# 第三章 矢量数据与栅格数据

2021 shapefile数据

2019 栅格数据重采样的最近邻法和双线性插值法

2017 栅格数据的表达、取值、组织

2015 矢量数据与栅格数据

2006 空间数据压缩

2020 根据河流分布图，对矢量数据进行栅格化，并说明栅格大小对结果的影响。

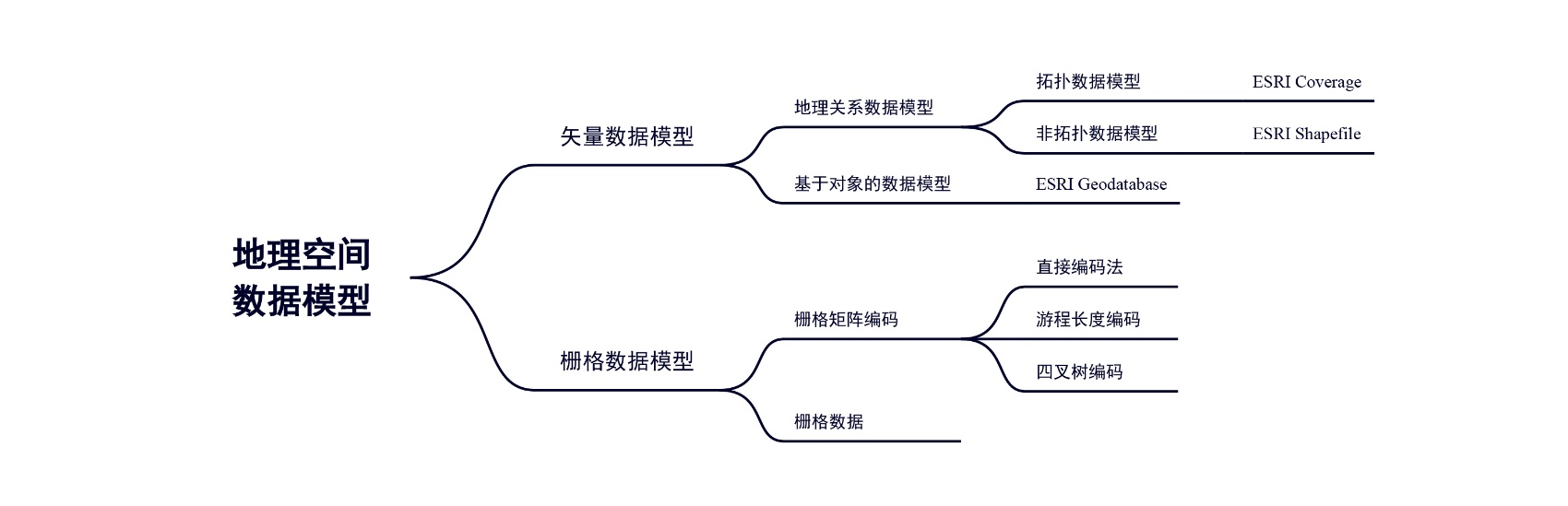
2011 请比较矢量数据模型与栅格数据模型在表达空间对象上的不同点及优缺点，并完成计算。

2009 什么是游程编码？叙述游程编码的方法并对以下栅格数据进行游程编码。

2009 请写出下图的DIME数据文件，并写出对多边形112进行连接编辑的算法步骤。

2005 目前地理信息系统常用的空间数据(矢量数据)格式有哪些？不同格式空间数据如何融合？

2004 叙述游程长度编码的基本思想，并以此设计一种具体的游程长度编码方法对以下数据进行编码。

2001 叙述矢量编码的三种方法及各自特点？

## 一、矢量数据模型

矢量数据模型通过记录地理空间实体坐标的方式精确地表示点、线、面等实体的空间位置和形状。这种数据模式能够很好表达地理数据的空间分布特征,且精度高,数据存储冗余度较低，属性隐含与定位明显。矢量数据模型按照对于空间（图形）和属性数据的存储方式可分为

* 地理关系数据模型
* 基于对象的数据模型

地理关系数据模型

用两个独立的系统分别存储空间和属性数据：用图形文件存储空间数据，用关系数据库存储属性数据，用要素标识码（ID）对两者进行链接。矢量型数据模型中的**图形文件**，按其是否明确表示地理实体之间的空间相互关系可分为实体型和拓扑型两大类。Coverage模型和Shapefile数据是典型的地理关系模型。

基于对象的数据模型

基于对象的数据模型将地理空间数据作为对象（Object），GIS中几乎所有组成部分都可以作为对象表示，其把空间数据和属性数据存储在一个系统中。采用BLOB（Binary large object）数据类型存储空间数据（图形数据）；图形数据通常采用拓扑型数据结构。

基于对象的数据模型允许一个空间要素（对象）与一系列属性和方法相联系。属性描述对象的性质或特征，方法执行特定的操作。例如：作为一个要素类对象，一个公路图层可以具有形状和范围的属性，也可以有复制和删除的方法。属性和方法直接影响GIS操作如何执行。

矢量数据结构

矢量数据结构是指矢量数据中图形文件(或图形记录)的存储方式，可以分为实体式数据结构、Arc/Node、拓扑型数据结构。

实体数据模型

边界坐标数据和多边形单元实体一一对应，各个多边形边界点都单独编码并记录坐标  
**优点：** 结构简单、直观、编码容易。  
**缺点：** 相邻多边形的公共边界被数字化并存储两次，造成数据冗余和不一致。

