

第一章 概 论

§ 1.1 化学计量学技术及发展概况

科学的发展与技术的进步使得化学量测工作逐步仪器化、自动化和计算机化。现代分析仪器能迅速、准确地为人们提供大量可靠的量测数据。化学工作者面临着如何选择合适的实验方法和最优量测过程,对原始量测数据进行再加工,从而最大限度地提取有用的化学及其相关信息。随着计算机科学、应用数学和统计学方法在化学中应用的日益广泛和深入,一门新的化学分支学科即化学计量学(Chemometrics)诞生了。20多年的学科发展实践证明,化学计量学理论和方法已渗透到化学中的各个领域,已经成为化学量测的基础理论和方法及非常实用的技术。

化学计量学诞生于20世纪70年代初期。1971年,瑞典化学家Wold在为—项基金项目定名时,首次提出了“Chemometrics(化学计量学)”—词,标志化学计量学这门新兴学科的诞生。三年后(1974年),他与美国华盛顿大学的Kowalski教授在美国西雅图成立了国际化学计量学学会(ICS)。80年代随着计算机的普及应用,化学工作者不仅应用现有的数学和统计学方法,而且根据化学学科的特殊性要求创建了一系列化学量测数据的处理、分类、解析与预测等一大批化学计量学方法,编制了许多优秀的化学计量学软件,很多软件已成为现代化学量测仪器,主要是分析仪器的有机组成部分。80年代后期,化学计量学课程开始进入化学教学大纲,在一些分析化学杂志,如*Anal. Chem.*, *Anal. Chim. Acta.*, *Anal. Lett.*, *Trend. Anal. Chem.*, *Talanta*, *Mikrochim. Acta.* 和 *Appl. Spectrosc.* 等,出现了大量化学计量学方面的文章,专门刊登化学计量学学术研究成果的学术期刊“*Journal of Chemometrics*”, “*Chemometrics & Intelligent Laboratory Systems*”问世,交流化学计量学研究成果的国际学术会议也陆续召开。进入90年代,由于计算机及软件技术的飞速发展,符号处理高级语言的普及,使许多过去认为过于复杂,难于普及的化学计量学算法逐步得到推广与应用,特别是信息高速公路的建立,广大化学计量学工作者可以更快、更及时地了解化学计量学学科发展动态,在1994年和1996年已分别召开了两届Internet国际化学计量学学术会议,化学计量学作为与计算机科学等多学科相联系的交叉学科已得到了极大的发展。

§ 1.2 化学计量学主要研究内容和应用情况

化学计量学是数学和统计学、化学及计算机科学三者相互交缘而形成的一门边缘学科,是化学中很具有实力和应用前景十分广泛的新兴分支学科。按照国际化学计量学学会(International Chemometrics Society)的定义,化学计量学是化学的一门分支学科。它应用数学和统计学方法,设计或选择最优量测程序和实验方法,并通过解析化学量测数据而获取最大限度的信息。理解这个定义应考虑以下几点:

(1)要选择最优量测程序并获取最大限度信息,必须也只有借助计算机技术;(2)化学计量学研究 and 探讨各种化学量测过程的共性问题,如化学试验设计与优化、化学数据解析及有用信息的提取等,因而它是有关化学量测的基础理论和方法学;(3)化学计量学是化学、数学和统计学以及计算机科学诸多学科的“接口”,但同时应该注意到化学计量学又是一个学科总体,有其自身的学科体系。另一方面,化学计量学可以更具体地表达为研究应用数学和统计学方法,借助计算机技术,进行化学量测的试验设计、数据处理、分类、解析和预测的一门学科。

化学计量学的研究范围极为广泛,内容非常丰富。化学试验设计与优化、定量校正理论、分析信号处理、化学模式识别、模型与参数估计、数据解析、过程模拟、人工智能、情报检索、实验室自动化等等都是化学计量学的研究范围。化学计量学作为化学量测的基础理论与方法学,其应用非常广泛。可以说,凡是施行化学量测的所有领域,如工业过程采样、过程分析化学、过程控制、食品工业、海洋化学、地球化学、环境化学、造纸工业、石油勘探、临床诊断、制药工业、染料工业、有机合成化学、生物工程、材料工程等等都可以应用化学计量学。

化学计量学包含了一些非常实用的技术,应用该技术可从大量的数据中抽提出有用的信息,并尽可能地滤除随机噪声的影响。如在分析化学中采用各种滤波、平滑、变换、卷积技术和最优化技术,消除干扰,解析重叠峰信息,可提高灵敏度和准确性,改善选择性。

§ 1.3 本书主要内容

在概论中介绍化学计量学技术及发展概况,主要研究内容和应用情况及本书主要内容。

在第一篇中介绍化学信号处理:第二章包括化学信号与信息特征、信号的符号表示、信号与噪声、信号分辨和信号的属性。第三章介绍信噪比软件增强技术,包括化学信号的拾取、模拟信号的傅里叶级数和傅里叶变换、信号的采样与复原、信号与噪声的频谱分析、傅里叶变换滤波器及一般实用数字滤波器。第四、五章中介绍

