数据分析与机器学习中的 矩阵理论及应用

刘维湘—深圳大学

20200606

摘要

线性空间与内积空间：在非空集合的基础上，引入代数结构（线性运算：加法与数乘）得线性空间；在线性空间的基础上，引入拓扑(度量) 结构 (内积) 得内积空间，由此可以度量线性空间中元素的长度、向量间夹角与距离等拓扑属性。其中向量正交的概念非常重要，在此基础上的正交投影与最佳逼近在信号分析、数据处理与机器学习中有广泛的应用。

线性映射与线性变换：一个线性空间到另一个线性空间的线性映射在理论与实践问题中非常重要。有限维线性空间到自身的线性变换，可以通过矩阵来表示。在此基础上认识矩阵的本质就是线性变换。突破日常数据中对数据表示成”矩阵”的认识。研究线性变换 (矩阵) 的性质，比如矩阵的特征值与特征向量及其性质，以及特殊的线性变换如正交 (酉) 变换。

《冬夜读书示子聿》

（南宋）陆游

古人学问无遗力，少壮工夫老始成。

纸上得来终觉浅，绝知此事要躬行。

目录

1 线性空间与内积空间

1.1 引言

对一般工科大学生来讲，已经具备有关集合、函数及线性代数的基本概念。其中有关空间的概念，主要是从日常的三维空间推广到了 $n(n \geq 3)$ 维实向量空间。在" 大数据" 的时代，产生的数据呈现高维数、多维度、多模态的趋势。比如，在生物信息学领域，一个基因表达谱数据反映的是上万个基因的表达水平；在生物医学的功能磁共振成像数据中，单个被试的数据则呈现为一个四阶张量的形式[1]。这些数据有其自身的特点，比如在网络科学 (生物基因网络、脑网络、社交网络) 中，数据可能以一个对称矩阵（甚至是正定矩阵）的形式出现。在对这些信号或数据进行处理时，特别在基于数据驱动的机器学习领域，如何刻画它们之间的距离、相似性或相关性，就有必要把原有的空间概念进行推广，并引入更多的数学结构便于分析[2]。

1.2 预备知识

1.2.1 集合及其运算

集合与关系是现代数学[3]的基础，我们这里先回顾相关的基本概念：

[1]这里关于张量的概念，是来自数据的表示，从实际工程中直观的理解如下：单个的数是零阶张量，多个元素构成的向量是一阶张量，矩阵是二阶张量，一个三通道的彩色图像则表现为是一个三阶张量，彩色的视频就是四阶张量了。若要从物理或数学上去理解有关张量的概念，需要参考相关的专业教材。通常 (我) 在机器学习中遇到的张量分解和流形学习，会把非数学物理专业的工科学生 (我) 搞懵。

[2]关于数学结构，就是把数学中的各种研究对象给出一个统一的表现形式。法国布尔巴基学派采用全局观点，着重分析各个数学分支之间的结构差异和内在联系，他们认为数学的基本结构有三种母结构：代数结构，拓扑结构和序结构。通过以上三种母结构的变化、复合、交叉形成各种数学分支 [百度百科]。对我们工科研究生来讲，如何结合自己的课题方向，充分发挥现有数学工具的作用，完美解决自己的实践问题。

[3]数学是研究数和形的科学。大致内容包括三个方面，即代数，几何与分析。这三个方面在发展中也相互影响和渗透，让数学这门学科变得异常的生动而美丽多彩！现代数学教

- 集合，数域，数集的上/下界（确界）；
- 笛卡尔积 (Cartesian product)，二元关系，等价关系；
- 映射，代数运算，二元运算，运算封闭。

定义 1.1. 集合与元素　具有某种特定属性的事物的全体称之为集合 *(Set)*，常用大写字母表示。这个集合里单独的一个事物，称之为集合的元素 *(Element)*，常用小写字母表示。对一个集合的描述通常用列举法或描述法进行。集合里的元素与集合的关系可以用 $\in$ 来表示; 若某个事物不属于某个集合，则用 $\notin$ 表示。集合中的元素个数，称之为集合的基数 *(Cardinality)*，常用简写 *card(A)* 来表示集合 A 的基数。含有有限个集合的集合为有限集；否则为无限集。不含任何元素的集合称为空集，记为 $\emptyset$.

例 1.1. 包含有三个元素的集合 $A=\{a_1,a_2,a_3\}$, $a_1\in A$; 二维平面上与原点的距离小于等于 1 的点的集合 $C=\{(x,y)|x^2+y^2\le 1\}$,$(1,2)\notin C$。这里的 A 是有限集，C 是无限集。

定义 1.2. 数域 *(Field)*　包含 0 和 1 的数集，且其中元素的四则运算 (除数不为 0) 封闭[4]。

例 1.2. 根据定义，有理数的全体构成一个数域 Q。常用的数域：实数域 R 和复数域 C，通常用 $F/K/P$ 来表示。

练习 1.1. 证明数集 $Q_{\sqrt{2}}=a+\sqrt{2}b|a,b\in Q$ 是数域。

定义 1.3. 非空实数集的上、下（确）界　设 A 是非空实数集，若存在 u，使得任意 $a\in A$，有 $a\le u$, 则称 u 为 A 的上界（*upper bound*）；A 的最小上界称为上确界 *(supremum)*, 记为 $\sup A$；若存在 l，使得任意 $a\in A$，有 $a\ge l$, 则称 l 为 A 的下界（*lower bound*）；A 的最大下界称为下确界 *(infimum)*, 记为 $\inf A$。

材在大量使用公理化方法的同时，一定要告诉其发展历史。往往是这些数学发展历史蕴藏着先人无穷的智慧，这个或许比那些抽象的数学知识来的更珍贵。

[4]运算封闭就是指运算的结果还在原来集合里，参考后续的代数运算封闭的定义。

例 1.3. 由定义可知，若数集有最大值，则必是其上确界；有最小值，则必是其下确界。比如集合 $A=[0,1]$，则 $\sup A=1, \inf A=0$。

练习 1.2. 考虑集合 $A=(0,1)$，求其上、下确界。

定义 1.4. 集合运算　设有 A,B 两个集合,如果集合 A 的元素都是集合 B 的元素,则称 A 为 B 的子集(*subset*),或 B 包含 A,记为 $A\subseteq B$ 或 $B\supseteq A$. 我们规定空集是任意集合的子集。如果 $A\subseteq B$ 且 $A\supseteq B$, 则称集合 A 与 B 相等,记为 $A=B$. 由属于 A 或 B 的所有元素组成的集合称为 A 与 B 的并集(*union*),记为 $A\cup B=\{x|x\in A \text{或} x\in B\}$;由同时属于 A 和 B 的元素构成的集合称为 A 与 B 的交集 *(intersection)*, 记为 $A\cap B=\{x|x\in A \text{且} x\in B\}$. 集合 A 与 B 的差集定义为 $A-B=\{x|x\in A, x\notin B\}$. 相对全集 S 来说，A 的补集就是 $\overline{A}=S-A$. 集合的并、交、补运算，满足以下运算规律。

