红外图像处理技术
之红外光谱信号处理(下)

航空航天学院
陆哲明/郑阳明

内容

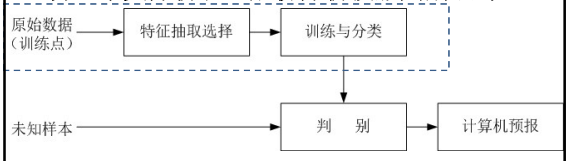
- 9.1模式识别判别
- 9.2主成份分析
- 9.3偏最小二乘法
- 9.4人工神经网络
- 9.5遗传算法
- 9.6小波变换

9.1 模式识别判别分析

- 所谓模式是事物或事件的各种可测量和不可测量的集合。人们对客观世界的认识就是对模式的识别。
- 一个物体（的性质）是由其结构、状态、组成和所处的环境等因素决定的，这些因素称为特征，人们对事物的认识就是依靠对这些特征的识别。
- 模式识别方法主要分成有管理识别和无管理识别两种方法。

- 有管理识别方法也称有监督识别方法或称有教师识别方法；无管理识别方法也称无监督识别方法或称无教师识别方法。
- 有管理识别方法要求有一训练集或学习集（类似与校正理论中的校正集），在训练中，各样本类别是已知的，如已知一些样本属于A类，而另一些样本属于B类等。将这些训练集样本的性质及类别输入计算机，让计算机通过训练或学习掌握一定的识别规律后，再去识别未知样本。

- 有管理识别方法有：因子分析、线性判别分析、K最近邻法、SIMCA方法、神经网络法等。
- 无管理识别方法不需要训练集，如最小生成树、聚类分析和主成份分析投影判别法等。

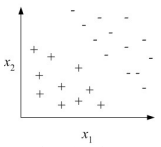


模式识别计算过程示意图

- 模式空间的几种距离与相似性度量

✓模式空间

□一个参量就构成了模式识别空间的一个点，一组参量即是一个*n*维的模式，这样的模式所构成的*n*维空间，称为模式空间。一般来说，因高维模式空间提供了更多的信息，故有可能解决一些低维空间中难以解决的问题。这种情况可容易通过下图说明。



在二维空间可能区分一维空间不可分的二类样本

✓模式空间的距离与相似性度量

- 模式识别的基本假设是相似的样本在模式空间处于相近的位置，如何度量两个样本在模式空间的远近的问题。可应用几何中距离的概念。在模式识别中是作为与“相似性”相反的“相异性”的度量使用的，任意满足下述三个条件的“相异性”均可以定义模式空间的距离:

(1) $D_{ij}\geq 0$

(2) $D_{ij}=D_{ji}$

(3) $D_{ik}+D_{jk}\geq D_{ij}$

- 条件1和2表明距离应取正值且是对称的，条件3是作为“度量”距离与“非度量”距离的差异。在欧氏空间中。两点间的直线距离恒小于或等于其他径的长度。

✓ 测量数据预处理和特征选取

- 模式识别中，所有分类的识别过程都基于样本集数据矩阵$X_{n\times p}$，数据集x的质量直接影响计算结果，决定判别的成败。“质量”一方面是指数据集的质量，另一方面是指数据本身的质量。
- 影响数据本身质量的原因主要是在采集数据过程中，人为地扩大了或缩小了数据的变化域（比如均值、方差等）。数据集的质量是由描述样本的各个特征或变量的类型差别是否较大所决定的。
- 另外，在数据集中，有些变量可能为谱图数据，有些则为化合物的物理化学性质或结构方面的的信息，数据类型不仅量纲不一样，其绝对值的大小有时也可能会有几个数量级的差别。

- 对测量数据预处理是必要的，数据标准化（autoscaling），也称均值方差化或归一化，是常用的预处理方法，该方法将数据阵中各元素减去该列元素的均值后再除以该列元素的标准偏差，变换公式为：

$$z_{ij}=\frac{x_{ij}-\bar{x}_j}{s_j}$$

- 式中，x_{ij}为样本i的变量j的值；$\bar{x}_j$为变量j（即第j列）的平均值；s_j为变量j的标准偏差。经过标准化预处理的变量（第一列元素）权重相同，均值为0，方差或标准偏差均为1。

- 在某些情况下，也可对数据标准化变化公式的分母作如下的变换：

$$z_{ij}=\frac{x_{ij}-\bar{x}_j}{x_{j.\max}-x_{j.\min}}$$

- 式中，$x_{j.max}$，$x_{j.min}$是数据矩阵x中第j列的最大值和最小值。
- 数据正规化变换（normalization）也是常用的数据处理方法，这种方法可将量纲不一、范围不同的各种变量或特征数据表达为0~1范围内的数据，其变换公式如下：

$$z_{ij}=\frac{x_{ij}-x_{j.\min}}{x_{j.\max}-x_{j.\min}}$$

- 该方法不仅适用于同类型范围大小的原始数据阵，也适用于不同数据类型和范围大小差别较大的数据阵的预处理，有时称之为值域调整。

- 对数变换也可以用来处理数据，当$x_{ij}>0$时，对每个元素进行对数运算，可以是常用对数或自然对数也可以是其他正整数为底的对数：$z_{ij}=\lg(x_{ij})$ 当变量的动态范围较大，比如相差数个数量级时，采用这种对数变换法是恰当的。
- 特征变量数的压缩在特征选择过程中是非常重要的一步。通过压缩，应使特征变量数降低到最小。因为当有些不必要的特征变量引入模式中时。不仅与分类无关，而且往往导致分类结果变差，在特征选择过程中，一些方法属于统计方法，如数据偏差计算、因子分析等，以所得结果为依据来确定与选择比较重要的特征。
- 另一些方法是考察特征变量对分类结果的影响，影响大者选之，否则弃之。

- 特征变量的选择在模式识别中尽管研究较为广泛，但至今尚无通用的理论可以遵循。应用较多的有偏差权重法，即不同样本取值差异较大的特征变量，对样本分类应当贡献较大。反之，如果某一变量的偏差较小，这种变量对于分类而言显然意义不大。偏差通常以特征变量j的标准偏差s_j或方差v_j来描述：

$$v_j=s_j^2=\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2/(n-1)$$

- 在选择变量时，将v_j小的特征变量先行弃去。对于训练集包含已知分类的样本，可采用一种简单而有效的特征抽取方法，它可以很容易辨别出特征向量对分类贡献的大小，这种评估特征变量的标准称为Fisher比率法：

$$F_j = \frac{(\bar{x}_{j.1} - \bar{x}_{j.2})^2}{s_{j.1} + s_{j.2}}$$

- 式中，$\bar{x}_{j.1}$和$\bar{x}_{j.2}$分别为类1和类2中变量j的均值；$s_{j.1}$和$s_{j.2}$分别为类1和类2中变量j的标准偏差（也可以采用方差）。F_j值越大，表示该变量j对于该两类差别的影响越大，应优先选用。

