《聚合物流变学基础》讲义

孙尉翔

mswxsun@scut.edu.cn

2021-03-21

前言

本讲义是华南理工大学材料科学与工程学院硕士研究生选修课《聚合物流变学基础》的辅助资料。该课共 32 学时，面向本科是化学类专业背景的学生。本课拟重点讲述以下内容：

- 流变学的连续介质力学基础（包括必要的数学基础）
- 线性粘弹性本构方程
- 非线性粘弹性本构方程
- 流变测量学

连续介质力学是流变学的基础。而必要的线性代数和向量函数微积分又是连续介质力学的数学基础。流变学专注于非线性粘弹性本构关系的研究，是连续介质力学的一个小分支。然而，由于流变学跟高分子科学和工业联系十分紧密，因此高分子学科的学生往往有学习流变学的强烈需求。由于历史的缘故，高分子专业往往是开设在一所大学的化学与化学工程、材料科学与工程等学院，学生普遍不具备学习连续介质力学的数学基础。高分子专业背景的学生，要成为一名流变学领域的研究者需要依次跨跃数学、力学和流变学三道槛。

在化学和材料类专业本科的一般课程设置中，学生从《工程力学》、《材料力学》、《化工原理》等课程中接触过的思想，其实都属于连续介质力学。

例如，《工程力学》或《材料力学》一开始就声明了学科的基本假设包括：连续性假设、均匀性假设、各向同性假设和小形变假设。其中的连续性假设，就是连续介质力学的一般性假设。均匀性、各向同性假设使我们可以考虑一个不随位置和方向变化的标量值模量。小形变假设使得课程覆盖的问题均假设虎克弹性。但是，这两门课主要关注的对象是刚体或虎克固体的静力学，既物体处于静止且形变恒定状态下的应力和应变与载荷的关系，因此可进一步利用力系简化原则和平衡条件解题，使得这些题目只需运用标量的代数公式或含简单积分的公式就可以解决。因此，《工程力学》和《材料力学》的知识无法解决更一般的形变问题。其实《工程力学》和《材料力学》概念上还包括运动学和动力学内容，研究的是刚体或虎克固体的运动与形变与载荷的关系，机械能转化效率与机械波的传递等问题，常明显地称作《工程动力学》和《材料动力学》。由于这两门课常常直接基于分析力学和连续介质力学的基础讲述，于是少见于化学和材料类专业的课程中。

又例如，《化工原理》中关于传质部分的内容也声明了流体力学的基本假设：连续性假设和不可压缩假设。事实上，在这一课程中，还默认了流体的均匀性假设和各向同性假设（反过来看，《工程力学》和《材料力学》课程中其实允许了可压缩性，因此介绍了柏松比参数）。流体力学也分为静力学和动力学（称为“流体静力学”和“流体动力学”），二者在《化工原理》课程中都有涉及。这门课关心的对象是理想流体与牛顿流体，原则上既考虑低雷诺数的问题（层

流假设）也考虑高雷诺数问题（层流假设失效）。在介绍流体动力学部分时，明确介绍了“质量守恒”和连续性方程，并以直观的方式引入了动量守恒因素，最终推出纳维–斯托克斯方程（Navier–Stokes equation）。这些都是连续介质力学针对牛顿流体的应用。然而，由于《化工原理》的研究目标仅为解决传质传热问题，加上数学基础设置上的限制，《化工原理》课程中的应用例题一般仅限于层流（线性化的纳维–斯托克斯方程）、一维、定态流问题，甚至可以越过纳维–斯托克斯方程的明显引入，直接采用微积分思想列出微分方程解题；往往只关心压差和总流量的关系问题（例如 Hagen–Poisseuille 方程），避免细化到流速分布场等涉及场函数微积分定理的数学知识的题目，以保持全书篇幅在合理的范围内。

上述两门课中的共同假设就是连续性假设，实际上就是连续介质力学的应用。一般地，连续介质力学在牛顿力学定律的基础上仅需增加这一条假设。但是上述这两门课分别只关心虎克固体和牛顿流体。而流变学关心的非牛顿流动和塑性流动，则属于粘弹性流体或粘塑性流体。这些力学或流动行为，高分子专业的同学在《高分子物理》中也接触过一些基本概念和基本现象，但仅限于应力应变关系，书中没有进一步介绍用于预测和解决实际问题的连续介质力学理论框架。同时对于这些特殊流动现象本身，又往往急于直接从分子层面定性解释，却又略去了基本的统计力学推导。实际上，大部分高分子专业本科的《高分子物理》课程只是一种定性程度的科普。

流变学研究的对象往往被称为复杂流体或软物质（包括但不仅限于高分子）。这类体系的统计力学研究属于（软）凝聚态物理的内容。一般而言，凝聚态物理的研究对像是大量遵循量子力学或经典力学的微观粒子形成的离散的整体（因此并不假定连续性）。这些微观粒子除了满足量子力学或经典力学的假设之外，还需要承认统计力学的公理，包括等概率假设、系综平均的假设和各态遍历假设等等，才能预测系统的整体宏观性质。为了解决复杂的多体相互作用问题，不同体系的统计力学模型还常常引入很多重要的近似思想，如等效介质、平均场、重整化群等，大量运用了近代数学和理论物理的思想和手段，事实上难以完全地嵌入到化学与材料学科的本科教学体系中。

近代软凝聚态物理的发展提供了大量复杂流体体系的微观模型，大大丰富了流变学的应用范畴，因此这些与流变学相关的统计力学模型也常被纳入流变学的教材当中。很多流变学教材都包括了特定体系的统计力学模型的介绍，例如聚合物流变学的内容常常包括从链状分子的动力学出发的流变学模型介绍。流变学需要的连续介质力学和数学基础则常常只在开头花一章左右的篇幅进行很简要的介绍，并在后续章节中少用或不用。至于统计力学模型涉及的更广阔的平衡态统计力学基础，在一个流变学教材的有限篇幅限制内就更无法一一陈列了。专门的《连续介质力学》、《平衡态统计力学》课本又都自成体系，相关的内容常常超出软凝聚态物理和流变学一般关注的范畴（例如连续介质力学中针对晶态固体弹性的的对称性问题、

平衡态统计力学常常注重量子统计、电磁场响应等）。这些专门的课程本是物理学专业的学生的必修或选修课程，不能指望大部分化学或材料类专业的学生都去学习。因此，今天的高分子研究面临的尴尬是，高分子专业培养的本科乃至研究生的数理基础无法满足高分子学科前沿研究的需要。高分子专业的价值越来越局限于为已有的高分子工业输出职业人才。事实上，今天写在在高分子物理教科书中的统计物理理论成果，当初都是物理学家做出来的；当时的大学还鲜有专门的高分子专业，高分子问题属于物理学的前沿问题。无奈的是当高分子发展到自己成为独立本科专业的今天，反而发现除高分子专业自己培养的人才之外，已少有理论物理背景的人士加入到高分子物理的研究中来了。这也许是恰恰是因为"高分子学科有自己的本科专业"已成为所有人的印象，其他学科的人鲜有"越殂代疱"者。因此，高分子物理前沿研究的人才，今后一段时间仍需靠高分子专业自己的培养方案来解决。面向高分子背景乃至材料学背景的教科书不仅不应该回避必要的数学和物理基础去作简化，而且还要不照搬面向物理学背景的教材惯例，而重新为化类学生专门去撰写。

综合上述考虑，本讲义仅介绍流变学的连续介质力学基础以及面向化类学生的相关的数学基础。为此本讲义包括了一个比较详细的数学准备，希望讲义的力学涉及到的数学知识都能找到相关定理和证明。如果由于笔者时间和水平不足、暂未纳入讲义的数学准备部分的数学知识，也至少给出证明和推导的参考资料，力求做到全课程中的理论推导不发生中断。本讲义的数学准备还尝试在不引入"协变和逆变"的概念的条件下保持自洽性和完整性，因为针对以流体和无定形固体为主的软物质的特点，可以不引入。这也使得无需正式地介绍张量的概念和高阶张量分析知识，使讲义事实上要求的数学概念不超出本科《高等数学》和《线性代数》的范围过多。具体地，本讲义数学部分是以 Hoffman and Kunze 的线性代数和 William, Crowell and Trotter 的向量函数微积分为蓝本编写的（见 §I.3的建议参考书目）。

本讲义的连续介质力学部分，比某些连续介质力学课本花费了稍多一些的篇幅去讲解事件世界和标架等时空概念。这些较为近代的时空概念，物理专业的学生也许在其他基础力学课程中已经建立了，因此在很多连续介质力学课本中，这类知识多是作为回顾、复习而被简述。但是，要正确理解物理客观性和物质客观性——本构关系的重要要求，必须对不依赖观察者的那些抽象概念十分清楚，而这些是化学专业的学生普遍并不具备的。

在准备此门课的过程中我受到了很多无私的帮助。在此我需重点感谢：我的博士生导师童真教授把我引入到了流变学的世界；李俊杰老师和罗锦添博士牺牲大量私人时间为我回答了大量数学的问题。

孙尉翔

2020 年 10 月

目录

第一部分 引入

I.1 流变学的定义和研究内容

流变学（rheology）在牛津语言网站上的释义是：研究物质的变形和流动——特别是液体的非牛顿流动和固体的塑性流动——的科学（The branch of science that deals with the deformation and flow of matter, esp. the non-Newtonian flow of liquids and the plastic flow of solids）。

流变学是一门力学，属于经典力学范畴。力学包括静力学（statics）、运动学（kinematics）和动力学 (dynamics）三种问题。静力学集中考虑物体处于力的平衡之下的问题。运动学定量描述物体的位置和状态随时间的变化关系。动力学研究和分析物体运动状态改变的原因。[1]p. 1

在经典力学的运动学中，为了描述物体的位置，我们需要先选择参考系（frame of reference），建立坐标系（coordinates），使用位置向量来表示质点的位移。我们假设时间是连续的，把质点的位移写成时间的函数，在可导的前提下，其速度和加速度分别是位移的一阶和二阶时间导数[1] 附录 A,p. 422 （如图I.1.1所示）：

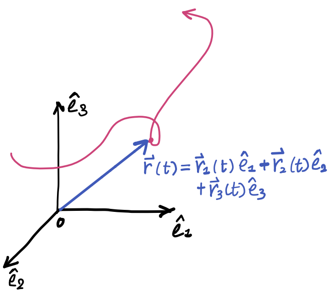

图 I.1.1: 质点运动学

$$\mathbf{r} = \mathbf{r}(t) = r_1\widehat{\mathbf{e}}_1 + r_2\widehat{\mathbf{e}}_2 + r_3\widehat{\mathbf{e}}_3,$$
$$\mathbf{v} = \frac{d\mathbf{r}}{dt}, \mathbf{a} = \frac{d^2\mathbf{r}}{dt^2}$$

经典力学的动力学遵循牛顿运动定律，引入了“力”的概念作为物体运动状态变化的原因（第一定律），并建立了定量关系（第二定律），以及一个守恒律（第三定律）。牛顿运动定律在且仅在惯性系下成立。作为质点的动力学，在选定惯性系下，

$$\mathbf{F} = m\frac{d\mathbf{v}}{dt}$$

，其中 m 是质点的质量，是惯性（改变物定运动状态的代价）大小的量度。

对于一个可以发生形变和流动的物体，上述的质点运动学和动力学是不够用的。

以下是两种适用的力学：

- 连续介质力学（continuum mechanics）：把宏观物体视为可无限微分的连续体。从热力学基本定律和牛顿运动定律（惯性系）出发，建立物体运动与形变的本构关系。
- 统计力学（statistical mechanics）：考虑组成物质的微观粒子的运动规律，基于统计力学的假设推出宏观物体的本构关系。高分子物理中，从链统计出发建立的粘流模型和橡胶弹性模型就是统计力学在高分子物理中的应用。

描述宏观物体形变规律的语言是连续介质力学。统计力学只是进一步给出微观原因，它关于宏观规律的预测仍旧要使用连续介质力学的语言来表述。

在连续介质力学中，一个物体的独特力学性质，是由它的本构关系/方程（constitutive relation/equation）来描述的。在一般物理学中，本构关系是指描述物质对外场激励的响应的定量关系：外场 $\mathcal{S}(\mathbf{r},t) \to$ 材料特性函数 $\mathcal{G}(\Delta\mathbf{x},\Delta t) \to$ 材料响应 $\mathcal{R}(\mathbf{x},t)$。其中，$\mathbf{r}$ 是空间任一位置，$\mathbf{x}$ 是处于材料内部的某位置。$\mathcal{G}(\Delta\mathbf{x},\Delta t)$ 是材料本身的性质，称为物料函数（material function）。我们留意到，$\mathcal{S}$ 和 $\mathcal{R}$ 都是场函数。物料函数是对材料观测的时间和空间间隔的函数，在数学形式上也是场函数[2]§ 9.7,p.192。以下我们通过两个例题来理解本构关系的概念。这两个例题都是以前的课程中介绍过的，请注意参见例中标注的引用。

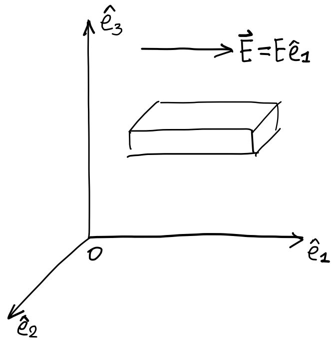

图 I.1.2: 例I.1.1

例 I.1.1. 此例中，外场是稳恒电场 $\mathbf{E}(\mathbf{r},t)=\mathbf{E}=E\widehat{\mathbf{e}}_1$，材料的物性函数是电极化率 $\chi(\Delta\mathbf{x},\Delta t)$（如图*I.1.2*所示）。对于各向同性均匀材料 $\chi = const$。材料的响应是电极化强度密度 $\mathbf{P}(\mathbf{x},t)$，表示材料内 $\mathbf{x}$ 处单位体系的电极化强度，即 $\mathbf{P}(\mathbf{x},t)\,dV$ 表示 $\mathbf{x}$ 处的微体积 dV 的电极化强度。实验表明，当电场不太大时[3]§18.2.3,p. 57，

$$\mathbf{P}(\mathbf{x},t)=\chi\varepsilon_0\mathbf{E}(\mathbf{r}=\mathbf{x},t)$$

其中 ε_0 是真空电容率。故本例中

$$\mathbf{P}(\mathbf{x},t)=\chi\varepsilon_0 E\widehat{\mathbf{e}}_1$$

这是该材料的电极化本构方程。

例 I.1.2. 此例中外场是一个梯度场，例如温度场 $T(\mathbf{r},t)=kr_1\widehat{\mathbf{e}}_1, \nabla T(\mathbf{r},t)=k\widehat{\mathbf{e}}_1$。材料的物料函数是传输物料函数，例如热导率 $\kappa = const$（假设各向同性均匀材料）。材料的响应是流（*flux*），即单位时间经过 $\mathbf{x}$ 处面积元 $\mathbf{n}dS$ 的物理量。例如热流：$\mathbf{q}(\mathbf{x},t)$，$\mathbf{q}(\mathbf{x},t)\cdot\mathbf{n}dS$ 是单位

时间通过面积元 $\mathbf{n}dS$ 的热量。$\mathbf{q}$ 的方向与等温面垂直并指向温度减小的方向。实验表明，当温度梯度不太大时，热流满足傅立叶热传导定律[4]§1.1.3.1,p. 9：

$$\mathbf{q}(\mathbf{x},t) = -\kappa\nabla T(\mathbf{r}=\mathbf{x},t) = -k\kappa\widehat{\mathbf{e}}_1$$

这是该材料的热传导本构方程。

流变学中的本构关系主要是应力与应变（或应变速率）的关系。

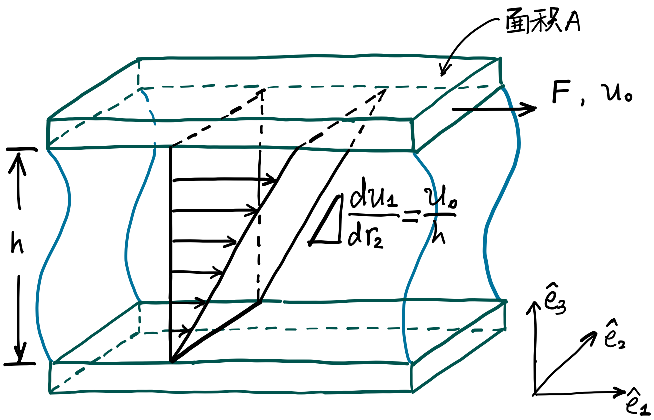

图 I.1.3: 例I.1.3

例 I.1.3 (简单剪切下的牛顿流体本构关系). 如图*I.1.3*所示的间距为 h、面积足够大的两平行板间充满粘度为 μ 牛顿流体。按照下板固定建立参考系，建立如图所示的坐标系。上板以速率 u_0 往平行于下板的方向运动。假设流体与平板接触的界面处相对流速为零；液体的流速 $\mathbf{u}(\mathbf{r},t) = u_0 r_2 \widehat{\mathbf{e}}_1/h$。若上板面积为 A、所受的拖曳力大小为 F（方向与上板运动速度相同）。F 与 u_0 的关系是

$$\tau = \frac{F}{A} = \mu\frac{u_0}{h}$$

这是一个本构关系：u_0 是外场，F 是材料响应，μ 是物料函数。

例 I.1.4 (轴向拉伸下的虎克固体本构关系). 如图*I.1.4*所示的虎克固体直圆杆，弹性模量为 E，初始截面积为 A，初始长度为 l_0。在杆的轴向受到大小为 F 的力时，杆的长度为 l，则 F 与 l 的关系是

$$\sigma = \frac{F}{A} = E\frac{l-l_0}{l_0}$$

这是一个本构关系：F 是外场，l 是材料响应，E 是物料函数。

上面两例中的本构方程形式只适用于特定形变方式。要完整描述一般形变过程的运动学和动力学，需要使用向量和张量以及它们的代数和微积分数学基础。

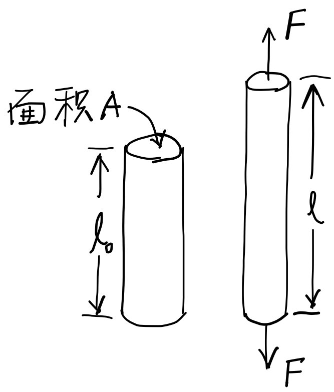

图 I.1.4: 例I.1.4

实际物体的运动和形变是在一定的初条件（initial condition）和边界条件（boundary condition）下，遵循守恒律（balance law）和本构关系的共同限定下才被具体确定的。在采用质点来简化的问题中，这些要素往往是隐含的条件。下面以一个例子说明，在解决一个具体流动问题的过程中，初条件、边界条件、守恒律和本构关系的角色。

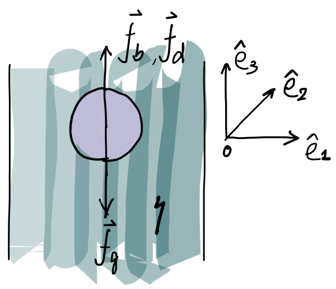

图 I.1.5: 落球粘度计原理图

例 I.1.5 (落球粘度计). 半径为 a 的球浸没在粘度为 η 的牛顿流体中。以液体的容器为参考系，建立如图*I.1.5*所示的坐标系。球的运动满足牛顿第二定律，即：$\mathbf{F} = m d\mathbf{v}/dt$（动量守恒），其中 $m = \text{const}$ 是球的质量（质量守恒）。此时，我们并不能确定球的具体运动。

设 $t = 0$ 时刻球速 $\mathbf{v} = \mathbf{0}$（初条件），液体的密度为 ρ_w、球的密度为 ρ_s，则球受到的重力是 $\mathbf{f}_\text{g} = -4\pi\rho_\text{s} g a^3 \hat{\mathbf{e}}_3/3$，浮力 $\mathbf{f}_\text{b} = 4(\rho_\text{w} - \rho_\text{s})\pi g a^3 \hat{\mathbf{e}}_3$，液体对球的拖曳力 $\mathbf{f}_\text{d} = -k\pi\eta a\mathbf{v}$（本构关系）。其中，当球面处的相对流速总为零时，$k = 6$（边界条件）。由牛顿第二定律，

$$\frac{4/3}{\pi}\rho_\text{s} g a^3 \frac{d\mathbf{v}}{dt} = \frac{4}{3}\pi g a^3 (\rho_\text{w} - \rho_\text{s}) - 6\pi\eta a\mathbf{v}$$

解方程得:

$$
\begin{aligned}
\mathbf{v}(t) &= v_1(t)\,\widehat{\mathbf{e}}_1 + v_2(t)\,\widehat{\mathbf{e}}_2 + v_3(t)\,\widehat{\mathbf{e}}_3 \\
v_1(t) &= v_2(t) \equiv 0 \\
v_3(t) &= \frac{2(\rho_{\mathrm{w}} - \rho_{\mathrm{s}})\,ga^3}{9\eta}\left(1 - e^{\frac{-9\eta t}{2\rho_{\mathrm{s}} a^2 g}}\right)
\end{aligned}
$$

当 $t \to \infty$ 时,

$$
\begin{aligned}
v_3(t) \equiv v_\infty &= \frac{2(\rho_{\mathrm{w}} - \rho_{\mathrm{s}})\,ga^2}{9\eta} \\
\eta &= \frac{2(\rho_{\mathrm{w}} - \rho_{\mathrm{s}})\,ga^2}{9v_\infty}
\end{aligned}
$$

上例中的考察对象（球）是刚体，在运动过程中形状不变。在流变学中，我们要描述物体在形变过程中如何保持质量、动量和能量守恒，这需要借助场函数的一系列积分定理，相关的知识将会在数学准备部分介绍。

I.2 聚合物流体的复杂流动现象

虎克固体与牛顿流体的本构关系不足以描述聚合物流体的流变学性质。聚合物流体是一种复杂流体（complex fluid），即具有粘弹性（viscoelasticity）和非线性（nonlinearity）的流体。

虎克固体和牛顿流体的本构关系的“简单”之处在于:

1. 应力只与应变或应变的时间导数有关
2. 应力与应变或应变的时间导数的关系是线性的*。

复杂流体的“复杂”之处分别是对上述两条理想情况的偏离:

1. 本构关系中包括应力、应力的时间导数、应变、应变的时间导数这四类函数中的三类或三类以上的函数†;
2. 物料函数依赖应变、应力或它们的时间导数；或者说应力与应变名应变的时间导数的关系是非线性的。

第一种复杂性就是粘弹性，第二种复杂性就是非线性。理论上一个复杂流体可以只满足这两种复杂性质的其中一种。但真实的复杂流体常常同时具有两种复杂性质，即非线性粘弹性（non-linear viscoelasticity）。

*线性（linearity）是指，对于一个系统，如果输入 $x_1(t)$ 和 $x_2(t)$ 的输出是 $y_1(t)$ 和 $y_2(t)$，则输入 $k_1x_1(t) + k_2x_2(t)$ 的输出就是 $k_1y_1(t) + k_2y_2(t)$ 的性质。虎克固体和牛顿流体的本构关系满足此性质。何曼君 等[5] §8.4 中所介绍的元件模型均有此性质，可自行验证。

†读者可自行验证，何曼君 等[5] §8.5 中的各种元件模型的方程均满足这一描述。

复杂流体的这两方面复杂性还可以更粗略的说成：复杂流体的流变学行为依赖外场的作用时间（粘弹性）和作用强度（非线性）。这一表述更符合我们的直观印象。

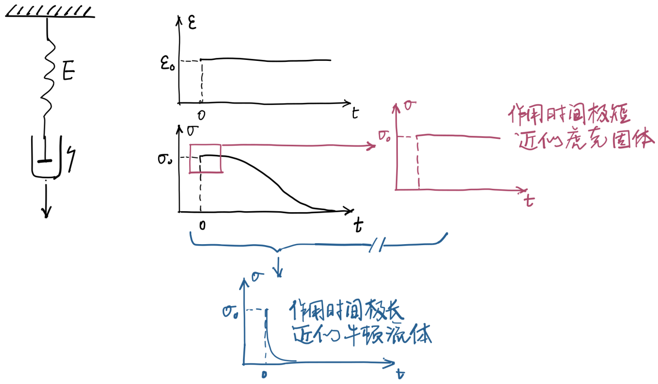

图 I.2.1: Maxwell 模型的应力松弛实验

例 I.2.1 (应力松弛实验). *Maxwell 模型的标量本构关系*[5]§8.5.1,p. 240：

$$\sigma(t) = \frac{\eta}{\tau} e^{-\frac{t}{\tau}}$$

测试结果如图I.2.1所示。当应变时间极短时，模型的响应近似虎克固体；当应变时间极长时，模型的响应近似牛顿流体。可定义 *Deborah* 数 $De \equiv \tau/t$，当 $De \to 0$ 时材料呈液体响应；当 $De \to \infty$ 时材料呈固体响应。

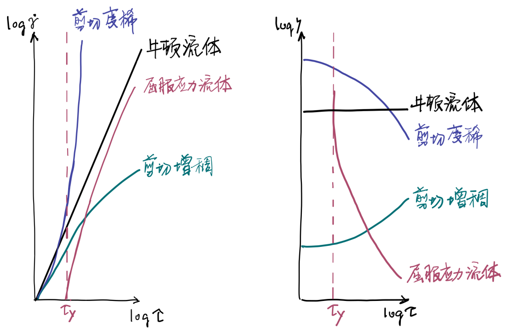

图 I.2.2: 非牛顿流体的流动曲线

例 I.2.2. 各类非牛顿流体在如例I.1.3中的简单剪切下的定常流动规律如图I.2.2所示：剪切变稀（*shear-thinning*），剪切增稠（*shear-thickening*），屈服应力流体（*yield-stress fluid*）。

非线性粘弹性——Weissenberg 效应（爬杆效应）：Weissenberg 数 $Wi \equiv \tau\dot{\gamma} = \tau\frac{\gamma}{t} = De\gamma$。

