生物医学信号处理与分析

- 现根据课件过一遍，重点结合好课后题，记好笔记。
- 再完成课后习题。
- 根据cohen的书实践练习。

绪论

什么是生物医学信号

> 生物医学信号是反映**生命状态**（健康、疾病）和**功能**及其**变化**的数据，范围十分广泛。

生物医学信号的特点

- 信号特征
  - 微弱信号
  - 强干扰噪声
  - 频率低
  - 干扰与信号的混叠
  - 强随机性（时变、复杂、个体差异）
  - 非平稳特性
  - 非线性特性
  - 信息量大
- 影响信号特征的因素
  - 生理系统的复杂性
  - 测量环境的复杂性
  - 测量环境的性能限制

信号特性	处理方法
微弱信号	低噪声高增益放大器
高噪声	软硬件方法消除噪声、抑制干扰
干扰与信号的混叠	根据干扰和信号的特征设计待定的测量与信号处理方法。如：陷波器消除工频干扰；非等间隔诱发，多次叠加抵消诱发脑电中的自发脑电。
频率低	时间稳定性高的放大器
强随机性、非平稳特性	一般不能用确定数学函数表达，选哟从大量数据中提取拥挤结果。——**采用统计信号处理理论与方法**。
非线性特性	非线性建模，分形（自相似）、混沌（无限循环但不完全重复）
信息量大	数字化、计算机自动算法、机器学习

生物医学信号分类

1. 按产生方式

信号类型	定义	举例
内源性信号（诊断、监护和治疗的依据）	有生理过程自发产生的**主动信号**	心电、脑电、眼电、肌电、呼吸、脉搏...
外源性信号（临床应用）	外界信号施加于人体，人体作为通道通过外界信号后产生的的信号特征，是**被动信号**。	入射超声波-人体-出射超声波、PRT、X-ray、MRI...
感生信号（重要的诊断手段、科研工具）	指检测到的信号由外部刺激所诱发的生理变化，从而产生的内源性信号、诱发信号。	施感信号（是刺激-光学信号）、视觉诱发脑电-电信号

2. 按数学特性分类

信号类型	特性
确定性信号	
非确定性信号	
分形信号	无限复杂但一定意义下有自相似性（心率信号）
混沌信号	不能准确预测其未来的确定性信号。

3. 按物理性质分类

信号类型	特性
生物电信号	从离子通道到体表电位
非生物电信号	心音信号、CT信号、超声信号、MRI信号
有限能量信号	
有限功率信号	

常见信号类型幅度和频率

信号类型	幅度范围	频率范围	传感器类型
心电（ECG）	0.01~5mV	0.05~100Hz	表面电极
脑电（EEG）	2~200μV	0.1~100Hz	帽状/表面/针状电极
肌电（EMG）	0.02~5mV	5~2000Hz	表面电极
眼电（EOG）	0.01~4mV	0.1~100Hz	表面电极
胃电	0.01~1mV	DC~1Hz	表面电极

生物医学信号的价值

- 临床诊断价值
  - 原理：生理信号 - 生理状态 – 心理状态 – 行为，这些变量之间存在特定的关系
  - 方法：通过观测生理信号对人的生理、心理的健康状况进行评估，对人的某些特定行为进行解释。
- 科研价值

信号的时空层次

行为——网络连接——皮层区——细胞群——细胞——细胞膜——基因

生物医学信号有关定义

- 生理信号：从人体采集到反映人体生理功能和状态的各种信号。
- 医学信号：用于临床检验诊断的人体信号。

- 生物医学信号：生命体携带货发出的信号

实验数据、建模与仿真

科学问题的提出

什么是科学的问题?

> 科学问题是可以通过提出**假设**并进行**实验**、系统观察和分析来回答的问题。
>
> A scientific question is a questio that can be answered by forming a hyopthesis and conducting an experiment/systematic observation and analysis.

> 科学问题应该旨在寻求能够验证、解释或理论化信息的答案。
>
> A scientific question should aim for an answer that verifies, explains or theorizes information.

科学问题来自哪里?

- 观察+思考
  - 好奇心
  - 需求
- 科学文献查阅
  - 研究的空白、已有研究的不足
- 上述两者结合

科学数据

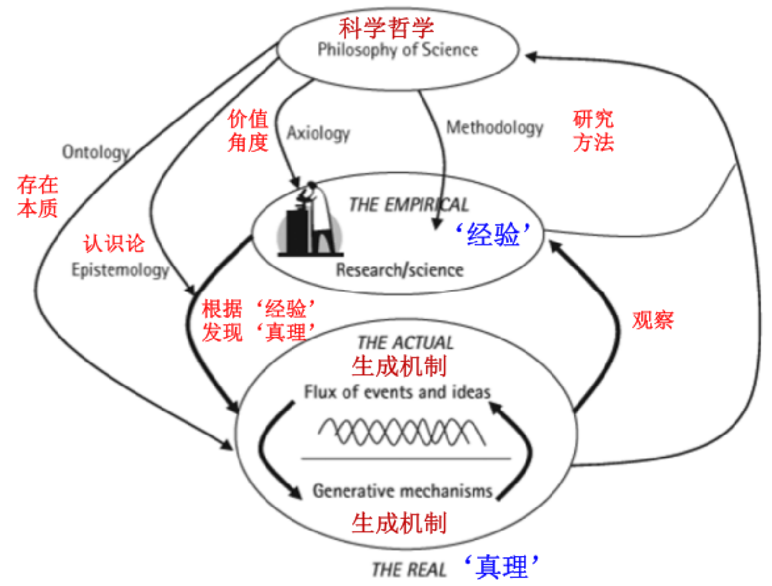
什么是科学的数据?

> 科学数据通常是由有意识的行为产生的。通常情况下，科学家会自觉地测量、观察或计算科学数据，以解答科学问题。

> Scientific data is usually created by a conscious act. Normally the scientist consciously measures , observes or computes scientific data to answer scientific questions

注：与科学问题无关的数据也是数据，但对于科学问题来说这是噪声。

数据的本质来源?



John Mingers (2011) The Truth, the Whole Truth, and Nothing but the Truth: High-Quality Research in Information Systems.

科学数据哪里来?

- 实验设计
  - 有效地获取特定条件下够量的特定数据
- 数据采集
  - 用合适的设备和方法获取所需的数据。

为什么要做实验设计?

1. 目标角度
   - 合理的实验设计是有效回答科学问题的基础
     - 提升研究结论的可靠性
     - 经济性
   - 生物医学数据的性质要求良好的实验设计，充分利用统计分析的方法。

2. 需求角度
    - 生物医学数据：广泛变异性
        - 数据受个体差异影响而有随机性，个体集合的取值有规律性
        - 个体状态的变化，非目标变量对目标变量的影响
    - 仪器：不可避免的测量误差。
    - 数据量：有限的观察
        - 受条件限制，仅能测量研究对象部分数据。

> 一个精心设计的试验是认识世界的有效方法 。

统计学对数据分析的重要性

- 描述性统计学
- 推断性统计学

什么是随机误差？抽样的原则是什么？为什么需要建立这样的抽样原则?

什么是随机误差?

- 不可控的诸微小因素之总和，称为随机误差（random error）。
- 同样条件下的两次实验结果可能不同。
- 随机误差存在于一切试验之中。

抽样的原则是什么?

- 代表性
- 可实施性
- 代表性和可实施性两者的平衡

为什么需要建立这样的抽样?

