

目录

第一章 线性代数	3
1.1 矩阵空间	3
1.1.1 Subspace 子空间	3
1.1.2 Columnspace 列空间	3
1.1.3 Overdetermined Homogeneous Linear System	4
1.2 方阵的特征值与特征向量	4
1.3 奇异值分解	5
第二章 信号处理	7
2.1 MUSIC 算法	7
2.2 统计信号处理的估计理论	8
2.2.1 参数估计的数学模型	8
2.2.2 估计量的性能	9
2.2.3 估计量的最小均方误差准则	12
2.2.4 Cramer-Rao 下限	13
第三章 深度学习	17
3.1 数据处理	17
3.2 模型设计	19
3.2.1 多层感知机	19
3.2.2 卷积神经网络	19
3.3 训练配置	19
3.3.1 反向传播	19
3.3.2 特征标准化	19
3.4 模型保存	20
3.5 模型预测	20
第四章 通信原理	21
4.1 数字频带传输	21
4.1.1 二进制通断键控 (On-Off Keying)/二进制幅度键控 (2ASK)	21
4.1.2 二进制相移键控 (2PSK)	22
4.1.3 二进制差分相移键控 (2DPSK)	23
4.2 移动通讯	23
4.2.1 5G	23
4.2.2 6G	23

第五章 电子电路	25
5.1 电源	25
5.1.1 同步 Buck	25
5.1.2 PFC	26

第一章 线性代数

1.1 矩阵空间

1.1.1 Subspace 子空间

对于空间 \mathbb{R}^3 ，认为：

- 任何过原点的平面是 \mathbb{R}^3 的子空间；
- 任何过原点的直线是 \mathbb{R}^3 的子空间；
- \mathbb{R}^3 中的两个子空间的并集不能构成子空间；
- \mathbb{R}^3 中的两个子空间的交集可以构成子空间——零空间。

考虑矩阵 A

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix}$$

则矩阵 A 的列空间是 \mathbb{R}^4 的二维子空间。因为 A 的列的线性组合表示了 \mathbb{R}^4 中某个过原点的平面。

1.1.2 Columnspace 列空间

由线性无关的 n 个 m 维列向量组成矩阵 A ，可以把 A 看做是由这些向量为基所张成的空间。 Ax 就是把这些向量按特定的权重（列向量 x ）组合起来，得到的在这个列空间内的向量。

这给认识“方程 $Ax = b$ 无解”提供了一个新的角度：因为 b 不落在 A 的列空间中。所以无法找到一个基的组合，使组合与 b 重合，但仍可以与 b 近似——即组合 Ax 是 b 向矩阵 A 的列空间做的投影，此时一般认为 Ax 最接近 b 。这就是最小二乘法的本质——投影的目的，实际上是使向量 b 到空间 $C(A)$ 的欧式距离最短。¹

若设该投影向量为 p ，则可知方程 $Ax = p$ 有唯一解， p 与 b 的误差向量即 $e = Ax - b$ 。 e 垂直于 $C(A)$ ，或者说： e 位于 A 的左零空间。

利用向量垂直，可以得到使 $\|e\|_2$ 最小的近似解，记作 \hat{x} 。

$$p^T e = 0 \tag{1.1}$$

因而得到

$$(A\hat{x})^T (A\hat{x} - b) = 0$$

¹以 $b \in \mathbb{R}^{3 \times 1}$, $A \in \mathbb{R}^{3 \times 2}$ 为例，则表示平面外一点到平面的最短距离是过该点作到平面的垂线段的长度。

发现, 用 $p \perp e$ 会使方程中含有两个 x 而不易于求解。注意到 A 的每一列都应当与 e 垂直, 因此有

$$A^T(A\hat{x} - b) = 0 \quad (1.2)$$

这里, 由于 A 列满秩, 所以方阵 $A^T A$ 满秩、可逆, 因此解为

$$\hat{x} = (A^T A)^{-1} A^T b \quad (1.3)$$

至此, 问题似乎已经解决: 我们只需一些矩阵乘法和求逆, 即可得出原非齐次超定方程的最小二乘解。

但是, 还有一些值得研究的问题, 例如: 是否有一种特定形式的矩阵, 可以作为一种变换, 把原来的 b 向量投影到 $C(A)$ 上? 这个把 b 投影为 p 的矩阵, 和 A 的关系表示为:

$$p = A\hat{x} = \overset{\text{投影矩阵 } P}{A(A^T A)^{-1} A^T} b \quad (1.4)$$

易发现, 投影矩阵 P 有以下性质: (证明方法直接代入其表达式即可)

1. 是对称方阵。 $P^T = P$
2. “可折叠”。 $P^k = P$ ($k \in \mathbb{N}^+$), 其直观解释: 对于已经在 $C(A)$ 上的向量, 再次作投影, 不发生任何改变。
3. 降秩。 $P \in \mathbb{R}^{m \times m}$, 但其秩不超过 $R(A) = n < m$ 。
4. 有固定的特征值 0 和 1。因为:

$$Px = \begin{cases} x, & x \in C(A) \\ 0, & x \perp C(A) \end{cases} \quad (1.5)$$

1.1.3 Overdetermined Homogeneous Linear System

对于过定齐次方程组

$$[A]_{m \times n} x = 0 \quad (m > n) \quad (1.6)$$

它倾向于只有唯一零解, 因为此时 A 通常是非奇异 (或列满秩) 的。

我们希望得到一个不平凡解 (non-trivial solution) \hat{x} , 要求 $\hat{x} \neq 0$ 但使 Ax 几近于 0 。

这个目标可以建模为如下形式:

$$\begin{aligned} \arg \min_x \|Ax\|_2 \\ \text{s.t. } \|x\|_2 = 1 \end{aligned} \quad (1.7)$$

这个问题有封闭解: \hat{x} 为 $A^T A$ 最小特征值对应的特征向量; 或者说是 A 的最小奇异值对应的奇异向量。

1.2 方阵的特征值与特征向量

特征值/特征向量的一个重要思想, 在于它处理的是这样一种特殊情况——以方阵列为基和以坐标轴 (单位阵的列) 为基, 分别按特征向量加权后组合出的两个向量是平行的——比值为 λ 。即

