**生物医学信号处理知识点部分总结**

1. **周期图**（Periodogram）：一个经典功率谱估计方法。

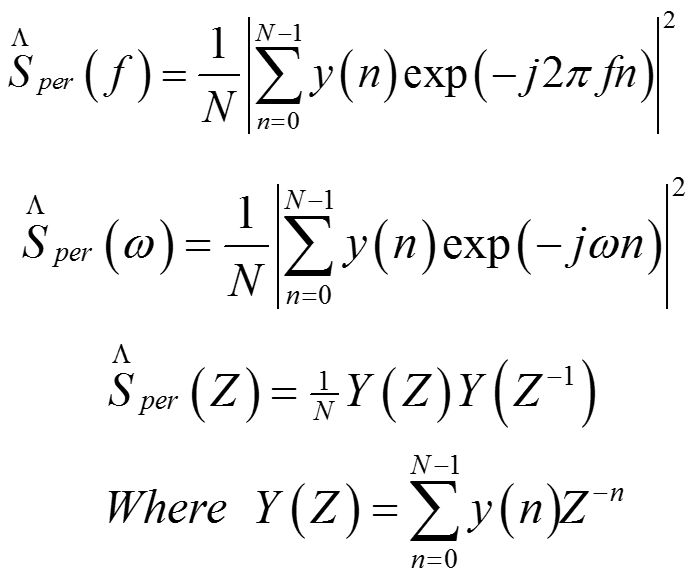
特性：(1)一个有偏估计量;

(2)不一致的估计量;

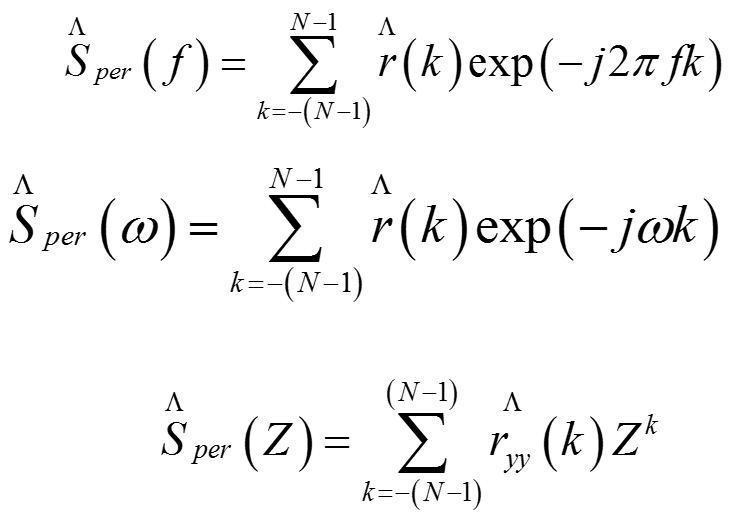
(3)频率分辨率∝ *N*

功率谱密度（Power Spectral Density）：假设y(n)是一个有界数字序列长度n。

直接方法：



间接方法：



直接法和间接法的关系：在M=N-1时，直接法和间接法结果一样；间接法往往取M<<N-1，此时结果不一样。

**平均周期图**是一种改进周期图的方法，也叫做Welch’s method，



与周期图方法相比:

(1)平均周期图方法的方差下降了1 / L;

(2)频率分辨率较差;

(3)通过补零,它可以增加频率的观察点,但没有增加频率分辨率。

1. **Blackman-Tukey method**: 改进周期图的性能方法。





1. **听觉诱发电位分析**Analysis of Auditory Evoked Potentials(AEPs)

听觉系统有非线性特征；当刺激的带宽增加或听力损失的程度增加时, 该系统的相关函数降低。

1. Analysis of Heart Rate Variability
2. 相关分子生物学知识

DNA: 脱氧核糖核酸（DNA，为英文Deoxyribonucleic acid的缩写）,又称去氧核糖核酸，是染色体的主要化学成分，同时也是组成基因的材料。有时被称为“遗传微粒”

基因: DNA分子上具有遗传效应的特定核苷酸序列的总称

(1) 具有遗传效应的DNA分子片段;

