A-H 八单元 16 个两页短篇及其简介 导读和参考文献

何 炳 生　maths.nju.edu.cn/~hebma

我把最优化理论与方法, 定位成最接地气的应用数学。在这个领域里我偏重算法研究, 主要兴趣是单调变分不等式的投影收缩算法和凸优化的 (ADMM类) 分裂收缩算法, 组织发表了 (包括一些浅陋之作的) 百多篇文章。这些方法得到了学术界的一些好评, 也偶尔被工程领域的学者知悉, 用来成功解决了他们的一些问题。我曾经写过两页的介绍, 企图概括自己一生的工作, 但显得过于简单, 不尽如人意。退休前后我还做过总结, 用 PPT 写成二十讲的系列讲义, 挂在个人主页, 但篇幅较长, 通读要花不少时间。若精力允许, 也应该修订更新。转眼 10 多年过去, 时而还有年轻朋友向我咨询学术问题。随着年龄的增长, 我又考虑, 一生中自己最看重的工作是哪些？能否从自己零零散散的收获, 以及七旬以后的进展中, 选一些每个(包括主要推导过程)都能用A4 两页讲清的东西, 介绍给那些想对我的工作有所了解的年轻人。写成以后, 请几位在读博士生看看, 回馈都说看懂基本没有问题。目前完成的这 **A–H** 八章的 16 篇, 对我自己而言, 犹如老人自制的文玩小把件, 每件都能勾起我一段美好的回忆。这些短篇本身多半就是披着高等数学外衣的初等数学, 读懂这些材料所需的预备知识, 是中学的数理基础和普通的大学数学, 同时需要认可一般的优化原理。只是对近 20 年来热门的 ADMM 类分裂收缩算法比较感兴趣的学者, 也可以从这里的 **C** 章开始读起。最后的 **G** 和 **H** 两章中的内容, 说明我古稀之后没有马上洗手是对的。能用这种形式写下来, 原因也是这些工作需要的基础都是初等的。

为展示研究路径, 也给出主要参考文献, 但读懂这些短篇, 阅读所列文献并非必需的。

A-从变分不等式的投影收缩算法到凸优化的分裂收缩算法 2013年, 我从南京大学退休, 试图用两页纸概括总结一下职业生涯的科研成果, 戏称要为自己写一篇“墓志铭”, 于是就有了这个短篇 **A**。说起来, 我从单调变分不等式的投影收缩算法做起, 到 (包括ADMM的) 线性约束凸优化的分裂收缩算法, 可以说是一条主线, 一个模式, 短篇 **A** 主要是把这两类算法的关系说清楚。如需进一步了解关于投影收缩算法的成果, 建议阅读[1, 2]。对于如何从单调变分不等式的投影收缩算法过渡到线性约束凸优化的分裂收缩算法的, 我在论文[3, 4]中做了说明。短篇 **A** 第二节已经包含的变分不等式意义下凸优化分裂收缩算法的统一框架, 以及方法收敛的关键不等式。中文读者对 **A** 中§2中内容感兴趣, 建议阅读[4] 。

这个两页的 **A** 有点过于简单, 大概只适合自己的学生有兴趣时重温而已。后面 **B** 和 **C** 两章中的短篇, 分别介绍并证明变分不等式的投影收缩算法和热门的 ADMM 方法的主要收敛性质。

B-变分不等式投影收缩算法中的孪生方向和姊妹方法 在单调变分不等式的投影收缩算法中, 我们根据投影得到一对孪生方向建立的姊妹方法, 被一些工程界学者用来解决了长期困扰他们的问题。姊妹方法中用孪生方向和相同步长, 内涵的数学之美常常令我自己陶醉不已！回头看, 虽然发现这些关系的过程有点漫长, 但其中用到的数学就是投影的基本性质, 犹如平面几何中的余弦定理。**B** 中两个两页纸的短篇, 分别详细介绍了单调线性和非线性变分不等式投影收缩算法中孪生方向的由来, 也给出了姊妹方法关键收敛性质的证明。如果有进一步了解这方面工作的需求, 建议阅读论文[1, 2, 3, 5], 也可以阅读我系列讲义[6] 的 1-5 讲。论文[1] 和[3] 中提到了三个基本不等式并指出孪生方向也可以从这些基本不等式的不同组合导出。一般情形下求解单调变分不等式的预测－校正框架和一些可供参考的算例可见[7, 8]。将这些想法推广到凸优化问题的求解上, 文献[9]做了一些探讨。

C-利用变分不等式证明乘子交替方向法 (ADMM) 的关键收敛性质 我们从 1997 年开始在变分不等式投影收缩算法的基础上研究 ADMM (乘子交替方向法)[10]。直到 2006 年, 我们研究的目标主要针对管理科学中的一些结构型变分不等式, 并在自调比准则[11]、子问题不精确求解[12] 以及和预测-校正实现[13] 等方面做了些工作。十年后, 不少数学工作者发现 ADMM 在优化领域有广泛的应用。凭借已有的基础, 我们迅速投入, 继续做了一些研究工作, 包括 ADMM 收敛速率的文章[14, 15] 和不精确ADMM方法[16]。虽然我们是在 Gowinski[17] 工作的基础上研究 ADMM, 最初的论文也采用他们的步长法则[10, 18], 为了让初学者更好地理解 ADMM 方法的收敛性原理, 对工程界普遍采用的 ADMM 的最简版本, **C** 中的第一个短篇, 在变分不等式框架下写下关键收敛性质的简短证明, 得到不少好评。**C-2** 在预测-校正统一框架下证明 ADMM 的收敛性, 这种处理方式对收敛速率证明提供的方便写在 **C3** 中。我自己很欣赏的最优线性化ADMM的成果[19], 因篇幅关系这里没有介绍, 有兴趣的可以参阅相关文献。

D-凸优化问题对应的变分不等式和因需定制的PPA算法 min-max 问题的解点和线性约束凸优化问题 Lagrange 函数的鞍点, 都可以等价地转换成一个(混合)变分不等式 (VI) 的解点。求解这类变分不等式, 我们提出了一类通过构造适当的分块正定矩阵, 分裂求解优化子问题实现的邻近点算法[20], 并称其为因需定制的邻近点算法(Customized PPA)[21, 22]。**D** 中的两个两页纸短篇, 包含了对一些典型问题如何构造 Customized PPA 的例子。利用 PPA 中正定矩阵的分块结构, 使迭代得以用分裂的形式实现, 是这些方法被称为分裂算法的原因。关于求解变分不等式的正定矩阵模下的邻近点算法, 进一步的资料可以参阅[23, 24, 37]。通过合理设计PPA中的正定矩阵, 构造一些分摊子问题难度的均困方法, 这些工作可以在[25, 26, 39]中找到。

E-效率更高的求解两个可分离块凸优化问题的ADMM类方法 在短篇 **C** 里我们用两页纸给出了 ADMM 的收敛性证明。在常用的 ADMM 的基础上, 做一些简单修改(包括迭代中改变变量更新顺序的 PPA 型 ADMM 算法[27]和两次修正乘子的对称 ADMM 算法[28])。这些想法比较自然的算法, 在几乎不增加工作量的情况下, 效率一般都有30%以上的提高。**E** 中短篇介绍的这两个算法, 用在短篇 **A** 的§2 中已经提及的统一框架下验证了收敛性。关于统一框架算法的两条简单而又重要的性质, 我们会在 **G** 和 **H** 的短篇中给出更系统的证明。

F-求解三个可分离块凸优化问题的 ADMM 类分裂收缩算法 ADMM 是求解两个可分离块的凸优化问题的有效方法, 但直接推广到用来求解三个可分离块的凸优化问题是并不能保证收敛的[29]。在这个事实得到明确证明之前, 根据社会生活中的一些机理, 我们提出了两个修正的 ADMM 类算法[30, 31], 并且都在统一框架下验证了算法的收敛性。我们将算法及其收敛性证明写成 **F** 中的两个短篇。读者如需理解这些算法的机理, 建议阅读文献[32]。处理三个可分离块问题最优正则化的文章[33], 是处理两个可分离块最优线性化ADMM[19] 同类思想的推广, 也非常值得介绍, 有兴趣的读者可以参阅相关文献。

G-从好不容易凑出一个方法到并不费劲构造一簇算法 线性约束的凸优化问题的拉格朗日函数的鞍点, 都可以转换成一个单调变分不等式的解点。为求解这类变分不等式, 十多年前, 我们就提出了一个预测-校正的统一框架[3], 并给出了这个算法框架下保证收敛的条件。这对验证算法收敛性带来极大的方便, 有时我们也利用这个框架的收敛性条件去凑一些方法[34, 35]。这些给人一种凑得很巧的神秘感觉, 直到 2022 年, 我们才对算法框架的收敛性条件给出了一个等价表示, 据此便可以并不费劲地构造一簇算法[36, 37, 38]。对这类构造性方法, 包括统一的收敛性证明, 用两页纸写成了短篇 **G-1**。**G-2** 以三个可分离块的线性约束凸优化为例, 解释如何基于同一预测, 采用不同的校正, 构造一簇算法。

这样的成果没有在晚年的我手边溜走, 是值得庆幸的!

H-求解线性约束凸优化问题的预测-校正型广义邻近点算法 对由凸优化转换得来的变分不等式, 短篇 **D** 中已经介绍了因需定制的 PPA 算法。在短篇 **G**中, 我们根据预测-校正统一框架的收敛性条件的等价表示, 对确定的合格预测, 可以选择不同的校正矩阵, 从而并不费劲地构造一簇算法。这里, 在短篇 **G-1** 的基础上, 采用由预测唯一确定的特殊的校正矩阵, 得到预测-校正的广义邻近点算法[38], 具有短篇 **D** 中经典 PPA 算法的所有漂亮性质。我们将个结果及其证明用两页纸写成短篇 **H-1**。**H-2** 以多个可分离块的线性约束凸优化问题为例, 解释如何构造预测-校正的广义 PPA 算法。其每步迭代中的预测和校正, 犹如高斯消去法中的“消去和回代”, 程序非常流畅。不同的只是这里的预测是通过求解形式比较简单的子问题实现的。当然, 高斯消去法是求解线性方程组的直接方法。需要强调的是：凸优化问题的求解一般没有直接法, 这里介绍的广义 PPA 的一次迭代, 相当于高斯消去法求解线性方程组的消去和回代.

利用变分不等式和邻近点算法这两大法宝, 才有了我们自成体系又颇有特色的工作。几位从论文中了解到我们工作的欧美学者, 称赞我们的方法“Very Simple yet Powerful”和“Elegant”, 也是基于我们在 VI 框架下构造了 PPA。我们的统一框架, 不仅涵盖了 PPA, ALM 和 ADMM 等耳熟能详的基础算法, 还为这些方法无法胜任的(三块和多于三块的)可分离凸优化问题提供了系列求解方法。有朋友说我用的主要数学工具, 本质上就是中学数学的余弦定理, 我也觉得大差不离。要说我在学术生涯中主导和带领学生做了点什么, 多半已经在这八章16个短篇中提及。对希望了解我这一生究竟做了些什么的数学界朋友, 我要说, 浏览一下这些“文玩小把件”, 大体上也就可以!

何炳生　2024 年 7 月

A-从变分不等式的投影收缩算法到凸优化的分裂收缩算法

2013年前总结的主要研究工作 maths.nju.edu.cn/~hebma 何炳生

1 变分不等式的投影收缩算法 (变分不等式应用与算法建议参阅本人主页系列讲义前三讲)

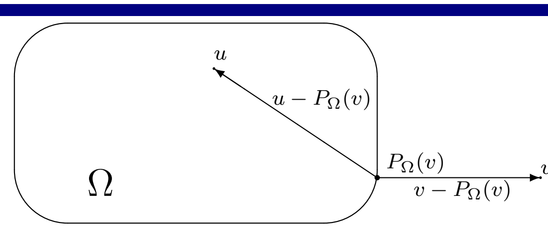
设 $\Omega \subset \Re^n$ 是一个非空闭凸集, F 是 $\Re^n \to \Re^n$ 的一个映射. 考虑求解单调变分不等式

$$\text{(VI)} \qquad u^* \in \Omega, \quad (u-u^*)^T F(u^*) \ge 0, \quad \forall\, u \in \Omega. \tag{A1}$$

当 $\Omega = \Re^n$, 问题(A1)就退化成非线性方程组 $F(u)=0$.

我们说一个变分不等式单调, 是指其中的算子 F 满足 $(u-v)^T(F(u)-F(v)) \ge 0$. 在求解变分不等式 (A1) 的投影收缩算法中, 对给定的当前点 u^k 和常数 $\beta_k > 0$, 利用投影

$$\tilde u^k = P_\Omega[u^k - \beta_k F(u^k)] \tag{A2}$$



生成一个预测点 $\tilde u^k$. 同时假设用投影 (A2) 产生 $\tilde u^k$ 时选取的 β_k 能使得预测点 $\tilde u^k$ 满足

$$\beta_k \|F(u^k) - F(\tilde u^k)\| \le \nu \|u^k - \tilde u^k\|, \quad \nu \in (0,1). \tag{A3}$$

假如 $\tilde u^k = u^k$, u^k 就是 (A1) 的解. 否则, 就要基于以下的分析进行新的迭代. 首先, 凸集上投影有如右上图所示的重要性质: 对任意的 $v \in \Re^n$, 用 $P_\Omega(v)$ 表示 v 到 Ω 上的投影, 有

$$(v - P_\Omega(v))^T(u - P_\Omega(v)) \le 0, \quad \forall u \in \Omega. \tag{A4}$$

在 (A4) 中令 $v = u^k - \beta_k F(u^k)$, 那么由 (A4) 得到 $\tilde u^k = P_\Omega(v)$, 并从 (A4) 得到

$$\tilde u^k \in \Omega, \quad (u - \tilde u^k)^T\{\tilde u^k - [u^k - \beta_k F(u^k)]\} \ge 0, \quad \forall u \in \Omega. \tag{A5}$$

在上面的 (A5) 式两边都加上 $(u-\tilde u^k)^T d(u^k,\tilde u^k)$, 其中

$$d(u^k,\tilde u^k) = (u^k - \tilde u^k) - \beta_k[F(u^k) - F(\tilde u^k)], \tag{A6}$$

从而由投影 (A2) 得到我们需要的预测公式

$$\textbf{[预测]} \qquad \tilde u^k \in \Omega, \ \ (u-\tilde u^k)^T \beta_k F(\tilde u^k) \ge (u-\tilde u^k)^T d(u^k,\tilde u^k), \ \ \forall u \in \Omega. \tag{A7}$$

将 (A7) 中任意的 $u \in \Omega$ 选成某个解点 u^*, 利用 (A1) 和 $F(u)$ 的单调性, 有 $(\tilde u^k - u^*)^T d(u^k,\tilde u^k) \ge 0$, 因此

$$(u^k - u^*)^T d(u^k,\tilde u^k) \ge (u^k - \tilde u^k)^T d(u^k,\tilde u^k). \tag{A8}$$

由 $d(u^k,\tilde u^k)$ 的表达式和假设 (A3), 利用 Cauchy-Schwarz 不等式, 推得

$$(u^k - \tilde u^k)^T d(u^k,\tilde u^k) \ge (1-\nu)\|u^k - \tilde u^k\|^2. \tag{A9}$$

当 $u^k \ne \tilde u^k$ 时, (A8) 说明 $d(u^k,\tilde u^k)$ 是距离函数 $\frac12\|u-u^*\|^2$ 在 u^k 处的一个上升方向. 可以用

$$\textbf{[单位步长的校正]} \qquad u^{k+1} = u^k - d(u^k,\tilde u^k) \qquad \text{生成新的迭代点.} \tag{A10}$$

由 (A6) 可得, $\quad 2(u^k-\tilde u^k)^T d(u^k,\tilde u^k) - \|d(u^k,\tilde u^k)\|^2 = \|u^k - \tilde u^k\|^2 - \beta_k^2\|F(u^k) - F(\tilde u^k)\|^2. \tag{A11}$

因此,

$$\begin{aligned} \|u^k-u^*\|^2 - \|u^{k+1}-u^*\|^2 &= 2(u^k-u^*)^T d(u^k,\tilde u^k) - \|d(u^k,\tilde u^k)\|^2 \\ &\ge 2(u^k-\tilde u^k)^T d(u^k,\tilde u^k) - \|d(u^k,\tilde u^k)\|^2 \ge (1-\nu^2)\|u^k-\tilde u^k\|^2. \end{aligned} \tag{A12}$$

$$\textbf{[计算步长的校正]} \qquad u^{k+1} = u^k - \alpha_k^* d(u^k,\tilde u^k) \qquad \text{产生离 } u^* \text{ 更近的迭代点.} \tag{A13}$$

$$\text{其中步长} \quad \alpha_k^* \quad \text{由} \quad \alpha_k^* = (u^k-\tilde u^k)^T d(u^k,\tilde u^k)/\|d(u^k,\tilde u^k)\|^2 \quad \text{给出.} \tag{A14}$$

由 (A11) 和 (A3), 容易推得 $2(u^k-\tilde u^k)^T d(u^k,\tilde u^k) > \|d(u^k,\tilde u^k)\|^2$, 因而 $\alpha_k^* > \frac12$. 接着就有

$$\begin{aligned} \|u^k-u^*\|^2 - \|u^{k+1}-u^*\|^2 &= \|u^k-u^*\|^2 - \|u^k-u^*-\alpha_k^* d(u^k,\tilde u^k)\|^2 \\ &= 2\alpha_k^*(u^k-u^*)^T d(u^k,\tilde u^k) - (\alpha_k^*)^2\|d(u^k,\tilde u^k)\|^2 \quad \text{(利用 (A8) 和 (A14))} \\ &\ge \alpha_k^*(u^k-\tilde u^k)^T d(u^k,\tilde u^k) \ge \tfrac12(1-\nu)\|u^k-\tilde u^k\|^2. \quad \text{(最后一个不等式用了 (A9))} \end{aligned} \tag{A15}$$

收缩不等式 (A12) 和 (A15) 是证明算法收敛的关键式子. 分处不等式 (A7) 两端的 $\beta_k F(\tilde u^k)$ 和 $d(u^k,\tilde u^k)$ 称为一对孪生方向. 将在 **B** 中证明的美妙之处是：采用方向 $\beta_k F(\tilde u^k)$ 和 (A14) 中给出的步长 α_k^*, 用

[投影校正] $\quad u^{k+1} = P_\Omega[u^k - \alpha_k^* \beta_k F(\tilde u^k)]$ 与方法 (A13) 具有同样的收敛性质 (A15), 详见 **B**.

✠ 孪生方向, 相同步长. 姊妹方法, 投影校正效果更好！ 真所谓 兄弟齐心, 其利断金！ ✠

2 线性约束凸优化分裂收缩算法的统一框架 (参见何炳生 2015, 2018 运筹学报文章)

我们在变分不等式框架下讨论线性约束凸优化的分裂收缩算法, 以三块可分离凸优化问题

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z) \mid Ax+By+Cz=b,\ x\in\mathcal{X}, y\in\mathcal{Y}, z\in\mathcal{Z}\} \tag{A16}$$

为例做介绍. 问题 (A16) 的拉格朗日函数是 $L(x,y,z,\lambda)=\theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^T(Ax+By+Cz-b)$.

拉格朗日函数的鞍点 (x^*,y^*,z^*,λ^*) （其中 (x,y,z) 为原始变量, λ 为对偶变量）满足

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

利用鞍点 $w^*=(x^*,y^*,z^*,\lambda^*)$ 的极大极小性质 (推导可参见 **C** 或 **D**), 它可以表述为如下的变分不等式:

$$\text{(VI)}\qquad w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega, \tag{A17}$$

其中, $\Omega=\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}\times\Re^m$, $\quad\theta(u)=\theta_1(x)+\theta_2(y)+\theta_3(z)$, 并且

$$u=\begin{pmatrix}x\\y\\z\end{pmatrix},\quad w=\begin{pmatrix}x\\y\\z\\\lambda\end{pmatrix},\quad F(w)=\begin{pmatrix}-A^T\lambda\\-B^T\lambda\\-C^T\lambda\\Ax+By+Cz-b\end{pmatrix}. \tag{A18}$$

注意到, (A18) 中的 $F(w)$ 恰有 $(w-\tilde{w})^T(F(w)-F(\tilde{w}))\equiv 0$.

加之 $\theta(u)$ 是凸函数(不一定可微), 我们称 (A17) 为单调(混合)变分不等式, 简称单调变分不等式.

我们首先介绍与投影收缩算法类似的求解单调变分不等式问题 (A17) 的预测-校正收缩算法.

[预测] 对给定的 v^k(在两块可分离问题的ADMM中 $v=(y,\lambda)$, 有些算法中 $v=w$), 求得点 $\tilde{w}^k$, 使得

$$\tilde{w}^k\in\Omega,\ \ \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(v-\tilde{v}^k)^TQ(v^k-\tilde{v}^k),\ \ \forall w\in\Omega, \tag{A19}$$

如果矩阵 Q^T+Q 正定(Q 不一定对称), 我们称 (A19) 为合格的预测. 例见短篇 **F** 的 (4.2)

利用预测 (A19), 可以建立求解单调变分不等式 (A17) 的预测-校正的统一框架. (ADMM为其一个特例).

[单位步长的校正] 采用单位步长, 新迭代点的校正公式为 例见短篇 **F** 的 (4.3)

$$v^{k+1}=v^k-M(v^k-\tilde{v}^k). \tag{A20}$$

由合格预测 (A19) 和单位步长的校正 (A20), 组成一个预测-校正算法框架, 其收敛性条件是:

[收敛性条件] 要求存在正定矩阵 H, 能够使得 例见短篇 **F** 的 (4.4)-(4.6)

$$HM=Q \qquad \text{并且} \qquad G=Q^T+Q-M^THM\succ 0.\quad (\text{正定}) \tag{A21}$$

将 (A19) 中任意的 $w\in\Omega$ 选成某个解点 w^*, 利用$(\tilde{w}-w^*)^TF(\tilde{w})=(\tilde{w}-w^*)^TF(w^*)$ 和 w^* 是解这一事实,

得到 $\quad(\tilde{v}^k-v^*)^TQ(v^k-\tilde{v}^k)\ge 0.\quad$ 进而有 $\quad(v^k-v^*)^TQ(v^k-\tilde{v}^k)\ge(v^k-\tilde{v}^k)^TQ(v^k-\tilde{v}^k).\quad$ (A22)

$$\begin{aligned}\|v^k-v^*\|_H^2-\|v^{k+1}-v^*\|_H^2&=\|v^k-v^*\|_H^2-\|(v^k-v^*)-M(v^k-\tilde{v}^k)\|_H^2 \quad (\text{利用 } HM=Q)\\ &\ge\ (v^k-\tilde{v}^k)^T(Q^T+Q)(v^k-\tilde{v}^k)-\|M(v^k-\tilde{v}^k)\|_H^2 \overset{(A21)}{=} \|v^k-\tilde{v}^k\|_G^2.\end{aligned} \tag{A23}$$

不等式 (A26) 说明统一框架算法在条件 (A21) 满足的情况下, 迭代方法生成的序列 $\{v^k\}$ 具有收缩性质

$$\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2-\|v^k-\tilde{v}^k\|_G^2. \qquad (\text{这是证明方法收敛性的关键不等式}) \tag{A24}$$

结论: 由预测 (A19) 和校正 (A20) 构成的求解 VI 问题 (A17) 的算法统一框架, 条件 (A21)满足时方法就收敛.

根据 (A19), 也可以如同 §1 中考虑 H-模下计算步长的校正, 这里的 H 为任意的正定矩阵 (也可以是单位阵). 由 (A22) 和 Q^T+Q 正定, 向量 $H^{-1}Q(v^k-\tilde{v}^k)$ 是距离函数 $\frac{1}{2}\|v-v^*\|_H^2$ 在 v^k 处的一个上升方向. 如同 §1 的投影收缩算法中有了 (A8), 可以用 (A13)-(A14) 进行校正. 根据 (A22) 实行计算步长的校正:

[计算步长的校正] 对选定的正定矩阵 H, 由以下法则生成新的迭代点:

$$v^{k+1}=v^k-\alpha_k^*M(v^k-\tilde{v}^k),\qquad \text{其中}\quad M=H^{-1}Q, \tag{A25a}$$

$$\text{步长}\ \ \alpha_k^*\ \ \text{由}\ \ \alpha_k^*=(v^k-\tilde{v}^k)^TQ(v^k-\tilde{v}^k)/\|M(v^k-\tilde{v}^k)\|_H^2\quad\text{给出}. \tag{A25b}$$

$$\begin{aligned}\|v^k-v^*\|_H^2-\|v^{k+1}-v^*\|_H^2 &\overset{(A25a)}{=} \|v^k-v^*\|_H^2-\|(v^k-v^*)-\alpha_k^*M(v^k-\tilde{v}^k)\|_H^2 \quad (\text{利用 } HM=Q)\\ &\ge\alpha_k^*(v^k-\tilde{v}^k)^T(Q^T+Q)(v^k-\tilde{v}^k)-\alpha_k^*\big(\alpha_k^*\|M(v^k-\tilde{v}^k)\|_H^2\big) \overset{(A25b)}{=} \tfrac{1}{2}\alpha_k^*\|v^k-\tilde{v}^k\|_{(Q^T+Q)}^2.\end{aligned} \tag{A26}$$

✠ 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！ ✠

B1-变分不等式投影收缩算法中的孪生方向和姊妹方法

奇妙的孪生方向和姊妹方法 maths.nju.edu.cn/~hebma 何炳生

1　孪生方向和相同步长的姊妹方法 (VI的应用及PC算法建议参阅本人主页系列讲义前五讲)

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. 考虑求解单调变分不等式

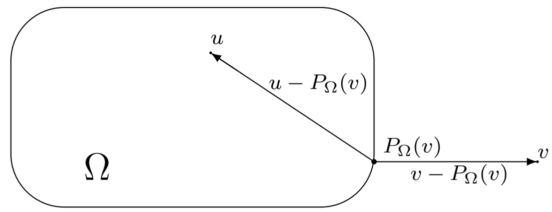
$$\text{(VI)} \qquad u^* \in \Omega, \quad (u-u^*)^T F(u^*) \ge 0, \quad \forall\, u \in \Omega. \tag{B1-1}$$

如果上式中 $\Omega = \Re^n_+$ (非负卦限), 变分不等式 (B1-1) 就是一个互补问题 (Complementarity Problem):

$$\text{(CP)} \qquad u^* \ge 0, \qquad F(u^*) \ge 0, \qquad (u^*)^T F(u^*) = 0. \tag{B1-2}$$

换句话说, 互补问题是 $\Omega = \Re^n_+$ 的变分不等式. 我们说一个变分不等式单调, 是指其中的算子 F 满足 $(u-v)^T(F(u)-F(v)) \ge 0$. 在求解变分不等式 (B1-1) 的投影收缩算法中, 对给定的当前点 u^k 和常数$\beta_k > 0$, 我们利用投影

$$\tilde{u}^k = P_\Omega[u^k - \beta_k F(u^k)] \tag{B1-3}$$



生成一个预测点 $\tilde{u}^k$. 假如 $\tilde{u}^k = u^k$, u^k 就是 (B1-1) 的解. 否则, 就要进行基于以下分析的新的迭代.

