凸优化 线性规划，单纯形法，
外切椭球法

凸优化方法的进展

0224

V.M. Tikhomirov

Tikh.. VM 冯慈璜

凸优化的故事仅开始于大约五十年前是很奇怪的．更令人惊奇的是变分学的大分析学家和专门家都没有考虑不等式的约束．在 1940 年前很少有文章致力于不等式．这时期 Fourier(1823) 与 Vallée-Poussin (1911) 的两篇文章最有重要意义 (见 [1])．从此成千上万篇论文致力于这门学科，其中大量的是讨论凸问题．

L.V. Kantorovich 的文章最早涉及不等式约束的非平凡问题．事情是这样的： 1939 年春 Kantorovich 是列宁格勒大学 26 岁的教授，接到一个三夹板厂工程师的报告，他们想要更有效地运用机器，但缺乏问题的数学背景．他们所提出的这个尚未获得诺贝尔奖的问题是很简单的，实际上是个大学生的问题．但年轻的学者处理得非常严谨．在 35 年之后他写道： "总之，这不是一个偶然的问题，我注意到大量具有不同内容而有很多共同数学特征的问题" 它们具有模式

$$\begin{array}{ccc} (c,x) \to \sup & & \displaystyle\sum_{i=1}^{n} c_i x_i \to \sup \\ & \Leftrightarrow & \\ Ax \le b & & \displaystyle\sum_{i=1}^{n} a_{ji} x_i \le b_j, 1 \le j \le m \end{array} \tag{1}$$

归结于众所周知的线性规划问题．

同年 (1939) 年 Kantorovich 写了解决问题 (1) 的专著 [2]．他试图用他的研究帮助苏维埃政权，因为他认为这些对于发展苏维埃经济有用．但是，据某些意识形态的教条，象数学那样抽象的学科不可能用于象经济那种与生活密切相关的学科．最终 Kantorovich 被粗暴地告之取消这方面的研究．

线性规划以及更广泛意义下凸优化的各种进展在美国开始于 1947 年． G.B.Dantzig 在 [3] 的一篇文章叙述了数学规划的起源与进展．依据 Dantzig, 线性规划的迅速发展主要是由于象 von Neumann, Kantorovich,Leont'ev 与 Koopmans 等杰出学者的努力．在这些名单中无疑应添上 Dantzig 的名字．他的巨大贡献之一是解线性规划问题的著名的单纯形法．

单纯形法 Dantzig 的方法实质上很简单．细察问题 (1) 可知，问题中的可行性向量即满足 $Ax \le b$ 的 x, 组成 $\mathbb{R}^n$ 中的多面体 (有界或无界，在前一情形称为凸多胞形)．多面

原题: The Evolution of Methods of Convex Optimization. 译自: *Amer. Math. Monthly* Vol. 103, No. 1, 1996, pp. 65–71.

体是有限个半空间的交. 在这类集(满足附加假设, 不包含直线)上的线性函数在其一个顶点之处达到极大. 这意味着只要找(1)中线性函数在顶点集上的极大, 并选取其最大者. 然而在应用问题中顶点个数可能极其巨大, 因而需要系统地探究. Dantzig 提供了探索的方法. 在非退化情形, 这就是 Dantzig 的单纯形法.

假设已知某一顶点(找顶点有有效的方法). 假设它是非退化的, 这表明在该顶点处恰有 n 个 "线性独立" 的不等式变为等式.

不失一般性, 假定前 n 个不变式变为等式, 即

$$\langle a^j,\bar{x}\rangle = b_j,\ 1\le j\le n,\ \langle a^j,\bar{x}\rangle < b_j,\ j\ge n+1$$
$$a^j := (a_{j1},\cdots,a_{jn}),\ j=1,\cdots,m \tag{i}$$

顶点的非退化性表明向量 $\{a^1,\cdots.a^n\}$ 构成 $\boldsymbol{R}^n$ 的基, 亦即矩阵 $A_n := (a_{ij})_{1\le i,j\le n}$ 是非奇异的. 令 $\bar{b}=(b_1,\cdots,b_n)\in\boldsymbol{R}^n$ 则

$$A_n\bar{x}=\bar{b}=(b_1,\cdots,b_n)\Leftrightarrow \bar{x}=A_n^{-1}\bar{b} \tag{ii}$$

解方程

$$A_n^T\bar{\lambda}=c \tag{iii}$$

(A_n^T 表示 A_n 的转置). 有两种情况: (I) $\bar{\lambda}=(\lambda_1,\cdots,\lambda_n)\ge 0$ 以及 (II) $\bar{\lambda}$ 的某些分量为负.