相关化学信号处理,包括信号的相关性、交叉相关和锁相方法、卷积和退卷积、相关化学测量、波谱解谱技术,包括复杂信号的测量、信号组分的分离.第六、七章中介绍时-频分布处理,整体平均值和局部平均值方法及短-时傅里叶变换和小波变换,包括短-时傅里叶变换等.在第二篇中主要介绍多变量数据处理分析技术:第八章中介绍回归分析技术,包括一元线性回归、多元线性回归、岭回归、逐步回归和主成分回归方法.第九章中介绍模式识别判别分析技术,包括多重判别矢量法、Fisher判别矢量法、主成分分析法、非线性映射法.第十章介绍偏最小二乘法,包括偏最小二乘法中噪声过滤原理和 PRESS 判据,偏最小二乘方法原理和基本算式及用于 PLS 和 PCR 的简单迭代算法.第十一章中介绍变量扩维-筛选方法,包括该方法的原理及一些新的变量筛选方法,其中有正交递归选择法、遗传算法、模拟退火算法与遗传算法结合的方法.第十二章中介绍多变量分析方法在工业诊断和优化中应用的一些实例.第三篇介绍其他化学计量学技术:其中第十三章介绍人工神经网络方法,包括基本原理,网络的结构与优化,人工神经网络的各种模型和人工神经网络的应用举例.第十四章中介绍在电感耦合等离子体原子发射光谱法干扰校正中应用的一些化学计量学方法,包括光谱干扰的校正技术,有广义标准加入法、次因子重叠设计分析法、矩阵投影算法、卡尔曼滤波法及光谱线及其相互干扰的计算机模拟方法。

第一篇 化学信号处理^[1~8]

第二章 信号处理概述

§ 2.1 化学信号与信息特征

顾名思义,化学信息是指与化学过程有关的信息.信息可以通过测量化学过程产生的物理和化学信号而获取.广义地说,一个化学过程可以被简述成:



该过程表示化学物质 A 和 B 反应生成了产物 C, A 和 B 都是比原子大的化学物质,包括原子、分子、高分子及聚合、簇等化合物.有时,价电子,包括外层和内层的核内质子、中子的振动跃迁及核反应也被作为化学反应来处理,但实际上,这类反应也可以归结到原子物理的范畴.

化学信号大都是由物理变化引起的,比如说,由于化学变化而引起了颜色的变化,溶液电位的变化,原子和分子能量的变化等等,这些物理变化被某一些特定的物理和化学传感器所感知而转换成电信号.这些电信号经过测量和处理后便组成化学数据.从这些化学数据的性质及组成便可知道所需要的化学信息.

哪些类的信号可以被称为化学信号或者可以用来获取到化学信号呢?通常来说,化学信号可以分为以下几类.

§ 2.1.1 第一类化学信号为波谱信号

广义地波谱信号横跨低频电磁波至 γ 射线,包括甚低频、低频、中频、高频、甚高频、微波、远红外、中红外、近红外、可见、紫外、远紫外、X 射线、 β 射线和 γ 射线.波谱信号与化学过程密切相连,如原子价电子光谱,分子价电子光谱,原子分子内层电子光谱或能谱,原子核谱.价电子光谱一般在红外到紫外,分子的振动带可延长到微波.内层电子的跃迁产生 X 射线光谱.

§ 2.1.2 第二类化学信号为能谱信号

能谱包括光电子能谱,广义地讲,一般 X 射线以上的都可称能谱,因粒子的波动性已大大减弱,而粒子的质量性质则越来越明显.质谱信号是根据原子分子质量的大小而进行质量分散的,但就其测量原理来说,它利用了原子和分子离子在外加电磁场下其加速能量的差异,故也可看作是与能量大小有关的质谱信号.

§ 2.1.3 第三类化学信号为电化学信号

电化学信号比较接近电子学中的测量信号,通常以测量反应中的电势、电位、电导或阻抗及电流的变化为主,其结果经常可以表示为伏安特性曲线图。

§ 2.1.4 第四类化学信号为色谱分离信号

色谱分离的信号与其检测系统有关,可以是第一至第三类化学信号之中的一种,但由于分离信号在时间上的依赖关系,所以可以单独成一类相对色谱值-时间函数关系的信号。

从化学信号到化学信息有许多中间过程,当然,有些化学信号本身可以作为化学信息,如一个反应从无色到有色,表明反应已经发生。通常,化学信号首先要经传感器和探测器等敏感器件转变成电信号。电信号可以经过电学处理(如去噪声、取比值等)后读出,或者再被模数转换成数字信号。数字信号可以被进一步处理,如滤波、测面积或峰值。至此,被处理后的数字信号可以被当做数据点存贮起来。由若干个和集团数据组成的数据块可以被进一步处理(如一般校正和特别的化学计量学校正)以得到化学信息,如浓度、组分、结构等。因此,化学信息的处理可分为两大部分:化学信号处理和化学数据处理。

本篇内容主要讨论化学信号的数值处理,有别于化学信号的电子学测量和最终数据的化学计量学处理。因此,本篇内容中假设化学信号已经被转变成了数字信号,我们对这些化学数字信号进行处理,使之成为化学数据。化学数据处理将在其他篇中涉及。

§ 2.2 信号的符号表示

一般来说,在化学信号测量中我们得到的信号多数可以表示为时域和频域的信号。有关时域与频域的信号特征在以后章节详细介绍。

一些化学信号会随时间变化,从而产生了随时间变化的起伏或称波形,即时域信号,我们用 $f(t)$ 来表示时域上的信号。在原理上,时间信号可以有任何函数形式,而且能够产生非常丰富而复杂的信号,研究复杂的信号要从简单的信号开始,因为从数学分析的角度出发,任何复杂的信号都可以分解为简单的信号组合。因此有必要研究和表达简单的信号。

