陆 恺 田蔚风 编

最优估计理论及其在导航中应用

上海交通大学出版社

最优估计理论及其在导航中应用

陆 恺　田蔚风　编

上 海 交 通 大 学 出 版 社

内 容 简 介

本书是船舶工业总公司船舶类专业教材。全书共分九章，第一至第七章主要介绍最优估计原理，包括绪论，估计理论的数学基础，随机线性动力学系统的有关知识，基本估计方法，离散线性系统的最优滤波和预测，连续线性系统的最优滤波和预测，最优估计的误差分析和稳定性问题，连续和离散线性系统的最优平滑，自适应滤波和解决非线性滤波的几种近似方法等。第八章简要介绍了几种主要导航方法的原理。第九章着重阐述最优估计理论在组合导航和惯性导航中的应用。本书可作为惯性技术及导航设备专业高年级学生和研究生教材，亦可作为控制类专业的教学参考书，还可供有关工程技术人员和高等院校师生参考。

最优估计理论及其在导航中应用

出　版：上海交通大学出版社
（淮海中路1984弄10号）
发　行：新华书店上海发行所
印　刷：常熟市文化印刷厂
开　本：787×1092(毫米)　1/16
印　张：17
字　数：416000
版　次：1990年7月　第1版
印　次：1990年8月　第1次
印　数：1—1000
科　目：223—317
ISBN7-313-00669-1/U·66
定 价：3.30 元

前　言

在科学研究和工程实践中经常遇到下述问题：即如何从一个物理系统的输出中尽可能精确地提取系统状态的真实信息。通常称这一信息提取过程为对系统状态的估计。根据系统的动力学特性和作用在系统上的随机干扰，量测中的随机噪声以及系统随机初始状态等的统计特征，按某种性能指标达到最优的原则，实现系统状态的最优估计是对系统进行最优控制不可缺少的一步。因而最优估计理论是现代控制理论的一个重要分支。它在导航、制导、姿态控制、飞机和宇宙飞船飞行后的数据处理，以及大规模生产过程和化工过程的控制等方面都得到了广泛应用，所以它是导航和控制类专业学生的知识结构中不可缺少的组成部分。这也是本书作为船舶类专业教材的原因。本书按船舶类专业教材编审委员会所审定的大纲进行编写，全书共分九章。第一章为绪论，简要地阐述了最优估计理论的发展史和最小方差估计的物理概念。由于本书的内容是以一定的数学水平为基础的，因此读者必须具备矩阵分析、常微分方程、概率论等方面的知识，所以在本书的第二章中对估计理论所涉及的必要的数学知识进行了回顾。这对已经具备上述知识的读者是一个复习，对不具备上述数学知识的读者通过这一章的学习，可以掌握阅读本书必要的数学工具。本书的第三章介绍了估计理论所涉及的线性动力学系统方面的知识，其中包括连续和离散线性系统的状态方程及其解，系统的能观性和能控性，具有随机干扰的线性动力学系统，高斯马尔可夫过程和序列，几种常见的噪声模型等。在本书的第四章中系统地介绍了几种常用的统计估计方法，推导了离散线性系统最优滤波和最优预测的算式，讨论了系统噪声和观测噪声在同一时刻相关、系统噪声和观测噪声均为有色噪声时的滤波方程，以及用多个观测器观测同一物理过程的滤波方法。第五章阐述了高斯马尔可夫过程模型及其化为等效离散时间系统时的表达式，推导了连续线性系统的最优滤波和预测的方程式，介绍了求解利卡蒂矩阵微分方程的方法，还讨论了最优估计的误差分析和稳定性问题。在本章的最后一部分，从维纳-霍夫积分方程出发推导了连续线性系统的最优滤波方程，介绍了维纳滤波和卡尔曼滤波的异同。第六章介绍了固定区间平滑，固定点平滑和固定滞后平滑等三种平滑估计的含义，推导了三种平滑估计的算式。第七章讨论了滤波的发散现象，介绍了自适应滤波原理，列举了自适应滤波两种常用方法，并扼要地介绍了几种解决非线性滤波的近似方法。

为了阐述最优滤波理论在导航中应用，必须先对几种主要导航方法进行扼要介绍，因此在本书第八章中对推算船位、天文导航、惯性导航、无线电导航等导航方法从原理上作了简要阐述。本书第九章则着重介绍最优估计理论在组合导航中的应用，诸如：推算船位/劳兰C/卫导组合导航系统，推算船位/奥米加组合导航系统，GPS/惯性组合导航系统，惯性/天文组合导航系统，惯性/位置信息组合导航系统，以及在惯性导航初始对准和综合校正中应用等。本书由上海交通大学陆恺主编并撰写了第一至第七章，上海交通大学田蔚风撰写了第八和第九章并绘制全部插图。哈尔滨工业大学胡恒章教授、田自耘副教授对本书进行了仔细审阅并提出了不少有益的意见，编者对他们表示诚挚的感谢。在本书的编写过程中得到上海交通大学惯性技术与检测系统研究室全体同志的关心和协助，对此编者表示深切的谢

