最优化方法I

Operations Research

|  |  |  |
| --- | --- | --- |
| **周次** | **教学内容** | **教学方式** |
| 1 | 最优化问题简介、发展史、分类  最优化发展和数据驱动问题的需求  无约束优化理论  无约束优化的最优性条件、泰勒展开式和中值定理  课程要求：考试和作业 | 讲授 |
| 2 | 约束优化理论  约束优化的最优性条件: 几何条件、一阶条件、切锥、法锥、约束规范（正则点）；Lagrange函数 | 讲授 |
| 3 | 约束优化的最优性条件: 二阶条件、约束规范、临界锥 | 讲授 |
| 4 | 对偶原理Lagrange乘子、鞍点定理、扰动分析  特殊优化问题：线性优化、凸优化的历史地位 | 讲授 |
| 5 | 凸集、凸函数的定义及其基本性质  凸规划的定义与性质、对偶理论、Wolfe dual | 讲授 |
| 6 | 线性规划的基本性质，单纯形方法 | 讲授 |
| 7 | 算法概述和基础、一维搜索、收敛性  使用导数的最优化方法 | 讲授 |
| 8 | 最速下降法，牛顿法 | 曲文涛 |
| 9 | 拟牛顿法 | 任世青 |
| 10 | 共轭梯度法 | 庄丽 |
| 11 | 约束优化问题的罚函数法 | 丹娅静 |
| 12 | ALM增广Lagrangian方法 | 尚盼 |
| 13 | ADMM交替迭代乘子法 | 孙军 |
| 14 | PPA 和PGD近端梯度法 | 李梅 |
| 15 | CDM 坐标下降法 | 陈黄岳 |
| 16 | 最小二乘法、压缩感知和变量选择  总复习 | 讲授 |

# 第0讲 绪论

* 运筹学发展简史
* 最优化简介
* 最优化问题模型和分类

## 一、运筹学发展简史

运筹学是自二十世纪三四十年代发展起来的一门新兴交叉学科。它主要研究人类对各种资源的运用及筹划活动，以期通过了解和发展其基本规律和方法，发挥有限资源的最大效益，达到总体最优的目标。从问题的形成开始，到构造模型、提出解案、进行检验、建立控制，直至付诸实施为止的所有环节构成了运筹学研究的全过程。运筹学研究对象的客观普遍性，以及强调研究过程完整性的重要特点，决定了运筹学应用的广泛性，它的应用范围遍及工农业生产、经济管理、工程技术、国防安全、自然科学等各个方面和领域。

运筹学从创建时期开始起就表现出其理论与实践结合的鲜明特点，在它的发展过程中还充分表现出多学科的交叉结合，物理学家、化学家、数学家、经济学家、工程师等联合组织成研究队伍，各自从不同学科的角度出发提出各自对实际问题的认识和见解，促使解决大型复杂现实问题的新途径、新方法、新理论更快地形成。

运筹学的学科体系主要包含三大部分：模型、理论和算法。无论是早期解决二战中的兵力部署和武器调配，还是生产组织问题或交通、通讯问题，相关领域的运筹学工作者都建立了各种各样的模型，在这些模型下逐步地建立了比较完整的理论体系，提出了求解相应问题的各种类型的高效算法。

运筹学经过七十多年的发展，已经逐步形成了一套系统的研究和解决实际问题的方法，它可以概括为以下五个阶段：

（1）构建所关心问题的数学模型，将一个实际问题表示为一个运筹学问题。

（2）分析问题（最优）解的性质和求解问题的难易程度，寻求合适的求解方案。

（3）设计求解相应问题的算法，并对算法的性能进行理论分析。

（4）编程实现算法，并分析模拟数值结果。

（5）判断模型和解法的有效性，提出解决原始实际问题的具体实施方案。

诚然，以上这五个阶段并不是相互独立的，也决非依次进行的。正如邦德（美国工程院院士；曾任美国军事运筹学会主席和美国运筹学会主席）[1] 在谈到他几十年建模和分析的体会时指出的那样：“对于模型的开发应该是一种连续的研究、开发、分析、改进 …… 的过程，是一个原型化和呈螺旋状发展的过程，而不是一个单个事件！在短期内建造一个原型（假若有必要，加上一些不切实际的假设），然后通过去除那些不切实际的假设，增加过程，增加系统等等不断地将模型改进”。

线性规划是运筹学模型、理论和算法的最典型的代表之一。上个世纪四十年代前学者们缺乏对事物进行优化的兴趣和动力，在文献中虽然有四五十篇关于线性不等式系统的文章，但其中没有一篇提及目标函数。1947年，是丹齐格基于其二战时担任实践计划规划者的经历，认识到多数的实际计划关系都可以用一组线性不等式来刻画，并用一个目标函数来取代为选取一个较好计划而设定的一组基本规则，从而提出了线性规划模型和求解方法——单纯形法。运筹学以后的发展表明，线性规划及单纯形法不仅是证明理论的一个有力的分析工具，还是一个强有力的计算工具，更是运筹学研究的一个催化剂。学者们对线性规划和单纯形法的计算复杂性的持续研究，最终产生了椭球算法和内点算法等一系列理论成果，并形成了新的研究课题。线性规划的产生和巨大成功极大地推动了数学规划发展。

