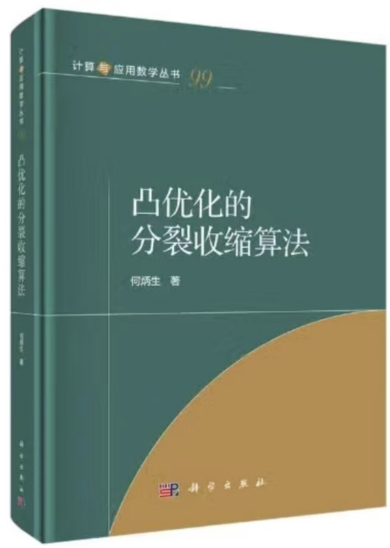
# 从变分不等式的投影收缩算法
# 到 ADMM 为代表的
# 凸优化的分裂收缩算法

用八单元20个两页短篇给读者提供




凸优化分裂收缩算法的辅助读物

南京大学　数学学院

何 炳 生

**Homepage: maths.nju.edu.cn/∼hebma**

# 八单元20个两页短篇 试述一生主要学术历程与收获

## 目 录


序言与导读: **A-H** 八单元20个两页短篇的内容简介 Pref. 1-2

1. **A1**-从变分不等式的投影收缩算法到凸优化的分裂收缩算法概论 1
2. **A2**-变分不等式与优化问题的关系及变分不等式的一些重要性质 3
3. **B1**-单调变分不等式中的三个基本不等式和基本的投影收缩算法 5
4. **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 7
5. **B3**-非线性变分不等式投影收缩算法中的孪生方向和姊妹方法 9
6. **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 11
7. **C2**-预测-校正统一框架下乘子交替方向法 (ADMM) 的收敛性证明 13
8. **C3**-预测-校正统一框架下交替方向法 (ADMM) 收敛速率的关键性质 15
9. **D1**-凸优化问题对应的变分不等式和因需定制的PPA算法 17
10. **D2**-两块可分离凸优化问题平行求解原始子问题的PPA算法 19
11. **E1**-求解两个可分离块凸优化问题的PPA类ADMM方法 21
12. **E2**-求解两个可分离块凸优化问题的对称型ADMM方法 23
13. **F1**-求解三个可分离块凸优化问题的Gauss回代型ADMM算法 25
14. **F2**-求解三个可分离块凸优化问题部分平行正则化的ADMM算法 27
15. **G1**-从好不容易凑出一个方法到并不费劲构造一簇算法 29
16. **G2**-根据收敛条件的等价表示构造求解三可分离块问题的一簇算法 31
17. **G3**-三个可分离块问题分裂收缩算法的预测矩阵设计和方法实现 33
18. **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 35
19. **H2**-求解多块可分离凸优化问题的预测-校正型广义邻近点算法 37
20. **H3**-求解多块可分离凸优化问题预设预测核矩阵的分裂收缩算法 39

参考文献 Ref. 1-2


**说明**. 把单调变分不等式的投影收缩算法安排在 **A B** 两单元, 是因为我们的研究工作是从变分不等式的投影收缩算法开始的。上世纪90年代末我们能较早地开展对 **ADMM** 的研究, 是受前期投影收缩算法研究工作的影响, 这从概论 **A1** 中就可以看出。只是对以ADMM为代表的凸优化的分裂收缩算法感兴趣的读者, 可以浏览下概论 **A1** 的第2页后就直接跳到从 **C** 单元开始读起。

整理这20讲的每篇两页, 我犹如在大教室上课, 考虑黑板怎样布局才算合理。

如果你对我的工作感兴趣, 又只有半天时间, 建议阅读参考文献后的最后两页, 那两页相当于我主页上的《如果只给我两页—那我就写上这些》。

# 序言　A-H 八单元 20 个两页短篇的简介与导读

我心目中的最优化理论与方法, 是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 (Projection and Contraction) 算法和凸优化的 ADMM 类分裂收缩 (Splitting and Contraction) 算法, 组织发表了 (包括一些浅陋之作的) 百余篇文章。这些方法得到了学术界的一些好评, 也碰巧被工程界用来成功地解决了一些他们的问题。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。临近退休我做过总结, 用 PPT 写成二十讲的系列讲义, 挂在个人主页, 但总的篇幅较长, 通读要花不少时间。若精力允许, 我也应该对讲义充实内容, 调整修订。转眼 10 多年过去, 时而还有年轻朋友向我咨询学术问题。对自己一生的学术收获, 我也写了《凸优化的分裂收缩算法》一本专著[45]。关于读书, 华罗庚先生讲过从“越读越厚”到“越读越薄”的道理。随着年龄的增长, 我常问自己, 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中, 比较系统地选一些每个 (包括主要推导过程) 都能用 A4 两页 (犹如在职时写在大黑板上) 写下来, 给人讲清。做成了, 就相当于给那些想了解我工作的后来人, 提供一个跟我的专著配套的一个“薄”的版本。写成几篇以后, 请在读研究生看看, 回馈说看懂基本没有问题。这 **A–H** 八单元的 20 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。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。只是对这 15 年多来热门的 ADMM 类分裂收缩算法感兴趣的学者, 也可以直接从 **C** 单元开始读起。最后的 **G** 和 **H** 两单元, 说明我古稀之后没有马上“洗手”是对的。

参考文献帮助我回顾研究历程。从顺序上说, 读懂这些自洽短篇, 回头再读需要的文献也是可以的。

**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概论**. 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收获, 戏称要为自己写一篇“墓志铭”, 于是就有了这个短篇 **A1** 的最初版本。从单调变分不等式的投影收缩算法开始, 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。后者的研究得益于前者, 是因为 Lagrange 函数的鞍点相当于变分不等式的解点, 论文[1, 2]做了说明。作为概论, 短篇 **A1** 主要是把这两类算法的关系说清楚。如需进一步了解投影收缩算法的成果, 建议阅读[3]和[4]。短篇 **A1** 的第二节已经包含了变分不等式意义下凸优化分裂收缩算法的统一框架, 以及方法收敛关键不等式的证明。这个统一框架, 在 [2, 5, 6] 中都已经提及。短篇 **A2** 讲述了变分不等式包含互补问题, 它与凸优化的关系与区别, 以及变分不等式的一些重要性质。

作为概论的两页 **A1**, 大概只适合曾经关注过我工作的学生和朋友有兴趣时重新拾起。后面 **B** 和 **C** 两单元中的短篇, 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。从单元 **D** 到单元 **H**, 阐述在短篇 **A1** §2 中给出的分裂收缩算法统一框架思想指导下不断取得的进展。

**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, 根据投影得到一对孪生方向建立起的姊妹方法, 曾经被工程界的学者用来解决了一些长期困扰他们的问题。相关的工作还被 M.I. Jordan 教授在他们当年的论文中介绍, 有关定理写进了他们报告的附录. 短篇 **B1** 介绍了投影收缩算法中的三个基本不等式, 投影收缩算法中的寻查方向都源自这些不等式的几种不同组合[1, 3] 。**B2** 和 **B3**, 分别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。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！虽然发现这些关系的过程有点漫长, 但其中主要用到的就是投影的基本性质, 犹如平面几何中的余弦定理。读者如果有进一步了解这方面工作的需求, 建议阅读论文[1, 3, 4, 7], 也可以阅读我z主页上系列讲义[8] 的前五讲。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[9, 10]。将这些想法推广到凸优化问题的求解, 文献[11]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开展 ADMM 的研究 [12], 直到 2006 年, 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[13]、求解实现[14, 15]等方面做了些工作。2009 年, 我们才发现一大批 ADMM 的最新应用, 适逢第一届华东地区运筹学与控制论博士生学术论坛在上海大学召开, 我在大会报告中以“信息技术中的凸优化问题及交替方向法求解”为题呼请年轻人注意 ADMM。2010 年 Boyd 他们关于 ADMM 的综述文章 [16], 让 ADMM 的研究和应用在全世界得到了令人瞩目的重视。受此鼓舞的我们, 凭借已有的基础, 顺势而为, 做了些包括 ADMM 收敛速率[17, 18], 不精确 ADMM [19] 和 (在 **D** 单元开始讲述的) 方法拓展方面的工作。虽然我们最初是在 Gowinski [20] 工作的基础上研究 ADMM, 早期的论文也采用他们的乘子更新步长法则 [12, 21], 为了让初学者更好地理解 ADMM 的收敛性质, **C1** 以变分不等式为框架对工程界普遍采用的 ADMM 版本给出了关键收敛性质的简短而自洽的证明。**C2** 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关键性质证明[17, 18]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[22], 因篇幅关系这里没有介绍, 有兴趣的可以参阅相关文献。此后的所有单元, 讲的都是我们在 ADMM 研究过程中得到的进一步拓展的成果。

**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的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[23], 这个观点被一些美国名校教授们写进了授课讲义。我们称这类方法为因需定制的邻近点算法(Customized PPA) [24, 25]。**D** 中的两个两页纸的短篇, 包含了对一些典型问题如何构造Customized PPA的例子。合理设计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 [5, 26, 27]。设计PPA中的正定矩阵, 使得子问题求解难度平摊的均困邻近点算法(Balanced PPA), 这些工作可以在 [28, 29]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 **C1** 中我们用两页纸给出了ADMM的关键收敛性质的证明。在常用的ADMM的基础上, 我们做一些简单修改(包括迭代中改变变量 $y$ 和 $\lambda$ 更新顺序的PPA型ADMM算法 [30]和两次均等修正乘子的对称ADMM算法 [31])。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，更多的计算实践可参见 [32]。**E** 单元中两个短篇 介绍的算法, 都用在短篇 **A1** 的§2中已经提及的统一框架验证其收敛性。关于统一框架算法的两条简单而又重要的性质, 我们会在单元 **G** 和 **H** 的短篇中给出更系统和详细的证明。

**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 [33]。在这个事实得到明确证实之前, 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 [34, 35], 并且都在统一框架下验证了算法的收敛性。这里将算法及其收敛性证明写成 **F** 单元中的两个短篇。中文读者如需理解这些算法最初的机理, 建议阅读论文 [36]。将 **F1** 中回代方法扩展到多块问题可参阅 [37]。**F2** 中是部分平行加正则化的方法, 有关最优正则化因子的文章 [38], 是处理两个可分离块最优线性化ADMM [22] 同类思想的推广, 个人觉得也很有价值, 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, 都可以等价地转换成一个单调变分不等式的解点。对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 [5, 1, 2], 给出了保证框架算法收敛的条件。这个框架对验证算法收敛性带来极大的方便, 有时也被我们用来根据收敛性条件反向去凑出一些方法 [5, 6, 39]。那些凑成的方法, 往往给人一种神秘的感觉。直到2022年, 我们才对算法框架的收敛性条件给出了一个明确的等价表示, 据此便可以并不费劲地构造一簇算法 [27, 40, 41]。对这类构造性方法, 包括统一的收敛性证明, 写成了短篇 **G1**。短篇 **G2** 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Gauss型预测, 采用不同的校正, 构造一簇算法。短篇 **G3** 走倒过来的路线, 从设定不同的预测矩阵去构造一簇算法。这些成果没有在晚年的我手边溜走, 是非常值得庆幸的!

**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, 单元 **D** 中已经介绍了因需定制的PPA算法。在单元 **G**中, 我们根据预测-校正统一框架收敛性条件的等价表示, 对确定的合格预测, 通过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 **G1** 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[41], 其迭代序列具有短篇 **D** 中经典PPA算法的所有漂亮性质。我们将这个结果及其证明写成短篇 **H1**。短篇 **H2** 以多个可分离块的线性约束凸优化问题为例, 解释如何构造一类预测-校正的广义PPA算法: 它的每步迭代的预测和校正, 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, 相当于高斯消去法完成求解一个线性方程组的消去和回代。跟 **G3** 类似, 短篇 **H3** 讲解如何从设定预测矩阵出发构建求解多个可分离块问题的分裂收缩算法。

结语. 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。几位初步了解我们工作的欧美学者, 称赞我们的方法“Very Simple yet Powerful”和“Elegant” [42, 43], 指的是我们在VI框架下构造了PPA算法。事实上, 对线性约束的凸优化问题, 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求解的(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。不同短篇之间内容出现一些重复, 那是为了每个短篇基本上都能自洽。有朋友说我用的主要数学工具, 就是大学数学中的一些东西, 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, 科学出版社出了一本我的专著[45], 其中的关键思想多半已经在这八单元20个短篇里提及。对希望了解我这一生究竟做了些什么的数学界朋友, 浏览一下我的这些“文玩小把件”, 大体上也就可以！穷其一生, 我主要也就做成了这些东西而已。

何炳生　2025年3月

# A1-从变分不等式的投影收缩算法到凸优化的分裂收缩算法概论

2013 年退休前总结的主要研究工作 (参阅文献[3, 4, 1, 2, 7])

## 1 变分不等式的投影收缩算法 (参阅 [3, 4], 变分不等式的由来建议参阅系列讲义[8]的第一讲)

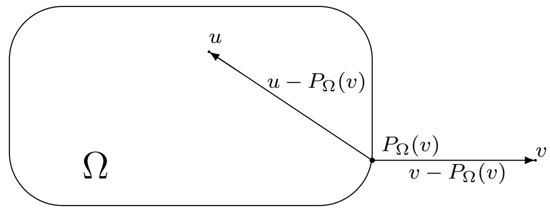
设 $\Omega \subset \Re^n$ 是一个非空闭凸集, $F$ 是 $\Re^n \to \Re^n$ 的一个映射. 考虑求解单调变分不等式

$$\text{(VI)} \qquad u^* \in \Omega, \quad (u-u^*)^T F(u^*) \ge 0, \quad \forall\, u \in \Omega. \tag{A1.1}$$

当 $\Omega = \Re^n$, 问题(A1.1)就是非线性方程组 $F(u) = 0$.

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 (A1.1) 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

$$\tilde{u}^k = P_\Omega[u^k - \beta_k F(u^k)] \tag{A1.2}$$




生成一个预测点 $\tilde{u}^k$, 假如 $\tilde{u}^k = u^k$, $u^k$ 就是 (A1.1) 的解 (参见 **A2** 中的定理 A1 及其证明). 否则, 就要基于以下分析进行新的迭代. 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 $\Omega$ 上的投影, 有

$$(v - P_\Omega(v))^T(u - P_\Omega(v)) \le 0, \quad \forall u \in \Omega. \tag{A1.3}$$

证明可见 **A2** §2 中引理 A. 我们假设用投影 (A1.2) 产生 $\tilde{u}^k$ 时选取的 $\beta_k$ 能使得预测点 $\tilde{u}^k$ 满足

$$\beta_k\|F(u^k) - F(\tilde{u}^k)\| \le \nu\|u^k - \tilde{u}^k\|, \quad \nu \in (0,1). \tag{A1.4}$$

在 (A1.3) 中令 $v = u^k - \beta_k F(u^k)$, 由 (A1.2) 则知 $P_\Omega(v) = \tilde{u}^k$, (A1.3) 告诉我们

$$(u - \tilde{u}^k)^T\{\tilde{u}^k - [u^k - \beta_k F(u^k)]\} \ge 0, \quad \forall u \in \Omega. \tag{A1.5}$$

在上面的 (A1.5) 式两边都加上 $(u - \tilde{u}^k)^T d(u^k, \tilde{u}^k)$, 其中

$$d(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k) - \beta_k[F(u^k) - F(\tilde{u}^k)]. \tag{A1.6}$$

由投影 (A1.2) 我们得到了下面需要的预测公式

$$\text{[预测]} \qquad (u - \tilde{u}^k)^T \beta_k F(\tilde{u}^k) \ge (u - \tilde{u}^k)^T d(u^k, \tilde{u}^k), \ \ \forall u \in \Omega. \tag{A1.7}$$

将上式中的 $u$ 选成 $\Omega$ 中的 $u^*$, 有 $(\tilde{u}^k - u^*)^T d(u^k, \tilde{u}^k) \ge (\tilde{u}^k - u^*)^T \beta_k F(\tilde{u}^k)$. 再由单调性和 VI 定义, 得到

$$(\tilde{u}^k - u^*)^T d(u^k, \tilde{u}^k) \ge 0. \qquad \text{因此} \qquad (u^k - u^*)^T d(u^k, \tilde{u}^k) \ge (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \tag{A1.8}$$

由 $d(u^k, \tilde{u}^k)$ 的表达式和假设 (A1.4), 利用 Cauchy-Schwarz 不等式, 推得

$$(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k)^T\big\{(u^k - \tilde{u}^k) - \beta_k[F(u^k) - F(\tilde{u}^k)]\big\} \ge (1-\nu)\|u^k - \tilde{u}^k\|^2. \tag{A1.9}$$

当 $u^k \ne \tilde{u}^k$ 时, (A1.8) 结合 (A1.9) 说明 $d(u^k, \tilde{u}^k)$ 是距离函数 $\frac{1}{2}\|u - u^*\|^2$ 在 $u^k$ 处的一个上升方向. 我们

可以用 **[单位步长的校正]** $\quad u^{k+1} = u^k - d(u^k, \tilde{u}^k) \quad$ 生成离 $u^*$ 更近的迭代点. (A1.10)

从 (A1.10) 通过简单计算得到

$$\begin{aligned}
\|u^k - u^*\|^2 - \|u^{k+1} - u^*\|^2 &= 2(u^k - u^*)^T d(u^k, \tilde{u}^k) - \|d(u^k, \tilde{u}^k)\|^2 && \text{(利用(A1.8))}\\
&\ge 2(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \|d(u^k, \tilde{u}^k)\|^2 && \text{(利用(A1.6) 和 (A1.4))}\\
&= \|u^k - \tilde{u}^k\|^2 - \beta_k^2\|F(u^k) - F(\tilde{u}^k)\|^2 \ge (1-\nu^2)\|u^k - \tilde{u}^k\|^2.
\end{aligned} \tag{A1.11}$$

也可用 **[计算步长的校正]** $\quad u^{k+1} = u^k - \alpha_k^* d(u^k, \tilde{u}^k) \quad$ 产生离 $u^*$ 更近的迭代点. (A1.12)

其中步长 $\alpha_k^*$ 由 $\alpha_k^* = (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)/\|d(u^k, \tilde{u}^k)\|^2$ 给出. (A1.13)

利用 (A1.12) 和 (A1.13) 通过简单计算即得

$$\begin{aligned}
\|u^k - u^*\|^2 - \|u^{k+1} - u^*\|^2 &= \|u^k - u^*\|^2 - \|(u^k - u^*) - \alpha_k^* d(u^k, \tilde{u}^k)\|^2 && \text{(利用了 (A1.12))}\\
&= 2\alpha_k^*(u^k - u^*)^T d(u^k, \tilde{u}^k) - (\alpha_k^*)^2\|d(u^k, \tilde{u}^k)\|^2 && \text{(利用 (A1.8) 和 (A1.13))}\\
&\ge (\alpha_k^*)^2\|d(u^k, \tilde{u}^k)\|^2 = \|u^k - u^{k+1}\|^2. && \text{(最后的等式再次利用了 (A1.12))}
\end{aligned} \tag{A1.14}$$

收缩不等式 (A1.11) 和 (A1.14) 是证明算法收敛的关键式子. 不等式 (A1.7) 两端的 $\beta_k F(\tilde{u}^k)$ 和 $d(u^k, \tilde{u}^k)$ 称为一对孪生方向. 将在短篇 **B3** 中证明的美妙之处是：采用方向 $\beta_k F(\tilde{u}^k)$ 和 (A1.13) 中给出的步长 $\alpha_k^*$, 用

**[投影校正]** $\quad u^{k+1} = P_\Omega[u^k - \alpha_k^* \beta_k F(\tilde{u}^k)] \quad$ 产生 $u^{k+1}$ 的方法比 (A1.12) 有更好的收敛效果, 详见 **B3**.

✠ 孪生方向, 相同步长. 姊妹方法, 采用投影校正效果更好！ 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 (参阅文献[5, 1, 2, 6])

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, 以三个可分离块的凸优化问题

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z) \mid Ax+By+Cz=b,\ x\in\mathcal{X}, y\in\mathcal{Y}, z\in\mathcal{Z}\} \tag{A1.15}$$

为例做介绍. 问题 (A1.15) 的拉格朗日函数是 $L(x,y,z,\lambda)=\theta_1(x)+\theta_2(y)+\theta_3(z)-\lambda^T(Ax+By+Cz-b)$.

设拉格朗日函数的鞍点是 $(x^*,y^*,z^*,\lambda^*)$ (其中 $(x,y,z)$ 为原始变量, $\lambda$ 为对偶变量), 则其满足

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

把鞍点 $w^*=(x^*,y^*,z^*,\lambda^*)$ 的极大极小性质写出来(具体可参考 **C1**), 它就表述为如下变分不等式的解点:

$$\text{(VI)}\qquad w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega, \tag{A1.16}$$

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

$$u=\begin{pmatrix}x\\y\\z\end{pmatrix},\quad w=\begin{pmatrix}x\\y\\z\\\lambda\end{pmatrix},\quad F(w)=\begin{pmatrix}-A^T\lambda\\-B^T\lambda\\-C^T\lambda\\Ax+By+Cz-b\end{pmatrix}. \tag{A1.17}$$

注意到, (A1.17) 中的 $F(w)$ 恰有 $(w-\tilde{w})^T(F(w)-F(\tilde{w}))\equiv 0$ (也可看作$\ge 0$). 因此, 这个 $F$ 也是单调的.

加之 $\theta(u)$ 是凸函数 (不一定可微), 我们称 (A1.16) 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 (A1.16) 的预测-校正收缩算法统一框架.

[预测] 由给定的 $v^k$ ( $w$ 的核心部分, 在 ADMM 中 $v=(y,\lambda)$, 有些算法中 $v=w$), 求得点 $\tilde{w}^k$, 使得

$$\tilde{w}^k\in\Omega,\ \ \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(v-\tilde{v}^k)^TQ(v^k-\tilde{v}^k),\ \ \forall w\in\Omega, \tag{A1.18}$$

如果矩阵 $Q^T+Q$ 正定 ($Q$ 不一定对称), 我们称 (A1.18) 为合格的预测. 亦见短篇 **G1** 的 (G1.2)

利用 (A1.18) 得到的预测点, 通过以下校正得到新的核心变量. **C2** 中会看到, ADMM 可以解释为这样的预测-校正

可采用 [单位步长的校正] 新迭代点的校正公式为 亦见短篇 **G1** 的 (G1.3)

$$v^{k+1}=v^k-M(v^k-\tilde{v}^k). \tag{A1.19}$$

由合格预测 (A1.18) 和单位步长的校正 (A1.19), 组成一个《预测-校正算法统一框架》, 其收敛性条件是:

[收敛性条件] 存在正定矩阵 $H$, 能够使得 亦见短篇 **G1** 的 (G1.4)

$$HM=Q\qquad \text{并且}\qquad G=Q^T+Q-M^THM\succ 0.\quad (\text{正定}) \tag{A1.20}$$

将 (A1.18) 的 $w$ 选成属于 $\Omega$ 的解点 $w^*$, 利用$(\tilde{w}-w^*)^TF(\tilde{w})=(\tilde{w}-w^*)^TF(w^*)$ 和 $w^*\in\Omega^*$, 得到

$$(\tilde{v}^k-v^*)^TQ(v^k-\tilde{v}^k)\ge 0.\quad \text{进而有}\quad (v^k-v^*)^TQ(v^k-\tilde{v}^k)\ge(v^k-\tilde{v}^k)^TQ(v^k-\tilde{v}^k). \tag{A1.21}$$

$$\begin{aligned}\|v^k-v^*\|_H^2-\|v^{k+1}-v^*\|_H^2 &= \|v^k-v^*\|_H^2-\|(v^k-v^*)-M(v^k-\tilde{v}^k)\|_H^2 \qquad (\text{利用 } HM=Q)\\ &\ge\ (v^k-\tilde{v}^k)^T(Q^T+Q)(v^k-\tilde{v}^k)-\|M(v^k-\tilde{v}^k)\|_H^2 \overset{(A1.20)}{=} \|v^k-\tilde{v}^k\|_G^2.\end{aligned} \tag{A1.22}$$

不等式 (A1.22) 说明若条件 (A1.20) 满足, 统一框架算法生成的序列 $\{v^k\}$ 有收缩性质

$$\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2-\|v^k-\tilde{v}^k\|_G^2.\qquad (\text{这是证明方法收敛性的关键不等式}) \tag{A1.23}$$

结论: 由预测 (A1.18) 和校正 (A1.19) 构成的求解 VI (A1.16) 的算法框架, 条件 (A1.20)满足时方法收敛!

根据 (A1.18), 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 (A1.21) 和 $(v^k-\tilde{v}^k)^TQ(v^k-\tilde{v}^k)>0$, 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 (A1.8), 可以用 (A1.12)-(A1.13) 那样实行计算步长的校正:

也可用 [计算步长的校正] 选定一个正定矩阵 $H$, 然后由以下法则生成新的迭代点:

$$v^{k+1}=v^k-\alpha_k^*M(v^k-\tilde{v}^k),\quad \text{其中}\quad M=H^{-1}Q,\ \boxed{\text{见 (A1.20) 中} HM=Q} \tag{A1.24}$$

$$\text{步长}\ \ \alpha_k^*\quad\text{由}\quad \alpha_k^*=(v^k-\tilde{v}^k)^TQ(v^k-\tilde{v}^k)/\|M(v^k-\tilde{v}^k)\|_H^2\quad\text{给出}. \tag{A1.25}$$

利用上面两个式子, 就可以得到下面的关于算法收敛的关键不等式

$$\begin{aligned}&\|v^k-v^*\|_H^2-\|v^{k+1}-v^*\|_H^2=\|v^k-v^*\|_H^2-\|(v^k-v^*)-\alpha_k^*M(v^k-\tilde{v}^k)\|_H^2\ \ (\text{利用 } HM=Q \text{ 和 (A1.21)})\\ &\ge 2\alpha_k^*(v^k-\tilde{v}^k)^TQ(v^k-\tilde{v}^k)-(\alpha_k^*)^2\|M(v^k-\tilde{v}^k)\|_H^2)\overset{(A1.25)}{=}(\alpha_k^*)^2\|M(v^k-\tilde{v}^k)\|_H^2\overset{(A1.24)}{=}\|v^k-v^{k+1}\|_H^2.\end{aligned}$$

✠ 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, 同一种思维! ✠

# A2-变分不等式与优化问题的关系及变分不等式的一些重要性质

变分不等式 互补问题 及 优化问题的变分不等式表示 何炳生

## 1 变分不等式与可微凸优化及互补问题的关系 (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

$$\mathrm{VI}(\Omega, F) \qquad u^* \in \Omega, \quad (u-u^*)^T F(u^*) \ge 0, \quad \forall\, u \in \Omega. \tag{A2.1}$$

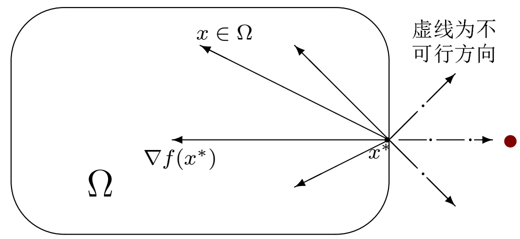
我们说一个变分不等式单调, 是指其中的算子 $F$ 满足

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

考虑可微凸优化问题 $\quad \min\{f(x) \mid x \in \Omega\}$. (A2.2)

$x^*$ 是问题 (A2.2) 的最优解 $\iff$ $x^* \in \Omega$ 并且
从 $x^*$ 出发的任何可行方向都不是下降方向.

记可行方向集: $S_f(x^*) = \{s \in \Re^n \mid s = x - x^*,\, x \in \Omega\}$.
下降方向集: $S_d(x^*) = \{s \in \Re^n \mid s^T \nabla f(x^*) < 0\}$. 因此



Differential Convex Optimization and VI

$$x^* \in \Omega^* \iff \{\, x^* \in \Omega \ \ \& \ \ S_f(x^*) \cap S_d(x^*) = \emptyset. \,\} \iff \{\, x^* \in \Omega,\ (x-x^*)^T \nabla f(x^*) \ge 0,\ \forall x \in \Omega. \,\}$$

上面最右边的就是可微凸优化最优解的变分不等式表示. 可微凸优化有 $(x-y)^T(\nabla f(x) - \nabla f(y)) \ge 0$. 梯度 $\nabla f(x)$ 若可微则其雅可比矩阵一定是对称的, 而一般算子 $F(u)$ 的雅可比矩阵(存在)也不一定对称. 因此, 可微凸优化可以化成相应的单调变分不等式, 但是变分不等式不一定能转化成一个优化问题.

[变分不等式与互补问题的关系]. 非负卦限 $\Re^n_+$ 上变分不等式的形式是

$$\mathrm{VI}(\Re^n_+, F) \quad x \ge 0, \quad (x'-x)^T F(x) \ge 0, \quad \forall x' \ge 0. \tag{A2.3}$$

最优化理论与方法中一类非常重要的问题 – 非线性互补问题的数学形式是

$$(\mathrm{NCP}) \qquad x \ge 0, \quad F(x) \ge 0, \quad x^T F(x) = 0. \tag{A2.4}$$

事实上, 互补问题 NCP 与非负卦限上的变分不等式 $\mathrm{VI}(\Re^n_+, F)$ 是等价的.