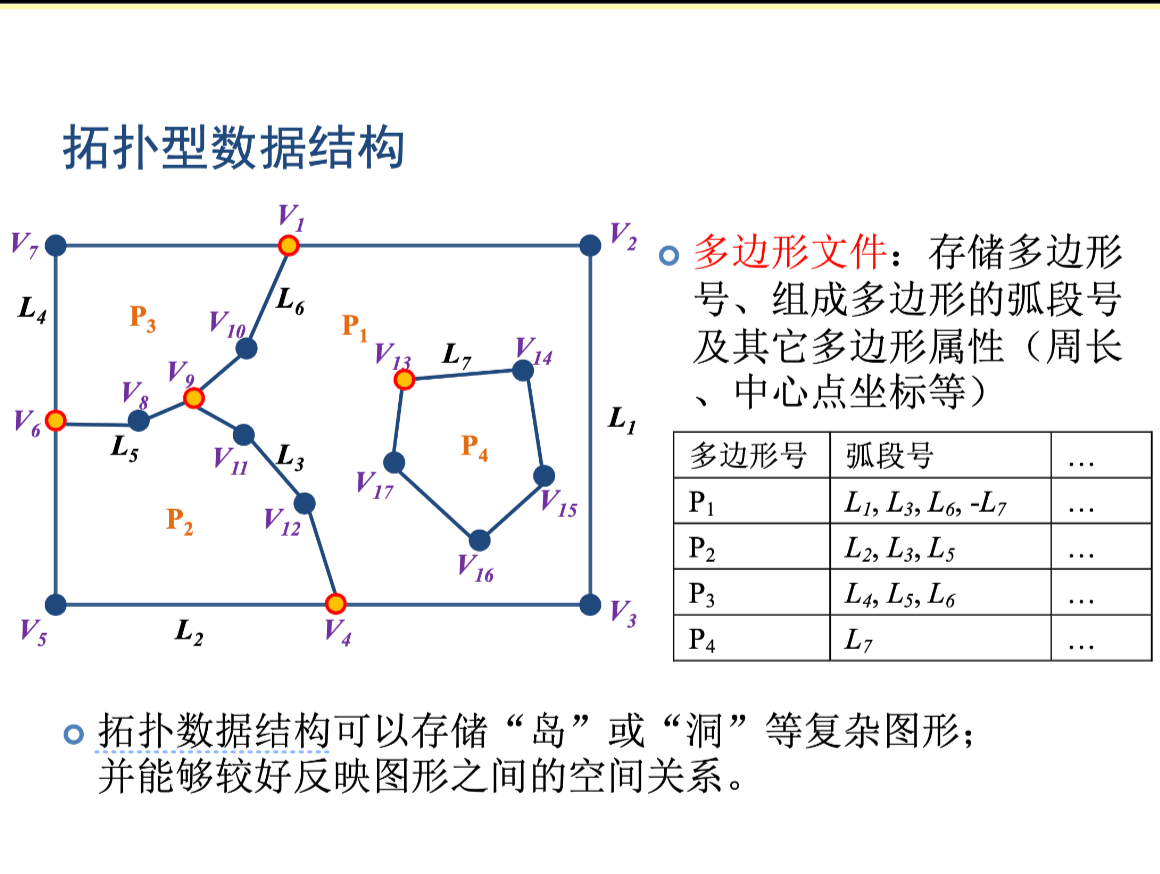
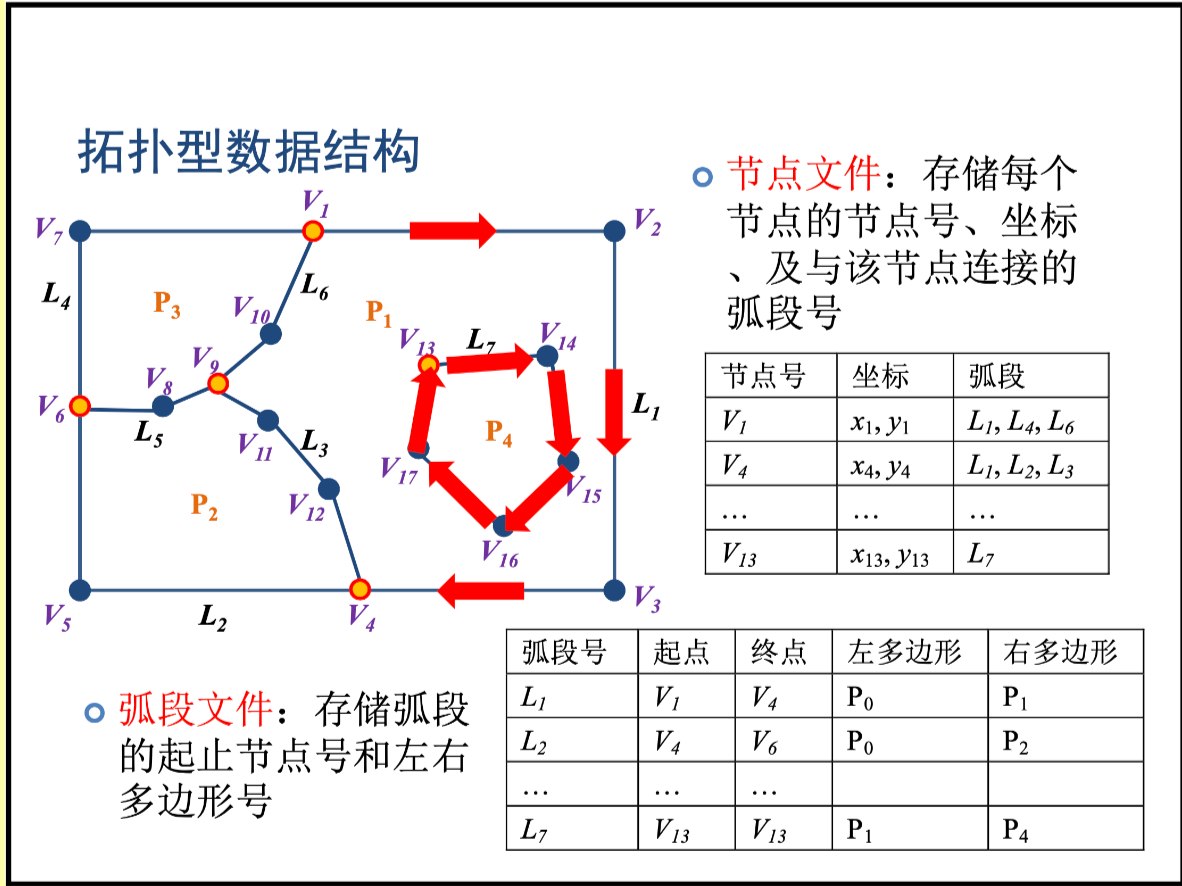
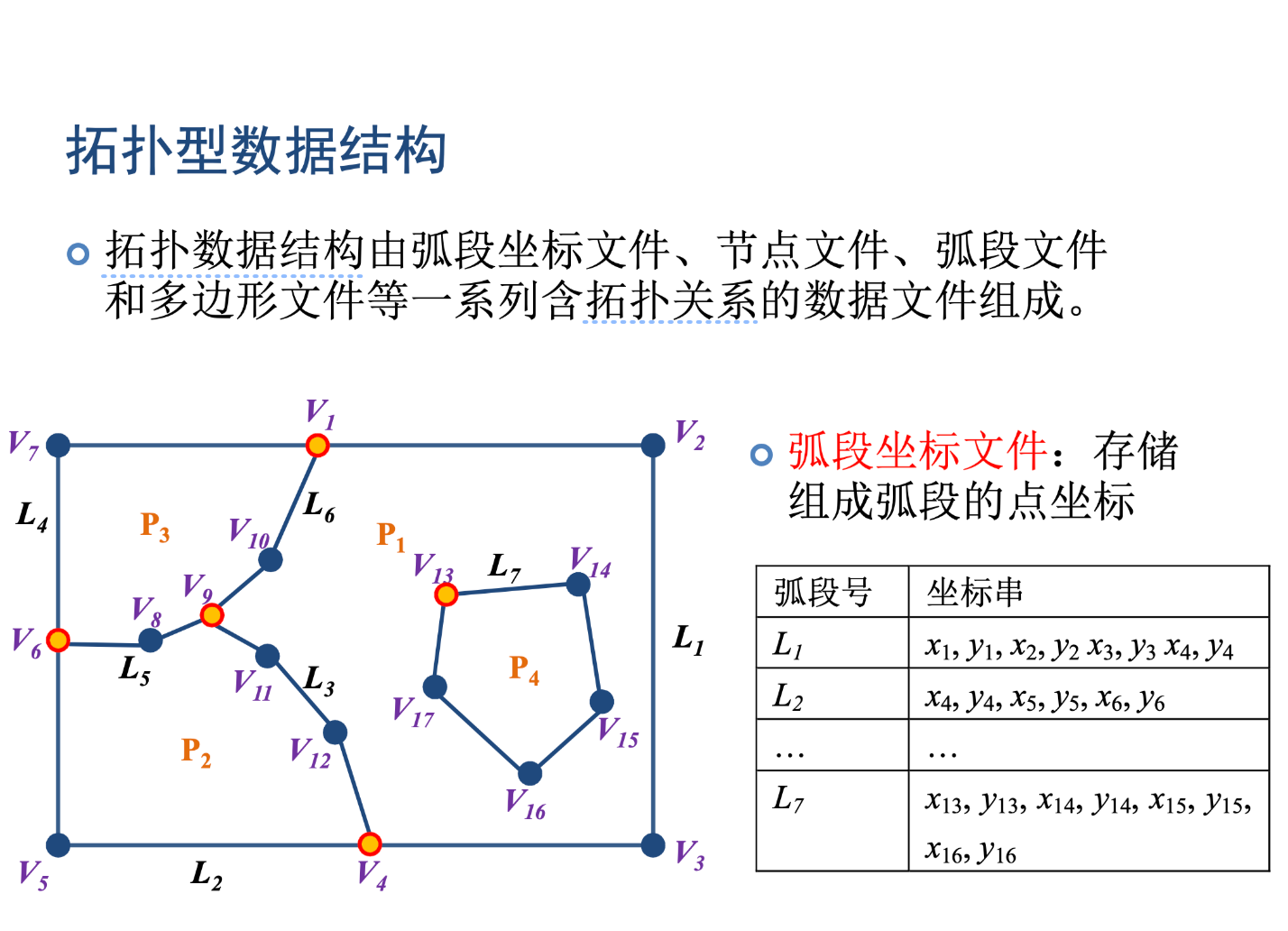
·实体互相独立缺乏联系，无法表示拓扑关系。岛图形拓扑难以检查