我们先证明在第一种情形 $\bar{x}$ 是(1)的解. 事实上, 令 x 为可行向量 ($Ax\le b$ (iv)). 记 $\lambda=(\lambda_1,\cdots,\lambda_n,0,\cdots,0)$ 则

$$\langle c,x\rangle \overset{(iii)}{=} \langle A_n^T\bar{\lambda},x\rangle \overset{Id}{=} \langle\bar{\lambda},A_nx\rangle \overset{Id}{=} \langle\lambda,A_x\rangle \overset{(iv)及\lambda\ge 0}{\le}$$
$$\le \langle\lambda,b\rangle \overset{Id}{=} \langle\bar{\lambda},\bar{b}\rangle, \overset{(ii)}{=} \langle\bar{\lambda},A_n\bar{x}\rangle \overset{Id}{=} \langle A_n^T\bar{\lambda},\bar{x}\rangle \overset{(iii)}{=} \langle c,\bar{x}\rangle,$$

这就是所要证的. 下面考虑另一种情形

假设 $\lambda_1<0$, 我们求齐次方程组

$$\langle a^2,y\rangle=\cdots=\langle a^n,y\rangle=0. \tag{v}$$

的非平凡解.

由于 A_n 非奇异可知 $\langle a^1,y\rangle\ne 0$. 适当改变符号有

$$\langle a^1,y\rangle=-\varepsilon<0 \tag{vi}$$

于是对小的值 $t>0$, 有

$$\langle a^j,\bar{x}+ty\rangle<b_j, j=1, j\ge n+1, \langle a^j,\bar{x}+ty\rangle=b_j, 2\le j\le n,$$

即 $\bar{x}+ty$ 是可行向量. 又

$$\langle c,\bar{x}+ty\rangle \overset{(v),(vi)}{=} \langle c,\bar{x}\rangle + t < c, A_n^{-1}(\varepsilon,0,\cdots,0) > \overset{Id}{=} \langle c,\bar{x}\rangle$$
$$+ t\langle (A_n^{-1})_c^T, (-\varepsilon,0,\cdots,0)\rangle \overset{Id}{=} \langle c,\bar{x}\rangle - t\varepsilon\lambda_1.$$

这意味着对一切 $t>0$ 有 $\langle c,\bar{x}+ty\rangle > \langle c,\bar{x}\rangle$.

若对一切 $t>0, \bar{x}+ty$ 是可行的，则问题的上确界为 $+\infty$. 否则，对 t 的某个值 t_0 及 $j\geq n+1$ 有 $\langle a^j,\bar{x}+t_0y\rangle = b_j$. 然后以 $\bar{x}+t_0y$ 代替 $\bar{x}$ 的位置进行另一次迭代. 依次进行下去，这就是非奇异非退化情形的单纯形方法.

单纯形法在优化数值方法史上占有重要地位. 在相当长时间里，不知道 (1) 中的问题是不是多项式的，也就是说，在各种情形下，这类问题的算法是否只需要多项式数目的运算 (依据输入量的大小).1970 年 Klee 与 Minty 造出一个例子来表明在某些情况下，单纯形法需要指数数目的步数. 但人们有理由认为这类问题不会在实际中碰到！很多数学家 (包括 Dantzig) 曾经说过，单纯形法对无数个应用问题顺利地使用了五十年，这真正是个奇迹！ Dantzig 自己也说："单纯形法的巨大威力常常使我惊奇".

在发现单纯形法之后若干年， Dantzig 决定写一本线性规划详尽的专著 (除 [1] 以外). 它是有关不等式及其在极值问题中应用的全部文章的综述. 特别是 Kantorovich 的专著 [2] 吸引了他的注意，对这部专著 Dantzig 加以高度赞扬，并看准把它翻译出来. 这样在科学界就知道了 Kantorovich 和 Koopman 对线性规划论的贡献. 1975 年这两位科学家由于发展线性规划及其在经济中应用的贡献而获诺贝尔经济学奖. 接下去我们讨论线性及凸规划的其它方法.

