深蓝学院
shenlanxueyuan.com

机器人学中的状态估计

Timothy Barfoot 著　高翔，谢晓佳等 译
Slides by Xiang Gao

2020年春

深蓝学院
shenlanxueyuan.com

第1讲 概述与基础知识

- 状态估计简介
- 概率密度函数
- 高斯概率密度函数

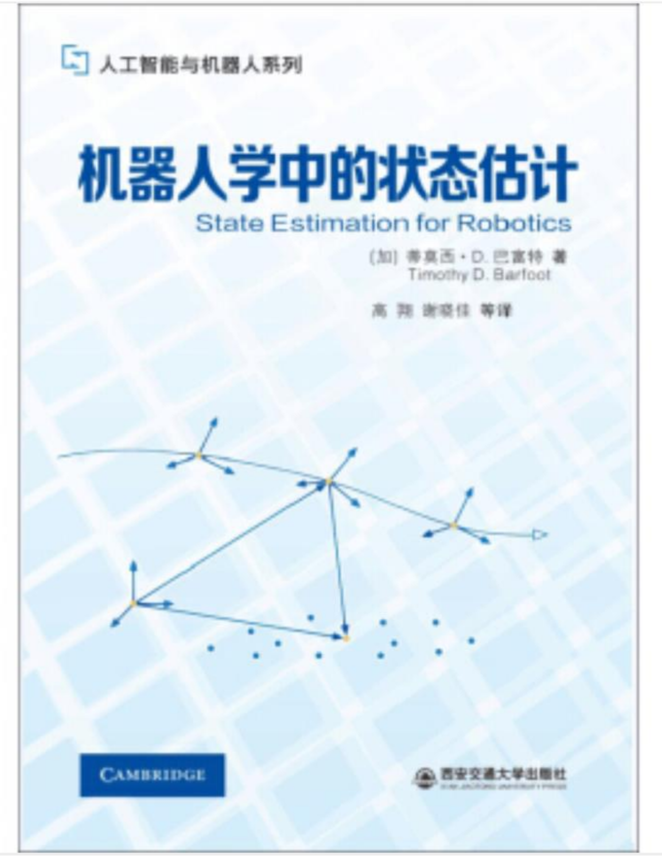
深蓝学院
shenlanxueyuan.com

状态估计简介

深蓝学院
shenlanxueyuan.com

状态估计简介

- 本课程内容：
    - 状态估计问题的概念、方法和理解
    - 三维空间下的状态估计问题
    - 矩阵李群与李代数
    - 状态估计的应用
- 课程使用的教材：《机器人学中的状态估计》，Tim Barfoot著
- 重点的内容：状态估计理论、李群李代数
- 忽略的内容：连续时间部分
- 在原书基础上补充一些推导过程和前后内容的串讲
- 我会根据自己体验，介绍各部分内容的用处，略过一些复杂的、实用性不大的内容
- 需要时会手推一部分公式
- 请尽可能买一本英文或中文版教材，习题主要来自教材



深蓝学院
shenlanxueyuan.com

状态估计简介

- 课程大纲
  - 第1讲 概述与基础知识
  - 第2讲 线性高斯系统的状态估计问题
  - 第3讲 非线性非高斯系统中的状态估计
  - 第4讲 偏差和外点的处理
  - 第5讲 三维几何学
  - 第6讲 矩阵李群
  - 第7讲 应用：位姿估计问题
  - 第8讲 应用：位姿和点的估计问题

本课程的一些忠告：

- 课程有很多公式（全书有一千多个，每章两三百个）
- 课上用的大部分会推出来（但是不保证你能马上看懂）
- 最好留一点时间预习和复习
- 课外内容请根据自己时间酌量补充

进一步，

- 有耐心
- 掌握推导的方法而不是过程
- 培养对公式的感觉
- 记住结论和结论的意义

深蓝学院
shenlanxueyuan.com

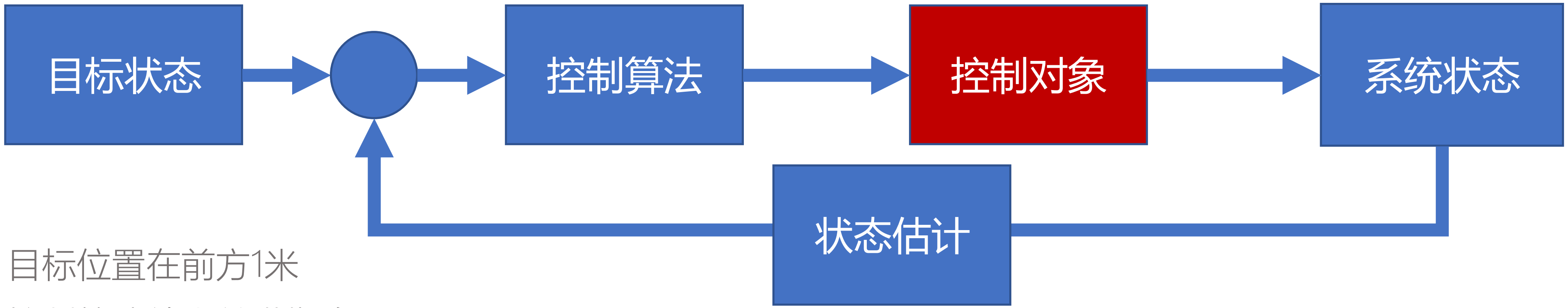
第1讲 概述与基础知识

- 状态估计简介
- 概率密度函数
- 高斯概率密度函数
- 高斯过程

深蓝学院
shenlanxueyuan.com

状态估计简介

- 机器人、无人机、自动驾驶汽车是如何运动的?