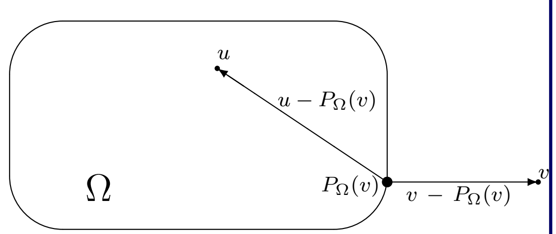
- (A2.3)⇒(A2.4). 如果 $x$ 是 $\mathrm{VI}(\Re^n_+, F)$ (A2.3) 的一个解, 那么 $x \ge 0$. 将 $x' = 2x$ 和 $x' = 0$ 代入 $(x'-x)^T F(x) \ge 0$, 我们得到 $\pm x^T F(x) \ge 0$. 因此 $x^T F(x) = 0$. 只剩下 $F(x) \ge 0$ 需要证明. 对此采用反证法. 若 $F(x)$ 的某个分量 $F_j(x) < 0$, 我们取 $x'$, 使得 $x'_i = \begin{cases} x_i, & \text{if } i \ne j \\ x_j + 1, & \text{if } i = j \end{cases}$ 这样的 $x' \ge 0$. 但 $(x'-x)^T F(x) = F_j(x) < 0$, 这与 $x$ 是 $\mathrm{VI}(\Re^n_+, F)$ 的解矛盾.
- $(A2.4) \Rightarrow (A2.3)$. 如果 $x$ 是 NCP (A2.4) 的解, 那么有 $x \ge 0$ 和 $F(x) \ge 0$. 对于任意的 $x' \ge 0$, 由于 $F(x) \ge 0$, 两个非负向量的内积非负, 有 $(x')^T F(x) \ge 0$. 又因 $x^T F(x) = 0$, 得
  $$(x'-x)^T F(x) = (x')^T F(x) \ge 0. \qquad \text{所以 } x \text{ 是 } \mathrm{VI}(\Re^n_+, F) \text{ (A2.3) 的一个解.}$$

## 2 变分不等式与投影方程 (求解变分不等式等同于求不动点: $u = P_\Omega[u - \beta F(u)]$)

我们需要研究到闭凸集 $\Omega$ 上欧氏模下投影 $\quad \boldsymbol{P_\Omega(v) = \arg\min\{\|v-u\| \mid u \in \Omega\}}\quad$ 的性质.

**引理 A** 设 $\Omega \subset \Re^n$ 为闭凸集, 对任意的 $v \in \Re^n$ 都有

$$(v - P_\Omega(v))^T (u - P_\Omega(v)) \le 0, \ \forall u \in \Omega. \tag{A2.5}$$




**证明**. 首先, 根据 $P_\Omega(v)$ 的定义, 有

$$\|v - P_\Omega(v)\| \le \|v - w\|, \quad \forall w \in \Omega. \tag{A2.6}$$

注意到对任意的 $v \in \Re^n$, 都有 $P_\Omega(v) \in \Omega$.
由于 $\Omega \subset \Re^n$ 是闭凸集, 则对 $u \in \Omega$ 和 $\alpha \in (0,1)$, 都有 $w := P_\Omega(v) + \alpha(u - P_\Omega(v)) \in \Omega$. 代入 (A2.6),

$$\|v - P_\Omega(v)\|^2 \le \|v - P_\Omega(v) - \alpha(u - P_\Omega(v))\|^2.$$

对上式展开, 对任意的 $u \in \Omega$ 和 $\alpha \in (0,1)$, 都有

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

这里的不等式 (A2.5) 就是短篇 **A1** 中的 (A1.3)

令 $\alpha \to 0_+$, 投影性质 (A2.5) 得证.

**定理 A1** 设 $\beta > 0$, $e(u,\beta) = u - P_\Omega[u - \beta F(u)]$, $u^*$ 是 VI$(\Omega, F)$ 的解当且仅当 $e(u^*,\beta) = 0$.

换句话说, 求解变分不等式VI$(\Omega, F)$ 可以归结为求 $e(u,\beta) = u - P_\Omega[u - \beta F(u)$ 的一个零点.

**证明** 先证必要性. 若 $u^*$ 是 VI$(\Omega, F)$ 的解, 则 $u^* \in \Omega$. 由于 $\Omega \subset \Re^n$ 是闭凸集, 利用 (A2.5) 我们得到

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

上式中取 $v := u^* - \beta F(u^*)$, 则有 $\big(e(u^*,\beta) - \beta F(u^*)\big)^T e(u^*,\beta) \le 0$ ,即

$$\|e(u^*,\beta)\|^2 \le \beta e(u^*,\beta)^T F(u^*). \tag{A2.7}$$

由于$P_\Omega[u^* - \beta F(u^*)] \in \Omega$, 而且 $u^*$ 是变分不等式的解, 根据 (A2.1)可以得到

$$\{P_\Omega[u^* - \beta F(u^*)] - u^*\}^T F(u^*) \ge 0, \qquad \text{即} \qquad e(u^*,\beta)^T F(u^*) \le 0. \tag{A2.8}$$

由不等式 (A2.7) 和 (A2.8) 可得 $e(u^*,\beta) = 0$.

再证充分性. 在 (A2.5) 中设 $v = u^* - \beta F(u^*)$, 并利用 $e(u^*,\beta)$ 的表达式, 有

$$\{e(u^*,\beta) - \beta F(u^*)\}^T\{u - P_\Omega[u^* - \beta F(u^*)]\} \le 0, \quad \forall u \in \Omega. \tag{A2.9}$$

根据条件 $e(u^*,\beta) = 0$, 有 $u^* = P_\Omega(\cdot) \in \Omega$ 和 $P_\Omega[u^* - \beta F(u^*)] = u^*$. 代入不等式 (A2.9), 可以得到

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

即 $u^*$ 是VI$(\Omega, F)$ 的解. 定理得证. □

**定理 A2** 对所有的 $u \in \Re^n$ 和 $\tilde\beta \ge \beta > 0$, 我们有 俞志刚在用初等数学证明该定理中起了关键作用 [44]

$$\|e(u,\tilde\beta)\| \ge \|e(u,\beta)\| \qquad \text{和} \qquad \{\|e(u,\tilde\beta)\|/\tilde\beta\} \le \{\|e(u,\beta)/\beta\}. \tag{A2.10}$$

换句话说: $\|e(u,\beta)\|$ 是$\beta$的不减函数, 而 $\{\|e(u,\beta)\|/\beta\}$ 是$\beta$的不增函数. 证明参见 [8] 中第二讲定理 1.2.

## 3 凸优化的变分不等式表示 (分裂收缩算法的数学基础: 拉格朗日函数的鞍点性质 + 定理 A3)

**定理 A3** 设 $\mathcal{X} \subset \Re^n$ 是闭凸集, $\theta(x)$ 和 $f(x)$ 都是凸函数, 其中 $f(x)$ 在包含 $\mathcal{X}$ 的一个开集上可微. 那么

$$x^* \in \operatorname{argmin}\{\theta(x) + f(x)\,|\,x \in \mathcal{X}\} \tag{A2.11}$$

的充分必要条件是

$$x^* \in \mathcal{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^T \nabla f(x^*) \ge 0, \quad \forall x \in \mathcal{X}. \tag{A2.12}$$

**证明** 首先, 如果 (A2.11) 真, 那么对任意的 $x \in \mathcal{X}$ 和 $x_\alpha = (1-\alpha)x^* + \alpha x$, $\forall \alpha \in (0,1]$, 都有

$$\frac{\theta(x_\alpha) - \theta(x^*)}{\alpha} + \frac{f(x_\alpha) - f(x^*)}{\alpha} \ge 0, \tag{A2.13}$$

因为$\theta(x)$ 是 $x$ 的凸函数, 根据 $x_\alpha$ 的定义和凸函数的性质有

$$\theta(x_\alpha) \le (1-\alpha)\theta(x^*) + \alpha\theta(x) \qquad \text{并且因此} \qquad \theta(x) - \theta(x^*) \ge \frac{\theta(x_\alpha) - \theta(x^*)}{\alpha}, \quad \forall \alpha \in (0,1].$$

将此结果代入 (A2.13) 的左边, 我们就有

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

注意到上式中 $f(x_\alpha) = f(x^* + \alpha(x - x^*))$, 由于 $f(x)$ 可微, 令 $\alpha \to 0_+$, 得到

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

这样就从 (A2.11) 得到了凸优化的变分不等式形式 (A2.12). 再考虑从 (A2.12) 到 (A2.11). $f$ 是凸函数, 就有

$$f(x_\alpha) \le (1-\alpha)f(x^*) + \alpha f(x), \qquad \text{这可以写成} \qquad f(x_\alpha) - f(x^*) \le \alpha(f(x) - f(x^*)).$$

因此, 对所有的 $\alpha \in (0,1]$, 我们有

$$f(x) - f(x^*) \ge \frac{f(x_\alpha) - f(x^*)}{\alpha} \quad \equiv \quad \frac{f(x^* + \alpha(x - x^*)) - f(x^*)}{\alpha}.$$

在上式中取 $\alpha \to 0_+$, 利用 $f(x)$ 的可微性, 有

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

将此代入 (A2.12) 的左边, 得 下面的式子说明 $x^*$ 是 (A2.11) 的解

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

这就从 (A2.12) 得到 (A2.11). 完成了全部证明. □ 分裂收缩算法都根据定理 **A3** 中的对应关系设计优化子问题

# B1-单调变分不等式中的三个基本不等式和基本的投影收缩算法

基本不等式的组合提供寻查方向 maths.nju.edu.cn/~hebma 何炳生

## 1 由投影得到的三个基本不等式 (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

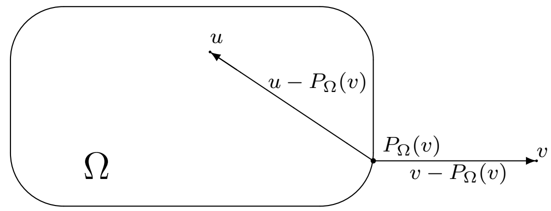
$$\text{(VI)} \qquad u^* \in \Omega, \quad (u-u^*)^T F(u^*) \geq 0, \quad \forall\, u \in \Omega. \tag{B1.1}$$

我们说一个变分不等式单调, 是指其中的算子 $F$ 满足

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

在求解变分不等式 (B1.1) 的投影收缩算法中, 对给定的当前点 $u^k$ 和常数 $\beta > 0$, 我们利用投影

$$\tilde{u}^k = P_\Omega[u^k - \beta F(u^k)] \tag{B1.2}$$




生成一个预测点 $\tilde{u}^k$. 在短篇 **A2** 中已经证明: 假如 $\tilde{u}^k = u^k$, $u^k$ 就是 (B1.1) 的解.

此外, 短篇 **A2** 中同样已经证明了凸集上投影有如右上图所示的重要性质: 对任意的 $v \in \Re^n$, 有

$$(P_\Omega(v) - u)^T(v - P_\Omega(v)) \geq 0, \quad \forall u \in \Omega. \tag{B1.3}$$

由于 $\tilde{u}^k \in \Omega$, 根据变分不等式 (B1.1) 的定义, 对任意的 $\beta > 0$, 都有

$$\text{(FI-1)} \qquad (\tilde{u}^k - u^*)^T \beta F(u^*) \geq 0, \quad \forall u^* \in \Omega^*. \tag{B1.4}$$

在 (B1.3) 中令 $v = u^k - \beta_k F(u^k)$, 那么由 (B1.2) 得到 $\tilde{u}^k = P_\Omega(v)$, 将那个属于 $\Omega$ 的 $u$ 取成 $u^*$, 得到

$$\text{(FI-2)} \qquad (\tilde{u}^k - u^*)^T \big(u^k - \beta F(u^k) - \tilde{u}^k\big) \geq 0, \quad \forall u^* \in \Omega^*. \tag{B1.5}$$

根据单调性, 我们有

$$\text{(FI-3)} \qquad (\tilde{u}^k - u^*)^T \beta\big(F(\tilde{u}^k) - F(u^*)\big) \geq 0, \quad \forall u^* \in \Omega^*. \tag{B1.6}$$

我们称它们为投影收缩算法中的三个基本不等式.

## 2 由 (FI-1)+(FI-2) 生成求解线性变分不等式的寻查方向

线性变分不等式中, $F(u) = Mu + q$, 其中 $M$ 为方阵, $q$ 为相应的向量. (FI-1) 和 (FI-2) 就分别成了

$$(\tilde{u}^k - u^*)^T \beta(Mu^* + q) \geq 0, \quad \forall u^* \in \Omega^* \tag{B1.7}$$

和

$$(\tilde{u}^k - u^*)^T \big\{(u^k - \tilde{u}^k) - \beta(Mu^k + q)\big\} \geq 0, \quad \forall u^* \in \Omega^*. \tag{B1.8}$$

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

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

我们考虑的是单调变分不等式, 利用算子的单调性, 有 $(u^k - u^*)^T M(u^k - u^*) \geq 0$, 从上式得到

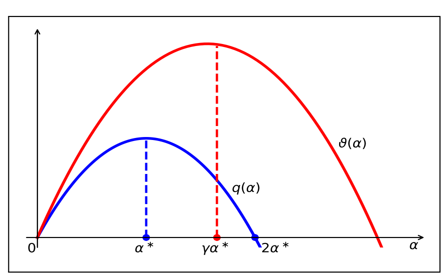
$$(u^k - u^*)^T\{(I + \beta M^T)(u^k - \tilde{u}^k)\} \geq \|u^k - \tilde{u}^k\|^2, \quad \forall u^* \in \Omega^*. \tag{B1.9}$$

不等式 (B1.9) 表明由 (FI-1) + (FI-2) 导出的 $(I + \beta M^T)(u^k - \tilde{u}^k)$ 是距离函数 $\|u - u^*\|^2$ 在 $u^k$ 处的上升方向. 我们可以选择适当的 $\alpha > 0$, 通过下式生成欧氏模下比 $u^k$ 更靠近解集的, 依赖于步长 $\alpha$ 的新的迭代点

$$u^{k+1}(\alpha) = u^k - \alpha(I + \beta M^T)(u^k - \tilde{u}^k). \tag{B1.10}$$

如何选取步长 $\alpha$? 合理的想法是通过选择 $\alpha$ 最大化迭代收益 $\vartheta_k(\alpha) = \|u^k - u^*\|^2 - \|u^{k+1}(\alpha) - u^*\|^2$. 利用 (B1.10) 计算

$$\begin{aligned}
\vartheta_k(\alpha) &= \|u^k - u^*\|^2 - \|u^k - u^* - \alpha(I + \beta M^T)(u^k - \tilde{u}^k)\|^2 \\
&= 2\alpha(u^k - u^*)^T(I + \beta M^T)(u^k - \tilde{u}^k) \\
&\qquad -\alpha^2\|(I + \beta M^T)(u^k - \tilde{u}^k)\|^2 \quad (\text{利用(B1.9)}) \\
&\geq 2\alpha\|u^k - \tilde{u}^k\|^2 - \alpha^2\|(I + \beta M^T)(u^k - \tilde{u}^k)\|^2 \\
&= q_k(\alpha).
\end{aligned} \tag{B1.11}$$




$\vartheta_k(\alpha)$ 和 $q_k(\alpha)$ 都是 $\alpha$ 的二次函数, 但前者含有未知的 $u^*$. 后者是前者的下界函数, 计算 $q_k(\alpha)$ 取得最大值的 $\alpha_k^* = \|u^k - \tilde{u}^k\|^2 \|(I + \beta M^T)(u^k - \tilde{u}^k)\|^2$, 然后用

上图说明 (B1.12) 中为什么对 $\alpha^*$ 要乘适当的 $\gamma > 1$.

$$u^{k+1} = u^k - \gamma\alpha_k^*(I + \beta M^T)(u^k - \tilde{u}^k), \quad \gamma \in [1.2, 1.8] \subset (0, 2) \qquad \text{生成新的迭代点.} \tag{B1.12}$$

如果考虑只用 (FI-1), 对线性变分不等式来说也就是只用 (B1.7), 可以得到

$$\{(u^k-u^*)-(u^k-\tilde{u}^k)\}^T\{\beta(Mu^k+q)-\beta M(u^k-u^*)\}\ge 0. \tag{B1.13}$$

再利用算子的单调性, 有 $(u^k-u^*)^TM(u^k-u^*)\ge 0$, 进而就能从 (B1.13) 得到

$$(u^k-u^*)^T\{\beta[(Mu^k+q)+M^T(u^k-\tilde{u}^k)]\}\ge(u^k-\tilde{u}^k)^T\beta(Mu^k+q),\quad \forall u^*\in\Omega^*. \tag{B1.14a}$$

假如 $u^k\in\Omega$, 在 (B1.3) 中令 $v=u^k-\beta_kF(u^k)$, 由于 $\tilde{u}^k=P_\Omega(v)$, 我们就有

$$(u^k-\tilde{u}^k)^T\beta(Mu^k+q)\ge\|u^k-\tilde{u}^k\|^2. \tag{B1.14b}$$

比较 (B1.14) 和 (B1.9), 求解线性单调变分不等式, 由于少用一个基本不等式 (FI-2), 似乎提供了一个比 $(I+\beta M^T)(u^k-\tilde{u}^k)$ 更好的寻查方向 $\beta[(Mu^k+q)+M^T(u^k-\tilde{u}^k)]$.
在短篇 **B2** 中, 我们将证明: 采用 (B1.12) 中同样的 $\alpha_k^*$, 由

$$u^{k+1}=P_\Omega\{u^k-\gamma\alpha_k^*\beta[(Mu^k+q)+M^T(u^k-\tilde{u}^k)]\},\quad \gamma\in[1.2,1.8]\subset(0,2) \tag{B1.15}$$

生成新的迭代点的方法, 比方法 (B1.12) 更有优势. 取 $\gamma\in[1.2,1.8]$ 是依赖计算经验.

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

对非线性单调变分不等式, 在生成预测点 $\tilde{u}^k$ 的投影 (B1.2) 中, 要求参数 $\beta_k$ 取得满足

$$\beta_k\|F(u^k)-F(\tilde{u}^k)\|\le\nu\|u^k-\tilde{u}^k\|,\quad \nu\in(0,1). \tag{B1.16}$$

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

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

利用记号

$$d(u^k,\tilde{u}^k)=(u^k-\tilde{u}^k)-\beta_k(F(u^k)-F(\tilde{u}^k), \tag{B1.17}$$

我们得到 $$(u^k-u^*)^Td(u^k,\tilde{u}^k)\ge(u^k-\tilde{u}^k)^Td(u^k,\tilde{u}^k),\quad \forall u^*\in\Omega^*. \tag{B1.18}$$

根据 (B1.16) 和 Cauchy-Schwarz 不等式, (B1.18) 右端的 $(u^k-\tilde{u}^k)^Td(u^k,\tilde{u}^k)\ge(1-\nu)\|u^k-\tilde{u}^k\|^2$, 这表明 $d(u^k,\tilde{u}^k)$ 是由 (FI-1) + (FI-2) + (FI-3) 生成的欧氏模下距离函数 $\|u-u^*\|^2$ 在 $u^k$ 处的上升方向. 我们可以通过选择适当的 $\alpha$, 生成比 $u^k$ 更靠近解集的新的迭代点

$$u^{k+1}(\alpha)=u^k-\alpha d(u^k,\tilde{u}^k). \tag{B1.19}$$

考虑最大化 $\vartheta_k(\alpha)=\|u^k-u^*\|^2-\|u^{k+1}(\alpha)-u^*\|^2$, 由于 $\vartheta_k(\alpha)$ 中含有需求的 $u^*$, 利用 (B1.18), 我们得到

$$\begin{array}{rcl}\vartheta_k(\alpha) &=& \|u^k-u^*\|^2-\|u^{k+1}(\alpha)-u^*\|^2 \;=\; \|u^k-u^*\|^2-\|u^k-u^*-\alpha d(u^k,\tilde{u}^k)\|^2\\ &=& 2\alpha(u^k-u^*)^Td(u^k,\tilde{u}^k)-\alpha^2\|d(u^k,\tilde{u}^k)\|^2 \qquad (\text{利用 (B1.18)})\\ &\ge& 2\alpha(u^k-\tilde{u}^k)d(u^k,\tilde{u}^k)-\alpha^2\|d(u^k,\tilde{u}^k)\|^2 \quad =:\; q_k(\alpha).\end{array} \tag{B1.20}$$

注意到 $\vartheta_k(\alpha)$ 和 $q_k(\alpha)$ 都是 $\alpha$ 的二次函数, 但前者含有未知的 $u^*$. 后者是前者的下界函数, 计算 $q_k(\alpha)$ 取得最大值的 $\alpha_k^*=(u^k-\tilde{u}^k)^Td(u^k,\tilde{u}^k)\|d(u^k,\tilde{u}^k)\|^2$, 然后用

$$u^{k+1}=u^k-\gamma\alpha_k^*d(u^k,\tilde{u}^k),\quad \gamma\in[1.2,1.8]\subset(0,2)\quad \text{生成新的迭代点.} \tag{B1.21}$$

如果考虑只用 (FI-1)+(FI-3), 对非线性变分不等式来说可以得到

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

进而有 $$(u^k-u^*)^T\beta_kF(\tilde{u}^k)\ge(u^k-\tilde{u}^k)^T\beta_kF(\tilde{u}^k),\quad \forall u^*\in\Omega^*. \tag{B1.22a}$$

假如 $u^k\in\Omega$, 在 (B1.3) 中令 $v=u^k-\beta_kF(u^k)$, 由于 $\tilde{u}^k=P_\Omega(v)$, 我们就有

$$(\tilde{u}^k-u^k)^T\{u^k-\beta_kF(u^k)-\tilde{u}^k\}\ge 0,$$

从而得到

$$(u^k-\tilde{u}^k)^T\beta_kF(u^k)\ge(u^k-\tilde{u}^k)^T(u^k-\tilde{u}^k)$$

再在上式两端加上 $(u^k-\tilde{u}^k)^T\beta_k[F(\tilde{u}^k)-F(u^k)]$, 并利用 $d(u^k,\tilde{u}^k)$ (见 (B1.17)), 就得到

$$(u^k-\tilde{u}^k)^T\beta_kF(\tilde{u}^k)\ge(u^k-\tilde{u}^k)^Td(u^k,\tilde{u}^k). \tag{B1.22b}$$

比较 (B1.22) 和 (B1.18), 求解非线性单调变分不等式, 由于少用一个基本不等式 (FI-2) 似乎提供了一个比 $d(u^k,\tilde{u}^k)$ 更好的寻查方向 $\beta_kF(\tilde{u}^k)$. 在短篇 **B3** 中, 我们将证明: 采用 (B1.21) 中同样的 $\alpha_k^*$, 由

$$u^{k+1}=P_\Omega\{u^k-\gamma\alpha_k^*\beta_kF(\tilde{u}^k\},\quad \gamma\in[1.2,1.8]\subset(0,2) \tag{B1.23}$$

生成新的迭代点的方法, 比方法 (B1.21) 更有优势. 同样, 取 $\gamma\in[1.2,1.8]$ 是依赖计算经验.

# 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

## 漂亮的孪生方向和姊妹方法 I (参阅文献[3, 4, 1, 2, 7])

### 1 线性 VI 中的孪生方向和相同步长的姊妹方法 (参阅 [3, 4, 1, 7] 和系列讲义 [8] 的前五讲)

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

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

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

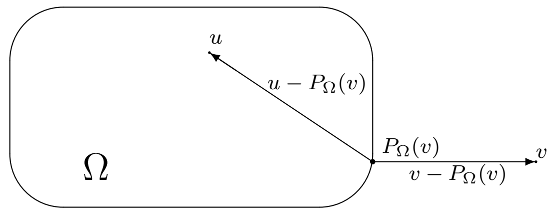
$$\text{(LCP)} \qquad u^* \ge 0, \qquad Mu^*+q \ge 0, \qquad (u^*)^T(Mu^*+q) = 0. \tag{B2.2}$$

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

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

生成预测点 $\tilde{u}^k$. 在投影 (B2.3) 中, 对参数 $\beta > 0$ 理论上没有要求, 但要大致调成 $\beta\|M(u^k-\tilde{u}^k)\| \approx \|u^k-\tilde{u}^k\|$.




假如 $\tilde{u}^k = u^k$, $u^k$ 就是 (B2.1) 的解, 这在定理 A2 中已经证明. 否则, 就要基于以下分析进行新的迭代.

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 $\Omega$ 上的投影, 有

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

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

$$\tilde{u}^k \in \Omega, \quad (u-\tilde{u}^k)^T\{(u^k-\tilde{u}^k) - \beta(Mu^k+q)\} \le 0, \quad \forall u \in \Omega. \tag{B2.5}$$

上式两边都加上 $(u-\tilde{u}^k)^T\beta[M^T(u^k-\tilde{u}^k)+(Mu^k+q)]$, 由此我们基于投影 (B2.3) 得到预测公式

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

**定义 1** (线性 **VI** 中的孪生方向) 在基于投影 (B2.3) 的预测变分不等式 (B2.6) 中, 分处两端的

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

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

**定义 2** (线性 **VI** 中的姊妹方法) 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 $\alpha$ 是相同的待定步长.

**定义 3** (效益函数) 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 $\alpha$ 的 $k$ 次迭代的收益, 是步长 $\alpha$ 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 $\alpha$, $\zeta_k(\alpha)$ 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 **B3** 中非线性情形的证明要大一些. 由 $\tilde{u}^k \in \Omega$, 根据 LVI 的定义 (B2.1), 我们有

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

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

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

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

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

将 (B2.12) 和 (B2.14) 相加得到 $\{(u^k-u^*)-(u^k-\tilde{u}^k)\}^T\{(u^k-\tilde{u}^k)-\beta M(u^k-u^*)\} \ge 0$, 进一步化简,

利用 (B2.8) 中记号 $d(u^k,\tilde{u}^k)$ 得到
$$(u^k-u^*)^Td(u^k,\tilde{u}^k) \ge \|u^k-\tilde{u}^k\|^2, \quad \forall u^* \in \Omega^*. \tag{B2.15}$$

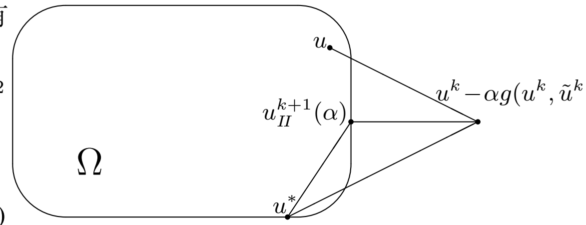
## 2　姊妹方法的收缩性质证明　(参阅 [1, 7, 9, 10] 和系列讲义 [8]中第三讲)

**定理 B1** 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 (B2.11) 给出.

**证明** 根据 (B2.10) 对 $\vartheta_k(\alpha)$ 的定义, 利用 (B2.15) 就有




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

我们证明了定理结论的第一部分. 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

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

下面要经过两个分拆过程. **1)**. 首先将 (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, 因为 $u_{II}^{k+1}(\alpha) \in \Omega$, 将 (B2.6) 中属于 $\Omega$ 的 $u$ 设成 $u_{II}^{k+1}(\alpha)$, 我们得到

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

**2)** 再对 (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

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

**3)** 最后, 将 (B2.24) 代入 (B2.19), 进一步就可以化得

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

注意到最后一式中的 $(u^k - \alpha d(u^k, \tilde{u}^k))$ 就是 $u_I^{k+1}(\alpha)$, 这样就完成了定理结论 (B2.16) 的完整证明. □

这个定理说明, $q_k(\alpha)$ 是 $\vartheta_k(\alpha)$ 和 $\zeta_k(\alpha)$ 的下界. 对同样的 $\alpha$, $\zeta_k(\alpha)$ 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 $\alpha_k^*$ 的取法). 在实际计算中, 我们采用校正公式

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

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

**B2**中对线性VI性质的证明, 比**B3**中非线性的更需技巧, 因为预测(B2.3)只要求$\beta > 0$, 比非线性的条件宽松.

# B3-非线性变分不等式投影收缩算法中的孪生方向和姊妹方法

## 漂亮的孪生方向和姊妹方法 II (参阅文献[3, 4, 1, 2, 7, 8, 9, 10])

### 1 非线性 VI 中的孪生方向和相同步长的姊妹方法 (参阅 [3, 4, 1, 7] 和系列讲义 [8]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

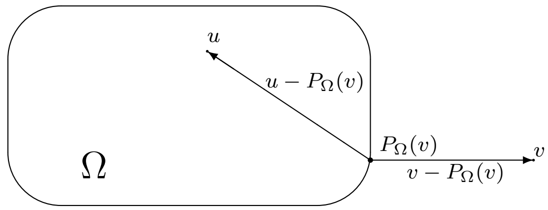
$$\text{(VI)} \qquad u^* \in \Omega, \quad (u-u^*)^T F(u^*) \geq 0, \quad \forall\, u \in \Omega. \tag{B3.1}$$

如果上式中 $\Omega = \Re^n_+$ (非负卦限), 变分不等式 (B3.1) 就是一个互补问题 (Complementarity Problem):

$$\text{(CP)} \qquad u^* \geq 0, \qquad F(u^*) \geq 0, \qquad (u^*)^T F(u^*) = 0. \tag{B3.2}$$

换句话说, 互补问题是变分不等式中 $\Omega = \Re^n_+$ 的子类. 我们说一个变分不等式单调, 是指其中的算子 $F$ 满足 $(u-v)^T(F(u)-F(v)) \geq 0$. 在求解变分不等式 (B3.1) 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

$$\tilde{u}^k = P_\Omega[u^k - \beta_k F(u^k)] \tag{B3.3}$$




生成一个预测点 $\tilde{u}^k$. 假如 $\tilde{u}^k = u^k$, $u^k$ 就是 (B3.1) 的解 (该结论可见定理 A1, 证明也可参见[8]的第二讲). 否则, 就要基于以下分析进行新的迭代.

