GEO-SPATIAL INFORMATION SCIENCE

普通高等教育测绘类规划教材

广义测量平差

（新版）

崔希璋 於宗俦 陶本藻 刘大杰
于正林 孙海燕 王新洲 编著

WUHAN UNIVERSITY PRESS
武汉大学出版社

普通高等教育测绘类规划教材

广义测量平差

（新版）

崔希璋 於宗俦 陶本藻 刘大杰
于正林 孙海燕 王新洲 编著

WUHAN UNIVERSITY PRESS
武汉大学出版社

图书在版编目(CIP)数据

广义测量平差/崔希璋等编著.—新版.—武汉:武汉大学出版社.
2001.2
普通高等教育测绘类规划教材
ISBN 7-307-04830-2

Ⅰ.广… Ⅱ.崔…[等] Ⅲ.测量平差 Ⅳ.P207

中国版本图书馆 CIP 数据核字(2000)第 78699 号

责任编辑:龙方明 王金龙

出版发行:**武汉大学出版社** (430072 武昌 珞珈山)
(电子邮件:wdp4@whu.edu.cn 网址:www.wdp.com.cn)
印刷:安陆市鼎鑫印务有限责任公司
开本:787×1092 1/16 印张:11.25 字数:282 千字
版次:2001 年 2 月第 1 版 2005 年 11 月第 2 次印刷
ISBN 7-307-04830-2/P·75 定价:15.00 元

前　言

《广义测量平差》作为测绘专业研究生教材已近 20 年，1982 年 6 月第一次出版，1992 年 10 月经较大修改出版了第二版，至今仍作为许多高等院校测绘专业的研究生教材或参考书以及本科生高年级选修教材。根据教师和读者对教材内容的修改意见，考虑教学用书的需要，决定修订再版本书。

在一些学校的研究生教学计划中，将原开设的《广义测量平差》课程改名为《现代测量数据处理理论》，以原来的广义测量平差基本内容为主并进行某些扩充，本书的修订就考虑了这一变化因素。为了本书的连续性和共知性，我们没有更改书名，但书名后用括号加注“新版”字样，其内容可认为是《现代测量数据处理理论》课程的基本教材。

本版对第二版在教学体系上作了较大修改，并以少而精原则精选内容。修订主要之处是：将第二版的第 3 章秩亏自由网平差、第 5 章滤波与配置及第 6 章若干平差问题的补充等这三章所精选的教学内容合并为一章，取名为最小二乘平差的统一理论和方法；删除了第二版中第 1 章随机过程基础和附录等纯数学内容，删除了第二版的第 8 章卡尔曼滤波的渐近性质和误差分析；基本保留了第 2 章估计方法和广义测量平差原理、第 4 章平差随机模型的验后估计、第 7 章动态线性系统的卡尔曼滤波等三章内容；新增了稳健估计的基本理论和有偏估计两章。经上述修订，已将第二版中 8 章 53 节，缩成为 6 章 36 节，字数也大为减少。但仍保留了介绍近代测量平差理论和方法最新成果为主线的特色。

本书的第一版，由崔希璋主持，於宗俦、陶本藻、刘大杰参与编写；第二版由刘大杰、于正林主编，负责全面修订，於宗俦、陶本藻参编；本版由陶本藻主编、负责全面修订，孙海燕、王新洲参编，孙海燕编写了第 5 章，王新洲编写了第 6 章。刘经南参与了本书编写大纲的讨论，所提出的修订意见已反映在本版中。

《广义测量平差原理》第一版、第二版均由测绘出版社出版，经协商同意，本书改由武汉测绘科技大学出版社出版，我们对这两个出版社的领导和编辑的大力支持表示衷心感谢。

武汉测绘科技大学教务处和全国测绘类高等学校教学指导委员会对本书出版作了大量工作，我们在此也一并表示深切感谢。

希望能继续得到读者支持，不断提供改进意见。

编著者

2000 年 6 月于武汉

目　录

GENERALIZED SURVEYING ADJUSTMENT

Contents

第 1 章　估计方法和广义测量平差原理

1-1　概　述

在测量、通讯和控制等学科中，为了求得某些未知参数，常常要进行一系列的观测. 由于测量上的局限性，往往只能观测未知量的某些函数，且观测值中必然含有误差(或称为噪声). 这就产生了根据含有误差的观测值求定未知参数估值的问题. 下面举几个例子.

(1)为了确定平面或三维控制网中各点的坐标，对控制网的边长和方向进行了观测，当然，观测值包含有误差. 设各点的坐标为未知参数向量 X，而包括边长和方向的观测值向量为 L，则 L 和 X 之间有函数关系

$$L = F(X) + \Delta$$

式中 Δ 表示误差向量. 通过含有误差 Δ 的观测向量 L 来求定待定点坐标的最佳估值，就是一个估计问题. 在测量中，就是一个平差问题.

