



中国研究生创新实践系列大赛
“华为杯”第十八届中国研究生
数学建模竞赛

学 校

武汉大学

参赛队号

21104860006

队员姓名

1. 谢鑫

2. 陈洲

3. 朱建桦

中国研究生创新实践系列大赛

“华为杯”第十八届中国研究生

数学建模竞赛

题 目 信号干扰下的超宽带（UWB）精确定位问题

摘 要：

本文主要对信号干扰下的超宽带（UWB）精确定位问题展开研究，在研究了数据预处理方法、TOA 定位原理、基于最小二乘法的方程组求解、遍历反演思想、约束条件下的非线性规划模型的基础上，提出了一种新的定位模型。

本文在分析相关文献的基础上，首先对原始数据进行了预处理，在得到有效数据后，基于最小二乘法的思想求解坐标方程组，得到靶点的计算位置坐标。然后分析有信号干扰和无信号干扰条件下，靶点与锚点的测量距离和真实距离的误差分布，对误差进行补偿，接着反演出新的位置坐标，输出误差最小的一组位置坐标。最后，基于改进后的非线性规划模型，能有效分辨测量结果是否存在信号干扰。由于新模型的主要采用了遍历可能解、反演初始条件和误差修正的步骤，因此本文将新模型称为“遍历-反演-自适应定位模型”。

对于问题一，清洗数据包括以下几类，超过物理边界的异常数据、含有粗大误差值的坏值以及相同或相似数据，对于异常数据和坏值依据了拉依达准则进行删除，相同或相似数据采用了 K-均值聚类来保留其中之一。

对于问题二，对正常数据分析，针对真实距离和测量距离的线性关系，将 xoz 面上分成 4×9 个方格来作拟合，使得测量距离更贴近真实距离；对异常数据分析，存在两类干扰情况，一类对距离的相应类似阶跃信号，另一类类似波动信号，均对距离有加性的作用。而后建立了“遍历-反演-自适应定位模型”，基于该模型利用问题一处理后的数据求解出靶点位置坐标，与真实值相比误差较小，最终 3 个直角坐标的平均绝对误差可以达到 46.23mm、42.26mm 及 121.14mm。

对于问题三，发现实验场景 2 相对实验场景 1 锚点 z 方向下降 100mm，空间 y 方向减小 2000mm，并不影响问题二建立的拟合关系，故直接套用问题二建立的模型，对实验场景 2 的靶点坐标进行求解。

对于问题四，本文分析了干扰的特征，同一时间只对一个锚点产生影响。建立了 4 组约束条件下的非线性规划模型，决策变量每次只选取其中 3 个锚点与靶点的距离，使用序列二次规划算法求解，由于干扰和无干扰情况下坐标解具有差异性，因此利用附件一给定的正常数据，采取 MSE 量化差异，从而可以判断数据是否为干扰数据。

对于问题五，本文利用上述的模型求解出靶点运动时的各个位置，针对 z 方向 1 波动较大的问题，利用了物体运动的规律，对 z 进行了平滑处理，拟合曲线与计算结果贴近，实现了近似最优的精确定位。

关键字：TOA 定位；厘米级定位；最小二乘法；非线性规划；

目 录

一、问题重述	3
二、模型假设与说符号说明	4
三、问题一 原始数据预处理方法研究	4
3.1 问题分析	4
3.2.数据清洗预处理	5
3.2.1 处理超出边界的异常数据	5
3.2.2 处理含有粗大误差值的坏值	7
3.2.3 处理相同或相似数据	8
3.3.特定数据文件处理后的结果	10
3.4.问题一总结	11
四、问题二：基于最小二乘思想的定位模型	11
4.1 问题的描述及分析	11
4.2 定位模型的建立	15
4.3 问题二模型的求解	17
4.4 对 z 坐标修正	18
4.5 结果及分析	20
五、问题三：不同场景下定位模型的应用	21
5.1 问题的描述及分析	21
5.2 模型的建立和求解	22
六、问题四：基于含约束非线性规划模型的分类算法	22
6.1 问题分析	22
6.2 模型准备	24
6.3 基于非线性规划模型的分类算法	25
6.4 模型求解	25
6.5 模型验证	27
七、问题五：运动轨迹定位问题求解	28
7.1 问题分析	28
7.2 问题求解	29
7.2.1 靶点运动规律分析	29
7.2.1 根据运动规律进行拟合	30
7.1 模型的评价	32
7.1.1 优点	32
7.1.2 缺点	32
八 参考文献	32

一、问题重述

近年来，随着物联网技术不断发展，对高精度室内定位技术的需求不断增加。但是传统的测距定位方法，如蓝牙、超声波、GPS 等有一些局限性。在新方法中，UWB（Ultra-Wideband）技术凭借无载波，低功率，高精度成为被广泛使用的技术。

UWB 的定位技术有多种方法，本文仅考虑基于飞行时间（Time of Flight, TOF）的测距原理，它是 UWB 定位法中最常见的定位方法之一。TOF 测距技术属于双向测距技术，其通过计算信号在两个模块的飞行时间，再乘以光速求出两个模块之间的距离，这个距离肯定有不同程度的误差，但其精度已经比较高。在室内定位的应用中，UWB 技术可以实现厘米级的定位精度（一般指 2 维平面定位），并具有良好的抗多径干扰和衰弱的性能以及具有较强的穿透能力。但由于室内环境复杂多变 UWB 通信信号极易受到遮挡，虽然 UWB 技术具有穿透能力，但仍然会产生误差，在较强干扰时，数据会发生异常波动（通常是时间延时），基本无法完成室内定位，甚至会造成严重事故。

请解决下面的问题：

（1）完成数据的抓取和预处理，将 Tag 在同一坐标上采集的多组数据统计出来，并用二维表二维表（矩阵）形式（txt、Excel 或其他数据格式），每一行代表一组数据（即一个样品），然后对这些数据文件进行预处理（清洗），删除掉一些“无用”（异常、缺失、相同或相似）的数据（样品）。

（2）利用任务 1 处理后的数据，分别对“正常数据”和“异常数据”，设计合适的数学模型（或算法），估计（或预测）出 Tag 的精确位置，并说明你所建立的定位模型（或算法）的有效性；同时请利用你的定位模型（或算法）分别对附件 2 中提供的前 5 组（信号无干扰）数据和后 5 组（信号有干扰）数据进行精确定位（3 维坐标）；

（3）所建立的定位模型能够应用于不同场景。附件 3 中 10 组数据采集于下面实验场景 2（前 5 组数据信号无干扰，后 5 组数据信号有干扰），请分别用上述建立的定位模型，对这 10 组数据进行精确定位（3 维坐标）

（4）UWB 在采集数据时并不知道信号有无干扰，所以判断信号有无干扰是 UWB 精确定位问题的重点和难点。利用任务 1 处理后的数据，建立数学模型(或算法)，以便区分哪些数据是在信号无干扰下采集的数据，哪些数据是在信号干扰下采集的数据。并说明你所建立的分类模型(或算法)的有效性；同时请用你所建立的分类模型（或算法）判断附件 4

中提供的 10 组数据（这 10 组数据同样采集于实验场景

（5）运动轨迹定位是 UWB 重要应用之一，利用静态点的定位模型，加上靶点自身运动规律，希望给出动态靶点的运动轨迹。附件 5 是对动态靶点采集的数据（一段时间内连续采集的多组数据），请注意，在采集这些数据时，会随机出现信号干扰，请对这个运动轨迹进行精确定位，最终画出这条运动轨迹图（数据采集来自实验场景 1）

二、模型假设与说符号说明

- （1）假设样本量可以认为足够多；
- （2）同一时间可以认为仅有一个锚点被遮拦

符号	变量说明
(X_i, Y_i, Z_i)	锚点 A_i 的三维坐标
(x_i, y_i, z_i)	靶点 Tag_j 的真实三维坐标
a, b, c	靶点 (Tag) 范围
d_{ij}	锚点 A_i 与靶点 Tag_j 的理论距离
s_{ij}	锚点 A_i 与靶点 Tag_j 的正常测得距离
s'_{ij}	锚点 A_i 与靶点 Tag_j 的异常测得距离

三、问题一 原始数据预处理方法研究

3.1 问题分析

针对问题一，要求对题目所提供的的数据进行清洗预处理。首先，对原始数据进行分析：附件 1 提供的超宽带（UWB）数据集分为正常数据和异常数据两种情况，正常数据是指在无信号干扰的情况下采集的，异常数据是指在有信号干扰的情况下采集的。每种情况采集 324 份数据，同一份数据代表靶点（Tag）在同一坐标点上采集，被存储在同一个 txt 文件下。由于 Tag 在同一位置会停留一段时间，所以每份数据包含多个时刻测量的多组数据，每组数据提供了 Tag 到四个锚点（anchor，A0、A1、A2、A3）的距离数据。

由于 Tag 在每个位置测量的数据数量不同、数据文件中存在一些“无用”（异常、缺

失、相同或相似)数据,这些“无用”数据会对精确定位问题的求解造成误差,所以在后续建模分析前,需要对原始数据进行清洗预处理。

本文对上述数据的预处理包含如下过程:首先,对不满足物理边界条件和缺失的异常数据进行删除处理;其次,基于拉依达准则(3σ 准则)对含有粗大误差值的坏值进行清洗处理;最后,利用 K-均值聚类方法对数据进行聚类操作,对同类数据取众数,以达到剔除相同或相似数据的效果。问题一的思路流程图如图 3.1 所示:

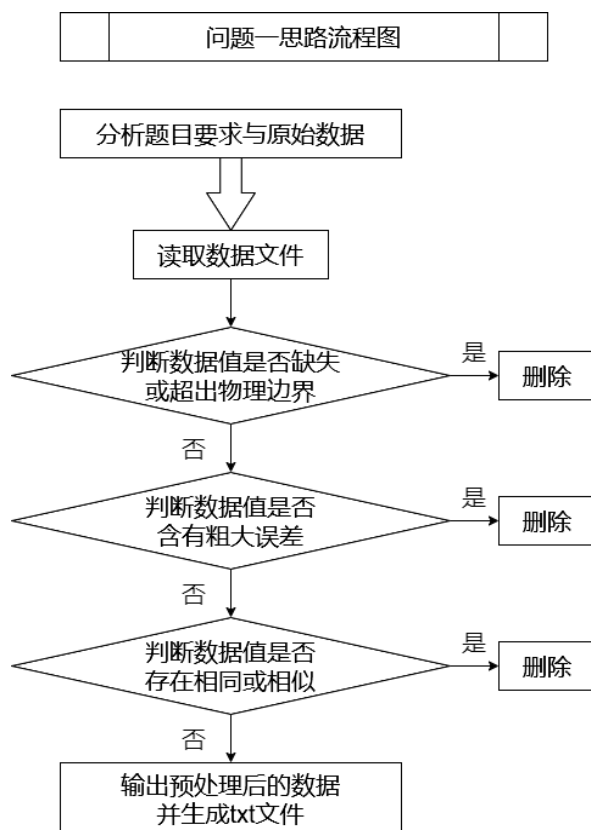


图 3.1 问题一思路流程图

3.2.数据清洗预处理

3.2.1 处理超出边界的异常数据

实测环境示意图如图 3.2 所示,实验场景 1 的位置图如图 3.3 所示。

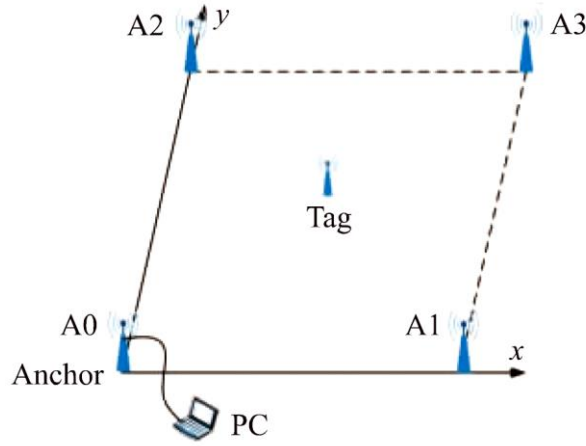


图 3.2 实测环境示意图

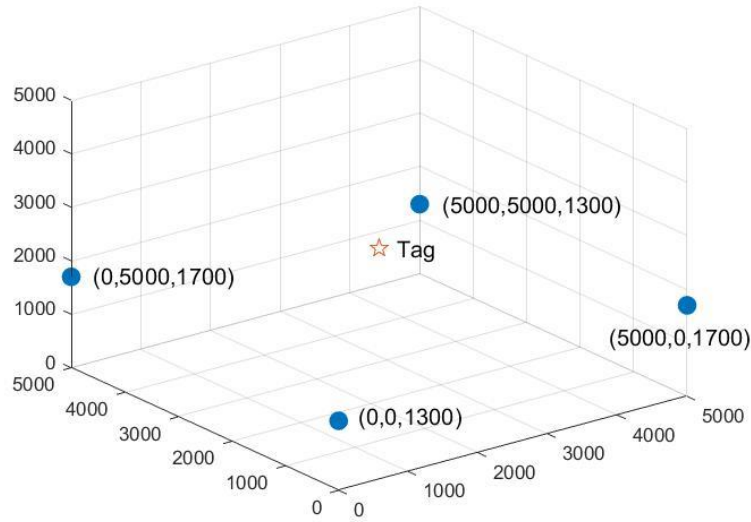


图 3.3 实验场景 1 中靶点（五角星）和锚点（圆形）的位置图

虽然靶点在测试环境内随机出现，但由于测试环境存在物理边界，所以靶点与锚点的距离存在上限 s_{max} （单位， mm ）。观测位置图不难发现锚点 A0（0，0，1300）到对位顶点（5000，5000，5000）的距离最远， s_{max} 的计算如公式 XX 所示：

$$s_{\max} = \max_{\substack{1 \leq i < 4 \\ 1 \leq j < 324}} \sqrt{(X_i - x_j)^2 + (Y_i - y_j)^2 + (Z_i - z_j)^2} \leq \sqrt{(0 - 5000)^2 + (0 - 5000)^2 + (1300 - 5000)^2} = 7980$$

对于实测数据值超过上限 s_{max} 的异常数据值应采取删除处理。另外，由于一组数据应包含靶点到四个锚点的距离信息，对于缺少某个距离信息的数据，已无法满足后续建模的需要，应采取删除处理。

3.2.2 处理含有粗大误差值的坏值

拉依达准则（ 3σ 准则）

拉依达准则又称为 3σ 准则，先假设一组数据只有随机误差，对它进行计算处理得到标准偏差，再按一定的概率确定一个范围，凡是超过了这个范围的误差，就不属于随机误差，含有该误差的数据就应该被删除掉。

在正态分布里面， σ 表示的标准差， μ 表示均值， $x = \mu$ 是图像的对称轴。 3σ 准则为：

数值分布在 $(\mu - \sigma, \mu + \sigma)$ 的概率是0.6826；

数值分布在 $(\mu - 2\sigma, \mu + 2\sigma)$ 的概率是0.9544；

数值分布在 $(\mu - 3\sigma, \mu + 3\sigma)$ 的概率是 0.9974。

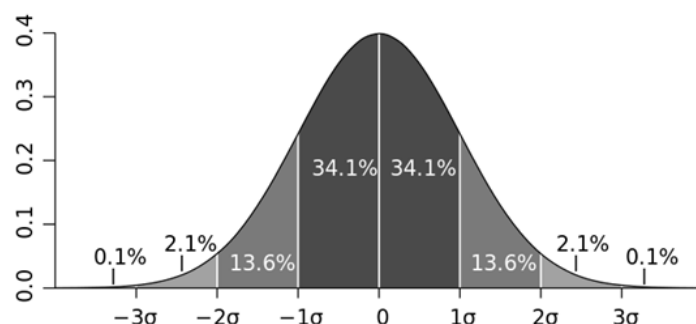


图 3.4 标准正态分布曲线

可以看出， y 的取值基本上全部集中在 $(\mu - 3\sigma, \mu + 3\sigma)$ 范围内，超出这个区间的可能性小于0.3%。

为验证拉依达准则在本文中的合理性，对附件 1 中正常情况下不同靶点测量的所有数据进行归一化处理，得到正常情况下距离测量值归一化后的分布情况，如图 3.5 所示。不难发现四个距离的测量值均满足正态分布，采用拉依达准则能有效剔除含有粗大误差值的坏值，能有效提高后续建模的准确性。

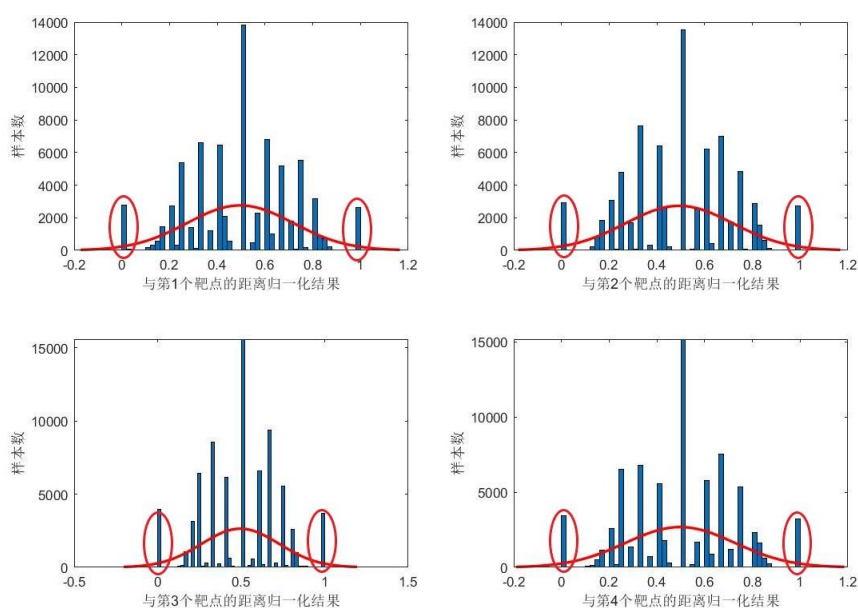


图 3.5 正常情况下距离测量值归一化后的分布情况

3.2.3 处理相同或相似数据

K-均值聚类方法

聚类是一个将数据集中在某些方面相似的数据成员进行分类组织的过程，聚类就是一种发现这种内在结构的技术，聚类技术经常被称为无监督学习。**K-均值聚类**是最著名的划分聚类算法，由于简洁和效率使得他成为所有聚类算法中最广泛使用的。给定一个数据点集合和需要的聚类数目 m ， m 由用户指定，**K-均值**算法根据某个距离函数反复把数据分入 m 个聚类中。

K-均值聚类方法的具体步骤是，预将数据分为 m 组，则随机选取 m 个对象作为初始的聚类中心，然后计算每个对象与各个种子聚类中心之间的距离，把每个对象分配给距离它最近的聚类中心。聚类中心以及分配给它们的对象就代表一个聚类。每分配一个样本，聚类的聚类中心会根据聚类中现有的对象被重新计算。这个过程将不断重复直到满足某个终止条件。终止条件可以是没有（或最小数目）对象被重新分配给不同的聚类、没有（或最小数目）聚类中心再发生变化、误差平方和局部最小。

本文处理相同或相似数据的方法是，对每份数据文件下的所有测量数据进行聚类，每类数据选取数值出现次数最多的一组数据，及实现了去除相同和相似数据的任务。由于靶点位置已知，可以计算靶点到锚点的真实距离，将真实距离和正常数据进行比对，可以发现测量距离是沿真实值上下波动，如图 3.6 所示，靶点对四个锚点分别测量得到的

距离值的分布，可以发现一般存在三种情况，接近真实值，大于真实值，小于真实值，故进行 K-均值聚类时，聚类数目取 3；

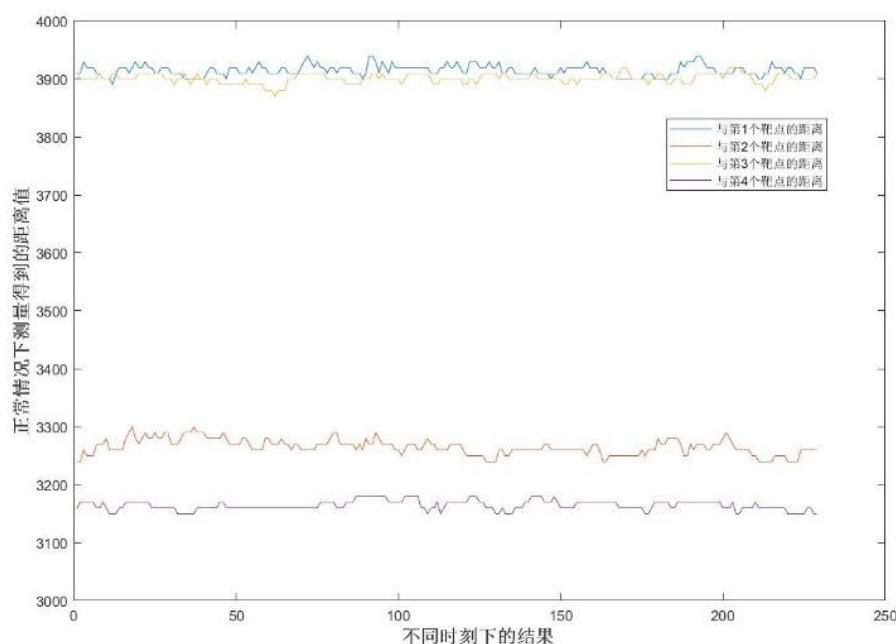


图 3.6 正常情况下测量得到的四个距离值的典型分布

在异常情况下，如图 3.7、3.8 所示，依然存在测量距离在真实值附近波动的情况，并且在某一时刻，会出现一个跳变，或者是在真实值的基础上附加一个加性噪声，结合上下波动的情况，为了能有效的筛选出相似数据，又不至于聚类数量太多，对异常情况进行 K-均值聚类时，聚类数目取 6。