> 抽样——从总体中获得一个样本，通过样本推断总体的特征。

实验设计的Fisher原则是什么？为什么需要遵守这三项原则？

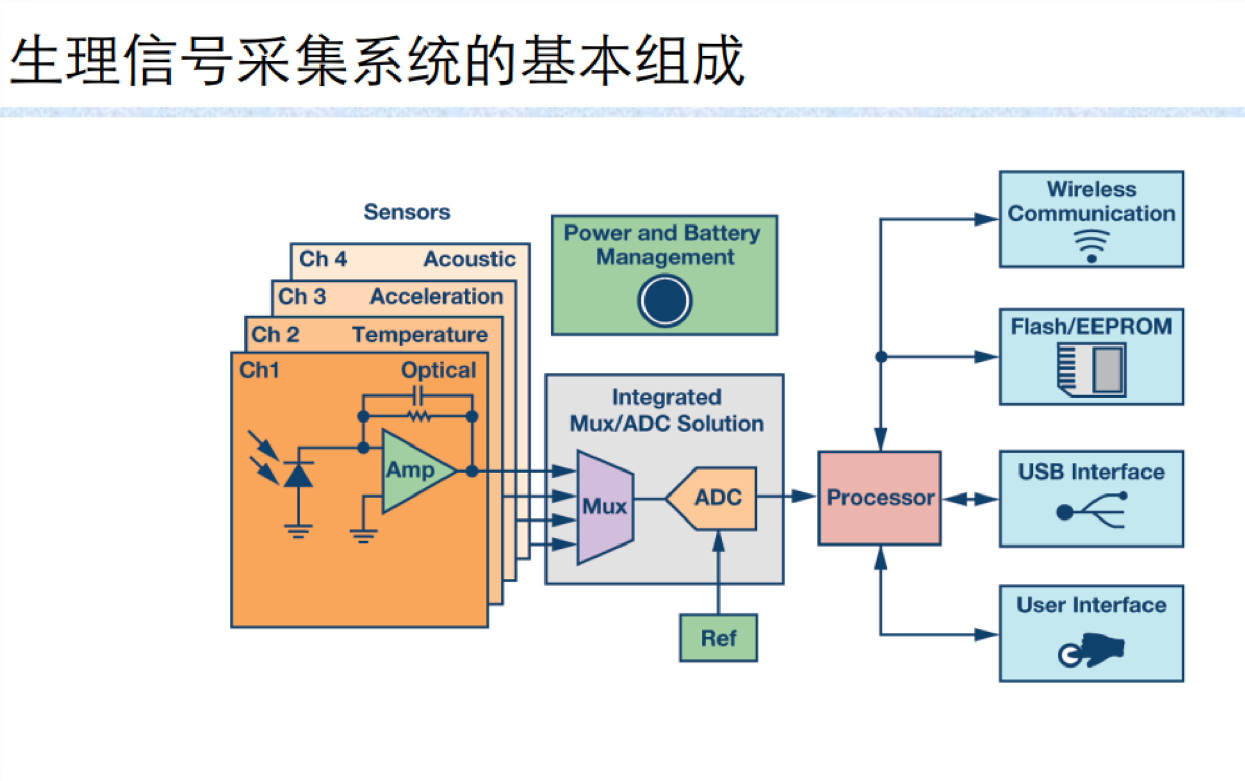
- 重复
- 随机化
- 局部控制

1. 重复的意义：
    - 只有设置重复才能获得实验误差的估计
    - 有效减少实验误差的影响
    - 推断出处理效应
2. 随机化的意义：
    - 样本的配置与实验处理的顺序随机确定。
    - 获得无偏的误差估计，避免主观和偏敬 。
3. 局部控制的意义：
    - 局部控制是指试验环境或单位差异大时，可将整个试验环境或试验单位分成若干个小环境或小组，在小环境或小组内使非处理因素尽量一致。
    - 区组效应可以在方差分析中分离出来，从而有效提高检验精度。

统计试验设计方法

- 简单实验设计
    - 成组比较实验设计
    - 配对比较实验设计
- 单因素实验设计
    - 完全随机化设计
    - 随机化完全区组设计
- 多因素实验设计

一个生物医学信号采集系统包含哪些基本模块？每个模块的作用是什么？为了起到这些作用，需要满足哪些特性？



模块	主要作用	关键特性 / 性能要求
1. 传感器（Sensors）	将生理信号（如光信号、温度、加速度、声学信号等）转化为可测量的电信号。	- 高灵敏度：能检测微弱的生理变化（如心率、脑电、肌电）- 高信噪比：抗干扰、低噪声- 合适的频带响应：匹配信号频率范围- 安全性与生物相容性（对人体无害）
2. 信号调理电路（Amplifier，前置放大与滤波）	对传感器输出的微弱信号进行放大、滤波、阻抗匹配。	- 高输入阻抗（防止信号衰减）- 低噪声、高共模抑制比（CMRR）- 合适的增益与带宽- 有效滤除工频干扰与噪声（如 50/60 Hz）
3. 多路选择与模数转换模块（Mux + ADC）	将多通道模拟信号按顺序输入 ADC，转换为数字信号供处理器使用。	- 多通道切换速度快（足够高的采样率）- ADC 分辨率高（≥12 bit，常用16 bit或24 bit）- 采样频率符合奈奎斯特准则（信号最高频率的两倍以上）- 转换精度高、量化误差小

模块	主要作用	关键特性 / 性能要求
4. 参考电压模块（Ref）	为 ADC 提供稳定的参考电压，保证转换准确性。	- 稳定性高、温漂小- 噪声低
5. 处理器（Processor / MCU / DSP）	负责数据接收、分析、存储与通信控制，是系统的“核心大脑”。	- 处理速度快、功耗低- 支持多任务（数据采集、AI算法、通信）- 拥有足够的存储空间与接口（UART、I^2C、SPI等）
6. 存储模块（Flash / EEPROM）	存放采集到的数据或系统参数（如用户设置、设备配置）。	- 存储容量适当- 支持断电保存（非易失性）- 读写速度快
7. 通信模块（Wireless / USB）	将数据传输至外部设备（手机、电脑、云端）。	- 支持多种通信方式（蓝牙、Wi-Fi、USB等）- 传输速率稳定- 低功耗通信协议（特别是可穿戴设备）
8. 用户接口（User Interface）	提供用户与系统交互（显示状态、按键控制、指示灯）。	- 友好的人机交互界面- 实时反馈- 易于操作
9. 电源与电池管理模块（Power & Battery Management）	为系统各部分提供稳定电源，管理充放电。	- 稳压、低噪声供电- 电池容量合适- 支持低功耗管理模式（Sleep / Standby）
10. 校准与参考系统（Ref + Self-Check）	保证系统测量精度与一致性。	- 自校准机制- 温度补偿- 定期参考信号检测

用生物医学仪器实现信号采集的时候，主要注意哪些影响信号质量的因素？一般采用哪些方案进行解决，以获取高质量的生物医学信号?

- 仪器的噪声与稳定性
  - 微弱信号——低噪声仪器
  - 慢变信号，如皮肤电导率对稳定性要求高
- 环境干扰
  - 声音干扰——声音屏蔽
  - 电磁干扰——电磁屏蔽和合适而可靠的接地
  - 振动干扰——减振系统

- 被试观测对象
  - 保持状态稳定性，睡眠状态与清醒状态的生理信号不同
  - 非目标信号对目标信号的干扰。

> 同时要注意：电气安全和信息安全。

什么是建模和仿真？建模与仿真有哪些作用或优点?

什么是建模和仿真?

- 建模：对实际系统特性或行为的抽象、近似表达，可用于理论验证，帮助提出科学假设。
- 仿真：对模型状态和行为的数值呈现，产生某种特性的数据集，测试算法。

> 研究对象：系统内部个元素之间及系统内外作用的过程和结果。

建模与仿真的作用

- 产生和验证科学假设，抽象出特定的生命机制
- 检验针对某种信号特征的信号处理算法和性能
- 用来获取一般情况下很难获取的数据
- 代替动物研究，实现生物医学研究伦理“好科学”的“3Rs”的代替原则

好科学的3Rs原则是什么？为什么要遵从这三个原则

- 改进refinement
- 减少reduction
- 代替replacement

遵从这三个原则是为了**实现生物医学研究伦理**

模型的精确性和准确性分别指代什么？

- 精确性——结果的可重复性
- 准确性——模型仿真结果与真实之间的差距。

白噪声

> 白噪声white noise：所有频率上具有相等功率的不相关随机过程。

高斯白噪声

- 定义：功率谱函数是常数（即白噪声），且信号的时域值服从高斯分布的噪声信号。

白噪声与有色噪声频域特点

- 白噪声：功率谱是**平坦的**（常数）。即所有频率上功率相等，**与频率无关**。
- 有色噪声：功率谱密度在感兴趣的带宽内呈现不均匀分布的噪声，噪声功率通常随着频率增高而递减 。

带限白噪声

- 功率密度谱图表明，**理想的白噪声**需要无限大的功率来覆盖无穷带宽的频率范围，物理上不可以实现。
- 信号采集系统具有有限的带宽，且研究者仅对某些带宽内的信号感兴趣。
- 实际的噪声信号只在一段频段内可以用高斯噪声特性来近似处理。