(2)通讯理论中的一个重要问题是从接收到的信号中，提取被发送的信号. 设被发送的信息调制成信号 $S(t)$，而接收到的信号也就是信号的观测值 $L(t)$，由于大气噪声和电路噪声的干扰，因此有

$$L(t) = S(t) + n(t)$$

其中 $n(t)$是噪声，t 表示时间. 通讯中的主要问题就是从 $L(t)$中将有用的信号 $S(t)$分离出来，也就是由 $L(t)$求定 $S(t)$的最佳估值. 信号 $S(t)$也是一种未知参数.

(3)生产过程的自动化可以达到高效率和高精度. 在实现生产过程的控制中，需要通过对生产系统进行状态的不断测量，得到与系统运行状态有关的观测值；然后对观测值进行分析处理得到控制信号，实时地控制生产系统按要求运行. 但由于观测值中存在误差，所以，为了得到控制信号，就要求由观测值来估计系统的运行状态.

(4)卫星(或其它运动体)的轨道往往可以由如下微分方程确定

$$X(t) = f(X(t),\ U(t),\ \Omega(t))$$

式中 t 表示时间；$X(t)$表示卫星的轨道参数，在此处称为状态向量；$U(t)$为控制向量；$\Omega(t)$是随机的状态噪声. 为了精确估计或预测卫星的轨道，就需要对卫星进行观测，从而得到大量的观测数据 $L(t)$，然后实时地由含有误差的观测值 $L(t)$来估计卫星的轨道，即估计卫星的轨道参数.

以上例中所述的信号或状态都可以说是一种未知参数. 在测量平差中，通常称非随机的未知参数向量为参数，而称随机参数向量为信号，而称随时间 t 变化的动态系统中的未知参数向量为状态向量，或简称为状态. 可以看到，在上面的例子中，都存在一个对未知参数进行估计的问题.

一般说来，若设 x 为 t 阶未知参数向量(简称为参数)，L 为 n 阶观测向量(或称观测值)，Δ 表示 n 维误差(或噪声)向量. 那么，所谓估计问题，就是根据含有误差 Δ 的观测值 L，构造

一个函数 $\hat{X}(L)$，使 $\hat{X}(L)$ 成为未知参数向量 X 的最佳估计量，其具体数值称为最佳估值(以后一般不区分其含义)．通常将 $\hat{X}(L)$ 简记为 $\hat{X}$，并记

$$\Delta_X = X - \hat{X}(L) = X - \hat{X}$$

称 Δ_X 为 $\hat{X}(L)$ 的估计误差．

可以看到，当 Δ_X 的数学期望等于零时，Δ_X 的方差就等于 $E(\Delta_X\Delta_X^T)$；而当 X 为非随机量时，未知参数的估值 $\hat{X}$ 的方差 $D_{\hat{X}}$ 也就等于其误差方差 $D(\Delta_X)$．在估计理论中，通常是用估计量 $\hat{X}$ 的误差方差 $D(\Delta_X)$ 来衡量其精度的．但在经典的最小二乘平差中，由于 X 一般都是非随机参数，所以习惯上都用估值(平差值)的方差衡量精度．

在根据观测值 L 求未知参数 X 的估值 $\hat{X}(L)$ 时，总是希望所得到的估值是最优的．由估计理论知道，最优估计量主要应具有以下几个性质：

(1)一致性．由观测值得到的估值 $\hat{X}(L)$ 通常与其真值是不同的，我们希望当观测值个数 n 增加时，估计量变得更好些；当 n 无限增大时，估计量向被估计的参数趋近的概率等于 1．即如果对于任意 $\varepsilon>0$，有

$$\lim_{n\to\infty} P(X - \varepsilon < \hat{X} < X + \varepsilon) = 1 \tag{1-1-1}$$

则称估计量 $\hat{X}$ 具有一致性；若有

$$\lim_{n\to\infty}((X - \hat{X})(X - \hat{X})^{\mathrm{T}}) = 0 \tag{1-1-2}$$

则称此估计量是均方一致的．估计量的一致性是从它的极限性质来看的．

(2)无偏性．若估计量 $\hat{X}$ 的数学期望等于被估计量 X 的数学期望，即

$$E(\hat{X}) = E(X) \tag{1-1-3}$$

如果 X 是非随机量，上式即为

$$E(\hat{X}) = X \tag{1-1-4}$$

则称 $\hat{X}$ 为无偏估计量．如果 $E(\hat{X})\to X(n\to\infty)$，则称 $\hat{X}$ 为渐近无偏．

(3)有效性．若由观测向量 L 得到无偏估计量 $\hat{X}$ 的误差方差 $E((X-\hat{X})(X-\hat{X})^{\mathrm{T}})$，小于由 L 得到的任何其它无偏估计量 X^* 的误差方差 $E((X-X^*)(X-X^*)^{\mathrm{T}})$，即

$$E((X - \hat{X})(X - \hat{X})^{\mathrm{T}}) < E((X - X^*)(X - X^*)^{\mathrm{T}})$$