对非线性单调变分不等式, 在生成预测点 $\tilde{u}^k$ 的投影 (B3.3) 中, 要求参数 $\beta_k$ 取得满足

$$\beta_k\|F(u^k) - F(\tilde{u}^k)\| \leq \nu\|u^k - \tilde{u}^k\|, \quad \nu \in (0,1). \tag{B3.4}$$

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 $\Omega$ 上的投影, 有

$$(u - P_\Omega(v))^T(v - P_\Omega(v)) \leq 0, \quad \forall u \in \Omega. \tag{B3.5}$$

这就是引理 A 的结论. 在 (B3.5) 中令 $v = u^k - \beta_k F(u^k)$, 那么由 (B3.3) $P_\Omega(v) = \tilde{u}^k$, 并从 (B3.5) 得到

$$\tilde{u}^k \in \Omega, \quad (u-\tilde{u}^k)^T\{[u^k - \beta_k F(u^k)] - \tilde{u}^k\} \leq 0, \quad \forall u \in \Omega. \tag{B3.6}$$

上式两边都加上 $(u-\tilde{u}^k)^T\beta_k F(\tilde{u}^k)$, 由此我们基于投影 (B3.3) 得到预测公式

$$\text{[预测]} \qquad \tilde{u}^k \in \Omega, \;\; (u-\tilde{u}^k)^T d(u^k,\tilde{u}^k) \leq (u-\tilde{u}^k)^T \beta_k F(\tilde{u}^k), \;\; \forall u \in \Omega. \tag{B3.7}$$

其中 $$d(u^k,\tilde{u}^k) = (u^k - \tilde{u}^k) - \beta_k[F(u^k) - F(\tilde{u}^k)]. \tag{B3.8}$$

**定义 1** (非线性 VI 中的孪生方向) 基于投影 (B3.3) 得到的预测变分不等式 (B3.7) 中, 分处两端的

$$d(u^k,\tilde{u}^k) \qquad \text{和} \qquad \beta_k F(\tilde{u}^k), \qquad \text{称为一对孪生方向.} \tag{B3.9}$$

将 (B3.7) 中的属于 $\Omega$ 的 $u$ 选成任意的解点 $u^*$, 利用 $F(u)$ 的单调性和 (B3.1), 有

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

因此, 进而有

$$(u^k - u^*)^T d(u^k,\tilde{u}^k) \geq (u^k - \tilde{u}^k)^T d(u^k,\tilde{u}^k). \tag{B3.10}$$

**定义 2** (姊妹方法) 采用 (B3.9) 中一对孪生方向 $d(u^k,\tilde{u}^k)$ 和 $\beta_k F(\tilde{u}^k)$ 分别进行校正

$$u_I^{k+1}(\alpha) = u^k - \alpha d(u^k,\tilde{u}^k), \tag{B3.11}$$

和

$$u_{II}^{k+1}(\alpha) = P_\Omega[u^k - \alpha\beta_k F(\tilde{u}^k)], \tag{B3.12}$$

的方法称为一对姊妹方法. 其中 $\alpha$ 是相同的待定步长.

**定义 3** (效益函数) 对任意给定的解点 $u^* \in \Omega^*$, 我们记

$$\vartheta_k(\alpha) = \|u^k - u^*\|^2 - \|u_I^{k+1}(\alpha) - u^*\|^2 \tag{B3.13}$$

和

$$\zeta_k(\alpha) = \|u^k - u^*\|^2 - \|u_{II}^{k+1}(\alpha) - u^*\|^2. \tag{B3.14}$$

它们是依赖于步长 $\alpha$ 的 $k$ 次迭代的收益, 是步长 $\alpha$ 的函数.

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

$$q_k(\alpha) = 2\alpha(u^k - \tilde{u}^k)^T d(u^k,\tilde{u}^k) - \alpha^2\|d(u^k,\tilde{u}^k)\|^2 \tag{B3.15}$$

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 (B3.3) 中, 要求参数 $\beta_k$ 使得 (B3.4) 成立. 在这篇注记的 §2, 我们将会证明, 对同样的 $\alpha$, $\zeta_k(\alpha)$ 优于 $\vartheta_k(\alpha)$.

## 2 姊妹方法的收缩性质证明 (参阅 [1, 9, 10] 和系列讲义 [8]中第三讲)

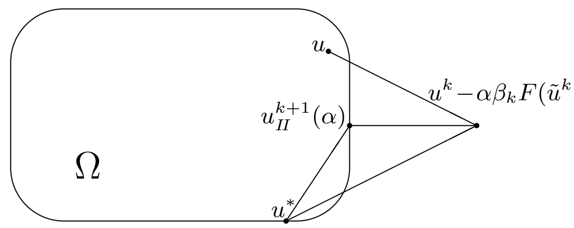
**定理 B2** 设 $u_I^{k+1}(\alpha)$ 和 $u_{II}^{k+1}(\alpha)$ 由姊妹方法生成. 对分别由 (B3.13) 和 (B3.14) 定义的效益函数 $\vartheta_k(\alpha)$ 和 $\zeta_k(\alpha)$, 我们有

$$\vartheta_k(\alpha) \ge q_k(\alpha) \quad 和 \quad \zeta_k(\alpha) \ge q_k(\alpha) + \|u_I^{k+1}(\alpha) - u_{II}^{k+1}(\alpha)\|^2, \tag{B3.16}$$

其中 $q_k(\alpha)$ 由 (B3.15) 给出.

**证明** 根据 (B3.13) 中对 $\vartheta_k(\alpha)$ 的定义, 利用 (B3.10) 我们有

$$\begin{aligned}
\vartheta_k(\alpha) &= \|u^k - u^*\|^2 - \|u_I^{k+1}(\alpha) - u^*\|^2 \\
&= \|u^k - u^*\|^2 - \|u^k - u^* - \alpha d(u^k, \tilde{u}^k)\|^2 \\
&= 2\alpha(u^k - u^*)^T d(u^k, \tilde{u}^k) - \alpha^2 \|d(u^k, \tilde{u}^k)\|^2 \\
&\ge 2\alpha(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \alpha^2 \|d(u^k, \tilde{u}^k)\|^2 \\
&= q_k(\alpha).
\end{aligned} \tag{B3.17}$$




我们证明了定理结论的第一部分. 对定理结论的第二部分, 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

$$\|u_{II}^{k+1}(\alpha) - u^*\|^2 \le \|u^k - \alpha\beta_k F(\tilde{u}^k) - u^*\|^2 - \|u^k - \alpha\beta_k F(\tilde{u}^k) - u_{II}^{k+1}(\alpha)\|^2. \tag{B3.18}$$

因此, 利用 $\zeta_k(\alpha)$ 的定义(见 (B3.14)), 我们有

$$\begin{aligned}
\zeta_k(\alpha) &\ge \|u^k - u^*\|^2 - \|(u^k - u^*) - \alpha\beta_k F(\tilde{u}^k)\|^2 + \|(u^k - u_{II}^{k+1}(\alpha)) - \alpha\beta_k F(\tilde{u}^k)\|^2 \\
&= 2\alpha(u^k - u^*)^T \beta_k F(\tilde{u}^k) + 2\alpha(u_{II}^{k+1}(\alpha) - u^k)^T \beta_k F(\tilde{u}^k) + \|u^k - u_{II}^{k+1}(\alpha)\|^2 \\
&= \|u^k - u_{II}^{k+1}(\alpha)\|^2 + 2\alpha(u_{II}^{k+1}(\alpha) - u^*)^T \beta_k F(\tilde{u}^k).
\end{aligned} \tag{B3.19}$$

将 (B3.19) 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 (B3.19), 进而得到

$$\zeta_k(\alpha) \ge \|u^k - u_{II}^{k+1}(\alpha)\|^2 + 2\alpha(u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T \beta_k F(\tilde{u}^k). \tag{B3.20}$$

因为 $u_{II}^{k+1}(\alpha) \in \Omega$, 用它替代 (B3.7) 中的任意 $u \in \Omega$, 得到

$$(u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T \beta_k F(\tilde{u}^k) \ge (u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \tag{B3.21}$$

将 (B3.21) 代入 (B3.20) 的右端, 就有

$$\zeta_k(\alpha) \ge \|u^k - u_{II}^{k+1}(\alpha)\|^2 + 2\alpha(u_{II}^{k+1}(\alpha) - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \tag{B3.22}$$

对上式右端, 利用 $q_k(\alpha)$ 的形式 (见(B3.15)), 就可以进一步化成

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

利用 (B3.11), 上式右端的第一部分就是 $\|u_I^{k+1}(\alpha) - u_{II}^{k+1}(\alpha)\|^2$. 这就完成了结论 (B3.16) 的全部证明. □

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

$$\alpha_k^* = \arg\max\{q_k(\alpha)\} = (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) / \|d(u^k, \tilde{u}^k)\|^2 \tag{B3.23}$$

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

$$(\text{收缩算法–1}) \quad u^{k+1} = u^k - \gamma\alpha_k^* d(u^k, \tilde{u}^k) \tag{B3.24}$$

或者

$$(\text{收缩算法–2}) \quad u^{k+1} = P_\Omega[u^k - \gamma\alpha_k^* \beta_k F(\tilde{u}^k)] \tag{B3.25}$$

产生新的迭代点 $u^{k+1}$, 其中的 $\alpha_k^*$ 都由 (B3.23) 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 (B3.24), 好处是生成 $u^{k+1}$ 不用再做投影. 实际问题中, 到 $\Omega$ 上的投影代价往往不高 (例如 $\Omega$ 是一个正卦限或者框形), 因此常采用校正公式 (B3.25). 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. 细心的读者会注意到 **B3** 的证明比 **B2** 更显简单.

# 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 (参阅[12, 13, 14, 15] 和讲义 [8] 的第 11 讲)

## 1 两个可分离块的凸优化问题和相应的变分不等式 (参阅 [12, 13] 和 [8] 的 11-15 讲)

乘子交替方向法 (ADMM) 处理的是等式约束的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

$$L(x,y,\lambda)=\theta_1(x)+\theta_2(y)-\lambda^T(Ax+By-b).\tag{C1.2}$$

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

$$(x^*,y^*,\lambda^*)\in\Omega,\quad L(x,y,\lambda^*)\ge L(x^*,y^*,\lambda^*)\ge L(x^*,y^*,\lambda),\quad \forall(x,y,\lambda)\in\Omega,\tag{C1.3}$$

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.

可以把刻画鞍点的不等式 (C1.3) 中分别关于 $x,y,\lambda$ 的部分分开来写成: $(x^*,y^*,\lambda^*)\in\mathcal{X}\times\mathcal{Y}\times\Re^m$,

$$\left\{\begin{array}{l} L(x,y^*,\lambda^*)-L(x^*,y^*,\lambda^*)\ge 0,\ \ \forall x\in\mathcal{X},\\ L(x^*,y,\lambda^*)-L(x^*,y^*,\lambda^*)\ge 0,\ \ \forall y\in\mathcal{Y},\\ L(x^*,y^*,\lambda^*)-L(x^*,y^*,\lambda)\ge 0,\ \ \forall\lambda\in\Re^m.\end{array}\right. \Leftrightarrow \left\{\begin{array}{r} \theta_1(x)-\theta_1(x^*)+(x-x^*)^T(-A^T\lambda^*)\ge 0,\ \ \forall x\in\mathcal{X},\\ \theta_2(y)-\theta_2(y^*)+(y-y^*)^T(-B^T\lambda^*)\ge 0,\ \ \forall y\in\mathcal{Y},\\ (\lambda-\lambda^*)^T(Ax^*+By^*-b)\ge 0,\ \ \forall\lambda\in\Re^m.\end{array}\right.$$

以上的等价关系是由拉格朗日函数 $L(x,y,\lambda)$ 的表达式得来的. 右侧的紧凑表示就是变分不等式

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

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 (聚焦求得变分不等式 (C1.4) 的解)

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

[该写法用到优化中改变目标函数中的常数项并不改变问题解这一性质]

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

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

根据凸优化的最优性原理 (定理 A3), ADMM $k$ 步迭代 (C1.5) 的三个子问题的最优性条件分别是

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

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

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

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

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

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

对上式进行整合 (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

分别把上式中 下波纹线部分 和 下划线部分 组装在一起

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

## 3 基于变分不等式表示的交替方向法收敛性关键性质证明

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

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

若用记号 $\quad v=\begin{pmatrix}y\\\lambda\end{pmatrix}\quad$ 和 $\quad H=\begin{pmatrix}\beta B^TB & 0\\0 & \frac{1}{\beta}I_m\end{pmatrix},\qquad$ (C1.9)

不等式 (C1.8) 就可以改写成

$$\begin{aligned}(v^{k+1}-v^*)^TH(v^k-v^{k+1})\quad\ge\quad&\begin{pmatrix}x^{k+1}-x^*\\y^{k+1}-y^*\end{pmatrix}^T\beta\begin{pmatrix}A^T\\B^T\end{pmatrix}B(y^k-y^{k+1})\\&+\{\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^TF(w^{k+1})\}.\end{aligned}\qquad\text{(C1.10)}$$

下面我们证明 (C1.10) 式右端两部分都非负. 由于 $(w^{k+1}-w^*)^TF(w^{k+1})=(w^{k+1}-w^*)^TF(w^*)$,

$$\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^TF(w^{k+1})=\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^TF(w^*)\ge 0.$$

后面的不等式源于 $w^*$ 是问题的解. 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 推导的最后用到等式 $\beta(Ax^{k+1}+By^{k+1}-b)=(\lambda^k-\lambda^{k+1})$

$$\begin{aligned}\begin{pmatrix}x^{k+1}-x^*\\y^{k+1}-y^*\end{pmatrix}^T\beta\begin{pmatrix}A^T\\B^T\end{pmatrix}B(y^k-y^{k+1})&=(y^k-y^{k+1})^TB^T\beta(A,B)\begin{pmatrix}x^{k+1}-x^*\\y^{k+1}-y^*\end{pmatrix}\\&=\ (y^k-y^{k+1})^TB^T\beta\big(Ax^{k+1}+By^{k+1}-(Ax^*+By^*)\big)\qquad\text{利用}(Ax^*+By^*=b)\\&=\ (y^k-y^{k+1})^TB^T\beta\big(Ax^{k+1}+By^{k+1}-b\big)=(y^k-y^{k+1})^TB^T(\lambda^k-\lambda^{k+1}).\end{aligned}\qquad\text{(C1.12)}$$

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

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

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

(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. 从而得到需要的

$$(v^{k+1}-v^*)^TH(v^k-v^{k+1})\ge 0.\qquad\text{(C1.13)}$$

注意到若有 $b^TH(a-b)\ge 0$ 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 ADMM 收敛性的关键不等式

$$\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2-\|v^k-v^{k+1}\|_H^2,\quad\forall v^*\in\mathcal{V}^*.\qquad\text{(核心变量的收缩性质)}\quad\text{(C1.14)}$$

# C2-预测-校正统一框架下乘子交替方向法 (ADMM) 的收敛性证明

## 在分裂收缩算法统一框架下证明 ADMM 收敛性 (参阅[17, 18])

## 1 两个可分离块的凸优化问题和相应的变分不等式 (参阅 [2, 12, 13] 和 [8] 的 11-15 讲)

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. 从短篇 **C1** 我们已经知道: 问题 (C2.1) 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 $w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})$:

[该写法用到优化中改变目标函数中的常数项并不改变问题解这一性质]

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

## 2 乘子交替方向法分拆成统一框架下的预测-校正

在 **A1** 的 §2 中我们已经提及线性约束凸优化分裂收缩算法的统一框架, 并证明了收敛关键的收缩不等式. 这一篇, 我们在统一框架下演绎 ADMM 方法. 根据 (C2.3) 产生的 $w^{k+1}$, 我们定义预测点 $\tilde w^k$:

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

这样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, 根据 (C2.3) 和 (C2.4), 预测就是

ADMM 的 $k$ 步迭代的预测是从给定的 $v^k=(y^k,\lambda^k)$ 出发, 按顺序求得预测点 $\tilde w^k=(\tilde x^k,\tilde y^k,\tilde\lambda^k)$:

[预测]

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

根据凸优化的最优性原理 (参见定理 A3), 第 $k$ 迭代预测 (C2.5) 的三个子问题的变分不等式形式是

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

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

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

将同类型的并在一起, 并利用变分不等式 (C2.2) 中的记号, 得到预测 (C2.5) 的 VI 表示: $\tilde w^k\in\Omega$,

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

由 (C2.3) 和 (C2.4) 之间的关系, $y^{k+1}=\tilde y^k$, $\lambda^{k+1}=\tilde\lambda^k+\beta(y^k-\tilde y^k)$, 因此新的核心变量 $v^{k+1}$ 由

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

这里的 (C2.7) 和 (C2.8) 分别相当于短篇 **A1 §2** 中统一框架中的预测 (A1.18) 和校正 (A1.19).

## 3 预测-校正分拆的ADMM收敛性质证明

**定理 C1** 采用预测-校正方法 (C2.7)-(C2.8) 求解变分不等式 (C2.2), 产生的序列 $\{w^k\}, \{\tilde{w}^k\}$ 满足

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

其中
$$H=\begin{pmatrix}\beta B^TB & 0\\ 0 & \frac{1}{\beta}I\end{pmatrix},\qquad G=Q^T+Q-M^THM. \quad \text{(C2.10)}$$

进而序列 $\{\|v^k-v^*\|_H^2\}$ 具有收缩性质

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

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

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(v-\tilde{v}^k)^THM(v^k-\tilde{v}^k),\quad \forall w\in\Omega.$$

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

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

将恒等式

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

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

$$(v-\tilde{v}^k)^TH(v^k-v^{k+1})=\tfrac{1}{2}(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2)+\tfrac{1}{2}(\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2). \quad \text{(C2.13)}$$

对 (C2.13) 的右端, 利用 $HM=Q$ 和 $2v^TQv=v^T(Q^T+Q)v$, 我们有

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

上式最后一个等式用了 (C2.10) 中矩阵 $G$ 的定义. 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) 中任意的 $w\in\Omega$ 设成一个解点 $w^*$, 就得到

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

由于 $(\tilde{w}^k-w^*)^TF(\tilde{w}^k)=(\tilde{w}^k-w^*)^TF(w^*)$, 上式右端非负, 因此得到收缩结论 (C2.11). □

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

$$\|v^k-\tilde{v}^k\|_G^2\ge\|v^k-v^{k+1}\|_H^2. \quad \text{(C2.15)}$$

进而根据 (C2.11) 有

(C2.16) 是 ADMM 收敛的关键不等式

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

**证明** 由于 $G=Q^T+Q-M^THM$, 根据由 (C2.7) 和 (C2.10) 分别给出的 $Q$ 和 $H$, 我们有

$$G=Q^T+Q-M^TQ=\begin{pmatrix}0 & 0\\ 0 & \frac{1}{\beta}I\end{pmatrix},\quad \text{因此}\quad \|v^k-\tilde{v}^k\|_G^2=\tfrac{1}{\beta}\|\lambda^k-\tilde{\lambda}^k\|^2. \quad \text{(C2.17)}$$

由 (C2.6), 我们知道 $y$-子问题的最优性条件是

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

根据校正公式 (C2.8), 我们有

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

因此变分不等式 (C2.18) 可以写成

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

上述变分不等式对前一次迭代同样成立, 所以我们同样有

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

将 (C2.19) 和 (C2.20) 中的 $y$ 分别设成 $y=y^k$ 和 $y=y^{k+1}$, 然后再将它们相加, 从而得到

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

利用关系式 $\lambda^k-\tilde{\lambda}^k=(\lambda^k-\lambda^{k+1})+\beta B(y^k-y^{k+1})$ 和不等式 (C2.21), 以及 (C2.10) 中 $H$ 的表达式, 得到

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

从 (C2.17) 和 (C2.22) 直接得到结论 (C2.15). 再由 (C2.11) 和 (C2.15) 得到结论 (C2.16). 定理得证. □

# C3-预测-校正统一框架下交替方向法(ADMM)收敛速率的关键性质

分裂收缩算法统一框架下的收敛速率关键性质证明 (参阅文献[3, 4, 1, 2, 7])

## 1 ADMM 在预测-校正分拆下的性质 (参阅 [2, 17, 18])

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, $A$, $B$, $b$ 是相应的矩阵和向量, $\mathcal{X}, \mathcal{Y}$ 是闭凸集. 短篇 **C1** 中已经说明问题 (C3.1) 的拉格朗日函数的鞍点 $w^* = (x^*, y^*, \lambda^*)$ 等价于变分不等式

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

的解点, 其中

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

短篇 **C2** 中已经说明: 将经典的 ADMM 生成的 $w^{k+1}$, 通过

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

定义预测点 $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$, 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] $\quad \tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega,$ (C3.4)

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

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

**C2 中的主要结论:** 采用预测-校正方法 (C3.4)-(C3.5) 求解变分不等式 (C3.2), 产生的序列具有性质

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

其中

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

此外还有 (见 (C2.16))

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

## 2 遍历意义下收敛速率的关键性质 (参阅[2]和[17] He,Yuan 2012 SIAM Numr. Analysis)

我们证明 ADMM 在遍历意义下的收敛性质. 满足 $\tilde{w} \in \Omega, \ \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(\tilde{w}) \ge 0, \ \forall\, w \in \Omega$ 的 $\tilde{w}$ 是 (C3.2) 的解. 由于 $(w - \tilde{w})^T F(w) = (w - \tilde{w})^T F(\tilde{w})$, 因此满足

$$\tilde{w} \in \Omega, \ \ \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \ge 0, \ \ \forall\, w \in \Omega \tag{C3.8}$$

的 $\tilde{w}$ 也是 (C3.2) 的解. 据此, 对给定的 $\epsilon > 0$, 如果 $\tilde{w}$ 满足

$$\tilde{w} \in \Omega, \ \ \theta(u) - \theta(\tilde{u}) + (w - \tilde{w})^T F(w) \ge -\epsilon, \ \ \forall\, w \in \mathcal{D}_{(\tilde{w})} \tag{C3.9}$$

其中 $\mathcal{D}_{(\tilde{w})} = \{w \in \Omega \,|\, \|w - \tilde{w}\| \le 1\}$, 那么 $\tilde{w}$ 就可以看作变分不等式 (C3.2) 的一个 $\epsilon$-近似解. 以下证明: 对给定的 $\epsilon > 0$, 经过 $t$ 次迭代, 交替方向法能够提供一个 $\tilde{w} \in \mathcal{W}$, 使得

$$\sup_{w \in \mathcal{D}_{(\tilde{w})}} \{\theta(\tilde{u}) - \theta(u) + (\tilde{w} - w)^T F(w)\} \le \epsilon. \tag{C3.10}$$

证明的依据就是 (C3.6). 利用 $(w - \tilde{w}^k)^T F(w) = (w - \tilde{w}^k)^T F(\tilde{w}^k)$ 和 $G \succeq 0$, 从 (C3.6) 我们得到

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(w) + \tfrac{1}{2}\|v - v^k\|_H^2 \ge \tfrac{1}{2}\|v - v^{k+1}\|_H^2, \quad \forall w \in \Omega. \tag{C3.11}$$

**定理 C3** 设 $\{\tilde{w}^k\}$, $\{v^k\}$ 是由预测-校正方法 (C3.4)-(C3.5) 产生的序列. 令

$$\tilde{w}_t = \frac{1}{t+1}\sum_{k=0}^{t} \tilde{w}^k. \tag{C3.12}$$

那么, 对任何正整数 $t > 0$, 我们都有 $\tilde{w}_t \in \Omega$ 和

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^T F(w) \le \frac{1}{2(t+1)}\|v - v^0\|_H^2, \quad \forall w \in \Omega. \tag{C3.13}$$

**证明** 首先, 对每个 $k>0$ 都有 $\tilde{w}^k\in\Omega$. 由于 $\mathcal{X}$ 和 $\mathcal{Y}$ 是闭凸集, (C3.12) 隐含了 $\tilde{w}_t\in\Omega$. 将 $k=0,1,\ldots,t$ 的不等式 (C3.11) 相加, 我们得到

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

利用 $\tilde{w}_t$ 的定义, 从上式得到

$$\frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k)-\theta(u)+(\tilde{w}_t-w)^TF(w)\le\frac{1}{2(t+1)}\|v-v^0\|_H^2,\quad \forall w\in\Omega. \tag{C3.14}$$

因为 $\theta(u)$ 是凸函数并且

$$\tilde{u}_t=\frac{1}{t+1}\sum_{k=0}^{t}\tilde{u}^k,\qquad \text{我们有}\qquad \theta(\tilde{u}_t)\le\frac{1}{t+1}\sum_{k=0}^{t}\theta(\tilde{u}^k).$$

将其代入 (C3.14), 就直接得到定理的结论. □

得益于 **C2** 预测-校正得到的不等式 (C3.11), 证明ADMM遍历意义下收敛速率(C3.13)是如此简单 !

## 3 ADMM 点列意义下收敛性的关键性质 (参阅 [2] 和 [18] He and Yuan 2015 Numer Math.)

**定理 C4** 设 $\{w^k\}$, $\{\tilde{w}^k\}$ 是由预测-校正方法 (C3.4)-(C3.5) 产生的序列. 我们有

$$\|v^{k+1}-v^{k+2}\|_H\le\|v^k-v^{k+1}\|_H,\quad \forall\, k>0. \tag{C3.15}$$

**证明** 首先, 我们证明以下辅助结论:

$$(v^k-v^{k+1})^TH\{(v^k-v^{k+1})-(v^{k+1}-v^{k+2})\}\ge\tfrac{1}{2}\|(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\|_{(Q^T+Q)}^2. \tag{C3.16}$$

在 (C3.4) 中令 $w=\tilde{w}^{k+1}$, 我们有

$$\theta(\tilde{u}^{k+1})-\theta(\tilde{u}^k)+(\tilde{w}^{k+1}-\tilde{w}^k)^TF(\tilde{w}^k)\ge(\tilde{v}^{k+1}-\tilde{v}^k)^TQ(v^k-\tilde{v}^k). \tag{C3.17}$$

将 (C3.4) 中的 $k$ 替换成 $k+1$, 就有

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

在上面的不等式中的 $w$ 设成 $\tilde{w}^k$, 我们得到

$$\theta(\tilde{u}^k)-\theta(\tilde{u}^{k+1})+(\tilde{w}^k-\tilde{w}^{k+1})^TF(\tilde{w}^{k+1})\ge(\tilde{v}^k-\tilde{v}^{k+1})^TQ(v^{k+1}-\tilde{v}^{k+1}). \tag{C3.18}$$

将 (C3.17) 和 (C3.18) 加在一起并利用$(\tilde{w}^k-\tilde{w}^{k+1})^T(F(\tilde{w}^k)-F(\tilde{w}^{k+1}))\equiv 0$, 我们就有

$$(\tilde{v}^k-\tilde{v}^{k+1})^TQ\{(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\}\ge 0. \tag{C3.19}$$

在不等式 (C3.19) 的两边都加上 $\{(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\}^TQ\{(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\}$, 并利用恒等式 $v^TQv=\frac{1}{2}v^T(Q^T+Q)v$, 我们得到

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

对上式左端使用 $Q=HM$ 和 $M(v^k-\tilde{v}^k)=(v^k-v^{k+1})$, 就得到后面要用到的结论 (C3.16).

在恒等式 $\|a\|_H^2-\|b\|_H^2=2a^TH(a-b)-\|a-b\|_H^2$ 中设 $a=(v^k-v^{k+1})$ 和 $b=(v^{k+1}-v^{k+2})$, 得到

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

将已经证明了的结论 (C3.16) 用到上式右端的第一部分, 我们得到

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

其中最后一个不等式成立是因为矩阵 $(Q^T+Q)-M^THM\succeq 0$. 结论(C3.15) 得证. □

利用 $\|v^k-v^{k+1}\|_H^2\le\|v^k-v^*\|_H^2-\|v^{k+1}-v^*\|_H^2$ 和 $\|v^k-v^{k+1}\|_H^2\le\|v^{k-1}-v^k\|_H^2$, 容易得到

$$\|v^k-v^{k+1}\|_H^2\le\tfrac{1}{k+1}\|v^0-v^*\|_H^2,\quad \forall v^*\in\mathcal{V}^*. \tag{C3.20}$$

得益于 **C2** 的预测-校正结论(C3.6)和(C3.7), 是证明ADMM点历意义下收敛速率性质的主要工具.

