前言

本讲义不是一个完整、深入的流变学教材，仅为了帮助化工类本科专业的人士建立必要的数学基础，为理解其他正规的流变学教材和论文作准备。

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. 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，我假定读者已完整地、熟练地掌握华南理工大学xxx，故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、页码）。希望读者在比较后能够同意，我在此讲义中新介绍的数学知不是完全陌生的。
2. 第一优先介绍第二部分的物理理论基础需要用到的数学概念。例如，在引入时空的概念（第X章）时，需要用到集合的划分与等价类的知识，因此我在集合论的章节中介绍了关系的知识。又例如，流变学中的各种张量，在本讲义已简化成有限维向量空间上的线性变换。按照上一条假定考虑读者的数学基础，仍需要比较详尽地介绍有限维向量空间及其上的线性变换的知识。
3. 第二优先介绍应用物理理论进行计算所需要的数学概念。例如，如果完全不进行任何应用题的计算的话，仅理解向量空间的知识就可以理解流变学理集结中的方程了，因为我在本讲义采用不依赖坐标的表达风格（不介绍也不使用求和约定）。我希望这种做法本身能彰显流变学理论的建立者希望实现的物理客观性，既材料的响应规律（数学表达式）不依赖包括坐标系的选择在内的任何主观选择。但是面临具体应用问题时，常需要建立曲线坐标系。此时向量和线性变换的坐标关系需要曲线坐标系的知识。
4. 仅为了证明某定理所需的数学概念和引理，要么不正式介绍、要么放在附录中介绍。
5. 定理的证明过程，仅供有兴趣的读者参考，故未必都提供。不提供时，我尽可能提出其他提供了证明过程的参考资料。例如：雷诺传输定理的证明（附录XX）、等距变换的表示定理的证明（附录XX）。
6. 本讲义用不到的数学知识，不作介绍。例如，在集合论的章节中提到了部分公理集合论的公理，但又假定关于自然数、算术运算、偏序、全序、数学归纳法等知识为读者已知，包括皮亚诺公设在内的相关的公设就不作介绍了。故集合论的章节的最终状态，深度上似乎要介绍公理集合论，但广度上又不完整。这是我故意为之的。因为本讲义的目标不是要向化学专业的读者大肆介绍数学知识，而是补充看懂流变学理论的最少必要的数学

知识。

一般情况下，概念的定义仅通过字体的改变来暗示。例如，集合（set）是具有某种特性的事物的整体。仅在需要时，定义才以带编号的方式引入。而定理、引理和例子则均带编号。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。不采用带编号式的引入，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，而代价则是使定义失去了引用链接的便利，故在此敬请读者在学习时要不厌其烦地翻阅回顾定义。

关于物理理论基础部分的内容深度和广度，我作了如下考虑：

更多说明将在讲义完成之后再在此解释。

孙尉翔

2020年10月

目录

第一部分　引言

第二部分 数学部分

II.1 集合与映射

II.1.1 集合的基础概念

集合是近代数学的基本语言。用集合与映射重述人的直观经验，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。

定义 II.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。

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B，记为$A = B$[*]。换言之，如果集合A与B的元素都相同，则A与B就是同一个集合。再换言之，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$A \subsetneq B$。

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

[*]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。

四个元素组成的集合B。一般地，我们将此记作：

$$\{x|x \in A \wedge (x\text{需要满足的条件})\}$$

其中符号$\wedge$表示“且”的意思。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；“$\wedge$”后的语句形式是对A的元素x的规定，而不能是其他形式。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$x \in V$。集合V的存在性来自分类公理；V可表示成：

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

V的唯一性简证如下：若另有一$V' = \{x|x \in A \wedge x \in B\}$且$V' \neq V$，则必存在$x \in V'$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$V = A \cap B$。

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

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

其中$\Leftrightarrow$是“当且仅当”的意思。这一集合V的唯一性可类似上一段那样得证，此略†。此时称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）
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。

---

*由集合定义II.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。

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$\mathscr{C}$的至少一个元素的元素[*]。换言之，只要有一个$X\in\mathscr{C}$满足$x\in X$，就有$x\in U$。我们称U是$\mathscr{C}$的所有元素的**并集**（*union*），记作

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

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：

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

集合的并集有如下性质：

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$C=\{x|x\in A且x\notin B\}$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。

[*]$\mathscr{C}$的元素是集合。

[†]至此，我们可以回过头来重新理解集合的定义II.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II.1.1中的“确定性”，只是逻辑上排中律的要求。

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，

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

给定一个集合A，我们规定存在一个集合，暂记作$\mathscr{P}'(A)$，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设给出恰好包括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：

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

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

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

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II.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$(a,b)\subset\mathscr{P}(\mathscr{P}(A\cup B))$。于是我们可以遵守分类公理，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$B\subset Y$

---

*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。

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

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

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

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⇒的意义：（语句1）⇒（语句2）表示“若（语句1 ），则（语句2）。”

II.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，我们没有其他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，而是把所有具有婚姻关系的公民对全部列出来组成一个集合，来作为关于“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)属于关系R，则记为xRy。习惯上，关系常用符号“$\sim$”表示，故更常记为$x\sim y$。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。
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
$$\bigcup_{X\in\bigcup_{X'\in\sim}X'}X\times\bigcup_{X\in\bigcup_{X'\in\sim}X'}X$$
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 II.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$a\sim 1,a\nsim 2$。

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。

*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。

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。

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

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

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

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

为关系$\sim$的值域（*range*），记作$\mathrm{ran} \sim$。不难留意到，对任一关系$\sim$有$\sim \equiv (\mathrm{dom} \sim) \times (\mathrm{ran} \sim)$；对于任一集合上的等于关系，有$(\mathrm{dom} =) = (\mathrm{ran} =)$；对于任一集合上的属于关系，若$(\mathrm{dom} \in) = X$，则$(\mathrm{ran} \in) = \mathscr{P}(X) \setminus \{\emptyset\}$[†]。

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*）。

例 II.1.2. 设$X = \{a, b, c\}$。请验证，X上的等于关系是集合

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

共有*3*个元素；$X \times X$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。

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。

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II.1.1所示）。不难留意到$\mathscr{C} \subset \mathscr{P}(X)$。

[*]式中的记法“$\exists b$, (关于b的语句)表示“存在符合关于b的语句的一个b”。例如“$\exists b, b \in U_\sim$”表示“存在一个属于集合$U_\sim$的b”。

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

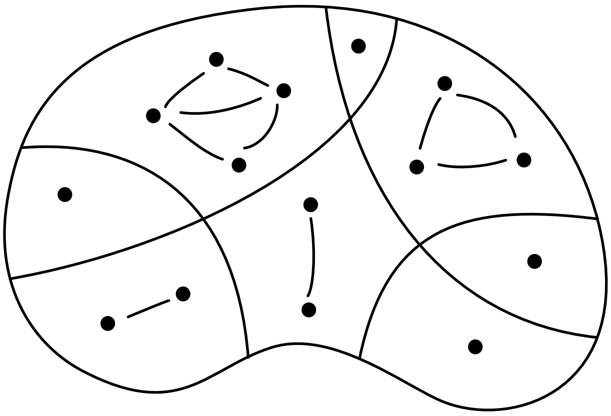

图 II.1.1: 集合的划分示意图

设关系$\sim$是集合X上的一个等价关系，则集合$[\![x]\!]_\sim \equiv \{y|y \in X$且$\exists x \in X, y \sim x\}$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*）‡。仍以图II.1.1为例。圆点表示集合X的元素。在集合X上定义了某关系，圆点之间的连线表示这两个元素之间具有这种关系。由图可见，这一关系是等价关系（连线是双向的）。每个元素，都能通过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。

若集合$\mathscr{C}$是集合X的一个划分，我们可以定义一个关系$\sim\equiv X/\mathscr{C}$§，使得当且仅当X的元素x,y属于$\mathscr{C}$的同一个元素时，$x \sim y$。此时我们称关系$\sim$是由集合X的划分$\mathscr{C}$引出的关系（*equivalent relation induced by the partition $\mathscr{C}$ of X*）。¶ 可以证明，集合X上的任一划分都唯一地引出一个X上的等价关系‖。仍以图II.1.1为例，我们可以让集合X上的这一划分下每个子集内的元素之间都建立关系。这种关系可描述为“X的两个元素同属X的划分的同一个元素”**。易验这种关系是等价关系。因此X的任一划分，都能用这种描述使得每一划分都是X关于

*“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，

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

§该记法与刚刚介绍完的$X/\sim$无关，是符号“/”的滥用。

¶正式地，

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

‖证明过程

**理解此句时注意X的划分是X的子集的集合，其元素是X的子集。

某等价关系的等价类。故我们总是可以直接说$X/\mathscr{C}$是由集合X的划分$\mathscr{C}$引出的等价关系。这一关系的“唯一性”，意思是说，不会有另一个X上的不同的等价关系*。

定理 II.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。

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†。 □

由等价关系基本定理，我们可以写

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

这使得等价关系、划分和商集有类似“集合的除法”的意义。该定理告诉我们，集合上的一个等价关系必定可以划分这个集合。

II.2 向量空间

为一个集合添加一些性质和运算法则，可以形成不同类型的空间，乃至不同的数学分支。在本节中我们介绍其中一种空间——向量空间（又称线性空间[1]）。

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

1. $\mathscr{V}$的元素的加法运算。$\forall\mathbf{a},\mathbf{b},\mathbf{c}\in\mathscr{V}$：
   (a) 封闭性：$\mathbf{a}+\mathbf{b}\in\mathscr{V}$
   (b) 交换律：$\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}$
   (c) 结合律：$(\mathbf{a}+\mathbf{b})+\mathbf{c}=\mathbf{a}+(\mathbf{b}+\mathbf{c})$
   (d) 恒等元素：$\exists\mathbf{0}\in\mathscr{V}:\mathbf{a}+\mathbf{0}=\mathbf{a}$
   (e) 逆：$\forall\mathbf{a}\in\mathscr{V},\exists-\mathbf{a}:\mathbf{a}+(-\mathbf{a})=\mathbf{0}$
2. $\mathbb{F}$与$\mathscr{V}$的元素间的乘法运算。$\forall\alpha,\beta\in\mathbb{F},\mathbf{a},\mathbf{b}\in\mathscr{V}$：
   (a) $\alpha\mathbf{a}\in\mathscr{V}$

*等价关系是集合。故等价关系“相同”即集合的相等；“不同”既集合的不相等。集合的相等已经在上一节定义过。

†证明过程

(b) $\alpha(\beta\mathbf{a}) = (\alpha\beta)\mathbf{a}$

(c) $1\mathbf{a} = \mathbf{a}$

(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）。

在定义II.2.1中，数域$\mathbb{F}$可以是实数域$\mathbb{R}$或复数域$\mathbb{C}$，它是关于标量乘的运算规定中的标量所属的数域。任何非空集合，只要具备定义II.2.1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。

读者可尝试利用定义证明$0\mathbf{a} = \mathbf{0}\forall\mathbf{a}\in\mathscr{V}$。

例 II.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. *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。*

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。*

5. *验证：记$\mathscr{C}^\infty(a,b)$为开区间(a,b)上的所有光滑实值一元函数的集合，它是实数域上的向量空间。*

接下来，我们将逐步发现，定义II.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

*见[1]§2.1矩阵与矩阵的运算

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，$\mathbf{a}_1,\cdots,\mathbf{a}_n\in\mathscr{V}$，$\alpha_1,\cdots,\alpha_n\in\mathbb{F}$，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。由此可引入如下定义。

定义 II.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，$\mathbf{a}_1,\cdots,\mathbf{a}_n\in\mathscr{V}$。若$\alpha_1,\cdots,\alpha_n\in\mathbb{F}$，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II.2.2易得如下结论（证明略）：

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[1]定理3.1、3.2, p. 98。
2. 任何线性无关向量组的子集也是线性无关向量组[1]§7.2 “(2)” ,p. 171。
3. 任何含有$\mathbf{0}$向量的向量组线性相关，因为总有$1\neq0$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。

向量空间定义中的封闭性要求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II.2.1）。

定义 II.2.3 (子空间). 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[1]§7.1定理1.1,p.169。

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 II.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）。

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。

*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$(\mathbf{a}_i)$。

例 II.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 II.2.5 (基). 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。因为按定义，$\mathscr{V}$内可能可以找出不止一组基，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，我们才能通过把这它们的个数统一定义为$\mathscr{V}$的维数，来使得向量空间具有唯一确定的维数。下面的定理解决了这个问题。

定理 II.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[1] “(3)的证明”，p. 171 [2]§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 II.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，记为$\dim\mathscr{V}$，是它的任一组基的向量个数。规定：零向量空间（只有一个零向量组成的向量空间）是0维。

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

推论 II.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[1] “(3)的证明”，p. 171。

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II.2.2。

例 II.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。

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。"[1]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 II.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II.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个标量若只是一个集合，那么这一标量集合变换不同的顺序去与基向量组合，可以表出不同的向量。因此我们必须给基向量规定顺序，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的基向量，而非基向量的一个集合。所以定义II.2.7才特别增加了"有序"的要求。在本讲义中，我们都有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"）。

*这里的唯一性可再次参考[1]"(3)的证明"，p. 171。

在这一定义中，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。沿用定义中的设定，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$\mathbf{x}=(\chi_1,\cdots,\chi_n)$†。

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“转置”的概念进行正式的定义。

II.3 内积空间与赋范向量空间

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。

定义 II.3.1 (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$(\mathbf{a}|\mathbf{b})\,,(\cdot|\cdot):\mathscr{V}\times\mathscr{V}\to\mathbb{F}$，满足‡：$\forall\mathbf{a},\mathbf{b},\mathbf{c}\in\mathscr{V},\alpha\in\mathbb{F}$

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$(\mathbf{a}|\mathbf{0})=0\forall\mathbf{a}\in\mathscr{V}$。因为$\forall\mathbf{a},\mathbf{b}\in\mathscr{V}$均有$(\mathbf{a}|\mathbf{b})=(\mathbf{a}|\mathbf{0}+\mathbf{b})=$

* “利用基和坐标可把线性空间的运算变得更具体。” [1]p.173。

†这里的概念区分可参见[1]例题2.1,p. 173。

‡上划线表示复数共轭

$(\mathbf{a}|\mathbf{0})+(\mathbf{a}|\mathbf{b})\Rightarrow(\mathbf{a}|\mathbf{0})=0$。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，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[3]§1.3.1。

例 II.3.1.

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。

定理 II.3.1. 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：$|(\mathbf{a}|\mathbf{b})|^2\leq(\mathbf{a}|\mathbf{a})\,(\mathbf{b}|\mathbf{b})$
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

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

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 II.3.2 (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$\|\mathbf{x}\|\,,\forall\mathbf{x}\in\mathscr{V}$，满足：

1. 非负性：$\|\mathbf{x}\|\ge 0,\forall\mathbf{x}\in\mathscr{V}$
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：$\|\mathbf{x}+\mathbf{y}\|\le\|\mathbf{x}\|+\|\mathbf{y}\|\,,\forall\mathbf{x},\mathbf{y}\in\mathscr{V}$

带有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$\hat{\mathbf{x}}=\mathbf{x}/\|\mathbf{x}\|$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。

*请验证它满足定义II.3.2。

由于有限维向量空间上定义的不同的范之间是等价的（见§IV.1.1），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 II.3.2. 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，若定义内积：

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

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

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II.3.1，柯西–施瓦茨不等式是对任一内积空间均成立的。由定义II.3.2，三角不等式是对任一赋范向量空间均成立的。而由定理II.3.2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 II.3.3 (正交). 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 II.3.2. *[1] “例题2.2” ,p. 174* *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$E^{pq},p=1,\cdots,n,q=1,\cdots,n$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 II.3.3. 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

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

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

由定理II.3.3以及子空间、线性生成空间的定义（II.2.2、II.2.4），内积空间$\mathscr{V}$的正交集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$\mathscr{V}$的一个规范正交基（orthonormal basis）。

定理II.3.3的证明也给出了格拉姆–施密特正交化过程（Gram–Schmidt process），作为定理如下。

定理 II.3.4. 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II.3.3，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 II.3.4.1. 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 II.3.3. 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$\hat{\mathbf{a}}_1=\frac{1}{5}\mathbf{a}_1,\hat{\mathbf{a}}_2=\frac{1}{5}\mathbf{a}_2,\hat{\mathbf{a}}_3=(0,1,0)^\intercal$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$\{\hat{\mathbf{e}}_i\}$叫标准基（*standard basis*），其中

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

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$G_{ij}=(\mathbf{e}_i|\mathbf{e}_j)$，则由内积定义有$G_{ij}=\overline{G_{ji}}$。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$G_{ij}=\delta_{ij}$。

II.4　线性变换

II.4.1　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。

向量空间之间的映射不一定保留其运算规则，即$f(\alpha\mathbf{a}+\beta\mathbf{b})\neq\alpha f(\mathbf{a})+\beta f(\mathbf{b})$。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，则称这种映射为同态映射（homomorphism）。

例 II.4.1. *以下映射是否同态映射？*

- $f:\mathbb{R}\to\mathbb{R}:f(x)=x^2$
- $g:\mathbb{R}^2\to\mathbb{R}:g(x,y)=x^2+y^2-4$
- $F:\mathbb{R}^2\to\mathbb{C}:F(x,y)=U(x,y)+iV(x,y),U,V:\mathbb{R}^2\to\mathbb{R}$
- $T:\mathbb{R}\to\mathbb{R}^2:T(t)=(t+3,2t-5)$
- *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 II.4.1 (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

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

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：$T(\mathbf{a})=\mathbf{Ta}$。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*。†

易验，任何一个线性变换都总把零向量映射为零向量。

易验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II.2.1定义的向量那般，定义II.4.1定义的线性变换也是抽象概念。例II.2.1中的向量空间上都可以建立线性变换。

例 II.4.2. 验证以下是例*II.2.1*中的$\mathscr{C}^\infty(a,b)$上的线性变换。

1. 一阶导数映射$D:\mathscr{C}^\infty(a,b)\to\mathscr{C}^\infty(a,b),Df(x)=f'(x),\forall f\in\mathscr{C}^\infty(a,b)$
2. 积分上限函数映射$T:\mathscr{C}^\infty(a,b)\to\mathscr{C}^\infty(a,b),Tf(x)=\int_0^x f(x')\,dx',\forall f\in\mathscr{C}^\infty(a,b)$
3. $F:\mathscr{C}^\infty(a,b)\to\mathscr{C}^\infty(a,b),Ff(x)=\sin(x)f(x),\forall f\in\mathscr{C}^\infty(a,b)$

定理 II.4.1. 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

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

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

则$(\mathbf{T}+\mathbf{U})$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II.2.1、线性变换的定义II.4.1和本命题中的运算定义证明。略。 □

上面这一定理，事实上使得线性变换本身也能形成一个向量空间，作为推论如下。

推论 II.4.1.1. 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$\mathscr{L}(\mathscr{V},\mathscr{W})$，零变换是该向量空间的零向量。

证明. 前一句由定理II.4.1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$\mathbf{a}\in\mathscr{V}$，$(\mathbf{T}_0+\mathbf{T})\,\mathbf{a}=\mathbf{T}_0\mathbf{a}+\mathbf{Ta}=\mathbf{0}_{\mathscr{W}}+\mathbf{Ta}=\mathbf{Ta}$，即$\mathbf{T}_0+\mathbf{T}=\mathbf{T}\forall\mathbf{T}\in\mathscr{L}(\mathscr{V},\mathscr{W})$。 □

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。我们在这里考察$\mathscr{L}(\mathscr{V},\mathscr{W})$的基和维数。

引理 II.4.1. 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$\mathbf{Te}_i=\mathbf{b}_i,i=1,\cdots,N$。

*此处要注意区分不同向量空间中的零向量。

†我们可以说线性变换就是向量空间之间的同态映射。

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$\mathbf{T} : \mathscr{V}_N \to \mathscr{W}, \mathbf{Ta} \equiv \sum_{i=1}^{N} \alpha_i \mathbf{b}_i$，则易验$\mathbf{Te}_i = \mathbf{b}_i, i = 1, \cdots, N$。我们还需验证映射$\mathbf{T}$是线性变换，按定义II.4.1只需验证$T(\alpha\mathbf{a} + \beta\mathbf{b}) = \alpha(\mathbf{Ta}) + \beta(\mathbf{Tb})$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

$$\mathbf{T}(\gamma\mathbf{a} + \mathbf{b}) = \sum_{i=1}^{N} (\gamma\alpha_i + \beta_i)\, \mathbf{b}_i \tag{II.4.1}$$

$$\gamma(\mathbf{Ta}) + \mathbf{Tb} = \gamma \sum_{i=1}^{N} \alpha_i \mathbf{b}_i + \sum_{i=1}^{N} \beta_i \mathbf{b}_i \tag{II.4.2}$$

$$= \sum_{i=1}^{N} (\gamma\alpha_i + \beta_i)\, \mathbf{b}_i \tag{II.4.3}$$

所以$\mathbf{T}(\gamma\mathbf{a} + \mathbf{b}) = \gamma(\mathbf{Ta}) + \mathbf{Tb}$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$\mathbf{Ue}_i = \mathbf{b}_i$，则对任一$\mathbf{a} \in \mathscr{V}_N$，

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

即$\mathbf{U}$就是$\mathbf{T}$。 □

定理 II.4.2. 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. 证明的过和，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$(p, q), 1 \leq p \leq N, 1 \leq q \leq M$，定义一个线性变换$\mathbf{E}^{pq} : \mathscr{V}_N \to \mathscr{W}_M$*，

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

*这一线性变换与例II.3.2中的矩阵很类似，可以比较理解。

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II.4.1，对$\{\mathbf{f}_j\}_{j=1}^{M}$，$\mathbf{T}$是唯一的。且上式中的$A_{qi}$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II.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，

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

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$M \times N$。 □

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$N \times M$差别。在本讲义中我们会注意线性变换维数写法的顺序。

讨论完线性变换本身作为向量的性质，我们接下来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，下面先给出这一定理。

定义 II.4.2 (零空间、零化度、秩). （如图II.4.1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），记作$\ker\mathbf{T}$，是所有满足$\mathbf{T}\mathbf{a} = \mathbf{0}_{\mathscr{W}}$的向量$\mathbf{a}$的集合。零空间的维数称为该线性变换的零化度（nullity），记作$\mathrm{nullity}\,\mathbf{T} \equiv \dim(\ker\mathbf{T})$。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$\mathrm{rank}\,\mathbf{T} \equiv \dim(\mathrm{ran}\,\mathbf{T})$。

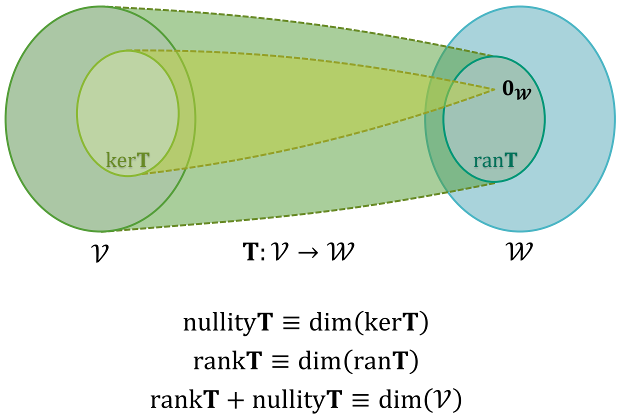

图 II.4.1: 线性变换的零空间、值域、零化度和秩以及线性变换的维数定理表达式

特别地，如果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）。

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，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[1]§7.3 “2.线性变换的简单性质(4)” p. 177。基于定义II.4.2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（§7.3 “2.线性变换的简单性质(5)” [1]p. 178）。

定理 II.4.3 (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

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

证明. 设$\dim\mathscr{V}=n$，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$\{\mathbf{T}\mathbf{a}_1,\cdots,\mathbf{T}\mathbf{a}_n\}$线性生成$\mathbf{T}$的值域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$\mathbf{T}\mathbf{a}_j=\mathbf{0}_{\mathscr{W}}\forall j\leq k$，所以实际上仅$\{\mathbf{T}\mathbf{a}_{k+1},\cdots,\mathbf{T}\mathbf{a}_n\}$就线性生成ran$\mathbf{T}$了。我们进一步验证它们是线性无关的。设标量$\gamma_i$满足$\sum_{i=k+1}^n\gamma_i\mathbf{T}\mathbf{a}_i=\mathbf{0}_{\mathscr{W}}$，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II.4.2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。我们首先考虑一个线性变换是单射的情况。

定理 II.4.4. 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

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

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$\mathbf{Ta}$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II.4.3可直接得到。 □

定理II.4.4联系了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 II.4.4.1. 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II.2.1的推论3、定理II.4.3和II.4.4，$\mathbf{T}$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??，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。

定理 II.4.5. 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“同构线性变换”（isomorphic linear transformation）。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§II.2中我们了解了向量与其在给定某基下的坐标是一一对应的，且易证这一对应关系是一个同构线性变换。因此，事实上任一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$g\circ f:A\to C$是双射，故所有同维数的向量空间之间相互同构。以下定理及其推论从一个不同的出发点证明了这上述的结论。

定理 II.4.6. 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$\mathbf{a}=\sum_{i=1}^N\alpha_i\mathbf{e}_i$，$\mathbf{T}\mathbf{a}=\sum_{i=1}^N\alpha_i\mathbf{f}_i$。易验$\mathbf{a}=\mathbf{0}_{\mathscr{V}}\Leftrightarrow\mathbf{T}\mathbf{a}=\mathbf{0}_{\mathscr{W}}$，即$\mathbf{T}$是单射。由定理II.4.4及其推论可知$\mathbf{T}$是双射。 □

推论 II.4.6.1. 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是恒等映射。我们于是要问：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。

定理 II.4.7. 设$\mathscr{V},\mathscr{W},\mathscr{X}$是数域$\mathbb{F}$上的向量空间，$\mathbf{T}:\mathscr{V}\to\mathscr{W},\mathbf{U}:\mathscr{W}\to\mathscr{X}$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。

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：向量空间的恒等映射总是同构线性变换，称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。

例 II.4.3. 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$\mathbf{T}^{-1}\mathbf{T}=\mathbf{I}_{\mathscr{V}},\mathbf{T}\mathbf{T}^{-1}=\mathbf{I}_{\mathscr{W}}$。

定义 II.4.3 (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）。

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。同态的双射叫同构映射（isomorphism），前面也提到过双射的线性变换叫同构线性变换，故双射的线性算符又称自同构（endomorphic）线性算符。由定理II.4.4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（自同构）的线性算符才是可逆映射。自同构线性算符与其逆映射同属于同一个线性变换的空间$\mathscr{L}(\mathscr{V},\mathscr{V})$（简写为$\mathscr{L}(\mathscr{V})$）。

对于一般的线性变换，本节开头介绍了其作为向量的线性代数。现在我们可以通过复合变换的定义，为线性算符引入“乘法”的代数*。此时恒等变换记号$\mathbf{I}$无需指明作用空间。