或写为

$$D(\Delta_X) < D(\Delta_X{}^*) \tag{1-1-5}$$

则称 $\hat{X}$ 是有效估计量，也称 $\hat{X}$ 具有有效性或方差最小性．

以不同的准则来求定未知参数的最佳估值，可得到不同的估计方法．估计方法主要有极大似然估计，最小二乘估计，极大验后估计，最小方差估计和线性最小方差估计等；经典的测量平差法都是以最小二乘估计或极大似然估计为根据导出的；而滤波、配置和动态系统的卡尔曼滤波等，最初是以极大验后估计或最小方差估计为根据导出的．因此，概率统计中的估计理论是广义测量平差的理论基础．

1-2　多维正态分布

正态分布是测量平差理论中最常用的误差分布，是最小二乘平差误差理论的基础．本节在已学过一元正态分布基础上，对多维正态分布作全面阐述．广义测量平差理论中还牵涉到其它分布，则将分别在相应章节中一一介绍．

1. 多维正态分布的定义和性质

设有 m 个互相独立的标准正态随机变量构成的随机向量 $Z=[Z_1 \quad Z_2 \quad \cdots \quad Z_m]^{\mathrm{T}}$，则称它们的有限个线性函数

$$\underset{n\times 1}{X}=\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}=\underset{n\times m}{A}\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_m \end{bmatrix}+\underset{n\times 1}{\mu} \tag{1-2-1}$$

为 n 维正态随机向量.此时，X 的数学期望和方差阵为

$$\left.\begin{aligned} E(X) &= \mu \\ D_X &= AA^{\mathrm{T}} \end{aligned}\right\} \tag{1-2-2}$$

X 的分布函数和概率密度都简称为 n 维(或 n 元)正态分布，简记为 $X\sim N_n(\mu, AA^{\mathrm{T}})$，或写为 $X\sim N(\mu, D_X)$.

由互相独立的标准正态随机变量组成的随机向量 Z，可写为 $Z\sim N(0, E_n)$. E_n 为 n 阶单位阵.

多维正态分布具有以下性质:

(1)正态随机向量的线性函数还是正态的.例如，设 $X\sim N_n(\mu, AA^{\mathrm{T}})$，$Y=BX+b$，则

$$Y \sim N(B\mu + b, BAA^{\mathrm{T}}B^{\mathrm{T}})$$

(2)设 $X\sim N_n(\mu, AA^{\mathrm{T}})$，记

$$X=\begin{bmatrix} X_1 \\ X_2 \end{bmatrix},\ \mu=\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix},\ AA^{\mathrm{T}}=\begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}$$

则

$$\underset{r\times 1}{X_1}\sim N_r(\mu_1, D_{11}),\ X_2\sim \underset{n-r}{N}(\mu_2, D_{22}).$$

2. 多维正态分布

设有 n 维正态随机向量 $N_n(\mu_X, D_X)$，其中方差阵 D_X 为可逆阵，即 $\det(D_X)\neq 0$，则它的概率密度为

$$f(x)=2(\pi)^{-\frac{\pi}{2}}\mid D_X\mid^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(x-\mu_X)^{\mathrm{T}}D_X^{-1}(x-\mu_X)\right\} \tag{1-2-3}$$

式中 $|D_X|$ 表示 D_X 的行列式.

对于二维正态随机向量 $[X\ Y]^{\mathrm{T}}$，若它有可逆方差阵和数学期望为

$$\begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix} \text{和} \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}$$

则由(1-2-3)式可得其概率密度为

$$f(x,y)=\frac{1}{2\pi\sqrt{\sigma_X^2\sigma_Y^2-\sigma_{XY}^2}}\cdot$$

$$\exp\left\{-\frac{(x-\mu_X)^2\sigma_Y^2-2(x-\mu_X)(y-\mu_Y)\sigma_{XY}+(y-\mu_Y)^2\sigma_X^2}{2(\sigma_X^2\sigma_Y^2-\sigma_{XY}^2)}\right\}$$

因相关系数 $\rho_{XY}=\dfrac{\sigma_{XY}}{\sigma_X\sigma_Y}$，所以上式可写为

$$f(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}}\exp\left\{-\frac{1}{2(1-\rho_{XY}^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2}-2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right\} \tag{1-2-4}$$

这就是二维正态随机向量概率密度.

