BRNO UNIVERSITY OF TECHNOLOGY

Faculty of Electrical Engineering
and Communication

DOCTORAL THESIS

Brno, 2024

Ing. DOMINIK FRIML

BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ

DEPARTMENT OF CONTROL AND INSTRUMENTATION

ÚSTAV AUTOMATIZACE A MĚŘICÍ TECHNIKY

BAYESIAN IDENTIFICATION VIA GENERALIZED TOTAL LEAST SQUARES: APPLIED TO LINEAR MOTOR DRIVES

BAYESOVSKÁ IDENTIFIKACE PROSTŘEDNICTVÍM ZOBECNĚNÝCH METOD NEJMENŠÍCH ČTVERCŮ: APLIKACE NA LINEÁRNÍ POHONY

DOCTORAL THESIS

DIZERTAČNÍ PRÁCE

AUTHOR Ing. Dominik Friml

AUTOR PRÁCE

SUPERVISOR prof. Ing. Pavel Václavek, Ph.D.

ŠKOLITEL

BRNO 2024

ABSTRACT

This doctoral thesis reformulates the model of linear motor with permanent magnets as an errors-in-variables problem and derives identification methods that quantify parameter uncertainty using Bayesian approaches. The study offers an alternative to classical least squares methods, with the primary contribution being the Bayesian Generalized Total Least Squares framework. This framework identifies errors-in-variables problems with arbitrary noise covariance matrices. Appropriate approximations allow the maximum a posteriori solution to be obtained using existing algorithms, enabling immediate application in various practical problems. An analytical equation provides a normal approximation of the posterior distribution, encompassing parameter uncertainty for advanced diagnostic and control methods. For the Bayesian Total Least Squares framework, special case of Bayesian Generalized Total Least Squares, previously unrecognized connection with directional statistics is revealed. The proposed framework is applied for recursive identification of linear motors with permanent magnets, supporting continuous parameter updates. Five algorithms are presented, demonstrating fast convergence and practical viability through numerical analysis. The practicality and diverse applications of the results are illustrated through both numerical and practical experiments.

KEYWORDS

PMLSM, PMSM, Statistical Identification, Errors-in-variables, EIV, Total Least Squares, TLS, Generalized Total Least Squares, GTLS, Bayesian Inference, Bayesian Identification, Bingham Distribution, Directional Statistics, Laplace Approximation, Variational Methods

ABSTRAKT

Tato dizertační práce nachází novou formulaci modelu lineárního motoru s permanentními magnety ve tvaru problému s chybami v proměnných. Odvozuje identifikační metody, které kvantifikují nejistotu identifikovaných parametrů pomocí bayesovských přístupů. Studie nabízí alternativu ke klasickým metodám nejmenších čtverců, přičemž hlavním přínosem je bayesovský algoritmus obecných nejmenších totálních čtverců. Jedná se o algoritmus pro identifikaci problémů s chybou v proměnných s libovolnými kovariančními maticemi šumu. Vhodné aproximace umožňují získat odhad maxima posteriorní distribuce pomocí existujících algoritmů, což umožňuje okamžité použití v mnoha praktických problémech. Analytická rovnice poskytuje normální aproximaci posteriorní distribuce, která obsahuje informaci o nejistotě parametrů, čehož může být využito v pokročilých diagnostických a řídicích metodách. V rámci analýzy vláštního případu bayesovských obechných nejmenších totálních čtverců je odhalena dříve nepoznaná souvislost se směrovou statistikou. Navrhované algoritmy jsou aplikovány pro rekurzivní identifikaci parametrů lineárního motoru s permanentními magnety, podporující průběžnou aktualizaci parametrů. V rámci práce je odvozeno pět algoritmů, jejich rychlá konvergence a využitelnost je demonstrována prostřednictvím numerické analýzy. Praktická použitelnost a různé využití výsledků jsou ilustrovány prostřednictvím numerických i experimentálních testů.

KLÍČOVÁ SLOVA

PMLSM, PMSM, Statistická identifikace, Problém s chybou v proměnných, EIV, Nejmenší totální čtverce, TLS, Obecné nejmenší totální čtverce, GTLS, Bayesovská inference, Bayesovská identifikace, Binghamova distribuce, Směrová statistika, Laplaceova aproximace, Variační metody

FRIML, Dominik. *Bayesian Identification via Generalized Total Least Squares: Applied to Linear Motor Drives*. Brno: Brno University of Technology, Faculty of Electrical Engineering and Communication, Department of Control and Instrumentation, 2024, 130 p. Doctoral thesis. Advised by prof. Ing. Pavel Václavek, Ph.D.

Author's Declaration

Author:	Ing. Dominik Friml
Author's ID:	186060
Paper type:	Doctoral thesis
Academic year:	2023/24
Topic:	Bayesian Identification via Generalized Total Least Squares: Applied to Linear Motor Drives

I declare that I have written this paper independently, under the guidance of the advisor and using exclusively the technical references and other sources of information cited in the paper and listed in the comprehensive bibliography at the end of the paper.

As the author, I furthermore declare that, with respect to the creation of this paper, I have not infringed any copyright or violated anyone's personal and/or ownership rights. In this context, I am fully aware of the consequences of breaking Regulation § 11 of the Copyright Act No. 121/2000 Coll. of the Czech Republic, as amended, and of any breach of rights related to intellectual property or introduced within amendments to relevant Acts such as the Intellectual Property Act or the Criminal Code, Act No. 40/2009 Coll. of the Czech Republic, Section 2, Head VI, Part 4.

Brno ....

author's signature*

*The author signs only in the printed version.

ACKNOWLEDGEMENT

First and foremost, I would like to express my sincere gratitude to my supervisor, prof. Ing. Pavel Václavek, Ph.D. His guidance and the freedom he granted me in choosing my research focus were invaluable. His support has been instrumental in the completion of this thesis. He provided essential consultations that enriched my view on problems, secured funding that made my research possible, and handled the majority of administrative tasks, allowing me to dedicate myself entirely to my research.

I am deeply grateful to Ing. Jakub Dokoupil, Ph.D., for introducing me to Bayesian analysis and the Total Least Squares. His patience and willingness to share his extensive knowledge through numerous productive consultations often led to significant improvements in my publication outcomes. His insights and critiques were crucial in refining my methodologies and ensuring the quality of my results.

My thanks extend to Ing. Lukáš Otava, Ph.D., for his invaluable help with the hardware aspects of my research. His technical expertise ensured that all equipment functioned optimally, which was critical for the successful implementation of my experiments. His readiness to assist and troubleshoot at any hour was truly commendable.

I also wish to thank Ing. Lukáš Zezula for the many fascinating conversations that frequently helped me overcome the research obstacles I faced. His intellectual curiosity and problem-solving skills were a great asset, particularly his assistance with variational approximation, which is presented in this thesis. His ability to readily help with complex mathematical problems made a significant difference in my work process.

Additionally, I would like to acknowledge Ing. Ondřej Bartík, Ing. Michal Kozubík, and Ing. Lukáš Zezula for creating a pleasant working environment in the doctoral office and for our many fruitful discussions. Their camaraderie and joyful spirit made the long hours of research more enjoyable and less solitary.

Ing. Libor Veselý, Ph.D., deserves special mention for his moral support during difficult times. His friendship and encouragement were a constant source of strength, helping me persevere through challenging periods. His positivity and unwavering belief in my capabilities often lifted my spirits when I needed it most.

I would also like to extend my heartfelt thanks to all members of the CEITEC and FEEC control teams. Your friendship has been invaluable throughout my research journey. Your support and camaraderie made the long hours of work more enjoyable and provided a sense of community that I sincerely appreciate. Thank you for being a vital part of this experience.

I also want to thank my parents for instilling in me the ability to self-learn. They taught me to strive for excellence and to endure through failures. Their lessons in self-reliance have been invaluable throughout my academic journey and in life. Their sacrifices and encouragement laid the groundwork for my academic pursuits. Their support throughout my life has been fundamental to my achievements.

Despite her indirect influence on my research results, my deepest gratitude goes to my wife, Eva Frimlová. She was a great source of moral support in both good times and bad and assisted me with crucial decisions. Her patience, understanding, and unwavering support were essential in balancing the demands of research and personal life. She provided a stable foundation that allowed me to focus on my work with peace of mind.

My daughter, Alžběta Frimlová, provided emotional support and motivation throughout most of this journey. Her presence reminded me of the importance of perseverance and dedication. Lastly, I am thankful to my unborn child for giving me the strong motivation to submit my dissertation on time. The anticipation of his arrival provided a powerful incentive to complete my work and move forward to the next chapter of my life.

In conclusion, this thesis would not have been possible without the support, guidance, and encouragement of many individuals. To each of you, I extend my deepest gratitude. Thank you for being part of this journey.

