优化方法导论

梁礼明 主编

北京理工大学出版社
BEIJING INSTITUTE OF TECHNOLOGY PRESS

优化方法导论

主　编　梁礼明
副主编　杨国亮

北京理工大学出版社
BEIJING INSTITUTE OF TECHNOLOGY PRESS

内容简介

本书是一本关于优化方法的入门教材。全书共6章，主要内容包括优化问题的数学模型与数学基础、线性和非线性规划、有约束和无约束优化方法以及现代优化方法等。

本书取材得当、难易适度，注重思想性，推导过程伴以几何解释，同时对一些优化方法或算法给出了定性评价，便于读者理解与掌握优化方法的基本理论和基本算法。

本书可作为高等学校工科相关专业硕士研究生的教材，也可作为工程技术领域的科研人员的参考书。

本书的编写得到江西理工大学研究生教材建设项目资助。

图书在版编目(CIP)数据

优化方法导论/梁礼明主编. —北京：北京理工大学出版社，2017.9
ISBN 978-7-5682-4877-8

Ⅰ.①优…　Ⅱ.①梁…　Ⅲ.①最优化算法　Ⅳ.①O242.23

中国版本图书馆CIP数据核字(2017)第238143号

出版发行／北京理工大学出版社有限责任公司
社　　址／北京市海淀区中关村南大街5号
邮　　编／100081
电　　话／(010)68914775(总编室)
　　　　　(010)82562903(教材售后服务热线)
　　　　　(010)68948351(其他图书服务热线)
网　　址／http://www.bitpress.com.cn
经　　销／全国各地新华书店
印　　刷／三河市天利华印刷装订有限公司
开　　本／787毫米×1092毫米　1/16
印　　张／13
字　　数／314千字
版　　次／2017年9月第1版　2017年9月第1次印刷
定　　价／42.00元

责任编辑／李志敏
文案编辑／赵　轩
责任校对／周瑞红
责任印制／施胜娟

前　言

“优化理论与方法”是工科研究生学习的一门重要的基础课。编者结合教学实践编写了本书，在编写过程中主要体现以下特点：

（1）书名。关于本书的命名，主要出于两点考虑：一是现有的优化理论与方法是基于一定条件或意义下的“最优”，而非普适性，故名“优化方法”；二是突出“导论”，本书专门辟出部分章节介绍相关的基础知识，如线性代数、微积分、空间集合与多元函数等，适合基础知识程度不一的研究生更好地学习。

（2）重在传授优化思想。本书遵循认识规律，注意阐述优化理论的系统性和理论溯源，培养工程技术人员优化思想，使其具备创新性优化处理工程实际问题的基本素养。

（3）注重抽象理论形象化。本书力求用简明的语言尤其是几何语言，总括性地描述抽象的优化理论与方法的基本思想，深入浅出，直观通俗，以便于读者从全局的高度掌握优化理论与方法的本质特征。

（4）给出优化方法的定性评价。在全面介绍优化思想、原理方法、推导证明、基本步骤和算法收敛性的基础上进一步进行分析与讨论，对优化方法或算法给出定性评价，有利于读者在实际问题中选用。

（5）吸收了部分现代优化理论的研究成果。在介绍相关基础知识、传统优化方法的基础上，适度地介绍部分优化理论的新成果和发展趋势。

《优化方法导论》是一本既注重实际应用又有理论深度的优化方法教材。优化所涉及的内容较多，大致分为经典方法和现代方法两部分，经典方法主要包含线性规划、非线性规划、无约束规划和有约束规划等；现代方法主要包括模拟退火算法、神经网络算法、遗传算法、支持向量机、深度学习等。

