课程概况

教师 梁鑫

办公室 近春园西楼 132 室

Email liangxinslm@tsinghua.edu.cn

课程 高等线性代数选讲

时间地点 周二 08:00–09:35 五教 5105

答疑时间 周三 19:20–20:55

参考书

- 蓝以中编著，高等代数简明教程，下册，第二版，2007 年.
- A.I. Kostrikin and Yu.I. Manin, Linear Algebra and Geometry, translated by M.E. Alferieff, Gordan and Breach Science Publishers, 1996.
- Ray M. Bowen and C.-C. Wang, Introduction to Vectors & Tensors, 2nd Ed., Dover, 2008.

目　　录

第一部分　线性代数续编 **1**

第 9 章　商空间和同构 **2**

第 10 章　度量空间和微积分 **3**

10.1 引子 3
10.2 矩阵序列和矩阵级数 6
10.2.1 极限与连续 6
10.2.2 矩阵函数 10
10.3 矩阵微积分 14
10.3.1 一元微积分 14
10.3.2 多元微分 18

第 11 章　多重线性代数 **24**

11.1 双线性函数 24
11.1.1 双线性函数空间 25
11.1.2 合同标准形 29
11.2 多线性函数 31
11.2.1 基本概念 31
11.2.2 对偶空间 35
11.2.3 多线性映射 38
11.3 张量 39
11.3.1 张量积和张量 39
11.3.2 张量及其运算的向量化表示 43
11.3.3 协变张量和逆变张量 46
11.4 反对称张量和外积 48

第二部分　线性几何 **56**

概述 **57**

第 12 章　仿射空间 **58**

12.1 基本概念 58

12.2 欧氏几何 58

12.3 仿射几何 58

第 13 章　射影空间 **59**

13.1 基本概念 59

13.2 射影几何 59

13.3 椭圆几何 59

第 14 章　闵氏空间和辛空间 **60**

14.1 闵氏空间和双曲几何 60

14.2 辛空间和辛几何 60

小结 **61**

第一部分

线性代数续编

第 9 章　商空间和同构

第 10 章　度量空间和微积分

本章重点考虑向量和矩阵与微积分的结合，我们将会看到两种强有力工具的融合带来的巨大影响.

10.1　引子

微积分学的重心是函数的微分学和积分学，其理论基础是序列与函数的极限. 而极限的严格定义本质上依赖于集合中元素的距离或者粗略说是元素的远近.

在实数集 $\mathbb{R}$ 和复数集 $\mathbb{C}$ 上，元素的距离用绝对值来定义：$\operatorname{dist}(a,b) := |a-b|$. 在第 3 章中，我们为数组向量空间定义了距离：$\operatorname{dist}(\boldsymbol{a},\boldsymbol{b}) := \|\boldsymbol{a}-\boldsymbol{b}\|$. 而在第 6 章中，我们为矩阵定义了范数来衡量不同矩阵的差别：$\operatorname{dist}(A,B) := \|A-B\|$ 或 $\|A-B\|_{\mathrm{F}}$.

事实上，这些都是度量空间的实例.

定义 10.1.1 (度量空间)　*给定集合 $\mathcal{M}$ 及其上的二元函数 $\operatorname{dist}\colon (x,y)\mapsto \operatorname{dist}(x,y)$，若函数 $\operatorname{dist}(\cdot,\cdot)$ 满足如下三条：*

1. *正定性：$\operatorname{dist}(x,y)\geqslant 0$，且 $\operatorname{dist}(x,y)=0 \Leftrightarrow x=y$；*
2. *对称性：$\operatorname{dist}(x,y)=\operatorname{dist}(y,x)$；*
3. *三角不等式：$\operatorname{dist}(x,z)\leqslant \operatorname{dist}(x,y)+\operatorname{dist}(y,z)$，*

*则称 $\operatorname{dist}(\cdot,\cdot)$ 是集合 $\mathcal{M}$ 上的一个***距离***或***度量***，称集合 $\mathcal{M}$ 是一个***度量空间***.*

*度量空间的子集关于该度量构成一个度量空间，称为***度量子空间***.*

因此，数组向量空间 $\mathbb{F}^n$ 和矩阵空间 $\mathbb{F}^{m\times n}$ 都是度量空间. 但度量空间不局限于这两种简单情形.

例 10.1.2 (Grassmann 流形)　线性空间 $\mathbb{F}^n$ 的所有 k 维子空间组成的集合是度量空间，记为 $\operatorname{Gr}_k(\mathbb{F}^n)$，常称为一个 **Grassmann 流形**.

线性空间 $\mathbb{F}^n$ 的所有子空间组成的集合也是度量空间，记为 $\operatorname{Gr}(\mathbb{F}^n) := \cup_{k=0}^{n}\operatorname{Gr}_k(\mathbb{F}^n)$.

下面说明后者成立，由于前者是后者的子集，如果后者是度量空间，则前者自然是一个度量子空间.

回忆第 3 章（或第 8 章）中的正交投影. 对任意 $\mathbb{F}^n$ 的子空间 $\mathcal{M}$，都存在唯一的 n 阶正交投影矩阵 $P_{\mathcal{M}}$；反之，对任意 n 阶正交投影矩阵 P，都存在 $\mathbb{F}^n$ 的子空间 $\mathcal{M}$，使得 $P = P_{\mathcal{M}}$. 这就说明 $\mathrm{Gr}(\mathbb{F}^n)$ 与正交投影矩阵的全体组成的集合 $\mathbb{F}_P^{n\times n}$ 之间有一个双射. 而 $\mathbb{F}_P^{n\times n}$ 是度量空间 $\mathbb{F}^{n\times n}$ 的子集，因而是度量空间. 这就提示我们可以利用双射来定义 $\mathrm{Gr}(\mathbb{F}^n)$ 中元素之间的距离：

$$\mathrm{dist}(\mathcal{M},\mathcal{N}) = \|P_{\mathcal{M}} - P_{\mathcal{N}}\|.$$

可以验证上述二元函数确实是 $\mathrm{Gr}(\mathbb{F}^n)$ 上的一个度量，从而 $\mathrm{Gr}(\mathbb{F}^n)$ 是度量空间：

1. 正定性：$\mathrm{dist}(\mathcal{M},\mathcal{N}) = \|P_{\mathcal{M}} - P_{\mathcal{N}}\| \geqslant 0$，且 $\mathrm{dist}(\mathcal{M},\mathcal{N}) = 0 \Leftrightarrow \|P_{\mathcal{M}} - P_{\mathcal{N}}\| = 0 \Leftrightarrow P_{\mathcal{M}} = P_{\mathcal{N}} \Leftrightarrow \mathcal{M} = \mathcal{N}$；
2. 对称性：$\mathrm{dist}(\mathcal{M},\mathcal{N}) = \|P_{\mathcal{M}} - P_{\mathcal{N}}\| = \|P_{\mathcal{N}} - P_{\mathcal{M}}\| = \mathrm{dist}(\mathcal{N},\mathcal{M})$；
3. 三角不等式：$\mathrm{dist}(\mathcal{M},\mathcal{L}) = \|P_{\mathcal{M}} - P_{\mathcal{L}}\| \leqslant \|P_{\mathcal{M}} - P_{\mathcal{N}}\| + \|P_{\mathcal{N}} - P_{\mathcal{L}}\| = \mathrm{dist}(\mathcal{M},\mathcal{N}) + \mathrm{dist}(\mathcal{N},\mathcal{L})$. ☺

度量空间上容易定义序列的极限和映射的极限.

定义 10.1.3 (极限) *给定度量空间 $\mathcal{M}$ 及其上度量 $\mathrm{dist}(\cdot,\cdot)$. 对度量空间 $\mathcal{M}$ 上的序列 $\{a_k\}, a_k \in \mathcal{M}$，如果存在 $a \in \mathcal{M}$，使得 $\lim_{k\to\infty} \mathrm{dist}(a_k, a) = 0$，则称序列 $\{a_k\}$* **收敛**到 *a，又称 a 是序列 $\{a_k\}$ 的***极限**，*记作 $\lim_{k\to\infty} a_k = a$，或 $k \to \infty$ 时，$a_k \to a$.*

给定度量空间 $\mathcal{M},\mathcal{N}$ 及分别的度量 $\mathrm{dist}_{\mathcal{M}}(\cdot,\cdot), \mathrm{dist}_{\mathcal{N}}(\cdot,\cdot)$. 对 $\mathcal{M}$ 的子集 S 到 $\mathcal{N}$ 的映射 $f\colon S \to \mathcal{N}$，如果存在 $y_* \in \mathcal{N}$，使得 $\mathrm{dist}_{\mathcal{M}}(x, x_*) \to 0 \Rightarrow \mathrm{dist}_{\mathcal{N}}(f(x), y_*) \to 0$，则称 $x \to x_*$ 时，$f(x)$ **收敛**到 *y_*，又称 y_* 是 $x \to x_*$ 时 $f(x)$ 的***极限**，*记作 $\lim_{x\to x_*} f(x) = y_*$，或 $x \to x_*$ 时，$f(x) \to y_*$.*

当然，距离还使我们可以定义邻域：

$$B(x_*, \delta) := \{x \in \mathcal{M} \mid \mathrm{dist}(x, x_*) < \delta\}.$$

从而极限也可以用 ε-N 或 ε-δ 语言来描述，不赘言.

和一元微积分一样，有了极限，我们就可以定义开子集、闭子集、稠密子集、连续性等一系列微积分中的基本概念.

定义 10.1.4 (开、闭、稠密) *给定度量空间 $\mathcal{M}$ 及其上度量 $\mathrm{dist}(\cdot,\cdot)$. 对子集 $S \subseteq \mathcal{M}$：*

1. *若对任意 $x \in S$，都存在 $\delta > 0$，使得 $B(x,\delta) \subseteq S$，则称 S 是 $\mathcal{M}$ 的***开子集**.

2. 若对任意 $\mathcal{M}$ 中有极限的序列 $\{x_k\}, x_k \to x$，都有 $x_k \in S \Rightarrow x \in S$，则称 S 是 $\mathcal{M}$ 的**闭子集**.
3. 若对任意 $x \in \mathcal{M}$，都存在 S 中的序列 $\{x_k\}, x_k \in S$，使得 $x_k \to x$，则称 S 是 $\mathcal{M}$ 的**稠密子集**.

定义 10.1.5 (连续) 对度量空间 $\mathcal{M}$ 的子集 S 到 $\mathcal{N}$ 间定义的映射 $f\colon S \to \mathcal{N}$，若在 $x_* \in S$ 的邻域上，有 $x \to x_*$ 时，$f(x) \to f(x_*)$，则称 f 在 x_* 处**连续**.

若 f 在子集 $S \subseteq \mathcal{M}$ 的每一点处连续，则称 f 在子集 S 上**连续**.

对于度量空间上的可微性及其关联概念，由于需要加法，因此我们假定涉及的度量空间也是线性空间.

定义 10.1.6 (收敛级数) 对度量线性空间 $\mathcal{M}$ 上的序列 $\{a_k\}, a_k \in \mathcal{M}$，如果其**部分和**$S_n := \sum_{k=1}^{n} a_k$ 收敛到 S，则称级数 $\sum_{k=1}^{\infty} a_k$ **收敛**到 S，记作 $\sum_{k=1}^{\infty} a_k = S$.

定义 10.1.7 (可微) 对度量线性空间 $\mathcal{M}$ 的子集 S 到 $\mathcal{N}$ 的映射 $f\colon S \to \mathcal{N}$，若在 $x_* \in S$ 的某个邻域 $B(x_*, \delta)$ 上，存在线性映射 $g_{x_*}\colon B(x_*, \delta) \to \mathcal{N}$，使得

$$x \to x_* \text{时，} \frac{\operatorname{dist}_{\mathcal{N}}(f(x), f(x_*) + g_{x_*}(x - x_*))}{\operatorname{dist}_{\mathcal{M}}(x, x_*)} \to 0,$$

则称 f 在 x_* 处**可微**[1]，而 $g_{x_*}(x - x_*)$ 称为 f 在 x_* 处的**微分**，简记为 $\mathrm{d}f(x_*) = \mathrm{d}f\,|_{x=x_*} = g_{x_*}(\mathrm{d}x)$. 线性映射 $g_{x_*}\colon B(x_*, \delta) \to \mathcal{N}$ 称为 f 在 x_* 处的**导数**.

若 f 在开子集 $S \subseteq \mathcal{M}$ 的每一点处可微，则称 f 在子集 S 上**可微**，映射 $g\colon S \to \mathcal{N}, x_* \mapsto g_{x_*}$ 称为 f 在子集 S 上的**导映射**，一般也直接称为 f 在子集 S 上的**导数**，记为 $\frac{\mathrm{d}f}{\mathrm{d}x}$ 或 $f^{(1)}$.

若 f 在 x_* 处可微，且导数在 x_* 处连续，则称 f 在 x_* 处**连续可微**，或称 f 在 x_* 处是 C^1 类映射.

而其他概念，如高阶可微性、光滑性、解析性、可积性等，在一般情形下需要利用第 11 章中讲述的张量来完整描述. 下面只描述一个简单情形，即定义域是 $\mathbb{F}$ 的子集的情形.

定义 10.1.8 (一元高阶可微) 若 f 在 x_* 处可微，且导数在 x_* 处连续，则称 f 在 x_* 处**连续可微**，或称 f 在 x_* 处是 C^1 类映射.

若 f 在 x_* 处连续可微，其导数在 x_* 处仍连续可微，则称 f 在 x_* 处**二阶连续可微**，或称 f 在 x_* 处是 C^2 类映射，而导数的导数称为**二阶导数**，记为 $f^{(2)}$.

---

[1] 严格来说，应称为 f 在 x_* 处 Frechét 可微. 类似地，下文的微分和导数应称为 Frechét 微分和 Frechét 导数. 本书中仅使用这一种可微、微分、导数的定义，故省略.

类似地，对任意正整数 k，可以定义 k 阶连续可微、C^k 类映射以及 k 阶导数 $f^{(k)}$.

特别地，若 f 在 x_* 处无穷阶连续可微，则称 f 在 x_* 处**光滑**，或称 f 在 x_* 处是 C^∞ 类映射.

若 f 在 x_* 处光滑，且在 x_* 的邻域上满足 $f(x_*+d)=f(x_*)+\sum_{k=1}^{\infty}\frac{1}{k!}f_{x_*}^{(k)}d^k$，则称 f 在 x_* 处**解析**，或称 f 在 x_* 处是 C^ω 类映射.

类似地，可以定义 f 在 S 上连续可微、k 阶连续可微、光滑、解析.

定义 10.1.9 (一元可积) 给定度量线性空间 $\mathcal{N}$ 与区间 $[a,b]\subset\mathbb{R}$ 和映射 $f\colon[a,b]\to\mathcal{N}$，对区间 $[a,b]$ 的任意分割 $a=x_0<x_1<\cdots<x_{n-1}<x_n=b$ 中的任意节点 $x_i<y_i\leqslant x_{i+1}$，称 $\sum_{i=1}^{n}(x_i-x_{i-1})f(y_i)$ 为 f 的对应分割对应节点的 ***Riemann* 和**. 若 $\max\{x_i-x_{i-1}\mid i=1,\cdots,n\}\to 0$ 时，f 的 *Riemann* 和收敛，则称 f 在区间 $[a,b]$ 上**可积**[1]，称其极限为 f 在区间 $[a,b]$ 上的**积分**，记为 $\int_a^b f(t)\,\mathrm{d}t$ 或 $\int_{[a,b]} f(t)\,\mathrm{d}t$.

定义无穷区间上的积分为

- $\int_a^{+\infty} f(t)\,\mathrm{d}t=\lim_{b\to+\infty}\int_a^b f(t)\,\mathrm{d}t$;
- $\int_{-\infty}^{b} f(t)\,\mathrm{d}t=\lim_{a\to-\infty}\int_a^b f(t)\,\mathrm{d}t$;
- $\int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t=\lim_{a\to-\infty}\int_a^0 f(t)\,\mathrm{d}t+\lim_{b\to+\infty}\int_0^b f(t)\,\mathrm{d}t$.

自变量为复数的映射在路径上的积分也可以类似定义.

为简化讨论，**本章中的数域 $\mathbb{F}$ 特指 $\mathbb{R}$ 或 $\mathbb{C}$**. 同时，**本章将出现的度量空间都是数组向量空间或矩阵空间**. 下文提到这两类度量空间时，一般都默认其度量为 $\|\cdot-\cdot\|$.

10.2 矩阵序列和矩阵级数

10.2.1 极限与连续

命题 10.2.1 对矩阵序列 $A_k=\left[a_{ij}^{(k)}\right]$ 和矩阵 $A=\left[a_{ij}\right]$，$A_k\to A$ 当且仅当矩阵序列**逐元素**收敛，即对任意 i,j，$a_{ij}^{(k)}\to a_{ij}$.

证 先来说明在度量 $\|\cdot-\cdot\|_{\mathrm{F}}$ 下，二者等价.

$$A_k\to A\Leftrightarrow \operatorname{dist}(A_k,A)=\|A_k-A\|_{\mathrm{F}}=\sqrt{\sum_{i,j}\left|a_{ij}^{(k)}-a_{ij}\right|^2}\to 0\Leftrightarrow \forall i,j,\left|a_{ij}^{(k)}-a_{ij}\right|\to 0.$$

① 严格来说，应称为 f 在区间 $[a,b]$ 上 Riemann 可积. 类似地，下文的积分应称为 Riemann 积分. 本书中仅使用这一种可积的定义，故省略.

另一方面，对任意 B，$\|B\| \leqslant \|B\|_{\mathrm{F}} \leqslant n\|B\|$，不难说明，在度量 $\|\cdot - \cdot\|$ 下，二者也等价. $\square$

这就是在前面多个章节中直接认定向量或矩阵的极限就是元素极限组成的向量或矩阵的原因.

容易得到如下结论.

命题 10.2.2 *1.* $\lim_{k\to\infty}(pA_k + qB_k) = p\lim_{k\to\infty} A_k + q\lim_{k\to\infty} B_k$；

2. $\lim_{k\to\infty} A_kB_k = \lim_{k\to\infty} A_k \lim_{k\to\infty} B_k$；

3. $\lim_{k\to\infty} A_k^{\mathrm{T}} = (\lim_{k\to\infty} A_k)^{\mathrm{T}}$；

4. $\lim_{k\to\infty} A_k^{\mathrm{H}} = (\lim_{k\to\infty} A_k)^{\mathrm{H}}$；

证 **留为作业**. $\square$

由于方阵有乘法和幂，因此可以直接考虑方阵的幂的极限.

定理 10.2.3 对矩阵 $A \in \mathbb{C}^{n\times n}$，

1. 当 $k\to\infty$ 时，A^k 收敛，当且仅当 A 只有两种特征值，一是绝对值小于 1 的特征值，二是半单的特征值 1；

2. 当 $k\to\infty$ 时，$A^k \to O$，当且仅当 A 的任意特征值 λ 都满足 $|\lambda| < 1$.

证 设 A 的 Jordan 分解为 $A = XJX^{-1}$，则 $A^k = XJ^kX^{-1}$. 根据命题 10.2.2 ，A^k 收敛当且仅当 J^k 收敛.

先看 J 只含一个 Jordan 块的情形. 对 $J = \lambda I + J_0 = \begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}$，$J^k = (\lambda I + J_0)^k = \sum_{j=0}^{k} \mathrm{C}_k^j \lambda^{k-j} J_0^j$. 而当 $j \geqslant n$ 时，$J_0^j = 0$，因此 $J^k = \sum_{j=0}^{n} \mathrm{C}_k^j \lambda^{k-j} J_0^j = \begin{bmatrix} \lambda^k & \mathrm{C}_k^1\lambda^{k-1} & \cdots & \mathrm{C}_k^n\lambda^{k-n} \\ & \ddots & \ddots & \vdots \\ & & \lambda^k & \mathrm{C}_k^1\lambda^{k-1} \\ & & & \lambda^k \end{bmatrix}$. 分情形讨论：

1. 当 $|\lambda| < 1$ 时，任意元素都有 $\left|\mathrm{C}_k^j\lambda^{k-j}\right| \leqslant \frac{k^j}{j!}|\lambda|^{k-j} \to O$，即 $J^k \to O$.
2. 当 $|\lambda| > 1$ 时，对角元素 λ^k 不收敛，因此 J^k 不收敛.
3. 当 $\lambda \neq 1, |\lambda| = 1$ 时，对角元素 λ^k 不收敛，因此 J^k 不收敛.
4. 当 $\lambda = 1$ 时，若 Jordan 块阶大于 1，则次对角元素 $k\lambda^{k-1} = k$ 不收敛，因此 J^k 不收敛；若 Jordan 块阶为 1 时，则 $J^k = 1$.

综上，$J^k \to O$ 当且仅当 $|\lambda| < 1$；J^k 收敛但极限不是零矩阵当且仅当 $J = 1$.

一般地，若 J 含有不止一个 Jordan 块，类似上述讨论可得结论. □

为方便记忆，引入如下定义.

定义 10.2.4 (谱半径) 方阵 A 所有特征值的绝对值的最大值，称为 A 的谱半径，记为 $\rho(A)$.

定理 10.2.3 中满足第 2 条的矩阵常称为**收敛矩阵**. 因此 定理 10.2.3 第 2 条就可以重述为：**矩阵是收敛矩阵，当且仅当矩阵谱半径小于** 1.

练习 10.2.1 证明：$\rho(A) \leqslant \|A\|$.

命题 10.2.5 *1.* n 阶可逆矩阵的全体组成的集合是 $\mathbb{C}^{n\times n}$ 的稠密子集.

2. 所有特征值都是单特征值的 n 阶矩阵的全体组成的集合是 $\mathbb{C}^{n\times n}$ 的稠密子集.

3. 所有特征值都是单特征值的 n 阶可逆矩阵的全体组成的集合是 $\mathbb{C}^{n\times n}$ 的稠密子集.

4. n 阶可逆矩阵的全体组成的集合是 $\mathbb{C}^{n\times n}$ 的开子集.

证 第 1 条：对任意矩阵 A，设奇异值分解为 $A = U\Sigma V^{\mathrm{H}}$，其中 $\Sigma = \operatorname{diag}(\sigma_1, \cdots, \sigma_r, 0, \cdots, 0)$. 令 $A_k = U\Sigma_k V^{\mathrm{H}}$，其中 $\Sigma_k = \operatorname{diag}(\sigma_1, \cdots, \sigma_r, \frac{1}{k}, \cdots, \frac{1}{k})$，则 $A_k \to A$.

第 2 条：对任意矩阵 A，设其 Jordan 分解为 $A = XJX^{-1}$，其中 $J = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_p \end{bmatrix}$，而 $J_s = \begin{bmatrix} \lambda_s & 1 & & \\ & \lambda_s & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_s \end{bmatrix}$，$s = 1, \cdots, p$ 是 n_s 阶 Jordan 块. 记 A 的任意两特征值的距离最小值为 γ. 令 $A_k = XJ^{(k)}X^{-1}$，其中 $J^{(k)} = \begin{bmatrix} J_1^{(k)} & & \\ & \ddots & \\ & & J_p^{(k)} \end{bmatrix}$，而 $J_s^{(k)}$ 按如下方法构造：若 $n_s = 1$，则令 $J_s^{(k)} = \lambda_s^{(k)} = \lambda_s + \frac{\gamma}{sk}$；若 $n_s > 1$，则令 $J_s^{(k)} = \begin{bmatrix} \lambda_s & 1 & & \\ & \lambda_s & \ddots & \\ & & \ddots & 1 \\ \frac{\gamma}{sk} & & & \lambda_s \end{bmatrix}$.

不难得到 A_k 是特征值互不相同的矩阵，且 $A_k \to A$.

第 3 条：$\boxed{\textbf{留为作业}.}$

第 4 条：对任意可逆矩阵 A，设其最小奇异值为 $\sigma_n > 0$. 取 $\delta = \frac{\sigma_n}{2}$，则对任意满足 $\|A - B\| \leqslant \delta$ 的矩阵 B，利用命题 6.3.9 即得 B 的最小奇异值

$$
\begin{aligned}
\sigma_n(B) &= \min_{\boldsymbol{x}\neq 0} \frac{\|B\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \\
&\geqslant \min_{\boldsymbol{x}\neq 0} \frac{\|A\boldsymbol{x}\| - \|(B-A)\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \\
&\geqslant \min_{\boldsymbol{x}\neq 0} \frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|} - \max_{\boldsymbol{x}\neq 0} \frac{\|(B-A)\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \\
&= \sigma_n - \|B - A\| \\
&\geqslant \sigma_n - \delta > 0,
\end{aligned}
$$

故 B 可逆，由此可知可逆矩阵构成开子集. □

练习 10.2.2 1. 证明：n 阶可对角化矩阵的全体组成的集合是 $\mathbb{F}^{n\times n}$ 的稠密子集.

2. 举例说明，上述集合不一定是开子集.

命题 10.2.6 求逆映射 $\cdot^{-1}\colon \{A \in \mathbb{F}^{n\times n} \mid A \text{ 可逆}\} \to \mathbb{F}^{n\times n}, A \mapsto A^{-1}$ 是连续映射.

证 设可逆矩阵 A 的最小奇异值为 $\sigma_n > 0$，则 $\|A^{-1}\| = \frac{1}{\sigma_n}$. 利用命题 10.2.5 第 4 条的证明可知，只要 $\|A - B\| \leqslant \frac{\sigma_n}{2}$，就有 B 的最小奇异值不小于 $\frac{\sigma_n}{2}$，即 $\|B^{-1}\| \leqslant \frac{2}{\sigma_n}$. 因此 $\|B^{-1} - A^{-1}\| = \|B^{-1}(A-B)A^{-1}\| \leqslant \|B^{-1}\|\|A-B\|\|A^{-1}\| \leqslant \frac{2}{\sigma_n^2}\|A-B\|$. 由此即得求逆映射连续. □

练习 10.2.3 证明：$A \mapsto \|A\| = \sigma_{\max}(A)$ 和 $A \mapsto \sigma_{\min}(A)$ 是连续函数.

练习 10.2.4 证明：$A \mapsto \det(A)$ 和 $A \mapsto \operatorname{trace}(A)$ 是连续函数.

注 10.2.7 *事实上，特征值关于矩阵也是连续的*[①]*. 利用这一性质还可以证明命题 10.2.5 第 2 条和第 3 条中涉及的集合也是开子集.*

命题 10.2.8 *1.* *映射* $\{A \in \mathbb{F}^{n\times n} \mid A \text{ 的所有顺序主子阵可逆}\} \to \mathbb{F}^{n\times n}, A \mapsto A$ *的 LU 分解中的 L 是连续映射.*

2. *映射* $\{A \in \mathbb{F}^{n\times n} \mid A \text{ 的所有顺序主子阵可逆}\} \to \mathbb{F}^{n\times n}, A \mapsto A$ *的 LU 分解中的 U 是连续映射.*

3. *映射* $\{A \in \mathbb{F}^{m\times n} \mid A \text{ 列满秩}\} \to \mathbb{F}^{n\times n}, A \mapsto A$ *的 QR 分解中的 Q 是连续映射.*

4. *映射* $\{A \in \mathbb{F}^{m\times n} \mid A \text{ 列满秩}\} \to \mathbb{F}^{n\times n}, A \mapsto A$ *的 QR 分解中的 R 是连续映射.*

---

① 证明这一结论需要较深入的矩阵分析知识，感兴趣的读者请查阅矩阵分析的教材或专著.

证 $\boxed{\textbf{选做题，可不做}}$ $\square$

10.2.2 矩阵函数

下面来讨论矩阵级数和矩阵函数. 所谓矩阵级数，指的是方阵的幂级数；所谓矩阵函数，指的是方阵的解析函数. 其先决条件在于方阵有线性运算和乘法.

在一元微积分中，最简单的函数是多项式函数 $p(x) = \sum_{k=0}^{d} p_k x^k$，由此把有限项的和推广到无限项的和，就得到了性质很好的解析函数. 具体说来，$\mathbb{F}$ 上的在 $x = 0$ 处解析的函数 $f(x)$，一定在 $x = 0$ 的邻域上无穷次连续可微，且在其收敛圆盘内，有 $f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k$，其中幂级数在收敛圆盘内绝对收敛.

类似地，在第 1 章中，我们已经给出了矩阵多项式 $p(A) = \sum_{k=0}^{d} p_k A^k$，上节又给出了矩阵级数收敛的定义. 这就提示我们定义矩阵函数 $f(A) := \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} A^k$. 立刻我们就面临着一个问题:

- 级数 $\sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} A^k$ 是否收敛?

先来看矩阵级数的收敛性.

定理 10.2.9 给定方阵 $A \in \mathbb{C}^{n\times n}$ 和在 $B(0,\rho) = \{z \mid |z| < \rho\}$ 上绝对收敛的幂级数 $\sum_{k=0}^{\infty} p_k z^k$. 若 $\rho(A) < \rho$，则矩阵级数 $\sum_{k=0}^{\infty} p_k A^k$ 收敛.