意。限于水平，缺点错误在所难免，敬请读者不吝指正俾便改进。

编 者

1989 年 12 月于上海交通大学

目　录

第一章 绪 论

§1-1 历史回顾

关于随机观测量的处理方法可以追溯到高斯(C.F.Gauss)，在1800年左右，他发明了最小二乘法，并用于天体轨道测量数据的处理中。大约100年之后，从事概率密度函数研究的费歇(R.A.Fisher)引入了极大似然估计方法，在估计理论方面作出了又一个重要贡献。

在1940年左右，维纳(N.Wiener)利用随机过程理论提出了一个在频率域设计最优滤波器的方法。此方法局限于统计平稳过程，并只提供稳态的最优估值。在同一时期内柯尔莫哥洛夫(A.N.Kolmogorov)处理了离散时间问题。

于1960年前后，卡尔曼(R.E.Kalman)和布西(R.S.Bucy)等人从时间域的状态方程出发提出了最优递推估计。这种方法通常称之为卡尔曼滤波。由于它适于用计算机来实现，因此随着计算机技术的发展得到了广泛运用。值得注意的是，卡尔曼滤波实质上是高斯最小二乘法进一步发展。在后面章节中可以看到，离散线性系统卡尔曼滤波公式与高斯最小平方估计的递推解具有类似的形式。

§1-2 最优估计概述

一个物理系统常常受到两个输入变量集的作用。一个是控制输入，通常它是确定性的。另一个是干扰输入，它是系统内部和外部的一些不能控制的随机干扰。例如：通讯系统中的天电干扰；电子线路中的噪声；惯性导航系统中的陀螺仪随机漂移等。一个物理系统的状态可以通过某些传感器观测得到，通常称这类传感器为观测器。由于观测器同样受到随机干扰的影响，使观测值产生误差。一个受控制作用的物理系统和用观测系统得到系统状态的观测输出的过程，如图1-1所示。估计问题的实质就是如何从观测值中较精确地确定系统状态的真实值及其变化过程。如果引入一个性能指标来评价估计的好坏，当估计值使性能指标取极值(极大或极小)时，该估计称之为最优估计。因此最优估计实质上是一种处理测量数据的算法。以后我们将看到，卡尔曼滤波就是利用物理系统和观测系统的动力学知识，系统噪声和观测噪声的统计特性，以及初始条件等信息对观测数据进行处理，从而得到系统状态的最小误差估计的一种算法。

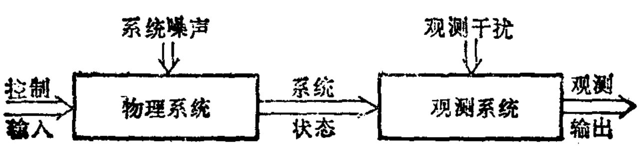

图1-1 物理系统和观测系统

根据获得状态最优估值的时间与获得观测值的时间的不同关系，有下面三种估计问题：

① 当得到状态估值的时间 τ 与最后观测时间 T 相重合时，称为滤波问题。② 当希望得到的状态估值的时间 τ 处于所得到的观测数据的时间间隔 $0\text{-}T$ 之间时，称为平滑问题。③ 当希望得到状态估值的时间 τ 在最后观测时间 T 之后时，称为预测问题。以上三种估计可以用图 1-2 示意。

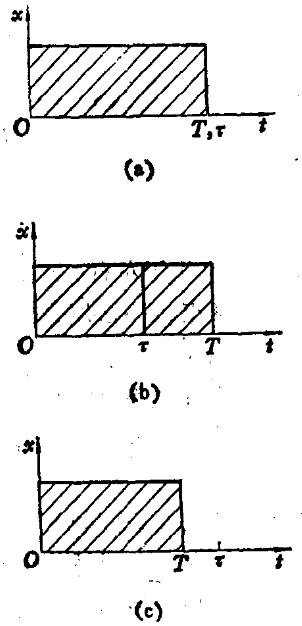

图 1-2 三种估计的示意图

(a) 滤波； (b) 平滑； (c) 预测

§1-3 最优估计理论的应用

最优估计理论具有广泛的应用。例如统计图像处理，交通密度估计，化学过程控制，河水流量估计，发电站负荷预测，卫星轨道估计等。下面用几个具体例子说明其应用情况。

一、估计理论在通讯系统中的应用

通讯过程中的一个重要问题是如何从接收信号中提取真实信息，图 1-3 表示通讯系统的框图。发射信号不仅包含信息而且包括发射过程中产生的误差。此外信号在传输过程中又受到讯道噪声的污染，如大气噪声、天电干扰等。因而，接收端的任务是要从接收信号中估计出发射端的真实信号。