本书可作为工科研究生教材，也可作为科研人员以及从事实际应用的工程技术人员的参考书。

编者由于水平有限，尽管在编写过程中努力做到术语准确、表达全面和语句通顺，但书中难免存在不足之处，敬请广大专家和读者批评指正。

编　者

2017 年 6 月

本书主要符号

$\mathbb{R}^n$	n 维实向量空间
$\boldsymbol{x},\boldsymbol{x}^k(k=1,2,\cdots,n)$ $\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}$; $\boldsymbol{x}^k=[x_1^k,x_2^k,\cdots,x_n^k]^{\mathrm{T}}$。	n 维实列向量
$\boldsymbol{x}^{\mathrm{T}}$	向量 $\boldsymbol{x}$ 的转置
$\boldsymbol{x}>\boldsymbol{0}$	向量 $\boldsymbol{x}$ 的各分量均为正数
$\boldsymbol{x}\geqslant\boldsymbol{0}$	向量 $\boldsymbol{x}$ 的各分量均为非负数
$\boldsymbol{x}<\boldsymbol{0}$	向量 $\boldsymbol{x}$ 的各分量均为负数
$\boldsymbol{x}\leqslant\boldsymbol{0}$	向量 $\boldsymbol{x}$ 的各分量均为非正数
$\boldsymbol{x}<\boldsymbol{y}$	向量 $\boldsymbol{x}$ 的各分量均小于向量 $\boldsymbol{y}$ 的各分量
$\boldsymbol{x}\in D$	向量 $\boldsymbol{x}$ 属于向量集合 D
$\boldsymbol{x}\notin D$	向量 $\boldsymbol{x}$ 不属于向量集合 D
$\boldsymbol{0}$	零向量
$\boldsymbol{Q}>0$	矩阵 $\boldsymbol{Q}$ 正定
$\boldsymbol{Q}\geqslant0$	矩阵 $\boldsymbol{Q}$ 半正定
$\boldsymbol{Q}<0$	矩阵 $\boldsymbol{Q}$ 负定
$\boldsymbol{Q}\leqslant0$	矩阵 $\boldsymbol{Q}$ 半负定
$\forall\boldsymbol{x}\in\mathbb{R}^n$	$\boldsymbol{x}$ 属于 $\mathbb{R}^n$ 的任意 n 维实向量
$\Vert\boldsymbol{x}\Vert$	向量 $\boldsymbol{x}$ 的欧氏范数
s. t.	"subject to"的缩写,即"满足约束"
$\nabla f(\boldsymbol{x})$	$f(\boldsymbol{x})$在点 $\boldsymbol{x}$ 处的梯度
$\nabla^2 f(\boldsymbol{x})$	$f(\boldsymbol{x})$在点 $\boldsymbol{x}$ 处的 Hesse 矩阵
α	搜索步长
$\boldsymbol{d}$	搜索方向

目　录

第 1 章　优化问题的数学模型与数学基础

“优化”一词来自英文 Optimization,其本意是寻优的过程,即指从一切可能的备选方案中选择一个“最好的”方案以达到最优目标。备选方案的好坏程度,可以根据实际问题对象的需求采用目标函数或性能指标来度量。

优化理论与方法是一个重要的数学分支。优化思想源远流长,最早可追溯到古希腊的欧几里得(Euclid,公元前 300 年左右)。他指出:在周长相同的一切矩形中,以正方形的面积为最大;17 世纪时,牛顿(Newton,1643—1727)和莱布尼茨(Leibniz,1646—1716)所创立的微积分中已包含优化思想,如极值问题;后来伯努利(Bernoulli,1654—1705)、欧拉(Euler,1707—1783)和拉格朗日(Lagrange,1736—1813)等对变分法的创立与发展做出了各自的贡献;柯西(Cauchy,1789—1857)则最早提出求解多元函数最优值的最速下降法。1939 年,苏联数学家康托诺维奇(Cantolovch)首次提出了下料问题和运输问题这两种线性规划的求解方法;1947 年,美国数学家丹茨格(Dantzig)在研究美国空军资源的优化配资时,提出了求解线性规划的通用解法——单纯形法,促使线性规划理论趋于成熟;1950 年至 1965 年,匈牙利数学家库恩(Kuhn)和塔克(Tucker)等建立了线性规划的对偶理论,为求解鞍点问题提供了数学工具,他们建立的约束极值问题的最优性条件被称为 K-T 条件(也称 KKT 条件),为求解非线性规划问题奠定了基础。随着科学技术的迅猛发展,特别是计算机科学与技术的发展,使得优化问题研究不仅成为一种迫切需要,而且有了求解的有力工具,促使优化理论与方法得以迅速发展,并不断完善,逐步成为一门系统的学科。如今,优化方法已成为工程技术人员必须具备的研究工具。

本章简要地介绍了优化方法的学科发展和优化问题的数学模型,重点阐述了求解优化问题的相关基础知识,包括线性代数、微积分、空间和几何、多元函数等,为后续优化理论与方法的分析和讨论做铺垫。

1.1　最优化问题的数学模型

1.1.1　最优化问题的例子

例 1 - 1(运输问题)　设某地有 m 个水泥厂 $A_1,A_2,\cdots,A_m$,年产量分别为 $a_1,a_2,\cdots,a_m$(吨);有 n 个销售地 $B_1,B_2,\cdots,B_n$,每个销售地的需求量为 $b_1,b_2,\cdots,b_n$(吨)。假定产销是平衡的,即 $\sum_{i=1}^{m}a_i=\sum_{j=1}^{n}b_j$,由 A_i 地到 B_j 地的运费为每吨 $c_{ij}(i=1,2,\cdots,m;j=1,2,\cdots,n)$元。试设计一个调运方案,既可保证满足需求量,又可使运费最省。