Contents

Introduction **13**

1 Linear Motor Drives Today **15**
- 1.1 Linear Motor Drive 15
- 1.2 Control of Permanent Magnet Linear Synchronous Drives 16
- 1.3 Model of Permanent Magnet Synchronous Motor 17
- 1.4 Model of Permanent Magnet Linear Synchronous Motor 19
- 1.5 Identification of Permanent Magnet Linear Synchronous Drive 20

2 Preliminaries **23**
- 2.1 Model definition 23
- 2.2 Dataset notation 23
- 2.3 Generalized singular value decomposition 24
- 2.4 Laplace Approximation 24
- 2.5 Bingham Distribution 25
- 2.6 Least Squares 28
- 2.7 Total Least Squares 30
- 2.8 Generalized Total Least Squares 34

3 Problem definition **39**
- 3.1 Bayesian Errors-in-Variables 40
- 3.2 Errors-in-Variables Approach to PMLSM Identification 42

4 Bayesian Generalized Total Least Squares **45**
- 4.1 The Expectation of the Hidden Variables 46
- 4.2 The Likelihood Function of the Dataset 47
- 4.3 Establishing Prior Distribution 49
- 4.4 Functional Form of the Posterior Distribution 50
- 4.5 Gaussian Approximation of the Posterior 52
- 4.6 Recursive Approximate B-GTLS Algorithm 53
- 4.7 Recursive PMLSM Identification Algorithm 54

5 Bayesian Total Least Squares **64**
- 5.1 Derivation of the Bayesian Total Least Squares 64
- 5.2 Bayesian Total Least Squares properties 66
- 5.3 Normal Prior 68
- 5.4 Laplace Approximation of the Posterior 69
- 5.5 Variational Approximation of the Posterior Distribution 70

5.6 Design of Recursive Identification Approximation Algorithms 72

6 Numerical Analysis **76**

6.1 Comparison of Approximation Methods 76

6.2 Numerical analysis of B-TLS-based algorithms 79

6.3 Numerical analysis of B-GTLS-based algorithms 82

6.4 Forgetting Factor 85

7 Examples **87**

7.1 Quality of Identification using B-TLS 87

7.2 Quality of Identification using B-TLS in practice 90

7.3 Approximate Recursive Identification with Confidence Intervals . . . 92

7.4 Recursive PMLSM Identification Simulation 94

7.5 Recursive PMLSM Identification 101

Conclusion **109**

List of Figures

1.1 Differences between ideal and non-ideal linear motor drives. 19
2.1 Orthographic projection of Bingham distribution contour plot on sphere. 26
2.2 Graphical illustration of orthogonal regression. 31
3.1 Classical EIV (left), Berkson Model (right) 41
4.1 Structure of the PMLSM-RGTLS algorithm. 63
5.1 Geometrical interpretation of the Bingham distribution 67
6.1 The visual comparison of the posterior $p_{\mathcal{N}}(\theta|D)$. 77
6.2 The visual comparison of the posteriors and surrogate posteriors. . . 78
6.3 Numerical analysis of Algorithm 7. 80
6.4 Numerical analysis of Algorithms 8 and 9. 81
6.5 Numerical analysis of Algorithm 4.6. 83
6.6 Numerical analysis of Algorithm 6. 84
6.7 Relative error analysis of Algorithm 6. 85
6.8 Identification of parameters with and without forgetting factor. . . . 86
7.1 Comparison of simulated data for PRGS (left) and PRBS (right). . . 88
7.2 Posterior Probability Distribution of PRGS and PRBS Simulations. . 89
7.3 Step transition Characteristics of True and Sampled Parameters. . . . 90
7.4 Comparison of measured data for PRGS (left) and PRBS (right). . . 91
7.5 Posterior Probability Distribution of PRGS and PRBS Measurements. 92
7.6 The battery cell total capacity estimation. 94
7.7 PMLSM identification and validation data. 98
7.8 PMLSM identification posterior distributions. 99
7.9 Evolution of PMLSM identified parameters. 100
7.10 Photo of used linear dynamometer. 101
7.11 Schematic view on the used linear dynamometer. 102
7.12 Detail of experiment data. 104
7.13 Evolution of PMLSM identified parameters. 105
7.14 Comparison of mover speed approximations. 107

List of Tables

4.1 Available algorithms for MAP of special cases. 53
6.1 B-TLS numerical analysis simulation parameters. 79
7.1 PMLSM simulation parameters. 97
7.2 Tested PMLSM drive parameters. 103
7.3 Comparison of the original and PMLSM-RGTLS estimates. 106

List of Algorithms

1 Recursive Least Squares (RLS) 28
2 Inverse Iteration (II) 33
3 Recursive Total Least Squares (RTLS) 34
4 Recursive Generalized Total Least Squares (RGTLS) 37
5 Bayesian Recursive Generalized Total Least Squares (B-RGTLS) . . . 54
6 Bayesian Recursive PMLSM GTLS (PMLSM-RGTLS) 62
7 Inverse Iteration Laplace RTLS (IIL-RTLS) 72
8 Recursive Laplace TLS (L-RTLS) 73
9 Recursive VB TLS (VB-RTLS) 74

Introduction

Electric drives are widely used in industry. While rotary drives have been utilized for a relatively long time, linear electric drives have only recently become prevalent due to several advantages they offer over rotary drives in many applications. Linear electric drives provide benefits such as higher precision, direct linear motion, and improved efficiency, making them suitable for a wide range of industrial applications.

Given the variety of different types of linear motors, it is not feasible to cover all of them in this work. Therefore, the focus will be on the linear motor with permanent magnets, which has recently gained significant popularity in the industry. This type of motor is favored for its high force density, excellent dynamic performance, and lower maintenance requirements compared to other types.

The aim of this work is to analyze the issues of linear drives identification. This is beneficial for field of identification, control and diagnostics, as most control algorithms rely directly on the knowledge of the model and its parameters and diagnostics often rely on comparing model predictions with measured data. Improving the quality of both control and diagnostics can be achieved by obtaining the model and its parameters.

Modern identification methods provide not only the identified parameters but also information about the uncertainty of each parameter, often using Bayesian statistics to express the probability distributions of the parameters. Bayesian methods enhance the robustness and reliability of the identification process by quantifying uncertainty, which is critical for high-precision and safety-critical applications. The vast majority of statistical identification and estimation methods are based on the Bayesian formulation of the least squares inference. This thesis focuses on analyzing the possibility of reformulating the model of a linear motor with permanent magnets as an errors-in-variables problem and deriving identification methods for such a model while providing uncertainty information about the identified parameters.

In order to achieve this, the errors-in-variables model needs to be introduced, and connections with classical least squares methods must be revealed. This involves a detailed comparison of the two approaches, highlighting their similarities and differences in handling measurement errors.

Then, the model of the linear drive needs to be formulated, explored, and consequently reformulated as an errors-in-variables problem. This step is crucial for developing a more accurate representation of the system, considering both input and output measurement errors.

As existing methods for errors-in-variables identification do not provide parameter uncertainty information, a thorough Bayesian analysis of the problem needs to be conducted, and novel identification algorithms need to be proposed and verified. These new algorithms will be designed to integrate uncertainty quantification, enhancing their applicability in real-world scenarios.

Finally, the proposed methods can be used for the identification of the reformulated linear drive model. Numerous simulations and experimental validations will be performed to demonstrate the effectiveness and accuracy of the proposed identification techniques, ensuring their practical viability in industrial settings.

1 Linear Motor Drives Today

Linear motor drives offer many advantages for industry. A detailed list of the ways linear motor drives are used, their control options, modeling, and associated complications will be covered in the following chapters.

1.1 Linear Motor Drive

Linear electric drives find their applications, particularly in scenarios requiring linear motion. Such applications are extensively represented in the industrial sector, for instance, in processes involving machining, the movement of products along production lines, specific robotic movements, or in the wafer bonding process to chip leads [6, 7, 8, 9, 10].

In such applications, using linear motor drives is advantageous, particularly for eliminating the need to convert rotary motion to linear motion. The conversions in existing solutions pose challenges due to their cost, susceptibility to failure, maintenance demands, energy losses, and especially the nonlinearities they introduce, such as backlash in gears [11, 12, 13, 6], which complicate their effective control. Since many applications require high positional accuracy with large accelerations and decelerations under varying loads, the demands on control are significant. This can be addressed by enhancing knowledge of the model and its parameters, which allows for improved control. As an extension, knowledge of parameter uncertainty enables advanced control methods and diagnostics of the drives.

Additional properties of linear motors, such as the ability to mechanically isolate the primary and secondary parts to create corrosion, water, and dust-resistant linear motor drives, their capacity to exert significant force, or the possibility of having multiple independent sliders [14, 15, 16] on a linear track [17], enable the development of new applications that were not achievable with rotary motors.

Linear motors are manufactured with various structural designs. The different types of constructions generally derive from the designs of rotary motors. Consequently, it is possible to find linear motors that are synchronous, asynchronous, excited by permanent magnets, or externally excited [18]. This thesis will focus solely on synchronous linear motors with permanent magnets, often abbreviated in the literature as permanent magnet linear synchronous motor (PMLSM) or linear permanent magnet synchronous motor (LPMSM).

Due to the similarity with rotary drives, linear motor drives often adopt terminology such as rotor and stator, which can lead to confusion. Therefore, the following terminology will be introduced in the subsequent chapters: The powered part of the PMLSM drive will be referred to as the primary part, and the part containing only permanent magnets will be referred to as the secondary part of the motor. The stator consists of the track on which the slider moves. It is generally possible to create either a primary stator or a primary slider, so the terms are not universally interchangeable. The following chapters will primarily focus on the case of a powered slider, where the slider is the primary part of the motor. At the same time, the track contains the permanent magnets and thus becomes the secondary part of the motor.