✓ 聚类分析

- 聚类分析（clustering）是一种无管理模式识别方法，常用于目标观测对象的分类，即利用表征观测对象的一组变量对目标进行分类。
- 有两种方法可以用来表示观测目标的聚类。一种称为树状图，它可将一个大组分类成较小的组，直至每个子集仅含一个观察目标。另一种方法是采用表格形式来表示观测对象的聚类。
- 聚类分析不需要训练集，是无管理模式识别方法的典型代表，有很大的实用价值，特别适用于样品归属不清楚的情况。聚类分析所讨论的对象是n个样本组成的样本测量数据集合或样本数据矩阵。

- 假如样本集合是不连续的，则这个样本集合可以看作是包含未知的若干个性质不同的子集，即不同的类。聚类分析的目的就是寻找这些子集，其基本思想是在多维模式空间中，任何一个子集内部样本之间的相似性即同类内相似性大于不同子集样本之间的相似性，即类与类之间的相似性。
- 要实现“物以类聚”，首先必须选定相似性的度量标准。前面介绍的各种距离常作为聚类分析中样本之间相似度的度量，距离越小表示相似性越大，越容易聚在一起形成一类。选择不同的相似性度量对样本矩阵进行聚类分析的结果可能不同，甚至相差很大。

✓聚类分析算法基本原理

- 只要事先给定了样本与样本之间或样本与类之间或类与类之间的相似性度量即可进行聚类分析。其基本思想是将待聚类的样本集的n个样本各自看成一类，然后规定或定义样本之间的距离或相似度度量以及类与类之间的距离后开始进行聚类。聚类开始时，因每个样本各自形成一类，类与类之间的距离和样本与样本之间的距离是相同的，选择距离最小的一对即相似性最大的一对样本合并成一个新类；
- 进而计算该新类和其他所有类间的距离。比较各个距离之后，将距离最小的两类合并成另一新类；再计算类间或样本与样本或类与样本间的距离，按距离大小合并成新类。

- 如此下去，直到所有的样本归为一类为止。整个聚类过程，进行了n-1步合并新类的操作，并得到n-1个并类距离。这n-1个合并过程也可以称为树图的聚类图(dendrogram)来表示。

✓类间距离的定义与系统聚类方法

- 定义类与类之间距离的方法有很多种，相应的系统聚类分析法也有很多种，常用的有最短距离关联法、最长距离关联法、中间距离关联法、重心法、类平均关联法及方差平方和关联法等。总的来说，选择平均距离关联法是恰当的。在使用平均距离关联法的过程中，假如需要关联两个大小不同的类，有时可引入权重系数对关联距离进行处理。
- 另一种可以给出较好的结果的方法叫Ward方法，它采用不均匀的判断规则，其思想来源与方差分析，即如果类分得比较正确，同类样本的方差平方和 $s_i^2 = \sum_{i=1}^{m}(x_{ij} - \bar{x}_j)^2$ 应当较小，类与类之间的方差平方和应当较大。

- 由于在给定样本数和给定分类数后，可能的分类方案仍然是一个非常大的数，目前还没有很好的方法以较少的计算求得精确的最优解。即在一切可能的方案中求得同类内方差平方和s^2的极小值，一般只能找到一个局部最优解，寻找局部最优解的思想是：
  - 先将n个样本自成一类
  - 然后计算每两两合并为一类时的方差平方和，选择方差平方和s^2增加最小的两类合并成新类
  - 又计算新类中两两组合的方差平方和；并选择合并，直至所有样本归为一类。

- 从上述讨论中看，似乎方差平方和法与前面介绍的数种系统聚类法有较大的差异，其实可以证明，由样本或类i与新类r（新类r由两个样本或类p和q合并而成）关联的方差平方和法的类距离递推计算公式为：

$$D_{ir}^2=\frac{n_i+n_p}{n_i+n_r}D_{ip}^2+\frac{n_i+n_q}{n_i+n_r}D_{iq}^2+\frac{n_i}{n_i+n_r}D_{pq}^2$$

- 式中，n_i为类i中的样本数，而n_p和n_q分别为类p和类q中的样本数（$n_r=n_p+n_q$）。
- 最近距离关联法有将样本组成条形聚集的倾向，甚至某些性质非常不同的样本最后聚集在一条狭长的条形范围内，具有空间收缩的特性。最长距离关联法有将样本聚集在较小且过于紧凑的范围内的倾向。
- 总的来说，平均距离关联法和Ward法是最有效也是最常用的方法，一般能给出更好的结果。

✓最小生成树法

- 最短距离法在数学计算方面较为简单，倘若样本数较少，还可以进一步采用一种称为最小生成树（minimal spanning tree）的方法对样本进行分类。尽管最小生成树法与最短距离法步骤不尽相同，但可得相同的聚类结果。
- 以表1和2中的数据为例对最小生成树法进行说明。

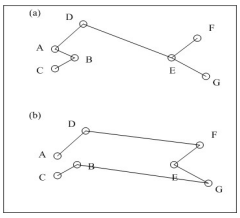
表1 7个样本的特征x_1和x_2

样本	x_1	x_2
A	45	24
B	24	43
C	14	23
D	64	52
E	36	121
F	56	140
G	20	148

表2 表1中各样本间的相似性数据

	A	B	C	D	E	F	G
A	0						
B	28	0					
C	32	23	0				
D	35	40	60	0			
E	100	80	103	76	0		
F	119	104	128	90	29	0	
G	127	105	126	105	30	35	0

- 将样本A,B,C…想像成7个小镇，现要用高速公路将它们连接起来，显然要求总长度最短是合理的。



以最小生成树法处理表1和2的数据结果表明，显然最小最小生成树法的分类结果(a)好于一般方法的结果(b)

- 最小生成树法的算法很多，Kruskal算法是常用的一种，该算法的实质在于将距离最短的两个样本连接但不能组成三角形。显然表2中最短距离为23，即可先将BC连接起来，接下去的最短距离为28（AB)。然后可以发现29和30(EF和EG)为最短的距离。32(AC)是第5个最短的距离，但如果将AC连接起来便组成了三角形ABC，故只能取消AC，最后可先后连接AD和DE(35和76)，由此所得的最小生成树见图中(a)。如果断开此图中最长的距离DE，便可得到两类（ABCD和EFG）。假如要作进一步研究，可以切断第二长的连线AD，便可得到第三类（ABC，EFG和D）。这样的过程可以继续下去，甚至获得所希望的分类结果。

✓ 基于主成分分析的投影判别法

- 在一些模式识别中，含有p个特征的一个样本，是p维空间中的一个点，若p等于2或3，则可直接用图形显示这些模式向量。
- 图形显示具有直观性，现代计算机提供了强有力的屏幕图形显示功能，使人眼很容易地识别图形，在二维或三维空间中显示模式分布情况，人眼便能识别样本的分类情况，也就是说人眼是强有力的模式识别器。
- 然而，在一些模式识别中，绝大多数测量数据矩阵的每一个样本都是由多个变量来描述的，特征空间都是高维空间。