在生成预测点 $\tilde{u}^k$ 的投影 (B1-3) 中, 要求参数 β_k 取得满足

$$\beta_k \|F(u^k) - F(\tilde{u}^k)\| \le \nu \|u^k - \tilde{u}^k\|, \quad \nu \in (0,1). \tag{B1-4}$$

凸集上投影有如右上图所示的重要性质: 对任意的 $v \in \Re^n$, 用 $P_\Omega(v)$ 表示 v 到 Ω 上的投影, 有

$$(u - P_\Omega(v))^T (v - P_\Omega(v)) \le 0, \quad \forall u \in \Omega. \tag{B1-5}$$

在 (B1-5) 中令 $v = u^k - \beta_k F(u^k)$, 那么由 (B1-3) 得到 $\tilde{u}^k = P_\Omega(v)$, 并从 (B1-5) 得到

$$\tilde{u}^k \in \Omega, \quad (u-\tilde{u}^k)^T \{[u^k - \beta_k F(u^k)] - \tilde{u}^k\} \le 0, \quad \forall u \in \Omega. \tag{B1-6}$$

上式两边都加上 $(u-\tilde{u}^k)^T \beta_k F(\tilde{u}^k)$, 由此得到我们基于投影 (B1-3) 的预测公式

$$\textbf{[预测]} \qquad \tilde{u}^k \in \Omega, \;\; (u-\tilde{u}^k)^T d(u^k,\tilde{u}^k) \le (u-\tilde{u}^k)^T \beta_k F(\tilde{u}^k), \;\; \forall u \in \Omega. \tag{B1-7}$$

$$\text{其中} \qquad d(u^k,\tilde{u}^k) = (u^k - \tilde{u}^k) - \beta_k [F(u^k) - F(\tilde{u}^k)]. \tag{B1-8}$$

定义 1 (孪生方向) 在基于投影 (B1-3) 得到的预测变分不等式 (B1-7) 中, 分处两端的

$$d(u^k,\tilde{u}^k) \qquad \text{和} \qquad \beta_k F(\tilde{u}^k), \qquad \text{称为一对孪生方向.} \tag{B1-9}$$

将 (B1-7) 中的 $u \in \Omega$ 选成任意的解点 u^*, 利用 $F(u)$ 的单调性和 (B1-1), 有

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

因此, 进而有

$$(u^k - u^*)^T d(u^k,\tilde{u}^k) \ge (u^k - \tilde{u}^k)^T d(u^k,\tilde{u}^k). \tag{B1-10}$$

定义 2 (姊妹方法) 采用 (B1-9) 中一对孪生方向进行校正的

$$u_I^{k+1}(\alpha) = u^k - \alpha d(u^k,\tilde{u}^k), \tag{B1-11}$$

和

$$u_{II}^{k+1}(\alpha) = P_\Omega[u^k - \alpha\beta_k F(\tilde{u}^k)], \tag{B1-12}$$

称为一对姊妹方法. 其中 α 是相同的待定步长.

对任意给定的解点 $u^* \in \Omega^*$, 我们记

$$\vartheta_k(\alpha) = \|u^k - u^*\|^2 - \|u_I^{k+1}(\alpha) - u^*\|^2 \tag{B1-13}$$

和

$$\zeta_k(\alpha) = \|u^k - u^*\|^2 - \|u_{II}^{k+1}(\alpha) - u^*\|^2. \tag{B1-14}$$

它们是依赖于步长 α 的 k 次迭代的收益, 是步长 α 的函数.

我们不能直接极大化 $\vartheta_k(\alpha)$ 和 $\zeta_k(\alpha)$, 因为它含有未知的 u^*. 下面将证明二次函数

$$q_k(\alpha) = 2\alpha (u^k - \tilde{u}^k)^T d(u^k,\tilde{u}^k) - \alpha^2 \|d(u^k,\tilde{u}^k)\|^2 \tag{B1-15}$$

是它们的下界. 为了保证 $q_k(\alpha) > 0$, 我们在生成预测点 $\tilde{u}^k$ 的投影 (B1-3) 中, 要求参数 β_k 使得 (B1-4) 成立. 在这篇注记的 §2, 我们将会证明, 对同样的 α, $\zeta_k(\alpha)$ 优于 $\vartheta_k(\alpha)$.

2 姊妹方法的收缩性质 (参见何炳生 2016 高校计算数学学报 38 卷第 1 期)

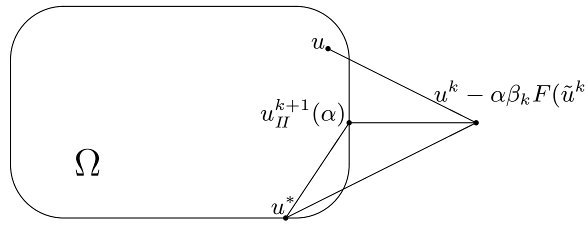
定理 设 $u_I^{k+1}(\alpha)$ 和 $u_{II}^{k+1}(\alpha)$ 由姊妹方法生成. 对分别由 (B1-13) 和 (B1-14) 定义的 $\vartheta_k(\alpha)$ 和 $\zeta_k(\alpha)$, 有

$$\vartheta_k(\alpha) \ge q_k(\alpha) \quad 和 \quad \zeta_k(\alpha) \ge q_k(\alpha) + \|u_I^{k+1}(\alpha) - u_{II}^{k+1}(\alpha)\|^2, \tag{B1-16}$$

其中 $q_k(\alpha)$ 由 (B1-15) 给出.

证明. 根据 (B1-13) 中对 $\vartheta_k(\alpha)$ 的定义, 利用 (B1-10) 我们有

$$\begin{aligned}
\vartheta_k(\alpha) &= \|u^k - u^*\|^2 - \|u_I^{k+1}(\alpha) - u^*\|^2 \\
&= \|u^k - u^*\|^2 - \|u^k - u^* - \alpha d(u^k, \tilde{u}^k)\|^2 \\
&= 2\alpha(u^k - u^*)^T d(u^k, \tilde{u}^k) - \alpha^2 \|d(u^k, \tilde{u}^k)\|^2 \\
&\ge 2\alpha(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \alpha^2 \|d(u^k, \tilde{u}^k)\|^2 \\
&= q_k(\alpha).
\end{aligned} \tag{B1-17}$$



我们证明了定理的第一部分. 对定理的第二部分, 因为 $u_{II}^{k+1}(\alpha) = P_\Omega[u^k - \alpha\beta_k F(\tilde{u}^k)]$ 和 $u^* \in \Omega$, 根据投影的性质和余弦定理 (参考右上图), 有

$$\|u_{II}^{k+1}(\alpha) - u^*\|^2 \le \|u^k - \alpha\beta_k F(\tilde{u}^k) - u^*\|^2 - \|u^k - \alpha\beta_k F(\tilde{u}^k) - u_{II}^{k+1}(\alpha)\|^2. \tag{B1-18}$$

因此, 利用 $\zeta_k(\alpha)$ 的定义(见 (B1-14)), 我们有

$$\begin{aligned}
\zeta_k(\alpha) &\ge \|u^k - u^*\|^2 - \|(u^k - u^*) - \alpha\beta_k F(\tilde{u}^k)\|^2 + \|(u^k - u_{II}^{k+1}(\alpha)) - \alpha\beta_k F(\tilde{u}^k)\|^2 \\
&= 2\alpha(u^k - u^*)^T \beta_k F(\tilde{u}^k) + 2\alpha(u_{II}^{k+1}(\alpha) - u^k)^T \beta_k F(\tilde{u}^k) + \|u^k - u_{II}^{k+1}(\alpha)\|^2 \\
&= \|u^k - u_{II}^{k+1}(\alpha)\|^2 + 2\alpha(u_{II}^{k+1}(\alpha) - u^*)^T \beta_k F(\tilde{u}^k).
\end{aligned} \tag{B1-19}$$

将 (B1-19) 中右端的最后一项 $(u_{II}^{k+1}(\alpha) - u^*)^T \beta_k F(\tilde{u}^k)$ 分拆成

$$(u_{II}^{k+1}(\alpha) - u^*)^T \beta_k F(\tilde{u}^k) = (u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T \beta_k F(\tilde{u}^k) + (\tilde{u}^k - u^*)^T \beta_k F(\tilde{u}^k).$$

利用单调性, $(\tilde{u}^k - u^*)^T \beta_k F(\tilde{u}^k) \ge (\tilde{u}^k - u^*)^T \beta_k F(u^*)$, 上式右端第二部分非负. 代入 (B1-19), 进一步得到

$$\zeta_k(\alpha) \ge \|u^k - u_{II}^{k+1}(\alpha)\|^2 + 2\alpha(u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T \beta_k F(\tilde{u}^k). \tag{B1-20}$$

因为 $u_{II}^{k+1}(\alpha) \in \Omega$, 用它替代 (B1-7) 中的任意 $u \in \Omega$, 得到

$$(u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T \beta_k F(\tilde{u}^k) \ge (u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \tag{B1-21}$$

将 (B1-21) 代入 (B1-20) 的右端, 就有

$$\zeta_k(\alpha) \ge \|u^k - u_{II}^{k+1}(\alpha)\|^2 + 2\alpha(u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \tag{B1-22}$$

对上式右端, 利用 $q_k(\alpha)$ 的形式 (见(B1-15)), 就化成

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

利用 (B1-11), 上式右端的第一部分就是 $\|u_I^{k+1}(\alpha) - u_{II}^{k+1}(\alpha)\|^2$. 这样就完成了结论 (B2-16) 的证明. □

这个定理说明, $q_k(\alpha)$ 是 $\vartheta_k(\alpha)$ 和 $\zeta_k(\alpha)$ 的下界. 对同样的 α, $\zeta_k(\alpha)$ 优于 $\vartheta_k(\alpha)$. 由于 $q_k(\alpha)$ 在

$$\alpha_k^* = \arg\max\{q_k(\alpha)\} = (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) / \|d(u^k, \tilde{u}^k)\|^2 \tag{B1-23}$$

处取得最大值(这也是短篇 **A** § **1**中步长) α_k^* 的取法). 在实际计算中, 我们采用校正公式

$$(\text{收缩算法–1}) \quad u_I^{k+1} = u^k - \gamma\alpha_k^* d(u^k, \tilde{u}^k) \tag{B1-24}$$

或者

$$(\text{收缩算法–2}) \quad u_{II}^{k+1} = P_\Omega[u^k - \gamma\alpha_k^* \beta_k F(\tilde{u}^k)] \tag{B1-25}$$

产生新的迭代点 u^{k+1}, 其中的 α_k^* 都由 (B1-23) 给出, 松弛因子 $\gamma \in (0, 2)$, 我们往往取 $\gamma \in [1.2, 1.8]$.

采用校正公式 (B1-24), 它的好处是生成 u^{k+1} 不用再做投影. 实际问题中, 到 Ω 上的投影代价往往不高 (例如 Ω 常常是一个正卦限或者框形), 因此常采用校正公式 (B1-25). 这方面的理由在我主页系列讲义的前三讲中有更详细的说明. 相关的收敛速率的结论和证明可以参考第五讲.

B2-线性变分不等式投影收缩算法中的孪生方向和姊妹方法

奇妙的孪生方向和姊妹方法 maths.nju.edu.cn/~hebma 何炳生

1 孪生方向和相同步长的姊妹方法（VI的应用及PC算法建议参阅本人主页系列讲义前五讲）

设 $\Omega \subset \Re^n$ 是一个非空闭凸集, $M \in \Re^{n\times n}, q \in \Re^n$. 考虑求解线性单调变分不等式

$$\text{(LVI)} \qquad u^* \in \Omega, \quad (u-u^*)^T(Mu^*+q) \geq 0, \quad \forall\, u \in \Omega. \tag{B2-1}$$

如果上式中 $\Omega = \Re^n_+$, LVI (B2-1) 就是一个线性互补问题 (Linear Complementarity Problem):

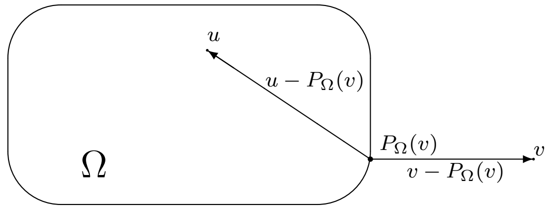
$$\text{(LCP)} \qquad u^* \geq 0, \qquad Mu^*+q \geq 0, \qquad (u^*)^T(Mu^*+q)=0. \tag{B2-2}$$

换句话说, LCP 是 $\Omega = \Re^n_+$ 的 LVI. 说一个 LVI 单调, 是指其中不一定对称的矩阵 M 满足 $(M^T+M) \succeq 0$.

在求解线性变分不等式 (B2-1) 的投影收缩算法中, 对给定的当前点 u^k 和 常数$\beta > 0$, 我们利用投影

$$\tilde{u}^k = P_\Omega[u^k - \beta(Mu^k+q)] \tag{B2-3}$$

生成一个预测点 $\tilde{u}^k$. 在投影 (B2-3) 中, 要求参数 β 大致调整成 $\beta\|M(u^k-\tilde{u}^k)\| \approx \|u^k-\tilde{u}^k\|$.



假如 $\tilde{u}^k = u^k$, u^k 就是 (B2-1) 的解. 否则, 就要进行基于以下分析的新的迭代.

凸集上投影有如右上图所示的重要性质: 对任意的 $v \in \Re^n$, 用 $P_\Omega(v)$ 表示 v 到 Ω 上的投影, 有

$$(u-P_\Omega(v))^T(v-P_\Omega(v)) \leq 0, \quad \forall u \in \Omega. \tag{B2-4}$$

在 (B2-4) 中令 $v = u^k - \beta(Mu^k+q)$, 那么由 (B2-3) 得到 $\tilde{u}^k = P_\Omega(v)$, 并从 (B2-4) 得到

$$\tilde{u}^k \in \Omega, \quad (u-\tilde{u}^k)^T(u^k-\tilde{u}^k) \leq (u-\tilde{u}^k)^T\beta(Mu^k+q), \quad \forall u \in \Omega. \tag{B2-5}$$

上式两边都加上 $(u-\tilde{u}^k)^T\beta M^T(u^k-\tilde{u}^k)$, 由此得到我们基于投影 (B2-3) 的预测公式

$$\textbf{[预测]} \quad (u-\tilde{u}^k)^T(I+\beta M^T)(u^k-\tilde{u}^k) \leq (u-\tilde{u}^k)^T\beta[M^T(u^k-\tilde{u}^k)+(Mu^k+q)], \ \ \forall u \in \Omega. \tag{B2-6}$$

定义 1（孪生方向） 在基于投影 (B2-3) 的预测变分不等式 (B2-6) 中, 分处两端的

$$(I+\beta M^T)(u^k-\tilde{u}^k) \quad \text{和} \quad \beta[M^T(u^k-\tilde{u}^k)+(Mu^k+q)] \quad \text{称为一对孪生方向.} \tag{B2-7}$$

$$\text{记} \quad d(u^k,\tilde{u}^k) = (I+\beta M^T)(u^k-\tilde{u}^k), \qquad g(u^k,\tilde{u}^k) = \beta[M^T(u^k-\tilde{u}^k)+(Mu^k+q)]. \tag{B2-8}$$

定义 2（姊妹方法） 称采用 (B2-7) 中一对孪生方向进行校正的方法:

$$u_I^{k+1}(\alpha) = u^k - \alpha d(u^k,\tilde{u}^k) \qquad \text{和} \qquad u_{II}^{k+1}(\alpha) = P_\Omega[u^k - \alpha g(u^k,\tilde{u}^k)] \tag{B2-9}$$

为一对姊妹方法. 其中 α 是相同的待定步长.

相应地, 对任意给定的解点 $u^* \in \Omega^*$, 我们记

$$\vartheta_k(\alpha) = \|u^k-u^*\|^2 - \|u_I^{k+1}(\alpha)-u^*\|^2 \quad \text{和} \quad \zeta_k(\alpha) = \|u^k-u^*\|^2 - \|u_{II}^{k+1}(\alpha)-u^*\|^2. \tag{B2-10}$$

它们是依赖于步长 α 的 k 次迭代的收益, 是步长 α 的函数.

我们不能直接极大化 $\vartheta_k(\alpha)$ 和 $\zeta_k(\alpha)$, 因为它含有未知的 u^*. 下面将证明二次函数

$$q_k(\alpha) = 2\alpha\|u^k-\tilde{u}^k\|^2 - \alpha^2\|d(u^k,\tilde{u}^k)\|^2 \tag{B2-11}$$

是它们的下界. 在 §2 中, 我们将证明, 对同样的 α, $\zeta_k(\alpha)$ 优于 $\vartheta_k(\alpha)$. 用姊妹方法中的第二种方法为好!

这个结果的证明难度比**B1**中非线性的情形要大一些. 由于 $\tilde{u}^k \in \Omega$, 根据 LVI 的定义 (B2-1), 我们有

$$(\tilde{u}^k-u^*)^T\beta(Mu^*+q) \geq 0, \quad \forall u^* \in \Omega^*. \tag{B2-12}$$

上式可以改写成 $\{(u^k-u^*)-(u^k-\tilde{u}^k)\}^T\beta\{(Mu^k+q)-M(u^k-u^*)\} \geq 0$. 由 $M^T+M \succeq 0$, 得到

$$(u^k-u^*)^T\beta[M^T(u^k-\tilde{u}^k)+(Mu^k+q)] \geq (u^k-\tilde{u}^k)^T\beta(Mu^k+q), \quad \forall u^* \in \Omega^*. \tag{B2-13}$$

将 (B2-5) 中任意的 $u \in \Omega$ 设成 u^*, 则有

$$(\tilde{u}^k-u^*)^T\{(u^k-\tilde{u}^k)-\beta(Mu^k+q)\} \geq 0, \quad \forall u^* \in \Omega^*. \tag{B2-14}$$

将 (B2-12) 和 (B2-14) 相加得到 $\{(u^k-u^*)-(u^k-\tilde{u}^k)\}^T\beta\{(u^k-\tilde{u}^k)-M(u^k-u^*)\} \geq 0$, 利用 (B2-8) 中记号 $d(u^k,\tilde{u}^k)$ 得到

$$(u^k-u^*)^Td(u^k,\tilde{u}^k) \geq \|u^k-\tilde{u}^k\|^2, \quad \forall u^* \in \Omega^*. \tag{B2-15}$$

2 姊妹方法的收缩性质 (何炳生 2016 高校计算数学学报 38 卷第 1 期)

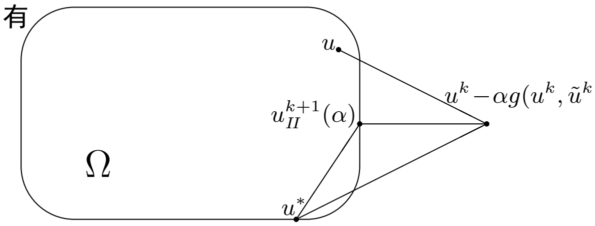
定理 设 $u_I^{k+1}(\alpha)$ 和 $u_{II}^{k+1}(\alpha)$ 由姊妹方法生成. 对由 (B2-10) 定义的 $\vartheta_k(\alpha)$ 和 $\zeta_k(\alpha)$, 有

$$\vartheta_k(\alpha) \ge q_k(\alpha) \qquad \text{和} \qquad \zeta_k(\alpha) \ge q_k(\alpha) + \|u_I^{k+1}(\alpha) - u_{II}^{k+1}(\alpha)\|^2, \tag{B2-16}$$

其中 $q_k(\alpha)$ 由 (B1-15) 给出.

证明. 根据 (B2-10) 对 $\vartheta_k(\alpha)$ 的定义, 利用 (B2-15) 我们有

$$\begin{aligned}
\vartheta_k(\alpha) &= \|u^k - u^*\|^2 - \|u_I^{k+1}(\alpha) - u^*\|^2 \\
&= \|u^k - u^*\|^2 - \|u^k - u^* - \alpha d(u^k, \tilde{u}^k)\|^2 \\
&= 2\alpha(u^k - u^*)^T d(u^k, \tilde{u}^k) - \alpha^2\|d(u^k, \tilde{u}^k)\|^2 \\
&\ge 2\alpha\|u^k - \tilde{u}^k\|^2 - \alpha^2\|d(u^k, \tilde{u}^k)\|^2 \\
&= q_k(\alpha). \quad \text{(see(B2-11))}
\end{aligned} \tag{B2-17}$$



我们证明了定理的第一部分. 对定理的第二部分, 因为 $u_{II}^{k+1}(\alpha) = P_\Omega[u^k - \alpha\beta_k F(\tilde{u}^k)]$ 和 $u^* \in \Omega$, 根据投影的性质和余弦定理 (参考右上图), 有

$$\|u_{II}^{k+1}(\alpha) - u^*\|^2 \le \|u^k - \alpha g(u^k, \tilde{u}^k) - u^*\|^2 - \|u^k - \alpha g(u^k, \tilde{u}^k) - u_{II}^{k+1}(\alpha)\|^2. \tag{B2-18}$$

因此, 利用 $\zeta_k(\alpha)$ 的定义(见 (B2-10)), 我们有

$$\begin{aligned}
\zeta_k(\alpha) &\ge \|u^k - u^*\|^2 - \|(u^k - u^*) - \alpha g(u^k, \tilde{u}^k)\|^2 + \|(u^k - u_{II}^{k+1}(\alpha)) - \alpha g(u^k, \tilde{u}^k)\|^2 \\
&= \|u^k - u_{II}^{k+1}(\alpha)\|^2 + 2\alpha(u_{II}^{k+1}(\alpha) - u^*)^T g(u^k, \tilde{u}^k).
\end{aligned} \tag{B2-19}$$

将 (B2-19) 中右端的后一项 $(u_{II}^{k+1}(\alpha) - u^*)^T g(u^k, \tilde{u}^k)$ 分拆成

$$(u_{II}^{k+1}(\alpha) - u^*)^T g(u^k, \tilde{u}^k) = (u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T g(u^k, \tilde{u}^k) + (\tilde{u}^k - u^*)^T g(u^k, \tilde{u}^k). \tag{B2-20}$$

先看 (B2-20) 右端的第一部分, 利用 $u_{II}^{k+1}(\alpha) \in \Omega$ 和 (B2-6), 我们得到

$$\begin{aligned}
(u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T g(u^k, \tilde{u}^k) &\ge (u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T d(u^k, \tilde{u}^k) \\
&= (u_{II}^{k+1}(\alpha) - u^k)^T d(u^k, \tilde{u}^k) + (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k).
\end{aligned} \tag{B2-21}$$

再看 (B2-20) 右端的第二部分, $(\tilde{u}^k - u^*)^T g(u^k, \tilde{u}^k)$, 继续分拆

$$(\tilde{u}^k - u^*)^T g(u^k, \tilde{u}^k) = (u^k - u^*)^T g(u^k, \tilde{u}^k) - (u^k - \tilde{u}^k)^T g(u^k, \tilde{u}^k). \tag{B2-22}$$

利用记号 $g(u^k, \tilde{u}^k)$ 和 $(u^k - u^*)^T g(u^k, \tilde{u}^k) \ge (u^k - \tilde{u}^k)^T \beta(Mu^k + q)$ (见 (B2-13)), 我们得到

$$\begin{aligned}
(\tilde{u}^k - u^*)^T g(u^k, \tilde{u}^k) &= (u^k - u^*)^T g(u^k, \tilde{u}^k) - (u^k - \tilde{u}^k)^T g(u^k, \tilde{u}^k) \\
&\ge (u^k - \tilde{u}^k)^T \beta(Mu^k + q) - (u^k - \tilde{u}^k)^T \beta[M^T(u^k - \tilde{u}^k) + (Mu^k + q)] \\
&= -\beta(u^k - \tilde{u}^k)^T M^T (u^k - \tilde{u}^k).
\end{aligned} \tag{B2-23}$$

将 (B2-20), (B2-21) 和 (B2-23) 相加, 就有

$$(u_{II}^{k+1}(\alpha) - u^*)^T g(u^k, \tilde{u}^k) \ge (u_{II}^{k+1}(\alpha) - u^k)^T d(u^k, \tilde{u}^k) + \|u^k - \tilde{u}^k\|^2. \tag{B2-24}$$

再将 (B2-24) 代入 (B2-19), 便得

$$\begin{aligned}
\zeta_k(\alpha) &\ge \|u^k - u_{II}^{k+1}(\alpha)\|^2 + 2\alpha(u_{II}^{k+1}(\alpha) - u^k)^T d(u^k, \tilde{u}^k) + 2\alpha\|u^k - \tilde{u}^k\|^2 \\
&= \|(u^k - u_{II}^{k+1}(\alpha)) - \alpha d(u^k, \tilde{u}^k)\|^2 - \alpha^2\|d(u^k, \tilde{u}^k)\|^2 + 2\alpha\|u^k - \tilde{u}^k\|^2 \\
&= \|u_I^{k+1}(\alpha) - u_{II}^{k+1}(\alpha)\|^2 + q_k(\alpha).
\end{aligned}$$

这样就完成了定理结论 (B2-16) 的证明. □

这个定理说明, $q_k(\alpha)$ 是 $\vartheta_k(\alpha)$ 和 $\zeta_k(\alpha)$ 的下界. 对同样的 α, $\zeta_k(\alpha)$ 优于 $\vartheta_k(\alpha)$. 由于 $q_k(\alpha)$ 在

$$\alpha_k^* = \arg\max\{q_k(\alpha)\} = \|u^k - \tilde{u}^k)\|^2 / \|d(u^k, \tilde{u}^k)\|^2 \tag{B2-25}$$

处取得最大值(这也是短篇 **A** § **1**中步长) α_k^* 的取法). 在实际计算中, 我们采用校正公式

$$u_I^{k+1} = u^k - \gamma\alpha_k^* d(u^k, \tilde{u}^k) \qquad\qquad u_{II}^{k+1} = P_\Omega[u^k - \gamma\alpha_k^* g(u^k, \tilde{u}^k)] \tag{B2-26}$$

产生新的迭代点 u^{k+1}, 其中的 α_k^* 都由 (B2-25) 给出, 松弛因子 $\gamma \in (0, 2)$, 我们往往取 $\gamma \in [1.2, 1.8]$.

