计算方法实习题二

161220072， 李颖

1. **实习任务**

本次作业的主要任务是用最小二乘法拟合室内定位系统中的轨迹。给出一组TDOA数据，每一条数据对应一个时刻待定位点到各基站的距离差信息，可以计算出一个坐标，随后根据这些点的分布情况选择合适的直线或曲线函数进行最小二乘逼近，形成运动轨迹。

1. **处理过程**

**1、由TDOA数据计算坐标**

（1）Chan算法

基于所测得TDOA，可以建立定位方程组



其中，



(，)为基站坐标，（x, y）为待测移动体坐标，为移动体与基站之间的距离，N为参与定位的基站数目，c为光速，为测得的服务基站与第i个基站之间的信号到达时间差。在几何上，每个方程表现出一条双曲线，由于存在开方运算，该方程组是非线性的。求解非线性方程组等价于无约束非线性最优化问题，首先必须对其进行线性处理。

对非线性方程组（1）进行处理，因为

 （3）

由式（2）

 （4）

其中，将（4）代入（3）得

 （5）

由式（4），当i = 1时

 （6）

将（6）代入（5）得

 （7）

如果采用测距模块的话，未知数只有两个：x和y，二元线性方程组很方便求解。如果不是，将x，y，看成未知数，仍为线性方程组。令，并假设x，y，线性无关，在存在TDOA观测误差的实际情况下，通常用最小二乘法或者加权最小二乘法对方程组求解。

当基站数超过3个时，TDOA得到的线性方程组个数要多于未知变量的个数，Chan算法采用加权最小二乘方法得到初始解，在用第一次得到的初始解等约束变量进行第二次加权最小二乘估计，最后得到改进的位置估计。

题目中**直接给出**了Chan算法得到的位置坐标，散点图如下所示。

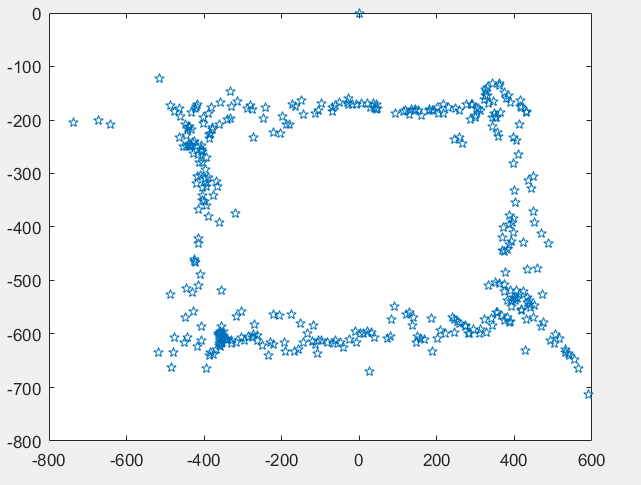


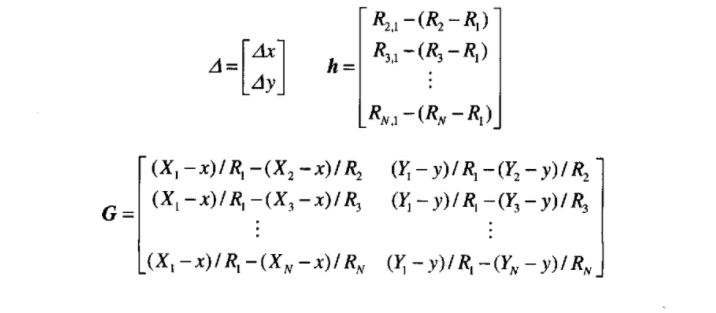
图 1 Chan算法得到的定位散点图

1. 基于Chan算法确定初值的Taylor级数展开法

将非线性方程组转换为线性方程组，可以用Taylor展开的方法。于是，对方程组（1）进行Taylor级数展开并忽略二阶以上分量，得到

 （8）

其中，



Taylor级数展开法通过递归来求解TDOA测量误差的局部线性最小二乘解，并以此来改进估计位置。初始值的选取对定位结果影响很大，如果初始值选取不合适，可能导致算法不收敛。所以需要先通过某种算法，对测量数据进行定位运算，定位算法应该运算速度快且相对精确。

