Semidefinite Programming

半定规划

修乃华　罗自炎　编著

北京交通大学出版社
http://www.bjtup.com.cn

半定规划

修乃华　罗自炎　编著

北京交通大学出版社
·北京·

内容简介

本书主要介绍半定规划的基本理论与典型应用，包括线性半定规划的基本理论、非线性半定规划的基本理论、线性与非线性半定规划的若干应用实例，如图的着色、量子计算、压缩感知，力求反映最新和最重要的成果，以期把读者引向该研究领域的最前沿。

本书可作为运筹学、系统科学、管理科学、计算机科学等有关专业的高校师生、科研人员、工程技术人员的参考书，同时也可作为大学有关专业的研究生和高年级本科生的教材。

图书在版编目（CIP）数据

半定规划／修乃华，罗自炎编著. —北京：北京交通大学出版社，2014.6
ISBN 978-7-5121-1949-9

Ⅰ. ①半…　Ⅱ. ①修…②罗…　Ⅲ. ①运筹学－最佳化－研究　Ⅳ. ①O224

中国版本图书馆CIP数据核字（2014）第129943号

责任编辑：王晓春
出版发行：北京交通大学出版社　　电话：010-51686414
北京市海淀区高梁桥斜街44号　　邮编：100044
印 刷 者：北京艺堂印刷有限公司
经　　销：全国新华书店
开　　本：185×260　印张：8.5　字数：181千字
版　　次：2014年6月第1版　2014年6月第1次印刷
书　　号：ISBN 978-7-5121-1949-9/O·135
印　　数：1～1 000册　定价：38.00元

本书如有质量问题，请向北京交通大学出版社质监组反映。对您的意见和批评，我们表示欢迎和感谢。
投诉电话：010-51686043，51686008；传真：010-62225406；E-mail：press@bjtu.edu.cn。

前　言

半定规划是近 20 年发展起来的一个新的数学规划分支，它与经典的线性与非线性规划模型的主要区别就在于其决策变量服从一个对称半正定矩阵锥约束. 线性半定规划不仅“形如”经典的线性规划，具有良好的数学性质，以及有多项式时间的内点算法，而且它与组合优化、多项式代数与几何、统计分析、不确定性优化等其他应用数学分支有着天然的内在联系，更为重要的是它在经济、金融、工程、管理等许多领域有着广泛的应用. 因此，它受到广大应用数学工作者和实际工作者的青睐，被誉为 21 世纪的线性规划和现代数学规划的摇篮.

本书主要介绍半定规划的基本理论与典型应用，包括线性半定规划的基本理论、非线性半定规划的基本理论、线性与非线性半定规划的若干应用实例，力求反映最新和最重要的成果. 此外，本书还扼要介绍了当今数学规划领域的一个热点研究课题——矩阵秩极小问题的松弛理论与方法，以期把读者引向该研究方向的最前沿.

本书可作为运筹学、系统科学、管理科学、计算机科学等有关专业的高校师生、科研人员、工程技术人员的参考书，同时也可作为大学有关专业的研究生和高年级本科生的教材.

借此机会，作者非常感谢中国科学院数学与系统科学研究院韩继业教授的长期指导与帮助，感谢大连理工大学张立卫教授和曲阜师范大学王宜举教授对本书认真细致的审阅；感谢国家 973 计划（2010CB732501）、国家自然科学基金（71271021，11301022）的资助，使得研究工作得以长期进行；感谢三位硕士生白霜华、何真强、张倩对于手稿的录入与校对；此外，作者感谢家人长期无私的理解与支持.

囿于作者水平，本书存在不妥或错误之处，欢迎读者批评指正.

修乃华　罗自炎

2014 年 6 月于北京

目　　录

第1章 预备知识

半定规划是带有半定锥约束的一类特殊矩阵规划，是经典的线性与非线性规划在矩阵空间中的推广. 本章给出学习半定规划所需要的矩阵理论与凸分析基础. 为求简洁，只给出一些关键的或有特色的定义和定理，并选择性地给出了一些重要证明或者证明纲要，省略了许多一般化的证明过程.

1.1 矩阵空间

记所有 $m\times n$ 阶实矩阵组成的集合为 $\mathbf{R}^{m\times n}$，由矩阵加法与数乘运算可构成一个维数为 mn 的线性空间. 进一步，引入如下形式的内积

$$\langle \boldsymbol{A},\boldsymbol{B}\rangle := \mathrm{tr}(\boldsymbol{B}^{\top}\boldsymbol{A})=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}b_{ij}, \forall \boldsymbol{A}=[a_{ij}], \boldsymbol{B}=[b_{ij}]\in \mathbf{R}^{m\times n}, \tag{1-1}$$

则该集合可构成一个欧氏空间，称为 $m\times n$ 阶实矩阵空间，简记为 $\mathbf{R}^{m\times n}$，其中 $\mathrm{tr}(\bullet)$ 为方阵的迹算子，表示相应矩阵的所有主对角线元素之和，$\boldsymbol{B}^{\top}$ 为矩阵 $\boldsymbol{B}$ 的转置，$\boldsymbol{B}^{\top}\boldsymbol{A}$ 为矩阵乘积. 容易验证，对于任意 $\boldsymbol{A}$，$\boldsymbol{B}$，$\boldsymbol{C}\in\mathbf{R}^{n\times n}$ 迹算子满足如下性质

$$\mathrm{tr}(\boldsymbol{A})=\mathrm{tr}(\boldsymbol{A}^{\top}), \mathrm{tr}(\boldsymbol{AB})=\mathrm{tr}(\boldsymbol{BA}),\ \mathrm{tr}(\boldsymbol{ABC})=\mathrm{tr}(\boldsymbol{BCA})=\mathrm{tr}(\boldsymbol{CAB}).$$

由式（1-1）中定义的内积可以诱导出矩阵空间 $\mathbf{R}^{m\times n}$ 的一种重要范数，称为 Frobenius 范数，定义为

$$\|\boldsymbol{A}\|_{\mathrm{F}} := \sqrt{\langle \boldsymbol{A},\boldsymbol{A}\rangle}=\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}^{2}}. \tag{1-2}$$

利用欧氏空间同构的性质可知，矩阵空间 $\mathbf{R}^{m\times n}$ 与向量空间 $\mathbf{R}^{mn}$ 是同构的，其相应的同构映射为

$$\mathrm{vec}(\boldsymbol{A}) := (a_{11}, a_{21},\cdots,a_{m1},a_{12},\cdots,a_{mn})^{\top}, \forall \boldsymbol{A}=[a_{ij}]\in\mathbf{R}^{m\times n}.$$

显然，对于任意 A， $B\in\mathbf{R}^{m\times n}$，

$$\langle A,B\rangle=\operatorname{vec}(A)^{\top}\operatorname{vec}(B),\ \|A\|_{\mathrm{F}}=\|\operatorname{vec}(A)\|,$$

其中$\|\bullet\|$为向量的欧氏范数.