C1-利用变分不等式证明乘子交替方向法 (ADMM) 的关键收敛性质

以变分不等式为工具的证明方法 maths.nju.edu.cn/~hebma 何炳生

1 两个可分离块的凸优化问题和相应的变分不等式

乘子交替方向法(ADMM)处理的是两个可分离块的凸优化问题

$$\min\{\theta_1(x)+\theta_2(y) \mid Ax+By=b,\ x\in\mathcal{X}, y\in\mathcal{Y}\}, \tag{C1-1}$$

其中 $\theta_1(x), \theta_2(y)$ 是凸函数, A, B, b 是相应的矩阵和向量, $\mathcal{X},\mathcal{Y}$ 是闭凸集. 该问题的拉格朗日函数是

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

我们致力于求问题 (C1-1) 的拉格朗日函数的鞍点 $\big(x^*,y^*,\lambda^*\big)$. 鞍点满足

$$(x^*,y^*,\lambda^*)\in\Omega,\quad L(x,y,\lambda^*)\ge L(x^*,y^*,\lambda^*)\ge L(x^*,y^*,\lambda),\quad \forall(x,y,\lambda)\in\Omega, \tag{C1-2}$$

其中 $\Omega=\mathcal{X}\times\mathcal{Y}\times\Re^m$. 鞍点中的 (x^*,y^*) 就是原问题 (C1-1) 的解点.

刻画鞍点的不等式 (C1-2), 对 x,y,λ 分开来写就是

$$\left\{\begin{array}{lll} x^*\in\mathcal{X}, & L(x,y^*,\lambda^*)-L(x^*,y^*,\lambda^*)\ge 0, & \forall x\in\mathcal{X},\\ y^*\in\mathcal{Y}, & L(x^*,y,\lambda^*)-L(x^*,y^*,\lambda^*)\ge 0, & \forall y\in\mathcal{Y},\\ \lambda^*\in\Re^m, & L(x^*,y^*,\lambda^*)-L(x^*,y^*,\lambda)\ge 0, & \forall \lambda\in\Re^m. \end{array}\right. \tag{C1-3}$$

利用拉格朗日函数 $L(x,y,\lambda)$ 的表达式, 将 (C1-3) 的每个不等式写出来, 就得到下面紧凑的变分不等式

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega, \tag{C1-4a}$$

其中

$$w=\begin{pmatrix}x\\y\\\lambda\end{pmatrix},\quad u=\begin{pmatrix}x\\y\end{pmatrix},\quad \theta(u)=\theta_1(x)+\theta_2(y),\quad F(w)=\begin{pmatrix}-A^T\lambda\\-B^T\lambda\\Ax+By-b\end{pmatrix}. \tag{C1-4b}$$

2 乘子交替方向法及其变分不等式表示

ADMM 的 k 步迭代是从给定的 $v^k=(y^k,\lambda^k)$ 出发, 按下面的顺序求得 $(x^{k+1},y^{k+1},\lambda^{k+1})$:

$$\left\{\begin{array}{ll} x^{k+1}\in\arg\min\big\{\theta_1(x)-x^TA^T\lambda^k+\frac12\beta\|Ax+By^k-b\|^2 \;\big|\; x\in\mathcal{X}\big\}, & \text{(C1-5a)}\\ y^{k+1}\in\arg\min\big\{\theta_2(y)-y^TB^T\lambda^k+\frac12\beta\|Ax^{k+1}+By-b\|^2 \;\big|\; y\in\mathcal{Y}\big\}, & \text{(C1-5b)}\\ \lambda^{k+1}=\ \lambda^k-\beta(Ax^{k+1}+By^{k+1}-b). & \text{(C1-5c)} \end{array}\right.$$

由于 k 步迭代从给定的 (y^k,λ^k) 开始, x^{k+1} 是计算得来的, 我们把 (y,λ) 称作核心变量. 称 x 为中间变量.

根据凸优化的最优性原理 (参见 **D** 中定理), ADMM 第 k 步迭代 (C1-5) 的三个子问题最优性条件分别是

$$\left\{\begin{array}{ll} x^{k+1}\in\mathcal{X}, & \theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\{-A^T\lambda^k+\beta A^T(Ax^{k+1}+By^k-b)\}\ge 0,\ \forall x\in\mathcal{X},\\ y^{k+1}\in\mathcal{Y}, & \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\{-B^T\lambda^k+\beta B^T(Ax^{k+1}+By^{k+1}-b)\}\ge 0,\ \forall y\in\mathcal{Y},\\ \lambda^{k+1}\in\Re^m, & (\lambda-\lambda^{k+1})^T\{(Ax^{k+1}+By^{k+1}-b)+\frac1\beta(\lambda^{k+1}-\lambda^k)\}\ge 0,\ \forall\lambda\in\Re^m. \end{array}\right.$$

跟任何一个向量内积都非负的只能是零向量, 最后一行相当于 $(Ax^{k+1}+By^{k+1}-b)+\frac1\beta(\lambda^{k+1}-\lambda^k)=0$.

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

$$\left\{\begin{array}{ll} x^{k+1}\in\mathcal{X}, & \theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\{-A^T\lambda^{k+1}+\beta A^TB(y^k-y^{k+1})\}\ge 0,\ \forall x\in\mathcal{X}, \quad\text{(C1-6a)}\\ y^{k+1}\in\mathcal{Y}, & \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\{-B^T\lambda^{k+1}\}\ge 0,\ \forall y\in\mathcal{Y}, \quad\text{(C1-6b)}\\ \lambda^{k+1}\in\Re^m, & (\lambda-\lambda^{k+1})^T\{(Ax^{k+1}+By^{k+1}-b)+\frac1\beta(\lambda^{k+1}-\lambda^k)\}\ge 0,\ \forall\lambda\in\Re^m. \quad\text{(C1-6c)} \end{array}\right.$$

在不等式 (C1-6b) 的左端后面加上和为零的两项 $(\beta B^TB(y^k-y^{k+1})+\beta B^TB(y^{k+1}-y^k))$, 上式改写成

$$\left\{\begin{array}{l} \underline{\theta_1(x)-\theta_1(x^{k+1})}+(x-x^{k+1})^T\{\underline{-A^T\lambda^{k+1}}+\beta A^TB(y^k-y^{k+1})\}\ge 0,\\ \underline{\theta_2(y)-\theta_2(y^{k+1})}+(y-y^{k+1})^T\{\underline{-B^T\lambda^{k+1}}+\beta B^TB(y^k-y^{k+1})+\beta B^TB(y^{k+1}-y^k)\}\ge 0,\\ (\lambda-\lambda^{k+1})^T\{(\underline{Ax^{k+1}+By^{k+1}-b})+\frac1\beta(\lambda^{k+1}-\lambda^k)\}\ge 0. \end{array}\right.$$

对上式进行合理整合 (同一类型的并在一起), 利用变分不等式 (C1-4) 中的记号, 得到ADMM的VI表示

$$\underline{\theta(u)-\theta(u^{k+1})}+(w-w^{k+1})^T\underline{F(w^{k+1})}$$
$$+\begin{pmatrix} x-x^{k+1}\\ y-y^{k+1}\end{pmatrix}^T \underbrace{\beta\begin{pmatrix}A^T\\ B^T\end{pmatrix}B(y^k-y^{k+1})}+\begin{pmatrix} y-y^{k+1}\\ \lambda-\lambda^{k+1}\end{pmatrix}^T\begin{pmatrix}\beta B^TB & 0\\ 0 & \frac{1}{\beta}I_m\end{pmatrix}\begin{pmatrix} y^{k+1}-y^k\\ \lambda^{k+1}-\lambda^k\end{pmatrix}\ge 0,\ \forall w\in\Omega. \quad \text{(C1-7)}$$

3 基于变分不等式表示的交替方向法收敛性证明

将 ADMM 的变分不等式表示 (C1-7) 中属于 Ω 的 w, 设成某个确定的解点 w^*, 便有

$$\theta(u^*)-\theta(u^{k+1})+(w^*-w^{k+1})^TF(w^{k+1})$$
$$+\begin{pmatrix} x^*-x^{k+1}\\ y^*-y^{k+1}\end{pmatrix}^T \beta\begin{pmatrix}A^T\\ B^T\end{pmatrix}B(y^k-y^{k+1})+\begin{pmatrix} y^*-y^{k+1}\\ \lambda^*-\lambda^{k+1}\end{pmatrix}^T\begin{pmatrix}\beta B^TB & 0\\ 0 & \frac{1}{\beta}I_m\end{pmatrix}\begin{pmatrix} y^{k+1}-y^k\\ \lambda^{k+1}-\lambda^k\end{pmatrix}\ge 0. \quad \text{(C1-8)}$$

若用记号
$$v=\begin{pmatrix} y\\ \lambda\end{pmatrix} \quad 和 \quad H=\begin{pmatrix}\beta B^TB & 0\\ 0 & \frac{1}{\beta}I_m\end{pmatrix}, \quad \text{(C1-9)}$$
不等式 (C1-8) 就可以改写成
$$(v^{k+1}-v^*)^TH(v^k-v^{k+1}) \ \ge\ \begin{pmatrix} x^{k+1}-x^*\\ y^{k+1}-y^*\end{pmatrix}^T \beta\begin{pmatrix}A^T\\ B^T\end{pmatrix}B(y^k-y^{k+1})$$
$$+\{\underbrace{\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^TF(w^{k+1})}\}. \quad \text{(C1-10)}$$

下面我们证明 (C1-10) 式右端非负. 首先, 由于
$$\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^TF(w^{k+1})=\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^TF(w^*)\ge 0.$$
这就是说, (C1-10) 式右端第二部分 (下波纹线那部分) 非负.

因此, 从 (C1-10) 得到 $(v^{k+1}-v^*)^TH(v^k-v^{k+1}) \ \ge\ \begin{pmatrix} x^{k+1}-x^*\\ y^{k+1}-y^*\end{pmatrix}^T \beta\begin{pmatrix}A^T\\ B^T\end{pmatrix}B(y^k-y^{k+1}).$ (C1-11)

我们接着对 (C1-11) 的右端化简, 注意到
$$\begin{pmatrix} x^{k+1}-x^*\\ y^{k+1}-y^*\end{pmatrix}^T \beta\begin{pmatrix}A^T\\ B^T\end{pmatrix}B(y^k-y^{k+1})=(y^k-y^{k+1})^TB^T\beta(A,B)\begin{pmatrix} x^{k+1}-x^*\\ y^{k+1}-y^*\end{pmatrix}$$
$$=\ (y^k-y^{k+1})^TB^T\beta\big(Ax^{k+1}+By^{k+1}-(Ax^*+By^*)\big) \qquad 利用(Ax^*+By^*=b)$$
$$=\ (y^k-y^{k+1})^TB^T\underbrace{\beta\big(Ax^{k+1}+By^{k+1}-b\big)}=(y^k-y^{k+1})^TB^T\underbrace{(\lambda^k-\lambda^{k+1})}. \quad \text{(C1-12)}$$

下面证明 (C1-12) 式的右端 $(y^k-y^{k+1})^TB^T(\lambda^k-\lambda^{k+1})\ge 0$, 要用到一点小技巧.

利用 (C1-6b) 有
$$\theta_2(y)-\theta_2(y^{k+1}) \ +\ (y-y^{k+1})^T(-B^T\lambda^{k+1}) \ \ge\ 0,\ \forall\, y\in\mathcal{Y},$$
$$和\ \ \theta_2(y)-\theta_2(y^{k}) \ +\ (y-y^{k})^T(-B^T\lambda^{k}) \ \ge\ 0,\ \forall\, y\in\mathcal{Y}.$$

(将上面两式中任意的 y 分别设成 y^k 和 y^{k+1})
$$\theta_2(y^k)-\theta_2(y^{k+1}) \ +\ (y^k-y^{k+1})^T(-B^T\lambda^{k+1}) \ \ge 0.$$
$$\theta_2(y^{k+1})-\theta_2(y^{k}) \ +\ (y^{k+1}-y^{k})^T(-B^T\lambda^{k}) \ \ge 0.$$

(将上面两式相加, 就有) $\quad (y^k-y^{k+1})B^T(\lambda^k-\lambda^{k+1})\ge 0.$ $\quad$ ((C1-12) 式右端非负)

上面证明了 $(y^k-y^{k+1})^TB^T(\lambda^k-\lambda^{k+1})\ge 0$, 由 (C1-12) 得到 (C1-11) 式右端非负. 从而得到我们需要的
$$(v^{k+1}-v^*)^TH(v^k-v^{k+1})\ge 0. \quad \text{(C1-13)}$$

若有 $b^TH(a-b)\ge 0$ 就有 $\|b\|_H^2=\|a\|_H^2-2b^TH(a-b)-\|a-b\|_H^2\le\|a\|_H^2-\|a-b\|_H^2.$

置 $a=(v^k-v^*)$ 和 $b=(v^{k+1}-v^*)$, 根据 (C1-13) 就得到证明 ADMM 收敛性的关键不等式
$$\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2-\|v^k-v^{k+1}\|_H^2, \quad \forall v^*\in\mathcal{V}^*. \qquad (核心变量的收缩性质)$$

C2-预测-校正统一框架下乘子交替方向法 (ADMM) 的收敛性证明

以变分不等式为工具的证明方法 maths.nju.edu.cn/~hebma 何炳生

1 两个可分离块的凸优化问题和相应的变分不等式

乘子交替方向法(ADMM)处理的是两个可分离块的凸优化问题

$$\min\{\theta_1(x)+\theta_2(y) \mid Ax+By=b,\, x\in\mathcal{X}, y\in\mathcal{Y}\}, \tag{C2-1}$$

其中 $\theta_1(x),\theta_2(y)$ 是凸函数, A, B, b 是相应的矩阵和向量, $\mathcal{X},\mathcal{Y}$ 是闭凸集. 问题 (C3-1) 的拉格朗日函数的鞍点 $w^*=(x^*,y^*,\lambda^*)$ 等价于变分不等式

$$w^*\in\Omega, \quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0, \quad \forall\, w\in\Omega, \tag{C2-2a}$$

的解点, 其中

$$w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix}, \quad u=\begin{pmatrix} x\\ y\end{pmatrix}, \quad \theta(u)=\theta_1(x)+\theta_2(y), \quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix}. \tag{C2-2b}$$

ADMM 的 k 步迭代是从给定的 $v^k=(y^k,\lambda^k)$ 出发, 按下面的顺序求得 $(x^{k+1},y^{k+1},\lambda^{k+1})$:

$$\left\{\begin{array}{ll} x^{k+1}\in \arg\min\big\{\theta_1(x)-x^TA^T\lambda^k+\frac12\beta\|Ax+By^k-b\|^2 \;\big|\; x\in\mathcal{X}\big\}, & \text{(C2-3a)}\\ y^{k+1}\in \arg\min\big\{\theta_2(y)-y^TB^T\lambda^k+\frac12\beta\|Ax^{k+1}+By-b\|^2 \;\big|\; y\in\mathcal{Y}\big\}, & \text{(C2-3b)}\\ \lambda^{k+1}= \lambda^k-\beta(Ax^{k+1}+By^{k+1}-b). & \text{(C2-3c)}\end{array}\right.$$

2 乘子交替方向法分拆成统一框架下的预测-校正

根据 (C2-3) 产生的 w^{k+1}, 我们定义预测点

$$\tilde{x}^k=x^{k+1}, \quad \tilde{y}^k=y^{k+1}, \quad \tilde{\lambda}^k=\lambda^k-\beta(A\tilde{x}^k+By^k-b). \tag{C2-4}$$

就可以把交替方向法 (C2-3) 故意分拆成一个预测-校正方法. 首先, 由 (C2-3) 和 (C3-3), 预测就是

ADMM 的 k 步迭代是从给定的 $v^k=(y^k,\lambda^k)$ 出发, 按下面的顺序求得预测点 $(\tilde{x}^k,\tilde{y}^k,\tilde{\lambda}^k)$:

$$\left\{\begin{array}{ll} \tilde{x}^{k}\in \arg\min\big\{\theta_1(x)-x^TA^T\lambda^k+\frac12\beta\|Ax+By^k-b\|^2 \;\big|\; x\in\mathcal{X}\big\}, & \text{(C2-5a)}\\ \tilde{y}^{k}\in \arg\min\big\{\theta_2(y)-y^TB^T\lambda^k+\frac12\beta\|A\tilde{x}^{k}+By-b\|^2 \;\big|\; y\in\mathcal{Y}\big\}, & \text{(C2-5b)}\\ \tilde{\lambda}^{k}= \lambda^k-\beta(A\tilde{x}^{k}+By^{k}-b). & \text{(C2-5c)}\end{array}\right.$$

根据凸优化的最优性原理 (参见 **D** 中定理), 第 k 迭代预测 (C2-7) 的三个子问题的变分不等式形式是

$$\left\{\begin{array}{ll} \tilde{x}^k\in\mathcal{X}, & \theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T\big\{-A^T\lambda^k+\beta A^T(A\tilde{x}^k+By^k-b)\big\}\ge 0, \quad \forall x\in\mathcal{X},\\ \tilde{y}^k\in\mathcal{Y}, & \theta_2(y)-\theta_2(\tilde{y}^k)+(y-\tilde{y}^k)^T\big\{-B^T\lambda^k+\beta B^T(A\tilde{x}^k+B\tilde{y}^k-b)\big\}\ge 0, \quad \forall y\in\mathcal{Y},\\ \tilde{\lambda}^k\in\Re^m, & (\lambda-\tilde{\lambda}^k)^T\{(A\tilde{x}^k+By^k-b)+\frac1\beta(\tilde{\lambda}^k-\lambda^k)\}\ge 0, \quad \forall\lambda\in\Re^m.\end{array}\right.$$

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

$$\left\{\begin{array}{lll} \tilde{x}^k\in\mathcal{X}, & \underline{\theta_1(x)-\theta_1(\tilde{x}^k)}+(x-\tilde{x}^k)^T\big\{\underline{-A^T\tilde{\lambda}^k}\big\}\ge 0, \quad \forall x\in\mathcal{X}, & \text{(C2-6a)}\\ \tilde{y}^k\in\mathcal{Y}, & \underline{\theta_2(y)-\theta_2(\tilde{y}^k)}+(y-\tilde{y}^k)^T\big\{\underline{-B^T\tilde{\lambda}^k}+\beta B^TB(\tilde{y}^k-y^k)\big\}\ge 0, \quad \forall y\in\mathcal{Y}, & \text{(C2-6b)}\\ \tilde{\lambda}^k\in\Re^m, & (\lambda-\tilde{\lambda}^k)^T\{\underline{(A\tilde{x}^k+B\tilde{y}^k-b)}-B(\tilde{y}^k-y^k)+\frac1\beta(\tilde{\lambda}^k-\lambda^k)\}\ge 0, \quad \forall\lambda\in\Re^m. & \text{(C2-6c)}\end{array}\right.$$

对上式整合 (把同一类型的并在一起), 并利用变分不等式 (C3-2) 中的记号, 得到ADMM预测的VI表示

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(v-\tilde{v}^k)^TQ(v^k-\tilde{v}^k), \quad \forall w\in\Omega, \quad \text{其中}\;\; Q=\begin{pmatrix}\beta B^TB & 0\\ -B & \frac1\beta I_m\end{pmatrix}. \tag{C2-7}$$

由 (C2-3) 和 (C3-3) 之间的关系, $y^{k+1}=\tilde{y}^k$, $\lambda^{k+1}=\tilde{\lambda}^k+\beta(y^k-\tilde{y}^k)$, 新的核心变量 v^{k+1} 由

$$v^{k+1}=v^k-M(v^k-\tilde{v}^k) \quad \text{生成, 其中} \quad M=\begin{pmatrix} I & 0\\ -\beta B & I_m\end{pmatrix}. \tag{C2-8}$$

这里的 (C2-7) 和 (C2-8) 相当于短篇 **A** §**2** 中统一框架的预测和校正

3 预测-校正分拆的ADMM收敛性质证明

定理 C-2A 采用预测-校正方法 (C2-7)-(C2-8) 求解变分不等式 (C3-2), 则产生的序列 $\{v^k\}$ 具有收缩性质

$$\theta(u)-\theta(\tilde u^k)+(w-\tilde w^k)^TF(\tilde w^k)\ge \tfrac12(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2)+\tfrac12\|v^k-\tilde v^k\|_G^2,\quad \forall w\in\Omega, \tag{C2-9}$$

其中

$$H=\begin{pmatrix}\beta B^TB & 0\\ 0 & \frac1\beta I_m\end{pmatrix},\qquad G=Q^T+Q-M^THM \tag{C2-10}$$

进而有

$$\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2-\|v^k-\tilde v^k\|_G^2,\quad \forall v^*\in\mathcal V^*. \tag{C2-11}$$

证明 对 (C2-10) 中的 H 和 (C2-8) 中的 M, 恰有 $HM=Q$, 因此, 由预测得到的 (C2-7) 可以改写成

$$\theta(u)-\theta(\tilde u^k)+(w-\tilde w^k)^TF(\tilde w^k)\ge(v-\tilde v^k)^THM(v^k-\tilde v^k).$$

由于 $M(v^k-\tilde v^k)=(v^k-v^{k+1})$ (见校正公式(C2-8)), 上式可以写成

$$\theta(u)-\theta(\tilde u^k)+(w-\tilde w^k)^TF(\tilde w^k)\ge(v-\tilde v^k)^TH(v^k-v^{k+1}). \tag{C2-12}$$

将恒等式

$$(a-b)^TH(c-d)=\tfrac12\{\|a-d\|_H^2-\|a-c\|_H^2\}+\tfrac12\{\|c-b\|_H^2-\|d-b\|_H^2\}$$

用于不等式 (C2-12) 的右端, 并设 $a=v,\ b=\tilde v^k,\ c=v^k$ 和 $d=v^{k+1}$, 得到

$$(v-\tilde v^k)^TH(v^k-v^{k+1})=\tfrac12(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2)+\tfrac12(\|v^k-\tilde v^k\|_H^2-\|v^{k+1}-\tilde v^k\|_H^2), \tag{C2-13}$$

对 (C2-13) 的右端, 利用 $HM=Q$ 和 $2v^\top Qv=v^\top(Q^T+Q)v$, 我们有

$$\begin{aligned}\|v^k-\tilde v^k\|_H^2-\|v^{k+1}-\tilde v^k\|_H^2 &= \|v^k-\tilde v^k\|_H^2-\|(v^k-\tilde v^k)-(v^k-v^{k+1})\|_H^2\\ &= \|v^k-\tilde v^k\|_H^2-\|(v^k-\tilde v^k)-M(v^k-\tilde v^k)\|_H^2 \qquad \boxed{\text{利用 } HM=Q}\\ &= (v^k-\tilde v^k)^T(Q^T+Q-M^THM)(v^k-\tilde v^k) = \|v^k-\tilde v^k\|_G^2.\end{aligned} \tag{C2-14}$$

将 (C2-14) 代入 (C2-13), 再代入 (C2-12), 就得到定理的结论 (C2-9). 将 (C2-9) 中的任意的 $w\in\Omega$ 设成一个解点 w^*, 就得到定理的另一结论 (C2-11). □

定理 C-2B 采用预测-校正方法 (C2-7)-(C2-8) 求解变分不等式 (C3-2), 产生的序列 $\{v^k\}$ 和 $\{\tilde v^k$ 具有性质

$$\|v^k-\tilde v^k\|_G^2\ge\|v^k-v^{k+1}\|_H^2. \tag{C2-15}$$

进而有

$$\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2-\|v^k-v^{k+1}\|_H^2,\quad\forall v^*\in\mathcal V^*. \tag{C2-16}$$

Proof. Because $G=Q^T+Q-M^THM$, we have

$$G=Q^T+Q-Q^TM=\begin{pmatrix}0 & 0\\ 0 & \frac1\beta I_m\end{pmatrix}\quad\text{and thus}\quad \|v^k-\tilde v^k\|_G^2=\tfrac1\beta\|\lambda^k-\tilde\lambda^k\|^2. \tag{C2-17}$$

From (C2-6), we know that the optimal condition of the y-subproblem is

$$\tilde y^k\in\mathcal Y,\quad \theta_2(y)-\theta_2(\tilde y^k)+(y-\tilde y^k)^T\{-B^T\tilde\lambda^k+\beta B^TB(\tilde y^k-y^k)\}\ge0,\quad\forall y\in\mathcal Y. \tag{C2-18}$$

According to (C3-3), we get

$$\lambda^{k+1}=\tilde\lambda^k-\beta B(\tilde y^k-y^k)\quad\text{and}\quad\tilde y^k=y^{k+1},$$

and thus the inequality (C2-18) can be written as

$$y^{k+1}\in\mathcal Y,\quad\theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\{-B^T\lambda^{k+1}\}\ge0,\quad\forall y\in\mathcal Y. \tag{C2-19}$$

The above inequality is hold also for the last iteration, *i. e.*, we have

$$y^k\in\mathcal Y,\quad\theta_2(y)-\theta_2(y^k)+(y-y^k)^T\{-B^T\lambda^k\}\ge0,\quad\forall y\in\mathcal Y. \tag{C2-20}$$

Setting $y=y^k$ and $y=y^{k+1}$ in in (C2-19) and (C2-20), respectively, and then adding them, we get

$$(\lambda^k-\lambda^{k+1})^TB(y^k-y^{k+1})\ge0. \tag{C2-21}$$

Using $\lambda^k-\tilde\lambda^k=(\lambda^k-\lambda^{k+1})+\beta B(y^k-y^{k+1})$ and the inequality (C2-21), we obtain

$$\begin{aligned}\tfrac1\beta\|\lambda^k-\tilde\lambda^k\|^2 &= \tfrac1\beta\|(\lambda^k-\lambda^{k+1})+\beta B(y^k-y^{k+1})\|^2\\ &\ge \tfrac1\beta\|\lambda^k-\lambda^{k+1}\|^2+\beta\|B(y^k-y^{k+1})\|^2 = \|v^k-v^{k+1}\|_H^2.\end{aligned} \tag{C2-22}$$

The assertion (C2-15) follows from (C2-17) and (C2-22) directly. Consequently, the assertion (C2-16) follows from (C2-11) and (C2-15). The theorem is proved. □

C3-预测-校正统一框架下乘子交替方向法(ADMM)的收敛速率证明

以变分不等式为工具的证明方法 maths.nju.edu.cn/~hebma 何炳生

1 ADMM 在预测-校正分拆下的性质

乘子交替方向法(ADMM)处理的是两个可分离块的凸优化问题

$$\min\{\theta_1(x)+\theta_2(y) \mid Ax+By=b,\, x\in\mathcal{X}, y\in\mathcal{Y}\}, \tag{C3-1}$$

其中 $\theta_1(x),\theta_2(y)$ 是凸函数, A, B, b下 是相应的矩阵和向量, $\mathcal{X},\mathcal{Y}$ 是闭凸集. 问题 (C3-1) 的拉格朗日函数的鞍点 $w^*=(x^*,y^*,\lambda^*)$ 等价于变分不等式

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega, \tag{C3-2a}$$

的解点, 其中

$$w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix},\quad u=\begin{pmatrix} x\\ y\end{pmatrix},\quad \theta(u)=\theta_1(x)+\theta_2(y),\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix}. \tag{C3-2b}$$

根据经典 ADMM 产生的 w^{k+1}, 通过定义预测点

$$\tilde{x}^k=x^{k+1},\quad \tilde{y}^k=y^{k+1},\quad \tilde{\lambda}^k=\lambda^k-\beta(A\tilde{x}^k+By^k-b). \tag{C3-3}$$

就可以把经典的交替方向法故意分拆成一个预测-校正方法.

