临时前言

笔者在多年的学习和交流过程中发现，流变学有一些基本问题是许多教科书没有强调的。对这些问题的生疏，是部分流变学同行产生各种误解的根源。以下是笔者觉得需要先在前言中抛出的两种说法。

一、在流变仪上测得的“应力”、“应变”，跟流变学本构关系里的“应力”、“应变”是两码事。

本构关系里的“应力”、“应变”都是张量场函数。“场函数”是指自变量是 3 维空间位置和时间的函数；“张量”是指采用具有几何意义的矢量和线性变换来表示的物理量，这些物理量要遵守不同坐标系选取之间的坐标变换法则。因此，具有本构关系意义的“应力–应变关系”是函数与函数之间的关系（泛函关系）。

流变学本构关系呈现出这样的数学形式，并非小部分人的特殊趣味，而是物理客观性的自然结果，即“材料的流变学性质不应该依赖参考系和坐标系的选取方式，亦不应该依赖被研究体系的规模和形变方式”。这一客观性要求并非一句哲学口号；它对定量表征应力、应变的物理量的构建方式，乃至本构关系的可以取的形式，都提出了非常具体的要求或限制，因此是流变学理论所不可缺少的内容。

而流变仪上测得的“应力”、“应变”则是限于某种理想流场（例如简单剪切、扭转、单轴拉伸、3 点弯曲等）假定之下才能比较完整地表征材料的应力和应变状态的标量值变量，它们往往不依赖空间位置。用这样的“应力”和“应变”表示的关系是变量与变量之间的关系（函数关系）；它们虽然简单，但仅限于在所假定的流场下才适用。这样的“应力–应变关系”不应被称为“本构关系”。

二、流变学研究分为软物质物理研究和非牛顿流体力学研究两类。

流变学的发展和高分子科学的发展不仅在时间上几乎重合，在研究内容上也高度重合。在二十世纪，高分子科学是一个物理理论和工业技术都同时具有极其丰富的发展的学科，流变学也一样。近几十年，胶体体系或软物质物理的研究也与流变学研究结合得非常紧密。流变学研究的“上下文”（context），可以大致分为各自比较独立两类，一类是软凝聚态物理研究，一类是工业实际问题研究。软凝聚态物理与流变学的结合，主要是从非平衡态统计力学的角度。工业实际问题对流变学理论的需求，主要从非牛顿流体力学的角度。流变学求问于非平衡态统计力学，主要是为材料的本构关系探求微观机理。但流变学要能解决工业实际，并不能停留在本构关系本身，还需要求解不同实际情况的流动问题。这两种研究需的实验条件和实验内容是很不同的。为前一种研究建立的实验条件，未必能做后一种研究。

想要理解上述两种说法，并进行批判得出自己的观点，就需要严格掌握流变学的理论基础。但是这超出了大部分化学类或高分子专业本科的数学基础。这是从流变学诞生开始就面

临的问题*。因此，已有大量作者致力于撰写“化学家看得懂的流变学教材”。但是，若想要成为能够利用流变学知识解决实际问题的专业人士，是无法回避上列两个问题的，而此时我们仍然面临着数学和物理基础缺口的问题。

流变学理论涉及到的数学基础，常常在面向的物理系学生的教材中介绍，是适应物理学人才的培养体系写就的。为了物理系人才往量子物理、凝聚态物理、天体物理……等不同方向发展的可能性，面向物理系学生的数学教材往往形成了远超流变学必需的广度和深度。因此当化学背景人士仅为了流变学的兴趣，去找数学书补习时，往往会发现他要么选择放弃，要么就只能先变成一个数学基础扎实的物理系毕业生。事实上，化学学科有其特有的旨趣和不可替代的角色，成为一名化学家——无论是一名有机合成专家还是一名高分子材料专家——所需要经受的学术和工程训练已经十分丰富和系统。流变学应该是化学家不得不关心的，需要用到最多数学和物理的学科了。仅为流变学裁剪出来的最少必要数学和物理学讲义，也许对任何一个物理系学生的发展是没有帮助的，却是面临流变学问题的化学家所需要的。这样的讲义，当然需要针对化学类专业本科的数学和物理学教育背景去帮读者补习数学和物理学，或至少指出适应化学类专业人士的自学路径。

笔者从攻读研究生时期至今，自己闭门造车地完成了这件事情，没有机会受到批评。写就这本讲义，第一目的就是为了广泛接受批评，并不断修改它。如果它现在还不能对他人有正面的益处，也希望它有朝一日能有。请把批评意见发至邮箱 mswxsun@scut.edu.cn。

基于上述的思想，本讲义的内容有很多个人做法，在此作部分解释。

讲义主要内容分两部分。第一部分是必要的数学基础。第二部分是流变学的物理理论基础。关于数学基础部分的内容深度和广度，我作了如下考虑：

1. 从集合论开始介绍。在物理学中，数学不仅用于表示数量关系，还用于传达**物理概念**†，因此我们才需要采用近世的、基于集合论的语言来铺叙流变学原理。有很多常见的困惑，使用简单的数学语言确实是无法回答的；数学语言太简单本身恰恰就是导致这些困惑的原因。不过，数学语言的抽象化和一般化是无止境的；我们不盲目追求这个。那么，一个起码的数学语言深度，应当要恰好能满足物理思想的准确表达。在这一宗旨下，集合和映射的语言是必须的。近几年工科大学的高等数学教学以为集合论知识已下沉到中学，就普遍不再作正面介绍‡。笔者不完全调查发现也不是所有中学都教了集合论；就算教了，

*I said, “This branch of physics already exists: it is called mechanics of continuous media, or mechanics of continua.”“No, this will not do,” Bingham replied, “Such a designation will frighten away the chemists.” So he consulted the professor of classical languages and arrived at the designation of rheology, taking as the motto of the subject Heraclitus’ $\pi\alpha\nu\tau\alpha\rho\varepsilon\iota$ or “everything flows.”[1]

†A perhaps unorthodox feature of this book is its use of mathematics as a conceptual tool rather than merely as a device to express complex numerical relations.—Coleman, Markovitz and Noll (1966)[2].

‡例如同济大学《高等数学（第七版）》详见其前言。此外，一般工科数学还会修《概率论与数理统计》课程，可能会在

内容也非常浅。因此本讲义从集合论开始介绍。

2. 以化学类工科专业大一学习的微积分和线性代数为基础。具体地，我假定读者已掌握华南理工大学数学系编著的《高等数学（上、下册）》和《线性代数与解析几何》，原则上只介绍上述课本中没有涉及到的数学知识。我在讲义中尽可能多地提示了所介绍的新内容跟上述课本中已经介绍过哪些内容相关（精确到章节和小标题）。这么做的目的是为了让化学类专业的同学具体地衔接数学上的缺口，顺利实现数学语言的近世化。
3. 没有完全建立在张量分析的基础上。张量分析是为了处理在不仅限于 3 维、不仅限于欧几里得空间、更不仅限于直角坐标系下的物理问题的。这在基于非相对论经典力学的流变学当中是不必要的。因此本讲义打算仅介绍一个不涉及度规张量的 3 维欧几里得空间上的曲线坐标系理论。事实上，在这样的语境下，流变学中使用的张量，在数学上都可仅称作线性算符。按照流变学学科的惯例，它们都仍被称作张量，这表示当这些线性算符被用于表示 3 维欧几里得空间上的物理概念时，必须不依赖坐标系的选择（即遵循相应的曲线坐标变换法则）。
4. 直接有助于理解物理概念的数学基础写在正文，仅为了证明数学定理所需的数学基础写在附录。定理的证明过程用浅灰字体。没有习题。这些做法都是因为数学内容在本讲义所扮演的角色不是正式的数学教学，而是展示流变学理论的概念体系。讲义引用到的教材，都是充当正式学习这些数学知识的优秀教材。

概念在被定义时会通过字体的改变来突出。例如，集合（*set*）是具有某种特性的事物的整体。有时，定义是以带编号的方式独立列出的，有时则在正文叙述当中引入。定义是极其重要的。它在文中只出现一次，因此难免要经常反复回顾。有时我不采用带编号式的定义，只是因为需要定义的概念很多。如果每个定义都带编号，定义将会严重打断行文的流畅性，让本来就抽象的内容更难以阅读。但是这么做的代价则是使定义失去了引用链接的便利，故在此敬请读者在阅读时贴上标签，不厌其烦地翻阅和回顾定义。将来笔者将会编制名词索引，彻底解决概念定义的定位问题，并进一步减少定义以编号的方式出现现象。

如同许多不尽人意的教科书那般，我在前言中宣称要做的事，可能恰恰做得特别差，的确是水平和时间有限所致，请见谅！请同学们学习时不要只信一本书，要带着质疑同时看几本书，并让它们相互“对质”，以防被不学无术的作者坑害！下列书籍是在同类主题的书中我特别推荐的，包括了数学、力学和流变学。如能直接看懂这些书，就完全不必使用本讲义（序号同时代表着建议的学习顺序）。

1. P. Halmos (1960), *Naive Set Theory*, D. Van Nostrand Company, Inc.
2. K. Hoffman & R. Kunze (1971), *Linear Algebra, 2nd ed.*, Prentice-Hall Inc.

学习“概率的抽象定义”时再涉及到一些集合论基础。但是化学专业不修该课的也很普遍。

3. R. Williamson, R. Crowell & H. Trotter (1972), *Calculus of Vector Functions, 3rd ed.*, Prentice-Hall Inc.
4. D. Smith (1993), *An Introduction to Continuum Mechanics—after Truesdell and Noll*, Springer
5. A. Murdoch (2012), *Physical Foundations of Continuum Mechanics*, Cambridge University Press
6. W. Schowalter (1978), *Mechanics of Non-Newtonian Fluids*, Pergamon Press
7. R. Tanner (2000), *Engineering Rheology, 2nd ed.*, Oxford University Press
8. 许元泽（1988），*高分子结构流变学*，四川教育出版社
9. 王玉忠、郑长义（1993），*高聚物流变学导论*，四川大学出版社

虽然这个讲义还没写完，这只是临时前言，但也要列出早应表达的感恩。感谢我的博士生导师童真教授在我工作后的继续庇护，为我营造了（一定程度上）不用理会“科研产出”的宽松环境！这样的环境离不开钱，感谢国家的资助！这样的环境也离不开我所在单位的宽容，感谢华南理工大学人事处和材料科学与工程学院的领导和同事！

更多说明和致谢将在讲义完成之后补充。

孙尉翔

2023-12-04

目录

第一部分

数学部分

第一章 集合与映射

1.1 集合

集合是近代数学的基本语言。用集合论重述人的直观经验，是近世数学和基于其构建的理论物理的普遍特点。连续介质力学中的大量概念都是依赖集合和映射来引入的。如果不熟悉集合和映射的语言和符号，就很难读懂后面的内容。

一般资料中常见的对集合的定义方式如下：

定义 1.1. 集合(*set*)是具有某种特性的事物的整体。构成集合的事物或对象称为元素(*element*)。集合还必须满足：

- 无序性：一个集合中，每个元素的地位是相同的，元素之间是无序的
- 互异性：一个集合中，任何两个元素都不相同，即每个元素只出现一次
- 确定性：给定一个集合及一个事物，该事物要么属于要么不属于该集合，不允许模棱两可。

定义1.1的“确定性”，其实是逻辑学的排中律。具体地，若 A 是集合，x 是 A 的一个元素，则记 $x \in A$；若 y 不是 A 的元素，则记为 $y \notin A$。记号“$x \in A$”规定了“A 是集合且 x 是 A 的元素”。

如果只要有 $x \in A$ 就有 $x \in B$，则称 A 是 B 的子集（*subset*)，或称 A 包含于 B、B 包含 A，记为 $A \subset B$。显然，任一集合都是它自己的子集。

如果 $A \subset B$ 且 $B \subset A$，则集合 A 就是（*is identical to*）B，记为 $A = B$*。如果集合 A 与 B 的元素都相同，则 A 与 B 就是同一个集合。一个集合由其所有元素唯一确定。这是集合论的外延公理（*axiom of extension*)。

如果 $A \subset B$ 且 $A \neq B$，则称 A 是 B 的真子集（*proper subset*)，或称 A 真包含于 B、B 真包含 A，记作 $A \subsetneqq B$。

---

*此处等号“$=$”的意义应理解为：等号两边的字母是同一集合的不同代号。若写 $A = B$，那么 A 就是 B。相应地，不等号“$\neq$”两边的字母是不同集合的代号。若写 $A \neq B$，那么 A 不是 B。

没有元素的集合称为空集（*empty set*），记作 $\emptyset$。空集是唯一的。简要证明：若 $\emptyset_1$ 和 $\emptyset_2$ 都是空集且 $\emptyset_1 \neq \emptyset_2$，则由空集的上述定义，要么存在 $x \in \emptyset_1$ 且 $x \notin \emptyset_2$，要么存在 $y \in \emptyset_2$ 且 $y \notin \emptyset_1$；无论哪种情况与 $\emptyset_1$ 和 $\emptyset_2$ 是空集相矛盾。因此要么 $\emptyset_1$ 和 $\emptyset_2$ 有一个不是空集，要么它们是同一个集合，即“空集是唯一的”。

给定一个集合 A，我们可以根据 A 的元素所需要满足的附加要求，构建出 A 的子集。例如，设 A 是所有偶数的集合，附加的要求是“比 1 大、比 9 小”，我们就从 A 中找出了 2、4、6、8 这四个元素组成的集合 B。一般地，我们将此记作：

$$\left\{x \in A | x \text{ 需要满足的条件}\right\}$$

注意，预先给定一个基本集合 A 这一步原则上是不可省略的；“|”后的语句形式是对 A 的元素 x 的规定，而不能是其他意义。这是集合论的分类公理（*axiom of specification*）。

给定两个集合 A 和 B，总存在唯一一个这样的集合 V：只要 $x \in V$，就有 $x \in A$ 且 $x \in B$；反之，若 $x \in A$ 且 $x \in B$，则有 $x \in V$。这样的集合 V 的存在性来自分类公理；V 可表示成：

$$V = \{x \in A | x \in B\}$$

V 的唯一性简证如下：若另有一 $V' = \{x \in A | x \in B\}$ 且 $V' \neq V$，则由外延公理必存在 $x \in V'$ 且 $x \notin V$。若 $x \in V'$，则 $x \in A$ 且 $x \in B$；若 $x \notin V$，则 $x \notin A$ 或 $x \notin B$，这两个推论相互矛盾*。因此 $V' = V$ 或 V' 不存在，即 V 是唯一的。我们把这样的集合 V 称作集合 A 与 B 的交集（*interset*），记作 $V = A \cap B$。

我们经常还要讨论“集合的集合”，从而有“元素的元素”。

设 $\mathscr{C}$ 是集合的集合，且 $\mathscr{C} \neq \emptyset$。设 $A \in \mathscr{C}$，则由分类公理存在以下集合

$$V = \{x \in A | x \in X \Leftrightarrow X \in \mathscr{C}\}$$

其中 $\Leftrightarrow$ 是当且仅当（*if and only if*, *iff*）的意思。这一集合 V 的唯一性可类似上一段那样得证，此略†。我们称 V 是 $\mathscr{C}$ 的元素的交集，记作

$$V = \bigcap_{X \in \mathscr{C}} X$$

集合的交集有如下性质：

1. $A \cap \emptyset = \emptyset$
2. $A \cap B = B \cap A$（交换律）

*由集合定义1.1中的“确定性”。

†下文构建的集合的唯一性，不作说明时，都由外延公理保证。

3. $A \cap (B \cap C) = (A \cap B) \cap C$（结合律）
4. $A \cap A = A$（幂等）
5. $A \subset B \Leftrightarrow A \cap B = A$

如果集合 A 和 B 的交集是空集，即 $A \cap B = \emptyset$，则称 A 与 B 是不相交的（*disjoint*）。比如，我们有时会说，某集合的集合 $\mathscr{C}$ 中的元素两两不相交（*pair-wise disjoint*）。

分类公理只允许我们“收窄”一个给定的集合。以下规定的原则将允许我们从已有集合构建出“更大的”集合。

我们可以把任意两个集合 a、b 组成一对，变成一个新的集合，记作 $\{a, b\}$，并规定这样的集合可以存在。这是集合论的配对公理（*axiom of paring*）。于是有 $a \in \{a, b\}$ 和 $b \in \{a, b\}$。特别地，一个集合 a 可与其自身“成对”，得到“$\{a, a\}$”，但由于集合的元素要满足互异性，故实际所得到的集合应是 $\{a\}$。这种只有一个元素的集合，称为单元素集（*singleton*）。注意理解以下事实：$\emptyset \in \{\emptyset\}$，$\{\emptyset\} \neq \emptyset$。

我们可以让若干个集合形成并集。具体地，我们规定对任意集合的集合 $\mathscr{C}$，总存在一个集合 U，它含有 $\mathscr{C}$ 的至少一个元素的所有元素*。这样规定存在的集合 U，还可能包含不属于 $\mathscr{C}$ 中任一集合的元素。但是我们总是可以再利用分类公理，把 U 收窄为恰好只包含 $\mathscr{C}$ 中的所有元素的所有元素。具体地，给定一个集合的集合 $\mathscr{C}$，由并集公理必存在上述的 U，利用分类公理构建以下集合†

$$\{x \in U | \exists X \, (X \in \mathscr{C} \wedge x \in X)\}$$

使得对每一 $x \in U$，当且仅当 x 属于 $\mathscr{C}$ 的某个元素 X，x 属于上列集合。因此上列集合包括且仅包括 $\mathscr{C}$ 中的所有元素的所有元素。我们称上列集合为 $\mathscr{C}$ 的所有元素的并集（*union*），记作

$$\bigcup_{X \in \mathscr{C}} X$$

这是集合论的并集公理（*axiom of unions*）。特别地，若 $\mathscr{C} = \emptyset$，则 $\bigcup_{X \in \mathscr{C}} X = \emptyset$，简单证明：由并集的定义，若存在 $a \in \bigcup_{X \in \mathscr{C}} X$，则至少存在一个 $X \in \mathscr{C} = \emptyset$ 满足 $x \in X$，但是显然不存在属于空集的元素 X，因此不存在所述的 a，即 $\bigcup_{X \in \mathscr{C}} X$ 是空集。

如果 $\mathscr{C}$ 是由两个集合 A 和 B 配对而成，即 $\mathscr{C} = \{A, B\}$，则 $\mathscr{C}$ 的元素的并集常以中缀的记法写成：

$$\bigcup_{X \in \{A, B\}} X = A \cup B$$

集合的并集有如下性质：

*$\mathscr{C}$ 的元素是集合。

†符号 $\exists$ 表示“存在一个”或“给定一个”的意思。符号 $\wedge$ 表示“且”的意思。

1. $A \cup \emptyset = A$
2. $A \cup B = B \cup A$（交换律）
3. $A \cup (B \cup C) = (A \cup B) \cup C$（结合律）
4. $A \cup A = A$（幂等）
5. $A \subset B \Leftrightarrow A \cup B = B$

我们利用并集操作把若干个单元素集结合成一个含有限个元素的集合。例如，

$$\{a, b, c\} = \{a\} \cup (\{b\} \cup \{c\}) = (\{a\} \cup \{b\}) \cup \{c\} = \bigcup_{X \in \{\{a\}, \{b\}, \{c\}\}} X$$

依此类推，任意有限个元素的集合都可由此方法构建*。

给定集合 A 和 B，集合 $C = \{x \in A | x \notin B\}$ 称 B 在 A 中的相对补集（*relative complement of B in A*），记为 $C = A \setminus B$。注意，此处 B 不必包含于 A。

我们常在给定一个全集 E 之下讨论其子集在 E 中的相对补集。如果默认这一前提，则可简称任一 E 的子集 $A \subset E$ 在 E 中的相对补集为 A 的补集，记为 $A^{\complement} \equiv E \setminus A$†。

给定全集 E 下集合的补集有如下性质：

1. $\left(A^{\complement}\right)^{\complement} = A$
2. $\emptyset^{\complement} = E$
3. $E^{\complement} = \emptyset$
4. $A \cap A^{\complement} = \emptyset$
5. $A \cup A^{\complement} = E$
6. $A \subset B \Leftrightarrow B^{\complement} \subset A^{\complement}$

关于集合的并集、交集，以及给定全集下的补集还有一条重要定律——德摩根定律（*De Morgan's Laws*）：对任意集合 A、B，有

$$(A \cup B)^{\complement} = A^{\complement} \cap B^{\complement}$$
$$(A \cap B)^{\complement} = A^{\complement} \cup B^{\complement}$$

给定一个集合 A，我们规定存在一个集合，暂记作 $\mathscr{P}'(A)$，它包含所有 A 的子集。这是集合论的幂集公理（*axiom of powers*）。我们于是可遵守分类公理给出恰好包括 A 的所有子集（包括空集和集合 A 本身），而不包括其他元素的集合：

$$\mathscr{P}(A) = \{X \in \mathscr{P}'(A) | X \subset A\}$$

*至此，我们可以回过头来重新理解集合的定义1.1中的“无序性”、“互异性”和“确定性”。头两个规定，其实是外延公理的推论。例如，若集合 $A = \{a, b, c\}$，$B = \{c, b, a\}$，则由外延公理 $A = B$（无序性）。若集合 $A = \{a, a\}$，$B = \{a\}$，则由外延公理 $A = B$（互异性）。最后，定义1.1中的“确定性”，只是逻辑上排中律的要求。

†有时，我们使用恒等号“$\equiv$”来表示新引入的符号或者记法代表什么表达式。

我们称 $\mathscr{P}(A)$ 为集合 A 的幂集（*power set*）。

为什么把这样的集合称为“幂集”呢？若 $A=\emptyset$，则 $\mathscr{P}(A)=\{\emptyset\}$，有 $1=2^0$ 个元素；若 $A=\{a\}$，则 $\mathscr{P}(A)=\{\emptyset,\{a\}\}$，有 $2=2^1$ 个元素；若 $A=\{a,b\}$，则 $\mathscr{P}(A)=\{\emptyset,\{a\},\{b\},\{a,b\}\}$，有 $4=2^2$ 个元素；……

注意到，若 $X\in\mathscr{P}(E)$，则有 $X^{\complement}\in\mathscr{P}(E)$。于是，默认以 E 为全集时，我们不必逐个讨论 E 的子集的补集。设 $\mathscr{C}\subset\mathscr{P}(E)$，则 $\mathscr{C}$ 中的元素的补集的集合为

$$\mathscr{D}=\left\{X\in\mathscr{P}(E)\,|X^{\complement}\in\mathscr{C}\right\}$$

这时，德摩根定律有如下更一般的形式：

$$\left(\bigcup_{X\in\mathscr{C}}X\right)^{\complement}=\bigcap_{X\in\mathscr{C}}X^{\complement}$$
$$\left(\bigcap_{X\in\mathscr{C}}X\right)^{\complement}=\bigcup_{X\in\mathscr{C}}X^{\complement}$$

其中我们引入了以下记法惯例：

$$\bigcap_{X\in\mathscr{C}}X^{\complement}\equiv\bigcap_{X\in\mathscr{D}}X,\bigcup_{X\in\mathscr{C}}X^{\complement}\equiv\bigcup_{X\in\mathscr{D}}X$$

由集合的定义1.1所要求的无序性，给定两个集合 a、b，它们的配对集合 $\{a,b\}$ 是不表示顺序的。我们可以用 a 的单元素集 $\{a\}$ 和 $\{a,b\}$ 这两个集合，配对得到集合 $\{\{a\},\{a,b\}\}\equiv(a,b)$，来表示有序对（*ordered pair*）。在 (a,b) 的元素中，$\{a,b\}$ 表明我们讨论哪两个元素的有序对，而 $\{a\}$ 表明哪一个元素放在前面。所以，$(a,b)\neq(b,a)$。

给定两个集合 A、B，是否可以给出所有 $a\in A$ 且 $b\in B$ 的有序对 (a,b) 的集合？我们留意到，对任意 $a\in A$ 和 $b\in B$ 有 $\{a\}\subset A$，$\{b\}\subset B$，$\{a,b\}\subset A\cup B$，$\{a\}\subset A\cup B$。可见，$\{a\}$ 和 $\{a,b\}$ 都是集合 $A\cup B$ 的子集，故 $(a,b)\in\mathscr{P}(A\cup B)$，$(a,b)\subset\mathscr{P}(\mathscr{P}(A\cup B))$。换言之，只要 $a\in A$ 且 $b\in B$，就有 $(a,b)\subset\mathscr{P}(\mathscr{P}(A\cup B))$。于是我们可以遵循分类公理，把所有 $a\in A$ 且 $b\in B$ 的有序对 (a,b) 的集合规定出来，且其唯一性由外延公理保证*。故我们可

*用分类公理、并集公理、配对公理和幂集构建集合 A 与 B 的笛卡尔集的过程，用自然语言描述将十分繁琐。以下是采用合式公式（*well-formed formula*）表达的结果，仅供熟悉此知识的读者参考。

$$A\times B=\{X\in\mathscr{P}(\mathscr{P}(A\cup B))\,|\varphi(X)\}$$

其中，$\varphi(X)$ 表示关于 X 的语句，

$$\varphi(X)=\exists U\exists V\exists W\exists Y\,(U\in A\wedge V\in B\wedge\phi(W,U,U)\wedge\phi(Y,U,V)\wedge\phi(X,W,Y))$$

而关于 X、U 和 V 的语句

$$\phi(X,U,V)=\forall Z\,(Z\in X\leftrightarrow((X=U)\vee(X=V)))$$

定义由集合 A 和 B 形成的所有满足 $a \in A$ 且 $b \in B$ 的有序对的集合为 A 与 B 的笛卡尔积（*Cartesian product*），记为 $A \times B$。

以下是与笛卡尔集有关的一些性质：

1. $(A \cup B) \times X = (A \times X) \cup (B \times X)$
2. $(A \cap B) \times (X \cap Y) = (A \times X) \cap (B \times Y)$
3. $(A \setminus B) \times X = (A \times X) \setminus (B \times X)$
4. $A = \emptyset$或$B = \emptyset \Leftrightarrow A \times B = \emptyset$
5. $A \subset X$且$B \subset Y \Rightarrow A \times B \subset X \times Y$*
6. $A \times B \subset X \times Y$且$A \times B \neq \emptyset \Rightarrow A \subset X$且$B \subset Y$

不难留意到，笛卡尔积运算不满足交换律和结合律。

多于两个集合的笛卡尔集的严格定义，需要在介绍完映射的概念后才能给出。

1.2 关系

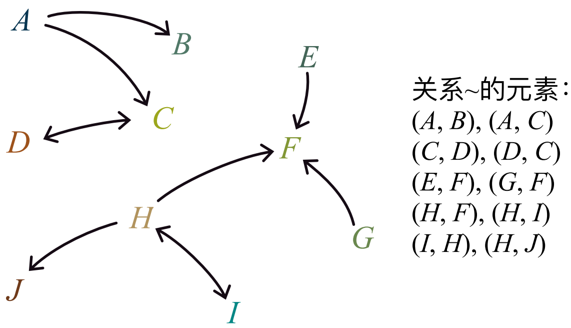
一个集合的元素可与另一个集合的元素形成对应关系。例如，设 A 是所有成年公民的集合，那么“婚姻”就是定义在 A 的任意两个不同元素之间的关系。每对夫妻都是一个有序对 $(a,b) \in A \times A$。在实际社会生活中，我们是先用其他概念对婚姻关系进行定义（例如当地的《婚姻法》），再辨别任意两个公民之间是否具有婚姻关系的。但是在集合论中，我们没有超出集合论的其他概念以供我们独立地定义元素间的一种关系。我们只能视符合某关系的所有有序对的集合为这一关系的定义。例如，我们不采用既有的《婚姻法》来定义何谓婚姻关系，而是把所有具有合法婚姻关系的公民对全部列出来组成一个集合，作为关于“何谓婚姻关系”的一种完整的界定。要辩认 $a, b \in A$ 是否婚姻关系，就只看有序对 (a,b) 是否属于上述集合。这种定义关系的方法才是集合论可以普适地采用的。

正式地，若集合 R 的元素都是有序对，则集合 R 就是一个关系（*relation*）。若有序对 $(x,y) \in R$，则记为 xRy。习惯上，一般的关系常用符号“$\sim$”表示，各种特殊的关系会用特定的符号表示。设 $\sim$ 是一个关系，若 $(x,y) \in \sim$，则 $x \sim y$；若 $(x,y) \notin \sim$，则记为 $x \nsim y$。关系的定义告诉我们：

1. 关系 $\sim$ 是一个有序对的集合。图1.1的箭头表示法可协助我们把“有序对的集合”联系到“关系”一词的日常意义。

其中，记号 $\forall a$ (关于 a 的语句) 表示“对每一/任一满足关于 a 的语句规定的 a”。注意到，语句 $\phi(X,U,V)$ 表示的就是 $X = \{U, V\}$ 这件事，故语句 $\varphi(X)$ 就是让 $X = \{\{U\}, \{U, V\}\}$。

*符号 $\Rightarrow$ 的意义：（语句 1）$\Rightarrow$（语句 2）表示“若（语句 1 ），则（语句 2）。”



$$\operatorname{dom}\sim=\{A,C,D,E,G,H,I\}$$
$$\operatorname{ran}\sim=\{B,C,D,F,H,I,J\}$$

图 1.1: 图中展示了一个关系 $\sim$。箭头表示一个有序对中两个元素的先后次序。

2. 任一关系 $\sim$ 总可以写成两个集合的笛卡尔积的子集。证明的方法：验证关系 $\sim$ 至少可以是下列笛卡尔积

$$\left(\bigcup_{X\in\bigcup_{X'\in\sim}X'}X\right)\times\left(\bigcup_{X\in\bigcup_{X'\in\sim}X'}X\right)$$

的子集*

3. 关系的元素未必是同一个集合与其自身的笛卡尔集。两个不同集合 X 与 Y 之间也可以定义某关系 $\sim\subset X\times Y$。只要 $x\in X, y\in Y, (x,y)\in\sim$，则 $x\sim y$。若 $\sim\subset X\times X$，则称"$\sim$ 是集合 X 上的关系"；若 $\sim\subset X\times Y$，则称"$\sim$ 是从集合 X 到集合 Y 的关系"。

例 1.1. 设 $A=\{a,b\}, B=\{1,2\}$，则 $A\times B=\{(a,1),(a,2),(b,1),(b,2)\}$。设关系 $\sim=\{(a,1),(b,2)\}$。我们不难留意到，$\sim\subset A\times B$。按照关系 $\sim$ 的定义，我们可以写 $a\sim 1, a\not\sim 2$。

留意到，$A\times B$ 本身就是一个关系。若 $\sim=A\times B$，则 A 的任一元素与 B 的任一元素之间都有 $\sim$ 关系，即 $\forall a\in A\forall b\in B, a\sim b$。

等于"$=$"关系是任一集合与其自身的笛卡积的子集。具体地，设 X 是一个非空集合，则 $X\times X$ 中所有满足 $x=y$ 的有序对 (x,y) 的集合就是等于关系。

属于"$\in$"也是一个关系。具体地，它是 $X\times\mathscr{P}(X)$ 满足 $x\in A$ 的所有有序对 (x,A) 的集合。

*提示：回顾有序对的定义 $(a,b)\equiv\{\{a\},\{a,b\}\}$，把表达式

$$\bigcup_{X\in\bigcup_{X'\in\sim}X'}X$$

所形成的集合写出来，可以发现它是关系 $\sim$ 的所有有序对中的元素的集合。读者可以尝试以图1.1的例子写出图中关系的 $\bigcup_{X\in\bigcup_{X'\in\sim}X'}X$；它就是 $\{A,B,C,D,E,F,G,H,I,J\}$。

空集是有序对的集合（因为空集是集合，且空集不含有任何不是有序对的元素），因此空集也可以是一个关系。

接上列的第 2 条，给定一个关系 $\sim$，记 $U_\sim \equiv \bigcup_{X\in\bigcup_{X'\in\sim} X'} X$，遵循分类公理所构建的集合

$$\{a \in U_\sim | \exists b\,(b \in U_\sim \wedge a \sim b)\}$$

为关系 $\sim$ 的定义域（*domain*），记作 $\operatorname{dom} \sim$。集合

$$\{b \in U_\sim | \exists a\,(a \in U_\sim \wedge a \sim b)\}$$

为关系 $\sim$ 的值域（*range*），记作 $\operatorname{ran} \sim$。图1.1给出了所示关系的定义域和值域。不难留意到，对于任一集合上的等于关系，有 $(\operatorname{dom} =) = (\operatorname{ran} =)$；对于任一集合上的属于关系，若 $(\operatorname{dom} \in) = X$，则 $(\operatorname{ran} \in) = \mathscr{P}(X) \setminus \{\emptyset\}$[*]。

定义 1.2 (等价关系). 设 $\sim$ 是集合 X 上的一个关系。若

1. 对任一 $x \in X$ 都有 $x \sim x$，则称关系 $\sim$ 是自反的（*reflextive*）。
2. 对任意 $x \in X$ 和 $y \in X$，只要 $x \sim y$ 就有 $y \sim x$，则称关系 $\sim$ 是对称的（*symmetric*）。
3. 对任意 $x \in X$、$y \in X$ 和 $z \in X$，只要 $x \sim y$且$y \sim z$ 就有 $x \sim z$，则称关系 $\sim$ 是传递的（*transitive*）

若集合 X 上的一个关系 $\sim$ 同时满足上述 3 个性质，则称 $\sim$ 是 X 上的一个等价关系（*equivalent relation*）。

以图1.1为例，如果图1.1中的每个元素都有一个从自己回到自己的箭头，那么图中的关系就是自反的；如果图1.1中的所有箭头都是双向箭头，那么图中的关系就是对称的。读者可尝试把图1.1的关系修改成非自反、非对称，但传递的关系[†]。

例 1.2. 设 $X = \{a, b, c\}$。X 上的等于关系是 X 上的等价关系。它是集合

$$\{(a,a),(b,b),(c,c)\}$$

这一集合共有 *3* 个元素。同时，$X \times X$ 也是 X 上的等价关系，它有 $2^3 = 8$ 个元素。

一般地，任一非空集合 X 上的等于关系是 X 上的（除空集外）“最小”的等价关系，$X \times X$ 是 X 上的最大等价关系。

不等于 $\neq$ 只满足对称性。

[*]因为 $\emptyset \in \mathscr{P}(X)$，但没有元素能够属于 $\emptyset$。

[†]注意这几条定义所要求的“对任意……”，只要有一个例外就可以失效。

任意集合的包含关系具有自反性和传递性，但集合的包含关系没有对任意集合均成立的对称性。事实上，在上一节关于外延公理的段落中已经介绍过，具有对称性的包含关系就是集合的等于关系。

显然，“婚姻”不是“全体成年公民”集合上的等价关系。“婚姻”只满足对称性。

此时我们回过头讨论集合论最基本的关系——从属关系的自反性、对称性和传递性。从属关系的自反性是指，一个集合 A 属于它自己，$A \in A$。从属关系的对称性是指，$A \in B \Leftrightarrow B \in A$。从属关系的传递性是指，$A \in B \wedge B \in C \Rightarrow A \in C$。直觉上，这三种性质都不令人舒适，但它们未被直至目前的讨论所禁止。若想禁止上述性质，需要正则公理（*axiom of regularity*）：给定任一非空集合 X，则 X 中必含有一个元素 $y \in x$ 满足 $y \cup X = \emptyset$，用符号表述为 $X \neq \emptyset \Rightarrow \exists y\,(y \in X \wedge y \cup X = \emptyset)$。正则公理同时禁止了三种性质（证明从略）。本讲义所使用的集合论遵循正则公理。

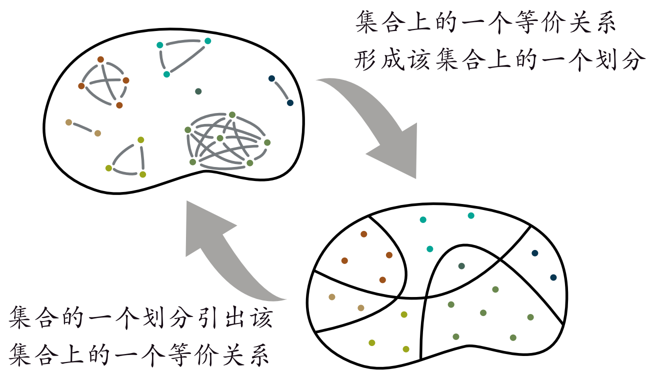


图 1.2: 集合的划分与集合上的等价关系

回到等价关系的讨论。如果集合 X 的非空子集的集合 $\mathscr{C}$ 满足 $\bigcup_{Y \in \mathscr{C}} Y = X$ 且 $\mathscr{C}$ 的元素两两不相交，则称集合 $\mathscr{C}$ 是 X 的一个划分（*partition*）。换言之，如果 X 的若干个非空子集两两不相交，但它们的并集又恰好得到 X，那么这些子集就好像对集合 X 进行“切蛋糕”所得到结果（如图1.2右下的情况）。

若集合 $\mathscr{C}$ 是集合 X 的一个划分，我们可以由此定义一个关系 $\sim \equiv X/\mathscr{C}$*，使得当且仅当 X 的元素 x, y 属于 $\mathscr{C}$ 的同一个元素时，$x \sim y$。正式地，

$$X/\mathscr{C} = \{(x, y) \in X \times X | \exists A\,(A \in \mathscr{C} \wedge \{x, y\} \subset A)\} \tag{1.1}$$

图1.2中，将右下所示的集合的划分中同属一个子集的元素两两连线，就能得到左上所示的关系。这一关系在给定划分 $\mathscr{C}$ 下的唯一性由外延公理保证。

*该记法与刚刚介绍完的 $X/\sim$ 无关，是符号“/”的滥用。

可以证明，这一关系是等价关系：

证明. 验证关系 $\sim \equiv X/\mathscr{C}$ 是等价关系，需一一验证其自反性、对称性和传递性。

1. 自反性：需证明对任意 $x \in X$，有序对 $(x,x) \in \sim$。这需要：
   (a) $(x,x) \in X \times x$。由 $x \in X$ 这显然满足。
   (b) $\exists A \in \mathscr{C}, x \in A$。由并集的定义（式(1.1)），$\mathscr{C}$ 作为一个集合的集合，对任一 $x \in \bigcup_{A\in\mathscr{C}} A$ 必存在 $A \in \mathscr{C}$ 满足 $x \in A$。现在 $\mathscr{C}$ 是集合 X 的一个划分，即 $\bigcup_{A\in\mathscr{C}} A = X$，故对任一 $x \in X$ 必存在 $A \in \mathscr{C}$ 满足 $x \in A$。自反性证毕。
2. 对称性：需证明 $(x,y) \in \sim \Leftrightarrow (y,x) \in \sim$，其中 $x,y \in X, x \neq y$。显然，由于 $x,y \in X$，$(x,y) \in X \times X$ 且 $(Y,x) \in X \times X$。若 $(x,y) \in \sim$，则由 $\sim$ 的定义 $\exists A \in \mathscr{C} \{x,y\} \subset A$，故自然有 $(y,x) \in \sim$；反之亦然。对称性证毕。
3. 传递性：需证明 $(x,y) \in \sim \wedge (y,z) \in \sim \Rightarrow (x,z) \in \sim$，其中 $x,y,z \in X, x \neq y, x \neq z, y \neq z$。显然，由于 $x,z \in X$，$(x,z) \in X \times X$。由 $\sim$ 的定义，$x \sim y \Rightarrow \exists A \in \mathscr{C}, \{x,y\} \subset A$，$y \sim z \Rightarrow \exists A'\mathscr{C}, \{y,z\} \subset A'$。由于 $\mathscr{C}$ 是 X 的一个划分，由划分的定义，要么 $A = A'$，要么 $A \cap A' = \emptyset$，故 $A = A'$，即 $\{x,y,z\} \subset A$。再由关系 $\sim$ 的定义有 $x \sim z$。传递性证毕。

□

因此我们称式(1.1)定义的等价关系是由集合 X 的划分 $\mathscr{C}$ 引出的等价关系（*equivalent relation induced by the partition* $\mathscr{C}$ *of* X）。

上一段介绍的是图1.2右下到左上的定理，下面我们将介绍相当于图1.2中从左上到右下的定理。

设关系 $\sim$ 是集合 X 上的一个等价关系，则集合 $[\![x]\!]_\sim \equiv \{y | y \in X \wedge (\exists x \in X, y \sim x)\}$ 称 x 关于 $\sim$ 的等价类（*equivalent class*）[*]。X 的元素关于 $\sim$ 的所有等价类的集合，记作 $X/\sim$[†]，称为集合 X 在等价关系 $\sim$ 下的商集（*quotient set*）[‡]。如图1.2左上的情况所示，在一个集合上定义了等价关系。每个元素，都能通过这一等价关系的传递性联系若干个共同关联的元素，而形成 X 的一个子集。每个这样的子集，都是 X 关于这一等价关系的等价类。

定理 1.1 (等价关系基本定理). 设 $\sim \subset X \times X$ 是集合 X 上的一个等价关系，则 X 在 $\sim$ 下的商集 $S/\sim$ 是 S 的一个划分。

---

[*]“类”与“集合”在概念上无实质区别。

[†]注意与相对补集的符号相区别。

[‡]正式地，

$$X/\sim \equiv \{A \in \mathscr{P}(X) | \forall x (x \in A \wedge [\![x]\!]_\sim = A)\}$$

证明. 根据划分的定义，要使 $S/\sim$ 是 S 的一个划分，以下 3 条必须同时满足：

1. $S/\sim$ 的元素都不是空集，即 $\forall [\![x]\!]_\sim \in S/\sim, [\![x]\!]_\sim \neq \emptyset$；
2. $S/\sim$ 的元素两两不交，即 $[\![x]\!]_\sim \neq [\![y]\!]_\sim \Leftrightarrow [\![x]\!]_\sim \cap [\![y]\!]_\sim = \emptyset$；
3. 所有 $S/\sim$ 的元素并集得到集合 S，即 $\bigcup_{Y\in S/\sim} Y = S$。

具体证明过程暂略*。 □

推论 1.1.1. 由 $X/\sim$ 引出的等价关系就是 $\sim$。

证明. 证明过程是十分直接的，暂略。提示：利用外延公理，即集合相等的概念。 □

由等价关系基本定理及其推论，我们可以写

$$\sim = X/\mathscr{C} \Leftrightarrow \mathscr{C} = X/\sim$$

可见，符号 / 的用法使得等价关系、划分和商集之间有类似“集合的除法”的意义（故称“商集）。

1.3　映射

定义 1.3 (映射). 设 X 和 Y 是集合，如果由 X 到 Y 的关系 f 同时满足：

1. $\mathrm{dom} f = X$；
2. 对每一 X 的元素 $x \in X$，有且只有一个 Y 的元素 $y \in Y$ 满足 $(x,y) \in f$，

则称 f 是由 X 到 Y 的映射（*mapping*），记作 $f : X \to Y$†。对每一 $(x,y) \in f$，称 y 是 x 在映射 f 下的值（*value*），记作 $f(x)$。

映射定义的第 1 条要求如果违反了，可通过对集合 X 的改动重新得到满足，而无需改动关系 f 本身。例如若 $\mathrm{dom} f \subsetneqq X$，则令 $X' = \mathrm{dom} f$ 并改为讨论由 X' 到 Y 的关系 f，就可通过映射定义的第 1 条。然而映射定义的第 2 条如果违反了，想要重新满足就不得不对关系 f 本身进行改动。图1.3中的第一个例子就只是一个关系，而不是一个映射，除非我们从这一关系中拿掉一个有序对。

给定映射 $f : X \to Y$，我们继续以下讨论：

---

*证明过程

†记号 $f : X \to Y$ 包含的信息是：

1. X、Y 是集合；
2. f 是由 X 到 Y 的映射。

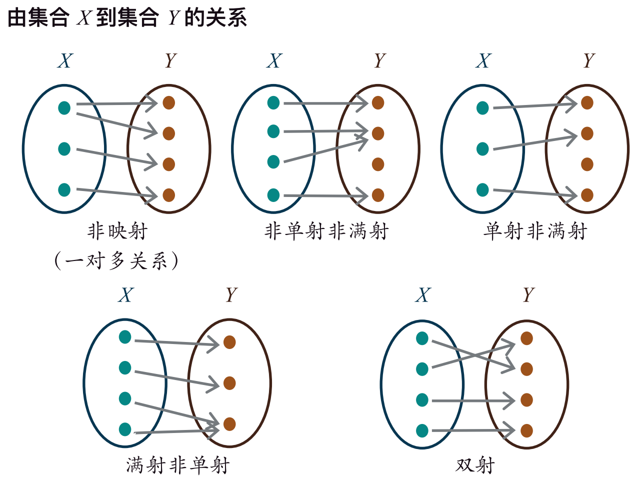


图 1.3: 映射的不同概念示意图。

- 一般地，Y 未必等于 $\mathrm{ran}f$，集合 Y 称映射 f 的陪域（*codomain*）。
- 若 E 是 X 的一个子集，若由 E 到 Y 的映射 $f|_E : E \to Y$ 满足 $f|_A(x) = f(x), \forall x \in E$，则称映射 $f|_E$ 是映射 f 在 E 上的限制（*restriction*）（记法亦已表明）*。
- 若 $\mathrm{ran}f = Y$ 则称映射 f 是满射（*surjective mapping*）。图1.3中的第 4 和第 5 个例子都是满射。
- 若 $A \subset X$，则集合 $\{y \in Y | \forall x (x \in A \wedge y = f(x))\}$ 称集合 A 在映射 f 下的像（*image*），简记为 $f(A)$。这一集合可用语言描述为：由集合 A 的所有元素在映射 f 下的值组成的集合。易证它是 Y 的子集。
- 若对任意 $x_1, x_2 \in X$，$f(x_1) = f(x_2) \Rightarrow x_1 = x_2$，则称 f 是单射（*injective mapping*）。可用语言描述为，“单射的输出唯一地确定其输入”。图1.3中的第 3 和第 5 个例子都是单射。
- 如果 f 既是满射又是单射，则称 f 是一个双射（*bijective mapping*）。图1.3中的第 5 个例子是双射。
- 若另一映射 $g : Y \to Z$，可与映射 f 构成从 X 到 Z 的映射 $g \circ f : X \to Z$，且

  $$g \circ f(x) = g(f(x)), \forall x \in X$$

  则称 $g \circ f$ 是 f 和 g 的复合映射（*composite mapping*）。
- 如果 $f(x) = x, \forall x \in X$，则称映射 f 是恒等映射（*identity mapping*）。
- 如果映射 $g : Y \to X$ 使得复合映射 $g \circ f$ 是恒等映射，则称映射 f 是可逆的（*invertible*），映射 g 是 f 的逆映射（*inverse mapping*）。常将 f 的逆映射记作 f^{-1}。

关于逆映射，有一条重要的定理——

*通俗说 f 在 A 上的限制跟 f 是同一个映射，只不过定义在了“更小的”一个集合 A 上罢了。

定理 1.2. 双射必存在唯一逆映射。双射的逆映射也是双射。

证明. 为了证明这一定理，我们首先证明一个引理：任一单射非满射均存在逆映射。

设 $f:X\to Y$ 是一个单射非满射，即 $\exists y\notin \mathrm{ran}f, y\in Y$。由集合的相关定义此处必有 $\{y|y\in \mathrm{ran}f\}\cup\{y|y\notin \mathrm{ran}f, y\in Y\}=Y$。

现定义 $g:Y\to X$，为使 g 为一个映射，它必须对 $y\in\mathrm{ran}f$ 和 $y\notin\mathrm{ran}f, y\in Y$ 均有定义。现将其定义为：

$$g(y)=\begin{cases} x|_{f(x)=y}, & y\in\mathrm{ran}f \\ \text{任一}x\in X, & y\notin\mathrm{ran}f, y\in Y\end{cases}$$

则有如下几条结论：

1. $g(y)$ 是映射。因为它对每一 $y\in Y$ 均有定义且一个 $y\in Y$ 只对应一个 $x\in X$。
2. g 是满射。因为，仅 $y\in\mathrm{ran}f$ 情况的定义式就已决定了 $\mathrm{ran}g=X$。
3. g 是非单射。因为 g 是满射，再考虑 $y\notin\mathrm{ran}f, y\in Y$ 情况的定义式，就可知 $\exists x\in X$ 满足 $x=g(y)=g(y')$，其中 $y\neq y', y\in\mathrm{ran}f, y'\notin\mathrm{ran}f, y'\in Y$。
4. g 是 f 的逆映射。因为，对于任一 $x\in X$ 均有 $g\circ f(x)\equiv g(f(x))=x$，即 $g\circ f=\mathrm{id}_X$。
5. 一般地，g 是不唯一的。因为 $y\notin \mathrm{f}, y\in Y$ 的情况可定义 $g(x)$ 等于任一 $x\in X$，故只要集合 X 不是只有一个元素，那么 g 都不唯一。

该引理证毕。

现在正式证定理1.2。从上面定义的这个 g 继续，如果 g 是双射，则 g 不仅是满射，还是单射。由刚刚证完的引理，可用类似方法给 g 找一个逆映射 $f':X\to Y$。而且，由于 $\mathrm{ran}g\equiv X$，我们无需像定义 g 那样为 f' 分出 $x\notin\mathrm{ran}g, x\in X$ 的情况，因为不存在这种情况。故

$$f'(x)=y|_{g(y)=x}$$

是 g 的逆映射，且 f' 是满射。而且，把 g 的定义代入上式有 $f'(x)=y|_{g(y)=x}=y|_{x|_{f(x)=y}}=f(x)$，即 f' 不是别的映射而恰为 $f(x)$。即 g 的逆映射是唯一的。因 f' 是满射故 f 是满射，而 f 本身就是单射，故 f 是双射。 □

我们常把映射写成另一种形式，并给以另一个名称。具体地，当我们把由 I 到 X 的映射 $x:I\to X$ 改称为索引族（*indexed family*）时，定义域 I 就称为索引集（*indexing set*），其元素 $i\in I$ 称为索引（*indexes*）。映射 x 的关于 $i\in I$ 的值，记作 x_i，称为该索引族的一项（*term*）。映射 x 的值域，称作由 I 索引的集合（*set indexed by I*），或笼统地称其为一个索引集（*indexed set*）。我们经常不加分辨地直接把 $\{x_i\}_{i\in I}$ 称为一个由 I 索引的族（*I-indexed family*）。可见，我们无非把映射原有概念的名称换了一套新的名称。我们经常讨论的是以集合为项的族，称为集合的索引族（*indexed family of sets*）。

设 $\mathscr{C}$ 是集合的集合。如果有 $\mathscr{C}$ 恰好还是一个由集合 I 索引的集合，$\mathscr{C}=\{X_i\}_{i\in I}$，那么第一节介绍的 $\mathscr{C}$ 的元素的交集和并集就相应有新的表示方式：

$$\bigcap_{X\in\mathscr{C}} X=\bigcap_{i\in I} X_i,\quad \bigcup_{X\in\mathscr{C}} X=\bigcup_{i\in I} X_i$$

下面我们介绍多于两个集合的笛卡尔集的定义*

定义 1.4 (笛卡尔积). 设 $\{S_i\}_{i\in I}$ 是由 I 索引的族，且它是集合的索引族，$\{s_i\}_{i\in I}$ 也是由 I 索引的族，且 $s_i\in S_i,\forall i\in I$。我们称 $\{S_i\}_{i\in I}$ 的笛卡尔积是由 $\{S_i\}_{i\in I}$ 得出的所有族 $\{s_i\}_{i\in I}$ 的集合，记为

$$\prod_{i\in I} S_i$$

若 $S_i=S,\forall i\in I$，则记 $\prod_{i\in I} S_i\equiv S^I$。

例 1.3. *设 $I=\{a,b\},X_a=\{a_\alpha,a_\beta\},X_b=\{b_\alpha,b_\beta\}$，则按照笛卡尔积的老定义，*

$$X_a\times X_b=\{(a_\alpha,b_\alpha),(a_\alpha,b_\beta),(a_\beta,b_\alpha),(a_\beta,b_\beta)\}$$

其中按照有序对的老定义，$(x,y)=\{\{x\},\{x,y\}\}$。

按照新定义1.4的要求，我们要在 X_a 和 X_b 中各选一个元素组成由 $\{a,b\}$ 索引的族 $\{x_i\}_{i\in\{a,b\}}$。例如，令 $x_a=a_\beta,x_b=b_\alpha$ 所形成的族 $\{x_i\}_{i\in\{a,b\}}\equiv\{x_a,x_b\}=\{a_\beta,b_\alpha\}$ 就是其中一个符合要求的族。所有这样的族，一共有 4 个。因此有

$$\prod_{i\in\{a,b\}} X_i=\{\{a_\alpha,b_\alpha\},\{a_\alpha,b_\beta\},\{a_\beta,b_\alpha\},\{a_\beta,b_\beta\}\}$$

笛卡尔积的老定义与新定义是不冲突的。一般地，老定义下的笛卡尔积 $X_a\times X_b$ 和新定义下的笛卡尔集 $\prod_{i\in\{a,b\}} X_i$ 之间总可以定义一个映射 $f:\prod_{i\in\{a,b\}} X_i\to X_a\times X_b, f(z)=(z_a,z_b),\forall z\equiv\{z_i\}_{i\in\{a,b\}}\in\prod_{i\in I} X_i$。记 $z=\{z_i\}_{i\in\{a,b\}}$，则按定义1.4有 $z_i\in X_i,\forall i\in\{a,b\}$。可以证明 f 是双射：

证明. 设 $z=\{z_i\}_{i\in\{a,b\}},z'=\{z'_i\}_{i\in\{a,b\}}$ 且 $z_a,z'_a\in X_a,z_b,z'_b\in X_b$。若 $z\neq z'$。由映射 f 的定义，$f(z)=(z_a,z_b),f(z')=(z'_a,z'_b)$。由集合相等的定义，$z\neq z'\Rightarrow(z_a,z_B)\neq(z'_a,z'_b)$，即 $f(z)\neq f(z')$，即 f 是非单射。

*第一节中引入的两个集合的笛卡尔集定义，无法推广至不可数无穷多个集合间的笛卡尔积。所以，我们在介绍了映射之后，在族的基础上可重新定义笛卡尔集。这个新的定义，在集合个数为两个的情况下，也导致一种与老定义不同的“有序对”，但是新定义和老定义得出有序对之间总是一一对应的，因此无所谓从哪种定义去理解有序对。虽然笛卡尔积的新定义既兼容两个集合间的情况（甚至把“一个集合的笛卡尔集”也定义了），又能够推广到不可数无穷个集合间的笛卡尔集，但是它却不能在一开始就采用，因为它依赖映射的定义，而映射是一种关系，关系的定义依赖有序对的定义。所以我们至少需要先以老定义获得有序对的概念，才能走到现在这一步。

设 $(u,v) \in X_a \times X_b$[*]，则以 $\{a,b\}$ 索引的族 $x \equiv \{x_i\}_{i\in\{a,b\}}, x_a = u, x_b = v$ 满足 $x_a \in X_a, x_b \in X_b$（按照笛卡尔积的老定义），按照定义1.4，有 $x \in \operatorname{dom} f$。由于所选取的 $X_a \times X_b$ 的元素 (u,v) 是任意的，上述性质对任一 $X_a \times X_b$ 的元素都成立，故 f 是满射。

由双射的定义，f 是双射。 □

因此，笛卡尔集的新定义1.4在集合数为 2 的情况下所得到的集合，跟笛卡尔集的老定义所得到的集合，它们的元素之间是一一对应的。在此情况下老定义和新定义没有本质差别。根据笛卡尔积的新定义，"有序对"又可以定义成索引集有两个元素的族。即 $(x,y) = \{z_i\}_{i\in\{a,b\}}, z_a = x, z_b = y$。这时"序"的性质仍被保留，因为 $(y,x) = \{z_i\}_{i\in\{a,b\}}, z_a = y, z_b = x$ 是不同的族，故有 $(x,y) \neq (y,x)$。而且，按照新定义，仅通过改变索引集元素的个数，就可方便地推广出"有序三元组"、"有序四元组"……的定义，这是比老定义更有利的地方。从今以后，笛卡尔积就不再采用老定义，而采用定义1.4。就算若干个两两不同的集合 $X, Y, \cdots$ 尚未成为一个索引族，由于易验任一集合的非空集合 $\mathscr{C}$ 均可充当其自己的索引集而成为一个索引族，故由 $X, Y, \cdots$ 组成的集合 $\{X, Y, \cdots\}$ 总能成为一个索引族，从而它们的笛卡尔集仍可通过定义1.4得到定义。更一般地，给定任一集合的集合 $\mathscr{C}$，若 $\mathscr{C} \neq \emptyset$，则它的元素的笛卡尔积可由定义1.4记为

$$\prod_{X\in\mathscr{C}} X \equiv \prod_{i\in\mathscr{C}} A_i, \quad A_i = i, \forall i \in \mathscr{C}$$

我们常讨论索引集 I 为自然数集 $\mathbb{N}$ 的子集[†]的族，即 $I \subset \mathbb{N}$。具体地，若

$$I = \{a, a+1, a+2, \cdots, b-2, b-1, b\}, a, b \in \mathbb{N}, a < b$$

则由 I 索引的族可记为 $\{X_i\}_{i=a}^{b}$。相应地，若该族是集合的族，则该族的交集、并集和笛卡尔积可分别记为：

$$\bigcap_{i=a}^{b} X_i, \quad \bigcup_{i=a}^{b} X_i, \quad \prod_{i=a}^{b} X_i$$

特别地，若 $X_i = X, \forall i \in I$，令 $n = b - a + 1$，$\{X_i\}_{i=a}^{b}$ 的笛卡尔积可记为 X^n。例如，$\mathbb{R}^n$ 是所有有序实数 n 元组 $(x_1, \cdots, x_n), x_1, \cdots, x_n \in \mathbb{R}$ 的集合。在数学中我们常常会称 $\mathbb{R}^n$ 是"n 维"的。因此，定义1.4的好处是它可以支持各种无穷维空间的构造。

[*] 此处我们需要规定，只要集合 X 和 Y 都是非空集合，那么它们的笛卡尔集 $X \times Y$ 也是非空集合，即必存在一 $(x,y) \in X \times Y, x \in X, y \in Y$。这是集合论的选择公理（*axiom of choice*）。

[†] 本讲义默认读者常识上理解各种数集，而不再介绍它们的集合论引入。

第二章　线性代数

2.1　向量空间

2.1.1　数域

读者应该已经熟悉自然数、整数、有理数、实数、复数等数的集合，以及定义在它们之上的四则运算法则。现在我们特别关心这些数的集合是否具有对四则运算的封闭性，即集合的元素经过四则运算之后的结果仍属于原集合。

具体地，设 $\mathbb{F}$ 是一个集合，且在 $\mathbb{F}$ 上的两种二元“+”、“×”满足以下要求：

1. 加法交换律：
$$x+y=y+x,\quad \forall x,y\in\mathbb{F}$$
2. 加法结合律：
$$x+(y+z)=(x+y)+z,\quad \forall x,y,z\in\mathbb{F}$$
3. 加法的单位元（零）：$\mathbb{F}$ 中存在一个零元素 $0\in\mathbb{F}$，使得
$$x+0=x,\quad \forall x\in\mathbb{F}$$
4. 对相反数封闭：对任一 $x\in\mathbb{F}$，总存在唯一 $\mathbb{F}$ 的元素，记作 $-x$，满足 $x+(-x)=0$
5. 乘法交换律：
$$x\times y=y\times x,\quad \forall x,y\in\mathbb{F}$$
6. 乘法结合律：
$$x\times(y\times z)=(x\times y)\times z,\quad \forall x,y,z\in\mathbb{F}$$
7. 乘法的单位元（壹）：$\mathbb{F}$ 中存在一个非零元素 $1\in\mathbb{F}$，满足
$$x\times 1=x,\quad \forall x\in\mathbb{F}$$
8. 对倒数封闭：对 $\mathbb{F}$ 中的任一非零元素 x，总存在唯一元素，记作 x^{-1} 或 $1/x$，满足 $x\times x^{-1}=1$

9. 乘法对加法的分配律：

$$x \times (y+z) = x \times y + x \times z, \quad \forall x, y, z \in \mathbb{F}$$

则称集合 $\mathbb{F}$ 是一个域（*field*）。例如，复数 $\mathbb{C}$、实数 $\mathbb{R}$ 和有理数 $\mathbb{Q}$ 是域。整数 $\mathbb{Z}$、自然数 $\mathbb{N}$ 不是域。

我们常将 $a+(-b)$ 记为 $a-b$，将 $a \times b$ 记为 ab。

如果一个域 $\mathbb{F}$ 的子集 $\mathbb{F}'$ 也是域，则称 $\mathbb{F}'$ 是 $\mathbb{F}$ 的子域（*subfield*）。例如，$\mathbb{C}$ 是 $\mathbb{C}$ 的子域；$\mathbb{R}$ 是 $\mathbb{C}$ 的子域；所有形如 $x+y\sqrt{2}, x, y \in \mathbb{C}$ 的复数集合是 $\mathbb{C}$ 的子域。

只要一个集合上所谓的加法和所谓的乘法分别满足上列 9 个要求，它就是一个域。在这种一般性下若不作更多规定，将会允许一些奇怪的域，本讲义不深入探讨。在接下来的内容里，每当提到“数域 $\mathbb{F}$”，都具体指复数域 $\mathbb{C}$ 的任一子域。

2.1.2 向量空间的定义和基本性质

定义 2.1 (向量空间). 给定一个数域 $\mathbb{F}$ 和一个非空集合 $\mathscr{V}$，如果它们满足：

1. $\mathscr{V}$ 的元素的二元运算（*binary operation*）规则，称为加法，记为 $\mathbf{a}+\mathbf{b}, \quad \mathbf{a}, \mathbf{b} \in \mathscr{V}$。该运算满足以下规定：$\forall \mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathscr{V}$：
   (a) 封闭性（*closure*）：$\mathbf{a}+\mathbf{b} \in \mathscr{V}$
   (b) 交换律（*commutative*）[*]：$\mathbf{a}+\mathbf{b} = \mathbf{b}+\mathbf{a}$
   (c) 结合律（*associative*）[†]$(\mathbf{a}+\mathbf{b})+\mathbf{c} = \mathbf{a}+(\mathbf{b}+\mathbf{c})$
   (d) 单位元（*identity elemnet*）：$\exists \mathbf{0} \in \mathscr{V} : \mathbf{a}+\mathbf{0} = \mathbf{a}$
   (e) 逆元（*inverse element*）：$\forall \mathbf{a} \in \mathscr{V}, \exists -\mathbf{a} : \mathbf{a}+(-\mathbf{a}) = \mathbf{0}$
2. 数域 $\mathbb{F}$ 与 $\mathscr{V}$ 的元素间的二元运算规则，称为标量乘法，记为 $\alpha\mathbf{a}, \quad \alpha \in \mathbb{F}, \mathbf{a} \in \mathscr{V}$。该运算满足以下规定：$\forall \alpha, \beta \in \mathbb{F}, \mathbf{a}, \mathbf{b} \in \mathscr{V}$：
   (a) $\alpha\mathbf{a} \in \mathscr{V}$
   (b) $\alpha(\beta\mathbf{a}) = (\alpha\beta)\mathbf{a}$
   (c) $1\mathbf{a} = \mathbf{a}$，其中 1 是数域 $\mathbb{F}$ 的单位元
   (d) $\alpha(\mathbf{a}+\mathbf{b}) = \alpha\mathbf{a}+\alpha\mathbf{b}$
   (e) $(\alpha+\beta)\mathbf{a} = \alpha\mathbf{a}+\beta\mathbf{a}$

则称 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的向量空间（*vector space*）；$\mathscr{V}$ 的元素 $\mathbf{a}, \mathbf{b}, \cdots$ 是向量（*vector*）；$\mathbb{F}$ 中的数是 $\mathscr{V}$ 的标量（*scalar*）。

[*] 对应的英文单词是形容词。

[†] 对应的英文单词是形容词。

读者可尝试利用定义证明 $0\mathbf{a}=\mathbf{0},\forall\mathbf{a}\in\mathscr{V}$。由于这条可由定义证明，故尽管它看上去也很基本，但并不写在定义中。

在定义2.1中，数域 $\mathbb{F}$ 是关于标量乘的运算规定 (2-a) 中的标量所属的数域。任何非空集合，只要配备了定义2.1所规定的运算法则，就是一个向量空间，其元素就是向量。因此，向量是抽象的一般概念。大一的《线性代数与解析几何》课本中的“行/列向量”和“矩阵”，都只是特例。例2.1给出了更多向量空间的例子。这种抽象的定义方式，使得只要是在抽象的层面上论证得到的性质和定理，就必然适用于所有具体的例子当中，使我们学习和发展数学事半功倍。

定义2.1中还有若干需要注意的细节。比如，(1-d) 和 (1-e) 都只规定了存在性，而没有规定唯一性。其实单位元是唯一的；相对于一个向量的逆元也是唯一的。这两个命题可以由定义2.1证明出来，故无需添加到定义中作为规定。请回顾《线性代数与解析几何》第七章“2. 线性空间的性质”下的所有内容。

例 2.1. *以下是一些向量空间的例子。*

1. $\mathbb{R}^n$ 是所有有序实数 n 元组 $(a_i)=(\alpha_1,\cdots,\alpha_n),\alpha_i\in\mathbb{R},i=1,\cdots,n$ 的集合。若 $\forall(a_i),(b_i)\in\mathbb{R}^n$：
   (a) $(a_i)+(b_i)=(\alpha_1+\beta_1,\cdots,\alpha_n+\beta_n)$；
   (b) $\alpha(a_i)=(\alpha\alpha_1,\cdots,\alpha\alpha_n),\forall\alpha\in\mathbb{R}$；
   (c) $(0)=(0,\cdots,0)$；
   (d) $-(a_i)=(-\alpha_1,\cdots-\alpha_n)$；

   则 $\mathbb{R}^n$ 连同上述的运算规定形成数域 $\mathbb{R}$ 上的一个向量空间，又称为实坐标空间。符号 $\mathbb{R}$ 或 $\mathbb{C}$ 既可表示数域，又可表示一个一维实或复坐标空间。可以验证，一维实坐标空间 $\mathbb{R}$ 不是数域 $\mathbb{C}$ 上的向量空间。
2. 数域 $\mathbb{F}$ 上的所有 $m\times n$ 矩阵的集合 $\mathbb{F}^{m\times n}$（连同矩阵的加法和矩阵的标量乘法规定*）是一个向量空间。其零向量是 $m\times n$ 全零矩阵[3]§7.1 例 1.3。
3. 验证：在开区间 (a,b) 上的所有实值一元连续函数的集合是实数域上的向量空间[3]§7.1。
4. 验证：记 $\mathscr{C}^n(a,b)$ 为开区间 (a,b) 上的所有 n 阶连续可导实值一元函数的集合，它是实数域上的向量空间。

接下来直至定义2.6，我们将逐步发现，定义2.1中的性质和运算法则将进一步导致向量空间有一定的维数。

由于“封闭性”的要求，一个向量空间内的任一向量总能被这一向量空间中的其他向量按所规定的运算法则表达出来。具体地，若 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的向量空间，$\mathbf{a}_1,\cdots,\mathbf{a}_n\in\mathscr{V}$，

*见[3]§2.1 矩阵与矩阵的运算

$\alpha_1,\cdots,\alpha_n \in \mathbb{F}$，则依据向量空间的封闭性，$\sum_{i=1}^n \alpha_i\mathbf{a}_i$ 也属于 $\mathscr{V}$，即这一求和的结果也是一个向量，该向量可以用前面这个求和表达式来表达。由此引出线性组合、线性表出和线性无关的概念。

定义 2.2 (线性组合、线性表出、线性无关). * 若 $\mathscr{V}$ 是 $\mathbb{F}$ 上的向量空间，$\mathbf{a}_1,\cdots,\mathbf{a}_n \in \mathscr{V}$，$\alpha_1,\cdots,\alpha_n \in \mathbb{F}$，则 $\sum_{i=1}^n \alpha_i\mathbf{a}_i$ 称为这 n 个向量 $\{\mathbf{a}_1,\cdots,\mathbf{a}_n\}$ 的线性组合（*linear combination*）。令 $\mathbf{b}=\sum_{i=1}^n \alpha_i\mathbf{a}_i$，则称 $\mathbf{b}$ 被 $\{\mathbf{a}_i\}$ 线性表出†。若 $\sum_{i=1}^n \alpha_i\mathbf{a}_i=\mathbf{0}$ 当且仅当 $\alpha_i=0\forall i$，则称向量 $\{\mathbf{a}_i\}$ 线性无关（*linear independent*）。反之，若存在某不全为零的一组 $\{\alpha_i\},\alpha_i\in\mathbb{F},i=1,\cdots,n$ 使得 $\sum_{i=1}^n \alpha_i\mathbf{a}_i=\mathbf{0}$ 则称向量 $\{\mathbf{a}_i\}$ 线性相关（*linear dependent*）。

由定义2.2可证得如下结论[3]§7.2 定义 2.3 下的“常用结果”：

1. 任何真包含一组线性无关向量的向量集合是线性相关的。
2. 任何线性无关向量组的子集也是线性无关向量组。
3. 任何含有 $\mathbf{0}$ 向量的向量组线性相关，因为总有 $1\neq 0$ 使得 $1\mathbf{0}=\mathbf{0}$。
4. 一个向量组 S 是线性无关向量组，当且仅当 S 的所有子集都是线性无关向量组。

向量空间定义中所要求的封闭性保证了向量可类似于我们习惯的数字那样被用作数学表达和运算。因此，封闭性是一个很重要的性质。那么，一个向量空间之内，会不会有一部分子集本身就满足了封闭性呢（就好像复数与实数之间的关系）？我们通过考察 $\mathbb{C}^n$ 和 $\mathbb{R}^n$ 可以举出很多正面的例子。一般地，如果一个向量空间的子集本身也满足封闭性，那么它自己也是一个向量空间（即满足定义2.1）。

定义 2.3 (子空间). [3]§7.1 定义 1.2 及其后例题 令 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的一个向量空间，如果 $\mathscr{V}$ 的子集 $\mathscr{W}\subseteq\mathscr{V}$ 也是一个向量空间（并与 $\mathscr{V}$ 采用相同的加法和标量乘定义），则称 $\mathscr{W}$ 是 $\mathscr{V}$ 的一个子空间（*subspace*）。

按照这一定义，易证 $\mathscr{W}$ 的任意两个向量 $\mathbf{a},\mathbf{b}\in\mathscr{W}$ 和任一标量 $\alpha\in\mathbb{F}$ 构成的线性组合 $\alpha\mathbf{a}+\mathbf{b}$ 也在 $\mathscr{W}$ 内[3]§7.1 定理 1.1。

就算向量空间 $\mathscr{V}$ 的某子集 S 因不满足封闭性而成为不了 $\mathscr{V}$ 的子空间，S 内的向量的所有线性组合表出的向量可以形成一个比 S 更大的集合，且满足封闭性，因而我们可以说由 S 生成了一个 $\mathscr{V}$ 的子空间，定义如下。

定义 2.4 (线性生成空间). 若 S 是向量空间 $\mathscr{V}$ 的非空子集，即 $S\subseteq\mathscr{V},S\neq\emptyset$，那么 S 内的向量的所有线性组合的集合 $\mathscr{W}_S$ 也是一个向量空间，称为由 S 线性生成的子空间（*the subspace spanned by S*），记为 $\mathscr{W}_S\equiv\mathrm{span}S$。

*[3]§7.2 定义 2.2、2.3

†集合 $\{\mathbf{a}_1,\cdots,\mathbf{a}_n\}\subset\mathscr{V}$ 可写为 $\{\mathbf{a}_i\}_{i=1}^n$ 或 $\{\mathbf{a}_i\}$，$\{\mathbf{a}_i\}$ 的一个有序序列则记为 $(\mathbf{a}_i)$。

换句话说，$\mathscr{W}_S$ 中的向量都能由 S 的向量线性表出。一个直接的结论就是 $S \subseteq \mathscr{W}_S$，因为一个向量总能被它自己线性表出。

例 2.2.