定理 1.1. 设 A，B，C 三个集合，则

(1) $A\cup B=B\cup A$, $A\cap B=B\cap A$ ；

(2) $A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$, $A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$ ；

(3) $\overline{A\cup B}=\overline{A}\cap\overline{B}, \overline{A\cap B}=\overline{A}\cup\overline{B}$.

例 1.4. 通常用集合的文氏图来表示其运算。

定义 1.5. 笛卡尔积 *(Cartesian product)*，二元及多元关系　设 A、B 两个非空集合，由他们的元素构成的序对集合 $\{(a,b)|a\in A, b\in B\}$ 称为 A 与 B 的笛卡尔积 *(*或直积*)*, 记为 $A\times B$。可以推广到多个集合，比如 $A\times B\times C=\{(a,b,c)|a\in A, b\in B, c\in C\}$. 我们称 $A\times B$ 的一个子集为其中的一个二元关系 *(Binary relation)*；若 $(a,b)\in R\subset A\times B$，记为 aRb。同理，$A\times B\times C$ 的一个子集则是一个三元关系。

定义 1.6. 等价 *(Equivalence)* 关系　集合 A 上的一个二元关系 R 满足如下三个条件:

(1) 自反性 *(reflexivity),* 对任意 $a \in A$, 有 aRa;

(2) 对称性 *(symmetry),* 对任意 $a, b \in A$, 如果 aRb, 则 bRa;

(3) 传递性 *(transitivity),* 对任意 $a, b, c \in A$, 如果 aRb, 则 bRc, 则 aRc;

则称 R 是 A 上的一个等价关系。

例 1.5. 实数集 $\mathbb{R}$ 上的相等 ”=” 是 $\mathbb{R}$ 上的等价关系。

例 1.6. 设集合 T 表示所有的等边三角形的集合，则该集合上的相似 ”$\sim$” 是 T 上的等价关系。

1.2.2 映射

定义 1.7. 映射　映射是函数概念的推广. 对两个非空集合，在规则 f 的作用下, 把 A 中的每一个元素 x 都能唯一的对应到 B 中的一个元素 y，则称 f 是 A 到 B 的一个映射，$y = f(x)$, 常用记号 $f : A \to B$ 表示. 一个映射有三个要素 (f, A, B): 定义域 *(Domain)*A，值域（*Range*）B，映射规则 f. 有时候也用 $D(f), R(f)$ 分别表示映射 f 的定义域和值域。

定义 1.8. 单位映射/恒等映射　$I_a : a \to a, a \in A$.

定义 1.9. 单映射，满映射，双映射

定义 1.10. 映射的乘积，复合映射　$f_1 : A \to B, f_2 : B \to C, f_3 = f_2 \cdot f_1 : A \to C$.

定义 1.11. 逆映射　$f : A \to B, g : B \to A, f \cdot g = I_B, g \cdot f = I_A, g = f^{-1}$. g 为 f 的逆映射。

定义 1.12. 变换　集合 A 到自身的映射称为 A 的变换。

定义 1.13. 代数运算, 运算封闭　设 A, B, C 为三个非空集合,$A \times B$ 到 C 的映射称为 A 与 B 到 C 的一个代数运算 (有时简称运算)。若 A=B=C，则称该映射为 A 的代数运算或二元运算，并称 A 对代数运算封闭。[5]

[5]代数，就是用字母符号代替具体的数（包括已知量和未知量），这里的数泛指一切可研究的数据 (data)。

例 1.7. 比如实数集 $\mathbb{R}$ 上的加法运算，就是实数集上的一个二元运算：$+:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$，对任意 $x,y\in\mathbb{R}$，有 $x+y\in\mathbb{R}$；数乘运算也是一个二元运算 $*:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$，对任意 $x,y\in\mathbb{R}$，有 $x*y\in\mathbb{R}$.

例 1.8. 考虑 2 维空间 $\mathbb{R}^2$ 上的向量加法运算"+"与数乘运算"*"：对任意 $x=[x_1,x_2]^T, y=[y_1,y_2]^T, k\in\mathbb{R}$,

$$+: x+y=[x_1+y_1, x_2+y_2]^T,$$

$$*: k*x=[kx_1, kx_2]^T.$$

注 1.1. 由定义可知，代数运算本质上是一种映射。

1.3 线性空间

给定一定的集合，我们可以在集合上定义一些数学结构，进一步研究集合的性质[6]。这里主要的基本概念包括：

- 线性 (向量) 空间的定义 (注意零元素的定义)，常见的线性空间，线性空间的性质；
- 线性组合，线性相关/无关，线性空间的维数，基与坐标，过渡矩阵；
- 线性子空间，张成空间，子空间的交与和、直和。

1.3.1 线性空间的定义

随着研究对象的广泛深入，人们通过归纳抽象，并以公理化的方式给出线性空间的定义，即引入集合上的代数结构（两个线性运算及运算规则）.