> 带限白噪声：带宽为whz的带限白噪声指带限范围内的功率谱恒定，幅值符合高斯分布的噪声。

其他随机噪声

- 有色噪声
- 电磁噪声
- 热噪声

- 散粒噪声

良好的编程风格

> 良好的编程风格:
>
> 1.可以提高代码的可读性，一致的编程风格
>
> 2.减少bug

```
%变量和命名规范
%清理工作空间： 使用 clear all 开始一个脚本程序
%所有变量在脚本开头定义。
%充分注释
```

确定性信号

> 定义：可以准确地用一个数学上的时间函数描述的信号，可以准确地重现。

随机信号基础

概率密度函数与概率密度分布函数

概率密度分布函数：$F(a) = P(x \le a)$

概率密度函数：$f(a) = dF(a)/da$

离散变量, 用**概率质量函数 (PMF)**, P(X=x)。它直接给出了 X 等于某个特定值 x 的概率。

连续变量, 用**概率密度函数 (PDF)**, f(x)。它的值本身不是概率，但它描述了变量在 x 点附近的“概率密度”。f(x) 的值越高，意味着变量落在这个点周围一个小区间的概率就越大。

高斯概率密度函数

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

> 优点：均值和方差完全确定概率密度函数

生物医学信号是大量独立随机事件相互作用的结果，如果每个因素对信号的贡献都很小，总的贡献构成测量到的信号，这个信号的幅值特性可以用正态分布描述。

随机变量的矩

矩(moment)：

- 可以用来刻画随机变量的性质，随机变量的一种数字特征。
- 随机变量x的函数g(x)的广义矩定义为其数学期望算子

$$E[g(x)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

- $g(x)$ 常常是多项式形式，例如x, x^2, $(x\text{-}m)^2$
- $f(x)$ 是概率密度函数

常见的矩

- 均值（一阶矩）
- 均方值（二阶矩）
- 中心矩
- 方差（二阶中心矩）
- 偏度（三阶中心矩）

概率论&统计学

概率论	统计学
随机变量	随机样本（→推断总体，对应随机变量）
分布函数（分布律）、概率密度函数	抽样分布—统计量的概率分布
数字特征	基于统计量的推断（参数估计和假设检验）
前提：**已知**随机变量的分布，在此前提下研究**随机变量**的性质、特点和规律	条件：假设随机变量分布，但是**具体参数未知**，需要根据**随机样本**对其（即总体性质）进行**推测**

- 描述性统计学
  - 解决如何将原始数据整理成有用形式。
  - 包括收集、整理、概括、描述数据信息。
- 推断性统计学
  - 从有限、不确定的样本信息对总体做出判断及决策。
  - 前提：有限样本能很好的反应总体特征。

统计推断

参数估计

假设已知总体X的分布函数的形式，例如高斯分布N(μ, δ^2)，但是函数中的参数未知。参数估计的任务即是找到根据观察到的随机样本计算这些未知参数。

点估计：计算未知参数的值。

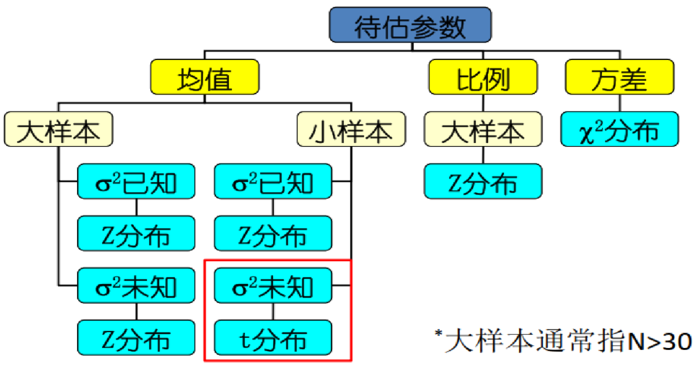
区间估计：计算未知参数所在区间。

– 均值 $E[x] \approx \hat{m} = \frac{1}{N}\sum_{j=1}^{N} x_j$

– 方差 $\hat{\sigma}^2 = \frac{1}{N}\sum_{j=1}^{N}(x_j - m)^2$ $\hat{\sigma}^2 = \frac{1}{N-1}\sum_{j=1}^{N}(x_j - \hat{m})^2$

– 偏度 $\hat{\mu}_3 = \frac{1}{N}\sum_{j=1}^{N}(x_j - \hat{m})^3$

常见的区间估计条件和抽样分布

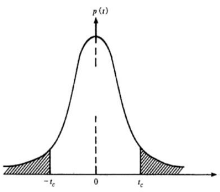


方差的置信区间估计
标准差的区间估计
(在PPT上)

均值估计的置信区间

- 小样本方差未知情况下，均值估计量符合$t(n\text{-}1)$分布
- 均值估计的置信区间 $[m_l, m_u]$
  - 满足 的t_c是临界值
    - 1-α 为置信 $P[|t| \geq t_c] = 0.05 = \alpha$
  - 非阴影区域为置信区
  - 对于 α 的置信下界和置信上界

$$\begin{cases} m_l = \hat{m} - t_c \frac{\hat{\sigma}}{\sqrt{N}} \\ m_u = \hat{m} + t_c \frac{\hat{\sigma}}{\sqrt{N}} \end{cases}$$



区间估计要点

· 统计量的分布的对应关系
– 与参数（均值、比例、方差）有关
– 与样本大小（自由度）有关
– 与方差已知与否有关
• 查分布表对应的临界值
– 与临界水平/置信水平有关
– 与样本大小有关
• 代入区间估计公式计算置信区间的边界

随机过程

- 数学表达：设T 是一个实数集，依赖于参数$t \in T$的一簇随机变量X(t)称为随机过程，记为 $X(t)$, $t \in T$
  - 对于任意时刻t=ti， $X(ti)$是一个随机变量，可用概率论研究
  - X(t)是随机过程在t时刻的状态

- 对于一切$t \in T$， X(t)所有可能取值的全体是随机过程的状态空间

随机过程指代随机时间演变的随机现象本身

	概率论	随机过程
研究对象	随机变量	随时间变化的随机现象，引入了时间维度的随机变量
举例	心率	心率随时间的变化
研究内容	概率分布【与时间无关】	概率分布【是时间的函数】
	数字特征（矩）【如均值、方差、偏度等，与时间无关】	数字特征（矩函数）【是时间（或其它变量）的函数】

随机过程的样本

- n个随机样本在随机过程中称为随机函数
- 样本函数仅仅是这个随机过程的具体实例
- 随机过程包含这样样本在内的全体“曲线”样本，X(t, ξ)
- 选取特定时刻，例如t_0对应一个随机变量X(t_0)

随机信号的概率分布函数

一维随机信号的概率分布函数及概率密度分布函数

$$F_X(x_n, t_n) = P(X(t_n) < x_n)$$

$$f_X(x_n, t_n) = df_x/dx_n$$

x1是随机过程在时刻t1的取值（即随机变量X(t1)）的一个具体数值

> n维分布函数簇
>
> $$F_X(x_1, x_2, ..., x_n; t_1, t_2, ..., t_n) = P(X(t_1) < x_1, X(t_2) < x_2, ..., X(t_n) < x_n$$

随机信号在不同时刻的取值组成的高维随机变量。

当n足够大，n维分布函数簇可以接近完全刻画随机过程的特性

随机过程的平稳性

> 随机信号的概率密度分布是时间的函数。

一阶平稳：$f_{x(t_1)}(a)=f_{x(t_2)}(a)$

二阶平稳：$f_{x(t_1+T)x(t_2+T)}(a_1,a_2)=f_{x(t_1)x(t_2)}(a_1,a_2)$

强平稳（严平稳、狭义平稳）：$f_{x(t_1+T)\ldots x(t_n+T)}(a_1,\ldots,a_n)=f_{x(t_1)\ldots x(t_n)}(a_1,\ldots,a_n)$

> 很多随机信号分析方法的基本要求是随机信号具有弱平稳性 .
>
> 平稳性判定的最简单方法 – **考察产生信号的物理机制和环境** 。
>
> 统计检验方法 。

