凸优化和单调变分不等式的收缩算法
— 统一框架与应用

前言目录与简要说明

南京大学数学系　　何炳生

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

一个科学家最大的本领就在于化复杂为简单, 用简单的方法去解决复杂的问题。 —— 冯康

✠ 先贤名言, 铭记在心。纵然可以没有本领, 也不迷失价值标准 ✠

爱美之心, 人皆有之。数学往往被人认为是枯燥的, 计算更是被人看作是繁琐的。领悟了数学之美的数学工作者, 他的职业生涯才可能是充满乐趣的。

✠ 世上三百六十行, 应是同一道理 ✠

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

✠ 发现了其中的数学之美, 研究才变得满怀激情和欲罢不能 ✠

This lecture series study the algorithms for the following problems

- Convex Optimization. $\min\{f(x)|x\in\mathcal{X}\}$, f convex function, $\mathcal{X}$ convex set.
- Variational Inequality. $$u\in\Omega,\quad (u'-u)^TF(u)\ge 0,\quad \forall u'\in\Omega$$
- $\min$-$\max$ problem $$\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\{\mathcal{L}(x,y)=\theta_1(x)-y^TAx-\theta_2(y)\}$$
- Constrained Convex Optimization $$\min\{\theta(x)|Ax=b\,(\text{or}\ge b),\ x\in\mathcal{X}\}$$
- Separable COP $$\min\{\theta_1(x)+\theta_2(y)|Ax+By=b,x\in\mathcal{X},y\in\mathcal{Y}\}$$
- Multi-blocks separable convex optimization problems $$\min\{\theta_1(x)+\theta_2(y)+\theta_3(z)|Ax+By+Cz=b,x\in\mathcal{X},y\in\mathcal{Y},z\in\mathcal{Z}\}$$

变分不等式和邻近点算法是我们的两大法宝.

变分不等式(VI) 是瞎子爬山的数学表达形式

邻近点算法(PPA) 是步步为营 稳扎稳打的求解方法.

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

中学的数理基础	必要的社会实践
普通的大学数学	一般的优化原理

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

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

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

投影收缩算法的基本操作是通过投影

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

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

这一部分的最后两讲研究算法的收敛速率, 初学者可以跳过这些内容.

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

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

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

引进 Lagrange 乘子 λ, 将求 Lagrange 函数

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

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

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

的解, 其中

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

凸优化和单调变分不等式的收缩算法 — 目录

第一部分: 单调变分不等式的求解方法

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

第八讲 基于增广 Lagrange 乘子法的 PPA 收缩算法

第九讲 基于LVI -PC 方法求解复合凸优化的收缩算法

第十讲 基于梯度投影求解凸优化的收缩算法和下降算法

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

第十一讲 结构型优化的交替方向法(ADMM)

第十二讲 线性化的交替方向收缩算法 (Linearized ADMM)

第十三讲 定制 PPA 意义下的交替方向法及其线性化方法

第十四讲 自变量 x-y 地位平等的对称型化交替方向法

第十五讲 在统一框架下研究交替方向法的收敛速率

第四部分: 多个可分离算子凸优化问题带简单校正的分裂方法

- **第十六讲** 三块可分离凸优化问题的平行分裂 ALM 乘子法
- **第十七讲** 三块可分离凸优化问题的略有改动的交替方向法
- **第十八讲** 多块可分离凸优化带高斯回代的交替方向算法
- **第十九讲** 多块可分离凸优化问题部分平行正则化的交替方向法
- **第二十讲** 变分不等式意义下凸优化分裂收缩算法的统一框架

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

懂些变分不等式的基本概念, 对凸优化收缩算法的设计和收敛性分析很有帮助.

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