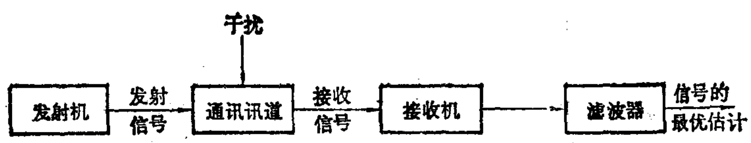

图 1-3 通讯系统示意图

二、估计理论在导航系统中的应用

导航系统中，利用最优估计理论可以得到导航参数的最优估值。以惯性导航为例，通过陀螺仪和加速度计等敏感元件测量运载体的加速度和角位移，经过计算，得到运载体在某一导航坐标系中的位置、速度和角速度，以及运载体的姿态等信息，这些信息对飞机、宇宙飞船的飞行过程和水面舰船、水下潜艇的航行过程的导航是十分重要的。图 1-4 简要地表示出惯性导航系统的工作过程。

三、估计理论在验后数据分析中的应用

当一个试验完成后，为了判断试验结果通常须对试验过程中所得到的数据进行处理。由于测量数据本身含有误差，因此这类验后数据处理就要使用估计理论。例如对一个空间飞行器从发射到进入轨道以及在轨道上运行期间所得到的观测数据进行处理，以期得到真实的飞行路径，从而判定制导系统的误差、进入轨道的精度，以发现系统内部的缺点，用以改

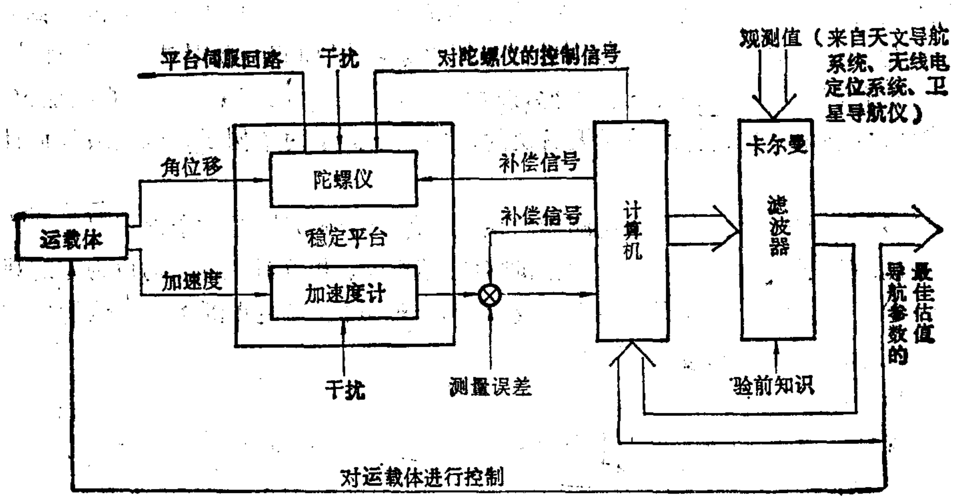

图 1-4　惯导系统工作过程示意图

善以后的飞行试验。而这些遥测数据的处理就要使用最优估计理论。

四、估计理论在过程控制中应用

一个复杂的过程或者大规模的系统在其运行过程中为了提高效率或者提高产品质量、降低原材料的消耗，往往需要进行控制。通常对运行过程中的测量数据进行处理以便得到状态估计，然后把这一信息提供给具有一定计算方法的控制器，经过计算，输出一个控制量施加给系统。其结果能有效地抵销系统内部或外部的干扰，使系统按预期的状态运行。