解 设从 A_i 地到 B_j 地运送水泥量为 x_{ij}吨,由题意可画出如图 1-1 所示的运输调运费用图,则总运费 S 为

$$S = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}$$

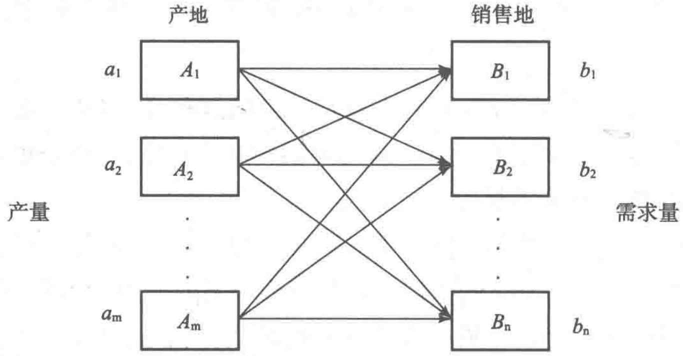

图 1-1 运输调运费示意图

此时,x_{ij}应满足

$$\text{s. t.} \begin{cases} \sum\limits_{j=1}^{n} x_{ij} = a_i (i = 1,2,\cdots,m) \\ \sum\limits_{i=1}^{m} x_{ij} = b_j (j = 1,2,\cdots,n) \\ x_{ij} \geqslant 0 (i = 1,2,\cdots,m; j = 1,2,\cdots,n) \end{cases}$$

因此,该问题的数学模型可表示为

$$\min S = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}$$

$$\text{s. t.} \begin{cases} \sum\limits_{j=1}^{n} x_{ij} = a_i (i = 1,2,\cdots,m) \\ \sum\limits_{i=1}^{m} x_{ij} = b_j (j = 1,2,\cdots,n) \\ x_{ij} \geqslant 0 (i = 1,2,\cdots,m; j = 1,2,\cdots,n) \end{cases}$$

注意:平衡条件 $\sum\limits_{i=1}^{m} a_i = \sum\limits_{j=1}^{n} b_j$ 作为已知条件并不出现在约束条件中。

例 1-2(生产计划问题) 某工厂有 m 种资源 $B_1, B_2, \cdots, B_m$,数量分别为 $b_1, b_2, \cdots, b_m$,利用这些资源生产 n 种产品 $A_1, A_2, \cdots, A_n$。每生产一个单位的 A_j 产品需要消耗资源 B_i 的量为 a_{ij}。根据合同规定,产品 A_j 的量不少于 d_j,且每生产一个单位 A_j 产品的单价为 c_j。问如何安排生产计划,才能既完成合同,又能使该厂总收入最多?

解 假设产品 A_j的计划产量是 x_j,由题意可画出如图 1-2 所示的生产与消耗关系图,则其数学模型为

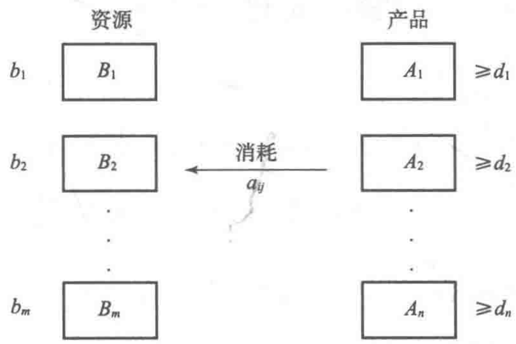

图 1-2　生产与消耗关系图

$$\max S = \sum_{j=1}^{n} c_j x_j$$

$$\text{s. t.} \begin{cases} \sum_{j=1}^{n} a_{ij} x_j \leqslant b_i (i = 1,2,\cdots,m) \\ x_j \geqslant d_j (j = 1,2,\cdots,n) \end{cases}$$

例 1-3(指派问题)　设有四项任务 B_1、B_2、B_3、B_4，需派四个人 A_1、A_2、A_3、A_4 去完成。每个人都可以承担四项任务中的任何一项，但所消耗的资金不同。设 A_i 完成 B_j 所需资金为 c_{ij}。请问如何分配任务，使总支出最少?

解　设变量

$$x_{ij} = \begin{cases} 1 & (\text{指派 } A_i \text{ 完成 } B_j \text{ 任务}) \\ 0 & (\text{不指派 } A_i \text{ 完成 } B_j \text{ 任务}) \end{cases}$$

则总支出可表示为

$$S = \sum_{i=1}^{4} \sum_{j=1}^{4} c_{ij} x_{ij}$$

其数学模型为

$$\min S = \sum_{i=1}^{4} \sum_{j=1}^{4} c_{ij} x_{ij}$$

$$\text{s. t.} \begin{cases} \sum_{j=1}^{4} x_{ij} = 1 \quad (i = 1,2,3,4) \\ \sum_{i=1}^{4} x_{ij} = 1 \quad (j = 1,2,3,4) \\ x_{ij} \in \{0,1\} \quad (i,j = 1,2,3,4) \end{cases}$$