运筹学作为一门新兴的交叉学科，也已在军事国防、企业民生、科技工程、经济金融等领域中产生了深刻而广泛的影响。近年来，世界科技发展突飞猛进，经济全球化愈演愈烈，市场竞争日趋激烈；中国经济已经从计划体制转入市场体制的轨道，并持续高速发展，举国正在实施建设创新型国家的发展战略。在此形势下，我们希望本研究报告能对我国从事运筹学研究、教学和应用的学者、师生和实践者有所启迪，使得我国运筹学工作者能为现代科技的日新月异的发展、社会和经济的可持续发展做出贡献。

### 1、运筹学发展的历程

数学既是所有学科的共同语言，也是有力的工具。数学作为一级学科包括了五个二级学科分支：基础数学、应用数学、计算数学、概率统计和运筹与控制论。尽管运筹学作为一门有着不长历史的新兴交叉学科，已形成了比较完整的学科体系，但它通常还是作为数学的一个分支。实际上，数学与运筹学有着紧密的关系。数学是解决运筹问题和实现运筹思想的最基本的工具之一。运筹学工作者主要用数学方法构造问题模型，建立相应理论，设计和分析求解算法。在这个过程中，他们不仅可以用数学方法解决实际问题，也可以发现新的数学问题，丰富数学的内涵，推动数学的发展。卡斯蒂在其著作中 [2] 列出了二十世纪数学的五大指导理论，其中四个属于运筹学或与运筹学密切相关，对偶定理、极大极小定理、停机定理和不动点定理。下面我们对运筹学的发展做一个简要的回顾[3,4]，籍此可以更好的理解运筹学的内涵和特征。

1.1 运筹学发展简史

“就技术发明对战争的影响或就为未来事件的筹划来说，运筹学可认为是起源于古中国或古埃及”[1]。的确，朴素的运筹思想在中国古代历史发展中源远流长。公元前六世纪的著作《孙子兵法》研究如何筹划兵力以争取全局胜利，是我国古代军事运筹思想最早的典籍。同一时期，我国创造的轮作制、间作制与绿肥制等先进的耕作技术暗含了现代运筹学中二阶段决策问题的雏形。总之，统筹、多阶段决策、多目标优化、合理运输、选址问题、都市规划、资源综合利用等运筹思想方法在中国古代的生产活动和日常生活种屡见不鲜，但很少有人从数学的角度将这些运筹思想和方法进行提升。

西方国家的科学家一方面试图从朴素的运筹问题和运筹思想中发展新的数学内涵，另一方面又试图利用已经建立的数学概念和方法解决实际问题。1736年，欧拉用图论思想成功地解决了哥尼斯堡七桥问题。1738年，贝努利首次提出了效用的概念，并以此作为决策的标准。1777年，布冯发现了用随机投针试验来计算*π* 的方法，这是随机模拟方法最古老的试验。1896年，帕累托首次从数学角度提出多目标优化问题，引进了帕累托最优的概念。1909年，丹麦电话工程师埃尔朗利用概率论，开展了关于电话局中继线数目的话务理论的研究，开创了排队论研究的先河。1912年，策梅洛首次用数学方法来研究博弈问题。

现代运筹的思想萌芽于一战时期，这段时间人们开始用数学的方法探讨各种运筹问题，只是由于人力和经费不足，资料有限等原因限制了运筹学研究的深度。1915年哈里斯对商业库存问题的研究是库存论模型最早的工作。1916年兰彻斯特开展了关于战争中兵力部署的理论，这是现代军事运筹最早提出的战争模型。1921年博雷尔引进了博弈论中最优策略的概念，对某些博弈问题证明了最优策略的存在。1926年博鲁夫卡最早发现了拟阵与组合优化算法之间的关系。1928年，冯·诺依曼提出了二人零和博弈的一般理论。1932年威布尔研究了维修问题和替换问题，这是可靠性数学理论最早的工作。1939年康托罗维奇开创性地提出线性规划，并据此模型研究了工业生产的资源合理利用和计划等问题，因而在1975年获得了诺贝尔经济奖。上述这些先驱性的成就对运筹学的发展有着深远的影响。

现代运筹学真正起源于20世纪二次大战期间，并因其在军事作战方面的大量成功运用而得到蓬勃发展。1935—1938年被视作运筹学基本概念酝酿期。英国为了有效地运用新研制的雷达系统来对付德国飞机的空袭，在皇家空军中组织了一批科学家，进行新战术试验和战术效率的研究，并取得了满意的效果。他们把自己从事的这种工作叫做“Operational Research”(译作“运筹学”)。二战期间英军的每一个大的指挥部大都成立了这种运筹研究小组。在美国和加拿大的军事部门也相继成立了若干运筹研究小组，称之为“Operations Research”。他们广泛地研究有关战果评价、战术革新、技术援助、战略决策和战术计划等问题。美国运筹学会创始人之一莫尔斯在上个世纪五十年代初给运筹学做出了如下定义：“运筹学是为领导机构对其控制下的业务活动作决策时提供定量依据的科学方法”，它反映出运筹学初期的主要作用。

1949年，美国成立了著名的兰德公司，与此同时，许多运筹学工作者逐步从军方转移到政府及产业部门进行研究。在新的、更宽阔的环境中，运筹学的理论和应用研究得到了蓬勃的发展。随之产生的理论成果主要有线性规划、整数规划、图论、网络流、几何规划、非线性规划、大型规划、最优控制理论等；同时也为欧美等国创造了巨大的经济效益和社会财富。