**Arc/Node**

* 线的端点（起点、终点）称为节点（Node）
* 组成线的其它点称为顶点或拐点（Vertex）
* 节点与节点之间的线称为弧段（Arc）

拓扑型数据结构

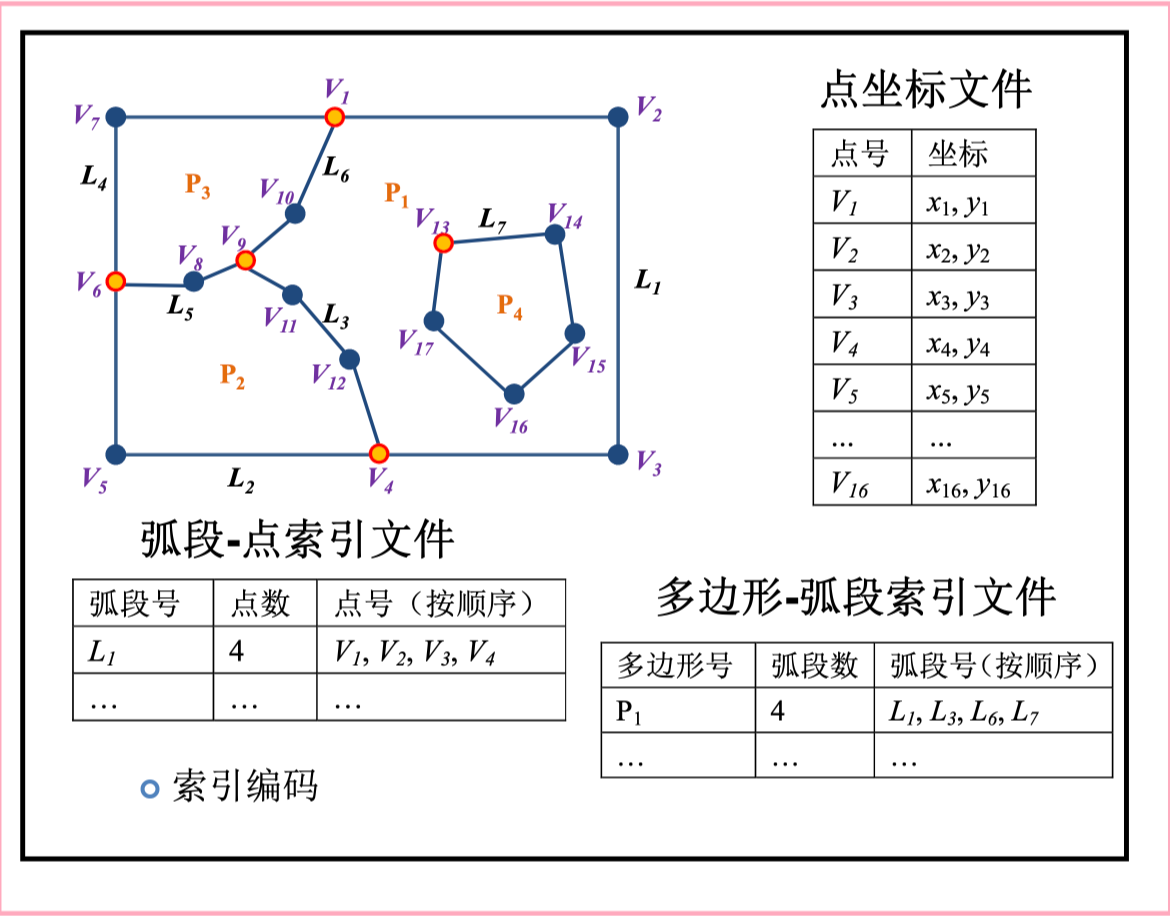
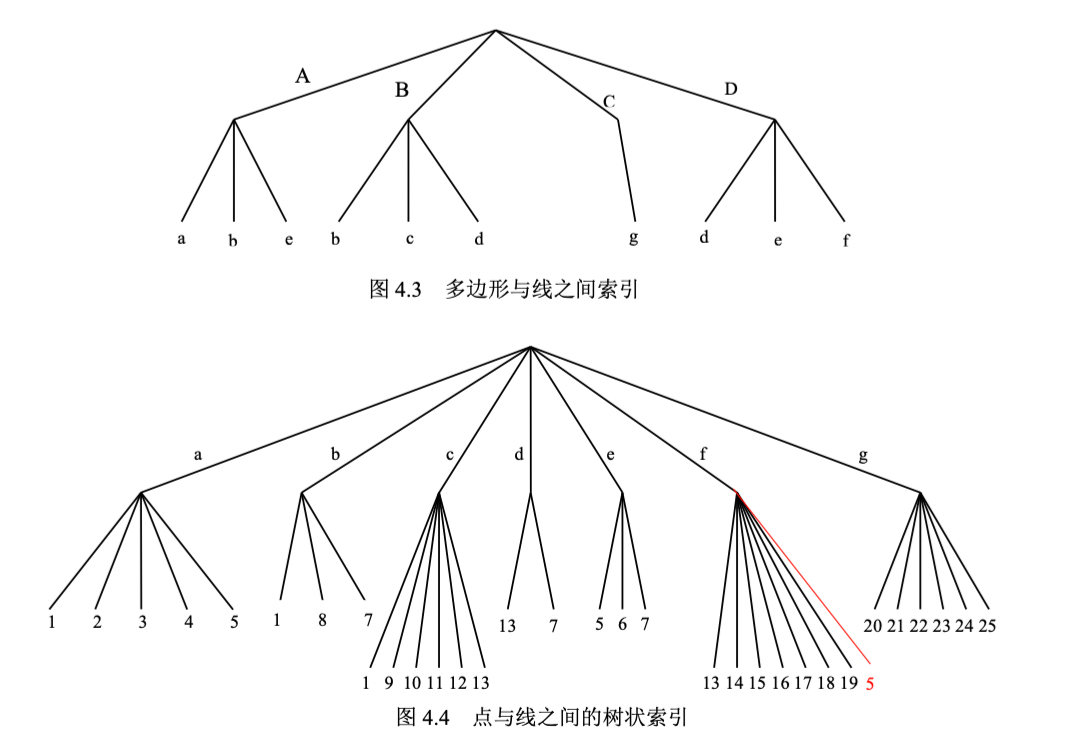
拓扑数据结构由弧段坐标文件、节点文件、弧段文件和多边形文件等一系列含拓扑关系的数据文件组成。

