

分类号: TM74

单位代码: 10363

密 级: 公开

学 号: 2210330105

题 目 智能电网状态估计中不良数据检测方法研究

题 目 Research on Bad Data Detection Methods in Smart
Grid State Estimation

论文作者: 乔礼娜

校内导师: 高文根教授

校外导师: 刘建

专 业: 电子信息

研究方向: 人工智能与系统集成

论文答辩日期: 2024 年 5 月 19 日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名： 乔礼娜

日期： 2024 年 5 月 30 日

安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 ☐，在 _____ 年解密后适用本版权书。

本学位论文属于

不保密 ☐。

学位论文作者签名：

乔礼娜

指导教师签名：

高彬

日期：2024 年 5 月 30 日

日期：2024 年 5 月 30 日

智能电网状态估计中不良数据检测方法研究

摘 要

在智能电网中,通过实时收集、分析和处理电网中的各种测量数据,实现状态估计,从而为电网运行提供准确的实时信息,全面监控电网的运行状态。然而,由于系统设备老化、传感器故障、线路连接错误等原因,电力系统中可能会产生不良数据。这些不良数据的存在会对电力系统的安全运行造成威胁。因此,及时识别和处理不良数据对于确保电力系统稳定可靠运行至关重要。

传统的不良数据检测方法通常以残差为主要依据,然而这些方法存在着“残差淹没”和“残差污染”的问题,同时计算量较大,这些问题限制了传统方法在实际应用中的效果。为了解决这些问题,本文提出了两种新的不良数据检测方法,分别是基于改进扩展卡尔曼算法(Extended Kalman Filter, EKF)的不良数据检测方法和基于次优化-最大化算法(Minorization-Maximization, MM)的不良数据检测方法。主要的研究工作如下:

(1) 构建出电力系统测量模型,阐述电力系统中的直流状态估计和交流状态估计方法。对传统的不良数据检测方法进行了研究,阐述了其检测原理,并分析了传统方法存在的局限性。这为本文的后续研究奠定了理论基础。

(2) 针对智能电网中扩展卡尔曼滤波算法在不良数据检测方面鲁棒性不足的问题,提出一种改进扩展卡尔曼滤波算法,提高了电力

系统状态估计的鲁棒性。通过在扩展卡尔曼滤波算法的预测步骤和更新步骤之间引入投影统计算法，实现对不良数据的有效检测。采用指数加权函数对检测到的不良数据进行处理，降低其权重，以有效减小不良数据对电力系统状态估计的影响，保障状态估计过程的准确性和稳定性。并在 IEEE-14 节点系统中进行仿真实验，验证了所提方法的有效性。

(3) 针对智能电网中不良数据检测与定位问题，提出一种基于次优化-最大化算法的不良数据检测方法。通过分析正常测量值和异常测量值之间的分布特征差异，构造出高斯混合模型，利用次优化-最大化算法来估计高斯混合模型中的未知参数。根据得到的参数估计对测量值进行分类，从而实现对智能电网中不良数据的检测与定位。通过在 IEEE-14 节点系统中进行仿真实验，验证了所提方法的有效性。

关键词：智能电网，状态估计，不良数据检测，扩展卡尔曼滤波器，投影统计算法，次优化-最大化算法

RESEARCH ON BAD DATA DETECTION METHODS IN SMART GRID STATE ESTIMATION

ABSTRACT

In smart grids, state estimation is achieved by collecting, analyzing, and processing various measurement data in the grid in real-time, to provide accurate real-time information for grid operation and comprehensively monitor the operating state of the grid. However, bad data may be generated in the power system due to aging system equipment, faulty sensors, and incorrect line connections. The existence of these bad data can pose a threat to the safe operation of the power system. Therefore, timely identification and processing of bad data is essential to ensure stable and reliable operation of the power system.

Traditional bad data detection methods are usually based on residuals, but these methods suffer from the problems of residual flooding and residual contamination, as well as a large amount of computation, which limit the effectiveness of traditional methods in practical applications. To solve these problems, two new bad data detection methods are proposed in this paper, namely the bad data detection method based on the Improved Extended Kalman Filter (EKF) algorithm and the bad data detection method based on the Minorization-Maximization (MM) algorithm. The main research work is as follows:

(1) A power system measurement model is constructed to illustrate the DC state estimation and AC state estimation methods in power systems. The traditional bad data detection methods are studied, their detection principles are described, and the limitations of the traditional methods are analyzed. This lays a theoretical foundation for the subsequent research in this paper.

(2) Aiming at the problem of insufficient robustness of the extended Kalman filter algorithm in detecting bad data in smart grids, an improved extended Kalman filter algorithm is proposed to enhance the robustness of power system state estimation. By introducing the projection statistics algorithm between the prediction step and the update step of the extended Kalman filter algorithm, to achieve the effective detection of bad data. An exponential weighting function is used to process the detected bad data and reduce its weight, to effectively reduce the influence of bad data on the power system state estimation and guarantee the accuracy and stability of the state estimation process. Simulation experiments are carried out in the IEEE-14 bus system to verify the effectiveness of the proposed method.

(3) Aiming at the problem of bad data detection and localization in smart grids, a bad data detection method based on the Minorization-Maximization algorithm is proposed. A Gaussian mixture model is constructed by analyzing the difference in distribution characteristics between normal and abnormal measurements, and the Minorization-

Maximization algorithm is used to estimate the unknown parameters in the Gaussian mixture model. The measurements are classified according to the obtained parameter estimates, thus achieving the detection and location of undesirable data in the smart grid. The effectiveness of the proposed method is verified by simulation experiments in an IEEE-14 bus system.

KEY WORDS: Smart Grid, State Estimation, Bad Data Detection, Extended Kalman Filter, Projection Statistical Algorithm, Minorization-Maximization Algorithm

目录

摘 要.....	III
ABSTRACT.....	V
第 1 章 绪论.....	1
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	3
1.2.1 智能电网状态估计的研究现状.....	3
1.2.2 不良数据检测方法的研究现状.....	6
1.2.3 研究现状分析.....	9
1.3 本文主要研究内容及创新点.....	9
第 2 章 智能电网状态估计与不良数据检测.....	12
2.1 引言.....	12
2.2 智能电网状态估计.....	12
2.2.1 模型构建.....	12
2.2.2 交流状态估计.....	14
2.2.3 直流状态估计.....	15
2.3 传统不良数据检测方法.....	15
2.3.1 不良数据相关概念.....	15
2.3.2 残差方程.....	17
2.3.3 目标函数极值检测法.....	18
2.3.4 加权残差检测法.....	18
2.3.5 标准化残差检测法.....	19
2.3.6 量测突变检测法.....	20
2.4 本章小结.....	21
第 3 章 基于改进扩展卡尔曼滤波算法的不良数据检测方法.....	22
3.1 引言.....	22
3.2 扩展卡尔曼滤波算法原理.....	22
3.2.1 卡尔曼滤波算法.....	22

3.2.2 扩展卡尔曼滤波算法.....	24
3.3 投影统计算法原理.....	25
3.3.1 马氏距离.....	25
3.3.2 投影统计.....	26
3.4 改进扩展卡尔曼滤波算法.....	28
3.4.1 面向不良数据的状态估计.....	28
3.4.2 不良数据检测.....	29
3.5 仿真系统与结果分析.....	31
3.5.1 仿真系统.....	31
3.5.2 仿真结果分析.....	32
3.6 本章小结.....	36
第 4 章 基于次优化-最大化算法的不良数据检测方法	37
4.1 引言.....	37
4.2 模型构建.....	37
4.2.1 不良数据测量模型.....	37
4.2.2 高斯混合测量模型.....	38
4.3 次优化-最大化算法.....	39
4.3.1 次优化-最大化算法基本思想	39
4.3.2 次优化-最大化算法基本原理	39
4.4 基于次优化-最大化算法的不良数据检测	40
4.4.1 詹森不等式.....	40
4.4.2 替代函数的构造.....	41
4.4.3 高斯混合模型中的未知参数估计.....	42
4.4.4 算法流程.....	44
4.5 算法分析.....	45
4.5.1 算法收敛性分析.....	45
4.5.2 算法复杂度分析.....	47
4.6 仿真结果分析.....	48
4.7 本章小结.....	57

第 5 章 总结与展望.....	58
5.1 全文总结.....	58
5.2 工作展望.....	59
参考文献.....	60
攻读硕士期间取得的学术成果.....	66
致谢.....	67

第 1 章 绪论

1.1 研究背景及意义

随着全球能源需求的持续增长以及对可再生能源的广泛应用，如图 1-1 所示传统的中央化、单向供电模式已经不能满足现代电力系统的需求。在这一背景下，智能电网的概念应运而生。智能电网是一种基于先进通信和信息技术的电力系统，旨在通过引入双向能源传输、实现去中心化和去层级化、提高灵活性和抗干扰能力等措施，以更好地适应快速变化的电力需求和应对环境挑战。传统电网中央化结构易受单点故障影响，能源分配和传输缺乏灵活性，难以整合大规模可再生能源。智能电网建设首先可提高能源利用效率，通过更灵活、智能的能源分配和调度，实现有效利用，减少浪费。其次，发展智能电网有助减小环境影响，通过整合可再生能源和智能调度降低对传统化石能源依赖，减排温室气体，有利于应对气候变化等环境问题。此外，智能电网建设还能提高电力系统可靠性和稳定性，减少单点故障引发的停电事故，提高供电质量，保障用户用电需求。

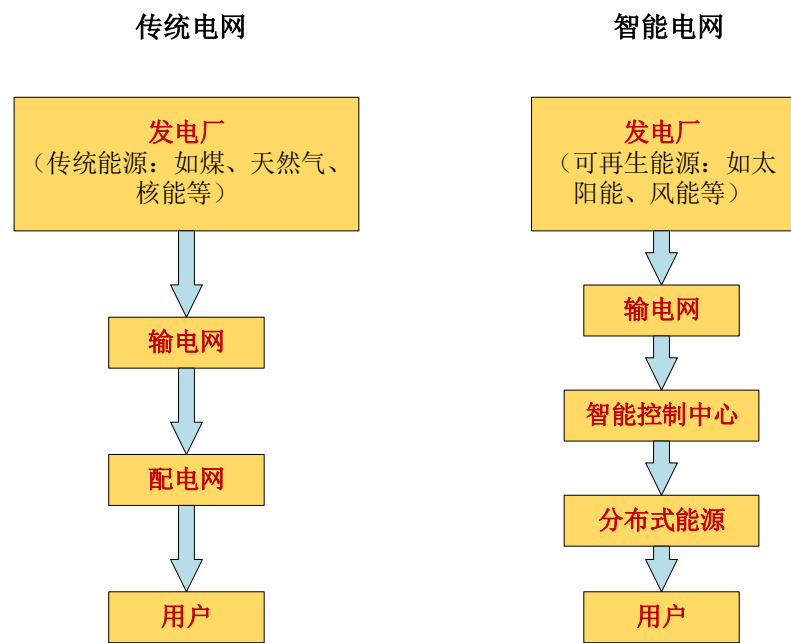


图 1-1 传统电网与智能电网对比图

智能电网的核心组成部分包括能量管理系统（Energy Management System, EMS）^[1]与数据采集与监控系统（Supervisory Control And Data Acquisition, SCADA）

[2]。SCADA 系统在电力系统中负责实时监测和数据采集，但单独的测量值并不能直接反映电力系统的全面运行状态。这些测量值可能受到噪声、测量误差或数据丢失等因素的影响，因此需要通过进一步处理和分析来获得准确的电网运行状态信息。

EMS 在智能电网中扮演着关键的角色，它通过对 SCADA 系统采集到的数据进行状态估计，得到更为准确和全面的电力系统运行状态估计值。利用先进的数学模型、算法和优化技术，EMS 对来自 SCADA 系统的数据进行处理和分析，包括潮流分析、状态估计、故障诊断等，从而实现对电力系统运行状态的实时监控和精确评估。准确的状态估计是确保电力系统稳定运行的基石，因为它提供了关键的信息，支持电力系统的实时监控、调度和控制。通过状态估计，能够充分利用传感器测量数据，准确估计系统中各节点的状态变量，如电压和相角，从而实现对整个系统运行状态的精准监测，为电力系统的稳定性和可靠性提供了必要的基础。一旦 EMS 获得了准确的电网运行状态估计值，它就能够根据系统的实际情况发出相应的实时命令。这些命令可能涉及电力设备的控制、调整发电机的输出功率、调度电力负载等。通过 EMS 的智能调度和控制，可以保障电网的安全稳定运行，最大程度地提高电力系统的效率和可靠性。

在智能电网的复杂环境中，数据的多样性和复杂性为电网状态估计带来了潜在的挑战。测量数据从采集到调度中心，要经过数据传输等多个通道，在这一过程中可能会产生不良数据。这些不良数据可能源于测量设备精度不足、设备故障导致的异常数据输出，或是通信故障引发的数据传输丢失与失真。若未能对这些不良数据进行及时有效的识别与处理，必将对电网状态估计的准确性产生严重影响，进而威胁到电网的可靠性和稳定性。

对不良数据检测方法的深入研究对于确保智能电网状态估计的准确性至关重要。在智能电网中，准确的数据是正确运行和管理的基础。通过有效的不良数据检测方法，可以提高数据质量，确保数据的可靠性，为系统提供可信的信息支持，从而增强电网系统的安全性。通过识别可能是由设备故障、通信攻击或其他恶意行为引起的不良数据，系统能够提高对潜在威胁的防御能力，降低因数据异常而引发的安全风险。具体而言，不良数据的检测需要综合运用先进的数据分析技术、机器学习算法以及模型预测方法。通过对历史数据的深入分析，可以建立

模型来识别正常和不良数据的模式，从而在实时数据中进行快速准确的检测。此外，还可以采用实时监测技术，对电力系统各个环节的数据进行持续监控，及时发现和响应不良数据。这样的实时性和准确性是确保电网状态估计的关键因素。因此，不良数据的有效检测和处理是智能电网安全稳定运行的重要保障之一。

深入研究不良数据检测方法不仅关乎电网的安全和可靠运行，更为智能电网的发展提供了关键支持。通过确保状态估计的精准性和可信度，能够更好地迎接未来电力系统面临的挑战，实现智能电网的可持续发展。准确的状态估计不仅可以帮助电力系统运营商及时发现和解决潜在的故障和问题，还可以提高电力系统的运行效率，降低运行成本。此外，精确的状态估计还为优化调度策略、提升电力系统的整体性能提供了坚实基础。通过不断改进和完善不良数据检测方法，可以提高电力系统数据的质量和可靠性，进而促进智能电网的可持续发展。这将为电力系统的智能化、高效化和可靠化提供关键支持，推动智能电网技术的不断创新和应用。因此，深入研究不良数据检测方法对于智能电网的发展具有重要意义，将为构建更加安全、可靠、高效的电力系统奠定坚实基础。

1.2 国内外研究现状

1.2.1 智能电网状态估计的研究现状

状态估计是电力系统能量管理系统中的一个重要功能。为了执行状态估计，控制中心利用 SCADA 系统收集测量数据，这些数据通常包括有功注入和无功功率注入、传输线路中的有功功率流和无功功率流，以及母线上的电压幅值。通过状态估计，将测量数据转换为系统状态变量，这对于实现电力系统的实时监测、故障诊断和运行控制至关重要。随着智能电网的发展，状态估计变得更为复杂，需要更高的准确性和实时性。智能电网引入了更多的可再生能源、分布式能源和储能设备，使得电力系统的结构更加复杂多样。此外，智能电网还面临着更大规模的数据处理挑战，需要处理来自各种传感器和设备的大量数据，并在瞬息万变的环境中做出及时准确的决策。

为了应对这些挑战，需要不断改进状态估计算法和技术，提高其准确性和实时性。同时，还需要加强电力系统的通信和信息技术基础设施，以支持大规模数

据的快速传输和处理。只有如此，才能更好地实现电力系统的实时监测、故障诊断和运行控制，推动智能电网的可持续发展。

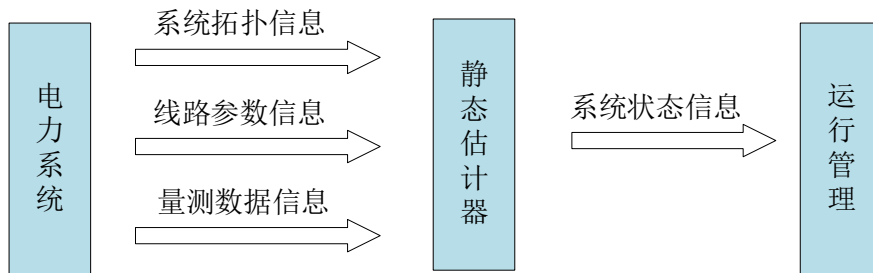


图 1-2 静态电力系统状态估计

状态估计方法一般分为静态状态估计和动态状态估计。最初，大多数状态估计研究都集中在静态状态估计上^[3]，如图 1-2 所示。静态状态估计是利用系统的拓扑信息、线路参数信息以及实时测量数据来估计电力系统在特定时刻的状态。这种估计主要基于系统的稳态数学模型。其中，加权最小二乘法（Weighted Least Square, WLS）^[4]是一种常用的静态状态估计方法，但其在实际应用中存在一些限制。

WLS 方法的计算复杂度相对较高，这使得在实时应用中可能效率较低。此外，该方法的性能也高度依赖于对系统模型的准确性。如果系统模型存在误差或不准确，可能导致状态估计结果不够准确。针对这些问题，研究人员提出了各种改进的算法和技术，以提高状态估计的精确性和实时性。文献[5]中提出的解决方案通过优化相量测量方程的结构，简化了 WLS 线性估计公式，从而显著减少了计算时间，并提高了计算性能。这种优化方案为静态状态估计提供了更高效的实现途径，同时保持了较高的求解精度。这种改进对于电力系统的安全稳定运行和智能电网的发展具有积极的影响，为提高电力系统的运行效率和可靠性提供了有力支持。

实际上，电力系统的运行状态是时刻在变化的，测量量和状态量也随之持续更新。因此，需要进行实时的、动态的状态估计，以跟踪系统的真实运行状态。动态状态估计^[6]是指在电力系统运行过程中，根据最新的实时测量数据，利用数学模型和算法对系统的状态进行估计和更新的过程。

在动态状态估计中，系统的状态不断随着时间变化而更新，以反映电力系统实际的运行情况。通过动态状态估计，可以更准确地把握电力系统的运行状态，

及时发现潜在的问题和故障，并采取相应的措施进行调整和处理，从而保障电力系统的安全稳定运行。动态状态估计的过程如图 1-3 所示，它参考前一时刻的数据信息，利用某种状态方程来预测下一时刻的系统状态，并通过滤波对估计结果进行优化。这种方法有效地结合了历史数据和预测模型，能够更准确地预测未来的系统状态，为系统运行提供更可靠的指导。此外，动态状态估计还能通过预测提供实时数据，用于监测并提前警示电力系统的运行状态。通过连续地更新和优化系统状态的估计结果，可以及时发现潜在的问题，并采取相应的措施进行干预和调整，从而避免电力系统出现严重故障或失控情况。

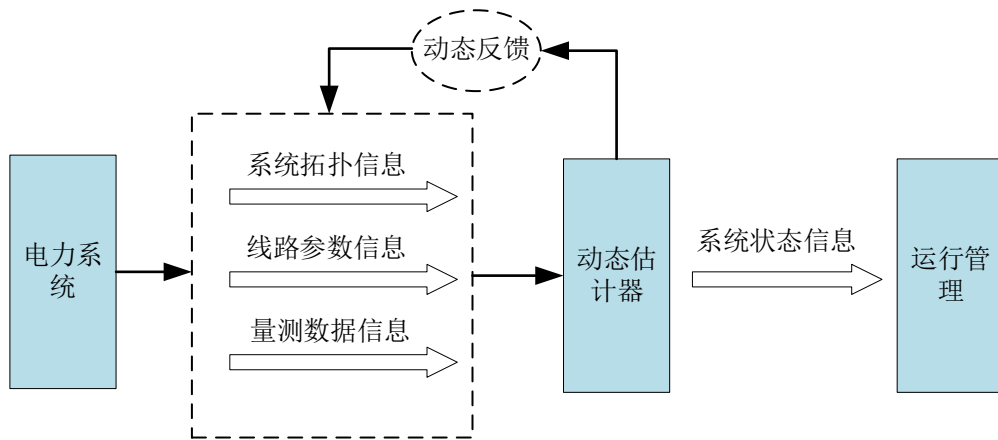


图 1-3 动态电力系统状态估计

为了实现电力系统的动态状态估计，通常采用卡尔曼滤波器(Kalman Filter, KF)^{[7]-[8]}，如扩展卡尔曼滤波器(Extended Kalman Filter, EKF)^{[9]-[11]}、无迹卡尔曼滤波器(Unscented Kalman Filter, UKF)^{[12]-[15]}和粒子滤波器(Particle Filter, PF)^{[16]-[17]}。相较于普通的卡尔曼滤波器，EKF 能够处理非线性系统，并且通常具有较高的计算效率。EKF 通过一阶泰勒展开来近似非线性函数，以估计电力系统的状态。尽管文献[18]中详细介绍了基于 EKF 算法的动态状态估计，但文献[19]指出该算法的缺点，即对非线性系统动态和观测模型进行线性化可能会引入误差，并在某些情况下导致滤波器对高度非线性函数的发散。

文献[20]中的 UKF 方法将无迹变换(Unscented Transformation, UT)与卡尔曼滤波理论相结合，能够提供更好的估计结果。无迹变换用于生成一组称为 Sigma 点的向量，以捕捉原始状态分布的均值和协方差。相较于 EKF，UKF 不对状态和测量模型的非线性进行线性化处理，这使得 UKF 在处理高度非线性的系统时更为适用，但也存在数值不稳定性问题。

另一种方法是粒子滤波器^[21]，它是一种基于概率的完全非线性估计器，最早在 1940 年代提出。粒子滤波器不仅不对噪声类型做出任何特定假设，也不对高度非线性系统进行线性化处理。这使得粒子滤波器在处理复杂系统和非线性问题时具有独特的优势。粒子滤波器也被称为自举滤波器^[22]、顺序蒙特卡罗滤波器^[23]和凝聚算法^[24]。由于粒子滤波器不对系统进行线性化处理，因此适用于广泛的非线性系统，并在实时定位和跟踪、目标追踪、机器人导航等领域得到广泛应用。然而，粒子滤波器的计算复杂性随着粒子数量的增加而增加，因此在高维状态空间或需要实时性能的情况下可能会面临挑战。