因而想到将Chan算法的定位结果作为Taylor展开法的初始值，因为Chan算法的定位结果反映了与真实值之间的某种近似，以此为初值也能加速Taylor的收敛。为了避免少数情况下Taylor不收敛的情况，需要综合Taylor和Chan算法的结果，可以对两者进行加权计算，也可以默认Taylor不收敛的情况下使用Chan的结果。程序中为了减轻计算的复杂度，采用后者。

将**Taylor级数展开法**应用于实验所给的TDOA数据计算坐标，得出散点图如图示。**之后的处理过程**都将**基于该散点图**完成。

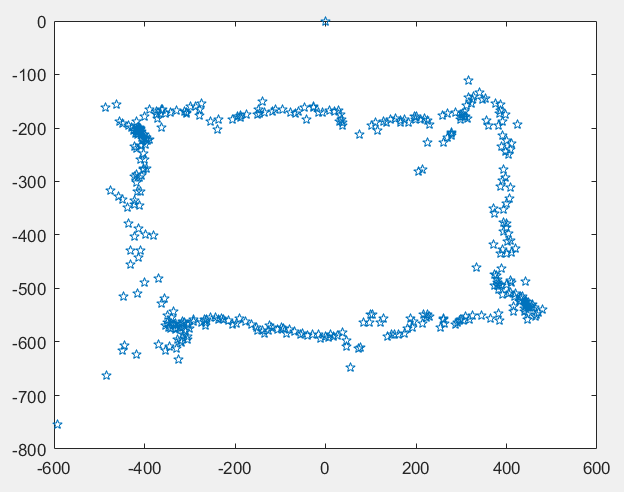


图 2 基于Chan算法确定初值的Taylor级数展开法得到的定位散点图

1. **去除异常点**
2. 基于数据时序性的异常点检测

不论是Chan算法还是Taylor级数展开算法得到的散点图，都存在一些点与上一位置的距离异常远，在实际情况下物体不可能具有“瞬间漂移”的能力，所以完全可以认定这些点是异常点，属于观测误差很大。这些观测误差很大的点对拟合产生的影响很大，所以应当要先去除。

因为采样的数据是在连续时间内按序得到的，如果后一个点距上一点的距离超过某一阈值，那么检测到异常点。对384个坐标分别进行检测，一旦检测到异常点便去除，这里的距离为两点间的绝对距离。

如何设置阈值是一个比较关键的问题，阈值设置的好坏将直接影响异常点的个数，从而影响曲线拟合的效果。从原理上看，距离的阈值肯定不超过图中连续两点距离的最大值；从直觉上看，阈值不能过低，否则去除的异常点太多，减少了原始数据提供的信息，对曲线拟合有较大影响。