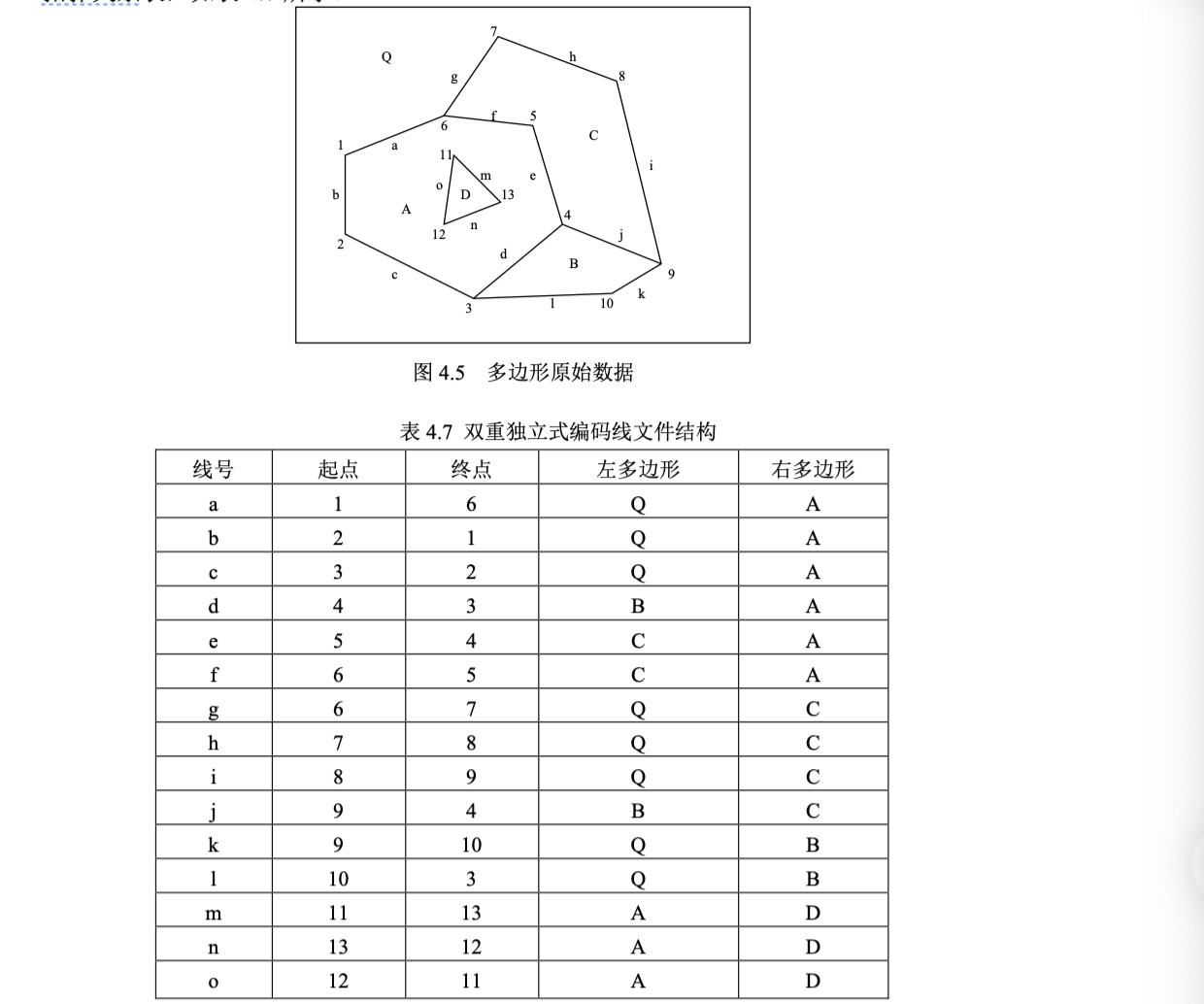
**拓扑关系：**是地理空间实体相互之间空间关系的一种，是物体在空间中相互之间位置关系的定性描述，并且这些关系不会因为物体的弯曲和拉伸等弹性形变（指对象未被撕裂）发生改变。主要用来描述地理实体间的相邻、连通、包含和相交等关系。

矢量数据编码

索引式数据结构

其先建立点坐标文件，再建点与边(线)、线与多边形的索引文件。

**优点：** 用建索引的方法消除多边形数据的冗余和不一致，邻接信息、岛信息可在多边形文件中通过是否公共弧段号的方式查询。  
**缺点：** 表达拓扑关系较繁琐，计算困难。



双重独立编码(DIME, Dual Independent Map Encoding)

双重独立编码是一种拓扑编码结构，其由点文件、线文件(线文件以线段为记录单位)与面文件定义数据。

**优点：** 利用这种拓扑关系可以来组织数据，可以有效地进行数据存储正确性检查（如多边形是否封闭），同时便于对数据进行更新和检索。  
**缺点：** 在表示线段的时候，一条线只能用直线两端点的序号及相邻的多边形来表示。不能显示线段中间部分的细节。

链状双重独立编码

链状双重独立式数据结构是DIME数据结构的一种改进，在链状数据结构中，将若干直线段合为一个弧段（或链段），每个弧段可以有许多中间点。链状双重独立式数据结构主要由四个文件组成：**多边形，弧段文件，弧段坐标文件，节点文件**。比双重独立编码更加精简。  
**优点：**拓扑关系明确，能表达岛信息，其以弧段为记录单位。

**缺点：** 当图形数据修改、删除、增加点、线、面要素后，其拓扑关系也发生改变，需要重新建立拓扑关系。其也是Coverage数据模型的数据编码形式

矢量数据文件格式

Coverage数据模型

该空间数据结构是ESRI公司于1980年开发的空间数据文件存储形式，该数据模型将空间坐标数据存储与一系列的数据文件中(称为Arc)，而属性数据另外存储与关系数据库中(称为Info),空间要素和对应的属性记录采用标识码(ID)进行逻辑链接。采用拓扑型数据结构来管数据，其支持以下三种拓扑关系：

* 连通性(Connectivity)：弧段(Arc)在节点(Node)处彼此相连;
* 面定义(Area Definition)：面(Area)用围绕它相连的弧段(Arc)来定义;
* 邻接性(Contiguity)：弧段(Arc)有方向性以及左右两侧的面(Area) 。

点文件由坐标和点ID组成，线文件由弧段-节点表和弧段-坐标表组成。面文件由左/右多边形表、多边形/弧段表、弧段坐标表。 **优点：**

* 有利于数据文件的组织，并减少数据冗余。每个空间对象都是有更基本的对象，只有结点的坐标是被实际存储的，其他复杂对象的坐标实际上是逻辑构成。两个多变形之间的共享边界在弧段坐标表中只列一次，而不是两次。
* 共享边界定义两个多边形，所以更新多边形就变得相对容易

**缺点：**

* 对单个地理实体操作效率不高，难以表达复杂的地理实体，难以实现快速查询和复杂的空间分析。
* 局部更新困难，系统难以维护与扩充。

Shapefile数据模型

Shapefile 是ESRI于1990年推出的一种基于地理关系数据模型的矢量数据格式。shapefile 中的地理要素可通过点、线或面（区域）来表示，其采用实体型数据结构的方式存储空间数据。点是用一对坐标、线是用一系类点、多边形用一系列的线来存储，但是没有描述几何对象拓扑关系的文件。shapefile文件由多个同名不同后缀的多个文件构成，其中主文件(.shp)、索引文件(.shx)和表文件(.dbf)必不可少。

* **主文件**是一个直接存取、变长记录的二进制文件，其中每个记录包含一个要素(点、线或面)的所有顶点坐标。
* **索引文件**主要包含主文件数据的索引信息。
* **表文件**储存要素属性信息的dBase文件(关系数据库文件），几何和属性间的对应关系是基于一个不重复的记录顺序代码来实现的。

