凸分析讲义

李庆娜　李萌萌　于盼盼　编

科学出版社

凸分析讲义

李庆娜　李萌萌　于盼盼　编

科学出版社
北京

内 容 简 介

本书主要讲述了最优化理论的基础——凸分析的主要内容，是结合作者多年来在最优化课程中的经验及凸分析讨论班涉及的内容总结整理而成的. 本书融入了大量研究最优化理论用的应用案例及图片，使得对知识点的理解更加简单形象，便于本科生及研究生作为教材及优化的参考书. 本书基本内容包括仿射集、凸集及凸集上的运算、凸集的拓扑性质、凸函数及其运算等.

本书可作为应用数学、运筹学及相关学科的高年级本科生及研究生的教材和参考书.

图书在版编目(CIP)数据

凸分析讲义/李庆娜，李萌萌，于盼盼编. —北京：科学出版社，2019.2

ISBN 978-7-03-060504-7

Ⅰ. ①凸…　Ⅱ. ①李…　②李…　③于…　Ⅲ. ①凸分析-高等学校-教材　Ⅳ. ①O174.13

中国版本图书馆 CIP 数据核字（2019）第 020981 号

责任编辑：胡庆家／责任校对：邹慧卿
责任印制：吴兆东／封面设计：陈　敬

科学出版社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

北京中石油彩色印刷有限责任公司 印刷
科学出版社发行　各地新华书店经销
*
2019 年 2 月第　一　版　开本：720 × 1000　B5
2019 年 2 月第一次印刷　印张：9 3/4
字数：150 000

定价：58.00 元

（如有印装质量问题，我社负责调换）

前　言

运筹学产生于第二次世界大战期间. 作为运筹学的一个重要而活跃的部分 —— 最优化理论与方法在近半个世纪以来得到了蓬勃发展. 凸分析作为最优化理论与方法的重要理论基础, 也越来越为人们所重视.

本书主要对凸分析的基本概念和内容进行了介绍. 主要内容包括仿射集、凸集、凸函数、凸集及凸函数的运算、拓扑性质. 这些内容看似简单, 实则构成了最优化理论及算法设计的基本的分析工具. 为了增强可读性, 本书将最优化方法部分前沿研究内容与凸分析的概念、理论相结合，辅以大量例子、图片及练习, 以期读者对本书内容有更形象深刻地理解和把握.

在本书的编写过程中, 得到了国内同行专家的支持和鼓励, 在此一并表示衷心的感谢! 感谢优化课题组每一位成员积极参加讨论班, 没有他们的激烈讨论、认真校对, 就没有本书的出版. 最后，感谢国家自然科学基金 (11671036) 的资助及北京理工大学 "十三五" 教材规划资助.

本书可作为数学、运筹学及相关学科的高年级本科生及研究生的教材和参考书.

因作者水平所限, 本书难免有不足之处. 恳请读者不吝赐教. 来信请发至: qnl@bit.edu.cn.

李庆娜　李萌萌　于盼盼

2019 年 1 月

目　　录

第 1 章　仿　射　集

1.1　预备知识: 内积

定义 1.1　设 x 和 y 为 n 维列向量, x 与 y 的内积定义为

$$\langle x, y\rangle = x_1y_1 + x_2y_2 + \cdots + x_ny_n = x^{\mathrm{T}}y,$$

其中, x^{T} 表示 x 的转置, 即

$$x^{\mathrm{T}} = (x_1,\ x_2,\ \cdots,\ x_n), \quad y^{\mathrm{T}} = (y_1,\ y_2,\ \cdots,\ y_n).$$

设 A 为一个 $m \times n$ 的实矩阵, 即 $A : \mathbb{R}^n \to \mathbb{R}^m$. 把 A 的转置矩阵以及相应的伴随线性变换记为 $A^* : \mathbb{R}^m \to \mathbb{R}^n$, 则有 $\langle Ax, y\rangle = \langle x, A^*y\rangle$. 可以验证:

$$\langle Ax, y\rangle = (Ax)^{\mathrm{T}}y = x^{\mathrm{T}}A^*y = \langle x, A^*y\rangle.$$