非线性粘弹性——触变性（thixotropy）：屈服应力流体的屈服流动与恢复明显依赖时间的性质。

流变学主要研究复杂流体的非线性粘弹性本构关系。

I.3 学习方法建议与参考书目

I.3.1 本课的内容

由前两节的介绍可以预见，流变学课程将会：

- 补习向量、张量和场函数微积分的数学知识（数学准备）
- 描述一般形变的运动学和动力学的方法（连续介质力学基础）
- 复杂流体的本构关系（线性粘弹性和非线性粘弹性本构方程）

与例I.1.5相同的问题，但在更一般的流变学层面上讨论的例子是 Zheng et al.[6]。可以认为，流变学课程的其中一个学习目标就是看懂这种论文。

I.3.2 学习资料

数学基础：

- Hoffman and Kunze, Linear Algebra, 2nd ed., Prentice Hall (1971)
- William, Crowell and Trotter, Calculus of Vector Functions, 2nd ed., Prentice Hall (1968)
- Itskov, Tensor Algebra and Tensor Analysis for Engineers—with Applications to Continuum Mechanics, 5th ed., Springer (2019)

连续介质力学：

- Rudnicki, Fundamentals of Continuum Mechcanics, John Wiley and Sons (2015)
- Reddy, An Introduction to Continuum Mechanics with Applications, Cambridge University Press (2008)
- Smith, An Introduction to Continuum Mechanics—After Truesdell and Noll, Springer (1993)
- 赵亚溥, 近代连续介质力学, 科学出版社 (2016)
- 网站Continuum Mechanics
- 网课视频Continuum Mechanics, İlker Temizer, Bilkent University
- 网课视频Continuum Mechanics, website, Xavier Oliver, Technical University of Catalonia
- 网课视频Lectures on Continuum Physics, website, Krishna Garikipati, University of Michigan

流变学：

- Macosko, Rheology: Principles, Measurements, and Applications, Wiley (1994)
- Bird, Stewart and Lightfoot, Transport Phenomena, 2nd ed., John Wiley and Sons (2002)

- 网课视频Rheology of Complex Materials, Abhijit Deshpande, Indian Institute of Technology Madras

I.3.3 学习方法建议

本课的课堂学习目标仅为:

- 帮助材料类本科专业背景的同学克服学习流变学的思想障碍和实际困难
- 带领入门，为将来更加正式、深入的自学打下基础
- 介绍基本假设、基本概念、基本结论，对独立推算和解题能力不作要求

学习建议:

- 多本同类资料比较学习，相互印证，积极寻找更多学习资料
- 建群互助。QQ 群：5069284
- 了解学科历史、人物传记、思想文化

第二部分 数学准备

II.1 集合与映射

集合是近代数学的基本语言。连续介质力学中的大量概念都依赖集合来定义。而集合本身却是难以借助更基础的概念进行定义的“最原始概念”。因此在下述集合的定义只能假定每一位读者都能一致地理解。

定义 II.1.1 (集合)**.** *集合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或成员。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。*

集合的表示方式有列举法和描述法。例：以下都是集合。

$$A = \{2, 4, 6, 8\}$$
$$B = \{n | n\text{是偶数且}1 < n < 9\}$$
$$1 \notin A, 2 \in B$$

我们能否说上例的两个集合是“相同”的或“相等”的？下一条定义将明确这一问题的答案。

定义 II.1.2 (子集、包含、集合的相等)**.** *如果只要 $a \notin B$，则必有 $a \in A$，则称 B 是 A 的子集，或称 A 包含 B，记为 $B \subseteq A$，$A \supseteq B$。如果 $A \subseteq B$ 且 $B \subseteq A$ 则称 $A = B$。*

可验证如下定理：

定理 II.1.1. *1. $S \subseteq S, \forall S$，*

2. 若 $A \subseteq B$ 且 $B \subseteq C$ 则 $A \subseteq C$。

定义 II.1.3 (真子集、真包含)**.** *若 $B \subseteq A$ 且 $B \neq A$，则称 B 是 A 的真子集，或称 A 真包含 B，记为 $B \subset A$，$A \supset B$。*

可验证如下定理：

定理 II.1.2. *1. $\forall S$，$S \subset S$ 均不成立。*

2. 若 $A \subset B$ 则 $B \subset A$ 不成立。
3. 若 $A \subset B$ 且 $B \subset C$，则 $A \subset C$。

定义 II.1.4 (空集)**.** 空集 $\emptyset = \{a|a \neq a\}$

可验证如下定理：

定理 II.1.3. 1. $\emptyset \subseteq A, \forall A$。
2. $\emptyset \subset A, \forall A \neq \emptyset$。

定义 II.1.5 (并集)**.** 若 $a \in A$ 或 $a \in B, \forall a, A, B$，则 $a \in A \cup B$，称 A 与 B 的并集。

可验证：

定理 II.1.4. 并集关系具有以下性质：
1. 结合律：$A \cup (B \cup C) = (A \cup B) \cup C$。
2. 交换律：$A \cup B = B \cup A$。
3. $\emptyset \cup A = A, \forall A$。

定义 II.1.6 (交集)**.** 若 $a \in A$ 且 $a \in B, \forall a, A, B$，则 $a \in A \cap B$，称 A 与 B 的交集。

定义 II.1.7 (笛卡尔积)**.** 两个集合 A 和 B 的笛卡尔积是所有有序对 $(a, b), a \in A, b \in B$ 的集合：$A \times B = \{(a, b) | a \in A \text{且} b \in B\}$

我们可以推广这一定义至任意个集合的笛卡尔积。集合 $A_1, A_2, \cdots, A_n$ 的笛卡尔积是 $A_1 \times A_2 \times \cdots \times A_n = \{(a_1, a_2, \cdots, a_n) | a_i \in A_i\}$，一个有序 n 元组的集合。如果 $A_1 = A_1 = \cdots = A_n = A$，则记 $A_1 \times A_2 \times \cdots \times A_n = A^n$。例如，$\mathbb{R}$ 是实数集，$\mathbb{R}^n$ 是 n 元有序实数组的集合，在解析几何中常用以表示 n 维欧几里德空间中一个点的坐标。

一个集合的元素可以与另一个集合的元素建立对应关系。我们主要关心的是满足某种规定的对应关系，称为映射，定义如下。

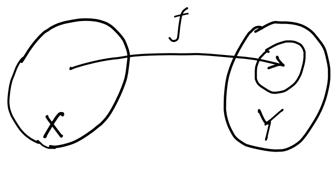

图 II.1.1: 映射

定义 II.1.8 (映射). （如图II.1.1所示）从集合 X 到集合 Y 的映射 f，记为 $f: X \to Y$，是 X 的所有元素与 Y 的部分或所有元素之间的对应关系，且每个 X 的元素只对应 Y 的一个元素。如果 $y \in Y$ 是 $x \in X$ 通过映射 f 的对应，则可写成：

$$y = f(x) \text{或} x \mapsto f(x) \text{或} f: x \mapsto y$$

并称 $f(x)$ 是 x 的像。按上述定义，x 只有一个像。X 称为该映射 f 的定义域，记为 $\mathrm{dom} f$，Y 称为该映射 f 的陪域（*codomain*）或到达域（*target domain*）。x 的像 $f(x)$ 是 Y 的子集，称为该映射 f 的值域（*range*），记为 $\mathrm{ran} f$。

当映射的陪域是实数集 $\mathbb{R}^n$ 或复数集 $\mathbb{C}^n$ 时，我们常常也称其为函数。

定义 II.1.9 (映射的相等). 若映射 $f: X \to Y$ 和 $g: X \to Y$ 满足 $f(x) = g(x), \forall x \in X$，则这两个映射相等。

定义 II.1.10 (恒等映射). 恒等映射 $\mathrm{id}_A: A \to A$ 是由集合 A 到其自身的映射：$\mathrm{id}_A(a) = a, \forall a \in A$。

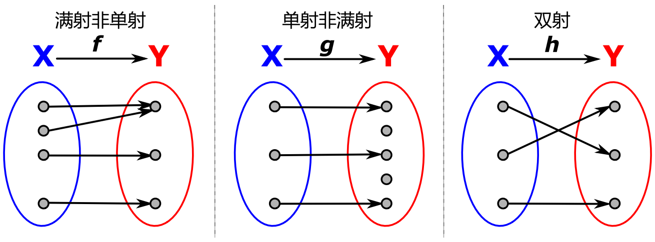

图 II.1.2: 满射、单射和双射

定义 II.1.11 (单射、双射、满射). （如图II.1.2所示）对于映射 $f: X \to Y$，若 $f(x_1) = f(x_2) \Rightarrow x_1 = x_2$，则映射 f 是单射（*injective mapping*）。若 $f(X) = Y$，则称 f 是满射（*surjective mapping*）。既是单射又是满射的映射叫双射（*bijective mapping*）。

读者可自行画图表示非满射非单射的情况。

定义 II.1.12 (复合映射). 令 $f: X \to Y$ 和 $g: Y \to Z$ 是两个映射。则 f 和 g 的复合映射，记为 $g \circ f$，是从 X 到 Z 的映射：

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

定理 II.1.5. 对于映射 $f: X \to Y$ 和 $g: Y \to Z$，

- 如果 f 和 g 都是满射，则 $g \circ f$ 是满射

- 如果 $g \circ f$ 是满射，则 g 是满射

证明. 留作练习。 □

定理 II.1.6. 对于映射 $f: X \to Y$ 和 $g: Y \to Z$，

- 如果 f 和 g 都是单射，则 $g \circ f$ 是单射
- 如果 $g \circ f$ 是满射，则 f 是单射

证明. 留作练习。 □

定义 II.1.13 (逆映射)**.** 对于映射 $f: X \to Y$，如果存在一个映射 $g: Y \to X$，使得复合映射 $g \circ f$ 是恒等映射 id_X，则称映射 f 是可逆的，映射 g 是 f 的逆映射。

关于逆映射，有一条重要的定理，使得在今后的数学陈述和推理中，我们可以默认——

定理 II.1.7. 双射必存在唯一逆映射。双射的逆映射也是双射。

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

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

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

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

则有如下几条结论：

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

该引理证毕。

现在正式证定理II.1.7。从上面定义的这个 g 继续，如果 g 是双射，则 g 不仅是满射，还是单射。由刚刚证完的引理，可用类似方法给 g 找一个逆映射 $f': X \to Y$。而且，由于 $\text{rang} \equiv X$，我们无需像定义 g 那样为 f' 分出 $x \notin \text{rang}, x \in X$ 的情况，因为不存在这种情况。故

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

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

II.2 向量空间

为一个的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[7]）。

定义 II.2.1 (向量空间). *数域 $\mathbb{F}$ 上的向量空间 $\mathcal{V}$ 是向量的集合，且满足：*

1. *加法运算。$\forall \mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathcal{V}$:*
   (a) *封闭性：$\mathbf{a} + \mathbf{b} \in \mathcal{V}$*
   (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：$\exists \mathbf{0} \in \mathcal{V} : \mathbf{a} + \mathbf{0} = \mathbf{a}$*
   (e) *逆：$\forall a \in \mathcal{V}, \exists -\mathbf{a} : \mathbf{a} + -\mathbf{a} = \mathbf{0}$*
2. *标量乘法运算。$\forall \alpha, \beta \in \mathbb{F}, \mathbf{a}, \mathbf{b} \in \mathcal{V}$:*
   (a) *$\alpha \mathbf{a} \in \mathcal{V}$*
   (b) *$\alpha (\beta \mathbf{a}) = (\alpha\beta) \mathbf{a}$*
   (c) *$1\mathbf{a} = \mathbf{a}$*
3. *标量乘法满足分配律：*
   (a) *$\alpha (\mathbf{a} + \mathbf{b}) = \alpha \mathbf{a} + \alpha \mathbf{b}$*
   (b) *$(\alpha + \beta) \mathbf{a} = \alpha \mathbf{a} + \beta \mathbf{a}$*

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。

例 II.2.1. *请根据定义II.2.1验证*

- $\mathbb{R}$ 是 $\mathbb{R}$ 上的向量空间。
- $\mathbb{C}$ 是 $\mathbb{R}$ 上的向量空间。
- $\mathbb{C}$ 是 $\mathbb{C}$ 上的向量空间。
- $\mathbb{R}$ 不是 $\mathbb{C}$ 上的向量空间。
- 函数 $f:(a,b)\subset\mathbb{R}\to\mathbb{C}$ 的集合是数域 $\mathbb{C}$ 上的向量空间。
- $\mathbb{R}^n$ 是所有实数 n 元组 $\{a\}=(\alpha_1,\cdots,\alpha_n),\alpha_i\in\mathbb{R},i=1,\cdots,n$ 的集合。若 $\forall\{a\},\{b\}\in\mathbb{R}^n$:
  1. $\{a\}+\{b\}=(\alpha_1+\beta_1,\cdots,\alpha_n+\beta_n)$;
  2. $\alpha\{a\}=(\alpha\alpha_1,\cdots,\alpha\alpha_n),\forall\alpha\in\mathbb{R}$;
  3. $\{0\}=(0,\cdots,0)$;
  4. $-\{a\}=(-\alpha_1,\cdots-\alpha_n)$;

  则 $\mathbb{R}^n$ 连同上述的运算规定形成 $\mathbb{R}$ 上的向量空间，又称为实坐标空间。
- 数域 $\mathbb{F}$ 上的所有 $m\times n$ 矩阵的集合 $\mathcal{M}^{m\times n}$（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。

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。首先，由于“封闭性”的要求，一个向量空间内的任一向量总能被这一向量空间中的其他向量按所规定的运算法则表达出来。具体地，若 $\mathcal{V}$ 是数域 $\mathbb{F}$ 上的向量空间，$\mathbf{a}_1,\cdots,\mathbf{a}_n\in\mathcal{V}$，$\alpha_1,\cdots,\alpha_n\in\mathbb{F}$，则依据线性空间的封闭性，$\sum_{i=1}^n\alpha_i\mathbf{a}_i$ 也属于 $\mathcal{V}$，即也是一个向量，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 $\mathcal{V}$ 是 $\mathbb{F}$ 上的向量空间，$\mathbf{a}_1,\cdots,\mathbf{a}_n\in\mathcal{V}$。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。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。反之，若存在某不全为零的 $\alpha_i,i=1,\cdots,n$ 使得 $\sum_{i=1}^n\alpha_i\mathbf{a}_i=\mathbf{0}$ 则称向量 $\{\mathbf{a}_i\}$ 线性相关。

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[7] 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[7]§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 $1\neq 0$ 使得 $1\mathbf{0}=\mathbf{0}$。
4. 一个向量组 S 是线性无关向量组当且仅当 S 的所有子集都是线性无关向量组。

可见，封闭性是一个好性质。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

*见[7]§2.1 矩阵与矩阵的运算

†集合 $\{\mathbf{a}_1,\cdots,\mathbf{a}_n\}\subset\mathcal{V}$ 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 $(\mathbf{a}_i)$。

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 $\mathcal{V}$ 是数域 $\mathbb{F}$ 上的一个向量空间，如果 $\mathcal{V}$ 的子集 $\mathcal{W} \subseteq \mathcal{V}$ 也是一个向量空间（并与 $\mathcal{V}$ 采用相同的加法和标量乘定义），则称 $\mathcal{W}$ 是 $\mathcal{V}$ 的一个子空间（*subspace*）。

接这一定义，易证 $\mathcal{W}$ 的任意两个向量 $\mathbf{a}, \mathbf{b} \in \mathcal{W}$ 和任一标量 $\alpha \in \mathbb{F}$ 构成的线性组合 $\alpha\mathbf{a}+\mathbf{b}$ 也在 $\mathcal{W}$ 内[7]§7.1 定理 1.1,p.169。

就算向量空间 $\mathcal{V}$ 的某子集 S 因不满足封闭性而成为不了 $\mathcal{V}$ 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 $\mathcal{V}$ 的子空间。

定义 II.2.4 (线性生成空间). 若 S 是向量空间 $\mathcal{V}$ 的非空子集，即 $S \subseteq \mathcal{V}, S \neq \emptyset$，那么 S 内的向量的所有线性组合的集合 $\mathcal{W}_S$ 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，$\mathcal{W}_S$ 中的向量都能由 S 的向量线性表出。一个直接的结论就是 $S \subseteq \mathcal{W}_S$，因为一个向量总能被它自己线性表出。

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 $Q = \mathrm{span}\{(2,1,3), (1,0,1)\}$），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 $\mathcal{V}$ 是其子空间 $\mathcal{B}$ 的线性生成空间（即 $\mathcal{V} = \mathrm{span}\mathcal{B}$，且 $\mathcal{B}$ 内的所有向量线性无关，则称 $\mathcal{B}$ 是 $\mathcal{V}$ 的一组基（*basis*）。如果 $\mathcal{B}$ 含有有限个向量，则称 $\mathcal{V}$ 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，$\mathcal{V}$ 内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，我们才能通过把这一向量个数定义为 $\mathcal{V}$ 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[7] "(3) 的证明", p. 171 [8] §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. *设 $\mathcal{V}$ 是数域 $\mathbb{F}$ 上的有限维向量空间，它的维数，记为 $\dim\mathcal{V}$，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. *令 $\mathcal{V}$ 是一个有限维向量空间，其维数 $n = \dim\mathcal{V}$，则*

1. *$\mathcal{V}$ 的任一含有多于 n 个向量的子集都是线性相关向量组[7] "(3) 的证明", p. 171。*
2. *$\mathcal{V}$ 的任一向量个数少于 n 的子集都不能线性生成整个 $\mathcal{V}$（即这样的子集的线性生成空间总是 $\mathcal{V}$ 的真子集）。*
3. *$\mathcal{V}$ 的任一子空间 $\mathcal{W}$ 的维数不大于 $\mathcal{V}$ 的维数，即 $\dim\mathcal{W} \le \dim\mathcal{V}$；当且仅当 $\mathcal{W} = \mathcal{V}$ 时取等号。*

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 和 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。*

现在我们已经明确了向量空间的维数的概念。那么一个向量空间 $\mathcal{V}$ 的子空间 $\mathcal{W}$ 本身也是一个向量空间，也有它的维数。$\dim\mathcal{V}$ 与 $\dim\mathcal{W}$ 的关系是怎样的呢？回顾例II.2.2中，三维实坐标空间的子空间是一个平面。这是否暗示，$\dim\mathcal{W} \le \dim\mathcal{V}$ 呢？答案是肯定的，且当且仅当 $\mathcal{W} = \mathcal{V}$ 时取等号，证明此略，在此提到这件事只为了让读者放心地按照例II.2.2所暗示那般，一般性地理解向量空间及其子空间的维数关系。

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。"[7] p. 167 不过，

通过引入“坐标”的概人记，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 $\mathcal{V}$ 是一个 n 维向量空间，$\mathcal{V}$ 内的一组线性无关的有序向量序列 $(\mathbf{a}_i)$ 线性生成整个 $\mathcal{V}$，则称这组有序向量序列 $\mathcal{B}$ 为 $\mathcal{V}$ 的一组有序基（ordered basis）。由定义II.2.5，任一向量 $\mathbf{x} \in \mathcal{V}$ 均可表达为 $\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{a}_i$，进而，任一 $\mathbf{x} \in \mathcal{V}$ 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 $\mathcal{B}$ 下的坐标（coordinate）或分量（component）。*

基的原始定义仅要求基是一个集合，没有有序性的规定。用一组基表出一个 n 维向量所使用的 n 个标量也只是一个集合。在给定一组基下，向量与这个标量集合也是唯一对应的。只不过，我们希望所定义的“坐标”是一个有序序列，所以定义II.2.7才特别增加了“有序”的要求。

有了坐标的定义，在给定基下，n 维向量空间 $\mathcal{V}$ 中的一个向量就与 $\mathbb{F}^n$ 中的一个有序 n 元数组形成了一一对应的关系。由封闭性，没有一个向量不对应一个数组，反之亦然。

易验，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 $\mathbb{F}^n$ 某组基 $\mathcal{B} = \{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ 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 $\mathbf{x}$ 在基 $\mathcal{B}$ 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，无论是向量 $\mathbf{x}$ 本身还是其中某基的下的坐标，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“正交基”、“单位向量”的概念得以引入，同时介绍其他重要概念。

*这里的唯一性可再次参考[7] “(3) 的证明”，p. 171。

† “利用基和坐标可把线性空间的运算变得更具体。”[7]p.173。

‡这里的概念区分可参见[7] 例题 2.1,p. 173。

定义 II.3.1 (内积). 数域 $\mathbb{F}$ 上的向量空间 $\mathcal{V}$ 中两向量 $\mathbf{a},\mathbf{b}\in\mathcal{V}$ 的内积（*inner product*）记为 $(\mathbf{a}|\mathbf{b}),(\cdot|\cdot):\mathcal{V}\times\mathcal{V}\to\mathbb{F}$，满足*：$\forall\mathbf{a},\mathbf{b},\mathbf{c}\in\mathcal{V},\alpha\in\mathbb{F}$

- $(\mathbf{a}|\mathbf{b})=\overline{(\mathbf{b}|\mathbf{a})}$
- $(\alpha\mathbf{a}|\mathbf{b})=\alpha(\mathbf{a}|\mathbf{b})$
- $(\mathbf{a}+\mathbf{b}|\mathbf{c})=(\mathbf{a}|\mathbf{b})+(\mathbf{a}|\mathbf{c})$
- $(\mathbf{a}|\mathbf{a})\in\mathbb{R}$ 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*）。

由内积运算可推出以下性质：$\forall\mathbf{a},\mathbf{b},\mathbf{c}\in\mathcal{V},\alpha\in\mathbb{F}$

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

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，在数学上有时会造成重大的后果，但在流变学中并不重要。我们将会在线性泛函与对偶空间的时候再讲这件事，不感兴趣的同学可以忽略。

例 II.3.1. 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*）$\mathbf{a}\cdot\mathbf{b}$。

由内积的一般定义可证明柯西–施瓦茨不等式（Cauchy–Schwarz inequality），故该不等式对任一满足内积定义的一般要求的具体定义都成立。

定理 II.3.1. 设 $\mathcal{V}$ 是数域 $\mathbb{F}$ 上的一个内积空间，则有 $\forall\mathbf{a},\mathbf{b}\in\mathcal{V}$，

1. 柯西–施瓦茨不等式 $|(\mathbf{a}|\mathbf{b})|\leq(\mathbf{a}|\mathbf{a})(\mathbf{b}|\mathbf{b})$
2. $(\mathbf{a}|\mathbf{b})=\operatorname{Re}(\mathbf{a}|\mathbf{b})+\operatorname{Re}(\mathbf{a}|i\mathbf{b})$

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

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

*上划线表示复数共轭

则可验证 $(\mathbf{c}|\mathbf{a})=0$，且

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

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（norm）的规定，得到赋范向量空间(normed vector space)。

定义 II.3.2 (向量的范)**.** *设 $\mathcal{V}$ 是数域 $\mathbb{F}$ 向量空间，则有 $\forall\mathbf{x}\in\mathcal{V},\alpha\in\mathbb{F}$,*

1. *非负性：$\|\mathbf{x}\|\ge0$*
2. *调和性：$\|\alpha\mathbf{x}\|=|\alpha|\,\|\mathbf{x}\|$*
3. *三角不等式：$\|\mathbf{x}+\mathbf{y}\|\le\|\mathbf{x}\|+\|\mathbf{y}\|$*

带有范的定义的向量空间叫赋范向量空间。

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，而无法法作为定理由之前两个规定证明出来。这说明我们主观上就希望向量的范满足这样的性质。这是因为，“范”是我们给于向量以“长度”的概念，并希望它能于欧几里德几何公设规定的性质相兼容。

一个赋范向量空间 $\mathcal{V}$ 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：$\|\hat{\mathbf{a}}\|=1,\hat{\mathbf{a}}\in\mathcal{V}$。赋范向量空间 $\mathcal{V}$ 的任一向量 § 均可通过 $\hat{\mathbf{x}}=\mathbf{x}/\|\mathbf{x}\|$ 归一化为一个单位向量（由范的定义易验)。

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。只要满足相应定义中的一般要求的任一种定义，都可作为内积或范。向量的范的其中一种常用的定义是：设 $\mathcal{V}$ 是内积空间，向量 $\mathbf{a}\in\mathcal{V}$ 的范 $\|\mathbf{a}\|\equiv(\mathbf{a}|\mathbf{a})^{\frac{1}{2}}$。我们把这一定义称为欧几里德范。其他范的定义则为非欧几里德范，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。

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。

定理 II.3.2. *一个赋范向量空间是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 $\mathcal{V}$ 是数域 $\mathbb{F}$ 上的一个赋范向量空间，若定义内积：

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

可验证上式满足内积定义，且 $\|\mathbf{a}\|^2=(\mathbf{a}|\mathbf{a})^{\frac{1}{2}}$。 □

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 $\mathbf{a},\mathbf{b}\in\mathbb{R}^n$ 则有 $\mathbf{a}\cdot\mathbf{b}=\frac{1}{4}\|\mathbf{a}+\mathbf{b}\|^2-\frac{1}{4}\|\mathbf{a}-\mathbf{b}\|^2$。

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把两种空间统一起来。特别地，欧几里德范总满足柯西–施瓦茨不等式、三角不等式和极化恒等式。

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 $\mathcal{V}$ 中的向量 $\mathbf{a},\mathbf{b}\in\mathcal{V}$，若 $(\mathbf{a}|\mathbf{b})=0$，则称 $\mathbf{a}$ 与 $\mathbf{b}$ 是正交的。若 S 是 $\mathcal{V}$ 的一个子集，且 S 中的向量两两正交，则称 S 是 $\mathcal{V}$ 的一个正交集（orthogonal set）。若 $\mathcal{V}$ 还是一个赋范向量空间，且 $\mathcal{V}$ 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）。*

例 II.3.2. [7] *“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 $(A|B)\equiv\sum_{j,k}A_{jk}\overline{B}_{jk},\forall A,B\in\mathbb{F}^{n\times n}$，$\overline{B}$ 表示矩阵 B 的每个分量均取复数共轭后的矩阵。*

以下定理证明任意一组两两正交的向量线性无关，但需注意的是线性无关向量组却未必两两正交。

定理 II.3.3. *正交集中的所有非零向量线性无关。*

证明. 设 $\mathcal{V}$ 是内积空间，S 是 $\mathcal{V}$ 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 $\mathcal{V}$ 的正交集 S 总能线性生成 $\mathcal{V}$ 的一个子空间。若内积空间 $\mathcal{V}$ 的一组基 $\mathcal{B}$ 是正交集，则称 $\mathcal{B}$ 为 $\mathcal{V}$ 的正交基（orthogonal basis）。如果 $\mathcal{V}$ 是赋范内积空间，其一组基 $\mathcal{B}$ 是规范正交集，则称 $\mathcal{B}$ 是 $\mathcal{V}$ 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 $\mathcal{V}$ 是内积空间，$\mathbf{b}_1, \cdots, \mathbf{b}_n \in \mathcal{V}$ 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 $\mathbf{a}_1, \cdots, \mathbf{a}_n \in \mathcal{V}$ 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 ($\mathbb{R}^n$ 内积空间). *n 元有序实数对的集合 $\mathbb{R}^n$ 中可定义如下加法运算：*

$$(x_1, \cdots, x_n)^\intercal + (y_1, \cdots, y_n)^\intercal = (x_1 + y_1, \cdots, x_n + y_n)^\intercal,$$
$$\alpha\,(x_1, \cdots, x_n)^\intercal = (\alpha x_1, \cdots, \alpha x_n)^\intercal, \forall x_i, y_i, \alpha \in \mathbb{R}, i = 1, \cdots, n$$

可验证，带有上述运算规则的 $\mathbb{R}^n$ 是一个向量空间。再引入如下内积运算：

$$(x_1, \cdots, x_n)^\intercal \cdot (y_1, \cdots, y_n)^\intercal = \sum_{j=1}^{n} x_j y_j$$

可验证，带有上述运算规则满足内积定义，$\mathbb{R}^n$ 成变一个内积空间。

例 II.3.4. *考虑 $\mathbb{R}^3$ 中的三个向量 $\mathbf{b}_1 = (3, 0, 4)^\intercal, \mathbf{b}_2 = (-1, 0, 7)^\intercal, \mathbf{b}_3 = (2, 9, 11)^\intercal$。首先可以验证它们线性无关，即*

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

只有唯一解 $x_1 = x_2 = x_3 = 0$。

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

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

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

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

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

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

其中

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

叫克劳内克符号（*Kronecker symbol*）。

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

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

标准基是一个规范正交基。

在选定任一规范正交基 $\{\hat{\mathbf{a}}\}$ 下，两向量 $\mathbf{x}$ 和 $\mathbf{y} \in \mathbb{R}^3$ 的标准内积计算，按定义有 $\mathbf{x} \cdot \mathbf{y} =$

$\sum_i x_i y_i$。另一方面，表达成给定基下的坐标后，则有

$$\begin{aligned}
\mathbf{x} &= \sum_i \mu_i \mathbf{a}_i, \mathbf{y} = \sum_i \nu_i \mathbf{a}_i \\
\mathbf{x}\cdot\mathbf{y} &= \sum_i \mu_i \mathbf{a}_i \cdot \left(\sum_j \nu_j \mathbf{a}_j\right) \\
&= \sum_i \sum_j \left(\mu_i \mathbf{a}_i\right)\cdot\left(\nu_j \mathbf{a}_j\right) \\
&= \sum_i \sum_j \mu_i \nu_j \left(\mathbf{a}_i \cdot \mathbf{a}_j\right) \\
&= \sum_i \sum_j \mu_i \nu_j \delta_{ij} \\
&= \sum_i \mu_i \nu_i
\end{aligned}$$

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映射——线性变换

例 II.4.1.