**优点：**

1、非拓扑矢量数据能比拓扑数据更快速地在计算机上加载。

2、非拓扑数据具有非专业性和互操作性，可以在不同软件包之间通用。目前，该文件格式已经成为了地理信息软件界的开放标准。

Geodatabase数据模型

Geodatabase是一种基于关系数据库、采用面向对象技术来组织和管理空间数据的空间数据模型。其支持存储多种数据类型(如矢量地图要素类（Feature Class）、表格（Table）、栅格图像（Raster Dataset）以及元数据文档、空间处理工具等其它特殊数据)，可以将多个具有相同空间参照系统（坐标系统）的要素类存放在一个要素数据集（Feature Dataset）中。在要素数据集中，还可以定义拓扑（Topology）和要素连通性规则等。

**优点**：

**1、空间数据统一存储**

所有的数据的几何结构与属性数据由一个数据库存储和集中管理，便于空间查询任务。栅格数据、矢量数据、表格等存储在一个Geodatabase中。所有归属于同一个要素数据中的空间数据都共享一个空间参照系统。

**2、空间数据录入和编辑更加精确**

geodatabase提供即时拓扑，用户可以建立拓扑规则，如数据存在错误，可以通过验证拓扑规则或智能验证行为加以防止。

**目前存在三种****Geodatabase模型**

Personal Geodatabase：使用微软的Access数据库来管理数据，单个文件最大为2GB

File Geodatabase：将所有数据存储在一个文件夹中

ArcSDE Geodatabase：通过ArcSDE连接企业级数据库平台，利用企业级数据库来管理数据。

## 二、栅格数据模型

是指将地理空间划分为形状规则的网格，每个栅格单元上的数值表示地理要素的数量或质量特征，每个栅格的位置由阵列中每个单元的行列号确定。

|  |  |
| --- | --- |
| 优点 | 缺点 |
| * 数据结构较为简单,便于算法实现 * 属性明显、定位隐含 * 空间数据的叠置和组合较为容易,有利于与遥感数据的匹配与分析. * 输出方法快速简便,成本较低廉 | * 图形数据量较大,用大像元减小数据量时,精度和信息量受损失. * 投影变换较难实现 * 图形数据质量低,地图输出不精美. |

栅格数据模型的特征

* 栅格数据中像元值的变化，可以表现连续变化的空间现象。
* 对于点、线和多边形，也可以采用栅格数据表示：用单个像元代表点，用一系列相邻像元代表线，用连续的像元集合代表面。
* 栅格数据可能单波段或者多波段，其必须有空间参照信息(坐标系统)，在GIS软件中才能与其他数据进行叠合。

栅格单元的取值方法

栅格单元取值是唯一的，但由于受到栅格大小的限制，栅格单元中可能会出现多个地物，那么在确定该单元的属性取值，可采用下列方法：中心点法、面积占优法、重要性法等。

* **中心点法**：用位于栅格中心的处的地物类型决定其取值，这种方法常用于有连续分布特性的地理现象。
* **面积占优法**：以占矩形区域面积最大的地物类型作为栅格单元的代码。
* **重要性法**：根据栅格内不同地物的重要性，选取最重要的地物类型作为相应栅格单元代码，这种方法常用于有特殊意义而面积较小的地理要素，特別是点状和线状地理要素）城镇、文通线、水系等。在栅格代中应尽量表示这些重要地物。
* **百分比法**：根据矩形区域内各地理要素所占面积的百分比确定栅格单元的取值。

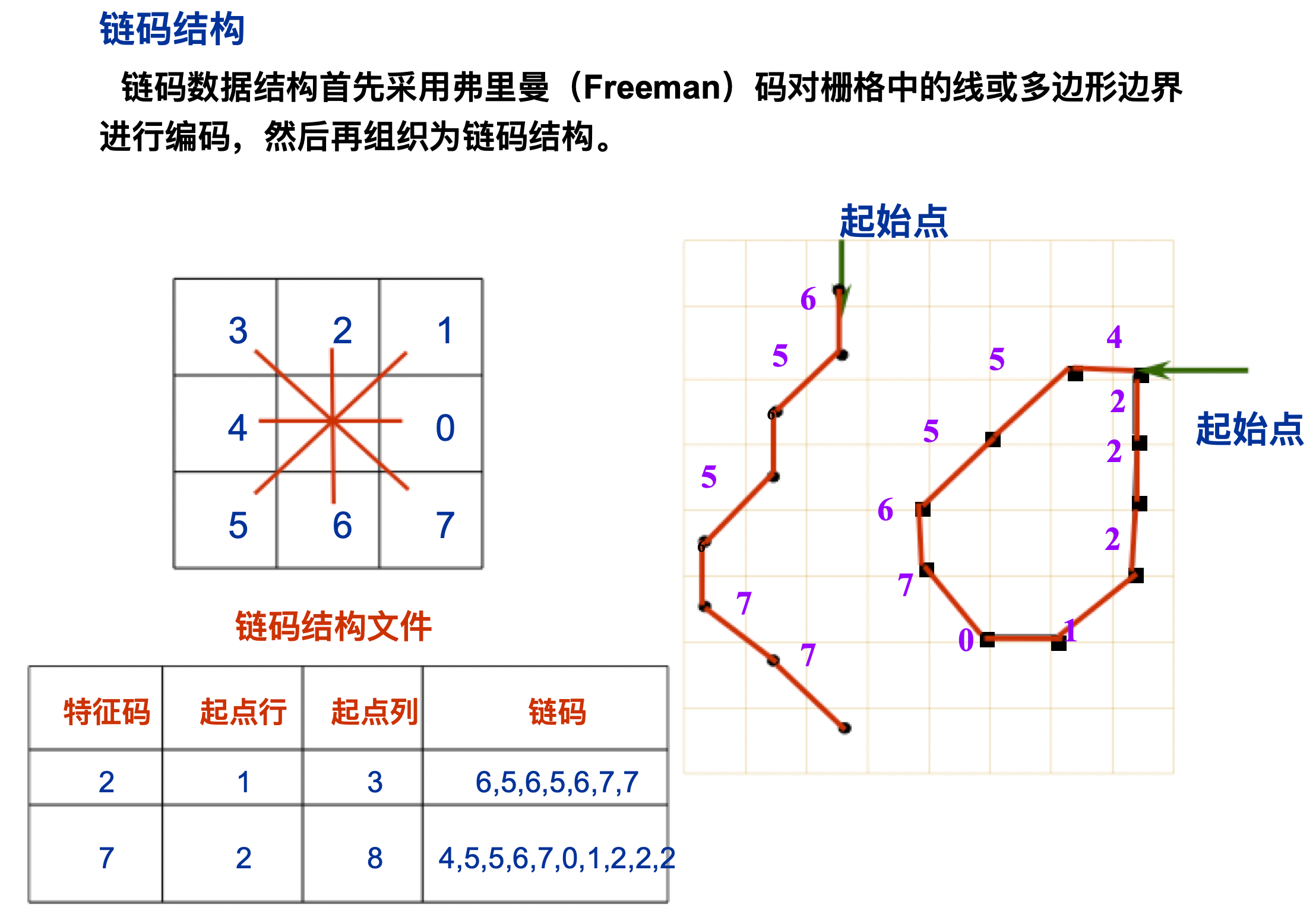
栅格数据的编码方法

* 直接编码法(完全栅格数据结构)
* 链式编码法(Chain Encoding)
* 游程长度编码(Run-length Encoding)
* 块式编码 (Block Encoding)
* 四叉树编码(Quarter-tree Encoding)