1.2.2 不良数据检测方法的研究现状

在电力系统能量管理系统中，状态估计的准确性对于实时监测、故障诊断和运行控制至关重要。然而，不良数据的存在可能导致状态估计结果与实际系统运行之间出现较大的偏差，严重影响估计器的性能。因此，不良数据检测工作显得尤为重要，以确保状态估计的可靠性和准确性。

在电力系统中，不良数据可能由于各种原因而产生，包括传感器故障、通信故障、数据采集错误等。为了应对这些问题，通常采用不良数据检测技术来识别并剔除潜在的不良数据。传统的不良数据检测方法包括目标函数极值检测法^[25]、加权残差检测法^[26]、标准化残差检测法^[27]、量测突变检测法^[28]、最大归一化残差检测法^[29]、残差搜索法^[30]以及非二次准则法^[31]等多种方法。

在文献[25]中，目标函数极值检测法通过对目标函数进行极值分析，能够快速识别潜在的异常数据。这一方法相对简单直观，易于理解和实施。然而，其对数据分布的假设较为敏感，若与实际数据分布不符，可能导致误报或漏报。此外，该方法在应用于复杂数据集和多变量问题时的适用性受到限制，难以充分捕捉数据之间的复杂关系。同时，在处理大规模数据时，由于需要进行全局搜索对目标函数进行分析，而导致检测效率较低。

文献[26][27]中的加权残差检测法和标准化残差检测法与目标函数检测法有所不同。加权残差检测法和标准化残差检测法并非全局型检测的方法，而是采用局部化的策略。这种方法在处理复杂数据集时可能更为灵活，具有更强的适应性。然而，采用这种方法可能存在残差淹没和残差污染的风险。文献[28]中的量测突

变检测法能够有效地克服由残差淹没引起的检测障碍。

文献[29]中的最大归一化残差检测法具有高度敏感性，能够有效地识别数据中的不良数据。该方法通过归一化残差，并选取归一化残差的最大值来判断是否存在不良数据。由于归一化过程能够将残差值转化为相对于数据集整体的异常程度，因此这种方法对于异常数据的识别相对较为敏感。

文献[30]中的残差搜索方法使用加权残差和标准残差对测量数据进行排序。剔除残差较大的数据后，重新进行状态估计，以达到最优估计的目的。然而，这种方法的缺点是计算量大，容易产生残差污染和残差淹没，导致误检或漏检。

而文献[31]中的非二次准则法采用的策略是逐步降低可疑数据的权重，而不是逐个消除可疑数据。这一方法避免了重新估计以减少计算量，但可能导致收敛困难，并且不能保证最终的估计是最佳的。

文献[32]中解释并测试了另一种处理不良数据的方法，即假设检验。这种方法通过对数据集的统计特征进行假设检验，从而判断是否存在异常数据。假设检验方法能够在一定程度上提供对数据异常性的量化评估，并且具有一定的理论基础支持。通过在实际数据集上进行测试，可以评估这种方法在不同情况下的有效性和适用性。

文献[33]中提出了加权最小绝对值（Weighted Least Absolute Value, WLAV）算法，该算法旨在通过加权最小绝对值的优化准则来进行状态估计。相比于传统的最小二乘法，WLAV 算法能够更好地处理异常值，并且具有更强的鲁棒性。

文献[34]中的主成分分析（Principal Component Analysis, PCA）用于处理虚假数据攻击。PCA 是一种常用的数据降维技术，通过寻找数据中的主成分来减少数据的维度。在处理虚假数据攻击时，PCA 可以帮助识别和剔除攻击性数据，从而提高状态估计的准确性和可靠性。

文献[35]中提出了另一种方法，该方法基于研究测量值在空间和时间上的相关性，以减少状态估计所需的数据量。然后，开发一种算法来使用该压缩测量值来估计系统的状态。这种方法利用了测量值之间的相关性信息，通过压缩数据量来提高状态估计的效率，并且能够保持较高的准确性。

文献[36]使用投影统计算法处理所有类型的异常值，包括垂直异常值和不良杠杆点，并且还开发了其他算法来减轻杠杆点效应^[37]。该算法能够有效降低数据

集的维度，提高对不良数据的鲁棒性。此外，该算法在处理不良数据的同时，也能够降低数据的维度，从而提高计算效率。这种方法在处理各种类型的不良数据时都表现出很好的效果，并且具有较强的适应性和鲁棒性。

随着科学技术的不断发展，现今在处理异常值和数据质量问题方面出现了一些新的理论方法，其中包括基于数据挖掘算法、基于模糊数学理论、以及基于新息向量等方法。

文献[38]提出了一项创新性的方法，将数据挖掘中的关联规则应用于不良数据检测与辨识。通过采用 **Apriori** 算法和周期性关联规则，研究人员从系统历史数据库中获取信息，有助于在每一时刻获取测量数据，从而为状态估计决策提供科学依据，增强了系统的可靠性。这种方法利用了关联规则的挖掘技术，为不良数据检测提供了一种更为有效和可靠的途径。通过挖掘历史数据中的关联规则，可以发现测量数据之间的潜在关系，进而识别异常数据点。这种方法不仅能够提高异常数据检测的准确性，还能够为系统状态估计提供更加科学的依据，从而增强系统的可靠性和稳定性。

为克服传统检测方法中可能存在的残差污染和残差淹没问题，刘浩等研究人员^[39]根据模糊数学理论提出了一种基于模糊动态的不良数据检测方法。该方法运用了标准残差和相邻两采样时刻的量测数据差值作为特征量，通过模糊聚类分析来有效解决问题。这种方法能够更好地处理不良数据，提高了检测的准确性和鲁棒性。

文献[40]在这一方法的基础上进行了进一步的改进，采用聚类分析方法如 **ISODATA** 或遗传算法获取聚类矩阵，从而实现全局最优解。通过引入聚类分析，可以更好地对数据进行分类和处理，进一步提高了异常数据检测的效果。这种改进使得方法能够更好地适应不同类型和复杂程度的数据集，并且能够更加全面地考虑数据之间的关系，从而提高了检测的精度和可靠性。

文献[41][42]采用新息图方法，综合考虑动态和静态两方面理论，通过将具有具体物理意义的数据（新息向量）转化为虚拟假设条件下的等效表示。这一方法结合了基本的电路理论，旨在实现对不良数据和拓扑错误的检测与辨识。通过将数据转换为虚拟条件下的等效表示，并结合电路理论，这种方法有效地实现了对不良数据和拓扑错误的检测和辨识。

1.2.3 研究现状分析

现有研究在智能电网状态估计方法上取得了一定进展,但仍存在一些挑战和待解决的问题。静态状态估计方面,虽然 WLS 等方法已经取得了一定成就,但在实时应用中的效率和准确性仍有待提高,尤其是对于复杂系统的建模和数据处理方面还存在挑战。动态状态估计方面,虽然 KF 及其变种在处理非线性系统时取得了一些进展,但仍然存在数值稳定性和计算复杂性等问题,如何在保证估计精度的同时提高计算效率是未来研究的重点之一。

在电力系统能量管理系统中,状态估计对于实时监测、故障诊断和运行控制至关重要。然而,不良数据可能导致状态估计结果与实际系统运行存在较大偏差,因此不良数据检测成为保障状态估计准确性和可靠性的重要环节。目前,已经提出了多种不良数据检测方法,包括目标函数极值检测、加权残差检测、标准化残差检测、量测突变检测、最大归一化残差检测、残差搜索、非二次准则等。这些方法各具特点,适用于不同的数据特征和应用场景。最新研究还提出了一些新的方法和改进策略,如基于数据挖掘算法的检测、基于模糊数学理论的方法以及基于新息向量的处理等。虽然取得了一定成果,但不良数据检测仍面临一些挑战,如处理复杂数据集、提高检测效率、降低误报率等。未来的研究可能集中在进一步改进现有方法、探索新的技术手段以及应对实际应用中的挑战等方面,以提高状态估计的可靠性和准确性,支持电力系统的安全稳定运行。

1.3 本文主要研究内容及创新点

不良数据可能导致电网状态估计的不准确性,从而影响电网的运行稳定性。通过深入研究不良数据检测方法,可以有效减少不良数据的影响,确保电网运行的可靠性。此外,精确的不良数据检测有助于提高智能电网系统对实时数据的可信度,为智能决策提供更为可靠的基础。因此,研究不良数据检测对于智能电网的正常运行和智能决策的支持至关重要。本文的主要研究内容如下:

第一章:首先介绍了研究背景和意义,突出智能电网状态估计与不良数据检测在电力系统运行与管理中的重要性。随后对智能电网状态估计与不良数据检测方法的国内外研究现状进行了综述。在本章中,明确了本文的主要研究内容,即

针对智能电网中不良数据的检测与处理问题展开研究。综述了国内外在状态估计和不良数据检测方面的研究成果和进展，为后续章节的展开提供了基础。

第二章：首先引入了智能电网状态估计的概念，详细介绍了模型构建以及交流和直流状态估计的关键步骤。随后，对传统不良数据检测方法进行了探讨，包括不良数据的相关概念、残差方程以及各种检测方法的原理。

第三章：提出了基于改进扩展卡尔曼滤波算法的不良数据检测方法。首先介绍了卡尔曼滤波（Kalman Filter, KF）算法和扩展卡尔曼滤波（Extended Kalman Filter, EKF）算法的基本思想。随后，引入了投影统计（Projection Statistics, PS）算法原理，有效检测不良数据。为了降低不良数据对电力系统的不良影响，引入加权指数函数，有效减轻不良数据在系统中的权重。并引入加权指数函数降低不良数据权重，减轻其对电力系统的影响，最后通过仿真系统和结果分析，验证了该方法的有效性。

第四章：提出了基于次优化-最大化（Minorization-Maximization, MM）算法的不良数据检测方法。首先构建了不良数据的测量模型和高斯混合测量模型（Gaussian Mixture Model, GMM）。引入次优化-最大化算法，并应用于不良数据检测中。通过构造替代函数优化高斯混合模型中的未知参数，对样本数据进行迭代分类，进而检测和定位出不良数据。最后，通过仿真结果分析，验证了该方法在不良数据检测方面的性能。

第五章：对全文的工作内容和研究成果进行总结，分析目前存在的不足之处，对智能电网还可能存在的安全风险做了进一步的研究和探讨。

本文主要创新之处如下：

（1）针对智能电网中扩展卡尔曼滤波算法在不良数据检测方面鲁棒性不足的问题，结合扩展卡尔曼滤波算法与投影统计算法提出了一种改进的扩展卡尔曼滤波算法。扩展卡尔曼滤波算法擅长处理系统动态模型的状态估计，对于系统的时序变化具有较好的适应性。通过将其与投影统计算法结合，可以在状态估计的同时，利用投影统计算法的多类型不良数据处理能力，提高了对动态系统中不良数据的检测精度。并采用指数加权函数处理检测出的不良数据，改进后的算法显著提升了智能电网状态估计的鲁棒性和准确性。

（2）针对智能电网状态估计中可能存在的不良数据，提出采用统计学习领

域中的次优化-最大化 (Minorization-Maximization, MM) 算法来进行检测与定位。与传统的不良数据检测方法相比, MM 算法能更准确地检测异常值, 特别是在面对复杂的数据分布时, 其性能更为优越。MM 算法是一种优化算法, 其核心思想是通过找到更易于优化的函数来替代原始函数, 从而间接优化目标函数。在电力系统中, 基于正常测量值与异常测量值的分布差异构建了高斯混合模型。利用 MM 算法, 对该模型中的未知参数进行优化, 以实现不良数据的精准检测和定位。

第 2 章 智能电网状态估计与不良数据检测

2.1 引言

在智能电网中，状态估计与不良数据检测等关键技术扮演着至关重要的角色，旨在显著提升电网运行的可靠性与安全性。随着电力系统规模的扩大以及新能源和分布式能源的不断引入，智能电网面临着前所未有的挑战，使得电力系统的运行变得更加复杂和动态。状态估计作为电力系统的核心组成部分，通过融合测量值、控制命令和模型等多方面信息，运用先进的数学方法对电力系统各个节点的电压、相角等状态变量进行推断。提供了准确而全面的系统状态信息，为实现电网的稳定运行提供了有力支持。然而，电力系统常常面临各种不良数据的干扰，例如错误的测量值、通信故障等。这些不良数据可能导致对系统状态的误判，进而影响电网的正常运行。因此，不良数据检测成为保障电力系统运行的另一重要环节。通过先进的算法和技术，及时发现并排除错误数据，确保状态估计的准确性和可靠性，从而进一步提升电网的整体性能。本章主要介绍了智能电网状态估计的基本概念以及传统的不良数据检测方法。

2.2 智能电网状态估计

2.2.1 模型构建

智能电网状态估计是一项基于数据采集和分析的技术，旨在实现对电网各个组件运行状态的实时监测和准确估计。该系统利用先进的传感器技术，包括电流传感器、电压传感器等，实时采集电网节点的关键参数。通过建立高效的通信网络，采集到的数据传输到中心服务器或控制中心，进行综合处理和分析。数据融合技术被用于整合不同传感器的信息，提高数据的准确性和完整性。在进行状态估计时，智能电网系统采用多种方法，包括卡尔曼滤波、最小二乘法等，以确保对电网状态的准确估计。同时，通过实时监测系统，能够快速检测电网中的异常情况，并采取相应的措施，以维护电网的稳定性和可靠性。电力系统状态估计功能流程框图如图 2-1 所示。

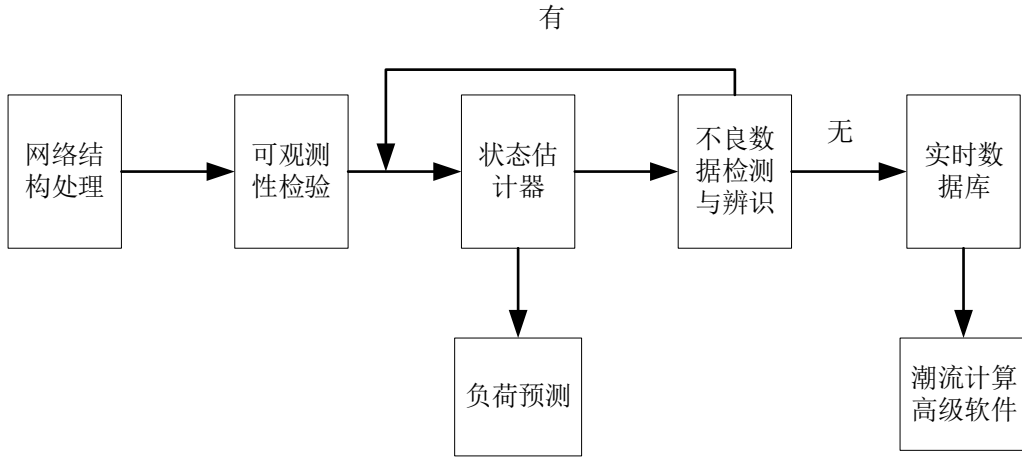


图 2-1 状态估计功能流程图

在智能电网状态估计中，通常采用非线性模型描述测量向量 $\mathbf{z} \in \mathbf{R}^{m \times 1}$ 与状态向量 $\mathbf{x} \in \mathbf{R}^{n \times 1}$ 之间的关系。通过对电压、电流、功率等关键参数的测量，状态估计算法能够推断电力系统中各个节点和支路的状态信息，从而实现对系统当前状态的准确估计，为运行监控、故障诊断和优化控制等应用提供关键信息。这种关系可以用非线性方程^[43]表示为

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{e} \quad (2-1)$$

其中 $\mathbf{e} = [e_1, e_2, \dots, e_m]^T \in \mathbf{R}^{m \times 1}$ 为测量误差向量，假定 $\mathbf{e}$ 为均值为零、协方差为 σ^2 的高斯测量噪声，即 $\mathbf{e} \sim \mathcal{N}(0, \sigma^2)$ 。 $\mathbf{h} = [h_1(\mathbf{x}), h_2(\mathbf{x}), \dots, h_m(\mathbf{x})]^T \in \mathbf{R}^{m \times 1}$ ， $h_i(\mathbf{x})$ 是将第 i 次测量与状态向量 $\mathbf{x}$ 联系起来的非线性函数。状态向量和测量向量之间的非线性关系 $\mathbf{h}(\cdot)$ 为电力系统的网络拓扑结构。如图 2-2 所示， π 型等效电路描绘了电力系统交流基本模型网络拓扑结构的特征。

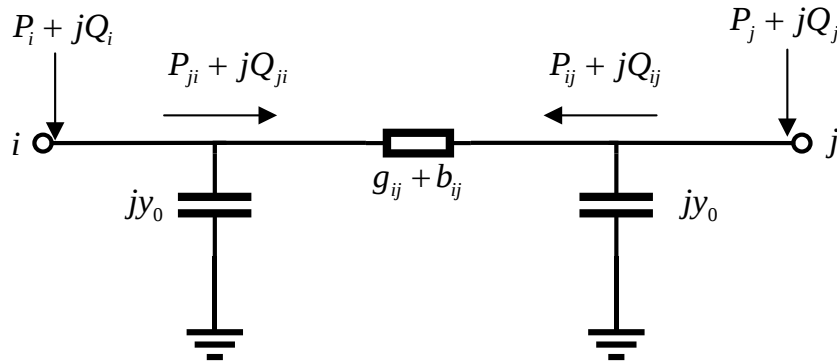


图 2-2 电力系统分支等效电路

具体数学模型为

$$P_i = V_i \sum_{j \in \Omega_i} V_j (g_{ij} \cos \theta_{ij} + b_{ij} \sin \theta_{ij}) \quad (2-2)$$

$$Q_i = V_i \sum_{j \in \Omega_i} V_j (g_{ij} \sin \theta_{ij} - b_{ij} \cos \theta_{ij}) \quad (2-3)$$

$$P_{ij} = V_i^2 (g_{si} + g_{ij}) - V_i V_j (g_{ij} \cos \theta_{ij} + b_{ij} \sin \theta_{ij}) \quad (2-4)$$

$$Q_{ij} = -V_i^2 (b_{si} + b_{ij}) - V_i V_j (g_{ij} \sin \theta_{ij} - b_{ij} \cos \theta_{ij}) \quad (2-5)$$

其中， P_i 和 Q_i 分别表示总线 i 上的有功功率注入和无功功率注入； P_{ij} 和 Q_{ij} 分别表示从总线 i 到总线 j 的有功功率流和无功功率流； V_i 表示总线 i 上的电压； θ_i 表示总线 i 上的相位角； θ_{ij} 表示总线 i 和总线 j 之间的相位差， $\theta_{ij} = \theta_i - \theta_j$ ； g_{ij} 和 b_{ij} 表示总线 i 和总线 j 之间的电纳和阻抗； y_0 为电力系统输电线路的充电电纳； Ω_i 是与总线 i 相邻的总线集。

2.2.2 交流状态估计

在电力系统中，交流状态估计的主要目标是通过利用来自电网的 m 维观测值 $\mathbf{z}$ 估计 n 维状态向量 $\mathbf{x}$ 中包含的电压幅值和相角等重要电力系统状态参数。目前电力系统常用加权最小二乘法 (Weighted Least Square, WLS) 来进行状态估计。利用加权最小二乘估计器最小化式 (2-6) 中的目标函数。

$$J(\mathbf{x}) = \sum_{i=1}^m (z_i - h_i(\mathbf{x}))^2 / R_{ii} = [\mathbf{z} - \mathbf{h}(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{z} - \mathbf{h}(\mathbf{x})] \quad (2-6)$$

其中， $\mathbf{R} = \text{Cov}(\mathbf{e}) = \text{diag}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2\}$ 。

对目标函数取一阶偏导数并令其等于零，通过这个过程可以得到一阶最优条件。具体步骤为

$$g(\mathbf{x}) = \frac{\partial J(\mathbf{x})}{\partial \mathbf{x}} = -\mathbf{H}^T(\mathbf{x}) \mathbf{R}^{-1} [\mathbf{z} - \mathbf{h}(\mathbf{x})] = 0 \quad (2-7)$$

其中， $\mathbf{H}(\mathbf{x}) = \frac{\partial \mathbf{h}(\mathbf{x})}{\partial \mathbf{x}}$ 。

将非线性函数 $g(\mathbf{x})$ 展开成围绕状态向量 $\mathbf{x}^k$ 的泰勒级数，得到

$$g(\mathbf{x}) = g(\mathbf{x}^k) + G(\mathbf{x}^k)(\mathbf{x} - \mathbf{x}^k) + \dots = 0 \quad (2-8)$$

忽略高阶项，那么采用 Gauss-Newton 法^[44]的迭代步骤可以表达为

$$\mathbf{x}^{k+1} = \mathbf{x}^k - [G(\mathbf{x}^k)]^{-1} g(\mathbf{x}^k) \quad (2-9)$$

$$G(\mathbf{x}^k) = \frac{\partial g(\mathbf{x}^k)}{\partial \mathbf{x}} = H^T(\mathbf{x}^k) R^{-1} H(\mathbf{x}^k) \quad (2-10)$$

$$g(\mathbf{x}^k) = -H^T(\mathbf{x}^k) R^{-1} [\mathbf{z} - h(\mathbf{x}^k)] \quad (2-11)$$

其中， k 为迭代次数， $\mathbf{x}^k$ 为第 k 次迭代时的状态向量。因此，第 $k+1$ 次迭代时的状态向量为

$$\mathbf{x}^{k+1} = \mathbf{x}^k + [H^T(\mathbf{x}^k) R^{-1} H(\mathbf{x}^k)]^{-1} H^T(\mathbf{x}^k) R^{-1} [\mathbf{z} - h(\mathbf{x}^k)] \quad (2-12)$$

2.2.3 直流状态估计

在直流电力系统模型中，假定系统中所有节点的电压幅值均相等，且为 1 标么值。同时，相邻节点的电压相角差被假设非常小，仅考虑线路的电抗，则不存在无功功率的测量值，系统的状态变量仅包括各节点的电压相角。因此，测量模型简化为

$$\mathbf{z} = H\mathbf{x} + \mathbf{e} \quad (2-13)$$

构建的目标函数为

$$J(\mathbf{x}) = [\mathbf{z} - H\mathbf{x}]^T \mathbf{R}^{-1} [\mathbf{z} - H\mathbf{x}] \quad (2-14)$$

通过加权最小二乘估计器得到的估计值为

$$\hat{\mathbf{x}} = [H^T \mathbf{R}^{-1} H]^{-1} H^T \mathbf{R}^{-1} \mathbf{z} \quad (2-15)$$