研究优化模型的规划论，研究排队或服务模型的排队论（亦称随机服务系统），及研究博弈模型的博弈论是运筹学最早的三个重要分支，通常称为运筹学早期的三大支柱。 随着学科的发展和计算机的出现，现在分支更细，名目更多；例如线性与整数规划、图与网络、组合优化、非线性规划、多目标规划、动态规划、随机规划、博弈论、随机服务系统、库存论、可靠性理论、决策分析、马尔可夫决策过程、搜索论、随机模拟、管理信息系统等应用基础性学科分支，工程技术运筹学、管理运筹学、工业运筹学、农业运筹学、军事运筹学等交叉与应用学科分支也先后形成。

1.2 中国运筹学发展简史

现代运筹学被引入中国是在上个世纪五十年代后期。中国第一个运筹学小组是在钱学森和许国志两位先生的积极推动下，在1956年于中国科学院力学研究所成立。钱学森先生在麻省理工学院取得硕士学位，在加州理工大学取得博士学位后成为该校的第一位戈达德讲座教授。许国志先生在堪萨斯大学取得博士学位后，在马里兰大学流体力学和应用数学研究所当研究员。他们两人于1955年回到祖国致力于新中国的科技事业。可见在中国运筹学一开始就被理解为与工程有密切联系的学科。

1959年，第二个运筹学部门在中国科学院数学研究所成立，这是大跃进中数学家们投身于国家经济建设的一个产物。力学所小组与数学所的小组于1960年合并成为数学研究所的一个研究室，当时的主要研究方向为排队论、非线性规划和图论，还有人专门研究运输理论、动态规划和经济分析（如投入产出方法）。1963年是中国运筹学教育史上值得一提的一年，数学研究所的运筹学研究室为中国科技大学应用数学系的第一届学生（1958届）开设了较为系统的运筹学专业课，这是第一次在中国的大学里开设运筹学专业和讲授运筹学的课程。今天在中国，运筹学的课程已成为大多数大学的商学院、工学院乃至数学系和计算机系的基本课程了。

上个世纪五十年代后期，运筹学在中国的应用集中在运输问题上。其中一个代表性工作是“打麦场的选址问题”，主要研究和解决在手工收割为主的情况下如何节省人力。此外，国际上著名的“中国邮路问题”模型也是在那个时期由管梅谷教授提出的。可以看出现在非常热门的“物流学”，在当时就形成一些研究雏形，但可惜中国在计划经济体制下，现代化工业水平不高，使我国在相当长的时期中远离了当代“物流学”的发展主流。

中国运筹学早期普及与推广工作的亮点是由华罗庚先生点燃的。在文化大革命期间，他身为中国数学会理事长和中国科学院数学所所长，亲自率领一个小组，大家称为“华罗庚小分队”，到农村、工厂讲解基本的优化技术和统筹方法，应用于日常的生产和生活中。自1965年起的十年中，他到了约二十个省和无数个城市，受到各界人士的欢迎，他的辛勤劳动得到了毛泽东主席的肯定和表扬。华罗庚先生这一时期的推广工作播下了运筹学哲学思想的种子，大大推动了运筹学在中国的普及和发展。直到今天，许多中国人还记得“优选法”和“统筹法”。

上个世纪六、七十年代，中国各种政治运动不断，机关单位、科研院所、工矿企业都受到了巨大冲击。令人钦佩的是，在如此动荡和艰苦的环境下，许国志和越民义等中国运筹学的开拓者们始终未停止运筹学的研究和实践。他们在排队论的瞬时概率性态问题、非线性规划梯度算法收敛问题、组合优化中的排序问题等方面取得了一批重要成果，得到了国外同行的关注和好评。美国数学会1977年出版的访华报告中指出：“在应用数学方面，中国在诸如排队论等领域已十分讯捷地达到了这些领域的前列”。相关成果在1978年全国科学大会上获得大会奖和中国科学院重大成果奖；也为中国运筹学的发展打下了坚实的基础，同时培养了一批运筹学的学科带头人和研究骨干。

自上个世纪八十年代以来，随着改革开放，国内外学术交流不断增加。中国运筹学有了快速地发展，运筹学工作者取得了一批有国际影响的理论和应用成果。例如，将全局最优化、图论、神经网络等运筹学理论及方法应用于分子生物信息学中的若干应用基础性问题的研究中；将优化及决策分析方法，应用于金融风险控制与管理、资产评估与定价分析模型等相关问题研究中；将随机过程方法应用于排队网络的数量指标分析中；将随机动态规划模型应用于供应链管理中的多重决策的最优策略计算中。特别是，运筹学工作者因在组合优化、生产系统优化、图论、非线性规划和城市交通领域的突出贡献曾先后获得国家自然科学奖二等奖五项，因在经济信息系统评估和粮食产量预测方面取得突出成绩曾先后获得国际运筹学会联合会运筹学进展奖一等奖二项。

此外，中国运筹学工作者继续坚持运筹学研究与国民经济建设等重大项目和问题紧密结合。他们在山东省与大连市经济发展计划的制定，兰州铁路局铁路运输的优化安排，中外合资经营项目经济评价，宝钢和武钢等大型企业的调度优化，若干国家重大工程中的综合风险分析等方面，都发挥了积极的作用，产生了良好的经济效益和社会效益。

最后值得一提的是，在中国运筹学几十年的发展过程中，中国运筹学会起到了非常重要的作用。中国运筹学会于文化大革命后的1980年成立，当时它作为中国数学会的一个分会。第一届全国大会在山东省济南市召开，华罗庚被选为第一届理事长，副理事长有许国志和越民义。中国运筹学会在1982年加入国际运筹学联合会，成为其成员。在时任中国科协主席钱学森先生的大力支持下，1992年中国运筹学会获批从中国数学会独立出来，成为国家一级学会。这是中国运筹学发展史上的一个重要事件，它凸显了运筹学以数学为基础，但与数学学科有本质不同的特征。目前，中国运筹学会有十四个专业分会，涵盖了现今运筹学的大多数分支方向，它们在未来中国运筹学的发展中将起到更大的作用。