中心切割法 我谈谈亲身经历的故事. 1962 年我在俄罗斯中心的大城市 Voronezh 工作. 那时由 M.A.Krasnosel'ski 领导的 Voronezh 数学学院欣欣向荣. 学院追求重要的科学目标并试图进入各种应用领域. 尤其是在 60 年代早期， Krasnosel'ski 与由 David B.Yudin 领导的应用数学所签有合同， Yudin 的名字由于他与 E.G.Gol'shtein 合写的专著 [4] 而为优化专家们所熟知. Yudin 提出如下问题：找一有效算法使在列紧多面体上 (以正权) 的指数和为最小.

这个问题吸引了本故事的英雄和 Krasnosel'ski 组的成员 Anatoli Yu Levin. 和对待其它问题一样， Levin 考虑了很长时间. 他与伙伴们 (包括我在内) 共同讨论. 有一天，他想到了一个能用于更一般问题的值得注意的想法：在有限维凸体 A 上，找凸 (以及拟凸) 函数 f 的最小值，即

$$f(x)\to \inf; x\in A.$$

实质上这是凸优化的一般问题. 下面叙述一下光滑凸函数的 Levin 算法.

记 A 为 A_1. 确定 A_1 的重心 $x_1 = grA_1$. 然后计算 $f'(x_1)$. 如果它是零向量，那么问题就解决了. 否则，去除 A_1 在半空间 $\prod_1' = \{x|\langle f'(x_1), x-x_1\rangle > 0\}$ 的部分 (这一步证明如下：对凸的 f 不难证明 $f(x)-f(x_1)\geq \langle f'(x_1), x-x_1\rangle$. 其次由 $x\in A_1\cap \prod_1'$, 有 $f(x)>f(x')>\min$). 记 A_1 的留下部分为 A_2, 重复上述步骤，依次下去.

若取 ξ_m 为 $\{x_1,\cdots,x_m\}$ 的一点，且 $f(\xi_m)$ 不小于任何 $f(x_i), 1\le i\le m$，则可以证明 $f(\xi_m)$ 趋向于 f 在 A 上的最小值，并且 f 之值的误差依几何级数减少．此外 A_m 的体积指数地递减．这个事实是凸几何学中 Grünbaum 的下述结果的推论：设 A 为 $\boldsymbol{R}^n$ 中的凸体，$\xi = grA$ 为它的重心，则过 ξ 的任一超平面分 A 为两部分，每部分的体积不少于 A 体积的 $(1-1/e)$ 倍．

我们指出在一维情形此法就是对分区间法．

Levin 延误了将结果成文发表（其中的一些原因下面将提到）．与此同时，美国的 D.J.Newman 独立地想出了中心切割法．Levin 与 Newman 的文章 [5,6] 同时发表于 1965 年．

不久，我的学生 A.I.Kuzovkin 与我补充了 Levin 的算法，并证明用计算 f 的值来代替它的梯度，可得到指数的收敛速度．

此后进入了冬眠期．下一个重要的事件距发现中心切割法后约 15 年．

Nemirovski-Yudin-Shor 的外切椭球法． 我向读者介绍一位在推进凸优化中起过重要作用的人，这就是莫斯科大学研究生，G.E.Shilov 的关门弟子之一 Arkadi Nemirovski.