The advantages of linear motor drives have been summarized in the previous paragraphs. PMLSM drives have simple construction and are more expensive, but they can generate high force, high speed, and high precision. However, it is also necessary to highlight their disadvantages. Unlike rotary motors, they have a limited range of motion and cannot be operated in speed control mode [10]. Additionally, gears cannot be used, so changes in load often directly affect the inverter, necessitating adjustments in control [19, 20, 21]. Another disadvantage, which will be discussed in more detail in the following chapters, is the so-called end effect. This occurs due to magnetic flux leakage at the edges of the slider, resulting in ripples of the electromotive force. This effect must be considered and mitigated in most control applications [12, 13].

1.2 Control of Permanent Magnet Linear Synchronous Drives

As previously mentioned, linear motor drives differ from rotary motors in control aspects almost exclusively by methods mitigating additional parasitic influence of the end effect. Consequently, the same control methods used for rotary motors are applied to linear motors. Applications include both scalar and vector classical control using two-degree-of-freedom controllers [22, 23, 24] or feedforward control [25, 26, 27, 28, 29, 30]. Examples also include the direct torque control [31], sensorless control [32], backstepping control [33, 34], and adaptive backstepping control [25, 26].

Modern methods based on neural networks are also popular, such as fuzzy neural networks [27], Elman neural networks [22, 35, 28, 36], modified Elman networks [37, 29], and static neural networks [23, 24, 38]. Optimal control is not specific to rotary motors either, as evidenced by applications utilizing genetic algorithms [39], Grey Wolf optimization [40], or particle swarm optimization algorithms [5, 41, 42, 43].

Apart from these methods, which primarily utilize vector control, optimal control design is also used. These designs can employ various motor modeling techniques beyond the state-space description. An example is optimal control based on magnetic equivalent circuits [17].

The quality of control is usually directly dependent on the quality of the motor model used. Therefore, the following section will focus on modeling linear motor drives and their identification. Furthermore, methods providing stochastic estimation of unknown parameters are introduced later in the thesis. The knowledge of uncertainty allows for the development of diagnostic methods for the PMLSM drive.

1.3 Model of Permanent Magnet Synchronous Motor

Modern control, filtering, estimation, and diagnostic algorithms are founded on the knowledge of the motor model. Linear motors are very similar to rotary motors, on whose principles they are often based. Therefore, the same modeling methods used for rotary motors can be applied to linear motors. Although other modeling methods offer specific advantages and disadvantages, such as equivalent magnetic circuit modeling [44, 45, 46, 47, 48, 49, 50], which allows for the modeling of certain parasitic phenomena, this work will focus on the most widely used model. Specifically, it will consider the state-space model describing currents in the dq coordinates and speed. The model will first be derived for a rotary motor and then adapted for a linear motor.

The model of a PMSM motor in dq coordinates is widespread, so detailed derivation is not necessary. For more detailed derivation and examples of use, one can refer to [51, 52, 53, 54, 55, 56, 57, 58].

The derivation is based on the voltage equations of the d and q components of stator voltage.

$$u_d = Ri_d + \dot{\lambda}_d - \omega_e \lambda_q, \tag{1.1}$$

$$u_q = Ri_q + \dot{\lambda}_q + \omega_e \lambda_d, \tag{1.2}$$

where $\lambda_d = L_d i_d + \Psi_{PM}$ and $\lambda_q = L_q i_q$ are the stator magnetic fluxes in dq coordinates, R is the stator resistance, and ω_e is the rotor electrical angular velocity, L_d and L_q are direct, and quadrature inductances respectively, i_d and i_q are direct and quadratic stator currents respectively and lastly, Ψ_{PM} is the permanent magnet flux; the dot above the magnetic fluxes denotes time derivative.

The electromechanical torque T_e is then

$$T_e = \frac{3p_p}{2} \left(\Psi_{PM} i_q + (L_d - L_q) i_q i_d \right) = T_L + B\omega_m + J\dot{\omega}_m, \tag{1.3}$$

where p_p is the number of pole pairs, T_L is the load torque, B is the damping coefficient, and ω_m is the mechanical angular velocity, i.e., the angular velocity of the rotor; $\dot{\omega}_m$ is the time derivative of mechanical angular velocity, the mechanical angular acceleration.

This model holds under the assumptions [51]:

1. neglecting saturation, which can be modeled by changing the model parameters,
2. the electromotive force is harmonic,
3. eddy current and hysteresis losses are negligible,
4. the rotor is not cage-type[1].

The model can be adjusted into the form of state equations

$$\dot{i}_d = \frac{1}{L_d}\left(L_q i_q \omega_e - R i_d + u_d\right), \tag{1.4a}$$

$$\dot{i}_q = \frac{1}{L_q}\left(-L_d i_d \omega_e - \Psi_{PM}\omega_e - R i_q + u_q\right), \tag{1.4b}$$

$$\dot{\omega}_m = \frac{1}{J}\left(\frac{3p_p}{2}\left(\Psi_{PM} i_q + (L_d - L_q) i_q i_d\right) - T_L - B\omega_m\right), \tag{1.4c}$$

which hold under the condition

$$\omega_e = p_p \omega_m. \tag{1.5}$$

The problem with this model is that conditions 1 to 3 are practically never met, which is especially evident in torque ripple. Multiple sources of ripple are included under the term cogging torque, which can act in either direction of the motor's rotation. The sources of ripple can be divided into two groups.

The first group includes ripple sources caused by the motor's design. Improved properties can be achieved, for example, by using magnets of different qualities [60], shapes [61, 62], or heights [63]. The analysis and analytical modeling of the magnetic field are usually too complex to include all parasitic phenomena, such as saturation or the distribution of magnetic flux [64, 65, 66, 67, 68, 61, 60, 63, 62]. Therefore, other methods are used, such as finite element analysis [69, 70, 71] or the method of equivalent magnetic circuits [45, 46, 47, 48, 49, 50]. The design and associated analysis of the motor are not the focus of this work, and it will be assumed that the design cannot be changed.

The second group consists of cogging torques caused by imperfections or faults in the motor's construction. These torques cannot be prevented. They manifest as torque fluctuations with a period of at most one rotation. Most commonly, however, the torque fluctuations occur with a period that is an integer multiple of the electrical rotation, which happens, for example, when there is slot curvature that manifests as any pole approaches. In PMSM drives, cogging torque lead to periodic disturbances that can be estimated and suppressed, for example, by an inverse periodic function describing one mechanical rotation of the motor.

[1]Cage-type PMSM motors (LS-PMSM) are used as an alternative to ACIM motors [59]

1.4 Model of Permanent Magnet Linear Synchronous Motor

The same model applies to a PMLSM motor under the same conditions. Since it is a linear motor drive with linear motion, it is necessary to convert the angular velocity dependence ω_e to the linear speed of the mover v, the moment of inertia J to the mass of the moving part M, and the dependence of the electrical torque T_e to the dependence on the electrical force F_e. The voltage equations in dq coordinates are then given by

$$u_d = Ri_d + \dot{\lambda}_d - \frac{\pi}{\tau_s} v \lambda_q, \tag{1.6a}$$

$$u_q = Ri_q + \dot{\lambda}_q + \frac{\pi}{\tau_s} v \lambda_d, \tag{1.6b}$$

where τ_s is the pole pitch. The electrical force is expressed as

$$F_e = \frac{3\pi p_p}{2\tau_s} \left(\Psi_{PM} i_q + (L_d - L_q) i_q i_d \right). \tag{1.7}$$

It is again possible to derive the state equations.

$$\dot{i}_d = \frac{1}{L_d} \left(\frac{L_q \pi}{\tau_s} i_q v - R i_d + u_d \right), \tag{1.8a}$$

$$\dot{i}_q = \frac{1}{L_q} \left(-\frac{L_d \pi}{\tau_s} i_d v - \frac{\Psi_{PM} \pi}{\tau_s} v - R i_q + u_q \right), \tag{1.8b}$$

$$\dot{v} = \frac{1}{M} \left(\frac{3\pi p_p}{2\tau} \left(\Psi_{PM} i_q + (L_d - L_q) i_q i_d \right) - F_L - Bv \right), \tag{1.8c}$$

where M is the mass of the entire moving part, which in most cases includes the mover and the load, and F_L is the load force. As evident, the models are equivalent with the substitutions $\omega_e = \frac{\pi}{\tau_s} v$, $J = M$, and $T_e = F_e \frac{\tau_s}{\pi}$.

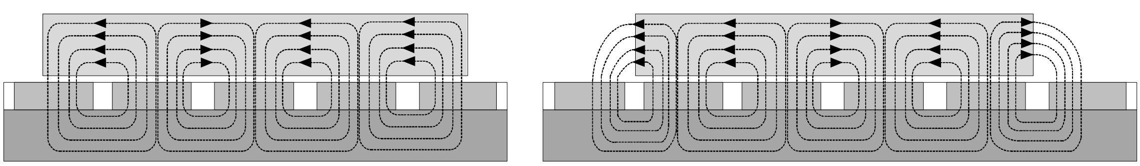


Fig. 1.1: Differences between an ideal (left) and non-ideal (right) linear motor drive with permanent magnets and a primary mover. Source: [72]

This model type is again prevalent, as evidenced by for example [73, 74, 75, 76, 77, 78, 79, 8]. Just as with rotary drives, the conditions for the accuracy of this model are not met, resulting in ripples in the simulated force compared to the true generated force. In the case of PMLSM drives, both sources of ripples occur, as with PMSM. The group caused by the motor's design is again periodic and can be suppressed in the same way as with PMSM motors. The second group is not periodic and requires describing the ripple with an aperiodic function that accounts for the entire drive path. In the case of PMLSM drives, there is an additional source of ripple called the end effect caused by the drive's design, as shown in Figure 1.1. The end effect leads to a periodic disturbance that cannot be distinguished from the disturbances in the first group. The end effect, like the cogging force, can be minimized by the motor's design [80].