$$Ax = \lambda x \quad (1.8)$$

特别指出，特征向量不能为零向量。

给定任意一个方阵，我们希望得出其特征值/特征向量，就需要解上式矩阵方程——但是未知量有两个，已知量只有 \mathbf{A} ，怎么办呢？

我们知道，非奇异矩阵的零空间只有零向量。因此把上式整理为关于 \mathbf{x} 的齐次线性方程组，要使得

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0}$$

存在非零解，那么 $(\mathbf{A} - \lambda \mathbf{I})$ 只能是奇异矩阵。

因此

$$|\mathbf{A} - \lambda \mathbf{I}| = 0 \quad (1.9)$$

该式被称为特征方程 (eigenvalue equation)。显然，它是关于 λ 的一元 n 次方程。

好消息是这意味着 n 阶方阵有且仅有 n 个特征值，坏消息是，特征值可能很难求——五次及以上的方程没有根式解。

零空间的求法：对矩阵 \mathbf{A} 进行消元求得主变量和自由变量；给自由变量赋值得到特解；对特解进行线性组合

1.3 奇异值分解

矩阵的奇异值分解 (singular value decomposition) 将是最重要的矩阵分解方式。

- 矩阵是实对称方阵：特征向量正交
- 矩阵是共轭对称方阵 (Hermitian Matrix)：特征值为实数，且特征向量正交
-

第二章 信号处理

2.1 MUSIC 算法

设有远场区信号源 $X_1, X_2 \in \mathbb{C}$ ，其中信号 $X_i, i = 1, 2$ 到达 $N_a = 3$ 的 ULA，由阵元间距 d 和入射角 θ_i 引起的相位差为 ϕ_i ，则

$$\phi_i = \frac{2\pi}{\lambda} d \cos \theta_i = 2\pi f \frac{d \cos \theta_i}{c} \quad (2.1)$$

为计算方便，把这个相邻阵元接收信号的相位差用复数表示：

$$\Phi_i = e^{j\phi_i} \quad (2.2)$$

设 $\gamma_i \in \mathbb{C}$ 表示信源 i 传播到 ULA 的参考阵元（通常就取第一个阵元）的过程中引起的幅值和相位的变化，设阵元 j 处收到的信号为 Y_j ，则

$$Y_j = \gamma_1 X_1 \Phi_1^{(j-1)} + \gamma_2 X_2 \Phi_2^{(j-1)} \quad j = 1, \dots, N_a = 3 \quad (2.3)$$

注意， Y_j 和 X_i 本身都是时域信号的采样点，因此是向量的形式。

写成矩阵形式：

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \Phi_1 & \Phi_2 \\ \Phi_1^2 & \Phi_2^2 \end{bmatrix} \begin{bmatrix} \gamma_1 X_1 \\ \gamma_2 X_2 \end{bmatrix} \quad (2.4)$$

Y 矩阵 Φ 矩阵

在实际问题中，往往已知 \mathbf{Y} ，而信源和传播路径未知，要由此直接解出 Φ 似乎不可能。

回想一下学过的数学知识，在满足什么样的条件时，我们可以仅通过一个已知量 b ，就确定方程

$$k_1 x_1 + k_2 x_2 = b$$

的参数 k 呢？

在线性代数里，如果两个向量线性无关，那么这两个向量按任意不全为零的系数 k_1, k_2 加权求和都不可能得到零向量。

因此，如果这里 b 是零向量，而 x_1, x_2 满足线性无关时，就只有 $k_1 = k_2 = 0$ 能满足原方程。换言之，在涉及线性无关向量组的运算中，如果把方程整理（或寻找一种变换）使等式右边为 0，左边为线性无关向量做 linear combination 的形式，那么就必然得出加权系数为 0 的结论。

受以上启发，我们对问题增加新的限制条件：**各个信源信号彼此是线性无关的**。则此处 γX_1 和 γX_2 线性无关。对式 (2.4) 两边左乘加权系数向量 $\mathbf{C} \in \mathbb{C}^{N_a}$ ：

$$c_1 Y_1 + c_2 Y_2 + c_3 Y_3 = \begin{bmatrix} c_1 & c_2 & c_3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \Phi_1 & \Phi_2 \\ \Phi_1^2 & \Phi_2^2 \end{bmatrix} \begin{bmatrix} \gamma_1 X_1 \\ \gamma_2 X_2 \end{bmatrix} \quad (2.5)$$

我们需要找到合适的 $\mathbf{C} \neq 0$, 使 $\mathbf{C}\mathbf{Y} = 0$ 。如果这样的 \mathbf{C} 存在, 那么加权系数向量 $\mathbf{C}\Phi$ 就是零向量, 进而求出 Φ 。

那么这样的 \mathbf{C} 存不存在呢? 事实上, 由于天线数量 N_a 大于信源数量, 即 Φ 的行数大于列数, 因此矩阵 $\mathbf{Y} = \Phi\mathbf{X}$ 的秩一定不大于 Φ 的列数。即 $R(\mathbf{Y}) \leq \min\{R(\Phi), R(\mathbf{X})\} \leq 2$ 。¹ 因此矩阵 \mathbf{Y} 行降秩, 所以 Y_1, Y_2, Y_3 必然是线性相关的, 存在不全为零的系数 c_1, c_2, c_3 , 使

$$c_1 Y_1 + c_2 Y_2 + c_3 Y_3 = 0$$

对于有 N_s 个采样点的接收信号组成的矩阵 $\mathbf{Y}_{N_s \times 3}$, 如何求解 $\mathbf{C}\mathbf{Y} = 0$? 显然, 该方程

$$\begin{bmatrix} Y_1[1] & Y_2[1] & Y_3[1] \\ Y_1[2] & Y_2[2] & Y_3[2] \\ \vdots & \vdots & \vdots \\ Y_1[N_s] & Y_2[N_s] & Y_3[N_s] \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (2.6)$$

是关于 \mathbf{C} 的超定 (overdetermined) 方程组, 而且还是齐次 (homogenous) 的。

超定方程组通常无解, 我们只能得到近似解, 并选取一个量化误差的指标使近似解最优。最常用的指标是平方和 (或者欧氏距离), 这种求近似解的方法就是**最小二乘法**。

2.2 统计信号处理的估计理论

本章节介绍从含有随机性的时间序列 (以下简称随机信号) 中提取未知参数的问题。

2.2.1 参数估计的数学模型

几乎任何来自物理世界的的数据都具有随机性, 一组时间序列可以看作是参数服从特定分布的随机过程 \mathbf{x} 的一个实现, 记作 \mathbf{x} 。为了简化问题, 通常认为随机过程 \mathbf{x} 是各态历经(Ergodic) 的, 即随机过程中的任何一个样本函数 \mathbf{x} 都含有随机过程的全部统计信息。

通常问题为: 需要通过观测到的随机信号 \mathbf{x} 来估计 \mathbf{x} 的某个或某些参数, 参数用向量 θ 表示。

为了得到更优的估计, 我们应确定一个合适的估计量(Estimator), 记作 $\hat{\theta}$, 它是关于样本数据的函数, 并应当充分利用样本的信息。

至此, 问题描述为: 随机信号 \mathbf{x} 服从参数为 θ 的某分布, 已知 \mathbf{x} 的样本 $[x(0), x(1), \dots, x(N-1); \theta]$, 求尽可能²接近真值 θ 的估计量 $\hat{\theta}$ 。

求解参数估计问题的两个关键:

1. 对随机信号 \mathbf{x} 建模。对 \mathbf{x} 的分布类型提出一个合理的假设, 即得出 \mathbf{x} 服从的 PDF。模型的选择要求其既方便数学推导出闭合形式的估计量, 又贴合实际问题。
2. 确定最佳/准最佳估计量。估计量 $\hat{\theta}$ 通常是关于样本数据和其他已知参数的函数, 它应当按你所希望的方式接近其真值 θ 。

对于第一个问题, 通常有两种方式得出随机信号的 PDF。

经典方法是认为参数 θ 是一组确定的未知数。与之相对的是贝叶斯 (Bayesian) 估计, 认为参数本身是服从某种分布 (需要依据先验知识假定, 例如均匀分布) 的随机变量, 因此它具有先验 PDF。这里用两个例子说明之。

¹ Φ 可以看作是范德蒙矩阵的一部分, 此处只要 $\theta_1 \neq \theta_2$ 则 Φ 就是列满秩的; 而由于线性无关所以 \mathbf{X} 行满秩, 因此它们乘积的秩一般而言为 2, 即信源个数。

²我们希望估计量越接近真值越好, 但怎样评估估计量的性能, 以得到最适合当前问题的估计量, 这个问题将在下一节讨论。

例 2.1: 服从正态分布的单随机变量均值估计概率模型

对于长度为 1 的随机信号 $\mathbf{x} = x(0)$, 其值服从方差为 σ^2 的正态分布, 现需要估计其均值。求随机信号 \mathbf{x} 的 PDF。

解:

在经典方法中, 认为分布的参数是未知的确定量, 因此这里无论是估计均值还是估计方差, 或者两者, 都不影响 PDF 的形式——把所有参数当做已知量即可。PDF 为

$$p(\mathbf{x}; \boldsymbol{\theta}) = p(x(0); \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x(0) - \theta)^2}{2\sigma^2} \right] \quad (2.7)$$

须指出, $p(\mathbf{x}; \boldsymbol{\theta})$ 的分号表示该 PDF 是依赖于给定向量 $\boldsymbol{\theta}$ 的一族 PDF 曲线。而 $p(\mathbf{x}, \boldsymbol{\theta})$ 中的逗号表示这是一个联合概率密度函数, 联合的随机变量用逗号分隔。

例 2.2: 服从正态分布的多变量均值估计概率模型

现收集有某地连续 N 天的道琼斯 (Dow-Jones) 工业平均指数的数据 $x(0), x(1), \dots, x(N-1)$, 数据整体随时间呈上升趋势, 但细节仍有较大起伏。求该地区道琼斯工业平均指数服从的 PDF。

解:

我们用直线和高斯白噪声 (White Gaussian Noise, WGN) 拟合数据, 设

$$x(n) = \theta_1 + \theta_2 n + w(n) \quad n = 0, 1, \dots, N-1 \quad (2.8)$$

从经典估计的角度来看, 直线 $\theta_1 + \theta_2 n$ 是时域一条确定的直线, 每个数据点都是在这条定直线的基础上加一个零均值、方差为 σ^2 的随机数得到, 换言之, 每个样本点中随机的成分正是正态随机变量 $w(n)$ 的实现。而从贝叶斯估计的角度, 由于参数是随机变量, 每次对随机信号的观测都对应一条不同的直线, 每一个样本点应当视作是 $\theta_1, \theta_2, w(n)$ 的共同实现。

须指出, 模型如式 (2.8) 不依赖于估计方法, 这是为了贴合实际问题作出的人为假设。不同估计方法造成的差异体现在 PDF 的推导上。

对于这种含有加性白噪声的数字信号模型, 通常随机信号看作一组联合的随机变量。

因此经典估计认为 \mathbf{x} 服从的 PDF 为

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x(n) - \theta_1 - \theta_2 n)^2 \right] \quad (2.9)$$

而贝叶斯估计认为 \mathbf{x} 服从的 PDF 为

$$p(\mathbf{x}, \boldsymbol{\theta}) = p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \quad (2.10)$$

其中 $p(\boldsymbol{\theta})$ 是先验 PDF, 表示在数据观测以前关于 $\boldsymbol{\theta}$ 的先验知识; $p(\mathbf{x}|\boldsymbol{\theta})$ 是条件 PDF, 表示在给定 $\boldsymbol{\theta}$ 的前提下由数据 \mathbf{x} 提供的知识。

2.2.2 估计量的性能

已经指出, 估计量是样本值的函数:

$$\hat{\boldsymbol{\theta}} = g(\mathbf{x})$$

因此估计量是一个随机变量，它的性能由其 PDF 完全描述，并且也只能从统计的角度描述。

通常从以下几个方面考察估计量的性能：

1. 从估计量与样本数量 n 的关系考虑：

估计量的 PDF 随观测次数的变化的性质需引起重视，因为这决定了我们增加试验次数是否真的能使估计的结果——如果是，那么可以适当增加观测数；否则只是在浪费时间。

我们用随机变量序列 $\hat{\theta}_n$ 来表示估计量对观测数 n 的依赖关系。例如其中 $\hat{\theta}_i$ 表示用 $n = i$ 个样本来估计真值 θ 的一个估计量。

定义 2.1: 一致估计量

给定 n 个观测 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ ，估计量 $\hat{\theta}_n$ 是这 n 个观测的函数。若对于真值的所有可能取值 θ ， $\hat{\theta}_n$ 都满足依概率收敛于真值，则称该估计量为一致 (consistent) 估计量（或称为相合估计量）。即

$$\lim_{n \rightarrow \infty} \Pr \left\{ |\hat{\theta}_n - \theta| \geq \epsilon \right\} = 0 \quad (2.11)$$

其中 ϵ 为任意小的正数。

“ $\hat{\theta}_n$ 依概率收敛于 θ ”有时也记作：

$$\text{p} \lim_{n \rightarrow \infty} (\hat{\theta}_n) = \theta \quad (2.12)$$

在思考这个概念时，应当把估计量的 PDF 看做是依赖于 n 的一族曲线，而一致性 (consistency)则是描述了这一族曲线随着 $n \rightarrow \infty$ 逐渐压缩成一个冲激函数（即收敛于一个具体的数值 θ ）或者逐渐形成某个特殊分布的形状²（即收敛于一个已知参数的随机变量 θ ）。