直接编码法

将栅格看作是一个数据矩阵,逐行或者逐列记录栅格单元的值,这种栅格基于像元的数据组织方式简单，属于无损编码，无数据丢失。但是无数据压缩，数据冗余严重。

链码数据结构

 首先采用弗里曼（Freeman）码对栅格中的线或多边形边界进行编码，然后再组织为链码结构的文件。链式编码将线状地物或区域边界表示为：由某一起始点和在某些基本方向上的单位矢量链组成。  
**优点**：可以有效地压缩栅格数据，对计算面积、长度、转折方向和凹凸度等运算十分方便。  
**缺点**：对边界进行合并和插入等修改编辑工作比较困难，对局部的修改将改变整体结构，效率较低，而且由于链码以每个区域为单位存储边界，相邻区域的边界将被重复存储而产生冗余。

游程长度（run-length）编码

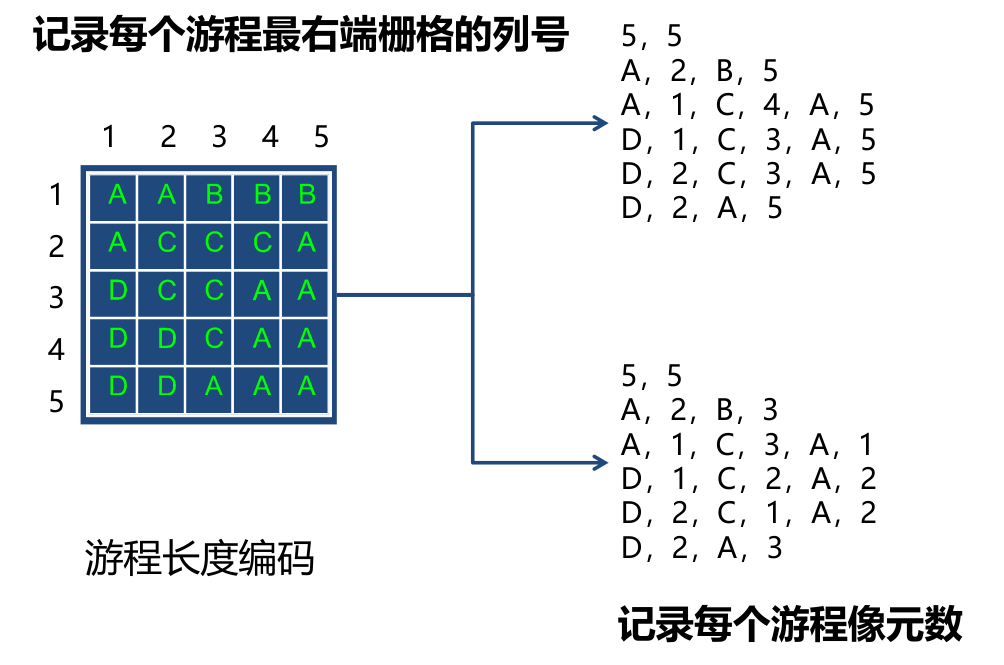
游程指栅格矩阵一行内相邻同值栅格的数量.也称为行程。游程编码结构是逐行将相邻同值的栅格合并,记录合并后栅格的值及合并栅格的数量(即游程)。其目的是压缩栅格数据量,消除数据间的冗余。

**优点：**无损压缩，对于大图斑而言数据压缩明显，适用于类型区面积较大的专题图、分区集中的分类图、二值图等

**缺点** ：对属性变化较大的栅格数据压缩效率不显著。

采用记录游程长度的记录方式有两种

* 记录每个游程最右端栅格的列号
* 记录每个游程像元数



块式编码

块式编码（Block Encoding）是将游程扩大到二维情况，把多边形范围划分成若干具有同一属性的正方形，然后对各个正方形进行编码。块式编码的数据结构由初始位置（行列号）、半径和属性代码组成。

**优点：**无损编码，块码在合并、插入、检査延伸性、计算面积等操作时有明显的优越性。

**缺点**

* 块码与游程长度编码相似，随着图形复杂程度的提高而降低效率，就是说图斑越大，压缩比越高，图斑越碎，压缩比越低
* 在某些操作时，则必须把游程长度编码和块码解码，转换为基本栅格结构进行。

四叉树编码（quarter-tree Encoding）

是一种对栅格数据的压缩编码方法。将整个图像区域按四个象限逐步分解为一系列方形区域，且每一个方形区域具有单一的属性。最小的方形区域为原始栅格数据的一个像元。

分割过程：将欲分解区域等分为四个象限，再根据各个象限的像元值是否单一决定要不要再分。如果单一则不再分割，否则同法再分，直到所有象限的像元属性值相同为止。

**优点**：无损编码，是最有效的栅格数据压缩编码方法之一，绝大部分图形操作和运算都可以直接在四叉树结构上实现，既压缩了数据量，又可大大提高图形操作的效率。   
**缺点**：不利于形状分析和模式识别，即具有图形编码的不定性，如同一形状和大小的多边形可得出完全不同的四叉树结构。

**常规四叉树:**

每个节点通常存储6个量，即 4 个子节点指针，一个父节点指针和一个节点值。常规四叉树对栅格按照莫顿码顺序进行检测,这种方法除了要记录叶节点外,还要记录中间节点,即每记录四个不同节点时,即生产其父节点作为中间节点,如果四个节点相同,则合并生产上一级的一个父节点，这些指针不仅增加了数据存储量,而且增加了操作的复杂性,主要在数据索引和图幅索引等方面应用。

**线性四叉树:**

只存储 3个量,即莫顿码，深度和节点值。编码的基本思想: 不需要记录中间节点，0值节点,也不使用指针,仅仅记录非0值叶节点，并用莫顿码表示叶节点的位置。由于栅格数据常常并不恰好是方针，为了能对不同行列数的栅格数据进行四叉树编码，对不足部分以 0 补足，在建树时，对于补足部分生成的叶节点可不存储,这样存储不增加。

**线性四叉树优点**

* 常规四叉树在处理上简单灵活，但是到栅格数据很大时，存储和处理整个矩阵较困难，可以采用线性四叉树存储法：
* 线性四叉树相比常规四叉树节省存储空间；
* 由于记录节点的地址，既能直接找到其在四叉树中的走向路径，又可以换算出在整个栅格区域内的行列位置，压缩和解压较方便。
* 各部分的分辨率不同，既可以精确表示图形结构，又可以减少存储量，易于进行大部分图形操作和运算。

栅格数据压缩

数据压缩是指减少数据量，节约存储空间，降低数据传输成本。有多种技术可以用于图像压缩，压缩技术又可以分为无损压缩和有损压缩。

* 无损压缩可以保留像元值，允许原始栅格数据被精确重构。
* 有损压缩会损失一部分像元值，不能完全重构原始图像，但是可以达到很高的压缩率。

压缩的栅格格式：

* MrSID（Multi-resolution Seamless Image Database），后缀名为sid，采用小波变换（Wavelet Transform）
* JPEG（Joint Photographic Experts Group）

影像金字塔结构

**影像金字塔是指在统一的空间参照下，根据用户需要以不同分辨率进行存储与显示，形成分辨率由粗到细、数据量由小到大的金字塔结构**。影像金字塔结构用于图像编码和进式图像传输，是一种典型的分层数据结构形式，适合于栅格数据和像数据的多分辨率组织，也是一种栅格数据或影像数据的有损压缩方式。

在金字塔结构里图像被分层表示。在金字塔结构的最顶层，存储最低分辨率的数据：随着金字塔层数的加，数据的分辨率依次降低。在金塔结构的底层，则存储能满足用户需要的最高分率的数据。