在实矩阵空间 $\mathbf{R}^{m\times n}$ 中，由所有 $n\times n$ 阶实对称矩阵组成的子空间简记为 S^n．容易验证：S^n 是 $\dfrac{n(n+1)}{2}$ 维欧氏空间，且通过如下同构映射与 $\mathbf{R}^{\frac{n(n+1)}{2}}$ 同构

$$\operatorname{svec}(A):=(a_{11},\sqrt{2}a_{21},\cdots,\sqrt{2}a_{n1},a_{22},\sqrt{2}a_{32},\cdots,a_{nn})^{\top},\ \forall A=[a_{ij}]\in S^n. \quad (1\text{-}3)$$

此外，对于任意 A， $B\in S^n$，有如下等式关系成立

$$\langle A,B\rangle=\operatorname{svec}(A)^{\top}\operatorname{svec}(B),\|A\|_{\mathrm{F}}=\|\operatorname{svec}(A)\|.$$

尽管空间 S^n 中的内积与相应的 Frobenius 范数均能通过同构映射式（1-3）转化为向量空间 $\mathbf{R}^{\frac{n(n+1)}{2}}$ 的内积与欧氏范数，但该空间也具有自身的一些特有性质与结构，而这些性质与结构是 $\mathbf{R}^{\frac{n(n+1)}{2}}$ 无法表现的.

定理 1.1.1 （谱分解定理） 对于任意矩阵 $A\in S^n$，存在不减实数序列 $\lambda_1(A),\cdots,\lambda_n(A)$，以及一组标准正交向量组 $v_1,\cdots,v_n\in\mathbf{R}^n$，其中 $P=[v_1,\cdots,v_n]\in\mathbf{R}^{n\times n}$，使得

$$A=\sum_{i=1}^{n}\lambda_i(A)v_iv_i^{\top}=P\operatorname{Diag}(\lambda_1(A),\cdots,\lambda_n(A))P^{\top}, \quad (1\text{-}4)$$

其中 $P:=[v_1,\cdots,v_n]$ 为正交矩阵， $\operatorname{Diag}(\lambda_1(A),\cdots,\lambda_n(A))$ 表示以 $\lambda_1(A),\cdots,\lambda_n(A)$ 为主对角元素的对角矩阵. 称实数 $\lambda_i(A)$ 为矩阵 A 的特征值，v_i 为相应的特征向量. 式(1-4)为对称矩阵 A 的特征值分解或谱分解.

利用矩阵的特征值可以刻画矩阵的迹、行列式及秩函数如下

$$\operatorname{tr}(A)=\sum_{i=1}^{n}\lambda_i(A),\ \det(A)=\prod_{i=1}^{n}\lambda_i(A),\ \operatorname{rank}(A)=|\{i:\lambda_i(A)\neq 0\}|,$$

其中$|\cdot|$表示相应集合所含元素的个数. 显然上述性质是 $\mathbf{R}^{\frac{n(n+1)}{2}}$ 无法表现的. 这也就从某种程度上说明了独立研究空间 S^n 的必要性. 需要指出的是，在矩阵空间 $\mathbf{R}^{m\times n}$ 中存在奇异值分解定理，通过奇异值所反映出来的性质同样能体现出 $\mathbf{R}^{m\times n}$ 具有与 $\mathbf{R}^{mn}$ 不同的结构与性质，具体的内容可参考文献［5］.

在实对称矩阵空间 S^n 中还有一类特殊的矩阵，称为半正定矩阵，组成的集合简记为 S_+^n，其定义为

$$S_+^n := \{A \in S^n : \boldsymbol{x}^{\mathrm{T}} A\boldsymbol{x} \geqslant 0, \forall \boldsymbol{x} \in \mathbf{R}^n\}. \tag{1-5}$$

特别地，若对于任意 $\boldsymbol{x} \in \mathbf{R}^n \setminus \{\mathbf{0}\}, \boldsymbol{x}^{\mathrm{T}} A\boldsymbol{x} > 0$，则称对称矩阵 A 是正定矩阵. 所有 n 阶实对称正定矩阵组成的集合简记为 S_{++}^n. 下面回顾半正定矩阵的一些重要性质.

定理 1.1.2　设 $A \in S^n$，则下列结论是等价的:

（i）$A \in S_+^n$；

（ii）$\lambda_i(A) \geqslant 0, i = 1, \cdots, n$；

（iii）$\boldsymbol{x}^{\mathrm{T}} A\boldsymbol{x} \geqslant 0, \forall \boldsymbol{x} \in \mathbf{R}^n$；

（iv）A 的所有主子式非负;

（v）存在矩阵 $L \in \mathbf{R}^{m\times n}$ 使得 $A = L^{\mathrm{T}} L$，其中 $\mathrm{rank}(L) = \mathrm{rank}(A)$.

类似的，正定矩阵也具有如下性质.

定理 1.1.3　设 $A \in S^n$，则下列结论是等价的:

（i）$A \in S_{++}^n$；

（ii）$\lambda_i(A) > 0, i = 1, \cdots, n$；

（iii）$\boldsymbol{x}^{\mathrm{T}} A\boldsymbol{x} > 0, \forall \boldsymbol{x} \in \mathbf{R}^n \setminus \{\mathbf{0}\}$；

（iv）A 的所有主子式为正数;

（v）存在矩阵 $L \in \mathbf{R}^{m\times n}$ 使得 $A = L^{\mathrm{T}} L$，其中 $\mathrm{rank}(L)=n$.