Further reduction of detent forces, that is, cogging force and end effect, can be achieved using an appropriate control algorithm. Typically, detent forces are measured as deviations from the model and described by a function, whose inverse function is then forwardly injected into the control algorithm [81]. As there is usually no way of distinguishing the effect from other detent forces, cogging mitigation methods are successfully used in PMLSM drives to suppress both cogging and effect. It is, however, essential to acknowledge this difference between the model and actual motor for identification.

1.5 Identification of Permanent Magnet Linear Synchronous Drive

Previous chapters focused on formulating the linear motor model, emphasizing the critical importance of an accurate model for the control of the motor. However, an accurate model formulation alone is insufficient for a control algorithm. It is necessary to determine the correct model parameters to reflect real-world conditions accurately. While some parameters can be considered known, as they can be directly measured or are documented in relevant technical literature, other parameters, such as inductances, exhibit nonlinear dependencies on the operating point and are typically identified only for specific operating point. The process of determining such model parameters is known as model identification, and it can be performed using various approaches.

Identification methods can be categorized into two main groups: offline and online. Offline identification methods are conducted before the controlled process, necessitating preparation before the initial operation and process interruption if additional identification is required. The primary drawback of these methods is their inability to adjust the model parameters in real time, rendering them incapable of responding to parameter changes and disturbances. This limitation is addressed by online identification methods, which enable realtime identification and continuous adjustment of the estimated parameters during motor operation. Due to the similarities between PMSM and PMLSM drives, PMSM identification methods can be applied to PMLSM systems. This means that although this thesis refers almost exclusively to PMLSM drives, results directly apply to PMSM drives.

An example of an offline method is the current decay test under direct current voltage [82, 83, 84, 85, 86]. The test is conducted by positioning the rotor at an electrical angle of $\alpha_e = \frac{\pi}{2}\,rad$. Subsequently, a direct current voltage is applied to two phases of the motor, causing the rotor to align to an electrical angle of $\alpha_e = 0\,rad$, during which the current profile is measured. From the voltage and current profiles, the inductances of the model in dq coordinates are then calculated.

Another example of an offline method is the measurement of mutual inductance with the rotor at a standstill [87, 88, 89, 90]. This method leverages the fundamental properties of the coils in the motor. By applying an alternating voltage to two coils and observing the phase shift in the third coil, information about the self and mutual inductance of the coils can be obtained. The issue with both methods is that the inductance is not measured for the specific operating point at which the motor will be operated. Additional offline identification methods can be studied in [91].

Today's computational technology enables continuous, real-time identification, which offers undeniable advantages over offline identification, primarily when the motor is operated across multiple operating points. A popular example of online identification algorithms includes those utilizing neural networks [92, 93, 94, 95]. However, these methods often provide results that are difficult to interpret.

Popular alternatives are methods based on least squares, for example, [96]. These methods involve solving a system of equations derived from measurements of a suitably excited motor. Subsequently, optimization is used to find parameters that minimize the squared difference between the model output estimate and the actual measured values of the motor output. These methods allow minor adjustments to perform offline identification, online identification, and online identification with forgetting.

From the perspective of Bayesian statistics, least squares methods can be interpreted as maximum likelihood methods. Such an interpretation further enables extending results to statistical results and defining additional estimators, such as maximum a posteriori probability estimators or the Kalman filter. Bayesian statistics opens up possibilities for new identification methods and is a highly advantageous perspective that has been widely utilized in recent times.

However, applying these methods often leads to ignorance of some measurement errors, which can be prevalent in the case of the PMLSM. This thesis aims to formulate the identification of linear motor drives as an errors-in-variables problem, which allows for seamless incorporation of measurement errors and explanation of other phenomenons, for example, detention forces, as measurement uncertainty.

2 Preliminaries

The preliminary chapter is included in this work to establish a necessary foundational understanding that enables readers to navigate and understand the research presented in the subsequent chapters and its connection with existing methods.

2.1 Model definition

Throughout this work, we will consistently use a linear model

$$\overline{y}_i = \overline{x}_i^T \theta, \tag{2.1}$$

where $\overline{y}_i \in \mathbb{R}$ and $\overline{x}_i \in \mathbb{R}^m$ are dependent variable (output) and independent variable (input), respectively; $\theta \in \mathbb{R}^m$ is a sought-after vector of unknown parameters; the m denotes the dimension of the problem.

This model provides a simplified and effective framework for analyzing and understanding a wide range of systems, including the PMLSM, the system under study. This choice allows us to focus on the core aspects of the problem while maintaining a transparent and manageable structure in our research. Reformulation of PMLSM electrical state-space equations (1.8a) and (1.8b) into linear model will be presented in the following chapter.

2.2 Dataset notation

Let us introduce a new notation for a dataset of N samples s available at time k. This dataset will be further denoted as sequence

$$S(k) := \{s_i\}_{i=0}^{N}. \tag{2.2}$$

In most cases, time specifications are not required and will be omitted for improved readability.

Without loss of generality, the dataset $S(k)$ can also be represented as a matrix of individual transposed vectors stacked underneath each other.

For example, assuming $s_i = [s_{i,1},\, s_{i,2},\, s_{i,3}]^T \in \mathbb{R}^3$, the resulting matrix representation is

$$S(k) = \begin{bmatrix} s_{i,1} & s_{i,2} & s_{i,3} \\ s_{i,1} & s_{i,2} & s_{i,3} \\ & \vdots & \\ s_{N,1} & s_{N,2} & s_{N,3} \end{bmatrix}. \tag{2.3}$$

As intended representation is always apparent from the context, the representations will be used interchangeably.

2.3 Generalized singular value decomposition

The singular value decomposition (SVD) is a well-known, extensively applied decomposition method used in both theoretical and applied contexts, as well as in this thesis. The SVD decomposes given matrix $A \in \mathbb{R}^{m\times n}$ into three components: $A = U\Sigma V^T$, where $U \in \mathbb{R}^{m\times m}$ is an orthogonal matrix of left singular vectors, $V \in \mathbb{R}^{n\times n}$ is an orthogonal matrix of right singular vectors and $\Sigma \in \mathbb{R}^{m\times n} = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r)$ is a diagonal matrix of singular values. Mathematical operator $diag\,(x_1, x_2, \ldots)$ represents a diagonal matrix with the specified elements $x_1, x_2, \ldots$ on its main diagonal. The singular values are usually sorted in descending order $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$, where $r = \min(m, n)$.

Less known is the generalization of the SVD, the generalized SVD (GSVD). As the GSVD is less known, let us define it here, using Theorem from [97]. To comprehensively explore GSVD, including its properties and relationships with other decompositions, refer to [98].

Theorem 1. *GSVD of a matrix pair (A, B). If $A \in \mathbb{R}^{m\times n}$ $(m \geq n)$ and $B \in \mathbb{R}^{p\times n}$, then there exist orthogonal $T \in \mathbb{R}^{m\times m}$ and $W \in \mathbb{R}^{p\times p}$ and a nonsingular $Z \in \mathbb{R}^{n\times n}$ such that*

$$T^T AZ = D_A \quad \textit{and} \quad W^T BZ = D_B \tag{2.4a}$$

with

$$D_A = \text{diag}(\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{R}^{m\times n}, \quad \alpha_i \geq 0 \tag{2.4b}$$

and

$$D_B = \text{diag}(\beta_1, \beta_2, \ldots, \beta_q) \in \mathbb{R}^{p\times n}, \quad \beta_i \geq 0, \quad q = \min(p, n) \tag{2.4c}$$

$$\beta_1 \geq \cdots \geq \beta_r > \beta_{r+1} = \cdots = \beta_q = 0, \quad r = rank(B). \tag{2.4d}$$

The elements of set $\{\alpha_i/\beta_i\}_{i=0}^{r}$ are known as the ordinary generalized singular values of the matrix pair (A, B), the remaining generalized singular values in which α_i are nonzero are called infinite, and in the case when the α_i is zero, the singular values are called indefinite. The invertible square matrix Z represents the right singular vectors associated with both A and B.

The GSVD is a generalization of the SVD in that if $B = \mathbf{I}$, then the generalized singular values are equal to the singular values of A.

2.4 Laplace Approximation

The Laplace approximation is a powerful method used in Bayesian statistics to approximate complex posterior distributions or problematic probability distributions in general. Named after the mathematician Pierre-Simon Laplace, this technique leverages the simplicity of Gaussian distributions to simplify the often intractable integrals involved in Bayesian inference. By approximating the probability distribution with a Gaussian centered at the mode, the Laplace approximation transforms the challenging problem of computing the posterior into a more manageable one. This approach is particularly useful in scenarios where the exact posterior is challenging to work with, providing a balance between computational efficiency and accuracy. It is widely used in various applications, including machine learning, which helps derive insights from complex models. For a more detailed analysis and derivation of the Laplace approximation method, refer to [99, 100].