- 在这种p大于3的高维空间中样本点的分布比聚类情况呈什么图形是抽象的，远超过了人类视觉的判别能力，因此必须借助于降维技术将p维空间的图形在二维或三维空间中显示出来，并尽可能减少p维空间中的分类信息。
- 基于主成份分析的投影判别法采用多元统计分析中的主成分分析法，先直接对样本测量数据矩阵x进行分解，只取其中的主成分（得分向量）来投影，然后进行判别分析，故有主成分分析的投影判别法之称。
- 主成分分析所得的主成分轴是该数据矩阵的最大方差方向，且这些主成分轴相互正交，这样就可保证从高维向低维空间投影时尽量多的保留有效信息。

- 如果对样本测量矩阵x的构成作如下规定。可以很清楚地看出主成份分析的投影性质。样本数据阵x可表示为：

$$X=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}=\begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}=\left[x_1 x_2 ... x_p\right]$$

- 式中，n为样本数；p为样本变量数；$x_j \quad j=1,2,...,p$为样本向量；$x_i^T \quad i=1,2,...,n$为变量向量。有奇异值分解$x=USV^T$，即$XV=US$：

$$XV=\begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}\left[v_1 v_2 ... v_d\right]=\begin{bmatrix} x_1^T v_1 & x_1^T v_2 & \cdots & x_1^T v_d \\ x_2^T v_1 & x_2^T v_2 & \cdots & x_2^T v_d \\ \vdots & \vdots & \ddots & \vdots \\ x_n^T v_1 & x_n^T v_2 & \cdots & x_n^T v_d \end{bmatrix}=US$$

- 有$T=US$，T即为主成分分析中的得分矩阵，其中每个元素实际上即是每一个样本向量：

$$x_i^T \quad i=1,2,...,n$$

在载荷V中的各个相互正交的载荷向量上的投影坐标（内积本质上就是投影），它反映了样本与样本之间的相互空间关系。

- 因此，二维得分投影图（PC_1~ PC_2）或三维得分投影图（PC_1~ PC_2 ~ PC_3）可直接由主成份分析程序计算得到的前两个得分向量（t_1和t_2）或前三个得分向量（t_1、t_2和t_3）中的相应元素在对应坐标中相互作图而得。

- 复原原始矩阵需要准确的主成分数，而在此所述的投影判别法中仅取其中2个或3个主成分作图，倘若实际主成分数大于2或大于3时，必然会产生较大误差。在低维空间中得到的分类结果的可信度受到一定影响，因此一般应计算前2个或前3个主成分时可信度的大小。前2个和前3个主成分的可信度可分别用下式表示：

$$R=\frac{\lambda_1+\lambda_2}{\sum_{j=1}^{p}\lambda_j}\times 100 \qquad R=\frac{\lambda_1+\lambda_2+\lambda_3}{\sum_{j=1}^{p}\lambda_j}\times 100$$

- R值越大，则前面两个或三个主成分对原始数据中的聚类信息贡献越大，代表性越强。为保证一定的准确度，R值应大于80~85%。

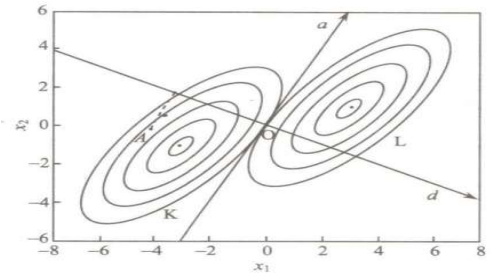
- 有管理模式识别

✓ 有管理模式识别方法总体思路是用一组已知类型的样本作为训练集，即用已知的样本进行训练，让计算机向这些已知样本学习，这种求取分类器的模式识别方法称为“有管理的学习”，这里的训练集便是管理者，并由这个训练集得到判别模式。

✓ 常用的方法有线性判别分析（linear discriminant analysis）或称线性学习机（linear learning machine），K-最近邻法(k-nearest neighbour method)及SIMCA(多元数据分析软件)等。

✓ 线性判别分析

- 线性判别分析（linear discriminant analysis, LDA）是应用较广的模式识别方法，假设在二维空间(x_1, x_2)中有两个类K和L，见下图。
- 图中绘出的两个椭圆形分别表示正态双参数样本概率分布范围，称为双参数概率分布椭圆(bi-variates probability ellipses)，即表示样本从属的范围。两组椭圆数据相等，各自对应的椭圆表示相同的概率水平。最外的两个椭圆外相切，直线*a*为公切线。直线*a*便可作区分K和L的分界直线，即*a*左边的样本属于K类而右边的样本数据L类。



两类样本K和L的线性判别
A为属于K类的样本

- 对于多维空间（例如*W*维）中的判别分析，每个样本被看作*W*维空间中的一个点，要在两类样本点群之间寻找一个最优的分割面，使得两类样本点在该分割面上的投影达到最大限度的分离，并且使每类样品内部的离散度最小。
- 在实际中，一般采用代数方法来定义区分上述两种样本的界线。定义垂直于*a*的直线*d*，将各个样本点在该直线上投影，A点在直线*d*上的投影可由它的得分*D*来表示：

$$D = \omega_0 + \omega_1 x_1 + \omega_2 x_2$$

- 假设数据已作标准化则$\omega_0 = 0$，在点0处*D*为0。对于属于L类的样本有*D*>0，而属于K类的样本则*D*<0。

- 从上式来看*D*为一个具有隐含性并类似于主成分分析中的变量，因此可以将LDA看作为主成分分析的一种特殊情况，即将高维空间中的变量降维为低维空间中的变量，然后，将样本点在低维空间中投影。
- LDA与主成分分析不同的是：主成分分析选择数据点变异最大的方向作为其主成分方向以达到降维的目的，而LDA往往选择能够最大程度分开所给出各类的分割面作为它的方向，即找到一个函数，该函数的变量（隐含）为原始变量的线性组合，称之为正规变量（canonical variable）。在存在*k*个分类的情况下，需要确定*k*-1个正规变量以建立*k*-1个分割面对样本进行分类。

- 在多变量的情况下，找到一个多变量线性组合的函数*D*同样是可能的，但要求各个样本组之间存在足够大的差异，同时每个样本组中的样本均比较紧凑和密集。用统计数学的术语来说，即组间方差大于组内方差。

