2014年3月 运 筹 学 学 报 第18卷 第1期
Mar., 2014 Operations Research Transactions Vol.18 No.1

线性与非线性规划算法与理论*

戴彧虹[1,†] 刘新为[2]

摘要 线性规划与非线性规划是数学规划中经典而重要的研究方向. 主要介绍该研究方向的背景知识，并介绍线性规划、无约束优化和约束优化的最新算法与理论以及一些前沿与热点问题. 交替方向乘子法是一类求解带结构的约束优化问题的方法，近年来倍受重视. 全局优化是一个对于应用优化领域非常重要的研究方向. 因此也试图介绍这两个方面的一些最新研究进展和问题.

关键词 线性规划，非线性规划，无约束优化，约束优化，交替方向乘子法，全局优化
中图分类号 O221.1, O221.2
2010 数学分类号 90C05, 90C30

Advances in linear and nonlinear programming*

DAI Yuhong[1,†] LIU Xinwei[2]

Abstract Linear and nonlinear programming is a classical branch in mathematical programming. We introduce some backgrounds on linear and nonlinear programming, and some new methods and new research advances in linear programming, unconstrained and constrained optimization. The alternating direction method of multipliers is very efficient in solving some constrained optimization problems with special structure and has been attracted much attentions in recent years. Global optimization is specially important for applications of optimization. These two topics are also involved.

Keywords linear programming, nonlinear programming, unconstrained optimization, constrained optimization, alternating direction method of multipliers, global optimization
Chinese Library Classification O221.1, O221.2
2010 Mathematics Subject Classification 90C05, 90C30

0 研究领域概述

线性规划旨在线性约束条件下求线性目标函数的最小值. 它是运筹学中研究较早、发展较快、应用广泛、方法较为成熟的一个重要分支，广泛应用于军事作战、经济分析、经营管理和工程技术等各个方面. 其想法最早可以溯源到Fourier (1832). Kantorovich (1939) 在《生产组织与计划中的数学方法》中正式提出线性规划问题，Dantzig (1947) 提出了线性规划的一般数学模型和求解线性规划问题的通用方法——单纯形法，为这门学科奠定了基础. Von Neumann (1947) 提出对偶理论，开创了线性规划许多新的研究领域，

收稿日期：2014年1月16日
* 基金项目：国家杰出青年科学基金 (No. 11125107), 国家自然科学基金资助项目 (Nos. 11331012, 81173633, 11271107)
1. 中国科学院数学与系统科学研究院, 北京 100190；Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
2. 河北工业大学理学院, 天津 300401；Faculty of Science, Hebei University of Technology, Tianjin 300401, China
† 通讯作者 Corresponding author, Email: dyh@lsec.cc.ac.cn

扩大了它的应用范围和解题能力. Koopmans (1951) 将线性规划应用到经济领域，为此与Kantorovich一起获1975年诺贝尔经济学奖. 1950年后对线性规划进行大量的理论研究，并涌现出一大批新的算法. Khachian (1979) 第一次提出了求解线性规划的基于椭球算法的多项式时间算法，Karmarkar (1984) 提出了被称为内点法的新的多项式时间算法. 现已形成线性规划多项式算法理论. 线性规划的研究直接推动了其他数学规划问题包括整数规划、随机规划和非线性规划的算法研究. 由于计算机的发展，出现了许多线性规划软件，如MPSX，OPHEIE，UMPIRE，CPLEX，XPRESS，GUROBI等，可以很方便地求解几万个以上变量的线性规划问题.

非线性规划研究多元实函数在一组等式或不等式的约束条件下的极值问题，且目标函数和约束条件至少有一个是未知量的非线性函数. 根据问题特点和对解的要求，它又可以分为无约束优化、约束优化、凸规划、二次规划、非光滑优化、全局优化、几何规划、分式规划和稀疏优化等研究方向. 它是20世纪50年代开始形成的一门学科. 1951年Kuhn 和Tucker发表的关于最优性条件 (后来称为Karush-Kuhn-Tucher条件或Kuhn-Tucker条件) 的论文是非线性规划正式诞生的一个重要标志. 在50年代还得出了可分离规划和二次规划的几种解法，它们大都是基于Dantzig提出的解线性规划的单纯形法. 50年代末到60年代末出现了以拟牛顿方法、罚函数方法等为首的许多解非线性规划问题的有效的算法，70年代又得到进一步的发展，包括增广拉格朗日乘子法和SQP算法. 非线性规划在工程、管理、经济、科研、军事等方面也得到广泛的应用，为最优化设计提供了有力的工具. 20世纪80年代以来，在信赖域法、内点法、无导数方法、稀疏优化、交替方向法等诸多方向取得了丰硕的成果. 目前主要的非线性规划软件和测试环境有CUTEr, IPOPT, LINGO, MOSEK, NLPQLP, EASY-FIT, CVX等.

1 线性规划

标准的线性规划可以表述为

$$\begin{aligned} \min \quad & c^{\mathrm{T}} x \\ \text{s. t.} \quad & Ax = b, \\ & x \geqslant 0, \end{aligned} \tag{1.1}$$

其中$A \in \mathbb{R}^{n\times m}$, $c, x \in \mathbb{R}^m$, $b \in \mathbb{R}^n$. 问题的输入字长L定义为输入该问题所需的计算机字符串长度. 考虑A, b 和c中的数据都为整数或有理数的情形，此时每个数据的长度是其二进制表示的长度，而L相当于所有数据的总和. 如果算法对于任何情况总能在$O(f(n,m,L))$时间内得到该问题的最优解或者确认该问题无解，而$f(n,m,L)$ 为多项式函数，则称此算法为多项式时间算法. 线性规划已有算法的步数均与问题的数据输入字长有关. Kachiyan的椭球算法在最坏情形下需要$O(n^4L)$ 步求解线性规划问题，而每步为字长不超过$O(L)$ 的数据的四则运算，相应的时间复杂度为$O(n^4L^2 \ln L \ln\ln L)$. Karmarkar的内点法在最坏情形下需要$O(n^4L)$ 步求解线性规划问题，每步仍为字长不超过$O(L)$ 的数据的四则运算，其总体时间复杂度改进为$O(n^{3.5}L^2 \ln L \ln\ln L)$, 并且其通常表现远胜于其在最坏情形下的预测. Renegar[1]提出的路径跟踪法只需$O(n^{3.5}L)$ 步，总体时间复杂度进一步改进为$O(n^3L^2 \ln L \ln\ln L)$. 类似的 (原始对偶) 路径跟踪法在实际应用中表现出色. Spielman和Teng[2]证明了两阶段影子顶点单纯形算法的平均时间复杂度为多项式.

当一个算法能在$O(f(n,m))$ 步内求解线性规划，而其中每一步都为至多$O(g(n,m,L))$长的数据的四则运算，而且f和g均为多项式函数，那么这个算法就被称为强多项式时间算法. 至于是否存在强多项式时间算法，即是否存在计算步数与问题数据输入字长无关，而只与问题维数和约束个数有关的多项式时间算法，是目前线性规划研究领域最基础的问题. 该问题与黎曼猜想，P是否等于NP等问题并列，1998年被Smale[3]列为21世纪18个尚未解决的数学问题. Tardos[4]首次提出一个求解组合线性规划的强多项式时间方法 (组合

线性规划是指线性规划的约束矩阵的元素的大小能够被关于问题维数和约束个数的某个多项式所界定)，Vavasis和Ye[5]基于分层最小二乘的思想，提出了步数只与问题维数、约束个数及约束矩阵有关的多项式时间原始对偶内点算法. Ye[6]对于一类特殊的马尔可夫决策问题证明了使用Dantzig (1947) 的最小简约价格转轴规则的单纯形法具有强多项式计算复杂性. Chubanov[7]对于具有0-1解的一类线性规划问题给出了强多项式时间算法，它可作为用一般线性规划问题构造多项式算法的基础.

在实用算法方面，目前流行的内点法是Mehrotra[8]提出的预测校正方法，尽管在理论上对它所知甚少，但在实际应用中表现非常出色. 单纯形算法的转轴规则对算法效率影响很大，基于[9]和[10]提出的最陡边规则 (steepest edge) 能够大大减少迭代步数，但在每步需要额外的计算量，因此在实际计算中常常使用近似的最陡边规则. 如何设计有效的算法求解更大规模的线性规划问题一直备受关注. 很多来自组合优化的线性规划问题，其变量甚至可能有指数多个. 对于一类特殊的大规模线性规划问题，Muter等[11]提出了一种行与列同时生成的计算方法.

一个与线性规划相关的猜想是Hirsch猜想，它是1957年Hirsch给Dantzig的一封信中提出的(见[12]第521页). 设P是$\mathbb{R}^d$中的一个凸多面体，$\delta_P(x,y)$ 表示P中两个顶点x和y之间的距离，即连接x和y的所有路径中边数的最小值. 凸多面体P的直径指所有顶点之间的距离$\delta_P(x,y)$ 的最大值. 对于$n>d\geqslant 2$, 记$\Delta(d,n)$为$\mathbb{R}^d$中所有正好有n个面 ($d-1$维的边界) 的凸多面体P的直径. Hirsch猜想如下：对于$n>d\geqslant 2$, 均有$\Delta(d,n)\leqslant n-d$. 令人惊讶的是，Santos[13]给出了一个43维具有86个面的凸多面体的例子，其直径是44，因此否定性解决了Hirsch猜想.

2 无约束优化

梯度法、共轭梯度法和拟牛顿方法是三类基本的求解无约束优化问题

$$\min_{x\in\mathbb{R}^n} f(x) \tag{2.1}$$

的一阶导数方法. 我们本节主要回顾与介绍这三类一阶优化方法以及信赖域方法、无导数优化方法的研究进展.

2.1 梯度法

设目标函数在x_k点的梯度为$g_k=\nabla f(x_k)$, 梯度法的迭代格式如下：

$$x_{k+1}=x_k-\alpha_k\, g_k, \tag{2.2}$$

其中，$\alpha_k>0$是按某种方式给出的步长. Cauchy在1847年提出的最速下降法通过精确线搜索选取步长，但收敛速度很慢，受问题的条件数影响很大，并且会出现锯齿现象. BB方法是加拿大数学会前会长Borwein和他的同事Barzilai在1988年提出的一种新型梯度法[14]，其计算效果远好于最速下降法，现已发展为一种求解大规模问题的竞争力很强的方法，在稀疏优化、支撑向量机、图像处理、矩阵方程、数字通讯等许多方面得到应用和推广. 令$s_{k-1}=x_k-x_{k-1}$, $y_{k-1}=g_k-g_{k-1}$, BB方法的基本思想是使矩阵$D=\alpha^{-1}I$在最小二乘意义下满足拟牛顿关系式，从而达到较好的逼近目标函数在x_k点的海色矩阵$\nabla^2 f(x_k)$的效果. 通过求解一元二次函数$\|y_{k-1}-Ds_{k-1}\|^2$ 极小，便得

$$\alpha_k^{\mathrm{BB}}=\frac{s_{k-1}^{\mathrm{T}}s_{k-1}}{s_{k-1}^{\mathrm{T}}y_{k-1}}. \tag{2.3}$$

对于严格凸二次函数的极小化问题，

$$\min_{x\in\mathbb{R}^n}\frac{1}{2}x^{\mathrm{T}}Ax+b^{\mathrm{T}}x, \tag{2.4}$$

其中$A \in \mathbb{R}^{n\times n}$对称正定，$b \in \mathbb{R}^n$, 可证BB方法全局收敛[15]，而且$R$-线性收敛[16]. 然而，由于BB方法强烈的非单调表现，这一R-线性收敛的理论结果远不如最速下降法的Q-线性速度估计，很难反映它远优于最速下降法的计算效果. 更多的理论分析表明，梯度法的有效性与目标函数海色矩阵的特征值相关联. 事实上，对于二次极小化问题，注意到$g_k = Ax_k + b$, 可从 (2.2) 易知

$$g_{k+1} = (I - \alpha_k A)g_k = \cdots = \prod_{j=1}^{k}(I - \alpha_j A)g_1. \tag{2.5}$$

根据线性代数中的Cayley-Halmiton定理，若取$\{\alpha_i : 1 \leqslant i \leqslant n\} = \{\lambda_i^{-1} : 1 \leqslant i \leqslant n\}$, 其中$\lambda_i(1 \leqslant i \leqslant n)$ 为A的n个特征值，则对任意的初始g_1，均有$g_{k+1} = 0$对某$k \leqslant n$成立. 梯度法的该二次有限终止性质最早可见Lai (1983). 与此同时，对A做谱分解，即$A = \sum_{i=1}^{n} \lambda_i u_i u_i^{\mathrm{T}}$，其中$u_i(1 \leqslant i \leqslant n)$ 为n个相互正交的向量，并令$w_i^k = u_i^{\mathrm{T}} g_k$, 则从 (2.5) 知

$$w_i^{k+1} = \prod_{j=1}^{k}(1 - \alpha_j \lambda_i)w_i^1, \quad \forall\, k \geqslant 1, 1 \leqslant i \leqslant n. \tag{2.6}$$

从上式可看出，欲使梯度法具有R-超线性收敛性，对任意的$i \in [1, n]$, 必须有某个无穷子列$\{\alpha_{k_i}\}$, 使之趋向于λ_i^{-1}. [14]证明了当$n = 2$时，BB方法对二次函数具有R-超线性收敛性. [17]分析了这一结果对$n = 3$也成立，同时对任意n 维情形，只要$m \geqslant \frac{n}{2}+1$, 则循环$m$步最速下降步长的梯度法在合适的假定下也具有$R$-超线性收敛性. Fletcher[18]的综述文章给出了关于BB方法很多有趣的洞察.

