数学物理方法基础
答疑讲义

呜哩天才琪露诺
华中科技大学物理学院

日期：2024 年 6 月

目录

1　线性空间

1.1　线性空间的定义

定义 1.1. 设 V 是一个非空集合，K 是一个数域. 若在 V 上规定了加法 $V\times V\to V,(\alpha,\beta)\mapsto\alpha+\beta$ 和数乘 $K\times V\to V,(k,\alpha)\mapsto k\alpha$，且满足：

1. 加法交换律：$\alpha+\beta=\beta+\alpha$；
2. 加法结合律：$(\alpha+\beta)+\gamma=\alpha+(\beta+\gamma)$；
3. 存在零元：存在 $\mathbf{0}\in V$ 满足 $\forall\alpha\in V$，都有 $\alpha+\mathbf{0}=\alpha$；
4. 存在负元：$\forall\alpha\in V$，$\exists\beta\in V$，满足 $\alpha+\beta=\mathbf{0}$；
5. 分配律 ：$k(\alpha+\beta)=k\alpha+k\beta$；
6. 分配律 ：$(k_1+k_2)\alpha=k_1\alpha+k_2\alpha$；
7. 数乘结合律：$(kl)\alpha=k(l\alpha)$；
8. 单位律：$1\alpha=\alpha$

则称 V 是数域 K 上的线性空间.

我们最常见的线性空间就是 $\mathbb{R}^n$，它的元素是排成一列的实数

$$\begin{pmatrix}x_1 & \cdots & x_n\end{pmatrix}^T$$

或排成一行的实数，称为列（行）向量. 两列向量相加，对应实数相加；列向量与一实数 k 相乘，列向量中的每一个实数都乘 k.

但请大家记住，线性空间里头不一定就是我们熟知的列向量. 线性空间里面是什么不重要，重要的是加法和数乘. 只要规定了加法和数乘，猫、狗、核桃夹、指甲剪，都可以是矢量.

1.2　子空间，张成空间

定义 1.2. 设 V 是数域 F 上的线性空间，W 是 V 的一个非空子集合. 若 W 对于 V 中同样的加法和数乘运算也构成 F 上的线性空间，则称 W 是 V 的一个子空间.

而事实上，我们只要验证 W 对 V 上的加法和数乘封闭，即可证明 W 是 V 的子空间. 不能认为随便拎出来 V 的一个子集就是 V 的子空间，比如第一象限全体向量的集合是 $\mathbb{R}^2$ 的子集，但它对 $\mathbb{R}^2$ 上的加法和数乘并不封闭，所以不是 $\mathbb{R}^2$ 的子空间.

向量 $k_i\alpha_i$（这里用了 Einstein 求和约定）称为向量组 $\alpha_1,\cdots,\alpha_n$ 的线性组合. 若 α 是 $\alpha_1,\cdots,\alpha_n$ 的线性组合，则称 α 可由 $\alpha_1,\cdots,\alpha_n$ 线性表出.

向量组 $\alpha_1,\cdots,\alpha_n$ 全体线性组合的集合构成 V 的一个子空间，即这组向量的张成子空间，记作 $L(\alpha_1,\cdots,\alpha_n)$.

1.3　向量组的等价，线性相关与线性无关

如果两组向量张成相同的子空间，我们就说它们是等价的. 等价的向量组不一定有相同数目的向量，比如，向量 $\hat{e}_x$ 和 $\hat{e}_y$ 张成 $\mathbb{R}^2$，我们可以在这组向量中随便加入 $3\hat{e}_x+2\hat{e}_y$ 或者别的什么向量，只要这个向量在 $\mathbb{R}^2$ 中，新的向量组张成的子空间还是 $\mathbb{R}^2$. 但是，等价的向量组，其中一个向量组中的每一个向量必须可以由另一个向量组线性表出.

容易验证，向量组的等价满足自反律、对称律和传递律.

对 $\alpha_1,\cdots,\alpha_n$，如果存在不全为 0 的 $k_1,\cdots,k_n$ 使得 $k_i\alpha_i=\mathbf{0}$，就说 $\alpha_1,\cdots,\alpha_n$ 是线性相关的，否则称它们是线性无关的或线性独立的. 一组向量线性相关的充要条件是其中某个向量可以由其它向量线性表示，这很容易从反证法得到证明. 其逆否命题是，一组向量线性无关的充要条件是其中任一向量都不能由其它向量线性表示. 这也说明了线性无关的含义：如果一组向量是线性无关的，那么其中任何一个向量都是独立于其它向量的，即不能被其它向量线性表出.

设 T 是向量组 S 的一个部分组，如果 T 是线性无关的，而再加入 S 中的任何一个向量得到的向量组 $T\cup\{u\}$ 都是线性相关的，则称 T 为 S 的极大线性无关组. 对 V 中不全为零的有限向量组，其极大线性无关组都是存在的.

可以证明，S 中的任意向量都可以由 T 中的向量线性表示，而 T 的向量显而易见地可以由 S 中的向量线性表示，所以向量组和它的极大线性无关组是等价的.

一个很重要的命题是，若线性无关的向量组 $T=\{\alpha_1,\cdots,\alpha_t\}$ 可由另一组向量 $S=\{\beta_1,\cdots,\beta_s\}$ 线性表示，则 $t\leq s$. 这个命题不难理解：倘若 $t>s$，比如 $S=\{(0,1)\}$，$T=\{(0,1),(1,0)\}$，S 中的向量只能表出平面上一条直线上的向量，而 T 却包含了两个独立的向量，S 怎么可能线性表出 T 呢?

该命题一个明显的推论是，若两个线性无关的向量组等价，则它们所含向量个数相等. 而任一向量组和它的极大线性无关组都是等价的. 由此可知，等价的两向量组，其极大线性无关组所包含的向量个数相同. 特别地，向量组 S 的任一极大线性无关组所包含的向量个数都是相同的，我们把这定义为 S 的秩，记作 $r(S)$.

两个向量组 S 和 T 等价，当且仅当 $r(S)=r(T)=r(S\cup T)$.

1.4 线性空间的基和维数

如果存在有限个线性无关的向量 $\alpha_1,\cdots,\alpha_n\in V$ 使得 $L(\alpha_1,\cdots,\alpha_n)=V$，则称 V 是有限维线性空间，否则称 V 是无限维线性空间. 我们只研究有限维线性空间. 对有限维线性空间，称前面的 $\alpha_1,\cdots,\alpha_n$ 为 V 的一组基，n 称为 V 的维度，记为 $n=\dim V$.

对 n 维线性空间 V，其子空间的维数不能超过 n，V 中的线性无关向量组的向量数目不能超过 n. 且 V 中任意线性无关的向量组都构成 V 的一组基.

对 V 的一组基 $\alpha_1,\cdots,\alpha_n$（以后默认线性空间的基指有序基，即向量的前后顺序是一定的）和任一向量 α，α 可唯一地表示成 $x_i\alpha_i$ 的形式，我们把 $x_1,\cdots,x_n$ 排成列的形式 $\mathbf{x}=\begin{pmatrix}x_1 & \cdots & x_n\end{pmatrix}^T$，称为 α 关于 $\alpha_1,\cdots,\alpha_n$ 的坐标.

给定了一组基，则空间中的向量和它的坐标是一一对应的. 并且，如果 α 和 β 在同一组基下的坐标分别为 $\mathbf{x}$ 和 $\mathbf{y}$，则 $\alpha+\beta$ 及 $k\alpha$ 的坐标分别为 $\mathbf{x}+\mathbf{y}$ 及 $k\mathbf{x}$. 即，向量之间的线性组合关系和坐标之间的线性组合关系完全一致. 因此，有限维线性空间的问题可以归结为 $\mathbb{R}^n$ 中的问题.

我们引入线性空间的同构的概念. 对数域 F 上的线性空间 U 和 V，如果有双射 $f:U\to V$ 使得 $f(k\alpha+l\beta)=kf(\alpha)+lf(\beta)$，则称 U 和 V 是同构的，记作 $U\cong V$，称 f 为同构映射. 容易证明，线性空间的同构满足自反律、对称律和传递律，且同构映射的逆映射仍为同构映射.

数域 F 上的两个有限维线性空间同构当且仅当它们的维数相同，且任意 n 维空间都与 $\mathbb{R}^n$ 同构. 因此，有限维线性空间的唯一特性就是它的维数，一切维数相同的线性空间都可以看成相同的，$\mathbb{R}^n$ 可以看成 n 维线性空间的基本模型.

1.5 子空间的交与和，子空间的直和

直观上很容易认识到，子空间的交集一定还是子空间，而子空间的并集不一定是子空间，比如过原点的平面和不在这个平面上的过原点的直线的并集显然不是 $\mathbb{R}^3$ 的子空间.

设 S 和 T 是 V 的子空间，则它们的和定义为 $S+T=\alpha+\beta \mid \alpha\in S,\beta\in T$，仍是 V 的子空间. 同时，$S+T$ 还是包含 $S\cup T$ 的最小子空间. $S+T$ 和 $S\cap T$ 都是 V 的子空间，如果它们是有限维的，则它们的维数有如下关系：

$$\dim S+\dim T=\dim(S+T)+\dim(S\cap T)$$

如果 $S\cap T\in\{\mathbf{0}\}$，可以证明，此时 $S+T$ 中的每一个向量都可以唯一地表示为 $\alpha+\beta$，其中 $\alpha\in S,\beta\in T$，这时称 $S+T$ 为 S 和 T 的直和，改记 $S\oplus T$. 如果 S 和 T 是有限维的，则 S 的一个基和 T 的一个基合在一起是 $S\oplus T$ 的一个基，且 $\dim(S\oplus T)=\dim S+\dim T$.

若 $S\oplus T=V$，则 V 中的任一向量都可以唯一地表示为 S 和 T 中某两个向量的和，这时称 S 和 T 互为补空间.

1.6 商空间 *

设 $\alpha\in V$，W 是 V 的一个子空间，则我们可以定义 W 的陪集

$$\alpha+W\equiv\{\alpha+\beta \mid \beta\in V\}$$

称 α 为 $\alpha+W$ 的一个陪集代表元，对一个陪集，其代表元不是唯一的. 定义 V 对于 W 的商集为

$$V/W\equiv\{\alpha+W \mid \alpha\in V\}$$

即 V/W 是由 W 的所有陪集组成的. 在 V/W 上规定加法和数乘

$$(\alpha+W)+(\beta+W)\equiv(\alpha+\beta)+W$$
$$k(\alpha+W)\equiv k\alpha+W$$

可以证明 V/W 对这样的运算构成线性空间，称为 V 对于 W 的商空间.

2 矩阵与方程组

2.1 矩阵的秩

因为有限维线性空间总与 $\mathbb{R}^n$ 同构，所以我们重点研究 $\mathbb{R}^n$ 中向量的性质. $\mathbb{R}^n$ 中向量总是一列实数或者一行实数的形式. 我们把 $\mathbb{R}^n$ 中的 m 个列向量排成一排，得到一个数表，称为矩阵

$$\mathbf{A}_{mn}=\begin{pmatrix}\alpha_1 & \cdots & \alpha_m\end{pmatrix}=\begin{pmatrix}a_{11} & \cdots & a_{1m}\\ \vdots & & \vdots\\ a_{n1} & \cdots & a_{nm}\end{pmatrix}$$

它也可以看成 $\mathbb{R}^m$ 中的 n 个行向量排成一列. 我们称矩阵各行张成的空间为其行空间，各列张成的空间为其列空间. 称矩阵的行向量组的秩为其行秩，列向量组的秩为其列秩.

对矩阵，我们可以执行一些操作，比如把它的第 i 行和第 j 行互换，或者把它的第 j 行的 k 倍加到第 i 行上. 对矩阵的以下操作属于行（列）初等变换：

1. 交换矩阵的两行（列）;

2. 用非零的数乘以矩阵的某一行（列）；

3. 用某个数乘矩阵的某一行（列）后加到另一行（列），

通过一系列初等行变换，矩阵总可以化成类似下面的形式：

$$\mathbf{J}=\begin{pmatrix}1 & * & 0 & * & 0 & *\\ 0 & 0 & 1 & * & 0 & *\\ 0 & 0 & 0 & 0 & 1 & *\\ 0 & 0 & 0 & 0 & 0 & 0\end{pmatrix}$$

其特征为：

1. $\mathbf{J}$ 的零行都在下方；
2. $\mathbf{J}$ 的每个非零行的第一个元素（称为主元）都是 1；
3. $\mathbf{J}$ 的每个主元所在列的其余元素都是 0.

这叫做矩阵的简化行阶梯型. 本例中，$\mathbf{J}$ 中有 3 个主元 1. 显然主元所在的各行是线性无关的，且剩余的行（都是零行）可以由这些行线性表出，于是 $\mathbf{J}$ 的行秩为 3. 另一方面，主元所在的各列也是线性无关的，且星号元素所在的列都可以由这些列线性表出（比如第二列是第一列的 * 倍），于是 $\mathbf{J}$ 的各列的极大线性无关组为第一、三和五列，$\mathbf{J}$ 的列秩也为 3. $\mathbf{J}$ 的行秩等于列秩，这对一切简化行阶梯型都是成立的.