证 设 A 的 Jordan 分解为 $A = XJX^{-1}$，则 $\sum_{k=0}^{\infty} p_k A^k$ 收敛等价于 $\sum_{k=0}^{\infty} p_k J^k$ 收敛，且收敛时有 $\sum_{k=0}^{\infty} p_k A^k = X\left(\sum_{k=0}^{\infty} p_k J^k\right) X^{-1}$. 因此只需证明对 Jordan 标准形级数收敛，而这又只需证明对单一 Jordan 块收敛. 考虑 Jordan 块 $J = \lambda I + J_0 = \begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}$，下证部分和 $f_m(J) := \sum_{k=0}^{m} p_k J^k$ 收敛.

类似定理 10.2.3 的证明，

$$f_m(J) = \sum_{k=0}^{m} p_k \left(\sum_{j=0}^{k} \mathrm{C}_k^j \lambda^{k-j} J_0^j \right) = \sum_{j=0}^{m} \left(\sum_{k=j}^{m} p_k \mathrm{C}_k^j \lambda^{k-j} \right) J_0^j = \sum_{j=0}^{n-1} \left(\sum_{k=j}^{m} p_k \mathrm{C}_k^j \lambda^{k-j} \right) J_0^j.$$

而 $\mathrm{C}_k^j \lambda^{k-j} = \frac{1}{j!} \frac{\mathrm{d}^j}{\mathrm{d}z^j} z^k \Big|_{z=\lambda}$，于是

$$\begin{aligned} f_m(J) &= \sum_{j=0}^{n-1} \left(\sum_{k=j}^{m} p_k \frac{1}{j!} \frac{\mathrm{d}^j}{\mathrm{d}z^j} z^k \Big|_{z=\lambda} \right) J_0^j \\ &= \sum_{j=0}^{n-1} \frac{1}{j!} \frac{\mathrm{d}^j}{\mathrm{d}z^j} \left(\sum_{k=0}^{m} p_k z^k \right) \Bigg|_{z=\lambda} J_0^j \end{aligned}$$

$$
= \sum_{j=0}^{n-1} \frac{1}{j!} f_m^{(j)}(\lambda) J_0^j = \begin{bmatrix} f_m(\lambda) & f_m'(\lambda) & \cdots & \frac{f_m^{(n-1)}(\lambda)}{(n-1)!} \\ & \ddots & \ddots & \vdots \\ & & f_m(\lambda) & f_m'(\lambda) \\ & & & f_m(\lambda) \end{bmatrix}.
$$

注意到在 $B(0,\rho)$ 内，由于 $f(z) := \sum_{k=0}^{\infty} p_k z^k$ 绝对收敛，故 f 是解析函数，且对任意 $z \in B(0,\rho)$，$f_m(z) \to f(z), f_m^{(k)}(\lambda) \to f^{(k)}(\lambda)$. 根据命题 10.2.1 ，当 $\lambda \in B(0,\rho)$ 时，$f_m(J) \to \begin{bmatrix} f(\lambda) & f'(\lambda) & \cdots & \frac{f^{(n-1)}(\lambda)}{(n-1)!} \\ & \ddots & \ddots & \vdots \\ & & f(\lambda) & f'(\lambda) \\ & & & f(\lambda) \end{bmatrix}$. 由此即得结论. □

定理 10.2.9 保证了下面定义的合理性.

定义 10.2.10 (矩阵函数) 给定方阵 $A \in \mathbb{C}^{n\times n}$ 和在 $B(0,\rho) = \{z \mid |z| < \rho\}$ 上解析的函数 $f(z) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} z^k$，且有 $\rho(A) < \rho$，则**矩阵函数**$f(A)$ 定义为 $f(A) := \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} A^k$.

常见的矩阵函数包括：

1. 矩阵指数：$\exp(A) = \mathrm{e}^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k$；

2. 矩阵正弦：$\sin(A) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} A^{2k+1}$；

3. 矩阵余弦：$\cos(A) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} A^{2k}$；

4. 矩阵对数：$\ln(I+A) = \sum_{k=0}^{\infty} \frac{(-1)^{k+1}}{k} A^k$，其中 $\rho(A) < 1$；

5. 矩阵逆：$(I-A)^{-1} = \sum_{k=0}^{\infty} A^k$，其中 $\rho(A) < 1$；

6. 矩阵幂：$(I+A)^t = \sum_{k=0}^{\infty} \frac{t(t-1)\cdots(t-k+1)}{k!} A^k$，其中 $\rho(A) < 1$.

推论 10.2.11 给定方阵 $A \in \mathbb{C}^{n\times n}$ 和在 $B(0,\rho)$ 上解析的函数 $f(z)$，且有 $\rho(A) < \rho$.

1. 若 A 是 *Jordan* 块 $\lambda I + J_0 = \begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}$，则 $f(A) = \sum_{j=0}^{n-1} \frac{1}{j!} f^{(j)}(\lambda) J_0^j =$

$$
\begin{bmatrix} f(\lambda) & f'(\lambda) & \cdots & \frac{f^{(n-1)}(\lambda)}{(n-1)!} \\ & \ddots & \ddots & \vdots \\ & & f(\lambda) & f'(\lambda) \\ & & & f(\lambda) \end{bmatrix}.
$$

2. 若 A 的 Jordan 分解为 $A = TJT^{-1}$，其中 $J = \mathrm{diag}(J_1, \cdots, J_r)$ 而 $J_i, i = 1, \cdots, r$ 是 Jordan 块，则 $f(A) = T\,\mathrm{diag}(f(J_1), \cdots, f(J_r))T^{-1}$.
3. 若 A 的所有特征值为 $\lambda_1, \cdots, \lambda_n$，则 $f(A)$ 的所有特征值为 $f(\lambda_1), \cdots, f(\lambda_n)$.

矩阵函数中最重要的是矩阵指数，下面给出其基本性质.

定理 10.2.12 给定方阵 A，则

1. e^A 可逆，且 $(\mathrm{e}^A)^{-1} = \mathrm{e}^{-A}$；
2. 对任意可逆矩阵 X，$\mathrm{e}^{X^{-1}AX} = X^{-1}\mathrm{e}^A X$；
3. $\det(\mathrm{e}^A) = \mathrm{e}^{\mathrm{trace}\,A}$；
4. $\lim_{k\to\infty}(I + \frac{A}{k})^k = \mathrm{e}^A$；
5. 若方阵 B 与 A 可交换，则 $\mathrm{e}^A\mathrm{e}^B = \mathrm{e}^{A+B} = \mathrm{e}^B\mathrm{e}^A$.

证 **留为作业.** □

练习 10.2.5 证明：

1. 反 Hermite 矩阵的指数是酉矩阵；
2. 酉矩阵一定是某个反 Hermite 矩阵的指数；
3. （极分解）对任意方阵 A，存在 Hermite 半正定矩阵 R, Θ 使得 $A = R\mathrm{e}^{\mathrm{i}\Theta}$.
   注意：参见练习 6.3.6.
4. 实反对称矩阵的指数是正交矩阵；
5. 行列式为 1 的正交矩阵一定是某个实反对称矩阵的指数；
6. 写出实矩阵的极分解并证明.

下面来看矩阵函数的运算.

定理 10.2.13 给定方阵 $A \in \mathbb{C}^{n\times n}$ 与 $B(0, \rho_1)$ 上的解析函数 $f_1(z)$ 和 $B(0, \rho_2)$ 上的解析函数 $f_2(z)$. 则

1. 若 $\rho(A) < \min\{\rho_1, \rho_2\}$，$h(z) = k_1 f_1(z) + k_2 f_2(z)$，其中 $k_1, k_2 \in \mathbb{C}$，则 $h(A) = k_1 f_1(A) + k_2 f_2(A)$；
2. 若 $\rho(A) < \min\{\rho_1, \rho_2\}$，$h(z) = f_1(z)f_2(z)$，则 $h(A) = f_1(A)f_2(A)$；
3. 若 $\rho(A) < \rho_2, \rho(f_2(A)) < \rho_1$，$h(z) = f_1(f_2(z))$，则 $h(A) = f_1(f_2(A))$.

证 与定理 10.2.3 和 10.2.9 类似，只需考虑 $A = J$ 是一个 Jordan 块的情形.

第 1 条：利用导数的线性性质，不难证明 $h^{(k)}(z) = kf_1^{(k)}(z) + lf_2^{(k)}(z)$，因此

$$h(J) = \begin{bmatrix} h(\lambda) & h'(\lambda) & \cdots & \frac{h^{(n-1)}(\lambda)}{(n-1)!} \\ & \ddots & \ddots & \vdots \\ & & h(\lambda) & h'(\lambda) \\ & & & h(\lambda) \end{bmatrix}$$

$$
\begin{aligned}
&= \begin{bmatrix} k_1 f_1(\lambda)+k_2 f_2'(\lambda) & k_1 f_1'(\lambda) & \cdots & \frac{k_1 f_1^{(n-1)}(\lambda)+k_2 f_2^{(n-1)}(\lambda)}{(n-1)!} \\ & \ddots & \ddots & \vdots \\ & & k_1 f_1(\lambda)+k_2 f_2(\lambda) & k_1 f_1'(\lambda)+k_2 f_2'(\lambda) \\ & & & k_1 f_1(\lambda)+k_2 f_2(\lambda) \end{bmatrix} \\
&= k_1 \begin{bmatrix} f_1(\lambda) & f_1'(\lambda) & \cdots & \frac{f_1^{(n-1)}(\lambda)}{(n-1)!} \\ & \ddots & \ddots & \vdots \\ & & f_1(\lambda) & f_1'(\lambda) \\ & & & f_1(\lambda) \end{bmatrix} + k_2 \begin{bmatrix} f_2(\lambda) & f_2'(\lambda) & \cdots & \frac{f_2^{(n-1)}(\lambda)}{(n-1)!} \\ & \ddots & \ddots & \vdots \\ & & f_2(\lambda) & f_2'(\lambda) \\ & & & f_2(\lambda) \end{bmatrix} \\
&= k_1 f_1(J)+k_2 f_2(J).
\end{aligned}
$$

由此即得.

第 2 条：利用导数的 Leibniz 法则 $h^{(k)}(z)=\sum_{j=0}^{k} \mathrm{C}_k^j f_1^{(j)}(z) f_2^{(k-j)}(z)$,

$$
\begin{aligned}
h(J) &= \begin{bmatrix} h(\lambda) & h'(\lambda) & \cdots & \frac{h^{(n-1)}(\lambda)}{(n-1)!} \\ & \ddots & \ddots & \vdots \\ & & h(\lambda) & h'(\lambda) \\ & & & h(\lambda) \end{bmatrix} \\
&= \begin{bmatrix} f_1(\lambda)f_2(\lambda) & f_1'(\lambda)f_2(\lambda)+f_1(\lambda)f_2'(\lambda) & \cdots & \frac{f_1^{(n-1)}(\lambda)f_2(\lambda)+\cdots+f_1(\lambda)f_2^{(n-1)}(\lambda)}{(n-1)!} \\ & \ddots & \ddots & \vdots \\ & & f_1(\lambda)+f_2(\lambda) & f_1'(\lambda)f_2(\lambda)+f_1(\lambda)f_2'(\lambda) \\ & & & f_1(\lambda)f_2(\lambda) \end{bmatrix} \\
&= \begin{bmatrix} f_1(\lambda) & f_1'(\lambda) & \cdots & \frac{f_1^{(n-1)}(\lambda)}{(n-1)!} \\ & \ddots & \ddots & \vdots \\ & & f_1(\lambda) & f_1'(\lambda) \\ & & & f_1(\lambda) \end{bmatrix} \begin{bmatrix} f_2(\lambda) & f_2'(\lambda) & \cdots & \frac{f_2^{(n-1)}(\lambda)}{(n-1)!} \\ & \ddots & \ddots & \vdots \\ & & f_2(\lambda) & f_2'(\lambda) \\ & & & f_2(\lambda) \end{bmatrix} \\
&= f_1(J)f_2(J).
\end{aligned}
$$

由此即得.

第 3 条：利用导数的链式法则可得结论，细节参见阅读 10.2.6 . □

阅读 10.2.6 (定理 10.2.13 第 3 条的证明) TBD

例 10.2.14 反复利用定理 10.2.13 ，可以构造很多矩阵函数及之间的关系：

1. $\sin^2 A+\cos^2 A=I, \mathrm{e}^{\mathrm{i}A}=\cos A+\mathrm{i}\sin A.$
2. $\sin(2A)=2\sin A\cos A, \cos(2A)=2\cos^2 A-I.$

3. 若 $f(z)=\frac{1}{z}$，则 $f(A)=A^{-1}$，这是因为我们可以从 $f(z)z=1$ 得到 $f(A)A=I$. ☺

10.3 矩阵微积分

10.3.1 一元微积分

首先考虑度量空间 $\mathbb{F}^{m\times n}$ 上的一元微积分.

这时从 $\mathbb{F}$ 的子集到度量空间 $\mathbb{F}^{m\times n}$ 的映射 $\boldsymbol{F}$ 可以明确表示为

$$\begin{bmatrix} f_{11}(t) & f_{12}(t) & \cdots & f_{1n}(t) \\ f_{21}(t) & f_{22}(t) & \cdots & f_{2n}(t) \\ \vdots & \vdots & & \vdots \\ f_{m1}(t) & f_{m2}(t) & \cdots & f_{mn}(t) \end{bmatrix} =: F(t).$$

这种元素是函数的矩阵称为**函数矩阵**.

命题 10.3.1 对函数矩阵 $A(t)=\left[a_{ij}(t)\right]$，

1. 它在 $t\to t_*$ 时收敛当且仅当函数矩阵**逐元素**收敛，即对任意 i,j，$a_{ij}(t)$ 在 $t\to t_*$ 时收敛；
2. 它在 t_* 处连续当且仅当函数矩阵**逐元素**连续，即对任意 i,j，$a_{ij}(t)$ 在 t_* 处连续；
3. 它在 t_* 处可微当且仅当函数矩阵**逐元素**可微，即对任意 i,j，$a_{ij}(t)$ 在 t_* 处可微；
4. 它在区间 I 上可积当且仅当函数矩阵**逐元素**可积，即对任意 i,j，$a_{ij}(t)$ 在区间 I 上可积.

而 $A(t)$ 可微或可积时，导数 $A'(t)=\left[a'_{ij}(t)\right]$，微分 $\mathrm{d}A(t)=A'(t)\,\mathrm{d}t$，积分 $\int_I A(t)\,\mathrm{d}t=\left[\int_I a_{ij}(t)\,\mathrm{d}t\right]$.

进一步地，函数矩阵 k 阶连续可微，当且仅当函数矩阵**逐元素** k 阶连续可微；函数矩阵光滑，当且仅当函数矩阵**逐元素**光滑；函数矩阵解析，当且仅当函数矩阵**逐元素**解析.

证 利用与命题 10.2.1 的证明相同的思路，不难证明第 1 条. 其他结论利用第 1 条可得. □

因此前面多个章节中直接认定了向量或矩阵的导数和微分就是元素的导数和微分组成的向量或矩阵.

函数矩阵的极限、微分、积分的运算法则和函数的法则完全一样，稍加计算即可证明其正确性，最需要注意的是矩阵乘法没有交换律，因此计算时不能交换因子的顺序，如

1. $\mathrm{d}\,(A(t)B(t)) = \mathrm{d}A(t)B(t) + A(t)\,\mathrm{d}B(t)$;
2. $\int_I A(t)B'(t)\,\mathrm{d}t = (A(t)B(t))\,\big|_I - \int_I A'(t)B(t)\,\mathrm{d}t$.

命题 10.3.2 $\frac{\mathrm{d}(A(t)^{-1})}{\mathrm{d}t} = -A(t)^{-1}\frac{\mathrm{d}A(t)}{\mathrm{d}t}A(t)^{-1}$.

证 对 $A(t)A(t)^{-1} = I$ 两边求导数，有 $A(t)\frac{\mathrm{d}(A(t)^{-1})}{\mathrm{d}t} + \frac{\mathrm{d}A(t)}{\mathrm{d}t}A(t)^{-1} = O$. 立得. □

命题 10.3.3 函数矩阵的行列式的导数公式：

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\det(\boldsymbol{a}_1(t),\boldsymbol{a}_2(t),\cdots,\boldsymbol{a}_n(t)) = \quad & \det(\boldsymbol{a}_1'(t),\boldsymbol{a}_2(t),\cdots,\boldsymbol{a}_n(t)) \\ & + \det(\boldsymbol{a}_1(t),\boldsymbol{a}_2'(t),\cdots,\boldsymbol{a}_n(t)) \\ & \qquad\qquad \vdots \\ & + \det(\boldsymbol{a}_1(t),\boldsymbol{a}_2(t),\cdots,\boldsymbol{a}_n'(t)).\end{aligned}$$

特别地，当 $A(t)$ 可逆时，$\frac{\mathrm{d}}{\mathrm{d}t}\det(A(t)) = \det(A(t))\,\mathrm{trace}\,(A(t)^{-1}A'(t))$.

证 1 利用完全展开：

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\det(\boldsymbol{a}_1(t),\boldsymbol{a}_2(t),\cdots,\boldsymbol{a}_n(t)) \\ &\qquad = \frac{\mathrm{d}}{\mathrm{d}t}\sum_{\text{排列 }\sigma} a_{\sigma_1 1}(t)a_{\sigma_2 2}(t)\cdots a_{\sigma_n n}(t) \\ &\qquad = \sum_{\sigma}\frac{\mathrm{d}}{\mathrm{d}t}\Big(a_{\sigma_1 1}(t)a_{\sigma_2 2}(t)\cdots a_{\sigma_n n}(t)\Big) \\ &\qquad = \sum_{\sigma}\Big(a_{\sigma_1 1}'(t)a_{\sigma_2 2}(t)\cdots a_{\sigma_n n}(t) + a_{\sigma_1 1}(t)a_{\sigma_2 2}'(t)\cdots a_{\sigma_n n}(t) \\ &\qquad\qquad\qquad + \cdots + a_{\sigma_1 1}(t)a_{\sigma_2 2}(t)\cdots a_{\sigma_n n}'(t)\Big) \\ &\qquad = \sum_{\sigma} a_{\sigma_1 1}'(t)a_{\sigma_2 2}(t)\cdots a_{\sigma_n n}(t) + \sum_{\sigma} a_{\sigma_1 1}(t)a_{\sigma_2 2}'(t)\cdots a_{\sigma_n n}(t) \\ &\qquad\qquad\qquad + \cdots + \sum_{\sigma} a_{\sigma_1 1}(t)a_{\sigma_2 2}(t)\cdots a_{\sigma_n n}'(t) \\ &\qquad = \det(\boldsymbol{a}_1'(t),\boldsymbol{a}_2(t),\cdots,\boldsymbol{a}_n(t)) + \det(\boldsymbol{a}_1(t),\boldsymbol{a}_2'(t),\cdots,\boldsymbol{a}_n(t)) \\ &\qquad\qquad\qquad + \cdots + \det(\boldsymbol{a}_1(t),\boldsymbol{a}_2(t),\cdots,\boldsymbol{a}_n'(t)).\end{aligned}$$

然后按对应列展开来化简，其中 $C_{ij}(t)$ 为 $a_{ij}(t)$ 的代数余子式，而 $C(t) = \big[C_{ij}(t)\big]$：

$$\frac{\mathrm{d}}{\mathrm{d}t}\det(A(t)) = \sum_{i=1}^{n}\det(\cdots,\boldsymbol{a}_{i-1}(t),\boldsymbol{a}_i'(t),\boldsymbol{a}_{i+1}(t),\cdots)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a'_{ij}(t) C_{ij}(t) = \operatorname{trace}(C(t)^{\mathrm{T}} A'(t)).$$

注意 $C(t)^{\mathrm{T}}$ 是 $A(t)$ 的补矩阵, 当 $A(t)$ 可逆时, $C(t)^{\mathrm{T}} = \det(A(t))A(t)^{-1}$, 由此即得. □

证 2 对 n 用归纳法. $n=1$ 时显然成立，假设命题对 $n-1$ 阶函数矩阵成立，下证命题对 n 阶函数矩阵成立.

行列式按第一列展开，其中 $\boldsymbol{a}_{i,j}$ 是 $\boldsymbol{a}_j$ 去掉第 i 个元素得到的向量,

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t} \det(\boldsymbol{a}_1(t), \boldsymbol{a}_2(t), \cdots, \boldsymbol{a}_n(t)) \\
&\quad = \frac{\mathrm{d}}{\mathrm{d}t} \sum_{i=1}^{n} a_{i1}(t) \det(\boldsymbol{a}_{i,2}(t), \cdots, \boldsymbol{a}_{i,n}(t)) \\
&\quad = \sum_{i=1}^{n} a'_{i1}(t) \det(\boldsymbol{a}_{i,2}(t), \cdots, \boldsymbol{a}_{i,n}(t)) + \sum_{i=1}^{n} a_{i1}(t) \frac{\mathrm{d}}{\mathrm{d}t} \det(\boldsymbol{a}_{i,2}(t), \cdots, \boldsymbol{a}_{i,n}(t))
\end{aligned}$$

利用归纳假设,

$$\begin{aligned}
&= \sum_{i=1}^{n} a'_{i1}(t) \det(\boldsymbol{a}_{i,2}(t), \cdots, \boldsymbol{a}_{i,n}(t)) \\
&\qquad + \sum_{i=1}^{n} a_{i1}(t) \left[\det(\boldsymbol{a}'_{i,2}(t), \cdots, \boldsymbol{a}_{i,n}(t)) + \cdots + \det(\boldsymbol{a}_{i,2}(t), \cdots, \boldsymbol{a}'_{i,n}(t))\right] \\
&= \sum_{i=1}^{n} a'_{i1}(t) \det(\boldsymbol{a}_{i,2}(t), \cdots, \boldsymbol{a}_{i,n}(t)) \\
&\qquad + \sum_{i=1}^{n} a_{i1}(t) \det(\boldsymbol{a}'_{i,2}(t), \cdots, \boldsymbol{a}_{i,n}(t)) + \cdots + \sum_{i=1}^{n} a_{i1}(t) \det(\boldsymbol{a}_{i,2}(t), \cdots, \boldsymbol{a}'_{i,n}(t)) \\
&= \det(\boldsymbol{a}'_1(t), \boldsymbol{a}_2(t), \cdots, \boldsymbol{a}_n(t)) \\
&\qquad + \det(\boldsymbol{a}_1(t), \boldsymbol{a}'_2(t), \cdots, \boldsymbol{a}_n(t)) + \cdots + \det(\boldsymbol{a}_1(t), \boldsymbol{a}_2(t), \cdots, \boldsymbol{a}'_n(t))
\end{aligned}$$

特别情形同证 1. □

上节中的矩阵函数加入变量也可以得到函数矩阵.

给定方阵 $A \in \mathbb{C}^{n\times n}$ 和在 $B(0,\rho)$ 上解析的函数 $f(z)$. 则 $f(At)$ 在合适范围内都有定义:

$$f(At) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!}(At)^k = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} A^k t^k, \qquad \text{其中} \quad \rho(A)|t| < \rho.$$

利用绝对收敛和微分或积分次序的可交换性，立得如下结论.

定理 10.3.4 *1.* $\mathrm{d}f(At) = f'(At)A\,\mathrm{d}t = Af'(At)\,\mathrm{d}t$;

2. $\int_{t_1}^{t_2} Af'(At)\,\mathrm{d}t = f(At_2) - f(At_1)$.

特别地，

1. $\mathrm{d}\mathrm{e}^{At} = A\mathrm{e}^{At}\,\mathrm{d}t$;

2. $A\int_0^t \mathrm{e}^{A\tau}\,\mathrm{d}\tau = \mathrm{e}^{At} - I$.

矩阵指数常用于定常线性常微分方程的求解. 回顾例 7.5.15 ，其结论总结得到命题 10.3.5 . 稍加引申还可得到定理 10.3.6 .

命题 10.3.5 常微分方程 $\begin{cases}\dfrac{\mathrm{d}\boldsymbol{u}(t)}{\mathrm{d}t} = A\boldsymbol{u}(t),\\ \boldsymbol{u}(0) = \boldsymbol{u}_0,\end{cases}$ 的唯一解是 $\boldsymbol{u}(t) = \mathrm{e}^{At}\boldsymbol{u}_0$.

定理 10.3.6 对矩阵 $A \in \mathbb{C}^{n\times n}$，

- 当 $t \to +\infty$ 时，e^{At} 收敛，当且仅当 A 只有两种特征值，一是实部小于 0 的特征值，二是半单的特征值 0；
- 当 $t \to +\infty$ 时，$\mathrm{e}^{At} \to O$，当且仅当 A 是**稳定矩阵**，即 A 的特征值实部都小于 0.

证 **留为作业.** □

推论 10.3.7 若 A 是稳定矩阵，则 $A^{-1} = -\int_0^{+\infty} \mathrm{e}^{At}\,\mathrm{d}t$.

证 **留为作业.** □

下面来考虑定常非齐次线性常微分方程.

定理 10.3.8 常微分方程 $\begin{cases}\dfrac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = A\boldsymbol{x}(t) + B\boldsymbol{u}(t),\\ \boldsymbol{x}(0) = \boldsymbol{x}_0,\end{cases}$ 的唯一解是 $\boldsymbol{x}(t) = \mathrm{e}^{A(t-t_0)}\boldsymbol{x}_0 + \int_0^t \mathrm{e}^{A(t-\tau)}B\boldsymbol{u}(\tau)\,\mathrm{d}\tau$.

证 **留为作业.** □

利用定理 10.3.4 ，还可以求解一些矩阵方程，例如控制论中常见的 Sylvester 方程（参见练习 5.2.23 和练习 5.4.7） $A_1X - XA_2 = B$.

命题 10.3.9 *1.* 当 $-A_1$ 和 A_2 是稳定矩阵时，*Sylvester* 方程 $A_1X - XA_2 = B$ 的唯一解是 $\int_0^{+\infty} \mathrm{e}^{-A_1t}B\mathrm{e}^{A_2t}\,\mathrm{d}t$.

2. 当 $A \in \mathbb{R}^{n\times n}$ 是稳定矩阵时，*Lyapunov* 方程 $A^{\mathrm{T}}X + XA = -B$ 的唯一解是 $\int_0^{+\infty} \mathrm{e}^{A^{\mathrm{T}}t}B\mathrm{e}^{At}\,\mathrm{d}t$.

3. 特别地，当 B 是实对称正定矩阵时，上述 *Lyapunov* 方程的解也是实对称正定矩阵.

证 只证明第 1 条, 第 2 条是前者的简单推论. 第 3 条留给读者验证. **留为作业**.

首先，由 $-A_1$ 和 A_2 稳定，二者没有公共的特征值，根据练习 5.4.7，Sylvester 方程存在唯一解，下面就来找到这个解.

考虑函数矩阵 $B(t)=\mathrm{e}^{-A_1t}B\mathrm{e}^{A_2t}$. 则 $B(0)=B, \lim_{t\to+\infty}B(s)=O$. 两边微分可得

$$\mathrm{d}B(t)=(-A_1\mathrm{e}^{-A_1t}B\mathrm{e}^{A_2t}+\mathrm{e}^{-A_1t}B\mathrm{e}^{A_2t}A_2)\,\mathrm{d}t.$$

因此

$$\begin{aligned}\int_0^{+\infty}\mathrm{d}B(t)&=\int_0^{+\infty}(-A_1\mathrm{e}^{-A_1t}B\mathrm{e}^{A_2t}+\mathrm{e}^{-A_1t}B\mathrm{e}^{A_2t}A_2)\,\mathrm{d}t\\&=-A_1\left(\int_0^{+\infty}\mathrm{e}^{-A_1t}B\mathrm{e}^{A_2t}\,\mathrm{d}t\right)+\left(\int_0^{+\infty}\mathrm{e}^{-A_1t}B\mathrm{e}^{A_2t}\,\mathrm{d}t\right)A_2.\end{aligned}$$

另一方面，$\int_0^{+\infty}\mathrm{d}B(t)=\lim_{s\to+\infty}B(s)-B(0)=O-B=-B$. 立得结论. □

10.3.2 多元微分

下面考虑自变量为向量或矩阵的微分.

一般来说，使用增量法来分析涉及向量或矩阵的微分总是有益的，而且当函数有显式表示时，增量法能够给出微分的显式表示. 下面就给出几个增量法的例子.

命题 10.3.10 *求逆映射* $\cdot^{-1}\colon\{A\in\mathbb{F}^{n\times n}\mid A\text{ 可逆}\}\to\mathbb{F}^{n\times n}, A\mapsto A^{-1}$ *是可微映射，且其微分为* $\mathrm{d}A^{-1}=-A^{-1}\,\mathrm{d}AA^{-1}$.