# D1-凸优化问题对应的变分不等式和因需定制的 PPA 算法

## 凸优化问题的 VI 和因需定制的 PPA (参阅 [23, 24, 25] 和 [26])

## 1 预备定理和引理 (定理 D 已经在短篇 **A2** 中作为 定理 A3 证明. 也可参见 [8] 中系列讲义第六讲)

**定理 D** 设 $\mathcal{X}\subset\Re^n$ 是闭凸集, $\theta(x)$ 和 $f(x)$ 是凸函数, $f(x)$ 在包含 $\mathcal{X}$ 的某个开集上可微. 那么

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

当且仅当

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

**引理 D** 设向量 $a,b\in\Re^n$, $H\in\Re^{n\times n}$ 是半正定矩阵. 如果 $b^TH(a-b)\ge 0$, 则有

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

论断能从 $\|a\|_H^2=\|b+(a-b)\|_H^2\ =\|b\|_H^2+2b^TH(a-b)+\|a-b\|_H^2$ 和 $b^TH(a-b)\ge 0$ 直接推得.

## 2 凸优化和相对应的变分不等式 (参阅 [23, 24, 25] 和 [26])

**1). min-max** 问题. 设 $\theta_1(x),\theta_2(y)$ 是凸函数, $\mathcal{X},\mathcal{Y}$ 是闭凸集, $(x^*,y^*)$ 是 min-max 问题 (鞍点问题)

$$\min_x\max_y\{\Phi(x,y)=\theta_1(x)-y^TAx-\theta_2(y)\,|\,x\in\mathcal{X},y\in\mathcal{Y}\}\tag{D1.1}$$

的解. 利用 $\Phi(x,y)$ 的表达式, 以上关系可以写成

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

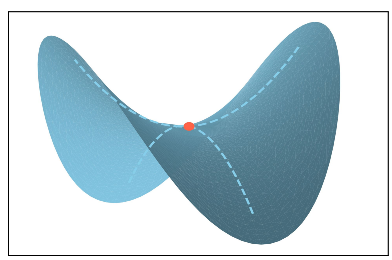
进一步, 可以写成单调变分不等式形式:

$$u^*\in\Omega,\quad \theta(u)-\theta(u^*)+(u-u^*)^TF(u^*)\ge 0,\quad \forall u\in\Omega,\tag{D1.2}$$

$$\text{其中}\quad u=\begin{pmatrix}x\\y\end{pmatrix},\qquad \theta(u)=\theta_1(x)+\theta_2(y),\qquad F(u)=\begin{pmatrix}-A^Ty\\Ax\end{pmatrix}\qquad\text{和}\qquad \Omega=\mathcal{X}\times\mathcal{Y}.\tag{D1.3}$$

注意到 (D1.3) 中的 $F(u)$ 具有性质 $(u-\tilde{u})^T(F(u)-F(\tilde{u}))\equiv 0$, 因此变分不等式 (D1.2) 是单调的.

**2). 线性约束的凸优化问题**



考虑线性约束的凸优化问题

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

它的拉格朗日函数是定义在 $\mathcal{U}\times\Re^m$ 上的

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

如果一对 $(u^*,\lambda^*)\in\mathcal{U}\times\Re^m$ 满足

$$L(u,\lambda^*)\ge L(u^*,\lambda^*)\ge L(u^*,\lambda),\quad (u^*,\lambda^*)\in\mathcal{U}\times\Re^m,$$

则称作拉格朗日函数 (D1.5) 的鞍点. 上面的关系式可以写成$(u^*,\lambda^*)\in\mathcal{U}\times\Re^m$,

$$\begin{cases}L(u,\lambda^*)-L(u^*,\lambda^*)\ge 0, & \forall u\in\mathcal{U},\\ L(u^*,\lambda^*)-L(u^*,\lambda)\ge 0, & \forall\lambda\in\Re^m.\end{cases}\iff\begin{cases}\theta(u)-\theta(u^*)+(u-u^*)^T(-\mathcal{A}^T\lambda^*)\ge 0, & \forall u\in\mathcal{U},\\ \qquad\qquad(\lambda-\lambda^*)^T(\mathcal{A}u^*-b)\ge 0, & \forall\lambda\in\Re^m.\end{cases}\tag{D1.6}$$

利用更紧致的格式, 鞍点可以表示成下述变分不等式

$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall w\in\Omega\tag{D1.7}$$

$$\text{的解, 其中}\qquad w=\begin{pmatrix}u\\\lambda\end{pmatrix},\qquad F(w)=\begin{pmatrix}-\mathcal{A}^T\lambda\\\mathcal{A}u-b\end{pmatrix}\qquad\text{和}\qquad\Omega=\mathcal{U}\times\Re^m.\tag{D1.8}$$

这里的 $F(w)$ 同样具有性质 $(w-\tilde{w})^T(F(w)-F(\tilde{w}))\equiv 0$. 凸优化问题 (D1.4) 转换成了变分不等式 (D1.7).

问题 (D1.1) 和 (D1.4) 分别转换成了变分不等式 (D1.2) 和 (D1.7). **PPA** 聚焦于求解这些相应的变分不等式.

## 3 单调变分不等式的邻近点 (PPA) 算法 (参阅 [5, 23, 26])

**定义 D** (PPA 算法的 $k$次迭代) 设 $\boldsymbol{H}$ 为给定对称正定矩阵. 从给定的 $w^k$ 开始, 求得 $w^{k+1}$, 使得

$$w^{k+1}\in\Omega,\quad \theta(u)-\theta(u^{k+1})+(w-w^{k+1})^TF(w^{k+1})\ge(w-w^{k+1})^TH(w^k-w^{k+1}),\quad \forall\, w\in\Omega. \tag{D1.9}$$

将(D1.9) 中的 $w$ 设为 $w^*$, 并利用$(w^{k+1}-w^*)^TF(w^{k+1})=(w^{k+1}-w^*)^TF(w^*)$, 我们得到

$$(w^{k+1}-w^*)^TH(w^k-w^{k+1})\ge\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^TF(w^*)\ge0. \tag{D1.10}$$

上面最后一个不等号是因为 $w^*\in\Omega^*$. 用引理 D, 设 $a=(w^k-w^*)$ 和 $b=(w^{k+1}-w^*)$, 从 (D1.10) 就得到

(PPA 收敛的收缩不等式) $$\|w^{k+1}-w^*\|_H^2\le\|w^k-w^*\|_H^2-\|w^k-w^{k+1}\|_H^2,\quad \forall w^*\in\Omega^*. \tag{D1.11}$$

用 PPA 算法求解 VI, 建议采用进一步松弛的方法 $w^{k+1}:=w^k-\alpha(w^k-w^{k+1})$, $\alpha\in[1.2,1.8]\subset(0,2)$.

## 4 求解变分不等式的因需定制的PPA(C-PPA)和均困的PPA(B-PPA) (如何实现(D1.9))

**1). 求解 VI (D1.2) 的 Customized-PPA** (参阅 [23])

从给定的 $u^k$ 出发, 求得满足 (D1.9) 的 $u^{k+1}$, 其中 (矩阵 $H$ 可选下面的 $H_1$ 或 $H_2$)

$$u=\begin{pmatrix}x\\y\end{pmatrix},\quad w=u,\quad F(w)=\begin{pmatrix}-A^Ty\\Ax\end{pmatrix},\quad H_1=\begin{pmatrix}rI_n & A^T\\ A & sI_m\end{pmatrix},\quad H_2=\begin{pmatrix}rI_n & -A^T\\ -A & sI_m\end{pmatrix},\ rs>\|A^TA\|. \tag{D1.12}$$

对 (D1.12) 给出的算子 $F$ 和第一个正定矩阵 $H$, PPA 算法 (D1.9) 的具体格式是 下波纹线部分是 $F(w^{k+1})$

$$\begin{cases}\theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\{\underline{-A^Ty^{k+1}}+rI_n(x^{k+1}-x^k)+A^T(y^{k+1}-y^k)\}\ge0,\ \forall x\in\mathcal{X},\\ \theta_2(y)-\theta_2(y^{k+1})\ +\ (y-y^{k+1})^T\{\underline{Ax^{k+1}}+A(x^{k+1}-x^k)+sI_m(y^{k+1}-y^k)\}\ge0,\ \forall y\in\mathcal{Y}.\end{cases}$$

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

**2). 求解 VI (D1.7) 的 Customized-PPA 和 Balanced-PPA** (参阅 [28, 29])

从给定的 $w^k$ 出发, 求得满足 (D1.9) 的 $w^{k+1}$, 其中 (矩阵 $H$ 可选下面的 $H_1$ 或 $H_2$)

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

因需定制的 **C-PPA**. 利用 $H=H_1$ (见(D1.14)) 的 PPA (D1.9) 的具体格式是 下波纹线部分是 $F(w^{k+1})$

$$\begin{cases}\theta(u)-\theta(u^{k+1})+(u-u^{k+1})^T\{\underline{-\mathcal{A}^T\lambda^{k+1}}+\ (\beta\mathcal{A}^T\mathcal{A}+\delta I_n)(u^{k+1}-u^k)+\mathcal{A}^T(\lambda^{k+1}-\lambda^k)\}\ge0,\\ \qquad\qquad\qquad\qquad\qquad\qquad\underline{(\mathcal{A}u^{k+1}-b)}\qquad+\mathcal{A}(u^{k+1}-u^k)\quad+\quad(1/\beta)(\lambda^{k+1}-\lambda^k)=0.\end{cases}$$

根据定理 D, 完成上面的第 $k$-次 PPA 迭代可以通过求解下面的

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

均(分)困(难)的 **B-PPA**. 利用 $H=H_2$ (见(D1.14)) 的 PPA (D1.9) 的具体格式是 下波纹线部分是 $F(w^{k+1})$

$$\begin{cases}\theta(u)-\theta(u^{k+1})+(u-u^{k+1})^T\{\underline{-\mathcal{A}^T\lambda^{k+1}}+\ rI_n(u^{k+1}-u^k)\quad+\quad\mathcal{A}^T\quad(\lambda^{k+1}-\lambda^k)\}\ge0,\\ \qquad\qquad\qquad\qquad\qquad\qquad\underline{(\mathcal{A}u^{k+1}-b)}+\mathcal{A}(u^{k+1}-u^k)+(\frac{1}{r}\mathcal{A}\mathcal{A}^T+\delta I_m)(\lambda^{k+1}-\lambda^k)=0.\end{cases}$$

根据定理 D, 完成上面的第 $k$-次 PPA 迭代, 可以通过求解下面的子问题完成:

$$\begin{cases}u^{k+1}=\text{argmin}\big\{\theta(u)-u^T\mathcal{A}^T\lambda^k+\frac{r}{2}\|u-u^k\|^2\,|\,u\in\mathcal{U}\big\}, & \boxed{\text{原始变量子问题比 (D1.15a)简单}}\quad\text{(D1.16a)}\\ \lambda^{k+1}=\lambda^k-(\frac{1}{r}\mathcal{A}\mathcal{A}^T+\delta I_m)^{-1}\big(\mathcal{A}[2u^{k+1}-u^k]-b\big). & \boxed{\text{求解中只需做一次 Cholesky 分解}}\quad\text{(D1.16b)}\end{cases}$$

跟 (D1.15) 相比, 算法 (D1.16) 降低了 原始变量子问题的迭代难度, 而略微增加了对偶变量更新的工作量.

# D2-两块可分离凸优化问题平行求解原始子问题的 PPA 算法

## 平行处理 PRIME 子问题的 PPA (参阅 [23, 24, 25, 26, 5] 和 [27])

## 1 两个可分离块的凸优化问题和相应的变分不等式

两个可分离块的凸优化问题

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

短篇 **C1** 中已经证明, 凸优化问题 (D2.1) 的 Lagrange 函数的鞍点可以表示成一个变分不等式:

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

其中

$$u=\begin{pmatrix} x\\ y\end{pmatrix},\quad \theta(u)=\theta_1(x)+\theta_2(y),\quad w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix},\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix}, \tag{D2.2b}$$

$\Omega=\mathcal{X}\times\mathcal{Y}\times\Re^m$. 对这里的 $F$, 我们有 $(w-\tilde{w})^T(F(w)-F(\tilde{w}))\equiv 0$.

这类问题可以用 **C1** 中介绍的 ADMM 方法求解, 该篇我们介绍 $x, y$ 子问题可以平行求解的 PPA 算法.

## 2 求解变分不等式 (D2.2) 的 PPA 算法

**定义 D** (PPA 算法的 $k$次迭代) 设 $H$ 为正定矩阵. 从给定的 $w^k$ 开始, 求得 $w^{k+1}$, 使得

$$w^{k+1}\in\Omega,\ \ \theta(u)-\theta(u^{k+1})+(w-w^{k+1})^TF(w^{k+1})\ge(w-w^{k+1})^TH(w^k-w^{k+1}),\ \ \forall\, w\in\Omega. \tag{D2.3}$$

将 (D2.3) 中的 $w$ 设为 $w^*$, 并利用 $(w^{k+1}-w^*)^TF(w^{k+1})=(w^{k+1}-w^*)^TF(w^*)$, 我们得到

$$(w^{k+1}-w^*)^TH(w^k-w^{k+1})\ge\theta(u^{k+1})-\theta(u^*)+(w^{k+1}-w^*)^TF(w^*)\ge 0. \tag{D2.4}$$

上式中最后的不等式是因为 $w^*$ 是解点. 在短篇 **D1** 的引理 D 中设 $a=(w^k-w^*)$ 和 $b=(w^{k+1}-w^*)$, 得到

(PPA 收敛的收缩不等式) $$\|w^{k+1}-w^*\|_H^2\le\|w^k-w^*\|_H^2-\|w^k-w^{k+1}\|_H^2,\quad \forall w^*\in\Omega^*. \tag{D2.5}$$

用 PPA 算法求解 VI, 建议采用进一步松弛的方法 $w^{k+1}:=w^k-\alpha(w^k-w^{k+1})$, $\alpha\in[1.2,1.8]\subset(0,2)$.

根据 PPA算法的 $k$-次迭代的定义, 从给定的 $w^k$ 出发, 求得满足 (D2.3) 的 $w^{k+1}$, 其中

$$u=\begin{pmatrix} x\\ y\end{pmatrix},\quad w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix},\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ Ax+By-b\end{pmatrix},\qquad H=H_1 \text{ 或 } H_2. \tag{D2.6}$$

$H_1$ 和 $H_2$ 都是对称正定矩阵. 为使实现 PPA 迭代的子问题容易求解, 我们取以下的具体形式:

$$H_1=\begin{pmatrix} \beta A^TA+\delta I_{n_1} & 0 & A^T\\ 0 & \beta B^TB+\delta I_{n_2} & B^T\\ A & B & \frac{2}{\beta}I\end{pmatrix},\quad H_2=\begin{pmatrix} \beta A^TA+\delta I_{n_1} & 0 & -A^T\\ 0 & \beta B^TB+\delta I_{n_2} & -B^T\\ -A & -B & \frac{2}{\beta}I\end{pmatrix}. \tag{D2.7}$$

**1).** 采用 (D2.7) 中的 $H=H_1$ 的 **PPA** 算法. 按需定制的 (Customized) PPA 求解变分不等式(D2.2).

采用 (D2.7) 中的 $H=H_1$, PPA 迭代 (D2.3) 的具体形式是: 下波纹线部分组合在一起是 $F(w^{k+1})$

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

根据短篇 **D1** 中的定理 D, 上面的第 $k$-次 PPA 迭代, 只需要通过求解下面的子问题实现: 先原始, 后对偶

$$\begin{cases} x^{k+1}=\text{argmin}\big\{\theta_1(x)-x^TA^T\lambda^k+\frac{1}{2}(x-x^k)^T(\beta A^TA+\delta I_{n_1})(x-x^k)|x\in\mathcal{X}\big\}, & \text{(D2.8a)}\\ y^{k+1}=\text{argmin}\big\{\theta_2(y)-y^TB^T\lambda^k+\frac{1}{2}(y-y^k)^T(\beta B^TB+\delta I_{n_2})(y-y^k)|y\in\mathcal{Y}\big\}, & \text{(D2.8b)}\\ \lambda^{k+1}=\lambda^k-\frac{1}{2}\beta\big[2(Ax^{k+1}+By^{k+1}-b)-(Ax^k+By^k-b)\big]. & \text{(D2.8c)}\end{cases}$$

**2).** 采用 (D2.7) 中的 $H=H_2$ 的 **PPA** 算法　按需定制的 (Custormized) PPA 求解变分不等式(D2.2).

采用 (D2.7) 中的 $H=H_2$, PPA 迭代 (D2.3) 的具体形式是: $\boxed{\text{下波纹线部分组合在一起是 } F(w^{k+1})}$

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

根据短篇 **D1** 中的定理 D, 上面的第 $k$-次 PPA 迭代, 只需要通过求解下面的子问题实现: $\boxed{\text{先对偶, 后原始}}$

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

## 4　求解变分不等式(D2.2)的均困的 B-PPA　(参阅 [24, 25] 和 [28, 29])

当 (D2.8) 和 (D2.9) 中 $y$-子问题求解比较困难的时候, 采用均困方法, 把部分求解困难转移到乘子更新上.

根据 PPA算法的 $k$-次迭代的定义, 从给定的 $w^k$ 出发, 求得满足 (D2.3) 的 $w^{k+1}$, 其中

$$u=\begin{pmatrix}x\\y\end{pmatrix},\quad w=\begin{pmatrix}x\\y\\\lambda\end{pmatrix},\quad F(w)=\begin{pmatrix}-A^T\lambda\\-B^T\lambda\\Ax+By-b\end{pmatrix},\qquad H=H_1 \text{ 或 } H_2, \qquad \text{(D2.10)}$$

$H_1$ 和 $H_2$ 都是对称正定矩阵. 为使实现 PPA 迭代的子问题容易求解, 我们取以下的具体形式:

$$H_1=\begin{pmatrix}\beta A^TA+\delta I & 0 & A^T\\ 0 & (s+\delta)I & B^T\\ A & B & \frac{1}{\beta}I_m+\frac{1}{s}BB^T\end{pmatrix},\quad H_2=\begin{pmatrix}\beta A^TA+\delta I & 0 & -A^T\\ 0 & (s+\delta)I & -B^T\\ -A & -B & \frac{1}{\beta}I_m+\frac{1}{s}BB^T\end{pmatrix}. \quad \text{(D2.11)}$$

**1**). 采用 (D2.11) 中的 $H=H_1$ 的 B-PPA 算法.　均(分)困(难)的 (Balanced) PPA 求解变分不等式(D2.2).

采用 (D2.11) 中的 $H=H_1$, PPA 迭代 (D2.3) 的具体形式是: $\boxed{\text{下波纹线部分组合在一起是 } F(w^{k+1})}$

$$\begin{cases} \theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\{\underline{-A^T\lambda^{k+1}}+(\beta A^TA+\delta I_{n_1})(x^{k+1}-x^k)+A^T(\lambda^{k+1}-\lambda^k)\}\ge 0,\\ \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\{\underline{-B^T\lambda^{k+1}}\ +\ (s+\delta)(y^{k+1}-y^k)\ +\ B^T(\lambda^{k+1}-\lambda^k)\}\ge 0,\\ \qquad (\underline{Ax^{k+1}+By^{k+1}-b})+A(x^{k+1}-x^k)+B(y^{k+1}-y^k)+\big(\frac{1}{\beta}I_m+\frac{1}{s}BB^T\big)\,(\lambda^{k+1}-\lambda^k)=0.\end{cases}$$

根据短篇 **D1** 中的定理 D, 上面的第 $k$-次 PPA 迭代, 只需要通过求解下面的子问题实现: $\boxed{\text{先原始, 后对偶}}$

$$\begin{cases} x^{k+1}=\text{argmin}\big\{\theta_1(x)-x^TA^T\lambda^k+\frac{1}{2}(x-x^k)^T(\beta A^TA+\delta I_{n_1})(x-x^k)|x\in\mathcal{X}\big\}, & \text{(D2.12a)}\\ y^{k+1}=\text{argmin}\big\{\theta_2(y)-y^TB^T\lambda^k+\frac{1}{2}(s+\delta)\|y-y^k\|^2|y\in\mathcal{Y}\big\}, & \text{(D2.12b)}\\ \lambda^{k+1}=\lambda^k-\big(\frac{1}{\beta}I_m+\frac{1}{s}BB^T\big)^{-1}\big[2(Ax^{k+1}+By^{k+1}-b)-(Ax^k+By^k-b)\big]. & \text{(D2.12c)}\end{cases}$$

**2**). 采用 (D2.11) 中的 $H=H_2$ 的 B-PPA 算法.　均困的 (Balanced) PPA 求解变分不等式(D2.2).

采用 (D2.11) 中的 $H=H_2$, PPA 迭代 (D2.3) 的具体形式是: $\boxed{\text{下波纹线部分组合在一起是 } F(w^{k+1})}$

$$\begin{cases} \theta_1(x)-\theta_1(x^{k+1})+(x-x^{k+1})^T\{\underline{-A^T\lambda^{k+1}}+(\beta A^TA+\delta I_{n_1})(x^{k+1}-x^k)-A^T(\lambda^{k+1}-\lambda^k)\}\ge 0,\\ \theta_2(y)-\theta_2(y^{k+1})+(y-y^{k+1})^T\{\underline{-B^T\lambda^{k+1}}\ +\ (s+\delta)(y^{k+1}-y^k)\ -\ B^T(\lambda^{k+1}-\lambda^k)\}\ge 0,\\ \qquad (\underline{Ax^{k+1}+By^{k+1}-b})-A(x^{k+1}-x^k)-B(y^{k+1}-y^k)+\big(\frac{1}{\beta}I_m+\frac{1}{s}BB^T\big)\,(\lambda^{k+1}-\lambda^k)=0.\end{cases}$$

根据短篇 **D1** 中的定理 D, 上面的第 $k$-次 PPA 迭代, 只需要通过求解下面的子问题实现: $\boxed{\text{先对偶, 后原始}}$

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

# E1-求解两个可分离块凸优化问题的 PPA 类 ADMM 方法

## 交换 ADMM 方法中 $y$ 与 $\lambda$ 实现顺序的方法 (参阅[2, 30])

## 1 两块可分离凸优化问题所对应的变分不等式

从短篇 **C1**, 我们已经知道两个可分离块的线性约束凸优化问题 (其中 $\theta_1, \theta_2$ 是凸函数, $\mathcal{X}, \mathcal{Y}$ 是闭凸集)

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

的拉格朗日函数是 $L(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax + By - b)$, 并已经将它的鞍点化成变分不等式

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega \tag{E1.2}$$

的解点, 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, 还是以 $v = (y, \lambda)$ 为核心变量.

## 2 交换顺序的 ADMM 方法 (参阅 [30] Science China Mathematics 2013)

修正的一种自然的想法是交换经典 ADMM 迭代中 $y$ 和 $\lambda$ 更新的顺序. 就是说, 我们同样从给定的核心变量 $v^k = (y^k, \lambda^k)$ 出发, 通过

$$\begin{bmatrix} \boldsymbol{x}\text{-}\boldsymbol{\lambda}\text{-}\boldsymbol{y} \text{ 顺序} \\ \text{的}\textbf{ADMM} \end{bmatrix} \quad \begin{cases} x^{k+1} \in \operatorname{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 \,|\, x \in \mathcal{X}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^k - b), \\ y^{k+1} \in \operatorname{argmin}\{\theta_2(y) - y^T B^T \lambda^{k+1} + \frac{1}{2}\beta\|Ax^{k+1} + By - b\|^2 \,|\, y \in \mathcal{Y}\} \end{cases} \tag{E1.4}$$

直接求得 $w^{k+1}$, 该方法跟短篇 **C1** 中经典的 ADMM 数值表现几乎相当. 附加策略 (E1.8) 可明显提高效率.

根据凸优化的最优性原理 (参见定理 A3), 第 $k$ 迭代预测 (E1.4) 的三个子问题的变分不等式形式是

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

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 $w^{k+1} \in \Omega$,

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

对上面的式子进行整合 (把同一类型的并在一起, 注意到下划线部分是 $\theta(u) - \theta(u^{k+1})$, 下波纹线部分是 $F(w^{k+1})$), 并利用变分不等式 (E1.2) 中的记号, 得到紧凑的形式

$$\theta(u) - (u^{k+1}) + (w - w^{k+1})^T F(w^{k+1}) \geq (v - v^{k+1})^T H_0(v^k - v^{k+1}), \ \ \forall w \in \Omega, \tag{E1.6a}$$

其中

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

这里的 $H_0$ 是半正定的, 跟单元 **D** 中的 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

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

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. 由于 $H_0$ 是对称半正定的, 可以对方法 (E1.4) 产生的 $v^{k+1}$ 做如下的延拓:

$$v^{k+1} := v^k - \alpha(v^k - v^{k+1}), \quad \alpha \in [1.2, 1.8] \subset (0, 2), \tag{E1.8}$$

效率比经典的 ADMM 方法一般有 30% 的提高.

读完该篇就会知道延拓步长 $\alpha$ 的理由

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. 由 $H_0$ 半正定, 从 (E1.7) 式得到

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

利用 $H_0$ 的表达式 (见(E1.6b)) 和 (E1.4) 中的 $\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^k - b)$, 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 $\|v^k - v^{k+1}\|_{H_0}^2 = 0$, 从上式得到 $B(y^k - y^{k+1}) - \frac{1}{\beta}(\lambda^k - \lambda^{k+1}) = 0$ 同时也是 $Ax^{k+1} + By^{k+1} - b = 0$, 将此代入 (E1.5) 就直接得到 $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ 是变分不等式 (E1.2) 的解这一结论.

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

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

利用 (E1.10b) 中 $\tilde{\lambda}^k$ 的表达式, 问题(E1.10c) 的等价形式 优化问题中可以随意改变目标函数中的常数项

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

利用 $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b)$, 由 (E1.10) 得到的 $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \Omega$ 的变分不等式形式是:

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

对上式整合 (把同一类型的并在一起), 利用变分不等式 (E1.2) 中的记号, 得到 (E1.11) 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

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

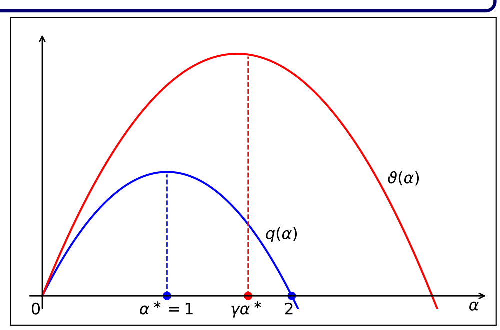
在 (E1.12a) 中令 $w = w^*$, 并利用 $\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^TF(\tilde{w}^k) = \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^TF(w^*) \ge 0$, 得到

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

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




对任意给定的解点 $v^*\in\mathcal{V}^*$, 我们定义

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

为依赖于步长 $\alpha$ 的 $k$ 次迭代的收益(进步). 利用 (E1.13) 得到

$$\begin{aligned} \vartheta_k(\alpha) &= \|v^k - v^*\|_H^2 - \|v^{k+1}(\alpha) - v^*\|_H^2\\ &= \|v^k - v^*\|_H^2 - \|(v^k - v^*) - \alpha(v^k - \tilde{v}^k)\|_H^2\\ &= 2\alpha(v^k - v^*)^TH(v^k - \tilde{v}^k) - \alpha^2\|v^k - \tilde{v}^k\|_H^2\\ &\ge \alpha(2 - \alpha)\|v^k - \tilde{v}^k\|_H^2 =: q_k(\alpha). \end{aligned} \tag{E1.16}$$

我们无法极大化二次函数 $\vartheta_k(\alpha)$, 因为它含有未知的 $v^*$. 由 (E1.16), $\vartheta_k(\alpha)$ 的下界函数 $q_k(\alpha)$ 在 $\alpha^* = 1$ 取得极大值. 右上图所示, 是我们在延拓校正 (E1.8) 和 (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

ADMM 中求解 $x$, $y$ 子问题后都校正一次乘子 $\lambda$ 的方法 (参阅[2, 31, 32])

## 1 两块可分离凸优化问题所对应的变分不等式

从短篇 **C1**, 我们已经知道两个可分离块的线性约束凸优化问题(其中 $\theta_1,\theta_2$ 是凸函数, $\mathcal{X},\mathcal{Y}$ 是闭凸集)

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

的拉格朗日函数是 $L(x,y,\lambda)=\theta_1(x)+\theta_2(y)-\lambda^T(Ax+By-b)$, 并已经将它的鞍点化成变分不等式

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

的解点, 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 (参阅 [2] 2018 运筹学报综述文章)

我们2015年在 [5] 中首次提出, 2018年综述文章中再次提及的算法统一框架每步迭代包括预测与校正.