随机信号的数字特征

课程仅以高斯信号为例理解医学信号处理，则我们仅需考虑一阶矩和二阶矩

- 随机过程$\{X(t), t\in T\}$
- **随机过程的均值函数**:$\mu_X(t)=E[X(t)]$ 均值函数是所有样本函数的平均值，是时间的函数；是随机过程X(t)在各个时刻的摆动中心
- **随机过程的二阶中心矩**：$\sigma_X^2(t)=E[[X(t)-\mu_X(t)]^2]$ 是随机变量围绕 “ 摆动中心 ” 变化范围的一个度量

随机信号的统计量

- 随机信号的数字特征
- 自相关函数 设 x(t1)和x(t2)是随机过程X(t)在t1和t2时刻的随机变量， x(t1)和x(t2)的乘积的统计均值为自相关函数
- 自协方差函数 设 x(t1)和x(t2)是随机过程X(t)在t1和t2时刻的随机变量，称x(t1)和x(t2)的二阶联合中心矩为X(t)的自协方差函数

随机信号广义平稳的定义及其判别

广义平稳（宽平稳）：均值和方差不随时间变化。

$$E[x(t_1)]=E[x(t_2)]$$

$$Cov[x(t_1),x(t_2)]=Cov[x(t_1),x(t_1+T)]$$

平稳性的统计检验方法:

- **步骤**
  - **将记录信号分成多个片段**
    - 片段要足够长，对于包含振荡分量的随机信号，每个片段都必须包含多个振荡周期
    - 例如心电包含振荡分量，过短则不能反映信号的真正特性
  - **估计每个片段的均值和方差**
  - **用标准统计检验方法检验各个片段的均值和方差异同**
    - 分两段的均值检验 (T统计量) 和方差检验 (F统计量)
    - 分多段比较的方差分析法 (ANOVA，此课程不要求)
    - 对于非高斯信号，则需要建立各个信号分段的概率密度函数或者累计分布函数模型，比较这些模型的稳定性 。

随机过程各态遍历性的意义，与数据采集方法的关系

定义：如果所有样本函数在某一特定时刻的统计特性与单一样本函数在长时间内的统计特性一致，则称该随机过程各态遍历。

意义：

- 获得大量样本函数很难，常常只能观察一个样本函数
- 测量生物医学信号，都假设其具有各态遍历性，通过多次测量估计出随机信号的特征和参数。

随机信号采集

- 理想情况：完全观测数据
- 实际情况：
  - 放大器数量N有限。
  - 大多数实验数据采集中N=1。
  - 假设随机遍历性
  - 分时间先后顺序，多次重复测量M次

随机信号处理的三项数学基础：

	概率论	随机过程[1]	统计学	随机信号
研究对象	随机变量：无时间维度	随机函数：随某时间[2]变化的随机现象	随机样本（以及样本矩）	随机信号的样本（以及随机信号的样本矩）
举例	心率作为随机变量	视觉诱发电位	样本：对50人的一次心率检测，或者一段时间的即时心率数据	测量50次的视觉诱发电位
研究内容	概率分布【与时间无关】	概率分布【是时间的函数】	抽样分布—统计量的概率分布	随机信号统计量的概率分布(抽样分布)
	数字特征（矩）【特定数值，如均值】	数字特征（矩函数）【是时间的函数】	基于统计量的推断（参数估计-点估计和区间估计，假设检验）	基于统计量的推断（参数估计-点估计和区间估计，假设检验）

时域分析与处理

互相关函数CCF

$R_{yx}(k) = E[y(n)x(x+k)]$

k表示信号x相对于信号y的延时点数。对于采样间隔T，延迟时间为$T_d = kT$

互协方差函数CCVF

样本协方差 $\hat{\sigma}_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \hat{m}_x)(y_i - \hat{m}_y)$

$C_{yx} = E[(y(n) - m_y)(x(n+k) - m_x)] = R_{yx}(k) - m_y m_x$

归一化互协方差函数NCCF

$\rho_{y,x}(k) = \frac{C_{yx}(k)}{\sigma_y \sigma_x}$

归一化协方差 $\hat{\rho} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y}$

性质

- 互相关函数具有对称性，关于x=0正负半轴对称
- 归一化的互协方差函数的取值范围 $-1 =< \rho_{y,x}(k) =< 1$
- $R_{y,x}(k)$最大值处，信号最相似

互相关函数估计

- 互协方差函数

$$\hat{C}_{yx}(k) = \frac{1}{N}\sum_{n=0}^{N-k-1}(y(n) - \hat{m}_y)(x(n+k) - \hat{m}_x) \quad \text{for } k \geq 0$$

$$\hat{C}_{yx}(k) = \frac{1}{N}\sum_{n=0}^{N-k-1}(y(n+k) - \hat{m}_y)(x(n) - \hat{m}_x) \quad \text{for } k \leq 0$$

  - 信号分析时，数据常常去趋势和直流成分，互协方差函数本身就是相关函数，也记做 $\hat{R}_{yx}(k) = \hat{C}_{yx}(k)$ $c_{yx}(k)$
  - 用于时延分析，但是无法有效评估归一化的互相关强度
- 归一化的互协方差函数的估计，估计值范围[-1 1]

$$\hat{\rho}_{yx}(k) = r_{yx}(k) = \frac{c_{yx}(k)}{s_y s_x}$$

互相关函数应用

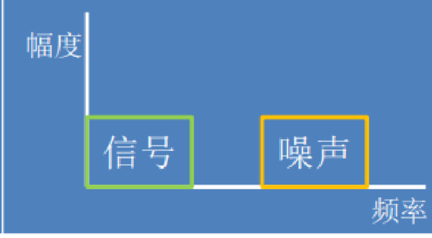
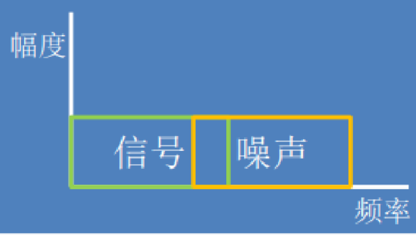
- 优先市场信号时延、距离、速度
- 信号极性变化检测
- 微弱信号处理（模板匹配法去除确定性强干扰）

有限时间信号的传播时延

理想情况

- 模板信号x(t)，时间有限，记为w
- 待测信号y(t)
- 互相关函数 $$R_{xy}(\tau)=\frac{1}{W}\int_{0}^{W}x(t)y(t+\tau)dt$$
- y(t)是x(t)的一个时延信号y(t)=x(t-T_d)
- 互 相 关 函 数 等 价 为 自 相 关 函 数 $$=\frac{1}{W}\int_{0}^{W}x(t)x(t-\tau_d+\tau)dt=R_{xx}(\tau-\tau_d)$$
- $T=T_d$达到最大自相关系数$R_{xx}(0)$

叠加平均原理与应用

目标信号	噪声与信号 频谱不重叠 	噪声与信号 频谱重叠 
确定信号	滤波	叠加平均
随机信号	滤波	例如独立成分分析 （研究生课程或自学）

叠加去噪的原理

- 目标信号为确定信号S(t)
- 测量的信号D(t)包含目标信号和高斯噪声n(t):D(t）=S(t)+n(t)
- 可以重复测量D_i（$i=1,2,...,n$）
- 每次测量的确定信号一致，即$S_i(t)$都相等。
- 每次测量包含的噪声信号满足高斯分布（独立同分布）

叠加平均算法的假设前提

- 数据模型D(t) = S(t) +nh(t)
  - 信号和噪声是线性叠加关系
  - 满足遍历性，可以通过在同等条件下多次测量，获得多个样本函数Di(t)
- 待提取信号S(t)为确定信号，即
  - S1(t)=S2(t)=…=S_K(t)= S(t)
  - 相对于每次测量的起始点，时间锁定，相位锁定
- 随机噪声h(t)为高斯噪声，符合独立同分布，即
  - n(t) ~ N(0, s^2)

叠加去噪的实际情况及对策

条件	策略
锁时，锁相	波形叠加平均，功率叠加平均
锁时，不锁相	不可波形叠加平均，功率叠加平均
不锁时，不锁相	不可波形叠加平均，可功率叠加平均，但模糊
不锁时，锁相	不存在的情况