[6]现在的数学教材大多以公理化的形式对相关的概念进行介绍，然后引出公理、定理、推论、性质，最后给出应用案例。要搞清楚其中的来龙去脉，特别是概念发生的背景和历史的演变过程，对数学修养要求就拔高了。很多的概念和方法是经过前人不停的积累、批判、继承和发扬，有的数学家甚至为此而奉献自己的生命（比如传说中现代群论 (group theory) 的创始人伽罗瓦 Évariste Galois）。

定义 1.14. 线性空间 *(*向量空间*)*：四个要素 $\{V, F; +, \cdot\}$，即一个非空集合 V，一个数域 F，两个 *(*代数*)* 运算（加法运算与数乘运算，这两个运算就是线性运算），八条运算规则（加法运算与数乘运算各 *4* 条）。线性空间 = 非空集合 V + 数域 F + 运算 *(*加法与数乘及运算规则*)*。

注 1.2. 根据习惯，线性空间 *(linear space)* 也叫向量空间 *(vector space)*, 但是这里的向量是一种广义的说法了。向量空间 V 中的元素可能是 n 维向量，可能是矩阵，甚至可能是多项式或其它函数。

注 1.3. 线性空间的零元素*(*用 0，或 θ 表示，根据上下文可判断*)* 和负元素的特殊含义。

常见的线性空间：

例 1.9. R^n，其中的向量就是我们通常说的 n 维向量。

例 1.10. $R^{m\times n}$, 其中的向量是一个 $m \times n$ 的矩阵。

例 1.11. $P[x]$, 所有一元多项式的全体；$P_n[x]$ 多项式空间，全体次数不超过 n 的一元多项式的集合。其中的向量是一个多项式，比如 $p_1(x) = 1, p_2(x) = 1 + x^3$。

例 1.12. $C[a, b]$ 在区间 $[a, b]$ 上连续的全体实函数的集合。其中的向量是一个连续函数，比如 $f(x) = \sin x, g(x) = e^x$.

练习 1.3. $V = R_+, F = R, \forall x, y \in V, k \in F$, 并定义加法和数乘运算如下：

加法运算 $\oplus$:

$$x \oplus y = xy$$

数乘运算 $\odot$:

$$k \odot x = x^k$$

试证明 $\{V, F; +, \cdot\}$ 是线性空间。注意其中的零元素和负元素。

定理 1.2. 线性空间的性质

注 1.4. 现代数学的崛起, 呈现出一定的规律，即一方面在理论上通过对大量研究对象的归纳、抽象与推广, 从公理化的方法进行定义、讨论性质，给出推论等，然后就是各种应用；另一方面则是大量的实践需求促进人们对现有理论和方法的改进和革命。

1.3.2 线性表示与线性相关

定义 1.15. 向量的线性表示与线性相关　设 V 是数域 P 上的线性空间，$\alpha_i \in V, k_r \in P, i = 1, \ldots, r$，如果 $\alpha \in V$ 可以表示为

$$\alpha = \sum_{i=1}^{r} k_i \alpha_i,$$

则称 α 可由 $\alpha_i, i = 1, \ldots, r$ 线性表示，或者说 α 是 $\alpha_i, i = 1, \ldots, r$ 的线性组合。如果存在不全为 *0* 的 $k_i \in P, i = 1, \ldots, r$ 使得

$$\sum_{i=1}^{r} k_i \alpha_i = 0$$

则称 $\alpha_i, i = 1, \ldots, r$ 线性相关；否则称为线性无关。

例 1.13. 支持向量机的最终决策函数就是支持向量的线性组合函数 *[]*。

例 1.14. 人脸的稀疏表示分类器 *[]*

定义 1.16. 线性空间的维数　如果线性空间 V 中有 n 个线性无关的向量，但是没有更多数目的线性无关的向量，则称 V 是 n 维的，记为 $dim(V) = n$; 如果 V 中可以找到任意多个线性无关的向量，则称 V 是无限维的.

例 1.15. 无限维线性空间的例子. 所有一元多项式全体定义在实数域上的线性空间 *P[x]*;$C[a, b]$ 为 $[a, b]$ 上所有连续实函数的集合, 数域 $\mathbb{R}$, 定义连续函数的加法 *(+)* 及数乘 *(·)*, 则 $\{C[a, b], R; +, \cdot\}$ 是无限维线性空间.[7]

---

[7]泛函分析是研究无限维空间的理论. 关于有限维空间和无限维空间的不同, 可以参考 [].

例 1.16. 流形学习（*Manifold learning*）的兴起 *[]*，其中一个本质的假设就是高维数据是嵌入在低维空间中的。其中的一个主要任务就是给定一组观测数据，估算该样本数据的本质维度 *[]*。

例 1.17. 另外一个任务就是，对高维数据进行低维的可视化，比如早期的 *PCA[]*，*MDS[]* 降维技术，流形学习的代表 *LLE[]*，*Isomap []*，以及其扩展，比如基于深度学习的 *auto-encoder[]*。

注 1.5. 在大数据时代 *("BIG DATA ERA")*，数据的高维化带来了所谓的"维数诅咒（*Curse of Dimensionality*）"，给现代数据分析带来了挑战。参考 *David L. Donoho* 的讲座 *[]*，*Jianqin Fan* 的综述 *[]*.

注 1.6. 关于流形上的优化 *[]*，矩阵上的优化 *[]* 为现代大数据和机器学习提供工具。

定义 1.17. 向量组等价

在有了线性组合的基础上，我们可以给出线性空间的凸集概念了。

定义 1.18. 凸集　设 V 时数域 P 上的线性空间，W 是 V 的子集，如果任意 $\alpha, \beta \in W$，以及任意 $\lambda \in [0,1]$，都有 $\lambda\alpha + (1-\lambda)\beta \in W$，则称 W 是凸集。

注 1.7. 凸集的概念是凸优化中最基本的概念之一 *[]*. 深度学习中一种用于数据扩增的方法 *mix-up* 就是采用训练集的凸组合完成 *[]*。为什么不试试其它组合方式呢？

1.3.3　线性空间的基与坐标

定义 1.19. 基　设 V 是数域 P 上的 n 维线性空间，向量组 $\epsilon_i \in V, i = 1, \ldots, n$ 线性无关，则成为 V 中的一组基。V 中的任意向量 α 都可以唯一的表示为该组基的线性组合，即

$$\alpha = \sum_{i=1}^{n} x_i \epsilon_i = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n) x,$$

其中，$x = (x_1, x_2, \ldots, x_n)^T \in R^n$ 称为 α 在基 $(\epsilon_1, \epsilon_2, \ldots, \epsilon_n)$ 下的坐标.