**[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:

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

预测 (E2.4) 和校正 (E2.5) 分别是 **A1** 中的 (A1.18) 和 (A1.19). 收敛性条件 (E2.6) 与 **A1** 中的 (A1.20) 相同.

因此, 根据 **A1**§2 的知识, 只要条件 (E2.6) 成立, 就有 下式中的 $H$ 和 $G$ 分别称为范数矩阵和效益矩阵

$$\|v^{k+1}-v^*\|_H^2\le\|v^k-v^*\|_H^2-\|v^k-\tilde{v}^k\|_G^2. \tag{E2.7}$$

在后面的 **G-H** 单元中, 有统一框架算法收敛性的更详细证明. **E** 和 **F** 中的短篇只是用算法框架验证收敛性.

## 3 对称的ADMM类方法 (参阅 [2] 和 [31] 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, 各自等同(对称)地校正一次Lagrange乘子. 换句话说, 从给定的 $\mu\in(0,1)$ 和 $v^k=(y^k,\lambda^k)$ 出发, 按照下述方式求得 $w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})$:

$$\text{[对称的}\textbf{ADMM}\text{]}\left\{\begin{array}{ll}x^{k+1}=\text{argmin}\{\theta_1(x)-x^TA^T\lambda^k+\frac{1}{2}\beta\|Ax+By^k-b\|^2\,|\,x\in\mathcal{X}\}, & \text{(E2.8a)}\\ \lambda^{k+\frac{1}{2}}=\lambda^k-\mu\beta(Ax^{k+1}+By^k-b),\qquad\qquad (\mu\in(0,1)) & \text{(E2.8b)}\\ y^{k+1}=\text{argmin}\{\theta_2(y)-y^TB^T\lambda^{k+\frac{1}{2}}+\frac{1}{2}\beta\|Ax^{k+1}+By-b\|^2\,|\,y\in\mathcal{Y}\}, & \text{(E2.8c)}\\ \lambda^{k+1}=\lambda^{k+\frac{1}{2}}-\mu\beta(Ax^{k+1}+By^{k+1}-b). & \text{(E2.8d)}\end{array}\right.$$

我们下面的任务是把方法 (E2.8) 分拆成预测 (E2.4) 和校正 (E2.5), 然后用条件 (E2.6) 去验证方法收敛性.

**1)**. 首先讨论预测. 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, 要使预测 (E2.4) 的右端只有 $v$ 的元素, 中间变量 $x$ 预测的最优性条件要能写成

$$\tilde{x}^k\in\mathcal{X},\quad\theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T(-A^T\tilde{\lambda}^k)\ge 0,\quad\forall x\in\mathcal{X} \tag{E2.9}$$

这种形式, $\tilde{x}^k$ 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, 并令 $\tilde{\lambda}^k=\lambda^k-\beta(A\tilde{x}^k+By^k-b)$ 才行. 将 (E2.8) 生成的新的迭代点 $w^{k+1}$, 刻意用

$$\tilde{w}^k=(\tilde{x}^k,\tilde{y}^k,\tilde{\lambda}^k)=(x^{k+1},y^{k+1},\lambda^k-\beta(Ax^{k+1}+By^k-b))$$ 表示一个预测点, 也就是

$$
\text{[预测]}\quad \begin{cases} \tilde{x}^k \in \text{argmin}\{\theta_1(x) - x^TA^T\lambda^k + \frac{1}{2}\beta\|Ax + By^k - b\|^2 \mid x \in \mathcal{X}\}, & \text{(E2.10a)} \\ \tilde{y}^k \in \text{argmin}\{\theta_2(y) - y^TB^T\lambda^{k+\frac{1}{2}} + \frac{1}{2}\beta\|A\tilde{x}^k + By - b\|^2 \mid y \in \mathcal{Y}\}, & \text{(E2.10b)} \\ \tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k - b), & \text{(E2.10c)} \end{cases}
$$

接着讨论 $\lambda^{k+\frac{1}{2}}$. 利用 $\tilde{x}^k = x^{k+1}$ 和 (E2.10c) 中的 $\tilde{\lambda}^k$, (E2.8b) 中的 $\lambda^{k+\frac{1}{2}}$ 可以表示成

$$
\lambda^{k+\frac{1}{2}} = \lambda^k - \mu\beta(A\tilde{x}^k + By^k - b) = \lambda^k - \mu(\lambda^k - \tilde{\lambda}^k) = \tilde{\lambda}^k - (1-\mu)(\tilde{\lambda}^k - \lambda^k). \tag{E2.11}
$$

利用 (E2.11) 和 (E2.10c) 中的 $\tilde{\lambda}^k$, (E2.10b) 中 $\tilde{y}^k$ 可以表示成

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

这样, 由预测 (E2.10) 得到的这些最优性条件就是:

注意下式中下划线部分是 $\theta(u) - \theta(\tilde{u}^k)$, 波纹线部分是 $F(\tilde{w}^k)$.

$$
\begin{cases}
\tilde{x}^k \in \mathcal{X},\ \underline{\theta_1(x) - \theta_1(\tilde{x}^k)} + (x - \tilde{x}^k)^T\{\underline{-A^T\tilde{\lambda}^k} \} \ge 0,\ \forall x \in \mathcal{X}, & \text{(E2.12a)} \\
\tilde{y}^k \in \mathcal{Y},\ \underline{\theta_2(y) - \theta_2(\tilde{y}^k)} + (y - \tilde{y}^k)^T\{\underline{-B^T\tilde{\lambda}^k} + \beta B^TB(\tilde{y}^k - y^k) - \mu B^T(\tilde{\lambda}^k - \lambda^k)\} \ge 0,\ \forall y \in \mathcal{Y}, & \text{(E2.12b)} \\
\tilde{\lambda}^k \in \Re^m,\quad (\lambda - \tilde{\lambda}^k)^T\{\underline{(A\tilde{x}^k + B\tilde{y}^k - b)} - B(\tilde{y}^k - y^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\} \ge 0,\ \forall \lambda \in \Re^m. & \text{(E2.12c)}
\end{cases}
$$

对上面的式子进行整合 (把同一类型的并在一起, 注意到下划线部分是 $\theta(u) - \theta(u^{k+1})$, 下波纹线部分是 $F(w^{k+1})$), 并利用变分不等式 (E2.2)-(E2.3) 中的记号, 得到 (E1.11) 的紧凑表示

$$
\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^TF(\tilde{w}^k) \ge (v - \tilde{v}^k)^TQ(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \tag{E2.13a}
$$

其中

$$
Q = \begin{pmatrix} \beta B^TB & -\mu B^T \\ -B & \frac{1}{\beta}I \end{pmatrix}. \tag{E2.13b}
$$

**2**). 然后讨论校正. 把由对称的 ADMM (E2.8) 生成的 $w^{k+1}$ 看作由 (E2.10) 生成的 $\tilde{w}^k$ 经校正得来的. 由于 (E2.10) 中的 $\tilde{x}$ 和 $\tilde{y}^k$ 等于 (E2.8) 中的 $x^{k+1}$ 和 $y^{k+1}$. 对 (E2.8d) 中的 $\lambda^{k+1}$, 我们有

$$
\begin{aligned}
\lambda^{k+1} \quad &= \quad \lambda^{k+\frac{1}{2}} - \mu\beta[-B(y^k - \tilde{y}^k) + (A\tilde{x}^k + By^k - b)] \\
&= \quad \lambda^k - \mu[-\beta B(y^k - \tilde{y}^k) + 2\beta(A\tilde{x}^k + By^k - b)].
\end{aligned} \tag{E2.14}
$$

再利用 (E2.10c), 有 $\mu\beta(A\tilde{x}^k + By^k - b) = \mu(\lambda^k - \tilde{\lambda}^k)$, 因此得到

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

换句话说, $w^{k+1}$ 中的核心变量部分 $v^{k+1} = (y^{k+1}, \lambda^{k+1})$ 由

$$
\text{[校正]} \quad \begin{pmatrix} y^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} y^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I & 0 \\ -\mu\beta B & 2\mu I \end{pmatrix} \begin{pmatrix} y^k - \tilde{y}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix} \quad \text{产生.} \tag{E2.15}
$$

**3**). 最后验证收敛性条件. 对于 (E2.15) 中的校正矩阵

$$
M = \begin{pmatrix} I & 0 \\ -\mu\beta B & 2\mu I \end{pmatrix}, \qquad \text{令} \qquad H = \begin{pmatrix} (1 - \frac{1}{2}\mu)\beta B^TB & -\frac{1}{2}B^T \\ -\frac{1}{2}B & \frac{1}{2\mu\beta}I \end{pmatrix}, \tag{E2.16}
$$

可以验证当 $\mu \in (0,1)$ 时, $H$ 正定并有 $HM = Q$ (矩阵 $Q$ 见 (E2.13b)). 此外, 矩阵

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

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.

由自然合理的想法建立的方法, 容易根据预测-校正统一框架验证收敛条件, 最终具有性质 (E2.7).

# F1-求解三个可分离块凸优化问题的Gauss回代型ADMM算法

直接推广的 ADMM 为预测再做回代校正的方法 (参阅文献[2, 34, 36, 37])

## 1 直接推广的ADMM求解三块可分离问题不能保证收敛 (参阅 [33] 2016 Math. Program.)

我们把一般的线性约束三可分离块的凸优化问题 ($\theta_1, \theta_2, \theta_3$ 是凸函数, $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ 是闭凸集)

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

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

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

的解, 其中

$$w = \begin{pmatrix} x \\ y \\ z \\ \lambda \end{pmatrix}, \quad u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ -C^T\lambda \\ Ax + By + Cz - b \end{pmatrix}. \tag{F1.2b}$$

直接推广的 ADMM 处理三个可分离块凸优化问题 (F1.1) 的做法是:

$$\begin{bmatrix}\text{直接推广的} \\ \textbf{ADMM}\text{求解} \\ \text{问题}\textbf{(F1.1)}\end{bmatrix} \left\{\begin{array}{ll} x^{k+1} = \text{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k + Cz^k - b\|^2 \,|\, x \in \mathcal{X}\}, & \text{(F1.3a)} \\ y^{k+1} = \text{argmin}\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\beta\|Ax^{k+1} + By + Cz^k - b\|^2 \,|\, y \in \mathcal{Y}\}, & \text{(F1.3b)} \\ z^{k+1} = \text{argmin}\{\theta_3(z) - z^T C^T \lambda^k + \frac{1}{2}\beta\|Ax^{k+1} + By^{k+1} + Cz - b\|^2 \,|\, z \in \mathcal{Z}\}, & \text{(F1.3c)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). & \text{(F1.3d)} \end{array}\right.$$

我们在 2016 年的文章 [33] 中指出, 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 $B^TC = 0$ 时方法一定收敛. 在一般情形下**,** 不能保证收敛.

## 2 可以用来验证算法收敛的统一框架 (参阅 [36], 2015 年运筹学报)

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

$$\tilde{w}^k \in \Omega, \ \ \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \ \forall\, w \in \Omega, \tag{F1.4}$$

成立. 其中矩阵 $Q^T + Q$ 是正定的. 如果 (F1.4) 中 $v^k = \tilde{v}^k$, $\tilde{w}^k$ 相当于 (F1.2) 的 $w^*$, 就是问题的解.

在短篇 **F1** 和 **F2** 中处理三块的可分离问题时, 都用 $u = (x, y, z)$, $w = (x, y, z, \lambda)$, $v = (y, z, \lambda)$.

[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

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{F1.5}$$

我们分别称 (F1.4) 中的 $Q$ 为预测矩阵, (F1.5) 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:

$$\text{存在正定矩阵} \ \ H \succ 0 \ \ \text{使得} \ \ HM = Q, \ \ \text{并且} \ \ G := Q^T + Q - M^T HM \succ 0. \tag{F1.6}$$

[关键性收缩性质] 核心变量的序列 $\{v^k\}$ 具备收缩性质: 下式中的 $H$ 和 $G$ 分别称为范数矩阵和效益矩阵

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{F1.7}$$

如果条件 (F1.6) 满足, 算法就是收敛的, 在后面的 **G-H** 单元中, 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 (参阅 [34] He, Tao and Yuan 2012 SIAM Opt. 和 [36])

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

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

生成. 要使 (F1.4) 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 $\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + By^k + Cz^k - b)$. 因此, 我们先用

$$
[\text{预测}]\quad
\begin{cases}
\tilde{x}^k \in \arg\min\big\{\theta_1(x) - x^TA^T\lambda^k + \frac{1}{2}\beta\|Ax + By^k + Cz^k - b\|^2 \mid x\in\mathcal{X}\big\}, & \text{(F1.8a)}\\
\tilde{y}^k \in \arg\min\big\{\theta_2(y) - y^TB^T\lambda^k + \frac{1}{2}\beta\|A\tilde{x}^k + By + Cz^k - b\|^2 \mid y\in\mathcal{Y}\big\}, & \text{(F1.8b)}\\
\tilde{z}^k \in \arg\min\big\{\theta_3(z) - z^TC^T\lambda^k + \frac{1}{2}\beta\|A\tilde{x}^k + B\tilde{y}^k + Cz - b\|^2 \mid z\in\mathcal{Z}\big\}, & \text{(F1.8c)}\\
\tilde{\lambda}^k = \lambda^k - \beta\big(A\tilde{x}^k + By^k + Cz^k - b\big). & \text{(F1.8d)}
\end{cases}
$$

得到预测点 $\tilde{w}^k$. 利用优化问题的最优性条件, 得到预测点的 $\tilde{w}^k$ 的 (F1.8) 的变分不等式形式是:

$$
\begin{cases}
\tilde{x}^k\in\mathcal{X},\ \underline{\theta_1(x)-\theta_1(\tilde{x}^k)}+(x-\tilde{x}^k)^T\big\{-A^T\tilde{\lambda}^k\big\}\ge 0,\ \forall x\in\mathcal{X}, & \text{(F1.9a)}\\
\tilde{y}^k\in\mathcal{Y},\ \underline{\theta_2(y)-\theta_2(\tilde{y}^k)}+(y-\tilde{y}^k)^T\big\{-B^T\tilde{\lambda}^k+\beta B^TB(\tilde{y}^k-y^k)\big\}\ge 0,\ \forall y\in\mathcal{Y}, & \text{(F1.9b)}\\
\tilde{z}^k\in\mathcal{Z},\ \underline{\theta_3(z)-\theta_3(\tilde{z}^k)}+(z-\tilde{z}^k)^T\big\{-C^T\tilde{\lambda}^k+\beta C^TB(\tilde{y}^k-y^k)+\beta C^TC(\tilde{z}^k-z^k)\big\}\ge 0,\ \forall z\in\mathcal{Z}, & \text{(F1.9c)}\\
\tilde{\lambda}^k\in\Re^m,\quad (\lambda-\tilde{\lambda}^k)^T\left\{\begin{array}{c}(A\tilde{x}^k+B\tilde{y}^k+C\tilde{z}^k-b)\\ -B(\tilde{y}^k-y^k)-C(\tilde{z}^k-z^k)+\frac{1}{\beta}(\tilde{\lambda}^k-\lambda^k)\end{array}\right\}\ge 0,\ \forall\lambda\in\Re^m. & \text{(F1.9d)}
\end{cases}
$$

对上式整合 (把同类型的并在一起), 并利用变分不等式 (F1.2) 中的记号, 得到 (F1.9) 的紧凑表示

$$
\tilde{w}^k\in\Omega,\quad \underline{\theta(u)-\theta(\tilde{u}^k)}+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(v-\tilde{v}^k)^TQ(v^k-\tilde{v}^k),\quad \forall w\in\Omega, \tag{F1.10a}
$$

其中

$$
Q=\begin{pmatrix}\beta B^TB & 0 & 0\\ \beta C^TB & \beta C^TC & 0\\ -B & -C & \frac{1}{\beta}I\end{pmatrix}. \tag{F1.10b}
$$

$Q^T+Q$ 是本质上正定的 (当 $B, C$ 列满秩时是正定的). 利用这样的预测点, 用校正生成新的核心变量.

$$
[\text{校正}]\quad v^{k+1}=v^k-M(v^k-\tilde{v}^k),\ \ \text{其中}\ \ M=\begin{pmatrix}\nu I & -\nu(B^TB)^{-1}B^TC & 0\\ 0 & \nu I & 0\\ -\beta B & -\beta C & I\end{pmatrix},\ \ \nu\in(0,1). \tag{F1.11}
$$

对于给定的预测矩阵 $Q$ 和校正矩阵 $M$, 存在正定矩阵

$$
H=\begin{pmatrix}\frac{1}{\nu}\beta B^TB & \frac{1}{\nu}\beta B^TC & 0\\ \frac{1}{\nu}\beta C^TB & \frac{1}{\nu}\beta[C^TC+C^TB(B^TB)^{-1}B^TC] & 0\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix} \tag{F1.12}
$$

满足 $HM=Q$. 此外, 当 $\nu\in(0,1)$ 时, 矩阵

$$
G=(Q^T+Q)-M^TQ=\begin{pmatrix}(1-\nu)\beta B^TB & 0 & 0\\ 0 & (1-\nu)\beta C^TC & 0\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix}\succ 0, \tag{F1.13}
$$

收敛性条件 (F1.6) 满足. 收敛性关键不等式 (F1.7) 成立. 计算过程中, 也可以避免校正中的矩阵求逆.

事实上, 由于预测 (F1.8) 只需要 $(By^k, Cz^k, \lambda^k)$ 就可以开始, 校正 (F1.11) 也可以通过

$$
\begin{pmatrix}By^{k+1}\\ Cz^{k+1}\end{pmatrix}=\begin{pmatrix}By^k\\ Cz^k\end{pmatrix}-\begin{pmatrix}\nu I & -\nu I\\ 0 & \nu I\end{pmatrix}\begin{pmatrix}B(y^k-\tilde{y}^k)\\ C(z^k-\tilde{z}^k)\end{pmatrix},\ \ \lambda^{k+1}=\lambda^k-\beta\big(A\tilde{x}^k+B\tilde{y}^k+C\tilde{z}^k-b\big) \tag{F1.14}
$$

实现, 这样就为开始下一次预测 (F1.8) 提供了 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$. 我们将这些理由做一个简要的说明.

将 (F1.10) 中的 $v$ 设为 $v^*$, 就得到 $(\tilde{v}^k-v^*)^TQ(v^k-\tilde{v}^k)\ge 0$. 再做变换, 设

$$
\xi=\begin{pmatrix}By\\ Cz\\ \lambda\end{pmatrix},\quad \text{得到相应的}\quad (\tilde{\xi}^k-\xi^*)^T\mathcal{Q}(\xi^k-\tilde{\xi}^k)\ge 0,\quad \text{其中}\ \mathcal{Q}=\begin{pmatrix}\beta I & 0 & 0\\ \beta I & \beta I & 0\\ -I & -I & \frac{1}{\beta}I\end{pmatrix}. \tag{F1.15}
$$

做了这样的变换以后, 对 (F1.15) 中的 $\mathcal{Q}$, 用 

请留意矩阵 $Q$ 和 $\mathcal{Q}$ 的差别

$$
\mathcal{H}=\begin{pmatrix}\frac{1}{\nu}\beta I & \frac{1}{\nu}\beta I & 0\\ \frac{1}{\nu}\beta I & \frac{2}{\nu}\beta I & 0\\ 0 & 0 & \frac{1}{\beta}I_m\end{pmatrix},\ \ \mathcal{M}=\begin{pmatrix}\nu I & -\nu I & 0\\ 0 & \nu I & 0\\ -\beta I & -\beta I & I\end{pmatrix}\ \ \text{和}\ \ \mathcal{G}=\begin{pmatrix}(1-\nu)\beta I & 0 & 0\\ 0 & (1-\nu)\beta I & 0\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix},
$$

就有 $\mathcal{H}\mathcal{M}=\mathcal{Q}$ 和 $\mathcal{G}=\mathcal{Q}^T+\mathcal{Q}-\mathcal{M}^T\mathcal{H}\mathcal{M}\succ 0$. 这样, 变换下的校正由 $\xi^{k+1}=\xi^k-\mathcal{M}(\xi^k-\tilde{\xi}^k)$ 实现.

校正 $\xi^{k+1}=\xi^k-\mathcal{M}(\xi^k-\tilde{\xi}^k)$ 写开来的具体形式就是 (F1.14). 如同 (F1.7), 序列 $\{\xi^k\}$ 具备收缩性质

$$
\|\xi^{k+1}-\xi^*\|_{\mathcal{H}}^2\le\|\xi^k-\xi^*\|_{\mathcal{H}}^2-\|\xi^k-\tilde{\xi}^k\|_{\mathcal{G}}^2,\quad \forall\xi^*\in\Xi^*. \tag{F1.16}
$$

# F2- 求解三个可分离块凸优化问题部分平行正则化的 ADMM 算法

## 对 $y$ 和 $z$ 子问题平行求解并加正则项的方法 (参阅文献 [2, 35, 36])

## 1 直接推广的 ADMM 求解三块可分离问题不能保证收敛 (参阅 [33] 2016 Math. Program)

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

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

的解, 其中

$$w = \begin{pmatrix} x \\ y \\ z \\ \lambda \end{pmatrix}, \quad u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ -C^T\lambda \\ Ax + By + Cz - b \end{pmatrix}. \tag{F2.2b}$$

我们已经知道 [33], 用直接推广的 ADMM 求解三块可分离凸优化问题 (F2.1), 一般情形下, 不保证收敛.

## 2 可以用来验证算法收敛的统一框架 (参阅 [5] J. Oper. Res. Soc. China (2015))

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 [2] 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

$$\tilde{w}^k \in \Omega, \ \ \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \ \ \forall\, w \in \Omega, \tag{F2.3}$$

成立. 其中矩阵 $Q^T + Q$ 是本质上正定的. 如果上式中 $v^k = \tilde{v}^k$, $\tilde{w}^k$ 相当于 (F2.2) 中的 $w^*$, 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

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{F2.4}$$

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:

$$\text{存在正定矩阵} \ \ H \succ 0 \ \ \text{使得} \ \ HM = Q, \ \ \text{并且} \ \ G := Q^T + Q - M^T HM \succ 0. \tag{F2.5}$$

这样, 序列 $\{v^k\}$ 具备收缩性质
$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{F2.6}$$

如果条件 (F2.5) 满足, 算法就是收敛的, 后面的短篇 **G1** 中的定理 G, 有统一框架算法收敛性更详细的证明.

## 3 部分平行并加正则项的 ADMM 方法 (参阅 [35] IMA Numr. Analysis, 2015)

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, 求解 $x$ 子问题以后, 平行处理 $y$ 和 $z$ 子问题. 通过

$$\left\{\begin{array}{ll} x^{k+1} \in \text{argmin}\{\theta_1(x) - x^T A^T \lambda^k + \frac{\beta}{2}\|Ax + By^k + Cz^k - b\|^2 \,|\, x \in \mathcal{X}\}, & \text{(F2.7a)} \\ y^{k+1} \in \text{argmin}\{\theta_2(y) - y^T B^T \lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By + Cz^k - b\|^2 + \frac{\tau\beta}{2}\|B(y - y^k)\|^2 \,|\, y \in \mathcal{Y}\}, & \text{(F2.7b)} \\ z^{k+1} \in \text{argmin}\{\theta_3(z) - z^T C^T \lambda^k + \frac{\beta}{2}\|Ax^{k+1} + By^k + Cz - b\|^2 + \frac{\tau\beta}{2}\|C(z - z^k)\|^2 \,|\, z \in \mathcal{Z}\}, & \text{(F2.7c)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b) & \text{(F2.7d)} \end{array}\right.$$

直接求得 $w^{k+1}$. 当 $\tau > 1$ 时, 我们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 (F2.7) 中的 $y^{k+1}$ 可以通过

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

同理, (F2.7) 中的 $z^{k+1}$ 可以通过

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

记 $\boldsymbol{\mu = \tau + 1}$, 方法 (F2.7) 可以改写成等价的便于编程处理的

$$\begin{cases} x^{k+1} \in \text{argmin}\{\theta_1(x) - x^T A^T\lambda^k + \frac{1}{2}\beta\|Ax + By^k + Cz^k - b\|^2 \,|\, x \in \mathcal{X}\}, & \text{(F2.8a)} \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^k + Cz^k - b) & \text{(F2.8b)} \\ y^{k+1} \in \text{argmin}\{\theta_2(y) - y^T B^T\lambda^{k+\frac{1}{2}} + \frac{\mu}{2}\beta\|B(y - y^k)\|^2 \,|\, y \in \mathcal{Y}\}, & \text{(F2.8c)} \\ z^{k+1} \in \text{argmin}\{\theta_3(z) - z^T C^T\lambda^{k+\frac{1}{2}} + \frac{\mu}{2}\beta\|C(z - z^k)\|^2 \,|\, z \in \mathcal{Z}\}, & \text{(F2.8d)} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} + Cz^{k+1} - b). & \text{(F2.8e)} \end{cases}$$

用 **A1** §2中的统一框架, 收敛性证明才比较简单, 为此我们把方法 (F2.8) 故意拆解成等价的预测-校正方法.