2.3 传统不良数据检测方法

2.3.1 不良数据相关概念

在电力系统中，不良数据是指与真实电力系统状态不符或可能导致系统错误

决策或故障的数据。这些数据可能由多种因素引起，以下是产生不良数据的一些常见原因：

（1）设备故障：传感器、测量设备或通信设备出现故障，导致采集到的数据不准确或缺失。

（2）通信问题：数据在传输过程中发生错误，可能由于网络故障、通信协议问题或恶劣的通信条件引起。

（3）人为错误：操作员或系统用户误操作或错误配置系统，导致错误的数据输入或处理。

（4）恶意攻击：攻击者可能通过操纵数据或对通信通道进行攻击，引入虚假的信息，干扰智能电网系统的正常运行。

（5）环境因素：不良天气、自然灾害等环境因素可能导致设备损坏或数据采集异常。

（6）数据干扰：不良数据可能在数据采集过程中受到其他电磁干扰或噪声的影响。

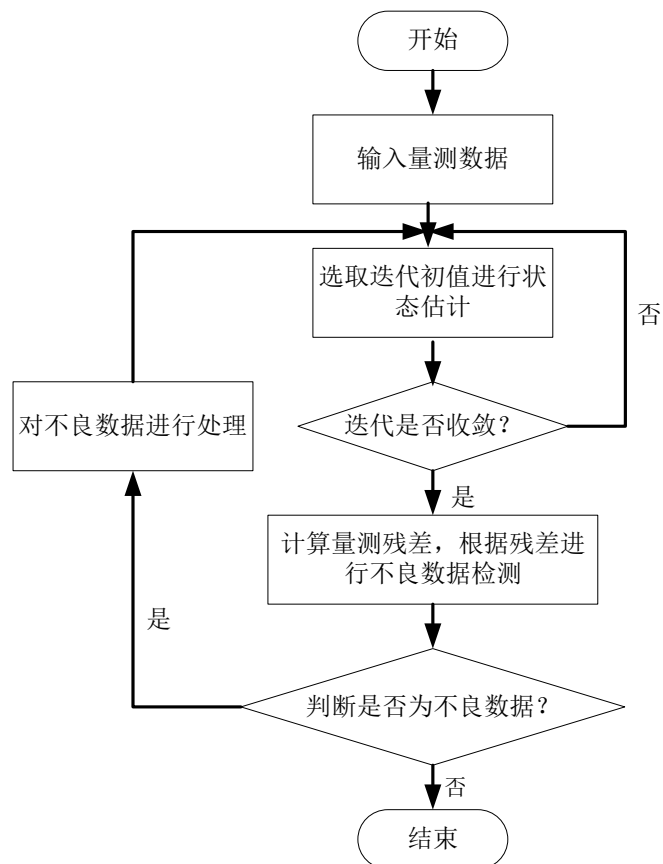


图 2-3 不良数据检测基本流程图

电力系统中存在的不良数据给状态估计带来严峻挑战。状态估计在确保电力系统的稳定运行中扮演着至关重要的角色，然而，当系统中存在不良数据时，状态估计的准确性和可靠性就面临严重的威胁。为确保电力系统状态估计的准确性和可靠性，需要采取一系列有效的措施。传统的不良数据检测方法有目标函数 $J(\hat{\mathbf{x}})$ 检测法、加权残差 r_w 检测法、标准化残差 r_N 检测法以及量测量突变检测法。传统的不良数据检测方法流程图如图 2-3 所示。

2.3.2 残差方程

在电力系统中，量测残差被定义为实际测量值与量测估计值之间的差异。传统的不良数据检测方法主要集中在对量测残差的处理上，通过观察量测残差的大小和分布，可以评估状态估计的可靠性和准确性。

利用最小二乘算法得到的状态估计量为 $\hat{\mathbf{x}}$ ，则量测估计量 $\hat{\mathbf{z}}$ 为

$$\hat{\mathbf{z}} = h(\hat{\mathbf{x}}) \quad (2-16)$$

将式 (2-16) 进行线性化处理可得

$$\hat{\mathbf{z}} \approx h(\mathbf{x}) + \mathbf{H}(\mathbf{x} - \hat{\mathbf{x}}) = h(\mathbf{x}) + \mathbf{H}\tilde{\mathbf{x}} \quad (2-17)$$

那么量测误差 $\tilde{\mathbf{z}}$ 可表示为

$$\tilde{\mathbf{z}} = h(\mathbf{x}) - \hat{\mathbf{z}} = -\mathbf{H}\tilde{\mathbf{x}} \quad (2-18)$$

其中， $h(\mathbf{x})$ 为量测真值向量； $\tilde{\mathbf{x}}$ 为状态估计误差向量。

量测残差向量定义为量测向量与量测估计向量之差，即 $\mathbf{r} = \mathbf{z} - \hat{\mathbf{z}}$ ，其中 $\mathbf{z} = h(\mathbf{x}) + \mathbf{e}$ ，通过计算可得

$$\begin{aligned} \mathbf{r} &= \mathbf{z} - \hat{\mathbf{z}} = h(\mathbf{x}) + \mathbf{e} - h(\mathbf{x}) - \mathbf{H}\tilde{\mathbf{x}} \\ &= \mathbf{e} - \mathbf{H}(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{e} \\ &= [\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1}] \mathbf{e} \\ &= \mathbf{W} \mathbf{e} \end{aligned} \quad (2-19)$$

及

$$\mathbf{W} = \mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \quad (2-20)$$

其中， $\mathbf{W}$ 为残差灵敏度矩阵， $\mathbf{I}$ 为单位矩阵。式 (2-19) 为残差方程。

2.3.3 目标函数极值检测法

在目标函数 $J(\hat{\mathbf{x}})$ 检测方法中，通常使用假设检验的方法来判断 $J(\hat{\mathbf{x}})$ 的值是否超出正常范围，以确定量测值中是否存在不良数据。

将状态估计值 $\hat{\mathbf{x}}$ 代入目标函数中，得到目标函数极值 $J(\hat{\mathbf{x}})$ 为

$$J(\hat{\mathbf{x}}) = [\mathbf{z} - \mathbf{h}(\hat{\mathbf{x}})]^T \mathbf{R}^{-1} [\mathbf{z} - \mathbf{h}(\hat{\mathbf{x}})] = \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r} \quad (2-21)$$

在正常情况下，假设量测误差服从正态分布，并且在无不良数据的情况下，目标函数 $J(\hat{\mathbf{x}})$ 服从自由度为 K 的卡方分布。假设有 m 个量测量， n 个状态变量，那么 $J(\hat{\mathbf{x}})$ 的数学期望为 $m - n = K$ ，方差为 $2K$ 。当存在不良数据时， $J(\hat{\mathbf{x}})$ 的数学期望变为 $K + \alpha_{w_i}^2 w_{w,ii}$ ，方差变为 $2K + 4\alpha_{w_i}^2 w_{w,ii}$ 。其中， α_{w_i} 为不良数据的值， $w_{w,ii}$ 为加权残差灵敏度矩阵的对角元。

使用目标函数 $J(\hat{\mathbf{x}})$ 进行不良数据检测时，可以采用假设检验的方法。

$$\begin{cases} \mathbf{H}_0 \text{假设: } J(\hat{\mathbf{x}}) < \gamma_j, \text{无不良数据, } \mathbf{H}_0 \text{属真;} \\ \mathbf{H}_1 \text{假设: } J(\hat{\mathbf{x}}) \geq \gamma_j, \text{有不良数据, } \mathbf{H}_0 \text{不真。} \end{cases} \quad (2-22)$$

其中， $\mathbf{H}_0$ 为零假设； $\mathbf{H}_1$ 为备择假设， γ_j 为检测阈值，对应于某一自由度为 K 的卡方分布。

目标函数 $J(\hat{\mathbf{x}})$ 检测法是一种全局性的不良数据检测方法，其主要作用是判断整个量测数据集中是否存在不良数据，而无法精确定位到具体是哪一个数据点出现了异常。尽管该方法能够指示整体系统可能存在问题，但却无法提供对个别异常数据点的具体定位信息。

2.3.4 加权残差检测法

加权残差定义为

$$\mathbf{r}_w = \sqrt{\mathbf{R}^{-1}} \mathbf{e} = \mathbf{W}_w \mathbf{e}_w \quad (2-23)$$

式中的 $\mathbf{W}_w$ 为加权残差灵敏度矩阵，表示为

$$\mathbf{W}_w = \sqrt{\mathbf{R}^{-1}} \mathbf{W} \sqrt{\mathbf{R}} = \mathbf{I} - \sqrt{\mathbf{R}^{-1}} \mathbf{H} (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \sqrt{\mathbf{R}^{-1}} \quad (2-24)$$

加权残差检测方法是通过逐维地对量测量应用假设检验，可表示为

$$\begin{cases} \mathbf{H}_0 \text{假设: } |\mathbf{r}_{w,i}| < \gamma_{w,i}, \text{无不良数据, } \mathbf{H}_0 \text{属真, 接受} \mathbf{H}_0 ; \\ \mathbf{H}_1 \text{假设: } |\mathbf{r}_{w,i}| \geq \gamma_{w,i}, \text{有不良数据, } \mathbf{H}_0 \text{不真, 接受} \mathbf{H}_1 。 \end{cases} \quad (2-25)$$

其中， $\mathbf{r}_{w,i}$ 为第 i 个加权残差， $\gamma_{w,i}$ 为第 i 个加权残差的检测阈值。在正常量测条件下，加权残差呈现零均值的正态分布。如果误检概率为 0.005，则正常加权残差取值范围为

$$|\mathbf{r}_{w,i}| < 2.81 \sqrt{\mathbf{W}_{w,ii}} \quad (i=1, 2, \dots, m) \quad (2-26)$$

那么，加权残差的检测阈值 $\gamma_{w,i}$ 为

$$\gamma_{w,i} = 2.81 \sqrt{\mathbf{W}_{w,ii}} \quad (i=1, 2, \dots, m) \quad (2-27)$$

加权残差 $\mathbf{r}_w$ 检测法的优势在于它综合考虑了量测值的相对影响，更准确地反映了量测值的异常性。

2.3.5 标准化残差检测法

标准化残差定义为

$$\mathbf{r}_N = \sqrt{\mathbf{D}^{-1}} \mathbf{r} \quad (2-28)$$

其中，对角阵 $\mathbf{D} = \text{diag}[\Sigma_r] = \text{diag}[\mathbf{W}\mathbf{R}]$ 。标准化残差灵敏度矩阵则为

$$\mathbf{W}_N = \sqrt{\mathbf{D}^{-1}} \mathbf{W} \quad (2-29)$$

与加权残差 $\mathbf{r}_w$ 检测法类似，标准化残差 $\mathbf{r}_N$ 检测方法也采用假设检验的方式进行异常检测，可表示为

$$\begin{cases} \mathbf{H}_0 \text{假设: } |\mathbf{r}_{N,i}| < \gamma_{N,i}, \text{无不良数据, } \mathbf{H}_0 \text{属真, 接受} \mathbf{H}_0 ; \\ \mathbf{H}_1 \text{假设: } |\mathbf{r}_{N,i}| \geq \gamma_{N,i}, \text{有不良数据, } \mathbf{H}_0 \text{不真, 接受} \mathbf{H}_1 。 \end{cases} \quad (2-30)$$

其中， $\mathbf{r}_{N,i}$ 为第 i 个标准化残差； $\gamma_{N,i}$ 为第 i 个标准化残差的检测阈值。当误检概率为 0.005 时，标准化残差的检测阈值 $\gamma_{N,i}$ 同样为 2.81。

在数据检测过程中，加权残差 $\mathbf{r}_w$ 检测法与标准化残差 $\mathbf{r}_N$ 检测法是常用的不良数据检测方法，它们在识别异常值和异常模式方面具有一定优势。然而，这些方法在实际应用中常常面临一些挑战，其中包括“残差污染”和“残差淹没”等问题，这些问题可能对检测方法的准确性和稳健性造成不良影响。

2.3.6 量测突变检测法

为了有效应对“残差淹没”导致的检测困难，充分利用前一采样时刻的测量信息被认为是一种有效的方法。量测突变检测法是一种用于监测电力系统量测数据中是否存在异常变化的方法。该方法通过比较相邻两次采样时刻的量测信息，判断下一时刻的量测数据是否正常。其基本假设包括：相邻两次采样时间间隔内，电力系统的网络拓扑结构未发生变化，并且前一时刻的量测数据中不存在异常数据或者已经得到了处理。

若某一测量值发生异常突变，即超出了预先设定的阈值，系统将识别其为可疑数据。判定该量测量发生异常突变的依据为

$$|\Delta \mathbf{z}_{t,i}| = |\mathbf{z}_{t,i} - \mathbf{z}_{t|t-1,i}| \geq \gamma_i \quad (2-31)$$

其中， $\mathbf{z}_{t,i}$ 为 t 时刻的第 i 个实际测量值； $\mathbf{z}_{t|t-1,i}$ 为基于 $t-1$ 时刻第 i 个测量值对 t 时刻的预测值； $\Delta \mathbf{z}_{t,i}$ 为第 i 个量测增量； γ_i 为检测阈值。

将前一时刻的测量值表示为 $\mathbf{z}_{t-1}$ ，当前时刻的测量值表示为 $\mathbf{z}_t$ ，那么 t 时刻的测量值 $\mathbf{z}_t$ 为

$$\mathbf{z}_t = h(\mathbf{x}_t) + \mathbf{e}_t = h(\mathbf{x}_{t-1}) + H(\mathbf{x}_{t-1})\Delta_{t-1} + \mathbf{e}_t \quad (2-32)$$

$t-1$ 时刻的测量值 $\mathbf{z}_{t-1}$ 为 $\mathbf{z}_{t-1} = h(\mathbf{x}_{t-1}) + \mathbf{e}_{t-1}$ ，代入式（2-32）可得

$$\mathbf{z}_t - \mathbf{z}_{t-1} - H(\mathbf{x}_{t-1})\Delta_{t-1} = \mathbf{e}_t - \mathbf{e}_{t-1} \quad (2-33)$$

在表示两个采样时刻的测量误差时，使用了 $\mathbf{e}_t$ 和 $\mathbf{e}_{t-1}$ 两个变量。这两个变量是对于同一量测数据的误差，它们均呈现相同的正态分布。因此，由（2-33）式可得

$$\Delta \mathbf{z}_{t,i}^N = \frac{\mathbf{z}_{t,i} - \mathbf{z}_{t-1,i} - H_i(\mathbf{x}_{t-1})\Delta_{i-1,i}}{\sqrt{2}\sigma_i} \sim \mathcal{N}(0,1), i=1,2,\dots,m \quad (2-34)$$

同样利用假设检验的方法进行检测，可表示为

$$\begin{cases} \mathbf{H}_0 \text{假设: } |\Delta \mathbf{z}_{t,i}^N| < \gamma_i, \text{无不良数据, } \mathbf{H}_0 \text{ 属真, 接受 } \mathbf{H}_0 ; \\ \mathbf{H}_1 \text{假设: } |\Delta \mathbf{z}_{t,i}^N| \geq \gamma_i, \text{有不良数据, } \mathbf{H}_0 \text{ 不真, 接受 } \mathbf{H}_1 。 \end{cases} \quad (2-35)$$

当误检概率为 0.005 时，检测阈值 γ_i 同样为 2.81。

量测突变检测法虽然在监测电力系统中的异常变化方面具有一定的优势，但也存在一些缺点。该方法在假设相邻两次采样时间内电网拓扑结构保持不变的前提下才能有效运作，如果系统中的拓扑结构发生变化，可能导致误报或漏报。并且量测突变检测法通常假定前一时刻的量测数据是准确的，没有测量误差。然而，在实际情况中，测量误差可能存在，这可能影响到对不良数据的准确判断。

2.4 本章小结

本章主要介绍了智能电网状态估计以及传统的不良数据检测方法的基本概念。首先分析了常用的加权最小二乘估计算法，其次介绍了基于残差的不良数据检测算法，详细推导了不同检测方法的数学描述过程。不仅如此，还详细阐述了每种检测方法的工作原理，并对它们各自的优缺点进行了深入剖析。本章的内容不仅为后续章节介绍不同研究方法提供了坚实的理论基础，同时也在思想上为进一步研究提供了重要的启示。

第3章 基于改进扩展卡尔曼滤波算法的不良数据检测方法

3.1 引言

在智能电网的运行和监控中，状态估计发挥着关键作用，能够提供实时监控系统运行状态的电力系统状态信息，为决策制定、调度优化以及故障诊断和定位提供支持。智能电网的状态估计可视为智能电网运行和监控的中枢神经，为运营人员和调度员提供准确的电力系统状态信息，确保电力系统稳定运行和最优调度。但智能电网中不良数据的存在可能会影响电力系统状态估计的精度，从而影响电力系统的稳定运行。如果不及时发现和处理这些不良数据，可能会误导操作人员对电力系统状态的判断，导致错误的运行决策，进而影响电力系统的安全经济运行。因此，电力系统不良数据检测至关重要。

然而，由于系统非线性和高度动态的特性，传统的扩展卡尔曼滤波(Extended Kalman Filter, EKF)在应对智能电网中的非线性系统动力学时可能面临一些限制。为了提高状态估计的精度和鲁棒性，本章提出一种基于投影统计(Projection Statistics, PS)的改进扩展卡尔曼滤波算法，以便有效地检测和处理智能电网中的不良数据。

3.2 扩展卡尔曼滤波算法原理

3.2.1 卡尔曼滤波算法

卡尔曼滤波算法是一种用于估计动态系统状态的递归算法，卡尔曼滤波的核心思想在于利用当前时刻的状态对下一时刻状态的进行预测。首先进行先验估计，获取一个初始的状态值。接着，将这个先验估计与实际的测量结果相结合，通过卡尔曼滤波增益的计算，得到下一时刻更为准确的状态估计，即后验估计。并通过卡尔曼滤波增益的调整，使后验估计无限接近系统状态的实际值。这样，通过不断的迭代，卡尔曼滤波能够有效地融合先验估计和测量信息，提供对系统状态的最优估计。该算法适用于线性系统，状态方程和状态转移方程分别为

$$\mathbf{z}_t = \mathbf{H}\mathbf{x}_t + \mathbf{v}_t \quad (3-1)$$

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\mathbf{u}_t + \boldsymbol{\omega}_t \quad (3-2)$$

其中， $\mathbf{x}_t$ 为系统在 t 时刻的状态实际值； $\mathbf{z}_t$ 为系统在 t 时刻的测量向量； $\mathbf{A}$ 为状态转移矩阵； $\mathbf{B}$ 为外部输入矩阵； $\mathbf{u}_t$ 为外部输入向量； $\mathbf{H}$ 为观测矩阵；而 $\boldsymbol{\omega}_t$ 和 $\mathbf{v}_t$ 分别为系统的过程噪声和观测噪声； $\mathbf{Q}$ 和 $\mathbf{R}$ 分别为过程噪声的协方差矩阵和观测噪声的协方差矩阵。

卡尔曼滤波的基本原理包括预测和更新两个步骤。在预测步骤中，系统状态通过动态模型进行预测

$$\hat{\mathbf{x}}_{t|t-1} = \mathbf{A}\hat{\mathbf{x}}_{t-1|t-1} + \mathbf{B}\mathbf{u}_t \quad (3-3)$$

$$\mathbf{P}_{t|t-1} = \mathbf{A}\mathbf{P}_{t-1|t-1}\mathbf{A}^T + \mathbf{Q} \quad (3-4)$$

其中， $\hat{\mathbf{x}}_{t-1|t-1}$ 为在 $t-1$ 时刻的状态估计值； $\hat{\mathbf{x}}_{t|t-1}$ 为在 t 时刻的状态预测值； $\mathbf{P}_{t|t-1}$ 是在 $t-1$ 时刻基于先前信息和系统模型进行的对 t 时刻系统状态的预测误差协方差矩阵。

在更新步骤中，在接收到新的测量值时，通过测量更新修正预测的状态估计。这一步基于测量方程和卡尔曼增益 $\mathbf{K}_t$ ，可表示为

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{H}^T(\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^T + \mathbf{R})^{-1} \quad (3-5)$$

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{z}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1}) \quad (3-6)$$

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H})\mathbf{P}_{t|t-1} \quad (3-7)$$

以上步骤形成了卡尔曼滤波算法的迭代过程。卡尔曼滤波算法框图如图 3-1 所示。

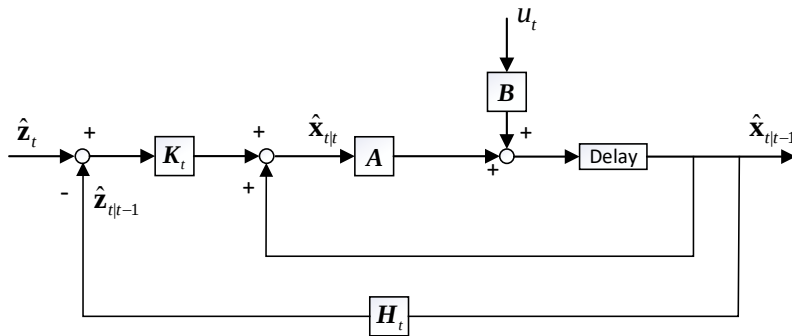


图 3-1 卡尔曼滤波算法框图

通过不断的时间更新和测量更新，卡尔曼滤波算法可以递归地提供系统状态的最优估计，同时考虑了过程噪声和测量噪声的影响，以及先前状态估计的不确定性。算法的优势在于其最小均方误差的估计原则，使其在动态系统的状态估计问题中得到广泛应用。需要注意的是，卡尔曼滤波算法的前提是系统是线性的，而在非线性系统中，扩展卡尔曼滤波等算法可能更为适用。

3.2.2 扩展卡尔曼滤波算法

扩展卡尔曼滤波算法是卡尔曼滤波的一种变种，用于处理非线性系统。与标准的卡尔曼滤波适用于线性系统不同，扩展卡尔曼滤波算法通过对非线性系统的状态转移方程和测量方程进行线性化处理，再基于基本卡尔曼滤波原理进行状态估计。式(2-1)表明了测量向量 $\mathbf{z}$ 与状态向量 $\mathbf{x}$ 之间的非线性相关关系，因此，状态转移方程和用于动态估计的测量方程为

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t) + \mathbf{w}_t \quad (3-8)$$

$$\mathbf{z}_t = h(\mathbf{x}_t) + \mathbf{e}_t \quad (3-9)$$

其中， t 表示采样时间， $\mathbf{x}_t$ 表示 t 时刻的状态向量， $\mathbf{z}_t$ 表示时刻 t 的测量向量， $f(\cdot)$ 和 $h(\cdot)$ 分别表示状态转移函数和测量函数。 $\mathbf{w}_t$ 和 $\mathbf{e}_t$ 是均值为零、协方差分别为 $\mathbf{Q}$ 和 $\mathbf{R}$ 的不相关高斯白噪声。扩展卡尔曼滤波算法是解决非线性状态估计问题最直接的方法之一，其状态估计过程可以分解为两个关键步骤。