例 1.18. 在 n 维线性空间 P^n 中。

例 1.19. 在线性空间 $P_n[x]$ 中（实数域上次数小于 n 的一元多项式的线性空间），$dim(P_n[x]) = n$, 一组基可以是 $\epsilon_1 = 1, \epsilon_2 = x, \ldots, \epsilon_n = x^{n-1}$, 则多项式 $f(x) = \sum_{i=0}^{n-1} a_i x^i$ 在这组基下的坐标为 $(a_0, a_1, \ldots, a_{n-1})^T$. 考虑 $f(x)$ 的 *Taylor* 展开式，可以得到另外一组基及其坐标。

注 1.8. 在 n 维线性空间 V 中，任意 n 个线性无关的向量都可以取作 V 的基。同一向量在不同的基下坐标也不一样，它们的关系可以通过过渡矩阵表示出来。

注 1.9. 在数据分析中，选择一组基并考虑数据在该基上的投影是一个艺术。目前基的选择大致可分为三类：固定的一组基（*DFT*, *WT*），基于现有数据的一组基（*PCA*），以及随机的一组基（*RP*，适于高维数据）。

课后思考与练习 1.1. 给出在实数域上线性空间 $R^{2\times 2}$ 中的一组基，并给出其中一个向量在该组基下的坐标。

1.3.4 线性子空间

定义 1.20. 线性子空间　设 W 是数域 P 上线性空间 V 的非空子集，如果 W 对于 V 中的两种运算也构成数域 P 上的线性空间，则称 W 是 V 的一个线性子空间，简称子空间 *(subspace)*。

定理 1.3. 数域 P 上的线性空间 V 的非空集合 W 是 V 的一个线性子空间的充要条件是 W 对 V 的两种运算封闭，即

(1) 如果 $\forall \alpha, \beta \in W$, 则 $\alpha + \beta \in W$;

(2) 如果 $k \in P, \alpha \in W$, 则 $k\alpha \in W$。

注 1.10. 线性空间的每个子空间都是凸集。

定义 1.21. 张成的子空间 $W = span\{\alpha_1, \ldots, \alpha_m\}$

定理 1.4.

例 1.20. 设矩阵 $A \in R^{\{}m \times n\} = [\alpha_1, \ldots, \alpha_n]$, 称 $span(A) = R(A) = \{y|y = Ax, x \in R^n\}$ 是矩阵 A 的列空间 *(*值域*)*。

定义 1.22. 子空间的交与和

定理 1.5.

定义 1.23. 子空间的直和

定理 1.6.

例 1.21.

定理 1.7. 维数公式　设 V_1, V_2 是数域 P 上的线性空间 V 的两个有限维子空间，则 $V_1 \cap V_2$, $V_1 + V_2$ 都是有限维的，并且

$$dim(V_1) + dim(V_2) = dim(V_1 + V_2) + dim(V_1 \cap V_2).$$

注 1.11. 信号处理中的小波 *(wavelet)* 分析, 多分辨率 *(multi-resolution)* 分析就是考虑空间的直和分解。

注 1.12. 机器学习中的子空间方法。

1.3.5 线性空间的同构

定义 1.24. 同构映射

定理 1.8.

例 1.22.

注 1.13. 同构的 n 维向量空间之间的关系

1.4 内积 (度量) 空间

实数集上两点之间的距离, 三维空间中两点的直线距离, 都是我们最熟悉的距离度量. 除了代数结构, 线性空间的拓扑结构是另外一种数学结构. 它是在我们通常说的度量结构基础上的推广, 其中的度量是描述极限和连续性这些分析属性的一般框架[8]。度量在信号与信息处理，图像处理与计算机视觉，人工智能，模式识别与机器学习等应用中扮演非常重要的作用。我们这里先从实函数的连续性开始说起, 以此说明距离度量的重要性. 后面再继续介绍的内积就是一种度量。

1.4.1 函数的连续性

定义 1.25. 函数的连续性与一致连续性　设 $f(x)$ 是区间 $[a,b]\in R$ 上的实函数[9]，在 $x_0\in[a,b]$ 处，若对任意 $\epsilon>0$, 存在 $\delta>0$, 使得当 $|x-x_0|<\delta$ 时，有 $|f(x)-f(x_0)|<\epsilon$, 则称 $f(x)$ 在点 x_0 是连续的（*continuous*）; 若 $f(x)$ 在每个点 $x\in[a,b]$ 都是连续的，则称其在区间 $[a,b]$ 上是连续的；若对任意 $\epsilon>0$, 存在 $\delta>0$, 使得当 $x_1,x_2\in[a,b],|x_1-x_2|<\delta$ 时，有 $|f(x_1)-f(x_2)|<\epsilon$, 则称 $f(x)$ 在区间 $[a,b]$ 是一致连续的（*uniformaly continuous*）.

注 1.14. 由以上定义可以看出，连续性是一个局部概念，而一致连续性是整体概念。这里的 $|x_1-x_2|$ 以及 $|f(x_1)-f(x_2)|$ 都用到了距离的概念。当我们的研究对象不再是实数时，就有必要拓展距离的概念了。

例 1.23. 连续函数 $f(x)=\frac{1}{x}$ 在区间 $(0,1)$ 上不是一致连续的。请读者自行证明。

定义 1.26. *Lipschitz* 函数　对于在实数集的子集的函数 $f:I\in R\to R$ ，若存在常数 K，使得 $|f(a)-f(b)|<K|a-b|,\forall a,b\in I$ ，则称 f 符合

[8]想想黎曼度量 [] 和 Finsler 度量 [] 吧。然而没有一定的基础直接入门拓扑学的专业书籍（比如 []），感觉很抽象。只有假以时日，遍访名家/书，自己慢慢去领悟吧。

[9]这里的区间可能不是闭区间，需要根据上下文进行确定。这里采用闭区间的表示法只是为了简单示意而已。

利普希茨连续条件, 称 f 为 *Lipschitz* 函数，或 *L*-函数*;* 对于 f 最小的常数 K 称为 f 的利普希茨常数。若 $K<1$ ，f 称为收缩映射。

注 1.15. 利普希茨连续条件（*Lipschitz continuity*），以德国数学家鲁道夫 • 利普希茨命名，是一个比通常连续更强的光滑性条件。直觉上，利普希茨连续函数限制了函数改变的速度，符合利普希茨条件的函数的斜率（即导数），必小于一个称为利普希茨常数的实数（该常数依函数而定）。符合利普希茨条件的函数是一致连续的，也是连续的。在优化领域，经常考虑同时具有 *Lipschitz* 性和凸性的函数 *[]*。在深度学习中，也有关于深度学习中的 *Lipschitz* 模型的研究 *[]*。

例 1.24. 可以证明导函数有界的函数都满足 *Lipschitz* 条件。比如 $f(x) = x^2$ 在区间 $[-1,1]$ 上的导数有界，则其在该区间上是满足 *Lipschitz* 条件的。

1.4.2 度量空间的定义

定义 1.27. 度量空间 设 V 为一个非空集合，一个映射 $d: V\times V \to R_+)$。若对于任意 $x,y,z\in V$，有

(1) 非负性*:* $d(x,y)\geq 0$，且 $d(x,y)=0$ 当且仅当 $x=y$；

(2) 对称性*:* $d(x,y)=d(y,x)$；

(3) 三角不等式*:* $d(x,z)\leq d(x,y)+d(y,z)$.