例 1-4(数据拟合问题)　假设两个统计量 x 和 y 之间存在函数关系，但其解析表达式未知。对于给定的一系列成对数据 $\{(x_1,y_1),(x_2,y_2),\cdots,(x_m,y_m)\}$，根据给定的数据导出函数 $y=f(x)$ 的一个简单而近似的解析表达式。

解　曲线拟合问题是在实际数据处理或统计资料分析中经常碰到的情形，求解这种问题常用的方法是最小二乘法(又称最小平方法)。最小二乘法是从误差拟合角度对回归模型进

行参数估计或系统辨识，并在科学实验、分析计算、预测预报和工程技术等诸多领域中得到极其广泛的应用。

最小二乘法的数学思想如下：

对于给定的一组数据$\{(x_i,y_i),(i=1,2,\cdots,m)\}$，若拟合曲线模型为$y=f(x)$，则第$i$组数据误差距离为$f(x_i)-y_i$，因此所有点的误差平方和为$\sum\limits_{i=1}^{m}[f(x_i)-y_i]^2$，进而求出$\sum\limits_{i=1}^{m}[f(x_i)-y_i]^2$的最小值对应的参数，从而找到拟合曲线$y=f(x)$。

因此，数据拟合问题的数学模型为

$$\min\sum_{i=1}^{m}[f(x_i)-y_i]^2$$

其中，x_i、$y_i(i=1,2,\cdots,m)$及$f(x_i)$为已知的某一函数类型。

1.1.2 数学模型

从1.1.1节中四个不同类型的例子可以看出，凡是追求最优目标的数学问题都属于最优化问题。

1.1.2.1 数学模型的一般形式

最优化问题的数学模型的一般形式为

$$\min f(\boldsymbol{x}) \tag{1-1}$$

$$\text{s.t.}\begin{cases}g_i(\boldsymbol{x})\geqslant 0 & (i=1,2,\cdots,m \quad \text{或} \quad i\in I=\{1,2,\cdots,m\})\\ h_j(\boldsymbol{x})=0 & (j=1,2,\cdots,l \quad \text{或} \quad j\in J=\{1,2,\cdots,l\})\end{cases}$$

其中，$\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}\in\mathbb{R}^n$称为决策变量或设计向量，$n$为该问题的维数；$f:\mathbb{R}^n\to\mathbb{R}$称为目标函数或者评价函数；$g_i(\boldsymbol{x})\geqslant 0(i\in I)$，$h_j(\boldsymbol{x})=0(j\in J)$称为约束函数或约束条件；$g_i:\mathbb{R}^n\to\mathbb{R}$与$h_j:\mathbb{R}^n\to\mathbb{R}$分别称为不等式约束和等式约束。

令$D=\{\boldsymbol{x}\mid g_i(\boldsymbol{x})\geqslant 0,i\in I;\quad h_j(\boldsymbol{x})=0,j\in J\}$，称$D$为最优化问题(1-1)的可行域，即满足所有约束条件的点所构成的集合。D中的点，即满足约束条件的点称为可行点或可行解。

综上可知，最优化问题的数学模型包括三个要素：目标函数、决策变量和约束条件。

1.1.2.2 优化问题的分类

对于优化问题常采用以下方法进行分类。

1. 根据优化问题约束类型的不同进行分类

1)无约束问题

若一个最优化问题的可行域是整个空间，即求$\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}$使目标函数$f(\boldsymbol{x})=f(x_1,x_2,\cdots,x_n)$达到最小值，记为$\min f(\boldsymbol{x})$。

2)约束问题

根据约束函数的类型，约束问题可进一步分为以下几类。

(1)等式约束问题：求$\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}$在满足$l$个等式约束条件$h_j(\boldsymbol{x})=0,j\in J$的情

况下,使目标函数 $f(\boldsymbol{x})=f(x_1,x_2,\cdots,x_n)$ 达到最小值,记为

$$\min f(\boldsymbol{x})$$
$$\text{s.t.}\quad h_j(\boldsymbol{x})=0\quad (j\in J=\{1,2,\cdots,l\})$$

(2)不等式约束问题:求 $\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}$ 在满足 m 个不等式约束条件 $g_i(\boldsymbol{x})\geqslant 0,i\in I$ 的情况下,使目标函数 $f(\boldsymbol{x})=f(x_1,x_2,\cdots,x_n)$ 达到最小值,记为

$$\min f(\boldsymbol{x})$$
$$\text{s.t.}\quad g_i(\boldsymbol{x})\geqslant 0\quad (i\in I=\{1,2,\cdots,m\})$$