1. 目标位置在前方1米
2. 控制算法给出前进指令
3. 控制对象实际前进
4. 观测到系统实际前进的距离，反馈给控制算法

两大实际问题：**控制**（control）和 **状态估计**（state estimation）

深蓝学院
shenlanxueyuan.com

状态估计简介

- 状态估计简史
  - 状态估计是个古老的问题，早期的航海图、指南针、六分仪都是状态估计的鼻祖
  - 1801年，高斯提出了最小二乘法，用于估计谷神星（Ceres）的位置
  - 1960年，卡尔曼发表线性高斯系统下的卡尔曼滤波器
  - NASA在阿波罗登月着陆过程中，使用卡尔曼滤波器估计自身位置

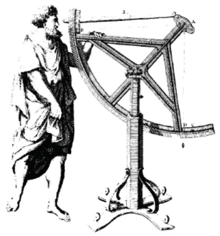
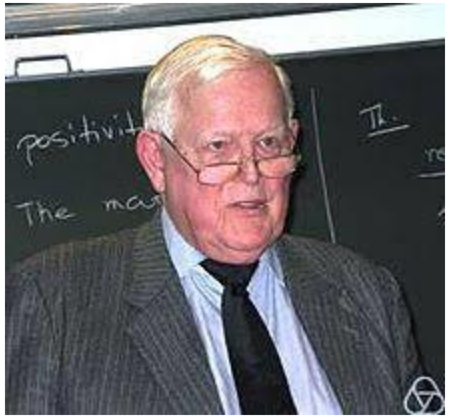


图 1–1 四分仪，测量角度的一种工具

图 1–2 Harrison 制造的 H4，第一个用于准确航海计时的手持式钟表，用于确定经度

卡尔·弗里德里希·高斯（Carl Friedrich Gauss，1777—1855），德国数学家，在很多领域都有重大贡献，包括统计学和估计理论。

鲁道夫·埃米尔·卡尔曼（Rudolf Emil Kálmán，1930—2016），匈牙利裔美国人，电子工程师、数学家和发明家。



深蓝学院
shenlanxueyuan.com

状态估计简介

表 1–1 估计理论的里程碑

1654	帕斯卡和费马奠定了概率论的基础
1764	贝叶斯法则
1801	高斯用最小二乘法估计出谷神星的轨道
1805	勒让德发表“最小二乘法”
1913	马尔科夫链（Markov chains）
1933	查普曼-柯尔莫哥洛夫等式（Chapman-Kolmogorov equations）
1949	维纳滤波器（Wiener filter）
1960	卡尔曼滤波器（Kalman (Bucy) Filter）
1965	Rauch-Tung-Striebel 平滑算法
1970	Jazwinski 提出“贝叶斯滤波”

- 什么是状态估计
  - 状态估计就是理解传感器的本质
- 传感器
  - 以一定精度测量物理变量
  - 状态估计问题：如何以最好的方式利用已有传感器

深蓝学院
shenlanxueyuan.com

状态估计简介

- **传感器分类**
  - 内感受型（interoceptive）
    - 测量自身的数据，例子：加速度计、陀螺、轮式编码器
  - 外感受型（extroceptive）
    - 测量外部数据，例子：GPS、相机、激光
  - 外感受型测量角度和位置，内感受型测量速度、角速度、加速度等
  - 实用的状态估计系统同时利用内感受型和外感受型传感器
- **状态估计，是根据系统的先验模型和测量序列，对系统内在状态进行重构的问题**

深蓝学院
shenlanxueyuan.com

状态估计简介

- 课程内容与书本内容对应关系
  - 第一部分 状态估计机理 第1-4课
  - 第二部分 三维空间运动机理 第5-6课
  - 第三部分 应用 第7-8课
  - 由于每节课内容可能不一样多，部分课程可能分割成两节，视讲课时间而定
  - 课后布置一定量作业，课内外时间占比约1:5-1:7左右，请留出一定时间完成作业
  - 课程和作业均以理论推导为主，不含编程

深蓝学院
shenlanxueyuan.com

状态估计简介

- 数学符号说明

一般符号

a	标量
$\mathbf{a}$	向量
$\mathbf{A}$	矩阵
$p(\mathbf{a})$	$\mathbf{a}$ 的概率密度函数
$p(\mathbf{a}\|\mathbf{b})$	在条件 $\mathbf{b}$ 下 $\mathbf{a}$ 的概率密度函数
$\mathcal{N}(\mathbf{a}, \mathbf{B})$	均值为 $\mathbf{a}$，协方差为 $\mathbf{B}$ 的高斯概率密度函数
$\mathcal{GP}(\boldsymbol{\mu}(t), \boldsymbol{\mathcal{K}}(t, t'))$	均值函数为 $\boldsymbol{\mu}(t)$，协方差函数为 $\boldsymbol{\mathcal{K}}(t, t')$ 的高斯过程
$\boldsymbol{\mathcal{O}}$	能观性矩阵
$(\cdot)_k$	k 时刻的值
$(\cdot)_{k_1:k_2}$	由 k_1 时刻到 k_2 时刻的值的集合
$\underrightarrow{\mathcal{F}}_a$	三维空间中的参考系
$\underrightarrow{a}$	三维空间中的向量
$(\cdot)^\times$	叉积运算符，可将 3×1 的向量生成反对称矩阵
$\mathbf{1}$	单位矩阵
$\mathbf{0}$	零矩阵
$\mathbb{R}^{M \times N}$	$M \times N$ 的矩阵
$\hat{(\cdot)}$	后验
$\check{(\cdot)}$	先验

矩阵李群符号

$SO(3)$	特殊正交群，表示旋转
$\mathfrak{so}(3)$	$SO(3)$ 对应的李代数
$SE(3)$	特殊欧几里得群，表示位姿
$\mathfrak{se}(3)$	$SE(3)$ 对应的李代数
$(\cdot)^\wedge$	与李代数相关的运算符
$(\cdot)^\curlywedge$	与李代数的伴随相关的运算符
$\mathrm{Ad}(\cdot)$	生成李群的伴随的运算符
$\mathrm{ad}(\cdot)$	生成李代数的伴随的运算符
$\mathbf{C}_{ba}$	3×3 的旋转矩阵，可以计算在 $\underrightarrow{\mathcal{F}}_a$ 中的点在 $\underrightarrow{\mathcal{F}}_b$ 下的表示
$\mathbf{T}_{ba}$	4×4 的变换矩阵，可以计算在 $\underrightarrow{\mathcal{F}}_a$ 中的点在 $\underrightarrow{\mathcal{F}}_b$ 下的表示
$\boldsymbol{\mathcal{T}}_{ba}$	变换矩阵的 6×6 伴随矩阵

深蓝学院
shenlanxueyuan.com

第1讲 概述与基础知识

- 状态估计简介
- 概率密度函数
- 高斯概率密度函数
- 高斯过程

深蓝学院
shenlanxueyuan.com

概率密度函数

- 为什么要从概率密度函数开始谈?
- 系统的状态通常由一个随机变量描述，我们从它的分布情况了解系统的状态信息
- 多数系统使用高斯分布来刻画系统状态，因此高斯分布在状态估计中占据重要地位
- 我们先讨论概率密度函数的一般性质，再介绍高斯分布里对应的性质
- 本章是后续所有章节的基础，请务必掌握