在推广BB方法用于求解一般无约束优化问题时，常需引入某种非单调线搜索. [19]引入的自适应非单调线搜索允许目标函数值在某些迭代点列任意坏，只要求历史以来最好的目标函数值在几步之内得到改善，因此相比[20]使用要求M步之内最坏目标函数有所改善的传统非单调线搜索，能够很好地保持了BB方法的本性. BB方法的R-线性收敛性能够用于说明，如果恰当选取非单调线搜索中的参数，则当迭代点列靠近解点时，BB试探步长总能够被线搜索接收，因此在每一个迭代步只需计算一次目标函数值. 这一性质类似于 (拟) 牛顿法中的单位步长往往能够被相应的线搜索接收，在很多的数值试验中能够观测得到，可以说是BB方法的一个计算优点. 对于约束优化问题，如果可行集为盒子约束、球约束或者盒子约束再加单个线性约束等形式的集合，这时点到可行集的投影较容易，也已经形成了数值效果较好的投影BB算法.

目前看来，BB方法至少有四个方面的问题仍然值得研究. 第一个方面就是对于一般n元二次目标函数，如何建立BB方法优于最速下降法的理论证据. 第二个方面就是何种BB类型的方法计算效果最好. 目前，已经有许多研究者从特征值的角度、从延迟的角度或者从多点信息的角度等推出过各种类型的BB类型步长，但与一般认为BFGS方法是计算效果最好的拟牛顿法不同，对最优的BB类型步长还未达成共识. 第三个方面是具有单调性质的梯度法研究. 大多数优化应用工作者往往青睐于单调的方法. 文献[21—23]给出了保持单调并且可与BB方法计算效果相媲美的梯度算法，但此方面仍有待更多研究. 第四个方面就是如何推广BB类型方法到各种合适的优化问题，这时候步长的选取也往往需要针对不同的场合加以针对性地研究.

2.2 非线性共轭梯度法

这类方法所需存储少、计算速度快，自从其提出之后就成为求解大规模优化问题一类很有效的方法. 令$d_0 = 0$, 共轭梯度法具有迭代格式

$$x_{k+1} = x_k + \alpha_k d_k, \tag{2.7}$$

$$d_k = -g_k + \beta_k d_{k-1}, \tag{2.8}$$

其中α_k为步长，组合系数β_k为共轭梯度参数. 对于二次函数极小化问题，如果采取精确线搜索，这类方法回退为由Hestenes和Stiefel在1952年提出的求解对称正定线性方程组的(线性)共轭梯度法. 此时，方法所生成的搜索方向将相互共轭，从而具有二次有限终止性质. 对于非线性优化问题，在线搜索非精确时，不同的共轭梯度参数所对应的共轭梯度法其理论性质和数值表现可能很不相同. 以下是选取共轭梯度参数β_k较为有名的4个公式：

$$\beta_k^{\mathrm{FR}}=\frac{\|g_k\|^2}{\|g_{k-1}\|^2},\quad \beta_k^{\mathrm{DY}}=\frac{\|g_k\|^2}{d_{k-1}^{\mathrm{T}}y_{k-1}},\quad \beta_k^{\mathrm{PRP}}=\frac{g_k^{\mathrm{T}}y_{k-1}}{\|g_{k-1}\|^2},\quad \beta_k^{\mathrm{HS}}=\frac{g_k^{\mathrm{T}}y_{k-1}}{d_{k-1}^{\mathrm{T}}y_{k-1}},$$

其中$\|\cdot\|$为二模，$y_{k-1}=g_k-g_{k-1}$.

假设共轭梯度法每步产生的搜索方向均为下降方向，与拟牛顿法一样，共轭梯度法也具有一定的自调节性. 事实上，设$\phi_1=\|g_1\|^2$以及$\phi_k=\prod_{j=2}^k\beta_j^2\ (k\geqslant 2)$, 利用(2.8)可证对所有$k\geqslant 1$,

$$\frac{\|d_k\|^2}{\phi_k^2}=-2\sum_{i=1}^k\frac{g_i^{\mathrm{T}}d_i}{\phi_i^2}-\sum_{i=1}^k\frac{\|g_i\|^2}{\phi_i^2}\leqslant 2\sum_{i=1}^k\frac{(g_i^{\mathrm{T}}d_i)^2}{\|g_i\|^2\phi_i^2}-\frac{1}{2}\sum_{i=1}^k\frac{\|g_i\|^2}{\phi_i^2}. \tag{2.9}$$

从上式可看出，反映方向长度的量$t_k=\frac{\|d_k\|^2}{\phi_k^2}$被反映方向下降程度的和式$S_1=\sum_{i=1}^k\frac{(g_i^{\mathrm{T}}d_i)^2}{\|g_i\|^2\phi_i^2}$的两倍控制，即$t_k\leqslant 2S_1$；而因前者非负，和式$S_1$又不能太小，它总大于或等于和式$S_2=\sum_{i=1}^k\frac{\|g_i\|^2}{\phi_i^2}$的$\frac{1}{4}$，即$S_1\geqslant\frac{1}{4}S_2$. 特别地，如果

$$\sum_{k\geqslant 1}\phi_k^{-2}=+\infty, \tag{2.10}$$

且收敛关系式$\liminf_{k\to\infty}\|g_k\|\neq 0$, 即$\|g_k\|\geqslant\gamma$对所有$k\geqslant 1$和某常数$\gamma>0$成立，则必有$S_2=+\infty$, 从而$S_1=+\infty$. 据此和$t_k\leqslant 2S_1$可推得$\sum_{k\geqslant 1}\frac{(g_k^{\mathrm{T}}d_k)^2}{\|d_k\|^2}=+\infty$, 这就导致与Zoutendijk条件的矛盾. 故运用反证法知，若 (2.10) 成立，方法必给出收敛关系式$\liminf_{k\to\infty}\|g_k\|=0$. Dai[24]严格地建立该一般性收敛结果，并举例说明了它在一定条件下的必要性. 条件 (2.10) 的优点在于它只跟$\{\beta_k\}$有关. 对于FR方法，易知$\phi_k=\frac{\|g_k\|^4}{\|g_1\|^4}$ $(k\geqslant 2)$, 据此和上述一般收敛结果，可以很快建立FR方法的收敛结果. 对于DY方法，利用其等价公式$\beta_k^{\mathrm{DY}}=\frac{g_k^{\mathrm{T}}d_k}{g_{k-1}^{\mathrm{T}}d_{k-1}}$知，$\phi_k=\frac{(g_k^{\mathrm{T}}d_k)^2}{\|g_1\|^4}$ $(k\geqslant 2)$. 另外在条件$0<\gamma\leqslant\|g_k\|\leqslant\bar{\gamma}$ $(\forall k\geqslant 1)$下可证(例见[25]第73页), 对任意的$p\in(0,1)$, 存在正常数δ, 使得对所有$k\geqslant 1$, 关系式$-g_i^{\mathrm{T}}d_i\geqslant\delta$对$[1,k]$中至少$[pk]$个$i$成立. 据此知 (2.10) 成立, 从而建立DY方法的收敛性. Dai[24]通过引入一类新的性质Property (#), 扩充了该一般性收敛结果的适用范围，使之能用于分析PRP方法等几乎所有的共轭梯度法.

前段所述假设共轭梯度法产生的搜索方向均为下降方向，即$g_k^{\mathrm{T}}d_k<0$ $(\forall k\geqslant 1)$. 然而，与梯度法中负梯度方向总是下降方向不同，我们需要对共轭梯度法是否产生下降方向给予特殊的关注. Dai[26]据此将非线性共轭梯度方法归为三类：早期的共轭梯度法、具有下降性质的共轭梯度法以及具有充分下降性质的共轭梯度法. 早期的共轭梯度法包括FR方法、PRP方法以及HS方法等. 这个阶段对于共轭梯度法的研究比较支离破碎，在1991年之前“可能是最未被理解的优化方法”[27]. 早期方法的一个缺点就是，即使采取强Wolfe搜索，它们也可能因为产生上升方向而失败. 这就促使了具有下降性质的共轭梯度法的研究. 一个典型的工作是Dai和Yuan[28]在1999年提出的DY方法，它只需要Wolfe搜索，即可保证每步产生下降方向，并且全局收敛；它和HS方法杂交，可以得到计算效果明显优于PRP方法的共轭梯度算法. 利用共轭梯度法和拟牛顿法的内在联系，可以构造具有充分下降性质的共轭梯度法，即存在常数$c>0$, 使得$g_k^{\mathrm{T}}d_k\leqslant -c\|g_k\|^2$ $(\forall k\geqslant 1)$. Hager和Zhang[29]通过考

虑将无记忆自调比BFGS方法的搜索方向中的第三项简单删去，得到了一个能保证充分下降条件成立的共轭梯度法，并设计了CG_Descent算法. Yu等[30]比较直接地推广了[29]的结果. Cheng和Liu[31, 32]试图先通过$-g_k$和d_{k-1}组合得到与g_k相互垂直的方向$d_k^{\perp}$, 然后再通过将$-g_k$与$d_k^{\perp}$结合定义搜索方向，给出了一种保证充分下降条件成立的共轭梯度法的新技巧. Zhang等[33]通过在PRP方法产生的搜索方向添加一个新的y_{k-1}项，得到了一种保证充分下降条件成立的三项共轭梯度法. Dai和Kou[34]通过将一种最优无记忆自调比BFGS方法之搜索方向投影到由当前梯度方向和前一步搜索方向张成的二维流形，得到了一个如下具有充分下降性质的共轭梯度法：

$$\beta_k^{\mathrm{DK}} = \frac{y_{k-1}^{\mathrm{T}} g_k}{d_{k-1}^{\mathrm{T}} y_{k-1}} - \frac{d_{k-1}^{\mathrm{T}} g_k}{d_{k-1}^{\mathrm{T}} y_{k-1}} \frac{\|y_{k-1}\|^2}{d_{k-1}^{\mathrm{T}} y_{k-1}}. \tag{2.11}$$

基于上述公式，[34]提出了一种非单调的Wolfe型搜索和新的重开始技巧，设计了更加有效的CGOPT算法软件. 公式 (2.11) 最近也被CG_Descent算法采用.

为了得到更为有效的非线性共轭梯度算法，探讨新的共轭条件是一种可能的途径. 通常共轭梯度方向可以写成类似拟牛顿方向的形式，即$d_k = -H_k g_k$. 如果要求H_k对称而且满足拟牛顿关系式，$H_k y_{k-1} = s_{k-1}$, 则有$d_k^{\mathrm{T}} y_{k-1} = -g_k^{\mathrm{T}} H_k y_{k-1} = -g_k^{\mathrm{T}} s_{k-1}$. 引入参变量$r_k$, 我们就可考虑如下一般性共轭条件[35]：$d_k^{\mathrm{T}} y_{k-1} = -t_k g_k^{\mathrm{T}} s_{k-1}$. 如果要求$d_k$具有形式(2.8), 则$\beta_k$具有如下一般的形式：

$$\beta_k^{\mathrm{DL}} = \frac{g_k^{\mathrm{T}} y_{k-1}}{d_{k-1}^{\mathrm{T}} y_{k-1}} - t_k \frac{g_k^{\mathrm{T}} s_{k-1}}{d_{k-1}^{\mathrm{T}} y_{k-1}}. \tag{2.12}$$

Yabe和Takano[36]，Li等[37]通过考虑修正拟牛顿关系式考虑了更多的共轭条件. 一般来说，如何在线性共轭梯度法的基础上，适当地引入某些量，以尽量摄取所要优化问题的非线性性质，将是设计更为有效的非线性共轭梯度法的重要研究方向. 从这个角度，基于子空间极小思想的共轭梯度法[38]、多项共轭梯度法以及非线性预条件共轭梯度法仍有很多的研究空间. 目前高效的共轭梯度算法仍然依赖于相对较为精确的Wolfe类型的线搜索，这使得每步平均需要计算两次目标函数值. 如何减少每步迭代计算量也是值得考虑的重要问题.