利用定理 1.1.2（iii）与定理 1.1.3（iii）可知，矩阵的半正定性（正定性）在合同变换下是不变的.

命题 1.1.4　若矩阵 $A \in S_+^n$，则对于任意 $B \in \mathbf{R}^{n\times m}$，$B^{\mathrm{T}} AB \in S_+^m$. 特别地，$A \in S_+^n \Leftrightarrow B^{\mathrm{T}} AB \in S_+^n, \forall B \in \mathbf{R}^{n\times n}$ 且 B 非奇异; $A \in S_{++}^n \Leftrightarrow B^{\mathrm{T}} AB \in S_{++}^n, \forall B \in \mathbf{R}^{n\times n}$ 且 B 非奇异.

证明　若 $A \in S_+^n$，则对于 $B \in \mathbf{R}^{n\times m}$，任意 $\boldsymbol{x} \in \mathbf{R}^m$，有

$$\boldsymbol{x}^{\mathrm{T}} (B^{\mathrm{T}} AB)\boldsymbol{x} = (B\boldsymbol{x})^{\mathrm{T}} A(B\boldsymbol{x}) \geqslant 0,$$

从而 $B^{\mathrm{T}} AB \in S_+^m$. 特别地，当矩阵 B 是非奇异矩阵时，对于任意 $\boldsymbol{y} \in \mathbf{R}^n$，有

$$\boldsymbol{y}^{\mathrm{T}} A\boldsymbol{y} = (B^{-1}\boldsymbol{y})^{\mathrm{T}} B^{\mathrm{T}} AB(B^{-1}\boldsymbol{y}),$$

从而可得到命题后半部分的等价性. ■

利用命题 1.1.4 的等价性结论，可以得到半正定矩阵的一个重要定理——Schur 补定理.

定理 1.1.5　设 $A \in S_{++}^m, C \in S^n, B \in \mathbf{R}^{m\times n}$，则

$$\begin{bmatrix} A & B \\ B^{\mathrm{T}} & C \end{bmatrix} \in S_+^{m+n} (\text{或} S_{++}^{m+n}) \Leftrightarrow C - B^{\mathrm{T}} A^{-1} B \in S_+^n (\text{或} S_{++}^n) \tag{1-6}$$

证明 设矩阵 $\boldsymbol{D}:=-\boldsymbol{A}^{-1}\boldsymbol{B}$，由矩阵乘法可知

$$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{D}^{\mathrm{T}} & \boldsymbol{I} \end{bmatrix}\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^{\mathrm{T}} & \boldsymbol{C} \end{bmatrix}\begin{bmatrix} \boldsymbol{I} & \boldsymbol{D} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}=\begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}-\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{B} \end{bmatrix}. \tag{1-7}$$

由命题 1.1.4、定理 1.1.2（iv）及定理 1.1.3（iv）可以得到相应的等价性. ■

利用定理 1.1.2 （ii）中半正定矩阵特征值的非负性及谱分解定理，可以定义半正定矩阵的对称平方根矩阵

$$\boldsymbol{A}^{\frac{1}{2}}:=\boldsymbol{P}\mathrm{Diag}(\sqrt{\lambda_1(\boldsymbol{A})},\cdots,\sqrt{\lambda_n(\boldsymbol{A})})\boldsymbol{P}^{\mathrm{T}},$$

其中 $\boldsymbol{A}$ 具有式（1-4）的谱分解形式. 类似地，对所有正整数 k，也可以基于相应的谱分解来定义 $\boldsymbol{A}^{\frac{1}{k}}(\boldsymbol{A}\in S_{+}^{n})$和$\boldsymbol{A}^{-\frac{1}{k}}(\boldsymbol{A}\in S_{++}^{n})$.

定理 1.1.6 （Rayleigh-Ritz 定理）设 $\boldsymbol{A}\in S^n$，则

$$\lambda_{\min}(\boldsymbol{A})=\min_{x\in\mathbf{R}^n}\{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}:\|\boldsymbol{x}\|^2=1\}=\min_{0\neq x\in\mathbf{R}^n}\left\{\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\|\boldsymbol{x}\|^2}\right\},$$

$$\lambda_{\max}(\boldsymbol{A})=\max_{x\in\mathbf{R}^n}\{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}:\|\boldsymbol{x}\|^2=1\}=\max_{0\neq x\in\mathbf{R}^n}\left\{\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\|\boldsymbol{x}\|^2}\right\}.$$

定理 1.1.7 对于任意给定矩阵 $\boldsymbol{A}=[a_{ij}]\in S^n$，称 $\boldsymbol{A}$ 为对角占优矩阵，若对于任意 $i\in\{1,\cdots,n\}$，$a_{ii}\geqslant\sum_{j\neq i}|a_{ij}|$；称矩阵 $\boldsymbol{A}$ 为严格对角占优矩阵，若对于任意 $i\in\{1,\cdots,n\}$，$a_{ii}>\sum_{j\neq i}|a_{ij}|$. 如果 $\boldsymbol{A}\in S^n$ 为对角（或严格对角）占优矩阵，则 $\boldsymbol{A}\in S_{+}^{n}$(或$\boldsymbol{A}\in S_{++}^{n}$).

证明 利用实对称矩阵的 Gerschgorin 定理可知，矩阵 $\boldsymbol{A}$ 的任意特征值 $\lambda_k(\boldsymbol{A})(k=1,\cdots,n)$ 均满足

$$\lambda_k(\boldsymbol{A})\in\bigcup_{i=1}^{n}\{t\,\|\,t-a_{ii}\,|\leqslant\sum_{i\neq j}|a_{ij}|\}.$$

若 $\boldsymbol{A}$ 为对角占优矩阵，由定义式 $a_{ii}\geqslant\sum_{i\neq j}|a_{ij}|$ 可推导出 $\lambda_k(\boldsymbol{A})\geqslant 0,\forall k=1,\cdots,n$. 利用定理 1.1.2（ii）有 $\boldsymbol{A}\in S_{+}^{n}$. 类似地，可以证明严格对角占优的情形. ■

此外，半正定矩阵还具有如下简单性质.