### 2、运筹学发展趋势

数学内部各分支的相互交叉与融合曾经带来意想不到的成就，数学和应用领域之间的大量相互影响也在科学工程、经济发展、国防安全等方面发挥了重要作用。科学技术近些年的进步与巨大发展，从未像现在这样让我们认识到运筹学对交叉科学研究与应用已经和将带来的深刻影响， 特别是运筹学与生命科学、网络科学、管理科学、服务科学、经济学、大数据科学与技术等领域的交叉。运筹学的广泛应用使得它和其他科学领域的交叉将日益加强。这些交叉不仅为运筹学的应用提供了很好的舞台，同时也为运筹学的新兴分支的产生和发展提供了土壤。其中产生的众多问题和相关研究极大地推动了运筹学的发展。

运筹学经过七十多年的发展，其理论越来越艰深，应用愈来愈广泛，目前已经没有任何一个人可以是运筹学所有方向的专家。因而对未来运筹学的任何一个具有挑战性的课题的研究，尤其是对出现在新的学科交叉领域的重大问题的探索，就更需要一组具有运筹学的不同专长的人才组成的类似于运筹学发展初期时的研究团队，其中还应该包含概率论、统计学、经济学、工商管理、计算机科学、行为科学等学科背景的人才，才能做出重要的科学发现和贡献。

十三五规划以来，我国实施国家大数据战略，精心谋划、超前布局、力争主动，加快建设数字中国。大数据发展日新月异，各部委从战略规划、技术能力提升、应用与管理三个层面积极落实推进大数据发展政策，科研单位及大公司相继成立了大数据研究中心，在多个层面，深入了解大数据发展现状和趋势及其对经济社会发展的影响，分析我国大数据发展取得的成绩和存在的问题，更好服务我国经济社会发展和人民生活改善。现代科学技术的发展推动了大数据产业的应用和发展，特别是人工智能的快速发展和广泛应用，包括工业互联网、物联网、智能交通、信息和金融经济等领域。人工智能仿佛一夜之间遍地开花，目前达到了前所未有的高度。例如，AlphaGo 击败围棋大师李世石。现阶段，人工智能已经应用到生活中的各个领域，谷歌、苹果、亚马逊、微软、阿里、腾讯、百度纷纷成立人工智能实验室，应对人工智能时代的到来。以实际应用为背景，产生了一系列大规模数据理论分析和计算等复杂数学优化问题，这涉及机器学习、统计学和运筹学最优化学科的深入交叉融合。机器学习、优化与统计学等学科交叉融合必能推动大数据科学计算和应用的快速发展，为国家经济发展和应用学科的进步提供技术支持，为建设数字中国和实施大数据战略做出贡献。

美国国家工程学院院士、科学院院士和艺术与科学学院院士Michael Jordan 和机器学习大师Tom Mitchell 认为机器学习是计算机科学和统计学的交叉，同时是人工智能和数据科学的核心。机器学习就是从数据里面挖掘出有用的价值。在信息与科学技术高速发展的今天，特别是人工智能的快速发展和广泛应用，大数据无处不在，不仅在人工智能领域[10]，在经济[1]、金融[2]、生物[16,21]、工业[3]、医学[17]、神经科学[26]、图像处理[4]、风险管理[12]等几乎各个领域都有大量的数据出现。这些实际应用产生了一系列大规模数据理论分析和计算等涉及机器学习、统计学和运筹学最优化学科的深入交叉融合复杂问题。越来越多的问题往往是非凸的，很难设计具有全局收敛性的算法。与此同时，当问题规模超大时，需要用几百台机器投入几天到几个月的时间。这其中有许多值得研究的挑战性的前沿交叉课题。简单讲，这些问题具有以下特征：（1）研究问题新：传统的一些模型已经不能很好的解释数据，需要建立新的、以数据为驱动的模型；（2）研究问题难：数据规模比较大，且往往带有一些复杂的约束，如非线性约束，非凸非光滑约束等；（3）快速有效算法缺乏：随着问题规模的增大和发展的需求，需要设计快速有效的算法进行求解。我们知道，数据分析一般先要给数据一个抽象的表示，接着基于表示进行建模和估计模型的参数。统计是建模的主要工具和途径，而模型求解大多被定义为一个优化问题，且问题的解具有某种性质，如稀疏性（低秩性），组效应，组稀疏性等。目前研究比较广泛的统计模型可分为凸优化模型和非凸优化模型。凸模型包括Ridge[8]、LASSO[18]、Elastic-net[26]，Fused LASSO[19]，Group LASSO[22]，Adaptive LASSO[27]，Adaptive Elastic-net[28]，核范数正则化模型[13, 23]，多变量组Lasso 模型[15]，核范数正则化矩阵回归模型[14]，组LASSO 正则化矩阵最小二乘模型[20]，基于奇异值的矩阵回归模型[24]，组稀疏正则化矩阵最小二乘模型[29]，行自适应弹性网模型[30], 行列稀疏低秩矩阵回归模型[31]等。非凸模型包括Bridge[7,9]，SCAD[6]，MCP[25]，极大似然LASSO 模型[11]，低秩半正定约束矩阵回归模型[5]等。以上模型大都以最小二乘为损失函数。在实际问题中，还可以选取最小一乘，分位数，Huber 函数等为损失函数。建立了这些模型之后，就需要算法来求解模型中的未知参数，进而发现隐藏在数据中的有用信息，为我们的决策提供一些帮助。实际上，上述模型和算法问题是目前机器学习、统计和优化等多个学科领域的热点研究对象，例如： 统计和机器学习侧重统计性质、算法模拟和实践数据应用效果分析。目前已经形成了一门新的比较热门的学科领域即统计学习， 其先驱有著名的美国国家工程院院士Vapnik 教授和英国皇家统计学会、国际数理统计协会和美国统计学会会士Hastie、1996 年COPSS 总统奖得主Tibshirani 和美国科学院院士Friedman 教授，代表作有1998 年出版《统计学习理论》和2001年出版《统计学习基础》，以及Hastie、Tibshirani 和Wainwright在2015 年总结出版《稀疏之统计学习:Lasso 和拓展》[32-34] 。优化专家已经致力于大规模优化问题的理论和算法研究，形成了矩阵优化、稀疏优化和大数据的优化等方向，研究成果丰富，有大量的文献，可以参阅莱斯大学网站http://dsp.rice.edu/cs 和博客http：//nuit-blanche.blogspot.com 等。最近，对于统计和机器学习实际问题有效的算法的研究取得了新的进展，如2018 年发表在优化领域顶级期刊《SIAM J. OPTIM.》的文章[35]，指出算法大致可以分为两类，一类是只使用梯度信息的算法，包括大家熟知的交替方向法（ADMM）、加速梯度方法（APG）、GPSR、SPGL1、SpaRSA、FPC\_AS 和 NESTA 等。另一类算法利用具体问题的二阶信息设计算法提高收敛性能，包括半光滑牛顿法、分布式内点算法等。国内许多专家做出了一系列突出贡献，发表了多篇高水平论文，引起了国内外的广泛关注，如最新文献[35-40]。优化与统计学和机器学习等应用学科之间的整合与交叉是当务之急。越来越多的统计和优化专家都以实际问题为出发点，着眼于新的见解、速度和稳健性。机器学习等应用科学领域的专家越来越理性地看待优化工具。目前以大数据统计优化为主题的国内外会议日渐增大，例如2017 年2 月份在美国杜克大学召开的“Workshop on theInterface of Statistics and Optimization (WISO)”上由Michael Jordan、 Arkadi Nemirovski, Alex Shapiro、Martin Wainwright、范剑青、Eldar 教授等最优化、统计学、压缩感知交叉领域领军科学家的大会报告。 2017 年5 月中国运筹学会数学规划年会中安排统计优化短训课程讲座。