(3)一般约束问题(或称混合约束问题):求 $\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}$ 在满足 m 个不等式约束条件 $g_i(\boldsymbol{x})\geqslant 0,i\in I$ 以及 l 个等式约束条件 $h_j(\boldsymbol{x})=0,j\in J$ 的情况下,使目标函数 $f(\boldsymbol{x})=f(x_1,x_2,\cdots,x_n)$ 达到最小值,记为

$$\min f(\boldsymbol{x})$$
$$\text{s.t.}\begin{cases} g_i(\boldsymbol{x})\geqslant 0 & (i\in I=\{1,2,\cdots,m\})\\ h_j(\boldsymbol{x})=0 & (j\in J=\{1,2,\cdots,l\})\end{cases}$$

以上给出的是标准的优化问题。实际上某些不标准的优化问题可以转化为标准的优化问题。

(1)约束条件:

$$g_i(\boldsymbol{x})\leqslant 0\Rightarrow -g_i(\boldsymbol{x})\geqslant 0$$

(2)最大值问题化为最小值问题:

$$\max_{\boldsymbol{x}\in D} f(\boldsymbol{x})\Rightarrow -\min_{\boldsymbol{x}\in D} f(\boldsymbol{x})$$

(3)等式约束化为不等式约束:

$$h_j(\boldsymbol{x})=0\Rightarrow\begin{cases} h_j(\boldsymbol{x})\geqslant 0\\ h_j(\boldsymbol{x})\leqslant 0\end{cases}$$

因此,上述分类可简化为两类问题。

(1)无约束最优化问题:

$$\min f(\boldsymbol{x})$$

(2)约束最优化问题:

$$\min f(\boldsymbol{x})$$
$$\text{s.t.}\ g_i(\boldsymbol{x})\geqslant 0\quad (i\in I=\{1,2,\cdots,m\})$$

2. 根据目标函数及约束函数的类型进行分类

最优化问题也称为规划问题。

(1)当目标函数 $f(\boldsymbol{x})$ 和约束条件 $g_i(\boldsymbol{x})$ 均为线性函数时,称此最优化问题为线性规划;

(2)当目标函数 $f(\boldsymbol{x})$ 和约束条件 $g_i(\boldsymbol{x})$ 不全为线性函数时,称此最优化问题为非线性规划;

(3)当目标函数 $f(\boldsymbol{x})$ 为二次函数,而 $g_i(\boldsymbol{x})$ 全为线性函数时,称此最优化问题为二次规划。

3. 根据变量的类型进行分类

对于最优化问题：

(1)如果变量 $\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}$ 的各分量仅能取整数，则相应的最优化问题称为整数规划；

(2)如果变量 $\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}$ 的部分分量只能取整数，则相应的最优化问题称为混合整数规划；

(3)如果变量 $\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}$ 的各分量只能取 0 和 1，则相应的最优化问题称为 0－1 规划。

4. 其他分类方法

(1)如果某个优化问题只包含一个目标函数，则称为单目标优化问题；如果优化问题的目标函数不止一个，则称该优化问题为多目标规划。

(2)如果决策变量与时间无关，则该问题称为静态最优化问题；如果最优化问题需要根据其特性将决策过程按时间或者空间分为若干个相互联系又相互区别的阶段，在它的每一个阶段都需要作出决策，从而使整个决策过程达到最好的效果，则这样的最优化问题称为动态最优化问题(或称动态规划)。

1.2　相关数学基础

1.2.1　向量与矩阵

n 维向量定义为含有 n 个数的数组，记为

$$\boldsymbol{a}=\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

其中，$a_i(i=1,2,\cdots,n)$表示向量 $\boldsymbol{a}$ 的第 i 个元素。定义$\mathbb{R}$ 为由全体实数组成的集合，则由实数组成的 n 维列向量可表示为$\mathbb{R}^n$，称为 n 维实数向量空间。通常将$\mathbb{R}^n$的元素用大写或小写粗体字母表示(如 $\boldsymbol{x}$、$\boldsymbol{a}$ 等)。

n 维行向量记为

$$\boldsymbol{b}=[b_1,b_2,\cdots,b_n]$$

列向量 $\boldsymbol{a}$ 的转置记为 $\boldsymbol{a}^{\mathrm{T}}$，行向量 $\boldsymbol{b}$ 的转置记为 $\boldsymbol{b}^{\mathrm{T}}$，即

$$\boldsymbol{a}^{\mathrm{T}}=[a_1,a_2,\cdots,a_n],\quad \boldsymbol{b}^{\mathrm{T}}=\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

相应地,可以记为 $\boldsymbol{a}=[a_1,a_2,\cdots,a_n]^{\mathrm{T}}$。在本书中,如果不进行特别说明,只要提到向量,均为列向量。

