclusion Remarks

Theorem 2.1 is the base of this note. Using $\alpha M(v^k - \tilde{v}^k) = (v^k - v^{k+1})$ and $Q = HM$, we get

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

The elementary identity

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

plays a key role in the proof. we thus obtain

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

This basic technique has also been adopted by A. Beck [1] in his monographs. He has made a special footnote (see Page 428 in [1]). The technique is simple yet powerful !

References

[1] A. Beck, First-Order Methods in Optimization, MOS-SIAM Seriers on Optimization, 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*, manuscript, 2010.

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

[5] R. Glowinski and P. Le Tallec, Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics, SIAM Studies in Applied Mathematics, Philadelphia, PA, 1989.

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

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

[8] B.S. He, M. Tao and X.M. Yuan, Alternating Direction Method with Gaussian Back Substitution for Separable Convex Programming, SIAM J. Optim. 22, 313-340, 2012

[9] B. S. He, M. H. Xu and X. M. Yuan, Solving large-scale least squares covariance matrix problems by alternating direction methods, SIAM J. Matrix Analy. Appli., 32, (2011), pp. 136-152.

[10] B.S. He and X.M. Yuan, On the $O(1/n)$ Convergence Rate of the Douglas-Rachford Alternating Direction Method, SIAM J. Numer. Anal. 50, 700-709, 2012.

[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. Sci., 5, 119-149, 2012.

[12] B.S. He and X.M. Yuan, On non-ergodic convergence rate of Douglas-Rachford alternating directions method of multipliers, Numer. Math., (2015) 130:567–577.

[13] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appli., 4(1969), pp. 303-320.

[14] S. Kontogiorgis and R. R. Meyer, A variable-penalty alternating directions method for convex optimization, Math. Program., 83 (1998), pp. 29-53.

[15] A. Nemirovski, *Prox-method with rate of convergence $O(1/t)$ for variational inequality with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems*, SIAM J. Optim. 15 (2005), pp. 229-251.

[16] Y. E. Nesterov, *A method for unconstrained convex minimization problem with the rate of convergence O($1/k^2$)*, Doklady AN SSSR, 269 (1983), pp. 543-547.

[17] P. Tseng, *On accelerated proximal gradient methods for convex-concave optimization*, manuscript, University of Washington, USA, 2008.

凸优化和单调变分不等式的收缩算法

第十六讲: 三块可分离凸优化问题的平行分裂增广 Lagrange 乘子法

Parallel splitting augmented Lagrangian methods for convex optimization with three separable blocks

南京大学数学系　何炳生
hebma@nju.edu.cn

The context of this lecture is based on the publication [6]

1 Convex Optimization with 3 separable operators

We consider the linearly constrained convex optimization with 3 separable operators:

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z) \mid Ax+By+Cz=b,\ x\in\mathcal{X},\ y\in\mathcal{Y},\ z\in\mathcal{Z}\} \quad (1.1)$$

where $\theta_1(x):\Re^{n_1}\to\Re,\ \theta_2(y):\Re^{n_2}\to\Re,\ \theta_3(z):\Re^{n_3}\to\Re$ are convex; $A\in\Re^{m\times n_1}, B\in\Re^{m\times n_2}, C\in\Re^{m\times n_3},\ b\in\Re^m; \mathcal{X}\subset\Re^{n_1}, \mathcal{Y}\subset\Re^{n_2}$ and $\mathcal{Z}\subset\Re^{n_3}$ are given convex set. Let $n=n_1+n_2+n_3$.

This optimization problem is equivalent to find $(x^*,y^*,z^*,\lambda^*)\in\Omega$, such that

$$\left\{\begin{array}{l}\theta_1(x)-\theta_1(x^*)+(x-x^*)^T(-A^T\lambda^*)\ge 0,\\ \theta_2(y)-\theta_2(y^*)+(y-y^*)^T(-B^T\lambda^*)\ge 0,\\ \theta_3(z)-\theta_3(z^*)+(z-z^*)^T(-C^T\lambda^*)\ge 0,\\ (\lambda-\lambda^*)^T(Ax^*+By^*+Cz^*-b)\ge 0,\end{array}\right. \quad \forall\ (x,y,z,\lambda)\in\Omega, \quad (1.2)$$

where

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

By denoting

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

and

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

the optimal condition (1.2) can be written as

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

For convenience, in this paper, we use the notations

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

2 Full Parallel Splitting ALM

Full parallel splitting augmented Lagrangian method is a prediction-correction method. First it generates a predictor:

1. From given $(x^k, y^k, z^k, \lambda^k)$, obtain $\tilde{x}^k$, $\tilde{y}^k$ and $\tilde{z}^k$, by the following parallel manner:

$$\tilde{x}^k = \text{Argmin} \left\{ \begin{array}{c} \theta_1(x) - (\lambda^k)^T(Ax + By^k + Cz^k - b) \\ + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 \end{array} \middle| x \in \mathcal{X} \right\} \quad (2.1a)$$

$$\tilde{y}^k = \text{Argmin} \left\{ \begin{array}{c} \theta_2(y) - (\lambda^k)^T(Ax^k + By + Cz^k - b) \\ + \frac{\beta}{2}\|Ax^k + By + Cz^k - b\|^2 \end{array} \middle| y \in \mathcal{Y} \right\} \quad (2.1b)$$

$$\tilde{z}^k = \text{Argmin} \left\{ \begin{array}{c} \theta_3(z) - (\lambda^k)^T(Ax^k + By^k + Cz - b) \\ + \frac{\beta}{2}\|Ax^k + By^k + Cz - b\|^2 \end{array} \middle| z \in \mathcal{Z} \right\} \quad (2.1c)$$

2. Update $\tilde{\lambda}^k$ by

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b). \quad (2.1d)$$

Note that the solution $(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ of (2.1a)-(2.1c) satisfies

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{-A^T\lambda^k + \beta A^T\big(A\tilde{x}^k + By^k + Cz^k - b\big)\} \geq 0 \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{-B^T\lambda^k + \beta B^T\big(Ax^k + B\tilde{y}^k + Cz^k - b\big)\} \geq 0 \\ \theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^T \{-C^T\lambda^k + \beta C^T\big(Ax^k + By^k + C\tilde{z}^k - b\big)\} \geq 0 \end{cases} \tag{2.2}$$

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

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b)$$

and by a manipulation, (2.2) can be rewritten as $(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$,

$$\theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \\ z - \tilde{z}^k \end{pmatrix}^T \begin{pmatrix} -A^T\tilde{\lambda}^k + \beta A^T[B(y^k - \tilde{y}^k) + C(z^k - \tilde{z}^k)] \\ -B^T\tilde{\lambda}^k + \beta B^T[A(x^k - \tilde{x}^k) + C(z^k - \tilde{z}^k)] \\ -C^T\tilde{\lambda}^k + \beta C^T[A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k)] \end{pmatrix} \geq 0, \tag{2.3}$$

for all $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. Combining the last inequality with (2.1d), we

have

$$\begin{aligned} \theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \\ z - \tilde{z}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \Bigg\{ \begin{pmatrix} -A^T\tilde{\lambda}^k \\ -B^T\tilde{\lambda}^k \\ -C^T\tilde{\lambda}^k \\ A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b \end{pmatrix} \\ + \beta \begin{pmatrix} A^T \\ B^T \\ C^T \\ 0 \end{pmatrix} \big(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k) + C(z^k - \tilde{z}^k)\big) \\ + \begin{pmatrix} \beta A^TA & 0 & 0 & 0 \\ 0 & \beta B^TB & 0 & 0 \\ 0 & 0 & \beta C^TC & 0 \\ 0 & 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} \tilde{x}^k - x^k \\ \tilde{y}^k - y^k \\ \tilde{z}^k - z^k \\ \tilde{\lambda}^k - \lambda^k \end{pmatrix} \Bigg\} \geq 0, \ \forall w \in \Omega. \end{aligned} \tag{2.4}$$

Based on the above analysis, we have the following lemma.

Lemma 2.1 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (2.1) *from the given* $w^k = (x^k, y^k, z^k, \lambda^k)$*. Then, we have*

$$(\tilde{w}^k - w^*)^T H(w^k - \tilde{w}^k) \geq (\tilde{w}^k - w^*)^T \big(F(\tilde{w}^k) + \eta(u^k, \tilde{u}^k)\big), \tag{2.5}$$

where

$$\eta(u^k, \tilde{u}^k) = \beta \begin{pmatrix} A^T \\ B^T \\ C^T \\ 0 \end{pmatrix} \big[A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k) + C(z^k - \tilde{z}^k)\big] \tag{2.6}$$

and

$$H = \begin{pmatrix} \beta A^T A & 0 & 0 & 0 \\ 0 & \beta B^T B & 0 & 0 \\ 0 & 0 & \beta C^T C & 0 \\ 0 & 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{2.7}$$

Proof. Setting $(x, y, z, \lambda) = (x^*, y^*, z^*, \lambda^*)$ in (2.4), the assertion follows directly. $\square$

Since F is monotone and $\tilde{w}^k \in \Omega$, it follows that

$$(\tilde{w}^k - w^*)^T F(\tilde{w}^k) \geq (\tilde{w}^k - w^*)^T F(w^*) \geq 0. \tag{2.8}$$

In addition, by using $Ax^* + By^* + Cz^* = b$ and

$$\beta(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b) = (\lambda^k - \tilde{\lambda}^k),$$

we have

$$\begin{aligned} &(\tilde{w}^k - w^*)^T \eta(u^k, \tilde{u}^k) \\ &\quad = \quad (A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k) + C(z^k - \tilde{z}^k))^T \beta(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b) \\ &\quad = \quad (\lambda^k - \tilde{\lambda}^k)^T [A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k) + C(z^k - \tilde{z}^k)]. \end{aligned} \tag{2.9}$$

Lemma 2.2 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (2.1) *from the*

given $w^k = (x^k, y^k, z^k, \lambda^k)$. *Then, we have*

$$(w^k - w^*)^T H(w^k - \tilde{w}^k) \geq \varphi(w^k, \tilde{w}^k), \quad \forall\, w^* \in \Omega^*, \tag{2.10}$$

where

$$\begin{aligned}\varphi(w^k, \tilde{w}^k) &= \|w^k - \tilde{w}^k\|_H^2 \\ &\quad +(\lambda^k - \tilde{\lambda}^k)^T \big(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k) + C(z^k - \tilde{z}^k)\big). \end{aligned} \tag{2.11}$$

In addition, we have

$$\varphi(w^k, \tilde{w}^k) \geq \frac{2-\sqrt{3}}{2}\|w^k - \tilde{w}^k\|_H^2. \tag{2.12}$$

Proof. First, using (2.5), (2.8) and (2.9) we obtain that

$$\begin{aligned}&(\tilde{w}^k - w^*)^T H(w^k - \tilde{w}^k) \\ &\quad \geq \quad (\lambda^k - \tilde{\lambda}^k)^T \big(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k) + C(z^k - \tilde{z}^k)\big).\end{aligned}$$

The first assertion of this lemma follows from the last inequality and the definition of $\varphi(w^k, \tilde{w}^k)$ directly.

Now, we turn to the second assertion (2.12). Notice that

$$\begin{aligned}\varphi(w^k, \tilde{w}^k) &= \|w^k - \tilde{w}^k\|_H^2 \\ &\qquad +(\lambda^k - \tilde{\lambda}^k)^T \big(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k) + C(z^k - \tilde{z}^k)\big) \\ &= \begin{pmatrix} x^k - \tilde{x}^k \\ y^k - \tilde{y}^k \\ z^k - \tilde{z}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} \beta A^T A & 0 & 0 & \frac{1}{2}A^T \\ 0 & \beta B^T B & 0 & \frac{1}{2}B^T \\ 0 & 0 & \beta C^T C & \frac{1}{2}C^T \\ \frac{1}{2}A & \frac{1}{2}B & \frac{1}{2}C & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} x^k - \tilde{x}^k \\ y^k - \tilde{y}^k \\ z^k - \tilde{z}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{\beta}A(x^k - \tilde{x}^k) \\ \sqrt{\beta}B(y^k - \tilde{y}^k) \\ \sqrt{\beta}C(z^k - \tilde{z}^k) \\ \sqrt{1/\beta}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix}^T \begin{pmatrix} I_m & 0 & 0 & \frac{1}{2}I_m \\ 0 & I_m & 0 & \frac{1}{2}I_m \\ 0 & 0 & I_m & \frac{1}{2}I_m \\ \frac{1}{2}I_m & \frac{1}{2}I_m & \frac{1}{2}I_m & I_m \end{pmatrix} \begin{pmatrix} \sqrt{\beta}A(x^k - \tilde{x}^k) \\ \sqrt{\beta}B(y^k - \tilde{y}^k) \\ \sqrt{\beta}C(z^k - \tilde{z}^k) \\ \sqrt{1/\beta}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix} \\ &\quad .\end{aligned} \tag{2.13}$$

Because the eigenvalues of the matrix

$$\begin{pmatrix} 1 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 1 \end{pmatrix} \quad \text{are } 1,\ 1 \text{ and } 1 \pm \frac{\sqrt{3}}{2},$$

The smallest eigenvalue is $\frac{2-\sqrt{3}}{2}$. Therefore,

$$\begin{aligned} \varphi(w^k, \tilde{w}^k) &\geq \frac{2-\sqrt{3}}{2} \begin{pmatrix} \sqrt{\beta}A(x^k - \tilde{x}^k) \\ \sqrt{\beta}B(y^k - \tilde{y}^k) \\ \sqrt{\beta}C(z^k - \tilde{z}^k) \\ \sqrt{1/\beta}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix}^T \begin{pmatrix} \sqrt{\beta}A(x^k - \tilde{x}^k) \\ \sqrt{\beta}B(y^k - \tilde{y}^k) \\ \sqrt{\beta}C(z^k - \tilde{z}^k) \\ \sqrt{1/\beta}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix} \\ &= \frac{2-\sqrt{3}}{2}\|w^k - \tilde{w}^k\|_H^2. \end{aligned}$$

The lemma is proved. □

According to Lemma 2.2, we have

$$(w^k - w^*)^T H(w^k - \tilde{w}^k) \geq \varphi(w^k, \tilde{w}^k) \geq \frac{2-\sqrt{3}}{2}\|w^k - \tilde{w}^k\|_H^2.$$

This give us the possibility to produce a new iterate which is more closed to the

solution set by the following correction.

Correction: Based on the predictor by (2.1), update the new iterate w^{k+1} by:

$$w^{k+1} = w^k - \alpha_k(w^k - \tilde{w}^k), \qquad \alpha_k = \gamma\alpha_k^*, \qquad \gamma \in (0,2) \tag{2.14a}$$

where

$$\alpha_k^* = \frac{\varphi(w^k, \tilde{w}^k)}{\|w^k - \tilde{w}^k\|_H^2}. \tag{2.14b}$$

By using (2.10) and (2.14), we obtain

$$\begin{aligned} \|w^{k+1} - w^*\|_H^2 &= \|(w^k - w^*) - \gamma\alpha_k^*(w^k - \tilde{w}^k)\|_H^2 \\ &= \|w^k - w^*\|_H^2 - 2\gamma\alpha_k^*(w^k - w^*)^T H(w^k - \tilde{w}^k) \\ &\quad + \gamma^2(\alpha_k^*)^2\|w^k - \tilde{w}^k\|_H^2 \\ &\leq \|w^k - w^*\|_H^2 - 2\gamma\alpha_k^*\varphi(w^k, \tilde{w}^k) \\ &\quad + \gamma^2(\alpha_k^*)^2\|w^k - \tilde{w}^k\|_H^2 \\ &\leq \|w^k - w^*\|_H^2 - \gamma(2-\gamma)\alpha_k^*\varphi(w^k, \tilde{w}^k). \end{aligned} \tag{2.15}$$

In comparison with the computational load for obtaining $(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k)$, the calculation cost for step-size α_k^* is slight.

Convergence Using (2.14) and (2.14b), we obtain

$$\|w^{k+1} - w^*\|_H^2 \le \|w^k - w^*\|_H^2 - \frac{\gamma(2-\gamma)(7-4\sqrt{3})}{4}\|w^k - \tilde{w}^k\|_H^2.$$

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

3 Partially Parallel Splitting ALM

In the partially parallel splitting ALM, x is an intermediate variable. The iteration begins with $v^k = (y^k, z^k, \lambda^k)$ and produces new iterate v^{k+1}. It is still a prediction-correction method.

Prediction:

1. From given (y^k, z^k, λ^k), obtain $\tilde{x}^k, \tilde{y}^k$ and $\tilde{z}^k$, by the following partially parallel manner:

$$\tilde{x}^k = \text{Argmin}\left\{ \begin{array}{c} \theta_1(x) - (\lambda^k)^T(Ax + By^k + Cz^k - b) \\ +\frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 \end{array} \middle| x \in \mathcal{X} \right\} \quad (3.1a)$$

$$\tilde{y}^k = \text{Argmin}\left\{ \begin{array}{c} \theta_2(y) - (\lambda^k)^T(A\tilde{x}^k + By + Cz^k - b) \\ +\frac{\beta}{2}\|A\tilde{x}^k + By + Cz^k - b\|^2 \end{array} \middle| y \in \mathcal{Y} \right\} \quad (3.1b)$$

$$\tilde{z}^k = \text{Argmin}\left\{ \begin{array}{c} \theta_3(z) - (\lambda^k)^T(A\tilde{x}^k + By^k + Cz - b) \\ +\frac{\beta}{2}\|A\tilde{x}^k + By^k + Cz - b\|^2 \end{array} \middle| z \in \mathcal{Z} \right\} \quad (3.1c)$$

2. Update $\tilde{\lambda}^k$ by

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b). \quad (3.1d)$$

Note that in (3.1b) and (3.1c), we use $\tilde{x}^k$ generated by (3.1a).

Analysis The solution $(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ of (3.1a)-(3.1c) satisfies

$$\begin{cases} \theta_1(x) - \theta_1(\tilde{x}^k) + (x - \tilde{x}^k)^T \{-A^T\lambda^k + \beta A^T\big(A\tilde{x}^k + By^k + Cz^k - b\big)\} \geq 0 \\ \theta_2(y) - \theta_2(\tilde{y}^k) + (y - \tilde{y}^k)^T \{-B^T\lambda^k + \beta B^T\big(A\tilde{x}^k + B\tilde{y}^k + Cz^k - b\big)\} \geq 0 \\ \theta_3(z) - \theta_3(\tilde{z}^k) + (z - \tilde{z}^k)^T \{-C^T\lambda^k + \beta C^T\big(A\tilde{x}^k + By^k + C\tilde{z}^k - b\big)\} \geq 0 \end{cases} \tag{3.2}$$

for all $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. Using $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b)$ and by a manipulation, (3.2) can be rewritten as $(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$,

$$\theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \\ z - \tilde{z}^k \end{pmatrix}^T \begin{pmatrix} -A^T\tilde{\lambda}^k + \beta A^T[B(y^k - \tilde{y}^k) + C(z^k - \tilde{z}^k)] \\ -B^T\tilde{\lambda}^k + \beta B^T[\quad 0 \quad + C(z^k - \tilde{z}^k)] \\ -C^T\tilde{\lambda}^k + \beta C^T[B(y^k - \tilde{y}^k) + \quad 0 \quad] \end{pmatrix} \geq 0,$$

for all $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. Combining the last inequality with (3.1d), we

have

$$\theta(u) - \theta(\tilde{u}^k) + \begin{pmatrix} x - \tilde{x}^k \\ y - \tilde{y}^k \\ z - \tilde{z}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\tilde{\lambda}^k \\ -B^T\tilde{\lambda}^k \\ -C^T\tilde{\lambda}^k \\ A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b \end{pmatrix} \right.$$

$$+\beta \begin{pmatrix} A^T \\ B^T \\ C^T \\ 0 \end{pmatrix} \big(B(y^k - \tilde{y}^k) + C(z^k - \tilde{z}^k)\big)$$

$$\left. + \begin{pmatrix} 0 & 0 & 0 \\ \beta B^TB & 0 & 0 \\ 0 & \beta C^TC & 0 \\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} \tilde{y}^k - y^k \\ \tilde{z}^k - z^k \\ \tilde{\lambda}^k - \lambda \end{pmatrix} \right\} \geq 0, \ \forall w \in \Omega. \tag{3.3}$$

Based on the above analysis, we have the following lemma.

Lemma 3.1 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (3.1) *from the given* $v^k = (y^k, z^k, \lambda^k)$*. Then, we have*

$$(\tilde{v}^k - v^*)^T H(v^k - \tilde{v}^k) \ge (\tilde{w}^k - w^*)^T \big(F(\tilde{w}^k) + \eta(u^k, \tilde{u}^k)\big), \tag{3.4}$$

where

$$\eta(u^k, \tilde{u}^k) = \beta \begin{pmatrix} A^T \\ B^T \\ C^T \\ 0 \end{pmatrix} \big(B(y^k - \tilde{y}^k) + C(z^k - \tilde{z}^k)\big) \tag{3.5}$$

and

$$H = \begin{pmatrix} \beta B^T B & 0 & 0 \\ 0 & \beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{3.6}$$

Proof. Setting $(x, y, z, \lambda) = (x^*, y^*, z^*, \lambda^*)$ in (3.3), and using $v = (y, z, \lambda)$ the assertion follows directly. $\square$

Since F is monotone and $\tilde{w}^k \in \Omega$, it follows that

$$(\tilde{w}^k - w^*)^T F(\tilde{w}^k) \ge (\tilde{w}^k - w^*)^T F(w^*) \ge 0. \tag{3.7}$$

In addition, by using $Ax^* + By^* + Cz^* = b$ and

$$\beta(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b) = (\lambda^k - \tilde{\lambda}^k),$$

we have

$$\begin{aligned} &\eta(u^k, \tilde{u}^k)^T(\tilde{w}^k - w^*) \\ &\quad = \quad (B(y^k - \tilde{y}^k) + C(z^k - \tilde{z}^k))^T \beta(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b) \\ &\quad = \quad (B(y^k - \tilde{y}^k) + C(z^k - \tilde{z}^k))^T (\lambda^k - \tilde{\lambda}^k). \end{aligned} \tag{3.8}$$

Lemma 3.2 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (3.1) *from the given* $v^k = (y^k, z^k, \lambda^k)$*. Then, we have*

$$(v^k - v^*)^T H(v^k - \tilde{v}^k) \ge \varphi(v^k, \tilde{v}^k), \quad \forall\, w^* \in \Omega^*, \tag{3.9}$$

where

$$\varphi(v^k, \tilde{v}^k) = \|v^k - \tilde{v}^k\|_H^2 \\ +(\lambda^k - \tilde{\lambda}^k)^T \big(B(y^k - \tilde{y}^k) + C(z^k - \tilde{z}^k)\big). \tag{3.10}$$

Proof. First, using (3.4), (3.7) and (3.8) we obtain that

$$\begin{aligned} &(\tilde{v}^k - v^*)^T H(v^k - \tilde{v}^k) \\ &\quad \geq \quad (\lambda^k - \tilde{\lambda}^k)^T \big(B(y^k - \tilde{y}^k) + C(z^k - \tilde{z}^k)\big). \end{aligned}$$

The assertion of this lemma follows from the last inequality and the definition of $\varphi(v^k, \tilde{v}^k)$ directly. $\square$

Lemma 3.3 *Let* $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (3.1) *from the given* $v^k = (y^k, z^k, \lambda^k)$. *For* H *defined in* (3.6) *and* $\varphi(v^k, \tilde{v}^k)$ *defined in* (3.10)*, we have*

$$\varphi(v^k, \tilde{v}^k) \geq \frac{2 - \sqrt{2}}{2} \|v^k - \tilde{v}^k\|_H^2. \tag{3.11}$$

Proof. According to the definition of (3.10), we obtain

$$\begin{aligned} \varphi(v^k, \tilde{v}^k) &= \|v^k - \tilde{v}^k\|_H^2 + (\lambda^k - \tilde{\lambda}^k)^T \big(B(y^k - \tilde{y}^k) + C(z^k - \tilde{z}^k)\big) \\ &= \begin{pmatrix} y^k - \tilde{y}^k \\ z^k - \tilde{z}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} \beta B^T B & 0 & \frac{1}{2} B^T \\ 0 & \beta C^T C & \frac{1}{2} C^T \\ \frac{1}{2} B & \frac{1}{2} C & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ z^k - \tilde{z}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{\beta} B(y^k - \tilde{y}^k) \\ \sqrt{\beta} C(z^k - \tilde{z}^k) \\ \sqrt{1/\beta}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix}^T \begin{pmatrix} I_m & 0 & \frac{1}{2} I_m \\ 0 & I_m & \frac{1}{2} I_m \\ \frac{1}{2} I_m & \frac{1}{2} I_m & I_m \end{pmatrix} \begin{pmatrix} \sqrt{\beta} B(y^k - \tilde{y}^k) \\ \sqrt{\beta} C(z^k - \tilde{z}^k) \\ \sqrt{1/\beta}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix}. \end{aligned}$$

Because the matrix

$$\begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 \end{pmatrix}$$

is positive definite and its eigenvalues are 1 and $1 \pm \frac{\sqrt{2}}{2}$, thus, we obtain

$$\begin{aligned}
\varphi(v^k, \tilde{v}^k) &\geq \frac{2-\sqrt{2}}{2} \begin{pmatrix} \sqrt{\beta}B(y^k - \tilde{y}^k) \\ \sqrt{\beta}C(z^k - \tilde{z}^k) \\ \sqrt{1/\beta}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix}^T \begin{pmatrix} \sqrt{\beta}B(y^k - \tilde{y}^k) \\ \sqrt{\beta}C(z^k - \tilde{z}^k) \\ \sqrt{1/\beta}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix} \\
&= \frac{2-\sqrt{2}}{2} \|v^k - \tilde{v}^k\|_H^2,
\end{aligned}$$

and the assertion is proved. $\square$

Correction Based on the predictor by (3.1), update the new iterate v^{k+1} by

$$v^{k+1} = v^k - \alpha_k(v^k - \tilde{v}^k), \qquad \alpha_k = \gamma\alpha_k^*, \qquad \gamma \in (0,2) \tag{3.12a}$$

where

$$\alpha_k^* = \frac{\varphi(v^k, \tilde{v}^k)}{\|v^k - \tilde{v}^k\|_H^2} \tag{3.12b}$$

By using (3.9) and (3.12), we obtain

$$\begin{aligned}
\|v^{k+1} - v^*\|_H^2 &= \|(v^k - v^*) - \gamma\alpha_k^*(v^k - \tilde{v}^k)\|_H^2 \\
&= \|v^k - v^*\|_H^2 - 2\gamma\alpha_k^*(v^k - v^*)^T H(v^k - \tilde{v}^k) + \gamma^2(\alpha_k^*)^2\|v^k - \tilde{v}^k\|_H^2 \\
&\leq \|v^k - v^*\|_H^2 - 2\gamma\alpha_k^*\varphi(v^k, \tilde{v}^k) + \gamma^2(\alpha_k^*)^2\|v^k - \tilde{v}^k\|_H^2 \\
&= \|v^k - v^*\|_H^2 - \gamma(2-\gamma)\alpha_k^*\varphi(v^k, \tilde{v}^k).
\end{aligned} \tag{3.13}$$

In comparison with the computational load for obtaining $(\tilde{x}^k, \tilde{y}^k, \tilde{z}^k)$, the calculation cost for step-size α_k^* is slight.

Convergence Using (3.11) and (3.12b), it follows from (3.13) that

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \frac{\gamma(2-\gamma)(3-2\sqrt{2})}{2}\|v^k - \tilde{v}^k\|_H^2.$$

From (3.11) we know that $\alpha_k^* \geq \frac{2-\sqrt{2}}{2}$. In the correction step, we can also take the fixed step size $\alpha_k \equiv \alpha \in (0, 2-\sqrt{2})$, the method is still convergent.