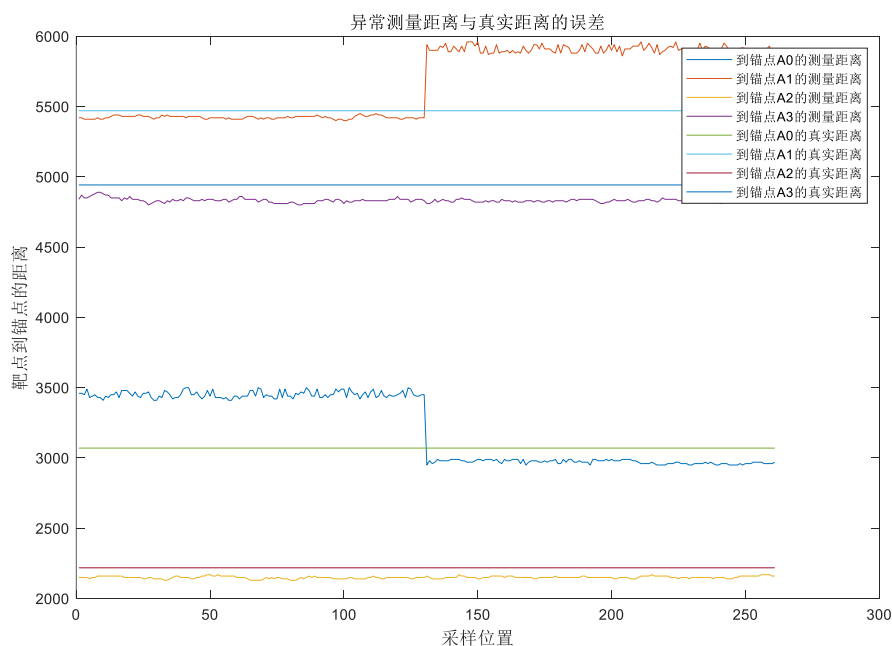


图 3.7 异常情况下测量得到的四个距离值的典型分布

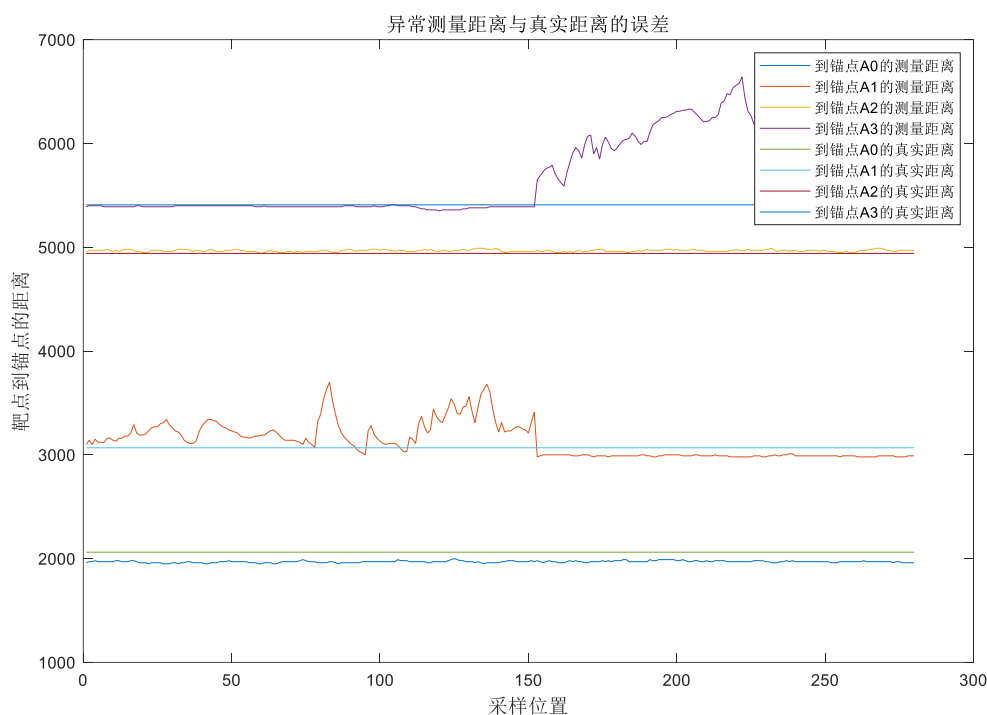


图 3.8 异常情况下测量得到的四个距离值的典型分布

3.3.特定数据文件处理后的结果

在完成上述处理超出边界的异常数据、处理含有粗大误差值的坏值和处理相同或相似数据后，正常情况下每个 Tag 测量目标保存三组数据，异常情况下每个 Tag 测量目标保存六组数据。

表 3.1 按问题一要求列出 24.正常.txt、109.正常.txt、1.异常.txt 和 100.异常.txt 数据文件经处理后保留的数据。

表 3.1 问题一结果表

文件名	s_1	s_2	s_3	s_4
24.正	3280	4660	2610	3910
常.txt	3270	4660	2600	3910
	3280	4660	2600	3900
109.正	4880	5330	2030	2880
常.txt	4890	5310	2010	2860
	4910	5340	2040	2880

文件名	s'_1	s'_2	s'_3	s'_4
1.异常.txt	760	5010	4550	6310
	1230	4540	4550	6310
	770	5030	4550	6290
	770	5010	4550	6300
	1250	4550	4560	6300
	1240	4550	4550	6300
100.异常.txt	1510	3820	4680	5700
	1510	3560	4930	5700
	1510	3740	4680	5710
	1520	3560	4930	5690
	1520	3750	4680	5710
	1510	3560	5110	5700

3.4. 总结

通过对附件 1 提供的超宽带（UWB）数据集进行分析，我们对超过物理边界的异常数据、含有粗大误差值的坏值和相同或相似数据采取了删除处理，其中删除含有粗大误差值的坏值采用了拉依达准则，删除相同或相似数据采用了 K-均值聚类的方法。得到预处理后的数据更接近理想情况，为后续建模分析打下基础。

四、问题二：基于最小二乘思想的定位模型

4.1 问题的描述及分析

任务二中，要求利用任务 1 处理完成的数据，分别对“正常数据”和“异常数据”，设计合适的数学模型（或算法），估计（或预测）出 Tag 的精确位置。并利用定位模型（或算法）分别对附件 2 中提供的前 5 组（信号无干扰）数据和后 5 组（信号有干扰）数据进行精确定位（3 维坐标）

为了提高定位精度，需要在计算中引入适当的方法削弱误差对于结果的影响，所以需

要对两种情况进行讨论，确定误差模型，从而将误差模型引入计算过程中，进而提高定位模型的准确度。

正常数据是指接收过程中没有受到信号干扰的数据。在任务一中，已经分析了正常情况下测量距离是沿真实值上下波动的。未来得到误差模型，这里做进一步的分析。以 0 号接收机为例，将采样位置作为自变量，靶点到锚点的真实距离和正常数据分别为因变量，得到图 4.1。

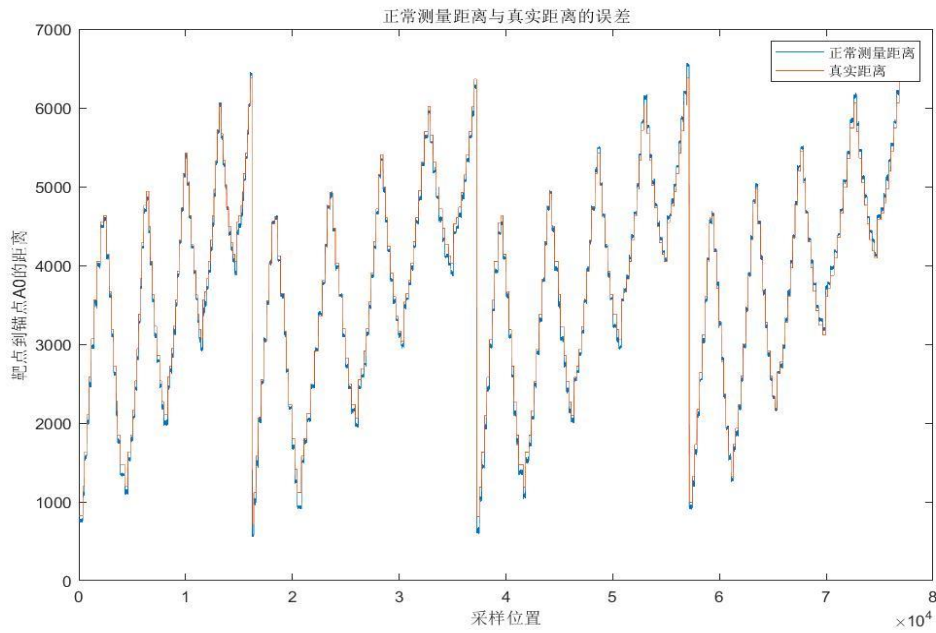


图 4.1 靶点到锚点的真实距离和正常数据示意

由图 4.1 可见，正常数据和实际距离基本符合，但是仍然存在一定的偏差。通过信道知识，可以得知这种误差主要由为钟差和环境噪声引起。而且这两类误差都符合高斯分布，属于独立同分布的误差模型。因此可以将其合并成一种均值为 μ_{wgn} ，标准差为 σ_{wgn} 的高斯分布。

针对这样的分析，我们通过将正常数据和实际距离作差，并统计误差值的样本数量，构造了频数直方图，得到了验证，如图 4-2 所示。根据已知信息对误差模型进行拟合，如图中红线所示，较为符合高斯分布的模型，进而对这个模型进行非参数假设检验，经过计算其符合高斯分布。因此可以认为在正常数据测量中，即无信号干扰下的误差模型符合高斯分布。

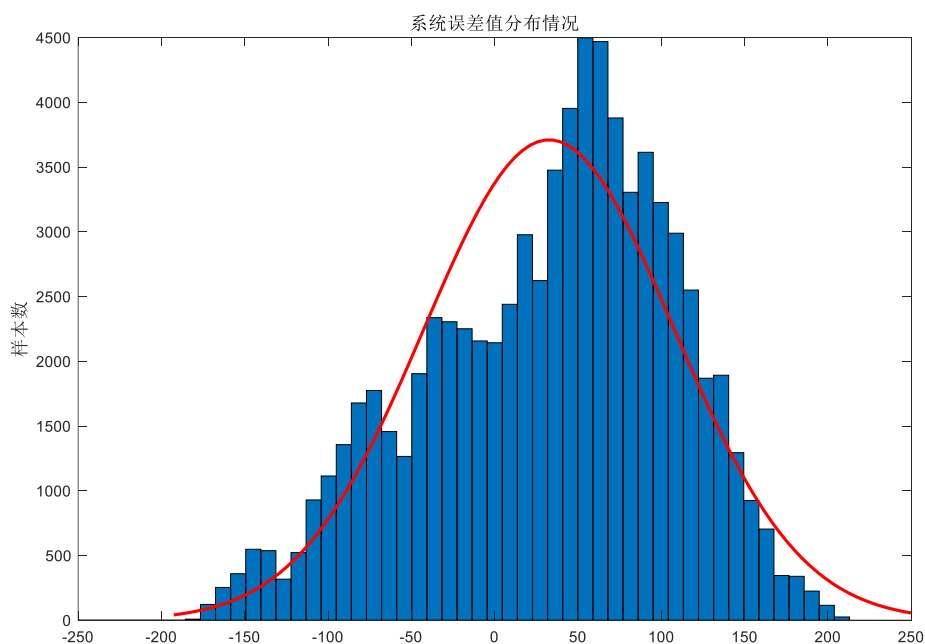


图 4-2 正常数据的误差频率直方图

在有信号干扰的情况下，情况就变得复杂起来，因为除了有上述的系统误差，同时还可能有其他因素造成的误差。调查相关信道文献可知，这种误差可能有多径效应，天线的方向，人体运动遮挡等因素，将产生很多随机误差。任务 1 中已经对异常数据作了初步分析，可以发现其中干扰主要分为两类。一类是阶跃噪声，一类是随机信号噪声，并且在同一时刻，仅有一个锚点会受到噪声的影响，其余三个锚点都是处于正常工作状态。

下面将进一步分析有信号干扰下的误差模型，与分析无信号干扰的模型类似，通过分析异常数据，可以发现数据中既包含系统误差信息也包含信号干扰信息。选取其中的无干扰数据，其误差频数直方图与图 4.2 一致。