给定向量 $\boldsymbol{a}=[a_1,a_2,\cdots,a_n]^{\mathrm{T}}$ 和 $\boldsymbol{b}=[b_1,b_2,\cdots,b_n]^{\mathrm{T}}$,如果 $a_i=b_i(i=1,2,\cdots,n)$,那么这两个向量相等。

向量 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 的和记为 $\boldsymbol{a}+\boldsymbol{b}$,计算方式为

$$\boldsymbol{a}+\boldsymbol{b}=[a_1+b_1,a_2+b_2,\cdots,a_n+b_n]^{\mathrm{T}}$$

向量的相加运算具有如下性质。

(1)交换性:$\boldsymbol{a}+\boldsymbol{b}=\boldsymbol{b}+\boldsymbol{a}$

(2)结合性:$(\boldsymbol{a}+\boldsymbol{b})+\boldsymbol{c}=\boldsymbol{a}+(\boldsymbol{b}+\boldsymbol{c})$

(3)存在零向量 $\boldsymbol{0}=[0,0,\cdots,0]^{\mathrm{T}}$,使得

$$\boldsymbol{a}+\boldsymbol{0}=\boldsymbol{0}+\boldsymbol{a}=\boldsymbol{a}$$

向量$[a_1-b_1,a_2-b_2,\cdots,a_n-b_n]^{\mathrm{T}}$ 称为 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 之间的差,记为 $\boldsymbol{a}-\boldsymbol{b}$。向量 $\boldsymbol{0}-\boldsymbol{b}$ 记为 $-\boldsymbol{b}$,有如下公式成立:

$$\boldsymbol{b}+(\boldsymbol{a}-\boldsymbol{b})=\boldsymbol{a}$$
$$-(-\boldsymbol{b})=\boldsymbol{b}$$
$$-(\boldsymbol{a}-\boldsymbol{b})=\boldsymbol{b}-\boldsymbol{a}$$

向量 $\boldsymbol{b}-\boldsymbol{a}$ 是向量方程 $\boldsymbol{a}+\boldsymbol{x}=\boldsymbol{b}$ 的唯一解。假定 $\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}$ 是

$$\boldsymbol{a}+\boldsymbol{x}=\boldsymbol{b}$$

的解,则有

$$\begin{gathered}a_1+x_1=b_1\\a_2+x_2=b_2\\\vdots\\a_n+x_n=b_n\end{gathered}$$

从而有

$$\boldsymbol{x}=\boldsymbol{b}-\boldsymbol{a}$$

向量 $\boldsymbol{a}\in\mathbb{R}^n$ 与标量 $\alpha\in\mathbb{R}$ 的乘积定义为

$$\alpha\boldsymbol{a}=[\alpha a_1,\alpha a_2,\cdots,\alpha a_n]^{\mathrm{T}}$$

该运算具有如下性质:

(1)分配性:对于任意实数 α 和 β,有

$$\alpha(\boldsymbol{a}+\boldsymbol{b})=\alpha\boldsymbol{a}+\alpha\boldsymbol{b}$$
$$(\alpha+\beta)\boldsymbol{a}=\alpha\boldsymbol{a}+\beta\boldsymbol{a}$$

(2)结合性:　$\alpha(\beta\boldsymbol{a})=(\alpha\beta)\boldsymbol{a}$

(3)标量 1 满足:　$1\boldsymbol{a}=\boldsymbol{a}$

(4)任意标量 α 满足:　$\alpha\boldsymbol{0}=\boldsymbol{0}$

(5)标量 0 满足:　$0\boldsymbol{a}=\boldsymbol{0}$

(6)标量 -1 满足:　$(-1)\boldsymbol{a}=-\boldsymbol{a}$

注意：当且仅当 $\alpha=0$ 或 $\boldsymbol{a}=\mathbf{0}$ 时，$\alpha\boldsymbol{a}=0$。可以看出 $\alpha\boldsymbol{a}=0$ 等价于

$$\alpha a_1+\alpha a_2+\cdots+\alpha a_n=0$$

如果 $\boldsymbol{a}\neq\mathbf{0}$，那么至少其中一个元素 $a_i\neq 0$，对于 $\alpha\boldsymbol{a}=\mathbf{0}$，则必有 $\alpha=0$。类似地，$\alpha\neq 0$ 时，对于 $\alpha\boldsymbol{a}=\mathbf{0}$，必有 $\boldsymbol{a}=\mathbf{0}$。

如果方程

$$\alpha_1\boldsymbol{a}_1+\alpha_2\boldsymbol{a}_2+\cdots+\alpha_k\boldsymbol{a}_k=\mathbf{0}$$

中所有的系数 $\alpha_i(i=1,2,\cdots,k)$ 都等于零，那么称向量集或向量组 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_k\}$ 是线性无关的。如果向量组 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_k\}$ 不是线性无关的，那么称其为线性相关的。