初等行变换显然不改变矩阵的行秩. 而我们也可以证明，初等行变换不改变矩阵的列秩，且不改变矩阵各列的线性组合关系. 而我们把矩阵变成简化行阶梯型用到的只有初等行变换. 由此，我们知道，任何矩阵 $\mathbf{A}$，其行秩和列秩都是相同的，我们把它定义为矩阵的秩，记作 $r(\mathbf{A})$. 于是，求 $\mathbb{R}^n$ 中的向量组的秩以及极大线性无关组的手续被我们确定了下来：

1. 把向量写成列向量的形式，然后排成一行，成为一个矩阵 $\mathbf{A}$；
2. 对 $\mathbf{A}$ 施行行初等变换使其变成简化行阶梯型 $\mathbf{J}$；
3. $\mathbf{J}$ 的主 1 个数即为向量组的秩，主 1 所在的各列对应 $\mathbf{A}$ 的列即为向量组的极大线性无关组.

若设矩阵 $\mathbf{A}$ 有 m 行 n 列，则 $r(\mathbf{A})=m$ 或 $r(\mathbf{A})=n$ 时，称其为行满秩或列满秩的，否则称其为行降秩或列降秩的. 列满秩的矩阵，其列向量线性无关.

2.2 矩阵和行列式的运算

主要有：矩阵的加法与乘法，矩阵的转置，方阵的求逆，分块矩阵的计算，行列式的计算. 我们不打算在这里展开这部分内容.

2.3 线性方程组

求 β 如何用向量组 $\alpha_1,\cdots,\alpha_n$ 线性表示，即求 n 个数 $x_1,\cdots x_n$ 使得 $\alpha=x_i\alpha_i$，这就是一个线性方程组. 当 $\alpha_1,\cdots,\alpha_n$ 都是 $\mathbb{R}^m$ 中的列向量时，方程组就是我们熟知的 n 元方程组：

$$\begin{cases}a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1\\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2\\ \cdots\\ a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n=b_m\end{cases}$$

我们先来探讨线性方程组有解的条件. 把 $\alpha_1,\cdots,\alpha_n$ 排成一行，成为系数矩阵 $\mathbf{A}$，则方程组有解等价于 β 可以由 $\mathbf{A}$ 的列向量线性表出，于是 $\{\alpha_1.\cdots,\alpha_n,\beta\}$ 与 $\{\alpha_1.\cdots,\alpha_n\}$ 等价，前面讨论过，这等价于

$r(\alpha_1.\cdots,\alpha_n,\beta)=r(\alpha_1.\cdots,\alpha_n)$. 我们把 $\alpha_1.\cdots,\alpha_n,\beta$ 排成一行，叫做方程组的增广矩阵. 于是，线性方程组有解的充要条件是其系数矩阵 $\mathbf{A}$ 与增广矩阵 $[\mathbf{A},\beta]$ 的秩相同.

$r(\mathbf{A})=n$ 时，矩阵的各列向量线性无关，所以线性表示的系数是唯一的，则方程组有唯一解. $r(\mathbf{A})<n$ 时，矩阵的各列向量线性相关，线性表示的系数是不唯一的，则方程组有无穷多解.

我们把方程组的解 $x_1,x_2,\cdots,x_n$ 排成一列，成为列向量 $\mathbf{x}$，则方程组记作 $\mathbf{Ax}=\beta$ 的形式. 当 $\beta=\mathbf{0}$ 时，称为齐次方程组；当 $\beta\neq\mathbf{0}$ 时，称为非齐次方程组.

对齐次线性方程组，可以证明，它的任意两个解的线性组合仍是它的解. 也就是说，齐次线性方程组的解的集合构成一个线性空间，称为系数矩阵 $\mathbf{A}$ 的零空间，记作 $N(\mathbf{A})$.

于是，我们只要求出 $N(\mathbf{A})$ 的一组基，就可以确定齐次方程组的解的结构. 这样的一组基称为方程组的基础解系. 一般，我们通过行初等变换将 $\mathbf{A}$ 变成简化行阶梯型，通过主 1 的位置得出 $\mathbf{A}$ 的极大线性无关组，其中有 $r(\mathbf{A})$ 个线性无关的向量，将剩下的 $n-r(\mathbf{A})$ 个向量用极大线性无关组线性表出，容易发现，每组表出系数都对应着原方程组的一组解，即求得方程组的基础解系，并证明了：

$$r(\mathbf{A})+\dim(N(\mathbf{A}))=n$$

对非齐次线性方程组 $\mathbf{Ax}=\mathbf{b}$，称 $\mathbf{Ax}=\mathbf{0}$ 为其导出组. 可以证明：

1. $\mathbf{Ax}=\mathbf{b}$ 的任意两个解之差都是其导出组的解；
2. $\mathbf{Ax}=\mathbf{b}$ 的一个解与其导出组的一个解之和，仍为 $\mathbf{b}$ 的一个解；
3. 若 γ 为 $\mathbf{Ax}=\mathbf{b}$ 的一个解，则 $\mathbf{Ax}=\mathbf{b}$ 的所有解都可以写为 $\gamma+\alpha$ 的形式，其中 α 是导出组的解.

可以认为，$\mathbf{Ax}=\mathbf{b}$ 的解集就是其导出组的解空间整体沿 γ 平移得到的.

3 线性映射

3.1 线性映射与线性变换

U 和 V 是 F 上的线性空间，线性映射是指映射 $\sigma:U\to V$，满足：

1. $\forall\alpha,\beta,\sigma(\alpha+\beta)=\sigma(\alpha)+\sigma(\beta)$；
2. $\forall\alpha,k\in F,\sigma(k\alpha)=k\sigma(\alpha)$.

设 $\sigma:V\to U$. 集合 $\{\sigma(v)\mid v\in V\}$ 构成 U 的一个子空间，称作 σ 的像，记作 $\mathrm{Im}(\sigma)$. 集合 $\{v\in V\mid\sigma(v)=\mathbf{0}\}$ 构成 V 的一个子空间，称为 σ 的核，记作 $\mathrm{Ker}(\sigma)$. σ 的像与核的维度分别称为其秩和零度. 我们有重要的秩-零度定理，即

$$r(\sigma)+N(\sigma)=\dim(V)$$

这个定理是很直观的：线性变换的零度度量了 V 的多少个维度被“压缩”为 0，剩下的维数自然就是 U 的维数.

线性映射的加法和数乘的定义是明确的(参考课本),可以验证 F 上线性空间 V 到 U 的线性映射的全体构成 F 上的一个线性空间, 记作 $\hom(V,U)$. 线性映射的乘法即为线性映射的复合, 即 $(\tau\sigma)(\alpha)=\tau(\sigma(\alpha))$,$\forall\alpha\in V$. 一般而言，线性映射满足乘法结合律，而不满足乘法交换律，即线性映射通常是不可对易的.

线性空间 V 到它自己的线性映射称为 V 上的线性变换. 线性变换仍满足秩-零度定理，且 V 上线性变换的全体也构成线性空间，记作 $L(V)$.

可以证明，若 V 上的线性变换 σ 为双射，则它的逆映射 σ^{-1} 也是线性变换. 此时说 σ 是可逆线性变换. 对线性变换而言，单、满、双、可逆是互相等价的，这是一般映射所不具有的性质.

3.2 线性变换的矩阵表示

由于 $\mathbb{R}^n$ 是 n 维线性空间的代表，所以 $\mathbb{R}^n \to \mathbb{R}^m$ 中的线性映射也应是 $V \to U$ 的线性映射的代表（其中 $\dim V = n, \dim U = m$）. 而 $\mathbb{R}^n \to \mathbb{R}^m$ 总能写成 $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$ 的形式，其中 $\mathbf{A}$ 是一个 $n \times m$ 的矩阵.

由于线性映射的线性性质，要研究 σ 怎样作用于 V，只要知道 σ 怎么作用于 V 的每一个基底就可以了. 我们假设已经给出了 V 中的一组基 $\{\alpha_1, \cdots, \alpha_n\}$ 和 U 中的一组基 $\{\beta_1, \cdots, \beta_m\}$，那么，只要确定 $\sigma(\alpha_i)$ 在 $\{\beta_1, \cdots, \beta_m\}$ 下的坐标，就确定了 $\sigma(\alpha_i)$. 有：

$$\begin{bmatrix}\sigma(\alpha_1) & \sigma(\alpha_2) & \cdots & \sigma(\alpha_n)\end{bmatrix} = \begin{bmatrix}\beta_1 & \beta_2 & \cdots & \beta_m\end{bmatrix} \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

因为 $\{\alpha_i\}$ 和 $\{\beta_j\}$ 都不必是列向量，所以等式右边只是形式上的写法，表示 $\sigma_i = \beta_j a_{ji}$，而非矩阵乘法. $\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$ 称为线性映射 σ 关于基 $\{\alpha_i\}$ 和 $\{\beta_j\}$ 的矩阵. 若 γ 在 $\{\alpha_i\}$ 下的坐标为 $\mathbf{x}$，则 $\sigma(\gamma)$ 在 $\{\beta_j\}$ 下的坐标为 $\mathbf{A}\mathbf{x}$.

我们可以证明，给定 V 和 U 各自的一组基后，$V \to U$ 的线性映射与 $\mathbb{R}^{m\times n}$ 中的矩阵一一对应，其中的线性运算也一一对应. 于是 $\hom(V, U) \cong \mathbb{R}^{m\times n}$. 此外，若在给定的基下，$\sigma : V \to U$ 和 $\tau : U \to W$ 的矩阵分别为 $\mathbf{A}$ 和 $\mathbf{B}$，则 $\sigma\tau$ 在相同的基下的矩阵为 $\mathbf{AB}$.

对线性变换，则有 $L(V) \cong \mathbb{R}^{n\times n}$. 若在给定的基下，$\sigma$ 和 τ 的矩阵分别为 $\mathbf{A}$ 和 $\mathbf{B}$，则 $\sigma\tau$ 在相同的基下的矩阵为 $\mathbf{AB}$.

3.3 相似矩阵

然而，线性空间的基并不是唯一的，如果线性空间的两组基 $\{\alpha_i\}$ 和 $\{\beta_i\}$ 满足：

$$\begin{bmatrix}\beta_1 & \beta_2 & \cdots & \beta_n\end{bmatrix} = \begin{bmatrix}\sigma(\alpha_1) & \sigma(\alpha_2) & \cdots & \sigma(\alpha_n)\end{bmatrix} \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix}$$

则称 $\mathbf{C} = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix}$ 为 $\{\alpha_i\}$ 到 $\{\beta_j\}$ 的过渡矩阵. 同时，$\mathbf{C}^{-1}$ 是 $\{\beta_j\}$ 到 $\{\alpha_i\}$ 的过渡矩阵. 若一向量在 $\{\alpha_i\}$ 和 $\{\beta_j\}$ 下的坐标分别为 $\mathbf{X}$ 和 $\mathbf{Y}$，则有 $\mathbf{X} = \mathbf{CY}$.

现在我们考虑同一线性变换在不同基上的矩阵表示的关系. 若 $\{\alpha_i\}$ 到 $\{\beta_j\}$ 的过渡矩阵为 $\mathbf{C}$，σ 在 $\{\alpha_i\}$ 和 $\{\beta_j\}$ 下的矩阵分别为 $\mathbf{A}$ 和 $\mathbf{B}$，则：

$$\mathbf{B} = \mathbf{C}^{-1}\mathbf{AC}$$

我们称同一线性变换在不同基下的矩阵是相似的. 容易验证矩阵的相似满足自反律、对称律和传递律.

相似的矩阵，其迹和行列式相同. 据此可定义线性变换的迹和行列式.

3.4 本征值，本征向量

设 $\sigma \in L(V)$，$\lambda \in F$ 若存在非零向量 $\sigma \in V$ 使得

$$\sigma(\alpha) = \lambda\alpha$$

则称 λ 是 α 的本征值，α 是对应的本征向量. 因为线性变换的线性性可知，同一本征值对应的任意两个本征向量的线性组合仍是这个本征值的本征向量. 于是属于 λ 的全部本征向量与零向量构成 V 的子空间，称为 λ 的本征子空间. 对矩阵，定义是完全类似的. 且线性变换的本征值、本征向量与矩阵的本征值、本征向量是一一对应的.

本征向量表示了空间的这样一些方向：当线性变换作用于空间上时，这些方向上的向量只有长度的伸缩（当然，也会有方向的反转），而无别的变化.

对 $\mathbf{Ax} = \lambda\mathbf{x}$ 变形得 $(\lambda\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$. $\mathbf{x} \neq \mathbf{0}$，故 $|\lambda\mathbf{I} - \mathbf{A}| = 0$，解此方程即得 $\mathbf{A}$ 的本征值，再代入 $(\lambda\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$ 可解出对应的本征向量. 这是求矩阵的本征值及本征向量的一般手续. $\lambda\mathbf{I} - \mathbf{A}$ 称为 $\mathbf{A}$ 的本征多项式.

可以证明，相似的矩阵有相同的本征多项式、本征值、迹、行列式及秩，且有一一对应的本征向量.

相似对角化

我们称一个线性变换 σ 可对角化，如果存在一组基，使得在这组基下的矩阵是对角阵. 称一个矩阵 $\mathbf{A}$ 可对角化，如果存在一个可逆矩阵 $\mathbf{P}$，使得 $\mathbf{P}^{-1}\mathbf{AP}$ 为对角阵.

n 阶矩阵 $\mathbf{A}$ 可对角化的充要条件是它在 $\mathbb{R}^n$ 中有 n 个线性无关的本征向量，且 $\mathbf{P}$ 就是这些本征向量排成一排后形成的矩阵（各本征向量的顺序与对角阵中本征值的顺序一致）.

属于不同本征值的向量线性无关. 于是有 n 个特征向量的 n 阶矩阵必然可对角化.