注 1.1　记 $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ 为线性变换, $\mathcal{A}^* : \mathbb{R}^p \to \mathbb{R}^{m\times n}$ 为其伴随线性变换, 则同样有

$$\langle \mathcal{A}(X), y\rangle = \langle X, \mathcal{A}^*y\rangle, \quad X \in \mathbb{R}^{m\times n},\ y \in \mathbb{R}^p. \tag{1.1}$$

注 1.2　矩阵间的内积有如下定义: $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times n}$,

$$\langle A, B\rangle = \mathrm{tr}(A^{\mathrm{T}}B) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij}.$$

例子 1.1　设 $X \in \mathbb{R}^{m\times n}$, 定义线性算子 $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ 如下:

$$\mathcal{A}(X)=\begin{bmatrix}\langle A^{(1)},X\rangle\\ \vdots\\ \langle A^{(p)},X\rangle\end{bmatrix}, \tag{1.2}$$

其中, $A^{(i)}\in\mathbb{R}^{m\times n}$, $i=1,\cdots,p$. 则由式 (1.1) 可知

$$\langle\mathcal{A}(X),y\rangle=\left\langle\begin{bmatrix}\langle A^{(1)},X\rangle\\ \vdots\\ \langle A^{(p)},X\rangle\end{bmatrix},y\right\rangle=\sum_{i=1}^{p}y_i\langle A^{(i)},X\rangle=\langle X,\mathcal{A}^*y\rangle,$$

因此

$$\mathcal{A}^*y=\sum_{i=1}^{p}y_iA^{(i)}.$$

例子 1.2[7]　如上例, 对于 $X\in\mathbb{R}^{n\times n}$, 若定义 $\mathcal{A}(X):=\mathrm{diag}(X)=(X_{11},\cdots,X_{nn})^{\mathrm{T}}$, 这里 X_{ij} 是 X 中第 i 行 j 列的元素. 则 $\mathcal{A}(X)$ 可写为

$$\mathcal{A}(X)=\begin{bmatrix}\langle A^{(1)},X\rangle\\ \vdots\\ \langle A^{(n)},X\rangle\end{bmatrix}$$

的形式. 其中

$$A^{(1)}=\begin{pmatrix}1&&\\&\ddots&\\&&0\end{pmatrix},\ \cdots,\ A^{(n)}=\begin{pmatrix}0&&\\&\ddots&\\&&1\end{pmatrix}\in\mathbb{R}^{n\times n}.$$

可以得出, 对向量 $y=(y_1,\cdots,y_n)^{\mathrm{T}}$,

$$\mathcal{A}^*y=\begin{pmatrix}y_1&&&\\&y_2&&\\&&\ddots&\\&&&y_n\end{pmatrix}:=\mathrm{Diag}(y).$$

例子 1.3 设 $D \in \mathbb{R}^{3\times 3}$. 已知 $\mathcal{A}(D) = \begin{bmatrix} D_{11} - D_{12} \\ D_{23} - D_{31} \\ D_{33} \end{bmatrix}$, 则对应的 $A^{(1)}$, $A^{(2)}$, $A^{(3)}$ 分别为

$$A^{(1)} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad A^{(2)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \end{pmatrix}, \quad A^{(3)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

对于 $y \in \mathbb{R}^3$, 可计算得

$$\mathcal{A}^* y = \sum_{i=1}^{3} y_i A^{(i)} = \begin{bmatrix} y_1 & -y_1 & 0 \\ 0 & 0 & y_2 \\ -y_2 & 0 & y_3 \end{bmatrix}.$$