证 使用增量法. 取满足 $\|\delta A\|\|A^{-1}\|<1$ 的 δA，当 A 可逆时，根据命题 10.2.5 第 4 条，$A+\delta A$ 也可逆. 于是

$$\begin{aligned}(A+\delta A)^{-1}&=(I+A^{-1}\delta A)^{-1}A^{-1}\\&=\sum_{k=0}^{\infty}(-1)^{k+1}(A^{-1}\delta A)^kA^{-1}\\&=A^{-1}-A^{-1}\delta AA^{-1}+A^{-1}\delta AA^{-1}\delta A\sum_{k=0}^{\infty}(-1)^{k+1}(A^{-1}\delta A)^kA^{-1},\\&=A^{-1}-A^{-1}\delta AA^{-1}+A^{-1}\delta AA^{-1}\delta A\sum_{k=0}^{\infty}(-1)^{k+1}(A^{-1}\delta A)^kA^{-1},\end{aligned}$$

则当 $\delta A\to O$ 时,

$$\frac{\left\|(A+\delta A)^{-1}-A^{-1}+A^{-1}\delta AA^{-1}\right\|}{\|\delta A\|}=\frac{1}{\|\delta A\|}\left\|A^{-1}\delta AA^{-1}\delta A\sum_{k=0}^{\infty}(-1)^{k+1}(A^{-1}\delta A)^kA^{-1}\right\|$$

$$\leqslant \left\|A^{-1}\right\|^3 \|\delta A\| \sum_{k=0}^{\infty} \left\|A^{-1}\right\|^k \|\delta A\|^k$$
$$= \frac{\left\|A^{-1}\right\|^3}{1-\|A^{-1}\|\|\delta A\|}\|\delta A\| \to 0,$$

这就说明求逆是可微映射，且微分为 $\mathrm{d}A^{-1} = -A^{-1}\,\mathrm{d}AA^{-1}$. □

命题 10.3.11 行列式 $\det\colon A \mapsto \det(A)$ 是可微映射，且其微分为

$$\mathrm{d}\left(\det(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n)\right) = \det(\mathrm{d}\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n) + \det(\boldsymbol{a}_1, \mathrm{d}\boldsymbol{a}_2, \cdots, \boldsymbol{a}_n) + \cdots + \det(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \mathrm{d}\boldsymbol{a}_n).$$

特别地，当 $\det(A) \neq 0$ 时，$\mathrm{d}\left(\det(A)\right) = \det(A)\,\mathrm{trace}(A^{-1}\,\mathrm{d}A)$.

证 使用增量法，有

$$\begin{aligned}
\det(A+\delta A) = \det(&\boldsymbol{a}_1+\delta\boldsymbol{a}_1, \boldsymbol{a}_2+\delta\boldsymbol{a}_2, \cdots, \boldsymbol{a}_n+\delta\boldsymbol{a}_n) \\
= \quad & \det(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n) \\
& + \det(\delta\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n) + \det(\boldsymbol{a}_1, \delta\boldsymbol{a}_2, \cdots, \boldsymbol{a}_n) + \cdots + \det(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \delta\boldsymbol{a}_n) \\
& + \det(\delta\boldsymbol{a}_1, \delta\boldsymbol{a}_2, \cdots, \boldsymbol{a}_n) + \text{其他含有两个向量增量的项} \\
& + \det(\delta\boldsymbol{a}_1, \delta\boldsymbol{a}_2, \delta\boldsymbol{a}_3, \cdots) + \text{其他含有三个向量增量的项} \\
& + \cdots \\
& + \det(\delta\boldsymbol{a}_1, \delta\boldsymbol{a}_2, \cdots, \delta\boldsymbol{a}_n).
\end{aligned}$$

由 Hadamard 不等式（参见练习 4.2.28 和练习 6.2.7），每一个含有超过两个向量增量的行列式都满足

$$\left|\det(\cdots, \delta\boldsymbol{a}_i, \cdots, \delta\boldsymbol{a}_j, \cdots)\right| \leqslant \cdots \|\delta\boldsymbol{a}_i\| \cdots \left\|\delta\boldsymbol{a}_j\right\| \cdots \leqslant \|\delta A\|^2 \|A\|^{n-2},$$

其中省略号中的向量可能是 A 或 δA 的列：$\|\boldsymbol{a}_k\| = \|A\boldsymbol{e}_k\| \leqslant \|A\|$，$\|\delta\boldsymbol{a}_k\| \leqslant \|\delta A\| \leqslant \|A\|$. 因此，当 $\delta A \to O$ 时，

$$\begin{aligned}
&\frac{1}{\|\delta A\|}\left|\det(A+\delta A) - \det(A) - \sum_{i=1}^{n} \det(\cdots, \boldsymbol{a}_{i-1}, \delta\boldsymbol{a}_i, \boldsymbol{a}_{i+1}, \cdots)\right| \\
&\qquad\qquad \leqslant (2^n - n - 1)\|\delta A\|\|A\|^{n-2} \to 0,
\end{aligned}$$

这就说明行列式是可微映射.

若 $\det(A) \neq 0$，类似于命题 10.3.3 中特殊情形的证明可得. □

注 10.3.12 命题 *10.3.10* 和命题 *10.3.2* 、命题 *10.3.11* 和命题 *10.3.3* 的类似，本质上是一阶微分形式的不变性，可以参考分析学的相关教材或专著.

例 10.3.13 (矩阵单特征值的微分) 事实上，还能证明矩阵的单特征值关于矩阵也可微[1]．这里只考虑实对称矩阵这一简单情形．记 $A\boldsymbol{x} = \lambda\boldsymbol{x}$，其中 $\boldsymbol{x}$ 是 λ 的特征向量．考虑到特征向量的长度不确定，用 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = 1$ 来确定其长度，那么 $\boldsymbol{x}$ 关于矩阵可微[1]．下面来计算 λ 和 $\boldsymbol{x}$ 的微分．

首先

$$\begin{aligned}
A\boldsymbol{x} = \lambda\boldsymbol{x} \quad &\Rightarrow \quad \mathrm{d}A\boldsymbol{x} + A\,\mathrm{d}\boldsymbol{x} = \mathrm{d}\lambda\boldsymbol{x} + \lambda\,\mathrm{d}\boldsymbol{x}, \qquad (10.3.1)\\
\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = 1 \quad &\Rightarrow \quad \mathrm{d}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{y}^{\mathrm{T}}\,\mathrm{d}\boldsymbol{x} = 0 \quad \Rightarrow \quad \boldsymbol{x}^{\mathrm{T}}\,\mathrm{d}\boldsymbol{x} = 0.
\end{aligned}$$

在 (10.3.1) 式两边同乘 $\boldsymbol{x}^{\mathrm{T}}$，则有 $\boldsymbol{x}^{\mathrm{T}}\,\mathrm{d}A\boldsymbol{x} + \boldsymbol{x}^{\mathrm{T}}A\,\mathrm{d}x = \mathrm{d}\lambda\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} + \lambda\boldsymbol{x}^{\mathrm{T}}\,\mathrm{d}\boldsymbol{x}$，可得

$$\mathrm{d}\lambda = \boldsymbol{x}^{\mathrm{T}}\,\mathrm{d}A\boldsymbol{x}.$$

把 $\boldsymbol{x}$ 扩充成正交矩阵 $\begin{bmatrix}\boldsymbol{x} & Q_x\end{bmatrix}$，故 $I = \begin{bmatrix}\boldsymbol{x} & Q_x\end{bmatrix}\begin{bmatrix}\boldsymbol{x} & Q_x\end{bmatrix}^{\mathrm{T}} = \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}} + Q_xQ_x^{\mathrm{T}}$. 再由 (10.3.1) 式有，$(\lambda I - A)\,\mathrm{d}\boldsymbol{x} = \mathrm{d}A\boldsymbol{x} - \mathrm{d}\lambda\boldsymbol{x}$．两边同乘 Q_x^{T}，则有

$$Q_x^{\mathrm{T}}(\lambda I - A)(\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}} + Q_xQ_x^{\mathrm{T}})\,\mathrm{d}\boldsymbol{x} = Q_x^{\mathrm{T}}(\mathrm{d}A\boldsymbol{x} - \mathrm{d}\lambda\boldsymbol{x}),$$

故 $Q_x^{\mathrm{T}}\,\mathrm{d}\boldsymbol{x} = [Q_x^{\mathrm{T}}(\lambda I - A)Q_x]^{-1}Q_x^{\mathrm{T}}(\mathrm{d}A\boldsymbol{x} - \mathrm{d}\lambda\boldsymbol{x})$，于是

$$\begin{aligned}
\mathrm{d}\boldsymbol{x} &= (\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}} + Q_xQ_x^{\mathrm{T}})\,\mathrm{d}\boldsymbol{x}\\
&= \boldsymbol{0} + Q_x[Q_x^{\mathrm{T}}(\lambda I - A)Q_x]^{-1}Q_x^{\mathrm{T}}(\mathrm{d}A\boldsymbol{x} - \mathrm{d}\lambda\boldsymbol{x})\\
&= Q_x[Q_x^{\mathrm{T}}(\lambda I - A)Q_x]^{-1}Q_x^{\mathrm{T}}\,\mathrm{d}A\boldsymbol{x}.
\end{aligned}$$

☺

最后来看一下求逆和行列式的导映射．根据命题 10.3.10 ，导映射就是线性映射

$$\boldsymbol{D}\colon \mathbb{F}^{n\times n} \to \mathbb{F}^{n\times n}, X \mapsto -A^{-1}XA^{-1};$$

根据命题 10.3.11 ，导映射就是线性映射

$$\boldsymbol{W}\colon \mathbb{F}^{n\times n} \to \mathbb{F}, X \mapsto \sum_{k=1}^{n}\det(\cdots, A\boldsymbol{e}_{k-1}, X\boldsymbol{e}_k, A\boldsymbol{e}_{k+1}, \cdots).$$

这两个线性映射表示起来颇为不易．按第 7 章的思路，把它们在一组基下表示将会使问题简单．

取 $\mathbb{F}^{n\times n}$ 的一组标准基 $E_{11}, E_{21}, \cdots, E_{n1}, E_{12}, E_{22}, \cdots, E_{n2}, \cdots, \cdots, E_{1n}, E_{2n}, \cdots, E_{nn}$，其中 $E_{ij} = \boldsymbol{e}_i\boldsymbol{e}_j^{\mathrm{T}}, i, j = 1, 2, \cdots, n$．那么矩阵 $X = \begin{bmatrix}\boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n\end{bmatrix}$ 在这组基下的坐标是

[1] 证明这一结论需要较深入的矩阵分析知识，感兴趣的读者请查阅矩阵分析的教材或专著.

$\begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_n \end{bmatrix}$. 记 $A^{-1} = B = \begin{bmatrix} b_{ij} \end{bmatrix}$，则

$$\boldsymbol{D}(E_{ij}) = -A^{-1}E_{ij}A^{-1} = -B\boldsymbol{e}_i\boldsymbol{e}_j^{\mathrm{T}}B = -\sum_{i',j'}(\boldsymbol{e}_{i'}^{\mathrm{T}}B\boldsymbol{e}_i\boldsymbol{e}_j^{\mathrm{T}}B\boldsymbol{e}_{j'})E_{i'j'} = -\sum_{i',j'}b_{i'i}b_{jj'}E_{i'j'},$$

因此映射 $\boldsymbol{D}$ 在这组基下的矩阵是 $-\begin{bmatrix} b_{11}B & \cdots & b_{n1}B \\ \vdots & & \vdots \\ b_{1n}B & \cdots & b_{nn}B \end{bmatrix}$. 而

$$\boldsymbol{W}(E_{ij}) = \sum_{i=1}^{n}\det(\cdots, A\boldsymbol{e}_{i-1}, \boldsymbol{e}_i\boldsymbol{e}_j^{\mathrm{T}}\boldsymbol{e}_k, A\boldsymbol{e}_{i+1}, \cdots) = \det(\cdots, A\boldsymbol{e}_{j-1}, \boldsymbol{e}_i, A\boldsymbol{e}_{j+1}, \cdots) = C_{ij}$$

是 a_{ij} 的代数余子式，因此映射 $\boldsymbol{W}$ 在基 $E_{ij}, i, j = 1, \cdots, n; 1$ 下的矩阵是

$$\begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} & C_{12} & C_{22} & \cdots & C_{n2} & \cdots\cdots & C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}.$$

上述讨论启发我们作如下定义.

定义 10.3.14 (向量化、Kronecker 积) 矩阵的**向量化算子**定义为

$$\mathrm{vec}\colon \begin{bmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \end{bmatrix} \mapsto \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_n \end{bmatrix}.$$

矩阵 $A = \begin{bmatrix} a_{ij} \end{bmatrix}_{m\times n}$ 和 B 的 ***Kronecker* 积**或**张量积**定义为

$$A \otimes B := \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{mn}B & \cdots & a_{mn}B \end{bmatrix}.$$

注意向量化算子 vec 其实是一系列定义在不同的线性空间上的映射，但由于其功能的一致性，这里采用了相同的记号而不加区分，读者自能从自变量中得到定义域. 另外，当矩阵的行列数确定后，vec 的逆也是确定的.

为说明这样处理的有效性，我们把上面的映射 $\boldsymbol{D}$ 推广到任意矩阵空间.

例 10.3.15 考虑线性映射 $f\colon \mathbb{F}^{m\times n} \to \mathbb{F}^{m'\times n'}, X \mapsto BXA^{\mathrm{T}}$. 在定义域和陪域上分别有一组标准基 $\boldsymbol{e}_i\boldsymbol{i}_j^{\mathrm{T}}; \tilde{\boldsymbol{e}}_{i'}\tilde{\boldsymbol{i}}_{j'}^{\mathrm{T}}$，其中四组向量基 $\boldsymbol{e}_i, \boldsymbol{i}_j, \tilde{\boldsymbol{e}}_{i'}\tilde{\boldsymbol{i}}_{j'}$ 都是标准基，而矩阵空间上基的顺序按照下标 $11, 21, 31, \cdots, 12, 22, 32, \cdots, \cdots$ 排列. 不难得到这个线性映射在对

应基下的矩阵就是 $m'm \times n'n$ 矩阵 $\begin{bmatrix} a_{11}B & \cdots & a_{1m}B \\ \vdots & & \vdots \\ a_{m'1}B & \cdots & a_{m'm}B \end{bmatrix} = A \otimes B$. 而任意自变量 $X = \begin{bmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \end{bmatrix}$ 在相应基下的坐标恰好是 $\begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_n \end{bmatrix} = \operatorname{vec}(X)$，像 $Y = \begin{bmatrix} \boldsymbol{y}_1 & \cdots & \boldsymbol{y}_n \end{bmatrix}$ 在相应基下的坐标恰好是 $\begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_n \end{bmatrix} = \operatorname{vec}(Y)$. 有了定义域和陪域中的向量的坐标表示和二者之间线性映射的矩阵表示，我们就可以用矩阵乘法来计算任意向量的像，即 $\operatorname{vec}(Y) = (A \otimes B)\operatorname{vec}(X)$. ☺

容易得到如下的运算法则和简单性质.

命题 10.3.16 *1.* $\operatorname{vec}(AXB) = (B^{\mathrm{T}} \otimes A)\operatorname{vec}(X)$;

2. 结合律：$A \otimes (B \otimes C) = (A \otimes B) \otimes C$;

3. 双线性律：$(kA + lC) \otimes (pB + qD) = kpA \otimes B + kqA \otimes D + lpC \otimes B + lqC \otimes D$;

4. $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$;

5. 若 A, B 可逆，则 $A \otimes B$ 可逆，且 $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$;

6. $(A \otimes B)^{\mathrm{T}} = A^{\mathrm{T}} \otimes B^{\mathrm{T}}, (A \otimes B)^{\mathrm{H}} = A^{\mathrm{H}} \otimes B^{\mathrm{H}}$;

7. $\operatorname{rank}(A \otimes B) = \operatorname{rank}(A)\operatorname{rank}(B)$;

8. 若 $\boldsymbol{a}, \boldsymbol{b}$ 是向量，则 $\boldsymbol{b}\boldsymbol{a}^{\mathrm{T}} = \boldsymbol{b} \otimes \boldsymbol{a}^{\mathrm{T}} = \boldsymbol{a}^{\mathrm{T}} \otimes \boldsymbol{b}$;

9. 若 A 是 m 阶方阵，B 是 n 阶方阵，二者特征值分别为 $\lambda_1, \cdots, \lambda_m; \mu_1, \cdots, \mu_n$，则：

   (a) $\operatorname{trace}(A \otimes B) = \operatorname{trace}(A)\operatorname{trace}(B)$;

   (b) $\det(A \otimes B) = (\det(A))^n(\det(B))^m$;

   (c) $A \otimes B$ 的特征值为 $\lambda_i\mu_j, i = 1, \cdots, m, j = 1, \cdots, n$.

   (d) $I_n \otimes A + B \otimes I_m$ 的特征值为 $\lambda_i + \mu_j, i = 1, \cdots, m, j = 1, \cdots, n$.

证 **留为作业**. □

利用上述运算法则，求逆的导映射 $\boldsymbol{D}$ 就可以直接写作

$$\operatorname{vec}(X) \mapsto -A^{-T} \otimes A^{-1}\operatorname{vec}(X), \quad \text{或} \quad \frac{\mathrm{d}\operatorname{vec}(A^{-1})}{\mathrm{d}\operatorname{vec}(A)} = -A^{-T} \otimes A^{-1};$$

行列式的导映射 $\boldsymbol{W}$ 就直接写作

$$\operatorname{vec}(X) \mapsto \operatorname{vec}(C)^{\mathrm{T}} \operatorname{vec}(X), \quad \text{或} \quad \frac{\mathrm{d}(\det(A))}{\mathrm{d} \operatorname{vec}(A)} = \operatorname{vec}(C)^{\mathrm{T}}.$$

向量化和 Kronecker 积对线性矩阵方程也很有帮助.

命题 10.3.17 *1.* *Sylvester 方程* $A_1X - XA_2 = B$ *存在唯一解，当且仅当* A_1, A_2 *没有公共特征值. 此时唯一解* X *满足* $\operatorname{vec}(X) = (I \otimes A_1 - A_2^{\mathrm{T}} \otimes I)^{-1} \operatorname{vec}(B)$.

2. *Lyapunov 方程* $A^{\mathrm{T}}X + XA = B$ *存在唯一解，当且仅当* A *的任意两特征值之和不为* 0. *此时唯一解* X *满足* $\operatorname{vec}(X) = (I \otimes A^{\mathrm{T}} + A^{\mathrm{T}} \otimes I)^{-1} \operatorname{vec}(B)$.

3. *Stein 方程* $X - A_1XA_2 = B$ *存在唯一解，当且仅当* A_1 *的非零特征值的倒数不是* A_2 *的特征值. 此时唯一解* X *满足* $\operatorname{vec}(X) = (I - A_2^{\mathrm{T}} \otimes A_1)^{-1} \operatorname{vec}(B)$.

最后作一简单注记. 从命题 10.3.10 的证明中不难看到，求逆映射的二阶微分应当是

$$\mathrm{d}^2 A^{-1} = A^{-1}\,\mathrm{d}A A^{-1}\,\mathrm{d}A A^{-1}.$$

但即使利用 Kronecker 积，此式也难以写成 $\frac{\mathrm{d}^2 \operatorname{vec}(A^{-1})}{[\mathrm{d} \operatorname{vec}(A)]^2}$ 的形式. 这说明表示高阶微分需要更强有力的数学工具，即下章中介绍的张量. 而且张量不仅可以表出高阶微分，还可以表出多元积分. 事实上，张量是现代微积分的常用描述语言，从而在几何学、力学、天文学等领域得到了广泛应用.

第 11 章　多重线性代数

第 7 章至第 8 章为第 1 章至第 6 章中最基本的概念进行了抽象化，使得线性代数不再仅仅局限于向量和矩阵，使得线性代数可以被应用到不同领域. 本章将把第 1 章至第 6 章中的其他一些概念抽象化，并得到更强有力的新概念.

为简化讨论，**如果不加特别说明，本章所涉及的线性空间都是有限维线性空间**. 事实上，本章定义的概念都可以直接在无限维线性空间上建立而不会引起问题，只有由基的存在性推出的性质才局限于有限维线性空间.

11.1　双线性函数

为简化说法，首先给出如下定义.

定义 11.1.1 (集合的直积)　给定集合 $S_1,\cdots,S_n$，集合

$$S_1\times\cdots\times S_n := \{(x_1,\cdots,x_n)\mid x_i\in S_i, i=1,\cdots,n\}$$

称为 $A_1,\cdots,A_n$ 的**直积**或 ***Descartes* 积**.

有了直积,多个自变量的映射的定义域就容易写出,如 n 阶行列式除写法 $\det\colon \mathbb{F}^{n\times n}\to\mathbb{F}$ 外，还可写作 $\det\colon \mathbb{F}^n\times\cdots\times\mathbb{F}^n\to\mathbb{F}$. 而欧氏空间的内积可写作 $\langle\cdot,\cdot\rangle\colon \mathcal{V}\times\mathcal{V}\to\mathbb{R}$；而酉空间的内积可写作 $\langle\cdot,\cdot\rangle\colon \mathcal{V}\times\mathcal{V}\to\mathbb{C}$.

欧氏空间的内积经推广可以得到双线性函数.

定义 11.1.2 (双线性函数)　设 $\mathcal{U},\mathcal{V}$ 是 $\mathbb{F}$ 上的线性空间，如果函数 $\varphi\colon \mathcal{U}\times\mathcal{V}\to\mathbb{F}$ 满足：对任意 $k_1,k_2\in\mathbb{F},\boldsymbol{a}_1,\boldsymbol{a}_2\in\mathcal{U},\boldsymbol{b}_1,\boldsymbol{b}_2\in\mathcal{V}$，

$$\begin{aligned}\varphi(k_1\boldsymbol{a}_1+k_2\boldsymbol{a}_2,\boldsymbol{b}_1)&=k_1\varphi(\boldsymbol{a}_1,\boldsymbol{b}_1)+k_2\varphi(\boldsymbol{a}_2,\boldsymbol{b}_1),\\ \varphi(\boldsymbol{a}_1,k_1\boldsymbol{b}_1+k_2\boldsymbol{b}_2)&=k_1\varphi(\boldsymbol{a}_1,\boldsymbol{b}_1)+k_2\varphi(\boldsymbol{a}_1,\boldsymbol{b}_2),\end{aligned}$$

则称 φ 是 $\mathcal{U}\times\mathcal{V}$ 上的一个**双线性函数**. $\mathcal{U}\times\mathcal{V}$ 上双线性函数的全体记为 $\mathscr{L}(\mathcal{U},\mathcal{V})$.

特别地，$\mathcal{U}\times\mathcal{U}$ 上的双线性函数简称 $\mathcal{U}$ 上的**双线性函数**.

对 $\mathcal{U}$ 上双线性函数 φ，如果对任意 $\boldsymbol{a},\boldsymbol{b}\in\mathcal{U}$，$\varphi(\boldsymbol{a},\boldsymbol{b})=\varphi(\boldsymbol{b},\boldsymbol{a})$，则称 φ 是 $\mathcal{U}$ 上的一个**对称双线性函数**；如果对任意 $\boldsymbol{a},\boldsymbol{b}\in\mathcal{U}$，$\varphi(\boldsymbol{a},\boldsymbol{b})=-\varphi(\boldsymbol{b},\boldsymbol{a})$，则称 φ 是 $\mathcal{U}$ 上的一个**反对称双线性函数**或**交错双线性函数**.

为形式计，有时我们也写作 $\varphi(\boldsymbol{x},\boldsymbol{y})=\varphi\left(\boldsymbol{x}\mid\boldsymbol{y}\right)$.

不难看出，欧氏空间 $\mathcal{V}$ 上的内积是 $\mathcal{V}$ 上一个对称双线性函数.

例 11.1.3 1. 对 $m\times n$ 矩阵 A, 两个数组向量空间 $\mathbb{F}^n\times\mathbb{F}^m$ 上的函数 $\boldsymbol{b}^{\mathrm{T}}A\boldsymbol{a}$ 是双线性函数.

2. 对 $m\times n$ 矩阵 A，两个矩阵空间 $\mathbb{F}^{n\times p}\times\mathbb{F}^{m\times p}$ 上的函数 $\operatorname{trace}(Y^{\mathrm{T}}AX)$ 是双线性函数.

3. 给定 $x_1,x_2\in\mathbb{F}$，多项式空间 $\mathbb{F}[x]_n\times\mathbb{F}[x]_m$ 上的函数 $\varphi(p,q)\mapsto p(x_1)q(x_2)$ 是双线性函数.

4. 给定区间 $[a,b]$,光滑函数空间 $C^\infty[a,b]$ 上的二元函数 $\int_a^b f(x)g'(x)\,\mathrm{d}x=\int_a^b f(x)\,\mathrm{d}g(x)$ 是双线性函数. ☺

酉空间的内积可以推广出倍半线性函数.

定义 11.1.4 (倍半线性函数) 设 $\mathcal{U},\mathcal{V}$ 是 $\mathbb{F}$ 上的线性空间，如果函数 $\varphi\colon\mathcal{U}\times\mathcal{V}\to\mathbb{F}$ 满足：对任意 $k_1,k_2\in\mathbb{F},\boldsymbol{a}_1,\boldsymbol{a}_2\in\mathcal{U},\boldsymbol{b}_1,\boldsymbol{b}_2\in\mathcal{V}$，

$$\begin{aligned}\varphi(k_1\boldsymbol{a}_1+k_2\boldsymbol{a}_2,\boldsymbol{b}_1)&=k_1\varphi(\boldsymbol{a}_1,\boldsymbol{b}_1)+k_2\varphi(\boldsymbol{a}_2,\boldsymbol{b}_1),\\ \varphi(\boldsymbol{a}_1,k_1\boldsymbol{b}_1+k_2\boldsymbol{b}_2)&=\overline{k}_1\varphi(\boldsymbol{a}_1,\boldsymbol{b}_1)+\overline{k}_2\varphi(\boldsymbol{a}_1,\boldsymbol{b}_2),\end{aligned}$$

则称 φ 是 $\mathcal{U}\times\mathcal{V}$ 上的一个**倍半线性函数**.

特别地，如果 $\mathcal{U}=\mathcal{V}$，则直称 $\mathcal{V}\times\mathcal{V}$ 上的倍半线性函数为 $\mathcal{V}$ 上的倍半线性函数.

二者的性质十分类似，因此下面重点讲述双线性函数的性质，倍半线性函数的性质留给读者思考.

11.1.1 双线性函数空间

双线性函数可以自然地定义线性运算.

定义 11.1.5 (线性运算) 给定数域 $\mathbb{F}$ 上的线性空间 $\mathcal{U},\mathcal{V}$，$\mathcal{U}\times\mathcal{V}$ 上双线性函数全体是 $\mathscr{L}(\mathcal{U},\mathcal{V})$. 规定

1. $\mathscr{L}(\mathcal{U},\mathcal{V})$ 上的**加法**：给定 $\varphi,\psi\in\mathscr{L}(\mathcal{U},\mathcal{V})$，定义

$$\begin{aligned}\varphi+\psi\colon\ \ &\mathcal{U}\times\mathcal{V}\to\mathbb{F},\\ &(\boldsymbol{x},\boldsymbol{y})\mapsto\varphi(\boldsymbol{x},\boldsymbol{y})+\psi(\boldsymbol{x},\boldsymbol{y}),\end{aligned}$$

2. $\mathscr{L}(\mathcal{U},\mathcal{V})$ 上的**数乘**：给定 $\varphi\in\mathscr{L}(\mathcal{U},\mathcal{V}), k\in\mathbb{F}$，定义

$$\begin{aligned} k\varphi\colon\ & \mathcal{U}\times\mathcal{V}\to\mathbb{F},\\ & (\boldsymbol{x},\boldsymbol{y})\mapsto k\varphi(\boldsymbol{x},\boldsymbol{y}). \end{aligned}$$

命题 11.1.6 集合 $\mathscr{L}(\mathcal{U},\mathcal{V})$ 关于加法和数乘两种运算构成线性空间.

证明留为作业.

在 8.1 节中我们曾把内积用其度量矩阵表示出来. 下面我们来说明，有限维线性空间上的双线性函数都可以类似地用矩阵来表示.

命题 11.1.7 给定线性空间 $\mathcal{U},\mathcal{V}$ 及各自一组基 $\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n;\boldsymbol{i}_1,\boldsymbol{i}_2,\cdots,\boldsymbol{i}_m$，则：

1. $\mathcal{U}\times\mathcal{V}$ 上双线性函数 φ 由它在两组基处的函数值 $\varphi(\boldsymbol{e}_i,\boldsymbol{i}_j), i=1,\cdots,n, j=1,\cdots,m$ 唯一确定，亦即，如果又有 $\mathcal{U}\times\mathcal{V}$ 上双线性函数 ψ 使得 $\psi(\boldsymbol{e}_i,\boldsymbol{i}_j)=f(\boldsymbol{e}_i,\boldsymbol{i}_j), i=1,\cdots,n, j=1,\cdots,m$，那么对任意 $\boldsymbol{a}\in\mathcal{U},\boldsymbol{b}\in\mathcal{V}$，$\psi(\boldsymbol{a},\boldsymbol{b})=\varphi(\boldsymbol{a},\boldsymbol{b})$；

2. 任给 $\mathbb{F}$ 上 $m\times n$ 矩阵 $F=\begin{bmatrix}f_{ij}\end{bmatrix}$，必存在唯一的 $\varphi\in\mathscr{L}(\mathcal{U},\mathcal{V})$，使得 $\varphi(\boldsymbol{e}_i,\boldsymbol{i}_j)=f_{ji}, i=1,\cdots,n, j=1,\cdots,m$.