把 k 个不同本征值的 k 组各自线性无关的本征向量并在一起，其依然线性无关. 于是，矩阵可对角化的充要条件应为 $\sum_{i=1}^{s} \dim E_{\lambda_i} = n$ 或 $\bigoplus_{i=1}^{s} V_{\lambda_i} = \mathbb{R}^n$.

本征子空间的维数小于等于本征值的代数重数. 于是，矩阵可对角化的充要条件是，每个本征子空间的维数都等于其本征值的代数重数.

线性变换的矩阵表示实际上是度量了线性变换对基的作用. 线性变换在某组基下的矩阵为对角矩阵，说明线性变换作用于这组基上，只使它们乘上一个数，而不改变其它东西，于是这组基只能是本征向量. 于是，线性变换可对角化，必然要求它有 n 个线性无关的本征向量. 这也解释了矩阵 $\mathbf{A}$ 到对角矩阵的转换矩阵一定是它的本征向量排成一排：我们假定 $\mathbf{A}$ 是某线性变换 σ 在自然基下的矩阵表示，而对角矩阵 Λ 是 σ 在其本征向量构成的基下的矩阵表示，过渡矩阵自然是本征向量排成一排.

3.5 从线性变换的角度看线性方程组

之前研究线性方程组时，我们是把 $\mathbf{Ax}$ 看成 $\mathbf{A}$ 各列的线性组合. 现在我们认识到，$\sigma : \mathbb{R}^n \to \mathbb{R}^n, \mathbf{x} \mapsto \mathbf{Ax}$ 其实是一个线性映射. 以这种观点来看，求 $\mathbf{Ax} = \mathbf{0}$ 的解，其实就是看 σ 把哪些向量映为 $\mathbf{0}$，即求 σ 的核. $N(\mathbf{A})$ 的维数就是 σ 的零度，而 $\mathbf{A}$ 的秩当然就是 σ 的秩，由秩-零度定理，$\dim N(\mathbf{A}) + r(\mathbf{A}) = n$ 是显然成立的.

若 $r(\mathbf{A}) = n$，则 σ 作用后的 $\mathbb{R}^n$ 仍是 n 维的，没有任何一个维度被压缩，也就只有 $\mathbf{0}$ 被映为 $\mathbf{0}$，方程组只有零解. 而若 $r(\mathbf{A}) < n$，则 σ 作用后的 $\mathbb{R}^n$ 小于 n 维，也就是有些方向上的向量被压缩到原点了，方程组有非零解.

3.6 不变子空间 *

设 σ 是 V 上的线性变换，W 是 V 的子空间. 若任意 $\alpha \in W$ 都有 $\sigma(\alpha) \in W$. 则称 W 是 σ 的不变子空间，简称 σ-子空间. 显然，σ-子空间的交与和都是 σ-子空间. 且 $\mathrm{span}(\alpha_1, \alpha_2, \cdots, \alpha_m)$ 是 σ-子空间的充要条件是 $i = 1, 2, \cdots, m$ 都有 $\sigma(\alpha_i) \in W$. 对于线性变换 σ，$\mathrm{im}(\sigma), \ker(\sigma)$ 和 σ 关于某本征值 λ 的本征子空间 E_λ 都是 σ-子空间.

$\sigma|_W : W \to V, \alpha \mapsto \sigma(\alpha)$ 称为 σ 在 W 上的限制. 设 $\{\alpha_1, \cdots, \alpha_m\}$ 是 W 的一组基，$\sigma|_W$ 在 $\{\alpha_1, \cdots, \alpha_m\}$ 下的矩阵为 A. 将 $\{\alpha_1, \cdots, \alpha_m\}$ 扩充为 V 的一组基 $\{\alpha_1, \cdots, \alpha_m, \beta_{m+1}, \cdots, \beta_n\}$，则在这组基下，$\sigma$ 的矩阵具有以下结构

$$A = \begin{bmatrix} A_1 & A_3 \\ O & A_2 \end{bmatrix}$$

反之亦成立，即若 σ 在某个基下的矩阵为上三角分块矩阵，则与左上对角块对应从一组基向量张成的子空间是 V 的 σ-子空间. 进一步，若 $V = V_1 \oplus V_2$，且 V_1 和 V_2 都是 σ-子空间，则 V_1 和 V_2 的基合起来构成 V 的一组基，这组基下 σ 的矩阵具有对角分块的形式

$$A = \begin{bmatrix} A_1 & O \\ O & A_2 \end{bmatrix}$$

其中 A_1 和 A_2 是 $\sigma|_{V_1}$ 和 $\sigma|_{V_2}$ 在给定基下的矩阵. 这个结论可以推广到 σ 的多个不变子空间的直和的情形.

3.7 矩阵的同时对角化，线性映射的对易

如果 A 和 B 可以用同一个相似变换矩阵 P 相似对角化，即如果 A 和 B 有共同的完备的本征矢量组，则称 A 和 B 可同时对角化. 可以证明，A 和 B 可同时对角化的充要条件是 $AB = BA$.

若已知 $AB = BA$，要将 A 和 B 同时对角化，具体地，我们先将 A 对角化，然后在 A 的每个本征子空间上求出 B 的矩阵形式，进一步在子空间中将 B 对角化，求出该子空间中 B 的本征矢量，这些本征矢量也是 A 的本征矢量，即找到 A 和 B 共同的完备本征向量组. 将这些本征向量按列排成矩阵，即为相似变换的变换矩阵.

对于线性映射 σ 和 ρ，若 $\sigma\rho = \rho\sigma$，则称 σ 和 ρ 可对易. 在某一组基下将它们表示为矩阵形式 A 和 B，再用以上手续，即可找到 σ 和 ρ 共同的完备本征矢. 这一点在量子力学中很重要，如果两个力学量算符可对易，则它们共同的本征矢完备，我们可以取这组本征矢为希尔伯特空间的基底.

3.8 Cayley-Hamilton 定理，根子空间 *

前面我们看到，算子 T 总能表示为对角分块的形式，这需要我们作 V 的直和分解，为做到此，我们先引入算子的多项式理论. V 上算子的多项式是 $f(T) = a_0 I + a_1 T + \cdots + a_n T^n$，仍是 V 上的算子. 且算子的多项式一定可对易，即 $p(T)q(T) = q(T)p(T)$.

可以证明 $\ker P(T)$ 一定是 T 的不变子空间，且若 $p(\lambda)$ 可以因式分解为 $p_1(\lambda)p_2(\lambda)\cdots p_s(\lambda)$，则有直和分解

$$\ker p(T) = \bigoplus_{i=1}^{s} \ker p_i(T)$$

现在我们希望找到 V 的直和分解，可以寻找一类 $p(\lambda)$ 使得 $\ker p(T) = V$，也就要求 $p(T) = O$，也就是要找到多项式 $p(\lambda)$ 使得 $p(T) = O$，这样的多项式称为 T 的零化多项式.

我们有 Cayley-Hamilton 定理,即算子 T 的零化多项式就是它的本征多项式. 设 T 有互异本征值 $\lambda_1, \cdots, \lambda_s$，其重数分别为 $d_1, \cdots, d_s$，则本征多项式可以分解为

$$p(\lambda) = (\lambda - \lambda_1)^{d_1} \cdots (\lambda - \lambda_s)^{d_s} \tag{1}$$

于是，V 可以作直和分解

$$V = \bigoplus_{i=1}^{m} \ker(T - \lambda_i I)^{d_i}$$

其中 $\ker(T - \lambda_i I)^{d_i}$ 称为 λ_i 对应的根子空间.

3.9 Jordan 标准型 *

Jordan 块是指

$$\mathbf{P}_i = \begin{pmatrix} \lambda_i & 1 & 0 & 0 \\ 0 & \lambda_1 & \ddots & 0 \\ 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \lambda_i \end{pmatrix}$$

Jordan 型矩阵是由 Jordan 块组成的矩阵

$$\mathbf{J} = \begin{pmatrix} \mathbf{J_1} & & \\ & \ddots & \\ & & \mathbf{J}_s \end{pmatrix}$$

我们直接了当地给出结论，对任何复向量空间上的算子 T，都能找到一组基，使得 T 在这组基下的矩阵为 Jordan 标准型. 其手续如下:

1. 求出 T 在任意一组基下的矩阵 $\mathbf{A}$；
2. 计算本征多项式 $p(\lambda) = \det(\lambda I - \mathbf{A})$；
3. 求出 $p(\lambda)$ 的互异根 $\lambda_1, \cdots, \lambda_m$；
4. 对每个 λ_i，由公式

$$n_i(t) = \text{rank}(\mathbf{A} - \lambda_i \mathbf{I})^{t-1} + \text{rank}(\mathbf{A} - \lambda_i \mathbf{I})^{t+1} - 2\text{rank}(\mathbf{A} - \lambda_i \mathbf{I})^{t}$$

   算出 λ_i 对应的 t 阶 Jordan 块的个数；
5. 把 Jordan 块随意地排成准对角型，即为所求 Jordan 标准型.

4 线性常微分方程组简介 *

本章讨论齐次常微分线性方程组（后面记为 1）:

$$\mathbf{x}' = \mathbf{A}(t)\mathbf{x}$$

以及非齐次线性常微分方程组（后面记为 2）:

$$\mathbf{x}' = \mathbf{A}(t)\mathbf{x} + \mathbf{f}(t)$$

并默认 $\mathbf{A}(t)$ 是 $n \times n$ 矩阵，$\mathbf{f}(t)$ 是 n 维向量值函数. 它们在 $a \le t \le b$ 有定义且连续可导. 我们由解的存在唯一性定理：对任意 $t_0 \in [a, b]$ 以及 n 维列向量 η，方程组 (2) 存在唯一的解 $\phi(t)$ 满足 $\phi(t_0) = \eta$.

4.1 齐次常微分方程组

同线性方程组一样, 齐次线性常微分方程组也有叠加原理. 即若 $\mathbf{u}(t)$ 和 $\mathbf{v}(t)$ 是 (1) 的解, 则 $\alpha\mathbf{u}(t)+\beta\mathbf{v}(t)$ 也是方程组 (1) 的解. 于是，我们可以用线性代数的方法研究线性常微分方程组解的结构.

若存在不全为零的数 $c_1, c_2, \cdots, c_n$ 使得:

$$c_1\mathbf{x}_1(t) + c_2\mathbf{x}_2(t) + \cdots + c_n\mathbf{x}_n(t) = \mathbf{0}$$

则称向量值函数 $\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x}_n$ 线性相关. 否则，称 $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ 线性无关. 行列式

$$W(t) = \begin{bmatrix} \mathbf{x}_1(t) & \mathbf{x}_2(t) & \cdots & \mathbf{x}_n(t) \end{bmatrix}$$

称为向量值函数 $\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x}_n$ 的 Wronsky 行列式.

我们可以证明:

1. 若向量值函数 $\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x}_n$ 线性相关，则其 Wronsky 行列式 $W(t) \equiv 0$;
2. 若 $\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x}_n$ 是方程组 (1) 的 n 个线性无关的解，则其 Wronsky 行列式 $W(t) \neq 0$;
3. 方程组 (1) 一定有 n 个线性无关的解;
4. 若 $\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x}_n$ 是方程组 (1) 的 n 个线性无关的解，则方程组 (1) 的任一解可由 $\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x}_n$ 线性表出.

由上所述，我们看到，方程组 (1) 的解对加法和数乘构成一向量空间，其维度为 n，方程组 (1) 的任意线性无关的 n 个解构成它的一组基.

若矩阵 $\varPhi$ 的每一列都是方程组 (1) 的解，则称 $\varPhi$ 为方程组 (1) 的解矩阵. 进一步，若 $\varPhi$ 的各列线性无关，则称它是方程组 (1) 的基解矩阵. 若还满足 $\varPhi(t_0) = \mathbf{E}$，则称它为方程组 (1) 的标准基解矩阵.

于是，由定理 4 和定理 5 可得，方程组 (1) 的基解矩阵 $\varPhi(t)$ 一定存在. 且如果 $\varphi(t)$ 是方程组 (1) 的解，则一定存在 n 维列向量 $\mathbf{c}$ 使得 $\varphi(t) = \varPhi(t)\mathbf{c}$

由定理 2 和定理 3 可得，方程组 (1) 的解矩阵 $\varPhi(t)$ 为基解矩阵当且仅当 $\det \varPhi(t) \neq 0$. 而且，只要在 $[a, b]$ 内的某一点 t_0 处 $\det \varPhi(t_0) \neq 0$ 成立，则 $\det \varPhi(t) \neq 0$ 必然成立.

对常微分方程组的解空间，不同的基之间也存在类似的过渡关系.$\varPhi(t)$ 为方程组 (1) 的基解矩阵，$\mathbf{C}$ 为非奇异矩阵，则 $\varPhi(t)\mathbf{C}$ 也是方程组 (1) 的基解矩阵. $\varPhi(t)$ 和 $\varPsi(t)$ 都是方程组 (1) 的基解矩阵，存在非奇异矩阵 $\mathbf{C}$，使得 $\varPsi(t) = \varPhi(t)\mathbf{C}$.