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

一般地，建立了映射的两个向量空间是在同一数域上的，但向量空间之间的映射不一定保留其运算规则，即 $f(\alpha\mathbf{a}+\beta\mathbf{b})\neq\alpha f(\mathbf{a})+\beta f(\mathbf{b})$*。

定义 II.4.1 (线性变换)**.** 设 $\mathcal{V}$ 和 $\mathcal{W}$ 是 $\mathbb{F}$ 上的向量空间。如果从 $\mathcal{V}$ 到 $\mathcal{W}$ 的映射 $T:\mathcal{V}\to\mathcal{W}$ 满足

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

则称 T 是从 $\mathcal{V}$ 到 $\mathcal{W}$ 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 $\mathbf{T}:\mathcal{V}\to\mathcal{W}$ 和 $\mathbf{U}:\mathcal{V}\to\mathcal{W}$ 满足 $\mathbf{Ta}=\mathbf{Ua}\forall\mathbf{a}\in\mathcal{V}$ 则称这两个线性变换相

*在抽象代数中，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（isomorphism）。

等，$\mathbf{U}=\mathbf{T}$。零变换 $\mathbf{T}_0:\mathcal{V}\to\mathcal{W}$ 定义为 $\mathbf{T}_0\mathbf{a}=\mathbf{0}_{\mathcal{W}}\forall\mathbf{a}\in\mathcal{V}$，其中 $\mathbf{0}_{\mathcal{W}}\in\mathcal{W}$ 表示 $\mathcal{W}$ 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 $\mathcal{V}$ 到其自身的线性变换 $\mathbf{I}_{\mathcal{V}}:\mathcal{V}\to\mathcal{V},\mathbf{I}_{\mathcal{V}}\mathbf{a}=\mathbf{a}\forall\mathbf{a}\in\mathcal{V}$ 又可称为恒等线性变换。

定理 II.4.1. 设 $\mathcal{V}$、$\mathcal{W}$ 是 $\mathbb{F}$ 上的向量空间，$\mathbf{T}$、$\mathbf{U}$ 是从 $\mathcal{V}$ 到 $\mathcal{W}$ 的线性变换。若定义

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

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

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

证明. 分别用 $\mathbf{T}+\mathbf{U}$ 和 $\alpha\mathbf{T}$ 作用于向量 $\beta\mathbf{a}+\gamma\mathbf{b},\alpha,\beta\in\mathbb{F},\mathbf{a},\mathbf{b}\in\mathcal{V}$，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 $\mathcal{V},\mathcal{W}$，所有由 $\mathcal{V}$ 到 $\mathcal{W}$ 的线性变换的集合组成一个向量空间，记为 $\mathcal{L}(\mathcal{V},\mathcal{W})$，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 $\mathcal{V}$ 到 $\mathcal{W}$ 的零变换，对任一 $\mathbf{T}\in\mathcal{L}(\mathcal{V},\mathcal{W})$ 和 $\mathbf{a}\in\mathcal{V}$，$(\mathbf{T}_0+\mathbf{T})\,\mathbf{a}=\mathbf{T}_0\mathbf{a}+\mathbf{Ta}=\mathbf{0}_{\mathcal{W}}+\mathbf{Ta}=\mathbf{Ta}$，即 $\mathbf{T}_0+\mathbf{T}=\mathbf{T}\forall\mathbf{T}\in\mathcal{L}(\mathcal{V},\mathcal{W})$。 □

线性变换的空间 $\mathcal{L}(\mathcal{V},\mathcal{W})$ 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 $\mathcal{L}(\mathcal{V},\mathcal{W})$ 中，括号表示有序对。我们在这里考察 $\mathcal{L}(\mathcal{V},\mathcal{W})$ 的基和维数。

引理 II.4.1. 设 $\mathcal{V}_N$ 是数域 $\mathbb{F}$ 上的 N 维向量空间，$B_{\mathcal{V}}=\{\mathbf{e}_i\}_{i=1}^N$ 是 $\mathcal{V}_N$ 的一组基，$\mathcal{W}$ 是同数域上的另一向量空间。对 $\mathcal{W}$ 中给定的任意一组 N 个不同向量 $\{\mathbf{b}_i\}_{i=1}^N\in\mathcal{W}$，有且只有一个线性变换 $\mathbf{T}:\mathcal{V}_N\to\mathcal{W}$ 满足 $\mathbf{Te}_i=\mathbf{b}_i,i=1,\cdots,N$。

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

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

*此处要注意区分不同向量空间中的零向量。

†由上一脚注我们可以称线性变换是向量空间之间的同态映射。

定义映射 $\mathbf{T}:\mathcal{V}_N\to\mathcal{W}$ $\mathbf{Ta}\equiv\sum_{i=1}^{N}\alpha_i\mathbf{b}_i$，则有 $\mathbf{Te}_i=\mathbf{b}_i, i=1,\cdots,N$。我们还需验证 $\mathbf{T}$ 是否线性变换，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。此略。

唯一性的证明，设另有一线性变换 $\mathbf{U}:\mathcal{V}_N\to\mathcal{W}$ 也满足 $\mathbf{Ue}_i=\mathbf{b}_i$，然后只需显示 $\mathbf{Ua}=\mathbf{Ta}$ 即 $\mathbf{U}=\mathbf{T}$ 即可。此略。 □

定理 II.4.2. *若 $\mathcal{V}_N,\mathcal{W}_M$ 分别为数域 $\mathbb{F}$ 上的 N,M 维向量空间，则 $\mathcal{L}(\mathcal{V}_N,\mathcal{V}_M)$ 的维数是 $N\times M$。*

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

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

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

则对于每对 (p,q)，$\mathbf{E}^{pq}\in\mathcal{L}(\mathcal{V}_N,\mathcal{W}_M)$ 均唯一存在。给定一线性变换 $\mathbf{T}\in\mathcal{L}(\mathcal{V}_N,\mathcal{W}_M)$，且令

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

由引理II.4.1，对 $\{\mathbf{f}_j\}_{j=1}^{M}$，$\mathbf{T}$ 是维一的。且上式中的 A_{qi} 就是向量 $\mathbf{Te}_i$ 的坐标。

定义 $\mathbf{U}\in\mathcal{L}(\mathcal{V}_N,\mathcal{W}_M)$，$\mathbf{Ue}_i=\sum_p\sum_q A_{qp}\mathbf{E}^{pq}\mathbf{e}_i$，则

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

则可见 $\mathbf{U}=\mathbf{T}$，同时有

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

即 $\mathbf{T}$ 可由 $\{\mathbf{E}^{pq}\}$ 线性表出，换句话说 $\mathcal{L}(\mathcal{V}_N,\mathcal{W}_M)$ 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 $\mathcal{L}(\mathcal{V}_N,\mathcal{W}_M)$ 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，

$$\mathbf{Te}_i=\sum_q A_{qi}\mathbf{f}_q=\mathbf{0}_\mathcal{W}\Leftrightarrow A_{qi}=0\forall q=1,\cdots,M, i=1,\cdots,N$$

*这一线性变换与例II.3.2中的矩阵很类似，可以比较理解。

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

注意，给定两个有限维向量空间 $\mathcal{V}, \mathcal{W}$，$\mathcal{L}(\mathcal{V}, \mathcal{W})$ 中的线性变换和 $\mathcal{L}(\mathcal{W}, \mathcal{V})$ 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 $N \times M$ 和 $M \times N$ 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又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). 设 $\mathcal{V}$ 和 $\mathcal{W}$ 是数域 $\mathbb{F}$ 上的向量空间。线性变换 $\mathbf{T}: \mathcal{V} \to \mathcal{W}$ 的零空间 (*null space*) 或核空间 (*kernel*)，记作 $\mathrm{ker}\mathbf{T}$，是所有满足 $\mathbf{T}\mathbf{a} = \mathbf{0}_\mathcal{W}$ 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 $\mathrm{nullity}\mathbf{T} \equiv \dim(\mathrm{ker}\mathbf{T})$。如果 $\mathcal{V}$ 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 $\mathrm{rank}\mathbf{T} \equiv \dim(\mathrm{ran}\mathbf{T})$。特别地，如果 $\mathrm{nullity}\mathbf{T} = 0$，即 $\mathbf{T}$ 的零空间只有零向量 $\mathbf{0}_\mathcal{V}$ 一个元素，则称 $\mathcal{T}$ 是非奇异的 (*non-singular*)。

上述定义默认了两个易验事实的成立: $\mathrm{ker}\mathbf{T}$ 是 $\mathcal{V}$ 的子空间, $\mathrm{ran}\mathbf{T}$ 是 $\mathcal{W}$ 的子空间 (否则不能直接谈它这两个子集的“维数”), 证明从略[7]§7.3 “2. 线性变换的简单性质 (4)” p. 177。基于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]§7.3 “2. 线性变换的简单性质 (5)” p. 178。

定理 II.4.3. 设 $\mathcal{V}$ 和 $\mathcal{W}$ 是数域 $\mathbb{F}$ 上的向量空间，其中 $\mathcal{V}$ 是有限维向量空间，则对线性变换 $\mathbf{T}: \mathcal{V} \to \mathcal{W}$ 有

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

证明. 设 $\dim\mathcal{V} = n$，$\mathcal{N}$ 是 $\mathbf{T}$ 的零空间，$\mathrm{nullity}\mathbf{T} = \dim\mathcal{N} = k$，$\{\mathbf{a}_1, \cdots, \mathbf{a}_k\}$ 是 $\mathcal{N}$ 的一组基。则可在 $\mathcal{V}$ 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 $\mathcal{V}$ 的基。

由于 $\{\mathbf{T}\mathbf{a}_1, \cdots, \mathbf{T}\mathbf{a}_n\}$ 线性生成 $\mathbf{T}$ 的值域 $\mathrm{ran}\mathbf{T}$（由线性变换性质易证），其中由于 $\{\mathbf{a}_1, \cdots, \mathbf{a}_k\} \in \mathcal{N}$，故有 $\mathbf{T}\mathbf{a}_j = \mathbf{0}_\mathcal{W} \forall j \leq k$，所以实际上仅 $\{\mathbf{T}\mathbf{a}_{k+1}, \cdots, \mathbf{T}\mathbf{a}_n\}$ 就线性生成 $\mathrm{ran}\mathbf{T}$ 了。我们进一步验证它们是线性无关的。设标量 γ_i 满足 $\sum_{i=k+1}^n \gamma_i \mathbf{T}\mathbf{a}_i = \mathbf{0}_\mathcal{W}$，即

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

即 $\sum_{i=k+1}^n \gamma_i \mathbf{a}_i \equiv \mathbf{a} \in \mathcal{N}$。向量 $\mathbf{a}$ 在 $\mathcal{N}$ 的基 $\{\mathbf{a}_1, \cdots, \mathbf{a}_k\}$ 下表示为 $\mathbf{a} = \sum_{i=1}^k \alpha_i a i$。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 $\mathcal{W}$ 的基，即 $\mathcal{W}$ 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 $\mathcal{V}, \mathcal{W}$ 是数域 $\mathbb{F}$ 上的有限维向量空间，$\mathbf{T}: \mathcal{V} \to \mathcal{W}$ 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 $\mathcal{V}$ 中的任意一组线性无关向量组映射为 $\mathcal{W}$ 的一组线性无关向量组；*
4. *$\mathrm{rank}\mathbf{T} = \dim\mathcal{V}$。*

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

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

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

假设 $\mathbf{T}$ 总把 $\mathcal{V}$ 的一组线性无关向量映射为 $\mathcal{W}$ 的一组线性无关向量，令 $\mathbf{a} \neq \mathbf{0}_{\mathcal{V}}$ 是 $\mathcal{V}$ 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 $\mathbf{0}_{\mathcal{W}}$，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 $\mathbf{0}_{\mathcal{V}}$ 映射为 $\mathbf{0}_{\mathcal{W}}$，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：$\dim\mathcal{V} = \dim\mathcal{W}$，那么 $\mathcal{T}$ 就是双射。

推论 II.4.4.1. *设 $\mathcal{V}, \mathcal{W}$ 是数域 $\mathbb{F}$ 上的有限维向量空间且 $\dim\mathcal{V} = \dim\mathcal{W}$，$\mathbf{T}: \mathcal{V} \to \mathcal{W}$ 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 $\mathcal{V}$ 中的任意一组基映射为 $\mathcal{W}$ 的一组基；*

证明. $1 \Leftrightarrow 2$：设 $n = \dim\mathcal{V} = \dim\mathcal{W}$，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 $\Leftrightarrow \operatorname{rank}\mathbf{T} = n = \dim\mathcal{W} \Leftrightarrow \operatorname{ran}\mathbf{T} = \mathcal{W}$。

$1 \Leftrightarrow 3$：留作练习。 □

由定理II.1.7，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 $\mathcal{V}, \mathcal{W}$ 是数域 $\mathbb{F}$ 上的向量空间，$\mathbf{T} : \mathcal{V} \to \mathcal{W}$ 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 $\mathcal{W}$ 到 $\mathcal{V}$ 的线性变换。

证明. 对任意 $\mathbf{b}_1, \mathbf{b}_3 \in \mathcal{W}$ 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 $\gamma\mathbf{a}_1 + \mathbf{a}_2 \in \mathcal{V}$ 就是向量 $\gamma\mathbf{b}_1 + \mathbf{b}_2 \in \mathcal{W}$ 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 $\mathbb{F}^n$ 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. 首先证明两个维数相等的向量空间之间必存在一个同构线性变换。设 $\mathcal{V}_N$、$\mathcal{W}_N$ 是数域 $\mathbb{F}$ 上的两个 N 维向量空间。给定 $\mathcal{V}_N$ 的一组基 $B_{\mathcal{V}} = \{\mathbf{e}_i\}_{i=1}^N$，则向量 $\mathbf{a} \in \mathcal{V}_N$ 可表示为 $\mathbf{a} = \sum_{i=1}^N \alpha_i\mathbf{e}_i$，故 $\mathbf{a} = \mathbf{0}_{\mathcal{V}} \Leftrightarrow \alpha_1 = \cdots = \alpha_N = 0$。任一线性变换 $\mathbf{T} : \mathcal{V}_N \to \mathcal{W}_N$ 满足

$$\mathbf{T}\mathbf{a} = \mathbf{T}\sum_{i=1}^N \alpha_i\mathbf{e}_i = \sum_{i=1}^N \alpha_i\mathbf{T}\mathbf{e}_i$$

我们先考察 $\{\mathbf{T}\mathbf{e}_i\}_{i=1}^N$ 是否线性无关。设 $\{\mathbf{T}\mathbf{e}_i\}_{i=1}^N$ 线性相关，即存在 $\{\beta_i\}_{i=1}^N \subseteq \mathbb{F}$ 使得 $\sum_{i=1}^N \beta_i\mathbf{T}\mathbf{e}_i = \mathbf{0}_{\mathcal{W}}$ 且至少有一个 $\beta_j \neq 0$。若是如此，$\dim\mathcal{W}_N > N$，与已知条件矛盾，故 $\{\mathbf{T}\mathbf{e}_i\}_{i=1}^N$ 线性无关。

然后再考察 $\{\mathbf{T}\mathbf{e}_i\}_{i=1}^N$ 是否线性生成 $\mathcal{W}_N$。设 $\mathcal{W}'$ 是 $\{\mathbf{T}\mathbf{e}_i\}_{i=1}^N$ 的线性生成空间且 $\mathcal{W}' \subset \mathcal{W}_N$（是真子集）。由于已证 $\{\mathbf{T}\mathbf{e}_i\}_{i=1}^N$ 线性无关，则存在 $m > 0$ 个线性无关向量 $\{\mathbf{b}_i\}_{i=1}^m$ 使得

$\{\mathbf{Te}_i\}_{i=1}^N \cup \{\mathbf{b}_i\}_{i=1}^m$ 是 $\mathcal{W}_N$ 的一组基，即 $\dim\mathcal{W}_N = N + m > N$，这又与已知条件矛盾，故 $\mathcal{W}' = \mathcal{W}_N$，即 $\{\mathbf{Te}_i\}_{i=1}^N$ 线性生成 $\mathcal{W}_N$

所以 $\{\mathbf{Te}_i\}_{i=1}^N$ 是 $\mathcal{W}_N$ 的一组基。

有了这一结论，我们得以考察 $\mathbf{T}$ 是否双射。作为线性变换的 $\mathbf{T}$ 首先是满射。对任一 $\mathbf{c} \in \mathcal{W}_N$，可由 $\{\mathbf{Te}_i\}_{i=1}^N$ 线性表出，即

$$\mathbf{c} = \sum_{i=1}^N \gamma_i (\mathbf{Te}_i) \equiv \mathbf{T}\left(\sum_{i=1}^N \gamma_i \mathbf{e}_i\right) = \mathbf{Ta}', \mathbf{a}' \in \mathbf{V}_N$$

即 $\forall \mathbf{c} \in \mathcal{W}_N, \mathbf{c} \in \operatorname{ran}\mathbf{T}$，故 $\mathbf{T}$ 又是单射。总之 $\mathbf{T}$ 是双射，$\mathcal{V}_N$ 与 $\mathcal{W}_N$ 同构。 □

推论 II.4.6.1. 任一 $\mathbb{F}$ 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 $\mathcal{V}, \mathcal{W}, \mathcal{Z}$ 是数域 $\mathbb{F}$ 上的向量空间，$\mathbf{T}: \mathcal{V} \to \mathcal{W}, \mathbf{U}: \mathcal{W} \to \mathcal{Z}$ 是线性变换，则复合映射 $\mathbf{U} \circ \mathbf{T}$（记为 $\mathbf{UT}$）也是线性变换。

证明. 用 $\mathbf{UT}$ 作用于 $\gamma\mathbf{a} + \mathbf{b}$，其中 $\mathbf{a}, \mathbf{b} \in \mathcal{V}, \gamma \in \mathbb{F}$ 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 $\mathbf{I}_\mathcal{V}$，其中 $\mathcal{V}$ 是这个恒等变换所作用的向量空间。

例 II.4.2. 设 $\mathcal{V}, \mathcal{W}$ 是数域 $\mathbb{F}$ 上的向量空间，$\mathbf{T}: \mathcal{V} \to \mathcal{W}$ 是同构线性变换。则 $\mathbf{T}^{-1}\mathbf{T} = \mathbf{I}_\mathcal{V}, \mathbf{TT}^{-1} = \mathbf{I}_\mathcal{W}$。

如果一个同构线性变换是从一个向量空间到其自己的映射，则称该线性变换是自同构线性变换（endomorphic linear transformation）。由于维数相等的任意两个向量空间同构，对同构线性变换的研究总能等价于对自同构线性变换的研究。

定义 II.4.3 (线性算符)**.** 设 $\mathcal{V}$ 是数域 $\mathbb{F}$ 上的向量空间，由 $\mathcal{V}$ 到其自身的线性变换 $\mathbf{T}: \mathcal{V} \to \mathcal{V}$ 称为线性算符（*linear operator*）。

显然，线性算符都可逆。而且线性算符与其逆映射都同属同一个线性变换的空间 $\mathcal{L}(\mathcal{V}, \mathcal{V})$（简写为 $\mathcal{L}(\mathcal{V})$），故线性算符等义于自同构线性变换。对于一般的线性变换，本节开头介绍了其向量代数。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。

*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，既域上的运算结果还在域内。一般的线性变换的复合操作不满足此条件。

定理 II.4.8. *设 $\mathcal{V}$ 是数域 $\mathbb{F}$ 上的线性变换，$\mathbf{U}, \mathbf{T}_1, \mathbf{T}_2 \in \mathcal{L}(\mathcal{V})$，$\gamma \in \mathbb{F}$，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 $\mathbf{a} \in \mathcal{V}$，

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

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

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

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

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，一般意义的 n 维向量与 n 维坐标空间 $\mathbb{F}^n$ 建立一一对应关系的方式是通过选定的某组基，使得后者为前者的坐标。此时，一个 n 维向量与一个 $\mathbb{F}^n$ 的 n 元组的具体对应关系是依赖基的选择的。在未指定基的时候，不能直接用一个 n 元数组指代一个向量。

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[7]§7.3 “3” ,p. 178。

*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 $\mathbf{U}$ 是线变换的条件都没用到。

考虑数域 $\mathbb{F}$ 上的 N 维向量空间 $\mathcal{V}_N$ 的一组基 $B=\{\mathbf{a}_i\}_{i=1}^N$，任一向量 $\mathbf{x}\in\mathcal{V}$ 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 $\mathcal{V}_N,\mathcal{W}_M$ 分别是数域 $\mathbb{F}$ 上的 N、M 维向量空间，线性变换 $\mathbf{A}\in\mathcal{L}(\mathcal{V}_N,\mathcal{W}_M)$ 将 $\mathcal{V}_N$ 的一组基 $B_{\mathcal{V}}=\{\mathbf{a}_i\}_{i=1}^N$ 映射到 $\mathcal{W}_M$ 中的 N 个向量 $\mathbf{w}_k=\mathbf{A}\mathbf{a}_k,k=1,\cdots,N$。如果我们选取 $\mathcal{W}_M$ 的一组基 $B_{\mathcal{W}}=\{\mathbf{b}_j\}_{j=1}^M$，则 $\mathbf{w}_k$ 又可表示为 $\mathbf{w}_k=\sum_{j=1}^M\alpha_{jk}\mathbf{b}_j,k=1,\cdots,N$。此时，向量 $\mathbf{w}_k$ 的坐标 α_{jk} 需要两个下标来统一表示，它们构成一个 $M\times N$ 矩阵*

$$(\mathbf{A})=(\alpha_{jk})=\begin{pmatrix}\alpha_{11}&\cdots&\alpha_{1N}\\ \vdots&&\vdots\\ \alpha_{M1}&\cdots&\alpha_{MN}\end{pmatrix}$$

我们称矩阵 $(\mathbf{A})$ 或 (α_{jk}) 是线性变换 $\mathbf{A}$ 在基 $B_{\mathcal{V}}$ 与 $B_{\mathcal{W}}$ 下的表示矩阵。α_{jk} 称为 $\mathbf{A}$ 的在基 $B_{\mathcal{V}}$ 与 $B_{\mathcal{W}}$ 下的坐标或分量。

$$(\alpha_{ij})=\begin{pmatrix}\alpha_{11}&\cdots&\alpha_{1M}\\ \vdots&&\vdots\\ \alpha_{N1}&\cdots&\alpha_{NM}\end{pmatrix}$$