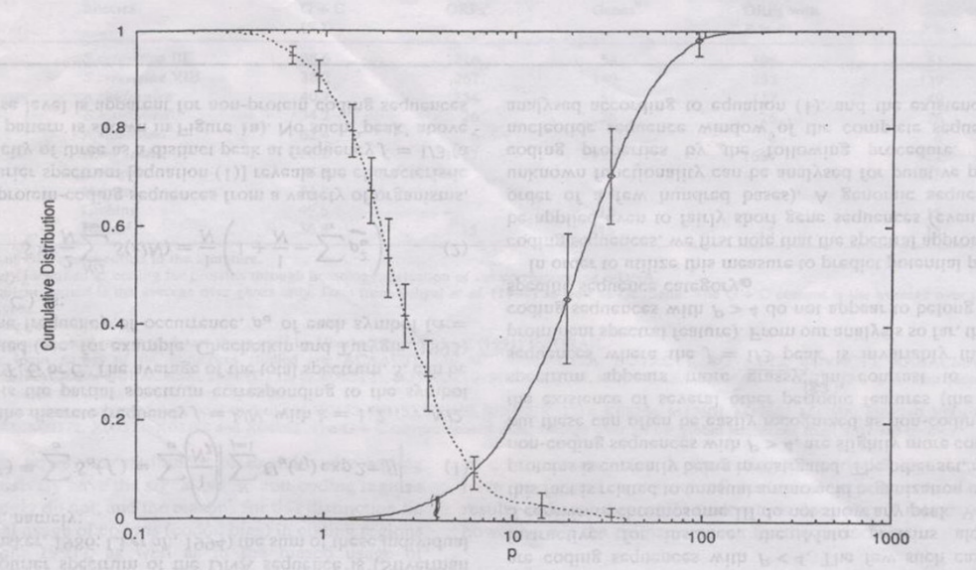
(2) 位于染色体上，并在染色体上呈线性排列;

(3) 能够表达一定基因产物的DNA序列。

基因组:生物细胞中所有的DNA,包括所有的基因和基因间区间。每个生物都具有基因组。绝大部分基因组，由DNA组成，但有一些病毒具有RNA基因组。

1. **Prediction of probable gene by Fourier analysis**

在该文中是使用蛋白质编码区的1/3周期性（原因：1、密码子的三联体性2、密码子的不规则使用，即密码子偏置）



优点:

(1)容易实现;

(2)不依赖训练组基因;

(3)不需要基因的先验信息;

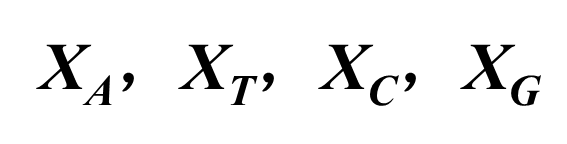
(4)在碱基中独立变化

(5)和排序错误

缺点:

(1)位置误差很大,约54个碱基;

(2) 在小比例的基因中缺乏1/3周期性;

本文的方法步骤：1、DNA序列的映射，把他们映射成四个子序列

2、通过傅里叶分析计算功率谱

3、定义局部信噪比PM (j) = SM (1/3) / AP，其中M(< N)是滑动窗口的长度，SM(1/3)在1/3频点的功率谱密度（PSD），AP是平均功率谱密度. PM (j)是编码区和非编码序列的分类器，PM (j)> 4时是编码序列，PM (j) < 4时是非编码序列。4、做出预测

1. **最小均方误差的自适应噪声抵消算法**（LMS adaptiv noise canceling algorithm）



1. **用AR来模拟癫痫脑电**（AR Modeling of Seizure EEG）

传统方法：（1）FFT：不适合EEG信号的非平稳的特点

（2）周期图法（Periodogram method）：需要平均,但实掩盖了频域的动态变化

AR分析方法（AR analysis method），用来分析局灶性癫痫大鼠的多通道EEG

AR方法描述：（1）动物模型

（2）使用左额叶和左和右枕叶通道记录表面脑电图

（3）采样率为128 hz,十天之后

（4）记录2 - 14天。

（5）使用AR和priodogram方法得到1 - 2秒PSD片段

1. **基于卡尔曼滤波理论的脑电逆问题反演**

　引　言

（1）脑电图( EEG)是由分布在头皮上的电极无创测量得到的脑电信号。

(2)脑电技术的一个关键问题就是如何进行脑电逆问题的求解,即如何通过测量得到的电位来推断出大脑内神经活动源的信息。

(3)脑电逆问题的解决对脑功能的研究有着重要的科学意义和临床应用价值。

（4）在求解脑电逆问题的过程中存在以下难点:

一是解的非唯一性,即从头皮测量得到的电位分布不能

唯一地确定脑电源的数量、位置和分布, 这是逆问题

求解过程中最主要的问题;

二是解的不稳定性,即输入很小的噪声或扰动都会引起

解的振荡。

三是测量数据的不完备、观测误差和随机干扰,都不可

避免地会造成出现多个解。

（5）现有方法：

最小模法(MNE )：

最小模法是分布源模型中出现最早且被普遍采用的成像方法, 其出发点就是选择具有最小模的值为最终解,它对应于一幅具有最小能量的电流密度分布图像。 该算法是寻找最小能量解的过程, 其代价函数为:

min | J |2 =min J T J , J 表示由多个点的电流密度组成的矩阵, J T 是J 的转置矩阵。

多信号分类算法(MUSIC )：

是一种典型的偶极源定位方法,这种算法是在信号子空间算法的基础上,运用单个偶极子进行一次3维空域网格点的搜索扫描来完成对多个脑电源的空域定位。

（6）现有方法存在的问题：

传统方法很少考虑到脑电系统中不确定因素影响

的问题, 这跟实际情况有很大的差距。

由于在求逆过程中,如果不对噪声做任何分析,那么势必会影响到脑电源定位的精确度,而且在运用中也将会受到很大限制,因此有必要寻求一种更为有效的求逆方法。

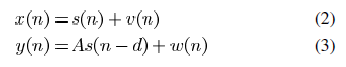
作者提出了求解脑电逆问题的新途径,即将卡尔曼滤波法应用到脑电逆问题的求解过程中。这种基于状态空间的求逆算法,不仅可以很好地处理脑电系统中的不确定因素,而且还可以将静态和动态脑电逆问题的求解统一到同一框架下。

1. **Establishing Causality With Whitened Cross-Correlation Analysis**

**建立与美白交叉相关分析的因果关系**

有许多信号处理和时间序列分析技术来统计上描述两个随机过程之间的关系，如：互相关分析、传递函数估计、参数化建模、一致性分析。然而,能够应用在确定两个过程之间是否存在因果关系的技术很少。

关于延迟的方法：信号的统计模型：



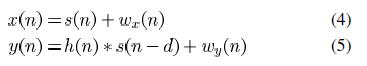
x(n) , y(n):观测信号

v(n),w(n):互不相关的白噪声过程

s(n):原始信号

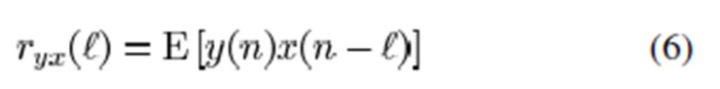
d：延迟率

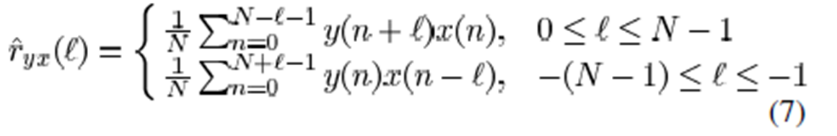
更一般的统计模型：



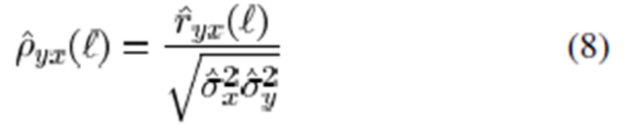
h(n): 一个未知的LTI系统的脉冲响应

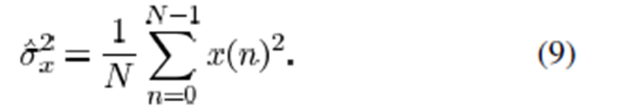
计算互相关（Computing Cross-correlation）





Cross-correlation function (CCF)：





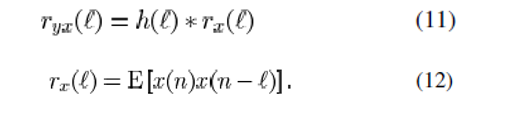
**不相关或相关是什么?**

答：（1）对所有延迟，CCF都等于0时表示两个过程是独立的或不相关的

（2）在一个特定的延迟CCF接近1，表示两个过程之间存在一个纯延迟d样品。

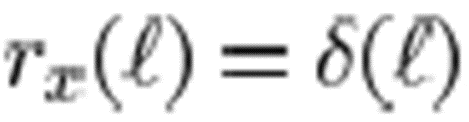
为什么要白化？

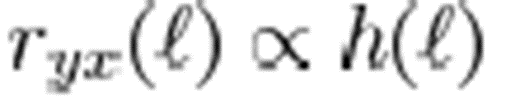
答：



(1）The cross-correlation depend on both *rx(l)* and *h(l)*;