4 Extension to problems with 4 operators

We consider the following optimization problem with 4 separable operators:

$$\min\{\sum_{i=1}^{4}\theta_i(x_i) \mid \sum_{i=1}^{4} A_i x_i = b, x_i \in \mathcal{X}_i\} \tag{4.1}$$

where $\theta_i(x_i) : \Re^{n_i} \to \Re$, is convex, $A_i \in \Re^{m\times n_i}$, $\mathcal{X}_i \subset \Re^{n_i}$ is given convex set, $i = 1, \ldots, 4$; $b \in \Re^m$. This optimization problem is equivalent to find $w^* = (x_1^*, x_2^*, x_3^*, x_4^*, \lambda^*) \in \Omega$, such that

$$\left\{\begin{array}{l} \theta_1(x_1) - \theta_1(x_1^*) + (x_1 - x_1^*)^T(-A_1^T\lambda^*) \ge 0, \\ \theta_2(x_2) - \theta_2(x_2^*) + (x_2 - x_2^*)^T(-A_2^T\lambda^*) \ge 0, \\ \theta_3(x_3) - \theta_3(x_3^*) + (x_3 - x_3^*)^T(-A_4^T\lambda^*) \ge 0, \\ \theta_4(x_4) - \theta_4(x_4^*) + (x_4 - x_4^*)^T(-A_4^T\lambda^*) \ge 0, \\ (\lambda - \lambda^*)^T(A_1x_1^* + A_2x_2^* + A_3x_3^* + A_4x_4^* - b) \ge 0, \end{array}\right. \qquad \forall\, w \in \Omega, \tag{4.2}$$

where

$$\Omega = \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3 \times \mathcal{X}_4 \times \Re^m.$$

By denoting

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}, \quad w = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \lambda \end{pmatrix} \quad F(w) = \begin{pmatrix} -A_1^T\lambda \\ -A_2^T\lambda \\ -A_3^T\lambda \\ -A_4^T\lambda \\ \sum_{i=1}^4 A_i x_i - b, \end{pmatrix},$$

and $\quad \theta(x) = \theta_1(x_1) + \theta_2(x_2) + \theta_3(x_3) + \theta_4(x_4),$

the optimal condition (4.2) can be written as

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

Full parallel splitting augmented Lagrangian method

If we extend the full parallel splitting augmented Lagrangian method to solve (4.1), similarly as in Lemma 2.2 (see (2.10) and (2.11)), we have

$$(w^k - w^*)^T H(w^k - \tilde{w}^k) \ge \varphi(w^k, \tilde{w}^k), \quad \forall\, w^* \in \Omega^*, \tag{4.3}$$

where

$$\varphi(w^k, \tilde{w}^k) = \|w^k - \tilde{w}^k\|_H^2 + (\lambda^k - \tilde{\lambda}^k)^T \Big(\sum_{i=1}^{4} A_i(x_i^k - \tilde{x}_i^k)\Big) \qquad (4.4)$$

and

$$H = \textbf{diag}(\beta A_1^T A_1, \beta A_2^T A_2, \beta A_3^T A_3, \beta A_4^T A_4, \frac{1}{\beta} I_m).$$

Note that, similar as (2.13), we have

$$\varphi(w^k, \tilde{w}^k) = \begin{pmatrix} \sqrt{\beta} A_1(x_1^k - \tilde{x}_1^k) \\ \sqrt{\beta} A_2(x_2^k - \tilde{x}_2^k) \\ \sqrt{\beta} A_3(x_3^k - \tilde{x}_3^k) \\ \sqrt{\beta} A_4(x_4^k - \tilde{x}_4^k) \\ \sqrt{1/\beta}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix}^T \begin{pmatrix} I & 0 & 0 & 0 & I/2 \\ 0 & I & 0 & 0 & I/2 \\ 0 & 0 & I & 0 & I/2 \\ 0 & 0 & 0 & I & I/2 \\ I/2 & I/2 & I/2 & I/2 & I \end{pmatrix} \begin{pmatrix} \sqrt{\beta} A_1(x_1^k - \tilde{x}_1^k) \\ \sqrt{\beta} A_2(x_2^k - \tilde{x}_2^k) \\ \sqrt{\beta} A_3(x_3^k - \tilde{x}_3^k) \\ \sqrt{\beta} A_4(x_4^k - \tilde{x}_4^k) \\ \sqrt{1/\beta}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix}.$$

In other words, for the problem with 4 separable operators, if we use the full

parallel splitting augmented Lagrangian method, we will met a matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 1 \end{pmatrix}.$$

The above matrix is positive semi-definite and its eigenvalues are $0, 1, 1, 1$ and 2. The vector $(1, 1, 1, 1, -2)^T$ is the eigenvector related to the eigenvalue 0. This means, even through $\|w^k - \tilde{w}^k\|_H \neq 0$, it can not guarantee that $\varphi(w^k, \tilde{w}^k) > 0$. Therefore, we suggest to use the partially splitting augmented Lagrangian method.

Partially parallel splitting augmented Lagrangian method

In the partially parallel splitting augmented Lagrangian method, x_1 is an intermediate variable during the iterations. After solving the x_1-subproblem, then solve the x_2, x_3, x_4-subproblems in the parallel manner !

Prediction Similarly as the prediction step (3.1) for problems with 3 operators

From given $(x_2^k, x_3^k, x_4^k, \lambda^k)$, obtain $\tilde{w}^k$ by the following partially parallel manner:

$$\tilde{x}_1^k = \text{Argmin}\left\{\begin{array}{c}\theta_1(x_1) - (\lambda^k)^T(A_1x_1 + A_2x_2^k + A_3x_3^k + A_4x_4^k - b)\\ +\frac{\beta}{2}\|A_1x_1 + A_2x_2^k + A_3x_3^k + A_4x_4^k - b\|^2\end{array}\middle| x_1 \in \mathcal{X}_1\right\} \quad (4.5a)$$

$$\tilde{x}_2^k = \text{Argmin}\left\{\begin{array}{c}\theta_2(x_2) - (\lambda^k)^T(A_1\tilde{x}_1^k + A_2x_2 + A_3x_3^k + A_4x_4^k - b)\\ +\frac{\beta}{2}\|A_1\tilde{x}_1^k + A_2x_2 + A_3x_3^k + A_4x_4^k - b\|^2\end{array}\middle| x_2 \in \mathcal{X}_2\right\} \quad (4.5b)$$

$$\tilde{x}_3^k = \text{Argmin}\left\{\begin{array}{c}\theta_3(x_3) - (\lambda^k)^T(A_1\tilde{x}_1^k + A_2x_2^k + A_3x_3 + A_4x_4^k - b)\\ +\frac{\beta}{2}\|A_1\tilde{x}_1^k + A_2x_2^k + A_3x_3 + A_4x_4^k - b\|^2\end{array}\middle| x_3 \in \mathcal{X}_3\right\} \quad (4.5c)$$

$$\tilde{x}_4^k = \text{Argmin}\left\{\begin{array}{c}\theta_4(x_4) - (\lambda^k)^T(A_1\tilde{x}_1^k + A_2x_2^k + A_3x_3^k + A_4x_4 - b)\\ +\frac{\beta}{2}\|A_1\tilde{x}_1^k + A_2x_2^k + A_3x_3^k + A_4x_4 - b\|^2\end{array}\middle| x_4 \in \mathcal{X}_4\right\} \quad (4.5d)$$

Update $\tilde{\lambda}^k$ by

$$\tilde{\lambda}^k = \lambda^k - \beta(A_1\tilde{x}_1^k + A_2\tilde{x}_2^k + A_3\tilde{x}_3^k + A_4\tilde{x}_4^k - b). \quad (4.5e)$$

Note that in (4.5b), (4.5c) and (4.5d), we use $\tilde{x}_1^k$ generated by (4.5a).

Similar as (3.3), for the $\tilde{w}^k$ generated by (4.5), we have

$$\theta(x) - \theta(\tilde{x}^k) + \begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ x_3 - \tilde{x}_3^k \\ x_4 - \tilde{x}_4^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A_1^T\tilde{\lambda}^k \\ -A_2^T\tilde{\lambda}^k \\ -A_3^T\tilde{\lambda}^k \\ -A_4^T\tilde{\lambda}^k \\ A_1\tilde{x}_1^k + A_2\tilde{x}_2^k + A_3\tilde{x}_3^k + A_4\tilde{x}_4^k - b \end{pmatrix} \right.$$

$$+\beta \begin{pmatrix} A_1^T \\ A_2^T \\ A_3^T \\ A_4^T \\ 0 \end{pmatrix} \left[A_2(x_2^k - \tilde{x}_2^k) + A_3(x_3^k - \tilde{x}_3^k) + A_4(x_4^k - \tilde{x}_4^k)\right]$$

$$\left. + \begin{pmatrix} 0 & 0 & 0 & 0 \\ \beta A_2^TA_2 & 0 & 0 & 0 \\ 0 & \beta A_3^TA_3 & 0 & 0 \\ 0 & 0 & \beta A_4^TA_4 & 0 \\ 0 & 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix} \begin{pmatrix} \tilde{x}_2^k - x_2^k \\ \tilde{x}_3^k - x_3^k \\ \tilde{x}_4^k - x_4^k \\ \tilde{\lambda}^k - \lambda \end{pmatrix} \right\} \geq 0, \ \forall w \in \Omega. \quad (4.6)$$

For convenience we use the notations

$$v = \begin{pmatrix} x_2 \\ x_3 \\ x_4 \\ \lambda \end{pmatrix} \quad \text{and} \quad \mathcal{V}^* = \{(x_2^*, x_3^*, x_4^*, \lambda^*) \,|\, (x_1^*, x_2^*, x_3^*, x_4^*, \lambda^*) \in \Omega^*\}.$$

Based on the above analysis, we have the following lemma.

Lemma 4.1 *Let* $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \tilde{x}_3^k, \tilde{x}_4^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (4.5) *from the given* $v^k = (x_2^k, x_3^k, x_4^k, \lambda^k)$. *Then, we have*

$$(\tilde{v}^k - v^*)^T H(v^k - \tilde{v}^k) \ge (\tilde{w}^k - w^*)^T \big(F(\tilde{w}^k) + \eta(x^k, \tilde{x}^k)\big), \tag{4.7}$$

where

$$H = \textbf{diag}(\beta A_2^T A_2, \beta A_3^T A_3, \beta A_4^T A_4, \frac{1}{\beta} I_m) \tag{4.8}$$

and

$$\eta(x^k, \tilde{x}^k) = \beta \begin{pmatrix} A_1^T \\ A_2^T \\ A_3^T \\ A_4^T \\ 0 \end{pmatrix} \big[A_2(x_2^k - \tilde{x}_2^k) + A_3(x_3^k - \tilde{x}_3^k) + A_4(x_4^k - \tilde{x}_4^k)\big] \tag{4.9}$$

Lemma 4.2 *Let* $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \tilde{x}_3^k, \tilde{x}_4^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (4.5) *from the given* $v^k = (x_2^k, x_3^k, x_4^k, \lambda^k)$. *Then, we have*

$$(v^k - v^*)^T H(v^k - \tilde{v}^k) \ge \varphi(v^k, \tilde{v}^k), \quad \forall\, v^* \in \mathcal{V}^*, \tag{4.10}$$

where

$$\begin{aligned} \varphi(v^k, \tilde{v}^k) &= \|v^k - \tilde{v}^k\|_H^2 \\ &\quad + (\lambda^k - \tilde{\lambda}^k)^T \big[A_2(x_2^k - \tilde{x}_2^k) + A_3(x_3^k - \tilde{x}_3^k) + A_4(x_4^k - \tilde{x}_4^k)\big]. \end{aligned} \tag{4.11}$$

The proofs of Lemma 4.2 is similar as those in Lemma 3.2.

Lemma 4.3 *Let* $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \tilde{x}_3^k, \tilde{x}_4^k, \tilde{\lambda}^k) \in \Omega$ *be generated by* (4.5) *from the given* $v^k = (x_2^k, x_3^k, x_4^k, \lambda^k)$. *For* H *defined in* (4.8) *and* $\varphi(v^k, \tilde{v}^k)$ *defined in* (4.11)*, we have*

$$\varphi(v^k, \tilde{v}^k) \geq \frac{2-\sqrt{3}}{2}\|v^k - \tilde{v}^k\|_H^2. \tag{4.12}$$

Proof. According to the definition of (4.11), we have

$$\begin{aligned}\varphi(v^k, \tilde{v}^k) &= \|v^k - \tilde{v}^k\|_H^2 \\ &\quad +(\lambda^k - \tilde{\lambda}^k)^T\left[A_2(x_2^k - \tilde{x}_2^k) + A_3(x_3^k - \tilde{x}_3^k) + A_4(x_4^k - \tilde{x}_4^k)\right] \\ &= \begin{pmatrix} \sqrt{\beta}A_2(x_2^k - \tilde{x}_2^k) \\ \sqrt{\beta}A_3(x_3^k - \tilde{x}_3^k) \\ \sqrt{\beta}A_4(x_4^k - \tilde{x}_4^k) \\ \sqrt{1/\beta}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix}^T \begin{pmatrix} I_m & 0 & 0 & \frac{1}{2}I_m \\ 0 & I_m & 0 & \frac{1}{2}I_m \\ 0 & 0 & I_m & \frac{1}{2}I_m \\ \frac{1}{2}I_m & \frac{1}{2}I_m & \frac{1}{2}I_m & I_m \end{pmatrix} \begin{pmatrix} \sqrt{\beta}A_2(x_2^k - \tilde{x}_2^k) \\ \sqrt{\beta}A_3(x_3^k - \tilde{x}_3^k) \\ \sqrt{\beta}A_4(x_4^k - \tilde{x}_4^k) \\ \sqrt{1/\beta}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix}.\end{aligned}$$

Note that the eigenvalues of the symmetric matrix

$$\begin{pmatrix} 1 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 1 \end{pmatrix} \quad \text{are} \quad 1,\ 1 \text{ and } 1 \pm \frac{\sqrt{3}}{2},$$

and the smallest eigenvalue is $\frac{2-\sqrt{3}}{2}$. Therefore,

$$\begin{aligned}\varphi(v^k, \tilde{v}^k) &\geq \frac{2-\sqrt{3}}{2} \begin{pmatrix} \sqrt{\beta}A_2(x_2^k - \tilde{x}_2^k) \\ \sqrt{\beta}A_3(x_3^k - \tilde{x}_3^k) \\ \sqrt{\beta}A_4(x_4^k - \tilde{x}_4^k) \\ \sqrt{1/\beta}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix}^T \begin{pmatrix} \sqrt{\beta}A_2(x_2^k - \tilde{x}_2^k) \\ \sqrt{\beta}A_3(x_3^k - \tilde{x}_3^k) \\ \sqrt{\beta}A_4(x_4^k - \tilde{x}_4^k) \\ \sqrt{1/\beta}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix} \\ &= \frac{2-\sqrt{3}}{2}\|v^k - \tilde{v}^k\|_H^2,\end{aligned}$$

and the assertion is proved. □

Correction Based on the predictor by (4.5), update the new iterate v^{k+1} by

Update the new iterate v^{k+1} by

$$v^{k+1} = v^k - \alpha_k(v^k - \tilde{v}^k), \qquad \alpha_k = \gamma\alpha_k^*, \qquad \gamma \in (0,2) \tag{4.13a}$$

where

$$\alpha_k^* = \frac{\varphi(v^k, \tilde{v}^k)}{\|v^k - \tilde{v}^k\|_H^2} \tag{4.13b}$$

By using (4.10) and (4.13), we obtain

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

Convergence Using (4.12) and (4.13b), we obtain

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \frac{\gamma(2-\gamma)(7-4\sqrt{3})}{4}\|v^k - \tilde{v}^k\|_H^2. \tag{4.14}$$

This inequality (4.14) is essential for the convergence.

5 Conclusion Remarks

All the methods in this note are prediction-correction methods. The predictor offers us a descent direction of the unknown function $\frac{1}{2}\|w - w^*\|_H^2$ and the correction produces the new iterate which is more closed to the solution set.

The first ADMM-based prediction-correction method was published by Ye and Yuan [13]. For solving the problem (1.1), we suggest to use the method in Section 3, whose prediction and correction is (3.1) and (3.12), respectively. Even though it needs to compute the step size in the correction step in each iteration, the cost is usually small in comparison with the computational load for solving the subproblem in the prediction step.

Since 2012, we have some new publication about the multi-block problems. First, without correction, the direct extension of the ADMM for the problem (1.1) is not necessarily convergent [2]. For multi-block problems, we suggest to use the methods [7, 8, 9, 11].

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, Nov. 2010.

[2] 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 necessary convergent, Mathematical Programming, 155 (2016) 57-79.

[3] D. Gabay, "Applications of the Method of Multipliers to Variational Inequalities," in *Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems*, M. Fortin and R. Glowinski, eds., North Holland, Amsterdam, The Netherlands, 299-331, 1983.

[4] B.S. He, A class of projection and contraction methods for monotone variational inequalities, Applied Mathematics and Optimization, **35**, pp. 69-76, 1997.

[5] B.S. He, "Inexact implicit methods for monotone general variational inequalities", Math. Program. Vol. 86, Series A, pp. 199-217, 1999

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

[7] B.S. He, L.S. Hou, and X.M. Yuan, On Full Jacobian Decomposition of the Augmented Lagrangian Method for Separable Convex Programming, SIAM J. Optim., 25 (2015) 2274 − 2312.

[8] B.S. He, M. Tao and X.M. Yuan, Alternating Direction Method with Gaussian Back Substitution for Separable Convex Programming, SIAM J. Optim. 22(2012), 313-340.

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

[10] B. S. He and M.-H. Xu, A general framework of contraction methods for monotone variational inequalities, Pacific J. Optimization, 4, 195-212, 2008.

[11] B.S. He and X.M. Yuan, Block-wise Alternating Direction Method of Multipliers for Multiple-block Convex Programming and Beyond, SMAI J. Computational Mathematics 1 (2015) 145-174.

[12] R.T. Rockafellar, Monotone operators and the proximal point algorithm, SIAM J. Control Optim., 14 (1976), pp. 877-898.

[13] C. H. Ye and X. M. Yuan, A Descent Method for Structured Monotone Variational Inequalities, Optimization Methods and Software, Vol. 22, pp. 329-338, 2007.

凸优化和单调变分不等式的收缩算法

第十七讲: 三块可分离凸优化问题的略有改动的交替方向法

A slightly changed ADMM for convex optimization with three separable blocks

南京大学数学系　何炳生
hebma@nju.edu.cn

The context of this lecture is based on the publication [10] and [13]

Abstract. The classical alternating direction method of multipliers (ADMM) has been well studied in the context of linearly constrained convex programming and variational inequalities where the involved operator is formed as the sum of two individual functions without crossed variables. Recently, ADMM has found many novel applications in diversified areas such as image processing and statistics. However, it is still not clear whether ADMM can be extended to the case where the operator is the sum of more than two individual functions. In this lecture, we present a little changed ADMM for solving the linearly constrained separable convex optimization whose involved operator is separable into three individual functions. The $\mathcal{O}(1/t)$ convergence rate of the proposed methods is demonstrated.

Keywords: Alternating direction method, convex programming, linear constraint, separable structure, contraction method

1 Introduction

An important case of structured convex optimization problem is

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

where $\theta_1 : \Re^n \to \Re$ and $\theta_2 : \Re^m \to \Re$ are closed proper convex functions (not necessarily smooth); $A \in \Re^{m\times n}$; $\mathcal{X} \subseteq \Re^n$ and $\mathcal{Y} \subseteq \Re^m$ are closed convex sets. The alternating direction method of multipliers (ADMM), which dates back to [6] and is closely related to the Douglas-Rachford operator splitting method [2], is perhaps the most popular method for solving (1.1). More specifically, for given (y^k, λ^k) in the k-th iteration, it produces the new iterate in the following order:

$$\left\{\begin{array}{l} x^{k+1} = \text{Argmin}\{\theta_1(x) - (\lambda^k)^T Ax + \frac{\beta}{2}\|Ax + y^k - b\|^2 \mid x \in \mathcal{X}\}; \\ y^{k+1} = \text{Argmin}\{\theta_2(y) - (\lambda^k)^T y + \frac{\beta}{2}\|Ax^{k+1} + y - b\|^2 \mid y \in \mathcal{Y}\}; \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + y^{k+1} - b). \end{array}\right. \tag{1.2}$$

Therefore, ADMM can be viewed as a practical and structured-exploiting variant (split form or relaxed form) of ALM for solving the separable problem (1.1), with the adaption of minimizing the involved separable variables x and y separably in an alternating order. In fact, the iteration (1.2) is from (y^k, λ^k) to (y^{k+1}, λ^{k+1}), x is only an auxiliary variable in the iterative process. The sequence $\{(y^k, \lambda^k)\}$ generated by the recursion (1.2) satisfies

(see Theorem 1 in [12] by setting fixed β and $\gamma \equiv 1$)

$$\begin{aligned} &\|\beta(y^{k+1} - y^*)\|^2 + \|\lambda^{k+1} - \lambda^*\|^2 \\ &\quad \le \quad \|\beta(y^k - y^*)\|^2 + \|\lambda^k - \lambda^*\|^2 - \big(\|\beta(y^k - y^{k+1})\|^2 + \|\lambda^k - \lambda^{k+1}\|^2\big). \end{aligned}$$

Because of its efficiency and easy implementation, ADMM has attracted wide attention of many authors in various areas, see e.g. [1, 7]. In particular, some novel and attractive applications of ADMM have been discovered very recently, e.g. the total-variation problem in image processing, the covariance selection problem and semidefinite least square problem in statistics [11], the semidefinite programming problems, the sparse and low-rank recovery problem in Engineering [14], and the matrix completion problem [1].

In some practical applications [4], the model is slightly more complicated than (1.1). The mathematical form of the problem is

$$\min\{\theta_1(x) + \theta_2(y) + \theta_3(z) | Ax + y + z = b,\ x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\}, \tag{1.3}$$

where $\theta_1 : \Re^n \to \Re$, $\theta_2, \theta_3 : \Re^m \to \Re$ are closed proper convex functions (not necessarily smooth); $A \in \Re^{m\times n}$; $\mathcal{X} \subseteq \Re^n$, $\mathcal{Y}, \mathcal{Z} \subseteq \Re^m$ are closed convex sets. It is then natural to manage to extend ADMM to solve the problem (1.3), resulting in the

following scheme:

$$\begin{cases} x^{k+1} = \text{Argmin}\{\theta_1(x) - (\lambda^k)^T Ax + \frac{\beta}{2}\|Ax + y^k + z^k - b\|^2 \mid x \in \mathcal{X}\}; \\ y^{k+1} = \text{Argmin}\{\theta_2(y) - (\lambda^k)^T y \ + \ \frac{\beta}{2}\|Ax^{k+1} + y + z^k - b\|^2 \mid y \in \mathcal{Y}\}; \\ z^{k+1} = \text{Argmin}\{\theta_3(z) - (\lambda^k)^T z \ + \ \frac{\beta}{2}\|Ax^{k+1} + y^{k+1} + z - b\|^2 \mid z \in \mathcal{Z}\}; \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + y^{k+1} + z^{k+1} - b), \end{cases} \tag{1.4}$$

and the involved subproblems of (1.4) are solved consecutively in the ADMM manner. Unfortunately, with the $(y^{k+1}, z^{k+1}, \lambda^{k+1})$ offered by (1.4), the convergence of the extended ADMM (1.4) is still open.

In this paper, we present a little changed alternating direction method for the problem (1.3). Again, based on $(y^{k+1}, z^{k+1}, \lambda^{k+1})$ offered by (1.4), we set

$$(y^{k+1}, z^{k+1}, \lambda^{k+1}) := (y^{k+1} + (z^k - z^{k+1}), z^{k+1}, \lambda^{k+1}). \tag{1.5}$$

Note that the change of (1.5) is small. In addition, for the problem with two separable operators, by setting $z^k = 0$ for all k, the proposed method is just reduced to the algorithm (1.2) for the problem (1.1). Therefore, we call the proposed method *a little*

changed alternating direction method of multipliers for convex optimization with three separable operators.

The outline of this paper is as follows. In Section 2, we convert the problem (1.3) to the equivalent variational inequality and characterize its solution set. Section 3 shows the contraction property of the proposed method. In Section 4, we define an auxiliary vector and derive its main associated properties, and show the $\mathcal{O}(1/t)$ convergence rate of the proposed method. Finally, some conclusions are made in Section 6.

2 The variational inequality characterization

Throughout, we assume that the solution set of (1.3) is not empty. The convergence analysis is based on the tool of variational inequality. For this purpose, we define

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

It is easy to verify that the convex programming problem (1.3) is characterized by the following variational inequality: Find $w^* = (x^*, y^*, z^*, \lambda^*) \in \mathcal{W}$ such that

$$\left\{ \begin{array}{l} \theta_1(x) - \theta_1(x^*) + (x - x^*)^T(-A^T\lambda^*) \geq 0, \\ \theta_2(y) - \theta_2(y^*) + (y - y^*)^T(-\lambda^*) \geq 0, \\ \theta_3(z) - \theta_3(z^*) + (z - z^*)^T(-\lambda^*) \geq 0, \\ \qquad\qquad (\lambda - \lambda^*)^T(Ax^* + y^* + z^* - b) \geq 0, \end{array} \right. \qquad \forall\, w \in \mathcal{W}, \tag{2.1}$$

or in the more compact form:

$$\text{VI}(\mathcal{W}, F, \theta) \qquad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \mathcal{W}, \tag{2.2}$$

where

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

and

$$u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \qquad w = \begin{pmatrix} x \\ y \\ z \\ \lambda \end{pmatrix}, \qquad F(w) = \begin{pmatrix} -A^T\lambda \\ -\lambda \\ -\lambda \\ Ax + y + z - b \end{pmatrix}. \tag{2.3}$$

Note that $F(w)$ defined in (2.3) is monotone. Under the nonempty assumption on the solution set of (1.3), the solution set of (2.2)-(2.3), denoted by $\mathcal{W}^*$, is also nonempty.

The Theorem 2.3.5 in [5] provides an insightful characterization for the solution set of a generic VI. This characterization actually provides us a novel and simple approach which enables us to derive the $O(1/t)$ convergence rate for the original ADMM in [13]. In the following theorem, we specify this result for the derived VI$(\mathcal{W}, F, \theta)$. Note that the proof of the next theorem is an incremental extension of Theorem 2.3.5 in [5] and also Theorem 2.1 in [13]. But, we include all the details because of its crucial importance in our analysis.

Theorem 2.1 *The solution set of VI$(\mathcal{W}, F, \theta)$ is convex and it can be characterized as*

$$\mathcal{W}^* = \bigcap_{w \in \mathcal{W}} \big\{\bar{w} \in \mathcal{W} : \big(\theta(u) - \theta(\bar{u})\big) + (w - \bar{w})^T F(w) \geq 0\big\}. \tag{2.4}$$

Proof. Indeed, if $\bar{w} \in \mathcal{W}^*$, according to (2.2) we have

$$\theta(u) - \theta(\bar{u}) + (w - \bar{w})^T F(\bar{w}) \geq 0, \quad \forall\, w \in \mathcal{W}.$$

By using the monotonicity of F on $\mathcal{W}$, this implies

$$\theta(u) - \theta(\bar{u}) + (w - \bar{w})^T F(w) \geq 0, \quad \forall\, w \in \mathcal{W}.$$

Thus, $\bar{w}$ belongs to the right-hand set in (2.4).