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。

继续上面的讨论，若给定线性变换 $\mathbf{A}:\mathcal{V}_N\to\mathcal{W}_M$、向量 $\mathbf{x}\in\mathcal{V}_N$、$\mathbf{y}\in\mathcal{W}_M$ 和基向量 $\{\mathbf{a}_i\}_{i=1}^N\subset\mathcal{V}_N,\{\mathbf{b}_j\}_{j=1}^M\subset\mathcal{W}_M$，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 $\mathbf{A}$ 在基 $\{\mathbf{a}_i\},\{\mathbf{b}_j\}$ 下的表示矩阵为 (α_{ji})，则线性关系式 $\mathbf{y}=\mathbf{A}\mathbf{x}$ 恰好可以写成关于 $\mathbf{x}$、$\mathbf{y}$ 和 $\mathbf{A}$

*我们用双下标的标量数组 $\{\alpha_{ij}\},i=1,\cdots,N,j=1,\cdots,M$ 来表示如下 $N\times M$ 矩阵

$$(\alpha_{ij})=\begin{pmatrix}\alpha_{11}&\cdots&\alpha_{1M}\\ \vdots&&\vdots\\ \alpha_{N1}&\cdots&\alpha_{NM}\end{pmatrix}$$

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：

$$
\begin{pmatrix} \eta_1 \\ \vdots \\ \eta_M \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1N} \\ \vdots & & \vdots \\ \alpha_{M1} & \cdots & \alpha_{MN} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_N \end{pmatrix}
$$

或可写成

$$(\mathbf{y}) = (\mathbf{A})(\mathbf{x})$$

这就是线性变换 $\mathbf{y} = \mathbf{A}\mathbf{x}$ 在给定基下的坐标运算法则。

上面的讨论也同时说明，任一个数域 $\mathbb{F}$ 上的一个 $M \times N$ 矩阵 A 都通过

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

定义了一个线性变换 $\mathbf{A}: \mathcal{V}_N \to \mathcal{W}_M, \mathbf{A} \in \mathcal{L}(\mathcal{V}_N, \mathcal{W}_M)$，使后者在 $\mathcal{V}_N$ 的某组基 $\{\mathbf{a}_i\}_{i=1}^N$ 和 $\mathcal{W}_N$ 的某组基 $\{\mathbf{b}_j\}_{j=1}^M$ 下的矩阵表示恰为矩阵 A。总结成定理如下。

定理 II.4.9. 设 $\mathcal{V}_N$ 和 $\mathcal{W}_M$ 是数域 $\mathbb{F}$ 上的有限维向量空间。$B_{\mathcal{V}}$ 和 $B_{\mathcal{W}}$ 分别是 $\mathcal{V}_N$ 和 $\mathcal{W}_M$ 的一组基。对每个线性变换 $\mathbf{T}: \mathcal{V}_N \to \mathcal{W}_M$ 都存在唯一一个 $\mathbb{F}$ 上的 $M \times N$ 矩阵 T 使得 $(\mathbf{Ta})_{B_{\mathcal{W}}} = T(\mathbf{a})_{B_{\mathcal{V}}} \forall \mathbf{a} \in \mathcal{V}_N$。其中 $(\cdot)_B$ 表示以 B 为基的矩阵表示。

定理 II.4.10. 设 $\mathcal{V}_N$ 和 $\mathcal{W}_M$ 是数域 $\mathbb{F}$ 上的有限维向量空间。在给定任意 $\mathcal{V}_N$ 的基 $B_{\mathcal{V}}$ 和 $\mathcal{W}_M$ 的基 $B_{\mathcal{W}}$ 下，从线性变换 $\mathbf{T}: \mathcal{V}_N \to \mathcal{W}_M$ 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 $\mathcal{L}(\mathcal{V}_N, \mathcal{W}_M)$ 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。通过下面的讨论，我们进一步获得复合变换与矩阵乘法的对应。

定理 II.4.11. 设 $\mathcal{V}, \mathcal{W}, \mathcal{Z}$ 是 $\mathbb{F}$ 上的有限维向量空间，$\{\mathbf{e}_i\}, \{\mathbf{f}_j\}, \{\mathbf{g}_k\}$ 分别是 $\mathcal{V}, \mathcal{W}, \mathcal{Z}$ 的基，$\mathbf{T}: \mathcal{V} \to \mathcal{W}, \mathbf{U}: \mathcal{W} \to \mathcal{Z}$ 是线性变换。则复合线性变换 $\mathbf{C} = \mathbf{TU}$ 在 $\{\mathbf{e}_i\}, \{\mathbf{g}_k\}$ 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 $\mathbf{T}, \mathbf{U} \in \mathcal{L}(\mathcal{V})$，由定理II.4.8有 $(\mathbf{U})(\mathbf{T}) = (\mathbf{T})(\mathbf{U}) = (\mathbf{I})$。易证在给定 $\mathcal{V}$ 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），其中

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

总结为如下定理:

定理 II.4.12. 恒等变换 $\mathbf{I}: \mathcal{V}_N \to \mathcal{V}_N$ 在任意一组基下的矩阵表示都是单位矩阵 I_N。

因此，$(\mathbf{U})(\mathbf{T}) = (\mathbf{T})(\mathbf{U}) = I$。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 $(\mathbf{T}^{-1}) = (\mathbf{T})^{-1}$。

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 $\mathcal{V}_N$ 中的两组基 $\{\mathbf{e}_i\}, \{\mathbf{e}'_j\}$。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:

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

称为从基 $\{\mathbf{e}_i\}$ 到基 $\{\mathbf{e}'_j\}$ 的基变换公式。矩阵 (S_{ij}) 称为从基 $\{\mathbf{e}_i\}$ 到基 $\{\mathbf{e}'_j\}$ 的过渡矩阵。

定理 II.4.13. 设 $\mathcal{V}_N$ 是数域 $\mathbb{F}$ 上的 N 维向量空间，$\{\mathbf{e}_i\},\{\mathbf{e}'_j\}$ 是 $\mathcal{V}_N$ 的两组基，则从基 $\{\mathbf{e}_i\}$ 到基 $\{\mathbf{e}'_j\}$ 的过渡矩阵是从基 $\{\mathbf{e}'_j\}$ 到基 $\{\mathbf{e}_i\}$ 的过渡矩阵的逆矩阵。若 $\mathbf{e}'_j = \sum_{i=1}^N S_{ij}\mathbf{e}_i, \mathbf{e}_i = \sum_{i=1}^N T_{ji}\mathbf{e}'_j$，则 $S = T^{-1}$。

证明. 如果我们把基向量 $\{\mathbf{e}_i\}$ 排成一个列向量，则基变换关系可表示为*：

$$\begin{pmatrix} \mathbf{e}'_1 \\ \vdots \\ \mathbf{e}'_N \end{pmatrix} = \begin{pmatrix} S_{11} & \cdots & S_{1N} \\ \vdots & \ddots & \vdots \\ S_{N1} & \cdots & S_{NN} \end{pmatrix} \begin{pmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_N \end{pmatrix}$$

由此易见 $S = T^{-1}$。 □

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 $\mathbf{v} \in \mathcal{V}_N$ 在两组基 $\{\mathbf{e}_i\},\{\mathbf{e}'_j\}$ 下的坐标之间的关系：

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

这 N 个式子称为向量 $\mathbf{v}$ 从基 $\{\mathbf{e}'_j\}$ 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 $\{\mathbf{e}'_j\}$ 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 $\{\mathbf{e}'_j\}$ 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 $\{\mathbf{e}'_j\}$

*虽然这严格来说不是一个矩阵，但矩阵乘以向量的式子本质上是线性方程组的一种表示，故此入可视为一个线性方程组的简化表示。矩阵的互逆，本质上也是线性方程组的求解，故本命题也可通过矩阵运算法则得到的互逆性得以证明。

到 $\{\mathbf{e}_i\}$ 的坐标变换公式是 $v'_j = \sum_{i=1}^{N} T_{ji} v_i, j = 1, \cdots, N$，其中 $T_{ji} = S_{ij}^{-1}$ 是从 $\{\mathbf{e}'_j\}$ 到 $\{\mathbf{e}_i\}$ 的过渡矩阵。

接下来，我们看线性变换的矩阵在不同基下的变换公式。

定理 II.4.14. 设 $\mathcal{V}_N$、$\mathcal{W}_M$ 分别是数域 $\mathbb{F}$ 上的 N、M 维向量空间，$\{\mathbf{a}_i\}, \{\mathbf{a}'_i\} \in \mathcal{V}_N$ 是 $\mathcal{V}_N$ 的两组基，基变换公式为 $\mathbf{a}'_j = \sum_{i=1}^{N} S_{ij} \mathbf{a}_i$；$\{\mathbf{b}_j\}, \{\mathbf{b}'_j\}$ 是 $\mathcal{W}_M$ 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 $\mathbf{A} : \mathcal{V}_N \to \mathcal{W}_M$ 在 $\{\mathbf{a}_i\}, \{\mathbf{b}_i\}$ 下的矩阵表示为 $(\mathbf{A})$，在 $\{\mathbf{a}'_i\}, \{\mathbf{b}'_i\}$ 下的矩阵表示为 $(\mathbf{A})'$。则有

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

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

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

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：

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

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

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

由该定义，若映射 $f:\mathcal{V}\to\mathbb{F}$ 满足 $f(\alpha\mathbf{a}+\mathbf{b})=\alpha f(\mathbf{a})+f(\mathbf{b}),\forall\mathbf{a},\mathbf{b}\in\mathcal{V},\alpha\in\mathbb{F}$，则 f 是 $\mathcal{V}$ 上的一个线性泛函，且 $\mathcal{V}$ 的对偶空间 $\mathcal{V}^*$ 就是线性变换空间 $\mathcal{L}(\mathcal{V},\mathbb{F})$，$\dim\mathcal{V}^*=\dim\mathcal{V}$。

例 II.4.3. 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。

虽然一个向量空间 $\mathcal{V}$ 的对偶空间 $\mathcal{V}^*$ 是 $\mathcal{V}$ 上的一种线性变换——线性泛函本身的空间，但由于，$\dim\mathcal{V}^*=\dim\mathcal{V}$，实际上 $\mathcal{V}^*$ 与 $\mathcal{V}$ 更像是一对并列概念。下面我们考察 $\mathcal{V}$ 的基与 $\mathcal{V}^*$ 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。我们先介绍一个关于线性变换的一般定理，然后介绍它关于线性泛函这一特例上的推论，最终支持我们给出“对偶基”的定义。

定理 II.4.15. 设 $\mathcal{V}$ 是数域 $\mathbb{F}$ 上的有限维向量空间，$\{\mathbf{a}_1,\cdots,\mathbf{a}_n\}$ 是 $\mathcal{V}$ 的一组有序基 ($\dim\mathcal{V}=n$)，$\mathcal{W}$ 是相同数域上的向量空间。给定 n 个 $\mathcal{W}$ 的向量 $\{\mathbf{b}_1,\cdots,\mathbf{b}_n\}$，有且只有一个线性变换 $\mathbf{T}:\mathcal{V}\to\mathcal{W}$ 使得 $\mathbf{T}\mathbf{a}_i=\mathbf{b}_i,\forall i=1,\cdots,n$。

证明. 首先证明存在性。在所给定的基下，$\mathcal{V}$ 的向量 $\mathbf{x}$ 与其坐标 $(\xi_1,\cdots,\xi_n)^\intercal$ 是一一对应的；任一 $\mathcal{V}$ 中的向量 $\mathbf{x}$ 都可表示为 $\mathbf{x}=\sum_{i=1}^n\xi\mathbf{a}_i$。我们将线性变换 $\mathcal{T}$ 具体定义为 $\mathbf{T}\mathbf{x}=\sum_{i=1}^n\xi_i\mathbf{b}_i$，即该线性变换效果是把 $\mathcal{V}$ 的向量 $\mathbf{x}$ 对应到用 $\mathbf{x}$ 的坐标与所给定的 $\mathcal{W}$ 的 n 个向量来表出的一个向量。我们可以验证 $\mathbf{T}$ 确实是一个线性变换。且这一线性变换满足定理的要求。

然后我们证明唯一性。设另有一个线性变换 $\mathbf{U}:\mathcal{V}\to\mathcal{W}$ 满足定理的要求，则 $\mathbf{U}\mathbf{x}=\mathbf{U}\left(\sum_{i=1}^n\xi_i\mathbf{a}_i\right)=\sum_{i=1}^n\xi_i(\mathbf{U}\mathbf{a}_i)=\sum_{i=1}^n\xi_i\mathbf{b}_i=\mathbf{T}$。 □

推论 II.4.15.1. 设 $\mathcal{V}$ 是数域 $\mathbb{F}$ 上的有限维向量空间，$\{\mathbf{a}_i\}$ 是 $\mathcal{V}$ 的一组基，则对每个 $i=1,\cdots,n$，有且只有一个 $\mathcal{V}$ 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"由原定理易证，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 $\mathcal{V}^*$ 的零向量）$\Leftrightarrow c_j=0\forall j=1,\cdots,n$。 □

由于 $\dim\mathcal{V}^*=\dim\mathcal{V}$，故 $\dim\mathcal{V}$ 个 $\mathcal{V}^*$ 中的线性无关向量就是 $\mathcal{V}^*$ 的一组基，因而上面的定理和推论告诉我们，给定向量空间 $\mathcal{V}$ 上的一组有序基，总能在其对偶空间 $\mathcal{V}^*$ 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 $\mathcal{V}$ 是数域 $\mathbb{F}$ 上的有限维向量空间。给定 $\mathcal{V}$ 的一组基 $B=\{\mathbf{a}_i\}$，$\mathcal{V}$ 的对偶空间 $\mathcal{V}^*$ 中的（唯一一组）满足 $f_i(\mathbf{a}_j)=\delta_{ij}, i,j=1,\cdots,\dim\mathcal{V}$ 的基 $\{f_i\}$ 称为 $\{\mathbf{a}_i\}$ 的对偶基（*dual basis*）。

由基的概念，$\mathcal{V}^*$ 中的任一线性泛函 $f\in\mathcal{V}^*$ 都可以用 $\mathcal{V}^*$ 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 $\mathcal{V}$ 的一组基 $B=\{\mathbf{a}_i\}$ 的对偶基，那么类似定理II.4.15的推论的证明过程，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 $\mathcal{V}$ 中的向量 $\mathbf{x}\in\mathcal{V}$ 都可表示为 $\mathbf{x}=\sum_{i=1}^{n}\xi_i\mathbf{a}_i$。若用 B 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"。

我们曾经在 §II.3节提过，内积定义中，我们必须给形成内积的两个向量之一赋予共轭调和的规则。本讲义是给后一个向量定义了这一规则，即 $\forall\mathbf{a},\mathbf{b}\in\mathcal{V},\alpha\in\mathbb{F},(\mathbf{a}|\alpha\mathbf{b})=\overline{\alpha}(\mathbf{a}|\mathbf{b})$，暂称第一种惯例。第二种惯例是给前一个向量定义这一规则，即 $\langle\alpha\mathbf{a}|\mathbf{b}\rangle=\overline{\alpha}\langle\mathbf{a}|\mathbf{b}\rangle$，我们在此把第二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：

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

第二种惯例与量子力学中与狄拉克 bra-ket 标记的规定是相同的。如果内积采用第二种惯例，则线性泛函和内积还有更直接的关系。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[9]§1.3.1。

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 $\mathcal{V},\mathcal{W}$ 是数域 $\mathbb{F}$ 上的向量空间，$\mathbb{T}:\mathcal{V}\to\mathcal{W}$ 是线性变换，对于 $\mathcal{W}$ 上的每一个线性泛函 $g\in\mathcal{W}^*$，我们都可以定义一个 $\mathcal{V}$ 上的线性泛函 $f\in\mathcal{V}^*$ 使其满足 $f(\mathbf{a})=g(\mathbf{T}\mathbf{a}),\forall\mathbf{a}\in\mathcal{V}$，由 $g\in\mathcal{W}^*$ 到 $f\in\mathcal{V}^*$ 的这一对应规则定义了一个由 $\mathcal{W}^*$ 到 $\mathcal{V}^*$ 的映射 $\mathbf{T}^\intercal:\mathcal{W}^*\to\mathcal{V}^*$。令 $(\mathbf{T}^\intercal g)(\cdot)=g(\mathbf{T}\cdot)$，易验对任一 $\mathbf{T}\in\mathcal{L}(\mathcal{V},\mathcal{W})$ 有且只有一个满足上述性质的 $\mathbf{T}^\intercal\in\mathcal{L}(\mathcal{W}^*,\mathcal{V}^*)$，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 $\mathbf{T} \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ 有且只有一个符合定义的映射 $\mathcal{T}^\intercal$；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 $\mathbf{U}: \mathcal{W}^* \to \mathcal{V}^*$ 满足 $(\mathbf{U}g)\,\mathbf{a} = g\mathbf{U}\mathbf{a} \forall \mathbf{a} \in \mathcal{V}$，则有 $\mathbf{U}g = \mathbf{T}^\intercal g \Rightarrow \mathbf{U}\mathbf{T}^\intercal$。关于 2)，我们可以按照线性变换的定义去验证映射 $\mathcal{T}^\intercal$ 作用于一个线性组合的结果：对任意 $g_1, g_2 \in \mathcal{W}^*, \gamma \in \mathbb{F}$，由 $[\mathbf{T}^\intercal(\gamma g_1 + g_2)](\mathbf{a}) = (\gamma g_1 + g_2)(\mathbf{T}\mathbf{a}) = \gamma g_1(\mathbf{T}\mathbf{a}) + g_2(\mathbf{T}\mathbf{a}) = \gamma(\mathbf{T}^\intercal g_1)(\mathbf{a}) + (\mathbf{T}^\intercal g_2)(\mathbf{a}) \forall \mathbf{a} \in \mathcal{V}$ 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.4.16. 设 $\mathcal{V}, \mathcal{W}$ 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：$B = \{\mathbf{a}_i\} \subset \mathcal{V}, B^* = \{f_i\} \subset \mathcal{V}^*, B' = \{\mathbf{b}_i\} \subset \mathcal{W}, B'^* = \{g_i\} \subset \mathcal{W}^*$，且令 $\mathbf{T}: \mathcal{V} \to \mathcal{W}$ 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，

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

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

由线性变换转置的定义和对偶基的定义，

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

另一方面，由于 $\mathbf{T}^\intercal g_i \in \mathcal{V}^*$ 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.4.4 线性算符的不变量

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行列式、特征值、不变量这三种取值不依赖基的选择的标量参数。

II.5 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.5.1 伴随算符

定理 II.5.1. 设 $\mathcal{V}$ 是数域 $\mathbb{F}$ 上的有限维内积空间，f 是 $\mathcal{V}$ 上的线性泛函，则存在唯一向量 $\mathbf{b} \in \mathcal{V}$ 满足 $f(\mathbf{a}) = (\mathbf{a}|\mathbf{b})\,\forall \mathbf{a} \in \mathcal{V}$。

证明. 令 $\{\mathbf{a}_1, \cdots, \mathbf{a}_n\}$ 是 $\mathcal{V}$ 的一组正交基，令 $\mathbf{b} = \sum_{j=1}^{n} \overline{f(\mathbf{a}_j)}\mathbf{a}_j$，令线性泛函 $f_{\mathbf{b}}(\mathbf{a}) = (\mathbf{a}|\mathbf{b})$，则有 $f_{\mathbf{b}}(\mathbf{a}_k) = (\mathbf{a}_k|\mathbf{b}) = \left(\mathbf{a}_k|\sum_{j=1}^{n} \overline{f(\mathbf{a}_j)}\mathbf{a}_j\right) = f(\mathbf{a}_k), k = 1, \cdots, n, \therefore f_{\mathbf{b}} = f$，即 f 是唯一的。若 $\mathbf{c} \in \mathcal{V}$ 满足 $(\mathbf{a}|\mathbf{b}) = (\mathbf{a}|\mathbf{c})\,\forall \mathbf{a} \in \mathcal{V}$，则 $(\mathbf{b}-\mathbf{c}|\mathbf{b}-\mathbf{c}) = 0 \Leftrightarrow \mathbf{b} = \mathbf{c}$，故所找到的 f 和 $\mathbf{b}$ 满足命题条件，命题得证。 □

上述证明也给出了 $\mathbf{b}$ 的找法。由上述定理的证明过程显见 $\mathbf{b}$ 在 f 的核空间的正交补集中。若 $\mathcal{W} = \ker f$ 则 $\mathcal{V} = \mathcal{W} + \mathcal{W}^{\perp}$，$f$ 完全由其在 $\mathcal{W}^{\perp}$ 的值决定。若 $\mathbf{P}$ 是由 $\mathcal{V}$ 到 $\mathcal{W}^{\perp}$ 的正交投影算符，则 $f(\mathbf{a}) = f(\mathbf{Pa})\,\forall \mathbf{a} \in \mathcal{V}$。若 $f \neq 0$，则 f 的秩为 1，$\dim\mathcal{W}^{\perp} = 1$。若 $\mathbf{c} \neq \mathbf{0} \in \mathcal{W}^{\perp}$ 则

$$\mathbf{Pa} = \frac{(\mathbf{a}|\mathbf{c})}{\|\mathbf{c}\|^2}\mathbf{c}\,\forall \mathbf{a} \in \mathcal{V}$$

$$\therefore f(\mathbf{a}) = (\mathbf{a}|\mathbf{c})\frac{f(\mathbf{c})}{\|\mathbf{c}\|^2}\forall \mathbf{a} \in \mathcal{V}, \mathbf{b} = \frac{\overline{f(\mathbf{c})}}{\|\mathbf{c}\|^2}\mathbf{c}$$

下面，我们将引入 $\mathbf{T}$ 的伴随算符 $\mathbf{T}^*$ 的定义。但在此之前先通过两个定理来明确其存在性、唯一性。

定理 II.5.2. 设 $\mathcal{V}$ 是有限维内积空间，$\mathbf{T}$ 是 $\mathcal{V}$ 上的线性算符，则存在唯一线性算符 $\mathbf{T}^*$ 满足

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

证明. □

定理 II.5.3. 设 $\mathcal{V}$ 是有限维内积空间，$B=\{\widehat{\mathbf{e}}_i\}$ 是 $\mathcal{V}$ 的一组规范正交基，$\mathbf{T}$ 是 $\mathcal{V}$ 的线性算符，则

$$(\mathbf{T})_{kj}=(\mathbf{T}\widehat{\mathbf{e}}_j|\mathbf{e}_k)$$

证明. □

定理 II.5.4. 设 $\mathcal{V}$ 是有限维内积空间，$B=\{\widehat{\mathbf{e}}_i\}$ 是 $\mathcal{V}$ 的一组规范正交基，$\mathbf{T},\mathbf{T}^*$ 是 $\mathcal{V}$ 上的线性算符满足

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

则 $\mathbf{T}^*$ 在 B 下的矩阵是 $\mathbf{T}$ 在 B 下的矩阵的共轭转置。

证明. □

定义 II.5.1 (伴随算符)**.** 设 $\mathcal{V}$ 是有限维内积空间，$\mathbf{T}$ 是 $\mathcal{V}$ 的线性算符，如果存在 $\mathcal{V}$ 的另一线性算符 $\mathbf{T}^*$ 满足 $(\mathbf{Ta}|\mathbf{b})=(\mathbf{a}|\mathbf{T}^*\mathbf{b})\,\forall\mathbf{a},\mathbf{b}\in\mathcal{V}$，则称 $\mathbf{T}^*$ 是 $\mathbf{T}$ 的伴随算符。若 $\mathbf{T}=\mathbf{T}^*$ 则称 $\mathbf{T}$ 是厄米算符或自伴随算符。

定理 II.5.5. 设 $\mathcal{V}$ 是数域 $\mathbb{F}$ 上的有限维内积空间，$\mathbf{T},\mathbf{U}$ 是 $\mathcal{V}$ 的线性算符，则

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

算符的伴随类似于复数的共轭。且由上面的相关定理可知算符的伴随对应于矩阵的共轭转置。

线性算符及其伴随算符也有类似于“对于 $\alpha\in\mathbb{F}$, 当且仅当 $\alpha=\overline{\alpha}$ 时 $\alpha\in\mathbf{R}$”的结论。数域 $\mathbb{F}$ 上的内积空间的线性算符 $\mathbf{T}$ 可表示为 $\mathbf{T}=\mathbf{U}_1+i\mathbf{U}_2,\mathbf{U}_1=\frac{1}{2}\left(\mathbf{T}+\mathbf{T}^*\right),\mathbf{U}_2=\frac{1}{2i}\left(\mathbf{T}-\mathbf{T}^*\right)$, 其中 $\mathbf{U}_1,\mathbf{U}_2$ 都是厄米算符。

线性算符的伴随算符与线性算符的转置之间的区别与联系：设 $\mathcal{V}$ 是数域 $\mathbb{F}$ 上的有限维内积空间，$\mathbf{T}$ 是 $\mathcal{V}$ 是的线性算符，则由定义，

$$\mathbf{T}^\intercal:\mathcal{V}^*\to\mathcal{V}^*,\mathbf{T}^\intercal\left(g\right)\left(\cdot\right)=g\left(\mathbf{T}\cdot\right)\forall g\in\mathcal{V}^*$$

$$\mathbf{T}^*:\mathcal{V}\to\mathcal{V},(\mathbf{Ta}|\mathbf{b})=(\mathbf{a}|\mathbf{T}^*\mathbf{b})\,\forall\mathbf{a},\mathbf{b}\in\mathcal{V}$$

注意到，由内积与线性泛函的关系，对每一 $\mathbf{a}\in\mathcal{V}$ 均可定义线性泛函 $(\mathbf{a}|\cdot)\in\mathcal{V}^*$，此外 $(\mathbf{T}^\intercal)=(\mathbf{T})^\intercal,(\mathbf{T}^*)=(\mathbf{T})^*$，其中 $\intercal,*$ 作用于矩阵分别表示矩阵的转置和共轭转置。由此可见，线性变换的转置是比线性算符的伴随更一般的概念（后者是前者的特例）。

II.5.2 幺正算符

定义 II.5.2 (保持内积)**.** 设 $\mathcal{V},\mathcal{W}$ 是同数域上的两内积空间，若线性变换 $\mathbf{T}:\mathcal{V}\to\mathcal{W}$ 满足 $(\mathbf{Ta}|\mathbf{Tb})=(\mathbf{a}|\mathbf{b})\,\forall\mathbf{a},\mathbf{b}\in\mathcal{V}$ 则称 $\mathbf{T}$ 保持内积。