- 易验证，三维实坐标空间 $\mathbb{R}^3$ 的子集 $P = \{(x_1, x_2, x_3) \,|x_1 = 0, x_2, x_3 \in \mathbb{R}\}$ 具有封闭性，因此它是 $\mathbb{R}^3$ 的子空间。
- 设 Q 是 $\mathbb{R}^3$ 的子集 $\{(2,1,3), (1,0,1)\}$ 线性生成的子空间，常记为

$$Q = \operatorname{span}\{(2,1,3), (1,0,1)\}$$

  则可验 $\{(7,2,9)\} \in Q$。
- 如果用有序实数三元组表示从原点 $(0,0,0)$ 出发的矢量，则上例中的 $(2,1,3), (1,0,1)$ 的两个矢量不共线（易验它们线性无关），子空间 Q 是由这两个矢量所确定的平面。

从上面的例子看到，一个向量空间与其子空间之间的关系，暗示了某种维度的概念。我们首先可以明确“一个向量空间维数”的一般意义，但需要先引入“基”的概念。

定义 2.5 (向量空间的基). [3]§7.2 定义 2.4 如果向量空间 $\mathscr{V}$ 是其子空间 B 的线性生成空间（即 $\mathscr{V} = \operatorname{span}B$，且 B 内的所有向量线性无关，则称 B 是 $\mathscr{V}$ 的一组基（*basis*）。如果 B 含有有限个向量，则称 $\mathscr{V}$ 是有限维向量空间。

上一定义仅引入了“有限维”的概念，但没有具体涉及到“维数是几”的问题。因为按目前已有的定义，$\mathscr{V}$ 内可能可以找出不止一组满足定义的基，而这些基是否必然都具有相同个数的向量？只有当这个问题的答案是肯定的，我们才能通过把它们的个数统一定义为 $\mathscr{V}$ 的维数，来使得向量空间具有确定的维数。下面的定理解决了这个问题。

定理 2.1. 有限维向量空间的每组基具有相同个数的线性无关向量。

证明. 此略[3]“(3) 的证明”，p. 171[4]§2.3,Theorem 4,p. 44。 □

有了这一定理，我们就可以直接把有限维向量空间的维数定义为它的任一组基的向量个数——

定义 2.6 (有限维向量空间的维数). 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的有限维向量空间，它的维数（*dimension*），记为 $\dim\mathscr{V}$，是它的任一组基的向量个数。规定：零向量空间（只由一个零向量组成的向量空间）是的维数是 *0*。

由定理2.1还可直接得到如下推论，它们的证明与定理2.1的证明过程很类似，故从略。

推论 2.1.1. 设 $\mathscr{V}$ 是一个有限维向量空间，其维数 $\dim\mathscr{V} = n$，则

1. $\mathscr{V}$ 的任一含有多于 n 个向量的子集都是线性相关向量组。
2. $\mathscr{V}$ 的任一向量个数少于 n 的子集都不能线性生成整个 $\mathscr{V}$（即这样的子集的线性生成空间总是 $\mathscr{V}$ 的真子集）。
3. $\mathscr{V}$ 的任一子空间 $\mathscr{W}$ 的维数不大于 $\mathscr{V}$ 的维数，即 $\dim\mathscr{W} \le \dim\mathscr{V}$；当且仅当 $\mathscr{W} = \mathscr{V}$ 时取等号。

这一推论的第 3 条的一个例子就是之前的例2.2。

例 2.3. 验证以下命题——

- 如果把一维实坐标空间 $\mathbb{R}$ 看作是实数域 $\mathbb{R}$ 上的向量空间，则 $\{1\}$ 是其一组基，故一维实坐标空间 $\mathbb{R}$ 是实数域 $\mathbb{R}$ 上的一维向量空间。
- 如果把一维复坐标空间 $\mathbb{C}$ 看作是实数域 $\mathbb{R}$ 上的向量空间，则 $\{1, i\}$ 是其一组基，故一维复坐标空间 $\mathbb{C}$ 是实数域 $\mathbb{R}$ 上的二维向量空间。
- 如果把一维复坐标空间 $\mathbb{C}$ 看作是复数域 $\mathbb{C}$ 上的向量空间，则 $\{1\}$ 是其一组基，故一维复坐标空间 $\mathbb{C}$ 是复数域 $\mathbb{C}$ 上的一维向量空间。

2.1.3 向量的坐标

到此为止，我们未具体地阐明“向量”是什么，也未具体地规定加法和标量乘法如何进行。“这种抽象性使我们可以把不同的数学对象统一到线性空间这一概念之下。”[3]p. 167不过，通过引入“坐标”的概念，我们又使得任一抽象向量都能用一组有序数组来唯一地表示，从而使抽象的向量之间的运算得以由具体的有序数组的运算来代替（就像我们以往在《线性代数》课中所熟悉的那样）。

定义 2.7 (向量在给定有序基下的坐标). 若 $\mathscr{V}$ 是一个 n 维向量空间，$\mathscr{V}$ 内的一组线性无关的有序向量序列 $(\mathbf{a}_i)$ 线性生成整个 $\mathscr{V}$，则称这组有序向量序列 B 为 $\mathscr{V}$ 的一组*有序基*（*ordered basis*）。由定义2.5，任一向量 $\mathbf{x} \in \mathscr{V}$ 均可表达为 $\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{a}_i$，进而，任一 $\mathbf{x} \in \mathscr{V}$ 在给定有序基下都唯一对应 $\mathbb{F}^n$ 中的一个有序 n 元数组 (x_i)*，我们称这一有序 n 元数组 (x_i) 为向量 $\mathbf{x}$ 在有序基 B 下的*坐标*（*coordinate*）。

基的原始定义（2.5）仅要求基是一个集合，没有有序性的规定。故我们可以书写“$B = \{\mathbf{a}_1, \cdots \mathbf{a}_n\}$ 是某向量空间的一组基”。但由于集合的元素是无序的，用一组基向量表出任意一个 n 维向量时所使用的 n 个标量，若只形成一个集合，那么这一标量集合的元素变换不同的

*这里的唯一性可参考[3]§7.2 定义 2.3 下“(3) 的证明”。

顺序去与基向量组合，将线性表出不同的向量。因此我们必须给基向量规定顺序，形成有序基，才能说出“一个向量的第 i 个坐标”。换句话说，向量坐标的定义需要基于有序基，而非仅是基向量的一个集合。所以定义2.7才特别增加了“有序”的要求。在本讲义中，我们都用同一个符号来表示作为集合而言的一组基（例如“$B=\{\mathbf{a}_1,\cdots,\mathbf{a}_n\}$ 是某向量空间的一组基”）和一组有序基（例如“$B=(\mathbf{a}_1,\cdots,\mathbf{a}_n)$ 是某向量空间的一组有序基”）。

在定义2.7中提到的向量 $\mathbf{x}$ 与其在有序基 B 下的坐标 $(x_1,\cdots,x_n)$ 之间的唯一对应性，可以很方便地证明。沿用定义2.7中的记法，若另有某 $(y_1,\cdots,y_n)\in\mathbb{F}^n$ 满足 $\mathbf{x}=\sum_{i=1}^n y_i\mathbf{a}_i$，其中至少有一 $y_k\notin\{x_1,\cdots,x_n\}$，则 $\mathbf{x}-\mathbf{x}=\sum_{i=1}^n(x_i-y_i)\,\mathbf{a}_i=\mathbf{0}$ 且 $(x_1-y_1,\cdots,x_n-y_n)$ 中有一个数 $x_k-y_k\neq 0$，与 $\{\mathbf{a}_1,\cdots,\mathbf{a}_n\}$ 线性无关矛盾。故 $(x_1,\cdots,x_n)$ 对 $\mathbf{x}$ 是唯一的。

有了坐标的定义，在给定基下，n 维向量空间 $\mathscr{V}$ 中的每一个向量就都与 $\mathbb{F}^n$ 中的一个有序 n 元数组形成了一一对应的关系。由 $\mathscr{V}$ 和 $\mathbb{F}$ 的封闭性，没有一个向量不对应一个数组，反之亦然。易验，通过向量的加法和标量乘法所得到的新向量所对应的坐标，就是原向量的坐标按照 $\mathbb{F}^n$ 上的加法和标量乘法运算的结果*。但是要注意，n 维实坐标空间 $\mathbb{R}^n$ 中的一个向量 $\mathbf{x}\in\mathbb{R}^n$ 本身就是一个有序实数 n 元组 $\mathbf{x}=(x_1,\cdots,x_n)\,,x_i\in\mathbb{R},i=1,\cdots,n$。在选定 $\mathbb{F}^n$ 某组有序基 $B=(\mathbf{e}_1,\cdots,\mathbf{e}_n)$ 下（这些基向量本身也都是有序实数 n 元组），向量 $\mathbf{x}$ 的坐标可能又是另一个不同的有序实数 n 元组 $(\chi_1,\cdots,\chi_n)$。正确的表示是 $\mathbf{x}=\sum_{i=1}^n\chi_i\mathbf{e}_i$，或称向量 $\mathbf{x}$ 在有序基 B 下的坐标是 $(\chi_1,\cdots,\chi_n)$，但不能写成 $\mathbf{x}=(\chi_1,\cdots,\chi_n)$†。

例 2.4. 设 $B=\{\alpha_1,\alpha_2,\alpha_3\}$ 是 $\mathbb{R}^3$ 的一组有序基，其中

$$\alpha_1=(1,0,-1)\,,\quad\alpha_2=(1,1,1)\,,\quad\alpha_3=(1,0,0)$$

请问向量 (a,b,c) 在 B 下的坐标表达式？

在本讲义中，无论是 n 维向量 $\mathbf{x}$ 本身还是其在某基的下的坐标，都一律写成 $n\times 1$ 矩阵（列向量）；为方便，在文字段落中表示为 $1\times n$ 矩阵（行向量）的转置，即 $\mathbf{x}=(x_1,\cdots,x_n)^\intercal$。这里的转置可直接按以前在《线性代数》课中学过的意义来理解。但本讲义会对向量转置的概念进行正式的定义。

2.1.4　向量的内积与范

在本节中，我们在向量空间的基础上再增加一些性质和运算法则，使得“正交”、“单位向量”的概念得以引入，同时也介绍其他相关的概念。

* “利用基和坐标可把线性空间的运算变得更具体。”[3]p.173。

†这里的概念区分可参见[3]§7.1 例题 2.1。

定义 2.8 (内积). 数域 $\mathbb{F}$ 上的向量空间 $\mathscr{V}$ 中两向量 $\mathbf{a}, \mathbf{b} \in \mathscr{V}$ 的内积（*inner product*）记为 $(\mathbf{a}|\mathbf{b})$，其中运算 $(\cdot|\cdot)$ 是由 $\mathscr{V} \times \mathscr{V}$ 到 $\mathbb{F}$ 的映射，满足以下规定*：$\forall \mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathscr{V}, \alpha \in \mathbb{F}$

1. $(\mathbf{a}|\mathbf{b}) = \overline{(\mathbf{b}|\mathbf{a})}$
2. $(\alpha\mathbf{a}|\mathbf{b}) = \alpha (\mathbf{a}|\mathbf{b})$
3. $(\mathbf{a}+\mathbf{b}|\mathbf{c}) = (\mathbf{a}|\mathbf{c}) + (\mathbf{b}|\mathbf{c})$
4. $(\mathbf{a}|\mathbf{a}) \in \mathbb{R}$ 且 $(\mathbf{a}|\mathbf{a}) \geq 0$，当且仅当 $\mathbf{a} = \mathbf{0}$ 时取等号。

带有一种内积运算规定的向量空间叫做内积空间（*inner product space*）。

由内积运算要求的第 3 条易知 $\forall \mathbf{a} \in \mathscr{V}, (\mathbf{a}|\mathbf{0}) = 0$，因为 $\forall \mathbf{a}, \mathbf{b} \in \mathscr{V}$ 均有 $(\mathbf{a}|\mathbf{b}) = (\mathbf{a}|\mathbf{0}+\mathbf{b}) = (\mathbf{a}|\mathbf{0}) + (\mathbf{a}|\mathbf{b}) \Rightarrow (\mathbf{a}|\mathbf{0}) = 0$。由内积定义还可以推出：$\forall \mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathscr{V}, \alpha \in \mathbb{F}$

$$(\mathbf{a}|\alpha\mathbf{b}+\mathbf{c}) = \overline{\alpha} (\mathbf{a}|\mathbf{b}) + (\mathbf{a}|\mathbf{c})$$

即从内积的后一个向量提出标量到内积之外时，这一标量要取复数共轭。

内积的定义中设置复数共轭是必要的。否则将面临如下的矛盾：由 $(\mathbf{a}|\mathbf{a}) > 0 \forall \mathbf{a} \neq \mathbf{0}$，竟然有 $(i\mathbf{a}|i\mathbf{a}) = -1 (\mathbf{a}|\mathbf{a}) > 0$。但是，到底是规定从内积的前一个向量提出的标量要取复数共轭（很多物理书习惯），还是从内积的后一个向量提出的标量要取复数共轭（本讲义的定义方式，也是很多数学书的习惯）？这个惯例的不同，有时会造成重要的差别。我们可以把两种惯例的内积定义用不同的括号来区分，使得这两种惯例的内积的关系是：

$$(\mathbf{x}|\mathbf{y}) \equiv \langle \mathbf{y}|\mathbf{x} \rangle$$

上面的三角括号定义与量子力学中的狄拉克 bra-ket 标记的规定是相同的。由于本讲义的流变学部分不涉及复数域上的向量空间，因此不再详述，可参考[5]§1.3.1。

例 2.5.

1. *在* $\mathbb{F}^n$ *上可定义这样的内积：对于* $\mathbf{a} = (\alpha_1, \cdots, \alpha_n)^\intercal, \mathbf{b} = (\beta_1, \cdots, \beta_n)^\intercal \in \mathbb{F}$，$(\mathbf{a}|\mathbf{b}) \equiv \sum_j \alpha_j \overline{\beta_j}$，*称为* $\mathbb{F}^n$ *上的标准内积*（*standard inner product*）。$\mathbb{R}^n$ *上的标准内积又可记为点乘*（*dot product*）$\mathbf{a} \cdot \mathbf{b}$。
2. *记* $\mathbb{C}(a,b)$ *为所有定义在实开区间* (a,b) *上的复数值一元函数的集合，若通过* (a,b) *上的恒正函数* $w(x)$ *定义内积运算* $(f|g) = \int_a^b w(x) f(x) \overline{g(x)} dx, \forall f, g \in \mathbb{C}(a,b)$，*验证* $\mathbb{C}(a,b)$ *是一个内积空间。*

由内积的一般定义可直接证明柯西–施瓦茨不等式（Cauchy–Schwarz inequality），故该不等式在任一内积空间上都成立。

*上划线表示复数共轭

定理 2.2. 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的一个内积空间，则有 $\forall \mathbf{a}, \mathbf{b} \in \mathscr{V}$，

1. 柯西–施瓦茨不等式：$\left|(\mathbf{a}|\mathbf{b})\right|^2 \leq (\mathbf{a}|\mathbf{a})\,(\mathbf{b}|\mathbf{b})$
2. $(\mathbf{a}|\mathbf{b}) = \mathrm{Re}\,(\mathbf{a}|\mathbf{b}) + \mathrm{Re}\,(\mathbf{a}|i\mathbf{b})$

证明. 当 $\mathbf{a} = \mathbf{0}$ 时，不等式取等号成立。当 $\mathbf{a} \neq \mathbf{0}$ 时，令

$$\mathbf{c} = \mathbf{b} - \frac{(\mathbf{b}|\mathbf{a})}{(\mathbf{a}|\mathbf{a})}\mathbf{a}$$

则可验证 $(\mathbf{c}|\mathbf{a}) = 0$，且

$$\begin{aligned} 0 \leq (\mathbf{c}|\mathbf{c}) &= \left(\mathbf{b} - \frac{(\mathbf{b}|\mathbf{a})}{(\mathbf{a}|\mathbf{a})}\mathbf{a}\,\middle|\, \mathbf{b} - \frac{(\mathbf{b}|\mathbf{a})}{(\mathbf{a}|\mathbf{a})}\mathbf{a}\right) \\ &= (\mathbf{b}|\mathbf{b}) - \frac{\left|(\mathbf{b}|\mathbf{a})\right|^2}{(\mathbf{a}|\mathbf{a})} \\ \Leftrightarrow \left|(\mathbf{a}|\mathbf{b})\right|^2 &\leq (\mathbf{a}|\mathbf{a})\,(\mathbf{b}|\mathbf{b}) \end{aligned}$$

由 $(\mathbf{a}|\mathbf{b}) = \mathrm{Re}\,(\mathbf{a}|\mathbf{b}) + i\mathrm{Im}\,(\mathbf{a}|\mathbf{b})$ 和 $\mathrm{Im}\,(\alpha) = \mathrm{Re}\,(-i\alpha)\,\forall\alpha \in \mathbb{F}$，有 $\mathrm{Im}\,(\mathbf{a}|\mathbf{b}) = \mathrm{Re}\,(-i\,(\mathbf{a}|\mathbf{b})) = \mathrm{Re}\,(\mathbf{a}|i\mathbf{b})$，故 $(\mathbf{a}|\mathbf{b}) = \mathrm{Re}\,(\mathbf{a}|\mathbf{b}) + i\mathrm{Re}\,(\mathbf{a}|i\mathbf{b})$。 □

除了内积空间，我们还可以为一个向量空间引入范的规定，得到赋范向量空间。

定义 2.9 (向量的范). 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 向量空间，$\mathscr{V}$ 上的范（*norm*）作用于 $\mathscr{V}$ 中的任一向量，记为 $\|\mathbf{x}\|$，并满足：

1. 非负性：$\|\mathbf{x}\| \geq 0, \forall \mathbf{x} \in \mathscr{V}$
2. 调和性：$\|\alpha\mathbf{x}\| = |\alpha|\,\|\mathbf{x}\|\,, \forall \mathbf{x} \in \mathscr{V}, \alpha \in \mathbb{F}$
3. 三角不等式：$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|\,, \forall \mathbf{x}, \mathbf{y} \in \mathscr{V}$

带有一种范的定义的向量空间叫赋范向量空间（*normed vector space*）。

我们注意到，赋范向量空间也有一个总成立的不等式——三角不等式。与内积空间的柯西–施瓦茨不等式不同，赋范向量空间的三角不等式是在范的定义中直接规定的，而无法作为定理由之前两个规定证明出来。我们主观上就希望向量的范满足这样的性质，因为“范”是我们为向量定义的一种“长度”的概念，所以希望它能与欧几里得几何公设规定的性质相兼容。

一个赋范向量空间 $\mathscr{V}$ 中，范为 1 的向量称为单位向量（unit vector），在本讲义中单位向量会加一个小帽子来特别表示：$\|\hat{\mathbf{a}}\| = 1, \hat{\mathbf{a}} \in \mathscr{V}$。赋范向量空间 $\mathscr{V}$ 的任一向量 $\mathbf{x}$ 均可通过 $\hat{\mathbf{x}} = \mathbf{x}/\left\|\mathbf{x}\right\|$ 归一化为一个单位向量（由范的调和性易验）。

不管是内积的定义、还是范的定义，都没有具体规定计算方法。只要满足定义的一般要求，我们可以为同一个向量空间赋予不同内积或范的规定。向量的范的其中一种常用的定义是：设

$\mathscr{V}$ 是内积空间，向量 $\mathbf{a}\in\mathscr{V}$ 的范 $\|\mathbf{a}\|\equiv(\mathbf{a}|\mathbf{a})^{\frac{1}{2}}$。我们把这一定义称为欧几里得范（*Euclidean norm*）。其他范的定义则为非欧几里得范，例如在 $\mathbb{R}^n$ 上，还可以有如下范的定义。对任一 $\mathbf{x}=(x_1,\cdots,x_n)$，$\|\mathbf{x}\|=\max\{|x_1|,\cdots,|x_n|\}$*。一般地，由于给一个向量空间引入内积定义的方式本身就可以有多种，依赖内积的范的定义也会有多种。对于有些向量空间，范的定义可以不依赖内积。

由于有限维向量空间上定义的不同的范之间是等价的（见 §A.1），故我们实际只需以欧几里得范为代表进行后续的讨论，未经说明的话，一般提到有限维向量空间上的范都指欧几里得范。

以下定理说明，在一定条件下，我们能够用范的一般定义构造一个内积，使任何一个尚未定义内积的赋范空间成为一个内积空间。而且特别地，这一种范就是欧几里得范。

定理 2.3. 一个赋范向量空间是一个内积空间，当且仅当该空间的范满足极化恒等式（polarization identity），即

$$\|\mathbf{a}+\mathbf{b}\|^2+\|\mathbf{a}-\mathbf{b}\|^2=2\|\mathbf{a}\|^2+2\|\mathbf{b}\|^2$$

证明. 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的一个赋范向量空间，若定义二元运算：

$$\begin{aligned}(\mathbf{a}|\mathbf{b})&=\frac{1}{4}\|\mathbf{a}+\mathbf{b}\|^2-\frac{1}{4}\|\mathbf{a}-\mathbf{b}\|^2+\frac{i}{4}\|\mathbf{a}+i\mathbf{b}\|^2-\frac{i}{4}\|\mathbf{a}-i\mathbf{b}\|^2\\&=\frac{1}{4}\sum_{n=1}^{4}i^n\|\mathbf{a}+i^n\mathbf{b}\|^2,\forall\mathbf{a},\mathbf{b}\in\mathscr{V}\end{aligned}$$

可验证上式满足内积定义，且 $\|\mathbf{a}\|^2=(\mathbf{a}|\mathbf{a})^{\frac{1}{2}}$，即该范就是欧几里得范。 □

上面的等式在几何上等价于*平行四边形法则*（*parallelogram law*）。特别地，对于 $\mathbf{a},\mathbf{b}\in\mathbb{R}^n$ 有 $\mathbf{a}\cdot\mathbf{b}=\frac{1}{4}\|\mathbf{a}+\mathbf{b}\|^2-\frac{1}{4}\|\mathbf{a}-\mathbf{b}\|^2$。

柯西–施瓦茨不等式、三角不等式和极化恒等式在很多定理的证明中经常用到，但它们的含义及适用范围需要区分清楚。由定理2.2，柯西–施瓦茨不等式是对任一内积空间均成立的。由定义2.9，三角不等式是对任一赋范向量空间均成立的。而由定理2.3可知，满足极化恒等式的赋范空间必然也是一个内积空间，从而两种空间得到了统一。特别地，*欧几里得范是同时满足三角不等式和极化恒等式的范*。

下面我们由内积空间的性质引入正交（orthogonal）及相关的概念。

定义 2.10 (正交)**.** 给定内积空间 $\mathscr{V}$ 中的向量 $\mathbf{a},\mathbf{b}\in\mathscr{V}$，若 $(\mathbf{a}|\mathbf{b})=0$，则称 $\mathbf{a}$ 与 $\mathbf{b}$ 是正交的（*orthogonal*）。若 S 是 $\mathscr{V}$ 的一个子集，且 S 中的向量两两正交，则称 S 是 $\mathscr{V}$ 的一个正

*请验证它满足定义2.9。

交集（*orthogonal set*）。若 $\mathscr{V}$ 还是一个赋范向量空间，且 $\mathscr{V}$ 的一个正交集 S 中的向量均满足 $\|\hat{\mathbf{e}}\| = 1 \forall \hat{\mathbf{e}} \in S$，则称 S 是规范正交集（*orthonormal set*）。

易验，零向量与同一内积空间的任意向量都正交。

例 2.6. *[3]§7.2 例题 2.2 在数域 $\mathbb{F}$ 上的 $n \times n$ 矩阵的空间 $\mathbb{F}^{n\times n}$ 中，记 E^{pq} 为仅第 p 行、第 q 列为 1，其余为 0 的 $n \times n$ 矩阵。则由这 n^2 个矩阵 $E^{pq}, p = 1, \cdots, n, q = 1, \cdots, n$ 组成的集合是规范正交集。其中 $\mathbb{F}^{n\times n}$ 上的内积定义是 $(A|B) \equiv \sum_{j,k} A_{jk}\overline{B_{jk}}, \forall A, B \in \mathbb{F}^{n\times n}$。*

以下定理证明任意一组两两正交的非零向量线性无关。限定“非零”是因为，含零向量的任何向量组线性相关，但零向量又与任一向量正交。还需注意的是线性无关向量组未必都是两两正交的。

定理 2.4. 正交集中的所有非零向量线性无关。

证明. 设 $\mathscr{V}$ 是内积空间，S 是 $\mathscr{V}$ 的一个正交集，$\mathbf{a}_1, \cdots, \mathbf{a}_m \in S$。令 $\mathbf{b} = \beta_1\mathbf{a}_1 + \cdots + \beta_m\mathbf{a}_m$，则

$$
\begin{aligned}
(\mathbf{b}|\mathbf{a}_k) &= \left(\sum_j \beta_j \mathbf{a}_j \middle| \mathbf{a}_k\right) \\
&= \sum_j \beta_j \left(\mathbf{a}_j|\mathbf{a}_k\right) \\
&= \beta_j \left(\mathbf{a}_k|\mathbf{a}_k\right), k = 1, \cdots, m \\
\because (\mathbf{a}_k|\mathbf{a}_k) &\neq 0 \\
\therefore \beta_k &= \frac{(\mathbf{b}|\mathbf{a})}{\|\mathbf{a}\|^2}, k = 1, \cdots, m
\end{aligned}
$$

考察上式可验证 $\mathbf{b} = \mathbf{0} \Leftrightarrow \beta_1 = \cdots = \beta_m = 0$。 □

由定理2.4以及子空间、线性生成空间的定义（2.3、2.4），内积空间 $\mathscr{V}$ 的任一正交集 S 总能线性生成 $\mathscr{V}$ 的一个子空间。若内积空间 $\mathscr{V}$ 的一组基 B 是正交集，则称 B 为 $\mathscr{V}$ 的正交基（*orthogonal basis*）。如果 $\mathscr{V}$ 是赋范内积空间，其一组基 B 是规范正交集，则称 B 是 $\mathscr{V}$ 的一组规范正交基（*orthonormal basis*）。

定理2.4的证明也给出了由一个赋范内积空间的任一组基经格拉姆–施密特正交化过程*获得一组规范正交基的方法，作为定理如下。

*Gram–Schmidt process

定理 2.5. 设 $\mathscr{V}$ 是内积空间，$\mathbf{b}_1,\cdots,\mathbf{b}_n\in\mathscr{V}$ 是一组线性无关向量。那么总是可以由它们构建一组两两正交的向量 $\mathbf{a}_1,\cdots,\mathbf{a}_n\in\mathscr{V}$ 使得对于每一 $k=1,\cdots,n$，向量组 $\{\mathbf{a}_1,\cdots,\mathbf{a}_k\}$ 都是由 $\{\mathbf{b}_1,\cdots,\mathbf{b}_k\}$ 线性生成的空间的一组基。

证明. 采用数学归纳法。作为 $k=1$ 的情况，令 $\mathbf{a}_1=\mathbf{b}_1$，则命题显然成立。假设当 $k=m$ 时命题成立，即 $\{\mathbf{a}_1,\cdots,\mathbf{a}_m\},m<n$ 是已经构建好的满足命题要求的正交向量，则对每一 $k=1,\cdots,m$，$\{\mathbf{a}_1,\cdots,\mathbf{a}_k\}$ 是由 $\{\mathbf{b}_1,\cdots,\mathbf{b}_k\}$ 线性生成的子空间的正交基。令

$$\mathbf{a}_{m+1}=\mathbf{b}_{m+1}-\sum_{k=1}^{m}\frac{(\mathbf{b}_{m+1}|\,\mathbf{a}_k)}{(\mathbf{a}_k|\mathbf{a}_k)}\mathbf{a}_k$$

则有 $\mathbf{a}_{m+1}\neq\mathbf{0}$，否则 $\mathbf{b}_1,\cdots,\mathbf{b}_m,\mathbf{b}_{m+1}$ 线性相关，因 $\mathbf{b}_{m+1}$ 可由 $\mathbf{b}_1,\cdots,\mathbf{b}_m$ 线性表出。由上式还可知，对每一 $j=1,\cdots,m$ 均有

$$\begin{aligned}(\mathbf{a}_{m+1}|\mathbf{a}_j)&=(\mathbf{b}_{m+1}|\mathbf{a}_j)-\sum_{k=1}^{m}\frac{(\mathbf{b}_{m+1}|\mathbf{a}_k)}{(\mathbf{a}_k|\mathbf{a}_k)}(\mathbf{a}_k|\mathbf{a}_j)\\&=(\mathbf{b}_{m+1}|\mathbf{a}_j)-(\mathbf{b}_{m+1}|\mathbf{a}_j)\\&=0\end{aligned}$$

所以 $\{\mathbf{a}_1,\cdots,\mathbf{a}_{m+1}\}$ 是一个非零正交集。由定理2.4，它们线性无关。故 $\{\mathbf{a}_1,\cdots,\mathbf{a}_{m+1}\}$ 也是由 $\{\mathbf{b}_1,\cdots\mathbf{b}_{m+1}\}$ 线性生成的子空间的正交基。 □

特别地，对 $n=4$，

$$\begin{aligned}\mathbf{a}_1&=\mathbf{b}_1\\\mathbf{a}_2&=\mathbf{b}_2-\frac{(\mathbf{b}_2|\mathbf{a}_1)}{(\mathbf{a}_1|\mathbf{a}_1)}\mathbf{a}_1\\\mathbf{a}_3&=\mathbf{b}_3-\frac{(\mathbf{b}_3|\mathbf{a}_1)}{(\mathbf{a}_1|\mathbf{a}_1)}\mathbf{a}_1-\frac{(\mathbf{b}_3|\mathbf{a}_2)}{(\mathbf{a}_2|\mathbf{a}_2)}\mathbf{a}_2\\\mathbf{a}_4&=\mathbf{b}_4-\frac{(\mathbf{b}_4|\mathbf{a}_1)}{(\mathbf{a}_1|\mathbf{a}_1)}\mathbf{a}_1-\frac{(\mathbf{b}_4|\mathbf{a}_2)}{(\mathbf{a}_2|\mathbf{a}_2)}\mathbf{a}_2-\frac{(\mathbf{b}_4|\mathbf{a}_3)}{(\mathbf{a}_3|\mathbf{a}_3)}\mathbf{a}_3\end{aligned}$$

推论 2.5.1. 每个有限维赋范内积空间都有一组规范正交基。

证明. 只需要对采用格拉姆–施密特正交化过程得到的正交基，再对其基向量归一化即可。 □

例 2.7. 考虑 $\mathbb{R}^3$ 中的三个向量 $\mathbf{b}_1=(3,0,4)^\intercal,\mathbf{b}_2=(-1,0,7)^\intercal,\mathbf{b}_3=(2,9,11)^\intercal$。首先可以验证它们线性无关，即

$$\begin{cases}3x_1-x_2+2x_3&=0\\9x_3&=0\\4x_1+7x_2+11x_3&=0\end{cases}$$

只有唯一解 $x_1 = x_2 = x_3 = 0$。

通过格拉姆–斯密特正交化过程可由 $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ 得到 $\mathbf{R}^3$ 的一组正交基：

$$
\begin{aligned}
\mathbf{a}_1 &= (3,0,4)^\intercal \\
\mathbf{a}_2 &= (-1,0,7)^\intercal - \frac{(-1,0,7)^\intercal \cdot (3,0,4)^\intercal}{(3,0,4)^\intercal \cdot (3,0,4)^\intercal}(3,0,4)^\intercal \\
&= (-1,0,7)^\intercal - (3,0,4)^\intercal \\
&= (-4,0,3)^\intercal \\
\mathbf{a}_3 &= (2,9,11)^\intercal - \frac{(2,9,11)^\intercal \cdot (3,0,4)^\intercal}{(3,0,4)^\intercal \cdot (3,0,4)^\intercal}(3,0,4)^\intercal - \frac{(2,9,11)^\intercal \cdot (-4,0,5)^\intercal}{(-4,0,3)^\intercal \cdot (-4,0,3)^\intercal}(-4,0,3)^\intercal \\
&= (2,9,11)^\intercal - 2\,(3,0,4)^\intercal - (-4,0,3)^\intercal \\
&= (0,9,0)^\intercal
\end{aligned}
$$

可见 $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ 均为非零向量，故它们是 $\mathbb{R}^3$ 的一组正交基。归一化后得到 $\widehat{\mathbf{a}}_1 = \frac{1}{5}\mathbf{a}_1, \widehat{\mathbf{a}}_2 = \frac{1}{5}\mathbf{a}_2, \widehat{\mathbf{a}}_3 = (0,1,0)^\intercal$ 是一组规范正交基。任一向量 $\mathbf{x} = (x_1, x_2, x_3)^\intercal \in \mathbb{R}^3$ 在基 $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ 下的坐标为：

$$
(x_1, x_2, x_3)^\intercal = \frac{3x_1 + 4x_3}{25}\mathbf{a}_1 + \frac{-4x_1 + 3x_3}{25}\mathbf{a}_2 + \frac{x_2}{9}\mathbf{a}_3
$$

由格拉姆–斯密特正交化过程可知，一般地，对于规范正交基 $\{\mathbf{a}_i\}$ 有

$$
\mathbf{a}_i \cdot \mathbf{a}_j = \delta_{ij}
$$

其中

$$
\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}
$$

叫克劳内克符号（*Kronecker symbol*）。

在选取什么基之下，向量 $\mathbf{x} = (x_1, \cdots, x_n)$ 在该基下的坐标就恰好是 $x_1, \cdots, x_n$ 呢？这样的基 $\{\widehat{\mathbf{e}}_i\}$ 叫标准基（*standard basis*），其中

$$
\widehat{\mathbf{e}}_i = (e_{i,1}, \cdots, e_{i,n})^\intercal, e_{i,j} = \delta_{ij}
$$

标准基是一个规范正交基。

以下推导向量内积在给定基下的坐标计算公式。设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的 n 维内积空间，$\{\mathbf{e}_i\}$ 是 $\mathscr{V}$ 的一组基，则任意两个向量 $\mathbf{a}, \mathbf{b}$ 可表示为

$$
\mathbf{a} = \sum_{i=1}^{n} \alpha_i \mathbf{e}_i, \mathbf{b} = \sum_{i=1}^{n} \beta_i \mathbf{e}_i
$$

它们的内积

$$\begin{aligned}(\mathbf{a}|\mathbf{b}) &= \left(\sum_{i=1}^{n}\alpha_i\mathbf{e}_i\middle|\sum_{j=1}^{n}\beta_j\mathbf{e}_j\right)\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\overline{\beta_j}\left(\mathbf{e}_i|\mathbf{e}_j\right)\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\overline{\beta_j}G_{ij}\end{aligned}$$

其中，G 为基 $\{\mathbf{e}_i\}$ 的格拉姆矩阵（Gram matrix），即 $G_{ij}=(\mathbf{e}_i|\mathbf{e}_j)$，由内积定义有 $G_{ij}=\overline{G_{ji}}$。若 $\{\mathbf{e}_i\}$ 是正交基，则 $G_{ij}=\overline{G_{ji}}\delta_{ij}$。若 $\{\mathbf{e}_i\}$ 是规范正交基，则 $G_{ij}=\delta_{ij}$。

2.2 线性变换

2.2.1 线性变换的定义和基本性质

在本节我们考虑由一个向量空间到另一个向量空间的一种映射。我们先限定，建立了映射的两个向量空间是在同一数域上的。

向量空间是含有运算规则规定的集合。由一个向量空间到另一个向量空间的映射不一定能保持运算规则不变。所谓保持运算规则不变，意思就是定义域所在空间的运算法则，经过映射之后的陪域也具备。比如说，如果 $f(\alpha\mathbf{a}+\beta\mathbf{b})=\alpha f(\mathbf{a})+\beta f(\mathbf{b})$，则在一个向量空间上的映射 f 的陪域也是一个向量空间。这当然不是一定能保证的。如果两个同类代数结构之间的映射保持这类代数结构的关系定义不变，则称这种映射为这类代数结构上的同态映射（*homomorphism*）。

例 2.8. 以下映射是否向量空间上的同态映射？

- $f:\mathbb{R}\to\mathbb{R}:f(x)=x^2$ 讨论过程：$x\in\mathbb{R}$，视 $\mathbb{R}$ 为数域 $\mathbb{R}$ 上的向量空间，想要映射 f 是同态的，需要其满足 $f(\alpha x+y)=\alpha f(x)+f(y)$， $\forall x,y,\alpha\in\mathbb{R}$。显然，本例中的映射 f 并不满足。
- $g:\mathbb{R}^2\to\mathbb{R}:g(x,y)=x^2+y^2-4$
- $F:\mathbb{R}^2\to\mathbb{C}:F(x,y)=U(x,y)+iV(x,y),U,V:\mathbb{R}^2\to\mathbb{R}$
- $T:\mathbb{R}\to\mathbb{R}^2:T(t)=(t+3,2t-5)$
- 一个点粒子在空间中的运动可视为一个映射 $M:[a,b]\to\mathbb{R}^3$，具体定义为对每一 $t\in[a,b],M(t)=(x(t),y(t),z(t))$，其中 $x(t),y(t),z(t)$ 是 t 的实值函数。如果视 t 为时间，则 $M(t)$ 描述了粒子的运动轨迹，a 和 b 是运动的起始和终止时间。

定义 2.11 (线性变换). 设 $\mathscr{V}$ 和 $\mathscr{W}$ 是 $\mathbb{F}$ 上的向量空间。如果从 $\mathscr{V}$ 到 $\mathscr{W}$ 的映射 $T:\mathscr{V}\to\mathscr{W}$ 满足

$$T\left(\alpha\mathbf{a}+\beta\mathbf{b}\right)=\alpha T\left(\mathbf{a}\right)+\beta T\left(\mathbf{b}\right)$$

则称 T 是从 $\mathscr{V}$ 到 $\mathscr{W}$ 的线性变换（linear transformation），常写成算符的形式：$T\left(\mathbf{a}\right)=\mathbf{Ta}$。如果两个线性变换 $\mathbf{T}:\mathscr{V}\to\mathscr{W}$ 和 $\mathbf{U}:\mathscr{V}\to\mathscr{W}$ 满足 $\mathbf{Ta}=\mathbf{Ua}\forall\mathbf{a}\in\mathscr{V}$ 则称这两个线性变换相等，$\mathbf{U}=\mathbf{T}$。零变换 $\mathbf{T}_0:\mathscr{V}\to\mathscr{W}$ 定义为 $\mathbf{T}_0\mathbf{a}=\mathbf{0}_{\mathscr{W}}\forall\mathbf{a}\in\mathscr{V}$，其中 $\mathbf{0}_{\mathscr{W}}\in\mathscr{W}$ 表示 $\mathscr{W}$ 中的零向量*。

定义2.11就是在说线性变换是向量空间上的同态映射。由定义2.11，可直接证得以下结论：

- 任何一个线性变换都总把零向量映射为零向量。
- 两个向量空间之间的恒等映射是 $\mathscr{V}$ 到其自身的线性变换 $\mathbf{I}_{\mathscr{V}}:\mathscr{V}\to\mathscr{V},\mathbf{I}_{\mathscr{V}}\mathbf{a}=\mathbf{a}\forall\mathbf{a}\in\mathscr{V}$ 又可称为恒等线性变换。

正如由定义2.1所定义的向量那般，定义2.11定义的线性变换也是抽象概念。在例2.1中的向量空间上都可以建立线性变换。

例 2.9.
- *回顾课本《线性代数与解析几何》相关例题[3][§7.1 例题 1.4、§7.3 例题 3.3]。这些例题说明求导运算是线性的。*
- *设 $\mathscr{V}$ 是所有由 $\mathbb{R}$ 到 $\mathbb{R}$ 的连续函数的集合，验证 $\mathscr{V}$ 是实数域 $\mathbb{R}$ 上的向量空间。定义 $\mathscr{V}$ 上的映射 T，对任意 $\mathscr{V}$ 中的向量 f，$\left(Tf\right)\left(x\right)\equiv\int_0^x f\left(t\right)\mathrm{d}t,\forall x\in\mathbb{R},f\in\mathscr{V}$，验证 T 是由 $\mathscr{V}$ 到 $\mathscr{V}$ 的线性变换。这一例子说明积分运算是线性的。*

下面这个定理，使得线性变换本身也能形成一个向量空间（见推论）。

定理 2.6. 设 $\mathscr{V}$、$\mathscr{W}$ 是 $\mathbb{F}$ 上的向量空间，$\mathbf{T}$、$\mathbf{U}$ 是从 $\mathscr{V}$ 到 $\mathscr{W}$ 的线性变换。若定义

$$\left(\mathbf{T}+\mathbf{U}\right)\mathbf{a}=\mathbf{Ta}+\mathbf{Ua},\forall\mathbf{a}\in\mathscr{V}$$

$$\left(\alpha\mathbf{T}\right)\mathbf{a}=\alpha\left(\mathbf{Ta}\right),\forall\alpha\in\mathbb{F}$$

则 $\left(\mathbf{T}+\mathbf{U}\right)$ 和 $\alpha\mathbf{T}$ 也是从 $\mathscr{V}$ 到 $\mathscr{W}$ 的线性变换。

证明. 分别用 $\mathbf{T}+\mathbf{U}$ 和 $\alpha\mathbf{T}$ 作用于向量 $\beta\mathbf{a}+\gamma\mathbf{b},\alpha,\beta\in\mathbb{F},\mathbf{a},\mathbf{b}\in\mathscr{V}$，使用向量空间的定义2.1、线性变换的定义2.11和本命题中的运算定义证明。略。 □

推论 2.6.1. 给定数域 $\mathbb{F}$ 上的两个向量空间 $\mathscr{V},\mathscr{W}$，所有由 $\mathscr{V}$ 到 $\mathscr{W}$ 的线性变换的集合是一个数域 $\mathbb{F}$ 上的向量空间，记为 $\mathscr{L}\left(\mathscr{V},\mathscr{W}\right)$，零变换是该向量空间的零向量。

*此处要注意区分不同向量空间中的零向量。

证明. 前一句由定理2.6易证。设 $\mathbf{T}_0$ 是由 $\mathscr{V}$ 到 $\mathscr{W}$ 的零变换，对任一 $\mathbf{T} \in \mathscr{L}(\mathscr{V}, \mathscr{W})$ 和 $\mathbf{a} \in \mathscr{V}$，$(\mathbf{T}_0 + \mathbf{T})\,\mathbf{a} = \mathbf{T}_0\mathbf{a} + \mathbf{T}\mathbf{a} = \mathbf{0}_{\mathscr{W}} + \mathbf{T}\mathbf{a} = \mathbf{T}\mathbf{a}$，即 $\mathbf{T}_0 + \mathbf{T} = \mathbf{T} \forall \mathbf{T} \in \mathscr{L}(\mathscr{V}, \mathscr{W})$。 □

请读者按照向量空间的定义验证，线性变换的空间 $\mathscr{L}(\mathscr{V}, \mathscr{W})$ 作为一个向量空间，拥有一切向量空间的一般性质。注意：在记法 $\mathscr{L}(\mathscr{V}, \mathscr{W})$ 中，括号表示有序对。

接下来，我们先考察 $\mathscr{L}(\mathscr{V}, \mathscr{W})$ 的基和维数。不关心证明的读者可直接朗读定理2.7。

引理 2.1. 设 $\mathscr{V}_N$ 是数域 $\mathbb{F}$ 上的 N 维向量空间，$B_{\mathscr{V}} = \{\mathbf{e}_i\}_{i=1}^N$ 是 $\mathscr{V}_N$ 的一组基，$\mathscr{W}$ 是同数域上的另一向量空间。对 $\mathscr{W}$ 中给定的任意一组 N 个不同向量 $\{\mathbf{b}_i\}_{i=1}^N \in \mathscr{W}$，有且只有一个线性变换 $\mathbf{T} : \mathscr{V}_N \to \mathscr{W}$ 满足 $\mathbf{T}\mathbf{e}_i = \mathbf{b}_i, i = 1, \cdots, N$。

证明. 存在性的证明，只需找出这一线性变换。任一 $\mathbf{a} \in \mathscr{V}_N$ 可用基 $B_{\mathscr{V}}$ 表示成

$$\mathbf{a} = \sum_{i=1}^{N} \alpha_i \mathbf{e}_i, \alpha_i \in \mathbb{F}, i = 1, \cdots, N$$

特别地，

$$\mathbf{e}_i = \sum_{j=1}^{N} \delta_{ij} \mathbf{e}_j, i = 1, \cdots, N$$

定义映射 $\mathbf{T} : \mathscr{V}_N \to \mathscr{W}, \mathbf{T}\mathbf{a} \equiv \sum_{i=1}^N \alpha_i \mathbf{b}_i$，则易验 $\mathbf{T}\mathbf{e}_i = \mathbf{b}_i, i = 1, \cdots, N$。我们还需验证映射 $\mathbf{T}$ 是线性变换，按定义2.11只需验证 $T(\alpha\mathbf{a} + \beta\mathbf{b}) = \alpha(\mathbf{T}\mathbf{a}) + \beta(\mathbf{T}\mathbf{b})$ 即可。给定任意 $\mathbf{b} \in \mathscr{V}_N$ 和标量 $\gamma \in \mathbb{F}$，其中 $\mathbf{b}$ 可表示为 $\mathbf{b} = \sum_{i=1}^N \beta_i \mathbf{e}_i$，且有

$$\gamma\mathbf{a} + \mathbf{b} = \sum_{i=1}^{N} (\gamma\alpha_i + \beta_i)\,\mathbf{e}_i$$

由 $\mathbf{T}$ 的定义有

$$\mathbf{T}(\gamma\mathbf{a} + \mathbf{b}) = \sum_{i=1}^{N} (\gamma\alpha_i + \beta_i)\,\mathbf{b}_i \tag{2.1}$$

$$\gamma(\mathbf{T}\mathbf{a}) + \mathbf{T}\mathbf{b} = \gamma \sum_{i=1}^{N} \alpha_i \mathbf{b}_i + \sum_{i=1}^{N} \beta_i \mathbf{b}_i \tag{2.2}$$

$$= \sum_{i=1}^{N} (\gamma\alpha_i + \beta_i)\,\mathbf{b}_i \tag{2.3}$$

所以 $\mathbf{T}(\gamma\mathbf{a} + \mathbf{b}) = \gamma(\mathbf{T}\mathbf{a}) + \mathbf{T}\mathbf{b}$ 对任意向量 $\mathbf{a}, \mathbf{b} \in \mathscr{V}_N$ 和标题 $\gamma \in \mathbb{F}$ 均成立，即 $\mathbf{T}$ 是一个线性变换。存在性证毕。

唯一性的证明，设另有一线性变换 $\mathbf{U} : \mathscr{V}_N \to \mathscr{W}$ 也满足 $\mathbf{U}\mathbf{e}_i = \mathbf{b}_i$，则对任一 $\mathbf{a} \in \mathscr{V}_N$，

$$\mathbf{U}\mathbf{a} = \mathbf{U}\left(\sum_{i=1}^{N} \alpha_i \mathbf{e}_i\right) = \sum_{i=1}^{N} \alpha_i \mathbf{U}\mathbf{e}_i = \sum_{i=1}^{N} \alpha_i \mathbf{b}_i$$

即 $\mathbf{U}$ 就是 $\mathbf{T}$。 □

定理 2.7. 若 $\mathscr{V}_N, \mathscr{W}_M$ 分别为数域 $\mathbb{F}$ 上的 N, M 维向量空间，则 $\mathscr{L}(\mathscr{V}_N, \mathscr{W}_M)$ 的维数是 $M \times N$。

证明. 证明的过程，相当于找出 $\mathscr{L}(\mathscr{V}_N, \mathscr{W}_M)$ 的基*

设 $B_{\mathscr{V}} = \{\mathbf{e}_i\}_{i=1}^N, B_{\mathscr{W}} = \{\mathbf{f}_j\}_{j=1}^M$ 分别为 $\mathscr{V}_N, \mathscr{W}_M$ 的基。对于每一对整数 $(p,q), 1 \le p \le N, 1 \le q \le M$，定义一个线性变换 $\mathbf{E}^{pq} : \mathscr{V}_N \to \mathscr{W}_M$

$$\mathbf{E}^{pq}\mathbf{e}_i = \begin{cases} \mathbf{0}_{\mathscr{W}}, & i \neq p \\ \mathbf{f}_q, & i = p \end{cases}$$

给定任一线性变换 $\mathbf{T} \in \mathscr{L}(\mathscr{V}_N, \mathscr{W}_M)$，且令

$$\mathbf{T}\mathbf{e}_i = \sum_{q=1}^{M} A_{qi}\mathbf{f}_q$$

由引理2.1，对 $\{\mathbf{f}_j\}_{j=1}^M$，满足上式的 $\mathbf{T}$ 是唯一的，且根据向量的坐标表达式，上式中的 A_{qi} 就是向量 $\mathbf{T}\mathbf{e}_i$ 在有序基 $B_{\mathscr{W}}$ 下的坐标。

下面我们证明 $\mathbf{T}$ 能被 $\{\mathbf{E}^{pq}\}$ 线性表出。定义 $\mathbf{U} \in \mathscr{L}(\mathscr{V}_N, \mathscr{W}_M)$，$\mathbf{U}\mathbf{e}_i = \sum_p \sum_q A_{qp}\mathbf{E}^{pq}\mathbf{e}_i$，则

$$\begin{aligned} \mathbf{U}\mathbf{e}_i &= \sum_p \sum_q A_{qp}\mathbf{E}^{pq}\mathbf{e}_i \\ &= \sum_p \sum_q A_{qp}\delta_{ip}\mathbf{f}_q \\ &= \sum_q A_{qi}\mathbf{f}_q \\ &= \mathbf{T}\mathbf{e}_i \end{aligned}$$

则可见 $\mathbf{U} = \mathbf{T}$，同时有

$$\mathbf{T} = \sum_{p=1}^{N}\sum_{q=1}^{M} A_{qp}\mathbf{E}^{pq}$$

由于 $\mathbf{T}$ 是 $\mathscr{L}(\mathscr{V}_N, \mathscr{W}_M)$ 中的任意一个线性变换，故 $\mathscr{L}(\mathscr{V}_N, \mathscr{W}_M)$ 是 $\{\mathbf{E}^{pq}\}$ 的线性生成空间。由基的定义2.5，现在我们只需再证 $\{\mathbf{E}^{pq}\}$ 是线性无关的，它们就是 $\mathscr{L}(\mathscr{V}_N, \mathscr{W}_M)$ 的基了。按照线性无关的定义，设 $\mathbf{T}$ 是零变换，由于 $\{\mathbf{f}_j\}_{j=1}^M$ 是线性无关的，故有

$$\mathbf{T} = \mathbf{0} = \sum_{p=1}^{N}\sum_{q=1}^{M} A_{qp}\mathbf{E}^{pq} \Leftrightarrow \mathbf{T}\mathbf{e}_i = \sum_q A_{qi}\mathbf{f}_q = \mathbf{0}_{\mathscr{W}} \Leftrightarrow A_{qi} = 0 \forall q = 1, \cdots, M, i = 1, \cdots, N$$

即 $\{\mathbf{E}^{pq}\}$ 线性无关。故 $\{\mathbf{E}^{pq}\}$ 是 $\mathscr{L}(\mathscr{V}_N, \mathscr{W}_M)$ 的一组基，$\mathscr{L}(\mathscr{V}_N, \mathscr{W}_M)$ 的维数就是 $M \times N$。 □

*我们在例2.6中找过数域 $\mathbb{F}$ 上的 $n \times n$ 矩阵空间的一组规范正交基 $\{E^{pq}\}_{p,q=1}^n$。定理2.7的证明过程与例2.6很类似。

注意，给定两个有限维向量空间 $\mathscr{V},\mathscr{W}$，$\mathscr{L}(\mathscr{V},\mathscr{W})$ 中的线性变换和 $\mathscr{L}(\mathscr{W},\mathscr{V})$ 中的线性变换是不同空间的元素，没有一般关系——虽然它们的维数无非只是 $M\times N$ 和 $N\times M$ 的差别。

讨论完线性变换本身作为向量的性质——它所在向量空间的基与维数，我们接下来转而考虑线性变换作为映射的性质：单射、满射、双射……。线性变换作为映射的性质与它所映射的两个向量空间的维数有关，这是线性代数最重要的定理之一——线性变换的维数定理。下面我们先为该定理的引入介绍一些必要概念。

定义 2.12 (零空间、零化度、秩)**.**（如图2.1所示）设 $\mathscr{V}$ 和 $\mathscr{W}$ 是数域 $\mathbb{F}$ 上的向量空间。线性变换 $\mathbf{T}:\mathscr{V}\to\mathscr{W}$ 的零空间（*null space*）或核空间（*kernel*）是所有满足 $\mathbf{Ta}=\mathbf{0}_{\mathscr{W}}$ 的向量 $\mathbf{a}$ 的集合，记作 $\ker\mathbf{T}$。零空间的维数称为该线性变换的零化度（*nullity*），记作 $\text{nullity}\mathbf{T}\equiv\dim(\ker\mathbf{T})$。如果 $\mathscr{V}$ 是有限维向量空间，则 $\mathbf{T}$ 的秩（*rank*）是 $\mathbf{T}$ 的值域的维数，记作 $\text{rank}\mathbf{T}\equiv\dim(\text{ran}\mathbf{T})$。特别地，如果 $\text{nullity}\mathbf{T}=0$，即 $\mathbf{T}$ 的零空间只有零向量 $\mathbf{0}_{\mathscr{V}}$ 一个元素，则称 $\mathbf{T}$ 是非奇异的（*non-singular*）。

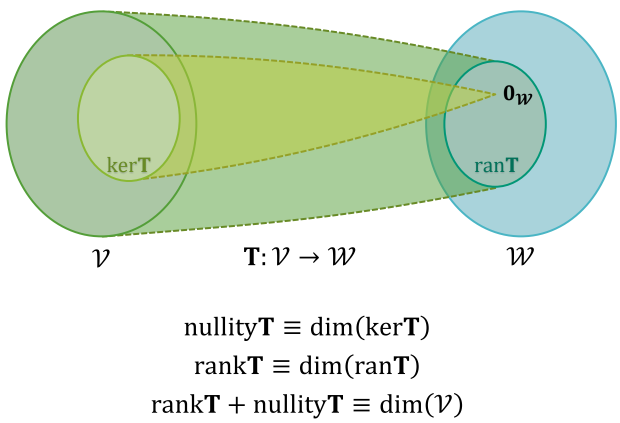


图 2.1: 线性变换的零空间、值域、零化度和秩以及线性变换的维数定理

上述定义默认了两个易验事实的成立：$\ker\mathbf{T}$ 是 $\mathscr{V}$ 的子空间，$\text{ran}\mathbf{T}$ 是 $\mathscr{W}$ 的子空间（否则不能直接谈它这两个子集的“维数”），证明从略[3]§7.3 “2. 线性变换的简单性质 (4)”。

基于定义2.12的概念，我们给出线性代数中非常重要的定理——线性变换的维数定理*。

定理 2.8 (线性变换的维数定理)**.** 设 $\mathscr{V}$ 和 $\mathscr{W}$ 是数域 $\mathbb{F}$ 上的向量空间，其中 $\mathscr{V}$ 是有限维向量空间，则对线性变换 $\mathbf{T}:\mathscr{V}\to\mathscr{W}$ 有

$$\text{rank}\mathbf{T}+\text{nullity}\mathbf{T}=\dim\mathscr{V}$$

*[3]§7.3 “2. 线性变换的简单性质 (5)”

证明. 设 $\dim\mathscr{V}=n$，$\mathrm{nullity}\mathbf{T}=\dim(\ker\mathbf{T})=k$，$\{\mathbf{a}_1,\cdots,\mathbf{a}_k\}$ 是 $\ker\mathbf{T}$ 的一组基。则可在 $\mathscr{V}$ 中继续找到 $\{\mathbf{a}_{k+1},\cdots,\mathbf{a}_n\}$ 与 $\{\mathbf{a}_1,\cdots,\mathbf{a}_k\}$ 合并成线性无关向量组，从而成为 $\mathscr{V}$ 的基。

由于 $\{\mathbf{Ta}_1,\cdots,\mathbf{Ta}_n\}$ 线性生成 $\mathbf{T}$ 的值域 $\mathrm{ran}\mathbf{T}$（由线性变换性质易证），其中由于

$$\{\mathbf{a}_1,\cdots,\mathbf{a}_k\}\in\ker\mathbf{T}$$

故有

$$\mathbf{Ta}_j=\mathbf{0}_{\mathscr{W}}\forall j\leq k$$

所以实际上仅 $\{\mathbf{Ta}_{k+1},\cdots,\mathbf{Ta}_n\}$ 就线性生成 $\mathrm{ran}\mathbf{T}$ 了。我们进一步验证它们是线性无关的。设标量 γ_i 满足

$$\sum_{i=k+1}^{n}\gamma_i\mathbf{Ta}_i=\mathbf{0}_{\mathscr{W}}$$

即

$$\mathbf{T}\left(\sum_{i=k+1}^{n}\gamma_i\mathbf{a}_i\right)=\mathbf{0}_{\mathscr{W}}$$

即 $\sum_{i=k+1}^{n}\gamma_i\mathbf{a}_i\equiv\mathbf{a}\in\mathscr{N}$。向量 $\mathbf{a}$ 在 $\ker\mathbf{T}$ 的基 $\{\mathbf{a}_1,\cdots,\mathbf{a}_k\}$ 下表示为 $\mathbf{a}=\sum_{i=1}^{k}\alpha_i\mathbf{a}_i$。故

$$\mathbf{a}=\sum_{i=1}^{k}\alpha_i\mathbf{a}_i=\sum_{j=k+1}^{n}\gamma_j\mathbf{a}_j$$

由于 $\{\mathbf{a}_1,\cdots,\mathbf{a}_n\}$ 是线性无关的，故有且只有 $\alpha_1=\cdots=\alpha_k=\gamma_{k+1}=\cdots=\gamma_n=0$，即 $\{\mathbf{Ta}_{k+1},\cdots,\mathbf{Ta}_n\}$ 是线性无关的。因此 $\{\mathbf{Ta}_{k+1},\cdots,\mathbf{Ta}_n\}$ 是 $\mathscr{W}$ 的基，即 $\mathscr{W}$ 的维数是 $n-k$，由定义2.12即 $\mathrm{rank}\mathbf{T}=n-k$。 □

线性变换的维数定理是我们继续讨论线性变换的映射性质的重要定理。利用它，我们首先可以得到以下命题相互等价（即只要其中一条成立，剩余皆成立）。

定理 2.9. 设 $\mathscr{V},\mathscr{W}$ 是数域 $\mathbb{F}$ 上的有限维向量空间，$\mathbf{T}:\mathscr{V}\to\mathscr{W}$ 是一个线性变换，则以下命题互相等价：

1. $\mathbf{T}$ 是非奇异的；
2. $\mathbf{T}$ 是单射；
3. $\mathbf{T}$ 将 $\mathscr{V}$ 中的任意一组线性无关向量组映射为 $\mathscr{W}$ 的一组线性无关向量组；
4. $\mathrm{rank}\mathbf{T}=\dim\mathscr{V}$。

证明. $1\Leftrightarrow 2$：设 $\mathbf{T}$ 是非奇异的（即有 $\mathbf{Ta}=\mathbf{0}_{\mathscr{W}}\Leftrightarrow\mathbf{a}=\mathbf{0}_{\mathscr{V}}\forall\mathbf{a}\in\mathscr{V}$），则对任意 $\mathbf{a},\mathbf{b}\in\mathscr{V}$，$\mathbf{Ta}=\mathbf{Tb}\Leftrightarrow\mathbf{T}(\mathbf{a}-\mathbf{b})=\mathbf{0}_{\mathscr{W}}\Leftrightarrow\mathbf{a}-\mathbf{b}=\mathbf{0}_{\mathscr{V}}\Leftrightarrow\mathbf{a}=\mathbf{b}$。

$1 \Leftrightarrow 3$：设 $\mathbf{T}$ 是非奇异的。令 S 是 $\mathscr{V}$ 的一个线性无关向量组，如果向量 $\mathbf{a}_1, \cdots, \mathbf{a}_k \in S$，则

$$\sum_{i=1}^{k} c_i (\mathbf{T}\mathbf{a}_i) = \mathbf{0}_{\mathscr{W}} \Leftrightarrow \mathbf{T}\sum_{i=1}^{k} c_i\mathbf{a}_i = \mathbf{0}_{\mathscr{W}} \Leftrightarrow \sum_{i=1}^{k} c_i\mathbf{a}_i = \mathbf{0}_{\mathscr{W}} \Leftrightarrow c_i = 0 \forall i$$

假设 $\mathbf{T}$ 总把 $\mathscr{V}$ 的一组线性无关向量映射为 $\mathscr{W}$ 的一组线性无关向量，令 $\mathbf{a} \neq \mathbf{0}_{\mathscr{V}}$ 是 $\mathscr{V}$ 的一个非零向量，则只含 $\mathbf{a}$ 的向量组 $\{\mathbf{a}\}$ 是线性无关向量组，这时 $\mathbf{T}\mathbf{a}$ 必不为 $\mathbf{0}_{\mathscr{W}}$，因为单一个零向量的向量组是线性相关的，违反了假设。因此 $\mathbf{T}$ 只能把 $\mathbf{0}_{\mathscr{V}}$ 映射为 $\mathbf{0}_{\mathscr{W}}$，即为非奇异的。

$1 \Leftrightarrow 4$：由定理2.8可直接得到。 □

定理2.9联系了一个线性变换的非奇异性与其单射性。下面的推论进一步说明，如果在此基础上再加上一个条件：$\dim\mathscr{V} = \dim\mathscr{W}$，那么 $\mathbf{T}$ 要么是双射，要么是非单射非满射。

推论 2.9.1. 设 $\mathscr{V}, \mathscr{W}$ 是数域 $\mathbb{F}$ 上的有限维向量空间且 $\dim\mathscr{V} = \dim\mathscr{W}$，$\mathbf{T} : \mathscr{V} \to \mathscr{W}$ 是一个线性变换，则以下命题相互等价：

1. $\mathbf{T}$ 是单射；
2. $\mathbf{T}$ 是满射；
3. $\mathbf{T}$ 将 $\mathscr{V}$ 中的任意一组基映射为 $\mathscr{W}$ 的一组基；

证明. $1 \Leftrightarrow 2$：设 $n = \dim\mathscr{V} = \dim\mathscr{W}$，则由定理2.1的推论 3、定理2.8和2.9，$\mathbf{T}$ 是单射 $\Leftrightarrow \mathrm{rank}\mathbf{T} = n = \dim\mathscr{W} \Leftrightarrow \mathrm{ran}\mathbf{T} = \mathscr{W}$。

$1 \Leftrightarrow 3$：留作练习。 □

由定理1.2，作为双射的线性变换必存在唯一逆映射，那么这个逆映射会不会也是一个线性变换呢？答案是肯定的，即如下定理。

定理 2.10. 设 $\mathscr{V}, \mathscr{W}$ 是数域 $\mathbb{F}$ 上的向量空间，$\mathbf{T} : \mathscr{V} \to \mathscr{W}$ 是一个线性变换。如果 $\mathbf{T}$ 可逆，则其逆映射 $\mathbf{T}^{-1}$ 是一个由 $\mathscr{W}$ 到 $\mathscr{V}$ 的线性变换。

证明. 对任意 $\mathbf{b}_1, \mathbf{b}_3 \in \mathscr{W}$ 和 $\gamma \in \mathbb{F}$，令 $\mathbf{a}_i = \mathbf{T}^{-1}\mathbf{b}_i, i = 1, 2$，由于 $\mathbf{T}$ 是线性变换，则有 $\mathbf{T}(\gamma\mathbf{b}_1 + \mathbf{b}_2) = \gamma\mathbf{T}\mathbf{a}_1 + \mathbf{T}\mathbf{a}_2 = \gamma\mathbf{b}_1 + \mathbf{b}_2$。因此向量 $\gamma\mathbf{a}_1 + \mathbf{a}_2 \in \mathscr{V}$ 就是向量 $\gamma\mathbf{b}_1 + \mathbf{b}_2 \in \mathscr{W}$ 在映射 $\mathbf{T}$ 下的原像，由逆映射性质有 $\mathbf{T}^{-1}(\gamma\mathbf{b}_1 + \mathbf{b}_2) = \gamma\mathbf{a}_1 + \mathbf{a}_2 = \gamma(\mathbf{T}^{-1}\mathbf{b}_1) + \mathbf{T}^{-1}\mathbf{b}_2$，即 $\mathbf{T}^{-1}$ 满足线性变换定义的性质（对任意选取的 $\mathbf{b}_1, \mathbf{b}_2, \gamma$ 均成立），故 $\mathbf{T}^{-1}$ 是线性变换。 □

双射（可逆）线性变换，经常称作同构（*isomorphic*）线性变换。如果两个向量空间之间存在一个同构线性变换，则称这两个向量空间是同构（*isomorphic*）的*。在 §二中我们已了解，

*请同一形容词的修饰对象的区别：形容一个线性变换，还是跟形容两个向量空间。

一个向量与其在给定某有序基下的坐标是一一对应的，且易证由一个向量到其坐标的关系在给定有序基时是一个同构线性变换。因此，任一数域 $\mathbb{F}$ 上的 n 维向量空间均与 n 维坐标空间 $\mathbb{F}^n$ 同构。由于双射的可逆性具有传递性（即当 $f : A \to B$ 是双射、$g : B \to C$ 是双射，则 $g \circ f : A \to C$ 也是双射*），故所有同维数的向量空间之间都相互同构。以下定理及其推论则从另一个不同的出发点证明了这上述的结论。

定理 2.11. 数域 $\mathbb{F}$ 上的两个向量空间同构，当且仅当它们维数相等。

证明. 由"两个向量空间同构"的定义，需要证明两个维数相同的向量空间中存在一个双射线性变换。设 $\mathscr{V}_N$、$\mathscr{W}_N$ 是数域 $\mathbb{F}$ 上的两个 N 维向量空间。分别给定 $\mathscr{V}_N, \mathscr{W}_N$ 的各一组有序基 $B_{\mathscr{V}} = \{\mathbf{e}_i\}_{i=1}^N, B_{\mathscr{W}} = \{\mathbf{f}_i\}_{i=1}^N$，构建一个线性变换 $\mathbf{T} : \mathscr{V}_N \to \mathscr{W}_N$，对 $\mathscr{V}_N$ 中的任一向量 $\mathbf{a} = \sum_{i=1}^N \alpha_i \mathbf{e}_i$，$\mathbf{Ta} = \sum_{i=1}^N \alpha_i \mathbf{f}_i$。易验 $\mathbf{a} = \mathbf{0}_{\mathscr{V}} \Leftrightarrow \mathbf{Ta} = \mathbf{0}_{\mathscr{W}}$，即 $\mathbf{T}$ 是单射。由定理2.9及其推论可知 $\mathbf{T}$ 是双射。 □

推论 2.11.1. 任一数域 $\mathbb{F}$ 上的 N 维向量空间均与 $\mathbb{F}^N$ 同构

按照逆映射的性质，一个同构线性变换与其逆映射的复合映射必是恒等映射。这个映射能否进一步是一个线性变换？我们于是先要问：一般地，两个线性变换形成的复合映射是线性变换吗？——答案是肯定的，即以下定理。

定理 2.12. 设 $\mathscr{V}, \mathscr{W}, \mathscr{Z}$ 是数域 $\mathbb{F}$ 上的向量空间，$\mathbf{T} : \mathscr{V} \to \mathscr{W}, \mathbf{U} : \mathscr{W} \to \mathscr{Z}$ 是线性变换，则复合映射 $\mathbf{U} \circ \mathbf{T}$（记为 $\mathbf{UT}$）也是线性变换。它是由 $\mathscr{V}$ 到 $\mathscr{Z}$ 的线性变换。

证明. 用 $\mathbf{UT}$ 作用于 $\gamma \mathbf{a} + \mathbf{b}$，其中 $\mathbf{a}, \mathbf{b} \in \mathscr{V}, \gamma \in \mathbb{F}$ 是任意的，证明 $\gamma (\mathbf{UT}) \mathbf{a} + (\mathbf{UT}) \mathbf{b}$ 即可，留作练习。 □

由此定理易得推论：向量空间上的恒等映射必是一个同构线性变换，我们称之为恒等线性变换或恒等变换，记为 $\mathbf{I}_{\mathscr{V}}$，其中 $\mathscr{V}$ 是这个恒等变换所作用的向量空间。通过验证下例，可以注意到恒等线性变换需要注明所作用的向量空间；作用于不同向量空间上的恒等变换是不同的东西。

例 2.10. 设 $\mathscr{V}, \mathscr{W}$ 是数域 $\mathbb{F}$ 上的向量空间，$\mathbf{T} : \mathscr{V} \to \mathscr{W}$ 是同构线性变换。则 $\mathbf{T}^{-1}\mathbf{T} = \mathbf{I}_{\mathscr{V}}, \mathbf{T}\mathbf{T}^{-1} = \mathbf{I}_{\mathscr{W}}$。

我们实际经常面临的是从一个向量空间到其自身的线性变换，故我们特别给这种线性变换起个名字，正式定义如下。

*回顾定义1.2。

定义 2.13 (线性算符). 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的向量空间，由 $\mathscr{V}$ 到其自身的线性变换 $\mathbf{T}:\mathscr{V}\to\mathscr{V}$ 称为线性算符（*linear operator*）。

之前说过，线性变换是向量空间之间的同态映射，故线性算符又可称为自同态（*automorphic*）线性变换。前面也提到过，双射线性变换叫同构线性变换，故双射线性算符又称自同构（*endomorphic*）线性算符。由定理2.9的推论，线性算符要么是非单射非满射，要么是双射。也就是说，并非所有线性算符都是可逆映射；只有是双射（即自同构）线性算符才是可逆映射。另外，自同构线性算符与其逆映射同属于一个线性变换的空间 $\mathscr{L}(\mathscr{V},\mathscr{V})$（简写为 $\mathscr{L}(\mathscr{V})$）。

在本节刚开始我们了解了线性变换作为向量的性质。现在又了解了线性变换作为映射的一些性质。具体对于线性算符，我们可以通过关于复合线性变换的定理2.12，为线性算符引入“乘法”的运算*。作为线性算符的恒等变换记号 $\mathbf{I}$ 无需指明作用空间。以下定理具体给出了线性算符之间的乘法所满足的性质。

定理 2.13. 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的向量空间，$\mathbf{U},\mathbf{T}_1,\mathbf{T}_2\in\mathscr{L}(\mathscr{V})$，$\gamma\in\mathbb{F}$，则有：

1. $\mathbf{IU}=\mathbf{UI}=\mathbf{U}$
2. $\mathbf{U}(\mathbf{T}_1+\mathbf{T}_2)=\mathbf{UT}_1+\mathbf{UT}_2$，$(\mathbf{T}_1+\mathbf{T}_2)\mathbf{U}=\mathbf{T}_1\mathbf{U}+\mathbf{T}_2\mathbf{U}$
3. $\gamma\mathbf{UT}_1=(\gamma\mathbf{U})\mathbf{T}_1=\mathbf{U}(\gamma\mathbf{T}_1)$

证明. 第 1 条由相关定义是易证的。

对任一向量 $\mathbf{a}\in\mathscr{V}$，

$$\begin{aligned}[\mathbf{U}(\mathbf{T}_1+\mathbf{T}_2)]\mathbf{a}&=\mathbf{U}[(\mathbf{T}_1+\mathbf{T}_2)\mathbf{a}]\quad\text{（由定理2.6中定义的运算法则）}\\&=\mathbf{U}(\mathbf{T}_1\mathbf{a}+\mathbf{T}_2\mathbf{a})\quad\text{（}\mathbf{U}\text{ 是线性变换）}\\&=(\mathbf{UT}_1)\mathbf{a}+(\mathbf{UT}_2)\mathbf{a}\quad\text{（由复合映射的定义）}\end{aligned}$$

故由两映射相等的定义，$\mathbf{U}(\mathbf{T}_1+\mathbf{T}_2)=\mathbf{UT}_1+\mathbf{UT}_2$。类似地，

$$\begin{aligned}[(\mathbf{T}_1+\mathbf{T}_2)\mathbf{U}]\mathbf{a}&=(\mathbf{T}_1+\mathbf{T}_2)(\mathbf{Ua})\quad\text{（由复合映射的定义）}\\&=\mathbf{T}_1(\mathbf{Ua})+\mathbf{T}_2(\mathbf{Ua})\quad\text{（由定理2.6中定义的运算法则）}\\&=(\mathbf{T}_1\mathbf{U})\mathbf{a}+(\mathbf{T}_2\mathbf{U})\mathbf{a}\quad\text{（由复合映射的定义）}\end{aligned}$$

故由两映射相等的定义，$(\mathbf{T}_1+\mathbf{T}_2)\mathbf{U}=\mathbf{T}_1\mathbf{U}+\mathbf{T}_2\mathbf{U}$，第 2 条证毕†。第 3 条证明留作练习。 □