考虑测量距离与真实距离之间的相关性，二者符合线性相关，但是并不相等，依然存在约 100mm 的偏差，为机器自身的测量偏差，故考虑将二者拟合成直线，减小后续计算误差。对其进行拟合操作得到两者间的关系是：

$$y = Kx + B$$

数据采集时 x , y , z 的变动步长分别为 9,9,4, $9*9*4=324$, 为总采集数据数量，并且满足 x , z 不变, y 坐标变动的规律，为得到更好地拟合效果，利于后续模型在有信号干扰的情况下的误差模型的调整和修正，将对 xz 平面切为 36 个小块，每个小块均进行一组拟合。为了验证拟合的准确性和有效性，本文也选取所有异常数据进行验证，如图 4.3

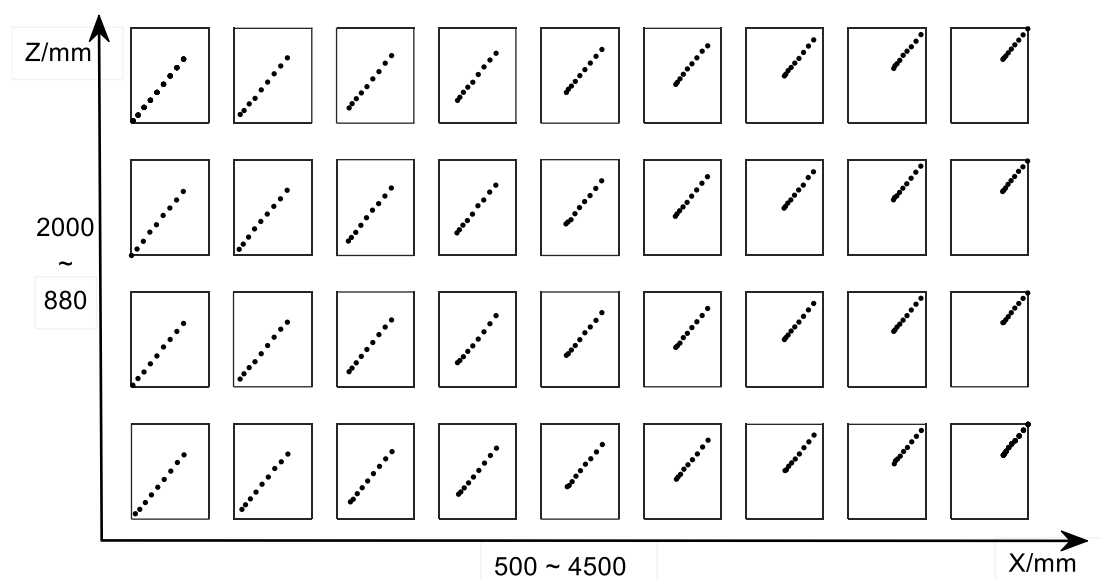


图 4.3 xoz 上是 4*9 方格单独拟合

从图 4.3 中可以看出，在所有的场景切面中，靶点和锚点之间的测量距离与真实距离成非常好的线性关系，每个切面的系数 K 和 B 由表 4.1，4.2 可见

表 4.1 拟合系数 k

Z\X	1	2	3	3	5	6	7	8	9
1	0.9899	1.0015	0.9619	0.9515	0.9473	1.0102	0.9913	1.0061	0.9815
2	0.9573	1.0049	0.9786	1.0104	0.9986	0.9874	0.9923	0.9958	0.9709
3	0.9951	0.9920	0.9519	0.9627	0.9687	0.9973	0.9730	0.9806	0.9634
4	0.9862	1.0024	1.0232	0.9740	1.0303	0.9952	0.9605	0.9834	0.9780

表 4.2 拟合系数 b

Z\X	1	2	3	3	5	6	7	8	9
1	109.58	68.42	233.06	257.68	272.41	20.20	46.89	33.95	70.52
2	214.65	-30.78	131.94	52.71	101.88	160.23	138.05	130.88	106.61
3	74.39	107.04	220.95	188.89	175.59	41.08	134.17	161.16	211.21
4	108.64	-41.53	-126.3	140.32	-4.70	131.27	232.86	134.36	138.70

综上所述，可以建立误差模型：

无信号干扰条件:

$$s_{ij} = d_{ij} + \gamma_i^{(wgn)}$$

有信号干扰条件:

$$s'_{ij} = d_{ij} + \gamma_i^{(wgn)} + \gamma_{ij}^{(delay)}$$

其中, $\gamma_i^{(wgn)}$ 为锚点误差, 即系统误差, 且符合高斯分布, $\gamma_{ij}^{(delay)}$ 是室内信号干扰带来的误差, 且在每个场景切面上都和测量距离成线性关系。

4.2 定位模型的建立

TOF 数据测量的准确性是后续精确定位重要前提, 故需对已测的数据进行修正。通过上文分析得出的误差模型, 来对 TOF 数据进行适当的补偿修正。

设第 j 个靶点 Tag_j 的坐标为 (x_j, y_j, z_j)

第 i 个锚点 A_i 的坐标为 (X_i, Y_i, Z_i)

显然, 靶点 Tag_j 到锚点 A_i 的实际距离 d_{ij} 为:

$$d_{ij} = \sqrt{(x_j - X_i)^2 + (y_j - Y_i)^2 + (z_j - Z_i)^2}$$

结合式得到等式

$$\sqrt{(x_j - X_i)^2 + (y_j - Y_i)^2 + (z_j - Z_i)^2} = s'_{ij} - \gamma_i^{(wgn)} + \gamma_{ij}^{(delay)}$$

在没有信号干扰的情况下, $\gamma_{ij}^{(delay)} = 0$, 等式可以简写成

$$\sqrt{(x_j - X_i)^2 + (y_j - Y_i)^2 + (z_j - Z_i)^2} = s_{ij} - \gamma_i^{(wgn)}$$

从上述关系式可以看出, 实际距离与测量距离中往往会因为测量系统误差及信号干扰的影响而产生差距。而精确定位的主要目的即要满足预测终端位置与实际终端位置间的差距最小, 故可根据该目标建立非线性终端定位规划模型

决策变量: 可决定非线性规划方程中的决策变量如下

- [1] 靶点 Tag_j 的坐标为 (x_j, y_j, z_j) ;
- [2] 系统误差 $\gamma_i^{(wgn)}$;
- [3] 干扰信号带来的误差 $\gamma_{ij}^{(delay)}$;

目标函数: 已知靶点 Tag_j 到锚点 A_i 的实际距离 d_{ij} 为

$$d_{ij} = \sqrt{(x_j - X_i)^2 + (y_j - Y_i)^2 + (z_j - Z_i)^2}$$

而锚点到靶点的测量距离为

$$\hat{s}_{ij} = \bar{s}_{ij} + \gamma_{ij}$$

精确定位的目标即为使得测量距离与实际距离的差距最小，即

$$\min \sum_{i=1}^m \sum_{j=1}^n (\bar{s}_{ij} - d_{ij})$$

又考虑到迭代过程的复杂性，为方便后面计算和迭代，避免迭代过程中出现 $d_{ij} > \bar{s}_{ij}$ 引起发散的情况，可将目标函数改为

$$\min \sum_{i=1}^m \sum_{j=1}^n (\bar{s}_{ij} - d_{ij})^2$$

约束条件：在测量模型中建立了测试实际距离与 TOF 测量距离的关系表达式。由于实验场景中存在的信号干扰和可能存在的系统误差，使得测量出的距离数值上要大于靶点与锚点之间的实际距离，即

$$\sqrt{(x_j - X_i)^2 + (y_j - Y_i)^2 + (z_j - Z_i)^2} \leq s_{ij}$$

当且仅当无信号干扰且无系统误差时等号成立。

即主要的约束条件为

$$d_{ij} \leq s_{ij}$$

同时应满足的约束条件为

$$\begin{cases} 0 \leq x_j \leq 5000 \\ 0 \leq y_j \leq 5000 \\ 0 \leq z_j \leq 3000 \end{cases}$$

综合上述条件，可得定位的非线性规划方程组为：

$$\begin{aligned} & \min \sum_{i=1}^m \sum_{j=1}^n (\bar{s}_{ij} - d_{ij})^2 \\ & \begin{cases} d_{ij} \leq s_{ij} \\ s_{ij} = d_{ij} + \gamma_i^{(wgn)} \\ s'_{ij} = d_{ij} + \gamma_i^{(wgn)} + \gamma_{ij}^{(delay)} \\ 0 \leq x_j \leq 5000 \\ 0 \leq y_j \leq 5000 \\ 0 \leq z_j \leq 3000 \\ d_{ij} = \sqrt{(x_j - X_i)^2 + (y_j - Y_i)^2 + (z_j - Z_i)^2} \end{cases} \end{aligned}$$

4.3 问题二模型的求解

通过前文分析可知，误差已经可以线性化，为了计算方便，下面将方程组也实现线性化。锚点的空间位置为 (x_j, y_j, z_j) ，则由 TOF 计算从靶点到四个锚点的距离，分别为 d_{0j} 、 d_{1j} 、 d_{2j} 和 d_{3j} 。然后，根据以下计算方法，可以计算出靶点的位置坐标。根据已知条件列出方程

$$\begin{cases} (x_j - X_0)^2 + (y_j - Y_0)^2 + (z_j - Z_0)^2 = d_{0j}^2 \\ (x_j - X_1)^2 + (y_j - Y_1)^2 + (z_j - Z_1)^2 = d_{1j}^2 \\ (x_j - X_2)^2 + (y_j - Y_2)^2 + (z_j - Z_2)^2 = d_{2j}^2 \\ (x_j - X_3)^2 + (y_j - Y_3)^2 + (z_j - Z_3)^2 = d_{3j}^2 \end{cases}$$

显然，我们可以把方程通过作差简化，变成线性方程

$$\begin{cases} 2(X_0 - X_1)x + 2(Y_0 - Y_1)y + 2(Z_0 - Z_1)z = P_0 \\ 2(X_0 - X_2)x + 2(Y_0 - Y_2)y + 2(Z_0 - Z_2)z = P_1 \\ 2(X_0 - X_3)x + 2(Y_0 - Y_3)y + 2(Z_0 - Z_3)z = P_2 \end{cases}$$

其中： $P_n = d_{(n+1)j}^2 - d_{0j}^2 - X_{(n+1)}^2 + X_0^2 - Y_{(n+1)}^2 + Y_0^2 - Z_{(n+1)}^2 + Z_0^2$

写成向量形式为

$$2 \begin{bmatrix} X_0 - X_1 & Y_0 - Y_1 & Z_0 - Z_1 \\ X_0 - X_2 & Y_0 - Y_2 & Z_0 - Z_2 \\ X_0 - X_3 & Y_0 - Y_3 & Z_0 - Z_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} P_0 \\ P_1 \\ P_2 \end{bmatrix}$$

可以写成 $AX = B$ 的形式，其中

$$A = 2 \begin{bmatrix} X_0 - X_1 & Y_0 - Y_1 & Z_0 - Z_1 \\ X_0 - X_2 & Y_0 - Y_2 & Z_0 - Z_2 \\ X_0 - X_3 & Y_0 - Y_3 & Z_0 - Z_3 \end{bmatrix}, \quad B = \begin{bmatrix} P_0 \\ P_1 \\ P_2 \end{bmatrix}$$