定理 II.5.6. 设 $\mathcal{V},\mathcal{W}$ 是同数域上的两内积空间，关于线性变换 $\mathbf{T}:\mathcal{V}\to\mathcal{W}$ 的以下命题相互等价：

1. $\mathbf{T}$ 保持内积
2. $\mathbf{T}$ 是同构映射
3. $\mathbf{T}$ 将 $\mathcal{V}$ 的某个规范正交基映射为 $\mathcal{W}$ 的一个规范正交基
4. $\mathbf{T}$ 将 $\mathcal{V}$ 的每个规范正交基映射为 $\mathcal{W}$ 的一个规范正交基
5. $\{\mathbf{Ta}\|=\|\mathbf{A}\|\,\forall\mathbf{a}\in\mathcal{V}$

定义 II.5.3 (幺正算符)**.** 内积空间 $\mathcal{V}$ 的同构线性算符称该内积空间的幺正算符。

若 $\mathbf{U}_1$ 和 $\mathbf{U}_2$ 是内积空间 $\mathcal{V}$ 的幺正算符，则 $\mathbf{U}_2\mathbf{U}_1$ 可逆且 $\{\mathbf{U}_2\mathbf{U}_1\mathbf{a}\|=\|\mathbf{U}_1\mathbf{a}\|=\|\mathbf{a}\|\,\forall\mathbf{a}\in\mathcal{V}$。幺正算符的逆也是幺正的，$\|\mathbf{Ua}\|=\|\mathbf{a}\|\Rightarrow\|\mathbf{U}^{-1}\mathbf{b}\|=\|\mathbf{b}\|\,,\mathbf{b}=\mathbf{Ua}$。由这些事实可见，一个内积空间的所有幺正算符及其复合操作组成一个群。

定理 II.5.7. 设 $\mathbf{U}$ 是内积空间 $\mathcal{V}$ 上的线性算符，则 $\mathbf{U}$ 是幺正算符当且令当 $\mathbf{UU}^*=\mathbf{U}^*\mathbf{U}=\mathbf{I}$。

定义 II.5.4 (幺正矩阵)**.** 若 $n\times n$ 矩阵 A 满足 $AA^*=I$ 则称 A 是幺正矩阵。

定理 II.5.8. 设 $\mathcal{V}$ 是有限维内积空间，$\mathbf{U}$ 是 $\mathcal{V}$ 上的线性算符，则 $\mathbf{U}$ 是幺正算符当且仅当 $\mathbf{U}$ 在 $\mathcal{V}$ 的某一（或每一）基下的矩阵是幺正矩阵。

定义 II.5.5 (正交矩阵)**.** 若 $n\times n$ 矩阵 A 满足 $A^\intercal A=I$ 则称 A 是正交矩阵。

实正交矩阵是幺正矩阵。

II.5.3 正规算符

内积空间上的线性算符在对角化方面有新的性质。

定义 II.5.6 (正规算符)**.** 设 $\mathcal{V}$ 是有限维内积空间，$\mathbf{T}$ 是 $\mathcal{V}$ 上的线性算符，若 $\mathbf{TT}^*=\mathbf{T}^*\mathbf{T}$ 则称 $\mathbf{T}$ 是正规算符。

厄米算符和幺正算符都是正规算符（见以下定理），但正规算符的和或乘积（复合）未必是正规算符。

以下一连串定理关注的是厄米算符的可对角化性质。

定理 II.5.9. 设 $\mathbf{T}$ 是内积空间 $\mathcal{V}$ 的一个厄米算符，则 $\mathbf{T}$ 的所有特征值全是实数，且两两不同特征值对应的特征向量两两正交。

定理 II.5.10. 有限非零维内积空间的任一个厄米算符均有一个（非零）特征向量。

定理 II.5.11. 设 $\mathcal{V}$ 是有限维内积空间，$\mathbf{T}$ 是 $\mathcal{V}$ 上任一线性算符。设 $\mathcal{W}$ 是 $\mathcal{V}$ 的子空间，且 $\forall \mathbf{a} \in \mathcal{W} \mathbf{T}\mathbf{a} \in \mathcal{W}$，那么就有 $\forall \mathbf{b} \in \mathcal{W}^{\perp} \mathbf{T}^{*}\mathbf{b} \in \mathcal{W}^{\perp}$。

定理 II.5.12. 设 $\mathcal{V}$ 是有限维内积空间，$\mathbf{T}$ 是 $\mathcal{V}$ 上的厄米算符，则存在一组规范正交基的基向量就是 $\mathbf{T}$ 的特征向量。

推论 II.5.12.1. 设 A 是 $n \times n$ 厄米矩阵，则存在一个幺正矩阵 P 使得 $P^{-1}AP$ 是对角矩阵。若 A 是实对称矩阵，则存在正交矩阵 P 使得 $P^{-1}AP$ 是对角矩阵

由这些定理，我们得出结论：对于有限维实内积空间的线性算符 $\mathbf{T}$，只要它是厄米的，则必有一组规范正交基是它的特征向量。

定理 II.5.13. 设 $\mathcal{V}$ 有限维内积空间，$\mathbf{T}$ 是正规算符。则 $\mathbf{a} \in \mathcal{V}$ 是 $\mathbf{T}$ 的特征值 c 对应的特征向量当且仅当 $\mathbf{a}$ 是 $\mathbf{T}^{*}$ 的特征值 $\bar{c}$ 对应的特征向量。

定义 II.5.7 (正规矩阵)**.** 若复数值 $n \times n$ 矩阵 A 满足 $AA^{*} = A^{*}A$ 则称 A 是正规矩阵。

定理 II.5.14. 设 $\mathcal{V}$ 是有限维内积空间，B 是 $\mathcal{V}$ 的一组规范正交基。若 $\mathcal{V}$ 上的线性算符 $\mathbf{T}$ 在该基下的矩阵 (T) 是上三角矩阵，则 $\mathbf{T}$ 是正规算符当且仅当 $(\mathbf{T})$ 是对角矩阵。

定理 II.5.15. 设 $\mathcal{V}$ 是有限维内积空间，$\mathbf{T}$ 是 $\mathcal{V}$ 上的线性算符，则 $\mathcal{V}$ 中总存在一组规范正交基使得 $\mathbf{T}$ 在该其下的矩阵 $(\mathbf{T})$ 是上三角矩阵。

推论 II.5.15.1. 对每一复数值 $n \times n$ 矩阵 A 均有一幺正矩阵 U 使得 $U^{-1}AU$ 是上三角矩阵。

由上面的几条定理，我们把之前只对厄米算符成立的对角化性质推广至对正规算符都成立。

定理 II.5.16. 设 $\mathcal{V}$ 是有限维内积空间，$\mathbf{T}$ 是 $\mathcal{V}$ 上的正规算符，则 $\mathcal{V}$ 中存在一组规范正交基的基向量就是 T 的特征向量。

推论 II.5.16.1. 对每一正规矩阵 A 总有幺正矩阵 P 使得 $P^{-1}AP$ 是对角矩阵。

II.6 正规算符的谱分解

II.7 欧几里德几何

II.7.1 准备概念

定义 II.7.1 (度量空间)**.** 度量空间是个有序对 (M,d)，其中 M 是集合，d 是在 M 上的度量，$d : M \times M \to \mathbb{R}$，使得对于任何在 M 内的 x, y, z，下列条件均成立：

1. 不可区分者的同一性：$d(x,y) = 0$ 当且仅当 $x = y$
2. 对称性：$d(x,y) = d(y,x)$
3. 三角不等式：$d(x,z) \leq d(x,y) + d(y,z)$

定理 II.7.1. 设 (M,d) 是一个度量空间，则 $d(x,y) \geq 0 \forall x, y \in M$

证明. 由度量空间定义的条 3 有：$d(x,y) + d(y,x) \geq d(x,x)$ 又由条件 2 有 $d(x,y) + d(x,y) \geq d(x,x)$，再由条件 1 有 $2d(x,y) \geq 0, d(x,y) \geq 0$。 □

例 II.7.1. 设 $\mathcal{V}$ 是数域 $\mathbb{F}$ 上的一个赋范向量空间，则 $(\mathcal{V}, \|\cdot\|)$ 是一个度量空间。

定义 II.7.2 (序列的收敛性)**.** 设 (M,d) 是一个度量空间，如果对任一 $\epsilon > 0$ 总存在正整数 N 使得对任一序列 $x_1, \cdots, x_n, \cdots \in M$ 总有 $x \in M$ 满足 $d(x, x_n) < \epsilon$，则称该序列收敛于 x。

定义 II.7.3 (柯西序列)**.** 设 (M,d) 是一个度量空间，如果对任一 $\epsilon > 0$ 总存在正整数（即自然数）N 使得序列 $x_1, \cdots, x_n, \cdots \in M$ 对任意 $m, n > N$ 总有 $d(x_m, x_n) < \epsilon$，则称该序列是一个柯西序列。

例 II.7.2. 设 $\{\mathbf{x}_i\}, \mathbf{x} \in \mathcal{V}$ 是赋范向量空间 $(\mathcal{V}, \|\cdot\|)$ 的一个序列，则 $\{\mathbf{x}_i\}$ 是柯西序列当且仅当对每一正实数 $\epsilon > 0$ 总存在自然数 $N \in \mathbb{N}$ 使得只要 $m, n > N, m, n \in \mathbb{N}$ 就有 $\|\mathbf{x}_n - \mathbf{x}_m\| < \epsilon$。

定义 II.7.4 (度量空间的完备性)**.** 设 (M,d) 是一个度量空间，如果 M 中的每一柯西序列都收敛，则称 (M,d) 是完备的。

定理 II.7.2. $(\mathbb{R}, |\cdot|)$ 是完备的。

证明. 该基本定理的证明见 □

定义 II.7.5 (等距变换)**.** 设 $(X, d_X), (Y, d_Y)$ 是度量空间，等距变换是映射 $i : X \to Y$ 满足 $d_Y(i(a), i(b)) = d_X(a,b) \forall a, b \in X$。如果两个度量空间之间存在一个等距变换，则称这两个度量空间是等距的。

II.7.2 欧几里德空间

由于度量的规定，等距变换总是单射。由一个度量空间 $(\mathcal{E}, d)$ 到其自身的等距变换形成一个基于映射复合操作的变换群 $\mathcal{I}$，它是集合

$$\mathcal{I} = \{i : \mathcal{E} \to \mathcal{E} | i\text{可逆}, d(X,Y) = d(i(X), i(X)) \forall X, Y \in \mathcal{E}\}$$

易证 $i_2 \circ i_1 \in \mathcal{I} \forall i_1, i_2 \in \mathcal{I}$。

如果度量 d 满足欧几里德几何，它必须拥有一些特定的性质。我们直觉上应当同意，欧几里德空间上的平移操作是一种等距变换。以下按照平移操作来定义一类等距变换，它将是 $\mathcal{I}$ 的一个子集 $\mathcal{V}$，满足 $\forall i_1, i_2 \in \mathcal{V} \subset \mathcal{I}$

1. 交换律：$i_1 \circ i_2 (X) = i_2 \circ i_1 (X), \forall X \in \mathcal{E}$
2. 传递性：$\forall X, Y \in \mathcal{E} \exists i \in \mathcal{V}, i(X) = Y$
3. 零向量：若 $i(X_0) = X_0, X_0 \in \mathcal{E}$ 则 $i(X) = X \forall X \in \mathcal{E}$
4. 向量加：$i_1 \circ i_2 \in \mathcal{V}$
5. 标量乘：存在操作 $\mathbb{R} \times \mathcal{V} \to \mathcal{V}$
6. 内积：$(i_1 | i_1) = d^2(X,Y)$ 当 $i_1(X) = Y$*

这几条实际使 $\mathcal{V}$ 成为了一个赋范向量空间。此外，可证一个度量空间的等距变换群 $\mathcal{I}$ 至多只有一个满足上述规定的子群 $\mathcal{V}$（证明见[11]）。我们可定义由 $\mathcal{E}$ 中一点 X 到另一点 Y 的“平移向量” $\Phi : \mathcal{E}^2 \to \mathcal{V}, \Phi(\cdot, i(\cdot)) \equiv i(\cdot), i \in \mathcal{V}$，并规定一些记法：若 $i \in \mathcal{V}, X \in \mathcal{E}$，则记 $Y = i(X)$ 为 $Y = X + \Phi(X,Y)$ 或者 $\Phi(X,Y) = Y - X$。如果 $Y = i_1(X), Z = i_2(Y)$，则记 $i_1 \circ i_2 (X) = \Phi(X,Y) + \Phi(Y,Z)$。如果 $i(X) = Y$ 则内积 $(i|i) = \|\Phi(X,Y)\|$。注意到，由上述的内积形成的范是欧几里德范。如果一个度量空间 $(\mathcal{E}, d)$ 的度量 d 满足上述条件，则称 $\mathcal{E}$ 是一个欧几里德空间，$X \in \mathcal{E}$ 称为点，d 是一个欧几里德度量，$d(X,Y)$ 称为两点间的距离，常记作 $|XY| \equiv d(X,Y)$，$\mathcal{V}$ 是 $\mathcal{E}$ 的平移空间。可以验证，：$\forall O, X, Y \in \mathcal{E}$

1. $\Phi(X,Y) = \mathbf{0}$ 当且仅当 $X = Y$
2. $\Phi(Y,X) = -\Phi(X,Y)$
3. $\Phi(O,X) - \Phi(O,Y) = \Phi(X,Y)$

定理 II.7.3. 欧几里德度量空间 $(\mathcal{E}, d)$ 是完备的。

证明. 由于欧几里德度量空间的平移空间为实数域上的赋范空间，而后者是完备的。 □

*这是 W. Noll 原本的表述[10]。从字面上看，这一条规定只是赋范。若将上一条的“标量乘”具体规定为：$\alpha i(X)$ 满足 $d(X, \alpha i(X)) = \alpha d(X, i(X)) \forall X \in \mathcal{E}, \alpha \in \mathbb{R}$，则由度量 d 的一般要求可知该范的定义满足范的一般要求。再规定该范满足平行四边行法则，则可由该范定义一个内积，且范的定义可表述成内积的平方根。因此，这一表述其实是人为赋予了度量空间 $(\mathcal{E}, d)$ 以欧几里德几何。

下面我们引入，欧几里德空间中的角。由于欧几里德范总满足三角不等式和平行四边行法则，故给定 $X, O, Y \in \mathcal{E}, X \neq O \neq Y$，有

$$\|\Phi(O,X)\|^2 + \|\Phi(O,Y)\|^2 = \|\Phi(O,X) - \Phi(O,Y)\|^2 + 2(\Phi(O,X)|\Phi(O,Y))$$

由三角不等式，

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

定义 $\mathcal{E}$ 中的角为映射 $a : \mathcal{E}^3 \to \mathbb{R}$ 满足

$$\angle XOY \equiv a(X,O,Y) = \cos^{-1}\left[\frac{(\Phi(O,X)|\Phi(O,Y))}{\|\Phi(O,X)\| \|\Phi(O,Y)\|}\right], \forall X, O, Y \in \mathcal{E}, X \neq O \neq Y$$

其中 $\cos : [0,\pi] \to \mathbb{R}, \cos x = \frac{1}{2}(e^{ix} + e^{-ix})$。注意到，此处定义的函数 cos 为 $[0, pi]$ 到 $[-1,1]$ 的双射，故 $\mathrm{ran}a = [0,\pi]$，$\angle XOY = -\angle YOX$。

设 $i : \mathcal{E} \to \mathcal{E}$ 是一个等距变换，则易验有 $\angle XOY = \angle i(X)\, i(O)\, i(Y)$。

下面介绍欧几里德空间中的直线、两直线垂直的引入，并得出欧几里德空间的维数概念。给定两点 $X, Y \in \mathcal{E}, X \neq Y$，$\mathcal{E}$ 的子集

$$L_{XY} = \{C | C = X + \alpha\Phi(X,Y), \alpha \in \mathbb{R}\}$$

称为过 X, Y 两点的直线。由 Φ 定义要求可知，过 X, Y 有且只有一条直线。

如果 $\angle XOY = \frac{\pi}{2}$，则称直线 L_{OX} 与 L_{OY} 垂直。由角的定义，可知直线 L_{OX} 与 L_{OY} 垂直当且仅当 $(\Phi(O,X)|\Phi(O,Y)) = 0$，故过 $\mathcal{E}$ 中任一点的两两垂直的直线的最大数量相等且都等于 $\mathcal{V}$ 的维数，故定义 $\mathcal{E}$ 的维数就是 $\mathcal{V}$ 的维数。

过 $\mathcal{E}$ 中任意两点 X, Y 的直线 $L_{XY} = \{C | C = X + \alpha\Phi(X,Y), \alpha \in \mathbb{R}\}$ 实际上也同时定义了一个由 $\mathcal{V}$ 的子集 $\{\mathbf{a} | \mathbf{a} = \alpha\Phi(X,Y), \alpha \in \mathbb{R}\}$ 到 L_{XY} 的双射，以及 d 的像集 $d(X, L_{XY})$ 到 $\mathbb{R}$ 的双射。由于实数集的完备性，度量空间 (L_{XY}, d) 是完备的。

如果选定 $O \in \mathcal{E}$ 为原点，则对任一 $X \in \mathcal{E}, X \neq O$，度量空间 (L_{OX}, d) 都具有完备性。定义映射：$\Phi_O : \mathcal{E} \to \mathcal{V}$ 满足

$$\Phi_o(X) = \Phi(O,X), \forall X \in \mathcal{E}$$

则可由上述的完备性证明 Φ_O 是双射。我们称 Φ_O 是选定 O 为原点的 $\mathcal{E}$ 的位置向量。若 $\{\widehat{\mathbf{e}}_i\}$ 是 $\mathcal{V}$ 的一组规范正交基，则直线集 $\{L_{OX_i} | \Phi_O(X_i) = \widehat{\mathbf{e}}_i\}$ 称为 $\mathcal{E}$ 的一个直角坐标系（或称笛卡尔坐标系），我们也可以等价地称组合 $\{O, \{\widehat{\mathbf{e}}_i\}\}$ 为一个直角坐标系。

特别地，设 $\mathcal{V}$ 是 $\mathbb{R}^n$，选定原点 $O \in \mathcal{E}$ 和 $\mathbb{R}^n$ 的一组规范正交基 $\{\widehat{\mathbf{e}}_i\} \subset \mathbb{R}^n$，则称组合 $\{\mathcal{E}, \mathbb{R}^n, O, \Phi_O, \{\widehat{\mathbf{e}}_i\}\}$ 是一个欧几里德空间。我们常简记这一组合为 $\mathcal{E}$，其中直角坐标系

$\{O, \{\widehat{\mathbf{e}}_i\}\}$ 称为这个欧几里德空间的基本坐标系。欧几里德空间可因维数 n 和基本坐标系的选择不同而不同。两个欧几里德空间 $\mathcal{E}, \mathcal{E}^*$ 相同（即 $\mathcal{E} = \mathcal{E}^*$）当且仅当它们的维数相同且基本坐标系重合。

II.7.3 等距变换的表示

引理 II.7.1. *若关于某点 $X \in \mathcal{E}$ 的映射 $\mathbf{Q}_X : \mathcal{V} \to \mathcal{V}$ 满足 $\mathbf{Q}_X\mathbf{u} = \Phi\left(i\left(X\right), i\left(X + \mathbf{u}\right)\right), \forall \mathbf{u} \in \mathcal{V}$，则 $\mathbf{Q}_X$ 与 X 无关，且 $\mathbf{Q}$ 是正交张量。*

证明. 由 $\mathbf{Q}_X$ 的定义有 $\mathbf{Q}_X\Phi\left(X, Y\right) \equiv \Phi\left(i\left(X\right), i\left(Y\right)\right)$，且

$$\left(\mathbf{Q}_X\mathbf{u}|\mathbf{Q}_X\mathbf{v}\right) = \left(\Phi\left(i\left(X\right), i\left(X + \mathbf{u}\right)\right)|\Phi\left(i\left(X\right), i\left(X + \mathbf{v}\right)\right)\right) = \left(\mathbf{u}|\mathbf{v}\right), \forall \mathbf{u}, \mathbf{v} \in \mathcal{V}$$

证明 $\mathbf{Q}_X$ 是线性变换：$\forall \alpha_1, \alpha_2 \in \mathbb{R}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{v} \in \mathcal{V}$,

$$\begin{aligned}
&\left(\mathbf{Q}_X\left(\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2\right) - \mathbf{Q}_X\left(\alpha_1\mathbf{u}_1\right) - \mathbf{Q}_X\left(\alpha_2\mathbf{u}_2\right)|\mathbf{Q}_X\mathbf{v}\right)\\
=&\left(\Phi\left(i\left(X\right), i\left(X + \alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2\right)\right) - \Phi\left(i\left(X\right), i\left(X + \alpha_1\mathbf{u}_1\right)\right) - \Phi\left(i\left(X\right), i\left(X + \alpha_2\mathbf{u}_2\right)\right)\right.\\
&\left.|\Phi\left(i\left(X\right), i\left(X + \mathbf{v}\right)\right)\right)\\
=&\left(\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2|\mathbf{v}\right) - \alpha_1\left(\mathbf{u}_1|\mathbf{v}\right) - \alpha_2\left(u_2|\mathbf{v}\right) = 0\\
\Leftrightarrow&\mathbf{Q}_X\left(\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2\right) = \alpha_1\mathbf{Q}_X\mathbf{u}_1 + \alpha_2\mathbf{Q}_X\mathbf{u}_2\\
&= i\left(X_2\right) + \mathbf{Q}_{X_2}\Phi\left(X_2, X_1 + \mathbf{u}\right)
\end{aligned}$$

故 $\mathbf{Q}_X$ 是线性变换。

证明 $\mathbf{Q}_X$ 是可逆的：令 $\mathbf{v} = \widehat{\mathbf{e}}_i, \{\widehat{\mathbf{e}}_i\}$ 是 $\mathcal{V}$ 的一组规范正交基，则易验 $\{\mathbf{Q}_X\widehat{\mathbf{e}}_i\}$ 也是一组规范正交基，故 $\left(\mathbf{Q}_X\mathbf{u}|\mathbf{Q}_X\mathbf{u}\right) = \left(\mathbf{u}|\mathbf{u}\right) = 0$ 当且仅当 $\mathbf{u} = \mathbf{0} \Leftrightarrow \mathbf{Q}_X\mathbf{u} = \mathbf{u}$ 当且仅当 $\mathbf{u} = \mathbf{0}$，所以 $\mathbf{Q}_X$ 是同构线性变换。

证明 $\mathbf{Q}_X$ 是正交的：由于 $\forall \mathbf{u}, \mathbf{v} \in \mathcal{V}$ 有 $\left(\mathbf{Q}_X\mathbf{u}|\mathbf{Q}_X\mathbf{v}\right) = \left(\mathbf{u}|\mathbf{v}\right)$，且由转置的定义，又有 $\left(\mathbf{Q}_X\mathbf{u}|\mathbf{Q}_X\mathbf{v}\right) = \left(\mathbf{u}|\mathbf{Q}_X^\intercal\mathbf{Q}_X\mathbf{v}\right)$，故有 $\mathbf{Q}_X^\intercal\mathbf{Q}_X\mathbf{v} = \mathbf{v}$，即 $\mathbf{Q}_X^\intercal\mathbf{Q}_X = \mathbf{I}$，所以 $\mathbf{Q}_X$ 是正交张量。

证明 $\mathbf{Q}_X$ 与 X 的选择无关：$\forall X_1, X_2 \in \mathcal{E}, \mathbf{u} \in \mathcal{V}$，则有恒等式

$$\begin{aligned}
i\left(X_1\right) &\equiv i\left(X_2\right)+\Phi\left(i\left(X_2\right), i\left(X_2+\Phi\left(X2, X1\right)\right)\right) \\
&= i\left(X_2\right)+\mathbb{Q}_{X_2} \Phi\left(X_2, X_1\right) \\
i\left(X_1+\mathbf{u}\right) &= i\left(X_2\right)+\Phi\left(i\left(X_2\right), i\left(X_2+\Phi\left(X_2, X_1+\mathbf{u}\right)\right)\right) \\
&= i\left(X_2\right)+\mathbf{Q}_{X_2} \Phi\left(X_2, X_1+\mathbf{u}\right) \\
\therefore \mathbf{Q}_{X_1} \mathbf{u} &= \Phi\left(i\left(X_1\right), i\left(X_1+\mathbf{u}\right)\right) \\
&= \Phi\left[i\left(X_2\right)+\mathbf{Q}_{X_2} \Phi\left(X_2, X_1\right), i\left(X_2\right)+\mathbf{Q}_{X_2} \Phi\left(X_2, X_1+\mathbf{u}\right)\right] \\
&= \mathbf{Q}_{X_2} \Phi\left(X_2, X_1+\mathbf{u}\right)-\mathbf{Q}_{X1} \Phi\left(X_2, X_1\right) \\
&= \mathbf{Q}_{X_2}\left(\Phi\left(X_2, X_1+\mathbf{u}\right)-\Phi\left(X_2, X_1\right)\right)=\mathbf{Q}_{X_2} \mathbf{u}
\end{aligned}$$

□

推论 II.7.3.1. 设 $i: \mathcal{E} \rightarrow \mathcal{E}$ 是欧几里德度量空间 $(\mathcal{E}, d)$ 上的一个等距变换，则存在 $\mathcal{E}$ 的平移空间 $\mathcal{V}$ 上的一个正交线性变换 $\mathbf{Q}$ 满足 $\mathbf{Q}\mathbf{u}=\Phi\left(i\left(X\right), i\left(X+\mathbf{u}\right)\right) \forall X \in \mathcal{E}, \mathbf{u} \in \mathcal{V}$。