2.3 拟牛顿法

Davidon (1959)、Fletcher和Powell (1963) 提出的拟牛顿法，大大地促进了非线性规划的研究. 该工作被Trefethen认为是20世纪数值分析中13个经典工作之一. 目前认为最有效的拟牛顿法是由Broyden, Fletcher, Goldfarb和Shanno在1970年独立提出的BFGS方法. BFGS方法以及求解大规模问题的有限内存BFGS方法，已经被工业与应用数学界广泛地使用. 关于拟牛顿法好的综述可见[27, 39, 40].

关于拟牛顿法的研究工作很多，然而多年以来大家一直关心两个收敛性问题 (如见[27, 40]：第一个是采取Wolfe搜索的DFP方法对一致凸函数是否收敛？另一个是采取Wolfe搜索的BFGS方法对非凸目标函数是否收敛？Dai[41]通过进一步考察Powell[42]关于PRP方法的例子，给出了在6个聚点之间循环的反例，说明采取Wolfe搜索的BFGS方法对非凸函数可能不收敛. Powell[43]证明了采取总取第一个极小点的线搜索时，BFGS方法对二维非凸函数全局收敛. Mascarenhas[44]构造了具有8个聚点的例子，说明采取 (全局) 精确搜索的BFGS方法对非凸函数也可能不收敛. Dai[45]给出了一个性质较为优美的具有8个聚点的反例，它具有如下性质：(a) 总是采取单位步长，而且单位步长总满足Wolfe线搜索条件和Armijo线搜索条件等；(b) 尽管目标函数非凸，每个线搜索函数为强凸函数，并且单位步长是其唯一的极小点, 故此反例也适用于各种类型的精确线搜索 (包括总取第一个极小点的线搜索)；(c) 目标函数是一个38次多项式函数，因此无穷次连续可微, 而如果不要求每个线搜索函数为凸函数，反例中的目标函数可取为六次多项式.

对于非凸目标函数，一些学者试图通过适当修正BFGS方法使之收敛，例如Li和Fukushima[46]. 关于DFP方法的收敛性，我国学者Pu和Yu[47], Yuan[48]等也曾有深入精道的分析，但目前尚不清楚采取Wolfe非精确搜索的DFP 方法对一致凸函数是否收敛. 很多学者引入修正的拟牛顿关系式，并在不同的程度上可以改善原来的BFGS方法. 这方面的第一个工作是Yuan[49]. 目前已从三种途径推广拟牛顿法求解大规模问题: 一是有限内存拟牛顿法; 二是考虑目标函数具有部分可分的结构; 三是假定目标函数的海色矩阵具有稀疏性，即存在某个集合F, 使得$[\nabla^2 f(x)]_{i,j}=0\ (\forall(i,j)\in F)$. 对于第三种途径，早期由Toint以及Fletcher提出的稀疏拟牛顿法要求拟牛顿迭代既满足拟牛顿方程，又满足稀疏性，但常常会数值不稳定. Yamashita[50]提出了一种新颖的思路. 记矩阵函数$\psi(\cdot)=\mathrm{tr}(\cdot)-\ln\det(\cdot)$, 仍记$s_{k-1}=x_k-x_{k-1}$, $y_{k-1}=g_k-g_{k-1}$. 为从H_{k-1}得到新的对$[\nabla^2 f(x_k)]^{-1}$的逼近H_k, 首先我们利用H_{k-1}以及曲率信息 (s_{k-1},y_{k-1}) 做一个普通的DFP拟牛顿迭代

$$H^{\mathrm{QN}}=H_{k-1}-\frac{H_{k-1}y_{k-1}y_{k-1}^{\mathrm{T}}H_{k-1}}{y_{k-1}^{\mathrm{T}}H_{k-1}y_{k-1}}+\frac{s_{k-1}s_{k-1}^{\mathrm{T}}}{s_{k-1}^{\mathrm{T}}y_{k-1}}. \tag{2.13}$$

然后再通过求解以下关于H的子问题得到H_k:

$$\begin{array}{ll}\min & \psi(H_{k-1}^{-\frac{1}{2}}HH_{k-1}^{-\frac{1}{2}})\\ \text{s. t.} & H_{ij}=H_{i,j}^{\mathrm{QN}},\quad (i,j)\in F,\\ & (H^{-1})_{ij}=0,\ (i,j)\notin F,\\ & H\text{半正定}.\end{array} \tag{2.14}$$

当集合F使得对该类型的矩阵做Cholesky分解不会带来新的填充时，问题 (2.14) 可以较为容易求解，并等价于部分位置的元素已知，而添满缺失位置的元素使得矩阵行列式最大的问题. 令人惊讶的是，这种稀疏牛顿法只保证稀疏性，放松了严格满足拟牛顿方程的要求，但Yamashita仍然能够为这种方法建立超线性收敛性. 对这类的方法的研究，仍然值得深入.

2.4 信赖域方法

和线搜索方法一样，信赖域方法也是求解最优化问题的一类重要方法. 但与线搜索方法先确定搜索方向然后选择适当的步长不同，信赖域方法是先确定信赖域半径的大小，然后在以当前迭代点为中心的信赖域中寻找最好的试探步. 如果试探步仍然不够满意，则需要缩小信赖域半径并寻找新的试探步. 这一过程被袁亚湘院士形象地比喻为“瞎子爬山”.

信赖域方法最早出现可以追溯到60多年前的非线性参数估计方法[51]. 在1970年，Powell把它引入无约束最优化算法来保证算法的全局收敛性. Yuan[52]研究了非光滑优化的信赖域方法. 当前，信赖域技术已经被广泛应用于求解约束最优化，许多重要的约束最优化方法如Byrd等[53]的内点方法、 Fletcher和Leyffer[54]的滤子方法和Gould 和Toint[55]的不使用罚函数或滤子的逐步二次规划方法等均通过求解信赖域子问题来产生迭代的试探步. 由于使用了信赖域约束，使得信赖域方法比线搜索方法具有更好的鲁棒性，主要表现在它能够在算法中直接使用最优化问题的二阶信息，即使这些信息不具有线搜索方法所要求的正定性质.

在信赖域方法中，信赖域子问题的有效求解至关重要. 考虑无约束最优化的信赖域子问题

$$\begin{array}{ll}\min & g_k^{\mathrm{T}}d+\frac{1}{2}d^{\mathrm{T}}B_kd\\ \text{s. t.} & \|d\|\leqslant\Delta_k,\end{array} \tag{2.15}$$

向量d_k是它的全局最优解当且仅当存在非负数λ，使得矩阵 $(B_k+\lambda I)$ 是正半定的，且

$$(B_k+\lambda I)d_k=-g_k,\quad \lambda(\|d_k\|-\Delta_k)=0. \tag{2.16}$$

一些直接求解问题 (2.15) 或等价地求解 (2.16) 的有效的方法已经提出. Yuan[56]表明，在B_k正定情况下，使用截断共轭梯度法获得的问题(2.15)的近似最优目标函数值一定不会超过最优目标函数值的$\frac{1}{2}$.

对于约束最优化,信赖域子问题的模型和求解变得更加复杂. Yuan[57],Chen和Yuan[58]等研究了下列等式约束信赖域子问题的性质：

$$\begin{array}{ll} \min & g_k^{\mathrm{T}} d + \frac{1}{2} d^{\mathrm{T}} B_k d \\ \text{s. t.} & \|c_k + A_k^{\mathrm{T}} d\| \leqslant \xi, \\ & \|d\| \leqslant \Delta_k, \end{array} \tag{2.17}$$

得到了许多很有意义的结果. Ni[59]等提出和研究了求解非线性约束最优化的锥信赖域模型.

除了应用于求解经典的无约束和约束最优化问题以外，信赖域方法也广泛应用于求解非线性方程[60]、对称锥规划[61]等大规模特殊结构的约束最优化和Nash平衡问题[62].

信赖域方法的实际应用例子可见文献[51]. 王彦飞和袁亚湘应用信赖域方法研究了不适定反问题中的参数选取和大气图像重构问题.

最新研究进展和研究方向 Cartis等[63]提出了一个求解无约束最优化的使用非线性步长控制的非标准的信赖域方法. 该方法每次迭代求解一个三次正则化子问题

$$\min_{d \in \mathbb{R}^n} \quad f_k + g_k^{\mathrm{T}} d + \frac{1}{2} d^{\mathrm{T}} B_k d + \frac{1}{3} \sigma_k \|d\|^3, \tag{2.18}$$

其中σ_k是一个自适应的正则化参数. 由 (2.18) 可得

$$\|d_k\| \leqslant 3 \max \left\{ \frac{\|B_k\|}{\sigma_k}, \sqrt{\frac{\|g_k\|}{\sigma_k}} \right\}. \tag{2.19}$$

Fan和Yuan[64]建议在信赖域子问题 (2.15) 中选取

$$\Delta_k = \mu_k \|g_k\|, \tag{2.20}$$

其中μ_k的修正类似于经典信赖域半径的修正方法. 一些非标准的信赖域方法更早也出现在求解非线性最小二乘问题和非线性方程中[65−68]. 这类方法都享有信赖域方法所具有的好的全局收敛性质以及最坏情况下的复杂性结果. 最为关键的是，它们都有令人满意的数值表现. 但这些方法却不能像经典的信赖域方法那样来对它们进行全局收敛性分析.

非线性步长控制技术正扩展来求解复合非光滑最优化和无约束多目标最优化问题[69]. Cartis等除了在[70]中给出了无约束最优化信赖域方法最坏情况下的复杂性结果外，在[71]中还研究了非线性约束最优化信赖域方法最坏情况下的复杂性. 有关非线性步长控制算法的局部收敛性可能是下一个研究方向.

2.5 无导数优化方法

很多的优化方法需要利用目标函数的导数信息，然而实际应用的很多问题函数值由实际生产、物理实验或计算机模拟给出，其导数信息几乎不可能得到. 对于无导数优化问题，依赖导数信息的优化方法不能直接应用，需要设计无导数优化方法 (derivative-free optimization methods)，或者称为直接方法 (direct methods). 数学建模的精细化和计算机模拟的广泛应用，对于无导数优化方法的求解速度以及解的质量提出了更高的要求. 为引起大家的注意，这里介绍Audet和Orban[72]利用无导数优化方法调节算法参数的例子. 设某段计算机程序含有一组参数$p \in \mathbb{R}^K$, 其中K是参数个数. 当参数p给定值，我们就可

实现该段计算机程序，并观察到其性能表现. 假设这个性能可以被量化为f, 那么f就构成了p的一个函数. 于是为了最大化程序的性能，我们需要求解如下参数调节优化问题：

$$\max\{f(p) : p \in \mathbb{R}^K \text{并满足可能的某些约束条件}\}.$$

由于f是通过实际测试得到，很难有显示表达式，其梯度更难直接给出，这就是一个典型的无导数优化问题. Audet和Orban[72]利用这一途径研究了带有四个参数的信赖域算法，得到了比经典参数选取更好的参数选取方法.