由于存在测距误差，所以合理的线性模型应该是

$$AX + N = B$$

其中， N 是3维的随机误差向量，根据最小误差原理， X 的值应当误差 $N = B - AX$ 达到最小，即用最小化 $Q(X) = \|B - AX\|^2$ 求 X 的估计，对 $Q(X)$ 关于 X 求导并令其等于0，就可以得到 X 的最小二乘估计：

$$X = (A^T A)^{-1} A^T B$$

从而得到靶点坐标。

由于接收机是在一个靶点位置多次采样，所以需要考虑多次采样取平均，即在问题一的结果基础上，求得平均值。但是正如前文所考虑到的，真实距离 d_{ij} 和测量距离 s_{ij} 之间是有误差的，而这会影响定位的精度。基于最小二乘法可以提高定位的准确性，可以发现， B 的微小变化，对结果影响非常大。所以当采集到异常数据时，系统定位精度也会大大降低，这在实际测量中，是不希望遇到的情况。

所以通过对误差模型的建立和求解，将测量距离进行一定程度的补偿，就可以比较合适的解决这个问题了。对于正常数据来说，算法比较简单，就是减去高斯分布的均值，就可以视为补偿，并带入计算了。而异常数据则相对复杂，将异常数据的算法思路流程图，如图 4.4 所示

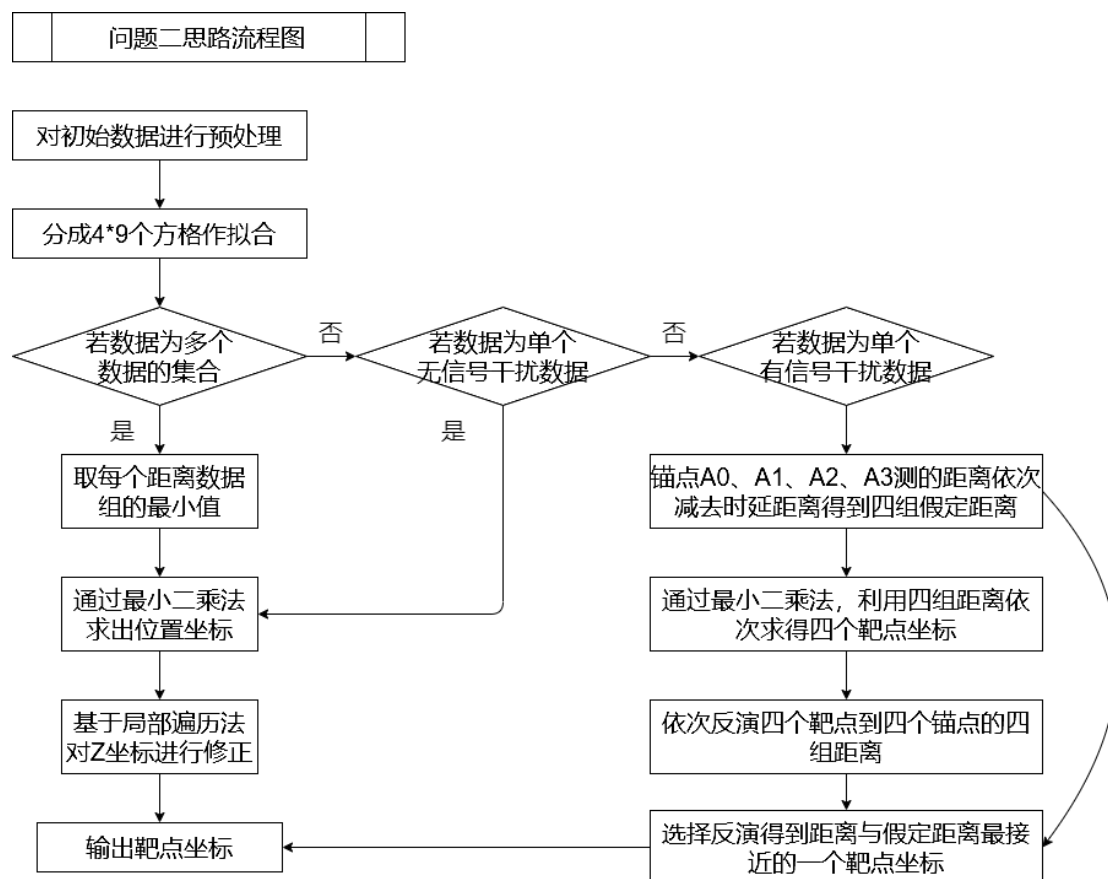


图 4.4 问题二思路流程图

4.4 对 z 坐标修正

当定位出的靶点的竖坐标 z 为负值时，显然不符合实际意义，有必要对其进行修正，考虑到前面对测试数据进行误差分析时发现，对靶点坐标的定位结果中，横坐标 x 和纵坐标 y 对误差结果不敏感，并且和靶点真实坐标对比，也可以发现误差非常小，因为可以认

为对横纵坐标的定位是准确的，只是高度 z 相对不准确。那就说明了正确的定位点应该在过点 $(x,y,0)$ 且垂直地面向上的射线上。因此可以在 $|z|$ 的附近区间做局部搜索，以期找到一个合适的值作为 z 的修正值。实际求解中，从 $|z|$ 开始，以 $\Delta z=1$ 为步长，分别向上，和向下进行 1000 次搜索，从而得到修正值 \tilde{z} ，使得以 (x, y, \tilde{z}) 作为靶点位置是所有遍历过程中的靶点位置与锚点距离跟测量计算的距离的总误差最小。

此时通过问题 1 中的测量数据，即正常数据和异常数据所求得的三维坐标和已知靶点的真实坐标作比对，作出图 4.3

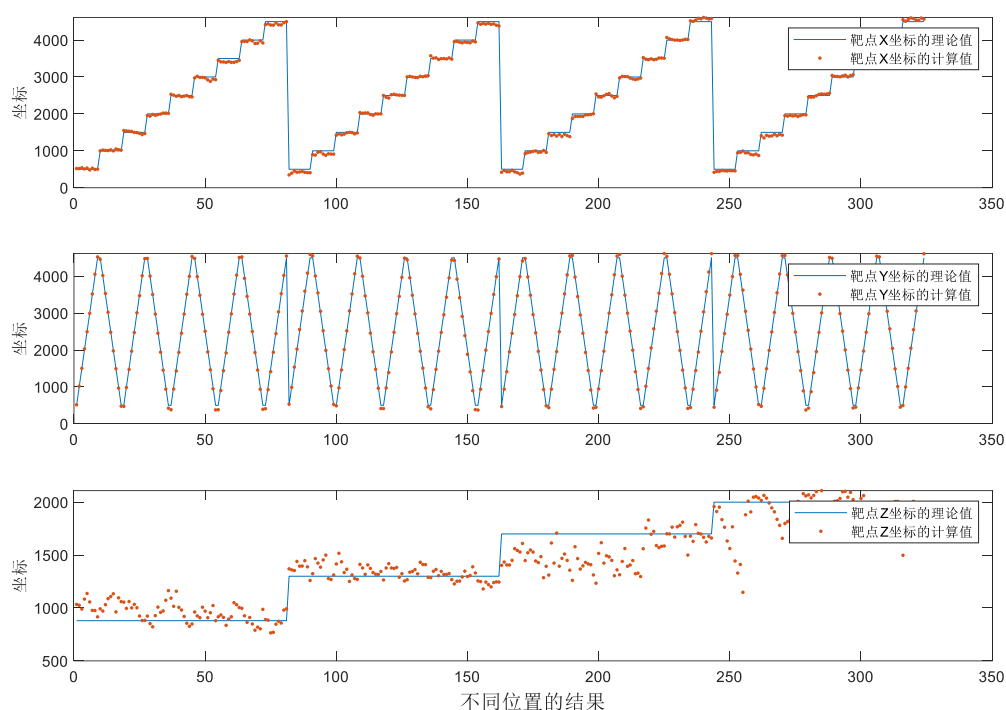


图 4.3 无信号干扰下的比对结果

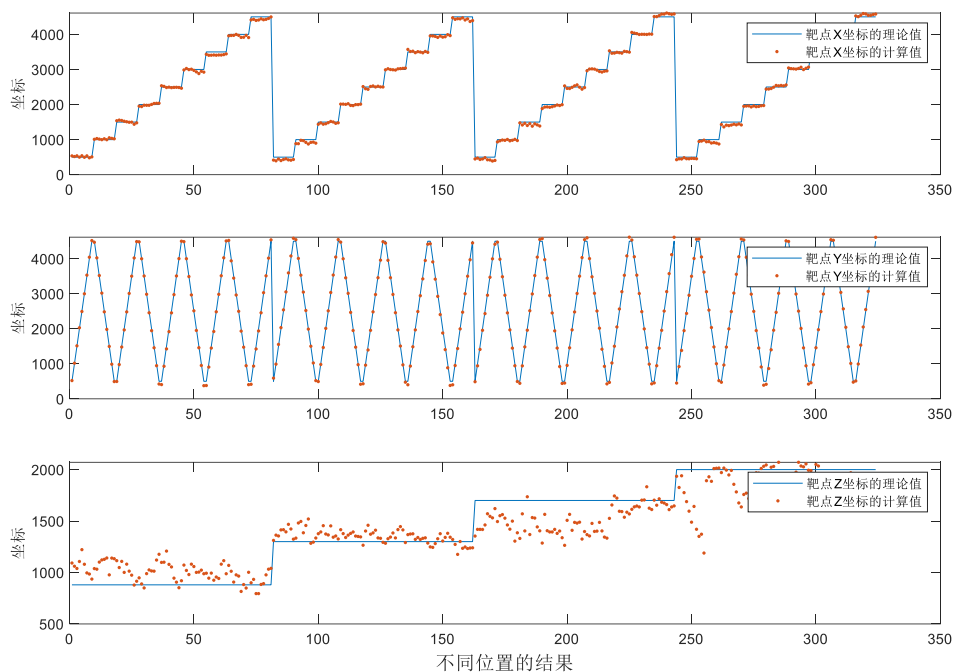


图 4.4 有信号干扰下的比对结果

通过图 4.4 可知，定位结果基本分布在真实值附近，定位结果较好。

4.5 结果及分析

对问题 2 中提供的数据根据数据类型，利用不同的误差模型进行修正，并用定位模型得到靶点坐标，结果记录在表 4.3~4.5

表 4.3 问题 2 定位坐标结果表

	无信号干扰	有信号干扰
1	(1155, 645, 1046)	(2072,685,597)
2	(3160, 1678, 953)	(4318,1542,715)
3	(2722, 1128, 1061)	(1750,1238,2494)
4	(2431, 969, 2005)	(3542,1967,2388)
5	(1448, 2520, 1803)	(4892,1940,989)

表 4.4 无信号干扰的精确度

	x	y	z	(x,y)	整体
绝对误差平均值(mm)	46.23	42.26	121.14	69.12	152.40
相对误差平均值(%)	3.15	3.13	8.50	2.30	4.28
相关性	0.9993	0.9995	0.9218	0.9994	0.9962

表 4.6 有信号干扰的精确度

	x	y	z	(x,y)	整体
绝对误差平均值 (mm)	44.76	40.85	135.48	66.27	162.95
相对误差平均值(%)	2.97	2.95	9.81	2.16	4.59
相关性	0.9993	0.9995	0.9168	0.9994	0.9957