根据大数据时代的需求和国际统计和优化交叉学科的发展趋势，北京交通大学优化团队组织2013全国统计优化高级讲习班暨国际研讨会, 由孙捷教授任主席，会议邀请到COPSS 奖得主蔡天文教授(美国宾夕法尼亚大学)、国际优化大师Rockafellar 教授(美国华盛顿大学)、Uryasev 教授（美国佛罗里达大学)、孙德锋教授(新加坡国立大学)等。接着，2015 年组织统计优化国际小型研讨会，会议邀请到戚厚铎教授(南安普顿大学)、孙德锋教授(新加坡国立大学)、Toh Kim-Chuan 教授(新加坡国立大学)、Hui Zou教授(明尼苏达大学)等。2017年组织统计优化国际小型研讨会，邀请到美国斯坦福大学Guenther Walther教授、美国卡耐基梅隆大学Jiashun Jin教授、香港理工大学Defeng Sun教授、美国明尼苏达大学双城分校Hui Zou教授、加拿大的曼尼托巴大学Liqun Wang教授等。

## 二、最优化方法简介

人在优化。投资者寻求创造投资组合，避免过度风险，同时获得高回报率。制造商的目标是在生产过程的设计和操作中发挥最大的效率。物理系统趋向于最小能量的状态，孤立化学系统中的分子相互反应，直到它们的电子的总势能最小化。光线跟随路径，使其旅行时间最小化。生活需要优化，自然界适者生存的本质在某种意义上可以说是优化过程。

最优化作为运筹学的一个分支是近几十年形成的，它主要运用[数学方法](https://baike.baidu.com/item/%E6%95%B0%E5%AD%A6%E6%96%B9%E6%B3%95" \t "_blank)研究各种系统的优化途径及方案，为决策者提供科学决策的依据。从数学意义上说，最优化方法是一种求极值的方法，即在一组约束为等式或不等式的条件下，使系统的目标函数达到极值，即最大值或最小值。从经济意义上说，是在一定的人力、物力和财力资源条件下，使经济效果达到最大（如产值、利润），或者在完成规定的生产或经济任务下，使投入的人力、物力和财力等资源为最少。