## 三、栅格化

将点、线或面的矢量数据，转换成对应的栅格数据，这一过程称为栅格化。其分别针对点、线和面来进行，点的栅格化是线和面的栅格化的基础。

首先是建立矢量坐标的平面直角坐标系和栅格行列坐标系之间的对应关系。

1、确定栅格矩阵的大小

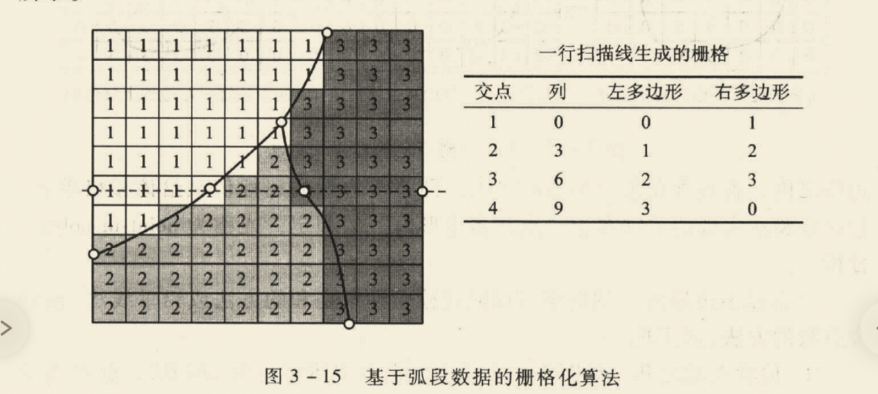
需要先确定栅格**像元的大小**、**像元的分配类型及分配原则**。 栅格像元的大小决定了数据的输出精度。像元值的分配类型主要用于确定多个对象落入同一像元中时像元值的计算类型。

2、分要素类型栅格化

* **点的栅格化：**矢量点在建立矢量坐标与栅格坐标后可以直接转换。
* **线的栅格化：**由于 GIS 中的线实体是由顺次连接一组坐标点形成的折线段表达的，所以线的栅格化可以分解成对组成折线的每一个线段的栅格化。对于一条线段的栅格化，先使用点栅格化的方法，栅格化线段的两个端点，然后利用DDA或者Bresenham算法转化为栅格坐标。
* **面的栅格化：**依据多边形的不同结构，有两种算法，基于**弧段数据的栅格化和基于多边形数据的栅格化**

**基于弧段数据的栅格化**

基于弧段数据的栅格化方法是针对**拓扑矢量数据结构**进行栅格化的方法。对应于Coverage或者是Geodatabase数据模型。  
 对整个要进行栅格化的范围按行或者按列作中心扫描线，对其中的任一条扫描线，求与所有矢量多边形的边界弧段的交点，记录其坐标，并用点的栅格化方法求出交点的栅格坐标行列值，再根据弧段的左右多边形信息判断并记录交点左右多边形的数值。最后通过对一行所有交点按其坐标 x 值从小到大进行排序，并参照左右多边形配对情况，逐段生成栅格数据。直到全部扫描线都完成从矢量向栅格的转换为止。



**基于多边形数据的栅格化方法**

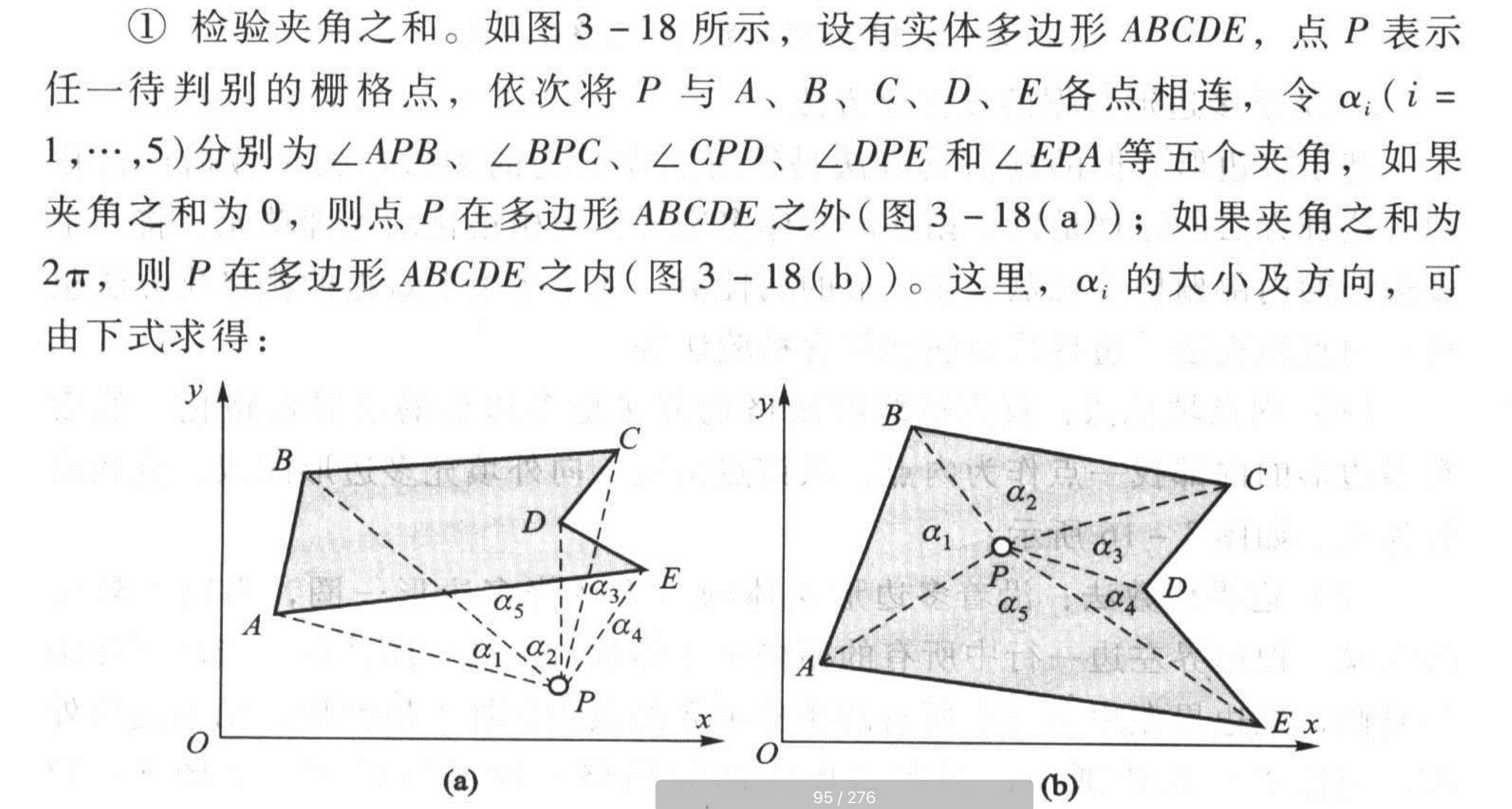
基于多边形数据的栅格化方法是针对**实体结构的多边形**矢量数据栅格化的。这种方法的原理是以非拓扑的实体多边形作为栅格化的处理单元，将一个多边形的内部栅格单元赋予多边形的属性值。基于多边形数据的栅格化方法包括：**内点填充法、边界代数法和包含检验法等。**