2. 从估计量的期望考虑：

定义 2.2: 无偏估计量

若对于真值的所有可能取值 θ 以及任意有限的观测次数 n ，估计量 $\hat{\theta}_n$ 的期望¹都与之相等，则称该估计量为无偏 (Unbiased) 估计量。即

$$E [\hat{\theta}_n] = \theta, \quad n = 1, 2, \dots \quad (2.13)$$

反之，不满足以上条件的估计量为有偏 (Biased) 估计量，称其期望与真值的差值为偏差（或称偏误），写作

$$\text{b}(\hat{\theta}_n) = E [\hat{\theta}_n] - \theta \quad (2.14)$$

3. 综合以上 1、2 考虑：

²TODO: 这是贝叶斯学派的观点，是这样吗？

¹TODO: 又一说为已知 θ 的条件期望？

定义 2.3: 渐进无偏估计量

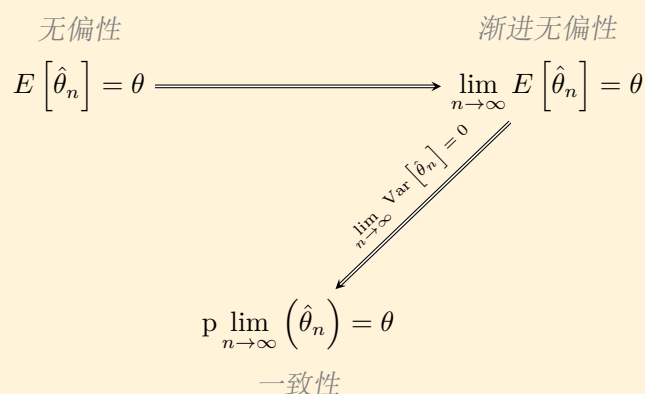
给定 n 个观测 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, 估计量 $\hat{\theta}_n$ 是这 n 个观测的函数。若对于真值的所有可能取值 θ , 估计量的期望在 $n \rightarrow \infty$ 时收敛于真值, 则称该估计量为渐进无偏 (asymptotic unbiased) 估计量。即

$$\lim_{n \rightarrow \infty} E[\hat{\theta}_n] = \theta \quad (2.15)$$

❗ 注意

无偏、一致、渐进无偏三者的区别和联系。

- 估计量的无偏性与样本个数 n 无关, 它是一个有限样本特性, 也称小样本特性。
- 估计量的渐进无偏性需要在 $n \rightarrow \infty$ 时根据 $E[\hat{\theta}]$ 的极限值判断, 与一致性一同被称为无限样本特性 (或大样本特性)。
- 若估计量的期望等于真值 (无偏性), 则其极限必然也为真值 (渐进无偏性)。
- “依概率收敛”是一个相对宽松的条件。估计量依概率收敛于真值 (一致性), 并不能得出估计量的期望在 $n \rightarrow \infty$ 的极限等于真值 (渐进无偏性)! (可以举一个让期望趋于无穷的反例, 见《概率导论》P298 例 5.7)
- 若估计量的方差满足在 $n \rightarrow \infty$ 的极限为 0, 且估计量是渐进无偏的, 则估计量具有一致性。这是因为“依均方收敛”的条件强于“依概率收敛”。



4. 从估计量的方差考虑:

估计量的方差越小, 说明估计结果越稳定; 但估计量仍可能是有偏的、也可能不以概率收敛。

5. 从估计量到真值的均方误差考虑:

定义 2.4: 估计量的均方误差

估计量 $\hat{\theta}$ 到真值 θ 的均方误差 (Mean Square Error, MSE), 记作 $\text{mse}(\hat{\theta})$, 定义为随机变量 $(\hat{\theta} - \theta)^2$ 的期望。即

$$\text{mse}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] \quad (2.16)$$

均方误差同时考虑到了估计量的偏差和方差所带来的误差。因此能更全面地体现估计量的性能。

2.2.3 估计量的最小均方误差准则

目的：找到用来估计未知确定性参数的最佳估计量。

如何定义“最佳”？我们需要给出一个量化的准则，这里使用的就是最小均方误差准则。使估计量到真值的 MSE 最小的估计量，可以称为最小均方误差 (Minimum MSE) 估计量，但为什么研究的却是最小方差无偏 (Minimum Variance Unbiased, MVU) 估计量呢？

这是因为最小均方误差 (MMSE) 估计量通常是不可实现的：

$$\begin{aligned} \text{mse}(\hat{\theta}) &= E \left\{ \left[(\hat{\theta} - E(\hat{\theta})) + (E(\hat{\theta}) - \theta) \right]^2 \right\} \\ &= E \left\{ (\hat{\theta} - E(\hat{\theta}))^2 \right\} + 2E \left\{ \hat{\theta} - E(\hat{\theta}) \right\} \cdot b(\hat{\theta}) + b^2(\hat{\theta}) \\ &= \text{Var}(\hat{\theta}) + b^2(\hat{\theta}) \end{aligned} \quad (2.17)$$

这是一个很重要的公式。它说明估计量到真值的 MSE 是估计量的方差和偏差引起的。在很多统计问题中都存在等式右边两项的权衡——方差的减少总是伴随着偏差的增大。

但也可以看到，MMSE 准则因为引入真值参与运算而使之成为一般不可实现的估计量¹，它不能写成只关于样本数据的函数——估计量的偏差还是关于真值的函数。

为了避免不可实现的估计量，我们约束偏差为零，即在无偏估计的前提下使估计量的方差最小，得到最小方差无偏 (MVU) 估计量。

MVU 估计量的 MSE 就是它的方差，而方差作为一个统计量可以由样本数据得出¹，因此 MUV 估计量是可实现的估计量，下面讨论其存在性。

MVU 估计量的存在性

MVU 估计量要存在，首先应满足无偏估计量的存在性，然后在所有的无偏估计量中，若对所有可能的真值取值 θ ，都存在某个无偏估计量的方差始终比其他无偏估计量的方差小，则这个估计量就是 MVU 估计量。

你可能会问：估计量的方差为什么会随真值 θ 取值变化而变化？前面已经提到，观测次数 n 可能会影响估计量的 PDF 曲线，这是因为估计量的构造和样本数目有关。而事实上参数真值影响了样本的分布，因此估计量的 PDF 也会随之变化。

MVU 估计量的求法

目前有一种总是能得到 MVU 估计量的“转动摇把”(turn-the-crank) 的方法。后面的章节将详细讨论如何在已知存在性时，求出 MVU 估计量，这里只作为一个引子简单介绍。

法一：确定 Cramer-Rao 下限 (Cramer-Rao lower bound, CRLB)，然后检查估计量是否满足 Cramer-Rao 下限。

法二：应用 Rao-Blackwell-Lehmann-Scheffe(RBLS) 定理。

法三：进一步限制无偏估计量是线性的，然后找到线性无偏约束下的最小方差估计。

¹在几种极端情况下是可实现的。见《统计信号处理》P17

¹样本方差是方差的无偏估计量。

2.2.4 Cramer-Rao 下限

我们希望对任何无偏估计量，都能确定其方差的下限。在最理想的情况下，对于未知参数的所有取值，如果某个无偏估计量的方差都达到了此下限，那么它就是 MVU 估计量。在最恶劣的情况下，即使不存在 MVU 估计量，它也能比较无偏估计量的性能提供一个标准。

Cramer-Rao 下限 (CRLB)，由 Harald Cramer 和 Callyampudi Radhakrishna Rao 得名。事实上，历史上的研究者提出了很多种估计量的方差下限，但是 CRLB 是最容易确定的。