[预测] $\quad \tilde{w}^k\in\Omega,\quad \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge v-\tilde{v}^k)^TQ(v^k-\tilde{v}^k),\quad \forall w\in\Omega,$ (C3-4)

[校正] $\quad v^{k+1}=v^k-M(v^k-\tilde{v}^k).$ (C3-5)

其中 $\quad Q=\begin{pmatrix} \beta B^TB & 0\\ -B & \frac{1}{\beta}I_m\end{pmatrix},\qquad M=\begin{pmatrix} I & 0\\ -\beta B & I_m\end{pmatrix}.$

C2 中的主要结论 采用预测-校正方法 (C2-7)-(C2-8) 求解变分不等式 (C3-2), 产生的序列具有性质

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge \tfrac12(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2)+\tfrac12\|v^k-\tilde{v}^k\|_G^2,\ \ \forall w\in\Omega, \tag{C3-6}$$

其中

$$H=\begin{pmatrix} \beta B^TB & 0\\ 0 & \frac{1}{\beta}I_m\end{pmatrix},\qquad G=Q^T+Q-M^THM=\begin{pmatrix} 0 & 0\\ 0 & \frac{1}{\beta}I_m\end{pmatrix}.$$

此外还有

$$\|v^k-v^{k+1}\|_H^2\le\|v^k-v^*\|_H^2-\|v^{k+1}-v^*\|_H^2,\quad \forall\, v^*\in\mathcal{V}^*. \tag{C3-7}$$

2 遍历意义下的收敛性 (He and Yuan 2012 SIAM Numr. Analysis)

Since $(w-\tilde{w})^TF(w)=(w-\tilde{w})^TF(\tilde{w})$, if $\tilde{w}\in\Omega,\ \ \theta(u)-\theta(\tilde{u})+(w-\tilde{w})^TF(w)\ge 0,\ \ \forall\, w\in\Omega$, then $\tilde{w}$ is a solution of (C3-2). For given $\epsilon>0$, $\tilde{w}\in\Omega$ is called an ϵ-approximate solution of (C3-2), if it satisfies

$$\tilde{w}\in\Omega,\ \ \theta(u)-\theta(\tilde{u})+(w-\tilde{w})^TF(w)\ge-\epsilon,\quad \forall\, w\in\mathcal{D}_{(\tilde{w})}\ =\ \{w\in\Omega\,|\,\|w-\tilde{w}\|\le 1\}.$$

We need to show that for given $\epsilon>0$, after t iterations, it can offer a $\tilde{w}\in\mathcal{W}$, such that

$$\tilde{w}\in\mathcal{W}\quad\text{and}\quad \sup_{w\in\mathcal{D}_{(\tilde{w})}}\big\{\theta(\tilde{u})-\theta(u)+(\tilde{w}-w)^TF(w)\big\}\le\epsilon. \tag{C3-8}$$

Using $(w-\tilde{w}^k)^TF(w)=(w-\tilde{w}^k)^TF(\tilde{w}^k)$ and $G\succeq 0$, from (C2-9) we obtain

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(w)+\tfrac12\|v-v^k\|_H^2\ge\tfrac12\|v-v^{k+1}\|_H^2,\quad \forall w\in\Omega. \tag{C3-9}$$

Theorem 2.1 *Let $\{w^k\}$, $\{\tilde{w}^k\}$ be the sequences generated the prediction-correction method* (C2-7)-(C2-8). *Let $\tilde{w}_t$ be defined by*

$$\tilde{w}_t=\frac{1}{t+1}\sum_{k=0}^{t}\tilde{w}^k. \tag{C3-10}$$

Then, for any integer number $t>0$, $\tilde{w}_t\in\Omega$ and

$$\theta(\tilde{u}_t)-\theta(u)+(\tilde{w}_t-w)^TF(w)\le\frac{1}{2(t+1)}\|v-v^0\|_H^2,\quad \forall w\in\Omega. \tag{C3-11}$$

Proof. First, it holds that $\tilde{w}^k \in \Omega$ for all $k \geq 0$. Together with the convexity of $\mathcal{X}$ and $\mathcal{Y}$, (C3-10) implies that $\tilde{w}_t \in \Omega$. Summing the inequality (C3-9) over $k = 0, 1, \ldots, t$, we obtain

$$(t+1)\theta(u) - \sum_{k=0}^{t} \theta(\tilde{u}^k) + \Big((t+1)w - \sum_{k=0}^{t} \tilde{w}^k\Big)^T F(w) + \frac{1}{2}\|v - v^0\|_H^2 \geq 0, \quad \forall w \in \Omega.$$

Using the notation of $\tilde{w}_t$, it can be written as

$$\frac{1}{t+1}\sum_{k=0}^{t} \theta(\tilde{u}^k) - \theta(u) + (\tilde{w}_t - w)^T F(w) \leq \frac{1}{2(t+1)}\|v - v^0\|_H^2, \quad \forall w \in \Omega. \tag{C3-12}$$

Since $\theta(u)$ is convex and

$$\tilde{u}_t = \frac{1}{t+1}\sum_{k=0}^{t} \tilde{u}^k, \qquad \text{we have that} \qquad \theta(\tilde{u}_t) \leq \frac{1}{t+1}\sum_{k=0}^{t} \theta(\tilde{u}^k).$$

Substituting it in (C3-12), the assertion of this theorem follows directly. $\Box$

由结论 (C3-6) 而得到的 (C3-9), 是证明ADMM遍历意义下收敛速率结论 (C3-11) 的主要工具.

3 点列意义下的收敛性 (He and Yuan 2015 Numr. Mathematik)

Theorem 3.1 *Let* $\{w^k\}$, $\{\tilde{w}^k\}$ *be the sequences generated the prediction-correction method* (C2-7)-(C2-8). *Then we have*

$$\|v^{k+1} - v^{k+2}\|_H \leq \|v^k - v^{k+1}\|_H, \quad \forall\, k > 0. \tag{C3-13}$$

Proof. First, we prove the following result:

$$(v^k - v^{k+1})^T H\{(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\} \geq \tfrac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T+Q)}^2. \tag{C3-14}$$

Set $w = \tilde{w}^{k+1}$ in (C2-7), we have

$$\theta(\tilde{u}^{k+1}) - \theta(\tilde{u}^k) + (\tilde{w}^{k+1} - \tilde{w}^k)^T F(\tilde{w}^k) \geq (\tilde{v}^{k+1} - \tilde{v}^k)^T Q(v^k - \tilde{v}^k). \tag{C3-15}$$

Note that (C2-7) is also true for $k := k+1$ and thus we have

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

Set $w = \tilde{w}^k$ in the above inequality, we obtain

$$\theta(\tilde{u}^k) - \theta(\tilde{u}^{k+1}) + (\tilde{w}^k - \tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \geq (\tilde{v}^k - \tilde{v}^{k+1})^T Q(v^{k+1} - \tilde{v}^{k+1}). \tag{C3-16}$$

Combining (C3-15) and (C3-16) and using $(\tilde{w}^k - \tilde{w}^{k+1})^T(F(\tilde{w}^k) - F(\tilde{w}^{k+1}))$, we get

$$(\tilde{v}^k - \tilde{v}^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq 0. \tag{C3-17}$$

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 (C3-17), and using $v^T Q v = \frac{1}{2}v^T(Q^T + Q)v$, we obtain

$$(v^k - v^{k+1})^T Q\{(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\} \geq \tfrac{1}{2}\|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T+Q)}^2.$$

Substituting $(v^k - v^{k+1}) = M(v^k - \tilde{v}^k)$ in the left-hand side of the last inequality and using $Q = HM$, we obtain (C3-14). In the identity $\|a\|_H^2 - \|b\|_H^2 = 2a^T H(a-b) - \|a-b\|_H^2$ setting $a = (v^k - v^{k+1})$ and $b = (v^{k+1} - v^{k+2})$, we obtain

$$\begin{aligned} &\|v^k - v^{k+1}\|_H^2 - \|v^{k+1} - v^{k+2}\|_H^2 \\ = \;& 2(v^k - v^{k+1})^T H\{(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})\} - \|(v^k - v^{k+1}) - (v^{k+1} - v^{k+2})]\|_H^2. \end{aligned}$$

Inserting (C3-14) into the first term of the right-hand side of the last equality, we obtain

$$\begin{aligned} &\|v^k - v^{k+1}\|_H^2 - \|v^{k+1} - v^{k+2}\|_H^2 \\ \geq \;& \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T+Q)}^2 - \|M[(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})]\|_H^2 \\ = \;& \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_{(Q^T+Q-M^T HM)}^2 \geq 0, \end{aligned}$$

where the last inequality is because of the positive definiteness of the matrix $(Q^T + Q) - M^T HM \succeq 0$. The assertion (C3-13) follows immediately. $\Box$

By using $\|v^k - v^{k+1}\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2$ and $\|v^k - v^{k+1}\|_H^2 \leq \|v^{k-1} - v^k\|_H^2$, we get

$$\|v^k - v^{k+1}\|_H^2 \leq \tfrac{1}{k+1}\|v^0 - v^*\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{C3-18}$$

结论 (C3-6)-(C3-7), 是证明ADMM点历意义下收敛速率 (C3-18) 的主要工具. 预测-校正带来许多方便!

D1-凸优化问题对应的变分不等式和因需定制的PPA算法

1 Preliminaries

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

$$x^* \in \arg\min\{\theta(x) + f(x) \mid x \in \mathcal{X}\}$$

if and only if

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

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

$$\|b\|_H^2 \leq \|a\|_H^2 - \|a-b\|_H^2.$$

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

2 Convex optimization and its related variational inequality

1). min − max problem. Let (x^*, y^*) be the solution of the $\min - \max$ problem

$$\min_x \max_y \{\Phi(x,y) = \theta_1(x) - y^T A x - \theta_2(y) \mid x \in \mathcal{X}, y \in \mathcal{Y}\}. \qquad \text{(D1-1)}$$

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

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

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

$$u^* \in \Omega, \quad \theta(u) - \theta(u^*) + (u-u^*)^T F(u^*) \geq 0, \quad \forall\, u \in \Omega, \qquad \text{(D1-2)}$$

where $u = \begin{pmatrix} x \\ y \end{pmatrix}$, $\theta(u) = \theta_1(x) + \theta_2(y)$, $F(u) = \begin{pmatrix} -A^T y \\ Ax \end{pmatrix}$ and $\Omega = \mathcal{X} \times \mathcal{Y}$. (D1-3)

Please notice that for the $F(u)$ in (D1-3), we have $(u-\tilde{u})^T(F(u) - F(\tilde{u})) \equiv 0$.

2). Linearly constrained convex optimization

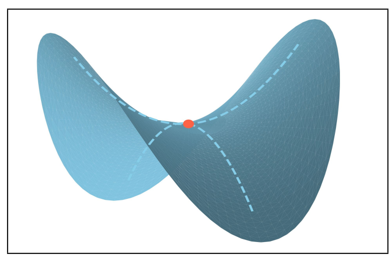
We consider the convex optimization problem

$$\min\{\theta(u) \mid \mathcal{A}u = b, \; u \in \mathcal{U}\}. \qquad \text{(D1-4)}$$

The Lagrangian function of the problem (D1-4) is

$$L(u,\lambda) = \theta(u) - \lambda^T(\mathcal{A}u - b), \qquad \text{(D1-5)}$$

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



A pair of $(u^*, \lambda^*) \in \mathcal{U} \times \Re^m$ is called a saddle point of the Lagrange function (D1-5), if

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

The above inequalities can be written as

$$\left\{\begin{array}{ll} u^* \in \mathcal{U}, & L(u,\lambda^*) - L(u^*,\lambda^*) \geq 0, \;\; \forall\, u \in \mathcal{U}, \qquad \text{(D1-6a)} \\ \lambda^* \in \Re^m, & L(u^*,\lambda^*) - L(u^*,\lambda) \geq 0, \;\; \forall\, \lambda \in \Re^m. \qquad \text{(D1-6b)} \end{array}\right.$$

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

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w-w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega \qquad \text{(D1-7)}$$

where $w = \begin{pmatrix} u \\ \lambda \end{pmatrix}$, $F(w) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{pmatrix}$ and $\Omega = \mathcal{U} \times \Re^m$. (D1-8)

Please notice that $(w-\tilde{w})^T(F(w) - F(\tilde{w})) \equiv 0$. The problem (D1-4) is translated to VI (D1-7).

The problem (D1-1) **and** (D1-4) **are translated to the equivalent VI** (D1-2) **and** (D1-7)**, respectively.**

3 Proximal point algorithm for monotone variational inequalities

Definition (k-th iteration of PPA) $\boldsymbol{H \succ 0}$. Start with a given w^k, find a w^{k+1}, such that

$$w^{k+1} \in \Omega, \;\; \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \ge (w - w^{k+1})^T H(w^k - w^{k+1}), \;\; \forall\, w \in \Omega. \quad \text{(D1-9)}$$

Setting $w = w^*$ in (D2-4), and using $(w^{k+1} - w^*)^T F(w^{k+1}) = (w^{k+1} - w^*)^T F(w^*)$, we obtain

$$(w^{k+1} - w^*)^T H(w^k - w^{k+1}) \ge \theta(u^{k+1}) - \theta(u^*) + (w^{k+1} - w^*)^T F(w^*) \ge 0. \quad \text{(D1-10)}$$

Using Lemma D with $a = (w^k - w^*)$ and $b = (w^{k+1} - w^*)$ in the above inequality, we get

(Convergence Key of PPA) $\quad \|w^{k+1} - w^*\|_H^2 \le \|w^k - w^*\|_H^2 - \|w^k - w^{k+1}\|_H^2, \quad \forall w^* \in \Omega^*. \quad$ (D1-11)

By using PPA to solve VI, usually, we take $\boldsymbol{w^{k+1} := w^k - \alpha(w^k - w^{k+1})}$ with $\boldsymbol{\alpha = 1.5 \in (0, 2)}$.

4 Customized PPA (C-PPA) and Balanced PPA (B-PPA) for VI

1). C-PPA for VI (D1-2) (He, Yuan 2012 SIAM Im., He, Yuan, Zhang 2013 COA, Gu, He, Yuan 2014 COA)

Start with given u^k, find a u^{k+1}, such that (D2-4) is satisfied, where

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad w = u, \quad F(w) = \begin{pmatrix} -A^T y \\ Ax \end{pmatrix}, \quad H = \begin{pmatrix} rI_n & A^T \\ A & sI_m \end{pmatrix}, \quad rs > \|A^T A\|. \quad \text{(D1-12)}$$

The concrete formula of (D2-4) with F and H given by (D1-12) is $\boxed{\text{the uwave parts is } F(w^{k+1})}$

$$\left\{ \begin{array}{l} \theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^T \{ \underset{\sim\sim\sim}{-A^T y^{k+1}} + \boldsymbol{rI_n}(x^{k+1} - x^k) + \boldsymbol{A^T}(y^{k+1} - y^k)\} \ge 0, \; \forall x \in \mathcal{X}, \\ \theta_2(y) - \theta_2(y^{k+1}) \; + \; (y - y^{k+1})^T \{ \underset{\sim\sim\sim}{Ax^{k+1}} + \boldsymbol{A}(x^{k+1} - x^k) + \boldsymbol{sI_m}(y^{k+1} - y^k)\} \ge 0, \; \forall y \in \mathcal{Y}. \end{array} \right.$$

According to Theorem D, to complete PPA iteration, we need only to solve the following sub-problems:

$$\left\{ \begin{array}{ll} x^{k+1} = \text{argmin}\{\theta_1(x) - x^T A^T y^k + \frac{r}{2}\|x - x^k\|^2 \,|\, x \in \mathcal{X}\}, & \text{(D1-13a)} \\ y^{k+1} = \text{argmin}\{\theta_2(y) + y^T A[2x^{k+1} - x^k] + \frac{s}{2}\|y - y^k\|^2 \,|\, y \in \mathcal{Y}\}. & \text{(D1-13b)} \end{array} \right.$$

2). C-PPA and B-PPA for VI (D1-7) (He and Yuan, arXiv:2108.08554; S.J. Xu, 2023, JAMC)

Start with given w^k, find a w^{k+1}, such that (D2-4) is satisfied, where

$$F(w) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{pmatrix}, \; H = H_1 \text{ or } H_2, \; H_1 = \begin{pmatrix} \beta\mathcal{A}^T\mathcal{A} + \delta I_n & \mathcal{A}^T \\ \mathcal{A} & \frac{1}{\beta}I_m \end{pmatrix}, \; H_2 = \begin{pmatrix} rI_n & \mathcal{A}^T \\ \mathcal{A} & \frac{1}{r}\mathcal{A}\mathcal{A}^T + \delta I_m \end{pmatrix}. \quad \text{(D1-14)}$$

Customized PPA. The concrete formula of (D2-4) with $H = H_1$ given by (D1-14) is $\boxed{\text{the uwave parts is } F(w^{k+1})}$

$$\left\{ \begin{array}{l} \theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T \{ \underset{\sim\sim\sim}{-\mathcal{A}^T\lambda^{k+1}} + \; \boldsymbol{(\beta\mathcal{A}^T\mathcal{A} + \delta I_n)}(u^{k+1} - u^k) + \boldsymbol{\mathcal{A}^T}(\lambda^{k+1} - \lambda^k)\} \ge 0, \\ \qquad\qquad\qquad\qquad (\underset{\sim\sim\sim}{\mathcal{A}u^{k+1} - b}) \quad + \boldsymbol{\mathcal{A}}(u^{k+1} - u^k) \quad + \quad \boldsymbol{(1/\beta)}(\lambda^{k+1} - \lambda^k) = 0. \end{array} \right.$$

According to Theorem D, to complete PPA iteration, we need only to solve the following sub-problems:

$$\left\{ \begin{array}{ll} u^{k+1} = \text{argmin}\big\{\theta(u) - u^T\mathcal{A}^T\lambda^k + \frac{1}{2}(u - u^k)^T(\beta\mathcal{A}^T\mathcal{A} + \delta I_n)(u - u^k) \,|\, u \in \mathcal{U}\big\}, & \text{(D1-15a)} \\ \lambda^{k+1} = \lambda^k - \beta\big(\mathcal{A}[2u^{k+1} - u^k] - b\big). & \text{(D1-15b)} \end{array} \right.$$

Balanced PPA. The concrete formula of (D2-4) with $H = H_2$ given by (D1-14) is $\boxed{\text{the uwave parts is } F(w^{k+1})}$

$$\left\{ \begin{array}{l} \theta(u) - \theta(u^{k+1}) + (u - u^{k+1})^T \{ \underset{\sim\sim\sim}{-\mathcal{A}^T\lambda^{k+1}} + \boldsymbol{rI_n}(u^{k+1} - u^k) \quad + \quad \boldsymbol{\mathcal{A}^T} \quad (\lambda^{k+1} - \lambda^k)\} \ge 0, \\ \qquad\qquad\qquad\qquad (\underset{\sim\sim\sim}{\mathcal{A}u^{k+1} - b}) + \boldsymbol{\mathcal{A}}(u^{k+1} - u^k) + \boldsymbol{(\tfrac{1}{r}\mathcal{A}\mathcal{A}^T + \delta I_m)}(\lambda^{k+1} - \lambda^k) = 0. \end{array} \right.$$

According to Theorem D, to complete PPA iteration, we need only to solve the following sub-problems:

$$\left\{ \begin{array}{lll} u^{k+1} = \text{argmin}\big\{\theta(u) - u^T\mathcal{A}^T\lambda^k + \frac{r}{2}\|u - u^k\|^2 \,|\, u \in \mathcal{U}\big\}, & \boxed{\text{This problem is easier than(D1-15a)}} & \text{(D1-16a)} \\ \lambda^{k+1} = \lambda^k - (\frac{1}{r}\mathcal{A}\mathcal{A}^T + \delta I_m)^{-1}\big(\mathcal{A}[2u^{k+1} - u^k] - b\big). & \boxed{\text{Only once Cholesky-Decomposition}} & \text{(D1-16b)} \end{array} \right.$$

In comparison with (D1-15), Algorithm (D1-16) is called a balanced algorithm, which shared the difficulties.

D2-两块可分离凸优化问题平行求解原始子问题的PPA算法

1 Two blocks separable convex optimization and VI

Two blocks separable convex optimization

$$\min\{\theta_1(x)+\theta_2(y) \mid Ax+By=b,\ x\in\mathcal{X}, y\in\mathcal{Y}\}. \tag{D2-1}$$

The saddle point of its Lagrange function is a solution of the following variational inequality:

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^T F(w^*)\ge 0,\quad \forall\, w\in\Omega \tag{D2-2}$$

where

$$u=\begin{pmatrix} x\\ y\end{pmatrix},\quad \theta(u)=\theta_1(x)+\theta_2(y),\quad w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix},\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix} \tag{D2-3}$$

and $\Omega=\mathcal{X}\times\mathcal{Y}\times\Re^m$. Again, we have $(w-\tilde{w})^T(F(w)-F(\tilde{w}))\equiv 0$.

2 PPA for monotone variational inequalities

Definition (k-th iteration of PPA) $\boldsymbol{H}\succ\mathbf{0}$. Start with a given w^k, find a w^{k+1}, such that

$$w^{k+1}\in\Omega,\ \ \theta(u)-\theta(u^{k+1})+(w-w^{k+1})^T F(w^{k+1})\ge (w-w^{k+1})^T H(w^k-w^{k+1}),\ \ \forall\, w\in\Omega. \tag{D2-4}$$

Setting $w=w^*$ in (D2-4), and using $(w^{k+1}-w^*)^T F(w^{k+1})=(w^{k+1}-w^*)^T F(w^*)$, we obtain

$$(w^{k+1}-w^*)^T H(w^k-w^{k+1})\ge \theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^T F(w^*)\ge 0. \tag{D2-5}$$

Using Lemma D with $a=(w^k-w^*)$ and $b=(w^{k+1}-w^*)$ in the above inequality, we get

(Convergence Key of PPA) $$\|w^{k+1}-w^*\|_H^2\le\|w^k-w^*\|_H^2-\|w^k-w^{k+1}\|_H^2,\quad \forall w^*\in\Omega^*. \tag{D2-6}$$

By using PPA to solve VI, usually, we take $\boldsymbol{w^{k+1}:=w^k-\alpha(w^k-w^{k+1})}$ **with** $\boldsymbol{\alpha=1.5\in(0,2)}$**.**

3 Customized PPA for VI (D2-2) (Gu, He, Yuan 2014 COA; He, Yuan, Zhang 2014 COA)

Start with a given w^k, find a w^{k+1}, such that (D2-4) is satisfied, where

$$u=\begin{pmatrix} x\\ y\end{pmatrix},\quad w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix},\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix},\qquad H=H_1 \text{ or } H_2, \tag{D2-7}$$

and

$$H_1=\begin{pmatrix} \beta A^TA+\delta I_{n_1} & 0 & A^T\\ 0 & \beta B^TB+\delta I_{n_2} & B^T\\ A & B & \frac{2}{\beta}I_m\end{pmatrix},\ H_2=\begin{pmatrix} \beta A^TA+\delta I_{n_1} & 0 & -A^T\\ 0 & \beta B^TB+\delta I_{n_2} & -B^T\\ -A & -B & \frac{2}{\beta}I_m\end{pmatrix}. \tag{D2-8}$$

1). Customized PPA for VI (D2-2) with $H=H_1$ given by (D2-8).

The concrete formula of (D2-4) with $H=H_1$ given by (D2-8) is

the underline parts is $F(w^{k+1})$

$$\begin{cases} \theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\{\underline{-A^T\lambda^{k+1}}+\boldsymbol{(\beta A^TA+\delta I_{n_1})}(x^{k+1}-x^k)+\boldsymbol{A^T}(\lambda^{k+1}-\lambda^k)\}\ge 0,\\ \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\{\underline{-B^T\lambda^{k+1}}+\boldsymbol{(\beta B^TB+\delta I_{n_2})}(y^{k+1}-y^k)+\boldsymbol{B^T}(\lambda^{k+1}-\lambda^k)\}\ge 0,\\ \qquad (\underline{Ax^{k+1}+By^{k+1}-b})+\boldsymbol{A}(x^{k+1}-x^k)+\boldsymbol{B}(y^{k+1}-y^k)+\boldsymbol{(2/\beta)}\,(\lambda^{k+1}-\lambda^k)=0.\end{cases}$$

According to Theorem D, to complete PPA iteration, we need only to solve the following sub-problems:

$$\begin{cases} x^{k+1}=\text{argmin}\{\theta_1(x)-x^TA^T\lambda^k+\frac{1}{2}(x-x^k)^T(\beta A^TA+\delta I_{n_1})(x-x^k)|x\in\mathcal{X}\}, & \text{(D2-9a)}\\ y^{k+1}=\text{argmin}\{\theta_2(y)-y^TB^T\lambda^k+\frac{1}{2}(y-y^k)^T(\beta B^TB+\delta I_{n_2})(y-y^k)|y\in\mathcal{Y}\}, & \text{(D2-9b)}\\ \lambda^{k+1}=\lambda^k-\frac{1}{2}\beta\big[2(Ax^{k+1}+By^{k+1}-b)-(Ax^k+By^k-b)\big]. & \text{(D2-9c)}\end{cases}$$

2). Customized PPA for VI (D2-2) with $H = H_2$ given by (D2-8).