*为何要在线性算符之间才能引入“乘法”？直接按照定理2.12给一般的线性变换定义“乘法”会有什么问题？

†我们留意到第 2 条的第一部分证明没有用到 $\mathbf{T}_1$ 和 $\mathbf{T}_2$ 是线性变换的条件，第二部分的证明连 $\mathbf{U}$ 是线性变换的条件都没用到。

2.2.2　线性变换的坐标矩阵

在之前的内容里，我们主要介绍了向量和线性变换的代数定义。在以往的《线性代数》课里我们学习“向量”都是 $n\times 1$ 矩阵（“列向量”）或 $1\times n$ 矩阵（“行向量”），实际上这些只是 3 维实坐标空间 $\mathbb{R}^3$ 空间的向量。在 §二中我们明确了，要在一个数域 $\mathbb{F}$ 上的 n 维向量空间 $\mathscr{V}$ 与 n 维坐标空间 $\mathbb{F}^n$ 之间建立一一对应关系，可以通过选定 $\mathscr{V}$ 的某组基，把 $\mathscr{V}$ 中的向量表示成该基下的坐标——$\mathbb{F}^n$ 中的一个 n 元组。于是，这一一一对应关系是依赖 $\mathscr{V}$ 的基的选择的。在未指定基的时候，不能直接用 $\mathbb{F}^n$ 中的一个 n 元数组指代 $\mathscr{V}$ 中的一个向量。

我们还知道，线性变换本身也是一个向量；线性变换的空间也有基和维数，因此相应地也应该有选定基下的坐标。下面我们将会看到，线性变换在选定基下的坐标可以表示为一个矩阵[3]§7.3 “3” 。

考虑数域 $\mathbb{F}$ 上的 n 维向量空间 $\mathscr{V}_n$ 的一组有序基 $B=(\mathbf{a}_i)_{i=1}^n$，按照向量的坐标表达式，任一向量 $\mathbf{x}\in\mathscr{V}$ 均可表示成 $\mathbf{x}=\sum_{i=1}^n\xi_i\mathbf{a}_i$，其中 $\xi_i\in\mathbb{F}, i=1,\cdots,n$。向量 $\mathbf{x}$ 在有序基 B 下的坐标可表示为由 $\{\xi_i\}$ 组成的 $n\times 1$ 矩阵 $(\xi_1,\cdots,\xi_n)^\intercal$。

设 $\mathscr{V}_n,\mathscr{W}_m$ 分别是数域 $\mathbb{F}$ 上的 n、m 维向量空间，且线性变换 $\mathbf{A}\in\mathscr{L}(\mathscr{V}_n,\mathscr{W}_m)$ 将 $\mathscr{V}_n$ 的一组基向量 $B_{\mathscr{V}}=\{\mathbf{a}_i\}_{i=1}^n$ 映射到 $\mathscr{W}_m$ 中的 n 个向量 $\mathbf{w}_k=\mathbf{A}\mathbf{a}_k, k=1,\cdots,n$。如果我们选取 $\mathscr{W}_m$ 的一组有序基 $B_{\mathscr{W}}=(\mathbf{b}_j)_{j=1}^m$，则 $\mathbf{w}_k$ 又可表示为 $\mathbf{w}_k=\sum_{j=1}^m\alpha_{jk}\mathbf{b}_j, k=1,\cdots,n$。此时，向量 $\mathbf{w}_k$ 的坐标 α_{jk} 需要两个下标来统一表示，它们构成一个 $m\times n$ 矩阵

$$(\mathbf{A})=\begin{pmatrix}\alpha_{11} & \cdots & \alpha_{1n}\\ \vdots & & \vdots\\ \alpha_{m1} & \cdots & \alpha_{mn}\end{pmatrix}$$

我们称矩阵 $(\mathbf{A})$ 是线性变换 $\mathbf{A}$ 在有序基 $B_{\mathscr{V}}$ 与 $B_{\mathscr{W}}$ 下的坐标矩阵（*coordinate matrix*）。α_{jk} 称为 $\mathbf{A}$ 的在有序基 $B_{\mathscr{V}}$ 与 $B_{\mathscr{W}}$ 下的坐标（*coordinates*）。

由于向量和线性变换都是抽象的一般概念，而基和坐标矩阵分别是向量空间和线性变换的一般属性，故无论把什么具体数学对象视作有限维向量空间的向量或线性变换并满足相应的定义，都可以在给定有序基下变成相应的矩阵运算*。若给定线性变换 $\mathbf{A}:\mathscr{V}_n\to\mathscr{W}_m$、向量 $\mathbf{x}\in\mathscr{V}_n$、$\mathbf{y}\in\mathscr{W}_m$ 和基向量 $\{\mathbf{a}_i\}_{i=1}^n\subset\mathscr{V}_n,\{\mathbf{b}_j\}_{j=1}^m\subset\mathscr{W}_m$，则 $\mathbf{x}$ 和 $\mathbf{y}$ 可分别表示为 $\mathbf{x}=\sum_{i=1}^n\xi_i\mathbf{a}_i$、$\mathbf{y}=\sum_{j=1}^m\eta_j\mathbf{b}_j$。若 $\mathbf{A}$ 在有序基 $(\mathbf{a}_i),(\mathbf{b}_j)$ 下的表示矩阵为 (α_{ji})，则线性关系式 $\mathbf{y}=\mathbf{A}\mathbf{x}$ 恰好

*像例2.1和例2.9中诸如 $\mathscr{C}^n$ 之类元素是函数的向量空间，维数是无穷。$\mathscr{C}^n$ 空间的任一组基都有无穷个基向量。把 $\mathscr{C}^n$ 空间的向量（函数）用一组基向量（函数）线性表出，将得到一个无穷级数。这种向量空间上的向量和线性变换的坐标无法写成有限个数。更多关于函数空间作为无穷维向量空间的知识可参见其他数学资料[5]。本讲义今后不再涉及无穷维向量空间，而默认只讨论有限维向量空间。

可以写成关于 $\mathbf{x}$、$\mathbf{y}$ 和 $\mathbf{A}$ 的矩阵之间的乘法关系，推算如下：

$$\begin{aligned}
\mathbf{y} &= \mathbf{A}\mathbf{x} \\
&= \mathbf{A}\sum_{i=1}^{n}\xi_i\mathbf{a}_i = \sum_{i=1}^{n}\xi_i\left(\sum_{j=1}^{m}\alpha_{ji}\mathbf{b}_j\right) \quad \text{仅利用线性变换定义中规定的性质} \\
&= \sum_{j=1}^{m}\left(\sum_{i=1}^{n}\xi_i\alpha_{ji}\right)\mathbf{b}_j \quad \text{变换求和顺序} \\
&= \sum_{j=1}^{m}\eta_j\mathbf{b}_j
\end{aligned}$$

$$\Leftrightarrow$$

$$\eta_j = \sum_{i=1}^{n}\alpha_{ji}\xi_i, j = 1, \cdots, m$$

最后这个表达式恰为以下矩阵乘法的计算法则：

$$\begin{pmatrix} \eta_1 \\ \vdots \\ \eta_m \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix}$$

这就是式子 $\mathbf{y} = \mathbf{A}\mathbf{x}$ 在给定基下的坐标运算法则。

上面的讨论也同时说明，数域 $\mathbb{F}$ 上的任一 $m \times n$ 矩阵 A 都通过下列关系

$$\mathbf{A}\left(\sum_{i=1}^{n}\xi_i\mathbf{a}_i\right) = \sum_{i=1}^{m}\left(\sum_{j=1}^{n}\alpha_{ji}\xi_i\right)\mathbf{b}_j$$

唯一确定了一个线性变换 $\mathbf{A} : \mathscr{V}_n \to \mathscr{W}_m, \mathbf{A} \in \mathscr{L}(\mathscr{V}_n, \mathscr{W}_m)$，使 $\mathbf{A}$ 在 $\mathscr{V}_n$ 的某组有序基 $(\mathbf{a}_i)_{i=1}^{n}$ 和 $\mathscr{W}_n$ 的某组有序基 $(\mathbf{b}_j)_{j=1}^{m}$ 下的矩阵表示恰为矩阵 A。用较严格的语言可总结成如下两个定理。

定理 2.14. 设 $\mathscr{V}_n$ 和 $\mathscr{W}_m$ 是数域 $\mathbb{F}$ 上的有限维向量空间。$B_{\mathscr{V}}$ 和 $B_{\mathscr{W}}$ 分别是 $\mathscr{V}_n$ 和 $\mathscr{W}_m$ 的一组基。对每个线性变换 $\mathbf{T} : \mathscr{V}_n \to \mathscr{W}_m$ 都存在唯一一个 $\mathbb{F}$ 上的 $m \times n$ 矩阵 T 使得 $(\mathbf{T}\mathbf{a}) = T(\mathbf{a})\, \forall \mathbf{a} \in \mathscr{V}_n$。其中 $(\cdot)$ 表示给定有序基下的矩阵表示。

定理 2.15. 设 $\mathscr{V}_n$ 和 $\mathscr{W}_m$ 是数域 $\mathbb{F}$ 上的有限维向量空间。在给定任意 $\mathscr{V}_n$ 的基 $B_{\mathscr{V}}$ 和 $\mathscr{W}_m$ 的基 $B_{\mathscr{W}}$ 下，从线性变换 $\mathbf{T} : \mathscr{V}_n \to \mathscr{W}_m$ 到其在上述基下的矩阵表示的对应关系是一个同构映射。

证明. 定理2.14中的关系式定义了一个由 $\mathscr{L}(\mathscr{V}_n, \mathscr{W}_m)$ 到 $\mathbb{F}^{m\times n}$ 的单射。再由矩阵运算法则易证满射。此略。此外，由于 $\mathbb{F}^{m\times n}$ 在通常的矩运算定义下是一个向量空间，故这一映射是同态映射 $\wedge$ 双射 $\Leftrightarrow$ 同构映射。 □

其实，以上两条定理几乎是与定理2.9及其推论重复的。总之我们可以简单地说，当确定了基的选择时，每个线性变换都唯一对应一个相应维数的矩阵，反之亦然。而且，线性变换的向量代数运算结果与矩阵的加法和标量乘法运算结果直接对应。需要注意的是，同一个向量或同一个线性变换在不同的基下的坐标一般是不同的。

通过以下定理，我们进一步获得线性变换的复合与矩阵乘法的对应。

定理 2.16. 设 $\mathscr{V},\mathscr{W},\mathscr{Z}$ 是 $\mathbb{F}$ 上的有限维向量空间，$\{\mathbf{e}_i\},\{\mathbf{f}_j\},\{\mathbf{g}_k\}$ 分别是 $\mathscr{V},\mathscr{W},\mathscr{Z}$ 的基，$\mathbf{T}:\mathscr{V}\to\mathscr{W},\mathbf{U}:\mathscr{W}\to\mathscr{Z}$ 是线性变换。则复合线性变换 $\mathbf{C}=\mathbf{TU}$ 在有序基 $(\mathbf{e}_i),(\mathbf{g}_k)$ 下的表示矩阵

$$(\mathbf{C})=(\mathbf{U})(\mathbf{T})$$

其中 $(\mathbf{T})$ 是 $\mathbf{T}$ 在 $(\mathbf{e}_i),(\mathbf{f}_j)$ 下的表示矩阵,$(\mathbf{U})$ 是 $\mathbf{U}$ 在 $(\mathbf{f}_j),(\mathbf{g}_k)$ 下的表示矩阵。

证明. 证明留作练习。 □

定理2.16就是复合线性变换在给定基下的坐标运算法则。

对于线性算符 $\mathbf{T},\mathbf{U}\in\mathscr{L}(\mathscr{V})$，由定理2.13有 $(\mathbf{U})(\mathbf{T})=(\mathbf{T})(\mathbf{U})=(\mathbf{I})$。易证在给定 $\mathscr{V}$ 的任意一组基下，恒等变换的矩阵表示都是单位矩阵 I，即 $(\mathbf{I})\equiv I$（不依赖基的选择）。用严格语言可总结为如下定理：

定理 2.17. 恒等变换 $\mathbf{I}:\mathscr{V}_n\to\mathscr{V}_n$ 在任意一组有序基下的矩阵表示都是 $n\times n$ 单位矩阵 I_n。

据此易验，可逆线性变换（线性算符）的矩阵与其逆变换的矩阵之间也互逆，即 $(\mathbf{T}^{-1})=(\mathbf{T})^{-1}$。

在下面的讨论中我们将看到，线性变换的秩等于其坐标矩阵的秩。设 $\mathscr{V}$、$\mathscr{W}$ 是数域 $\mathbb{F}$ 上的两个有限维向量空间，$\dim\mathscr{V}=n$，$\dim\mathscr{W}=m$。在分别给定 $\mathscr{V}$、$\mathscr{W}$ 的一组基 $B=\{\mathbf{e}_1,\cdots,\mathbf{e}_n\}$、$B'=\{\mathbf{e}'_1,\cdots,\mathbf{e}'_m\}$ 下，若向量 $\mathbf{y}\in\mathscr{W}$ 满足

$$\mathbf{y}=\mathbf{T}\mathbf{x},\quad \mathbf{x}\in\mathscr{V}$$

则

$$\mathbf{y}=\sum_{i=1}^{m}\sum_{j=1}^{n}T_{ij}x_j\mathbf{e}'_i$$

其中 T_{ij} 和 x_j 是 $\mathbf{T}$ 和 $\mathbf{x}$ 在所选定的基下的坐标。若以 $\mathbf{T}$ 的坐标矩阵的 n 个列为坐标形成 n 个 $\mathscr{W}$ 的向量

$$\mathbf{c}_j=\sum_{i=1}^{m}T_{ij}\mathbf{e}'_i,\quad j=1,\cdots n$$

则

$$\mathbf{y}=\sum_{j=1}^{n} x_j\mathbf{c}_j$$

这对任一 $\mathbf{x}\in\mathscr{V}$ 都成立。也就是说，$\mathrm{ran}\mathbf{T}$ 中的任一向量都能被向量组 $\{\mathbf{c}_1,\cdots,\mathbf{c}_n\}$ 线性表出，故 $\mathrm{ran}\mathbf{T}=\mathrm{span}\{\mathbf{c}_1,\cdots,\mathbf{c}_n\}$。$\dim(\mathrm{ran}\mathbf{T})$ 就是线性变换 $\mathbf{T}$ 的秩，而 $\dim(\mathrm{span}\{\mathbf{c}_1,\cdots,\mathbf{c}_n\})$ 就是 $\mathbf{T}$ 的坐标矩阵的列秩，矩阵的列秩就等于矩阵的秩[3]§4.4 定理 4.3。再加上一个矩阵的左乘或右乘一个可逆矩阵后的秩不变[3]§4.4 推论 4.6，故线性变换的坐标矩阵的秩在坐标变换下不变。我们于是可把线性变换的秩定义为它在任何选定基下的坐标矩阵的秩。

2.2.3 线性变换的转置

由之前的介绍可知，抽象定义的向量和线性变换都不是矩阵，但是在给定基下，它们又都能唯一地表示为矩阵。线性变换的转置，也需要在更一般的角度得到定义，而且需要为这件事引入更多新概念。

我们首先引入线性泛函的概念，以便定义线性变换的转置。

定义 2.14 (线性泛函与对偶空间)**.** 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的向量空间，从 $\mathscr{V}$ 到 $\mathbb{F}$ 的线性变换称为向量空间 $\mathscr{V}$ 上的*线性泛函*（*linear functional*），$\mathscr{V}$ 上的所有线性泛函的集合称为向量空间 $\mathscr{V}$ 的*对偶空间*（*dual space*），记为 $\mathscr{V}^*$。

线性泛函把一个向量对应为同数域上的一个数。由该定义，若映射 $f:\mathscr{V}\to\mathbb{F}$ 满足 $f(\alpha\mathbf{a}+\mathbf{b})=\alpha f(\mathbf{a})+f(\mathbf{b}),\forall\mathbf{a},\mathbf{b}\in\mathscr{V},\alpha\in\mathbb{F}$，则 f 是 $\mathscr{V}$ 上的一个线性泛函，且 $\mathscr{V}$ 的对偶空间 $\mathscr{V}^*$ 就是线性变换空间 $\mathscr{L}(\mathscr{V},\mathbb{F})$，且有 $\dim\mathscr{V}^*=\dim\mathscr{V}$。

例 2.11. *验证：*

- *函数* $f(x,y,z)=3x+5y-z,x,y,z\in\mathbb{R}$ *是 3 维实坐标空间* $\mathbb{R}^3$ *上的线性泛函。*
- *函数* $f(\mathbf{x})=3(\mathbf{x}\cdot\mathbf{x}),\mathbf{x}\in\mathbb{R}^n$ *是* n *维实坐标空间* $\mathbb{R}^n$ *上的线性泛函。*
- *设* $\mathbb{F}^{M\times M}$ *是数域* $\mathbb{F}$ *上的* $M\times M$ *矩阵的集合，矩阵的迹* $\mathrm{tr}A=A_{11}+\cdots+A_{MM},A\in\mathbb{F}^{M\times M}$ *是* $\mathbb{F}^{M\times M}$ *上的线性泛函。*
- *设* $a_1,\cdots,a_n\in\mathbb{F}$*，并定义映射* $f:\mathbb{F}^n\to\mathbb{F}$,

$$f(x_1,\cdots,x_n)=a_1x_1+\cdots+a_nx_n,\quad\forall(x_1,\cdots,x_n)\in\mathbb{F}$$

验证 f *是* $\mathbb{F}^n$ *上的线性泛函。这事实上既是两个* $n\times 1$ *列向量的“点乘”，又是一个* $1\times n$ *“行向量”与* $n\times 1$ *“列向量”的矩阵乘。*

- $\mathscr{C}[a,b]$ 是区间 $[a,b]$ 上的所有连续实值函数的集合，可验证它是数域 $\mathbb{R}$ 上的向量空间[3][§7.1 例 1.3]。定义 $L:\mathscr{C}[a,b]\to\mathbb{R}, L(g)=\int_a^b f(t)\,\mathrm{d}t,\quad \forall g\in\mathscr{C}[a,b]$，验证 L 是 $\mathscr{C}[a,b]$ 上的线性泛函。

从概念上可以说，线性泛函是 $\mathscr{V}$ 上的线性变换，$\mathscr{V}$ 的对偶空间是 $\mathscr{V}$ 上的线性变换的空间。但我们没有用 $\mathscr{L}(\mathscr{V},\mathbb{F})$ 这种记号，而是用 $\mathscr{V}^*$，是因为考虑到 $\dim\mathscr{V}=\dim\mathscr{V}^*$，$\mathscr{V}$ 与 $\mathscr{V}^*$ 更像是并列的概念。我们注意到，作为空间 $\mathscr{L}(\mathscr{V},\mathbb{F})$，维数应该写成 $\dim\mathbb{F}\times\dim\mathscr{V}$。也就是说，设 $\dim\mathscr{V}=N$，则给定 $\mathscr{V}$ 的某组有序基，$\mathscr{V}$ 的向量的坐标是 $N\times 1$“列向量”，而 $\mathscr{V}^*$ 中的线性泛函（它们也是向量）在给定 $\mathscr{V}^*$ 的某组有序基下的坐标是 $1\times N$“行向量”。那么，我们能不能进一步在 $\mathscr{V}$ 与 $\mathscr{V}^*$ 之间找到更明确的对应关系，看看 $\mathscr{V}$ 中的某一向量的坐标，“转置”成行向量之后，在 $\mathscr{V}^*$ 中找到的是哪一个线性泛函？这一联系当然需要给定 $\mathscr{V}$ 和 $\mathscr{V}^*$ 的基之后才能被确定。下面我们考察 $\mathscr{V}$ 的基与 $\mathscr{V}^*$ 的基的关系，通过一系列定理的证明找到两者之间的唯一对应性。

定理 2.18. 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的有限维向量空间，$\{\mathbf{a}_i\}$ 是 $\mathscr{V}$ 的一组基，则对每个 $i=1,\cdots,n$，有且只有一个 $\mathscr{V}$ 上的线性泛函 f_i 满足 $f_i(\mathbf{a}_j)=\delta_{ij},\quad i,j=1,\cdots,n$，且 $\{f_i\}$ 线性无关。

证明. “有且只有”由引理2.1易证，略。下面简要证明“线性无关”。

设 $f=\sum_{i=1}^n\gamma_i f_i$，则 $f(\mathbf{a}_j)=\sum_{i=1}^n\gamma_i f_i(\mathbf{a}_j)=\sum_{i=1}^n\delta_{ij}=c_j, j=1,\cdots,n$。特别地，$f=0$（这里的 0 是 $\mathscr{V}^*$ 的零向量）$\Leftrightarrow c_j=0\forall j=1,\cdots,n$。 □

由于 $\dim\mathscr{V}^*=\dim\mathscr{V}=n$，故 n 个 $\mathscr{V}^*$ 中的线性无关向量就是 $\mathscr{V}^*$ 的一组基，因而上面的定理和推论告诉我们，给定向量空间 $\mathscr{V}$ 上的一组有序基，总能在其对偶空间 $\mathscr{V}^*$ 中找到唯一对应的一组有序基，且 $\mathscr{V}^*$ 的基向量作为 $\mathscr{V}$ 上的线性泛函作用于 $\mathscr{V}$ 的基向量等于克劳内克符号。既然 $\mathscr{V}$、$\mathscr{V}^*$ 的上述两种基总是相互唯一对应的，我们就可以将一者定义为另一者的“对偶基”。

定义 2.15 (对偶基). 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的有限维向量空间。给定 $\mathscr{V}$ 的一组基 $B=\{\mathbf{a}_i\}$，$\mathscr{V}$ 的对偶空间 $\mathscr{V}^*$ 中的（唯一一组）满足 $f_i(\mathbf{a}_j)=\delta_{ij},\quad i,j=1,\cdots,\dim\mathscr{V}$ 的基 $\{f_i\}$ 称为 $\{\mathbf{a}_i\}$ 的对偶基（*dual basis*）。

由基的概念，$\mathscr{V}^*$ 中的任一线性泛函 $f\in\mathscr{V}^*$ 都可以用 $\mathscr{V}^*$ 的任一组基 $\{f_i\}$ 表出，即 $f=\sum_{i=1}^n c_i f_i, c_i\in\mathbb{F}$。如果已知 $\{f_i\}$ 是 $\mathscr{V}$ 的一组基 $B=\{\mathbf{a}_i\}$ 的对偶基，那么通过与定理2.18的证明过程类似的方法，可知 $c_i=f(\mathbf{a}_i), i=1,\cdots,n$，故 $f=\sum_{i=1}^n f(\mathbf{a}_i)f_i$，即 f 的第 i 个坐标可通过用 f 作用于有序基 B 的第 i 个向量来获得。同时，任一 $\mathscr{V}$ 中的向量 $\mathbf{x}\in\mathscr{V}$ 都可表示为 $\mathbf{x}=\sum_{i=1}^n\xi_i\mathbf{a}_i$。若用 B 的对偶基向量一一作用于 $\mathbf{x}$，$f_i(\mathbf{x})=\sum_{i=1}^n\xi_i f_i(\mathbf{a}_i)=$

$\sum_{i=1}^{n}\xi_i\delta_{ij}=\xi_i, i=1,\cdots,n$，即得到 $\mathbf{x}$ 在 B 下的相应坐标，故 $\mathbf{x}=\sum_{i=1}^{n}f_i(\mathbf{x})\,\mathbf{a}_i$，即 $\mathbf{x}$ 的第 i 个坐标可通过用 f_i 作用于向量 $\mathbf{x}$ 来获得。简而言之，对偶基其实就是一种“取坐标的函数”；一个向量在给定基下的第 i 个向量，就用这组基的第 i 个对偶基作用于这个向量来得到。

例 2.12. *设 $\mathbb{F}^{n\times 1}$ 是数域 $\mathbb{F}$ 上的所有 $n\times 1$“列向量”的集合。由例2.6可知，仅第 i 个分量为 1，其余分量均为零的 n 个列向量 E^i 是 $\mathbb{F}^{n\times 1}$ 的一组规范正交集。由于它们正好有 n 个，且由定理2.4它们线性无关，故它们就 $\mathbb{F}^{n\times 1}$ 的一组规范正交基。任一 $\mathscr{F}^{n\times 1}$ 中的列向量 $(a_1,\cdots,a_n)$ 在这组规范正交基下的坐标就恰为 $(a_1,\cdots,a_n)$。*

由例2.11最后一例的启发，我们发现 $E^{i\intercal}$ 与任一 $\mathscr{F}^{n\times 1}$ 中的列向量进行矩阵乘，可得到该列向量的第 i 个坐标，可见 $E^{i\intercal}$ 是 E^i 的对偶基。诚然，$E^{i\intercal}E^j=\delta_{ij}$。

我们在以往的线性代数课中知道，两个 $n\times 1$“列向量”的点乘 $\mathbf{a}\cdot\mathbf{b}$ 等于第一个向量转置成一个 $1\times n$“行向量”之后再与第二个向量作矩阵乘。在前面的段落中我们暗示了向量和线性泛函的概念与“列/行向量”类似。那么，我们能否再找出线性泛函与内积的联系呢？以下定理提供了具体的答案。

定理 2.19 (有限维里斯（Riesz）表示定理). 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的有限维内积空间，给定 $\mathscr{V}$ 上的任意一个线性泛函 $f\in\mathscr{V}^*$，必存在唯一一个向量 $\mathbf{b}\in\mathscr{V}$ 满足 $f(\mathbf{a})=(\mathbf{a}|\mathbf{b}), \forall\mathbf{a}\in\mathscr{V}$。

证明. 设 $\{\hat{\mathbf{e}}_i\}$ 是 $\mathscr{V}$ 的一个规范正交基，$\mathbf{a}=\sum_i\alpha_i\hat{\mathbf{e}}_i$ 是 $\mathscr{V}$ 中的任一向量，f 是 $\mathscr{V}^*$ 中的任一线性变换，则 $f(\mathbf{a})=\sum_i\alpha_i f(\hat{\mathbf{e}}_i)$。若要找到一个向量 $\mathbf{b}\in\mathscr{V}$ 满足 $(\mathbf{a}|\mathbf{b})=f(\mathbf{a})$，则要求等号左边 $(\mathbf{a}|\mathbf{b})=\sum_i\alpha_i\overline{\beta_i}$ 等于等号右边 $f(\mathbf{a})=\sum_i\alpha_i f(\hat{\mathbf{e}}_i)$，且需对任意 $\mathbf{a}\in\mathbf{V}$ 成立。这相当于要求 $\overline{\beta_j}=f(\hat{\mathbf{e}}_j)$，即 $\mathbf{b}=\sum_j\overline{f(\hat{\mathbf{e}}_j)}\hat{\mathbf{e}}_j$。这就证明了 $\mathbf{b}$ 的存在性。

设 $\mathbf{c}\in\mathscr{V}$ 也满足 $(\mathbf{a}|\mathbf{b})=(\mathbf{a}|\mathbf{c}), \forall\mathbf{a}\in\mathscr{V}$，则 $(\mathbf{a}|\mathbf{b}-\mathbf{c})=0, \forall\mathbf{a}\in\mathscr{V}\Leftrightarrow\mathbf{b}=\mathbf{c}$，故 $\mathbf{b}$ 是唯一的。 □

这就是说，在内积运算这件事情上，可以说向量空间 $\mathscr{V}$ 中的向量与它的对偶空间 $\mathscr{V}^*$ 中的线性泛函是一一对应的。基于内积的定义，内积空间 $\mathscr{V}$ 中的一个向量在对偶空间 $\mathscr{V}^*$ 中唯一对应的那个线性泛函，称为这个向量的对偶（*adjoint*）。在内积定义的讨论中，我们介绍过本讲义的内积与量子力学惯例（狄拉克 bra-ket 标记）的区别：$(\mathbf{x}|\mathbf{y})\equiv\langle\mathbf{y}|\mathbf{x}\rangle$。如果按照狄拉克 bra-ket 标记的规定，那么向量表示为 $|\mathbf{x}\rangle$，其对偶则直接表示为 $\langle\mathbf{x}|$，形象更加直接。

接下来我们完成定义线性变换的转置的任务。如图2.2所示，设数域 $\mathbb{F}$ 上的向量空间 $\mathscr{V}$ 上有一个线性变换 $\mathbf{T}$。给定任意向量 $\mathbf{a}\in\mathscr{V}$，线性变换后的向量是 $\mathbf{Ta}$。如果有一个线性泛函 g 把向量 $\mathbf{Ta}$ 对应为一个数，我们想知道，有没有一个线性泛函 f 能直接作用于 $\mathbf{a}$ 得到相同的数？f 与 g 的关系是什么？线性变换 $\mathbf{T}$ 的转置 $\mathbf{T}^\intercal$ 就是充当这一问题的答案的一种概念。你想问什么样的 f 能满足 $f(\mathbf{a})=g(\mathbf{Ta}), \forall\mathbf{a}\in\mathscr{V}$？答案是 $f=\mathbf{T}^\intercal g$。用严格语言的定义如下——

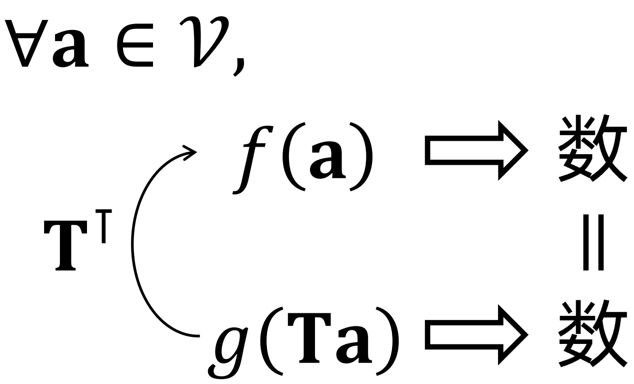


图 2.2: 线性变换的转置是怎么定义的？设数域 $\mathbb{F}$ 上的向量空间 $\mathscr{V}$ 上有一个线性变换 $\mathbf{T}$。给定任意向量 $\mathbf{a}\in\mathscr{V}$，线性变换后的向量是 $\mathbf{Ta}$。如果有一个线性泛函 g 把向量 $\mathbf{Ta}$ 对应为一个数，我们想知道，有没有一个线性泛函 f 能直接作用于 $\mathbf{a}$ 得到相同的数？f 与 g 的关系是什么？线性变换 $\mathbf{T}$ 的转置 $\mathbf{T}^{\intercal}$ 就是充当这一问题的答案的一种概念。你想问什么样的 f 能满足 $f(\mathbf{a})=g(\mathbf{Ta})\,\forall\mathbf{a}\in\mathscr{V}$？答案是 $f=\mathbf{T}^{\intercal}g$。

定义 2.16 (线性变换的转置)**.** 设 $\mathscr{V},\mathscr{W}$ 是数域 $\mathbb{F}$ 上的向量空间，$\mathbf{T}:\mathscr{V}\to\mathscr{W}$ 是线性变换，对于 $\mathscr{W}$ 上的每一个线性泛函 $g\in\mathscr{W}^*$，我们都可以定义一个 $\mathscr{V}$ 上的线性泛函 $f\in\mathscr{V}^*$ 使其满足 $f(\mathbf{a})=g(\mathbf{Ta}),\forall\mathbf{a}\in\mathscr{V}$。由 $g\in\mathscr{W}^*$ 到 $f\in\mathscr{V}^*$ 的这一对应规则定义了一个由 $\mathscr{W}^*$ 到 $\mathscr{V}^*$ 的映射 $\mathbf{T}^{\intercal}:\mathscr{W}^*\to\mathscr{V}^*,(\mathbf{T}^{\intercal}g)(\cdot)=g(\mathbf{T}\cdot),\forall g\in\mathscr{W}^*$，易验对任一 $\mathbf{T}\in\mathscr{L}(\mathscr{V},\mathscr{W})$ 有且只有一个满足上述性质的 $\mathbf{T}^{\intercal}\in\mathscr{L}(\mathscr{W}^*,\mathscr{V}^*)$，且 $\mathbf{T}^{\intercal}$ 也是线性变换。我们称 $\mathbf{T}^{\intercal}$ 为 $\mathbf{T}$ 的**转置**（*transpose*）。

以上定义中隐含默认了两件事：1）每个线性变换 $\mathbf{T}\in\mathscr{L}(\mathscr{V},\mathscr{W})$ 有且只有一个符合定义的映射 $\mathbf{T}^{\intercal}$；2）$\mathbf{T}^{\intercal}$ 也是一个线性变换。1）是比较显然的：假设另有一映射 $\mathbf{U}:\mathscr{W}^*\to\mathscr{V}^*$ 满足 $(\mathbf{U}g)\,\mathbf{a}=g\mathbf{Ua}\forall\mathbf{a}\in\mathscr{V}$，则有 $\mathbf{U}g=\mathbf{T}^{\intercal}g\Rightarrow\mathbf{U}\mathbf{T}^{\intercal}$。关于 2)，我们可以按照线性变换的定义去验证映射 $\mathbf{T}^{\intercal}$ 作用于一个线性组合的结果：对任意 $g_1,g_2\in\mathscr{W}^*,\gamma\in\mathbb{F}$，由 $[\mathbf{T}^{\intercal}(\gamma g_1+g_2)](\mathbf{a})=(\gamma g_1+g_2)(\mathbf{Ta})=\gamma g_1(\mathbf{Ta})+g_2(\mathbf{Ta})=\gamma(\mathbf{T}^{\intercal}g_1)(\mathbf{a})+(\mathbf{T}^{\intercal}g_2)(\mathbf{a})\,\forall\mathbf{a}\in\mathscr{V}$ 可知 $\mathbf{T}^{\intercal}(\gamma g_1+g_2)=\gamma(\mathbf{T}^{\intercal}g_1)+\mathbf{T}^{\intercal}g_2$，故 $\mathbf{T}^{\intercal}$ 是线性变换。

下面的定理告诉我们，一个线性变换 $\mathbf{T}$ 与其转置 $\mathbf{T}^{\intercal}$ 在给定基和对偶基下的矩阵之间的关系就是矩阵转置。

定理 2.20. 设 $\mathscr{V},\mathscr{W}$ 分别是数域 $\mathbb{F}$ 上的 n,m 维有限维向量空间，给定以下基及其对偶基：$B=\{\mathbf{a}_i\}\subset\mathscr{V},B^*=\{f_i\}\subset\mathscr{V}^*,B'=\{\mathbf{b}_i\}\subset\mathscr{W},B'^*=\{g_i\}\subset\mathscr{W}^*$，且令 $\mathbf{T}:\mathscr{V}\to\mathscr{W}$ 是线性变换，A 是 $\mathbf{T}$ 在 B,B' 下的坐标矩阵，B 是 $\mathbf{T}^{\intercal}$ 在 B^*,B'^* 下的坐标矩阵，则 $B_{ij}=A_{ji},i=1,\cdots,m,j=1,\cdots,n$。

证明. 根据向量的坐标表达式，可以写出

$$\mathbf{T}\mathbf{a}_i = \sum_{j=1}^{m} A_{ji}\mathbf{b}_j, i = 1, \cdots, n$$

$$\mathbf{T}^\intercal g_i = \sum_{j=1}^{n} B_{ji} f_j, i = 1, \cdots, m$$

由线性变换转置的定义和对偶基的定义，

$$\begin{aligned}
\left(\mathbf{T}^\intercal g_i\right)\left(\mathbf{a}_j\right) &= g_i\left(\mathbf{T}\mathbf{a}_j\right) \\
&= g_i\left(\sum_{k=1}^{m} A_{kj}\mathbf{b}_k\right) \\
&= \sum_{k=1}^{m} A_{kj} g_i\left(\mathbf{b}_k\right) \\
&= \sum_{k=1}^{m} A_{kj}\delta_{ik} \\
&= A_{ij}, i = 1, \cdots, m, j = 1, \cdots, n
\end{aligned}$$

另一方面，由于 $\mathbf{T}^\intercal g_i \in \mathscr{V}^*$，故它可用 B^* 表出，再利用关系 $f = \sum_{i=1}^{n} f\left(\mathbf{a}_i\right) f_i$，有

$$\begin{aligned}
\mathbf{T}^\intercal g_i &= \sum_{j=1}^{n} \left(\mathbf{T}^\intercal g_i\right)\left(\mathbf{a}_j\right) f_j \\
&= \sum_{j=1}^{n} A_{ij} f_j, i = 1, \cdots, m
\end{aligned}$$

与之前的结果比较可得 $A_{ij} = B_{ji}, i = 1, \cdots, m, j = 1, \cdots, n$。 □

2.3 基变换与坐标变换公式

一个向量空间可以有不止一组基。同一个向量，在不同基下的坐标是不同的，因此向量并不一般地对应一组数；线性变换与矩阵的关系也类似。这是本讲义中的线性代数部分与大学一年级的线性代数课程不同的地方。之所以要建立这种更抽象的线性代数结构，是为了用它们描述物理规律时，能拥有不依赖坐标系选择的不变形式。

但是，光说“向量并不一般地对应一组数”，仍然无法保证这种数学对象能够具有“不依赖基的选择的不变性”。比如，同在一个向量空间 $\mathscr{V}$ 内，向量 $\mathbf{a}$ 在某两组基下的坐标分别是 $(a_1, \cdots, a_n)^\intercal$ 和 $(a_1', \cdots, a_n')^\intercal$；向量 $\mathbf{b}$ 在选择某第三组基下的坐标是 $(b_1, \cdots, b_n)$。这三个有序 n 元数组各不相等，但有没有可能其实 $\mathbf{a} = \mathbf{b}$？我们必须有能力判断这件事，才一般性地拥有

“不依赖基的选择而辨认不同向量”的能力，从而才有资格作出“向量独立于坐标系的选择而存在”的判断。这样的能力就基于基变换与坐标变换公式的知识。基变换与坐标变换公式，使得每个向量在选择不同基下的坐标之间形成相互联系的整体，使得尽管不同向量各自都可以有不同的坐标，但它们仍然能相互区别。

我们先推导基变换公式。考虑数域 $\mathbb{F}$ 上的 N 维向量空间 $\mathscr{V}_N$ 中的两组有序基 $(\mathbf{e}_i),(\mathbf{e}'_j)$，用第一组基表示第二组基的每个基向量，可列出以下 N 个等式：

$$\mathbf{e}'_j=\sum_{i=1}^{N}S_{ij}\mathbf{e}_i,\quad j=1,\cdots,N$$

它们称为从有序基 $(\mathbf{e}_i)$ 到有序基 $(\mathbf{e}'_j)$ 的**基变换公式**（*change of basis formula*），其中矩阵 (S_{ij}) 称为从有序基 $(\mathbf{e}_i)$ 到有序基 $(\mathbf{e}'_j)$ 的**过渡矩阵**（*change-of-basis matrix*）。以下定理告诉我们，反过来从 $(\mathbf{e}'_i)$ 到 $(\mathbf{e}_i)$ 的过渡矩阵就是矩阵 S 的逆（S 可逆的证明见[3]§2.4 例题 4.4。）

定理 2.21. 设 $\mathscr{V}_N$ 是数域 $\mathbb{F}$ 上的 N 维向量空间，$\{\mathbf{e}_i\},\{\mathbf{e}'_j\}$ 是 $\mathscr{V}_N$ 的两组基，则从有序基 $(\mathbf{e}_i)$ 到有序基 $(\mathbf{e}'_j)$ 的过渡矩阵是从 $(\mathbf{e}'_j)$ 到 $(\mathbf{e}_i)$ 的过渡矩阵的逆矩阵。具体地，若 $\mathbf{e}'_j=\sum_{i=1}^{N}S_{ij}\mathbf{e}_i$，则 $\mathbf{e}_i=\sum_{j=1}^{N}S^{\mathrm{inv}}_{ji}\mathbf{e}'_j$*。

证明. 两边同乘 S^{-1} 立刻得证，略。 □

特别地，由 n 维向量空间的一组基到它自身的过渡矩阵是 $n\times n$ 单位矩阵 I_n。

我们通过过渡矩阵可以写出一个向量 $\mathbf{v}\in\mathscr{V}_N$ 在两组有序基 $(\mathbf{e}_i),(\mathbf{e}'_j)$ 下的坐标之间的关系，共有 N 条式子：

$$\begin{aligned}
\mathbf{v}&=\sum_{j=1}^{N}v'_j\mathbf{e}'_j\\
&=\sum_{j=1}^{N}v'_j\left(\sum_{i=1}^{N}S_{ij}\mathbf{e}_i\right)\\
&=\sum_{i=1}^{N}\left(\sum_{j=1}^{N}S_{ij}v'_j\right)\mathbf{e}_i\\
&=\sum_{i=1}^{N}v_i\mathbf{e}_i\\
\Leftrightarrow v_i&=\sum_{j=1}^{N}S_{ij}v'_j,,\quad i=1,\cdots,N
\end{aligned}$$

*由于矩阵的分量是标量，为防与倒数的记法混淆，矩阵 S 的逆矩阵 S^{-1} 的分量记为 S^{inv}_{ij}，矩阵 S 的分量的倒数记为 S^{-1}_{ij}。

这 N 个式子称为向量 $\mathbf{v}$ 从有序基 $\left(\mathbf{e}'_j\right)$ 到 $(\mathbf{e}_i)$ 的坐标变换公式。它们也可以写成一个矩阵乘法表达式：

$$\begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix} = \begin{pmatrix} S_{11} & \cdots & S_{1N} \\ \vdots & & \vdots \\ S_{N1} & \cdots & S_{NN} \end{pmatrix} \begin{pmatrix} v'_1 \\ \vdots \\ v'_N \end{pmatrix}$$

注意到，对于同一个矩阵 S_{ij}，它是从 $\left(\mathbf{e}'_j\right)$ 到 $(\mathbf{e}_i)$ 的过渡矩阵，但却用于向量 $\mathbf{v}$ 从 $(\mathbf{e}_i)$ 下的坐标到 $\left(\mathbf{e}'_j\right)$ 下的坐标的变换公式中。按照相同的推算方法还可以得到，向量 $\mathbf{v}$ 从 $\left(\mathbf{e}'_j\right)$ 到 $(\mathbf{e}_i)$ 的坐标变换公式是

$$v'_j = \sum_{i=1}^{N} S^{\text{inv}}_{ji} v_i, j = 1, \cdots, N$$

接下来，我们讨论线性变换在给定不同有序基下的坐标矩阵之间的变换公式。

定理 2.22. 设 $\mathscr{V}_N$、$\mathscr{W}_M$ 分别是数域 $\mathbb{F}$ 上的 N、M 维向量空间，$\{\mathbf{a}_i\}, \{\mathbf{a}'_i\} \in \mathscr{V}_N$ 是 $\mathscr{V}_N$ 的两组基，基变换公式为 $\mathbf{a}'_j = \sum_{i=1}^{N} S_{ij}\mathbf{a}_i$；$\{\mathbf{b}_j\}, \{\mathbf{b}'_j\}$ 是 $\mathscr{W}_M$ 的两组基，基变换公式为 $\mathbf{b}'_j = \sum_{i=1}^{M} T_{ij}\mathbf{b}_i$。线性变换 $\mathbf{A} : \mathscr{V}_N \to \mathscr{W}_M$ 在 $(\mathbf{a}_i), (\mathbf{b}_i)$ 下的矩阵表示为 $(\mathbf{A})$，在 $(\mathbf{a}'_i), (\mathbf{b}'_i)$ 下的矩阵表示为 $(\mathbf{A})'$。则有

$$(\mathbf{A}) = T\,(\mathbf{A})'\,S^{-1}$$
$$(\mathbf{A})' = T^{-1}\,(\mathbf{A})\,S$$

证明. 对于任一向量 $\mathbf{v} \in \mathscr{V}_N$，记 $\mathbf{w} = \mathbf{A}\mathbf{v} \in \mathscr{W}_M$。我们从向量 $\mathbf{w}$ 的坐标变换出发：

$$\begin{aligned} w_i &= \sum_{j=1}^{M} T_{ij} w'_j \\ &= \sum_{j=1}^{M} T_{ij} \left(\sum_{k=1}^{N} A'_{jk} v'_k \right) \\ &= \sum_{j=1}^{M} \sum_{k=1}^{N} T_{ij} A'_{jk} \left(\sum_{l=1}^{N} S^{\text{inv}}_{kl} v_l \right) \\ &= \sum_{j=1}^{M} \sum_{k=1}^{N} \sum_{l=1}^{N} T_{ij} A'_{jk} S^{\text{inv}}_{kl} v_l \\ &= \sum_{j=1}^{N} \sum_{l=1}^{M} \sum_{k=1}^{N} T_{il} A'_{lk} S^{\text{inv}}_{kl} v_j, \quad i = 1, \cdots, M \end{aligned}$$

其中 A_{ij}, A'_{ij} 分别是 $(\mathbf{A}), (\mathbf{A})'$ 的坐标。

另一方面，$w_i = \sum_{j=1}^{N} A_{ij} v_j, i = 1, \cdots, M$，与上面的结果比较可得：

$$A_{ij} = \sum_{l=1}^{M} \sum_{k=1}^{N} T_{il} A'_{lk} S^{\text{inv}}_{kj} \Leftrightarrow (\mathbf{A}) = T\,(\mathbf{A})'\,S^{-1}$$

由 $T^{-1}(\mathbf{A})S = T^{-1}T(\mathbf{A})'S^{-1}S = (\mathbf{A})'$，可得 $(\mathbf{A})' = T^{-1}(\mathbf{A})S$。 □

有了基变换和坐标变换公式，我们可以验证任何关于抽象的向量和线性变换的运算结果是否依赖基的选择。以下定理及其证明可作为一个示例。

定理 2.23. 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的有限维内积空间，则 $\mathscr{V}$ 上的内积运算结果不依赖基的选择。

证明. 设 $(\mathbf{e}),(\mathbf{e}')$ 是 $\mathscr{V}$ 的任意两组有序基，任意两个向量 $\mathbf{u},\mathbf{v}\in\mathscr{V}$ 在这两组有序基下的坐标表示为：$\mathbf{u}=\sum_{i=1}^n u\mathbf{e}_i=\sum_{i=1}^n u'\mathbf{e}'_i,\mathbf{v}=\sum_{i=1}^n v\mathbf{e}_i=\sum_{i=1}^n v'\mathbf{e}'_i$。设由 $(\mathbf{e})$ 到 $(\mathbf{e}')$ 的过渡矩阵坐标是 S_{ij}，即

$$\mathbf{e}'_j=\sum_{i=1}^n S_{ij}\mathbf{e}_i, j=1,\cdots,n$$

则有：

$$u_i=\sum_{i=1}^n S_{ij}u'_j, v_i=\sum_{i=1}^n S_{ij}v'_j, i=1,\cdots,n$$

记 $G_{ij}=(\mathbf{e}_i|\mathbf{e}_j),G'_{ij}=\left(\mathbf{e}'_i|\mathbf{e}'_j\right),i,j=1,\cdots,n$，分别是基 $\{\mathbf{e}_i\},\{\mathbf{e}'_i\}$ 的格拉姆矩阵的分量，则两组基之间的格拉姆矩阵变换关系为：

$$\begin{aligned}
G'_{ij}&=\left(\mathbf{e}'_i|\mathbf{e}'_j\right)\\
&=\left(\sum_{k=1}^n S_{ki}\mathbf{e}_k\middle|\sum_{l=1}^n S_{lj}\mathbf{e}_l\right)\\
&=\sum_{k=1}^n\sum_{l=1}^n S_{ki}\overline{S_{lj}}\left(\mathbf{e}_k|\mathbf{e}_l\right)\\
&=\sum_{k=1}^n\sum_{l=1}^n S_{ki}\overline{S_{lj}}G_{kl}
\end{aligned}$$

故 $\mathbf{u},\mathbf{v}\in\mathscr{V}$ 的内积

$$\begin{aligned}
(\mathbf{u}|\mathbf{v})&=\sum_{i=1}^n\sum_{j=1}^n u_i\overline{v_j}G_{ij}\\
&=\sum_{i=1}^n\sum_{j=1}^n\left(\sum_{k=1}^n S_{ik}u'_k\right)\overline{\left(\sum_{l=1}^n S_{jl}v'_l\right)}G_{ij}\quad\text{（利用了坐标变换公式。）}\\
&=\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n u'_k\overline{v'_k}S_{ik}\overline{S_{jl}}G_{ij}\\
&=\sum_{k=1}^n\sum_{l=1}^n u'_k\overline{v'_l}G'_{kl}\quad\text{（利用格拉姆矩阵的坐标变换公式。）}
\end{aligned}$$

可见两向量内积在任意两组基下的计算结果是相等的。 □

推论 2.23.1. 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的有限维赋范向量空间，则 $\mathscr{V}$ 中向量的范不依赖基的选择。

证明. 由 $\mathscr{V}$ 上的内积不依赖基的选择易证 $\mathscr{V}$ 上的欧几里得范不依赖基的选择。再由范的等价性（定理A.1）易证该命题。 □

2.4 线性算符

在线性变换的章节中，我们已经了解线性算符的定义（2.13）。除了满足所有线性变换的性质外，由于线性算符的定义域和值域都是同一个向量空间，所以在线性算符之上能够定义比较符合直觉的“乘法”，从而使得其求逆运算也更普遍。线性算符的可逆性是十分重要的性质。在大学一年级的线性代数课程中我们知道，矩阵的可逆性对应于线性方程组是否有非全零解的问题。在例2.9中我们知道求导和积分运算也是线性变换。它们所对应的方程是微分方程和积分方程。它们是否可逆，对应的是微分方程和积分方程的解的存在性问题。线性代数的研究者很早就发现，线性算符的可逆性与其行列式、对角化、特征值等性质密切相关，这些也是我们已经在大学一年级的线性代数课程中以矩阵为例有所了解的。在第2.4.1节，我们将从给出行线性算符的行列式、迹和特征值的不依赖基的选择的抽象定义，但是它们在给定有序基下的坐标运算，跟以往我们所学过的知识无异。

2.4.1 线性算符的行列式、迹和特征值

线性算符的行列式、迹和特征值，都是关于一个线性算符的标量，而且这些标量的取值不依赖基的选择。

线性算符的行列式

定义 2.17 ($n \times n$ 矩阵的行列式)**.** 设 A 是数域 $\mathbb{F}$ 上的 $n \times n$ 矩阵（记为 $A \in \mathbb{F}^{n\times n}$），$a_i = (A_{i1}, \cdots, A_{in}), i = 1, \cdots, n$ 是 A 的第 i 行的有序数组，如果函数 $D : \mathbb{F}^{n\times n} \to \mathbb{F}$ 满足：

1. D 是关于 $(a_1, \cdots, a_n)$ 的 n 重线线性函数，即
$$D(a_1, \cdots, \lambda a_i + a_i', \cdots, a_n) = \lambda D(a_1, \cdots, a_i, \cdots, a_n) + D(a_1, \cdots, a_i', \cdots, a_n), \forall \lambda \in \mathbb{F}$$
2. 当 A 的任意两行相等则 $D(A) = 0$
3. 设 A' 是 A 的任意两行调换后的矩阵，则 $D(A') = -D(A)$
4. $D(I) = 1$，其中 I 是 $n \times n$ 单位矩阵

则称 D 是矩阵 A 的行列式（*determinant*），记为 $\det A$。

上述定义中的条件 1 是本身是n 重线性函数（*n-linear function*）的定义，条件 1 和条件 2 一起则是交替（*alternating*）n 重线性函数的定义。条件 3 是可由条件 1 和 2 独立证明得到的（此略[4]§5.2），加到了定义中只是为了便于理解。我们可以说，行列式就是一个关于 $n\times n$ 矩阵的交替 n 重线性标量值函数且附加满足条件 4。

上面的定义隐含默认了任意一个 $n\times n$ 矩阵的行列式是唯一存在的要求（因为“函数”的定义本身要求这一条），但这未被证明。本讲义暂不列出详细证明的证明过程，只简述证明的思路。证明存在性往往相当于于找到这一存在。回顾本科的线性代数课本，我们发现那里的定义方式[3]§1.3 定义 3.1就帮我们找出了这一存在：

$$\det A=\sum_{\sigma}(\mathrm{sgn}\sigma)\,A\left(1,\sigma_1\right)\cdots A\left(n,\sigma_n\right)$$

其中 $A\left(i,j\right)\equiv A_{ij}$；$\sigma_i,i=1,\cdots,n$ 表示 n 阶排列 σ 的一种；$\mathrm{sgn}\sigma$ 是根据 σ 的奇偶性取值 1 或 -1*。作为存在性的证明，提出这一公式，然后验证其满足行列式定义中的条件，就完成了。而唯一性的证明则需要如下引理：任意 $\mathbb{F}^{n\times n}$ 上的交替 n 重线性函数 D 均满足 $D\left(A\right)=\det A D\left(I\right),\forall A\in\mathbb{F}^{n\times n}$，其中 $\det A$ 在此只需存在即可。该引理的证明也可以从上述的行列式公式直接得出，使得行列式定义中的条件 4 实际成为行列式唯一性的必要条件。

以下定理列出了 $n\times n$ 矩阵的行列式的一些性质, 在本讲义中也不作证明而直接承认其成立，因为它们都已经大学一年级的线性代数课程中介绍过了。

定理 2.24.

- $\det\left(AB\right)=\det A\det B,\forall A,B\in\mathbb{F}^{n\times n}$
- $\det\left(A^{\intercal}\right)=\det A,\forall A,B\in\mathbb{F}^{n\times n}$
- A 可逆 $\Leftrightarrow\det\left(A\right)\neq 0$
- $\det\left(A^{-1}\right)=\left(\det A\right)^{-1},\forall A,B\in\mathbb{F}^{n\times n}$
- 如果 B 是 A 的第 i 行加上第 j 行的倍数得到的矩阵，则 $\det B=\det A$
- 克拉默法则（*Cramer's rule*）[3]§1.5

定义在数域 $\mathbb{F}$ 上的有限维向量空间 $\mathscr{V}$ 上的线性算符 $\mathbf{T}\in\mathscr{L}\left(\mathscr{V}\right)$ 在不同有序基下的坐标矩阵之间有变换公式。设 B,B' 是 $\mathscr{V}$ 的两组有序基，S 是从 B 到 B' 的过渡矩阵，则由定理2.22有 $\left(\mathbf{T}\right)=S\left(\mathbf{T}\right)'S^{-1}$，其中 $\left(\mathbf{T}\right),\left(\mathbf{T}\right)'$ 分别是 $\mathbf{T}$ 在 B,B' 下的坐标矩阵。由行列式的性质有 $\det\left(\mathbf{T}\right)=\det\left[S\left(\mathbf{T}\right)S^{-1}\right]=\det\left(\mathbf{T}\right)'$，因此一个线性算符在任意基下的坐标矩阵的行列式都相等。由此我们可以直接定义“线性算符的行列式”，记为 $\det\mathbf{T}$，为其在任一基下的坐标矩阵的行列式。正式定义如下。

*关于 n 阶排列的知识可参见[3]§1.2。

定义 2.18 (线性算符的行列式). 设 $\mathbf{T}$ 是 n 维向量空间 $\mathscr{V}$ 上的线性算符，B 是 $\mathscr{V}$ 的一组有序基，则 $\mathbf{T}$ 的行列式 $\det\mathbf{T} \equiv \det(\mathbf{T})$，其中 $(\mathbf{T})$ 是 $\mathbf{T}$ 在 B 下的坐标矩阵。

线性算符的迹

从线性算符的行列式的定义过程，我们发现，与本科线性代数课上直接定义成一个运算公式不同，我们总是先定义一组代数规则，再去证明这样的代数规则唯一对应一个具体的运算公式。线性算符的迹也可按类似的方式重新定义。不过，既然已经通过行列式的例子来了解这种思想，此处关于迹的定义仍采用简化的方式。值得注意的是，迹是定义在内积空间上的线性算符上的。

定义 2.19 (线性算符的迹). 设 $\mathbf{A}$ 是数域 $\mathbb{F}$ 上的内积空间 $\mathscr{V}$ 上的线性算符，则 $\mathbf{A}$ 的迹（*trace*）为 $\mathrm{tr}\mathbf{A} \equiv \sum_k (\mathbf{A}\hat{\mathbf{e}}_k|\hat{\mathbf{e}}_k)$，其中 $\{\hat{\mathbf{e}}_k\}$ 是 $\mathscr{V}$ 的任意一组规范正交基。

易验，上述定义的迹的值不依赖基的选择而改变，从而任一线性算符唯一对应一个迹。我们还能进一步获得如下性质。

定理 2.25. 设 $\mathbf{A}$、$\mathbf{B}$ 是数域 $\mathbb{F}$ 上的内积空间 $\mathscr{V}$ 上的任意两个线性算符，则

1. $\mathrm{tr}(\alpha\mathbf{A}+\mathbf{B}) = \alpha\mathrm{tr}\mathbf{A} + \mathrm{tr}\mathbf{B}$
2. $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$

我们在此还可通过迹来定义 $\mathscr{L}(\mathscr{V})$ 空间上的一种内积。

定义 2.20 (线性算符的标准内积). 设 $\mathbf{A}, \mathbf{B}$ 是数域 $\mathbb{F}$ 上的内积空间 $\mathscr{V}$ 上的任意两个线性算符，令 $(\mathbf{A}|\mathbf{B}) \equiv \mathrm{tr}(\mathbf{A}^*\mathbf{B})$，可验证，该定义满足内积规定，称为线性算符的标准内积（*standard inner product*），经常记作 $\mathbf{A} : \mathbf{B}$。

读者可尝试自行推导两个线性算符的标准内积在给定基下的坐标运算公式。

线性算符的特征值

我们已经知道，向量空间上的线性算符在选定某组基下，它的性质就全由其坐标矩阵决定。线性算符的特征值分析，其实是对一个线性算符的“内部结构”像对钟表那般进行解剖，搞清楚它的运作方式。

尽管线性算符的特征值分析能给出很丰富的结论，但目前我们也许可以先从这样一种动机去引入线性算符的特征值定义。比如我们关心一个线性算符是否可逆，从线性变换的维数定理的角度，需要知道线性算符的秩，即其在任一选定基下的坐标矩阵的秩。从定理2.24的角

度，需要知道线性算符的行列式，即其在任一选定基下的坐标矩阵的行列式。如果一个数域 $\mathbb{F}$ 上的 n 维向量空间 $\mathscr{V}$ 上的线性算符 $\mathbf{T}$ 在基 $\{\mathbf{a}_i\}$ 下的坐标矩阵是对角矩阵，即

$$(\mathbf{T}) = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}, \lambda_i \in \mathbb{F}, i = 1, \cdots, n$$

就称线性算符 $\mathbf{T}$ 是可对角化的（*diagonalizable*）。此时有

$$\mathbf{T}\mathbf{a}_i = \lambda_i \mathbf{a}_i, \quad i = 1, \cdots, n$$

$\mathbf{T}$ 的值域就是由 $\lambda_i \neq 0$ 对应的 $\mathbf{a}_i$ 线性生成的，$\lambda_i \neq 0$ 的个数就是 $\mathrm{rank}\mathbf{T}$，要想 $\mathbf{T}$ 可逆，必须使得 $\mathrm{rank}\mathbf{T} = n$（满秩），或由 $\det\mathbf{T} = \prod_{i=1}^{n} \lambda_i \neq 0$ 要求 $(\mathbf{T})$ 的对角元素全不为零。总之矩阵 $\mathbf{T}$ 的很多性质就能确定了。

为了这一目标，我们先讨论一般的形如 $\mathbf{T}\mathbf{a} = \lambda\mathbf{a}$ 的情况。

定义 2.21 (特征值、特征向量、特征空间). 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的有限维向量空间，$\mathbf{T} \in \mathscr{L}(\mathscr{V})$ 是一个线性算符。若存在 $\mathbf{a} \neq \mathbf{0}, \mathbf{a} \in \mathscr{V}, \lambda \in \mathbb{F}$ 满足 $\mathbf{T}\mathbf{a} = \lambda\mathbf{a}$，则称 λ 是 $\mathbf{T}$ 的一个特征值（*characteristic value*），$\mathbf{a}$ 是 $\mathbf{T}$ 对应于特征值 λ 的一个特征向量（*characteristic vector*）。$\mathbf{T}$ 关于同一特征值 λ 的所有不同的特征向量的集合称为 $\mathbf{T}$ 的特征空间（*characteristic space*）。

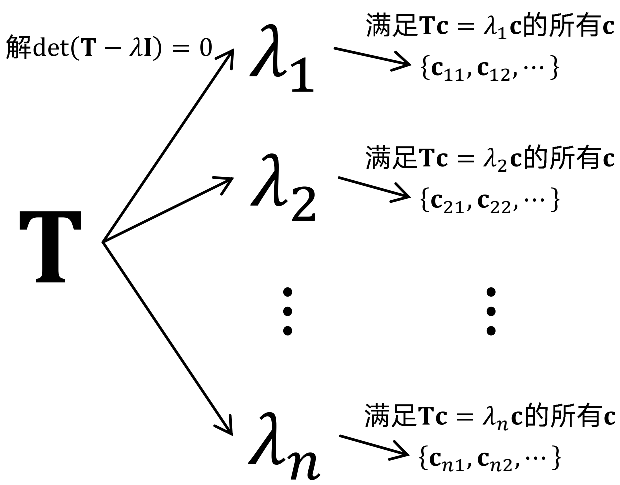


图 2.3: 复数域上的 n 维向量空间上的线性算符 $\mathbf{T}$ 有 n 个特征值，对应每个特征值都可能有一个或以上的特征向量。对应于每个特征值的所有特征向量的集合称为关于该特征值的特征空间。特征空间具有向量运算的封闭性（需证明），故它们都是所在向量空间的子空间。

由特征值和特征向量的定义可知，复数域上的 n 维向量空间上的线性算符可以有 n 个特征值。对应于同一个特征值，可能有多个特征向量。能够证明，关于同一特征值的特征空间是向量空间的一个子空间，作为定理如下。

定理 2.26. 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的有限维向量空间，$\mathbf{T} \in \mathscr{L}(\mathscr{V})$ 是一个线性算符，则 $\mathbf{T}$ 的特征空间是 $\mathscr{V}$ 的子空间。

证明. 设 $\mathbf{a}, \mathbf{b} \in \mathscr{V}$ 是关于 $\mathbf{T}$ 的特征值 λ 的其中两个特征向量，则对任意 $\alpha \in \mathbb{F}$ 有 $\mathbf{T}(\alpha\mathbf{a} + \mathbf{b}) = \alpha\mathbf{Ta} + \mathbf{Tb} = \alpha\lambda\mathbf{a} + \lambda\mathbf{b} = \lambda(\alpha\mathbf{a} + \mathbf{b})$。 □

定义2.21只告诉我们，如果有满足定义的东西，则把它称为什么；它没有告诉我们这样的东西存不存在，如何去求它们。为使特征值的概念更实用，我们进一步考虑一个重要的线性算符表达式。设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的有限维向量空间，$\mathbf{T} \in \mathscr{L}(\mathscr{V})$ 是一个线性算符，λ 是 $\mathbf{T}$ 的一个特征值，$\mathbf{a}$ 是 $\mathbf{T}$ 关于 λ 的任一特征向量，则总有

$$(\mathbf{T} - \lambda\mathbf{I})\,\mathbf{a} = \mathbf{Ta} - \lambda\mathbf{Ia} = \mathbf{0}$$

即 $\mathbf{T}$ 关于 λ 的特征空间就是线性算符 $\mathbf{T} - \lambda\mathbf{I}$ 的零空间。根据线性变换的维数定理，$\mathbf{T}$ 的特征空间的维数就是 $\mathbf{T} - \lambda\mathbf{I}$ 的零化度，进而有如下定理。

定理 2.27. 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的有限维向量空间，$\mathbf{T} \in \mathscr{L}(\mathscr{V})$ 是一个线性算符，则以下命题相互等价：

1. λ 是 $\mathbf{T}$ 的特征值
2. $\mathbf{T} - \lambda\mathbf{I}$ 是奇异的
3. $\det(\mathbf{T} - \lambda\mathbf{I}) = 0$

证明. 我们在本科的线性代数课中已经了解，矩阵行列式为零，就是该矩阵相对就的一个线性方程组无非全零解，亦即该矩阵是不可逆的。这些结论的详细证明就是对 $2 \Leftrightarrow 3$ 的证明。此略。$1 \Leftrightarrow 3$ 可利用行列式的性质简单证得，此略。 □

这一定理的第 3 个命题实际提供了一个求算给定线性算符的特征值的方法。因为

$$\det(\mathbf{T} - \lambda\mathbf{I})$$

必是一个 $\dim\mathscr{V}$ 阶首一多项式*。因此可将其定义为 **T 的特征多项式**（*characteristic polynomial*），从而 $\mathbf{T}$ 的任一特征值均为其特征多项式的一个根。在复数域上，由代数基本定理，n

*这一命题的结论需要依靠多项式代数的知识来证明。后文还要用到代数基本定理，这个定理也是通过多项式代数证明的重要结论，此略[4]。

阶首一多项式有且必有 n 个复根（可能有重根）*。因此我们可以说复数域上任一 n 维线性空间上的线性算符必有 n 个特征值；解其特征多项式就可以得到它们全部。

本节开头提出的，希望线性变换在某组基下的坐标矩阵是对角矩阵，这相当于要求线性变换的特征向量直接就是一组基。我们可以从这一要求出发，找出与其等价的一系列命题，如以下定理所示。

定理 2.28. 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的 n 维向量空间，$\mathbf{T} \in \mathscr{L}(\mathscr{V})$ 是一个线性算符，$\{\theta_1, \cdots, \theta_k\}$ 是 $\mathbf{T}$ 的两两不同的特征值，$\mathscr{W}_i$ 是线性算符 $\mathbf{T} - \theta_i \mathbf{I}$ 的零空间，则以下命题相互等价：

1. $\mathbf{T}$ 是可对角化的；
2. $\mathbf{T}$ 的特征多项式是
$$(\theta - \theta_1)^{d_1} \cdots (\theta - \theta_k)^{d_k}$$
且
$$\dim \mathscr{W}_i = d_i, \quad i = 1, \cdots, k$$
3. $\dim \mathscr{V} = \sum_{i=1}^{k} \dim \mathscr{W}_i$

这一定理第 1、2 条之间的等价性可见[3]§5.2 定理 2.1，第 2、3 条之间的等价性要参考其他线性代数教材[4]§6.2。

若考察第 1、2 条之间等价的证明过程，我们将看到，复数域上的向量空间上的任一线性算符必有 n 个特征值。但只有当这一线性算符可对角化时，才能从这些特征值对应的特征空间中找到 n 个线性无关的向量。如果线性算符 $\mathbf{T}$ 可对角化，它的对角坐标矩阵的对角元素就是 $\mathbf{T}$ 的 n 个特征值。如果 $\mathbf{T}$ 不可对角化，那么 $\mathbf{T}$ 仍可以有 n 个特征值，但就无法找到任何一组基，使 $\mathbf{T}$ 在其下的坐标矩阵是一个对角矩阵，更别说想要是以 $\mathbf{T}$ 的特征值作为对角元素的对角矩阵了。也就是说，我们不存在这样一种情况：$\mathbf{T}$ 可对角化，但其对角矩阵的对角元素却不全是 $\mathbf{T}$ 的特征值。

若考察第 2、3 条之间等价的证明，我们将看到，在复数域上，无论一个线性算符 $\mathbf{T}$ 是否可对角化，它的 n 个特征值中都有可能有重复的（即 $\mathbf{T}$ 的特征多项式有重根）。设 $\theta_1, \cdots, \theta_k$ 为 $\mathbf{T}$ 的两两不同特征值，第 j 个特征值重复 d_j 次。如果 $\mathbf{T}$ 可对角化，则 $\mathbf{T}$ 的对角化坐标矩阵可以写成
$$(\mathbf{T}) = \begin{pmatrix} \theta_1 I_1 & & \\ & \ddots & \\ & & \theta_k I_k \end{pmatrix}$$

*这样的数域又叫代数闭域。例如实数域就不是代数闭域。所以线性算符所有关于特征值的性质，在实数域上都需要小心是否会丧失。

其中 I_j 表示 $d_j \times d_j$ 单位矩阵。证明过程明确了这一结论：每个两两不同特征值的重复次数 d_j 就是它对应的特征空间的维数，进而有 $d_1 + \cdots + d_k = n$。此外，对应于不同特征值的特征向量线性无关。也就是说从若干个两两不同的特征值所对应的特征空间中分别各取一个向量所组成的向量组必为一个线性无关向量组[3]§5.1 性质 1.4，但是反过来我们不能说任意两个线性无关的特征向量必对应于两个不同特征值。

以下定理[3]§5.1 性质 1.1、1.2给出了特征多项式最重要的三项的系数：

定理 2.29. 设 $\{\lambda_i\}$ 是线性算符 $\mathbf{T} \in \mathscr{L}(\mathscr{V})$ 的 n 个特征值，则 $\mathrm{tr}\mathbf{T} = \sum_{i=1}^{n} \lambda_i$，$\det\mathbf{T} = \prod_{i=1}^{n} \lambda_i$，$\det(\mathbf{T} - \lambda\mathbf{I}) = \lambda^n - \mathrm{tr}\mathbf{T}\lambda^{n-1} + \cdots + (-1)^n \det\mathbf{T}$

证明. 利用 n 阶行列式的计算公式证明，略。 □

特别地，当 $n = 3$ 时，$\det(\mathbf{T} - \lambda\mathbf{I}) = -\lambda^3 + \mathrm{tr}\mathbf{T}\lambda^2 - \frac{1}{2}\left(\mathrm{tr}^2\mathbf{T} - \mathrm{tr}\mathbf{T}^2\right)\lambda + \det\mathbf{T}$。我们令 $I_\mathbf{T} = \mathrm{tr}\mathbf{T}, II_\mathbf{T} = \frac{1}{2}\left(\mathrm{tr}^2\mathbf{T} - \mathrm{tr}\mathbf{T}^2\right), III_\mathbf{T} = \det\mathbf{T}$，称为 $\mathbf{T}$ 的第一、第二和第三主不变量（*principal invariants*）。易见这些主不变量都是标量，且它的值不依赖基的选择，与 $\mathbf{T}$ 唯一对应。我们还常用 $J_1 = I_\mathbf{T}, J_2 = I_\mathbf{T}^2 - 2II_\mathbf{T}, J_3 = I_\mathbf{T}^3 - 3I_\mathbf{T}II_\mathbf{T} + 3III_\mathbf{T}$，中文也常译为主不变量（*main invariants*）。

定理2.29和2.24告诉我们验证线性算符是否可逆的一个方法；可逆线性算符所有特征值不为零。

2.4.2 内积空间上的线性算符

当我们为一个向量空间赋予了内积定义，这个空间上的线性算符的性质也相应增加了。由于向量的内积与“投影”、“正交”等几何概念密切相关，向量的范与“长度”这一几何概念概念密切相关，欧几里得范与内积之间也密切相关，因此内积空间上的线性算符也富有几何意义。我们先分别介绍伴随算符（§2.4.2伴随算符）和幺正算符（§2.4.2幺正算符）。它们都属于正规算符（§2.4.2正规算符）。在正式的线性代数教材中，这些算符的内容十分繁杂。本讲义在此只列举一些流变学理论用到的结论。

伴随算符

按照内积的定义，我们有 $(\mathbf{a}|\alpha\mathbf{b}) = (\overline{\alpha}\mathbf{a}|\mathbf{b})$（$\alpha \neq 0$）。如果把等号左边的标量 α 换成一个非零线性算符 $\mathbf{T}$，那么处于等号右边的“α 的复数共轭”这一角色的，又会是怎样的一个线性算符？我们先给这个线性算符一个名称，然后再分析它的特点。

定义 2.22 (伴随算符). 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的内积空间，$\mathbf{T} \in \mathscr{L}(\mathscr{V})$ 是 $\mathscr{V}$ 上的线性算符。若存在另一线性算符 $\mathbf{T}^* \in \mathscr{L}(\mathscr{V})$ 满足 $(\mathbf{Ta}|\mathbf{b}) = (\mathbf{a}|\mathbf{T}^*\mathbf{b}), \forall \mathbf{a}, \mathbf{b} \in \mathscr{V}$，则称 $\mathbf{T}^*$ 是 $\mathbf{T}$ 的伴随算符（*adjoint operator*）。若 $\mathbf{T} = \mathbf{T}^*$，则称 $\mathbf{T}$ 是一个自伴随（*self-adjoint*）或厄米（*hermitian*）算符。若 $\mathbf{T} = -\mathbf{T}^*$，则称 $\mathbf{T}$ 是一个反厄米（*skew hermitian*）算符。