(1) 状态预测步

进行整个电力系统状态的一步预测，同时计算与该预测相对应的状态预测误差的协方差矩阵

$$\hat{\mathbf{x}}_t^- = \mathbf{F}_{t-1} \hat{\mathbf{x}}_{t-1}^+ \quad (3-10)$$

$$\mathbf{P}_t^- = \mathbf{F}_{t-1} \mathbf{P}_{t-1} \mathbf{F}_{t-1}^T + \mathbf{Q}_{t-1} \quad (3-11)$$

(2) 状态更新步

通过以下公式计算卡尔曼滤波器的增益，并更新状态估计以及估计误差协方差矩阵

$$\hat{\mathbf{x}}_t^+ = \hat{\mathbf{x}}_t^- + \mathbf{K}_t [\mathbf{z}_t - h(\hat{\mathbf{x}}_t^-)] \quad (3-12)$$

$$K_t = P_t^- H_t^T (H_t P_t^- H_t^T + R_t)^{-1} \quad (3-13)$$

$$P_t^+ = (I - K_t H_t) P_t^- \quad (3-14)$$

其中， $\hat{\mathbf{x}}_t^+$ 和 $\hat{\mathbf{x}}_t^-$ 分别表示 t 时刻的状态估计向量和状态预测向量； $F_{t-1} = \partial f(\mathbf{x}_{t-1}) / \partial \mathbf{x}_{t-1}$ 是 $t-1$ 时刻状态方程的雅可比矩阵； $H_t = \partial h(\mathbf{x}_t) / \partial \mathbf{x}_t$ 是 t 时刻测量方程的雅可比矩阵。 P_t^- 和 P_t^+ 分别表示 t 时刻的状态预测误差协方差矩阵和误差协方差矩阵。 K_t 表示 t 时刻的卡尔曼增益。 $\mathbf{r}^- = \mathbf{z}_t - h(\hat{\mathbf{x}}_t^-)$ 被定义为 t 时刻的预测估计残差。图 3-2 为扩展卡尔曼滤波算法流程图。

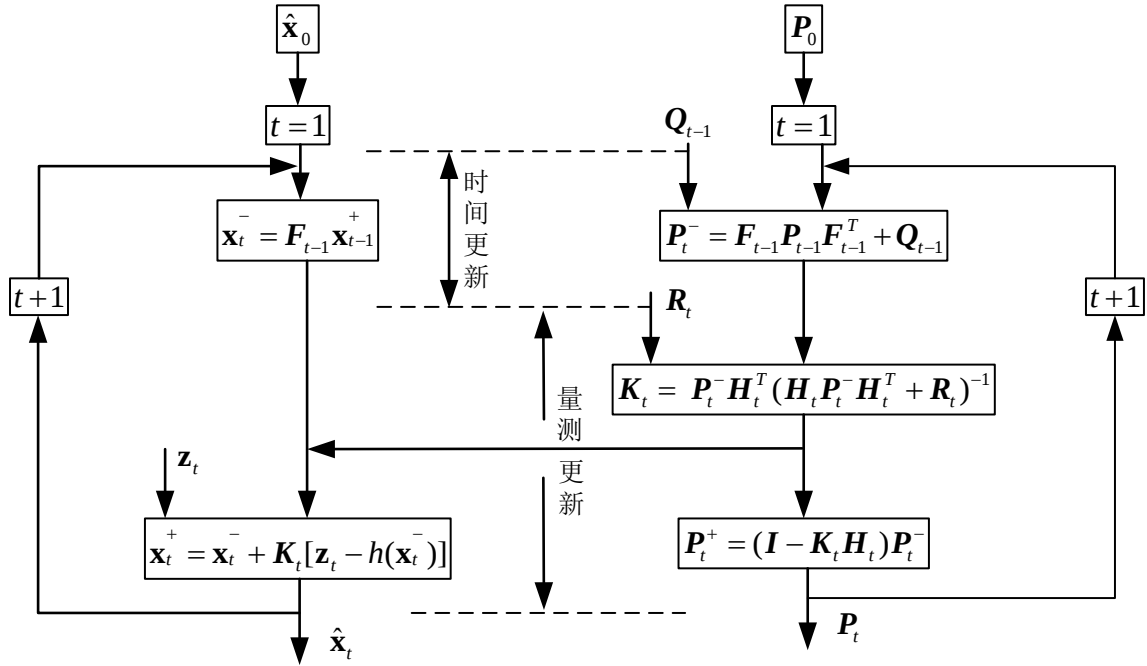


图 3-2 扩展卡尔曼滤波算法流程图

3.3 投影统计算法原理

3.3.1 马氏距离

马氏距离（Mahalanobis Distance）^{[45]-[46]}是由印度统计学家帕尔特里亚·丘达姆巴拉姆·马哈拉诺比斯（P. C. Mahalanobis）于 1936 年提出的。在电力系统运行中，通过使用马氏距离来监测系统异常情况，可以提前发现潜在的故障或异

常。具体而言，如果实时测量值与正常运行状态之间的马氏距离超过预先设定的阈值，可能表明系统存在潜在的故障或异常，需要进行进一步的调查和诊断。

考虑由 m 维向量集 $\{\mathbf{z}_1, \dots, \mathbf{z}_l\}$ 定义的 l 个测量向量，假设集合 $\{\mathbf{z}_1, \dots, \mathbf{z}_l\}$ 服从均值为 μ ，协方差矩阵为 Σ 的多元正态分布 $\mathcal{N}(\mu, \Sigma)$ 。

马氏距离的计算公式为

$$MD_i = \sqrt{(\mathbf{z}_i - \mu)^T \Sigma^{-1} (\mathbf{z}_i - \mu)} \quad (3-15)$$

其中， $\mathbf{z}_i$ 为第 i 个测量向量； MD_i 为 $\mathbf{z}_i$ 的马氏距离。 MD_i^2 近似符合自由度为 n 的卡方分布 χ^2 。将马氏距离大于阈值的所有测量向量标记为不良数据，即当 $MD_i^2 \geq \chi_{n,0.975}^2$ 时，该测量向量为不良数据。马氏距离考虑了数据的协方差矩阵，因此能够更准确地测量多维数据点之间的距离，考虑了各个维度之间的相关性。

3.3.2 投影统计

投影统计 (Projection Statistics, PS) 的公式灵感来源于马氏距离的计算方式。在马氏距离中，通过考虑数据点与均值之间的关系以及协方差矩阵，得到了一种有效的多维度距离度量。在投影统计中，借鉴了这一思想，并通过将数据点投影到某个方向上，定义了一种投影统计的公式

$$PS_i = \max_{\|\phi\|=1} \frac{|\mathbf{z}_i^T \phi - \text{med}_j(\mathbf{z}_j^T \phi)|}{S_m} \quad (3-16)$$

其中， $\mathbf{z}_i$ 为 m 维向量集 $\{\mathbf{z}_1, \dots, \mathbf{z}_l\}$ 中第 i 个测量向量； u_j 为投影方向； ϕ 为标准化投影方向； $\text{med}_j(\mathbf{z}_j^T \phi)$ 为数据集的中位数； $S_m = 1.4826 \text{med}_i |\mathbf{z}_i^T \phi - \text{med}_j(\mathbf{z}_j^T \phi)|$ 为中位数绝对偏差。

投影统计算法的主要步骤如下：

步骤一：首先，对于 m 维数据集中的第 i 个测量向量 $\mathbf{z}_i$ ，计算其分量中位数

$$M = \{\text{med}_{j=1, \dots, l}(z_{j1}), \dots, \text{med}_{j=1, \dots, l}(z_{jl})\} \quad (3-17)$$

其中， l 表示为数据集中测量向量的个数。 $\text{med}_{j=1, \dots, l}(z_{jl})$ 表示对数据集中所有测量向量的第 l 个分量 z_{jl} 取中位数。

步骤二：选择一个特定的投影方向，通常是从数据集的中位数出发，沿着每个数据点的方向进行投影，计算投影方向

$$u_j = \mathbf{z}_j - M; j = 1, \dots, l \quad (3-18)$$

步骤三：标准化 u_j 得到

$$\varphi_j = \frac{u_j}{\|u_j\|} = \frac{u_j}{\sqrt{u_{j1}^2 + \dots + u_{jm}^2}}; j = 1, \dots, l \quad (3-19)$$

步骤四：计算向量集 $\{\mathbf{z}_1, \dots, \mathbf{z}_l\}$ 在 φ_j 上的标准化投影，即

$$\gamma_{1j} = \mathbf{z}_1^T \varphi_j; \gamma_{2j} = \mathbf{z}_2^T \varphi_j; \dots; \gamma_{lj} = \mathbf{z}_l^T \varphi_j \quad (3-20)$$

步骤五：计算 $\{\gamma_{1j}, \dots, \gamma_{lj}\}$ 的中位数 $\gamma_{med,j}$

步骤六：计算中位数绝对偏差 S_m

$$S_m = 1.4826 \cdot med_i |\gamma_{ij} - \gamma_{med,j}| \quad (3-21)$$

步骤七：计算标准化投影

$$PS_{ij} = \frac{|\gamma_{ij} - \gamma_{med,j}|}{1.4826 \cdot med_i |\gamma_{ij} - \gamma_{med,j}|}; i = 1, \dots, l \quad (3-22)$$

步骤八：重复步骤 4 至步骤 7，以获取标准化投影 $\{PS_{i1}, \dots, PS_{il}\}$ ，其中 $i = 1, \dots, l$ 。

步骤九：计算投影统计量

$$PS_i = \max\{PS_{i1}, \dots, PS_{il}\}; i = 1, \dots, l \quad (3-23)$$

步骤十：设定一个阈值，用于确定哪些标准化投影值被视为异常。通过比较每个数据点的标准化投影值与设定的不良数据阈值，将超过阈值的数据点标记为不良数据。

投影统计方法使用中位数和中位数绝对偏差（MAD）等鲁棒性统计量，相较于均值和标准差，更能有效地处理数据集中存在的离群值。投影统计方法的输出，即标准化投影值，是相对于中位数和中位数绝对偏差的度量，这些指标对于异常值的影响相对较小，这使得它在高维数据中更具鲁棒性。

3.4 改进扩展卡尔曼滤波算法

3.4.1 面向不良数据的状态估计

在电力系统的监测和测量过程中，由于传感器故障、通信问题或其他外部干扰导致的异常数据。当电力系统中存在不良数据时，测量值产生偏差，即

$$\mathbf{z}_a = \mathbf{z} + \mathbf{a} \quad (3-24)$$

其中， $\mathbf{z}_a$ 表示带有不良数据偏差的测量值， $\mathbf{z}$ 为在没有异常的情况下的真实测量值，而 $\mathbf{a}$ 为不良数据引起的偏差。

假设在 t 时刻，系统中出现了不良数据，那么在 EKF 算法中， t 时刻的状态估计可以表示为

$$\begin{aligned} \mathbf{x}_t^{a+} &= \mathbf{F} \mathbf{x}_{t-1}^+ + \mathbf{K}_t [\mathbf{z}_t^a - h(\mathbf{x}_t^-)] \\ &= \mathbf{F} \mathbf{x}_{t-1}^+ + \mathbf{K}_t [\mathbf{z}_t + \mathbf{a}_t - h(\mathbf{x}_t^-)] \\ &= \mathbf{x}_t^+ + \mathbf{K}_t \mathbf{a}_t \end{aligned} \quad (3-25)$$

假设在 t 时刻，由于不良数据引起的状态估计偏差为 $\Delta \mathbf{x}_t^{a+} = \mathbf{x}_t^{a+} - \mathbf{x}_t^+$ 因此， $t+1$ 时刻的状态估计表示如下

$$\begin{aligned} \mathbf{x}_{t+1}^{a+} &= \mathbf{F} \mathbf{x}_t^{a+} + \mathbf{K}_{t+1} [\mathbf{z}_{t+1}^a - h(\mathbf{x}_{t+1}^{a-})] \\ &= \mathbf{F} \mathbf{x}_t^{a+} + \mathbf{K}_{t+1} [\mathbf{z}_{t+1} + \mathbf{a}_{t+1} - h(\mathbf{x}_{t+1}^{a-})] \\ &= \mathbf{F} \Delta \mathbf{x}_t^{a+} + \mathbf{K}_{t+1} [\Delta h^a(\mathbf{x}_{t+1}^-) + \mathbf{a}_{t+1}] + \mathbf{x}_{t+1}^+ \end{aligned} \quad (3-26)$$

其中， $\Delta h^a(\mathbf{x}_{t+1}^-)$ 被定义为在 $t+1$ 时刻存在不良数据的情况下的预测估计测量偏差，即

$$\Delta h^a(\mathbf{x}_{t+1}^-) = h(\mathbf{x}_{t+1}^-) - h(\mathbf{x}_{t+1}^{a-}) \quad (3-27)$$

因此， $t+1$ 时刻不良数据的存在导致状态估计的偏差为

$$\mathbf{x}_{t+1}^{a+} - \mathbf{x}_{t+1}^+ = \mathbf{F} \Delta \mathbf{x}_t^{a+} + \mathbf{K}_{t+1} [\Delta h^a(\mathbf{x}_{t+1}^-) + \mathbf{a}_{t+1}] \quad (3-28)$$

当前时刻的状态估计偏差受到先前时刻状态估计偏差和不良数据存在的影响。由于不良数据的影响，获得的最优状态估计 $\mathbf{x}_{t+1}$ 与真实状态 $\mathbf{x}_t$ 之间存在显著偏差，这对电力系统的安全运行构成了重大风险。因此，考虑在预测步骤和更新

步骤之间引入不良数据检测方法，以提高状态估计的鲁棒性。

3.4.2 不良数据检测

在电力系统实际应用中，由于传感器故障、电力设备故障、损坏以及数据传输或处理中的错误等原因，可能导致观测值中出现观测异常值。在采用 EKF 算法时，发电机参数的不准确性（如饱和效应）、控制器故障（如励磁器、电力系统稳定器等）以及突发的系统过程噪声等因素可能导致状态预测变量的不正确估计，从而引发创新异常值。创新异常值是指观测值和状态预测向量之间的残差明显偏离了正常的期望行为。

为了检测这些不良数据，将投影统计（PS）应用于以下基于两个相邻时刻预测结果构建的检测矩阵 Φ_t

$$\Phi_t = \begin{bmatrix} \hat{\mathbf{x}}_{t-1}^- & \hat{\mathbf{x}}_t^- \\ z_{t-1} - h(\hat{\mathbf{x}}_{t-1}^-) & z_t - h(\hat{\mathbf{x}}_t^-) \end{bmatrix} \quad (3-29)$$

Φ_t 为二维矩阵，它们的坐标可以由 $\phi_i^T = [\phi_{i1}, \phi_{i2}]^T$ 给出， ϕ_i 是 Φ_t 的第 i 行向量。其中， $z_{t-1} - h(\hat{\mathbf{x}}_{t-1}^-)$ 和 $z_t - h(\hat{\mathbf{x}}_t^-)$ 为创新向量， $\hat{\mathbf{x}}_{t-1}^-$ 和 $\hat{\mathbf{x}}_t^-$ 分别为 $t-1$ 时刻和 t 时刻的状态预测向量。

创新向量和状态预测向量是电力系统响应的时间序列数据，它们之间应该存在强烈的时序相关性。正常情况下，它们的变化趋势应该保持一致。然而，如果出现不良数据，可能会破坏这种时序关系，为了检测这种异常，引入了矩阵 Φ_t 并应用投影统计算法

$$PS_i = \max_{\|\phi\|=1} \frac{|\phi_i^T \phi - \text{med}_j(\phi_j^T \phi)|}{1.4826 \text{med}_t |\phi_i^T \phi - \text{med}_j(\phi_j^T \phi)|} \quad (3-30)$$

其中， ϕ_i^T 是矩阵 Φ_t 的第 i 行向量。

计算出投影统计量后，将 PS_i 值与预先确定的阈值进行比较， PS_i 近似为一个卡方随机变量，其分布假设为符合 2 个自由度的卡方分布。因此，检测阈值设为 $\chi_{2,0.975}^2$ 。若 $PS_i > \chi_{2,0.975}^2$ ，则认为对应的测量向量为不良数据。

为了在 EKF 算法中有效地检测并处理不良数据，引入加权函数。当检测到不良数据时，EKF 算法通过应用加权函数来调整卡尔曼增益，从而为不良数据分

配较低的权重，以减少不良数据对估计结果的影响，提高系统鲁棒性。同时，对于正常的测量值，EKF 算法仍然保持较高的权值，以确保它们能够在状态估计中得到充分利用。从而保证了状态估计的准确性和稳定性。测量加权矩阵 Γ_t 可以视为测量噪声协方差矩阵 R 的逆矩阵，表示为 $\Gamma_t = R^{-1}$ 。引入的加权函数可表示为

$$(\Gamma_t^a)^{-1} = \Gamma_t^{-1} \cdot e^{|z_t - h(\bar{x}_t)|} \quad (3-31)$$

通过采用这种方法，EKF 算法可以更有效地适应不良数据的存在，提高状态估计的鲁棒性和准确性。因此，即使在处理包含不良数据的情况下，该算法依然能够获得可靠的结果。为了更好的理解采用改进 EKF 算法检测不良数据的算法流程，给出的算法流程图如图 3-3 所示。

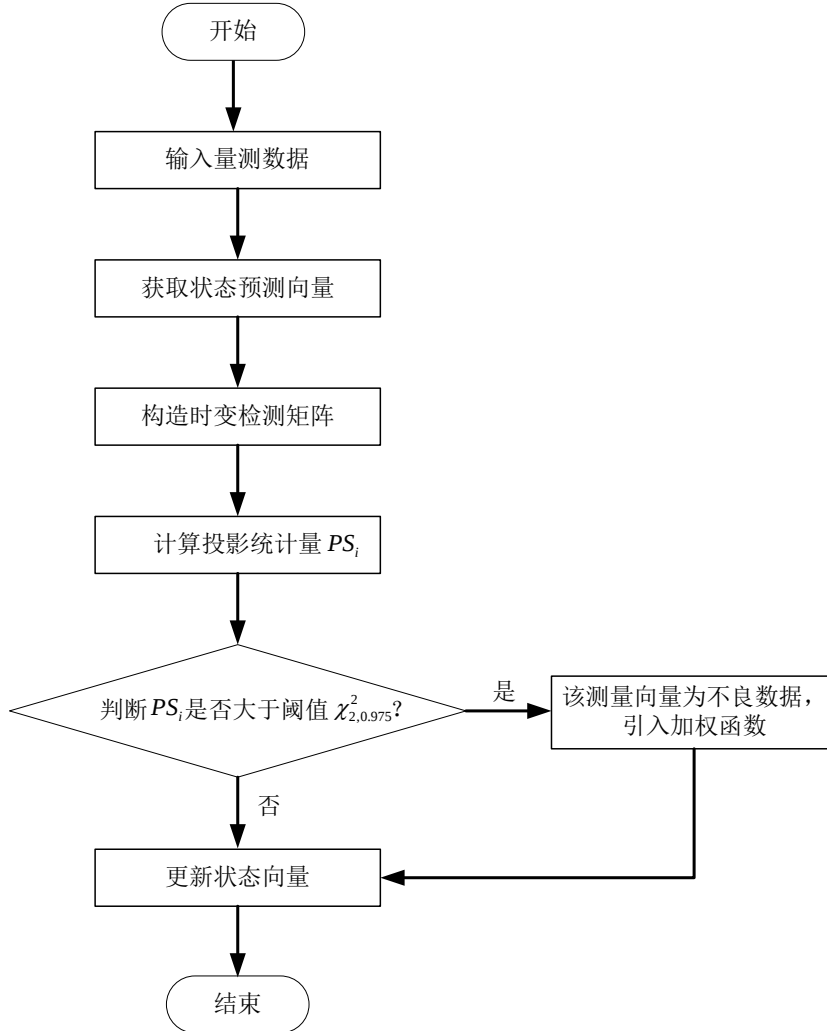


图 3-3 基于改进 EKF 算法的不良数据检测流程图

进行 IEEE-14 节点测试系统中仿真实验时，共使用了 40 个功率测量值，其中包括节点的有功功率注入 P_i 、无功功率注入 Q_i ，以及输电线路的有功功率流 P_{ij} 和无功功率流 Q_{ij} 。如表 3-1 和表 3-2 所示。

表 3-2 IEEE-14 系统输电线路功率流

起始节点 i	终止节点 j	有功功率 P_{ij}	无功功率 Q_{ij}
1	2	1.5792	-0.6622
2	3	0.7267	0.1923
4	2	-0.5486	-0.051
4	7	0.2707	-0.1540
4	9	0.1546	-0.0264
5	2	-0.4081	-0.0193
5	4	0.6006	-0.1006
5	6	0.4589	-0.2084
6	13	0.1834	0.0998
7	9	0.2707	0.1480
11	6	-0.0816	-0.0864
12	13	0.0188	0.0141

3.5.2 仿真结果分析

对于 IEEE-14 节点测试系统，假设 $\mathbf{a}$ 为不良数据产生的偏差向量，即 $\mathbf{a} = [\Delta P_3, \Delta Q_3, \Delta P_{1-2}, \Delta P_{2-3}, \Delta P_{4-2}, \Delta Q_{1-2}, \Delta Q_{2-3}, \Delta Q_{4-2}]^T$ 。其他正常测量值的增量均为零。图 3-5 展示了不良数据存在前后系统的数据分布变化。

为了检测和识别这些不良数据，采用了基于投影统计量的方法。通过 100 次采样，可以获得足够多的数据样本来计算投影统计量。利用公式 (3-30)，可以计算出每个样本的投影统计量值。这些投影统计量值反映了样本数据与正常数据分布之间的差异程度。接下来，将这些投影统计量与卡方检验的阈值 $\gamma = \chi_{2,0.975}^2$ 进行比较，卡方表中的检测阈值设定为 7.378。