证 第 1 条: 对任意 $\boldsymbol{a}\in\mathcal{U},\boldsymbol{b}\in\mathcal{V}$, 如果 $\boldsymbol{a}=a_1\boldsymbol{e}_1+\cdots+a_n\boldsymbol{e}_n, \boldsymbol{b}=b_1\boldsymbol{i}_1+\cdots+b_m\boldsymbol{i}_m$, 那么 $\varphi(\boldsymbol{a},\boldsymbol{b})=\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m}a_ib_j\varphi(\boldsymbol{e}_i,\boldsymbol{i}_j)=\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m}a_ib_j\psi(\boldsymbol{e}_i,\boldsymbol{i}_j)=\psi(\boldsymbol{a},\boldsymbol{b})$.

第 2 条：先定义二元函数

$$\begin{aligned} \varphi\colon\quad & \mathcal{U}\times\mathcal{V}\to\mathbb{F},\\ & \left(x_1\boldsymbol{e}_1+\cdots+x_n\boldsymbol{e}_n, y_1\boldsymbol{i}_1+\cdots+y_m\boldsymbol{i}_m\right)\mapsto\sum_{i=1}^{n}\sum_{j=1}^{m}x_iy_jf_{ji},\\ \text{或者写成}\quad & \Big((\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n)\widehat{\boldsymbol{x}},(\boldsymbol{i}_1,\cdots,\boldsymbol{i}_m)\widehat{\boldsymbol{y}}\Big)\mapsto\widehat{\boldsymbol{y}}^{\mathrm{T}}F\widehat{\boldsymbol{x}}. \end{aligned}$$

只需证明它是一个双线性函数，因为显然 $\varphi(\boldsymbol{e}_i,\boldsymbol{i}_j)=f_{ji}, i=1,\cdots,n, j=1,\cdots,m$，而唯一性由第一条保证. **其余留为作业.** □

类似于线性映射的表示矩阵，称

$$F=\begin{bmatrix}\varphi(\boldsymbol{e}_1,\boldsymbol{i}_1) & \cdots & \varphi(\boldsymbol{e}_n,\boldsymbol{i}_1)\\ \vdots & & \vdots\\ \varphi(\boldsymbol{e}_1,\boldsymbol{i}_m) & \cdots & \varphi(\boldsymbol{e}_n,\boldsymbol{i}_m)\end{bmatrix}=:\varphi\left(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n\,|\,\boldsymbol{i}_1,\cdots,\boldsymbol{i}_m\right)$$

为双线性函数 φ 在基 $\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n;\boldsymbol{i}_1,\cdots,\boldsymbol{i}_m$ 下的**度量矩阵**.

不难看出对称双线性函数的度量矩阵是对称矩阵，反对称双线性函数的度量矩阵是反对称矩阵.

类似于线性映射的矩阵参与的形式计算，这里也有双线性函数的形式计算：

$$\varphi\left((\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n)\widehat{\boldsymbol{x}} \mid (\boldsymbol{i}_1,\cdots,\boldsymbol{i}_m)\widehat{\boldsymbol{y}}\right)=\varphi\left(\boldsymbol{x}\mid\boldsymbol{y}\right)=\widehat{\boldsymbol{y}}^{\mathrm{T}}F\widehat{\boldsymbol{x}}=\widehat{\boldsymbol{y}}^{\mathrm{T}}\varphi\left(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n\mid\boldsymbol{i}_1,\cdots,\boldsymbol{i}_m\right)\widehat{\boldsymbol{x}}.$$

例 11.1.8 下面给出一些双线性函数的度量矩阵的例子.

1. 考虑线性空间 $\mathbb{R}^4$ 上的双线性函数 $\varphi(\begin{bmatrix}x_1\\y_1\\z_1\\t_1\end{bmatrix},\begin{bmatrix}x_2\\y_2\\z_2\\t_2\end{bmatrix})=x_1x_2+y_1y_2+z_1z_2-t_1t_2$.

   它在基 $\begin{bmatrix}1\\0\\0\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\\0\end{bmatrix},\begin{bmatrix}0\\0\\1\\0\end{bmatrix},\begin{bmatrix}0\\0\\0\\1\end{bmatrix}$ 下的度量矩阵就是 $\begin{bmatrix}1&&&\\&1&&\\&&1&\\&&&-1\end{bmatrix}$. 这个双线性函数就是 Minkowski 时空空间[①]用来定义距离的函数.

2. 考虑 $\mathbb{R}^{2\times 2}$ 上的双线性函数 $\varphi(X,Y)=\operatorname{trace}(Y^{\mathrm{T}}AX)$, 它在基 $\begin{bmatrix}1&0\\0&0\end{bmatrix},\begin{bmatrix}0&0\\1&0\end{bmatrix},$ $\begin{bmatrix}0&1\\0&0\end{bmatrix},\begin{bmatrix}0&0\\0&1\end{bmatrix}$ 下的度量矩阵是 $\begin{bmatrix}a_{11}&a_{12}&0&0\\a_{21}&a_{22}&0&0\\0&0&a_{11}&a_{12}\\0&0&a_{21}&a_{22}\end{bmatrix}=\begin{bmatrix}A&\\&A\end{bmatrix}$.

3. 给定 $x_1,x_2\in\mathbb{F}$，考虑多项式空间 $\mathbb{F}[x]_2\times\mathbb{F}[x]_3$ 上的双线性函数 $\varphi(p,q)=p(x_1)q(x_2)$，它在基 $1,x-1;1,x,x^2$ 下的度量矩阵是 $\begin{bmatrix}1&x_1-1\\x_2&(x_1-1)x_2\\x_2^2&(x_1-1)x_2^2\end{bmatrix}$.

4. 考虑多项式空间 $\mathbb{R}[x]_2\times\mathbb{R}[x]_3$ 上的双线性函数 $\varphi(p,q)=\int_0^1 p(x)q'(x)\,\mathrm{d}x$，它在它在基 $1,x-1;1,x,x^2$ 下的度量矩阵是 $\begin{bmatrix}0&0\\1&-\frac{1}{2}\\1&-\frac{1}{3}\end{bmatrix}$. ☺

定义映射

$$\begin{aligned}\tau=\tau_{\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n;\boldsymbol{i}_1,\boldsymbol{i}_2,\cdots,\boldsymbol{i}_m}:\quad \mathscr{L}(\mathcal{U},\mathcal{V})&\to\mathbb{F}^{m\times n},\\ \varphi&\mapsto\varphi\left(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n\mid\boldsymbol{i}_1,\cdots,\boldsymbol{i}_m\right).\end{aligned}$$

类似于线性映射组成的线性空间 $\operatorname{Hom}(\mathcal{U},\mathcal{V})$ 与 $\mathbb{F}^{m\times n}$ 的同构，可有如下结论.

定理 11.1.9 映射 $\tau=\tau_{\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n;\boldsymbol{i}_1,\boldsymbol{i}_2,\cdots,\boldsymbol{i}_m}$ 是 $\mathscr{L}(\mathcal{U},\mathcal{V})$ 到 $\mathbb{F}^{m\times n}$ 的同构映射. 特别地，$\dim\mathscr{L}(\mathcal{U},\mathcal{V})=\dim\mathbb{F}^{m\times n}=mn$.

① 细节请参看狭义相对论.

证 首先说明 τ 是双射. 命题 11.1.7 第 1 条说明不同双线性函数的度量矩阵不同，即 τ 是单射；命题 7.5.1 第 2 条说明对任意矩阵都可以找到一个双线性函数使其矩阵就是给定矩阵，即 τ 是满射. 其次说明 τ 是线性映射. 设 $\tau(\varphi)=F,\tau(\psi)=G$，注意到 $(k\varphi+l\psi)(\boldsymbol{e}_i,\boldsymbol{i}_j)=k\varphi(\boldsymbol{e}_i,\boldsymbol{i}_j)+l\psi(\boldsymbol{e}_i,\boldsymbol{i}_j)=kf_{ji}+lg_{ji},i=1,\cdots,n,j=1,\cdots,m$，就有 $\tau(k\varphi+l\psi)=kF+lG=k\tau(\varphi)+l\tau(\psi)$. □

下面讨论双线性函数的度量矩阵在基变换下的变化规律.

命题 11.1.10 给定数域 $\mathbb{F}$ 上线性空间 $\mathcal{U},\mathcal{V}$，和 $\mathcal{U}$ 的两组基 $\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n;\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_n$ 与 $\mathcal{V}$ 的两组基 $\boldsymbol{i}_1,\boldsymbol{i}_2,\cdots,\boldsymbol{i}_m;\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots,\boldsymbol{s}_m$. 记二者的过渡矩阵分别为 T,S，即

$$(\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_n)=(\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n)T,\quad(\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots,\boldsymbol{s}_m)=(\boldsymbol{i}_1,\boldsymbol{i}_2,\cdots,\boldsymbol{i}_m)S.$$

如果 $\mathcal{U}\times\mathcal{V}$ 上双线性函数 φ 在基 $\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n;\boldsymbol{i}_1,\boldsymbol{i}_2,\cdots,\boldsymbol{i}_m$ 下的矩阵为 F，则 φ 在基 $\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_n;\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots,\boldsymbol{s}_m$ 下的矩阵为 $S^{\mathrm{T}}FT$.

证 设 φ 在基 $\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_n;\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots,\boldsymbol{s}_m$ 下的度量矩阵为 $\widetilde{F}$. 利用形式上的结合律，有

$$\begin{aligned}
\tilde{\boldsymbol{y}}^{\mathrm{T}}\widetilde{F}\widetilde{\boldsymbol{x}}&=\tilde{\boldsymbol{y}}^{\mathrm{T}}\varphi\left(\boldsymbol{t}_1,\cdots,\boldsymbol{t}_n\mid\boldsymbol{s}_1,\cdots,\boldsymbol{s}_m\right)\widetilde{\boldsymbol{x}}\\
&=\varphi\left((\boldsymbol{t}_1,\cdots,\boldsymbol{t}_n)\widetilde{\boldsymbol{x}}\mid(\boldsymbol{s}_1,\cdots,\boldsymbol{s}_m)\tilde{\boldsymbol{y}}\right)\\
&=\varphi\left((\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n)T\widetilde{\boldsymbol{x}}\mid(\boldsymbol{i}_1,\cdots,\boldsymbol{i}_n)S\tilde{\boldsymbol{y}}\right)\\
&=(S\tilde{\boldsymbol{y}})^{\mathrm{T}}\varphi\left(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n\mid\boldsymbol{i}_1,\cdots,\boldsymbol{i}_m\right)(T\widetilde{\boldsymbol{x}})\\
&=\tilde{\boldsymbol{y}}^{\mathrm{T}}S^{\mathrm{T}}FT\widetilde{\boldsymbol{x}}.
\end{aligned}$$

因此 $\widetilde{F}$ 和 $S^{\mathrm{T}}FT$ 是同一个 $\mathbb{F}^n\times\mathbb{F}^m$ 上双线性函数的度量矩阵. 根据定理 11.1.9，即知 $\widetilde{F}=S^{\mathrm{T}}FT$. □

命题 11.1.10 给出了如下形式计算：

$$\varphi\left((\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n)X\mid(\boldsymbol{i}_1,\cdots,\boldsymbol{i}_m)Y\right)=Y^{\mathrm{T}}\varphi\left(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n\mid\boldsymbol{i}_1,\cdots,\boldsymbol{i}_m\right)X.$$

容易验证，这对任意方阵 X,Y 都成立.

推论 11.1.11 给定数域 $\mathbb{F}$ 上线性空间 $\mathcal{U}$ 和它的两组基 $\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n;\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_n$. 记二者的过渡矩阵分别为 T，即 $(\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_n)=(\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n)T$. 如果 $\mathcal{U}$ 上双线性函数 φ 在基 $\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n$ 下的矩阵为 F，则 φ 在基 $\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_n$ 下的矩阵为 $T^{\mathrm{T}}FT$.

可以看到，**线性空间做基变换，其上双线性函数的度量矩阵就做合同变换**. 事实上，有如下结论.

命题 11.1.12 数域 $\mathbb{F}$ 上两个 n 阶方阵 A,B 合同，当且仅当 A,B 是 n 维线性空间 $\mathcal{U}$ 上某个双线性函数在两组基下的度量矩阵.

证 “$\Leftarrow$”：由推论 11.1.11 立得.

“$\Rightarrow$”：设 $B=T^{\mathrm{T}}AT$，其中 T 可逆. 首先根据命题 11.1.7 第 2 条的证明，我们可以找到 $\mathcal{V}$ 上双线性函数 φ 使得在它在 $\mathcal{V}$ 的某组基，记为 $\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n$，下的度量矩阵为 F. 令 $(\boldsymbol{t}_1,\cdots,\boldsymbol{t}_n)=(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n)T$，由于 T 可逆，根据命题 7.4.6 ，$\boldsymbol{t}_1,\cdots,\boldsymbol{t}_n$ 是 $\mathcal{V}$ 的一组基. 根据命题 11.1.10 ，f 在这组基下的矩阵就是 $T^{\mathrm{T}}AT=B$. □

命题 11.1.12 说明，计算 n 阶矩阵 A 的合同标准形，本质上就是希望为 $\mathbb{F}^n$ 找到一组基，使得 A 对应的双线性函数在这组基下的度量矩阵尽可能简单.

11.1.2 合同标准形

如果 A 是对称矩阵，根据定理 11.1.13 ，这个尽可能简单的度量矩阵一定是对角阵.

定理 11.1.13 给定对称矩阵 $A\in\mathbb{F}^{n\times n}$，则存在可逆矩阵 $X\in\mathbb{F}^{n\times n}$，使得 $X^{\mathrm{T}}AX=D$ 是对角矩阵.

证 采用数学归纳法. $n=1$ 时显然成立. 假设命题对 $n-1$ 阶对称矩阵成立，考察 n 阶矩阵 $A=\begin{bmatrix}a_{ij}\end{bmatrix}$. 分如下情形讨论.

1. 如果 $a_{11}\neq 0$, 那么对分块矩阵用 Gauss 消元法，直接验证或者利用 (1.6.5) 式，有
$$\begin{bmatrix}1 & 0\\ -\frac{1}{a_{11}}A_{21} & I\end{bmatrix}\begin{bmatrix}a_{11} & A_{21}^{\mathrm{T}}\\ A_{21} & A_{22}\end{bmatrix}\begin{bmatrix}I & -\frac{1}{a_{11}}A_{21}^{\mathrm{T}}\\ 0 & I\end{bmatrix}=\begin{bmatrix}a_{11} & 0\\ 0 & A_{22}-\frac{1}{a_{11}}A_{21}A_{21}^{\mathrm{T}}\end{bmatrix}.$$
利用归纳假设，存在可逆矩阵 X_2，使得 $X_2^{\mathrm{T}}(A_{22}-\frac{1}{a_{11}}A_{21}A_{21}^{\mathrm{T}})X_2=D_2$ 是对角矩阵. 那么令 $X=\begin{bmatrix}I & -\frac{1}{a_{11}}A_{21}^{\mathrm{T}}\\ 0 & I\end{bmatrix}\begin{bmatrix}1 & \\ & X_2\end{bmatrix}$，即得 $X^{\mathrm{T}}AX=\begin{bmatrix}a_{11} & \\ & D_2\end{bmatrix}$ 是对角矩阵.
2. 如果 $a_{11}=0$，但存在 $a_{ii}\neq 0$，那么利用置换矩阵 P_{1i}，即可使得 $P_{1i}^{\mathrm{T}}AP_{1i}$ 的左上角元素不为 0，转化为第一种情形.
3. 如果 $a_{ii}=0,i=1,\cdots,n$，但 $a_{21}\neq 0$，那么取 $X_{11}=\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$，则有
$$\begin{bmatrix}X_{11} & \\ & I_{n-2}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}A_{11} & A_{21}^{\mathrm{T}}\\ A_{21} & A_{22}\end{bmatrix}\begin{bmatrix}X_{11} & \\ & I_{n-2}\end{bmatrix}=\begin{bmatrix}X_{11}^{\mathrm{T}}A_{11}X_{11} & X_{11}A_{21}^{\mathrm{T}}\\ A_{21}X_{11} & A_{22}\end{bmatrix}$$

其中 $X_{11}^{\mathrm{T}}A_{11}X_{11}=\begin{bmatrix}1&1\\1&-1\end{bmatrix}\begin{bmatrix}0&a_{21}\\a_{21}&0\end{bmatrix}\begin{bmatrix}1&1\\1&-1\end{bmatrix}=\begin{bmatrix}2a_{21}&\\&-2a_{21}\end{bmatrix}$. 转化为第一种情形.

4. 如果 $a_{ii}=0,i=1,\cdots,n,a_{12}=0$，但存在 $a_{1i}\neq 0$，那么利用置换矩阵 P_{2i}，转化为上一种情形.

5. 若不然，则 A 的对角元和 A 的第一行第一列全为 0，即 $A=\begin{bmatrix}0&0\\0&A_{22}\end{bmatrix}$，利用归纳假设，存在可逆矩阵 X_2，使得 $X_2^{\mathrm{T}}A_{22}X_2=D_2$ 是对角矩阵. 那么令 $X=\begin{bmatrix}1&\\&X_2\end{bmatrix}$，即得 $X^{\mathrm{T}}AX=\begin{bmatrix}0&\\&D_2\end{bmatrix}$ 是对角矩阵.

综上即得结论. □

显然，对实对称矩阵，定理 11.1.13 可以直接由定理 6.1.2 得到.

然后我们就可以给出合同变换下对称矩阵的标准形.

推论 11.1.14 给定对称矩阵 $A\in\mathbb{C}^{n\times n}$，则存在可逆矩阵 $X\in\mathbb{C}^{n\times n}$，使得 $X^{\mathrm{T}}AX=\begin{bmatrix}I_r&\\&0\end{bmatrix}$，其中 $r=\operatorname{rank}(A)$.

证 根据定理 11.1.13 ，存在可逆矩阵 X_1，使得 $X_1^{\mathrm{T}}AX_1=D=\begin{bmatrix}d_{11}&&\\&\ddots&\\&&d_{nn}\end{bmatrix}$. 首先利用置换矩阵，记为 P，可以把所有为 0 的 D 的对角元换到右下角. 由于秩是合同变换下的不变量，我们知道为 0 的对角元的个数是 $n-r$. 因此 $P^{\mathrm{T}}X_1^{\mathrm{T}}AX_1P=\begin{bmatrix}D_1&\\&0\end{bmatrix}$，其中 $D_1=\begin{bmatrix}\tilde{d}_{11}&&\\&\ddots&\\&&\tilde{d}_{rr}\end{bmatrix}$ 的对角元都不为 0. 令 $X_2=\begin{bmatrix}r_{11}&&\\&\ddots&\\&&r_{rr}\end{bmatrix}$，其中 $r_{ii}^2=\tilde{d}_{ii},i=1,\cdots,r$，在复数域上，这样的 r_{ii} 总存在. 注意到 X_2 对称且可逆，因此 $\begin{bmatrix}X_2^{-1}&\\&I\end{bmatrix}P^{\mathrm{T}}X_1^{\mathrm{T}}AX_1P\begin{bmatrix}X_2^{-1}&\\&I\end{bmatrix}=\begin{bmatrix}I_r&\\&0\end{bmatrix}$. 令 $X=X_1P\begin{bmatrix}X_2^{-1}&\\&I\end{bmatrix}$ 即得结论. □

推论 11.1.14 中的 $\begin{bmatrix}I_r&\\&0\end{bmatrix}$ 称为复对称矩阵 A 的**合同标准形**. 复对称矩阵的合同标准形由它的秩唯一决定. 根据推论 11.1.14 ，n 阶复对称矩阵在合同变换下的等价类数目有限，只有 $n+1$ 类.

如下推论就是命题 6.2.8 ，这里给出了另一种证法.

推论 11.1.15 给定对称矩阵 $A \in \mathbb{R}^{n\times n}$，则存在可逆矩阵 $X \in \mathbb{R}^{n\times n}$，使得 $X^{\mathrm{T}}AX = J = \begin{bmatrix} I_p & & \\ & -I_{r-p} & \\ & & 0 \end{bmatrix}$，其中 $r = \operatorname{rank}(A), 0 \leqslant p \leqslant r$.

证 $\boxed{\textbf{留为作业.}}$ □

推论 11.1.15 中的 J 称为实对称矩阵 A 的**合同标准形**. 根据推论 11.1.15，n 阶实对称矩阵在合同变换下的等价类数目有限，只有 $\frac{(n+1)(n+2)}{2}$ 类.

注意，对任意的数域，其合同标准形与能否在数域中找到平方根有很大关系.

再来看反对称矩阵在合同变换下的标准形. 注意反对称矩阵秩一定为偶数（参见练习 2.3.21）.

定理 11.1.16 给定反对称矩阵 $A \in \mathbb{R}^{n\times n}$，设 $\operatorname{rank}(A) = 2r$. 则存在可逆矩阵 $X_1 \in \mathbb{R}^{n\times n}$，使得 $X_1^{\mathrm{T}}AX_1 = J_1 = \begin{bmatrix} J & & & \\ & \ddots & & \\ & & J & \\ & & & 0 \end{bmatrix}$，其中 $J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ 共有 r 个；也存在可逆矩阵 $X_2 \in \mathbb{R}^{n\times n}$，使得 $X_2^{\mathrm{T}}AX_2 = J_2 = \begin{bmatrix} & I_r & \\ -I_r & & \\ & & 0 \end{bmatrix}$.

$\boxed{\textbf{证明留为作业.}}$

11.2 多线性函数

上节阐述了两个线性空间上的双线性函数，而一个线性空间上的线性函数就是之前学习过的线性映射的特例. 本节就来认识多个线性空间上的多线性函数.

11.2.1 基本概念

定义 11.2.1 (多线性函数) 给定 $\mathbb{F}$ 上的线性空间 $\mathcal{U}_1,\cdots,\mathcal{U}_r$，若映射 $f\colon \mathcal{U}_1 \times \cdots \times \mathcal{U}_r \to \mathcal{R}, (\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r) \mapsto f(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r)$ 对每一个变元都有线性性质，即固定任意 $k-1$ 个变元 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_{p-1},\boldsymbol{a}_{p+1},\cdots,\boldsymbol{a}_r$ 的任意取值后，$\boldsymbol{w} \mapsto f(\cdots,\overset{p}{\boldsymbol{w}},\cdots)$ 是 $\mathcal{U}_p$ 上的线性函数，则称 f 是 $\mathcal{U}_1 \times \cdots \times \mathcal{U}_r$ 上的一个**多线性函数**或r **重线性函数**. $\mathcal{U}_1 \times \cdots \times \mathcal{U}_r$ 上的多线性函数的全体记为 $\mathscr{L}(\mathcal{U}_1,\cdots,\mathcal{U}_r)$.

特别地，$r=2$ 时就是定义 *11.1.2* 中的双线性函数，其全体是 $\mathscr{L}(\mathcal{U}_1,\mathcal{U}_2)$；$r=1$ 时就是线性函数，其全体是 $\mathscr{L}(\mathcal{U}_1)$.

特别地，$\mathscr{U}\times\cdots\times\mathscr{U}$ 上的多线性函数简称 $\mathscr{U}$ 上的**多线性函数**.

对 $\mathscr{U}$ 上多线性函数 φ, 如果对任意 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r\in\mathscr{U}$ 和任意排列 σ, $f(\boldsymbol{a}_{\sigma_1},\cdots,\boldsymbol{a}_{\sigma_r})=f(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r)$，则称 f 是 $\mathscr{U}$ 上的一个**对称多线性函数**；如果对任意 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r\in\mathscr{U}$ 和任意排列 σ，$f(\boldsymbol{a}_{\sigma_1},\cdots,\boldsymbol{a}_{\sigma_r})=\operatorname{sign}(\sigma)f(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r)$，则称 f 是 $\mathscr{U}$ 上的一个**反对称多线性函数**或**交错多线性函数**.

例 11.2.2 1. 如下函数是 $\mathbb{F}^4\times\mathbb{F}^3\times\mathbb{F}^2$ 上的三重线性函数：

$$\begin{aligned}f(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})=\ & f_{111}x_1y_1z_1+f_{211}x_2y_1z_1+f_{311}x_3y_1z_1+f_{411}x_4y_1z_1+\\ & f_{121}x_1y_2z_1+f_{221}x_2y_2z_1+f_{321}x_3y_2z_1+f_{421}x_4y_2z_1+\\ & f_{131}x_1y_3z_1+f_{231}x_2y_3z_1+f_{331}x_3y_3z_1+f_{431}x_4y_3z_1+\\ & f_{112}x_1y_1z_2+f_{212}x_2y_1z_2+f_{312}x_3y_1z_2+f_{412}x_4y_1z_2+\\ & f_{122}x_1y_2z_2+f_{222}x_2y_2z_2+f_{322}x_3y_2z_2+f_{422}x_4y_2z_2+\\ & f_{132}x_1y_3z_2+f_{232}x_2y_3z_2+f_{332}x_3y_3z_2+f_{432}x_4y_3z_2.\end{aligned}$$

2. 如下函数是 $\mathbb{F}^2$ 上的三重线性函数：

$$\begin{aligned}g(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})=\ & g_{111}x_1y_1z_1+g_{211}x_2y_1z_1+g_{121}x_1y_2z_1+g_{221}x_2y_2z_1+\\ & g_{112}x_1y_1z_2+g_{212}x_2y_1z_2+g_{122}x_1y_2z_2+g_{222}x_2y_2z_2.\end{aligned}$$

   函数 g 对称，当且仅当 $g_{112}=g_{121}=g_{211}, g_{122}=g_{212}=g_{221}$.
   函数 g 反对称，当且仅当 $g_{ijk}=0$，即 g 是零函数.

3. n 阶行列式函数 $\det$ 是 $\mathbb{F}^n$ 上的反对称多线性函数.
4. 若 $\mathscr{U}_1,\cdots,\mathscr{U}_r$ 都是欧氏空间，则如下函数是 $\mathscr{U}_1\times\mathscr{U}_2\times\cdots\times\mathscr{U}_r$ 上的多线性函数：

$$h\colon(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_r)\mapsto\langle\boldsymbol{x}_1,\boldsymbol{a}_1\rangle\cdots\langle\boldsymbol{x}_r,\boldsymbol{a}_r\rangle.$$

5. 矩阵空间 $\mathbb{F}^{m_1\times n_1}\times\cdots\times\mathbb{F}^{m_r\times n_r}$ 上的函数 $\operatorname{trace}(X_1A_1X_2A_2\cdots X_nA_n)$ 是多线性函数.
6. 给定区域上的无穷次可微的向量变元向量值函数的 Jacobi 行列式是 $\det\left[\frac{\partial f_j}{\partial x_i}\right]$ 是反对称多线性函数. ☺

多线性函数的线性运算可以类比于双线性函数去定义，并构成线性空间，且若各个线性空间都是有限维空间，则给定各个线性空间的基后，多线性函数也可以和一组数一一对应.

定义 11.2.3 (线性运算) 给定数域 $\mathbb{F}$ 上的线性空间 $\mathscr{U}_1,\cdots,\mathscr{U}_r$，$\mathscr{U}_1\times\cdots\times\mathscr{U}_r$ 上双线性函数全体是 $\mathscr{L}(\mathscr{U}_1,\cdots,\mathscr{U}_r)$. 规定

1. $\mathscr{L}(\mathcal{U}_1,\cdots,\mathcal{U}_r)$ 上的**加法**：给定 $f,g\in\mathscr{L}(\mathcal{U}_1,\cdots,\mathcal{U}_r)$，定义

$$\begin{aligned} f+g\colon\ &\mathcal{U}_1\times\cdots\times\mathcal{U}_r\to\mathbb{F},\\ &(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r)\mapsto f(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r)+g(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r),\end{aligned}$$

2. $\mathscr{L}(\mathcal{U}_1,\cdots,\mathcal{U}_r)$ 上的**数乘**：给定 $f\in\mathscr{L}(\mathcal{U}_1,\cdots,\mathcal{U}_r),k\in\mathbb{F}$，定义

$$\begin{aligned} kf\colon\ &\mathcal{U}_1\times\cdots\times\mathcal{U}_r\to\mathbb{F},\\ &(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r)\mapsto kf(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r).\end{aligned}$$

命题 11.2.4 集合 $\mathscr{L}(\mathcal{U}_1,\cdots,\mathcal{U}_r)$ 关于加法和数乘两种运算构成线性空间.