定理 II.4.8. 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$\mathbf{a}\in\mathscr{V}$，

$$\begin{aligned}[\mathbf{U}(\mathbf{T}_1+\mathbf{T}_2)]\mathbf{a}&=\mathbf{U}[(\mathbf{T}_1+\mathbf{T}_2)\mathbf{a}] &&\text{（由定理II.4.1中的定义）}\\&=\mathbf{U}(\mathbf{T}_1\mathbf{a}+\mathbf{T}_2\mathbf{a}) &&\text{（}\mathbf{U}\text{是线性变换）}\\&=(\mathbf{UT}_1)\mathbf{a}+(\mathbf{UT}_2)\mathbf{a} &&\text{（由复合映射的定义）}\end{aligned}$$

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，

$$\begin{aligned}[(\mathbf{T}_1+\mathbf{T}_2)\mathbf{U}]\mathbf{a}&=(\mathbf{T}_1+\mathbf{T}_2)(\mathbf{Ua}) &&\text{（由复合映射的定义）}\\&=\mathbf{T}_1(\mathbf{Ua})+\mathbf{T}_2(\mathbf{Ua}) &&\text{（由定理II.4.1中的定义）}\\&=(\mathbf{T}_1\mathbf{U})\mathbf{a}+(\mathbf{T}_2\mathbf{U})\mathbf{a} &&\text{（由复合映射的定义）}\end{aligned}$$

*代数（algebra）的定义此略，此处只强调，要成为代数的运算，其中一个必要条件是要具有封闭性。一般的线性变换的复合操作不满足此条件。

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。 □

II.4.2 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》课里我们主要学习的“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§II.2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[1]§7.3 “3” ,p. 178。

考虑数域$\mathbb{F}$上的n维向量空间$\mathscr{V}_n$的一组有序基$B=(\mathbf{a}_i)_{i=1}^n$，任一向量$\mathbf{x}\in\mathscr{V}$可表示成$\mathbf{x}=\sum_{i=1}^n\xi_i\mathbf{a}_i,\xi_i\in\mathbb{F},i=1,\cdots,n$。我们在§II.2中说明过，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，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$\alpha_{jk}$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。

由于向量和线性变换是抽象的一般概念，但基和坐标又是向量空间和线性变换的一般属性，故不管把什么具体数学对象视作向量和线性变换，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$\{\mathbf{a}_i\}_{i=1}^n\subset$

[†]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。

[*]像例II.2.1和例II.4.2中的那种$\mathscr{C}^\infty$空间的维数是无穷，$\mathscr{C}^\infty$空间的任一组基都有无穷个基向量。把$\mathscr{C}^\infty$空间的向量（函数）用一组基向量（函数）来表出，将得到一个无穷级数。因此这种空间上的向量和线性变换无法表示有限维矩阵。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[3]。本讲义不再涉及，默认只讨论有限维向量空间。

$\mathscr{V}_n, \{\mathbf{b}_j\}_{j=1}^m \subset \mathscr{W}_m$，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$(\alpha_{ji})$，则线性关系式$\mathbf{y} = \mathbf{A}\mathbf{x}$恰好可以写成关于$\mathbf{x}$、$\mathbf{y}$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

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

唯一确定了一个线性变换$\mathbf{A} : \mathscr{V}_n \to \mathscr{W}_m, \mathbf{A} \in \mathscr{L}(\mathscr{V}_n, \mathscr{W}_m)$，使后者在$\mathscr{V}_n$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$A$。总结成定理如下。

定理 II.4.9. 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 II.4.10. 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II.4.9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+双射=同构映射。 □

其实，以上两条定理几乎是与定理II.4.4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 II.4.11. 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II.4.11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II.4.8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）。总结为如下定理：

定理 II.4.12. 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$(\mathbf{T}^{-1})=(\mathbf{T})^{-1}$。

II.4.3 线性变换的转置

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 II.4.4 (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$\mathscr{V}^*$。

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$\mathscr{L}(\mathscr{V},\mathbb{F})$，$\dim\mathscr{V}^*=\dim\mathscr{V}$。

例 II.4.4. 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。

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。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 II.4.13. 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$f_i(\mathbf{a}_j)=\delta_{ij}, i,j=1,\cdots,n$。且$\{f_i\}$线性无关。

证明. “有且只有”由定理II.4.1易证，略。下面证明“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，且后者的基向量作为线性泛函作用于前者的基向量等于克劳内克符号。我们因此可以定义对偶基的概念。

定义 II.4.5 (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$f_i(\mathbf{a}_j)=\delta_{ij}, i,j=1,\cdots,\dim\mathscr{V}$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，$f=\sum_{i=1}^n c_i f_i, c_i\in\mathbb{F}$。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II.4.13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。作为线性泛函的对偶基向量其实就是一种“取坐标的函数”。

接下来我们完成定义线性变换的转置的任务。

定义 II.4.6 (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$\mathbf{T}^\intercal:\mathscr{W}^*\to\mathscr{V}^*$。令$(\mathbf{T}^\intercal g)(\cdot)=g(\mathbf{T}\cdot)$，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$\mathbf{U}g=\mathbf{T}^\intercal g\Rightarrow\mathbf{UT}^\intercal$。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 II.4.14. 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. 由§II.4.2的知识，

$$\begin{aligned}\mathbf{Ta}_i&=\sum_{j=1}^m A_{ji}\mathbf{b}_j,i=1,\cdots,n\\ \mathbf{T}^\intercal g_i&=\sum_{j=1}^n B_{ji}f_j,i=1,\cdots,m\end{aligned}$$

由线性变换转置的定义和对偶基的定义，

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

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

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

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

II.5 基变换与坐标变换公式

不管是向量还是线性变换，它们的本质都独立于它们在选定基下的坐标矩阵。同一个向量或线性变换，在不同基下将对应为不同的坐标矩阵。下面我们讨论同一个向量或线性变换在不同基下的矩阵之间的关系——基变换与坐标变换公式。

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，可列出如下的N个等式：

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

称为从有序基$(\mathbf{e}_i)$到有序基$(\mathbf{e}'_j)$的基变换公式。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。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[1]p. 52,例题4.4。）

定理 II.5.1. 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$\mathbf{e}_i=\sum_{j=1}^{N} S_{ji}^{\mathrm{inv}}\mathbf{e}'_j$*。

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单位矩阵I_n。

我们通过基的过渡矩阵，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：

*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_{ij}^{inv}，矩阵S的分量的倒数记为S_{ij}^{-1}。

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

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，也可以写成矩阵乘：

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

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$v_j' = \sum_{i=1}^{N} S_{ji}^{\text{inv}} v_i, j = 1, \cdots, N$。

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 II.5.2. 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$\mathbf{a}_j' = \sum_{i=1}^{N} S_{ij} \mathbf{a}_i$；$\{\mathbf{b}_j\}, \left\{\mathbf{b}_j'\right\}$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

$$
\begin{aligned}
(\mathbf{A}) &= T (\mathbf{A})' S^{-1} \\
(\mathbf{A})' &= T^{-1} (\mathbf{A}) S
\end{aligned}
$$

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：

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

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 II.5.3. 设$\mathscr{V}$是数域$\mathbb{F}$上的有限维内积空间，则$\mathscr{V}$上的内积不依赖基的选择。

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：

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

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

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

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 II.5.3.1. 设$\mathscr{V}$是数域$\mathbb{F}$上的有限维赋范向量空间，则$\mathscr{V}$上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IV.1.1）易证该命题。 □

II.6 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。

II.6.1 伴随算符

定义 II.6.1 (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。若另一线性算符$\mathbf{T}^*\in\mathscr{V}$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。

对于有限维内积空间，每个线性算符有且只有一个伴随算符（证明见附录§IV.1）。易验，若$\mathbf{T}$是厄米算符，则$i\mathbf{T}$就是反厄米算符。

以下定理列举一些伴随算符的运算规律。

定理 II.6.1. 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. $(\mathbf{T}\mathbf{U})^* = \mathbf{U}^*\mathbf{T}^*, \forall \mathbf{T}, \mathbf{U} \in \mathscr{L}(\mathscr{V})$
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$\mathbf{U}_1 = \frac{1}{2}(\mathbf{T}+\mathbf{T}^*), \mathbf{U}_2 = \frac{1}{2i}(\mathbf{T}-\mathbf{T}^*)$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，且在证明过程顺便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某有序基下的坐标。

定理 II.6.2. 设$\mathscr{V}$是数域$\mathbb{F}$上的有限维内积空间，$B = \{\hat{\mathbf{e}}\}$是$\mathscr{V}$的一组规范正交基，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$\mathbf{a} = \sum_{i=1}^n \alpha_i \hat{\mathbf{e}}_i$，其中$\alpha_i, i = 1,\cdots,n$是$\mathbf{a}$在有序基$B$下的坐标，则$\alpha_i = (\mathbf{a}|\hat{\mathbf{e}}_i)$，因为

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

由§II.4.2可知，线性算符$\mathbf{T}$在给定有序基B下的坐标T_{ij}满足$\mathbf{T}\hat{\mathbf{e}}_i = \sum_{j=1}^n T_{ji}\hat{\mathbf{e}}_j, i = 1,\cdots,n$。同时每个$\mathbf{T}\hat{\mathbf{e}}_i$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$\mathbf{T}\hat{\mathbf{e}}_i = \sum_{j=1}^n (\mathbf{T}\hat{\mathbf{e}}_i|\hat{\mathbf{e}}_j)\hat{\mathbf{e}}_j$。由于$\{\hat{\mathbf{e}}_j\}$线性无关，当$\mathbf{T} \neq \mathbf{0}$时比较上述两结果可得$T_{ij} = (\mathbf{T}\hat{\mathbf{e}}_j|\hat{\mathbf{e}}_i)$。

由伴随算符定义，

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

□

注意，这一定理证明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含有该组基的格拉姆矩阵分量。

回顾线性变换的转置（§II.4.3），如果由$\mathscr{V}$的每个向量$\mathbf{a}\in\mathscr{V}$定义一个相应的线性算符$f_{\mathbf{a}}\in\mathscr{V}^*, f_{\mathbf{a}}(\mathbf{b})\equiv(\mathbf{a}|\mathbf{b})$，则$\mathscr{V}$上的任一线性算符$\mathbf{T}\in\mathscr{L}(\mathscr{V})$与其转置$\mathbf{T}^\intercal\in\mathscr{L}(\mathscr{V}^*)$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。比较而言，线性算符$\mathbf{T}$与其伴随算符$\mathbf{T}^*$满足$(\mathbf{T}\mathbf{a}|\mathbf{b}) = (\mathbf{a}|\mathbf{T}^*\mathbf{b}), \forall\mathbf{a},\mathbf{b}\in\mathscr{V}$，$\mathbf{T}$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。在这一比较下，线性算符的转置与伴随在是否需要共轭上的差别，来自它们生效的是内积的第一个向量还是第二个向量；内积定义中规定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。然而，线性变换的转置和伴随有本质的不同，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，作用于不同空间的向量。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$\mathbf{T}^\intercal$。

在矩阵代数中，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（symmetric operator）和斜称算符（skew-symmetrix operator）。

II.6.2 幺正算符

定义 II.6.2 (幺正算符). 设$\mathscr{V}$是数域$\mathbb{F}$上的内积空间，若线性算符$\mathbf{Q}\in\mathscr{L}(\mathscr{V})$保持内积，即$(\mathbf{Q}\mathbf{a}|\mathbf{Q}\mathbf{b}) = (\mathbf{a}|\mathbf{b}), \forall\mathbf{a},\mathbf{b}\in\mathscr{V}$，则称$\mathbf{Q}$是一个幺正算符（unitary operator）。

以下定理给出幺正算符的双射性。

定理 II.6.3. 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，则以下命题相互等价：

1. $\mathbf{T}$保持内积

2. $\mathbf{T}$是双射
3. $\mathbf{T}$把$\mathscr{V}$的某个规范正交基映射为另一个规范正交基
4. $\mathbf{T}$把$\mathscr{V}$的每个规范正交基映射为一个规范正交基

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

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

3)⇒4)：显然易证；

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

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$\mathbf{T} \in \mathscr{L}(\mathscr{V})$是一个幺正算符时，就有$\|\mathbf{Ta}\| = \|\mathbf{a}\| \, \forall \mathbf{a} \in \mathscr{V}$，即幺正算符不改变向量的长度。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$\mathbf{I}$是幺正算符。

下一条定理告诉我们幺正算符的伴随算符有何性质。

定理 II.6.4. 设$\mathscr{V}$是数域$\mathbb{F}$上的内积空间，$\mathbf{U} \in \mathscr{L}(\mathscr{V})$是线性算符，则$\mathbf{U}^*\mathbf{U} = \mathbf{I} \Leftrightarrow \mathbf{U}$是幺正算符。

证明. 设$\mathbf{U}$是幺正算符，则$\mathbf{U}$可逆，且$(\mathbf{Ua}|\mathbf{b}) = \left(\mathbf{Ua}|\mathbf{UU}^{-1}\mathbf{b}\right) = (\mathbf{a}|\mathbf{U}^{-1}\mathbf{b}), \forall \mathbf{a}, \mathbf{b} \in \mathscr{V}$。因此$\mathbf{U}^{-1} = \mathbf{U}^*$即$\mathbf{UU}^* = \mathbf{I} = \mathbf{U}^*\mathbf{U}$。

设$\mathbf{U}^*\mathbf{U} = \mathbf{UU}^* = \mathbf{I}$，则$\mathbf{U}^{-1} = \mathbf{U}^*, (\mathbf{Ua}|\mathbf{Ub}) = (\mathbf{a}|\mathbf{U}^*\mathbf{Ub}) = (\mathbf{a}|\mathbf{Ib}) = (\mathbf{a}|\mathbf{b}), \forall \mathbf{a}, \mathbf{b} \in \mathscr{V}$。 □

在矩阵代数中，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）。由上面的定理可知，只有在实数域$\mathbb{R}$上的内积空间上，幺正算符在给定基下的坐标矩阵是才是一个正交矩阵。因此，我们又把实数域上的幺正算符称为正交算符（orthogonal operator）。

*这里用到了线性变换的维数定理及其推论。

II.7 线性算符的行列式、迹和特征值

在§II.4.1中我们介绍了线性算符（定义II.4.3）。本节我们以引入线性算符的几种取值不依赖基的选择的标量参数。

II.7.1 线性算符的行列式

定义 II.7.1 ($n\times n$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

$$D\left(a_1,\cdots,\lambda a_i+a_i',\cdots,a_n\right)=\lambda D\left(a_1,\cdots,a_i,\cdots,a_n\right)+D\left(a_1,\cdots,a_i',\cdots,a_n\right),\forall\lambda\in\mathbb{F}$$

2. 当A的任意两行相等则$D\left(A\right)=0$
3. 设A'是A的任意两行调换后的矩阵，则$D\left(A'\right)=-D\left(A\right)$
4. $D\left(I\right)=1$，其中I是$n\times n$单位矩阵

则称D是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[2]§5.2,pp. 144̃146），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其还满足条件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（因函数的定义）的要求，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。证明存在性往往等于找到这一存在。回顾本科的线性代数课本，我们发现那里的定义方式[1]§1.3 “定义3.1” ,p. 7是直接列出公式

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

其中$A\left(i,j\right)\equiv A_{ij}$；$\sigma_i,i=1,\cdots,n$表示$n$阶排列$\sigma$的一种；$\mathrm{sgn}\sigma$是根据$\sigma$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$D\left(A\right)=\det AD\left(I\right)\forall A\in\mathbb{F}^{n\times n}$，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。以下行列式的性质在本讲义中也不作证明而直接承认其成立。

*关于n阶排列的知识可参见[1]§1.2,p. 4。

定理 II.7.1.

- $\det(AB) = \det A \det B, \forall A, B \in \mathbb{F}^{n\times n}$
- $\det(A^{\intercal}) = \det A, \forall A, B \in \mathbb{F}^{n\times n}$
- $\det(A^{-1}) = (\det A)^{-1}, \forall A, B \in \mathbb{F}^{n\times n}$
- 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[1]§1.5,p. 15

定义在数域$\mathbb{F}$上的有限维向量空间$\mathscr{V}$上的线性算符$\mathbf{T} \in \mathscr{L}(\mathscr{V})$在不同有序基下的坐标矩阵之间有变换公式（见§II.4.2）。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II.5.2有$(\mathbf{T}) = S(\mathbf{T})' S^{-1}$，其中$(\mathbf{T}), (\mathbf{T})'$分别是$\mathbf{T}$在$B, B'$下的坐标矩阵。由行列式的性质有$\det(\mathbf{T}) = \det[S(\mathbf{T}) S^{-1}] = \det(\mathbf{T})'$，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，我们因此可以定义“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。

定义 II.7.2 (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。

II.7.2 线性算符的迹

从线性算符的行列式定义过程我们发现，与本科线性代数课上直接定义成一个运算公式不同，我们总是先定义代数规则，再去证明这样的代数规则只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 II.7.3 (线性算符的迹)**.** 设$\mathbf{A}$是数域$\mathbb{F}$上的内积空间$\mathbf{V}$上的线性算符，$\mathbf{A}$的迹（trace）$\mathrm{tr}\mathbf{A} = \sum_k (\mathbf{A}\hat{\mathbf{e}}_k|\hat{\mathbf{e}}_k)$，其中$\{\hat{\mathbf{e}}_k\}$是$\mathscr{V}$的一组规范正交基。

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 II.7.2. 设$\mathbf{A}$是数域$\mathbb{F}$上的内积空间$\mathscr{V}$上的线性算符，

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 II.7.4 (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$\mathbf{V}$上的任意两个线性算符，令$(\mathbf{A}|\mathbf{B}) \equiv \mathrm{tr}\left(\mathbf{A}^*\mathbf{B}\right)$，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），以经常记作$\mathbf{A}:\mathbf{B}$

II.7.3 线性算符的特征值

如果一个数域$\mathbb{F}$上的有限维向量空间$\mathscr{V}$上的线性算符$\mathbf{T}$在基$\{\mathbf{a}_i\}$下的坐标矩阵是对角矩阵，

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

那么就有$\mathbf{T}\mathbf{a}_i = \lambda_i\mathbf{a}_i, i = 1, \cdots, n$，$\mathbf{T}$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$\mathrm{rank}\mathbf{T}$，$\mathbf{T}$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 II.7.5 (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$\lambda$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）。

我们马上且迅速地解决一条定理。

定理 II.7.3. 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$\lambda$的其中两个特征向量，则对任意$\alpha \in \mathbb{F}$有$\mathbf{T}(\alpha\mathbf{a} + \mathbf{b}) = \alpha\mathbf{T}\mathbf{a} + \mathbf{T}\mathbf{b} = \alpha\lambda\mathbf{a} + \lambda\mathbf{b} = \lambda(\alpha\mathbf{a} + \mathbf{b})$。 □

由特征值和特征向量的定义，一个线性算符可能有多个特征值，对应同一个特征值也可以有多个特征向量。上述特征值的定义无法告诉我们回答“一个线性算符到底有多少个特征值”的方法，它甚至没有告诉我们除了碰运气之外“如何找到线性算符的任一个特征值”。为此我们进一步考虑一个重要的线性算符式。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，$\lambda$是$\mathbf{T}$的一个特征值，$\mathbf{a}$是$\mathbf{T}$关于$\lambda$的任一特征向量，则有

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

即$\mathbf{T}$关于λ的特征空间是线性算符$\mathbf{T} - \lambda\mathbf{I}$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 II.7.4. 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给定线性变换的特征值的方法。因为$\det(\mathbf{T}-\lambda\mathbf{I})$必是一个$\dim\mathscr{V}$阶首一多项式*。因此可将此列式$\det(\mathbf{T}-\lambda\mathbf{I})$定义为$\mathbf{T}$的特征多项式（characteristic polynomial）。于是$\mathbf{T}$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 II.7.5. 设$\mathscr{V}$是数域$\mathbb{F}$上的n维向量空间，$\mathbf{T}\in\mathscr{L}(\mathscr{V})$是一个可逆线性算符，则以下命题相互等价：

1. $\mathscr{V}$中存在一组基恰好就是$\mathbf{T}$的特征向量
2. $\mathbf{T}$的特征向量中有n个线性无关
3. $\mathbf{T}$的特征向量线性生成$\mathscr{V}$
4. $\mathbf{T}$的特征空间维数是n
5. $\mathbf{T}$是可对角化的（diagonalizable）

以上定理中的命题之间的等价性几乎是直接的[1]§5.2“矩阵可对角化的条件”,p.123，第5条命题是为了定义“可对角化的线性变换”而设。如果$\mathbf{T}$可对角化，我们就有n个特征值$\{\lambda_i\}_{i=1}^n$，但它们之间有可能有重复的（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

*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。

†这样的数域又叫代数闭域。例如实数域就不是代数闭域。

特征值重复d_j次，则$\mathbf{T}$的对角化坐标矩阵可以写成

$$(\mathbf{T}) = \begin{pmatrix} \theta_1 I_1 & & \\ & \ddots & \\ & & \theta_k I_k \end{pmatrix}$$

其中I_j是$d_j \times d_j$单位矩阵。由之前关于特征值的知识可知，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$d_1 + \cdots + d_k = n$。易验在相同的基下，算符$\mathbf{T} - \theta_j \mathbf{I}, j = 1, \cdots, k$的坐标矩阵也都是对角矩阵，它们各将含有$d_j$个零在对角线上。$\mathbf{T} - \theta_i \mathbf{I}$的零化度就是$d_i$。

以下介绍两个与特征值相关的定理[1]“性质1、性质2、p. 118。

定理 II.7.6. 1. 线性算符的的转置的特征值与原线性算符的特征值相同

2. 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）。

II.8 正规算符及其谱分解

定义 II.8.1 (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$\mathbf{T}\mathbf{T}^* = \mathbf{T}^*\mathbf{T}$，则称$\mathbf{T}$是正规算符（normal operator）。

自伴随算符和幺正算符都是正规算符。正规算符的特征值分解有十分方便的性质，见如下定理。

定理 II.8.1. 设$\mathscr{V}$是数域$\mathbb{F}$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。

证明. 见附录。 □

由此定理可知，正规算符必可对角化。此外，对于幺正算符$\mathbf{U}$，其行列式的绝对值为1：由$\mathbf{U}^*\mathbf{U}=\mathbf{I}$，$\det(\mathbf{U}\mathbf{U}^*)=1=\det\mathbf{U}\det(\mathbf{U}^*)=(\det\mathbf{U})^2$。

在内积空间上，我们可以把正规算符展开成一系列正交投影算符的线性组合，由此我们可以定义这类线性算符的初等函数，扩大线性算符的运算性质。我们先引入正交投影算符的概念。

定义 II.8.2 (最好近似). 设$\mathscr{V}$是数域$\mathbb{F}$上的赋范内积空间，$\mathscr{W}$是$\mathscr{V}$的子空间，如果对$\mathscr{V}$中的向量$\mathbf{b}\in\mathscr{V}$，有$\mathscr{W}$中的一个向量$\mathbf{a}\in\mathscr{W}$满足$\|\mathbf{b}-\mathbf{a}\|\leq\|\mathbf{b}-\mathbf{c}\|\ \forall\mathbf{c}\mathscr{W}$，则称$\mathbf{a}$是$\mathbf{b}$在$\mathscr{W}$中的最好近似（best approximation）。

定理 II.8.2. 设$\mathscr{V}$是数域$\mathbb{F}$上的赋范内积空间，$\mathscr{W}$是$\mathscr{V}$的子空间，则

1. $\mathbf{a}\in\mathscr{W}$是$\mathbf{b}\in\mathscr{V}$在$\mathscr{W}$中的最好近似$\Leftrightarrow$ $\mathbf{b}-\mathbf{a}$与$\mathscr{W}$中所有向量都正交
2. 若$\mathbf{a}\in\mathscr{W}$是$\mathbf{b}\in\mathscr{V}$的一个最好近似，则$\mathbf{a}$是唯一的
3. 若$\mathscr{W}$是有限维的，且$\{\mathbf{e}_k\}$是$\mathscr{W}$的一组正交基，则向量

$$\mathbf{a}=\sum_k\frac{(\mathbf{b}|\mathbf{e}_k)}{(\mathbf{e}_k|\mathbf{e}_k)}\mathbf{e}_k$$

   是$\mathbf{b}$在$\mathscr{W}$的（唯一）一个最好近似，$\forall\mathbf{b}\in\mathscr{V}$。

上面的定理解决了存在唯一性问题，使得我们可以进一步把“最好近似”定义为“正交投影”。

定义 II.8.3 (正交投影算符). 设$\mathscr{V}$是数域$\mathbb{F}$上的赋范内积空间，$\mathscr{W}$是$\mathscr{V}$的子空间，若$\mathbf{b}\in\mathscr{V}$在$\mathscr{W}$中有最好近似$\mathbf{a}\in\mathscr{W}$，则称$\mathbf{a}$是$\mathbf{b}$在$\mathscr{W}$中的正交投影（orthogonal projection）。由$\mathscr{V}$到$\mathscr{W}$的映射$\mathbf{E}:\mathscr{V}\to\mathscr{W},(\mathbf{E}\mathbf{b}-\mathbf{b}|\mathbf{b})=0\forall\mathbf{b}\in\mathscr{V}$称向量空间$\mathscr{V}$到其子空间$\mathscr{W}$的正交投影。

我们可证明映射$\mathbf{E}$是一个线性算符，而且是幂等的，即$\mathbf{E}^n=\mathbf{E}$。

以下定理给出正规算符的谱分解。

定理 II.8.3. 设$\mathscr{V}$是数域$\mathbb{F}$上的有限维内积空间，$\mathbf{T}\in\mathscr{L}(\mathscr{V})$是一个正规算符，$\theta_1,\cdots,\theta_k$为$\mathbf{T}$的两两不同特征值，分别对应特征空间$\mathscr{W}_1,\cdots,\mathscr{W}_k$，记$\mathbf{E}_i$为$\mathscr{V}$到$\mathscr{W}_i$的正交投影，则$\mathbf{T}=\theta_1\mathbf{E}_1+\cdots+\theta_k\mathbf{E}_k$，称$\mathbf{T}$的一个谱分解（spectral decomposition）

推论 II.8.3.1. 设表达式$e_j(x)\equiv\prod_i\left(\frac{x-\theta_i}{\theta_j-\theta_i}\right)$，则该表达式应用于正规算符就得到其谱分解的各个正交投影算符，$\mathbf{E}_j=e_j(\mathbf{T})$。

上面的定理和推论分别定义了一个正规算符的谱分解形式，以及具体获得其中的正交投应算符的计算公式。值得注意的是，正规算符的谱分解一般没有唯一性。有了正规算符的一

个谱分解，我们证明对任一初等表达式$f(x)$，$f(\mathbf{T})=\sum_{i=1}^{k}f(\theta_i)\,\mathbf{E}_i$，可视为“正规算符的函数”的计算定义。

针对正规算符的特征值性质，我们还可以由如下定理进一步区分自伴随算符、幺正算符和非负算符。

定理 II.8.4. 设$\mathscr{V}$是数域$\mathbb{F}$上的n维内积空间，$\mathbf{T}\in\mathscr{L}(\mathscr{V})$是一个正规算符，$\{\lambda_i\}$是$\mathbf{T}$的特征值，则