1974 年 Nemirovski 完成研究生学习之后开始跟 Yudin 工作． Yudin 把解凸优化问题的复杂性放在了年轻的合作者面前．1974 年 11 月（正在森林里散步时），Nemirovski 想到了解凸优化问题的另外的方法，这就是外切椭球法．这个方法叙述在一篇文章里，其关键的思想也被著名的凸优化专家，基辅数学家 Naum Z.Shor 独立地（稍微迟一点）得到（见 [8]）．因此外切椭球法也称为 Nemirovski-Yudin-Shor 法．

这个方法联合了两个思想：其一是前述的切割法，另一个则是几何事实：半椭球能置入于体积比原椭球体积更小的椭球里．

现在较详细叙述 N-Y-S 算法．以 $E_\circ$ 表示外切于 A 的椭球．若该椭球的中心 $c_\circ$ 在 A 之外，则过 $c_\circ$ 有不含 A 的点的半空间，并去掉与 A 不相交的半椭球．若 $c_\circ \in A$ 则计算 $f'(c_\circ)$ 进行 Levin-Newman 的切割，将得到的半椭球记作 $E'_\circ$，再作体积比 $E_\circ$ 更小的一个 $E'_\circ$ 的外切椭球 E_1，重复地做下去．

所述方法的解的收敛速度是几何的． Yudin 与 Nemirovski 在文中指出， Levin-Newman 中心切割法在凸函数最小化的收敛算法类中不能有实质性的改进．外切椭球法在收敛速度方面稍逊于中心切割法，但它具有无需找多面体重心的优点．

在各种讨论班上， Nemirovski 与 Yudin 讲了这一切．有一次，在中央经济数学所 E.G.Gol'shtein 的讨论班上他们讲这个方法． Gol'shtein 很喜欢 Nemirovski-Yudin 讲学，但他发现以前有过类似的思想． "在什么地方？" 讲演人问． "在 A.Yu Levin 的文章里"."哪一个 Levin?"Yudin 紧张地问． Golshtein 回答说： "15 年前解决你问题的那个 Levin …….” 因此 Nemirovski 与 Yudin 在 [8] 中把 A.Yu. Levin 的文章 [5] 列为参考文献．

L.G.Khachian 的文章． 过了若干年，只有一件事值得一提． Nemirovski 的同学 L.Levin 决定移居美国， Nemirovski 去送行，他将椭球法一文 [8] 的复印件送给 Levin. 这

时 Levin 的心情不在数学上，将它放在箱底里.

一天早上，我碰巧经过在莫斯科的科学院计算中心，看到在那里工作的老同学正在跑步．我大声问他："有什么事吗？"他一边跑一边断断续续地大喊着："Khachian ……纽约时报……新闻记者招待会……我迟了……."

事情是这样的：计算中心的一位年轻同事 Leonid G.Khachian 在苏联科学院进展上发表过题为"线性规划中的多项式算法"的短文 [9], 很长时间没有人对它有什么反应.后来在美国举行的一个凸优化专家会议上，与会者之一叙述了该领域的新进展，提到了 Khachian 的文章．他评论到："Khachian 证明了线性规划问题可在多项式时间内求解……". 他正要继续讲下去，却被匈牙利数学家 Peter Gács 打断了："请你重复一遍刚才说的话好吗？"报告人重述了一遍之后，Gács 宣布："这就解决了一个著名的问题：关于线性规划问题的多项式性质!"Khachian 的结果可解释作线性规划问题多项式性质的证明．这个算法将椭球法应用于线性规划问题．新闻界对 Khachian 的文章感兴趣，这就导致了在苏联科学院计算中心的纽约时报记者招待会……

Khachian 的文章包含 Shor 的文章作为文献，这使 L. Levin 想起在他赴美时 Nemirovski 给他的复印件．从那里他学会了外切椭球法，有了在解凸规划问题时步数的多项式估计，尤其是在实用上的线性规划问题．Khachian 走了下一步，这是关键的一步，即他发现必需的计算复杂性是有理的 (从而是多项式的).

总之，Khachian 及 Nemirovski 的文章为众所周知，1982 年，这些文章的作者获得了数学规划国际学会与美国数学会授予的有威望的 Fulkerson 奖.

Peter Gács 提醒 L.Levin 注意与他同名的 A. Yu. Levin 的文章 [5]. 现在我们回到前面提到的，文 [5] 发表的延误原因之一是因为 A.Yu. Levin 想要克服找多面体重心的困难 (后来知道这个问题需要指数的复杂性). 在 [5] 里 Levin 更进一步，有时可对 A_n 作外切单形，并且不改变其体积依指数的减少．他未给出专门的算法，而是这类算法的存在定理．从这一想法出发，Levin 与 Boris Yamnitsky 在 [10] 证明，A.Yu. Levin 的想法在原则上是正确的．文 [10] 叙述了"外切单形"法，它与外切椭球法相符，只需运行多项式时间；在某些方面还优于椭球法.