Assuming probability distribution

$$p(x|\mathcal{D}), \tag{2.5}$$

where x is the random variable and $\mathcal{D}$ is set of distribution parameters, the surrogate distribution $q(x|\mathcal{D}')$ is obtained as

$$q(x|\mathcal{D}') \approx p(x|\mathcal{D}). \tag{2.6}$$

The surrogate, normal posterior probability distribution is

$$q(x|\mathcal{D}') = \mathcal{N}(x|\hat{\mu}, H(\hat{\mu})), \tag{2.7}$$

where the mean $\hat{\mu}$ is the modus of the original probability distribution and the covariance matrix $H(\hat{\mu})$ is the Fisher information of the sample data, or identically, the Hessian matrix of the log-probability $\ln p(x|\mathcal{D})$ evaluated at the point $\hat{\mu}$

$$H(\hat{\mu}) = -\nabla\nabla^T \ln p(x|\mathcal{D})\Big|_{x=\hat{\mu}}. \tag{2.8}$$

2.5 Bingham Distribution

The Bingham distribution emerges from a zero-mean multivariate normal distribution in $\mathbb{R}^{m+1}$, constrained to the unit hyper-sphere $\mathbb{S}^m$. Introduced by [101] as an antipodally symmetric distribution on the sphere $\mathbb{S}^2$, it was later generalized to higher-dimensional sphere $\mathbb{S}^m$. This section is adapted from the author's earlier research presented in [2].

The density function is given by

$$\mathcal{B}(x|A) = c(A)^{-1} \exp\left(-x^T A x\right), \tag{2.9}$$

where $x \in \mathbb{R}^{m+1}$, $x^T x = 1$, and A is a $(m+1) \times (m+1)$ symmetric, positive definite matrix. The term $c(A)$ is an intractable integration constant, which poses the main complication of the Bingham distribution [102].

This distribution finds applications in various fields, such as paleomagnetic studies [103], wind speed modeling [104], biomedical image analysis [105], crystal orientation analysis [106], and orientation estimation [107].

The covariance matrix A can be spectrally decomposed, for example using singular value decomposition, into $A = \Gamma\Lambda\Gamma^T$, where Γ is an orthogonal matrix constructed from the eigenvectors of A, and $\Lambda = diag(\lambda_1, \ldots, \lambda_{m+1})$ is the diagonal matrix of the eigenvalues of A. Without loss of generality, we assume that the concentration parameters Λ satisfy the identifiability constraint $\lambda_1 \geq \cdots \geq \lambda_m \geq \lambda_{m+1} = 0$. These constraints ensure identifiability, as the density remains unchanged if a positive constant is added to the concentration parameter vector.

By decomposing the covariance matrix A, we obtain the concentration parameter matrix Λ and the matrix of concentration vectors $\Gamma = [\gamma_1, \gamma_2, \ldots, \gamma_{m+1}]$, where $\gamma_1, \gamma_2, \ldots, \gamma_m$ are concentration axes vectors with corresponding concentration parameters $\lambda_1, \lambda_2, \ldots, \lambda_m$. The vector γ_{m+1} is the mean concentration vector, also known as the modus vector.

The Bingham distribution can also be defined by the parameters Γ and Λ as

$$\mathcal{B}(x|\Gamma, \Lambda) = c(\Lambda)^{-1} \exp\left(-x^T \Gamma\Lambda\Gamma^T x\right). \tag{2.10}$$

While this notation is more widely used, the version (2.9) is used in this thesis for improved readability.

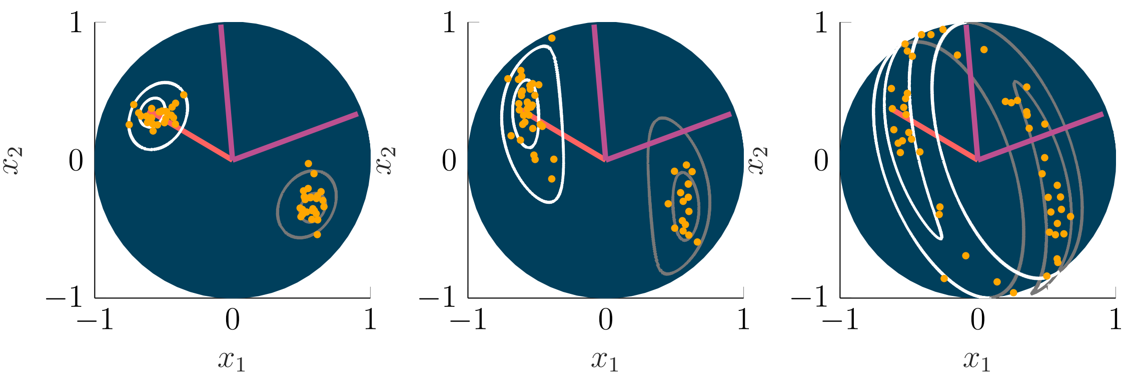


Fig. 2.1: Orthographic projection of Bingham distribution contour plot on the sphere $\mathbb{S}^2$. The red line is the mean concentration vector; purple lines are concentration vectors, and concentration parameters are $\Lambda = diag(-60, -60, 0)$, $\Lambda = diag(-60, -10, 0)$ and $\Lambda = diag(-60, -1, 0)$ from left to right, respectively. Yellow dots are samples drawn from the distributions.

The standard form of the distribution is

$$p(x|\Lambda) = c(\Lambda)^{-1} \exp\left\{-\sum_{i=1}^{m} \lambda_i x_i^2\right\} \tag{2.11}$$

with respect to a uniform measure on the sphere and

$$c(\Lambda) = \int_{x\in\mathbb{S}^m} \exp\left\{\sum_{i=1}^{m} \lambda_i x_i^2\right\} d\mathbb{S}^m(x). \tag{2.12}$$

Notably, the standard form does not include the concentration vectors Γ, as it conveniently follows from the rotation of the distribution such that the concentration vectors align with the main axes. If Z follows a Bingham distribution with density $p(x|A)$, then $W = \Gamma Z$ follows a Bingham distribution with density $p(x|\Lambda)$ (see [108] and [101] for a detailed explanation).

Let $Z^{(N)}$ be a set of N samples of unit vectors in $\mathbb{S}^m$, sampled from a Bingham distribution with density $p(x|\Lambda)$. The likelihood function is obtained as

$$\mathcal{L}(\Lambda) = c(\Lambda)^{-N} \exp\left\{-\sum_{i=1}^{m} \lambda_i \sum_{j=1}^{N} \left(x_j^i\right)^2\right\} \tag{2.13}$$

$$= c(\Lambda)^{-N} \exp\left\{-\sum_{i=1}^{m} \lambda_i \tau_i\right\}, \tag{2.14}$$

where $\tau_i = \frac{1}{N}\sum_{j=1}^{N}\left(x_j^i\right)^2$. The sufficient statistics for Λ is the set of $N, \tau_1, \ldots, \tau_m$.

The maximum likelihood estimator of the matrix of concentration vectors Γ is $\sum_{i=1}^{N} x_i x_i^T$.

To the author's knowledge, analytical formulation of moments of the Bingham distribution have not yet been derived. However, moments for similar directional probability distributions [109], the von Mises and the Fisher distributions have been derived using a method expected to be extendable to the Bingham distribution [110]. Numerically approximated moments can be obtained using a fast algorithm introduced by [111].

Establishing confidence limits based on concentration parameters has proven to be complicated. An approximate formula was discovered by [101]. Another notable method for establishing confidence limits, popular in paleomagnetism, is presented in [112].

Although the cumulative distribution function of the Bingham distribution cannot be expressed due to the doubly intractable normalizing constant [113, 114], sampling using the Metropolis-Hastings algorithm is possible by approximating the normalization constant [115]. Other methods bypassing the normalizing constant altogether are presented in [116] and [117]. In this paper, we utilize the reversible-jump MCMC [118] method introduced by [115] for the convenience of MATLAB implementation in libDirectional [119].

2.6 Least Squares

The least squares (LS) is a well-known, widely used identification method. Assuming samples from linear model (2.1), obtained output samples are burdened with model error

$$\overline{y}_i = y_i + \varepsilon_i; \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2), \tag{2.15}$$

where ε_i is the additive noise component caused by model imperfections. The usual noise assumption, also used in this thesis, is that the ε is independent and identically, normally distributed random noise with variance σ^2 and zero mean.

In this case, the least squares solution is obtained by finding such correction $\hat{Y} = \{\hat{y}_i\}_{i=0}^N$ of noisy outputs $\overline{Y} = \{\overline{y}_i\}_{i=0}^N$, such that the errors between true and noisy outputs are minimal in the sense of the L_2 norm, subject to the linear model (2.1). This can be achieved by solving the following constrained optimization problem [120]:

$$\begin{aligned} [\hat{\theta}_{LS}, \hat{Y}] := \min_{\theta, \hat{Y}} \quad & \|\overline{Y} - \hat{Y}\|_2 \\ \text{s.t.} \quad & \theta \overline{X} = \hat{Y} \end{aligned} \tag{2.16}$$

The vector of corrected outputs $\hat{Y}$ is typically unnecessary for the analysis. Since the sought-after parameter vector θ can be derived analytically without the explicit computation of $\hat{Y}$, it is often omitted from the process altogether.