The concrete formula of (D2-4) with $H = H_2$ given by (D2-8) is

the underline parts is $F(w^{k+1})$

$$\left\{\begin{array}{r}\theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\{\underline{-A^T\lambda^{k+1}}+\boldsymbol{(\beta A^T A+\delta I_{n_1})}(x^{k+1}-x^k)-\boldsymbol{A^T}(\lambda^{k+1}-\lambda^k)\}\ge 0,\\ \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\{\underline{-B^T\lambda^{k+1}}+\boldsymbol{(\beta B^T B+\delta I_{n_2})}(y^{k+1}-y^k)-\boldsymbol{B^T}(\lambda^{k+1}-\lambda^k)\}\ge 0,\\ (\underline{Ax^{k+1}+By^{k+1}-b})-\boldsymbol{A}(x^{k+1}-x^k)-\boldsymbol{B}(y^{k+1}-y^k)+\boldsymbol{(2/\beta)}\,(\lambda^{k+1}-\lambda^k)=0.\end{array}\right.$$

According to Theorem D, to complete PPA iteration, we need only to solve the following sub-problems:

$$\left\{\begin{array}{ll}\lambda^{k+1}=\lambda^k-\frac{1}{2}\beta(Ax^k+By^k-b), & \text{(D2-10a)}\\ x^{k+1}=\text{argmin}\{\theta_1(x)-x^TA^T(2\lambda^{k+1}-\lambda^k)+\frac{1}{2}(x-x^k)^T(\beta A^TA+\delta I_{n_1})(x-x^k)|x\in\mathcal{X}\}, & \text{(D2-10b)}\\ y^{k+1}=\text{argmin}\{\theta_2(y)-y^TB^T(2\lambda^{k+1}-\lambda^k)+\frac{1}{2}(y-y^k)^T(\beta B^TB+\delta I_{n_2})(y-y^k)|y\in\mathcal{Y}\}. & \text{(D2-10c)}\end{array}\right.$$

4 Balanced PPA for VI (D2-2) (He, Yuan 2012 SIAM Im., Gu, He, Yuan 2014 COA)

Start with given w^k, find a w^{k+1}, such that (D2-4) is satisfied, where

$$u=\begin{pmatrix}x\\y\end{pmatrix},\quad w=\begin{pmatrix}x\\y\\\lambda\end{pmatrix},\quad F(w)=\begin{pmatrix}-A^T\lambda\\-B^T\lambda\\Ax+By-b\end{pmatrix},\qquad H=H_1 \text{ or } H_2, \tag{D2-11}$$

and

$$H_1=\begin{pmatrix}\beta A^TA+\delta I_{n_1} & 0 & A^T\\ 0 & (s+\delta)I_{n_2} & B^T\\ A & B & \frac{1}{\beta}I_m+\frac{1}{s}BB^T\end{pmatrix},\quad H_2=\begin{pmatrix}\beta A^TA+\delta I_{n_1} & 0 & -A^T\\ 0 & (s+\delta)I_{n_2} & -B^T\\ -A & -B & \frac{1}{\beta}I_m+\frac{1}{s}BB^T\end{pmatrix}. \tag{D2-12}$$

1). Balanced PPA for VI (D2-2) with $H = H_1$ given by (D2-12).

The concrete formula of (D2-4) with $H = H_1$ given by (D2-12) is

the underline parts is $F(w^{k+1})$

$$\left\{\begin{array}{r}\theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\{\underline{-A^T\lambda^{k+1}}+\boldsymbol{(\beta A^T A+\delta I_{n_1})}(x^{k+1}-x^k)+\boldsymbol{A^T}(\lambda^{k+1}-\lambda^k)\}\ge 0,\\ \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\{\underline{-B^T\lambda^{k+1}}\;+\;\boldsymbol{(s+\delta)}(y^{k+1}-y^k)\;+\;\boldsymbol{B^T}(\lambda^{k+1}-\lambda^k)\}\ge 0,\\ (\underline{Ax^{k+1}+By^{k+1}-b})+\boldsymbol{A}(x^{k+1}-x^k)+\boldsymbol{B}(y^{k+1}-y^k)+\big(\boldsymbol{\tfrac{1}{\beta}I_m+\tfrac{1}{s}BB^T}\big)\,(\lambda^{k+1}-\lambda^k)=0.\end{array}\right.$$

According to Theorem D, to complete PPA iteration, we need only to solve the following sub-problems:

$$\left\{\begin{array}{ll}x^{k+1}=\text{argmin}\{\theta_1(x)-x^TA^T\lambda^k+\frac{1}{2}(x-x^k)^T(\beta A^TA+\delta I_{n_1})(x-x^k)|x\in\mathcal{X}\}, & \text{(D2-13a)}\\ y^{k+1}=\text{argmin}\{\theta_2(y)-y^TB^T\lambda^k+\frac{1}{2}(s+\delta)\|y-y^k\|^2|y\in\mathcal{Y}\}, & \text{(D2-13b)}\\ \lambda^{k+1}=\lambda^k-\big(\frac{1}{\beta}I_m+\frac{1}{s}BB^T\big)^{-1}\big[2(Ax^{k+1}+By^{k+1}-b)-(Ax^k+By^k-b)\big]. & \text{(D2-13c)}\end{array}\right.$$

2). Balanced PPA for VI (D2-2) with $H = H_2$ given by (D2-12).

The concrete formula of (D2-4) with $H = H_2$ given by (D2-12) is

the underline parts is $F(w^{k+1})$

$$\left\{\begin{array}{r}\theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\{\underline{-A^T\lambda^{k+1}}+\boldsymbol{(\beta A^T A+\delta I_{n_1})}(x^{k+1}-x^k)-\boldsymbol{A^T}(\lambda^{k+1}-\lambda^k)\}\ge 0,\\ \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\{\underline{-B^T\lambda^{k+1}}\;+\;\boldsymbol{(s+\delta)}(y^{k+1}-y^k)\;-\;\boldsymbol{B^T}(\lambda^{k+1}-\lambda^k)\}\ge 0,\\ (\underline{Ax^{k+1}+By^{k+1}-b})-\boldsymbol{A}(x^{k+1}-x^k)-\boldsymbol{B}(y^{k+1}-y^k)+\big(\boldsymbol{\tfrac{1}{\beta}I_m+\tfrac{1}{s}BB^T}\big)\,(\lambda^{k+1}-\lambda^k)=0.\end{array}\right.$$

According to Theorem D, to complete PPA iteration, we need only to solve the following sub-problems:

$$\left\{\begin{array}{ll}\lambda^{k+1}=\lambda^k-\big(\frac{1}{\beta}I_m+\frac{1}{s}BB^T\big)^{-1}(Ax^k+By^k-b), & \text{(D2-14a)}\\ x^{k+1}=\text{argmin}\{\theta_1(x)-x^TA^T(2\lambda^{k+1}-\lambda^k)+\frac{1}{2}(x-x^k)^T(\beta A^TA+\delta I_{n_1})(x-x^k)|x\in\mathcal{X}\}, & \text{(D2-14b)}\\ y^{k+1}=\text{argmin}\{\theta_2(y)-y^TB^T(2\lambda^{k+1}-\lambda^k)+\frac{1}{2}(s+\delta)\|y-y^k\|^2|y\in\mathcal{Y}\}. & \text{(D2-14c)}\end{array}\right.$$

E1-求解两个可分离块凸优化问题的 PPA 类 ADMM 方法

1 两块可分离凸优化问题所对应的变分不等式

在短篇 **C** 中, 我们已经把线性约束两块的可分离凸优化 (不要求 θ_1, θ_2 中任何一个强凸) 问题

$$\min\{\theta_1(x)+\theta_2(y) | Ax+By=b,\ x\in\mathcal{X}, y\in\mathcal{Y}\} \tag{E1-1}$$

的拉格朗日函数 $L(x,y,\lambda)=\theta_1(x)+\theta_2(y)-\lambda^T(Ax+By-b)$ 的鞍点化为变分不等式

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall w\in\Omega, \tag{E1-2}$$

其中
$$w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix},\quad u=\begin{pmatrix} x\\ y\end{pmatrix},\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix}. \tag{E1-3}$$

经典的 ADMM 中, 我们以 $v=(y,\lambda)$ 作为核心变量. 这个短篇介绍的方法, 也以 $v=(y,\lambda)$ 作为核心变量.

2 交换顺序的 ADMM 方法 (Cai, Gu, He and Yuan 2013 Science China Mathematics)

经典的 ADMM 中, 我们以 $v=(y,\lambda)$ 作为核心变量. 如果仍然以 $v=(y,\lambda)$ 作为核心变量, 修正的一种自然的想法是交换经典ADMM中 y 和 λ 更新的顺序. 从给定的 $v^k=(y^k,\lambda^k)$ 出发, 通过

[x-λ-y 顺序的ADMM]
$$\begin{cases} x^{k+1}\in \text{argmin}\{\theta_1(x)-x^TA^T\lambda^k+\frac{1}{2}\beta\|Ax+By^k-b\|^2\,|\,x\in\mathcal{X}\},\\ \lambda^{k+1}=\lambda^k-\beta(Ax^{k+1}+By^k-b),\\ y^{k+1}\in \text{argmin}\{\theta_2(y)-y^TB^T\lambda^{k+1}+\frac{1}{2}\beta\|Ax^{k+1}+By-b\|^2\,|\,y\in\mathcal{Y}\}\end{cases} \tag{E1-4}$$

直接求得 w^{k+1}, 这个方法跟 短篇 **C** 中介绍的经典ADMM收敛表现几乎相当.

根据凸优化的最优性原理 (参见 **D** 中定理), 第 k 迭代预测 (E1-4) 的三个子问题的变分不等式形式是

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

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

$$\begin{cases} \underline{\theta_1(x)-\theta_1(x^{k+1})}+(x-x^{k+1})^T\{\underline{-A^T\lambda^{k+1}}\}\ge 0,\ \forall x\in\mathcal{X}, & \text{(E1-5a)}\\ \underline{\theta_2(y)-\theta_2(y^{k+1})}+(y-y^{k+1})^T\{\underline{-B^T\lambda^{k+1}}+\beta B^TB(y^{k+1}-y^k)-B^T(\lambda^{k+1}-\lambda^k)\}\ge 0,\ \forall y\in\mathcal{Y}, & \text{(E1-5b)}\\ (\lambda-\lambda^{k+1})^T\{\underline{(Ax^{k+1}+By^{k+1}-b)}-B(y^{k+1}-y^k)+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)\}\ge 0,\ \forall\lambda\in\Re^m. & \text{(E1-5c)}\end{cases}$$

对上式整合 (把同一类型的并在一起), 并利用变分不等式 (E1-2) 中的记号, 得到 (E1-5) 的紧凑表示

$$\theta(u)-(u^{k+1})+(w-w^{k+1})^TF(w^{k+1})\ge (v-v^{k+1})^TH_0(v^k-v^{k+1}),\quad \forall w\in\Omega, \tag{E1-6a}$$

其中
$$H_0=\begin{pmatrix} \beta B^TB & -B^T\\ -B & \frac{1}{\beta}I_m\end{pmatrix}. \tag{E1-6b}$$

这里的 H_0 是半正定的, 跟 PPA 一样, (E1-4) 产生的 $\{w^k\}$ 中的核心变量的序列 $\{v^k\}$ 具备性质

$$\|v^{k+1}-v^*\|_{H_0}\le\|v^k-v^*\|_{H_0}^2-\|v^k-v^{k+1}\|_{H_0}^2,\quad \forall v^*\in\mathcal{V}^*. \tag{E1-7}$$

计算实践说明, ADMM方法 (E1-4) 和经典的ADMM方法计算效果完全一样. 若对方法 (E1-4) 产生的 v^{k+1} 做如下的延拓:

$$v^{k+1}:=v^k-\alpha(v^k-v^{k+1}),\quad \alpha=1.5\in(0,2), \tag{E1-8}$$

效率经典的ADMM方法一般有 30% 的提高.

对 $\boldsymbol{x}$-$\boldsymbol{\lambda}$-$\boldsymbol{y}$ 顺序的 ADMM 方法 (E1-4), 我们只列出收敛性的关键式子. 注意到从 (E1-7) 式可以得到

$$\sum_{k=0}^{\infty}\|v^k - v^{k+1}\|_{H_0}^2 < +\infty \qquad \text{和} \qquad \lim_{k\to\infty}\||v^k - v^{k+1}\|_{H_0}^2 = 0. \tag{E1-9}$$

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

$$\|v^k - v^{k+1}\|_{H_0}^2 = \beta\|B(y^k - y^{k+1}) - \tfrac{1}{\beta}(\lambda^k - \lambda^{k+1})\|^2 = \beta\|Ax^{k+1} + By^{k+1} - b\|^2.$$

如果 $\|v^k - v^{k+1}\|_{H_0}^2 = 0$, 从 (E1-5) 可以直接得到 w^{k+1} 就是变分不等式 (E1-2) 的解.

3 略作修正得到 H 正定的 PPA 方法

为了让前一节的方法严格具备收敛性质, 我们将方法 (E1-4) 稍做改造, 并将输出记为预测点 $\tilde{w}^k$.

[预测]
$$\begin{cases} \tilde{x}^k \in \operatorname{argmin}\{\theta_1(x) - x^TA^T\lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 \,|\, x\in\mathcal{X}\}, & \text{(E1-10a)}\\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), & \text{(E1-10b)}\\ \tilde{y}^k \in \operatorname{argmin}\{\theta_2(y) - y^TB^T\tilde{\lambda}^k + \frac{1}{2}\beta\|A\tilde{x}^k + By - b\|^2 + \frac{1}{2}\delta\|y - y^k\|^2 \,|\, y\in\mathcal{Y}\}. & \text{(E1-10c)}\end{cases}$$

跟 (E1-4) 相比, 这里的 y-子问题的目标函数中多了正则项 $\frac{1}{2}\delta\|y - y^k\|^2$.

利用 (E1-10b), (E1-10c) 的等价形式是

$$\begin{aligned}\tilde{y}^k \;&\in\; \arg\min\{\theta_2(y) - y^TB^T\tilde{\lambda}^k + \tfrac{1}{2}\beta\|(A\tilde{x}^k + By^k - b) + B(y - y^k)\|^2 + \tfrac{1}{2}\delta\|y - y^k\|^2 \,|\, y\in\mathcal{Y}\}\\ &=\; \arg\min\{\theta_2(y) - y^TB^T\tilde{\lambda}^k + y^TB^T\beta(A\tilde{x}^k + By^k - b) + \tfrac{1}{2}\|y - y^k\|_{(\delta I + \beta B^TB)}^2 \,|\, y\in\mathcal{Y}\}\\ &=\; \arg\min\{\theta_2(y) - y^TB^T[2\tilde{\lambda}^k - \lambda^k] + \tfrac{1}{2}\|y - y^k\|_{(\delta I + \beta B^TB)}^2 \,|\, y\in\mathcal{Y}\}.\end{aligned}$$

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

$$\begin{cases} \underline{\theta_1(x) - \theta_1(\tilde{x}^k)} + (x - \tilde{x}^k)^T\{-A^T\tilde{\lambda}^k \} \ge 0, \;\forall x\in\mathcal{X}, & \text{(E1-11a)}\\ \underline{\theta_2(y) - \theta_2(\tilde{y}^k)} + (y - \tilde{y}^k)^T\{-B^T\tilde{\lambda}^k + (\delta I + \beta B^TB)(\tilde{y}^k - y^k) - B^T(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \;\forall y\in\mathcal{Y}, & \text{(E1-11b)}\\ (\lambda - \tilde{\lambda}^k)^T\{(A\tilde{x}^k + B\tilde{y}^k - b) - B(\tilde{y}^k - y^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \;\forall\lambda\in\Re^m. & \text{(E1-11c)}\end{cases}$$

对上式整合 (把同一类型的并在一起), 并利用变分不等式 (E1-2) 中的记号, 得到 (E2-10) 的紧凑表示

$$\underline{\theta(u) - \theta(\tilde{u}^k)} + (w - \tilde{w}^k)^TF(\tilde{w}^k) \ge (v - \tilde{v}^k)^TH(v^k - \tilde{v}^k), \quad \forall w\in\Omega, \tag{E1-12a}$$

其中

$$H = \begin{pmatrix} \delta I_{n_2} + \beta B^TB & -B^T\\ -B & \frac{1}{\beta}I_m\end{pmatrix} \succ 0. \tag{E1-12b}$$

在 (E2-11a) 中令 $w = w^*$, 并利用 $\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^TF(\tilde{w}^k) = \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^TF(w^*) \ge 0$, 得到

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

因而

$$(v^k - v^*)^TH(v^k - \tilde{v}^k) \ge \|v^k - \tilde{v}^k\|_H^2,, \quad \forall v^*\in\mathcal{V}^*. \tag{E1-13}$$

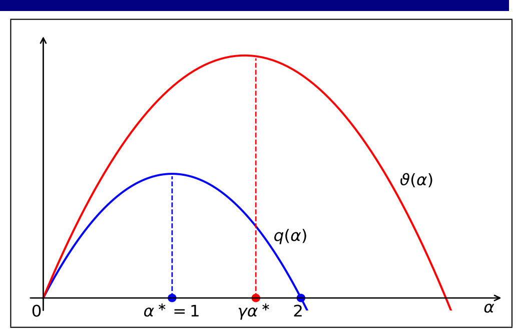
[延拓校正] $\qquad v^{k+1}(\alpha) = v^k - \alpha(v^k - \tilde{v}^k), \quad \alpha = [1.2, 1.8] \subset (0, 2)$ (E1-14)

对任意给定的解点 $v^*\in\mathcal{V}^*$, 我们定义

$$\vartheta_k(\alpha) = \|v^k - v^*\|^2 - \|v^{k+1}(\alpha) - v^*\|^2. \tag{E1-15}$$

为依赖于步长 α 的 k 次迭代的收益. 利用 (E1-13),

$$\begin{aligned}\boldsymbol{\vartheta_k(\alpha)} \;&=\; \|v^k - v^*\|^2 - \|v^{k+1}(\alpha) - v^*\|^2\\ &=\; \|v^k - v^*\|^2 - \|(v^k - v^*) - \alpha(v^k - \tilde{v}^k)\|^2\\ &=\; 2\alpha(v^k - v^*)^T(v^k - \tilde{v}^k) - \alpha^2\|v^k - \tilde{v}^k\|^2\\ &\ge\; \alpha(2 - \alpha)\|v^k - \tilde{v}^k\|^2 =: \boldsymbol{q_k(\alpha)}.\end{aligned}$$



我们无法极大化二次函数 $\vartheta_k(\alpha)$, 因为它含有未知的 v^*, 利用 (E1-13), 得到 $\vartheta_k(\alpha)$ 的在 $\alpha^* = 1$ 取得极大值的下界函数 $q_k(\alpha)$, 如右上图所示. 这是我们在延拓校正 (E1-14) 中一般取 $\alpha = \gamma\alpha^*, \gamma\in[1.2, 1.8]$ 的原因.

E2-求解两个可分离块凸优化问题的对称型 ADMM 方法

1 两块可分离凸优化问题所对应的变分不等式

在短篇 **C** 中, 我们已经把线性约束两块的可分离凸优化 (不要求 θ_1, θ_2 中任何一个强凸) 问题

$$\min\{\theta_1(x)+\theta_2(y) | Ax+By=b,\ x\in\mathcal{X}, y\in\mathcal{Y}\} \tag{E2-1}$$

的拉格朗日函数 $L(x,y,\lambda)=\theta_1(x)+\theta_2(y)-\lambda^T(Ax+By-b)$ 的鞍点化为变分不等式

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega, \tag{E2-2}$$

其中
$$w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix},\quad u=\begin{pmatrix} x\\ y\end{pmatrix},\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix}. \tag{E2-3}$$

2 分裂收缩算法的预测-校正统一框架 (何炳生 2015、2018 运筹学报)

预测-校正的统一框架, 我们 2015 年在"运筹学学报"文章中首次提出, 2018 年综述文章中再次提及.

[预测] 第 k 步迭代从给定的核心变量 v^k 开始, 求得预测点 $\tilde{w}^k$, 使得

$$\tilde{w}^k\in\Omega,\ \ \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(v-\tilde{v}^k)^TQ(v^k-\tilde{v}^k),\ \ \forall\, w\in\Omega, \tag{E2-4}$$

成立. 其中矩阵 Q^T+Q 是正定的. 如果 (E2-4) 中 $v^k=\tilde{v}^k$, $\tilde{w}^k$ 相当于 (E2-2) 的 w^*, 就是问题的解.

在这里处理的两块可分离凸优化问题的变分不等式中, $u=(x,y)$, $w=(x,y,\lambda)$, $v=(y,\lambda)$.

[校正] 根据预测得到的 $\tilde{v}^k$, 给出核心变量 v 的新迭代点 v^{k+1} 的公式为

$$v^{k+1}=v^k-M(v^k-\tilde{v}^k). \tag{E2-5}$$

我们称 (E2-4) 中的 Q 为预测矩阵, (E2-5) 中的 M 为校正矩阵.

收敛性条件 对给定的预测矩阵 Q, 收敛方法要求设计的校正矩阵 M 满足如下条件:

$$\text{存在正定矩阵}\ \ H\succ 0\ \ \text{使得}\ \ HM=Q,\ \ \text{并且}\ \ G:=Q^T+Q-M^THM\succ 0. \tag{E2-6}$$

这节的预测 (E2-4) 和校正 (E2-5) 分别是短篇 **A** 中的 (2.2) 和 (2.8). 收敛性条件 (E2-6) 与短篇 **A** 中的 (2.9) 相同. 在后面的 短篇 **G-H** 中, 有统一框架算法收敛性的更详细证明. 短篇 **E** 和 **F** 只是用算法框架验证收敛性.

3 对称的 ADMM 类方法 (He, Liu, Wang and Yuan, 2014 SIAM Opt.)

修正ADMM的另一种自然的想法是分别求解 x,y 子问题后, 各自(对称地)校正一次 Lagrange 乘子. 换句话说, 从给定的 $\mu>0$ 和 $v^k=(y^k,\lambda^k)$ 出发, 按照下述方式求得 $w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})$:

[对称的ADMM]
$$\left\{\begin{array}{ll}
x^{k+1}=\text{argmin}\{\theta_1(x)-x^TA^T\lambda^k+\frac{1}{2}\beta\|Ax+By^k-b\|^2\,|\,x\in\mathcal{X}\}, & \text{(E2-7a)}\\
\lambda^{k+\frac12}=\lambda^k-\mu\beta(Ax^{k+1}+By^k-b), \qquad \big(\mu\in(0,1)\big) & \text{(E2-7b)}\\
y^{k+1}=\text{argmin}\{\theta_2(y)-y^TB^T\lambda^{k+\frac12}+\frac{1}{2}\beta\|Ax^{k+1}+By-b\|^2\,|\,y\in\mathcal{Y}\}, & \text{(E2-7c)}\\
\lambda^{k+1}=\lambda^{k+\frac12}-\mu\beta(Ax^{k+1}+By^{k+1}-b). & \text{(E2-7d)}
\end{array}\right.$$

把 (E2-7) 分拆成预测与校正, 然后用§2中的框架验证方法的收敛性. 两块可分离的凸优化问题 (E2-1), 每次迭代从核心变量 $v^k=(y^k,\lambda^k)$ 开始, 中间变量 x 的预测需要用

$$\tilde{x}^k\in\text{argmin}\big\{\theta_1(x)-x^TA^T\lambda^k+\tfrac{1}{2}\beta\|Ax+By^k-b\|^2\ |\ x\in\mathcal{X}\big\}$$

生成. 要使 (E2-4) 的右端只有 v 的元素, 就要求上述 x 预测的最优性条件能够写成

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

这种形式, 乘子预测必须用 $\boldsymbol{\tilde{\lambda}^k=\lambda^k-\beta(A\tilde{x}^k+By^k-b)}$. 将 (E2-7) 生成的 w^{k+1}, 用

$$\boldsymbol{\tilde{w}^k=(\tilde{x}^k,\tilde{y}^k,\tilde{\lambda}^k)=(x^{k+1},y^{k+1},\lambda^k-\beta(Ax^{k+1}+By^k-b))}\quad \text{生成预测点, 也就是}$$

[预测]

$$\begin{cases} \tilde{x}^k \in \text{argmin}\big\{\theta_1(x) - x^TA^T\lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\big\}, & \text{(E2-8a)}\\ \tilde{y}^k \in \text{argmin}\big\{\theta_2(y) - y^TB^T\lambda^{k+\frac{1}{2}} + \frac{1}{2}\beta\|A\tilde{x}^k + By - b\|^2 \mid y \in \mathcal{Y}\big\}, & \text{(E2-8b)}\\ \tilde{\lambda}^k = \lambda^k - \beta\big(A\tilde{x}^k + By^k - b\big), & \text{(E2-8c)} \end{cases}$$

利用 $\tilde{x}^k = x^{k+1}$, (E2-7b) 和 (E2-8c), 我们得到

$$\lambda^{k+\frac{1}{2}} = \lambda^k - \mu\beta(A\tilde{x}^k + By^k - b) = \lambda^k - \mu(\lambda^k - \tilde{\lambda}^k) = \tilde{\lambda}^k - (1-\mu)(\tilde{\lambda}^k - \lambda^k). \tag{E2-9}$$

利用 (E2-9) 和 (E2-8c), (E2-8b) 中的 $\tilde{y}^k$ 可以通过

$$\begin{aligned} \tilde{y}^k &\in \arg\min\big\{\theta_2(y) - y^TB^T\lambda^{k+\frac{1}{2}} + \tfrac{1}{2}\beta\|(A\tilde{x}^k + By^k - b) + B(y-y^k)\|^2 \,|\, y \in \mathcal{Y}\big\} \\ &= \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^TB^T[\tilde{\lambda}^k - (1-\mu)(\tilde{\lambda}^k - \lambda^k)] \\ +y^TB^T\beta(A\tilde{x}^k + By^k - b) + \frac{1}{2}\beta\|B(y-y^k)\|^2 \end{array} \middle| y \in \mathcal{Y} \right\} \\ &= \arg\min\big\{\theta_2(y) - y^TB^T[\tilde{\lambda}^k + \mu(\tilde{\lambda}^k - \lambda^k) + \tfrac{1}{2}\beta\|B(y-y^k)\|^2 \,|\, y \in \mathcal{Y}\big\} \qquad \text{得到.} \end{aligned}$$