在有限维内积空间上，每一个线性算符有且只有一个伴随算符（证明见附录 §A.3）。

易验，若 $\mathbf{T}$ 是厄米算符，则 $i\mathbf{T}$ 就是反厄米算符。

以下定理列举了一些伴随算符的运算规律。

定理 2.30. 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的有限维内积空间，

1. $(\mathbf{T}+\mathbf{U})^* = \mathbf{T}^* + \mathbf{U}^*, \forall \mathbf{T}, \mathbf{U} \in \mathscr{L}(\mathscr{V})$
2. $(\alpha\mathbf{T})^* = \overline{\alpha}\mathbf{T}^*, \forall \mathbf{T} \in \mathscr{L}(\mathscr{V}), \alpha \in \mathbb{F}$
3. $(\mathbf{TU})^* = \mathbf{U}^*\mathbf{T}^*, \forall \mathbf{T}, \mathbf{U} \in \mathscr{L}(\mathscr{V})$
4. $(\mathbf{T}^*)^* = \mathbf{T}, \forall \mathbf{T} \in \mathscr{L}(\mathscr{V})$

证明. 利用相关定义易证，留作练习。 □

从上面的运算规律可以看出，线性算符的伴随与复数的共轭有些类似。例如在复数域 $\mathbb{C}$ 上的内积空间 $\mathscr{V}$ 上，任一线性算符 $\mathbf{T}$ 都可以写成“虚部与实部”的形式，即 $\mathbf{T} = \mathbf{U}_1 + i\mathbf{U}_2$，其中 $\mathbf{U}_1 = \frac{1}{2}(\mathbf{T}+\mathbf{T}^*), \mathbf{U}_2 = \frac{1}{2i}(\mathbf{T}-\mathbf{T}^*)$ 都是自伴随算符。注意到，$i\mathbf{U}_2$ 是反厄米算符，故我们也常说任一线性算符 $\mathbf{T}$ 都可以分解成一个厄米算符和一个反厄米算符。

一对伴随算符在给定基下的坐标矩阵之间的关系是矩阵的共轭转置。以下我们给出证明。从证明的过程可以注意到，内积空间上的向量和线性算符如何通过内积来取得它们在给定有序基下的坐标。

定理 2.31. 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的 n 维内积空间，$B = \{\widehat{\mathbf{e}}_i\}_{i=1}^n$ 是 $\mathscr{V}$ 的一组规范正交基，$\mathscr{V}$ 上的线性算符 $\mathbf{T}$ 及其伴随算符 $\mathbf{T}^*$ 在有序基 B 下的坐标分别是 $T_{ij}, T'_{ij}, i,j = 1,\cdots,n$，则 $T_{ij} = \overline{T'_{ji}}$。

证明. 设 $\mathbf{a} \in \mathscr{V}$ 是 $\mathscr{V}$ 中的任一向量，则 $\mathbf{a} = \sum_{i=1}^n \alpha_i \widehat{\mathbf{e}}_i$，其中 $\alpha_i, i = 1,\cdots,n$ 是 $\mathbf{a}$ 在有序基 B 下的坐标，则 $\alpha_i = (\mathbf{a}|\widehat{\mathbf{e}}_i)$，因为

$$\begin{aligned}(\mathbf{a}|\widehat{\mathbf{e}}_i) &= \left(\sum_{j=1}^n \alpha_j \widehat{\mathbf{e}}_j \middle| \widehat{\mathbf{e}}_i\right) = \sum_{j=1}^n \alpha_j (\widehat{\mathbf{e}}_i|\widehat{\mathbf{e}}_j) \\ &= \sum_{j=1}^n \alpha_j \delta_{ij} = \alpha_i\end{aligned}$$

由 §2.2.2可知，线性算符 $\mathbf{T}$ 在给定有序基 B 下的坐标 T_{ij} 满足 $\mathbf{T}\widehat{\mathbf{e}}_i=\sum_{j=1}^{n}T_{ji}\widehat{\mathbf{e}}_j, i=1,\cdots,n$。同时每个 $\mathbf{T}\widehat{\mathbf{e}}_i$ 作为一个向量在有序基 B 下的坐标满足上面刚刚证明结论，故 $\mathbf{T}\widehat{\mathbf{e}}_i=\sum_{j=1}^{n}(\mathbf{T}\widehat{\mathbf{e}}_i|\widehat{\mathbf{e}}_j)\,\widehat{\mathbf{e}}_j$。由于 $\{\widehat{\mathbf{e}}_j\}$ 线性无关，当 $\mathbf{T}\neq\mathbf{0}$ 时比较上述两结果可得 $T_{ij}=(\mathbf{T}\widehat{\mathbf{e}}_j|\widehat{\mathbf{e}}_i)$。

由伴随算符定义，

$$\begin{aligned}T_{ji}=(\mathbf{T}\widehat{\mathbf{e}}_i|\widehat{\mathbf{e}}_j)&=(\widehat{\mathbf{e}}_i|\mathbf{T}^*\widehat{\mathbf{e}}_j)\\&=\overline{(\mathbf{T}^*\widehat{\mathbf{e}}_j|\widehat{\mathbf{e}}_i)}\\&=\overline{T'_{ij}}\end{aligned}$$

□

注意，这一定理讨论的是厄米算符仅限于在规范正交基下的坐标规律。在一般有序基下的坐标规律表达式将额外含有该组基的格拉姆矩阵分量。

回顾线性变换的转置（§2.2.3），如果由 $\mathscr{V}$ 的每个向量 $\mathbf{a}\in\mathscr{V}$ 定义一个相应的线性泛函来实现内积，即 $f_{\mathbf{a}}\in\mathscr{V}^*, f_{\mathbf{a}}(\mathbf{b})\equiv(\mathbf{a}|\mathbf{b})$，则 $\mathscr{V}$ 上的任一线性算符 $\mathbf{T}\in\mathscr{L}(\mathscr{V})$ 与其转置 $\mathbf{T}^{\intercal}\in\mathscr{L}(\mathscr{V}^*)$ 将满足

$$(\mathbf{T}^{\intercal}f_{\mathbf{a}})(\mathbf{b})=(\mathbf{a}|\mathbf{T}\mathbf{b}),\forall\mathbf{a},\mathbf{b}\in\mathscr{V}$$

同时 $\mathbf{T}$ 与 $\mathbf{T}^{\intercal}$ 在给定一对基 $B\subset\mathscr{V}$ 与对偶基 $B^*\subset\mathscr{V}^*$ 下的坐标矩阵互为矩阵的转置。而线性算符 $\mathbf{T}$ 与其伴随算符 $\mathbf{T}^*$ 满足

$$(\mathbf{T}\mathbf{a}|\mathbf{b})=(\mathbf{a}|\mathbf{T}^*\mathbf{b}),\forall\mathbf{a},\mathbf{b}\in\mathscr{V}$$

且 $\mathbf{T}$ 与 $\mathbf{T}^*$ 在给定基 B 下的坐标矩阵之间互为矩阵的共轭转置。比较可发现，线性算符的转置与伴随似乎仅存在“是否需要共轭”的差别，这个差别来自于它们扮演的角色是放在在内积的第一个向量前面还是第二个向量前面。本讲义的内积定义规定对第二向量具有共轭线性，所以坐标矩阵需要共轭的是定义在内积的第二个向量的伴随算符。但是，线性变换的转置和伴随概念根本不同。因为 $\mathbf{T}^{\intercal}\in\mathscr{L}(\mathscr{V}^*)$ 而 $\mathbf{T}^*\in\mathscr{L}(\mathscr{V})$，即它们属于不同的空间，作用于不同空间中的对象。在实数域中（经典力学语境下），线性算符的伴随的概念不如在复数域中（量子力学语境下）重要。在本讲义后续内容中凡涉及到实数域上的线性算符，都暂不区分其转置和伴随，而均写成 $\mathbf{T}^{\intercal}$。

此外，在以往所学习的矩阵代数中，有“对称矩阵”的概念。如果数域 $\mathbb{F}$ 上的 $n\times n$ 矩阵 $A\in\mathbb{F}^{n\times n}$ 满足 $A=A^{\intercal}$，则称矩阵 A 是对称矩阵（*symmetric matrix*）；若 $A=-A^{\intercal}$，则称矩阵 A 是斜称矩阵（*skew-symmetric matrix*）。由厄米与反厄米算符的定义可知，只有在实数域 $\mathbb{R}$ 上的内积空间上，厄米和反厄米算符在给定基于的坐标矩阵才是对称和斜称矩阵。因

此，我们又把实数域上的厄米和反厄米算符*称为对称算符（*symmetric operator*）和斜称算符（*skew-symmetrix operator*）。

最后我们关心一下厄米和反厄米算符在特征值分解时的性质。归入以下定理——

定理 2.32. 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的 n 维内积空间，$\mathbf{T}$ 是 $\mathscr{V}$ 上的一个线性算符，则

1. 若 $\mathbb{F}=\mathbb{C}$，$\mathbf{T}$ 是厄米算符，则 $\mathbf{T}$ 的特征值全为实数。
2. 若 $\mathbf{T}$ 是厄米算符，则 $\mathbf{T}$ 的两个不同特征值分别对应的两个特征向量正交。
3. 若 $\mathscr{V}$ 是有限维的，$\mathbf{T}$ 是厄米算符，则 $\mathbf{T}$ 至少有一个（非零）特征向量。
4. 若 $\mathscr{V}$ 是有限维的，$\mathbf{T}$ 是厄米算符，则 $\mathscr{V}$ 中存在一组规范正交基是 $\mathbf{T}$ 的特征向量。
5. 若 $\mathbb{F}=\mathbb{C}$，$\mathbf{T}$ 是反厄米算符，则 $\mathbf{T}$ 的所有特征值都是纯虚数或 0。

证明. “1”的证明：设 λ 是 $\mathbf{T}$ 的任意一个特征值，$\mathbf{a}$ 是 $\mathbf{T}$ 关于 λ 的任意一个特征向量，即 $\mathbf{Ta}=\lambda\mathbf{a}$，则

$$
\begin{aligned}
&\left(\mathbf{a}|\mathbf{Ta}\right)=\overline{\lambda}\left(\mathbf{a}|\mathbf{a}\right)=\left(\mathbf{Ta}|\mathbf{a}\right)=\lambda\left(\mathbf{a}|\mathbf{a}\right)\\
\Leftrightarrow&\overline{\lambda}=\lambda
\end{aligned}
$$

其中用到了 $\mathbf{a}$ 是特征向量故 $\mathbf{a}\neq\mathbf{0}$。

“2”的证明：设 $\lambda_1\neq\lambda_2$ 是 $\mathbf{T}$ 的两个特征值，$\mathbf{a}_1,\mathbf{a}_2$ 分别是 λ_1,λ_2 对应的特征向量，则

$$
\begin{aligned}
&\left(\mathbf{a}_1|\mathbf{Ta}_2\right)=\lambda_2\left(\mathbf{a}_1|\mathbf{a}_2\right)=\left(\mathbf{Ta}_1|\mathbf{a}_2\right)=\lambda_1\left(\mathbf{a}_1|\mathbf{a}_2\right)\\
\Leftrightarrow&\left(\lambda_1-\lambda_2\right)\left(\mathbf{a}_1|\mathbf{a}_2\right)=0\\
\Leftrightarrow&\left(\mathbf{a}_1|\mathbf{a}_2\right)=0
\end{aligned}
$$

其中用到了“1”。

“3”的证明：我们建立以下联系，以便在矩阵代数的层面证有关于 $\mathbf{T}$ 的命题。

设 $\mathscr{V}$ 是复数域 $\mathbb{C}$ 上的 n 维内积空间，$B=\{\hat{\mathbf{e}}_i\}$ 是 $\mathscr{V}$ 的一组规范正交基。记 T_{ij} 是 $\mathbf{T}$ 在 B 下的坐标，则矩阵 T 是厄米矩阵。

设 $\mathscr{W}$ 是 $\mathbb{C}$ 上所有 $n\times 1$ 矩阵（“列向量”），并定义了内积 $(X|Y)=\overline{Y}^{\intercal}X,\forall X,Y\in\mathscr{W}$。在 $\mathscr{W}$ 上定义映射 $U\left(X\right)\equiv TX$，则 U 是 $\mathscr{W}$ 上的厄米算符，因为

$$
\left(X|U\left(Y\right)\right)=\left(X|TY\right)=\left(TX|Y\right)=\left(U\left(x\right)|Y\right)
$$

其中用到了 $T=\overline{T}^{\intercal}$。

*即在实数域上的内积空间，满足定义2.22中的内积运算性质的线性算符。

由“1”，矩阵 T 的特征多项式应全为实根。设 λ_1 是其中一个实根（即 λ_1 是矩阵 T 的一个特征值），则存在非零 $X \in \mathscr{W}, X = (x_1, \cdots, x_n)^\intercal$ 满足线性方程组

$$(T - \lambda_1 I)\, X = (0, \cdots, 0)^\intercal$$

则 $\mathscr{V}$ 中有非零向量 $\mathbf{x} = \sum_{i=1}^n x_i \widehat{\mathbf{e}}_i$ 满足 $\mathbf{Tx} = \lambda_1 \mathbf{x}$。

若 $\mathscr{V}$ 是实数域 $\mathbb{R}$ 上的 n 维内积空间，由于 T 的所有特征值（包括 λ_1）都是实数，在实数域上 $\det(T - \lambda_1 I) = 0$ 仍成立，故在实数域上，上述线性方程组仍有非零解。

“4”的证明见附录。

“5”的证明：与“1”的证明类似，略。 □

在特征值的章节中曾提到，在复数域上，对于一般向量空间上的线性算符，对应于不同特征值的特征向量线性无关。定理2.32“2”其实是这一结论中的线性算符是内积空间上的厄米算符时的延伸结论；这种情况下对应于不同特征值的特征向量不仅线性无关，还两两正交。

第 4 条再加上定理2.28可知，

推论 2.32.1. 可逆厄米算符必可对角化。

推论 2.32.2. 在实数域 $\mathbb{R}$ 上的有限维内积空间 $\mathscr{V}$ 上，某线性算符 $\mathbf{T}$ 是对称算符当且仅当 $\mathscr{V}$ 中存在一组规范正交基是 $\mathbf{T}$ 的特征向量。

证明. “充份性”：定理2.32“4”的证明过程并不依赖复数域的特殊性，故其结论在实数域上仍成立。

“必要性”：设 $\{\widehat{\mathbf{e}}_i\}$ 是 $\mathscr{V}$ 的一组规范正交基且它们是 $\mathbf{T}$ 的特征向量。记 $\lambda_i \in \mathbb{R}, i = 1, \cdots, n$ 为 $\widehat{\mathbf{e}}_i$ 所对应的 $\mathbf{T}$ 的特征值，则有 $\mathbf{T}\widehat{\mathbf{e}}_i = \lambda_i \widehat{\mathbf{e}}_i, i = 1, \cdots, n$。故有

$$\begin{aligned}
(\mathbf{T}\widehat{\mathbf{e}}_i|\widehat{\mathbf{e}}_j) &= \lambda_i \delta_{ij}, \quad i, j = 1, \cdots, n \\
(\mathbf{T}^*\widehat{\mathbf{e}}_i|\widehat{\mathbf{e}}_j) &= (\widehat{\mathbf{e}}_i|\mathbf{T}\widehat{\mathbf{e}}_j) \\
&= \lambda_j \delta_{ij}, \quad i, j = 1, \cdots, n
\end{aligned}$$

因此在在某组基下，$\mathbf{T}^*$ 的坐标矩阵等于 $\mathbf{T}$ 的坐标矩阵，故 $\mathbf{T} = \mathbf{T}^*$ □

幺正算符

幺正算符有许多相互等价的性质，选择哪一个作为其定义在数学上也都是等价的。本讲义仍倾向于从其在内积运算中扮演的角色——保持内积运算结果不变——来定义。

引理 2.2. 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的内积空间。若双射 $Q:\mathscr{V}\to\mathscr{V}$ 满足

$$(Q(\mathbf{a})\,|Q(\mathbf{b}))=(\mathbf{a}|\mathbf{b})\,,\quad \forall\mathbf{a},\mathbf{b}\in\mathscr{V}$$

（又可说 Q 保持内积，或说 Q 是 $\mathscr{V}$ 上的自同构映射（*automorphism*），则 Q 是 $\mathscr{V}$ 上的线性算符。

证明. 因 Q 是双射，故对任意给定的 $\mathbf{a}'\in\mathscr{V}$ 必存在唯一 $\mathbf{a}\in\mathscr{V}$ 满足 $\mathbf{a}'=Q(\mathbf{a})$，给定任意 $\mathbf{u},\mathbf{v}\in\mathscr{V}$ 和 $\alpha\in\mathbb{F}$，有

$$\begin{aligned}(Q(\alpha\mathbf{u}+\mathbf{v})-\alpha Q(\mathbf{u})-Q(\mathbf{v})\,|\mathbf{a}') &= (Q(\alpha\mathbf{u}+\mathbf{v})\,|\mathbf{a}')-\alpha(Q(\mathbf{u})\,|\mathbf{a}')-(Q(\mathbf{v})\,|\mathbf{a}')\\ &=(\alpha\mathbf{u}+\mathbf{v}|\mathbf{a})-\alpha(\mathbf{u}|\mathbf{a})-(\mathbf{v}|\mathbf{a})\\ &=0\\ \Rightarrow Q(\alpha\mathbf{u}+\mathbf{v}) &=\alpha Q(\mathbf{u})+Q(\mathbf{v})\,,\quad \forall\mathbf{u},\mathbf{v}\in\mathscr{V},\alpha\in\mathbb{F}\end{aligned}$$

□

定义 2.23 (幺正算符). 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的内积空间，若映射 $\mathbf{Q}:\mathscr{V}\to\mathscr{V}$ 是保持内积的双射，就称 $\mathbf{Q}$ 是 $\mathscr{V}$ 上的一个幺正算符（*unitary operator*）。

因此定义2.23又可以简洁地表述为：内积空间上的自同构映射叫做幺正算符。

幺正算符的其他等价的定义，可见以下定理。

定理 2.33. 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的有限维内积空间，$\mathbf{Q}\in\mathscr{L}(\mathscr{V})$ 是一个线性算符，则以下命题相互等价：

1. $\mathbf{Q}$ 保持内积（即 $\mathbf{Q}$ 是幺正算符）
2. $\mathbf{Q}$ 是双射
3. $\mathbf{Q}$ 把 $\mathscr{V}$ 的某个规范正交基映射为另一个规范正交基
4. $\mathbf{Q}$ 把 $\mathscr{V}$ 的每个规范正交基映射为一个规范正交基
5. $\mathbf{Q}\mathbf{Q}^*=\mathbf{I}$

证明. 1)⇒2)：由 1)，$(\mathbf{Qa}|\mathbf{Qa})=(\mathbf{a}|\mathbf{a})\geq 0\forall\mathbf{a}\in\mathscr{V}$，当且仅当 $\mathbf{a}=\mathbf{0}$ 时取等号。故 $\mathbf{Q}$ 是非奇异的。由于 $\mathbf{Q}:\mathscr{V}\to\mathscr{V}$，故 $\mathbf{Q}$ 是双射*。

*这里用到了线性变换的维数定理及其推论。

2)⇒3)：由于 $\mathbf{Q}$ 是内积空间上的同构映射，令 $\{\widehat{\mathbf{e}}_i\}$ 是 $\mathscr{V}$ 的一组规范正交基，则由向量空间上的同构映射性质，$\{\mathbf{Q}\widehat{\mathbf{e}}_i\}$ 是 $\mathscr{V}$ 的一组基，且有 $\left(\mathbf{Q}\widehat{\mathbf{e}}_i|\mathbf{Q}\widehat{\mathbf{e}}_j\right)=(\widehat{\mathbf{e}}_i|\widehat{\mathbf{e}}_j)=\delta_{ij}, i,j=1,\cdots,\dim\mathscr{V}$，故 $\{\mathbf{Q}\widehat{\mathbf{e}}_i\}$ 是 $\mathscr{V}$ 的一组规范正交基。

3)⇒4)：显然易证；

4)⇒1)：由 4)，若已知 $\{\widehat{\mathbf{e}}_i\}$ 是 $\mathscr{V}$ 的一组规范正交基，且 $\{\mathbf{Q}\widehat{\mathbf{e}}_i\}$ 也是规范正交基，则有 $\left(\mathbf{Q}\widehat{\mathbf{e}}_i|\mathbf{Q}\widehat{\mathbf{e}}_j\right)=\delta_{ij}=(\widehat{\mathbf{e}}_i|\widehat{\mathbf{e}}_j)$。对任意 $\mathbf{a},\mathbf{b}\in\mathscr{V}$，又有 $\mathbf{a}=\sum_{i=1}^n\alpha_i\widehat{\mathbf{e}}_i, \mathbf{b}=\sum_{i=1}^n\beta_i\widehat{\mathbf{e}}_i, n\equiv\dim\mathscr{V}$，则 $(\mathbf{a}|\mathbf{b})=\sum_{i=1}^n\alpha_i\overline{\beta_j}, (\mathbf{Qa}|\mathbf{Qb})=\left(\sum_{i=1}^n\mathbf{Q}\widehat{\mathbf{e}}_i|\sum_{j=1}^n\mathbf{Q}\widehat{\mathbf{e}}_j\right)=\sum_{i=1}^n\alpha_i\overline{\beta_j}=(\mathbf{a}|\mathbf{b})$，故 $\mathbf{Q}$ 保持内积。

1)⇒5)：设 $\mathbf{Q}$ 是幺正算符，则 $\mathbf{Q}$ 可逆，且 $(\mathbf{Qa}|\mathbf{b})=\left(\mathbf{Qa}|\mathbf{QQ}^{-1}\mathbf{b}\right)=\left(\mathbf{a}|\mathbf{Q}^{-1}\mathbf{b}\right), \forall\mathbf{a},\mathbf{b}\in\mathscr{V}$。因此 $\mathbf{Q}^{-1}=\mathbf{Q}^*$ 即 $\mathbf{QQ}^*=\mathbf{I}=\mathbf{Q}^*\mathbf{Q}$。

设 $\mathbf{Q}^*\mathbf{Q}=\mathbf{QQ}^*=\mathbf{I}$，则 $\mathbf{Q}^{-1}=\mathbf{Q}^*, (\mathbf{Qa}|\mathbf{Qb})=(\mathbf{a}|\mathbf{Q}^*\mathbf{Qb})=(\mathbf{a}|\mathbf{Ib})=(\mathbf{a}|\mathbf{b}), \forall\mathbf{a},\mathbf{b}\in\mathscr{V}$。 □

上述定理告诉我们，幺正算符必可逆，且幺正算符的逆算符也是幺正算符。

如果在内积空间 $\mathscr{V}$ 上还定义了欧几里得范 $\|\mathbf{a}\|^2\equiv(\mathbf{a}|\mathbf{a}), \forall\mathbf{a}\in\mathscr{V}$，那么当 $\mathbf{Q}\in\mathscr{L}(\mathscr{V})$ 是一个幺正算符时，就有 $\|\mathbf{Qa}\|=\|\mathbf{a}\|, \forall\mathbf{a}\in\mathscr{V}$，即幺正算符不改变向量的“长度”。如果一个算符作用于一个向量后，不改变其“长度”，那就只剩改变其“方向”的效果了。这强烈地暗示了幺正算符的几何意义，将在后面正式介绍。

此外还易证，两个幺正算符的复合也是幺正算符。特别地，恒等算符 $\mathbf{I}$ 本身就是一个幺正算符。

定理2.33的第 5 条，常被作为幺正算符的定义。在以往所学习的矩阵代数中，有“正交矩阵”的概念。如果数域 $\mathbb{F}$ 上的 $n\times n$ 矩阵 $A\in\mathbb{F}^{n\times n}$ 满足 $A^\intercal A=I$，其中 I 是 $n\times n$ 单位矩阵，则称矩阵 A 是*正交矩阵*（*orthogonal matrix*）。对于抽象的线性算符，我们知道一个线性算符与其伴随算符在给定基下的坐标矩阵之间的关系，在实数域上，就是矩阵转置。因此2.33的第 5 条在实数域上相当于说幺正算符在给定基下的坐标矩阵是一个正交矩阵。因此，我们又把实数域上的幺正算符称为*正交算符*（*orthogonal operator*）。

现在我们说说定理2.33的第 3、4 条。回顾线性变换的坐标变换公式（定理2.22）。在那里，我们只考虑在两组一般的基之间的变换公式。现设 $B=\{\widehat{\mathbf{e}}_i\}$、$B'=\{\widehat{\mathbf{e}}'_i\}$ 是数域 $\mathbb{F}$ 上的 n 维内积空间 $\mathscr{V}$ 的两组*规范正交基*。Q 是由 B 到 B' 的过渡矩阵，即 Q 满足 $\widehat{\mathbf{e}}'_j=\sum_{i=1}^n Q_{ij}\widehat{\mathbf{e}}_i,\quad j=1,\cdots,n$。如果 Q 是某线性算符 $\mathbf{Q}$ 在基 $\mathbf{B}$ 下的坐标矩阵，那么 $\mathbf{Q}$ 满足定理2.33的第 3 条，于是 $\mathbf{Q}$ 是幺正算符且满足定理2.33的其他几条命题。对于 $\mathscr{V}$ 上的另一任意线性算符 $\mathbf{T}$，我们有坐标变换公式：$(\mathbf{T})_{B'}=Q(\mathbf{T})_B Q^{-1}$。但由于 $\mathbf{Q}$ 是幺正算符，$QQ^*=I$，因此有 $(\mathbf{T})_{B'}=Q(\mathbf{T})_B Q^*$。

我们再联系定理2.32的第 4 条：若 $\mathbf{T}$ 是厄米算符，则 $\mathscr{V}$ 中存在一组规范正交基——不妨记为 B——恰为 $\mathbf{T}$ 的特征向量。因此，任一厄米矩阵 $T\equiv(\mathbf{T})_{B'}$ 与它的对角矩阵 $D\equiv\operatorname{diag}(\lambda_1,\cdots,\lambda_n)$ 之间就相差一个形如 $T=QDQ^*$ 的形式，其中 Q 是一个幺正矩阵[3]§5.3 定理 3.6。

我们把对某一矩阵进行对角化时所使用的可逆矩阵同时是幺正矩阵的情况称作*幺正地对角化*（*unitarily diagonalization*）。因此可以说，可逆厄米算符跟一般可对角化的算符的区别是：与后者相比，前者不仅是可对角化的，还是可幺正地对角化的。特别地，在有限维实内积空间上，一个线性算符是对称算符当且仅当该内积空是中有一组规范正交基是该算符的特征向量。也就是说，定理2.32命题“4”在复数域上作为厄米算符时逆命题不成立，但在实际数域上作为对称算符时逆命题也成立。

最后我们关心一下幺正算符在特征值分解时的性质*——

定理 2.34. 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的 n 维内积空间，$\mathbf{Q}$ 是 $\mathscr{V}$ 上的一个线性算符，$\{\lambda_1,\cdots,\lambda_n\}$ 是其特征值。若 $\mathbf{Q}$ 是幺正算符，则 $|\lambda_i|=1, i=1,\cdots,n$。

证明. 设 λ 是 $\mathbf{Q}$ 的任一特征值，且 $\mathbf{c}$ 是 $\mathbf{Q}$ 关于这一特征值的任一特征向量。由幺正算符定义有

$$(\mathbf{c}|\mathbf{c}) = (\mathbf{Qc}|\mathbf{Qc}) = \lambda\bar{\lambda}\,(\mathbf{c}|\mathbf{c})$$

由于特征向量都是非零向量，故 $(\mathbf{c}|\mathbf{c})>0$，上式成立只需 $\lambda\bar{\lambda}=1 \Leftrightarrow |\lambda|=1$。 □

幺正算符可对角化。但这一命题的证明在引入了*正规算符*（*normal operator*）之后再证明更加直接。本讲义不打算详细介绍正规算符。在此简要提及其定义和最重要的性质之一。

正规算符

定义 2.24 (正规算符)**.** 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的内积空间，$\mathbf{T}\in\mathscr{L}(\mathscr{V})$ 是一个线性算符，如果 $\mathbf{TT}^*=\mathbf{T}^*\mathbf{T}$，则称 $\mathbf{T}$ 是*正规算符*（*normal operator*）。

很容易验证，厄米算符和幺正算符都是正规算符。在复数域上，正规算符必可对角化，因为如下的定理

定理 2.35. 设 $\mathscr{V}$ 是复数域 $\mathbb{C}$ 上的内积空间，$\mathbf{T}\in\mathscr{L}(\mathscr{V})$ 是一个正规算符，则 $\mathscr{V}$ 中必有一组规范正交基是 $\mathbf{T}$ 的特征向量。

该定理的证明路径较长，此略[4]§8.5 Theorem 22。但值得注意的是定理2.32“4”已证明该结论至少对厄米算符是成立的。

---

*在大一的线性代数课中只介绍了考虑实数域上的正交矩阵的对角化[3]§5.3 定理 3.7，得到一系列性质。但是这些性质需要在复数域的更一般情况下得到确认。

结合此定理与定理2.28可知，在复数域上，正规算符必可幺正地对角化。而由定理2.32的推论，在实数域上，对称算符必可正交地对角化。

其逆命题也成立，即在复数域 $\mathbb{C}$ 上，可幺正地对角化的算符必为正规算符。因为说一个线性算符 $\mathbf{T}$ 可幺正地对角化，就是说存在一组规范正交基 B 满足 $(T)_B = \operatorname{diag}(\lambda_1, \cdots, \lambda_n)$，其中 $\{\lambda_i\}$ 是 $\mathbf{T}$ 的特征值。若 Q 是从 B 到另一规范正交基 B' 的过渡矩阵，则 $(\mathbf{T})_{B'} = Q(\mathbf{T})_B Q^{-1}$。易验 $(\mathbf{T})_{B'} \overline{(\mathbf{T})^{\intercal}_{B'}} = \overline{(\mathbf{T})^{\intercal}_{B'}} (\mathbf{T})_{B'}$，亦即 $\mathbf{T}\mathbf{T}^* = \mathbf{T}^*\mathbf{T}$，故 $\mathbf{T}$ 为正规算符。

还有两个定理不是关于正规算符的，但却需要引入正规算符及其谱定理才可证明，本讲义无法详细介绍。不过，这两个定理使得复数域上的线性算符很像复数。比如，厄米算符的所有特征值都是实数，因此类比于“实数”，我们可以给厄米算符定义“正”或“非负”的概念如下。

定义 2.25. 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的有限维内积空间，$\mathbf{T}$ 是 $\mathscr{V}$ 上的自伴随算符。

1. 如果 $(\mathbf{Ta}|\mathbf{a}) \leq 0, \quad \forall \mathbf{a} \in \mathscr{V}$，则称 $\mathbf{T}$ 为非负（*non-negative*）算符；
2. 如果 $(\mathbf{Ta}|\mathbf{a}) > 0, \quad \forall \mathbf{a} \in \mathscr{V}$，则称 $\mathbf{T}$ 为正（*positive*）算符。

以下定理类似“非负实数的平方根是非负实数”。

定理 2.36. 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的有限维内积空间，$\mathbf{T} \in \mathscr{L}(\mathscr{V})$ 是一个非负算符，则必存在唯一非负算符 $\mathbf{N} \in \mathscr{L}(\mathscr{V})$ 满足 $\mathbf{T} = \mathbf{N}^2$。

以下定理则类似于：任一复数 z 可分解为 $z = \rho e^{i\theta}$，其中 ρ 是非负实数，表示伸缩；$\left|e^{i\theta}\right| = 1$，表示旋转。对于复数，这叫“极坐标分解”。对于线性算符，这叫极分解。

定理 2.37 (极分解)**.** 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的有限维内积空间，$\mathbf{T} \in \mathscr{L}(\mathscr{V})$ 是一个线性算符，则总存在幺正算符 $\mathbf{Q}$ 和唯一一个非负算符 $\mathbf{N}$ 满足 $\mathbf{T} = \mathbf{UN}$，称为 $\mathbf{T}$ 的一个极分解（*polar decomposition*）。如果 $\mathbf{T}$ 可逆，则连 $\mathbf{Q}$ 也是唯一的。如果 $\mathbf{T}$ 是正规算符，则 $\mathbf{QN} = \mathbf{NQ}$。

值得注意的是，对于非正规算符 $\mathbf{T}$，可极分解为 $\mathbf{T} = \mathbf{N}_1\mathbf{Q} = \mathbf{Q}\mathbf{N}_2$，且一般地 $\mathbf{N}_1 \neq \mathbf{N}_2$。在连续介质力学部分，讲到形变梯度张量的时候，就需要用到极分解定理。

第三章　欧几里得空间

在经典力学中, 我们总假定物理事件发生欧几里得空间中。在本节我们将以集合的语言重新引入欧几里得空间。

3.1　欧几里得空间的构建

为了建立一个几何空间，我们需要先让一个集合的两个元素间有“距离”的概念。

定义 3.1 (度量空间). 设 $\mathscr{E}$ 是一个非空集合，如果映射 $d:\mathscr{E}\times\mathscr{E}\to[0,+\infty)\subset\mathbb{R}$ 满足

1. 不可区分者的同一性：$d(x,y)=0\Leftrightarrow x=y,\quad\forall x,y\in\mathscr{E}$
2. 对称性：$d(x,y)=d(y,x),\quad\forall x,y\in\mathscr{E}$
3. 三角不等式：$d(x,z)\le d(x,y)+d(y,z),\quad\forall x,y,z\in\mathscr{E}$

则称 d 是定义在 $\mathscr{E}$ 上的一个度量（*metric*），有序对 $(\mathscr{E},d)$ 是一个度量空间（*metric space*）。

由定义易证，度量总是非负的，即 $d(x,y)\ge 0,\forall x,y\in\mathscr{E}$。由第 3 条，$d(x,y)+d(y,x)\ge d(x,x)$；再由第 2 条，$d(x,y)+d(x,y)\ge d(x,x)$；最后由第一条有 $2d(x,y)\ge 0\Rightarrow d(x,y)\ge 0$（当且仅当 $x=y$ 时取等号）。由于这是定义3.1中的规定能够推出的，因此就算它是我们对距离的最直观要求，但却无需写进定义3.1中。

例 3.1. *1. 设 $M=\{0,1\}$，定义 $d(x,y),\forall x,y\in M$*

$$d(x,y)=\begin{cases}1, & x\neq y\\ 0, & x=y\end{cases}$$

则 (M,d) 是一个度量空间。把 M 改成实数集 $\mathbb{R}$，结合上列定义的 d，$(\mathbb{R},d)$ 也是一个度量空间。

2. 数域 $\mathbb{F}$ 上的赋范向量空间 $\mathscr{V}$，连同 $d(\mathbf{a},\mathbf{b})\equiv\|\mathbf{a}-\mathbf{b}\|$，构成一个度量空间 $(\mathscr{V},d)$。

3. 在实区间 $[0,\infty)$ 上定义的度量 $d(x,y)=|x-y|$ 形成一个度量空间。

易验，若 (M,d) 是一个度量空间，N 是 M 的一个子集，则 (N,d) 也是一个度量空间。

度量空间之间的同态映射将保持距离。具体地，

定义 3.2 (等距变换). 设 $(A, d_A), (B, d_B)$ 是两个度量空间，若映射 $i: A \to B$ 满足

$$d_B(i(a), i(b)) = d_A(a, b), \forall a, b \in A$$

则称 i 是由 (A, d_A) 到 (B, d_B) 的一个保距映射（*distance-preserving mapping*）；若 i 是双射，则称 i 是由 (A, d_A) 到 (B, d_B) 的一个等距变换（*isometry*）。若两个度量空间之间可以定义出至少一个等距变换，则称这两个度量空间是等距的（*isometric*）。

所有保距映射都是单射，因为当 $a = b, d_B(i(a), i(b)) = d_A(a, b) = 0 \Rightarrow i(a) = i(b)$。因此也可以说，等距变换是满射的保距映射。

一个集合 M 上可以定义不止一种度量映射，而形成不同的度量空间。两个度量空间之间也可以存在不止一个保距映射或等距变换。由一个度量空间到另一个度量空间的一个映射是否保距映射或等距变换，依赖这两个度量空间的度量定义。

例 3.2. 等距变换的一些例子：

1. 给定两个度量空间：

$$\left(\mathbb{R}^+, d_1\right), d_1(x, y) = |\log x - \log y|, \quad \forall x, y \in \mathbb{R}^+$$

和

$$(\mathbb{R}, d_2), d_2(x, y) = |x - y| \, \forall x, y \in \mathbb{R}$$

则映射 $\log: \mathbb{R}^+ \to \mathbb{R}$ 是这两个度量空间的一个等距变换。

2. $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的赋范向量空间，度量空间 $(\mathscr{V}, \|\cdot\|)$ 到 $(\mathscr{V}, \|\cdot\|_r)$ 的映射 $i: \mathscr{V} \to \mathscr{V}, i(\mathbf{a}) = r^{-1}\mathbf{a}, \quad \forall \mathbf{a} \in \mathscr{V}$ 是一个等距变换，其中 $\|\cdot\|_r = r\|\cdot\|, r \in \mathbb{F}$。

3. 接着例3.1中的第 1 个例子，$\mathscr{V}$ 上的幺正算符是该例的度量空间中的等距变换。

4. 接着例3.1中的第 2 个例子，定义函数 $f(x) = x + 1, \forall x \in [0, \infty)$，则 f 是该例的度量空间中的一个等距变换。

为了建立一个几何空间，我们特别关心的是从一个度量空间 (M, d) 到其自身的等距变换（度量空间的自同态映射）。特别地，若 (M, d) 上的所有等距变换的集合 $\mathcal{I}$ 满足：

G1 封闭性：$\forall i_1, i_2 \in \mathcal{I}, \quad i_1 \circ i_2 \in \mathcal{I}$

G2 结合律：$\forall i_1, i_2, i_3 \in \mathcal{I}, \quad , (i_1 \circ i_2) \circ i_3 = i_1 \circ (i_2 \circ i_3)$

G3 单位元：集合 M 上的恒等映射 $\mathrm{id}_M: M \to M, \quad \mathrm{id}_M(x) = x, \forall x \in M$ 也是等距变换，即 $\mathrm{id}_M \in \mathcal{I}$，且满足 $\forall i \in \mathcal{I}, \quad \mathrm{id}_M \circ i = i \circ \mathrm{id}_M = i$

G4 逆元：$\forall i \in \mathcal{I} \exists i^{-1} \in \mathcal{I} \quad i^{-1} \circ i = \mathrm{id}_M$

则称 $\mathcal{I}$ 是度量空间 (M, d) 的等距群（*isomtric group*）。事实上，条件 G1 至 G4 是群的一般定义——

定义 3.3 (群)**.** 设 G 是一个非空集合，若为 G 的元素规定一个二元运算（binary operation），记为 $x \circ y, \quad x, y \in G$[*]，且该运算具有以下性质：

1. 封闭性：$x \circ y \in G, \quad \forall x, y \in G$；
2. 结合律：$(x \circ y) \circ z = x \circ (y \circ z), \quad \forall x, y, z \in G$；
3. 单位元：$\exists e \in G, \quad e \circ x = x \circ c = x, \quad \forall x \in G$；
4. 逆元：$\forall x \in G, \exists x^{-1} \in G, xx^{-1} = x^{-1}x = e$。

因此，一个度量空间上的等距群是由该度量空间上的等距变换基于映射的复合运算形成的群。

易证，群的单位元和逆元总是唯一的，这也跟向量空间的情况类似。一般地，群的定义不包括交换律。满足交换率的群叫*交换群*（*commutative group*）。交换群的二元操作比较像平时的“加法”。向量空间的定义中，若去除标量乘的规定，剩下的规定实际上定义了一个交换群。

例 3.3. *设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的 n 维向量空间，$\mathscr{V}$ 上的所有可逆算符的集合，基于算符的复合操作，形成一个群。我们称这个群为数域 $\mathbb{F}$ 上的 n 维*一般线性群（generalized linear group），*记作 $GL(n, \mathbb{F})$；在所讨论的数域维持不变时，简记为 $GL(n)$。*

为了使空间两点间不仅有距离的概念，还能有“有向线段”的概念（即以往我们所习惯的，用“向量”来描述几何对象的数学语言），我们需要从一个度量空间的等距群中定义出一个向量空间。因此首先要使一个度量空间上的等距群或等距子群成为一个交换群。

设 $(\mathscr{E}, d)$ 是一个度量空间，$\mathcal{I}$ 是 $(\mathscr{E}, d)$ 上的等距群，若 $\mathcal{I}$ 的子群 $\mathscr{V}$ 满足：

G5 交换律：$i_1 \circ i_2 = i_2 \circ i_1, \quad \forall i_1, i_2 \in \mathscr{V}$；

G6 $\mathscr{V}$ 在 $\mathscr{E}$ 上的*作用*（*action*）的传递性，即对 $\mathscr{E}$ 中的任一点 X 和一点 Y，总存在 $\mathcal{I}$ 中的一个等距变换 i 满足 $Y = i(x)$。

则 $\mathscr{V}$ 是 $\mathcal{I}$ 的一个交换子群。一般地，度量空间上的等距群 $\mathcal{I}$ 的满足条件 G5 至 G6 的子群 $\mathscr{V}$ 最多只有一个。

接下来我们赋予 $\mathscr{V}$ 以实数域上的标量乘法规定，即使 $\mathscr{V}$ 具有满足以下规定的形式运算：

S1 $\forall \alpha \in \mathbb{R}, i \in \mathscr{V}, \alpha i \in \mathscr{V}$。特别地，$\forall i \in \mathscr{V}, 1i = i$；

S2 $\forall \alpha, \beta \in \mathbb{R}, i \in \mathscr{V}, \quad \alpha(\beta i) = (\alpha\beta) i$；

S3 $\forall \alpha \in \mathbb{R}, i, j \in \mathscr{V}, \quad \alpha(i \circ j) = (\alpha i) \circ (\alpha j)$

S4 $\forall \alpha, \beta \in \mathbb{R}, i \in \mathscr{V}, \quad (\alpha + \beta) i = (\alpha i) \circ (\beta i)$

带有上述规定的 $\mathscr{V}$ 形成一个实数域 $\mathbb{R}$ 上的向量空间。

[*]跟复合映射的符号一样，可算是一种推广意义的使用，因为映射的复合就是一种二元运算。

向量空间 $\mathscr{V}$ 的向量作为度量空间 $(\mathscr{E},d)$ 上的等距变换，可天然具有一种范的定义。由条件 G5 和 G6 易验 $\forall i\in\mathscr{V},X,Y\in\mathscr{E},d\left(X,i\left(X\right)\right)=d\left(Y,i\left(Y\right)\right)$，故每个等距变换自带一个特定的长度。可定义 $\mathscr{V}$ 上的范 $\|\cdot\|:\mathscr{V}\to[0,\infty)\subset\mathbb{R}$,

$$\forall i\in\mathscr{V},\|i\|=d\left(X,i\left(X\right)\right),X\in\mathscr{E}$$

再附加一条运算的形式规定：

N1 调和性：$\forall\alpha\in\mathbb{R},i\in\mathscr{V},\|\alpha i\|=|\alpha|\,\|i\|$

则易验该范满足范的定义。再通过极化恒等式构建 $\mathscr{V}$ 上的内积 $(\cdot|\cdot):\mathscr{V}^2\to\mathbb{R}$,

$$(i|j)=\frac{1}{4}\|i\circ j\|^2-\frac{1}{4}\left\|i\circ j^{-1}\right\|^2$$

则 $\mathscr{V}$ 形成一个实数域上的内积空间，$\|\cdot\|$ 是 $\mathscr{V}$ 上的欧几里得范。

引理A.4告诉我们，如果一个度量空间上的等距群存在一个满足条件 G1 至 G6、S1 至 S4 和 N1 的子群，则这样的子群只有一个。我们就已经做好正式构建欧几里得空间的准备。

定义 3.4 (欧几里得空间)**.** 若一个度量空间 $(\mathscr{E},d)$ 上的度量 d 所确定的等距群可定义出满足条件 G1 至 G6、S1 至 S4 和 N1 的子群 $\mathscr{V}$，则 $\mathscr{V}$ 自然形成一个实数域上的内积空间。若 $\mathscr{V}$ 是有限维内积空间，则称 $(\mathscr{E},d,\mathscr{V})$ 为欧几里得空间（*Euclidean space*）。$\mathscr{E}$ 称为欧几里得空间的点空间（*point space*），$\mathscr{E}$ 的元素称为点（*points*）；$\mathscr{V}$ 称为欧几里得空间的平移空间（*translation space*），$\mathscr{V}$ 中的向量称为欧几里得空间的平移向量（*translation vectors*）；$\mathscr{V}$ 的维数就是欧几里得空间的维数。我们常简记欧几里得空间为 $\mathscr{E}$。

最后，我们建立一个记法。设 $\mathscr{E}$ 是欧几里得空间，$X,Y\in\mathscr{E}$，由 $X,Y\in\mathscr{E}$ 确定的平移向量 $\mathbf{u}$ 可记作 $\mathbf{u}=Y-X$；由点 X 经平移向量 $\mathbf{u}$（它是一个等距变换）平移为点 Y 的事实可记作 $Y=X+\mathbf{u}$。注意，我们没有定义两个点“相加”的意义。

3.2 角、直线、位置向量和坐标系

本节我们依次在欧几里得空间中引入角、直线、位置向量和坐标系。

为了引入角，我们考虑欧几里得空间 $\mathscr{E}$ 中的给定三个不同的点 $X,O,Y\in\mathscr{E}$，由于 $\mathscr{E}$ 的平移空间 $\mathscr{V}$ 是一个赋范内积空间，故有极化恒等式，

$$\begin{aligned}\|X-O\|^2+\|Y-O\|^2&=\|(X-O)-(Y-O)\|^2+2\left(X-O|Y-O\right)\\&=\|X-Y\|^2+2\left(X-O|Y-O\right)\end{aligned}$$

再应用柯西–施瓦茨不等式，有

$$\|X-O\|^2\,\|Y-O\|^2 \geq |(X-O|Y-O)|^2 \Leftrightarrow -1 \leq \frac{(X-O|Y-O)}{\|X-O\|\,\|Y-O\|} \leq 1$$

我们就做好了引入角的准备。

定义 3.5 (角). 设 $\mathscr{E}$ 是欧几里得空间，$\mathscr{E}$ 中的角是一个映射 $\angle:\mathscr{E}^3\to\mathbb{R}$ 满足

$$\angle XOY \equiv \cos^{-1}\frac{(X-O|Y-O)}{\|X-O\|\,\|Y-O\|},\quad \forall X,O,Y\in\mathscr{E}, X\neq O\neq Y$$

称点 XOY 所夹的角，或简称角 XOY。其中余弦函数 $\cos:[0,\pi]\to\mathbb{R}$ 定义为

$$\cos(x)=\frac{1}{2}\left(e^{ix}+e^{-ix}\right),\quad \forall x\in[0,\pi)$$

注意到，上述定义中的余弦函数是一个双射，故其逆映射 $\cos^{-1}$ 也是双射，所以上述定义的角的取值范围是 $\operatorname{ran}\angle=[0,\pi]$。同时 $\angle XOY$ 的顺序是重要的，$\angle YOX=-\angle XOY$。

设 $i:\mathscr{E}\to\mathscr{E}$ 是一个等距变换，可验证 $\angle i(X)\,i(O)\,i(Y)=\angle XOY,\forall X,O,Y\in\mathscr{E},X\neq O\neq Y$，即等距变换前后角不变。

定义 3.6 (过两点的直线). 设 $(\mathscr{E},d)$ 是欧几里得空间，给定两点 $X,Y\in\mathscr{E},X\neq Y$，则 $\mathscr{E}$ 的子集 $L_{XY}=\{C\in\mathscr{E}|\forall\alpha(\alpha)\in\mathbb{R}\wedge C=X+\alpha(Y-X))\}$ 是过 X,Y 两点的一条直线。如果 $\angle XOY=\frac{\pi}{2}$，则直线 L_{OX} 与 L_{OY} 垂直，记为 $L_{OX}\perp L_{OY}$。

由角的定义，如果 $L_{OX}\perp L_{OY}$，则 $(X-O|Y-O)=0$。再由内积空间的格拉姆–施密特正交化过程可知，过 $\mathscr{E}$ 中任一点 O 的两两垂直的直线最大条数都相等且等于 $\dim\mathscr{V}$，故欧几里得空间的维数就可被自然地定义为其平移空间的维数。

L_{XY} 又可记为 $L_{XY}=\{C\in\mathscr{E}|\forall\alpha(\alpha\in\mathbb{R}\wedge C-X=\alpha(Y-X))\}$，它对应着平移向量空间 $\mathscr{V}$ 的子集 $L_{XY}^{\mathscr{V}}=\{\mathbf{u}\in\mathscr{V}|\forall\alpha(\alpha\in\mathbb{R}\wedge\mathbf{u}=\alpha(X-Y))\}$，易知该子集是 $\mathscr{V}$ 的子空间，维数是 1*。

我们将一个选定的原点 $O\in\mathscr{E}$ 和 $\mathscr{V}$ 的一组规范正交基的组合 $(O,\{\hat{\mathbf{e}}_i\})$ 称为欧几里得空间 $\mathscr{E}$ 的一个直角坐标系（*rectangular coordinates*），又称笛卡尔坐标系（*Cartesian coordinates*）。我们常常默认一个 n 维欧几里得空间必然已经自带一个直角坐标系，称为基本坐标系（*common coordinates*），从而直接采用 $\mathbb{R}^n$ 来表示任意一点的坐标。在基本坐标系下，原点坐标为 $(0,\cdots,0)$，第 i 个基向量为 $(0,\cdots,1,\cdots,0)^\intercal$，也就是除第 i 个分量为 1 外其他分量均为零的有序实数 n 元组。选定了原点 O 后，对任一点 $X\in\mathscr{E}$ 可定义映射 $\mathbf{r}_O:\mathscr{E}\to\mathscr{V},\mathbf{r}(X)\equiv\mathbf{r}_X=X-O,\forall X\in\mathscr{E}$，我们称这个向量值函数 $\mathbf{r}_X$ 就是选定原点 O 下点

*这里需要实数集的完备性概念。

X 的位置向量（*position vector*）。注意，当且仅当选定了原点后，欧几里得空间 $\mathscr{E}$ 中的点才与其平移空间 $\mathscr{V}$ 的向量通过位置向量这个映射一一对应。

总而言之，一个（有限维）欧几里得空间 $(\mathscr{E}, d, O, \mathscr{V}, \{\hat{\mathbf{e}}_i\})$ 包括：

1. 一个度量空间 $(\mathscr{E}, d)$
2. 一个实数域上的 n 维内积空间 $\mathscr{V}$。它是由度量空间 $(\mathscr{E}, d)$ 上的等距群 $\mathcal{I}$ 经过 G1 至 G6、S1 至 S4 和 N1 构建的。n 同时定义了该欧几里得空间的维数。我们还规定了记法：

$$\forall (X, Y) \in \mathscr{E}^2 \exists \mathbf{u} \in \mathscr{V}, \quad \mathbf{u} = Y - X$$

3. 选定了原点 $O \in \mathscr{E}$
4. 选定了一组规范正交基 $\{\hat{\mathbf{e}}_i\} \subset \mathscr{V}$

   最后两项同时也使欧几里得空间默认带有一个直角坐标系，$\mathscr{V}$ 的向量是 $\mathscr{E}$ 的点在此直角坐标系下的位置向量。

明确了欧几里得空间的完整概念之后，为了简便我们仍只用 $\mathscr{E}$ 表示一个欧几里得空间。

基于《几何原本》的公设得到的大量欧氏几何定理仍然成立，因为这些公设的要求已蕴含在了实数域的、向量内积的性质以及度量的性质中了[6]。更重要的是，明确了这一线性结构后，我们能够用统一的数学语言推导出更多几何结论[7]。

3.3 几种线性算符在 3 维欧几里得空间中的几何意义

实数域上的内积空间上的厄米算符、反厄米算符和幺正算符分别特称为对称算符、斜称算符和正交算符。它们的定义方式与其在复数域上的对应是一样的，但它们的部分性质与其在复数域上的对应不同。最突出的表面是那些通过特征多项式和行列式得出的性质。

我们在上一节看到，n 维欧几里得空间的平移空间是一个实数域 $\mathbb{R}$ 上的 n 维内积空间；欧几里得空间上的几何规律已经由向量内积的性质和欧几里得范的性质得以确定。当实数域 $\mathbb{R}$ 上的一个 n 维内积空间 $\mathscr{V}$ 是 n 维欧几里得空间 $\mathscr{E}$ 的平移空间时，$\mathscr{V}$ 上的对称算符、斜称算符和正交算符对 $\mathscr{V}$ 中的向量作用的结果，将对应于欧几里得空间中的几何形状的变换。

本节我们讨论这三种算符在实内积空间上的特殊性质和由此得出的几何意义。

3.3.1 正交算符

首先，通过与定理2.34类似的证明方法可知，在 n 维实内积空间上，正交算符的行列式要么是 1 要么是 -1。于是正交算符可按此分为两类，需要分开讨论。

我们先在 2 维的情况上认识正交算符。

2 维欧几里得平面上的几何意义

设 $\mathscr{V}$ 是实数域 $\mathbb{R}$ 上的 2 维内积空间，$\mathbf{Q}$ 是 $\mathscr{V}$ 上的一个正交算符，其在给定某一有序规范正交基下的坐标矩阵是

$$(\mathbf{Q}) = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}$$

则由线性算符在给定基下的坐标计算式、正交算符在内积运算中的性质、以及正交算符行列式的性质可得

$$\begin{aligned} Q_{11}^2 + Q_{12}^2 = Q_{21}^2 + Q_{22}^2 = 1 \\ Q_{11}Q_{21} + Q_{12}Q_{22} = 0 \\ Q_{11}Q_{22} - Q_{12}Q_{21} = \pm 1 \end{aligned}$$

其中“$\pm$”表示当 $\det\mathbf{Q} = 1$ 时取正号，当 $\det\mathbf{Q} = -1$ 时取负号。下同。

设

$$Q_{11} = \cos\theta, \quad Q_{12} = \sin\theta, \quad Q_{21} = \cos\phi, \quad Q_{22} = \sin\phi$$

则有

$$\begin{aligned} Q_{11}Q_{21} + Q_{12} + Q_{22} = 0 \Leftrightarrow \cos(\phi - \theta) = 0 \\ Q_{11}Q_{22} - Q_{12}Q_{21} = \pm 1 \Leftrightarrow \sin(\phi - \theta) = \pm 1 \end{aligned}$$

即 $\phi = \theta \pm \pi/2$。因此，给定某一有序规范正交基，2 维实内积空间上的对称算符的坐标矩阵总可由某角 θ 表示成以下形式

$$(\mathbf{Q}) = \begin{pmatrix} \cos\theta & \sin\theta \\ \mp\sin\theta & \pm\cos\theta \end{pmatrix}$$

当 $\det\mathbf{Q} = 1$ 时，采用角度 θ 的矩阵表达式是有明快的几何意义的。设在相同的有序规范正交基下，某向量 $\mathbf{a}$ 的坐标是 (α_1, α_2)，则 $\mathbf{Qa}$ 的坐标由下式计算得到

$$(\mathbf{Qa}) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \alpha_1\cos\theta + \alpha_2\sin\theta \\ -\alpha_1\sin\theta + \alpha_2\cos\theta \end{pmatrix}$$

由 $\mathbf{Q}$ 的性质，$\|\mathbf{Qa}\| = \|\mathbf{a}\|$，故我们仅需关心 $\mathbf{Qa}$ 与 $\mathbf{a}$ 的夹角 ω，由角的定义，利用矩阵运算，可得

$$\cos\omega = \frac{(\mathbf{Qa}|\mathbf{a})}{(\mathbf{a}|\mathbf{a})} = \cos\theta$$

注意到，在欧几里得空间中的角的定义中的余弦函数是双射，故上式 $\Leftrightarrow \omega=\theta$。换言之，当 $\det\mathbf{Q}=1$ 时，任一向量 $\mathbf{a}$ 在被 $\mathbf{Q}$ 作用之后，长度不变，方向变化角度 θ（如图3.1）。若向量 $\mathbf{a}$ 是给定了原点后点 A 的位置向量，则被 $\mathbf{Q}$ 作用后，点 A 到原点的距离不变，绕原点旋转角度 θ。特别地，当 $\theta=0$ 时，$\mathbf{Q}=\mathbf{I}$。

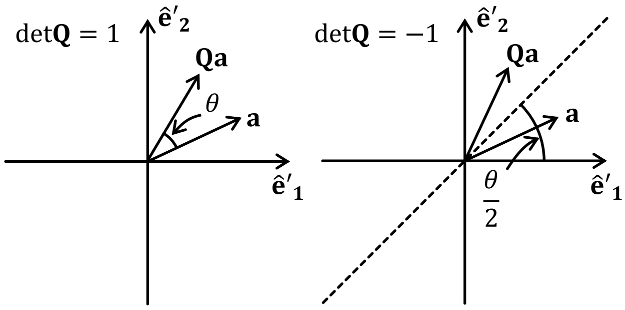


图 3.1: 正交算符在 2 维欧几里得空间中的几何意义。

接下来我们考虑 $\det\mathbf{Q}=-1$ 时的情况。首先，$\mathbf{Q}$ 的两个特征值是 1 和 -1。这相当于说 $\det(\mathbf{Q}\pm\mathbf{I})=0$。诚然，由算符的伴随的计算性质（定理2.30），$(\mathbf{Q}\pm\mathbf{I})^*=\mathbf{Q}^*\pm\mathbf{I}$。由行列式的性质，$\det(\mathbf{Q}\pm\mathbf{I})^*=\det(\mathbf{Q}\pm\mathbf{I})$。再由

$$\begin{aligned}\det(\mathbf{Q}\pm\mathbf{I})&=\det(\mathbf{Q}\pm\mathbf{Q}^*\mathbf{Q})\\&=\det(\mathbf{Q})\det(\mathbf{I}\pm\mathbf{Q}^*)\\&=(-1)(-1)^2\det(\mathbf{Q}^*\pm\mathbf{I})\\&=-\det(\mathbf{Q}^*\pm\mathbf{I})\end{aligned}$$

可得 $\det(\mathbf{Q}\pm\mathbf{I})=-\det(\mathbf{Q}\pm\mathbf{I})\Leftrightarrow\det(\mathbf{Q}\pm\mathbf{I})=0$。

我们可进一步得到结论：行列式为 -1 的正交算符是对称算符。诚然，当 $\det\mathbf{Q}=-1$ 时，设 $\widehat{\mathbf{e}}_1,\widehat{\mathbf{e}}_2$ 分别是对应其特征值 1 和 -1 的单位特征向量，则 $(\widehat{\mathbf{e}}_1|\widehat{\mathbf{e}}_2)=(\mathbf{Q}\widehat{\mathbf{e}}_1|\mathbf{Q}\widehat{\mathbf{e}}_2)=-(\widehat{\mathbf{e}}_1|\widehat{\mathbf{e}}_2)\Leftrightarrow(\widehat{\mathbf{e}}_1|\widehat{\mathbf{e}}_2)=0$。再由定理2.32的推论可得 $\mathbf{Q}$ 是对称算符。

下面我们开始考察行列式为 -1 的对称算符的几何性质。沿用上一段的设定，在规范正交基 $B=\{\widehat{\mathbf{e}}_1,\widehat{\mathbf{e}}_2\}$ 下，$\mathbf{Q}$ 的坐标矩阵将是以下对角矩阵

$$(\mathbf{Q})_B=\begin{pmatrix}1&0\\0&-1\end{pmatrix}$$

但 B 是由 $\mathbf{Q}$ 的性质规定的，一般不会恰巧就是我们讨论几何问题时选用的坐标系的基。设我们选用的是另一个有序规范正交基 $B'=\{\widehat{\mathbf{e}}_1',\widehat{\mathbf{e}}_2'\}$，$P$ 是由 $\mathbf{B}$ 到 $\mathbf{B}'$ 的过渡矩阵，则 P 也是一个正交矩阵。按照之前的结论，矩阵 P 可以某角度 ϕ 表示成

$$P=\begin{pmatrix}\cos\phi & \sin\phi\\ \mp\sin\phi & \pm\cos\phi\end{pmatrix}$$

事实上 ϕ 是 $\widehat{\mathbf{e}}_1'$ 与 $\widehat{\mathbf{e}}_1$ 的夹角。诚然，

$$\begin{aligned}(\widehat{\mathbf{e}}_1'|\widehat{\mathbf{e}}_1)&=\left(\sum_{k=1}^{2}P_{k1}\widehat{\mathbf{e}}_k|\widehat{\mathbf{e}}_1\right)\\&=\sum_{k=1}^{2}P_{1k}\left(\widehat{\mathbf{e}}_k|\widehat{\mathbf{e}}_1\right)\\&=P_{1k}\delta_{k1}=P_{11}=\cos\phi\end{aligned}$$

由之前的结论，$\mathbf{Q}$ 在 B' 下的坐标可用某角 θ 表示，同时又是通过矩阵 P 从其在 B 下的坐标变换而来的，故有以下关系

$$\begin{aligned}(\mathbf{Q})_{B'}&=\begin{pmatrix}\cos\theta & \sin\theta\\ \sin\theta & -\cos\theta\end{pmatrix}\\&=P^{\intercal}\left(\mathbf{Q}\right)_B P\\ \Leftrightarrow\theta&=2\phi\end{aligned}$$

设 $\mathbf{a}=\alpha_1\widehat{\mathbf{e}}_1'+\alpha_2\widehat{\mathbf{e}}_2'$ 是二维欧几里得平面的任一平移向量。则 $\mathbf{Qa}$ 在 B' 下的坐标可由以下矩阵运算得到

$$(\mathbf{Q})_{B'}\left(\alpha_1,\alpha_2\right)^{\intercal}=\begin{pmatrix}\alpha_1\cos\theta+\alpha_2\sin\theta\\ -\alpha_2\cos\theta+\alpha_1\sin\theta\end{pmatrix}$$

记 δ 为向量 $\mathbf{a}$ 与 $\widehat{\mathbf{e}}_1'$ 的夹角，则上式说明 $\mathbf{Qa}$ 与 $\widehat{\mathbf{e}}_1'$ 的夹角是 $2\phi-\delta$。也就是说，在 2 维欧几里得平面上，任一向量被行列式为 -1 的正交算符 $\mathbf{Q}$ 作用后的结果，是原向量关于角度为 $\theta/2$ 的轴的镜象翻转。其中“角度”是指与所选定规范正交基的第 1 个单位基向量的夹角（图3.1）。θ 是 $\mathbf{Q}$ 在所选定的规范正交基下的坐标矩阵取前文所述的表达式时的角；它是 $\mathbf{Q}$ 本身的特性。也就是说，当 $\det\mathbf{Q}=-1$ 时，$\mathbf{Q}$ 在 2 维欧几里得空间中确定了一条镜象对称轴。特别地，当 $\theta=0$ 时，意味着我们选取的直角坐标系第一条轴恰好是 $\mathbf{Q}$ 确定的镜象对称轴。

3 维欧几里得平面上的几何意义

设 $\mathbf{Q}$ 是 3 维实内积空间 $\mathscr{V}$ 中的正交算符，由其行列式的基本性质 $\det\mathbf{Q}=\pm1$，可知其 3 个特征值 $\lambda_i=\pm1, i=1,2,3$。设 $\{\widehat{\mathbf{e}}_1,\widehat{\mathbf{e}}_2,\widehat{\mathbf{e}}_3\}$ 是 $\mathscr{V}$ 的一组规范正交基，且 $\widehat{\mathbf{e}}_1$ 是 $\mathbf{Q}$ 的一个特征

向量，即 $\mathbf{Q}\widehat{\mathbf{e}}_1 = \pm 1$。我们考察 $\mathbf{Q}$ 在这组规范正交基下的坐标 $Q_{ij} = (\mathbf{Q}\widehat{\mathbf{e}}_i|\widehat{\mathbf{e}}_j)$，有如下结论

$$\begin{aligned}(\mathbf{Q}\widehat{\mathbf{e}}_1|\widehat{\mathbf{e}}_1) &= \pm 1 \\ (\mathbf{Q}\widehat{\mathbf{e}}_1|\widehat{\mathbf{e}}_2) &= (\mathbf{Q}\widehat{\mathbf{e}}_1|\widehat{\mathbf{e}}_3) = 0, \\ (\mathbf{Q}\widehat{\mathbf{e}}_2|\widehat{\mathbf{e}}_1) = (\widehat{\mathbf{e}}_2|\mathbf{Q}^*\widehat{\mathbf{e}}_1) &= \pm(\widehat{\mathbf{e}}_2|\widehat{\mathbf{e}}_1) = 0,\end{aligned}$$

其中用到了事实 $\mathbf{Q}^*\widehat{\mathbf{e}}_1 = \pm\widehat{\mathbf{e}}_1$，因为

$$\widehat{\mathbf{e}}_1 = \mathbf{Q}^*\mathbf{Q}\widehat{\mathbf{e}}_1 = \mathbf{Q}^*(\pm\widehat{\mathbf{e}}_1) = \pm\mathbf{Q}^*\widehat{\mathbf{e}}_1$$

同理有

$$(\mathbf{Q}\widehat{\mathbf{e}}_3|\widehat{\mathbf{e}}_1) = 0$$

换言之，我们得到了

$$Q_{11} = \pm 1, \quad Q_{12} = Q_{13} = Q_{21} = Q_{31} = 0$$

另外注意到，$(\mathbf{Q}\widehat{\mathbf{e}}_i|\widehat{\mathbf{e}}_j)$ 同时也是向量 $\mathbf{Q}\widehat{\mathbf{c}}_i$ 的第 j 个坐标，故可知

$$\mathbf{Q}\widehat{\mathbf{e}}_2 = Q_{22}\widehat{\mathbf{e}}_2 + Q_{23}\widehat{\mathbf{e}}_3, \quad \mathbf{Q}\widehat{\mathbf{e}}_3 = Q_{32}\widehat{\mathbf{e}}_2 + Q_{33}\widehat{\mathbf{e}}_3$$

故由 $\|\mathbf{Q}\widehat{\mathbf{e}}_2\| = \|\widehat{\mathbf{e}}_2\| = \|\mathbf{Q}\widehat{\mathbf{e}}_3\| = \|\widehat{\mathbf{e}}_3\| = 1$ 有

$$Q_{22}^2 + Q_{23}^2 = Q_{32}^2 + Q_{33}^2 = 1$$

由 $(\widehat{\mathbf{e}}_2|\widehat{\mathbf{e}}_3) = 0$ 有

$$Q_{22}Q_{32} + Q_{23}Q_{33} = 0$$

由 $\det\mathbf{Q} = \pm 1$ 有

$$-Q_{23}Q_{32} + Q_{22}Q_{33} = \pm 1$$

类似 2 维的情况，上述结论使我们可以把 $\mathbf{Q}$ 在 $\mathscr{V}$ 的任何一个以 $\mathbf{Q}$ 的一个特征值为第一个单位基向量的有序基 $B = \{\widehat{\mathbf{e}}_1, \widehat{\mathbf{e}}_2, \widehat{\mathbf{e}}_3\}$ 下的坐标矩阵用某角 θ 表达成

$$(\mathbf{Q})_B = \begin{pmatrix} \pm 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & \mp\sin\theta & \pm\cos\theta \end{pmatrix}$$

其中“$\pm$”表示当 $\det\mathbf{Q} = 1$ 时取正号，当 $\det\mathbf{Q} = -1$ 时取负号。

以下我们考察，对任意向量 $\mathbf{a} \in \mathscr{V}$，$\mathbf{Qa}$ 相比于 $\mathbf{a}$ 在以 $\mathscr{V}$ 为平移空间的欧几里得空间中的效果。

当 $\det\mathbf{Q}=1$ 时，易验 $(\mathbf{Qa}|\widehat{\mathbf{e}}_1)=(\mathbf{a}|\widehat{\mathbf{e}}_1)$，也就是说，$\mathbf{Q}$ 的作用不改变向量与 $\widehat{\mathbf{e}}_1$ 的夹角；$\mathbf{Qa}$ 与 $\mathbf{a}$ 相比仅绕 $\widehat{\mathbf{e}}_1$ 旋转了一定的角度。具体地，设 $\mathbf{a}=\sum_{i=1}^{3}\alpha_i\widehat{\mathbf{e}}_i$，则

$$\mathbf{Qa}=\alpha_1\widehat{\mathbf{e}}_1+(\alpha_2\cos\theta+\alpha_3\sin\theta)\,\widehat{\mathbf{e}}_2+(-\alpha_2\sin\theta+\alpha_3\cos\theta)\,\widehat{\mathbf{e}}_3$$

比较正交算符在 2 维欧几里得平面的分析结果，上式的后两项表示向量 $\alpha_2\widehat{\mathbf{e}}_2+\alpha_3\widehat{\mathbf{e}}_3$ 在 $(\widehat{\mathbf{e}}_2,\widehat{\mathbf{e}}_3)$ 所确定的平面上绕原点旋转了角 θ。联系格拉姆–施密特正交化过程可知，$\alpha_2\widehat{\mathbf{e}}_2+\alpha_3\widehat{\mathbf{e}}_3$ 就是向量 $\mathbf{a}$ 在 $(\widehat{\mathbf{e}}_2,\widehat{\mathbf{e}}_3)$ 所确定的平面上的投影向量。因此，当 $\det\mathbf{Q}=1$ 时，$\mathbf{Qa}$ 在 3 维欧几里得空间确定了一个规范正交基 $\{\widehat{\mathbf{e}}_1,\widehat{\mathbf{e}}_2,\widehat{\mathbf{e}}_3\}$，其中 $\widehat{\mathbf{e}}_1$ 是一个旋转轴。对任意平移向量 $\mathbf{a}$，$\mathbf{Qa}$ 在 $\widehat{\mathbf{e}}_2,\widehat{\mathbf{e}}_3$ 所张的平面上的投影相比于 $\mathbf{a}$ 在这一平面上的投影绕旋转轴旋转角 θ（图3.2）。值得注意的是，$\{\widehat{\mathbf{e}}_1,\widehat{\mathbf{e}}_2,\widehat{\mathbf{e}}_3\}$ 是由 $\mathbf{Q}$ 的性质确定的。我们在讨论几何问题时未必恰好选择了这组特殊的规范正交基来建立直角坐标系。所以一般地，由 $\mathbf{Q}$ 所定义的旋转轴在预先选定的直角坐标系中是“斜放”的。特别地，当 $\theta=0$ 时，$\mathbf{Q}=\mathbf{I}$。

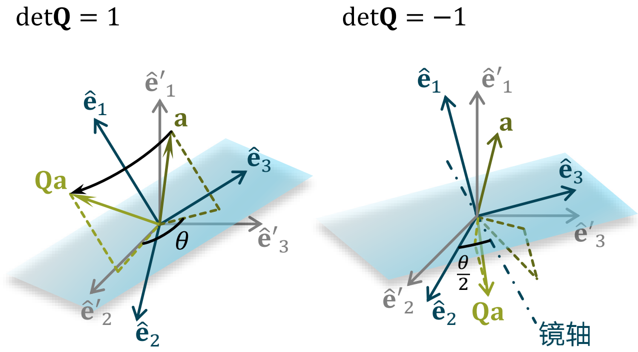


图 3.2: 正交算符在 3 维欧几里得空间中的几何意义。

当 $\det\mathbf{Q}=-1$ 时，由类似的讨论可得

$$\mathbf{Qa}=-\alpha_1\widehat{\mathbf{e}}_1+(\alpha_2\cos\theta+\alpha_3\sin\theta)\,\widehat{\mathbf{e}}_2+(-\alpha_3\cos\theta+\alpha_2\sin\theta)\,\widehat{\mathbf{e}}_3$$

首先，上式第一项说明，$\mathbf{Qa}$ 与 $\widehat{\mathbf{e}}_1$ 的夹角 ϕ' 跟 $\mathbf{a}$ 与 $\widehat{\mathbf{e}}_1$ 的夹角 ϕ 保持 $\phi'=\pi-\phi$ 的关系，也就是说 $\mathbf{Qa}$ 把 $\mathbf{a}$ 关于由 $(\widehat{\mathbf{e}}_2,\widehat{\mathbf{e}}_3)$ 确定的平面进行了镜象翻转。然后，比较上式后两项与 2 维内积空间上的正交算符的分析结果可知，它们是 $\mathbf{a}$ 在由 $(\widehat{\mathbf{e}}_2,\widehat{\mathbf{e}}_3)$ 所确定的平面上，以与 $\widehat{\mathbf{e}}_2$ 夹角为 $\theta/2$ 的直线为对称轴进行平面内的镜向翻转的结果（图3.2）。总之，当 $\det\mathbf{Q}=-1$ 时，$\mathbf{Q}$ 的几何效果就是上述两个操作的组合。$\mathbf{Q}$ 在 3 维欧几里得空间确定了一个镜面翻转的平面以及在该面内镜象翻转的一条对称轴。在已选定的直角坐标系中，这个平面和这条对称轴一般是“斜放”的。