定理 1.1.8 对称半正定矩阵的所有主对角线元素均为非负数，特别地，对称正定矩阵的所有主对角线元素均为正数；由有限多个实对称矩阵组成的块对角矩阵是半正定（或正定）的当且仅当每个实对称矩阵是半正定（或正定）的；若实对称半正定矩阵的某个对角元素为零，则该对角元素所在列与行的元素均

为零.

利用上述定理，容易判定如下矩阵的半正定性.

$$\boldsymbol{A}=\begin{bmatrix}5&1&-2&1\\1&1&7&-1\\-2&7&0&0\\1&-1&0&5\end{bmatrix},\boldsymbol{B}=\begin{bmatrix}5&1&0&0\\1&0&0&0\\0&0&-1&0\\0&0&0&5\end{bmatrix}$$

显然矩阵 $\boldsymbol{A}$ 不是半正定矩阵，因为第三个主对角元素为 0，而相应的行元素不都为 0；矩阵 $\boldsymbol{B}$ 也不是半正定的，因为存在负的主对角元素.

1.2　凸集与半定锥

凸集是凸分析的一个基本概念，而半定锥作为矩阵空间中的一个特殊凸集，它是半定规划的基础之一. 本节主要回顾与探讨凸集和半定锥的一些基本性质，为半定规划的理论分析做准备.

定义 1.2.1　设 C 是有限维 Hilbert 空间 X 中的一个非空子集，若对于 C 中任意两点，连接它们的线段也在 C 中，即

$$\alpha\boldsymbol{x}+(1-\alpha)\boldsymbol{y}\in C,\forall\boldsymbol{x},\boldsymbol{y}\in C,\forall\alpha\in[0,1],\tag{1-8}$$

则称集合 C 是空间 X 中的一个凸集.

图 1-1 给出了 $\mathbf{R}^2$ 空间中的两个凸集（左侧）和两个非凸集（右侧）.
下面给出矩阵空间 S^n 中几个重要的凸集.

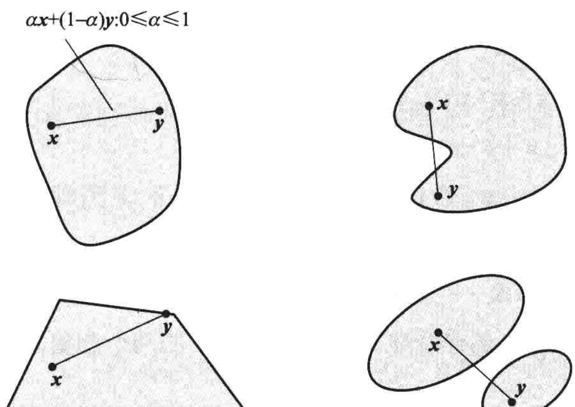

图 1-1　$\mathbf{R}^2$ 中的凸集与非凸集

例 1.2.2 S_+^n 是 S^n 中的凸集. 事实上，对于任意半正定矩阵 $\boldsymbol{A}, \boldsymbol{B} \in S_+^n$，任意 $\lambda \in [0,1]$，

$$\boldsymbol{x}^{\mathrm{T}}(\lambda \boldsymbol{A} + (1-\lambda)\boldsymbol{B})\boldsymbol{x} = \lambda \boldsymbol{x}^{\mathrm{T}} \boldsymbol{A}\boldsymbol{x} + (1-\lambda)\boldsymbol{x}^{\mathrm{T}} \boldsymbol{B}\boldsymbol{x} \geqslant 0, \forall \boldsymbol{x} \in \mathbf{R}^n.$$

因此，$\lambda \boldsymbol{A} + (1-\lambda)\boldsymbol{B} \in S_+^n$，即 S_+^n 是一个凸集.

例 1.2.3 $C_n := \{\boldsymbol{A} \in S^n : \boldsymbol{x}^{\mathrm{T}} \boldsymbol{A}\boldsymbol{x} \geqslant 0, \forall \boldsymbol{x} \in \mathbf{R}_+^n\}$ 是 S^n 中的凸集.

例 1.2.4 $C_n^* := \{\boldsymbol{A} \in S^n : \boldsymbol{A} = \sum_{i=1}^{m} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}}, \boldsymbol{x}_i \in \mathbf{R}_+^n, i = 1, \cdots, m, m \geqslant 1\}$ 是 S^n 中的凸集.

例 1.2.5 $N_n := \{\boldsymbol{A} \in S^n : \boldsymbol{A} = [a_{ij}], a_{ij} \geqslant 0, \forall i, j = 1, \cdots, n\}$ 是 S^n 中的凸集.

例 1.2.6 $M := \{\boldsymbol{S} \in S^n : \boldsymbol{S} = \boldsymbol{C} - \sum_{i=1}^{m} y_i \boldsymbol{A}_i, y_i \in \mathbf{R}, i = 1, 2, \cdots, m\}$ 是 S^n 中的凸面，其中 $\boldsymbol{C}$，$\boldsymbol{A}_1, \cdots, \boldsymbol{A}_m \in S^n$. 事实上，对于任意 $\boldsymbol{S}, \boldsymbol{S}' \in M$，存在 $\boldsymbol{y}, \boldsymbol{y}' \in \mathbf{R}^m$ 使得

$$\boldsymbol{S} = \boldsymbol{C} - \sum_{i=1}^{m} y_i \boldsymbol{A}_i, \boldsymbol{S}' = \boldsymbol{C} - \sum_{i=1}^{m} y_i' \boldsymbol{A}_i.$$

对于任意 $\lambda \in [0,1]$，有

$$\begin{aligned}\lambda \boldsymbol{S} + (1-\lambda)\boldsymbol{S}' &= \lambda(\boldsymbol{C} - \sum_{i=1}^{m} y_i \boldsymbol{A}_i) + (1-\lambda)(\boldsymbol{C} - \sum_{i=1}^{m} y_i' \boldsymbol{A}_i) \\ &= \boldsymbol{C} - \sum_{i=1}^{m} [\lambda y_i + (1-\lambda) y_i'] \boldsymbol{A}_i,\end{aligned}$$

从而，$\lambda \boldsymbol{S} + (1-\lambda)\boldsymbol{S}' \in M$. 于是 M 是一个凸集.

凸集具有如下基本性质.