The analytical solution is obtained by [121]

$$\hat{\theta}_{LS} = (\overline{X}^T \overline{X})^{-1} \overline{X}^T \overline{Y}, \tag{2.17}$$

where $\overline{X} = \{\overline{x}_i\}_{i=0}^N$ is sequence of inputs.

The simple analytical solution in connection with the Sherman–Morrison formula leads to a well-known recursive algorithm, allowing for sequential incorporation of new samples [120].

Algorithm 1 Recursive Least Squares (RLS)

Input: $\overline{X}$, $\overline{Y}$, $\hat{\theta}_{LS,0}$, P_0, λ
Output: $\hat{\theta}_{LS}$

```
θ̂_LS ← θ̂_LS,0
P ← P_0
for i ← 1 to N do
    L ← (P x̄_i)/(λ + x̄_i^T P x̄_i)
    θ̂_LS ← θ̂_LS + L(ȳ_i − x̄_i^T θ̂_LS)
    P ← (I − L x̄_i^T) P (1/λ)
end for
```

In the Algorithm 1, $\overline{X}$ and $\overline{Y}$ are data sequences, $\lambda \in \mathbb{R}$ is forgetting factor, $\hat{\theta}_{LS,0}$ is initial guess of the $\hat{\theta}_{LS}$ and $P_0 \in \mathbb{R}^{m\times m}$ is initial value of the precision matrix of the parameter estimates. It is usually set to $c\mathbf{I}_d$, where c is a large number.

Assuming that the underlying model is autoregressive, this definition of least squares aligns with the well-known autoregressive model with exogenous inputs (ARX)

$$y_i = -\sum_{n=1}^{N} a_n \overline{y}_{i-n} + \sum_{m=0}^{M} b_m \overline{u}_{i-m} + \varepsilon_i, \tag{2.18}$$

which ignores the measurement error, and ε represents an error in the model. As the erroneous value $\overline{y}_i$ enters the model, the ARX can be expressed with the equations (2.1) and (2.15) by following substitutions

$$\overline{x}_i^T = [\overline{y}_{i-1}, \overline{y}_{i-2}, \dots, \overline{y}_{i-N}, \overline{u}_{i-0}, \overline{u}_{i-1}, \dots, \overline{u}_{i-M}]^T, \tag{2.19}$$
$$\theta^T = [-a_1, -a_2, \dots, -a_N, b_0, b_1, \dots, b_M]^T. \tag{2.20}$$

While this is a correct method, one could argue that the model error can be neglected, as the sample periods are oftentimes small. On the other hand, small sample periods can amplify the measurement errors, and the measurement errors can become dominant. In such a case, it is more appropriate to assume that the model is true, i.e., ignoring the model error. The measurement errors and parameter uncertainties can explain all model-data discrepancies (like for example detention forces). In such case, the output error should be modeled as

$$y_i = \overline{y}_i + \varepsilon_i; \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_y^2), \tag{2.21}$$

Additional uncertainty can also be expressed for the input variable

$$x_i = \overline{x}_i + \varepsilon_i; \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_x^2). \tag{2.22}$$

This model is called the errors-in-variables model and should be described as

$$\overline{x}_i^T = [\overline{y}_{i-1}, \overline{y}_{i-2}, \dots, \overline{y}_{i-N}, \overline{u}_{i-0}, \overline{u}_{i-1}, \dots, \overline{u}_{i-M}]^T, \tag{2.23}$$
$$\theta^T = [-a_1, -a_2, \dots, -a_N, b_0, b_1, \dots, b_M]^T. \tag{2.24}$$

However, as the true data $\overline{y}_i$ and sometimes even $\overline{u}_i$ is only known with some uncertainty, the best approximation of $\overline{x}_i \approx \hat{\overline{x}}_i$ is

$$\hat{\overline{x}}_i^T = [y_{i-1}, y_{i-2}, \dots, y_{i-N}, u_{i-0}, u_{i-1}, \dots, u_{i-M}]^T, \tag{2.25}$$

which is usually the case for PMLSM identification problems, where the regressor $\hat{\overline{x}}_i$ usually consists solely of noisy current, voltage and position measurements.

The presence of delayed, noisy measurements in the $\hat{\overline{x}}_i$ violates the assumptions of the LS approach. Thus, the utilization of the LS method is considered inappropriate by definition. This issue will be addressed in subsequent sections by introducing the Total Least Squares method and its generalization, enabling the resolution of more complex problems.

2.7 Total Least Squares

The total least squares (TLS) is a well-studied but less popular identification method compared to LS [4, 122, 123]. The TLS also allows the identification of unknown vector θ from data coming from the linear model (2.1). In the case of TLS, all samples are burdened with noise:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} \overline{x}_i \\ \overline{y}_i \end{bmatrix} + \varepsilon_i, \quad \varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}), \tag{2.26}$$

which makes this model an errors-in-variables model.

Similarly to LS, the ε_i is an independent, identically, normally distributed random additive noise with zero mean and diagonal covariance $\sigma^2 \mathbf{I}$.

The TLS estimate is shown in [124] to be an unbiased maximum likelihood estimator of θ within first-order error terms. For a comprehensive review of the TLS, refer to [123].

As all variables are noisy, there is no need to distinguish between inputs and outputs, and sequence of all measurements can be concatenated to

$$d_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}, \quad \overline{d}_i = \begin{bmatrix} \overline{x}_i \\ \overline{y}_i \end{bmatrix}, \quad D(k) := \{d_i\}_{i=0}^{N}, \quad \overline{D}(k) := \{\overline{d}_i\}_{i=0}^{N}. \tag{2.27}$$

The solution is obtained by minimizing the distance of measured data to the flat affine subspace of dimension m, defined by the normal vector $\vartheta = [\hat{\theta}_{\mathrm{TLS}}^T, -1]^T$. In other words, the orthogonal distance is minimized. For this reason, total least squares is sometimes referred to as an orthogonal regression.

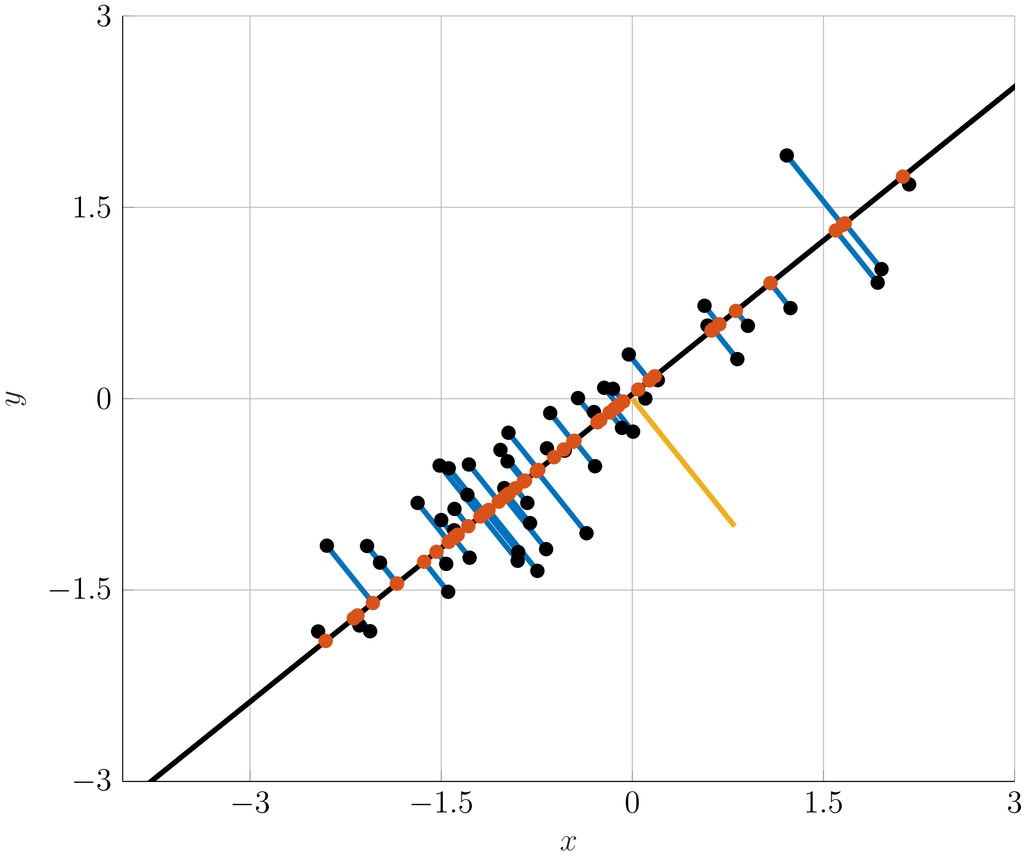


Fig. 2.2: Graphical illustration of orthogonal regression, also known as the total least squares method. The estimated line (black line) is obtained from data (black dots) by minimizing the orthogonal distance (blue lines) from the solution. Analogously, distances from the orthogonal projection (red dots) to the measured data (black dots) are minimized. The solution plane is fully specified by the yellow vector $[\hat{\theta}_{\mathrm{TLS}}^T, -1]^T$.