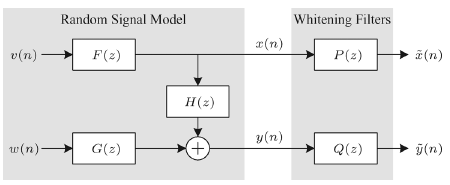
(2) The cross-correlation is affected by *rx(l)* as much as it is affected by the system H(z);

If x(n) is a white noise process such that

Then 

The cross-correlation fully characterizes the relationship that we are interested in.

白化的方法：



**主成分分析(PCA)原理**

PCA 的目标是寻找 p 个新变量，使它们反映事物的主要特征，压缩原有数据矩阵的规模。每个新变量是原有变量的线性组合，体现原有变量的综合效果，具有一定的实际含义。

这 p 个新变量称为“主成分”，它们可以在很大程度上反映原来变量的影响，并且这些新变量是互不相关的，也是正交的。

假设原始数据为data,那么它的协方差：

 (5-1)

对协方差矩阵R进行奇异值分解：

[u,d,v]=SVD(R) (5-2)

其中，u和v分别是特征向量，d是对角线为A阵奇异值的对角阵，且奇异值是按从大到小的顺序排列的。

把原数据向特征空间投影，得到原数据data的主成分

PCA＝ （5-3）

若需要取原数据A的前k个主成分，则

PCAk＝ (5-4)

投影到原始数据

Data’=data\*U(:,1:k)\*U(:,1:k)’

如果需要计算这k个主成分在原数据中的所占比例，即k个主成分的累计方差贡献率，则

 (5-5)

通常当>90％时，可以确定保留这k个主成分比较合适，而其余的主成分则可以略去。

PCA存在的问题及限制：

* PCA方法存在容易丢失重要信息的弱点，即一些包含有重要信息的主成分由于其方差的贡献率很低，而排列在其前面的主成分累积贡献率已达到很高的比例（如85%以上），因此该主成分极易被当作噪声信号丢弃。
* 要求“主成分”是互不相关的，也是正交的。
* 原始变量到“主成分”之间的变换是一种线性变换，即线性空间变换。

奇异值分解（SVD）:

假设M是一个m×n阶矩阵，其中的元素全部属于域 K，也就是 实数域或复数域。如此则存在一个分解使得

M = UΣV\*，

其中U是m×m阶酉矩阵；Σ是半正定m×n阶对角矩阵；而V\*，即V的共轭转置，是n×n阶酉矩阵。这样的分解就称作M的奇异值分解。Σ对角线上的元素Σi,i即为M的奇异值。

SVD与PCA: 奇异值分解在统计中的主要应用为主成分分析（PCA），种数据分析方法，用来找出大量数据中所隐含的“模式”,它可以用在模式识别，数据压缩等方面。PCA算法的作用是把数据集映射到低维空间中去。 数据集的特征值（在SVD中用奇异值表征）按照重要性排列，降维的过程就是舍弃不重要的特征向量的过程，而剩下的特征向量组成的空间即为降维后的空间。

**NMF，非负矩阵分解**，它的目标很明确，就是将大矩阵分解成两个小矩阵，使得这两个小矩阵相乘后能够还原到大矩阵。而非负表示分解的矩阵都不包含负 值。 从应用的角度来说，矩阵分解能够用于发现两种实体间的潜在特征，一个最常见的应用就是协同过滤中的预测打分值，而从协同过滤的这个角度来说，非负也很容易 理解：打分都是正的，不会出现负值。

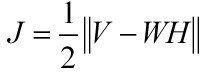
给定一个非负矩阵V，我们可以找到2个非负分解因子矩阵W，H.其中，W称为权重矩阵，H称为特征矩阵。公式如下：

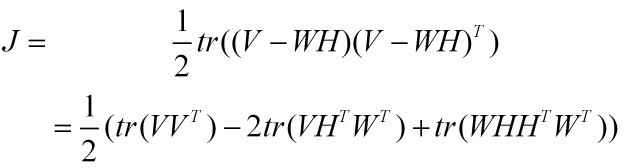
[NMF](http://photo.blog.sina.com.cn/showpic.html#blogid=642c9bdd01010iun&url=http://s1.sinaimg.cn/orignal/642c9bdd4b625e45eff30)