3.3.2 对称算符与向量的拉伸

由定理2.32，复数域上的厄米算符特征值全为实数。也就是说，厄米算符的特征多项式全为实根。因此，定理2.32中关于复数域上厄米算符的性质，也适用于在实数域上的对称算符，包括：特征值全都是实数[3]§5.3 定理 3.4；对应于不同特征值的特征向量必正交[3]§5.3 定理 3.5；必可对角化[3]§5.3 定理 3.6；存在一组规范正交基是其特征向量。

一个对称算符的 3 个特征向量两两正交。设对称算符 $\mathbf{U}$ 的 3 个特征值为 $\lambda_1,\lambda_2,\lambda_3$，$\mathbf{U}$ 的 3 个特征向量为 $\widehat{\mathbf{c}}_1,\widehat{\mathbf{c}}_2,\widehat{\mathbf{c}}_3$，在规范正交基 $C=\{\widehat{\mathbf{c}}_i\}$ 下，$(\mathbf{U})_C$ 是一个对称矩阵

$$(\mathbf{U})_C=\begin{pmatrix}\lambda_1&0&0\\0&\lambda_2&0\\0&0&\lambda_3\end{pmatrix}$$

在实际问题中我们未必方便选择 $\mathbf{U}$ 的特征向量作为基。在其他规范正交基 $B=\{\widehat{\mathbf{e}}_i\}$ 下，$(\mathbf{U})_B$ 是一个对称矩阵。

给定任一向量 $\mathbf{a}=\sum_{i=1}^3\alpha_i\widehat{\mathbf{c}}_i$，则有 $\mathbf{U}\mathbf{a}=\sum_{i=1}^3\lambda_i\alpha_i\widehat{\mathbf{c}}_i$。也就是说，与 $\mathbf{a}$ 相比，$\mathbf{U}\mathbf{a}$ 是分别在 $\widehat{\mathbf{c}}_i$ 方向进行了比例为 λ_i 的拉伸或压缩，其中 $i=1,2,3$。因此，任一对称算符 $\mathbf{U}$ 均在 3 维空间中确定了 1 套（3 个）两两正交的拉伸方向，以及在相应方向的拉伸比。我们把对称算符 $\mathbf{U}$ 的单位特征向量 $\{\mathbf{c}_i\}$ 在 $\mathscr{E}$ 中所表示的方向称为该对称算符 $\mathbf{U}$ 的*主方向*（*principal direction*）。一组向量*在对称算符的作用下的几何效果是在主方向上的拉伸。具体地，在 $\mathbf{c}_i$ 方向的拉伸比就是 λ_i。在预先选定的直角坐标系中，对称算符的主方向可能是“斜放”的。

特别地，当 $\lambda_i=\pm1,i=1,2,3$ 时，对称算符不改变形状的尺寸，但仍可能使形状发生翻转；因为此时对称算符要么是恒等算符，要么是行列式为 -1 的正交算符。

3.3.3 斜称算符与向量的“叉乘”

我们先比定理2.32更详细地考察斜称算符的性质。设 $\mathscr{W}$ 是数域 $\mathbb{F}$ 上的 n 维内积空间，$\mathbf{W}$ 是 $\mathscr{W}$ 上的一个斜称算符，$\{\lambda_i\}$ 是 $\mathbf{W}$ 的特征值。由定理2.32，$\{\lambda_i\}$ 是纯虚数或零。在 $\mathbb{R}^n$ 上，当 n 是奇数时，

$$\begin{aligned}&\det\mathbf{W}^\intercal=\det\mathbf{W}=\det(-\mathbf{W})=(-1)^n\det\mathbf{W}=-\det\mathbf{W}\\ \Leftrightarrow&\det\mathbf{W}=\prod_{i=1}^n\lambda_i=0\end{aligned}$$

故当 n 为奇数时，斜称算符必有一特征值为零。而且，上列结果也说明，当 n 为奇数时，斜称算符不可逆（不满秩）。

*在 $\mathscr{E}$ 中，任何几何形状都是点的集合，而点又与位置向量一一对应，故任何几何形状都是一个向量组。

现考虑 3 维的情况，改设 $\mathbf{W}$ 是 $\mathscr{V}$ 上的一个斜称算符，则对任意 $\mathbf{a} \in \mathscr{V}$，

$$(\mathbf{W}\mathbf{a}|\mathbf{a}) = -(\mathbf{a}|\mathbf{a}) = (\mathbf{a}|\mathbf{W}\mathbf{a}) = -(\mathbf{a}|\mathbf{a}) \Leftrightarrow (\mathbf{W}\mathbf{a}|\mathbf{a}) = 0$$

故被 $\mathbf{W}$ 作用过的向量，都被投影到了与原向量垂直的平面上。至于投影了之后，有没有伸缩或旋转，要看 $\mathbf{W}$ 的具体取值。这十分类似于以往所学过的一个向量被另一个向量“叉乘”的效果*。事实上，我们可以从斜称算符定义“叉乘”†。我们可通过如下定理联系二者：

定理 3.1. 设 $\mathscr{V}$ 是实数域 $\mathbb{R}$ 上的 3 维内积空间，$\mathbf{W}$ 是 $\mathscr{V}$ 上的一个斜称算符，$\hat{\mathbf{e}}_1$ 是 $\mathbf{W}$ 的关于特征值 $\lambda = 0$ 的一个特征单位向量。$\{\hat{\mathbf{e}}_i\}$ 是由 $\hat{\mathbf{e}}_1$ 生成的规范正交基。对任意 $\mathbf{a} \in \mathscr{V}$，可定义“叉乘”运算

$$\mathbf{W}\mathbf{a} = \omega\hat{\mathbf{e}}_1 \times \mathbf{a}$$

其中 $\omega = (\mathbf{W}\hat{\mathbf{e}}_2|\hat{\mathbf{e}}_3)$。

证明. 若记向量 $\mathbf{W}\hat{\mathbf{e}}_2 = \sum_{i=1}^{3}\beta_i\hat{\mathbf{e}}_i$，即 $\beta_i = (\mathbf{W}\hat{\mathbf{e}}_2|\hat{\mathbf{e}}_i)$，　$i = 1, 2, 3$，则有

$$\begin{aligned}
\beta_1 &= (\mathbf{W}\hat{\mathbf{e}}_2|\hat{\mathbf{e}}_1) = -(\hat{\mathbf{e}}_2|\mathbf{W}\hat{\mathbf{e}}_1) = (\hat{\mathbf{e}}_2|\mathbf{0}) = 0\\
\beta_2 &= (\mathbf{W}\hat{\mathbf{e}}_2|\hat{\mathbf{e}}_2) = -(\mathbf{e}_2|\mathbf{W}\hat{\mathbf{e}}_2) \Leftrightarrow \beta_2 = -\beta_2 \Leftrightarrow \beta_2 = 0\\
\beta_3 &= (\mathbf{W}\hat{\mathbf{e}}_2|\hat{\mathbf{e}}_3) = \omega
\end{aligned}$$

因此向量 $\mathbf{W}\hat{\mathbf{e}}_2 = \omega\hat{\mathbf{e}}_3$。类似的方法可得到 $\mathbf{W}\hat{\mathbf{e}}_3 = -\omega\hat{\mathbf{e}}_2$。故对任意 $\mathbf{a} \in \mathscr{V}$，

$$\mathbf{W}\mathbf{a} = \alpha_1\mathbf{W}\hat{\mathbf{e}}_1 + \alpha_2\mathbf{W}\hat{\mathbf{e}}_2 + \alpha\mathbf{W}\hat{\mathbf{e}}_3 = \alpha_2\omega\hat{\mathbf{e}}_3 - \alpha_3\omega\hat{\mathbf{e}}_2$$

这恰为 $\omega\hat{\mathbf{e}}_1 \times \mathbf{a}$ 的结果 □

要使用这个定理，就需要先已知 $\mathbf{W}$，并找到 $\mathbf{W}$ 在关于其零特征值的一个单位特征向量。如果我们先给定两个向量的“叉乘” $\mathbf{a} \times \mathbf{b}$，那么什么样的斜称算符 $\mathbf{W}$ 满足 $\mathbf{W}\mathbf{b} = \mathbf{a} \times \mathbf{b}$ 呢？答案作为3.1的推论如下。

推论 3.1.1. 设 $\{\hat{\mathbf{e}}_i\}$ 是 3 维欧几里得空间 $\mathscr{E}$ 的基本坐标系下已给定的一组有序规范正交基。给定向量 $\mathbf{a} \in \mathscr{V}, \mathbf{a} = \sum_{i=1}^{3}\alpha_i\hat{\mathbf{e}}_i$，则坐标矩阵为

$$(\mathbf{W}) = \begin{pmatrix} 0 & \alpha_3 & \alpha_2 \\ -\alpha_3 & 0 & \alpha_1 \\ -\alpha_2 & -\alpha_1 & 0 \end{pmatrix}$$

*3 维实内积空间上的“叉乘”在以前的线性代数课本中又称作“向量的外积”[3]§3.2。请复习其计算方法。

†在本讲义中，将保持使用“叉乘”这种不正式的表述。因为正式起来，3 维实内积空间上的“叉乘”是抽象代数中的不同东西在 3 维实内积空间上的巧合。在连续介质力学的数学语言中出现的“叉乘”，有时真的就是本节所述的几何意义，有时则是外代数/外积/楔积（如向量场的旋度）。本讲义暂时不介绍基于外代数和微分型知识，故“叉乘”运算就都由本节引入了。

的斜称算符 $\mathbf{W}$ 满足 $\mathbf{W}\mathbf{b} = \mathbf{a} \times \mathbf{b}, \forall \mathbf{b} \in \mathscr{V}$。

该推论可利用坐标变换公式证明，可留作练习。另外，“叉乘”的性质 $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$ 也可通过在给定规范正交基下的坐标矩阵运算得到验证（仅限 3 维）。

一个斜称算符 $\mathbf{W}$ 所对应的叉乘运算的第一个向量 $\mathbf{a}$ 是由算符 $\mathbf{W}$ 本身确定的，称为 $\mathbf{W}$ 的轴向量（*axial vetor*）。“叉乘”得到的向量，在坐标变换中的有特殊的性质。对任意 $\mathbf{b} \in \mathscr{V}$，其在基本坐标系下的坐标是 $(\beta_1, \beta_2, \beta_3)$。当我们反转各坐标轴的方向，即在 $\{-\widehat{\mathbf{e}}_i\}$ 下，$\mathbf{b}$ 的坐标将变号，变成 $(-\beta_1, -\beta_2, -\beta_3)$，但是易验向量 $\mathbf{a} \times \mathbf{b}$ 在坐标轴反转前后坐标不变号（请读者用坐标变换公式验证这两个结论）。同样的性质也导致“混合积” $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$ 这一“标量”在坐标轴反转前后变号（而“真正的”标量的值不依赖坐标变换）。这是由于向量 $\mathbf{a} \times \mathbf{b}$ 不是一个任意给定的孤立的向量；它是向量 $\mathbf{b}$ 在以 $\mathbf{a}$ 为轴向量的斜称算符 $\mathbf{W}$ 的操作下的结果。由“叉乘”所得到的向量，其看上去特殊的坐标变换性质，实际上是这个投影操作带来的。有的资料称这种“叉乘得到的向量”为赝向量（*pseudo-vector*）（但这不是该概念的正式定义）。

3.4 等距变换的表示定理

在一个欧几里得空间中，任一针对点空间的点的等距变换结果都是相对某另一点的平移再叠加一个正交变换。这个定理也是后面介绍物理定律的标架变换不变性时的理论基础。我们把它正式表述如下：

定理 3.2 (等距变换的表示定理). 设 $\mathscr{E}$ 是一个欧几里得空间，$\mathscr{V}$ 是其平移空间，选定任一点 $X_0 \in \mathscr{E}$，则 $\mathscr{E}$ 上的任一等距变换 $i : \mathscr{E} \to \mathscr{E}, i \in \mathscr{V}$ 都可表示为

$$i(X) = i(X_0) + \mathbf{Q}_i(X - X_0)$$

其中 $\mathbf{Q}_i$ 是一个正交算符，由 i 唯一确定。

证明. 见附录。 □

由定理3.2，给定欧几里得空间上的任一等距变换 i，则仅需知道 i 关于某一参考点 X_0 的像是哪个点，以及由 i 确定的某特征正交算符 $\mathbf{Q}_i$ 的取值，就可以知道 i 对任意一点 $X \in \mathscr{E}$ 的像。享有表示定理的等距变换 i 未必需要是 $\mathscr{V}$ 中的向量，但由于任意两点的“差”唯一对应一个 $\mathscr{V}$ 中的平移向量，故 $i(X) - i(X_0)$ 和 $X - X_0$ 分别表示由点 $i(X_0)$ 到点 $i(X)$ 和点 X_0 到点 X 的平移向量。由等距变换的表示定理，这两个平移向量只相差由正交算符 $\mathbf{Q}_i$ 规定的几何变换，具体可见 §3.3。

例 3.4. 考虑欧几里得空间 $(\mathscr{E}, d)$ 上的以下等距变换，其中 $\mathbf{Q}$ 是一个正交算符，X_0, C 是 $\mathscr{E}$ 中固定的点：

$$
\begin{aligned}
i_1(X) &= X + (C - X_0) \\
i_2(X) &= X_0 + \mathbf{Q}(X - X_0) \\
i_3(X) &= X + \mathbf{Q}^{-1}(C - X_0)
\end{aligned}
$$

i_1 把任一点向固定的方向平移固定距离（$i_1(X) = X + \mathbf{u}, \mathbf{u} \equiv C - X_0$）。

$i_1 \circ i_2 = i_2 \circ i_3$（自行验证作为练习。）

当 $\mathbf{Q} = \mathbf{I}$ 时，i_2 是恒等映射。当 $\mathbf{Q} \neq \mathbf{I}$ 时，由正交算符性质 $\det\mathbf{Q} = \pm 1$。当 $\det\mathbf{Q} = 1$ 时，i_2 是一种旋转操作；当 $\det\mathbf{Q} = -1$ 时，由 $\mathbf{Q} = (-\mathbf{I})(-\mathbf{Q})$ 和 $\det(-\mathbf{Q}) = 1$ 可知 i_2 是先进行了一个旋转（$-\mathbf{Q}$）再进行了反转（$-\mathbf{I}$）的操作。

由该例最后的结论可知，$\det\mathbf{Q}$ 为 $+1$ 和 -1 的两种情况，只差一个翻转 $-\mathbf{I}$ 变换。设 $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3\}$ 是 3 维欧几里得空间的平移空间 $\mathscr{V}$ 的一组规范正交基，则它们经过翻转变换后形成的另一组基 $\{-\hat{\mathbf{e}}_1, -\hat{\mathbf{e}}_2, -\hat{\mathbf{e}}_3\}$，在画法惯例上只存在“左手规则”与“右手规则”的差别（图3.3）。用形象的语言说，就是我们的物理世界是存在“镜子外”和“镜子内”两套的，而且它们的运动学规律只相差一个翻转变换。在本讲义内，我们只关心“镜子外的物理世界”的规律，并规定用“右手规则”来建立坐标系，对应于 $\det\mathbf{Q} = 1$ 的情况。

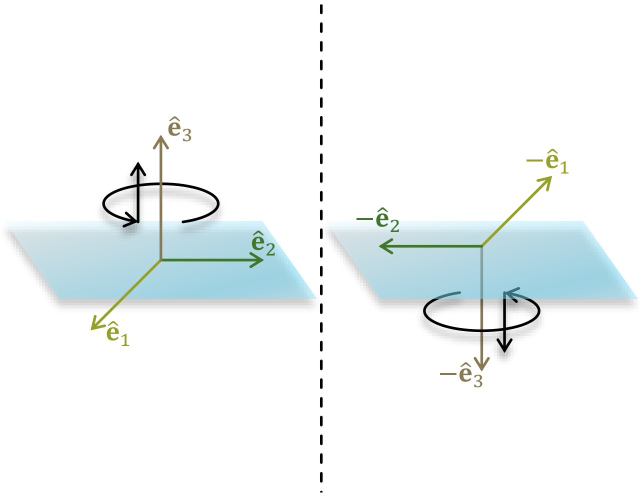


图 3.3: 翻转变换在纸面画法上相当于左、右手规则之差别。

第四章　向量函数的微积分

4.1　向量函数及其可视化

在线性代数部分的介绍中，我们知道实数域 $\mathbb{R}$ 上的 n 维向量空间的向量在给定基下总是唯一对应于 n 维实坐标空间 $\mathbb{R}^n$ 中的一组有序实数 n 元组。因此，如不另作说明，向量函数都直接讨论 $\mathbb{R}^n$ 上的向量。

定义 4.1 (向量函数). 从 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的映射 $\mathbf{f}: \mathbb{R}^n \supset D \to \mathbb{R}^m$ 是一个向量函数（*vector function*）。其变量是一个 n 维向量空间的向量 $\mathbf{r} \in \mathbb{R}^n$，$\mathbf{r} = (r_1, \cdots, r_n)^\intercal$；其像（函数值）$\mathbf{f}(\mathbf{r})$ 是一个 m 维向量空间的向量 $\mathbf{f}(\mathbf{r}) \in \mathbb{R}^m$，$\mathbf{f}(\mathbf{r}) = (f_1(\mathbf{r}), \cdots, f_m(\mathbf{r}))^\intercal$。$f_j : \mathbb{R}_n \to \mathbb{R}, j = 1, \cdots, m$ 称为 $\mathbf{f}$ 的坐标函数（*coordinate function*）。

例 4.1. 函数 $\mathbf{f}: \mathbb{R}^3 \to \mathbb{R}^2$，

$$\mathbf{f}(\mathbf{r}) = \begin{pmatrix} r_1^2 + r_2^2 + r_3^2 \\ r_1 + r_2 + r_3 \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix},$$

$$f_1 \le 0, f_2 \in \mathbb{R}$$

其中 $\mathbf{r} = (r_1, r_2, r_3)^\intercal \in \mathbb{R}^3$，$\mathbf{f} = (f_1, f_2)^\intercal \in \mathbb{R}^2$。以往我们更习惯把上述函数写成：

$$\begin{cases} P(x, y, z) = x^2 + y^2 + z^2 \\ Q(x, y, z) = x + y + z \end{cases}$$

例 4.2. 函数 $\mathbf{A}: \mathbb{R}^3 \to \mathbb{R}^2$，

$$\mathbf{A}(\mathbf{r}) = \begin{pmatrix} 3r_1 + 4r_2 \\ 3r_2 + 5r_3 \end{pmatrix} = \begin{pmatrix} 3 & 4 & 0 \\ 0 & 3 & 5 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix}$$

是一个线性变换，又称线性函数。线性函数也可以按线性代数的惯例记为：$\mathbf{Ar}$。

定义 4.2 (隐函数). 考虑函数 $\mathbf{F}:\mathbb{R}^{n+m}\to\mathbb{R}^m$，若将 $\mathbb{R}^{n+m}$ 中的元素 $(x_1,\cdots,x_n,y_1,\cdots,y_m)^\intercal$ 写成 $(\mathbf{x},\mathbf{y})^\intercal$，其中 $\mathbf{x}=(x_1,\cdots,x_n)^\intercal$，$\mathbf{y}=(y_1,\cdots,y_m)^\intercal$，则 $\mathbf{F}(\mathbf{r}),\mathbf{r}\in\mathbb{R}^{n+m}$ 也可写成 $\mathbf{F}(\mathbf{x},\mathbf{y})$。若函数 $\mathbf{f}:\mathbb{R}^n\supset D\to\mathbb{R}^m$ 满足方程 $\mathbf{F}(\mathbf{x},\mathbf{f}(\mathbf{x}))=\mathbf{0},\forall\mathbf{x}\in D$ 则称这一方程隐含地定义了（*implicitly defined*）函数 $\mathbf{f}$。用这种方式表示的函数叫隐函数（*implicit function*）。

例 4.3. 设函数 $F(x,y)=x^2+y^2-1=0$，则"$F(x,f(x))=x^2+(f(x))^2-1=0,\forall x\in\mathrm{dom}f$"隐含地定义了以下任一函数：

$$
\begin{aligned}
f_1(x)&=\sqrt{1-x^2},\quad -1\le x\le 1\\
f_2(x)&=-\sqrt{1-x^2},\quad -1\le x\le 1\\
f_3(x)&=\begin{cases}\sqrt{1-x^2}, & -\frac{1}{2}\le x\le 0\\ -\sqrt{1-x^2}, & 0\le x\le 1\end{cases}
\end{aligned}
$$

由例4.3可见，不同定义域的函数 $\mathbf{f}$ 都可由同一函数 $\mathbf{F}$ 隐含地定义。函数的隐含定义并不完全确定该函数。

定义 4.3 (函数的图像). 函数 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$ 的图像（*graph*）是指所有有序对 $(\mathbf{r},\mathbf{f}(\mathbf{r}))$ 的集合[8]§6.4~6.7。

图像的数学概念是一个集合，我们把图像画在纸上的方式只是一种惯例。具体地，函数 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$ 的图像是由所有满足

$$
\begin{cases}
y_1=f_1(x_1,\cdots,x_n)\\
\qquad\vdots\\
y_m=f_m(x_1,\cdots,x_n)
\end{cases}
$$

的点 $(x_1,\cdots,x_n,y_1,\cdots,y_m)$ 的集合（是 $\mathbb{R}^{n+m}$ 的子集）。其中 $f_1,\cdots,f_m$ 是 $\mathbf{f}$ 的坐标函数。

我们只懂在纸上画出维数小于等于 3 的图形，即图上的任一点的坐标最多为 3 个实数。因此我们懂得在纸上画出的函数图像仅限 $n+m\le 3$ 的情况。

例 4.4. 函数 $y=x^2-2$ 的图象是所有有序对 (x,y)。同时，由于 $(x,y)\in\mathbb{R}^2$，故这是平面上的图形。具体地，它是如图*4.1*所示的一条二次曲线。

例 4.5. 函数 $f:\mathbb{R}^2\to\mathbb{R},f(\mathbf{r})=\|\mathbf{r}\|^2$ 的图像是有序 *3* 元组 $\left(r_1,r_2,\|\mathbf{r}\|^2\right)$ 的集合，是 *3* 维空间的一个如图*4.2*所示的曲面。

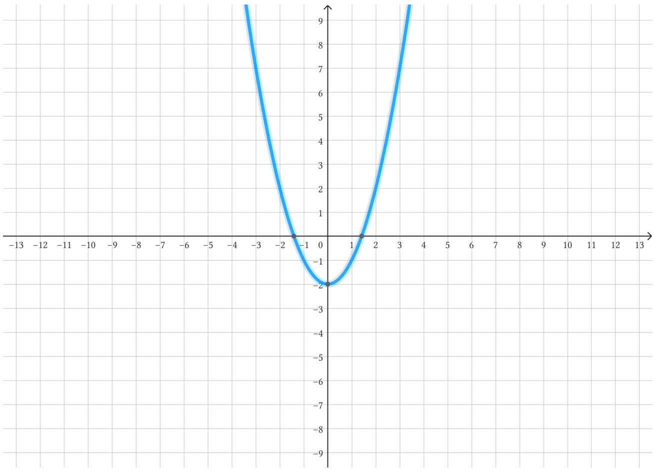


图 4.1: 函数 $y = x^2 - 2$ 的图像。

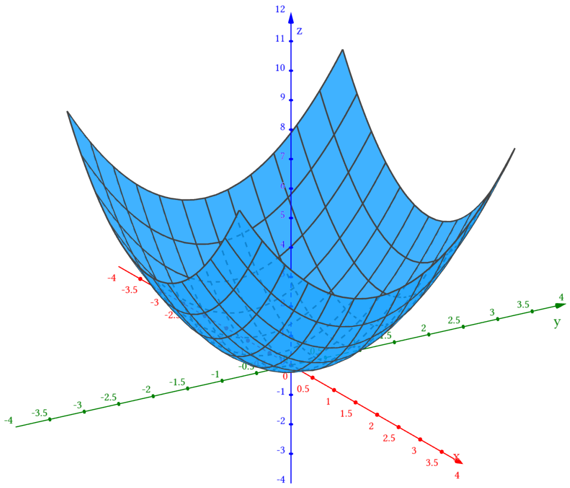


图 4.2: 函数 $f : \mathbb{R}^2 \to \mathbb{R}, f(\mathbf{r}) = \|\mathbf{r}\|^2$ 的图像在 $-2 < r_1 < 2, -2 < r_2 < 2$ 范围内的部分。

例 4.6. 函数 $\mathbf{f} : \mathbb{R} \to \mathbb{R}^2$，

$$\mathbf{f}(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}, \quad \forall t \in \mathbb{R}$$

的图像是所有有序对 $(t, (\cos t, \sin t))$。我们把 *2* 维向量 $(\cos t, \sin t)$ 作在实数轴上 t 对应的位置，得到图*4.3*所示的螺线。

例 4.7. 函数 $f : \mathbb{R}^2 \supset D \to \mathbb{R}^3$，

$$\mathbf{f}(\mathbf{r}) = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} = \begin{pmatrix} r_1 \cos r_2 \\ r_1 \sin r_2 \\ r_2 \end{pmatrix}, \quad D : 0 \le r_1 \le 4, \le r_2 \le 2\pi$$

该函数的图像无法作成 *3* 维欧几里得空间中的点集，但我们仍可在 *3* 维欧几里得空间中对该

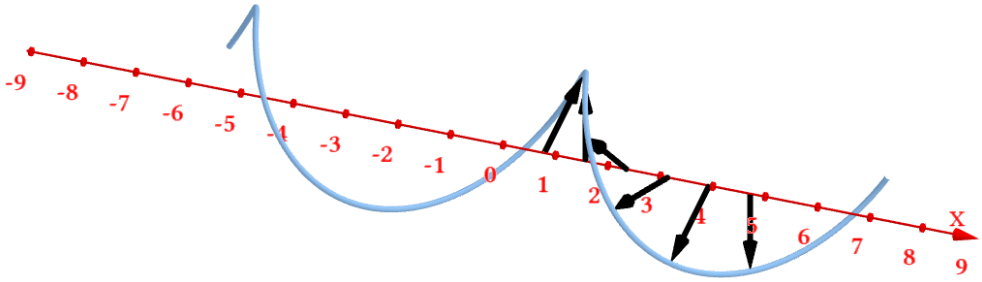


图 4.3: 例4.6函数的图象。

函数进行可视化。我们看到，$\mathbf{f}$ 的定义域是 2 维平面上的一个矩形区域 D。令 $r_1 = a$ 有

$$\mathbf{f} = \begin{pmatrix} a\cos r_2 \\ a\sin r_2 \\ r_2 \end{pmatrix}, \quad 0 \le r_2 \le 2\pi$$

且 $f_1^2 + f_2^2 = a^2$。视 a 为常数，则有序 3 元组 $(a\cos r_2, a\sin r_2, r_2)$ 是一个螺线，其在 $f_1 - f_2$ 面上的投影是以原点为圆心、半径为 a 的圆。若令 $r_2 = \theta$，且 θ 是常数，则有

$$\mathbf{f} = \begin{pmatrix} r_1\cos\theta \\ r_1\sin\theta \\ \theta \end{pmatrix}, \quad 0 \le r_2 \le 2\pi$$

且 $f_1^2 + f_2^2 = r_1^2$。在同样的空间坐标上，这是螺线上某点到 z 轴的线段。随着 a 在 $[0,4]$ 上变化、θ 在 $[0,2\pi]$ 上变化，以原函数 $\mathbf{f}(\mathbf{r})$ 的坐标函数为坐标的点集是如图4.4所示的螺带曲面。但按定义4.3，这个曲面不是函数 $\mathbf{f}$ 的图像。

例4.7中的函数按照图4.4可视化时，我们称这个三维空间中的螺带曲面是由一个参数方程（*parametric equation*）（即 $\mathbf{f} = \mathbf{f}(\mathbf{r})$）所定义的曲面，它的参数变量（*parametric variable*）$\mathbf{r} \in \mathbb{R}^2$ 取值的区域就是如图4.4所示的矩形区域，称为参数域（*domain of parameter variable*）。我们所画出来的 3 维曲面只是函数 $\mathbf{f}$ 的值域 $\mathrm{ran}\mathbf{f}$ 所形成的点集。一般地，由参数方程 $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ 定义的几何形状，当 $n = 1$ 时，是 m 维空间中的曲线，$m = 2, 3$；当 $n = 2$ 时是 $m = 3$ 维空间中的曲面。

定义 4.4 (函数的水平集). 函数 $\mathbf{f}\colon \mathbb{R}^n \supset D \to \mathbb{R}^m$ 的水平集（*level set*）是其定义域 D 的子集 $S = \{\mathbf{r} | \mathbf{f}(\mathbf{r}) = \mathbf{a}\}$，其中 $\mathbf{a} \in \mathrm{ran}\mathbf{f}$ 是某常向量[9]§7.1。

例 4.8. 设函数 $f\colon \mathbb{R}^3 \to \mathbb{R}, f(\mathbf{r}) = r_1r_2 + r_2r_3 + r_3r_1$ 及其一个水平集 $S = \{\mathbf{r} | f(\mathbf{r}) = 1\}$。以 S 的元素 (r_1, r_2, r_3) 为坐标的点集构成的三维图形是怎样的？设 $r_3 = 0$ 得到方程 $r_1r_2 = 1$，这定义了 $\hat{\mathbf{e}}_1$-$\hat{\mathbf{e}}_2$ 平面上的一条双曲线，这条双曲线是 S 的图像与 $\hat{\mathbf{e}}_1 - \hat{\mathbf{e}}_2$ 平面的截线。类似地，S

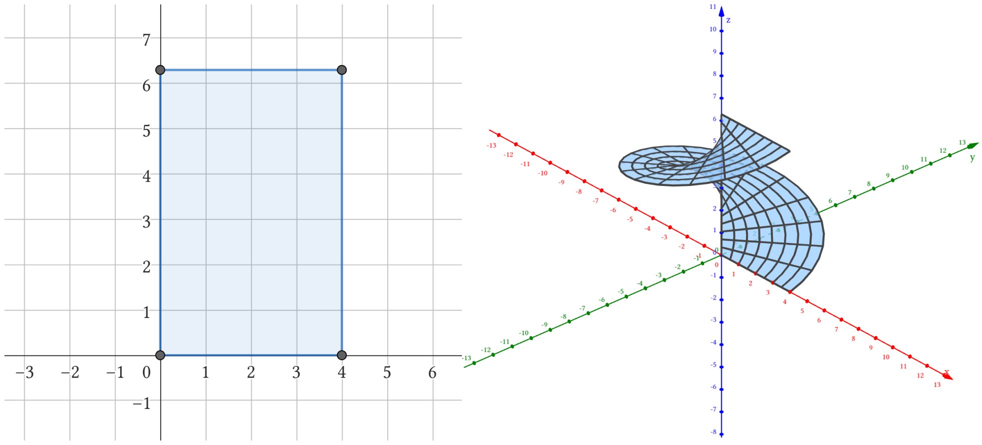


图 4.4: 由例4.7的参数方程定义的曲面。

与 $\widehat{\mathbf{e}}_2-\widehat{\mathbf{e}}_3$、$\widehat{\mathbf{e}}_1-\widehat{\mathbf{e}}_3$ 平面的截线也是类似的双曲线。更一般地，S 是一个由一系列双曲线构成的曲面（如图*4.5*所示）。但是，按定义*4.3*，这不是函数 $\mathbf{f}$ 的图像。

总结以上例子，我们给一个函数画出来的图有以下三种情况：

1. 如果 $\mathbb{R}^{n+m}$ 的子集 S 是函数 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$ 的图像，则称 S 是由显函数定义的图像（例4.4,4.5,4.6）。
2. 如果 $\mathbb{R}^m$ 的子集 S 是函数 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$ 的值域，则 S 是由参数方程定义的图像（例4.7）。
3. 如果 S 是函数 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$ 的一个水平集，则 S 是由隐函数定义的图像（例4.8）。

注意：只有第 1 种情况中 S 才是函数 $\mathbf{f}$ 的图像，但我们经常通过第 2、3 种情况中的 S 来认识函数 $\mathbf{f}$。

4.2　向量函数的极限与连续性

我们用“$\varepsilon-\delta$ 语言”来定义函数极限。

定义 4.5 (向量函数的极限). 给定函数 $\mathbf{f}:\mathbb{R}^n\supset D\to\mathbb{R}^m$，如果对任意正实数 ε 总存在正实数 δ 使得只要 $0<\|\mathbf{r}-\mathbf{r}_0\|<\delta,\mathbf{r},\mathbf{r}_0\in D$，就总有 $\|\mathbf{f}(\mathbf{r})-\mathbf{y}_0\|<\varepsilon,\mathbf{y}_0\in\mathbb{R}^m$，就称 $\mathbf{y}_0$ 是函数 $\mathbf{f}(\mathbf{r})$ 在 $\mathbf{r}_0$ 处的极限（*limit*），记为

$$\lim_{\mathbf{r}\to\mathbf{r}_0}\mathbf{f}(\mathbf{r})=\mathbf{y}_0$$

定理 4.1. 函数 $\mathbf{f}:\mathbb{R}^n\supset D\to\mathbb{R}^m$ 在 $\mathbf{r}_0\in D$ 处存在极限 $\lim_{\mathbf{r}\to\mathbf{r}_0}\mathbf{f}(\mathbf{r})=\mathbf{a},\mathbf{a}=(a_i)\in\mathbb{R}^m$ 的充要条件是其坐标函数的极限 $\lim_{\mathbf{r}\to\mathbf{r}_0}f_i(\mathbf{r})=a_i,\forall i=1,\cdots,m$ 都存在。

证明. 由向量函数定义，$\mathbf{f}$ 与 $f_i,i=1,\cdots,m$ 的定义域都相同。

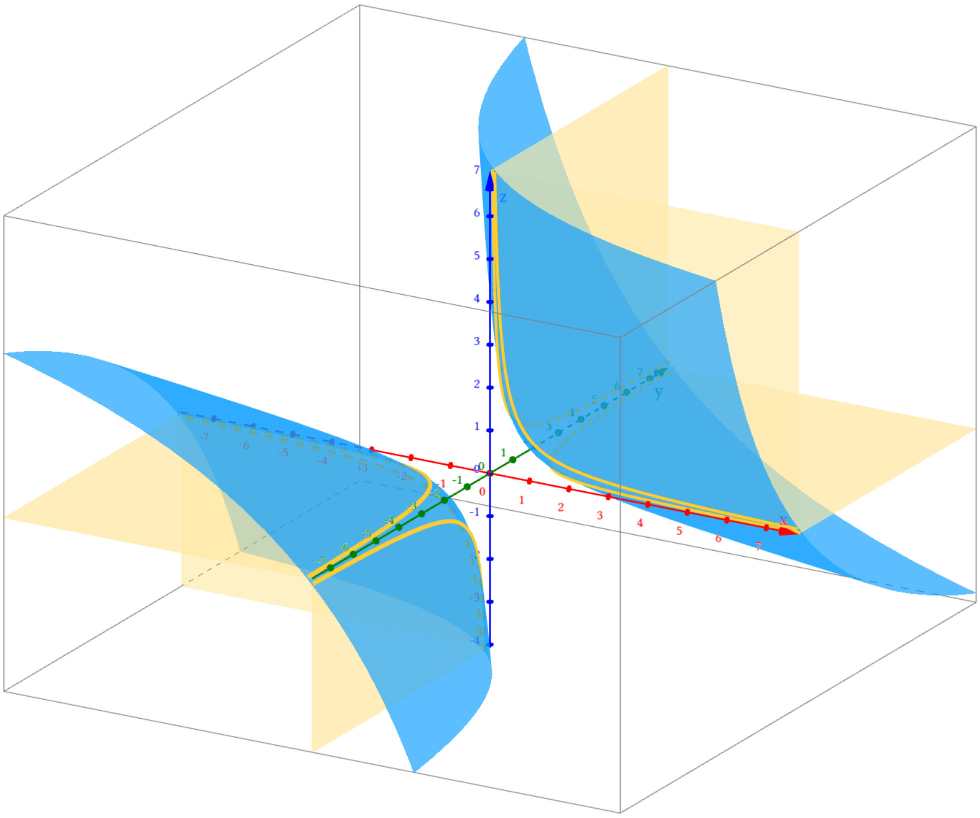


图 4.5: 函数 $f(\mathbf{r}) = r_1r_2 + r_2r_3 + r_3r_1$ 的水平集 $S = \{(x, y, z) \,|xy + yz + xz = 1\}$。

若已知 $\lim_{\mathbf{r}\to\mathbf{r}_0} \mathbf{f}(\mathbf{r}) = \mathbf{a}$，即对任一实数 $\varepsilon > 0$ 都存在实数 $\delta > 0$ 使得只要 $0 < \|\mathbf{r} - \mathbf{r}_0\| < \delta$ 就有 $\|\mathbf{f}(\mathbf{r}) - \mathbf{a}\| < \varepsilon$。那么，取同一对正实数 ε 和 δ，对任一 $i \in \{1, \cdots, m\}$ 都有

$$
\begin{aligned}
|f_i(\mathbf{r}) - \alpha_i| &\leq \sum_{i=1}^{m} |f_i(\mathbf{r}) - a_i| \quad (\text{当}m = 1\text{或}f_j = a_j \forall j \neq i\text{时取等号。}) \\
&\leq \left|\sum_{i=1}^{m} (f_i(\mathbf{r}) - \alpha_i)\right| \quad (\text{三角不等式。}) \\
&= \|\mathbf{f}(\mathbf{r}) - \mathbf{a}\| < \varepsilon
\end{aligned}
$$

故按照极限的定义，$\lim_{\mathbf{r}\to\mathbf{r}_0} f_i(\mathbf{r}) = \alpha_i, i = 1, \cdots, m$，其中 $a_i, i = 1, \cdots, n$ 是 $\mathbf{a}$ 的坐标。

反之，若极限 $\lim_{\mathbf{r}\to\mathbf{r}_0} f_i(\mathbf{r}) = a_i, i = 1, \cdots, m$ 都存在，即对任一实数 $\varepsilon > 0$ 都能找到实数 $\delta_i > 0, i = 1, \cdots, m$，使得只要 f_i 的定义域 D 内一向量 $\mathbf{r}$ 到 $\mathbf{r}_0$ 的距离满足 $0 < \|\mathbf{r} - \mathbf{r}_0\| < \delta_i$，就有 $|f_i(\mathbf{r}) - a_i| < \varepsilon/\sqrt{m}$，那么令 $\delta = \min\{\delta_1, \cdots, \delta_m\}$，则当 $0 < \|\mathbf{r} - \mathbf{r}_0\| < \delta$，就有 $\max\{|f_i(\mathbf{r}) - a_i|\} < \varepsilon/\sqrt{m}$。

利用事实*

$$\|\mathbf{x}\| \leq \sqrt{n}\max\{|x_1|, \cdots, |x_n|\}, \forall \mathbf{x} \in \mathbb{R}^n$$

*由 $\|\mathbf{x}\|^2 = \sum_{i=1}^{n} x_i^2 \leq n\max\{x_1^2, \cdots, x_n^2\}$ 得到。

可得

$$\|\mathbf{f}(\mathbf{r})-\mathbf{a}\| \le \sqrt{m}\max\{|f_i(\mathbf{r})-a_i|\}<\varepsilon$$

这样的 δ 是总能取到的，故由函数极限的定义，$\lim_{\mathbf{r}\to\mathbf{r}_0}\mathbf{f}(\mathbf{r})=\mathbf{a}$。 □

定义4.5表示，若 $\lim_{\mathbf{x}\to\mathbf{x}_0}\mathbf{f}(\mathbf{x})=\mathbf{y}_0$，则只要点 $\mathbf{x}\neq\mathbf{x}_0$ 足够靠近 $\mathbf{x}_0$，则 $\mathbf{f}(\mathbf{x})$ 总可以任意近地靠近点 $\mathbf{y}_0$，而且 $\mathbf{f}(\mathbf{x}_0)$ 未必等于 $\mathbf{y}_0$。结合 §B.1定义的概念，点 $\mathbf{r}_0$ 是 $\mathrm{dom}\mathbf{f}$ 中的一个极限点，$0<\|\mathbf{r}-\mathbf{r}_0\|<\delta$ 是点 $\mathbf{r}_0$ 的一个去心邻域。若 $\lim_{\mathbf{x}\to\mathbf{x}_0}\mathbf{f}(\mathbf{x})=\mathbf{y}_0$，则对 $\mathbf{y}_0$ 的任一邻域 $B_\varepsilon(\mathbf{y}_0)\subset\mathbb{R}^m$，总存在点 $\mathbf{x}_0$ 的邻域 $B_\delta(\mathbf{x}_0)\subset\mathbf{R}^n$，使得集合 $B_\delta(\mathbf{x}_0)\cup D$ 在 $\mathbf{f}$ 下的像集是 $B_\varepsilon(\mathbf{y}_0)$ 的子集（如图4.6）。

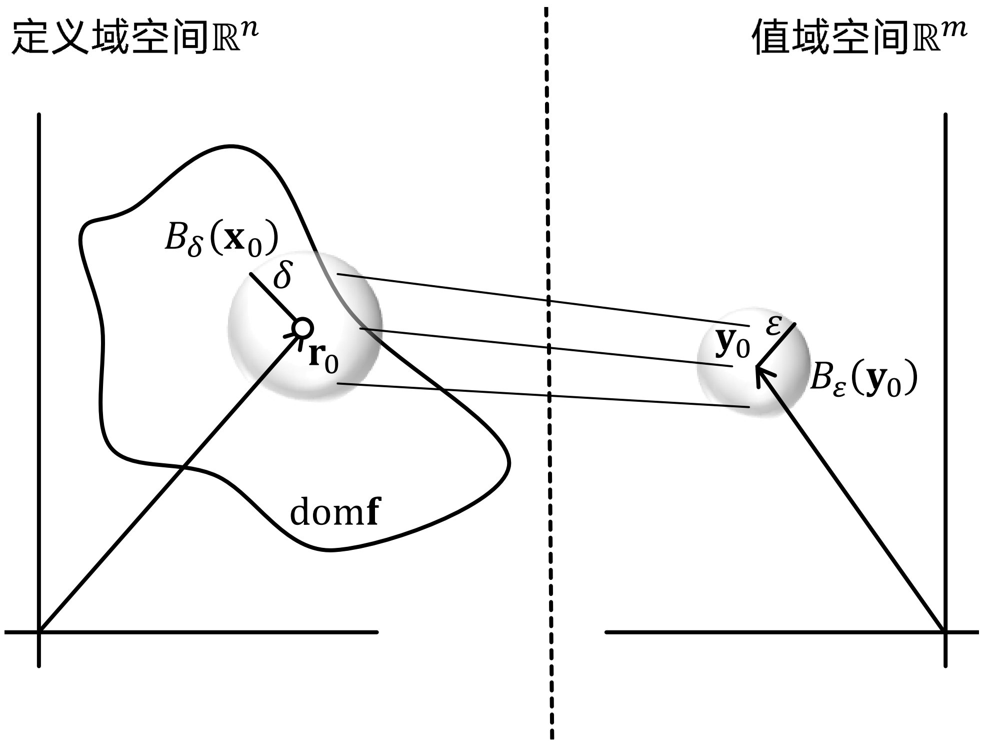


图 4.6: 向量函数的极限（定义4.5）的图像化理解。

定义 4.6 (函数的连续性). 给定函数 $\mathbf{f}:\mathbb{R}^n\supset D\to\mathbb{R}^m$ 和其定义域中一点 $\mathbf{r}_0\in D$。如果极限 $\lim_{\mathbf{r}\to\mathbf{r}_0}\mathbf{f}(\mathbf{r})=\mathbf{f}(\mathbf{r}_0)$ 存在，则称 $\mathbf{f}(\mathbf{r})$ 在 $\mathbf{r}_0$ 处连续（*continuous at* $\mathbf{r}_0$）。

定理 4.2. 函数 $\mathbf{f}:\mathbb{R}^n\supset D\to\mathbb{R}^m$ 在 $\mathbf{r}_0\in D$ 处连续的充要条件是 $\mathbf{f}$ 的坐标函数在 $\mathbf{r}_0$ 处都连续。

证明. 由定理4.1直接得证。 □

如果一个函数 $\mathbf{f}$ 在其定义域 D 内处处都连续，我们就直接称函数 $\mathbf{f}$ 在 D 内连续（*continuous in D*），或称函数 $\mathbf{f}$ 是 D 上的连续函数（*continuous function on D*）。以下给出的定理及其推论告诉我们：线性变换总是连续的。

定理 4.3. 设 $\mathscr{V}$、$\mathscr{W}$ 分别是数域 $\mathbb{F}$ 上的 n, m 维赋范向量空间，$\mathbf{L}: \mathscr{V} \to \mathscr{W}$ 是一个线性变换，则总存在正实数 $k > 0$ 使得 $\|\mathbf{Lx}\| \leq k \|\mathbf{x}\|, \forall \mathbf{x} \in \mathscr{V}$，当且仅当 $n = 1$ 时取等号。

证明. 设 $\{\mathbf{e}_i\}$ 是 $\mathscr{V}$ 的一组基，则任一向量 $\mathbf{x} \in \mathscr{V}$ 可表示为 $\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i$。设 $\|\cdot\|_{\mathrm{E}}$ 是欧几里得范，即 $\|\mathbf{x}\|_{\mathrm{E}} \equiv (\sum_{i=1}^n x_i^2)^{\frac{1}{2}}$。由定理2.23及其推论，包括欧几里得范在内的任意范都不依赖基的选择。故在任一基下总有 $|x_i| \leq (\sum_{i=1}^n x_i^2)^{\frac{1}{2}} = \|\mathbf{x}\|_{\mathrm{E}}$。

由引理A.1，对 $\mathscr{V}$ 上的任一范的定义 $\|\cdot\|$，总存在 $K > 0$ 使得 $\|\cdot\|_{\mathrm{E}} \leq K \|\cdot\|$。故

$$\begin{aligned} \|\mathbf{Lx}\| &= \left\| \sum_{i=1}^n x_i \mathbf{Le}_i \right\| \\ &\leq \sum_{i=1}^n |x_i| \|\mathbf{Le}_i\| \\ &\leq \sum_{i=1}^n \|\mathbf{x}\|_{\mathrm{E}} \|\mathbf{Le}_i\| \\ &\leq \sum_{i=1} K \|\mathbf{x}\| \|\mathbf{Le}_i\| = k \|\mathbf{x}\| \end{aligned}$$

其中 $k = K \sum_{i=1}^n \|\mathbf{Le}_i\|$。上面的第 2 个不等号用到范的三角不等式和范的调和性，第三个不等号用到证明开头提到的事实，第四个不等号用到引理A.1。由于引理A.1的等号当且仅当 $n = 1$ 时成立，这一条件已经强于其余两个不等号的取等充要条件，故整个不等式的取等充要条件就是 $n = 1$。 □

推论 4.3.1. 线性变换是连续函数。

证明. 由定理4.3，对任意 $\mathbf{x}, \mathbf{x}_0 \in \mathscr{V}$ 有 $\|\mathbf{Lx} - \mathbf{Lx}_0\| = \|\mathbf{L}(\mathbf{x} - \mathbf{x}_0)\| \leq k \|\mathbf{x} - \mathbf{x}_0\|$，故对任一 $\varepsilon > 0$ 总可取 $\delta = \frac{\varepsilon}{k}$ 使得只要 $\|\mathbf{x} - \mathbf{x}_0\| < \delta$ 就有 $\|\mathbf{Lx} - \mathbf{Lx}_0\| \leq k \|\mathbf{x} - \mathbf{x}_0\| < k\delta = \varepsilon$。 □

由定理4.1和4.2，向量函数的极限存在性和连续性总可以通过其坐标函数的极限存在性和连续性得到证明，使得问题回到本科高等数学课的范围内。以下函数极限的性质，也已在本科的高等数学课中教过了，在此列出而不作证明。

定理 4.4 (函数极限的基本性质). 在实数域内，以下命题成立:

1. 设 a, b 是常数，则 $\lim_{x\to a} b = b$。
2. 若 a 是常数，则 $\lim_{x\to a} x = a$。
3. 设 $\lim_{x\to a} f(x) = L$，c 是常数，则 $\lim_{x\to a} cf(x) = c \lim_{x\to a} f(x) = cL$。
4. 设 $\lim_{x\to c} f(x) = L, \lim_{x\to c} g(x) = M$，则 $\lim_{x\to c} [f(x) + g(x)] = L + M$。
5. 设 $\lim_{x\to c} f(x) = L, \lim_{x\to c} g(x) = M$，则 $\lim_{x\to c} [f(x) g(x)] = LM$。

6.（夹逼定理）若 c 是常数，$g(x) \le f(x) \le h(x)$ 至少在点 $x = c$ 之外都成立，且 $\lim_{x\to c} g(x) = \lim_{x\to c} h(x) = L$，则 $\lim_{x\to c} f(x) = L$。

7. 设 $\lim_{x\to c} f(x) = L$，则 $|\lim_{x\to c} f(x)| = |L| = \lim_{x\to c} |f(x)|$。

4.3　向量函数的微分与导数

4.3.1　一元函数的导数与微分（回顾）

我们在本科的高等数学课上已经学过一元函数的导数与微分。我们复述它们，一是为了明确符号的记法，二是为了与后面介绍的多维的情况作对比。

回顾高阶无穷小定义，简述为：若函数 $f(x)$ 在点 x_0 处有极限 $\lim_{x\to x_0} f(x) = 0$，则称函数 $f(x)$ 是当 $x \to x_0$ 时的一个无穷小。若函数 $f(x), g(x)$ 是当 $x \to x_0$ 时的无穷小，且 $\lim_{x\to x_0} f(x)/g(x) = 0$，则称 $f(x)$ 是当 $x \to x_0$ 时 $g(x)$ 的一个高阶无穷小。

定义 4.7 (一元函数的导数). [8]定义 2.1.1 设函数 $y = f(x)$ 在点 x_0 的某个邻域内有定义，当自变量 x 在点 x_0 处取得改变量 Δx，且点 $x_0 + \Delta x$ 在上述邻域内时，相应地，函数的改变量为

$$\Delta y = f(x_0 + \Delta x) - f(x_0),$$

如果极限

$$\lim_{\Delta x\to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x\to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

存在，称函数 $f(x)$ 在点 x_0 处*可导*（或*存在导数*），上述极限值称为*函数 f 在点 x_0 处的导数*，记为

$$f'(x_0), \quad y'|_{x=x_0}, \quad \left.\frac{df}{dx}\right|_{x=x_0}, \quad 或 \left.\frac{dy}{dx}\right|_{x=x_0},$$

即

$$f'(x_0) = \lim_{\Delta x\to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}。$$

若该极限不存在，称*函数 $f(x)$ 在 x_0 处不可导*。

定义 4.8 (一元函数的微分). [8]定义 2.5.1 设函数 $y = f(x)$ 在点 x_0 的某邻域内有定义，当 x 在点 x_0 处获得增量 Δx 时，如果相应的函数增量 $\Delta y = f(x_0 + \Delta x) - f(x_0)$ 可以表示为

$$\Delta y = A\Delta x + o(\Delta x),$$

其中，常数 A 与 Δx 无关（仅与 x_0 有关），而 $o(\Delta x)$ 是较 $\Delta x\,(\Delta x \to 0)$ 的高阶无穷小，则称函数 $y = f(x)$ *在点 x_0 处可微*（*differentiable at x_0*）。且称 $A\Delta x$ 为函数 $y = f(x)$ 在点 x_0

处对应于自变量增量 Δx 的微分（*differential of* $y = f(x)$ *at* x_0），记作 dy，即

$$dy = A\Delta x,$$

也常把 $A\Delta x\,(A \neq 0)$ 称为函数增量 Δy 的线性主部。

在 Δy 的分解式中，由于其值主要取决于线性主部 $A\Delta x$，因此当 $\Delta x \to 0$ 时，我们可以写

$$\Delta y \approx dy = A\Delta x。$$

定理 4.5. [8]定理 2.5.1 函数 $y = f(x)$ 在点 x_0 处可微的充分必要条件是：函数 $y = f(x)$ 在点 x_0 处可导，且当 $y = f(x)$ 在点 x_0 处可微时，其微分是

$$dy = f'(x_0)\,\Delta x。$$

证明. 略[8]p. 103。 □

关于符号"d"的意义，这里给出一个比高等数学课本[8]p. 104更仔细的说明。

由函数微分的定义可知，函数 $f(x)$ 在点 x_0 处对应于自变量增量 Δx 的微分 $dy = f'(x_0)\Delta x$，实际上是一个由点 x_0 引出的函数，记为 $df_{x_0} : \mathbb{R} \to \mathbb{R}, df_{x_0}(x) = f'(x_0)\,x, \forall x \in I$，其中区域 I 是所有满足 $x_0 + x$ 在函数可微的 x_0 的邻域的所有实数 x 的集合。

若视 x 本身为一个恒等映射，即 $x(p) = p, \forall p \in \mathbb{R}$，则函数 x 在点 p 处的微分也是一个由点 p 引出的函数，且 $dx_p(u) = x'(p)\,u = 1 \times u, \forall u \in \mathbb{R}$，也是一个恒等映射。同时我们看到，函数 $dx_p(u)$ 是 $u \to 0$ 的无穷小。

考虑复合映射 $f(x(p))$ 在点 p_0 处的微分，它是由点 p_0 引出的函数，

$$d(f \circ x)_{p_0}(u) = \left.\frac{d(f \circ x)}{dp}\right|_{p=p_0} u = f'(x(p_0))x'(p_0)\,u = f'(p_0)\,dx_{p_0}(u)$$

而且函数 $d(f \circ x)_{p_0}(u)$ 与 $dx_{p_0}(u)$ 是 $u \to 0$ 时的同阶无穷小，$f'(p_0)$ 就是这两个无穷小的比值，该结论对所有 $p_0, u \in \mathbb{R}$ 都成立，因此若去掉 p_0 和 u 简记：

$$df(x) = f'dx$$

则

$$f' = \frac{df}{dx}。$$

因此，当我们把导数当作分数来处理时，实际是在上述的意义上视导数为两个微分作为同阶无穷小的比值。

定义 4.9 (一元向量函数的导数)**.** 设函数 $\mathbf{f}:\mathbb{R}\supset D\to\mathbb{R}^n$ 的极限

$$\lim_{h\to 0}\frac{\mathbf{f}(t+h)-\mathbf{f}(t)}{h}$$

存在，则称函数 $\mathbf{f}(x)$ 在 $x=t$ 处可导。该极限是函数 $\mathbf{f}(x)$ 在 $x=t$ 处的导数，记为

$$\left.\frac{d\mathbf{f}(x)}{dx}\right|_{x=t}\equiv\lim_{h\to 0}\frac{\mathbf{f}(t+h)-\mathbf{f}(t)}{h}$$

由一元标量函数求导的法则可证有以下一元向量函数求导法则：对任意函数 $\mathbf{f}:\mathbb{R}\supset D\to\mathbb{R}^n,\mathbf{g}:D\to\mathbb{R}^n$，

- $\frac{d\mathbf{c}}{dt}=0,\mathbf{c}\in\mathbb{R}^n$ 是常向量
- $\frac{d}{dt}(\alpha\mathbf{f}+\beta\mathbf{g})=\alpha\frac{d}{dt}\mathbf{f}+\beta\frac{d}{dt}\mathbf{g},\forall\alpha,\beta\in\mathbb{R}$
- $\frac{d}{dt}[u(t)\,\mathbf{f}(t)]=\frac{du}{dt}\mathbf{f}+u\frac{d}{dt}\mathbf{f},\forall u:D\to\mathbb{R}$
- $\frac{d}{dt}(\mathbf{f}\cdot\mathbf{g})=\frac{d\mathbf{f}}{dt}\cdot\mathbf{g}+\mathbf{f}\cdot\frac{d\mathbf{g}}{dt}$
- $\frac{d}{dt}(\mathbf{f}\times\mathbf{g})=\frac{d\mathbf{f}}{dt}\times\mathbf{g}+\mathbf{f}\times\frac{d\mathbf{g}}{dt}$

定义 4.10 (多元标量值函数的偏导数)**.** 给定函数 $f:\mathbb{R}^n\supset D\to\mathbb{R}$，若对某一 $i\in\{1,\cdots,n\}$，极限

$$\lim_{t\to 0}\frac{f(\cdots,x_i+h,\cdots)-f(\cdots,x_i,\cdots)}{t}$$

当 $x_i=x_{i0}$ 时存在，则称该极限为函数 $f(\mathbf{x})$ 对第 i 个变量 x_i 在 $x_i=x_{i0}$ 处的的偏导数（*partial derivative*），记为

$$\left.\frac{\partial f(\mathbf{x})}{\partial x_i}\right|_{x_i=x_{i0}}\equiv\lim_{t\to 0}\frac{f(\cdots,x_i+h,\cdots)-f(\cdots,x_i,\cdots)}{t}$$

定理 4.6. * 如果函数 $f:\mathbb{R}^n\to\mathbb{R}$ 的一个二阶混合偏导数 $\frac{\partial^2 f}{\partial x_i\partial x_j}$ 在 $\mathbb{R}^2$ 的一个开子集 S 上处处连续，则二阶混合偏导数 $\frac{\partial^2 f}{\partial x_j\partial x_i}$ 在 S 上处处存在，且 $\frac{\partial^2 f}{\partial x_i\partial x_j}=\frac{\partial^2 f}{\partial x_j\partial x_i}$。

证明. 略† □

定义 4.11 (向量函数的偏导数)**.** 函数 $\mathbf{f}:\mathbb{R}^n\supset D\to\mathbb{R}^m$ 对第 i 个变量在 $x_i=x_{i0}$ 处的偏导数定义如下向量

$$\left.\frac{\partial\mathbf{f}(\mathbf{x})}{\partial x_i}\right|_{x_i=x_{i0}}\equiv\begin{pmatrix}\left.\frac{\partial f_1(\mathbf{x})}{\partial x_i}\right|_{x_i=x_{i0}}\\ \vdots\\ \left.\frac{\partial f_m(\mathbf{x})}{\partial x_i}\right|_{x_i=x_{i0}}\end{pmatrix}\in\mathbb{R}^m$$

其中，$\mathbf{x}=(x_1,\cdots,x_n)^\intercal\in\mathbb{R}^n$。

*即高等数学[9]p. 16定理 7.2.1 向 n 维的推广。

†进一步了解：Symmetry of second derivatives。

4.3.2 向量函数的微分和导数

回顾多元标量函数的全微分的定义[9]“定义 7.3.1”,p. 19——

定义 4.12 (多元标量值函数的微分). 若函数 $z=f(x,y)$ 在点 $P_0(x_0,y_0)$ 的全增量

$$\Delta z=f(x_0+\Delta x,y_0+\Delta y)-f(x_0,y_0)$$

可表示为

$$\Delta z=A\Delta x+B\Delta y+o(\rho)$$

其中 A,B 只与点 (x_0,y_0) 有关，而与 $\Delta x,\Delta y$ 无关。又 $\rho=\sqrt{(\Delta x)^2+(\Delta y)^2}$,$o(\rho)$ 是当 $\rho\to 0$ 时 ρ 的高阶无穷小，则称函数 $z=f(x,y)$ 在点 $P_0(x_0,y_0)$ 可微分，且把 Δz 的线性主部 $A\Delta x+B\Delta y$ 称为函数 $z=f(x,y)$ 在点 $P_0(x_0,y_0)$ 的全微分（*differential*），记作

$$dz|_{x=x_0,y=y_0}=A\Delta x+B\Delta y\text{或}df(x_0,y_0)=A\Delta x+B\Delta y$$

如果函数 $z=f(x,y)$ 在区域 D 内每一点都可微，则称这函数在 D 内可微（*differentiable in D*）。

如果函数 $z=f(x,y)$ 在点 $P_0(x_0,y_0)$ 处可微，则有

$$\begin{aligned}
&f(x_0+\Delta x,y_0+\Delta y)-f(x_0,y_0)=A\Delta x+B\Delta y+o(\rho)\\
\Leftrightarrow & o(\rho)=f(x_0+\Delta x,y_0+\Delta y)-f(x_0,y_0)-(A\Delta x+B\Delta y)\\
\Leftrightarrow & \lim_{\rho\to 0}\frac{f(x_0+\Delta x,y_0+\Delta y)-f(x_0,y_0)-(A\Delta x+B\Delta y)}{\sqrt{\Delta x^2+\Delta y^2}}=0\\
\Leftrightarrow & \lim_{\rho\to 0}\frac{\Delta z-dz}{\rho}=0
\end{aligned}$$

其中 $\rho=\sqrt{\Delta x^2+\Delta y^2}$。上面的极限式是定义4.12的等价定义式。我们其实可以不引入某个高阶无穷小 $o(\rho)$，直接用使该极限式成立的 A,B 的存在性来定义函数的可微性。下面我们按照把函数微分的定义推广到一般的向量函数。

定义 4.13 (向量值函数的微分). 若函数 $\mathbf{f}:\mathbb{R}^n\supset D\to\mathbb{R}^m$ 在其定义域内某点 $\mathbf{x}_0\in D$ 处，存在一个线性变换 $\mathbf{L}:\mathbb{R}^n\to\mathbb{R}^m$ 使得对任意 $\mathbf{x}_0$ 的邻域 $N(\mathbf{x}_0)$ 中的点 $\mathbf{x}\in N(\mathbf{x}_0)$，

$$\lim_{\mathbf{x}\to\mathbf{x}_0}\frac{\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0)-\mathbf{L}(\mathbf{x}-\mathbf{x}_0)}{\|\mathbf{x}-\mathbf{x}_0\|}=\mathbf{0}$$

就称函数 $\mathbf{f}(\mathbf{x})$ 在 $\mathbf{x}_0$ 处可微分（*differentiable at* $\mathbf{x}_0$）。向量 $\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})\equiv\mathbf{L}\mathbf{x}-\mathbf{L}\mathbf{x}_0$，称函数 $\mathbf{f}(\mathbf{x})$ 在 $\mathbf{x}_0$ 处的微分（*the differential of* $\mathbf{f}(\mathbf{x})$ *at* $\mathbf{x}_0$）。如果函数 $\mathbf{f}(\mathbf{x})$ 在开集 $S\subseteq D$ 内的每一点上都可微分，则称函数是S 上的可微函数（*differentiable function on* S）。

我们可以根据向量函数的定义来看出，定义4.13就是定义4.12的推广。延用定义4.13的设定，设 $\mathbf{f}=(f_1,\cdots,f_m)^\intercal$，$\mathbf{x}=(x_1,\cdots,x_n)^\intercal$，$\mathbf{x}_0=(x_{01},\cdots,x_{0n})^\intercal$，则函数 $\mathbf{f}(\mathbf{x})$ 在 $\mathbf{x}_0$ 处的全增量是：

$$\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0)=\begin{pmatrix} f_1(x_1,\cdots,x_n)-f_1(x_{01},\cdots,x_{0n}) \\ \vdots \\ f_m(x_1,\cdots,x_n)-f_m(x_{01},\cdots,x_{0n}) \end{pmatrix}$$

可见我们实际考虑的是 m 个 n 元标量值函数分别在点 $x_{01},\cdots,x_{0n}$ 处的全增量。如果在这些点上这些标量值函数分别都可微，则 $\mathbf{f}$ 的每个坐标函数的全增量都可按定义4.12写成

$$f_i(x_1,\cdots,x_n)-f_i(x_{01},\cdots,x_{0n})=\sum_{j=1}^{n}L_{ji}(x_j-x_{0j})+o(\rho),i=1,\cdots,m$$

其中，L_{ij} 是“线性主部”的系数，只与 $(x_{01},\cdots,x_{0n})$ 有关，$\rho=\left(\sum_{j=1}^{n}(x_j-x_{0j})^2\right)^{1/2}=\|\mathbf{x}-\mathbf{x}_0\|$，$o(\rho)$ 是当 $\rho\to 0$ 时 ρ 的高阶无穷小。这等价于如下极限式

$$\lim_{\rho\to 0}\frac{1}{\|\mathbf{x}-\mathbf{x}_0\|}\begin{pmatrix} f_1(x_1,\cdots,x_n)-f_1(x_{01},\cdots,x_{0n})-\sum_{j-1}^{n}L_{1j}(x_j-x_{0j}) \\ \vdots \\ f_m(x_1,\cdots,x_n)-f_m(x_{01},\cdots,x_{0n})-\sum_{j-1}^{n}L_{mj}(x_j-x_{0j}) \end{pmatrix}=\begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix}$$

若线性变换 $\mathbf{L}$ 在标准基下的矩阵坐标就是 L_{ij}，则上式等价于定义4.13的极限式。

我们以一个易于作图的例子说明函数微分的几何意义。函数 $f:\mathbb{R}^2\to\mathbb{R}$ 的图像是 3 维欧几里得空间中的一个曲面（如图4.7）。f 在点 $\mathbf{x}_0$ 处的微分 $(\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}f(\mathbf{x}))(\mathbf{x}-\mathbf{x}_0)$ 定义了该曲面在 $\mathbf{x}=\mathbf{x}_0$ 处的切平面。随着 $\mathbf{x}$ 接近 $\mathbf{x}_0$，函数 f 的全增量 $f(\mathbf{x})-f(\mathbf{x}_0)$ 将比 $\mathbf{x}$ 接近 $\mathbf{x}_0$ 更快地接近切平面。

我们接下来考察函数可微与可导的关系，即函数可微的充分和必要条件*。首先，以下定理是函数在某点处可微分的必要条件，它同时也给出了 $\mathbf{L}$ 或其坐标矩阵 L_{ij} 的计算方法。

定理 4.7. 设函数 $\mathbf{f}:\mathbb{R}^n\supset D\to\mathbb{R}^m$ 在点 $\mathbf{x}_0\in D$ 处可微分，即存在线性变换 $\mathbf{L}\in\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$ 满足

$$\lim_{\Delta\mathbf{x}\to\mathbf{0}}\frac{\mathbf{f}(\mathbf{x}_0+\Delta\mathbf{x})-\mathbf{f}(\mathbf{x}_0)-\mathbf{L}\Delta\mathbf{x}}{\|\Delta\mathbf{x}\|}=\mathbf{0}$$

则 $\mathbf{f}$ 的每个坐标函数在 $\mathbf{x}_0$ 处的每个偏导数

$$\left.\frac{\partial f_i(\mathbf{x})}{\partial x_j}\right|_{\mathbf{x}=\mathbf{x}_0},i=1,\cdots,m,j=1,\cdots,n$$

*这部分内容可以与本科高等数学课上的相应内容对比学习，特别是关于必要非充份条件和充份非必要条件的例子（“二、全微分存在的条件”[9]p. 20。）

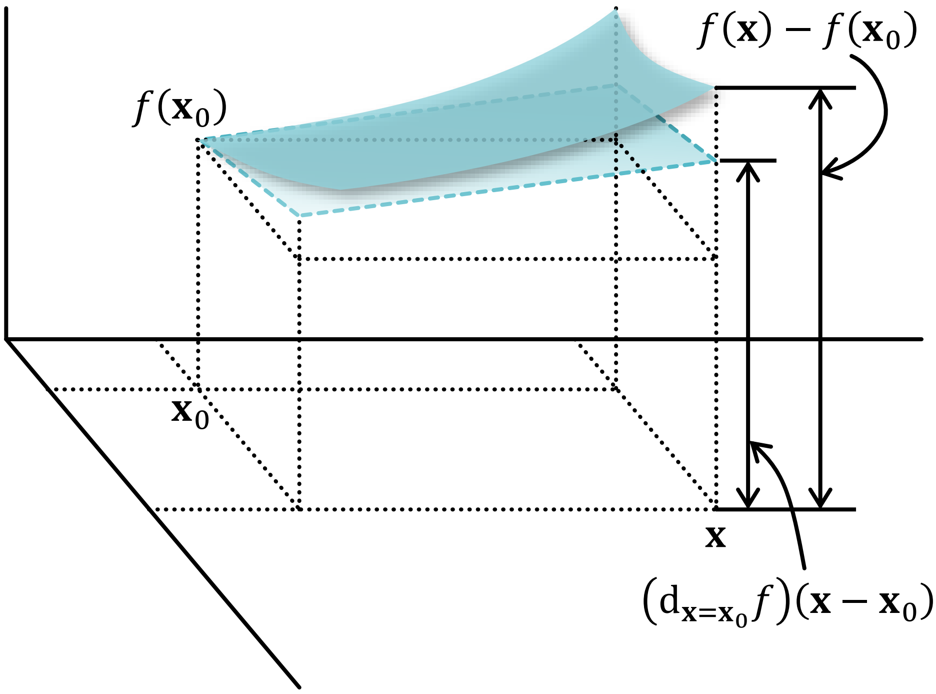


图 4.7: 向量函数的极限（定义4.5）的图像化理解。

都存在。若 $\{\hat{\mathbf{e}}_1,\cdots\hat{\mathbf{e}}_n\},\{\hat{\mathbf{u}}_1,\cdots,\hat{\mathbf{u}}_m\}$ 分别是 $\mathbb{R}^n,\mathbb{R}^m$ 的标准基，则

$$\mathbf{L}\hat{\mathbf{e}}_j=\sum_{i=1}^{m}\left(\left.\frac{\partial f_i(\mathbf{x})}{\partial x_j}\right|_{\mathbf{x}=\mathbf{x}_0}\right)\hat{\mathbf{u}}_i, j=1,\cdots,n$$

证明. 见附录。 □

注意到 $\mathbf{L}\hat{\mathbf{e}}_j$ 其实就是线性变换的第 j 列，故上述定理说明，如果函数在某点处可微分，则其微分的线性变换 $\mathbf{L}$ 就是函数的偏导数所形成的矩阵：

$$(\mathbf{L})=\begin{pmatrix}\left.\frac{\partial f_1(\mathbf{x})}{\partial x_1}\right|_{\mathbf{x}=\mathbf{x}_0} & \cdots & \left.\frac{\partial f_1(\mathbf{x})}{\partial x_n}\right|_{\mathbf{x}=\mathbf{x}_0}\\ \vdots & \ddots & \vdots\\ \left.\frac{\partial f_m(\mathbf{x})}{\partial x_1}\right|_{\mathbf{x}=\mathbf{x}_0} & \cdots & \left.\frac{\partial f_m(\mathbf{x})}{\partial x_n}\right|_{\mathbf{x}=\mathbf{x}_0}\end{pmatrix}$$

以下定理解决了函数微分的线性变换 $\mathbf{L}$ 的唯一性。

定理 4.8. 设函数 $\mathbf{f}:\mathbb{R}^n\supset D\to\mathbb{R}^m$ 在点 $\mathbf{x}_0\in D$ 处可微分，即存在线性变换 $\mathbf{L}\in\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$ 满足

$$\lim_{\Delta\mathbf{x}\to\mathbf{0}}\frac{\mathbf{f}(\mathbf{x}_0+\Delta\mathbf{x})-\mathbf{f}(\mathbf{x}_0)-\mathbf{L}\Delta\mathbf{x}}{\|\Delta\mathbf{x}\|}=\mathbf{0}$$

则 $\mathbf{L}$ 是唯一的。

证明. 见附录。 □

定理4.7和4.8共同构成了函数可微分的必要非充份条件。也就是说，并非每当函数在某点处的所有偏导数都存在，该函数就一定在该点可微[9]“例 2”,p. 21。不过，有了唯一性，我们至少可以把函数微分的线性变换 $\mathbf{L}$ 定义为函数的导数，具体地——

定义 4.14 (向量函数的导数). 设函数 $\mathbf{f}: \mathbb{R}^n \supset D \to \mathbb{R}^m$ 在点 $\mathbf{x}_0 \in D$ 处可微分，即存在线性变换 $\mathbf{L} \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$ 满足

$$\lim_{\Delta\mathbf{x}\to\mathbf{0}} \frac{\mathbf{f}(\mathbf{x}_0+\Delta\mathbf{x})-\mathbf{f}(\mathbf{x}_0)-\mathbf{L}\Delta\mathbf{x}}{\|\Delta\mathbf{x}\|} = \mathbf{0}$$

则称线性变换 $\mathbf{L}$ 是函数 $\mathbf{f}(\mathbf{x})$ 在 $\mathbf{x}_0$ 处的导数（*derivative of* $\mathbf{f}(\mathbf{x})$ *at* $\mathbf{x}_0$），记为

$$\mathbf{L} \equiv \mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})$$

$\mathbf{L}$ 在标准基下的坐标矩阵称为函数 $\mathbf{f}(\mathbf{x})$ 在 $\mathbf{x}_0$ 处的雅可比矩阵（*Jacobian matrix*），$\mathbf{L}$ 的行列式 $\det\mathbf{L}$ 称函数 $\mathbf{f}(\mathbf{x})$ 在 $\mathbf{x}_0$ 处的雅可比行列式（*Jacobian determinant*）。