例子 1.4 记 $\mathcal{S}^n$ 为 $n \times n$ 的对称矩阵空间. 设 $D \in \mathcal{S}^3$. 已知 $\mathcal{A}(D) = \begin{bmatrix} D_{11} - D_{12} \\ D_{13} - D_{23} \\ D_{33} \end{bmatrix}$, 则对应的 $A^{(1)}$, $A^{(2)}$, $A^{(3)} \in \mathcal{S}^3$ 分别为

$$A^{(1)} = \begin{bmatrix} 1 & -0.5 & 0 \\ -0.5 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A^{(2)} = \begin{bmatrix} 0 & 0 & 0.5 \\ 0 & 0 & -0.5 \\ 0.5 & -0.5 & 0 \end{bmatrix},$$

$$A^{(3)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

对于 $y \in \mathbb{R}^3$, 可计算得

$$\mathcal{A}^* y = \sum_{i=1}^{3} y_i A^{(i)} = \begin{bmatrix} y_1 & -0.5y_1 & 0.5y_2 \\ -0.5y_1 & 0 & -0.5y_2 \\ 0.5y_2 & -0.5y_2 & y_3 \end{bmatrix}.$$

例子 1.5[8]　设算子 $\mathcal{A}$ 如例子 1.2 定义, 则

$$\mathcal{A}\mathcal{A}^* y = \operatorname{diag}(\operatorname{Diag}(y)) = y.$$

更进一步, 若已知

$$\mathcal{A}\mathcal{A}^* y = b,$$

其中 $b \in \mathbb{R}^n$ 为给定的向量, 则可求得

$$y = b.$$

1.2　仿　射　集

定义 1.2　设 x 和 y 为 $\mathbb{R}^n$ 中两个不同的点, $\lambda \in \mathbb{R}$, 称 $(1-\lambda)x+\lambda y$ 为过 x 与 y 的直线. 当 $\lambda \in [0,1]$ 时, $(1-\lambda)x+\lambda y$ 称为 x 与 y 的凸组合.

定义 1.3 (仿射集)　设 M 是 $\mathbb{R}^n$ 的子集. 如果对任意的 $x,\ y \in M$, 任意的 $\lambda \in \mathbb{R}$, 有 $(1-\lambda)x+\lambda y \in M$, 则称 M 为仿射集. 仿射集也叫仿射流形、仿射簇、线性簇等. 当 $\lambda \in [0,1]$ 时, 称 M 为凸集.

显然, 仿射集一定是凸集.

例子 1.6　在 $\mathbb{R}^3$ 中, 任意的点、没有终点的直线、x 轴、y 轴、整个 $\mathbb{R}^3$ 空间都是仿射集.

x 的正半轴、起点为原点的射线、$y \geqslant 0$ 的半空间、第三个轴没有终点的圆柱 $A = \{(x,y,z) | x^2+y^2=1\}$ 都不是仿射集.

注 1.3 $\mathbb{R}^n$ 和 $\varnothing$ 是特殊的仿射集, 单点集也是仿射集, 但有限点集 (点的数目大于 1) 不是仿射集. 从另外的角度, $(1-\lambda)x+\lambda y = x+\lambda(y-x)$ 可以看作从 x 出发, 沿着 $y-x$ 方向移动 λ 步长. 如图 1.1 所示.

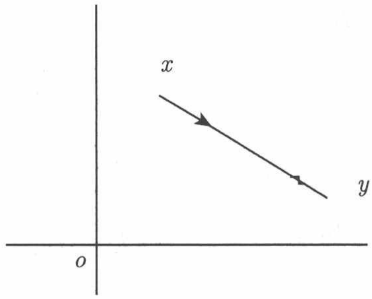

图 1.1 对仿射集的理解

例子 1.7 实圆盘 $C=\{(x,y) \mid x^2+y^2 \leqslant r^2\} \triangleq \{z \in \mathbb{R}^2 \mid \|z\|_2 \leqslant r\}$.

例子 1.8 球体 $B = \{(x,\ y,\ z) \mid x^2 + y^2 + z^2 \leqslant r^2\} \triangleq \{z \in \mathbb{R}^3 \mid \|z\|_2 \leqslant r\}$.

例子 1.9 n 维球体 $B^n = \{z \in \mathbb{R}^n \mid \|z\|_2 \leqslant r\}$.

来检验 B^n 是否是凸集、仿射集、子空间. 类似地可以判断 C, B.

(1) 对任意的 $x,\ y \in B^n,\ \lambda \in [0,1]$, 因为

$$\|(1-\lambda)x+\lambda y\|_2 \leqslant (1-\lambda)\|x\|_2 + \lambda\|y\|_2 \leqslant (1-\lambda)r + \lambda r = r,$$

所以

$$(1-\lambda)x+\lambda y \in B^n.$$

因此, B^n 是凸集.