这样, 由预测 (E2-8) 得到的这些最优性条件就是:

$$\begin{cases} \tilde{x}^k \in \mathcal{X}, \ \underline{\theta_1(x) - \theta_1(\tilde{x}^k)} + (x - \tilde{x}^k)^T\{\underset{\sim}{-A^T\tilde{\lambda}^k} \qquad\qquad\qquad\qquad\qquad\qquad \} \geq 0, \ \forall x \in \mathcal{X}, & \text{(E2-10a)} \\ \tilde{y}^k \in \mathcal{Y}, \ \underline{\theta_2(y) - \theta_2(\tilde{y}^k)} + (y - \tilde{y}^k)^T\{\underset{\sim}{-B^T\tilde{\lambda}^k} + \beta B^TB(\tilde{y}^k - y^k) - \mu B^T(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \ \forall y \in \mathcal{Y}, & \text{(E2-10b)} \\ \tilde{\lambda}^k \in \Re^m, \qquad (\lambda - \tilde{\lambda}^k)^T\{\underset{\sim}{(A\tilde{x}^k + B\tilde{y}^k - b)} - B(\tilde{y}^k - y^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\} \geq 0, \ \forall \lambda \in \Re^m. & \text{(E2-10c)} \end{cases}$$

对上式整合 (把同一类型的并在一起), 并利用变分不等式 (E2-2)-(E2-3) 中的记号, 得到 (E2-10) 的紧凑表示

$$\tilde{w}^k \in \Omega, \quad \underline{\theta(u) - \theta(\tilde{u}^k)} + (w - \tilde{w}^k)^T\underset{\sim}{F(\tilde{w}^k)} \geq (v - \tilde{v}^k)^TQ(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \tag{E2-11a}$$

其中

$$Q = \begin{pmatrix} \beta B^TB & -\mu B^T \\ -B & \frac{1}{\beta}I_m \end{pmatrix}. \tag{E2-11b}$$

由于 (E2-8) 中的 $\tilde{x}$ 和 $\tilde{y}^k$ 等于 (E2-7) 中的 x^{k+1} 和 y^{k+1}. 对 (G2-1d) 中的 λ^{k+1}, 利用 (E2-8c), 我们有

$$\begin{aligned} \lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - \mu\big[-\beta B(y^k - \tilde{y}^k) + \beta(A\tilde{x}^k + By^k - b)\big] \\ &= \lambda^k - \mu\big[-\beta B(y^k - \tilde{y}^k) + 2\beta(A\tilde{x}^k + By^k - b)\big]. \end{aligned} \tag{E2-12}$$

再利用 (E2-8c), 即 $\mu\beta(A\tilde{x}^k + By^k - b = \mu(\lambda^k - \tilde{\lambda}^k)$. 得

$$\lambda^{k+1} = \lambda^k - \big[-\mu\beta B(y^k - \tilde{y}^k) + 2\mu(\lambda^k - \tilde{\lambda}^k)\big].$$

因此, 新的核心变量 $v^{k+1} = (y^{k+1}, \lambda^{k+1})$ 由

[校正]

$$\begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I & 0 \\ -\mu\beta B & 2\mu I \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}, \tag{E2-13}$$

产生. 对于 (E2-13) 中的校正矩阵

$$M = \begin{pmatrix} I & 0 \\ -\mu\beta B & 2\mu I \end{pmatrix}, \qquad \text{令} \qquad H = \begin{pmatrix} (1-\frac{1}{2}\mu)\beta B^TB & -\frac{1}{2}B^T \\ -\frac{1}{2}B & \frac{1}{2\mu\beta}I \end{pmatrix}, \tag{E2-14}$$

可以验证当 $\mu \in (0,1)$ 时, H 正定并有 $HM = Q$ (矩阵 Q 见 (F2-6b)). 此外, 矩阵

$$\begin{aligned} G &= (Q^T + Q) - M^THM = (Q^T + Q) - M^TQ \\ &= \begin{pmatrix} 2\beta B^TB & -(1+\mu)B^T \\ -(1+\mu)B & \frac{2}{\beta}I \end{pmatrix} - \begin{pmatrix} I & -\mu\beta B^T \\ 0 & 2\mu I \end{pmatrix} \begin{pmatrix} \beta B^TB & -\mu B^T \\ -B & \frac{1}{\beta}I_m \end{pmatrix} \\ &= \begin{pmatrix} 2\beta B^TB & -(1+\mu)B^T \\ -(1+\mu)B & \frac{2}{\beta}I \end{pmatrix} - \begin{pmatrix} (1+\mu)\beta B^TB & -2\mu B^T \\ -2\mu B & \frac{2\mu}{\beta}I \end{pmatrix} = (1-\mu)\begin{pmatrix} \beta B^TB & -B^T \\ -B & \frac{2}{\beta}I \end{pmatrix}. \end{aligned}$$

跟经典的ADMM一样, 当 B 列满秩时, H 和 G 都正定, 此时称它们本质上正定. 收敛性条件 (E2-6) 满足.

这说明: 由自然合理的想法建立的方法, 容易验证其符合预测-校正的统一框架及其收敛性条件.

F1-求解三个可分离块凸优化问题的 Gauss 回代型 ADMM 算法

不附加强凸条件的收敛方法　maths.nju.edu.cn/~hebma　何炳生

1　直接推广的 ADMM 求解三块可分离问题不能保证收敛 (Chen, He, Ye, Yuan 2016 MP)

我们把一般的线性约束三块的可分离凸优化 (不要求 $\theta_1, \theta_2, \theta_3$ 中任何一个强凸) 问题

$$\min\{\theta_1(x) + \theta_2(y) + \theta_3(z) | Ax + By + Cz = b,\ x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\} \tag{F1-1}$$

的拉格朗日函数 $L(x, y, z, \lambda) = \theta_1(x) + \theta_2(y) + \theta_3(z) - \lambda^T(Ax + By + Cz - b)$ 的鞍点化为变分不等式

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{F1-2}$$

其中

$$w = \begin{pmatrix} x \\ y \\ z \\ \lambda \end{pmatrix}, \quad u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ -C^T\lambda \\ Ax + By + Cz - b \end{pmatrix}. \tag{F1-3}$$

直接推广的ADMM 处理问题 (F1-1)

[直接推广的ADMM]

$$\left\{\begin{array}{ll} x^{k+1} = \text{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k + Cz^k - b\|^2 \,|\, x \in \mathcal{X}\}, & \text{(F1-4a)} \\ y^{k+1} = \text{argmin}\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\beta\|Ax^{k+1} + By + Cz^k - b\|^2 \,|\, y \in \mathcal{Y}\}, & \text{(F1-4b)} \\ z^{k+1} = \text{argmin}\{\theta_3(z) - z^T C^T \lambda^k + \frac{1}{2}\beta\|Ax^{k+1} + By^{k+1} + Cz - b\|^2 \,|\, z \in \mathcal{Z}\}, & \text{(F1-4c)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^k - b). & \text{(F1-4d)} \end{array}\right.$$

我们在 2016 年的 MP 文章中已经指出, 用直接推广的 ADMM 求解凸优化问题 (F1-1), 当矩阵 A, B, C 中有两个矩阵的列互相正交, 即 $A^TB = 0$, $A^TC = 0$ 或者 $B^TC = 0$ 时方法收敛. 在一般情形下, 并不能保证收敛.

2　可以用来验证算法收敛的统一框架　(何炳生 2015 年运筹学报第 19 卷第 3 期)

预测-校正的统一框架我们 2015 年在“运筹学学报”文章中首次提出, 2018 年综述文章中再次提及.

[预测] 第 k 步迭代从给定的核心变量 v^k 开始, 求得预测点 $\tilde{w}^k$, 使得

$$\tilde{w}^k \in \Omega, \ \ \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \ \forall\, w \in \Omega, \tag{F1-5}$$

成立. 其中矩阵 $Q^T + Q$ 是正定的. 如果 (F1-5) 中 $v^k = \tilde{v}^k$, $\tilde{w}^k$ 相当于 (F1-2) 的 w^*, 就是问题的解.

在我们这里处理的三块的可分离问题中, $u = (x, y, z), w = (x, y, z, \lambda), v = (y, z, \lambda)$.

[校正] 根据预测得到的 $\tilde{v}^k$, 给出核心变量 v 的新迭代点 v^{k+1} 的公式为

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{F1-6}$$

我们称 (F1-5) 中的 Q 为预测矩阵, (F1-12) 中的 M 为校正矩阵.

[收敛性条件] 对给定的预测矩阵 Q, 要求设计的校正矩阵 M 满足如下条件:

$$\text{存在正定矩阵}\ \ H \succ 0\ \ \text{使得}\ \ HM = Q, \ \ \text{并且}\ \ G := Q^T + Q - M^T H M \succ 0. \tag{F1-7}$$

序列 $\{v^k\}$ 具备收缩性质

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{F1-8}$$

如果条件 (F1-7) 满足, 算法就是收敛的, 在后面的 短篇 **G-H** 中, 有统一框架算法收敛性的更详细证明.

3　采用 Gauss 回代的预测-校正方法　(He, Tao and Yuan 2012 SIAM Opt.)

三块可分离的凸优化问题 (F1-1), 每次迭代从核心变量 $v^k = (y^k, z^k, \lambda^k)$ 开始, 中间变量 x 的预测需要用

$$\tilde{x}^k \in \text{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \tfrac{1}{2}\beta\|Ax + By^k + Cz^k - b\|^2 \,|\, x \in \mathcal{X}\}$$

生成. 要使 (F1-5) 的右端只有 v 的元素, 就要求上述 x 预测的最优性条件能够写成

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

这种形式, 乘子预测必须用 $\boldsymbol{\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k + Cz^k - b)}$. 因此, 我们先用

$$
\textbf{[预测]}\quad
\begin{cases}
\tilde{x}^k \in \arg\min\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\}, & \text{(F1-9a)}\\
\tilde{y}^k \in \arg\min\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\beta\|A\tilde{x}^k + By + Cz^k - b\|^2 \mid y \in \mathcal{Y}\}, & \text{(F1-9b)}\\
\tilde{z}^k \in \arg\min\{\theta_3(z) - z^T C^T \lambda^k + \frac{1}{2}\beta\|A\tilde{x}^k + B\tilde{y}^k + Cz - b\|^2 \mid z \in \mathcal{Z}\}, & \text{(F1-9c)}\\
\tilde{\lambda}^k = \lambda^k - \beta\big(A\tilde{x}^k + By^k + Cz^k - b\big). & \text{(F1-9d)}
\end{cases}
$$

得到预测点 $\tilde{w}^k$. 利用优化问题的最优性条件, 我们得到预测的 (F1-9) 的变分不等式形式:

$$
\begin{cases}
\tilde{x}^k \in \mathcal{X},\ \underline{\theta_1(x) - \theta_1(\tilde{x}^k)} + (x - \tilde{x}^k)^T\{-A^T\tilde{\lambda}^k\} \ge 0,\ \forall x \in \mathcal{X}, & \text{(F1-10a)}\\
\tilde{y}^k \in \mathcal{Y},\ \underline{\theta_2(y) - \theta_2(\tilde{y}^k)} + (y - \tilde{y}^k)^T\{-B^T\tilde{\lambda}^k + \beta B^T B(\tilde{y}^k - y^k)\} \ge 0,\ \forall y \in \mathcal{Y}, & \text{(F1-10b)}\\
\tilde{z}^k \in \mathcal{Z},\ \underline{\theta_3(z) - \theta_3(\tilde{z}^k)} + (z - \tilde{z}^k)^T\{-C^T\tilde{\lambda}^k + \beta C^T B(\tilde{y}^k - y^k) + \beta C^T C(\tilde{z}^k - z^k)\} \ge 0,\ \forall z \in \mathcal{Z}, & \text{(F1-10c)}\\
\tilde{\lambda}^k \in \Re^m,\quad (\lambda - \tilde{\lambda}^k)^T\left\{\begin{array}{c}(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b)\\ -B(\tilde{y}^k - y^k) - C(\tilde{z}^k - z^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\end{array}\right\} \ge 0,\ \forall \lambda \in \Re^m & \text{(F1-10d)}
\end{cases}
$$

对上式整合 (把同类型的并在一起), 并利用变分不等式 (F1-2)-(F1-3) 中的记号, 得到 (F1-10) 的紧凑表示

$$
\tilde{w}^k \in \Omega,\quad \underline{\theta(u) - \theta(\tilde{u}^k)} + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k),\quad \forall w \in \Omega, \tag{F1-11a}
$$

其中

$$
Q = \begin{pmatrix} \beta B^T B & 0 & 0\\ \beta C^T B & \beta C^T C & 0\\ -B & -C & \frac{1}{\beta}I_m \end{pmatrix}. \tag{F1-11b}
$$

$Q^T + Q$ 本质上是正定的 (当 B, C 列满秩时是正定的). 利用这样的预测点, 我们采用校正

$$
\textbf{[校正]}\quad v^{k+1} = v^k - M(v^k - \tilde{v}^k),\quad \text{其中}\quad M = \begin{pmatrix} \nu I & -\nu(B^T B)^{-1}B^T C & 0\\ 0 & \nu I & 0\\ -\beta B & -\beta C & I \end{pmatrix}. \tag{F1-12}
$$

生成. 对于矩阵

$$
H = \begin{pmatrix} \frac{1}{\nu}\beta B^T B & \frac{1}{\nu}\beta B^T C & 0\\ \frac{1}{\nu}\beta C^T B & \frac{1}{\nu}\beta[C^T C + C^T B(B^T B)^{-1}B^T C] & 0\\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix} \tag{F1-13}
$$

可以验证 H 正定并有 $HM = Q$. 此外, 当 $\nu \in (0,1)$ 时, 矩阵

$$
G = (Q^T + Q) - M^T Q = \begin{pmatrix} (1-\nu)\beta B^T B & 0 & 0\\ 0 & (1-\nu)\beta C^T C & 0\\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix} \succ 0, \tag{F1-14}
$$

收敛性条件 (F1-7) 满足. 校正 (F1-12) 中的求逆是可以避免的.

事实上, 由于预测 (F1-9) 只需要 (By^k, Cz^k, λ^k) 就可以开始, 校正 (F1-12) 也可以通过

$$
\begin{pmatrix} By^{k+1}\\ Cz^{k+1} \end{pmatrix} = \begin{pmatrix} By^k\\ Cz^k \end{pmatrix} - \begin{pmatrix} \nu I & -\nu I\\ 0 & \nu I \end{pmatrix}\begin{pmatrix} B(y^k - \tilde{y}^k)\\ C(z^k - \tilde{z}^k) \end{pmatrix},\quad \lambda^{k+1} = \lambda^k - \beta\big(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b\big) \tag{F1-15}
$$

实现, 这样就为开始下一次迭代提供了 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$. 我们将理由做一个简要的说明.

将 (F1-11) 中的 v 设为 v^*, 就得到 $(\tilde{v}^k - v^*)^T Q(v^k - \tilde{v}^k) \ge 0$. 再做变换, 设

$$
\xi = \begin{pmatrix} By\\ Cz\\ \lambda \end{pmatrix},\quad \text{得到相应的}\quad (\tilde{\xi}^k - \xi^*)^T \mathcal{Q}(\xi^k - \tilde{\xi}^k) \ge 0,\quad \text{其中}\ \mathcal{Q} = \begin{pmatrix} \beta I_{n_2} & 0 & 0\\ \beta I_{n_2} & \beta I_{n_3} & 0\\ -I_{n_2} & -I_{n_3} & \frac{1}{\beta}I_m \end{pmatrix}. \tag{F1-16}
$$

做了这样的变换以后, 对 (F1-16) 中的 $\mathcal{Q}$, 用

$$
\mathcal{H} = \begin{pmatrix} \frac{1}{\nu}\beta I_{n_2} & \frac{1}{\nu}\beta I_{n_3} & 0\\ \frac{1}{\nu}\beta I_{n_2} & \frac{2}{\nu}\beta I_{n_3} & 0\\ 0 & 0 & \frac{1}{\beta}I_m \end{pmatrix}.\qquad \mathcal{M} = \begin{pmatrix} \beta I_{n_2} & 0 & 0\\ \beta I_{n_2} & \beta I_{n_3} & 0\\ -I_{n_2} & -I_{n_3} & I_m \end{pmatrix}
$$

就有 $\mathcal{H}\mathcal{M} = \mathcal{Q}$ 和 $\mathcal{G} = \mathcal{Q}^T + \mathcal{Q} - \mathcal{M}^T\mathcal{H}\mathcal{M} \succ 0$. 变换下的校正由

$\xi^{k+1} = \xi^k - \mathcal{M}(\xi^k - \tilde{\xi}^k)$ 实现. 写开来的具体形式就是 (F1-15).

如同 (F1-8), 序列 $\{\xi^k\}$ 具备收缩性质 $\quad \|\xi^{k+1} - \xi^*\|_{\mathcal{H}}^2 \le \|\xi^k - \xi^*\|_{\mathcal{H}}^2 - \|\xi^k - \tilde{\xi}^k\|_{\mathcal{G}}^2,\quad \forall \xi^* \in \Xi^*.$ (F1-17)

F2- 求解三个可分离块凸优化问题部分平行正则的 ADMM 算法

不附加强凸条件的收敛方法 maths.nju.edu.cn/~hebma 何炳生

1 直接推广的 ADMM 求解三块可分离问题不能保证收敛 (Chen, He, Ye, Yuan 2016 MP)

我们把一般的线性约束三块的可分离凸优化 (不要求 $\theta_1, \theta_2, \theta_3$ 中任何一个强凸) 问题

$$\min\{\theta_1(x) + \theta_2(y) + \theta_3(z) | Ax + By + Cz = b, \ x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\} \tag{F2-1}$$

的拉格朗日函数 $L(x, y, z, \lambda) = \theta_1(x) + \theta_2(y) + \theta_3(z) - \lambda^T(Ax + By + Cz - b)$ 的鞍点化为变分不等式

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega, \tag{F2-2}$$

其中 $$w = \begin{pmatrix} x \\ y \\ z \\ \lambda \end{pmatrix}, \quad u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ -C^T\lambda \\ Ax + By + Cz - b \end{pmatrix}. \tag{F2-3}$$

我们已经知道, 用直接推广的 ADMM 求解三块可分离凸优化问题 (F2-1), 在一般情形下, 并不能保证收敛.

2 可以用来验证算法收敛的统一框架 (何炳生 2015 年运筹学报第 19 卷第 3 期)

预测-校正的统一框架我们 2015 年在“运筹学学报”文章中首次提出, 2018 年综述文章中再次提及.

[预测] 第 k 步迭代从给定的核心变量 v^k 开始, 求得预测点 $\tilde{w}^k$, 使得

$$\tilde{w}^k \in \Omega, \ \ \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \ \forall\, w \in \Omega, \tag{F2-1}$$

成立. 其中矩阵 $Q^T + Q$ 是正定的. 如果 (F2-3) 中 $v^k = \tilde{v}^k$, $\tilde{w}^k$ 相当于 (F2-2) 的 w^*, 就是问题的解.

[校正] 根据预测得到的 $\tilde{v}^k$, 给出核心变量 v 的新迭代点 v^{k+1} 的公式为

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{F2-2}$$

收敛性条件 对给定的预测矩阵 Q, 要求设计的校正矩阵 M 满足如下条件:

$$\text{存在正定矩阵} \ \ H \succ 0 \ \ \text{使得} \ \ HM = Q, \ \ \text{并且} \ \ G := Q^T + Q - M^T HM \succ 0. \tag{F2-3}$$

序列 $\{v^k\}$ 具备收缩性质 $$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{F2-4}$$

如果条件 (F2-3) 满足, 算法就是收敛的, 在后面的 短篇 **G-H** 中, 有统一框架算法收敛性的更详细证明.

3 部分平行并加正则项的 ADMM 方法 (He, Tao and Yuan 2015 IMA Numr. Analysis)

求解三块可分离凸优化问题 (F2-1) 的部分平行并加正则项的 ADMM 方法中, 仍然以 $v = (y, z, \lambda)$ 作为核心变量. 每步迭代从给定的 $v^k = (y^k, z^k, \lambda^k)$ 开始平行处理 y 和 z 子问题. 通过

$$\left\{\begin{array}{ll}
x^{k+1} \in \text{argmin}\{\theta_1(x) - x^T A^T\lambda^k + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 \,|\, x \in \mathcal{X}\}, & \text{(F2-1a)} \\
y^{k+1} \in \text{argmin}\{\theta_2(y) - y^T B^T\lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By + Cz^k - b\|^2 + \frac{\tau\beta}{2}\|B(y - y^k)\|^2 \,|\, y \in \mathcal{Y}\}, & \text{(F2-1b)} \\
z^{k+1} \in \text{argmin}\{\theta_3(z) - z^T C^T\lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By^k + Cz - b\|^2 + \frac{\tau\beta}{2}\|C(z - z^k)\|^2 \,|\, z \in \mathcal{Z}\}, & \text{(F2-1c)} \\
\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). & \text{(F2-1d)}
\end{array}\right.$$

直接求得 w^{k+1}, 当 $\tau > 1$ 时, 我们将证明这个方法是收敛的. 若记 $\lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b)$

方法中 (F2-1) 的 y^{k+1} 可以通过

$$\begin{aligned}
y^{k+1} &\in \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T\lambda^k + \frac{1}{2}\beta\|(Ax^{k+1} + By^k + Cz^k - b) + B(y - y^k)\|^2 \\ +\frac{\tau}{2}\beta\|B(y - y^k)\|^2 \end{array} \,\Big|\, y \in \mathcal{Y}\right\} \\
&= \arg\min\left\{ \begin{array}{c} \theta_2(y) - y^T B^T\lambda^k + y^T B^T\beta(Ax^{k+1} + By^k + Cz^k - b) \\ +\frac{1}{2}\beta\|B(y - y^k)\|^2 + \frac{\tau}{2}\beta\|B(y - y^k)\|^2 \end{array} \,\Big|\, y \in \mathcal{Y}\right\} \\
&= \arg\min\{\theta_2(y) - y^T B^T\lambda^{k+\frac{1}{2}} + \tfrac{1+\tau}{2}\beta\|B(y - y^k)\|^2 \,|\, y \in \mathcal{Y}\} \qquad \text{得到.}
\end{aligned}$$

同理, (F2-1) 中的 z^{k+1} 可以通过

$$z^{k+1} \in \arg\min\{\theta_3(z) - z^T C^T\lambda^{k+\frac{1}{2}} + \tfrac{1+\tau}{2}\beta\|C(z - z^k)\|^2 \,|\, z \in \mathcal{Z}\} \qquad \text{得到.}$$

记 $\mu = \tau + 1$, 方法 (F2-1) 可以改写成等价的便于编程处理的

$$\begin{cases} x^{k+1} \in \text{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k + Cz^k - b\|^2 \,|\, x \in \mathcal{X}\}, & \text{(F2-2a)} \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b) & \text{(F2-2b)} \\ y^{k+1} \in \text{argmin}\{\theta_2(y) - y^T B^T \lambda^{k+\frac{1}{2}} + \frac{\mu}{2}\beta\|B(y - y^k)\|^2 \,|\, y \in \mathcal{Y}\}, & \text{(F2-2c)} \\ z^{k+1} \in \text{argmin}\{\theta_3(z) - z^T C^T \lambda^{k+\frac{1}{2}} + \frac{\mu}{2}\beta\|C(z - z^k)\|^2 \,|\, z \in \mathcal{Z}\}, & \text{(F2-2d)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). & \text{(F2-2e)} \end{cases}$$

用§2中的统一框架, 收敛性证明才比较简单, 为此我们把方法 (F2-2) 故意拆解成等价的预测-校正方法.