当 $\rho_{XY}=0$ 或 $\sigma_{XY}=0$ 时,即当 X 和 Y 是互不相关的两个正态随机变量时,则有

$$\begin{aligned}f(x,y)&=\frac{1}{2\pi\sigma_X\sigma_Y}\exp\left\{-\frac{(x-\mu_X)^2}{2\sigma_X^2}-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right\}\\&=\frac{1}{\sqrt{2\pi}\sigma_X}\exp\left\{-\frac{(x-\mu_Y)^2}{2\sigma_X^2}\right\}\cdot\\&\quad\frac{1}{\sqrt{2\pi}\sigma_Y}\exp\left\{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right\}\\&=f_x(x)f_y(y)\end{aligned} \tag{1-2-5}$$

这就是说,当 $\rho_{XY}=0$ 时,X 和 Y 是互相独立的.所以,对于正态分布来说,随机变量的“互不相关”与“互相独立”是等价的.

根据(1-2-4)式绘制二维正态曲面(密度曲面)如图 1-1.曲面在点(μ_x,μ_y)处取得最大值.如果用平行于 XOY 面的平面 $Z=Z_0$(常数)截此曲面,即得到一族椭圆,椭圆上所有点的概率密度值均相等,因此,称这些椭圆为等密度椭圆.

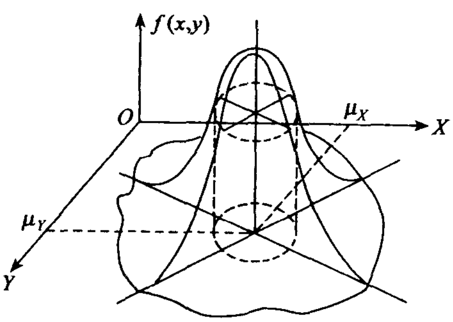

图 1-1

3. 正态随机向量的条件概率密度

设有 $n+t$ 维正态随机向量 X,且设

$$X=\begin{bmatrix}X_1\\X_2\end{bmatrix},\ \mu_X=\begin{bmatrix}\mu_1\\\mu_2\end{bmatrix},\ D_X=\begin{bmatrix}D_{11}&D_{12}\\D_{21}&D_{22}\end{bmatrix}$$

X_1 和 X_2 分别是由 X 的前 n 个分量和后 t 个分量构成的正态随机向量,即 $X_1\sim N_n(\mu_1,D_{11})$,$X_2\sim N_t(\mu_2,D_{22})$.$X$ 的概率密度是

$$f(x)=(2\pi)^{-\frac{n+t}{2}}|D_X|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\begin{bmatrix}x_1-\mu_1\\x_2-\mu_2\end{bmatrix}^{\mathrm{T}}D_X^{-1}\begin{bmatrix}x_1-\mu_1\\x_2-\mu_2\end{bmatrix}\right\} \tag{1-2-6}$$

按分块矩阵求逆公式,有

$$D_X^{-1} = \begin{bmatrix} D_{11}^{-1} + D_{11}^{-1}D_{12}\tilde{D}_{22}^{-1}D_{21}D_{11}^{-1} & -D_{11}^{-1}D_{12}\tilde{D}_{22}^{-1} \\ -\tilde{D}_{22}^{-1}D_{21}D_{11}^{-1} & \tilde{D}_{22}^{-1} \end{bmatrix} \tag{1-2-7}$$

或为

$$D_X^{-1} = \begin{bmatrix} \tilde{D}_{11}^{-1} & -\tilde{D}_{11}^{-1}D_{12}D_{22}^{-1} \\ -D_{22}^{-1}D_{21}\tilde{D}_{11}^{-1} & D_{22}^{-1} + D_{22}^{-1}D_{21}\tilde{D}_{11}^{-1}D_{12}D_{22}^{-1} \end{bmatrix} \tag{1-2-8}$$

其中

$$\left.\begin{aligned} \tilde{D}_{11} &= D_{11} - D_{12}D_{22}^{-1}D_{21} \\ \tilde{D}_{22} &= D_{22} - D_{21}D_{11}^{-1}D_{12} \end{aligned}\right\} \tag{1-2-9}$$

可将(1-2-7)和(1-2-8)两式分别写为

$$D_X^{-1} = \begin{bmatrix} -D_{11}^{-1}D_{12} \\ E_t \end{bmatrix}\tilde{D}_{22}^{-1}[-D_{21}D_{11}^{-1} \quad E_t] + \begin{bmatrix} D_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \tag{1-2-10}$$

$$D_X^{-1} = \begin{bmatrix} E_n \\ -D_{22}^{-1}D_{21} \end{bmatrix}\tilde{D}_{11}^{-1}[E_n \quad -D_{12}D_{22}^{-1}] + \begin{bmatrix} 0 & 0 \\ 0 & D_{22}^{-1} \end{bmatrix} \tag{1-2-11}$$