最优化将其根源归结为变分法和欧拉和拉格朗日的工作。20世纪40年代线性规划的发展拓宽了优化领域，促进了现代优化理论和实践的进步。公元前500年[古希腊](https://baike.baidu.com/item/%E5%8F%A4%E5%B8%8C%E8%85%8A" \t "_blank)在讨论[建筑美学](https://baike.baidu.com/item/%E5%BB%BA%E7%AD%91%E7%BE%8E%E5%AD%A6" \t "_blank)中就已发现了长方形长与宽的最佳比例为0.618，称为黄金分割比。其倒数至今仍在优选法中得到广泛应用。在微积分出现以前，已有许多学者开始研究用数学方法解决最优化问题。例如[阿基米德](https://baike.baidu.com/item/%E9%98%BF%E5%9F%BA%E7%B1%B3%E5%BE%B7" \t "_blank)证明：给定周长，圆所包围的面积为最大。这就是欧洲古代城堡几乎都建成圆形的原因。但是最优化方法真正形成为科学方法则是在17世纪以后。17世纪，I.牛顿和G.W.莱布尼茨在他们所创建的微积分中，提出求解具有多个自变量的实值函数的最大值和最小值的方法。以后又进一步讨论具有未知函数的函数极值，从而形成[变分法](https://baike.baidu.com/item/%E5%8F%98%E5%88%86%E6%B3%95" \t "_blank)。这一时期的最优化方法可以称为古典最[优化方法](https://baike.baidu.com/item/%E4%BC%98%E5%8C%96%E6%96%B9%E6%B3%95" \t "_blank)。第二次世界大战前后，由于军事上的需要和科学技术和生产的迅速发展，许多实际的最优化问题已经无法用古典方法来解决，这就促进了近代最优化方法的产生。近代最优化方法的形成和发展过程中最重要的事件有：以苏联Л.В.康托罗维奇和美国G.B.丹齐克为代表的[线性规划](https://baike.baidu.com/item/%E7%BA%BF%E6%80%A7%E8%A7%84%E5%88%92" \t "_blank)；以美国[库恩](https://baike.baidu.com/item/%E5%BA%93%E6%81%A9" \t "_blank)和塔克尔为代表的[非线性规划](https://baike.baidu.com/item/%E9%9D%9E%E7%BA%BF%E6%80%A7%E8%A7%84%E5%88%92" \t "_blank)；以美国R.贝尔曼为代表的[动态规划](https://baike.baidu.com/item/%E5%8A%A8%E6%80%81%E8%A7%84%E5%88%92" \t "_blank)；以苏联Л.С.庞特里亚金为代表的[极大值原理](https://baike.baidu.com/item/%E6%9E%81%E5%A4%A7%E5%80%BC%E5%8E%9F%E7%90%86" \t "_blank)等。这些方法后来都形成体系，成为近代很活跃的学科，对促进[运筹学](https://baike.baidu.com/item/%E8%BF%90%E7%AD%B9%E5%AD%A6" \t "_blank)、管理科学、[控制论](https://baike.baidu.com/item/%E6%8E%A7%E5%88%B6%E8%AE%BA" \t "_blank)和系统工程等学科的发展起了重要作用。

最优化是决策科学和物理系统分析中的一个重要工具。为了利用这个工具，我们必须首先确定目标来刻画正在研究的系统性能的定量度量。这个目标可以是利润、时间、势能，或者可以用单个数字表示的任何数量或数量的组合。目标取决于系统的某些特性，称为变量或未知数。我们的目标是找到优化目标的变量的值。通常，变量在某种程度上受到限制或约束。识别给定问题的目标、变量和约束的过程称为建模。构建合适的模型是第一步，有时是优化过程中最重要的一步。如果模型过于简单化，它将不能给实际问题提供有益的见解。如果它太复杂，可能太难解决。一旦模型被公式化，通常借助于计算机可以使用优化算法来找到它的解。没有通用的优化算法，而是一组算法，每个算法都是针对特定类型的优化问题而定制的。选择适合于特定应用程序的算法的责任通常落在用户身上。这个选择很重要，因为它可以确定问题是快速地还是缓慢地解决，以及是否找到了解决方案。在将优化算法应用于该模型之后，我们必须能够识别它是否成功地完成了求解的任务。在许多情况下，存在称为最优性条件的数学表达式，用于检查当前变量集是否确实是问题的解决方案。如果最优性条件不满足，它们可以提供关于如何改进解的当前估计的有用信息。可以通过应用诸如灵敏度分析之类的技术来改进模型，这些技术揭示了解决方案对模型和数据变化的敏感性。根据应用对解决方案的解释还可以建议改进或改进（或校正）模型的方法。如果对模型进行任何更改，则重新求解优化问题，并且该过程可以重复。

### 1.最优化模型

最优化问题在我们的学习生活中经常遇到，中小学的课本中就已经出现最优化问题。下一部分我们给出最优化问题的实际且简单的几种简单实例来帮助读者更好的理解最优化问题。

问题1、给定一根长为L的绳子，将这根绳子围成一个长方形，问如何设计长方形的长和宽使得该矩形的面积最大？

答：设该矩形的长和宽分别为和，则和满足：2(+)=L，。则题目可转换为如下模型：

问题2、假设我们想要用一根绳子围成一个面积为的长方形，问如何选择长方形的长和宽使得绳子的长度最短？

答：设该矩形的长和宽分别为和，则和满足：=，。则题目可转换为如下模型：

问题3、在以原点为圆心，半径为圆上找到一个点使得该点的两个坐标和最小。

问题4、食谱（配食）问题

假设市场上有n种不同的食物，第j中食物每个单位的销售价为（j=1,2，…，n）。研究表明，人体在正常生命活动过程中需要m种基本的营养成分。威力保证人体的健康，一个人每天至少需要摄入第i种营养成分（i=1,2，…，m）个单位。此外，人们还知道第j中食物的每个单位包含第i种营养成分,（i=1,2，…，m，j=1,2，…，n）个单位。

假设一个人摄入的洋洋成分会被人体完全吸收，每天不同书屋的配给量构成一种配食方案（食谱）。食谱问题就是要求在满足人体基本营养需求的前提下，寻找最经济的配食方案。

类似于上述问题中的模型在我们的生活工作中经常碰到，比如：产品设计中，如何搭配各种原料的比例才能既降低成本，又能提高产品质量？金融投资中，如何记性证券组合才能在可接受的风险范围内获得最大的投资回报？