1. $\lambda_i\in\mathbb{R}\forall i=1,\cdots,n\Leftrightarrow\mathbf{T}$是厄米算符
2. $\overline{\lambda_i}\in\mathbb{R}\forall i=1,\cdots,n\Leftrightarrow\mathbf{T}$是反厄米算符。
3. $\lambda_i\geq0\forall i=1,\cdots,n\Leftrightarrow(\mathbf{Ta}|\mathbf{a})\geq0\forall\mathbf{a}\in\mathscr{V}$，称为非负算符（non-negative operator）
4. $|\lambda_i|=1\forall i=1,\cdots,n\Leftrightarrow\mathbf{T}$是幺正算符

以下性质类似“正实数的平方根是非负实数”。

定理 II.8.5. 设$\mathscr{V}$是数域$\mathbb{F}$上的有限维内积空间，$\mathbf{T}\in\mathscr{L}(\mathscr{V})$是一个正规算符，则必存在唯一非负算符$\mathbf{N}\in\mathscr{L}(\mathscr{V})$满足$\mathbf{T}=\mathbf{N}^2$。

以下性质类似“复数z可分解为$z=\rho e^{i\theta}$”。

定理 II.8.6. 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$\mathbf{U}$和唯一一个非负算符$\mathbf{N}$满足$\mathbf{T}=\mathbf{UN}$，称为$\mathbf{T}$的极分解（polar decomposition）。如果$\mathbf{T}$可逆，则连$\mathbf{U}$也是唯一的。如果$\mathbf{T}$是正规算符，则$\mathbf{UN}=\mathbf{NU}$。

II.9　欧几里得空间

在经典力学中我们假设物理事件所发生的几何空间是欧几里得空间。在本节我们将以集合的语言重新描述欧几里得空间。

II.9.1　准备知识：度量空间与等距变换

我们把一个非空集合的元素称为“点”，然后我们为该集合中任意两个点定义“距离”。

定义 II.9.1 (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：$d(x,y)=0\Leftrightarrow x=y\forall x,y\in\mathscr{E}$
2. 对称性：$d(x,y)=d(y,x)\forall x,y\in\mathscr{E}$
3. 三角不等式：$d(x,z)\leq d(x,y)+d(x,y)\forall x,y,z\in\mathscr{E}$

则称d是定义在$\mathscr{E}$上的一种距离（distance）或度量（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$d(x,y) \geq 0 \forall x,y \in \mathscr{E}$。由3)，$d(x,y)+d(y,x) \geq d(x,x)$；再由2)，$d(x,y)+d(x,y) \geq d(x,x)$；最后由1)有$2d(x,y) \geq 0 \Rightarrow d(x,y) \geq 0$当且仅当$x=y$时取等号。由于这是定义中的规定能够推出的，因此就算它就是我们对距离的最直观要求，但无需写进定义中。

例 II.9.1. 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$(\mathscr{V},d)$。

度量空间的完备性（completeness）是重要的概念，但在本讲义不受度量空间的完备性问题所影响，故暂不作介绍。

在引入度量概念后，由一个度量空间到另一个度量空间有一类特殊的映射叫做等距变换，定义如下。

定义 II.9.2 (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$(A,d_A),(B,d_B)$之间的一个等距变换（isometry）。

由度量的定义易验等距变换都是单射，因为当$a=b, d_B(i(a),i(b))=d_A(a,b)=0 \Rightarrow i(a)=i(b)$。

在欧几里得空间中，对几何对象的平移、旋转一定的角度、镜象这几种操作，都属于等距变换，因为在这些操作前后，任意两点间的距离是不变的。

例 II.9.2. 等距变换的一些例子：

1. 给定两个度量空间：$(\mathbb{R}^+,d_1),d_1(x,y)=|\log x-\log y|\,\forall x,y\in\mathbb{R}^+$，$(\mathbb{R},d_2),d_2(x,y)=|x-y|\,\forall x,y\in\mathbb{R}$，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$i:\mathscr{V}\to\mathscr{V}, i(\mathbf{a})=r^{-1}\mathbf{a}\forall\mathbf{a}\in\mathscr{V}$是等距变换，其中$\|\cdot\|_r=r\|\cdot\|,r\in\mathbb{F}$。

定义 II.9.3 (等距群). 由一个度量空间$(\mathscr{E},d)$到其自身的所有等距变换的集合：

$$\mathscr{I}=\{i:\mathscr{E}\to\mathscr{E}|i\text{是等距变换}\}$$

加上映射的复合操作$i_1\circ i_2,i_1,i_2\in\mathscr{I}$，形成一个交换群，称为等距群（isometry group）

我们通过说明为什么$\mathscr{I}$是一个“群”来同时介绍“群”的定义。类似向量空间的概念，群是一个非空集合加上一种满足某些规则的运算形成的代数结构。按照等距变换的定义，易验$\mathscr{I}$中的元素满足：

1. 封闭性：$i_1 \circ i_2 \in \mathscr{I}, \forall i_1, i_2 \in \mathscr{I}$
2. 结合律：$i_1 \circ (i_2 \circ i_3) = (i_1 \circ i_2) \circ i_3, \forall i_1, i_2, i_3 \in \mathscr{I}$
3. 恒等元素：恒等映射$\mathrm{id}_{\mathscr{E}} \in \mathscr{I}$满足$\mathrm{id}_{\mathscr{E}} \circ i = i \forall i \in \mathscr{I}$。
4. 逆：对任一$i \in \mathscr{I}$存在唯一$i^{-1} \in \mathscr{I}$满足$i^{-1} \circ i = \mathrm{id}_{\mathscr{E}}$
5. 交换律：$i_1 \circ i_2 = i_2 \circ i_1, \forall i_1, i_2 \in \mathscr{I}$

   以上的要求1~4是群（group）的一般定义要求，第5条的满足使该群是一个交换群（commutative group）。我们发现，交换群的这些规定跟向量空间中的向量加法部分是相同的。因此，作为一个交换群的等距群$\mathscr{I}$只要再定义一种“标量乘”，就能形成一个向量空间了。我们规定在$\mathscr{I}$的某个交换子群（即$\mathscr{I}$的一个满足交换群定义的子集）$\mathscr{V}$中，可有如下操作
6. 标量乘法：$\alpha i \in \mathscr{V}$；$\alpha (\beta i) = (\alpha\beta) i$；$1i = i, \forall \alpha, \beta \in \mathbb{R}, i \in \mathscr{V}$
7. 分配律：$\alpha (i_1 \circ i_2) = (\alpha i_1) \circ (\alpha i_2)$；$(\alpha + \beta) i = (\alpha i) \circ (\beta i), \forall \alpha, \beta \in \mathbb{R}, i, i_1, i_2 \in \mathscr{V}$

   则$\mathscr{V}$是一个向量空间。如果我们再定义一个欧几里得范：
8. $\|i\|^2 = (i|i) = d^2 (x, y), \forall x, y \in \mathscr{E}$其中$i$满足$i(x) = y$。

   则$\mathscr{V}$就是一个赋范内积空间，且该范为欧几里得范。

已证明，任一度量空间上的等距群最多只有一个满足上述要求向量空间子群[4]。

II.9.2 欧几里得空间及其平移向量空间

定义 II.9.4 (欧几里得空间). 若度量$d : \mathscr{E} \times \mathscr{E} \to [0, +\infty) \subset \mathbb{R}$上的等距群$\mathscr{I}$有满足上述条件1)到8)的子群$\mathscr{V}$，则称$d$是一个欧几里得度量（Euclidean metric），d赋予集合$\mathscr{E}$以欧几里得空间的结构，或称$\mathscr{E}$是一个欧几里得空间（Euclidean space），内积空间$\mathscr{V}$称$\mathscr{E}$的平移空间（translation space），$\mathscr{V}$中的向量称为$\mathscr{E}$的平移向量。

我们从这一定义可以看出，欧几里得空间本质上是一个度量空间。任何一个的度量空间，总可在其等距群的向量空间子群上形成一个欧几里得空间。

从此，我们把一个欧几里得空间$\mathscr{E}$的元素$X, Y, \dots \in \mathscr{E}$称为点（point），并用向量的记法表示$\mathscr{E}$的平移空间$\mathscr{V}$中的等距变换$\mathbf{u}, \mathbf{v}, \dots \in \mathscr{V}$，称为平移向量。同时我们把$\mathscr{V}$中的一个平移向量$\mathbf{u}$作用于$\mathscr{E}$中一个点$X$得到另一个点$Y$表示为$Y = X + \mathbf{u}, \mathbf{u} = Y - X, X - Y = -\mathbf{u}$，故两个$\mathscr{E}$中的点“相减”的结果是$\mathscr{V}$中的一个平移向量。且由于平移向量本质上是一个等距变换，故$d(X, Y) = d(Y, X) = \|\mathbf{u}\| = \|Y - X\| = \|X - Y\|$。注意，我们没有定义两个点“相加”（$X + Y$）的意义。

下面我们在上述这种定义下的欧几里得空间中引入角、直线、位置向量和基本坐标系。

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，

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

再应用柯西–施瓦茨不等式，有

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

定义 II.9.5 (角)**.** 设$(\mathscr{E},d)$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

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

称“点XOY所夹的锐角”，或简称“角XOY”。其中余弦函数$\cos:[0,\pi]\to\mathbb{R},\cos(x)=\frac{1}{2}\left(e^{ix}+e^{-ix}\right)$。

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，角的取值范围$\mathrm{ran}\angle=[0,\pi]$。同时$\angle XOY$的顺序是重要的，$\angle YOX=-\angle XOY$。等距变换不改变角度。设$i:\mathscr{E}\to\mathscr{E}$是一个等距变换，可验证$\angle i(X)\,i(O)\,i(Y)=\angle XOY\forall X,O,Y\in\mathscr{E},X\neq O\neq Y$。

定义 II.9.6 (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$L_{XY}=\{C|C=X+\alpha\,(Y-X)\,,\alpha\in\mathbb{R}\}$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自然地定义为其平移空间的维数。

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

如果选定某点$O\in\mathscr{E}$为原点（origin），则对任一点$X\in\mathscr{E}$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$X$的位置向量（position vector）。注意，当且只当选定了原点后，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形成双射（一一对应关系）。

设$\{\hat{\mathbf{e}}_i\}$是$\mathscr{V}$的一组规范正交基，则$\{L_{OX_i}|\mathbf{r}\,(X_i)=\hat{\mathbf{e}}_i,i=1,\cdots,\dim\mathscr{V}\}$称为$\mathscr{E}$的一个直角坐标系（rectangular coordinates），又称笛卡尔坐标系（Cartesian coordinates）。点$X\in\mathscr{E}$对应的位置向量$\mathbf{r}_X$在这组基下的坐标称为点X在该坐标系下的坐标。我们将一个选定的原点和

*由欧几里得空间的完备性和实数集的完备性可知L_{XY}和$L_{XY}^{\mathscr{V}}$是同构的。

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。我们常常默认一个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$\mathbb{R}^n$来表示点的坐标，$n=\dim\mathscr{V}$。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。

II.9.3 等距变换的表示定理

下面我们介绍等距变换的一个重要定理：等距变换的表示定理。这个定理使得等距变换可以具体地表达成一个通式。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。

定理 II.9.1 (等距变换的表示定理). 设$(\mathscr{E},d)$是一个欧几里德空间，$\mathscr{V}$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$i:\mathscr{E}\to\mathscr{E},i\in\mathscr{I}$都可表示为

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

其中$\mathbf{Q}_i$是一个正交算符，关于i唯一存在。

证明. 见附录。 □

推论 II.9.1.1. 欧几里得空间上的等距变换都是双射。

证明. 定理II.9.1已经暗示欧几里得空间上的等距变换都是单射。故仅需再证明对任一$i:\mathscr{E}\to\mathscr{E}$和$Y\in\mathscr{E}$总存在一个$X\in\mathscr{E}$满足$i\left(X\right)=Y$。我们可直接找出这样的$X$：

$$X=X_0+\mathbf{Q}^{-1}\left(Y-i\left(X_0\right)\right)$$

验证这就是满足条件的X：

$$\begin{aligned}i\left(X\right)&=i\left(X_0\right)+\mathbf{Q}\left(\mathbf{X}_0+\mathbf{Q}^{-1}\left(Y-i\left(X_0\right)\right)-X_0\right)\\&=i\left(X_0\right)+Y-i\left(X_0\right)\\&=Y\end{aligned}$$

□

定理II.9.1的通俗解释：给定任一等距变换i，仅需知道它对某一参考点X_0的像是哪个点，以及该变换的特征正交算符$\mathbf{Q}_i$，就可以知道它对任意点X的像。因此这一定理给出了等距变换的通用表达式。

这里的等距变换i不一定是欧几里得空间$\mathscr{E}$的平移向量空间$\mathscr{V}$中的元素（前面提到过$\mathscr{V}$至多是$\mathscr{I}$的子群）。例如旋转和镜向反转都是等距变换，却不满足向量空间对向量的要求。

一般$i\left(X_0\right)$是容易找到的，但是$\mathbf{Q}_i$不是直接易得的。我们可以举例认识$\mathbf{Q}_i$的一般意义。

例 II.9.3. *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：*

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

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

$i_1\circ i_2=i_2\circ i_3$（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（幺正）算符性质$\det\mathbf{Q}=\pm1$。当$\det\mathbf{Q}=1$时，i_2是一种旋转操作；当$\det\mathbf{Q}=-1$时，由$\mathbf{Q}=\left(-\mathbf{I}\right)\left(-\mathbf{Q}\right)$和$\det\left(-\mathbf{Q}\right)=1$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。

在连续介质力学中，我们只考虑$\det\mathbf{Q}=1$的情况，即等距变换中的正交算符仅表旋转。在此限定下，定理II.9.1说的就是，欧几里得空间的任一等距变换（镜像除外）都是平移加旋转。

II.10　向量函数及其图像

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 II.10.1 (向量函数)**.** 从$\mathbb{R}^n$到$\mathbb{R}^m$的映射$\mathbf{f}:\mathbb{R}^n\supseteq D\to\mathbb{R}^m$是一个向量函数（vector function）。其变量是一个n维向量空间的向量$\mathbf{r}\in\mathbb{R}^n$，$\mathbf{r}=\left(r_1,\cdots,r_n\right)^\intercal$；其像（函数值）$\mathbf{f}\left(\mathbf{r}\right)$是一个$m$维向量空间的向量$\mathbf{f}\left(\mathbf{r}\right)\in\mathbb{R}^m$，$\mathbf{f}\left(\mathbf{r}\right)=\left(f_1\left(\mathbf{r}\right),\cdots,f_m\left(\mathbf{r}\right)\right)^\intercal$。$f_j:\mathbb{R}_n\to\mathbb{R},j=1,\cdots,m$称为$\mathbf{f}$的坐标函数（coordinate function）。

严格而言，函数的记法$\mathbf{f}:\mathbb{R}^n\supseteq D\to\mathbb{R}^m$表示函数$f$的定义域$\mathrm{dom}f=D$，而不是整个$\mathbb{R}^n$。这一记法中箭头右侧则一向仅指陪域。但是在后面的讨论当中，当函数的定义域待定时，我们也常粗略记为$\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$。

例 II.10.1. 函数$\mathbf{f}:\mathbb{R}^3\to\mathbb{R}^2$，

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

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

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

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

例 II.10.2. 函数$\mathbf{A}:\mathbb{R}^3\to\mathbb{R}^2$，

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

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 II.10.2 (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$\left(x_1,\cdots,x_n,y_1,\cdots,y_m\right)^\intercal$写成$\left(\mathbf{x},\mathbf{y}\right)^\intercal$，其中$\mathbf{x}=\left(x_1,\cdots,x_n\right)^\intercal$，$\mathbf{y}=\left(y_1,\cdots,y_m\right)^\intercal$，则$\mathbf{F}\left(\mathbf{r}\right),\mathbf{r}\in\mathbb{R}^{n+m}$也可写成$\mathbf{F}\left(\mathbf{x},\mathbf{y}\right)$。若函数$\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$满足方程$\mathbf{F}\left(\mathbf{x},\mathbf{f}\left(\mathbf{x}\right)\right)=\mathbf{0},\forall\mathbf{x}\in\mathbb{R}^n$则称这一方程隐含地定义了（implicitly defined）函数$\mathbf{f}$。

例 II.10.3. 设函数$F\left(x,y\right)=x^2+y^2-1=0$，则“$F\left(x,f\left(x\right)\right)=x^2+\left(f\left(x\right)\right)^2-1=0\forall x\in\mathrm{dom}f$”隐含地定义了以下任一函数：

$$f_1\left(x\right)=\sqrt{1-x^2},-1\leq x\leq 1$$

$$f_2\left(x\right)=-\sqrt{1-x^2},-1\leq x\leq 1$$

$$f_3\left(x\right)=\begin{cases} \sqrt{1-x^2}, & -\frac{1}{2}\leq x\leq 0 \\ -\sqrt{1-x^2}, & 0\leq x\leq 1 \end{cases}$$

定义 II.10.3 (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$\left(\mathbf{r},\mathbf{f}\left(\mathbf{r}\right)\right)$的集合。

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

$$\begin{cases} y_1=f_1\left(x_1,\cdots,x_n\right) \\ \vdots \\ y_m=f_m\left(x_1,\cdots,x_n\right) \end{cases}$$

*请复习高等数学相关内容[5]§6.4 6.7。

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 II.10.4. 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*II.10.1*所示的一条二次曲线。

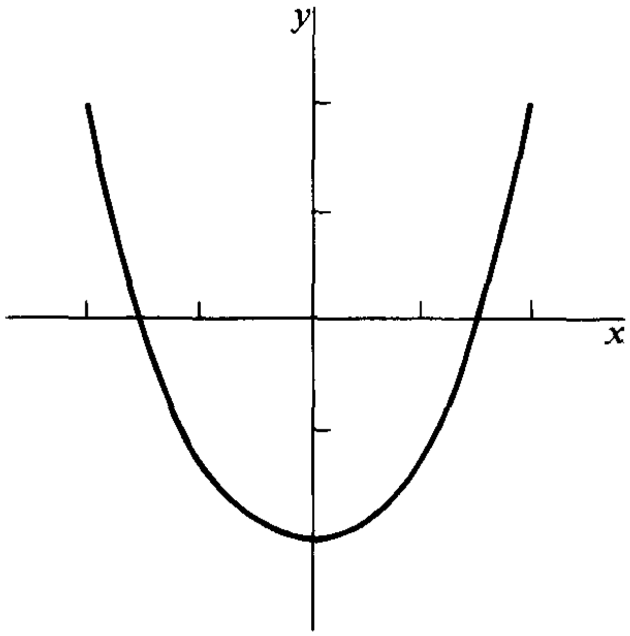

图 II.10.1: $y=x^2-2$

例 II.10.5. $f:\mathbb{R}^2\to\mathbb{R}, f(\mathbf{r})=\|\mathbf{r}\|^2$的图像是有序*3*元组$(r_1,r_2,f)$的集合，是*3*维空间的一个如图*II.10.2*所示的曲面。

例 II.10.6. 函数$\mathbf{f}:\mathbb{R}\to\mathbb{R}^2$，

$$\mathbf{f}(t)=\begin{pmatrix}\cos t\\ \sin t\end{pmatrix}, t\in\mathbb{R}$$

的图像是$\mathbb{R}^3$的子集。如何画出来？首先由函数定义式有$\|\mathbf{f}(t)\|=\sqrt{\cos^2 t+\sin^2 t}=1$，故$\mathbf{f}$的长度是恒定值。$t$是向量$\mathbf{f}$与$\hat{\mathbf{e}}_1$轴夹角的弧度。当$t$在$\mathbb{R}$内取不同值时，$\mathbf{f}$就在$\hat{\mathbf{e}}_1$-$\hat{\mathbf{e}}_2$平面上绕原点划出单位圆。但是，这个单位圆不是$\mathbf{f}$的图像。$\mathbf{f}$的图像是*3*元组$(t,\cos t,\sin t)$的集合，$\mathbb{R}^3$的子集，即如图*II.10.3*所示的螺线。

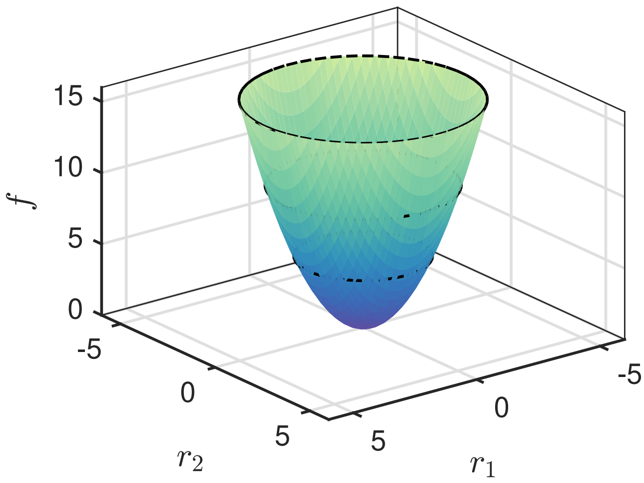

图 II.10.2: $f(\mathbf{r}) = \|\mathbf{r}\|^2$

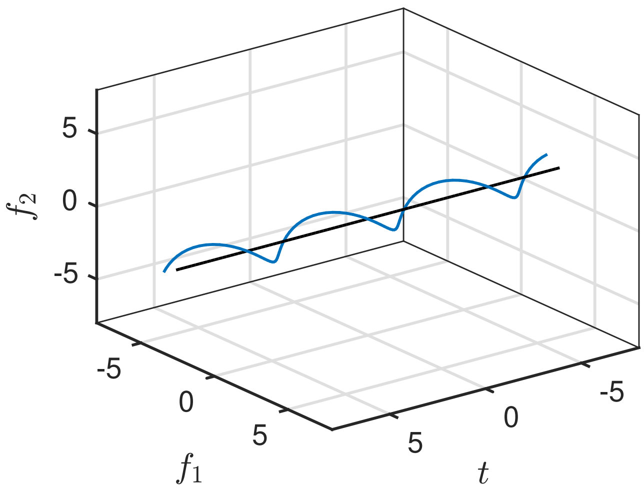

图 II.10.3: $\mathbf{f}(t) = (\cos t, \sin t)^\intercal$

例 II.10.7. 函数$f : \mathbb{R}^2 \to \mathbb{R}^3$,

$$\mathbf{f}(\mathbf{r}) = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} = \begin{pmatrix} r_1 \cos r_2 \\ r_1 \sin r_2 \\ r_2 \end{pmatrix}, 0 \le r_1 \le 4, 0 \le r_2 \le 2\pi$$

$\mathbf{f}$的定义域是2维平面上的一个矩形（图II.10.4）。令$r_1 = a$，a是常数。则有

$$\mathbf{f} = \begin{pmatrix} a \cos r_2 \\ a \sin r_2 \\ r_2 \end{pmatrix}, 0 \le r_2 \le 2\pi$$

且$f_1^2 + f_2^2 = a^2$。把r_2作为f_3轴，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

的投影是圆心在原点、半径为a的圆。令$r_2=\theta$，θ是常数，则有

$$\mathbf{f}=\begin{pmatrix} r_1\cos\theta \\ r_1\sin\theta \\ \theta \end{pmatrix}, 0\le r_2\le 2\pi$$

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II.10.4所示的螺带曲面。但按定义它不是函数$\mathbf{f}$的图像。我们说这个三维曲面是由一个参数方程$\mathbf{f}=\mathbf{f}(\mathbf{f})$定义的曲面，它的参数$\mathbf{r}\in\mathbb{R}^2$域就是如图II.10.4所示的矩形区域，称为参数域。我们所画出来的3维曲面只是函数$\mathbf{f}$的值域。

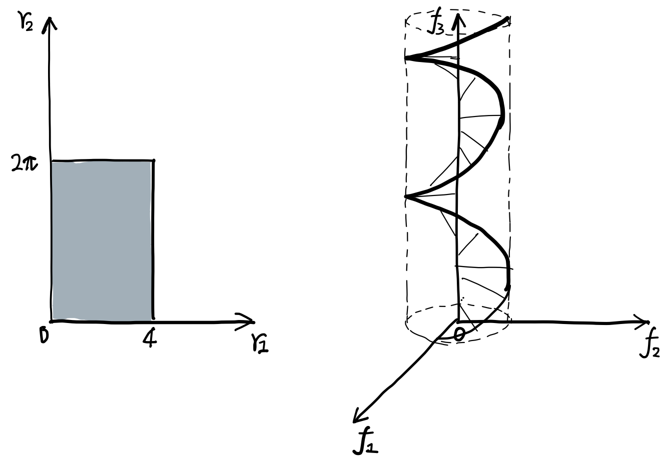

图 II.10.4: 函数$\mathbf{f}(\mathbf{r})=(f_1,f_2,f_3)(r_1\cos r_2, r_1\sin r_2, r_2)^\intercal, 0\le r_1\le 4, 0\le r2\le 2\pi$

定义 II.10.4 (函数的水平集). * 函数$\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$的水平集（level set）$S=\{\mathbf{r}|\mathbf{f}(\mathbf{r})=\mathbf{a}\}, \mathbf{a}\in\mathbb{R}^m$。

例 II.10.8. 设函数$f:\mathbb{R}^3\to\mathbb{R}, f(\mathbf{r})=r_1r_2+r_2r_3+r_3r_1$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，$S$与$\hat{\mathbf{e}}_2$-$\hat{\mathbf{e}}_3$、$\hat{\mathbf{e}}_1$-$\hat{\mathbf{e}}_3$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II.10.5所示）。但是，按定义，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。

*在高等数学课里，我们学过水平集的概念[6]§7.1,p 1。

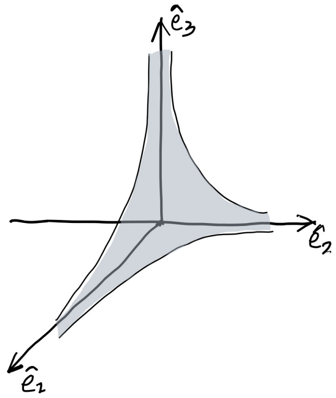

图 II.10.5: 函数$f(\mathbf{r}) = r_1r_2 + r_2r_3 + r_3r_1$

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。
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。

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$\mathbf{f}$。

II.11 向量函数的极限与连续性

II.11.1 向量函数的极限与连续性

我们用“$\epsilon - \delta$语言”来定义函数极限

定义 II.11.1 (函数的极限). 给定函数$\mathbf{f}: \mathbb{R}^n \supseteq D \to \mathbb{R}^m$，如果对任意正实数$\epsilon$总存在正实数$\delta$使得只要$0 < \|\mathbf{r} - \mathbf{r}_0\| < \delta, \mathbf{r}, \mathbf{r}_0 \in \mathbb{R}^n$总有$\|\mathbf{f}(\mathbf{r}) - \mathbf{y}_0\| < \epsilon, \mathbf{y}_0 \in \mathbb{R}^m$，则称$\mathbf{y}_0$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 II.11.1. 函数$\mathbf{f}: \mathbb{R}^n \supseteq D \to \mathbb{R}^m$在$\mathbf{r}_0 \in \mathbb{R}^n$处存在极限$\lim_{\mathbf{r} \to \mathbf{r}_0} \mathbf{f}(\mathbf{r}) = \mathbf{a}$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。

若已知$\lim_{\mathbf{r} \to \mathbf{r}_0} \mathbf{f}(\mathbf{r}) = \mathbf{a}$，即对任一实数$\epsilon > 0$都存在实数$\delta > 0$使得只要$0 < \|\mathbf{r} - \mathbf{r}_0\| < \delta$就