✓ K-最近邻法

- K-最近邻法(k-nearest neighbour method, KNN)是一种直接以模式识别的基本假设——同类样本在模式空间相互较靠近为依据的分类方法，它计算在最近邻域中*k*个已知样本到未知待判样本的距离，即使在所研究的体系线性不可分的情况下，此法仍可适用。KNN算法上来讲较为直观，KNN实际要将训练集的全体样本数据存储在计算机内，对每个待判别的未知样本，逐一计算与训练集样本之间的距离（一般取欧式距离），找出其中最近的*k*个进行判别。
- 如果*k*=1，显然这一个最近邻样本属于哪一类，未知待判样本就属于哪一类；

- 如果k>1，则这个最近邻样本不一定都属于同一类。应采用“表决”的办法，如果在这个最小距离区域中，包含第1类训练集的样本数较多，距离又小，则可判待分类样本为第1类；若包含第2类训练集的样本数较多，距离小，则可判待分类样本为第2类；以此类推。
- 在仅考虑两类判别问题时，可按下式计算判别函数S：

$$S = \sum_{i=1}^{k} S_i / D_i$$

- 式中，S_i为训练集中第i个样本的取值，当i属于第1类时取“+1”，当i属于第2类时取“-1”。D_i是待判样本对第i个样本的距离，可理解为权因子，即距离小的训练集样本给予较大的权，而距离大的训练集样本给予较小的权。显然，在相同距离的情况下，D_i值越大即距离越远，相似性越小，总S值的绝对值越大；在样本数相同的情况下，D_i值越大即距离越远、相似性越小，对总S的贡献越小。

- KNN优点是：不要求模式集合是线性可分的，也不需要单独的训练过程，新的模式或样本直接加入到模式集合中也很容易，而且还能处理多类问题，因此应用较为方便。
- 但该法不能压缩变量的数据，比较适用于低维的数据集，此外，选用的k值不同时，待判样本的分类结果可能不同，且k值的选择尚无规律可循，只能由具体情况或由经验而确定。一般k值不宜过小。

✓SIMCA分类法

- SIMCA(soft independent modeling of class analogy) 分类法又称相似分析法，它除了有分类和判别等功能外，还可以用于建立数学模型。
- SIMCA分类法是建立在主成分分析基础上的一种模式识别方法，其基本思路是先利用主成分分析的显示结果得到一个样本分类基本印象，然后分别对各类样本建立相应的模型，继而应用这些模型来对未知样本进行判别分析，以确定其属于哪一类，或不属于哪一类。
- 由于整个SIMCA计算过程可在投影图上直接进行，所以只要计算机程序的人机对话界面建立的较好，此方法可在投影图上直接操作完成，使用起来非常方便。

9.2 主成分分析

- 主成分分析（Principal component analysis, PCA）是对多变量数据进行统计处理的一种数据线形投影方法，它在尽可能保留原有信息的基础上将高维空间中的样本映射到较低维的主成分空间中。其基本思想是以一种最优化方法浓缩量测数据信息，使数据矩阵简化，降低维数，寻找少数几个由原始变量线形组合的组成分，以揭示数据结构特征，提取基本信息。

- 主成分回归

✓主成分回归（principal component regression, PCR）在校正模型的建立中也得到较多的应用，其优点是不需要专门寻找建模的波长，而是利用全谱数据通过因子分析来建立模型。

✓它是把原始数据进行线性组合来建立新的变量。在这种方法中，产生的变量（主成分）可用一个二维系统描述，每一个主成分的总方差是它的特征值。

✓PCA方法的优越之处在于所有主成分是相互垂直的，这样在不减少光谱信息的情况下就可消除共线性，得到一个更好的模型。处理原始光谱得到每一个主成分的量叫得分，它对应于主成分轴上每一样品的位置。

- PCR是将样品的得分对参考含量进行回归来建立定量模型。PCR不像多元线性回归那样要花很长时间进行波长选择，而且会降低噪声影响。因为噪声是分布在所有主成分中的，而PCR只用前几个主成分。
- PCR能够有效地解决多元线性回归中遇到的共线问题、变量数使用限制问题和在一定程度上解决了噪声滤除问题。

✓ 主成分分析主要应用领域

（1）降维（或称数据压缩），寻找几个主成分（也称潜变量）在低维空间表示高维数据。

（2）数据的可视化和分类聚集，主成分的投影显示法即可用于分类判别又可用于聚类，可以从投影图中看出样本与样本之间的关系，变量和变量之间大的关系；

（3）降低随机误差，主成分分析的过程是寻找少数几个相互正交、方差最大的新变量，来重新构造数据，能够有效地去除抽取误差；

（4）确定组分数，从数学意义上主成分分析的实质是特征值问题，主成分分析所得的非零特征值的个数就是矩阵的秩，就是构成数据的化学组分数，确定了矩阵的秩就可以确定体系的组分数。

✓ PCR的优点

（1）可以使用整体量测数据（原始全谱或部分数据），能充分利用数据信息，使用更多的数据则能利用数据的平均效应，增强模型的抗干扰能力。

（2）解决了共线问题。

（3）适用于复杂分析体系，不需要知道干扰组分的存在就可以预测被测组分。

✓ PCR的缺点

（1）模型优化需要PCA,对模型的理解不如多元线性回归直观；

（2）并不能保证参与回归的主成分一定与被测组分或性质相关。

✓ 奇异值分解

- 现有数据测量矩阵$X_{n\times p}$，n表示测量目标数目，p表示测量通道数。将奇异值分解（singular value decomposition，SVD）这种矩阵代数算法应用于数据阵X的分解可得：

$$X = U\Lambda V^T or \Lambda = U^T XV$$

- 式中，U为$n\times r$阶行正交矩阵；V为$p\times r$阶列正交矩阵；而 Λ 为$r\times r$阶对角阵；r为维数，它的最大值可为n或p中的较小者。对角阵 Λ 中的对角元素均为正值，而非对角元素均为零。SVD分解还可用矩阵U和V的列向量来表示：

$$X = \lambda_1 u_1 v_1^T + \lambda_2 u_2 v_2^T + ... + \lambda_r u_r v_r^T = \sum_{i=1}^{r} \lambda_i u_i v_i^T$$

由于矩阵U和V中列向量的正交性，有

$$U^T U = V^T V = I_r$$

式中，I_r为$r\times r$阶单位阵，该式表示矩阵U和V各列元素平方和为1，而列向量间的内积为0，向量的这种性质称为正交性。

在SVD分解中，U称为行奇异向量矩阵，V称为列奇异向量矩阵，而Λ称为奇异值组成的对角矩阵。

✓ 特征值和特征向量

- 设测量矩阵X的行数n大于列数p，则有$r \le p < n$，这里r被称为矩阵X的秩，它表示测量数据矩阵X的无关变量数目。有：

$$C_n = XX^T = U\Lambda V^T V\Lambda U^T = U\Lambda^2 U^T$$
$$C_p = X^T X = V\Lambda U^T U\Lambda V^T = V\Lambda^2 V^T$$