4.2 非齐次方程组解的结构

若 $\varphi(t)$ 是方程组 (1) 的解，$\psi(t)$ 是方程组 (2) 的解，则 $\varphi(t) + \psi(t)$ 是方程组 (2) 的解. 若 $\varphi(t)$ 和 $\psi(t)$ 都是方程组 (2) 的解，则 $\varphi(t) - \psi(t)$ 是方程组 (1) 的解. 方程组 (2) 的解均可以写成

$$\psi(t) = \varPhi(t)\mathbf{c} + \bar{\psi}(t)$$

其中 $\varPhi(t)$ 是方程组 (1) 的基解矩阵，$\mathbf{c}$ 是任一 n 维列向量，$\bar{\psi}(t)$ 是方程组 (2) 的某一解. 这些都与非齐次线性方程组相同.

要求出特解，可以用常数变易法. 即方程组 (2) 的一个满足 $\psi(t_0) = \mathbf{0}$ 的解是:

$$\psi(t) = \varPhi(t) \int_{t_0}^{t} \varPhi^{-1}(s)\mathbf{f}(s), \mathrm{d}s$$

.

4.3 常系数方程组的解法

常系数方程组，是指：

$$\mathbf{x}'(t) = \mathbf{A}\mathbf{x}$$

其中 $\mathbf{A}$ 是 $n \times n$ 常数矩阵.

与分析学一致，我们用幂级数的形式定义矩阵指数：

$$\exp \mathbf{A} = \sum_{k=0}^{\infty} \frac{\mathbf{A}^k}{k!}$$

矩阵指数的运算服从以下定律：

1. 若 $\mathbf{A}, \mathbf{B}$ 为可对易，则

$$\exp(\mathbf{A} + \mathbf{B}) = \exp \mathbf{A} \exp \mathbf{B}$$

2.

$$(\exp \mathbf{A})^{-1} = \exp(-\mathbf{A})$$

3. 若 $\mathbf{T}$ 为非奇异 $n \times n$ 方阵，则

$$\exp(\mathbf{T}^{-1}\mathbf{A}\mathbf{T}) = \mathbf{T}^{-1}(\exp \mathbf{A})\mathbf{T}$$

则 $\varPhi(t) = \exp \mathbf{A}t$ 是方程组 (3) 的基解矩阵，且满足 $\varPhi(0) = \mathbf{E}$.

若 $\mathbf{A}$ 可相似对角化，即 $\mathbf{A}$ 有 n 个线性无关的本征向量：

$$\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$$

其对应的本征值分别为 $\lambda_1, \lambda_2, \cdots, \lambda_n$，则

$$\varPhi(t) = \begin{bmatrix} e^{\lambda_1 t}\mathbf{v}_1 & e^{\lambda_2 t}\mathbf{v}_2 & \cdots & e^{\lambda_n t}\mathbf{v}_n \end{bmatrix}$$

是方程组 (3) 的基解矩阵.

$\varPhi(t)$ 不一定是 $\exp \mathbf{A}t$，但一定有 $\exp \mathbf{A}t = \varPhi(t)\varPhi^{-1}(0)$.

对一般的 $\mathbf{A}$，设其有本征值 $\lambda_1, \lambda_2, \cdots, \lambda_k$，其代数重数分别为 $n_1, n_2, \cdots, n_k$，$\sum_{i=1}^{k} n_i = n$.

设 $\mathbf{V}$ 为 n 维 Euclid 空间，

$$\mathbf{V}^{\lambda_i} = \{\mathbf{v}_i \in \mathbf{V} \mid (\mathbf{A} - \lambda_i)^{r_i}\mathbf{v} = \mathbf{0}\}$$

称为 $\mathbf{A}$ 的属于 λ_i 的根子空间. 则：

$$\mathbf{V} = \bigoplus_{i=1}^{k} \mathbf{V}^{\lambda_i}$$

下面我们求方程组 (3) 的满足 $\varphi(0) = \eta$ 的解.

设 $\eta = \sum_{i=1}^{k} \mathbf{v}_i$，其中 $\mathbf{v}_i \in \mathbf{V}^{\lambda_i}$. 则对任意 $l \geq n_i$ 有 $(\mathbf{A} - \lambda_i\mathbf{E})^l\mathbf{v}_i = \mathbf{0}$. 容易发现 $\mathbf{E} = e^{\lambda_i t}\exp(-\lambda_i\mathbf{E}t)$，则

$$\begin{aligned} (\exp \mathbf{A}t)\mathbf{v}_i &= (\exp \mathbf{A}t)e^{\lambda_i t}[\exp(-\lambda_i\mathbf{E}t)]\mathbf{v}_i \\ &= e^{\lambda_i t}[\exp(\mathbf{A} - \lambda_i\mathbf{E})t]\mathbf{v}_i \end{aligned}$$

于是

$$
\begin{aligned}
\varphi(t) &= (\exp \mathbf{A}t)\eta \\
&= (\exp \mathbf{A}t)\sum_{i=1}^{k}\mathbf{v}_i \\
&= \sum_{i=1}^{k} e^{\lambda_i t}\left[\sum_{j=0}^{n_i-1}\frac{t^j}{j!}(\mathbf{A}-\lambda_i\mathbf{E})^j\right]\mathbf{v}_i
\end{aligned}
$$

要求 $\exp \mathbf{A}t$，只需要分别令 $\eta = \mathbf{e}_1, \mathbf{e}_2, \cdots \mathbf{e}_n$，求得 n 个解，再排成一行即可.

对非齐次方程组，仍可以用常数变易法：对方程组 $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{f}(t)$，它的满足 $\varphi(t_0) = \eta$ 的一个解是

$$\varphi(t) = \exp[(t-t_0)\mathbf{A}]\eta + \int_{t_0}^{t}\exp[(t-s)\mathbf{A}]\mathbf{f}(s)\,\mathrm{d}s$$

5 内积空间

5.1 欧几里得空间，幺正空间，范数

我们已经研究了线性空间和线性空间之间的映射. 但是，我们的空间还是太单调了，我们没法度量向量之间的距离或者夹角. 于是，我们需要赋予线性空间某种结构，故我们引入内积的概念.

设 V 是 $\mathbb{R}$ 上的线性空间，若有映射 $(\cdot,\cdot) : V \times V \to \mathbb{R}$，满足：

1. 对称性：$(\alpha,\beta) = (\beta,\alpha)$；
2. 左线性性：$(k_1\alpha + k_2\beta, \gamma) = k_1(\alpha,\gamma) + k_2(\beta,\gamma)$；
3. 正定性：$(\alpha,\alpha) \geq 0$，当且仅当 $\alpha = \mathbf{0}$ 取等.

则称 (α,β) 为 α 与 β 的内积，称 V 为装备了内积的 Euclid 空间.

我们可以推出：

$$(\alpha, k_1\beta + k_2\gamma) = k_1(\alpha,\beta) + k_2(\alpha,\gamma)$$

也就是说，我们固定左边或者右边的向量，则内积对剩下的那个向量而言是线性映射，这叫做内积的双线性性.

设 V 是 $\mathbb{C}$ 上的线性空间，若有映射 $(\cdot,\cdot) : V \times V \to \mathbb{C}$，满足：

1. 共轭对称性：$(\alpha,\beta) = (\beta,\alpha)^*$；
2. 左线性性：$(k_1\alpha + k_2\beta, \gamma) = k_1(\alpha,\gamma) + k_2(\beta,\gamma)$；
3. 正定性：$(\alpha,\alpha) \geq 0$，当且仅当 $\alpha = \mathbf{0}$ 取等.

则称 (α,β) 为 α 与 β 的内积，称 V 为装备了内积的幺正空间.

我们可以推出：

$$(\alpha, k_1\beta + k_2\gamma) = k_1^*(\alpha,\beta) + k_2^*(\alpha,\gamma)$$

可以看到，复线性空间的内积不再有双线性性.

下面我们用内积空间指代 Euclid 空间或幺正空间.

X 是线性空间，若映射 $\|\cdot\| : X \to \mathbb{R}$ 满足

1. 正定性：任意 $\alpha \in X$ 都有 $\|\alpha\| \geq 0$，且 $\|\alpha\| = 0$ 当且仅当 $\alpha = \mathbf{0}$；
2. 正齐次性：$\|k\alpha\| = |k| \cdot \|\alpha\|$；
3. 三角不等式：$\|\alpha+\beta\| \leq \|\alpha\| + \|\beta\|$.

则称 $\|\cdot\|$ 为 X 上的范数，称 V 为赋范线性空间. 内积空间 V 上的范数可以由内积诱导

$$\|\alpha\| = \sqrt{(\alpha, \alpha)}$$

可以证明这确满足范数的性质.

向量的夹角定义为 $\langle \alpha, \beta \rangle = \arccos \frac{(\alpha,\beta)}{\|\alpha\| \cdot \|\beta\|}$. 若 $(\alpha, \beta) = 0$，则称它们是正交的.

5.2 度规矩阵

内积具有双线性性或厄米双线性性，故要研究任意向量的内积，只需要研究基的内积就好了. 任意给定 Euclid 空间或幺正空间的一组基 $\alpha_1, \cdots, \alpha_n$，设 $\mathbf{u} = a_i \alpha_i$, $\mathbf{v} = b_i \alpha_i$，则：

$$(\mathbf{u}, \mathbf{v}) = (a_i \alpha_i, b_i \alpha_i) = a_i b_j (\alpha_i, \alpha_j)$$

记 $\mathbf{G} = (g_{ij})_{n \times n}$，其中 $g_{ij} = (\alpha_i, \alpha_j)$，称为内积在基 $\{\alpha_i\}$ 下的度规矩阵或 Gram 矩阵，则：

$$(\mathbf{u}, \mathbf{v}) = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} \mathbf{G} \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

可见，取定 V 上的一组基，则 V 上的内积结构与 n 阶实对称正定矩阵一一对应.

若 $\{\alpha_i\}$ 到 $\{\beta_j\}$ 的过渡矩阵为 $\mathbf{P}$，$\mathbf{A}$ 和 $\mathbf{B}$ 分别是内积在基 $\{\alpha_i\}$ 和 $\{\beta_j\}$ 下的度规矩阵. 则对欧几里得空间有 $\mathbf{B} = \mathbf{P}^T \mathbf{G} \mathbf{P}$，对幺正空间有 $\mathbf{B} = \mathbf{P}^\dagger \mathbf{G} \mathbf{P}$. 如果存在可逆矩阵 $\mathbf{P}$ 使得 $\mathbf{B} = \mathbf{P}^T \mathbf{G} \mathbf{P}$，我们就说矩阵 $\mathbf{A}$ 和 $\mathbf{B}$ 是合同的. 容易验证矩阵的合同满足自反律、对称律和传递律.

5.3 标准正交基

我们常用 $\mathbb{R}^3$ 中的 $\hat{i}, \hat{j}, \hat{k}$ 当作基，它们是两两正交的，而且都是单位向量. 而只有定义了内积之后，我们才有可能定义“正交”和“单位向量”的概念.

Euclid 空间或幺正空间中的向量组，若其所有向量都是单位向量，且两两正交，则称它是标准正交向量组. 容易验证正交向量组的向量必然线性无关. 于是可以在 V 中取相互正交的向量构成的基，且这些向量都是单位向量，称该基为 V 的标准正交基. 标准正交基的条件可以写成：

$$(\varepsilon_i, \varepsilon_j) = \delta_{ij}$$

标准正交基有很多好的性质：

1. 度规矩阵 $\mathbf{G} = \mathbf{I}$，由此，若两向量 α 和 β 在标准正交基下的坐标分别为 $\mathbf{X}$ 和 $\mathbf{Y}$，则它们的内积为 $(\alpha, \beta) = \mathbf{X}^T \mathbf{Y} = x_i y_i$；
2. 若 $\{\varepsilon_i\}$ 是 Euclid 空间 V 的一组标准正交基，则对任意 α，$\alpha = (\varepsilon_i, \alpha) \varepsilon_i$，这叫做 α 的 Fourier 展开；
3. 线性变换 σ 在标准正交基 $\{\varepsilon_i\}$ 下的矩阵 $\mathbf{A}$ 满足 $(\mathbf{A})_{ij} = (\varepsilon_i, \sigma(\varepsilon_j))$.

给定 Euclid 空间或幺正空间 V 的任意一组基 $\{\alpha_i\}$，总可以通过下面的施密特正交化方法得到一组正交基 $\{\beta_i\}$.

$$\begin{cases} \beta_1 = \alpha_1 \\ \beta_2 = \alpha_2 - \frac{(\alpha_2, \beta_1)}{(\beta_1, \beta_1)} \beta_1 \\ \cdots \\ \beta_n = \alpha_n - \sum_{i=1}^{n-1} \frac{(\alpha_n, \beta_i)}{(\beta_i, \beta_i)} \beta_i \end{cases}$$

再令 $\gamma_i = \frac{\beta_i}{\|\beta_i\|}, i = 1, 2 \cdots, n$，就得到一组标准正交基.

对有限维幺正空间内的线性映射 φ，存在一组基使得 φ 为上三角矩阵. 这叫做 Schur 定理.

5.4 对偶基自然地生成内积（Riesz 表示定理）*

内积具有双线性性或厄米双线性性，于是，要求 (α,β)，我们实际上只需要求 (e_i,β)，这里 $\{e_i\}$ 是 V 的一组基. 如果我们知道 V 在 V^* 的对偶基为 $\{e^i\}$，而 e^i 又是 V 到 $\mathbb{R}$ 的映射，那么，我们可以很自然地规定 $(e_i,\beta)=e^i(\beta)$. 于是，通过对偶基，很自然地确定了 V 上的内积. 回忆一下，我们构造对偶基时，要求 $e^i(e_j)=\delta_{ij}$，于是，在这样产生的内积下，$\{e_i\}$ 正好是标准正交基！

我们当然可以按照别的什么方法确定内积，但这样弄出来的内积下，$\{e_i\}$ 不是标准正交基，我们何必搬起石头砸自己的脚. 对 V 的一切可能的基，我们都能规定一种内积结构，使得 $\{e_i\}$ 是标准正交基. 之所以有这种任意性，是因为 V 和 V^* 不存在自然，V^* 中基的选择还是依赖 V 的基.