则称 d 为集合 V 的一个度量（或距离）。称偶对 (V,d) 为一个度量空间，或者称 V 为一个对于度量 d 而言的度量空间。

例 1.25. 在 R^n 上定义距离, $x,y\in R^n$, $d(x,y)=(\sum_{i=1}^n |x_i-y_i|^p)^{\frac{1}{p}}, p\in [1,+\infty)$. 如果 $p\in[0,1)$，该距离会怎么样？试试在二维平面的情况，画图进行说明，并讨论当 $p\to 0$ 时的极限情况。

例 1.26. 在无限维线性空间 $C[a,b]$ 上定义不同的距离，空间的性质不一样。

定义 1.28. 序列的收敛性与一致收敛性

定义 1.29. 开集和闭集

定义 1.30. 稠密性与可分性

定义 1.31. 序列的极限与空间的完备性

定义 1.32. 紧性

1.4.3 内积空间的定义与性质

同三维空间相比，向量空间的长度和夹角等度量需要进行推广。在线性空间中引入内积结构，就得到内积空间。

- 内积空间，常见内积空间，内积空间的性质，向量长度，单位向量，常用不等式，*Hermite* 矩阵
- 标准正交基与正交化，正交投影，最佳逼近的定义、特征、存在性、构造
- 最佳逼近的应用实例

定义 1.33. 内积空间　内积空间 $\{V, P; +, \cdot, <>\}$ = 线性空间 + 内积运算 $< \cdot, \cdot >$

定理 1.9. 内积空间的性质

例 1.27. 常见的内积空间

定义 1.34. 向量的长度、单位向量　向量空间 V，其上元素 $\alpha \in V$，则向量的长度为 $||\alpha|| = \sqrt{< \alpha, \alpha >}$；进行单位化，得 $u = \frac{\alpha}{||\alpha||}$.

注 1.16. 向量的长度与定义的内积有关，这里的长度的推广就是范数；向量的单位化在机器学习里经常作为一个数据的预处理，也叫归一化 *(normalization)*，根据不同的范数，就有不同的单位向量. 在深度学习里，有很多不同的规范化策略 *[]*。

定理 1.10. 向量长度的性质

定义 1.35. 长度导出距离 $d(\alpha,\beta)=||\alpha-\beta||$

注 1.17. 内积空间按照此距离成为度量空间，进一步可以讨论极限、开集、闭集等拓扑特性。

定义 1.36. 向量夹角、向量正交 $\theta=\angle(\alpha,\beta)=\arccos\frac{<\alpha,\beta>}{||\alpha||\,||\beta||}$. 如果 $<\alpha,\beta>=0$, 则两向量正交。

练习 1.4. 考虑线性空间 $P_2[x]$, 定义内积如下：$\forall f(x),g(x)\in P_2[x],<f,g>=\int_0^1 f(x)g(x)dx$。设 $p_1(x)=1,p_2(x)=x,p_3(x)=x^2$,

(1) 求向量 $p_i,i=1,2,3$ 的长度；

(2) 求向量 $p_i,i=1,2,3$ 相互之间的夹角，并判断它们是否正交；

(3) 求向量 $p_i,i=1,2,3$ 相互之间的距离。

注 1.18. 向量的正交与投影在数据分析和机器学习中应用非常广泛。高维空间中的两个向量几乎都近似正交，如何利用这个特点？

课后思考与练习 1.2. 考虑两幅灰度图像（比如人脸、手写数字、建筑物或自然景色照片），如何计算它们各自的长度，两者之间的距离？给出实验验证结果。

定义 1.37. 度量矩阵、共轭转置矩阵 设 $\alpha_i,i=1,2,\ldots,n$ 是 n 维内积空间 V 的一组基, 对任意的向量 $\alpha,\beta\in V$, 有 $\alpha=\sum_{i=1}^n x_i\alpha_i,\beta=\sum_{i=1}^n y_i\alpha_i$, 则

$$<\alpha,\beta>=y^HAx$$

其中 $A=(a_{ij}=<\alpha_i,\alpha_j>)$ 为基 $\alpha_i,i=1,2,\ldots,n$ 的度量矩阵。如果基是标准正交基，则 A 退化为单位阵 E。

定义 1.38. 共轭转置矩阵 $A^H=(\bar{A})^T$ 称为矩阵 A 的共轭转置矩阵。

注 1.19. 在利用 *Matlab* 计算矩阵的转置时，提供了点转置操作，注意不要搞混了。

注 1.20. 在机器学习领域，如何学习两个样本间的度量，就是典型的度量学习问题 *(metric learning)* 以及近期基于深度学习的方法 *(deep metric learning).*

练习 1.5. 度量矩阵的计算及应用。定义 $P_2[x]$（实数域上次数不超过 *2* 的一元多项式集合构成的线性空间）中的内积 $< f, g >= \int_{-1}^{1} f(x)g(x)dx.$

(1) 求基 $1, x, x^2$ 的度量矩阵；

(2) 设 $f(x) = 1 - x + x^2, g(x) = 1 - 4x - 5x^2$, 求其内积。（直接法和间接利用度量矩阵来求）

定理 1.11. 共轭转置运算的性质

定义 1.39. *Hermite* 矩阵

1.4.4 标准正交基及正交化

定义 1.40. 正交向量组、标准正交向量组

定义 1.41. *Gram* 矩阵

定理 1.12.

定义 1.42. 正交基与标准正交基

定理 1.13. 设 $\alpha_i, i = 1, 2, \ldots, n$ 是 n 维内积空间 V 的一组基，则可以找到 V 中的一组标准正交基 $\epsilon_i, i = 1, 2, \ldots, n$, 使得 *span*$(\alpha_i, i = 1, 2, \ldots, n)$ = *span*$(\epsilon_i, i = 1, 2, \ldots, n)$.