- 式中，C_p为$p\times p$阶数据矩阵X列乘积的对称方阵，C_n为$n\times n$阶数据阵X行乘积的对称方阵，同时将矩阵 Λ^2主对角线上的元素平方和Λ后即可得矩阵 。
- 上式所表示的矩阵分解方法常称为特征值分解(eigenvalue decomposition, EVD)。特征值分解EVD还可由下式表示：

$$U^T C_n U = V^T C_p V = \Lambda^2$$

- 显然通过正交矩阵U，对称方阵C_n可被对角化而转换成Λ^2。同样通过正交矩阵V，对称方阵C_p也可被对角化为Λ^2。
- 假如u_1和v_1为最大特征值λ_1^2所对应的特征向量，有：

$$u_1^T C_n u_1 = v_1^T C_p v_1 = \lambda_1^2$$

- 将上式中的行向量写成列向量形式，并进行重排可得：

$$(C_n - \lambda_1^2 I_n)u_1 = O_n \quad or \quad (C_p - \lambda_1^2 I_p)v_1 = O_p$$

- 式中，O_n和O_p为元素值为零的向量。对于所有的正交向量$u_1,u_2,\ldots$和$v_1,v_2,\ldots$，可以将上式推广写作：

$$(C_n - \lambda_k^2 I_n)u_k = O_k \quad or \quad (C_p - \lambda_k^2 I_p)v_k = O_k \quad k = 1,2,...,r$$

- 式中，I_n为$n\times n$阶单位阵；I_p为$p\times p$阶单位阵。为了求解上式中的向量u_k和v_k，要满足其系数矩阵、为0，有特征方程：

$$|C_n-\lambda_k^2 I_n|=0 \quad or \quad |C_p-\lambda_k^2 I_p|=0 \quad k=1,2,...,r$$

- 上式为一个r次方程，可解得r个正实根 λ_k^2，式中$r\le p$，且假设$p\le n$。特征值分解EVD方法中，C_n、C_p和 Λ^2的迹均等于矩阵X中各元素的平方和C有：

$$\sum_{k=1}^{r}\lambda_k^2=tr(\Lambda^2)=tr(C_n)=tr(C_p)=\sum_{i=1}^{n}\sum_{j=1}^{p}x_{ij}^2=C$$

- 式中迹表示一个方阵中主对角线元素之和。从上式中可知，各个特征向量u_k或v_k以 λ_k^2 的形式对数据阵X各元素平方和作出贡献，因此特征值可按

其对应的 λ_k^2 的大小进行排序。一般来说矩阵U和V中各个列向量均按其贡献的大小由大到小排列。注意到主成分分析常被写作：

$$X=TP^T$$

- 若用奇异值分解（SVD）的结果来表示，有 $X=U\Lambda V^T$

即有：

$$T=U\Lambda, P^T=V^T$$

- 式中，矩阵T，称之为得分矩阵；而矩阵P，称之为载荷矩阵。

✓ NIPALS算法

- 利用非线性迭代最小二乘法(nonlinear iterative partial least squares, NIPALS)来对测量矩阵X进行主成分分解，步骤：

（1）取X中任意一列作为初始的向量t；

（2）由此t及矩阵X计算v': $v'=t^TX/t^Tt$

（3）将v'归一化：$v_{new}^f=v_{old}^f/\left\|v_{old}^f\right\|$

（4）计算新的t:$t=Xv/v^Tv$

（5）比较步骤4中所得t和初始t，若两者相等（在给定的误差范围内）则按t^Tt计算特征值，并转第6步继续进行；如不相等，将t返回第2步继续迭代计算。

（6）从X矩阵中减去tv^T: $X=X-tv^T$，返回步骤1继续进行计算，直到X趋向于0。

- 步骤5计算得到的特征值的大小与原始数据阵X是否进行预处理（中心化和归一化）有关。即对测量数据做不做预处理以及进行怎么样的预处理对得分矩阵T和特征值的大小有一定影响，但不影响载荷矩阵V，也不影响数据X的分解结果。

✓ 主成分数（主因子数）的判别

- 按前述的SVD和NIPALS都可将数据矩阵X进行分解求得p个特征值λ，由于在迭代算法中各特征值是按由大到小的次序排列的，因此可以写作：

$$\sum_{k=1}^{p}\lambda_k=\sum_{k=1}^{r}\lambda_k+\sum_{k=r+1}^{p}\lambda_k$$

- 式中右边第一项是主要因子（主要成分）对特征值的贡献，后面一项是误差及其他微量成分等因素对特征值的贡献。前面已指出，λ的大小与原始数据阵所进行的预处理有关，因而一般不能直接依据特征值的大小来判别主因子数或组分数，而应该根据其相对大小及其变化规律来加以判别。

- 常用于判断组分数的判别指标有比率法（RATIO）和RSD函数法。比率法从组分数等于1开始计算r取不同值时的RATIO值，当该分组百分数大于某确定值（根据情况可取85~95%）时的r值即为主成分数

$$RATIO=\frac{\sum_{k=1}^{r}\lambda_k}{\sum_{k=1}^{p}\lambda_p}\times 100\%$$

- 注意上式中仅表示列数p小于行数n的情况，否则可将加和号中的p以n取代。可以计算特征值的残差（RE）值：

$$RE=\sqrt{\sum_{k=r+1}^{p}\lambda_k/n(p-r)}$$

- RE表示表示提取r个较大的特征值之后，剩余的p-r个因子所含的均方差。

- 显然当RE值大于测量方法的误差时，说明残差RE中尚包含其他因子。当RE的大小与方法误差大小相当时，对应的*r*值即为主因子数。
- 因此，若测量方法的误差大小可以估计，则可用RE的计算值与测量值进行比较，当两者接近时的*r*值即为主因子数。但在实际的测量中，大多数情况下方法误差为未知，这时就难于直接用RE做依据来确定主因子数。
- 还可以通过计算特征值的嵌入误差IE来对主因子数进行估计。嵌入误差IE是指取*r*个主因子后得到的复制数据X^*与不含误差的纯数据矩阵之间的误差，

- 它是在因子分析中不可排除的误差，即：

$$IE = RE\sqrt{r/p}$$

- 嵌入误差IE随主因子数*r*的取值不同而不同。当*r*的取值小于体系中真实存在的因子数目时，由于残差RE值较大，IE值也较大。随着*r*取值的增加，IE逐渐减小到一极小值，随后又随*r*值的增加而增大，因而IE取得极小值时的*r*值为主因子数。
- IND函数法是常用的一种判断主因子数的方法，它由下式表示：

$$IND = RE/(p-r)^2 = \sqrt{\sum_{k=r+1}^{p}\lambda_k/n(p-r)^4}$$

- IND函数指示的灵敏度优于IE，IND随着*r*值的增加开始减小，随后又增大，因而有一极小值。IND函数取得极小值的*r*值才是主因子数。

✓ 交叉验证法

- 交叉验证法（cross-validation）是一种以数据内部验证为基础的方法，它意味着测量数据阵中每一个元素将被其余元素的数学模型所预报。
- 交叉验证法将一组已知的标准的测量矩阵*X*及其对应的组成矩阵*C*中的数据对应分成*n*个子集，例如可将每个样品作为一个子集，将这*n*个子集中的一个作为预报集，其余(*n*-1)个作为校正集。因此用(*n*-1)个标准样品经校正计算建立合理的校正模型，然后以此模型对留出的那个预报集样品进行预报。
- 注意，在建立模型的过程中，分别取*r*=1,2,…个因子数，例如选取*r*=1个因子进行校正计算。