有$\|\mathbf{f}(\mathbf{r})-\mathbf{a}\|<\epsilon$。那么，给定任一$\mathbb{R}^m$的标准基$\{\hat{\mathbf{u}}_i\}$，对任一$i\in\{1,\cdots,m\}$都有

$$\begin{aligned}|f_i(\mathbf{r})-\alpha_i| &\le \sum_{i=1}^{m}|f_i(\mathbf{r})-\alpha_i|\,\|\hat{\mathbf{u}}_i\| \qquad \text{（当且仅当}m=1\text{时取等号。）}\\ &=\sum_{i=1}^{m}\|(f_i(\mathbf{r})-\alpha_i)\,\hat{\mathbf{u}}_i\|\\ &\le\left\|\sum_{i=1}^{m}(f_i(\mathbf{r})-\alpha_i)\,\hat{\mathbf{u}}_i\right\| \qquad \text{（三角不等式。）}\\ &=\|\mathbf{f}(\mathbf{r})-\mathbf{a}\|<\epsilon\end{aligned}$$

故$\lim_{\mathbf{r}\to\mathbf{r}_0}f_i(\mathbf{r})=\alpha_i, i=1,\cdots,m$，其中$\alpha_i, i=1,\cdots,n$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$\epsilon>0$都能找到实数$\delta_i>0, i=1,\cdots,m$使得只要$f_i$的定义域内一向量$\mathbf{r}$到$\mathbf{r}_0$的距离$0<\|\mathbf{r}-\mathbf{r}_0\|<\delta_i$便有$|f_i(\mathbf{r})-a_i|<\frac{\epsilon}{\sqrt{m}}$。令$\delta=\min\{\delta_1,\cdots,\delta_m\}$，则当$0<\|\mathbf{r}-\mathbf{r}_0\|<\delta$则有$\max\{|f_i(\mathbf{r})-a_i|\}<\frac{\epsilon}{\sqrt{m}}$。

利用事实*

$$\|\mathbf{x}\|\le\sqrt{n}\max\{|x_1|,\cdots,|x_n|\},\forall\mathbf{x}\in\mathbb{R}^n$$

可得

$$\|\mathbf{f}(\mathbf{r})-\mathbf{a}\|\le\sqrt{m}\max\{|f_i(\mathbf{r})-a_i|\}<\epsilon$$

□

定义 II.11.2 (函数的连续性). 给定函数$\mathbf{f}:\mathbb{R}^n\supseteq D\to\mathbb{R}^m$，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）。

定理 II.11.2. 函数$\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$在$\mathbf{r}_0\in\mathbb{R}^n$处连续的充要条件是其坐标函数在$\mathbf{r}_0$处都连续。

证明. 由定理II.11.1直接得证。 □

如果一个函数在其定义域内处处都连续，我们就直接称该函数在其定义域内是连续的，或称该函数是连续函数。以下给出的定理及其推论，既是一个例子，也是一个重要结论。

定理 II.11.3. 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$\|\mathbf{L}\mathbf{x}\|\le k\|\mathbf{x}\|,\forall\mathbf{x}\in\mathscr{V}$，当且仅当$n=1$时取等号。

*由$\|\mathbf{x}\|^2=\sum_{i=1}^{n}x_i^2\le n\max\{x_1^2,\cdots,x_n^2\}$得到。

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II.5.3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IV.1.1，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

$$\begin{aligned}
\|\mathbf{L}\mathbf{x}\| &= \left\| \sum_{i=1}^n x_i \mathbf{L}\mathbf{e}_i \right\| \\
&\leq \sum_{i=1}^n |x_i| \|\mathbf{L}\mathbf{e}_i\| \\
&\leq \sum_{i=1}^n \|\mathbf{x}\|_{\mathrm{E}} \|\mathbf{L}\mathbf{e}_i\| \\
&\leq \sum_{i=1} K \|\mathbf{x}\| \|\mathbf{L}\mathbf{e}_i\| = k \|\mathbf{x}\|
\end{aligned}$$

其中$k = K \sum_{i=1}^n \|\mathbf{L}\mathbf{e}_i\|$。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IV.1.1。由于引理IV.1.1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 II.11.3.1. 线性变换是连续函数。

证明. 由定理II.11.3，对任意$\mathbf{x}, \mathbf{x}_0 \in \mathscr{V}$有$\|\mathbf{L}\mathbf{x} - \mathbf{L}\mathbf{x}_0\| = \|\mathbf{L}(\mathbf{x} - \mathbf{x}_0)\| \leq k \|\mathbf{x} - \mathbf{x}_0\|$，故对任一$\epsilon > 0$总可取$\delta = \frac{\epsilon}{k}$使得只要$\|\mathbf{x} - \mathbf{x}_0\| < \delta$就有$\|\mathbf{L}\mathbf{x} - \mathbf{L}\mathbf{x}_0\| \leq k \|\mathbf{x} - \mathbf{x}_0\| < k\delta = \epsilon$。 □

II.11.2 极限的一些性质

定理 II.11.4 (极限基本性质). 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，$g(x) \leq f(x) \leq h(x)$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$|\lim_{x \to c} f(x)| = |L| = \lim_{x \to c} |f(x)|$。

证明. 留作练习或者待补充。 □

II.12 向量函数的微分与导数

II.12.1 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 II.12.1 (一元函数的导数). [5]"定义2.1.1",p. 70 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），极限值称为函数在点x_0处的导数（或微商），记为

$$f'(x_0), y'|_{x=x_0}, \left.\frac{df}{dx}\right|_{x=x_0}, 或 \left.\frac{dy}{dx}\right|_{x=x_0},$$

即

$$f'(x_0) = \lim_{\Delta x\to 0}\frac{f(x_0+\Delta x)-f(x_0)}{\Delta x}。$$

若极限不存在，称函数$f(x)$在x_0处不可导。

定义 II.12.2 (一元函数的微分). [5]"定义2.5.1",p. 103 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。且称$A\Delta x$为函数$y=f(x)$在点$x_0$处对应于自变量增量$\Delta x$的微分，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 II.12.1. [5] "定理2.5.1" ,p. 103 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[5]p. 103。 □

关于符号"d"的意义，这里给出一个比高等数学课本[5]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 II.12.3 (一元向量函数的导数)**.** 设函数$\mathbf{f} : \mathbb{R} \supseteq D \to \mathbb{R}^n$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以下一元向量函数求导法则成立：对任意函数$\mathbf{f}:\mathbb{R}\supseteq D\to\mathbb{R}^n,\mathbf{g}:D\to\mathbb{R}^n$，

- $\frac{d\mathbf{c}}{dt}=0,\mathbf{c}\in\mathbb{R}^n$是常向量
- $\frac{d}{dt}\left(\alpha\mathbf{f}+\beta\mathbf{g}\right)=\alpha\frac{d}{dt}\mathbf{f}+\beta\frac{d}{dt}\mathbf{g},\forall\alpha,\beta\in\mathbb{R}$
- $\frac{d}{dt}\left[u\left(t\right)\mathbf{f}\left(t\right)\right]=\frac{du}{dt}\mathbf{f}+u\frac{d}{dt}\mathbf{f},\forall u:D\to\mathbb{R}$
- $\frac{d}{dt}\left(\mathbf{f}\cdot\mathbf{g}\right)=\frac{d\mathbf{f}}{dt}\cdot\mathbf{g}+\mathbf{f}\cdot\frac{d\mathbf{g}}{dt}$
- $\frac{d}{dt}\left(\mathbf{f}\times\mathbf{g}\right)=\frac{d\mathbf{f}}{dt}\times\mathbf{g}+\mathbf{f}\times\frac{d\mathbf{g}}{dt}$

定义 II.12.4 (多元标量值函数的偏导数). 给定函数$f:\mathbb{R}^n\supseteq D\to\mathbb{R}$，若对某一$i\in\{1,\cdots,n\}$极限

$$\lim_{t\to0}\frac{f\left(\cdots,x_i+h,\cdots\right)-f\left(\cdots,x_i,\cdots\right)}{t}$$

当$x_i=x_{i0}$时存在，则称该极限为函数$f\left(\mathbf{x}\right)$对第i个变量x_i在$x_i=x_{i0}$处的的偏导数（partial derivative），记为

$$\left.\frac{\partial f\left(\mathbf{x}\right)}{\partial x_i}\right|_{x_i=x_{i0}}\equiv\lim_{t\to0}\frac{f\left(\cdots,x_i+h,\cdots\right)-f\left(\cdots,x_i,\cdots\right)}{t}$$

定理 II.12.2. * 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 II.12.5 (向量函数的偏导数). 函数$\mathbf{f}:\mathbb{R}^n\supseteq D\to\mathbb{R}^m$对第$i$个变量在$x_i=x_{i0}$处的偏导数

$$\left.\frac{\partial\mathbf{f}\left(\mathbf{x}\right)}{\partial x_i}\right|_{x_i=x_{i0}}\equiv\begin{pmatrix}\left.\frac{\partial f_1(\mathbf{x})}{\partial x_i}\right|_{x_i=x_{i0}}\\ \vdots\\ \left.\frac{\partial f_m(\mathbf{x})}{\partial x_i}\right|_{x_i=x_{i0}}\end{pmatrix}\in\mathbb{R}^m$$

其中，$\mathbf{x}=\left(x_1,\cdots,x_n\right)^\intercal\in\mathbb{R}^n$。

II.12.2 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[6] “定义7.3.1” ,p. 19：

*即高等数学[6]p. 16定理7.2.1向n维的推广。

†进一步了解：Symmetry of second derivatives。

定义 II.12.6 (函数的全微分（多元标量值函数）). 若函数$z=f\left(x,y\right)$在点$P_0\left(x_0,y_0\right)$的全增量

$$\Delta z=f\left(x_0+\Delta x,y_0+\Delta y\right)-f\left(x_0,y_0\right)$$

可表示为

$$\Delta z=A\Delta x+B\Delta y+o\left(\rho\right)$$

其中A,B只与点(x_0,y_0)有关，而与$\Delta x,\Delta y$无关。又$\rho=\sqrt{\left(\Delta x\right)^2+\left(\Delta y\right)^2}$,$o\left(\rho\right)$是当$\rho\to 0$时$\rho$的高阶无穷小，则称函数$z=f\left(x,y\right)$在点$P_0\left(x_0,y_0\right)$可微分，且把$\Delta z$的线性主部$A\Delta x+B\Delta y$称为函数$z=f\left(x,y\right)$在点$P_0\left(x_0,y_0\right)$的全微分，记作

$$dz|_{x=x_0,y=y_0}=A\Delta x+B\Delta y \text{或} df\left(x_0,y_0\right)=A\Delta x+B\Delta y$$

如果函数$z=f\left(x,y\right)$在区域D内每一点都可微，则称这函数在D内可微。

如果函数$z=f\left(x,y\right)$在点$P_0\left(x_0,y_0\right)$处可微，则有

$$\begin{aligned}
&f\left(x_0+\Delta x,y_0+\Delta y\right)-f\left(x_0,y_0\right)=A\Delta x+B\Delta y+o\left(\rho\right)\\
\Leftrightarrow &o\left(\rho\right)=f\left(x_0+\Delta x,y_0+\Delta y\right)-f\left(x_0,y_0\right)-\left(A\Delta x+B\Delta y\right)\\
\Leftrightarrow &\lim_{\rho\to 0}\frac{f\left(x_0+\Delta x,y_0+\Delta y\right)-f\left(x_0,y_0\right)-\left(A\Delta x+B\Delta y\right)}{\sqrt{\Delta x^2+\Delta y^2}}=0\\
\Leftrightarrow &\lim_{\rho\to 0}\frac{\Delta z-dz}{\rho}=0
\end{aligned}$$

其中$\rho=\sqrt{\Delta x^2+\Delta y^2}$。上面的极限式是定义II.12.6的等价定义式。我们其实可以不引入某个高阶无穷小$o\left(\rho\right)$，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 II.12.7 (函数的微分（向量函数）). 若函数$\mathbf{f}:\mathbb{R}^n\supseteq D\to\mathbb{R}^m$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$N\left(\mathbf{x}_0\right)$中的点$\mathbf{x}\in N\left(\mathbf{x}_0\right)$，

$$\lim_{\mathbf{x}\to\mathbf{x}_0}\frac{\mathbf{f}\left(\mathbf{x}\right)-\mathbf{f}\left(\mathbf{x}_0\right)-\mathbf{L}\left(\mathbf{x}-\mathbf{x}_0\right)}{\left\|\mathbf{x}-\mathbf{x}_0\right\|}=\mathbf{0}$$

则称函数$\mathbf{f}\left(\mathbf{x}\right)$在$\mathbf{x}_0$处可微分（differentiable）。记$d\mathbf{f}_{\mathbf{x}_0}=\mathbf{L}\left(\mathbf{x}-\mathbf{x}_0\right)$是函数$\mathbf{f}\left(\mathbf{x}\right)$在$\mathbf{x}_0$处的微分（differential）。如果函数$\mathbf{f}\left(\mathbf{x}\right)$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）。

我们可以根据向量函数的定义来看出，定义II.12.7就是定义II.12.6的推广。延用定义II.12.7的设定，设$\mathbf{f}=\left(f_1,\cdots,f_m\right)^\intercal$，$\mathbf{x}=\left(x_1,\cdots,x_n\right)^\intercal$，$\mathbf{x}_0=\left(x_{01},\cdots,x_{0n}\right)^\intercal$，则函数$\mathbf{f}\left(\mathbf{x}\right)$在$\mathbf{x}_0$处的全

增量：

$$\mathbf{f}\left(\mathbf{x}\right)-\mathbf{f}\left(\mathbf{x}_0\right)=\begin{pmatrix} f_1\left(x_1,\cdots,x_n\right)-f_1\left(x_{01},\cdots,x_{0n}\right)\\ \vdots \\ f_m\left(x_1,\cdots,x_n\right)-f_m\left(x_{01},\cdots,x_{0n}\right)\end{pmatrix}$$

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II.12.6写成

$$f_i\left(x_1,\cdots,x_n\right)-f_i\left(x_{01},\cdots,x_{0n}\right)=\sum_{j=1}^{n}L_{ji}\left(x_j-x_{0j}\right)+o\left(\rho\right),i=1,\cdots,m$$

其中，L_{ij}是“线性主部”的系数，只与$(x_{01},\cdots,x_{0n})$有关，$\rho=\left(\sum_{j=1}^{n}\left(x_j-x_{0j}\right)^2\right)^{1/2}=\|\mathbf{x}-\mathbf{x}_0\|$，$o\left(\rho\right)$是当$\rho\to 0$时$\rho$的高阶无穷小。这等价于如下极限式

$$\lim_{\rho\to 0}\frac{1}{\|\mathbf{x}-\mathbf{x}_0\|}\begin{pmatrix} f_1\left(x_1,\cdots,x_n\right)-f_1\left(x_{01},\cdots,x_{0n}\right)-\sum_{j-1}^{n}L_{1j}\left(x_j-x_{0j}\right)\\ \vdots \\ f_m\left(x_1,\cdots,x_n\right)-f_m\left(x_{01},\cdots,x_{0n}\right)-\sum_{j-1}^{n}L_{mj}\left(x_j-x_{0j}\right)\end{pmatrix}=\begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix}$$

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II.12.7的极限式。

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 II.12.3. 设函数$\mathbf{f}:\mathbb{R}^n\supseteq D\to\mathbb{R}^m$在点$\mathbf{x}_0\in D$处可微分，即存在线性变换$\mathbf{L}\in\mathscr{L}\left(\mathbb{R}^n,\mathbb{R}^m\right)$满足

$$\lim_{\Delta\mathbf{x}\to\mathbf{0}}\frac{\mathbf{f}\left(\mathbf{x}_0+\Delta\mathbf{x}\right)-\mathbf{f}\left(\mathbf{x}_0\right)-\mathbf{L}\Delta\mathbf{x}}{\|\Delta\mathbf{x}\|}=\mathbf{0}$$

则$\mathbf{f}$的每个坐标函数在$\mathbf{x}_0$处的每个偏导数

$$\left.\frac{f_i\left(\mathbf{x}\right)}{\partial x_j}\right|_{\mathbf{x}=\mathbf{x}_0}\text{ﬁ}i=1,\cdots,m,j=1,\cdots,n$$

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

$$\mathbf{L}\hat{\mathbf{e}}_j=\sum_{i=1}^{m}\left(\left.\frac{\partial f_i\left(\mathbf{x}\right)}{\partial x_j}\right|_{\mathbf{x}=\mathbf{x}_0}\right)\hat{\mathbf{u}}_i,j=1,\cdots,n$$

证明. 见附录。 □

*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”[6]p. 20。）

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 II.12.4. 设函数$\mathbf{f}:\mathbb{R}^n \supseteq D \to \mathbb{R}^m$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II.12.3和II.12.4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[6] "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 II.12.8 (向量函数的导数)**.** 设函数$\mathbf{f}:\mathbb{R}^n \supseteq D \to \mathbb{R}^m$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），记为

$$\mathbf{L} \equiv \left.\frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0}$$

$\mathbf{L}$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）[6] "例3" ,p. 23。

定理 II.12.5. 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。 □

我们看到，函数导数的定义引入了类似微商的记法，但是我们没有定义过“矢量的无穷小”和“矢量的除法”。这里我们作一些说明。若函数$\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$在$D\subset\mathbb{R}^n$上可微分，其在某点$\mathbf{x}_0\in D$处的全微分是由点$\mathbf{x}_0$引出的函数$d\mathbf{f}_{\mathbf{x}_0}:\mathbb{R}^n\to\mathbb{R}^m, d\mathbf{f}_{\mathbf{x}_0}(\mathbf{x})=\mathbf{L}_{\mathbf{x}_0}\mathbf{x},\forall\mathbf{x}\in\mathbf{R}^n$*，其中$\mathbf{L}_{\mathbf{x}_0}$是函数$\mathbf{f}$在$\mathbf{x}_0$处的导数。但是我们直接看到这个引出的函数就是函数的导数$\mathbf{L}_{\mathbf{x}_0}$本身。若视$\mathbf{x}\in D$为$\mathbb{R}^n$上的恒等变换，即$\mathbf{x}(\mathbf{p})=\mathbf{p},\forall\mathbf{p}\in\mathbb{R}^n$，则由点$\mathbf{p}_0$上的全微分引出的函数$d\mathbf{x}_{\mathbf{p}_0}:\mathbb{R}^n\to\mathbb{R}^n$满足$d\mathbf{x}_{\mathbf{p}_0}(\mathbf{u})=\mathbf{M}_{\mathbf{p}_0}\mathbf{u}=\mathbf{I}_n\mathbf{u}=\mathbf{u}$，其中$\mathbf{M}_{\mathbf{p}_0}$是$\mathbf{x}$在$\mathbf{p}_0$处的导数，显然这个导数总为恒等线性变换$\mathbf{I}_n$，故$d\mathbf{x}_{\mathbf{p}_0}$也是$\mathbb{R}^n$上的恒等映射。我们把它的分量写出来：

$$d\mathbf{x}_{\mathbf{p}_0}(\mathbf{u})=\begin{pmatrix}dx_{1,\mathbf{p}_0}(\mathbf{u})\\ \vdots\\ dx_{n,\mathbf{p}_0}(\mathbf{u})\end{pmatrix}=\begin{pmatrix}u_1\\ \vdots\\ u_n\end{pmatrix}$$

其中$dx_{i,\mathbf{p}_0}(\mathbf{u})=u_i,i=1,\cdots,n$是取坐标函数，且它们都是当$\mathbf{u}\to\mathbf{0}$时的无穷小。因此$d\mathbf{x}_{\mathbf{p}_0}$是由这$n$个无穷小量组成的无穷小向量。现再考虑复合函数$\mathbf{f}\circ\mathbf{x}$在点$\mathbf{p}_0$处的全微分引出的函数，它满足*

$$d(\mathbf{f}\circ\mathbf{x})_{\mathbf{p}_0}(\mathbf{u})=\mathbf{L}_{\mathbf{x}(\mathbf{p}_0)}\mathbf{M}_{\mathbf{p}_0}\mathbf{u}=\mathbf{L}_{\mathbf{x}(\mathbf{p}_0)}d\mathbf{x}_{\mathbf{p}_0}(\mathbf{u}),\forall\mathbf{u}\in\mathbb{R}^n$$

由于线性变换是连续函数，故可证函数$d(\mathbf{f}\circ\mathbf{x})_{\mathbf{p}_0}(\mathbf{u})$与函数$\mathbf{x}_{\mathbf{p}_0}(\mathbf{u})$当$\mathbf{u}\to\mathbf{0}$时都趋于零，且对任意$\mathbf{p}_0,\mathbf{u}\in\mathbb{R}^n$均成立，故可略去$\mathbf{p}_0$和$\mathbf{u}$简记为

$$d\mathbf{f}(\mathbf{x})=\mathbf{L}_{\mathbf{x}}d\mathbf{x}$$

其中

$$d\mathbf{x}=\begin{pmatrix}dx_1\\ \vdots\\ dx_n\end{pmatrix}$$

该式就是函数的微分的定义式，由上述的讨论我们知道这个式子表达了一个无穷小向量与另一个无穷小向量之间的线性关系。

我们进一步把$\mathbf{L}_{\mathbf{x}}$记为$\frac{d\mathbf{f}}{d\mathbf{x}}$确实是滥用了符号，因为线性变换并非“两个向量的商”。但这种记法是很多资料都使用的惯例。与$\mathbf{L}_{\mathbf{x}}$相比更利于直接告诉我们这是一个函数的导数而无需另作文字说明，因此本讲义也经常使用这种记法，即

$$d\mathbf{f}(\mathbf{x})=\frac{d\mathbf{f}}{d\mathbf{x}}d\mathbf{x}$$

*为了讲述简洁这里略去了对函数$\mathbf{f}$在一点处可微分与相应的增量与该点邻域的关系问题。

*这里提前用到了复合函数求导法则。

II.12.3 向量函数的导函数、连续可微函数

定义 II.12.9 (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$\mathbf{L}:S\to\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$,

$$\mathbf{L}(\mathbf{x})\equiv\left.\frac{d\mathbf{f}(\mathbf{x}')}{d\mathbf{x}'}\right|_{\mathbf{x}'=\mathbf{x}},\forall\mathbf{x}\in S$$

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。

由定理II.12.4知，导函数是单射。

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II.11.3，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$\|\mathbf{L}\mathbf{x}\|\leq k\|\mathbf{x}\|,\forall\mathbf{x}\in\mathbb{R}^n$。于是我们定义——

定义 II.12.10 (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$\|\mathbf{L}\mathbf{x}\|\leq k\|\mathbf{x}\|,\forall\mathbf{x}\in\mathbb{R}^n$的正实数$k$的最大下界（infimum），即

$$\|\mathbf{L}\|\equiv\inf\{k|k>0,\|\mathbf{L}\mathbf{x}\|\leq k\|\mathbf{x}\|\,\forall\mathbf{x}\}$$

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IV.1.1），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：$\|\mathbf{L}\mathbf{x}\|\leq\|\mathbf{L}\|\,\|\mathbf{x}\|$。

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$\epsilon>0$总存在$\delta>0$使得只要$\|\mathbf{x}-\mathbf{x}_0\|<\delta,\mathbf{x},\mathbf{x}_0\in S$就有$\|\mathbf{L}(\mathbf{x})-\mathbf{L}(\mathbf{x}_0)\|<\epsilon$。

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 II.12.11 (连续可微函数). 若函数$\mathbf{f}:\mathbb{R}^n\supseteq D\to\mathbb{R}^m$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。

推论 II.12.5.1. 函数$\mathbf{f}:\mathbb{R}^n\supseteq D\to\mathbb{R}^m$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II.12.5可知函数连续可微是函数可微的充份非必要条件。

II.12.4 对向量的导数与方向导数

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：

$$\lim_{t\to 0}\frac{\mathbf{f}(\mathbf{x}_i)-\mathbf{f}(\mathbf{x}_0)}{t},i=1,\cdots,n$$

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$\Leftrightarrow$

$$\lim_{t\to 0}\frac{\mathbf{f}(\mathbf{x}_0+t\hat{\mathbf{e}}_i)-\mathbf{f}(\mathbf{x}_0)}{t},i=1,\cdots,n$$

现在我们把$\hat{\mathbf{e}}_j$改为任意向量$\mathbf{y}$，引入方向导数。

定义 II.12.12 (对向量的导数、方向导数)**.** 函数$\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$对向量$\mathbf{y}\in\mathbb{R}^n$的导数是

$$\frac{\partial\mathbf{f}(\mathbf{x})}{\partial\mathbf{y}}=\lim_{t\to 0}\frac{\mathbf{f}(\mathbf{x}+t\mathbf{y})-\mathbf{f}(\mathbf{x})}{t}$$

其中作为导函数的$\frac{\partial\mathbf{f}(\mathbf{x})}{\partial\mathbf{y}}$的定义域是原函数$\mathbf{f}$的定义域的使该导数存在的子集。特别地，函数对单位向量$\mathbf{u},\|\mathbf{u}\|=1$的导数称为方向导数（directional derivative）。

以下定理使得函数对任意向量的导数可用该函数的全导数来计算。

定理 II.12.6. 设函数$\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$在$\mathbf{x}\in\mathbb{R}^n$处可微，则

$$\frac{\partial\mathbf{f}(\mathbf{x})}{\partial\mathbf{y}}=\frac{d\mathbf{f}(x)}{d\mathbf{x}}\mathbf{y},\forall\mathbf{y}\in\mathbb{R}^n$$

证明. 对$\mathbf{y}=\mathbf{0}$显然成立。若$\mathbf{y}\neq\mathbf{0}$，由全微分的近似意义，由自变量的增量$t\mathbf{y}$造成的函数增量$\mathbf{f}(\mathbf{x}+t\mathbf{y})-\mathbf{f}(\mathbf{x})$满足

$$\begin{aligned}\lim_{t\to 0}\frac{\mathbf{f}(\mathbf{x}+t\mathbf{y})-\mathbf{f}(\mathbf{x})-\frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}}(t\mathbf{y})}{\|t\mathbf{y}\|}&=\mathbf{0}\\ \Leftrightarrow\lim_{t\to 0}\frac{1}{\|\mathbf{y}\|}\left\|\frac{\mathbf{f}(\mathbf{x}+t\mathbf{y})-\mathbf{f}(x)}{t}-\frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}}\mathbf{y}\right\|&=\mathbf{0}\\ \Leftrightarrow\lim_{t\to 0}\frac{\mathbf{f}(\mathbf{x}+t\mathbf{y})-\mathbf{f}(\mathbf{x})}{t}&=\frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}}\mathbf{y}\end{aligned}$$

□

我们比较一个函数$\mathbf{f}(\mathbf{x})$在$\mathbf{x}_0$处的全微分及其在$\mathbf{x}_0$处对某向量$\mathbf{y}$的导数可以发现后者就是令前者中的$\mathbf{x}-\mathbf{x}_0=\mathbf{y}$。在全微分中，我们强调的是对任意$\mathbf{x}$（即点$\mathbf{x}_0$邻近的每处），而在对$\mathbf{y}$的导数中我们强调的是$\mathbf{x}_0$朝某个选定的向量$\mathbf{y}$的某处（$\mathbf{x}=\mathbf{x}_0+\mathbf{y}$）。留意到，函数在某点处对标准基向量的导数就是函数对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II.12.6表明，拿一个函数在某点处的全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。这也是函数的全导数的重要几何意义。