最简单的时间变化的信号是正弦波,如图 2.2.1 所示。实际上,它是许多基本方程的解。正弦波的特点是具有恒定的幅度 a_0 和固定频率 ω , 即

$$f(t) = a_0 \cos(\omega_0 t + \varphi_0) \quad (2.2.1)$$

信号的强度也可以用能量密度来表示,信号有什么能量,特别是消耗多少能量产生

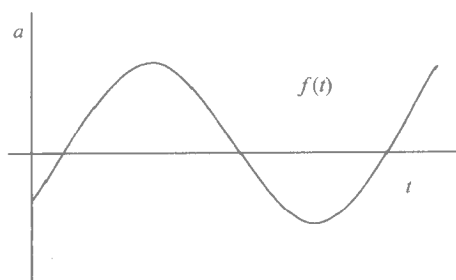


图 2.2.1 正弦波信号

信号是一个重要问题, 能量密度一般等于 $|f(t)|^2$, 也称为瞬时功率. 在时间间隔 Δt 内所消耗的能量可以表示为 $|f(t)|^2 \Delta t$, 那么总能量则可以由下式求得:

$$E = \int |f(t)|^2 dt \quad (2.2.2)$$

对于一个测量值 x , 可以是时间信号、电压信号、或者电流信号, 它的测量平均值可以表示为

$$\langle x \rangle = \frac{\int x |f(t)|^2 dt}{\int |f(t)|^2 dt} = \int x |f(t)|^2 dt \quad (2.2.3)$$

此处用了归一化条件

$$\int |f(t)|^2 dt = 1$$

测量值的标准偏差则为

$$\sigma_x^2 = \langle x^2 \rangle - \langle x \rangle^2 \quad (2.2.4)$$

随时间变化的信号也可以用复数形式来表示:

$$f(t) = A(t)e^{j\varphi(t)} \quad (2.2.5)$$

式中, $A(t)$ 称为幅度函数, 而 $\varphi(t)$ 称为时间相位函数. 应当指出, 实信号的相位与幅度一般与复信号的相位与幅度不尽相同.

式(2.2.1)所表示的时域信号隐含了频域信号的特征 ω_0 , 表示该简单的正弦波具有特定的振动频率, 如图 2.2.2 所示. 实际上, 在频域上研究一个波形信号可以简化对波形的理解, 特别是在时域上多个复杂波形的重叠的信号可以非常简便的在频域上进行研究. 在以后章节将叙述的频谱分析便是分析复杂重叠信号的有力工具. 简单正弦波事实上为一些基本运动方程所共有, 因而是它们的可能解.

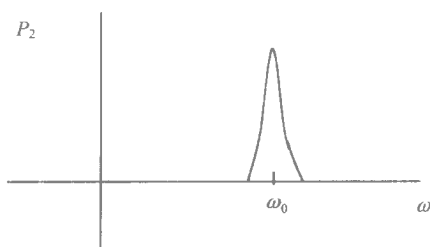


图 2.2.2 正弦波信号的频谱

如果控制化学变量的方程是线性的, 那么正弦波的和也是它们的可能解, 一般来讲, 信号是复杂的, 但实际上, 这种杂乱常常是正弦波的简单叠加, 而叠加对理解和表示复杂信号是比较简单的. 信号可以用于不同频率的正弦波展开为

$$f(t) = \frac{1}{\sqrt{2\pi}} \int G(\omega) e^{j\omega t} d\omega \quad (2.2.6)$$

这个波形由简单正弦波 $e^{j\omega t}$ 线性叠加组成, 每一个正弦波都有一个频率 ω 和用系数函数 $G(\omega)$ 表示的有关量. $G(\omega)$ 可由信号 $f(t)$ 得到:

$$G(\omega) = \frac{1}{\sqrt{2\pi}} \int f(t) e^{-j\omega t} dt \quad (2.2.7)$$

式(2.2.6)和式(2.2.7)描述了傅里叶变换或者时频变换(频谱). 因为 $G(\omega)$ 和 $f(t)$ 是惟一相关的, 所以可以把频谱看作是在频域空间或者频率表示的信号.

频谱也可以表示成幅度与相位的表示式:

$$G(\omega) = B(\omega) e^{j\Psi(\omega)} \quad (2.2.8)$$

式中, $B(\omega)$ 称为频谱幅度, $\Psi(\omega)$ 叫做频谱相位, 以便把它们与信号的相位与幅度区别开来.

与能量密度函数 $|f(t)|^2$ 相对应, $|G(\omega)|^2$ 称为能量密度频谱并满足:

$$E = \int |G(\omega)|^2 d\omega = \int |f(t)|^2 dt \quad (2.2.9)$$

对于频域上的测量值, 也有平均频率 $\langle \omega \rangle$ 及其标准偏差 σ_ω

$$\langle \omega \rangle = \int \omega |G(\omega)|^2 d\omega \quad (2.2.10)$$

和

$$\sigma_\omega^2 = \langle \omega^2 \rangle - \langle \omega \rangle^2 \quad (2.2.11)$$

以及任何频率函数 $g(\omega)$ 的平均值:

$$\langle g(\omega) \rangle = \int g(\omega) |G(\omega)|^2 d\omega \quad (2.2.12)$$

应当指出 $|f(t)|^2$ 和 $|G(\omega)|^2$ 都是密度. 我们称密度的傅里叶变换特征函数, 特征函数是研究密度的一种强有力的方法. 由于傅里叶变换是线性算子, 我们可以定义时域上的平移量 $\tau = t_0$ 和频域上的平移量 $\theta = \omega_0$, 则能量密度的精确函数是:

$$\begin{aligned} R(\tau) &= \int \overline{f(t)} e^{j\omega\tau} f(t) dt \\ &= \int \overline{f(t)} f(t + \tau) dt \end{aligned} \quad (2.2.13)$$