证明：利用归纳法对该 *Gram-Schmidt(G-S)* 标准正交化给出构造性证明。

令

$$\epsilon_1 = \frac{\alpha_1}{||\alpha_1||}$$

则 ϵ_1 是单位向量，且 $span(\epsilon_1) = span(\alpha_1)$.

现在假设求出标准正交基 $\epsilon_i, i = 1, 2, \ldots, m$, 且

$$span(\alpha_i, i = 1, 2, \ldots, r) = span(\epsilon_i, i = 1, 2, \ldots, r), r = 1, \ldots m$$

接着求 ϵ_{m+1}: 作向量

$$\beta_{m+1} = \alpha_{m+1} - \sum_{i=1}^{m} k_i \epsilon_i$$

其中 $k_i, i = 1, \ldots, m$ 是待定常数。我们选择

$$k_i =< \alpha_m, \epsilon_i >, i = 1, \ldots, m$$

则 $< \beta_{m+1}, \epsilon_i >= 0, i = 1, \ldots, m$. 因为 $\alpha_i, i = 1, \ldots, m+1$ 线性无关，且 $span(\alpha_i, i = 1, 2, \ldots, m) = span(\epsilon_i, i = 1, 2, \ldots, m)$, 所以 $\beta_{m+1} \neq 0$, 并且 $\epsilon_i, i = 1, 2, \ldots, m, \beta_{m+1}$ 是正交向量组。令

$$\epsilon_{m+1} = \frac{\beta_{m+1}}{||\beta_{m+1}||}$$

则 $\epsilon_i, i = 1, 2, \ldots, m+1$ 是标准正交向量组，且 $span(\alpha_i, i = 1, 2, \ldots, m+1) = span(\epsilon_i, i = 1, 2, \ldots, m+1)$.

由归纳法原理得证。□

注 1.21. 从上述过程可以看出，主要包括正交化和标准化 (单位化) 两个步骤。其中正交化的几何意义，可以通过二维平面上两个向量的例子来理解。假设 α_1, α_2 线性无关，根据上述过程有：

$$\epsilon_1 = \frac{\alpha_1}{||\alpha_1||} \tag{1}$$

$$\beta_2 = \alpha_2 - < \alpha_2, \epsilon_1 > \epsilon_1 \tag{2}$$

$$\epsilon_2 = \frac{\beta_2}{||\beta_2||} \tag{3}$$

其中式 (2) 可以写成 $\alpha_2 = \beta_2 + < \alpha_2, \epsilon_1 > \epsilon_1$, 相当于是对 α_2 沿 ϵ_1 方向做了一个正交分解，且有 $< \beta_2, \epsilon_1 >= 0$. 如图所示。这正是其几何意义。这个思想在计算主成分时也得到了体现。

注 1.22. 在标准正交基下，向量的坐标和内积有很简洁的表达式。

例 1.28. R^n 中向量组的单位正交化。将如下 R^4 中的向量组化成标准正交向量组。并将向量 $\alpha=(1,2,3,4)^T$ 用这组标准正交基进行表示。

$$\alpha_1=\begin{bmatrix}1\\1\\0\\0\end{bmatrix},\alpha_2=\begin{bmatrix}1\\0\\1\\0\end{bmatrix},\alpha_3=\begin{bmatrix}-1\\0\\0\\1\end{bmatrix},\alpha_4=\begin{bmatrix}1\\-1\\-1\\1\end{bmatrix}$$

例 1.29. 在 $P_2[x]$ 上定义内积 $<f,g>=\int_{-1}^{1}f(x)g(x)dx$, 求其一组标准正交基。

解: 确定 $1,x,x^2$ 是 $P_2[\mathbf{x}]$ 的一组无关向量，是一组基。对其应用格拉姆-施密特过程.

首先，令 $\alpha_1=1,\alpha_2=x,\alpha_3=x^2$. 我们有

$$\|\alpha_1\|^2=<\alpha_1,\alpha_1>=\int_{-1}^{1}1^2\mathrm{d}x=2$$

于是 $\|\alpha_1\|=\sqrt{2}$, 所以 $\overline{e_1=\frac{\alpha_1}{\|\alpha_1\|}=\sqrt{\frac{1}{2}}}$.

现在 e_2 的表达式中的分子为

$$\beta_2=\alpha_2-\langle\alpha_2,e_1\rangle e_1=x-\left(\int_{-1}^{1}x\sqrt{\frac{1}{2}}\mathrm{d}x\right)\sqrt{\frac{1}{2}}=x$$

我们有

$$\|\beta_2\|^2=<\beta_2,\beta_2>=\int_{-1}^{1}x^2\mathrm{d}x=\frac{2}{3}$$

于是 $\|\beta_2\|=\sqrt{\frac{2}{3}}$, 所以 $\overline{e_2=\sqrt{\frac{3}{2}}x}$.

而 e_3 表达式中的分子为

$$\begin{aligned}\beta_3&=\alpha_3-\langle\alpha_3,e_1\rangle e_1-\langle\alpha_3,e_2\rangle e_2\\&=x^2-\langle x^2,e_1\rangle e_1-\langle x^2,e_2\rangle e_2\\&=x^2-\left(\int_{-1}^{1}x^2\sqrt{\tfrac{1}{2}}\mathrm{d}x\right)\sqrt{\tfrac{1}{2}}-\left(\int_{-1}^{1}x^2\sqrt{\tfrac{3}{2}}x\mathrm{d}x\right)\sqrt{\tfrac{3}{2}}x=x^2-\tfrac{1}{3}\end{aligned}$$

我们有

$$\|\beta_3\|^2 =< \beta_3, \beta_3 >= \int_{-1}^{1} \left(x^4 - \frac{2}{3}x^2 + \frac{1}{9}\right) \mathrm{d}x = \frac{8}{45}$$

于是 $\|\beta_3\| = \sqrt{\frac{8}{45}}$, 所以 $\underline{e_3 = \sqrt{\frac{45}{8}}\left(x^2 - \frac{1}{3}\right)}$.

因此

$$\underline{\sqrt{\frac{1}{2}}, \sqrt{\frac{3}{2}}x, \sqrt{\frac{45}{8}}\left(x^2 - \frac{1}{3}\right)}$$

是 $P_2[x]$ 中长度为 *3* 的规范正交组，是 $P_2[x]$ 的规范正交基.

练习 1.6. 接上题，

- 尝试求出向量 $f(x) = x^2 + 2x + 3$ 在该规范正交基下的坐标；
- 改变次序，对结果的影响.