II.12.5 复合函数求导的链式法则、反函数定理、隐函数定理

定理 II.12.7 (复合函数求导的链式法则). 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，且其全导数

$$\left.\frac{d\mathbf{g}\circ\mathbf{f}(\mathbf{x})}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0}=\left.\frac{d\mathbf{g}(\mathbf{y})}{d\mathbf{y}}\right|_{\mathbf{y}=\mathbf{f}(\mathbf{x}_0)}\left.\frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0}$$

证明. 见附录。 □

定理 II.12.8 (反函数定理). 设函数$\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^n$是连续可微函数，且在$\mathbf{x}_0$处其导数$\mathbf{L}(\mathbf{x}_0)$可逆。则在$\mathbf{x}_0$的某邻域$N$上，函数$\mathbf{f}$有连续可微的逆函数$\mathbf{f}^{-1}$；$N$的像集$\mathbf{f}(N)$是开集；且$\mathbf{f}^{-1}$在$\mathbf{f}(\mathbf{x}_0)$处的导数是$\mathbf{f}$在$\mathbf{x}_0$的导数的逆变换，即

$$\left.\frac{d\mathbf{f}^{-1}(\mathbf{y})}{d\mathbf{y}}\right|_{\mathbf{y}=\mathbf{f}(\mathbf{x}_0)}=\mathbf{L}^{-1}(\mathbf{x}_0)$$

证明. 见附录 □

如果$\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$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

$$\left.\frac{\partial\mathbf{F}(\mathbf{x},\mathbf{y})}{\partial\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0,\mathbf{y}=\mathbf{y}_0}=\left.\frac{d\mathbf{F}(\mathbf{x},\mathbf{y}_0)}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0}$$

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 II.12.1. 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

$$\begin{aligned}\frac{d\mathbf{f}(\mathbf{g}(\mathbf{x}),\mathbf{x})}{d\mathbf{x}}&=\frac{d\mathbf{h}(\mathbf{x})}{d\mathbf{x}}\\&=\left.\frac{d\mathbf{f}(\mathbf{y})}{d\mathbf{y}}\right|_{\mathbf{y}=\mathbf{h}(\mathbf{x})}\frac{d\mathbf{h}(\mathbf{x})}{d\mathbf{x}}\\&=\begin{pmatrix}\frac{\partial f_1}{\partial h1} & \cdots & \frac{\partial f_1}{\partial h_n} & \frac{\partial f_1}{\partial h_{n+1}} & \cdots & \frac{\partial f_1}{\partial h_{n+m}}\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ \frac{\partial f_p}{\partial h_1} & \cdots & \frac{\partial f_p}{\partial h_n} & \frac{\partial f_p}{\partial h_{n+1}} & \cdots & \frac{\partial f_p}{\partial h_{n+m}}\end{pmatrix}\begin{pmatrix}\frac{\partial h_1}{\partial y_1} & \cdots & \frac{\partial h_1}{\partial y_m}\\ \vdots & \ddots & \vdots\\ \frac{\partial h_n}{\partial y_1} & \cdots & \frac{\partial h_1}{\partial y_m}\\ \frac{\partial h_{n+1}}{\partial y_1} & \cdots & \frac{\partial h_{n+1}}{\partial y_m}\\ \vdots & \ddots & \vdots\\ \frac{\partial h_{n+m}}{\partial y_1} & \cdots & \frac{\partial h_{n+m}}{\partial y_m}\end{pmatrix}=C\end{aligned}$$

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

$$\begin{aligned}C_{ij}&=\sum_{k=1}^{n}\frac{\partial f_i}{\partial g_k}\frac{\partial g_k}{\partial x_j}+\sum_{j=1}^{m}\frac{\partial f_i}{\partial x_j}\\ \Leftrightarrow\frac{d\mathbf{f}(\mathbf{g}(\mathbf{x}),\mathbf{x})}{d\mathbf{x}}&=\left.\frac{\partial\mathbf{f}(\mathbf{y},\mathbf{x})}{\partial\mathbf{y}}\right|_{\mathbf{y}=\mathbf{g}(\mathbf{x})}\frac{\partial\mathbf{g}(\mathbf{x})}{\partial\mathbf{x}}+\frac{\partial\mathbf{f}(\mathbf{y},\mathbf{x})}{\partial\mathbf{x}}\end{aligned}$$

定理 II.12.9 (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

- $\mathbf{F}(\mathbf{x}_0,\mathbf{y}_0)=\mathbf{0}$；
- 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

$$\left.\frac{d\mathbf{f}\left(\mathbf{x}\right)}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}}=-\left[\left.\frac{\partial\mathbf{F}\left(\mathbf{x},\mathbf{y}\right)}{\partial\mathbf{y}}\right|_{\mathbf{x}=\mathbf{x},\mathbf{y}=\mathbf{f}(\mathbf{x})}\right]^{-1}\left.\frac{\partial\mathbf{F}\left(\mathbf{x},\mathbf{y}\right)}{\partial\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x},\mathbf{y}=\mathbf{f}(\mathbf{x})}$$

上式中的“-1”是指线性变换的逆。

证明. 待补充[7]p. 593。 □

例 II.12.2. 设函数$\mathbf{F}:\mathbb{R}^4\to\mathbb{R}^2$，$\mathbf{F}\left(u,v,x,y\right)=\left(F_1,F_2\right)$，$\mathbf{F}=\mathbf{0}$隐含定义了$\left(x,y\right)=\mathbf{f}\left(u,v\right),\mathbf{f}:\mathbf{R}^2\to\mathbf{R}^2$，则

$$\begin{aligned}
&\frac{\partial\mathbf{F}}{\partial u}=\mathbf{0},\frac{\partial\mathbf{F}}{\partial v}=\mathbf{0}\\
\Leftrightarrow&\begin{cases}\frac{\partial F_1}{\partial u}+\frac{\partial F_1}{\partial x}\frac{\partial x}{\partial u}+\frac{\partial F_1}{\partial y}\frac{\partial y}{\partial u}=0\\ \frac{\partial F_2}{\partial u}+\frac{\partial F_2}{\partial x}\frac{\partial x}{\partial u}+\frac{\partial F_2}{\partial y}\frac{\partial y}{\partial u}=0\end{cases},\begin{cases}\frac{\partial F_1}{\partial v}+\frac{\partial F_1}{\partial x}\frac{\partial x}{\partial v}+\frac{\partial F_1}{\partial y}\frac{\partial y}{\partial v}=0\\ \frac{\partial F_2}{\partial v}+\frac{\partial F_2}{\partial x}\frac{\partial x}{\partial v}+\frac{\partial F_2}{\partial y}\frac{\partial y}{\partial v}=0\end{cases}\\
\Leftrightarrow&\begin{pmatrix}\frac{\partial F_1}{\partial u}&\frac{\partial F_1}{\partial v}\\ \frac{\partial F_2}{\partial u}&\frac{\partial F_2}{\partial v}\end{pmatrix}_+\begin{pmatrix}\frac{\partial F_1}{\partial x}&\frac{\partial F_1}{\partial y}\\ \frac{\partial F_2}{\partial x}&\frac{\partial F_2}{\partial y}\end{pmatrix}\begin{pmatrix}\frac{\partial x}{\partial u}&\frac{\partial x}{\partial v}\\ \frac{\partial y}{\partial u}&\frac{\partial y}{\partial v}\end{pmatrix}=0\\
\Leftrightarrow&\begin{pmatrix}\frac{\partial x}{\partial u}&\frac{\partial x}{\partial v}\\ \frac{\partial y}{\partial u}&\frac{\partial y}{\partial v}\end{pmatrix}=-\begin{pmatrix}\frac{\partial F_1}{\partial x}&\frac{\partial F_1}{\partial y}\\ \frac{\partial F_2}{\partial x}&\frac{\partial F_2}{\partial y}\end{pmatrix}^{-1}\begin{pmatrix}\frac{\partial F_1}{\partial u}&\frac{\partial F_1}{\partial v}\\ \frac{\partial F_2}{\partial u}&\frac{\partial F_2}{\partial v}\end{pmatrix}
\end{aligned}$$

II.13 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[6]第九章，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$\mathbb{R}^3$一一对应。

II.13.1 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$\mathbf{g}\left(t\right),t\in I$为参数方程的像集$\mathbf{g}\left(I\right)$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$\mathbf{g}\left(t_0\right)$处的切向量。可定义$\hat{\mathbf{t}}\equiv\frac{d\mathbf{g}/dt}{\|d\mathbf{g}/dt\|}$为曲线$\mathscr{C}$的单位切向量（unit tangent vector）。

曲线有两种弧微元：$d\mathbf{l}\equiv\left(d\mathbf{g}/dt\right)dt$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$d\mathbf{l}=\hat{\mathbf{t}}dl$。

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[6]p. 133,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[6]p. 140,定理9.2.1

$$\int_{\mathscr{C}} \mathbf{h}(\mathbf{g}) \cdot d\mathbf{l} = \int_a^b \mathbf{h}(\mathbf{g}(t)) \cdot \frac{d\mathbf{g}}{dt} dt = \int_{\mathscr{C}} \mathbf{h} \cdot \hat{\mathbf{t}} dl$$

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

$$W(\mathscr{C}) = \int_{\mathscr{C}} \mathbf{F} \cdot d\mathbf{l} = \int_a^b \mathbf{F}(\mathbf{g}(t), t) \cdot \frac{d\mathbf{g}}{dt} dt$$

II.13.2 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

$$\left.\frac{d\mathbf{g}(\mathbf{u})}{d\mathbf{u}}\right|_{\mathbf{u}=\mathbf{u}_0}$$

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，

$$\hat{\mathbf{n}} \equiv \left.\left(\frac{\frac{\partial \mathbf{g}}{\partial u_1} \times \frac{\partial \mathbf{g}}{\partial u_2}}{\left\|\frac{\partial \mathbf{g}}{\partial u_1} \times \frac{\partial \mathbf{g}}{\partial u_2}\right\|}\right)\right|_{\mathbf{u}=\mathbf{u}_0}$$

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$d\boldsymbol{\sigma}=\hat{\mathbf{n}}d\sigma$。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

$$A\left(\mathscr{S}\right)=\int_{\mathscr{S}}d\sigma=\int_D\left\|\frac{\partial\mathbf{g}}{\partial u_1}\times\frac{\partial\mathbf{g}}{\partial u_2}\right\|d\sigma_D$$

“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[6]p. 165,定理9.4.1

$$\int_{\mathscr{S}}\mathbf{f}\left(\mathbf{g}\right)d\sigma=\int_D\mathbf{f}\left(\mathbf{g}\left(\mathbf{u}\right)\right)\left\|\frac{\partial\mathbf{g}}{\partial u_1}\times\frac{\partial\mathbf{g}}{\partial u_2}\right\|d\sigma_D$$

“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[6]p. 170,第五节

$$\int_{\mathscr{S}}\mathbf{h}\cdot d\boldsymbol{\sigma}=\int_D\mathbf{h}\left(\mathbf{g}\left(\mathbf{u}\right)\right)d\sigma_D=\int_{\mathscr{S}}\mathbf{h}\cdot\hat{\mathbf{n}}d\sigma$$

II.13.3　积分换元公式[6]p. 116,“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$\Omega=\mathbf{g}\left(\mathscr{U}\right)$是一个经过形变后的三维区域。当且仅当导数

$$\left.\frac{d\mathbf{g}\left(\mathbf{r}\right)}{d\mathbf{r}}\right|_{\mathbf{r}=\mathbf{r}_0}$$

存在且满秩时，区域Ω在点$\mathbf{r}_0$处是光滑的。此时

$$\left|\frac{\partial\mathbf{g}}{\partial r_3}\cdot\left(\frac{\partial\mathbf{g}}{\partial r_1}\times\frac{\partial\mathbf{g}}{\partial r_2}\right)\right|\equiv\left|\det\left(\frac{\partial\mathbf{g}}{\partial\mathbf{r}}\right)\right|$$

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。$dV_\Omega=\left|\det\left(d\mathbf{g}/d\mathbf{r}\right)\right|dV_{\mathscr{U}}$是$\mathscr{U}$的体积元$dV_{\mathscr{U}}$与$\Omega$的体积元$dV_\Omega$之间的转换关系。区域$\Omega$的体积

$$V\left(\Omega\right)=\int_\Omega dV_\Omega=\int_{\mathscr{U}}\left|\det\left(\frac{d\mathbf{g}}{d\mathbf{r}}\right)\right|dV_{\mathscr{U}}$$

“单位体积的性质”$\mathbf{f}:\mathbb{R}^3\supset\mathscr{U}\to\mathbb{R}^3$在区域$\Omega$的总和为

$$\mathbf{F}\left(\Omega\right)=\int_\Omega\mathbf{f}\left(\mathbf{g}\right)dV_\Omega=\int_{\mathscr{U}}\mathbf{f}\left(\mathbf{g}\left(\mathbf{r}\right)\right)\left|\det\left(\frac{d\mathbf{g}}{d\mathbf{r}}\right)\right|dV_{\mathscr{U}}$$

上式就是积分换元公式。

II.13.4　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：

$$\int_D\left(\frac{\partial Q}{\partial x_1}-\frac{\partial P}{\partial x_2}\right)d\sigma_D=\int_{\partial D}Pdx_1+Qdx_2$$

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

$$\mathrm{curl}\mathbf{F} = \left(\frac{\partial F_3}{\partial x_2} - \frac{\partial F_2}{\partial x_3}, \frac{\partial F_1}{\partial x_3} - \frac{\partial F_3}{\partial x_1}, \frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2} \right)^\intercal$$

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$\Omega$中的连续可微向量场,，则有高斯定理

$$\int_\Omega \mathrm{div}\mathbf{F} dV_\Omega = \int_{\partial \Omega} \mathbf{F} \cdot d\boldsymbol{\sigma}$$

第三部分 连续介质力学基础

III.1 标架与参考系

III.1.1 新古典时空

在经典力学中，我们认为物理事件的发生是不依赖于人的观察与否或观察方式的客观事实。我们希望，同一件物理事件，经不同的观察者观察后作出的报告，能通过一套互译系统而统一成一个结论；不同观察者通过对实验观测的总结归纳，可以达成一个不依赖具体观察者的情境和观察方式的、关于大自然规律性的共识。这种情况称为物理客观性*。

经典力学中的观察者对时间的感受依赖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。

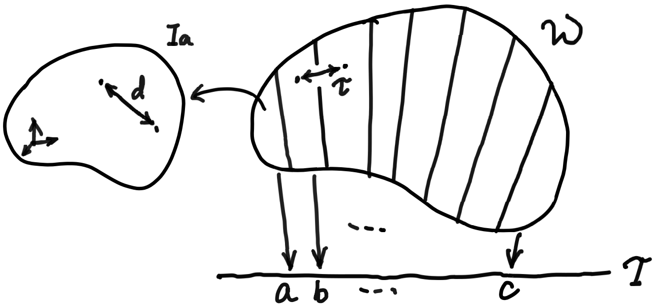

图 III.1.1: 事件世界与同时等价类示意图

定义 III.1.1 (事件世界、时间). $\mathscr{W}$是一个非空集，其元素$w \in \mathscr{W}$称为一个事件。我们给予$\mathscr{W}$一个函数$\tau : \mathscr{W}^2 \to \mathbb{R}$，满足：

- $\forall a, b \in \mathscr{W} : \tau(a, b) = -\tau(b, a)$
- $\forall a, b, c \in \mathscr{W} : \tau(a, b) + \tau(b, c) = \tau(a, c)$
- $\forall a \in \mathscr{W}, t \in \mathbb{R}, \exists b \in \mathscr{W} : \tau(a, b) = t$

我们称τ为（两事件的）时间间隔或时长。我们还能称

- 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同时当且仅当$\tau(a, b) = 0$

上面的第三条给出了两事件的同时性。易验同时性是一个等价关系，即$S = \{(a, b) \,|\, \tau(a, b) = 0\}$满

*科学研究的客观性有不同层面的意义[8]。本讲义中讨论的“客观性”概念仅限于经典力学理论基础的范畴。

足自反性、传递性、对称性。这一等价关系将事件世界$\mathscr{W}$划分为等价类，即：

$$\mathscr{W} = \bigcup_{a \in \mathscr{T}} I_a, I_a = \{x \in \mathscr{W} | \tau(a, x) = 0\}$$

其中等价类I_a称为时刻a对应的同时等价类，它是所有与事件a同时的事件的子集，而a在此变成了时刻的标记，集合$\mathscr{T}$称为时间，是所有（不同）时刻的集合（如图III.1.1所示）。

我们还给予事件世界$\mathscr{W}$以一个度量$d : I_a \times I_a \to \mathbb{R}, a \in \mathscr{T}$，称两事件间的“距离”，它还要满足：

- d是I_a上的一个欧几里德度量，故I_a是一个欧几里德空间，带有一个平移空间$\mathscr{V}_a$。
- $\forall a \in \mathscr{T} : \dim \mathscr{V}_a = 3$

注意到，这一度量只能作用在同时发生的两个事件上的（如图III.1.1所示）。

结合度量τ的定义，$(\mathscr{T}, |\tau|)$形成一个度量空间。在此度量空间上的等距变换拥有之前介绍过的一切性质。值得注意的是，由τ的定义，可知$\mathscr{T}$的平移空间是一维的、完备的。我们可直接使用实数集$\mathbb{R}$作为其平移空间。选定某时刻$t_0 \in \mathscr{T}$作为原点，$\mathbb{R}$的正或负作为时间流逝的方向以及单位时长（相当于在作为向量空间的$\mathbb{R}$中选择了一个规范正交基），则由等距变换的表示定理（定理II.9.1）对任一$\mathscr{T}$上的等距变换$i : \mathscr{T} \to \mathscr{T}$有$i(t) = i(t_0) + Q(t - t_0)$。其中$Q = \pm 1$时间正方向的选择。由于我们很少遇到两个观察者记录时间的方式是反号的情况，因此常常只讨论（默认）$Q = 1$，即

$$i(t) = i(t_0) + (t - t_0)$$

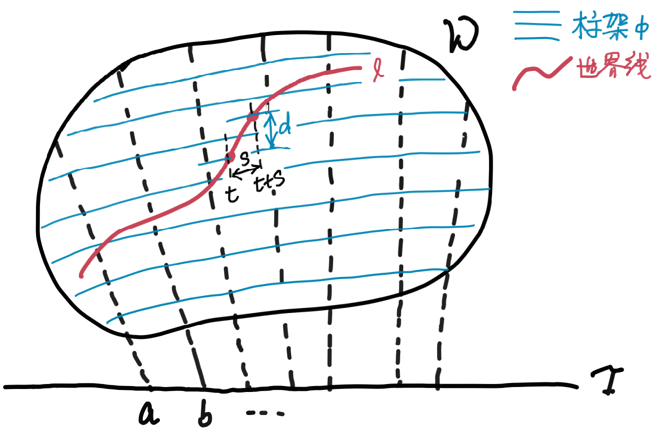

图 III.1.2: 世界线及事件世界的标架示意图

III.1.2 世界线与标架

引入世界线的概念。

定义 III.1.2 (世界线). 世界线是由时间$\mathscr{T}$到事件世界$\mathscr{W}$的单射$l:\mathscr{T}\to\mathscr{W}$。

一条世界线$l\left(t\right),t\in\mathscr{T}$的任意两个值都是属于不同时刻的事件（如图III.1.2所示）。一条世界线是某质点运动轨迹的客观存在（即不依赖人的观察与否与观察方式的独立概念）。

依赖世界线的概念，我们可以引入“事件世界的标架”的概念。

定义 III.1.3 (事件世界的标架). 事件世界的标架ϕ是一组“平行”的世界线$\{l,\cdots\}$（如图III.1.2所示），其元素$l\in\phi$称为标架线。“平行”的意思是对任意两时刻$t_1,t_2\in\mathscr{T}$两条世界线$l_1,l_2\in\phi$有$d\left(l_1\left(t_1\right),l_2\left(t_!\right)\right)=d\left(l_1\left(t_2\right),l_2\left(t_2\right)\right)$。一个标架的所有标架线还要覆盖整个事件世界的所有事件，即

$$\bigcup_{l_i\in\phi}\mathrm{ran}l_i=\mathscr{W}$$

使得没有一个事件是不属于任何标架线的。映射$l_\phi:\mathscr{W}\to\phi,l_\phi\left(a\right)=\{l|l\in\phi,a\in\mathrm{ran}l\}\,\forall a\in\mathscr{W}$是为任一事件找到其所属的标架线的映射。由标架线“平行”的性质易验$\{l|l\in\phi,a\in\mathrm{ran}l\}$有且只有一个元素，即任一事件$a\in\mathscr{W}$对应且只对应一条标架线，且有$d\left(l_\phi\left(a\right)\left(t\right),l_\phi\left(b\right)\left(t\right)\right)=$常数$\forall a,b\in\mathscr{W},t\in\mathscr{T}$，故可由此定义“标架线之间的距离”：

$$d\left(l_\phi\left(a\right),l_\phi\left(b\right)\right)=d\left(l_\phi\left(a\right)\left(t\right),l_\phi\left(b\right)\left(t\right)\right),t\in\mathscr{T}$$

映射l_ϕ还具有性质$l_\phi\left(l_\phi\left(a\right)\left(t\right)\right)=l_\phi\left(a\right)\forall t\in\mathscr{T},a\in\mathscr{W}$。

总结标架的特点：

- 由于标架线是世界线，故它贯穿了不同时刻；
- 任一事件必属于某标架线；
- 我们可以谈论两条标架线之间的距离；

可知，标架的重要意义是使得我们可以讨论不同时刻的任意两个事件$a\in I_{t_1},b\in I_{t_2},t_1,t_2\in\mathscr{T}$之间的距离，选定标架$\phi$，$a$与$b$的距离就是$d\left(l_\phi\left(a\right),l_\phi\left(b\right)\right)$。

例如，既然一条世界线所代表的某质点的运动过程，那么我们直觉上会想要通过形如下式的导数来定义“速度”：

$$\lim_{s\to0}\frac{l\left(t+s\right)-l\left(t\right)}{s}$$

但是，上式的分子中相减的两个事件是属于不同时刻的同时等价类的（$l\left(t+s\right)\in I_{t+s},l\left(t\right)\in I_t,t+s,t\in\mathscr{T}$，如图III.1.2所示），因此这个减法的意义是没有定义的。通过事件世界的标架ϕ，我们可以解决这个问题，从而定义速度$\mathbf{v}:\mathscr{T}\to\mathscr{V}_\phi$：

$$\mathbf{v}\left(t_0\right)=\lim_{t\to t_0}\frac{l_\phi\left(l\left(t\right)\right)-l_\phi\left(l\left(t_0\right)\right)}{t}$$

上式的速度定义利用了(ϕ, d)是一个欧几里德度量空间的事实，分子中相减的两个元素是标架ϕ中的两条世界线。它们的差对应$\mathscr{V}_\phi$中的一个平移向量。故速度是一个向量且依赖标架ϕ的选择。注意到，在速度的定义过程中我们默认了一切需要的连续性。加速度也可以类似地被定义，且由于相同的原因加速度也依赖标架ϕ的选择。我们称标架ϕ的选择是主观的。依赖ϕ定义的概念都是不具有物理客观性的（除非增加某些限定条件）。

考虑任一时刻$t \in \mathscr{T}$对应的同时等价类I_t。选定某标架ϕ后，易证I_t中的任意两个不同的事件不可能属于同一条标架线，反之亦然。故映射l_ϕ限定在某同时等价类I_t上是双射。同为欧几里德度量空间的(I_t, d)和(ϕ, d)，它们对应的平移空间$\mathscr{V}_t, \mathscr{V}_\phi$同构，故$\dim \mathscr{V}_\phi = \dim \mathscr{V}_t = 3$。对于一系列不同时刻的同时等价类$I_{t_1}, I_{t_2}, \cdots$，它们各自都对应一个平移空间$\mathscr{V}_{t_1}, \mathscr{V}_{t_2}, \cdots$，且$\dim \mathscr{V}_{t_1} = \dim \mathscr{V}_{t_2} = \cdots = \dim \mathscr{V}_\phi$，即标架的平移空间$\mathscr{V}_\phi$与每个时刻的同时等价类的平移空间都同构。我们于是可以统一采用$\mathscr{V}_\phi$作为每个时刻的同时类的平移空间，不同时刻下的事件的位置，我们都用统一的一套平移向量描述。

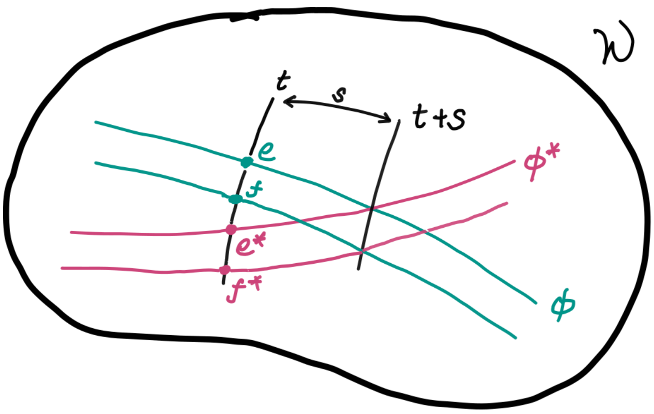

图 III.1.3: 事件世界的标架变换示意图

考虑事件世界$\mathscr{W}$中的两组不同的标架ϕ, ϕ^*。事件$e \in I_t$的时刻是$t \in \mathscr{T}$，其所属的ϕ中的标架线是$l_\phi(e)$。考虑距时刻t间隔为$s \in \mathbb{R}$的另一时刻$t+s$，该时刻在标架线$l_\phi(e)$上的事件是$l_\phi(e)(t+s)$。该事件在标架ϕ^*中所属的标架线是$l_{\phi^*}(l_\phi(e)(t+s))$，在该标架线上$t$时刻的事件是$l_{\phi^*}(l_\phi(e)(t+s))(t) \in I_t$（如图III.1.3所示）。令

$$e^* = l_{\phi^*}(l_\phi(e)(t+s))(t)$$

则有$l_{\phi^*}(e^*) = l_{\phi^*}(l_\phi(e)(t+s))$。设$e, f \in I_t, e \neq f$，则有

$$d(e^*, f^*) = d(l_{\phi^*}(l_\phi(e)(t+s)), l_{\phi^*}(l_\phi(f)(t+s)))$$

由恒等关系$l_{\phi^*}\left(l_\phi\left(e\right)\left(t\right)\right)\left(t\right)=l_\phi\left(e\right)\left(t\right)\forall e\in\mathscr{W},t\in\mathscr{T}$，上式$\Leftrightarrow$

$$\begin{aligned}d\left(l_\phi\left(e\right)\left(t+s\right),l_\phi\left(f\right)\left(t+s\right)\right)&=d\left(l_\phi\left(e\right)\left(t\right),l_\phi\left(f\right)\left(t\right)\right)\\&=d\left(e,f\right)\end{aligned}$$