而频域上的特征函数是:

$$\begin{aligned} R(\theta) &= \int \overline{G(\omega)} e^{j\theta\omega} G(\omega) d\omega \\ &= \int \overline{G(\omega)} G(\omega - \theta) d\omega \end{aligned} \quad (2.2.14)$$

它们的傅里叶变换为

$$|G(\omega)|^2 = \frac{1}{2\pi} \int R(\tau) e^{-j\omega\tau} d\tau \quad (2.2.15)$$

和

$$|f(t)|^2 = \frac{1}{2\pi} \int R(\theta) e^{-j\theta t} d\theta \quad (2.2.16)$$

式(2.2.13)和(2.2.14)表示函数在两个不同时间时不同频率比较,或者使信号相关,所以通常也称为自相关函数。

§ 2.3 信号与噪声^[1]

化学定义的信号与噪声与无线电通信中定义的有所区别,简单地讲,信号是每位化学家期望得到的,除此以外,均被认为是噪声。例如,光谱扫描得到的谱图,紧靠 $\lambda = \lambda_0$ 波长处的信号可能是测量信号,而其他波长的峰和基线噪声均被认为是噪声。如图 2.3.1 所示。

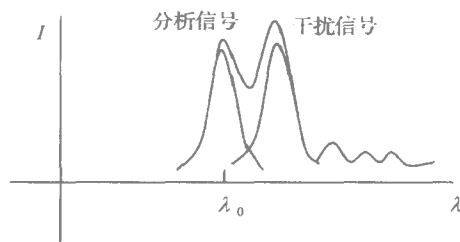


图 2.3.1 化学测量中的分析信号与噪声

数学上信号可表示为

$$S = f(t) \quad (2.3.1)$$

$f(t)$ 表示 S 是时间的函数,这在一般测量中是准确的,因为在大多数情况下我们测量得到的信号都是时间的函数。

$f(t)$ 可以是周期的或非周期的信号,我们在处理周期与非周期的信号时会采用不同的方法。化学测量中信号有

一些是非周期的(如色谱、光谱、光电子等),也有一些是周期的(如 NMR, FT 质谱 FT-红外光谱)。在无线电与通信中噪声可表示为

$$N = n(t) \quad (2.3.2)$$

$n(t)$ 也是时间的函数。幸运的是,噪声通常可以用周期信号的方法处理,因此噪声可以用傅里叶变换即频谱分析的方法进行分析。

信号强度 S 和噪声强度 N 的比值称为信噪比,以 S/N 表示。 S/N 是一项非常重要的指标,提高 S/N 比或信噪比增强技术在化学信号处理中是相当重要的。信号功率可表示为

$$W = V^2/R = IV \quad (2.3.3)$$

其中 V 是在检测器负载电抗 R 上得到的电压,而 I 是检出器的输出电流。信号的平均值可以表示为

$$\langle S \rangle = \bar{S} = \int S |f(t)|^2 dt \quad (2.3.4)$$

信号的标准方差,也称为信号的噪声水平 σ_s (σ_s 为标准偏差)

$$\sigma_s^2 = \langle S^2 \rangle - \langle S \rangle^2 \quad (2.3.5)$$

信号的相对标准偏差

$$R_{SD} = \frac{\sigma_s}{\bar{S}} = \frac{(\langle S^2 \rangle - \langle S \rangle^2)^{\frac{1}{2}}}{\langle S \rangle} \quad (2.3.6)$$

信噪比的另一种表示方法为

$$\frac{S}{N} = \frac{\bar{S}}{\sigma_s} = \frac{1}{R_{SD}} \quad (2.3.7)$$

信号的统计学表示方法与电子学表示方法有着本质的联系，在数学上等同的。在讨论信号与噪声的问题时，我们还注意到信号与噪声的统计学表示方法，信号的平均值为

$$\bar{x} = \left(\sum_{i=1}^n x_i \right) / n \quad (2.3.8)$$

用于量度平均值的离散度或信号的噪声水平是标准偏差 σ

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (2.3.9)$$

标准方差的另一种表示方法为

$$\sigma^2 = \frac{\sum_{i=1}^n x_i^2}{n - 1} - \frac{(\sum_{i=1}^n x_i)^2}{n - 1} \quad (2.3.10)$$

读者可以看到式(2.3.8)和式(2.3.9)与式(2.3.3)和式(2.3.4)的内在联系。

根据信噪比理论，噪声满足高斯分布（正态分布）（见图 2.3.2）。

在论述信噪比增强技术时，我们还将利用噪声的高斯分布来抑制噪声水平。

噪声水平的快速估计，为从测量的信号中迅速估算出噪声水平，可以采用下列公式（注： rms 表示均方差）：

$$N_{\text{rms}} = \sigma_N \approx 1/5(N_{\text{max}} - N_{\text{min}}) \quad (2.3.11)$$

上式的根据可用图 2.3.3 表示，在噪声

最可几分布 $\pm 2.5\sigma$ 处观察信号，可以覆盖 98% 的噪声出现概率，因此在较短时间内测量噪声，噪声值的最大值和最小值（注：噪声值 N 不等于噪声的水平 σ ）落在 $\pm 2.5\sigma$ 之间。