**[预测]**

$$\begin{cases} \tilde{x}^k \in \arg\min\big\{\theta_1(x) - x^T A^T\lambda^k + \frac{1}{2}\beta\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\big\}, & \text{(F2.9a)} \\ \tilde{y}^k \in \arg\min\big\{\theta_2(y) - y^T B^T\tilde{\lambda}^k + \frac{\mu}{2}\beta\|B(y - y^k)\|^2 | y \in \mathcal{Y}\big\}, & \text{(F2.9b)} \\ \tilde{z}^k \in \arg\min\big\{\theta_3(z) - z^T C^T\tilde{\lambda}^k + \frac{\mu}{2}\beta\|C(z - z^k)\|^2 | z \in \mathcal{Z}\big\}, & \text{(F2.9c)} \\ \tilde{\lambda}^k = \lambda^k - \beta\big(A\tilde{x}^k + By^k + Cz^k - b\big). \qquad (\mu = \tau + 1) & \text{(F2.9d)} \end{cases}$$

方法 (F2.8) 中的 $w^{k+1}$ 和预测 (F2.9) 中 $\tilde{w}^k$ 之间的关系是：

$$x^{k+1} = \tilde{x}^k, \quad y^{k+1} = \tilde{y}^k, \quad z^{k+1} = \tilde{z}^k, \quad \text{和} \quad \lambda^{k+1} = \tilde{\lambda}^k + \beta B(y^k - \tilde{y}^k) + \beta C(z^k - \tilde{z}^k) \quad \text{(F2.10)}$$

这样, 由预测 (F2.9) 得到的这些最优性条件就是:

$$\begin{cases} \tilde{x}^k \in \mathcal{X}, & \underline{\theta_1(x) - \theta_1(\tilde{x}^k)} + (x - \tilde{x}^k)^T\{\underset{\sim}{-A^T\tilde{\lambda}^k} \qquad\qquad\qquad\qquad \} \ge 0, \ \forall x \in \mathcal{X}, & \text{(F2.11a)} \\ \tilde{y}^k \in \mathcal{Y}, & \underline{\theta_2(y) - \theta_2(\tilde{y}^k)} + (y - \tilde{y}^k)^T\{\underset{\sim}{-B^T\tilde{\lambda}^k} + \mu\beta B^T B(\tilde{y}^k - y^k)\} \ge 0, \ \forall y \in \mathcal{Y}, & \text{(F2.11b)} \\ \tilde{z}^k \in \mathcal{Z}, & \underline{\theta_3(z) - \theta_3(\tilde{z}^k)} + (z - \tilde{z}^k)^T\{\underset{\sim}{-C^T\tilde{\lambda}^k} + \mu\beta C^T C(\tilde{z}^k - z^k)\} \ge 0, \ \forall z \in \mathcal{Z}, & \text{(F2.11c)} \\ \tilde{\lambda}^k \in \Re^m, & (\lambda - \tilde{\lambda}^k)^T \left\{ \begin{array}{c} \underset{\sim}{(A\tilde{x}^k + B\tilde{y}^k + C\tilde{z}^k - b)} \\ -B(\tilde{y}^k - y^k) - C(\tilde{z}^k - z^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) \end{array} \right\} \ge 0, \ \forall \lambda \in \Re^m. & \text{(F2.11d)} \end{cases}$$

把上式中同一类型的并在一起, 下划线的是 $\theta(\tilde{u}^k)$, 下波纹线的是 $F(\tilde{w}^k)$. 得到 (F2.11) 的紧凑表示:

$$\tilde{w}^k \in \Omega, \quad \underline{\theta(u) - \theta(\tilde{u}^k)} + (w - \tilde{w}^k)^T \underset{\sim}{F(\tilde{w}^k)} \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \qquad \text{(F2.12a)}$$

其中

$$Q = \begin{pmatrix} \mu\beta B^T B & 0 & 0 \\ 0 & \mu\beta C^T C & 0 \\ -B & -C & \frac{1}{\beta}I \end{pmatrix}. \qquad \text{(F2.12b)}$$

根据 (F2.10), 从预测 (F2.9) 得到的 $\tilde{w}^k$ 再回到 (F2.8) 中核心变量 $v^{k+1}$, 校正公式是

**[校正]** $$v^{k+1} = v^k - M(v^k - \tilde{v}^k), \quad \text{其中} \quad M = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\beta B & -\beta C & I \end{pmatrix}. \qquad \text{(F2.13)}$$

注意到 (F2.12) 中的预测矩阵和 (F2.13) 中的校正矩阵分别是

$$Q = \begin{pmatrix} \mu\beta B^T B & 0 & 0 \\ 0 & \mu\beta C^T C & 0 \\ -B & -C & \frac{1}{\beta}I \end{pmatrix} \quad \text{和} \quad M = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\beta B & -\beta C & I \end{pmatrix}. \qquad \text{(F2.14)}$$

选取对称矩阵

$$H = \begin{pmatrix} \mu\beta B^T B & 0 & 0 \\ 0 & \mu\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix}, \qquad \text{(F2.15)}$$

可以验证 $HM = Q$. 当 $B, C$ 列满秩时 $H$ 正定. 此外,

$$G = (Q^T + Q) - M^T Q = \begin{pmatrix} (\mu-1)\beta B^T B & -\beta B^T C & 0 \\ -\beta C^T B & (\mu-1)\beta C^T C & 0 \\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix}. \qquad \text{(F2.16)}$$

由于 $\mu = \tau + 1 > 2$, 矩阵 $G$ 本质上正定. 收敛性条件 (F2.5) 满足. 在 He and Yuan, Optimization Online 6325 中 (此文正式发表在 [38] Springer Proc. Math. Stat., 360, Springer, Singapore, 2021), 我们已经证明:

事实上, 将方法 (F2.8) 中 $\mu > 2$ 改成 $\mu > 1.5$ (实际计算中取 $\mu$ 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

## 统一框架收敛性条件的等价表示 (参阅文献[39, 40, 41, 45])

## 1 能够用来验证方法保证收敛的统一框架 (古稀前总结出的算法统一框架[2, 6, 36])

我们把线性约束凸优化问题的 Lagrange 函数的鞍点归结为变分不等式

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

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

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

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 $w$ 相同, 也可以是 $w$ 的部分分量. 例如, 在 ADMM 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：

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

如果条件 (G1.4) 满足, 算法就是收敛的. 读者将会在这一篇的 §3 中看到, 关键收敛性质证明相当容易.
以往我们是把方法看成(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 $G$ 本质上正定, 这相当于只要求它们半正定, 而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本质上正定的矩阵就一定正定.
就像短篇 **C1** 中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 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, 则方法生成的迭代序列 $\{v^k\}$ 和 $\{\tilde{w}^k\}$ 具有性质

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge \tfrac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \tfrac{1}{2}\|v^k - \tilde{v}^k\|_G^2, \ \ \forall w \in \Omega \tag{G1.5}$$

和

$$\|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{G1.6}$$

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

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T HM(v^k - \tilde{v}^k).$$

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

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T H(v^k - v^{k+1}). \tag{G1.7}$$

将恒等式

$$(a - b)^T H(c - d) = \tfrac{1}{2}\{\|a - d\|_H^2 - \|a - c\|_H^2\} + \tfrac{1}{2}\{\|c - b\|_H^2 - \|d - b\|_H^2\} \tag{G1.8}$$

用于不等式 (G1.7) 的右端, 并设 $a = v, \ b = \tilde{v}^k, \ c = v^k$, 和 $d = v^{k+1}$, 我们得到

$$(v - \tilde{v}^k)^T H(v^k - v^{k+1}) = \tfrac{1}{2}\big(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2\big) + \tfrac{1}{2}\big(\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2\big). \tag{G1.9}$$

对 (G1.9) 右端第二部分, 利用 (G1.3), $HM = Q$ 和 $2v^T Qv = v^T(Q^T + Q)v$, 我们有

$$\begin{aligned}
\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 &\overset{(G1.3)}{=} \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\
&= (v^k - \tilde{v}^k)^T (HM + M^T H)(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\
&= (v^k - \tilde{v}^k)^T (Q + Q^T - M^T HM)(v^k - \tilde{v}^k) \quad (\text{利用 (G1.4) 中矩阵 } G \text{ 的定义}) \\
&= \|v^k - \tilde{v}^k\|_G^2.
\end{aligned} \tag{G1.10}$$

以 (G1.10) 的结果代入 (G1.9), 就得到

$$(v-\tilde{v}^k)^T H(v^k - v^{k+1}) = \tfrac{1}{2}\big(\|v-v^{k+1}\|_H^2 - \|v-v^k\|_H^2\big) + \tfrac{1}{2}\|v^k-\tilde{v}^k\|_G^2$$

把这个结果代入 (G1.7), 就得到结论 (G1.5).

为证明定理结论得第二部分, 只要将 (G1.5) 中的 $w$ 设为任意解点 $w^*$, 就有

$$\|v^k - v^*\|_H^2 - \|v^{k+1}-v^*\|_H^2 \ge \|v^k-\tilde{v}^k\|_G^2 + 2\{\theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^T F(\tilde{w}^k)\}. \tag{G1.11}$$

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

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

因此从 (G1.11) 得到

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

这就是定理的结论 (G1.6). □

将 ADMM 做预测-校正拆解, 短篇 **C2** 的定理 C1 证明了与 (G1.5) 同样的性质.

就像短篇 **C3** 中根据定理 C1 得到 ADMM 遍历意义下收敛速率性质的定理 C3, 从结论 (G1.5) 同样可以推得统一框架算法遍历意义下的收敛速率性质. 关于 ADMM 的遍历意义下收敛速率就是基于不等式 (G1.5) 证明的. 见 B. S. He and X. M. Yuan, SIAM J. Numer. Anal, 2012, 50: 700-709.
我们称 (G1.6) 中的 $H$ 为范数矩阵, $G$ 为效益矩阵. 结论 (G1.6) 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.

## 3 可以用来构造一簇算法的等价的统一框架 (古稀后发现等价的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).

我们推导出一个与之等价的构造校正矩阵 $M$ 和 $H$ 的方法:

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

$$\iff \left\{\begin{array}{r}D\succ 0,\quad G\succ 0,\\ D+G=Q^T+Q,\\ Q^TM=D,\\ HM=Q.\end{array}\right. \iff \left\{\begin{array}{r}D\succ 0,\quad G\succ 0,\\ D+G=Q^T+Q,\\ M=Q^{-T}D,\\ H=QD^{-1}Q^T.\end{array}\right. \tag{G1.12}$$

选校正矩阵 $M$, 要求存在 $H\succ 0$, 使 $HM=Q$ 和保证 $G\succ 0$. 上式得到的 $H$ 和 $M$ 满足这些条件.

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.
现在的做法是: 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 $M$, 实行校正 (G1.3), 收敛性条件自然满足. 而 $Q^T$ 的求逆往往是容易的!
校正 $v^{k+1}=v^k-M(v^k-\tilde{v}^k)$ 也可以通过 $Q^T(v^{k+1}-v^k)=D(\tilde{v}^k-v^k)$ 实现, $Q$ 往往具有特殊结构.

## 4 从验证方法收敛的框架到自由设计算法的方针 (合格预测提供了足够多的校正方法)

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(G1.12)和(G1.4) 等价, 可根据 (G1.12) 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§3中的等价性条件为算法设计提供了方针.

设计算法的方针. 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 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!
本单元的 **G2**, **G3**, 以及 **H** 单元中的 **H2**, **H3** 会有不少例子说明如何用这里的方针去设计算法.

# G2-根据收敛条件的等价表示构造求解三可分离块问题的一簇算法

## 以直接推广的 ADMM 当预测的不同校正方法 (参阅文献[39, 40, 41, 45])

### 1 能够用来验证方法保证收敛的统一框架 (古稀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:

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

这里的条件 (G2.4) 主要用来验证方法收敛性, 而短篇 **G1** 中 §3 的等价性证明为开发算法设计提供了方针.

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

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, 本质上容易求逆.

### 2 三个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 VI 在短篇 **F1** 中已经做了介绍. 我们采用其中的 (F1.8) 做预测.

[预测]

$$\begin{cases}
\tilde{x}^k \in \arg\min\big\{\theta_1(x) - x^T A^T \lambda^k + \frac{1}{2}\beta\|Ax + By^k + Cz^k - b\|^2 \mid x \in \mathcal{X}\big\}, & \text{(G2.5a)}\\
\tilde{y}^k \in \arg\min\big\{\theta_2(y) - y^T B^T \lambda^k + \frac{1}{2}\beta\|A\tilde{x}^k + By + Cz^k - b\|^2 \mid y \in \mathcal{Y}\big\}, & \text{(G2.5b)}\\
\tilde{z}^k \in \arg\min\big\{\theta_3(z) - z^T C^T \lambda^k + \frac{1}{2}\beta\|A\tilde{x}^k + B\tilde{y}^k + Cz - b\|^2 \mid z \in \mathcal{Z}\big\}, & \text{(G2.5c)}\\
\tilde{\lambda}^k = \lambda^k - \beta\big(A\tilde{x}^k + By^k + Cz^k - b\big). & \text{(G2.5d)}
\end{cases}$$

利用优化问题的最优性条件, 得到预测 (G2.5) 的变分不等式形式 (G2.2), 其预测矩阵 $Q$ 为 (参见 (F1.10))

[预测矩阵]

$$Q = \begin{pmatrix} \beta B^T B & 0 & 0 \\ \beta C^T B & \beta C^T C & 0 \\ -B & -C & \frac{1}{\beta} I \end{pmatrix}. \tag{G2.6}$$

### 3 基于(G2.6)的预测矩阵 $Q$ 选取不同的校正 (获取满足 $Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k)$ 的 $v^{k+1}$)

若记 $P = \begin{pmatrix} B & 0 & 0 \\ 0 & C & 0 \\ 0 & 0 & I \end{pmatrix}$, 则有

$$Q = \begin{pmatrix} \beta B^T B & 0 & 0 \\ \beta C^T B & \beta C^T C & 0 \\ -B & -C & \frac{1}{\beta} I \end{pmatrix} = P^T \begin{pmatrix} \beta I & 0 & -I \\ \beta I & \beta I & 0 \\ -I & -I & \frac{1}{\beta} I \end{pmatrix} P. \tag{G2.7}$$

同样的道理, 对 $Q^T + Q$, 我们有

$$Q^T + Q = \begin{pmatrix} 2\beta B^T B & \beta B^T C & -B^T \\ \beta C^T B & 2\beta C^T C & -C^T \\ -B & -C & \frac{2}{\beta} I \end{pmatrix} = P^T \begin{pmatrix} 2\beta I & \beta I & -I \\ \beta I & 2\beta I & -I \\ -I & -I & \frac{2}{\beta} I \end{pmatrix} P. \tag{G2.8}$$

我们将 (G2.7) 和 (G2.8) 右端中心部分的矩阵分别称为矩阵 $Q$ 和 $(Q^T + Q)$ 的核, 记成 $\mathcal{Q}$ 和 $(\mathcal{Q}^T + \mathcal{Q})$. 即

$$\mathcal{Q} = \begin{pmatrix} \beta I & 0 & 0 \\ \beta I & \beta I & 0 \\ -I & -I & \frac{1}{\beta} I \end{pmatrix}, \quad \mathcal{Q}^T + \mathcal{Q} = \begin{pmatrix} 2\beta I & \beta I & -I \\ \beta I & 2\beta I & -I \\ -I & -I & \frac{2}{\beta} I \end{pmatrix}. \quad \text{注意到} \quad \mathcal{Q}^{-T} = \begin{pmatrix} \frac{1}{\beta} I & -\frac{1}{\beta} I & 0 \\ 0 & \frac{1}{\beta} I & I \\ 0 & 0 & \beta I \end{pmatrix}. \tag{G2.9}$$

有了 (G2.6) 中给定的预测矩阵 $Q$, 我们可以通过选择不同的 $D$, 构造不一样的校正方法. 因为 $k$-次迭代从给定的 $(By^k, Cz^k, \lambda^k)$ 进行预测 (G2.5), 为了下一次迭代, $k$-次校正只要提供 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$. 基于 (G2.6) 给出的预测矩阵的核矩阵, 我们举例说明选择不同的满足条件的矩阵 $\mathcal{D}$, 得到满足 $Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k)$ 的 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$, 完成相应的校正.

### 3.1 **将核矩阵 $\mathcal{Q}^T+\mathcal{Q}$ 分解成两个正定矩阵方法–I** [为下一次迭代预测准备 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$]

$$\mathcal{Q}^T+\mathcal{Q}=\begin{pmatrix} 2\beta I & \beta I & -I \\ \beta I & 2\beta I & -I \\ -I & -I & \frac{2}{\beta}I \end{pmatrix}=\begin{pmatrix} (1+\nu)\beta I & \beta I & -I \\ \beta I & (1+\nu)\beta I & -I \\ -I & -I & \frac{1}{\beta}I \end{pmatrix}+\begin{pmatrix} (1-\nu)\beta I & 0 & 0 \\ 0 & (1-\nu)\beta I & 0 \\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix}, \quad \text{(G2.10)}$$

其中 $0<\nu<1$. 这样就把 $Q^T+Q$ 的核矩阵 $\mathcal{Q}^T+\mathcal{Q}$ 分解成两个正定矩阵.

我们把 (G2.10) 右端的第一个矩阵为 $\mathcal{D}$, 则 $D=P^T\mathcal{D}P$, 方程

$$Q^T(v^{k+1}-v^k)=D(\tilde{v}^k-v^k) \quad \text{(G2.11)}$$

就可以写成

$$P^T\mathcal{Q}^TP(v^{k+1}-v^k)=P^T\mathcal{D}P(\tilde{v}^k-v^k). \quad \text{(G2.12)}$$

通过

$$\mathcal{Q}^TP(v^{k+1}-v^k)=\mathcal{D}P(\tilde{v}^k-v^k)$$

得到的 $Pv^{k+1}$ 就满足 (G2.12). 把上式具体写开来就是

$$\begin{pmatrix} \beta I & \beta I & -I \\ 0 & \beta I & -I \\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix}\begin{pmatrix} By^{k+1}-By^k \\ Cz^{k+1}-Cz^k \\ \lambda^{k+1}-\lambda^k \end{pmatrix}=\begin{pmatrix} (1+\nu)\beta I & \beta I & -I \\ \beta I & (1+\nu)\beta I & -I \\ -I & -I & \frac{1}{\beta}I \end{pmatrix}\begin{pmatrix} B\tilde{y}^k-By^k \\ C\tilde{z}^k-Cz^k \\ \tilde{\lambda}^k-\lambda^k \end{pmatrix}. \quad \text{(G2.13)}$$

利用 (G2.9) 中的 $\mathcal{Q}^{-T}$ 和这里的 $\mathcal{D}$, 求得 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$ 的公式为:

$$\begin{pmatrix} By^{k+1} \\ Cz^{k+1} \\ \lambda^{k+1} \end{pmatrix}=\begin{pmatrix} By^k \\ Cz^k \\ \lambda^k \end{pmatrix}-\begin{pmatrix} \nu I & -\nu I & 0 \\ 0 & \nu I & 0 \\ -\beta I & -\beta I & I \end{pmatrix}\begin{pmatrix} B(y^k-\tilde{y}^k) \\ C(z^k-\tilde{z}^k) \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}. \quad \text{(G2.14)}$$

这恰好是短篇 **F1** 的校正 (F1.14). 对同一预测, 我们可以通过选择不同的符合条件的 $D$, 构造不同的方法.

### 3.2 **将核矩阵 $\mathcal{Q}^T+\mathcal{Q}$ 分解成两个正定矩阵方法–II** [为下一次迭代预测准备 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$]

$$\mathcal{Q}^T+\mathcal{Q}=\begin{pmatrix} 2\beta I & \beta I & -I \\ \beta I & 2\beta I & -I \\ -I & -I & \frac{2}{\beta}I \end{pmatrix}=\begin{pmatrix} \nu\beta I & 0 & 0 \\ 0 & \nu\beta I & 0 \\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix}+\begin{pmatrix} (2-\nu)\beta I & \beta I & -I \\ \beta I & (2-\nu)\beta I & -I \\ -I & -I & \frac{1}{\beta}I \end{pmatrix}, \quad \text{(G2.15)}$$

其中 $\nu\in(0,1)$. 同样 (G2.15) 把 $Q^T+Q$ 的核矩阵 $\mathcal{Q}^T+\mathcal{Q}$ 分解成了两个正定矩阵, 把上式右端的第一个矩阵选为 $\mathcal{D}$. 同样可以通过

$$\mathcal{Q}^TP(v^{k+1}-v^k)=\mathcal{D}P(\tilde{v}^k-v^k)$$

得到满足 $P^T\mathcal{Q}^TP(v^{k+1}-v^k)=P^T\mathcal{D}P(\tilde{v}^k-v^k)$ 的 $Pv^{k+1}$. 具体写开来就是

$$\begin{pmatrix} \beta I & \beta I & -I \\ 0 & \beta I & -I \\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix}\begin{pmatrix} By^{k+1}-By^k \\ Cz^{k+1}-Cz^k \\ \lambda^{k+1}-\lambda^k \end{pmatrix}=\begin{pmatrix} \nu\beta I & 0 & 0 \\ 0 & \nu\beta I & 0 \\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix}\begin{pmatrix} B\tilde{y}^k-By^k \\ C\tilde{z}^k-Cz^k \\ \tilde{\lambda}^k-\lambda^k \end{pmatrix}. \quad \text{(G2.16)}$$

利用 (G2.9) 中的 $\mathcal{Q}^{-T}$ 和这里的 $\mathcal{D}$, 求得 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$ 的公式为:

$$\begin{pmatrix} By^{k+1} \\ Cz^{k+1} \\ \lambda^{k+1} \end{pmatrix}=\begin{pmatrix} By^k \\ Cz^k \\ \lambda^k \end{pmatrix}-\begin{pmatrix} \nu I & -\nu I & 0 \\ 0 & \nu I & \frac{1}{\beta}I \\ 0 & 0 & I \end{pmatrix}\begin{pmatrix} B(y^k-\tilde{y}^k) \\ C(z^k-\tilde{z}^k) \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}. \quad \text{(G2.17)}$$

### 3.3 **将核矩阵 $\mathcal{Q}^T+\mathcal{Q}$ 分解成两个成比例的正定矩阵** [为下一次迭代预测准备 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$]

$$\mathcal{Q}^T+\mathcal{Q}=\alpha\begin{pmatrix} 2\beta I & \beta I & -I \\ \beta I & 2\beta I & -I \\ -I & -I & \frac{2}{\beta}I \end{pmatrix}+(1-\alpha)\begin{pmatrix} 2\beta I & \beta I & -I \\ \beta I & 2\beta I & -I \\ -I & -I & \frac{2}{\beta}I \end{pmatrix}, \quad \alpha\in(0,1). \quad \text{(G2.18)}$$

我们把 (G2.18) 右端的第一个矩阵为 $\mathcal{D}$, 求得 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$ 的方程为

$$\begin{pmatrix} \beta I & \beta I & -I \\ 0 & \beta I & -I \\ 0 & 0 & \frac{1}{\beta}I \end{pmatrix}\begin{pmatrix} By^{k+1}-By^k \\ Cz^{k+1}-Cz^k \\ \lambda^{k+1}-\lambda^k \end{pmatrix}=\alpha\begin{pmatrix} 2\beta I & \beta I & -I \\ \beta I & 2\beta I & -I \\ -I & -I & \frac{2}{\beta}I \end{pmatrix}\begin{pmatrix} B(y^k-\tilde{y}^k) \\ C(z^k-\tilde{z}^k) \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}. \quad \text{(G2.19)}$$

利用 (G2.9) 中的 $\mathcal{Q}^{-T}$ 和这里的 $\mathcal{D}$, 求得 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$ 的公式为:

$$\begin{pmatrix} By^{k+1} \\ Cz^{k+1} \\ \lambda^{k+1} \end{pmatrix}=\begin{pmatrix} By^k \\ Cz^k \\ \lambda^k \end{pmatrix}-\alpha\begin{pmatrix} I & -I & 0 \\ 0 & I & \frac{1}{\beta}I \\ -\beta I & -\beta I & 2I \end{pmatrix}\begin{pmatrix} B(y^k-\tilde{y}^k) \\ C(z^k-\tilde{z}^k) \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}, \quad \alpha\in(0,1). \quad \text{(G2.20)}$$

# G3-三个可分离块问题分裂收缩算法的预测矩阵设计和方法实现

## 从设计合理的预测核矩阵出发构造一簇算法 (参阅文献[39, 40, 41, 45])

## 1　统一框架下三个可分离块的分裂收缩算法

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

$$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z) \mid Ax+By+Cz=b,\ x\in\mathcal{X}, y\in\mathcal{Y}, z\in\mathcal{Z}\} \tag{G3.1}$$

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

**(VI)**
$$w^*\in\Omega,\quad \theta(u)-\theta(u^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega, \tag{G3.2a}$$

的解, 其中

$$w=\begin{pmatrix} x\\ y\\ z\\ \lambda \end{pmatrix},\quad u=\begin{pmatrix} x\\ y\\ z \end{pmatrix},\quad F(w)=\begin{pmatrix} -A^T\lambda\\ -B^T\lambda\\ -C^T\lambda\\ Ax+By+Cz-b \end{pmatrix}. \tag{G3.2b}$$

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 (G3.1), 在一般情形下并不能保证收敛. 为求解 (G3.2) 这样的变分不等式, 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

$$\tilde{w}^k\in\Omega,\ \ \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^TF(\tilde{w}^k)\ge(v-\tilde{v}^k)^TQ(v^k-\tilde{v}^k),\ \ \forall\, w\in\Omega, \tag{G3.3}$$

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

$$v^{k+1}=v^k-M(v^k-\tilde{v}^k). \tag{G3.4}$$

**[收敛性条件]** 对满足 $Q^T+Q\succ 0$ 的预测矩阵 $Q$, 要求设计的校正矩阵 $M$ 满足如下条件:

$$\text{存在正定矩阵}\ \ H\succ 0\ \ \text{使得}\ \ HM=Q,\ \ \text{并且}\ \ G:=Q^T+Q-M^THM\succ 0. \tag{G3.5}$$

等价的实现. 对满足 $Q^T+Q\succ 0$ 的预测矩阵 $Q$, 选 $D\succ 0, G\succ 0$, 使得 $D+G=Q^T+Q$, 再求解

$$Q^T(v^{k+1}-v^k)=D(\tilde{v}^k-v^k)\quad \text{完成}. \tag{G3.6}$$

若预测(G3.3)从给定的 $(By^k, Cz^k, \lambda^k)$ 就可以开始, 校正只要求能提供满足(G3.6)的 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$.

## 2　不同预测矩阵的设计和校正方法的实现

在求解变分不等式 (G3.2) 的过程中, 若仍然以 $v=(y,z,\lambda)$ 为核心变量, 则总要求预测同时满足

$$\theta_1(x)-\theta_1(\tilde{x}^k)+(x-\tilde{x}^k)^T(-A^T\tilde{\lambda}^k)\ge 0\qquad \text{和}\qquad \tilde{\lambda}^k=\lambda^k-\beta\big(A\tilde{x}^k+By^k+Cz^k-b\big).$$

这样的 $\tilde{x}^k$ 可以通过对给定的 $(By^k, Cz^k, \lambda^k)$ 求解 $x$-子问题

$$\tilde{x}^k\in\arg\min\big\{\theta_1(x)-x^TA^T\lambda^k+\tfrac{1}{2}\beta\|Ax+By^k+Cz^k-b\|^2 \mid x\in\mathcal{X}\big\}\quad \text{得到}.$$

### 2.1　设定核矩阵 $\mathcal{Q}$ 满足 $\mathcal{Q}^T+\mathcal{Q}\succ 0$ [预测按 $x,\lambda,y,z$ 的顺序进行]

$$\text{设定[核矩阵]}\ \ \mathcal{Q}=\begin{pmatrix} 2\beta I & 0 & -I\\ \beta I & 2\beta I & -I\\ -I & -I & \frac{1}{\beta}I \end{pmatrix},\ \text{相应的预测矩阵}\ Q=\begin{pmatrix} 2\beta B^TB & 0 & -B^T\\ \beta C^TB & 2\beta C^TC & -C^T\\ -B & -C & \frac{1}{\beta}I \end{pmatrix}. \tag{G3.7}$$

利用变分不等式 (G3.2) 和 (G3.7) 中矩阵 $Q$ 的表达式, 此时预测 (G3.3) 的具体形式是

$$\left\{\begin{array}{ll}
\underline{\theta_1(x)-\theta_1(\tilde{x}^k)}+(x-\tilde{x}^k)^T\{-A^T\tilde{\lambda}^k\}\ge 0,\ \forall x\in\mathcal{X}, & \text{(G3.8a)}\\
\underline{\theta_2(y)-\theta_2(\tilde{y}^k)}+(y-\tilde{y}^k)^T\{-B^T\tilde{\lambda}^k+2\beta B^TB(\tilde{y}^k-y^k)-B^T(\tilde{\lambda}^k-\lambda^k)\}\ge 0,\ \forall y\in\mathcal{Y}, & \text{(G3.8b)}\\
\underline{\theta_3(z)-\theta_3(\tilde{z}^k)}+(z-\tilde{z}^k)^T\left\{\begin{array}{r} -C^T\tilde{\lambda}^k+\beta C^T[B(\tilde{y}^k-y^k)+2C(\tilde{z}^k-z^k)]\\ -C^T(\tilde{\lambda}^k-\lambda^k) \end{array}\right\}\ge 0,\ \forall z\in\mathcal{Z}, & \text{(G3.8c)}\\
(\lambda-\tilde{\lambda}^k)^T\left\{\begin{array}{l} (A\tilde{x}^k+B\tilde{y}^k+C\tilde{z}^k-b)\\ -B(\tilde{y}^k-y^k)-C(\tilde{z}^k-z^k)+\frac{1}{\beta}(\tilde{\lambda}^k-\lambda^k) \end{array}\right\}\ge 0,\ \forall\lambda\in\Re^m. & \text{(G3.8d)}
\end{array}\right.$$

根据定理 A3, 这里的 $\tilde{w}^k=(\tilde{x}^k,\tilde{y}^k,\tilde{z}^k,\tilde{\lambda}^k)$ 可以通过求解下面的子问题生成:

$$
[\text{预测}]\left\{\begin{array}{ll}
\tilde{x}^k \in \arg\min\big\{\theta_1(x) - x^T A^T\lambda^k + \frac{1}{2}\beta\|Ax + By^k + Cz^k - b\|^2 \mid x\in\mathcal{X}\big\}, & \text{(G3.9a)}\\
\tilde{\lambda}^k = \lambda^k - \beta\big(A\tilde{x}^k + By^k + Cz^k - b\big), & \text{(G3.9b)}\\
\tilde{y}^k \in \arg\min\big\{\theta_2(y) - y^T B^T(2\tilde{\lambda}^k - \lambda^k) + \beta\|B(y-y^k)\|^2 \mid y\in\mathcal{Y}\big\}, & \text{(G3.9c)}\\
\tilde{z}^k \in \arg\min\big\{\theta_3(z) + z^T C^T[\beta B(\tilde{y}^k - y^k) - (2\tilde{\lambda}^k - \lambda^k)] + \beta\|C(z-z^k)\|^2 \mid z\in\mathcal{Z}\big\}. & \text{(G3.9d)}
\end{array}\right.
$$

将正定核矩阵 $\mathcal{Q}^T+\mathcal{Q}$ (见 (G3.7)) 分解成两个正定矩阵的和, 有无穷多的可能性. 例如, 我们将其分解成

$$
\mathcal{Q}^T+\mathcal{Q} = \begin{pmatrix} 4\beta I & \beta I & -2I\\ \beta I & 4\beta I & -2I\\ -2I & -2I & \frac{2}{\beta}I\end{pmatrix} = \begin{pmatrix} \frac{9}{4}\beta I & 0 & -I\\ 0 & \frac{9}{4}\beta I & -I\\ -I & -I & \frac{1}{\beta}I\end{pmatrix} + \begin{pmatrix} \frac{7}{4}\beta I & \beta I & -I\\ \beta I & \frac{7}{4}\beta I & -I\\ -I & -I & \frac{1}{\beta}I\end{pmatrix}. \tag{G3.10}
$$

选上述右端两个正定矩阵之一为 $\mathcal{D}$, 就可以通过下式得到下一次迭代开始需要的 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$:

$$
\begin{pmatrix} By^{k+1}\\ Cz^{k+1}\\ \lambda^{k+1}\end{pmatrix} = \begin{pmatrix} By^k\\ Cz^k\\ \lambda^k\end{pmatrix} - \mathcal{Q}^{-T}\mathcal{D}\begin{pmatrix} B(y^k-\tilde{y}^k)\\ C(z^k-\tilde{z}^k)\\ \lambda^k-\tilde{\lambda}^k\end{pmatrix}, \quad \text{其中}\quad \mathcal{Q}^{-T} = \begin{pmatrix} \frac{1}{\beta}I & 0 & I\\ \frac{1}{\beta}I & \frac{1}{\beta}I & 2I\\ 2I & I & 4\beta I\end{pmatrix}. \tag{G3.11}
$$