5.5 内积空间的同构

内积空间之间若存在一个双射 σ，它保持空间的线性结构和内积结构，即：

1. $\sigma(\alpha+\beta)=\sigma(\alpha)+\sigma(\beta)$；
2. $\sigma(k\alpha)=k\sigma(\alpha)$；
3. $(\sigma(\alpha),\sigma(\beta))=(\alpha,\beta)$.

则称它们同构，σ 为同构映射. 同构满足自反律、对称律和传递律. 可以证明，内积空间同构的充要条件是它们的维数相同.

5.6 正交补空间与正交投影

设 V_1,V_2 是内积空间 V 的两个子空间，若 $\forall\alpha\in V_1,\forall\beta\in V_2$ 都有 $(\alpha,\beta)=0$，则称 V_1,V_2 是正交的，记作 $V_1\perp V_2$.

设 $V=V_1+V_2$，且 $V_1\perp V_2$，则可证明 $V=V_1\oplus V_2$，我们称 V_1 是 V_2 的正交补空间.

线性空间的子空间的补空间不是唯一的，而我们可以证明，内积空间的子空间的正交补空间是唯一的. 这个从直观上很好理解，$\mathbb{R}^3$ 中，平面外的直线可以有无数条，它们都是这个平面的补空间（当然，平面和直线都是过原点的）. 然而，垂直于这个平面的直线当然只有一个.

下面我们再来审视线性方程组 $\mathbf{A}\mathbf{x}=\mathbf{0}$，以

$$\begin{cases}a_{11}x+a_{12}y+a_{13}z=0\\a_{21}x+a_{22}y+a_{23}z=0\\a_{31}x+a_{32}y+a_{33}z=0\end{cases}$$

为例，我们发现，三个方程组实际上说明了 $\begin{pmatrix}a_{i1}&a_{i2}&a_{i3}\end{pmatrix}^T$ $(i=1,2,3)$ 与 $\begin{pmatrix}x&y&z\end{pmatrix}^T$ 是正交的，即矩阵的各行向量与方程组的解是正交的. 于是，矩阵 $\mathbf{A}$ 的行空间和方程组的解空间是正交的，而由秩-零度定理又知道它们的维数之和等于矩阵的列数. 故，$\mathbf{A}$ 的行空间在 $\mathbb{R}^n$ 的正交补空间是 $\mathbf{A}\mathbf{x}=\mathbf{0}$ 的解空间.

设 U 是 V 的子空间，定义 V 到 U 的正交投影 $P_U\in\hom(V,U)$. 使得任意 $v=u+w\in V$，其中 $u\in U,w\in U^\perp$，有 $P_U(v)=u$. 正交投影的性质开列如下

1. $P_U\in L(V)$;
2. 对 $u\in U$ 有 $P_U(u)=u$，对 $v\in U^\perp$ 有 $P_U(v)=\mathbf{0}$;
3. $\mathrm{im}(P_U)=U,\ker(P_U)=U^\perp$;

4. $V = \mathrm{im}(P_U) \oplus \ker(P_U)$;
5. $P_U^2 = P_U$;
6. $\|P_U(u)\| \leq \|u\|$.

5.7 最小二乘法与线性回归

给定一个线性方程组 $\mathbf{A}\mathbf{x} = \mathbf{b}$，它未必有解，然而，我们可以考虑这样一个问题：如何找到这样一个 $\mathbf{x}$，使得 $\mathbf{A}\mathbf{x}$ 和 $\mathbf{b}$ 的距离尽可能地小，也就是使 $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$ 尽可能地小. 这称为最小二乘问题. 下面我们从 Eculid 空间的角度考虑这个问题.

对任意的 $\mathbf{x}$，$\mathbf{A}\mathbf{x}$ 是 $\mathbf{A}$ 的列空间中的元素. 所以我们需要在 $\mathbf{A}$ 的列空间中找到一个元素 $\mathbf{A}\mathbf{y}$，与 $\mathbf{b}$ 的距离最小，从几何上这个问题很好解决，只需要把 $\mathbf{b}$ 投影到 $\mathbf{A}$ 的列空间上就行. 于是，$\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}$ 必须与 $\mathbf{A}$ 的列空间正交，对任意的 $\mathbf{y}$ 都要有 $(\mathbf{A}\mathbf{y})^T(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) = \mathbf{0}$，取 $\mathbf{y} = \mathbf{A}^T(\mathbf{b} - \mathbf{A}\hat{\mathbf{X}})$，即得正规方程

$$\mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^T\mathbf{b}$$

投影是唯一存在的，故正规方程的解必然存在且唯一.

变量 x 和 y 之间的关系往往是线性关系 $y = kx + b$，假设现在我们通过实验测得了 x 和 y 的数据列 $\{x_1, x_2, \cdots, x_n\}$ 和 $\{y_1, y_2, \cdots, y_n\}$，要求出 k 和 b. 则 k 和 b 满足方程

$$\mathbf{X}\beta = \mathbf{y}$$

其中 $\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \beta = \begin{pmatrix} k \\ b \end{pmatrix}, \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$. 这个方程未必有解，我们需要找到 β 的最小二乘解，对应的正规方程为

$$\mathbf{X}^T\mathbf{X}\hat{\beta} = \mathbf{X}^T\mathbf{b}$$

5.8 正交矩阵和正交变换

满足 $\mathbf{A}^T\mathbf{A} = \mathbf{I}$ 的 n 阶实矩阵 $\mathbf{A}$ 叫做正交矩阵. 它的性质开列如下：

1. $|\mathbf{A}| = \pm 1$；
2. $\mathbf{A}$ 可逆，$\mathbf{A}^{-1} = \mathbf{A}^T$；
3. $\mathbf{A}$ 为正交矩阵，则 $\mathbf{A}^{-1}$ 或 $\mathbf{A}^T$ 也是正交矩阵；
4. 正交矩阵的乘积仍是正交矩阵；
5. 正交矩阵的行（列）向量为 $\mathbb{R}^n$ 的一组标准正交基；
6. $\mathbb{R}^n$ 的标准正交基之间的过渡矩阵为正交矩阵；
7. 正交矩阵把 $\mathbb{R}^n$ 的一组标准正交基变为另一组标准正交基.

所谓正交变换，是指满足 $\|\sigma(\alpha)\| = \|\alpha\|$ 的线性变换. 即，正交变换保持向量的模长不变.

σ 是正交变换与以下陈述等价：

1. $(\alpha, \beta) = (\sigma(\alpha), \sigma(\beta))$；
2. σ 在 V 的标准正交基下的矩阵为正交矩阵；
3. σ 把 V 的一组标准正交基变为另一组标准正交基.

可见正交变换不仅保向量的模长不变，而且保向量的内积不变. 正交变换是保内积结构的变换.

n 阶实正交矩阵的集合对矩阵的乘法构成一个群，称 n 阶实正交矩阵群；n 维 Euclid 空间上正交变换的集合对线性变换的乘法构成一个群，称 n 维实正交变换群. 两个群都可记作 $O(n)$.

幺模实正交矩阵群以及对应的正交变换群记作 $SO(n)$，这对应 n 维空间中的转动. 如果我们把关于原点对称的变换（宇称变换）记作 $\mathbf{P}$，则 $O(n) = SO(n) \sqcup \mathbf{P}SO(n)$，即任意正交变换要么是旋转变换，要么是旋转变换再加上一个宇称变换. 从这个角度很容易理解正交变换的性质. 旋转变换或宇称变换不改变任何两个矢量的夹角，也不改变矢量的模长，故保持任何两个矢量的内积. 同理，正交变换保持标准正交基的模长为 1，也保持它们两两之间仍然正交，故正交变换把一组标准正交基变为另一组.

5.9 幺正变换，幺正矩阵

幺正矩阵是满足 $\mathbf{A}^\dagger\mathbf{A} = \mathbf{I}$ 的矩阵. 它有如下性质：

1. $|\mathbf{A}| = e^{i\theta}$；
2. $\mathbf{A}$ 可逆，且 $\mathbf{A}^{-1} = \mathbf{A}^\dagger$；
3. $\mathbf{A}^{-1}$ 或 $\mathbf{A}^\dagger$ 仍为幺正矩阵；
4. 两幺正矩阵的乘积仍为幺正矩阵；
5. 幺正矩阵的 n 个行（列）构成 $\mathbb{C}^n$ 的标准正交基；
6. 幺正矩阵将 V 的一组标准正交基变为另一组标准正交基；
7. V 的标准正交基之间的过渡矩阵一定是幺正矩阵.

满足 $\|\sigma(\alpha)\| = \|\alpha\|$ 的线性变换 σ 为幺正变换，σ 为幺正变换等价于：

1. $(\sigma(\alpha), \sigma(\beta)) = (\alpha, \beta)$；
2. σ 在标准正交基下的矩阵为幺正矩阵；
3. σ 将一组标准正交基变为另一组标准正交基；

n 阶幺正矩阵的集合对矩阵的乘法构成一个群，称 n 阶幺正矩阵群；n 维幺正空间上正交变换的集合对线性变换的乘法构成一个群，称 n 维幺正变换群. 两个群都可记作 $U(n)$.

5.10 伴随算子 *

线性映射 $T \in \hom(W, V)$ 的伴随为 $T^* : W \to V$. 对于任意 $v \in V, w \in W$，有 $\langle T(v), w\rangle = \langle v, T^*(w)\rangle$. 伴随也是线性映射. 伴随的性质开列如下：

1. 对任意的线性映射 S 和 T 有 $(S+T)^* = S^* + T^*$；
2. 对任意的线性映射 T 和数 λ 有 $(\lambda T)^* = \lambda^* T^*$；
3. 对任意的线性映射 T 有 $(T^*)^* = T$；
4. 对任意的 $S \in \hom(U, V)$ 和 $T \in \hom(V, W)$ 有 $(S \circ T)* = T^* \circ S^*$；
5. $I^* = I$.

设线性映射 T，它的伴随为 T^*，则 $\ker(T)$ 与 $\mathrm{im}(T^*)$ 互为正交补，$\ker(T^*)$ 与 $\mathrm{im}(T)$ 互为正交补. 设 W 的标准正交基 $\{\alpha_i\}$，V 的标准正交基为 $\{\beta_i\}$，线性映射 $T : W \to V$ 在这组基底下的矩阵为 $\mathbf{A}$，则 T^* 在这组基底下的矩阵为 $\mathbf{A}^* \equiv \overline{\mathbf{A}^T}$.

5.11 自伴算子

$T \in L(V)$ 称为自伴算子，若 $T = T^*$，或对任意的 $u, v \in V$ 都有 $\langle T(u), v\rangle = \langle u, T(v)\rangle$.

本讲义中将 Eculid 空间上的自伴算子称为对称变换，幺正空间上的自伴算子称为 Hermite 变换. Hermite 变换有相当良好的性质，σ 是 Hermite 变换与以下陈述等价：

1. $(\alpha, \sigma(\alpha)) \in \mathbb{R}$；
2. σ 的本征值为实数，其本征向量组完备且可选为相互正交的单位向量.

V 上的 Hermit 变换 σ 在 V 的任一标准正交基下的矩阵为 Hermite 矩阵，进而，Hermite 矩阵有如下性质：

1. Hermite 矩阵的本征值全为实数；
2. Hermite 矩阵属于不同本征值的本征向量互相正交；
3. Hermite 矩阵可以通过幺正矩阵相似对角化.

5.12 实谱定理，复谱定理

下面我们来讨论内积空间上算子的本征矢量构成一组标准正交基的条件，即算子可正交对角化的条件.

在 Eculid 空间上，我们有实谱定理：σ 的本征值为实数，其本征向量组完备且可选为相互正交的单位向量，当且仅当 σ 是对称变换.

对称变换在标准正交基下的矩阵为是实称矩阵，实对称矩阵有如下性质：

1. 实对称矩阵的本征值全为实数；
2. 实对称矩阵属于不同本征值的本征向量互相正交；
3. 实对称矩阵可以通过实正交矩阵相似对角化.

用正交矩阵将矩阵 $\mathbf{A}$ 对角化的手续如下：

1. 解 $|\lambda \mathbf{I} - \mathbf{A}| = 0$ 得 $\mathbf{A}$ 的本征值 $\lambda_1, \cdots, \lambda_s$；
2. 对每一个 λ_i，算出对应的本征向量，用施密特标准正交化方法得到一组相互正交的单位本征向量；
3. 将这些相互正交的单位本征向量按顺序排成一列，得到正交矩阵 $\mathbf{P}$，则 $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{\Lambda}$.

对幺正空间，可正交对角化的条件相对不那么严格. 我们引入正规算子的概念.$T \in L(V)$ 称为正规算子，若 $T^*T = TT^*$，显然自伴算子是正规算子，而正规算子未必是自伴算子.T 是正规算子的等价表述是 $\|T(v)\| = \|T^*(v)\|$ 对一切 $v \in V$ 成立.

于是复谱定理告诉我们：在幺正空间上，σ 的本征值为实数，其本征向量组完备且可选为相互正交的单位向量，当且仅当 σ 是正规变换.