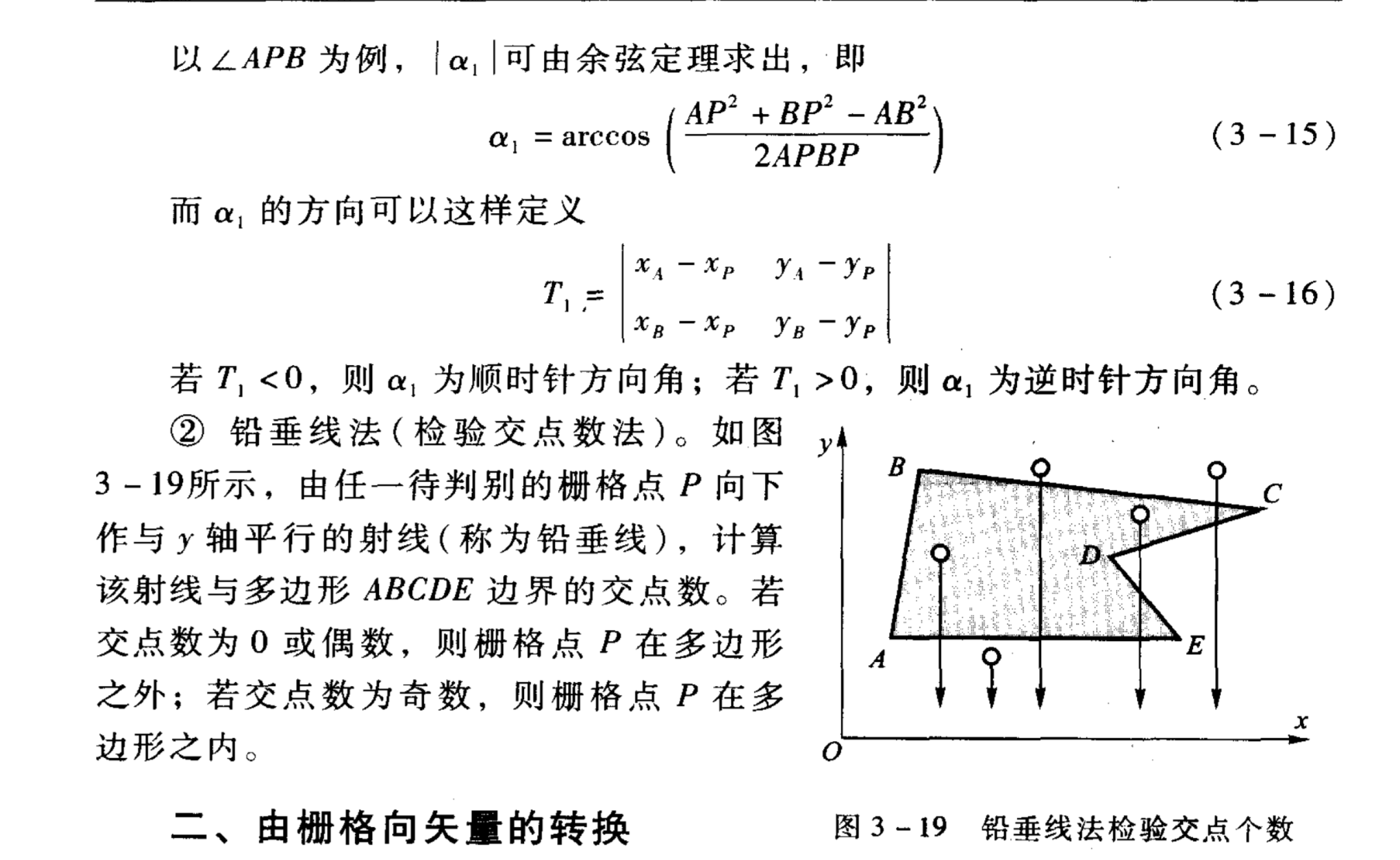
内点填充法(种子填充法)  
 首先按线的栅格化方法把多边形的边界栅格化，然后在多边形的内部找一点作为内点，从该点出发向周围4个或者是8个方向的相邻栅格点扩散填充多边形区域,直到边界为止。

边界代数法  
 围着多边形实体的边界环绕多边形一圈，当向上环绕的时候，把边界左边一行中所有的栅格单元的数值都减去属性值，当向下环绕的时候，把边界左边一行中所有的栅格单元的数值都加上属性值，则多边形外部的栅格正负数值抵消，而多边形内部的栅格被赋予属性值。

包含检验法  
 对每个栅格单元，逐个判定其是否包含在某个实体多边形之内，若包含在某个多边形之内，则将多边形的属性值赋给该单元。

* 检验夹角之和：为0在多边形内，为2则在多边形外。
* 铅垂线线法(检验交点数法)：由任意待判别的栅格点做平行或垂直线，计算其与多边形的交点数，若交点为0或者为偶数，则栅格点P在多边形之外，若为奇数，则栅格点在多边形之外。

****

****

## 四、矢量化

矢量化处理的目的

* 将自动扫描仪获取的栅格数据加入矢量形式的数据库
* 将栅格数据分析的结果，通过矢量绘图装置输出
* 为了数据压缩的需要，将大量的面状栅格数据转换为少量数据表示的多边形边界

流程

* 图像预处理
* 图形二值化
* 细化
* 追踪
* 去除多余点和线段平滑
* 建构拓扑关系

**二值图像预处理**：对于扫描输入的图幅，由于数据源等问题，会出现一些污损、分辨率过低地物识别不清等问题，对此可以依靠图像编辑等人机交互处理。用低通滤波合并破碎地物和高通滤波提取区域范围。

**二值化**：从原始扫描图计算得到黑白二值图像，二值化的关键是在灰度级的最大值和最小值之间选取一个阈值，当灰度级小于阈值时取值为 0, 当灰度级大于阈值时取值为 1。

* 当扫描图纸较清晰时，可以采用人工设置的方法。
* 当扫描图不清晰时，需由灰度级直方图来确定阈值。

**细化**：将二值图像像元线划图的线划宽度细化为一个像元的骨架图形。细化后的图形骨架既保留了原图形的绝大部分特征，又便于下一步的跟踪处理。

**追踪**：细化后的二值图像形成了二值线划图，追踪就是把线划图转换为矢量数据。其基本步骤为：对每个边界弧段由一个结点向另一个结点搜索，通常对每个已知边界点需要从8个方向进行搜索下一个边界点，直至连接成边界弧段。

**去除多余点及曲线圆滑**：由于追踪是逐个栅格进行的，必须去除由此造成的多余点记录，以减少数据冗余。可采用线性迭代法、样条函数插值法等进行平滑处理。

**建立拓扑关系**：对于矢量表示的边界弧段数据，判断其与原图上各多边形的空间关系，以形成完整的拓扑结构并建立与属性数据的联系。

ArcGIS中矢量化步骤

**第一步:地图导入与配准**

获取原始底图(无空间坐标信息) 可以是扫描纸质地图或者原始遥感影像

* 在ArcMap中，利用【定义投影】工具，定义数据框标的地理坐标系统与投影坐标系统与底图保持一致。
* 导入原始地图，利用【Georeferencing】工具对原始地图进行地理配准，一般如地形图选取经纬网交汇处为控制点，选取四个以上的控制点，检验RMS是否符合限差，一般在1以内即可。选取几何变换与重采样方法如仿射变换与双线性插值等，进行地理配准。

**第二步:准备矢量数据文件**

* 利用ArcCatalog新建矢量数据文件(空白文件)，选择Shapefile或Geodatabase等ArcGIS所支持的数据格式，创建要素类型(点、线、多边形) ，定义这些数据的坐标系与底图保持一致。
* 在ArcMap或ArcCatalog中,为矢量图层的属性表添加需要的字段
* 建立不同数据图层，如建立等高线图层，建筑物图层，高程点图层等。

**第三步:矢量化**

加载需要编辑的要素类(矢量数据)及参考用的底图，在ArcMap中加载【Editor Toolbar】,数字化工具条支持新增要素、要素的移动、旋转和删除、要素的分割与合并、要素形状编辑等，对点、线、面数据分别进行矢量化操作。

2014 论述题

已有一张我国某区域的1：5万纸质地形图，现需要从地形图上数字化高程点(点)、等高线（线）以及居民地（多边形）等数据，要求数字化得到的矢量数据的坐标系统为高斯-克吕格投影坐标系统，请叙述实现过程。