

何炳生教授四十年学术路

何炳生



编者按：从本期开始，本栏目将扩展为数学家访谈和数学家自述。作为自述栏目，我们很高兴请到知名计算数学家何炳生教授作为开篇作者。

何炳生，1948 年出生于江苏无锡，1966 年高中毕业后回乡当了 11 年农民。1977 年恢复高考被南京大学数学系录取，本科毕业被公派去德国留学，取得维尔茨堡大学博士学位后回国在南京大学数学系任教。1997 年晋升为南京大学教授，退休后于 2015 年加入南方科技大学。长期从事最优化理论与方法的研究，主要成果是变分不等式的投影收缩算法和 ADMM 为代表的凸规划的分裂收缩算法，代表性算法被成功应用于求解多类工程计算问题，研究成果被国内外学者广泛引用。

先后获得江苏省科技进步一等奖、中国运筹学会科学技术奖运筹研究奖、首届江苏省工业与应用数学学会突出贡献奖和教育部科学研究优秀成果自然科学二等奖。

文化大革命耽误了我 10 多年才上大学

我 1966 年就读完了高中，农村人，读的是当年江苏省十八所示范性中学之一的南菁高级中学，在江阴城里。正准备考大学的时候爆发了文化大革命，最初说这个革命“闹”半年后还可以考大学，最后的结果大家都知道。读中学时我在班上年龄最小，又很能考试，校长和老师们都比较爱护和器重。有老同学还记得，当我眼睛近视的时候，学校还出钱替我配了眼镜，这里应该是两方面原因，一是校领导对我的厚爱，另外也是因为我家经济比较困难。教我们高三解析几何的薛潮老师，她是学校里最年长的女教师，教学一丝不苟，1966 年夏天我们毕业考试结束，文化大革命山雨欲来，地主家庭出身的她预感到自己不再有上讲台的机会，把自己的由俄文翻译的教学参考书都送给了我。那个年代参考书很少，我爱不释手。文化大革命开始，我也参与“破四旧”，后来造反派学生开始把斗争矛头针对校长和一些“有问题的”老师，我就成了“保皇派”。再到后来我成了学校里和“造反派”对立的“保皇派”学生组织的

一号头目，原因我一直没想清楚。保皇派头头这个“桂冠”本不该轮到我，文化大革命之前，我只是高三年级六个班中一个班的团支部委员。是因为成绩好，跟我走的人多，还是互掐中我针对性的大字报写的不错？1966年10月作为学生代表到北京串联之后，看到国家主席都能被打倒，就知道自己“错”了，从此偃旗息鼓。等队伍解散得差不多，我就溜回家种田。文化大革命的混乱，使得我离校的时候居然一本念过的书都没有能带回家。

造反派得势之后，作为战果，要整几个“反动学生”，我必须是其中的一个。本来已经在离校20多公里的老家的我被押解回学校，关在一个窗户有钢棚的办公室，睡（有地板的）地铺。我成了待宰的羔羊，虽然提心吊胆，但皮肉没有受苦。所幸“造反派”很快分裂成两派，打斗愈演愈烈，以致无暇顾及我。机会来了，在一些朋友的帮助下，我侥幸逃脱。文化大革命开始时当“保皇派”，完全是念着校长和老师对我的好。后来“识了时务”，我也真以为自己犯了大错。最消极的时候，认为这一生能在农村安分守己种田，没有人再来抓我就算不错了。

回到农村，当年读完高三的我就算是文化程度最高的。慢慢地，农村的基层（生产大队）领导认为我可以帮忙做些事情，先是安排我当了大队团支部书记，年龄大了点以后又让我当了大队办厂的负责人。但是，我不能入党。按照文化大革命中的说法，我的家庭和社会关系都有一些问题，爸爸本人也不知道，他的名字怎么会在国民党的花名册上，这就应该算是历史问题隐藏得够深的了；姨父是新四军内部的反革命，姑父是叛徒，其实都是不实之词。由于这些问题，我不可能被推荐上大学，不是党员，那时候也就没有上升空间。但是，多一点文化，接受新技术也快一点。当年华东师大生物系推广的生物治虫、固氮菌肥等技术，我能比较快地接受，在农村是有那么一点被需要的感觉。文化大革命期间推广华罗庚“优选法”的小分队到我们县城宣讲，县里的科技干部也通知我去听讲。学习优选法的过程中，我知道了函数只是一种对应关系，不像中学书本上说的总有显式表达式，而是给了一个自变量，能观察到相应的函数值。实际生活中，这种观察往往是代价不菲的。优选法就是要通过尽可能少的观察次数，近似地找到对应最大函数值的自变量。我虽然没有从学校带回书本，但由于工作需要，学过的数理化还常常在脑海里盘旋。



直到1977年10月，2002年母校120年校庆见到高中时的康校长和他夫人



南菁高中原副校长吴新萃

我到刚刚办成正在试产的海山石化总厂，洽谈购买他们的试车废料。那是个从日本全套引进的现代化工厂，厂区非常安静，因工作需要我在厂招待所住了几天。那时的招待所大多是一个大房间，随机地安排几个互不相识的人同住。我住的房间是三张单人床，一张办公桌，但只住了两个人。同房间的那个年轻人在做10多年前的高考数学题，我问他做这些东西干什么，他说马上要恢复高考了，听后我心情非常复杂。上大学一直是我的愿望，可是我当时已经超过29岁，没听说年近30还能考大学的。除了羡慕年轻人机会比自己好以外，我也安慰自己，心想总可以把自己的孩子培养得好一点，让他们以后圆我

上大学的梦。谁知第二天在厂区就听到大喇叭播新闻说要恢复高考了，而且66、67届高中毕业生都可以考，30岁以内，婚否不限。这时候我简直有点惊呆了，有这等好事？赶快去邮局买份报纸看看。知道文化大革命期间的高中生学的东西不多，我深信考试自己是有竞争力的。文化大革命后期在江苏省教材编写组工作的原南菁中学主管教学的吴新萃副校长也写信给我，要我“站出来让国家挑选”。我决定参加高考，当时已经是无锡农村有点像样的大队办企业的负责人，工作不可能马上找人接手。我白天忙于工作，晚上抽点时间复习，资料呢，都是人家主动送上门来的。刚恢复高考，家长们都想让自己的孩子试试，不知道有些家长是通过什么渠道弄来一些油印的复习资料，带着子女来问我一些数学、物理和化学的题目，那些题目都不算太难，我基本上都能解答。这个为他人答疑的过程，同时也“复习了我”。那年报名参加考试的人太多，就由地区初考，合格的再参加省里统考。初试成绩不公布，只说通过了没有（我现在也没有想通有什么保密的必要）。因为一些家庭和社会关系的原因，统考前我填报的是江苏师范学院（现在的苏州大学）数学系，当时的无锡县也属于苏州地区，想毕业后回家乡当一名中学数学老师。

好在省里把考得好但没有报考重点大学的考生资料整理了出来，南京大学负责录取的老师就要了我，我很幸运地到南京大学数学系就读。那时我高中毕业在农村已经待了十多年，对数学的理解也很肤浅，以为有那种只要今天好好学，明天就能立竿见影派用场的应用数学。带着那种期待，我走进了南京大学。

我考上大学，常年贫病的母亲似乎第一次站直了腰。妻子淌眼泪了，当农民的我上大学后没有了收入，家庭和两个孩子要靠她一人务农负担几年。

读大学认真是当农民多年的勤劳惯性

上了大学以后，我们这些老三届学生，学习都有自己的特点。虽说我们原来基础比较好，但很快就尝到“数学是年轻人的学问”那味道。好在长期在农村一线劳动的人能吃苦，就像一种惯性，我读大学时最大的特点大概就是用功。在农村，我还会打打牌，上了大学我几乎没有再玩过，显得有点不合群。在乡下我也学会了抽烟，上大学以后，我就彻底把烟戒掉了。

学数学分析的时候，书上有二元函数在某一点有偏导数，而在这一点函数不连续的例子，我怎么也记不住。倒是10多年的种田经历帮助了我，我家在江南水乡，是稻麦两季，水稻要水，麦子怕渍。水稻收割前，就把田里的水排掉，让地凉干。待水稻收割后种麦子，要在地里纵横开沟。为多打一些粮食，这沟开的还有些讲究。沟要开的直，沟底要平，才便于沥水；也要开的窄，那是为了增加麦子生长的有效面积。精耕细作，是我30岁以前的那十多年里年复一年的工作。水田是平的，把它看成二维平面，那些开得笔直的沟，就近似地看成线，沟的纵横交叉处，就可以看成偏导数为零而不连续的点。刚洗净脚上的污泥，我就进大学念数学系，想到这样的例子，是最自然不过的。其实，不管怎样的经历，对人生总会有帮助的。