将图中连续两点距离的阈值设置为95，对384个距离均进行检测。得到图2中散点检测有7个为异常，为下图红色标记部分。

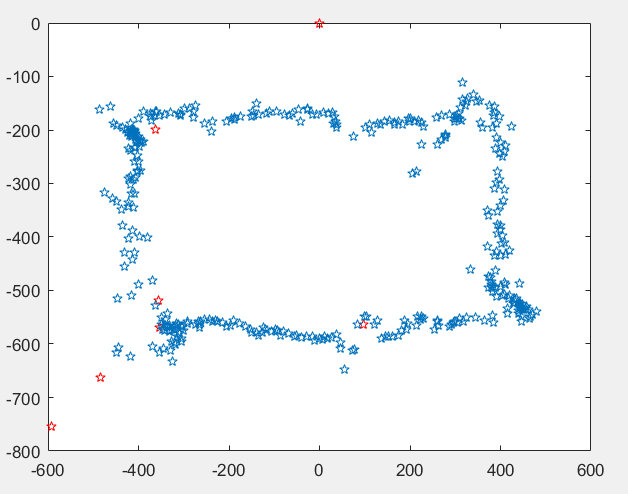


图 3 基于数据时序性的异常点检测

去除异常点后，得到了处理后的散点图如下：

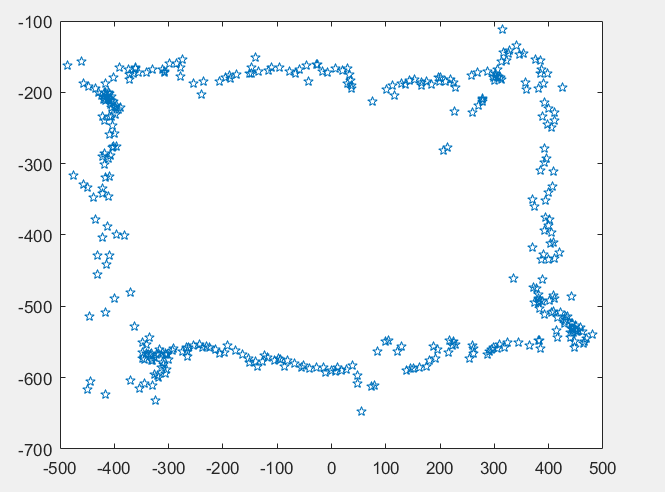


图 4 基于数据时序性异常点处理后的散点图

1. 基于邻近度的异常点检测

观测散点图，发现有一些点与其他的点距离很远，相对来说比较孤立，像离群之雁，与“大部队”不一致，这些点如果被称为“离群点”，那么显然离群点很可能是由观测误差过大而产生的。

我们可以认为那些没有足够多“邻居”的点是离群点，这里的“邻居”是用距离（绝对距离）来定义。更进一步，我们查找每个点在半径d范围内的邻居数，假设对于一个异常点来说，在d领域内最多只能有M个邻居，那么对于点x，如果发现了M+1个邻居，那么x就不是一个异常点。

由于样本总量不多，对于每个点查找邻居个数便去遍历样本总体。（，M）则是一组需要设定的阈值。确定阈值组（，M）是该问题的关键，对于多个阈值可以通过控制变量法，在确定的情况下观察每个点的邻居数得出M的大致范围。

若将设置为1e+3，并且将M设置为1时，检测到11个离群点。为下图红色标记。

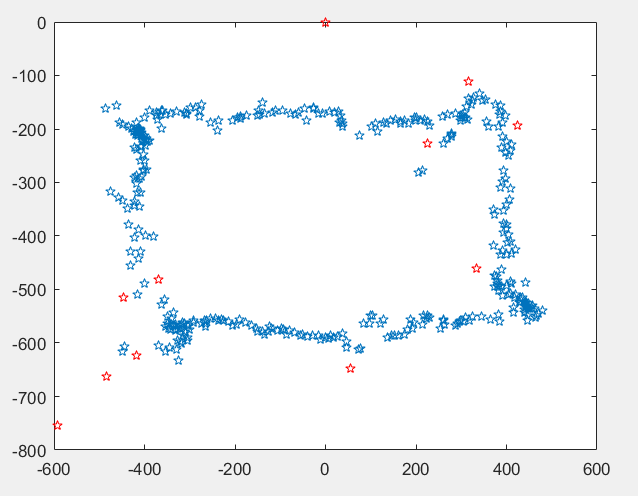


图 5 基于邻近度的异常点检测

去除异常点后，得到了处理后的散点图如下.