- 这样将*n*个子集轮流做一次预报集样品，经过*n*次校正-预报过程后，就可以预报*r*取不同值时的误差的平方和PRESS(predictive residual error sum of squares):

$$PRESS(r) = \sum_{i=1}^{n}\sum_{j=1}^{m}(c_{ij} - \hat{c}_{ij})^2$$

- 式中，c_{ij}是第*j*个组分在第*i*个样品子集中的浓度，而$\hat{c}_{ij}$是在因子数取*r*时所建立的矫正模型的对应预报值，*m*是组分数，*n*为样品数。

- 分别取*r*=1,2,….*p*个因子数来构造校正模型，并计算各个*r*值时的PRESS值。显然当*r*取值小于体系中实际存在的独立组分数时，由于校正模型丢失某些组分的贡献，因而PRESS值较大。
- 而当*r*的取值大于体系中独立组分数时，由于引入了多余的误差信息，PRESS值也将变大。只有*r*当取值等于体系中实际存在的独立组分时，PRESS取得极小值。为找出这个最佳的*r*值，可将*r*取不同值时的PRESS值加以比较，有:

$$W(r) = PRESS(r+1)/PRESS(r)$$

- 当*W*(*r*)大于1时，所得*r*即为主成份数（主因子数）。

- 以上所述的交叉验证法以测量数据预报其相应浓度组成，这是在多组分分析中常用的方法。还可以直接用测量数据矩阵*X*的残差平方和*RSS*(*r*)来确定主成分数：

$$RSS(r) = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_{ij} - \hat{x}_{ij})^2$$

- 式中，$\hat{x}_{ij}$为因字数取*r*时又校正模型预报后重建的x_{ij}值，注意到在该法中不必将数据矩阵分成若干子集，而是同时利用所有的数据进行校正计算，因此*RSS*(*r*)值实际上是测量数据矩阵提取*r*个主成分后所得残余矩阵中各元素的平方和。
- 这种方法得到的数据-残差平方和*RSS*随*r*的增大具有逐渐减小的趋势，最后趋于平缓，但一般得不到极小值。

这时可以假设一个较小的判别值，当前后两个RSS的差小于或者等于该预先确定的值时的因子数，便可以认为是最佳因子数。

- 通过计算不同数目因子数时的交叉检验模型的PRESS值，可以确定最佳的因子数目。当模型给出最小的PRESS值，设这时的主因子数为r^*，理论上r^*是最佳的。
- 但一般也会发现，采用其他因子数$r(r<r^*)$时，所建立的模型的PRESS值与r^*所对应的PRESS有时非常接近，特别是r与r^*较为接近时。

9.3 偏最小二乘法

- 偏最小二乘法（partial least squares regression, PLS）是一种基于因子分析的多变量矫正方法。偏最小二乘是目前在近红外光谱中应用最多的多元校正方法。与主成分回归不同的是，偏最小二乘在光谱降维的同时考虑了性质阵的作用。
- 但当校正集样本中出现较大的奇异点，或个别样品的性质范围已超出校正集的样本的范围，则可能出现较大的偏差。
- PLS中，光谱和含量的分解同时进行，并将含量信息引入到光谱数据分解过程中，在每计算一个新主成分之前，交换光谱与含量的得分，从而使光谱主成分直接与被分析组分含量关联。

- 以红外光谱分析在化学测量和有关研究中的应用为例，来阐述偏最小二乘法。
- 最小二乘法首先将校正集的混合标准溶液的吸光度矩阵A和浓度矩阵C分别进行主成分分解，计提取r个主因子，使组分的贡献与误差得以分离：

$$A = TP^T + E = \sum_{i=1}^{r} t_i p_i^T + E$$

$$C = UQ^T + F = \sum_{i=1}^{r} u_i q_i^T + F$$

- 式中，T和P分别是测量矩阵A的得分矩阵或载荷矩阵；U和Q分别为浓度矩阵C的得分矩阵和载荷矩阵；E和F分别是测量矩阵和浓度矩阵的测定误差矩阵（亦称为残差矩阵）。T和U之间可由如下回归式予以关联：

$$u_i = b_i t_i$$

- 式中，u_i和t_i分别是U和T的第i列，然后对T进行旋转变换，使T即可以描述C矩阵，也可描述A矩阵。对T进行旋转变换的基本思想是在从A矩阵分解出T时引入U的信息，或从C分解U时引入T的信息。具体的实现方法一般是在迭代过程中交换迭代变量，即以U代替T计算P^T，以T代替U计算Q^T。得到T与U之间关系后，即可由下式进行预报：

$$C_{new} = T_{new} BQ^T = A_{new} PBQ^T$$

- 式中，C_{new}表示一组未知混合溶液的浓度矩阵；A_{new}和T_{new}分别表示未知溶液的测量数据矩阵和其对应的得分矩阵。对于单个未知溶液其浓度向量C_{new}^T由下式表示：

$$C_{new}^T = t_{new}^T BQ^T = a_{new}^T PBQ^T$$

- Otto等曾建议采用下式对未知物进行浓度预报，但应用该式对未知样品进行预报的报告并不多见。

$$C_{new}^T = a_{new}^T (U^T A)^T BQ^T$$

- 实际上的偏最小二乘法运算并不独立对测量矩阵和浓度矩阵分别进行主成分分析，而是通过迭代的方法，彼此之间交换信息，将两个独立的主成分分析过程合二为一，得到PLS解。
- 偏最小二乘法最常用的算法是非线性迭代偏最小二乘法（NIPALS），其步骤如下（不失一般性，以Y表示浓度组成即自变量矩阵，以X表示其对应的测量数据矩阵）：
  - NIPALS算法对测量数据矩阵X和浓度矩阵Y进行交替迭代运算，有：

$$\omega \propto X^T u$$
$$t = X\omega$$
$$q \propto Y^T t$$
$$u = Yq$$

  - 在Y矩阵中任选一列向量作为初始向量u后便可开始上述的循环迭代，直至收敛。这时有

$$\omega \propto X^T u \propto X^T Yq \propto X^T YY^T t \propto X^T YY^T X\omega$$