外语一直是我学习路上的拦路虎，中学我学了六年俄语。上了大学以后，因为没有基础，英语就得从认字母开始，一天只能背几个单词。我的英语比同学们差多了，特别是口语。大学三年级时候，学校开设了“二外”选修课，我当时想着自己年岁大，被选拔出国的机会不可能轮到我，就把刚学了一点的英语丢一边，再去把中学时期学过的外语捡回来，二外就选了俄语。

大学四年级以前我没有考虑要读研究生，因为我本科毕业就要接近34周岁了，当时认为出国留学或者考研都是年轻人的事，况且家庭已经靠妻子单独负担了好几年。快要毕业的时候，又觉得自己学的东西还解决不了什么问题，心想，是不是要读了研究生，才会有点真本事？上大学之前，推广华罗庚优选法小分队的报告对我影响很大，上了大学以后，又在南京大学礼堂听了华罗庚教授关于优选法推广应用的报告。当年南京大学的何旭初教授，是国内优化方向的领军人物，也希望我考他的研究生。可是我英语根本考不起来，老师说如果考他的研究生，外语科目也可以选考俄语，这样我才能报名参加硕士研究生考试。



力主录取我进南京大学的周树棠老师

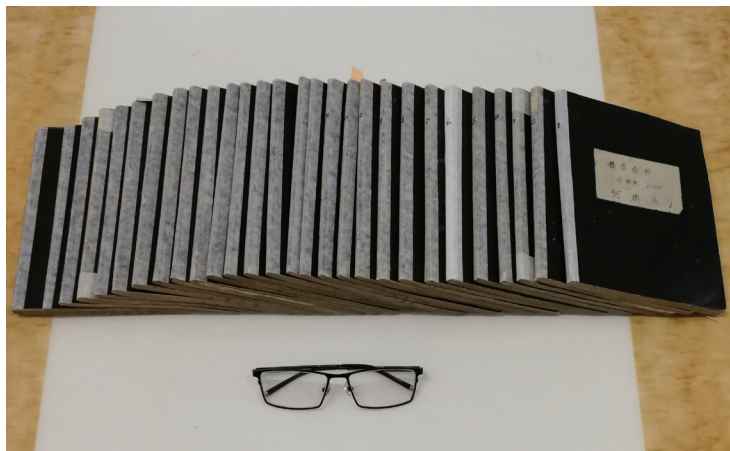


大学同班同学中人才济济，四排右五为中科院院士田刚，右四是何炳生

研究生考试主要考一二年级的基础科目，作为“老三届”大学生的我，前两年的学习还有优势，因此我考研又是比较顺利。考上南京大学的研究生以后又有机会出国学习，这我预先一点都没有想到。我的研究生考试成绩也不是最高，考得最好的是一位女生，她年龄也偏大。南京大学在录取的研究生中，每个系都能安排1、2名学生出国留学，这个名额是分配给导师的。何旭初先生名下有一个，他在考上的研究生中选了我，说不清是不是我也姓何的原因（当年我父亲是这样说的）。全校有二十几人得到这样的机会，这些人考研究生时外语考的几乎都是英语，只有两位例外，考日语的一位和考俄语的我。考英语的都去了美国和英国，考日语的去了日本。学校规定我只能在德国、法国和日本三个国家中选一个，因为去美国的学生已经超过教育部要求的比例了。这个时候我最后悔的是，大学三四年级不该去读俄语，而是应该把英语补一补。导师何旭初先生的意见是去德国也很好，并且要我师从德国维尔茨堡大学的 Stoer 教授。为此，我就得去上海外国语学院突击学习德语。回想起来，我上大学读研究生都不是一帆风顺，路是坎坷的，但也很



何旭初先生（1921-1990）



一些品相还可以的大学时代作业本

幸运。回头想想，也觉得很有意思。

大学时代，我做了不少作业，也整理了一些笔记。这些本本，有的还躺在那里，时间长的已经超过了四十年。去德国留学，行前，我把这些本本和用过的书捆在一起，从南京的大学生宿舍带回无锡农村的家里。那是1983年，我35岁，恰好是现在年龄的二分之一（本文初稿撰于2018年）。

我农村的家，房子比较简陋，没有几件家具。那年头，农村人，家里必须有地方做猪圈和堆放农具，一般都没有能放书本的柜子。我把这些本本装在一个纸箱里，告诉老婆不能随便丢弃。几年后我回国，它们还静静地躺在那里，只是有了点霉味。

回国后我在南京大学任教，这些本本就和其他材料一起回到城里，在家里不大的空间，它们占了一席之地。老婆一直有点意见，劝我把没用的趁早清理。年过古稀的我，也慢慢想通同意。到它们被付之一炬的时候，难道还会再有40年？丢弃，也要排个次序，翻翻，犹如与之告别。喔，由于专业关系，我自己都有点怀疑：那些年，我还曾经懂过这些？品相好一点的本本，犹犹豫豫中我还是决定让它多留几年。封皮快要脱落的，再用胶带纸粘在一起。不为再有使用价值，只为给自己余生多留一份回忆。

德国潜心读博 为后继研究打好了基础

德语考试合格后我才去了德国。当年，全国来说，最优化也是比较受重视的一个方向——中科院送了袁亚湘去英国跟 Powell 读研究生，西交大有徐成贤到英国跟 Fletcher 读研究生，南京大学就送了我到德国维尔茨堡大学跟随 Stoer。Powell, Fletcher, Stoer 都是那时国际优化界巨头。Stoer 是巴伐利亚科学院院士，当年（1982）刚开过的世界数学规划大会（ISMP）的主题报告人，他的著作 *Introduction to Numerical Analysis* 是著名的研究生通用教材。导师对我帮助很大，我到德国的时候，是1983年下半年，已经过了35岁，而我



德国导师 Stoer 教授 80 岁生日照片

的业务就是刚刚在国内念完本科的水平。那时大陆到德国留学的学生还很少，导师给我安排了一些课程，让我参加讨论班，也给了我一些他们同行交流的预印本（当时还没有电子文档），这些对启发我独立思考非常重要。他根据我的能力，给我合适的题目，容我细细琢磨，大概由于我年龄偏大，导师有意放我一马？到头来，由于我一直在学术界工作，我导师这一生最欣赏的学生中，可能我也算一个。

我们中国学生比较能做题，德国高年级学生中也不乏强手。记得在德国学习的时候，一次导师出的题目，大概 20 到 30 个学生，普遍做的不好。我在习题课的前一个晚上，从学校回家的雪地里，边走边想，想到了一个

简洁的方法，给负责习题课的助教和同学们都留下了深刻的印象。2015 年我去德国母校参加一个校友会组织的学术活动，有几个德国同学知道后，就到活动场所来看我，其中一对夫妇都是当年的同学，他们的女儿也念数学，即将大学毕业。这对夫妇还记得当年我让他们刮目相看的那堂习题课，要我回国后把题目和我的解答整理一下给他们，说要用来看考他们自己的孩子。那个时候我们中国公派出去的留学生不多，被国内单位选上不太容易，到了国外学习一般都比较优秀。

在专业学习的同时，为了提高自己的语言能力，我需要找德文的书籍学习。回想起来，对我专业发展影响最大的是 Blum 和 Oettli 由斯普林格出版的德文专著 *Mathematische Optimierung*，其中第六章讲到投影收缩算法，跟导师建议我考虑的投影梯度法很有关系。Blum 和 Oettli 都是苏黎世大学毕业的博士，Oettli 教授生前还接受过我的访问。我后来研究的算法，符合他们的专著中关于投影收缩算法的定义和算法框架，所不同的只是构造了未知距离函数的不同下降方向和步长法则，使得能在计算机上真正实现。出于对历史的尊重和对前人工作的一种敬畏，我还是把这些方法称为投影收缩方法（Projection and Contraction Methods）。

我在德国留学期间，我妻子和家乡的多数农民一样，有一份责任田，同时又在镇上的社办厂工作，叫忙时务农，闲时做工。1985 年春节前，她下班后在回家的公路上出了一次车祸，昏迷了三天。那个春节，我的一位中学同学和他夫人一起到我家探望，看看有什么可以帮助的，谁知我老婆还躺在县城医院，没有碰面。过了两个月，同学不放心，再去我家，到村边就看到我家正在办丧事，原来是我母亲去世了。我妈本来体弱多病，被我妻子车祸的事又受了大惊



我还保存的一些导师寄给我的贺卡

吓，每况愈下，刚过 60 就告别了世界。那时也不通电话，我父亲委托我的老同学把家里发生的事写信告诉我，同学说他本来被吓的那一跳还要更大，那没有说完的意思我也明白。那时，国家给我们的生活费中，是省不出一张来回机票的。导师知道我家里出了这么多事，要出旅费让我回家探望，为了让我好意思接受他的私人馈赠，他说资助困难学生的费用是可以在税前列支的。