The problem can be defined as minimizing Frobenius norm between corrected data $\hat{D} = \{\hat{d}_i\}_{i=0}^N$ and the measured data D subject to the linear model (2.1). It is, therefore possible to find the solution by solving the following constrained optimization problem:

$$\begin{aligned} [\hat{\theta}_{\mathrm{TLS}}, \hat{D}] := \min_{\theta, \hat{D}} \quad & \|D - \hat{D}\|_F \\ \text{s.t.} \quad & 0 = \hat{D}\vartheta. \end{aligned}$$

Once again, the corrected data sequence $\hat{D}$ is usually not required. As the dimension of $\hat{D}$ grows with a number of samples, identification methods try to omit calculating it.

In the case when corrected data sequence $\hat{D}$ is needed, it can be calculated using orthogonal projection of the measured data D to the hyperplane defined by ϑ:

$$\hat{D} = \{\hat{d}_i\}_{i=0}^N = \left\{ \frac{d_i^T \vartheta}{\vartheta^T \vartheta} \vartheta \right\}_{i=0}^N . \tag{2.28}$$

The following theorem, formulated by [123], provides the TLS solution and outlines the assumptions necessary for its uniqueness:

Theorem 2. *Solution of the total least squares problem.*
Let

$$D = V \Sigma V^T, \text{ where } \Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_{m+1})$$

be a singular value decomposition of D and $\sigma_i,\ i \in \{1, 2, \dots, m+1\}$ be the singular values of D. After defining the partitioning

$$V = \begin{array}{c} \begin{array}{cc} m & 1 \end{array} \\ \begin{bmatrix} V_{11} & v_{12} \\ v_{12}^T & v_{22} \end{bmatrix} \end{array}, \quad \Sigma = \begin{array}{c} \begin{array}{cc} m & 1 \end{array} \\ \begin{bmatrix} \Sigma_1 & \mathbf{0} \\ \mathbf{0} & \sigma_{m+1} \end{bmatrix} \end{array},$$

a TLS solution exists if and only if v_{22} is non-zero. In addition, it is unique if and only if $\sigma_m \neq \sigma_{m+1}$. In the case when the TLS solution exists and is unique, it is given by

$$\hat{\theta}\, TLS = \frac{-v_{12}}{v_{22}},$$

the corresponding TLS correction matrix is

$$D - \hat{D} = V \ \operatorname{diag}(\mathbf{0}, \sigma_{m+1})\, V^T$$

and nuisance variables are

$$\hat{D} = V \ \operatorname{diag}(\Sigma_1, \mathbf{0})\, V^T.$$

Calculation of full singular Value Decomposition (SVD) is often avoided in practical applications due to its high computational complexity. Instead, only the singular vector corresponding to the smallest singular value $[v_{12}^T, v_{22}]^T$ is computed by minimizing the Rayleigh quotient

$$\begin{bmatrix} v_{12} \\ v_{22} \end{bmatrix} := \arg\min_v \frac{v^T \left(D^T D\right) v}{v^T v}, \tag{2.29}$$

of which the minimal value is the smallest singular value.

The optimum can be achieved using the inverse iteration algorithm [98]. The inverse iteration, also termed the inverse power method, constitutes an iterative procedure frequently employed to compute the eigenvalues and eigenvectors of a matrix.

Distinguished among its advantageous features are its rapid convergence and numerical stability. Moreover, its straightforward implementation makes the algorithm suitable for online applications.

Algorithm 2 Inverse Iteration (II)

Input: D, v_0

Output: v

1: $v \leftarrow v_0$

2: **while** Stopping condition **do**

3: $\quad v \leftarrow \dfrac{(D^T D)^{-1} v}{\|(D^T D)^{-1} v\|}$

4: **end while**

In Algorithm 2, v_0 represents the initial estimate for the singular vector, and D denotes the matrix in consideration. The algorithm is typically terminated when the change in v between two successive iterations falls below a specified threshold.

The convergence rate can be further improved by using so-called shifts, which means replacing line 3 in Algorithm 2 with

$$v \leftarrow \frac{(D^T D - \mu \mathbf{I})^{-1} v}{\|(D^T D - \mu \mathbf{I})^{-1} v\|}. \tag{2.30}$$

In the particular case of finding the smallest eigenvector, μ is the smallest eigenvalue. As it is usually not known in advance, it can be approximated in each iteration by the Rayleigh quotient

$$\mu = \frac{v^T \left(D^T D\right) v}{v^T v}. \tag{2.31}$$

This extension of the inverse iteration is known as the Rayleigh quotient iteration.

The algorithm exhibits sufficiently rapid convergence, allowing for only a single iteration to be executed for each sample. Incorporating each sample into the inverted matrix can be achieved using the Woodbury matrix inversion lemma. The resulting algorithm proposed in [120] is presented in Algorithm 3.

Algorithm 3 Recursive Total Least Squares (RTLS)

Input: D, $\hat{\theta}_{TLS,0}$, P_0, λ

Output: $\hat{\theta}_{TLS}$

$\hat{\theta}_{TLS} \leftarrow \hat{\theta}_{TLS,0}$

$P \leftarrow P_0$

for $i \leftarrow 1$ to N **do**

  $L \leftarrow (Pd_i)/(\lambda + d_i^T P d_i)$

  $P \leftarrow (\mathbf{I} - Ld_i^T)P\frac{1}{\lambda}$

  $Z \leftarrow P[\hat{\theta}_{TLS}^T, -1]^T$

  $\hat{\theta}_{TLS} \leftarrow -z_{12}/z_{22}$

end for

In Algorithm 3, the d_i represents a value from the data sequence D as proposed in (2.27). The parameter $\lambda \in \mathbb{R}$ is the forgetting factor, $\hat{\theta}_{TLS,0}$ denotes initial guess for $\hat{\theta}_{TLS}$ and $P_0 \in \mathbb{R}^{(m+1)\times(m+1)}$ is the initial value of precision matrix; it is typically set to $c\mathbf{I}_{m+1}$, where c is a large constant. The vector Z is $[z_{12}^T, z_{22}]^T$, where z_{22} is the last element of the vector.

2.8 Generalized Total Least Squares

The total least squares problem has several notable generalizations that account for different restrictions on the covariance of the noise ε. Among these, there is the mixed least squares total least squares (LS-TLS) problem and the element-wise weighted TLS (EW-TLS) problem; lastly, there is the most comprehensive of these generalizations, the generalized total least squares (GTLS) problem. For an overview of the mentioned and other generalizations, refer to [125, 122] and [126].

The LS-TLS problem assumes

$$\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \Sigma_{\text{LS-TLS}}), \tag{2.32}$$

where $\Sigma_{\text{LS-TLS}}$ is a diagonal matrix consisting exclusively of zeros and ones. Typically, the variables are reorganized such that $\Sigma_{\text{LS-TLS}} = \operatorname{diag}\left(\left[\mathbf{0}_a^T, \mathbf{I}_b^T\right]^T\right)$, where $a + b = m + 1$. This problem addresses scenarios where only some measurements are corrupted with uncorrelated noise. It is apparent that the extreme cases of LS-TLS are LS and TLS. The LS-TLS structure implies that the problem can be divided into two subproblems. Usually, the solution is obtained sequentially. First, the a-dimensional LS subproblem is solved, and then a correction is obtained from the solution of the b-dimensional TLS subproblem. For more information about the LS-TLS problem, refer [127].

The EW-TLS problem still restricts the noise to be uncorrelated, but the variances can vary. The noise can be expressed as

$$\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \operatorname{diag}(\sigma_{\text{EW-TLS}})), \tag{2.33}$$

where $\sigma_{\text{EW-TLS}} \in \mathbb{R}^{m+1}$ is a vector whose elements are positive values. This problem is fit for varying levels of precision across data points, making it particularly useful when some measurements are more reliable than others. The classical approach to addressing the EW-TLS problem involves rescaling the data to equalize variances. By doing this, the problem reduces the TLS subproblem. The solution to the EW-TLS problem is generally obtained by first solving the TLS subproblem using established TLS techniques, followed by rescaling the TLS subproblem solution to obtain solution of the original problem.

It is apparent that TLS is a particular case of the EW-TLS. Conversely, by definition, the EW-TLS problem can not be simplified to the LS-TLS problem.

The most comprehensive generalization of the TLS presumes an unconstrained structure for the noise covariance matrix

$$\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \Sigma_{\text{GTLS}}), \tag{2.34}$$

except for obligatory requirements of symmetry and positive semi-definiteness.

The flexibility of the GTLS accommodates scenarios where the noise may exhibit varying degrees of variances and correlation, making GTLS particularly useful for identifying systems with complex error structures, such as in PMLSM. It is apparent that all the mentioned problems (LS, TLS, LS-TLS, EW-TLS) are special cases of GTLS. While restrictions imposed by the specific scenarios allow for simpler algorithms, the solution can always be obtained by solving the GTLS formulation of the problem.

A comprehensive analysis of GTLS algorithms is beyond the scope of this thesis. Therefore, this section focuses exclusively on the properties of well-behaved GTLS problems. Similar to other problems discussed therein, several algorithms for solving GTLS exist, many of which address special considerations for zero variance measurements, efficient computation, numerical stability, and under-determined systems, as discussed in [97]. For the examples provided, we utilize an implementation based on the algorithm presented in [97], sourced from `github.com/iwoodsawyer/gtls`.