深蓝学院
shenlanxueyuan.com

概率密度函数

- 本节我们回顾概率学的基础知识，简单说明在后面用到的地方
- 概率密度函数：

我们定义 x 为区间 $[a, b]$ 上的**随机变量**（random variable），服从某个概率密度函数 $p(x)$，那么这个非负函数[1] 必须满足：

$$\int_a^b p(x)\mathrm{d}x = 1 \tag{2.1}$$

- 这个称为全概率公理
- 注：(1) 这是PDF不是概率 (2) 这里不讨论概率定义问题

深蓝学院
shenlanxueyuan.com

概率密度函数

- 概率密度函数在某区间的积分即为概率：

$$\Pr(c \leqslant x \leqslant d) = \int_c^d p(x)\mathrm{d}x \tag{2.2}$$

- 如果x表达某种状态，也称为x在该区间下的可能性/似然（likehood）：

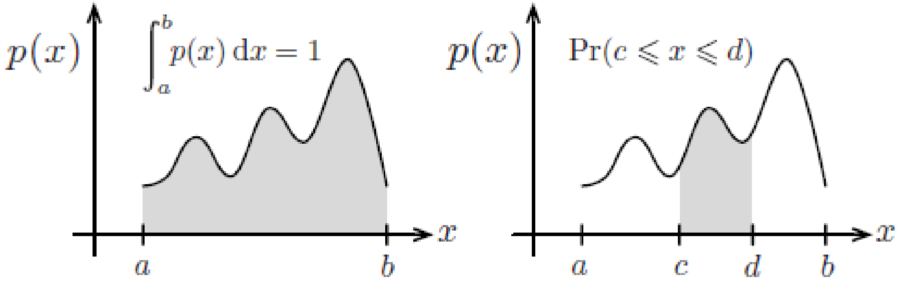


图 2–1　左图为定义在一个有限区域上的概率密度；右图为在一个子区间上的概率

深蓝学院
shenlanxueyuan.com

概率密度函数

- 条件概率： $(\forall y) \quad \int_a^b p(x|y)\mathrm{d}x = 1 \qquad (2.3)$
- 联合概率： $p(x_1, x_2, \cdots, x_N)$
- 联合概率也满足全概率公理：

$$\int_a^b p(x)\mathrm{d}x = \int_{a_N}^{b_N} \cdots \int_{a_2}^{b_2} \int_{a_1}^{b_1} p(x_1, x_2, \cdots, x_N)\mathrm{d}x_1\mathrm{d}x_2 \cdots \mathrm{d}x_N = 1 \qquad (2.6)$$

概率密度函数

- 贝叶斯公式：联合=条件*边缘

- 另一形式：

$$p(x, y) = p(x|y)p(y) = p(y|x)p(x) \tag{2.7}$$

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} \tag{2.8}$$

  - 赋予该式物理意义：x=状态，y=传感器读数，p(y|x)=传感器模型，p(x|y)=状态估计

Thomas Bayes
1702 - 1761

托马斯·贝叶斯（Thomas Bayes，1701—1761）是一位英国统计学家、哲学家和长老会牧师。贝叶斯以其在概率论领域的研究闻名于世，他提出的贝叶斯定理对于现代概率论和数理统计的发展有重要的影响。他生前没有发表任何科学论著；他的笔记在他死后由理查德·普莱斯（Richard Price）编辑出版[14]。

深蓝学院
shenlanxueyuan.com

概率密度函数

- 贝叶斯公式:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} \tag{2.8}$$

- 分母p(y)难以直接计算:

$$\begin{aligned} p(y) &= p(y)\underbrace{\int p(x|y)\mathrm{d}x}_{1} = \int p(x|y)p(y)\mathrm{d}x \\ &= \int p(x,y)\mathrm{d}x = \int p(y|x)p(x)\mathrm{d}x \end{aligned} \tag{2.10}$$

  - 物理意义为所有读数的分布（例：一张图像有猫的概率是多大？）
  - 实用当中不会直接描述p(x|y)的分布，通常取某些近似形式（特指高斯）

深蓝学院
shenlanxueyuan.com

概率密度函数

- 复杂的分布难以描述，奈何?
- 使用一些简单的性质来刻画这个分布
- 矩（moments）
  - 0阶矩恒等于1
  - 1阶矩称为期望（Expectation）：
    $$\mu = E[x] = \int x p(x)\, dx \tag{2.11}$$
  - 2阶矩称为协方差（Covariance）：
    $$\Sigma = E\left[(x-\mu)(x-\mu)^T\right] \tag{2.14}$$
  - 3阶和4阶称为偏度（skewness）和峰度（kurtosis）

深蓝学院
shenlanxueyuan.com

概率密度函数

- 若只用1阶和2阶矩刻画某个分布，实际上等同于将它近似为高斯分布
- 如何估计某个随机变量的矩？——通过样本
- 每次传感器测量数据就是一个变量的样本（measurement）

- 如果有很多样本，就可以计算该随机变量的1和2阶矩：

$$x_{\text{meas}} \leftarrow p(x) \tag{2.15}$$

$$\mu_{\text{meas}} = \frac{1}{N}\sum_{i=1}^{N} x_{i,\text{meas}} \tag{2.16a}$$

$$\Sigma_{\text{meas}} = \frac{1}{N-1}\sum_{i=1}^{N}(x_{i,\text{meas}} - \mu_{\text{meas}})(x_{i,\text{meas}} - \mu_{\text{meas}})^T \tag{2.16b}$$

N-1称为贝塞尔修正

深蓝学院
shenlanxueyuan.com

概率密度函数

- 随机变量的统计独立性

如果两个随机变量 x 和 y 的联合概率密度可以用以下的因式分解，那么我们称这两个随机变量是**统计独立**（statistically independent）的：

$$p(x, y) = p(x)p(y) \tag{2.17}$$

- 以及不相关性：

同样地，如果两个变量的期望运算满足下式，我们称它们是**不相关的**（uncorrelated）：

$$E[xy^{\mathrm{T}}] = E[x]E[y]^{\mathrm{T}} \tag{2.18}$$

- 独立性可推出不相关性，反之不行

- 对于高斯分布来说，独立性=不相关性

深蓝学院
shenlanxueyuan.com

概率密度函数

- 归一化积（Normalized product）
- 有时候我们会考虑状态x的多个估计分布之融合
- 若 $p_1(x)$ 和 $p_2(x)$ 是关于x的两个分布，那么它们的归一化积为

$$p(x) = \eta p_1(x) p_2(x)$$

- 其中η为归一化因子。
- 归一化积在多传感器融合中十分有用！

- 这种因子我们之后还会见到。它通常是一个数，某些情况下为常数，用于保证概率积分为1。
- 大部分情况下我们不关心这个常数的具体取值。

深蓝学院
shenlanxueyuan.com

概率密度函数

- 为什么状态的多次估计分布之融合可以用归一化积描述?
- 假设状态为x，且 y_1，y_2 是两次测量，那么根据这两次测量可以估计x的分布：

$$p(\boldsymbol{x}|\boldsymbol{y}_1,\boldsymbol{y}_2) = \eta p(\boldsymbol{x}|\boldsymbol{y}_1)p(\boldsymbol{x}|\boldsymbol{y}_2)$$