以Nelder-Mead单纯形方法和Powell共轭方向法为代表的经典无导数方法，在工业界得到广泛使用，吸引了许多著名学者的注意. 尤其是Nelder和Mead在1965年提出的单纯形方法[73]，思想直观，简单易行，截止到2013年12月24日已经被引用17 000多次. 它从$n+1$个初始点构成的单纯形开始迭代，通过将最坏的函数值点对其余n个点构成的剩图的重心作可能的反射 (reflection)、扩张 (expansion)、外收缩 (outside contraction)、内收缩 (inside contraction) 或者整体聚拢 (shrinkage)，从而得到一个新的单纯形. 这一方法基于Spendley等[74]的单纯形方法，但能通过扩张和收缩等步使单纯形的形状更好地符合目标函数的特征：当目标函数的变化比较平缓时，单纯形能够变长；当遇到狭长山谷时，单纯形能够改变搜索方向；当快到解点时，单纯形能够收缩. Woods的博士论文[75]给出了二维非凸函数的例子，使得Nelder-Mead单纯形方法的每步迭代均为收缩步，从而使单纯形所有顶点的聚点都不是最小点. Wright[76]的分析表明，即使对于形态良好的函数，该方法也会停滞不前. McKinnon[77]给出了一族二维严格凸函数的例子，说明单纯形方法可能收敛于一个非稳定点. Lagarias等[78]对一类具有有界水平集和正定海色矩阵的二维函数建立了没有扩张步的单纯形方法的收敛性. 很多数值观察表明，当靠近最优解时，原始的单纯形方法并不出现扩张步，但缺少严格数学证明. 此外，原始的单纯形方法在实际问题中的广泛应用，使得很多学者希望能够对哪怕最简单的二次函数极小化问题

$$\min\ x_1^2 + x_2^2$$

证明其收敛性，但这一问题尚未解决. 原始单纯形方法可能不收敛，究其原因, 一是因为单纯形的几何形状在迭代过程中可能会退化，二是每步迭代缺乏充分下降. 由此出发，有不少的研究者考虑了修正的单纯形方法，使它能够全局收敛到稳定点, 其中包括我国学者早期的工作Yu[79]和华人学者的工作Tseng[80]. 如何在单纯形方法中结合非单调策略使得既保证算法收敛又能提高算法性能是一个值得探索的方向. 关于更多的模式搜索方法和基于线搜索的方法，读者可参考Powell的综述文章[81]以及丁晓东和张在坤的博士论文[82, 83]，在此不再一一赘述.

另一种不计算目标函数导数的优化方法，是在一阶优化方法中的导数用目标函数的向前差分或中心差分代替，但用差分近似导数的计算中需要用到许多的目标函数值的计算，这些目标函数值的计算只用于逼近，显得非常浪费. Powell[84]认为，Nelder-Mead单纯形方法并没有充分利用各个顶点的函数值信息，只是在计算最坏函数值点时加以了利用，而正好可以利用这$n+1$个顶点和这些顶点的函数值进行插值得到唯一的线性插值函数. 当然，线性插值函数的最小点很可能在无穷远达到，因此Powell[84]结合信赖域法对约束优化问题建立了基于线性插值的无导数优化方法. [85]考虑了基于二次插值的方法，但由于插值点集可能不适定而出现病态的情况. Conn等[86]基于[87]的方法思想通过二次插值设计了DFO算法，而Powell[88]同样以二次插值为基础构造了UOBYQA算法. 这些方法因为在进行二次插值时较好地保证插值点集的适定性，因此它们尤其后者的计算表现非常出色. 目前一般认为[89, 90], 基于插值的信赖域算法在迭代中必要时引入某种几何步改善插值点集的适定性，可以改善插值模型，从而改进算法的计算表现. Scheinberg和Toint[91]从理论上证明了基于多项式插值模型的信赖域方法具有内在的自我调节机制，使得插值点集的适定性能在迭代中自动得到改善，但同时也指出，要保证算法全局收敛，几何步不能完全去掉. 基于插值模型的无导数优化方法，目前已经取得了丰硕的成果，并有专著[92]. 但值得指出的是，这些新的无导数方法目前对于无约束优化问题和界约束优化问

题[93, 82]取得成功，如何推广到一般的非线性规划仍然需要很多的研究. 张在坤[83]利用偏微分方程理论中经典的Sobolev范数理论阐述了一类最小范数插值模型，并进而研究了扩展的对称Broyden修正，给出了简单直观的参数选取方式，得到了优于Powell[94]的计算方法.

可以这样说，上面所述的无导数优化方法在计算时往往满足于某个稳定点，尽管数值试验表明它们往往求得一个目标函数值较好的稳定点，例如Powell的UOBYQA算法. 一些智能算法，例如模拟退火算法[95, 96]、遗传算法[97, 98]、禁忌搜索[99]、神经网络方法[100]、粒子群算法[101]等方法, 旨在求全局最优解，但通常需要很长的求解过程. 对于维数高 (Highly dimensional)、函数值计算非常昂贵 (computationally Expensive)、通过黑箱子求得 (Black-box) 的HEB函数，Shan和Wang[102]给了一个较好的总结. 我们在这里着重介绍一个计算表现较好的SRS (Stochastic Response Surface) 算法[103]和一个具有很好理论性质的EGO (Efficient Global Optimization) 算法[104, 105]. SRS算法首先根据正交拉丁方等方法在感兴趣的区域$\mathcal{D}$上产生若干个初始点，并计算这些点的函数值. 假设在第k步，SRS算法通过拟合m个数据点$\{(x_i, f(x_i)) : i = 1, \cdots, m\}$得到了某个响应面函数 (例如径基插值函数) $S_m(x)$，接着算法按照某种分布 (例如均匀分布或正态分布) 在区域$\mathcal{D}$上随机产生p个点$z_1, \cdots, z_p$, 然后根据这些点和$\{x_1, \cdots, x_m\}$的距离以及响应面函数值$S_m(z_j)(j = 1, \cdots, p)$ 的大小从该p个点中挑选一个作为x_{m+1}, 并计算实际函数值$f(x_{m+1})$. 这样就完成一个迭代步. SRS算法每步迭代的计算量并不大，在一定的条件下能够证明以概率1收敛. 数值结果显示, 当目标函数值的计算次数不能很多的情况下, 能得到比Powell的UOBYQA算法更好的解. 目前该算法已经成功推广到约束优化和混合整数规划. 值得提到的是，电子科技大学的王亦伦[106]最近构造了一个最优化HEB函数的统计优化分析与响应面的框架. EGO算法基于以下随机过程方法：

$$y(x_i) = \mu + \varepsilon(x_i) \quad (i = 1, \cdots, m),$$

其中μ是随机过程的期望，m是观察数据的个数，每个数据点的误差$\varepsilon(x_i)$ 服从正态分布，但两个数据点的误差函数之间的协方差并不为零，而是与两个数据点的某种距离$d(x_i, x_j)$有关，即

$$\mathrm{Corr}[\varepsilon(x_i), \varepsilon(x_j)] = \exp[-d(x_i, x_j)].$$

也就是说，当x_i和x_j的距离较小时，协方差将接近于1；相反，如果距离较大，协方差将接近于零. 在对函数$d(x_i, x_j)$ 做某种假定后，可以根据m个数据观察值得到每个点x上的预测值$y(x)$ 以及均方误差$s^2(x)$. 进一步地，我们可以据此求得每个点x相比于目前最好的函数值点的期望改善$EI(x)$. EGO算法于是挑选期望改善最大的点做为下一步待观测的点，即

$$x_{m+1} = \arg\max\{EI(x) : x \in \mathcal{D}\}.$$

在适当的假定下，Bull[105]给出了EGO算法的最优速度分析. Jones, Schonlau和Welch[104]给出了EGO算法的一个近似实现，但目前来看该算法每步的计算量仍然很大，究其原因主要是以上求解x_{m+1}的子问题本身是一个非凸函数的全局最优化问题，比较难于求解. 如何构造算法使之既有EGO算法良好的数值表现，又具有EGO算法优美的理论性质是一个饶有兴趣的问题.

3 约束优化

约束最优化是在约束可行集$\Omega \subset \mathbb{R}^n$上极小化目标函数$f(x)$：

$$\begin{array}{ll} \min & f(x) \\ \text{s. t.} & x \in \Omega. \end{array}$$

一般地，Ω可以表示为一组等式和 (或) 不等式的解集：$\Omega = \{x \in \mathbb{R}^n : c_i(x) = 0, i = 1, \cdots, m;\ c_{m+j}(x) \leqslant 0, j = 1, \cdots, p\}$，且为了区别于线性规划和凸规划，常常假定$f$, $c_i(i = 1, \cdots, m)$ 和$c_{m+j}(j = 1, \cdots, p)$ 中至少有一个是非线性非凸函数.

3.1 逐步二次规划方法

逐步二次规划方法是目前认为求解非凸约束最优化问题最有效的方法，它也曾被称为约束拟牛顿法和Han-Powell-Wilson方法. 在每次迭代，它需要求解一个二次规划子问题

$$\begin{array}{ll} \min & g_k^{\mathrm{T}} d + \frac{1}{2} d^{\mathrm{T}} B_k d \\ \mathrm{s.\,t.} & c_i(x_k) + \nabla c_i(x_k)^{\mathrm{T}} d = 0, i = 1, \cdots, m, \\ & c_{m+j}(x_k) + \nabla c_{m+j}(x_k)^{\mathrm{T}} d \leqslant 0, j = 1, \cdots, p, \end{array}$$

其中B_k是拉格朗日函数的海色矩阵或者它的一阶拟牛顿矩阵近似. 理论上，算法的全局收敛性只需要B_k是有界的，并且在积极约束梯度的零空间正定. 局部只要在解x^*附近满足

$$\lim_{k \to \infty} \frac{\|P_k(B_k - W^*)d_k\|}{\|d_k\|} = 0$$

(其中P_k是x_k点处的积极约束梯度投影矩阵，d_k是第k次迭代的搜索方向，W^*是x^*处的拉格朗日函数的海色矩阵)，算法还可获得快速超线性收敛速率.

罚函数在发展约束最优化方法过程中一直扮演着十分重要的角色. 许多经典的约束最优化方法，如增广拉格朗日方法，就是通过直接求一系列罚函数的 (近似) 最优解来逼近原问题的解. 逐步二次规划方法是以罚函数作为效益函数，在选择步长时使得罚函数充分下降，从而实现算法的全局收敛. 但在实际计算过程中，由于初始罚参数选取不适当，可以导致逐步二次规划方法找不到解. 由Fletcher和Leyffer[54]首创的滤子方法正是为克服这个失败提出的. 随后，Fletcher等[107]、Ulbrich等[108]、Wächter和Biegler[109]、Ribeiro等[110]，以及国内最优化专家孙文瑜教授、濮定国教授的团队等相继提出和发展了各种求解非线性约束最优化问题的有效的滤子方法及其相关的全局收敛性理论，一些滤子方法也被证明具有局部快速收敛性质. 牛凌峰与袁亚湘、王晓与袁亚湘研究了使用增广拉格朗日函数作为效益函数的逐步二次规划方法.

Gould和Toint[55]在2009年提出和发展了一个求解等式约束最优化的不使用罚函数或滤子技术的信赖域逐步二次规划方法及其全局收敛性理论，并报告了大量数值结果. 数值测试表明该方法对求解标准测试问题集CUTE中的问题非常有效. Liu和Yuan[111]在2011年提出和发展了一个求解等式约束最优化的不使用罚函数或滤子技术的使用线搜索的逐步二次规划方法及其全局和局部收敛性理论. 考虑等式约束最优化

$$\begin{array}{ll} \min & f(x) \\ \mathrm{s.\,t.} & c(x) = 0, \end{array}$$

其中$c(x) = (c_i(x), i = 1, \cdots, m)$. 假设$d_k$是当前迭代点$x_k$处的搜索方向，令$v(x) = \|c(x)\|$，我们选择步长$\alpha_k \in (0, 1]$满足

$$f(x_k + \alpha_k d_k) - f(x_k) \leqslant \min\{\sigma \alpha_k g_k^{\mathrm{T}} d_k, -\xi_1 v(x_k + \alpha_k d_k)\}, \tag{3.1}$$

$$v(x_k + \alpha_k d_k) \leqslant \tau_k v_k^{\max} \quad \text{如果} \quad v_k^{\max} \neq 0; \tag{3.2}$$

或者

$$v(x_k + \alpha_k d_k) - v(x_k) \leqslant \min\{\sigma \alpha_k (\|c(x_k) + \nabla c(x_k)^{\mathrm{T}} d_k\| - \|c(x_k)\|), -\xi_2 \alpha_k^2 \|d_k\|^2\}, \tag{3.3}$$

其中$0 < \xi_1 < \xi_2 < 1$，$\tau_k \in [0.95, 1)$. 选取$v_0^{\max} = 0$，并在α_{k-1}满足 (3.1) 和 (3.2) 但α_k满足 (3.3) 时修正$v_{k+1}^{\max} = v(x_k)$，这样做的目的是尽可能避免因为 (3.2) 导致步长变得更小.