因 D_X 还可分解为

$$D_X = \begin{bmatrix} D_{11} & 0 \\ D_{21} & \tilde{D}_{22} \end{bmatrix}\begin{bmatrix} E & D_{11}^{-1}D_{12} \\ 0 & E \end{bmatrix} = \begin{bmatrix} \tilde{D}_{11} & D_{12} \\ 0 & D_{22} \end{bmatrix}\begin{bmatrix} E & 0 \\ D_{22}^{-1}D_2 & E \end{bmatrix} \tag{1-2-12}$$

所以，D_X 的行列式之值为

$$|D_X| = |D_{11}||\tilde{D}_{22}| = |D_{22}||\tilde{D}_{11}| \tag{1-2-13}$$

利用(1-2-10)、(1-2-9)和(1-2-13)式，可将概率密度式(1-2-6)改写为

$$\begin{aligned} f(x) &= f(x_1, x_2) \\ &= (2\pi)^{-\frac{n}{2}}|D_{11}|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(x_1 - \mu_1)^T D_{11}^{-1}(x_1 - \mu_1)\right\} \cdot \\ &\quad (2\pi)^{-\frac{t}{2}}|\tilde{D}_{22}|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(x_2 - \tilde{\mu}_2)^T \tilde{D}_{22}^{-1}(x_2 - \tilde{\mu}_2)\right\} \end{aligned} \tag{1-2-14}$$

或

$$\begin{aligned} f(x) &= f(x_1, x_2) \\ &= (2\pi)^{-\frac{t}{2}}|D_{22}|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(x_2 - \mu_2)^T D_{22}^{-1}(x_2 - \mu_2)\right\} \cdot \\ &\quad (2\pi)^{-\frac{n}{2}}|\tilde{D}_{11}|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(x_1 - \tilde{\mu}_1)^T \tilde{D}_{11}^{-1}(x_1 - \tilde{\mu}_1)\right\} \end{aligned} \tag{1-2-15}$$

其中

$$\left.\begin{aligned} \tilde{\mu}_1 &= \mu_1 + D_{12}D_{22}^{-1}(x_2 - \mu_2) \\ \tilde{\mu}_2 &= \mu_2 + D_{21}D_{11}^{-1}(x_1 - \mu_1) \end{aligned}\right\} \tag{1-2-16}$$

根据边际概率密度和多维正态分布的性质可知

$$f_1(x_1) = (2\pi)^{-\frac{n}{2}}|D_{11}|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(x_1 - \mu_1)^T D_{11}^{-1}(x_1 - \mu_1)\right\} \tag{1-2-17}$$

$$f_2(x_2) = (2\pi)^{-\frac{t}{2}}|D_{22}|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(x_2 - \mu_2)^T D_{22}^{-1}(x_2 - \mu_2)\right\} \tag{1-2-18}$$

又由条件概率密度公式知

$$f(x_2/x_1) = \frac{f(x_1, x_2)}{f_1(x_1)} \tag{1-2-19}$$

$$f(x_1/x_2) = \frac{f(x_1, x_2)}{f_2(x_2)} \tag{1-2-20}$$

将(1-2-14)和(1-2-17)两式代入(1-2-19)式,得

$$f(x_2/x_1) = (2\pi)^{-\frac{t}{2}} \mid \tilde{D}_{22} \mid^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(x_2 - \tilde{\mu}_2)^{\mathrm{T}} \tilde{D}_{22}^{-1}(x_2 - \tilde{\mu}_2)\right\} \tag{1-2-21}$$

而将(1-2-15)和(1-2-18)两式代入(1-2-20)式,即得

$$f(x_1/x_2) = (2\pi)^{-\frac{n}{2}} \mid \tilde{D}_{11} \mid^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(x_1 - \tilde{\mu}_1)^{\mathrm{T}} \tilde{D}_{11}^{-1}(x_1 - \tilde{\mu}_1)\right\} \tag{1-2-22}$$

显然,上两式仍然是正态概率密度,根据条件期望和条件方差的定义和正态概率密度的性质可得

$$\left.\begin{aligned} E(X_1/x_2) &= \tilde{\mu}_1 = \mu_1 + D_{12}D_{22}^{-1}(x_2 - \mu_2) \\ E(X_2/x_1) &= \tilde{\mu}_2 = \mu_2 + D_{21}D_{11}^{-1}(x_1 - \mu_1) \end{aligned}\right\} \tag{1-2-23}$$

$$\left.\begin{aligned} D(X_1/x_2) &= \tilde{D}_{11} = D_{11} - D_{12}D_{22}^{-1}D_{21} \\ D(X_2/x_1) &= \tilde{D}_{22} = D_{22} - D_{21}D_{11}^{-1}D_{12} \end{aligned}\right\} \tag{1-2-24}$$