频域分析与处理——傅里叶变换与功率谱估计

振荡

> 振荡是基本的生命现象。

周期性活动是物理现象、生命现象、工程系统中的基本信息。

> 傅里叶分析是最基本的谐波分析方法，是揭示振荡规律最重要的工具。
>
> 应用：提取信号的频率特征。
>
> 在频域内进行信号处理更加直观、高效。

确定性信号的傅里叶分析

傅里叶变换存在的条件

- 函数f(t)在正无穷到负无穷或者周期[0,T]上满足狄利克雷条件
  - 连续或只有有限个第一类间断点。
  - 极大值和极小值的数目应是有限个。
  - 信号是绝对可积的。
- 一般的周期信号都满足狄利克雷条件
- 狄利克雷条件是一个信号存在傅里叶变换的充分不必要条件。

傅里叶变换与逆变换

傅里叶变换：

$F(\omega)=\int_{-\infty}^{\infty} f(t)e^{-j\omega t}\,dt$

傅里叶逆变换：

$f(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{j\omega t}\,d\omega$

离散傅里叶变换

周期延拓，把一个有限时间定义的信号，**重复无穷次**，从而得到一个周期信号。变换成**周期信号**，经过傅里叶变换，获得频域内离散的线谱。

对频谱进行周期延拓，傅里叶逆变换，获得离散时域信号。

信号类型和傅里叶分析形势

	傅里叶级数	傅里叶变换
时域	$\tilde{x}(t)=\sum_{k=-\infty}^{+\infty} a_k e^{jw_0kt}$ 时域是连续周期的	$x(t)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}X(jw)e^{jwt}dt$ 时域是连续非周期的
频域	$a_k=\frac{1}{T_0}\int_{T_0}\tilde{x}(t)e^{-jw_0kt}dt$ 频域是离散非周期的	$X(jw)=\int_{-\infty}^{+\infty}x(t)e^{-jwt}dt$ 频域是连续非周期的
时域	$\tilde{x}(n)=\sum_{k=\langle N\rangle}\tilde{a}_k e^{jk(2\pi/N)n}$ 时域是离散周期的	$x(n)=\frac{1}{2\pi}\int_{2\pi}X(e^{jw})e^{jwn}dw$ 时域是离散非周期的
频域	$\tilde{a}_k=\frac{1}{N}\sum_{n=\langle N\rangle}\tilde{x}(n)e^{-jk(2\pi/N)n}$ 频域是离散周期的	$X(e^{jw})=\sum_{n=-\infty}^{+\infty}x(n)e^{-jwn}$ 频域是连续周期的

采样率、采样时间、频率分辨率

- **采样率**Fs，**采样时间间隔**$t_{resol}=1/Fs(sec)$
- 信号处理的时间窗口长度T(sec)——又称信号分析的时间窗口宽度、时间分辨率
- 时域数据长度$N_t=T*Fs$
- 频域数据长度$N_f=N_t$
- **频域分辨率**$f_{resol}=Fs/N_f=Fs/N_t=Fs/(T*Fs)=1/T$

为什么不同域内观察频率成分的难易度不同?

> 在时域中，周期信号被随机噪声掩盖而不明显；而在频域中，傅里叶变换把信号能量分解到各个频率上，使得周期成分的能量在特定频率上集中形成谱峰，从而揭示出隐藏的周期特征。

随机信号的功率谱估计

随机信号的谱估计概念

- 确定信号 x(t)唯一确定，谱也唯一确定
- 对于随机信号，我们仅能获得样本函数 x(t)，样本函数是一个随机函数，不是唯一确定，因此X(f)也是一个随机函数。
  - 样本函数x(t)的傅里叶变换X(f)仅是随机信号谱的一个样本函数
  - 需要获得多个样本函数才能估计随机信号的谱
- 时域内，通过一系列的样本函数x(t)，可以估计随机信号的矩函数
- 频域内，通过获得一系列样本函数的谱X(f)，可以估计随机信号的谱 —— 谱估计

随机信号谱分析间接法

> - 随机信号
>   - 无限市场，非周期性，非平方可积。
>   - 不能离散傅里叶变换，因为周期无限长，对应连续谱。
> - 样本函数
>   - 对一段样本进行傅里叶分析。
>   - 单一样本不能代表总集。
> - 间接法
>   - 找到用样本描述的总集的统计量。
>   - 如果该样本满足傅里叶变换的条件，对该统计量进行傅里叶分析。

维纳-辛钦定理：

对于一个**平稳随机过程** x(t)x(t)x(t),
其**自相关函数**定义为：

$$R_x(\tau) = E[x(t)\, x*(t+\tau)]$$

则其**功率谱密度（Power Spectral Density, PSD）**定义为：

$S_x(f) == \int_{-\infty}^{+\infty} R_x(\tau)e^{-j2\pi f\tau}\, d\tau$

即：

> **功率谱密度是自相关函数的傅里叶变换。**

反之：

$Rx(\tau) = \int_{-\infty}^{+\infty} S_x(f)e^{j2\pi f\tau}\, df$

> **自相关函数是功率谱密度的反傅里叶变换。**

间接法步骤

- 先估计随机样本函数的自相关函数。
- 再根据维纳辛钦定理，计算自相关函数的傅里叶变换获得功率谱。

随机信号谱分析直接法

> 直接法：对样本函数直接进行傅里叶变换，对获得的各频率系数平方求平均以获得功率谱。
>
> 又称——周期图法。

- 根据自相关函数的傅里叶变换获得的功率谱

$$P_{xx}(e^{j\omega}) = \sum_{m=-\infty}^{+\infty} R_{xx}(m)e^{j\omega n}$$

- 上式带入自相关函数$R_{xx}(m)$变形

> 存在问题：直接法功率谱估计不是一致估计，即功率谱估计的方差不会随样本量的增大而减小。

直接法的改进

> 改进方法:
>
> 平均周期图法，又称Bartlett法。

> 原理:
>
> 方差为σ^2的独立高斯随机变量的k个样本的均值的方差等于σ^2/k
>
> 随着数据量增加，样本个数k增加，方差减小

平均周期图算法

- 将长度为N点的样本函数分为K段，每段长度为M (M=N/K)
- 计算每一段M点数据的功率谱$P_i(m)$
- 对k段周期图取平均P(m)=1/K×$\Sigma_{i=1}^{K}P_i(m)$
  - 优点：改善了方差特性——σ^2/K
  - 缺点：每一段的数据长度M远远小于N——更短的窗口，更大的频谱泄漏

> 频谱泄露的改进方法:
>
> ——加窗

平均周期图法M与K值的选取

- 将N点数据分为K段，每段长度为 M = N/K
- K段间无重叠，如果每段独立，则估计均值不变，方差减少K倍，即σ^2/K
- K越大，方差越小，但是M 越小，估计偏差越大
- K和M的取值的相互妥协
  - M需要满足频率分辨率的需求
  - 在此基础上选择K
  - 对于有窄峰的谱， M值的选择非常重要

功率谱故居改进的welch法

> 应用更广泛

实质：**平均法**

分段： K = N/L

- 每段可以重叠
- 从而获得更大段数K
- 每段数据时域加窗
- 每段数据傅里叶变换
- 求所有分段功率谱的均值

使用matlab函数：`pwelch()`

短时傅里叶变换与连续小波变换

为什么要做时频分析

- 生物医学信号常为非平稳信号。（反映了生理系统的动态特性）
- 非平稳信号的均值、方差、自相关函数等统计特征在不同时间区间有显著差异。
- 功率谱分析非平稳随机信号的困难。
  - 随机信号的平稳遍历性时功率谱分析的基本假设。
  - 非平稳信号不满足功率谱估计的基本操作。
  - 功率谱分析将时间域变换到频域，弯曲丢失信号的时间信息没无法体现信号特性随时间的变化
  - 某些非平稳随机信号经过预处理后满足平稳性质
  - 并非所有场合都可以做以上处理。

时频分析的概念

> 能够有效体现随机信号的频域特性随时间变化规律的信号分析方法。

时频分析的种类

- 线性时频分析方法: 短时傅里叶 /Gabor 变换/小波变换
- 非线性时频分析方法: Wigner-Ville / Cohen类时频分析

  > 不同的时频分析方法具有各自特点和局限性,根据信号和分析需求，选择合适的时频分析方法。

近似平稳

> 慢变的非平稳信号在较短时间内，可以近似认为时平稳的，从而借助平稳信号的分析方法进行分析。

局部频谱

> 使用一个窄的时间窗口函数取出随机信号的一段，按照平稳随机信号的处理方法估计其频谱特性，由于剔除了窗口之外的信号，这种分析被称为局部频谱。

短时傅里叶变换

基本原理

对于慢变得非平稳信号x(t)，使用延时间轴滑动的窗函数w(t)进行加窗，加窗信号的傅里叶变换称为短时傅里叶变换（short time Fourier transform STFT）