容易注意到，上述步长准则允许约束违背度量非单调下降，但对于目标函数值在整个迭代过程中是否下降没有明确的要求. 这与使用非单调线搜索的逐步二次规划方法是不同的[112]. 而且，使用步长准则 (3.1)—(3.3)，使得我们可以在算法的全局收敛性分析中移除要求迭代点序列$\{x_k\}$有界的普遍性假设. 对于经典的使用罚函数为效益函数的逐步二次规划方法的收敛性分析要移除这个假设是非常困难的[113].

结合求解松弛的可行二次规划子问题

$$\begin{array}{ll} \min & g_k^{\mathrm{T}} d + \frac{1}{2} d^{\mathrm{T}} B_k d \\ \text{s. t.} & \nabla c_k^{\mathrm{T}} d = \nabla c_k^{\mathrm{T}} q_k, \end{array}$$

其中q_k是约束最小二乘问题

$$\begin{array}{ll} \min & \|c_k + \nabla c_k^{\mathrm{T}} d\| \\ \text{s. t.} & \|d\| \leqslant \xi \|\nabla c_k c_k\| \quad (\xi \geqslant 1) \end{array}$$

的一个近似解，可以证明，算法具有下列强收敛性质：如果$\|\nabla c_k c_k\| \geqslant \eta \|c_k\|$ (其中$\eta > 0$是一个常数)，那么一定存在迭代点列$\{x_k\}$的一个聚点，它是一个Karush-Kuhn-Tucker(KKT)点; 否则，存在$\{x_k\}$的一个聚点，它或是一个不满足线性无关约束规范的可行的Fritz-John(FJ)点，或是$\min \|c(x)\|$的一个稳定点 (也称为约束最优化的一个不可行稳定点).

上述全局收敛性结果的重要意义之一在于，它表明了算法是否收敛到一个KKT点，主要取决于一些困难约束的梯度的线性无关性，以及用户所能接受的约束Jacobi矩阵最小奇异值η的大小. 该方法与内点方法结合可推广来求解带有不等式约束最优化问题. 其他相关的研究还有Ulbrich和Ulbrich[114]. 近年来国内濮定国教授和陈中文教授等也提出了一些新的既不使用罚函数也不使用滤子的求解约束最优化问题的算法.

3.2 内点方法

内点方法是除逐步二次规划方法之外的另一种求解不等式约束最优化的方法. 从某种意义上说，它也是结合内点技术的逐步二次规划方法. 内点方法最早源于20世纪60年代提出的求解约束最优化的对数障碍方法. 由于在最优解附近可能遇到坏条件问题，该方法没有引起足够的重视. 这种情况直到20世纪80年代，Karmarkar提出线性规划的内点方法并证明该方法具有多项式时间复杂性以后才发生改变. El-Bakry等[115]基于最优性条件提出了求解非凸约束最优化的内点方法并建立了算法的全局收敛性，Byrd等[53]提出并发展了求解非线性规划的信赖域内点方法，Liu和Sun[116]提出和发展了基于线搜索的求解不等式约束最优化的原始对偶内点方法. 目前的研究表明，在数值表现上内点方法与逐步二次规划方法具有可比性.

约束正则性包含积极约束梯度的线性无关性、积极约束函数凸性和不等式约束严格可行性等性质. 约束最优化基本理论是在约束正则性假设下建立的. 约束正则性影响拉格朗日乘子的存在性和有界性，因此在算法设计和收敛性分析中起着举足轻重的作用.

Andreani等[117]考虑了弱约束正则性条件下的增广拉格朗日方法，Andreani等[118]提出了弱约束正则性条件下的算法终止条件并考虑了相应的算法和收敛性. 考虑没有约束正则性假设下的算法有助于开发新的、更具鲁棒性的约束最优化算法，进一步认清算法的实质和正则性在约束最优化算法中的作用.

Liu和Sun[119]提出了一个求解非凸不等式约束最优化的原始对偶内点方法，并在没有约束正则性假设下给出了算法的全局收敛性分析. Liu和Yuan[116]在[119]的基础上，进一步给出了一个求解非凸等式和不等式约束最优化的原始对偶内点方法的统一的框架，并在没有约束正则性假设下给出了算法的全局和局部收敛性分析.

考虑不等式约束最优化

$$\begin{array}{ll} \min & f(x) \\ \text{s. t.} & c(x) \leqslant 0, \end{array} \tag{3.4}$$

其中$c(x)=(c_i(x), i=1,\cdots,m)$. 通过引进非负松弛向量$y=(y_i, i=1,\cdots,m)\in\mathbb{R}^m$，并结合使用对数-障碍松弛方法，得到下列对数-障碍问题：

$$\begin{array}{ll} \min & f(x)-\mu\sum_{i=1}^m \log y_i \\ \text{s. t.} & c_i(x)+y_i=0,\ i=1,\cdots,m. \end{array} \tag{3.5}$$

注意到，原始问题 (3.4) 的KKT条件为

$$g^*+\sum_{i=1}^m \lambda_i^*\nabla c_i^*=0; \tag{3.6}$$

$$\lambda_i^* c_i^*=0,\ c_i^*\leqslant 0,\ \lambda_i^*\geqslant 0,\ i=1,\cdots,m. \tag{3.7}$$

而对数-障碍问题 (3.5) 的KKT条件为

$$g^*+\sum_{i=1}^m \lambda_i^*\nabla c_i^*=0; \tag{3.8}$$

$$y_i^*\lambda_i^*=\mu,\ c_i^*+y_i^*=0,\ y_i^*\geqslant 0,\ \lambda_i^*\geqslant 0,\ i=1,\cdots,m. \tag{3.9}$$

比较 (3.7) 和 (3.9)，可以看出，当μ是充分小的正数时，对数-障碍问题 (3.5) 的KKT点可以看成原始问题 (3.4) 的近似KKT点.

为了保证要求所有迭代点是内点以及缺少约束正则性不会影响算法的表现，[116, 119]在搜索方向的产生和步长选取方面采取了一些与等式约束最优化不同，而与内点技术有关的措施. 我们求解二次规划子问题

$$\begin{array}{ll} \min & g_k^{\mathrm T}\Delta x-\mu\sum_{i=1}^m y_{ki}^{-1}\Delta y_i+\frac{1}{2}\Delta x^{\mathrm T}B_k\Delta x+\frac{1}{2}\sum_{i=1}^m y_{ki}^{-1}\lambda_{ki}\Delta y_i^2 \\ \text{s. t.} & \nabla c_{ki}^{\mathrm T}\Delta x+\Delta y_i=\nabla c_{ki}^{\mathrm T}q_k+v_{ki},\quad i=1,\cdots,m, \end{array}$$

其中 $(q_k, v_k)\in\mathbb{R}^{n+m}$是约束最小二乘问题

$$\begin{array}{ll} \min & \|c_k+y_k+\nabla c_k^{\mathrm T}\Delta x+\Delta y\| \\ \text{s. t.} & \|(\Delta x, Y_k^{-1}\Delta y)\|\leqslant \xi\|(\nabla c_k(c_k+y_k), Y_k(c_k+y_k))\| \quad (\xi\geqslant 1) \end{array}$$

的一个近似解，其中$Y_k=\mathrm{diag}(y_{ki}, i=1,\cdots,m)$. 与线性规划的内点方法不同，在非线性约束最优化的内点方法中，我们对于原始变量和对偶变量分别使用不同的步长准则. 原始步长的选取除了要保证效益函数的充分下降以外，还要保证$y_{k+1}>0$ (即新的迭代点仍然是一个内点) 和$c_{k+1}+y_{k+1}\geqslant 0$. 对偶步长的选取是保证新的对偶变量满足

$$y_{(k+1)i}\lambda_{(k+1)i}\in[\beta_1\mu_k, \beta_2\mu_k],\quad i=1,\cdots,m,$$

其中$\beta_2>1>\beta_1>0$.

在适当假设下，文[116]已经证明，算法具有下列强收敛性质：如果

$$\|(\nabla c_k(c_k+y_k), Y_k(c_k+y_k))\|\geqslant \eta\|c_k+y_k\|$$

(其中$\eta>0$是一个常数)，那么迭代点列$\{(x_k, y_k)\}$的任何聚点都是一个对数-障碍问题 (3.5) 的KKT点；否则，存在序列$\{x_k\}$的一个聚点，它或是原始问题 (3.4) 的一个不满足积极约束线性无关约束规范的可行的FJ点，或是$\min\|\max(c(x),0)\|$的一个稳定点 (即一个不可行稳定点).

最新研究进展和研究方向 Byrd等[120]提出并研究了一个能够快速收敛到最优化问题不可行稳定点的逐步二次规划算法及其超线性收敛性理论. 文献[120]考虑求解不等式约束最优化问题

$$\begin{array}{ll} \min & f(x) \\ \text{s. t.} & c_i(x) \geqslant 0, i = 1, \cdots, m. \end{array}$$

每次迭代它求解两个或多个如下形式的可行的二次规划子问题 (第一个子问题令$\rho_k = 0$)：

$$\begin{array}{ll} \min & \rho_k g_k^{\mathrm{T}} d + \frac{1}{2} d^{\mathrm{T}} W(x_k, \lambda_k; \rho_k) d + \sum_{i=1}^m s_i \\ \text{s. t.} & c_k + \nabla c_k^{\mathrm{T}} d + s \geqslant 0, \quad s \geqslant 0, \end{array}$$

其中$\rho_k > 0$是当前罚参数，$W(x_k, \lambda_k; \rho_k) = \rho_k \nabla^2 f(x_k) - \sum_{i=1}^m \nabla^2 c_i(x_k)$是海色矩阵. 求解子问题的个数取决于由$\rho_k$的选取得到的子问题的解是否能够同时使得约束违背度量和效益函数获得充分下降.

线性和非线性半定规划本质上是一类大规模具有特殊结构的非线性约束最优化问题. 约束最优化方法是求解这些问题的主要的几类方法之一. 然而这方面的研究仅仅是一个开始，在算法和理论方面仍然有大量的工作要做. 另外，一些应用中的大规模非线性约束最优化的快速方法仍然需要结合一些具体的优化技术来处理，如孙聪和袁亚湘最近所做的信息与通讯中的非线性约束最优化问题.

3.3 交替方向乘子法

交替方向乘子法 (alternating direction method of multipliers) 是增广拉格朗日法 (augmented Lagrangian method) 的推广，最早由Glowinski和Marrocco[121]提出 (亦见[122, 123])，目前已经成为函数与变量具有两块或两块以上形式优化问题的基准算法. 考虑如下函数与变量具有K块形式的优化问题：

$$\begin{array}{ll} \min & f(x) = f_1(x_1) + f_2(x_2) + \cdots + f_K(x_K) \\ \text{s. t.} & Ax = A_1 x_1 + A_2 x_2 + \cdots + A_K x_K = b, \\ & x_i \in X_i,\ i = 1, 2, \cdots, K, \end{array} \tag{3.10}$$

其中，每个f_i为闭凸但不必光滑的函数，$x = (x_1^{\mathrm{T}}, \cdots, x_K^{\mathrm{T}})^{\mathrm{T}}$是变量$x$的一个划分，$X = \prod_{i=1}^K X_i$ 是x的可行集，$A = (A_1, A_2, \cdots, A_K) \in \mathbb{R}^{m \times n}$ 是矩阵A的一个划分 (与变量x的划分相匹配)，而$b \in \mathbb{R}^m$是一个向量. 引入增广拉格朗日函数

$$L_\beta(x; \lambda) = f(x) + \lambda^{\mathrm{T}}(b - Ax) + \frac{\beta}{2}\|b - Ax\|^2,$$

其中$\lambda \in \mathbb{R}^m$为拉格朗日乘子，$\beta > 0$为罚参数. 交替方向乘子法的迭代具有如下形式：

$$\begin{array}{rcl} x_i^{k+1} & = & \arg\min\limits_{x_i \in X_i} L_\beta(x_1^{k+1}, \cdots, x_{i-1}^{k+1}, x_i, x_{i+1}^k, \cdots, x_K^k; \lambda^k),\ i = 1, 2, \cdots, K, \\ \lambda^{k+1} & = & \lambda^k + \beta(b - Ax^{k+1}) = \lambda^k + \beta\left(b - \sum_{j=1}^K A_j x_j^{k+1}\right). \end{array} \tag{3.11}$$

对于$K = 1$的情形，交替方向乘子法回退到标准的增广拉格朗日乘子法. 在一定的假设下，可以证明这种类型的对偶梯度下降算法所产生的点列其聚点一定是原问题的最优解[124]. Luo和Tseng[125, 126]给出了对偶误差界的分析，并建立了R-线性收敛速度.