我博士毕业后回国，跟导师还保持着联系。上世纪 80 年代后期还没有电邮，交流还是靠信件邮寄。一次我将自己用打字机打的文稿寄给导师，他自己当时正在学用 TeX 软件写数学文章，索性帮我修改后，用 TeX 打了文本给我，我也从那个时候开始学习用 TeX 软件处理稿件。现今，我只用电邮问候导师，今年 85 岁的他却还是每年寄我一张贺卡，我能怎么报答？惟有忠诚科学，好好工作！学术生涯中，我总想着要对得起 Stoer 的学生这个称号。

跨越古稀的那天，收到导师祝贺我生日的邮件。来信中他用德文说，人生七十，一定是回顾一生成就和困苦的好时机（70 Jahre geben Ihnen sicher Anlaß, auf Ihr gesamtes Leben mit allen Erfolgen aber auch Schwierigkeiten zurückzuschauen），他对我是那么的理解！这也是我写本文的一个动机吧。

树木育人 就怕耽误门下学生

博士毕业后我回南京大学工作，讲课还比较受本科生欢迎。我的教学理念是，提倡学生多联想。我常说，抬起手，跳一跳，树上挂得高一点的桃子，味道一定更好！南京大学计算机学院院长周志华，当年是南大计算机系的学生，数学系安排我教计算机系的计算数学，周是我教的那个班的学生，学期结束他



年近 50 我有了带博士的资格

代表几位同学给我送了一张贺卡，表示感谢。这是我教书生涯中从外系学生中收到的唯一一张贺卡。周志华尊我为师，说我是他大学时代的好老师，最近几年，他还不时请我给他的研究生作些讲座。我在外面参加学术活动，提起周志华，做机器学习的，没有人不知道他学问做得好，有过交往的，一定再赞一句他为人谦虚厚道。人们都说他毕业以后发展很好，我说因为他学生时代给我送了贺卡，以后好运就一直跟着他！

到了能带研究生的时候，我进一步审视自己的研究方向。优化方法是最接地气的应用数学，越是普通大学生就能懂的算法才越有用，这是我的价值观。当年华罗庚教授推广“优选法”，就是看上了它的简单，一般技术人员就能学会，能在生产实践中发挥作用。华罗庚先生深入浅出诠释的优选法，当时是高中生的我都能学懂，尝试着在科学种田的实验中应用。现在全民族的文化素质提高了，大学生的数量不比当年的高中生少。从我们计算数学专业的优化工作者来说，是不是要把着眼点放在提出一些一般工程师们都能听懂，并在实践中应用的方法？这是我考虑问题的出发点。

研究往哪条路走比较好，一时我也说不清楚。其实，这跟哪块地适合种什么庄稼一样，因人而异。我有资格带博士生以后，既不招揽学生到自己门下就读，也没有强求学生做什么。几十年来，庆幸自己能当一名大学老师，与研究生们日常交流也能迫使自己不断学习，更新知识。

科学研究有一种不确定性，需要一颗奋发向上的心，更需要一颗平常的心。兴趣是最好的导师，自己觉得有意义，就去做，能不能做成是另外一回事。有人觉得某个方向他人已经领先，我不能当第一，就干脆不做，我认为有点过分功利。科学研究的结果不是能预先规划好的。

带研究生是有一定责任的，有人认为“跟了一位好老师，学生以后一定能怎样怎样的”，其实学生跟着老师做的东西，还只是一个阶段性成果。学生在

读期间，受老师阶段性的兴趣影响，有的能做出一些好的结果，有的并不出色。评价一个人的学术成就，还要看他离开导师后能不能做出一些独立的工作。

我自己不招研究生以后，还建议自己的有了“资格”的学生晚一点招，或者少招一些研究生为好。我们毕竟不是直接做工程的，不需要那么多的助手，自己没有好的想法会影响学生的发展。不带研究生，研究做得比较低迷是自己一个人的事情。我常担心自己的研究课题，能不能给学生带来足够有益的训练。

说到创新，年轻人首先要有自己的好奇心。数学研究实际上还是个体劳动，不要把研究人员往一条道上赶。当然最好是几个有共同意愿的人，互相激励，小范围讨论，能擦出一些思想的火花更好。一般来说，我们做的是自选项目。不是让某个人去组织某个团队把某个问题解决了。能算个动手派的我，特别鼓励好学生到国防科研第一线，默默无闻为国铸剑。我的好学生中，有在咱们国内重要研究所工作的，也有在国防科研单位挑大梁的，我引以为豪！

培养创新人才，要在好奇心强的学生中找苗子，又要让他们独立思考。学生自己一定要有好的眼界，能把遇到的问题看透，当然这个能力不是一天两天培养的。我大学同学中优秀人才很多，特别是搞基础数学的，在本专业计算数学的同学中，我的个体条件也比较差。最终我还做出一些或许能留下一点痕迹的东西，原因是我在研究的“入口”处多站了一会，仔细考虑了我能做什么，做成什么样的方法更容易被相关学科采用。

在数学系做优化方法的学者，往往不是去解决一个具体的问题，而是希望自己做的方法，能被不同领域拿去解决实际问题。能做到这一点很不容易，这只能算是我们计算数学工作者可遇而不可求的希望。所幸我在一定程度上做到了。在数学研究中，有人不提倡领头羊的说法，我也同样认可。可以有一些散羊，如果一群羊都跟着领头羊跑，哪怕找到新的草原和水源，大家也吃不饱，有一些散羊，散开了走，自己去找路，当然散羊会面临更严峻的生存问题，可能在半路中渴死了、饿死了。但是，如果有几头散羊找到新的草原和水源，就是不小的成功。对整个学科发展来说，思想的自由是非常非常重要的。

我对科研也一直保持着极大的热情。一生只取得很少的一点成绩，所以每一个好结果是在什么时候、什么情况下得到的，到现在还历历在目。研究的问题一有新进展，总是第一时间与学生交流。按照学校的规定，研究生要在老师毕业（喔，那应该叫退休）之前毕业，过了62岁就不能再招研究生。2010年，我不能再招收研究生的时候，写了一个关门感想，记录了我当时的心情，不妨抄录于后。

一个计算数学工作者的关门感想

无奈没有少时照片，但悔没能留下当年随笔。关门之际，写上几句，酿一壶小酒，供余生斟饮。

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

谨以冯先生的话开场。先贤名言，铭记在心。纵然可以没有本领，不能迷失价值标准。

一. 无论如何，要在工作找到美的享受

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。

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，他人才会拿来使用；统一，自己才有美的享受。只因自以为发现了其中的一些数学之美，研究才变得充满激情和欲罢不能。

他人看来未免纳闷，却是我长年自得其乐的原因。

二. 自己可以一事无成，就怕耽误门下学生

做了点自以为漂亮的东西，沾沾自喜。但也常问自己，或以平庸为神奇？并非戏言，自以为不错的东西，很可能一文不值。老一辈数学家越民义先生在一个报告中就说到：学生“很容易将老师所说作为数学的前沿作品来接受，工作后又将这些东西作为衣钵往下传。若因循守旧，则易造成谬种流传，误人不浅”。前辈良言，置我案前。

说到自己，年少时考试还行，被老师表扬几句，就以为比人家聪明。读完高中当了10多年农民，上了大学才知道数学是年轻人的学问，从此渐失自信。当耳闻议论，或实验不顺，更怀疑路是否走准。自己可以一事无成，就怕耽误门下学生。

三. 准备打烊关门，顿感压力减轻

窃以为：当今考上南大本科，一般是既要有实力，又要有运气。正儿八经考取南大读研究生，通常也是祖坟上风水好才行。师者，传道、授业、解惑。我不善织网，人脉不广，学生找工作，提职称，大都帮不上忙。

愚虽从名师大家，但入门迟，不开窍；学识浅，懂的少。当了教授，许有自知之明，无意接着申报博导。有了“资格”以后，也没有招徕学生到门下就读。深怕，开弓没有回头箭，引入入错行，回头无良方。

二十多年学术路，特别需要感谢的，除了导师，便是几个能一起平等讨论问题的学生。卖瓜路上，没有人听，恐怕也难维持（坚持）至今。

开张晚，关门早，经营时间不长，学生还是不少。肚中就那几两墨，又不善因材施教。虽说是，师父领进门，修行在个人；然年轻人最有创造力的几年都在我们身边度过，见毕业后学业上没有多大发展的，当导师的我总觉也有几分责任。