CRLB 给出了一个确定性参数的无偏估计量的方差下限。它的最简单形式是：任何无偏估计的方差都不小于 Fisher 信息的倒数。一个达到了方差下限的无偏估计被称为完全高效的 (fully efficient)。

给定估计量偏差，CRLB 还可以用于确定有偏估计的界限。在一些情况下，有偏估计的 MSE 可能小于无偏估计的 CRLB。

关于参数估计的精度考虑

前面已经提到，对于某个具体的估计量，它的 PDF 越集中于真值附近（对于一般估计量来说就是 MSE 越小，对无偏估计量来说就是方差越小），则估计精度越高。这说的是不同估计量的性能问题，但实际上除了估计量的选取影响了估计精度外，问题的概率模型自身——或者说随机信号的 PDF，才是限制估计精度的首要原因。

例如，如果数据服从的 PDF 对未知参数的依赖性较弱，或者考虑最极端的情况：PDF 根本与未知参数无关，那么我们不可能以任何精度来估计参数。一般而言，PDF 受未知参数的影响越大，所得估计越好。下面的例子说明了这一点。

例 2.3: 比较从不同功率的 WGN 中估计 DC 电平的精度

设有 WGN $w(n) \sim \mathcal{N}(0, \sigma^2)$ 。随机信号建模为一个未知但确定的 DC 量 A 与 WGN 的叠加。现每次观测单个样本：

$$x(0) = A + w(0)$$

提出 A 的一个无偏估计量 \hat{A} ，比较当 σ^2 分别取值 σ_1^2 和 σ_2^2 时 ($\sigma_1^2 < \sigma_2^2$)，估计量 \hat{A} 的精度。如何量化这种“精度”？

解：

$$E[x(0)] = A + E[w(0)] = A$$

因此 $\hat{A} = x(0)$ 是无偏估计量。

为了比较随机信号 PDF 对 \hat{A} 的估计精度的影响，先写出 PDF：

$$p_i(x(0); A) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left[-\frac{1}{2\sigma_i^2} (x(0) - A)^2 \right] \quad i = 1, 2$$

该式既是估计量 \hat{A} 的 PDF，也是随机信号的 PDF，它们都依赖于真值 A 。

从该式看出，当 σ^2 取值越大， A 对 PDF 的影响越不明显。极端情况当 $\sigma \rightarrow \infty$ 时，PDF 几乎与 A 无关。

下面引入似然函数的概念，以便于定量说明该问题。如果 $x(0)$ 给出，把 PDF 看作是关于 A 的函数，就得到了似然函数，如下图图 2.1 所示。

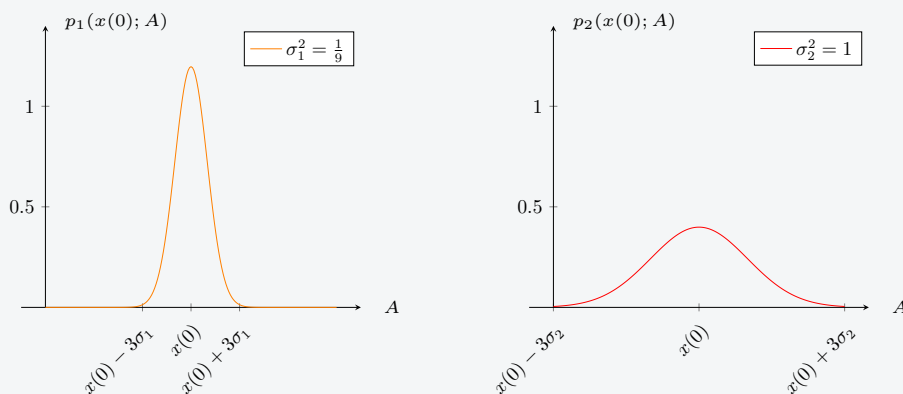


图 2.1: 不同噪声功率下未知参数的似然函数

直观地看，似然函数的“尖锐程度”体现了我们估计未知参数的精度。而“尖锐程度”可由函数的曲率刻画。因此提出用函数曲率来表示估计精度。

考察对数似然函数的曲率——即似然函数的对数的二阶负导数：

$$-\frac{\partial^2 \ln p(x(0); A)}{\partial A^2} = -\frac{\partial}{\partial A} \left[\frac{1}{\sigma^2} (x(0) - A) \right] = \frac{1}{\sigma^2} \quad (2.18)$$

该式说明正态分布曲线在极值点处的曲率等于方差的倒数。因此方差越大，曲率越小，曲线越平缓，估计的精度越差。

在这个例子中，曲率只和方差有关，与观测值无关，这只是因为 PDF 中两者是分离的，但一般来说对数似然函数的曲率也是关于样本 $x(n)$ 的函数——也就是说，似然函数是样本的随机变量函数，并且依赖于未知参数，所以它的对数的曲率也一般是关于样本和参数的函数。

因此我们先对 $p(x(0); A)$ 取期望，使它仅为 A 的函数。得到对数似然函数曲率更一般的度量：

$$-E \left[\frac{\partial^2 \ln p(x(0); A)}{\partial A^2} \right] \quad (2.19)$$

在这个例子中，估计量的方差

$$\text{Var} [\hat{A}] = \sigma^2 = \frac{1}{-E \left[\frac{\partial^2 \ln p(x(0); A)}{\partial A^2} \right]}$$

验证了“对数似然函数曲率越大，估计精度越高”的结论。

CRLB 定理

有例 2.3 的铺垫，我们已经知道了无偏估计量的方差与对数似然函数曲率有关，而 CRLB 定理正是从原理上证明在一定条件下方差和曲率的不等关系——后者的倒数是前者的 Cramer-Rao 下限。

定理 2.1: 标量参数的 CRLB 定理

若 PDF $p(\mathbf{x}; \theta)$ 满足正则条件: 对于所有可能的 θ , 有

$$E \left[\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right] = 0 \quad (2.20)$$

则任何无偏估计量 $\hat{\theta}$ 的方差必定满足:

$$\text{Var} [\hat{\theta}] \geq \frac{1}{-E \left[\frac{\partial^2 \ln p(\mathbf{x}; A)}{\partial A^2} \right]} \quad (2.21)$$

其中期望是对 $p(\mathbf{x}; \theta)$ 求的, 偏导数是在真值处求取的。

而且, 对于某个函数 g 和 I , 当且仅当

$$\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} = I(\theta)(g(\mathbf{x}) - \theta) \quad (2.22)$$

时, 存在 MVU 估计量 $\hat{\theta} = g(\mathbf{x})$, 对于所有可能的真值取值, 都能达到方差下限 $\frac{1}{I(\theta)}$ 。