下面我们给出一个函数在某点处可微的充分非必要条件（即未必一定要满足该条件函数才可微，但满足该条件函数必可微）[9]“例 3”,p. 23。

定理 4.9. 若函数 $\mathbf{f}: \mathbb{R}^n \supset D \to \mathbb{R}^m$ 的定义域 D 是开集，偏微分 $\frac{\partial f_i}{\partial x_j}, i=1,\cdots,n, j=1,\cdots,m$ 在 D 内都连续，则 $\mathbf{f}$ 在 D 内均可微分。

证明. 见附录。 □

我们不把 $\mathbf{L}_\mathbf{x}$ 记为 $\frac{d\mathbf{f}}{d\mathbf{x}}$，因为线性变换并非“两个向量的商”。

4.3.3　向量函数的导函数、连续可微函数

定义 4.15 (向量函数的导函数). 设函数 $\mathbf{f}: \mathbb{R}^n \subseteq D \to \mathbb{R}^m$ 是开集 $S \subseteq D$ 上的可微函数，则函数 $\mathbf{f}$ 在 S 上的导函数（*derivative function*）是一个线性变换值函数 $\mathbf{L}: S \to \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$,

$$\mathbf{L}(\mathbf{x}) \equiv d_{\mathbf{x}'=\mathbf{x}}\mathbf{f}(\mathbf{x}'), \quad \forall\mathbf{x}\in S$$

注意，向量函数的导函数的函数值是线性变换。向量函数在不同点上的导数是不同的线性变换。在上述定义中的“$\mathbf{L}(\mathbf{x})$”，不是指一个线性变换作用于一个向量，而是一个线性变换值函数及其自变量。此时 $\mathbf{L}(\mathbf{x})$ 可以作用于一个向量 $\mathbf{u}$ 得到另一个向量 $\mathbf{v}=\mathbf{L}(\mathbf{x})\mathbf{u}$。到底 $\mathbf{L}$ 后

面的括号表示哪种意义，在记法上无法区分，但是大多数情况下可根据语境区分。本讲义在必要处会说明一种记法到底表示哪种意义。

接下来我们准备讨论导函数的连续性，这需要先引入“线性变换的范”的概念。

由定理4.3，对任一线性变换 $\mathbf{L}:\mathbb{R}^n\to\mathbb{R}^m$ 皆存在一个正实数 $k>0$ 满足 $\|\mathbf{Lx}\|\leq k\|\mathbf{x}\|,\forall\mathbf{x}\in\mathbb{R}^n$。于是我们定义——

定义 4.16 (线性变换的范). 设 $\mathscr{V},\mathscr{W}$ 是数域 $\mathbb{F}$ 上的赋范向量空间，线性变换 $\mathbf{L}:\mathscr{V}\to\mathscr{W}$ 的范 $\|\mathbf{L}\|$ 为满足 $\|\mathbf{Lx}\|\leq k\|\mathbf{x}\|,\forall\mathbf{x}\in\mathbb{R}^n$ 的正实数 k 的最大下界（*infimum*），即

$$\|\mathbf{L}\|\equiv\inf\{k|k>0,\|\mathbf{Lx}\|\leq k\|\mathbf{x}\|\,\forall\mathbf{x}\}$$

这一最大下界的存在性和唯一性是显然的。我们无需知道它具体数值。可验证，以上定义的范符合范的一般要求。又由于不同的范的定义是等价的（定理A.1），故后续命题的证明过程每当需要线性变换的范的定义时都不妨通过上述这种范的定义来求证。由此定义我们可立即获得性质：$\|\mathbf{Lx}\|\leq\|\mathbf{L}\|\|\mathbf{x}\|$。

有了向量函数的导函数的定义以及线性变换的范的定义，我们可以很容易理解导函数的连续性是什么意思。按照函数连续性的定义，若函数 $\mathbf{L}:S\to\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$ 在 S 上连续，则对任意 $\varepsilon>0$ 总存在 $\delta>0$ 使得只要 $\|\mathbf{x}-\mathbf{x}_0\|<\delta,\mathbf{x},\mathbf{x}_0\in S$ 就有 $\|\mathbf{L}(\mathbf{x})-\mathbf{L}(\mathbf{x}_0)\|<\varepsilon$。

下面介绍函数“连续可微”的概念，它大致上说的是：导函数连续，则函数连续可微。准确定义如下。

定义 4.17 (连续可微函数). 若函数 $\mathbf{f}:\mathbb{R}^n\supset D\to\mathbb{R}^m$ 在开集 $S\subseteq D$ 上有导函数 $\mathbf{L}:S\to\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$，且其为 S 上的连续函数，则称函数 $\mathbf{f}(\mathbf{x})$ 是 S 上的连续可微（*continuously differentiable*）函数。若函数 $\mathbf{f}$ 的前 k 阶导数在 S 上都存在并连续，则称 $\mathbf{f}$ 属于 D 上的 k 阶连续可微函数的集合 $\mathcal{C}^k$。

推论 4.9.1. 函数 $\mathbf{f}:\mathbb{R}^n\supset D\to\mathbb{R}^m$ 在开集 D 上连续可微当且仅当函数 $\mathbf{f}$ 在 D 上的偏导数

$$\left.\frac{\partial f_i(\mathbf{x})}{\partial x_j}\right|_{\mathbf{x}\in D},i=1,\cdots,m,j=1,\cdots,n$$

都存在且连续。

结合这一推论和定理4.9可知函数连续可微是函数可微的充份非必要条件。

4.3.4 方向导数与梯度

刚才在介绍雅可比矩阵的时候，我们利用全微分的性质考虑过以下极限：

$$\lim_{t\to0}\frac{\mathbf{f}(\mathbf{x}_i)-\mathbf{f}(\mathbf{x}_0)}{t},i=1,\cdots,n$$

并知道它就是 $\mathbf{f}(\mathbf{x})$ 在 $\mathbf{x}_0=(x_{01},\cdots,x_{0n})^\intercal$ 处对 x_{0i} 的偏导数，因为上式 $\Leftrightarrow$

$$\lim_{t\to 0}\frac{\mathbf{f}(\mathbf{x}_0+t\widehat{\mathbf{e}}_i)-\mathbf{f}(\mathbf{x}_0)}{t}, i=1,\cdots,n$$

现在我们把 $\widehat{\mathbf{e}}_j$ 改为任意单位向量 $\widehat{\mathbf{n}}$，引入方向导数的定义。

定义 4.18 (方向导数). 设函数 $\mathbf{f}:\mathbb{R}^n\supset D\to\mathbb{R}^m$ 在点 $\mathbf{x}_0\in D$ 的某邻域 N 有定义，给定单位向量 $\widehat{\mathbf{n}}\in\mathbb{R}^n$，若极限

$$\lim_{t\to 0^+}\frac{\mathbf{f}(\mathbf{x}_0+t\widehat{\mathbf{n}})-\mathbf{f}(\mathbf{x}_0)}{t}$$

存在，则称此极限值为函数 $\mathbf{f}(\mathbf{x})$ 在点 $\mathbf{x}_0$ 处沿方向 $\widehat{\mathbf{n}}$ 的方向导数（*directional derivative*），记作

$$\left.\frac{\partial\mathbf{f}(\mathbf{x})}{\partial\widehat{\mathbf{n}}}\right|_{\mathbf{x}=\mathbf{x}_0}$$

以下定理使得函数的方向导数可用该函数的导数来计算。

定理 4.10. 若函数 $\mathbf{f}:\mathbb{R}^n\supset D\to\mathbb{R}^m$ 在点 $\mathbf{x}_0\in D$ 的某邻域 N 可微分，则给任一单位向量 $\widehat{\mathbf{n}}\in\mathbb{R}^n$，就有

$$\left.\frac{\partial\mathbf{f}(\mathbf{x})}{\partial\widehat{\mathbf{n}}}\right|_{\mathbf{x}=\mathbf{x}_0}=\left(\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})\right)\widehat{\mathbf{n}}$$

证明. 由函数 $\mathbf{f}$ 的可微条件，自变量的增量 $t\widehat{\mathbf{n}}$ 造成的函数增量 $\mathbf{f}(\mathbf{x}_0+t\widehat{\mathbf{n}})-\mathbf{f}(\mathbf{x}_0)$ 满足

$$\begin{aligned}\lim_{t\to 0}\frac{\mathbf{f}(\mathbf{x}_0+t\widehat{\mathbf{n}})-\mathbf{f}(\mathbf{x}_0)-\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})\,(t\widehat{\mathbf{n}})}{\|t\widehat{\mathbf{n}}\|}&=\mathbf{0}\\ \Leftrightarrow\lim_{t\to 0}\frac{1}{\|\widehat{\mathbf{n}}\|}\left\|\frac{\mathbf{f}(\mathbf{x}_0+t\widehat{\mathbf{n}})-\mathbf{f}(x_0)}{t}-\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})\,\widehat{\mathbf{n}}\right\|&=\mathbf{0}\\ \Leftrightarrow\lim_{t\to 0}\frac{\mathbf{f}(\mathbf{x}_0+t\widehat{\mathbf{n}})-\mathbf{f}(\mathbf{x}_0)}{t}&=\left(\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})\right)\widehat{\mathbf{n}}\end{aligned}$$

□

我们比较一个函数 $\mathbf{f}(\mathbf{x})$ 在 $\mathbf{x}_0$ 处的微分及其在 $\mathbf{x}_0$ 处关于 $\widehat{\mathbf{n}}$ 的方向导数可以发现，后者就是令前者中的 $\mathbf{x}-\mathbf{x}_0=\widehat{\mathbf{n}}$。注意到，函数在某点处关于标准基向量的方向导数就是函数关于该点相应分量的偏导数。

方向导数的几何意义是函数在某点处朝相应方向的变化率。定理4.10表明，拿一个函数在某点处的导数作用于一个单位向量，就可以得该函数在该点处朝该方向的变化率。这也是函数的导数的重要几何意义。

定义 4.19 (向量函数的梯度). 若函数 $\mathbf{f}:\mathbb{R}^n\supset D\to\mathbb{R}^m$ 在点 $\mathbf{x}_0\in D$ 处的导数 $\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})$ 存在，则记

$$\nabla_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})\equiv\left(\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})\right)^\intercal$$

称为函数 $\mathbf{f}(\mathbf{x})$ 在 $\mathbf{x}_0$ 处的梯度（*gradient*）。

由线性变换的转置的定义，若视 $\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x}) \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$，则 $\nabla_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x}) \in \mathscr{L}(\mathbb{R}^{m*}, \mathbb{R}^{n*})$。由有限维里斯表示定理（定理2.19），必存在一组 $\mathbb{R}^m$ 的规范正交基 $\{\widehat{\mathbf{e}}_i\}_{i=1}^m$，使得函数 $\mathbf{f}(\mathbf{x})$ 在点 $\mathbf{x}_0$ 的导数和梯度满足关系

$$(\widehat{\mathbf{e}}_i|\left(\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})\right)(\mathbf{x}-\mathbf{x}_0)) = \left(\nabla_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})\,\widehat{\mathbf{e}}_i|\mathbf{x}-\mathbf{x}_0\right), \quad i = 1, \cdots, m$$

其中 $\mathbf{x} \in N(\mathbf{x}_0)$，$N(\mathbf{x}_0)$ 是 $\mathbf{x}_0$ 某邻域。

用梯度表示定理4.10，则函数 $\mathbf{f}(\mathbf{x})$ 在 $\mathbf{x}_0$ 处的关于 $\widehat{\mathbf{n}}$ 的方向导数

$$\left.\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \widehat{\mathbf{n}}}\right|_{\mathbf{x}=\mathbf{x}_0} = \left(\nabla_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})\right)^{\intercal}\widehat{\mathbf{n}}$$

4.3.5 复合函数求导的链式法则、反函数定理、隐函数定理

定理 4.11 (复合函数求导的链式法则). 如果函数 $\mathbf{f}: \mathbb{R}^n \supset D \to \mathbb{R}^m$ 在 $\mathbf{x}_0 \in D$ 处可微分；函数 $\mathbf{g}: \mathbb{R}^n \supset E \to \mathbb{R}^p$ 在 $\mathbf{f}(\mathbf{x}_0) \in E \cap D$ 处可微分，则复合函数 $\mathbf{g} \circ \mathbf{f}$ 在 $\mathbf{x}_0$ 处可微分，且其导数

$$\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{g} \circ \mathbf{f}(\mathbf{x}) = \mathrm{d}_{\mathbf{y}=\mathbf{f}(\mathbf{x}_0)}\mathbf{g}(\mathbf{y})\,\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})$$

证明. 见附录。 □

定理 4.12 (反函数定理). 设函数 $\mathbf{f}: \mathbb{R}^n \supset D \to \mathbb{R}^n$ 在 D 上是连续可微函数，且其在 $\mathbf{x}_0 \in D$ 处的导数 $\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})$ 是可逆线性变换。则在 $\mathbf{x}_0$ 的某邻域 N 上，函数 $\mathbf{f}$ 有连续可微的逆函数 $\mathbf{f}^{-1}$；N 在函数 $\mathbf{f}$ 下的像集 $\mathbf{f}(N)$ 是开集；且 $\mathbf{f}^{-1}(\mathbf{y})$ 在 $\mathbf{y} = \mathbf{f}(\mathbf{x}_0)$ 处的导数是 $\mathbf{f}(\mathbf{x})$ 在 $\mathbf{x} = \mathbf{x}_0$ 的导数的逆变换，即

$$\mathrm{d}_{\mathbf{y}=\mathbf{f}(\mathbf{x}_0)}\mathbf{f}^{-1}(\mathbf{y}) = \left(\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})\right)^{-1}$$

证明. 见附录 □

定理4.12的证明过程依赖 $\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})$ 是 $\mathbb{R}^n$ 上的双射的事实。诚然，在定理4.12中所讨论的函数 $\mathbf{f}$ 的定义域和陪域是维数相同的。因此 $\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})$ 若存在逆变换（单射），则其必为双射。若将定理4.12推广至关于函数 $\mathbf{f}: \mathbb{R}^n \to \mathbb{R}^m$，则当 $n < m$ 时定理需明确要求 $\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})$ 是双射，且仅有 $\mathbf{f}$ 存在逆函数 $\mathbf{f}^{-1}$ 的结论，$\mathbf{f}^{-1}$ 未必连续可微（读者可验证 $f(x) = x^3$ 的情况）。作为推论正式表示如下。

推论 4.12.1. 设函数 $\mathbf{f}: \mathbb{R}^n \supset D \to \mathbb{R}^m$（$n < m$）在 D 上连续可微，且其在 $\mathbf{x}_0 \in D$ 处的导数 $\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})$ 是双射线性变换，则在 $\mathbf{x}_0$ 某邻域 N 上函数 $\mathbf{f}$ 存在逆函数 $\mathbf{f}^{-1}$。

证明. 仅列出证明概要。设 $\{\mathbf{e}_1, \cdots \mathbf{e}_n\}$ 是 $\mathbb{R}^m$ 上的一组有线性无关向量组，将其补全为 $\mathbb{R}^m$ 的一组基 $\{\mathbf{e}_1, \cdots, \mathbf{e}_n, \mathbf{e}_{n+1}, \ldots, \mathbf{e}_m\}$。定义函数 $\mathbf{g}: \mathbb{R}^m \to \mathbb{R}^n$，

$$\mathbf{g}\left(\sum_{i=1}^{m} a_i \mathbf{e}_i\right) = (a_1, \cdots, a_n)$$

即取 $\mathbb{R}^m$ 上任一向量在有序基 $\{\mathbf{e}_i\}$ 下的前 n 个坐标的函数。证明:复合函数 $\mathbf{g}\circ\mathbf{f}: \mathbb{R}^n \supset D \to \mathbb{R}^n$ 满足定理4.12的条件。

□

当 $n > m$ 时，由线性变换的维数定理 $\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})$ 必不满秩，故 $\mathbf{f}$ 不存在逆函数。

反函数定理的逆命题一般不成立（仍可通过考虑 $f(x) = x^3$ 理解）。以下推论成立。

推论 4.12.2. 设函数 $\mathbf{f}: \mathbb{R}^n \supset D \to \mathbb{R}^n$ 在 $\mathbf{x}_0 \in D$ 处的导数 $\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})$ 存在。若函数 $\mathbf{f}$ 在 $\mathbf{x}_0$ 的某邻域 N 存在逆函数 $\mathbf{f}^{-1}$ 且其在 N 上可微，则 $\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})$ 是可逆线性变换。

证明. 由于函数 $\mathbf{f}$ 和 $\mathbf{f}^{-1}$ 都可微，可应用定理4.11，由 $\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}\circ\mathbf{f}^{-1}(\mathbf{x}_0) = \mathbf{I}$ 自然得出结论。 □

如果函数 $\mathbf{f}: \mathbb{R}^n \to \mathbb{R}^m$ 由函数 $\mathbf{F}: \mathbb{R}^{n+m} \to \mathbb{R}^m$ 隐含定义，我们常常把 $\mathbf{F}$ 写成这样一种映射：$\mathbf{F}: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$，$\mathbf{F}(\mathbf{x}, \mathbf{f}(\mathbf{x})) = \mathbf{0}, \forall \mathbf{x} \in \mathbb{R}^n$。此时，我们记

$$\left.\frac{\partial \mathbf{F}(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0, \mathbf{y}=\mathbf{y}_0} = \mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{F}(\mathbf{x}, \mathbf{y}_0)$$

引入这个记法有两个用处，一是为了以下例子，这个例子是后文引入物质导数的一个基础；二是为了引入隐函数定理。

例 4.9. 设函数 $\mathbf{f}: \mathbb{R}^{n+m} \to \mathbb{R}^p$ 的定义域为 $D = \mathrm{dom}\mathbf{f}$，若函数 $\mathbf{g}: \mathbb{R}^m \to \mathbb{R}^n$ 满足 $\mathrm{dom}\mathbf{g} = D$，则总可以构建函数 $\mathbf{h}: \mathbb{R}^m \to \mathbb{R}^{m+n}$ 使得 $\mathrm{dom}\mathbf{h} = D$ 且

$$\mathbf{h}(\mathbf{x}) = (\mathbf{g}(\mathbf{x}), \mathbf{x})\, \forall \mathbf{x} \in D$$

$$h_i(\mathbf{x}) = \begin{cases} g_i(\mathbf{x}), & i = 1, \cdots, n \\ x_{i-n}, & i = n+1, \cdots, n+m \end{cases}$$

这时，$\mathbf{f}(\mathbf{g}(\mathbf{x}), \mathbf{x}) = \mathbf{f}(\mathbf{h}(\mathbf{x}))$。若 $\mathbf{g}$ 在某处可微分则 $\mathbf{h}$ 在该处也可微分。由链式法则定理，在

该处有

$$
\begin{aligned}
\mathrm{d}_{\mathbf{x}}\mathbf{f}\left(\mathbf{g}\left(\mathbf{x}\right),\mathbf{x}\right) &= \frac{d\mathbf{h}\left(\mathbf{x}\right)}{d\mathbf{x}} \\
&= \mathrm{d}_{\mathbf{y}=\mathbf{h}(\mathbf{x})}\mathbf{f}\left(\mathbf{y}\right)\mathrm{d}_{\mathbf{x}}\mathbf{h}\left(\mathbf{x}\right) \\
&= \begin{pmatrix} \frac{\partial f_1}{\partial h1} & \cdots & \frac{\partial f_1}{\partial h_n} & \frac{\partial f_1}{\partial h_{n+1}} & \cdots & \frac{\partial f_1}{\partial h_{n+m}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_p}{\partial h_1} & \cdots & \frac{\partial f_p}{\partial h_n} & \frac{\partial f_p}{\partial h_{n+1}} & \cdots & \frac{\partial f_p}{\partial h_{n+m}} \end{pmatrix} \begin{pmatrix} \frac{\partial h_1}{\partial y_1} & \cdots & \frac{\partial h_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial y_1} & \cdots & \frac{\partial h_1}{\partial y_m} \\ \frac{\partial h_{n+1}}{\partial y_1} & \cdots & \frac{\partial h_{n+1}}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_{n+m}}{\partial y_1} & \cdots & \frac{\partial h_{n+m}}{\partial y_m} \end{pmatrix} \quad = C
\end{aligned}
$$

其中矩阵 C 的分量

$$
C_{ij} = \sum_{k=1}^{n+m} \frac{\partial f_i}{\partial h_k}\frac{\partial h_k}{\partial x_j}, i = 1, \cdots, p, j = 1, \cdots, m
$$

注意到

$$
\frac{\partial f_i}{\partial h_k}\frac{\partial h_k}{\partial x_j} = \begin{cases} \frac{\partial f_i}{\partial g_k}\frac{\partial g_k}{x_j}, & k = 1, \cdots, n \\ \frac{\partial f_i}{\partial x_{k-n}}\delta_{k-n,j}, & k = n+1, \cdots, n+m \end{cases}, i = 1, \cdots, p, j = 1, \cdots, m
$$

故

$$
\begin{aligned}
C_{ij} &= \sum_{k=1}^{n} \frac{\partial f_i}{\partial g_k}\frac{\partial g_k}{\partial x_j} + \sum_{j=1}^{m} \frac{\partial f_i}{\partial x_j} \\
\Leftrightarrow \mathrm{d}_{\mathbf{x}}\mathbf{f}\left(\mathbf{g}\left(\mathbf{x}\right),\mathbf{x}\right) &= \left.\frac{\partial \mathbf{f}\left(\mathbf{y},\mathbf{x}\right)}{\partial \mathbf{y}}\right|_{\mathbf{y}=\mathbf{g}(\mathbf{x})} \frac{\partial \mathbf{g}\left(\mathbf{x}\right)}{\partial \mathbf{x}} + \frac{\partial \mathbf{f}\left(\mathbf{y},\mathbf{x}\right)}{\partial \mathbf{x}}
\end{aligned}
$$

定理 4.13 (隐函数定理)**.** 设 $\mathbf{F}: \mathbb{R}^{n+m} \to \mathbb{R}^m$ 是连续可导函数，且对 $\mathbf{x}_0 \in \mathbf{R}^n$ 和 $\mathbf{y}_0 \in \mathbf{R}^m$ 有

- $\mathbf{F}\left(\mathbf{x}_0, \mathbf{y}_0\right) = \mathbf{0}$;
- 导数 $\left.\frac{\partial \mathbf{F}(\mathbf{x},\mathbf{y})}{\partial \mathbf{y}}\right|_{\mathbf{x}=\mathbf{x}_0,\mathbf{y}=\mathbf{y}_0}$ 可逆;

则 $\mathbf{x}_0$ 的某邻域 N 存在由 $\mathbf{F}$ 隐函定义的连续可微函数 $\mathbf{f}: \mathbb{R}^n \to \mathbb{R}^m$ 满足 $\mathbf{F}\left(\mathbf{x}, \mathbf{f}\left(\mathbf{x}\right)\right)) = \mathbf{0}, \forall \mathbf{x} \in N$，且在 N 上有

$$
\mathrm{d}_{\mathbf{x}}\mathbf{f}\left(\mathbf{x}\right) = -\left[\left.\frac{\partial \mathbf{F}\left(\mathbf{x},\mathbf{y}\right)}{\partial \mathbf{y}}\right|_{\mathbf{x}=\mathbf{x},\mathbf{y}=\mathbf{f}(\mathbf{x})}\right]^{-1} \left.\frac{\partial \mathbf{F}\left(\mathbf{x},\mathbf{y}\right)}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x},\mathbf{y}=\mathbf{f}(\mathbf{x})}
$$

上式中的“-1”是指线性变换的逆。

证明. 待补充[10]p. 593。 □

例 4.10. 设函数 $\mathbf{F}:\mathbb{R}^4\to\mathbb{R}^2$，$\mathbf{F}(u,v,x,y)=(F_1,F_2)$，$\mathbf{F}=\mathbf{0}$ 隐含定义了 $(x,y)=\mathbf{f}(u,v)\,,\mathbf{f}:\mathbb{R}^2\to\mathbb{R}^2$，则

$$
\begin{aligned}
&\frac{\partial\mathbf{F}}{\partial u}=\mathbf{0},\quad \frac{\partial\mathbf{F}}{\partial v}=\mathbf{0}\\
\Leftrightarrow&\begin{cases}\frac{\partial F_1}{\partial u}+\frac{\partial F_1}{\partial x}\frac{\partial x}{\partial u}+\frac{\partial F_1}{\partial y}\frac{\partial y}{\partial u}=0\\ \frac{\partial F_2}{\partial u}+\frac{\partial F_2}{\partial x}\frac{\partial x}{\partial u}+\frac{\partial F_2}{\partial y}\frac{\partial y}{\partial u}=0\end{cases},\quad\begin{cases}\frac{\partial F_1}{\partial v}+\frac{\partial F_1}{\partial x}\frac{\partial x}{\partial v}+\frac{\partial F_1}{\partial y}\frac{\partial y}{\partial v}=0\\ \frac{\partial F_2}{\partial v}+\frac{\partial F_2}{\partial x}\frac{\partial x}{\partial v}+\frac{\partial F_2}{\partial y}\frac{\partial y}{\partial v}=0\end{cases}\\
\Leftrightarrow&\begin{pmatrix}\frac{\partial F_1}{\partial u}&\frac{\partial F_1}{\partial v}\\ \frac{\partial F_2}{\partial u}&\frac{\partial F_2}{\partial v}\end{pmatrix}+\begin{pmatrix}\frac{\partial F_1}{\partial x}&\frac{\partial F_1}{\partial y}\\ \frac{\partial F_2}{\partial x}&\frac{\partial F_2}{\partial y}\end{pmatrix}\begin{pmatrix}\frac{\partial x}{\partial u}&\frac{\partial x}{\partial v}\\ \frac{\partial y}{\partial u}&\frac{\partial y}{\partial v}\end{pmatrix}=0\\
\Leftrightarrow&\begin{pmatrix}\frac{\partial x}{\partial u}&\frac{\partial x}{\partial v}\\ \frac{\partial y}{\partial u}&\frac{\partial y}{\partial v}\end{pmatrix}=-\begin{pmatrix}\frac{\partial F_1}{\partial x}&\frac{\partial F_1}{\partial y}\\ \frac{\partial F_2}{\partial x}&\frac{\partial F_2}{\partial y}\end{pmatrix}^{-1}\begin{pmatrix}\frac{\partial F_1}{\partial u}&\frac{\partial F_1}{\partial v}\\ \frac{\partial F_2}{\partial u}&\frac{\partial F_2}{\partial v}\end{pmatrix}
\end{aligned}
$$

4.4 曲线、曲面和积分定理

本节的内容在大一的高等数学课中已经学过[9]§9，以下只是使用向量函数微积分的语言重新复述一次。本节默认讨论前题是在 3 维欧几里得空间中选取基本坐标系，使得任一点的位置向量直接可用平移空间中标准基下的坐标表示，从而与 $\mathbb{R}^3$ 的元素一一对应。

4.4.1 曲线积分

设函数 $\mathbf{g}:\mathbb{R}\supset I\to\mathbb{R}^3$ 在连通区域 I 上分段连续可微，则以 $\mathbf{g}(t)\,,t\in I$ 为参数方程的像集 $\mathbf{g}(I)$ 是 $\mathbb{R}^3$ 中的一条可求弧长（*rectifiable*）的曲线，可记为曲线 $\mathscr{C}$。

导数 $\left.\frac{d\mathbf{g}(t)}{dt}\right|_{t=t_0}$ 是曲线 $\mathscr{C}$ 在点 $\mathbf{g}(t_0)$ 处的切向量。可定义 $\widehat{\mathbf{t}}\equiv\frac{d\mathbf{g}/dt}{\|d\mathbf{g}/dt\|}$ 为曲线 $\mathscr{C}$ 的单位切向量（*unit tangent vector*）。

曲线有两种弧微元：$d\mathbf{l}\equiv(d\mathbf{g}/dt)\,dt$ 是曲线的弧向量微元；$dl\equiv\|d\mathbf{g}/dt\|\,dt$ 是弧长微元。它们的关系是 $d\mathbf{l}=\widehat{\mathbf{t}}dl$。

曲线的长度

$$L(\mathscr{C})=\int_{\mathscr{C}}dl=\int_a^b\|d\mathbf{g}/dt\|\,dt$$

如果定义在曲线 $\mathscr{C}$ 上的函数 $\mathbf{f}:\mathbb{R}\supset\mathscr{L}\to\mathbb{R}^n$ 是“单位弧长的性质”，则该性质在曲线 $\mathscr{C}$ 上的总和是对弧长的曲线积分[9]§9.1, 定理 9.1.1

$$\int_{\mathscr{C}}\mathbf{f}(\mathbf{g})\,dl=\int_a^b\mathbf{f}(\mathbf{g}(t))\,\|d\mathbf{g}/dt\|\,dt$$

例如，曲线 $\mathscr{C}$ 的线密度是函数 $\rho(\mathbf{g})$，则曲线 $\mathscr{C}$ 的总质量

$$m(\mathscr{C})=\int_{\mathscr{C}}\rho dl=\int_a^b\rho(\mathbf{g}(t))\,\|d\mathbf{g}/dt\|\,dt$$

如果定义在曲线 $\mathscr{C}$ 上的函数 $\mathbf{h}:\mathbb{R}^3\supset\mathscr{L}\to\mathbb{R}^3$ 是“作用在弧上的量”，则曲线 $\mathscr{C}$ 所受的总作用是对坐标的曲线积分[9]p. 140, 定理 9.2.1

$$\int_{\mathscr{C}}\mathbf{h}(\mathbf{g})\cdot d\mathbf{l}=\int_a^b\mathbf{h}(\mathbf{g}(t))\cdot\frac{d\mathbf{g}}{dt}dt=\int_{\mathscr{C}}\mathbf{h}\cdot\widehat{\mathbf{t}}dl$$

可见，曲线上的函数对坐标的曲线积分是这个函数与切向量点乘后对弧长的曲线积分。换句话说，作用 $\mathbf{h}$ 在曲线 $\mathscr{C}$ 上的总量是其在曲线每点的切方向上的投影分量的总和。例如，曲线 $\mathscr{C}$ 是一个质点的运动轨迹，力场 $\mathbf{F}(\mathbf{r},t)$ 对该质点做的总功

$$W(\mathscr{C})=\int_{\mathscr{C}}\mathbf{F}\cdot d\mathbf{l}=\int_a^b\mathbf{F}(\mathbf{g}(t),t)\cdot\frac{d\mathbf{g}}{dt}dt$$

4.4.2　曲面积分

设函数 $\mathbf{g}:\mathbb{R}^2\supset D\to\mathbb{R}^3$ 在由分段连续边界包围的连通区域 D 上连续可微，则以 $\mathbf{g}(\mathbf{u}),\mathbf{u}\in D$ 为参数方程的像集 $\mathbf{g}(D)$ 是 $\mathbb{R}^3$ 中的一个可求面积的曲面，记为曲面 $\mathscr{S}$。

当且仅当导数 $\mathrm{d}_{\mathbf{u}=\mathbf{u}_0}\mathbf{g}(\mathbf{u})$ 存在且满秩时，曲面 $\mathscr{S}$ 在点 $\mathbf{g}(\mathbf{u}_0)$ 处有切平面，或称曲面在此处是光滑的（*smooth*）。此时

$$\left.\left(\frac{\partial\mathbf{g}}{\partial u_1}\times\frac{\partial\mathbf{g}}{\partial u_2}\right)\right|_{\mathbf{u}=\mathbf{u}_0}$$

是曲面 $\mathscr{S}$ 在点 $\mathbf{g}(\mathbf{u}_0)$ 处的法向量，

$$\widehat{\mathbf{n}}\equiv\left.\left(\frac{\frac{\partial\mathbf{g}}{\partial u_1}\times\frac{\partial\mathbf{g}}{\partial u_2}}{\left\|\frac{\partial\mathbf{g}}{\partial u_1}\times\frac{\partial\mathbf{g}}{\partial u_2}\right\|}\right)\right|_{\mathbf{u}=\mathbf{u}_0}$$

是曲面 $\mathscr{S}$ 在点 $\mathbf{g}(\mathbf{u}_0)$ 处的单位法向量。

曲面有两种面微元：$d\boldsymbol{\sigma}=\left(\frac{\partial\mathbf{g}}{\partial u_1}\times\frac{\partial\mathbf{g}}{\partial u_2}\right)d\sigma_D$ 是有向曲面微元；$d\sigma=\left\|\frac{\partial\mathbf{g}}{\partial u_1}\times\frac{\partial\mathbf{g}}{\partial u_2}\right\|d\sigma_D$ 是面积微元。它们的关系是 $d\boldsymbol{\sigma}=\widehat{\mathbf{n}}d\sigma$。其中 $d\sigma_D=du_1du_2$ 是参数域 D 上的二重积分微元。

曲面 $\mathscr{S}$ 的面积

$$A(\mathscr{S})=\int_{\mathscr{S}}d\sigma=\int_D\left\|\frac{\partial\mathbf{g}}{\partial u_1}\times\frac{\partial\mathbf{g}}{\partial u_2}\right\|d\sigma_D$$

“单位面积的性质”$\mathbf{f}:\mathbb{R}^3\supset\mathscr{S}\to\mathbb{R}^n$ 在曲面 $\mathscr{S}$ 上的总和是第一型曲面积分[9]p. 165, 定理 9.4.1

$$\int_{\mathscr{S}}\mathbf{f}(\mathbf{g})d\sigma=\int_D\mathbf{f}(\mathbf{g}(\mathbf{u}))\left\|\frac{\partial\mathbf{g}}{\partial u_1}\times\frac{\partial\mathbf{g}}{\partial u_2}\right\|d\sigma_D$$

“作用场”$\mathbf{h}:\mathbb{R}^3\supset\mathscr{S}\to\mathbb{R}^3$ 在曲面上的总作用是第二型曲面积分或对坐标的曲面积分[9]§9.5

$$\int_{\mathscr{S}}\mathbf{h}\cdot d\boldsymbol{\sigma}=\int_D\mathbf{h}(\mathbf{g}(\mathbf{u}))d\sigma_D=\int_{\mathscr{S}}\mathbf{h}\cdot\widehat{\mathbf{n}}d\sigma$$

4.4.3 积分换元公式

本科高等数学已接触过 3 重积分的情况[9]§8.3 "五"。

设 $\mathscr{U} \subset \mathbb{R}^3$ 是由分段光滑边界包围的连通区域，则以参数方程 $\mathbf{g} : \mathbb{R}^3 \supset \mathscr{U} \to \Omega \subset \mathbb{R}^3$ 规定的像集 $\Omega = \mathbf{g}(\mathscr{U})$ 是一个经过形变后的三维区域。当且仅当导数 $\mathrm{d}_{\mathbf{r}=\mathbf{r}_0}\mathbf{g}(\mathbf{r})$ 存在且满秩时，区域 Ω 在点 $\mathbf{r}_0$ 处是光滑的。此时

$$\left|\frac{\partial \mathbf{g}}{\partial r_3} \cdot \left(\frac{\partial \mathbf{g}}{\partial r_1} \times \frac{\partial \mathbf{g}}{\partial r_2}\right)\right|_{\mathbf{r}=\mathbf{r}_0} \equiv \left|\det\left(\mathrm{d}_{\mathbf{r}=\mathbf{r}_0}\mathbf{g}(\mathbf{r})\right)\right|$$

是由向量 $\frac{\partial \mathbf{g}}{\partial r_1}, \frac{\partial \mathbf{g}}{\partial r_2}, \frac{\partial \mathbf{g}}{\partial r_3}$ 所搭成的平行六面体的体积。$\mathscr{U}$ 的体积元 $dV_{\mathscr{U}}$ 与 Ω 的体积元 dV_Ω 之间的转换关系是

$$dV_\Omega = \left|\det\left(\mathrm{d}_{\mathbf{r}}\mathbf{g}(\mathbf{r})\right)\right| dV_{\mathscr{U}}$$

区域 Ω 的体积是

$$V(\Omega) = \int_\Omega dV_\Omega = \int_{\mathscr{U}} \left|\det\left(\mathrm{d}_{\mathbf{r}}\mathbf{g}(\mathbf{r})\right)\right| dV_{\mathscr{U}}$$

"单位体积的性质" $\mathbf{f} : \mathbb{R}^3 \supset \mathscr{U} \to \mathbb{R}^3$ 在区域 Ω 的总和为

$$\mathbf{F}(\Omega) = \int_\Omega \mathbf{f}(\mathbf{g})\, dV_\Omega = \int_{\mathscr{U}} \mathbf{f}(\mathbf{g}(\mathbf{r})) \left|\det\left(\mathrm{d}_{\mathbf{r}}\mathbf{g}(\mathbf{r})\right)\right| dV_{\mathscr{U}}$$

上式其实就是积分换元公式。

4.4.4 积分定理

给定函数 $P, Q : \mathbb{R}^2 \supset D \to \mathbb{R}$ 和边界分段光滑的单连通区域 D，格林公式：

$$\int_D \left(\frac{\partial Q}{\partial x_1} - \frac{\partial P}{\partial x_2}\right) d\sigma_D = \int_{\partial D} P dx_1 + Q dx_2$$

可改写成

$$\int_D \mathrm{curl}\mathbf{F} d\sigma_D = \int_{\partial D} \mathbf{F} \cdot d\mathbf{l}$$

其中函数 $\mathbf{F} : \mathbb{R}^2 \supset D \to \mathbb{R}^2$ 在标准基下的坐标函数是 $\mathbf{F}(\mathbf{x}) = (P(\mathbf{x}), Q(\mathbf{x}))$。一般地，在标准基下函数 $\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), F_2(\mathbf{x}))$ 的旋度为

$$\mathrm{curl}\mathbf{F} = \frac{\partial F_2(\mathbf{x})}{\partial x_1} - \frac{\partial F_1(\mathbf{x})}{\partial x_2}$$

其中 $\mathbf{x} = (x_1, x_2)$。

由格林公式又有，

$$\int_D \left(\frac{\partial F_1}{\partial x_1} + \frac{\partial F_2}{\partial x_2}\right) d\sigma_D = \int_{\partial D} -F_2 dx_1 + F_1 dx_2$$

该式可改写成

$$\int_D \mathrm{div}\mathbf{F} d\sigma_D = \int_{\partial D} \mathbf{F} \cdot d\boldsymbol{\sigma}$$

一般地，在标准基下函数 $\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), F_2(\mathbf{x}))$ 函数 $\mathbf{F}$ 的散度为

$$\mathrm{div}\mathbf{F} = \frac{\partial F_1(\mathbf{x})}{\partial x_1} + \frac{\partial F_2(\mathbf{x})}{\partial x_2}$$

其中 $\mathbf{x} = (x_1, x_2)$。

以上 2 维空间的例子可推广到 3 维。在标准基下函数 $\mathbf{F} : \mathbb{R}^3 \supset \mathscr{U} \to \mathbb{R}^3$ 的旋度

$$\mathrm{curl}\mathbf{F} = \left(\frac{\partial F_3}{\partial x_2} - \frac{\partial F_2}{\partial x_3}, \frac{\partial F_1}{\partial x_3} - \frac{\partial F_3}{\partial x_1}, \frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2}\right)^{\mathsf{T}}$$

若函数 $\mathbf{F}$ 在 D 上连纪可导且 $\mathscr{U}$ 是开集则 $\mathscr{U}$ 也是旋度函数 $\mathrm{curl}\mathbf{F}(\mathbf{x})$ 的定义域。

设 $\mathscr{S}$ 是 $\mathbb{R}^3$ 中的光滑曲面，$\mathbf{g} : \mathbb{R}^2 \supset D \to \mathbb{R}^3$ 二阶连续可微，D 是由分段光滑边界围成的单连通区域，$\mathbf{F} : \mathbb{R}^3 \supset \mathscr{S} \to \mathbb{R}^3$ 是作用在曲面 $\mathscr{S}$ 上的连续可微向量场，则有斯托克斯定理

$$\int_{\mathscr{S}} \mathrm{curl}\mathbf{F} \cdot d\boldsymbol{\sigma} = \int_{\partial \mathscr{S}} \mathbf{F} \cdot d\mathbf{g}$$

在标准基下函数 $\mathbf{F} : \mathbb{R}^3 \supset \mathscr{U} \to \mathbb{R}^3$ 的散度

$$\mathrm{div}\mathbf{F} = \frac{\partial F_1}{\partial x_1} + \frac{\partial F_2}{\partial x_2} + \frac{\partial F_3}{\partial x_3}$$

若函数 $\mathbf{F}$ 在 D 上连纪可导且 $\mathscr{U}$ 是开集则 $\mathscr{U}$ 也是散度函数 $\mathrm{curl}\mathbf{F}(\mathbf{x})$ 的定义域。

设 Ω 是 $\mathbb{R}^3$ 中可数个简单区域的并集，由分段光滑边界围成。$\mathbf{F} : \mathbb{R}^3 \supset \Omega \to \mathbb{R}^3$ 是存在于 Ω 中的连续可微向量场,，则有高斯定理

$$\int_{\Omega} \mathrm{div}\mathbf{F} dV_{\Omega} = \int_{\partial \Omega} \mathbf{F} \cdot d\boldsymbol{\sigma}$$

第二部分

连续介质力学基础

第五章　时空观与参考系

在数学或物理学中，*不变性*（*invariance*）常常用于描述一个数学或物理对象的性质在某种变换操作前后的结果不变。这是一个比较笼统地一般叙述。每个具体的“不变性”概念都需要定义清楚，具体是何种“变换操作”，以及何谓“性质的变换结果不变”。

在物理学中，任意两个不同观察者对同一现象的观测记录，在了解了这两个观察者的相对状态后，就能够互相换算。这种从一个观察者的观测记录到另一个观察者的观测记录的换算操作就是一种变换操作。*物理客观性*（*physical objectivity*）要求，一切关于客观规律的陈述必须具有不依赖观察者的不变性。

在电磁学发展之前，物理学主要关心由宏观物体的运动现象总结出来的运动定律。它的完善版本是今天所说的经典力学。它已具有*伽俐略变换*（*Galilean transformation*）下的不变性，或称*伽俐略相对性*（*Galilean relativity*）。它说，物体运动定律在惯性系间（即相对运动是匀速直线*运动的观察者之间）的形式具有不变性。在非惯性系下，要保持物体运动定律在惯性系中的形式不变，需要引入*惯性力*（*intertial force*）。

在电磁学理论完善后，人们发现其不具有伽俐略变换下的不变性，但能在考虑了光速的修正——*洛伦兹变换*（*Lorentz transformation*）下保持惯性系中的形式不变性。爱因斯坦在进一步明确了所有物理定律在惯性系下的形式具有不变性的原则后，建立了*狭义相对论*（*spectial theory of relativity*）力学。在此基础上，他又通过提出物质和能量以某种方式使时空弯曲的假设，替代牛顿的万有引力理论，解决了后者在狭义相对论中的矛盾，建立了*广义相对论*（*general theory of relativity*）†。

参考标架（*frame of reference*）是任何时空理论的核心内容。它代表着不同观察者观测同一物理事件时，在“何谓静止”标准上的主观差异。不同的观察者所选择的参考标架可作为他们各自用于描述运动的“绝对时空”。如果物理事件是客观的，那么不同观察者对同一物理事件的观察结果的差别，应仅来自这些观察者之间的相对运动。因此，各类客观性概念都依赖参考标架的概念得以陈述。

* “直线”的概念依赖已有的欧几里得空间，后者隐含在了*伽俐略时空*（*Galilean spacetime*）的构建当中[11]。

†这里的相关知识应该已经在大学物理中介绍过了[12]§1,§2.5[13]§24。

在经典连续介质力学中，**物质客观性**（*material frame indifference*）一度是本构关系理论的基本原则之一[14]Sect.293[15]Sect.19。这一原则被用于约束材料本构关系的数学形式，得出了十分重要的流变学理论*。但是关于物质客观性原理的的具体含义、以及（无论在哪一种具体含义下）它是否必须作为本构关系的基本原则的讨论一直持续到今天†。不管是为了理解主流连续介质力学资料，还是为了跟上这一话题的最新讨论，都需要清晰而仔细地明确时空和参考标架的构建过程。

经典力学的时空观构建方式有很多版本[11]。本讲义无意陷入到时空观的历史回顾中，仅选择 W. Noll[20]构建的**新经典时空**（*neo-classical spacetime*），原因包括但不限于：大部分连续介质力学教科书默认或明确介绍的就是这一时空观。

5.1 新经典时空

经典力学中的观察者对时间的感受依赖他所观察到的物理事件发生的先后顺序。以下我们通过物理事件的概念引出时间的概念，请结合图5.1进行理解。

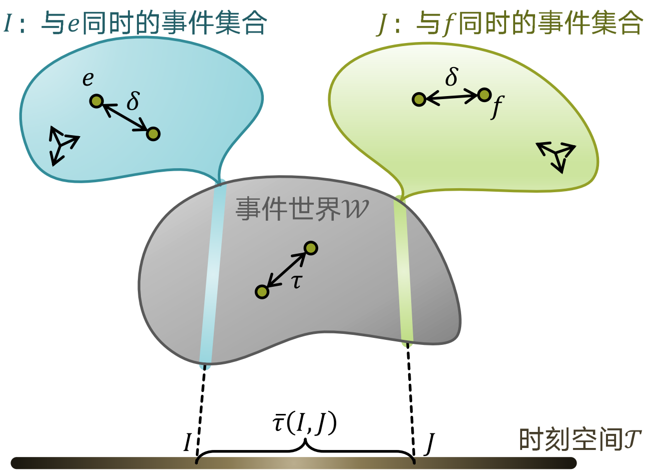


图 5.1: 事件世界概念示意图。图中字母符号与文中相同。

事件世界（*event-world*）$\mathcal{W}$ 是一个非空集合，其元素 $e\in\mathcal{W}$ 称为一个**事件**（*event*）。$\mathcal{W}$ 上还定义一个**时延函数** $\tau:\mathcal{W}^2\to\mathbb{R}$，满足：

T1 $\forall e,f\in\mathcal{W},\quad \tau\left(e,f\right)=-\tau\left(f,e\right)$

T2 $\forall e,f,g\in\mathcal{W},\quad \tau\left(e,f\right)+\tau\left(f,g\right)=\tau\left(e,g\right)$

T3 $\forall e\in\mathcal{W}\forall t\in\mathbb{R}\exists f\in\mathcal{W},\quad \tau\left(e,f\right)=t$

*例如简单流体（*simple fluids*）理论[16]。由其导出的测粘流（*viscometric flow*）理论是流变学测量得以实现的理论基础[2][17]。

†[18][19]

所谓“事件”或“物理事件”是物理观测经验事实的“最小单元”。一个事件需且仅需用时间中的一个时刻和空间中的一个点来表示，在经典条件下可理解为“某物质点在某时刻出现在空间某处”。时延函数的定义规定了任意两个事件的时间间隔取值是一个实数。实数域 $\mathbb{R}$ 具有完备性（*completeness*），即在任意两个不相等的实数间必有一个实数。把时延函数的值域定为实数，使得我们可以考虑“随时间连续发生的一系列事件”，比如一个质点的运动轨迹。在时延函数的 3 个规定中，T1 和 T2 的规定是符合直觉的，而 T3 的意思是，事件世界的元素在时延函数作用下具有封闭性。即无论从哪个事件开始算起，任意时延后都允许有一个事件的发生。这相当于说，事件世界包括了所有可发生的事件。

我们还能说：

- 事件 a 早于事件 b，当且仅当 $\tau(a,b)>0$；
- 事件 a 晚于事件 b，当且仅当 $\tau(a,b)<0$；
- 事件 a 与 b 同时（*simultaneous*）发生，当且仅当 $\tau(a,b)=0$；

事件的同时性（*simultaneity*）形成了一个等价关系

$$S\equiv\left\{(a,b)\in\mathcal{W}^2|\tau(a,b)=0\right\}$$

我们称 S 是物理事件的同时关系。由等价关系的基本定理1.1，通过 S 可对事件世界 $\mathcal{W}$ 形成划分 $\mathcal{T}\equiv\mathcal{W}\setminus S$，使得对任一 $e\in\mathcal{W}$，$I\equiv[\![e]\!]_S\in\mathcal{T}$ 是事件 e 关于同时关系 S 的等价类。我们称 $\mathcal{T}$ 为时刻空间（*instant space*）。时刻空间的元素 $I,J,\cdots\in\mathcal{T}$ 称为时刻（*instants*），它们是不同时刻下同时发生的所有事件的集合。

设 $I,J\in\mathcal{T}$ 是两个时刻，我们可以通过 I 和 J 的任一元素 $e\in I$ 和 $f\in J$ 来表示这两个时刻的间隔。具体地，定义在 $\mathcal{T}$ 上的函数 $\overline{\tau}:\mathcal{T}^2\to\mathbb{R}$，

$$\overline{\tau}(I,J)=\tau(e,f),\quad\forall e\in I\forall f\in J$$

的结果就是时刻 I、J 的时间间隔。我们可以设计符合我们以往习惯的记法：$\forall I,J\in\mathcal{T}$,

$$I-J\equiv\overline{\tau}(I,J)$$

若 $s=\overline{\tau}(I,J)$，则又可记 $I=J+s$。注意，我们没有定义两个时刻的“和”（$I+J$）的意义。事实上，时刻空间 $\mathcal{T}$ 与实数集 $\mathbb{R}$ 是自然同构的。因此在进行时刻之间的运算时，我们可以默认 $\mathcal{T}$ 就像 $\mathbb{R}$。但是每一 $\mathcal{T}$ 中的时刻又是同时事件构成的欧几里得空间，这时 $\mathcal{T}$ 比实数集 $\mathbb{R}$ 具有更多含义，故我们仍保持提及 $\mathcal{T}$。

接下来，我们定义两个同时发生的事件之间的距离。距离函数是定义在 S 上的映射 $\delta:S\to\mathbb{R}$，满足：

D1 对任一时刻 $I \in \mathcal{T}$，δ 在 $I^2 \subset S$ 上的限制 $\delta|_{I^2}$ 使作为事件集合的 I 成为一个以事件为点的欧几里得空间，其平移空间记为 $\mathscr{V}_I$；

D2 $\dim \mathscr{V}_I = 3, \quad \forall I \in \mathcal{T}$。

距离函数 δ 定度在同时关系 S 上，说明只有同时发生的事件之间才能讨论距离。S 中含有不同时刻下的同时事件，所以在不同时刻下，我们都用同一度量 δ 来测定距离。规定 D1 用通俗语言说就是每一时刻 I 作为同时事件的集合，在用了度量 δ 之后，能够形成欧几里得空间，即能够满足定义3.4。结合规定 D2，每一时刻下的欧几里得空间都是 3 维的。

规定 T1~T3、D1~D2 共同形成了新经典时空的概念。正式陈述如下——

定义 5.1 (新经典时空). 若集合 $\mathcal{W}$ 具有 T1~T3 规定的时延函数和 D1~D2 规定的距离函数，则称 $\mathcal{W}$ 是新经典时空（*neo-classical spacetime*）。

我们可以形象地把时延函数和距离函数理解为观察者所使用的时钟和直尺。在新经典时空的定义下，没有钟慢效应，因为“同时性”的定义是绝对的；也没有尺缩效应，因为距离函数的定义是绝对的。在这样的时空观下，本讲义介绍的连续介质力学基础适用的范围自然就是经典低速运动和形变。

事件世界 $\mathcal{W}$ 上的自同态映射 $\alpha : \mathcal{W} \to \mathcal{W}$ 应满足

$$\tau\left(\alpha\left(e\right), \alpha\left(f\right)\right) = \tau\left(e, f\right), \quad \forall e, f \in \mathcal{W}$$
$$\delta\left(\alpha\left(e\right), \alpha\left(f\right)\right) = \delta\left(e, f\right), \quad \forall e, f \in S$$

其中 S 是 $\mathcal{W}$ 上的同时关系。事实上映射 α 就是对其所作用的事件进行了一个时移。若记 U^α 为 $\mathcal{W}$ 上的自同态映射 α 在 $\mathcal{W}$ 的子集 U 上的限制的像集，即 $U^\alpha \equiv \operatorname{ran} \alpha|_U$，则易验

$$\overline{\tau}\left(I^\alpha, J^\alpha\right) = \overline{\tau}\left(I, J\right), \quad \forall I, J \in \mathcal{T}$$

或按之前规定的记法：

$$J^\alpha - I^\alpha = J - I, \quad \forall I, J \in \mathcal{T}$$

其中 $\mathcal{T}$ 是 $\mathcal{W}$ 上的时刻空间。上式等号两边同时加上 $I^\alpha - J$ 得

$$I^\alpha - I = J^\alpha - J, \quad \forall I, J \in \mathcal{T}$$

也就是说，映射 α 对不同时刻下的事件进行作用前后的时间间隔是只依赖 α 的固定值。因此可记由任一时刻 I 及其对应的 I^α 规定的实数 $s_\alpha \equiv I^\alpha - I, \forall I \in \mathcal{T}$ 为由 α 引出的时移（*time-shift*）。由规定 T3 可以证明，对应于每一个 $\mathcal{W}$ 上的自同态映射 α 有且只有一个实数 $s_\alpha \in \mathbb{R}$ 满足 $s_\alpha = I^\alpha - I$。留意到一个时刻 I 同时又是一个欧几里得空间，对任一时刻 I，α 是 I 到

I^α 关于共同的度量 δ 的保距映射，故上述论断亦表明 α 是一个等距变换（双射）。相应地也可证明，对应于每一 $\mathcal{W}$ 上的自同态映射 α 有且只有一个由 I 的平移空间 $\mathscr{V}_I$ 到 I^α 的平移空间 $\mathscr{V}_{I^\alpha}$ 的同态映射（即线性变换）$\mathbf{A}_I : \mathscr{V}_I \to \mathscr{V}_{I+s_\alpha}$ 满足 $\mathbf{A}_I\left(f-e\right) = \alpha\left(f\right) - \alpha\left(e\right), \forall e, f \in I$。

5.2 参考标架

事件世界含有的是一切可以发生的事件，它只是一个供我们讨论某一具体发生着的物理过程的背景概念。若某物体在某段时间之内发生了运动和形变，它将对应于事件世界里的一块子集。具体地，我们把一个*物体*（*body*）抽象为一个集合 B，其元素 $X, Y, \cdots \in B$ 称为*物质点*（*material points*）。在这里我们只讨论物体的*运动学*（*kinematics*），后续章节会再赋予物体以质量的概念。

现在我们准确地定义什么叫“在一段时间之内”的运动。我们把 $\mathcal{T}$ 的一个连通子集 $\Upsilon \subset \mathcal{T}$ 称作一个时间段。所谓“连通”是指，若 $t \in \Upsilon$ 且 $t+s \in \Upsilon$，则 $t + r \in \Upsilon, \forall r \in [0, t]$*。

以下内容可结合图5.2来理解。

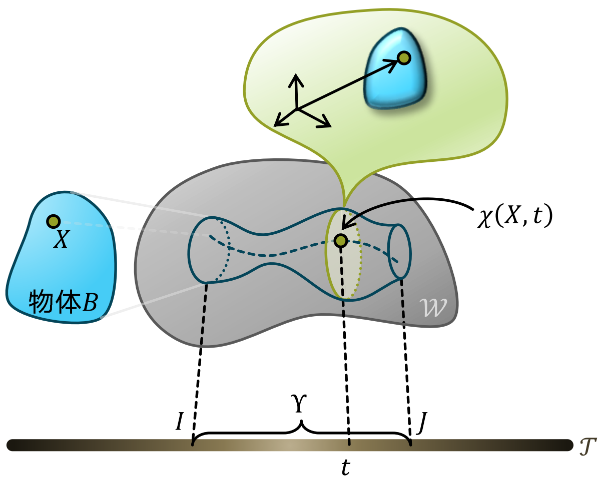


图 5.2: 物体的运动过程、时间线示意图。图中字母符号与文中相同。

一个物体 B 在时间段 Υ 内的*运动过程*（*kinematical process*）是一个映射 $\chi : B \times \Upsilon \to \mathcal{W}$,

$$\chi\left(X, t\right) \in t, \quad \forall X \in B, \quad \forall t \in \Upsilon$$

再次提醒，上式中的 t 既表示一个时刻又表示这个时刻下的所有同时事件构成的欧几里得空间点集。上述定义实际表达的意思是，想要描述一个由多个物质点构成的物体的运动过程，需

*这里我们改用小写字母来表示时刻以符合时刻计算的习惯，但需要注意每一时刻都是一个由同时事件组成的欧几里得空间。

且只需说出每个物质点在每个时刻下对应于哪一个物理事件。我们把一个物质点 $X \in B$ 在运动过程 χ 中的像集

$$\chi(X, \Upsilon) = \{e \in \mathcal{W} | \exists X \in B \forall t \in \Upsilon, e = \chi(X, t)\}$$

（即一个物质点在整个时间段 Υ 内经历的所有事件的集合）称 X 在运动过程 χ 中的世界线（*worldline*）。

可以留意到，一个物质点在一个运动过程中的世界线上的任意两个不同元素必属于不同时刻。诚然，由运动过程的定义可知，若 $t_1, t_2 \in \Upsilon$ 且 $t_1 \neq t_2$，则物质点 X 在时间段 Υ 内的运动过程 χ 中的世界线总满足

$$\chi(X, t_1) \in t_1, \quad \chi(X, t_2) \in t_2$$

由集合相等的定义，$t_1 \neq t_2$ 决定了 $\chi(X, t_1) \neq \chi(X, t_2)$。换言之，由 Υ 到 X 在 χ 中的世界线的映射 $t \mapsto \chi(X, t)$ 必是单射。

更形象地说，世界线是一条由参数方程定义的曲线*。一整个物体 B 的运动过程，是 B 的所有物质点形成的一束世界线。一个常常默认的公设是**经典时空下的信息守恒，即同一运动过程中的世界线不相交**。这等同于说，一个物体在运动过程中物质点的“数量”守恒†。这是因果决定论所要求的。任一当前时刻的发生的事件必由之前时刻的事件导致（没有“无因之果”），也必导致未来时刻的事件（没有“无果之因”）。

虽然世界线是一条曲线，但它无法直接理解为“质点的轨迹”。因为世界线上不同的事件属于不同时刻，亦属于不同的欧几里得空间的点，无法量度它们之间的距离（只有同时刻的事件之间者能讨论距离）。我们在现实世界中之所以能观测一个质点在两个时刻下的“位移”，必须依赖另一个在这个时间段内也一直出现的另个物体——参照物。就算如此，我们也只能在每一时刻下记录所关心的质点与这一参照物的**相对距离**，得到这个相对距离在两个不同时刻的的差别，来描述这个质点在这两个时刻所规定的时间间隔的标量值的相对位移变化，并进而得到它们的标量相对速率和相对加速率。具体地，我们可以考虑导数

$$\frac{\mathrm{d}}{\mathrm{d}t}\delta(\chi(X, t), \chi(Y, t)) = \lim_{t' \to t} \frac{\delta(\chi(X, t'), \chi(Y, t')) - \delta(\chi(X, t), \chi(Y, t))}{t' - t}$$

它是物质点 X 相对于物质点 Y 在时刻 t 下的**相对速率**（标量）。类似地可定义 2 阶导数为一个标量值的相对加速率。本讲义所考虑的运动过程，任意两个物质点的相对加速率都是连续函数。也就是说，任一运动过程中两物质点的相对距离随时间的变化是 2 阶光滑的。

*严格地应该说世界线是一个 1 维流形。

†事实上连续介质物体的物质点数量可能是无穷大的，所以“数量”这一不严谨的说法仅为了辅助理解。

现在我们只讨论了两个物质点的相对运动速率和加速率。若要讨论某一个物质点的向量值的位移、速度和加速度等等，就必须在同一欧几里得空间，建立统一的坐标系。这需要引入参考标架的概念。参考标架是在一个观察者的主观视角下的“静止的”3 维欧几里得空间*。我们习惯说的某物质点的“绝对运动”，事实上是相对于我们视角下的 3 维欧几里得空间的相对运动。利用上一段介绍的相对运动概念，我们不仅可以把我们所关心的物质点与这个参考标架任一点（物理事件）的相对速率和相对加速率当作这个物质点在这一参考标架下的“绝对”速率和“绝对”加速率，还能利用这一欧几里得空间的平移空间，通过选定原点来实现对这一物质点的向量值位移、速度和加速度的表征。

下面我们严格地给出参考标架的概念。

设一物体 B 在某时间段 Υ 内的运动过程 χ 中，任意两物质点 $X,Y\in B$ 的距离

$$\delta\left(\chi\left(X,t\right),\chi\left(Y,t\right)\right)$$

保持不依赖时刻 $t\in\Upsilon$ 变化的常数，则称 χ 是一个刚体运动过程（*rigid body process*）。由相对速率的概念易知，刚体运动过程中任意两物质点相对速率总为零。

若某物体 $\mathscr{E}$ 永远只能作刚体运动†，即 $\mathscr{E}$ 在整个时刻空间 $\mathcal{T}$ 内的运动过程 $\phi:\mathscr{E}\times\mathcal{T}\to\mathcal{W}$ 总是一个刚体运动过程——

$$\delta\left(\phi\left(X,t\right),\phi\left(Y,t\right)\right)=d\left(X,Y\right),\quad\forall X,Y\in\mathscr{E},t\in\mathcal{T}$$

其中 $d:\mathscr{E}^2\to\mathbb{R}$ 是按照上式定义的 $\mathscr{E}$ 上的一个映射‡，则称 $\mathscr{E}$ 是一个刚体系（*rigid system*）。上式同时也说明映射 d 的性质沿自 δ。因 δ 是 $\mathcal{T}$ 的每一时刻下的欧几里得度量（见规定 D1、D2），故 d 是 $\mathscr{E}$ 上的欧几里得度量，$(\mathscr{E},d)$ 由此形成一个欧几里得空间。我们称具有欧几里得空间构造的刚体系 $\mathscr{E}$ 及其某一运动过程 ϕ 的组合 $(\mathscr{E},\phi)$ 为参考标架（*frame of reference*）或简称标架（*frame*）；$\mathscr{E}$ 是这一参考标架的欧几里得空间；ϕ 本身称作参考运动过程（*kinematic process of reference*）。由于可选作参考标架的刚体系 $\mathscr{E}$ 可有不止一个，刚体运动 ϕ 也可有不止一种，故参考标架的建立方式也不止一种。参考标架反映的是一名观察者的主观视角或主观选择。

一名观察者在选定了一个参考标架 $(\mathscr{E},\phi)$ 后，仍然面临着在其欧几里得空间 $\mathscr{E}$ 上建立不同的坐标系的选择。至少要在欧几里得空间 $\mathscr{E}$ 的基本直角坐标系（选定的某原点 $O\in\mathscr{E}$ 和标准基 $\{\widehat{\mathbf{e}}_1,\widehat{\mathbf{e}}_2,\widehat{\mathbf{e}}_3\}$）中，任一时刻 t 的下的事件才对应于一组有序实数三元组 $(x_1,x_2,x_3)\in\mathbb{R}^3$。当然，他也可以在 $\mathscr{E}$ 上建立各种曲线坐标系，同样也能实现任一时刻 t 下使一个事件对应于

*在经典情况下我们不考虑“空间弯曲”，因此每个观察者视角下的几何空间都是 3 维欧几里得空间。

†它将会是一个欧几里得空间。

‡我们很快将说明它其实是 $\mathscr{E}$ 上的一个度量。

一组有序实数三元组。在最一般情况下，仅为了这一目的，他完全可以在每一时刻都选一个不同的坐标系。在实际问题中我们选定的坐标系是不依赖时间变化的。严格来说，时刻空间 $\mathcal{T}$ 虽与 $\mathbb{R}$ 同构，但任一时刻 $t \in \mathcal{T}$ 到底对应于 $\mathbb{R}$ 的哪个数值，也仍需在选定某时刻空间的原点 $t_0 \in \mathcal{T}$ 并记其为实数 0 后，才能得到确定。我们常常选择 Υ 的边界点作为 t_0，使 Υ 同构于 $\mathbb{R}$ 上的某连通区间 $[0, s]$。

总而言之，想要使一个物理事件 $e \in \mathcal{W}$ 对应于获得 4 个实数 $(t; x_1, x_2, x_3) \in \mathbb{R}^{1+3}$ 的对应，必须具备以下条件：

- 事件 e 属于某世界线，也就是说属于某一时间段 $\Upsilon \subset \mathcal{T}$ 中的运动 χ；
- 已选定某参考标架 $(\mathscr{E}, \phi)$；
- 已选定起始时刻 $t_0 \in \mathcal{T}$；
- 已选定 $\mathscr{E}$ 下的坐标系

5.3 标架变换

接下来为了讨论的方便，我们对时间段 Υ 下的给定的运动过程 χ 规定记法 $\chi_t : B \to \mathcal{W}$,

$$\chi_t(X) = \chi(X, t), \quad \forall X \in B, t \in \Upsilon$$

设一名观察者观察一个物体 B 在时间段 Υ 内的运动过程 χ，这名观察者要确定的是物体 B 相对于他选择的参考标架 $(\mathscr{E}, \phi)$ 的运动。在同一时段 Υ 内，定义映射 $\kappa : B \times \Upsilon \to \mathscr{E}$,

$$\kappa(X, t) = \kappa_t(X) = \phi_t^{-1} \circ \chi_t(X), \forall X \in B, t \in \Upsilon$$

则 κ 在形式上定义了“物体 B 在物体 $\mathscr{E}$ 内的运动过程”，而事实上 $\mathscr{E}$ 是这名观察者选择的参考标架的欧几里得空间。我们称 κ 为物体 B 相对于参考标架 $(\mathscr{E}, \phi)$ 的由 χ 确定的运动，它由物体 B 的客观运动过程 χ 和参考标架本身的客观运动过程 ϕ 共同确定。任一观察者在不与其他观察者交流的情况下，只能感知相对运动 κ。

假设有两名观察者分别选择了标架 $(\mathscr{E}, \phi)$ 和 $(\mathscr{E}^*, \phi^*)$，d、d^* 是按前面所述由 ϕ、ϕ^* 定义的度量。这两名观察者观察同一物体 B 在时间段 Υ 内的一个运动过程 χ，将各自感知为 B 相对标架 $(\mathscr{E}, \phi)$、$(\mathscr{E}^*, \phi^*)$ 的运动过程 κ 和 κ^*（可结合图5.3来理解）。按照前面介绍过的定义，

$$\kappa_t = \phi_t^{-1} \circ \chi_t, \quad \kappa_t^* = \phi_t^{*-1} \circ \chi_t, \quad \forall t \in \Upsilon$$

故有

$$\kappa_t^* = \left(\phi_t^{*-1} \circ \phi_t\right) \circ \kappa_t, \quad \forall t \in \Upsilon$$

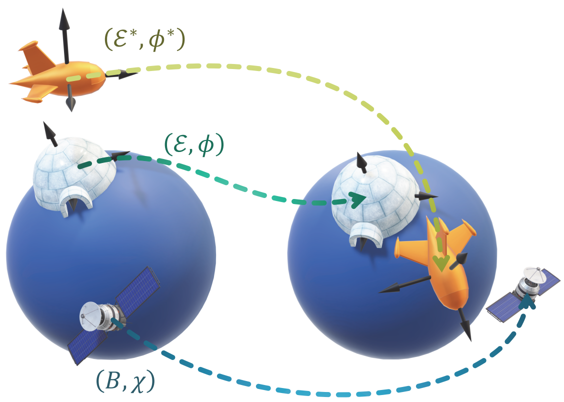


图 5.3: 不同观察者在各自的标架 $(\mathscr{E},\phi)$、$(\mathscr{E}^*,\phi^*)$ 下观察同一物体 B 的运动过程 χ。尽管刚体系 $\mathscr{E}$ 和 $\mathscr{E}^*$ 各自也有运动 ϕ、ϕ^*，但每个观察者都只能感知到 B 相对各自标架的运动 κ、κ^*（未标出）。观察者之间通过交流，也只能额外了解到两标架之间的相对运动 $\phi^{*-1}\circ\phi$（未标出）。

也就是说，两名观察者任一时刻所感知到的相对运动之间，只相差 $\phi_t^{*-1}\circ\phi_t$，而这是 $\mathscr{E}$ 作为一个刚体系相对于标架 $(\mathscr{E}^*,\phi^*)$ 的由 ϕ 确定的运动。两名观察者通过交流，也只能感受到这一标架间的相对运动，而无法单独感知到 ϕ 或 ϕ^*。这一标架间的相对运动 $\phi_t^{*-1}\circ\phi_t$ 是标架变换的讨论重点。

留意到，在时间段 Υ 内的每一时刻 t 下，(t,δ)、$(\mathscr{E},d)$、$(\mathscr{E}^*,d^*)$ 是三个欧几里得度量空间，ϕ_t、ϕ_t^* 分别是由 $\mathscr{E}$ 到 t、$\mathscr{E}^*$ 到 t 的等距变换，故复合映射 $\phi_t^{*-1}\circ\phi_t$ 是由 $\mathscr{E}$ 到 $\mathscr{E}^*$ 的、依赖时刻 t 的等距变换。在此我们可以利用等距变换的表示定理（定理3.2）写出复合映射 $\phi_t^{*-1}\circ\phi_t$ 的形式。具体地，物体 B 在时间段 Υ 内的运动过程 χ 中，时刻 $t\in\Upsilon$ 下，物质点 $P_X\in B$ 在时刻 t 下分别对应于欧几里得空间 $\mathscr{E}$、$\mathscr{E}^*$ 中的点 $X(t)$、$X^*(t)$，则有

$$X(t)=\kappa_t(P_X),\quad X^*(t)=\kappa_t^*(P_X),\quad t\in\Upsilon$$

其中 κ、κ^* 分别是物体 B 在时间段 Υ 内相对于标架 $(\mathscr{E},\phi)$、$(\mathscr{E}^*,\phi^*)$ 的运动。由之前介绍过的定义我们有

$$X^*(t)=\phi_t^{*-1}\circ\phi_t(X(t)),\quad \forall t\in\Upsilon$$

上式说明，两个观察者在各自的参考标架下，对同一物理事件的空间定位，相差一个等距变换。由定理3.2，再任意给定一物质点 $P_{X_0}\in B$，就有

$$X^*(t)=X_0^*(t)+\mathbf{Q}_t(X(t)-X_0(t))$$

其中

$$X_0(t)=\kappa_t(P_{X_0}),\quad X_0^*(t)=\kappa_t^*(P_{X_0}),\quad t\in\Upsilon$$

注意到，由于每一时刻 t 下的等距变换 $\phi_t^{*-1}\circ\phi_t$ 是不同的等距变换，因此上式中的正交算符 $\mathbf{Q}_t$ 依赖 t 而变化。另外，由 §3.4最后的讨论，$\det\mathbf{Q}_t\equiv 1$。

给定一个欧几里得空间，也就同时给定了它的平移空间和一个基本直角坐标系。因此每个欧几里得空间点集里的点，唯一对应于它在这一基本直角坐标系下的位置向量，也唯一对应于一个有序实数三元组。故上式的成立同时意味着下式的成立

$$\mathbf{r}_X^*(t)=\mathbf{r}_{X_0}^*(t)+\mathbf{Q}_t\left(\mathbf{r}_X(t)-\mathbf{r}_{X_0}(t)\right)$$

其中 $\mathbf{r}_{X_0}^*(t)=X_0^*(t)-O(t)\cdots\in\mathbb{R}^3$ 等都是相应的点在其所属的欧几里得空间中的基本直角坐标。

在时刻空间 $\mathcal{T}$ 中，只要选定一个时刻 t_0 作为时间原点（*time origin*），那么由之前关于时刻空间的讨论，任一时刻 t 都唯一对应 $\mathbb{R}$ 中的一个实数 $s=t-t_0$，任一实数 s 总有唯一一个时刻 t 满足 $s=t-t_0$。也就是说，只要选定了时间原点 t_0，那么时刻空间 $\mathcal{T}$ 的元素就与实数一一对应。假定两个观察者选定不同的时刻 t_0、$t_0^*\in\mathcal{T}$ 作为原点时刻，那么他们会相应使用不同的实数 s、$s^*\in\mathbb{R}$ 来标记同一时刻 $t\in\mathcal{T}$。正如 $\mathbf{r}_X^*$ 与 $\mathbf{r}_X$ 的关系那样，s 与 s^* 的换算关系也是被时空结构的定义确定的。为了得出这个关系，我们将一名观察者把一个时刻对应成一个实数的映射记为双射 $T:\mathcal{T}\to\mathbb{R}$。两个不同观察者分别采用映射 T、T^* 把同一时刻 $t\in\mathcal{T}$ 记作不同的实数。易验 T,T^* 满足

$$|T(t_1)-T(t_2)|=|T^*(t_1)-T^*(t_2)|=|t_1-t_2|,\quad\forall t_1,t_2\in\mathcal{T}$$