- 由贝叶斯公式：

$$p(x|y_1,y_2) = \frac{p(y_1,y_2|x)p(x)}{p(y_1,y_2)}$$

代回上式：$\eta = \dfrac{p(y_1)p(y_2)}{p(y_1,y_2)p(x)}$

那么，若p(x)为均匀分布， η就为常数

- 假设测量是独立的：

$$p(y_1,y_2|x) = p(y_1|x)p(y_2|x) = \frac{p(x|y_1)p(y_1)}{p(x)}\frac{p(x|y_2)p(y_2)}{p(x)}$$

深蓝学院
shenlanxueyuan.com

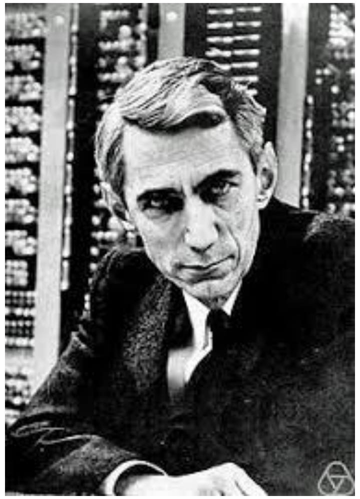
概率密度函数

- 香农信息（Shannon Information）：刻画某个随机变量的不确定性

$$H(x) = -E[\ln p(x)] = -\int p(x) \ln p(x) \mathrm{d}x$$

□ 或称负熵，描述分布的混乱程度

- 例子：明天90%概率下雨，香农信息=0.325，明天50%概率下雨，香农信息=0.693



- 两个随机变量之间还可以定义互信息（mutual information）

$$I(x, y) = E\left[\ln\left(\frac{p(x,y)}{p(x)p(y)}\right)\right] = \iint p(x,y) \ln\left(\frac{p(x,y)}{p(x)p(y)}\right) \mathrm{d}x\mathrm{d}y$$

- 显然，变量独立时，互信息为0；不独立时：

$$I(x, y) = H(x) + H(y) - H(x, y)$$

深蓝学院
shenlanxueyuan.com

概率密度函数

- 克拉美罗下界和费歇尔信息量（略）

深蓝学院
shenlanxueyuan.com

第1讲 概述与基础知识

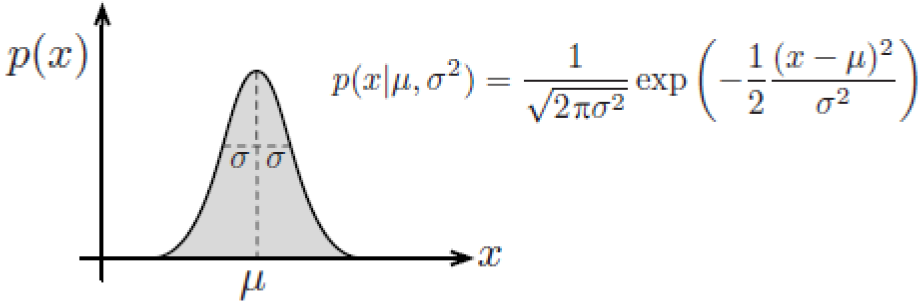
- 状态估计简介
- 概率密度函数
- 高斯概率密度函数
- 高斯过程

深蓝学院 shenlanxueyuan.com

高斯概率密度函数

- 下面来介绍高斯概率密度函数的对应性质，以及高斯分布自身的一些特性。
- 一维高斯分布：

$$p(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$



其中 μ 称为**均值**（mean），σ^2 为**方差**（variance），σ 称为**标准差**（standard deviation）。

- 思考：为何高斯分布积分为1？

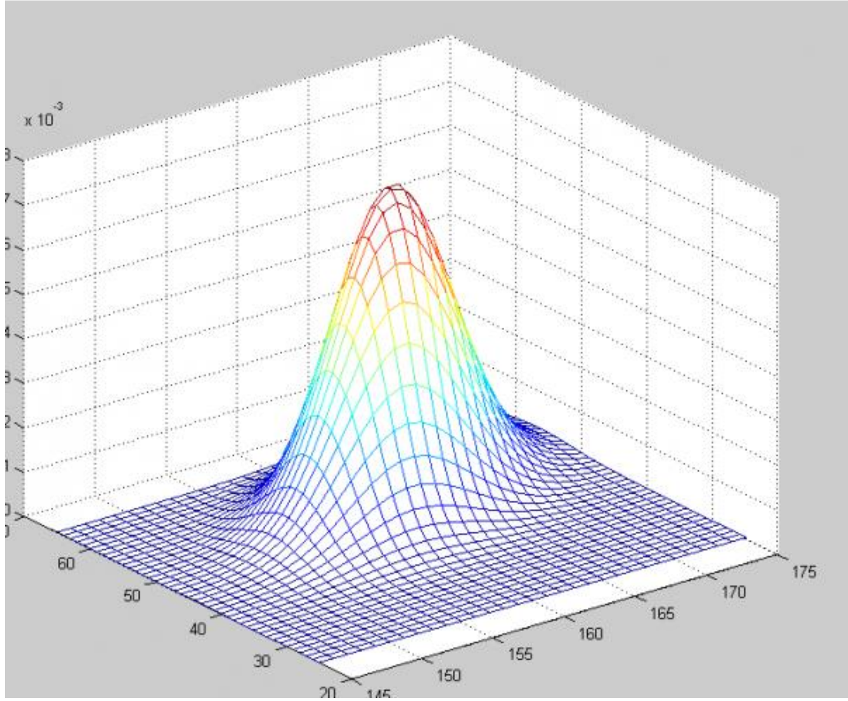
深蓝学院
shenlanxueyuan.com

高斯概率密度函数

- 多维高斯分布：

$$p(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^N \det \boldsymbol{\Sigma}}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