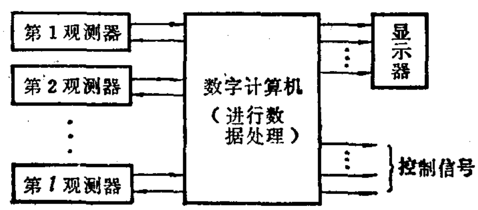

图 1-5　多观测器系统

估计问题可以由单个观测器观测单个过程，也可以由多个观测器观测单个过程或多个过程。后者称之为多观测器系统，如图 1-5 所示。假设有 l 个观测器对 m 个物理过程提供观测量，某些观测器提供相同的观测量，在这种情况下得到冗余观测值；另外一些观测器则提供与过程间接有关的量。在这种多观测器情况下，估计的任务是如何处理这些观测器的输出，以便得到对过程的最好估计。根据一定的算法来处理观测数据的任务是由计算机来完成，所得的最优估值通过显示器加以显示，也可以作为控制信号对系统进行控制。

多观测器系统的最突出例子为综合导航系统。同一导航过程其导航参数如位置、速度等可以由几个观测器给出，如用罗经和计程仪可以得到推算船位和航向、船速；利用惯性导航系统可以得到船位、航向和船速；利用天文导航、卫星导航以及其他无线电定位系统可以间断或连续地得到船位。惯性导航系统能连续给出导航参数，其缺点为由于陀螺漂移的不确定性导致定位误差随时间而积累。修正惯性导航系统误差的最简单的办法是修改惯性系统的位置指示，使之与其他导航系统所得到的位置观测值相一致。但这样做忽视了两个必须注意之点：第一，利用其他导航系统所得到的位置观测值包含着随机误差，这些误差与惯导系统的误差相比可能是不可忽视的。第二，惯导系统的误差源主要是陀螺仪的随机漂移，

而陀螺仪随机漂移的统计特性往往通过大量测试以后在一定程度上是已知的。因此如何把外部观测值和惯导系统所提供的数据进行合理的综合，从而得到舰船位置的最优估值，亦即所得到的船位与惯导系统的输出值相比，或与外部观测值相比都更接近于真实船位，这就是在导航系统中利用最优估计理论所要解决的问题。下面举两个例子来阐述最优估计的物理实质。

例 1-1 考虑一个含有两个观测器的系统，每一个观测器对一个恒定但未知的量 x 进行一次观测，观测值为 $z_i(i=1, 2)$，观测时存在随机的、相互独立的、均值为零的观测误差 $v_i(i=1, 2)$。试设计一种数据处理方法，组合两个观测值，以便得到 x 的无偏最优估值。

解：根据题意观测值 z_1、z_2 应为

$$z_1 = x + v_1, \qquad z_2 = x + v_2。 \tag{1-1}$$

在没有别的信息情况下，我们可以假定 x 的估计值应为两个观测值的线性函数，并以 $\hat{x}$ 表示 x 的估计值，则

$$\hat{x} = k_1 z_1 + k_2 z_2, \tag{1-2}$$

式中，k_1、k_2 为待定系数。估值误差定义为 $\tilde{x} = x - \hat{x}$，我们取 $\tilde{x}$ 的均方值为最小作为最优判据，而且选择的 k_1、k_2 与 x 值无关。根据估计值为无偏的要求可得：

$$E(\tilde{x}) = E[x - k_1(x+v_1) - k_2(x+v_2)] = 0。 \tag{1-3}$$

由于 $E(v_1) = E(v_2) = 0$，$E(x) = E(\hat{x})$，由式(1-3)得

$$k_2 = 1 - k_1。 \tag{1-4}$$

以式(1-4)代入 $\tilde{x} = x - \hat{x}$，求 $E(\tilde{x}^2)$，得

$$\begin{aligned} E(\tilde{x}^2) &= E\{[x - k_1(x+v_1) - (1-k_1)(x+v_2)]^2\} = E\{[k_1 v_1 + (1-k_1)v_2]^2\} \\ &= E[k_1^2 v_1^2 + (1-k_1)^2 v_2^2 + 2k_1(1-k_1)v_1 v_2]。 \end{aligned}$$

假定 v_1、v_2 的方差为 $E(v_1^2) = \sigma_1^2$，$E(v_2^2) = \sigma_2^2$，根据题意 v_1、v_2 相互独立，即 $E(v_1 v_2) = 0$，故得

$$E(\tilde{x}^2) = k_1^2\sigma_1^2 + (1-k_1)^2\sigma_1^2。$$

适当地选择 k_1，可使 $E(\tilde{x}^2)$ 为最小。为此令 $E(\tilde{x}^2)$ 对 k_1 的导数为零，即

$$\frac{\mathrm{d}E(\tilde{x}^2)}{\mathrm{d}k_1} = 2k_1\sigma_1^2 - 2(1-k_1)\sigma_2^2 = 0。$$