定理 1.2.7 设 C_1, C_2 是有限维 Hilbert 空间 X 的两个凸子集，则

（i）$\alpha C_1 := \{\alpha \boldsymbol{x} : \boldsymbol{x} \in C_1\}$ 是凸集，其中 $\alpha \in \mathbf{R}, \alpha \geqslant 0$；

（ii）$C_1 \cap C_2$ 是凸集；

（iii）$C_1 + C_2 := \{\boldsymbol{x} + \boldsymbol{y} : \boldsymbol{x} \in C_1, \boldsymbol{y} \in C_2\}$ 是凸集；

（iv）$C_1 - C_2 := \{\boldsymbol{x} - \boldsymbol{y} : \boldsymbol{x} \in C_1, \boldsymbol{y} \in C_2\}$ 是凸集；

（v）$\mathscr{A} C_1 := \{\mathscr{A} \boldsymbol{x} : \boldsymbol{x} \in C_1\}$ 是凸集，其中 $\mathscr{A}$ 是从 X 到有限维 Hilbert 空间 Y 的线性算子.

在凸集理论中还涉及一个重要的定理——凸集分离定理.

定理 1.2.8 设 C, D 是有限维 Hilbert 空间 X 中的两个凸集，则存在 $a \in X^* \setminus \{\boldsymbol{0}\}$ 使得

$$\inf_{x\in D}\langle a, \boldsymbol{x}\rangle \geqslant \sup_{y\in C}\langle a, \boldsymbol{y}\rangle \text{ 且 } \sup_{x\in D}\langle a, \boldsymbol{x}\rangle > \inf_{y\in C}\langle a, \boldsymbol{y}\rangle$$

当且仅当 $\mathrm{ri}(C)\cap \mathrm{ri}(D)=\varnothing$, 其中 $\mathrm{ri}(C)$ 表示凸集 C 的相对内部，即 $\mathrm{ri}(C):=\{\boldsymbol{x}\in C:\forall \tilde{\boldsymbol{x}}\in C, \exists \bar{\boldsymbol{x}}\in C, \exists \lambda\in(0,1)$，使得 $\boldsymbol{x}=\lambda\tilde{\boldsymbol{x}}+(1-\lambda)\bar{\boldsymbol{x}}\}$, X^* 是 X 的共轭空间.

【注】(i) 当集合 C 的内部非空时，相对内部即为内部，也就是说，$\mathrm{ri}(C)=\mathrm{int}(C)$ 如图 1-2 的情形；(ii) 上述凸集分离定理给出了两个非空凸集能够分离的充要条件，凸集分离实际上还包括真分离、严格分离、强分离等多种情形，具体内容可参考文献［14］中的 Theorem 11.1.

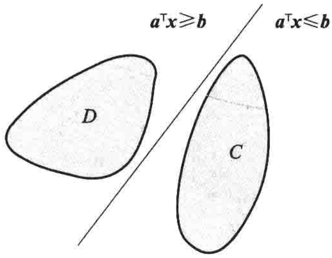

图 1-2　凸集分离

在凸集中还有一类重要的特殊情形——凸锥.

定义 1.2.9　称集合 $K\subseteq X$ 是一个锥，若对于任意 $\boldsymbol{x}\in K$ 和任意 $\lambda\geqslant 0$，总有 $\lambda\boldsymbol{x}\in K$. 进一步，若 K 还满足凸性，则称集合 K 是一个凸锥. 若 $K\cap(-K)=\{\boldsymbol{0}\}$，则称锥 K 为尖锥.

由定义可知，对于锥 K 而言，K 是凸锥当且仅当 $\boldsymbol{x}+\boldsymbol{y}\in K$，$\forall \boldsymbol{x},\boldsymbol{y}\in K$. 容易验证，例 1.2.2～例 1.2.5 中出现的凸集 S_+^n，C_n，C_n^*，N_n 均为 S^n 中的凸锥，分别称为半定锥、共正锥、完正锥、非负矩阵锥. 本书所考虑的半定规划就是基于半定锥而展开的，因此本节的后半部分主要介绍半定锥的一些重要代数特征与几何结构.

命题 1.2.10　S_+^n 是 S^n 中的一个闭尖凸锥且具有非空内部，即 $\mathrm{int}(S_+^n)=S_{++}^n$.

证明　由例 1.2.2 及锥的定义可知，S_+^n 是 S^n 的一个凸锥. 对于 $\forall \boldsymbol{A}\in S_+^n\cap(-S_+^n)$，有

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}\geqslant 0, -\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}\geqslant 0, \forall \boldsymbol{x}\in \mathbf{R}^n .$$

从而 $\boldsymbol{A}=\boldsymbol{0}$，因此 S_+^n 是尖锥. 由定理 1.1.2 (ii) 及特征值函数的连续性可知，S_+^n 是

闭集. 进一步，利用集合内部的定义及定理 1.1.3（ii）可知，$\operatorname{int}(S_+^n) = S_{++}^n$. ■

图 1-3 是二阶半定锥 S_+^2 的图像.

因为半定锥 S_+^n 是 S^n 空间中的一个闭尖凸锥,可以利用这一锥诱导出该空间中的一种偏序，称为 Löwner 偏序，即

$$\boldsymbol{A} \succeq \boldsymbol{B} \Leftrightarrow \boldsymbol{A}-\boldsymbol{B} \in S_+^n; \boldsymbol{A} \succ \boldsymbol{B} \Leftrightarrow \boldsymbol{A}-\boldsymbol{B} \in S_{++}^n.$$

有时将 $\boldsymbol{A} \in S_+^n$ 与 $\boldsymbol{A} \in S_{++}^n$ 分别简记为 $\boldsymbol{A} \succeq \mathbf{0}$与$\boldsymbol{A} \succ \mathbf{0}$.