算法步骤

1. 选定窗口类型和长度，设定窗口移动步长。
2. 窗口与信号的开始时刻对齐，截取一段信号。

3. 对接截取得信号进行傅里叶变换。
4. 按照设定好的步长，延时间轴移动窗口，截取一段新的信号，进行傅里叶变换。
5. 重复上述三四步，直到信号结束。

STFT的性质

- 短时傅里叶变换的结果是一个时间函数。
- 信号x(t)乘以窗函数w(t-T)等价于取出信号在T时刻附近的一个片段
- 结果表示为时间位置T和频率f的二维函数。（反应信号频谱随时间变化的特性）

韦伯定律 Weber's law

Weber’s law 表明心理量和物理量之间关系的定律

感觉的差别阈限（just noticeable difference）随刺激量的大小而变大变小，而且表现为一定的规律性，用公式来表示，就是△Φ/Φ = C 为以常数。

• 其中Φ为原刺激量

• △Φ为此时的差别阈限

• C作为一个常数，又称为韦柏率

Gabor函数

> **时域**内，一个**正弦函数**与一个**高斯函数**的**乘积**对应一个**Gabor函数**。

高斯函数的傅里叶变换

- **高斯函数**

$$g(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-x^2}{2\sigma^2}}$$

  - σ决定时间窗口的宽度

- **傅里叶变换**

$$G(\omega)=e^{-\frac{\omega^2\sigma^2}{2}}$$

  - σ也决定频带的宽度

- 高斯函数的傅里叶变换依旧是高斯函数。
- 高斯函数方差大，对应的傅里叶变换后的高斯函数方差小 。

Gabor函数的傅里叶变换

根据卷积定理，Gabor函数的傅里叶变换就是**高斯函数的傅里叶变换G(ω)**与**正弦函数傅里叶变换$F(\omega)=2\pi\delta(\omega-\omega_0)$**的卷积。

$$Gabor(\omega)=G(\omega)*F(\omega)=e^{-\frac{\omega^2\sigma^2}{2}}*2\pi\delta(\omega-\omega_0)$$

$$Gabor(\omega)=2\pi e^{-\frac{(\omega-\omega_0)^2\sigma^2}{2}}$$

> Gabor函数是带通滤波器，因为它在频域上表现为一个以**ω**为中心的高斯形频谱，仅允许附近频率成分通过；其**中心频率**由**ω0**决定，而**频带宽度**由**高斯窗宽度σ**控制。

> 如果确定时间窗口宽度σ与频率ω0的反比关系，例如：**σ=C/ω0**，其中C是正常数，则随着频率ω0增高，时间窗口宽度σ降低，从而实现时间-频率分辨率的自动调节。

Gabor函数又称**Morlet小波**

Gabor滤波

> 是复Morlet小波与信号的卷积在时域的一个卷积过程，实现时频分析 。

Gabor滤波本质：

- 从信号中找与复Morlet小波函数类似的频率特性成分。
- 是一个带通滤波器。

时间窗口的长度与分析的频率相关：

- 构造复Morlet小波的正弦波频率越低，窗口越长，分析的频率越精细。
- 构造复Morlet小波的正弦波频率越高，窗口越短，分析的时间精度越高，但是分析的频率分辨率越低。

小波变换

定义

小波（wavelet）是一类特殊的有限长度、正负交替、均值为零、两端衰减为零的波形函数。

性质

- 在频域也有衰减性质
- 小波变换没有核函数（对比：傅里叶变换有特定的复正弦函数作为核函数。）

条件

- 容许条件
  - 小波函数的傅里叶变换
- 正规性条件
  - 小波函数的前n阶原点矩等于0
  - n越大越好

典型基本小波

- Meyer小波
- Marr小波
- Morlet小波

基本小波与小波基函数

- 母小波
  - 满足相容性和正规性条件的小波可以作为基本小波，称为母小波ψ (t)
- 小波基函数-小波基
  - 母小波经伸缩平移获得的函数族$\{\psi_{a,b}(t)\}$为小波基
  - $\psi_{a,b}(t) = 1/a^{1/2}\psi((t-b)/a)$
  - a和b是实数，a>0
  - a是尺度因子，a>1时小波伸展；0<a<1时小波收缩。
  - b是平移因子，调整小波的时域位置，即时间窗口的中心
  - a,b联合确定分析的时间窗口宽度和中心位置。

连续小波变换

信号为f(t)，母小波函数为$\psi_{a,b}(t) = 1/a^{1/2}\psi((t-b)/a)$

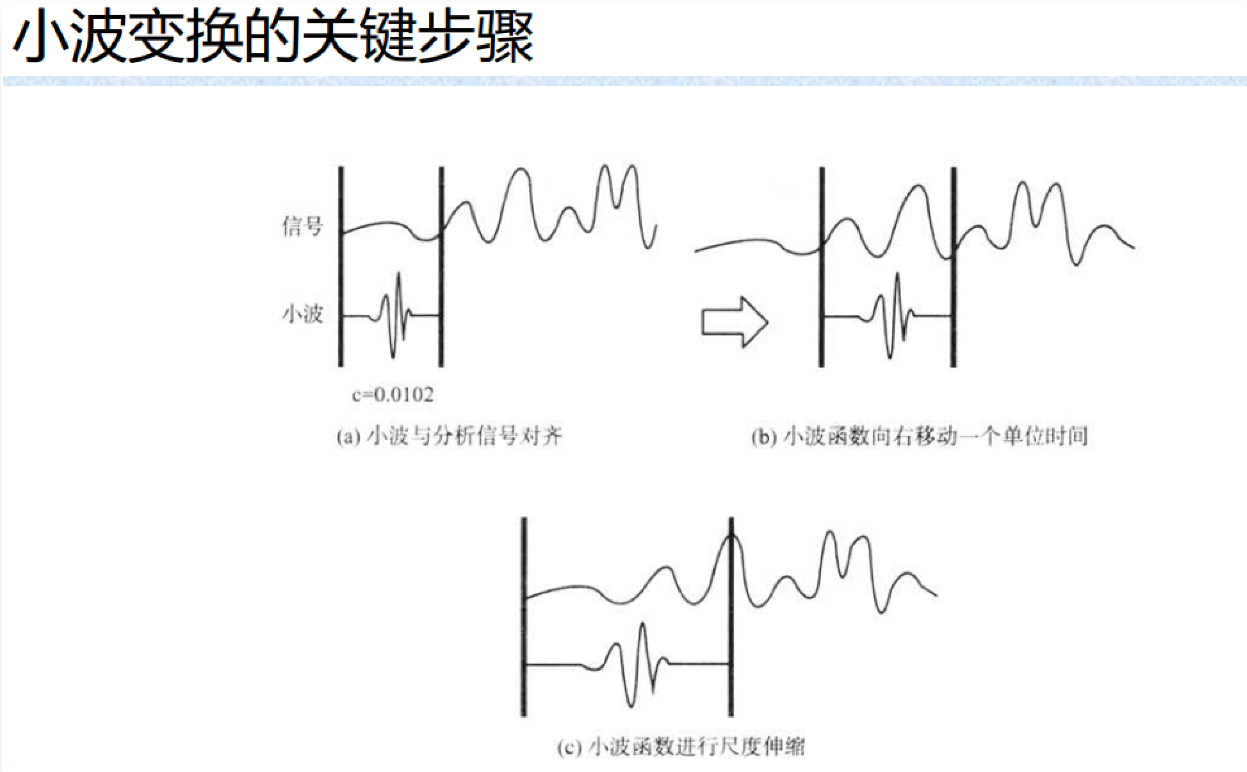
信号的小波变换：

$W_f(a,b) = 1/a^{1/2}\int_{-\infty}^{\infty}\psi((t-b)/a)f(t)dt$

> a是尺度因子，为正实数， b是平移因子，可是任意实数
>
> t, a和b的取值均为连续变量，上述变换成为**连续小波变换CWT**

尺度系数a与窗口长度的关系：

- 记得高斯窗的数学形式即容易理解
- 尺度系数a变小，则时间上压缩，获得高的时间分辨率，有助于分析瞬时变化的时间细节
- 尺度系数a变大，则时间上拉伸，获得高频率分辨率，有助于分析低频细节