Conversely, suppose $\bar{w}$ belongs to the latter set. Let $w \in \mathcal{W}$ be arbitrary. The vector

$$\tilde{w} = \tau\bar{w} + (1 - \tau)w$$

belongs to $\mathcal{W}$ for all $\tau \in (0, 1)$. Thus we have

$$\theta(\tilde{u}) - \theta(\bar{u}) + (\tilde{w} - \bar{w})^T F(\tilde{w}) \geq 0. \tag{2.5}$$

Because $\theta(\cdot)$ is convex and $\tilde{u} = \tau\bar{u} + (1 - \tau)u$, we have

$$\theta(\tilde{u}) \leq \tau\theta(\bar{u}) + (1 - \tau)\theta(u).$$

Substituting it in (2.5), we get

$$(\theta(u) - \theta(\bar{u})) + (w - \bar{w})^T F(\tau\bar{w} + (1 - \tau)w) \geq 0$$

for all $\tau \in (0, 1)$. Letting $\tau \to 1$ yields

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

Thus $\bar{w} \in \mathcal{W}^*$. Now, we turn to prove the convexity of $\mathcal{W}^*$. For each fixed but arbitrary $w \in \mathcal{W}$, the set

$$\{\bar{w} \in \mathcal{W} :\ \theta(\bar{u}) + \bar{w}^T F(w) \leq \theta(u) + w^T F(w)\}$$

is convex and so is the equivalent set

$$\{\bar{w} \in \mathcal{W} :\ \big(\theta(u) - \theta(\bar{u})\big) + (w - \bar{w})^T F(w) \geq 0\}.$$

Since the intersection of any number of convex sets is convex, it follows that the solution set of VI$(\mathcal{W}, F, \theta)$ is convex. $\square$

Theorem 2.1 thus implies that $\bar{w} \in \mathcal{W}$ is an approximate solution of VI$(\mathcal{W}, F, \theta)$ with the accuracy $\epsilon > 0$ if it satisfies

$$\theta(u) - \theta(\bar{u}) + F(w)^T(w - \bar{w}) \geq -\epsilon, \quad \forall\, w \in \mathcal{W}.$$

In this paper, we show that, for given $\epsilon > 0$ and a substantial compact set $\mathcal{D} \subset \mathcal{W}$, after t iterations of the proposed methods, we can find a $\bar{w} \in \mathcal{W}$ such that

$$\hat{w} \in \mathcal{W} \quad \text{and} \quad \sup_{w \in \mathcal{D}} \left\{ \theta(\hat{u}) - \theta(u) + (\hat{w} - w)^T F(w) \right\} \leq \epsilon. \tag{2.6}$$

The convergence rate $\mathcal{O}(1/t)$ of the proposed methods is thus established.

For convenience of coming analysis, we define the following matrices:

$$P = \begin{pmatrix} \beta I_m & 0 & 0 \\ \beta I_m & \beta I_m & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}, \qquad D = \begin{pmatrix} \beta I_m & 0 & 0 \\ 0 & \beta I_m & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{2.7}$$

Note that for the above defined matrices M and D, we have

$$D^{-1}P = \begin{pmatrix} I_m & 0 & 0 \\ I_m & I_m & 0 \\ 0 & 0 & I_m \end{pmatrix}, \quad P^{-T}D = \begin{pmatrix} I_m & -I_m & 0 \\ 0 & I_m & 0 \\ 0 & 0 & I_m \end{pmatrix}. \tag{2.8}$$

3 Contraction property of the proposed method

In the alternating direction method, x is only the auxiliary variable in the iteration process. For convenience of analysis, we use the notation

$$v = (y, z, \lambda),$$

which is a sub-vector of w. For $w^* \in \mathcal{W}^*$, we also define

$$\mathcal{V}^* := \{v^* = (y^*, z^*, \lambda^*) \,|\, (x^*, y^*, z^*, \lambda^*) \in \mathcal{W}^*\}.$$

In addition, we divide each iteration of the proposed method in two steps-the prediction step and the correction step. From a given $v^k = (y^k, z^k, \lambda^k)$, we use

$\bar{w}^k = (\bar{x}^k, \bar{y}^k, \bar{z}^k, \bar{\lambda}^k) \in \mathcal{W}$ to denote the solution of (1.4) and call it as the prediction step.

The prediction step in the k-th iteration (use ADMM manner):
Begin with given $v^k = (y^k, z^k, \lambda^k)$, generate $\bar{w}^k = (\bar{x}^k, \bar{y}^k, \bar{z}^k, \bar{\lambda}^k)$ in the following order:

$$\bar{x}^k = \text{Argmin}\{\theta_1(x) + \frac{\beta}{2}\|(Ax + y^k + z^k - b) - \frac{1}{\beta}\lambda^k\|^2 \mid x \in \mathcal{X}\}, \tag{3.1a}$$

$$\bar{y}^k = \text{Argmin}\{\theta_2(y) + \frac{\beta}{2}\|(A\bar{x}^k + y + z^k - b) - \frac{1}{\beta}\lambda^k\|^2 \mid y \in \mathcal{Y}\}, \tag{3.1b}$$

$$\bar{z}^k = \text{Argmin}\{\theta_3(z) + \frac{\beta}{2}\|(A\bar{x}^k + \bar{y}^k + z - b) - \frac{1}{\beta}\lambda^k\|^2 \mid z \in \mathcal{Z}\}, \tag{3.1c}$$

$$\bar{\lambda}^k = \lambda^k - \beta(A\bar{x}^k + \bar{y}^k + \bar{z}^k - b). \tag{3.1d}$$

For the new iterative loop, we need only to produce $v^{k+1} = (y^{k+1}, z^{k+1}, \lambda^{k+1})$. Use the notation of $\bar{w}^k$, the update form (1.5) can be written as

The correction step: Update the new iterate $v^{k+1} = (y^{k+1}, z^{k+1}, \lambda^{k+1})$ by

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ z^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I_m & -I_m & 0 \\ 0 & I_m & 0 \\ 0 & 0 & I_m \end{pmatrix} \begin{pmatrix} y^k - \bar{y}^k \\ z^k - \bar{z}^k \\ \lambda^k - \bar{\lambda}^k \end{pmatrix}. \tag{3.2}$$

Using the notation $P^{-T}D$, the correction step can be written as

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

We consider the general correction update form

$$v^{k+1} = v^k - \alpha P^{-T}D(v^k - \bar{v}^k), \quad \alpha \in (0, 1]. \tag{3.3}$$

In other words, the update form (3.2) is a special case of (3.3) with $\alpha = 1$. Taking $\alpha \in (0, 1)$, the method is a special case of the method proposed in [10].

3.1 Properties of the vector $\bar{w}^k$ by the prediction step

We establish the following lemma.

Lemma 3.1 *Let* $\bar{w}^k = (\bar{x}^k, \bar{y}^k, \bar{z}^k, \bar{\lambda}^k)$ *be generated by the alternating direction-prediction step* (3.1a)–(3.1d) *from the given vector* $v^k = (y^k, z^k, \lambda^k)$. *Then, we have* $\bar{w}^k \in \mathcal{W}$ *and*

$$\theta(u) - \theta(\bar{u}^k) + (w - \bar{w}^k)^T d(v^k, \bar{w}^k) \geq (v - \bar{v}^k)^T P(v^k - \bar{v}^k), \ \forall\, w \in \mathcal{W}, \quad (3.4)$$

where

$$d(v^k, \bar{w}^k) = F(\bar{w}^k) + \eta(v^k, \bar{v}^k), \quad (3.5)$$

$$\eta(v^k, \bar{v}^k) = \begin{pmatrix} A^T \\ I_m \\ I_m \\ 0 \end{pmatrix} \beta\big((y^k - \bar{y}^k) + (z^k - \bar{z}^k)\big), \quad (3.6)$$

and the matrix P *is defined in (2.7).*

Proof. The proof consists of some manipulations. Recall (3.1a)–(3.1c). We have $\bar{u}^k \in \mathcal{U}$ and

$$\begin{cases} \theta_1(x) - \theta_1(\bar{x}^k) + (x - \bar{x}^k)^T (A^T[\beta(A\bar{x}^k + y^k + z^k - b) - \lambda^k]) \geq 0, \\ \theta_2(y) - \theta_2(\bar{y}^k) + (y - \bar{y}^k)^T (\beta(A\bar{x}^k + \bar{y}^k + z^k - b) - \lambda^k) \geq 0, \\ \theta_3(z) - \theta_3(\bar{z}^k) + (z - \bar{z}^k)^T (\beta(A\bar{x}^k + \bar{y}^k + \bar{z}^k - b) - \lambda^k) \geq 0, \end{cases}$$

for all $u \in \mathcal{U}$. Using (3.1d), the above inequality can be written as

$$\theta(u) - \theta(\bar{u}^k) + \begin{pmatrix} x - \bar{x}^k \\ y - \bar{y}^k \\ z - \bar{z}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\bar{\lambda}^k \\ -\bar{\lambda}^k \\ -\bar{\lambda}^k \end{pmatrix} + \begin{pmatrix} \beta A^T(y^k - \bar{y}^k) + \beta A^T(z^k - \bar{z}^k) \\ \beta(z^k - \bar{z}^k) \\ 0 \end{pmatrix} \right\} \geq 0. \quad (3.7)$$

Adding the following term

$$\begin{pmatrix} x - \bar{x}^k \\ y - \bar{y}^k \\ z - \bar{z}^k \end{pmatrix}^T \begin{pmatrix} 0 \\ \beta(y^k - \bar{y}^k) \\ \beta(y^k - \bar{y}^k) + \beta(z^k - \bar{z}^k) \end{pmatrix}$$

to the both sides of the last inequality, we get $\bar{w}^k \in \mathcal{W}$ and

$$\theta(u)-\theta(\bar{u}^k)+\begin{pmatrix} x-\bar{x}^k \\ y-\bar{y}^k \\ z-\bar{z}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\bar{\lambda}^k \\ -\bar{\lambda}^k \\ -\bar{\lambda}^k \end{pmatrix} + \begin{pmatrix} A^T\beta\big((y^k-\bar{y}^k)+(z^k-\bar{z}^k)\big) \\ \beta\big((y^k-\bar{y}^k)+(z^k-\bar{z}^k)\big) \\ \beta\big((y^k-\bar{y}^k)+(z^k-\bar{z}^k)\big) \end{pmatrix} \right\}$$

$$\geq \begin{pmatrix} x-\bar{x}^k \\ y-\bar{y}^k \\ z-\bar{z}^k \end{pmatrix}^T \begin{pmatrix} 0 \\ \beta(y^k-\bar{y}^k) \\ \beta(y^k-\bar{y}^k)+\beta(z^k-\bar{z}^k) \end{pmatrix}, \quad \forall\, u \in \mathcal{U}. \quad (3.8)$$

Because

$$A\bar{x}^k+\bar{y}^k+\bar{z}^k-b=\frac{1}{\beta}(\lambda^k-\bar{\lambda}^k),$$

adding the equal terms

$$(\lambda-\bar{\lambda}^k)^T(A\bar{x}^k+\bar{y}^k+\bar{z}^k-b) \quad \text{and} \quad (\lambda-\bar{\lambda}^k)^T\frac{1}{\beta}(\lambda^k-\bar{\lambda}^k)$$

to the left side and right side of (3.8), respectively, we get $\bar{w}^k \in \mathcal{W}$ and

$$\theta(u)-\theta(\bar{u}^k)+\begin{pmatrix} x-\bar{x}^k \\ y-\bar{y}^k \\ z-\bar{z}^k \\ \lambda-\bar{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T\bar{\lambda}^k \\ -\bar{\lambda}^k \\ -\bar{\lambda}^k \\ A\bar{x}^k+\bar{y}^k+\bar{z}^k-b \end{pmatrix} + \begin{pmatrix} A^T\beta\big((y^k-\bar{y}^k)+(z^k-\bar{z}^k)\big) \\ \beta\big((y^k-\bar{y}^k)+(z^k-\bar{z}^k)\big) \\ \beta\big((y^k-\bar{y}^k)+(z^k-\bar{z}^k)\big) \\ 0 \end{pmatrix} \right\}$$

$$\geq \begin{pmatrix} y-\bar{y}^k \\ z-\bar{z}^k \\ \lambda-\bar{\lambda}^k \end{pmatrix}^T \begin{pmatrix} \beta(y^k-\bar{y}^k) \\ \beta(y^k-\bar{y}^k)+\beta(z^k-\bar{z}^k) \\ \frac{1}{\beta}(\lambda^k-\bar{\lambda}^k) \end{pmatrix}, \quad \forall\, w \in \mathcal{W}.$$

Using the notations of $F(w)$, $d(v^k,\bar{w}^k)$ and P, the assertion follows immediately and the lemma is proved. $\square$

Lemma 3.2 *Let $\bar{w}^k=(\bar{x}^k,\bar{y}^k,\bar{z}^k,\bar{\lambda}^k)$ be generated by the alternating direction prediction step (3.1a)–(3.1d) from the given vector $v^k=(y^k,z^k,\lambda^k)$. Then for all*

$v^* \in \mathcal{V}^*$, *we have*

$$\begin{aligned}&(v^k - v^*)^T P(v^k - \bar{v}^k)\\ &\quad\geq \quad \frac{1}{2}\|v^k - \bar{v}^k\|_D^2 + \frac{1}{2}\beta\|(y^k - \bar{y}^k) + (z^k - \bar{z}^k) + \frac{1}{\beta}(\lambda^k - \bar{\lambda}^k)\|^2, \end{aligned} \tag{3.9}$$

where the matrices P is defined in (2.7).

Proof. Setting $w = w^*$ in (3.4) (notice that v^* is a sub-vector of w^*), it yields that

$$\begin{aligned}(\bar{v}^k - v^*)^T P(v^k - \bar{v}^k) \quad \geq \quad & \big(\theta(\bar{u}^k) - \theta(u^*)\big) + (\bar{w}^k - w^*)^T F(\bar{w}^k)\\ & +(\bar{w}^k - w^*)^T \eta(v^k, \bar{v}^k). \end{aligned} \tag{3.10}$$

Now, we deal with the last term in the right hand side of the inequality (3.10). By using the notation of $\eta(v^k, \bar{v}^k)$, $Ax^* + y^* + z^* = b$ and (3.1d), we obtain

$$\begin{aligned}&(\bar{w}^k - w^*)^T \eta(v^k, \bar{v}^k)\\ &\quad = \quad \beta(A\bar{x}^k + \bar{y}^k + \bar{z}^k - Ax^* - y^* - z^*)^T\{(y^k - \bar{y}^k) + (z^k - \bar{z}^k)\}\\ &\quad = \quad \beta(A\bar{x}^k + \bar{y}^k + \bar{z}^k - b)^T\{(y^k - \bar{y}^k) + (z^k - \bar{z}^k)\}\\ &\quad = \quad (\lambda^k - \bar{\lambda}^k)^T\{(y^k - \bar{y}^k) + (z^k - \bar{z}^k)\}.\end{aligned}$$

Substituting it in (3.10), we get

$$\begin{aligned}(\bar{v}^k - v^*)^T P(v^k - \bar{v}^k) \;\geq\; & \big(\theta(\bar{u}^k) - \theta(u^*)\big) + (\bar{w}^k - w^*)^T F(\bar{w}^k)\\ & +(\lambda^k - \bar{\lambda}^k)^T\{(y^k - \bar{y}^k) + (z^k - \bar{z}^k)\}. \end{aligned} \tag{3.11}$$

Since F is monotone, we have

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

Substituting it in the right hand side of (3.11), we obtain

$$(\bar{v}^k - v^*)^T P(v^k - \bar{v}^k) \geq (\lambda^k - \bar{\lambda}^k)^T\{(y^k - \bar{y}^k) + (z^k - \bar{z}^k)\}.$$

It follows from the last equality that

$$\begin{aligned}&(v^k - v^*)^T P(v^k - \bar{v}^k)\\ &\quad\geq (v^k - \bar{v}^k)^T P(v^k - \bar{v}^k) + (\lambda^k - \bar{\lambda}^k)^T\{(y^k - \bar{y}^k) + (z^k - \bar{z}^k)\}. \end{aligned} \tag{3.12}$$

Observe the matrices P and D (see (2.7)), by a manipulation, the right hand side of (3.12) can be written as

$$
\begin{aligned}
&(v^k - \bar{v}^k)^T P(v^k - \bar{v}^k) + (\lambda^k - \bar{\lambda}^k)^T\{(y^k - \bar{y}^k) + (z^k - \bar{z}^k)\} \\
&= \begin{pmatrix} y^k - \bar{y}^k \\ z^k - \bar{z}^k \\ \lambda^k - \bar{\lambda}^k \end{pmatrix}^T \begin{pmatrix} \beta I_m & \frac{1}{2}\beta I_m & \frac{1}{2} I_m \\ \frac{1}{2}\beta I_m & \beta I_m & \frac{1}{2} I_m \\ \frac{1}{2} I_m & \frac{1}{2} I_m & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} y^k - \bar{y}^k \\ z^k - \bar{z}^k \\ \lambda^k - \bar{\lambda}^k \end{pmatrix} \\
&= \frac{1}{2}\|v^k - \bar{v}^k\|_D^2 + \frac{1}{2}\beta\|(y^k - \bar{y}^k) + (z^k - \bar{z}^k) + \frac{1}{\beta}(\lambda^k - \bar{\lambda}^k)\|^2.
\end{aligned}
$$

Substituting it in the right hand side of (3.12), the assertion of this lemma is proved. □

Whenever $v^k \neq \bar{v}^k$, the right hand side of (3.9) is positive. For any positive definite matrix H, (3.9) implies that

$$
\begin{aligned}
&\langle H(v^k - v^*), H^{-1}P(v^k - \bar{v}^k)\rangle \\
&\geq \frac{1}{2}\|v^k - \bar{v}^k\|_D^2 + \frac{1}{2}\beta\|(y^k - \bar{y}^k) + (z^k - \bar{z}^k) + \frac{1}{\beta}(\lambda^k - \bar{\lambda}^k)\|^2,
\end{aligned}
$$

and $H^{-1}P(v^k - \bar{v}^k)$ is an ascent direction of the distance function $\frac{1}{2}\|v - v^*\|_H^2$ at the point v^k. By choosing

$$
H = PD^{-1}P^T = \begin{pmatrix} \beta I_m & \beta I_m & 0 \\ \beta I_m & 2\beta I_m & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}, \tag{3.13}
$$

and using the matrix $P^{-T}D$ (see (2.8)), we have

$$
H^{-1}P = P^{-T}D = \begin{pmatrix} I_m & -I_m & 0 \\ 0 & I_m & 0 \\ 0 & 0 & I_m \end{pmatrix}.
$$

3.2 Correction and the contractive property

The alternating direction-prediction step begins with given $v^k = (y^k, z^k, \lambda^k)$. In fact x is only the auxiliary variable in the iteration process. For the new loop, we need only to produce $v^{k+1} = (y^{k+1}, z^{k+1}, \lambda^{k+1})$ and we call this step as the **correction step**.

Theorem 3.1 *Let $\bar{w}^k = (\bar{x}^k, \bar{y}^k, \bar{z}^k, \bar{\lambda}^k)$ be generated by the alternating*

direction-prediction step (3.1a)–(3.1d) *from the given vector* $v^k = (y^k, z^k, \lambda^k)$ *and* v^{k+1} *be given by* (3.3)*. Then we have*

$$\begin{aligned} \|v^{k+1} - v^*\|_H^2 \quad &\leq \quad \|v^k - v^*\|_H^2 - \alpha(1-\alpha)\|v^k - \bar{v}^k\|_D^2 \\ &-\alpha\beta\|(y^k - \bar{y}^k) + (z^k - \bar{z}^k) + \frac{1}{\beta}(\lambda^k - \bar{\lambda}^k)\|^2, \ \ \forall v^* \in \mathcal{V}^*, \end{aligned} \tag{3.14}$$

where H *is defined in* (3.13)*.*

Proof. By using (3.3) and the definition of the matrix H, we have

$$\begin{aligned} &\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\ &\quad = \quad \|v^k - v^*\|_H^2 - \|(v^k - v^*) - (v^k - v^{k+1})\|_H^2 \\ &\quad = \quad \|v^k - v^*\|_H^2 - \|(v^k - v^*) - \alpha P^{-T} D(v^k - \bar{v}^k)\|_H^2 \\ &\quad = \quad 2\alpha(v^k - v^*)^T P(v^k - \bar{v}^k) - \alpha^2\|v^k - \bar{v}^k\|_D^2. \end{aligned}$$

Substituting (3.9) in the last equality, we get

$$\begin{aligned} &\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\ &\quad \geq \quad \alpha(1-\alpha)\|v^k - \bar{v}^k\|_D^2 + \alpha\beta\|(y^k - \bar{y}^k) + (z^k - \bar{z}^k) + \frac{1}{\beta}(\lambda^k - \bar{\lambda}^k)\|^2. \end{aligned}$$

The assertion of this theorem is proved. $\square$

For any $\alpha \in (0,1)$, it follows (3.14) that

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \alpha(1-\alpha)\|v^k - \bar{v}^k\|_D^2, \quad \forall\, v^* \in \mathcal{V}^*. \tag{3.15}$$

The above inequality is essential for the global convergence of the method using update form (3.3) with $\alpha \in (0,1)$, see [10]. In fact, based on (3.15), the sequence $\{v^k\}$ is bounded, we have

$$\lim_{k\to\infty} \|v^k - \bar{v}^k\|_D^2 = 0. \tag{3.16}$$

Consequently the sequence $\{\bar{v}^k\}$ is also bounded and it converges to a limit point v^∞. On the other hand, due to Lemma 3.1, we have

$$\begin{aligned} &\theta(u) - \theta(\bar{u}^k) + (w - \bar{w}^k)^T F(\bar{w}^k) \\ &\quad \geq \quad -(w - \bar{w}^k)^T \eta(v^k, \bar{v}^k) + (v - \bar{v}^k)^T P(v^k - \bar{v}^k), \ \ \forall\, w \in \mathcal{W}. \end{aligned} \tag{3.17}$$

From (3.16) and (3.17) we can derive $v^\infty \in \mathcal{V}^*$ and the induced w^∞ is a solution of VI$(\mathcal{W}, F, \theta)$.

4 Convergence rate in an ergodic sense

The convergence in the last section is only for $\alpha \in (0,1)$ with update form (3.3). Since (1.5), namely, the little changed alternating direction method of multipliers, is equivalent to (3.2) with $\alpha = 1$, it follows from Theorem 3.1 that

$$\|v^{k+1}-v^*\|_H^2 \le \|v^k-v^*\|_H^2-\beta\|(y^k-\bar{y}^k)+(z^k-\bar{z}^k)+\frac{1}{\beta}(\lambda^k-\bar{\lambda}^k)\|^2, \ \forall v^* \in \mathcal{V}^*.$$

The sequence $\{\|v^k - v^*\|_H\}$ is monotonically non-increasing. However, we have not obtained the global convergence in the contraction sense. This section, however, shows that the method using update form (3.2) has convergence rate $O(1/t)$ for all $\alpha \in (0,1]$ in an ergodic sense.

According to Theorem 2.1, for given $\epsilon > 0$ and a substantial compact set $\mathcal{D} \subset \mathcal{W}$, our task is to find a $\tilde{w}$ such that (see (2.6))

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

in $O(1/\epsilon)$ iterations. Generally, our complexity analysis follows the line of [3, 13], but instead of using $\bar{w}^k$ directly, we need first to introduce an auxiliary vector.

The additional auxiliary vector

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{z}^k \\ \tilde{\lambda}^k \end{pmatrix}, \quad \text{where} \quad \tilde{u}^k = \bar{u}^k \tag{4.1a}$$

and

$$\tilde{\lambda}^k = \bar{\lambda}^k - \beta\big((y^k - \bar{y}^k) + (z^k - \bar{z}^k)\big). \tag{4.1b}$$

In order to rewrite the assertion in Lemma 3.1 in form of $\tilde{w}$, we need the following lemmas to express the terms $d(v^k, \bar{w}^k)$ and $P(v^k - \bar{v}^k)$ in form of w^k and $\tilde{w}^k$.

Lemma 4.1 *For the $\tilde{w}^k$ defined in* (4.1) *and the $\bar{w}^k$ generated by* (3.1)*, we have*

$$d(v^k, \bar{w}^k) = F(\tilde{w}^k), \tag{4.2}$$

where $d(v^k, \bar{w}^k)$ defined in (3.5)*. In addition, it holds that*

$$P(v^k - \bar{v}^k) = Q(v^k - \tilde{v}^k), \tag{4.3}$$

where

$$Q = \begin{pmatrix} \beta I_m & 0 & 0 \\ \beta I_m & \beta I_m & 0 \\ -I_m & -I_m & \frac{1}{\beta} I_m \end{pmatrix}. \tag{4.4}$$

Proof. Since $\tilde{u}^k = \bar{u}^k$ and $\tilde{\lambda}^k = \bar{\lambda}^k - \beta\big((y^k - \bar{y}^k) + (z^k - \bar{z}^k)\big)$, we have

$$\beta\big((y^k - \bar{y}^k) + (z^k - \bar{z}^k)\big) = \bar{\lambda}^k - \tilde{\lambda}^k. \tag{4.5}$$

Substituting it in the notation of $\eta(v^k, \bar{v}^k)$ (see (3.6)), we get

$$\begin{aligned} d(v^k, \bar{w}^k) &= F(\bar{w}^k) + \eta(v^k, \bar{v}^k) \\ &= \begin{pmatrix} -A^T\bar{\lambda}^k \\ -\bar{\lambda}^k \\ -\bar{\lambda}^k \\ A\bar{x}^k + \bar{y}^k + \bar{z}^k - b \end{pmatrix} + \begin{pmatrix} A^T \\ I \\ I \\ 0 \end{pmatrix} (\bar{\lambda}^k - \tilde{\lambda}^k) \;=\; F(\tilde{w}^k). \end{aligned}$$

The last equation is due to $(\bar{x}^k, \bar{y}^k, \bar{z}^k) = (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k)$. The equation (4.1b) implies

$$\begin{aligned} \frac{1}{\beta}(\lambda^k - \bar{\lambda}^k) &= \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) - \big((y^k - \bar{y}^k) + (z^k - \bar{z}^k)\big) \\ &= \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) - \big((y^k - \tilde{y}^k) + (z^k - \tilde{z}^k)\big). \end{aligned}$$

Using the matrix P (see (2.7)), (4.1a) and the last equation, we obtain

$$\begin{aligned} P(v^k - \bar{v}^k) &= \begin{pmatrix} \beta I_m & 0 & 0 \\ \beta I_m & \beta I_m & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} y^k - \bar{y}^k \\ z^k - \bar{z}^k \\ \lambda^k - \bar{\lambda}^k \end{pmatrix} \\ &= \begin{pmatrix} \beta I_m & 0 & 0 \\ \beta I_m & \beta I_m & 0 \\ -I_m & -I_m & \frac{1}{\beta} I_m \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ z^k - \tilde{z}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix} = Q(v^k - \tilde{v}^k). \end{aligned}$$

Thus (4.3) holds and the lemma is proved. $\square$

By using (4.2), the assertion in Lemma 3.1 can be rewritten accordingly in the following lemma.

Lemma 4.2 *Let* $\bar{w}^k = (\bar{x}^k, \bar{y}^k, \bar{z}^k, \bar{\lambda}^k)$ *be generated by the alternating direction-prediction step* (3.1a)–(3.1d) *from the given vector* $v^k = (y^k, z^k, \lambda^k)$ *and* $\tilde{w}^k$ *be defined by* (4.1)*. Then we have* $\tilde{w}^k \in \mathcal{W}$ *and*

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T P(v^k - \bar{v}^k), \ \forall w \in \mathcal{W}. \quad (4.6)$$

Proof. The assertion follows from (3.4), (4.2) and (4.1). □

Now, we are ready to prove the key inequalities for the convergence rate of the proposed method, which are given in the following lemmas.