五、问题三：不同场景下定位模型的应用

5.1 问题的描述及分析

任务 3 提供了一个新的实验场景 2，锚点位置发生了变化，靶点 也进行了调整，

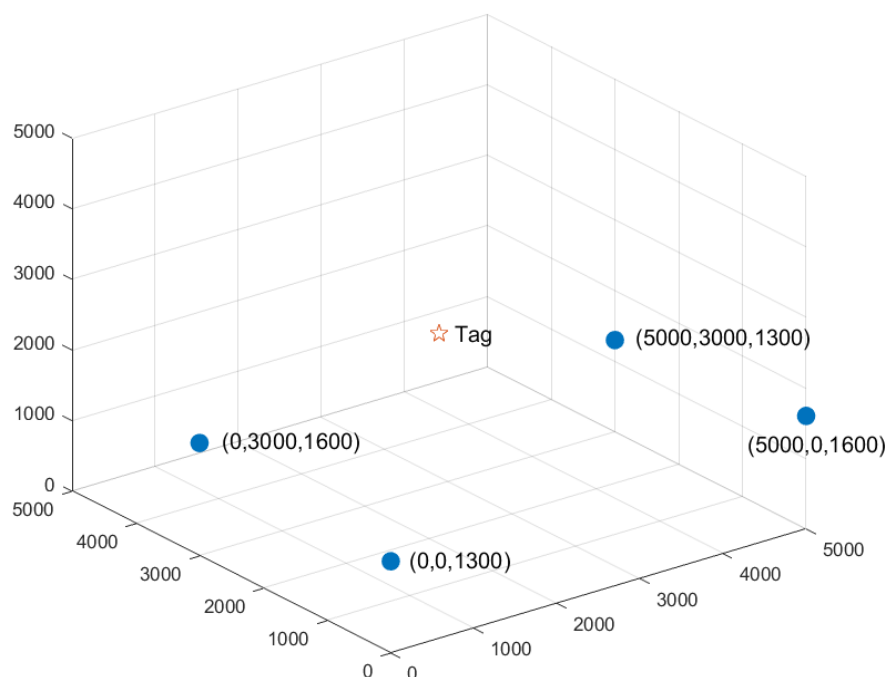


图 5.1 实验场景 2 的锚点分布

由图 5.1 可见，实验场景 2 只是在实验 1 的场景基础上改变了边界范围，和调整了锚

点位置，所以可以假设系统误差对锚点的影响并没有因为场景改变而发生变化，没有新的信号干扰类型出现，且信号干扰对测量数据的影响没有发生变化。

5.2 模型的建立和求解

调整问题 2 中的模型，可以得到问题 3 的定位模型

$$\min \sum_{i=1}^m \sum_{j=1}^n (\bar{s}_{ij} - d_{ij})^2$$

$$\begin{cases} d_{ij} \leq s_{ij} \\ s_{ij} = d_{ij} + \gamma_i^{(wgn)} \\ s'_{ij} = d_{ij} + \gamma_i^{(wgn)} + \gamma_{ij}^{(delay)} \\ 0 \leq x_j \leq 5000 \\ 0 \leq y_j \leq 3000 \\ 0 \leq z_j \leq 3000 \\ d_{ij} = \sqrt{(x_j - X_i)^2 + (y_j - Y_i)^2 + (z_j - Z_i)^2} \end{cases}$$

同理，按照问题 2 中的思路进行求解就可以得到定位结果。见表 5.1

表 5.1 问题 3 定位坐标结果表

	无信号干扰	有信号干扰
1	(3637, 2162, 1126)	(4753,1912,901)
2	(4188, 1642, 963)	(2462,1505,431)
3	(3154, 1683, 1141)	(1696,1352,2049)
4	(2565, 1837, 601)	(2023,785,705)
5	(493, 24, 748)	(1171,789,632)

六、问题四：基于含约束非线性规划模型的分类型算法

6.1 问题分析

针对问题四，给定了任务 1 的数据，要求建立数学模型，在 UWB 精确定位前先判断出是否为干扰信号。基于问题二中对正常和异常数据的分析，存在干扰信号时，4 个 UWB 锚点（A0, A1, A2, A3）始终仅只有一个锚点与靶点间有遮挡，即无论采集数据是否存

在干扰，至少有 3 个采集数据是正常的。无空间约束条件下，三维坐标的 TOA 定位至少需要 4 个锚点，但对于本问题，由于存在空间约束，事实上只需要 3 个就能完成定位。分别以一维、二维 TOA 定位进行类推，见图 6.1，可知，在约束空间下，4 个锚点对三位空间定位时仅 3 个锚点是可行的，模型验证时会具体说明这一点。

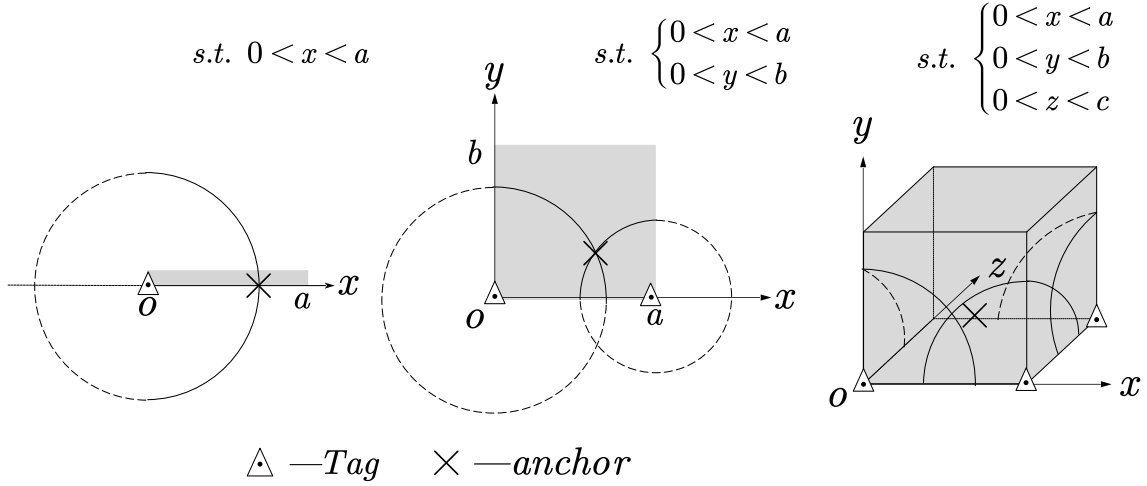


图 6.1 空间约束下一维、二维、三维 TOA 定位示意

通过上述分析，问题四的解决方法思路如下：对 4 个锚点任选取 3 个锚点，对代求靶点位置 $\text{Tagj}(x_j, y_j, z_j)$ 添加空间约束，建立其约束条件下的非线性规划模型，共有 4 种取法，对应需求解 4 次非线性规划模型。模型求解上，由于本问题建立的非线性模型是二次规划模型，考虑采用序列二次规划算法来搜索有效解。不同的锚点取法最终可以得到 4 个结果，将 4 个定位结果进行比较，如果结果相近，即可说明该组数据为无干扰下采集，如若结果相差过大，即可说明改组数据为信号干扰下采集的。思路流程图如下：

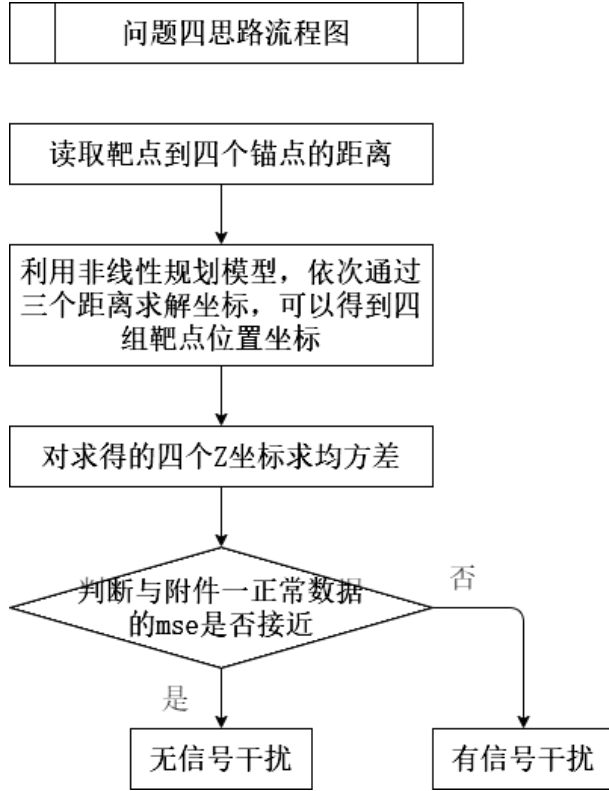


图 6.2 问题四思路流程图

6.2 模型准备

要判断提供的数据是否干扰下采集，根据上述分析的解题思路，需要先建立非线性规划模型，同问题 2 中的思路，可得：

目标函数：为方便叙述，定义 \hat{s}_{ij} 为根据等式约束计算的靶点 (x_j, y_j, z_j) 到各个锚点 (X_i, Y_i, Z_i) 的距离

$$\hat{s}_{ij} = f_s(X_i, Y_i, Z_i) = \sqrt{(x_j - X_i)^2 + (y_j - Y_i)^2 + (z_j - Z_i)^2}$$

因此，确定二次规划模型的目标函数为估计距离与测量距离的误差最小，其表达式如下：

$$\min Z = |\hat{s}_{ij} - s_{ij}|$$

决策变量：该非线性规划模型共需要建立 4 组，每一组都选取 4 个锚点的其中 3 的坐标作为决策变量，表示为 $(X_{i^{(k)}}, Y_{i^{(k)}}, Z_{i^{(k)}})$, $k=1,2,3$ ，满足 $i^{(k)} \in \{1,2,3,4\}$ ，共有 4 个组合。

约束条件：根据题目要求，靶点范围在一有限空间内，因此靶点坐标 (x_j, y_j, z_j) 不能超出边界 $x=0/a$, $y=0/b$, $z=0/c$ 围成的体积，即

$$0 < x < a; \quad 0 < y < b; \quad 0 < z < c;$$

那么可以建立如下二次规划模型

$$\begin{aligned} \min Z &= |\hat{s}_{ij} - s_{ij}| \\ s.t. &\begin{cases} 0 < x < a \\ 0 < y < b \\ 0 < z < c \\ f_s^2(X_{i^{(1)}}, Y_{i^{(1)}}, Z_{i^{(1)}}) = s_{i^{(1)}j}^2 \\ f_s^2(X_{i^{(2)}}, Y_{i^{(2)}}, Z_{i^{(2)}}) = s_{i^{(2)}j}^2 \\ f_s^2(X_{i^{(3)}}, Y_{i^{(3)}}, Z_{i^{(3)}}) = s_{i^{(3)}j}^2 \end{cases} \end{aligned}$$

6.3 基于非线性规划模型分类算法

问题四先要建立一个带约束的非线性规划模型，SQP 是优化问题(NLP) 的经典算法，对于中小规划约束最优化问题的解决特别有效。通过求解一系列优化子问题来获得最后的最优解，每个优化子问题在约束线性化的前提下优化目标的二次型模型。如果问题是无约束的，则该方法简化为寻找目标梯度为零的点的牛顿方法。对于有约束条件，该算法通过牛顿迭代法来寻找拉格朗日函数的临界点。由于拉格朗日乘数是附加变量，因此迭代形成一个系统。通过使用二次算法求解的二次极小化子问题。

针对二次规划模型求解，最终可以得到 4 组解 (x_j, y_j, z_j) ，可以表示为

$$P = \begin{bmatrix} x_j^{(1)} & y_j^{(1)} & z_j^{(1)} \\ x_j^{(2)} & y_j^{(2)} & z_j^{(2)} \\ x_j^{(3)} & y_j^{(3)} & z_j^{(3)} \\ x_j^{(4)} & y_j^{(4)} & z_j^{(4)} \end{bmatrix}$$