经典的交替方向乘子法研究的是$K = 2$的情形. 在与增广拉格朗日乘子法类似的条件下，可以证明算法的全局收敛性. 从计算效率的角度考虑，算法依赖于参数β的选取以及子问题的求解. 南京大学何炳生教授提出了β 随迭代变化的交替方向法[127]，以及子问

题非精确求解的算法[128]. 交替方向乘子法可以看作一种特殊的Douglas-Rachford分裂算法；对于两个极大单调算子之和求零点的问题，[129, 130]建立了Douglas-Rachford分裂算法的全局收敛性. 进一步地，在一定的假设下，文献[131—133]建立了$O(1/k)$和$O(1/k^2)$的复杂性结果. 这个方面更多的收敛结果需要假定f_1和f_2其中一个是强凸函数或误差界条件. 在此假定以及对约束矩阵秩的一些条件下，文献[134—136]给出了更多的线性收敛结果. 经典交替方向法受到高度重视要归因于其不但可以在压缩传感上有很好的应用，在矩阵完整化、影像处理、统计学中的协方差阵估计等问题中也大有用武之地. 美国Rice大学的张寅和印卧涛、香港浸会大学袁晓明[137]、南京大学杨俊锋[138]和美国科学院院士美国加州大学洛杉矶分校的Osher[139]在推广交替方向乘子法作出了重要的贡献. 关于经典交替方向乘子法的综述可见[140, 141].

对于一些复杂的优化问题，直接求解往往导致子问题非常困难，这时通过引入新的变量可以使目标与约束函数对于变量具有分块的形式. 换句话说，应用中的模型自然地或者经过人为地引入变量后，成为 (3.10) 中$K \geqslant 3$的情形. 对此问题，经典交替方向法的直接推广是否收敛一直是人们关心的一个重要问题. 对所有函数都强凸的情况，[142]给出了其收敛性. 最近，对于$K = 3$ 的凸问题，陈彩华等[143]给出了交替方向乘子算法不收敛的反例，回答了许多人长期关心的交替方向乘子法是否收敛的问题. 同时，他们也给出了该方法收敛的一个充分条件 (A_1, A_2, A_3三个矩阵中任意两个正交). 为利用交替方向乘子法的效率，同时保证算法收敛，南京大学的何炳生、南京师范大学的韩德仁、香港浸会大学的袁晓明等[128, 133, 142, 144]，通过变分不等式的角度提出了一些修正的交替方向算法，那些算法在对目标函数的凸性和可分性的要求下，任意块交替的收敛性和复杂度都是有保证的. 美国明尼苏达大学Hong和Luo[145]提出了如下的修正的交替方向算法：

$$
\begin{aligned}
x_i^{k+1} &= \arg\min_{x_i\in X_i} L_\beta(x_1^{k+1},\cdots,x_{i-1}^{k+1},x_i,x_{i+1}^k,\cdots,x_K^k;\lambda^k),\ i=1,2,\cdots,K,\\
\lambda^{k+1} &= \lambda^k+\tau\beta(b-Ax^{k+1})=\lambda^k+\tau\beta\left(b-\textstyle\sum_{j=1}^K A_j x_j^{k+1}\right).
\end{aligned}
\tag{3.12}
$$

在局部误差界条件以及步长τ充分小的假设下，给出交替方向乘子法的全局收敛性和线性收敛结果. 南京大学的杨俊锋、中科院计算数学所的刘歆和美国Rice大学的张寅，通过把交替方向乘子法看作是广义的Gauss-Seidel迭代与Uzawa方法耦合去求解鞍点方程，在分析中并不需要假定目标函数的凸性与可分离性，但目前只能得到局部收敛性质. 关于多块交替方向乘子法的最新工作可见[146, 147].

4 全局优化

全局优化是非线性规划的一个重要分支，主要研究求解非凸优化问题的全局最优或近似全局最优解. 由于非凸优化问题可能存在多个不同的局部最优点，基于导数信息的KKT最优性条件不再适用于刻画非凸问题的全局最优性，从而，经典的局部优化方法不能保证收敛到全局最优解. 全局优化较系统的研究始于20世纪70年代，Dixon和Szegö编辑的关于全局优化的书综述了全局优化的最近进展，为后来全局优化的深入研究起到了促进作用. Tuy和Horst是早期全局优化研究的先驱者，他们在凹规划和DC规划的系统研究成果使全局优化开始形成一个真正的学科. 由Pardalos在90年代初创立的《Journal of Global Optimization》在全局优化的发展中起到了重要的作用，全局优化作为非线性规划的一个分支开始受到广泛的关注，越来越多的研究者开始从事该领域的研究，特别是对一些具有特殊结构的非凸优化问题的研究取得了许多突破性的成果，如非凸二次规划、非凸多项式规划、机会约束问题的凸逼近，以及在实际应用中遇到的许多特殊形式的非凸优化问题的研究都有很多深刻的研究成果. 基于分枝-定界的全局优化通用软件BARON的发展和其在优化建模系统GAMS和AIMMS中的嵌入应用使学术界和工业界可以方便地求解一定规模的非凸问题的全局解.

全局优化问题的困难在于问题的非凸性使KKT条件一般不足以保证全局最优性，从而无法利用局部优化算法寻找全局最优点. 全局算法需要函数和问题的全局性质，但我们的数学理论很难或无法刻画一般多元函数的全局性质，这是全局优化问题的本质困难所在. 目前，全局优化问题的研究方向大体分为对具有一般特性非凸优化问题的研究和对具有特殊结构的非凸优化问题的研究. 基于剖分区域和凸松弛的分枝-定界算法是求解全局优化问题的基本算法框架. 一般认为，如不利用问题的特殊全局性质，分枝-定界算法在合理时间内能求解的问题规模不大，只有充分利用问题的特殊结构和信息，才有可能设计有效的全局优化算法或快速的近似算法.

凹函数在多面体上求最小的问题称为凹规划，由于凹函数在多面体上的最小值必在顶点达到，我们可以利用顶点枚举或部分枚举的算法策略来设计算法. 这个问题早在20世纪50年代由Hoang Tuy开始研究并提出了著名的凹割方法，割平面算法、外逼近算法和内逼近算法也分别被提出. DC函数是指可以写成两个凸函数之差的函数，DC规划主要研究含DC 函数的非凸优化问题的全局最优解. 目前在线性约束DC优化问题和反凸规划等特殊形式的DC规划研究上有一定的进展，相应的算法有基于单纯形剖分的分枝-定界算法等，但都只能求解很小规模的问题. 如果问题没有特殊的性质和结构可利用 (例如黑箱函数)，随机算法不失为一种选择，这方面的研究有基于各种投点策略的随机投点算法，积分水平集方法的实现也依赖于随机投点. 近年来利用模拟仿真技术提出的全局优化算法具有较好的随机收敛性质，对基于试验的工业设计有很好的应用前景. 当然，基于神经网络和遗传算法思想的随机 (智能) 全局优化算法在工程中也有一些应用. 非凸二次规划是形式上最简单的全局优化问题，具有广泛的应用背景，也是目前被研究得最多的具有特殊结构的非凸优化问题. 几类特殊的非凸二次规划是：箱子约束非凸二次规划、标准二次规划 (标准单纯形上极小化非凸二次函数)、二次指派问题和非凸二次约束二次规划. 非凸二次规划的研究重点是构造有效的凸松弛，利用凸松弛我们可以进一步发展有效的分枝-定界算法和近似算法. 非凸二次规划的半定规划松弛或二阶锥规划松弛是近年来的一个广受关注的研究课题，由于半定规划和二阶锥规划具有多项式时间算法，我们可以有效地得到问题的下界并可以利用随机化算法得到可行解.

以下是全局优化的一些前沿问题：(1) 凸逼近和凸松弛方法. 利用凸优化领域发展的成熟方法进行松弛和逼近是全局优化的一种自然也是有效的算法策略. 线性逼近、凸二次逼近、二阶锥规划逼近和半定规划逼近都是可用的逼近方法. 如何发展更有效的凸逼近方法是一个值得研究的课题. (2) 非凸二次规划. 这可能是在很长一段时间的一个重点研究方向，其原因是二次问题有丰富的全局性质，我们可以利用矩阵分解信息、特征值性质和二次规划与协正矩阵锥的关系来建立非凸二次规划的全局最优性条件或性质，进而设计更有效的精确和近似算法. (3) 基于模拟仿真技术的全局优化算法. 对于没有特殊信息和结构的全局优化问题，基于模拟仿真技术的算法思想可能是有希望的，特别是对低维问题，具有学习特性的模拟仿真随机方法往往简单实用，也能证明随机收敛性. (4) 特殊结构的全局优化问题. 具有特殊结构，尤其是有一定全局性质可以利用的问题将是一个有希望的研究方向，如分式规划、多项式规划、二次指派问题、可分离规划等. 另外，利用现代数学关于函数全局性质的最新成果，如代数几何和多项式代数理论等，也是一个值得注意的研究方向.

5 总结与展望

本文主要介绍了线性规划、无约束优化和约束优化三个方面的最新算法与理论以及一些前沿和热点问题. 对于较为困难的约束优化问题，往往可以通过适当变换化为带结构的约束优化问题，从而可以用交替方向乘子法进行求解. 全局优化是一个对于应用优化领域非常重要的研究方向. 对于一个实际优化问题，例如石油生产计划中的混流优化调度问题，哪怕不能得到全局最优解，一个好的局部解所带来的经济利润也是非常巨大的. 因此本文也试图介绍了这两个方面的一些最新研究进展和问题.

从整体来看，线性与非线性规划的发展呈现以下特点：(1)对于大规模优化问题，如果要提高收敛速度，改善数值算法，如何以较好的方式利用二阶曲率信息非常重要；(2)自适应技术和非单调策略在算法设计中越来越多地被使用；(3)在现有算法的基础上，试图进一步结合启发式算法与随机策略用于求解大规模、复杂优化问题的全局解；(4)设计算法，在较为合理的时间内求得一个不必十分精确但较为满意的解；(5)根据具体问题的实际特点展开理论和算法研究.

稀疏优化是指在某种线性或非线性约束条件下，求一个决策向量使其非零元素的个数达到极小；低秩矩阵优化是指在某种线性或非线性约束条件下，求一个决策矩阵使其秩达到极小. 它们是最优化领域中富有挑战性的热点研究课题，属于运筹学、统计学、管理科学、计算机科学、信息科学等学科的交叉与融合，具有重要的学术意义和广泛的应用背景. 由于大规模优化问题的解往往具有某种稀疏性质，近年来稀疏优化的发展非常迅速，我们已有许多学者在这个方面取得了很好的成果，并有许多综述文章，包括石光明等[148]、文再文等[149]和许志强[150]. 这方面的发展同时带动了非光滑优化和非凸优化方向的发展.

子空间方法也是大规模非线性约束最优化问题数值技术的另一个重要发展方向，关于这方面的一个综述可见[151]. 对于国内优化研究者来说，如何研究和设计较好的一些通用优化算法与软件，或者结合一些实际问题与合作部门共同建立一些行业领域的核心与关键的优化算法与软件也是一个很重要的问题. 由于作者研究领域的局限性以及时间有限，对一些研究方向和研究结果并未能作全面概括和介绍，这里敬请大家包涵与指正.

致谢 非常感谢韩德仁和孙小玲教授在交替方向乘子法和全局优化的撰写中分别提供的大力帮助，同时感谢两位审稿人提供的宝贵建议和意见.

参 考 文 献

[1] Renegar J. A polynomial-time algorithm based on Newton's method for linear programming [J]. *Mathematical Programming*, 1988, **40**: 59-93.

[2] Spielman D A, Teng S H. Smoothed analysis of algorithms: why the simplex algorithm usually takes polynomial time [J]. *Journal of the ACM*, 2004, **51**: 385-463.

[3] Smale S. Mathematical problems for the next century [J]. *Mathematical Intelligencer*, 1998, **20**: 7-15.

[4] Tardos E. A strongly polynomial algorithm to solve combinatorial linear programming [J]. *Operations Research*, 1986, **34**: 250-256.

[5] Vavasis S A, Ye Y. A primal-dual interior point method whose running time depends only on the constraint matrix [J]. *Mathematical Programming*, 1996, **74**: 79-120.

[6] Ye Y. The simplex and policy-iteration methods are strongly polynomial for the Markov decision problem with a fixed discount rate [J]. *Mathematics of Operations Research*, 2011, **36**: 593-603.