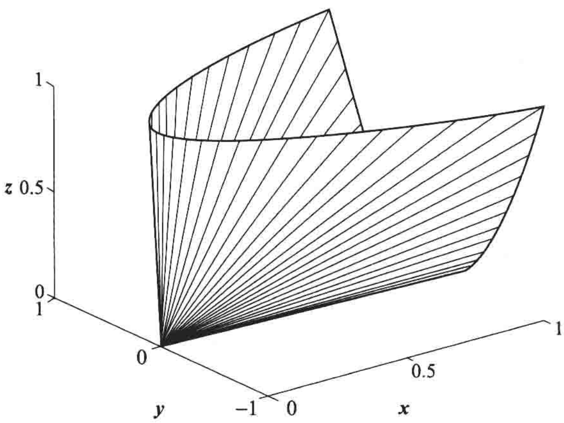

图 1-3　二阶半定锥 S_+^2

对于有限维 Hilbert 空间 X 的某个给定的锥 K，在其对偶空间 X^* 中可以引入对偶锥的概念. 称集合

$$K^* := \{\boldsymbol{y} \in X^* : \langle \boldsymbol{x}, \boldsymbol{y} \rangle \geqslant 0, \forall \boldsymbol{x} \in K\} \tag{1-9}$$

为锥 K 的对偶锥. 显然，对偶锥必定是非空闭凸锥. 若 $X = X^*$ 且 $K = K^*$，则称 K 是自对偶锥.

命题 1.2.11 $S_+^n = (S_+^n)^*$.

证明 对任意 $\boldsymbol{A} \in S_+^n$，由迹算子的性质可知，对 $\forall \boldsymbol{B} \in S_+^n$ 有

$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \operatorname{tr}(\boldsymbol{BA}) = \operatorname{tr}\left(\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{A}^{\frac{1}{2}}\right) = \operatorname{tr}\left(\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{A}^{\frac{1}{2}}\right) = \left\|\boldsymbol{B}^{\frac{1}{2}}\boldsymbol{A}^{\frac{1}{2}}\right\|_{\mathrm{F}}^2 \geqslant 0, \tag{1-10}$$

从而 $\boldsymbol{A} \in (S_+^n)^*$. 由 $\boldsymbol{A}$ 的任意性可知，$S_+^n \subseteq (S_+^n)^*$. 另一方面，对 $\forall \boldsymbol{A} \in (S_+^n)^*$，由对偶锥的定义可知 $\langle \boldsymbol{A}, \boldsymbol{B} \rangle \geqslant 0, \forall \boldsymbol{B} \in S_+^n$. 注意到对任意 $\boldsymbol{x} \in \mathbf{R}^n$, $\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}} \in S_+^n$. 因此，$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \langle \boldsymbol{A}, \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}} \rangle \geqslant 0$，即 $\boldsymbol{A} \in S_+^n$. 于是得到 $(S_+^n)^* \subseteq S_+^n$. ■

由半定锥的自对偶性可知，任意两个半定矩阵的内积是非负数. 进一步，还能得到如下代数性质.

命题 1.2.12 设 $\boldsymbol{A}, \boldsymbol{B} \in S^n$，则有

（i）$\boldsymbol{A} \in S_+^n \Leftrightarrow \langle \boldsymbol{A}, \boldsymbol{H} \rangle \geqslant 0, \forall \boldsymbol{H} \in S_+^n$；

（ii）$\boldsymbol{A} \in S_{++}^n \Leftrightarrow \langle \boldsymbol{A}, \boldsymbol{H} \rangle > 0, \forall \boldsymbol{H} \in S_+^n \setminus \{\boldsymbol{0}\}$；

（iii）若 $\boldsymbol{A}, \boldsymbol{B} \in S_+^n$，则 $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = 0 \Leftrightarrow \boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{0}$；

（iv）若 $\boldsymbol{A}, \boldsymbol{B} \in S_+^n$，则 $\lambda_{\min}(\boldsymbol{A})\lambda_{\max}(\boldsymbol{B}) \leqslant \lambda_{\min}(\boldsymbol{A})\mathrm{tr}(\boldsymbol{B}) \leqslant \langle \boldsymbol{A}, \boldsymbol{B} \rangle \leqslant \lambda_{\max}(\boldsymbol{A})\mathrm{tr}(\boldsymbol{B}) \leqslant n\lambda_{\max}(\boldsymbol{A})\lambda_{\max}(\boldsymbol{B})$.

证明 （i）可以由半定锥的自对偶性直接验证得到. 下面先证明（iii）. 由于 $\boldsymbol{A}, \boldsymbol{B} \in S_+^n$，利用式（1-10）可知 $0 = \langle \boldsymbol{A}, \boldsymbol{B} \rangle = \left\| \boldsymbol{A}^{\frac{1}{2}} \boldsymbol{B}^{\frac{1}{2}} \right\|_{\mathrm{F}}^2$，从而 $\boldsymbol{A}^{\frac{1}{2}} \boldsymbol{B}^{\frac{1}{2}} = \boldsymbol{0}$. 两边同时左乘 $\boldsymbol{A}^{\frac{1}{2}}$，右乘 $\boldsymbol{B}^{\frac{1}{2}}$，得到 $\boldsymbol{AB} = \boldsymbol{0}$. 同理可证 $\boldsymbol{BA} = \boldsymbol{0}$. 再证明（iv）. 不妨设 $\boldsymbol{A}$ 具有谱分解形式 $\boldsymbol{A} = \boldsymbol{P\Lambda P}^{\mathrm{T}}$，其中 $\boldsymbol{\Lambda}$ 为 $\boldsymbol{A}$ 的所有特征值组成的对角矩阵，$\boldsymbol{P}$ 为相应的特征向量组成的标准正交矩阵. 由定理 1.1.2（ii），结合迹算子及标准正交矩阵的性质可知

$$\begin{aligned}\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \mathrm{tr}(\boldsymbol{P\Lambda P}^{\mathrm{T}}\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{\Lambda P}^{\mathrm{T}}\boldsymbol{BP}) &\geqslant \lambda_{\min}(\boldsymbol{A})\mathrm{tr}(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{BP}) \\ &= \lambda_{\min}(\boldsymbol{A})\mathrm{tr}(\boldsymbol{B}) \geqslant \lambda_{\min}(\boldsymbol{A})\lambda_{\max}(\boldsymbol{B}).\end{aligned}$$

类似可证其他不等式. 最后来证（ii）. 若 $\boldsymbol{A} \in S_{++}^n$，用反证法，假设存在 $\boldsymbol{H} \in S_+^n \setminus \{\boldsymbol{0}\}$ 使得 $\langle \boldsymbol{A}, \boldsymbol{H} \rangle = 0$. 由定理 1.1.3（ii）可知 $\lambda_{\min}(\boldsymbol{A}) > 0$. 另一方面，由 $\boldsymbol{H} \in S_+^n \setminus \{\boldsymbol{0}\}$ 可得 $\mathrm{tr}(\boldsymbol{H}) > 0$. 利用（iv）可推出

$$0 = \langle \boldsymbol{A}, \boldsymbol{H} \rangle \geqslant \lambda_{\min}(\boldsymbol{A})\mathrm{tr}(\boldsymbol{H}) > 0,$$