第三章 深度学习

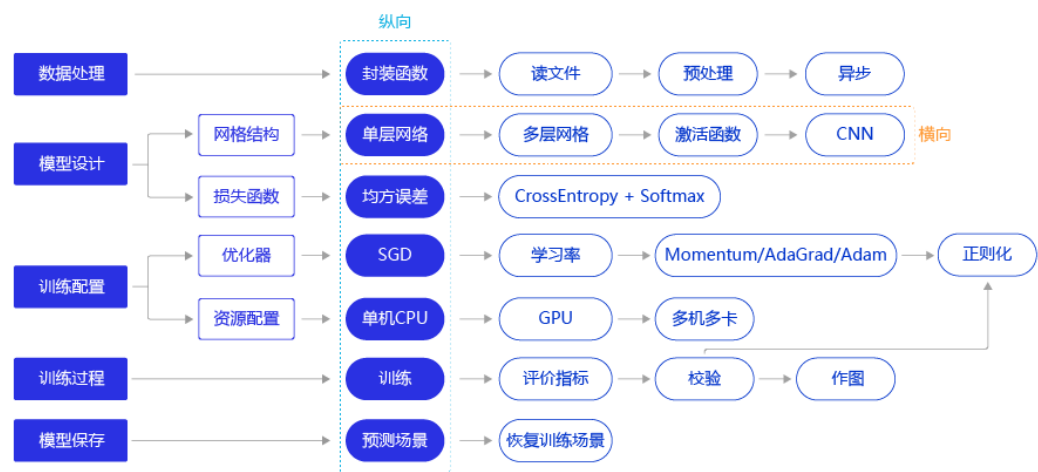


图 3.1: 深度学习模型实现流程

3.1 数据处理

- 读入数据
- 划分数据集
 - train_set (训练集): 用于确定模型参数。
 - val_set (验证集): 用于调节模型超参数 (如多个网络结构、正则化权重的最优选择)。
 - test_set (测试集): 用于估计应用效果 (尽量使用未在模型中应用过的数据)。
- 训练样本集乱序¹
 - 建立顺序编号的 ID 集合 index_list
 - 将 index_list 乱序
 - 按乱序后的顺序读取数据
- 生成批次数据
 - 设置合理的 batch_size
 - 将数据 shape 变为 (batch_size, image_size) 的 np.array
 - 使用 python 的 yield 生成器返回数据
- 校验数据有效性
 - 机器校验: assert 图片与标签维度相同
 - 人工校验: print 数据看格式是否符合预期

某基于 MNIST 数据集的数据生成器的代码实现如下。

```

1  # 定义数据生成器, 返回批次数据
2  def load_data(mode='train'):
3      datafile = './work/mnist.json.gz'
4      print('loading mnist dataset from {} .....'.format(datafile))
5      # 加载json数据文件
6      data = json.load(gzip.open(datafile))
7      print('mnist dataset load done')
8
9      # 读取到的数据区分训练集, 验证集, 测试集
10     train_set, val_set, eval_set = data
11     if mode=='train':
12         # 获得训练数据集
13         imgs, labels = train_set[0], train_set[1]
14     elif mode=='valid':
15         # 获得验证数据集
16         imgs, labels = val_set[0], val_set[1]
17     elif mode=='eval':
18         # 获得测试数据集
19         imgs, labels = eval_set[0], eval_set[1]
20     else:
21         raise Exception("mode can only be one of ['train', 'valid', 'eval']")
22     print("训练数据集数量: ", len(imgs))
23
24     # 校验数据
25     imgs_length = len(imgs)
26
27     assert len(imgs) == len(labels), \
28         "length of train_imgs({}) should be the same as train_labels({})".format(len(imgs), len(
29                                     labels))
30
31     # 获得数据集长度
32     imgs_length = len(imgs)
33
34     # 定义数据集每个数据的序号, 根据序号读取数据
35     index_list = list(range(imgs_length))
36     # 读入数据时用到的批次大小
37     BATCHSIZE = 100
38
39     # 定义数据生成器
40     def data_generator():
41         if mode == 'train':
42             # 训练模式下打乱数据
43             random.shuffle(index_list)
44             imgs_list = []
45             labels_list = []
46             for i in index_list:
47                 # 将数据处理成希望的类型
48                 img = np.array(imgs[i]).astype('float32')
49                 label = np.array(labels[i]).astype('float32')
50                 imgs_list.append(img)
51                 labels_list.append(label)
52                 if len(imgs_list) == BATCHSIZE:
53                     # 获得一个batchsize的数据, 并返回
54                     yield np.array(imgs_list), np.array(labels_list)
55                     # 清空数据读取列表
56                     imgs_list = []
57                     labels_list = []
58
59     # 如果剩余数据的数目小于BATCHSIZE,

```

```

59     # 则剩余数据一起构成一个大小为 len(imgs_list) 的 mini-batch
60     if len(imgs_list) > 0:
61         yield np.array(imgs_list), np.array(labels_list)
62     return data_generator

```

3.2 模型设计

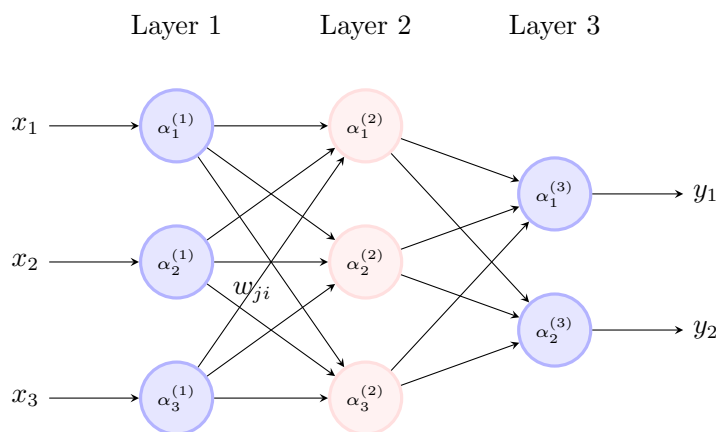
3.2.1 多层感知机

多层感知器 (Multi Layer Perceptron, 即 MLP) 除了一个输入层和一个输出层以外, 包括至少一个隐藏层。多层感知器用来模拟一个非线性函数。

3.2.2 卷积神经网络

3.3 训练配置

3.3.1 反向传播



- 计算梯度

$$\mathbf{g} \leftarrow \partial_{(\mathbf{W}, b)} \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} L(\mathbf{x}^{(i)}, \mathbf{y}^{(i)}, \mathbf{W}, b) \quad (3.1)$$

- 更新参数

$$(\mathbf{W}, b) \leftarrow (\mathbf{W}, b) - \eta \mathbf{g} \quad (3.2)$$

3.3.2 特征标准化

为了解决输入样本中各个维度的数值差异较大导致 error surface 形状扭曲, 给 gradient descent 带来困难的问题, 引入特征标准化 (Feature Normalization) 的方法。

- Batch Normalization
- Layer Normalization
- Instance Normalization
- Group Normalization