给定向量 $\boldsymbol{a}$，如果存在标量 $\alpha_1,\alpha_2,\cdots,\alpha_k$，使得

$$\boldsymbol{a}=\alpha_1\boldsymbol{a}_1+\alpha_2\boldsymbol{a}_2+\cdots+\alpha_k\boldsymbol{a}_k$$

那么称向量 $\boldsymbol{a}$ 为 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_k$ 的线性组合。

命题 1-1 向量集 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_k\}$ 是线性相关的，集合中有且仅有一个向量可以表示为其他向量的线性组合。 □

证明 (1)必要性。如果 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_k\}$ 是线性相关的，那么有

$$\alpha_1\boldsymbol{a}_1+\alpha_2\boldsymbol{a}_2+\cdots+\alpha_k\boldsymbol{a}_k=\mathbf{0}$$

其中至少存在一个标量 $\alpha_i\neq 0$，从而有

$$\boldsymbol{a}_i=-\frac{\alpha_1}{\alpha_i}\boldsymbol{a}_1-\frac{\alpha_2}{\alpha_i}\boldsymbol{a}_2-\cdots-\frac{\alpha_k}{\alpha_i}\boldsymbol{a}_k$$

(2)充分性。假定向量 $\boldsymbol{a}_1$ 可以被表示为其他向量的线性组合，即

$$\boldsymbol{a}_1=\alpha_2\boldsymbol{a}_2+\alpha_3\boldsymbol{a}_3+\cdots+\alpha_k\boldsymbol{a}_k$$

那么有

$$(-1)\boldsymbol{a}_1+\alpha_2\boldsymbol{a}_2+\cdots+\alpha_k\boldsymbol{a}_k=\mathbf{0}$$

由于第 1 个标量非零，故向量集 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_k\}$ 是线性相关的。 ■

令 v 表示 $\mathbb{R}^n$ 的一个子集，如果 v 在向量加和运算及标量乘积运算下是封闭的，那么称 v 是 $\mathbb{R}^n$ 的子空间。也就是说，如果 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 是 v 中的向量，那么 $\boldsymbol{a}+\boldsymbol{b}$ 和 $\alpha\boldsymbol{a}$(α 是任意标量)也是 v 中的向量。

每个子空间都包含向量 $\mathbf{0}$，这是因为如果 $\boldsymbol{a}$ 是子空间中的一个元素，那么有 $(-1)\boldsymbol{a}=-\boldsymbol{a}$，因此 $\boldsymbol{a}-\boldsymbol{a}=\mathbf{0}$ 也属于该子空间。

假定 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_k$ 是 $\mathbb{R}^n$ 中的任意向量，它们所有线性组合的集合称为 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_k$ 张成的子空间，记为

$$\operatorname{span}[\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_k]=\left\{\sum_{i=1}^{k}\alpha_i\boldsymbol{a}_i:\alpha_1,\alpha_2,\cdots,\alpha_k\in\mathbb{R}\right\}$$

对于向量 $\boldsymbol{a}$，子空间 $\operatorname{span}[\boldsymbol{a}]$ 由向量 $\alpha\boldsymbol{a}$ 组成，α 为任意实数($\alpha\in\mathbb{R}$)。同样地，如果 $\boldsymbol{a}$ 可表示为 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_k$ 的线性组合，则有

$$\operatorname{span}[\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_k,\boldsymbol{a}]=\operatorname{span}[\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_k]$$

因此，任意向量集合都能够张成一个子空间。

给定子空间 v，如果存在线性无关的向量集合 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_k\}\subset v$ 使得

$$v=\text{span}[\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_k]$$

那么称 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_k\}$ 是子空间 v 的一组基。子空间 v 中的所有基都包含相同数量的向量，这一数量称为 v 的维数，记为 $\dim v$。

命题 1-2　如果 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_k\}$ 是 v 的一组基，那么 v 中的任意向量 $\boldsymbol{a}$ 可以唯一地表示为

$$\boldsymbol{a}=\alpha_1\boldsymbol{a}_1+\alpha_2\boldsymbol{a}_2+\cdots+\alpha_k\boldsymbol{a}_k$$

其中 $\alpha_i\in\mathbb{R}\ (i=1,2,\cdots,k)$。 □

证明　假定 $\boldsymbol{a}$ 可以表示为以下两种形式：

$$\boldsymbol{a}=\alpha_1\boldsymbol{a}_1+\alpha_2\boldsymbol{a}_2+\cdots+\alpha_k\boldsymbol{a}_k$$

或

$$\boldsymbol{a}=\beta_1\boldsymbol{a}_1+\beta_2\boldsymbol{a}_2+\cdots+\beta_k\boldsymbol{a}_k$$