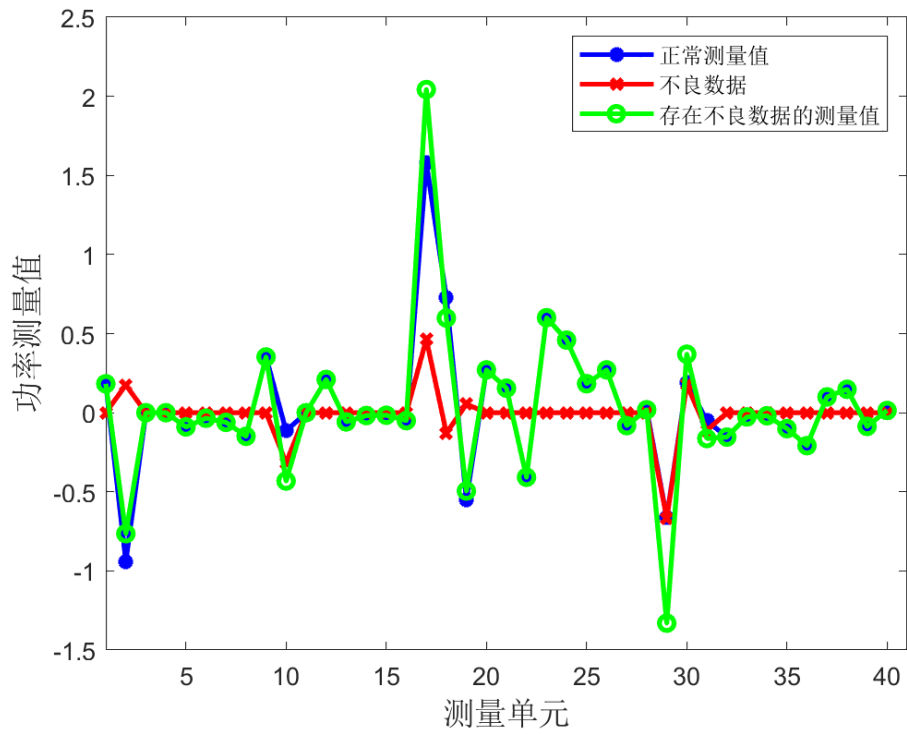


图 3-5 不良数据存在前后系统的数据分布变化

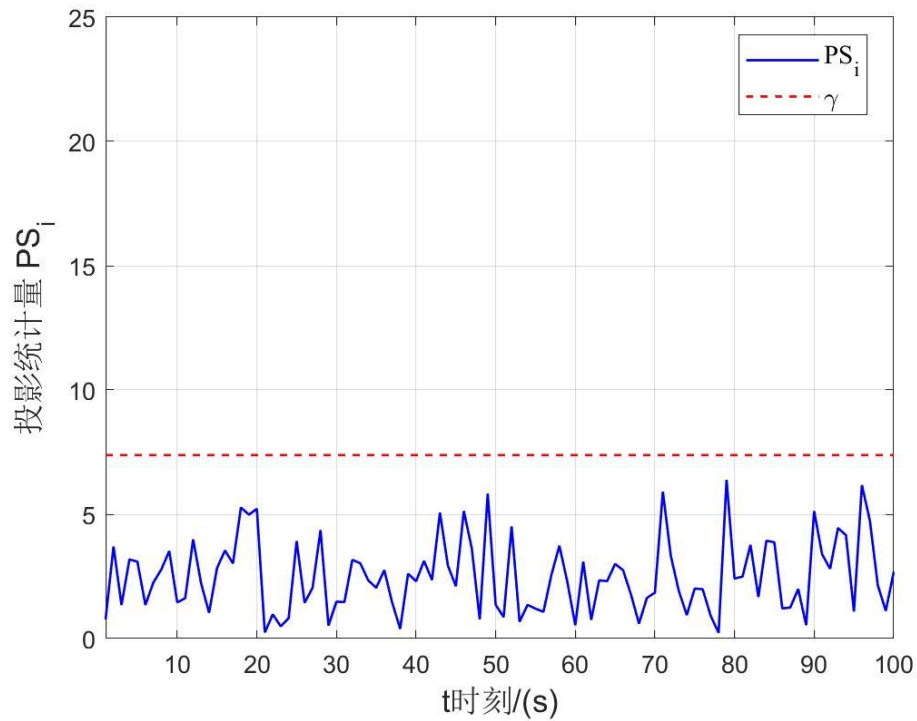


图 3-6 正常系统条件下 PS_i 与阈值的比较

如图3-6所示,通常情况下,每个样本的投影统计量都低于卡方检验的阈值,这表明系统中不存在不良数据。在这种情况下,投影统计量维持在正常水平,系

统处于稳定状态。这意味着系统中的测量数据没有明显的偏差或异常，系统运行正常，没有发生故障或其他异常情况。

然而，当系统中存在不良数据时，计算得到的投影统计量将超过卡方检测阈值，如图 3-7 所示。因此，监测投影统计量的变化能够及时发现系统中潜在的不良数据，并采取适当的措施进行处理。这样的实时监测和不良数据检测机制对于维持系统的准确性和可靠性至关重要。它们有助于提高系统的整体性能和稳定性，确保系统在运行过程中不受不良数据的影响，从而保障系统的正常运行，并且在发生异常时能够迅速做出响应，以最大程度地减少潜在的损失。

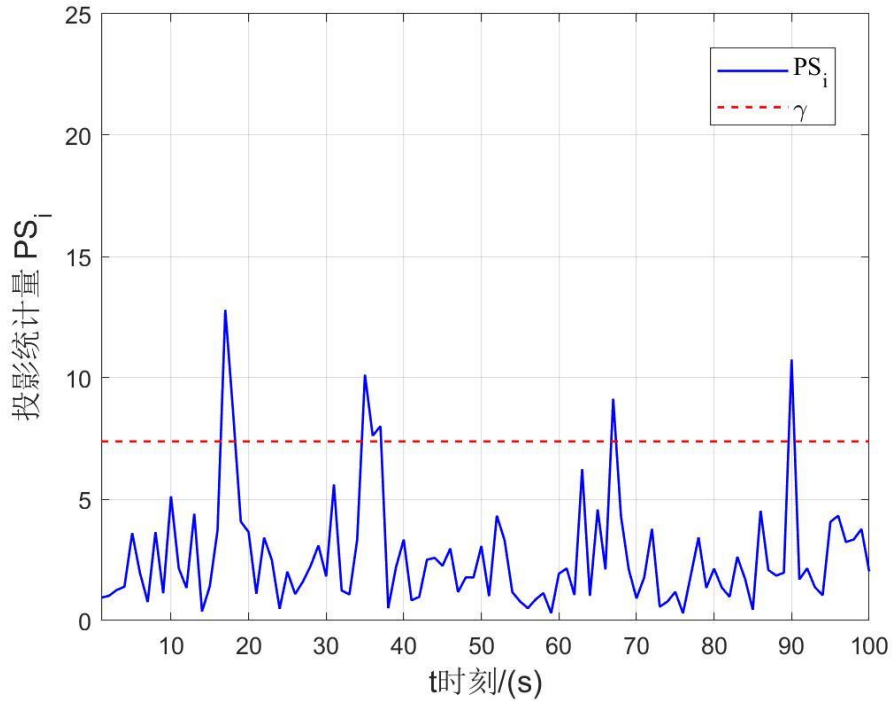


图 3-7 系统中存在不良数据时 PS_i 与阈值的比较

当电力系统中出现不良数据时，相较于传统的 EKF 算法，基于投影统计的加权 EKF 算法能够更有效地提升状态估计的准确性。该结论可通过图 3-8 和图 3-9 中的仿真结果得出。通过观察图 3-8 和图 3-9 所示的比较结果，显而易见，改进的 EKF 算法在节点电压幅度和相位角的估计方面表现出更大的稳定性。在电力系统面临不稳定环境的情况下，改进的 EKF 算法能为系统提供更为可靠的状态估计。

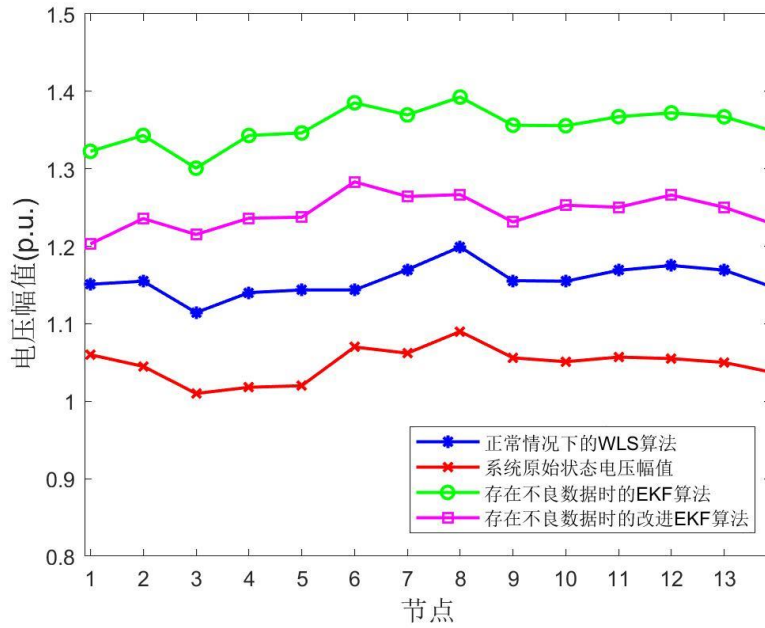


图 3-8 IEEE-14 节点系统电压幅值估计结果

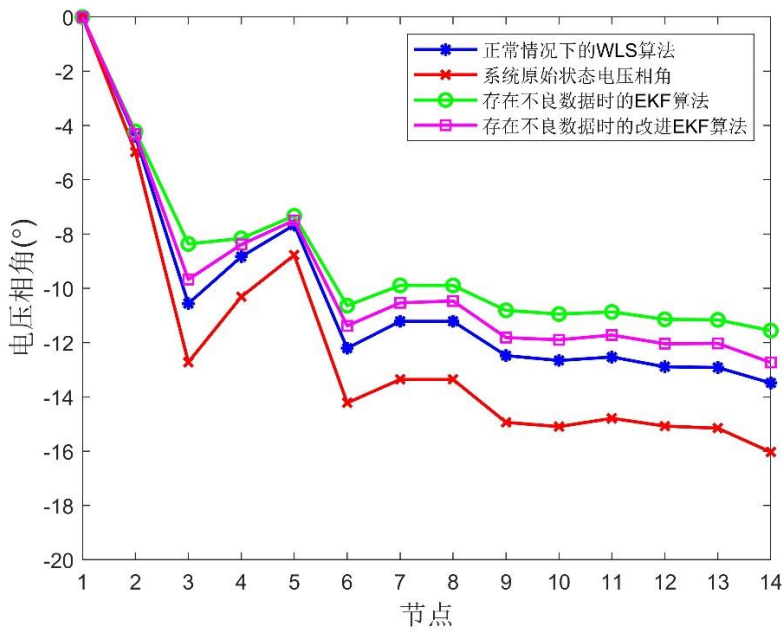


图 3-9 IEEE-14 节点系统电压相角估计结果

传统的 EKF 算法在面对不良数据时可能会产生较大的偏差。这突显了改进的 EKF 算法的重要性，因为它通过对不良数据进行加权，能够更灵活地适应各种不良数据的影响，从而提供更准确的状态估计结果。相较于传统的 EKF 算法，改进的 EKF 算法在处理不良数据方面表现出更强的鲁棒性，有效减轻了不良数据对状态估计精度的负面影响，进一步确保了系统状态估计的准确性和稳定性。

为了评估在存在不良数据时 EKF 算法和改进 EKF 算法状态估计的鲁棒性，

在本章中，使用均方根误差（Root Mean Square Error, RMSE）对这两种算法的性能进行量化评估。均方根误差是一种广泛用于度量预测模型准确性的指标，用于比较预测估计值与实际测量值之间的误差水平。均方根误差定义为

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\mathbf{x}_i - \hat{\mathbf{x}}_i)^2}{n}} \quad (3-32)$$

其中， n 表示系统的节点数， $\hat{\mathbf{x}}_i$ 为第 i 个节点的预测估计值， $\mathbf{x}_i$ 为第 i 个节点的实际值。在本章中，将传统的 EKF 算法和基于投影统计量的加权 EKF 算法的估计结果与实际测量值进行了对比。通过比较两种算法的 RMSE 值，可以评估它们在不良数据干扰下的性能。较小的 RMSE 值表明算法具有更好的鲁棒性和准确性，能够更可靠地估计电力系统的状态。从表 3-3 中呈现的结果可以明显观察到，在存在离群值的情况下，相较于传统的 EKF 算法，WEKF 算法在状态估计方面表现更为优越。

表 3-3 电压幅度的均方根误差

算法	电压幅度的均方根误差
WLS 算法	0.3112
EKF 算法	0.2693
改进 EKF 算法	0.1810

3.6 本章小结

在本章中，采用基于投影统计的加权 EKF 算法，对受到不良数据干扰的电力系统进行状态估计。在 EKF 算法的预测和更新步骤之间引入了不良数据检测算法，采用投影统计算法来识别异常数据。对于检测到的不良数据，使用指数加权函数来降低这些不良数据的权重，通过降低不良数据的权重，能够有效减小其对电力系统整体状态估计的影响，从而确保估计过程的准确性和稳定性。通过在 IEEE-14 节点系统的仿真实验，深入研究了改进 EKF 算法在电力系统中的适用性。实验结果表明，相较于传统的 EKF 算法，所提出的改进算法具备更为显著的鲁棒性，可为电力系统状态估计提供更加稳定的性能。

第4章 基于次优化-最大化算法的不良数据检测方法

4.1 引言

随着世界范围内电力工业的快速发展,电力系统规模逐渐扩大,其运行方式和网络结构也变得更加复杂。电网与现代信息的结合形成了智能电网,对电力系统运行的安全可靠性的提出了更大的挑战。为了提高电网的运行效率和可靠性,调度系统需要采集完整可靠的实时数据进行处理,便于高级应用软件进行在线分析和决策控制。电力系统状态估计是电力系统调度中心能源管理系统(EMS)的核心功能之一。随着电力系统自动化水平的不断提高,状态估计对实时测量数据的质量要求也越来越高。除了测量噪声之外,在获得的测量数据中可能存在不良数据。不良数据的存在会使状态估计结果与系统的实际运行有较大的偏差,使估计器的性能严重下降。

在第三章中,已经研究了一种基于投影统计的改进扩展卡尔曼滤波算法。该算法通过投影统计算法检测时变检测矩阵中的不良数据,并引入加权函数来减小这些不良数据的影响,从而提高了电力系统状态估计的稳定性。然而,仅仅进行不良数据的检测还不能满足所有的需求。因此,引入了第二种方法,即次优化-最大化(Minorization-Maximization, MM)算法^{[47]-[53]}。该算法不仅能够检测不良数据,还能够准确地定位这些不良数据,提高状态估计性能。将在本章中详细介绍 MM 算法的应用。

4.2 模型构建

4.2.1 不良数据测量模型

如果观测值完全符合式(2-1)中的测量模型,则可以使用加权最小二乘估计器来获得对未知参数 $\mathbf{x}$ 的估计。然而,当测量中存在不良数据时,可能会严重降低估计性能。不良数据是指由多种因素引起的数据偏差或异常,包括传感器故障、通信或人为错误、电力设备故障、损坏以及数据传输或处理错误。在不良数据存在的情况下,观测值并不完全遵循式(2-1)中的模型,而更趋向于符合式(4-1)

模型，表示为

$$\mathbf{z} = h(\mathbf{x}) + \mathbf{a} + \mathbf{e} \quad (4-1)$$

其中，向量 $\mathbf{a}$ 表示观测值中存在的不良数据。

4.2.2 高斯混合测量模型

假设测量向量 $\mathbf{z}$ 包含 M 个测量值，其中可能存在 S 个不良数据，为了进行不良数据检测，通过采集 L ($L \geq 1$) 个测量向量来识别潜在的不良数据。在测量采集过程中，不良数据的存在可能导致数据的噪声特性发生变化。因此，第 l 个测量矢量 $\mathbf{z}_l$ 的第 i 个测量值可表示为

$$z_{i,l} = \begin{cases} h_{i,l}(\mathbf{x}) + e_{i,l,1} \\ h_{i,l}(\mathbf{x}) + e_{i,l,2} \end{cases} \quad (4-2)$$

其中， $e_{i,l,1}$ 和 $e_{i,l,2}$ 分别服从高斯分布，即 $e_{i,l,1} \sim \mathcal{N}(\mu_1, \sigma_1^2)$ 和 $e_{i,l,2} \sim \mathcal{N}(\mu_2, \sigma_2^2)$ ， $e_{i,l,1}$ 表示在测量值中不存在不良数据时的测量误差向量。 $e_{i,l,2}$ 表示在测量值中存在不良数据时的测量误差向量。因此，为了描述测量误差向量 $\mathbf{e}$ 的概率密度，引入高斯混合模型。高斯混合模型 (Gaussian Mixture Model, GMM) ^{[54]-[56]} 是一种统计模型，用于描述由多个高斯分布组成的概率分布。该模型假设观测数据是由多个高斯分布生成的混合物，每个高斯分布代表数据的一个潜在子群。在高斯混合模型中，每个高斯分布由其均值 (μ)、方差 (σ^2)，以及权重 (π) 来描述，表示为

$$p(\mathbf{e}) = \sum_{k=1}^2 \pi_k \mathcal{N}(\mathbf{e} | \mu_k, \sigma_k^2) \quad (4-3)$$

其中， $\mathcal{N}(\mathbf{e} | \mu_k, \sigma_k^2)$ 表示混合模型中的第 k 个分量，其中 μ_k 和 σ_k^2 分别为该分量的均值和方差。 π_k 被称为混合系数，它反映了第 k 个分量在整个高斯混合模型中的权重。具体而言，混合系数满足

$$\sum_{k=1}^2 \pi_k = 1; 0 \leq \pi_k \leq 1 \quad (4-4)$$

考虑到 M 表示测量值的数量， S 表示不良数据的数量，因此可以通过以下方式定义混合模型的两个分量权重： $\pi_1 = (M - S) / M$ ， $\pi_2 = S / M$ 。在高斯混合模

型中，需要估计的参数为 $\theta = [\pi_1, \pi_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2]^T$ 。因此，可以构建关于 θ 的目标函数，以最大程度地拟合观测数据并估计模型参数，其数学表达式为

$$\begin{aligned}
 J(\theta, (\mathbf{z} - h(\mathbf{x}))) &= \ln p((\mathbf{z} - h(\mathbf{x})); \theta) \\
 &= \ln \left(\sum_{k=1}^2 \pi_k \mathcal{N}((\mathbf{z} - h(\mathbf{x})) | \mu_k, \sigma_k^2) \right) \\
 &= \sum_{i=1}^M \sum_{l=1}^L \ln \left(\sum_{k=1}^2 \pi_k \mathcal{N}((z_{i,l} - h_{i,l}(\mathbf{x})) | \mu_k, \sigma_k^2) \right)
 \end{aligned} \tag{4-5}$$

在进行未知参数的估计时，式(4-5)无法像处理单高斯模型（Single Gaussian Model, SGM）那样直接使用最大似然方法来推导出使似然函数最大化的参数。针对采集的观测数据，事先并不清楚其属于哪个子分布。每个子模型都包含未知参数，难以直接进行推导和计算。为了解决这一问题，可以采用次优化-最大化算法。这一方法为处理未知参数问题提供了一种有效的途径，能够更好地理解和拟合观测数据。

4.3 次优化-最大化算法

4.3.1 次优化-最大化算法基本思想

次优化-最大化算法是一种迭代优化算法，其核心思想在于通过构造一个目标函数的替代函数来逼近原始目标函数，并通过最大化这个替代函数来不断逼近原始问题的最优解。在每一次迭代中，算法交替执行两个关键步骤。首先，通过构造一个替代函数，即原始目标函数在当前参数估计值处的下界。这个替代函数通常被设计为简单且易于处理的形式，并且必须满足两个关键条件：始终处于原始目标函数下方，并在当前参数估计值处与原始目标函数相等。然后，算法通过最大化构造的替代函数来获取其最大值，从而改进当前的参数估计。通过反复执行这两个步骤，次优化-最大化算法希望逐渐逼近原始目标函数的最优解。这种迭代过程在满足一定收敛条件时会终止。

4.3.2 次优化-最大化算法基本原理

给定 $J(\theta, (\mathbf{z} - h(\mathbf{x})))$ 为测量误差的对数似然函数，参数向量

$\theta = [\pi_1, \pi_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2]^T$ ，根据 MM 算法的基本思想，替代函数需满足的条件为

$$\begin{cases} Q(\theta | \theta^{(s)}) \leq J(\theta, (\mathbf{z} - h(\mathbf{x}))) \\ Q(\theta^{(s)} | \theta^{(s)}) = J(\theta^{(s)}, (\mathbf{z} - h(\mathbf{x}))) \end{cases} \quad (4-6)$$

其中， $\theta^{(s)}$ 表示为 θ 的第 s 次迭代，函数 $Q(\theta | \theta^{(s)})$ 总是位于目标函数 $J(\theta, (\mathbf{z} - h(\mathbf{x})))$ 之下。当 $\theta = \theta^{(s)}$ 时，函数 $Q(\theta | \theta^{(s)})$ 与函数 $J(\theta, (\mathbf{z} - h(\mathbf{x})))$ 相切。在第 $s+1$ 次迭代中，将函数 $Q(\theta | \theta^{(s)})$ 作为目标函数 $J(\theta, (\mathbf{z} - h(\mathbf{x})))$ 的替代函数。随后，通过最大化替代函数 $Q(\theta | \theta^{(s)})$ 来获得参数 θ 的第 $s+1$ 次迭代，即

$$\theta^{(s+1)} = \arg \max_{\theta} Q(\theta | \theta^{(s)}) \quad (4-7)$$

且有

$$J(\theta^{(s+1)}, (\mathbf{z} - h(\mathbf{x}))) \geq Q(\theta^{(s+1)} | \theta^{(s)}) \geq Q(\theta^{(s)} | \theta^{(s)}) = J(\theta^{(s)}, (\mathbf{z} - h(\mathbf{x}))) \quad (4-8)$$

由此可见，在次优化-最大化算法的迭代过程中，目标函数的值呈逐渐增加的趋势，持续逼近目标函数的最优解。从次优化-最大化算法的基本原理来看，使用该算法估计未知参数的难点在于构造合适的函数作为目标函数的替代函数，以确保每次迭代都能够实现更好的优化结果。

4.4 基于次优化-最大化算法的不良数据检测

4.4.1 詹森不等式

詹森 (Jensen) 不等式^{[57]-[58]}是数学中的一个重要不等式，假设 χ 为随机变量， $\varphi(\cdot)$ 为凹函数，那么对于任意的随机变量 χ 和凹函数 $\varphi(\cdot)$ ，Jensen 不等式表示为

$$\varphi(E(\chi)) \geq E[\varphi(\chi)] \quad (4-9)$$

其中， $E(\chi)$ 为 $\varphi(\cdot)$ 的期望。

Jensen 不等式分为连续型和离散型两种形式。Jensen 不等式的连续形式为

$$\varphi\left(\int_{\chi} g(x) \cdot f(x) dx\right) \geq \int_{\chi} \varphi(g(x)) \cdot f(x) dx \quad (4-10)$$