因此,(1-2-21)和(1-2-22)又可写为

$$\left.\begin{aligned} f(x_2/x_1) = {} & (2\pi)^{-\frac{t}{2}} \mid D(X_2/x_1) \mid^{-\frac{1}{2}} \cdot \\ & \exp\left\{-\frac{1}{2}[x_2 - E(X_2/x_1)]^{\mathrm{T}} \cdot\right. \\ & \left. D^{-1}(X_2/x_1)[x_2 - E(X_2/x_1)]\right\} \\ f(x_1/x_2) = {} & (2\pi)^{-\frac{n}{2}} \mid D(X_1/x_2) \mid^{-\frac{1}{2}} \cdot \\ & \exp\left\{-\frac{1}{2}[x_1 - E(X_1/x_2)]^{\mathrm{T}} \cdot\right. \\ & \left. D^{-1}(X_1/x_2)[x_1 - E(X_1/x_2)]\right\} \end{aligned}\right\} \tag{1-2-25}$$

正态分布的条件期望具有以下性质:

(1) 由(1-2-23)式可知,$E(X_1/x_2)$是 x_2 的线性组合,所以,它是正态随机向量;当然,$E(X_2/x_1)$也是正态随机向量.

(2) 设 X 和 Y 为正态随机向量,且设

$$\left.\begin{aligned} \tilde{X} &= X - E(X/y) \\ Z &= AY \end{aligned}\right\} \tag{1-2-26}$$

则 $\tilde{X}$ 是与Z互相独立的随机向量.这是因为

$$\begin{aligned} \tilde{X} &= X - \{\mu_X + D_{XY}D_Y^{-1}(Y - \mu_Y)\} \\ &= X - D_{XY}D_Y^{-1}Y + D_{XY}D_Y^{-1}\mu_Y \end{aligned}$$

由协方差传播律可得

$$\begin{aligned} D(\tilde{X}, Z) &= \begin{bmatrix} E & -D_{XY}D_Y^{-1} \end{bmatrix} \begin{bmatrix} D_X & D_{XY} \\ D_{YX} & D_Y \end{bmatrix} \begin{bmatrix} 0 \\ A^{\mathrm{T}} \end{bmatrix} \\ &= D_{XY}A^{\mathrm{T}} - D_{XY}D_Y^{-1}D_YA^{\mathrm{T}} = 0 \end{aligned}$$

(3) 设 $X \sim N(\mu_X, D_X)$, $Y_1 \sim N(\mu_1, D_1)$, $Y_2 \sim N(\mu_2, D_2)$,且 $\mathrm{cov}(Y_1, Y_2) = 0$,而 $D(X, Y_1) = D_{XY_1} \neq 0$, $D(X, Y_2) = D_{XY_2} \neq 0$,则有

$$E(X/y)=E(X/y_1,y_2)=E(X/y_1)+E(X/y_2)-\mu_X \tag{1-2-27}$$

证 因为

$$E(X/y)=E(X/y_1,y_2)=\mu_X+D_{XY}D_Y^{-1}(y-\mu_Y)$$

$$=\mu_X+[D_{XY_1}\quad D_{XY_2}]\begin{bmatrix}D_1^{-1} & 0\\ 0 & D_2^{-1}\end{bmatrix}\begin{bmatrix}y_1 & -\mu_1\\ y_2 & -\mu_2\end{bmatrix}$$

所以

$$E(X/y_1,y_2)=\{\mu_X+D_{XY_1}D_1^{-1}(y_1-\mu_1)\}+$$

$$\{D_{XY_2}D_2^{-1}(y_2-\mu_2)+\mu_X-\mu_X\}$$

$$=E(X/y_1)+E(X/y_2)-\mu_X$$

(4) 设 $X\sim N(\mu_X,D_X)$, $Y\sim N(\mu_Y,D_Y)$,且

$$\mu_Y=\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix},\ D_Y=\begin{bmatrix}D_{11} & D_{12}\\ D_{21} & D_{22}\end{bmatrix},\ D_{YX}=\begin{bmatrix}D_{Y_1X}\\ D_{Y_2X}\end{bmatrix}=D_{XY}^{\mathrm{T}}$$

令 $\widetilde{Y}_2=Y_2-E(Y_2/y_1)$,则有

$$E(X/y_1,y_2)=E(X/y_1,\tilde{y}_2)$$

$$=E(X/y_1)+E(X/\tilde{y}_2)-\mu_X \tag{1-2-28}$$

证 因为

$$\widetilde{Y}_2=Y_2-\mu_2-D_{21}D_{11}^{-1}(Y_1-\mu_1)$$

$$=[-D_{21}D_{11}^{-1}\quad E]\begin{bmatrix}Y_1\\ Y_2\end{bmatrix}-\mu_2+D_{21}D_{11}^{-1}\mu_1$$