命题 11.2.5 给定线性空间 $\mathcal{U}_1,\cdots,\mathcal{U}_r$ 及各自一组基 $\boldsymbol{e}_1^{(1)},\boldsymbol{e}_2^{(1)},\cdots,\boldsymbol{e}_{n_1}^{(1)};\cdots;\boldsymbol{e}_1^{(r)},\boldsymbol{e}_2^{(r)},\cdots,\boldsymbol{e}_{n_r}^{(r)}$，则：

1. $\mathcal{U}_1\times\cdots\times\mathcal{U}_r$ 上多线性函数 f 由它在指定基处的函数值 $f(\boldsymbol{e}_{i_1}^{(1)},\cdots,\boldsymbol{e}_{i_r}^{(r)}),i_p=1,\cdots,n_p,p=1,\cdots,r$ 唯一确定，亦即，如果又有 $\mathcal{U}_1\times\cdots\times\mathcal{U}_r$ 上多线性函数 g 使得 $g(\boldsymbol{e}_{i_1}^{(1)},\cdots,\boldsymbol{e}_{i_r}^{(r)})=f(\boldsymbol{e}_{i_1}^{(1)},\cdots,\boldsymbol{e}_{i_r}^{(r)}),i_p=1,\cdots,n_p,p=1,\cdots,r$，那么对任意 $\boldsymbol{a}_p\in\mathcal{U}_p,p=1,\cdots,r$，$g(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r)=f(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r)$；
2. 任给 $\mathbb{F}$ 上 $n_1n_2\cdots n_r$ 个数 $f_{i_1i_2\cdots i_r},i_p=1,\cdots,n_p,p=1,\cdots,r$，必存在唯一的 $f\in\mathscr{L}(\mathcal{U}_1,\cdots,\mathcal{U}_r)$，使得 $f(\boldsymbol{e}_{i_1}^{(1)},\cdots,\boldsymbol{e}_{i_r}^{(r)})=f_{i_1i_2\cdots i_r}$.

证明留为作业.

我们常常把这 $n_1n_2\cdots n_r$ 个数码放成一个 r 维 $n_1\times n_2\times\cdots\times n_r$ 数组，即有 n_1 行 n_2 列 n_3 片 n_4 叠……的数组. 为指称方便，常称这种高维数组为r **阶张量**，而所有 $n_1\times n_2\times\cdots\times n_r$ 张量的全体记为 $\mathbb{F}^{n_1\times n_2\times\cdots\times n_r}$. 本书中用大写字母表示张量. 其中的数称为**元素**，张量 A 在第 i_1 行第 i_2 列第 i_3 片第 i_4 叠……的元素，简称 $(i_1,i_2,i_3,i_4,\cdots)$ 元素，记为 $a_{i_1i_2i_3i_4\cdots}$. 若张量的元素满足对任意排列 σ，$a_{\sigma_{i_1}\cdots\sigma_{i_n}}=a_{i_1\cdots i_n}$，则称其为**对称张量**；若张量的元素满足对任意排列 σ，$a_{\sigma_{i_1}\cdots\sigma_{i_n}}=\operatorname{sign}(\sigma)a_{i_1\cdots i_n}$，则称其为**反对称张量**或**交错张量**. 张量的线性运算可以和矩阵类似定义，即逐元素地作加法和数乘.

命题 11.2.6 集合 $\mathbb{F}^{n_1\times n_2\times\cdots\times n_r}$ 关于加法和数乘两种运算构成线性空间，而 $(j_1,\cdots,j_r)$ 元素为 1，其他元素都是 0 的张量 $E_{j1\cdots j_r},j_p=1,\cdots,n_p,p=1,\cdots,r$，是它的一组基，从而维数是 $n_1n_2\cdots n_r$.

证 **留为作业.** □

类似于双线性函数的度量矩阵，码放 $f(\boldsymbol{e}_{i_1}^{(1)},\cdots,\boldsymbol{e}_{i_r}^{(r)})$ 的张量称为多线性函数 f 在基 $\boldsymbol{e}_1^{(p)},\cdots,\boldsymbol{e}_{n_p}^{(p)},p=1,\cdots,r$ 下的**度量张量**. 若多线性函数对称，则其度量张量对称；若多线性函数反对称，则其度量张量反对称.

例 11.2.7 下面给出例 11.2.2 中多线性函数在标准基下的度量张量.

1. f 的度量张量为 $F = \begin{bmatrix} f_{111} & f_{121} & f_{131} \\ f_{211} & f_{221} & f_{231} \\ f_{311} & f_{321} & f_{331} \\ f_{411} & f_{421} & f_{431} \end{bmatrix}\begin{bmatrix} f_{112} & f_{122} & f_{132} \\ f_{212} & f_{222} & f_{232} \\ f_{312} & f_{322} & f_{332} \\ f_{412} & f_{422} & f_{432} \end{bmatrix}$. 为方便书写，将其改写成 $F =$
$$\begin{array}{cccccc} 1' & & & 2' & & \\ \end{array}$$
$$\left[\begin{array}{ccc|ccc} f_{111} & f_{121} & f_{131} & f_{112} & f_{122} & f_{132} \\ f_{211} & f_{221} & f_{231} & f_{212} & f_{222} & f_{232} \\ f_{311} & f_{321} & f_{331} & f_{312} & f_{322} & f_{332} \\ f_{411} & f_{421} & f_{431} & f_{412} & f_{422} & f_{432} \end{array}\right]$$
，其中竖线和小数字提示了第某片包含的 2 维数组（即矩阵).

2. g 的度量张量为 $G = \begin{array}{c} \begin{array}{cc|cc} 1' & & 2' & \end{array} \\ \left[\begin{array}{cc|cc} g_{111} & g_{121} & g_{112} & g_{122} \\ g_{211} & g_{221} & g_{212} & g_{222} \end{array}\right] \end{array}$,若 g 对称,则 $G = \begin{array}{c} \begin{array}{cc|cc} 1' & & 2' & \end{array} \\ \left[\begin{array}{cc|cc} a & b & b & c \\ b & c & c & d \end{array}\right] \end{array}$.

3. 二阶行列式的度量张量为 $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$，三阶行列式的度量张量为
$$\begin{array}{c} \begin{array}{ccc|ccc|ccc} 1' & & & 2' & & & 3' & & \end{array} \\ \left[\begin{array}{ccc|ccc|ccc} 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \end{array}.$$
☺

显见一阶张量是向量, 二阶张量是矩阵. 上面粗浅地写了几个三阶的张量的例子, 可见其繁琐，更高阶的张量将更加复杂.

下面就来找出线性空间上多线性函数空间的一组基.

命题 11.2.8 给定线性空间 $\mathcal{U}_1, \cdots, \mathcal{U}_r$ 及各自一组基 $\boldsymbol{e}_1^{(1)}, \boldsymbol{e}_2^{(1)}, \cdots, \boldsymbol{e}_{n_1}^{(1)}; \cdots; \boldsymbol{e}_1^{(r)}, \boldsymbol{e}_2^{(r)}, \cdots, \boldsymbol{e}_{n_r}^{(r)}$，则映射
$$\begin{aligned} \tau = \tau_{\boldsymbol{e}_1^{(1)}, \boldsymbol{e}_2^{(1)}, \cdots, \boldsymbol{e}_{n_1}^{(1)}; \cdots; \boldsymbol{e}_1^{(r)}, \boldsymbol{e}_2^{(r)}, \cdots, \boldsymbol{e}_{n_r}^{(r)}} : \quad & \mathscr{L}(\mathcal{U}_1, \cdots, \mathcal{U}_r) \to \mathbb{F}^{n_1 \times \cdots \times n_r}, \\ & f \mapsto \left[f(\boldsymbol{e}_{i_1}^{(1)}, \cdots, \boldsymbol{e}_{i_r}^{(r)}) \right]_{n_1 \times \cdots \times n_r}, \end{aligned}$$
是同构映射．特别地，$\dim \mathscr{L}(\mathcal{U}_1, \cdots, \mathcal{U}_r) = n_1 \cdots n_r$.

证 $\boxed{\textbf{留为作业.}}$ □

已经知道 $E_{j_1 \cdots j_r}, j_p = 1, \cdots, n_p, p = 1, \cdots, r$ 是 $\mathbb{F}^{n_1 \times \cdots \times n_r}$ 的一组基,利用命题 11.2.8 中的同构映射,就可以得到 $\mathscr{L}(\mathcal{U}_1, \cdots, \mathcal{U}_r)$ 的一组基 $\varphi_{j_1 \cdots j_r} = \tau^{-1}(E_{j_1 \cdots j_r}), j_p = 1, \cdots, n_p, p = 1, \cdots, r$.

立得如下结论.

命题 11.2.9 给定线性空间 $\mathcal{U}_1,\cdots,\mathcal{U}_r$ 及各自一组基 $\boldsymbol{e}_1^{(1)},\boldsymbol{e}_2^{(1)},\cdots,\boldsymbol{e}_{n_1}^{(1)};\cdots;\boldsymbol{e}_1^{(r)},\boldsymbol{e}_2^{(r)},\cdots,\boldsymbol{e}_{n_r}^{(r)}$，则满足如下条件的唯一一组多线性函数是 $\mathscr{L}(\mathcal{U}_1,\cdots,\mathcal{U}_r)$ 的一组基：

$$\varphi_{j_1\cdots j_r}(\boldsymbol{e}_{i_1}^{(1)},\cdots,\boldsymbol{e}_{i_r}^{(r)})=\delta_{i_1j_1}\cdots\delta_{i_rj_r},i_p=1,\cdots,n_p,p=1,\cdots,r,$$

其中 $\delta_{\cdot\cdot}$ 是例 *8.1.9* 中提及的 ***Kronecker* 符号**

$$\delta_{mn}=\begin{cases}1, & 若\ m=n,\\ 0, & 若\ m\neq n.\end{cases}$$

因此，$\dim\mathscr{L}(\mathcal{U}_1,\cdots,\mathcal{U}_r)=\dim\mathcal{U}_1\cdots\dim\mathcal{U}_r$.

证 张量 $E_{j_1\cdots j_r}$ 记作 $\left[\delta_{i_1j_1}\cdots\delta_{i_rj_r}\right]_{n_1\times\cdots\times n_r}$. 考察给定基在 $\tau^{-1}(E_{j_1\cdots j_r})$ 下的像，由命题 11.2.5 可知，这组基就是满足上述条件的唯一一组多线性函数. □

11.2.2 对偶空间

如果 $\mathcal{U}_1,\cdots,\mathcal{U}_r$ 是内积空间且给定基是标准正交基，则命题 11.2.9 中的基可以显式写出.

不难得到 $\left\langle\boldsymbol{e}_{i_p}^{(p)},\boldsymbol{e}_{j_p}^{(p)}\right\rangle_{\mathcal{U}_p}=\delta_{i_pj_p}$，因此

$$\varphi_{j_1\cdots j_r}(\boldsymbol{e}_{i_1}^{(1)},\cdots,\boldsymbol{e}_{i_r}^{(r)})=\left\langle\boldsymbol{e}_{i_1}^{(1)},\boldsymbol{e}_{j_1}^{(1)}\right\rangle_{\mathcal{U}_1}\cdots\left\langle\boldsymbol{e}_{i_r}^{(r)},\boldsymbol{e}_{j_r}^{(r)}\right\rangle_{\mathcal{U}_r},$$

利用多线性的性质，则有

$$\varphi_{j_1\cdots j_r}(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r)=\left\langle\boldsymbol{a}_1,\boldsymbol{e}_{j_1}^{(1)}\right\rangle_{\mathcal{U}_1}\cdots\left\langle\boldsymbol{a}_r,\boldsymbol{e}_{j_r}^{(r)}\right\rangle_{\mathcal{U}_r},$$

即

$$\varphi_{j_1\cdots j_r}\colon(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r)\mapsto\left\langle\boldsymbol{a}_1,\boldsymbol{e}_{j_1}^{(1)}\right\rangle_{\mathcal{U}_1}\cdots\left\langle\boldsymbol{a}_r,\boldsymbol{e}_{j_r}^{(r)}\right\rangle_{\mathcal{U}_r}.$$

事实上，内积空间上的多线性函数空间，其基用一系列内积的乘积写出和用一系列线性函数的乘积写出是一回事. 因为：显然内积是线性函数，即 $\left\langle\boldsymbol{a}_p,\boldsymbol{e}_{j_p}^{(p)}\right\rangle_{\mathcal{U}_p}\in\mathscr{L}(\mathcal{U}_p)$；另一方面，内积空间上的线性函数也是内积，这由下面定理保证.

命题 11.2.10 (Riesz 表示定理) 给定内积空间 $\mathcal{U}$，对任意 $f\in\mathscr{L}(\mathcal{U})$，都存在 $\boldsymbol{f}\in\mathcal{U}$，使得 $f(\boldsymbol{u})=\langle\boldsymbol{f},\boldsymbol{u}\rangle$ 对任意 $\boldsymbol{u}\in\mathcal{U}$ 成立.

证 **留为作业.** □

练习 11.2.1 证明：给定 n 维内积空间 $\mathcal{U}$ 和对应内积 $\langle\cdot,\cdot\rangle_{\mathcal{U}}$. 则：

1. 如下映射是同构映射：$\sigma\colon\mathscr{L}(\mathcal{U})\to\mathcal{U},f\mapsto\boldsymbol{f}$，其中 $\boldsymbol{f}$ 是由 Riesz 表示定理确定的向量.

2. $\mathscr{L}(\mathcal{U})$ 关于内积 $\langle f,g\rangle_{\mathscr{L}(\mathcal{U})}=\langle\sigma(f),\sigma(g)\rangle_{\mathcal{U}}$ 构成内积空间.
3. 若 $\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n$ 是 $\mathcal{U}$ 的一组标准正交基，则 $\sigma^{-1}(\boldsymbol{e}_1),\cdots,\sigma^{-1}(\boldsymbol{e}_n)$ 是 $\mathscr{L}(\mathcal{U})$ 的一组标准正交基.

相应地，线性空间上的多线性函数空间，其基能用一系列线性函数写出吗？
为此先作如下定义.

定义 11.2.11 (对偶空间) 数域 $\mathbb{F}$ 上线性空间 $\mathcal{U}$ 上的线性函数组成的线性空间 $\mathscr{L}(\mathcal{U})=\mathrm{Hom}(\mathcal{U},\mathbb{F})$ 称为 $\mathcal{U}$ 的**对偶空间**，记为 $\mathcal{U}^*$.

任取 $\mathcal{U}$ 的一组基 $\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n$，则对应同构映射是

$$
\begin{aligned}
\sigma=\sigma_{\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n;1}:\ \ & \mathcal{U}^*\to\mathbb{F}^{1\times n},\\
& f\mapsto\sigma(f)=\begin{bmatrix}f(\boldsymbol{e}_1) & f(\boldsymbol{e}_2) & \cdots & f(\boldsymbol{e}_n)\end{bmatrix},
\end{aligned}
$$

其中 $\sigma(f)$ 是 f 在基 $\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n;1$ 下的矩阵.

而 $\mathbb{F}^{1\times n}$ 有一组标准基 $E_{11},E_{12},\cdots,E_{1n}$，因此 $\sigma^{-1}(E_{11}),\cdots,\sigma^{-1}(E_{1n})$ 就是 $\mathcal{U}^*$ 的一组基. 容易看到 $f_j:=\sigma^{-1}(E_{1j})$ 是满足 $f_j(\boldsymbol{e}_i)=\delta_{ij},i=1,\cdots,n$ 的线性函数. 根据命题 7.5.1 ，f_j 被这一条件唯一确定.

定义 11.2.12 (对偶基) 在 n 维线性空间 $\mathcal{U}$ 中取定一组基 $\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n$，则由 $f_j(\boldsymbol{e}_i)=\delta_{ij},i=1,\cdots,n$ 确定的线性函数 $f_1,\cdots,f_n$ 构成 $\mathcal{U}^*$ 的一组基，称为 $\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n$ 在 $\mathcal{U}^*$ 中的**对偶基**，记为 $\boldsymbol{e}_1^*,\cdots,\boldsymbol{e}_n^*$.

观察线性函数 $f(\boldsymbol{a})$ 的形式，我们可以自然地定义一个双线性函数.

定义 11.2.13 (取值函数) 双线性函数

$$
\begin{aligned}
\langle\cdot\mid\cdot\rangle:\ \ & \mathcal{U}\times\mathcal{U}^*\to\mathbb{F},\\
& (\boldsymbol{u},f)\mapsto\langle\boldsymbol{u}\mid f\rangle=f(\boldsymbol{u}),
\end{aligned}
$$

称为线性空间 $\mathcal{U}\times\mathcal{U}^*$ 上的**取值函数**.

取值函数在基 $\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n;\boldsymbol{e}_1^*,\cdots,\boldsymbol{e}_n^*$ 下的度量矩阵就是单位矩阵 I_n.

例 11.2.14 下面给出一些对偶空间和取值函数的例子. TBD ☺

由于对偶空间 $\mathcal{U}^*$ 是线性空间,因此仍然可以定义对偶空间的对偶空间 $\mathcal{U}^{**}=\mathscr{L}(\mathcal{U}^*)$. 于是可以定义双线性函数 $\langle\cdot\mid\cdot\rangle^*:\mathcal{U}^*\times\mathcal{U}^{**},\langle f\mid\mathfrak{u}\rangle^*=\mathfrak{u}(f)$.

如果能在 $\mathcal{U}$ 和 $\mathcal{U}^{**}$ 上建立同构 ν，那么就可以把 $\langle f\mid\boldsymbol{u}\rangle$ 定义为 $\langle f\mid\nu(\boldsymbol{u})\rangle^*$. 而这确实是可以做到的.

命题 11.2.15 对应关系

$$\begin{aligned}\nu = \nu_{\mathcal{U}}\colon\ & \mathcal{U} \to \mathcal{U}^{**},\\ & \boldsymbol{u} \mapsto \Big(\mathfrak{u}\colon f \mapsto f(\boldsymbol{u})\Big),\end{aligned}$$

是 $\mathcal{U}$ 到 $\mathcal{U}^{**}$ 的同构映射.

证 首先 $\mathfrak{u} \in \mathcal{U}^{**}$. 只需验证 $\mathfrak{u}$ 是线性函数，即 $\mathfrak{u}(kf + lg) = (kf + lg)(\boldsymbol{u}) = kf(\boldsymbol{u}) + lg(\boldsymbol{u}) = k\mathfrak{u}(f) + l\mathfrak{u}(g)$.

其次 ν 是映射. 只需验证对每个 $\boldsymbol{u}$，最多只有一个 $\mathfrak{u}$ 与之对应. 若对任意 $f \in \mathcal{U}^*$，$\mathfrak{u}(f) = \mathfrak{v}(f) = f(\boldsymbol{u})$，则 $\mathfrak{u}, \mathfrak{v}$ 作为函数必然相等.

再次 ν 是线性映射：$\nu(k\boldsymbol{u} + l\boldsymbol{v}) = \Big(f \mapsto f(k\boldsymbol{u} + l\boldsymbol{v})\Big) = \Big(f \mapsto kf(\boldsymbol{u}) + lf(\boldsymbol{v})\Big) = k\Big(f \mapsto f(\boldsymbol{u})\Big) + l\Big(f \mapsto f(\boldsymbol{v})\Big) = k\nu(\boldsymbol{u}) + l\nu(\boldsymbol{v})$.

最后说明 ν 是双射. 若 $\nu(\boldsymbol{u}) = \mathfrak{u}$ 是零函数，则对任意 $f \in \mathcal{U}^*$，$f(\boldsymbol{u}) = \mathfrak{u}(f) = 0$，即得 $\boldsymbol{u} = 0$，这就说明 ν 是单射. 注意到有限维线性空间总和它的对偶空间同构，因而维数相同，因此 $\mathcal{U}$ 和 $\mathcal{U}^{**}$ 维数也相同. 这就使得线性映射 ν 必须是双射. □

这一同构显然与线性空间 $\mathcal{U}$ 的基的选择无关，因此称为**自然同构**. 为简化书写，自然同构的两个线性空间就直接写作相等，向量在自然同构下的像也认为与原像相等：$\mathcal{U}^{**} = \mathcal{U}, \boldsymbol{u}^{**} = \boldsymbol{u}, \langle f \mid \boldsymbol{u}\rangle = \langle \mathfrak{u} \mid f\rangle^*$.

记 $\mathfrak{e}_j = \nu(\boldsymbol{e}_j)$，由于 $\mathfrak{e}_j(\boldsymbol{e}_i^*) = \boldsymbol{e}_i^*(\boldsymbol{e}_j) = \delta_{ij}$，因此 $\mathfrak{e}_1, \cdots, \mathfrak{e}_n$ 是 $\boldsymbol{e}_1^*, \cdots, \boldsymbol{e}_n^*$ 在 $\mathcal{U}^{**}$ 中的对偶基，即 $\mathfrak{e}_i = \boldsymbol{e}_i^{**}$. 这是上述自然同构的一个体现.

下面考虑对偶基及对偶向量坐标在基变换下的变化规律.

命题 11.2.16 *给定数域 $\mathbb{F}$ 上线性空间 $\mathcal{U}$ 和它的两组基 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n; \boldsymbol{t}_1, \boldsymbol{t}_2, \cdots, \boldsymbol{t}_n$. 记过渡矩阵为 T，即 $(\boldsymbol{t}_1, \boldsymbol{t}_2, \cdots, \boldsymbol{t}_n) = (\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n)T$. 则*

1. $(\boldsymbol{t}_1^*, \boldsymbol{t}_2^*, \cdots, \boldsymbol{t}_n^*) = (\boldsymbol{e}_1^*, \boldsymbol{e}_2^*, \cdots, \boldsymbol{e}_n^*)T^{-\mathrm{T}}$，*即对偶基的过渡矩阵为* $T^{-\mathrm{T}}$；
2. $(\boldsymbol{t}_1^{**}, \boldsymbol{t}_2^{**}, \cdots, \boldsymbol{t}_n^{**}) = (\boldsymbol{e}_1^{**}, \boldsymbol{e}_2^{**}, \cdots, \boldsymbol{e}_n^{**})T$，*即对偶基的对偶基的过渡矩阵为* T；
3. $\mathcal{U}$ *上的坐标变换公式是* $(\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n)\widehat{\boldsymbol{u}} = (\boldsymbol{t}_1, \boldsymbol{t}_2, \cdots, \boldsymbol{t}_n)T^{-1}\widehat{\boldsymbol{u}}$；
4. $\mathcal{U}^*$ *上的坐标变换公式是* $(\boldsymbol{e}_1^*, \boldsymbol{e}_2^*, \cdots, \boldsymbol{e}_n^*)\widehat{\boldsymbol{u}} = (\boldsymbol{t}_1^*, \boldsymbol{t}_2^*, \cdots, \boldsymbol{t}_n^*)T^{\mathrm{T}}\widehat{\boldsymbol{u}}$.
5. $\mathcal{U}^{**}$ *上的坐标变换公式是* $(\boldsymbol{e}_1^{**}, \boldsymbol{e}_2^{**}, \cdots, \boldsymbol{e}_n^{**})\widehat{\boldsymbol{u}} = (\boldsymbol{t}_1^{**}, \boldsymbol{t}_2^{**}, \cdots, \boldsymbol{t}_n^{**})T^{-1}\widehat{\boldsymbol{u}}$.

证 第 1 条：记 $(\boldsymbol{t}_1^*, \boldsymbol{t}_2^*, \cdots, \boldsymbol{t}_n^*) = (\boldsymbol{e}_1^*, \boldsymbol{e}_2^*, \cdots, \boldsymbol{e}_n^*)S$. 类似于命题 11.1.10 的证明，考虑取值函数 $(\boldsymbol{u}, f) \mapsto f(\boldsymbol{u})$ 在两组基下的度量矩阵满足 $I_n = T^{\mathrm{T}} I_n S$. 因此 $S = T^{-\mathrm{T}}$.

第 2 条：对 $\mathcal{U}^*$ 利用第 1 条，即得过渡矩阵为 $(T^{-\mathrm{T}})^{-\mathrm{T}} = T$.

其他都显然. □

命题 11.2.16 再次体现了自然同构与基的选择无关的性质.

取值函数自此就有了对称性：$\langle f \mid \boldsymbol{u}\rangle = \langle f \mid \mathfrak{u}\rangle = \mathfrak{u}(f) = f(\boldsymbol{u}) = \langle \boldsymbol{u} \mid f\rangle$.

由于取值函数定义域是 $\mathcal{U}^* \times \mathcal{U}$（或 $\mathcal{U} \times \mathcal{U}^*$），我们不能期望它有正定性. 但是关于双线性函数取值为零有如下条件：对任意 $\boldsymbol{u} \in \mathcal{U}$，$\langle \boldsymbol{u} \mid f\rangle = 0 \Leftrightarrow f = 0$；对任意 $f \in \mathcal{U}^*$，$\langle \boldsymbol{u} \mid f\rangle = 0 \Leftrightarrow \boldsymbol{u} = \boldsymbol{0}$.

这样取值函数就和内积空间的内积几乎没有差别了.

内积可以诱导伴随映射，类似地取值函数也可以诱导出对偶映射.

定义 11.2.17 (对偶映射) 给定线性空间 $\mathcal{U},\mathcal{V}$，取值函数分别为 $\langle\cdot\mid\cdot\rangle_{\mathcal{U}},\langle\cdot\mid\cdot\rangle_{\mathcal{V}}$. 如果线性映射 $\boldsymbol{F} \in \mathrm{Hom}(\mathcal{U},\mathcal{V}),\boldsymbol{G} \in \mathrm{Hom}(\mathcal{V}^*,\mathcal{U}^*)$ 满足对任意 $\boldsymbol{x} \in \mathcal{U},\boldsymbol{y} \in \mathcal{V}$，都有 $\langle \boldsymbol{F}(\boldsymbol{x}) \mid \boldsymbol{y}^*\rangle_{\mathcal{V}} = \langle \boldsymbol{G} \mid \boldsymbol{x}\rangle_{\mathcal{U}}$ 成立，则称 $\boldsymbol{G}$ 是 $\boldsymbol{F}$ 的**对偶映射**，记为 $\boldsymbol{G} = \boldsymbol{F}^*$.

特别地，如果 $\mathcal{U} = \mathcal{V}$，则 $\boldsymbol{F}^*$ 又称为 $\boldsymbol{F}$ 的**对偶变换**.

不难看出，欧氏空间之间的伴随映射，是对偶映射的一个特例.

例 11.2.18 下面给出一些对偶映射的例子. TBD ☺

让我们回到多线性函数. 利用对偶基，仿照内积空间上的讨论，我们就可得到多线性函数的一组基可以用一系列取值函数的乘积写出：

$$\varphi_{j_1\cdots j_r}\colon (\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r) \mapsto \left\langle \boldsymbol{a}_1 \mid \boldsymbol{e}_{j_1}^{(1)*}\right\rangle_{\mathcal{U}_1} \cdots \left\langle \boldsymbol{a}_r \mid \boldsymbol{e}_{j_r}^{(r)*}\right\rangle_{\mathcal{U}_r}. \tag{11.2.1}$$

更进一步地，任意多线性函数都能用一系列内积或者线性函数的乘积写出吗?

首先，多线性函数总可以用线性函数乘积的线性组合写出，这是因为 $\varphi_{j_1\cdots j_r}$ 是多线性函数空间的一组基. 其次，多线性函数不一定可以用一组线性函数乘积给出，因为容易举出反例，例如二阶行列式函数就不能由线性函数的乘积给出. 事实上，若 $\det(\boldsymbol{a}_1,\boldsymbol{a}_2) = (\boldsymbol{b}_1^{\mathrm{T}}\boldsymbol{a}_1)(\boldsymbol{b}_2^{\mathrm{T}}\boldsymbol{a}_2)$，则 $a_{11}a_{22}-a_{12}a_{21} = (a_{11}b_{11}+a_{21}b_{21})(a_{12}b_{12}+a_{22}b_{22})$ 对任意 $a_{11},a_{21},a_{12},a_{22}$ 成立，则：取 $a_{11} = a_{22} = 0$，则 $-a_{12}a_{21} = a_{21}a_{12}b_{21}b_{12}$，于是 $b_{21}b_{12} = -1$；取 $a_{11} = a_{12} = 0$，则 $0 = a_{21}a_{22}b_{21}b_{22}$，于是 $b_{21}b_{22} = 0$，从而 $b_{22} = 0$；取 $a_{21} = a_{22} = 0$，则 $0 = a_{11}a_{12}b_{11}b_{12}$，于是 $b_{11}b_{12} = 0$，从而 $b_{11} = 0$；于是 $a_{11}a_{22} - a_{12}a_{21} = -a_{21}a_{12}$，这不可能对任意 $\boldsymbol{a}_1,\boldsymbol{a}_2$ 成立，矛盾.

11.2.3 多线性映射

最后将线性函数、双线性函数的概念推广到更一般情形.

定义 11.2.19 (多线性映射) 给定 $\mathbb{F}$ 上的线性空间 $\mathcal{U}_1,\cdots,\mathcal{U}_r,\mathcal{V}$，若映射 $f\colon \mathcal{U}_1\times\cdots\times\mathcal{U}_r \to \mathcal{V},(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r) \mapsto f(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r)$ 对每一个变元都有线性性质，即固定任意 $r-1$ 个变元 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_{p-1},\boldsymbol{a}_{p+1},\cdots,\boldsymbol{a}_r$ 的任意取值后，$\boldsymbol{w} \mapsto f(\cdots,\overset{p}{\boldsymbol{w}},\cdots)$ 是 $\mathcal{U}_p$ 到 $\mathcal{V}$ 的

线性映射, 则称 f 是 $\mathcal{U}_1\times\cdots\times\mathcal{U}_r$ 到 $\mathcal{V}$ 的一个**多线性映射**或r **重线性映射**. $\mathcal{U}_1\times\cdots\times\mathcal{U}_r$ 到 $\mathcal{V}$ 的多线性映射的全体记为 $\mathscr{L}(\mathcal{U}_1,\cdots,\mathcal{U}_r;\mathcal{V})$.