[7] Chubanov S. A strongly polynomial algorithm for linear systems have a binary solution [J]. *Mathematical Programming*, 2013, **134**: 533-570.

[8] Mehrotra S. On the implementation of a primal-dual interior point method [J]. *SIAM Journal on Optimization*, 1992, **2**: 575-601.

[9] Goldfarb D E, Reid J K. A practicable steepest-edge simplex algorithm [J]. *Mathematical Programming*, 1977, **12**: 361-371.

[10] Greenberg H J, Kalan J E. An exact update for Harris' TREAD [J]. *Mathematical Programming Study*, 1975, **4**: 26-29.

[11] Muter I, Birbil S. I, Bulbul K. Simultaneous column-and-row generation for large-scale linear programs with column-dependent rows [J]. *Mathematical Programming, Series* A, 2013, **142**: 47-82.

[12] “10000个科学难题”数学编委会. 10000个科学难题 [M]. 北京: 科学出版社, 2009.

[13] Santos F. A counterexample to the Hirsch conjecture [J]. *Annals of Mathematics*, 2012, **176**(2): 383-412.

[14] Barzilai J, Borwein J M. Two-point step size gradient methods [J]. *IMA Journal of Numerical Analysis*, 1988, **8**: 141-148.

[15] Raydan M. On the Barzilai and Borwein choice of steplength for the gradient method [J]. *IMA Journal of Numerical Analysis*, 1993, **13**: 321-326.

[16] Dai Y H, Liao L Z. *R*-linear convergence of the Barzilai and Borwein gradient method [J]. *IMA Journal of Numerical Analysis*, 2002, **26**: 1-10.

[17] Dai Y H, Fletcher R. On the asymptotic behavior of some new gradient methods [J]. *Mathematical Programming, Series* A, 2005, **103**: 541-559.

[18] Fletcher R. On the Barzilai-Borwein method [M]// *Optimization and Control with Applications*, Berlin: Springer-Verlag, 2005, 237-256.

[19] Dai Y H, Zhang H. An adaptive two-point stepsize gradient algorithm [J]. *Numerical Algorithms*, 2001, **27**: 377-385.

[20] Raydan M. The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem [J]. *SIAM Journal on Optimization*, 1997, **7**: 26-33.

[21] Yuan Y. A new stepsize for the steepest descent method [J]. *Journal of Computational Mathematics*, 2006, **24**: 149-156.

[22] Dai Y H, Yuan Y. Analyses of monotone gradient methods [J]. *Journal of Industry and Management Optimization*, 2005, **1**: 181-192.

[23] Zhou B, Gao L, Dai Y H. Gradient methods with adaptive step-sizes [J]. *Computational Optimization and Applications*, 2006, **35**: 69-86.

[24] Dai Y H. Convergence analysis of nonlinear conjugate gradient methods [M]// *Optimization and Regularization for Computational Inverse Problems and Applications*, Berlin: Springer-Verlag, 2010, 157-181.

[25] 戴彧虹, 袁亚湘. 非线性共轭梯度法 [M]. 北京: 上海科学技术出版社, 2000.

[26] Dai Y H. Nonlinear conjugate gradient methods [J]. *Wiley Encyclopedia of Operations Research and Management Science*, Doi: 10.1002/9780470400531.eorms0183.

[27] Nocedal J. Theory of algorithms for unconstrained optimization [J]. *Acta Numerica*, 1992, **1**: 199-242.

[28] Dai Y H, Yuan Y. A nonlinear conjugate gradient method with a strong global convergence property [J]. *SIAM Journal on Optimization,* 1999，**10**(1): 177-182.

[29] Hager W W, Zhang H. A new conjugate gradient method with guaranteed descent and an efficient line search [J]. *SIAM Journal on Optimization*, 2005, **16**: 170-192.

[30] Yu G H, Guan L T, Chen W. F. Spectral conjugate gradient methods with sufficient descent property for large-scale unconstrained optimization [J]. *Optimization Methods and Software*, 2008, **23**: 275-293.

[31] Cheng W Y. A two-term PRP-based descent method [J]. *Numerical Functional Analysis and Optimization*, 2007, **28**: 1217-1230.

[32] Cheng W Y, Liu Q F. Sufficient descent nonlinear conjugate gradient methods with conjugacy conditions [J]. *Numerical Algorithms*, 2010, **53**: 113-131.

[33] Zhang L, Zhou W, Li D. A descent modified Polak-Ribière-Polyak conjugate gradient method and its global convergence [J]. *IMA Journal of Numerical Analysis*, 2006, **26**: 629-640.

[34] Dai Y H, Kou C X. A nonlinear conjugate gradient algorithm with an optimal property and an improved Wolfe line search [J]. *SIAM Journal on Optimization*, 2013, **23**: 296-320.

[35] Dai Y H, Liao L Z. New conjugate conditions and related nonlinear conjugate gradient methods [J]. *Applied Mathematics and Optimization*, 2001, **43**: 87-101.

[36] Yabe H, Takano M. Global convergence properties of nonlinear conjugate gradient methods with modified secant condition [J]. *Computational Optimization and Applications*, 2004, **28**: 203-225.

[37] Li G Y, Tang C M, Wei Z X. New conjugacy condition and related new conjugate gradient methods for unconstrained optimization [J]. *Journal of Computational and Applied Mathematics*, 2007, **202**: 523-539.

[38] Yuan Y, Stoer J. A subspace study on conjugate gradient algorithms [J]. *Zeitschrift für Angewandte Mathematik und Mechanik*, 1995, **75**: 69-77.

[39] Dennis J E, Moré J J. Quasi-Newton method, motivation and theory [J]. *SIAM Review*, 1977, **19**: 46-89.

[40] 袁亚湘. 非线性规划数值方法 [M]. 上海: 上海科学技术出版社, 1993.

[41] Dai Y H. Convergence properties of the BFGS algorithm [J]. *SIAM Journal on Optimization*, 2002, **13**: 693-701.

[42] Powell M J D. Nonconvex minimization calculations and the conjugate gradient method [M]//*Numerical Analysis*, Berlin: Springer-Verlag, 1984, **1066**: 122-141.

[43] Powell M J D. On the convergence of the DFP algorithm for unconstrained optimization when there are only two variables [J]. *Mathematical Programming, Series* B, 2000, **87**: 281-301.

[44] Mascarenhas W F. The BFGS algorithm with exact line searches fails for nonconvex functions [J]. *Mathematical Programming, Series* A, 2004, **99**: 49-61.

[45] Dai Y H. A perfect example for the BFGS method [J]. *Mathematical Programming*, 2013, **138**: 501-530.

[46] Li D H, Fukushima M. On the global convergence of the BFGS method for nonconvex unconstrained optimization problems [J]. *SIAM Journal on Optimization*, 2001, **11**: 1054-1064.

[47] Pu D, Yu W. On the convergence property of the DFP algorithm [J]. *Annals of Operations Research*, 1990, **24**(1): 175-184.

[48] Yuan Y. Convergence of DFP algorithm [J]. *Science in China*, 1995, **38**: 1281-1294.

[49] Yuan Y. A modified BFGS algorithm for unconstrained optimization. [J] *IMA Journal of Numerical Analysis*, 1991, **11**: 325-332.

[50] Yamashita N. Sparse quasi-Newton updates with positive definite matrix completion [J]. *Mathematical Programming, Series* A, 2008, **115**: 1-30.

[51] Conn A R, Gould N I, Toint P L. *Trust-Region Methods* [M]. Philadelphia: Society for Industrial and Applied Mathematics and the Mathematical Programming Society, 2000.

[52] Yuan Y. Conditions for convergence of trust region algorithms for nonsmooth optimization [J]. *Mathematical Programming*, 1985, **31**: 220-228.

[53] Byrd R H, Gilbert J C, Nocedal J. A trust region method based on interior point techniques for nonlinear programming [J]. *Mathematical Programming*, 2000, **89**: 149-185.

[54] Fletcher R, Leyffer S. Nonlinear programming without a penalty function [J]. *Mathematical Programming*, 2002, **91**: 239-269.

[55] Gould N I, Toint P L. Nonlinear programming without a penalty function or a filter [J]. *Mathematical Programming*, 2009, **122**: 155-196.

[56] Yuan Y. On the truncated conjugate gradient method [J]. *Mathematical Programming*, 2000, **87**: 561-571.

[57] Yuan Y. On a subproblem of trust region algorithms for constrained optimization [J]. *Mathematical Programming*, 1990, **47**: 53-63.

[58] Chen X, Yuan Y. A note on quadratic forms [J]. *Mathematical Programming*, 1999, **86**: 187-197.

[59] Ni Q. Optimality conditions for trust-region subproblems involving a conic model [J]. *SIAM Journal on Optimization*, 2005, **15**: 826-837.

[60] Fan J Y, Yuan Y. On the quadratic convergence of the Levenberg-Marquardt method without nonsingularity assumption [J]. *Computing*, 2005, **74**: 23-39.

[61] Lu Y, Yuan Y. An interior-point trust-region algorithm for general symmetric cone programming [J]. *SIAM Journal on Optimization*, 2007, **18**: 65-86.

[62] Yuan Y. A trust region algorithm for Nash equilibrium problems [J]. *Pacific Journal of Optimization*, 2011, **7**: 125-138.

[63] Cartis C, Gould N I, Toint P L. Adaptive cubic overestimation methods for unconstrained optimization. Part I: motivation, convergence and numerical results [J]. *Mathematical Programming*, 2011, **127**: 245-295.

[64] Fan J Y, Yuan Y. A new trust region algorithm with trust region radius converging to zero [C]//*Proceedings of the 5th International Conference on Optimization: Techniques and Applications*, 2001, 786-794.

[65] Zhang J, Wang Y. A new trust region method for nonlinear equations [J]. *Mathematical Methods of Operations Research*, 2003, **58**: 283-298.

[66] Fan J Y. Convergence rate of the trust region method for nonlinear equations under local error bound condition [J]. *Computational Optimization and Applications*, 2006, **34**: 215-227.

[67] Nesterov Y, Polyak B T. Cubic regularization of Newton method and its global performance [J]. *Mathematical Programming*, 2006, **108**: 177-205.

[68] Nesterov Y. Modified Gauss-Newton scheme with worst-case guarantees for global performance [J]. *Optimization Methods and Software*, 2007, **22**: 469-483.

[69] Villacorta K, Oliveira P, Soubeyran A. A trust-region method for unconstrained multiobjective problems with applications in satisfying processes [J]. *Journal of Optimization Theory and Applications*, Doi: 10.1007/s10957-013-0392-7.

[70] Cartis C, Gould N I, Toint P L. Adaptive cubic overestimation methods for unconstrained optimization. Part II: worst-case function-evaluation complexity [J]. *Mathematical Programming*, 2011, **130**: 295-319.

[71] Cartis C, Gould N I, Toint P L. On the complexity of finding first-order critical points in constrained nonlinear optimization [J]. *Mathematical Programming*, Doi: 10.1007/s10107-012-0617-9.

[72] Audet C, Orban D. Finding optimal algorithmic parameters using derivative-free optimization [J]. *SIAM Journal on Optimization*, 2006, **17**(3): 642-664.

[73] Nelder J A, Mead R. A simplex method for function minimization [J]. *The Computer Journal*, 1965, **7**(4): 308-313.

[74] Spendley W, Hext G R, Himsworth F R. Sequential application of simplex designs in optimization and evolutionary operation [J]. *Technometrics*, 1962, **4**: 441-461.

[75] Woods D J. An interactive approach for solving multi-objective optimization problems [D]. Houston: Rice University, 1985.

[76] Wright M. The Nelder-Mead method: numerical experimentation and algorithmic improvements [R]. New Jersey: AT & T Bell Laboratories, 1996.

[77] Mckinnon K I M. Convergence of the Nelder-Mead simplex method to a nonstationary point [J]. *SIAM Journal on Optimization*, 1998, **9**: 148-158.

[78] Lagarias J C, Poonen B, Wright M H. Convergence of the restricted Nelder-Mead method in two dimensions [J]. *SIAM Journal on Optimization*, 2012, **22**: 501-532.

[79] Yu W C. The convergent property of the simplex evolutionary techniques [J]. *Sci Sinica*, 1979, **1**: 269-288.

[80] Tseng P. On the convergence of pattern search algorithms [J]. *SIAM Journal on Optimization*, 1999, **10**: 269-288.

[81] Powell M J D. Direct search algorithms for optimization calculations [J]. *Acta Numerica*, 1998, **7**: 287-336.