证明. 由前面的引理自动得证。 □

推论 II.7.3.2. 设 $(\mathcal{E}, d)$ 是一个欧几里德度量空间，$\mathcal{V}$ 是其平移空间。对 $\mathcal{E}$ 上的任一等距变换 $i: \mathcal{E} \rightarrow \mathcal{E}$，在都存在一个 $\mathcal{V}$ 上的唯一的正交线性变换 $\mathbf{Q}_i$ 满足 $i\left(X\right)-i\left(Y\right)=\mathbf{Q}_i\left(X-Y\right) \forall X, Y \in \mathcal{E}$。

证明. 由前面的引理可直接得证。 □

有了这些准备结论，以下定理成立。该定理称为等距变换的表示定理。

定理 II.7.4 (等距变换的表示定理)**.** 设 $(\mathcal{E}, d)$ 是一个欧几里德度量空间，$\mathcal{V}$ 是其平移空间。选定任一 $X_0 \in \mathcal{E}$，则任一 $\mathcal{E}$ 上的任一等距变换 $i: \mathcal{E} \rightarrow \mathcal{E}$ 都可表示为 $i\left(X\right)=i\left(X_0\right)+\mathbf{Q}_i\left(X-X_0\right)$。

注意，等距变换的表示式中的正交变换 $\mathbf{Q}$ 是依赖所讨论的等距变换 i 的，故在定理中 $\mathbf{Q}$ 带下标 i。但是后续讨论中我们将省略这一下标。

由于 $\mathbf{Q}$ 是同构的，欧几里德度量空间的任一等距变换 i 是同构变换。

具体地，我们可以显示任一 $Y\in\mathcal{E}$ 是点 $X_0+\mathbf{Q}^{-1}(Y-i(X_0))$ 的像。注意到 $\mathbf{Q}\Phi:\mathcal{E}^2\to\mathcal{V}$ 是一个复合映射，且由于 $\mathbf{Q}$ 和 Φ 都是双射，$\mathbf{Q}\Phi$ 也是双射，于是有：

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

例 II.7.3. 设 i_1、i_2 和 i_3 是欧几里德空间 $\mathcal{E}$ 上的等距变换，分别定义为：

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

i_1 是一种平移，即将任一点 $X\in\mathcal{E}$ 平移 $\mathbf{u}\equiv C-X_0$。

i_3 也是一种平移，将任一点 $X\in\mathcal{E}$ 平移 $\mathbf{v}\equiv\mathbf{Q}^{-1}(C-X_0)$。

可验证 $i_1\circ i_2(X)=i_2\circ i_3(X)$。

当 $\mathbf{Q}=\mathbf{I}$ 时，i_2 是恒等映射。当 $\mathbf{Q}\neq\mathbf{I}$ 时，由于 $\mathbf{Q}$ 是正交线性算符，故 $\det\mathbf{Q}=\pm1$。若 $\det\mathbf{Q}=1$，则 $\mathbf{Q}$ 是一种旋转。若 $\det\mathbf{Q}=-1$，则由 $\mathbf{Q}\equiv(-\mathbf{I})(-\mathbf{Q})$，$\det(-\mathbf{Q})=1$ 可知 $\mathbf{Q}$ 是旋转之后再进行了反转（$-\mathbf{I}$），使得 $\Phi(X,i(Y))=-\Phi(X,Y)\,\forall X,Y\in\mathcal{E}$。由于反转是实际物体无法发生的形变，因此我们在连续介质力学中只考虑 $\det\mathbf{Q}>0$ 的等距变换。

II.8 向量函数及其图像

在向量函数微积分部分的讨论，默认总是选择标准基作为 $\mathbb{R}^n$ 的基，即，向量 $\mathbf{f}=(f_1,\cdots,f_n)\in\mathbb{R}^n$ 的坐标就是 $f_1,\cdots,f_n$，其矩阵是 $(f_1,\cdots,f_n)^{\intercal}$。因此，是否把一个向量写成“列向量”（即 $n\times1$ 矩阵）的问题变得平凡。当我们想强调矩阵运算法则时，我们会注意明确“行、列向量”。我们会以专门的章节讨论选择非标准基乃至非规范正交基的情况。

定义 II.8.1 (向量函数)**.** 从 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的映射 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$ 是一个向量函数。其变量是一个 n 维向量空间的向量 $\mathbf{r}\in\mathbb{R}^n$，$\mathbf{r}=(r_1,\cdots,r_n)$；其像（函数值）是一个 m 维向量空间的向量 $\mathbf{f}\in\mathbb{R}^m$，$\mathbf{f}=(f_1,\cdots,f_m)$。在标准基下，$f_j:\mathbf{R}_n\to\mathbf{R},j=1,\cdots,m$ 称为 $\mathbf{f}$ 的坐标函数。

严格而言，函数的记法 $\mathbf{f}:\mathbb{R}^n\supset D\to\mathbb{R}^m$ 表示函数 f 的定义域 $\mathrm{dom}f=D$，而不是整个 $\mathbf{R}^n$。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，当定义域问题不重要时，我们也常粗略记为 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$。

例 II.8.1. 函数 $\mathbf{f}:\mathbb{R}^3\to\mathbb{R}^2$,

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

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

其中 $\mathbf{r}=(r_1,r_2,r_3)^\intercal\in\mathbb{R}^3$，$\mathbf{f}=(f_1,f_2)^\intercal\in\mathbb{R}^2$。

以往我们更习惯把上述函数写成：

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

例 II.8.2. 函数 $\mathbf{f}:\mathbb{R}^3\to\mathbb{R}^2$,

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

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

定义 II.8.2 (隐函数)**.** 考虑函数 $\mathbf{F}:\mathbb{R}^{n+m}\to\mathbb{R}^m$，若 $\mathbb{R}^{n+m}$ 中的元素 $(x_1,\cdots,x_n,y_1,\cdots,y_m)^\intercal$ 写成 $(\mathbf{x},\mathbf{y})^\intercal$, 其中 $\mathbf{x}=(x_1,\cdots,x_n)^\intercal$, $\mathbf{y}=(y_1,\cdots,y_m)^\intercal$, 则 $\mathbf{F}(\mathbf{r}),\mathbf{r}\in\mathbb{R}^{n+m}$ 也可写成 $\mathbf{F}(\mathbf{x},\mathbf{y})$。若函数 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$ 满足方程 $\mathbf{F}(\mathbf{x},\mathbf{y})=\mathbf{0},\forall\mathbf{x}\in\mathbb{R}^n$ 则称这一方程隐含地定义了函数 $\mathbf{f}$。

例 II.8.3. 设函数 $F(x,y)=x^2+y^2-1=0$, 则 "$F(x,f(x))=x^2+(f(x))^2-1=0\forall x\in\operatorname{dom}f$" 隐含地定义了以下任一函数：

$$f_1(x)=\sqrt{1-x^2},-1\leq x\leq 1$$
$$f_2(x)=-\sqrt{1-x^2},-1\leq x\leq 1$$
$$f_3(x)=\begin{cases}\sqrt{1-x^2}, & -\frac{1}{2}\leq x\leq 0\\ -\sqrt{1-x^2}, & 0\leq x\leq 1\end{cases}$$

定义 II.8.3 (函数的图像)**.** * 函数 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$ 的图像是指所有有序对 $(\mathbf{r},\mathbf{f}(\mathbf{r}))$ 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是一种惯例。具体地，由定义，函数 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$ 的图像是由所有满足

$$\begin{cases} y_1=f_1(r_1,\cdots,r_n)\\ \qquad\vdots\\ y_m=f_m(r_1,\cdots,r_n)\end{cases}$$

*请复习高等数学相关内容[12]§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包括 $n + m \leq 3$ 的情况。

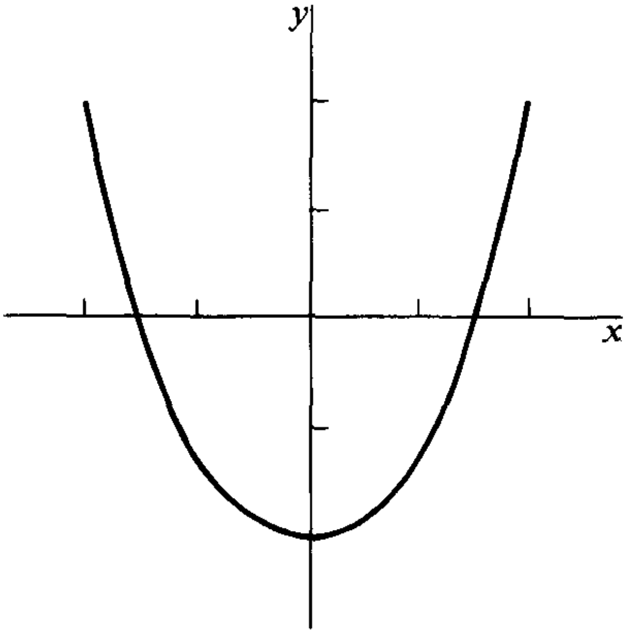

图 II.8.1: $y = x^2 - 2$

例 II.8.4. 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.8.1*所示的一条二次曲线。

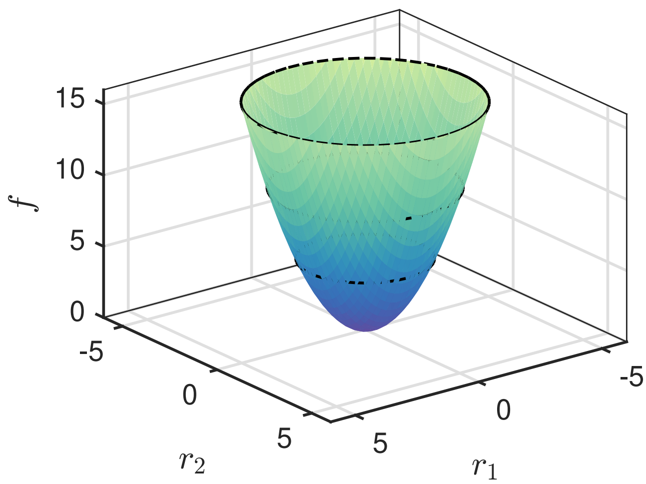

图 II.8.2: $f(\mathbf{r}) = \|\mathbf{r}\|^2$

例 II.8.5. $f : \mathbb{R}^2 \to \mathbb{R}, f(\mathbf{r}) = \|\mathbf{r}\|^2$ 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.8.2*所示的曲面。

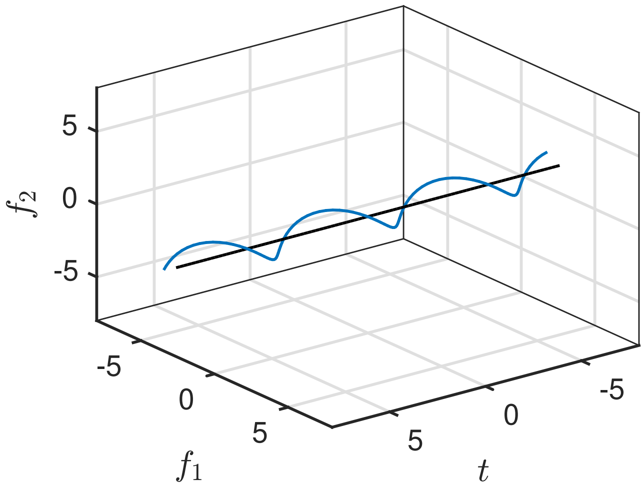

图 II.8.3: $\mathbf{f}(t) = (\cos t, \sin t)^\intercal$

例 II.8.6. 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 $\widehat{\mathbf{e}}_1$ 轴夹角的弧度。当 t 在 $\mathbb{R}$ 内取不同值时，$\mathbf{f}$ 就在 $\widehat{\mathbf{e}}_1$-$\widehat{\mathbf{e}}_2$ 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.8.3*所示的螺线。

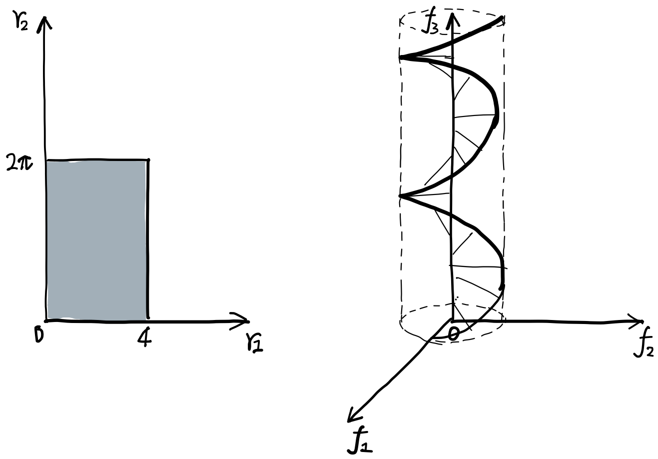

图 II.8.4: 函数 $\mathbf{f}(\mathbf{r}) = (f_1, f_2, f_3)\,(r_1 \cos r_2, r_1 \sin r_2, r_2)^\intercal, 0 \leq r_1 \leq 4, 0 \leq r2 \leq 2\pi$

例 II.8.7. 函数 $f : \mathbb{R}^2 \to \mathbb{R}^3$,

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

$\mathbf{f}$ 的定义域是 *2* 维平面上的一个矩形（图*II.8.4*）。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.8.4*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.8.4*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。

定义 II.8.4 (函数的水平集)**.** * 函数 $\mathbf{f}: \mathbb{R}^n \to \mathbb{R}^m$ 的水平集 $S = \{\mathbf{r}|\mathbf{f}(\mathbf{r}) = \mathbf{a}\}, \mathbf{a} \in \mathbb{R}^m$。

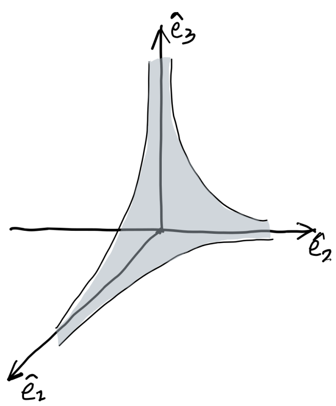

图 II.8.5: 函数 $f(\mathbf{r}) = r_1r_2 + r_2r_3 + r_3r_1$

例 II.8.8. 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.8.5*所示）。但是，按定义，这不是函数 $\mathbf{f}$ 的图像。

*在高等数学课里，我们学过水平集的概念[2]§7.1,p 1。

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。
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.9 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.9.1 (函数的极限)**.** 给定函数 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$，如果对任意正实数 ϵ 总存在正实数 δ 使得只要 $0<\|\mathbf{r}-\mathbf{r}_0\|<\delta,\mathbf{r},\mathbf{r}_0\in\mathbb{R}^n$ 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，记为

$$\lim_{\mathbf{r}\to\mathbf{r}_0}\mathbf{f}(\mathbf{r})=\mathbf{y}_0$$

在极限的定义中用到了向量的范。但是前面讲过，向量的范可以有不同定义。我们先讨论，不同定义的范在此是等价的。此处所谓“等价”是指，对任意两种定义的范 $\|\cdot\|_1,\|\cdot\|_2$，如果存在正实数 k 和 K 使得 $k\|\mathbf{x}\|_1\le\|\mathbf{x}\|_2\le K\|\mathbf{x}\|_1,\forall\mathbf{x}$，则这两种定义的范等价。可以验证，这一等价的定义满足等价关系的一般要求：自反性、对称性和传递性。

定理 II.9.1. 有限维向量空间 $\mathcal{V}$ 上的任意两种范的定义等价。

证明. 设 $n=\dim\mathcal{V}$，$\{\widehat{\mathbf{e}}_i\}$ 是 $\mathcal{V}$ 的一组基，记该基下的欧几里德范为 $\forall\mathbf{x}=\sum_{i=1}^n x_i\widehat{\mathbf{e}}_i\in\mathcal{V}$

$$\|\mathbf{x}\|_{\mathrm{E}}\equiv\sqrt{\sum_{i=1}^n x_i^2}$$

由范的定义II.3.3要求，

$$\begin{aligned}\|\mathbf{x}\|&=\left\|\sum_{i=1}^n x_i\widehat{\mathbf{e}}_i\right\|\\&\le\sum_{i=1}^n\|x_i\widehat{\mathbf{e}}_i\|=\sum_{i=1}^n x_i\|\widehat{\mathbf{e}}_i\|\\&\le\left(\sum_{i=1}^n\|\widehat{\mathbf{e}}_i\|\right)\max\{|x_i|\}\\&\le\left(\sum_{i=1}^n\|\widehat{\mathbf{e}}_i\|\right)\sqrt{\sum_{i=1}^n x_i^2}=K\|\mathbf{x}\|_{\mathrm{E}}\end{aligned}$$

其中 $K=\sum_{i=1}^{n}\|\widehat{\mathbf{e}}_i\|>0$。

下面考虑一般范作为欧几里德范的函数：$\|\cdot\|:\mathbb{R}\to\mathbb{R}$，并考察其连续性*。我们看到，对每一 $\epsilon>0$，选择 $\delta=\epsilon/K$，则只有 $\|\mathbf{x}-\mathbf{x}_0\|<\delta$，就有

$$\left|\|\mathbf{x}\|-\|\mathbf{x}_0\|\right|\le\|\mathbf{x}-\mathbf{x}_0\|\le K\|\mathbf{x}-\mathbf{x}_0\|_{\mathrm{E}}\le\epsilon,\forall\mathbf{x}_0$$

故一般范作为欧几里德范的函数是连续函数。令 k 等于一般范作为欧几里德范的函数在 $\|\mathbf{x}\|=1$ 的 $\mathcal{V}$ 的子集区域的极小值（闭区间上的连续函数必存在极小值），则对任一 $\mathbf{x}\neq\mathbf{0}$ 有 $\|\mathbf{x}/\|\mathbf{x}\|_{\mathrm{E}}\|\le k$, 故 $\|\mathbf{x}\|\le k\|\mathbf{x}\|_{\mathrm{E}},\forall\mathbf{x}\in\mathcal{V}$。 □

定理 II.9.2. 函数 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$ 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. 注意到 $\mathbf{f}$ 与 $f_i,i=1,\cdots,m$ 的定义域都相同。若极限 $\lim_{\mathbf{r}\to\mathbf{r}_0}\mathbf{f}(\mathbf{r})=\mathbf{a}$ 存在，则对任一实数 $\epsilon>0$ 都存在实数 $\delta>0$ 使得只要 $\mathbf{f}$ 的定义域内一向量 $\mathbf{r}$ 到 $\mathbf{r}_0$ 的距离 $0<\|\mathbf{r}-\mathbf{r}_0\|<\delta$ 便有 $\|\mathbf{f}(\mathbf{r})-\mathbf{a}\|<\epsilon$。由三角不等式有 $\|f_i(\mathbf{r})-a_i\|\le\|\mathbf{f}(\mathbf{r})-\mathbf{a}\|<\epsilon,i=1,\cdots,m$，即 $\lim_{\mathbf{r}\to\mathbf{r}_0}f_i(\mathbf{r})=a_i,i=1,\cdots,m$。

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.9.2 (函数的连续性)**.** 给定函数 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$，如果极限 $\lim_{\mathbf{r}\to\mathbf{r}_0}\mathbf{f}(\mathbf{r})=\mathbf{a}$ 存在，则称 $\mathbf{f}(\mathbf{r})$ 在 $\mathbf{r}_0$ 处连续。

定理 II.9.3. 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。

*这个函数相当于向量值参数方程：$\mathbf{y}=\mathbf{f}(\mathbf{x}),\mathbf{y}=\|\mathbf{t}\|,\mathbf{x}=\|\mathbf{t}\|_{\mathrm{E}},\mathbf{t}\in\mathcal{V}$

†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. 由定理II.9.2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。

定理 II.9.4. 设 $\mathcal{V}_N$、$\mathcal{W}_M$ 分别是数域 $\mathbb{F}$ 上的赋范向量空间，$\mathbf{L}:\mathcal{V}_N\to\mathcal{W}_M$ 是一个线性变换，则总存在 $k\in\mathbb{R}$ 使得不等式 $\|\mathbf{L}\mathbf{x}\|\le k\,\|\mathbf{x}\|\,,\forall\mathbf{x}\in\mathcal{V}_N$ 成立。

证明. 设 $\{\widehat{\mathbf{e}}_i\}$ 是 $\mathcal{V}_N$ 的一个规范正交基，$\mathbf{x}\in\mathcal{V}_N$，则 $\|\mathbf{L}\mathbf{x}\|=\left\|\sum_{i=1}^N x_i\mathbf{L}\widehat{\mathbf{e}}_i\right\|=\sum_{i=1}^N|x_i|\,\|\mathbf{L}\widehat{\mathbf{e}}_i\|$。又由柯西–施瓦茨不等式，$|x_i|=|(\mathbf{x}|\widehat{\mathbf{e}}_i)|\le\|\mathbf{x}\|\,\|\widehat{\mathbf{e}}_i\|=\|\mathbf{x}\|$，故 $\|\mathbf{L}\mathbf{x}\|\le\|\mathbf{x}\|\sum_{i=1}^N\|\mathbf{L}\widehat{\mathbf{e}}_i\|=k\,\|\mathbf{x}\|$，其中 $k=\sum_{i=1}^N\|\mathbf{L}\widehat{\mathbf{e}}_i\|$。 □

推论 II.9.4.1. 线性变换是连续函数。

证明. 由定理II.9.4有 $\|\mathbf{L}\mathbf{x}-\mathbf{L}\mathbf{x}_0\|=\|\mathbf{L}\,(\mathbf{x}-\mathbf{x}_0)\|\le k\,\|\mathbf{x}-\mathbf{x}_0\|$，故当 $\mathbf{x}$ 趋于 $\mathbf{x}_0$，$\mathbf{L}\mathbf{x}$ 趋于 $\mathbf{L}\mathbf{x}_0$*。 □

定义 II.9.3 (一元向量函数的导数). 设函数 $\mathbf{f}:\mathbb{R}\to\mathbb{R}^n$ 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

$$\frac{d\mathbf{f}}{dt}\equiv\lim_{h\to 0}\frac{\mathbf{f}\,(t+h)-\mathbf{f}\,(t)}{h}$$

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 $\mathbf{f}:\mathbb{R}\to\mathbb{R}^n,\mathbf{g}:\mathbb{R}\to\mathbb{R}^n$,

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

*这里使用到与不等式有关的极限定理。

定义 II.9.4 (多元标量函数的偏导数). 给定函数 $f:\mathbb{R}^n\to\mathbb{R}$，若对某一 $i=1,\cdots,n$ 极限

$$\lim_{t\to 0}\frac{f(\cdots,x_i+h,\cdots)-f(\cdots,x_i,\cdots)}{t}$$

存在，则称该极限为函数 $f(\mathbf{x})$ 对 $\mathbf{x}$ 的第 i 个分量 x_i 的偏导数，记为

$$\frac{\partial f}{\partial x_i}\equiv\lim_{t\to 0}\frac{f(\cdots,x_i+h,\cdots)-f(\cdots,x_i,\cdots)}{t}$$

定理 II.9.5. * 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.9.5 (向量函数的偏导数). 对于函数 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$ 的偏导数

$$\frac{\partial\mathbf{f}}{\partial x_i}\equiv\begin{pmatrix}\frac{\partial f_1}{\partial x_i}\\ \vdots\\ \frac{\partial f_m}{\partial x_i}\end{pmatrix}\in\mathbb{R}^m$$

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

II.10 全微分与全导数

回顾多元标量函数的全微分的定义[2]p. 19 的思想，推广至向量函数，需先引入仿射函数，作为一元标量线性函数 $y=kx+b$ 的推广。

定义 II.10.1 (仿射函数). 函数 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m,\mathbf{f}(\mathbf{x})=\mathbf{A}\mathbf{x}+\mathbf{y}_0,\mathbf{x}\in\mathbb{R}^n,\mathbf{y}_0\in\mathbb{R}^m$ 称为仿射函数，其中 $\mathbf{A}:\mathbb{R}^n\to\mathbb{R}^m$ 是一个线性变换（线性函数），$\mathbf{y}_0$ 是一个常向量。

注意，一般地仿射函数不是线性变换。

定义 II.10.2 (全微分与全导数). 设函数 $\mathbf{f}:\mathbb{R}^n\supset D\to\mathbb{R}^m$，若对 $\mathbf{x}_0\in D=\mathrm{dom}\mathbf{f}$，存在一个线性变换 $\mathbf{L}:\mathbb{R}^n\to\mathbb{R}^m$ 使得

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

则称 $\mathbf{L}(\mathbf{x}-\mathbf{x}_0)$ 是 $\mathbf{f}(\mathbf{x})$ 在 $\mathbf{X}_0$ 处的全微分。函数 $\mathbf{f}(x)$ 在 $\mathbf{x}=\mathbf{x}_0$ 处可微分。线性变换 $\mathbf{L}$ 称为函数 $\mathbf{f}(\mathbf{x})$ 在 $\mathbf{x}=\mathbf{x}_0$ 处的（全）导数，记为

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

*即高等数学[2]p. 16 定理 7.2.1 向 n 维的推广。

†进一步了解：Symmetry of second derivatives。

如果函数 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$ 在其定义域上都可微，则其导函数 $\mathbf{L}:\mathbf{R}^n\to\mathcal{L}(\mathbf{R}^n,\mathbf{R}^m)$ 是一个函数值为线性变换的函数 $\mathbf{L}(\mathbf{x})=\left.\frac{d\mathbf{f}(\mathbf{x}')}{d\mathbf{x}'}\right|_{\mathbf{x}'=\mathbf{x}}$。$\mathbf{f}$ 在点 $\mathbf{x}_0$ 处的导数是 $\mathbf{L}(\mathbf{x}_0)$，这里的括号表示函数的自变量。