由此求得 $k_1 = \sigma_2^2/(\sigma_1^2 + \sigma_2^2)$。于是 x 的最优估值为

$$\hat{x} = \left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right) z_1 + \left(\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\right) z_2,$$

对应的最小估值误差均方值为

$$E(\tilde{x}^2) = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^{-1}。$$

从题中可以看出，当缺乏其他信息并只能从观测值确定 x 时，x 的最优估值 $\hat{x}$ 应为观测值 z_1、z_2 的加权平均，加权系数 k_1、k_2 分别与观测误差的均方值成反比，即观测误差的均方值愈大，该观测值的加权系数愈小。反之，观测误差的均方值愈小，加权系数愈大。如果 $\sigma_1^2 = 0$，则 $\hat{x} = z_1$；如果 $\sigma_1^2 = \sigma_2^2$，则 $\hat{x} = (z_1 + z_2)/2$。在以后的讨论中，我们将看到对所获得的系统状态各种信息进行加权处理，来求得系统状态的最优估值，这是最优估计理论的基本思想。

例 1-2 在上例中如果观测误差彼此相关，并有 $E(v_1 v_2) = \rho\sigma_1\sigma_2$，$\rho$ 为相关系数，$|\rho| \leqslant$

1。试证明最优估计的加权系数为

$$k=\frac{\sigma_2^2-\rho\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2},$$

估计误差的均方值为

$$E(\tilde{x}^2)=\frac{\sigma_1^2\sigma_2^2(1-\rho^2)}{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}。$$

解： 仍令 $\hat{x}=k_1z_1+k_2z_2$。根据无偏估计的要求，得

$$\hat{x}=k_1z_1+(1-k_1)z_2。$$

令 $k_1=k$，则得：

$$\tilde{x}=k(x+v_1)+(1-k)(x+v_2)=kv_1+(1-k)v_2+x,$$

$$\tilde{x}=(k-1)v_2-kv_1,$$

$$E(\tilde{x}^2)=k^2\sigma_1^2+(1-k)^2\sigma_2^2+2k(1-k)\rho\sigma_1\sigma_2。\tag{1-5}$$

将式(1-5)对 k 求导数，并令其等于零，得

$$\frac{\mathrm{d}E(\tilde{x}^2)}{\mathrm{d}k}=2(\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2)k-2\sigma_2^2+2\rho\sigma_1\sigma_2=0。$$

由上式得

$$k=\frac{\sigma_2^2-\rho\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}。$$

将 k 的表达式代入式(1-5)，得

$$E(\tilde{x}^2)=\frac{\sigma_1^2\sigma_2^2(1-\rho^2)}{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}。$$

第二章　数学基础知识

本章主要介绍有关推导和应用现代估计理论一些主要公式时所常用的数学知识。包括向量和矩阵的运算，随机变量和随机过程。本章的取材并不是对以上内容作严格而详尽的推导和叙述，而是认为读者基本上已具备上述知识，只是对估计理论常常用到的这部分内容作简单的回顾。至于更详细的介绍，可以参阅有关文献。

§2-1　向量与矩阵

一、向量运算

定义：一组元素排成一列，用 $\boldsymbol{x}$ 表示，称之为向量。向量中所包含元素的数目称为向量的维数。这样一个 n 维向量 $\boldsymbol{x}$ 可写为

$$\boldsymbol{x}=\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}。\tag{2-1}$$

1. 向量加法

两个向量相加，定义为

$$\boldsymbol{x}+\boldsymbol{y}=\begin{pmatrix} x_1+y_1 \\ x_2+y_2 \\ \vdots \\ x_n+y_n \end{pmatrix}。\tag{2-2}$$

向量相减的定义与相加的相类似。在向量的相加或相减时，两个向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 必须有相同维数，而且 $x_1, x_2, \cdots, x_n, y_1, y_2, \cdots, y_n$ 为两向量在同一组基底上的分量。

2. 标量乘向量

一个向量乘以标量 k，则其每一元素均乘以 k，即

$$k\boldsymbol{x}=\begin{pmatrix} kx_1 \\ kx_2 \\ \vdots \\ kx_n \end{pmatrix}。\tag{2-3}$$

3. 零向量与向量的转置

每一个元素均为零的向量称为零向量，用 $\mathbf{0}$ 表示。向量 $\boldsymbol{x}$ 的转置用 $\boldsymbol{x}^{\mathrm{T}}$ 表示。向量转置后其元素由排成一列变为排成一行，亦即

$$\boldsymbol{x}=\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

$$\boldsymbol{x}^{\mathrm{T}}=(x_1,\ x_2,\ \cdots,\ x_n)。\tag{2-4}$$

4．向量的内积

具有相同维数的向量 $\boldsymbol{x}$，$\boldsymbol{y}$ 的内积或称点积定义为：

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=x_1y_1+x_2y_2+\cdots+x_ny_n,\tag{2-5}$$