### 2.2 设定核矩阵 $\mathcal{Q}$ 满足 $\mathcal{Q}^T+\mathcal{Q}\succ 0$ [预测按 $x,y,z,\lambda$ 的顺序进行]

$$
[\text{设定核矩阵}]\ \mathcal{Q} = \begin{pmatrix} \beta I & 0 & I\\ \beta I & \beta I & I\\ -I & -I & \frac{1}{\beta}I\end{pmatrix}, \quad \text{相应的预测矩阵}\quad Q = \begin{pmatrix} \beta B^TB & 0 & B^T\\ \beta C^TB & \beta C^TC & C^T\\ -B & -C & \frac{1}{\beta}I\end{pmatrix}. \tag{G3.12}
$$

利用变分不等式 (G3.2) 和 (G3.12) 中矩阵 $Q$的表达式, 此时预测 (G3.3) 的具体形式是

$$
\left\{\begin{array}{ll}
\underline{\theta_1(x)-\theta_1(\tilde{x}^k)} + (x-\tilde{x}^k)^T\big\{\underbrace{-A^T\tilde{\lambda}^k}\big\} \ge 0,\ \forall x\in\mathcal{X}, & \text{(G3.13a)}\\
\underline{\theta_2(y)-\theta_2(\tilde{y}^k)} + (y-\tilde{y}^k)^T\big\{\underbrace{-B^T\tilde{\lambda}^k} + \beta B^TB(\tilde{y}^k-y^k) + B^T(\tilde{\lambda}^k-\lambda^k)\big\} \ge 0,\ \forall y\in\mathcal{Y}, & \text{(G3.13b)}\\
\underline{\theta_3(z)-\theta_3(\tilde{z}^k)} + (z-\tilde{z}^k)^T\left\{\begin{array}{r}\underbrace{-C^T\tilde{\lambda}^k} + \beta C^T[B(\tilde{y}^k-y^k)+C(\tilde{z}^k-z^k)]\\ +C^T(\tilde{\lambda}^k-\lambda^k)\end{array}\right\} \ge 0,\ \forall z\in\mathcal{Z}, & \text{(G3.13c)}\\
(\lambda-\tilde{\lambda}^k)^T\left\{\begin{array}{l}\underbrace{(A\tilde{x}^k+B\tilde{y}^k+C\tilde{z}^k-b)}\\ -B(\tilde{y}^k-y^k)-C(\tilde{z}^k-z^k)+\frac{1}{\beta}(\tilde{\lambda}^k-\lambda^k)\end{array}\right\} \ge 0,\ \forall\lambda\in\Re^m. & \text{(G3.13d)}
\end{array}\right.
$$

根据定理 A3, 这里的 $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k, \tilde{\lambda}^k)$ 可以通过求解下面的子问题生成:

$$
[\text{校正}]\left\{\begin{array}{ll}
\tilde{x}^k \in \arg\min\big\{\theta_1(x) - x^TA^T\lambda^k + \frac{1}{2}\beta\|Ax+By^k+Cz^k-b\|^2 \mid x\in\mathcal{X}\big\}, & \text{(G3.14a)}\\
\tilde{y}^k \in \arg\min\big\{\theta_2(y) - y^TB^T\lambda^k + \frac{1}{2}\beta\|B(y-y^k)\|^2 \mid y\in\mathcal{Y}\big\}, & \text{(G3.14b)}\\
\tilde{z}^k \in \arg\min\big\{\theta_3(z) + z^TC^T[\beta B(\tilde{y}^k-y^k) - \lambda^k] + \frac{1}{2}\beta\|C(z-z^k)\|^2 \mid z\in\mathcal{Z}\big\}, & \text{(G3.14c)}\\
\tilde{\lambda}^k = \lambda^k - \beta\big(A\tilde{x}^k + By^k + Cz^k - b\big). & \text{(G3.14d)}
\end{array}\right.
$$

将正定核矩阵 $\mathcal{Q}^T+\mathcal{Q}$ (见 (G3.12)) 分解成两个正定矩阵的和, 有无穷多的可能性. 例如, 我们将其分解成

$$
\mathcal{Q}^T+\mathcal{Q} = \begin{pmatrix} 2\beta I & \beta I & 0\\ \beta I & 2\beta I & 0\\ 0 & 0 & \frac{2}{\beta}I\end{pmatrix} = \begin{pmatrix} \frac{1}{2}\beta I & 0 & 0\\ 0 & \frac{1}{2}\beta I & 0\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix} + \begin{pmatrix} \frac{3}{2}\beta I & \beta I & 0\\ \beta I & \frac{3}{2}\beta I & 0\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix}. \tag{G3.15}
$$

选上述右端两个正定矩阵之一为 $\mathcal{D}$, 就可以通过下式得到下一次迭代开始需要的 $(By^{k+1}, Cz^{k+1}, \lambda^{k+1})$:

$$
\begin{pmatrix} By^{k+1}\\ Cz^{k+1}\\ \lambda^{k+1}\end{pmatrix} = \begin{pmatrix} By^k\\ Cz^k\\ \lambda^k\end{pmatrix} - \mathcal{Q}^{-T}\mathcal{D}\begin{pmatrix} B(y^k-\tilde{y}^k)\\ C(z^k-\tilde{z}^k)\\ \lambda^k-\tilde{\lambda}^k\end{pmatrix}, \quad \text{其中}\quad \mathcal{Q}^{-T} = \begin{pmatrix} \frac{1}{\beta}I & -\frac{1}{\beta}I & 0\\ -\frac{1}{2\beta}I & \frac{1}{\beta}I & \frac{1}{2}I\\ -\frac{1}{2}I & 0 & \frac{1}{2}\beta I\end{pmatrix}. \tag{G3.16}
$$

### 2.3 方法 III 设定 $\mathcal{Q}$ 满足$\mathcal{Q}^T+\mathcal{Q}\succ 0$ [预测按 $x,\lambda$, 然后 $y,z$ 平行处理的顺序进行]

$$
[\text{设定核矩阵}]\ \mathcal{Q} = \begin{pmatrix} \frac{3}{2}\beta I & 0 & 0\\ 0 & \frac{3}{2}\beta I & 0\\ -I & -I & \frac{1}{\beta}I\end{pmatrix}, \ \text{相应的预测矩阵}\ Q = \begin{pmatrix} \frac{3}{2}\beta B^TB & 0 & 0\\ 0 & \frac{3}{2}\beta C^TC & 0\\ -B & -C & \frac{1}{\beta}I\end{pmatrix}. \tag{G3.17}
$$

读者可以自行考虑如何根据变分不等式 (G3.2) 和 (G3.17) 中矩阵 $Q$的表达式, 去实现预测 (G3.3).
将正定核矩阵 $\mathcal{Q}^T+\mathcal{Q}$ (见 (G3.12)) 分解成两个正定矩阵的和, 有无穷多的可能性. 例如, 我们将其分解成

$$
\mathcal{Q}^T+\mathcal{Q} = \begin{pmatrix} 3\beta I & 0 & -I\\ 0 & 3\beta I & -I\\ -I & -I & \frac{2}{\beta}I\end{pmatrix} = \begin{pmatrix} \frac{1}{2}\beta I & 0 & 0\\ 0 & \frac{1}{2}\beta I & 0\\ 0 & 0 & \frac{1}{\beta}I\end{pmatrix} + \begin{pmatrix} \frac{5}{2}\beta I & 0 & -I\\ 0 & \frac{5}{2}\beta I & -I\\ -I & -I & \frac{1}{\beta}I\end{pmatrix}. \tag{G3.18}
$$

对同一问题 (G3.1), 说明如何通过设计不同的预测核矩阵, 并由此构造一簇预测-校正的分裂收缩算法

# 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

## 进展 II–预测-校正实现的范数矩阵和效益矩阵 (参阅文献[39, 40, 41, 45])

## 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

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

的解点. 为求解这个变分不等式, 我们在 **D** 单元介绍了邻近点算法 (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. 短篇 **E1**, **E2** 介绍的交替方向法中 $u = (x, y), w = (x, y, \lambda), v = (y, \lambda)$.

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 $a = (v^k - v^*), b = (v^{k+1} - v^*)$, 利用引理 **D** $\boxed{b^T H(a - b) \ge 0 \ \Rightarrow\ \|b\|_H^2 \le \|a\|_H^2 - \|a - b\|_H^2}$ 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, 我们还在短篇 **A1** 的 §2 和 **G1** 中介绍了求解变分不等式 (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

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

我们分别称 (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:

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

短篇《从好不容易凑出一个算法到并不费劲构造一簇算法》的定理 G 中, 证明了算法的收缩性质

$$\text{[收缩性质]} \qquad \|v^{k+1} - v^*\|_H^2 \le \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2, \quad \forall v^* \in \mathcal{V}^*. \tag{H1.7}$$

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

$$\theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \geq (v-\tilde{v}^k)^T Q(v^k-\tilde{v}^k),\ \forall w\in\Omega \tag{H1.10}$$

和

$$\theta(u)-\theta(\tilde{u}^{k+1})+(w-\tilde{w}^{k+1})^T F(\tilde{w}^{k+1}) \geq (v-\tilde{v}^{k+1})^T Q(v^{k+1}-\tilde{v}^{k+1}),\ \forall w\in\Omega. \tag{H1.11}$$

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

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

和

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

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

$$(\tilde{v}^k-\tilde{v}^{k+1})^T Q\{(v^k-\tilde{v}^k)-(v^{k+1}-\tilde{v}^{k+1})\}\geq 0. \tag{H1.12}$$

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

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

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

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

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

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

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

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

这就是 (H1.8), 因此也有 (H1.9). 证明完毕. □

短篇 **C3** 的定理 C4, 视 ADMM 为特例证明了同样的性质.

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

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

然后令 $M^THM=D$, 由矩阵方程组解得

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

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, 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

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

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

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

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)！

# H2-求解多块可分离凸优化问题的预测-校正型广义邻近点算法

## 广义邻近点算法求解多块可分离凸优化问题 (参阅文献[39, 40, 41, 45])

## 1 优化问题的变分不等式及算法框架

考虑多块可分离凸优化问题

$$\min\{\textstyle\sum_{i=1}^p \theta_i(x_i) \mid \sum_{i=1}^p A_i x_i = b,\ x_i \in \mathcal{X}_i\}. \tag{H2.1}$$

它的 Lagrange 函数 $L(x_1,\ldots,x_p,\lambda) = \sum_{i=1}^p \theta_i(x_i) - \lambda^T(\sum_{i=1}^p A_i x_i - b)$ 定义在 $\Omega = \prod_{i=1}^p \mathcal{X}_i \times \Re^m$ 上. 设 $(x_1^*,\ldots,x_p^*,\lambda^*) \in \Omega$ 是 Lagrange 函数的鞍点, 那就像 **C1** 中分析的那样, 鞍点是变分不等式

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

的解点, 其中 $\Omega = \prod_{i=1}^p \mathcal{X}_i \times \Lambda$,

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}, \quad \theta(x) = \sum_{i=1}^p \theta_i(x_i), \quad w = \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A_1^T\lambda \\ \vdots \\ -A_p^T\lambda \\ \sum_{i=1}^p A_i x_i - b \end{pmatrix}. \tag{H2.2b}$$

我们仍以 $\Omega^*$ 表示变分不等式 (H2.2) 的解集, 用统一框架中的预测-校正方法求解凸优化问题 (H2.1).

**[预测]** 从给定的 $w^k$ 出发, 求得预测点 $\tilde{w}^k \in \Omega$, 使其满足

$$\tilde{w}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (w - \tilde{w}^k)^T Q(w^k - \tilde{w}^k), \quad \forall w \in \Omega, \tag{H2.3a}$$

其中 $Q$ 并不要求对称, 但要求 $Q^T + Q$ 的核矩阵是正定的. 这里核心变量为 $w = (x,\lambda)$

**[校正]** 新的迭代点 (校正点) $w^{k+1}$ 由公式

$$w^{k+1} = w^k - M(w^k - \tilde{w}^k) \tag{H2.3b}$$

给出, 其中 $M = Q^{-T}D$. 选择 $D$ 使其满足 $D \succ 0,\ G \succ 0,\ D + G = Q^T + Q$. 这是容易做到的

因为 $M = Q^{-T}D$, 校正也可以通过求解线性方程组 $Q^T(w^{k+1} - w^k) = D(\tilde{w}^k - w^k)$ 实现. (H2.4)

## 2 预测实现的方法 参阅 [39] Handbook of Numerical Analysis, 24 (2023) 511-557.

从给定的 $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ 出发, 通过

$$[\text{预测}]\ \left\{\begin{array}{l}
\tilde{x}_1^k \in \arg\min\{\theta_1(x_1) - x_1^T A_1^T \lambda^k + \frac{1}{2}\beta\|A_1(x_1 - x_1^k)\|^2 \mid x_1 \in \mathcal{X}_1\}; \\
\tilde{x}_2^k \in \arg\min\{\theta_2(x_2) - x_2^T A_2^T \lambda^k + \frac{1}{2}\beta\|A_1(\tilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \mid x_2 \in \mathcal{X}_2\}; \\
\qquad\vdots \\
\tilde{x}_p^k \in \arg\min_{x_p\in\mathcal{X}_p}\{\theta_p(x_p) - x_p^T A_p^T \lambda^k + \frac{1}{2}\beta\|\sum_{j=1}^{p-1} A_j(\tilde{x}_j^k - x_j^k) + A_p(x_p - x_p^k)\|^2\}; \\
\tilde{\lambda}^k = \lambda^k - \beta(\sum_{j=1}^p A_j\tilde{x}_j^k - b)
\end{array}\right. \tag{H2.5}$$

求得 $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \cdots, \tilde{x}_p^k, \tilde{\lambda}^k) \in \Omega$. 这种逐一向前推进的预测也可见 arXiv: 2107.01897v2 [math.OC].

注意到预测 (H2.5) 中对 $i = 1,\ldots,p$, 每个 $x_i$-子问题的最优性条件可以写成

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

生成对偶预测 $\tilde{\lambda}^k$ 的公式可以写成

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

根据以上分析, 利用 (H2.2) 中 $\theta(x)$ 和 $F(w)$ 的定义, 从预测 (H2.5) 我们得到以下的变分不等式

**引理H** 设 $\tilde{w}^k \in \Omega$ 是从给定的 $(A_1x_1^k, A_2x_2^k, \cdots, A_px_p^k, \lambda^k)$ 出发, 由 (H2.5) 生成的预测点. 我们有

$$\tilde{w}^k \in \Omega,\ \ \theta(x) - \theta(\tilde{x}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (w - \tilde{w}^k)^T Q(w^k - \tilde{w}^k),\ \ \forall\, w \in \Omega, \tag{H2.6a}$$

其中

$$Q = \begin{pmatrix}
\beta A_1^T A_1 & 0 & \cdots & 0 & A_1^T \\
\beta A_2^T A_1 & \beta A_2^T A_2 & \ddots & \vdots & A_2^T \\
\vdots & & \ddots & 0 & \vdots \\
\beta A_p^T A_1 & \beta A_p^T A_2 & \cdots & \beta A_p^T A_p & A_p^T \\
0 & 0 & \cdots & 0 & \frac{1}{\beta} I_m
\end{pmatrix}. \tag{H2.6b}$$

### 3 校正求满足 (H2.4) 的 $(A_1x_1^{k+1},\dots,A_px_p^{k+1},\lambda^{k+1})$ 参见 arXiv:2107.01897v2[math.OC].

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

其右端中间部分是 $Q^T+Q$ 的正定的核心矩阵. 同样地, 对矩阵 $Q^T$ 也写出相应的形式

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

将 $Q^T+Q$ 和 $Q^T$ 的核心矩阵分别记成 $\mathcal{Q}^T+\mathcal{Q}$ 和 $\mathcal{Q}^T$, 它们是

$$\mathcal{Q}^T+\mathcal{Q}=\begin{pmatrix} 2\beta I & \beta I & \cdots & \beta I & I\\ \beta I & 2\beta I & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & \beta I & I\\ \beta I & \cdots & \beta I & 2\beta I & I\\ I & \cdots & I & I & \frac{2}{\beta}I\end{pmatrix} \quad 和 \quad \mathcal{Q}^T=\begin{pmatrix} \beta I & \beta I & \cdots & \beta I & 0\\ 0 & \beta I & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & \beta I & 0\\ 0 & \cdots & 0 & \beta I & 0\\ I & \cdots & I & I & \frac{1}{\beta}I\end{pmatrix}. \tag{H2.7}$$

因为 $k$-次迭代的预测是从给定的 $(A_1x_1^k,A_2x_2^k,\dots,A_px_p^k,\lambda^k)$ 开始的, 校正只需要为下一次迭代提供 $(A_1x_1^{k+1},A_2x_2^{k+1},\dots,A_px_p^{k+1},\lambda^{k+1})$. 选 $D\succ 0$ 使满足 $G=Q^T+Q-D\succ 0$ 的途径是无穷的.

我们仅以 $\boldsymbol{D}$ 跟 $\boldsymbol{G}$ 成比例为例做校正. 也就是说, 取 $D=\alpha\big(Q^T+Q\big)$ 和 $G=(1-\alpha)\big(Q^T+Q\big)$, $\alpha\in(0,1)$.

$$D\!=\!\alpha\begin{pmatrix} A_1^T & 0 & \cdots & 0 & 0\\ 0 & A_2^T & \ddots & \vdots & 0\\ \vdots & & \ddots & 0 & \vdots\\ 0 & 0 & \cdots & A_p^T & 0\\ 0 & 0 & \cdots & 0 & I\end{pmatrix}\begin{pmatrix} 2\beta I & \beta I & \cdots & \beta I & I\\ \beta I & 2\beta I & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & \beta I & I\\ \beta I & \cdots & \beta I & 2\beta I & I\\ I & \cdots & I & I & \frac{2}{\beta}I\end{pmatrix}\begin{pmatrix} A_1 & 0 & \cdots & 0 & 0\\ 0 & A_2 & \ddots & \vdots & 0\\ \vdots & & \ddots & 0 & \vdots\\ 0 & 0 & \cdots & A_p & 0\\ 0 & 0 & \cdots & 0 & I\end{pmatrix} \tag{H2.8}$$

满足 $Q^T(v^{k+1}-v^k)=D(\tilde v^k-v^k)$ 的 $(A_1x_1^{k+1},A_2x_2^{k+1},\dots,A_px_p^{k+1},\lambda^{k+1})$ 可以通过

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

求得. 因为 $\mathcal{Q}^{-T}=\begin{pmatrix} \beta I & \beta I & \cdots & \beta I & 0\\ 0 & \beta I & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & \beta I & 0\\ 0 & \cdots & 0 & \beta I & 0\\ I & \cdots & I & I & \frac{1}{\beta}I\end{pmatrix}^{-1}=\begin{pmatrix} \frac{1}{\beta}I & -\frac{1}{\beta}I & 0 & \cdots & 0\\ 0 & \frac{1}{\beta}I & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & -\frac{1}{\beta}I & 0\\ 0 & \cdots & 0 & \frac{1}{\beta}I & 0\\ -I & 0 & \cdots & 0 & \beta I\end{pmatrix}$, 从上式得到

$$\begin{pmatrix} A_1x_1^{k+1}\\ A_2x_2^{k+1}\\ \vdots\\ A_px_p^{k+1}\\ \lambda^{k+1}\end{pmatrix}=\begin{pmatrix} A_1x_1^k\\ A_2x_2^k\\ \vdots\\ A_px_p^k\\ \lambda^k\end{pmatrix}-\alpha\begin{pmatrix} I & -I & 0 & \cdots & 0\\ 0 & \ddots & \ddots & \ddots & \vdots\\ 0 & 0 & I & -I & 0\\ I & \cdots & I & 2I & \frac{1}{\beta}I\\ -\beta I & 0 & \cdots & 0 & I\end{pmatrix}\begin{pmatrix} A_1x_1^k-A_1\tilde x_1^k\\ A_2x_2^k-A_2\tilde x_2^k\\ \vdots\\ A_px_p^k-A_p\tilde x_p^k\\ \lambda^k-\tilde\lambda^k\end{pmatrix} \tag{H2.9}$$

便可以开始新的迭代. 本篇用统一框架 (H2.3), 对求解 VI (H2.2) 的广义 PPA 的一次迭代做了完整的演绎.

# H3-求解多块可分离凸优化问题预设预测核矩阵的分裂收缩算法

## 从设计合理的预测核矩阵出发构造求解多块可分离问题的一簇算法

## 1 优化问题的变分不等式及算法框架

短篇 **H2** 中已经将多块可分离凸优化问题

$$\min\{\textstyle\sum_{i=1}^p \theta_i(x_i) \;\big|\; \sum_{i=1}^p A_i x_i = b,\; x_i\in\mathcal{X}_i\} \tag{H3.1}$$

的 Lagrange 函数的鞍点归结为变分不等式

(VI)
$$w^*\in\Omega,\quad \theta(x)-\theta(x^*)+(w-w^*)^TF(w^*)\ge 0,\quad \forall\, w\in\Omega, \tag{H3.2a}$$

的解点, 其中 $\Omega=\prod_{i=1}^p \mathcal{X}_i\times\Lambda$,

$$x=\begin{pmatrix}x_1\\ \vdots\\ x_p\end{pmatrix},\quad \theta(x)=\sum_{i=1}^p\theta_i(x_i),\quad w=\begin{pmatrix}x_1\\ \vdots\\ x_p\\ \lambda\end{pmatrix},\quad F(w)=\begin{pmatrix}-A_1^T\lambda\\ \vdots\\ -A_p^T\lambda\\ \sum_{i=1}^p A_ix_i-b\end{pmatrix}. \tag{H3.2b}$$

我们用 $\Omega^*$ 表示变分不等式 (H3.2) 的解集, 用统一框架中的预测-校正方法求解变分不等式 (H3.2).

[预测] 从给定的 $w^k$ 出发, 求得预测点 $\tilde w^k\in\Omega$, 使其满足

$$\tilde w^k\in\Omega,\quad \theta(x)-\theta(\tilde x^k)+(w-\tilde w^k)^TF(\tilde w^k)\ge (w-\tilde w^k)^TQ(w^k-\tilde w^k),\quad \forall w\in\Omega, \tag{H3.3a}$$

其中 $Q$ 并不要求对称, 但要求 $Q^T+Q$ 的核矩阵是正定的. 这里核心变量为 $w=(x,\lambda)$

[校正] 新的迭代点 (校正点) $w^{k+1}$ 由公式

$$w^{k+1}=w^k-M(w^k-\tilde w^k) \tag{H3.3b}$$

给出, 其中 $M=Q^{-T}D$. 选择$D$ 使其满足$D\succ 0,\ G\succ 0,\ D+G=Q^T+Q$. 这是容易做到的

因为 $M=Q^{-T}D$, 校正也可以通过求解线性方程组 $Q^T(w^{k+1}-w^k)=D(\tilde w^k-w^k)$ 实现. (H3.4)

若迭代预测能从 $(A_1x_1^k, A_2x_2^k,\ldots,A_px_p^k,\lambda^k)$ 开始, 校正也只需提供 $(A_1x_1^{k+1},A_2x_2^{k+1},\ldots,A_px_p^{k+1},\lambda^{k+1})$.

这样, 我们的首要任务是构造满足 $\mathcal{Q}^T+\mathcal{Q}\succ 0$ 的 $(p+1)\times(p+1)$ 分块核矩阵 $\mathcal{Q}$.

## 2 预测核矩阵的合理构建和预测的实现

在求解变分不等式 (H3.2) 的过程中, 我们以 $w$ 为核心变量. 对 $p$ 个可分离块的凸优化问题, 我们举一些构造简单的预测(核)矩阵的例子. 为了简化矩阵记号, 我们定义 $p\times p$ 分块矩阵和 $p\times 1$ 分块矩阵:

$$\mathcal{L}=\begin{pmatrix}I&0&\cdots&0\\ I&I&\ddots&\vdots\\ \vdots&\ddots&\ddots&0\\ I&\cdots&I&I\end{pmatrix},\quad \mathcal{I}=\begin{pmatrix}I&0&\cdots&0\\ 0&I&\ddots&\vdots\\ \vdots&\ddots&\ddots&0\\ 0&\cdots&0&I\end{pmatrix}\quad \text{和}\quad \mathcal{E}=\begin{pmatrix}I\\ I\\ \vdots\\ I\end{pmatrix}. \tag{H3.5}$$

利用上面的定义, 有

$$\mathcal{L}^T+\mathcal{L}=\mathcal{I}+\mathcal{E}\mathcal{E}^T. \tag{H3.6}$$

我们可以构造多种多样的满足 $\mathcal{Q}^T+\mathcal{Q}\succ 0$ 的预测(核)矩阵, 并让预测容易实现.

1. 取 $\mathcal{Q}_1=\begin{pmatrix}\mathcal{L}&\mathcal{E}\\ 0&I\end{pmatrix}$, 则有 $\mathcal{Q}_1^T+\mathcal{Q}_1=\begin{pmatrix}\mathcal{I}+\mathcal{E}\mathcal{E}^T&\mathcal{E}\\ \mathcal{E}^T&2I\end{pmatrix}=\begin{pmatrix}\mathcal{I}&0\\ 0&I\end{pmatrix}+\begin{pmatrix}\mathcal{E}\\ I\end{pmatrix}\begin{pmatrix}\mathcal{E}^T,&I\end{pmatrix}$.

2. 取 $\mathcal{Q}_2=\begin{pmatrix}\mathcal{L}&0\\ -\mathcal{E}^T&I\end{pmatrix}$, 则有 $\mathcal{Q}^T+\mathcal{Q}=\begin{pmatrix}\mathcal{I}+\mathcal{E}\mathcal{E}^T&-\mathcal{E}\\ -\mathcal{E}^T&2I\end{pmatrix}=\begin{pmatrix}\mathcal{I}&0\\ 0&I\end{pmatrix}+\begin{pmatrix}\mathcal{E}\\ -I\end{pmatrix}\begin{pmatrix}\mathcal{E}^T,&-I\end{pmatrix}$.

3. 取 $\mathcal{Q}_3=\begin{pmatrix}\mathcal{L}&\mathcal{E}\\ \mathcal{E}^T&\frac52 I\end{pmatrix}$, 则有 $\mathcal{Q}_3^T+\mathcal{Q}_3=\begin{pmatrix}\mathcal{I}+\mathcal{E}\mathcal{E}^T&2\mathcal{E}\\ 2\mathcal{E}^T&5I\end{pmatrix}=\begin{pmatrix}\mathcal{I}&0\\ 0&I\end{pmatrix}+\begin{pmatrix}\mathcal{E}\\ 2I\end{pmatrix}\begin{pmatrix}\mathcal{E}^T,&2I\end{pmatrix}$.

4. 取 $\mathcal{Q}_4=\begin{pmatrix}\mathcal{L}&-\mathcal{E}\\ -\mathcal{E}^T&\frac52 I\end{pmatrix}$, 则有 $\mathcal{Q}_4^T+\mathcal{Q}_4=\begin{pmatrix}\mathcal{I}+\mathcal{E}\mathcal{E}^T&-2\mathcal{E}\\ -2\mathcal{E}^T&5I\end{pmatrix}=\begin{pmatrix}\mathcal{I}&0\\ 0&I\end{pmatrix}+\begin{pmatrix}-\mathcal{E}\\ 2I\end{pmatrix}\begin{pmatrix}-\mathcal{E}^T,&2I\end{pmatrix}$.

5. 取 $\mathcal{Q}_5=\begin{pmatrix}\mathcal{L} & \mathcal{E}\\ -\mathcal{E}^T & I\end{pmatrix}$， 则有 $\mathcal{Q}_5^T+\mathcal{Q}_5=\begin{pmatrix}\mathcal{I}+\mathcal{E}\mathcal{E}^T & 0\\ 0 & 2I\end{pmatrix}=\begin{pmatrix}\mathcal{I} & 0\\ 0 & I\end{pmatrix}+\begin{pmatrix}\mathcal{E}\\ -I\end{pmatrix}\begin{pmatrix}\mathcal{E}^T, & -I\end{pmatrix}$.