其中 $\boldsymbol{\mu} \in \mathbb{R}^N$ 是均值，$\boldsymbol{\Sigma} \in \mathbb{R}^{N\times N}$ 是协方差矩阵（对称正定矩阵）



- 均值和协方差的计算：

$$\boldsymbol{\mu} = E[\boldsymbol{x}] = \int_{-\infty}^{\infty} \boldsymbol{x} \frac{1}{\sqrt{(2\pi)^N \det \boldsymbol{\Sigma}}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right) d\boldsymbol{x}$$

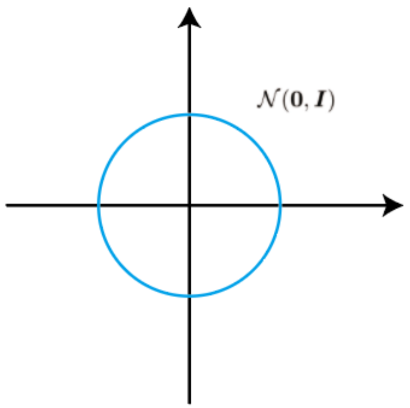
$$\begin{aligned}\boldsymbol{\Sigma} &= E\left[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\right] \\ &= \int_{-\infty}^{\infty} (\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} \frac{1}{\sqrt{(2\pi)^N \det \boldsymbol{\Sigma}}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right) d\boldsymbol{x}\end{aligned}$$

深蓝学院 shenlanxueyuan.com

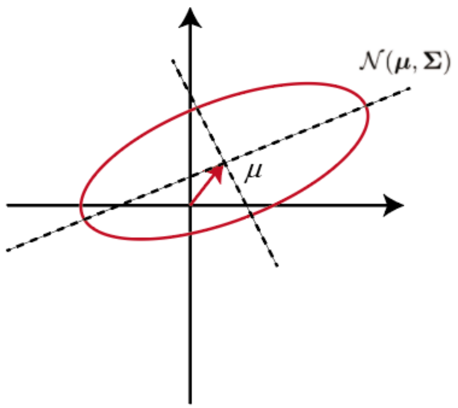
高斯概率密度函数

- 思考：为何协方差矩阵是对称正定矩阵？是人为规定的吗？
- 考虑标准正态分布： $p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^N}} \exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\right)$
- 对变量作可逆线性变换： $\boldsymbol{x} = \boldsymbol{A}(\boldsymbol{y} - \boldsymbol{\mu})$
- 代入后可发现： $\boldsymbol{\Sigma}^{-1} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$
- 由于A是可逆的，所以协方差是对称正定的
- 但是估计出来的协方差不能保证正定性

- 一般的多维分布可看成标准分布加上了一次线性变换
- 协方差描述了线性变换的情况



标准高斯分布



线性变换后的高斯分布

深蓝学院
shenlanxueyuan.com

高斯概率密度函数

- Isserlis定理（略，用到时再推）

深蓝学院
shenlanxueyuan.com

高斯概率密度函数

- 联合高斯分布：

  设有一对服从多元正态分布的变量 (x, y)，可以写出它们的联合概率密度函数：

$$p(x, y) = \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}\right)$$

- 由于协方差总是对称正定的，所以： $\boldsymbol{\Sigma}_{yx} = \boldsymbol{\Sigma}_{xy}^{\mathrm{T}}$

- 我们总可以把联合分布写成条件概率与边缘概率之乘积：

$$p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{x}|\boldsymbol{y})p(\boldsymbol{y})$$

- ☐ 这件事称为边缘化，或者高斯推断
- ☐ 它是后续很多内容的理论基础

深蓝学院 shenlanxueyuan.com

高斯概率密度函数

- 下面演示高斯分布如何进行边缘化

- 这个也称舒尔补或者矩阵打洞
- 可以看成对左侧矩阵进行高斯消元，然后把基础变换阵挪到右侧

- 对于联合的高斯协方差矩阵，进行对角化：

$$\begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix} = \begin{bmatrix} 1 & \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx} & 0 \\ 0 & \boldsymbol{\Sigma}_{yy} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx} & 1 \end{bmatrix}$$

- 两边取逆：

$$\begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ -\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx} & 1 \end{bmatrix} \times \begin{bmatrix} \left(\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}\right)^{-1} & 0 \\ 0 & \boldsymbol{\Sigma}_{yy}^{-1} \end{bmatrix} \begin{bmatrix} 1 & -\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1} \\ 0 & 1 \end{bmatrix}$$

深蓝学院
shenlanxueyuan.com

高斯概率密度函数

- 把取逆形式代入高斯分布：

$$
\begin{aligned}
&\left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}\right)^{\mathrm{T}} \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}^{-1} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}\right) \\
&= \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}\right)^{\mathrm{T}} \begin{bmatrix} 1 & 0 \\ -\Sigma_{yy}^{-1}\Sigma_{yx} & 1 \end{bmatrix} \begin{bmatrix} \left(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\right)^{-1} & 0 \\ 0 & \Sigma_{yy}^{-1} \end{bmatrix} \\
&\qquad \times \begin{bmatrix} 1 & -\Sigma_{xy}\Sigma_{yy}^{-1} \\ 0 & 1 \end{bmatrix} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}\right) \\
&= \left(x - \mu_x - \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)\right)^{\mathrm{T}} \left(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\right)^{-1} \\
&\qquad \times \left(x - \mu_x - \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)\right) + (y - \mu_y)^{\mathrm{T}} \Sigma_{yy}^{-1} (y - \mu_y)
\end{aligned}
$$

$$
\boxed{\begin{bmatrix} x - \mu_x - \Sigma_{xy}\Sigma_{yy}^{-1}\left(y - \mu_y\right) \\ y - \mu_y \end{bmatrix}^{T} \begin{bmatrix} \left(\Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\right)^{-1} & 0 \\ 0 & \Sigma_{yy}^{-1} \end{bmatrix} \begin{bmatrix} x - \mu_x - \Sigma_{xy}\Sigma_{yy}^{-1}\left(y - \mu_y\right) \\ y - \mu_y \end{bmatrix}}
$$

深蓝学院
shenlanxueyuan.com

高斯概率密度函数

- 简单写为：

$$p(x, y) = p(x|y)p(y)$$

$$p(x|y) = \mathcal{N}\left(\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y), \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\right)$$

$$p(y) = \mathcal{N}\left(\mu_y, \Sigma_{yy}\right)$$

- □ 联合分布
- □ 条件分布
- □ 边缘分布

- 实际用途：
  - 已知联合分布p(x,y)和p(y)，求p(x|y)
  - 这个过程中，均值作出了调整，方差变小了

例子：已知上个时刻状态和此时刻的观测，求此时刻的状态?