Lemma 4.3 *Let* $\bar{w}^k = (\bar{x}^k, \bar{y}^k, \bar{z}^k, \bar{\lambda}^k)$ *be generated by the alternating direction-prediction step* (3.1a)–(3.1d) *from the given vector* $v^k = (y^k, z^k, \lambda^k)$ *and* $\tilde{w}^k$ *be defined by* (4.1)*. If the new iterate* v^{k+1} *is updated by* (3.3)*, then we have*

$$\begin{aligned} &\big(\theta(u) - \theta(\tilde{u}^k)\big) + (w - \tilde{w}^k)^T F(\tilde{w}^k) + \frac{1}{2}(\|v - v^k\|_H^2 - \|v - v^{k+1}\|_H^2) \\ &\quad \ge \quad \frac{1}{2}(\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2), \quad \forall w \in \mathcal{W}, \end{aligned} \quad (4.7)$$

where matrix H *is defined in* (3.13)*.*

Proof. According to the update form (3.3), we have

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

Substituting it into the right hand side of (4.6) and using $PD^{-1}P^T = H$, we obtain

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T H(v^k - v^{k+1}), \ \ \forall\, w \in \mathcal{W}. \quad (4.8)$$

By setting

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

in the identity

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

we obtain

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

Substituting it in the right hand side of (4.8) and by a manipulation, we get (4.7) and the lemma is proved. □

Lemma 4.4 *Let* $\bar{w}^k = (\bar{x}^k, \bar{y}^k, \bar{z}^k, \bar{\lambda}^k)$ *be generated by the alternating direction prediction step* (3.1a)–(3.1d) *from the given vector* $v^k = (y^k, z^k, \lambda^k)$ *and* $\tilde{w}^k$ *be defined by* (4.1)*. If the new iterate* v^{k+1} *is updated by* (3.3)*, then we have*

$$\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 = \frac{\alpha}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2 + \alpha(1-\alpha)\|v^k - \bar{v}^k\|_D^2. \quad (4.9)$$

where matrix H *is defined in* (3.13)*.*

Proof. In view of the update form (3.3) and $H = PD^{-1}P^T$, we have

$$\begin{aligned}
&\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\
&\quad = \|v^k - \tilde{v}^k\|_H^2 - \|v^k - \tilde{v}^k + (v^{k+1} - v^k)\|_H^2 \\
&\quad = \|v^k - \tilde{v}^k\|_H^2 - \|v^k - \tilde{v}^k - \alpha P^{-T} D(v^k - \bar{v}^k)\|_H^2 \\
&\quad = 2\alpha(v^k - \tilde{v}^k)^T P(v^k - \bar{v}^k) - \alpha^2\|P^{-T} D(v^k - \bar{v}^k)\|_H^2 \\
&\quad = 2\alpha(v^k - \tilde{v}^k)^T P(v^k - \bar{v}^k) - \alpha^2\|v^k - \bar{v}^k\|_D^2. \qquad (4.10)
\end{aligned}$$

By using the relation in the equation (4.3) and the matrix Q (see (4.4)), we obtain

$$\begin{aligned}
&2(v^k - \tilde{v}^k)^T P(v^k - \bar{v}^k) \\
&\quad = \quad 2(v^k - \tilde{v}^k)^T Q(v^k - \tilde{v}^k) \\
&\quad = \quad (v^k - \tilde{v}^k)^T (Q + Q^T)(v^k - \tilde{v}^k) \\
&\quad = \quad \begin{pmatrix} y^k - \tilde{y}^k \\ z^k - \tilde{z}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} 2\beta I_m & \beta I_m & -I_m \\ \beta I_m & 2\beta I_m & -I_m \\ -I_m & -I_m & \frac{2}{\beta} I_m \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ z^k - \tilde{z}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix} \\
&\quad = \quad \beta\|(y^k - \tilde{y}^k) + (z^k - \tilde{z}^k) - \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k)\|^2 + \|v^k - \tilde{v}^k\|_D^2. \qquad (4.11)
\end{aligned}$$

Because $\tilde{y}^k = \bar{y}^k$ and $\tilde{z}^k = \bar{z}^k$, it follows from and (4.5) that

$$\beta\|(y^k - \tilde{y}^k) + (z^k - \tilde{z}^k) - \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k)\|^2 = \frac{1}{\beta}\|\lambda^k - \bar{\lambda}^k\|^2. \qquad (4.12)$$

Combining (4.10), (4.11) and (4.12) together, we obtain

$$\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 = \alpha\Big(\frac{1}{\beta}\|\lambda^k - \bar{\lambda}^k\|^2 + \|v^k - \tilde{v}^k\|_D^2\Big) - \alpha^2\|v^k - \bar{v}^k\|_D^2. \qquad (4.13)$$

Again, because $\tilde{y}^k = \bar{y}^k$ and $\tilde{z}^k = \bar{z}^k$, we have

$$\frac{1}{\beta}\|\lambda^k - \bar{\lambda}^k\|^2 + \|v^k - \tilde{v}^k\|_D^2 = \frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2 + \|v^k - \bar{v}^k\|_D^2.$$

Substituting it in the right hand side of (4.13), we get

$$\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 = \frac{\alpha}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2 + \alpha(1-\alpha)\|v^k - \bar{v}^k\|_D^2,$$

and thus the proof is complete. $\square$

Combining the assertions in Lemma 4.3 and Lemma 4.4, we have proved the key inequality for the proposed method, namely,

$$\big(\theta(u)-\theta(\tilde{u}^k)\big)+(w-\tilde{w}^k)^T F(\tilde{w}^k)+\frac{1}{2}(\|v-v^k\|_H^2-\|v-v^{k+1}\|_H^2) \ge 0, \ \forall w \in \mathcal{W}. \tag{4.14}$$

Note that the above inequality is true for any $\alpha \in (0,1]$. Having the key inequalities in the above lemmas, the $\mathcal{O}(1/t)$ rate of convergence (in an ergodic sense) can be obtained easily.

Theorem 4.1 *For any integer $t > 0$, we have a $\tilde{w}_t \in \mathcal{W}$ which satisfies*

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

where

$$\tilde{w}_t = \frac{1}{t+1}\sum_{k=0}^{t} \tilde{w}^k.$$

Proof. Since $F(\cdot)$ is monotone, it follows (4.14) that

$$\big(\theta(u)-\theta(\tilde{u}^k)\big)+(w-\tilde{w}^k)^T F(w)+\frac{1}{2}(\|v-v^k\|_H^2-\|v-v^{k+1}\|_H^2) \ge 0, \ \forall w \in \mathcal{W},$$

or, equivalently,

$$\big(\theta(\tilde{u}^k)-\theta(u)\big)+(\tilde{w}^k-w)^T F(w)+\frac{1}{2}(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2) \le 0, \ \forall w \in \mathcal{W}.$$

Summing the above inequality over $k = 0, \ldots, t$, we obtain

$$\sum_{k=0}^{t}\big(\theta(\tilde{u}^k)-\theta(u)\big)+\Big(\sum_{k=0}^{t}\tilde{w}^k-\sum_{k=0}^{t}w\Big)^T F(w)+\frac{1}{2}(\|v-v^{t+1}\|_H^2-\|v-v^0\|_H^2) \le 0.$$

Dropping the term $\|v^{t+1} - v^0\|_G^2$, we get

$$\Big(\frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k)-\theta(u)\Big)+\Big(\frac{1}{t+1}\sum_{k=0}^{t}\tilde{w}^k-w\Big)^T F(w) \le \frac{\|v-v^0\|_H^2}{2(t+1)}, \quad \forall w \in \mathcal{W}. \tag{4.15}$$

By incorporating the notation of $\tilde{w}_t$ and using

$$\theta(\tilde{u}_t^k) \le \frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k) \qquad \text{(due to the convexity of } \theta(u)\text{)}$$

it follows from (4.15) that

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

Hence, the proof is complete. $\square$

For given substantial compact set $\mathcal{D} \subset \mathcal{W}$, we define

$$d = \sup\{\|v - v^0\|_H \,|\, w \in \mathcal{D}\},$$

where $v^0 = (y^0, z^0, \lambda^0)$ is the initial point. Because $\varrho_k \ge \frac{1}{2}$, it follows that $\Upsilon_t \ge \frac{t+1}{2}$.

After t iterations of the proposed method, we can find a $\tilde{w} \in \mathcal{W}$ such that

$$\sup_{w\in\mathcal{D}}\big\{\theta(\tilde{u}) - \theta(u)\big) + (\tilde{w} - w)^T F(w)\big\} \le \frac{d^2}{2(t+1)}.$$

The convergence rate $\mathcal{O}(1/t)$ of the proposed method is thus proved.

5 Convergence rate in the non-ergodic sense

If we use (3.3) with $\alpha \in (0,1)$ to update the new iterate, it follows from (3.15) that

$$\sum_{k=0}^{\infty}\|v^k - \bar{v}^k\|_D^2 \le \frac{1}{\alpha(1-\alpha)}\|v^0 - v^*\|_H^2 \quad \forall\, v^* \in \mathcal{V}^*. \tag{5.1}$$

This section will show that the sequence $\{\|v^k - v^{k+1}\|_D^2\}$ is monotonically non-increasing, *i. e.*,

$$\|v^{k+1} - \bar{v}^{k+1}\|_D^2 \le \|v^k - \bar{v}^k\|_D^2, \quad \forall\, k \ge 0. \tag{5.2}$$

Based on (5.1) and (5.2), we drive

$$\|v^k - \bar{v}^k\|_D^2 \le \frac{1}{(k+1)\alpha(1-\alpha)}\|v^0 - v^*\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{5.3}$$

Since $\|v^k - \bar{v}^k\|_D^2$ is viewed as the stopping criterion, we obtain the worst-case $O(1/t)$ convergence rate in a non-ergodic sense. An important relation in the coming proof is (see (3.3))

$$P^T(v^k - v^{k+1}) = \alpha D(v^k - \bar{v}^k), \tag{5.4}$$

where the matrices P and D are given in (2.7). Lemma 4.2 enables us to establish an important inequality in the following lemma.

Lemma 5.1 *Let $\{v^k\}$ be the sequence generated by* (3.3)*, the associated sequence $\{\tilde{w}^k\}$ be defined by (4.1). Then, we have*

$$(\tilde{v}^k - \tilde{v}^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \ge 0, \tag{5.5}$$

where Q is given defined in (4.4).

Proof. Set $w = \tilde{w}^{k+1}$ in (4.6) and use (4.3), we obtain

$$\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T F(\tilde{w}^k) \ge (\tilde{v}^{k+1} - \tilde{v}^k)^T Q(v^k - \tilde{v}^k). \tag{5.6}$$

Note that (4.6) is also true for $k := k+1$ and thus we have

$$\theta(u) - \theta(\tilde{u}^{k+1}) + (w - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \ge (v - \tilde{v}^{k+1})^T P(v^{k+1} - \bar{v}^{k+1}),$$

for all $w \in \Omega$. Set $w = \tilde{w}^k$ in the above inequality and use (4.3), we get

$$\theta(\tilde{u}^k) - \theta(\tilde{u}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T F(\bar{w}^{k+1}) \ge (\tilde{v}^k - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}).$$

Adding this inequality with (5.6), we get assertion (5.5) due to the monotonicity of F. $\square$

Lemma 5.2 *Let the sequence $\{\bar{w}^k\}$ be generated by* (3.1)*, and $\{v^k\}$ be the sequence updated by* (3.3)*. Then, we have*

$$\begin{aligned} &\alpha(v^k - \bar{v}^k)^T D\{(v^k - \bar{v}^k) - (v^{k+1} - \bar{v}^{k+1})\} \\ &\quad \ge \quad \frac{1}{2}\|P(v^k - \bar{v}^k) - P(v^{k+1} - \bar{v}^{k+1})\|^2_{(Q^{-T}+Q^{-1})}, \end{aligned} \tag{5.7}$$

where the matrices P and D are given in (2.7) and Q is given defined in (4.4).

Proof. Adding the term

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

to the both sides of (5.5), we get

$$(v^k - v^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq \frac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|^2_{(Q^T+Q)}. \tag{5.8}$$

By using (5.4) and (4.3), we have

$$(v^k - v^{k+1})^T = \alpha(v^k - \bar{v}^k)^T DP^{-1},$$

and

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

Substituting them in (5.8), we obtain

$$\begin{aligned} &\alpha(v^k - \bar{v}^k)^T D\{(v^k - \bar{v}^k) - (v^{k+1} - \bar{v}^{k+1})\} \\ &\geq \quad \frac{1}{2}\|Q^{-1}P(v^k - \bar{v}^k) - Q^{-1}P(v^{k+1} - \bar{v}^{k+1})\|^2_{(Q^T+Q)}. \end{aligned} \tag{5.9}$$

Because

$$Q^{-T}(Q^T + Q)Q^{-1} = Q^{-T} + Q^{-1},$$

the both right hand sides of (5.7) and (5.9) are equal. The assertion is proved. □

Finally, we are ready to show the assertion (5.2) in the following theorem.

Theorem 5.1 *Let* $\{\bar{v}^k\}$ (3.1), $\{v^k\}$ *be the sequence generated by* (3.3)*. The sequence* $\{\|v^k - v^{k+1}\|^2_D\}$ *is monotonically non-increasing.*

Proof. Setting $a = (v^k - \bar{v}^k)$ and $b = (v^{k+1} - \bar{v}^{k+1})$ in the identity

$$\|a\|^2_D - \|b\|^2_D = 2a^T D(a - b) - \|a - b\|^2_D,$$

we obtain

$$\begin{aligned} &\|v^k - \bar{v}^k\|^2_D - \|v^{k+1} - \bar{v}^{k+1}\|^2_D \\ &\quad = 2(v^k - \bar{v}^k)^T D\{(v^k - \bar{v}^k) - (v^{k+1} - \bar{v}^{k+1})\} - \|(v^k - \bar{v}^k) - (v^{k+1} - \bar{v}^{k+1})\|^2_D. \end{aligned}$$

Inserting (5.7) into the first term of the right-hand side of the last equality, we obtain

$$\begin{aligned} &\|v^k - \bar{v}^k\|^2_D - \|v^{k+1} - \bar{v}^{k+1}\|^2_D \\ &\quad \geq \quad \frac{1}{\alpha}\|P(v^k - \bar{v}^k) - P(v^{k+1} - \bar{v}^{k+1})\|^2_{(Q^{-T}+Q^{-1})} \\ &\qquad\quad -\|P(v^k - \bar{v}^k) - P(v^{k+1} - \bar{v}^{k+1})\|^2_{(P^{-T}DP^{-1})} \\ &\quad \geq \quad \|P(v^k - \bar{v}^k) - P(v^{k+1} - \bar{v}^{k+1})\|^2_{\{(Q^{-T}+Q^{-1})-(P^{-T}DP^{-1})\}}. \end{aligned}$$

Notice that (see (4.4) and (2.7))

$$(Q^{-T}+Q^{-1})-(P^{-T}DP^{-1})=\begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{\beta}I_m & I_m \\ 0 & I_m & \beta I_m \end{pmatrix}$$

is a positive semidefinite matrix, the assertion (5.2) follows immediately and the lemma is proved. $\square$

With (5.1) and (5.2), we derived (5.3). The worst-case $O(1/t)$ convergence rate in a non-ergodic sense for the proposed method with $\alpha \in (0,1)$ is proved.

By using (5.4), we can also use $\|v^k - v^{k+1}\|_H \le \epsilon$ as the stop criterion. Since

$$H = PD^{-1}P^T,$$

we have

$$\|v^k - \bar{v}^k\|_D^2 = \frac{1}{\alpha^2}\|D^{-1}P^T(v^k - v^{k+1})\|_D^2 = \frac{1}{\alpha^2}\|v^k - v^{k+1}\|_H^2.$$

Therefore, the assertions (5.1) and (5.3) can be rewritten as

$$\sum_{k=0}^{\infty}\|v^k - v^{k+1}\|_H^2 \le \frac{\alpha}{(1-\alpha)}\|v^0 - v^*\|_H^2 \quad \forall\, v^* \in \mathcal{V}^*, \tag{5.10}$$

and

$$\|v^k - v^{k+1}\|_H^2 \le \frac{\alpha}{(k+1)(1-\alpha)}\|v^0 - v^*\|_H^2, \quad \forall v^* \in \mathcal{V}^*, \tag{5.11}$$

respectively.

6 Conclusions

Because of the attractive efficiency of the well-known alternating direction method (ADM), it is of strong desire to extend the ADMM to the linearly constrained convex programming problem with three separable operators. The convergence of the direct extension of the ADMM to the problem with 3 separable parts, however, is still open. The method proposed in this lecture (with update form (3.2)) is convergent and its variety to the direct extension of ADMM is tiny. We proved its $O(1/t)$ convergence rate in an ergodic sense. The

$O(1/t)$ non-ergodic convergence rate is also proved for the method using update form (3.3) with $\alpha \in (0,1)$.

Appendix. Why the direct extension of ADMM performs well in practice ?

In this appendix, we try to explain why the direct extension of ADMM performs well in practice. If we use the direct extension of ADMM, then $w^{k+1} = \bar{w}^k$ and thus the relation (3.11) can be written as

$$\begin{aligned}
&(v^{k+1} - v^*)^T P(v^k - v^{k+1}) \\
&\quad \geq \quad \big(\theta(u^{k+1}) - \theta(u^*)\big) + (w^{k+1} - w^*)^T F(w^{k+1}) \\
&\qquad\quad +(\lambda^k - \lambda^{k+1})^T\{(y^k - y^{k+1}) + (z^k - z^{k+1})\}.
\end{aligned} \tag{A.1}$$

Note that M is not symmetric, but $M^T + M$ is positive definite. Again, using $\bar{w}^k = w^{k+1}$, the third part of (3.7) is

$$z^{k+1} \in \mathcal{Z}, \;\; \theta_3(z) - \theta_3(z^{k+1}) + (z - z^{k+1})^T(-\lambda^{k+1}) \geq 0, \;\; \forall\, z \in \mathcal{Z}.$$

It holds also for the previous iteration, thus we have

$$z^k \in \mathcal{Z}, \;\; \theta_3(z) - \theta_3(z^k) + (z - z^k)^T(-\lambda^k) \geq 0, \;\; \forall\, z \in \mathcal{Z}.$$

Setting $z = z^k$ and $z = z^{k+1}$ in the above two sub-VIs, respectively, and then add the two resulting inequalities, we obtain

$$(\lambda^k - \lambda^{k+1})^T(z^k - z^{k+1}) \geq 0. \tag{A.2}$$

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

$$P = D + \begin{pmatrix} 0 & 0 & 0 \\ \beta I_m & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \text{(see (2.7))}$$

it follows from (A.1) and (A.2) that

$$(v^{k+1} - v^*)^T D(v^k - v^{k+1}) \geq (\lambda^k - \lambda^{k+1})^T(y^k - y^{k+1}) + \Delta_k \tag{A.3}$$

where

$$\Delta_k = \big(\theta(u^{k+1}) - \theta(u^*)\big) + (w^{k+1} - w^*)^T F(w^*) + (z^{k+1} - z^*)^T \beta(y^{k+1} - y^k). \tag{A.4}$$

Therefore, by using (A.3), we obtain

$$\begin{aligned}
&\|v^k - v^*\|_D^2 - \|v^{k+1} - v^*\|_D^2 \\
&\quad = \quad \|(v^{k+1} - v^*) + (v^k - v^{k+1})\|_D^2 - \|v^{k+1} - v^*\|_D^2 \\
&\quad = \quad 2(v^{k+1} - v^*)^T D(v^k - v^{k+1}) + \|v^k - v^{k+1}\|_D^2 \\
&\quad \geq \quad \|v^k - v^{k+1}\|_D^2 + 2(\lambda^k - \lambda^{k+1})^T(y^k - y^{k+1}) + 2\Delta_k. \qquad \text{(A.5)}
\end{aligned}$$

Because

$$D = \begin{pmatrix} \beta I_m & 0 & 0 \\ 0 & \beta I_m & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix} \quad \text{and} \quad v = \begin{pmatrix} y \\ z \\ \lambda \end{pmatrix}$$

it follows that

$$\begin{aligned}
&\|v^k - v^{k+1}\|_D^2 + 2(\lambda^k - \lambda^{k+1})^T(y^k - y^{k+1}) \\
&\quad = \quad \beta\|z^k - z^{k+1}\|^2 + \beta\|(y^k - y^{k+1}) + \frac{1}{\beta}(\lambda^k - \lambda^{k+1})\|^2. \qquad \text{(A.6)}
\end{aligned}$$

Substituting (A.6) and (A.3) in the right hand side of (A.5), we obtain

$$\begin{aligned}
&\|v^k - v^*\|_D^2 - \|v^{k+1} - v^*\|_D^2 \\
&\quad = \quad \beta\|z^k - z^{k+1}\|^2 + \beta\|(y^k - y^{k+1}) + \frac{1}{\beta}(\lambda^k - \lambda^{k+1})\|^2 \\
&\qquad +2\{\big(\theta(u^{k+1}) - \theta(u^*)\big) + (w^{k+1} - w^*)^T F(w^*)\} \\
&\qquad +2(z^{k+1} - z^*)^T \beta(y^{k+1} - y^k). \qquad \text{(A.7)}
\end{aligned}$$

In the right hand side of (A.7), the terms

$$\beta\|z^k - z^{k+1}\|^2 + \beta\|(y^k - y^{k+1}) + \frac{1}{\beta}(\lambda^k - \lambda^{k+1})\|^2,$$

and

$$2(\theta(u^{k+1}) - \theta(u^*)\big) + 2(w^{k+1} - w^*)^T F(w^*)$$

are non-negative. However, we do not know whether the last term of the right hand side of (A.7), *i.e.*

$$2(z^{k+1} - z^*)^T \beta(y^{k+1} - y^k)$$

is non-negative. It is pity that we can not show that the right hand side of (A.7) is positive. It seems that the direct extension of ADMM performs well because the right hand side of

(A.7) is positive in practice.

If the right hand side of (A.7) is positive, then the sequence $\{\|v^k - v^*\|_D\}$ is Fejèr monotone and has the contractive property.

References

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

[2] J. Douglas and H. H. Rachford, On the numerical solution of the heat conduction problem in 2 and 3 space variables, *Transactions of the American Mathematical Society* **82** (1956), 421–439.

[3] X.J. Cai, G.Y. Gu and B. S. He *On the $O(1/t)$ convergence rate of the projection and contraction methods for variational inequalities with Lipschitz continuous monotone operators*, Comput. Optim. Appl., 57(2014), 339-363.

[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, 2011, arXiv: 1102.0844v1 [stat.ML] 4 Feb 2011.

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

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

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

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

[9] 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.

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

[12] 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.

[13] 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.

[14] Z. C. Lin, M. M. Chen, L. Q. Wu, and Y. Ma, The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices, *manuscript*, 2009.

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

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

凸优化和单调变分不等式的收缩算法

第十八讲: 多块可分离凸优化问题 带回代的交替方向收缩算法

Alternating direction method of multipliers with back substitution for convex optimization containing more separable blocks

南京大学数学系　何炳生
hebma@nju.edu.cn

The context of this lecture is based on the publication [8]

1 Introduction

In the literature, the alternating direction method of multipliers (ADMM) proposed originally for the following linearly constrained separable convex programming whose objective function is separable into two individual convex functions without crossed variables:

$$\begin{array}{ll} \min & \theta_1(x_1)+\theta_2(x_2) \\ \text{s. t.} & A_1x_1+A_2x_2=b, \\ & x_1\in\mathcal{X}_1 \quad \text{and} \quad x_2\in\mathcal{X}_2, \end{array} \tag{1.1}$$

where $\theta_1:\Re^{n_1}\to\Re$ and $\theta_2:\Re^{n_2}\to\Re$ are closed proper convex functions (not necessarily smooth); $\mathcal{X}_1\subset\Re^{n_1}$ and $\mathcal{X}_2\subset\Re^{n_2}$ are closed convex sets; $A_1\in\Re^{l\times n_1}$ and $A_2\in\Re^{l\times n_2}$ are given matrices; and $b\in\Re^l$ is a given vector.

The augmented Lagrangian function of the problem (1.1) is

$$\mathcal{L}_\beta(x_1,x_2,\lambda)=\theta_1(x_1)+\theta_2(x_2)-\lambda^T(A_1x_1+A_2x_2-b)+\frac{\beta}{2}\|A_1x_1+A_2x_2-b\|^2,$$

where $\lambda^k\in\Re^l$ is the Lagrange multiplier associated with the linear constraints and $\beta>0$ is the penalty parameter for the violation of the linear constraints.

The iterative scheme of ADMM for solving (1.1) is as follows:

$$\begin{cases} x_1^{k+1} \in \arg\min\{\mathcal{L}_\beta(x_1, x_2^k, \lambda^k)\} \\ \qquad = \arg\min\{\theta_1(x_1) + \frac{\beta}{2}\|(A_1x_1 + A_2x_2^k - b) - \frac{1}{\beta}\lambda^k\|^2 \mid x_1 \in \mathcal{X}_1\}, \\ x_2^{k+1} \in \arg\min\{\mathcal{L}_\beta(x_1^{k+1}, x_2, \lambda^k)\} \\ \qquad = \arg\min\{\theta_2(x_2) + \frac{\beta}{2}\|(A_1x_1^{k+1} + A_2x_2 - b) - \frac{1}{\beta}\lambda^k\|^2 \mid x_2 \in \mathcal{X}_2\}, \\ \lambda^{k+1} = \lambda^k - \beta(A_1x_1^{k+1} + A_2x_2^{k+1} - b), \end{cases} \tag{1.2}$$

In this paper, we consider the general case of linearly constrained separable convex programming with $m \geq 3$:

$$\begin{array}{ll} \min & \sum_{i=1}^m \theta_i(x_i) \\ & \sum_{i=1}^m A_i x_i = b; \\ & x_i \in \mathcal{X}_i, \quad i = 1, \cdots, m; \end{array} \tag{1.3}$$

where $\theta_i : \Re^{n_i} \to \Re$ $(i = 1, \ldots, m)$ are closed proper convex functions (not necessarily smooth); $\mathcal{X}_i \subset \Re^{n_i}$ $(i = 1, \ldots, m)$ are closed convex sets; $A_i \in \Re^{l \times n_i}$ $(i = 1, \ldots, m)$ are given matrices and $b \in \Re^l$ is a given vector.

Because of the efficiency of ADMM for (1.1), a natural idea for solving (1.3) is to extend the ADMM (1.2) from the special case (1.1) to the general case (1.3).

In fact, even for the special case of (1.3) with $m = 3$, the extended ADMM is not necessarily convergent [2].

In this paper, we provide a novel approach towards the extension of ADMM for the problem (1.3). More specifically, we show that if a new iterate is generated by correcting the output of the ADMM with a Gaussian back substitution procedure, then the sequence of iterates is convergent to a solution of (1.3). In this sense, we prove the convergence of the extension of ADMM for (1.3). The resulting method is called the ADMM with Gaussian back substitution from now on. Alternatively, the ADMM with Gaussian back substitution can be regarded as a prediction-correction type method whose predictor is generated by the ADMM procedure and the correction is completed by a Gaussian back substitution procedure. We prove the convergence of the ADMM with Gaussian back substitution under the analytic framework of contractive type methods

Throughout, we assume that the solution set of (1.3) is nonempty [a].

[a] In our published paper [8], we assume that $A_i^T A_i$ is nonsingular, this assumption is not necessary.

2 The variational inequality characterization

In this section, we derive the first-order optimality condition of (1.3) and thus characterize (1.3) by a variational inequality (VI). As we will show, the VI characterization is convenient for the convergence analysis to be conducted.