其结果为一标量。如果

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=0,\tag{2-6}$$

则称 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 正交。向量 $\boldsymbol{x}$ 的长度的平方定义为：

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}=x_1^2+x_2^2+\cdots+x_n^2,\tag{2-7}$$

向量 $\boldsymbol{x}$ 的长度定义为：

$$|\boldsymbol{x}|=\sqrt{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}。\tag{2-8}$$

5．向量的外积

同维数的向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的外积定义为

$$\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}=\begin{pmatrix} x_1y_1 & x_1y_2\cdots x_1y_n \\ x_2y_1 & x_2y_2\cdots x_2y_n \\ & \cdots\cdots \\ x_ny_1 & x_ny_2\cdots x_ny_n \end{pmatrix},\tag{2-9}$$

其结果为一矩阵。类似地可以得到 $\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}$

$$\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}=\begin{pmatrix} x_1^2 & x_1x_2\cdots x_1x_n \\ x_2x_1 & x_2^2\ \ \cdots x_2x_n \\ & \cdots\cdots \\ x_nx_1 & x_nx_2\cdots x_n^2 \end{pmatrix},\tag{2-10}$$

$\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}$ 被称为向量 $\boldsymbol{x}$ 的扩散矩阵。

6．向量的导数

对于随时间连续变化的向量 $\boldsymbol{x}(t)$ 和 $\boldsymbol{x}(t+\Delta t)$，我们有：

$$\boldsymbol{x}(t+\Delta t)-\boldsymbol{x}(t)=\begin{pmatrix} x_1(t+\Delta t)-x_1(t) \\ x_2(t+\Delta t)-x_2(t) \\ \vdots \\ x_n(t+\Delta t)-x_n(t) \end{pmatrix},$$

用标量 $\dfrac{1}{\Delta t}$ 乘等号两边并取极限 $\Delta t\to 0$ 记为 $\dot{\boldsymbol{x}}(t)$，则有

$$\dot{\boldsymbol{x}}(t)=\begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \vdots \\ \dot{x}_n(t) \end{pmatrix},\tag{2-11}$$

即对向量求导为对其元素求导。同理，向量的积分可用其元素的积分表示：

$$\int\boldsymbol{x}(t)\mathrm{d}t=\begin{pmatrix} \int x_1\mathrm{d}t \\ \int x_2\mathrm{d}t \\ \vdots \\ \int x_n\mathrm{d}t \end{pmatrix}。\tag{2-12}$$

二、矩阵运算

$n\times m$个实数a_{ij}，$i=1, 2, \cdots, n$，$j=1, 2, \cdots, m$，作如下排列：

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ & \cdots\cdots & & \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}, \tag{2-13}$$

称为n行m列矩阵，或简称$n\times m$矩阵，并用A表示。有时简记$A=[a_{ij}]$。a_{ij}称为矩阵A的第i行第j列的元素。元素全为零的$n\times m$称为零矩阵，并以0表示。一个n维向量x

$$x=\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

可以看作n行一列的列矩阵。同理，一个n维的行向量可以看作n列一行的行矩阵。一个标量可以看作一行一列的矩阵。行列相等的矩阵称为方阵。行列数均为n的$n\times n$矩阵亦可称为n阶方阵。一个方阵若其非对角线元素均为零，称之为对角阵。

1．矩阵加法

两个$n\times m$矩阵$A=[a_{ij}]$，$B=[b_{ij}]$的和$A+B$定义为对应元素相加而成的$n\times m$矩阵：

$$A+B\overset{\text{def}}{=}[a_{ij}+b_{ij}]=\begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1m}+b_{1m} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2m}+b_{2m} \\ & \cdots\cdots\cdots\cdots\cdots\cdots & & \\ a_{n1}+b_{n1} & a_{n2}+b_{n2} & \cdots & a_{nm}+b_{nm} \end{pmatrix}。 \tag{2-14}$$

同理，两个$n\times m$矩阵相减定义为

$$A-B\overset{\text{def}}{=\!=}[a_{ij}-b_{ij}]=\begin{pmatrix} a_{11}-b_{11} & a_{12}-b_{12} & \cdots & a_{1m}-b_{1m} \\ a_{21}-b_{21} & a_{22}-b_{22} & \cdots & a_{2m}-b_{2m} \\ & \cdots\cdots\cdots\cdots\cdots\cdots & & \\ a_{n1}-b_{n1} & a_{n2}-b_{n2} & \cdots & a_{nm}-b_{nm} \end{pmatrix}。 \tag{2-15}$$