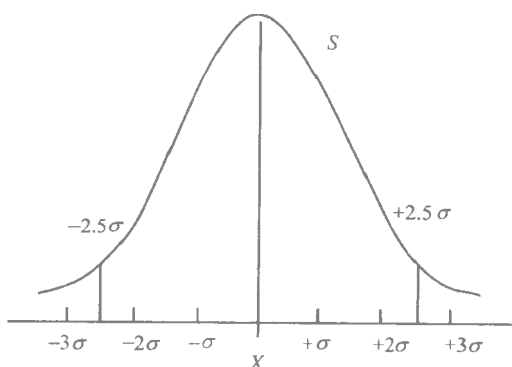
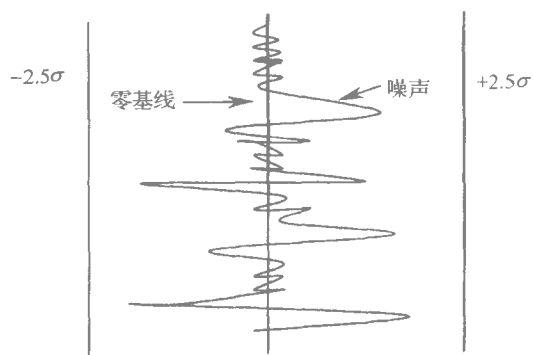


图 2.3.2 噪声的高斯分布（正态分布）

图 2.3.3 实际测量中噪声的 $\pm 2.5\sigma$ 起伏

噪声的频谱分析（有关频谱分布见以下章节）可以准确地告诉我们噪声的来源，一般来讲，噪声可分为三个部分：

$$\sigma^2 = \sigma_w^2 + \sigma_f^2 + \sigma_I^2 \quad (2.3.12)$$

其中 σ_w 称为白噪声， σ_f 称为 $1/f$ 噪声， σ_I 称为干扰频率噪声。平方和表示总功率等于各分功率之和。白噪声的频谱如图 2.3.4 所示。

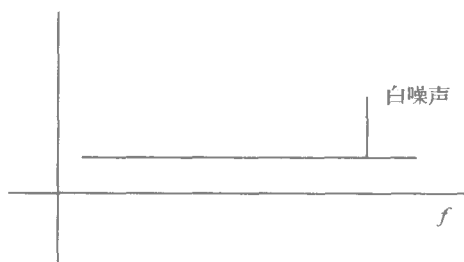


图 2.3.4 白噪声的频谱图

白噪声的频谱在频率轴方向上功率密度处处相等。在化学信号测量中，我们把接近 0 测量（直流）频率段处的噪声称为低频噪声，而把其他的称为高频噪声。在实际测量中，高频噪声反映了短时间功测量方法的噪声水平，可选用连续测量的重复性表示：

$$RSD|_{\text{短时}} = \frac{\sigma}{\bar{S}}|_{\text{长时}}$$

而测量中低频噪声反映了测量系统的漂移，即反映了长时间噪声水平，可用不连续测量的重现性表示：

$$RSD|_{\text{长时}} = \frac{\sigma}{\bar{S}}|_{\text{长时}}$$

测量系统的漂移也可以以漂移度表示（注：系统漂移测量时，化学信号为零值）

$$\begin{aligned} SHIFT|_T &= \langle |f'(t)|_T \rangle = \int_0^T |f'(t)| |f(t)|^2 dt \\ &\approx \frac{1}{T} \sum_0^T \left| \frac{\Delta S}{\Delta t} \right| \Delta t_i \end{aligned} \quad (2.3.13)$$

式中取变化率的绝对值 ($|f'|$ 或 $|\frac{\Delta S}{\Delta t}|$) 是要保证在每次 Δt 间隔中测量的 ΔS 不因符号相反而相互抵消。漂移度反映了系统在一定时间内的稳定程度。

白噪声来源于电子热运动、光子或离子的不规则运动、电子的自发运动等。噪声的第二个来源是 $1/f$ 噪声也称波动 (flicker) 和漂移 (drift) 噪声, 如图 2.3.5 所示。第三种噪声源称为工业干扰源, 其频谱如图 2.3.6 所示, 这种工业干扰中的有些干扰很讨厌, 如汽车点火器, 工业电 50Hz, 微波 2450Hz 发射, 高频加热 (10 ~ 100MHz), 工频 400Hz 等。

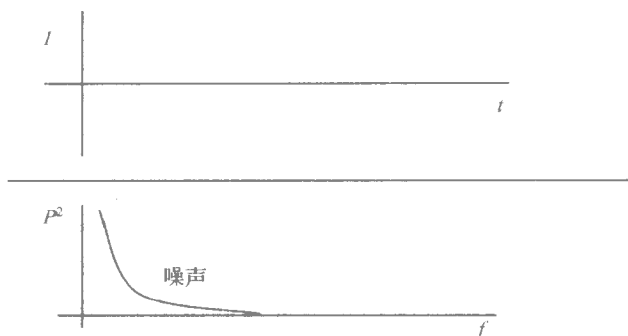


图 2.3.5 基线信号漂移噪声及频谱图

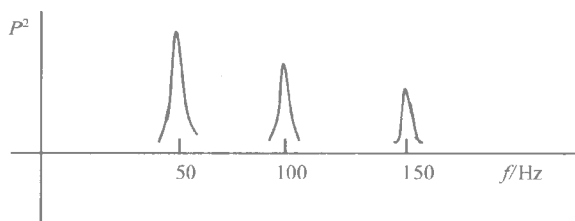


图 2.3.6 工业干扰噪声

在化学信号测量中, 经常还会测量背景值。背景值表示当不存在被测化学物质时, 系统所反映的空白信号平均值。空白信号可以包括化学物质的信号, 但不是被测的化学物质产生的, 这一类要特别注意:

$$S_{BG} = \langle S_B \rangle$$

在化学信号处理中, 还经常用信背比表示测量系统的空白水平

$$\frac{S}{B} = \frac{\langle S \rangle}{\langle S_B \rangle} \quad (2.3.14)$$

通常在化学信号测量中还用检出限表示测量系统的能力, 检出限与被测信号幅度有关, 其衡量值为噪声水平。在物理中经常采用 1 倍噪声水平 (此时置信度约为

67%), 在化学中采用 2 倍噪声水平(此时置信度约为 96%), 在分析化学中采用 3 倍噪声水平(此时置信度为 99%)。根据定义, 检出限应为相当于被测物质的空白的 3 倍标准偏差的信号所对应的被测物质浓度:

$$DOL = 3\sigma \frac{\Delta C}{\Delta S} \quad (2.3.15)$$

式中右边第二项的倒数定义为灵敏度。

$$Sens = \frac{\Delta S}{\Delta C} = \frac{S - S_B}{C - C_0} \quad (2.3.16)$$

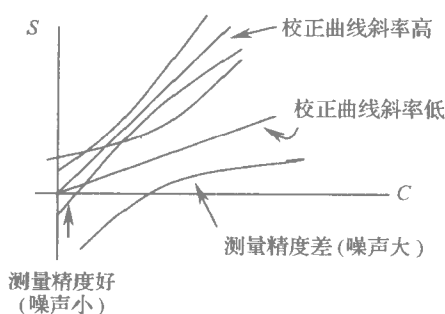


图 2.3.7 灵敏度、噪声与测量性能的关系

它表示在线性范围内, 单位被测量化学物质(浓度为 C , C_0 为空白浓度, 应 为零) 在检测器负载上引起的信号变化。因此, 灵敏度反映了测量系统的斜率。

不言而喻, 良好的检出限要求测量系统噪声水平低, 测量灵敏度高, 因此提高测量系统的灵敏度, 降低系统的噪声在化学信号测量和处理中是十分重要的这些关系可以由图 2.3.7 表

示。

§ 2.4 信号分辨^[2]

§ 2.4.1 理想情况

令 A 为真实测量信号, B 为真实背景信号, 理想情况时, 噪声水平在零上, 概率集中在零上, A 物质信号水平在真实值上, 概率为 1, 如图 2.4.1 所示。

§ 2.4.2 实际情况

噪声水平不在零上(存在背景, 即极低频噪声), 概率为高斯分布。信号水平在近似值上, 概率近似为高斯分布, 如图 2.4.2 所示。

§ 2.4.3 信号的检出

测量的信号 μ_A 和 μ_B 为 A 和 B 的近似值。 μ_A 和 μ_B 完全分辨, 如图 2.4.3 所示。

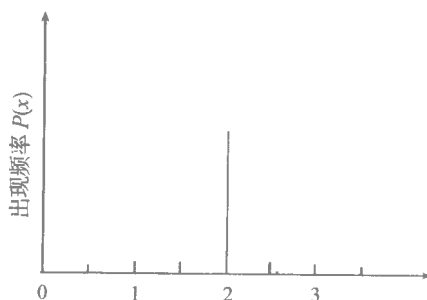


图 2.4.1 信号 x (任意单位)

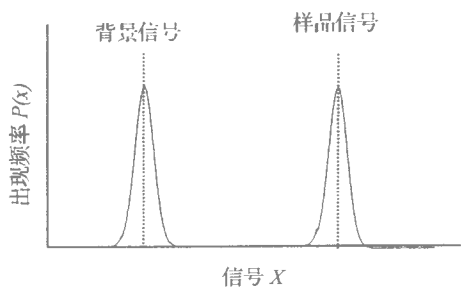


图 2.4.2 实际测量中的样品信号与背景信号

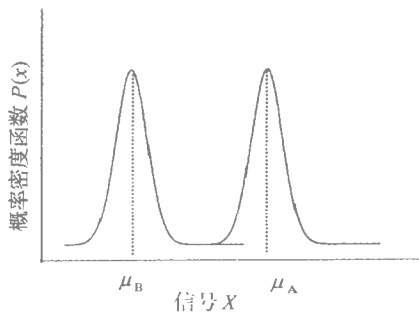


图 2.4.3 样品与背景信号的实际测量值

信号变小后，将向背景值靠拢，称为检出能力降低，如图 2.4.4 所示。当信号和噪声交替时，统计学变得重要起来，如图 2.4.5 所示。

检出信号

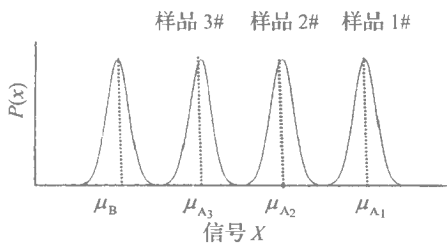


图 2.4.4 检出能力示意图

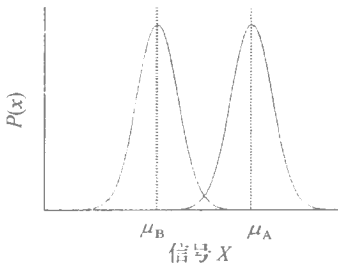


图 2.4.5 背景噪声与样品信号交盖示意图

§ 2.4.4 信号的可信度

根据误差理论，若 $\sigma_A = \sigma_B$ 时， μ_A 和 μ_B 相距 $\geq 6\sigma$ ， A 和 B 是统计学上独立的信号，如此， A 和 B 的测量互不干扰，如图 2.4.6 所示。

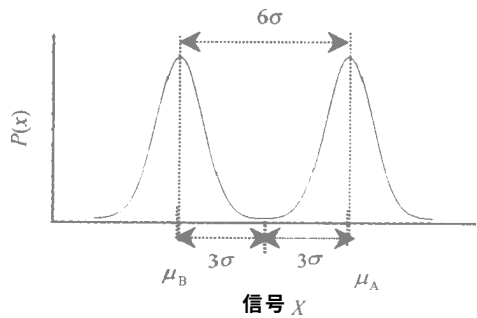


图 2.4.6 统计学上的相互独立的背景与样品的信号

若 B 为背景, 则 A 的检出是近乎绝对可信 ($\geq 99.8\%$), 或置信度为 99.8% .