如今耳顺又二，不再招收新生。准备打烊关门，顿感压力减轻。往后与毕业弟子切磋讨论，再无指导责任。数学兴趣之外，吟读优秀散文，如此美好生活，早已令我憧憬！

四. 个人经历严重影响自我定位和价值取向

有朋友建议，到了这个年龄，可以花些时间写写自己的数学思想。思想谈不上，感悟有一筐。我做的是最优化理论与方法，实际上又偏重算法，说到底在数学界也多半是个工匠。而立之前务农做工，为养家糊口，最想多点技术，当个能工巧匠。能有茅庐栖身，曾是当年最高理想。

农民当的时间长（因而狭隘），做工也在小作坊（所以肤浅）。日后上了大学，也未见变得高尚。一个再好的工匠，最多匠心独具，不能高瞻远瞩，岂能奢望思想。一知半解，似懂非懂，能不误导学生，就算功德无量。

回头看，做点所谓研究，多属瞎子摸象。翻阅以往论著，不乏浅陋文章。分解降低难度，整合把握方向，看似基本常识，屡屡帮我大忙。点滴经验感想，如能惠及他人，算我没有白忙。否则，君权当遇上个讨厌老头，随他喋喋不休，总有谢幕时光！

股市有风险，投资需谨慎。研究耗用时间，年华岂容浪费。尺有所短，寸有所长。找导师，选方向，扬长避短第一桩。听我宣讲，读我文章，是否真有理，敬请独立考量。

五. 文章能否站住脚，余生哪能不思考

若不半途改行，许会给国家多卖几年好粮，到了这个年龄，就担心养老有没有保障。只是后来又上了大学，写了文章，带了弟子，还得了个省级奖。有张教授“执照”，衣食会比兄弟强。君不知，文章能否站住脚，弟子曾否被误导，算法是否真有效，余生哪能不思考！偶见一些星星点点的关注与好评，权当没有误导学生的辩解与佐证材料。与生与友，分享快乐。

扬长避短 做与自己能力相适应的研究

我的研究历程非常简单，就是从变分不等式的投影收缩算法到以 ADMM 为代表的凸优化的分裂收缩算法，一个模式，一条主线。做这类研究，所需要的只是：

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。

我的论文和讲义里，大多是矩阵和向量的基本运算，再加上中学就学过的余弦定理。理论证明过程中用得最多的是 Cauchy-Schwarz 不等式，如果我要把自己的工作用两页纸表述¹，那就是本页下面的图示。

从变分不等式的投影收缩算法到凸优化的分裂收缩算法

— 我主导的主要研究工作 maths.nju.edu.cn/~hebma 何炳生

1 变分不等式的投影收缩算法

设 $\Omega \subset \mathbb{R}^n$ 是一个非空凸集, F 是 $\mathbb{R}^n \rightarrow \mathbb{R}^n$ 的一个映射. 考虑单调变分不等式

$$u^* \in \Omega, \quad (u - u^*)^\top F(u^*) \geq 0, \quad \forall u \in \Omega. \quad (1.1)$$

我们说变分不等式单调, 是指其中的算子 F 满足 $(u - v)^\top (F(u) - F(v)) \geq 0$. 在求解变分不等式 (1.1) 的投影收缩算法中, 对给定的当前点 u^k 和 $\beta_k > 0$, 我们利用投影

$$\bar{u}^k = P_\Omega[u^k - \beta_k F(u^k)] \quad (\text{也即 } \bar{u}^k = \text{Arg min } \{\frac{1}{2}\|u - [u^k - \beta_k F(u^k)]\|^2 \mid u \in \Omega\})$$

生成一个预测点 $\bar{u}^k$. 对任意的 $\beta_k > 0$, 假如 $\bar{u}^k = u^k$, u^k 就是 (1.1) 的解. 否则, 我们假设选取的 β_k 满足

$$\beta_k \|F(u^k) - F(\bar{u}^k)\| \leq \nu \|u^k - \bar{u}^k\|, \quad \nu \in (0, 1). \quad (1.2)$$

由于 $\bar{u}^k$ 是问题 $\min_{\{u \in \Omega \mid \|u - [u^k - \beta_k F(u^k)]\|^2 \leq \frac{1}{2}\|u^k - \bar{u}^k\|^2\}}$ 的解. 根据极小化问题的最优性条件(与“瞎子爬山”类似, 只是这里是探“底”), 可行方向和下降方向的交集是空集, 就有

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

上式两边都加上 $(u - \bar{u}^k)^\top d(u^k, \bar{u}^k)$, 其中

$$d(u^k, \bar{u}^k) = (u^k - \bar{u}^k) - \beta_k [F(u^k) - F(\bar{u}^k)]. \quad (1.4)$$

由此得到我们需要的预测公式

$$[\text{预测}] \quad \bar{u}^k \in \Omega, \quad (u - \bar{u}^k)^\top \beta_k F(\bar{u}^k) \geq (u - \bar{u}^k)^\top d(u^k, \bar{u}^k), \quad \forall u \in \Omega. \quad (1.5)$$

将 (1.5) 中任意的 $u \in \Omega$ 选成 u^* , 就有 $(\bar{u}^k - u^*)^\top d(u^k, \bar{u}^k) \geq 0$, 随后得到

$$(u^k - u^*)^\top d(u^k, \bar{u}^k) \geq (u^k - \bar{u}^k)^\top d(u^k, \bar{u}^k). \quad (1.6)$$

由假设 (1.2), 容易得

$$(u^k - \bar{u}^k)^\top d(u^k, \bar{u}^k) \geq (1 - \nu) \|u^k - \bar{u}^k\|^2. \quad (1.7)$$

当 u^k 不是 (1.1) 的解时, 不等式 (1.6) 右端为负, 说明 (1.4) 中定义的 $d(u^k, \bar{u}^k)$ 是未知函数 $\frac{1}{2}\|u - u^*\|^2$ 在 u^k 处欧氏模下的一个上升方向. 因此, 可以

$$[\text{校正}] \quad \text{用 } u^{k+1} = u^k - \alpha_k^* d(u^k, \bar{u}^k) \quad \text{产生离 } u^* \text{ 更近的迭代点.} \quad (1.8)$$

$$\text{其中步长 } \alpha_k^* \text{ 由 } \alpha_k^* = (u^k - \bar{u}^k)^\top d(u^k, \bar{u}^k) / \|d(u^k, \bar{u}^k)\|^2 \text{ 给出.} \quad (1.9)$$

由 (1.2) 和 (1.4), 可得 $2(u^k - \bar{u}^k)^\top d(u^k, \bar{u}^k) > \|d(u^k, \bar{u}^k)\|^2$, 因而 $\alpha_k^* > \frac{1}{2}$. 接着就有

$$\begin{aligned} \|u^k - u^*\|^2 - \|u^{k+1} - u^*\|^2 & \quad (\text{利用 (1.8), (1.6), (1.9) 和 (1.7)}) \\ & \geq \alpha_k^* (u^k - \bar{u}^k)^\top d(u^k, \bar{u}^k) \geq \frac{1}{2} (1 - \nu) \|u^k - \bar{u}^k\|^2. \end{aligned} \quad (1.10)$$

这是证明算法收敛的关键不等式. 分处不等式 (1.5) 两端的 $\beta_k F(\bar{u}^k)$ 和 $d(u^k, \bar{u}^k)$ 称为一对孪生方向. 这里最妙的是, 采用方向 $\beta_k F(\bar{u}^k)$ 和 (1.9) 中给出的步长 α_k^* 用

$$[\text{投影校正}] \quad u^{k+1} = P_\Omega[u^k - \alpha_k^* \beta_k F(\bar{u}^k)] \quad \text{与方法 (1.8) 具有同样的收敛性质 (1.10).}$$

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

2 凸优化的分裂收缩算法

我们以可分离凸优化

$$\min\{\theta_1(x) + \theta_2(y) \mid Ax + By = b, x \in \mathcal{X}, y \in \mathcal{Y}\} \quad (2.1)$$

为例, 它的拉格朗日函数是 $L(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^\top (Ax + By - b)$.

拉格朗日函数的鞍点 $((x^*, y^*), \lambda^*)$ 满足

$$L_{\lambda \in \mathbb{R}^m}(x^*, y^*, \lambda) \leq L(x^*, y^*, \lambda^*) \leq L_{x \in \mathcal{X}, y \in \mathcal{Y}}(x^*, y^*, \lambda^*).$$

利用鞍点 $w^* = ((x^*, y^*), \lambda^*)$ 的极大极小性质, 它可以表述为以下的变分不等式:

$$w^* \in \Omega, \quad \theta(w) - \theta(w^*) + (w - w^*)^\top F(w^*) \geq 0, \quad \forall w \in \Omega. \quad (2.2)$$

上面的变分不等式中, $\Omega = \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m$, $\theta(w) = \theta_1(x) + \theta_2(y)$, 并且

$$u = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^\top \lambda \\ -B^\top \lambda \\ Ax + By - b \end{pmatrix}. \quad (2.3)$$

注意到, (2.3) 中的 $F(w)$ 恰有 $(w - w^*)^\top (F(w) - F(w^*)) = 0$, 所以也是单调的.