其中， $g(\cdot)$ 为实值函数， $f(\cdot)$ 为随机变量 χ 的密度函数。

Jensen 不等式的离散形式为

$$\varphi\left(\sum_{i=1}^n \lambda_i x_i\right) \geq \sum_{i=1}^n \lambda_i \varphi(x_i) \quad (4-11)$$

其中, $\lambda_i \geq 0$ 且 $\sum_{i=1}^n \lambda_i = 1$, 则 $\sum_{i=1}^n \lambda_i \varphi(x_i)$ 为完全分离相加形式。

4.4.2 替代函数的构造

构建合适的替代函数是利用 MM 算法估计高斯混合模型中未知参数的关键。Tian 等人提出了一种新的组装与分解方法 (Assembly and Decomposition, AD) 来构建替代函数^[59]。AD 方法通过组装和分解过程, 有助于构建适用于 MM 算法的替代函数。其中, AD 方法中的 A 技术首先涉及将一系列简单的替代函数组装成一个整体函数, 确保构建的替代函数对原始目标函数进行有效的逼近。A 技术作为 MM 算法的基础, 对替代函数的构建起到了指导作用。而 D 技术则将目标函数分解为更简单的组成部分。通过分解, 能够更灵活地处理原始目标函数的结构, 并使得构建替代函数的过程更加可控。

根据 AD 方法, MM 算法在次优化步骤中借助离散形式的 Jensen 不等式构建替代函数。因此, 满足 (4-6) 式条件的替代函数表示为

$$\begin{aligned} Q(\theta | \theta^{(s)}) &= \sum_{i=1}^M \sum_{l=1}^L \left[\sum_{k=1}^2 w_{i,l,k} \ln \frac{\pi_k \mathcal{N}((z_{i,l} - h_{i,l}(x)) | \mu_k, \sigma_k^2)}{w_{i,l,k}} \right] \\ &= \sum_{i=1}^M \sum_{l=1}^L \sum_{k=1}^2 w_{i,l,k} \ln(\pi_k \mathcal{N}((z_{i,l} - h_{i,l}(x)) | \mu_k, \sigma_k^2)) + c^{(s)} \end{aligned} \quad (4-12)$$

式 (4-12) 中, $\theta^{(s)}$ 表示 θ 的第 s 次迭代, 权重函数 $w_{i,l,k}$ 和常数项 $c^{(s)}$ 为

$$w_{i,l,k} = \frac{\pi_k^{(s)} \mathcal{N}((z_{i,l} - h_{i,l}(\mathbf{x})) | \mu_k^{(s)}, \sigma_k^{2,(s)})}{p((z_{i,l} - h_{i,l}(\mathbf{x})); \theta^{(s)})} \quad (4-13)$$

$$c^{(s)} = - \sum_{i=1}^M \sum_{l=1}^L \sum_{k=1}^2 w_{i,l,k} \ln w_{i,l,k} \quad (4-14)$$

其中

$$p((z_{i,l} - h_{i,l}(x)); \theta^{(s)}) = \sum_{k=1}^2 \pi_k^{(s)} \mathcal{N}((z_{i,l} - h_{i,l}(\mathbf{x})) | \mu_k^{(s)}, \sigma_k^{2,(s)}) \quad (4-15)$$

定义 $\Phi_{i,l,k}^{(s)} = \pi_k^{(s)} \mathcal{N}((z_{i,l} - h_{i,l}(\mathbf{x})) | \mu_k^{(s)}, \sigma_k^{2,(s)})$ ，通过比较 $\Phi_{i,l,k}^{(s)}$ 的值，给出权重函数为

$$w_{i,l}^{(s)} = \begin{cases} 1, & \Phi_{i,l,1}^{(s)} \geq \Phi_{i,l,2}^{(s)} \\ 0, & \Phi_{i,l,1}^{(s)} < \Phi_{i,l,2}^{(s)} \end{cases} \quad (4-16)$$

即当 $w_{i,l}^{(s)}$ 为 1 时，该测量值为不良数据，且权重函数满足

$$w_{i,l,k} \geq 0; \sum_{k=1}^2 w_{i,l,k} = 1 \quad (4-17)$$

4.4.3 高斯混合模型中的未知参数估计

MM 算法的最大化步骤是通过运用最大似然估计(Maximum Likelihood Estimation, MLE)^[60]来最大化替代函数，从而对未知参数进行估计。式(4-12)中的替代函数可以分解为

$$\begin{aligned} Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(s)}) &= \sum_{i=1}^M \sum_{l=1}^L \sum_{k=1}^2 w_{i,l,k} \ln \pi_k \\ &+ \sum_{i=1}^M \sum_{l=1}^L \sum_{k=1}^2 w_{i,l,k} \ln(N((z_{i,l} - h_{i,l}(\mathbf{x})) | \mu_k, \sigma_k^2)) \\ &+ c^{(s)} \end{aligned} \quad (4-18)$$

其中，

$$J_1^{(s)}(\pi_k) = \sum_{i=1}^M \sum_{l=1}^L \sum_{k=1}^2 w_{i,l,k} \ln \pi_k \quad (4-19)$$

$$J_2^{(s)}(\mu_k, \sigma_k^2) = \sum_{i=1}^M \sum_{l=1}^L \sum_{k=1}^2 w_{i,l,k} \ln(N((z_{i,l} - h_{i,l}(\mathbf{x})) | \mu_k, \sigma_k^2)) \quad (4-20)$$

在第 $s+1$ 次迭代中，通过最大化替代函数来进行参数 $\boldsymbol{\theta}$ 的最大似然估计(MLE)，即

$$\boldsymbol{\theta}^{(s+1)} = \arg \max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(s)}) \quad (4-21)$$

公式(4-20)的展开形式为

$$\begin{aligned}
 & J_2^{(s)}(\mu_k, \sigma_k^2) \\
 &= \sum_{i=1}^M \sum_{l=1}^L [w_{i,l,1} \ln(\mathcal{N}((z_{i,l} - h_{i,l}(\mathbf{x})) | \mu_1, \sigma_1^2)) \\
 &+ w_{i,l,2} \ln(\mathcal{N}((z_{i,l} - h_{i,l}(\mathbf{x})) | \mu_2, \sigma_2^2))] \quad (4-22) \\
 &= \sum_{i=1}^M \sum_{l=1}^L [w_{i,l,1} (\ln \frac{1}{(2\pi\sigma_1^2)^{\frac{N}{2}}} - \frac{(z_{i,l} - h_{i,l}(\mathbf{x}) - \mu_1)^2}{2\sigma_1^2}) \\
 &+ w_{i,l,2} (\ln \frac{1}{(2\pi\sigma_2^2)^{\frac{N}{2}}} - \frac{(z_{i,l} - h_{i,l}(\mathbf{x}) - \mu_2)^2}{2\sigma_2^2})]
 \end{aligned}$$

替代函数关于未知参数的偏导数可以得到

$$\frac{\partial}{\partial \pi_k} [J_1^{(s)}(\pi_k)] = 0 \quad (4-23)$$

$$\frac{\partial}{\partial \mu_k} [J_2^{(s)}(\mu_k, \sigma_k^2)] = 0 \quad (4-24)$$

$$\frac{\partial}{\partial \sigma_k^2} [J_2^{(s)}(\mu_k, \sigma_k^2)] = 0 \quad (4-25)$$

通过求解方程可获得参数的迭代公式，可表示为

$$\pi_k^{(s+1)} = \frac{\sum_{i=1}^M \sum_{l=1}^L w_{i,l,k}}{ML} \quad (4-26)$$

$$\mu_k^{(s+1)} = \frac{\sum_{i=1}^M \sum_{l=1}^L (z_{i,l} - h_{i,l}(\mathbf{x}^{(s)})) w_{i,l,k}}{\sum_{i=1}^M \sum_{l=1}^L w_{i,l,k}} \quad (4-27)$$

$$\sigma_k^{2,(s+1)} = \frac{\sum_{i=1}^M \sum_{l=1}^L (z_{i,l} - h_{i,l}(\mathbf{x}^{(s)}) - \mu_k^{(s+1)})^2 w_{i,l,k}}{\sum_{i=1}^M \sum_{l=1}^L w_{i,l,k}} \quad (4-28)$$

基于 MM 算法估计 GMM 参数的详细工作流程如表 4-1 所示。

表 4-1 MM 算法工作流程

步骤	流程
1.	输入：测量数据 $\mathbf{z}$
2.	初始化：考虑 MM 算法的迭代过程，设置迭代指数为 $s = 0$ ；收敛容差为 $\Delta = 10^{-6}$ ；最大迭代次数为 $N_{itr}^{\max}$ 。给定初始值 $\theta^{(0)} = [\pi_k^{(0)}, \mu_k^{(0)}, \sigma_k^{2(0)}]^T$
3.	MM 算法循环：
	(1) 根据 (10) 式构造替代函数
	(2) 根据 (14) 式更新 $w_{i,l,k}$
	(3) 根据 (24) 式更新 $\pi_k^{(s+1)}$
	(4) 根据 (24) 式更新 $\mu_k^{(s+1)}$
	(5) 根据 (24) 式更新 $\sigma_k^{2,(s+1)}$
	(6) 如果满足收敛条件 $ J(\theta^{(s+1)}, (\mathbf{z} - h(\mathbf{x}))) - J(\theta^{(s)}, (\mathbf{z} - h(\mathbf{x}))) < \Delta$ 或 $s + 1 = N_{itr}^{\max}$ 时，MM 算法被终止。
4.	输出： $\pi_k^{(s+1)}$ ， $\mu_k^{(s+1)}$ ， $\sigma_k^{2,(s+1)}$

4.4.4 算法流程

图 4-1 展示了 MM 算法检测电力系统不良数据的全过程。首先，算法接收电力系统的测量数据作为输入。在初始化阶段，初始化模型参数和设置算法的收敛容差。接下来，算法采用高斯混合模型对测量误差进行聚类 and 建模，以更好地理解数据的分布特性。然后，MM 算法进入迭代优化过程，通过反复调整模型参数来逼近最优解。在每次迭代后，算法检查是否满足收敛准则，如果满足，则输出最终的模型参数和不良数据的检测结果。这些结果可以进一步用于系统分析或操作。在迭代结束后，根据设定的不良数据判定标准，系统可以精确识别可能存在的不良数据。此外，通过深入分析经过 MM 算法优化的高斯混合模型的参数，能够精准地定位这些不良数据在整个数据集中的具体位置。

通过这一流程，MM 算法能够有效地检测电力系统中的不良数据，为提高电力系统的安全性和可靠性提供了有力的工具。这一过程不仅能够有效识别不良数据，而且为进一步的数据分析和决策提供了强有力的支持。

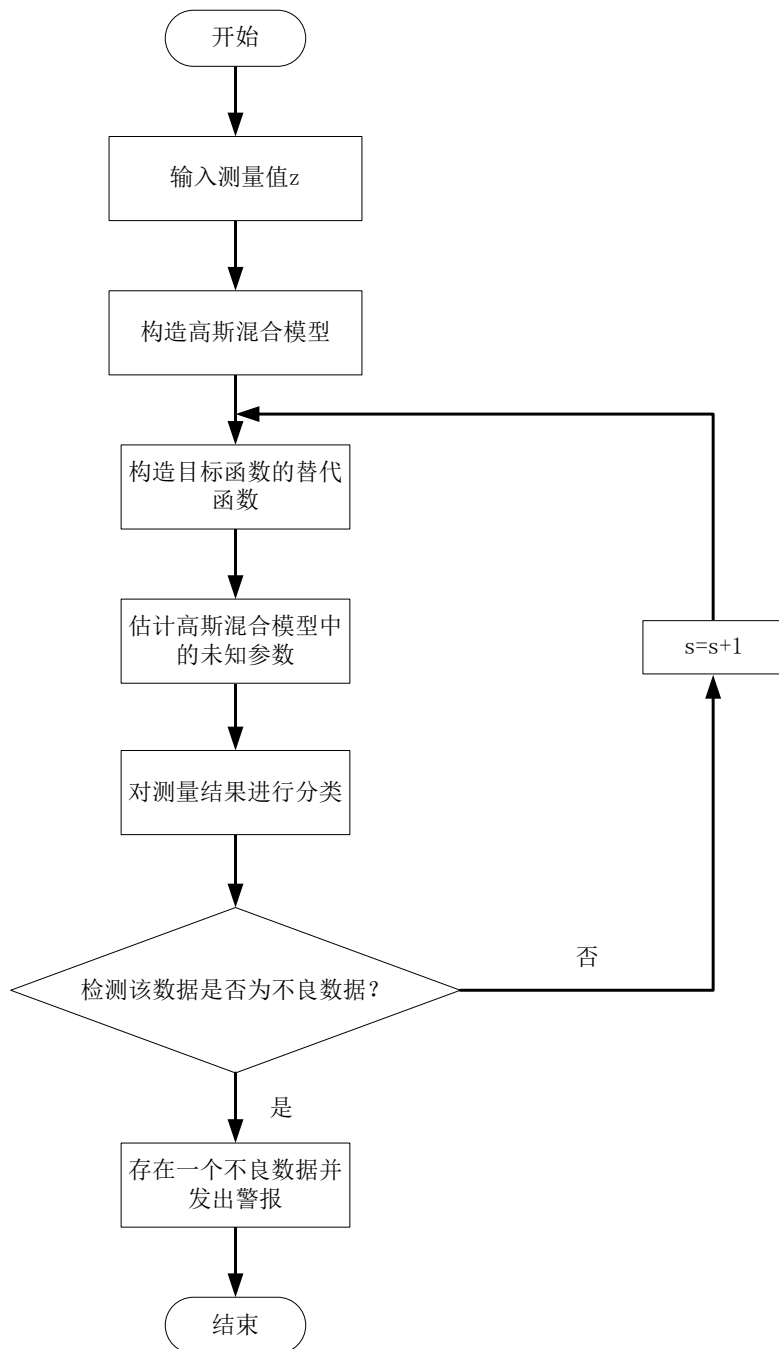


图 4-1 基于次优化-最大化算法的不良数据检测流程图

4.5 算法分析

4.5.1 算法收敛性分析

由于 MM 算法本质上是一个迭代优化算法。因此，有必要确保算法的收敛性，算法的收敛性是指 MM 算法最终单调收敛到对数似然函数 $J(\theta, (z - h(\mathbf{x})))$ 的

某个驻点 J^* ，为了验证这个问题，进行了相关性分析。首先需要证明 MM 算法的详细工作流程。

$$J(\boldsymbol{\theta}^{(s+1)}; \boldsymbol{\theta}^{(s)}) \geq J(\boldsymbol{\theta}^{(s)}; \boldsymbol{\theta}^{(s)}) \quad (4-29)$$

式 (4-29) 对其参数空间中的任意 $\boldsymbol{\theta}^{(s)}$ 都成立。

证明如下：对于给定的先验位置估计 $\mathbf{x}^{(s)}$ ，很容易证明对于相应的高斯分布，更新参数 $\pi_1^{(s+1)}$ ， $\pi_2^{(s+1)}$ ， $\mu_1^{(s+1)}$ ， $\mu_2^{(s+1)}$ ， $\sigma_1^{2,(s+1)}$ 和 $\sigma_2^{2,(s+1)}$ 是相应最大化问题的全局最优解。因此，可以得出以下结论

$$J(\boldsymbol{\theta}_m^{(s+1)}, \mathbf{x}^{(s)}; \boldsymbol{\theta}^{(s)}) \geq J(\boldsymbol{\theta}_m^{(s)}, \mathbf{x}^{(s)}; \boldsymbol{\theta}^{(s)}) \quad (4-30)$$

其中

$$J(\boldsymbol{\theta}_m^{(s)}, \mathbf{x}^{(s)}; \boldsymbol{\theta}^{(s)}) = J(\boldsymbol{\theta}^{(s)}; \boldsymbol{\theta}^{(s)}) \quad (4-31)$$

使用初始猜测集 $\mathbf{x}^{(s)}$ ，通过最小化目标函数来获取新的估计值 $\mathbf{x}^{(s+1)}$ ，其表达式为

$$f(\mathbf{x}) = \sum_{i=1}^M \sum_{l=1}^L \sum_{k=1}^2 \frac{(z_{i,l} - h_{i,l}(\mathbf{x}) - \mu_k^{(s+1)})^2}{\sigma_k^{2,(s+1)}} w_{i,l,k} \quad (4-32)$$

式(4-32)中使用了 BFGS 拟牛顿法。BFGS(Broyden-Fletcher-Goldfarb-Shanno) 拟牛顿法^{[61]-[62]}是一种用于数值优化的迭代算法，属于拟牛顿法的一种。该算法的主要目标是求解无约束的非线性优化问题，即最小化目标函数，而不考虑约束条件。BFGS 拟牛顿法的特性之一是保证在每次迭代的牛顿步骤中实现向局部最小值的下坡方向渐进。从而确保目标函数 $f(\mathbf{x})$ 在每次迭代中都趋向于减小。因此新的估计值 $\mathbf{x}^{(s+1)}$ 在第 $s+1$ 次迭代中不会导致 $J(\boldsymbol{\theta}^{(s)}; \boldsymbol{\theta}^{(s)})$ 减小，即

$$J(\boldsymbol{\theta}_m^{(s+1)}, \mathbf{x}^{(s+1)}; \boldsymbol{\theta}^{(s)}) \geq J(\boldsymbol{\theta}_m^{(s+1)}, \mathbf{x}^{(s)}; \boldsymbol{\theta}^{(s)}) \quad (4-33)$$

其中

$$J(\boldsymbol{\theta}_m^{(s+1)}, \mathbf{x}^{(s)}; \boldsymbol{\theta}^{(s)}) = J(\boldsymbol{\theta}^{(s+1)}; \boldsymbol{\theta}^{(s)}) \quad (4-34)$$

因此，式 (4-29) 得证。这意味着 $J(\boldsymbol{\theta}, (\mathbf{z} - h(\mathbf{x})))$ 的值随着迭代单调增加。由于其有上界，因此可以确保 $J(\boldsymbol{\theta}, (\mathbf{z} - h(\mathbf{x})))$ 收敛到某个驻点 J^* 。

4.5.2 算法复杂度分析

复杂度评估通常以浮点运算（FLOPs）作为一种度量标准。浮点运算是衡量执行算法所需计算量的常用指标，用于比较不同算法的计算复杂度和效率。在高斯混合模型的应用背景下，MM 算法的计算复杂度是一个值得关注的问题。为了更好地分析这一点，首先需要明确一些基本操作所需的浮点运算量。

- (1) ε_{add} ：加法所需的浮点运算（FLOPs）。
- (2) ε_{sub} ：减法所需的浮点运算（FLOPs）。
- (3) ε_{mul} ：乘法所需的浮点运算（FLOPs）。
- (4) ε_{div} ：除法所需的浮点运算（FLOPs）。
- (5) ε_{exp} ：指数运算所需的浮点运算（FLOPs）。
- (6) ε_{pow} ：对实数进行乘方运算所需的浮点运算（FLOPs）。
- (7) ε_{sqrt} ：平方根运算所需的浮点运算（FLOPs）。

由于 MM 算法是一个迭代算法，对第 $s+1$ 次迭代进行分析。在给定先验参数估计的情况下，估计的第一步是计算权重函数 $w_{i,l,k}$ 。其中 $i=1,2,\dots,M$ ； $l=1,2,\dots,L$ ； $k=1,2$ 。

计算权重函数 $w_{i,l,k}$ 所需的表达式为

$$z_{i,l} - h_{i,l}(\mathbf{x}^{(s)}) \quad (4-35)$$

其中， $i=1,2,\dots,M$ ； $l=1,2,\dots,L$ 。

$$\phi_{i,l,k} = \frac{\pi_k^{(s)}}{\sqrt{2\pi\sigma_k^{2,(s)}}} \cdot \exp\left[-\frac{(z_{i,l} - h_{i,l}(\mathbf{x}^{(s)}) - \mu_k^{(s)})^2}{2\sigma_k^{2,(s)}}\right] \quad (4-36)$$

其中， $i=1,2,\dots,M$ ； $l=1,2,\dots,L$ ； $k=1,2$ 。

$$w_{i,l,k} = \frac{\pi_k^{(s)} N((z_{i,l} - h_{i,l}(\mathbf{x})) | \mu_k^{(s)}, \sigma_k^{2,(s)})}{p((z_{i,l} - h_{i,l}(\mathbf{x})); \boldsymbol{\theta}^{(s)})} \quad (4-37)$$

其中， $i=1,2,\dots,M$ ； $l=1,2,\dots,L$ ； $k=1,2$ 。

因此，对于式（4-35）所需的浮点运算数量为 $ML\varepsilon_{sub}$ ；对于式(4-36)所需的

浮点运算数量为 $2((ML+3)\varepsilon_{mul} + (ML+1)\varepsilon_{pow} + ML\varepsilon_{sub} + ML\varepsilon_{div} + ML\varepsilon_{exp})$ ；对于式 (4-37) 所需的浮点运算数量为 $ML(1\varepsilon_{sub} + 1\varepsilon_{add} + 1\varepsilon_{div})$ 。

然后计算

$$\pi_k^{(s+1)} = \frac{\sum_{i=1}^M \sum_{l=1}^L w_{i,l,k}}{ML} \quad (4-38)$$

$$\mu_k^{(s+1)} = \frac{\sum_{i=1}^M \sum_{l=1}^L (z_{i,l} - h_{i,l}(\mathbf{x}^{(s)})) w_{i,l,k}}{\sum_{i=1}^M \sum_{l=1}^L w_{i,l,k}} \quad (4-39)$$

$$\sigma_k^{2,(s+1)} = \frac{\sum_{i=1}^M \sum_{l=1}^L (z_{i,l} - h_{i,l}(\mathbf{x}^{(s)}) - \mu_k^{(s+1)})^2 w_{i,l,k}}{\sum_{i=1}^M \sum_{l=1}^L w_{i,l,k}} \quad (4-40)$$