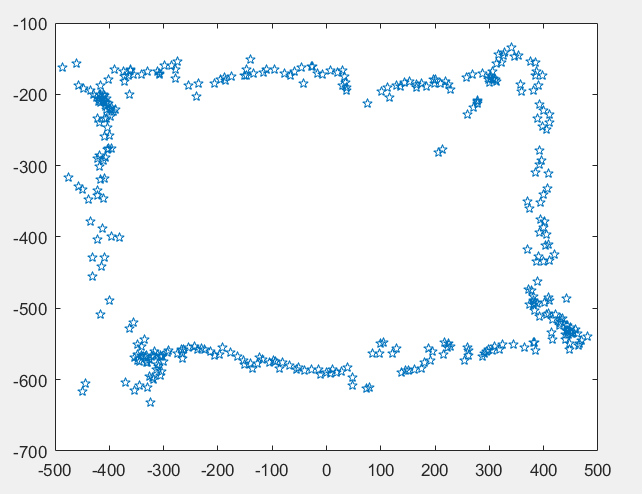


图 6 基于邻近度的异常点处理后的散点图

1. **曲线拟合**
2. 四条直线拟合

因为已知移动体的轨迹类似于矩形，最简单的方式便是用四个一次函数来拟合四条边。

根据点的分布情况将由TDOA得到的点分成四组，选择一次函数，对每组数据利用最小二乘拟合。拟合的难度在于如何将数据分组使得拟合效果最好，因为数据是在一段时间内的连续取样，所以按数据顺序进行分组是显然的选择。

观察原始的chan坐标数据，发现这385个坐标中有三个（0,0），是明显的异常点，这三个（0,0）恰好将数据分成了四部分：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Group1 | Group2 | Group3 | Group4 |
| Start | 1 | 67 | 159 | 300 |
| end | 65 | 157 | 298 | 385 |

如果对每一个部分用一个一次函数来拟合，可以得到以下的图。

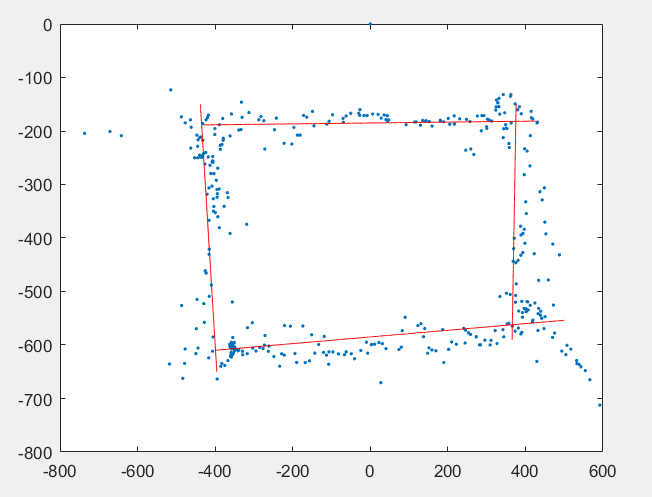


图 7 基于chan坐标（未处理坐标）的四个一次拟合

以上数据分组是一种情况，而且**比较特殊**。如果不是基于chan坐标数据拟合，而是**对基于Taylor级数展开法的坐标数据**进行拟合，则必须寻找一种策略将数据中的点分为四组，使得分别一次拟合的效果较好。有**两种**策略，一种是将所有的点分到四组内，组间无重叠点；一种是四个组的数据均来自坐标集，组间可以有重叠点。

如果采用第二种策略，采用一种较为合理的分组方式，得到的初步拟合曲线如图所示。

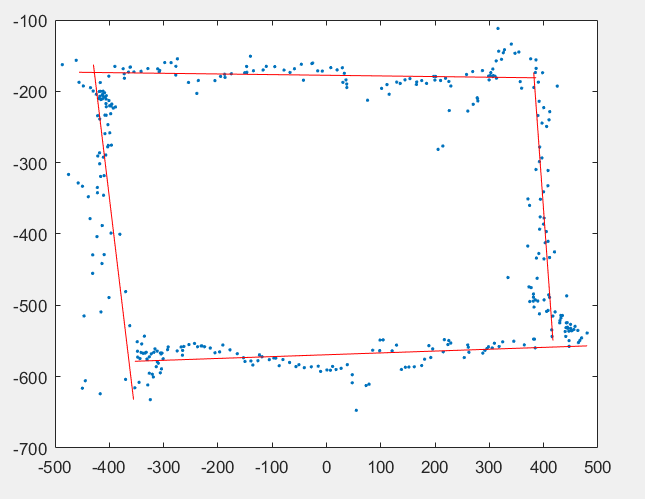


图 8 初步拟合（四条直线）

1. 弯角处理

从初步拟合的结果可以看出用四条直线拟合的效果能表现出轨迹的大概，但对于一些明显弯曲的部分，直线拟合效果并不好，尤其是在轨迹的四个明显的拐弯处，往往是有一些弧度，轨迹的变化非像直线相交那样，而是比较平缓，呈现弧形。所以为了优化拟合，对弯角处需要另用复杂函数进行拟合。