[82] 丁晓东. 基于插值模型的无导数优化方法及其应用 [D]. 北京: 中国科学院数学与系统科学研究院, 2010.

[83] 张在坤. 无导数优化方法的研究 [D]. 北京: 中国科学院数学与系统科学研究院, 2012.

[84] Powell M J D. A direct search optimization method that models the objective and constraint functions by linear interpolation [C]//*Advances in Optimization and Numerical Analysis*, Dordrecht: Kluwer Academic, 1994, 51-67.

[85] Winfield D. Function minimization by interpolation in a data table [J]. *IMA Journal of Applied Mathematics*, 1973, **12**: 339-437.

[86] Conn A R, Scheinberg K, Toint Ph L. A derivative free optimization algorithm in practice [C]//*Proceedings of the 7th AIAA/USAF/NASA/ISSMO symposium on multidisciplinary analysis and optimization*, St. Louis: American Institute of Aeronautics and Astronautics, 1998.

[87] Conn A R, Toint Ph L. An algorithm using quadratic interpolation for unconstrained derivative free optimization [C]//*Nonlinear optimization and Applications*, New York: Kluwer Academic Plenum Publishers, 1996.

[88] Powell M J D. UOBYQA: unconstrained optimization by quadratic approximation [J]. *Mathematical Programming*, 2002, **92**(3): 555-582.

[89] Fasano G, Morales J L, Nocedal J. On the geometry phase in model-based algorithms for derivative-free optimization [J]. *Optimization Methods and Software*, 2009, **24**(1): 145-154.

[90] Scheinberg K. Geometry in model-based algorithms for derivative-free unconstrained optimization [J]. *Mathematical Programming Society Newsletter*, 2009, **79**: 1-5.

[91] Scheiberg K, Toint Ph L. Self-Correcting geometry in model-based algorithms for derivative-free unconstrained optimization [J]. *SIAM Journal on Optimization*, 2010, **20**(6): 3512-3532.

[92] Conn A R, Scheinberg K, Vicente L N. *Introduction to Derivative-Free Optimization* [M]. Philadelphia: Society for Industrial and Applied Mathematics and the Mathematical Programming Society, 2009.

[93] Powell M J D. The BOBYQA algorithm for bound constrained optimization without derivatives [R]. Cambridge: University of Cambridge, 2009.

[94] Powell M J D. Beyond symmetric Broyden for updating quadratic models in minimization without derivatives [R]. Cambridge: University of Cambridge, 2010.

[95] Kirkpatrick S. Optimization by simulated annealing: quantitative studies [J]. *Journal of Statistical Physics*, 1984, **34**(5-6): 975-986.

[96] Kirkpatrick S, Gelatt C D Jr, Vecchi M P. Optimization by simulating annealing [J]. *Science*, 1983, **220**: 671-680.

[97] Goldberg D E. Genetic algorithms and machine learning [J]. *Machine Learning*, 1988, **3**(2-3): 95-99.

[98] Goldberg D E. *Genetic Algorithms in Search, Optimization and Machine Learning* [M]. Boston: Addison Wesley Longman, 1989.

[99] Glover F. Tabu search-Part I [J]. *ORSA Journal on computing*, 1989, **1**: 190-206.

[100] Haykin S. *Neural Networks: A Comprehensive Foundation* [M]. Upper Saddle River: Prentice Hall PTR, 1994.

[101] Kennedy J, Eberhar R t. Particle swarm optimization [C]//*Proceedings of the* 1995 *IEEE International Conference on Neural Networks*, Perth: IEEE Service Center, 1995, 1942-1948.

[102] Shan S, Wang G G. Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions [J]. *Structural and Multidisciplinary Optimization*, 2010, **41**: 219-241.

[103] Regis R G, Shoemaker C A. A stochastic radial basis function method for the global optimization of expensive functions [J]. *INFORMS Journal on Computing*, 2007, **19**(4): 497-509.

[104] Jones D R, Schonlau M, Welch W J. Efficient global optimization of expensive black-box functions [J]. *Journal of Global Optimization*, 1998, **13**: 455-492.

[105] Bull A D. Convergence rates of efficient global optimization algorithms [J]. *Journal of Machine Learning Research*, 2011, **12**: 2879-2904.

[106] Wang Y L. SOARS: Statistical and Optimization Analysis and Response Surfaces for Computationally Expensive Models [R]. Changchun: the 9th International Conference on Numerical Optimization and Numerical Linear Algebra, 2013.

[107] Fletcher R, Leyffer S, Toint P L. On the global convergence of a filter-SQP algorithm [J]. *SIAM Journal on Optimization*, 2002, **13**: 44-59.

[108] Ulbrich M, Ulbrich S, Vicente L N. A globally convergent primal-dual interior-point filter method for nonlinear programming [J]. *Mathematical Programming*, 2004, **100**: 379-410.

[109] Wächter A, Biegler L T. Line search filter methods for nonlinear programming: Motivation and global convergence [J]. *SIAM Journal on Optimization*, 2005, **16**: 1-31.

[110] Ribeiro A A, Karas E W, Gonzaga C C. Global convergence of filter methods for nonlinear programming [J]. *SIAM Journal on Optimization*, 2008, **19**: 1231-1249.

[111] Liu X W, Yuan Y. A sequential quadratic programming method without a penalty function or a filter for nonlinear equality constrained optimization [J]. *SIAM Journal on Optimization*, 2011, **21**: 545-571.

[112] Dai Y H, Schittkowski K. A sequential quadratic programming algorithm with non-monotone line search [J]. *Pacific Journal on Optimization*, 2008, **4**: 335-351.

[113] Solodov M. Global convergence of an SQP method without boundedness assumptions on any of the iterative sequences [J]. *Mathematical Programming*, 2009, **118**: 1-12.

[114] Ulbrich M, Ulbrich S. Non-monotone trust region methods for nonlinear equality constrained optimization without a penalty function [J]. *Mathematical Programming*, 2003, **95**: 103-135.

[115] El-Bakry A S, Tapia R A, Tsuchiya T, et al. On the formulation and theory of the Newton interior-point method for nonlinear programming [J]. *Journal of Optimization Theory and Applications*, 1996, **89**: 507-541.

[116] Liu X W, Yuan Y. A null-space primal-dual interior-point algorithm for nonlinear optimization with nice convergence properties [J]. *Mathematical Programming*, 2010, **125**: 163-193.

[117] Andreani R, Birgin E. G, Martinez J M, et al. Augmented Lagrangian methods under the constant positive linear dependence constraint qualification [J]. *Mathematical Programming*, 2008, **111**: 5-32.

[118] Andreani R, Martínez J M, Svaiter B F. A new sequential optimality condition for constrained optimization and algorithmic consequences [J]. *SIAM Journal on Optimization*, 2010, **20**: 3533-3554.

[119] Liu X W, Sun J. A robust primal-dual interior point algorithm for nonlinear programs [J]. *SIAM Journal on Optimization*, 2004, **14**: 1163-1186.

[120] Byrd R H, Curtis F E, Nocedal J. Infeasibility detection and SQP methods for nonlinear optimization [J]. *SIAM Journal on Optimization*, 2010, **20**: 2281-2299.

[121] Glowinski R, Marrocco A. Approximation par éléments finis d'ordre un et résolution par pénalisation-dualité dúne classe de problèmes non linéaires [J]. *R A I R O* , 1975, **R2**: 41-76.

[122] Chan T F, Glowinski R. Finite element approximation and iterative solution of a class of mildly non-linear elliptic equations [R]. California: Stanford University, 1978.

[123] Gabay D, Mercier B. A dual algorithm for the solution of nonlinear variational problems via finite element approximations [J]. *Comput Math Appli*, 1976, **2**: 17-40.

[124] Bertsekas D P. *Nonlinear Programming* [M]. [s.l.]: Athena Scientific, 2010.

[125] Luo Z Q, Tseng P. On the Linear convergence of descent methods for convex essentially smooth minimization [J]. *SIAM Journal on Control and Optimization*, 1992, **30**: 408-425.

[126] Luo Z Q, Tseng P. On the convergence rate of dual ascent methods for strictly convex minimization [J]. *Mathematics of Operations Research*, 1993, **18**: 846-867.

[127] He B S, Yang H. Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities [J]. *Operations Research Letters*, 1998, **23**: 151-161.

[128] He B S, Liao L, Han D, et al. A new inexact alternating directions method for monotone variational inequalities [J]. *Mathematical Programming*, 2002, **92**: 103-118.

[129] Douglas J, Rachford H H. On the numerical solution of the heat conduction problem in 2 and 3 space variables [J]. *Transactions of the American Mathematical Society*, 1956, **82**: 421-439.

[130] Eckstein J, Bertsekas D P. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators [J].*Mathematical Programming*, 1992, **55**: 293-318.

[131] Goldfarb D, Ma S. Fast multiple splitting algorithms for convex optimization [J]. *SIAM Journal on Optimization*, 2012, **22**: 533-556.

[132] Goldfarb D, Ma S, Scheinberg K. Fast alternating linearization methods for minimizing the sum of two convex functions [J]. *Mathematical Programming*, 2013, **141**: 349-382.

[133] He B S, Yuan X M. On the $O(1/n)$ convergence rate of Douglas-Rachford alternating direction method [J]. *SIAM Journal on Numerical Analysis*, 2012, **50**: 700-709.

[134] Boley D. Linear convergence of ADMM on a model problem [R]. Minnesota: University of Minnesota, 2012.

[135] Han D, Yuan X Y. Local linear convergence of the alternating direction method of multipliers for quadratic programs [J]. *SIAM Journal on Numerical Analysis*, 2013, **51**(6): 3446-3457.

[136] Goldstein T, O’Donoghue B, Setzer S. Fast alternating direction optimization methods [J]. CAM report 12-35, UCLA, 2012.

[137] Ng M K, Weiss P, Yuan X M. Solving constrained total-variation problems via alternating direction methods [J]. *SIAM Journal on Scientific Computing*, 2010, **32**: 2710-2736.

[138] Yang J F, Zhang Y. Alternating direction algorithms for L_1-problems in compressive sensing [J]. *SIAM Journal on Scientific Computing*, 2011, **33**: 250-278.

[139] Goldstein T, Osher S. The split Bregman method for 1-regularized problems [J]. *SIAM Journal on Imaging Sciences*, 2009, **2**: 323-343.

[140] Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers [J]. *Foundations and Trends in Machine*, 2010, **3**: 1-122.

[141] Glowinski R. On alternating direction methods of multipliers: a historical perspective [C]//*Proceedings of a Conference Dedicated to J Periaux*, 2014.

[142] Han D, Yuan X M. A note on the alternating direction method [J]. *Journal of Optimization Theory and Applications*, 2012, **155**: 227-238.

[143] Chen C, He B, Ye Y, et al. The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent [EB/OL]. [2013-10-29]. http://www.optimization-online.org/DB_FILE/2013/09/4059.pdf.

[144] He B S, Tao M, Yuan X M. Alternating direction method with Gaussian back substitution for Separable Convex Programming [J]. *SIAM Journal on Optimization*, 2012, **22**: 313-340.

[145] Hong M, Luo Z Q. On the linear convergence of the alternating direction method of multipliers [EB/OL]. [2013-11-10]. http://arxiv.org/pdf/1208.3922v3.pdf.

[146] He B S, Xu H K, Yuan X M. On the proximal Jacobian decomposition of ALM for multiple-block separable convex minimization problems and its relationship to ADMM [EB/OL]. [2013-11-21]. http://www.optimization-online.org/DB_FILE/2013/11/4142.pdf.

[147] Wang X F, Hong M Y, Ma S Q, et al. Solving multi-block separable convex minimization problems using two-block alternating direction method of multipliers [EB/OL]. [2013-11-23]. http://arxiv.org/pdf/1308.5294v1.pdf.

[148] 石光明, 刘丹华, 高大化, 等. 压缩感知理论及其研究进展 [J]. *电子学报*, 2009, **37**: 1070-1081.

[149] 文再文, 印卧涛, 刘歆, 等. 压缩感知和稀疏优化简介 [J]. *运筹学学报*, 2012, **16**: 49-64.

[150] 许志强. 压缩感知 [J]. *中国科学: 数学*, 2012, **42**: 865-877.

[151] Yuan Y. Subspace techniques for nonlinear optimization [M]//*Some Topics in Industrial and Applied Mathematics*, Beijing: Higher Education Press, 2007, 206-218.