所以

$$\left.\begin{aligned}
&E(\widetilde{Y}_2)=0,\\
&D(\widetilde{Y}_2)=[-D_{21}D_{11}^{-1}\quad E]\begin{bmatrix}D_{11} & D_{12}\\ D_{21} & D_{22}\end{bmatrix}\begin{bmatrix}-D_{11}^{-1}D_{12}\\ E\end{bmatrix}\\
&\qquad=D_{22}-D_{21}D_{11}^{-1}D_{12}=\widetilde{D}_{22}\\
&D(\widetilde{Y}_2,Y_1)=0,\\
&D(\widetilde{Y}_2,X)=[-D_{21}D_{11}^{-1}\quad E]\begin{bmatrix}D_{Y_1X}\\ D_{Y_2X}\end{bmatrix}\\
&\qquad=D_{Y_2X}-D_{21}D_{11}^{-1}D_{Y_1X}
\end{aligned}\right\} \tag{1-2-29}$$

利用分块求逆公式和(1-2-29)式得

$$E(X/y_1,y_2)=\mu_X+D_{XY}D_Y^{-1}(Y-\mu_Y)$$

$$=\mu_X+[D_{XY_1}\quad D_{XY_2}]\begin{bmatrix}D_{11} & D_{12}\\ D_{21} & D_{22}\end{bmatrix}^{-1}\begin{bmatrix}Y_1-\mu_1\\ Y_2-\mu_2\end{bmatrix}$$

$$=\mu_X+[D_{XY_1}\quad D_{XY_2}]\left\{\begin{bmatrix}-D_{11}^{-1}D_{12}\\ E\end{bmatrix}\cdot\right.$$

$$\left.\widetilde{D}_{22}^{-1}[-D_{21}D_{11}^{-1}\quad E]+\begin{bmatrix}D_{11}^{-1} & 0\\ 0 & 0\end{bmatrix}\right\}\begin{bmatrix}Y_1-\mu_1\\ Y_2-\mu_2\end{bmatrix}.$$

$$= \mu_X + \{(-D_{XY_1}D_{11}^{-1}D_{12} + D_{XY_2})D^{-1}\cdot(\widetilde{Y}_2)\cdot$$

$$[-D_{21}D_{11}^{-1}\quad E] + [D_{XY_1}D_{11}^{-1}\quad 0]\}\begin{bmatrix} Y_1-\mu_1 \\ Y_2-\mu_2 \end{bmatrix}$$

$$= \mu_X + D(X,\widetilde{Y}_2)D^{-1}(\widetilde{Y}_2)\{-D_{21}D_{11}^{-1}(Y_1-\mu_1) +$$

$$Y_2-\mu_2\} + D_{XY_1}D_{11}^{-1}(Y_1-\mu_1) + \mu_X - \mu_X$$

$$= E(X/y_1) + E(X/\tilde{y}_2) - \mu_X$$

4. 矩阵反演公式

由于正定矩阵的逆阵惟一,故由(1-2-7)、(1-2-8)两式直接可得:

$$(D_{11} - D_{12}D_{22}^{-1}D_{21})^{-1} = D_{11}^{-1} + D_{11}^{-1}D_{12}(D_{22} - D_{21}D_{11}^{-1}D_{12})^{-1}D_{21}D_{11}^{-1} \tag{1-2-30}$$

$$D_{11}^{-1}D_{12}(D_{22} - D_{21}D_{11}^{-1}D_{12})^{-1} = (D_{11} - D_{12}D_{22}^{-1}D_{21})^{-1}D_{12}D_{22}^{-1} \tag{1-2-31}$$

由此可知,对于任意矩阵 $\boldsymbol{A}$、$\boldsymbol{B}$ 和任意可逆阵 $\boldsymbol{C}$、$\boldsymbol{D}$,只要在下式中它们可以相乘,就有上两式关系,一般形式为

$$(D + ACB)^{-1} = D^{-1} - D^{-1}A(C^{-1} + BD^{-1}A)^{-1}BD^{-1} \tag{1-2-32}$$

$$CB(D + ACB)^{-1} = (C^{-1} + BD^{-1}A)^{-1}BD^{-1} \tag{1-2-33}$$

通常称(1-2-32)、(1-2-33)两式为矩阵反演公式,是两个非常重要的关系式,在测量平差推导公式时常要用到.

矩阵反演公式也可直接证明.令 $H=(D+ACB)^{-1}$,则有

$$\begin{aligned} &(D+ACB)H=E,\text{或 } DH+ACBH=E,\\ &H=D^{-1}-D^{-1}ACBH \end{aligned} \tag{1-2-34}$$

将上式左乘 B,得

$$BD^{-1} = (C^{-1} + BD^{-1}A)CBH,$$

或

$$(C^{-1} + BD^{-1}A)^{-1}BD^{-1} = CBH$$

此即(1-2-33)式,代入(1-2-34)式,即得(1-2-32)式.