我们将在给定条件下求最大或者最小问题的这种问题成为优化问题。从数学上讲，优化是对其变量进行约束的函数的最小化或最大化。我们使用以下符号：

——x是变量的向量，也称为未知量或参数；

——称为可行域，是决策变量x取值范围的界定。可行域中的点称为可行点。

——f是我们想要最大化或最小化的目标函数A（标量）函数；

——是约束函数，它是x的标量函数，定义了未知向量x必须满足的某些方程和不等式。

使用这个符号，优化问题可以写成如下形式的模型：

（1.1）

其中， 称为目标函数，又称费用函数。

模型（1）为优化问题的一般的数学模型，但是在研究过程中可行域的形式多种多样，为了方便研究，我们将可行域具体概括为一些等式和不等式约束的集合，因此得到有约束优化问题的标准模型为：

（1.2）

对于 称为不等式约束， 称为不等式约束指标集；对， 称为等式约束， 称为等式约束指标集。

对于优化问题（2.2），任意的，点处的不等式约束中有些能取到等号有些则为严格大于0，将点出所有约束函数为0的集合定义为积极约束集，即

。

可行域为满足约束条件的x的集合，即

对于优化模型（1.2），我们主要目的是想要找到可行域中的一个点使得目标函数达到最小，找到的这个点就称为全局最优解。

**定义0.1全局最优解（值）**

设，有，则为（2.1）的全局最优解， 为全

局最优值。

但是在计算过程中，我们发现有些问题不存在全局最优解或者全局最优解很难计算，这个时候我们就希望找到一个点使得该点为局部（邻域内）的最优解。为此，我们首先引入内积和范数的概念，然后给出邻域的定义。

**定义0.2（内积）**

设为维向量,则的内积定义为两个向量对应元素相乘的和，即

我们将向量的模长定义为显然。

**定义0.3（范数）**

设为维向量，则的p-范数（）定义为；

显然的2-范数。常用的范数为p=1，2 ，分别为

**定义0.4（邻域）**

设，则的邻域定义为：。

有了邻域的定义后，我们便可以定义局部最优解。

**定义0.5局部最优解（值）**

设若存在的邻域，使任意的,

有，则为（1.1）的局部最优解， 为局部最优值。

若局部最优解定义中不等号为严格不等号，则可以得到严格局部最优解的定义。

**定义0.6严格局部最优解（值）**

设若存在的邻域，使任意的,

有，则为（1.1）的严格局部最优解， 为严格局部最优值。

### 2．最优化问题分类

最优化问题的分类多种多样，主要根据目标函数以及可行域的不同来进行分类，也有根

据变量的性质进行分类的。优化问题常常根据不同的角度分为线性规划与非线性规划，连续、离散优化或混合整数规划，约束和无约束优化，全局优化与局部优化，随机与确定性优化，凸和非凸优化等。凸性是凸优化的基本概念，许多实际问题都具有这种性质，这使得这类问题在理论和实践上都更容易解决。

下面给出最优化问题的不同分类标准结果。

1、根据有无约束：

约束优化问题：

无约束优化问题：

2、根据约束函数和目标函数线性与否：

线性规划：目标函数和约束函数都是线性的

非线性规划：目标函数和约束函数中至少有一个为非线性的

特别地，若目标函数是二次函数，约束函数为线性函数，则（1.1）为二次规划问题。

3、根据目标函数与可行域的凸性与否：

凸规划：目标函数为凸函数，可行域为非空凸集

非凸规划：至少有一个条件不满足

凸集：若对，，有，则 为凸集。

凸函数： 为凸集， 为定义在上的函数，若对，，有

则 为上的凸函数。

4、根据函数的可微性质：

光滑优化问题：目标函数和约束函数都是连续可微的

非光滑优化问题：两个中至少有一个是不可微的

5、根据可行域中含有可行点的个数：

连续优化问题：可行域中含有无穷多个不可数的点且可行域中的点连续变化。

离散（组合）优化问题：可行域中含有有限或可数个点。

常见整数规划和0-1规划

6、根据变量的确定性：

确定性规划问题：（1.1）中所有系数都是确定的

不确定性规划问题：（1.1）中某些系数具有某种不确定性。

常见随机规划和模糊规划。

非线性最优化问题是最一般的最优化问题，而线性规划和二次规划问题都是相当重要的特殊的最优化问题，因为在实际中形成的许多最优化问题都是线性规划问题或二次规划问题，而且在用迭代法求非线性最优化问题的最优解时我们常常用线性规划或二次规划来局部近似原非线性最优化问题，并通过求所得近似问题的最优解来对已有最优解的估计进行改进。

除了线性规划和二次规划问题之外，目前广泛应用的优化问题还有，半定规划和锥规划，其中半定规划以矩阵作为决策变量，而锥规划以某种特殊结构的代数元素为变量，二者都是相对抽象和理论深刻的问题，在此省略，详细可参考有关文献。

### 3.优化算法简介

优化算法是迭代的。从变量的初始猜测开始，生成一系列改进的估计（称为“迭代”），直到算法终止，希望求得一个解决方案。用于从一个迭代到另一个迭代的策略将一个算法与另一个算法区分开来。大多数策略利用目标函数、约束函数的值，以及这些函数的一阶导数和二阶导数。一些算法累积在前一次迭代中收集的信息，而另一些算法只使用在当前点获得的信息。好的算法应该具有以下特性：鲁棒性、效率、精度,这些目标可能会发生冲突。收敛速度和存储要求、鲁棒性和速度之间的折衷是数值优化的中心问题，需要仔细考虑。