5.13 厄米矩阵与幺正矩阵的关系 *

若 F 为厄米矩阵，而 $U = e^{\mathrm{i}F}$，则 U 为幺正矩阵. 反之，若 U 为幺正矩阵，则必然存在厄米矩阵 F 使得 $U = e^{\mathrm{i}F}$. 这一点在量子力学中很重要，体现为某力学量算符是对应无穷小变换的生成元，这个读者会在后续学习中了解.

6 二次型

6.1 二次型的定义和标准型

n 元二次型是数域 F 上 n 个变量 $x_1, \cdots, x_n$ 的仅含二次项的多项式，即 $f(x_1, \cdots, x_n) = a_{ij}x_ix_j\,(a_{ij} = a_{ji})$（这里还是用了 Einstein 求和约定），$\mathbf{A} = \{a_{ij}\}_{n\times n}$ 称为二次型的矩阵. 我们只讨论实二次型，即 $F = \mathbb{R}$ 的情况.

$\mathbb{R}$ 上两组变量 $x_1, \cdots, x_n$ 与 $y_1, \cdots, y_n$ 之间的关系

$$x_i = c_{ij} y_j$$

称为由 $x_1, \cdots, x_n$ 到 $y_1, \cdots, y_n$ 的线性替换. 引入 $\mathbf{X} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix}^T$，$\mathbf{Y} = \begin{pmatrix} y_1 & \cdots & y_n \end{pmatrix}$ 及 $\mathbf{C} = (c_{ij})_{n\times n}$，则线性替换可写成矩阵形式 $\mathbf{X} = \mathbf{CY}$. 若 $\mathbf{C}$ 是可逆的，则称替换是非退化的.

一个仅含平方项的二次型 $f(x_1, \cdots, x_n) = a_i x_i^2$ 称为标准型. 任何 n 元二次型都可以通过非退化的线性替换变为标准型. 从矩阵的角度说，任一实对称矩阵都可以合同于对角矩阵.

化二次型为标准型一般有三种方法：

1. 行列对称初等变换法：对 $\begin{pmatrix} \mathbf{A} \\ \mathbf{I} \end{pmatrix}$ 每做一次行变换的同时做一次相同的列变换，直至变为 $\begin{pmatrix} \mathbf{\Lambda} \\ \mathbf{C} \end{pmatrix}$；
2. 配方法，这个我们在中学就很熟悉了；
3. 正交变换法：因为实对称矩阵一定可以通过正交矩阵相似对角化，所以，通过正交相似对角化的手续，也可以将二次型化为标准型. 这种方法通常也叫做主轴变换.

6.2 规范型, 惯性定理

实二次型的规范型是指仅含平方项且平方项系数只为 ±1 的二次型 $f(x_1, \cdots, x_n) = x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_r^2$（其中 r 为二次型的秩），任何实二次型都可以通过线性替换变为规范型，且规范型唯一，这是惯性定理.

实二次型的规范型中正项个数 p 称该二次型的正惯性指数，负项个数 $r - p$ 称该二次型的负惯性指数，$s = p - (r - p) = 2p - r$ 称二次型的号差.

惯性定理用矩阵来描述，即任一实对称矩阵都唯一合同于如下的规范型矩阵：

$$\begin{pmatrix} \mathbf{I}_p & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & -\mathbf{I}_{r-p} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{O} \end{pmatrix}$$

6.3 二次型的正定性

实二次型 $f(x_1, \cdots, x_n)$ 称为正定的，如果对不全为零的 $x_1, \cdots, x_n$ 恒有 $f(x_1, \cdots, x_n) > 0$. 实对称矩阵 $\mathbf{A}$ 称为正定的，如果其对应的二次型 $\mathbf{X}^T\mathbf{AX}$ 是正定的.

n 元实二次型正定的充要条件是其正惯性指数 $p = n$，这也是实对称矩阵正定的充要条件. 而二次型 $\mathbf{X}^T\mathbf{AX}$ 通过正交变换法得到的规范型的二次项系数实际上就是 $\mathbf{A}$ 的本征值，故 $\mathbf{A}$ 正定的充要条件也可以叙述为 $\mathbf{A}$ 的本征值全为正. $\mathbf{A}$ 一定合同于规范型矩阵，而正定实对称矩阵 $\mathbf{A}$ 的规范型矩阵只能是 $\mathbf{I_n}$，这又是一个充要条件. 此外，$\mathbf{A}$ 正定还有一个充要条件是它的顺序主子式都为正，这个在实际的判定中是很常用的.

实二次型 $f(x_1, \cdots, x_n) = \mathbf{X}^T\mathbf{AX}$，如果对不全为零的 $x_1, \cdots, x_n$ 恒有：

1. $f(x_1, \cdots, x_n) < 0$，则称其为负定二次型，$\mathbf{A}$ 为负定矩阵；
2. $f(x_1, \cdots, x_n) \geq 0$，则称其为准正定二次型，$\mathbf{A}$ 为准正定矩阵；
3. $f(x_1, \cdots, x_n) \leq 0$，则称其为准负定二次型，$\mathbf{A}$ 为准负定矩阵；
4. 若 $f(x_1, \cdots, x_n)$ 既不是准正定的，又不是准负定的，则称其为不定二次型.

6.4 矩阵的奇异值分解，主成分分析法 *

假设矩阵 $\mathbf{A}$ 按左乘作用在实矢量空间 $\mathbb{R}^n$ 上，$\mathbf{x}$ 变为 $\mathbf{Ax}$，则 $\frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|}$ 即为 $\frac{\mathbf{x}^T\mathbf{Ax}}{\mathbf{x}^T\mathbf{x}}$，左乘作用的效果可以完全由 $\mathbf{A}^T\mathbf{A}$ 决定.$\mathbf{A}^T\mathbf{A}$ 是实对称矩阵故可正交对角化，设其本征矢 $\{\mathbf{v}_i\}$ 构成 $\mathbb{R}^n$ 的一组标准正交基，对应的本征值为 $\{\lambda_i\}$ 且为降序排列，易见所有的本征值都非负.$\{\lambda_i\}$ 的平方根 $\{\sigma_i\}$ 称为 $\mathbf{A}$ 的奇异值.

设 $\mathbf{A}$ 有 r 个非零特征值，则可以证明 $\{\mathbf{Av}_i\}, i=1,\cdots,r$ 构成 $\mathbf{A}$ 的列空间的一组正交基. 对秩为 r 的矩阵 $\mathbf{A}$ 的 $m\times n$ 矩阵，存在 $m\times n$ 矩阵

$$\mathbf{\Sigma} = \begin{pmatrix} D & O \\ O & O \end{pmatrix}$$

其中 D 是 $\mathbf{A}$ 的非零奇异值组成的对角矩阵. 且存在正交矩阵 $\mathbf{U}$ 和 $\mathbf{V}$ 使得 $\mathbf{A}=\mathbf{U\Sigma V}$.

SVD 分解的几何意义很明显：任何一个 U 到 V 线性映射都可以分解为：

1. U 内的正交变换；
2. 从 U 到 V 的压缩变换；
3. V 内的正交变换.

SVD 的一般操作手续我们不再赘述. 下面我们看 SVD 分解的一个重要的应用：主成分分析法（PCA）. 假设我们有 $\{x_i\}, i=1,\cdots,p$ 共 p 个指标，但我们觉得用 p 个指标描述一个系统太过繁琐，于是，我们试图用一个加权平均数

$$s=\sum_{i=1}^{p} c_i x_i$$

来反映系统的性质，这样只需要一个指标. 其中我们要求 $c_1^2+\cdots+c_p^2=1$. 为了增加区分度，我们自然希望不同系统的 s 尽可能分散，在统计学意义上，也就是希望 s 的方差尽可能大. 有时一个指标的区分度会过小，也就是过多信息丢失了，我们就可以用几个互相独立的（这也就要求系数矢量 $\mathbf{c}_j$ 互相正交）指标来描述系统. 总之，我们是想要找到一个系数矩阵，来降低数据集的维度.

假设 $\begin{pmatrix}\mathbf{X}_1 & \mathbf{X}_2 & \cdots & \mathbf{X}_N\end{pmatrix}$ 是 N 个系统的观测矩阵，其中每一列包含 p 个指标. 观测矢量的样本均值 $\mathbf{M}=\frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_i$，并记 $\hat{\mathbf{X}}_k=\mathbf{X}_k-\mathbf{M}.\mathbf{B}=\begin{pmatrix}\hat{\mathbf{X}}_1 & \hat{\mathbf{X}}_2 & \cdots\hat{\mathbf{X}}_N\end{pmatrix}$ 称平均偏差形式. 样本的协方差矩阵定义为 $\mathbf{S}=\frac{\mathbf{BB}^T}{N-1}$，是半正定的. 容易验证，$\mathbf{S}$ 的对角元 S_{ii} 正是第 i 个指标 x_i 的方差，总方差 $\sum_{i=1}^{p}\mathrm{Var}(x_i)=\mathrm{tr}(\mathbf{S})$.

现在，我们对观测矩阵的每一列取变换 $\mathbf{X}=\mathbf{PY}$，则 $\{\mathbf{Y}_i\}$ 的协方差矩阵为 $\mathbf{P}^T\mathbf{SP}$，其迹不变. 我们当然将观测矩阵正交对角化为 $\mathbf{D}$，则样本的总方差为：

$$\sum_{i=1}^{p}\mathrm{Var}(x_i)=\mathrm{tr}(\mathbf{S})=\mathrm{tr}(\mathbf{D})=\sum_{i=1}^{p}\lambda_i$$

这里 λ_i 是 $\mathbf{S}$ 的本征值，假设其为降序排列. 则 $\mathbf{S}$ 的不同本征值对样本的方差有着不同的贡献，依次称为第一主成分、第二主成分... 于是，我们可以取前几个本征值对应的本征矢组成系数矩阵，来实现数据的降维.

7 多重线性代数 *

7.1 对偶空间

数域 F 其实就是自身上的一个线性空间，故 $f: V\to F$ 可以看成 V 上的一个特殊的线性映射，称为线性泛函. V 上所有线性泛函组成的集合构成 F 上的一个线性空间，记作 $\hom(V,F)$.

如果 V 是 n 维线性空间，则 $\dim\hom(V,F)=(\dim V)(\dim F)=n$，继而 $\hom(V,F)\cong V$. 记 $\hom(V,F)$ 为 V^*，称为 V 的对偶空间.

现在我们来求 V^* 的一组基. 因为 $\dim V^* = n$，故我们只要找到 V 上的 n 个线性泛函，且它们线性无关，就是 V^* 的一组基. 而确定 V 上的线性函数 f，只要知道 f 在 V 的基 $\{\alpha_i\}$ 上的作用就行了. 我们构造 V 上的 n 个线性泛函如下：

$$f_i(\alpha_j) = \delta_{ij}$$

可以验证它们线性无关，从而是 V^* 的一组基，称为 V 的 $\{\alpha_i\}$ 的对偶基.

对 V 中任意向量 α 和 V^* 中任一向量 f，分别有 $\alpha = f_i(\alpha)\alpha_i$ 和 $f = f(\alpha_j)f_j$.

设 V 的基 $\{\alpha_i\}$ 和 $\{\beta_j\}$ 在 V^* 中对应的对偶基分别为 $\{f_i\}$ 和 $\{g_j\}$. 从 $\{\alpha_i\}$ 到 $\{\beta_j\}$ 的过渡矩阵为 $\mathbf{A}$，则 $\{f_i\}$ 到 $\{g_j\}$ 的过渡矩阵为：

$$\mathbf{B} = (\mathbf{A}^{-1})^T$$

对 V 的对偶空间 V^*，还可以考虑其对偶空间 V^{**}，称为 V 的双重对偶空间，有 $V \cong V^{**}$.

在 V 中取基 $\{\alpha_i\}$，V^* 中的对偶基为 $\{f_j\}$，V^{**} 中的对偶基为 $\{\alpha_k^{**}\}$. 取 V 中的向量 $\alpha = x_i\alpha_i$，则 V^* 和 V^{**} 对应的向量分别为 $x_i f_i$ 和 $x_i\alpha_i^{**}$. 对任意的 f：

$$\begin{aligned}
\alpha^{**}(f) &= (x_i\alpha_i^{**})(f) \\
&= x_i\alpha_i^{**}(f) \\
&= x_i\alpha_i^{**}(f(\alpha_j)f_j) \\
&= x_i[f(\alpha_j)\alpha_i^{**}(f_j)] \\
&= x_i f(\alpha_i) \\
&= f(x_i\alpha_i) \\
&= f(\alpha)
\end{aligned}$$

于是，对任意的 α，我们不需要考虑 V 中的基即可以确定 α^{**} 的形式（α^{**} 把 $\hom(V, F)$ 中的每一个函数映到 $f(\alpha)$）. 我们称这种不依赖于基的选择的同构映射为自然同构. 所谓“自然”，就是可以通过人为约定让任何人都可以通过此约定独立地找到相同的数学对象.

7.2 Einstein 约定，协变与逆变

Einstein 约定是在张量分析和微分几何中常采用的记号系统：

1. 矢量的基的指标在下，坐标的指标在上;
2. 对偶矢量的基的指标在上，坐标的指标在下;
3. 矩阵的行指标在下，列指标在上;
4. Kronecker 符号写为 δ^i_j;
5. 求和后，未消去的指标保持上下位置不变;

按这套约定，矢量的傅里叶展开为 $v = x^i v_i$，对偶矢量的傅里叶展开为 $\varphi = x_i\varphi^i$. 矩阵乘法写作

$$C = AB \Leftrightarrow c^i_j = a^i_k b^k_j$$

且有 $\varphi(v_j) = \delta^i_j$.