By attaching a Lagrange multiplier vector $\lambda \in \Re^l$ to the linear constraint, the Lagrange function of (1.3) is:

$$L(x_1, x_2, \ldots, x_m, \lambda) = \sum_{i=1}^{m} \theta_i(x_i) - \lambda^T (\sum_{i=1}^{m} A_i x_i - b), \tag{2.1}$$

which is defined on

$$\Omega := \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_m \times \Re^l.$$

Let $\left(x_1^*, x_2^*, \ldots, x_m^*, \lambda^*\right)$ be a saddle point of the Lagrange function (2.1). Then we have

$$\begin{aligned} L_{\lambda \in \Re^l}(x_1^*, x_2^*, \cdots, x_m^*, \lambda) &\leq L(x_1^*, x_2^*, \cdots, x_m^*, \lambda^*) \\ &\leq L_{x_i \in \mathcal{X}_i \ (i=1,\ldots,m)}(x_1, x_2, \ldots, x_m, \lambda^*). \end{aligned}$$

It is evident that finding a saddle point of $L(x_1, x_2, \ldots, x_m, \lambda)$ is equivalent to finding

$w^* = (x_1^*, x_2^*, ..., x_m^*, \lambda^*) \in \Omega$, such that

$$\left\{ \begin{array}{l} \theta_1(x_1) - \theta_1(x_1^*) + (x_1 - x_1^*)^T \{-A_1^T \lambda^*\} \geq 0, \\ \theta_2(x_2) - \theta_2(x_2^*) + (x_2 - x_2^*)^T \{-A_2^T \lambda^*\} \geq 0, \\ \qquad\qquad \vdots \\ \qquad\qquad \vdots \\ \theta_m(x_m) - \theta_m(x_m^*) + (x_m - x_m^*)^T \{-A_m^T \lambda^*\} \geq 0, \\ (\lambda - \lambda^*)^T (\sum_{i=1}^m A_i x_i^* - b) \geq 0, \end{array} \right. \tag{2.2}$$

for all $w = (x_1, x_2, \cdots, x_m, \lambda) \in \Omega$. More compactly, (2.2) can be written into the following VI:

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

where

$$\theta(x) = \sum_{i=1}^{m} \theta_i(x_i),$$

$$w = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \\ \lambda \end{pmatrix} \quad \text{and} \quad F(w) = \begin{pmatrix} -A_1^T\lambda \\ -A_2^T\lambda \\ \vdots \\ -A_m^T\lambda \\ \sum_{i=1}^m A_i x_i - b \end{pmatrix}. \tag{2.3b}$$

Note that the operator $F(w)$ defined in (2.3b) is monotone. Especially, we have

$$(w - \tilde{w})^T (F(w) - F(\tilde{w})) \equiv 0. \tag{2.4}$$

In addition, since we have assumed that the solution set of (1.3) is not empty, the solution set of (2.3), denoted by Ω^*, is also nonempty. In addition to the notation of $w = (x_1, x_2, \cdots, x_m, \lambda)$, for any integer number k we also use the following notation:

$$v = (x_2, \cdots, x_m, \lambda).$$

Moreover, we define

$$\mathcal{V}^* = \{(x_2^*, \ldots, x_m^*, \lambda^*) \,|\, (x_1^*, x_2^*, \ldots, x_m^*, \lambda^*) \in \Omega^*\}.$$

3 ADMM with Gaussian back substitution

In this section, we show the combination of the extended ADMM scheme (1.2) with a Gaussian back substitution procedure, and derive the resulting ADMM with Gaussian back substitution for solving (1.3). We also elucidate how to realize the Gaussian back substitution for some special cases of (1.3).

To present the Gaussian back substitution procedure, we define the matrices:

$$L = \begin{pmatrix} I_l & 0 & \cdots & \cdots & \cdots & 0 \\ I_l & I_l & \ddots & & & \vdots \\ \vdots & \vdots & \ddots & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & & \vdots \\ I_l & I_l & \cdots & \cdots & I_l & 0 \\ 0 & 0 & \cdots & \cdots & 0 & I_l \end{pmatrix}_{m\text{-Blocks}}, \tag{3.1}$$

and

$$P = \text{diag}\big(\sqrt{\beta}A_2, \sqrt{\beta}A_3, \ldots, \sqrt{\beta}A_m, \sqrt{\frac{1}{\beta}}I_l\big), \qquad D = P^T P. \tag{3.2}$$

Note that for any $\beta > 0$, from Pv^k, we get $(A_2x_2^k, \cdots, A_mx_m^k, \lambda^k)$ very easily. The matrix L defined in (3.1) is a non-singular lower-triangular block matrix and D defined in (3.2) is a symmetric positive definite matrix. In addition, according to (3.1) and (3.2), we easily have:

$$L^{-T} = \begin{pmatrix} I_l & -I_l & 0 & \cdots & 0 & 0 \\ 0 & \ddots & \ddots & & \vdots & \vdots \\ \vdots & \ddots & \ddots & -I_l & & 0 \\ 0 & \cdots & 0 & I_l & -I_l & 0 \\ 0 & \cdots & 0 & 0 & I_l & 0 \\ 0 & \cdots & 0 & 0 & 0 & I_l \end{pmatrix}_{m\text{-Blocks}} . \tag{3.3}$$

which is a upper-triangular block matrix whose diagonal components are identity matrices.

Algorithm: ADMM with Gaussian back substitution for (1.3):

Let $\beta > 0$ and $\alpha \in [0.5, 1)$, With the given iterate $Pv^k = P(x_2^k, \cdots, x_m^k, \lambda^k)$.

Step 1. ADMM step (prediction step). For $i = 1, \ldots, m$, obtain $\tilde{x}_i^k$ in the forward (alternating) order by solving the following x_i-problem:

$$\min\big\{\theta_i(x_i) + \tfrac{\beta}{2}\|(\textstyle\sum_{j=1}^{i-1} A_j\tilde{x}_j^k + A_ix_i + \sum_{j=i+1}^{m} A_jx_j^k - b) - \tfrac{1}{\beta}\lambda^k\|^2 \mid x_i \in \mathcal{X}_i\big\} \tag{3.4a}$$

and set

$$\tilde{\lambda}^k = \lambda^k - \beta(\textstyle\sum_{j=1}^{m} A_j\tilde{x}_j^k - b). \tag{3.4b}$$

Step 2. Gaussian back substitution step (correction step). Correct the ADMM output $\tilde{w}^k$ in the backward order by the following Gaussian back substitution procedure and generate the new iterate Pv^{k+1}:

$$L^T P(v^{k+1} - v^k) = \alpha P\big(\tilde{v}^k - v^k\big). \tag{3.4c}$$

where the matrices L and P are defined by (3.1) and (3.2), respectively.

Note that the k-th iteration starts with a given $(A_2x_2^k, \cdots, A_mx_m^k, \lambda^k)$, and offers the

new iterate $(A_2x_2^{k+1}, \cdots, A_mx_m^{k+1}, \lambda^{k+1})$. Only for analysis convenience, we use Pv^k and Pv^{k+1}. In any way, the variable x_1 is only an intermediate variable.

Recall that the matrix L^T in (3.4c) is a upper-triangular block matrix. Thus, we call the correction step (3.4c) is a Gaussian back substitution, and it is very easy to execute. In fact, as we mentioned, after the predictor is generated by the ADMM scheme (3.4a) in the forward (alternating) order, the proposed Gaussian back substitution step corrects the predictor in the backward order. Since the Gaussian back substitution step is easy to perform, the computation of each iteration of the ADMM with Gaussian back substitution is dominated by the ADMM procedure (3.4a).

To show the main idea with clearer notation, we restrict our theoretical discussion to the case with fixed $\beta > 0$.

The main task of the Gaussian back substitution step (3.4c) can be rewritten into

$$Pv^{k+1} = Pv^k - \alpha L^{-T}P(v^k - \tilde{v}^k). \tag{3.5}$$

In fact, we can choose the step size dynamically based on some techniques in the literature (e.g. [10]), and the Gaussian back substitution procedure with the constant α can

be modified accordingly into the following variant with a dynamical step size:

$$Pv^{k+1} = Pv^k - \gamma\alpha_k^* L^{-T}P(v^k - \tilde{v}^k). \tag{3.6}$$

where

$$\alpha_k^* = \frac{\|v^k - \tilde{v}^k\|_D^2 + \|v^k - \tilde{v}^k\|_G^2}{2\|v^k - \tilde{v}^k\|_D^2}; \tag{3.7}$$

$$G = \begin{pmatrix} \beta A_2^TA_2 & \beta A_2^TA_3 & \cdots & \beta A_2^TA_m & A_2^T \\ \beta A_3^TA_2 & \beta A_3^TA_3 & \cdots & \beta A_3^TA_m & A_3^T \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \beta A_m^TA_2 & \beta A_m^TA_3 & \cdots & \beta A_m^TA_m & A_m^T \\ A_2 & A_3 & \cdots & A_m & \frac{1}{\beta}I_l \end{pmatrix}; \tag{3.8}$$

and $\gamma \in (0, 2)$. Indeed, for any $\beta > 0$, the symmetric matrix Q is positive semi-definite. Then, for given v^k and the $\tilde{v}^k$ obtained by the ADMM procedure (3.4a), we have that

$$\|v^k - \tilde{v}^k\|_D^2 = \beta\sum_{i=2}^m \|A_i(x_i^k - \tilde{x}_i^k)\|^2 + \frac{1}{\beta}\|\lambda^k - \tilde{\lambda}^k\|^2,$$

and

$$\|v^k - \tilde{v}^k\|_G^2 = \beta \left\| \sum_{i=2}^{m} A_i(x_i^k - \tilde{x}_i^k) + \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) \right\|^2,$$

where the norm $\|v\|_D^2$ ($\|v\|_G^2$, respectively) is defined as $v^T Dv$ ($v^T Gv$, respectively). In fact, it is easy to prove that the step size α_k^* defined in (3.7) satisfies $\frac{1}{2} \le \alpha_k^* \le \frac{m+1}{2}$.

4 Convergence of the proposed method

In this section, we prove the convergence of the proposed ADMM with Gaussian back substitution for solving (1.3). Our proof follows the analytic framework of contractive type methods, and it consists of the following three phases:

1.) Prove that the sequence generated by the proposed ADMM with Gaussian back substitution is contractive with respect to $\mathcal{V}^*$.

2.) Derive the convergence based on the Fej$\acute{e}$r monotonicity of the sequence generated by the proposed ADMM with Gaussian back substitution.

Accordingly, we divide this section into three subsections to address the tasks listed above.

4.1 Verification of the descent directions

For this purpose, we first prove two lemmas.

Lemma 4.1 *Let* $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \ldots, \tilde{x}_m^k, \tilde{\lambda}^k)$ *be generated by the ADMM step* (3.4a) *from the given vector* $Pv^k = P(x_2^k, \ldots, x_m^k, \lambda^k)$. *Then, we have*

$$\begin{aligned} \tilde{w}^k \in \Omega, \;\; & \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) + (x - \tilde{x}^k)^T d_2(x^k, \tilde{x}^k) \\ \ge \;\; & (v - \tilde{v}^k)^T d_1(v^k, \tilde{v}^k), \quad \forall\, w \in \Omega, \end{aligned} \tag{4.1}$$

where

$$\begin{aligned} d_1(v^k, \tilde{v}^k) &= \begin{pmatrix} \beta A_2^T A_2 & 0 & \cdots & \cdots & 0 \\ \beta A_3^T A_2 & \beta A_3^T A_3 & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \beta A_m^T A_2 & \beta A_m^T A_3 & \cdots & \beta A_m^T A_m & 0 \\ 0 & 0 & \cdots & 0 & \frac{1}{\beta} I_l \end{pmatrix} \begin{pmatrix} x_2^k - \tilde{x}_2^k \\ x_3^k - \tilde{x}_3^k \\ \vdots \\ x_m^k - \tilde{x}_m^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}, \\ &= P^T L P(v^k - \tilde{v}^k). \end{aligned} \tag{4.2}$$

and

$$d_2(x^k, \tilde{x}^k) = \beta \begin{pmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_m^T \end{pmatrix} \big(\sum_{j=2}^{m} A_j(x_j^k - \tilde{x}_j^k)\big). \tag{4.3}$$

Proof. Note that $\tilde{x}_i^k \in \arg\min\{\theta_i(x_i) + \frac{\beta}{2}\|A_i x_i - b_i\|^2 | x_i \in \mathcal{X}_i\}$ is equivalent to

$$\tilde{x}_i^k \in \mathcal{X}_i, \;\; \theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T \beta A_i^T (A_i \tilde{x}_i^k - b_i) \ge 0, \;\; \forall\, x_i \in \mathcal{X}_i.$$

Since $\tilde{x}_i^k$ is the solution of (3.4a), for $i = 1, 2, \ldots, m$, according to the optimality condition, we have

$$\begin{aligned} \tilde{x}_i^k \in \mathcal{X}_i, \quad & \theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T \{-A_i^T \lambda^k \\ & + \beta A_i^T (\textstyle\sum_{j=1}^{i} A \tilde{x}_j^k + \sum_{j=i+1}^{m} A_j x_j^k - b)\} \ge 0, \;\; \forall\, x_i \in \mathcal{X}_i. \end{aligned} \tag{4.4}$$

By using the fact (see (3.4b)) $\lambda^k = \tilde{\lambda}^k + \beta(\sum_{j=1}^{m} A_j \tilde{x}_j^k - b)$, thus, we have

$$A_i^T \lambda^k = A_i \tilde{\lambda}^k + \beta A_i (\sum_{j=1}^{m} A_j \tilde{x}_j^k - b).$$

Substituting it in (4.4), we obtain

$$\begin{aligned} \tilde{x}_i^k \in \mathcal{X}_i, \quad & \theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T \{-A_i^T \tilde{\lambda}^k \\ & + \beta A_i^T \big(\textstyle\sum_{j=i+1}^{m} A_j (x_j^k - \tilde{x}_j^k)\big)\} \ge 0, \;\; \forall\, x_i \in \mathcal{X}_i. \end{aligned} \tag{4.5}$$

It follows from (4.5) that $\tilde{x}^k \in \mathcal{X}$ and

$$\tilde{x}^k \in \mathcal{X}, \quad \theta(x) - \theta(\tilde{x}^k) +$$

$$\begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ x_3 - \tilde{x}_3^k \\ \vdots \\ x_{m-1} - \tilde{x}_{m-1} \\ x_m - \tilde{x}_m^k \end{pmatrix}^T \left\{ \begin{pmatrix} -A_1^T \tilde{\lambda}^k \\ -A_2^T \tilde{\lambda}^k \\ -A_3^T \tilde{\lambda}^k \\ \vdots \\ -A_{m-1}^T \tilde{\lambda}^k \\ -A_m^T \tilde{\lambda}^k \end{pmatrix} + \beta \begin{pmatrix} A_1^T \big(\sum_{j=2}^{m} A_j (x_j^k - \tilde{x}_j^k)\big) \\ A_2^T \big(\sum_{j=3}^{m} A_j (x_j^k - \tilde{x}_j^k)\big) \\ A_3^T \big(\sum_{j=4}^{m} A_j (x_j^k - \tilde{x}_j^k)\big) \\ \vdots \\ A_{m-1}^T \big(A_m (x_m^k - \tilde{x}_m^k)\big) \\ 0 \end{pmatrix} \right\} \ge 0, \tag{4.6}$$

for all $x \in \mathcal{X}$. Adding

$$\begin{pmatrix} x_2 - \tilde{x}_2^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \end{pmatrix}^T \beta \begin{pmatrix} A_2^T\big(A_2(x_2^k - \tilde{x}_2^k)\big) \\ A_3^T\big(\sum_{j=2}^3 A_j(x_j^k - \tilde{x}_j^k)\big) \\ \vdots \\ A_m^T\big(\sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k)\big) \end{pmatrix}$$

to the both sides of (4.6), we get $\tilde{x}^k \in \mathcal{X}$ and

$$\begin{aligned} \theta(x) - \theta(\tilde{x}^k) + & \begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \end{pmatrix}^T \begin{pmatrix} -A_1^T\tilde{\lambda}^k + \beta A_1^T\big(\sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k)\big) \\ -A_2^T\tilde{\lambda}^k + \beta A_2^T\big(\sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k)\big) \\ \vdots \\ -A_m^T\tilde{\lambda}^k + \beta A_m^T\big(\sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k)\big) \end{pmatrix} \\ \geq & \begin{pmatrix} x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \end{pmatrix}^T \begin{pmatrix} \beta A_2^T\big(\sum_{j=2}^2 A_j(x_j^k - \tilde{x}_j^k)\big) \\ \vdots \\ \beta A_m^T\big(\sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k)\big) \end{pmatrix}, \quad \forall\, x \in \mathcal{X}. \end{aligned} \tag{4.7}$$

Because that $\sum_{j=1}^m A_j\tilde{x}_j^k - b = \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k)$, we have

$$(\lambda - \tilde{\lambda}^k)^T\Big(\sum_{j=1}^m A_j\tilde{x}_j^k - b\Big) = (\lambda - \tilde{\lambda}^k)^T\frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k).$$

Adding (4.7) and the last equality together, we get $\tilde{w}^k \in \mathcal{W}$ and

$$\begin{aligned} \theta(x) - \theta(\tilde{x}^k) + & \begin{pmatrix} x_1 - \tilde{x}_1^k \\ x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} -A_1^T\tilde{\lambda}^k + \beta A_1^T\big(\sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k)\big) \\ -A_2^T\tilde{\lambda}^k + \beta A_2^T\big(\sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k)\big) \\ \vdots \\ -A_m^T\tilde{\lambda}^k + \beta A_m^T\big(\sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k)\big) \\ \sum_{i=1}^m A_i\tilde{x}_i^k - b \end{pmatrix} \\ \geq & \begin{pmatrix} x_2 - \tilde{x}_2^k \\ \vdots \\ x_m - \tilde{x}_m^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} \beta A_2^T\big(\sum_{j=2}^2 A_j(x_j^k - \tilde{x}_j^k)\big) \\ \vdots \\ \beta A_m^T\big(\sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k)\big) \\ \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix}, \quad \forall\, w \in \Omega. \end{aligned}$$

Use the notations of $F(w)$, $d_1(v^k, \tilde{v}^k)$ and $d_2(x^k, \tilde{x}^k)$, the left hand side and the right

hand side of the above inequality can be written as

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) + (x - \tilde{x}^k)^T d_2(x^k, \tilde{x}^k)$$

and

$$(v - \tilde{v}^k)^T d_1(v^k, \tilde{v}^k),$$

respectively. Thus, the assertion of this lemma is proved. $\square$

Note that the $d_1(v^k, \tilde{v}^k)$ depends only on v^k and $\tilde{v}^k$, while $d_2(x^k, \tilde{x}^k)$ is determined by both x^k and $\tilde{x}^k$.

Lemma 4.2 *Let $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \ldots, \tilde{x}_m^k, \tilde{\lambda}^k)$ be generated by the ADMM step (3.4a) from the given vector $Pv^k = P(x_2^k, \ldots, x_m^k, \lambda^k)$. Then, we have*

$$(\tilde{v}^k - v^*)^T d_1(v^k, \tilde{v}^k) \geq (\lambda^k - \tilde{\lambda}^k)^T \Big(\sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k)\Big), \quad \forall\, v^* \in \mathcal{V}^*, \tag{4.8}$$

where $d_1(v^k, \tilde{v}^k)$ is defined in (4.2).

Proof. Since $w^* \in \Omega$, it follows from (4.1) that

$$\begin{aligned}&(\tilde{v}^k - v^*)^T d_1(v^k, \tilde{v}^k)\\ &\quad\geq\quad \theta(\tilde{x}^k) - \theta(x^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}) + (\tilde{x}^k - x^*)^T d_2(x^k, \tilde{x}^k).\end{aligned} \tag{4.9}$$

We consider the right-hand side of (4.9). By using (4.3), we get

$$\begin{aligned}&\theta(\tilde{x}^k) - \theta(x^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}) + (\tilde{x}^k - x^*)^T d_2(x^k, \tilde{x}^k)\\ &\quad=\quad \theta(\tilde{x}^k) - \theta(x^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k)\\ &\qquad\quad +\Big(\sum_{j=2}^m A_j(x_j^k - \tilde{x}_j^k)\Big)^T \beta\Big(\sum_{j=1}^m A_j(\tilde{x}_j^k - x_j^*)\Big).\end{aligned} \tag{4.10}$$

Then, we look at the right-hand side of (4.10). First, by using the skew-symmetry of F (2.4) and the optimality of w^*, we have

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

For the last part of the right hand side of (4.10), because

$$\sum_{j=1}^{m} A_j x_j^* = b \qquad \text{and} \qquad \beta(\sum_{j=1}^{m} A_j \tilde{x}_j^k - b) = \lambda^k - \tilde{\lambda}^k,$$

it follows

$$\big(\sum_{j=2}^{m} A_j(x_j^k - \tilde{x}_j^k)\big)^T \beta\big(\sum_{j=1}^{m} A_j(\tilde{x}_j^k - x_j^*)\big) = \big(\sum_{j=2}^{m} A_j(x_j^k - \tilde{x}_j^k)\big)^T (\lambda^k - \tilde{\lambda}^k).$$

Finally, from (4.10) that

$$\begin{aligned} &\theta(\tilde{x}^k) - \theta(x^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}) + (\tilde{x}^k - x^*)^T d_2(x^k, \tilde{x}^k) \\ &\geq \quad (\lambda^k - \tilde{\lambda}^k)^T \big(\sum_{j=2}^{m} A_j(x_j^k - \tilde{x}_j^k)\big). \end{aligned} \tag{4.11}$$

Substituting (4.11) into (4.9), the assertion (4.8) follows immediately. □

Since (see (3.1) and (4.2))

$$d_1(v^k, \tilde{v}^k) = P^T L P(v^k - \tilde{v}^k), \tag{4.12}$$

consequently from (4.8) follows that

$$(\tilde{v}^k - v^*)^T P^T L P(v^k - \tilde{v}^k) \geq (\lambda^k - \tilde{\lambda}^k)^T \big(\sum_{j=2}^{m} A_j(x_j^k - \tilde{x}_j^k)\big), \ \forall\, v^* \in \mathcal{V}^*. \tag{4.13}$$

Now, based on the last two lemmas, we are at the stage to prove the main theorem.

Theorem 4.1 **(Main Theorem)** *Let* $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \ldots, \tilde{x}_m^k, \tilde{\lambda}^k)$ *be generated by the ADMM step* (3.4a) *from the given vector* $Pv^k = P(x_2^k, \ldots, x_m^k, \lambda^k)$*. Then, we have*

$$(v^k - v^*)^T P^T L P(v^k - \tilde{v}^k) \geq \frac{1}{2}\|v^k - \tilde{v}^k\|_D^2 + \frac{1}{2}\|v^k - \tilde{v}^k\|_G^2, \quad \forall\, v^* \in \mathcal{V}^*, \tag{4.14}$$

where L*,* P*,* D *and* G *are defined in* (3.1)*,* (3.2) *and* (3.8)*, respectively.*

Proof First, for all $v^* \in \mathcal{V}^*$, it follows from (4.13) that

$$\begin{aligned} &(v^k - v^*)^T P^T L P(v^k - \tilde{v}^k) \\ &\geq \ (v^k - \tilde{v}^k)^T P^T L P(v^k - \tilde{v}^k) + (\lambda^k - \tilde{\lambda}^k)^T \big(\sum_{j=2}^{m} A_j(x_j^k - \tilde{x}_j^k)\big). \end{aligned} \tag{4.15}$$

Now, we treat the first term of the right hand side of (4.15). Using the matrix L and P (see

(3.1) and (3.2)) , we have

$$
\begin{aligned}
&(v^k-\tilde{v}^k)^T P^T L P(v^k-\tilde{v}^k) \\
&= \begin{pmatrix} x_2^k-\tilde{x}_2^k \\ \vdots \\ x_m^k-\tilde{x}_m^k \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix} \begin{pmatrix} \beta A_2^T A_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \beta A_m^T A_2 & \cdots & \beta A_m^T A_m & 0 \\ 0 & \cdots & 0 & \frac{1}{\beta} I_l \end{pmatrix}^T \begin{pmatrix} x_2^k-\tilde{x}_2^k \\ \vdots \\ x_m^k-\tilde{x}_m^k \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}
\end{aligned} \tag{4.16}
$$

Let us deal with the second term of the right-hand side of (4.15). We have

$$
\begin{aligned}
&(\lambda^k-\tilde{\lambda}^k)^T \Big(\sum_{j=2}^m A_j(x_j^k-\tilde{x}_j^k)\Big) \\
&= \begin{pmatrix} x_2^k-\tilde{x}_2^k \\ \vdots \\ x_m^k-\tilde{x}_m^k \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} 0 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 0 \\ A_2 & \dots & A_m & 0 \end{pmatrix} \begin{pmatrix} x_2^k-\tilde{x}_2^k \\ \vdots \\ x_m^k-\tilde{x}_m^k \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}.
\end{aligned} \tag{4.17}
$$

Adding (4.16) and (4.17) together, it follows that

$$
\begin{aligned}
&(v^k-\tilde{v}^k)^T P^T L P(v^k-\tilde{v}^k) + (\lambda^k-\tilde{\lambda}^k)^T \Big(\sum_{j=2}^m A_j(x_j^k-\tilde{x}_j^k)\Big) \\
&= \begin{pmatrix} x_2^k-\tilde{x}_2^k \\ \vdots \\ x_m^k-\tilde{x}_m^k \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix} \begin{pmatrix} \beta A_2^T A_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \beta A_m^T A_2 & \cdots & \beta A_m^T A_m & 0 \\ A_2 & \cdots & A_m & \frac{1}{\beta} I_l \end{pmatrix}^T \begin{pmatrix} x_2^k-\tilde{x}_2^k \\ \vdots \\ x_m^k-\tilde{x}_m^k \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix} \\
&= \frac{1}{2} \begin{pmatrix} x_2^k-\tilde{x}_2^k \\ \vdots \\ x_m^k-\tilde{x}_m^k \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} 2\beta A_2^T A_2 & \cdots & \beta A_2^T A_m & A_2^T \\ \vdots & & \vdots & \vdots \\ \beta A_m^T A_2 & \cdots & 2\beta A_m^T A_m & A_m^T \\ A_2 & \cdots & A_m & \frac{2}{\beta} I_l \end{pmatrix} \begin{pmatrix} x_2^k-\tilde{x}_2^k \\ \vdots \\ x_m^k-\tilde{x}_m^k \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}.
\end{aligned}
$$

Use the notation of the matrices D and G to the right-hand side of the last equality, we

obtain

$$\begin{aligned}&(v^k-\tilde{v}^k)^T P^T L P(v^k-\tilde{v}^k)+(\lambda^k-\tilde{\lambda}^k)^T\big(\sum_{j=2}^m A_j(x_j^k-\tilde{x}_j^k)\big)\\ =\quad &\frac{1}{2}\|v^k-\tilde{v}^k\|_D^2+\frac{1}{2}\|v^k-\tilde{v}^k\|_G^2.\end{aligned}$$

Substituting the last equality in (4.15), this theorem is proved. □

It follows from (4.14) that

$$\big\langle L^T P(v^k-v^*), P(v^k-\tilde{v}^k)\big\rangle \ge \frac{1}{2}\|v^k-\tilde{v}^k\|_{(D+G)}^2. \tag{4.18}$$

Based on (4.18), in order to minimize $\|L^T P(v-v^*)\|^2$, we need only to let

$$L^T P(v^{k+1}-v^*) = L^T P(v^k-v^*) - \alpha P(v^k-\tilde{v}^k) \tag{4.19}$$

and by selecting a suitable step-size α.

$$Pv^{k+1} = Pv^k - \alpha L^{-T} P(v^k-\tilde{v}^k).$$

By setting

$$H = P^T L L^T P, \tag{4.20}$$

$\frac{1}{2}\|v-v^*\|_G^2$, and $L^{-T}P(\tilde{v}^k-v^k)$ is a descent direction of $\frac{1}{2}\|v-v^*\|_H^2$ at the current point v^k whenever $\tilde{v}^k \neq v^k$.

4.2 The contractive property

In this subsection, we mainly prove that the sequence generated by the proposed ADMM with Gaussian back substitution is contractive with respect to the set $\mathcal{V}^*$. Note that we follow the definition of contractive type methods. With this contractive property, the convergence of the proposed ADMM with Gaussian back substitution can be easily derived with subroutine analysis.