- Weight Normalization
- Spectrum Normalization

其中一个常用的标准化方法为 Batch Normalization (BN) 。BN 一次对一个 Batch (记作 \mathcal{B} , 其中第 i 个样本为 $N_e \times 1$ 列向量 $\mathbf{x}^{(i)}, i = 1, \dots, N_B$) 的输入样本的每一个维度作 standard deviation normalization:

$$\tilde{\mathbf{x}}^{(i)} = \frac{\mathbf{x}^{(i)} - \mathbb{E}[\mathbf{x}^{(\cdot)}]}{\sqrt{\text{Var}[\mathbf{x}^{(\cdot)}]}} \quad (3.3)$$

其中, $\mathbb{E}[\mathbf{x}^{(\cdot)}]$ 表示对一个 Batch 内全部样本在同一个维度上求均值, 得到一个 $N_e \times 1$ 列向量。求方差 $\text{Var}[\mathbf{x}]$ 的过程同理。

3.4 模型保存

3.5 模型预测

第四章 通信原理

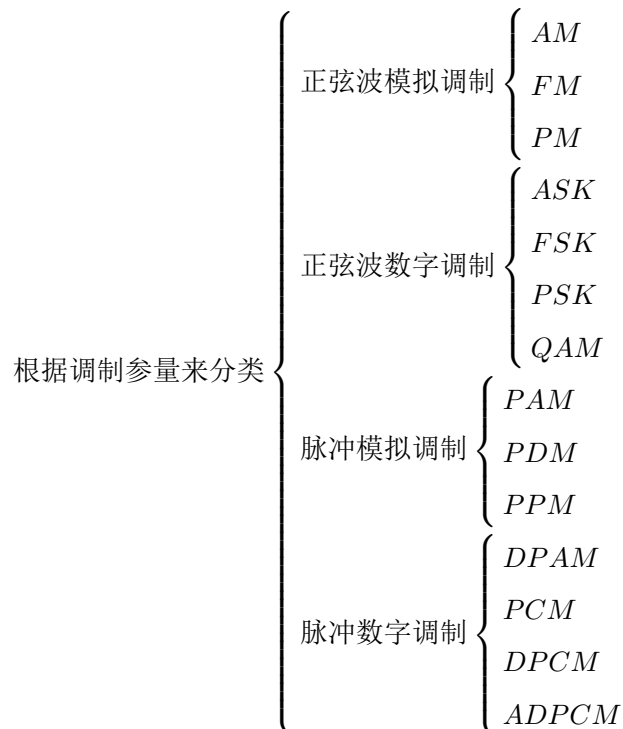
4.1 数字频带传输

Modulation: 用要传输的原始信号 $m(t)$ 控制高频载波信号的某一参量，使它随 $m(t)$ 发生变化。这个原始信号 $m(t)$ 称为 调制信号 或 基带信号。

- 模拟调制： $m(t)$ 为连续时间函数。
- 数字调制： 调制信号为离散时间序列。

调制可以根据高频载波信号 $C(t)$ 类型不同分类：

- 正弦波调制： 正弦信号作为载波
- 脉冲调制： 周期性脉冲作为载波



4.1.1 二进制通断键控 (On-Off Keying)/二进制幅度键控 (2ASK)

- 传号时有一定幅度，空号时幅度为 0。
- 可认为相位随机。

- 调制信号带宽是基带信号的 2 倍。
-

4.1.2 二进制相移键控 (2PSK)

2PSK: 用二进制数字基带信号控制正弦载波的相位。即码元的载波相位表示数字信息。

设数字基带信号为双极性非归零信号 $s(t)$, 则 2PSK 信号调制过程

$$S_{2PSK}(t) = s(t) \cos \omega_c t = \begin{cases} A \cos(\omega_c t) & \text{“传号”} \\ -A \cos(\omega_c t) & \text{“空号”} \end{cases} \quad 0 \leq t \leq T_b \quad (4.1)$$

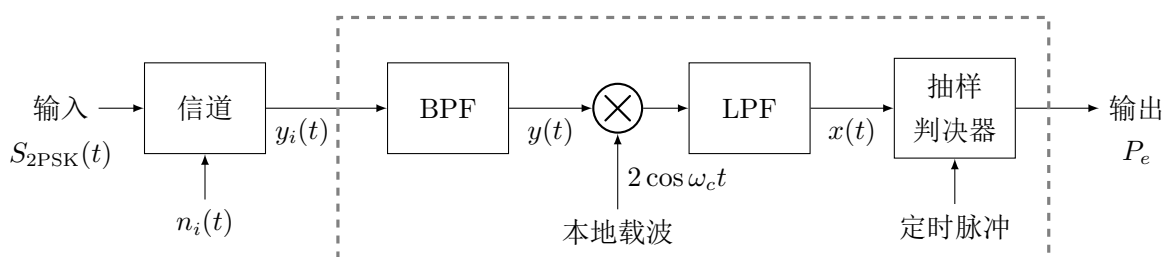


图 4.1: 2PSK 相干解调系统

经过 BPF, 加性高斯白噪声变为窄带高斯噪声, 有效信号的部分不变:

$$y(t) = \begin{cases} [k_B A + n_c(t)] \cos \omega_c t - n_s(t) \sin \omega_c t, & \text{“1”} \\ [-k_B A + n_c(t)] \cos \omega_c t - n_s(t) \sin \omega_c t, & \text{“0”} \end{cases} \quad 0 \leq t \leq T_b \quad (4.2)$$

本振信号与 BPF 输出相乘 (注意本振信号相位可能未对准), 并经过 LPF:

$$\begin{aligned} x(t) &= \text{LPF} \{ 2[\pm k_B A + n_c(t)] \cos(\omega_c t) \cos(\omega_c t + \Delta\theta) - 2n_s(t) \sin(\omega_c t) \cos(\omega_c t + \Delta\theta) \} \\ &= \text{LPF} \{ 2[\pm k_B A + n_c(t)] [\cos^2(\omega_c t) \cos \Delta\theta - \cos(\omega_c t) \sin(\omega_c t) \sin \Delta\theta] \\ &\quad - 2n_s(t) [-\sin^2(\omega_c t) \sin \Delta\theta + \sin(\omega_c t) \cos \omega_c t \cos \Delta\theta] \} \\ &= k_L [\pm k_B A + n_c(t)] \cos \Delta\theta + k_L n_s(t) \sin \Delta\theta \end{aligned} \quad (4.3)$$

理想情况下 $\Delta\theta = 0$,

$$x(t) = \begin{cases} k_L [k_B A + n_c(t)], & \text{“1”} \\ k_L [-k_B A + n_c(t)], & \text{“0”} \end{cases} \quad 0 \leq t \leq T_b \quad (4.4)$$

最恶劣情况下 $\Delta\theta = \pi$,

$$x(t) = \begin{cases} k_L [-k_B A - n_c(t)], & \text{“1”} \\ k_L [k_B A - n_c(t)], & \text{“0”} \end{cases} \quad 0 \leq t \leq T_b \quad (4.5)$$