[预测]

$$\begin{cases} \tilde{x}^k \in \arg\min\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\}, & \text{(F2-3a)} \\ \tilde{y}^k \in \arg\min\{\theta_2(y) - y^T B^T \tilde{\lambda}^k + \frac{\mu}{2}\beta\|B(y - y^k)\|^2 | y \in \mathcal{Y}\}, & \text{(F2-3b)} \\ \tilde{z}^k \in \arg\min\{\theta_3(z) - z^T C^T \tilde{\lambda}^k + \frac{\mu}{2}\beta\|C(z - z^k)\|^2 | z \in \mathcal{Z}\}, & \text{(F2-3c)} \\ \tilde{\lambda}^k = \lambda^k - \beta\big(A\tilde{x}^k + By^k + Cz^k - b\big). \qquad (\mu = \tau + 1) & \text{(F2-3d)} \end{cases}$$

方法 (F2-2) 中的 w^{k+1} 和预测 (F2-3) 中 $\tilde{w}^k$ 之间的关系是：

$$x^{k+1} = \tilde{x}^k, \quad y^{k+1} = \tilde{y}^k, \quad z^{k+1} = \tilde{z}^k, \quad \text{和} \quad \lambda^{k+1} = \tilde{\lambda}^k + \beta B(y^k - \tilde{y}^k) + \beta C(z^k - \tilde{z}^k) \tag{F2-4}$$

这样, 由预测 (F2-3) 得到的这些最优性条件就是:

$$\begin{cases} \tilde{x}^k \in \mathcal{X}, & \underline{\theta_1(x) - \theta_1(\tilde{x}^k)} + (x - \tilde{x}^k)^T\{\underset{\sim}{-A^T\tilde{\lambda}^k}\} \geq 0, \ \forall x \in \mathcal{X}, & \text{(F2-5a)} \\ \tilde{y}^k \in \mathcal{Y}, & \underline{\theta_2(y) - \theta_2(\tilde{y}^k)} + (y - \tilde{y}^k)^T\{\underset{\sim}{-B^T\tilde{\lambda}^k} + \mu\beta B^T B(\tilde{y}^k - y^k)\} \geq 0, \ \forall y \in \mathcal{Y}, & \text{(F2-5b)} \\ \tilde{z}^k \in \mathcal{Z}, & \underline{\theta_3(z) - \theta_3(\tilde{z}^k)} + (z - \tilde{z}^k)^T\{\underset{\sim}{-C^T\tilde{\lambda}^k} + \mu\beta C^T C(\tilde{z}^k - z^k)\} \geq 0, \ \forall z \in \mathcal{Z}, & \text{(F2-5c)} \\ \tilde{\lambda}^k \in \Re^m, & (\lambda - \tilde{\lambda}^k)^T \left\{ \begin{array}{c} \underset{\sim}{(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b)} \\ -B(\tilde{y}^k - y^k) - C(\tilde{z}^k - z^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) \end{array} \right\} \geq 0, \ \forall \lambda \in \Re^m. & \text{(F2-5d)} \end{cases}$$

把上式中同一类型的并在一起, 并利用变分不等式 (F2-2) -(F2-3) 中的记号, 得到 (F2-5) 的紧凑表示：

$$\tilde{w}^k \in \Omega, \quad \underline{\theta(u) - \theta(\tilde{u}^k)} + (w - \tilde{w}^k)^T \underset{\sim}{F(\tilde{w}^k)} \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \tag{F2-6a}$$

其中

$$Q = \begin{pmatrix} \mu\beta B^T B & 0 & 0 \\ 0 & \mu\beta C^T C & 0 \\ -B & -C & \frac{1}{\beta}I \end{pmatrix}. \tag{F2-6b}$$

根据 (F2-4), 从预测 (F2-3) 得到的 $\tilde{w}^k$ 再回到 (F2-2) 中核心变量 v^{k+1}, 校正公式是

[校正] $$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \quad \text{其中} \quad M = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\beta B & -\beta C & I \end{pmatrix}. \tag{F2-7}$$

注意到 (F2-6) 中的预测矩阵和 (F2-7) 中的校正矩阵分别是

$$Q = \begin{pmatrix} \mu\beta B^T B & 0 & 0 \\ 0 & \mu\beta C^T C & 0 \\ -B & -C & \frac{1}{\beta}I \end{pmatrix} \quad \text{和} \quad M = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\beta B & -\beta C & I \end{pmatrix}. \tag{F2-8}$$

选取对称矩阵

$$H = \begin{pmatrix} \mu\beta B^T B & 0 & 0 \\ 0 & \mu\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix}, \tag{F2-9}$$

可以验证 H 当 B, C 列满秩时正定并有 $HM = Q$. 此外,

$$G = (Q^T + Q) - M^T Q = \begin{pmatrix} (\mu - 1)\beta B^T B & -\beta B^T C & 0 \\ -\beta C^T B & (\mu - 1)\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix}. \tag{F2-10}$$

由于 $\mu = \tau + 1 > 2$, 矩阵 G 本质上正定. 收敛性条件 (F2-3) 满足. 在 He and Yuan, Optimization Online 6325 中 (此文正式发表在 Springer Proc. Math. Stat., 360, Springer, Singapore, 2021), 我们已经证明:

将方法 (F2-2) 中 $\mu > 2$ 改成 $\mu > 1.5$ (实际计算中取 μ 略大于1.5), 收敛速度有明显提高.

G1-从好不容易凑出一个方法到并不费劲构造一簇算法

古稀之后的新收获–I maths.nju.edu.cn/~hebma 何炳生

1 能够用来验证方法保证收敛的统一框架 (回顾一下退休前总结出的算法统一框架)

我们把线性约束凸优化问题的 Lagrange 函数的鞍点归结为求变分不等式

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega \tag{G1-1}$$

的解. 为求解这个变分不等式, 我们已经设计了下面的预测-校正的统一框架.

[预测] 第 k 步迭代从给定的核心变量 v^k 开始, 求得预测点 $\tilde{w}^k$, 使得

$$\tilde{w}^k \in \Omega, \ \ \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q (v^k - \tilde{v}^k), \ \ \forall\, w \in \Omega, \tag{G1-2}$$

成立. 其中矩阵 $Q^T + Q$ 是正定的. 如果 (G1-2) 中 $v^k = \tilde{v}^k$, $\tilde{w}^k$ 相当于 (G1-1) 的 w^*, 就是问题的解.

v 可以与 w 同, 也可以是 w 的部分分量. 例如, 在交替方向法中, $u = (x, y), w = (x, y, \lambda), v = (y, \lambda)$.

[校正] 根据预测得到的 $\tilde{v}^k$, 给出核心变量 v 的新迭代点 v^{k+1} 的公式为

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{G1-3}$$

我们称 (G1-2) 中的 Q 为预测矩阵, (G1-3) 中的 M 为校正矩阵.

收敛性条件 对给定的预测矩阵 Q, 要求设计的校正矩阵 M 满足如下条件:

$$\text{存在正定矩阵} \ \ H \succ 0 \ \ \text{使得} \ \ HM = Q, \ \ \text{并且} \ \ G := Q^T + Q - M^T H M \succ 0. \tag{G1-4}$$

如果条件 (G1-4) 满足, 算法就是收敛的, 读者将会在第二页的 §3 中看到, 关键收敛性质证明是相当容易的.
以往我们主要是把方法看作(或故意分拆成)预测 (G1-2) 和校正 (G1-3), 然后去验证收敛条件 (G1-4) 能否满足.

事实上, 我们只要求 $Q^T + Q$, H 和 G 本质上是正定的, 这相当于只要求它们半正定, 但是当相应的可分离凸优化问题的线性约束中被分离的分块矩阵都列满秩时就要求它们正定.

就像短篇 **C** 处理两块可分离问题的 ADMM 中, H 正定要求约束矩阵 B 是列满秩, 即 $B^T B$ 是正定的.

2 可以用来构造一簇算法的等价的统一框架 (古稀后发现框架能用来开发算法)

有了预测 (G1-2), 关键是怎样给出满足条件 (G1-4) 的矩阵 M, 才能去实行校正 (G1-3).

下面我们推导出一个跟条件 (G2-4) 等价的构造校正矩阵 M 的方法:

$$\left\{\begin{array}{l} \text{预测 (G1-2) 提供的 } Q \text{ 满足 } Q^T + Q \succ 0. \\ \text{收敛条件 (G1-4) 要求选出的校正矩阵 } M: \\ \textbf{存在 } H \succ 0, \text{ 使得 } \ HM = Q; \\ \textbf{并且 } G = Q^T + Q - M^T H M \succ 0. \end{array}\right. \Longleftrightarrow \left\{\begin{array}{r} D \succ 0, \quad G \succ 0, \\ D + G = Q^T + Q, \\ M^T H M = D, \\ HM = Q. \end{array}\right.$$

$$\Longleftrightarrow \left\{\begin{array}{r} D \succ 0, \quad G \succ 0, \\ D + G = Q^T + Q, \\ Q^T M = D, \\ HM = Q. \end{array}\right. \Longleftrightarrow \left\{\begin{array}{r} D \succ 0, \quad G \succ 0, \\ D + G = Q^T + Q, \\ M = Q^{-T} D, \\ H = Q D^{-1} Q^T. \end{array}\right. \tag{G1-5}$$

我们只要求存在一个对称正定矩阵 H, 保证 $HM = Q$ 和 $G \succ 0$. 上式得到的 H 和 M 确实满足 (G1-4).

有了合格的预测矩阵 Q, 以前是想办法去凑 H 和 M, 使其满足条件 (G1-4), 让人看起来似乎有点神秘.

现在的做法: 一旦有了合格的预测矩阵 Q, 就可以选择多种多样的 D, 使其满足 $0 \prec D \prec Q^T + Q$.

然后由 $M = Q^{-T} D$ 得到矩阵 M 进行 (G1-3) 校正, 收敛性条件自然满足. 而 Q^T 的求逆往往是容易的!

校正 $v^{k+1} = v^k - M(v^k - \tilde{v}^k)$ 可以通过 $Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k)$ 实现, Q 往往是块三角矩阵.

3 统一框架中方法收敛性关键收缩性质的证明

定理 G 采用预测-校正方法 (G1-2)-(G1-3) 求解变分不等式 (G1-1), 如果条件 (G1-4) 满足, 则对生成的序列 $\{v^k\}$ 和 $\{\tilde{v}^k\}$, 有

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge \tfrac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\tfrac{1}{2}\|v^k-\tilde{v}^k\|_G^2,\quad \forall w\in\Omega \qquad \text{(G1-6)}$$

和

$$\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2-\|v^k-\tilde{v}^k\|_G^2,\quad \forall v^*\in\mathcal{V}^*. \qquad \text{(G1-7)}$$

证明 在 (G1-2) 式的右端, 利用 $Q=HM$ (见(G1-4)中的前一式), 便有

$$\theta(u^*)-\theta(\tilde{u}^k)+(w^*-\tilde{w}^k)^TF(\tilde{w}^k)\ge(v^*-\tilde{v}^k)^THM(v^k-\tilde{v}^k).$$

由于 $M(v^k-\tilde{v}^k)=(v^k-v^{k+1})$ (见校正公式 (G1-3)), 上式可以写成

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(v-\tilde{v}^k)^TH(v^k-v^{k+1}). \qquad \text{(G1-8)}$$

将恒等式

$$(a-b)^TH(c-d)=\tfrac{1}{2}\{\|a-d\|_H^2-\|a-c\|_H^2\}+\tfrac{1}{2}\{\|b-c\|_H^2-\|b-d\|_H^2\} \qquad \text{(G1-9)}$$

用于不等式 (G1-8) 的右端, 并设 $a=v,\ b=\tilde{v}^k,\ c=v^k,$ 和 $d=v^{k+1}$, 我们得到

$$(v-\tilde{v}^k)^TH(v^k-v^{k+1})=\tfrac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\tfrac{1}{2}\big(\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2\big). \qquad \text{(G1-10)}$$

对 (G1-10) 右端第二部分, 利用 $HM=Q$ 和 $2v^TQv=v^T(Q^T+Q)v$, 我们有

$$\begin{aligned}
&\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2 && \text{(利用 (G1-3))}\\
&= \|v^k-\tilde{v}^k\|_H^2-\|(v^k-\tilde{v}^k)-M(v^k-\tilde{v}^k)\|_H^2\\
&= (v^k-\tilde{v}^k)^T(HM+M^TH)(v^k-\tilde{v}^k)-(v^k-\tilde{v}^k)^\top M^\top HM(v^k-\tilde{v}^k)\\
&= (v^k-\tilde{v}^k)^T(Q+Q^T-M^THM)(v^k-\tilde{v}^k) && \text{(利用 (G1-4)中的后一式)}\\
&= \|v^k-\tilde{v}^k\|_G^2. && \text{(G1-11)}
\end{aligned}$$

以 (G1-13) 的结果代入 (G1-10), 然后再代入 (G1-8), 就得到结论 (G1-6). 将 (G1-6) 中的 w 设为任意固定的 w^*, 就有

$$\|v^k-v^*\|_H^2-\|v^{k+1}-v^*\|^2\ge\|v^k-\tilde{v}^k\|_H^2+2\{\theta(\tilde{u}^k)-\theta(u^*)+(\tilde{w}^k-w^*)^TF(\tilde{w}^k)\}. \qquad \text{(G1-12)}$$

根据 $(\tilde{w}^k-w^*)^TF(\tilde{w}^k)=(\tilde{w}^k-w^*)^TF(w^*)$ 和最优性条件, 我们有

$$\theta(\tilde{u}^k)-\theta(u^*)+(\tilde{w}^k-w^*)^TF(\tilde{w}^k)=\theta(\tilde{u}^k)-\theta(u^*)+(\tilde{w}^k-w^*)^TF(w^*)\ge 0,$$

因此从 (G1-12) 得到

$$\|v^k-v^*\|_H^2-\|v^{k+1}-v^*\|_H^2\ge\|v^k-\tilde{v}^k\|_G^2. \qquad \text{(G1-13)}$$

这就是定理的结论 (G1-7). □

从结论 (G1-6) 容易推得算法遍历意义下的收敛速率. He & Yuan 关于 ADMM 的遍历意义下收敛速率就是基于不等式(G1-6) 证明的. 见 B. S. He and X. M. Yuan, SIAM J. Numerical Analysis, 2012, 50: 700-709.
我们称 (G1-7) 中的 H 为范数矩阵, G 为效益矩阵. 结论 (G1-7) 是证明方法总体收敛的关键不等式!

从证明过程可以知道, 定理 **G** 的结论当收敛条件 (G1-4) 中 Q^T+Q, H 和 G 弱化成半正定时仍然成立.

4 从验证方法收敛的框架到自由设计算法的纲领

定理G 是预测-校正方法(G1-2)-(G1-3)在条件(G1-4)下证明的. 条件(G1-5)和(G1-4) 等价, 可以根据 (G1-5) 构造算法. 换句话说, §1中的条件 (G1-4) 用来验证方法收敛性, 而§2中的等价性条件为算法设计提供了纲领.

设计算法的纲领. 对预测 (G1-2) 中满足 $Q^T+Q\succ 0$ 的预测矩阵 Q, 选 D, 使得

$$0\prec D\prec Q^T+Q,\qquad \text{然后取}\qquad M=Q^{-T}D\qquad \text{去完成校正 (G1-3).}$$

这样, 预测-校正方法 (G1-2)-(G1-3) 产生的序列 $\{v^k\}$ 满足定理 **G** 中的结论, 其中

$$H=QD^{-1}Q^T\qquad \text{和}\qquad G=Q^T+Q-D,$$

都是正定矩阵. 计算过程中并不要求给出显式的 H.

矩阵 Q 一般是由一个 容易求逆的块三角矩阵 加上 一个 广义秩一(或秩二)矩阵 组成.

直接给出 $M=Q^{-T}D$ 完成校正 (G1-3) 或者求解 $Q^T(v^{k+1}-v^k)=D(\tilde{v}^k-v^k)$ 都并不困难!

G2-根据收敛条件的等价表示构造三可分离块问题的一簇分裂收缩算法

古稀之后的新收获–I maths.nju.edu.cn/~hebma 何炳生

1 能够用来验证方法保证收敛的统一框架 (退休前总结出的算法统一框架)

我们把线性约束凸优化问题的 Lagrange 函数的鞍点归结为求变分不等式

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega \tag{G2-1}$$

的解. 为求解这个变分不等式, 我们已经设计了下面的预测-校正的统一框架.

[预测] 第 k 步迭代从给定的核心变量 v^k 开始, 求得预测点 $\tilde{w}^k$, 使得

$$\tilde{w}^k \in \Omega, \;\; \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \;\; \forall\, w \in \Omega, \tag{G2-2}$$

成立. 其中矩阵 $Q^T + Q$ 本质上是正定的.

[校正] 根据预测得到的 $\tilde{v}^k$, 给出核心变量 v 的新迭代点 v^{k+1} 的公式为

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{G2-3}$$

收敛性条件 对给定的预测矩阵 Q, 要求设计的校正矩阵 M 满足如下条件:

$$\text{存在正定矩阵}\;\; H \succ 0 \;\;\text{使得}\;\; HM = Q, \;\;\text{并且}\;\; G := Q^T + Q - M^T H M \succ 0. \tag{G2-4}$$

换句话说, §1 中的条件 (G2-4) 主要用来验证方法收敛性. §2 中的等价性证明为开发算法设计提供了纲领.

设计算法的纲领. 对预测 (G2-2) 中满足 $Q^T + Q \succ 0$ 的预测矩阵 Q, 选 D, 使得

$$0 \prec D \prec Q^T + Q, \;\; \text{然后取}\;\; M = Q^{-T} D \;\; \text{去完成校正 (G2-3).} \quad \text{求解}\; Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k)$$

矩阵 Q 一般是由一个 容易求逆的块三角矩阵 加上 一个 广义秩一(或秩二)矩阵 组成.

2 三可分块凸优化问题基于直接推广的 ADMM 的预测

三个可分离块凸优化问题及其相应的变分不等式在 **F** 篇章中已经做了介绍. 我们用以下方式生成预测点.

[预测]

$$\begin{cases} \tilde{x}^k \in \arg\min\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\}, & \text{(G2-1a)}\\ \tilde{y}^k \in \arg\min\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\beta\|A\tilde{x}^k + By + Cz^k - b\|^2 \mid y \in \mathcal{Y}\}, & \text{(G2-1b)}\\ \tilde{z}^k \in \arg\min\{\theta_3(z) - z^T C^T \lambda^k + \frac{1}{2}\beta\|A\tilde{x}^k + B\tilde{y}^k + Cz - b\|^2 \mid z \in \mathcal{Z}\}, & \text{(G2-1c)}\\ \tilde{\lambda}^k = \lambda^k - \beta\big(A\tilde{x}^k + By^k + Cz^k - b\big). & \text{(G2-1d)} \end{cases}$$

得到预测点 $\tilde{w}^k$. 利用优化问题的最优性条件, 我们得到预测的 (G2-1) 的变分不等式形式 (G2-2),

[预测矩阵]

$$Q = \begin{pmatrix} \beta B^T B & 0 & 0 \\ \beta C^T B & \beta C^T C & 0 \\ -B & -C & \frac{1}{\beta} I \end{pmatrix}. \tag{G2-2}$$

3 基于确定预测矩阵 Q的不同校正 Notice that

$$Q^T + Q = \begin{pmatrix} 2\beta B^T B & \beta B^T C & -B^T \\ \beta C^T B & 2\beta C^T C & -C^T \\ -B & -C & \frac{2}{\beta} I \end{pmatrix} = \begin{pmatrix} B^T & 0 & 0 \\ 0 & C^T & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} 2\beta I & \beta I & -I \\ \beta I & 2\beta I & -I \\ -I & -I & \frac{2}{\beta} I \end{pmatrix} \begin{pmatrix} B & 0 & 0 \\ 0 & C & 0 \\ 0 & 0 & I \end{pmatrix}$$

is positive definite (whenever B and C are both full column rank) and

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

We call the center part of the right hand sides of the matrices of $Q^T + Q$ and Q^T are kernel, *i. e.,*

$$\begin{pmatrix} 2\beta I & \beta I & -I \\ \beta I & 2\beta I & -I \\ -I & -I & \frac{2}{\beta} I \end{pmatrix} \text{ and } \begin{pmatrix} \beta I & \beta I & -I \\ 0 & \beta I & -I \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix}, \text{ resp.!} \quad \begin{pmatrix} \beta I & \beta I & -I \\ 0 & \beta I & -I \\ 0 & 0 & \frac{1}{\beta} I \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{\beta} I & -\frac{1}{\beta} I & 0 \\ 0 & \frac{1}{\beta} I & I \\ 0 & 0 & \beta I \end{pmatrix}. \tag{G2-3}$$

With the given prediction matrix Q in (G2-2), we can construct various corrections by choosing different matrix D. Because the k-th iteration begins with given (By^k, Cz^k, λ^k), the correction needs only to offer $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$ for starting the next iteration. See some examples !

3.1 构造方法 I By choosing

$$D=\begin{pmatrix} B^T & 0 & 0\\ 0 & C^T & 0\\ 0 & 0 & I\end{pmatrix}\begin{pmatrix}(1+\nu)\beta I & \beta I & -I\\ \beta I & (1+\nu)\beta I & -I\\ -I & -I & \frac{1}{\beta}I\end{pmatrix}\begin{pmatrix} B & 0 & 0\\ 0 & C & 0\\ 0 & 0 & I\end{pmatrix}, \tag{G2-4}$$

where $0<\nu<1$. Since

$$\begin{pmatrix}2\beta I & \beta I & -I\\ \beta I & 2\beta I & -I\\ -I & -I & \frac{2}{\beta}I\end{pmatrix}=\begin{pmatrix}(1+\nu)\beta I & \beta I & -I\\ \beta I & (1+\nu)\beta I & -I\\ -I & -I & \frac{1}{\beta}I\end{pmatrix}+\begin{pmatrix}(1-\nu)\beta I & 0 & 0\\ 0 & (1-\nu)\beta I & 0\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix},$$

the both kernel matrices in the right hand side of the above equation are positive definite. The solution of the system of equations $Q^T(v^{k+1}-v^k)=D(\tilde{v}^k-v^k)$ can be obtained by solving

$$\begin{pmatrix}\beta I & \beta I & -I\\ 0 & \beta I & -I\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix}\begin{pmatrix}By^{k+1}-By^k\\ Cz^{k+1}-Cz^k\\ \lambda^{k+1}-\lambda^k\end{pmatrix}=\begin{pmatrix}(1+\nu)\beta I & \beta I & -I\\ \beta I & (1+\nu)\beta I & -I\\ -I & -I & \frac{1}{\beta}I\end{pmatrix}\begin{pmatrix}B\tilde{y}^k-By^k\\ C\tilde{z}^k-Cz^k\\ \tilde{\lambda}^k-\lambda^k\end{pmatrix}.$$

By using the inverse of the kernel matrix of Q^T in (G2-3), the correction form can be simplified to

$$\begin{pmatrix}By^{k+1}\\ Cz^{k+1}\\ \lambda^{k+1}\end{pmatrix}=\begin{pmatrix}By^k\\ Cz^k\\ \lambda^k\end{pmatrix}-\begin{pmatrix}\nu I & -\nu I & 0\\ 0 & \nu I & 0\\ -\beta I & -\beta I & I\end{pmatrix}\begin{pmatrix}B(y^k-\tilde{y}^k)\\ C(z^k-\tilde{z}^k)\\ \lambda^k-\tilde{\lambda}^k\end{pmatrix}. \tag{G2-5}$$

This is just the correction described in **F1**.

3.2 构造方法 II By choosing

$$D=\begin{pmatrix}\nu\beta B^TB & 0 & 0\\ 0 & \nu\beta C^TC & 0\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix}=\begin{pmatrix} B^T & 0 & 0\\ 0 & C^T & 0\\ 0 & 0 & I\end{pmatrix}\begin{pmatrix}\nu\beta I & 0 & 0\\ 0 & \nu\beta I & 0\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix}\begin{pmatrix} B & 0 & 0\\ 0 & C & 0\\ 0 & 0 & I\end{pmatrix}, \tag{G2-6}$$

where $0<\nu<1$. Since

$$\begin{pmatrix}2\beta I & \beta I & -I\\ \beta I & 2\beta I & -I\\ -I & -I & \frac{2}{\beta}I\end{pmatrix}=\begin{pmatrix}\nu\beta I & 0 & 0\\ 0 & \nu\beta I & 0\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix}+\begin{pmatrix}(2-\nu)\beta I & \beta I & -I\\ \beta I & (2-\nu)\beta I & -I\\ -I & -I & \frac{1}{\beta}I\end{pmatrix},$$

the both kernel matrices in the right hand side of the above equation are positive definite. The solution of the system of equations $Q^T(v^{k+1}-v^k)=D(\tilde{v}^k-v^k)$ can be obtained by solving

$$\begin{pmatrix}\beta I & \beta I & -I\\ 0 & \beta I & -I\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix}\begin{pmatrix}By^{k+1}-By^k\\ Cz^{k+1}-Cz^k\\ \lambda^{k+1}-\lambda^k\end{pmatrix}=\begin{pmatrix}\nu\beta I & 0 & 0\\ 0 & \nu\beta I & 0\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix}\begin{pmatrix}B\tilde{y}^k-By^k\\ C\tilde{z}^k-Cz^k\\ \tilde{\lambda}^k-\lambda^k\end{pmatrix}.$$

By using the inverse of the kernel matrix of Q^T in (G2-3), the correction form can be simplified to

$$\begin{pmatrix}By^{k+1}\\ Cz^{k+1}\\ \lambda^{k+1}\end{pmatrix}=\begin{pmatrix}By^k\\ Cz^k\\ \lambda^k\end{pmatrix}-\begin{pmatrix}\nu I & -\nu I & 0\\ 0 & \nu I & \frac{1}{\beta}I\\ 0 & 0 & I\end{pmatrix}\begin{pmatrix}B(y^k-\tilde{y}^k)\\ C(z^k-\tilde{z}^k)\\ \lambda^k-\tilde{\lambda}^k\end{pmatrix}. \tag{G2-7}$$

3.3 构造方法 III Let us choose the following specific one as D.

$$D=\alpha\big(Q^T+Q\big)=\alpha\begin{pmatrix} B^T & 0 & 0\\ 0 & C^T & 0\\ 0 & 0 & I\end{pmatrix}\begin{pmatrix}2\beta I & \beta I & -I\\ \beta I & 2\beta I & -I\\ -I & -I & \frac{2}{\beta}I\end{pmatrix}\begin{pmatrix} B & 0 & 0\\ 0 & C & 0\\ 0 & 0 & I\end{pmatrix}, \quad \alpha\in(0,1). \tag{G2-8}$$

The solution of the system of equations $Q^T(v^{k+1}-v^k)=D(\tilde{v}^k-v^k)$ can be obtained by solving

$$\begin{pmatrix}\beta I & \beta I & -I\\ 0 & \beta I & -I\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix}\begin{pmatrix}By^{k+1}-By^k\\ Cz^{k+1}-Cz^k\\ \lambda^{k+1}-\lambda^k\end{pmatrix}=\alpha\begin{pmatrix}2\beta I & \beta I & -I\\ \beta I & 2\beta I & -I\\ -I & -I & \frac{2}{\beta}I\end{pmatrix}\begin{pmatrix}B(y^k-\tilde{y}^k)\\ C(z^k-\tilde{z}^k)\\ \lambda^k-\tilde{\lambda}^k\end{pmatrix}.$$

By using the inverse of the core matrix of Q^T in (G2-3), the correction form can be simplified to

$$\begin{pmatrix}By^{k+1}\\ Cz^{k+1}\\ \lambda^{k+1}\end{pmatrix}=\begin{pmatrix}By^k\\ Cz^k\\ \lambda^k\end{pmatrix}-\alpha\begin{pmatrix}I & -I & 0\\ 0 & I & \frac{1}{\beta}I\\ -\beta I & -\beta I & 2I\end{pmatrix}\begin{pmatrix}B(y^k-\tilde{y}^k)\\ C(z^k-\tilde{z}^k)\\ \lambda^k-\tilde{\lambda}^k\end{pmatrix}, \quad \alpha\in(0,1). \tag{G2-9}$$

H1-求解线性约束凸优化问题的预测-校正型广义邻近点算法

古稀之后的新收获–II　maths.nju.edu.cn/~hebma　何炳生

1　线性约束的凸优化问题及其变分不等式的 PPA 算法

我们把线性约束凸优化问题的 Lagrange 函数的鞍点归结为求变分不等式

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega \tag{H1-1}$$

的解. 为求解这个变分不等式, 我们有如下的邻近点算法 (PPA)

邻近点算法(PPA) 第 k-步迭代从给定的核心变量 v^k 开始, 求得新的迭代点 w^{k+1}, 使得

$$w^{k+1} \in \Omega,\ \ \theta(u) - \theta(u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \ge (v - v^{k+1})^T H(v^k - v^{k+1}),\ \forall\, w \in \Omega, \tag{H1-2}$$

成立. 其中 H 是对称正定矩阵. 交替方向法中 $u = (x, y), w = (x, y, \lambda), v = (y, \lambda)$. 有的情形中 $v = w$.