当两峰交盖增加 A 和 B 统计学上互不独立, 因此必须用统计方法检查信号的可信度.

设前提 H_0 信号的检出是真实的前提 H_1 : 信号的检出不能被置信, 而前提 $\neq H_0, \neq H_1$, 表明信号的检出在统计学上无法判断. 当不灵敏区 a, b 进一步下降后, 背景错被当作信号的概率为 P_{10} .

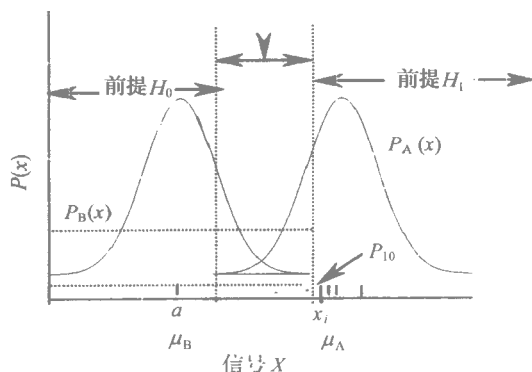


图 2.4.7 统计学上判断信号的归属

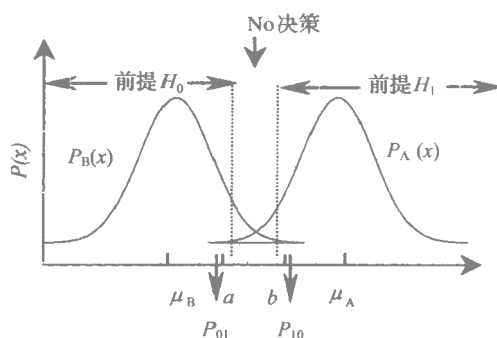


图 2.4.8 置信度较高的信号检出

图 2.4.7 表明了前提 H_1, H_0 和非 H_0 非 H_1 的情况. 图中 X 值可分为三个区域: H_0 真实区域, 无法判断区域以及 H_1 真实区域. 发现信号 A 为真实 [概率 $P_A(X)$] 的概率为

$$P_{11} = \int_{x_0}^0 P_A(X) dx \quad (2.4.1)$$

而发现背景 B 为真实 [概率 $P_B(X)$] 的概率为

$$P_{10} = \int_{x_i}^b P_B(X) dx \quad (2.4.2)$$

而错把信号当作背景的概率为

$$P_{01} = \int_a^{x_i} P_B(X) dx \quad (2.4.3)$$

事实上, 我们总希望 P_{10} 和 P_{01} 越小越好, 但实际上, 为了保证噪声不被误作信号, 我们要求 $P_{10} > P_{01}$ 或者至少 $P_{10} = P_{01}$. 如图 2.4.8 所示.

$$\text{此时 } P_{10} = P_{01} = \int_a^b P_A(X) dx = \int_a^b P_B(X) dx \quad (2.4.4)$$

式中 $b = \mu_B + 3\sigma_B$, $a = \mu_A - 3\sigma_A$

信号错被当作背景的概率 P_{01} , 在 ab 之间为不确定区信号检出的危险率下降。

没有不确定区的情况, 如图 2.4.9, 此时 $P_{11} = P_{01} = 0$. 但此时信号必须 6 倍于信号噪声水平, 显然, 也是一个过于苛刻的条件。

在信号趋近于背景时候, 信号的检出必须有统计学的定义, 如图 2.4.10, 是否 6σ 乃至到 3σ 都是正确的? 想法是既不要错当为信号, 也不要错漏掉真正的信号。

图 2.4.11 表示信号错过的概率较大的情况。由于只有 2σ 的分辨距离, 以此只有 16% 的信号分布 $\geq b = \mu_B + 3\sigma_B$