特别地，$r=1$ 时就是线性映射，其全体是 $\mathscr{L}(\mathcal{U}_1;\mathcal{V})=\mathrm{Hom}(\mathcal{U},\mathcal{V})$；$r=2$ 时称为**双线性映射**，其全体是 $\mathscr{L}(\mathcal{U}_1,\mathcal{U}_2;\mathcal{V})$.

特别地，$\mathcal{U}\times\cdots\times\mathcal{U}$ 到 $\mathcal{V}$ 的多线性映射简称 $\mathcal{U}$ 到 $\mathcal{V}$ 的**多线性映射**.

可以构造出自然同构

$$\begin{aligned}\mathscr{L}(\mathcal{U};\mathcal{V})&\to\mathscr{L}(\mathcal{U},\mathcal{V}^*),\\ f&\mapsto\Big((\boldsymbol{u},\boldsymbol{v}^*)\mapsto\langle f(\boldsymbol{u})\mid\boldsymbol{v}^*\rangle_{\mathcal{V}}\Big).\end{aligned}$$

这个同构帮助我们把 $\mathcal{U}\to\mathcal{V}$ 的映射转化为了 $\mathcal{U}\times\mathcal{V}^*\to\mathbb{F}$ 的映射. 为简化书写，下面就直接写作 $\mathscr{L}(\mathcal{U};\mathcal{V})=\mathscr{L}(\mathcal{U},\mathcal{V}^*)$.

作业 利用这个自然同构写出 $\mathscr{L}(\mathcal{U};\mathcal{V})$ 的一组基.

类似地，$\mathscr{L}(\mathcal{U}_1,\cdots,\mathcal{U}_r;\mathcal{V})$ 与 $\mathscr{L}(\mathcal{U}_1,\cdots,\mathcal{U}_r,\mathcal{V}^*)$ 也是自然同构.

作业 写出一个自然同构并证明之.

因此，r 重线性映射可以通过转化为 $r+1$ 重线性函数来研究.

例 11.2.20 TBD ☺

11.3 张量

本节主要研究多线性函数的一般形式，即张量.

11.3.1 张量积和张量

下面把 (11.2.1) 式中取值函数乘积改写成线性函数乘积的形式: 令 $\varphi_{j_p}^{(p)}(\boldsymbol{a}_p)=\left\langle\boldsymbol{a}_p\mid\boldsymbol{e}_{j_p}^{(p)}\right\rangle_{\mathcal{U}_p}$, 则

$$\varphi_{j_1\cdots j_r}(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r)=\varphi_{j_1}^{(1)}(\boldsymbol{a}_1)\cdots\varphi_{j_r}^{(r)}(\boldsymbol{a}_r).\tag{11.3.1}$$

这个乘积究竟是那种意义上的乘积呢? 我们可以说多线性函数在一个向量组合上的取值等于一系列线性函数在组合中各个向量的取值的乘积，但却不能说多线性函数是线性函数的乘积，因此函数作为映射的特例，乘积特指映射的复合，与这里的含义不同.

为此作如下定义.

定义 11.3.1 (多线性函数的张量积) 给定两个多线性函数

$$
\begin{aligned}
f\colon\ & \mathcal{U}_1\times\cdots\times\mathcal{U}_r\to\mathbb{F},\\
& (\boldsymbol{x}_1,\cdots,\boldsymbol{x}_r)\mapsto f(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_r),\\
g\colon\ & \mathcal{V}_1\times\cdots\times\mathcal{V}_s\to\mathbb{F},\\
& (\boldsymbol{y}_1,\cdots,\boldsymbol{y}_s)\mapsto g(\boldsymbol{y}_1,\cdots,\boldsymbol{y}_s),
\end{aligned}
$$

定义多线性函数 f 和 g 的**张量积**为多线性函数

$$
\begin{aligned}
f\otimes g\colon\ & \mathcal{U}_1\times\cdots\times\mathcal{U}_r\times\mathcal{V}_1\times\cdots\times\mathcal{V}_s\to\mathbb{F},\\
& (\boldsymbol{x}_1,\cdots,\boldsymbol{x}_r,\boldsymbol{y}_1,\cdots,\boldsymbol{y}_s)\mapsto f(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_r)g(\boldsymbol{y}_1,\cdots,\boldsymbol{y}_s).
\end{aligned}
$$

注意, 当 $f\in\mathscr{L}(\mathcal{U}_1,\cdots,\mathcal{U}_r), g\in\mathscr{L}(\mathcal{V}_1,\cdots,\mathcal{V}_s)$ 时, $f\otimes g\in\mathscr{L}(\mathcal{U}_1,\cdots,\mathcal{U}_r,\mathcal{V}_1,\cdots,\mathcal{V}_s)$.

反之, $h\in\mathscr{L}(\mathcal{U}_1,\cdots,\mathcal{U}_r,\mathcal{V}_1,\cdots,\mathcal{V}_s)$, 不一定能找到 $f\in\mathscr{L}(\mathcal{U}_1,\cdots,\mathcal{U}_r), g\in\mathscr{L}(\mathcal{V}_1,\cdots,\mathcal{V}_s)$, 使得 $h=f\otimes g$ (为什么?).

不难验证如下运算律.

命题 11.3.2 多线性函数的张量积满足:

1. 结合律: $(f\otimes g)\otimes h=f\otimes(g\otimes h)$;
2. 双线性律: $(k_1f_1+k_2f_2)\otimes g=k_1(f_1\otimes g)+k_2(f_2\otimes g)$;
$$f\otimes(k_1g_1+k_2g_2)=k_1(f\otimes g_1)+k_2(f\otimes g_2).$$

证 **留为作业.** □

反复利用张量积, (11.3.1) 式就可以写作

$$\varphi_{j_1\cdots j_r}=\varphi^{(1)}_{j_1}\otimes\cdots\otimes\varphi^{(r)}_{j_r}. \tag{11.3.2}$$

这就说明, $\mathscr{L}(\mathcal{U}_1,\cdots,\mathcal{U}_r)=\operatorname{span}\left(\varphi^{(1)}_{j_1}\otimes\cdots\otimes\varphi^{(r)}_{j_r}\mid j_p=1,\cdots,n_p,p=1,\cdots,r\right)$ 注意到 $\varphi_{j_p},p=1,\cdots,n_p$ 是 $\mathcal{U}_p$ 的一组基, 就有

$$\mathscr{L}(\mathcal{U}_1,\cdots,\mathcal{U}_r,\mathcal{V}_1,\cdots,\mathcal{V}_s)=\operatorname{span}(f\otimes g\mid f\in\mathscr{L}(\mathcal{U}_1,\cdots,\mathcal{U}_r),g\in\mathscr{L}(\mathcal{V}_1,\cdots,\mathcal{V}_s)).$$

于是可以借此来定义线性空间的张量积. 另一方面, 利用对偶空间的对偶空间可以直接看作原空间, 只要把对偶空间上的多线性函数看作原空间的向量, 就可以利用定义 11.3.1 来定义向量的张量积.

定义 11.3.3 (向量和线性空间的张量积) 两个向量 $\boldsymbol{a}\in\mathcal{U}$ 和 $\boldsymbol{b}\in\mathcal{V}$ 的张量积定义为

$$\boldsymbol{a}\otimes\boldsymbol{b}=\nu^{-1}_{\mathscr{L}(\mathcal{U}^*,\mathcal{V}^*)}\left(\nu_{\mathscr{L}(\mathcal{U}^*)}(\boldsymbol{a})\otimes\nu_{\mathscr{L}(\mathcal{V}^*)}(\boldsymbol{b})\right)=\nu^{-1}_{\mathscr{L}(\mathcal{U}^*,\mathcal{V}^*)}\left(\nu_{\mathcal{U}}(\boldsymbol{a})\otimes\nu_{\mathcal{V}}(\boldsymbol{b})\right),$$

其中 $\nu_\cdot$ 的定义见命题 *11.2.15* .

两个线性空间 $\mathcal{U},\mathcal{V}$ 的**张量积**定义为线性空间

$$\mathcal{U}\otimes\mathcal{V} := \mathrm{span}(\boldsymbol{a}\otimes\boldsymbol{b}\mid\boldsymbol{a}\in\mathcal{U},\boldsymbol{b}\in\mathcal{V}).$$

张量积空间 $\mathcal{U}\otimes\mathcal{V}$ 中的元素称为**张量**，其中能直接表示成两个向量的张量积 $\boldsymbol{a}\otimes\boldsymbol{b}$ 的元素，称为**可分张量**，否则称为**不可分张量**.

各种对象的张量积 $a_1\otimes\cdots\otimes a_r$ 中的每一个 $a_i, i=1,\cdots,r$，都称为该张量积的**张量因子**.

各种同一对象的张量积记作 $a^{\otimes k} = \underbrace{a\otimes a\otimes\cdots\otimes a}_{k\text{ 个}}$.

张量积空间的张量因子的个数，称为该空间中张量的**阶**.

特别地，如果张量积空间的张量因子是同一个线性空间 $\mathcal{U}$ 或其对偶空间 $\mathcal{U}^*$，则称该空间中的张量为 $\mathcal{U}$（或 $\mathcal{U}^*$）上的张量.

命题 11.3.4 *1.* 向量和线性空间的张量积满足结合律和双线性律.

2. 若 $\boldsymbol{u}_1,\cdots,\boldsymbol{u}_n;\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m$ 分别是 $\mathcal{U},\mathcal{V}$ 的一组基，则 $\boldsymbol{u}_i\otimes\boldsymbol{v}_j, i=1,\cdots,n, j=1,\cdots,m$ 是 $\mathcal{U}\otimes\mathcal{V}$ 的一组基，于是 $\dim\mathcal{U}\otimes\mathcal{V}=\dim\mathcal{U}\dim\mathcal{V}$.

3. 设 $\boldsymbol{u}\in\mathcal{U},\boldsymbol{v}\in\mathcal{V},\boldsymbol{u}\otimes\boldsymbol{v}=\boldsymbol{0}$，则 $\boldsymbol{u}=\boldsymbol{0}$ 或 $\boldsymbol{v}=\boldsymbol{0}$.

4. $\mathscr{L}(\mathcal{U},\mathcal{V})=\mathscr{L}(\mathcal{U}\otimes\mathcal{V})$，$(\mathcal{U}\otimes\mathcal{V})^*=\mathcal{U}^*\otimes\mathcal{V}^*$.

证 第 1 条直接验证即可.

第 2 条：显然 $\boldsymbol{u}_i\otimes\boldsymbol{v}_j$ 可以表出 $\mathcal{U}\otimes\mathcal{V}$ 中的任意向量，下面证其线性无关. 若 $\sum_{i,j} a_{ij}\boldsymbol{u}_i\otimes\boldsymbol{v}_j=\boldsymbol{0}$，记 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_m$ 的对偶基为 $\boldsymbol{v}_1^*,\cdots,\boldsymbol{v}_m^*$，$f\in\mathcal{U}^*$ 任意，两边作用在 $(f,\boldsymbol{v}_{j'}^*)$ 上，有 $\sum_i a_{ij'}\boldsymbol{u}_i(f)=0\Leftrightarrow\sum_i a_{ij'}\boldsymbol{u}_i=\boldsymbol{0}, j'=1,\cdots,m$. 再由 $\boldsymbol{u}_1,\cdots,\boldsymbol{u}_n$ 线性无关，即得 $a_{ij'}=0, i=1,\cdots,n, j'=1,\cdots,m$.

第 3 条和第 4 条：$\boxed{\textbf{留为作业.}}$ □

利用命题 11.3.4，我们可以把 $\mathcal{U}\times\mathcal{V}$ 中的向量组和向量的张量积等同起来，从而利用张量积的运算和运算律为向量组赋予运算，使之成为 $\mathcal{U}\otimes\mathcal{V}$：

$$(k_1\boldsymbol{u}_1+k_2\boldsymbol{u}_2,\boldsymbol{v}) := k_1(\boldsymbol{u}_1,\boldsymbol{v})+k_2(\boldsymbol{u}_2,\boldsymbol{v}),$$
$$(\boldsymbol{u},k_1\boldsymbol{v}_1+k_2\boldsymbol{v}_2) := k_1(\boldsymbol{u},\boldsymbol{v}_1)+k_2(\boldsymbol{u},\boldsymbol{v}_2).$$

现在 (11.3.1) 式又可以写成

$$\varphi_{j_1}^{(1)}(\boldsymbol{a}_1)\cdots\varphi_{j_r}^{(r)}(\boldsymbol{a}_r)=\varphi_{j_1}^{(1)}\otimes\cdots\otimes\varphi_{j_r}^{(r)}(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r)$$

$$
\begin{aligned}
&= \boldsymbol{a}_1 \otimes \cdots \otimes \boldsymbol{a}_r(\varphi_{j_1}^{(1)}, \cdots, \varphi_{j_r}^{(r)}) \\
&= \varphi_{j_1}^{(1)} \otimes \cdots \otimes \varphi_{j_r}^{(r)}(\boldsymbol{a}_1 \otimes \cdots \otimes \boldsymbol{a}_r) \\
&= \boldsymbol{a}_1 \otimes \cdots \otimes \boldsymbol{a}_r(\varphi_{j_1}^{(1)} \otimes \cdots \otimes \varphi_{j_r}^{(r)}) \\
&= \left\langle \varphi_{j_1}^{(1)} \otimes \cdots \otimes \varphi_{j_r}^{(r)} \mid \boldsymbol{a}_1 \otimes \cdots \otimes \boldsymbol{a}_r \right\rangle \\
&= \left\langle \boldsymbol{a}_1 \otimes \cdots \otimes \boldsymbol{a}_r \mid \varphi_{j_1}^{(1)} \otimes \cdots \otimes \varphi_{j_r}^{(r)} \right\rangle.
\end{aligned}
$$

由线性空间和向量的张量积可以诱导线性映射的张量积.

定义 11.3.5 (线性映射的张量积) 两个线性映射 $\boldsymbol{A} \in \mathrm{Hom}(\mathcal{U}_1, \mathcal{U}_2)$ 和 $\boldsymbol{B} \in \mathrm{Hom}(\mathcal{V}_1, \mathcal{V}_2)$ 的张量积定义为

$$
\begin{aligned}
\boldsymbol{A} \otimes \boldsymbol{B}\colon\ & \mathcal{U}_1 \otimes \mathcal{V}_1 \to \mathcal{U}_2 \otimes \mathcal{V}_2, \\
& \boldsymbol{x} \otimes \boldsymbol{y} \mapsto \boldsymbol{A}(\boldsymbol{x}) \otimes \boldsymbol{B}(\boldsymbol{y}), \\
\text{或} \quad & (\boldsymbol{x}, \boldsymbol{y}) \mapsto (\boldsymbol{A}(\boldsymbol{x}), \boldsymbol{B}(\boldsymbol{y}))\,.
\end{aligned}
$$

命题 11.3.6 *1.* 线性映射的张量积满足结合律和双线性律.

2. $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{AC}) \otimes (\boldsymbol{BD})$;

3. 如果 $\boldsymbol{A}, \boldsymbol{B}$ 可逆, 则 $\boldsymbol{A} \otimes \boldsymbol{B}$ 可逆, 且 $(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$;

例 11.3.7 下面讨论取值函数作为张量的例子.

对线性空间 $\mathcal{U}$, 考虑 $\mathcal{U} \otimes \mathcal{U}^*$ 中的可分张量 $\boldsymbol{u} \otimes f\colon \mathcal{U} \times \mathcal{U}^* \to \mathbb{F}, (g, \boldsymbol{v}) \mapsto \mathfrak{u}(g) f(\boldsymbol{v}) = g(\boldsymbol{u}) f(\boldsymbol{v})$. 注意它与取值函数的不同: 取值函数为 $\langle \cdot \mid \cdot \rangle\colon \mathcal{U} \times \mathcal{U}^* \to \mathbb{F}, (g, \boldsymbol{v}) \mapsto g(\boldsymbol{v})$.

取 $\mathcal{U}$ 的一组基 $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n$ 及其在 $\mathcal{U}^*$ 中的对偶基 $\boldsymbol{e}_1^*, \cdots, \boldsymbol{e}_n^*$, 则

$$
\boldsymbol{u} \otimes f(g, \boldsymbol{v}) = \sum_{i,j,i',j'} u_i f_j g_{j'} v_{i'} \boldsymbol{e}_{j'}^*(\boldsymbol{e}_i) \boldsymbol{e}_j^*(\boldsymbol{e}_{i'}) = \sum_{i,j,i',j'} u_i f_j g_{i'} v_{j'} \delta_{ij'} \delta_{i'j},
$$

而 $\langle g \mid \boldsymbol{v} \rangle = \sum_{i',j'} g_{i'} v_{j'} \boldsymbol{e}_{i'}^*(\boldsymbol{e}_{j'}) = \sum_{i',j'} g_{i'} v_{j'} \delta_{i'j'}$.

可以看到, 虽然内积或取值函数 $\langle \boldsymbol{u} \mid f \rangle \in \mathcal{U} \otimes \mathcal{U}^*$ 也是双线性函数, 但它并不是张量积 $\boldsymbol{u} \otimes f$, 而是一个尚未写出来的张量.

下面来写出取值函数对应的张量. 已经知道它由在 $(\boldsymbol{e}_{j'}^*, \boldsymbol{e}_{i'}), i', j' = 1, \cdots, n$ 上的取值 $\left\langle \boldsymbol{e}_{j'}^* \mid \boldsymbol{e}_{i'} \right\rangle = \boldsymbol{e}_{j'}^*(\boldsymbol{e}_{i'}) = \delta_{i'j'}$ 唯一确定. 设取值函数为 $\sum_i \sum_j \varepsilon_{ij} \boldsymbol{e}_i \otimes \boldsymbol{e}_j^*$, 则它在 $(\boldsymbol{e}_{j'}^*, \boldsymbol{e}_{i'}), i', j' = 1, \cdots, n$ 上的取值为

$$
\begin{aligned}
\sum_i \sum_j \varepsilon_{ij} \boldsymbol{e}_{j'}^*(\boldsymbol{e}_i) \boldsymbol{e}_j^*(\boldsymbol{e}_{i'}) &= \sum_i \sum_j \varepsilon_{ij} \delta_{ij'} \delta_{i'j} \\
&= \sum_{i \neq j'} \sum_j \varepsilon_{ij} \delta_{ij'} \delta_{i'j} + \sum_j \varepsilon_{j'j} \delta_{j'j'} \delta_{i'j}
\end{aligned}
$$

$$
\begin{aligned}
&= 0 + \sum_{j \neq i'} \varepsilon_{j'j}\delta_{i'j} + \varepsilon_{j'i'}\delta_{i'i'} \\
&= \varepsilon_{j'i'}.
\end{aligned}
$$

因此 $\varepsilon_{ij} = \delta_{ji}$，即取值函数在张量基下的表示为 $\sum_i \sum_j \delta_{ij} \boldsymbol{e}_i \otimes \boldsymbol{e}_j^* = \sum_i \boldsymbol{e}_i \otimes \boldsymbol{e}_i^*$.

当 $n > 1$ 时，取值函数不是可分张量.

取值函数其实是矩阵迹的推广. ☺

类似于例 11.3.7，我们把张量中某些张量因子作用取值函数定义为一个新的运算.

定义 11.3.8 (缩并) 对可分张量 $A = \cdots \otimes \boldsymbol{u}_i \otimes \cdots \otimes \boldsymbol{u}_j \otimes \cdots \in \cdots \otimes \mathcal{U}_i \otimes \cdots \otimes \mathcal{U}_j \otimes \cdots$，若 $\mathcal{U}_j = \mathcal{U}_i^*$，则该张量在 i, j 处的**缩并**为

$$
\begin{aligned}
\mathscr{C}_{ij}(A)\colon\ & \cdots \otimes \widehat{\mathcal{U}_i} \otimes \cdots \otimes \widehat{\mathcal{U}_j} \otimes \cdots \to \mathbb{F}, \\
& (\cdots, \widehat{\boldsymbol{v}_i}, \cdots, \widehat{\boldsymbol{v}_j}, \cdots) \mapsto \cdots \otimes \widehat{\boldsymbol{u}_i} \otimes \cdots \otimes \widehat{\boldsymbol{u}_j} \otimes \cdots (\cdots, \widehat{\boldsymbol{v}_i}, \cdots, \widehat{\boldsymbol{v}_j}, \cdots) \langle \boldsymbol{u}_i \mid \boldsymbol{u}_j \rangle,
\end{aligned}
$$

其中 $\widehat{\ }$ 表示不含该项.

不可分张量 $\sum_{k_1,\cdots,k_r} a_{k_1\cdots k_r} (\cdots \otimes \boldsymbol{u}_{k_i} \otimes \cdots \otimes \boldsymbol{u}_{k_j} \otimes \cdots)$ 的**缩并**利用多线性的性质定义：

$$
\begin{aligned}
&\mathscr{C}_{ij}\left(\sum_{k_1,\cdots,k_r} a_{k_1\cdots k_r}(\cdots \otimes \boldsymbol{u}_{k_i} \otimes \cdots \otimes \boldsymbol{u}_{k_j} \otimes \cdots)\right) \\
&:= \sum_{\cdots,\widehat{k_i},\cdots,\widehat{k_j},\cdots} \left(\sum_{k_i,k_j} a_{\cdots k_i \cdots k_j \cdots} \left\langle \boldsymbol{u}_{k_i} \mid \boldsymbol{u}_{k_j} \right\rangle\right) (\cdots \otimes \widehat{\boldsymbol{u}_{k_i}} \otimes \cdots \otimes \widehat{\boldsymbol{u}_{k_j}} \otimes \cdots).
\end{aligned}
$$

11.3.2 张量及其运算的向量化表示

首先考虑定义 11.3.3 中的张量的数组表示.

给定线性空间 $\mathcal{U}_1, \cdots, \mathcal{U}_r$ 及各自一组基 $\boldsymbol{e}_1^{(1)}, \boldsymbol{e}_2^{(1)}, \cdots, \boldsymbol{e}_{n_1}^{(1)}; \cdots; \boldsymbol{e}_1^{(r)}, \boldsymbol{e}_2^{(r)}, \cdots, \boldsymbol{e}_{n_r}^{(r)}$，考虑张量积空间 $\mathcal{U}_1 \otimes \cdots \otimes \mathcal{U}_r$.

张量积空间中的张量总是张量因子空间的基向量的张量积的线性组合，因此也是多线性函数的线性组合. 而多线性函数可以用高维数组表示，因此张量也就可以用高维数组表示. 这也是前面直接把高维数组称为张量的原因.

命题 11.2.8 可以重写为如下结论.

命题 11.3.9 给定线性空间 $\mathcal{U}_1, \cdots, \mathcal{U}_r$ 及各自一组基 $\boldsymbol{e}_1^{(1)}, \boldsymbol{e}_2^{(1)}, \cdots, \boldsymbol{e}_{n_1}^{(1)}; \cdots; \boldsymbol{e}_1^{(r)}, \boldsymbol{e}_2^{(r)}, \cdots, \boldsymbol{e}_{n_r}^{(r)}$，则映射

$$
\begin{aligned}
\tau = \tau_{\boldsymbol{e}_1^{(1)}, \boldsymbol{e}_2^{(1)}, \cdots, \boldsymbol{e}_{n_1}^{(1)}; \cdots; \boldsymbol{e}_1^{(r)}, \boldsymbol{e}_2^{(r)}, \cdots, \boldsymbol{e}_{n_r}^{(r)}}\colon\quad & \mathcal{U}_1 \otimes \cdots \otimes \mathcal{U}_r \to \mathbb{F}^{n_1 \times \cdots \times n_r}, \\
& \textstyle\sum_{i_1 \cdots i_r} f_{i_1 \cdots i_r} \boldsymbol{e}_{i_1}^{(1)} \otimes \cdots \otimes \boldsymbol{e}_{i_r}^{(r)} \mapsto \left[f_{i_1 \cdots i_r}\right]_{n_1 \times \cdots \times n_r},
\end{aligned}
$$

是同构映射.

因此，指定张量因子空间的基后，研究张量积空间中的张量就只需要研究对应的高维数组张量即可.

张量或者多线性函数的线性运算，显然对应于数组张量的线性运算. 那么张量的张量积对应于数组张量的什么运算?

注意到张量和对应多线性函数的关系，事实上张量只是刻画了多线性函数与基之间的线性关系，因此如何码放代表线性关系的数并不十分重要. 因此我们不如直接采用 1 维数组，即向量来安放这些数.

定义 11.3.10 (向量化) 作用于张量的**向量化算子**定义为

$$\mathrm{vec}\colon \left[x_{i_1\cdots i_r}\right]_{n_1\cdots n_r} \mapsto n_1\cdots n_r \text{ 个数 } x_{i_1\cdots i_r} \text{ 依自然顺序排列得到的向量},$$

其中**自然顺序**是按如下比较办法把指标按前后排列的顺序[①]：两个指标序列 $i_1\cdots i_r$，$j_1\cdots j_r$，若 $i_1-j_1,\cdots,i_r-j_r$ 这 r 个数自右而左第一个非零数为负，则 $i_1\cdots i_r$ 在前 $j_1\cdots j_r$ 在后.

注意向量化算子 vec 其实是一系列定义在不同的线性空间上的映射，但由于其功能的一致性，这里采用了相同的记号而不加区分，读者自能从自变量中得到定义域. 另外，当张量积空间的各个张量因子空间的维数确定后，vec 的逆也是确定的.

可以看到，张量的向量化是定义 10.3.14 中矩阵的向量化的推广.

例 11.3.11 回顾例 10.3.15 . $\mathbb{F}^{m\times n}$ 的这组标准基向量化后恰好得到 $\mathrm{vec}(\boldsymbol{e}_i\boldsymbol{i}_j^{\mathrm{T}}) = \boldsymbol{i}_j\otimes\boldsymbol{e}_i$. 另一方面，$\mathbb{F}^{m\times n}$ 作为线性映射空间的表示，可以分解为张量积

$$\mathbb{F}^{m\times n} = \mathscr{L}(\mathbb{F}^n;\mathbb{F}^m) = \mathscr{L}(\mathbb{F}^n,(\mathbb{F}^m)^*) = \mathscr{L}(\mathbb{F}^n\otimes(\mathbb{F}^m)^*) = (\mathbb{F}^n)^*\otimes\mathbb{F}^m \text{同构于} \mathbb{F}^m\otimes\mathbb{F}^n,$$

由张量因子空间上的标准基 $\boldsymbol{e}_i,\boldsymbol{i}_j$ 根据命题 11.3.4 生成的张量积空间的基，恰好是 $\boldsymbol{i}_j\otimes\boldsymbol{e}_i$. 也就是说，向量化和 Kronecker 积其实就是张量和张量积的具体表示形式（之一).

而每个矩阵 X 都可以被这组基表示：

$$\mathrm{vec}(X) = \sum_{i,j} x_{ij}\,\mathrm{vec}(\boldsymbol{e}_i\boldsymbol{i}_j^{\mathrm{T}}) = \sum_{i,j} x_{ij}\boldsymbol{i}_j\otimes\boldsymbol{e}_i.$$

注意，在实际运算中并没有真正地计算张量乘法和求和，因为 $\boldsymbol{e}_i\otimes\boldsymbol{i}_j$ 是某一特定元素为 1 其他元素为 0 的向量，这一转换在实际计算中的意义就是把矩阵（或者说 2 阶张量）的元素按照自然顺序排列.

① 通俗地说，自然顺序就是每一指标从小到大变化，越靠左的指标变化越快的顺序.

不难类似地计算出 $A\otimes B=\sum_{i,j,i',j'}a_{ij}b_{i'j'}(\boldsymbol{e}_i\boldsymbol{i}_j^{\mathrm{T}})\otimes(\tilde{\boldsymbol{e}}_{i'}\tilde{\boldsymbol{i}}_{j'}^{\mathrm{T}})$，其中

$$(\boldsymbol{e}_i\boldsymbol{i}_j^{\mathrm{T}})\otimes(\tilde{\boldsymbol{e}}_{i'}\tilde{\boldsymbol{i}}_{j'}^{\mathrm{T}})=(\boldsymbol{e}_i\otimes\tilde{\boldsymbol{e}}_{i'})(\boldsymbol{i}_j\otimes\tilde{\boldsymbol{i}}_{j'})^{\mathrm{T}}=\boldsymbol{e}_i\otimes\tilde{\boldsymbol{e}}_{i'}\otimes\boldsymbol{i}_j^{\mathrm{T}}\otimes\tilde{\boldsymbol{i}}_{j'}^{\mathrm{T}}.$$

由于 $\boldsymbol{e}_i\otimes\tilde{\boldsymbol{e}}_{i'}\otimes\boldsymbol{i}_j^{\mathrm{T}}\otimes\tilde{\boldsymbol{i}}_{j'}^{\mathrm{T}}$ 也是一个矩阵，因此也可以向量化：

$$\begin{aligned}\operatorname{vec}\left(\boldsymbol{e}_i\otimes\tilde{\boldsymbol{e}}_{i'}\otimes\boldsymbol{i}_j^{\mathrm{T}}\otimes\tilde{\boldsymbol{i}}_{j'}^{\mathrm{T}}\right)&=\operatorname{vec}\left((\boldsymbol{e}_i\otimes\tilde{\boldsymbol{e}}_{i'})\otimes(\boldsymbol{i}_j\otimes\tilde{\boldsymbol{i}}_{j'})^{\mathrm{T}}\right)\\&=(\boldsymbol{i}_j\otimes\tilde{\boldsymbol{i}}_{j'})\otimes(\boldsymbol{e}_i\otimes\tilde{\boldsymbol{e}}_{i'})\operatorname{vec}(1)=\boldsymbol{i}_j\otimes\tilde{\boldsymbol{i}}_{j'}\otimes\boldsymbol{e}_i\otimes\tilde{\boldsymbol{e}}_{i'}.\end{aligned}$$

于是

$$\operatorname{vec}(A\otimes B)=\sum_{i,j,i',j'}a_{ij}b_{i'j'}\boldsymbol{i}_j\otimes\tilde{\boldsymbol{i}}_{j'}\otimes\boldsymbol{e}_i\otimes\tilde{\boldsymbol{e}}_{i'}.$$

上式也没有真正求和，只是重新码放了张量的元素.