对于 2PSK 信号, 这种情况导致解调系统反相工作。因此本地载波相位会导致输出绝对码反相, 这种现象称为相位模糊。

4.1.3 二进制差分相移键控 (2DPSK)

为解决 2PSK 信号存在的相位模糊问题, 引入 2DPSK。

设数字基带信号为双极性非归零信号 $s(t)$, 对应的绝对码记作 $\{a_n\}$, 经过差分编码得到相对码 $\{b_n\}$:

$$b_n = a_n \oplus b_{n-1} \quad (4.6)$$

其中 \oplus 表示模二加, 这里等价于异或运算。 $\{b_n\}$ 的初始参考设为 0。

二进制下, 相对码的“0”和“1”分别表示

$$b_n = \begin{cases} 0, & \varphi_n - \varphi_{n-1} = 0 \\ 1, & \varphi_n - \varphi_{n-1} = \pm\pi \end{cases} \quad (4.7)$$

反过来由相对码得出绝对码的过程称为差分译码:

$$a_n = b_n \oplus b_{n-1} \quad (4.8)$$



图 4.2: 差分译码和差分解码

绝对码 $\{a_n\}$ 的 2DPSK 信号, 就是相对码 $\{b_n\}$ 的 2PSK 信号。即经过差分译码后, 可通过产生 PSK 信号的方式产生 DPSK 信号。

2DPSK 信号的解调可以使用相干解调加差分译码的方式, 也可以用差分相干解调法

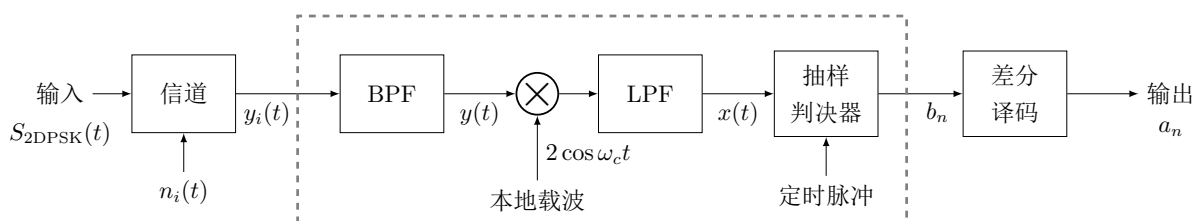


图 4.3: 2DPSK 相干解调系统

相干解调部分和 PSK 解调相同, 若只考虑一个码元内的波形:

4.2 移动通讯

4.2.1 5G

4.2.2 6G

覆盖率 (到 90%)、传输速度 (Tbps)、低延时 (1ms)、定位精度、可靠性、低功耗、智能水平。

第五章 电子电路

5.1 电源

5.1.1 同步 Buck

开关频率为 f_s ，则周期 $T_s = \frac{1}{f_s}$ ，MTOP（同步 Buck，上管称开关管，记 MTOP，下管称续流管，记 MBOT）导通时长为

$$T_{\text{on}} = D \cdot \frac{1}{f_s} \quad (5.1)$$

D 为占空比。电感开始储能，流过的电流增大，电感上电压左正（ V_{in} ）右负（ V_{out} ），伏秒积：

$$(V_{\text{in}} - V_{\text{out}})DT_s \quad (5.2)$$

（伏秒积反映了电感中电流的增量。这是因为电感特性：电感电压等于感量乘以电流的时间变化率。）其本质为流过的电流变化：

$$i_L(t) = i_L(0) + \frac{V_{\text{in}} - V_{\text{out}}}{L}t \quad (5.3)$$

因此电流增量：

$$\Delta I_L = i_L(T_{\text{on}}) - i_L(0) = \frac{(V_{\text{in}} - V_{\text{out}})D}{Lf_s} \quad (5.4)$$

MTOP 关断，MBOT 导通时长为：

$$T_{\text{off}} = (1 - D) \cdot \frac{1}{f_s} \quad (5.5)$$

电感向外输出电能，流过的电流逐渐减小，电感上电压左负（0）右正（ V_{out} ），伏秒积：

$$V_{\text{out}}(1 - D)T_s \quad (5.6)$$

在稳态下，电感上的电流纹波 ΔI ，应该等于 MTOP 开通期间输出电流的增加量，也等于其关断期间输出电流的减小量。所以开通和关断期间电感的伏秒积相等，

$$(V_{\text{in}} - V_{\text{out}})DT_s = V_{\text{out}}(1 - D)T_s \quad (5.7)$$

可得：

$$D = \frac{V_{\text{out}}}{V_{\text{in}}} \quad (5.8)$$

因此输出电流纹波：

$$\Delta I_L = \frac{(V_{\text{in}} - V_{\text{out}})D}{Lf_s} = \frac{V_{\text{out}} \left(1 - \frac{V_{\text{out}}}{V_{\text{in}}}\right)}{Lf_s} \quad (5.9)$$

由于输出电压的需求是确知的，开关频率由 IC 确定（为避开汽车 FM 频段，一般固定在 500 kHz 数量级或者大于 2MHz），我们可以利用的结论通常总结为：输入电压越高，纹波越大；电感越大，纹波越小。

有了电流纹波,就可以据此计算电感参数。一般原则是电流纹波为输出电流 I_{out} 的 20% ~ 40%。取小主要受到空间成本、动态响应等方面的约束,取大主要受到电容耐受、负载端 PSRR 指标的约束。

个人观点:在负载为较高精度的模拟电路下,尽量选大电感加大耐压的电容,并做好负载 IC 的瞬态抑制和浪涌保护。

电感值选取:

$$L = \frac{V_{out} \left(1 - \frac{V_{out}}{V_{in,max}}\right)}{\Delta I f_s} \quad (5.10)$$

取输入电压最大值就是为了考虑最坏情况——如果 ΔI 最大,电路能满足要求,那么其他情况纹波只会更小,更能满足要求。

电感选取还要注意饱和电流的问题:铁氧体电感存在硬饱和、磁漏、体积大的问题,当电流超过一定值后电感感量急剧下降;一体成型的粉芯电感软饱和,但可能存在感值小、散热慢等问题。通常取

$$I_{L(max)} = I_{out} + 0.5\Delta I \quad (5.11)$$

作为对电感饱和电流参数的标准。必须选择饱和电流比该值大的电感。

(饱和电流常定义为使 L 下降额定值的 30% 处的电流)

通过电流纹波计算电压纹波:

$$V_{ripple} = \Delta I \left(\frac{1}{8f_s C_{out}} + (ESR)_{C_{out}} \right) \quad (5.12)$$

输出电压纹波有容性分量和阻性分量,分别由输出滤波电容的容值(不足)和其 ESR(过大)引起。

5.1.2 PFC

功率因数校正 (Power Factor Correction, PFC) 电路,用于防止用电器产生的杂波影响到电网。

LG=logic ground 逻辑地,也是数字地

PE=power earth 是电源地

FG 是保护地,外壳接地