因此，对任意$e,f\in\mathscr{W}$，$d\left(l_\phi\left(e\right),l_\phi\left(f\right)\right)=d\left(l_{\phi^*}\left(e^*\right),l_{\phi^*}\left(f^*\right)\right)$，即由$\phi$到$\phi^*$存在一个等距映射。由等距映射的表示定理，选定$\phi$中的一标架线$l_0$，则$l*(s)=l_0^*\left(s\right)+\mathbf{Q}\left(s\right)\left(l-l_0\right),l\in\phi,l^*\in\phi^*$，其中$l_0^*=l_{\phi^*}\left(l_\phi\left(l_0\right)\left(t+s\right)\right)$。我们称这一变换为一个标架变换。一个标架变换不仅包括了上述的标架线之间的等距变换，还隐含包括了一个$\mathscr{T}$的等距变换（即选定了参考时刻t并将任一时刻表示为$t+s$）。

上述关于事件世界、世界线、标架的概念构建方式，代表着我们观测物理事件的实际方式。一位观察者总是把所观察到的空间理解为3维欧几里德空间。他用一把直尺测量同一时刻两物理事件的距离，用一个时钟测量两物理事件的时间间隔。在某时刻下，选定空间中一点作为参照点和过该点三条两两不共线的射线方向作为参考方向——即坐标系，则空间中任一点都能通过从参照点到该点的方向和距离（即平移向量）的得到唯一标示。选择某一时刻作为参考时刻，则任一时刻都可通过与参考时刻之间的时长得到标示。坐标系与这一参考时刻的选择，统称参考系。一个观察者在某时刻选择的坐标系会在其他时刻延用；这相当于选定了一个标架ϕ并实际统一采用$\mathscr{V}_\phi$中的向量作来任一时刻下的平移向量。

两名观察者，观察同一事件，他们首先要相互对表，即在某时刻两人用各自的钟表同时开始计时，经过相同的时长后，两人得出双方钟表读数的差别。这是在时间$\mathscr{T}$内的一个等距变换。对完表之后，两人选取同一时刻作为参考时刻，对同一物理事件进行观察。由于两人选取的坐标系可能是不同的，因此两人还需要在同一时刻确认对方所找的坐标系；然后，其中一人除了要观察所关注的物理事件在自己所选择的坐标系下随时间的变化之外，还要观察对方选择的坐标系在自己选择的坐标系下随时间的变化，才能知道任一时刻，同一物理事件在对方坐标系下的表示。每个时刻，两个坐标系的之间的关系都可能是不同的，故它们之间的转换是依赖时刻的等距变换。

III.1.3 标架的简化定义

一名观察者如果选定了参考时刻t_0和标架ϕ，则任一时刻都能对应$\mathbb{R}$中的一个实数。选定$\mathscr{V}_\phi$中的一向量$\hat{\mathbf{o}}$和一组基$\{\hat{\mathbf{e}}_i\}$，则任一时刻发生的物理事件的位置都能对应到$\mathbb{R}^3$中的一个坐标。为了实用性，我们可以把标架重新定义为一名观测者从$\mathscr{W}$直接到$\mathbb{R}^{3+1}$的对应过程，即$\phi:\mathscr{W}\to\mathbb{R}^{3+1}$。这时，标架$\phi$实际包括了一个观察者对$t_0$、$\hat{\mathbf{o}}$和$\{\hat{\mathbf{e}}_i\}$的主观选择。标架变换仍然包括时间和空间的两个等距变换，与原意义的标架变换是相同。

具体地，任一事件$w \in \mathscr{W}$经标架ϕ映射为$(\mathbf{x}, t)$, $\mathbf{x} \in \mathbb{R}^3, t \in \mathbb{R}$，经标架$\phi^*$映射为$(\mathbf{x}^*, t^*)$, $\mathbf{x}^* \in \mathbb{R}^3, t^* \in \mathbb{R}$。我们认为$\phi$和$\phi^*$都是可逆的，故有$(\mathbf{x}^*, t^*) = \phi^* \circ \phi^{-1}(\mathbf{x}, t)$。若$w$的时刻是$a \in \mathscr{T}$，标架$\phi$选择的时间原点是$t_0$，标架$\phi^*$选择的时间原点是$t_0^*$，则时间的度量空间上的一个等距变换$g : \mathscr{T} \to \mathscr{T}$可表示为

$$t^* = g(t) = t_0^* + \tau(t - t_0), t_0^* = g(t_0)$$

若标架ϕ和ϕ^*选择的空间原点分别是$\mathbf{x}_0, \mathbf{x}_0^*$，则在时刻$a$由标架$\phi$到标架$\phi^*$的变换是一个等距变换$i_a : \mathbb{R}^3 \to \mathbb{R}^3$，

$$\mathbf{x}^* = i_a(\mathbf{x}) = i_a(\mathbf{x}_0) + \mathbf{Q}_a(\mathbf{x} - \mathbf{x}_0), \mathbf{x}_0^* = i_a(\mathbf{x}_0)$$

所以，由ϕ到ϕ^*的标架变换包含一个时间的等距变换g和一个空间的等距变换i_a，其中a是事件的时刻。

III.1.4 场函数的标架不变性

设$(\mathbf{x}^*, t^*), (\mathbf{y}^*, t^*)$与$(\mathbf{x}, t), (\mathbf{y}, t)$是同一事件在标架$\phi, \phi^*$下的坐标和时标，则有$\mathbf{x}^* = \mathbf{x}_0^*(t) + \mathbf{Q}(t)[\mathbf{x} - \mathbf{x}_0], \mathbf{y}^* = \mathbf{x}_0^* + \mathbf{Q}(t)[\mathbf{y} - \mathbf{x}_0]$。两式相减得

$$\mathbf{y}^* - \mathbf{x}^* = \mathbf{Q}(t)(\mathbf{y} - \mathbf{x})$$

我们知道$(\mathbf{x}^*, t^*)$与$(\mathbf{x}, t)$对应是同一事件；同样地$(\mathbf{y}^*, t^*)$与$(\mathbf{y}, t)$也对应同一事件。故上式是同一向量在一个标架变换下的关系式。设$\mathbf{A} \in \mathscr{L}(\mathbb{R}^3)$是$\mathbb{R}^3$上的任一张量，$\mathbf{v} = \mathbf{y} - \mathbf{x}, \mathbf{v}^* = \mathbf{y}^* - \mathbf{x}^*$。则有

$$\mathbf{A}^*\mathbf{v}^* \equiv (\mathbf{A}\mathbf{v})^* = \mathbf{Q}(t)\,\mathbf{A}\mathbf{Q}^\intercal(t)\,\mathbf{v}^*$$

其中我们利用了$\mathbf{Q}^{-1} = \mathbf{Q}^\intercal$。上式是同一张量在一个标架变换下的关系式。

设在标架ϕ, ϕ^*下，标量场函数$h = h(\mathbf{x}, t), h^* = h^*(\mathbf{x}^*, t^*)$，向量场函数$\mathbf{v} = \mathbf{v}(\mathbf{x}, t), \mathbf{v}^* = \mathbf{v}^*(\mathbf{x}^*, t^*)$，张量场函数$\mathbf{A} = \mathbf{A}(\mathbf{x}, t), \mathbf{A}^* = \mathbf{A}^*(\mathbf{x}^*, t^*)$，则

- 当且仅当$h^*(\mathbf{x}^*, t^*) = h(\mathbf{x}, t)$时称$h$——
- 当且仅当$\mathbf{v}^*(\mathbf{x}^*, t^*) = \mathbf{Q}(t)\,\mathbf{v}(\mathbf{x}, t)$时称$\mathbf{v}$——
- 当且仅当$\mathbf{A}^*(\mathbf{x}^*, t^*) = \mathbf{Q}(t)\,\mathbf{A}(\mathbf{x}, t)\,\mathbf{Q}^\intercal(t)$时称$\mathbf{A}$——

具有标架不变性。

III.2 物体的运动

III.2.1 物体的运动

定义 III.2.1 (世界运动). 一组两两不相交的世界线的集合ψ称为一个世界运动。

一个世界运动中的世界线两两不相交是经典力学中的要求，称为“信息守恒定律”。通俗地说，任一时刻，某当前状态不可能有多于一种过去状态，也不可能有多于一种未来状态。

定义 III.2.2 (物体). 一个物体是一个非空集合$\mathscr{B}$，其元素$X \in \mathscr{B}$称为物质点。

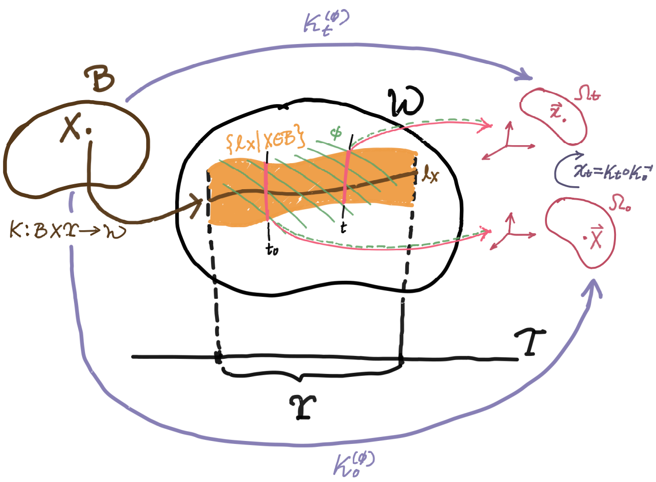

图 III.2.1: 物体的运动、运动学过程、放置、构型

定义 III.2.3 (运动学过程、物体的运动). 如图III.2.1所示，设$\mathscr{B}$是一个物体，$\mathscr{W}$是事件世界，Υ是时间$\mathscr{T}$的一个连通子集。映射$k: \mathscr{B} \times \Upsilon \to \mathscr{W}$称为一个运动学过程。对某物质点$X \in \mathscr{B}$，事件$k(X, a) = I_a, a \in \Upsilon$称$X$在时刻$a$经历的事件。集合$\{k(X, a) | a \in \Upsilon\}$是一条世界线的子集，记为$l_X$，称为物质点$X$的运动。$\mathscr{B}$的所有物质点的同一运动学过程$\{l_X | X \in \mathscr{B}\}$是一个世界运动的子集，记为$\psi_{\mathscr{B}}$，称为物体$\mathscr{B}$的运动。

作为世界运动的子集，物体的运动中各物质点的运动也是两两不相交的。

我们考虑物体$\mathscr{B}$的一个物质点$X \in \mathscr{B}$在一段时间Υ内的运动l_X。给定事件世界的标架ϕ，则l_X的每个事件$l_X(a), a \in \Upsilon$，都必属于ϕ的一条世界线$l_\phi(l_X(a))$。该物质点在某时刻$a \in \Upsilon$的速度是如下导数

$$\left.\frac{dl_X(t)}{dt}\right|_{t=a} \equiv \lim_{s \to 0} \frac{l_\phi(l_X(a+s)) - l_\phi(l_X(a))}{s}, s \in \mathbb{R}$$

同理可给出物质点X在时刻a的加速度。速度和加速度都是$\mathscr{V}_\phi$中的向量，它们都依赖参考系ϕ的选择。

定义 III.2.4 (构型). 如图III.2.1所示，设$\mathscr{W}$是事件世界，$\mathscr{T}$是时间，$\mathscr{B}$是物体，$\psi_{\mathscr{B}}$是其在时间间隔$\Upsilon \subset \mathscr{T}$的运动。选定标架$\phi$，则任一时刻$a \in \Upsilon$下集合$\Omega_a = \{l_\phi(l_X(a))\} \subset \phi$称为$\mathscr{B}$在时间$a$的构型。映射$\kappa_a^{(\phi)}: \mathscr{B} \to \Omega_a, \kappa_a^{(\phi)}(X) = l_\phi(l_X(a)) \, \forall X \in \mathscr{B}$称为物体$\mathscr{B}$在时刻$a$按标架$\phi$的放置映射。

放置映射$\kappa_a^{(\phi)}$是依赖标架ϕ的选择的。上一节我们对标架ϕ进行了简化的定义，相对应地，放置映射与构型的简化定义如下。

定义 III.2.5 (构型（简化）)**.** 给定标架$\phi : \mathscr{W} \to \mathbb{R}^{3+1}$，物体$\mathscr{B}$在任一时刻$t \in \mathbb{R}$占据的$\mathbb{R}^3$的区域$\Omega_t = \{\mathbf{x} | (\mathbf{x}, t) = \phi(X, t), X \in \mathscr{B}\}$称为$\mathscr{B}$在时刻$t$下的构型。映射$\kappa_t^{(\phi)} : \mathscr{B} \to \Omega_t, \kappa_t^{(\phi)}(X) = \mathbf{x}, X \in \mathscr{B}, \mathbf{x} \in \Omega_t$称为物体$\mathscr{B}$按标架$\phi$在时刻$t$的放置。

我们要求κ是可逆的。

设某一时刻物体$\mathscr{B}$在不同标架ϕ, ϕ^*下的两个放置映射是$\kappa_t, \kappa_{t^*}^*$，则任一物质点$X \in \mathscr{B}$在两个标架下的位置向量$\mathbf{x} = \kappa_t(X), \mathbf{x}^* = \kappa_{t^*}^*(X)$之间有标架变换关系：

$$\mathbf{x}^* = \mathbf{x}_0^* + \mathbf{Q}(\mathbf{x} - \mathbf{x}_0),$$

$$t^* = t_0^* + t - t_0 = t + a, a = t_0^* - t_0$$

其中$\mathbf{x}_0, \mathbf{x}_0^*$是选定某物质点$X_0 \in \mathscr{B}$在标架$\phi, \phi^*$下的位置向量。

我们考虑察时间导数。由上式可知$\frac{d}{dt^*} = \frac{d}{dt}$，故有

$$\frac{d\mathbf{x}^*}{dt^*} = \frac{d\mathbf{x}^*}{dt} = \frac{d\mathbf{x}_0^*}{dt} + \mathbf{Q}(t)\frac{d\mathbf{x}}{dt} + \frac{d\mathbf{Q}(t)}{dt}[\mathbf{x} - \mathbf{x}_0]$$
$$\mathbf{v}^* - \mathbf{Q}(t)\mathbf{v} = \frac{d\mathbf{x}_0^*}{dt} + \frac{d\mathbf{Q}(t)}{dt}[\mathbf{x} - \mathbf{x}_0] \neq \mathbf{0}$$

即速度不满足标架不变性。同理加速度也不满足：

$$\mathbf{a}^* - \mathbf{Q}(t)\mathbf{a} = \frac{d^2\mathbf{x}_0^*}{dt^2} + 2\mathbf{A}(t)\left[\frac{d\mathbf{x}^*}{dt} - \frac{d\mathbf{x}_0^*}{dt}\right] + \frac{d\mathbf{A}(t)}{dt}[\mathbf{x}^* - \mathbf{x}_0^*] - \mathbf{A}^2(t)[\mathbf{x}^* - \mathbf{x}_0^*] \neq \mathbf{0}$$

其中$\mathbf{A}(t) = \frac{d\mathbf{Q}(t)}{dt}\mathbf{Q}^\intercal(t)$。含$\mathbf{A}$的后三项分别为科里奥利加速度、角加速度和向心加速度。这三项是纯粹由于标架变换产生的，不是客观的运动。要使加速度满足标架不变性，除非$\frac{d^2\mathbf{x}_0^*}{dt^2} \equiv 0, \frac{d\mathbf{Q}_t}{dt} \equiv 0$。我们称满足这种标架变换的标架为伽俐略标架，这种标架变换称为伽俐略变换。任意两个伽俐略标架之间相对静止或相对作匀速直线运动。

III.3 物体的形变

III.3.1 物体的形变

设有两个时刻$a < b, a, b \in \mathbb{R}$，在给定标架$\phi$下，物体$\mathscr{B}$在这两个时刻的构型分别是$\Omega_a, \Omega_b$，则由$\Omega_a$到$\Omega_b$的映射为$\kappa_b^{(\phi)} \circ \kappa_a^{(\phi)-1} = \chi_{a\to b}$，称为物体$\mathscr{B}$在标架$\phi$下的形变*。如果在形变过程中

*此处使用了构型的简化定义，这与采用构型的原始定义（即ϕ从某标架替换为某事件世界的参考系）的结果无重要差别。

没有发生标架变换，我们常将标架符号从放置映射的上标省去，但应牢记物体的放置总是在选定某标架下的映射。

如图III.2.1所示，选定标架ϕ，物体$\mathscr{B}$在该标架的参考时刻t_0的构型Ω_0称作参考构型，在其他任意时刻t的构型Ω_t称作当前构型。我们用罗巴斜体大写英文字母$X, Y, \cdots$表示物质点，用粗体大写英文字母$\mathbf{X}, \mathbf{Y}, \cdots$表示同字母的物质点在参考构型下的位置向量，用粗体小写英文字母$\mathbf{x}, \mathbf{y}, \cdots$表示同字母的物质点在当前构型下的位置向量。则任一物质点$X \in \mathscr{B}$在标架ϕ下的运动轨迹是$\mathbf{x}(t), \mathbf{x} \in \Omega_t$。它的方程可通过形变映射$\chi_t(\mathbf{X})$表示为由参考构型$\Omega_0$到当前构型的形变。这时我们把当前时刻$t$也当作了变量，所以把形变映射写成$\chi(\mathbf{X}, t)$。$\mathbf{X}$是已经选定了的某时刻的构型中的位置，所以$\mathbf{X}$是不依赖时间的，对时间是常量。但它可作为表示$\Omega_0$区域内不同位置的一个变量。只要知道了形变映射$\chi(\mathbf{X}, t)$的形式，就知道了物体$\mathscr{B}$的运动在标架$\phi$下的总轨迹。由信息守恒定律，形变映射是可逆的。

III.3.2 形变梯度张量

设物质点$X, Y \in \mathscr{B}$在参考时刻$t_0 \in \Upsilon$经历的事件为$x_0, y_0 \in I_0$，在当前时刻$t \in \Upsilon$经历的事件为$x, y \in I_t$。若导数

$$\lim_{y_0 \to x_0} \frac{y - x}{y_0 - x_0}$$

存在，则称其为物体$\mathscr{B}$在时刻t下X所在处的形变梯度。上式的分子和分母分别是I_a和I_b的两个平移空间$\mathscr{V}_0, \mathscr{V}_t$中的向量。故该导数是由$\mathscr{V}_0$到$\mathscr{V}_t$的线性变换。在选定标架$\phi$下，则形变梯度张量可表示为形变映射$\chi$的导数，

$$\mathbf{F} = \frac{\partial \chi(\mathbf{X}, t)}{\partial \mathbf{X}} = \mathbf{F}(\mathbf{X}, t), \mathbf{X} \in \Omega_0$$

其中$\mathbf{X}$是物质点$X \in \mathscr{B}$在标架ϕ下的参考构型Ω_0中的位置向量。形变梯度张量作为一个全导数，可满足以下全微分的意义：对于两个物质点$X, Y \in \mathscr{B}$，

$$\mathbf{y}(t) - \mathbf{x}(t) = \chi(\mathbf{Y}, t) - \chi(\mathbf{X}, t) = \mathbf{F}(\mathbf{X}, t)(\mathbf{Y} - \mathbf{X}) + \|\mathbf{Y} - \mathbf{X}\| \mathbf{z}(\mathbf{Y} - \mathbf{X})$$

其中函数$\mathbf{z}(\mathbf{x})$具有性质$\lim_{\mathbf{x} \to \mathbf{0}} \mathbf{z}(\mathbf{x}) = \mathbf{0}$。

我们连续介质力学中所考虑的大部分问题，都假设形变梯度张量定义中的导数都存在，即物体$\mathscr{B}$有某种“可微性”，使得其在任一标架下任一时刻的构型都可微。

由于形变映射$\chi(\cdot, t)$是可逆的，由反函数定理，若$\mathbf{x} = \chi(\mathbf{X}, t)$，$\mathbf{F} = \frac{\partial \chi}{\partial \mathbf{X}}$，则$\mathbf{F}^{-1} = \frac{\partial \chi^{-1}}{\partial \mathbf{x}}$。

在基本坐标系下，若$\mathbf{x} = (x_1, x_2, x_3) = \chi(\mathbf{X}, t), \mathbf{X} = (X_1, X_2, X_3)$且记$x_i = \chi_i(\mathbf{X}, t)$，则形

变梯度张量的矩阵

$$(\mathbf{F})=\begin{pmatrix}\frac{\partial\chi_1}{\partial X_1} & \cdots & \frac{\partial\chi_1}{\partial X_3}\\ \vdots & \ddots & \vdots\\ \frac{\partial\chi_3}{\partial X_1} & \cdots & \frac{\partial\chi_3}{\partial X_3}\end{pmatrix}$$

称为形变χ的雅可比矩阵，行列式$\det\mathbf{F}=\det(\mathbf{F})$称为形变$\chi$的雅可比行列式。

设(ϕ,ϕ^*)是一个标架变换，物体$\mathscr{B}$在两个标架下的参考构型为Ω_0,Ω_0^*，当前构型为Ω_t,Ω_t^*。物质点$X,X_0\in\mathscr{B}$在参考时刻下的位置向量满足以下标架变换关系：

$$\mathbf{X}^*=\mathbf{X}_0^*+\mathbf{Q}_0(\mathbf{X}-\mathbf{X}_0)$$
$$\Leftrightarrow\mathbf{X}=\mathbf{X}_0+\mathbf{Q}_0^{\intercal}(\mathbf{X}^*-\mathbf{X}_0^*)$$

在当前时刻下的位置向量满足以下标架变换关系：

$$\mathbf{x}^*=\mathbf{x}_0^*+\mathbf{Q}_t(\mathbf{x}-\mathbf{x}_0)$$
$$t^*=t+a,a=t_0^*-t_0$$

设由构型Ω_0到Ω_t的形变映射为$\chi(\cdot,t)$，由构型Ω_0^*到ω_t^*的形变映射为$\chi^*(\cdot,t^*)$，则有

$$\mathbf{x}=\chi(\mathbf{X},t)$$
$$\mathbf{x}_0=\chi(\mathbf{X}_0,t)$$
$$\mathbf{x}^*=\chi^*(\mathbf{X}^*,t^*)$$
$$\mathbf{x}_0^*=\chi^*(\mathbf{X}_0^*,t^*)$$

代入上面的标架变换关系得

$$\begin{aligned}\chi^*(\mathbf{X}^*,t^*)&=\mathbf{x}^*=\mathbf{x}_0^*+\mathbf{Q}_t(\mathbf{x}-\mathbf{x}_0)\\&=\chi(\mathbf{X}_0^*,t^*)+\mathbf{Q}_t(\chi(\mathbf{X},t)-\chi(\mathbf{X}_0,t))\end{aligned}$$

标架ϕ^*下的形变梯度张量

$$\begin{aligned}\mathbf{F}=\frac{\partial\chi^*(\mathbf{X}^*,t^*)}{\partial\mathbf{X}^*}&=\mathbf{Q}_t\frac{\partial\chi(\mathbf{X},t)}{\partial\mathbf{X}^*}\\&=\mathbf{Q}_t\frac{\partial\chi(\mathbf{X},t)}{\partial\mathbf{X}}\frac{\partial\mathbf{X}}{\partial\mathbf{X}^*}\\&=\mathbf{Q}_t\mathbf{F}\mathbf{Q}_0\end{aligned}$$

上面利用了复合函数求导的链式法则。这一结果说明，一般地形变梯度张量不具有标架变换不变性。但当这个标架变换是一个伽俐略变换（即有$\mathbf{Q}_t\equiv\mathbf{Q}_0$）时，形变梯度张量具有标架变换不变性。

若$\mathbf{F}$对每一$\mathbf{X}\in\Omega_0$均相同，则称$\mathscr{B}$发生的是均匀形变。以下是一些二维均匀形变的例子。

例 III.3.1. 设对物体$\mathscr{B}$的每一物质点$X \in \mathscr{B}$，参考位置$\mathbf{X} = (X_1, X_2)$，当前位置$\mathbf{x} = (x_1, x_2)\,, \mathbf{X} = (X_1, X_2)$。

刚体平动：

$$x_1 = X_2 + 5$$
$$x2 = X_1 + 2$$

即$\mathbf{F} = \mathbf{I}$。一般地，没有形变和刚体转运时，$\mathbf{F} = \mathbf{I}$。

刚体转动：

$$x_1 = X_1 \cos\theta - X_2 \sin\theta$$
$$x2 = X_1 \sin\theta + X_2 \cos\theta$$

即

$$(\mathbf{F}) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

一般地，没有形变但有刚体转动时，$\mathbf{F} \neq \mathbf{I}$。设物体中两物质点之间的平移向量在物体运动前、后是$\mathbf{D}, \mathbf{d}$，

$$(\mathbf{D}) = \begin{pmatrix} a \\ b \end{pmatrix}$$

则有$\mathbf{d} = \mathbf{F}\mathbf{D}$，

$$(\mathbf{d}) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a\cos\theta - b\sin\theta \\ a\sin\theta + b\cos\theta \end{pmatrix}$$

膨胀：

$$x_1 = 2X_1$$
$$x2 = 1.5X_2$$

即

$$(\mathbf{F}) = \begin{pmatrix} 2 & 0 \\ 0 & 1.5 \end{pmatrix}$$

一般地，直角坐标系下$(\mathbf{F})$的对角元素表示相应方向的拉伸比例。

纯剪切：

$$x_1 = X_1 + 0.5X_2$$
$$x2 = 0.5X_1 + X_2$$

即

$$(\mathbf{F}) = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

一般地，直角坐标系下$(\mathbf{F})$的非对角元素表示剪切。

简单剪切：

$$x_1 = X_1$$
$$x2 = 0.5X_1 + X_2$$

即

$$(\mathbf{F}) = \begin{pmatrix} 1 & 0 \\ 0.5 & 1 \end{pmatrix}$$

可见，简单剪切是纯剪切和纯拉伸形变的复合形变。

一般形变：

$$x_1 = 1.3X_1 - 0.375X_2$$
$$x2 = 0.75X_1 + 0.65X_2$$

即

$$(\mathbf{F}) = \begin{pmatrix} 1.3 & -0.375 \\ 0.75 & 0.65 \end{pmatrix}$$

该形变可分为两步。第一步是拉伸：

$$x_1' = 1.5X_1$$
$$x_2' = 0.75X_2$$

这一步的形变梯度张量$\mathbf{U}$的矩阵

$$(\mathbf{U}) = \begin{pmatrix} 1.5 & 0 \\ 0 & 0.75 \end{pmatrix}$$

第二步是刚体旋转：

$$x_1 = x_1' \cos 30° - x_2' \sin 30° = 1.3X_1 - 0.375X_2$$
$$x_2 = x_1' \sin 30° + x_2' \cos 30° = 0.75X_1 + 0.65X_2$$

这一步的形变梯度张量$\mathbf{R}$的矩阵

$$(\mathbf{R}) = \begin{pmatrix} 0.866 & -0.5 \\ -0.5 & 0.866 \end{pmatrix}$$

故总的形变梯度张量可以写成$\mathbf{F} = \mathbf{RU}$。先旋转再拉伸也可以达到同样的结果，此时$\mathbf{F} = \mathbf{VR}'$并且$\mathbf{R}' = \mathbf{R}, \mathbf{V} = \mathbf{RUR}^\intercal, \mathbf{U} = \mathbf{R}^\intercal\mathbf{VR}$。