求解变分不等式问题 (2.2), ADMM 交替方向类算法都属于一个 预测-校正的统一框架.

[预测]. 对给定的 v^k (在 ADMM 中 $v = (y, \lambda)$, 有些算法中 $v = w$), 求得点 $\bar{w}^k$, 使其满足

$$\bar{w}^k \in \Omega, \quad \theta(\bar{w}^k) - \theta(v^k) + (v^k - \bar{w}^k)^\top F(\bar{w}^k) \geq (v^k - \bar{w}^k)^\top Q(v^k - \bar{w}^k), \quad \forall v \in \Omega, \quad (2.4)$$

如果矩阵 $Q^\top + Q$ 正定 (Q 不一定对称), 我们称 $\bar{w}^k$ 为预测点.

将 (2.4) 中任意的 $v \in \Omega$ 选成 w^* , 就得到 $(\bar{w}^k - w^*)^\top Q(v^k - \bar{w}^k) \geq 0$, 随后得到

$$[H(v^k - v^*)]^\top [H^{-1}Q(v^k - \bar{w}^k)] \geq (v^k - \bar{w}^k)^\top Q(v^k - \bar{w}^k). \quad (2.5)$$

这里 H 为任意正定矩阵 (也可以是单位阵). 由 (2.5) 和 $Q^\top + Q$ 正定, $H^{-1}Q(v^k - \bar{w}^k)$ 是未知函数 $\frac{1}{2}\|v - v^*\|_H^2$ 在 v^k 处 H -模下的一个上升方向. 如同投影收缩算法中利用 (1.6),

[计算步长的校正]. 可以对选定的正定矩阵 H , 由以下法则生成新的迭代点:

$$v^{k+1} = v^k - \alpha_k^* M(v^k - \bar{w}^k), \quad \text{其中 } M = H^{-1}Q, \quad \alpha_k^* = \frac{(v^k - \bar{w}^k)^\top Q(v^k - \bar{w}^k)}{\|M(v^k - \bar{w}^k)\|_H^2}.$$

[单位步长的校正]. 对正定矩阵 H , 同样记 $M = H^{-1}Q$, 如果矩阵

$$G = Q^\top + Q - M^\top H M \succ 0, \quad (\text{正定}) \quad (2.6)$$

也可以采用单位步长, 生成新迭代点的校正公式就简化为

$$v^{k+1} = v^k - M(v^k - \bar{w}^k). \quad (2.7)$$

$$\begin{aligned} \|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 & \quad (\text{利用 (2.7)}) \\ & = \|v^k - v^*\|_H^2 - \|(v^k - v^*) - M(v^k - \bar{w}^k)\|_H^2 \quad (\text{利用 } HM = Q \text{ 和 (2.5)}) \\ & \geq 2(v^k - \bar{w}^k)^\top Q(v^k - \bar{w}^k) - \|M(v^k - \bar{w}^k)\|_H^2 = \|v^k - \bar{w}^k\|_G^2. \end{aligned} \quad (2.8)$$

因此, 生成的序列 $\{v^k\}$ 具有收敛性质

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

✧ 从 变分不等式的投影收缩算法 到 凸优化的分裂收缩算法, 一条主线, 一个模式! ✧

从变分不等式的投影收缩算法到凸规划的分裂收缩算法——两页纸简述我的研究工作

¹ http://maths.nju.edu.cn/~hebma/thinking/Two_Page.pdf

我研究的第一类算法叫做变分不等式的投影收缩算法，粗略地说，我是30岁读大学，40岁研究生毕业，40岁到50岁之间在做投影收缩算法，50岁到60岁做以ADMM为代表的凸优化的分裂收缩算法。以后呢，60岁也快退休了，信息科学的发展，使得我以前重视的一些方法有了新的用武之地。这个时候相比其他学者，我有了比较好的研究基础，后来在这个方向就做得比较顺手，得到了一些经典的理论结果，并且提出了一些补缺的新的算法。

我们在变分不等式框架下研究优化方法，有人一听到变分不等式这个名字就会觉得难。其实，变分不等式用语言表达就是：所有可行方向都不再是下降方向，换句话说，就是找到了最低点。或者说，变分不等式是“瞎子爬山”（相对于爬高四周都不再有上升方向）的数学表达形式。大学生都知道，在中学求二次函数的极值，一般都用配方的方法，学了微积分以后，就用求导法则。一元二次函数求一次导就变成一个线性方程，求解容易多了。因此，用变分不等式处理优化问题，就像学了微积分以后，采用求导数的方法求二次函数的极小点一样，方便许多。

我们做的优化算法又都跟邻近点算法有关系，只不过经典的邻近点方法的二次正则项是欧氏距离，我们采用的是正定矩阵模下的距离。邻近点的思想是，在探索的前进道路上，下一步不要离开当前点太远，也就是“饭要一口口吃，路要一步步走”。看似保守，把握却大一些。

不少工程界的朋友给我写信，说我和我的学生的文章没有云山雾罩，犹如科普，容易读懂。听到这些赞扬，我特别欣慰。

算法是否真有用 要由用户说了算

我做变分不等式的投影收缩算法，跟当年在国外图书馆阅读 Blum 和 Oettli 的德文专著有关，多年没有放弃，也跟华罗庚先生推广优选法有关。现实生活中，我们往往关心路桥上车辆的密度（导数）而不在意究竟过了多少车，或者说，现实生活中，有的函数，梯度比函数值本身容易求，有的问题除了知道函数的凸性以外只能观察到它的梯度。投影收缩算法是求解变分不等式的只用导数（梯度）值的方法。上大学以前听报告弄懂了优选法，对自己一生影响特别大。我们做的变分不等式的投影收缩算法，近10年在工程领域也得到了应用，起到其它算法不能替代的作用！下面的两个应用特别令我欢欣鼓舞。

机器人控制里边有个特定的优化问题，要在三毫秒内给出结果。中山大学的教授和他的研究生们试用 MATLAB 库里的办法（或经他们自己改造）均不能满足要求，说是用我20多年前提出的求解线性变分不等式的方法才解决了问题。他们发的系列论文和出的专著都引用了我的文章，主持这项研究的学者现在是广东省珠江学者冠名教授。

中国科学院武汉岩土力学研究所的国家重点实验室副主任，一位老杰青，现在是北京工业大学特聘教授。他们用文献中宣传得比较多的方法求解一类接

触力学问题，让学生编程计算，说是消耗了几届博士生。后来发现了我 20 多年前的文章，让我去讲一讲，简单的方法很快就学懂了，由此解决了困扰他们多年的问题。项目负责人到欧洲讲学，还用几幅附上我照片的 PPT，向同行介绍说用了我的算法。告诉我说，有关工作他们还计划做成相应的商业软件。

尺有所短，寸有所长。优化界前辈 R. Fletcher 早就说过，“世界上没有一个算法是对所有问题都是最好的”。我的算法能找到实际应用，完全是因为简单。工程界的问题，我没有精力去弄得那么懂，最多也是一知半解。从这点上说，真得好好感谢那些用了我们的方法又说了实话的学者。

我是做优化通用算法的数学工作者，能在有生之年看到自己的方法被他人用来“救急”，解决一些他们原来解决不了的问题，吾复何求！吾复何求？

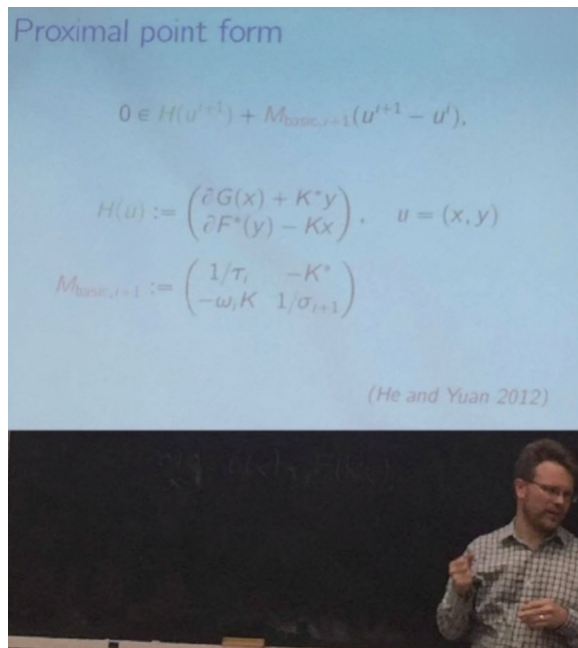
分裂收缩算法的研究成果扩大了我们的学术影响

我们同样是在变分不等式框架下做凸优化的分裂收缩算法。说起来，我们更多的是采用预测 - 校正的策略，这也跟我离开农村的前两年负责一个根据用户要求按图加工非标准设备的工厂有关。那些非标准件设备，个体大，做好每个产品都要动点脑筋。和几个能工巧匠商定方案以后，就把工作分解给相关的工人（小组），各个部件做成之后再整合。这种工作方式也影响到了我做应用数学研究，那就是，“分解”降低难度，“整合”把握方向。