  - 故实际上ω为矩阵X^TYY^TX的一个特征向量，同样q为矩阵Y^TXX^TY的一个特征向量。该两个矩阵为两个对称矩阵的积，即$(X^TY)(X^TY)^T$和$(X^TY)^T(X^TY)$，可以看到交叉矩阵X^TY为其基本计算矩阵。
  - （1）对测量矩阵X的各列进行中心化和归一化预处理：

$$X = (X - \bar{x}_j)/s_j \quad j = 1,2,...,n$$

- 对浓度矩阵*Y*的各列进行中心化和归一化预处理：

$$Y = (Y - \bar{y}_j)/s_j \quad j = 1,2,...,n$$

- 式中，$\bar{x}_j$ 和s_j分别表示测量矩阵*X*第*j*列的平均值和标准偏差，第二式中$\bar{y}_j$和s_j分别表示浓度组成矩阵*Y*第*j*列的平均值和标准偏差；
- （2）去*Y*中任意一列作为起始的*u*：

对于*X*矩阵，有：

- （3）由此*u*及*X*计算权重向量 $\omega^T : \omega^T = u^T X / u^T u$
- （4）对权重向量ω^T 进行归一化处理：$\omega^T_{new} = \omega^T_{old} / \|\omega^T_{old}\|$ (|| is norm)
- （5）计算 $t : t = X\omega / \omega^T \omega$

对于矩阵**Y**，有

- （6）计算矩阵Y的载荷向量

$$q^T : q^T = t^T Y / t^T t$$

- （7）对载荷向量q^T进行归一化处理： $q^T_{new} = q^T_{old} / \|q^T_{old}\|$
- （8）计算新的迭代值 $u : u = Yq / q^T q$
- （9）收敛检测：比较第5步所得的*t*值和前一轮迭代过程中的*t*值（第一轮迭代时选例初始值），若二者相等或其差在给定的误差范围内，则迭代停止，并至第10步继续运算；若两者不等，则将新*t*取上一轮迭代所得*t*，进入第3步继续迭代。若*Y*仅为一向量，则跳过6~9步，直接令q^T=1。
- 计算矩阵*X*的载荷，并对得分向量和权重向量重新进行标准化处理。
- （10）计算*X*矩阵的载荷向量 $p^T : p^T = t^T X / t^T t$
- （11）对载荷向量p^T进行归一化处理： $p^T_{new} = p^T_{old} / \|p^T_{old}\|$

- （12）对*X*的得分向量*t*进行标准化处理：$t_{new} = t_{old} / \|p^T_{old}\|$
- （13）对*X*的权重向量进行标准化处理：$\omega^T_{new} = \omega^T_{old} / \|p^T_{old}\|$

注意到上面计算所得的向量p^T，q^T和ω^T 应被用于预测计算，而*t*和*u*则可用于分类计算。

- （14）对于向量*t*和*u*采用系数*b*进行关联： $b = t^T u / t^T t$

对第*h*个主成分完成了迭代计算后，可分别进行测量矩阵*X*和浓度矩阵*Y*的残差：

- （15） $X_h = X_{h-1} - t_h p_h^T$
- （16） $Y_h = Y_{h-1} - b_h t_h q_h^T$

比较前后两次迭代计算所得到的残差矩阵是否相等（即前后两次计算所得残差矩阵中各元素的平方和是否小于一个给定的误差），如果是校正计算结束。否则以X_h取代X_{h-1}，Y_h取代Y_{h-1}。返回第2步进行下一主成分的迭代计算。

- ✓ 实际中，一般采用迭代算法对未知样品中的浓度进行估计。
- （1）将未知样本测量数据向量x_{unk}按校正中的方法进行数据预处理： $X^T_{unk} = (X^T_{unk} - \bar{x}_j)/s_j$
- （2）令*h*=0；
- （3）设*h*=*h*+1，并计算未知测量向量x_{unk}第*h*个主成分的得分值 $t_h : t_h = x^T_{unk}\omega_k$
- （4）完成所有的主成分运算之后(即*h*=*r*)，计算未知物的浓度的向量

$$y_{unk} : y^T_{unk} = \sum_{h=1}^{r} b_h t_h q_h^T$$

- （5）将所得的标准化的浓度向量 y^T_{unk} 恢复到原始状态：。

$$y^T_{unk} = y^T_{unk} + \bar{y}_j$$

- 如上所述，PLS法在构造校正模型时更充分地利用了*X*和*Y*矩阵中的信息，是比较完善的基于因子分析原理的校正，使用这种方法可以降低噪声对校正模型的影响。
- 通常情况下，非线性偏最小二乘法与偏最小二乘法的计算差别在于前者以多项关系式取代线性关系式来关联数据矩阵的得分向量*t*和浓度矩阵的得分矩阵*u*:

$$u = b_0 + b_1 t + b_2 t^2$$

9.4 人工神经网络

- 人工神经网络（artificial neural networks，ANN），是在现代生物学研究人脑组织所取得成果的基础上提出来的，利用由大量简单的处理单元广泛连接而组成的复杂网络，来模拟大脑的神经网络结构和行为，如记忆、联想、学习和归纳等功能。
- 人工神经网络介于常规计算机和人脑之间：一方面，ANN试图模拟人脑的功能；另一方面，许多实现技术是常规的，从而能够用计算机来实现人脑的的某些功能，解决许多常规计算机难以解决的问题。

9.5 遗传算法

- 遗传算法（genetic algorithm， GA）是一类借鉴生物界的进化规律演化而来的随机化搜索方法，其主要特点是直接对结构对象进行操作，不存在求导和函数连续性的界定。与传统的优化算法相比，遗传算法主要有以下一个特点：
- （1）通用性。目前使用较多的优化方法，大多是一些基于梯度信息的优化方法。这类优化方法不管是直接法还是间接法，都存在一些局限性。例如用这类优化方法求解得到的优化解，大多是在原始解附近的局部优化解，因而不同程度地存在着局部收敛等局限性。这类优化方法需利用目标和约束函数的导数信息，因此这类优化方法对函数的性能要求较高。

- 而遗传算法不是直接作用在参变量集上而是利用参变量集的某种编码和个体的适应度进行群体的进化。遗传算法仅需要计算目标函数值，而不需要优化模型中目标函数和约束函数的导数或其他辅助信息。因而GA能解决各种优化问题，不论其设计变量连续与否（为连续设计变量、离散设计变量或混合设计变量），目标函数和约束函数是否连续、可导。尤其还可以快速地搜索复杂、高度非线性和多维空间，可以解决许多传统优化方法难以解决的问题。