产生矛盾，从而得到 $\langle \boldsymbol{A}, \boldsymbol{H} \rangle > 0$. 反过来，若（ii）的右端式成立，用反证法，假设 $\boldsymbol{A} \in S_+^n \setminus S_{++}^n$，则 $\lambda_{\min}(\boldsymbol{A}) = 0$. 设

$$\boldsymbol{A} = [\boldsymbol{P}_1\ \boldsymbol{P}_2]\begin{bmatrix} \boldsymbol{\Lambda}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}[\boldsymbol{P}_1\ \boldsymbol{P}_2]^{\mathrm{T}},$$

其中，$\boldsymbol{P}_1$ 为 $\boldsymbol{A}$ 的非零特征值对应的特征向量按列组成的列正交矩阵，$\boldsymbol{P}_2$ 为 $\boldsymbol{A}$ 的零特征值对应的特征向量按列组成的列正交矩阵，$\boldsymbol{\Lambda}_1$ 为 $\boldsymbol{A}$ 的非零特征值构成的对角矩阵. 取 $\boldsymbol{H} = \boldsymbol{P}_2\boldsymbol{P}_2^{\mathrm{T}}$，显然 $\boldsymbol{H} \in S_+^n$ 且 $\boldsymbol{H} \neq \boldsymbol{0}$. 此时

$$\langle \boldsymbol{H}, \boldsymbol{A} \rangle = \mathrm{tr}(\boldsymbol{P}_2\boldsymbol{P}_2^{\mathrm{T}}\boldsymbol{P}_1\boldsymbol{\Lambda}_1\boldsymbol{P}_1^{\mathrm{T}}) = \mathrm{tr}(\boldsymbol{P}_2^{\mathrm{T}}\boldsymbol{P}_1\boldsymbol{\Lambda}_1\boldsymbol{P}_1^{\mathrm{T}}\boldsymbol{P}_2) = 0,$$

矛盾. ■

由上述命题的结论（i）不难得到，若 $\boldsymbol{S} \in S_+^n, \bar{\boldsymbol{S}} \in S^n$，则 $\bar{\boldsymbol{S}} \in S_+^n$ 当且仅当 $\boldsymbol{S} + t\bar{\boldsymbol{S}} \in S_+^n, \forall t \geqslant 0$. 后者说明在 $\boldsymbol{S}$ 处沿着 $\bar{\boldsymbol{S}}$ 方向生成的射线均位于半定锥 S_+^n 中. 该结果是半定锥的一个重要的几何特征. 半定锥的几何性质在刻画半定规划的最优性条件中起到关键作用. 下面简要介绍这方面的内容，包括半定锥的面结构、切锥、

临界锥及二阶切集的表达形式.

定义 1.2.13 称凸集 $F\subseteq C$ 是凸集 $C\subseteq X$ 的一个面，若对于任意 $\boldsymbol{x},\boldsymbol{y}\in C$ 满足 $\alpha\boldsymbol{x}+(1-\alpha)\boldsymbol{y}\in F$ 其中 $\alpha\in(0,1)$ 为某一实数，可以推出 $\boldsymbol{x},\boldsymbol{y}\in F$.

半定锥的面可以通过低阶半定锥来刻画.

定理 1.2.14 S_+^n 的任何面 F 必定具有如下形式之一:

（i）$F=\varnothing$；（ii）$F=\{\boldsymbol{0}\}$；（iii）$F=S_+^n$；

（iv）存在矩阵 $\boldsymbol{P}\in\mathbf{R}^{n\times k}$ ， $\operatorname{rank}(\boldsymbol{P})=k(1\leqslant k<n)$ ， $F=\{\boldsymbol{X}\in S^n:\boldsymbol{X}=\boldsymbol{PWP}^\top,\ \boldsymbol{W}\in S_+^k\}$.

证明 由定义可知（i）、（ii）或（iii）中定义的集合均为 S_+^n 的面. 下面考察由（iv）中定义的 F. 由锥的定义及 S_+^k 为凸锥易知 F 为凸锥. 用反证法，假设存在 $\boldsymbol{A},\boldsymbol{B}\in S_+^n$ 及 $\alpha\in(0,1)$ 使得 $\boldsymbol{X}=\alpha\boldsymbol{A}+(1-\alpha)\boldsymbol{B}\in F$ ，但是 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 中至少一个元素不在 F 中，不妨设 $\boldsymbol{A}\notin F$ ，即对于任意 $\boldsymbol{W}\in S_+^k$， $\boldsymbol{A}\neq\boldsymbol{PWP}^\top$. 于是存在非零向量 $\boldsymbol{y}\in\mathbf{R}^n$ 使得 $\boldsymbol{P}^\top\boldsymbol{y}=0$ 但 $\boldsymbol{y}^\top\boldsymbol{Ay}\neq0$. 结合 $\boldsymbol{A}$ 的半正定性得 $\boldsymbol{y}^\top\boldsymbol{Ay}>0$. 由假设条件 $\boldsymbol{X}\in F$ ，存在 $\boldsymbol{W}_0\in S_+^k$ ，使得 $\boldsymbol{X}=\boldsymbol{PW}_0\boldsymbol{P}^\top$ ，从而

$$0=\boldsymbol{y}^\top\boldsymbol{Xy}=\alpha\boldsymbol{y}^\top\boldsymbol{Ay}+(1-\alpha)\boldsymbol{y}^\top\boldsymbol{By}>0,$$

矛盾. 结合其凸性可知，形如（iv）的集合必是 S_+^n 的面.