我们从变分不等式的视角出发提出的一些方法，形式特别简单。友人到海外访问，参加会议，不经意间发现会议的第一位报告人讲的方法，就用一幅幻灯片说是用了我和袁晓明在 2012 年提出的变分不等式的邻近点形式 (Proximal

Point Form)。这说明简单的思想才会很快传播，容易被人接受。

交替方向法 (ADMM) 的一个重要理念是化繁为简，在每步迭代中把原本无从下手的问题分解成容易求解的子问题。我开始做交替方向法是 1997 年，当时这类方法并没有引起优化界足够的重视。我觉得这种化繁为简的思想方法很重要，ADMM 又跟我 97 年以前做了十多年的投影收缩算法有密切的关系。到了 2006 年，准备结束这个课题研究的时候



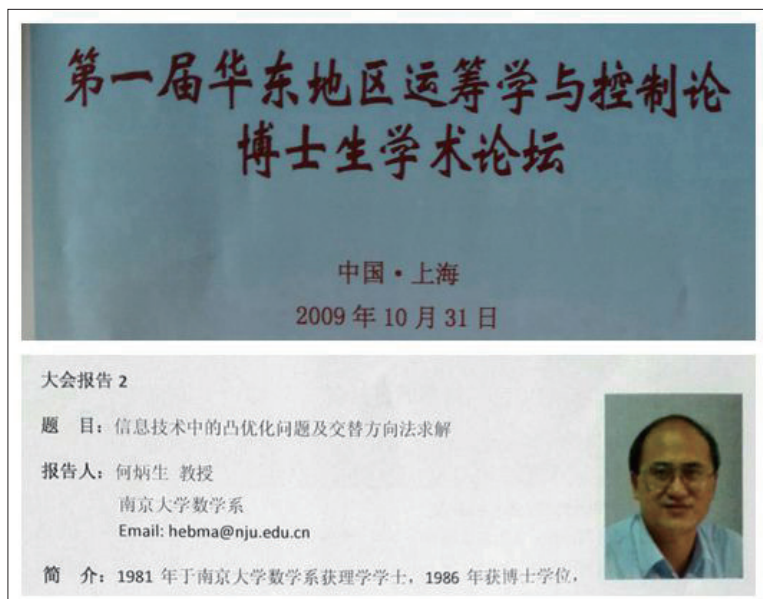
报告人指出用何 (He) 和袁 (Yuan) 2012 年提出的 PP 形式

候,发现这些工作在信息科学中用处很大,重燃了我新的研究热情。当时国际上已经开始用一些分裂方法求解信息科学中的问题,我读了一些文章,觉得人们用的交替极小化方法应该都比不上交替方向法,因为后者让对偶变量(拉格朗日乘子)也参加了迭代,犹如诚意谈判也要考虑对手的感受(这可能也跟我上大学之前在小企业负责,要跟甲方洽谈有关)。当2009年华东地区运筹学博士生论坛请我作大会报告的时候,我就讲信息科学中的一些问题和交替方向法(ADMM)求解。

交替方向法(ADMM)真正热起来,是斯坦福大学 Boyd 教授(他在2017年当选为中国科学院外籍院士)他们在2010年写了篇关于交替方向法的综述文章,其中提到我们2000年的工作。由于他们的宣传,国内外不少学者知道了我们在交替方向法方面早就开展了工作,并关心我们的后续研究。

提出交替方向法的人,都是从求解微分方程出发。我们看懂了这种方法后,觉得可以用来求解具有可分离结构的优化问题。交替方向法有两个起源,其中一个是从上世纪50年代就已经开始了,直到1978年才由大数学家 P. Lions 他们证明了收敛性,用的都是算子理论,那个方面的知识我到现在还比较生疏。另一个源头,是著名数学家 Glowinski 和他的学生们上世纪70年代发表的文章,是通过求泛函极小求解微分方程。看懂后我就认定,这个方法用来求解结构型优化问题是会有作为的,我们有的稿子到了 Glowinski 手里审,他很高兴,说真不知道交替方向法(也就是ADMM)还可以用来求解优化问题。

擅长算子理论的学者们为纪念 Jonathan M. Borwein,最近有篇回顾 Douglas-Rachford 算法六十年的综述文章(Suevey: Sixty Years of Douglas-Rachford),我也只能读懂个粗枝大叶。或许是他们看懂我们的更容易,这篇综述文章引用了我和袁晓明2012年以来发表的关于收敛速率的三篇文章,说我们给出了简单的证明。我要说的是,我们这三篇论文,加起来还不到30页!



2009年10月我在华东地区博士生论坛做大会报告

Stephen P. Boyd

- [Home](#)
- [Teaching](#)
- [Biography](#)
- Research**
- [Books](#)
- [Papers](#)
- [Software](#)
- [Students](#)
- Classes**

ADMM

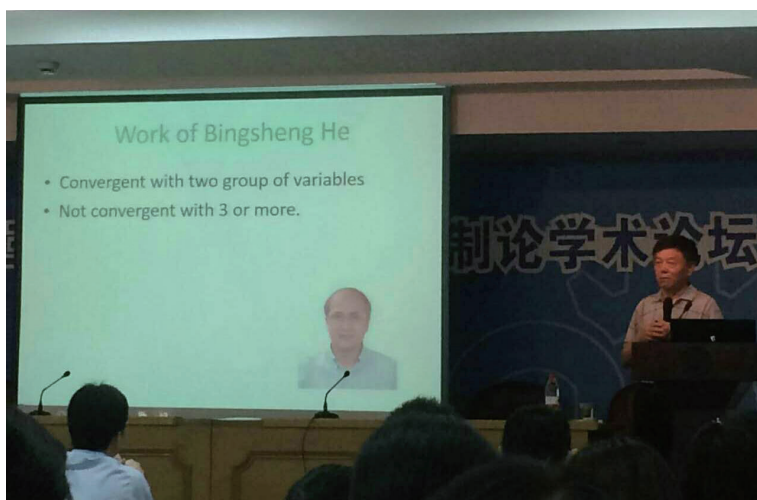
The *alternating direction method of multipliers* (ADMM) is an algorithm that solves convex optimization problems by breaking them into smaller pieces, each of which are then easier to handle. It has recently found wide application in a number of areas. On this page, we provide a few links to interesting applications and implementations of the method, along with a few primary references.

ADMM is used in a large number of papers at this point, so it is impossible to be comprehensive here. We only intend to highlight a few representative examples in different areas. To keep the listing light, we have only listed more detailed bibliographic information for papers that are not easy to find online; in any case, the information given

Theory and variations

- On the $O(1/t)$ convergence rate of the Douglas-Rachford alternating direction method
B. He and X. Yuan, 2012
- Fast alternating direction optimization methods
T. Goldstein, B. O'Donoghue, S. Setzer, and R. Baraniuk, 2012
- Augmented Lagrangian and alternating direction methods for convex optimization: a tutorial and some illustrative computational results
J. Eckstein, 2012

Boyd 的 ADMM 网站上链接我和袁晓明关于 ADMM 收敛速率的文章



知名优化学者堵丁柱在有关论坛上介绍我们的工作

简单，他人才容易看懂，才能帮助我们宣传。

学界似乎对算法的理论结果更加重视，我们关于 ADMM 方法收敛速率的文章，发表以来一直被广为关注。SIAM 系列的各个期刊都会根据文章在过去一年中被下载的次数排个序，列出前 20 篇文章。我们关于 ADMM 收敛速率的文章在 SIAM Numerical Analysis 的热点论文中曾经排名第二，现在仍然名列前茅。

中国科学院文献情报中心科学计量团队也有个《中国热点论文榜》，每个学科在前五年发表的文章中按引用次数选 10 篇论文。中国数学领域（2012-2016 年）的热点论文，袁晓明和我的这篇论文排名第一。我的看法是，现在提数据科学，首先是要把数据拿来。当然，有了数据，一定要分析，对数据去伪存真，去粗存精，否则，数据就有可能误导人们。