x　　　　y　　　　x|y

- 该方法称为高斯推断，是后续内容重要基础

深蓝学院
shenlanxueyuan.com

高斯概率密度函数

- 高斯分布的独立性和不相关性
  - 高斯分布的不相关性等价于独立性
  - 不相关性：由定义 $\Sigma_{xy} = E\left[(x-\mu_x)(y-\mu_y)^{\mathrm{T}}\right] = E\left[xy^{\mathrm{T}}\right] - E[x]E[y]^{\mathrm{T}}$
  - 为零，于是由上页第2行：$p(x,y) = p(x|y)p(y)$

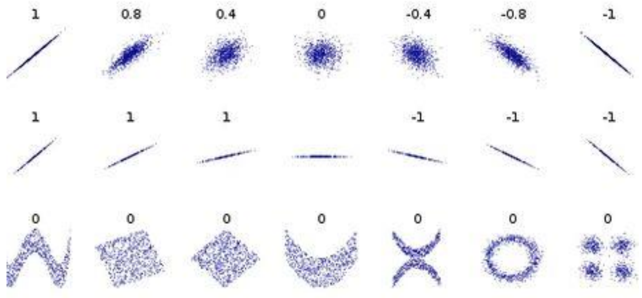
$$p(x|y) = \mathcal{N}\left(\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y-\mu_y), \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\right)$$

$$p(y) = \mathcal{N}\left(\mu_y, \Sigma_{yy}\right)$$

  - 得p(x|y)=p(x)，反之亦然，于是独立

- 所以对高斯分布，谈论不相关和独立是一个意思。

□ 但对于一般分布，独立性意味着x和y之间不存在任何关系，相关性则是存不存在线性关系



深蓝学院
shenlanxueyuan.com

高斯概率密度函数

- 高斯分布的线性变换
- 考虑一个高斯随机变量： $x \in \mathbb{R}^N \sim \mathcal{N}(\mu_x, \Sigma_{xx})$
- 它有一个线性变换： $y = Gx$
- 那么y的均值和方差为： $\mu_y = E[y] = E[Gx] = GE[x] = G\mu_x$

$$\Sigma_{yy} = E\left[(y-\mu_y)(y-\mu_y)^{\mathrm{T}}\right]$$
$$= GE\left[(x-\mu_x)(x-\mu_x)^{\mathrm{T}}\right]G^{\mathrm{T}} = G\Sigma_{xx}G^{\mathrm{T}}$$

深蓝学院
shenlanxueyuan.com

高斯概率密度函数

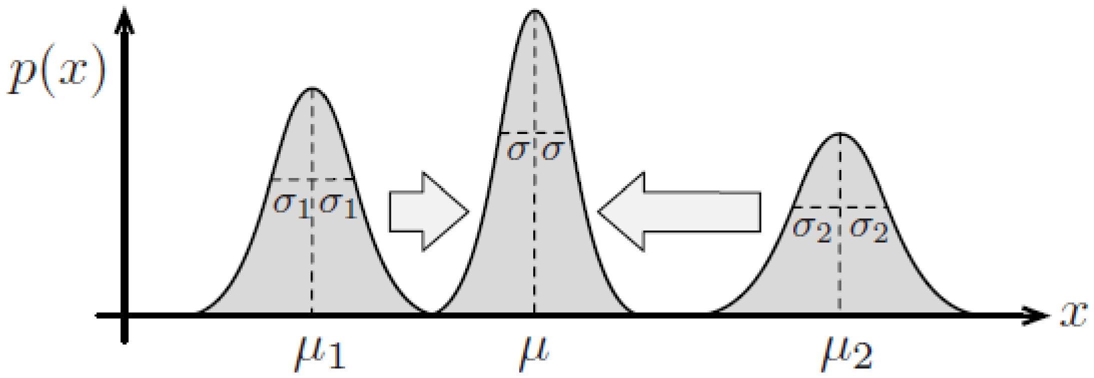
- 高斯分布的归一化积
- K个不同高斯分布的归一化积为：

$$\exp\left(-\frac{1}{2}(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)\right) \equiv \eta\prod_{k=1}^{K}\exp\left(-\frac{1}{2}(x-\mu_k)^{\mathrm{T}}\Sigma_k^{-1}(x-\mu_k)\right)$$

- 提取x的二次项系数和一次项系数，得：

$$\Sigma^{-1}=\sum_{k=1}^{K}\Sigma_k^{-1}$$

$$\Sigma^{-1}\mu=\sum_{k=1}^{K}\Sigma_k^{-1}\mu_k$$



$$\frac{1}{\sigma^2}=\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}$$

$$\frac{\mu}{\sigma^2}=\frac{\mu_1}{\sigma_1^2}+\frac{\mu_2}{\sigma_2^2}$$

- 该式可视为多个高斯分布之融合

深蓝学院
shenlanxueyuan.com

高斯概率密度函数

- Sherman-Morrison-Woodbury（SMW）恒等式
- SMW是一组常用的矩阵变换恒等式，在卡尔曼滤波中用来化简矩阵乘积之逆
- SMW的推导：对于任意可逆矩阵，总可以用初等变换将其变成LDU或UDL形式：

$$
\begin{aligned}
&\begin{bmatrix} A^{-1} & -B \\ C & D \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ CA & 1 \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & D + CAB \end{bmatrix} \begin{bmatrix} 1 & -AB \\ 0 & 1 \end{bmatrix} \qquad \text{(LDU)} \\
&= \begin{bmatrix} 1 & -BD^{-1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} A^{-1} + BD^{-1}C & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} 1 & 0 \\ D^{-1}C & 1 \end{bmatrix} \qquad \text{(UDL)}
\end{aligned}
$$

深蓝学院
shenlanxueyuan.com

高斯概率密度函数

- 对LDU和UDL式分别求逆：

$$\begin{bmatrix} A^{-1} & -B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} 1 & AB \\ 0 & 1 \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & (D+CAB)^{-1} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -CA & 1 \end{bmatrix} = \begin{bmatrix} A - AB(D+CAB)^{-1}CA & AB(D+CAB)^{-1} \\ -(D+CAB)^{-1}CA & (D+CAB)^{-1} \end{bmatrix}$$

$$\begin{bmatrix} A^{-1} & -B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ -D^{-1}C & 1 \end{bmatrix} \begin{bmatrix} (A^{-1}+BD^{-1}C)^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} 1 & BD^{-1} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} (A^{-1}+BD^{-1}C)^{-1} & (A^{-1}+BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A^{-1}+BD^{-1}C)^{-1} & D^{-1} - D^{-1}C(A^{-1}+BD^{-1}C)^{-1}BD^{-1} \end{bmatrix}$$