Theorem 4.2 *Let $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \ldots, \tilde{x}_m^k, \tilde{\lambda}^k)$ be generated by the ADMM step* (3.4a) *from the given vector $Pv^k = P(x_2^k, \ldots, x_m^k, \lambda^k)$. Let the matrix G be given by* (4.20)*. For the new iterate v^{k+1} produced by the Gaussian back substitution* (3.5)*, there exists a constant $c_0 > 0$ such that*

$$\|v^{k+1}-v^*\|_H^2 \le \|v^k-v^*\|_H^2 - c_0\big(\|v^k-\tilde{v}^k\|_D^2+\|v^k-\tilde{v}^k\|_G^2\big), \quad \forall\, v^* \in \mathcal{V}^*, \tag{4.21}$$

where H, D and G are defined in (4.20)*,* (3.2) *and* (3.8)*, respectively.*

Proof By using (3.5) and $H = P^T LL^T P$, we obtain

$$\begin{aligned}
&\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\
&\quad = \quad \|P(v^k - v^*)\|_{(LL^T)}^2 - \|P(v^{k+1} - v^*)\|_{(LL^T)}^2 \\
&\quad = \quad \|P(v^k - v^*)\|_{(LL^T)}^2 - \|P(v^k - v^*) - \alpha L^{-T} P(v^k - \tilde{v}^k)\|_{(LL^T)}^2 \\
&\quad = \quad 2\alpha(v^k - v^*)^T P^T LP(v^k - \tilde{v}^k) - \alpha^2 \|v^k - \tilde{v}^k\|_D^2. \qquad (4.22)
\end{aligned}$$

Substituting the result of Theorem 4.1 into the right-hand side of the last equation, we get

$$\begin{aligned}
&\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\
&\quad \geq \quad \alpha\big(\|v^k - \tilde{v}^k\|_D^2 + \|v^k - \tilde{v}^k\|_G^2\big) - \alpha^2 \|v^k - \tilde{v}^k\|_D^2 \\
&\quad = \quad \alpha(1-\alpha)\|v^k - \tilde{v}^k\|_D^2 + \alpha\|v^k - \tilde{v}^k\|_G^2,
\end{aligned}$$

and thus

$$\begin{aligned}
\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 \\
-\alpha\big((1-\alpha)\|v^k - \tilde{v}^k\|_D^2 + \|v^k - \tilde{v}^k\|_G^2\big), \quad \forall\, v^* \in \mathcal{V}^*. \qquad (4.23)
\end{aligned}$$

Set $c_0 = \alpha(1-\alpha)$. Recall that $\alpha \in [0.5, 1)$. Thus the assertion is proved.

Corollary 4.1 *The assertion of Theorem 4.2 also holds if the Gaussian back substitution update form is (3.6) with the calculated step length by (3.7) .*

Proof Analogous to the proof of Theorem 4.2, we have that

$$\begin{aligned}
&\|v^k - v^*\|_G^2 - \|v^{k+1} - v^*\|_G^2 \\
&\quad \geq \quad 2\gamma\alpha_k^*(v^k - v^*)^T P^T LP(v^k - \tilde{v}^k) - (\gamma\alpha_k^*)^2 \|v^k - \tilde{v}^k\|_D^2, \qquad (4.24)
\end{aligned}$$

where α_k^* is given by (3.7). According to (3.7), we have that

$$\alpha_k^*\big(\|v^k - \tilde{v}^k\|_D^2\big) = \frac{1}{2}\big(\|v^k - \tilde{v}^k\|_D^2 + \|v^k - \tilde{v}^k\|_G^2\big).$$

Then, it follows from the above equality and (4.14) that

$$\begin{aligned}
&\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\
&\quad \geq \quad \gamma\alpha_k^*\big(\|v^k - \tilde{v}^k\|_D^2 + \|v^k - \tilde{v}^k\|_G^2\big) - \tfrac{1}{2}\gamma^2\alpha_k^*\big(\|v^k - \tilde{v}^k\|_D^2 + \|v^k - \tilde{v}^k\|_G^2\big) \\
&\quad = \quad \tfrac{1}{2}\gamma(2-\gamma)\alpha_k^*\big(\|v^k - \tilde{v}^k\|_D^2 + \|v^k - \tilde{v}^k\|_G^2\big).
\end{aligned}$$

Because $\alpha_k^* \geq \frac{1}{2}$, it follows from the last inequality that

$$\begin{aligned}
\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 \\
-\tfrac{1}{4}\gamma(2-\gamma)\big(\|v^k - \tilde{v}^k\|_D^2 + \|v^k - \tilde{v}^k\|_G^2\big), \quad \forall v^* \in \mathcal{V}^*. \qquad (4.25)
\end{aligned}$$

Since $\gamma \in (0, 2)$, the assertion of this corollary follows from (4.25) directly. $\square$

4.3 Convergence

The proposed lemmas and theorems are adequate to establish the global convergence of the proposed ADMM with Gaussian back substitution, and the analytic framework is quite typical in the context of contractive type methods.

Theorem 4.3 *Let $\{v^k\}$ and $\{\tilde{w}^k\}$ be the sequences generated by the proposed ADMM with Gaussian back substitution. Then we have*

1. $\lim_{k\to\infty} \|A_i(x^k - \tilde{x}_i^k)\| = 0, i = 2, \ldots, m,$ *and* $\lim_{k\to\infty} \|\lambda^k - \tilde{\lambda}^k\| = 0.$
2. *If each A_i is full column rank matrix, then*
   (a) *any cluster point of $\{\tilde{w}^k\}$ is a solution point of* (2.3).
   (b) *The sequence $\{\tilde{v}^k\}$ converges to some $v^\infty \in \mathcal{V}^*$.*

Proof. From (4.21) we get

$$\sum_{k=0}^{\infty} c_0 \|v^k - \tilde{v}^k\|_D^2 \le \|v^0 - v^*\|_H^2$$

and thus we get $\lim_{k\to\infty} \|v^k - \tilde{v}^k\|_D^2 = 0$, and consequently

$$\lim_{k\to\infty} \|A_i(x^k - \tilde{x}_i^k)\| = 0, \quad i = 2, \ldots, m, \tag{4.26}$$

and

$$\lim_{k\to\infty} \|\lambda^k - \tilde{\lambda}^k\| = 0. \tag{4.27}$$

The first assertion is proved.

Now, we assume that each A_i is full column rank matrix. Substituting (4.26) into (4.5), for $i = 1, 2, \ldots, m$, we have

$$\tilde{x}_i^k \in \mathcal{X}_i, \ \lim_{k\to\infty} \{\theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T(-A_i^T\tilde{\lambda}^k)\} \ge 0, \ \ \forall\, x_i \in \mathcal{X}_i. \tag{4.28}$$

It follows from (3.4a) and (4.27) that

$$\lim_{k\to\infty} (\sum_{j=1}^{m} A_j \tilde{x}_j^k - b) = 0. \tag{4.29}$$

Combining (4.28) and (4.29) we get

$$\tilde{w}^k \in \Omega, \ \ \lim_{k\to\infty} \{\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k)\} \ge 0, \ \ \forall\, w \in \Omega, \tag{4.30}$$

and thus any cluster point of $\{\tilde{w}^k\}$ is a solution point of (2.3). The part (a) of the second assertion is proved.

It follows from the full column rank assumption, the matrix D and H are positive definite. Since $\lim_{k\to\infty}\|v^k-\tilde{v}^k\|_D^2=0$, then $\{\tilde{v}^k\}$ is bounded. Let v^∞ be a cluster point of $\{\tilde{v}^k\}$ and the subsequence $\{\tilde{v}^{k_j}\}$ converges to v^∞. It follows from (4.30)

$$\tilde{w}^{k_j}\in\Omega,\quad \lim_{k\to\infty}\{\theta(x)-\theta(\tilde{x}^{k_j})+(w-\tilde{w}^{k_j})^TF(\tilde{w}^{k_j})\}\ge 0,\quad \forall\, w\in\Omega \tag{4.31}$$

and consequently

$$\left\{\begin{array}{l}\theta_i(x_i)-\theta(x_i^\infty)+(x_i-x_i^\infty)^T\{-A_i^T\lambda^\infty\}\ge 0,\quad \forall x_i\in\mathcal{X}_i,\ \ i=1,\ldots,m,\\[2ex] \sum_{j=1}^m A_jx_j^\infty-b=0.\end{array}\right. \tag{4.32}$$

This means that $v^\infty\in\mathcal{V}^*$. Since $\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2$ and $\lim_{k\to\infty}\|v^k-\tilde{v}^k\|=0$, the sequence $\{\tilde{v}^k\}$ cannot have other cluster point and $\{\tilde{v}^k\}$ converges to $v^\infty\in\mathcal{V}^*$. $\square$

If we take $\alpha\equiv 1$ in the correction form (3.5), similarly as in the last lecture, the resulting method is convergent in the ergodic sense with the convergence rate $O(1/t)$.

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, Nov. 2010.

[2] 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 necessary convergent, Mathematical Programming, 155 (2016) 57-79.

[3] D. Gabay, “Applications of the Method of Multipliers to Variational Inequalities,” in *Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems*, M. Fortin and R. Glowinski, eds., North Holland, Amsterdam, The Netherlands, 299-331, 1983.

[4] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems viafinite element approximations, Computers and Mathematics with Applications, 2:17-40, 1976.

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

[6] B.S. He, A class of projection and contraction methods for monotone variational inequalities, Applied Mathematics and Optimization, **35**, pp. 69-76, 1997.

[7] B.S. He, “Inexact implicit methods for monotone general variational inequalities”, Math. Program. Series A, **86**, pp. 199-217, 1999

[8] B. S. He, M. Tao and X.M. Yuan, Alternating direction method with Gaussian back substitution for separable convex programming, SIAM J. Optim. **22**, pp. 13-34, 2012.

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

[10] C. H. Ye and X. M. Yuan, A Descent Method for Structured Monotone Variational Inequalities, Optimization Methods and Software, **22**, 329-338, 2007.

凸优化和单调变分不等式的收缩算法

第十九讲: 多块可分离凸优化问题部分平行正则化的交替方向类方法

Partially parallel and regularized Alternating direction method of multipliers for convex optimization containing more separable blocks

南京大学数学系　何炳生
hebma@nju.edu.cn

The context of this lecture is based on the publication [11]

1 Convex optimization problem with m-blocks

We consider the linearly constrained convex optimization with m separable operators

$$\min\Big\{\sum_{i=1}^{m}\theta_i(x_i)\;\Big|\;\sum_{i=1}^{m}A_ix_i=b,\;x_i\in\mathcal{X}_i\Big\}. \tag{1.1}$$

Its Lagrange function is

$$L(x_1,\ldots,x_m,\lambda)=\sum_{i=1}^{m}\theta_i(x_i)-\lambda^T(\sum_{i=1}^{m}A_ix_i-b), \tag{1.2}$$

which is defined on

$$\Omega:=\mathcal{X}_1\times\mathcal{X}_2\times\cdots\times\mathcal{X}_m\times\Re^l.$$

Let $\big(x_1^*,x_2^*,\ldots,x_m^*,\lambda^*\big)$ be a saddle point of the Lagrange function (1.2). Then we have

$$\begin{aligned}L_{\lambda\in\Re^l}(x_1^*,x_2^*,\cdots,x_m^*,\lambda) &\le L(x_1^*,x_2^*,\cdots,x_m^*,\lambda^*)\\ &\le L_{x_i\in\mathcal{X}_i\;(i=1,\ldots,m)}(x_1,x_2,\ldots,x_m,\lambda^*).\end{aligned}$$

It is evident that finding a saddle point of $L(x_1,x_2,\ldots,x_m,\lambda)$ is equivalent to finding

$w^* = (x_1^*, x_2^*, ..., x_m^*, \lambda^*) \in \Omega$, such that

$$\left\{\begin{array}{l} \theta_1(x_1) - \theta_1(x_1^*) + (x_1 - x_1^*)^T\{-A_1^T\lambda^*\} \geq 0, \\ \theta_2(x_2) - \theta_2(x_2^*) + (x_2 - x_2^*)^T\{-A_2^T\lambda^*\} \geq 0, \\ \qquad\qquad \vdots \\ \qquad\qquad \vdots \\ \theta_m(x_m) - \theta_m(x_m^*) + (x_m - x_m^*)^T\{-A_m^T\lambda^*\} \geq 0, \\ (\lambda - \lambda^*)^T(\sum_{i=1}^m A_i x_i^* - b) \geq 0, \end{array}\right. \tag{1.3}$$

for all $w = (x_1, x_2, \cdots, x_m, \lambda) \in \Omega$. More compactly, (1.3) can be written into the following VI:

$$w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \tag{1.4a}$$

where

$$\theta(x) = \sum_{i=1}^m \theta_i(x_i),$$

$$w = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \\ \lambda \end{pmatrix} \quad \text{and} \quad F(w) = \begin{pmatrix} -A_1^T\lambda \\ -A_2^T\lambda \\ \vdots \\ -A_m^T\lambda \\ \sum_{i=1}^m A_i x_i - b \end{pmatrix}. \tag{1.4b}$$

Note that the operator $F(w)$ defined in (1.4b) is an affine operator of and its matrix is skew-symmetric, thus, we have

$$(w - \tilde{w})^T\big(F(w) - F(\tilde{w})\big) \equiv 0, \quad \forall w, \tilde{w}. \tag{1.5}$$

Since we have assumed that the solution set of (1.1) is not empty, the solution set of (1.4), denoted by Ω^*, is also nonempty.

In addition to the notation of $w = (x_1, x_2, \cdots, x_m, \lambda)$, we also use the following notation:

$$v = (x_2, \cdots, x_m, \lambda).$$

Moreover, we define

$$\mathcal{V}^* = \{(x_2^*, \ldots, x_m^*, \lambda^*) \,|\, (x_1^*, x_2^*, \ldots, x_m^*, \lambda^*) \in \Omega^*\}.$$

The augmented Lagrangian function of the problem (1.1) is

$$\mathcal{L}_\beta(x_1,\ldots,x_m,\lambda) = L(x_1,\ldots,x_m,\lambda) + \frac{\beta}{2}\|\textstyle\sum_{i=1}^m A_i x_i - b\|^2. \tag{1.6}$$

Now, we are in the stage to describe the direct extension of ADMM to the problem (1.1).

Direct Extension of ADMM Start with given $(x_2^k,\ldots,x_m^k,\lambda^k)$,

$$\left\{\begin{array}{l}
x_1^{k+1} = \arg\min\{\mathcal{L}_\beta(x_1,x_2^k,x_3^k,\ldots,x_m^k,\lambda^k) \mid x_1\in\mathcal{X}_1\};\\
x_2^{k+1} = \arg\min\{\mathcal{L}_\beta(x_1^{k+1},x_2,x_3^k,\ldots,x_m^k,\lambda^k) \mid x_2\in\mathcal{X}_2\};\\
\qquad\vdots\\
x_i^{k+1} = \arg\min\{\mathcal{L}_\beta(x_1^{k+1},\ldots,x_{i-1}^{k+1},x_i,x_{i+1}^k,\ldots,x_m^k,\lambda^k) \mid x_i\in\mathcal{X}_i\};\\
\qquad\vdots\\
x_m^{k+1} = \arg\min\{\mathcal{L}_\beta(x_1^{k+1},\ldots,x_{m-1}^{k+1},x_m,\lambda^k) \mid x_m\in\mathcal{X}_m\};\\
\lambda^{k+1} = \lambda^k - \beta\big(\sum_{i=1}^m A_i x_i^{k+1} - b\big).
\end{array}\right. \tag{1.7}$$

There is counter example [3], it is not necessary convergent for the problem with $m \ge 3$.

2 ADMM + Prox-Parallel Splitting ALM

The following splitting method does not need correction. Its k-th iteration begins with given $v^k = (x_2^k,\ldots,x_m^k,\lambda^k)$, and obtain v^{k+1} via the following procedure:

$$\left\{\begin{array}{l}
x_1^{k+1} = \arg\min\{\mathcal{L}_\beta(x_1,x_2^k,x_3^k,\ldots,x_m^k,\lambda^k) \mid x_1\in\mathcal{X}_1\};\\
\text{for}\quad i=2,\ldots,m,\ \text{ do :}\\
x_i^{k+1} = \arg\min\limits_{x_i\in\mathcal{X}_i}\left\{\begin{array}{c}\mathcal{L}_\beta(x_1^{k+1},x_2^k\ldots,x_{i-1}^k,x_i,x_{i+1}^k,\ldots,x_m^k,\lambda^k)\\ +\frac{\tau\beta}{2}\|A_i(x_i-x_i^k)\|^2\end{array}\right\};\\
\lambda^{k+1} = \lambda^k - \beta\big(\sum_{i=1}^m A_i x_i^{k+1} - b\big)
\end{array}\right. \tag{2.1}$$

- **The $x_2\ldots x_m$-subproblems are solved in a parallel manner.**
- **To ensure the convergence, in the x_i-subproblem, $i=2,\ldots,m$, an extra proximal term $\frac{\tau\beta}{2}\|A_i(x_i-x_i^k)\|^2$ is necessary.**

An equivalent recursion of (2.1)

$\mu = \tau + 1$ and τ is given in (2.1).

$$\left\{\begin{array}{l} x_1^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x_1, x_2^k, x_3^k, \ldots, x_m^k, \lambda^k) \;\big|\; x_1 \in \mathcal{X}_1\big\}; \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta\big(A_1 x_1^{k+1} + \sum_{i=2}^m A_i x_i^k - b\big); \\ \text{for} \quad i = 2, \ldots, m, \;\; \text{do :} \\ x_i^{k+1} = \arg\min\left\{\begin{array}{c}\theta_i(x_i) - (\lambda^{k+\frac{1}{2}})^T A_i x_i \\ +\frac{\mu\beta}{2}\|A_i(x_i - x_i^k)\|^2\end{array}\;\middle|\; x_i \in \mathcal{X}_i\right\}; \\ \lambda^{k+1} = \lambda^k - \beta\big(\sum_{i=1}^m A_i x_i^{k+1} - b\big) \end{array}\right. \tag{2.2}$$

The method (2.2) is proposed in IMA Numerical Analysis [11]:

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

Equivalence of (2.1) and (2.2)

It needs only to check the optimization conditions of their x_i-subproblems for $i = 2, \ldots, m$. Note that the optimal condition of the x_i-subproblem of (2.1) is

$$\begin{aligned} x_i^{k+1} \in \mathcal{X}_i, \quad \theta_i(x_i) - \theta_i(x_i^{k+1}) + (x_i - x_i^{k+1})^T\big\{&-A_i^T\lambda^k + \\ &+\beta A_i^T\big[(A_1 x_1^{k+1} + \textstyle\sum_{j=2}^m A_j x_j^k - b) + A_i(x_i^{k+1} - x_i^k)\big] \\ &+\tau\beta A_i^T A_i(x_i^{k+1} - x_i^k)\big\} \;\geq\; 0. \end{aligned}$$

for all $x_i \in \mathcal{X}_i$. By using

$$\lambda^{k+\frac{1}{2}} = \lambda^k - \beta\big(A_1 x_1^{k+1} + \textstyle\sum_{j=2}^m A_j x_j^k - b\big); \tag{2.3}$$

it can be written as

$$\begin{aligned} x_i^{k+1} \in \mathcal{X}_i, \quad \theta_i(x_i) - \theta_i(x_i^{k+1}) + (x_i - x_i^{k+1})^T\big\{&-A_i^T\lambda^{k+\frac{1}{2}} \\ &+\beta A_i^T A_i(x_i^{k+1} - x_i^k) + \tau\beta A_i^T A_i(x_i^{k+1} - x_i^k)\big\} \;\geq\; 0. \end{aligned}$$

and consequently

$$x_i^{k+1} \in \mathcal{X}_i, \quad \theta_i(x_i) - \theta_i(x_i^{k+1}) + (x_i - x_i^{k+1})^T \{-A_i^T \lambda^{k+\frac{1}{2}} + (1+\tau)\beta A_i^T A_i (x_i^{k+1} - x_i^k)\} \geq 0, \ \forall\, x_i \in \mathcal{X}_i. \quad (2.4)$$

Setting $\mu = 1 + \tau$, (2.4) is the optimal condition of the x_i-subproblem of (2.2) !

Notice that the subproblems in (2.2) are

$$x_1^{k+1} = \arg\min\{\theta_1(x_1) + \tfrac{\beta}{2}\|A_1 x_1 + (\textstyle\sum_{i=2}^m A_i x_i^k - b) - \tfrac{1}{\beta}\lambda^k \,|\, x_1 \in \mathcal{X}_1\}$$

and

$$x_i^{k+1} = \arg\min\{\theta_i(x_i) + \tfrac{\mu\beta}{2}\|A_i(x_i - x_i^k) - \tfrac{1}{\mu\beta}\lambda^{k+\frac{1}{2}}\|^2 \,|\, x_i \in \mathcal{X}_i\},$$

for $i = 2, \ldots, m$. We assume that the above problems are not difficult to solve.

The convergence analysis is under the assumption $\mu \geq m - 1$. In the next section, Section 3, we prove the convergence of Algorithm (2.2). In Section 4, we present a prediction-correction method which uses the output of (2.2) as the predictor, and the new iterate is updated by a simple correction.

3 Convergence Analysis for the algorithm (2.2)

We use (2.2) to analyze the convergence and assume that $\boldsymbol{\mu \geq m - 1}$.

The optimal condition of the x_i-subproblems of (2.2) can be written as

$$\begin{cases} \theta_1(x_1) - \theta_1(x_1^{k+1}) + (x_1 - x_1^{k+1})^T(-A_1^T\lambda^{k+\frac{1}{2}}) \geq 0, \ \forall x_1 \in \mathcal{X}_1; \\ \theta_i(x_i) - \theta_i(x_i^{k+1}) + (x_i - x_i^{k+1})^T(-A_i^T\lambda^{k+\frac{1}{2}}) \\ \quad \geq (x_i - x_i^{k+1})^T \mu\beta A_i^T A_i (x_i^k - x_i^{k+1}), \ \forall x_i \in \mathcal{X}_i; \quad i = 2, \ldots, m. \end{cases} \quad (3.1)$$

Since

$$\lambda^{k+\frac{1}{2}} = \lambda^k - \beta\big(A_1 x_1^{k+1} + \sum_{j=2}^m A_j x_j^k - b\big)$$

and

$$\lambda^{k+1} = \lambda^k - \beta\big(\sum_{i=1}^m A_i x_i^{k+1} - b\big),$$

we have

$$-\lambda^{k+\frac{1}{2}} = -\lambda^{k+1} + \beta \sum_{j=2}^m A_j (x_j^k - x_j^{k+1}). \quad (3.2)$$

Substituting (3.2) in (3.1), we get

$$\begin{cases} \theta_1(x_1) - \theta_1(x_1^{k+1}) + (x_1 - x_1^{k+1})^T(-A_1^T\lambda^{k+1} + A_1^T p_k) \\ \qquad \geq 0, \ \forall x_1 \in \mathcal{X}_1; \\ \theta_i(x_i) - \theta_i(x_i^{k+1}) + (x_i - x_i^{k+1})^T(-A_i^T\lambda^{k+1} + A_i^T p_k) \\ \qquad \geq (x_i - x_i^{k+1})^T \mu\beta A_i^T A_i(x_i^k - x_i^{k+1}), \forall x_i \in \mathcal{X}_i; \\ \qquad i = 2, \ldots, m. \end{cases} \tag{3.3}$$

where

$$p_k = \beta \sum_{j=2}^{m} A_j(x_j^k - x_j^{k+1}). \tag{3.4}$$

Since

$$\Big(\sum_{i=1}^{m} A_i x_i^{k+1} - b\Big) = (1/\beta)(\lambda^k - \lambda^{k+1}),$$

It can be written as

$$(\lambda - \lambda^{k+1})^T\Big(\sum_{i=1}^{m} A_i x_i^{k+1} - b\Big) \geq (\lambda - \lambda^{k+1})^T(1/\beta)(\lambda^k - \lambda^{k+1}), \tag{3.5}$$

for all $\lambda \in \Re^l$.

Combining (3.3) and (3.5)

$$\begin{cases} \theta_1(x_1) - \theta_1(x_1^{k+1}) + (x_1 - x_1^{k+1})^T(-A_1^T\lambda^{k+1} + A_1^T p_k) \\ \qquad \geq 0, \ \forall x_1 \in \mathcal{X}_1; \\ \theta_i(x_i) - \theta_i(x_i^{k+1}) + (x_i - x_i^{k+1})^T(-A_i^T\lambda^{k+1} + A_i^T p_k) \\ \qquad \geq (x_i - x_i^{k+1})^T \mu\beta A_i^T A_i(x_i^k - x_i^{k+1}), \forall x_i \in \mathcal{X}_i; \\ \qquad i = 2, \ldots, m. \\ (\lambda - \lambda^{k+1})^T\big(\sum_{i=1}^{m} A_i x_i^{k+1} - b\big) \\ \quad \geq (\lambda - \lambda^{k+1})^T(1/\beta)(\lambda^k - \lambda^{k+1}). \quad \forall \lambda \in \Re^l. \end{cases} \tag{3.6}$$

By using the notations $\theta(x)$ and $F(w)$, it follows from (3.6) that

$$\begin{aligned}
&\theta(x) - \theta(x^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) + p_k^T \big(\sum_{i=1}^m A_i(x_i - x_i^{k+1})\big) \\
&\quad \geq \sum_{i=2}^m (x_i - x_i^{k+1})^T \mu\beta A_i^T A_i (x_i^k - x_i^{k+1}) \\
&\qquad + (\lambda - \lambda^{k+1})^T (1/\beta)(\lambda^k - \lambda^{k+1}), \quad \forall w \in \Omega.
\end{aligned}$$

Setting $w = w^*$ in the above inequality and by a manipulation, we get

$$\begin{aligned}
&\sum_{i=2}^m (x_i^{k+1} - x^*)^T \mu\beta A_i^T A_i (x_i^k - x_i^{k+1}) + (\lambda^{k+1} - \lambda^*)^T \frac{1}{\beta}(\lambda^k - \lambda^{k+1}) \\
&\quad \geq \quad \theta(x^{k+1}) - \theta(x^*) + (w^{k+1} - w^*)^T F(w^{k+1}) \\
&\qquad + p_k^T \big(\sum_{i=1}^m A_i(x_i^{k+1} - x_i^*)\big) \qquad (3.7)
\end{aligned}$$

Now, we treat the right hand side of (3.7). Using (1.5) and the optimality, we have

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

Because $\sum_{i=1}^m A_i x_i^* = b$ and $\sum_{i=1}^m A_i x_i^{k+1} - b = \frac{1}{\beta}(\lambda^k - \lambda^{k+1})$, using the definition of p_k (see (3.4)), we get

$$p_k^T \big(\sum_{i=1}^m A_i(x_i^{k+1} - x^*)\big) = (\lambda^k - \lambda^{k+1})^T \sum_{j=2}^m A_j(x_j^k - x_j^{k+1}).$$

Substituting it in (3.7), we get

$$\begin{aligned}
&\sum_{i=2}^m (x_i^{k+1} - x^*)^T \mu\beta A_i^T A_i (x_i^k - x_i^{k+1}) + (\lambda^{k+1} - \lambda^*)^T \frac{1}{\beta}(\lambda^k - \lambda^{k+1}) \\
&\quad \geq \quad (\lambda^k - \lambda^{k+1})^T \sum_{j=2}^m A_j(x_j^k - x_j^{k+1}). \qquad (3.8)
\end{aligned}$$

By denoting

$$v^k = \begin{pmatrix} x_2^k \\ \vdots \\ x_m^k \\ \lambda^k \end{pmatrix} \quad \text{and} \quad G = \begin{pmatrix} \mu\beta A_2^T A_2 & & & \\ & \ddots & & \\ & & \mu\beta A_m^T A_m & \\ & & & \frac{1}{\beta} I \end{pmatrix}, \tag{3.9}$$

we get the following assertion:

Lemma 3.1 *Let v^{k+1} be generated by* (2.2) *from the given vector v^k, then we have*

$$(v^k - v^*)^T G(v^k - v^{k+1}) \geq \varphi(v^k, v^{k+1}), \tag{3.10}$$

where

$$\begin{aligned} \varphi(v^k, v^{k+1}) = & \|v^k - v^{k+1}\|_G^2 \\ & + (\lambda^k - \lambda^{k+1})^T \big(\textstyle\sum_{i=2}^m A_i(x_i^k - x_i^{k+1})\big). \end{aligned} \tag{3.11}$$

Proof. Using the notations (3.9), the inequality (3.8) can be written as

$$(v^{k+1} - v^*)^T G(v^k - v^{k+1}) \geq (\lambda^k - \lambda^{k+1})^T \sum_{j=2}^m A_j(x_j^k - x_j^{k+1}).$$

Adding $(v^k - v^{k+1})^T G(v^k - v^{k+1})$ to the both sides of the above inequality, we get the assertion directly. $\square$

Now we consider the profit of the k-th iteration. Using (3.10) and (3.11), we have

$$\begin{aligned} & \|v^k - v^*\|_G^2 - \|v^{k+1} - v^*\|_G^2 \\ = \quad & \|v^k - v^*\|_G^2 - \|(v^k - v^*) - (v^k - v^{k+1})\|_G^2 \\ = \quad & 2(v^k - v^*)^T G(v^k - v^{k+1}) - \|v^k - v^{k+1}\|_G^2 \\ \geq \quad & \|v^k - v^{k+1}\|_G^2 + 2(\lambda^k - \lambda^{k+1})^T \sum_{j=2}^m A_j(x_j^k - x_j^{k+1}). \end{aligned} \tag{3.12}$$

In the following is to show that, **when $\mu > m - 1$, there is a constant $\sigma > 0$, such that the right hand side of** (3.12) **is greater than** $\sigma\|v^k - v^{k+1}\|_G^2$.