因此 T、T^* 都是度量空间 $(\mathcal{T},|\vec{\tau}|)$ 的上的等距变换。那么复合映射 $T^{*-1}\circ T$ 就是 $\mathbb{R}$ 上的一个等距变换。由等距变换的表示定理，对任意 $s\in\mathbb{R}$，

$$s^*=s_0^*+(s-s_0)$$

其中 $s_0^*=T^{*-1}\circ T(s_0)$；在 $\mathbb{R}$ 上，正交算符就是 ±1；行列式为 1 的正交算符就是 1。上式又可整理成

$$s^*=s+a$$

其中由 T、T^* 的等距变换事实可知 $a=s_0^*-s_0=t_0^*-t_0$ 是只依赖时间原点选择的常实数。。

把以上分析的结果进行总结：两名观察者 O、O^* 对同一物体的运动过程进行观测，在各自选定的时间原点 t_0、t_0^* 和参考标架 $(\mathscr{E},\phi)$、$(\mathscr{E}^*,\phi^*)$ 下，O 用于描述时刻 $t\in\mathcal{T}$ 下的两个事件 $e,e_0\in\mathcal{W}$ 的实数组 $(s,\mathbf{r}(s))$、$(s,\mathbf{r}_0(s))$ 和 O^* 用于描述同一时刻 t 的相同两个事件 e,e_0 的实数 $(s^*,\mathbf{r}^*(s^*))$、$(s^*,\mathbf{r}_0^*(s^*))$ 之间存在以下换算关系——

$$s^*=s+a$$

$$\mathbf{r}^*(s^*)=\mathbf{r}_0^*(s^*)+\mathbf{Q}_t\left(\mathbf{r}(s)-\mathbf{r}_0(s)\right)$$

其中 $a=t_0^*-t_0$ 是常实数，$\mathbf{Q}_t$ 是 $\mathbb{R}^3$ 上的正交变换（即一个 3×3 实正交矩阵）且 $\det\mathbf{Q}_t\equiv1$。我们称上述变换为一个标架变换（*change of frame*）。

上列的标架变换关系式，是由新经典时空的定义和参考标架的定义推导得出的数学结果。从它的形式上看，不同的参考标架之间，没有“钟慢效应”和“尺缩效应”，这亦是新经典时空结构本身蕴含的性质。回顾新经典时空的定义可以看到，同时性的定义方式是绝对的，欧几里得度量的定义方式也是绝对的，故有此结果。

我们对同一物理事件的观测，可能得到依赖这一事件的不同物理量。在选定了标架之后，这样的物理量是依赖时间和空间的场函数，而它的函数值可以是标量、向量和线性算符。具体地，在两个标架 $(\mathscr{E},\phi)$、$(\mathscr{E}^*,\phi^*)$ 下，对同一物理事件观测得到的标量、向量和线性算符场函数分别记为：

$$\alpha,\alpha^*:\mathbb{R}^{1+3}\supset D\to\mathbb{R},\quad \alpha\left(t,\mathbf{r}\right),\quad \alpha^*\left(t^*,\mathbf{r}^*\right)$$
$$\mathbf{a},\mathbf{a}^*:\mathbb{R}^{1+3}\supset D\to\mathbb{R}^3,\quad \mathbf{a}\left(t,\mathbf{r}\right),\quad \mathbf{a}^*\left(t^*,\mathbf{r}^*\right)$$
$$\mathbf{A},\mathbf{A}^*:\mathbb{R}^{1+3}\supset D\to\mathscr{L}\left(\mathbb{R}^3\right),\quad \mathbf{A}\left(t,\mathbf{r}\right),\quad \mathbf{A}^*\left(t^*,\mathbf{r}^*\right)$$

其中 $t,t^*\in\mathbb{R}$，$\mathbf{r},\mathbf{r}^*\in\mathbb{R}^3$。设 $\mathbf{Q}_t$ 是由 $(\mathscr{E},\phi)$ 到 $(\mathscr{E}^*,\phi^*)$ 的标架变换的正交算符。若对任意两标架 $(\mathscr{E},\phi)$、$(\mathscr{E}^*,\phi^*)$，上述物理量都满足

$$\alpha^*\left(t^*,\mathbf{r}^*\right)=\alpha\left(t,\mathbf{r}\right)$$
$$\mathbf{a}^*\left(t^*,\mathbf{r}^*\right)=\mathbf{Q}_t\mathbf{a}\left(t,\mathbf{r}\right)$$
$$\mathbf{A}^*\left(t^*,\mathbf{r}^*\right)=\mathbf{Q}_t\mathbf{A}\left(t,\mathbf{r}\right)\mathbf{Q}_t^{\intercal}$$

则称这些物理量具有标架变换下的不变性（*invariance under change of frames*）。

为什么要这样定义呢？客观的标量值物理性质，本应不依赖观察的参考标架。设某时刻下空间某处有某偶极矩 $\boldsymbol{\mu}$，这是一个客观现象。在不同的参考标架下，这一偶极矩将被“画”在不同坐标系下的欧几里得空间当中，各具有一组坐标 $\mathbf{u},\mathbf{u}^*\in\mathbb{R}^3$。在同一时刻下，两个标架的欧几里得空间 $\mathscr{E}$ 和 $\mathscr{E}^*$ 的差别，可视为同一欧几里得空间内的两个直角坐标系。那么，同一向量在这两种情况之间的差别，将只相差一个由该时刻下这两个直角坐标系决定的正交矩阵 $\mathbf{Q}_t$ 作为过渡矩阵的坐标变换公式，即 $\mathbf{u}^*=\mathbf{Q}_t\mathbf{u}$。同理，线性变换值的物理量，在每一时刻下的两个标架下的坐标矩阵之间也相差一个由这两个标架确定的正交矩阵 $\mathbf{Q}_t$ 作为过渡矩阵的坐标变换公式，即 $\mathbf{A}^*=\mathbf{Q}_t\mathbf{A}\mathbf{Q}_t^{\intercal}$。可见，标架变换下的不变性定义，道出的是具有客观的物理量应该满足的条件。但是，物理量的定义及其依附的经验规律（方程）是有无限可能的，并不一定都满足标架变换不变性。

但至少，同一观察者，对同一物理量的观测结果，必须具有不依赖坐标系选择的客观性。这里包括曲线坐标系选择。然而，标量、向量和线性算符物理量及其微积分在曲线坐标变换下

的数学形式遵守一套系统的理论——张量分析（*tensor analysis*）。当我们说“标量、向量和线性算符是都是张量”时，我们实际上是主张赋予作为代数结构的标量、向量和线性算符以曲线坐标变换的理论支撑。只要注意按照曲线坐标系理论来处理物理量的坐标变换，就能使物理量满足不依赖坐标系选择的客观性。这件事在一般的教材中又叫做“用张量来表示物理量”。这是之所以要求用张量来描述流变学的本构关系的根本原因。

第六章　连续物体的运动学

6.1　速度和加速度

设物体 B 在时间段 Υ 内的运动过程是 χ，在选定标架 $(\mathscr{E},\phi)$ 下，时间段 Υ 同构于一个实数 $\mathbb{R}$ 的连通子集 I。设物体 B 由 χ 决定的相对这一标架的运动是 κ，给定时刻 $t\in I$，我们把 $\kappa_t:B\to\mathscr{E}$,

$$\kappa_t\left(P_X\right)=\kappa\left(P_X,t\right)=X\left(t\right),\quad \forall P_X\in B,t\in I$$

称为该时刻下的*置放*（*placement*）映射，它把 B 的任一物质点 P_X 对应为 t 时刻下的欧几里得空间中的点 $X\left(t\right)$。像集 $\Omega_t=\kappa_t\left(B\right)\subset\mathscr{E}$ 称物体 B 在 t 时刻下的*构型*（*configuration*）。置放和构型都是依赖所选定的标架的。我们实际可观测到的不是物体的运动 χ 本身，而是其在我们所选定标架下的构型随时间的变化。

在相对于给定标架的运动过程中，一个物体的每个物质点，在每一时刻都有不同的速度。延用上一段的设定，我们定义

$$\begin{aligned}\mathbf{v}\left(P_X,t\right)&\equiv\lim_{s\to0}\frac{X\left(t+s\right)-X\left(t\right)}{s}\\&=\frac{\mathrm{d}}{\mathrm{d}t}X\left(t\right)\\&=\dot{X}\left(t\right)\end{aligned}$$

为物质点 P_X 在时刻 t 下的*速度*（*velocity*）。由这一定义可知，速度是欧几里得空间 $\mathscr{E}$ 的平移空间 $\mathscr{V}$ 的一个向量。我们再定义

$$\begin{aligned}\mathbf{a}\left(P_X,t\right)&\equiv\lim_{s\to0}\frac{\mathbf{v}\left(P_X,t+s\right)-\mathbf{v}\left(P_X,t\right)}{s}\\&=\dot{\mathbf{v}}\left(P_X,t\right)\\&=\ddot{X}\left(t\right)\end{aligned}$$

为物质点 P_X 在时刻 t 下的*加速度*（*acceleration*）。

速度和加速度的引入都依赖标架的选择。我们将上一节介绍的定义来考察速度和加速度是否具有标架变换下的不变性。

假定有两个观察者分别选定了标架 $(\mathscr{E},\phi)$ 和 $(\mathscr{E}^*,\phi^*)$，对同一时刻的实数标记为 $t,t^*\in\mathbb{R}$ 且 $t^*=t+a$。任选 $\mathscr{E}$ 中不依赖时间的固定一点 X_0，由标架变换式有

$$X^*(t^*)=X_0^*(t^*)+\mathbf{Q}_t(X(t)-X_0)$$

其中，X_0^* 是 X_0 作为刚体系的 $\mathscr{E}$ 的物质点在标架 $(\mathscr{E}^*,\phi^*)$ 下的点，它因 $\mathscr{E}$ 相对标架 $(\mathscr{E}^*,\phi^*)$ 的运动 $\phi_t^{-1}\circ\phi_t^*$ 而依赖 t^*。X 是物质点 P_X 在标架 $(\mathscr{E},\phi)$ 下的点，它因物体 B 相对标架 $(\mathscr{E},\phi)$ 的运动而依赖 t。X^* 是物质点 P_X 在标架 $(\mathscr{E}^*,\phi^*)$ 下的点，它因物体 B 相对标架 $(\mathscr{E}^*,\phi^*)$ 的运动而依赖 t^*。$\mathbf{Q}_t$ 是由 $\mathscr{E}$ 的平移空间 $\mathscr{V}$ 到 $\mathscr{E}^*$ 的平移空间 $\mathscr{V}^*$（$\mathscr{V}^*$ 与 $\mathscr{V}$ 同构）的正交算符，它依赖 t 是等距变换的表示定理和标架变换的原理决定的（见上一节）。

任一物质点 $P_X\in B$ 在两个标架下的速度之间的关系，可由标架变换式代入速度的定义推出*——

$$\begin{aligned}\mathbf{v}^*(P_X,t)&=\dot{X}(t^*)\\&=\dot{X}_0^*(t^*)+\dot{\mathbf{Q}}_t(X(t)-X_0)+\mathbf{Q}_t\mathbf{v}(P_X,t)\\&\neq\mathbf{Q}_t\mathbf{v}(P_X,t)\end{aligned}$$

可见，速度并不具有标架变换下的不变性。由标架变换式，

$$X(t)-X_0=\mathbf{Q}_t^\intercal(X^*(t^*)-X_0^*(t^*))$$

代入上式得（为简洁，对时间依赖略去不写了）

$$\mathbf{v}^*=\dot{X}^*+\mathbf{A}(X^*-X_0^*)+\mathbf{Qv}$$

其中 $\mathbf{A}(t)\equiv\dot{\mathbf{Q}}_t\mathbf{Q}_t^\intercal$ 称标架 $(\mathscr{E},\phi)$ 相对标架 $(\mathscr{E}^*,\phi^*)$ 的自旋（*spin*）。由定理3.1及其推论，$\mathbf{A}(t)$ 唯一地对应一个向量 (t) 满足 $\mathbf{Au}=\times\mathbf{u}$，故上式又可写成更多资料中的形式，

$$\mathbf{v}^*=\dot{X}^*+\boldsymbol{\Omega}\times(X^*-X_0^*)+\mathbf{Qv}$$

上式第一项是 $\mathscr{E}$ 相对标架 $(\mathscr{E}^*,\phi^*)$ 的平动速度，第二项是二者的相对转动角速度。要使 $\mathbf{v}^*=\mathbf{Qv}$，两标架间的相对运动必须对所有时刻都满足 $\dot{X}_0^*=\mathbf{0}$ 和 $\dot{\mathbf{Q}}_t=\mathbf{0}$，即两标架相对静止。

类似地，我们可得出标架 $(\mathscr{E}^*,\phi^*)$ 下的加速度

$$\begin{aligned}\mathbf{a}^*&=\ddot{X}_0^*+\dot{\mathbf{A}}(X^*-X_0^*)+\mathbf{A}\left(\mathbf{v}^*-\dot{X}_0^*\right)+\dot{\mathbf{Q}}\mathbf{v}+\mathbf{Qa}\\&\neq\mathbf{Qa}\end{aligned}$$

*用到了 $\mathrm{d}/\mathrm{d}t^*\equiv\mathrm{d}/\mathrm{d}t$。

可见，速度并不具有标架变换下的不变性。我们对 $\dot{\mathbf{Q}}\mathbf{v}$ 一项进行如下变化，由

$$\mathbf{v}^* = \dot{X}_0^* + \mathbf{A}\left(X^* - X_0^*\right) + \mathbf{Q}\mathbf{v}$$

有

$$\mathbf{Q}\mathbf{v} = \mathbf{v}^* - \dot{X}_0^* - \mathbf{A}\left(X^* - X_0^*\right)$$

两边同时（左）乘 $\mathbf{A} = \dot{\mathbf{Q}}\mathbf{Q}^\intercal$ 得

$$\dot{\mathbf{Q}}\mathbf{V} = \mathbf{A}\left(\mathbf{v}^* - \dot{X}_0^*\right) - \mathbf{A}^2\left(X^* - X_0^*\right)$$

代入原式得

$$\mathbf{a}^* = \ddot{X}_0^* + \dot{\mathbf{A}}\left(X^* - X_0^*\right) + 2\mathbf{A}\left(\mathbf{v}^* - \dot{X}_0^*\right) - \mathbf{A}^*\left(X^* - X_0^*\right) + \mathbf{Q}\mathbf{a}$$

使用 $\boldsymbol{\Omega}$ 的表达式为

$$\mathbf{a}^* = \ddot{X}_0^* + \dot{\boldsymbol{\Omega}}\left(X^* - X_0^*\right) + 2\boldsymbol{\Omega}\left(\mathbf{v}^* - \dot{X}_0^*\right) - \boldsymbol{\Omega}\times\left(\boldsymbol{\Omega}\times\left(X^* - X_0^*\right)\right) + \mathbf{Q}\mathbf{a}$$

除 $\mathbf{Q}\mathbf{a}$ 以外的项依次是（重新强调时间依赖性）

- $\ddot{X}_0^*\left(t^*\right)$：$\mathscr{E}$ 相对 $\mathscr{E}^*$ 的平动加速度
- $\dot{\mathbf{A}}\left(t\right)\left(X^*\left(t^*\right) - X_0^*\left(t^*\right)\right)$：$\mathscr{E}$ 相对 $\mathscr{E}^*$ 的转动角加速度
- $2\mathbf{A}\left(t\right)\left(\mathbf{v}^*\left(P_X, t^*\right) - \dot{X}_0^*\left(t^*\right)\right)$：科里奥利（*Coriolis*）加速度
- $-\mathbf{A}^2\left(t\right)\left(X^*\left(t^*\right) - X_0^*\left(t^*\right)\right)$：向心（*centripetal*）加速度*

要使 $\mathbf{a}^* = \mathbf{Q}\mathbf{a}$，两标架的相对运动需要对任意时刻满足 $\ddot{X}^*\left(t^*\right) = \mathbf{0}$ 且 $\dot{\mathbf{Q}}_t = \mathbf{0}$，即两标架作相对的匀速直线运动。我们把这种特殊的标架变换称作伽俐略变换（*Galilean transform*）。速度在伽俐略变换下并不具有不变性。所有两两之间是伽俐略变换的标架所组成的集合形成一个以标架变换为群操作的群，称为伽俐略群（*Galilean group*），群元素称惯性参考标架（*inertial frame of reference*）。

6.2　形变

在给定标架 $(\mathscr{E}, \phi)$ 下，设物体 B 在两时刻 $t_1, t_2 \in \mathbb{R}$ 下的构型分别为

$$\Omega_{t_1} = \kappa_{t_1}\left(B\right), \quad \Omega_{t_2} = \kappa_{t_2}\left(B\right)$$

则复合映射

$$\mu_{t_1\to t_2} : \Omega_{t_1} \to \Omega_{t_2}, \quad \mu_{t_1\to t_2} = \kappa_{t_2}\circ\kappa_{t_1}^{-1}$$

*其反号称离心（*centrifugal*）加速度。

称为物体 B 由 t_1 到 t_2 的形变（*deformation*）。我们规定形变映射是可逆的；通常情况下还是任意所需阶光滑的。

如图6.2所示，我们常选定一个固定的时刻 $t_0 \in \mathbb{R}$ 作为参考时刻（*reference instant*），t_0 时物体 B 的构型 Ω_0 称该物体的参考构型（*reference configuration*），任一时刻 t 下物体 B 的构型称为当前构型（*current configuration*）。在默认选定了参考构型的讨论中，我们依照以下记号惯例：

- 物体 B 的物质点为 $X, Y, \cdots \in B$
- 参考构型的位置向量：$\mathbf{X} = \kappa_0(X), \mathbf{Y} = \kappa_0(Y), \cdots \in \Omega_o$
- 当前构型的位置向量：$\mathbf{x}(t) = \kappa_t(X) = \mu_t(\mathbf{X}), \mathbf{y}(\mathbf{t}) = \kappa_t(Y) = \mu_t(\mathbf{Y}), \cdots \in \Omega_t$

在上列记法中，使用同一字母，表示来自同一物质点。置放映射的像本应是欧几里得空间的点，使用向量来表示置放映射的像，是指它在基本坐标系下的位置向量。形变映射 μ 的下标可以只写当前时刻 t，有时我们也把时间依赖性明显地表示出来，即 $\mu(\cdot, t) \equiv \mu_t(\cdot)$。

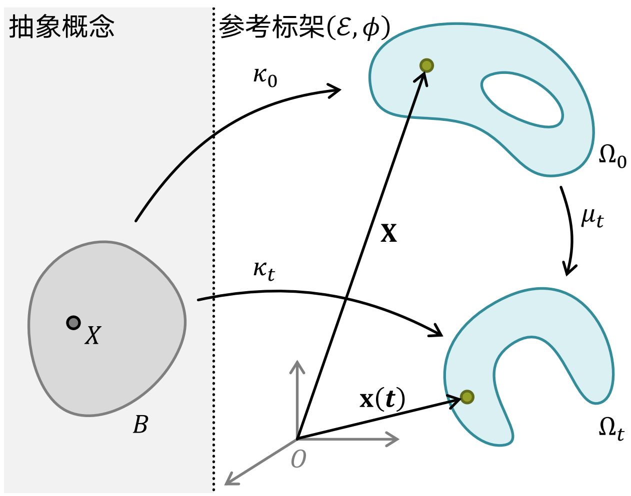


图 6.1: 物体 B 是物质点 $X, Y, \ldots$ 的集合，它只有在发生某种运动时，才能形成一系列物理事件，并在我们所选定的参考标架 $(\mathscr{E}, \phi)$ 下的欧几里得空间中表现出来。在参考时刻 t_0，置放映射 κ_0 将物体 B 的每一物质点 X 映射为参考构型 Ω_0 中的一点 $\mathbf{X}$。在当前时刻 t，置放映射 κ_t 将同一物质点 $X \in B$ 映射为当前构型 Ω_t 中的一点 $\mathbf{x}(t)$。观察者只能观测到从参考构型到当前构型的变化，由形变映射 $\mu_t = \kappa_t \circ \kappa_0^{-1}$ 描述。

在选定了坐标系后，欧几里得空间 $\mathscr{E}$ 中的任一位置都唯一对应 $\mathbb{R}^3$ 的一个有序实数 3 元组。形变映射 μ_t 的自变量和取值都是欧几里得空间中的点。在不同的坐标系选取方式下，同一形变 μ_t 将会是不同的 $\mathbb{R}^3$ 上的向量函数。以欧几里得空间 $\mathscr{E}$ 的基本直角坐标系为例，若

$\mathbf{X}=(X_1,X_2,X_3)$、$\mathbf{x}(t)=(x_1(t),x_2(t),x_3(t))$，则映射 μ_t 可由以下关系式完整确定

$$\mu_t:\begin{cases} x_1(t) &= \eta(X_1,X_2,X_3,t) \\ x_2(t) &= \xi(X_1,X_2,X_3,t) \\ x_3(t) &= \zeta(X_1,X_2,X_3,t) \end{cases}$$

其中 η、ξ、ζ 为选定某坐标系后，形变 μ_t 所对应的分量函数，它们的形式将依赖坐标系的选取，遵循 X_i 和 x_i 的曲线坐标系变换而相应地变化*。

物体 B 在 t 时刻相对参考构型的位移场（*displacement field*）是

$$\mathbf{u}(\mathbf{X},t)\equiv\mu_t(\mathbf{X})-\mathbf{X}=\mathbf{x}(t)-\mathbf{X}$$

速度场（*velocity field*）是

$$\mathbf{v}(\mathbf{X},t)=\frac{\mathrm{d}}{\mathrm{d}t}\mu(\mathbf{X},t)$$

加速度场（*acceleration field*）是

$$\mathbf{a}(\mathbf{X},t)=\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{v}(\mathbf{X},t)$$

由上一节的定义，我们有

$$\begin{aligned} &\mathbf{v}(\mathbf{X},t)=\mathbf{v}(X,t), \\ &\mathbf{a}(\mathbf{X},t)=\mathbf{a}(X,t),\quad \forall\mathbf{X}=\kappa_t(X),X\in B \end{aligned}$$

上列的这些运动学场函数都比较完整地以相对某标架的运动的形式描述了物体 B 的客观运动。但是连续介质力学关心的是材料的形变，它要通过物质点之间的相对运动来体现。我们的想法是，排除了平动的、“纯粹的”形变也许可以通过不同物质点在形变映射 μ_t 作用后的相对差别来显示，这一直觉指示我们对 $\mu_t(\mathbf{X})$ 求关于 $\mathbf{X}$ 的导数。如图6.2所示，参考构型位置 $\mathbf{X}$ 和 $\mathbf{X}+\Delta\mathbf{X}$ 被形变映射 μ_t 后的当前构型位置之差如果不等于 $\Delta\mathbf{X}$，则暗示在参考位置 $\mathbf{X}$ 邻近的物质点发生了超出平动的额外形变。

一般地，关于参考构型不同点 $\mathbf{X}$ 的邻域可能发生不同类型和程度的形变。对于连续物体，我们可以用形变映射的微分来表示每一点 $\mathbf{X}$ 邻域的形变：

$$\mu_t(\mathbf{X}+\Delta\mathbf{X})-\mu_t(\mathbf{X})=\mathbf{F}(\mathbf{X},t)\,\Delta\mathbf{X}+o(\|\Delta\mathbf{x}\|)$$

其中

$$\mathbf{F}(\mathbf{X},t)\equiv\mathrm{d}_{\mathbf{X}'=\mathbf{X}}\mu_t(\mathbf{X}')$$

*关于曲线坐标系的一般理论和本讲义大部分物理量的曲线坐标系变换结果都集中在附录中介绍。

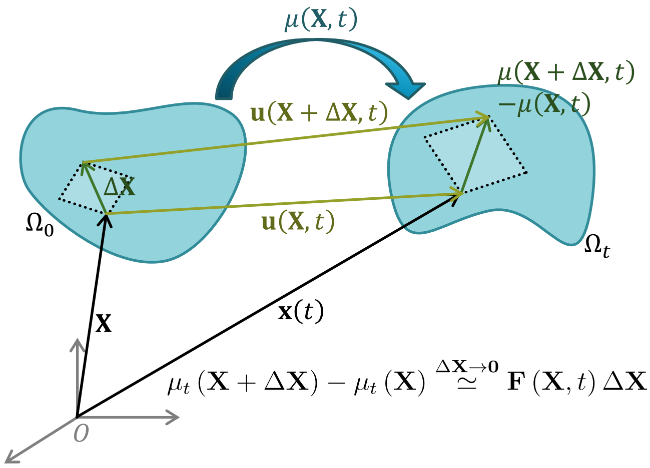


图 6.2: 在参考位置 $\mathbf{X}$ 和 $\mathbf{X}+\Delta\mathbf{X}$ 的两物质点，在形变 μ_t 后的当前位置差别，是映射 $\mu(\Delta\mathbf{X},t) \equiv \mu_t(\mathbf{X})$ 关于 $\mathbf{X}$ 的增量。由形变梯度张量 $\mathbf{F}$ 的定义，当 $\Delta\mathbf{X}\to\mathbf{0}$ 时 μ_t 的增量可由 $\mathbf{F}\Delta\mathbf{X}$ 近似。因此 $\mathbf{F}(\mathbf{X},t)$ 实际表示在参考位置 $\mathbf{X}$ 的邻域的形变。图中还标出了位移场 $\mathbf{u}(\mathbf{X},t)$，根据向量关系可知 $\mathrm{d}_{\mathbf{X}}\mathbf{u}(\mathbf{X},t)=\mathbf{F}-\mathbf{I}$。

是函数 μ_t 在 $\mathbf{X}$ 处的导数，它是平移空间 $\mathscr{V}$ 上的线性算符。由于 μ_t 是任意所需阶光滑的，由定理4.12的推论，$\det\mathbf{F}\neq 0$，即 $\mathbf{F}(\mathbf{X},t)$ 的取值总是可逆线性算符。

由 $\mathbf{F}$ 所对应的微分式和图6.2可知，线性算符 $\mathbf{F}(\mathbf{X},t)$ 实际表示在参考位置 $\mathbf{X}$ 的小邻域发生形变。一个线性算符如何表示形变，可回顾 §3.3。

形变梯度张量与位移场的梯度只相关一个单位算符 $\mathbf{I}$。由位移场的微分：

$$\begin{aligned}
\mathbf{u}(\mathbf{X}+\Delta\mathbf{X})-\mathbf{u}(\mathbf{X}) &= \mathrm{d}_{\mathbf{X}}\mathbf{u}(\mathbf{X})\,\Delta\mathbf{X}+o(\|\Delta\mathbf{X}\|)\\
&= \mu_t(\mathbf{X}+\Delta\mathbf{X})-(\mathbf{X}+\Delta\mathbf{X})-[\mu_t(\mathbf{X})-\mathbf{X}]\\
&= \mu_t(\mathbf{X}+\Delta\mathbf{X})-\Delta\mathbf{X}\\
&= \mathbf{F}(\mathbf{X},t)\,\Delta\mathbf{X}+o(\|\Delta\mathbf{X}\|)
\end{aligned}$$

可得 $\mathrm{d}_{\mathbf{X}}\mathbf{u}(\mathbf{X},t)=\mathbf{F}(\mathbf{X},t)-\mathbf{I}$。

在 $\mathscr{E}$ 的基本直角坐标系下，$\mathbf{F}(\mathbf{X},t)$ 的坐标矩阵是

$$\begin{pmatrix}
\frac{\partial\eta}{\partial X_1} & \frac{\partial\eta}{\partial X_2} & \frac{\partial\eta}{\partial X_3}\\
\frac{\partial\xi}{\partial X_1} & \frac{\partial\xi}{\partial X_2} & \frac{\partial\xi}{\partial X_3}\\
\frac{\partial\zeta}{\partial X_1} & \frac{\partial\zeta}{\partial X_2} & \frac{\partial\zeta}{\partial X_3}
\end{pmatrix}$$

一般地，$\mathbf{F}(\mathbf{X},t)$ 的取值依赖是依赖参考位置 $\mathbf{X}$ 的。若 $\mathbf{F}(\mathbf{X},t)$ 不依赖 $\mathbf{X}$，则称所关心物体的形变 μ_t 是均匀（*homogeneous*）的。特别地，当 $\mathbf{F}\equiv\mathbf{I}$ 时，所关心的物体的每个物质点都只发生大小和方向相同的位移（位移场梯度为零）。

由极分解定理2.37，结合 $\mathbf{F}$ 的值总是可逆算符，可知形变梯度张量总有极分解

$$\mathbf{F} = \mathbf{RU} = \mathbf{VR}$$

其中 $\mathbf{R}(\mathbf{X}, t)$ 的值是正交算符，$\mathbf{U}(\mathbf{X}, t)$、$\mathbf{V}(\mathbf{X}, t)$ 的值是非负对称算符，分别称为右、左拉伸张量（*right/left stretch tensor*），且有 $\mathbf{V} = \mathbf{R}^\intercal\mathbf{UR}$。这时，§3.3的内容就可以更加直接地用于理解 $\mathbf{F}$。若可以简单说正交算符表示刚体旋转、对称算符表示 3 轴拉伸、算符的映射复合是形变的叠加操作，那么极分解定理告诉我们任何一个局域的形变都是一个局域的 3 轴拉伸和一个局域的刚体旋转的叠加。由

$$\det\mathbf{F} = \det(\mathbf{RU}) = \det(\mathbf{VR})$$

以及正交算符性质 $\det\mathbf{R} = \pm 1$ 和规定 $\det\mathbf{F} \neq 0$ 可知，$\det\mathbf{U} = \mathbf{det}\mathbf{V} > 0$。由于与 §3.4最后所述类似的理由，物体无法发生镜象翻转的形变*，因此在连续介质力学中 $\mathbf{R}$ 只允许为行列式等于 1 的正交算符。这一规定往往等价地表述为：$\det\mathbf{F} > 0$。

由于 $\mathbf{U}$ 或 $\mathbf{V}$ 是正定对称张量且它们正交等价，故它们有相同的一组特征值 $\lambda_1, \lambda_2, \lambda_3$，称为主拉伸比（*principal stretch ratios*）。$\mathbf{U}$ 或 $\mathbf{V}$ 的相应的主轴称主拉伸方向（*principal stretch directions*）。

如果我们认为，一个物体在运动中除刚体平动外，只发生了刚体旋转，则“形变为零”，那么形变梯度张量 $\mathbf{F} = \mathbf{I}$ 并不意味着这种意义上的“无形变”，因为至少当一个物体发生由 $\mathbf{R}$ 描述的刚体旋转时，$\mathbf{F} = \mathbf{R} \neq \mathbf{I}$。满足我们的这一意图的形变的量度应该是拉伸张量 $\mathbf{U}$ 或 $\mathbf{V}$。我们对上述意义下的“零形变”的关注来自对弹性材料的直观感受。我们对弹性材料直观体验是，材料产生应力是由于材料形成了应变；无应变则无应力†。因此我们可以说，$\mathbf{U}$ 或 $\mathbf{V}$ 是应变量度（*strain measures*）。一切仅由三个主拉伸方向和主拉伸比确定的线性算符值函数都可等价地作为应变量度[14]Sect.32。比如由 $\mathbf{U}$ 或 $\mathbf{V}$ 分别定义的

$$\begin{aligned}\mathbf{C}(\mathbf{X}, t) &\equiv \mathbf{U}^2(\mathbf{X}, t)\\ &= (\mathbf{F}(\mathbf{X}, t))^\intercal\,\mathbf{F}(\mathbf{X}, t)\end{aligned}$$

称右柯西–格林应变张量（*right Cauchy–Green strain tensor*），

$$\begin{aligned}\mathbf{B}(\mathbf{X}, t) &\equiv \mathbf{V}^2(\mathbf{X}, t)\\ &= \mathbf{F}(\mathbf{X}, t)\,(\mathbf{F}(\mathbf{X}, t))^\intercal\end{aligned}$$

*右手无法形变为左手。

†在后面的章节将看到，严格定义下的应力并不依赖应变，更不一定在无应变时为零。

称左柯西–格林应变张量（*left Cauchy–Green strain tensor*）。当且仅当主拉伸比取值都为 1 时，$\mathbf{U}=\mathbf{V}=\mathbf{C}=\mathbf{B}=\mathbf{I}$。故物体存在的刚体平动或转动都不影响上述应变量度的取值。我们希望在相同的条件下，应变量度的取值为“零”，故再定义

$$\mathbf{E}\left(\mathbf{X},t\right)\equiv\frac{1}{2}\left(\mathbf{C}\left(\mathbf{X},t\right)-\mathbf{I}\right)$$
$$\mathbf{e}\left(\mathbf{X},t\right)\equiv\frac{1}{2}\left(\mathbf{I}-\left(\mathbf{B}\left(\mathbf{X},t\right)^{-1}\right)\right)$$

分别称格林应变张量（*Green strain tensor*）和阿尔曼西应变张量（*Almansi strain tensor*）*。这些应变的量度都常用于表示材料的本构关系。

*注意区分两种不同意义的“逆”的记号：$\mathbf{B}^{-1}\left(\mathbf{X},t\right)$ 表示关于自变量 $\left(\mathbf{X},t\right)$ 的逆映射，$\left(\mathbf{B}\left(\mathbf{X},t\right)\right)^{-1}$ 表示作为函数值的线性算符的逆变换。

第三部分

附录

附录 A　线性代数部分定理的证明

A.1　范的定义的等价

本节将证明有限维向量空间上的不同范的定义是等价的。

引理 A.1. 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的有限维向量空间，$n = \dim\mathscr{V}$，$\|\cdot\|$ 是定义在 $\mathscr{V}$ 的一个范，$\|\cdot\|_{\mathrm{E}} \equiv (\mathbf{a}|\mathbf{a})^{\frac{1}{2}}\,\forall\mathbf{a} \in \mathscr{V}$ 是定义在 $\mathscr{V}$ 上的欧几里得范，则总存在正实数 $k > 0, K > 0$ 使得 $k\|\mathbf{x}\|_{\mathrm{E}} \leq \|\mathbf{x}\| \leq K\|\mathbf{x}\|_{\mathrm{E}}, \forall\mathbf{x} \in \mathscr{V}$。

证明. 选择 $\mathscr{V}$ 的任意一组基 $\{\mathbf{e}_i\}$，任一向量 $\mathbf{x} \in \mathscr{V}$ 可表示成 $\mathbf{x} = \sum_{i=1}^{n} x_i\mathbf{e}_i$。由范的一般定义有

$$\begin{aligned}
\|\mathbf{x}\| &= \left\|\sum_{i=1}^{n} x_i\mathbf{e}_i\right\| \\
&\leq \sum_{i=1}^{n} |x_i|\,\|\mathbf{e}_i\| \quad \text{（三角不等式，当且仅当}x_1 = \cdots = x_n\text{时取等号。）} \\
&\leq \max\{|x_i|\}\sum_{i=1}^{n}\|\mathbf{e}_i| \quad \text{（非负实数求和的简易性质，当且仅当 }|x_1| = \cdots|x_n|\text{ 时取等号。）} \\
&\leq \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}\sum_{i=1}^{n}\|\mathbf{e}_i\| \quad \text{（当且仅当}n = 1\text{时取等号。）} \\
&= K\|\mathbf{x}\|_{\mathrm{E}}
\end{aligned}$$

其中 $K = \sum_{i=1}^{n}\|\mathbf{e}_i\| > 0$（基向量 $\mathbf{e}_i \neq \mathbf{0}\forall i = 1, \cdots, n$）。故总存在 $K > 0$ 使得 $\|\mathbf{x}\| \leq K\|\mathbf{x}\|_{\mathrm{E}}\,\forall\mathbf{x} \in \mathscr{V}$。

考虑一般范关于欧几里得范的函数：$f : \mathbb{R} \supset (0, +\infty] \to \mathbb{R}, f(\|\mathbf{x}\|_{\mathrm{E}}) = \|\mathbf{x}\|, \forall\mathbf{x} \in \mathscr{V}$。则对任意 $\mathbf{x}, \mathbf{x}_0 \in \mathscr{V}$，有 $|\|\mathbf{x}\|_{\mathrm{E}} - \|\mathbf{x}_0\|_{\mathrm{E}}| \leq \|\mathbf{x} - \mathbf{x}_0\|_{\mathrm{E}}$，故总存在 $k' \leq 1$ 使得 $|\|\mathbf{x}\|_{\mathrm{E}} - \|\mathbf{x}_0\|_{\mathrm{E}}| = k'\|\mathbf{x} - \mathbf{x}_0\|_{\mathrm{E}}$。前面又证明了，总存在 $K' > 0$ 使得 $\|\mathbf{x} - \mathbf{x}_0\| \leq K'\|\mathbf{x} - \mathbf{x}_0\|_{\mathrm{E}}$。故对任意 $\epsilon > 0$，总可取 $\delta = \frac{\epsilon k'}{K'}$，使得只要 $|\|\mathbf{x}\|_{\mathrm{E}} - \|\mathbf{x}_0\|_{\mathrm{E}}| < \delta$，就有 $|\|\mathbf{x}\| - \|\mathbf{x}_0\|| \leq \|\mathbf{x} - \mathbf{x}_0\| \leq K'\|\mathbf{x} - \mathbf{x}_0\|_{\mathrm{E}} = \frac{K'}{k'}|\|\mathbf{x}\|_{\mathrm{E}} - \|\mathbf{x}_0\|_{\mathrm{E}}| < \frac{K'\delta}{k'} = \epsilon$，即函数 f 是连续函数。

由函数的极值定理，闭区间上的连续函数必存在最小值。故在闭区间 $\|\mathbf{x}\|_{\mathrm{E}} \leq 1$ 范围内，f 必存在最小值，即必存在 $0 \leq a \leq 1$ 使得 $0 < f(a) \leq \|\mathbf{x}\|\ \forall \mathbf{x} \in \{\mathbf{x}|\mathbf{x} \in \mathscr{V}, \|\mathbf{x}\|_{\mathrm{E}} \leq 1\}$。又因为对任一 $\mathbf{x} \in \mathscr{V}$ 总有 $\hat{\mathbf{x}} \equiv \frac{\mathbf{x}}{\|\mathbf{x}\|_{\mathrm{E}}} \in \{\mathbf{x}|\mathbf{x} \in \mathscr{V}, \|\mathbf{x}\|_{\mathrm{E}} \leq 1\}$。故总存在 $k > 0$ 使得

$$
\begin{aligned}
& k \leq \|\hat{\mathbf{x}}\| \leq K \|\hat{\mathbf{x}}\|_{\mathrm{E}} \\
\Leftrightarrow & k \leq \frac{\|\mathbf{x}\|}{\|\mathbf{x}\|_{\mathrm{E}}} \leq K \\
\Leftrightarrow & k \|\mathbf{x}\|_{\mathrm{E}} \leq \|\mathbf{x}\| \leq K \|\mathbf{x}\|_{\mathrm{E}}
\end{aligned}
$$

□

引理 A.2. 如果存在 $k > 0, K > 0$ 使得数域 $\mathbb{F}$ 上的有限维向量空间 $\mathscr{V}$ 上的任意两种范的定义 $\|\cdot\|_1, \|\cdot\|_2$ 满足 $k\|\mathbf{x}\|_1 \leq \|\mathbf{x}\|_2 \leq K\|\mathbf{x}\|_1, \forall \mathbf{x} \in \mathscr{V}$，则该不等式成立条件定义了两种范的定义之间的等价关系。

证明. 自反性：显然满足。

对称性：只需令 $k' = \frac{1}{K}, K' = \frac{1}{k}$。

传递性：设 $\|\cdot\|_1, \|\cdot\|_2, \|\cdot\|_3$ 是 $\mathscr{V}$ 上的三种范的定义，正实数 k, k', K, K' 满足 $k\|\mathbf{x}\|_1 \leq \|\mathbf{x}\|_2 \leq K\|\mathbf{x}\|_1, k'\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_3 \leq K'\|\mathbf{x}\|_2, \forall \mathbf{x} \in \mathscr{V}$，则总可取 $0 < k'' = kk' \leq KK' = K''$ 使得 $k''\|\mathbf{x}\|_1 = kk'\|\mathbf{x}\|_1 \leq k'\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_3 \leq K'\|\mathbf{x}\|_2 \leq KK'\|\mathbf{x}\|_1 = K''\|\mathbf{x}\|_1$。 □

定理 A.1. 有限维向量空间上的任意两种范的定义等价。

证明. 由以上两引理可知有限维向量空间上的任意一种范的定义均与欧几里得范等价。由等价关系性质本命题成立。 □

A.2 内积空间中的正交投影

本节是 §2.1.4的补充知识，在不同的命题证明中将会用到。

A.2.1 线性无关子空间的直和

定义 A.1 (向量组的和). 若 $S_1, \cdots, S_k$ 是向量空间 $\mathscr{V}$ 的子集，则所有 $\sum_{i=1}^{k} \mathbf{a}_i, \mathbf{a}_i \in S_i$ 的集合 S，记作 $\sum_{i=1}^{k} S_i$，称向量组 $S_1, \cdots, S_k$ 的和（*sum of the subsects*）。

注意，定义A.1中的 S 的元素，是从每个 S_i 中抽一个向量 $\mathbf{a}_i$ 出来再求和的结果。求和式当然也可以写成加法：$\sum_{i=1}^{k} S_k = S_1 + \cdots + S_k$。若 $\mathscr{W}_1, \cdots, \mathscr{W}_k$ 是 $\mathscr{V}$ 的子空间，则易验 $\mathscr{W} = \sum_{i=1}^{k} \mathscr{W}_i$ 也是 $\mathscr{V}$ 的一个子空间，且具体为由 $\mathscr{W}_1, \cdots, \mathscr{W}_k$ 共同线性生成的子空间。我们还能为一系列子空间定义类似“线性无关”的概念。

定义 A.2 (线性无关子空间). 设 $\mathscr{W}_1, \cdots, \mathscr{W}_k$ 是向量空间 $\mathscr{V}$ 的子空间。若

$$\sum_{i=1}^{k} \mathbf{a}_i = \mathbf{0}, \quad \mathbf{a}_i \in \mathscr{W}_i \Leftrightarrow \mathbf{a}_i = \mathbf{0}, \quad \forall i = 1, \cdots, k$$

则称 $\mathscr{W}_1, \cdots, \mathscr{W}_k$ 线性无关（*linlearly independent*）。

沿用上述符号设定，记 $\mathscr{W} = \sum_{i=1}^{k} \mathscr{W}_i$，则任一向量 $\mathbf{a} \in \mathscr{W}$ 可表示为 $\sum_{i=1}^{k} \mathbf{a}_i, \mathbf{a}_i \in \mathscr{W}_i$。若 $\{\mathscr{W}_i\}$ 是线性无关的，那么 $\mathbf{a}$ 的这一表达式就是唯一的。因为若另有 $\mathbf{b}_1, \cdots, \mathbf{b}_k$ 满足 $\mathbf{a} = \sum_{i=1}^{k} \mathbf{b}_i, \mathbf{b}_i \in \mathscr{W}_i$，则

$$\mathbf{a} - \mathbf{a} = \mathbf{0} = \sum_{i=1}^{k} (\mathbf{a}_i - \mathbf{b}_i) \Rightarrow \mathbf{a}_i - \mathbf{b}_i = \mathbf{0}, \forall i = 1, \cdots, k$$

其中用到了 $\{\mathscr{W}_i\}$ 线性无关的定义。换言之，若 $\{\mathscr{W}_i\}$ 线性无关，则 $\mathscr{W} = \sum_{i=1}^{k}$ 中的任一向量 $\mathbf{a}$ 都可唯一对应于一个向量组 $\{\mathbf{a}_i\}, \mathbf{a}_i \in \mathscr{W}$，并由后者线性表出。但是，从线性无关的子空间 $\{\mathscr{W}_i\}$ 中各取一个向量 $\mathbf{a}_i \in \mathscr{W}_i$ 组成的向量组 $\{\mathbf{a}_i\}$ 未必线性无关。因此一般地 $\{\mathbf{a}_i\}$ 不是 $\mathbf{a}$ 的基。

在有限维向量空间下，以下定理与线性无关子空间的定义之间互为充要条件。

定理 A.2. 设 $\mathscr{V}$ 是有限维向量空间，$\mathscr{W}_1, \cdots, \mathscr{W}_k$ 是 $\mathscr{V}$ 的子集，且 $\mathscr{W} = \sum_{i=1}^{k} \mathscr{W}_i$，则以下命题相互等价：

1. $\{\mathscr{W}_i\}$ 线性无关；
2. $$\forall j = 2, \cdots, k, \quad \mathscr{W}_j \cup \sum_{i=1}^{j-1} \mathscr{W}_i = \{\mathbf{0}\}$$
3. 若 B_i 是 $\mathscr{W}_i$ 的一组有序基，则 $B = (B_1, \cdots, B_k)$ 是 $\mathscr{W}$ 的一组有序基。

证明. 1⇔2：设 $\mathbf{a} \in \mathscr{W}_j \cap (\mathscr{W}_1 + \cdots + \mathscr{W}_{j-1})$，则由定义A.1存在 $\mathbf{a}_i \in \mathscr{W}_i, i = 1, \cdots, j-1$ 使得 $\mathbf{a} = \sum_{i=1}^{j-1} \mathbf{a}_i$。故有 $\sum_{i=1}^{j-1} \mathbf{a}_i - \mathbf{a} = \mathbf{0}$。若命题 1 成立，则 $\sum_{i=1}^{j-1} \mathbf{a}_i - \mathbf{a} = \mathbf{0} \Leftrightarrow \mathbf{a}_i = \mathbf{a} = \mathbf{0} \Rightarrow$ 命题 2 成立。

若命题 2 成立，设 $\mathbf{a}_i \in \mathscr{W}_i, i = 1, \cdots, k$ 满足 $\sum_{i=1}^{k} \mathbf{a}_i = \mathbf{0}$ 且前 j 个 $\mathbf{a}_i$ 不为零，则有 $\sum_{i=1}^{j} \mathbf{a}_i = \mathbf{0}$，即 $-\mathbf{a}_j = \sum_{i=1}^{j-1} \mathbf{a}_i \neq \mathbf{0}$，即 $\mathbf{a}_j \in \mathscr{W}_1 + \cdots + \mathscr{W}_{j-1}$，故 $\mathbf{a}_j \in \mathscr{W}_j \cap (\mathscr{W}_1, \cdots + \mathscr{W}_{j-1})$，

与命题 2 矛盾。换言之，因命题 2 成立，则使 $\sum_{i=1}^{k}\mathbf{a}_i=\mathbf{0}$ 的任意前 j 个向量 $\mathbf{a}_i$ 均为零，即命题 1 成立。

1⇔3：

设命题 1 成立，$B_i=(\mathbf{e}_{1i},\cdots,\mathbf{e}_{n_ii})$ 是 $\mathscr{W}_i$ 的基，$n_i=\dim\mathscr{W}_i$，$i=1,\cdots,k$，则对每一 $\mathbf{b}\in\mathscr{W}$，都存在 $\mathbf{b}_i\in\mathscr{W}_i,i=1,\cdots,k$ 满足 $\mathbf{b}=\sum_{i=1}^{k}\mathbf{b}_i$ 且 $\{\mathbf{b}_1,\cdots,\mathbf{b}_k\}$ 线性无关。设 $\mathbf{b}_j=\sum_{i=1}^{n_i}\beta_{ji}\mathbf{e}_{ji},j=1,\cdots,k$,则 $\mathbf{b}_j=\mathbf{0}\Leftrightarrow\beta_{ji}=0,\forall i=1,\cdots,n_i$,故有 $\mathbf{b}=\sum_{j=1}^{k}\sum_{i=1}^{n_i}\beta_{ji}\mathbf{e}_{ji}=\mathbf{0}\Leftrightarrow\beta_{ji}=0,\forall i=1,\cdots,n_i,j=1,\cdots,k$，即 $B=B_1\cup\cdots\cup B_k$ 是 $\mathscr{W}$ 的一组基。命题 3 成立。

设命题 3 成立，由定理的大前提依然有 $\forall\mathbf{b}\in\mathscr{W},\exists\mathbf{b}_i\in\mathscr{W}_i,i=1,\cdots,k,\sum_{i=1}^{k}\mathbf{b}_i=\mathbf{b}$。故 $\mathbf{b}=\mathbf{0}\Leftrightarrow\beta_{ikj}=0,\forall i=1,\cdots,n_i,j=1,\cdots,k\Leftrightarrow\mathbf{b}_j=\mathbf{0},\forall j=1,\cdots k$，即命题 1 成立。 □

我们把线性无关子空间的和称作直和（*direct sum*）。记作 $\mathscr{W}=\mathscr{W}_1\bigoplus\cdots\bigoplus\mathscr{W}_k$。

定义 A.3 (投影算符). 设 $\mathscr{V}$ 是一个向量空间，若 $\mathscr{V}$ 上的线性算符 $\mathbf{E}$ 满足 $\mathbf{E}^2=\mathbf{E}$，则称 $\mathbf{E}$ 是 $\mathscr{V}$ 的一个投影算符（*projection operator*）。

定理 A.3. 若 $\mathscr{V}$ 是一个向量空间，$\mathscr{W}_1,\cdots,\mathscr{W}_k$ 是 $\mathscr{V}$ 的子空间且 $\mathscr{V}=\bigoplus_{i=1}^{k}\mathscr{W}_i$，则 $\mathscr{V}$ 上存在 k 个线性算符 $\mathbf{E}_1,\cdots,\mathbf{E}_k$ 满足：

1. $\mathbf{E}_i^2=\mathbf{E}_i,\forall i=1,\cdots,k$
2. 若 $i\neq j$ 则 $\mathbf{E}_i\mathbf{E}_j=\mathbf{0}$
3. $\mathbf{I}=\mathbf{E}_1+\cdots+\mathbf{E}_k$
4. $\mathrm{ran}\mathbf{E}_i=\mathscr{W}_i$

证明. 命题 1：对每一向量 $\mathbf{a}\in\mathscr{V}$，存在 $\mathbf{a}_i\in\mathscr{W}_i,i=1,\cdots,k$ 满足 $\mathbf{a}=\sum_{i=1}^{k}\mathbf{a}_i$。故对每一 $j=1,\cdots,k$ 均可定义映射 $\mathbf{E}:\mathscr{V}\to\mathscr{W}_j,\mathbf{E}_j\mathbf{a}=\mathbf{a}_j$。易验 $\mathbf{E}_j$ 是线性算符，$\mathbf{E}_j^2=\mathbf{E}_j$。命题 4 自动得证。

命题 2：由于 $\mathrm{ran}\mathbf{E}_j=\mathscr{W}_j$，且 $\{\mathscr{W}_1,\cdots,\mathscr{W}_k\}$ 线性无关（由直和的定义），故 $\mathbf{E}_j\mathbf{a}=\mathbf{a}_j=\mathbf{0}=\mathbf{E}_j\sum_{i\neq j}\mathbf{a}_i=\sum_{i\neq j}\mathbf{E}_j\mathbf{E}_i\mathbf{a}\Leftrightarrow\mathbf{E}_j\mathbf{E}_i\mathbf{a}=\mathbf{0},\forall i\neq j$，由于 $\mathbf{a}$ 可以不为零向量，故 $\mathbf{E}_j\mathbf{E}_i\neq\mathbf{0},\forall i\neq j$。

命题 3：由 $\mathbf{a}=\mathbf{I}\mathbf{a}=\sum_{i=1}^{k}\mathbf{a}_i=\sum_{i=1}^{k}\mathbf{E}_i\mathbf{a},\forall\mathbf{a}\in\mathscr{V}$ 可直接得证。 □

A.2.2 内积空间上的正交投影

定义 A.4 (最好近似). 设 $\mathscr{V}$ 是 $\mathbb{F}$ 上的内积空间并自然定义了欧几里得范，$\mathscr{W}$ 是 $\mathscr{V}$ 的子空间，对任一 $\mathbf{b}\in\mathscr{V}$，如果 $\mathbf{a}\in\mathscr{W}$ 满足

$$\|\mathbf{b}-\mathbf{a}\|\leq\|\mathbf{b}-\mathbf{c}\|,\quad\forall\mathbf{c}\in\mathscr{W},\text{当且仅当 }\mathbf{c}=\mathbf{a}\text{ 时取等号}$$

则称 $\mathbf{a}\in\mathscr{W}$ 是 $\mathbf{b}$ 在 $\mathscr{W}$ 中的最好近似（*best approximation*）。

向量 $\mathbf{b}$ 在子空间上的最好近似，最直观的理解就是空间向量在平面上的投影向量。找到这一投影向量的方法是“作垂线”，且该平面上，$\mathbf{b}$ 的投影向量总是唯一存在。以下定理是这些直观认知的严格阐述。

定理 A.4 (最好近似的表示). 设 $\mathscr{W}$ 是内积空间 $\mathscr{V}$ 的子空间，$\mathbf{b}$ 是 $\mathscr{V}$ 的任一向量，则

1. 当且仅当 $\mathbf{a} \in \mathscr{W}$ 与 $\mathbf{b}-\mathbf{a}$ 正交时，$\mathbf{a}$ 是 $\mathbf{b}$ 在 $\mathscr{W}$ 中的最好近似。
2. $\mathbf{b}$ 在 $\mathscr{W}$ 中总存在唯一一个最好近似。
3. 若 $\mathscr{W}$ 是有限维的，$\{\mathbf{e}_1, \cdots \mathbf{e}_n\}$ 是 $\mathscr{W}$ 的一组正交基，则向量
$$\mathbf{a} = \sum_{k=1}^{n} \frac{(\mathbf{b}|\mathbf{e}_k)}{(\mathbf{e}_k|\mathbf{e}_k)} \mathbf{e}_k$$
是 $\mathbf{b}$ 在 $\mathscr{W}$ 中的唯一最小近似。

证明. 由极化恒等式，以下关系式总成立

$$\|\mathbf{b}-\mathbf{c}\|^2 = \|\mathbf{b}-\mathbf{a}+\mathbf{a}-\mathbf{c}\|^2 = \|\mathbf{b}-\mathbf{a}\|^2 + 2\mathrm{Re}\,(\mathbf{b}-\mathbf{a}|\mathbf{a}-\mathbf{c}) + \|\mathbf{a}-\mathbf{c}\|^2$$

“1”的证明：若 $\mathbf{b}-\mathbf{a}$ 与 $\mathscr{W}$ 的所有向量都正交，且 $\mathbf{c} \in \mathscr{W}, \mathbf{c} \neq \mathbf{a}$，则由三角不等式有

$$\|\mathbf{b}-\mathbf{c}\|^2 = \|\mathbf{b}-\mathbf{a}\|^2 + \|\mathbf{a}-\mathbf{c}\|^2 > \|\mathbf{b}-\mathbf{a}\|^2$$

因此 $\mathbf{b}-\mathbf{a}$ 与 $\mathscr{W}$ 的所有向量都正交 $\Rightarrow$ $\mathbf{a}$ 是 $\mathbf{b}$ 在 $\mathscr{W}$ 的最好近似。

若 $\|\mathbf{b}-\mathbf{a}\| \leq \|\mathbf{b}-\mathbf{c}\|, \forall \mathbf{c} \in \mathscr{W}$，则由最开始的等式，有

$$\|\mathbf{b}-\mathbf{c}\|^2 = \|\mathbf{b}-\mathbf{a}\|^2 = 2\mathrm{Re}\,(\mathbf{b}-\mathbf{a}|\mathbf{a}-\mathbf{c}) + \|\mathbf{a}-\mathbf{c}\|^2 \geq 0$$

而实际上，由于 $\mathbf{c}$ 是 $\mathscr{W}$ 中任意一个向量，所以 $\mathbf{a}-\mathbf{c}$ 也是 $\mathscr{W}$ 中任意一个向量。因此我们实际证明了，

$$2\mathrm{Re}\,(\mathbf{b}-\mathbf{a}|\mathbf{t}) + \|\mathbf{t}\|^2 \geq 0, \quad \forall \mathbf{t} \in \mathscr{W}$$

现不妨取

$$\mathbf{t} = -\frac{(\mathbf{b}-\mathbf{a}|\mathbf{a}-\mathbf{c})}{\|\mathbf{a}-\mathbf{c}\|^2}(\mathbf{a}-\mathbf{c})$$

则 $\mathbf{t}$ 亦须满足

$$\begin{aligned}
&2\mathrm{Re}\,(\mathbf{b}-\mathbf{a}|\mathbf{t}) + \|\mathbf{t}\|^2 \geq 0 \\
\Leftrightarrow &2\mathrm{Re}\left[\left(\mathbf{b}-\mathbf{a}\,\middle|\,-\frac{(\mathbf{b}-\mathbf{a}|\mathbf{a}-\mathbf{c})}{\|\mathbf{a}-\mathbf{c}\|^2}(\mathbf{a}-\mathbf{c})\right)\right] + \frac{|(\mathbf{b}-\mathbf{a}|\mathbf{a}-\mathbf{c})|^2}{\|\mathbf{a}-\mathbf{c}\|^2} \geq 0 \\
\Leftrightarrow &-2\frac{|(\mathbf{b}-\mathbf{a}|\mathbf{a}-\mathbf{c})|}{\|\mathbf{a}-\mathbf{c}\|^2} + \frac{|(\mathbf{b}-\mathbf{a}|\mathbf{a}-\mathbf{c})|^2}{\|\mathbf{a}-\mathbf{c}\|^2} \geq 0 \\
\Leftrightarrow &-|(\mathbf{b}-\mathbf{a}|\mathbf{a}-\mathbf{c})|^2\|\mathbf{a}-\mathbf{c}\|^{-2} \geq 0 \Leftrightarrow (\mathbf{b}-\mathbf{a}|\mathbf{b}-\mathbf{c}) = 0
\end{aligned}$$

故 $\|\mathbf{b}-\mathbf{a}\| \le \|\mathbf{b}-\mathbf{c}\|, \forall \mathbf{c} \in \mathscr{W} \Rightarrow (\mathbf{b}-\mathbf{a}|\mathbf{a}-\mathbf{c}) = 0, \forall \mathbf{c} \in \mathscr{W}$，注意 $\mathbf{a}-\mathbf{c}$ 是 $\mathscr{W}$ 中的任意向量。“1”证毕。

“2”的证明：设 $\mathbf{a}, \mathbf{a}' \in \mathscr{W}$ 都是向量 $\mathbf{b} \in \mathscr{V}$ 在 $\mathscr{W}$ 中的最好近似，且一般地 $\mathbf{b} \neq \mathbf{a}, \mathbf{b} \neq \mathbf{a}'$，则由最好近似的定义有 $(\mathbf{b}-\mathbf{a}|\mathbf{a}') = (\mathbf{b}-\mathbf{a}|\mathbf{a}) = 0$，故 $\mathbf{a}' = \mathbf{a}$。“2”证毕。

“3”的证明：回顾格拉姆–施密特正交化过程，若 $\mathscr{V}$ 中有正交向量组 $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$，$\mathbf{b}$ 与它们线性无关，则“3”中的表达式，是对线性无关向量组 $\{\mathbf{e}_1, \cdots, \mathbf{e}_n, \mathbf{b}\}$ 进行格拉姆–施密特正交化，使 $\mathbf{b}-\mathbf{a}$ 与所有 $\mathbf{e}_1, \cdots, \mathbf{e}_n$ 都正交时所使用的表达式。换言之，若 $\mathbf{a}$ 满足命题，则 $\mathbf{b}-\mathbf{a}$ 与 $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ 都正交，亦与它们线性生成的子空间（即 $\mathscr{W}$）上的所有向量都正交。由“1”，$\mathbf{a}$ 是 $\mathbf{b}$ 在 $\mathscr{W}$ 中的最好近似。由“2”，$\mathbf{a}$ 是唯一的。“3”证毕。 □

定义 A.5 (正交补集). 设 $\mathscr{V}$ 是一个内积空间，S 是 $\mathscr{V}$ 的任一子集，$\mathscr{V}$ 中的与 S 中每个向量都正交的向量的集合记为 $S^\perp$，称 S 的正交补集（*orthogonal complement*）。

虽然我们定义了正交补集是什么，但仍担心未必总能为任一集合找到其非空的正交补集。但由于 $\mathbf{0}$ 与所有向量都正交，故甚至能说 $\mathscr{V}$ 的正交补集是 $\{\mathbf{0}\}$，反之 $\{\mathbf{0}\}^\perp = \mathscr{V}$。

$S^\perp$ 总是 $\mathscr{V}$ 的一个子空间。首先 $S^\perp$ 总是含 $\mathbf{0}$。其次，若 $\mathbf{a}, \mathbf{b} \in S^\perp$，则 $\forall \mathbf{c} \in S$，

$$(\mathbf{a}+\gamma\mathbf{b}|\mathbf{c}) = (\mathbf{a}|\mathbf{c}) + \gamma\,(\mathbf{b}|\mathbf{c}) = 0, \quad \forall \gamma \in \mathbb{F}$$

其中 $\mathbb{F}$ 是 $\mathscr{V}$ 所在的数域。

定义 A.6 (正交投影). 设 $\mathscr{V}$ 是一个内积空间，$\mathbf{b}$ 是 $\mathscr{V}$ 中的一个向量，$\mathscr{W}$ 是 $\mathscr{V}$ 的一个子空间。若 $\mathscr{W}$ 中存在向量 $\mathbf{a}$ 是 $\mathbf{b}$ 在 $\mathscr{W}$ 中的最好近似，则称向量 $\mathbf{a}$ 是向量 $\mathbf{b}$ 在 $\mathscr{W}$ 上的正交投影（*orthogonal projection of* $\mathbf{b}$ *on* $\mathscr{W}$）。若 $\mathscr{W}$ 的维数是 $m \le n$，$\{\mathbf{e}_k\}_{k=1}^m$ 是 $\mathscr{W}$ 的一组正交基，定义映射 $\mathbf{E}: \mathscr{V} \to \mathscr{W}$,

$$\mathbf{E}\mathbf{b} = \sum_{k=1}^{m} \frac{(\mathbf{b}|\mathbf{e}_k)}{(\mathbf{e}_k|\mathbf{e}_k)}\mathbf{e}_k, \quad \forall \mathbf{b} \in \mathscr{V}$$

则称映射 $\mathbf{E}$ 是内积空间 $\mathscr{V}$ 到其子空间 $\mathscr{W}$ 的正交投影算符（*orthogonal projection operator of* $\mathscr{V}$ *to* $\mathscr{W}$）。

定理A.4成立的内积空间 $\mathscr{V}$ 未必是有限维的，故基于定理A.4“3”的概念的正交投影映射，可以是无限维内积空间向其有限维子空间的投影。由定理A.4“3”，这样的投影映射是总存在的，作为推论如下。

推论 A.4.1. 延用定理A.4的设定，设映射 $\mathbf{E}': \mathscr{V} \to \mathscr{W}, \mathbf{E}'\mathbf{b} = \mathbf{b} - \mathbf{E}\mathbf{b}, \forall \mathbf{b} \in \mathscr{V}$ 是由 $\mathscr{V}$ 到 $\mathscr{W}^\perp$ 的正交投影。

证明. 由定义A.4和定理A.2，$\mathbf{Eb}$ 就是向量 $\mathbf{b}$ 在 $\mathscr{W}$ 中的最好近似，因此 $\mathbf{b}-\mathbf{Eb}$ 与所有 $\mathscr{W}$ 正交，即 $\mathbf{E'b}\in\mathscr{W}^{\perp}$。由于 $\mathbf{Eb}\in\mathscr{W}$，故 $\mathbf{Eb}=\mathbf{b}-\mathbf{E'b}$ 与 $\mathscr{W}^{\perp}$ 中的所有向量都正交。由定理A.4 “1”，$\mathbf{E'b}$ 是 $\mathbf{b}$ 在 $\mathscr{W}^{\perp}$ 上的最好近似。 □

我们把正交投影映射写成线性算符的样子，因为它确实是线性算符，作为定理如下。

定理 A.5. 设 $\mathscr{V}$ 是内积空间，$\mathscr{W}$ 是 $\mathscr{V}$ 的一个子空间，$\mathbf{E}$ 是由 $\mathscr{V}$ 到 $\mathscr{W}$ 的正交投影，则 $\mathbf{E}$ 是一个投影算符（或称其为幂等（*idempotent*）线性算符），$\mathscr{W}^{\perp}$ 是 $\mathbf{E}$ 的零空间，$\mathscr{V}=\mathscr{W}\bigoplus\mathscr{W}^{\perp}$。

证明. 设 $\mathbf{b}\in\mathscr{V}$，则有 $\mathbf{Eb}$ 是 $\mathbf{b}$ 在 $\mathscr{W}$ 上的最好近似，$\mathbf{E}(\mathbf{Eb})=\mathbf{Eb}$，故 $\mathbf{E}^2=\mathbf{E}$。

设 $\mathbf{a},\mathbf{b}\in\mathscr{V},\gamma\in\mathbb{F}$，其中 $\mathbb{F}$ 是 $\mathscr{V}$ 所在的数域，则向量 $\mathbf{a}-\mathbf{Ea}$ 和 $\mathbf{b}-\mathbf{Eb}$ 分别都与 $\mathscr{W}$ 的所有向量正交，即这两个向量都属于 $\mathscr{W}^{\perp}$。由于 $\mathscr{W}^{\perp}$ 必是 $\mathscr{V}$ 的子空间，故有

$$\mathbf{a}-\mathbf{Ea}+\gamma(\mathbf{b}-\mathbf{Eb})\in\mathscr{W}^{\perp}\Leftrightarrow\mathbf{a}+\gamma\mathbf{b}-\mathbf{E}(\mathbf{a}+\gamma\mathbf{b})\in\mathscr{W}^{\perp}$$

因此 $\mathbf{E}$ 是幂等线性算符。

前文说过，$\mathbf{Eb}$ 是 $\mathscr{W}$ 中唯一使 $\mathbf{b}-\mathbf{Eb}\in\mathscr{W}^{\perp}$ 的向量，故若 $\mathbf{Eb}=\mathbf{0}$，则 $\mathbf{b}-\mathbf{Eb}=\mathbf{b}\in\mathscr{W}^{\perp}$。反之，若 $\mathbf{b}\in\mathscr{W}^{\perp}$，设 $\mathbf{a}\in\mathscr{W}$ 是使 $\mathbf{b}-\mathbf{a}\in\mathscr{W}^{\perp}$ 的向量，则有 $(\mathbf{a}|\mathbf{b})=0$ 且 $(\mathbf{a}|\mathbf{b}-\mathbf{a})=0$，故只可能 $\mathbf{a}=\mathbf{0}$。而事实上 $\mathbf{Eb}$ 是 $\mathscr{W}$ 中唯一使 $\mathbf{b}-\mathbf{Eb}\in\mathscr{W}^{\perp}$ 的向量，故 $\mathbf{a}=\mathbf{Eb}=\mathbf{0}$。换言之，$\mathscr{W}^{\perp}$ 中的向量全是使 $\mathbf{Eb}=\mathbf{0}$ 的 $\mathbf{b}$，即 $\mathscr{W}^{\perp}=\ker\mathbf{E}$。

由 $\mathbf{Eb}$ 与 $\mathbf{b}$ 和 $\mathbf{b}-\mathbf{Eb}$ 的唯一性，有以下恒等式

$$\mathbf{b}=\mathbf{Eb}+\mathbf{b}-\mathbf{Eb},\forall\mathbf{b}\in\mathscr{V}$$

说明总有 $\mathscr{V}=\mathscr{W}+\mathscr{W}^{\perp}$。而由前面的证明又知 $\mathscr{W}\cup\mathscr{W}^{\perp}=\{0\}$，由定理A.2有 $\mathscr{V}=\mathscr{W}\bigoplus\mathscr{W}^{\perp}$。 □

由于一个内积空间 $\mathscr{V}$ 中的每一向量 $\mathbf{b}\in\mathscr{V}$ 在给定的一个 $\mathscr{V}$ 的有限维子空间 $\mathscr{W}$ 上的正交投影是唯一的（定理A.4），故由 $\mathscr{V}$ 到 $\mathscr{W}$ 的正交投影算符 $\mathbf{E}$（定义A.6）也由所给定的这一子空间 $\mathscr{W}$ 唯一确定。

A.3　内积空间上的线性算符相关证明

A.3.1　伴算算符的唯一存在性

定理 A.6. 设 $\mathscr{V}$ 是数域 $\mathbb{F}$ 上的有限维内积空间，对 $\mathscr{V}$ 上的任一线性算符 $\mathbf{T}\in\mathscr{L}(\mathscr{V})$ 有且只有一个伴随算符 $\mathbf{T}^*\in\mathscr{L}(\mathscr{V})$。

证明. 给定任意向量 $\mathbf{b}\in\mathscr{V}$，均可定义一个线性泛函 $f\in\mathscr{V}^*, f(\mathbf{a})\equiv(\mathbf{Ta}|\mathbf{b}),\forall\mathbf{a}\in\mathscr{V}$。由有限维 Riesz 表示定理（2.19），每一个这样的线性泛函都唯一对应一个 $\mathbf{b}'\in\mathscr{V}$ 满足 $f(\mathbf{a})=(\mathbf{a}|\mathbf{b}')$。因此，对每一个线性变换 $\mathbf{T}\in\mathscr{L}(\mathscr{V})$，都可以定义一个映射 $\mathbf{T}^*:\mathscr{V}\to\mathscr{V}$ 来将 $\mathscr{V}$ 中的每一个向量 $\mathbf{b}$ 如上所述地对应到 $\mathscr{V}$ 中的另一个向量 $\mathbf{b}'$，且这个映射 $\mathbf{T}^*$ 是一个线性算符，因为对任意 $\gamma\in\mathbb{F},\mathbf{b},\mathbf{c}\in\mathscr{V}$，

$$
\begin{aligned}
(\mathbf{a}|\mathbf{T}^*(\gamma\mathbf{b}+\mathbf{c}))&=(\mathbf{Ta}|\gamma\mathbf{b}+\mathbf{c})\\
&=\overline{\gamma}(\mathbf{Ta}|\mathbf{b})+(\mathbf{Ta}|\mathbf{c})\\
&=\overline{\gamma}(\mathbf{a}|\mathbf{T}^*\mathbf{b})+(\mathbf{a}|\mathbf{T}^*\mathbf{c})\\
&=(\mathbf{a}|\gamma\mathbf{T}^*\mathbf{b}+\mathbf{T}^*\mathbf{c}),\forall\mathbf{a}\in\mathscr{V}\\
\Leftrightarrow \mathbf{T}^*(\gamma\mathbf{b}+\mathbf{c})&=\gamma\mathbf{T}^*\mathbf{b}+\mathbf{T}^*\mathbf{c}
\end{aligned}
$$

现证明 $\mathbf{T}^*$ 是唯一的。设另一线性算符 $\mathbf{U}\in\mathscr{L}(\mathscr{V})$ 满足命题条件即 $(\mathbf{Ta}|\mathbf{b})=(\mathbf{a}|\mathbf{T}^*\mathbf{b})=(\mathbf{a}|\mathbf{Ub})$，则 $0=(\mathbf{a}|\mathbf{T}^*\mathbf{b})-(\mathbf{a}|\mathbf{Ub})=(\mathbf{a}|(\mathbf{T}^*-\mathbf{U})\mathbf{b}),\forall\mathbf{a},\mathbf{b}\in\mathscr{V}\Leftrightarrow\mathbf{T}^*=\mathbf{U}$，故 $\mathbf{T}^*$ 是唯一的。 □

A.3.2 定理2.32“4”的证明

若向量空间 $\mathscr{V}$ 上定义了线性算符 $\mathbf{T}$，则子空间 $\mathscr{W}$ 上的向量被 $\mathbf{T}$ 作用后可能不再属于 $\mathscr{W}$，于是我们定义以下的特殊情况。

定义 A.7 (子空间在线性算符的作用下不变). 设 $\mathbf{T}$ 是向量空间 $\mathscr{V}$ 上的一个线性算符，$\mathscr{W}$ 是 $\mathscr{V}$ 的一个子空间。如果 $\mathbf{Ta}\in\mathscr{W},\forall\mathbf{a}\in\mathscr{W}$，则称子空间 $\mathscr{W}$ 在线性算符 $\mathbf{T}$ 的作用下不变（*invariant under* $\mathbf{T}$）。