下面证明 S_+^n 的任意一个面必定具有（i）～（iv）之一的形式. 对于 S_+^n 的任意一个非空非零的面 $F\neq S_+^n$ ，因为 $\{\boldsymbol{0}\}\subset F$ ，所以存在非零矩阵 $\hat{\boldsymbol{X}}\in\operatorname{ri}(F)$. 记 $k=\operatorname{rank}(\hat{\boldsymbol{X}})$ ，记 $\boldsymbol{P}\in\mathbf{R}^{n\times k}$ 为矩阵 $\hat{\boldsymbol{X}}$ 的 k 个非零特征值对应的特征向量按列排成的矩阵，并记 $\hat{F}=\{\boldsymbol{X}\in S^n:\boldsymbol{X}=\boldsymbol{PWP}^\top,\boldsymbol{W}\in S_+^k\}$. 由前面的讨论可知 $\hat{F}$ 是 S_+^n 的一个面. 下证 $F=\hat{F}$. 由于 $\hat{\boldsymbol{X}}\in\operatorname{ri}(F)\cap\operatorname{ri}(\hat{F})$.，利用文献［14］的推论 18.1.2 可知 F 与 $\hat{F}$ 是同一个面. 于是结论得证. ■

在约束优化问题的最优性分析中，切锥在一阶最优性条件中发挥着重要作用. 下面给出半定锥这一特殊凸集的切锥的表达形式. 在此之前，先简单回顾一下切锥的基本定义. 切锥有多种类型，这里主要考虑 Bouligand 切锥.

定义 1.2.15 设 X 为一 Bannach 空间，对于任一给定集合 $S\subset X$ 及任意给定的点 $\boldsymbol{x}\in S$ ，集合 S 在点 $\boldsymbol{x}$ 处的切锥定义为

$$T_S(\boldsymbol{x}):=\limsup_{t\downarrow0}\frac{S-\boldsymbol{x}}{t}=\{\boldsymbol{h}\in X:\exists t_n\downarrow0,\operatorname{dist}(\boldsymbol{x}+t_n\boldsymbol{h},S)=o(t_n)\},$$

其中距离函数 $\operatorname{dist}(\boldsymbol{x},S):=\inf_{\boldsymbol{y}\in S}\|\boldsymbol{x}-\boldsymbol{y}\|$. 若 S 为凸集，则有

$$T_S(\boldsymbol{x})=\{\boldsymbol{h}\in X:\mathrm{dist}(\boldsymbol{x}+t\boldsymbol{h},S)=o(t),t\geqslant 0\}.$$

例 1.2.16　取 $X=\mathbf{R}^3$，$S=\{(x_1,x_2,x_3)^{\mathrm{T}}:\ x_1^2+x_2^2+x_3^2=1\}$，$\boldsymbol{x}=(1,0,0)^{\mathrm{T}}$，则 $T_S(\boldsymbol{x})=\{(0,u,v)^{\mathrm{T}}:u,v\in\mathbf{R}\}$.

特别地，若凸集 S 由连续可微函数的等式与不等式约束来刻画时，其切锥在一定条件下可由约束函数的一阶方向导数来表示.

命题 1.2.17　设 $h_1,\cdots,h_m$ 为 $\mathbf{R}^n$ 上的仿射函数，$g_1,\cdots,g_l$ 为 $\mathbf{R}^n$ 上的凸函数，若相应的约束集合 $S:=\{\boldsymbol{x}\in\mathbf{R}^n:h_i(x)=0,i=1,\cdots,m;g_j(x)\geqslant 0,j=1,\cdots,l\}$ 满足 Slater 条件（详见第 3 章），则对于任意 $\boldsymbol{x}\in S$，集合 S 在 $\boldsymbol{x}$ 处的切锥为

$$T_S(\boldsymbol{x})=\{\boldsymbol{d}\in\mathbf{R}^n:\nabla h_i(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{d}=0,i=1,\cdots,m;\nabla g_j(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{d}\geqslant 0,j\in I(\boldsymbol{x})\},$$

其中 $I(\boldsymbol{x}):=\{j:g_j(\boldsymbol{x})=0\}$.

证明　利用一阶泰勒展开及凸集的切锥定义，结合 Slater 条件可以直接验证得到相应的结论. ■

由定理 1.1.2（ii），半定锥可以通过最小特征值函数的不等式来刻画，即

$$S_+^n=\{\boldsymbol{X}\in S^n:\lambda_{\min}(\boldsymbol{X})\geqslant 0\}.$$

利用上述结论，结合矩阵最小特征值函数的微分性质，可以得到半定锥的切锥表达形式.

命题 1.2.18　对于任意给定的矩阵 $\boldsymbol{X}\in S_+^n\setminus S_{++}^n$，设 $\boldsymbol{E}\in\mathbf{R}^{n\times s}$ 为 $\boldsymbol{X}$ 的零特征值对应的标准正交特征向量按列排成的矩阵，s 为零特征值的重数，则 S_+^n 在 $\boldsymbol{X}$ 处的切锥为

$$T_{S_+^n}(\boldsymbol{X})=\{\boldsymbol{H}\in S^n:\boldsymbol{E}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{E}\succeq \boldsymbol{0}\}.$$

证明　由于 $\lambda_{\max}(\boldsymbol{Y})$ 是关于 $\boldsymbol{Y}\in S^n$ 的 Lipschitz 连续凸函数，从而 $\lambda_{\min}(\boldsymbol{Y})=-\lambda_{\max}(-\boldsymbol{Y})$ 是关于 $\boldsymbol{Y}\in S^n$ 的连续凹函数. 利用 Danskin 定理（文献［22］定理 2.7.1）可知

$$\lambda'_{\min}(\boldsymbol{X};\boldsymbol{H})=\min_{\|\boldsymbol{a}\|=1}(\boldsymbol{E}\boldsymbol{a})^{\mathrm{T}}\boldsymbol{H}(\boldsymbol{E}\boldsymbol{a})=\lambda_{\min}(\boldsymbol{E}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{E}).$$

结合命题 1.2.17 有

$$\begin{aligned}T_{S_+^n}(\boldsymbol{X})&=\{\boldsymbol{H}\in S^n:\lambda'_{\min}(\boldsymbol{X};\boldsymbol{H})\geqslant 0\}\\&=\{\boldsymbol{H}\in S^n:\lambda_{\min}(\boldsymbol{E}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{E})\geqslant 0\}\\&=\{\boldsymbol{H}\in S^n:\boldsymbol{E}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{E}\succeq\boldsymbol{0}\}.\end{aligned}$$

■