According to the definition of the matrix G, we have

$$\|v^k - v^{k+1}\|_G^2 + 2(\lambda^k - \lambda^{k+1})^T \Big(\sum_{i=2}^m A_i(x_i^k - x_i^{k+1})\Big)$$

$$= \begin{pmatrix} \sqrt{\beta}A_2(x_2^k - x_2^{k+1}) \\ \sqrt{\beta}A_3(x_3^k - x_3^{k+1}) \\ \vdots \\ \sqrt{\beta}A_m(x_m^k - x_m^{k+1}) \\ (1/\sqrt{\beta})(\lambda^k - \lambda^{k+1}) \end{pmatrix}^T \begin{pmatrix} \mu I_l & 0 & \cdots & 0 & I_l \\ 0 & \mu I_l & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & I_l \\ 0 & \cdots & 0 & \mu I_l & I_l \\ I_l & \cdots & I_l & I_l & I_l \end{pmatrix}$$

$$\begin{pmatrix} \sqrt{\beta}A_2(x_2^k - x_2^{k+1}) \\ \sqrt{\beta}A_3(x_3^k - x_3^{k+1}) \\ \vdots \\ \sqrt{\beta}A_m(x_m^k - x_m^{k+1}) \\ (1/\sqrt{\beta})(\lambda^k - \lambda^{k+1}) \end{pmatrix}. \tag{3.13}$$

Notice that the block-wise matrix

$$\begin{pmatrix} \mu I_l & 0 & \cdots & 0 & I_l \\ 0 & \mu I_l & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & I_l \\ 0 & \cdots & 0 & \mu I_l & I_l \\ I_l & \cdots & I_l & I_l & I_l \end{pmatrix}$$

in (3.13) have the same largest (resp. smallest) eigenvalues as the $m \times m$ symmetric matrix

$$M = \begin{pmatrix} \mu & 0 & \cdots & 0 & 1 \\ 0 & \mu & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 1 \\ 0 & \cdots & 0 & \mu & 1 \\ 1 & \cdots & 1 & 1 & 1 \end{pmatrix}_{m\times m}. \tag{3.14}$$

Lemma 3.2 *For $m \geq 2$, the $m \times m$ symmetric matrix M defined in* (3.14) *has $(m-2)$ multiple eigenvalues*

$$\nu_1 = \nu_2 = \cdots = \nu_{m-2} = \mu,$$

and another two eigenvalues

$$\nu_{m-1}, \nu_m = \frac{1}{2}\left[(\mu+1) \pm \sqrt{(\mu+1)^2 + 4((m-1)-\mu)}\right].$$

Proof. Let e be a $(m-1)$-vector whose each element equals 1. Thus

$$M = \begin{pmatrix} \mu I_{m-1} & e \\ e^T & 1 \end{pmatrix}.$$

Without loss of generality, we assume that the eigenvectors of M have forms

$$z = \begin{pmatrix} y \\ 0 \end{pmatrix} \quad \text{or} \quad z = \begin{pmatrix} y \\ 1 \end{pmatrix}, \quad \text{where } y \in \Re^{m-1}.$$

In the first case, we have

$$\begin{cases} \mu y = \nu y, \\ e^T y = 0. \end{cases} \tag{3.15}$$

It is clear that the $(m-1)$-vectors

$$y^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ -1 \end{pmatrix}, \quad y^2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ -1 \end{pmatrix} \quad \cdots\cdots \quad y^{m-2} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ -1 \end{pmatrix}$$

are linear independent and satisfy (3.15) with $\nu = \mu$. Thus,

$$z^i = \begin{pmatrix} y^i \\ 0 \end{pmatrix}, \quad i = 1, \ldots, m-2,$$

are eigenvectors of M and the related eigenvalue

$$\nu_1 = \nu_2 = \cdots = \nu_{m-2} = \mu.$$

In the second case, $z^T = (y^T, 1)$, we have

$$\begin{cases} \mu y + e = \nu y, \\ e^T y + 1 = \nu. \end{cases} \tag{3.16}$$

It follows from (3.16) that

$$(\nu - \mu)(\nu - 1) - (m - 1) = 0,$$

and thus

$$\nu_{m-1}, \nu_m = \frac{1}{2}\big[(\mu + 1) \pm \sqrt{(\mu + 1)^2 + 4((m - 1) - \mu)}\big].$$

The lemma is proved. $\square$

For $\mu \geq 1$, it is easy to verify that

$$\nu_m = \frac{(\mu + 1) - \sqrt{(\mu + 1)^2 + 4((m - 1) - \mu)}}{2} \tag{3.17}$$

is the smallest eigenvalue of M. For fixed $\mu > (m - 1)$, there is a σ such that $\nu_m > \sigma$. Together with (3.12) and (3.13), we have the following assetions:

Lemma 3.3 *Let $\mu > m - 1$, then there is a $\sigma > 0$ such that*

$$\|v^k - v^{k+1}\|_G^2 + 2(\lambda^k - \lambda^{k+1})^T \big(\sum_{i=2}^{m} A_i(x_i^k - x_i^{k+1})\big) \geq \sigma \|v^k - v^{k+1}\|_G^2, \tag{3.18}$$

where G is defined in (3.9)*.*

Theorem 3.1 *Let $\mu > m - 1$ and $\{v^k\}$ be the sequence generated by* (2.2)*, then there is a $\sigma > 0$ such that*

$$\|v^{k+1} - v^*\|_G^2 \leq \|v^k - v^*\|_G^2 - \sigma \|v^k - v^{k+1}\|_G^2, \quad \forall\, v^* \in \mathcal{V}^*. \tag{3.19}$$

where G is defined in (3.9)*.*

The inequality (3.19) is is the key for convergence of the method (2.1) and (2.2).

Implementation of the method for three block problems

For the problem with three separable operators

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

we have

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

For given $v^k=(y^k,z^k,\lambda^k)$, by using the method proposed in this subsection, the new iterate $v^{k+1}=(y^{k+1},z^{k+1},\lambda^{k+1})$ is obtained via ($\tau\geq 1$) :

$$\left\{\begin{array}{l} x^{k+1}=\text{Argmin}\{\mathcal{L}_\beta^3(x,y^k,z^k,\lambda^k)\,|\,x\in\mathcal{X}\},\\ y^{k+1}=\text{Argmin}\{\mathcal{L}_\beta^3(x^{k+1},y,z^k,\lambda^k)+\frac{\tau\beta}{2}\|B(y-y^k)\|^2\,|\,y\in\mathcal{Y}\},\\ z^{k+1}=\text{Argmin}\{\mathcal{L}_\beta^3(x^{k+1},y^k,z,\lambda^k)+\frac{\tau\beta}{2}\|C(z-z^k)\|^2\,|\,z\in\mathcal{Z}\},\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b),\end{array}\right. \tag{3.20}$$

An equivalent recursion of (3.20) is

$$\left\{\begin{array}{l} x^{k+1}=\text{Argmin}\{\mathcal{L}_\beta^3(x,y^k,z^k,\lambda^k)\,|\,x\in\mathcal{X}\},\\ \lambda^{k+\frac{1}{2}}=\lambda^k-\beta(Ax^{k+1}+By^k+Cz^k-b)\\ y^{k+1}=\text{Argmin}\{\theta_2(y)-(\lambda^{k+\frac{1}{2}})^TBy+\frac{\mu\beta}{2}\|B(y-y^k)\|^2\,|\,y\in\mathcal{Y}\},\\ z^{k+1}=\text{Argmin}\{\theta_3(z)-(\lambda^{k+\frac{1}{2}})^TCz+\frac{\mu\beta}{2}\|C(z-z^k)\|^2\,|\,z\in\mathcal{Z}\},\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b),\end{array}\right. \tag{3.21}$$

where $\mu=\tau+1\geq 2$. Implementation of (3.21) is via

$$\left\{\begin{array}{l} x^{k+1}=\text{Argmin}\{\theta_1(x)+\frac{\beta}{2}\|Ax+[By^k+Cz^k-b-\frac{1}{\beta}\lambda^k]\|^2\,|\,x\in\mathcal{X}\},\\ \lambda^{k+\frac{1}{2}}=\lambda^k-\beta(Ax^{k+1}+By^k+Cz^k-b)\\ y^{k+1}=\text{Argmin}\{\theta_2(y)+\frac{\mu\beta}{2}\|By-[By^k+\frac{1}{\mu\beta}\lambda^{k+\frac{1}{2}}]\|^2\,|\,y\in\mathcal{Y}\},\\ z^{k+1}=\text{Argmin}\{\theta_3(z)+\frac{\mu\beta}{2}\|Cz-[Cz^k+\frac{1}{\mu\beta}\lambda^{k+\frac{1}{2}}]\|^2\,|\,z\in\mathcal{Z}\},\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^{k+1}+Cz^{k+1}-b).\end{array}\right.$$

4 Method with the calculated stepsize

The iteration of the method (2.1) and/or (2.2) begin with $v^k = (x_2^k, \cdots, \lambda^k)$ and finish with $v^{k+1} = (x_2^{k+1}, \cdots x_m^{k+1}, \lambda^{k+1})$. In this section, we consider the method with the calculated step-size. In practice, we use the output of (2.2) as a predictor.

$$\left\{\begin{array}{l} x_1^{k+1} = \arg\min\big\{\mathcal{L}_\beta(x_1, x_2^k, x_3^k, \ldots, x_m^k, \lambda^k) \;\big|\; x_1 \in \mathcal{X}_1\big\}; \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta\big(A_1 x_1^{k+1} + \sum_{i=2}^m A_i x_i^k - b\big); \\ \text{for} \quad i = 2, \ldots, m, \;\; \text{do :} \\ \tilde{x}_i^k = \arg\min \left\{ \begin{array}{c} \theta_i(x_i) - (\lambda^{k+\frac{1}{2}})^T A_i x_i \\ + \frac{\mu\beta}{2}\|A_i(x_i - x_i^k)\|^2 \end{array} \middle| \; x_i \in \mathcal{X}_i \right\}; \\ \tilde{\lambda}^k = \lambda^k - \beta\big(A_1 x^{k+1} + \sum_{i=2}^m A_i \tilde{x}_i^k - b\big) \end{array}\right. \tag{4.1}$$

We only denote the output $v^{k+1} = (x_2^{k+1}, \cdots x_m^{k+1}, \lambda^{k+1})$ generated from (2.2) by using the new notations $\tilde{v}^k = (\tilde{x}_2^k, \cdots, \tilde{x}_m^k, \tilde{\lambda}^k)$. After getting $\tilde{v}^k$, we

offer thenew iterate v^{k+1} by $v^{k+1} = v^k - \alpha_k(v^k - \tilde{v}^k)$.

Algorithm 2: a prediction-correction splitting method for solving (1.1)

Step 1. Prediction step. From the given $v^k = (x_2^k, \cdots, x_m^k, \lambda^k)$, using (4.1) to produce the predictor $\tilde{v}^k = (\tilde{x}_2^k, \cdots, \tilde{x}_m^k, \tilde{\lambda}^k)$.

Step 2. Correction step. The new iterate $v^{k+1} = (x_2^{k+1}, \cdots x_m^{k+1}, \lambda^{k+1})$ is updated via:

$$v^{k+1} = v^k - \alpha_k(v^k - \tilde{v}^k), \tag{4.2}$$

where

$$\alpha_k = \gamma\alpha_k^*, \qquad \gamma_k \in (0, 2), \qquad \alpha_k^* = \frac{\varphi(v^k, \tilde{v}^k)}{\|v^k - \tilde{v}^k\|_G^2} \tag{4.3}$$

and

$$\varphi(v^k, \tilde{v}^k) = \|v^k - \tilde{v}^k\|_G^2 + (\lambda^k - \tilde{\lambda}^k)^T\big(\sum_{i=2}^m A_i(x_i^k - \tilde{x}_i^k)\big). \tag{4.4}$$

As we can see easily, Algorithm 1 (2.2) turns out to be a special case of Algorithm

2 where $\gamma_k \equiv 1/\alpha_k$ in (4.3). Thus, in the following, we prove the convergence for Algorithm 2, from which the convergence of Algorithm 1 becomes trivial.

Since the $\tilde{v}^k$ in (4.1) is the same of v^{k+1} in Algorithm (2.2), similarly as in Lemma 3.1, we have the following assertion directly.

Lemma 4.1 *Let $\tilde{v}^k$ be generated by* (4.1) *from the given vector v^k, then we have*

$$(v^k - v^*)^T G(v^k - \tilde{v}^k) \geq \varphi(v^k, \tilde{v}^k), \tag{4.5}$$

where $\varphi(v^k, \tilde{v}^k)$ is defined in (4.4)*.*

Lemma 4.2 *Under the assumption $\mu > m - 1$, it holds that*

$$\varphi(v^k, \tilde{v}^k) \geq \frac{1+\sigma}{2}\|v^k - \tilde{v}^k\|_G^2. \tag{4.6}$$

Proof. According to the definition of $\varphi(v^k, \tilde{v}^k)$ (see (4.4)) and the inequality

(3.18) in Lemma 3.3, we have

$$\begin{aligned} 2\varphi(v^k, \tilde{v}^k) &= 2\|v^k - \tilde{v}^k\|_G^2 + 2(\lambda^k - \tilde{\lambda}^k)^T\Big(\sum_{i=2}^m A_i(x_i^k - \tilde{x}_i^k)\Big) \\ &\geq (1+\sigma)\|v^k - \tilde{v}^k\|_G^2, \end{aligned}$$

and the assertion follows from the definitions of $\varphi(v^k, \tilde{v}^k)$ and α_k^* (see (4.3) and (3.11)) directly. $\square$

For determinate the step size α_k in (4.2), we define the step-size dependent new iterate by

$$v^{k+1}(\alpha) = v^k - \alpha(v^k - \tilde{v}^k), \tag{4.7}$$

In this way,

$$\vartheta(\alpha) = \|v^k - v^*\|_G^2 - \|v^{k+1}(\alpha) - v^*\|_G^2 \tag{4.8}$$

is the distance decrease functions in the k-th iteration by using updating form (4.7). By defining

$$q(\alpha) = 2\alpha\varphi(v^k, \tilde{v}^k) - \alpha^2\|v^k - \tilde{v}^k\|_G^2. \tag{4.9}$$

It follows from (4.7), (4.8) and (4.5) that

$$\begin{aligned}\vartheta(\alpha) &= \|v^k - v^*\|_G^2 - \|v^k - v^* - \alpha(v^k - \tilde{v}^k)\|_G^2 \\ &= 2\alpha(v^k - v^*)^T G(v^k - \tilde{v}^k) - \alpha^2\|v^k - \tilde{v}^k\|_G^2 \\ &\geq 2\alpha\varphi(v^k, \tilde{v}^k) - \alpha^2\|v^k - \tilde{v}^k\|_G^2 \\ &= q(\alpha). \end{aligned} \tag{4.10}$$

Note that $q(\alpha)$ is a quadratic function of α, it reaches its maximum at

$$\alpha_k^* = \frac{\varphi(v^k, \tilde{v}^k)}{\|v^k - \tilde{v}^k\|_G^2}, \tag{4.11}$$

and this is just the same as defined in (4.3). Usually, in practical computation, taking a relaxed factor $\gamma > 1$ is useful for fast convergence.

Theorem 4.1 *Let $\{v^k\}$ be the sequence generated by Algorithm 2. We have*

$$\|v^{k+1} - v^*\|_G^2 \leq \|v^k - v^*\|_G^2 - \frac{\gamma(2-\gamma)}{4}\|v^k - \tilde{v}^k\|_G^2, \quad \forall\, v^* \in \mathcal{V}^*. \tag{4.12}$$

Proof. It follows from (4.8) and (4.10) that

$$\|v^{k+1} - v^*\|_G^2 \leq \|v^k - v^*\|_G^2 - q(\gamma\alpha_k^*), \quad \forall v^* \in \mathcal{V}^*. \tag{4.13}$$

By using (4.9) and (4.11) we obtain

$$\begin{aligned} q(\gamma\alpha_k^*) &= 2\gamma\alpha_k^*\varphi(v^k, \tilde{v}^k) - (\gamma\alpha_k^*)^2\|v^k - \tilde{v}^k\|_G^2 \\ &= \gamma(2-\gamma)\alpha_k^*\varphi(v^k, \tilde{v}^k). \end{aligned} \tag{4.14}$$

Since (see (4.6))

$$\varphi(v^k, \tilde{v}^k) > \frac{1}{2}\|v^k - \tilde{v}^k\|_G^2$$

and consequently (see (4.3)),

$$\alpha_k^* > \frac{1}{2}.$$

Thus, we have

$$\alpha_k^*\varphi(v^k, \tilde{v}^k) \geq \frac{1}{4}\|v^k - \tilde{v}^k\|_G^2.$$

Substituting it in (4.14), the proof of this theorem is complete. □

Theorem 4.1 offers the key inequality for the convergence !

References

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

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

[3] C. H. Chen, B. S. He, Y. Y. Ye and X. M. Yuan, *The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent*, to appear in *Mathematical Programming, Series A*.

[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, A class of projection and contraction methods for monotone variational inequalities, Applied Mathematics and Optimization, **35**, 69-76, 1997.

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

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

[9] 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.

[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 J. Numerical Analysis* **31**(2015), 394-426.

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

[13] 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.

[14] 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.

[15] 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.

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

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

凸优化和单调变分不等式的收缩算法

第二十讲: 单调变分不等式意义下 凸优化分裂收缩算法的统一框架

A uniform framework of splitting contraction methods for convex optimization in sense of monotone variational inequality

南京大学数学系　　何炳生
hebma@nju.edu.cn

The context of this lecture is based on the publication [21]

1 Splitting Methods in a Unified Framework

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:

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

The analysis can be fund in [17] (Sections 4 and 5 therin). In order to illustrate the unified framework, let us restudy the augmented Lagrangian method.

1.1 Extended Augmented Lagrangian Method

For the convex optimization

$$\min\{\theta(x) \mid Ax = b,\, x \in \mathcal{X}\}. \tag{1.2}$$

Its augmented Lagrangian function is

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

where the quadratic term is the penalty for the linear constraints $Ax = b$. The k-th iteration of the **Augmented Lagrangian Method** [28, 32] begins with a given λ^k, obtain

$w^{k+1} = (x^{k+1}, \lambda^{k+1})$ via

$$\text{(ALM)} \quad \begin{cases} x^{k+1} = \arg\min\{\mathcal{L}_\beta(x, \lambda^k) \mid x \in \mathcal{X}\}, & \text{(1.3a)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} - b). & \text{(1.3b)} \end{cases}$$

In (1.3), x^{k+1} is only a computational result of (1.3a) from given λ^k, it is called the intermediate variable. If we denote the output of (1.3) by $\tilde{w}^k = (\tilde{x}^k, \tilde{\lambda}^k)$, then the optimal condition can be written as $\tilde{w}^k \in \Omega$ and

$$\begin{cases} \theta(x) - \theta(\tilde{x}^k) \;+\; (x - \tilde{x}^k)^T\{-A^T\lambda^k + \beta A^T(A\tilde{x}^k - b)\} \geq 0, \;\; \forall x \in \mathcal{X}, \\ \qquad\qquad\qquad\qquad (\lambda - \tilde{\lambda}^k)^T\{(A\tilde{x}^k - b) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \;\; \forall \lambda \in \Re^m. \end{cases}$$

The above relation can be written as

$$\theta(x) - \theta(\tilde{x}^k) + \begin{pmatrix} x - \tilde{x}^k \\ \lambda - \tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} -A^T\tilde{\lambda}^k \\ A\tilde{x}^k - b \end{pmatrix} \geq (\lambda - \tilde{\lambda}^k)^T \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k), \quad \text{(1.4a)}$$

for all $w \in \Omega$. In the classical augmented Lagrangian method, $\lambda^{k+1} = \tilde{\lambda}^k$. In practice, we can use relaxation techniques and offer the new iterate by

$$\lambda^{k+1} = \lambda^k - \alpha(\lambda^k - \tilde{\lambda}^k), \quad \alpha \in (0, 2). \quad \text{(1.4b)}$$

Setting $w = w^*$ in (1.4a), we get $(\tilde{\lambda}^k - \lambda^*)^T \frac{1}{\beta}(\lambda^k - \tilde{\lambda}^k) \geq 0$ and thus

$$(\lambda^k - \lambda^*)^T(\lambda^k - \tilde{\lambda}^k) \geq \|\lambda^k - \tilde{\lambda}^k\|^2. \quad \text{(1.5)}$$

By using (1.4b) and (1.5), we get

$$\|\lambda^{k+1} - \lambda^*\|^2 \leq \|\lambda^k - \lambda^*\|^2 - \alpha(2 - \alpha)\|\lambda^k - \tilde{\lambda}^k\|^2.$$

In order to describe the algorithm prototype, we give the following definition.

Definition (Intermediate variables and Essential Variables)

For an iterative algorithm solving VI(Ω, F, θ), if some coordinates of w are not involved in the beginning of each iteration, then these coordinates are called intermediate variables and those required by the iteration are called essential variables (denoted by v).

- The sub-vector $w \backslash v$ is called intermediate variables.
- In some Algorithms, v is a proper sub-vector of w; however, $v = w$ is also possible.

According to the above mentioned definition, in the the augmented Lagrangian method, x is an intermediate variable and λ is the essential variable.

1.2 Algorithms in a unified framework

A Prototype Algorithm for (1.1)

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

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall w \in \Omega, \quad (1.6a)$$

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

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

$$v^{k+1} = v^k - \alpha M(v^k - \tilde{v}^k). \quad (1.6b)$$

- Usually, we do not take the output of (1.6a), $\tilde{v}^k$, as the new iterate. Thus, $\tilde{v}^k$ is called the predictor. The new iterate v^{k+1} given by (1.6b) is called the corrector. The prototype algorithm is a Prediction-Correction Method.
- We use the extended ALM in Section 1.1 as an example. In (1.4a) we have $v = \lambda$, $Q = \frac{1}{\beta} I$, while in the correction step (1.4b), $M = I$ and $\alpha \in (0, 2)$.
- When $v^k = \tilde{v}^k$, it follows from (1.6a) directly that $\tilde{w}^k$ is a solution of (1.1). Thus, one can use $\|v^k - \tilde{v}^k\| < \epsilon$ as the stopping criterion in (1.6).

Convergence Conditions

For the matrices Q and M, and the step size α determined in (1.6), the matrices

$$H = QM^{-1} \quad (1.7a)$$

and

$$G = Q^T + Q - \alpha M^T H M. \quad (1.7b)$$

are positive definite (or $H \succ 0$ and $G \succeq 0$).

- We use the extended ALM in Section 1.1 as an example. Since $Q = \frac{1}{\beta} I$ in the prediction step, and $M = I$ and $\alpha \in (0, 2)$ in the correction step, it follows that
$$H = QM^{-1} = \frac{1}{\beta} I \quad \text{and} \quad G = Q^T + Q - \alpha M^T H M = \frac{2-\alpha}{\beta} I.$$
Therefore, the convergence conditions are satisfied.
- For $G \succeq 0$, it has the $O(1/t)$ convergence rate in a ergodic sense. If $G \succ 0$, the sequence $\{v^k\}$ is Fèjer monotone and converges to a $v^* \in \mathcal{V}^*$ in H-norm.
- Using the unified framework, the convergence proof is very simple. In addition, it will help us to construct more efficient splitting contraction method for convex optimization with different structures.

Given a positive definite matrix Q in (1.6a) ($Q^T + Q \succ 0$), for satisfying the convergence conditions (1.7), how to choose the matrix M and $\alpha > 0$ in the correction step (1.6b) ?

There are many possibilities, the principle is simplicity and efficiency. See an example:

- In order to ensure the symmetry and positivity of $H = QM^{-1}$, we take

$$H = QD^{-1}Q^T,$$

where D is a symmetric investable block diagonal matrix. Because

$$H = QD^{-1}Q^T \qquad \text{and} \qquad H = QM^{-1},$$

we only need to set $M^{-1} = D^{-1}Q^T$ and thus

$$M = Q^{-T}D \quad \text{satisfies the condition (1.7a).}$$

In this case, $M^T HM = Q^T M = Q^T Q^{-T} D = D$.

- After choosing the matrix M, let

$$\alpha_{\max} = \arg\max\{\alpha \,|\, Q^T + Q - \alpha M^T HM \succeq 0\},$$

the condition (1.7b) is satisfied for any $\alpha \in (0, \alpha_{\max})$.

1.3 Customized PPA satisfies the Convergence Condition

Recall the convex optimization problem discussed in Lecture 6, namely,

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

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

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

where

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

For given $v^k = w^k = (x^k, \lambda^k)$, let the output of the Customized PPA as a predictor and denote it as $\tilde{w}^k = (\tilde{x}^k, \tilde{\lambda}^k)$. Then, we have

$$\left\{ \begin{array}{l} \tilde{x}^k = \arg\min\big\{\theta(x) - (\lambda^k)^T(Ax - b) + \frac{r}{2}\|x - x^k\|^2 \;\big|\; x \in \mathcal{X}\big\}, \\ \tilde{\lambda}^k = \lambda^k - \frac{1}{s}[A(2\tilde{x}^k - x^k) - b]. \end{array} \right. \tag{1.8}$$

Similar as (1.4), the output $\tilde{w}^k \in \Omega$ of the iteration (1.8) satisfies

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

It is a form of (1.6a) where

$$Q = H = \begin{pmatrix} rI & A^T \\ A & sI \end{pmatrix}.$$

This matrix is positive definite when $rs > \|A^T A\|$. We take $M = I$ in the correction (1.6b) and the new iterate is updated by

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

Then, we have and

$$H = QM^{-1} = Q \succ 0 \quad \text{and} \quad G = Q^T + Q - \alpha M^T H M = (2 - \alpha) H \succ 0.$$

The convergence conditions (1.7) are satisfied. More about customized PPA, please see

♣ G.Y. Gu, B.S. He and X.M. Yuan, Customized Proximal point algorithms for linearly constrained convex minimization and saddle-point problem: a unified Approach, Comput. Optim. Appl., 59(2014), 135-161.

1.4 Primal-Dual relaxed PPA-based Contraction Methods

For solving (1.2), with given $v^k = w^k = (x^k, \lambda^k)$, let $\tilde{w}^k = (\tilde{x}^k, \tilde{\lambda}^k)$ be produced by

$$\tilde{x}^k = \arg\min\{\theta(x) - (\lambda^k)^T(Ax - b) + \tfrac{r}{2}\|x - x^k\|^2 \mid x \in \mathcal{X}\} \tag{1.9a}$$

and (according to equality constraints $Ax = b$ or inequality constranits $Ax \ge b$)

$$\tilde{\lambda}^k = \lambda^k - \frac{1}{s}(A\tilde{x}^k - b) \quad \text{or} \quad \tilde{\lambda}^k = [\lambda^k - \frac{1}{s}(A\tilde{x}^k - b)]_+. \tag{1.9b}$$

The predictor $\tilde{w}^k \in \Omega$ generated by (1.9) satisfies

$$\theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \forall w \in \Omega, \tag{1.10}$$

where the matrix

$$Q = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix}, \tag{1.11}$$

is not symmetric. However, (1.10) can be viewed as (1.6a). In this subsection, all the mentioned matrix Q is (1.11). There is example to show that the method is not necessary convergent if we directly take $w^{k+1} = \tilde{w}^k$.