当n维数据向量集放置在n\*m的矩阵V中，其中m是数据集的大小。V分解的的权重矩阵W：n\*r；V分解的特征矩阵H： r\*m； 通常r小于n，m。 r表示采用的特征数。这个可以认为是一种数据压缩方法。

这里，关键为题是如何计算出W，H矩阵？

首先我们构造一个代价函数：

[](http://photo.blog.sina.com.cn/showpic.html#blogid=642c9bdd01010iun&url=http://s10.sinaimg.cn/orignal/642c9bdd4b62629c3aaa9)代价函数J可以重写为：

[](http://photo.blog.sina.com.cn/showpic.html#blogid=642c9bdd01010iun&url=http://s16.sinaimg.cn/orignal/642c9bdd4b6264f3bd42f)

这是个典型的优化问题，这里采用拉格朗日乘法去优化J。

构造拉格朗日函数L：

[NMF](http://photo.blog.sina.com.cn/showpic.html#blogid=642c9bdd01010iun&url=http://s1.sinaimg.cn/orignal/642c9bdd4b6266f0c4db0)  
  
说明如下：W，H的每个元素必须大于等于0，这样设置

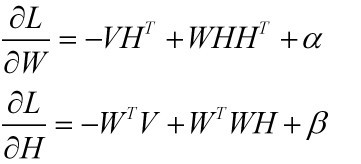
[NMF](http://photo.blog.sina.com.cn/showpic.html#blogid=642c9bdd01010iun&url=http://s7.sinaimg.cn/orignal/642c9bdd4b62679e07676)

为

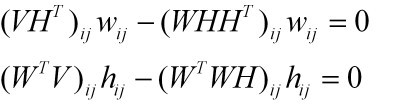
[NMF](http://photo.blog.sina.com.cn/showpic.html#blogid=642c9bdd01010iun&url=http://s1.sinaimg.cn/orignal/642c9bdd4b626edf5e930)拉格朗日乘法因子。其中

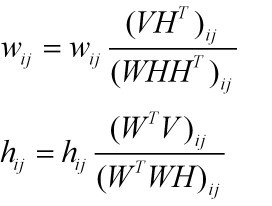
[[NMF](http://photo.blog.sina.com.cn/showpic.html#blogid=642c9bdd01010iun&url=http://s1.sinaimg.cn/orignal/642c9bdd4b626897b2b70)。](http://photo.blog.sina.com.cn/showpic.html" \l "blogid=642c9bdd01010iun&url=http://s1.sinaimg.cn/orignal/642c9bdd4b626897b2b70" \t "_blank)

这样，做L对W，H进行求偏导如下：

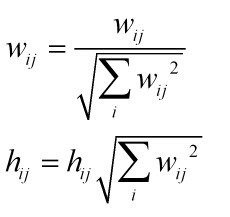
[](http://photo.blog.sina.com.cn/showpic.html#blogid=642c9bdd01010iun&url=http://s11.sinaimg.cn/orignal/642c9bdd0789d7be2506a)根据KKT条件：

[NMF](http://photo.blog.sina.com.cn/showpic.html#blogid=642c9bdd01010iun&url=http://s4.sinaimg.cn/orignal/642c9bdd4b626e86edb53)所以，得到如下：

[](http://photo.blog.sina.com.cn/showpic.html#blogid=642c9bdd01010iun&url=http://s3.sinaimg.cn/orignal/642c9bdd4b6270eec0622)  
根据以上等式，我们可以得出一个乘法迭代更新公式：

[](http://photo.blog.sina.com.cn/showpic.html#blogid=642c9bdd01010iun&url=http://s7.sinaimg.cn/orignal/642c9bdd4b6271e3e0e16)到目前为止，我们基本上描述完了利用乘法更新规则求解NMF的过程。其实，可以看出来，根据上面的规则：W，H不是唯一的。

若果想使W，H唯一，我们可以设定（当然这只是其中一种假设方法）W的列向量欧式距离长度为1，就对W进行列的归一化处理：

[](http://photo.blog.sina.com.cn/showpic.html#blogid=642c9bdd01010iun&url=http://s14.sinaimg.cn/orignal/642c9bdd4b6276c301d7d)

NMF应用：(1) 图像分析

(2) 文本聚类/数据挖掘 (3) 语音处理