两个矩阵相等，当且仅当它们的所有对应元素相等。这样$A=B$，亦即$a_{ij}=b_{ij}$。矩阵加法满足结合律和交换律：

$$(A+B)+C=A+(B+C),\ A+B=B+A。$$

2．矩阵乘标量

一个矩阵A乘以标量k，结果用kA表示：

$$kA=[ka_{ij}]。 \tag{2-16}$$

3．矩阵乘法

$n\times m$矩阵$A=[a_{ij}]$和$m\times l$矩阵$B=[b_{jk}]$的积AB定义为$n\times l$矩阵并以C表示：

$$AB=C\overset{\text{def}}{=\!=}\left[\sum_{j=1}^{m}a_{ij}b_{jk}\right]=[C_{ik}]。 \tag{2-17}$$

乘积AB只有A, B可相乘时才有意义，即要求A的列数等于B的行数。矩阵乘法一般不满足交换律，即

$$AB\neq BA。$$

因为 AB 有意义时，BA 不一定有意义，前者要求 A 的列数等于 B 的行数，并不要求 B 的列数等于 A 的行数。即使这一要求得到满足，AB 与 BA 不一定能进行比较。例如 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 都是 n 维向量时，

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=\sum_{i=1}^{n}x_iy_i$$

是标量，而 $\boldsymbol{y}\boldsymbol{x}^{\mathrm{T}}=[y_ix_j]$ 是一个 $n\times n$ 矩阵。而且即使 A 和 B 均为 $n\times n$ 方阵时，乘法交换律一般仍不能成立。例如

$$A=\begin{pmatrix}1 & 2\\ 0 & 1\end{pmatrix},\qquad B=\begin{pmatrix}1 & 0\\ 3 & 0\end{pmatrix},$$

则

$$AB=\begin{pmatrix}7 & 0\\ 3 & 0\end{pmatrix},\qquad BA=\begin{pmatrix}1 & 2\\ 3 & 6\end{pmatrix}。$$

矩阵乘法满足结合律

$$(\alpha\beta)A=\alpha(\beta A),\ (\alpha A)B=\alpha(AB),\ (AB)C=A(BC)。\tag{2-18}$$

对乘法和加法满足分配律

$$\left.\begin{aligned}&\alpha(A+B)=\alpha A+\alpha B,\ \ (\alpha+\beta)A=\alpha A+\beta A,\\ &A(B+C)=AB+AC\ \ (B+C)A=BA+CA。\end{aligned}\right\}\tag{2-19}$$

4．矩阵转置

把 $n\times m$ 矩阵 $A=[a_{ij}]$ 的行和列互换而成的矩阵称为 A 的转置，以 A^{T} 表示，它是一个 $n\times m$ 矩阵：

$$A^{\mathrm{T}}=\begin{pmatrix}a_{11} & a_{21} & \cdots & a_{n1}\\ a_{12} & a_{22} & \cdots & a_{n2}\\ & \cdots\cdots & \cdots\cdots & \\ a_{1m} & a_{2m} & \cdots & a_{nm}\end{pmatrix}。\tag{2-20}$$

不难证明矩阵转置具有下列关系：

$$\left.\begin{aligned}&(A+B)^{\mathrm{T}}=A^{\mathrm{T}}+B^{\mathrm{T}},\\ &(\alpha A)^{\mathrm{T}}=\alpha A^{\mathrm{T}},\\ &(AB)^{\mathrm{T}}=B^{\mathrm{T}}A^{\mathrm{T}}。\end{aligned}\right\}\tag{2-21}$$

5．矩阵的分块及其运算

在矩阵运算中，有时将其元素分块表示，然后进行运算，往往比较简便。例如将两个 $n\times m$ 矩阵 A 和 B 分成对应的四块：

$$A=\begin{pmatrix}A_{11} & A_{12}\\ A_{12} & A_{22}\end{pmatrix},\qquad B=\begin{pmatrix}B_{11} & B_{12}\\ B_{21} & B_{22}\end{pmatrix},$$