1-3 极大似然估计

设有参数向量 $\underset{t\times 1}{X}$,它可以是未知的非随机量,也可以是随机向量,为了估计 X,进行了 n 次观测,得到了观测向量 $\underset{n\times 1}{L}$ 的观测值 $\underset{n\times 1}{l}$,又假定对 X 的所有可能取值为 x,在 $X=x$ 的条件下得到的观测向量 L 的条件概率密度为 $f(l/x)$.容易理解,$f(l/x)$是 x 和 l 的函数,但对具体的观测值 l 来说,$f(l/x)$可以认为只是 x 的函数.因此,如果 $\hat{x}$ 是 x 中的一个,而 $f(l/\hat{x})$ 是 $f(l/x)$中的最大值,那么,$\hat{X}$ 是 X 的准确值的可能性最大.此时把 $\hat{X}$ 叫做 X 的极大似然估值,并记作 $\hat{X}_{ML}(L)$或 $\hat{X}_{ML}$.这就是说,极大似然估计是以

$$f(l/x) = \max \tag{1-3-1}$$

为准则求最佳估值 $\hat{X}$ 的方法.

显然,它满足于

$$\left.\frac{\partial f(l/x)}{\partial x}\right|_{x=\hat{X}_{ML}(L)} = 0 \tag{1-3-2}$$

由于对数是单调增加函数，因此 $\ln f(l/x)$ 与 $f(l/x)$ 在相同的 x 值达到最大，亦即(1-3-2)式等价于

$$\left.\frac{\partial \ln f(l/x)}{\partial x}\right|_{x=\hat{X}_{ML}(L)} = 0 \tag{1-3-3}$$

此方程称为似然方程，$f(l/x)$ 称为似然函数，而 $\ln f(l/x)$ 称为对数似然函数.

如果参数 X 是非随机量，则

$$f(l/x) = f(l,x)$$

而(1-3-1)式变为

$$f(l,x) = \max \tag{1-3-4}$$

此时，$f(l,x)$ 是 $\underset{n\times 1}{L}$ 的概率密度，其中的 x 只是表示函数与参数 X 有关.

由似然方程或(1-3-2)式可见，极大似然估值 $\hat{X}_{ML}$ 是观测值 L 的函数. 在采用极大似然估计求 $\hat{X}_{ML}$ 时，需要首先知道似然函数 $f(l/x)$ 或对数似然函数 $\ln f(l/x)$.

[例 1-3-1] 设有观测方程为

$$L_k = X_k,\ (k = 1,2,\cdots,n)$$

并且已知 X_k 与 $X_j(k\neq j)$ 互相独立，X_k 的概率密度为

$$f(x_k) = \frac{1}{\sqrt{2\pi}\sigma_x}\exp\left\{-\frac{1}{2\sigma_x^2}(x_k-\mu_x)^2\right\}$$

试由观测向量 $\underset{n\times 1}{L}$ 求 μ_x 和 σ_x^2 的极大似然估值.

解 由于 L_k 与 X_k 具有相同的分布，所以有似然函数

$$\begin{aligned} f(l/\mu_x,\sigma_x^2) &= f(l_1)f(l_2)\cdots f(l_n) \\ &= \frac{1}{(\sqrt{2\pi})^n\sigma_x^n}\exp\left\{-\frac{1}{2\sigma_x^2}\sum_{k=1}^{n}(l_k-\mu_x)^2\right\} \end{aligned}$$

下面分几种情况讨论：

(1) 当 σ_x^2 为已知，求 μ_x 的极大似然估值.

因对数似然函数为

$$\ln f(l/\mu_x,\sigma_x^2) = -\ln\{(2\pi)^{n/2}\sigma_x^n\} - \frac{1}{2\sigma_x^2}\sum_{k=1}^{n}(l_k-\mu_x)^2$$

故由似然方程可得：

$$\left.\frac{\partial \ln f(l/\mu_x,\sigma_x^2)}{\partial \mu_x}\right|_{\mu_x=\hat{\mu}_{xML}} = \left.\frac{1}{\sigma_x^2}\sum_{k=1}^{n}(l_k-\mu_x)\right|_{\mu_x=\hat{\mu}_{xML}} = 0$$

即

$$\sum_{k=1}^{n}(l_k-\hat{\mu}_{xML}) = 0$$

因此

$$\hat{\mu}_{xML} = \frac{1}{n}\sum_{k=1}^{n}l_k$$

又因为

$$E(\hat{\mu}_{xML}) = \frac{1}{n}\sum_{k=1}^{n}E(l_k) = \frac{1}{n}\sum_{k=1}^{n}E(X_k) = \mu_x$$