课后思考与练习 1.3. 给出实数域 R 上矩阵空间 $R^{2\times 2}$ 中一组线性无关的向量组，定义其上的内积，并将其进行标准正交化。

1.4.5 正交补与投影定理

定义 1.43. 向量与空间正交、空间与空间正交　设 V_1, V_2 是内积空间 V 的子空间. 向量 $\alpha \in V$, 如果对任意的 $\beta \in V_1$，都有 $< \alpha, \beta >= 0$, 则称 α 与子空间 V_1 正交，记为 $\alpha \perp V_1$。在此基础上定义两个子空间正交 $V_1 \perp V_2$, 如果 $< \alpha, \beta >= 0, \forall \alpha \in V_1, \forall \beta \in V_2$ 成立.

定理 1.14. 　内积空间中两个正交的子空间的和是直和。

定义 1.44. 正交和、正交补　线性空间两个正交的子空间的和就是两个子空间的正交和 $V_1 \perp V_2, V_1 \bigoplus V_2$ ；如果 $V_1 \subset V, V_1^{\perp} = \{\alpha \in V | \alpha \perp V_1\}$ 称为 V_1 的正交补。

例 1.30. 考虑二维空间中，x 轴为其子空间，那么与 y 轴平行的向量的全体就是 x 轴这个子空间的正交补。

定理 1.15. 内积空间 V 中的有限维子空间 V_1 存在唯一的正交补 $V_1^{\perp}$，使得 $V = V_1 \bigoplus V_1^{\perp}$。

课后思考与练习 1.4. 假设子空间 V_1 由有限的 m 个元素 $(x_1, x_2, x_3, \ldots, x_m)$ 构成，如何求其正交补？这个正交补的现实意义是啥？

定义 1.45. 正交投影　内积空间 V 及其元素 α, 其子空间 V_1 及元素 α_1, $\alpha_1 \in V_1 \subset V \ni \alpha$, 如果 $\exists \alpha_2 \perp V_1$ 使得 $\alpha = \alpha_1 + \alpha_2$ $(\alpha_2 \perp \alpha_1, \alpha_2 \in V)$，那么称 α_1 是 α 在 V_1 上的正交投影。

定理 1.16. 投影定理　内积空间中任意元素在其有限维子空间的正交投影存在且唯一。

定义 1.46. 最佳逼近　设 V_1 是内积空间的 V 的一个非空子集，给定向量 $\alpha \in V$, 如果存在 $\alpha_1 \in V_1$ 满足

$$||\alpha - \alpha_1|| = \inf_{\beta \in V_1} ||\alpha - \beta|| = d(\alpha, V_1) \tag{4}$$

则称 α_1 是 α 在 V_1 上的最佳逼近。

注 1.23. 从最佳逼近的定义可以看出，这是一个优化问题。后面我们在讨论的最佳逼近的实例中，可以从后续矩阵分析与最优化的角度来看这个问题。

定理 1.17. 最佳逼近的特征 *(充要条件)*　设 V_1 是内积空间的 V 的一个子集，给定向量 $\alpha \in V$, 则 α_1 是 α 在 V_1 上的最佳逼近的充要条件是 $(\beta = \alpha - \alpha_1) \perp V_1$.

注 1.24. 在二维空间（V）中点（α）直线（V_1）的距离或者三维空间 (V) 中的点 (α) 到平面（V_1）的距离是通过作垂线（$\beta = \alpha - \alpha_1$）完成。$\beta = \alpha - \alpha_1 (\alpha = \alpha_1 + \beta)$ 就是那个垂直向量（$\beta \perp \alpha_1$）。

定理 1.18.　最佳逼近存在且唯一，即正交投影。*(最佳逼近的存在性与构造)*

内积空间 V 中向量 α 在 m 维子空间 V_1最佳逼近的构造过程:
设 $\alpha_i, i=1,2,\ldots,m$ 是 V_1 的一组基，则对 $\beta \in V_1$ 有

$$\beta = \sum_{i=1}^{m} x_i \alpha_i$$

若 β 为 $\alpha \in V$ 在 V_1 中的最佳逼近，其充要条件为 $\alpha - \beta \perp V_1$, 即

$$\alpha - \beta \perp \alpha_i, i = 1,2,\ldots,m$$

从而有

$$< \alpha - \beta, \alpha_i > = < \alpha, \alpha_i > - \sum_{i=1}^{m} x_i < \alpha_j, \alpha_i > = 0, i = 1,2,\ldots,m$$

.

得到方程组

$$Ax = b$$

其中 $A = G(\alpha_1, \alpha_2, \ldots, \alpha_m), x = (x_1, x_2, \ldots, x_m)^T, b = (< \alpha, \alpha_1 >, \ldots, < \alpha, \alpha_m >)$, 解出 x, 即可得到 β. 特别地，如果选择 $\alpha_i, i = 1,2,\ldots,m$ 为标准正交基，则最佳逼近为 $\beta = \sum_{i=1}^{m} < \alpha, \alpha_i > \alpha_i$, 其中 $< \alpha, \alpha_i > = x_i, i = 1,\ldots,m$.

总结：求线性空间 V 中向量 α 在子空间 $V_1 \subset V$ 中的最佳逼近 α_1(三步曲)

(1) 确定 V 的子空间 V_1, 假设其维数为 m;

(2) 选择子空间 V_1 的一组基 $\{u_i, i = 1,\ldots,m\}$，并进行标准正交化得到 $\{e_i, i = 1,\ldots,m\}$;

(3) 求系数 $k_i = < \alpha, e_i >$，得逼近 $\alpha_1 = \sum_{i=1}^{m} < \alpha, e_i > e_i$。

定理 1.19. 最佳逼近定理 (唯一性） 设 V_1 是 n 维内积空间 V 的一个闭凸集，则 V 中任一向量在 V_1 上都有唯一的最佳逼近.