从本章第一节我们知道，矢量和对偶矢量的基底变换规则分别为 $\tilde{v}_j = T^i_j v_i$ 和 $\varphi^i = T^i_j\tilde{\varphi}^j$ 其中 T 为过渡矩阵. 对偶矢量的基变换恰好与矢量相反. 我们称基矢量为协变矢量，对偶基矢量为逆变矢量. 换基底造成的坐标变换为 $x^i = T^i_j\tilde{x}^j$ 以及 $\tilde{y}_j = T^i_j y_i$，可见，矢量的坐标变换规则与对偶基的相同，对偶矢量的坐标变换规则与矢量基底的相同，故称矢量的坐标为逆变坐标，对偶矢量的坐标为协变坐标.

7.3 张量

设 $V_1, V_2, \ldots, v_p, W$ 是矢量空间，则映射 $f : V_1 \times V_2 \times \cdots V_p \to W$ 称为多重线性映射，若对任意的 λ 和 μ 以及任意的 $i \in \{1, 2, \ldots, p\}$ 都有

$$f(v_1, \ldots, \lambda u_i + \mu v_i, \ldots, v_p) = \lambda f(v_1, \ldots, u_i, \ldots, v_p) + \mu f(v_1, \ldots, v_i, \ldots, v_p)$$

全体多重线性映射的集合记作 $\hom(V_1, \ldots, V_p; W)$. 对矢量空间 $V_1, \ldots, V_p$，矢量空间 T 配备多重线性映射 $\tau : \hom(V_1, \ldots, V_p; T)$ 称为矢量空间 $V_1, \ldots, V_p$ 的张量积，若对任意的矢量空间 W 和多重线性映射 $f \in \hom(V_1, \ldots, V_p; W)$，存在唯一的双线性映射 $\varphi \in \hom(T, W)$ 使得 $f = \varphi \circ \tau$.T 一般记作 $V_1 \otimes V_2 \otimes \cdots \otimes V_p$.

一般地，我们只把张量看成纯粹形式上的. 对任意的 $v_1 \in V_1, \cdots, v_i \in V_i$，它们的张量积形式地记作 $v_1 \otimes v_2 \otimes \cdots \otimes v_p$，满足多重线性性，即对任意的 λ 和 μ 以及任意的 $i \in \{1, 2, \ldots, p\}$ 都有

$$v_1 \otimes \cdots \otimes (\lambda u_i + \mu v_i) \otimes \cdots \otimes v_p = \lambda(v_1 \otimes \cdots \otimes u_i \otimes \cdots \otimes v_p) + \mu(v_1 \otimes \cdots \otimes v_i \otimes \cdots \otimes v_p)$$

全体 $v_1 \otimes \cdots \otimes v_i \otimes \cdots \otimes v_p$ 构成的空间记作 $V_1 \otimes \cdots \otimes V_i \otimes \cdots \otimes V_p$ 称为 $V_1, \cdots, V_i, \cdots, V_p$ 的张量积，其元素称为张量.

物理中，我们研究的是张量积 $T_q^p(V)$. 设 V 是一矢量空间，V^* 是它的对偶空间，则

$$T_q^p(V) = \underbrace{V \otimes \cdots \otimes V}_{p} \otimes \underbrace{V^* \otimes \cdots \otimes V^*}_{q}$$

称为 V 的 (p, q) 型张量积，其中的元素称为 (p, q) 型张量. 若 $q = 0$，则称为逆变张量积；若 $p = 0$，则称为协变张量积. 张量积 $T_q^p(V)$ 等同于多重线性映射 $\hom(V_1, \cdots, V_p, V_1^*, \cdots, V_q^*; F)$，就像一台机器，有 p 个上卡槽和 q 个下卡槽，装填完毕后，就输出一个数.

我们熟知的对偶矢量是 $(1, 0)$ 阶张量，因为对偶矢量是矢量空间上的线性泛函；而矢量是 $(0, 1)$ 阶张量，因为双重对偶空间与矢量空间之间可以建立起自然的同构；双线性形式则是 $(2, 0)$ 阶张量.

7.4 张量的基与坐标，张量运算

设 V 有基 $\{v_i\}$，V^* 有对偶基 $\{\varphi^i\}$，则 $T_q^p(V)$ 的基自然可以选为 $\{v_{i_1} \otimes \times \cdots \otimes v_{i_p} \otimes \varphi^{j_1} \otimes \cdots \varphi^{j_q}\}$. 张量 $t \in T_q^p(V)$ 可以展开为

$$t = \xi^{i_1 \ldots i_p}_{j_1 \ldots j_q} v_{i_1} \otimes \times \cdots \otimes v_{i_p} \otimes \varphi^{j_1} \otimes \cdots \varphi^{j_q}$$

$\xi^{i_1 \ldots i_p}_{j_1 \ldots j_q} v_{i_1}$ 称为张量的坐标.$i_1, \cdots, i_p$ 称为逆变指标，$j_1, \cdots, j_q$ 称为协变指标.

张量的加法和矢量几乎相同，有 $(a + b)^{i_1 \ldots i_p}_{j_1 \ldots j_q} = a^{i_1 \ldots i_p}_{j_1 \ldots j_q} + b^{i_1 \ldots i_p}_{j_1 \ldots j_q}$.

张量的乘法就是张量积，对 (p, q) 型张量 $t_1 = v_1 \otimes \cdots \otimes v_p \otimes \varphi^1 \otimes \cdots \otimes \varphi^q$ 和 (r, s) 型张量 $t_2 = u_1 \otimes \cdots \otimes u_r \otimes \psi^1 \otimes \cdots \otimes \psi^s$，其乘积 $t_1 \otimes t_2 = v_1 \otimes \cdots \otimes v_p \otimes u_1 \otimes \cdots \otimes u_r \otimes \varphi^1 \otimes \cdots \otimes \varphi^q \otimes \cdots \otimes \psi^1 \otimes \cdots \otimes \psi^s$. 若 t_1 的坐标为 $\{\xi^{i_1 \cdots i_p}_{j_1 \cdots j_p}\}$，$t_2$ 的坐标为 $\{\eta^{k_1 \cdots k_r}_{l_1 \cdots l_s}\}$，则 $t_1 \otimes t_2$ 的坐标为 $\{\xi^{i_1 \cdots i_p}_{j_1 \cdots j_p} \eta^{k_1 \cdots k_r}_{l_1 \cdots l_s}\}$.

设对 V 作基底变换，过渡矩阵为 $\mathbf{T} = \{t^i_j\}$，对偶基的过渡矩阵为 $\mathbf{S} = \{s^i_j\}$. 则基变换下张量的坐标变换为

$$\tilde{\xi}^{k_1 \cdots k_p}_{l_1 \cdots l_q} = t^{j_1}_{l_1} \cdots t^{j_q}_{l_q} s^{k_1}_{i_1} \cdots s^{k_p}_{i_p} \xi^{i_1 \cdots i_p}_{j_1 \cdots j_q}$$

7.5 张量的缩并

我们称线性映射

$$C_1^1 : T_q^p(V) \to T_{q-1}^{p-1}(V), v_1 \otimes \cdots \otimes v_p \otimes \varphi^1 \otimes \cdots \varphi^q \mapsto \varphi^1(v_1)(v_2 \otimes \cdots \otimes v_p \otimes \varphi^2 \otimes \cdots \varphi^q)$$

为对逆变指标 1 和协变指标 1 的缩并，对其它指标的缩并亦可以这样定义. 若 t 的坐标为 $\{\xi^{i_1\cdots i_p}_{j_1\cdots j_q}\}$，则 $C^1_1(t)$ 的坐标为 $\{\eta^{i_2\cdots i_p}_{j_2\cdots j_q}\} = \{\xi^{i_1\cdots i_p}_{j_1\cdots j_q}\delta^{j_1}_{i_1}\}$.

7.6 外积

设 X 和 Y 是线性空间，斜对称形式指多重线性映射 $L : X^k \to Y$，将它的任意两变量调换位置时其值变号：

$$L(\xi_1, \cdots, \xi_i, \cdots, \xi_j, \cdots, \xi_k) = -L(\xi_1, \cdots, \xi_j, \cdots, \xi_i, \cdots, \xi_k)$$

下文中我们以 A^p 表示 p 次斜对称形式，并简称为 p-形式.

引入斜对称形式的外积运算 $\wedge$，它将序偶 (A^p, B^q) 映为 $p + q$ -形式，且满足：

1. 结合性：$(A^p \wedge B^q) \wedge C^r = A^p \wedge (B^q \wedge C^r)$；
2. 分配性：$(A^p + B^p) \wedge C^q = A^p \wedge C^q + B^p \wedge C^q$；
3. 反交换性：$A^p \wedge B^q = (-1)^{pq}(B^q \wedge A^p)$

我们暂且只考虑实值形式，即 $Y = \mathbb{R}$.

1-形式就是矢量空间上的线性映射 $L^1 : X \to \mathbb{R}$，它根本就没有两个变量，因而也表现不出什么斜对称性. 我们定义 $L_1, \cdots L_k \in \mathfrak{L}(\mathbb{R}^n, \mathbb{R})$ 的外积，使得 $L_1 \wedge \cdots \wedge L_k$ 在 $\xi_1, \cdots, \xi_k \in \mathbb{R}^n$ 上有以下的值：

$$(L_1 \wedge \cdots \wedge L_k)(\xi_i, \cdots, \xi_k) = \det(L_j(\xi_i))$$

例 7.1. 投影映射 $\pi^i \in \mathfrak{L}(\mathbb{R}^n, \mathbb{R})$ 把 $\mathbb{R}^n$ 上的任意矢量 ξ 映为它的第 i 个分量 ξ^i. π^i 是线性映射，因而是 1-形式，于是有：

$$(\pi^1 \wedge \cdots \wedge \pi^k)(\xi_1, \cdots, \xi_k) = \det(\xi^j_i)$$

.

7.7 微分形式

我们先给出微分的严格定义：函数 $f : \mathbb{R}^m \supseteq E \to \mathbb{R}^n$ 称为在 $x \in E$ 点可微的，若：

$$f(x + h) - f(x) = L(x)(h) + \alpha(x, h)$$

其中 $L(x) : \mathbb{R}^m \to \mathbb{R}^n$ 是 h 的线性函数. 当 $h \to 0$ 时 $\alpha(x, h) = o(h)$.

注意到，微分是通过从 x 出发的位移 h 定义的. 我们把坐标原点在 x 的矢量空间 $\mathbb{R}^m$ 称为 x 处的切空间，记作 $T\mathbb{R}^m_x$. 切空间包含了从 x 出发所有可能的位移，亦即自变量所有可能的增量. 同样地 $T\mathbb{R}^n_{f(x)}$ 包含了因变量 $f(x)$ 一切可能的增量. 则微分 $L(x)$ (注意，这个符号整体是自变量为 h 的函数，而不是指 x 的函数 L) 实际上是在 $x \in \mathbb{R}^m$ 处定义的，从 x 的切空间到 $f(x)$ 的切空间的线性映射. 输入自变量的增量 h，输出因变量的增量 $L(x)(h)$. 这里 $L(x)h$ 并不一定是 h 导致的 $f(x)$ 的实际增量，但当 h 趋于 0 时，$[f(x) + L(x)(h)] - f(x + h)$ 以更快地速度趋于 0，这说明 x 附近 $L(x)(h)$ 能够很好地拟合 $f(x)$ 的实际增量. 于是我们得到了 x 附近 $f(x)$ 的良好的线性近似.

我们把 x 处的微分 $L(x)$ 记作 $\mathrm{d}f(x)$，并在以后的书写中略去自变量 h 的括号，即：

$$\mathrm{d}f(x)h$$

请千万不要认为这是将 $\mathrm{d}f(x)$ 与 h 相乘!

如果我们把 $\mathbb{R}^n$ 中的矢量都写成坐标形式，那定义 1 中的式子可以表示为：

$$f^i(x + h) - f^i(x) = L^i(x)h + \alpha^i(x, h)\ (i = 1, \cdots, n)$$

容易知道，$L^i(x):\mathbb{R}^m\to\mathbb{R}$ 也是线性函数，而 $h\to 0$ 时 $\alpha^i(x,h)=o(h)$. 即，映射 f 在 x 处可微当且仅当它的每个投影函数 f^i 在 x 处都可微.