连续小波变换的步骤⭐（必考）

- 1）选择一个**小波函数**，**尺度系数为a1**，将这个小波与待分析信号**起点对齐**。
- 2）计算**小波函数**与**信号**的**卷积**，获得小波变换系数C，小波系数越大，该段信号与小波函数波形越像。
- 3）将小波函数沿时间轴右移一个时间单位b，重复第2步，获得新的变换系数。
- 4）重复第2-3步，直到信号结束，获得尺度a1下的所有小波系数。
- 5）改变小波尺度系数成a2，重复第1-4步，获得尺度系数a2下的所有小波系数。
- 6）重复步骤1-5，获得所有感兴趣的尺度系数a_i（i=1, 2, … …, N）对应的小波系数。

小波的特点

- 能分析信号的局部时间频率特性
  - 例如能发现某个时间点附近的局部信号畸变
  - 这个畸变能体现在傅里叶谱上的某个尖峰，但傅里叶变换无法体现发生局部畸的时间信息
  - 小波分析可以准确地显示什么时间段发生畸变；因此小波变换也被称为信号分析的“数学显微镜”

- 小波变换在不违背不相容原理的基础上，根据需要实现可变分辨率

离散小波变换的基本原理

理论上，在连续小波变换中，尺度因子ai>0，位移因子bi，和时间ti都是连续的。

对于**离散数字信号**而言，**bi和ti的最小单位是采样间隔**，由**采样率**决定， a仍可以是任意正实数。

- bi的步长在理论上可以是≥1个采样间隔
- 过度密集采集ai和bi会导致巨大的计算量
- 过度密集采样会产生大量冗余数据
- 不足的采样导致信息的损失

> 对小波变换中的系数a和时移b参数离散化，获得离散小波变换dwt

离散方案：

- 尺度a按照幂级数离散化$a = a_0^j$，j时整数，可正可负。
- 对时移b的离散化为关于尺度系数的函数，如$b = ka_0^j$、$k = b/a = b/a_0^j$

`j` 确定小波**频率窗口**的位置； `k` 确定小波**时间窗口**的位置。

> 离散小波变换dwt的实质——**对时-频空间的采样**。

信号近似项和细节项的分解

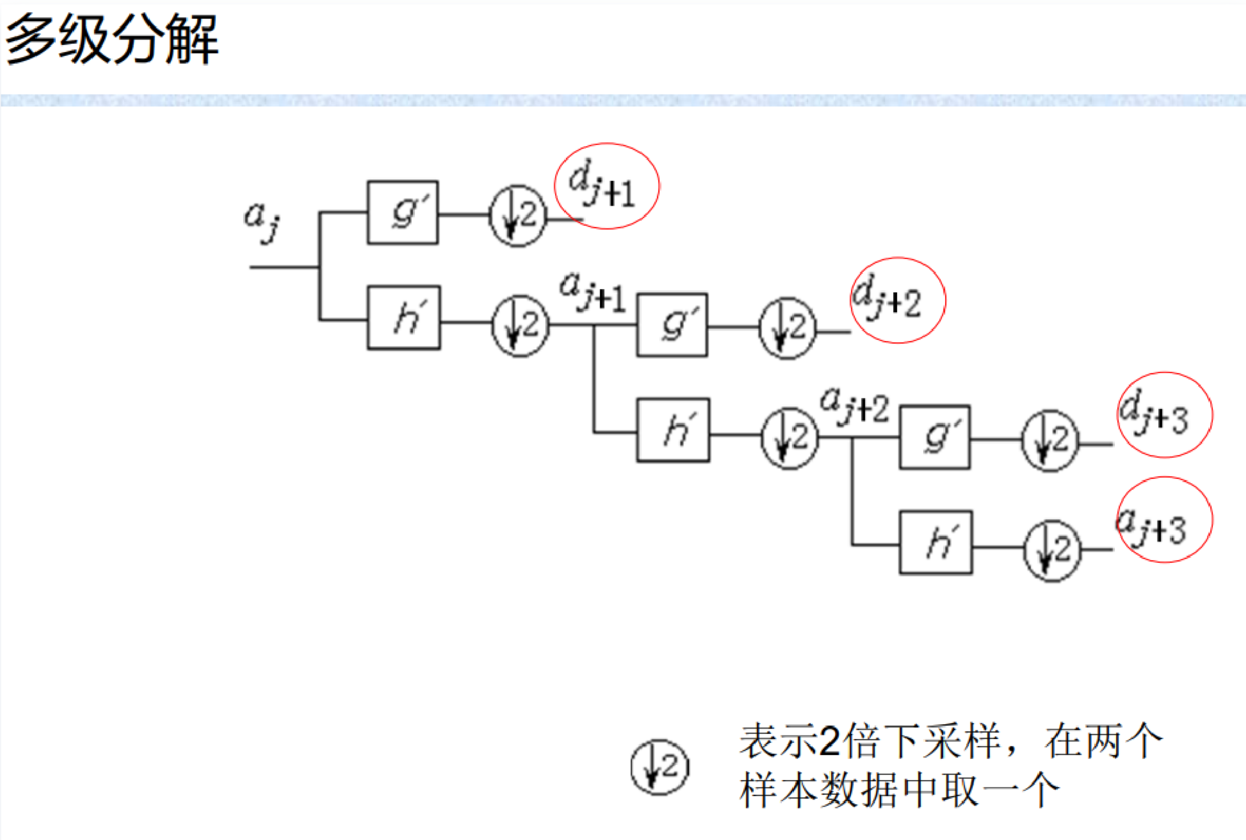
> Gabor函数本质上是一个带通滤波器，可提取信号中一系列频带的成分。
>
> 小波变换可实现信号从高频到低频的逐级分解。

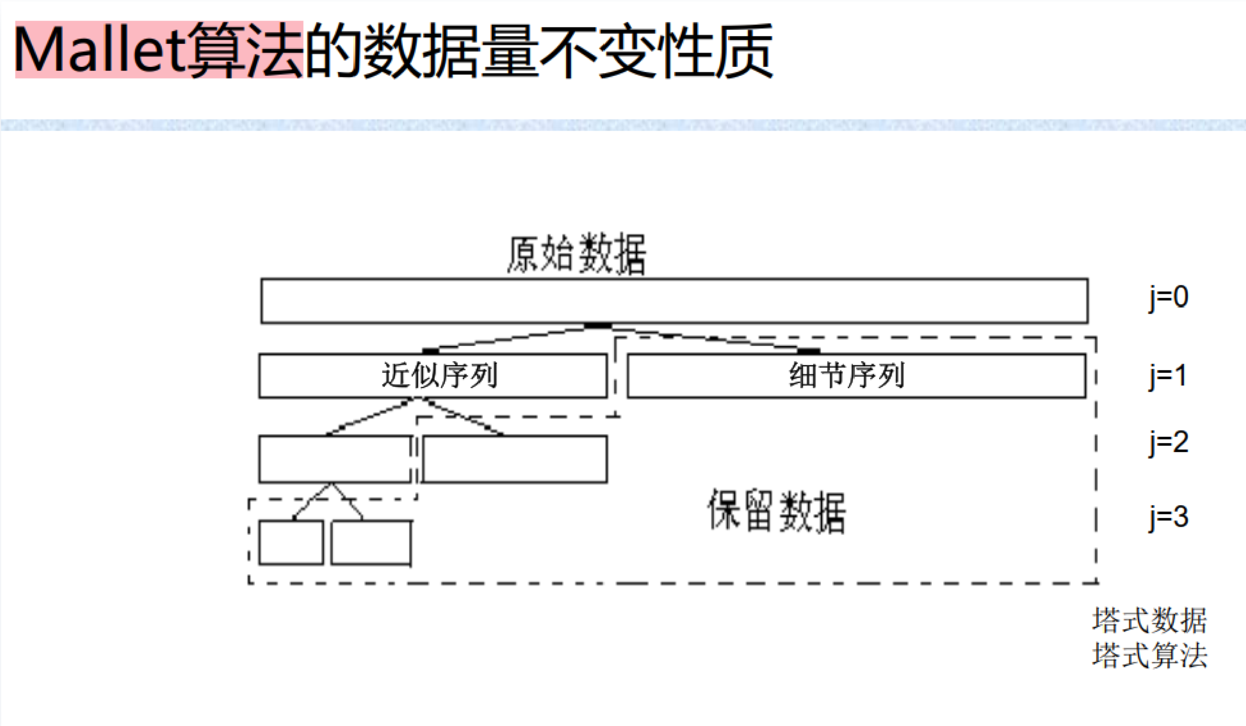
```
                |   |高通 —细节项detail_{j+1}
               /
信号appr_j
               \
                |低通|   |—近似项appr_{j+1}
```