一般地，对任一可逆线性算符总有唯一的厄米算符$\mathbf{U}, \mathbf{V}$和幺正算符$\mathbf{R}$满足$\mathbf{F} = \mathbf{RU} = \mathbf{VR}$，称为$\mathbf{F}$的极分解。特别地，在形变梯度张量$\mathbf{F}$存在唯一极分解，其中$\mathbf{U} = \mathbf{U}^\intercal, \mathbf{V} = \mathbf{V}^\intercal$是对称张量，分别称为右拉伸张量和左拉伸张量,$\mathbf{RR}^\intercal = \mathbf{I}$是正交张量，称为旋转张量。

我们希望能有一种关于物体形变的度量。我们认为刚体旋转不属于形变，但纯刚体旋转会使形变梯度张量取非平凡值，故一形变梯度张量并不是一个理想的形变度量。通过极分解，获得的拉伸张量，虽然是排除了物体的刚体旋转部分，但由于它们的计算不方便，故使用得很少。

例 III.3.2. 请验证：

$$\mathbf{F}^\intercal\mathbf{F} = (\mathbf{RU})^\intercal (\mathbf{RU}) = \mathbf{U}^\intercal\mathbf{R}^\intercal\mathbf{RU} = \mathbf{U}^\intercal\mathbf{U} = \mathbf{U}^2$$

$$\mathbf{FF}^\intercal = (\mathbf{VR}) (\mathbf{VR})^\intercal = \mathbf{VRR}^\intercal\mathbf{V}^\intercal = \mathbf{VV}^\intercal = \mathbf{V}^2$$

可见，将形变梯度张量与其转置相乘（注意顺序），刚体旋转部分会自行抵消掉。我们定义：

$$\mathbf{C} = \mathbf{F}^\intercal\mathbf{F} = \mathbf{U}^2 \qquad \text{右柯西–格林张量}$$

$$\mathbf{B} = \mathbf{FF}^\intercal = \mathbf{V}^2 \qquad \text{左柯西–格林张量}$$

$\mathbf{B}$和$\mathbf{C}$统称有限应变张量。进一步可定义

$$\mathbf{E} = \frac{1}{2}(\mathbf{C} - \mathbf{I}) \qquad \text{格林–拉格朗日应变张量}$$

III.4　物质描述与空间描述

物体$\mathscr{B}$的元素——物质点$X \in \mathscr{B}$不仅可以具有“位置”的性质（通过放置映射κ），还能具有密度、温度等物理性质。在选定的标架下，某时刻物质点的性质

$$f : \mathscr{B} \times \mathbb{R} \to \mathscr{F}$$

可以是标量，向量或张量。由于人无法直接认识物体$\mathscr{B}$及其物质点$X \in \mathscr{B}$，故函数$f(X,t)$需要转化为关于物质点X在物体$\mathscr{B}$在某标架下的构型中的位置向量的函数。设物体$\mathscr{B}$在某标架下，参考时刻t_0时的构型Ω_0是参考构型，当前时刻t的构型Ω_t是当前构型，形变映射是$\chi(\cdot,t)$，X在Ω_0中的位置向量是$\mathbf{X}$，则X在当前时刻t的位置向量是$\mathbf{x}(t) = \chi(\mathbf{X},t)$。如果把$f(X,t)$改写为参考构型中的位置$\mathbf{X}$的函数，即令

$$f_\mathrm{m} : \Omega_0 \times \mathbb{R} \to \mathscr{F}, f_\mathrm{m}(\mathbf{X},t) = f\left(\kappa_0^{-1}(\mathbf{X}),t\right), \forall \mathbf{X} \in \Omega_0$$

则称函数f_m是物体性质f的物质描述，又称拉格朗日描述。

如果，由于物体$\mathscr{B}$的运动，空间位置$\mathbf{r} \in \mathbf{R}^3$处某性质读数在不断变化，则这个由物体运导致的性质场函数

$$f_\mathrm{s} : \mathbb{R}^{3+1} \to \mathscr{F}, f_\mathrm{s}(\mathbf{r},t) = f\left(\kappa_t^{-1}(\mathbf{r}),t\right), \forall \mathbf{r} \in \Omega_t$$

称函数f_{s}是物体性质f的空间描述，又称欧拉描述。注意，f_{s}的定义域Ω_t是随时刻t变化的。

物质函数描述与空间描述函数之间的关系：

$$
\begin{aligned}
f_{\mathrm{m}}\left(\mathbf{X}, t\right) &= f_{\mathrm{s}}\left(\chi\left(\mathbf{X}, t\right), t\right) \\
f_{\mathrm{s}}\left(\mathbf{r}, t\right) &= f_{\mathrm{m}}\left(\chi^{-1}\left(\mathbf{r}, t\right), t\right)
\end{aligned}
$$

其中，为了注意到当前时刻的形变映射χ_t及其逆映射χ_t^{-1}均依赖时刻t，故将它们写成了$\chi\left(\cdot, t\right), \chi^{-1}\left(\cdot, t\right)$。

考虑物体在选定某标架下的运动的速度

$$
\mathbf{v}\left(X, t\right) = \frac{d\kappa\left(X, t\right)}{dt}, X \in \mathscr{B}
$$

其中为了注意到当前时刻的放置映射κ_t依赖时刻t故将其写成$\kappa\left(\cdot, t\right)$。由定义，速度的物质描述是

$$
\begin{aligned}
\mathbf{v}_{\mathrm{m}}\left(\mathbf{X}, t\right) &= \mathbf{v}\left(\kappa_0^{-1}\left(\mathbf{X}\right), t\right) \\
&= \frac{\partial}{\partial t}\kappa\left(\kappa_0^{-1}\left(\mathbf{X}\right), t\right) \\
&= \frac{\partial}{\partial t}\chi\left(\mathbf{X}, t\right)
\end{aligned}
$$

我们在进行粒子示踪测速（particle tracking velocimetry，PTV）的时候，粒子的速度就是物质描述的速度函数。速度的空间描述是

$$
\begin{aligned}
\mathbf{v}_{\mathrm{s}}\left(\mathbf{r}, t\right) &= \mathbf{v}\left(\kappa^{-1}\left(\mathbf{r}, t\right), t\right) \\
&= \frac{\partial}{\partial t}\kappa\left(\kappa^{-1}\left(\mathbf{r}, t\right), t\right)
\end{aligned}
$$

我们在进行粒子成像测速（particle imaging velocimetry，PIV）的时候，测量结果（速度场）就是物体运动速度的空间描述。

如果我们想知道速度场$\mathbf{v}_{\mathrm{s}}$在某处$\mathbf{r}$的变化率，直接对其求时间偏导数即可，但这不是物体$\mathscr{B}$的任一物质点的加速度。因为物体在运动中，所以$\mathbf{r}$处的物质点是一直变化的。而我们之所以想知道加速度，是为了将来运用牛顿第二定律。

加速度是物质点的速度变化，即

$$
\mathbf{a}\left(X, t\right) = \frac{\partial \mathbf{v}\left(X, t\right)}{\partial t} = \frac{\partial^2}{\partial t^2}\kappa\left(X, t\right), X \in \mathscr{B}
$$

速度的物质描述$\mathbf{a}_{\mathrm{m}}\left(\mathbf{X}, t\right) = \mathbf{a}\left(\kappa_0^{-1}\left(\mathbf{X}\right), t\right) = \frac{\partial^2}{\partial t}\chi\left(\mathbf{X}, t\right) = \frac{\partial}{\partial t}\mathbf{v}_{\mathrm{m}}\left(\mathbf{X}, t\right)$。如果我们只知道速度场$\mathbf{v}_{\mathrm{s}}\left(\mathbf{r}, t\right)$，要求加速度的物质描述，就需要用$\mathbf{v}_{\mathrm{s}}\left(\mathbf{r}, t\right)$来表出同一个物质点$X$（或关于参考构

型中的同一个位置$\mathbf{X}$）在不同时刻的速度，即$\mathbf{v}_{\mathrm{s}}(\chi(\mathbf{X},t),t)$。对这个复合函数求时间导数，才是物质点$X$的加速度。例II.12.1为我们提供了这类复合函数求导的法则，故：

$$\begin{aligned}\mathbf{a}(X,t)=\mathbf{a}_{\mathrm{m}}(\mathbf{X},t)&=\frac{\partial}{\partial t}\mathbf{v}_{\mathrm{s}}(\chi(\mathbf{X},t),t)\\&=\left.\frac{\partial\mathbf{v}_{\mathrm{s}}(\mathbf{r},t)}{\partial t}\right|_{\mathbf{r}=\chi(\mathbf{X},t)}+\left.\frac{\partial\mathbf{v}_{\mathrm{s}}(\mathbf{r},t)}{\partial\mathbf{r}}\right|_{\mathbf{r}=\chi(\mathbf{X},t)}\frac{\partial\chi(\mathbf{X},t)}{\partial t}\\&=\left.\frac{\partial\mathbf{v}_{\mathrm{s}}(\mathbf{r},t)}{\partial t}\right|_{\mathbf{r}=\chi(\mathbf{X},t)}+\left.\frac{\partial\mathbf{v}_{\mathrm{s}}(\mathbf{r},t)}{\partial\mathbf{r}}\right|_{\mathbf{r}=\chi(\mathbf{X},t)}\mathbf{v}_{\mathrm{m}}(\mathbf{X},t)\end{aligned}$$

一般地，在选定某标架下，如果空间存在的某性质场$f_{\mathrm{S}}(\mathbf{r},t)$是由具有该性质的物体$\mathscr{B}$的运动造成的（即$f_{\mathrm{s}}$是该物体性质的空间描述），那么，对$f_{\mathrm{s}}(\mathbf{r},t)$进行以下计算

可以得到该物体性质的变化率（物质描述）。我们称这一计算为对物体性质的空间描述的物质导数。

$$\begin{aligned}\frac{D}{Dt}f_{\mathrm{s}}(\mathbf{r},t)=f_{\mathrm{m}}(\mathbf{X},t)=f(X,t)=&\left.\frac{\partial}{\partial t}f_{\mathrm{s}}(\mathbf{r},t)\right|_{\mathbf{r}}\\&=\chi(\mathbf{X},t)+\left.\frac{\partial}{\partial\mathbf{r}}f_{\mathrm{s}}(\mathbf{r},t)\right|_{\mathbf{r}=\chi(\mathbf{X},t)}\mathbf{v}_{m}(\mathbf{X},t),\mathbf{X}\in\Omega_0,X\in\mathscr{B}\end{aligned}$$

例 III.4.1. 设物体$\mathscr{B}$的运动在某标架下满足

$$\chi(\mathbf{X},t)=\left(X_1+X_2t^2\right)\hat{\mathbf{e}}_1+(X_2-X_1t)\hat{\mathbf{e}}_2+X_3\hat{\mathbf{e}}_3$$

则速度的物质描述

$$\mathbf{v}_{\mathrm{m}}(\mathbf{X},t)=\frac{\partial\chi(\mathbf{X},t)}{\partial t}=2X_2t\hat{\mathbf{e}}_1-X_1\hat{\mathbf{e}}_2$$

速度的空间描述

$$\mathbf{v}_{\mathrm{s}}(\mathbf{r},t)=\mathbf{v}_{\mathrm{m}}\left(\chi^{-1}(\mathbf{r},t),t\right)$$

为求$\chi^{-1}(\mathrm{r},t)$，注意到，t时记刻，$\mathbf{r}=\chi(\mathbf{X},t)$，故由

$$\begin{cases}r_1=X_1+X_2t^2\\r_2=X_2-X_1t\\r_3=X_3\end{cases}$$

解得

$$\begin{cases}X_1=\frac{r_1-r_2t^2}{1+t^3}\\X_2=\frac{r_2+r_1t}{1+t^3}\\X_3=r_3\end{cases}$$

即

$$\chi^{-1}(\mathbf{r},t)=\frac{r_1-r_2t^2}{1+t^3}\hat{\mathbf{e}}_1+\frac{r_2+r_1t}{1+t^3}\hat{\mathbf{e}}_2+r_3\hat{\mathbf{e}}_3$$

故

$$\mathbf{v}_{\mathrm{s}}(\mathbf{r},t)=\frac{2r_2t+2r_1t^2}{1+t^3}\hat{\mathbf{e}}_1-\frac{r_1-r_2t^2}{1+t^3}\hat{\mathbf{e}}_2$$

加速度的物质描述

$$\mathbf{a}_{\mathrm{m}}(\mathbf{X},t)=\frac{\partial\mathbf{v}_m(\mathbf{X},t)}{\partial t}=2X_2\hat{\mathbf{e}}_1$$

若速度场$\mathbf{v}_{\mathrm{s}}(\mathbf{r},t)$是完全由物体$\mathscr{B}$的运动导致的，则加速度的物质描述

$$\begin{aligned}\mathbf{a}_{\mathrm{m}}(\mathbf{X},t)&=\frac{\partial\mathbf{v}_{\mathrm{s}}(\mathbf{r},t)}{\partial t}+\left.\frac{\partial\mathbf{v}_{\mathrm{s}}(\mathbf{r},t)}{\partial\mathbf{r}}\right|_{\mathbf{r}=\chi(\mathbf{X},t)}\mathbf{v}_{\mathrm{m}}(\mathbf{X},t)\\&=\begin{pmatrix}\frac{2\left(tX_1+X_2-t^3X_2\right)}{1+t^3}\\\frac{t(tX_1+2X_2)}{1+t^3}\\0\end{pmatrix}+\begin{pmatrix}\frac{2t^2}{1+t^3}&\frac{2t}{1+t^3}&0\\-\frac{1}{1+t^3}&\frac{t^2}{1+t^3}&0\\0&0&0\end{pmatrix}\begin{pmatrix}2X_2t\\-X_1\\0\end{pmatrix}\\&=2X_2\hat{\mathbf{e}}_1\end{aligned}$$

III.5 应变率张量

对速度的空间描述（即速度场）$\mathbf{v}_{\mathrm{s}}(\mathbf{r},t)$求当前构型下的空间导数得到的张量

$$\mathbf{L}=\frac{\partial}{\partial\mathbf{r}}\mathbf{v}_{\mathrm{s}}(\mathbf{r},t)=\mathbf{L}(\mathbf{r},t)$$

称为速度梯度张量。它是关于物体运动的一个空间描述的线性变换值函数。在标准基下，

$$(\mathbf{L})=\begin{pmatrix}\frac{\partial v_{\mathrm{s}1}}{\partial r_1}&\frac{\partial v_{\mathrm{s}1}}{\partial r_2}&\frac{\partial v_{\mathrm{s}1}}{\partial r_3}\\\frac{\partial v_{\mathrm{s}2}}{\partial r_1}&\frac{\partial v_{\mathrm{s}2}}{\partial r_2}&\frac{\partial v_{\mathrm{s}2}}{\partial r_3}\\\frac{\partial v_{\mathrm{s}3}}{\partial r_1}&\frac{\partial v_{\mathrm{s}3}}{\partial r_2}&\frac{\partial v_{\mathrm{s}3}}{\partial r_3}\end{pmatrix}$$

对速度的物质描述求参考构型下的空间导数

$$\begin{aligned}\frac{\partial}{\partial\mathbf{X}}\mathbf{v}_{\mathrm{m}}(\mathbf{X},t)&=\frac{\partial}{\partial\mathbf{X}}\left(\frac{\partial}{\partial t}\chi(\mathbf{X},)\right)\\&=\frac{\partial}{\partial t}\left(\frac{\partial}{\partial\mathbf{X}}\chi(\mathbf{X},t)\right)\\&=\frac{\partial}{\partial t}\mathbf{F}\end{aligned}$$

如果用速度的物质描述来表出速度场

$$\mathbf{v}_{\mathrm{s}}(\mathbf{r},t)=\mathbf{v}_{\mathrm{m}}\left(\chi^{-1}(\mathbf{r},t),t\right)$$

则

$$\begin{aligned}\mathbf{L} = \frac{\partial}{\partial \mathbf{r}} &= \frac{\partial}{\partial \mathbf{r}} \mathbf{v}_{\mathrm{m}} \left(\chi^{-1}(\mathbf{r}, t), t\right) \\ &= \frac{\partial}{\partial \mathbf{X}} \mathbf{v}_{\mathrm{m}} (\mathbf{X}, t)\bigg|_{\mathbf{X}=\chi^{-1}(\mathbf{r},t)} \frac{\partial}{\partial \mathbf{r}} \chi^{-1}(\mathbf{r}, t) \\ &= \frac{\partial}{\partial t} \mathbf{F}(\mathbf{X}, t)\bigg|_{\mathbf{X}=\chi^{-1}(\mathbf{r},t)} \mathbf{F}^{-1}(\mathbf{r}, t) = \dot{\mathbf{F}} \mathbf{F}^{-1}\end{aligned}$$

上式是速度梯度张量与形变梯度张量之间的关系。

如果形变梯度张量$\mathbf{F}$可逆则$\mathbf{L}$也可逆，故$\mathbf{L}$可以写成如下对称张量和斜称张量的和：

$$\mathbf{L} = \frac{1}{2}\left(\mathbf{L} + \mathbf{L}^{\intercal}\right) + \frac{1}{2}\left(\mathbf{L} - \mathbf{L}^{\intercal}\right) = \mathbf{D} + \mathbf{W}$$

其中定义

$$\mathbf{D} = \mathbf{D}^{\intercal} = \frac{1}{2}\left(\mathbf{L} + \mathbf{L}^{\intercal}\right) \qquad \text{应变速率张量}$$

$$\mathbf{W} = -\mathbf{W}^{\intercal} = \frac{1}{2}\left(\mathbf{L} - \mathbf{L}^{\intercal}\right) \qquad \text{旋度张量}$$

例 III.5.1. 请自行写出$\mathbf{D}$和$\mathbf{W}$在标准基下的矩阵式。

$\mathbf{L}$、$\mathbf{D}$、$\mathbf{W}$都是空间描述的函数（场函数）。由定理II.12.6，$\mathbf{L}(\mathbf{r}, t)$作用于单位向量$\mathbf{u}$可得到$t$时刻$\mathbf{r}$处速度场$\mathbf{v}_s$朝方向$\mathbf{u}$的变化率$\frac{\partial}{\partial \mathbf{u}} \mathbf{v}_s(\mathbf{r}, t)$。

第四部分 附录

IV.1 线性代数部分定理的证明

IV.1.1 范的等价性

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 IV.1.1. 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$k\|\mathbf{x}\|_{\mathrm{E}}\le\|\mathbf{x}\|\le K\|\mathbf{x}\|_{\mathrm{E}},\forall\mathbf{x}\in\mathscr{V}$。

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

$$\begin{aligned}
\|\mathbf{x}\| &= \left\|\sum_{i=1}^n x_i\mathbf{e}_i\right\| \\
&\le \sum_{i=1}^n |x_i|\,\|\mathbf{e}_i\| \qquad \text{（三角不等式，当且仅当}x_1=\cdots=x_n\text{时取等号。）} \\
&\le \max\{|x_i|\}\sum_{i=1}^n \|\mathbf{e}_i| \qquad \text{（非负实数求和的简易性质，当且仅当}\,|x_1|=\cdots|x_n|\,\text{时取等号。）} \\
&\le \left(\sum_{i=1}^n x_i^2\right)^{\frac{1}{2}}\sum_{i=1}^n\|\mathbf{e}_i\| \qquad \text{（当且仅当}n=1\text{时取等号。）} \\
&= K\|\mathbf{x}\|_{\mathrm{E}}
\end{aligned}$$

其中$K=\sum_{i=1}^n\|\mathbf{e}_i\|>0$（基向量$\mathbf{e}_i\ne\mathbf{0}\forall i=1,\cdots,n$）。故总存在$K>0$使得$\|\mathbf{x}\|\le K\|\mathbf{x}\|_{\mathrm{E}}\forall\mathbf{x}\in\mathscr{V}$。

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$|\|\mathbf{x}\|_{\mathrm{E}}-\|\mathbf{x}_0\|_{\mathrm{E}}|\le\|\mathbf{x}-\mathbf{x}_0\|_{\mathrm{E}}$，故总存在$k'\le 1$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$\|\mathbf{x}-\mathbf{x}_0\|\le K'\|\mathbf{x}-\mathbf{x}_0\|_{\mathrm{E}}$。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$|\|\mathbf{x}\|-\|\mathbf{x}_0\||\le\|\mathbf{x}-\mathbf{x}_0\|\le K'\|\mathbf{x}-\mathbf{x}_0\|_{\mathrm{E}}=\frac{K'}{k'}|\|\mathbf{x}\|_{\mathrm{E}}-\|\mathbf{x}_0\|_{\mathrm{E}}|<\frac{K'\delta}{k'}=\epsilon$，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$\|\mathbf{x}\|_{\mathrm{E}}\le 1$范围内，$f$必存在最小值，即必存在$0\le a\le 1$使得$0<f(a)\le\|\mathbf{x}\|\,\forall\mathbf{x}\in\{\mathbf{x}|\mathbf{x}\in\mathscr{V},\|\mathbf{x}\|_{\mathrm{E}}\le 1\}$。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 IV.1.2. 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 IV.1.1. 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。 □

IV.1.2 伴随算符的唯一存在性

引理 IV.1.3. 设$\mathscr{V}$是数域$\mathbb{F}$上的有限维内积空间，给定$\mathscr{V}$上的一个线性泛函$f\in\mathscr{V}^*$，则存在唯一一个向量$\mathbf{b}\in\mathscr{V}$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。

□

定理 IV.1.2. 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$f\in\mathscr{V}^*, f(\mathbf{a})\equiv(\mathbf{Ta}|\mathbf{b}), \forall\mathbf{a}\in\mathscr{V}$。由前一条引理，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。 □

IV.2 向量函数微积分部分的证明

IV.2.1 $\mathbb{R}^n$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[6]§7.1,p.1。

定义 IV.2.1 (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，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。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。

由定义，整个$\mathbb{R}^n$是一个开集。“空集是一个开集”虚真（vacuously true）。$\mathbb{R}^n$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。这些结论都需要证明但此略。

定义 IV.2.2 (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，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。S的所有极限点的集合称S的导集。如果集合S包含它的所有极限点，则称集合S是闭集。若集合S等于其闭包（$S = \mathrm{cl}S$），则称S是闭集。

由定义，整个$\mathbb{R}^n$是一个闭集。“空集是一个闭集”虚真。$\mathbb{R}^n$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。

定义 IV.2.3 (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。所有这样的点$\mathbf{x}_0$的集合称为集合$S$的边界，常记为$\partial S$。

定理 IV.2.1. 一个集合是闭集当且仅当它包含所有其边界上的点。

推论 IV.2.1.1. 一个集合是闭集当且仅当它包含其所有极限点。

推论 IV.2.1.2. 一个集合是开集当且仅当它不包含其任何边界上的点。

定义 IV.2.4 (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$\{\mathbf{x} | \|\mathbf{x} - \mathbf{x}_0\| = \delta\} \cap S = \{\mathbf{x}_0\}$，则称$\mathbf{x}_0$是$S$的一个孤立点。

例 IV.2.1. 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。

IV.2.2 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II.12.3：设函数$\mathbf{f} : \mathbb{R}^n \supseteq D \to \mathbb{R}^m$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

$$\left.\frac{f_i(\mathbf{x})}{\partial x_j}\right|_{\mathbf{x}=\mathbf{x}_0} \text{fi}\, i = 1, \cdots, m, j = 1, \cdots, n$$

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. 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$\Delta\mathbf{x} = t\hat{\mathbf{e}}_j, t \neq 0, t \in \mathbb{R}$，则有

$$\mathbf{f}(\mathbf{x}_0 + t\hat{\mathbf{e}}_j) - \mathbf{f}(\mathbf{x}_0) = \mathbf{L}(t\hat{\mathbf{e}}_j) + \|t\hat{\mathbf{e}}_j\| \, \mathbf{z}(t\hat{\mathbf{e}}_j)$$

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$\Leftrightarrow$

$$\begin{aligned}
&\lim_{t\to 0} \frac{\mathbf{f}(\mathbf{x}_0 + t\hat{\mathbf{e}}_j) - \mathbf{f}(\mathbf{x}_0) - t\mathbf{L}\hat{\mathbf{e}}_j}{t} = \lim_{t\to 0} t\mathbf{z}(t\hat{\mathbf{e}}_j) = \mathbf{0} \\
\Leftrightarrow &\lim_{t\to 0} \frac{\mathbf{f}(\mathbf{x}_0 + t\hat{\mathbf{e}}_j) - f_i(\mathbf{x}_0)}{t} = \mathbf{L}\hat{\mathbf{e}}_j \\
\Leftrightarrow &\lim_{t\to 0} \sum_{i=1}^{n} \frac{f_i(\mathbf{x}_0 + t\hat{\mathbf{e}}_j) - f_i(\mathbf{x}_0)}{t} \hat{\mathbf{u}}_j = \mathbf{L}\hat{\mathbf{e}}_j \\
\Leftrightarrow &\lim_{t\to 0} \sum_{i=1}^{n} \left.\frac{\partial f_i(\mathbf{x})}{\partial x_j}\right|_{\mathbf{x}=\mathbf{x}_0} \hat{\mathbf{u}}_j = \mathbf{L}\hat{\mathbf{e}}_j \qquad \text{（偏导数的定义II.12.4。）}
\end{aligned}$$

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II.12.4：设函数$\mathbf{f} : \mathbb{R}^n \supseteq D \to \mathbb{R}^m$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

$$\begin{aligned}
&\mathbf{A}\Delta\mathbf{x} = \mathbf{f}(\mathbf{x} + \Delta\mathbf{x}) - \mathbf{f}(\mathbf{x}) - \mathbf{L}\Delta\mathbf{x} - [\mathbf{f}(\mathbf{x} + \Delta\mathbf{x}) - \mathbf{f}(\mathbf{x}) - \mathbf{L}'(\Delta\mathbf{x})] \\
\Rightarrow &\|\mathbf{A}\Delta\mathbf{x}\| \leq \|\mathbf{f}(\mathbf{x} + \Delta\mathbf{x}) - \mathbf{f}(\mathbf{x}) - \mathbf{L}\Delta\mathbf{x}\| + \|\mathbf{f}(\mathbf{x} + \Delta\mathbf{x}) - \mathbf{f}(\mathbf{x}) - \mathbf{L}'\Delta\mathbf{x}\| \qquad \text{（三角不等式。）} \\
\Rightarrow &\lim_{\Delta\mathbf{x}\to\mathbf{0}} \frac{\|\mathbf{A}\Delta\mathbf{x}\|}{\|\Delta\mathbf{x}\|} = 0 \qquad \text{（夹逼定理。）}
\end{aligned}$$

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II.12.5：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：

$$\mathbf{f}(\mathbf{y}_k)-\mathbf{f}(\mathbf{y}_{k-1})=\sum_{j=1}^m(f_j(\mathbf{y}_k)-f_j(\mathbf{y}_{k-1}))\hat{\mathbf{e}}_j$$

其中$\{\hat{\mathbf{e}}_i\}$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$\hat{\mathbf{e}}_k$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：

$$\begin{aligned}\mathbf{f}(\mathbf{y}_k)-\mathbf{f}(\mathbf{y}_{k-1})&=\sum_{j=1}^m\left.\frac{\partial f_j}{\partial x_k}\right|_{x_k=c_k}(b_k-a_k)\hat{\mathbf{e}}_j\\&=(b_k-a_k)\left.\frac{\partial \mathbf{f}}{\partial x_k}\right|_{x_k=c_k}, k=1,\cdots,n\end{aligned}$$