过去我的在读研究生往往会问我，到底做什么好，我总跟学生说，我自己也没有十分的把握，只能告诉你我做成了什么和现在正在做什么。因为我也听一些著名数学家说过，他们也不一定知道明天该做什么。通常，我会建议我的学生去看看世界数学家大会上与我们研究方向关系密切的哪些学者在作报告，能不能看懂他们的报告？看不懂的话，再去浏览一下他们的主页，看看能不能跟他们的研究兴趣找到共同点。当年说这个话的时候，我根本没有想到，2006 年以来的四届世界数学家大会，都有大会报告人在他们的文章或讲义中提到了我们的工作。世界数学规划（ISMP）会议上，大会特邀报告人（Plenary Speaker）在报告中也提到我们 ADMM 方面最熟知的收敛速率（Best known convergence rate）的工作。

用来求解两块可分离结构凸优化问题的交替方向法，能否直接推广用来求解具有多块可分离结构的凸优化问题？大家本来以为也是可以的，只是没有可靠的证明。2013 年，我和合作者们给出了否定的答案。在做出这个否定答

案之前,我们已经率先提出了一些充分利用交替方向法优点、略作修正就能够处理多块可分离结构凸优化问题的分裂算法。我们的方法的特点是对原问题的假设不加任何额外条件,这些根据社会生活理念设计的算法一发表,很快就被加州大学洛杉矶分校 Osher 课题组采用。

对近代最优化发展,我经常讲几个里程碑,上世纪 40 年代中期的线性规划的单纯形算法,60 年代的拟牛顿法,80 年代的内点法,我总是把它归结为 20 年一个阶段。要是这么推算的话,到 2005 应该出现新的里程碑,但实际上没有一类标志性的算法出现。

做了 20 年交替方向法,我在 2018 年的《运筹学学报》发表了“我和乘子交替方向法 20 年”。这篇容易读懂的文章,总结了我们的 ADMM 算法方面,特别是近 10 年的工作。仔细阅读,也许就能掌握根据实际问题构造算法的基本技巧。

当自己认为很好的工作没有得到足够重视的时候,也应该以平常心对待。即使是一个不错的东西被越来越认可,曾经说过风言风语的人也不太可能歇手。这个时候,就用得着 Max Planck 的一句话:

“一个新的科学真理取得胜利,不是通过让它的反对者信服,而是随着这些反对者的最终老去,熟悉它的新一代终究会成长起来。”

年轻学者往往会抱怨自己结果不错的文章很难发表,其实大部分学者都有这个困扰。能发好杂志当然很好,现在网络信息传播很快,我们的被国际著名学者引用的结果,都是在正式发表之前就受到关注。Osher 课题组采用我们的方法求解三个可分离块的问题,提出原始-对偶一阶算法(图像学界也称为 CP 算法)的 Chambolle 和 Pock 等说了我们许多好话的文章,引的都是我们贴在 Optimization-Online 上还没有正式发表的文章。有关文章在网上公开后不久,E. Candès 的学生 S. Becker 在他 2011 的一篇研究报告中(后公开于 arXiv:1908.03633v1)第一句话就说,何袁的这项工作是分析优化方法的一个非常简单又强大的技术(Recent works such as [HY12] have proposed a very simple yet powerful technique for analyzing optimization methods)。我们几篇有较大影响的论文审稿周期都比较大,做研究真的要有比较好的心理素质。只要结果实在,事隔多年人家也会发现。Boyd 他们 2010 年的 ADMM 综述文章,以及 2017 年发表在 JMLR 上关于他们求解器的文章,引用我们 2000 年提出的调比准则。我们的文章发表在 JOTA,这个杂志也不是优化界的什么顶级刊物。

2018年3月
Mar., 2018

运筹学学报
Operations Research Transactions

第22卷 第1期
Vol.22 No.1

DOI: 10.15960/j.cnki.issn.1007-6093.2018.01.001

我和乘子交替方向法20年*

何炳生^{1,2,†}

摘要 1997 年, 交通网络分析方面的问题把作者引进乘子交替方向法 (ADMM) 的研究领域. 近 10 年来, 原本用来求解变分不等式的 ADMM 在优化计算中被广泛采用, 影响越来越大. 这里总结了 20 年来我们在 ADMM 方面的工作, 特别是近 10 年 ADMM 在凸优化分裂收缩算法方面的进展. 梳理主要结果, 说清来龙去脉. 文章利用变分不等式的形式研究凸优化的 ADMM 类算法, 论及的所有方法都能纳入一个简单的预测-校正统一框架. 在统一框架下证明算法的收缩性质特别简单. 通读, 有利于了解 ADMM 类算法的概貌. 仔细阅读, 也许就掌握了根据实际问题需要构造分裂算法的基本技巧. 也要清醒地看到, ADMM 类算法源自增广拉格朗日乘子法 (ALM) 和邻近点 (PPA) 算法, 它只是便于利用问题的可分离结构, 并没有消除 ALM 和 PPA 等一阶算法固有的缺点.

没想到古稀的我成了一个数学科普工作者

由于交替方向法被越来越重视,最近几年,邀请我参加学术交流的单位也比较多,我成了某种意义下的数学科普工作者。我自认为以变分不等式为工具,用两块黑板,两个小时,就可以把交替方向法连同其线性化的手段讲明白。

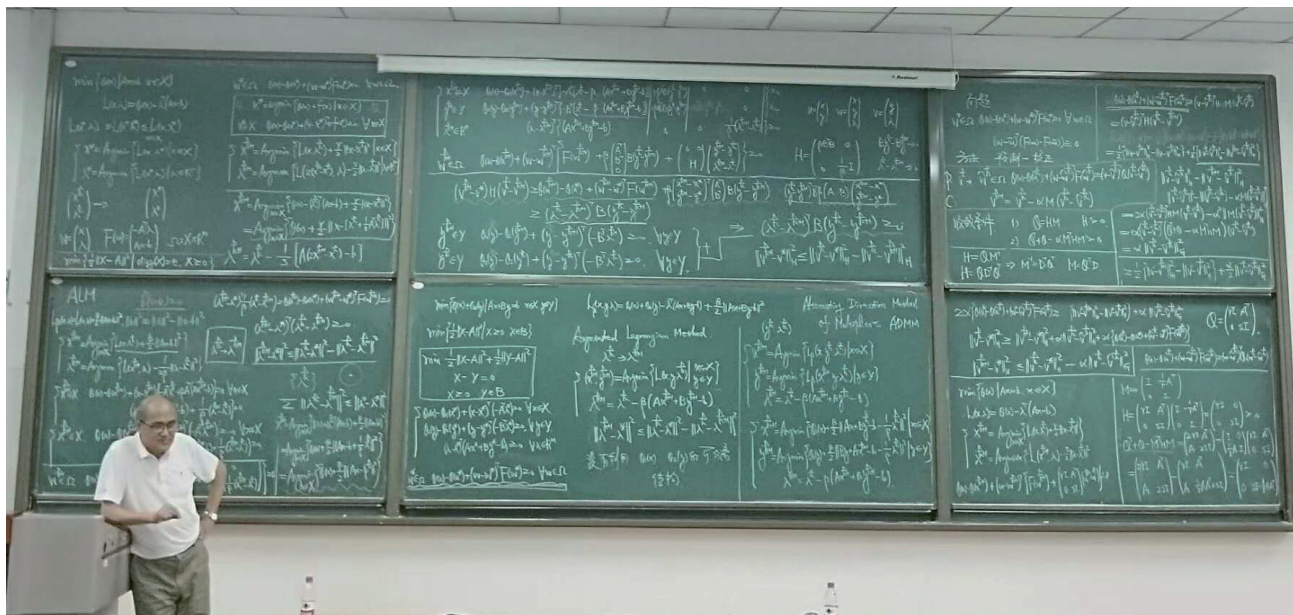
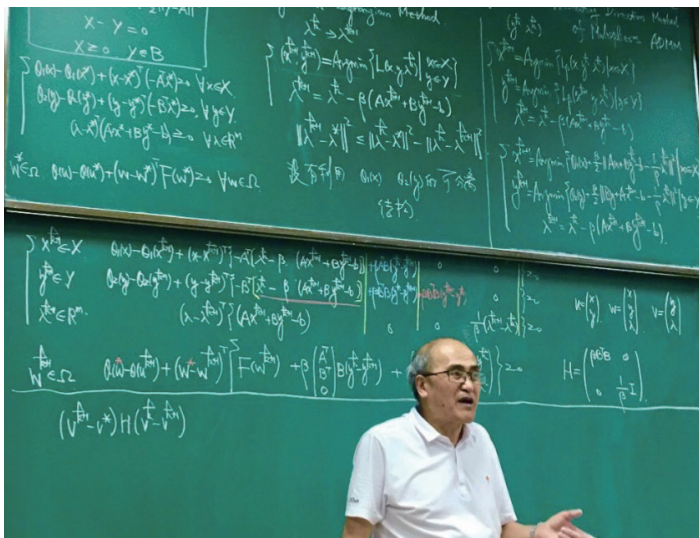
2018年的7月,是我跨越古稀的那个月,我应邀参加了好几个讲习班,作了不少讲座。我在70周岁那个暑假的几站演讲,一定是值得自己日后回味的。