结束语 前面描述的事情的结果是 80 年代凸优化法有突破性进展，涌现了许多切割法 (例如 Tarasov-Khachian-Ehrlich 的内接椭球法等，见 [19]). 许多方法作了改进和完善．1984 年，Karmarkar 根据另一种思路，提出了线性规划问题的多项式方法，它有许多优点，并引导出现了大量讨论凸规划算法的文章，使上述方法得到修正，发展和完善．有关的详情读者可在 Nemirovski 与 Nesterov 的专著 [11] 中找到.

英文方面的进一步信息见文献 [12-19].

(致谢部分从略)

参 考 文 献

[1] G.B.Dantzig, Linear Programming and Extensions, Princeton,Univ.Press, 1963.

[2] Kantorovich, L.V.Mathematical methods in the organization and planning of production, Leningrad University, 1939.[Russian] For an English version see *Management Science*, Vol. 6, 1960, 366–422.

[3] History of Mathematical Programming (A Collection of Personal Ren ιiscences). Ed. by J.K.Lenstra, F.H.G.Rinnooy Kan, A. Schrijver, North-Holland, 1991.

[4] Yudin, D.B.and Gol'shtein, E.G.Problems and methods of linear programming. Moscow. Sovradio, 1961. [Russian].

[5] Levin, A.Yu.On an algorithm for the minimization of convex functions. *DAN USSR* vol.160,issue 6,1965, 1244–1247. [Russian].

[6] Newman, D.J.Location of maximum on unimodal surfaces. *Journ. of the Assoc.for Computing Machinery*, **12**,1965, 395–398.

[7] Kuzovkin, A.I.and Tikhomirov, V.M.On the number of computations needed for finding the minimum of a convex function. *Economics and mathematical methods*, 3:1, 1967, 95–103. [Russian].

[8] Yudin, D.B. and Nemirovski, A.S.Informational complexity and effective methods of solution of convex extremal problems. *Economics and mathematical methods*, vol.12, issue 1, 1976, 357–369.[Russian].

[9] Khachian, L.G.A polynomial algorithm in linear programming. *DAN USSR* vol.244, issue 5, 1979, 1093–1096 [Russian]. For an English version see *Soviet Math. Doklady* **20**, 1979, 191–194.

[10] Yamnitsky, B., Levin, L.An Old Linear Programming Algorithm Runs in Polynomial Time. In 23rd Annual Symposium on Foundations of Computer Science, IEEE, New York, 1982, 327–328.

[11] Nemirovsky, A., Nesterov, Yu. Interior-Point Polynomial Algorithms in Convex programming. SIAM Studies in Applied Math.,1994.

[12] Nemirovsky, A. S., Yudin, D.B.Problem Complexity and Method Efficiency in Optimization. John Wiley and Sons, Chichester, 1983.

[13] Grötschel, M., Lovász, L., Schrijver, A. Geometric Algorithms and Combinatorial Optimization. Springer-Verlag, 1988.

[14] Schrijver, A. Theory of Linear and Integer Programming. John Wiley and Sons, 1986.

[15] Shor, N.Z.Minimization Methods for Nondifferentiable Functions. Springer-Verlag, Berlin, 1985.

[16] Bland, R. G., Goldfarb, D., Todd, M. J. The ellipsoid method: a survey. *Operations Research*, **29** (1981), 1039–1091.

[17] Goldfarb, D., Todd, M. J. Linear Programming in "Handbooks in Operations Research and Management Science," vol. 1, Optimization (G. L. Nemhauser, A. H. G. Rinnooy Kan and M. J. Todd, eds.), North Holland, Amsterdam, 1989, 73–170.

[18] Shor, N.Z.Cut-off method with space extension in convex programming problems, *Cybernetics*, **13** (1977), 94–96.

[19] Tarason, S.P., Khachiyan, L. G., Erlikh, I.I. The method of inscribed ellipsoids. *Soviet Math. Doklady*, **37** (1988), 226–230.

(冯慈璜 译　姚景齐 校)