其中， $i=1,2,\dots,M$ ； $l=1,2,\dots,L$ ； $k=1,2$ 。容易证明式 (4-38) 所需的浮点运算数量为 $(NK-1)\varepsilon_{add} + 1\varepsilon_{div} + 1\varepsilon_{sub}$ ；式 (4-39) 所需的浮点运算数量为 $2(NK\varepsilon_{mul} + (NK-1)\varepsilon_{add} + \varepsilon_{div})$ ；式 (4-40) 所需的浮点运算数量为 $2((NK+1)\varepsilon_{pow} + NK\varepsilon_{mul} + (NK-1)\varepsilon_{add} + \varepsilon_{div} + \varepsilon_{sub})$ 。将 $FL(\theta)$ 定义为在一个 MM 算法迭代中用于估计未知参数 θ 的总浮点运算数量。 $FL(\theta)$ 等于在式 (4-35) 到式 (4-40) 中所需的总浮点运算数量，表示为

$$\begin{aligned} FL(\theta) = & (6ML-5)\varepsilon_{add} + (4ML+3)\varepsilon_{sub} \\ & + (6ML+6)\varepsilon_{mul} + (4ML+4)\varepsilon_{pow} \\ & + (3ML+5)\varepsilon_{div} + 2ML\varepsilon_{exp} \end{aligned} \quad (4-41)$$

4.6 仿真结果分析

为了验证 MM 算法在检测不良数据方面的有效性，本章与第三章相同，采用 IEEE-14 节点电力系统作为仿真分析的对象。选择在 MATLAB R2018b 软中进行仿真实验。在仿真过程中，借助 Matpower 电力仿真包提供的相关数据，进行常规潮流计算，得到了系统的运行状态和参数。并将得到的系统运行数据用作电力系统的测量数据。

首先，选择采用 IEEE-14 节点电力系统进行仿真分析。通过采用高斯混合模型对电力系统节点的测量数据进行拟合，利用 MM 算法分别估计正常数据和异常数据所属模型的参数。在这一过程中，MM 算法的收敛性和稳定性得到了充分利用。基于获得的混合模型，将测量数据输入模型，对测量数据进行分类。因此，能够准确地对正常测量值和异常数据进行分类。

表 4-2 列出了整个仿真过程中所使用的关键仿真参数。其中， M 为测量向量 $\mathbf{z}$ 中的测量值数量； π_1 ， π_2 为混合系数，在高斯混合模型中，混合系数代表了不同高斯分布成分在总体分布中所占的比例； μ_1 ， μ_2 为高斯分布的平均值； σ_1^2 ， σ_2^2 为高斯分布的方差； Δ 为收敛容差； $N_{itr}^{\max}$ 为迭代的最大次数。

表 4-2 仿真参数

参数	值
M	40
π_1	0.8
π_2	0.2
μ_1	0
μ_2	0.03
σ_1^2	0.01
σ_2^2	0.0025
Δ	10^{-6}
$N_{itr}^{\max}$	100

在进行了包含 50 组测量向量的仿真研究后，观察到图 4-2 详细展示了这些测量向量中测量误差的分布情况。然而，在电力系统存在异常数据的情况下，需要考虑到部分测量误差会发生变化的可能性。因此，假设存在一个异常值向量 $\mathbf{a}$ ，其中的每个分量 $a_{i,l}$ 表示了电力系统中可能出现的异常数据的情况，且假设 $|a_{i,l}| > 0$ 。图 4-3 展示了在存在异常数据时测量误差的分布情况。与图 4-2 相比，可以明显看出异常数据对测量误差分布的影响。

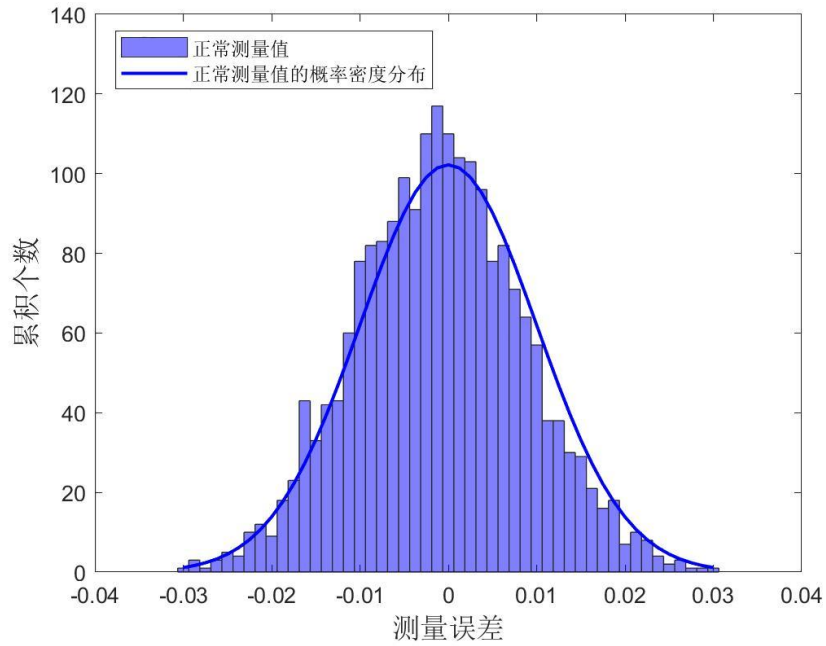


图 4-2 测量误差分布

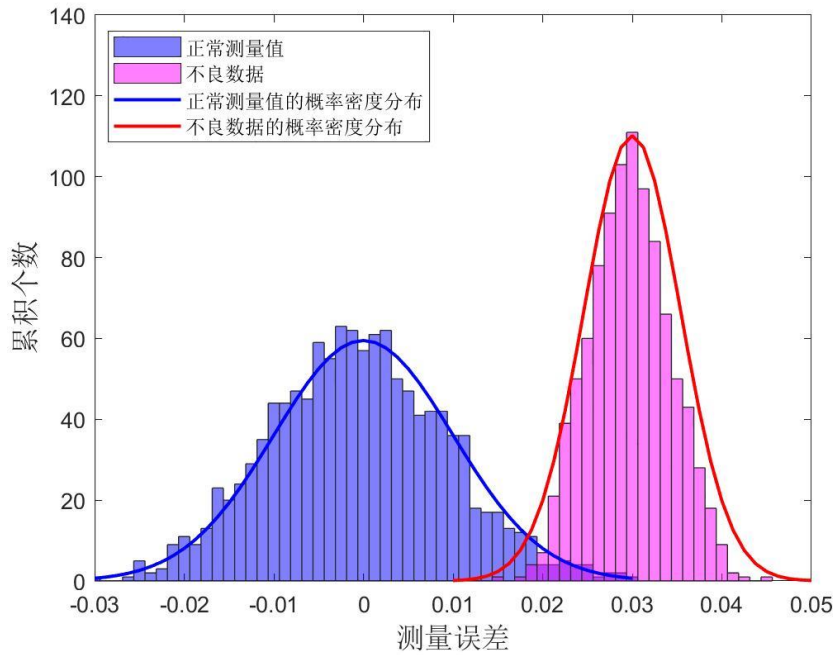


图 4-3 存在不良数据时测量误差分布

为了进一步验证所提出算法的有效性,采用MM算法对测量误差进行分类,并拟合误差分布,在图4-4中展示了分类结果。通过这一系列步骤,探讨了电力系统中存在不良数据的影响,并同时验证了算法的性能和对异常情况的有效处理能力。

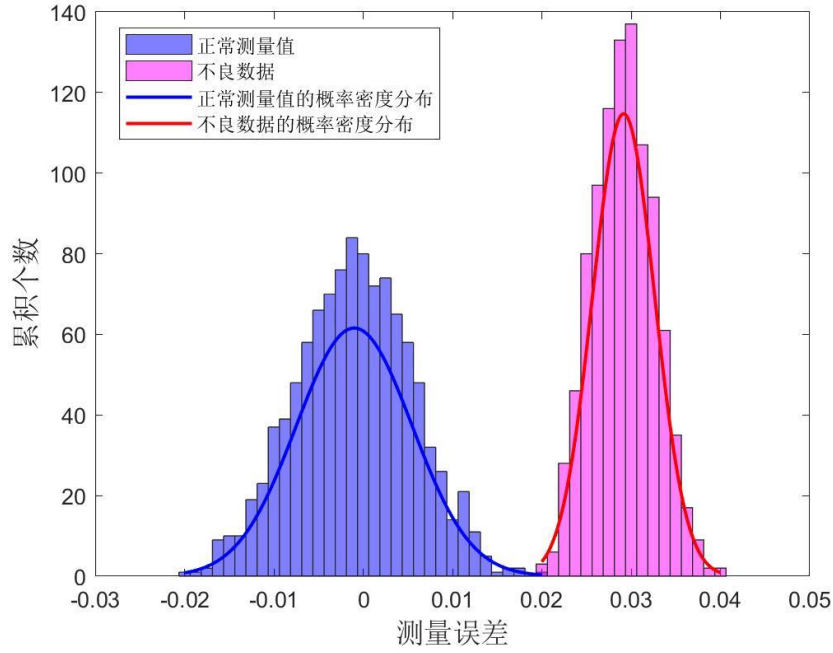


图 4-4 使用 MM 算法的测量误差分布

为验证算法的收敛性，假设 $M = 41$ ， $S = 8$ ， $L = 50$ 。使用蒙特卡罗方法进行了 1000 次独立实验，以检测参数 $\theta = [\pi_1, \pi_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2]$ 的值。针对每个参数，计算了估计值的均值与实际值之间的误差。

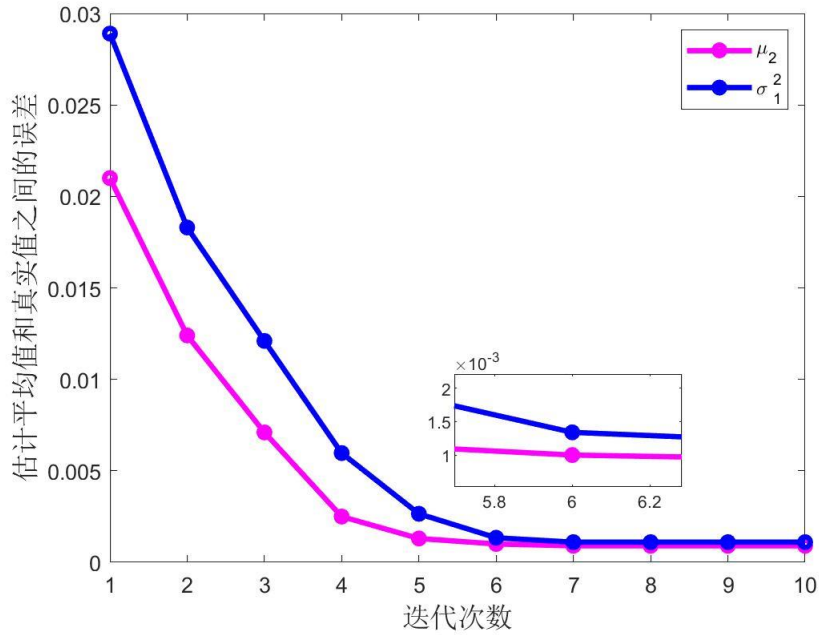


图 4-5 参数误差随迭代次数增加而变化

如图 4-5 所示，随着迭代次数的增加，参数 μ_2 和 σ_2^2 的估计均值逐渐接近真

实值，误差减小，并最终在第 6 次迭代时收敛到 0.00102 和 0.00134 显示出算法的良好收敛性。观察到随着具有不良数据的总线数量的增加，这些误差出现了变化。计算了从 1000 次独立实验中获得的参数估计均值与真实值之间的误差。方程表示为

$$E(\hat{\mu}_k) = \left| \frac{\sum_{n=1}^N \hat{\mu}_{k,n}}{N} - \mu_k \right| \quad (4-42)$$

$$E(\hat{\sigma}_k^2) = \left| \frac{\sum_{n=1}^N \hat{\sigma}_{k,n}^2}{N} - \sigma_k^2 \right| \quad (4-43)$$

$$E(\hat{\pi}_k) = \left| \frac{\sum_{n=1}^N \hat{\pi}_{k,n}}{N} - \pi_k \right| \quad (4-44)$$

$$E(\hat{\pi}_1) = \left| \frac{\sum_{n=1}^N \hat{\pi}_{1,n}}{N} - \pi_1 \right| \quad (4-45)$$

由于 $\pi_1 > 0$ ， $\pi_2 > 0$ 且 $\pi_1 + \pi_2 = 1$ ，因此 π_2 的误差可表示为

$$E(\hat{\pi}_2) = \left| \pi_1 - \frac{\sum_{n=1}^N \hat{\pi}_{1,n}}{N} \right| \quad (4-46)$$

其中， $N=1000$ ， $\hat{\mu}_{k,n}$ 表示在第 n 次实验中获得的 $\hat{\mu}_k$ 的估计值。通过对 1000 次独立实验中的 $\hat{\mu}_{k,n}$ 求和，并除以 N ，得到了估计均值。

由于 $E(\hat{\pi}_1) = E(\hat{\pi}_2)$ ，因此在图 4-6 中，红线与蓝线重合。正如图 4-6、图 4-7 和图 4-8 所示，随着具有不良数据的总线数量逐渐增加，能够清晰地观察到估计均值与参数 π_2 、 μ_2 和 σ_2^2 的实际值之间的误差逐渐减小。这是由于不良数据的数量在总体测量中所占的比例越来越大。在这种情况下，参数 π_2 、 μ_2 和 σ_2^2 的估

计均值逐渐增加,越来越接近真实值,从而导致误差逐渐减小。换句话说,随着具有不良数据的总线数量逐渐增加,观察到参数估计的准确性显著提高。

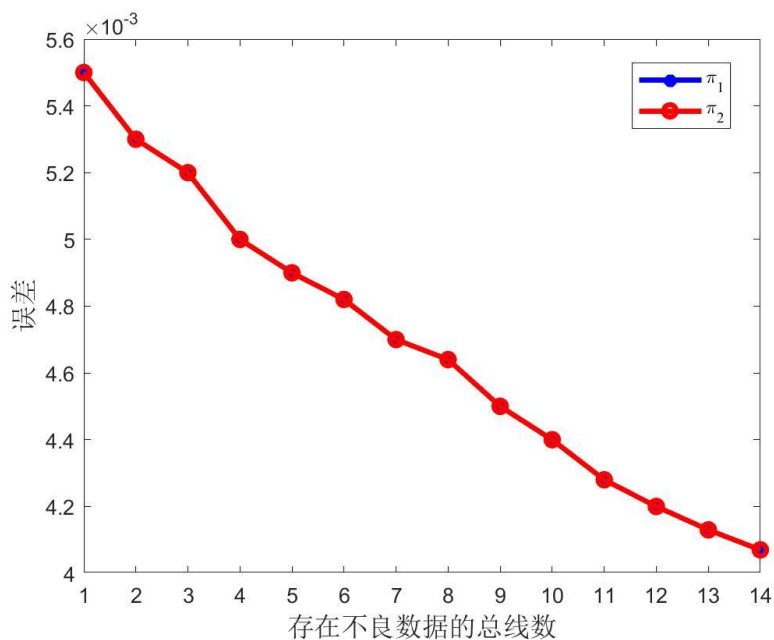


图 4-6 参数 π_k 的误差随存在不良数据的总线数量而变化

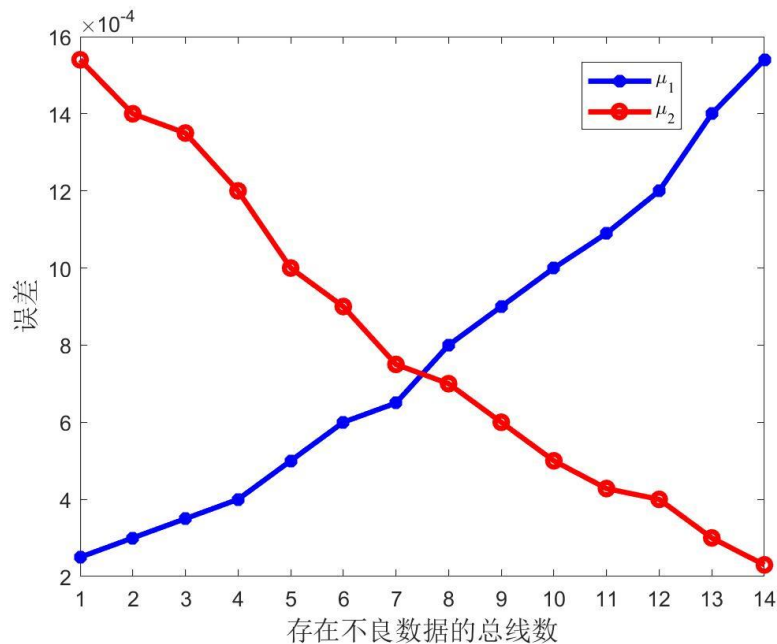


图 4-7 参数 μ_k 的误差随存在不良数据的总线数量而变化

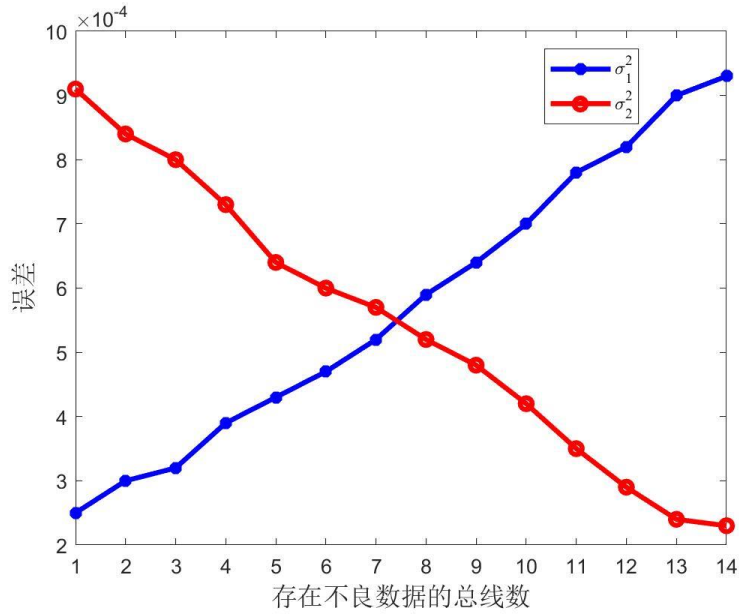


图 4-8 参数 σ_k 的误差随存在不良数据的总线数量而变化

如图 4-9 所示, 假设 40 个测量值中 P_3 、 Q_3 、 P_{1-2} 、 P_{2-3} 、 P_{4-2} 、 Q_{1-2} 、 Q_{2-3} 和 Q_{4-2} 的功率测量值存在异常值, 其余测量值的增量为 0。为了检测不良数据, 进行 1000 次独立实验, 并绘制了图 4-10, 显示了基于实验结果的不良数据检测结果。

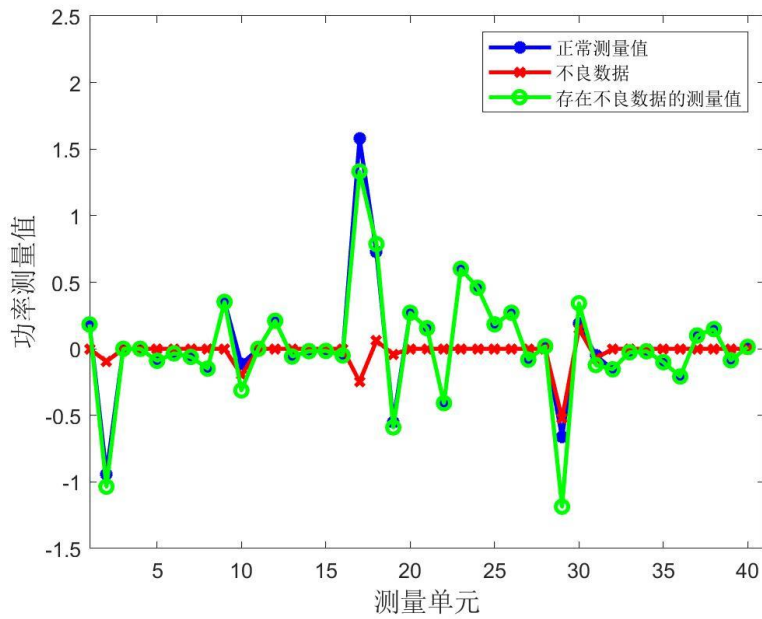


图 4-9 IEEE-14 总线系统中存在不良数据前后的数据分布

通过对图 4-9 和图 4-10 的观察, 可以注意到 Q_3 、 P_{1-2} 、 P_{4-2} 和 Q_{4-2} 中的测量

值与正常测量值更为接近。这种情况导致了少量的不良数据被错误地归类为正常数据。然而，通过应用 MM 算法进行不良数据检测，得到了超过 95% 的检测率。这意味着绝大多数不良数据可以被准确地检测到，从而有效地减少了误判的可能性。这一结果进一步证实了 MM 算法在识别不良数据方面的优越性，并为智能电网系统的安全性提供了重要保障。

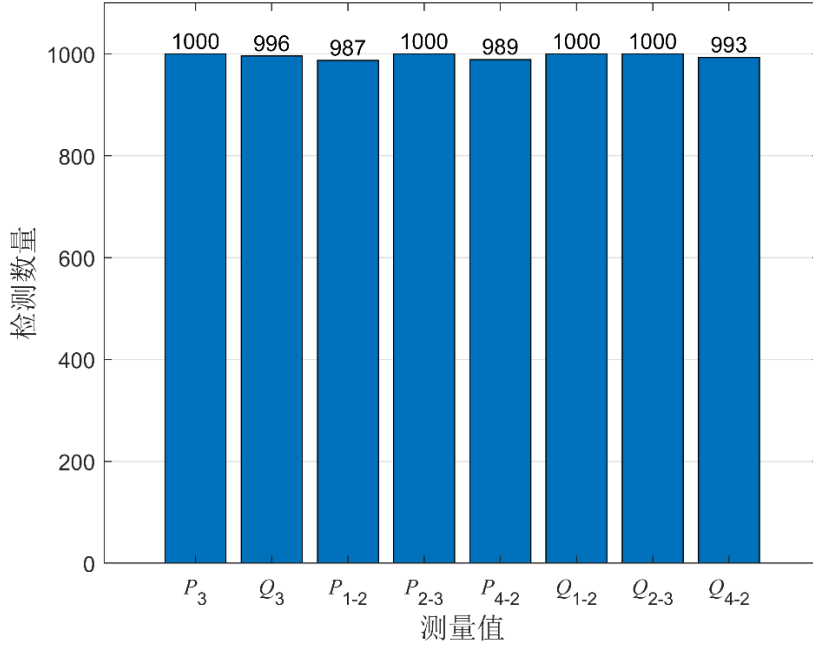


图 4-10 不良数据检测结果

为了评估 MM 算法、WLS 算法和鲁棒 Z-score 算法对不同离群值强度的检测性能，采用检测率作为性能指标。异常值强度被用来量化不良数据偏离正常测量值的程度。异常值强度越高，表示异常值对数据的影响越大，因此相应地，算法的检测性能也面临更大的挑战。通过评估这些算法在不同离群值强度下的检测率，能够更全面地了解算法在处理异常数据时的表现，并为电力系统或其他领域的数据分析提供更可靠的参考依据。异常值强度定义为

$$\eta = \frac{\|\mathbf{a}\|_2}{\|\mathbf{z}\|_2} \quad (4-47)$$

其中， $\mathbf{a} \in R^{m \times 1}$ 为异常值向量， $\mathbf{z} \in R^{m \times 1}$ 为测量向量。当异常值强度增大时，这意味着不良数据相对于正常测量值的偏离程度也随之增加，并且可能对系统产生更显著的影响。