第一站是7月15日在湖南第一师范学院,现任校长童小娇曾是我的博士后,

她说我在这个培训班上“圈了不少粉丝”,我说那得感谢你们提供了机会。的确,有人说已经读过我们的大部分论文,也有人参加一些会议想见我又没有见到,这次培训班终于如愿。学员中有自己就在讲授凸优化的老师,表示一定会在以后的课程中讲授我们的代表性成果,学员中也有从国内名校毕业的博士,他们甚至认为,几十年过后,我们凸优化方法的统一框架会走进教科书。

过誉的话,我照单全收。当然只能当成年轻人的祝福,到底有几分价值,只能由时间来检验。

我在湖南第一师范的24小时课,其中有



照片上的我在想,还有点时间,权当复习,我是否把准备的PPT也拉一下

接连三天，我上下午都讲，每天讲6小时。讲课的，只是体力劳动；听课的，是实实在在的脑力劳动。认真听课的学员，要动脑子消化，其实比讲课的还辛苦一些。我能把体力劳动坚持下来，或许跟年轻时常年超负荷的农业体力劳动有关。童小娇告诉我，培训班结束，学员总结几乎都提到听我课的收获，对年轻人有益就好！

第二站是7月25-26日在西安交通大学，每次讲三小时。有学员拍了张我讲解交替方向法的照片。看得出来，由于设施符合心意，我讲的很投入。我喜欢大黑板加颜色粉笔，有了这些，我就有了自己的标准讲法。对讲解效果，心里就已经有了底。三小时不到，把我准备讲的写满六黑板，听众也基本能跟上。

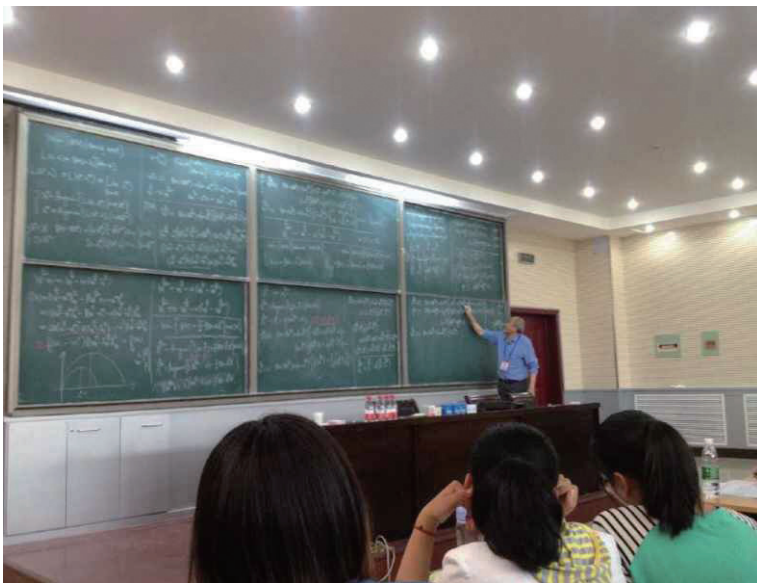
第三站是南京理工大学，计算机专业的部分研究生关心算法的应用，我的演讲则注重给学员讲一些方法的机理。在黑板上勾勾划划，说明这些方法的原理是简单的。掌握这些框架原理，做工程研究的就能够根据自己碰到的问题去设计适当的算法，事成之后，一定会变得更有成就感！个别参加暑期班的博士生，知道我会到暑期班上讲课，带上从我主页上下载的讲义资料向我请教。到各地走走，总发现有人关注我的主页，日后我应该对自己的主页好好维护和更新。

我在ADMM类方法的研究，在相关领域也有一定的影响。《图像处理和数字医药国际会议》已经两次邀请我去交流。2017年的第一届会议他们让我作了个报告，2018年的会议组织者让我去给系列讲座。这个理工医的跨学科的国际会议，邀请了一些老外牛人，有美国、英国、法国和荷兰等国家的一些院士。

三天的会议让我作四个半小时的讲座，也是有点太抬举我。组织者和我都没有想到有这么多人来分会场听我的讲座。有次开讲前座位就已经被占完，后来的人就只能站在会议室门口，听课的有在读研究生，也有博士毕业的教授和副教授。



2018年10月在《图像计算和数字医药》会议期间作系列讲座



2015 年河南大学，边讲边写犹如抹了一面墙

作讲座，传授知识，宾馆，PPT 加白板，效果哪里抵得上教室加黑板？

我讲一个半小时，站的人就站了一个半小时。图片上可以看出，他们往白板越走越近，是因为挤，也因为真心想学。其实，我已经把讲稿的 PPT 发到了他们的手机里。不过，数学的东西，黑板上推导更好一些，便于听众理解。有的与会年轻老师跟我说，他们做图像的都在用这些方法，整天就在调参数，不知所以然，听听我的讲座对以后的研究工作会有帮助。

感谢诸多单位给我提供机会，让我在过 70 岁生日的暑假做了些有意义的科普。能否把在数据科学和人工智能中有些有用的一些算法，用通俗简单的语言，给有点微积分基础的大学生们讲明白？我常常这样想，也这样努力着。

圈子里的人知道我喜欢说笑话，有同学认为已经超出了一个学者的界。我就用哲学家周国平的话为自己辩解：“我厌倦那种永远深刻的灵魂，它是狭窄的无底洞，里面没有光亮，没有新鲜空气，也没有玩笑和游戏。博大的深刻不避肤浅。走出深刻，这也是一种超脱。”

由浙大、上海交大、中山大学、国防科大、南理工、上海大学、河南大学等高校《图像处理》方向的老师发起、轮流举办的《图像处理理论方法及其应用》讲习班，2015 年 5 月由河南大学承办。这次讲习班为期两周，有 100 多位研究生和年轻教师参加，安排给我授课的时间最多。我讲凸优化的分裂收缩算法，讲了五个上午，每次三个小时。我是超级票友，主办方搭台，让我唱戏，唱了个够，很是享受。有听课的研究生照了照片，发给了他自己的老师，后来又传到了我的手里。事后，写下了这篇《教书打工都一样》，作为纪念。正文是：

外套好换，气质难装。我这个教授，农民当的时间长（9 年），做工也在小作坊（2 载）。虽说后来上了大学又留了几年洋，到如今，一抬腿，一开腔，



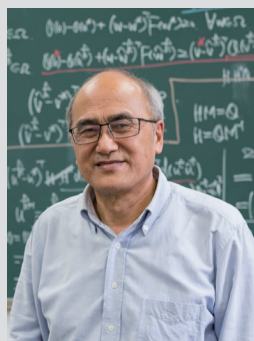
抹完一面墙，不少人照相，抹墙的人，汗水没有白淌

都与农民工大爷没啥两样。就连在家吃早饭，我盛粥还用大茶缸。不信你再看学生给我上课照的像，我手里拿个黑板擦，左看右看都像瓦工在抹墙。

我三十岁才上大学，既没有家学渊源，又不是聪慧过人。同是 77 级本科生的李国平先生的儿子就说过，老三届 77 级读数学的几乎全军覆没，当然，他的意思是说我们这些大龄学生很快就落伍了。即便如此，我认为前面要加个“当了十多年农民，上大学期间有乡下的老婆和年幼的孩子需要牵挂的”这个定语才准确。是啊，虽然我们在大学里也以悬梁刺股的精神学习，毕业后也兢兢业业工作，但毕竟丢失了读数学的黄金年龄，即使能取得一些成绩，也是非常有限的。

常常有相关领域的海外华人学者跟我讨论工作。最近收到素未谋面的芝加哥大学潘晓川教授的来信，潘教授是学物理的，目前是 IEEE TBME（生物医学工程）的主编。他告诉我说，他和他的实验室的同事都读过我和我的学生的一些最优化方面文章。信中除了赞扬我们文章的原创和质量，还说比优化领域的其他文章对他们更有帮助。这些评价，是对我人生最大的安慰。

有人问我学术生涯最深刻的体会是什么，那我要说：走自己的路，独立思考，扬长避短，不骄不躁。即使个人自然条件不好，也能做出一些有益的工作。



作者简介：

何炳生，南京大学数学系教授、博士生导师。2015 年应聘南方科技大学教授，研究领域为最优化理论与方法。