(2) 因为封闭的球不可能有直线, 过 B^n 中任意两点的直线不可能包含在封闭的球内, 故 B^n 不是仿射集.

(3) 子空间一定是仿射集 (见下面定理 1.1), 故不是仿射集一定不是子空间. 因此 B^n 不是子空间.

注 1.4　"封闭"的有界集合一定不是仿射集.

定理 1.1　$\mathbb{R}^n$ 的子空间是包含原点的仿射集.

证明　设 M 为包含原点的仿射集. 要证明 M 是子空间, 只要对加法和数乘进行验证即可. 对任意的 $x,\ y \in M,\ \lambda \in \mathbb{R},\ \lambda x = (1-\lambda)0+\lambda x$, 故 M 对数乘封闭. $\dfrac{1}{2}(x+y) = \dfrac{1}{2}x+\dfrac{1}{2}y \in M$, 则 $x+y = 2\left(\dfrac{1}{2}(x+y)\right) \in M$, 故 M 对加法封闭. 从而, M 是 $\mathbb{R}^n$ 的子空间.

反过来, 要证明子空间是包含原点的仿射集. 首先, 子空间一定包含原点, 且对加法和数乘封闭. 则对 $\forall x, y \in K$(子空间), $\lambda \in \mathbb{R},\ (1-\lambda)x \in K,\ \lambda y \in K$(数乘封闭), 都有 $(1-\lambda)x+\lambda y \in K$(加法封闭). 所以任意子空间一定是仿射集. 即子空间是包含原点的仿射集. □

定义 1.4　设集合 $M \subset \mathbb{R}^n,\ a \in \mathbb{R}^n$, 称 $M+a = \{x+a \mid x \in M\}$ 为 M 沿向量 a 的平移.

显然, 如果 M 是仿射集, 则 $M+a$ 也是仿射集.

证明　对任意的 $x,\ y \in M+a$, 存在 $x_1, y_1 \in M$, 使得 $x = x_1+a,\ y = y_1+a$,

因 M 是仿射集, 故对任意的 $\lambda \in \mathbb{R}$, 有 $(1-\lambda)x_1+\lambda y_1 \in M$, 故

$$\begin{aligned}(1-\lambda)x+\lambda y &= (1-\lambda)(x_1+a)+\lambda(y_1+a)\\ &= (1-\lambda)x_1+\lambda y_1+a \in M+a,\end{aligned}$$

即 $M+a$ 是仿射集. □

定义 1.5　如果存在 $a \in \mathbb{R}^n$, 使得 $M = L+a$, 则称仿射集 M 平行于仿射集 L.

"M 平行于 L" 在 $\mathbb{R}^n$ 的仿射子集构成的集族上是一个等价关系, 即满足自反性、对称性、传递性.

定理 1.2 任一非空仿射集 M 平行于唯一的子空间 L, 且

$$L = M - M = \{x - y \mid x \in M, y \in M\}.$$

证明 (唯一性) 设 L_1, L_2 为平行于 M 的子空间. 由传递性, L_1 平行于 L_2, 则存在 $a \in \mathbb{R}^n$, 使得 $L_2 = L_1 + a$. 而 $0 \in L_2$, 则 $-a \in L_1$, 即 $a \in L_1$, 故 $L_1 + a \subset L_1$, 即 $L_2 \subset L_1$. 同样可得 $L_2 \supset L_1$. 因此, $L_1 = L_2$.

(存在性) 设 M 是任一非空仿射集, 对任意 $y \in M$, $M - y = M + (-y)$ 是 M 沿 $-y$ 的平移, 且包含 0. 由定理 1.1 及上面所证, 该仿射集必为平行于 M 的唯一子空间, 记为 L. 由 $y \in M$ 的任意性可知

$$L = M - M.$$ □

以 $\mathbb{R}^2$ 为例, 仿射集与子空间的关系如图 1.2 所示.