Corrector—the new iterate For given v^k and the predictor $\tilde{v}^k$ by (1.9), we use

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \tag{1.12}$$

to produce the new iterate, where

$$M = \begin{pmatrix} I_n & \frac{1}{r}A^T \\ 0 & I_m \end{pmatrix}$$

is a upper triangular block matrix whose diagonal part is unit matrix. Note that

$$H = QM^{-1} = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \begin{pmatrix} I_n & -\frac{1}{r}A^T \\ 0 & I_m \end{pmatrix} = \begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix} \succ 0.$$

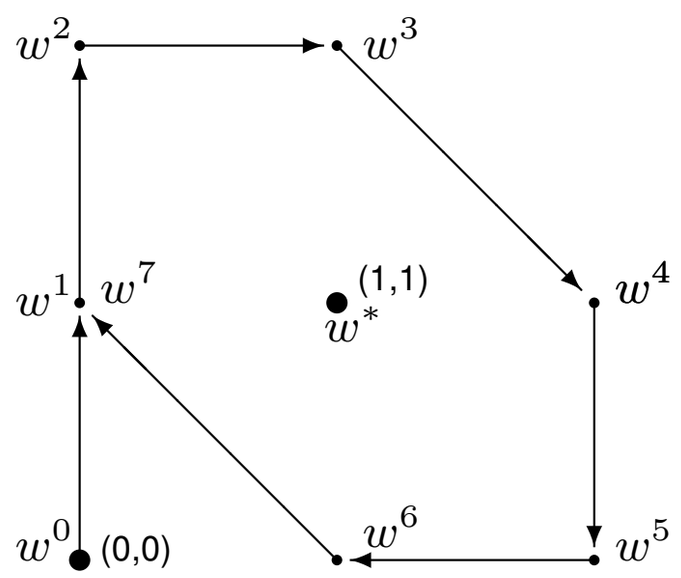
In addition,

$$\begin{aligned} G &= Q^T + Q - M^T H M = Q^T + Q - Q^T M \\ &= \begin{pmatrix} rI_n & 0 \\ 0 & sI_m - \frac{1}{r}AA^T \end{pmatrix}. \end{aligned}$$

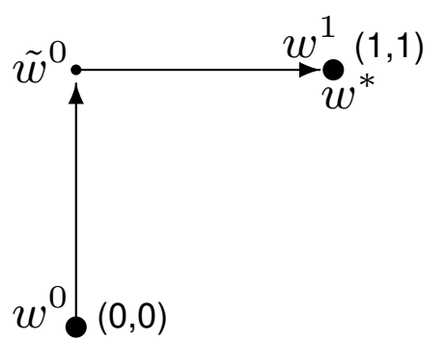
G is positive definite when $rs > \|A^T A\|$. The convergence conditions (1.7) are satisfied.

Convergence behaviors for LP

A Toy Example: $\min\{x \,|\, x = 1,\ x \geq 0\}$, $(x^*, y^*) = (1, 1)$.



Original PDHG



PDHG + Correction

This example shows, sometimes the correction has surprising effectiveness.

In the correction step (1.12), the matrix M is a upper-triangular matrix. We can also use the lower-triangular matrix

$$M = \begin{pmatrix} I_n & 0 \\ -\frac{1}{s}A & I_m \end{pmatrix}$$

According to (1.7a), $H = QM^{-1}$, by a simple computation, we have

$$H = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix} \begin{pmatrix} I_n & 0 \\ \frac{1}{s}A & I_m \end{pmatrix} = \begin{pmatrix} rI_n + \frac{1}{s}A^TA & A^T \\ A & sI_m \end{pmatrix}.$$

H is positive definite for any $r, s > 0$. In addition,

$$\begin{aligned} G &= Q^T + Q - M^THM = Q^T + Q - Q^TM \\ &= \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix} = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}. \end{aligned}$$

G is positive definite when $rs > \|A^TA\|$. The convergence conditions (1.7) are satisfied.

For a given prediction, there are different corrections which satisfy the convergence conditions (1.7). For example, we can take a convex combination of the above mentioned

matrices. Namely, for $\tau \in [0, 1]$

$$\begin{aligned} M &= (1-\tau)\begin{pmatrix} I_n & \frac{1}{r}A^T \\ 0 & I_m \end{pmatrix} + \tau \begin{pmatrix} I_n & 0 \\ -\frac{1}{s}A & I_m \end{pmatrix} \\ &= \begin{pmatrix} I_n & \frac{1-\tau}{r}A^T \\ -\frac{\tau}{s}A & I_m \end{pmatrix}. \end{aligned}$$

For this matrix M, we denote

$$\Pi = I + \frac{\tau(1-\tau)}{rs}AA^T.$$

Clearly, Π is positive definite. Let

$$H = \begin{pmatrix} rI_n + \frac{\tau^2}{s}A^T\Pi^{-1}A & \tau A^T\Pi^{-1} \\ \tau\Pi^{-1}A & s\Pi^{-1} \end{pmatrix}.$$

It is easy to verify that H is positive definite for any $r, s > 0$ and

$$HM = Q.$$

Now, we turn to observe the matrix G, it leads that

$$\begin{aligned} G &= Q^T + Q - M^T H M = Q^T + Q - Q^T M \\ &= \begin{pmatrix} 2rI_n & A^T \\ A & 2sI_m \end{pmatrix} - \begin{pmatrix} rI_n & 0 \\ A & sI_m \end{pmatrix} \begin{pmatrix} I_n & \frac{1-\tau}{r}A^T \\ \frac{-\tau}{s}A & I_m \end{pmatrix} \\ &= \begin{pmatrix} rI_n & \tau A^T \\ \tau A & s(I_m - \frac{1-\tau}{rs}AA^T) \end{pmatrix} \\ &= \begin{pmatrix} rI_n & \tau A^T \\ \tau A & \tau^2 sI \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & s(1-\tau)[(1+\tau)I_m - \frac{1}{rs}AA^T] \end{pmatrix}. \end{aligned}$$

For $\tau \in [0,1]$, G is positive definite when $rs > \|A^T A\|$. The convergence conditions (1.7) are satisfied. Especially, in the case $\tau = 1/2$, when $rs > \frac{3}{4}\|A^T A\|$,

$$G = \begin{pmatrix} rI_n & \frac{1}{2}A^T \\ \frac{1}{2}A & s(I_m - \frac{1}{2rs}AA^T) \end{pmatrix} \succ 0.$$

We do not need to calculate H and G, only verifying their positivity is necessary.

2 Convergence proof in the unified framework

In this section, assuming the conditions (1.7) in the unified framework are satisfied, we prove some convergence properties.

Theorem 2.1 *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*

$$\begin{aligned} &\alpha\big(\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k)\big) \\ &\quad \geq \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \frac{\alpha}{2}\|v^k - \tilde{v}^k\|_G^2, \quad \forall w \in \Omega. \end{aligned} \tag{2.1}$$

Proof. Using $Q = HM$ (see (1.7a)) and the relation (1.6b), the right hand side of (1.7a) can be written as $(v - \tilde{v}^k)^T \frac{1}{\alpha} H(v^k - v^{k+1})$ and hence

$$\alpha\{\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k)\} \geq (v - \tilde{v}^k)^T H(v^k - v^{k+1}), \ \forall w \in \Omega. \tag{2.2}$$

Applying the identity

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

to the right hand side of (2.2) with

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

we thus obtain

$$\begin{aligned}&(v - \tilde{v}^k)^T H(v^k - v^{k+1}) \\ &\quad = \frac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \frac{1}{2}\big(\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2\big). \end{aligned} \tag{2.3}$$

For the last term of (2.3), we have

$$\begin{aligned}&\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 \\ &\quad = \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\ &\quad \overset{(1.6b)}{=} \quad \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - \alpha M(v^k - \tilde{v}^k)\|_H^2 \\ &\quad = \quad 2\alpha(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - \alpha^2(v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\ &\quad = \quad \alpha(v^k - \tilde{v}^k)^T(Q^T + Q - \alpha M^T HM)(v^k - \tilde{v}^k) \\ &\quad \overset{(1.7b)}{=} \quad \alpha\|v^k - \tilde{v}^k\|_G^2. \end{aligned} \tag{2.4}$$

Substituting (2.3), (2.4) in (2.2), the assertion of this theorem is proved. □

2.1 Convergence in a strictly contraction sense

For the convergence in a strictly contraction sense, the matrix G should be positive definite.

Theorem 2.2 *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*

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \alpha\|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.5}$$

Proof. Setting $w = w^*$ in (2.1), we get

$$\begin{aligned}&\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\ &\quad \ge \quad \alpha\|v^k - \tilde{v}^k\|_G^2 + 2\alpha\{\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k)\}. \end{aligned} \tag{2.6}$$

By using the optimality of w^* and the monotonicity of $F(w)$, we have

$$\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k) \ge \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(w^*) \ge 0$$

and thus

$$\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \ge \alpha\|v^k - \tilde{v}^k\|_G^2. \tag{2.7}$$

The assertion (2.5) follows directly. □

2.2 Convergence rate in an ergodic sense

Equivalent Characterization of the Solution Set of VI

For the convergence rate analysis, we need another characterization of the solution set of VI (1.1). It can be described the following theorem and the proof can be found in [9] (Theorem 2.3.5) or [25] (Theorem 2.1).

Theorem 2.3 *The solution set of VI*(Ω, F, θ) *is convex and it can be characterized as*

$$\Omega^* = \bigcap_{w \in \Omega} \{\tilde{w} \in \Omega : \big(\theta(u) - \theta(\tilde{u})\big) + (w - \tilde{w})^T F(w) \geq 0\}. \tag{2.8}$$

Proof. Indeed, if $\tilde{w} \in \Omega^*$, we have

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

By using the monotonicity of F on Ω, this implies that

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

Thus, $\tilde{w}$ belongs to the right-hand set in (2.8). Conversely, suppose $\tilde{w}$ belongs to the latter set of (2.8). Let $w \in \Omega$ be arbitrary. The vector

$$\bar{w} = \alpha\tilde{w} + (1 - \alpha)w$$

belongs to Ω for all $\alpha \in (0, 1)$. Thus we have

$$\theta(\bar{u}) - \theta(\tilde{u}) + (\bar{w} - \tilde{w})^T F(\bar{w}) \geq 0. \tag{2.9}$$

Because $\theta(\cdot)$ is convex, we have

$$\theta(\bar{u}) \leq \alpha\theta(\tilde{u}) + (1 - \alpha)\theta(u) \quad \Rightarrow \quad (1 - \alpha)(\theta(u) - \theta(\tilde{u})) \geq \theta(u) - \theta(\tilde{u}).$$

Substituting it in (2.9) and using $\bar{w} - \tilde{w} = (1 - \alpha)(w - \tilde{w})$, we get

$$(\theta(u) - \theta(\tilde{u})) + (w - \tilde{w})^T F(\alpha\tilde{w} + (1 - \alpha)w) \geq 0$$

for all $\alpha \in (0, 1)$. Letting $\alpha \to 1$, it yields

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

Thus $\tilde{w} \in \Omega^*$. Now, we turn to prove the convexity of Ω^*. For each fixed but arbitrary $w \in \Omega$, the set

$$\{\tilde{w} \in \Omega : \theta(\tilde{u}) + \tilde{w}^T F(w) \le \theta(u) + w^T F(w)\}$$

and its equivalent expression

$$\{\tilde{w} \in \Omega : \big(\theta(u) - \theta(\tilde{u})\big) + (w - \tilde{w})^T F(w) \ge 0\}$$

is convex. Since the intersection of any number of convex sets is convex, it follows that the solution set of VI(Ω, F, θ) is convex. $\square$

In Theorem 2.3, we have proved the equivalence of

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

and

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

We use the late one to define the approximate solution of VI (1.1). Namely, for given $\epsilon > 0$, $\tilde{w} \in \Omega$ is called an ϵ-approximate solution of VI(Ω, F, θ), if it satisfies

$$\tilde{w} \in \Omega, \quad \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \ge -\epsilon, \quad \forall\, w \in \mathcal{D}_{(\tilde{w})},$$

where

$$\mathcal{D}_{(\tilde{w})} = \{w \in \Omega \,|\, \|w - \tilde{w}\| \le 1\}.$$

We need to show that for given $\epsilon > 0$, after t iterations, it can offer a $\tilde{w} \in \mathcal{W}$, such that

$$\tilde{w} \in \mathcal{W} \quad \text{and} \quad \sup_{w \in \mathcal{D}_{(\tilde{w})}} \big\{\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)\big\} \le \epsilon. \tag{2.10}$$

Theorem 2.1 is also the base for the convergence rate proof. Using the monotonicity of F, we have

$$(w - \tilde{w}^k)^T F(w) \ge (w - \tilde{w}^k)^T F(\tilde{w}^k).$$

Substituting it in (2.1), we obtain

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) + \frac{1}{2\alpha}\|v - v^k\|_H^2 \ge \frac{1}{2\alpha}\|v - v^{k+1}\|_H^2, \ \forall w \in \Omega. \tag{2.11}$$

Note that the above assertion is hold for $G \succeq 0$.

Theorem 2.4 *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

$$\tilde{w}_t = \frac{1}{t+1}\sum_{k=0}^{t}\tilde{w}^k. \tag{2.12}$$

Then, for any integer number $t > 0$, $\tilde{w}_t \in \Omega$ and

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2\alpha(t+1)}\|v - v^0\|_H^2, \quad \forall w \in \Omega. \tag{2.13}$$

Proof. First, it holds that $\tilde{w}^k \in \Omega$ for all $k \ge 0$. Together with the convexity of $\mathcal{X}$ and $\mathcal{Y}$, (2.12) implies that $\tilde{w}_t \in \Omega$. Summing the inequality (2.11) over $k = 0, 1, \ldots, t$, we obtain

$$(t+1)\theta(u) - \sum_{k=0}^{t}\theta(\tilde{u}^k) + \Big((t+1)w - \sum_{k=0}^{t}\tilde{w}^k\Big)^T F(w) + \frac{1}{2\alpha}\|v - v^0\|_H^2 \ge 0, \quad \forall w \in \Omega.$$

Use the notation of $\tilde{w}_t$, it can be written as

$$\frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2\alpha(t+1)}\|v - v^0\|_H^2, \quad \forall w \in \Omega. \tag{2.14}$$

Since $\theta(u)$ is convex and

$$\tilde{u}_t = \frac{1}{t+1}\sum_{k=0}^{t}\tilde{u}^k,$$

we have that

$$\theta(\tilde{u}_t) \le \frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k).$$

Substituting it in (2.14), the assertion of this theorem follows directly. □

Recall (2.10). The conclusion (2.13) 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.

- **For the unified framework and the convergence proof, the reader can consult: 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.**
- **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.**

2.3 Convergence rate in pointwise iteration-complexity

In this subsection, we show that if the matrix G defined in (1.7b) is positive definite, a worst-case $O(1/t)$ convergence rate in a nonergodic sense can also be established for the prototype algorithm (1.6). Note in general a nonergodic convergence rate is stronger than the ergodic convergence rate.

We first need to prove the following lemma.

Lemma 2.1 *For the sequence generated by the prototype algorithm (1.6) where the Convergence Condition is satisfied, we have*

$$\begin{aligned}&(v^k-\tilde{v}^k)^T M^T H M\{(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\}\\ \geq\ & \frac{1}{2\alpha}\|(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\|^2_{(Q^T+Q)}.\end{aligned} \tag{2.15}$$

Proof. First, set $w=\tilde{w}^{k+1}$ in (1.6a), we have

$$\theta(\tilde{u}^{k+1})-\theta(\tilde{u}^k)+(\tilde{w}^{k+1}-\tilde{w}^k)^T F(\tilde{w}^k)\geq(\tilde{v}^{k+1}-\tilde{v}^k)^T Q(v^k-\tilde{v}^k). \tag{2.16}$$

Note that (1.6a) is also true for $k:=k+1$ and thus we have

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

Set $w=\tilde{w}^k$ in the above inequality, we obtain

$$\theta(\tilde{u}^k)-\theta(\tilde{u}^{k+1})+(\tilde{w}^k-\tilde{w}^{k+1})^T F(\tilde{w}^{k+1})\geq(\tilde{v}^k-\tilde{v}^{k+1})^T Q(v^{k+1}-\tilde{v}^{k+1}). \tag{2.17}$$

Combining (2.16) and (2.17) and using the monotonicity of F, we get

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

Adding the term

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

to the both sides of (2.18), and using $v^T Qv=\frac{1}{2}v^T(Q^T+Q)v$, we obtain

$$(v^k-v^{k+1})^T Q\{(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\}\geq\frac{1}{2}\|(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\|^2_{(Q^T+Q)}.$$

Substituting $(v^k-v^{k+1})=\alpha M(v^k-\tilde{v}^k)$ in the left-hand side of the last inequality and using $Q=HM$, we obtain (2.15) and the lemma is proved. $\square$

Now, we are ready to prove (2.19), the key inequality in this section.

Theorem 2.5 *For the sequence generated by the prototype algorithm (1.6) where the Convergence Condition is satisfied, we have*

$$\|M(v^{k+1} - \tilde{v}^{k+1})\|_H \le \|M(v^k - \tilde{v}^k)\|_H, \quad \forall\, k > 0. \tag{2.19}$$

Proof. Setting $a = M(v^k - \tilde{v}^k)$ and $b = M(v^{k+1} - \tilde{v}^{k+1})$ in the identity

$$\|a\|_H^2 - \|b\|_H^2 = 2a^T H(a-b) - \|a-b\|_H^2,$$

we obtain

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

Inserting (2.15) into the first term of the right-hand side of the last equality, we obtain

$$\begin{aligned} &\|M(v^k - \tilde{v}^k)\|_H^2 - \|M(v^{k+1} - \tilde{v}^{k+1})\|_H^2 \\ &\ge \quad \frac{1}{\alpha}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T+Q)}^2 - \|M[(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})]\|_H^2 \\ &= \quad \frac{1}{\alpha}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_G^2 \ge 0, \end{aligned}$$

where the last inequality is because of the positive definiteness of the matrix $(Q^T + Q) - \alpha M^T H M \succeq 0$. The assertion (2.19) follows immediately. □

Note that it follows from $G \succ 0$ and Theorem 2.2 there is a constant $c_0 > 0$ such that

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - c_0\|M(v^k - \tilde{v}^k)\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{2.20}$$

Now, with (2.20) and (2.19), we can establish the worst-case $O(1/t)$ convergence rate in a nonergodic sense for the prototype algorithm (1.6).

Theorem 2.6 *Let $\{v^k\}$ and $\{\tilde{w}^k\}$ be the sequences generated by the prototype algorithm (1.6) under the Convergence Condition. For any integer $t > 0$, we have*

$$\|M(v^t - \tilde{v}^t)\|_H^2 \le \frac{1}{(t+1)c_0}\|v^0 - v^*\|_H^2. \tag{2.21}$$

Proof. First, it follows from (2.20) that

$$\sum_{k=0}^{\infty} c_0 \|M(v^k - \tilde{v}^k)\|_H^2 \le \|v^0 - v^*\|_H^2, \quad \forall\, v^* \in \mathcal{V}^*. \tag{2.22}$$

According to Theorem 2.5, the sequence $\{\|M(v^k - \tilde{v}^k)\|_H^2\}$ is monotonically non-increasing. Therefore, we have

$$(t+1)\|M(v^t - \tilde{v}^t)\|_H^2 \le \sum_{k=0}^{t} \|M(v^k - \tilde{v}^k)\|_H^2. \tag{2.23}$$

The assertion (2.21) follows from (2.22) and (2.23) immediately. □

Let $d := \inf\{\|v^0 - v^*\|_H \,|\, v^* \in \mathcal{V}^*\}$. Then, for any given $\epsilon > 0$, Theorem 2.6 shows that the it needs at most $\lfloor d^2/c_0\epsilon \rfloor$ iterations to ensure that $\|M(v^k - \tilde{v}^k)\|_H^2 \le \epsilon$. Recall that v^k is a solution of VI(Ω, F, θ) if $\|M(v^k - \tilde{v}^k)\|_H^2 = 0$ (see (1.6a) and due to $Q = HM$). A worst-case $O(1/t)$ convergence rate in pointwise iteration-complexity is thus established for the prototype algorithm (1.6).

Notice that, for a differentiable unconstrained convex optimization $\min f(x)$, it holds that

$$f(x) - f(x^*) = \nabla f(x^*)^T (x - x^*) + O(\|x - x^*\|^2) = O(\|x - x^*\|^2).$$

3 Conclusions and Remarks

3.1 ADMM vs AMA

We consider the following convex optimization problem

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

Solving (3.1) **by using the penalty function method**

$$u^{k+1} = \text{Argmin}\{\theta(u) + \frac{\beta_k}{2}\|\mathcal{A}u - b\|^2 \,|\, u \in \mathcal{U}\}$$

Solving (3.1) **by using the augmented Lagrangian method**

Begin with a given λ^k,

$$u^{k+1} = \text{Argmin}\{\theta(u) - (\lambda^k)^T(\mathcal{A}u - b) + \tfrac{\beta}{2}\|\mathcal{A}u - b\|^2 \,|\, x \in \mathcal{X}\}$$

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

我们现在要求解的是

$$\min\{\theta_1(x)+\theta_2(y)\,|\,Ax+By=b,\ x\in\mathcal{X},\ y\in\mathcal{Y}\} \tag{3.2}$$

求解问题 (3.2) 的罚函数方法

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

求解问题 (3.2) 的增广 Lagrange 乘子法 从给定的 λ^k 开始

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

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

两类不同方法, 子问题难能度完全一样.

增广 Lagrange 乘子法优于罚函数方法。

J. Nocedal, S. J. Wright: Numerical Optimization

处理具有可分离结构问题 (3.2) 时共同的缺点：

没有利用 x 和 y 的可分离结构！ 求解会无从着手。

- 两个可分离算子问题罚函数方法松弛后就是交替极小化算法 AMA。
- 两个可分离算子问题的增广 Lagrange 乘子法松弛就是乘子交替方向法 ADMM。

求解问题 (3.2) 的松弛的罚函数方法 — 交替极小化方法(AMA)

从给定的 y^k 开始

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

求解问题 (3.2) 的松弛的增广 Lagrange 乘子法 — ADMM

从给定的 (y^k, λ^k) 开始

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

都松弛, 乘子交替方向法 (ADMM) 应该优于交替极小化方法 (AMA)

3.2 两个算子问题 ADMM 方法的 (主要) 改进

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

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

(3.3a)

（松弛延拓） (3.3b)

这里 $\gamma \in (0, 2)$. 赋值号 $:=$ 表示 (3.3b) 右端的 (y^{k+1}, λ^{k+1}) 是由算法的前半部分 (3.3a) 产生的. 对多数问题, 这样往往能加快收敛速度.

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

2. Symmetric ADMM 对称的交替方向法

原始变量 x 和 y 本质上是平等的. 所以建议采用对称的交替方向法.

Symmetric Alternating Direction Method of Multipliers is described as

$$\begin{cases} x^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x, y^k, \lambda^k) \,|\, x \in \mathcal{X}\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \mu\beta(Ax^{k+1} + By^k - b), \\ y^{k+1} = \text{Argmin}\{\mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) \,|\, y \in \mathcal{Y}\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \mu\beta(Ax^{k+1} + By^{k+1} - b). \end{cases} \tag{3.4}$$

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

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

3.3 多个可分离算子的凸优化问题

我们以 3 个可分离算子的问题为例

$$\min\{\theta_1(x) + \theta_2(y) + \theta_3(z) | Ax + By + Cz = b,\ x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\}. \tag{3.5}$$

这个问题的增广 Lagrange 函数是

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

$$\boxed{\begin{cases} x^{k+1} &= \arg\min\left\{\mathcal{L}_\beta^3(x, y^k, z^k, \lambda^k) \,\middle|\, x \in \mathcal{X}\right\}, \\ y^{k+1} &= \arg\min\left\{\mathcal{L}_\beta^3(x^{k+1}, y, z^k, \lambda^k) \,\middle|\, y \in \mathcal{Y}\right\}, \\ z^{k+1} &= \arg\min\left\{\mathcal{L}_\beta^3(x^{k+1}, y^{k+1}, z, \lambda^k) \,\middle|\, z \in \mathcal{Z}\right\}, \\ \lambda^{k+1} &= \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). \end{cases}} \tag{3.6}$$

对 $m \geq 3$, 直接推广的交替方向法不能保证收敛.

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

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

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

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

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

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

值得继续研究的问题： 三个算子的实际问题中, 线性约束矩阵

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

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

举个简单的例子来说：

- 乘子交替方向法 (ADMM) 处理问题

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

- 将等式约束换成不等式约束, 问题就变成

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

- 再化成三个算子的等式约束问题

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

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

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

处理办法一：带高斯回代的 ADMM 方法

以 (3.6) 提供的 (y^{k+1}, z^{k+1}) 为预测, 取 $\alpha \in (0,1)$, 校正公式为

$$\begin{pmatrix} y^{k+1} \\ z^{k+1} \end{pmatrix} := \begin{pmatrix} y^k \\ z^k \end{pmatrix} - \alpha \begin{pmatrix} I & -(B^TB)^{-1}B^TC \\ 0 & I \end{pmatrix} \begin{pmatrix} y^k - y^{k+1} \\ z^k - z^{k+1} \end{pmatrix}. \quad (3.7)$$

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

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

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

对 y 和 z, 有先后, 不公平, 那就要做找补, 调整.

处理办法二：ADMM + Prox-Parallel Splitting ALM

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

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

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

太自由, 又不校正, 就加正则项, 不忘自己昨天的承诺.

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

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

人们因此有理由对 ADMM 格外关心。

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

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

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

这些方法，道理上能站住脚，计算表现也不俗。

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

ADMM 的广泛应用，人们自然想到向三个算子的问题推广。

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

(Computational Optimization and Applications, 2009).

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

这些方法的共同特点是不需要对问题加任何条件！对 β 不加限制，只对方法动手术！

后来我们又给出“直接推广的ADMM 方法处理三个算子问题不保证收敛”的例子(Math. Progr., 2016),说明:
以前提出的一些策略，手段上是必须的，机制上也是合理的。

- The analysis is guided by variational inequality.
- The most methods mentioned fall in a unified prediction-correction framework, in which the convergence analysis is quite simple.
- All the discussed methods are closed related to Proximal Point Algorithms.
- All the discussed ADMM-like splitting methods are rooted from Augmented Lagrangian Method.
- A through reading will acquaint you with the ADMM, while a more carefully reading may make you familiar with the tricks on constructing splitting methods according to the problem you met.
- The discussed first order splitting contraction methods are only appropriate for some structured convex optimization in some practical applications.

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] J. F. Cai, E. J. Candès and Z. W. Shen, A singular value thresholding algorithm for matrix completion, SIAM J. Optim. **20**, 1956-1982, 2010.

[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] 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.

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

[6] C. H. Chen, B. S. He, Y. Y. Ye and X. M. Yuan, *The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent*, to appear in *Mathematical Programming, Series A*.

[7] J. Douglas and H. H. Rachford, On the numerical solution of the heat conduction problem in 2 and 3 space variables, *Transactions of the American Mathematical Society* **82** (1956), 421–439.

[8] 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.

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

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

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

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

[13] B.S. He, *A new method for a class of linear variational inequalities*, Math. Progr., **66**, pp. 137-144, 1994.

[14] B.S. He, *Solving a class of linear projection equations*, Numerische Mathematik, **68**, pp. 71-80, 1994.

[15] B.S. He, A class of projection and contraction methods for monotone variational inequalities, Applied Mathematics and Optimization, **35**, 69-76, 1997.

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

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

[18] 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.

[19] 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.

[20] 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.

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

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

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

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

[29] 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 (2) (2015), pp. 287-325.

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

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

[32] 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.

[33] 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.

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

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

[36] 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.

[37] 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.