这类矩阵经适当调比后仍然是合格的核矩阵, 以 5. 中的核矩阵 $\mathcal{Q}_5$ 为例, 调比后的核矩阵为

$$\mathcal{Q}=\begin{pmatrix}\beta\mathcal{L} & \mathcal{E}\\ -\mathcal{E}^T & \frac{1}{\beta}I\end{pmatrix},\qquad \mathcal{Q}^T+\mathcal{Q}=\begin{pmatrix}\beta\mathcal{I}+\beta\mathcal{E}\mathcal{E}^T & 0\\ 0 & \frac{2}{\beta}I\end{pmatrix}. \tag{H3.7}$$

**我们以** (H3.7) **中的核矩阵为例**, 阐释如何构造预测方法. 首先, 对应于 (H3.7) 中的 $\mathcal{Q}$, 相应的预测矩阵为

$$Q=\begin{pmatrix}\beta A_1^TA_1 & 0 & \cdots & 0 & A_1^T\\ \beta A_2^TA_1 & \beta A_2^TA_2 & \ddots & \vdots & A_2^T\\ \vdots & & \ddots & 0 & \vdots\\ \beta A_p^TA_1 & \beta A_p^TA_2 & \cdots & \beta A_p^TA_p & A_p^T\\ -A_1 & -A_2 & \cdots & -A_p & \frac{1}{\beta}I_m\end{pmatrix}. \tag{H3.8}$$

根据变分不等式 (H3.2), 统一框架中 (H3.3) 预测的表达式, 和预测矩阵 (H3.8), 预测 (H3.3a) 的具体内容写开来就是：对 $i=1,\ldots,p$, 每个 $x_i$-子问题求解需要实现的变分不等式形式是

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

生成对偶预测 $\tilde{\lambda}^k$ 的公式则可以写成

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

上面带下波纹线的部分组装在一起就是 $F(\tilde{w}^k)$. 根据以上分析, 预测可以通过下面的 (H3.9) 完成.

从给定的 $(A_1x_1^k,A_2x_2^k,\cdots,A_px_p^k,\lambda^k)$ 出发, 通过

$$\text{[预测]}\begin{cases}\tilde{x}_1^k\in\arg\min\big\{\theta_1(x_1)-x_1^TA_1^T\lambda^k+\frac{1}{2}\beta\|A_1(x_1-x_1^k)\|^2\mid x_1\in\mathcal{X}_1\big\};\\ \tilde{x}_2^k\in\arg\min\big\{\theta_2(x_2)-x_2^TA_2^T\lambda^k+\frac{1}{2}\beta\|A_1(\tilde{x}_1^k-x_1^k)+A_2(x_2-x_2^k)\|^2\mid x_2\in\mathcal{X}_2\big\};\\ \qquad\vdots\\ \tilde{x}_p^k\in\arg\min_{x_p\in\mathcal{X}_p}\big\{\theta_p(x_p)-x_p^TA_p^T\lambda^k+\frac{1}{2}\beta\|\sum_{j=1}^{p-1}A_j(\tilde{x}_j^k-x_j^k)+A_p(x_p-x_p^k)\|^2\big\};\\ \tilde{\lambda}^k=\lambda^k-\beta\big(\sum_{j=1}^p A_jx_j^k-b\big).\end{cases} \tag{H3.9}$$

求得 $\tilde{w}^k=(\tilde{x}_1^k,\tilde{x}_2^k,\cdots,\tilde{x}_p^k,\tilde{\lambda}^k)\in\Omega$.

因此也有了相应的 $(A_1\tilde{x}_1^k,A_2\tilde{x}_2^k,\cdots,A_p\tilde{x}_p^k,\tilde{\lambda}^k)$

## 3 选定矩阵 $\mathcal{D}$ 和实现校正

对 (H3.7) 中的 $\mathcal{Q}^T+\mathcal{Q}$, 选择满足 $0\prec\mathcal{D}\prec\mathcal{Q}^T+\mathcal{Q}$ 的 $\mathcal{D}$ 有多种选法. 如果取 $\alpha\in(0,1)$,

$$\mathcal{D}=\begin{pmatrix}\alpha\beta\mathcal{I} & 0\\ 0 & \frac{1}{\beta}I\end{pmatrix},\quad \mathcal{G}=\begin{pmatrix}(1-\alpha)\beta\mathcal{I}+\beta\mathcal{E}\mathcal{E}^T & 0\\ 0 & \frac{1}{\beta}I\end{pmatrix}\quad \text{则有}\quad \mathcal{D}+\mathcal{G}=\mathcal{Q}^T+\mathcal{Q}. \tag{H3.10}$$

如此正定的 $\mathcal{D}$ 和 $\mathcal{G}$ 满足分拆要求. 对 (H3.7) 中的 $\mathcal{Q}$, 用普通的线性代数知识计算得到

$$\mathcal{Q}^{-T}=\begin{pmatrix}\frac{1}{\beta}\big(\mathcal{L}^{-T}-\frac{1}{2}\mathcal{L}^{-T}\mathcal{E}\mathcal{E}^T\mathcal{L}^{-T}\big) & \frac{1}{2}\mathcal{L}^{-T}\mathcal{E}\\ -\frac{1}{2}\mathcal{E}^T\mathcal{L}^{-T} & \frac{1}{2}\beta I\end{pmatrix}. \tag{H3.11}$$

由于

$$\mathcal{L}^{-T}=\begin{pmatrix}I & -I & 0 & 0\\ 0 & I & \ddots & 0\\ \vdots & \ddots & \ddots & -I\\ 0 & \cdots & 0 & I\end{pmatrix},\quad \mathcal{L}^{-T}\mathcal{E}=\begin{pmatrix}0\\ \vdots\\ 0\\ I\end{pmatrix},\quad \text{所以}\quad \mathcal{Q}^{-T}=\begin{pmatrix}\frac{1}{\beta}I & -\frac{1}{\beta}I & 0 & \cdots & 0\\ 0 & \ddots & \ddots & \ddots & \vdots\\ 0 & 0 & \frac{1}{\beta}I & -\frac{1}{\beta}I & 0\\ -\frac{1}{2\beta}I & \cdots & 0 & \frac{1}{\beta}I & \frac{1}{2}I\\ -\frac{1}{2}I & 0 & \cdots & 0 & \frac{1}{2}\beta I\end{pmatrix}.$$

利用 $\mathcal{Q}^{-T}$ 和 $\mathcal{D}$, 从下面的 (H3.12) 直接得到 $(A_1x_1^{k+1},A_2x_2^{k+1},\ldots,A_px_p^{k+1},\lambda^{k+1})$.

$$\text{[校正]}\begin{pmatrix}A_1x_1^{k+1}\\ A_2x_2^{k+1}\\ \vdots\\ A_px_p^{k+1}\\ \lambda^{k+1}\end{pmatrix}=\begin{pmatrix}A_1x_1^k\\ A_2x_2^k\\ \vdots\\ A_px_p^k\\ \lambda^k\end{pmatrix}-\begin{pmatrix}\alpha I & -\alpha I & 0 & \cdots & 0\\ 0 & \ddots & \ddots & \ddots & \vdots\\ 0 & 0 & \alpha I & -\alpha I & 0\\ -\frac{1}{2}\alpha & \cdots & 0 & \alpha I & \frac{1}{2\beta}I\\ -\frac{1}{2}\alpha\beta & 0 & \cdots & 0 & \frac{1}{2}I\end{pmatrix}\begin{pmatrix}A_1x_1^k-A_1\tilde{x}_1^k\\ A_2x_2^k-A_2\tilde{x}_2^k\\ \vdots\\ A_px_p^k-A_p\tilde{x}_p^k\\ \lambda^k-\tilde{\lambda}^k\end{pmatrix}. \tag{H3.12}$$

校正式 (H3.12) 右端中的 $(p+1)\times(p+1)$ 分块矩阵就是 $\mathcal{Q}^{-T}\mathcal{D}$. $\mathcal{Q}^{-T}$ 和 $\mathcal{D}$ 分别出自 (H3.11) 和 (H3.10).

## References


[1] 何炳生, 从变分不等式的投影收缩算法到凸优化的分裂收缩算法, 高等学校计算数学学报(2016) 38: 74-96.

[2] 何炳生, 我和乘子交替方向法20 年. 运筹学学报(2018) 22: 1-31. DOI: 10.15960/j.cnki.issn.1007-6093.2018.01.001

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

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

[5] B. S. He, PPA-like contraction methods for convex optimization: a framework using variational inequality approach, J. Oper. Res. Soc. China (2015) 3: 391-420. DOI 10.1007/s40305-015-0108-9

[6] B.S. He, and X. M. Yuan, A class of ADMM-based algorithms for three-block separable convex programming. Comput. Optim. Appl. (2018) 70: 791-826. DOI 10.1007/s10589-018-9994-1

[7] 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.

[8] 何炳生. 凸优化和单调变分不等式的分裂收缩算法, 二十讲系列讲义. 见南京大学数学学院何炳生个人主页 maths.nju.edu.cn/˜hebma.

[9] 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: 649-679, 2012. DOI 10.1007/s10589-010-9372-0

[10] 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. DOI 10.1007/s10589-010-9373-z

[11] 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. Sci Sin Math (2018) 48: 255 – 272. DOI: 10.1360/N012017-00034

[12] B.S. He and H. Yang, Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities, Operations Research Letters (1998) 3: 151-161.

[13] B.S. He, H. Yang, and S.L. Wang, Alternating directions method with self- adaptive penalty parameters for monotone variational inequalities, JOTA (2000) 106: 349-368.

[14] B.S. He, L.Z. Liao, D.R. Han, and H. Yang, A new inexact alternating directions method for monotone variational inequalities, Math. Progr, (2002) 92: 103-118. DOI 10.1007/s101070100280

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

[16] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, Foundations and Trends in Machine Learning (2010) 3: 1-122.

[17] 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. DOI 10.1137/110836936

[18] B.S. He and X.M. Yuan, On non-ergodic convergence rate of Douglas-Rachford alternating directions method of multipliers, Numerische Mathematik (2015) 130: 567-577. DOI 10.1007/s00211-014-0673-6

[19] G.Y. Gu, B. S. He and J. F. Yang, Inexact alternating-direction-based contraction methods for separable linearly constrained convex optimization. JOTA (2014) 163: 105-129. DOI 10.1007/s10957-013-0489-z

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

[21] M.H. Xu, Proximal alternating directions method for structured variational inequalities, JOTA (2007) 134: 107-117. DOI 10.1007/s10957-007-9192-2

[22] B.S. He, F. Ma and X.M. Yuan, Optimally linearizing the alternating direction method of multipliers for convex programming, COA (2020) 75: 361-388. DOI 10.1007/s10589-019-00152-3

[23] 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 (2012) 5: 119-149. DOI 10.1137/100814494

[24] B.S. He, X.M. Yuan and W.X. Zhang, A customized proximal point algorithm for convex minimization with linear constraints, COA (2013) 56: 559-572. DOI 10.1007/s10589-013-9564-5

[25] 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 (2014) 59: 135-161. DOI 10.1007/s10589-013-9616-x

[26] 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. Sci Sin Math, 2012, 42(5): 515 – 525, DOI 10.1360/012011-1049

[27] 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.

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

[29] S. J. Xu, A dual-primal balanced augmented Lagrangian method for lineraly constrained convex programming, J. Appl. Math. and Computing 69 (2023), 1015-1035. DOI 10.1007/s12190-022-01779-y

[30] X.J. Cai, G.Y. Gu, B.S. He and X.M. Yuan, A proximal point algorithms revisit on the alternating direction method of multipliers, Science China Math. (2013) 56: 2179-2186. DOI 10.1007/s11425-013-4683-0

[31] B.S. He, H. Liu, Z.R. Wang and X. M. Yuan, A strictly Peaceman-Rachford splitting method for convex programming, SIAM J. Optim. (2014) 24:1011-1040. DOI. 10.1137/13090849X

[32] E.X. Fang, B.S. He, H. Liu and X. M. Yuan, Generalized alternating direction method of multipliers: new theoretical insights and applications, Mathematical Programming Computation, 7 (2015) 149-187. DOI 10.1007/s12532-015-0078-2

[33] 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. DOI 10.1007/s10107-014-0826-5

[34] 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. DOI 10.1137/110822347

[35] B.S. He, M. Tao and X.M. Yuan, A splitting method for separable convex programming, IMA J. Numerical Analysis (2015) 31: 394-426. DOI 10.1093/imanum/drt060

[36] 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. DOI 10.15960/j.cnki.issn.1007-6093.2015.03.008

[37] B.S. He, M. Tao and X. M. Yuan, Convergence rate analysis for the alternating direction method of multipliers with a substitution procedure for separable convex programming, Math. OR (2017) 42: 662-691.

[38] 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.

[39] 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: 511-557, 2023. also see arXiv: 2107.01897v2 [math.OC].

[40] 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]

[41] 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.

[42] S. Becker, The Chen-Teboulle algorithm is the proximal point algorithm, arXiv: 1908. 03633[math.OC].

[43] A. Chambolle and T Pock, On the ergodic convergence rates of a first-order primal – dual algorithm， Math. Program., A (2016) 159: 253-287. DOI 10.1007/s10107-015-0957-3

[44] T. Zhu and Z.G Yu, A simple proof for some important properties of the projection mapping. Math. Inequal. Appl., (2004) 7: 453-456.

[45] 何炳生, 凸优化的分裂收缩算法. 北京：科学出版社, 2025. 4. ISBN 978-7-03-080804-2

# Splitting and contraction methods for convex optimization

If I have only two pages, then I will write the following © Bingsheng He

## 1 Convex optimization and its related variational inequality

We consider the linearly constrained convex optimization

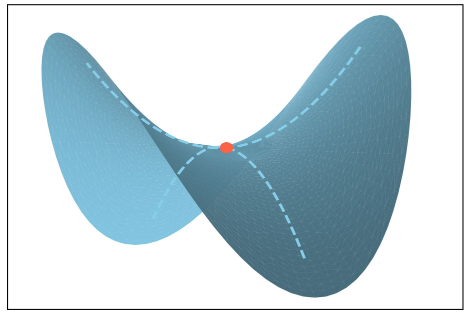
$$\min\{\theta(u) \mid \mathcal{A}u = b,\ u \in \mathcal{U}\}. \tag{1}$$

The Lagrangian function of the problem (1) is

$$L(u,\lambda) = \theta(u) - \lambda^T(\mathcal{A}u - b), \tag{2}$$

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 Lagrangian function (2), if

$$L_{u\in\mathcal{U}}(u,\lambda^*) \ge L(u^*,\lambda^*) \ge L_{\lambda\in\Re^m}(u^*,\lambda). \tag{3}$$



The two inequalities in (3) can be written as $(u^*,\lambda^*) \in \mathcal{U} \times \Re^m$,

$$\begin{cases} L(u,\lambda^*) - L(u^*,\lambda^*) \ge 0, & \forall u \in \mathcal{U}, \\ L(u^*,\lambda^*) - L(u^*,\lambda) \ge 0, & \forall \lambda \in \Re^m. \end{cases} \iff \begin{cases} \theta(u) - \theta(u^*) + (u-u^*)^T(-\mathcal{A}^T\lambda^*) \ge 0, & \forall u \in \mathcal{U}, \\ (\lambda - \lambda^*)^T(\mathcal{A}u^* - b) \ge 0, & \forall \lambda \in \Re^m. \end{cases} \tag{4}$$

Combining the above two inequalities, the saddle point is equivalent to the solution of the VI :

$$\text{(VI)} \qquad w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \ge 0, \quad \forall\, w \in \Omega, \tag{5}$$

where
$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \qquad F(w) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}u - b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{U} \times \Re^m. \tag{6}$$

Please notice that $(w - \tilde{w})^T(F(w) - F(\tilde{w})) \equiv 0$. The problem (1) is translated to VI (5).

For the multi-block separable convex optimization, we take the three-block problem

$$\min\{\theta_1(x) + \theta_2(y) + \theta_3(z) \,|\, Ax + By + Cz = b, x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\} \tag{7}$$

as an example. Its corresponding variational inequality has the form (5), where $\Omega = \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \Re^m$, and

$$u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \theta(u) = \theta_1(x) + \theta_2(y) + \theta_3(z), \quad w = \begin{pmatrix} x \\ y \\ z \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^T\lambda \\ -B^T\lambda \\ -C^T\lambda \\ Ax + By + Cz - b \end{pmatrix}. \tag{8}$$

For the $F(w)$ in (8), we still have $(w - \tilde{w})^T(F(w) - F(\tilde{w})) \equiv 0$, for any $w$ and $\tilde{w}$ in the space containing $\Omega$.

## 2 Algorithmic unified framework for monotone variational inequalities

We focus on how to solve the variational inequality (5), following is our algorithmic framework.

Algorithms in a unified framework (Each iteration of the method consists of a prediction and a correction)

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

$$\tilde{w}^k \in \Omega, \quad \theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \ge (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \tag{9a}$$

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

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

**[Correction Step].** Find the correction matrix $M$ which satisfied (10). New iteration $v^{k+1}$ is given by

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k). \tag{9b}$$

Convergence Conditions (It is easy to find the matrix $\boldsymbol{M}$ which satisfies the conditions, see the details in §4)

**[Convergence Conditions].** For the prediction matrix $Q$ in (9a) and the correction matrix $M$ in (9b), there is a positive definite matrix $H$, such that

$$HM = Q \quad \text{and} \quad G = Q^T + Q - M^T HM \succ 0. \tag{10}$$

# 3 Convergence proof of the methods in the algorithmic framework

**Theorem A**. For solving the VI (5), let $\{v^k\}$, $\{\tilde{w}^k\}$ be the sequences generated by (9). If the conditions (10) are satisfied, then we have

$$\tilde{w}^k \in \Omega, \quad \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \ge \tfrac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\tfrac{1}{2}\|v^k-\tilde{v}^k\|_G^2, \quad \forall w\in\Omega. \tag{11}$$

and

$$\|v^{k+1}-v^*\|_H^2 \le \|v^k-v^*\|_H^2 - \|v^k-\tilde{v}^k\|_G^2, \quad \forall v^*\in\mathcal{V}^*. \tag{12}$$

**Proof**. Treating the term $Q(v^k-\tilde{v}^k)$ in the RHS of (9a) by using $Q=HM$ (see (10)) and the correction formula (9b), we obtain $Q(v^k-\tilde{v}^k)=HM(v^k-\tilde{v}^k)=H(v^k-v^{k+1})$. Thus we get

$$\tilde{w}^k\in\Omega, \quad \theta(u)-\theta(\tilde{u}^k)+(w-\tilde{w}^k)^T F(\tilde{w}^k) \ge (v-\tilde{v}^k)^T H(v^k-v^{k+1}), \quad \forall w\in\Omega. \tag{13}$$

Applying the identity

$$(a-b)^T H(c-d) = \tfrac{1}{2}\big(\|a-d\|_H^2-\|a-c\|_H^2\big)+\tfrac{1}{2}\big(\|b-c\|_H^2-\|b-d\|_H^2\big) \tag{14}$$

to the RHS of (13) with $a=v$, $b=\tilde{v}^k$, $c=v^k$ and $d=v^{k+1}$, we obtain

$$(v-\tilde{v}^k)^T H(v^k-v^{k+1}) = \tfrac{1}{2}\big(\|v-v^{k+1}\|_H^2-\|v-v^k\|_H^2\big)+\tfrac{1}{2}\big(\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2\big). \tag{15}$$

To the second part of the RHS of (15), by using $HM=Q$ and $2v^TQv=v^T(Q^T+Q)v$, it follows that

$$\begin{aligned}\|v^k-\tilde{v}^k\|_H^2-\|v^{k+1}-\tilde{v}^k\|_H^2 &\overset{(9b)}{=} \|v^k-\tilde{v}^k\|_H^2-\|(v^k-\tilde{v}^k)-M(v^k-\tilde{v}^k)\|_H^2\\ &= (v^k-\tilde{v}^k)^T(Q^T+Q-M^THM)(v^k-\tilde{v}^k) \overset{(10)}{=} \|v^k-\tilde{v}^k\|_G^2.\end{aligned} \tag{16}$$

Substituting (16) in (15), and then in (13), we get the assertion (11) directly. Setting the $w\in\Omega$ in (11) by any fixed $w^*$, then using $(\tilde{w}^k-w^*)^T F(\tilde{w}^k)=(\tilde{w}^k-w^*)^T F(w^*)$ and $\theta(\tilde{u}^k)-\theta(u^*)+(\tilde{w}^k-w^*)^T F(w^*)\ge 0$, we obtain (12) and the theorem is completely proved. □

This theorem is proved under weak conditions: $Q^T+Q\succeq 0,\ H\succeq 0,\ HM=Q,\ G=Q^T+Q-M^THM\succeq 0.$

Assertion (11) is useful for the convergence rate proof of ADMM, see SIAM Numer. Anal. 2012, 50:700-709.

# 4 The equivalent convergence conditions and the generalized PPA

Under the condition that the prediction matrix $Q$ is nonsingular, it is easy to construct the correction matrix $M$ which satisfies the convergence conditions (10). In fact, because $Q^T+Q\succ 0$, we can take

$$D\succ 0, \qquad G\succ 0 \qquad \text{and} \qquad D+G=Q^T+Q. \tag{17}$$

Afterwards, we let

$$HM=Q \qquad \text{and} \qquad M^THM=D. \tag{18}$$

Because $\left\{\begin{array}{r} HM=Q,\\ M^THM=D.\end{array}\right. \Leftrightarrow \left\{\begin{array}{r} Q^TM=D,\\ HM=Q.\end{array}\right. \Leftrightarrow \left\{\begin{array}{l} M=Q^{-T}D,\\ H=QD^{-1}Q^T,\end{array}\right.$ we get the matrices $M$, $H$ and $G$,

which satisfy the conditions (10). There are infinite combinations of $D$ and $G$ which satisfy conditions (17).

Choosing matrix $D$ that satisfies condition (17), we get $M=Q^{-T}D$, and $H=QD^{-1}Q^T$ is positive definite. The correction $v^{k+1}=v^k-M(v^k-\tilde{v}^k)$ can be achieved by solving $Q^T(v^{k+1}-v^k)=D(\tilde{v}^k-v^k)$.

**The generalized PPA by choosing a special $D$.** We can take a special pair of $D$ and $G$ in (17) by

$$D=G=\tfrac{1}{2}(Q^T+Q). \tag{19}$$

In this case, $M=\frac{1}{2}Q^{-T}(Q^T+Q)$. Because $D=G$, the contractive inequality (12) becomes

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

Moreover, since $D=M^THM$ (see (18)) and $M(v^k-\tilde{v}^k)=v^k-v^{k+1}$ (see (9b)), it follows that

$$\|v^{k+1}-v^*\|_H^2 \le \|v^k-v^*\|_H^2-\|v^k-v^{k+1}\|_H^2, \quad \forall v^*\in\mathcal{V}^*. \tag{21}$$

The inequality (21) is just the main convergence result of the classical PPA, a favorable formula ! Because the each iteration consists of a prediction and a correction, we call it as a generalized PPA.

**In practice, in the generalized PPA, we suggest to take $D=\alpha(Q^T+Q)$ and $\alpha\in[0.5,1)$.**

# 对自己凸优化研究工作的一点认识

我做的是优化方法，主要兴趣是单调变分不等式和凸优化的求解。线性约束的凸优化问题，是科学与工程计算中经常遇到的一类优化问题，因为它具有组合方面的性质，一般没有直接方法，只能用迭代法求解。

邻近点算法 (PPA) 是凸优化的经典迭代算法。它的中心思想就是步步为营，稳扎稳打。

求解单块线性约束凸优化问题，增广拉格朗日乘子法 (ALM) 是公认的有效算法之一，ALM 恰是拉格朗日乘子 $\lambda$ 的邻近点算法。

处理极小–极大问题的混合梯度法 (PDHG)，人们想当然以为是正确的，是我们指出它不能保证收敛，并提出改造策略，使之成为单调变分不等式特定矩阵范数意义下的邻近点算法。由于这类算法在图像计算领域应用广泛，很快就引起一些欧美学者的注意，得到“Elegant”和“Very Simple yet Powerful”这样的赞誉！

处理两个可分离块的凸优化问题，往往需要利用其可分离的性质，否则会无从下手。五十年前人们提出的交替方向法 (ADMM)，最近 15 年才发现其重要价值，并变得如此风靡。我们从 1997 年开始做 ADMM 的研究，国内算最早，国际上也领先。交替方向法实质上是松弛了的增广拉格朗日乘子法，是变量 $y$ 和乘子 $\lambda$ 在某个特定矩阵范数下的邻近点算法。

有二就有三。然而，对三个可分离块的凸优化问题，采用对交替方向法继续松弛的方法去求解，有例子说明不能保证收敛，那就需要另想办法。

我们提出的算法统一框架，不仅涵盖了邻近点算法 (PPA)，增广拉格朗日乘子法 (ALM) 和交替方向法 (ADMM) 这些耳熟能详的经典算法，还为多块可分离凸优化问题的求解提供了一簇算法，关键技术是在原有的逐步推进的基础上加了个简单的校正。我们称这类方法为预测-校正的 ADMM 类分裂收缩算法，一是因为它保留了 ADMM 的收敛性质，二是又可以把已有的 ADMM 方法中的每步迭代故意拆解成预测和校正去看待。这样做，还让我们进一步发现 ADMM 的一些容易证明的漂亮性质。

预测-校正，是计算数学中的常用思想方法。在优化算法领域，我比较自觉地用校正策略。说起来，那还得感谢文化大革命期间的那段务农经历！当年的生产队，除了水稻和麦子，还有些副业。栽桑，要剪枝; 种瓜，要掐蔓。这样的活，我干了五、六年。那剪枝掐蔓，不就是校正！？学过和干过的，即便在其他领域，需要的时候，想法也会自然冒出来。

我们提出的这类方法的每步迭代，包含预测和校正两部分。用它处理多块的线性约束凸优化问题，有点像高斯消去法求解线性方程组。预测，相当于高斯消去法中自上而下的按序消元；而校正，则更像高斯消去法中自下往上的回代。

麻省理工学院教授 G. Strang 有本名为“Linear Algebra and its Applications”的线性代数教科书，配合这本教材，他还有个讲课视频。记得 Strang 在他的视频里有段话的大意是：高斯消去法这么重要，但高斯想出这个方法大概只花了他一生中的十分钟时间。

我们在优化方法中做出这一套思想体系上跟高斯消去法有联系的工作，前前后后大概花了 10 年。想起自己在美国数学评论的《数学谱系》中是 Gauss 的第 11 代，工作中还用到了高斯消去法的思想，自然尤感欣慰！

总结自己一生取得的学术成果 (最好称收获) 的原因，我给出了下面二十八个字：

**中学的数理基础，必要的社会实践，普通的大学数学，一般的优化原理。**

这里的一层意思是要告诉他人，听懂我的报告或者读懂我的论文著作所需要的基础知识，读了两年大学的理科学生一般都是具备的; 而一些想法的由来，跟个人经历是不无关系的。

另一层意思是要对自己学习和成长道路上四个阶段给过我帮助的人们表示感谢: 十二年完整的中小学教学给我打下了良好的基础; 十年多的务农期间，朋友的帮助，师长的鼓励，让我没有选择放弃; 恢复高考，我上了大学，老师的教诲，让我没有因为数学是年轻人的学问而气馁; 考上研究生后，何旭初先生的推荐和德国导师 Stoer 教授因势利导的指导，是我学成之后能独立地做些研究的关键。

何炳生　2025 年 5 月

# 凸优化分裂收缩算法统一框架的主要数学公式和收敛结果

线性约束凸优化问题的Lagrange函数的鞍点跟变分不等式

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

的解等价. 为求解这个变分不等式, 我们设计了下面的 **预测-校正算法统一框架**.

**[预测]** 第 $k$ 步迭代从给定的核心变量 $v^k$(向量 $v$ 可以是 $w$ 本身, 也可以是向量 $w$ 的部分分量)开始, 求得预测点 $\tilde{w}^k$, 使得

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

成立. 当矩阵 $Q^T + Q$ 正定的时候, 我们称 (2) 为一个合格的预测.

我们称 $Q$ 为预测矩阵. 如果 (2) 中 $v^k = \tilde{v}^k$, $\tilde{w}^k$ 相当于 (1) 中的 $w^*$, 就是问题 (1) 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

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

我们称 (3) 式中的 $M$ 为校正矩阵. 对由 (2) 预测, (3) 校正的求解变分不等式 (1) 的迭代方法, 有下面的收敛性条件.

**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:

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

由 (2) 预测, (3) 校正的求解问题 (1) 的迭代方法, 若收敛性条件 (4) 满足, 则有下面的收敛关键不等式

**迭代序列满足的收敛关键不等式**

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^{k+1} - v^*\|_H^2 - \|v^k - v^{k+1}\|_G^2. \tag{5}$$

对于一个确定的合格的预测, 选择符合收敛条件的校正矩阵 $M$ 并不神秘, 例如

合格预测 (2) 中的预测矩阵 $Q$ 满足 $Q^T + Q \succ 0$, 我们可以选 $D$, 使其满足

$$0 \prec D \prec Q^T + Q, \qquad \text{然后取} \qquad M = Q^{-T} D \tag{5}$$

做 (3) 那样的校正. 对这样选择的 $D$ 和 $M$, 记 $H = QD^{-}Q^T$ 就会有

$$HM = Q \quad \text{和} \quad M^T HM = D, \quad \text{因此} \quad G = Q^T + Q - D \succ 0. \tag{6}$$

我们在收敛性分析中要求存在正定的 $H$, 计算中不需要显式的 $H$. 由于 $M = Q^{-T}D$, 校正 $v^{k+1} = v^k - M(v^k - \tilde{v}^k)$ 也可以通过求解方程组 $Q^T(v^{k+1} - v^k) = D(\tilde{v}^k - v^k)$ 实现.