比如右上拐角处呈拱形变化，可用抛物线拟合，简单地可以选取二次函数。又比如右下拐角明显有另一条与两条边走向不一致的近似直线轨迹，可以利用分段的一次函数来拟合。

具体处理过程可以见下图。

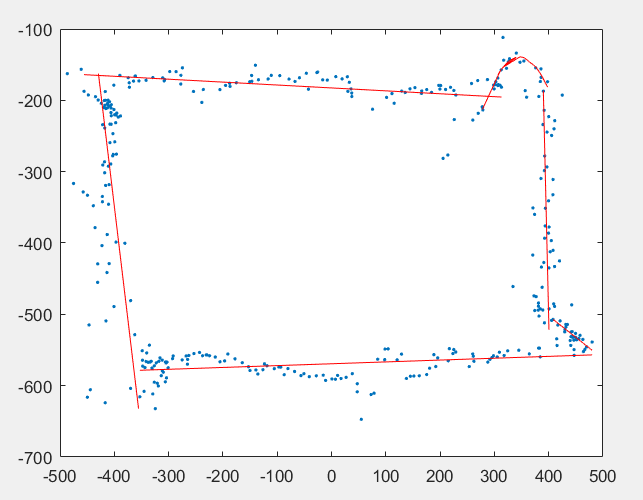


图 9 弯角处理过程图

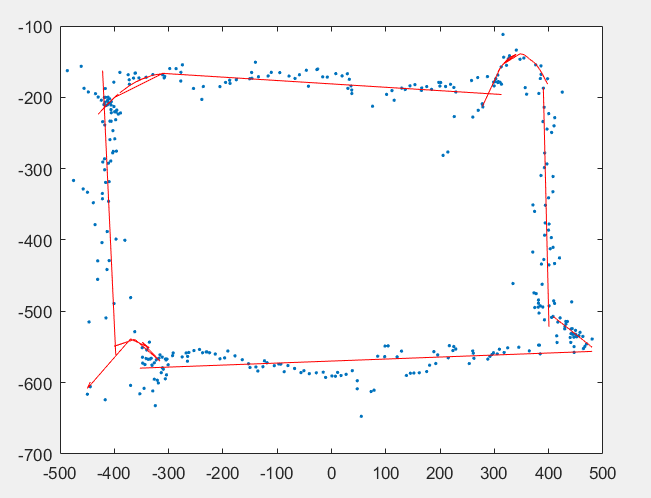


图 10 弯角处理过程图

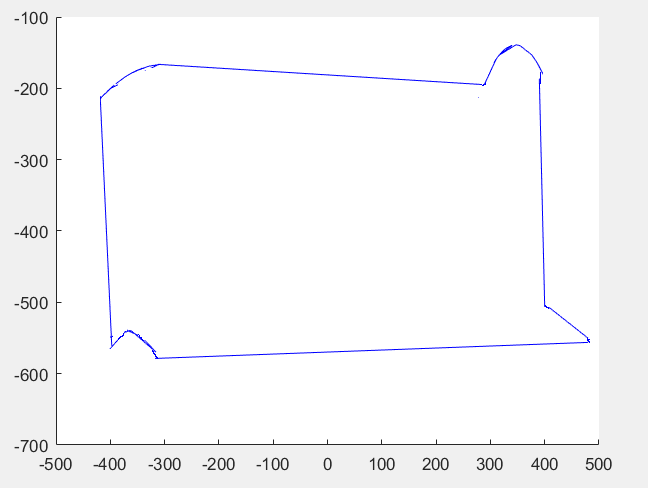