再用坐标形式表示 x、h 和 $L(x)$，则定义 1 中的式子可进一步写成：

$$f(x^1+h^1,\cdots,x^m+h^m)-f(x^1,\cdots,x^m)=a_i(x)h^i+o(h)$$

这里用了 Einstein 求和约定，即同一指标重复出现表示对其遍历求和. $a_i(x)$ 是依赖于 x 的实数. 为定出这些实数，依次取 $h_i=h^ie_i$（这不再是 Einstein 求和），得：

$$f(x^1,\cdots,x^{i-1},x^i+h^i,x^{i+1},\cdots,x^m)-f(x^1,\cdots,x^i,\cdots,x^m)=a_i(x)h^i+o(h^ie_i)$$

这时，除第 i 个自变量外，其余自变量皆固定不变. 定义：

$$a_i(x)=\lim_{h^i\to 0}\frac{f(x^1,\cdots,x^{i-1},x^i+h^i,x^{i+1},\cdots,x^m)-f(x^1,\cdots,x^i,\cdots,x^m)}{h^i}$$

为 $f(x)$ 在 x 处关于变量 x^i 的偏导数，本文中记作 $\partial_i f(x)$. 这是 $f(x)$ 的增量除以一个实数，为 $\mathbb{R}^n$ 中的矢量. 于是 f 在点 x 处可微等价于 f 在 x 处关于每个自变量有偏导数，且：

$$\mathrm{d}f(x)h=\partial_i f(x)h^i$$

需要注意，等号左边是线性映射 $\mathrm{d}f(x)$ 作用在矢量 h 上，右边是 m 个矢量 $\partial_i f(x)$ 的线性组合，其系数恰是 h 的坐标. 映射是映射，矢量是矢量，务必不要搞混！

考虑投影映射 $\pi^i:\mathbb{R}^m\to\mathbb{R},(x^1,\cdots,x^i,\cdots,x^m)\mapsto x^i$，它把点 $x\in\mathbb{R}^m$ 映为它的第 i 个坐标. 对这个函数有：

$$\pi^i(x+h)-\pi^i(x)=(x^i+h^i)-x^i=h^i$$

于是对任意的点有 $\mathrm{d}\pi^i(x)h=h^i$，这么来看映射 $\mathrm{d}\pi^i(x)$ 是不依赖于 x 的，于是把 $\mathrm{d}\pi^i(x)$ 写成 $\mathrm{d}x^i$. 则：

$$\begin{aligned}\mathrm{d}f(x)h &= \partial_i f(x)h^i\\ &= \partial_i f(x)(\mathrm{d}x^ih)\\ &= [\partial_i f(x)\mathrm{d}x^i]h\end{aligned}$$

于是，任何函数的微分都可以表示为它的自变量的投影映射的微分的线性组合，即：

$$\mathrm{d}f(x)=\partial_i f(x)\mathrm{d}x^i$$

$f(x)$ 的微分可以用其各分量的微分表示，而分量的微分又能用它们各自的偏导数线性表示，于是有：

$$\begin{aligned}\mathrm{d}f(x)h &= \begin{pmatrix}\mathrm{d}f^1(x)h\\ \vdots\\ \mathrm{d}f^n(x)h\end{pmatrix}\\ &= \begin{pmatrix}\partial_i f^1(x)h^i\\ \vdots\\ \partial_i f^n(x)h^i\end{pmatrix}\\ &= \begin{pmatrix}\partial_1 f^1(x)\cdots\partial_m f^1(x) & \\ \vdots & \vdots\\ \partial_1 f^n(x)\cdots\partial_m^n(x) & \end{pmatrix}\begin{pmatrix}h^1\\ \vdots\\ h^m\end{pmatrix}\\ &= (\partial_i f^j(x))_{mn}h\end{aligned}$$

$(\partial_i f^j(x))_{mn}$ 称为 f 在 x 的 Jacobian. 因为 $f^j(x)$ 都是因变量为实数的映射，所以 $\partial_i f^j(x)$ 为实数，Jacobian 为依赖于 x 的实数矩阵. $h=h^ie_i$，故 Jacobian 可以看作线性映射 $\mathrm{d}f:T\mathbb{R}^m_x\to T\mathbb{R}^n_{f(x)}$ 在基 $\{e_i\}$ 上的矩阵.

设 $D\in\mathbb{R}^n$，$f\in C^{(1)}(D,\mathbb{R})$. 则对 $x_0\in D$，该点的微分 $\mathrm{d}f(x_0)$ 是线性函数 $TD_x\to T\mathbb{R}_{f(x)}\cong\mathbb{R}$. 于是，微分 $\mathrm{d}f$ 也是 $TD_{x_0}\cong\mathbb{R}^n$ 上的 1-形式. 对每一个 $x\in D$，都产生一个线性形式，也就是说，f 在 D 内产生了一个线性形式场，其中每一点的线性形式都定义在相应的切空间上.

如果在每一点 $x\in D$ 上确定了一个斜对称形式 $\omega(x):(TD_x)^p\to\mathbb{R}$，则称在区域 $\mathbb{R}$ 内给出了一个实值的微分 $p-$ 形式 ω.

可以看到，微分形式实际上是与 x 相关的一族斜对称形式，对每一个 x，在其切空间 TD_x 上给出一个具体的斜对称形式.

下面我们给出物理上最常见的两种微分形式.

例 7.2. 在区域 D 内给出一个矢量场，即在 $x\in D$ 指定一个矢量 $\mathbf{F}(x)$. 如果 $\mathbb{R}^n$ 装备有内积结构，则 $\omega^1_{\mathbf{F}}(x)=\langle\mathbf{F}(x),\cdot\rangle$ 给出了 D 内的一个微分 1-形式. 我们很熟悉，在力场 $\mathbf{F}(x)$ 中走一个位移 ξ，场做的元功正是 $\omega^1_{\mathbf{F}(x)}(\xi)=\langle\mathbf{F}(x),\xi\rangle$，因此这种微分形式称为场的功形式.

例 7.3. 在区域 D 内给出一个矢量场 $\mathbf{V}(x)$，在 $x\in D$ 处，取对应的矢量 $\mathbf{V}(x)$ 以及 $\xi_1,\cdots\xi_{n-1}\in TD_x$，则以 $\mathbf{V}(x),\xi_1,\cdots,\xi_{n-1}$ 为棱的平行多面体的体积, 亦即 $\begin{vmatrix}\mathbf{V}(x) & \xi_1 & \cdots & \xi_{n-1}\end{vmatrix}$, 是变量 $\xi_1,\cdots,\xi_{n-1}$ 的 $(n-1)$-形式. 若 $\mathbf{V}(x)$ 为区域 D 内流体的流速场，则这恰是单位时间内流过以 $\xi_1,\cdots,\xi_{n-1}$ 为边的 $(n-1)$ 维平行四边形的流体体积，于是，我们把 $\omega^{n-1}_{\mathbf{V}(x)}=\begin{vmatrix}\mathbf{V}(x) & \cdot & \cdots & \cdot\end{vmatrix}$ 叫做流速场 $\mathbf{V}(x)$ 的流形式.

下面我们讨论斜对称形式及微分形式的坐标表示. L 是 $\mathbb{R}^n$ 内的 k-线性形式，对 $\mathbf{R}^n$ 内的一组基 $\{e_i\}$，每个矢量 $\xi\in\mathbb{R}^n$ 都能写成 $\xi=\xi^ie_i$ 的形式，则:

$$\begin{aligned}L(\xi_1,\cdots,\xi_k) &= L(\xi_1^{i_1}e_{i_1},\cdots,\xi_k^{i_k}e_{i_k})\\ &= L(e_{i_1},\cdots,e_{i_k})\xi_1^{i_1}\cdots\xi_k^{i_k}\end{aligned}$$

这里还是用了 Einstein 约定，对从 $1,\cdots,n$ 中取出的 $i_1,\cdots,i_k$ 遍历求和. 进一步计算可以知道:

$$\begin{aligned}L(\xi_1,\cdots,\xi_k) &= \sum_{1\le i_1<\cdots<i_k\le n}L(e_{i_1},\cdots,e_{i_k})\begin{vmatrix}\xi_1^{i_1} & \cdots & \xi_1^{i_k}\\ \vdots & & \vdots\\ \xi_k^{i_1} & \cdots & \xi_k^{i_k}\end{vmatrix}\\ &= \sum_{1\le i_1<\cdots<i_k\le n}a_{i_1\cdots i_k}[(\pi^{i_1}\wedge\cdots\wedge\pi^{i_k})(\xi_1,\cdots,\xi_k)]\end{aligned}$$

即:

$$L=\sum_{1\le i_1<\cdots<i_k\le n}a_{i_1\cdots i_k}(\pi^{i_1}\wedge\cdots\wedge\pi^{i_k})$$

也就是说，任何 k-形式都可以表示为形如 $\pi^{i_1}\wedge\cdots\wedge\pi^{i_k}$ 的 k-形式的线性组合.

设在区域 D 上给出 k-微分形式，在点 $x\in D$ 处，$\mathrm{d}x^i$ 就是 TD^n_x 上的投影映射，于是有:

$$\omega(x)=\sum_{1\le i_1<\cdots<i_k\le n}a_{i_1\cdots i_k}(x)(\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_k})$$

即任何微分 k-形式都是简单微分 k-形式 $\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_k}$ 的线性组合. 系数 $a_{i_1\cdots i_k}(x)$ 是 x 的函数.

例 7.4. 微分 1-形式 $\mathrm{d}f$ 可以明显地分解为 $\partial_i(x)\mathrm{d}x^i$.

例 7.5. 计算可知，在笛卡尔坐标系中，功形式 $\omega^1_{\mathbf{F}}(x)$ 可分解为 $\sum_{i=1}^nF^i(x)\mathrm{d}x^i$.

例 7.6. 计算可知，流形式 $\omega_{\mathbf{V}^{n-1}}$ 可展开为：

$$\sum_{i=1}^{n}(-1)^{i+1}V^i(x)\mathrm{d}x^1\wedge\cdots\wedge\overline{\mathrm{d}x^i}\wedge\cdots\mathrm{d}x^n$$

其中微分上的横杠表示这一项应舍去.

7.8 外微分形式以及微分形式的转移

以上我们都是从代数角度讨论微分形式，下面我们引入外微分运算，从而使我们得以用分析的手段研究微分形式.

定义 7.1. 约定区域 D 内定义的函数 $f\in\mathbb{R}$ 为 D 内的零次微分形式，f（在分析学中定义）的微分称为 0-形式 f 的外微分.

若在区域 D 内给出微分 p-形式：

$$\omega(x)=a_{i_1\cdots i_p}(x)\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}$$

则它的外微分形式定义为：

$$\mathrm{d}\omega(x)=\mathrm{d}a_{i_1\cdots i_p}\wedge\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}$$

由：

$$\mathrm{d}\omega(x)=\partial_i a_{i_1\cdots i_p}(x)\mathrm{d}x^i\wedge\mathrm{d}x^{i_1}\wedge\cdots\wedge\mathrm{d}x^{i_p}$$

可知 p-微分形式的外微分是 $(p+1)$-微分形式. 于是，我们可以从函数 f 出发，通过外微分运算，归纳地构造任意次（只要函数足够光滑）的微分形式.

下面给出最重要的三种外微分.

例 7.7. 设

$$\omega=P\mathrm{d}x+Q\mathrm{d}y$$

则

$$\begin{aligned}\mathrm{d}\omega&=\mathrm{d}P\wedge\mathrm{d}x+\mathrm{d}Q\wedge\mathrm{d}y\\&=(\partial_xP\mathrm{d}x+\partial_yP\mathrm{d}y)\wedge\mathrm{d}x+(\partial_xQ\mathrm{d}x+\partial_yQ\mathrm{d}y)\wedge\mathrm{d}y\\&=\partial_yP\mathrm{d}y\wedge\mathrm{d}x+\partial_xQ\mathrm{d}x\wedge\mathrm{d}y\\&=(\partial_xQ-\partial_yP)(\mathrm{d}x\wedge\mathrm{d}y)\end{aligned}$$

例 7.8. 设

$$\omega=P\mathrm{d}x+Q\mathrm{d}y+R\mathrm{d}z$$

则

$$\mathrm{d}\omega=(\partial_yR-\partial_zQ)\mathrm{d}y\wedge\mathrm{d}z+(\partial_zP-\partial_xR)\mathrm{d}z\wedge\mathrm{d}x+(\partial_xQ-\partial_yP)\mathrm{d}x\wedge\mathrm{d}y$$

例 7.9. 设

$$\omega=P\mathrm{d}y\wedge\mathrm{d}z+Q\mathrm{d}z\wedge\mathrm{d}x+R\mathrm{d}x\wedge\mathrm{d}y$$

则

$$\mathrm{d}\omega=(\partial_xP+\partial_yQ+\partial_zR)\mathrm{d}x\wedge\mathrm{d}y\wedge\mathrm{d}z$$

下面我们来讨论区域的映射下，微分形式发生怎样的变化.

设 $\phi: U \to V$ 是 $U \subseteq \mathbb{R}^m$ 到 $V \subseteq \mathbb{R}^n$ 的映射. f 是 V 上的函数，则可以产生函数

$$(\phi^* f)(t) \triangleq f(\phi(t))$$

因此，运算 $f \mapsto \phi^* f$ 把 V 上函数的集合映入 U 上的函数的集合，换言之，把 V 上定义的零形式映为 U 上定义的零形式.

假定 ϕ 光滑，$\phi'(t): TU_t \to TV_{\phi(t)}$ 为 ϕ 的切映射，而 ω 是 V 上的一个 p-形式. 这时，我们取 U 上的 p-形式 $\phi^*\omega$ 与之对应，其值由

$$(\phi^*\omega(t))(\tau_1, \cdots, \tau_p) = \omega(\phi(t))(\phi'(t)\tau_1, \cdots, \phi'(t)\tau_p)$$

决定.

这就是区域映射引起的微分形式的转移. 特别地，可以算出

$$\begin{aligned} &\phi^*\left(\sum_{1 \le i_1 < \cdots < i_p \le n} a_{i_1 \cdots i_p}(x) \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_p}\right) \\ = &\sum_{1 \le i_1 < \cdots < i_p \le n}^{1 \le j_1 < \cdots < j_p \le n} a_{i_1 \cdots i_p}(x(t)) \frac{\partial(x^{i_1}, \cdots, x^{i_p})}{\partial(t^{j_1}, \cdots, t^{j_p})} \mathrm{d}t^{j_1} \wedge \cdots \wedge \mathrm{d}t^{j_p} \end{aligned}$$