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：

$$\lim_{\mathbf{x}\to\mathbf{x}_0}\frac{\mathbf{f}\left(\mathbf{x}\right)-\mathbf{f}\left(\mathbf{x}_0\right)-\mathbf{L}\left(\mathbf{x}-\mathbf{x}_0\right)}{\left\|\mathbf{x}-\mathbf{x}_0\right\|}=\mathbf{0}$$

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。 □

IV.2.3 复合函数求导的链式法则

定理. 定理II.12.7：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$\mathbf{f}\left(\mathbf{x}_0\right)\in E\cap D$处可微分，则复合函数$\mathbf{g}\circ\mathbf{f}$在$\mathbf{x}_0$处可微分，且其全导数

$$\left.\frac{d\mathbf{g}\circ\mathbf{f}\left(\mathbf{x}\right)}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0}=\left.\frac{d\mathbf{g}\left(\mathbf{y}\right)}{d\mathbf{y}}\right|_{\mathbf{y}=\mathbf{f}\left(\mathbf{x}_0\right)}\left.\frac{d\mathbf{f}\left(\mathbf{x}\right)}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0}$$

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$\mathbf{f}\left(\mathbf{x}_0\right)\in\mathrm{dom}\mathbf{g}$且$\mathbf{g}$在$\mathbf{f}\left(\mathbf{x}_0\right)$处可微分，故总存在正实数$\delta'$使得只要$\left\|\mathbf{f}\left(\mathbf{x}\right)-\mathbf{f}\left(\mathbf{x}_0\right)\right\|<\delta'$就有$\mathbf{f}\left(\mathbf{x}\right)\in\mathrm{dom}\mathbf{g}$。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$\delta$使得只要$\left\|\mathbf{x}-\mathbf{x}_0\right\|<\delta$则$\mathbf{x}\in\mathrm{dom}\mathbf{f}$，同时还必存在$\delta'>0$使得这一$\delta$选择下的$\mathbf{x}$满足$\left\|\mathbf{f}\left(\mathbf{x}\right)-\mathbf{f}\left(\mathbf{x}_0\right)\right\|<\delta'$。所以任一满足$\left\|\mathbf{x}-\mathbf{x}_0\right\|<\delta$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$\mathbf{f}\left(\mathbf{x}_0\right)$可导，故存在函数$\mathbf{z}_1$、$\mathbf{z}_2$满足$\lim_{\mathbf{x}\to\mathbf{x}_0}\mathbf{z}_1\left(\mathbf{x}-\mathbf{x}_0\right)=\mathbf{0}$、$\lim_{\mathbf{f}(\mathbf{x})\to\mathbf{f}(\mathbf{x}_0)}\mathbf{z}_2\left(\mathbf{f}\left(\mathbf{x}\right)-\mathbf{f}\left(\mathbf{x}_0\right)\right)=\mathbf{0}$，且

$$\mathbf{f}\left(\mathbf{x}\right)-\mathbf{f}\left(\mathbf{x}_0\right)-\left.\frac{d\mathbf{f}\left(\mathbf{x}\right)}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0}\left(\mathbf{x}-\mathbf{x}_0\right)=\left\|\mathbf{x}-\mathbf{x}_0\right\|\mathbf{z}_1\left(\mathbf{x}-\mathbf{x}_0\right)$$
$$\mathbf{g}\left(\mathbf{f}\left(\mathbf{x}\right)\right)-\mathbf{g}\left(\mathbf{f}\left(\mathbf{x}_0\right)\right)-\left.\frac{d\mathbf{g}\left(\mathbf{y}\right)}{d\mathbf{y}}\right|_{\mathbf{y}=\mathbf{f}\left(\mathbf{x}_0\right)}\left(\mathbf{f}\left(\mathbf{x}\right)-\mathbf{f}\left(\mathbf{x}_0\right)\right)=\left\|\mathbf{f}\left(\mathbf{x}\right)-\mathbf{f}\left(\mathbf{x}_0\right)\right\|\mathbf{z}_2\left(\mathbf{f}\left(\mathbf{x}\right)-\mathbf{f}\left(\mathbf{x}_0\right)\right)$$

把$\mathbf{g}\left(\mathbf{f}\left(\mathbf{x}\right)\right)$记为$\mathbf{g}\circ\mathbf{f}\left(\mathbf{x}\right)$，并把上面的第一条式子代入第二条，得

$$\begin{aligned}
&\mathbf{g}\circ\mathbf{f}\left(\mathbf{x}\right)-\mathbf{g}\circ\mathbf{f}\left(\mathbf{x}_0\right)-\left.\frac{d\mathbf{g}\left(\mathbf{y}\right)}{d\mathbf{y}}\right|_{\mathbf{y}=\mathbf{f}\left(\mathbf{x}_0\right)}\left[\left.\frac{d\mathbf{f}\left(\mathbf{x}\right)}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0}\left(\mathbf{x}-\mathbf{x}_0\right)+\left\|\mathbf{x}-\mathbf{x}_0\right\|\mathbf{z}_1\left(\mathbf{x}-\mathbf{x}_0\right)\right]\\
=&\left\|\mathbf{f}\left(\mathbf{x}\right)-\mathbf{f}\left(\mathbf{x}_0\right)\right\|\mathbf{z}_2\left(\mathbf{f}\left(\mathbf{x}\right)-\mathbf{f}\left(\mathbf{x}_0\right)\right)\\
\Leftrightarrow&\mathbf{g}\circ\mathbf{f}\left(\mathbf{x}\right)-\mathbf{g}\circ\mathbf{f}\left(\mathbf{x}_0\right)-\left.\frac{d\mathbf{g}\left(\mathbf{y}\right)}{d\mathbf{y}}\right|_{\mathbf{y}=\mathbf{f}\left(\mathbf{x}_0\right)}\left.\frac{d\mathbf{f}\left(\mathbf{x}\right)}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0}\left(\mathbf{x}-\mathbf{x}_0\right)\\
=&\left\|\mathbf{x}-\mathbf{x}_0\right\|\left.\frac{d\mathbf{g}\left(\mathbf{y}\right)}{d\mathbf{y}}\right|_{\mathbf{y}=\mathbf{f}\left(\mathbf{x}_0\right)}\mathbf{z}_1\left(\mathbf{x}-\mathbf{x}_0\right)+\left\|\mathbf{f}\left(\mathbf{x}\right)-\mathbf{f}\left(\mathbf{x}_0\right)\right\|\mathbf{z}_2\left(\mathbf{f}\left(\mathbf{x}\right)-\mathbf{f}\left(\mathbf{x}_0\right)\right)
\end{aligned}$$

由三角不等式，又有*

$$\begin{aligned}\|\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0)\| &= \left\| \left.\frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0}(\mathbf{x}-\mathbf{x}_0)+\|\mathbf{x}-\mathbf{x}_0\|\,\mathbf{z}_1(\mathbf{x}-\mathbf{x}_0)\right\| \\ &\le \left\| \left.\frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0}(\mathbf{x}-\mathbf{x}_0)\right\| + \|\mathbf{x}-\mathbf{x}_0\|\,\|\mathbf{z}_1(\mathbf{x}-\mathbf{x}_0)\| \\ &\le k\,\|\mathbf{x}-\mathbf{x}_0\| + \|\mathbf{x}-\mathbf{x}_0\|\,\|\mathbf{z}_1(\mathbf{x}-\mathbf{x}_0)\|\end{aligned}$$

故

$$\begin{aligned}&\mathbf{g}\circ\mathbf{f}(\mathbf{x})-\mathbf{g}\circ\mathbf{f}(\mathbf{x}_0)-\left.\frac{d\mathbf{g}(\mathbf{y})}{d\mathbf{y}}\right|_{\mathbf{y}=\mathbf{f}(\mathbf{x}_0)}\left.\frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0}(\mathbf{x}-\mathbf{x}_0)\\ &\le \|\mathbf{x}-\mathbf{x}_0\|\left.\frac{d\mathbf{g}(\mathbf{y})}{d\mathbf{y}}\right|_{\mathbf{y}=\mathbf{f}(\mathbf{x}_0)}\mathbf{z}_1(\mathbf{x}-\mathbf{x}_0)+\left(k\,\|\mathbf{x}-\mathbf{x}_0\|+\|\mathbf{x}-\mathbf{x}_0\|\,\|\mathbf{z}_1(\mathbf{x}-\mathbf{x}_0)\|\right)\mathbf{z}_2(\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0))\\ &\le \|\mathbf{x}-\mathbf{x}_0\|\left\{\left\|\left.\frac{d\mathbf{g}(\mathbf{y})}{d\mathbf{y}}\right|_{\mathbf{y}=\mathbf{f}(\mathbf{x}_0)}\mathbf{z}_1(\mathbf{x}-\mathbf{x}_0)\right\|+\left(k+\|\mathbf{z}_1(\mathbf{x}-\mathbf{x}_0)\|\right)\mathbf{z}_2(\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0))\right\}\end{aligned}$$

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。 □

IV.2.4 反函数定理和隐函数定理

引理 IV.2.1. 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$\|\mathbf{L}\mathbf{x}\|\ge p\,\|\mathbf{x}\|\,\forall\mathbf{x}\in\mathscr{V}_n$。

证明. 如果$\mathbf{L}$不是单射，则存在$\mathbf{x}_0\ne\mathbf{0}$满足$\mathbf{L}\mathbf{x}_0=\mathbf{0}$，即有$\|\mathbf{L}\mathbf{x}_0\|=0<m\,\|\mathbf{x}\|$。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II.11.3，$\exists k>0,\|\mathbf{L}^{-1}\mathbf{y}\|\le k\,\|\mathbf{y}\|\,\forall\mathbf{y}\in\mathscr{W}_m$。令$p=1/k$则有$p\,\|\mathbf{x}\|=m\,\|\mathbf{L}^{-1}\mathbf{L}\mathbf{x}\|\le pk\,\|\mathbf{L}\mathbf{x}\|=\|\mathbf{L}\mathbf{x}\|$ □

引理 IV.2.2. 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$\left.\frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0}$是单射。则存在$\delta>0$和$M>0$使得只要$\|\mathbf{x}-\mathbf{x}_0\|<\delta$就有

$$\left\|\left.\frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0}\mathbf{y}\right\|\ge M\,\|\mathbf{y}\|\,\forall\mathbf{y}\in\mathbb{R}^n$$

*此处用到定理II.11.3。

证明. 记函数$\mathbf{f}$的导函数为$\mathbf{L}$。由引理IV.2.1，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 IV.2.3. 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$\left.\frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0}$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IV.2.1，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"$\delta$-球是凸的"。由于导函数连续是在整个$\delta$-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$\frac{d\pi_k(\mathbf{x})}{d\mathbf{x}} \equiv \pi_k(\mathbf{x})$。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

$$|\pi_k((\mathbf{L}(\mathbf{x} + t_k\mathbf{z}) - \mathbf{L}(\mathbf{x}_0))\mathbf{z})| + |\pi_k(\mathbf{L}(\mathbf{x}_0)\mathbf{z})| \leq |\pi_k(\mathbf{L}(\mathbf{x} + t_k\mathbf{z})\mathbf{z})|$$

以上不等式联合有

$$\|\mathbf{f}(\mathbf{x}') - \mathbf{f}(\mathbf{x})\| \geq |\pi_k((\mathbf{L}(\mathbf{x} + t_k\mathbf{z}) - \mathbf{L}(\mathbf{x}_0))\mathbf{z})| + |\pi_k(\mathbf{L}(\mathbf{x}_0)\mathbf{z})|$$

由事实$\|\mathbf{x}\| \leq \sqrt{n}\max\{|x_i|\} \equiv \sqrt{n}\max\{\pi_i(\mathbf{x})\}$（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

$$\sqrt{n}\,|\pi_k(\mathbf{L}(\mathbf{x}_0)\mathbf{z})| \geq \|\mathbf{L}(\mathbf{x}_0)\mathbf{z}\|$$

再次利用投影不大于原长，有

$$|\pi_k((\mathbf{L}(\mathbf{x} + t_k\mathbf{z}) - \mathbf{L}(\mathbf{x}_0))\mathbf{z})| \leq \|(\mathbf{L}(\mathbf{x} + t_k\mathbf{z}) - \mathbf{L}(\mathbf{x}_0))\mathbf{z}\|$$

再次联合这些不等式有

$$\begin{aligned}\|\mathbf{f}(\mathbf{x}') - \mathbf{f}(\mathbf{x})\| &\geq \frac{1}{\sqrt{n}}\|\mathbf{L}(\mathbf{x}_0)\mathbf{z}\| - \|(\mathbf{L}(\mathbf{x} + t_k\mathbf{z}) - \mathbf{L}(\mathbf{x}_0))\mathbf{z}\| \\ &\geq 2M\|\mathbf{z}\| - M\|\mathbf{z}\| = M\|\mathbf{x}' - \mathbf{x}\|\end{aligned}$$

□

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 IV.2.2 (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$\mathbf{L}(\mathbf{x})\equiv\left.\frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}}$。若$\mathbf{L}(\mathbf{x}_0)$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

$$\left.\frac{d\mathbf{f}^{-1}(\mathbf{y})}{d\mathbf{y}}\right|_{\mathbf{y}=\mathbf{f}(\mathbf{x})}=\mathbf{L}^{-1}(\mathbf{x})$$

其中$\mathbf{y}=\mathbf{f}(\mathbf{x})$。

证明. 我们先列出引理IV.2.2和IV.2.3的结论。由于$\mathbf{f}$在$\mathbf{x}_0$处连续可微，且$\mathbf{L}(\mathbf{x}_0)$是单射，故：

- 由引理IV.2.2，对$\mathbf{x}_0$的任一邻域$N=\{\mathbf{x}|\,\|\mathbf{x}-\mathbf{x}_0\|<\delta\}$（$\delta>0$为任一正实数），都能找到正实数$M(\mathbf{y})>0$满足$\|\mathbf{L}(\mathbf{x})\mathbf{y}\|\geq M\|\mathbf{y}\|\,\forall\mathbf{y}\in\mathbb{R}^n$。进一步地，再由引理IV.2.1可知导函数$\mathbf{L}(\mathbf{x})$在$N$上的每个值都是单射线性变换。再由于$\mathbf{L}(\mathbf{x})$在$N$上都是单射线性变换且其定义域和陪域维数相同，故$\mathbf{L}(\mathbf{x})$在$N$上的每个值都是双射（同构）线性变换。
- 由引理IV.2.3，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IV.2.3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$\delta_1$。

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IV.2.3，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$C = \left\{\mathbf{y} \middle| \|\mathbf{y} - \mathbf{y}_1\| < \widetilde{M}\right\}$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IV.2.3，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IV.2.3，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II.11.3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

$$\left.\frac{d\mathbf{f}^{-1}(\mathbf{y})}{d\mathbf{y}}\right|_{\mathbf{y}=\mathbf{y}}=\mathbf{L}^{-1}(\mathbf{x}),\mathbf{y}=\mathbf{f}(\mathbf{x}),\forall\mathbf{x}\in N$$

*这个基本定理可由极限的$\delta-\epsilon$语言证明，很多地方有，此略。

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IV.2.3，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$\delta$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

$$\begin{aligned}
\left\|\left(\mathbf{L}^{-1}(\mathbf{x}) - \mathbf{L}^{-1}(\mathbf{x}_1)\right)\mathbf{z}\right\| &= \left\|\mathbf{L}^{-1}(\mathbf{x})(\mathbf{L}(\mathbf{x}) - \mathbf{L}(\mathbf{x}_1))\mathbf{L}^{-1}(\mathbf{x}_1)\mathbf{z}\right\| \\
&\leq \frac{1}{M}\left\|(\mathbf{L}(\mathbf{x}) - \mathbf{L}(\mathbf{x}_1))\mathbf{L}^{-1}(\mathbf{x}_1)\mathbf{z}\right\| \\
&\leq \frac{1}{M}\|\mathbf{L}(\mathbf{x}) - \mathbf{L}(\mathbf{x}_1)\|\left\|\mathbf{L}^{-1}(\mathbf{x}_1)\mathbf{z}\right\| \\
&\leq \frac{1}{M^2}\|\mathbf{L}(\mathbf{x}) - \mathbf{L}(\mathbf{x}_1)\|\|\mathbf{z}\|
\end{aligned}$$

由线性变换的范的定义（最大下界），上述不等式⇔

$$\left\|\mathbf{L}^{-1}(\mathbf{x}) - \mathbf{L}^{-1}(\mathbf{x}_1)\right\| \leq \frac{1}{M^2}\|\mathbf{L}(\mathbf{x}) - \mathbf{L}(\mathbf{x}_1)\| \leq \frac{\epsilon'}{M^2}$$

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$\delta$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。 □

IV.2.5 等距变换的表示定理

引理 IV.2.4. 设$(\mathscr{E}, d)$是一个欧几里得空间，$\mathscr{V}$是其平移空间，$i : \mathscr{E} \to \mathscr{E}$是$\mathscr{E}$上的一个等距变换。记$\mathbf{u}' = i(X+\mathbf{u}) - i(X), \mathbf{v}' = i(X+\mathbf{v}) - i(X), \mathbf{u}, \mathbf{u}', \mathbf{v}, \mathbf{v}' \in \mathscr{V}, X \in \mathscr{E}$，则$(\mathbf{u}'|\mathbf{v}') = (\mathbf{u}|\mathbf{v})$对任意$\mathbf{u}, \mathbf{u}', \mathbf{v}, \mathbf{v}' \in \mathscr{V}$均成立。

证明. 利用欧几里得空间的定义和等距变换的定义，有

$$(\mathbf{u}'|\mathbf{u}') = d^2(i(X), i(X+\mathbf{u})) = d^2(X, X+\mathbf{u}) = (\mathbf{u}|\mathbf{u})$$

同理有$(\mathbf{v}'|\mathbf{v}') = (\mathbf{v}|\mathbf{v})$。由内积的性质，$(\mathbf{u}-\mathbf{v}|\mathbf{u}-\mathbf{v}) = (\mathbf{u}|\mathbf{u}) + (\mathbf{v}|\mathbf{v}) - 2(\mathbf{u}|\mathbf{v})$，则

$$\begin{aligned}
2(\mathbf{u}'|\mathbf{v}') &= (\mathbf{u}'|\mathbf{u}') + (\mathbf{v}'|\mathbf{v}') - (\mathbf{u}'-\mathbf{v}'|\mathbf{u}'-\mathbf{v}') \\
&= (\mathbf{u}|\mathbf{u}) + (\mathbf{v}|\mathbf{v}) \\
&\quad - (i(X+\mathbf{u}) - i(X) - i(X+\mathbf{v}) + i(X)\,|i(X+\mathbf{u}) - i(X) - i(X+\mathbf{v}) + i(X)) \\
&= (\mathbf{u}|\mathbf{u}) + (\mathbf{v}|\mathbf{v}) \\
&\quad - d^2(i(X+\mathbf{v}), i(X+\mathbf{v}) + (i(X+\mathbf{u}) - i(X+\mathbf{v}))) \\
&= (\mathbf{u}|\mathbf{u}) + (\mathbf{v}|\mathbf{v}) - d^2(X+\mathbf{v}, X+\mathbf{u}) \\
&= (\mathbf{u}|\mathbf{u}) + (\mathbf{v}|\mathbf{v}) - (X+\mathbf{u}-(X+\mathbf{V})\,|X+\mathbf{u}-(X+\mathbf{v})) \\
&= (\mathbf{u}|\mathbf{u}) + (\mathbf{v}|\mathbf{v}) - (\mathbf{u}-\mathbf{v}|\mathbf{u}-\mathbf{v}) \\
&= 2(\mathbf{u}|\mathbf{v})
\end{aligned}$$

□

引理 IV.2.5. 设$(\mathscr{E}, d)$是一个欧几里得空间，$\mathscr{V}$是其平移空间，$i:\mathscr{E}\to\mathscr{E}$是$\mathscr{E}$上的一个等距变换。定义关于一点$X\in\mathscr{E}$的映射$\mathbf{Q}_X:\mathscr{V}\to\mathscr{V}, \mathbf{Q}_X\mathbf{u} = i(X+\mathbf{u}) - i(X)\,\forall\mathbf{u}\in\mathscr{V}$，则

1. $\mathbf{Q}_X$与X的选择无关（故可略去下标$\mathbf{X}$）；
2. $\mathbf{Q}$是正交算符；
3. 对给定的i，Q唯一存在。

证明. 由命题中$\mathbf{Q}_X$的定义有

$$\mathbf{Q}_X(Y-X) = i(Y) - i(X)\,\forall X, Y\in\mathscr{E}$$

由引理IV.2.4有

$$\begin{aligned}
(\mathbf{Q}_X\mathbf{u}|\mathbf{Q}_X\mathbf{v}) &= (i(X+\mathbf{u}) - i(X)\,|i(X+\mathbf{u}) - i(X)) \\
&= (\mathbf{u}|\mathbf{v}), \forall\mathbf{u},\mathbf{v}\in\mathscr{V}
\end{aligned}$$

如果$\mathbf{Q}_X$是线性算符，则上述结论就证明了$\mathbf{Q}_X$是正交算符。以下进一步证明映射$\mathbf{Q}_X$是一个线

性算符。由

$$
\begin{aligned}
&\left(\mathbf{Q}_X\left(\alpha_1\mathbf{u}_1+\alpha_2\mathbf{u}_2\right)-\mathbf{Q}_X\left(\alpha_1\mathbf{u}_1\right)-\mathbf{Q}_X\left(\alpha_2\mathbf{u}_2\right)\middle|\mathbf{Q}_X\mathbf{v}\right)\\
=&\left(i\left(X+\alpha_1\mathbf{u}_1+\alpha_2\mathbf{u}_2\right)-i\left(X\right)-\left[i\left(X+\alpha_1\mathbf{u}_1\right)-i\left(X\right)\right]-\left[i\left(X+\alpha_2\mathbf{u}_2\right)-i\left(X\right)\right]\right.\\
&\left.\middle|i\left(X+\mathbf{v}\right)-i\left(X\right)\right)\\
=&\left(i\left(X+\alpha_1\mathbf{u}_1+\alpha_2\mathbf{u}_2\right)-i\left(X\right)\middle|i\left(X+\mathbf{v}\right)-i\left(X\right)\right)\\
&-\left(i\left(X+\alpha_1\mathbf{u}_1\right)-i\left(X\right)\middle|i\left(X+\mathbf{v}\right)-i\left(X\right)\right)\\
&-\left(i\left(X+\alpha_2\mathbf{u}_2\right)-i\left(X\right)\middle|i\left(X+\mathbf{v}\right)-i\left(X\right)\right)\quad\text{（用到引理IV.2.4）}\\
=&\left(\alpha_1\mathbf{u}_1+\alpha_2\mathbf{u}_2\middle|\mathbf{v}\right)-\left(\alpha_1\mathbf{u}_1\middle|\mathbf{v}\right)-\left(\alpha_2\mathbf{u}_2\middle|\mathbf{v}\right)\\
=&0
\end{aligned}
$$

得

$$\mathbf{Q}_X\left(\alpha_1\mathbf{u}_1+\alpha_2\mathbf{u}_2\right)-\mathbf{Q}_X\left(\alpha_1\mathbf{u}_1\right)-\mathbf{Q}_X\left(\alpha_2\mathbf{u}_2\right)=\mathbf{0}$$

即$\mathbf{Q}_X$是线性算符。结合上一结论$\mathbf{Q}_X$就是正交算符。第2个命题得证。

要证明$\mathbf{Q}_X$不依赖X的选择，设有另一$X'\in\mathscr{E}$，则对任意$\mathbf{u}\in\mathscr{V}$有

$$
\begin{aligned}
\mathbf{Q}_{X'}\mathbf{u}&=i\left(X'+\mathbf{u}\right)-i\left(X'\right)\\
&=i\left(X+\left[\left(X'-X\right)\mathbf{u}\right]\right)-i\left(X+\left(X'-X\right)\right)\\
&=\mathbf{Q}_X\left(\left(X'-X\right)+\mathbf{u}\right)-\mathbf{Q}_X\left(X'-X\right)\\
&=\mathbf{Q}_X\left[\left(X'-X\right)+\mathbf{u}-\left(X'-X\right)\right]\\
&=\mathbf{Q}_X\mathbf{u}
\end{aligned}
$$

即得证。

至此，对于给定的i，$\mathbf{Q}$的存在性已证明。$\mathbf{Q}$的唯一性也是显然的。设另有线性算符$\mathbf{P}:\mathscr{V}\to\mathscr{V},\mathbf{P}\mathbf{u}=i\left(X+\mathbf{u}\right)-i\left(X\right)\forall\mathbf{u}\in\mathscr{V}$，则易验

$$\left(\mathbf{P}-\mathbf{Q}\right)\mathbf{u}=\mathbf{P}\mathbf{u}-\mathbf{Q}\mathbf{u}=\mathbf{0}$$

□

定理 (等距变换的表示定理). 定理II.9.1：设$(\mathscr{E},d)$是一个欧几里德空间，$\mathscr{V}$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$i:\mathscr{E}\to\mathscr{E},i\in\mathscr{I}$都可表示为

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

其中$\mathbf{Q}_i$是一个正交算符，关于i唯一存在。

证明. 由引理IV.2.5，任一$\mathscr{E}$上的等距变换i都唯一对应一个$\mathscr{V}$上的正交算符$\mathbf{Q}$满足$\mathbf{Q}\mathbf{u} = i\left(X + \mathbf{u}\right) - i\left(X\right) \forall X \in \mathscr{E}, \mathbf{u} \in \mathscr{V}$。对任一$X \in \mathscr{E}$，可令$\mathbf{u} = X - X_0$，则对任一$i$存在唯一$\mathbf{Q}_i$满足$\mathbf{Q}_i\mathbf{u} = \mathbf{Q}_i\left(X - X_0\right) = i\left(X\right) - i\left(X_0\right) \Rightarrow i\left(X\right) = i\left(X_0\right) + \mathbf{Q}_i\left(X - X_0\right)$。 □

参考文献

[1] 周胜林, 刘西民. 线性代数与解析几何[M]. 高等教育出版社, 2012.

[2] HOFFMAN K, KUNZE R. Linear Algebra[M]. 2nd ed. Prentice-Hall, Inc., 1971.

[3] HASSANI S. Mathematical physics : a modern introduction to its foundations[M]. New York: Springer, 1999.

[4] NOLL W. Lectures on the Foundations of Continuum Mechanics and Thermodynamics[G] //The Foundations of Mechanics and Thermodynamics: Selected Papers. Berlin, Heidelberg: Springer Berlin Heidelberg, 1974: 293-324.

[5] 王全迪, 郭艾, 杨立洪. 高等数学（上册）[M]. 高等教育出版社, 2009.

[6] 王全迪, 郭艾, 杨立洪. 高等数学（下册）[M]. 高等教育出版社, 2009.

[7] WILLIAMSON R E, CROWELL R H, TROTTER H F. Calculus of Vector Functions[M]. 3rd ed. Prentice Hall, Inc., 1972.

[8] REISS J, SPRENGER J. Scientific Objectivity[G]//ZALTA E N. The Stanford Encyclopedia of Philosophy. Winter 2020. Metaphysics Research Lab, Stanford University, 2020.