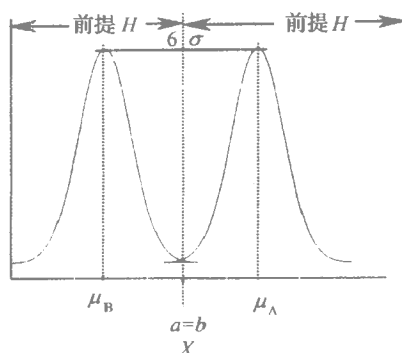


图 2.4.9 临界状态置信度的信号检出

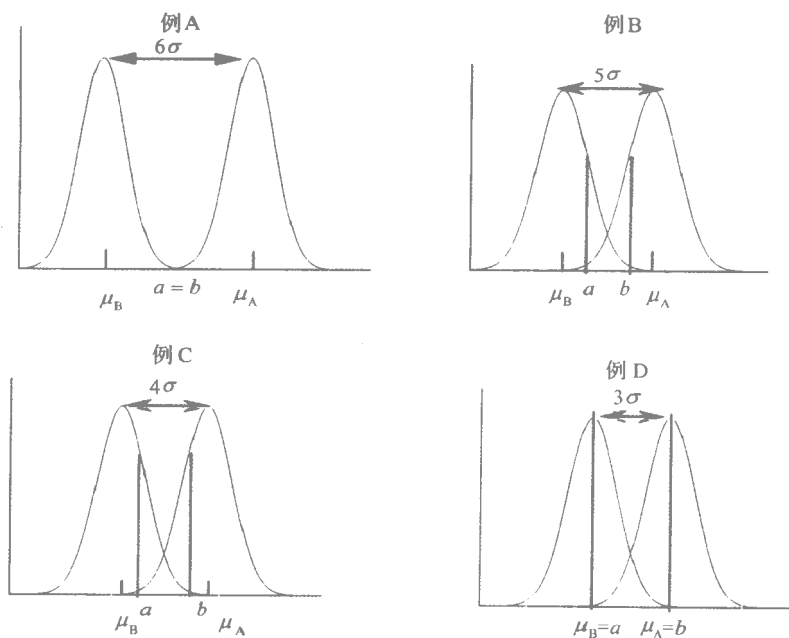


图 2.4.10 置信度的选择与信号的分辨

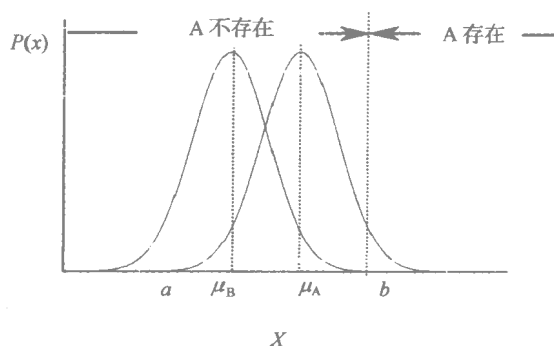


图 2.4.11 不适当的置信度选择

§ 2.5 信号的属性

§ 2.5.1 信号的时间属性

直观地讲,信号随时间的变化表明了信息的时间属性,或者讲,信号所包含的信息是时间的函数,如图 2.5.1 所示.图中 I 是测量仪器的响应, t 是时间,显然图

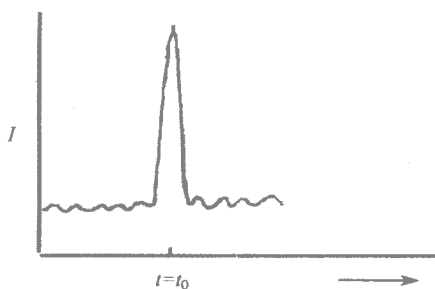


图 2.5.1 随时间变化的信号

示的信号在 $t = t_0$ 达到最大值.

$$I = f(t), I_{\max} = f(t = t_0) \quad (2.5.1)$$

§ 2.5.2 信号的频率属性

频率是与时间的倒数成正比,因为频率属性是依赖于时间的.但为何又研究其频率属性呢?这是因为有时用频率特性来研究信息的内涵来得更为清晰.频率属性可由图 2.5.2 表示.图中 P 是信号功率密度, ω 是频率,在 $\omega = \omega_0$ 时,功率密度有较大值.

$$P = F(\omega), P_{\max} = F(\omega = \omega_0) \quad (2.5.2)$$

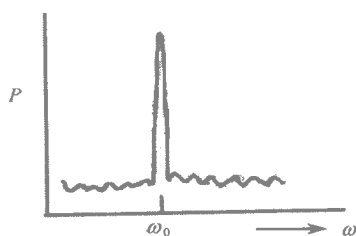


图 2.5.2 随频率变化的信号

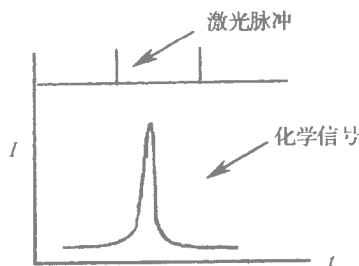


图 2.5.3 信号之间的相位关系

频率属性与时间属性的相互转变要运用傅里叶转换的技术。

与频率有关的另一附属性质是信息与参考信号之间的相位关系。比如用激光触发一化学反应，在过了半个周期之后观察到诱发的化学信号，如图 2.5.3 所示。

很明显，化学信号相对于脉冲触发有 Φ 的相位延迟。在频率表征中，通常用相位角来表示可能的相位关系：

$$I = I_0 \sin(\omega t + \Phi) \quad (2.5.3)$$

式中， ω 是角速度， Φ 为位相角。虽然上式以正弦为例表示了相位与角速度之间的联系以及相位角对信号强度的影响，但信号理论表明，任何信号都可以表征为简单正弦波的叠加，因此 Φ 相位角的普遍性是正确的。

§ 2.5.3 信号的波长属性(或能量属性)

当今信息理论已证明，由于粒子的波粒两重性，任何由粒子，包括从光子、电子到重 γ 射线是以粒子为主的信息波，可以由图 2.5.4 表示。

图中 E 为能量，如在光谱范围：

$$\Delta E = h\nu \quad (2.5.4)$$

h 为普朗克常量， ν 是粒子频率。根据测不准原理，凡是与能量有关的信息都会有自然宽度。另外观察到的变宽效应中自然宽度几乎微不足道，只有被观察点上粒子的碰撞，各种外场力作用等引起的变宽相对重要。这里不作详细介绍，读者可参考有关书籍。

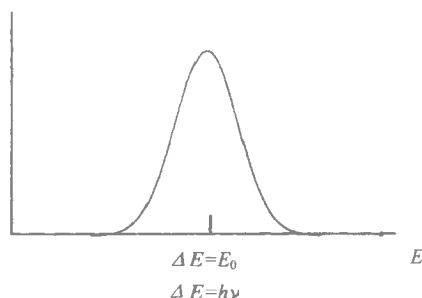


图 2.5.4 随能量变化的信号

§ 2.5.4 信号的空间属性

通常观察点和被观察点处于三维空