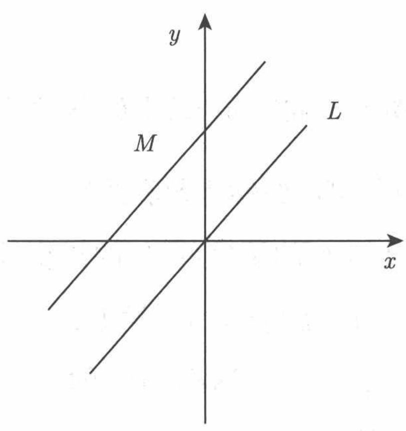

图 1.2 仿射集与子空间的关系

例子 1.10 仿射集 $M = \{(x, y) \mid x - y = -1\}$, $\alpha = (2, 3) \in \mathbb{R}^2$, $M - \alpha$ 是包含原点的仿射集, 即为一子空间 L.

定义 1.6 非空仿射集的维数定义为平行于它的子空间的维数.

规定空集的维数为 -1, 维数分别为 0, 1, 2 的仿射集分别是点、线、平面.

1.3 超 平 面

定义 1.7　$\mathbb{R}^n$ 的一个 $n-1$ 维仿射集叫作超平面.

$\mathbb{R}^2$ 上的超平面为直线, $\mathbb{R}^3$ 上的超平面是二维平面.

例子 1.11　在二维空间里, 线性方程 $x+y=1$ 的解集是一个仿射集, 也是一个超平面. 见图 1.3.

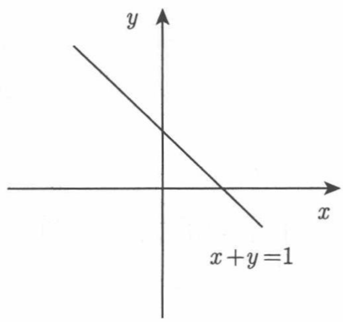

图 1.3　超平面与线性方程的关系

超平面及其他仿射集可由线性函数和线性方程来表示. 这可以从 $\mathbb{R}^n$ 空间上的正交理论导出. $\mathbb{R}^n$ 的子空间 L 的正交补 $L^{\perp}=\{x \mid x\perp y, y\in L\}$, 且 $\dim L+\dim L^{\perp}=n, (L^{\perp})^{\perp}=L$. 设 $b_1,\cdots,b_m$ 为 L 的一组基. 则 $x\perp L$ 等价于 $x\perp b_i$, $i=1,\cdots,m$. 这样, $\mathbb{R}^n$ 的一个 $n-1$ 维子空间 L 是一个一维子空间的正交补. 设这个一维子空间由非零向量 b 生成. 则与该一维子空间正交的 $n-1$ 维子空间可表示为 $\{x \mid \langle x,b\rangle=0\}$. 又超平面是 $n-1$ 维子空间的平移, 故存在 $a\in\mathbb{R}^n$, 使得

$$\{x \mid \langle x,b\rangle=0\}+a=\{y \mid \langle y-a,b\rangle=0\}=\{y \mid \langle y,b\rangle=\beta\}.$$

其中 $\beta=\langle a,b\rangle$. 由此可以得到如下超平面的刻画.

定理 1.3　$\mathbb{R}^n$ 中每个超平面 H 都可以表示为 $H=\{x \mid \langle x,b\rangle=\beta\}$,

其中, $b \in \mathbb{R}^n, \beta \in \mathbb{R}$. 在不计非零公因子意义下 b 与 β 是唯一的. b 叫作超平面 H 的法线.

例子 1.12 在 $\mathbb{R}^2$ 中, 令 $b = (0,1)^{\mathrm{T}}, \beta = 1$. 由 b 和 β 确定的超平面为 $H = \{x \mid \langle x, b\rangle = \beta\} = \{x = (x_1, x_2) \mid x_1 \cdot 0 + x_2 \cdot 1 = 1\} = \{x = (x_1, x_2) \mid x_1 \in \mathbb{R}, x_2 = 1\}$. 见图 1.4.