每一行对于不同组合求出来的位置， $P^{(k)} = (x_j^{(k)}, y_j^{(k)}, z_j^{(k)})$, $k = 1, 2, 3, 4$ ，对于正常数据，不同 k 值下的 $P^{(k)}$ 相差不大，而对于异常数据，由于 4 组中仅有一组含有由 3 个正确的锚点距离信息，其它的结果都是错误结果，因此 $P^{(k)}$ 相差较大。

6.4 模型求解

对于该分类模型的求解，需要解决两个问题。一是如何对建立的非线性模型进行求解，如果直接对满足空间约束的解空间求解，范围过大，消耗时间长。二是如何评估最终求得的 4 个解的相差程度。对于第一个问题，使用序列二次规划算法（SQP）对目标函数的最小值进行求解。

对于题目要求的 10 组数据，求解结果如表 6.1 所示：

表 6.1 附件四中 10 组数据的锚点不同组合下的求解位置(x,y,z)

数据	{A0,A1,A2} 组合	{A0,A1,A3} 组合	{A0,A2,A3} 组合	{A1,A2,A3} 组合
1	(1546,2545,1158)	(1503,2554,1689)	(1487,2570,958)	(1509,2517,1268)
2	(2329,4719,1864)	(2277,4545,2513)	(2135,4687,2189)	(2126,4510,2071)
3	(4069,2489, 1116)	(4045,2545,1420)	(4106,2484,1170)	(4162,2582,1879)
4	(3467,3625,447)	(3447,3694,698)	(3534,3607,676)	(3524,3682,549)
5	(1348,2584,857)	(1321,2680,1191)	(1436,2566,1084)	(1412,2648,785)
6	(2143,4506,1832)	(2129,4470,2015)	(2098,4500,1906)	(2093,4455,1837)
7	(3507,3565,1865)	(3506,3516,1882)	(3458,3565,1877)	(3432,3490,1546)
8	(3490,3713,745)	(3515,3627,428)	(3406,3736,456)	(3417,3641,595)
9	(4285,2514,0949)	(4310,2406,635)	(4150,2566,0298)	(4186,2415,750)
10	(2434,1020,2149)	(2443,1088,2030)	(2493,1038,1928)	(2530,1117,2393)

可以通过比较 MSE 大小来判断是否数据为干扰数据

$$mse_{(x,y,z)} = (\frac{1}{4} \sum (P^{(k)} - \bar{P})^2)^{1/2}$$

根据上式，求得 10 组数据求得的 4 组组合下的位置的 MSE，见表 6.2

表 6.2 附件四中 10 组数据的锚点的均方差

数据	mse_x	mse_y	mse_z
1	24.95	22.21	308.4
2	101.8	103.1	271.4
3	50.94	46.99	348.0
4	42.55	42.49	117.1
5	53.73	53.52	190.1
6	24.15	24.29	85.30
7	37.06	37.33	164.4
8	53.70	53.39	145.6
9	76.87	77.80	272.8
10	44.92	44.66	200.2

对附件一正常数据的作一样处理，可以得到

$$m\hat{se}_x = 23.8, m\hat{se}_y = 23.7, m\hat{se}_z = 160.6$$

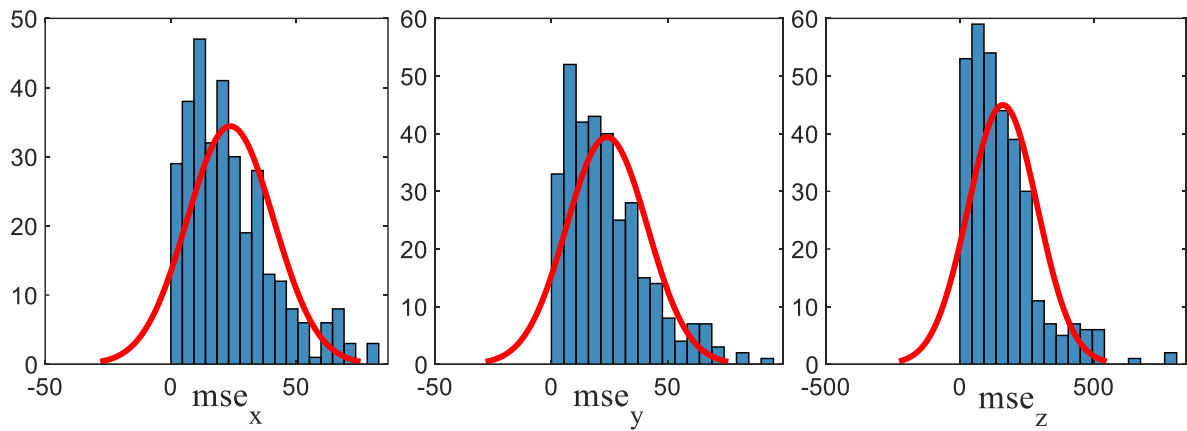


图 6.3 正常数据的均方差分布

因此可以认为正常数据和异常数组如下：

表 6.3 划分干扰和无干扰数据

无干扰数据	4、6、7、8
干扰数据	1、2、3、5、9、10

6.5 模型验证

为了验证上述模型的可靠性，对附件一的数据代入带约束的非线性模型中，求解出位置坐标，并于实际值比较，见图 6.4

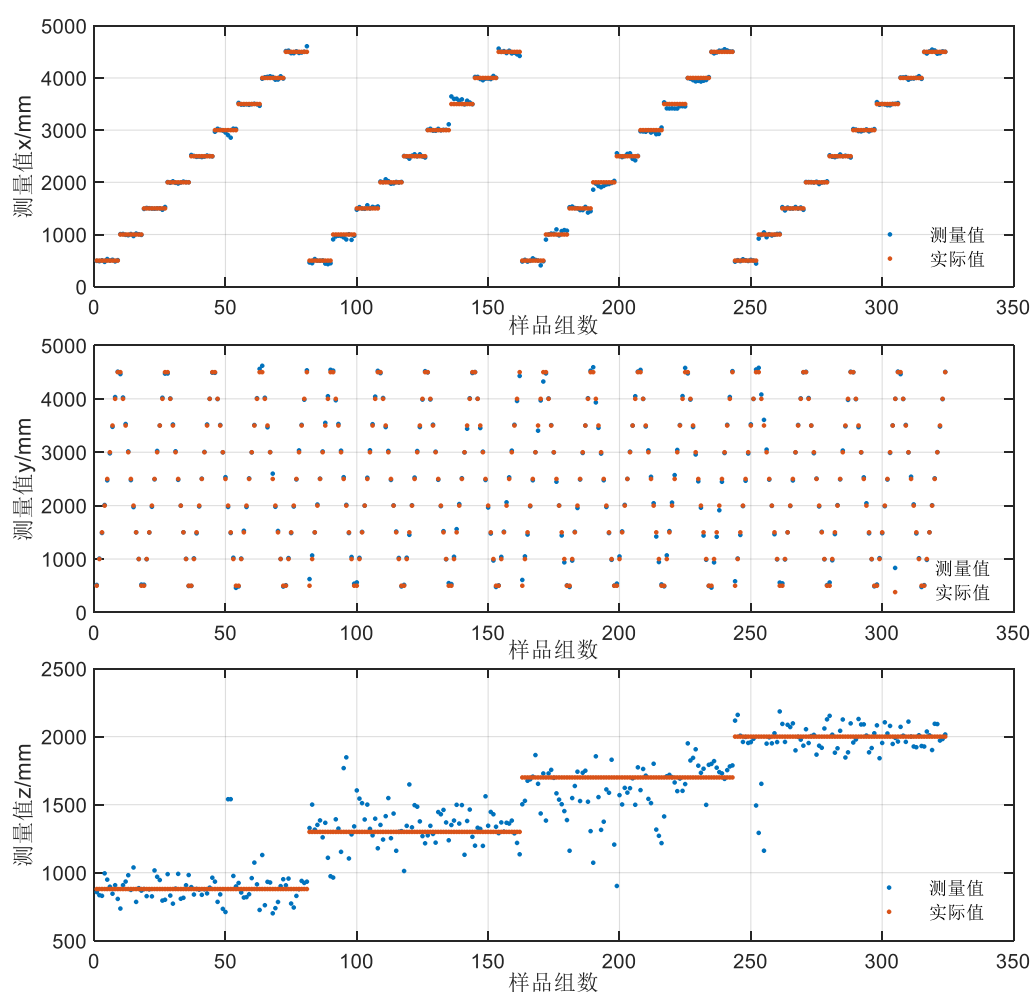


图 6.4 根据 3 锚点测量值与实际值对比

图 6.4 为 3 个锚点所求得的结果，其三个坐标的相关系数均达到 0.996, 0.998, 0.92，这也表明无干扰情况下，求得的坐标信息是相差不大的，故分成 4 种情况分别求解结果来判断是否存在干扰是有效的做法。

七、问题五：运动轨迹定位问题求解

7.1 问题分析

问题五要求实现对附件 5 对动态靶点采集的数据得到其精确的三维定位结果，要求加上靶点自身规律。本题解决思路为：先根据问题一~问题四的模型，求解出各个数据点的位置，然后结合采集数据的时间信息，先大致观察期运动规律。由于是运动靶点，其运动