图 11 弯角处理结果图

1. 用复杂函数替换一次函数拟合的优化

观察上图，发现其实左下角比较突兀，或许这里用一段一次函数和一段二次函数效果并不好，从而想到是否可以将这一段用一个高次函数来拟合，比如二次函数和三次函数。

又例如上下两条直线其实并非是真的直线，总会有一些上下波动近似均匀的部分，那么是否可以用复杂的函数，例如正弦函数等三角函数来拟合。这一部分有待完善。

附：（TDOA数据的实际轨迹图）

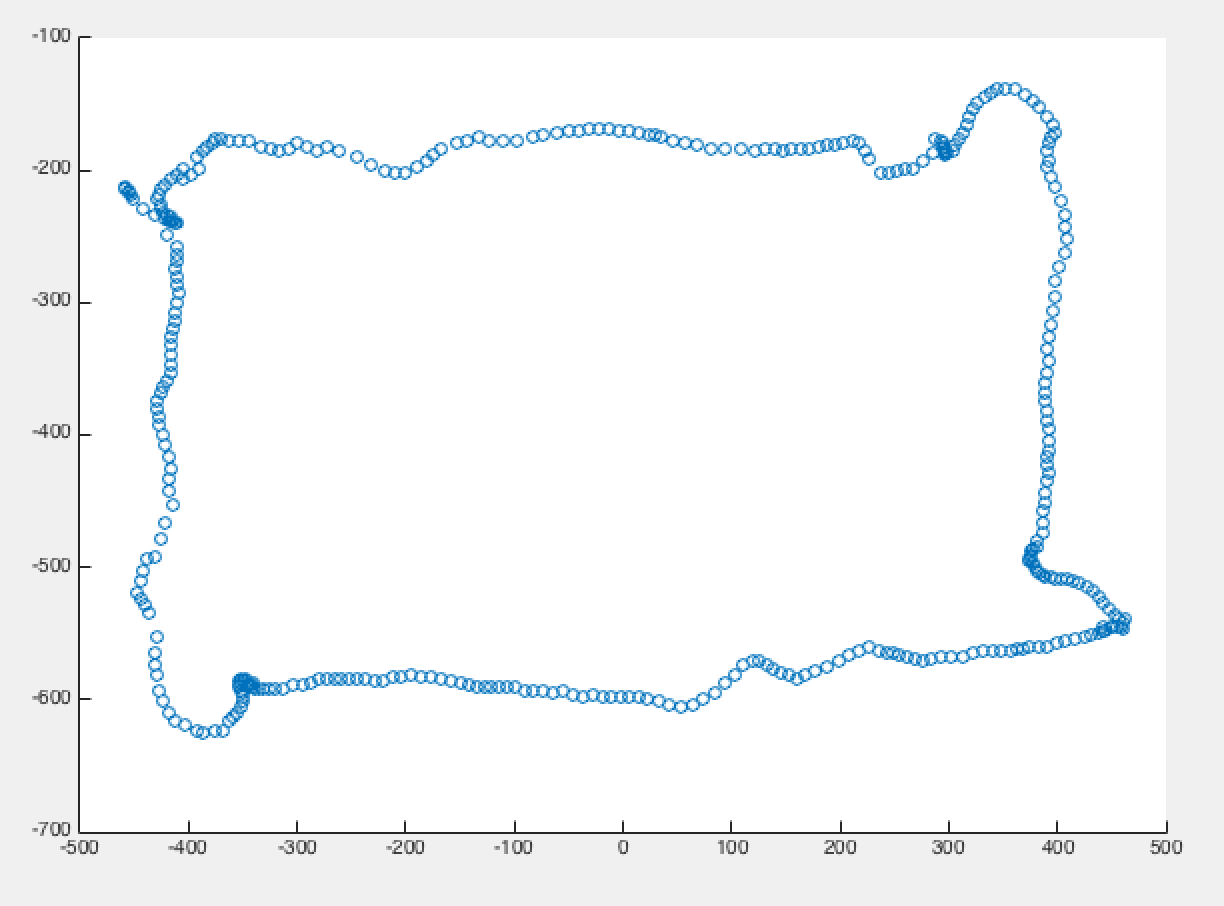


图 12 实际轨迹图