我们用以往熟悉的语言复述全导数的定义：如果函数 $\mathbf{f}(\mathbf{x})$ 在点 $\mathbf{x}_0$ 处的全增量 $\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0)$ 可以写成 $\mathbf{x}-\mathbf{x}_0$ 的一个线性主部 $\mathbf{L}(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0)$ 和一个高阶无穷小 $\|\mathbf{x}-\mathbf{x}_0\|\,\mathbf{z}(\mathbf{x}-\mathbf{x}_0)$（其中函数 $\mathbf{z}(\mathbf{x})\to\mathbf{0}$ 当 $\mathbf{x}\to\mathbf{0}$）的和，则称这个线性主部 $\mathbf{L}(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0)$ 是函数在 $\mathbf{x}=\mathbf{x}_0$ 处的全微分。如果函数 $\mathbf{f}(x)$ 在 $\mathbf{x}=\mathbf{x}_0$ 处可微分，则在 $\mathbf{x}_0$ 附近 $\mathbf{f}(\mathbf{x})$ 可由 $\mathbf{L}(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0)$ 近似，即 $\mathbf{f}(\mathbf{x})\approx\mathbf{L}(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0)$。

全微分的思想是，如果函数在某点处可微分，则在该点邻近函数可近似为一个仿射函数。给定一个函数 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$，若 $\mathbf{x}_0\in\mathrm{dom}\mathbf{f}$。如果函数在 $\mathbf{x}=\mathbf{x}_0$ 处恰好等于如下仿射函数：

$$\mathbf{f}(\mathbf{x}_0)=\mathbf{L}\mathbf{x}_0+\mathbf{a}_0\equiv\mathbf{a}(\mathbf{x}_0)$$

其中仿射函数 $\mathbf{a}:\mathbb{R}^n\to\mathbb{R}^m,\mathbf{a}(\mathbf{x})=\mathbf{L}\mathbf{x}+\mathbf{a}_0$，则 $\mathbf{a}_0$ 可用 $\mathbf{f}(\mathbf{x}_0)$ 表示，使得函数 $\mathbf{a}(\mathbf{x})$ 变为：

$$\begin{aligned}\mathbf{a}(\mathbf{x})&=\mathbf{L}\mathbf{x}+\mathbf{a}_0\\&=\mathbf{L}\mathbf{x}+\mathbf{f}(\mathbf{x}_0)-\mathbf{L}\mathbf{x}_0\\&=\mathbf{L}(\mathbf{x}-\mathbf{x}_0)+\mathbf{f}(\mathbf{x}_0)\end{aligned}$$

于是，我们就用 $\mathbf{f}(\mathbf{x})$ 在 $\mathbf{x}=\mathbf{x}_0$ 处的值构造了一个仿射函数。按照全微分的思想，我们进一步希望这一仿射函数 $\mathbf{a}(\mathbf{x})$ 在 $\mathbf{x}=\mathbf{x}_0$ 附近能近似代替原函数 $\mathbf{f}(\mathbf{x})$，即：

$$\lim_{\mathbf{x}\to\mathbf{x}_0}(\mathbf{f}(\mathbf{x})-\mathbf{a}(\mathbf{x}))=\mathbf{0}$$

使用 $\mathbf{f}(\mathbf{x}_0)$ 来表示 $\mathbf{a}(\mathbf{x})$，上式 $\Leftrightarrow$

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

其中，由定理II.9.4的推论 $\lim_{\mathbf{x}\to\mathbf{x}_0}\mathbf{L}(\mathbf{x}-\mathbf{x}_0)=\mathbf{0}$，故上式 $\Leftrightarrow$

$$\lim_{\mathbf{x}\to\mathbf{x}_0}(\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0))=\mathbf{0}$$

即函数 $\mathbf{f}(\mathbf{x})$ 在 $\mathbf{x}=\mathbf{x}_0$ 处连续。事实上，如果已知 $\mathbf{L}$ 就是 $\mathbf{f}$ 的导数，则以下定理已经得证。

定理 II.10.1. *如果函数 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$ 在点 $\mathbf{x}_0\in\mathrm{dom}\mathbf{f}$ 处可微分，则它在该点处连续。*

以上讨论明确了两个互为充要条件的命题：函数 $\mathbf{f}$ 在 $\mathbf{x}_0$ 处连续 $\Leftrightarrow$ 函数 $\mathbf{f}$ 在 $\mathbf{x}_0$ 处等于某仿射函数。但是我们还未讨论，这一仿射函数 $\mathbf{a}(\mathbf{x})$ 在 $\mathbf{x}=\mathbf{x}_0$ 附近能近似代替原函数 $\mathbf{f}(\mathbf{x})$ 的

条件，这需要再要求 $\mathbf{f}(\mathbf{x})-\mathbf{a}(\mathbf{x})$ 趋于 $\mathbf{0}$ 的速度快于 $\mathbf{x}$ 趋于 $\mathbf{x}_0$ 的速度（高阶无穷小的意义），即

$$\begin{aligned}&\lim_{\mathbf{x}\to\mathbf{x}_0}\frac{\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0)\mathbf{L}(\mathbf{x}-\mathbf{x}_0)}{\|\mathbf{x}-\mathbf{x}_0\|}=\mathbf{0}\\ \Leftrightarrow&\mathbf{f}(\mathbf{x})=\mathbf{f}(\mathbf{x}_0)+\mathbf{L}(\mathbf{x}-\mathbf{x}_0)+\|\mathbf{x}-\mathbf{x}_0\|\,\mathbf{z}(\mathbf{x}-\mathbf{x}_0)\end{aligned}$$

其中函数 $\mathbf{z}(\mathbf{x})$ 具有性质 $\lim_{\mathbf{x}\to\mathbf{0}}\mathbf{z}(\mathbf{x})=\mathbf{0}$。注意到上式就是全微分定义式。要使这一仿射函数 $\mathbf{a}(\mathbf{x})$ 满足要求，需且只需满足此要求的线性变换 $\mathbf{L}$ 存在。因此，一个函数在某点可微分与其在某点存在导数是一回事，即“可导必可微、可微必可导”。

定义 II.10.3 (雅可比矩阵)**.** *设函数 $\mathbf{f}:\mathbb{R}^n\supset D\to\mathbb{R}^m$ 在 $\mathbf{x}_0\in D$ 处可微分，则其全导数在标准基下的矩阵表示称为该函数的雅可比矩阵。*

我们来看函数 $\mathbf{f}(x)$ 的雅可比矩阵具体等于什么。若 $\mathbf{f}(\mathbf{x})$ 在 $\mathbf{x}=\mathbf{x}_0$ 处连续，则总存在足够小的正实数 t 使得 $\mathbf{x}_j=\mathbf{x}_0+t\widehat{\mathbf{e}}_j\in D, j=1,\cdots,n$，其中 $\{\widehat{\mathbf{e}}_j\}$ 是 $\mathbb{R}^n$ 的标准基。若 $\mathbf{f}(\mathbf{x})$ 在 $\mathbf{x}=\mathbf{x}_0$ 处还可微分，则存在全导数 $\mathbf{L}(\mathbf{x}_0)$ 使得极限

$$\lim_{t\to 0}\frac{\mathbf{f}(\mathbf{x}_j)-\mathbf{f}(\mathbf{x}_0)-\mathbf{L}(\mathbf{x}_0)(t\widehat{\mathbf{e}}_j)}{t}=\mathbf{0}, j=1,\cdots,n$$

即

$$\lim_{t\to 0}\frac{\mathbf{f}(\mathbf{x}_j)-\mathbf{f}(\mathbf{x}_0)}{t}=\mathbf{L}(\mathbf{x}_0)\widehat{\mathbf{e}}_j, j=1,\cdots,n$$

若 $\mathbf{x}_0=(x_{01},\cdots,x_{0n})^\intercal$，则 $\mathbf{x}_i=(\cdots,x_{0i}+t,\cdots), i=1,\cdots,n$，故上式左边的极限就是偏导数 $\frac{\partial\mathbf{f}}{\partial x_0 i}$。而上式右边是线性变换 $\mathbf{L}(\mathbf{x}_0)$ 在标准基下的矩阵的第 j 列的列向量。所以上式等价于如下等式

$$\begin{aligned}(\mathbf{L}(\mathbf{x}_0))&=\left(\left(\frac{\partial\mathbf{f}}{\partial x_1}\right),\cdots,\left(\frac{\partial\mathbf{f}}{\partial x_n}\right)\right)^\intercal\\&=\begin{pmatrix}\frac{\partial f_1}{\partial x_1}&\cdots&\frac{\partial f_1}{\partial x_n}\\ \vdots&\ddots&\vdots\\ \frac{\partial f_m}{\partial x_1}&\cdots&\frac{\partial f_m}{\partial x_n}\end{pmatrix}\end{aligned}$$

由于在给定基下，一个线性变换唯一对应一个矩阵，故只要函数 $\mathbf{f}(\mathbf{x})$ 在 $\mathbf{x}=\mathbf{x}_0$ 的全导数 $\mathbf{L}(\mathbf{x}_0)$ 存在，则它是唯一的。那么，$\mathbf{L}(\mathbf{x}_0)$ 到底何时存在呢?

定理 II.10.2. *若函数 $\mathbf{f}:\mathbb{R}^n\supset D\to\mathbb{R}^m$ 的定义域 D 是开集，偏微分 $\frac{\partial f_i}{\partial x_j}, i=1,\cdots,n, j=1,\cdots,m$ 在 D 内都连续，则 $\mathbf{f}$ 在 D 内均可微分。*

证明. 设 $\mathbf{x}=(b_1,\cdots,b_n)^\intercal,\mathbf{x}_0=(a_1,\cdots,a_n)^\intercal\in D$，令 $\mathbf{y}_k=(b_1,\cdots,b_k,a_{k+1},\cdots,a_n),k=0,\cdots,n$，则有 $\mathbf{y}_0=\mathbf{x}_0,\mathbf{y}_n=\mathbf{x}$，且 $\|\mathbf{y}_k-\mathbf{y}_{k-1}\|=|b_k-a_k|,k=1,\cdots,n$。故有 $\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0)=\sum_{k=1}^n(\mathbf{f}(\mathbf{y}_k)-\mathbf{f}(\mathbf{y}_{k-1})),k=1,\cdots,n$。考察上式等号右边的求和项：

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

其中 $\{\widehat{\mathbf{e}}_i\}$ 是 $\mathbb{R}^n$ 的标准基。注意到点 $\mathbf{y}_k$ 与 $\mathbf{y}_{k-1}$ 之间的连线是长度为 $|b_k-a_k|$、方向与第 k 个坐标轴 $\widehat{\mathbf{e}}_k$ 平行的有向线段。

对上式右边的坐标函数应用微分中值定理。由命题条件，坐标函数的偏导数 $\frac{\partial f_i}{\partial x_j}$ 在 D 内都连续，则坐标函数 f_i 本身在 D 内也连续，在 $\mathbf{y}_k$ 与 $\mathbf{y}_{k-1}$ 连线上必存在一点 c_k 使得

$$\frac{f_i(\mathbf{y}_k)-f_i(\mathbf{y}_{k-1})}{b_k-a_k}=\left.\frac{\partial f_i}{\partial x_k}\right|_{x_k=c_k}$$

代入上一个求和式得：

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

再代入上一个求和式得：

$$\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0)=\sum_{k=1}^n(b_k-a_k)\left.\frac{\partial \mathbf{f}}{\partial x_k}\right|_{x_k=c_k}$$

令线性变换 $\mathbf{L}:\mathbb{R}^n\to\mathbb{R}^m$ 满足

$$\begin{aligned}(\mathbf{L}(\mathbf{x}-\mathbf{x}_0))&=\left(\left(\left.\frac{\partial \mathbf{f}}{\partial x_1}\right|_{x_1=a_1}\right),\cdots,\left(\left.\frac{\partial \mathbf{f}}{\partial x_1}\right|_{x_n=a_n}\right)\right)(b_1-a_1,\cdots,b_n-a_n)^\intercal\\&=\sum_{k=1}^n(x_k-a_k)\left.\frac{\partial \mathbf{f}}{\partial x_k}\right|_{x_k=a_k}\end{aligned}$$

则有

$$\begin{aligned}\|\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{x}_0)-\mathbf{L}(\mathbf{x}-\mathbf{x}_0)\|&=\left\|\sum_{k=1}^n\left(\left(\frac{\partial \mathbf{f}}{\partial x_k}\right)|_{x_k=c_k}-\left.\frac{\partial \mathbf{f}}{\partial x_k}\right|_{x_k=a_k}\right)(x_k-a_k)\right\|\\&\leq\sum_{k=1}^n\left\|\left.\frac{\partial \mathbf{f}}{\partial x_k}\right|_{x_k=c_k}-\left.\frac{\partial \mathbf{f}}{\partial x_k}\right|_{x_k=a_k}\right\||x_k-a_k|\\&\leq\sum_{k=1}^n\left\|\left.\frac{\partial \mathbf{f}}{\partial x_k}\right|_{x_k=c_k}-\left.\frac{\partial \mathbf{f}}{\partial x_k}\right|_{x_k=a_k}\right\|\|\mathbf{x}-\mathbf{x}_0\|\end{aligned}$$

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 同都连续，即极限

$$\lim_{x_j\to a_k}\frac{\partial f_j}{\partial x_j}=\left.\frac{\partial f_i}{\partial x_j}\right|_{x_j=a_k},j=1,\cdots,n$$

都存在，故以下极限等式成立：

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

由全微分的定义，命题得证，且线性变换 $\mathbf{L}$ 就是函数的全导数。 □

现在我们引入“线性变换的范”的概念。由定理II.9.4，对任一线性变换 $\mathbf{L}$ 皆存在一个正实数 $k>0$ 满足 $\|\mathbf{Lx}\|\le k\|\mathbf{x}\|,\forall\mathbf{x}\in\mathbb{R}^n$。我们定义——

定义 II.10.4 (线性变换的范)**.** 设 $\mathcal{V},\mathcal{W}$ 是数域 $\mathbb{F}$ 上的赋范向量空间，线性变换 $\mathbf{L}:\mathcal{V}\to\mathbf{W}$ 的范 $\|\mathbf{L}\|$ 为满足 $\|\mathbf{Lx}\|\le k\|\mathbf{x}\|,\forall\mathbf{x}\in\mathbb{R}^n$ 的正实数 k 的最大下界（*infimum*），即

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

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。值得说明的是，线性变换的范的定义依赖它作用下的赋范空间的范的定义。然而，不同的范的具体定义均要满足范的一般要求，故在此意义上范的不同定义间具有等价性（见定理II.9.1）。由此定义我们可立即获得性质：$\|\mathbf{Lx}\|\le\|\mathbf{L}\|\,\|\mathbf{x}\|$。

至此，我们可以用“$\epsilon-\delta$ 语言”定义“函数连续可微即其导函数连续”——

定义 II.10.5 (连续可微)**.** 设 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$ 是可微函数，记其导函数为 $\mathbf{L}:\mathbb{R}^n\to\mathcal{L}(\mathbb{R}^n,\mathbb{R}^m)$，

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

如果对任一正实数 $\epsilon>0$ 均存在正实数 $\delta>0$ 使得只要 $\|\mathbf{x}-\mathbf{x}_0\|<\delta$ 就有 $\|\mathbf{L}(\mathbf{x})-\mathbf{L}(\mathbf{x}_0)\|<\epsilon$，即导函数 $\mathbf{L}(\mathbf{x})$ 在 $\mathbf{x}_0$ 处连续，则称函数 $\mathbf{f}$ 在 $\mathbf{x}_0$ 处连续可微。

定理 II.10.3. 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，则称函数 $\mathbf{f}$ 在 D 内连续可微。

若一个函数在一点处连续，且在该点的邻域内都可微分，该函数在该邻域内的导函数未必都连续。故上一定理给出的是函数连续的充份非必要条件。

刚才在介绍雅可比矩阵的时候，我们利用全微分的性质考虑过以下极限：

$$\lim_{t\to0}\frac{\mathbf{f}(\mathbf{x}_i)-\mathbf{f}(\mathbf{x}_0)}{t},i=1,\cdots,n$$

并知道它就是 $\mathbf{f}(\mathbf{x})$ 在 $\mathbf{x}_0=(x_{01},\cdots,x_{0n})^{\intercal}$ 处对 x_{0i} 的偏导数，因为上式 $\Leftrightarrow$

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

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

定义 II.10.6 (对向量的导数、方向导数). 函数 $\mathbf{f}\colon\mathbb{R}^n\to\mathbb{R}^m c$ 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。

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.10.4. 设函数 $\mathbf{f}\colon\mathbb{R}^n\to\mathbb{R}^m$ 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.10.4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.11 链式法则、反函数定理、隐函数定理

定理 II.11.1 (链式法则). 如果函数 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$ 在 $\mathbf{x}_0\in\mathrm{dom}\mathbf{f}$ 处可微分；函数 $\mathbf{g}:\mathbb{R}^n\to\mathbb{R}^p$ 在 $\mathbf{f}(\mathbf{x}_0)\in\mathrm{dom}\mathbf{g}$ 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. 首先证明 x_0 处于复合函数 $\mathrm{g}\circ\mathrm{f}$ 的定义域内。由于 $\mathrm{f}(\mathrm{x}_0)\in\mathrm{dom}\mathrm{g}$ 且 g 在 $\mathrm{f}(\mathrm{x}_0)$ 处可微分，故总存在正实数 δ' 使得只要 $\|\mathrm{f}(\mathrm{x})-\mathrm{f}(\mathrm{x}_0)\|<\delta'$ 就有 $\mathrm{f}(\mathrm{x})\in\mathrm{dom}\mathrm{g}$。又因为 $\mathrm{x}_0\in\mathrm{dom}\mathrm{f}$ 且 f 在 x_0 处可微分，故总存在正实数 δ 使得只要 $\|\mathrm{x}-\mathrm{x}_0\|<\delta$ 则 $\mathrm{x}\in\mathrm{dom}\mathrm{f}$，同时还必存在 $\delta'>0$ 使得这一 δ 选择下的 x 满足 $\|\mathrm{f}(\mathrm{x})-\mathrm{f}(\mathrm{x}_0)\|<\delta'$。所以任一满足 $\|\mathrm{x}-\mathrm{x}_0\|<\delta$ 的 x 均在复合函数 $\mathrm{g}\circ\mathrm{f}$ 的定义域内。

按照全微分和全导数的定义，由于函数 f 和 g 分别在 x_0 和 $\mathrm{f}(\mathrm{x}_0)$ 可导，故存在函数 z_1、z_2 满足 $\lim_{\mathrm{x}\to\mathrm{x}_0}\mathrm{z}_1(\mathrm{x}-\mathrm{x}_0)=\mathbf{0}$、$\lim_{\mathrm{f}(\mathrm{x})\to\mathrm{f}(\mathrm{x}_0)}\mathrm{z}_2(\mathrm{f}(\mathrm{x})-\mathrm{f}(\mathrm{x}_0))=\mathbf{0}$，且

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

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

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

由三角不等式，又有*

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

*此处用到定理II.9.4。

故

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

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。 □

定理 II.11.2 (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 $\mathbf{f}(\mathbf{N})$ 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. 详见 §VI.2。 □

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.11.1. *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*

$$
\begin{aligned}
&\mathbf{h}(\mathbf{x})=(\mathbf{g}(\mathbf{x}),\mathbf{x})\,\forall\mathbf{x}\in D\\
&h_i(\mathbf{x})=\begin{cases}g_i(\mathbf{x}), & i=1,\cdots,n\\ x_{i-n}, & i=n+1,\cdots,n+m\end{cases}
\end{aligned}
$$

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

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

定理 II.11.3 (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 $\mathbf{F}(\mathbf{x},\mathbf{f}(\mathbf{x})))=\mathbf{0}\forall\mathbf{x}\in N$，且在 N 上有

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

上式中的“-1”是指线性变换的逆。

例 II.11.2. 设函数 $\mathbf{F}:\mathbb{R}^4\to\mathbb{R}^2,\mathbf{F}(u,v,x,y)=(F_1,F_2),\mathbf{F}=\mathbf{0}$ 隐含定义了 $(x,y)=\mathbf{f}(u,v),\mathbf{f}:\mathbf{R}^2\to\mathbf{R}^2$，则

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

II.12 标量与向量场的积分定理

II.13 雷诺传输定理

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[13]。

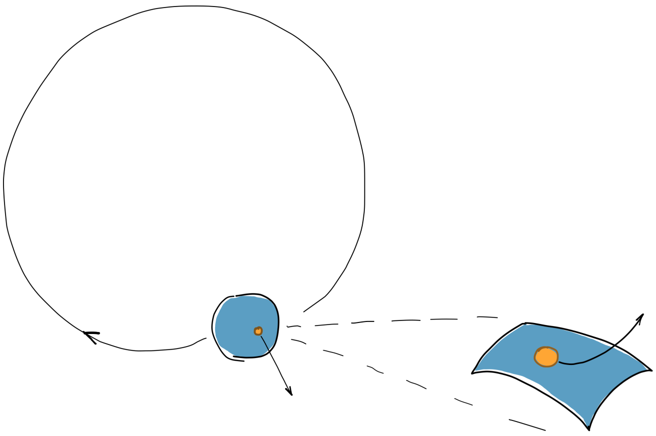

图 III.1.1: 非惯性系下的运动规律。

例 III.1.1 (“我们” vs “球上的人”). 在一个相对我们作匀速（低速）圆周运动的球上站着的人（他看不到我们，但能与我们通讯交流），通过抓稳该球保持与球相对静止。他以球作参照物，研究球上的物体的运动，会得出：保持物体的静止需要一个方向恒定的力，因为球上的物体都需要绑在球上才能保持相对于球静止，他通过弹簧秤可以测量出这一力的大小，进一步发现保持物体静止的力与该物体的质量成正比。如果把某物体松绑，使其不受任何力，物体会开始作一个曲线运动。我们则能不通过测量力就看到球和球上的物体的圆周运动，在我们眼中它们的运动状态是一直在改变中的，而不是静止的；给物体松绑后，物体以松绑时的速度作为初速度作匀速直线运动（如图*III.1.1*所示）。故我们的观察符合“物体在没有外力作用的情况静止或作匀速直线运动”的论断。如果球的匀速圆周运动的角速度发生变化，我们仍然能不通过力的测量而看到这一点，但球上的人只会发现，保持球上物体静止所需的力的大小在莫名其妙地变化。如果他在球上设置一个单摆，他将发现单摆的摆动方向在随时间变化，而我们认为这个单摆的摆动方向没有变化。总而言之，我们认为牛顿力学定律符合实际，但球上的人将不会同意。但他和我们通过通讯至少能确认双方所观察的现象是同一个物理现象，按照

物理客观性原则理应受同一套自然定律统治。摆在他和我们面前的任务只是如何构造一套理论统一地描述好这一自然定律。

在经典力学中，一个物理事件的发生，同时具有“发生的位置”和“发生的时刻”这两个属性。一名观察者使用时钟测量两个物理事件发生的时间间隔，使用直尺来测量两个物理事件发生位置之间的距离。其中距离的测量总是针对同时刻发生的两个事件。

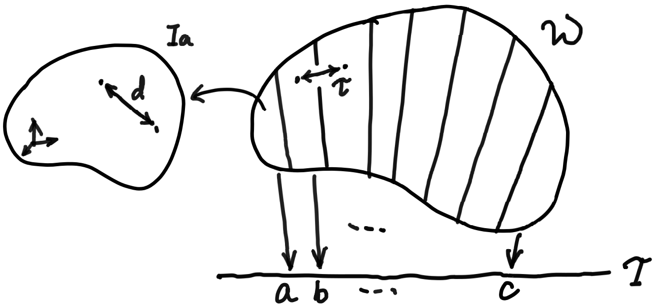

图 III.1.2: 事件世界与同时等价类示意图

引入事件世界 $\mathcal{W}$。

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

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

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 $S = \{(a, b) | \tau(a, b) = 0\}$ 满足自反性、传递性、对称性。这一等价关系将事件世界 $\mathcal{W}$ 划分为等价类，即：

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

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 $\mathcal{T}$ 称为时间，是所有（不同）时刻的集合（如图*III.1.2*所示）。

我们还给予事件世界 $\mathcal{W}$ 以一个度量 $d : I_a \times I_a \to \mathbb{R}, a \in \mathcal{T}$，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 $\mathcal{V}_a$。

- $\forall a \in \mathcal{T} : \dim \mathcal{V}_a = 3$

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.1.2*所示）。

结合度量 τ 的定义，$(\mathcal{T}, |\tau|)$ 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 $\mathcal{T}$ 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 $t_0 \in \mathcal{T}$ 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.7.4）对任一 $\mathcal{T}$ 上的等距变换 $i : \mathcal{T} \to \mathcal{T}$ 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

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

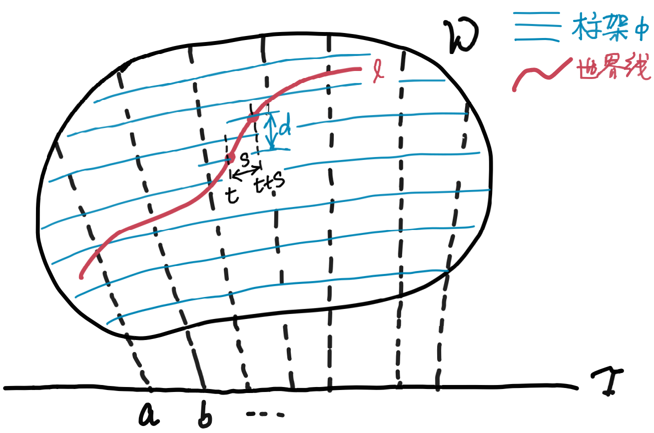

图 III.1.3: 世界线及事件世界的标架示意图

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 $\mathcal{T}$ 到事件世界 $\mathcal{W}$ 的单射 $l : \mathcal{T} \to \mathcal{W}$。

一条世界线 $l(t), t \in \mathcal{T}$ 的任意两个值都是属于不同时刻的事件（如图III.1.3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.1.3*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 $t_1, t_2 \in \mathcal{T}$ 两条世

界线 $l_1, l_2 \in \phi$ 有 $d(l_1(t_1), l_2(t_!)) = d(l_1(t_2), l_2(t_2))$。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

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

使得没有一个事件是不属于任何标架线的。映射 $l_\phi : \mathcal{W} \to \phi, l_\phi(a) = \{l | l \in \phi, a \in \mathrm{ran} l\} \forall a \in \mathcal{W}$ 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 $a \in \mathcal{W}$ 对应且只对应一条标架线，且有 $d(l_\phi(a)(t), l_\phi(b)(t)) =$ 常数$\forall a, b \in \mathcal{W}, t \in \mathcal{T}$，故可由此定义“标架线之间的距离”：