离散小波变换dwt的Mallet快速算法

步骤：

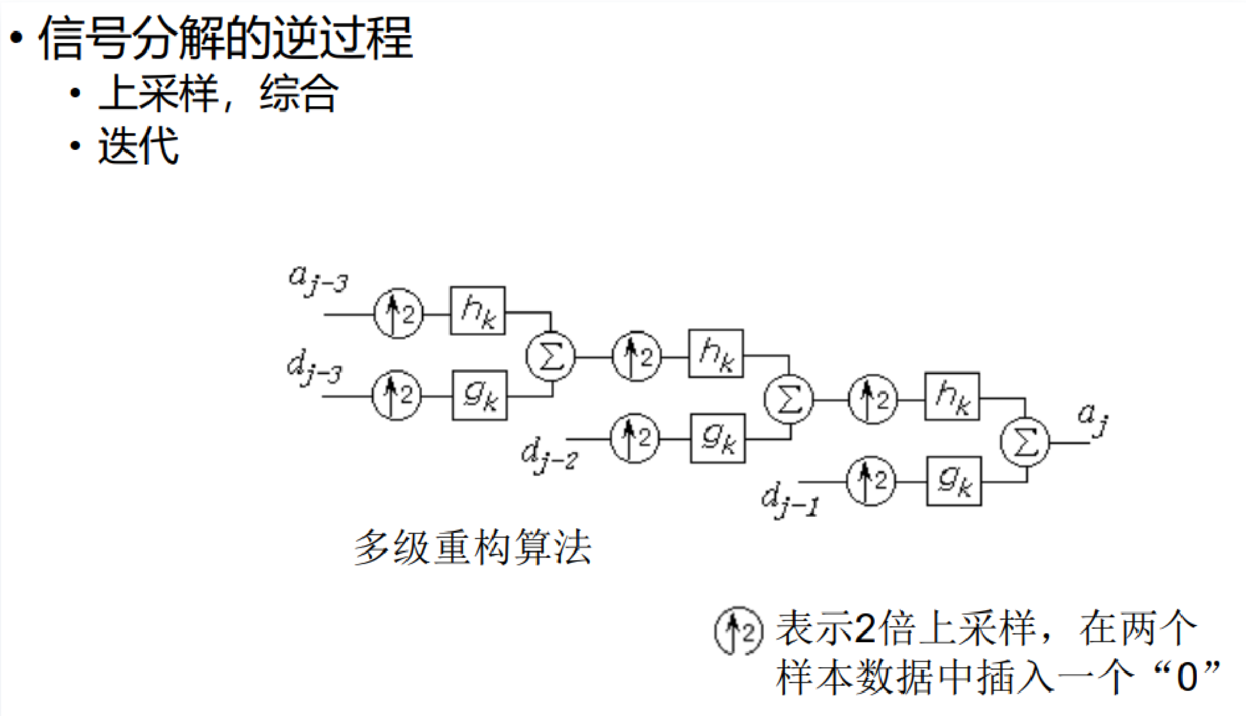
1. 信号通过低通和高通滤波，获得近似项和细节项。
2. 对滤波结果进行基-2的下采样。
3. 对上述近似项再次进行低通和高通滤波
4. 迭代步骤（2）和（3）直到获得预定的分解次数。





> 完整，没有冗余。

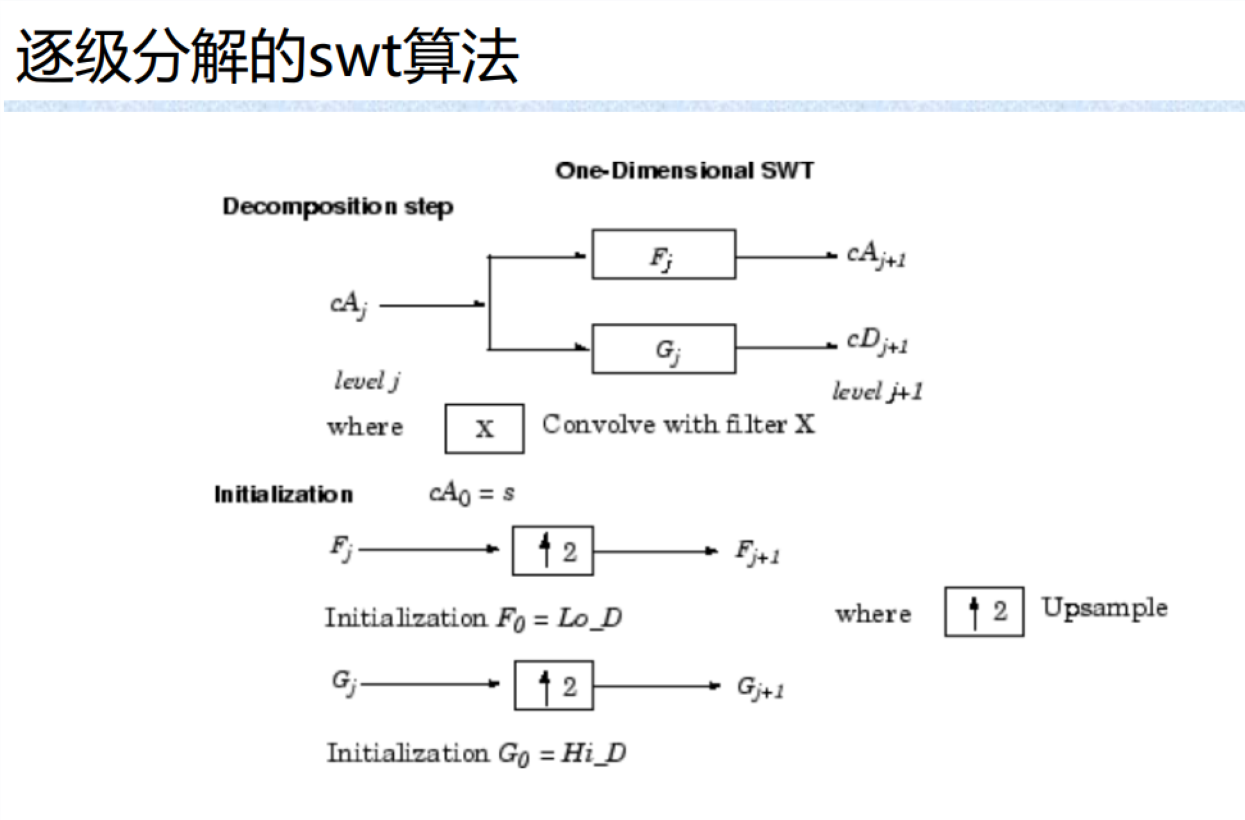
离散小波的逆变换



平稳小波变换SWT

> 原理：对小波函数进行上采样，从而增大小波尺度系数，获取低频部分的逐级分析。

- 优点：小波系数与信号具有一致的时间对应关系——时不变特性。
- 不足：比Mallat算法占用更多内存。



- 信号分解
  - 相比dwt，小波系数**无2倍下采样**的过程，保持小波系数与信号的长度一致（**时不变**）。
  - 但是要逐级对滤波器**g' h'** 进行**上采样**。
  - 迭代，实现多级分解。
- 信号分解的逆过程
  - 相比dwt，小波系数**无2倍上采样**的过程，保持小波系数与信号的长度一致（时不变）。
  - 但是要逐级对滤波器**h_k g_k**进行**下采样**
  - 迭代，实现多级重构

对比项	Mallat算法（DWT）	平稳小波变换（SWT）	关系
滤波结构	高通 + 低通 + 下采样	高通 + 低通 + 无下采样（上采样滤波器）	SWT 是 Mallat 的改进版
系数长度	每层减半	每层等长	SWT 系数包含更多冗余信息
时移敏感性	对平移敏感（时变）	平移不敏感（时不变）	SWT 改进 DWT 的缺陷
内存与计算量	较低	较高	SWT 占用更多资源
应用	信号压缩、快速分解	去噪、特征提取、模式识别	应用不同