显然，任一向一空间 $\mathscr{V}$ 在其上的任一线性算符的作用下不变。

定理 A.7 (定理2.32“4”). 给定有限维内积空间 $\mathscr{V}$ 上的自伴随线性算符 $\mathbf{T}$，$\mathscr{V}$ 中总存在一组规范正交基是 $\mathbf{T}$ 的特征向量。

证明. 数学归纳法。记 $n=\dim\mathscr{V}$。设 $\mathbf{a}_1$ 是 $\mathbf{T}$ 的一个（非零）特征向量（用到了定理2.32“3”），则易验单位向量 $\hat{\mathbf{a}}_1=\mathbf{a}_1/\|\mathbf{a}\|$ 与 $\mathbf{a}_1$ 同属于一个特征空间，故当 $n=1$ 时命题成立。

假定当维数大于 1 且小于 n 时命题成立，令 $\mathscr{W}$ 为由 $\hat{\mathbf{a}}_1$ 线性生成的 1 维子空间，则 $\hat{\mathbf{a}}_1$ 是 $\mathbf{T}$ 的一个特征向量 $\Leftrightarrow\mathscr{W}$ 在 $\mathbf{T}$ 的作用下不变。易验此时 $\mathscr{W}^\perp$ 在 $\mathbf{T}^*$ 的作用下不变。而 $\mathscr{W}^\perp$ 是 $n-1$ 维内积空间。设 $\mathbf{U}$ 是 $\mathbf{T}$ 限制在 $\mathscr{W}^\perp$ 上的线性算符，易验 $\mathbf{U}$ 是 $\mathscr{W}^\perp$ 上的自伴随算符。由数学归纳法假设，$\mathscr{W}^\perp$ 中必有一组规范正效基 $\{\hat{\mathbf{a}}_2,\cdots,\mathbf{a}_n\}$ 是 $\mathbf{U}$ 的特征向量，且易验它们

也都是 $\mathbf{T}$ 的特征向量。由定理A.5，$\mathscr{V}=\mathscr{W}\bigoplus\mathscr{W}^{\perp}$ 且 $\widehat{\mathbf{a}}_1\in\mathscr{W}$，故有 $\widehat{\mathbf{a}}_1$ 与 $\{\widehat{\mathbf{a}}_2,\cdots\widehat{\mathbf{a}}_n\}$ 都正交，即 $\{\widehat{\mathbf{a}}_1,\cdots\widehat{\mathbf{a}}_n\}$ 是一组规范正交基。 □

A.4　欧几里得空间相关证明

我们先引入关于子群的共轭子群的概念。

定义 A.8. 设 G 是一个群，对任意 $a,b\in G$，记其群操作为 $a\circ b$，单位元为 e，a 的逆元为 a^{-1}。若 G 中存在一个 $i\in G$ 使 G 中的两个元素 $a,b\in G$ 满足 $b=i\circ a\circ i^{-1}$，则称 b 为 a 的一个共轭（*conjugate*）。

给定 $i\in G$，设关系 $\sim\subset G\times G$ 满足 $b=i\circ a\circ i^{-1},\forall(a,b)\in\sim$，则易证 $\sim$ 是一个等价关系。给定一个 G 的元素 a，由关系 $\sim$ 可形成一个等价类 $[\![a]\!]_{\sim}$，称 a 的共轭类（*conjugacy class*）。

引理 A.3. 设 G 是一个群，对任意 $a,b\in G$，记其群操作为 $a\circ b$，单位元为 e，a 的逆元为 a^{-1}。若 V 是 G 的一个子群。给定 $i\in G$，定义映射 $Q:\mathscr{V}\to G$,

$$Q(v)=i\circ v\circ i^{-1},\quad\forall v\in V$$

则集合 $V^*\equiv\mathrm{ran}Q$ 是 G 的一个子群，且映射 Q 是由 V 到 V^* 的同构映射。

该引理的证明是直接的，留作练习。

定义 A.9 (共轭子群)**.** 设 G 是一个群，对任意 $a,b\in G$，记其群操作记为 $a\circ b$，单位元为 e，a 的逆元为 a^{-1}。若 V 是 G 的一个子群。给定 $i\in G$，定义映射 $Q:\mathscr{V}\to G$,

$$Q(v)=i\circ v\circ i^{-1}$$

则称 $\mathscr{V}^*$ 是 $\mathscr{V}$ 的一个共轭子群。

引理 A.4. 设 $(\mathscr{E},d)$ 是一个度量空间，若其上的等距群存在一个满足条件 G1 至 G6、S1 至 S4 和 N1 的子群 $\mathscr{V}$，则这样的子群只有一个。

证明. 设 $\mathscr{V}'$ 也是满足上述条件的子群，则给定任意两点 $X,Y\in\mathscr{E}$，$\mathscr{V}$ 中必存在唯一一个等距变换 $\mathbf{u}$、$\mathscr{V}'$ 中必存在唯一一个等距变换 $\mathbf{u}'$，分别满足 $\mathbf{u}=Y-X,\mathbf{u}'=Y-'X$，其中 $-$ 和 $-'$ 分别是由 $\mathscr{V}$ 和 $\mathscr{V}'$ 的传递性所定义的记号。相应地有 $Y=X+\mathbf{u}=X+'\mathbf{u}'$。

定义由 $\mathscr{V}$ 到 $\mathscr{V}'$ 的映射 $f:\mathscr{V}\to\mathscr{V}'$,

$$X+\mathbf{u}=X+'f(\mathbf{u}),\quad\forall X\in\mathscr{E},\forall\mathbf{u}\in\mathscr{V}$$

易证 f 是双射且 $f(\mathbf{0})=\mathbf{0}$，其中 $\mathbf{0}$ 是 $\mathscr{E}$ 上的恒等变换，故它同时是 $\mathscr{V}$ 和 $\mathscr{V}'$ 的单位元。

f 是由 $\mathscr{V}$ 到 $\mathscr{V}'$ 的同构映射：对任意 $\mathbf{u},\mathbf{v}\in\mathscr{V}$ 和 $X_0\in\mathscr{V}$ 必存在 $X,Y\in\mathscr{V}$ 满足

$$X=X_0+\mathbf{u},Y=X_0+\mathbf{v}=X+\mathbf{v}-\mathbf{u}$$

若记 $\mathbf{u}'=f(\mathbf{u}),\mathbf{v}'=f(\mathbf{v})$，则亦有

$$X=X_0+'\mathbf{u}',Y=X+'\mathbf{u}'-\mathbf{v}'$$

由 $\mathscr{V}$ 和 $\mathscr{V}'$ 上的范的定义方式都来自同一度量 d，故有

$$\|\mathbf{v}-\mathbf{u}\|=\|\mathbf{v}'-\mathbf{u}'\|$$

特别地，当 $\mathbf{u}=\mathbf{0}$ 时，$\mathbf{u}'=\mathbf{0}$，上式说明 $\|\mathbf{v}\|=\|\mathbf{v}'\|$ 对任意 $\mathbf{v}\in\mathscr{V}$ 成立。故

$$\begin{aligned}
&\|\mathbf{v}-\mathbf{u}\|=\|\mathbf{v}'-\mathbf{u}'\|\\
\Rightarrow\ &\|\mathbf{v}-\mathbf{u}\|^2=\|\mathbf{v}'-\mathbf{u}'\|^2\\
\Leftrightarrow\ &\|\mathbf{u}\|^2-2(\mathbf{u}|\mathbf{v})+\|\mathbf{v}\|^2=\|\mathbf{u}'\|^2-2(\mathbf{u}'|\mathbf{v}')+\|\mathbf{v}'\|^2\\
\Leftrightarrow\ &(\mathbf{u}|\mathbf{v})=(\mathbf{u}'|\mathbf{v}')
\end{aligned}$$

即 f 保持内积。由类似引理2.2的证明过程可知 f 是由 $\mathscr{V}$ 到 $\mathscr{V}'$ 的同构线性变换。

由外延公理，如果两个集合 A 和 B 之间存在一个双映射 f 满足 $f(x)=x,\forall x\in A$ 则 $A=B$。对任意 $X\in\mathscr{E}$ 和 $\mathbf{v}\in\mathscr{V}$，设 $\mathbf{u}=X-X_0,X_0\in\mathscr{E}$，则

$$\begin{aligned}
X+\mathbf{v}&=X_0+\mathbf{u}+\mathbf{v}\\
&=X_0+'f(\mathbf{u}+\mathbf{v})\\
&=X_0+'f(\mathbf{u})+f(\mathbf{v})\\
&=X_0+\mathbf{u}+'f(\mathbf{v})\\
&=X+f(\mathbf{v})
\end{aligned}$$

均成立，故 $f(\mathbf{v})=\mathbf{v},\forall\mathbf{v}\in\mathscr{V}$，即 $\mathscr{V}'$ 与 $\mathscr{V}$ 作为内积空间是同一的。 □

定理 (等距变换的表示定理3.2). 设 $\mathscr{E}$ 是一个欧几里得空间，$\mathscr{V}$ 是其平移空间，选定任一点 $X_0\in\mathscr{E}$，则 $\mathscr{E}$ 上的任一等距变换 $i:\mathscr{E}\to\mathscr{E},i\in\mathscr{V}$ 都可表示为

$$i(X)=i(X_0)+\mathbf{Q}_i(X-X_0)$$

其中 $\mathbf{Q}_i$ 是一个正交算符，由 i 唯一确定。

证明. 由引理A.3、A.4和2.2，给定任一 $\mathscr{E}$ 上的等距变换 i，$\mathscr{V}$ 的共轭子群都是它自己，且每个 i 引出的共轭映射 $\mathbf{Q}_i\mathbf{v} = i \circ \mathbf{v} \circ i^{-1}$ 就是 $\mathscr{V}$ 上的自同构映射，故 $\mathbf{Q}$ 是 $\mathscr{V}$ 上的正交算符。

注意到 $i \circ \mathbf{v} = \mathbf{Q}_i\mathbf{v} \circ i$，则对任一 $X_0 \in \mathscr{E}$，有

$$i \circ \mathbf{v}\left(X_0\right) = i\left(X + 0 + \mathbf{v}\right), \quad \mathbf{Q}_i\mathbf{v} \circ i\left(X_0\right) = i\left(X_0\right) + \mathbf{Q}_i\mathbf{v}$$

令 $X = X_0 + \mathbf{v}$，则有

$$i\left(X\right) = i\left(X_0\right) + \mathbf{Q}_i\left(X - X_0\right)$$

$\mathbf{Q}_i$ 由 i 唯一确定：设另有一正交算符 $\mathbf{P} : \mathscr{V} \to \mathscr{V}$ 满足 $\mathbf{P}\mathbf{u} = i\left(X + \mathbf{u}\right) - i\left(X\right), \forall \mathbf{u} \in \mathscr{V}, \forall X \in \mathscr{E}$，则 $\left(\mathbf{P} - \mathbf{Q}_i\right)\mathbf{u} = \mathbf{P}\mathbf{u} - \mathbf{Q}_i\mathbf{u} = \mathbf{0}, \forall \mathbf{u} \in \mathscr{V}$。 □

附录 B　向量函数微积分部分的证明

B.1　实空间上的一些拓扑概念

本附录的内容并不新。在高等数学中我们已经学过邻域、去心邻域、内点、外点、边界点、聚点、孤立点、闭集、开集、连通集、区域、有界区域等概念[9]§7.1，但当时学习的定义仅限于 $\mathbb{R}^2$ 的情况。

定义 B.1 (开集). 如果对于 $\mathbb{R}^n$ 的子集 S 中一个元素 $\mathbf{x}_0 \in S$，存在正实数 $\delta > 0$ 使得只要 $\|\mathbf{x} - \mathbf{x}_0\| < \delta$ 则 $\mathbf{x} \in S$，就称$\mathbf{x}_0$ *在 S 内*。所有这样的点 $\mathbf{x}_0$ 的集合称为集合 S 的*内部*（*interior*），记为 $\mathrm{int}S$。如果 S 的所有元素都在 S 内（$S = \mathrm{int}S$），就称 S 是*开集*（*open set*）。一个含有某元素 $\mathbf{x}_0$ 的开集 S 又可称为该点 $\mathbf{x}_0$ 的一个*邻域*（*neighborhood*）。

由定义，整个 $\mathbb{R}^n$ 是一个开集。$\mathbb{R}^n$ 中的开集的交集也是开集。$\mathbb{R}^n$ 中的有限个开集的并集也是开集。空集不含任何元素，故命题“空集是一个开集”*虚真*（*vacuously true*）。这些结论都需要证明但此略。

定义 B.2 (闭集). 如果集合 $S \subset \mathbb{R}^n$ 中的点 $\mathbf{x}_0$ 的每个邻域都含有至少一个 S 中的点（可以就是点 $\mathbf{x}_0$ 本身），则称点 $\mathbf{x}_0$ 是*S 的闭包中的点*，所有这样的点 $\mathbf{x}_0$ 的集合称 S 的*闭包*（*closure*），记为 $\mathrm{cl}S$。如果点 $\mathbf{x}_0$ 的每个邻域都含有至少一个 S 中的与 $\mathbf{x}_0$ 不同的点 $\mathbf{x}$，则称 $\mathbf{x}_0$ 是 S 的一个*极限点*（*limit point*）。S 的所有极限点的集合称 S 的*导集*（*derived set*）。如果集合 S 包含它的所有极限点，则称集合 S 是*闭集*（*closed set*）。

闭集有若干个等价定义。由定义B.2，可证——

定理 B.1. 集合 $\mathbf{S} \in \mathbb{R}^n$ 是闭集当且仅当：

1. 集合 S 等于其闭包（$S = \mathrm{cl}S$）；
2. 集合 S 相对于 $\mathbb{R}^n$ 的补集 $\mathbb{R}^n \setminus S$ 是开集；
3. 集合 S 包含它的所有边界上的点（见定义B.3）（$\partial S \subset S$）。

推论 B.1.1. 一个集合是开集当且仅当它不包含其任何边界上的点。

由定义B.2，整个 $\mathbb{R}^n$ 是一个闭集。$\mathbb{R}^n$ 中的闭集的交集也是闭集。$\mathbb{R}^n$ 中的有限个闭集的并集也是闭集。命题“空集是一个闭集”虚真。这些结论都需要证明但此略。

定义 B.3 (边界). 如果对于 $\mathbb{R}^n$ 的子集 S 中的一个元素 $\mathbf{x}_0 \in S$ 和任意正实数 $\delta > 0$，都存在至少一个 $\mathbf{x} \in S$ 满足 $\|\mathbf{x} - \mathbf{x}\| = \delta$（显然 $\mathbf{x} \neq \mathbf{x}_0$）和至少一个 $\mathbf{y} \notin S$ 满足 $\|\mathbf{y} - \mathbf{x}\| = \delta$，则称 $\mathbf{x}_0$ 是在 S 的边界上的点（*boundary point*）。所有这样的点 $\mathbf{x}_0$ 的集合称为集合 S 的边界（*boundary*），常记为 ∂S。

定义 B.4 (孤立点). 如果对于点 $\mathbf{x}_0 \in S \subset \mathbf{R}^n$，存在正实数 $\delta > 0$ 使得 $\{\mathbf{x} \in \mathbb{R}^n | \|\mathbf{x} - \mathbf{x}_0\| = \delta\} \cap S = \{\mathbf{x}_0\}$，则称 $\mathbf{x}_0$ 是 S 的一个孤立点。

例 B.1. *设 $S = (0,1] \cup \{2\}$，则 $\mathrm{int}S = (0,1)$，$\partial S = \{0,1,2\}$，S 的所有极限点是 $[0,1]$。2 是 S 的一个孤立点。*

B.2 向量函数可微分的必要条件与充分条件

定理 (必要条件之存在性). 定理4.7：设函数 $\mathbf{f} : \mathbb{R}^n \supset D \to \mathbb{R}^m$ 在点 $\mathbf{x}_0 \in D$ 处可微分，即存在线性变换 $\mathbf{L} \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$ 满足

$$\lim_{\Delta \mathbf{x} \to \mathbf{0}} \frac{\mathbf{f}(\mathbf{x}_0 + \Delta \mathbf{x}) - \mathbf{f}(\mathbf{x}_0) - \mathbf{L}\Delta \mathbf{x}}{\|\Delta \mathbf{x}\|} = \mathbf{0}$$

则 $\mathbf{f}$ 的每个坐标函数在 $\mathbf{x}_0$ 处的每个偏导数

$$\left.\frac{f_i(\mathbf{x})}{\partial x_j}\right|_{\mathbf{x}=\mathbf{x}_0}, i = 1, \cdots, m, j = 1, \cdots, n$$

都存在。若 $\{\widehat{\mathbf{e}}_1, \cdots \widehat{\mathbf{e}}_n\}, \{\widehat{\mathbf{u}}_1, \cdots, \widehat{\mathbf{u}}_m\}$ 分别是 $\mathbb{R}^n, \mathbb{R}^m$ 的标准基，则

$$\mathbf{L}\widehat{\mathbf{e}}_j = \sum_{i=1}^{m} \left(\left.\frac{\partial f_i(\mathbf{x})}{\partial x_j}\right|_{\mathbf{x}=\mathbf{x}_0} \right) \widehat{\mathbf{u}}_i, j = 1, \cdots, n$$

证明. 由于函数 $\mathbf{f}(\mathbf{x})$ 在 $\mathbf{x}_0$ 处可微分，对任一 $j \in \{1, \cdots, n\}$，令 $\Delta \mathbf{x} = t\widehat{\mathbf{e}}_j, t \neq 0, t \in \mathbb{R}$，则有

$$\mathbf{f}(\mathbf{x}_0 + t\widehat{\mathbf{e}}_j) - \mathbf{f}(\mathbf{x}_0) = \mathbf{L}(t\widehat{\mathbf{e}}_j) + \|t\widehat{\mathbf{e}}_j\| \mathbf{z}(t\widehat{\mathbf{e}}_j)$$

其中 $\mathbf{z} : \mathbb{R}^n \to \mathbb{R}^m$ 满足 $\lim_{\mathbf{x} \to \mathbf{0}} \mathbf{z}(\mathbf{x}) = \mathbf{0}$。上式 $\Leftrightarrow$

$$\begin{aligned}
&\lim_{t \to 0} \frac{\mathbf{f}(\mathbf{x}_0 + t\widehat{\mathbf{e}}_j) - \mathbf{f}(\mathbf{x}_0) - t\mathbf{L}\widehat{\mathbf{e}}_j}{t} = \lim_{t \to 0} t\mathbf{z}(t\widehat{\mathbf{e}}_j) = \mathbf{0} \\
\Leftrightarrow &\lim_{t \to 0} \frac{\mathbf{f}(\mathbf{x}_0 + t\widehat{\mathbf{e}}_j) - f_i(\mathbf{x}_0)}{t} = \mathbf{L}\widehat{\mathbf{e}}_j \\
\Leftrightarrow &\lim_{t \to 0} \sum_{i=1}^{n} \frac{f_i(\mathbf{x}_0 + t\widehat{\mathbf{e}}_j) - f_i(\mathbf{x}_0)}{t} \widehat{\mathbf{u}}_j = \mathbf{L}\widehat{\mathbf{e}}_j \\
\Leftrightarrow &\lim_{t \to 0} \sum_{i=1}^{n} \left.\frac{\partial f_i(\mathbf{x})}{\partial x_j}\right|_{\mathbf{x}=\mathbf{x}_0} \widehat{\mathbf{u}}_j = \mathbf{L}\widehat{\mathbf{e}}_j \quad \text{（偏导数的定义4.11。）}
\end{aligned}$$

上式对每一 $j \in \{1, \cdots, n\}$ 都成立。 □

定理 (必要条件之唯一性). 定理4.8：设函数 $\mathbf{f} : \mathbb{R}^n \supset D \to \mathbb{R}^m$ 在点 $\mathbf{x}_0 \in D$ 处可微分，即存在线性变换 $\mathbf{L} \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$ 满足

$$\lim_{\Delta\mathbf{x}\to\mathbf{0}} \frac{\mathbf{f}(\mathbf{x}_0 + \Delta\mathbf{x}) - \mathbf{f}(\mathbf{x}_0) - \mathbf{L}\Delta\mathbf{x}}{\|\Delta\mathbf{x}\|} = \mathbf{0}$$

则 $\mathbf{L}$ 是唯一的。

证明. 设线性变换 $\mathbf{L}' \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$ 也满足

$$\lim_{\Delta\mathbf{x}\to\mathbf{0}} \frac{\mathbf{f}(\mathbf{x}_0 + \Delta\mathbf{x}) - \mathbf{f}(\mathbf{x}_0) - \mathbf{L}'\Delta\mathbf{x}}{\|\Delta\mathbf{x}\|} = \mathbf{0}$$

令 $\mathbf{A} = \mathbf{L} - \mathbf{L}'$，则

$$\begin{aligned}
&\mathbf{A}\Delta\mathbf{x} = \mathbf{f}(\mathbf{x} + \Delta\mathbf{x}) - \mathbf{f}(\mathbf{x}) - \mathbf{L}\Delta\mathbf{x} - [\mathbf{f}(\mathbf{x} + \Delta\mathbf{x}) - \mathbf{f}(\mathbf{x}) - \mathbf{L}'(\Delta\mathbf{x})] \\
\Rightarrow\ &\|\mathbf{A}\Delta\mathbf{x}\| \leq \|\mathbf{f}(\mathbf{x} + \Delta\mathbf{x}) - \mathbf{f}(\mathbf{x}) - \mathbf{L}\Delta\mathbf{x}\| + \|\mathbf{f}(\mathbf{x} + \Delta\mathbf{x}) - \mathbf{f}(\mathbf{x}) - \mathbf{L}'\Delta\mathbf{x}\| \quad \text{（三角不等式。）} \\
\Rightarrow\ &\lim_{\Delta\mathbf{x}\to\mathbf{0}} \frac{\|\mathbf{A}\Delta\mathbf{x}\|}{\|\Delta\mathbf{x}\|} = 0 \quad \text{（夹逼定理。）}
\end{aligned}$$

由于 $\Delta\mathbf{x} \to 0$ 的路径不限，不仿考虑固定 $\Delta\mathbf{x} \neq \mathbf{0}$，下式亦成立

$$\lim_{t\to 0} \frac{\|\mathbf{a}(t\Delta\mathbf{x})\|}{|t\Delta\mathbf{x}\|} = 0 \Leftrightarrow \lim_{t\to 0} \frac{\|\mathbf{A}\Delta\mathbf{x}\|}{\|\Delta\mathbf{x}\|} = 0 \Rightarrow \mathbf{A} = \mathbf{0}$$

□

定理 (充分条件). 定理4.9：若函数 $\mathbf{f} : \mathbb{R}^n \supset D \to \mathbb{R}^m$ 的定义域 D 是开集，偏微分 $\frac{\partial f_i}{\partial x_j}, i = 1, \cdots, n, j = 1, \cdots, m$ 在 D 内都连续，则 $\mathbf{f}$ 在 D 内均可微分。

证明. 设 $\mathbf{x} = (b_1, \cdots, b_n)^\intercal, \mathbf{x}_0 = (a_1, \cdots, a_n)^\intercal \in D$，令 $\mathbf{y}_k = (b_1, \cdots, b_k, a_{k+1}, \cdots, a_n), k = 0, \cdots, n$，则有 $\mathbf{y}_0 = \mathbf{x}_0, \mathbf{y}_n = \mathbf{x}$，且 $\|\mathbf{y}_k - \mathbf{y}_{k-1}\| = |b_k - a_k|, k = 1, \cdots, n$。故有 $\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0) = \sum_{k=1}^{n} (\mathbf{f}(\mathbf{y}_k) - \mathbf{f}(\mathbf{y}_{k-1})), k = 1, \cdots, n$。考察上式等号右边的求和项：

$$\mathbf{f}(\mathbf{y}_k) - \mathbf{f}(\mathbf{y}_{k-1}) = \sum_{j=1}^{m} (f_j(\mathbf{y}_k) - f_j(\mathbf{y}_{k-1}))\,\widehat{\mathbf{e}}_j$$

其中 $\{\widehat{\mathbf{e}}_i\}$ 是 $\mathbb{R}^n$ 的标准基。注意到点 $\mathbf{y}_k$ 与 $\mathbf{y}_{k-1}$ 之间的连线是长度为 $|b_k - a_k|$、方向与第 k 个坐标轴 $\widehat{\mathbf{e}}_k$ 平行的有向线段。

对上式右边的坐标函数应用微分中值定理。由命题条件，坐标函数的偏导数 $\frac{\partial f_i}{\partial x_j}$ 在 D 内都连续，则坐标函数 f_i 本身在 D 内也连续，在 $\mathbf{y}_k$ 与 $\mathbf{y}_{k-1}$ 连线上必存在一点 c_k 使得

$$\frac{f_i(\mathbf{y}_k)-f_i(\mathbf{y}_{k-1})}{b_k-a_k}=\left.\frac{\partial f_i}{\partial x_k}\right|_{x_k=c_k}$$

代入上一个求和式得：

$$\begin{aligned}\mathbf{f}(\mathbf{y}_k)-\mathbf{f}(\mathbf{y}_{k-1})&=\sum_{j=1}^{m}\left.\frac{\partial f_j}{\partial x_k}\right|_{x_k=c_k}(b_k-a_k)\,\widehat{\mathbf{e}}_j\\&=(b_k-a_k)\left.\frac{\partial \mathbf{f}}{\partial x_k}\right|_{x_k=c_k},k=1,\cdots,n\end{aligned}$$

再代入上一个求和式得：

$$\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0)=\sum_{k=1}^{n}(b_k-a_k)\left.\frac{\partial \mathbf{f}}{\partial x_k}\right|_{x_k=c_k}$$

令线性变换 $\mathbf{L}:\mathbb{R}^n\to\mathbb{R}^m$ 满足

$$\begin{aligned}(\mathbf{L}(\mathbf{x}-\mathbf{x}_0))&=\left(\left(\left.\frac{\partial \mathbf{f}}{\partial x_1}\right|_{x_1=a_1}\right),\cdots,\left(\left.\frac{\partial \mathbf{f}}{\partial x_1}\right|_{x_n=a_n}\right)\right)(b_1-a_1,\cdots,b_n-a_n)^{\intercal}\\&=\sum_{k=1}^{n}(x_k-a_k)\left.\frac{\partial \mathbf{f}}{\partial x_k}\right|_{x_k=a_k}\end{aligned}$$

则有

$$\begin{aligned}\|\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0)-\mathbf{L}(\mathbf{x}-\mathbf{x}_0)\|&=\left\|\sum_{k=1}^{n}\left(\left(\frac{\partial \mathbf{f}}{\partial x_k}\right)|_{x_k=c_k}-\left.\frac{\partial \mathbf{f}}{\partial x_k}\right|_{x_k=a_k}\right)(x_k-a_k)\right\|\\&\le\sum_{k=1}^{n}\left\|\left.\frac{\partial \mathbf{f}}{\partial x_k}\right|_{x_k=c_k}-\left.\frac{\partial \mathbf{f}}{\partial x_k}\right|_{x_k=a_k}\right\||x_k-a_k|\\&\le\sum_{k=1}^{n}\left\|\left.\frac{\partial \mathbf{f}}{\partial x_k}\right|_{x_k=c_k}-\left.\frac{\partial \mathbf{f}}{\partial x_k}\right|_{x_k=a_k}\right\|\|\mathbf{x}-\mathbf{x}_0\|\end{aligned}$$

由于上述的不等式总成立，则有当 $\mathbf{x}\to\mathbf{x}_0$ 时 $c_k\to a_k$。又由命题条件 $\frac{\partial f_i}{\partial x_j}$ 在 D 都连续，即极限

$$\lim_{x_j\to a_k}\frac{\partial f_j}{\partial x_j}=\left.\frac{\partial f_i}{\partial x_j}\right|_{x_j=a_k},j=1,\cdots,n$$

都存在，故以下极限等式成立：

$$\lim_{\mathbf{x}\to\mathbf{x}_0}\frac{\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0)-\mathbf{L}(\mathbf{x}-\mathbf{x}_0)}{\|\mathbf{x}-\mathbf{x}_0\|}=\mathbf{0}$$

由全微分的定义，命题得证，且线性变换 $\mathbf{L}$ 就是函数的全导数。 □

B.3　复合函数求导的链式法则

定理. 定理4.11：如果函数 $\mathbf{f}:\mathbb{R}^n\supseteq D\to\mathbb{R}^m$ 在 $\mathbf{x}_0\in D$ 处可微分；函数 $\mathbf{g}:\mathbb{R}^n\supset E\to\mathbb{R}^p$ 在 $\mathbf{f}(\mathbf{x}_0)\in E\cap D$ 处可微分，则复合函数 $\mathbf{g}\circ\mathbf{f}$ 在 $\mathbf{x}_0$ 处可微分，且其导数

$$\mathbf{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{g}\circ\mathbf{f}(\mathbf{x})=\mathrm{d}_{\mathbf{y}=\mathbf{f}(\mathbf{x}_0)}\mathbf{g}(\mathbf{y})\,\mathbf{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})$$

证明. 首先证明 $\mathbf{x}_0$ 处于复合函数 $\mathbf{g}\circ\mathbf{f}$ 的定义域内。由于 $\mathbf{f}(\mathbf{x}_0)\in\mathrm{dom}\mathbf{g}$ 且 $\mathbf{g}$ 在 $\mathbf{f}(\mathbf{x}_0)$ 处可微分，故总存在正实数 δ' 使得只要 $\|\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0)\|<\delta'$ 就有 $\mathbf{f}(\mathbf{x})\in\mathrm{dom}\mathbf{g}$。又因为 $\mathbf{x}_0\in\mathrm{dom}\mathbf{f}$ 且 $\mathbf{f}$ 在 $\mathbf{x}_0$ 处可微分，故总存在正实数 δ 使得只要 $\|\mathbf{x}-\mathbf{x}_0\|<\delta$ 则 $\mathbf{x}\in\mathrm{dom}\mathbf{f}$，同时还必存在 $\delta'>0$ 使得这一 δ 选择下的 $\mathbf{x}$ 满足 $\|\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0)\|<\delta'$。所以任一满足 $\|\mathbf{x}-\mathbf{x}_0\|<\delta$ 的 $\mathbf{x}$ 均在复合函数 $\mathbf{g}\circ\mathbf{f}$ 的定义域内。

按照全微分和全导数的定义，由于函数 $\mathbf{f}$ 和 $\mathbf{g}$ 分别在 $\mathbf{x}_0$ 和 $\mathbf{f}(\mathbf{x}_0)$ 可导，故存在函数 $\mathbf{z}_1$、$\mathbf{z}_2$ 满足 $\lim_{\mathbf{x}\to\mathbf{x}_0}\mathbf{z}_1(\mathbf{x}-\mathbf{x}_0)=\mathbf{0}$、$\lim_{\mathbf{f}(\mathbf{x})\to\mathbf{f}(\mathbf{x}_0)}\mathbf{z}_2(\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0))=\mathbf{0}$，且

$$\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0)-\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})(\mathbf{x}-\mathbf{x}_0)=\|\mathbf{x}-\mathbf{x}_0\|\,\mathbf{z}_1(\mathbf{x}-\mathbf{x}_0)$$

$$\mathbf{g}(\mathbf{f}(\mathbf{x}))-\mathbf{g}(\mathbf{f}(\mathbf{x}_0))-\mathrm{d}_{\mathbf{y}=\mathbf{f}(\mathbf{x}_0)}\mathbf{g}(\mathbf{y})(\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0))=\|\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0)\|\,\mathbf{z}_2(\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0))$$

把 $\mathbf{g}(\mathbf{f}(\mathbf{x}))$ 记为 $\mathbf{g}\circ\mathbf{f}(\mathbf{x})$，并把上面的第一条式子代入第二条，得

$$\begin{aligned}
&\mathbf{g}\circ\mathbf{f}(\mathbf{x})-\mathbf{g}\circ\mathbf{f}(\mathbf{x}_0)-\mathrm{d}_{\mathbf{y}=\mathbf{f}(\mathbf{x}_0)}\mathbf{g}(\mathbf{y})\left[\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})(\mathbf{x}-\mathbf{x}_0)+\|\mathbf{x}-\mathbf{x}_0\|\,\mathbf{z}_1(\mathbf{x}-\mathbf{x}_0)\right]\\
=&\|\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0)\|\,\mathbf{z}_2(\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0))\\
\Leftrightarrow&\mathbf{g}\circ\mathbf{f}(\mathbf{x})-\mathbf{g}\circ\mathbf{f}(\mathbf{x}_0)-\mathrm{d}_{\mathbf{y}=\mathbf{f}(\mathbf{x}_0)}\mathbf{g}(\mathbf{y})\,\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})(\mathbf{x}-\mathbf{x}_0)\\
=&\|\mathbf{x}-\mathbf{x}_0\|\,\mathrm{d}_{\mathbf{y}=\mathbf{f}(\mathbf{x}_0)}\mathbf{g}(\mathbf{y})\,\mathbf{z}_1(\mathbf{x}-\mathbf{x}_0)+\|\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0)\|\,\mathbf{z}_2(\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0))
\end{aligned}$$

由三角不等式，又有*

$$\begin{aligned}
\|\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0)\|&=\|\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})(\mathbf{x}-\mathbf{x}_0)+\|\mathbf{x}-\mathbf{x}_0\|\,\mathbf{z}_1(\mathbf{x}-\mathbf{x}_0)\|\\
&\le\|\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})(\mathbf{x}-\mathbf{x}_0)\|+\|\mathbf{x}-\mathbf{x}_0\|\,\|\mathbf{z}_1(\mathbf{x}-\mathbf{x}_0)\|\\
&\le k\,\|\mathbf{x}-\mathbf{x}_0\|+\|\mathbf{x}-\mathbf{x}_0\|\,\|\mathbf{z}_1(\mathbf{x}-\mathbf{x}_0)\|
\end{aligned}$$

故

$$\begin{aligned}
&\mathbf{g}\circ\mathbf{f}(\mathbf{x})-\mathbf{g}\circ\mathbf{f}(\mathbf{x}_0)-\mathrm{d}_{\mathbf{y}=\mathbf{f}(\mathbf{x}_0)}\mathbf{g}(\mathbf{y})\,\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})(\mathbf{x}-\mathbf{x}_0)\\
\le&\|\mathbf{x}-\mathbf{x}_0\|\,\mathrm{d}_{\mathbf{y}=\mathbf{f}(\mathbf{x}_0)}\mathbf{g}(\mathbf{y})\,\mathbf{z}_1(\mathbf{x}-\mathbf{x}_0)+\left(k\,\|\mathbf{x}-\mathbf{x}_0\|+\|\mathbf{x}-\mathbf{x}_0\|\,\|\mathbf{z}_1(\mathbf{x}-\mathbf{x}_0)\|\right)\mathbf{z}_2(\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0))\\
\le&\|\mathbf{x}-\mathbf{x}_0\|\left\{\left\|\mathrm{d}_{\mathbf{y}=\mathbf{f}(\mathbf{x}_0)}\mathbf{g}(\mathbf{y})\,\mathbf{z}_1(\mathbf{x}-\mathbf{x}_0)\right\|+\left(k+\|\mathbf{z}_1(\mathbf{x}-\mathbf{x}_0)\|\right)\mathbf{z}_2(\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0))\right\}
\end{aligned}$$

*此处用到定理4.3。

由于函数 $\mathbf{f}$ 在 $\mathbf{x}_0$ 处连续，即极限 $\lim_{\mathbf{x}\to\mathbf{x}_0}\mathbf{f}(\mathbf{x})=\mathbf{f}(\mathbf{x}_0)$，故上式最后的大括号在 $\mathbf{x}\to\mathbf{x}_0$ 时趋于 $\mathbf{0}$。按照全微分和全导数的定义，命题得证。 □

B.4 反函数定理和隐函数定理

引理 B.1. 线性变换 $\mathbf{L}:\mathscr{V}_n\to\mathscr{W}_m$ 是单射当且仅当存在正实数 $p>0$ 满足 $\|\mathbf{Lx}\|\geq p\,\|\mathbf{x}\|\,\forall\mathbf{x}\in\mathscr{V}_n$。

证明. 如果 $\mathbf{L}$ 不是单射，则存在 $\mathbf{x}_0\neq\mathbf{0}$ 满足 $\mathbf{Lx}_0=\mathbf{0}$，即有 $\|\mathbf{Lx}_0\|=0<m\,\|\mathbf{x}\|$。故其逆否命题成立。

如果 $\mathbf{L}$ 是单射，则其存在逆 $\mathbf{L}^{-1}$ 满足 $\mathbf{L}^{-1}\mathbf{L}=\mathbf{I}_{\mathscr{V}}$ 且 $\mathbf{L}^{-1}$ 也是线性变换。由定理4.3，$\exists k>0,\|\mathbf{L}^{-1}\mathbf{y}\|\leq k\,\|\mathbf{y}\|\,\forall\mathbf{y}\in\mathscr{W}_m$。令 $p=1/k$ 则有 $p\,\|\mathbf{x}\|=m\,\|\mathbf{L}^{-1}\mathbf{Lx}\|\leq pk\,\|\mathbf{Lx}\|=\|\mathbf{Lx}\|$ □

引理 B.2. 设 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$ 是可微函数，且在 $\mathbf{x}_0$ 处连续可微。再假设 $\mathbf{f}$ 在 $\mathbf{x}_0$ 处的导数 $\left.\frac{d\mathbf{f}(\mathbf{x})}{dx}\right|_{\mathbf{x}=\mathbf{x}_0}$ 是单射。则存在 $\delta>0$ 和 $M>0$ 使得只要 $\|\mathbf{x}-\mathbf{x}_0\|<\delta$ 就有

$$\|(\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x}))\,\mathbf{y}\|\geq M\,\|\mathbf{y}\|\,,\quad\forall\mathbf{y}\in\mathbb{R}^n$$

证明. 记函数 $\mathbf{f}$ 的导函数为 $\mathbf{L}$。由引理B.1，存在 $p>0$ 使得 $\|\mathbf{L}(\mathbf{x}_0)\,\mathbf{y}\|\geq p\,\|\mathbf{y}\|\,,\forall\mathbf{y}\in\mathbb{R}^n$。同时，由于 $\mathbf{f}$ 在 $\mathbf{x}_0$ 处连续可微，故对任一正实数——此处选择 $p/2>0$——存在 $\delta>0$ 使得只要 $\|\mathbf{x}-\mathbf{x}_0\|<\delta$ 就有 $\|\mathbf{L}(\mathbf{x})-\mathbf{L}(\mathbf{x}_0)\|\leq p/2$。由线性变换的范的定义，有不等式

$$\|(\mathbf{L}(\mathbf{x})-\mathbf{L}(\mathbf{x}_0))\,\mathbf{y}\|\leq\|\mathbf{L}(\mathbf{x})-\mathbf{L}(\mathbf{x}_0)\|\,\|\mathbf{y}\|\leq\frac{p}{2}\,\|\mathbf{y}\|$$

由三角不等式又有

$$\begin{aligned}\|\mathbf{L}(\mathbf{x}_0)\,\mathbf{y}\|&=\|\mathbf{L}(\mathbf{x}_0)\,\mathbf{y}+\mathbf{L}(\mathbf{x})\,\mathbf{y}-\mathbf{L}(\mathbf{x})\,\mathbf{y}\|\\&\leq\|\mathbf{L}(\mathbf{x}_0)\,\mathbf{y}-\mathbf{L}(\mathbf{x})\,\mathbf{y}\|+\|\mathbf{L}(\mathbf{x})\,\mathbf{y}\|\\\Leftrightarrow\|\mathbf{L}(\mathbf{x}_0)\,\mathbf{y}\|-\|\mathbf{L}(\mathbf{x})\,\mathbf{y}-\mathbf{L}(\mathbf{x}_0)\,\mathbf{y}\|&\leq\|\mathbf{L}(\mathbf{x})\,\mathbf{y}\|\end{aligned}$$

上式不等号左边可代入刚刚确定的结论：$\|\mathbf{L}(\mathbf{x}_0)\,\mathbf{y}\|\leq p\,\|\mathbf{y}\|\,,-\|\mathbf{L}(\mathbf{x})\,\mathbf{y}-\mathbf{L}(\mathbf{x}_0)\,\mathbf{y}\|\geq-\frac{p}{2}$，得

$$\begin{aligned}\|\mathbf{L}(\mathbf{x})\,\mathbf{y}\|&\geq\|\mathbf{L}(\mathbf{x}_0)\,\mathbf{y}\|-\|\mathbf{L}(\mathbf{x})\,\mathbf{y}-\mathbf{L}(\mathbf{x}_0)\,\mathbf{y}\|\\&\geq p\,\|\mathbf{y}\|-\frac{p}{2}\,\|\mathbf{y}\|=\frac{p}{2}\,\|\mathbf{y}\|\end{aligned}$$

故存在 $M=p/2>0$ 满足命题。 □

引理 B.3. 设 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$ 是可微函数，且在 $\mathbf{x}_0$ 处连续可微。再假设 $\mathbf{f}$ 在 $\mathbf{x}_0$ 处的导数 $\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}(\mathbf{x})$ 是单射，则存在正实数 $\delta>0$ 和 $M>0$ 使得只要 $\|\mathbf{x}'-\mathbf{x}_0\|<\delta$ 就有 $\|\mathbf{f}(\mathbf{x}')-\mathbf{f}(\mathbf{x})\|\leq M\|\mathbf{x}'-\mathbf{x}\|$。

证明. 记函数 $\mathbf{f}$ 的导函数为 $\mathbf{L}$。由引理B.1，存在 $m>0$ 使得 $\|\mathbf{L}(\mathbf{x}_0)\mathbf{y}\|\geq m\|\mathbf{y}\|\,\forall\mathbf{y}\in\mathbb{R}^n$。

由于 $\mathbf{f}$ 在 $\mathbf{x}_0$ 处连续可微，故对任一正实数——此处选择 $M=m/(2\sqrt{n})>0$——存在 $\delta>0$ 使得只要 $\|\mathbf{x}-\mathbf{x}_0\|<\delta$ 就有 $\|\mathbf{L}(\mathbf{x})-\mathbf{L}(\mathbf{x}_0)\|\leq m/(2\sqrt{n})$。

按照命题叙述，设 $\mathbf{x}$ 和 $\mathbf{x}'$ 是 $\mathbb{R}^n$ 的任意两向量满足 $\|\mathbf{x}'-\mathbf{x}_0\|<\delta,\|\mathbf{x}-\mathbf{x}_0\|<\delta$，令 $\mathbf{z}=\mathbf{x}'-\mathbf{x}$，则对 $0\leq t\leq 1$ 有

$$\begin{aligned}\|\mathbf{x}+t\mathbf{z}-\mathbf{x}_0\|&=\|t\mathbf{x}'+(1-t)\mathbf{x}-\mathbf{x}_0\|\\&=\|t(\mathbf{x}'-\mathbf{x})+(1-t)(\mathbf{x}-\mathbf{x}_0)\|\\&\leq t\|\mathbf{x}'-\mathbf{x}_0\|+(1-t)\|\mathbf{x}-\mathbf{x}_0\|<t\delta+(1-t)\delta=\delta\end{aligned}$$

上述推导结论在几何上的意义是，只要点 $\mathbf{x}',\mathbf{x}$ 在由 $\|\mathbf{x}_0\|<\delta$ 的开集内部，则它们的连线上的点 $\mathbf{x}+t\mathbf{z}$ 都在此开集内部，或称“δ-球是凸的”。由于导函数连续是在整个 δ-球内都成立的，因此对由 $0\leq t\leq 1$ 定义的所有点 $\mathbf{x}+t\mathbf{z}$ 均有 $\|\mathbf{L}(\mathbf{x}+t\mathbf{z})-\mathbf{L}(\mathbf{x}_0)\|<m/(2\sqrt{n})$。又由线性变换的模的定义有 $\|(\mathbf{L}(\mathbf{x}+t\mathbf{z})-\mathbf{L}(\mathbf{x}_0))\mathbf{y}\|\leq\|\mathbf{L}(\mathbf{x}+t\mathbf{z})-\mathbf{L}(\mathbf{x}_0)|\,\|\mathbf{y}\|<M\|\mathbf{y}\|\,\forall\mathbf{y}\in\mathbb{R}^n$。

引入“取坐标函数”，$\pi_k(\mathbf{x})=x_k,\mathbf{x}=(x_1,\cdots,x_n)\in\mathbf{R}^n,k=1,\cdots,n$。易验证 $\frac{d\pi_k(\mathbf{x})}{dx}\equiv\pi_k(\mathbf{x})$。若定义 $g_k(t)=\pi_k(\mathbf{f}(\mathbf{x}+t\mathbf{z})),0\leq t\leq 1$，则由链式法则可得如下关系

$$\frac{dg_k}{dt}=\pi_k(\mathbf{L}(\mathbf{x}+t\mathbf{z})\mathbf{z})$$

由微分中值定理，存在 $t_k\in[0,1]$ 使得 $g_k(1)-g_k(0)=\frac{dg_k}{dt_k}$。代入 g_k、$\frac{dg_k}{dt_k}$ 的表达式得 $\pi_k(\mathbf{f}(\mathbf{x}'))-\pi_k(\mathbf{f}(\mathbf{x}))=\pi_k(\mathbf{L}(\mathbf{x}+t_k\mathbf{z})\mathbf{z})$。注意到，函数 $\pi_k(\mathbf{x})$ 就是向量 $\mathbf{x}$ 在第 k 个基上的投影长度。由投影长度不大于向量长度（代数意义是使用柯西–施瓦茨不等式），有

$$\begin{aligned}\|\mathbf{f}(\mathbf{x}')-\mathbf{f}(\mathbf{x})\|&\geq|\pi_k(\mathbf{f}(\mathbf{x}')-\mathbf{f}(\mathbf{x}))|\\&=|\pi_k(\mathbf{f}(\mathbf{x}'))-\pi_k(\mathbf{f}(\mathbf{x}))|\\&=|\pi_k(\mathbf{L}(\mathbf{x}+t_k\mathbf{z})\mathbf{z})|\end{aligned}$$

另有以下三角不等式成立：

$$\|(\mathbf{L}(\mathbf{x}+t_k\mathbf{z})-\mathbf{L}(\mathbf{x}_0))\mathbf{z}\|+\|\mathbf{L}(\mathbf{x}_0)\mathbf{z}\|\leq\|\mathbf{L}(\mathbf{x}+t_k\mathbf{z})\mathbf{z}\|$$

上式左右取投影也成立，即

$$\left|\pi_k\left(\left(\mathbf{L}\left(\mathbf{x}+t_k\mathbf{z}\right)-\mathbf{L}\left(\mathbf{x}_0\right)\right)\mathbf{z}\right)\right|+\left|\pi_k\left(\mathbf{L}\left(\mathbf{x}_0\right)\mathbf{z}\right)\right|\leq\left|\pi_k\left(\mathbf{L}\left(\mathbf{x}+t_k\mathbf{z}\right)\mathbf{z}\right)\right|$$

以上不等式联合有

$$\left\|\mathbf{f}\left(\mathbf{x}'\right)-\mathbf{f}\left(\mathbf{x}\right)\right\|\geq\left|\pi_k\left(\left(\mathbf{L}\left(\mathbf{x}+t_k\mathbf{z}\right)-\mathbf{L}\left(\mathbf{x}_0\right)\right)\mathbf{z}\right)\right|+\left|\pi_k\left(\mathbf{L}\left(\mathbf{x}_0\right)\mathbf{z}\right)\right|$$

由事实 $\|\mathbf{x}\|\leq\sqrt{n}\max\left\{|x_i|\right\}\equiv\sqrt{n}\max\left\{\pi_i\left(\mathbf{x}\right)\right\}$（之前在说明范的定义的等价性时证明过该事实）知，在 $k=1,\cdots,m$ 中至少有一个 k 满足

$$\sqrt{n}\left|\pi_k\left(\mathbf{L}\left(\mathbf{x}_0\right)\mathbf{z}\right)\right|\geq\left\|\mathbf{L}\left(\mathbf{x}_0\right)\mathbf{z}\right\|$$

再次利用投影不大于原长，有

$$\left|\pi_k\left(\left(\mathbf{L}\left(\mathbf{x}+t_k\mathbf{z}\right)-\mathbf{L}\left(\mathbf{x}_0\right)\right)\mathbf{z}\right)\right|\leq\left\|\left(\mathbf{L}\left(\mathbf{x}+t_k\mathbf{z}\right)-\mathbf{L}\left(\mathbf{x}_0\right)\right)\mathbf{z}\right\|$$

再次联合这些不等式有

$$\begin{aligned}\left\|\mathbf{f}\left(\mathbf{x}'\right)-\mathbf{f}\left(\mathbf{x}\right)\right\|&\geq\frac{1}{\sqrt{n}}\left\|\mathbf{L}\left(\mathbf{x}_0\right)\mathbf{z}\right\|-\left\|\left(\mathbf{L}\left(\mathbf{x}+t_k\mathbf{z}\right)-\mathbf{L}\left(\mathbf{x}_0\right)\right)\mathbf{z}\right\|\\&\geq 2M\left\|\mathbf{z}\right\|-M\left\|\mathbf{z}\right\|=M\left\|\mathbf{x}'-\mathbf{x}\right\|\end{aligned}$$

□

有了上面三个引理，我们可正式给出反函数定理的证明。

定理 B.2 (反函数定理). 设 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^n$ 是连续可微函数，记函数 $\mathbf{f}$ 的导函数为 $\mathbf{L}\left(\mathbf{x}\right)\equiv\mathrm{d}_{\mathbf{x}=\mathbf{x}_0}\mathbf{f}\left(\mathbf{x}\right)$。若 $\mathbf{L}\left(\mathbf{x}_0\right)$ 是单射，则总存在 $\mathbf{x}_0$ 的一个邻域 N 使得 $\mathbf{f}$ 在 N 上有连续可导的逆函数 $\mathbf{f}^{-1}$；$\mathbf{f}$ 的像的集合 $\mathbf{f}(N)$ 也是开集；对 N 内任意一点 $\mathbf{x}$ 都有

$$\mathrm{d}_{\mathbf{y}=\mathbf{f}(\mathbf{x})}\mathbf{f}^{-1}\left(\mathbf{y}\right)=\mathbf{L}^{-1}\left(\mathbf{x}\right)$$

其中 $\mathbf{y}=\mathbf{f}\left(\mathbf{x}\right)$。

证明. 我们先列出引理B.2和B.3的结论。由于 $\mathbf{f}$ 在 $\mathbf{x}_0$ 处连续可微，且 $\mathbf{L}\left(\mathbf{x}_0\right)$ 是单射，故：

- 由引理B.2，对 $\mathbf{x}_0$ 的任一邻域 $N=\left\{\mathbf{x}\middle|\,\|\mathbf{x}-\mathbf{x}_0\|<\delta\right\}$（$\delta>0$ 为任一正实数），都能找到正实数 $M\left(\mathbf{y}\right)>0$ 满足 $\left\|\mathbf{L}\left(\mathbf{x}\right)\mathbf{y}\right\|\geq M\left\|\mathbf{y}\right\|\,\forall\mathbf{y}\in\mathbb{R}^n$。进一步地，再由引理B.1可知导函数 $\mathbf{L}\left(\mathbf{x}\right)$ 在 N 上的每个值都是单射线性变换。再由于 $\mathbf{L}\left(\mathbf{x}\right)$ 在 N 上都是单射线性变换且其定义域和陪域维数相同，故 $\mathbf{L}\left(\mathbf{x}\right)$ 在 N 上的每个值都是双射（同构）线性变换。

- 由引理B.3，对 N 内部任一 $\mathbf{x}'$，总能找到正实数 $M'(\mathbf{x}') > 0$ 满足 $\|\mathbf{f}(\mathbf{x}') - \mathbf{f}(\mathbf{x})\| \geq M'\|\mathbf{x}' - \mathbf{x}\|$。

我们令 $M' = M$，这相当于联系了 $\mathbf{y}$ 和 $\mathbf{x}'$。

我们证明的任务包括：

I 函数 $\mathbf{f}$ 存在逆函数 $\mathbf{f}^{-1}$；

II 开集 N 经 $\mathbf{f}$ 的像集 $\mathbf{f}(N)$ 也是开集；

III $\forall \mathbf{x} \in N, \mathbf{L}^{-1}(\mathbf{x})$ 是 $\mathbf{f}^{-1}$ 的导数；

IV $\mathbf{f}^{-1}$ 连续可微。

I的证明：由引理B.3的结论，若 $\mathbf{x}' \neq \mathbf{x}$ 则 $\mathbf{f}(x) \neq \mathbf{f}(\mathbf{x}')$，即 $\mathbf{f}$ 是单射，故必存在逆 $\mathbf{f}^{-1}$。I证毕。

II的证明：首先我们确认一些比较直接的接论：

- 由于 N 是开集，故对任一 $\mathbf{x}_1 \in N$，总能找到足够小的 δ_1 使得 $B = \{\mathbf{x} | \|\mathbf{x} - \mathbf{x}_1\| \leq \delta_1\}$ 在 N 的内部。注意这里的 B 是一个闭集。
- 由于函数 $\mathbf{f}$ 存在逆函数 $\mathbf{f}^{-1}$，故 $\mathbf{x} \in N \Leftrightarrow N \ni \mathbf{x} = \mathbf{f}^{-1}(\mathbf{f}(\mathbf{x}))\mathbf{f}^{-1}(\mathbf{y}) \forall \mathbf{y} \in \mathbf{f}(N) = \{\mathbf{y} | \mathbf{y} = \mathbf{f}(\mathbf{x}), \mathbf{x} \in N\}$。即给定任一 $\mathbf{y}_1 \in \mathbf{f}(N)$ 有且只有一个 $\mathbf{x}_1 \in N$ 满足 $\mathbf{f}(\mathbf{x}_1) = \mathbf{y}_1$。

要证明 $\mathbf{f}(N)$ 是开集，就是要证明，对任一 $\mathbf{y}_1 \in \mathbf{f}(N)$，总能找到足够小的 $\widetilde{M} > 0$ 使得开集 $C = \left\{\mathbf{y} | \|\mathbf{y} - \mathbf{y}_1\| < \widetilde{M}\right\}$ 在 $\mathbf{f}(N)$ 的内部。

如何由已知条件来找到这个 $\widetilde{M}$ 呢？由于 N 是开集，我们通过 $\mathbf{x}_1 = \mathbf{f}^{-1}(\mathbf{y}_1) \in N$，可以找到使得闭集 $B = \{\mathbf{x} | \|\mathbf{x} - \mathbf{x}_1\| \leq \delta_1\}$ 在 N 的内部的一个正实数 δ_1。

如果 $\widetilde{M}$ 存在，则对任一 $\mathbf{y} \in C$，我们可以从 B 中找到一个 $\mathbf{x}'$ 使得 $\mathbf{y}' = \mathbf{f}(\mathbf{x}')$ 到 $\mathbf{y} \in C$ 的距离最短，并由引理B.3，总能找到足够小的正实数 M' 使得

$$\|\mathbf{f}(\mathbf{x}') - \mathbf{y}_1\| = \|\mathbf{f}(\mathbf{x}') - \mathbf{f}(\mathbf{x}_1)\| \geq M'\|\mathbf{x}' - \mathbf{x}_1\| = M'\delta_1$$

接下来我们将证明：

i 如果 $\widetilde{M} = M'\delta/2$，那么上述的 $\mathbf{x}'$ 在 B 的内部（即不在 B 的边界上）；

ii 这一 $\mathbf{y}'$ 就是 $\mathbf{y}$。

上面两条若得证，则给定任一 $\mathbf{y}_1 \in \mathbf{f}(N)$，总有正实数 $\widetilde{M}$（且具体地 $\widetilde{M} = M'\delta_1/2$）使得开集 $C = \left\{\mathbf{y} | \|\mathbf{y} - \mathbf{y}_1\| < \widetilde{M}\right\}$ 在 $\mathbf{f}(N)$ 的内部。II也就得证了。

i的证明：反证法。设 $\mathbf{x}'$ 在 B 的边界上，即 $\|\mathbf{x}' - \mathbf{x}_1\| = \delta_1$，则由引理B.3，总能找足够小的正实数 M' 使得

$$\|\mathbf{f}(\mathbf{x}') - \mathbf{y}_1\| = \|\mathbf{f}(\mathbf{x}') - \mathbf{f}(\mathbf{x}_1)\| \geq M'\|\mathbf{x}' - \mathbf{x}_1\| = M'\delta_1$$

那么，由三角不等式，对任一 $\mathbf{y}\in C$（即总有 $\|\mathbf{y}-\mathbf{y}_1\|<M'\delta_1/2$），

$$
\begin{aligned}
\|\mathbf{f}(\mathbf{x}')-\mathbf{y}\| &\geq \|\mathbf{f}(\mathbf{x}')-\mathbf{y}_1\|-\|\mathbf{y}-\mathbf{y}_1\|\\
&> M'\delta_1-\|\mathbf{y}-\mathbf{y}_1\|\\
&> M'\delta_1-\frac{M'\delta_1}{2}\\
&= \frac{M'\delta_1}{2}\\
&> \|\mathbf{y}-\mathbf{y}_2\|\\
&= \|\mathbf{f}(\mathbf{x}_1)-\mathbf{y}\|
\end{aligned}
$$

但这与“$\mathbf{y}'$ 到 $\mathbf{y}$ 的距离最短”相矛盾，故 $\mathbf{x}'$ 在 B 的内部。

ii的证明：设到 $\mathbf{y}$ 的距离平方函数

$$g(\mathbf{x})=\|\mathbf{f}(\mathbf{x})-\mathbf{y}\|^2=(\mathbf{f}(\mathbf{x})-\mathbf{y})\cdot(\mathbf{f}(\mathbf{x})-\mathbf{y})$$

则 $\mathbf{x}'$ 应使得该函数的一阶导数等于零，即 $\left.\frac{dg(\mathbf{x})}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}'}=\mathbf{0}$（零变换）。由零变换性质和链式法则，对任一 $\mathbf{z}\in\mathbb{R}^n$，

$$0=\left.\frac{dg(\mathbf{x})}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}'}\mathbf{z}=2(\mathbf{f}(\mathbf{x}')-\mathbf{y})\cdot(\mathbf{L}(\mathbf{x}')\mathbf{z})$$

由于 $\mathbf{L}(\mathbf{x})$ 在 N 上的每个值都是双射（同构）线性变换，故有且只有一个向量 $\mathbf{z}\in\mathbb{R}^n$ 满足 $\mathbf{L}(\mathbf{x}')\mathbf{z}=\mathbf{f}(\mathbf{x}')-\mathbf{y}$。故上式 $\Leftrightarrow$

$$0=2(\mathbf{f}(\mathbf{x}')-\mathbf{y})\cdot(\mathbf{f}(\mathbf{x}')-\mathbf{y})\Leftrightarrow\mathbf{f}(\mathbf{x}')-\mathbf{y}=\mathbf{0}$$

即，只要 $\mathbf{x}'\in N$ 是使 $\mathbf{y}'=\mathbf{f}(\mathbf{x}')$ 到任一 $\mathbf{y}\in\mathbf{f}(N)$ 的距离最短的点，则 $\mathbf{f}(\mathbf{x}')=\mathbf{y}'=\mathbf{y}$。ii证毕。II证毕。

III的证明：按照导数的定义，相当于要证明对任一 $\mathbf{x}\in N$，极限

$$\lim_{\mathbf{f}(\mathbf{x}')\to\mathbf{f}(\mathbf{x})}\frac{\mathbf{x}'-\mathbf{x}-\mathbf{L}^{-1}(\mathbf{x})(\mathbf{f}(\mathbf{x}')-\mathbf{f}(\mathbf{x}))}{\|\mathbf{f}(\mathbf{x}')-\mathbf{f}(\mathbf{x})\|}=\mathbf{0}$$

令未求极限前的比增量为 $\mathbf{s}$，即

$$\mathbf{s}=\frac{\mathbf{x}'-\mathbf{x}-\mathbf{L}^{-1}(\mathbf{x})(\mathbf{f}(\mathbf{x}')-\mathbf{f}(\mathbf{x}))}{\|\mathbf{f}(\mathbf{x}')-\mathbf{f}(\mathbf{x})\|}$$

由于 $\mathbf{L}(\mathbf{x})$ 在 $\mathbf{x}\in N$ 内都有定义，故极限

$$\lim_{\mathbf{x}'\to\mathbf{x}}\frac{\mathbf{f}(\mathbf{x}')-\mathbf{f}(\mathbf{x})-\mathbf{L}(\mathbf{x})(\mathbf{x}'-\mathbf{x})}{\|\mathbf{x}'-\mathbf{x}\|}=\mathbf{0}$$

令

$$\mathbf{r}=\frac{\mathbf{f}(\mathbf{x}')-\mathbf{f}(\mathbf{x})-\mathbf{L}(\mathbf{x})(\mathbf{x}'-\mathbf{x})}{\|\mathbf{x}'-\mathbf{x}\|}$$

则 $\lim_{\mathbf{x}'\to\mathbf{x}}\mathbf{r}=\mathbf{0}$。对 $\mathbf{x}'\in N,\mathbf{x}'\neq\mathbf{x}$，$\mathbf{s}$ 可由 $\mathbf{r}$ 表示为

$$\mathbf{s}=-\frac{\|\mathbf{x}'-\mathbf{x}\|}{\|\mathbf{f}(\mathbf{x}')-\mathbf{f}(\mathbf{x})\|}\mathbf{L}^{-1}(\mathbf{x})\,\mathbf{r}$$

由引理B.3，存在足够小正实数 M 使得 $\|\mathbf{f}(\mathbf{x}')-\mathbf{f}(\mathbf{x})\|\geq M\|\mathbf{x}'-\mathbf{x}\|$，故有

$$0\geq-\frac{\|\mathbf{x}'-\mathbf{x}\|}{\|\mathbf{f}(\mathbf{x}')-\mathbf{f}(\mathbf{x})\|}\geq-\frac{1}{M}$$

即 $-\frac{\|\mathbf{x}'-\mathbf{x}\|}{\|\mathbf{f}(\mathbf{x}')-\mathbf{f}(\mathbf{x})\|}$ 是有界的。

又由定理4.3的推论，线性变换都是连续函数，故由复合函数的连续性，极限

$$\lim_{\mathbf{x}'\to\mathbf{x}}\mathbf{L}^{-1}(\mathbf{x})\,\mathbf{r}=\mathbf{L}^{-1}(\mathbf{x})\lim_{\mathbf{x}'\to\mathbf{x}}\mathbf{r}=\mathbf{0}$$

由于一个有界函数与一个有极限的函数的积的极限等于那个有极限的函数的极限*，故 $\lim_{\mathbf{x}'\to\mathbf{x}}\mathbf{s}=\mathbf{0}$。又由于当 $\mathbf{x}'\to\mathbf{x}$ 时 $\mathbf{f}(\mathbf{x}')\to\mathbf{f}(\mathbf{x})$，由 $\mathbf{s}$ 的形式有 $\lim_{\mathbf{x}'\to\mathbf{x}}\mathbf{s}=\mathbf{0}\Leftrightarrow\lim_{\mathbf{f}(\mathbf{x}')\to\mathbf{f}(\mathbf{x})}\mathbf{s}=\mathbf{0}$。III证毕。

IV的证明：要证 $\mathbf{f}^{-1}$ 的导函数连续，即对任一 $\mathbf{x}_1\in N,\mathbf{y}_1=\mathbf{f}(\mathbf{x}_1)$ 有

$$\lim_{\mathbf{y}\to\mathbf{y}_1}\left.\frac{d\mathbf{f}^{-1}(\mathbf{y})}{d\mathbf{y}}\right|_{\mathbf{y}=\mathbf{y}}=\left.\frac{d\mathbf{f}^{-1}(\mathbf{y})}{d\mathbf{y}}\right|_{\mathbf{y}=\mathbf{y}_1}$$

由III的证明我们已经有

$$\left.\frac{d\mathbf{f}^{-1}(\mathbf{y})}{d\mathbf{y}}\right|_{\mathbf{y}=\mathbf{y}}=\mathbf{L}^{-1}(\mathbf{x}),\mathbf{y}=\mathbf{f}(\mathbf{x}),\forall\mathbf{x}\in N$$

故只需证

$$\lim_{\mathbf{x}\to\mathbf{x}_1}\mathbf{L}^{-1}(\mathbf{x})=\mathbf{L}^{-1}(\mathbf{x}_1)$$

由引理B.3，总存在足够小正实数 M 满足 $\|\mathbf{L}(\mathbf{x})\,\mathbf{y}\|\geq M\|\mathbf{y}\|\,\forall\mathbf{y}\in\mathbb{R}^n$，故令 $\mathbf{z}=\mathbf{L}(x)\,\mathbf{y}$，则 $\|\mathbf{z}\|\geq M\|\mathbf{L}^{-1}(\mathbf{x})\,\mathbf{z}\|$。

由于 $\mathbf{f}$ 是连续可微函数，设 $\mathbf{x}_1\in N$，对任一 $\epsilon'>0$，总存在 $\delta>0$，使得只要 $\|\mathbf{x}-\mathbf{x}_1\|<\delta$ 就有 $\|\mathbf{L}(\mathbf{x})-\mathbf{L}(\mathbf{x}_1)\|<\epsilon'$。具体的，设由 δ 定义的 $\mathbf{x}_1$ 的邻域 $N_1=\{\mathbf{x}|\,\|\mathbf{x}-\mathbf{x}_1\|\}<\delta$ 在 N 的内部，则对任一 $\mathbf{x}\in N_1$，以下不等式成立

$$\begin{aligned}\left\|\left(\mathbf{L}^{-1}(\mathbf{x})-\mathbf{L}^{-1}(\mathbf{x}_1)\right)\mathbf{z}\right\|&=\left\|\mathbf{L}^{-1}(\mathbf{x})\left(\mathbf{L}(\mathbf{x})-\mathbf{L}(\mathbf{x}_1)\right)\mathbf{L}^{-1}(\mathbf{x}_1)\,\mathbf{z}\right\|\\&\leq\frac{1}{M}\left\|\left(\mathbf{L}(\mathbf{x})-\mathbf{L}(\mathbf{x}_1)\right)\mathbf{L}^{-1}(\mathbf{x}_1)\,\mathbf{z}\right\|\\&\leq\frac{1}{M}\|\mathbf{L}(\mathbf{x})-\mathbf{L}(\mathbf{x}_1)\|\left\|\mathbf{L}^{-1}(\mathbf{x}_1)\,\mathbf{z}\right\|\\&\leq\frac{1}{M^2}\|\mathbf{L}(\mathbf{x})-\mathbf{L}(\mathbf{x}_1)\|\,\|\mathbf{z}\|\end{aligned}$$

*这个基本定理可由极限的 $\delta-\epsilon$ 语言证明，很多地方有，此略。

由线性变换的范的定义（最大下界），上述不等式 $\Leftrightarrow$

$$\left\|\mathbf{L}^{-1}(\mathbf{x})-\mathbf{L}^{-1}(\mathbf{x}_1)\right\| \leq \frac{1}{M^2}\|\mathbf{L}(\mathbf{x})-\mathbf{L}(\mathbf{x}_1)\| \leq \frac{\epsilon'}{M^2}$$

令 $\epsilon=\frac{\epsilon'}{M^2}$，我们就有对于任一 $\mathbf{x}_1 \in N$ 和任一 $\epsilon>0$，总有 $\delta>0$ 使得只要 $\|\mathbf{x}-\mathbf{x}_1\|<\delta$ 就有 $\|\mathbf{L}^{-1}(\mathbf{x})-\mathbf{L}^{-1}(\mathbf{x}_1)\|<\epsilon$。具体地，这个 δ 总存在是由于 M 总存在。这相当于说 $\lim_{\mathbf{x}\to\mathbf{x}_1}\mathbf{L}^{-1}(\mathbf{x})=\mathbf{L}^{-1}(\mathbf{x}_1)$，IV证毕。 $\square$

附录 C　曲线坐标系（未完成）

欧几里得空间中的一个点，在任一成立的坐标系下，可唯一对应于一个有序实数三元组，即这一点的坐标。在任一直角坐标系下，一个点的坐标就是它所对应的位置向量在相应的规范正交基下的坐标。点的坐标在不同直角坐标系之间的变换，只需要向量的坐标变换法则（§2.3）。但是我们还可以在同一个欧几里得空间中建立各种曲线坐标系。在任一曲线坐标系下，空间中任何一点仍然唯一对应于一个有序实数三元组作为其曲线坐标。此时直角坐标系只是一个特殊的曲线坐标系。一个点在曲线坐标系之间的坐标变换一般法则，仅按 §2.3的内容是不够的。

向量函数的定义域和值域都是向量的集合。如果这些向量都用于表示欧几里得空间中的物理过程（即它们是物理量），那么它们都必然与欧几里得空间中的点（位置）相联系。当我们的讨论不仅限于向量空间的数学讨论，而是关于在欧几里得空间中发生的物理过程的讨论时，作为向量函数的物理量在不同曲线坐标系下的取值和关系式也将面临不同曲线坐标系下的坐标变换问题。线性变换作为一种特殊的向量函数也不例外。而且不同的物理量的曲线坐标变换法则，还依赖这些物理量的定义，需要逐一讨论。流变学中常见的物理量在柱坐标和球坐标下的坐标变换公式，可以查表解决。

我们在大一的数学课中已经学过二维的极坐标，三维的柱坐标、球坐标。这里就不再重复这些知识了，直接介绍一般的情况。读者可以以 3 维欧几里得空间上的柱坐标或球坐标为例来辅助理解。另外，本节重复用到 §2.3的结论，就不每次都明显地引用了。

参考文献

[1] REINER M. The Deborah Number[J]. Physics Today, 1964, 17(1): 62-62.

[2] COLEMAN B D, MARKOVITZ H, NOLL W. Viscometric Flows of Non-Newtonian Fluids: vol. 5[M]. Springer-Verlag Berlin Heidelberg, 1966.

[3] 周胜林, 刘西民. 线性代数与解析几何[M]. 高等教育出版社, 2012.

[4] HOFFMAN K, KUNZE R. Linear Algebra[M]. 2nd ed. Prentice-Hall, Inc., 1971.

[5] HASSANI S. Mathematical physics : a modern introduction to its foundations[M]. New York: Springer, 1999.

[6] AUDIN M. Geometry[M]. Berlin: Springer, 2002.

[7] BERGER M. Geometry I[M]. Springer-Verlag, 1987.

[8] 王全迪, 郭艾, 杨立洪. 高等数学（上册）[M]. 高等教育出版社, 2009.

[9] 王全迪, 郭艾, 杨立洪. 高等数学（下册）[M]. 高等教育出版社, 2009.

[10] WILLIAMSON R E, CROWELL R H, TROTTER H F. Calculus of Vector Functions[M]. 3rd ed. Prentice Hall, Inc., 1972.

[11] WEATHERALL O, James. Classical Spacetime Structure[M]//KNOX E, WILSON A. The Routledge Companion to Philosophy of Physics. Taylor & Francis, 2022: 41.

[12] 邓文基. 大学物理（上册）[M]. 华南理工大学出版社, 2009.

[13] 邓文基. 大学物理（下册）[M]. 华南理工大学出版社, 2009.

[14] TRUESDELL C, TOUPIN R. The Classical Field Theories[M]//FLÜGGE S. Principles of Classical Mechanics and Field Theory / Prinzipien der Klassischen Mechanik und Feldtheorie. Berlin, Heidelberg: Springer Berlin Heidelberg, 1960: 226-858.

[15] TRUESDELL C, NOLL W. The Non-Linear Field Theories of Mechanics[M]. Ed. by ANTMAN S S. Third. Berlin, Heidelberg: Springer, 2004: 1-579.

[16] NOLL W. A mathematical theory of the mechanical behavior of continuous media[J/OL]. Archive for Rational Mechanics and Analysis, 1958, 2(1): 197-226. http://dx.doi.org/10.1007/BF00277929.

[17] WALTERS K. Rheometry[M]. Chapman, 1975.

[18] De GENNES P G, BIRD R B. Discussion[J/OL]. Physica A: Statistical Mechanics and its Applications, 1983, 118(1): 43-47. https://www.sciencedirect.com/science/article/pii/0378437183901759. DOI: https://doi.org/10.1016/0378-4371(83)90175-9.

[19] FREWER M. More clarity on the concept of material frame-indifference in classical continuum mechanics[J/OL]. Acta Mechanica, 2009, 202(1): 213-246. https://doi.org/10.1007/s00707-008-0028-4. DOI: 10.1007/s00707-008-0028-4.

[20] NOLL W. Lectures on the foundations of continuum mechanics and thermodynamics[J]. Archive for Rational Mechanics and Analysis, 1973, 52(1): 62-92. DOI: 10.1007/BF00249093.