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

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

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 $a \in I_{t_1}, b \in I_{t_2}, t_1, t_2 \in \mathcal{T}$ 之间的距离，选定标架 ϕ，a 与 b 的距离就是 $d(l_\phi(a), l_\phi(b))$。

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”：

$$\lim_{s \to 0} \frac{l(t+s) - l(t)}{s}$$

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（$l(t+s) \in I_{t+s}, l(t) \in I_t, t+s, t \in \mathcal{T}$，如图III.1.3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 $\mathbf{v} : \mathcal{T} \to \mathcal{V}_\phi$：

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

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 $\mathcal{V}_\phi$ 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 $t \in \mathcal{T}$ 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 $\mathcal{V}_t, \mathcal{V}_\phi$ 同构，故 $\dim\mathcal{V}_\phi = \dim\mathcal{V}_t = 3$。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 $\mathcal{V}_{t_1}, \mathcal{V}_{t_2}, \cdots$，且 $\dim\mathcal{V}_{t_1} = \dim\mathcal{V}_{t_2} = \cdots = \dim\mathcal{V}_\phi$，即标架的平移空间 $\mathcal{V}_\phi$ 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 $\mathcal{V}_\phi$ 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。

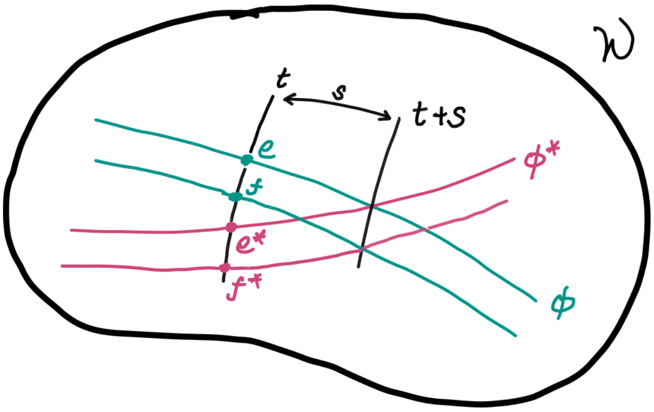

图 III.1.4: 事件世界的标架变换示意图

考虑事件世界 $\mathcal{W}$ 中的两组不同的标架 ϕ, ϕ^*。事件 $e \in I_t$ 的时刻是 $t \in \mathcal{T}$，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.1.4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 $l_{\phi^*}(l_\phi(e)(t))(t) = l_\phi(e)(t) \forall e \in \mathcal{W}, t \in \mathcal{T}$，上式 $\Leftrightarrow$

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

因此，对任意 $e, f \in \mathcal{W}$, $d(l_\phi(e), l_\phi(f)) = d(l_{\phi^*}(e^*), l_{\phi^*}(f^*))$，即由 ϕ 到 ϕ^* 存在一个等距映射。由等距映射的表示定理，选定 ϕ 中的一标架线 l_0，则 $l*(s) = l_0^*(s) + \mathbf{Q}(s)(l - l_0), l \in \phi, l^* \in \phi^*$，其中 $l_0^* = l_{\phi^*}(l_\phi(l_0)(t+s))$。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 $\mathcal{T}$ 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 $\mathcal{V}_\phi$ 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 $\mathcal{T}$ 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 $\mathcal{V}_\phi$ 中的一向量 $\widehat{\mathbf{o}}$ 和一组基 $\{\widehat{\mathbf{e}}_i\}$，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 $\mathcal{W}$ 直接到 $\mathbb{R}^{3+1}$ 的对应过程，即 $\phi : \mathcal{W} \to \mathbb{R}^{3+1}$。这时，标架 ϕ 实际包括了一个观察者对 t_0、$\widehat{\mathbf{o}}$ 和 $\{\widehat{\mathbf{e}}_i\}$ 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 $w \in \mathcal{W}$ 经标架 ϕ 映射为 $(\mathbf{x}, t), \mathbf{x} \in \mathbb{R}^3, t \in \mathbb{R}$，经标架 ϕ^* 映射为 $(\mathbf{x}^*, t^*), \mathbf{x}^* \in \mathbb{R}^3, t^* \in \mathbb{R}$。我们认为 ϕ, ϕ^* 是可逆的，故有 $(\mathbf{x}^*, t^*) = \phi^* \circ \phi^{-1}(\mathbf{x}, t)$。若 w 的时刻是 $a \in \mathcal{T}$，标架 ϕ 选择的时间原点是 t_0，标架 ϕ^* 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 $g : \mathcal{T} \to \mathcal{T}$ 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 ϕ 到标架 ϕ^* 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 ϕ, ϕ^* 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 $\mathbf{A} \in \mathcal{L}(\mathbb{R}^3)$ 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 $\mathcal{B}$，其元素 $X \in \mathcal{B}$ 称为物质点。

定义 III.2.3 (运动学过程、物体的运动)**.** 如图*III.2.1*所示，设 $\mathcal{B}$ 是一个物体，$\mathcal{W}$ 是事件世界，Υ 是时间 $\mathcal{T}$ 的一个连通子集。映射 $k : \mathcal{B} \times \Upsilon \to \mathcal{W}$ 称为一个运动学过程。对某物质点 $X \in \mathcal{B}$，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

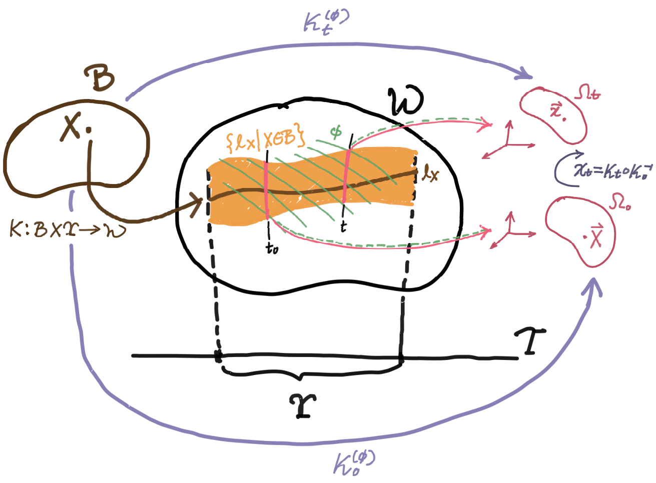

图 III.2.1: 物体的运动、运动学过程、放置、构型

子集，记为 l_X，称为物质点 X 的运动。$\mathcal{B}$ 的所有物质点的同一运动学过程 $\{l_X|X\in\mathcal{B}\}$ 是一个世界运动的子集，记为 $\psi_{\mathcal{B}}$，称为物体 $\mathcal{B}$ 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 $\mathcal{B}$ 的一个物质点 $X\in\mathcal{B}$ 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 $\mathcal{V}_\phi$ 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 $\mathcal{W}$ 是事件世界，$\mathcal{T}$ 是时间，$\mathcal{B}$ 是物体，$\psi_{\mathcal{B}}$ 是其在时间间隔 $\Upsilon\subset\mathcal{T}$ 的运动。选定标架 ϕ，则任一时刻 $a\in\Upsilon$ 下集合 $\Omega_a=\{l_\phi(l_X(a))\}\subset\phi$ 称为 $\mathcal{B}$ 在时间 a 的构型。映射 $\kappa_a^{(\phi)}:\mathcal{B}\to\Omega_a,\kappa_a^{(\phi)}(X)=l_\phi(l_X(a))\forall X\in\mathcal{B}$ 称为物体 $\mathcal{B}$ 在时刻 a 按标架 ϕ 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 $\phi:\mathcal{W}\to\mathbb{R}^{3+1}$，物体 $\mathcal{B}$ 在任一时刻 $t\in\mathbb{R}$ 占据的 $\mathbb{R}^3$ 的区域 $\Omega_t=\{\mathbf{x}|(\mathbf{x},t)=\phi(X,t),X\in\mathcal{B}\}$ 称为 $\mathcal{B}$ 在时刻 t 下的构型。映射 $\kappa_t^{(\phi)}:\mathcal{B}\to\Omega_t,\kappa_t^{(\phi)}(X)=\mathbf{x},X\in\mathcal{B},\mathbf{x}\in\Omega_t$ 称为物体 $\mathcal{B}$ 按标架 ϕ 在时刻 t 的放置。

我们要求 κ 是可逆的。

设某一时刻物体 $\mathcal{B}$ 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 $X \in \mathcal{B}$ 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 $X_0 \in \mathcal{B}$ 在标架 ϕ, ϕ^* 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 ϕ 下，物体 $\mathcal{B}$ 在这两个时刻的构型分别是 Ω_a, Ω_b，则由 Ω_a 到 Ω_b 的映射为 $\kappa_b^{(\phi)} \circ \kappa_a^{(\phi)-1} = \chi_{a \to b}$，称为物体 $\mathcal{B}$ 在标架 ϕ 下的形变*。如果在形变过程中没有发生标架变换，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 $\mathcal{B}$ 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 $X \in \mathcal{B}$

*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。

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 Ω_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 Ω_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 $\mathcal{B}$ 的运动在标架 ϕ 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 $X, Y \in \mathcal{B}$ 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 $\mathcal{B}$ 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 $\mathcal{V}_0, \mathcal{V}_t$ 中的向量。故该导数是由 $\mathcal{V}_0$ 到 $\mathcal{V}_t$ 的线性变换。在选定标架 ϕ 下，则形变梯度张量可表示为形变映射 χ 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 $X \in \mathcal{B}$ 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 $X, Y \in \mathcal{B}$，

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

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 $\mathcal{B}$ 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 χ 的雅可比行列式。

设 (ϕ, ϕ^*) 是一个标架变换，物体 $\mathcal{B}$ 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 $X, X_0 \in \mathcal{B}$ 在参考时刻下的位置向量满足以下标架变换关系：

$$\begin{aligned} \mathbf{X}^* &= \mathbf{X}_0^* + \mathbf{Q}_0\left(\mathbf{X} - \mathbf{X}_0\right) \\ \Leftrightarrow \mathbf{X} &= \mathbf{X}_0 + \mathbf{Q}_0^{\intercal}\left(\mathbf{X}^* - \mathbf{X}_0^*\right) \end{aligned}$$

在当前时刻下的位置向量满足以下标架变换关系：

$$\begin{aligned} \mathbf{x}^* &= \mathbf{x}_0^* + \mathbf{Q}_t\left(\mathbf{x} - \mathbf{x}_0\right) \\ t^* &= t + a, a = t_0^* - t_0 \end{aligned}$$

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

$$\begin{aligned} \mathbf{x} &= \chi\left(\mathbf{X}, t\right) \\ \mathbf{x}_0 &= \chi\left(\mathbf{X}_0, t\right) \\ \mathbf{x}^* &= \chi^*\left(\mathbf{X}^*, t^*\right) \\ \mathbf{x}_0^* &= \chi^*\left(\mathbf{X}_0^*, t^*\right) \end{aligned}$$

代入上面的标架变换关系得

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

标架 ϕ^* 下的形变梯度张量

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

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 $\mathcal{B}$ 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 $\mathcal{B}$ 的每一物质点 $X \in \mathcal{B}$，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 $\mathcal{B}$ 的元素——物质点 $X \in \mathcal{B}$ 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

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

可以是标量，向量或张量。由于人无法直接认识物体 $\mathcal{B}$ 及其物质点 $X \in \mathcal{B}$，故函数 $f(X,t)$ 需要转化为关于物质点 X 在物体 $\mathcal{B}$ 在某标架下的构型中的位置向量的函数。设物体 $\mathcal{B}$ 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

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

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 $\mathcal{B}$ 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

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

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：

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

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 $\chi(\cdot, t), \chi^{-1}(\cdot, t)$。

考虑物体在选定某标架下的运动的速度

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

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 $\kappa(\cdot, t)$。由定义，速度的物质描述是

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

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

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

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 $\mathcal{B}$ 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

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

速度的物质描述 $\mathbf{a}_{\mathrm{m}}(\mathbf{X}, t) = \mathbf{a}\left(\kappa_0^{-1}(\mathbf{X}), t\right) = \frac{\partial^2}{\partial t}\chi(\mathbf{X}, t) = \frac{\partial}{\partial t}\mathbf{v}_{\mathrm{m}}(\mathbf{X}, t)$。如果我们只知道速度场 $\mathbf{v}_{\mathrm{s}}(\mathbf{r}, t)$，要求加速度的物质描述，就需要用 $\mathbf{v}_{\mathrm{s}}(\mathbf{r}, t)$ 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.11.1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 $\mathcal{B}$ 的运动造成的（即 f_{s} 是该物体性质的空间描述），那么，对 $f_{\mathrm{s}}(\mathbf{r},t)$ 进行 r 以下计算

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

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。

例 III.4.1. 设物体 $\mathcal{B}$ 的运动在某标架下满足

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

则速度的物质描述

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

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

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

故

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

加速度的物质描述

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

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

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

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

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

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.10.4，$\mathbf{L}\left(\mathbf{r},t\right)$ 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 $\frac{\partial}{\partial\mathbf{u}}\mathbf{v}_s\left(\mathbf{r},t\right)$。

III.6　应力张量

III.7　守恒律

III.8　Navier–Stokes 方程

III.9　流变测量学

第四部分 线性粘弹性本构方程

IV.1 线性粘弹性本构方程的建立

IV.2 线性粘弹性本构关系的一般预测

IV.3 记忆函数的具体形式

IV.4 松弛时间谱

第五部分　非线性粘弹性本构方程概览

V.1　非线性粘弹性本构方程的构建原则

V.2　准线性粘弹性

V.3　广义牛顿流体

V.4　微分形本构方程

V.5　积分型本构方程

V.6　屈服应力流体

第六部分 附录

VI.1 实数集的拓扑概念

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[2]§7.1,p.1。

定义 VI.1.1 (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 VI.1.2 (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 VI.1.3 (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 ∂S。

定理 VI.1.1. 一个集合是闭集当且仅当它包含所有其边界上的点。

推论 VI.1.1.1. 一个集合是闭集当且仅当它包含其所有极限点。

推论 VI.1.1.2. 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 VI.1.4 (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 VI.1.1. 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。

VI.2 反函数定理的证明

本节我们将证明反函数定理。

引理 VI.2.1. 线性变换 $\mathbf{L}:\mathcal{V}_n\to\mathcal{W}_m$ 是单射当且仅当存在正实数 $m>0$ 满足 $\|\mathbf{Lx}\|\geq m\|\mathbf{x}\|\,\forall\mathbf{x}\in\mathcal{V}_n$。

证明. 如果 $\mathbf{L}$ 不是单射，则存在 $\mathbf{x}_0\neq\mathbf{0}$ 满足 $\mathbf{Lx}_0=\mathbf{0}$，即有 $\|\mathbf{Lx}_0\|=0<m\|\mathbf{x}\|$。故其逆否命题成立。

如果 $\mathbf{L}$ 是单射，则其存在逆 $\mathbf{L}^{-1}$ 满足 $\mathbf{L}^{-1}\mathbf{L}=\mathbf{I}$ 且 $\mathbf{L}^{-1}$ 也是线性变换。由定理II.9.4，$\exists k>0,\|\mathbf{L}^{-1}\mathbf{y}\|\leq k\|\mathbf{y}\|\,\forall\mathbf{y}\in\mathcal{W}_m$。令 $m=1/k$ 则有 $m\|\mathbf{x}\|=m\|\mathbf{L}^{-1}\mathbf{Lx}\|\leq mk\|\mathbf{Lx}\|=\|\mathbf{Lx}\|$ □

在之前的定义中，函数连续可微定义为其自然基下的所有偏导数皆连续。这里我们通过极限的 $\epsilon-\delta$ 语言重新定义一个向量函数连续可微的含义。

首先讨论“导函数”的正式定义。设 $\mathbf{f}:\mathbb{R}^n\to\mathbb{R}^m$ 是可微函数，记其导函数为 $\mathbf{L}:\mathbb{R}^n\to\mathcal{L}(\mathbb{R}^n,\mathbb{R}^m)$，

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

注意，$\mathbf{L}(\mathbf{x})$ 的函数值是一个线性变换，但未必是一个线性函数（即 $\mathbf{L}$ 未必是一个从 $\mathbb{R}^n$ 到 $\mathcal{L}(\mathbb{R}^n,\mathbb{R}^m)$ 的线性变换）。

引理 VI.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

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

证明. 记函数 $\mathbf{f}$ 的导函数为 $\mathbf{L}$。由引理VI.2.1，存在 $m>0$ 使得 $\|\mathbf{L}(\mathbf{x}_0)\mathbf{y}\|\geq m\|\mathbf{y}\|\,\forall\mathbf{y}\in\mathbb{R}^n$。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 $m/2>0$——存在 $\delta>0$ 使得只要 $\|\mathbf{x}-\mathbf{x}_0\|<\delta$ 就有 $\|\mathbf{L}(\mathbf{x})-\mathbf{L}(\mathbf{x}_0)\|\leq m/2$。由线性变换的范的定义，有不等式

$$\|(\mathbf{L}(\mathbf{x})-\mathbf{L}(\mathbf{x}_0))\mathbf{y}\|\leq\|\mathbf{L}(\mathbf{x})-\mathbf{L}(\mathbf{x}_0)\|\,\|\mathbf{y}\|\leq\frac{m}{2}\|\mathbf{y}\|$$

由三角不等式又有

$$\|\mathbf{L}(\mathbf{x}_0)\mathbf{y}\|-\|(\mathbf{L}(\mathbf{x})-\mathbf{L}(\mathbf{x}_0))\mathbf{y}\|\leq\|\mathbf{L}(\mathbf{x})\|$$

不等号左边可代入上面两个不等式关系，即

$$\|\mathbf{L}(\mathbf{x}_0)\mathbf{y}\|-\|(\mathbf{L}(\mathbf{x})-\mathbf{L}(\mathbf{x}_0))\mathbf{y}\|\geq m\|\mathbf{y}\|-\frac{m}{2}\|\mathbf{y}\|=\frac{m}{2}\|\mathbf{y}\|$$

故存在 $M=m/2>0$ 满足命题。 □

引理 VI.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 $\|\mathbf{f}(\mathbf{x}')-\mathbf{f}(\mathbf{x})\|\le M\|\mathbf{x}'-\mathbf{x}\|$。

证明. 记函数 $\mathbf{f}$ 的导函数为 $\mathbf{L}$。由引理VI.2.1，存在 $m>0$ 使得 $\|\mathbf{L}(\mathbf{x}_0)\mathbf{y}\|\ge m\|\mathbf{y}\|\,\forall\mathbf{y}\in\mathbb{R}^n$。

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 $\|\mathbf{L}(\mathbf{x})-\mathbf{L}(\mathbf{x}_0)\|\le m/(2\sqrt{n})$。

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 $0\le t\le 1$ 有

$$\begin{aligned}\|\mathbf{x}+t\mathbf{z}-\mathbf{x}_0\|&=\|t\mathbf{x}'+(1-t)\mathbf{x}-\mathbf{x}_0\|\\&=\|t(\mathbf{x}'-\mathbf{x})+(1-t)(\mathbf{x}-\mathbf{x}_0)\|\\&\le t\|\mathbf{x}'-\mathbf{x}_0\|+(1-t)\|\mathbf{x}-\mathbf{x}_0\|<t\delta+(1-t)\delta=\delta\end{aligned}$$

上述推导结论在几何上的意义是，只要点 $\mathbf{x}',\mathbf{x}$ 在由 $\|\mathbf{x}_0\|<\delta$ 的开集内部，则它们的连线上的点 $\mathbf{x}+t\mathbf{z}$ 都在此开集内部，或称"δ-球是凸的"。由于导函数连续是在整个 δ-球内都成立的，因此对由 $0\le t\le 1$ 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 $\|(\mathbf{L}(\mathbf{x}+t\mathbf{z})-\mathbf{L}(\mathbf{x}_0))\mathbf{y}\|\le\|\mathbf{L}(\mathbf{x}+t\mathbf{z})-\mathbf{L}(\mathbf{x}_0)|\,\|\mathbf{y}\|<M\|\mathbf{y}\|\,\forall\mathbf{y}\in\mathbb{R}^n$。

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 $g_k(t)=\pi_k(\mathbf{f}(\mathbf{x}+t\mathbf{z})),0\le t\le 1$，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

$$\begin{aligned}\|\mathbf{f}(\mathbf{x}')-\mathbf{f}(\mathbf{x})\|&\ge|\pi_k(\mathbf{f}(\mathbf{x}')-\mathbf{f}(\mathbf{x}))|\\&=|\pi_k(\mathbf{f}(\mathbf{x}'))-\pi_k(\mathbf{f}(\mathbf{x}))|\\&=|\pi_k(\mathbf{L}(\mathbf{x}+t_k\mathbf{z})\mathbf{z})|\end{aligned}$$

另有以下三角不等式成立：

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

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 VI.2.1 (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VI.2.2和VI.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VI.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VI.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VI.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VI.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 δ_1。

如果 $\widetilde{M}$ 存在，则对任一 $\mathbf{y}\in C$，我们可以从 B 中找到一个 $\mathbf{x}'$ 使得 $\mathbf{y}'=\mathbf{f}(\mathbf{x}')$ 到 $\mathbf{y}\in C$ 的距离最短，并由引理VI.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 $C=\left\{\mathbf{y}|\,\|\mathbf{y}-\mathbf{y}_1\|<\widetilde{M}\right\}$ 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VI.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

$$\lim_{\mathbf{f}(\mathbf{x}')\to\mathbf{f}(\mathbf{x})} \frac{\mathbf{x}'-\mathbf{x}-\mathbf{L}^{-1}(x)(\mathbf{f}(\mathbf{x}')-\mathbf{f}(\mathbf{x}))}{\|\mathbf{f}(\mathbf{x}')-\mathbf{f}(\mathbf{x})\|} = \mathbf{0}$$

令未求极限前的比增量为 $\mathbf{s}$，即

$$\mathbf{s} = \frac{\mathbf{x}'-\mathbf{x}-\mathbf{L}^{-1}(x)(\mathbf{f}(\mathbf{x}')-\mathbf{f}(\mathbf{x}))}{\|\mathbf{f}(\mathbf{x}')-\mathbf{f}(\mathbf{x})\|}$$

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VI.2.3，存在足够小正实数 $\mathbf{M}$ 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.9.4的推论，线性变换都是连续函数，故由复合函数的连续性，极限 $\lim_{\mathbf{x}'\to\mathbf{x}}\mathbf{L}^{-1}(\mathbf{x})\,\mathbf{r}=\mathbf{L}^{-1}(\mathbf{x})\lim_{\mathbf{x}'\to\mathbf{x}}\mathbf{r}=\mathbf{0}$。

由于一个有界函数与一个有极限的函数的积的极限等于那个有极限的函数的极限*，故 $\lim_{\mathbf{x}'\to\mathbf{x}}\mathbf{s}=\mathbf{0}$。又由于当 $\mathbf{x}'\to\mathbf{x}$ 时 $\mathbf{f}(\mathbf{x}')\to\mathbf{f}(\mathbf{x})$，由 $\mathbf{s}$ 的形式有 $\lim_{\mathbf{x}'\to\mathbf{x}}\mathbf{s}=\mathbf{0}\Leftrightarrow\lim_{\mathbf{f}(\mathbf{x}')\to\mathbf{f}(\mathbf{x})}\mathbf{s}=\mathbf{0}$。III证毕。

IV的证明：要证 $\mathbf{f}^{-1}$ 的导函数连续，即对任一 $\mathbf{x}_1\in N,\mathbf{y}_1=\mathbf{f}(\mathbf{x}_1)$ 有

$$\lim_{\mathbf{y}\to\mathbf{y}_1}\left.\frac{d\mathbf{f}^{-1}(\mathbf{y})}{d\mathbf{y}}\right|_{\mathbf{y}=\mathbf{y}}=\left.\frac{d\mathbf{f}^{-1}(\mathbf{y})}{d\mathbf{y}}\right|_{\mathbf{y}=\mathbf{y}_1}$$

由III的证明我们已经有

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

故只需证

$$\lim_{\mathbf{x}\to\mathbf{x}_1}\mathbf{L}^{-1}(\mathbf{x})=\mathbf{L}^{-1}(\mathbf{x}_1)$$

由引理VI.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 δ 定义的 $\mathbf{x}_1$ 的邻域 $N_1=\{\mathbf{x}|\,\|\mathbf{x}-\mathbf{x}_1\|\}<\delta$ 在 N 的内部，则对任一 $\mathbf{x}\in N_1$，以下不等式成立

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

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

由线性变换的范的定义（最大下界），上述不等式 $\Leftrightarrow$

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

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 $\left\|\mathbf{L}^{-1}\left(\mathbf{x}\right)-\mathbf{L}^{-1}\left(\mathbf{x}_1\right)\right\| < \epsilon$。具体地，这个 δ 总存在是由于 M 总存在。这相当于说 $\lim_{\mathbf{x}\to\mathbf{x}_1}\mathbf{L}^{-1}\left(\mathbf{x}\right)=\mathbf{L}^{-1}\left(\mathbf{x}_1\right)$，IV证毕。 $\square$

参考文献

[1] 邓文基. 大学物理（上册）[M]. 华南理工大学出版社, 2009.

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

[3] 邓文基. 大学物理（下册）[M]. 华南理工大学出版社, 2009.

[4] 钟理, 伍钦, 马四朋. 化工原理（上册）[M]. 化学工业出版社, 2008.

[5] 何曼君, 张红东, 陈维孝, 等. 高分子物理（第三版）[M]. 复旦大学出版社, 2007.

[6] ZHENG R, PHAN-THIEN N, TANNER R I. The flow past a sphere in a cylindrical tube: effects of intertia, shear-thinning and elasticity[J]. Rheologica Acta, 1991, 30(6): 499-510.

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

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

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

[10] 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.

[11] NOLL W. The Foundations of Mechanics[G]//GRIOLI G, TRUESDELL C. Non-linear Continuum Theories. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011: 159-200.

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

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