轨迹在连续时间应该是连续的，故其坐标在时间前面具有承袭特性，故可以采用拟合办法，减小干扰带来的扰动，提高定位的精确度。

7.2 问题求解

7.2.1 靶点运动规律分析

读取附件五中的时间和靶点到各个锚点的距离信息，调用问题一~问题四的模型，转化成直角坐标系下的位置信息。

分析时间与各个坐标位置的变化曲线，见图 7.1-7.3

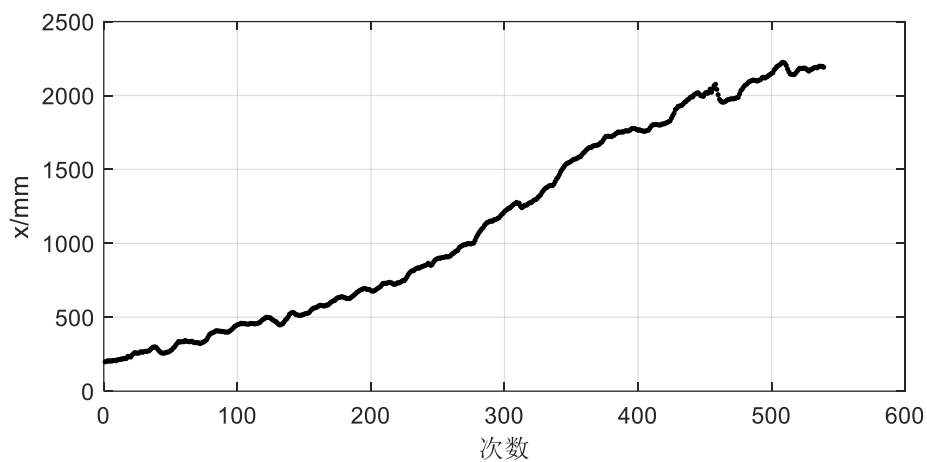


图 7.1 x 方向的坐标变化曲线

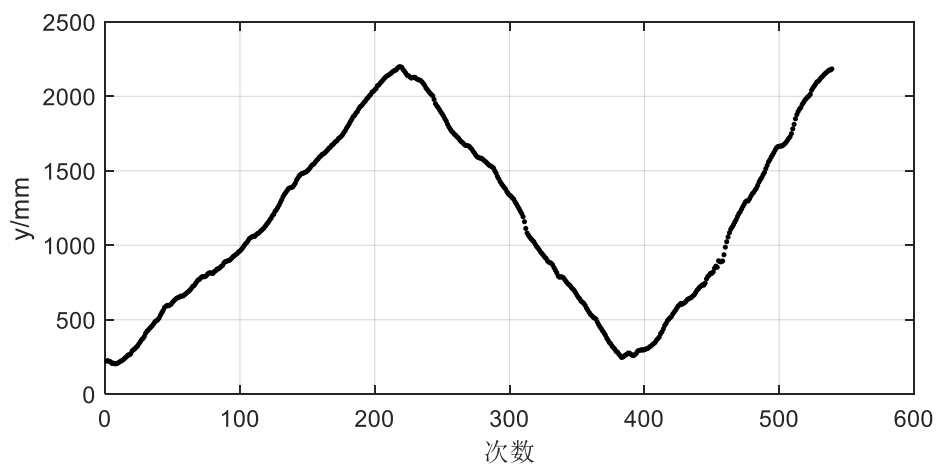


图 7.2 y 方向的坐标变化曲线

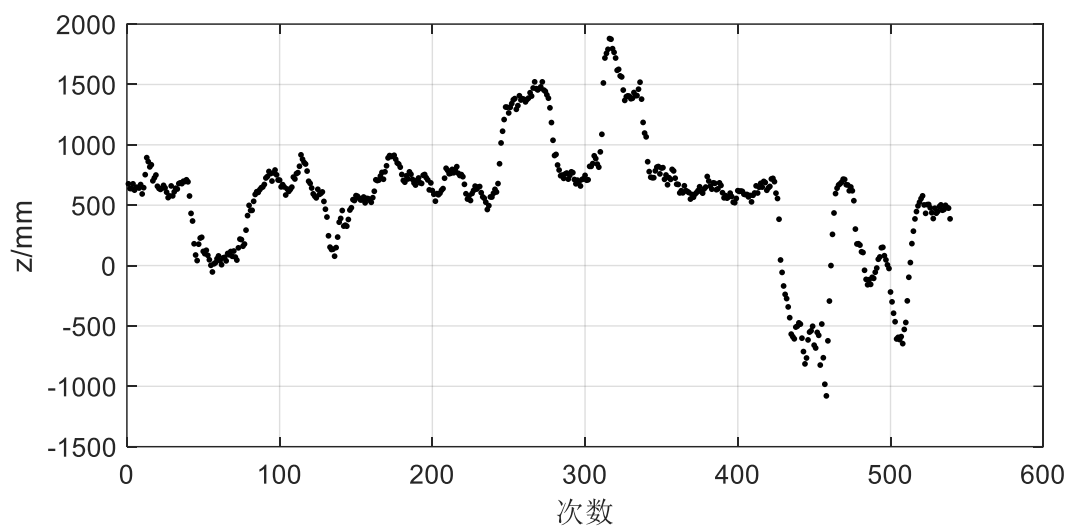


图 7.3 z 方向的坐标变化曲线

7.2.1 根据运动规律进行拟合

通过拟合，利用其运动规律平滑。对 z 方向单独平滑，平滑后 z 方向的结果如图 7.4，同时 xoy 平面的运动轨迹投影如图 7.5 所示

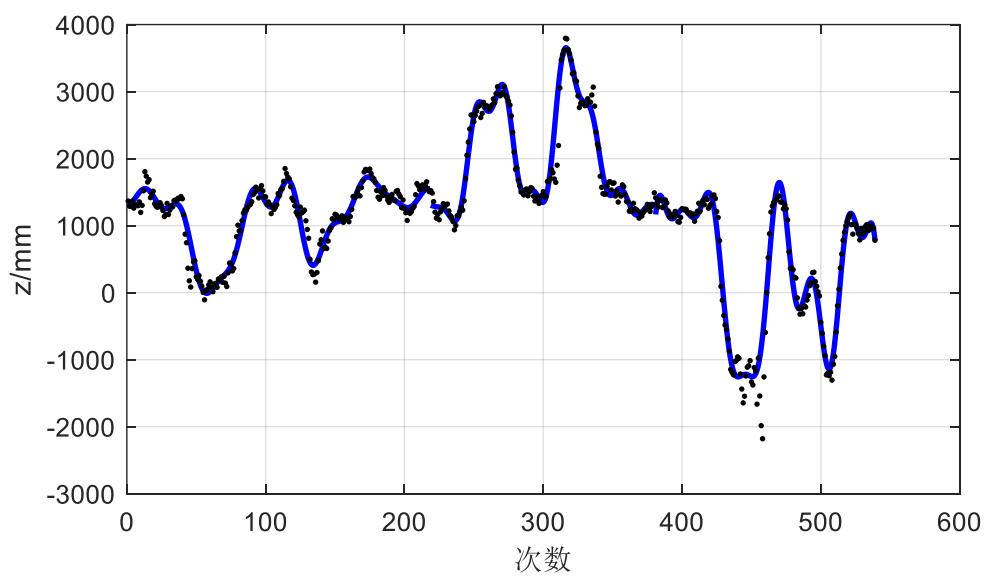


图 7.4 z 方向的坐标拟合后变化曲线

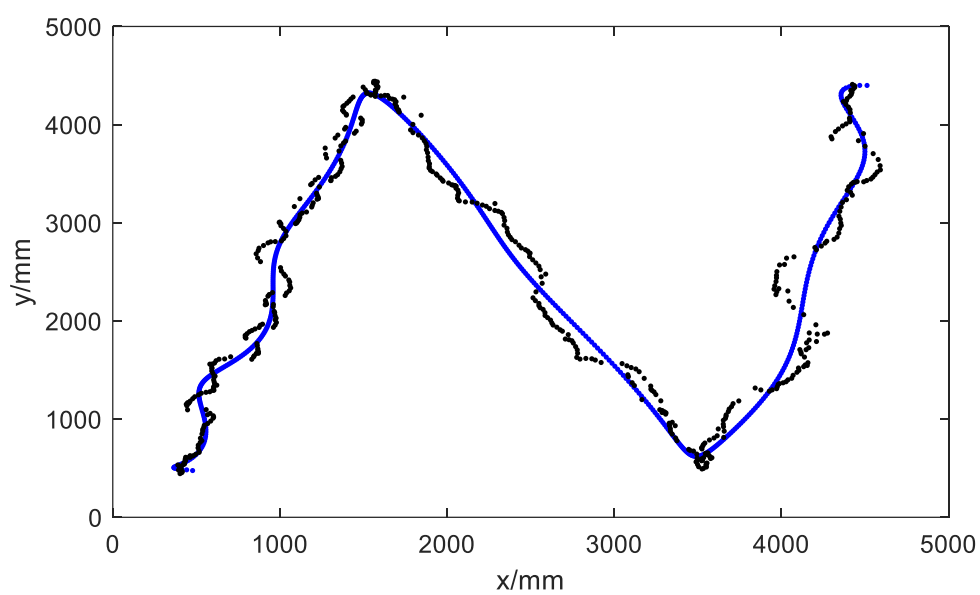


图 7.5 xoy 平面的运动轨迹投影

通过平滑后处理，最终的结果运动轨迹示意，最终效果如图 7.6

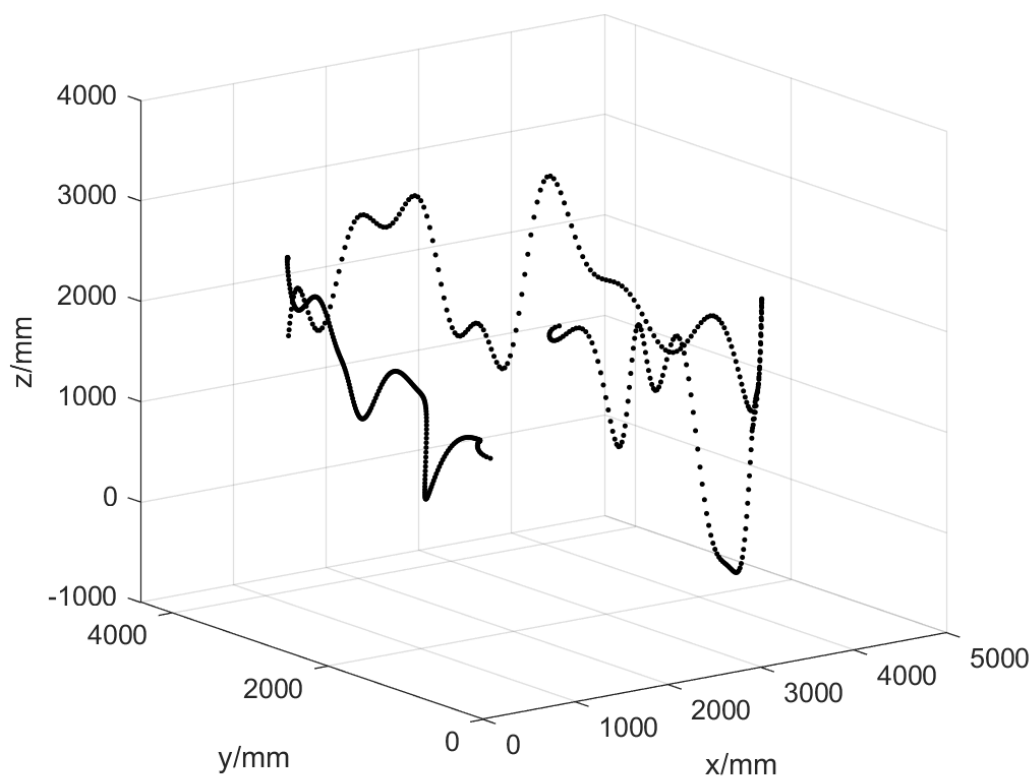


图 7.6 三维空间运动轨迹示意

七 模型评价

7.1 模型的评价

本文根据在实验场景 1 中测得的正常数据和异常数据，分别建立了系统环境的误差模型，和信号干扰产生的误差模型，和靶点的定位模型，和区分有无信号干扰的分类模型。第一种模型是非线性模型，第二种是线性模型，后两者都是非线性模型。在定位模型基础之上，对竖坐标 Z 做了额外修正，实现了三维坐标的厘米级定位。

7.1.1 优点

(1) 通过计算靶点与锚点的测量距离和真实距离的误差分布，对测量距离进行补偿，提高了定位精度

(2) 充分利用了场景信息，将误差模型线性化，大幅简化了计算

7.1.2 缺点

(1) 本文的模型通过迭代进行定位，算法的运行效率有待提高；

(2) 误差模型线性化的处理是基于实验场景 1 的室内环境，如果室内环境形状改变，从长方体变为圆柱体，或不规则体，譬如一些艺术馆，可能就无法适用。

八 参考文献

- [1]. 赵云轩. 非视距传播环境下移动台定位算法研究[D].吉林大学,2005.
- [2]. 段凯宇,张力军,高玲,等.两种 NLOS 误差消除及 TOA 定位算法[J].信号处理,2008.
- [3]. 李招华,汪毓,邵青.基于 Cha 的 TDOA 三维定位算法[J].现代电信科技,2014.

程序附件说明

根据题目输出要求，问题一以及所有代码文件汇总于附件，说明如下：

文件名	文件说明
Q1_output	问题一要求数据预处理结果
Code	处理数据的代码