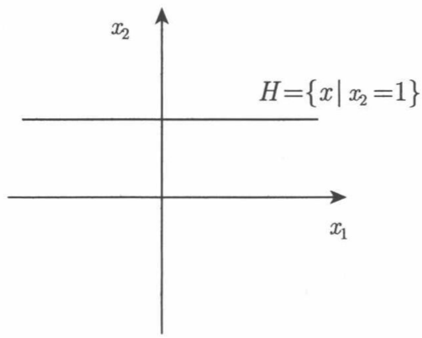

图 1.4 超平面示意图

例子 1.13 支持向量机 (Support Vector Machine, SVM[2]) 的原理, 就是找到合适的超平面 $H = \{x \in \mathbb{R}^n \mid x^{\mathrm{T}}w + b = 0\}$ 以实现样本分类.

例子 1.14 在 $\mathbb{R}^2$ 中, 令 $b = (0,1)^{\mathrm{T}}, \beta = 1$, 则超平面为 $H = \{x \in \mathbb{R}^2 \mid x_2 = 1\}$. 超平面与法向量如图 1.5 所示. 向量 $b, -b, \lambda b\ (\lambda \neq 0)$ 都是超平面的法向量.

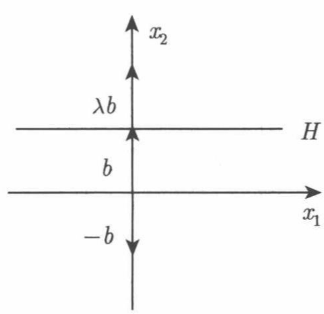

图 1.5 例子 1.14 示意图

例子 1.15 在 $\mathbb{R}^3$ 中, 令 $b=(0,1,2)^{\mathrm{T}}$, $\beta=2$, 则超平面为 $H=\{x\in\mathbb{R}^3 \mid x_2+2x_3=2\}$. 超平面与法向量如图 1.6 所示.

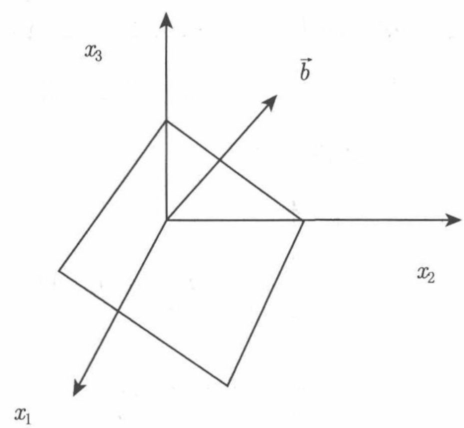

图 1.6 例子 1.15 示意图

定理 1.4 给定 $b\in\mathbb{R}^m$, $B\in\mathbb{R}^{m\times n}$, 则集合 $M=\{x\in\mathbb{R}^n \mid Bx=b\}$ 是一个仿射集. 反之, 每个仿射集都可以这样表示.

证明 由仿射集定义易证 M 是一个仿射集. 设 L 是平行于 M 的子空间, $b_1,b_2,\cdots,b_m\in\mathbb{R}^n$ 是 $L^{\perp}$ 的基, 则

$$L=\{x \mid \langle x,b_i\rangle=0, i=1,2,\cdots,m\}=\{x \mid Bx=0\},$$

其中, $B=[b_1,\cdots,b_m]^{\mathrm{T}}\in\mathbb{R}^{m\times n}$. 又 M 平行于 L, 故存在 $a\in\mathbb{R}^n$, 使得

$$\begin{aligned} M &= L+a=\{x+a \mid Bx=0\} \\ &= \{x \mid B(x-a)=0\} \\ &= \{x \mid Bx=b\}, \end{aligned}$$

其中, $b=Ba$. 若 $M=\mathbb{R}^n$, 令 $B=0, b=0$. 定理得证. □

M 也可以写成 $M=\{x \mid \langle x,b_i\rangle=\beta_i, i=1,2,\cdots,m\}=\bigcap\limits_{i=1}^{m} H_i$, b_i 是 B 的第 i 个行向量, β_i 是 b 的第 i 个分量, 每个 $H_i=\{x \mid \langle x,b_i\rangle=\beta_i\}$