- （2）GA具有全局寻优能力。采用一般的优化方法，要求得到优化问题的全局最优解，一般需要求解出优化问题的所有极值点，这就要求优化方法能遍历整个设计空间，但在实际计算中，由于设计变量空间很大，很难遍历整个设计空间，因此用现有的优化方法难以得到优化问题的全局解。GA利用设计变量编码在变量空间进行多点搜索，因而遗传算法在搜索的空间上将比现有优化方法要大。遗传算法中杂交算子能使群体进化不断向最优个体逼近；GA中的变异算子能避免杂交繁殖收敛于局部优良个体，保持群体搜索的多样性。这些确保了遗传算法中多点搜索一直处在不同的局部区域。因此，GA比现有优化方法具有更强的全局寻优能力。

- （3）并行性。GA从初始化群体，经过复制、杂交和变异等操作，产生新的群体。每次迭代计算，都是针对一组个体同时进行而不是针对某个个体进行，因而遗传算法具有隐含的并行性。

9.6 小波变换

- 小波变换（wavelet transform， WT）是一种新型的信号处理方法，是给出时间域和频率域方面信息的另一种技术，类似于傅立叶变换，但它可以做时域局部化分析，又具有时间窗口宽度随频率的变化而自动调节的特性，是傅里叶变换所不具备的。
- WT将测量信号分解为一组称之为小波基的基函数。在小波变换中，这种小波基函数被称为分析小波（analyzing wavelet）。通常使用较多的分析小波类型有Morlet小波和Daubechies小波。另外还有一种比较特殊的Haar小波（呈方波形）。小波族是对测量数据的小波进行伸缩(Stretching)和平移（shifting）形成的。

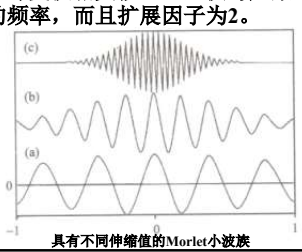
- 表示移动的参数为b，称为变位因子或平移因子（translation），而扩展参数a对小波有伸缩和拓宽的作用，a常被称作伸缩（dilation）因子或尺度因子，该分析小波用函数$h(T)$表示（称小波母函数），通过引入变位因子b和伸缩因子a，可得连续变化的小波族$h_{a,b}(T)$：

$$h_{a,b}(T)=\frac{1}{\sqrt{|a|}}h(\frac{t-b}{a})$$

- 其中有$a=2^n$, $b=2^n k$
- 式中，n和k分别为代表伸缩和变位的参数，且都是整数。这样便可得小波族的数学表达式(二进小波)：

$$h_{a,b}(T)=2^{-n/2}h(2^{-n}t-k)$$

- 变位参数确定小波在时域中的位置，而伸缩参数不仅确定其在频域中的位置，还确定时-频局部的尺度或程度。
- 下图表示伸缩因子a取不同值时的Morlet小波族。与傅里叶变换相类似，这里仅考虑某一确定测量时间处的频率，而且扩展因子为2。



具有不同伸缩值的Morlet小波族

- 小波变换要求小波基适合测量数据。小波基由对小波母函数$h(t)$进行伸缩和移动而获得。最狭窄的小波（水平a^{-1}）以较小的步长移动，而较宽的小波则以较大的步长移动。变位因子b常常是伸缩因子的倍数(k倍)。由适合数据的这些小波基可获得小波变换系数（wavelet transformation coefficient）
- 与较窄小波相联系的系数表示信号的局部特征，而较宽的小波表示信号的平滑特征。

- 将离散小波变换应用于离散形式的测量数据，要求数据的数目为2^n。在离散小波变换中，由小波滤波系数给出分析小波。例如Haar小波族第一个系数组（最小的伸缩因子a，而变位因子b=0）被定义为两个系数c_1=1和c_2=1。第二个系数组（伸缩因子2a ）的系数组为： c_1=1、c_2=1 、c_3=1和c_4=1。通常小波数由2^n个系数所确定。最宽的小波具有2^n =N个系数，即测量数据的数目。n的大小确定小波的水平。

- 例如，对于n=2，可获水平2的小波。对于各个水平，转换矩阵中小波滤波系数都按一定的方法排列。对于一个含8个数据点的测量向量（以8 × 1的列向量表示）在水平1时，变换矩阵为：

$$\begin{bmatrix} c_1 & c_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & c_1 & c_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & c_1 & c_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & c_1 & c_2 \end{bmatrix}$$

- 将此变换矩阵与上述数据信号列向量相乘便产生4个小波变换系数，即数据向量长度的一半(N/2)
- 对于 c_1=1、c_2=1 、c_3=1和c_4=1的情况，小波变换系数等于4个数据点信号的移动平均值。因此，小波滤波系数定义了一个低频通过的滤波器，所得的小波变换系数具有信号的“平滑”信息。

- 因此，改组小波滤波器系数被称作近似系数或逼近系数（approximation coefficient）,而所产生的变换系数为a-部分（approximation）。含有近似系数的变换矩阵称为G矩阵。上述8个测量点的例子中，可能的最高变换水平为水平3（2^3=8非零系数），这种变换的结果为测量信号的平均。而零水平（2^0=1个非零系数）的变换即是信号本身。
- 除去以上第一组系数之外，还定义了第二组滤波系数，相当于一个高通滤波器，它对测量信号作细节性的描述。高通滤波器利用上述的同一组小波系数，但其符号相反且顺序也相反。

- 这些系数被置于H矩阵中，信号长度为8而变换水平为2的H矩阵为H矩阵中的系数称为描述系数，或细节系数（detail coefficients），而H矩阵的输出为d-成分（detail）。在已知N/2个细节描述成分和N/2个近似描述成分的情况下，便有可能重构长度为N的信号。

$$\begin{bmatrix} c_2 & -c_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & c_2 & -c_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & c_2 & -c_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & c_2 & -c_1 \end{bmatrix}$$

- 离散小波变换能用向量-矩阵的形式来表示：

$$\alpha = W^T f$$

- 式中， α 含有N个小波变换系数；W^T为含有与指定小波相关的近似和细节系数的N×N阶正交矩阵；

- f为测量数据向量。矩阵W^T的作用是分别采用低通滤波器G和高通滤波器H进行两个相关的卷积计算。G的输出表示平滑方面的近似信息，而H的输出则给出较详细的细节信息。
- 小波变换的应用主要在波谱除噪方面。在小波变换计算中，如将某些小波系数置零，便可有选择地除去测量信号中某些区域中的噪声，而对其余区域无明显影响。
- 小波变换还用于声波信号、图像信号处理、地表波信号的测量和分析。小波变换还可被用于噪声信号中信号峰的检测，噪声中的某些变化较大的信号将引起同一位置小波信号的变化。

- 可通过小波变换进行数据压缩，而数据中的信息并不丢失。测量信号经小波分解以后将从原来的空间投影到小波空间，由于小波变换的特点，在小波空间的系数将有一部分特别小，对信号的表达没有显著的意义。如果将较小的系数去除，在重构的信号中将不会丢失有意义的信息。因此，小波变换可用于数据的压缩。
- 小波变换还可用于多元校正分析中信息量的提取。