优化的数学理论既用于刻画最优点，又为算法提供理论基础。不掌握支撑理论就不可能对数值优化有很好的理解。因此，我们将首先给出了一个坚实的（虽然不全面）最优性条件和收敛性分析，以此揭示了一些常用的重要算法的优缺点。本讲义将在后面对重要的优化算法详细分析讨论。

参考文献

[1] 中国运筹学会，中国运筹学发展研究报告[J]，《运筹学学报》，26 (3) (2012)， 1－48。

[2] John L. Casti, **Five Golden Rules: Great Theories of 20th Century Mathematics and Why They Matter**, John Wiley & Sons, Inc, 1996. （中译本：《二十世纪数学的五大指导理论：它们为什么至关重要》[M]，叶其孝、刘宝光译，上海教育出版社，2000。）

[3] 胡晓东、袁亚湘、章祥荪，运筹学发展的回顾与展望[J]，《中国科学院院刊》，27 (2) (2012)， 145－160。

《Convex Optimization Algorithm》 Dimitri P. Bertsekas, 2015《Convex Optimization Theory》Dimitri P. Bertsekas， 2009《Numerical Optimization》Jorge Nocedal, Stephen Wright， 1999(2006)《Convex Optimization》 Stephen Boyd, Lieven Vandenberghe， 2004《最优化理论与算法》陈宝林，1989(2005)编著

# 第一讲　无约束优化理论

* 无约束优化的最优性条件
* 一阶必要/充分条件
* 二阶必要/充分条件
* 泰勒展开式
* 中值定理

## 一．最优性条件基础知识

最优性条件是研究优化问题的基础理论， 其可以根据条件以及结论的关系分为：

必要性条件：若为优化问题（1.1）的（局部）最优解，则应满足的条件。

充分性条件：任意满足该条件，则为优化问题（1.1）的（局部）最优解。

另外，最优性条件可以根据利用的函数的性质分为：一阶最优性条件和二阶最优性条件。

我们首先从最简单的情况出发， 即考虑无约束优化问题。 其模型可表示为：

为此，我们回忆一元函数在处的Taylor展开式为：

一元函数在处的Taylor近似为：

Taylor中值定理

对于本章的多元函数，同样有Taylor展开、Taylor近似和中值定理，但是我们首先需要将一元函数中的导数和二阶导数推广到多元函数中，即为梯度和Hessian矩阵。

**定义1.1（多元函数的梯度）**

为多元函数，则在点的梯度定义为：



**定义1.2（多元函数的Hessian矩阵）**

为多元函数函数，则的Hessian矩阵定义为：

多元函数在处的Taylor展开式为：

多元函数在处的Taylor近似为：

中值定理





## 二．无约束优化问题的最优性条件

为了给出无约束优化问题的最优性条件，首先定义函数的下降方向.

**定义1.3（函数的下降方向）**

设是任给一点，,若存在，对任意，有则称d为在点处的下降方向。

**定理1.1：（一阶必要条件）**设函数在点处可微，若是局部极小点，则



**证明：**设,取,则有,

存在,使当时，根据函数的Taylor近似可得

与是局部极小点矛盾。 □

该定理给出的是函数极小值点的必要条件为梯度为0，但并非所有梯度为0的点都是函数的极小值点，因此引出下述定义。

**定义1.4（驻点、鞍点）**

若在点可微，并且则称为的一个驻点（或者平稳点，stationary point）。既不是极小点，也不是极大点的驻点称为鞍点（saddle point）。

**例**：对于二次函数,是它的驻点，但是该点既不是极小点，也不是极大点，所以是的鞍点。

**定理1.2：（二阶必要条件）**设在处二阶可微，若是局部极小点，则，且Hessian矩阵是半正定的。

**证明：**

由定理1.1可得。

设d是任意一个n维非零向量。

根据在二阶可微，且则有

其中当时，。

由是局部极小点，当充分小时，必有。故当时，有。

即是半正定的。 □

**定理1.3：（充分条件）**设函数在处二阶可微，若梯度.，且Hessian矩阵正定，则是严格局部极小点。

**证明：**

对任意的在处二阶可微且知

其中

再根据正定可得

。

因此，存在，使得当时有

。

由d的任意性知，是严格局部极小点。

**定理1.：**设函数在的邻域内二阶可微，若梯度，且Hessian矩阵在该邻域内半正定，则是局部极小点。特别地，对于邻域内的任意点,若是正定矩阵，则是严格的局部极小点。

自己举例？n=1? f=?思考：解？ What? how?Why?

习题：

1. 计算函数的梯度和和Hessian矩阵，并写出相应的二阶近似展开式。

+

1. 自己举例无约束优化问题并求解，可以分别取n=1，２，３?

参考文献

《Convex Analysis and Optimization》 Dimitri P. Bertsekas, 2015《Convex Optimization Theory》Dimitri P. Bertsekas， 2009《Numerical Optimization》Jorge Nocedal, Stephen Wright， 1999(2006)《Convex Optimization》 Stephen Boyd, Lieven Vandenberghe， 2004《最优化理论与算法》陈宝林，1989(2005)编著

# 第十三讲 Least-Squares Problems

## 10.3 ALGORITHMS FOR NONLINEAR LEAST-SQUARES PROBLEMS

(自学 Page254-269)

# 第十四—十五讲

# CHAPTER 17 Penalty and Augmented Lagrangian Methods

# 第十六讲

# 统计应用：CS and Variable Selections