上面计算也可用张量积来解释. 考虑到其实际意义，$A\otimes B$ 是 $f\in\mathrm{Hom}(\mathbb{F}^{m\times n},\mathbb{F}^{m'\times n'})$ 的矩阵表示. 由于

$$\begin{aligned}\mathrm{Hom}(\mathbb{F}^{m\times n},\mathbb{F}^{m'\times n'})&=\mathscr{L}(\mathbb{F}^{m\times n};\mathbb{F}^{m'\times n'})\\&=\mathscr{L}((\mathbb{F}^{n})^*\otimes\mathbb{F}^{m};(\mathbb{F}^{n'})^*\otimes\mathbb{F}^{m'})\\&=\mathscr{L}((\mathbb{F}^{n})^*\otimes\mathbb{F}^{m},\mathbb{F}^{n'}\otimes(\mathbb{F}^{m'})^*)\\&=\mathscr{L}((\mathbb{F}^{n})^*\otimes\mathbb{F}^{m}\otimes\mathbb{F}^{n'}\otimes(\mathbb{F}^{m'})^*)\\&=\mathbb{F}^{n}\otimes(\mathbb{F}^{m})^*\otimes(\mathbb{F}^{n'})^*\otimes\mathbb{F}^{m'}\\&\text{同构于 }\mathbb{F}^{m}\otimes\mathbb{F}^{m'}\otimes\mathbb{F}^{n}\otimes\mathbb{F}^{n'},\end{aligned}$$

而 $\boldsymbol{i}_j\otimes\tilde{\boldsymbol{i}}_{j'}\otimes\boldsymbol{e}_i\otimes\tilde{\boldsymbol{e}}_{i'}$ 就是它的一组基，$A\otimes B$ 的向量化就是对应张量的表示. ☺

上面想法显然不难扩展到任意阶张量. 也就是说，如果张量积空间 $\mathbb{F}^{n_1}\otimes\cdots\otimes\mathbb{F}^{n_r}$ 是张量空间 $\mathbb{F}^{n_1\times\cdots\times n_r}$，且 $\mathbb{F}^{n_p}$ 的一组基是 $\boldsymbol{e}_1^{(p)},\cdots,\boldsymbol{e}_{n_p}^{(p)},p=1,\cdots,r$，那么其基向量化后就是由 Kronecker 积计算得到的 $\boldsymbol{e}_{i_1}^{(1)}\otimes\cdots\otimes\boldsymbol{e}_{i_r}^{(r)},i_p=1,\cdots,n_p,p=1,\cdots,r$. 而线性映射的张量积的表示矩阵也可由 Kronecker 积表示.

向量化显然与线性运算是可以交换次序计算的. 那么其他矩阵运算，例如转置和乘法，如何用向量化合理地表示?

由于 $\operatorname{vec}(\boldsymbol{e}_i\boldsymbol{i}_j^{\mathrm{T}})=\boldsymbol{i}_j\otimes\boldsymbol{e}_i$，而 $\operatorname{vec}\left((\boldsymbol{e}_i\boldsymbol{i}_j^{\mathrm{T}})^{\mathrm{T}}\right)=\operatorname{vec}(\boldsymbol{i}_j\boldsymbol{e}_i^{\mathrm{T}})=\boldsymbol{e}_i\otimes\boldsymbol{i}_j$，可见转置对应于交换张量积中张量因子的次序.

两矩阵的乘积是

$$AB=\sum_{i,j,i',j'}a_{ij}b_{i'j'}(\boldsymbol{e}_i\boldsymbol{i}_j^{\mathrm{T}})(\tilde{\boldsymbol{e}}_{i'}\tilde{\boldsymbol{i}}_{j'}^{\mathrm{T}})=\sum_{i,j}\left(\sum_{j,i'}a_{ij}b_{i'j'}(\boldsymbol{i}_j^{\mathrm{T}}\tilde{\boldsymbol{e}}_{i'})\right)(\boldsymbol{e}_i\tilde{\boldsymbol{i}}_{j'}^{\mathrm{T}}),$$

两边向量化，有

$$\mathrm{vec}(AB) = \sum_{i,j} \left(\sum_{j,i'} a_{ij} b_{i'j'} (\boldsymbol{i}_j^{\mathrm{T}} \tilde{\boldsymbol{e}}_{i'}) \right) \tilde{\boldsymbol{i}}_{j'} \otimes \boldsymbol{e}_i = \mathrm{vec}\left(\mathscr{C}_{14}(A \otimes B) \right).$$

因此矩阵乘积对应于张量运算就是对应位置上的缩并．由于向量也是特殊的矩阵，因此矩阵与向量的乘积也是缩并.

11.3.3 协变张量和逆变张量

下面来讨论张量表示在基变换下的变化规律，为简单起见，我们只考虑内积空间 $\mathcal{U}$ 及其对偶空间 $\mathcal{U}^*$ 上的张量.

首先来重点介绍张量理论的传统记号法.

在上面计算中，我们已经看到对某一指标求和在涉及张量的计算总是频繁出现．为进一步简化写法，引入张量理论中常用的 **Einstein 记号**来省略求和号[①]：

$$a^i b_i := \sum_{i=1}^n a_i b_i = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n,$$

其记号含义在于：如果式子中乘积的因子中一个指标出现两次，且一个是上指标一个是下指标，则表示该乘积事实上需要对此指标所有的可能值求和.

给定 $\mathcal{U}$ 的一组基 $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n$，为利用 Einstein 记号，对偶基写成 $\boldsymbol{e}^1 = \boldsymbol{e}_1^*, \cdots, \boldsymbol{e}^n = \boldsymbol{e}_n^*$. 那么张量积空间中的基就可以写作 $\cdots \otimes \boldsymbol{e}_i \otimes \cdots \otimes \boldsymbol{e}^j \otimes \cdots$，其中上下指标与张量积空间中各张量因子的顺序有关．而任意张量

$$A = \sum_{i_1, \cdots, i_r} a_{\cdots\ \ \cdots i_q \cdots}^{\cdots i_p \cdots\ \ \cdots} (\cdots \otimes \boldsymbol{e}_{i_p} \otimes \cdots \otimes \boldsymbol{e}^{i_q} \otimes \cdots) = a_{\cdots\ \ \cdots i_q \cdots}^{\cdots i_p \cdots\ \ \cdots} (\cdots \otimes \boldsymbol{e}_{i_p} \otimes \cdots \otimes \boldsymbol{e}^{i_q} \otimes \cdots),$$

其中高维数组 $\left[a_{\cdots}^{\cdots} \right]$ 中对应指标的上下位置和基指标的上下位置相反，这是为了利用 Einstein 记号.

下面具体写出简单情形，便于读者熟悉.

$\mathcal{U}$ 中向量 $\boldsymbol{u}$ 在给定基下的坐标表示为 $\boldsymbol{u} = u^1 \boldsymbol{e}_1 + \cdots + u^n \boldsymbol{e}_n = u^i \boldsymbol{e}_i$；$\mathcal{U}^*$ 中向量 f 在对偶基下的坐标表示为 $f = f_1 \boldsymbol{e}^1 + \cdots + f_n \boldsymbol{e}^n = f_j \boldsymbol{e}^j$.

利用取值函数和对偶基，可以计算出 $\boldsymbol{u}$ 和 f 的坐标：$u^i = \langle \boldsymbol{u} \mid \boldsymbol{e}^i \rangle, f_j = \langle \boldsymbol{e}_j \mid f \rangle$. 注意，取坐标 $\boldsymbol{u} \mapsto \langle \boldsymbol{u} \mid \boldsymbol{e}^i \rangle = u^i$ 是 $\mathcal{U}$ 上线性函数，即 $\mathcal{U}^*$ 中向量，因此也和对偶基一样，使用上指标．类似地，取坐标 $f \mapsto f_i$ 是 $\mathcal{U}$ 中的向量，使用下指标．这也是 u^i, f_j 的指标位置安排的一个原因.

① 注意，在此写法中，上指标不是数的幂次，因此 a 的平方只能写成 aa，而不能写成 a^2，请读者注意文献的约定.

若 $\mathcal{U}$ 有另一组基 $\boldsymbol{t}_1,\cdots,\boldsymbol{t}_n$, 设过渡矩阵为 T, 即 $(\boldsymbol{t}_1,\cdots,\boldsymbol{t}_n)=(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n)T$. 把过渡矩阵看作 $\mathrm{Hom}(\mathcal{U},\mathcal{U})=\mathcal{U}^*\otimes\mathcal{U}$ 中的元素, 因此 T 的第一个指标是上指标, 第二个指标是下指标. 用 Einstein 记号写出, 即为 $\boldsymbol{t}_i=t^1{}_i\boldsymbol{e}_1+\cdots+t^n{}_i\boldsymbol{e}_n=t^k{}_i\boldsymbol{e}_k$. 类似地, $t^i{}_j=\langle\boldsymbol{t}_j\mid\boldsymbol{e}^i\rangle$. 从这个观点，矩阵的指标可以按如下规则安排：$T^{-1}=\left[(T^{-1})^i{}_j\right], T^{\mathrm{T}}=\left[(T^{\mathrm{T}})_i{}^j\right]$.

类似地，$I=\left[\delta^i{}_j\right]=\left[\delta_j{}^i\right]$，因此对偶基就满足 $\langle\boldsymbol{e}_i\mid\boldsymbol{e}^j\rangle=\delta_i{}^j$. 而 T^{-1} 的指标安排也和 $t^i{}_k(T^{-1})^k{}_j=\delta^i{}_j$ 对应.

而若向量 $\boldsymbol{u}$ 在这组基下的坐标表示为 $\boldsymbol{u}=\tilde{u}^1\boldsymbol{t}_1+\cdots+\tilde{u}^n\boldsymbol{t}_n=\tilde{u}^j\boldsymbol{t}_j$，则有 $u^i\boldsymbol{e}_i=\tilde{u}^j\boldsymbol{t}_j=\tilde{u}^jt^k{}_j\boldsymbol{e}_k=\tilde{u}^jt^i{}_j\boldsymbol{e}_i$，于是 $u^i=t^i{}_j\tilde{u}^j$.

根据命题 11.2.16 , 这组基的对偶基 $\boldsymbol{t}^1,\cdots,\boldsymbol{t}^n$ 满足 $(\boldsymbol{t}^1,\boldsymbol{t}^2,\cdots,\boldsymbol{t}^n)=(\boldsymbol{e}^1,\boldsymbol{e}^2,\cdots,\boldsymbol{e}^n)S$, 其中 $S=T^{-\mathrm{T}}$. 则有 $\boldsymbol{t}^i=s_1{}^i\boldsymbol{e}^1+\cdots+s_n{}^i\boldsymbol{e}^n=s_k{}^i\boldsymbol{e}^k$, 其中矩阵 $S=\left[s_i{}^j\right]$, 而 $s_i{}^j=\langle\boldsymbol{t}^j\mid\boldsymbol{e}_i\rangle$, $s_k{}^it^k{}_j=\delta^i{}_j$. 而若线性函数 f 在这组基下的坐标表示为 $f=\tilde{f}_1\boldsymbol{t}^1+\cdots+\tilde{f}_n\boldsymbol{t}^n=\tilde{f}_j\boldsymbol{t}^j$，则有 $f_i\boldsymbol{e}^i=\tilde{f}_j\boldsymbol{t}^j=\tilde{f}_js_k{}^j\boldsymbol{e}^k=\tilde{f}_js_i{}^j\boldsymbol{e}^i$，于是 $f_i=s_i{}^j\tilde{f}_j$.

另一方面，$T^{-1}=S^{\mathrm{T}}=\left[s^j{}_i\right], S^{-1}=T^{\mathrm{T}}=\left[t_j{}^i\right]$，则 $\tilde{u}^j=s^j{}_ku^k, \tilde{f}_j=t_j{}^kf_k, \boldsymbol{e}^j=t^j{}_k\boldsymbol{t}^k, \boldsymbol{e}_i=s_i{}^k\boldsymbol{t}_k$.

取值函数就是 $f(\boldsymbol{u})=\langle\boldsymbol{u}\mid f\rangle=\left\langle u^i\boldsymbol{e}_i\mid f_j\boldsymbol{e}^j\right\rangle=u^if_j\langle\boldsymbol{e}_i\mid\boldsymbol{e}^j\rangle=u^if_j\delta_i{}^j$，也是 $f(\boldsymbol{u})=\langle\boldsymbol{u}\mid f\rangle=\left\langle\tilde{u}^i\boldsymbol{t}_i\mid\tilde{f}_j\boldsymbol{t}^j\right\rangle=\tilde{u}^i\tilde{f}_j\langle\boldsymbol{t}_i\mid\boldsymbol{t}^j\rangle=\tilde{u}^i\tilde{f}_j\delta_i{}^j$，而换基给出 $u^if_j\delta_i{}^j=t^i{}_k\tilde{u}^ks_j{}^l\tilde{f}_l\delta_i{}^j=\tilde{u}^k\tilde{f}_lt^i{}_k\delta_i{}^js_j{}^l=\tilde{u}^k\tilde{f}_lt^i{}_ks_i{}^l=\tilde{u}^k\tilde{f}_l\delta_k{}^l$.

可以看到，随着基的变化，取值函数形式没有改变，而基或坐标的某个指标有两种变换规律. 不难看出 $\mathcal{U}$ 中向量的坐标，即 $\mathcal{U}^*$ 上线性函数逆变，而 $\mathcal{U}^*$ 中向量的坐标，即 $\mathcal{U}$ 上线性函数协变.

为此有如下定义.

定义 11.3.12 (协变与逆变) 随着基的变化，在基或坐标的某个指标的变换规律中，

1. 如果该指标是上指标而变换规律中会消去过渡矩阵的下指标，如 $\boldsymbol{t}^i=s_j{}^i\boldsymbol{e}^j, \tilde{u}^j=s^j{}_iu^i$，则称该指标关于基**逆变**或**反变**，又称该指标是**逆变指标**；
2. 如果该指标是下指标而变换规律中会消去过渡矩阵的上指标，如 $\boldsymbol{t}_i=t^j{}_i\boldsymbol{e}_j, \tilde{f}_j=t_j{}^iu_i$，则称该指标关于基**协变**或**共变**，又称该指标是**协变指标**.

若 p 阶张量的每个指标都是逆变指标，则称该张量为p **秩逆变张量**；若 q 阶张量的每个指标都是协变指标，则称该张量为q **秩协变张量**；若 $p+q$ 阶张量中有 p 个逆变指标和 q 个协变指标，则称该张量为 p 秩逆变 q 秩协变的**混合型张量**，简称(p,q) **型张量**.

利用 Einstein 记号和上述关于协变逆变的指标约定，就可以简单表述张量及其运算:

1. 当指定基后，张量直接表述为数组张量的元素，即 $A = a_{\cdots}{}^{\cdots i_p \cdots}{}_{\cdots i_q \cdots}{}^{\cdots}$；
2. 张量的线性运算：$C = kA + lB$ 表述为 $c_{\cdots}{}^{\cdots i_p \cdots}{}_{\cdots i_q \cdots}{}^{\cdots} = ka_{\cdots}{}^{\cdots i_p \cdots}{}_{\cdots i_q \cdots}{}^{\cdots} + lb_{\cdots}{}^{\cdots i_p \cdots}{}_{\cdots i_q \cdots}{}^{\cdots}$；
3. 张量的张量积：$C = A \otimes B$ 表述为 $c_{\cdots}{}^{\cdots i_p \cdots}{}_{\cdots i_q \cdots}{}^{\cdots}{}_{--}{}^{-i_r-}{}_{-i_s-}{}^{--} = a_{\cdots}{}^{\cdots i_p \cdots}{}_{\cdots i_q \cdots}{}^{\cdots} b_{--}{}^{-i_r-}{}_{-i_s-}{}^{--}$；
4. 张量的缩并 $C = \mathscr{C}_{i_p i_q}(A)$ 表述为 $c_{\cdots}{}^{\cdots \widehat{i_p} \cdots}{}_{\cdots \widehat{i_q} \cdots}{}^{\cdots} = a_{\cdots}{}^{\cdots k \cdots}{}_{\cdots k \cdots}{}^{\cdots}$，其中 $\widehat{\cdot}$ 表示不含该项.

如果张量积空间的各张量因子空间顺序不重要（在同构的意义下可以任意换位置），那么可以不使用错位的上下指标，例如 (p, q) 型张量 A 直接写为 $A = a^{i_1 \cdots i_p}{}_{j_1 \cdots j_q}$.

线性空间 $\mathcal{U}$ 是内积空间，因此对偶基可以用原空间的基线性表示 $(\boldsymbol{e}^1, \cdots, \boldsymbol{e}^n) = (\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n)G$. 注意到 G 是内积在基 $\boldsymbol{e}^1, \cdots, \boldsymbol{e}^n$ 下的度量矩阵：

$$
\begin{aligned}
\left[\langle \boldsymbol{e}^i, \boldsymbol{e}^j \rangle\right] &= \langle \boldsymbol{e}^1, \cdots, \boldsymbol{e}^n \mid \boldsymbol{e}^1, \cdots, \boldsymbol{e}^n \rangle \\
&= \langle (\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n)G \mid \boldsymbol{e}^1, \cdots, \boldsymbol{e}^n \rangle = \langle \boldsymbol{e}_1, \cdots, \boldsymbol{e}_n \mid \boldsymbol{e}^1, \cdots, \boldsymbol{e}^n \rangle G = G,
\end{aligned}
$$

因此，G 记为 $G = \left[g^{ij}\right]$，并有 $\boldsymbol{e}^j = g^{ij}\boldsymbol{e}_i$. 类似地，$G^{-1}$ 记为 $G^{-1} = \left[g_{ij}\right]$，并有 $\boldsymbol{e}_i = g_{ij}\boldsymbol{e}^j$，且 $g_{ik}g^{kj} = \delta_i{}^j$. 因此就可用通过度量矩阵 G 或 G^{-1} 来把下指标转换为上指标或把下指标转换为上指标，常称为升指标或降指标. 而

$$
\begin{aligned}
\gamma\colon \qquad & \mathcal{U} \to \mathcal{U}^*, \\
& (\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n)\hat{u} \mapsto (\boldsymbol{e}^1, \cdots, \boldsymbol{e}^n)\hat{u} = (\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n)G\hat{u},
\end{aligned}
$$

是同构映射，因此升一个指标其实就是利用同构映射把 (p, q) 型张量变换成了 $(p+1, q-1)$ 型张量；降一个指标其实就是利用同构映射把 (p, q) 型张量变换成了 $(p-1, q+1)$ 型张量.

11.4 反对称张量和外积

最后讨论线性空间 $\mathcal{U}$ 上的反对称张量. 记 $\mathcal{U}$ 上 r 阶反对称张量的全体为 $\mathcal{U}^{\wedge r}$. 易得如下结论.

命题 11.4.1 *1. $\mathcal{U}^{\wedge r}$ 是 r 阶张量积空间 $\mathcal{U}^{\otimes r}$ 的线性子空间；*
2. 若 $\dim \mathcal{U} = n < r$，则 $\mathcal{U}$ 上的 r 阶反对称张量是张量；
3. r 阶对称张量的全体是张量积空间 $\mathcal{U}^{\otimes r}$ 的线性子空间.

证 $\boxed{\textbf{留为作业}.}$ □

命题 11.4.1 说明，当 $r > n$ 时，$\dim \mathcal{U}^{\wedge r} = 0$.

根据行列式的定义和唯一性(参见定义 4.2.1 和命题 4.2.10), 当 $r=n$ 时, $(\mathbb{F}^n)^{\wedge n}$ 中的反对称的 n 重线性函数, 如果满足单位化条件, 则只能是行列式, 因此 $\dim(\mathbb{F}^n)^{\wedge n}=1$. 再利用 $\mathbb{F}^n$ 到 $\mathcal{U}$ 的同构可以自然地构造 $(\mathbb{F}^n)^{\wedge n}$ 到 $\mathcal{U}^{\wedge n}$ 的同构, 因此 $\dim\mathcal{U}^{\wedge n}=1$.

当 $r=1$ 时, $\mathcal{U}^{\wedge 1}$ 中的反对称的 1 重线性函数就是普通的线性函数, 因此 $\mathcal{U}^{\wedge 1}=\mathcal{U}^{\otimes 1}=\mathcal{U}$, 而 $\dim\mathcal{U}^{\wedge 1}=n$.

下面就来考虑 $1<r<n$ 时, $\mathcal{U}^{\wedge r}$ 的基和维数. 取定 $\mathcal{U}$ 的一组基 $\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n$, 则张量积空间 $\mathcal{U}^{\otimes r}$ 的一组基为

$$\boldsymbol{e}_{i_1}\otimes\cdots\otimes\boldsymbol{e}_{i_r}, i_p=1,\cdots,n, p=1,\cdots,r,$$

而 r 阶张量 A 必可表示为 $A=\sum_{i_1,\cdots,i_r}a_{i_1\cdots i_r}\boldsymbol{e}_{i_1}\otimes\cdots\otimes\boldsymbol{e}_{i_r}$. 若 A 反对称, 则其表示 $\left[a_{i_1\cdots i_r}\right]$ 是反对称张量, 其中的数有不少只是有与排列相关的正负号的差别. 由于这里的 $i_1\cdots i_r$ 这 r 个数可能不是连续的 1 到 r, 为此定义如下排列的运算.

定义 11.4.2 (排列的复合、直和) 任意 $1,\cdots,n$ 的排列 $\sigma_1,\cdots,\sigma_n$ 自然定义映射 $\sigma\colon i\mapsto\sigma_i$ 和置换矩阵 $P_\sigma=\begin{bmatrix}\boldsymbol{e}_{\sigma_1} & \boldsymbol{e}_{\sigma_2} & \cdots & \boldsymbol{e}_{\sigma_n}\end{bmatrix}$.

定义排列 σ 和 τ 的复合为对应映射的复合, 即 $\sigma\tau\colon i\to\sigma(\tau_i)$, 这也对应了置换矩阵的乘法 $P_{\sigma\tau}=P_\tau P_\sigma$.

若置换矩阵 $P_\sigma=\begin{bmatrix}P_{\sigma_1} & \\ & P_{\sigma_2}\end{bmatrix}$ 是分块对角矩阵, 则定义 σ 是 σ_1 和 σ_2 的直和, 记为 $\sigma=\sigma_1\oplus\sigma_2$.

利用上述运算就可以把 r 元排列 σ 在 r 个不连续的数 $i_1,\cdots,i_r$ 上的作用定义为 $\sigma(i_j)=(\sigma\oplus\mathrm{id}_{n-r})(i_j), j=1,\cdots,r$, 其中 n 是一个不小于所有 i_j 的正整数.

于是, 对任意排列 σ, $a_{\sigma(i_1)\cdots\sigma(i_r)}=\mathrm{sign}(\sigma)a_{i_1\cdots i_r}$. 由此可得, 若 $i_1,\cdots,i_r$ 中有两个及以上相同的, 则 $a_{i_1\cdots i_r}=0$, 于是

$$A=\sum_{i_1<\cdots<i_r}a_{i_1\cdots i_r}\sum_{\text{排列 }\sigma}\mathrm{sign}(\sigma)\boldsymbol{e}_{\sigma(i_1)}\otimes\cdots\otimes\boldsymbol{e}_{\sigma(i_r)}.$$

另一方面, 若 A 是零张量, 则必有对任意 $i_1,\cdots,i_r$, $a_{i_1\cdots i_r}=0$, 故 $\sum_{\text{排列 }\sigma}\mathrm{sign}(\sigma)\boldsymbol{e}_{\sigma(i_1)}\otimes\cdots\otimes\boldsymbol{e}_{\sigma(i_r)}$ 线性无关. 这就说明

$$\boldsymbol{e}_{i_1}\wedge\cdots\wedge\boldsymbol{e}_{i_r}:=\sum_{\text{排列 }\sigma}\mathrm{sign}(\sigma)\boldsymbol{e}_{\sigma(i_1)}\otimes\cdots\otimes\boldsymbol{e}_{\sigma(i_r)}, 1\leqslant i_1<\cdots<i_r\leqslant n, \tag{11.4.1}$$

是 r 阶反对称张量空间的一组基, 而维数 $\dim\mathcal{U}^{\wedge r}=\mathrm{C}_n^r$.

稍微扩展 (11.4.1) 式有如下定义.

定义 11.4.3 (向量的外积) 线性空间 $\mathcal{U}$ 中 r 个向量 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r$ 的**外积**定义为

$$\boldsymbol{a}_1\wedge\cdots\wedge\boldsymbol{a}_r:=\sum_{\text{排列 }\sigma}\mathrm{sign}(\sigma)\boldsymbol{a}_{\sigma(1)}\otimes\cdots\otimes\boldsymbol{a}_{\sigma(r)}.$$

命题 11.4.4 *1.* 向量的外积满足多线性律.

2. 对排列 σ，$\boldsymbol{a}_{\sigma(1)}\wedge\cdots\wedge\boldsymbol{a}_{\sigma(r)}=\operatorname{sign}(\sigma)\boldsymbol{a}_1\wedge\cdots\wedge\boldsymbol{a}_r$.

3. 若 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r$ 中有至少两个相同的向量，则 $\boldsymbol{a}_1\wedge\cdots\wedge\boldsymbol{a}_r=\boldsymbol{0}$.

4. $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_r$ 线性相关，当且仅当 $\boldsymbol{a}_1\wedge\cdots\wedge\boldsymbol{a}_r=\boldsymbol{0}$.

5. $$\langle\boldsymbol{a}_1\otimes\cdots\otimes\boldsymbol{a}_r\mid f_1\wedge\cdots\wedge f_r\rangle=\det(\begin{bmatrix}\langle\boldsymbol{a}_j\mid f_i\rangle\end{bmatrix})=\begin{vmatrix}\langle\boldsymbol{a}_1\mid f_1\rangle & \cdots & \langle\boldsymbol{a}_r\mid f_1\rangle\\ \vdots & & \vdots\\ \langle\boldsymbol{a}_1\mid f_r\rangle & \cdots & \langle\boldsymbol{a}_r\mid f_r\rangle\end{vmatrix}.$$

6. $$\langle\boldsymbol{a}_1\wedge\cdots\wedge\boldsymbol{a}_r\mid f_1\wedge\cdots\wedge f_r\rangle=r!\det(\begin{bmatrix}\langle\boldsymbol{a}_j\mid f_i\rangle\end{bmatrix})=r!\begin{vmatrix}\langle\boldsymbol{a}_1\mid f_1\rangle & \cdots & \langle\boldsymbol{a}_r\mid f_1\rangle\\ \vdots & & \vdots\\ \langle\boldsymbol{a}_1\mid f_r\rangle & \cdots & \langle\boldsymbol{a}_r\mid f_r\rangle\end{vmatrix}.$$

证 **留为作业.** □

反对称张量和向量的外积有很明确的几何含义，这是它得到关注的根本原因.

例 11.4.5 我们已经知道，三维空间 $\mathbb{R}^3$ 中的以 $\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3$ 为棱的平行六面体的有向体积是 $\det(\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{a}_3)$；二维空间 $\mathbb{R}^2$ 中的以 $\boldsymbol{a}_1,\boldsymbol{a}_2$ 为边的平行四边形的有向面积是 $\det(\boldsymbol{a}_1,\boldsymbol{a}_2)$. 那么三维空间 $\mathbb{R}^3$ 中的以 $\boldsymbol{a}_1,\boldsymbol{a}_2$ 为边的平行四边形的有向面积如何计算和表示?

练习 4.3.21 给出了这个平行四边形的有向面积的绝对值,为 $\sqrt{\det(\begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2\end{bmatrix})}$. 但其符号却难以直接确定. 因此我们还是从行列式表示有向面积或有向体积的几何含义出发来找出我们需要的有向面积的表达式. 设有向面积为 $S(\boldsymbol{a}_1,\boldsymbol{a}_2)$，从几何含义上看，$S$ 需要有如下性质:

1. 旋转不变：对任意行列式为 1 的 3 阶正交矩阵 Q，$S(Q\boldsymbol{a}_1,Q\boldsymbol{a}_2)=S(\boldsymbol{a}_1,\boldsymbol{a}_2)$，这是因为含参数的旋转可以构成一个连续运动；
2. 局部化：$S(\begin{bmatrix}\hat{\boldsymbol{a}}_1\\0\end{bmatrix},\begin{bmatrix}\hat{\boldsymbol{a}}_2\\0\end{bmatrix})=\det(\hat{\boldsymbol{a}}_1,\hat{\boldsymbol{a}}_2)$，即 xOy 平面内的图形在三维空间中的有向面积应与平面内的有向面积一致.