将 (H1-2) 中任意的 $w \in \Omega$ 设为 w^*, 就马上得到 $(v^{k+1} - v^*)^T H(v^k - v^{k+1}) \ge 0$.

设 $a = (v^k - v^*), b = (v^{k+1} - v^*)$, 利用 $\boxed{b^T H(a - b) \ge 0 \ \Rightarrow\ \|b\|_H^2 \le \|a\|_H^2 - \|a - b\|_H^2}$ 有

$$\text{(PPA 收敛性质)} \qquad \|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - v^{k+1}\|_H^2, \quad \forall v^* \in \mathcal{V}^*. \tag{H1-3}$$

2　预测-校正算法的统一框架及其收敛性

除了 PPA, 我们还有求解变分不等式 (H1-1) 的预测-校正的算法统一框架.

[预测] 第 k 步迭代从给定的核心变量 v^k 开始, 求得预测点 $\tilde{w}^k$, 使得

$$\tilde{w}^k \in \Omega,\ \ \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k),\ \ \forall\, w \in \Omega \tag{H1-4}$$

成立. 其中矩阵 $Q^T + Q$ 正定.

[校正] 根据预测得到的 $\tilde{v}^k$, 给出核心变量 v 的新迭代点 v^{k+1} 的公式为

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{H1-5}$$

我们称 (H1-4) 中的 Q 为预测矩阵, (H1-5) 中的 M 为校正矩阵.

收敛性条件 对给定的预测矩阵 Q, 要求设计的校正矩阵 M 满足如下条件:

$$\text{存在正定矩阵}\ \ H \succ 0\ \ \text{使得}\ \ HM = Q,\ \ \text{并且}\ \ G := Q^T + Q - M^T HM \succ 0. \tag{H1-6}$$

在《从好不容易凑出一个算法到并不费劲构造一簇算法》的短篇的定理 G 中, 我们证明了收缩性质

$$\text{(算法收敛性质)} \qquad \|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{H1-7}$$

预测-校正方法 (H1-4)-(H1-5) **和邻近点算法** (H1-2) **之间的关系:** 若 (H1-4) 中的矩阵 Q 对称, 就记其为 H, 并记其中的 $\tilde{w}^k$ 为 w^{k+1}, 便得到 (H1-2). 这也相当于在 (H1-5) 中取 $M = I$, (H1-6) 中的 G 就等于 H. 由于 $\tilde{v}^k = v^{k+1}$, $G = H$, 预测-校正算法的收敛性质 (H1-7) 就成了邻近点算法的收敛性质 (H1-3).

3　统一框架中方法收敛性另一重要性质的证明

满足收敛条件 (H1-6) 的统一框架算法不仅具有收敛性质 (H1-7), 序列 $\{\|v^k - v^{k+1}\|_H\}$ 还是单调不增的.

定理H 求解变分不等式 (H1-1), 设 $\{v^k\}, \{\tilde{w}^k\}$ 是由预测-校正框架 (H1-4)-(H1-5) 生成的. 如果收敛性条件 (H1-6) 满足, 则有

$$\|v^{k+1} - v^{k+2}\|_H^2 \le \|v^k - v^{k+1}\|_H^2 - \|(v^k - \tilde{v}^k) - (v^{k+1} - \tilde{v}^{k+1})\|_G^2. \tag{H1-8}$$

更简单地说, 有

$$\|v^{k+1} - v^{k+2}\|_H^2 \le \|v^k - v^{k+1}\|_H^2. \tag{H1-9}$$

证明 根据 (H1-4), 我们有

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(v-\tilde{v}^k)^TQ(v^k-\tilde{v}^k),\ \forall w\in\Omega \tag{H1-10}$$

和

$$\theta(u)-\theta(\tilde{u}^{k+1})+(w-\tilde{w}^{k+1})^TF(\tilde{w}^{k+1})\ge(v-\tilde{v}^{k+1})^TQ(v^{k+1}-\tilde{v}^{k+1}),\ \forall w\in\Omega. \tag{H1-11}$$

将不等式 (H1-10) 和 (H1-11) 中任意的 $w\in\Omega$ 分别设为 $\tilde{w}^{k+1}$ 和 $\tilde{w}^k$, 得到

$$\theta(\tilde{u}^{k+1})-\theta(\tilde{u}^k)+(\tilde{w}^{k+1}-\tilde{w}^k)^TF(\tilde{w}^k)\ge(\tilde{v}^{k+1}-\tilde{v}^k)^TQ(v^k-\tilde{v}^k)$$

和

$$\theta(\tilde{u}^k)-\theta(\tilde{u}^{k+1})+(\tilde{w}^k-\tilde{w}^{k+1})^TF(\tilde{w}^{k+1})\ge(\tilde{v}^k-\tilde{v}^{k+1})^TQ(v^{k+1}-\tilde{v}^{k+1}).$$

将上述两个不等式相加并利用 $(\tilde{w}^k-\tilde{w}^{k+1})^T\big(F(\tilde{w}^k)-F(\tilde{w}^{k+1})\big)=0$, 就有

$$(\tilde{v}^k-\tilde{v}^{k+1})^TQ\{(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\}\ge 0. \tag{H1-12}$$

在不等式 (H1-12) 两边加上 $\{(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\}^TQ\{(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\}$, 得到

$$(v^k-v^{k+1})^TQ\{(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\}\ge\tfrac{1}{2}\|(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\|^2_{(Q^T+Q)}.$$

利用 $HM=Q$ 和 $M(v^k-\tilde{v}^k)=(v^k-v^{k+1})$ (见(H1-5)), 就有

$$(v^k-v^{k+1})^TH\{(v^k-v^{k+1})-(v^{k+1}-v^{k+2})\}\ge\tfrac{1}{2}\|(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\|^2_{(Q^T+Q)}. \tag{H1-13}$$

最后, 利用恒等式 $\|a\|_H^2-\|b\|_H^2=2a^TH(a-b)-\|a-b\|_H^2$ 和 (H1-13), 我们得到

$$\begin{aligned}&\|v^k-v^{k+1}\|_H^2-\|v^{k+1}-v^{k+2}\|_H^2\\ &\quad= 2(v^k-v^{k+1})^TH\{(v^k-v^{k+1})-(v^{k+1}-v^{k+2})\}-\|(v^k-v^{k+1})-(v^{k+1}-v^{k+2})\|_H^2\\ &\quad\ge \|(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\|^2_{(Q^T+Q)}-\|(v^k-v^{k+1})-(v^{k+1}-v^{k+2})\|_H^2.\end{aligned} \tag{H1-14}$$

对上式右端最后一项利用 $(v^k-v^{k+1})=M(v^k-\tilde{v}^k)$ (见(H1-5)), 有

$$\|(v^k-v^{k+1})-(v^{k+1}-v^{k+2})\|_H^2=\|M(v^k-\tilde{v}^k)-M(v^{k+1}-\tilde{v}^{k+1})\|_H^2, \tag{H1-15}$$

由 (H1-14), (H1-15), 并利用 $Q^T+Q-M^THM=G$, 就得到

$$\|v^k-v^{k+1}\|_H^2-\|v^{k+1}-v^{k+2}\|_H^2\ \ \ge\ \ \|(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\|_G^2.$$

这就是 (H1-8), 证明完毕. □

定理H 是对预测-校正的统一框架的算法证明的. PPA 是统一框架算法的特例, 定理H 的结论同样成立.

4 根据收敛条件的等价性得到的预测-校正的广义邻近点算法

满足收敛条件 (H1-6) 的统一框架算法具有漂亮性质 (H1-7) 和 (H1-9). 如果统一框架中的某个算法能够使得 (H1-7) 右端的 $\|v^k-\tilde{v}^k\|_G^2$ 等于 $\|v^k-v^{k+1}\|_H^2$, 那就得到更漂亮的 PPA 的性质 (H1-3).

这是可以办到的！当预测 (H1-4) 中的预测矩阵 Q 满足 $Q^T+Q\succ 0$, 我们总可以取

$$D\succ 0,\quad G\succ 0\qquad 使得\qquad D+G=Q^T+Q. \tag{H1-16}$$

然后令 $M^THM=D$, 由矩阵方程组解得

$$\left\{\begin{array}{rcl}HM&=&Q,\\ M^THM&=&D.\end{array}\right.\quad\Longleftrightarrow\quad\left\{\begin{array}{rcl}HM&=&Q,\\ Q^TM&=&D.\end{array}\right.\quad\Longleftrightarrow\quad\left\{\begin{array}{rcl}H&=&QD^{-1}Q^T,\\ M&=&Q^{-T}D.\end{array}\right. \tag{H1-17}$$

得到满足收敛条件的 $M=Q^{-T}D$ 和 $H=QD^{-1}Q^T$. 再实行校正 (H1-5), 收敛性质 (H1-7) 仍然成立.

预测-校正的广义 PPA. 在 Q 非对称的预测-校正方法中, 如果在 (H1-16) 中取一对特殊的 D 和 G, 使得

$$D=G=\tfrac{1}{2}(Q^T+Q) \tag{H1-18}$$

由于 $D=G$, 收缩不等式 (H1-7) 就成为

$$\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2-\|v^k-\tilde{v}^k\|_D^2,\quad\forall v^*\in\mathcal{V}^*. \tag{H1-19}$$

因为 $D=M^THM$ (见 (H1-17)), 利用 $M(v^k-\tilde{v}^k)=v^k-v^{k+1}$ (见 (H1-5)), 不等式 (H1-19) 就成了

$$\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2-\|v^k-v^{k+1}\|_H^2,\quad\forall v^*\in\mathcal{V}^*. \tag{H1-20}$$

这是跟 (H1-3) 形式相同的收缩不等式, 因此我们称相应的算法为预测-校正的广义 PPA.

广义 PPA 的校正可以由 $Q^T(v^{k+1}-v^k)=\frac{1}{2}(Q^T+Q)(\tilde{v}^k-v^k)$ 完成. $\{v^k\}$ 具备性质 (H1-9) 和 (H1-20) !

1 The variational inequality and the algorithmic framework

Consider the multi-block convex optimization problem

$$\min\bigl\{\sum_{i=1}^p \theta_i(x_i) \bigm| \sum_{i=1}^p A_i x_i = b,\ x_i \in \mathcal{X}_i\bigr\}. \tag{H2-1}$$

The Lagrangian function of the problem (H2-1) is $L(x_1,\ldots,x_p,\lambda) = \sum_{i=1}^p \theta_i(x_i) - \lambda^T(\sum_{i=1}^p A_i x_i - b)$, which is defined on $\Omega = \prod_{i=1}^p \mathcal{X}_i \times \Re^m$. Note tha the saddle point of the Lagrangian function, say $(x_1^*,\ldots,x_p^*,\lambda^*) \in \Omega$, can be described the solution of the following variational inequality:

$$\text{(VI)} \qquad w^* \in \Omega, \quad \theta(x) - \theta(x^*) + (w-w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{H2-2}$$

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

$$w = \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \lambda \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A_1^T\lambda \\ \vdots \\ -A_p^T\lambda \\ \sum_{i=1}^p A_i x_i - b \end{pmatrix}. \tag{H2-3}$$

We still use Ω^* to represent the solution set of the variational inequality (H2-2). For solving VI (H2-2), we have proposed the following algorithmic unified framework which consists a prediction and a correction.

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

$$\tilde{w}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) + (w-\tilde{w}^k)^T F(\tilde{w}^k) \ge (w-\tilde{w}^k)^T Q(w^k-\tilde{w}^k), \quad \forall w \in \Omega, \tag{H2-4a}$$

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

[Correction]. The new iterate (corrector) w^{k+1} is given by

$$w^{k+1} = w^k - M(w^k - \tilde{w}^k). \tag{H2-4b}$$

where $M = Q^{-T}D$, D is chosen which is satisfied $D \succ 0,\ G \succ 0,\ D+G = Q^T+Q$.

Because $M = Q^{-T}D$, the correction can be achieved by $\quad Q^T(w^{k+1}-w^k) = D(\tilde{w}^k - w^k)$. (H2-5)

2 Prediction B.S.He, S.J.Xu, X.M.Yuan, Handbook of Numerical Analysis, 24 (2023) 511-557.

Start from a given $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$, obtain $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \cdots, \tilde{x}_p^k, \tilde{\lambda}^k)$ via:

$$\left\{\begin{array}{l} \tilde{x}_1^k \in \arg\min\bigl\{\theta_1(x_1) - x_1^TA_1^T\lambda^k + \frac12\beta\|A_1(x_1-x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\bigr\}; \\ \tilde{x}_2^k \in \arg\min\bigl\{\theta_2(x_2) - x_2^TA_2^T\lambda^k + \frac12\beta\|A_1(\tilde{x}_1^k-x_1^k) + A_2(x_2-x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\bigr\}; \\ \qquad\qquad \vdots \\ \tilde{x}_p^k \in \arg\min_{x_p\in\mathcal{X}_p}\bigl\{\theta_p(x_p) - x_p^TA_p^T\lambda^k + \frac12\beta\|\sum_{j=1}^{p-1}A_j(\tilde{x}_j^k-x_j^k) + A_p(x_p-x_p^k)\|^2\bigr\}; \\ \tilde{\lambda}^k = \lambda^k - \beta\bigl(\sum_{j=1}^p A_j\tilde{x}_j^k - b\bigr). \end{array}\right. \tag{H2-6}$$

The optimal condition of the x_i subproblem is

$$\tilde{x}_i^k \in \mathcal{X}_i, \ \ \theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\{-A_i^T\tilde{\lambda}^k + \beta\sum_{j=1}^i A_j(\tilde{x}_j^k - x_j^k) + A_i^T(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \ \ \forall x_i \in \mathcal{X}_i.$$

The formula which yields $\tilde{\lambda}^k$ can be rewritten as

$$\tilde{\lambda}^k \in \Re^m, \quad (\lambda - \tilde{\lambda}^k)^T\{\bigl(\sum_{j=1}^p A_j\tilde{x}_j^k - b\bigr) + \tfrac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\} \ge 0, \quad \forall \lambda \in \Re^m.$$

By using the notations $\theta(x)$ and $F(w)$, we obtain the following variational inequality for the prediction.

Lemma 2.1 *Let $\tilde{w}^k \in \Omega$ be generated by* (H2-6) *with given $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$. Then we have*

$$\tilde{w}^k \in \Omega, \ \ \theta(x) - \theta(\tilde{x}^k) + (w-\tilde{w}^k)^T F(\tilde{w}^k) \ge (w-\tilde{w}^k)^T Q(w^k - \tilde{w}^k), \ \ \forall\, w \in \Omega, \tag{H2-7a}$$

where

$$Q = \begin{pmatrix} \beta A_1^TA_1 & 0 & \cdots & 0 & A_1^T \\ \beta A_2^TA_1 & \beta A_2^TA_2 & \ddots & \vdots & A_2^T \\ \vdots & & \ddots & 0 & \vdots \\ \beta A_p^TA_1 & \beta A_p^TA_2 & \cdots & \beta A_p^TA_p & A_p^T \\ 0 & 0 & \cdots & 0 & \frac{1}{\beta}I_m \end{pmatrix}. \tag{H2-7b}$$

3 Correction (H2-5) by using the kernel matrix arXiv:2107.01897v2[math.OC].

$$Q^T+Q=\begin{pmatrix} A_1^T & 0 & \cdots & 0 & 0\\ 0 & A_2^T & \ddots & \vdots & 0\\ \vdots & & \ddots & 0 & \vdots\\ 0 & 0 & \cdots & A_p^T & 0\\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}\begin{pmatrix} 2\beta I & \beta I & \cdots & \beta I & I\\ \beta I & 2\beta I & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & \beta I & I\\ \beta I & \cdots & \beta I & 2\beta I & I\\ I & \cdots & I & I & \frac{2}{\beta}I \end{pmatrix}\begin{pmatrix} A_1 & 0 & \cdots & 0 & 0\\ 0 & A_2 & \ddots & \vdots & 0\\ \vdots & & \ddots & 0 & \vdots\\ 0 & 0 & \cdots & A_p & 0\\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}.$$

is positive definite (whenever each $A_i, i=1,\ldots,p$, is full column rank matrix) and

$$Q^T=\begin{pmatrix} A_1^T & 0 & \cdots & 0 & 0\\ 0 & A_2^T & \ddots & \vdots & 0\\ \vdots & & \ddots & 0 & \vdots\\ 0 & 0 & \cdots & A_p^T & 0\\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}\begin{pmatrix} \beta I & \beta I & \cdots & \beta I & 0\\ 0 & \beta I & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & \beta I & 0\\ 0 & \cdots & 0 & \beta I & 0\\ I & \cdots & I & I & \frac{1}{\beta}I \end{pmatrix}\begin{pmatrix} A_1 & 0 & \cdots & 0 & 0\\ 0 & A_2 & \ddots & \vdots & 0\\ \vdots & & \ddots & 0 & \vdots\\ 0 & 0 & \cdots & A_p & 0\\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}$$

The center part of the matrices Q^T+Q and Q^T are called their kernel matrix and denoted

$$\mathcal{Q}^T+\mathcal{Q}=\begin{pmatrix} 2\beta I & \beta I & \cdots & \beta I & I\\ \beta I & 2\beta I & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & \beta I & I\\ \beta I & \cdots & \beta I & 2\beta I & I\\ I & \cdots & I & I & \frac{2}{\beta}I \end{pmatrix} \quad\text{and}\quad \mathcal{Q}^T=\begin{pmatrix} \beta I & \beta I & \cdots & \beta I & 0\\ 0 & \beta I & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & \beta I & 0\\ 0 & \cdots & 0 & \beta I & 0\\ I & \cdots & I & I & \frac{1}{\beta}I \end{pmatrix}, \tag{H2-8}$$

respectively. Note that $\mathcal{Q}^{-T}$ has the simple form because $\mathcal{Q}^T$ is a nonsingular upper triangular matrix !

Since the k-th iteration begins with a given $(A_1x_1^k, A_2x_2^k,\ldots,A_px_p^k,\lambda^k)$, for starting the next iteration, the correction this iteration needs only to offer $(A_1x_1^{k+1}, A_2x_2^{k+1},\ldots,A_px_p^{k+1},\lambda^{k+1})$. There are infinite combinations of D and G that meet the conditions, we only take the following example for illustration.

$\boldsymbol{D}$ is proportional to $\boldsymbol{G}$ $\quad D=\alpha\big(Q^T+Q\big)$ and $G=(1-\alpha)\big(Q^T+Q\big)$, $\alpha\in(0,1)$

$$D=\alpha\begin{pmatrix} A_1^T & 0 & \cdots & 0 & 0\\ 0 & A_2^T & \ddots & \vdots & 0\\ \vdots & & \ddots & 0 & \vdots\\ 0 & 0 & \cdots & A_p^T & 0\\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}\begin{pmatrix} 2\beta I & \beta I & \cdots & \beta I & I\\ \beta I & 2\beta I & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & \beta I & I\\ \beta I & \cdots & \beta I & 2\beta I & I\\ I & \cdots & I & I & \frac{2}{\beta}I \end{pmatrix}\begin{pmatrix} A_1 & 0 & \cdots & 0 & 0\\ 0 & A_2 & \ddots & \vdots & 0\\ \vdots & & \ddots & 0 & \vdots\\ 0 & 0 & \cdots & A_p & 0\\ 0 & 0 & \cdots & 0 & I_m \end{pmatrix}. \tag{H2-9}$$

The solution of the system of equations $Q^T(v^{k+1}-v^k)=D(\tilde v^k-v^k)$ can be obtained by solving

$$\begin{pmatrix} \beta I & \beta I & \cdots & \beta I & 0\\ 0 & \beta I & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & \beta I & 0\\ 0 & \cdots & 0 & \beta I & 0\\ I & \cdots & I & I & \frac{1}{\beta}I \end{pmatrix}\begin{pmatrix} A_1x_1^{k+1}-A_1x_1^k\\ A_2x_2^{k+1}-A_2x_2^k\\ \vdots\\ A_px_p^{k+1}-A_px_p^k\\ \lambda^{k+1}-\lambda^k \end{pmatrix}=\alpha\begin{pmatrix} 2\beta I & \beta I & \cdots & \beta I & I\\ \beta I & 2\beta I & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & \beta I & I\\ \beta I & \cdots & \beta I & 2\beta I & I\\ I & \cdots & I & I & \frac{2}{\beta}I \end{pmatrix}\begin{pmatrix} A_1\tilde x_1^k-A_1x_1^k\\ A_2\tilde x_2^k-A_2x_2^k\\ \vdots\\ A_p\tilde x_p^k-A_px_p^k\\ \tilde\lambda^k-\lambda^k \end{pmatrix}.$$

Because
$$\begin{pmatrix} \beta I & \beta I & \cdots & \beta I & 0\\ 0 & \beta I & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & \beta I & 0\\ 0 & \cdots & 0 & \beta I & 0\\ I & \cdots & I & I & \frac{1}{\beta}I \end{pmatrix}^{-1}=\begin{pmatrix} \frac{1}{\beta}I & -\frac{1}{\beta}I & 0 & \cdots & 0\\ 0 & \frac{1}{\beta}I & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & -\frac{1}{\beta}I & 0\\ 0 & \cdots & 0 & \frac{1}{\beta}I & 0\\ -I & 0 & \cdots & 0 & \beta I \end{pmatrix},$$
finally, we get

$$\begin{pmatrix} A_1x_1^{k+1}\\ A_2x_2^{k+1}\\ \vdots\\ A_px_p^{k+1}\\ \lambda^{k+1} \end{pmatrix}=\begin{pmatrix} A_1x_1^k\\ A_2x_2^k\\ \vdots\\ A_px_p^k\\ \lambda^k \end{pmatrix}-\alpha\begin{pmatrix} I & -I & 0 & \cdots & 0\\ 0 & \ddots & \ddots & \ddots & \vdots\\ 0 & 0 & I & -I & 0\\ I & \cdots & I & 2I & \frac{1}{\beta}I\\ -\beta I & 0 & \cdots & 0 & I \end{pmatrix}\begin{pmatrix} A_1x_1^k-A_1\tilde x_1^k\\ A_2x_2^k-A_2\tilde x_2^k\\ \vdots\\ A_px_p^k-A_p\tilde x_p^k\\ \lambda^k-\tilde\lambda^k \end{pmatrix}. \tag{H2-10}$$

References

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

[2] B.S. He and L. Z. Liao, Improvements of some projection methods for monotone nonlinear variational inequalities, JOTA, 2002, 112: 111–128.

[3] 何炳生, 从变分不等式的统一收缩算法到凸优化的分裂收缩算法, 高等学校计算数学学报, 2016, 38: 74–96.

[4] 何炳生, 我和乘子交替方向法20 年. 运筹学学报, 2018, 22: 1–31.

[5] B.S. He, X.M. Yuan and J.J.Z. Zhang, Comparison of two kinds of prediction-correction methods for monotone variational inequalities, COA, 2004, 27: 247-267.

[6] 何炳生. 凸优化和单调变分不等式的收缩算法, 见 maths.nju.edu.cn/˜hebma 中讲义.

[7] 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, COA 51(2012), 649-679.

[8] 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, COA 51(2012), 681-708.

[9] 何炳生, 凸优化和单调变分不等式收缩算法的统一框架. 中国科学: 数学 2018 年第 48 卷第 2 期: 255－272.

[10] 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.

[11] 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).

[12] 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).

[13] B.S. He, L.Z. Liao and M.J. Qian, Alternating projection based prediction-correction method for structured variational inequalities, JCM 24: 693-710 (2006).

[14] B.S. He, X.M. Yuan, On the O(1/n) convergence rate of the Douglas-Rachford Alternating Direction Method, SIAM J. Numerical Analysis, 2012, 50: 700-709.

[15] 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.

[16] G.Y. Gu, B. S. He and J. F. Yang, Inexact alternating-direction-based contraction methods for separable linearly constrained convex optimization. JOTA, 163 (2014), 105 – 129.

[17] Glowinski R. *Numerical methods for nonlinear variational problems*, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.

[18] M.H. Xu, Proximal alternating directions method for structured variational inequalities, JOTA, 134: 107-117 (2007)

[19] B.S. He, F. Ma and X.M. Yuan, Optimally linearizing the alternating direction method of multipliers for convex programming, COA, 75 (2020), 361-388.

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

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

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

[23] 何炳生, 申远, 求解凸规划及鞍点问题定制的PPA 算法及其收敛速率, 中国科学: 数学 2012 年第 42 卷第 5 期: 515–525.

[24] 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.

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

[26] S. J. Xu, A dual-primal balanced augmented Lagrangian method for lineraly constrained convex programming, J. Appl. Math. and Computing 69 (2023), 1015-1035

[27] 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.

[28] B.S. He, H. Liu, Z.R. Wang and X. M. Yuan, A strictly Peaceman-Rachford splitting method for convex programming, SIAM J. Optim. 24 (2014),1011-1040.

[29] 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.

[30] 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.

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

[32] 何炳生, 修正乘子交替方向法求解三个可分离算子的凸优化. 运筹学学报, 2015, 19(3): 57–70.

[33] B.S. He and X.M. Yuan, On the optimal proximal parameter of an ADMM-like splitting method for separable convex programming. Mathematical methods in image processing and inverse problems, 139-163, Springer Proc. Math. Stat., 360. Springer, Singapore, 2021.

[34] B.S. He and X. M. Yuan, A class of ADMM-based algorithms for three-block separable convex programming. Comput. Optim. Appl. 70 (2018), 791-826.

[35] B.S. He, S.J. Xu and X.M. Yuan, Extensions of ADMM for separable convex optimization problems with linear equality or inequality constraints, Handbook of Numerical Analysis, 24 (2023) 511-557. See arXiv: 2107.01897v2 [math.OC].

[36] B.S He and X.M Yuan, On construction of splitting contraction algorithms in a prediction- correction framework for separable convex optimization, arXiv 2204.11522 [math.OC], to be appear in Communication on Optimization Theory.

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

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

[39] 何炳生，变分不等式框架下凸优化的分裂收缩算法. 科学出版社 2024 待出版.