- 比较各块，即得SMW恒等式：

$$(A^{-1}+BD^{-1}C)^{-1} \equiv A - AB(D+CAB)^{-1}CA$$

$$(D+CAB)^{-1} \equiv D^{-1} - D^{-1}C(A^{-1}+BD^{-1}C)^{-1}BD^{-1}$$

$$AB(D+CAB)^{-1} \equiv (A^{-1}+BD^{-1}C)^{-1}BD^{-1}$$

$$(D+CAB)^{-1}CA \equiv D^{-1}C(A^{-1}+BD^{-1}C)^{-1}$$

- 这四个都可以用
- 从左到右或从右到左都可以
- 很灵活
- 1和2，3和4是对称的，记两个即可

深蓝学院 shenlanxueyuan.com

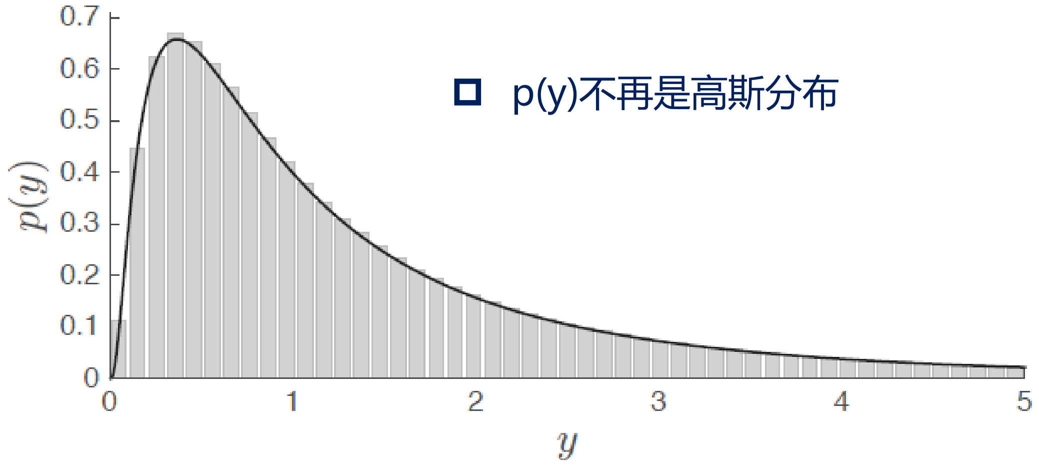
高斯概率密度函数

- 思考：为什么p(y|x)是高斯分布？为什么它的均值是g(x)?
- 这里默认R是对称正定的

- 高斯分布的非线性变换：考虑 y=g(x)，x为高斯分布
- 那么 $p\left(y|x\right)=\mathcal{N}\left(g\left(x\right),\boldsymbol{R}\right)$ ，其中g()是一个非线性映射，且受噪声为R的干扰
- 这种情况难以一概而论，试看简单的一维例子。

$$x\sim\mathcal{N}\left(0,\sigma^2\right) \qquad y=\exp(x) \qquad x=\ln(y)$$

- 显然 $$\begin{aligned}1&=\int_{-\infty}^{\infty}p(x)\mathrm{d}x\\&=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right)\mathrm{d}x\\&=\int_{0}^{\infty}\underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2}\frac{(\ln(y))^2}{\sigma^2}\right)\frac{1}{y}}_{p(y)}\mathrm{d}y\end{aligned}$$



- p(y)不再是高斯分布

深蓝学院
shenlanxueyuan.com

高斯概率密度函数

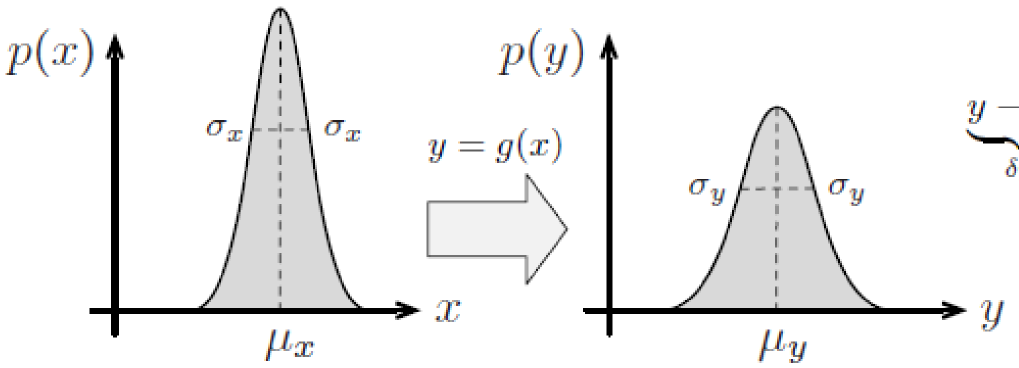
- 很多时候x甚至不能解析地写成y的表达式
- 通常的处理方式是：在某个工作点处（例如x的均值）进行线性化，近似地看成线性变换

$$g(\boldsymbol{x}) \approx \boldsymbol{\mu}_y + \boldsymbol{G}(\boldsymbol{x} - \boldsymbol{\mu}_x)$$

$$\boldsymbol{G} = \left.\frac{\partial \boldsymbol{g}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{\mu}_x}$$

$$\boldsymbol{\mu}_y = \boldsymbol{g}(\boldsymbol{\mu}_x)$$

- G是这个点的雅可比



$$\begin{aligned}
\mu_y &= g(\mu_x) \\
\underbrace{y - \mu_y}_{\delta y} &\approx \underbrace{\left.\frac{\partial g(x)}{\partial x}\right|_{x=\mu_x}}_{a} \underbrace{x - \mu_x}_{\delta x} \\
\sigma_y^2 &= E[\delta y^2] \\
&= a^2 E[\delta x^2] \\
&= a^2 \sigma_x^2
\end{aligned}$$