只需证明 $\alpha_i=\beta_i(i=1,2,\cdots,k)$ 即可。

由于

$$\alpha_1\boldsymbol{a}_1+\alpha_2\boldsymbol{a}_2+\cdots+\alpha_k\boldsymbol{a}_k=\beta_1\boldsymbol{a}_1+\beta_2\boldsymbol{a}_2+\cdots+\beta_k\boldsymbol{a}_k$$

即

$$(\alpha_1-\beta_1)\boldsymbol{a}_1+(\alpha_2-\beta_2)\boldsymbol{a}_2+\cdots+(\alpha_k-\beta_k)\boldsymbol{a}_k=\boldsymbol{0}$$

因为集合 $\{\boldsymbol{a}_i,i=1,2,\cdots,k\}$ 是线性无关的，故有

$$\alpha_1-\beta_1=\alpha_2-\beta_2=\cdots=\alpha_k-\beta_k=0$$

从而得到 $\alpha_i=\beta_i(i=1,2,\cdots,k)$。 ■

给定 v 的一组基 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_k\}$ 和向量 $\boldsymbol{a}\in v$，如果

$$\boldsymbol{a}=\alpha_1\boldsymbol{a}_1+\alpha_2\boldsymbol{a}_2+\cdots+\alpha_k\boldsymbol{a}_k$$

那么系数 $\alpha_i(i=1,2,\cdots,k)$ 称为 $\boldsymbol{a}$ 对应于基 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_k\}$ 的坐标。

$\mathbb{R}^n$ 的标准基为

$$\boldsymbol{e}_1=\begin{bmatrix}1\\0\\0\\\vdots\\0\\0\end{bmatrix},\boldsymbol{e}_2=\begin{bmatrix}0\\1\\0\\\vdots\\0\\0\end{bmatrix},\cdots,\boldsymbol{e}_n=\begin{bmatrix}0\\0\\0\\\vdots\\0\\1\end{bmatrix}$$

在标准基下，向量 $\boldsymbol{x}$ 可表示为

$$\boldsymbol{x}=\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=x_1\boldsymbol{e}_1+x_2\boldsymbol{e}_2+\cdots+x_n\boldsymbol{e}_n$$

这就是称其为“标准基”的缘由。

可按照类似的方式定义复向量空间。令 $\mathbb{C}$ 表示复数集合，$\mathbb{C}^n$ 表示 n 维复数向量。可以看出，集合 $\mathbb{C}^n$ 具有与 $\mathbb{R}^n$ 类似的属性，其中标量也可以用复数表示。

矩阵指的是行列数组，常用大写粗体字母表示（如 $\boldsymbol{A}$）。m 行 n 列的矩阵称为 $m\times n$ 矩阵，

记为

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

位于矩阵第 i 行第 j 列的实数 a_{ij} 称为矩阵 $\boldsymbol{A}$ 的第 (i,j) 个元素。如果认为 $\boldsymbol{A}$ 是由 n 个列向量组成的，那么它的每列都是 $\mathbb{R}^m$ 空间的一个列向量。类似地，如果认为 $\boldsymbol{A}$ 是由 m 个行向量组成的，那么它的每行都是一个 n 维的行向量。

考虑 $m\times n$ 矩阵 $\boldsymbol{A}$，其转置 $\boldsymbol{A}^{\mathrm{T}}$ 则是一个 $n\times m$ 矩阵：

$$A^{\mathrm{T}}=\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$$

即 $\boldsymbol{A}$ 的列是 $\boldsymbol{A}^{\mathrm{T}}$ 的行，反之亦然。

需要注意的是，行向量与 $1\times n$ 的矩阵记号之间存在少许差别：用逗号来分割行向量中的不同元素，而在矩阵中通常不使用逗号。但是，在同一行中使用逗号作为分割符号，区分效果比较明显。因此，在位于多个矩阵同一行时，有时候会使用逗号进行分割。

1.2.2　矩阵的秩

考虑 $m\times n$ 矩阵

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$\boldsymbol{A}$ 的第 k 列用 $\boldsymbol{a}_k$ 表示：

$$\boldsymbol{a}_k=\begin{bmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{bmatrix}$$

矩阵 $\boldsymbol{A}$ 中线性无关列的最大数目称为 $\boldsymbol{A}$ 的秩，记为 rank $\boldsymbol{A}$。可以看出，rank $\boldsymbol{A}$ 是 span$[\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n]$ 的维数。

命题 1－3　在如下运算中，矩阵 $\boldsymbol{A}$ 的秩保持不变：

(1)矩阵 $\boldsymbol{A}$ 的某个(些)列乘以非零标量；

(2)矩阵内部交换列次序；

(3)矩阵中加入一列，该列为其他列的线性组合。　□