下面来计算 $S(\boldsymbol{a}_1,\boldsymbol{a}_2)$. 计算 QR 分解 $\begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2\end{bmatrix}=QR$，其中 $R=\begin{bmatrix}r_{11} & r_{12}\\0 & r_{22}\\0 & 0\end{bmatrix}$. 不妨设 $\det(Q)=1$. 若不然，$\det(Q)=-1$，则取 $Q_1=\operatorname{diag}(1,1,-1)$，于是 $\begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2\end{bmatrix}=QQ_1Q_1R=(QQ_1)(Q_1R)=(QQ_1)R$，令 QQ_1 为所需的 Q 即可. 因此

$$S(\boldsymbol{a}_1,\boldsymbol{a}_2)=\begin{vmatrix}r_{11} & r_{12}\\ & r_{22}\end{vmatrix}=\begin{vmatrix}r_{11} & r_{12} & 0\\0 & r_{22} & 0\\0 & 0 & 1\end{vmatrix}=\det(Q\begin{bmatrix}R & \boldsymbol{e}_3\end{bmatrix})=\det(\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{q}_3).$$

从几何含义上看，$\boldsymbol{q}_3$ 是 $\mathrm{span}(\boldsymbol{a}_1, \boldsymbol{a}_2)$ 上的方向确定的单位法向量.

可见 S 是 $\mathcal{V} = \mathrm{span}(\boldsymbol{a}_1, \boldsymbol{a}_2)$ 上的满足单位化条件的反对称双线性函数. 记 $\mathcal{V}$ 的一组标准正交基为 $\boldsymbol{q}_1, \boldsymbol{q}_2$，设 S 在张量积空间 $\mathcal{V}^* \otimes \mathcal{V}^*$ 的对应基下的表示为 $S = \sum_{i,j=1,2} S_{ij} \boldsymbol{q}_i^* \otimes \boldsymbol{q}_j^*$，利用反对称和单位化条件有 $S = \boldsymbol{q}_1^* \otimes \boldsymbol{q}_2^* - \boldsymbol{q}_2^* \otimes \boldsymbol{q}_1^* = \boldsymbol{q}_1^* \wedge \boldsymbol{q}_2^*$. 而 $S(\boldsymbol{a}_1, \boldsymbol{a}_2) = \langle S \mid \boldsymbol{a}_1 \otimes \boldsymbol{a}_2 \rangle = r_{11} r_{22}$. 容易验证

$$
\begin{aligned}
\boldsymbol{a}_1 \wedge \boldsymbol{a}_2 &= \boldsymbol{a}_1 \otimes \boldsymbol{a}_2 - \boldsymbol{a}_2 \otimes \boldsymbol{a}_1 \\
&= (r_{11}\boldsymbol{q}_1) \otimes (r_{12}\boldsymbol{q}_1 + r_{22}\boldsymbol{q}_2) - (r_{12}\boldsymbol{q}_1 + r_{22}\boldsymbol{q}_2) \otimes (r_{11}\boldsymbol{q}_1) \\
&= r_{11} r_{22} (\boldsymbol{q}_1 \otimes \boldsymbol{q}_2 - \boldsymbol{q}_2 \otimes \boldsymbol{q}_1) \\
&= S(\boldsymbol{a}_1, \boldsymbol{a}_2)(\boldsymbol{q}_1 \wedge \boldsymbol{q}_2),
\end{aligned}
$$

因此 $\boldsymbol{a}_1 \wedge \boldsymbol{a}_2$ 就是 $\mathbb{R}^3$ 上的一个 2 阶反对称张量，其中 $\boldsymbol{q}_1 \wedge \boldsymbol{q}_2$ 表示 $\mathrm{span}(\boldsymbol{a}_1, \boldsymbol{a}_2)$ 上一个标准平行四边形，亦即表示平面 $\mathrm{span}(\boldsymbol{a}_1, \boldsymbol{a}_2)$，而 $S(\boldsymbol{a}_1, \boldsymbol{a}_2)$ 表示 $\boldsymbol{a}_1, \boldsymbol{a}_2$ 为边的平行四边形的有向面积. 这说明 $\boldsymbol{a}_1 \wedge \boldsymbol{a}_2$ 既表示平面上平行四边形的有向面积，又表示平行四边形所在平面的方向. 这比行列式仅仅代表有向面积又丰富了一些. ☺

接着来看反对称张量坐标在基变换下的变化规律.

为指称方便，对方阵 A，记 A 的第 $i_1, \cdots, i_r$ 行和第 $j_1, \cdots, j_r$ 列的交叉点组成的 r 阶方阵为

$$
A\left\{\begin{matrix} i_1 \cdots i_r \\ j_1 \cdots j_r \end{matrix}\right\} = \begin{bmatrix} a_{i_1 j_1} & \cdots & a_{i_1 j_r} \\ \vdots & & \vdots \\ a_{i_r j_1} & \cdots & a_{i_r j_r} \end{bmatrix},
$$

对应地，从 A 中划去第 $i_1, \cdots, i_r$ 行和第 $j_1, \cdots, j_r$ 列得到的 $n-r$ 阶方阵为 $A\left(\begin{smallmatrix} i_1 \cdots i_r \\ j_1 \cdots j_r \end{smallmatrix}\right)$.

命题 11.4.6 给定数域 $\mathbb{F}$ 上线性空间 $\mathcal{U}$ 和它的两组基 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n; \boldsymbol{t}_1, \boldsymbol{t}_2, \cdots, \boldsymbol{t}_n$. 记过渡矩阵为 T，即 $(\boldsymbol{t}_1, \boldsymbol{t}_2, \cdots, \boldsymbol{t}_n) = (\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n)T$. 则 $\mathcal{U}^{\wedge r}, r \leqslant n$ 的两组对应基的过渡矩阵为 $\left[\det\left(T\left\{\begin{smallmatrix} j_1 \cdots j_r \\ i_1 \cdots i_r \end{smallmatrix}\right\}\right)\right]$，即

$$
\begin{pmatrix} \boldsymbol{t}_{i_1} \wedge \cdots \wedge \boldsymbol{t}_{i_r}, \\ 1 \leqslant i_1 < \cdots < i_r \leqslant n \end{pmatrix} = \begin{pmatrix} \boldsymbol{e}_{j_1} \wedge \cdots \wedge \boldsymbol{e}_{j_r}, \\ 1 \leqslant j_1 < \cdots < j_r \leqslant n \end{pmatrix} \left[\det\left(T\left\{\begin{matrix} j_1 \cdots j_r \\ i_1 \cdots i_r \end{matrix}\right\}\right)\right].
$$

特别地，$\boldsymbol{t}_1 \wedge \cdots \wedge \boldsymbol{t}_n = \det(T) \boldsymbol{e}_1 \wedge \cdots \wedge \boldsymbol{e}_n$.

证 直接计算可得：

$$
\begin{aligned}
\boldsymbol{t}_{i_1} \wedge \cdots \wedge \boldsymbol{t}_{i_r} &= \sum_{\text{排列 } \sigma} \mathrm{sign}(\sigma) \boldsymbol{t}_{\sigma(i_1)} \otimes \cdots \otimes \boldsymbol{t}_{\sigma(i_r)} \\
&= \sum_{\text{排列 } \sigma} \mathrm{sign}(\sigma) \sum_{j_1} t_{j_1 \sigma(i_1)} \boldsymbol{e}_{j_1} \otimes \cdots \otimes \sum_{j_r} t_{j_r \sigma(i_r)} \boldsymbol{e}_{j_r}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{j_1,\cdots,j_r} \sum_{\text{排列 } \sigma} \operatorname{sign}(\sigma) t_{j_1\sigma(i_1)} \cdots t_{j_r\sigma(i_r)} \boldsymbol{e}_{j_1} \otimes \cdots \otimes \boldsymbol{e}_{j_r} \\
\text{记 } \widetilde{T} = T\begin{Bmatrix} j_1 \cdots j_r \\ i_1 \cdots i_r \end{Bmatrix}, \qquad &= \sum_{j_1,\cdots,j_r} \sum_{\text{排列 } \sigma} \operatorname{sign}(\sigma) \tilde{t}_{1\sigma(1)} \cdots t_{r\sigma(r)} \boldsymbol{e}_{j_1} \otimes \cdots \otimes \boldsymbol{e}_{j_r} \\
&= \sum_{j_1,\cdots,j_r} \det(\widetilde{T}) \boldsymbol{e}_{j_1} \otimes \cdots \otimes \boldsymbol{e}_{j_r} \\
&= \sum_{j_1,\cdots,j_r} \det\left(T\begin{Bmatrix} j_1 \cdots j_r \\ i_1 \cdots i_r \end{Bmatrix}\right) \boldsymbol{e}_{j_1} \otimes \cdots \otimes \boldsymbol{e}_{j_r} \\
&= \sum_{j_1<\cdots<j_r} \det\left(T\begin{Bmatrix} j_1 \cdots j_r \\ i_1 \cdots i_r \end{Bmatrix}\right) \boldsymbol{e}_{j_1} \wedge \cdots \wedge \boldsymbol{e}_{j_r}. \qquad \square
\end{aligned}
$$

例 11.4.7 继续分析例 11.4.5 .

$$
\boldsymbol{a}_1 \wedge \boldsymbol{a}_2 = S(\boldsymbol{a}_1, \boldsymbol{a}_2)(\boldsymbol{q}_1 \wedge \boldsymbol{q}_2) = (\boldsymbol{q}_1 \wedge \boldsymbol{q}_2)\langle \boldsymbol{q}_1^* \wedge \boldsymbol{q}_2^* \mid \boldsymbol{a}_1 \otimes \boldsymbol{a}_2 \rangle =: P(\boldsymbol{a}_1 \otimes \boldsymbol{a}_2),
$$

其中 P 是张量积空间上的线性映射. 由于

$$
\begin{aligned}
P^2(\boldsymbol{a}_1 \otimes \boldsymbol{a}_2) &= P\left(S(\boldsymbol{a}_1, \boldsymbol{a}_2)(\boldsymbol{q}_1 \wedge \boldsymbol{q}_2)\right) \\
&= (\boldsymbol{q}_1 \wedge \boldsymbol{q}_2)\langle \boldsymbol{q}_1^* \wedge \boldsymbol{q}_2^* \mid S(\boldsymbol{a}_1, \boldsymbol{a}_2)(\boldsymbol{q}_1 \wedge \boldsymbol{q}_2)\rangle \\
&= 2S(\boldsymbol{a}_1, \boldsymbol{a}_2)(\boldsymbol{q}_1 \wedge \boldsymbol{q}_2) = 2P(\boldsymbol{a}_1 \otimes \boldsymbol{a}_2),
\end{aligned}
$$

其中 $\langle \boldsymbol{q}_1^* \wedge \boldsymbol{q}_2^* \mid \boldsymbol{q}_1 \wedge \boldsymbol{q}_2 \rangle = 2$. 而 P 的对偶映射 $P^* = P$, 因此 $\frac{1}{2}P$ 是正交投影. 考虑到 $\frac{1}{\sqrt{2}}\boldsymbol{q}_1 \wedge \boldsymbol{q}_2$ 是 $\mathcal{V}^{\wedge 2}$ 的一组标准正交基, 因此 $\frac{1}{2}P$ 就是 $\mathcal{V}^{\wedge 2}$ 上的正交投影. ☺

例 11.4.5 提示如下结论.

命题 11.4.8 *给定线性空间 $\mathcal{U}$, 张量积空间 $\mathcal{U}^{\otimes r}$ 的子空间 $\mathcal{U}^{\wedge r}$ 上的正交投影映射是*

$$
P_{\wedge r} \colon A \mapsto \frac{1}{r!} \sum_{i_1<\cdots<i_r} (\boldsymbol{e}_{i_1} \wedge \cdots \wedge \boldsymbol{e}_{i_r}) \left\langle \boldsymbol{e}_{i_1}^* \wedge \cdots \wedge \boldsymbol{e}_{i_r}^* \mid A \right\rangle,
$$

其中 $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n$ 是 $\mathcal{U}$ 的一组基, 而 $\boldsymbol{e}_1^*, \cdots, \boldsymbol{e}_n^*$ 是其对偶基.

特别地,

1. *$P_{\wedge r}(\boldsymbol{a}_1 \otimes \cdots \otimes \boldsymbol{a}_r) = \frac{1}{r!}\boldsymbol{a}_1 \wedge \cdots \wedge \boldsymbol{a}_r$, 即向量的张量积在反对称张量空间的正交投影是向量的外积的 $\frac{1}{r!}$ 倍;*
2. *若 $A = \sum_{i_1,\cdots,i_r} a_{i_1\cdots i_r} \boldsymbol{e}_{i_1} \otimes \cdots \otimes \boldsymbol{e}_{i_r}$, 则*

$$
P_{\wedge r}(A) = \sum_{i_1<\cdots<i_r} \left(\frac{1}{r!} \sum_{\text{排列 } \sigma} \operatorname{sign}(\sigma) a_{\sigma(i_1)\cdots\sigma(i_r)} \right) \boldsymbol{e}_{i_1} \wedge \cdots \wedge \boldsymbol{e}_{i_r}.
$$

证 注意验证 $P_{\wedge r}$ 与基的选择无关. $\boxed{\textbf{留为作业.}}$ □

利用命题 11.4.8，可以定义任意两个反对称张量的外积.

对任意 p 阶张量 A 和 q 阶张量 B，$A\otimes B$ 是 $p+q$ 阶张量. 这些张量在反对称张量子空间的正交投影就分别为 $P_{\wedge p}(A), P_{\wedge q}(B), P_{\wedge(p+q)}(A\otimes B)$，如果张量外积 $\widetilde{\wedge}$ 有 $P_{\wedge p}(A)\,\widetilde{\wedge}\,P_{\wedge q}(B) = P_{\wedge(p+q)}(A\otimes B)$，那么外积会有比较好的性质. 另一方面，如果 A, B 都是可分张量，那么 $A\otimes B$ 也是可分张量，则上式化为

$$\frac{1}{p!}(\boldsymbol{a}_1\wedge\cdots\wedge\boldsymbol{a}_p)\,\widetilde{\wedge}\,\frac{1}{q!}(\boldsymbol{b}_1\wedge\cdots\wedge\boldsymbol{b}_q) = \frac{1}{(p+q)!}(\boldsymbol{a}_1\wedge\cdots\wedge\boldsymbol{a}_p\wedge\boldsymbol{b}_1\wedge\cdots\wedge\boldsymbol{b}_q),$$

这说明不考虑数乘时向量外积满足结合律. 为使外积直接满足结合律，我们在利用正交投影定义外积时应直接数乘：$\wedge = \frac{(p+q)!}{p!q!}\widetilde{\wedge}$. 这就得到了如下定义.

定义 11.4.9 (反对称张量的外积) 给定线性空间 $\mathcal{U}$ 上 p 阶反对称张量 A 和 q 阶反对称张量 B，定义 A 和 B 的**外积**为 $p+q$ 阶反对称张量

$$A\wedge B := \frac{(p+q)!}{p!q!}P_{\wedge(p+q)}(A\otimes B).$$

命题 11.4.10 若 $\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n$ 是 $\mathcal{U}$ 的一组基，则

$$\left(\sum_{i_1<\cdots<i_p} a_{i_1\cdots i_p}\boldsymbol{e}_{i_1}\wedge\cdots\wedge\boldsymbol{e}_{i_p}\right)\wedge\left(\sum_{i_{p+1}<\cdots<i_{p+q}} b_{i_{p+1}\cdots i_{p+q}}\boldsymbol{e}_{i_{p+1}}\wedge\cdots\wedge\boldsymbol{e}_{i_{p+q}}\right)$$
$$=\sum_{i_1<\cdots<i_{p+q}}\left(\sum_{\substack{i_1,\cdots,i_{p+q}\text{ 的子集}\\ j_1<\cdots<j_p}}\operatorname{sign}(\sigma_{j_1\cdots j_p})a_{j_1\cdots j_p}b_{j_{p+1}\cdots j_{p+q}}\right)\boldsymbol{e}_{i_1}\wedge\cdots\wedge\boldsymbol{e}_{i_{p+q}},$$

其中 $\sigma_{j_1\cdots j_p}$ 是从 $i_1,\cdots,i_p,i_{p+1},\cdots,i_{p+q}$ 到 $j_1,\cdots,j_p,j_{p+1},\cdots,j_{p+q}$ 的排列，而 $j_{p+1}<\cdots<j_{p+q}$ 是 $i_1,\cdots,i_{p+q}$ 中不在 $j_1,\cdots,j_p$ 中的数.

证 直接计算，

$$\begin{aligned}
A\otimes B &= \left(\sum_{i_1,\cdots,i_p} a_{i_1\cdots i_p}\boldsymbol{e}_{i_1}\otimes\cdots\otimes\boldsymbol{e}_{i_p}\right)\otimes\left(\sum_{j_1,\cdots,j_q} b_{j_1\cdots j_q}\boldsymbol{e}_{j_1}\otimes\cdots\otimes\boldsymbol{e}_{j_q}\right)\\
&= \sum_{i_1,\cdots,i_p,j_1,\cdots,j_q} a_{i_1\cdots i_p}b_{j_1\cdots j_q}\boldsymbol{e}_{i_1}\otimes\cdots\otimes\boldsymbol{e}_{i_p}\otimes\boldsymbol{e}_{j_1}\otimes\cdots\otimes\boldsymbol{e}_{j_q}\\
&= \sum_{i_1,\cdots,i_p,i_{p+1},\cdots,i_{p+q}} a_{i_1\cdots i_p}b_{i_{p+1}\cdots i_{p+q}}\boldsymbol{e}_{i_1}\otimes\cdots\otimes\boldsymbol{e}_{i_p}\otimes\boldsymbol{e}_{i_{p+1}}\otimes\cdots\otimes\boldsymbol{e}_{i_{p+q}},
\end{aligned}$$

根据命题 11.4.8，

$$A\wedge B = \frac{(p+q)!}{p!q!}\sum_{i_1<\cdots<i_{p+q}}\left(\frac{1}{(p+q)!}\sum_{\sigma}\operatorname{sign}(\sigma)a_{\sigma(i_1)\cdots\sigma(i_p)}b_{\sigma(i_{p+1})\cdots\sigma(i_{p+q})}\right)\boldsymbol{e}_{i_1}\wedge\cdots\wedge\boldsymbol{e}_{i_{p+q}}$$

$$= \sum_{i_1<\cdots<i_{p+q}} \underbrace{\left(\frac{1}{p!q!}\sum_{\text{排列 }\sigma} \operatorname{sign}(\sigma) a_{\sigma(i_1)\cdots\sigma(i_p)} b_{\sigma(i_{p+1})\cdots\sigma(i_{p+q})}\right)}_{=:c_{i_1\cdots i_{p+q}}} \boldsymbol{e}_{i_1}\wedge\cdots\wedge\boldsymbol{e}_{i_{p+q}}.$$

下面计算坐标. 将排列 σ 分为三个过程的复合: 把 $i_1,\cdots,i_{p+q}$ 分成不交的两组 $j_1<\cdots<j_p$ 和 $j_{p+1}<\cdots<j_{p+q}$, 前一组的排列 σ_1, 后一组的排列 σ_2. 故 $\sigma=(\sigma_1\oplus\sigma_2)\sigma_{j_1\cdots j_p}$. 因此

$$\begin{aligned}
c_{i_1\cdots i_{p+q}} &= \frac{1}{p!q!}\sum_{\substack{i_1,\cdots,i_{p+q}\text{ 的子集}\\ j_1<\cdots<j_p}}\sum_{\text{排列 }\sigma_1}\sum_{\text{排列 }\sigma_2} \operatorname{sign}\left((\sigma_1\oplus\sigma_2)\sigma_{j_1\cdots j_p}\right)\operatorname{sign}(\sigma_1) a_{j_1\cdots j_p}\operatorname{sign}(\sigma_2) b_{j_{p+1}\cdots j_{p+q}}\\
&= \frac{1}{p!q!}\sum_{\substack{i_1,\cdots,i_{p+q}\text{ 的子集}\\ j_1<\cdots<j_p}}\sum_{\text{排列 }\sigma_1}\sum_{\text{排列 }\sigma_2} \operatorname{sign}(\sigma_{j_1\cdots j_p}) a_{j_1\cdots j_p} b_{j_{p+1}\cdots j_{p+q}}\\
&= \frac{1}{p!q!}\sum_{\substack{i_1,\cdots,i_{p+q}\text{ 的子集}\\ j_1<\cdots<j_p}} p!q!\operatorname{sign}(\sigma_{j_1\cdots j_p}) a_{j_1\cdots j_p} b_{j_{p+1}\cdots j_{p+q}}\\
&= \sum_{\substack{i_1,\cdots,i_{p+q}\text{ 的子集}\\ j_1<\cdots<j_p}} \operatorname{sign}(\sigma_{j_1\cdots j_p}) a_{j_1\cdots j_p} b_{j_{p+1}\cdots j_{p+q}}.
\end{aligned}$$

即得结论. □

命题 11.4.11 *1.* 反对称张量的外积满足结合律和双线性律.

2. $A\wedge B=(-1)^{pq}B\wedge A$.

证 **留为作业**. □

下面来看两个在行列式方面的例子.

例 11.4.12 (Laplace 公式) 我们利用外积的结合律来重新分析命题 11.4.6 中 $\mathcal{U}^{\wedge n}$ 的情形.

注意 $\boldsymbol{t}_1\wedge\cdots\wedge\boldsymbol{t}_n=(\boldsymbol{t}_1\wedge\cdots\wedge\boldsymbol{t}_r)\wedge(\boldsymbol{t}_{r+1}\wedge\cdots\wedge\boldsymbol{t}_n)$. 而

$$\boldsymbol{t}_1\wedge\cdots\wedge\boldsymbol{t}_r=\sum_{i_1<\cdots<i_r}\det\left(T\begin{Bmatrix} i_1\cdots i_r\\ 1\cdots r\end{Bmatrix}\right)\boldsymbol{e}_{i_1}\wedge\cdots\wedge\boldsymbol{e}_{i_r},$$

$$\boldsymbol{t}_{r+1}\wedge\cdots\wedge\boldsymbol{t}_n=\sum_{i_{r+1}<\cdots<i_n}\det\left(T\begin{Bmatrix} i_{r+1}\cdots i_n\\ r+1\cdots n\end{Bmatrix}\right)\boldsymbol{e}_{i_{r+1}}\wedge\cdots\wedge\boldsymbol{e}_{i_n},$$

由命题 11.4.10 有,

$$\boldsymbol{t}_1\wedge\cdots\wedge\boldsymbol{t}_n=\sum_{i_1<\cdots<i_r<i_{r+1}<\cdots<i_n}\operatorname{sign}(\sigma_{i_1\cdots i_r})\det\left(T\begin{Bmatrix} i_1\cdots i_r\\ 1\cdots r\end{Bmatrix}\right)\det\left(T\begin{Bmatrix} i_{r+1}\cdots i_n\\ r+1\cdots n\end{Bmatrix}\right)\boldsymbol{e}_1\wedge\cdots\wedge\boldsymbol{e}_n,$$

因此

$$\begin{aligned}\det(T) &= \sum_{i_1<\cdots<i_r<i_{r+1}<\cdots<i_n} \operatorname{sign}(\sigma_{i_1\cdots i_r}) \det\left(T\begin{Bmatrix} i_1\cdots i_r \\ 1\cdots r\end{Bmatrix}\right) \det\left(T\begin{Bmatrix} i_{r+1}\cdots i_n \\ r+1\cdots n\end{Bmatrix}\right) \\ &= \sum_{i_1<\cdots<i_r} \operatorname{sign}(\sigma_{i_1\cdots i_r}) \det\left(T\begin{Bmatrix} i_1\cdots i_r \\ 1\cdots r\end{Bmatrix}\right) \det\left(T\begin{pmatrix} i_1\cdots i_r \\ 1\cdots r\end{pmatrix}\right),\end{aligned}$$

其中 $\operatorname{sign}(\sigma_{i_1\cdots i_r})$ 是 -1 的 $(r+1-i_{r+1})+\cdots+(n-i_n)=(r+1)+\cdots+n-(i_{r+1}+\cdots+i_n)=i_1+\cdots+i_r-(1+\cdots+r)$ 次方.

类似地，通过对列置换可得：对任意 r 个不大于 n 的正整数 $j_1<\cdots<j_r$ 有

$$\det(T) = \sum_{i_1<\cdots<i_r} (-1)^{i_1+\cdots+i_r+j_1+\cdots+j_r} \det\left(T\begin{Bmatrix} i_1\cdots i_r \\ j_1\cdots j_r\end{Bmatrix}\right) \det\left(T\begin{pmatrix} i_1\cdots i_r \\ j_1\cdots j_r\end{pmatrix}\right),$$

这就是行列式的 Laplace 展开式，即行列式可以按若干行展开. ☺

例 11.4.13 (Binet-Cauchy 公式) 利用命题 11.4.4 第 6 条计算矩阵乘积的行列式. 由

$$\begin{aligned}\boldsymbol{a}_1\wedge\cdots\wedge\boldsymbol{a}_r &= \sum_{i_1<\cdots<i_r} \det\left(A\begin{Bmatrix} i_1\cdots i_r \\ 1\cdots r\end{Bmatrix}\right) \boldsymbol{e}_{i_1}\wedge\cdots\wedge\boldsymbol{e}_{i_r}, \\ \boldsymbol{b}_1\wedge\cdots\wedge\boldsymbol{b}_r &= \sum_{j_1<\cdots<j_r} \det\left(B\begin{Bmatrix} j_1\cdots j_r \\ 1\cdots r\end{Bmatrix}\right) \boldsymbol{e}_{j_1}\wedge\cdots\wedge\boldsymbol{e}_{j_r},\end{aligned}$$

有

$$\begin{aligned}\det(A^{\mathrm{T}}B) &= \det(\left[\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{b}_j\right]) \\ &= \frac{1}{r!}\langle \boldsymbol{a}_1\wedge\cdots\wedge\boldsymbol{a}_r, \boldsymbol{b}_1\wedge\cdots\wedge\boldsymbol{b}_r\rangle \\ &= \frac{1}{r!}\left\langle \sum_{i_1<\cdots<i_r} \det\left(A\begin{Bmatrix} i_1\cdots i_r \\ 1\cdots r\end{Bmatrix}\right) \boldsymbol{e}_{i_1}\wedge\cdots\wedge\boldsymbol{e}_{i_r}, \sum_{j_1<\cdots<j_r} \det\left(B\begin{Bmatrix} j_1\cdots j_r \\ 1\cdots r\end{Bmatrix}\right) \boldsymbol{e}_{j_1}\wedge\cdots\wedge\boldsymbol{e}_{j_r}\right\rangle \\ &= \frac{1}{r!}\sum_{i_1<\cdots<i_r}\sum_{j_1<\cdots<j_r} \det\left(A\begin{Bmatrix} i_1\cdots i_r \\ 1\cdots r\end{Bmatrix}\right) \det\left(B\begin{Bmatrix} j_1\cdots j_r \\ 1\cdots r\end{Bmatrix}\right) \left\langle \boldsymbol{e}_{i_1}\wedge\cdots\wedge\boldsymbol{e}_{i_r}, \boldsymbol{e}_{j_1}\wedge\cdots\wedge\boldsymbol{e}_{j_r}\right\rangle \\ &= \frac{1}{r!}\sum_{i_1<\cdots<i_r}\sum_{j_1<\cdots<j_r} \det\left(A\begin{Bmatrix} i_1\cdots i_r \\ 1\cdots r\end{Bmatrix}\right) \det\left(B\begin{Bmatrix} j_1\cdots j_r \\ 1\cdots r\end{Bmatrix}\right) r!\det(\left[\boldsymbol{e}_{i_k}^{\mathrm{T}}\boldsymbol{e}_{j_l}\right])\end{aligned}$$

若 $i_1\cdots i_r$ 与 $j_1\cdots j_r$ 不完全相同，则 $\left[\boldsymbol{e}_{i_k}^{\mathrm{T}}\boldsymbol{e}_{j_l}\right]$ 有至少一列为零；反之矩阵为 I_r，

$$= \sum_{i_1<\cdots<i_r} \det\left(A\begin{Bmatrix} i_1\cdots i_r \\ 1\cdots r\end{Bmatrix}\right) \det\left(B\begin{Bmatrix} i_1\cdots i_r \\ 1\cdots r\end{Bmatrix}\right),$$

这称为 Binet-Cauchy 公式. ☺

第二部分

线性几何

概述

第 12 章　仿射空间

12.1　基本概念

12.2　欧氏几何

12.3　仿射几何

第 13 章　射影空间

13.1　基本概念

13.2　射影几何

13.3　椭圆几何

第 14 章　闵氏空间和辛空间

14.1　闵氏空间和双曲几何

14.2　辛空间和辛几何

小结