式中，A_{ij}，B_{ij} 为 $n_i\times m_j$ 矩阵；$i=1,\ 2$；$j=1,\ 2$；$n_1+n_2=n$；$m_1+m_2=m$，则有

$$A^{\mathrm{T}}=\begin{pmatrix}A_{11}^{\mathrm{T}} & A_{21}^{\mathrm{T}}\\ A_{12}^{\mathrm{T}} & A_{22}^{\mathrm{T}}\end{pmatrix},\qquad A+B=\begin{pmatrix}A_{11}+B_{11} & A_{12}+B_{12}\\ A_{21}+B_{21} & A_{22}+B_{22}\end{pmatrix},$$

$$\alpha A=\begin{pmatrix}\alpha A_{11} & \alpha A_{12}\\ \alpha A_{21} & \alpha A_{22}\end{pmatrix}。\tag{2-22}$$

如有 $m\times l$ 矩阵 C，将 C 分成如下四块：

$$C=\begin{pmatrix}C_{11} & C_{12}\\ C_{21} & C_{22}\end{pmatrix},$$

，C_{ij} 为 $m_i\times l_j$ 矩阵；$i=1,\ 2$；$j=1,\ 2$；$m_1+m_2=m$；$l_1+l_2=l$。容易验证有如下的乘法

$$AC=\begin{pmatrix} A_{11}C_{11}+A_{12}C_{21} & A_{11}C_{12}+A_{12}C_{22} \\ A_{21}C_{11}+A_{22}C_{21} & A_{21}C_{12}+A_{22}C_{22} \end{pmatrix}。\tag{2-23}$$

6．方阵的行列式、迹和逆

一个主对角线元素皆为1，其除元素均为零的 n 阶方阵称为 n 阶单位阵，记作 I_n 或 I，

$$I\overset{\text{def}}{=\!=}\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ & & \cdots\cdots & \cdots & \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}。\tag{2-24}$$

由矩阵乘法规则易知，对任一 $n\times m$ 矩阵 A 有

$$I_nA=A,\quad AI_m=A。$$

$|A|$ 表示方阵 A 的行列式，$|A|$ 是一个标量，其定义为

$$|A|=\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \cdots & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

$$=\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\cdots\sum_{\substack{l=1\\ l\neq i,j,\cdots,k}}^{n}\pm a_{1i}a_{2j}\cdots a_{nl}。$$

在每一项中第二个下标 $i,\ j,\ \cdots,\ l$ 是求1，2，…，n 的排列，下标为偶数排列的项取正号，奇数排列的项取负号。不难证明它有以下一些性质：

$$|A^{\mathrm{T}}|=|A|,\ |\alpha A|=\alpha^n|A|,\ |AB|=|A||B|。\tag{2-25}$$

方阵 A 的迹定义为它的主对角线元素之和，以 $\mathrm{tr}A$ 表示，

$$\mathrm{tr}A\overset{\text{def}}{=\!=}\sum_{i=1}^{n}a_{ii}。\tag{2-26}$$

不难证明：

$$\mathrm{tr}A^{\mathrm{T}}=\mathrm{tr}A,\ \mathrm{tr}[A+B]=\mathrm{tr}A+\mathrm{tr}B,\ \mathrm{tr}(\lambda A)=\lambda\mathrm{tr}A。\tag{2-27}$$

又如果 A、B 分别为 $n\times m$ 和 $m\times n$ 矩阵，则

$$\mathrm{tr}(AB)\cdots\sum_{i=1}^{n}\sum_{j=1}^{m}a_{ij}b_{ji}=\sum_{j=1}^{m}\sum_{i=1}^{n}b_{ji}a_{ij}=\mathrm{tr}(BA)。\tag{2-28}$$

设 A 为 n 阶方阵，如果 $|A|=0$，则称 A 为奇异阵；否则称 A 为非奇异阵或可逆阵。对于非奇异阵 A 可定义 A 的逆矩阵 A^{-1} 为

$$A^{-1}\overset{\text{def}}{=\!=}[(-1)^{i+j}|A_{ji}|/|A|],\tag{2-29}$$

式中，$(-1)^{i+j}|A_{ji}|/|A|$ 为逆矩阵 A^{-1} 的第 i 行第 j 列的元素，而 A_{ji} 表示在 A 中去掉第 j 行和第 i 列后的 $n-1$ 阶方阵。由行列式的性质可以证明下列关系式成立：

$$AA^{-1}=I=A^{-1}A。\tag{2-30}$$

容易证明，满足上述方程的 A^{-1} 是唯一的。从上述关系式出发，不难证明以下性质：

$$\left.\begin{aligned} &|A^{-1}|=|A|^{-1},\ (A^{-1})^{\mathrm{T}}=(A^{\mathrm{T}})^{-1},\ (A^{-1})^{-1}=A, \\ &(\alpha A)^{-1}=\alpha^{-1}A^{-1},\ (AB)^{-1}=B^{-1}A^{-1}。\end{aligned}\right\}\tag{2-31}$$