1.4.6 最佳逼近的应用实例

例 1.31. 最小二乘问题　这里从正交投影与最佳逼近的角度来看[10]。已知 y 与 $x_i, i=1,\dots,n$ 之间存在线性关系

$$y=\sum_{i=1}^{n}\lambda_i x_i$$

其中 $\lambda_i, i=1,\dots,n$ 是未知系数。为了确定这些系数，通过 $m(m\geq n)$ 次观测数据

1	2	$\cdots$	m
$x_1^{(1)}$	$x_1^{(2)}$	$\cdots$	$x_1^{(m)}$
$x_2^{(1)}$	$x_2^{(2)}$	$\cdots$	$x_2^{(m)}$
$\vdots$	$\vdots$		$\vdots$
$x_n^{(1)}$	$x_n^{(2)}$	$\cdots$	$x_n^{(m)}$
$y^{(1)}$	$y^{(2)}$	$\cdots$	$y^{(m)}$

按如下意义确定系数：设 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 使得

$$\min_{c_i\in\mathbf{P}}\sum_{j=1}^{m}\left|y^{(j)}-\sum_{i=1}^{n}c_i x_i^{(j)}\right|^2=\sum_{j=1}^{m}\left|y^{(j)}-\sum_{i=1}^{n}\lambda_i x_i^{(j)}\right|^2$$

记

$$a_i=\begin{bmatrix}x_i^{(1)}\\x_i^{(2)}\\\vdots\\x_i^{(m)}\end{bmatrix},i=1,\cdots,n;b=\begin{bmatrix}y^{(1)}\\y^{(2)}\\\vdots\\y^{(m)}\end{bmatrix},c=\begin{bmatrix}c_1\\c_2\\\vdots\\c_n\end{bmatrix},\lambda=\begin{bmatrix}\lambda_1\\\lambda_2\\\vdots\\\lambda_n\end{bmatrix}$$

有

$$\min_{c\in\mathbf{P}^n}\left\|b-\sum_{i=1}^{n}c_i a_i\right\|^2=\left\|b-\sum_{i=1}^{n}\lambda_i a_i\right\|^2$$

[10]在矩阵分析部分我们将从另外一个角度来看这个问题.

上式可以看作 b 在 $span(a_1, a_2, \ldots, a_n)$ 上的最佳逼近。记 $A = [a_1, a_2, \ldots, a_n], V_1 = span(a_1, a_2, \ldots, a_n)$, 则 $\alpha_1 = A\lambda \in V_1$ 就是 $\alpha = b \in P^m$ 在 $V_1 \subset P^m$ 上的最佳逼近。根据最佳逼近的充要条件，有 $\alpha - \alpha_1 \perp V_1$, 即有 $< b - A\lambda, a_i >= 0, i = 1, \ldots, n$. 从而得 $a_i^H(b - A\lambda) = 0, i = 1, \ldots, n$, 统一成矩阵形式即得 $A^H A\lambda = A^H b$.

例 1.32. 函数的最佳平方逼近　设 $f(x) \in \mathbf{C}[a, b], \varphi_{\mathrm{i}}(x)\, \varphi_2(x), \cdots, \varphi_n(x)$ 是 $\mathbf{C}[a, b]$ 中线性无关的函数组, 求系数 $a_1, a_2, \cdots, a_n$ 使函数 $p(x) = \sum_{i=1}^n a_i\varphi_i(x)$ 逼近 $f(x)$ 时 $\int_a^b [f(x) - p(x)]^2 \mathrm{d}x$ 最小.

解：因为 $C[a\ b]$ 按例 *1.6.4* 定义的内积成为欧氏空间。令

$$V_1 = \operatorname{span}(\varphi_1(x), \varphi_2(x), \cdots, \varphi_n(x))$$

则 V_1 是 $C\,[a, b]$ 的一个 n 维子空间

$$\|f - p\|^2 = \int_a^b [f(x) - p(x)]^2 \mathrm{d}x$$

问题转化为求 $C[a, b]$ 中向量 $f(x)$ 在子空间 V_1 上逼近 $p(x)$. 记 $A = [a_{ij} =< \phi_i, \phi_j >], x = [a_1, a_2, \ldots, a_n]^T, b = [b_i =< f, \phi_i >, i = 1, 2, \ldots, n]$, 解方程组 $Ax = b$ 即得。在实际问题中，通常选择子空间 $V_1 = P_n[x]$(次数不超过 n 的一元多项式), 并构建其一组标准正交基 $\phi_i, i = 1, 2, \ldots, n$, 下面给出两个具体的实例。

例 1.33. 函数的最佳平方逼近（续）实例 *1*　在 $[0, 1]$ 上，求 $f(x) = e^x$ 的最优线性逼近。*(*考虑 *Taylor* 展开，并进行对比*)*

解：令 $V_1 = P_1[x] \subset C[0, 1]$, 取 V_1 的一组基：$\phi_1 = 1, \phi_2 = x$, 并进行标准正交化。

$< \phi_1, \phi_1 >= \int_0^1 1 \cdot 1dx = 1, u_1 = \frac{\phi_1}{||\phi_1||} = 1,$

$\beta_2 = \phi_2 - < \phi_2, u_1 > u_1 = x - \frac{1}{2}, u_2 = \frac{\beta_2}{||\beta_2||} = \sqrt{12}(x - \frac{1}{2}),$

求系数

$a_1 =< 1, f(x) >= \int_0^1 1 \cdot e^x dx = e - 1,$

$a_2 =< 2, f(x) >= \int_0^1 u_2 \cdot e^x dx = \sqrt{3}(3 - e),$

得逼近函数
$p(x) = \sum_{i=1}^{2} a_i u_i = (4e - 10) + 6(3 - e)$。

例 1.34. 函数的最佳平方逼近（续）实例 *2*　在 $[-\pi, \pi]$ 上，寻找函数 $f(x) = \sin x$ 的最佳逼近[11]。（提示：需要先确定 $P_n[x]$ 中的 n，然后选择其中一组基进行，还可以考虑对比不同的 n 对结果的影响）

练习 1.7. 将以上实例中, 令 $f(x) = \cos x$，求其最佳逼近。

1.5　拓展阅读与思考

由内积可以诱导范数，进而可以讨论赋范线性空间，在此基础上可以讨论空间的完备化问题。

完备的内积空间是 *Hilbert* 空间；完备的赋范线性空间是 *Banahe* 空间。

从有限维空间进入到无限维空间，进入泛函分析的殿堂。

在后续的第二章引入线性映射之后，加上内积空间的标准正交基，就可以引入对偶空间 *(Dual space)* 的概念了，进一步可以引入多重线线性映射与张量的概念。

文献评注

有关线性空间的发展简史

代数学的发展历史

几何学的发展历史

分析学的发展历史

[11] 参考 http://linear.axler.net/MinimizationProblems.pdf