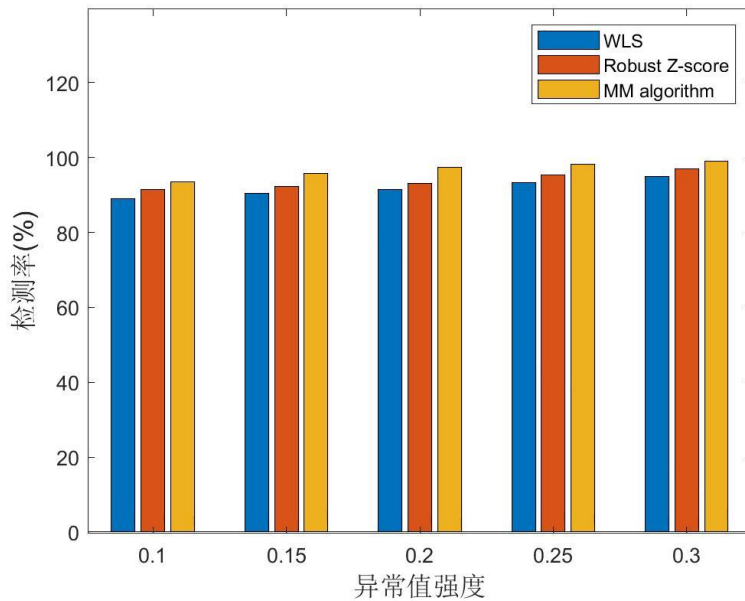


图 4-11 不同异常强度下的检测性能比较

根据图 4-11 所示的结果，可以观察到当异常值的影响相对较轻时，即异常值的强度较低时，不良数据的检测变得更加困难。尽管如此，本文提出的 MM 算法在异常值检测的性能方面表现出了显著的优越性，MM 算法的检测性能优于 WLS 算法和鲁棒 Z-score 算法。这是由于 MM 算法的迭代优化过程，使其能够逐步接近最优解，并在实际应用中表现出良好的收敛性、稳定性和鲁棒性。通过概率模型，MM 算法能够对参数进行更准确的估计。这有助于将正常数据和不良数据更好地区分开来，提高了不良数据检测的准确性。相对于某些不考虑数据分布的方法，MM 算法更具优势。相对而言，WLS 算法可能会受到不良数据的影响，导致参数估计出现偏差。与传统的 Z-score 相比，鲁棒 Z-score 在面对不良数据时表现出更高的鲁棒性。然而，MM 算法的概率模型可以帮助限制不良数据对结果的影响。通过合理的概率分布设定，MM 算法能够将不良数据的权重降低，从而减弱其对参数估计和最终结果的干扰。这使得 MM 算法在面对不良数据时更具鲁棒性。

随着异常值强度的变化，注意到误报率也会随之发生变化。如图 4-12 所示，随着异常值强度的增加，不良数据更容易被检测到，因此误报率相应地减少。这些观察结果表明，随着异常值强度的增加，异常数据的检测效果显著提高，进而导致误报率的明显下降。

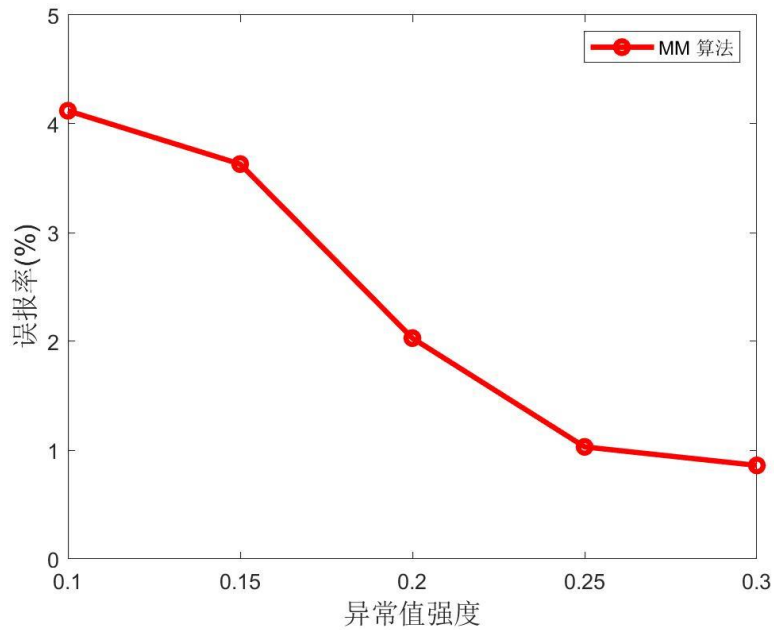


图 4-12 不同异常强度下 MM 算法的误报率

4.7 本章小结

考虑到不良数据的存在对智能电网安全的潜在影响，引入 MM 算法进行不良数据检测和定位是一种有效的方法。该算法使用高斯混合模型对测量误差进行聚类建模，并通过 MM 算法的迭代求解过程来优化模型参数。通过对 IEEE-14 总线系统的仿真分析，深入研究了 MM 算法在电力系统中的可行性和收敛性，并将其检测性能与在相同异常值强度下的 WLS 算法以及鲁棒 Z-score 进行了全面比较。实验结果表明，MM 算法能够有效地识别电力系统中的不良数据，其检测率可达 95%。研究结果强调了 MM 算法在电力系统不良数据检测中的优越性。相对于 WLS 算法和鲁棒 Z-score，MM 算法通过引入概率模型和迭代优化算法，能够更为精确地检测不良数据，进一步提升了电力系统的安全性和可靠性。

第 5 章 总结与展望

5.1 全文总结

在智能电网状态估计中,不良数据的存在对整个系统的性能和可靠性产生深远的影响。不良数据可能是由于设备故障、通信错误、测量误差或者恶意攻击等原因引起的异常数据。这些异常数据如果不及时识别和处理,可能会导致误导性的状态估计结果,进而影响电网的运行和管理。为了应对不良数据的影响,研究人员提出了各种不良数据检测和处理方法,包括基于统计学、机器学习和优化算法的方法。这些方法旨在从海量的数据中准确识别出不良数据,以确保电网状态估计的准确性和可靠性。因此,在智能电网的发展中,对不良数据的影响进行深入研究,并采取有效手段进行识别和处理,对于确保电网系统的安全、高效运行具有重要的意义。因此,本文针对智能电网状态估计中存在不良数据的问题展开研究,提出了相应的检测方法并进行仿真实验。本文主要完成以下几方面工作:

(1) 概述了在当前智能电网状态估计的发展背景下,深入研究不良数据检测方法,对提高电网运行的安全性、稳定性和可靠性的重要意义。

(2) 提出了基于改进 EKF 算法的不良数据检测方法。在 EKF 算法的预测步骤和更新步骤之间引入了不良数据检测算法,通过投影统计算法识别异常数据。对于检测到的不良数据,利用指数加权函数降低其权重,有效减缓其对电力系统整体状态估计的影响,确保估计过程的准确性和稳定性。并通过在 IEEE-14 节点系统的仿真实验中深入研究了改进 EKF 算法在电力系统中的适用性。实验结果表明,所提出的改进算法能够有效检测并降低不良数据的影响,提高了电力系统状态估计的准确性和稳定性。

(3) 提出了基于次优化-最大化算法的不良数据检测与定位方法。考虑了正常测量和异常测量之间的不同分布特征,并建立了相应的高斯混合模型。为了实现不良数据的精准检测和定位,采用次优化-最大化算法来估计高斯混合模型的未知参数。这一方法能够在不同测试场景下实现高精度的不良数据检测和定位。在 IEEE-14 节点系统中进行仿真实验,模拟了不同的测试场景,验证该算法的有效性。实验结果表明,该算法为电力系统异常数据的处理提供了一种有效的

方法,有助于提高电力系统的监测分析和控制能力,确保电力系统稳定可靠运行。

5.2 工作展望

本文针对智能电网状态估计中的不良数据问题进行了初步研究,虽然已取得一些研究成果,但仍然存在许多不足之处,需要进一步完善和深入研究:

(1) 多模型融合:探索多模型融合的方法,可以进一步完善不良数据检测方法,以应对智能电网技术发展中的更复杂的电力系统结构和更多样化的不良数据类型。

(2) 结合先进技术:结合机器学习、人工智能等先进技术,进一步提升不良数据检测的智能化水平,使系统能够更好地适应复杂多变的电网环境。

(3) 实验验证与应用:进行更多实际电网场景的验证实验,确保所提方法在真实环境中的有效性。同时,推动将这些方法应用于实际智能电网系统中,以验证其在实际运行中的可行性。

参考文献

- [1] Rathor S K, Saxena D. Energy management system for smart grid: An overview and key issues[J]. International Journal of Energy Research, 2020, 44(6): 4067-4109.
- [2] Upadhyay D, Sampalli S. SCADA (Supervisory Control and Data Acquisition) systems: Vulnerability assessment and security recommendations[J]. Computers & Security, 2020, 89: 101666-101678.
- [3] Schweppe F C, Handschin E J. Static state estimation in electric power systems[J]. Proceedings of the IEEE, 1974, 62(7): 972-982.
- [4] 于尔铿. 电力系统状态估计[M]. 水利电力出版社, 1985.
- [5] Göl M, Abur A. A fast decoupled state estimator for systems measured by PMUs[J]. IEEE Transactions on Power Systems, 2014, 30(5): 2766-2771.
- [6] Zhao J, Netto M, Huang Z, et al. Roles of dynamic state estimation in power system modeling, monitoring and operation[J]. IEEE Transactions on Power Systems, 2020, 36(3): 2462-2472
- [7] Kim Y, Bang H. Introduction to Kalman filter and its applications[J]. Introduction and Implementations of the Kalman Filter, 2018, 1: 1-16.
- [8] Nie X. Detection of grid voltage fundamental and harmonic components using Kalman filter based on dynamic tracking model[J]. IEEE Transactions on Industrial Electronics, 2019, 67(2): 1191-1200.
- [9] Karimipour H, Dinavahi V . Extended Kalman filter-based parallel dynamic state estimation[J]. IEEE transactions on smart grid, 2015, 6(3): 1539-1549.
- [10] Hartley R, Ghaffari M, Eustice R M, et al. Contact-aided invariant extended Kalman filtering for robot state estimation[J]. The International Journal of Robotics Research, 2020, 39(4): 402-430.
- [11] Watitwa J K, Awodele K O. Active distribution system state estimation: comparison between weighted least squares and extended kalman filter algorithms[C]. 2020 IEEE PES/IAS PowerAfrica. IEEE, 2020: 1-5.

- [12]Menegaz H M T, Ishihara J Y , Borges G A, et al. A systematization of the unscented Kalman filter theory[J]. IEEE Transactions on automatic control, 2015, 60(10): 2583-2598.
- [13]Daid A, Busvelle E, Aidene M. On the convergence of the unscented Kalman filter[J]. European Journal of Control, 2021, 57: 125-134.
- [14]Zhao J, Mili L, Gómez-Expósito A. Constrained robust unscented Kalman filter for generalized dynamic state estimation[J]. IEEE Transactions on Power Systems, 2019, 34(5): 3637-3646.
- [15]Wei L, Zhang Q. Detection of False Data Injection Attack in Smart Grid Based on Improved UKF[J]. Journal of System Simulation, 2023, 35(7): 1508-1516.
- [16]Yin S, Zhu X. Intelligent particle filter and its application to fault detection of nonlinear system[J]. IEEE Transactions on Industrial Electronics, 2015, 62(6): 3852-3861.
- [17]Jouin M, Gouriveau R, Hissel D, et al. Particle filter-based prognostics: Review, discussion and perspectives[J]. Mechanical Systems and Signal Processing, 2016, 72: 2-31.
- [18]Do Coutto Filho M B, de Souza J C S. Forecasting-aided state estimation—Part I: Panorama[J]. IEEE Transactions on Power Systems, 2009, 24(4): 1667-1677.
- [19]Amor N, Chebbi S. Performance comparison of particle swarm optimization and extended Kalman filter methods for tracking in non-linear dynamic systems[C]. 2017 International Conference on Control, Automation and Diagnosis (ICCAD). IEEE, 2017: 116-119.
- [20]Qi J, Sun K, Wang J, et al. Dynamic state estimation for multi-machine power system by unscented Kalman filter with enhanced numerical stability[J]. IEEE Transactions on Smart Grid, 2016, 9(2): 1184-1196.
- [21]Meng D, Zhou N, Lu S, et al. Estimate the electromechanical states using particle filtering and smoothing[C]. 2012 IEEE Power and Energy Society General Meeting. IEEE, 2012: 1-7.
- [22]Kong W, Dong Z, Xu Y, et al. An enhanced bootstrap filtering method for non-

- intrusive load monitoring[C]. 2016 IEEE Power and Energy Society General Meeting (PESGM). IEEE, 2016: 1-5.
- [23]Chopin N, Papaspiliopoulos O. An introduction to sequential Monte Carlo[M]. Berlin: Springer, 2020.
- [24]Luo X, Wan Y, He X, et al. Diversity-enhanced condensation algorithm and its application for robust and accurate endoscope three-dimensional motion tracking[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 1250-1257.
- [25]Schweppe F C, Handschin E J. Static state estimation in electric power systems[J]. Proceedings of the IEEE, 1974, 62(7): 972-982.
- [26]刘广一, 于尔铿, 夏祖治. 量测偏差的检测与辨识[J]. 中国电机工程学报, 1991, 11(3): 1-7.
- [27]卢志刚, 张宗伟. 基于量测量替换与标准化残差检测的不良数据辨识[J]. 电力系统自动化, 2007, 31(13): 52-56.
- [28]黄彦全, 肖建, 李云飞. 基于量测数据相关性的电力系统不良数据检测和辨识新方法[J]. 电网技术, 2006, 30(2): 70-74.
- [29]Garcia A, Monticelli A, Abreu P. Fast decoupled state estimation and bad data processing[J]. IEEE Transactions on Power apparatus and Systems, 1979 (5): 1645-1652.
- [30]Koglin,H.-J.,Neisius et al. Bad data detection and identification[J]. International Journal of Electrical Power & Energy Systems,1990,12(2):94-103.
- [31]魏强, 王凯, 韩学山. 不良数据识别发生误判和漏判时的处理[J]. 东北电力大学学报, 2003, 23(1): 34-38
- [32]Mili L, Van Cutsem T. Implementation of the hypothesis testing identification in power system state estimation[J]. IEEE Transactions on Power Systems, 1988, 3(3): 887-893..
- [33]马光, 朱文, 顾慧杰. 基于加权最小绝对值状态估计的有源配电网拓扑辨识方法[J]. 中国电力, 2024, 57(01): 167-174.
- [34]Liu Y, Ning P, Reiter M K. False data injection attacks against state estimation in

- electric power grids[J]. *ACM Transactions on Information and System Security (TISSEC)*, 2011, 14(1): 1-33.
- [35] Alam S M S, Natarajan B, Pahwa A. Distribution grid state estimation from compressed measurements[J]. *IEEE Transactions on Smart Grid*, 2014, 5(4): 1631-1642.
- [36] Mili L, Cheniae M G, Vichare N S, et al. Robust state estimation based on projection statistics of power systems[J]. *IEEE Transactions on Power Systems*, 1996, 11(2): 1118-1127.
- [37] Chen Y, Liu F, Mei S, et al. A robust WLAV state estimation using optimal transformations[J]. *IEEE Transactions on Power Systems*, 2014, 30(4): 2190-2191.
- [38] 吴越赢. 基于数据挖掘的电力系统不良数据检测与辨识算法研究[D]. 南京理工大学, 2017.
- [39] 刘浩, 刘玉田. 模糊数学在状态估计不良数据检测辨识中的应用[J]. *电力系统及其自动化学报*, 1996, 8(3): 50-57
- [40] 卫志农, 张云刚, 郑玉平. ISODATA 方法在配网状态估计不良数据识别中的应用[J]. *河海大学学报*, 2002, 30(2): 97-100
- [41] 刘佳芳. 配电网不良数据检测与辨识的改进新息图法研究[D]. 湖北民族学院, 2018.
- [42] 张艳军, 葛延峰, 高凯等. 基于广域测量系统的新息图辨识不良数据方法[J]. *电力系统自动化*, 2010, 34(23): 25-29.
- [43] Zhao J, Mili L. A framework for robust hybrid state estimation with unknown measurement noise statistics[J]. *IEEE Transactions on Industrial Informatics*, 2017, 14(5): 1866-1875.
- [44] Gratton S, Lawless A S, Nichols N K. Approximate Gauss–Newton methods for nonlinear least squares problems[J]. *SIAM Journal on Optimization*, 2007, 18(1): 106-132.
- [45] Ghorbani H. Mahalanobis distance and its application for detecting multivariate outliers[J]. *Facta Universitatis, Series: Mathematics and Informatics*, 2019: 583-595.

- [46]Smiti A. A critical overview of outlier detection methods[J]. Computer Science Review, 2020, 38: 100306-100317.
- [47]Sun Y, Babu P, Palomar D P. Majorization-minimization algorithms in signal processing, communications, and machine learning[J]. IEEE Transactions on Signal Processing, 2016, 65(3): 794-816.
- [48]Arora A, Tsinos C G, Rao B S M R, et al. Hybrid transceivers design for large-scale antenna arrays using majorization-minimization algorithms[J]. IEEE Transactions on Signal Processing, 2019, 68: 701-714.
- [49]Panwar K, Katwe M, Babu P, et al. A majorization-minimization algorithm for hybrid TOA-RSS based localization in NLOS environment[J]. IEEE Communications Letters, 2022, 26(5): 1017-1021.
- [50]Zhou G, Tward D, Lange K. A Majorization-Minimization Algorithm for Neuroimage Registration[J]. SIAM Journal on Imaging Sciences, 2024, 17(1): 273-300.
- [51]Tian G L, Ju D, Chuen Yuen K, et al. New expectation–maximization-type algorithms via stochastic representation for the analysis of truncated normal data with applications in biomedicine[J]. Statistical Methods in Medical Research, 2018, 27(8): 2459-2477.
- [52]Tavakoli S, Yooseph S. Learning a mixture of microbial networks using minorization–maximization[J]. Bioinformatics, 2019, 35(14): 23-30.
- [53]Shen K, Yu W, Zhao L, et al. Optimization of MIMO device-to-device networks via matrix fractional programming: A minorization–maximization approach[J]. IEEE/ACM Transactions on Networking, 2019, 27(5): 2164-2177.
- [54]Liu W, Cui D, Peng Z, et al. Outlier detection algorithm based on gaussian mixture model[C]. 2019 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS). IEEE, 2019: 488-492.
- [55]Yu B, Zhang Y, Xie W, et al. A network traffic anomaly detection method based on gaussian mixture model[J]. Electronics, 2023, 12(6): 1397-1411.
- [56]Choi H C, Deng C, Park H, et al. Gaussian Mixture Model-Based online anomaly

- detection for vectored area navigation arrivals[J]. *Journal of Aerospace Information Systems*, 2023, 20(1): 37-52.
- [57]Adil Khan M, Ullah H, Saeed T. Some estimations of the Jensen difference and applications[J]. *Mathematical Methods in the Applied Sciences*, 2023, 46(5): 5863-5892.
- [58]Khan M A, Pečarić Đ, Pečarić J. A new refinement of the Jensen inequality with applications in information theory[J]. *Bulletin of the Malaysian Mathematical Sciences Society*, 2021, 44(1): 267-278.
- [59]Tian G L, Huang X F, Xu J. An assembly and decomposition approach for constructing separable minorizing functions in a class of MM algorithms[J]. *Statistica Sinica*, 2019, 29(2): 961-982.
- [60]Ding D, Han Q L, Ge X, et al. Secure state estimation and control of cyber-physical systems: A survey[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2020, 51(1): 176-190.
- [61]Zhao W. A Broyden–Fletcher–Goldfarb–Shanno algorithm for reliability-based design optimization[J]. *Applied Mathematical Modelling*, 2021, 92: 447-465.
- [62]Li M. A three term Polak-Ribière-Polyak conjugate gradient method close to the memoryless BFGS quasi-Newton method[J]. *Journal of Industrial and Management Optimization*, 2018, 16(1): 245-260.

攻读硕士期间取得的学术成果

- [1] Qiao L, Gao W, Li Y, et al. Smart Grid Outlier Detection Based on the Minorization–Maximization Algorithm[J]. Sensors, 2023, 23(19): 8053.
- [2] Qiao L, Gao W, Guo X, et al. Power System Outlier Detection Based on Improved EKF Algorithm[C]. 2023 China Automation Congress (CAC). IEEE, 2023: 6512-6516.

致谢

三年一瞬，聚散有时。完成这篇毕业论文，也意味着我的学生生涯即将画上句号。这一刻，我百感交集。在此，我想向所有在我求学路上给予我帮助、支持和鼓励的人表达最真挚的感谢。

首先，我要感谢高老师和李老师，在两位老师的指导下，我不仅学到了专业知识，更学会了如何独立思考、如何面对挑战。老师的严谨治学、勤奋工作的精神深深地影响了我，将成为我今后工作和生活的宝贵财富。

其次，我要感谢同门师兄和师弟师妹们，我们一起度过了这段充实而难忘的研究生时光，你们的陪伴和帮助让我倍感温暖。我们共同学习、共同进步，也共同度过了许多难忘的时刻。

同时，我要感谢我的父母，父母之爱子，则为之计深远，我深知父母养育不易，感谢你们的一路支持。

此外，我要感谢我的三位室友和一直支持我的朋友。在我感到迷茫和困惑时，你们总是耐心开导和鼓励我，让我能够坚定信心、勇往直前。

最后，关关难过关关过，前路漫漫亦灿灿。