- 一维情形

深蓝学院
shenlanxueyuan.com

高斯概率密度函数

- 高维情形，设 $p(\boldsymbol{y})=\int_{-\infty}^{\infty} p(\boldsymbol{y}|\boldsymbol{x})\,p(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ $\quad p(\boldsymbol{y}|\boldsymbol{x})=\mathcal{N}(\boldsymbol{g}(\boldsymbol{x}),\boldsymbol{R})$ 有：

$$
\begin{aligned}
p(\boldsymbol{y}) &= \int_{-\infty}^{\infty} p(\boldsymbol{y}|\boldsymbol{x})\,p(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \\
&= \eta \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\boldsymbol{y}-\left(\boldsymbol{\mu}_y+\boldsymbol{G}(\boldsymbol{x}-\boldsymbol{\mu}_x)\right)\right)^{\mathrm{T}} \boldsymbol{R}^{-1}\left(\boldsymbol{y}-\left(\boldsymbol{\mu}_y+\boldsymbol{G}(\boldsymbol{x}-\boldsymbol{\mu}_x)\right)\right)\right) \\
&\qquad \times \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_x)^{\mathrm{T}} \boldsymbol{\Sigma}_{xx}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_x)\right)\mathrm{d}\boldsymbol{x} \\
&= \eta \boxed{\exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}_y)^{\mathrm{T}} \boldsymbol{R}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_y)\right)} \\
&\qquad \times \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_x)^{\mathrm{T}}\left(\boldsymbol{\Sigma}_{xx}^{-1}+\boldsymbol{G}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{G}\right)(\boldsymbol{x}-\boldsymbol{\mu}_x)\right) \\
&\qquad \times \exp\left((\boldsymbol{y}-\boldsymbol{\mu}_y)^{\mathrm{T}} \boldsymbol{R}^{-1}\boldsymbol{G}(\boldsymbol{x}-\boldsymbol{\mu}_x)\right)\mathrm{d}\boldsymbol{x}
\end{aligned}
$$

定义矩阵 $\boldsymbol{F}$，使得：

- ☐ 为什么要这样定义这个F?
- ☐ 为了凑平方的交叉项系数

$$
\boldsymbol{F}^{\mathrm{T}}\left(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{G}+\boldsymbol{\Sigma}_{xx}^{-1}\right)=\boldsymbol{R}^{-1}\boldsymbol{G}
$$

补全积分里的平方项，得：

$$
\begin{aligned}
&\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_x)^{\mathrm{T}}\left(\boldsymbol{\Sigma}_{xx}^{-1}+\boldsymbol{G}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{G}\right)(\boldsymbol{x}-\boldsymbol{\mu}_x)\right) \times \exp\left((\boldsymbol{y}-\boldsymbol{\mu}_y)^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{G}(\boldsymbol{x}-\boldsymbol{\mu}_x)\right) \\
&= \exp\left(-\frac{1}{2}\left((\boldsymbol{x}-\boldsymbol{\mu}_x)-\boldsymbol{F}(\boldsymbol{y}-\boldsymbol{\mu}_y)\right)^{\mathrm{T}}\left(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{G}+\boldsymbol{\Sigma}_{xx}^{-1}\right)\left((\boldsymbol{x}-\boldsymbol{\mu}_x)-\boldsymbol{F}(\boldsymbol{y}-\boldsymbol{\mu}_y)\right)\right) \\
&\quad \boxed{\times \exp\left(\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}_y)^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\left(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{G}+\boldsymbol{\Sigma}_{xx}^{-1}\right)\boldsymbol{F}(\boldsymbol{y}-\boldsymbol{\mu}_y)\right)}
\end{aligned}
$$

- ☐ 第二项与x无关，移出积分。
- ☐ 第1项是关于x的积分，积完后得一个数，与η合并
- ☐ 最后剩下蓝框里的部分

深蓝学院 shenlanxueyuan.com

高斯概率密度函数

- 注意有 $\boldsymbol{F}^{\mathrm{T}}\left(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{G}+\boldsymbol{\Sigma}_{xx}^{-1}\right)=\boldsymbol{R}^{-1}\boldsymbol{G}$
- 而且： $F=(G^TR^{-1}G+\Sigma_{xx}^{-1})^{-T}G^TR^{-T}$
  $=(G^TR^{-1}G+\Sigma_{xx}^{-1})^{-1}G^TR^{-1}$

- 于是得：
$$\begin{aligned}
p(\boldsymbol{y}) &= \rho\exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}_y)^{\mathrm{T}}\left(\boldsymbol{R}^{-1}-\boldsymbol{F}^{\mathrm{T}}\left(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{G}+\boldsymbol{\Sigma}_{xx}^{-1}\right)\boldsymbol{F}\right)(\boldsymbol{y}-\boldsymbol{\mu}_y)\right)\\
&= \rho\exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}_y)^{\mathrm{T}}\left(\underbrace{\boldsymbol{R}^{-1}-\boldsymbol{R}^{-1}\boldsymbol{G}\left(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{G}+\boldsymbol{\Sigma}_{xx}^{-1}\right)^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{R}^{-1}}_{\text{由式（2.75）知：}(\boldsymbol{R}+\boldsymbol{G}\boldsymbol{\Sigma}_{xx}\boldsymbol{G}^{\mathrm{T}})^{-1}}\right)(\boldsymbol{y}-\boldsymbol{\mu}_y)\right)\\
&= \rho\exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}_y)^{\mathrm{T}}\left(\boldsymbol{R}+\boldsymbol{G}\boldsymbol{\Sigma}_{xx}\boldsymbol{G}^{\mathrm{T}}\right)^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_y)\right)
\end{aligned}\tag{2.89}$$

- y的高斯分布为： $\boldsymbol{y}\sim\mathcal{N}\left(\boldsymbol{\mu}_y,\boldsymbol{\Sigma}_{yy}\right)=\mathcal{N}\left(\boldsymbol{g}\left(\boldsymbol{\mu}_x\right),\boldsymbol{R}+\boldsymbol{G}\boldsymbol{\Sigma}_{xx}\boldsymbol{G}^{\mathrm{T}}\right)$

- SMW(a)式： $\left(\boldsymbol{A}^{-1}+\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}\right)^{-1}\equiv\boldsymbol{A}-\boldsymbol{A}\boldsymbol{B}\left(\boldsymbol{D}+\boldsymbol{C}\boldsymbol{A}\boldsymbol{B}\right)^{-1}\boldsymbol{C}\boldsymbol{A}$

深蓝学院
shenlanxueyuan.com

高斯概率密度函数

- 高斯分布的香农信息
- 高斯分布的互信息
- 高斯分布的克拉美罗下界
- 高斯过程

- 以上均略，感兴趣同学可自行研读

深蓝学院
shenlanxueyuan.com

习题

1. 证明高斯分布积分为1，即：$1 = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$
2. 课本习题第1,4,5,6题。