The main idea is similar to the TLS. Let us recall that in the case of TLS, the solution is calculated from the singular vector corresponding to the smallest singular value of the dataset matrix D. Therefore, the singular vector can be obtained by minimizing the Rayleigh quotient (2.29).

The findings from [97] can be summed up in the following theorem:

Theorem 3. *Solution of the generalized total least squares problem.*
Let

$$T^T D Z = D_D \quad \textit{and} \quad W^T \Sigma_{GTLS} Z = D_\Sigma \tag{2.35}$$

with

$$D_D = \mathrm{diag}(\alpha_1,\, \alpha_2,\, \ldots,\, \alpha_n) \tag{2.36}$$

$$D_\Sigma = \mathrm{diag}(\beta_1,\, \beta_2,\, \ldots,\, \beta_q) \tag{2.37}$$

be a generalized singular value decomposition of the matrix pair (D, Σ_{GTLS})*; and* $\{\alpha_i/\beta_i\}_{i=0}^{m+1}$ *be the set of generalized singular values of the matrix pair* (D, Σ_{GTLS})*. After defining the partitioning*

$$Z = \begin{array}{cc} m & 1 \\ \end{array} \\ Z = \begin{bmatrix} Z_{11} & z_{12} \\ z_{12}^T & z_{22} \end{bmatrix}$$

a GTLS solution exists if and only if z_{22} *is non-zero. In addition, it is unique if and only if* $\alpha_d/\beta_d \neq \alpha_{m+1}/\beta_{m+1}$*. In the case when the GTLS solution exists and is unique, it is given by*

$$\hat{\theta}_{GTLS} = \frac{-z_{12}}{z_{22}}.$$

The similarity to the TLS solution is noticeable. The GTLS solution is calculated from the right singular vector corresponding to the smallest generalized eigenvalue associated with the matrix pair $(D, \Sigma_{\mathrm{GTLS}})$. The singular vector can be obtained by minimizing the generalized Rayleigh quotient:

$$\begin{bmatrix} z_{12} \\ z_{22} \end{bmatrix} := \arg\min_z \frac{z^T \left(D^T D \right) z}{z^T \Sigma_{\mathrm{GTLS}} z}. \tag{2.38}$$

This fact allows for the enhancement of the recursive, inverse iteration-based TLS algorithm 3 such that GTLS problems can be solved recursively. The resulting algorithm introduced in [120] is:

Algorithm 4 Recursive Generalized Total Least Squares (RGTLS)

Input: D, Σ_{GTLS}, $\hat{\theta}_{GTLS,0}$, P_0, λ

Output: $\hat{\theta}_{\text{GTLS}}$

$\hat{\theta}_{\text{GTLS}} \leftarrow \hat{\theta}_{GTLS,0}$

$P \leftarrow P_0$

for $i \leftarrow 1$ to N **do**

  $L \leftarrow (Pd_i)/(\lambda + d_i^T P d_i)$

  $P \leftarrow (\mathbf{I} - Ld_i^T)P\frac{1}{\lambda}$

  $Z \leftarrow P\Sigma_{\text{GTLS}}[\hat{\theta}_{\text{GTLS}}^T, -1]^T$

  $\hat{\theta}_{\text{GTLS}} \leftarrow -z_{12}/z_{22}$

end for

In Algorithm 4, the d_i is value from dataset D as defined in (2.27), $\lambda \in \mathbb{R}$ is forgetting factor, $\hat{\theta}_{GTLS,0}$ is initial value for $\hat{\theta}_{\text{GTLS}}$ and $P_0 \in \mathbb{R}^{m+1 \times m+1}$ is initial value of precision matrix, usually set to $c\mathbf{I}_{m+1}$, where c is a large number.

The alternative approach supporting only regular noise covariance matrix proposed by Leon Jay Gleser [128, 129] is summed up in the following theorem:

Theorem 4. *Let the dataset D be generated from an Errors-In-Variables process with positive definite noise covariance matrix $\Sigma_{GTLS} \succ 0$. Assuming the upper triangular factorization $E^{-T}E^{-1} = \Sigma_{GTLS}$, the GTLS solution $\hat{\theta}_{GTLS}$ is obtained as*

$$\hat{\theta}_{GTLS} = -\frac{E_{11}v'_{12} + e_{12}v'_{22}}{e_{22}v'_{22}}, \tag{2.39}$$

where v' is the singular vector corresponding to the smallest singular value of the transformed dataset $D' = DE$, with the following partitioning:

$$v' = \begin{bmatrix} v'_{12} \\ v'_{22} \end{bmatrix}, \qquad E = \begin{bmatrix} E_{11} & e_{12} \\ \mathbf{0} & e_{22} \end{bmatrix}, \tag{2.40}$$

where v'_{22} is the last element of v'.

Proof. The singular vector v' of the transformed dataset D' is obtained by minimizing the Rayleigh Quotient (2.29). By substitutions of $D' = DE$ and $v' = E^{-1}v$, we obtain following optimization problem

$$v' := \arg\min_v \frac{\left(v^T E^{-T}\right)\left(E^T D^T\right)(DE)\left(E^{-1}v\right)}{\left(v^T E^{-T}\right)\left(E^{-1}v\right)}. \tag{2.41}$$

This is equivalent to the optimization problem for the generalized total least squares (2.38).

By recovering the singular vector v corresponding to the smallest singular value of the original dataset

$$v = Ev' = \begin{bmatrix} E_{11}{v_{12}}' + E_{12}{v_{22}}' \\ E_{22}{v_{22}}' \end{bmatrix} \tag{2.42}$$

and plugging it into

$$\hat{\theta}_{\mathrm{GTLS}} = -v_{12}/v_{22}, \tag{2.43}$$

we obtain (2.39) which concludes the proof. □

Gleser's result allows for solving GTLS problems using TLS algorithms by transforming the dataset with a precomputed E. This is because the transformed dataset has a covariance matrix $\Sigma'_{\mathrm{GTLS}} \approx \mathbf{I}$, and the problem is thus reduced to a TLS problem. The TLS solution can then be transformed back to the original problem's solution. GTLS problems can, therefore, be solved using TLS algorithms, assuming that the inputs and outputs are transformed correspondingly. However, the problem with this method arises when the E^{-1} is ill-conditioned, which can lead to numerical issues.

It should be reiterated that the transformation is only possible for a positive definite covariance matrix, which results in some loss of generality. For a positive semi-definite covariance matrix, only an approximate solution can be obtained by replacing the covariance matrix with its nearest symmetric positive definite matrix [130]. However, E^{-1} of the approximate covariance matrix could be ill-conditioned.

3 Problem definition

In previous chapters, the modeling and identification of PMLSM were introduced. The identification process is a crucial step, as many other areas, such as control and diagnostics, directly depend on the quality of the identified parameters. However, current identification methods have several disadvantages.

Focusing on online identification, which is essential for PMLSM due to their time-variant parameters, most available methods provide only point estimates. While prior estimates of parameters are often available from datasheets or offline identification, the lack of uncertainty quantification for these estimates limits the potential to combine information into a single, reliable estimate. Moreover, it prevents the enhancement of online identification results using these prior estimates.

It should be noted that Bayesian methods which provide statistical results do exist for least squares identification. However, as previously discussed, least squares identification is often unsuitable because it typically relies on the least squares for approach to ARX model identification. As shown in Section 2.6, using least squares can violate the fundamental mathematical assumptions needed to ensure the validity and reliability of the solution when measurement errors are significant.

An alternative approach to identification is to formulate the problem in terms of errors-in-variables (EIV). This method assumes no error in the model itself, but parameter uncertainty can express discrepancies between the model and reality. However, existing EIV methods do not provide uncertainty quantification.

Addressing this issue requires the development of an EIV identification framework and algorithm that can handle the specific conditions of PMLSM, such as small sampling periods leading to large datasets and correlated noise across all variables. This is the first objective of this thesis. The second objective is to reformulate the PMLSM model as an EIV problem and apply the developed algorithm to successfully identify PMLSM parameters.

Given its advantageous properties, the analysis and derivation will be conducted from a Bayesian perspective. The intrinsic probabilistic interpretation of inputs, outputs, and parameters allows for the seamless incorporation of measurement errors while ensuring accurate determination of parameter uncertainty. Bayesian prior information permits the integration of prior knowledge while providing a systematic method to refine estimates as more data becomes available. This approach facilitates the development of recursive identification algorithms. Utilizing the Bayesian framework also allows for further extensions, such as incorporating findings into cost-benefit analyses in decision-making, leading to optimal decisions under uncertainty. Models can also be compared based on their posterior probabilities, offering a straightforward way to assess the relative plausibility of different models.

It is evident that the proposed EIV identification solution addresses many of the issues mentioned earlier.

Before proceeding, let us clearly define the two main objectives of this thesis.

3.1 Bayesian Errors-in-Variables

In order to derive a Bayesian understanding of Errors-in-Variables (EIV), let us first define EIV.

Let once again begin by assuming a linear model

$$\overline{y}_i = \overline{x}_i^T \theta, \tag{3.1a}$$

where $\overline{x}_i \in \mathbb{R}^m$ is i-th sample vector of linearly independent variables; $\overline{y}_i \in \mathbb{R}$ is i-th sample of linearly dependent variable and θ is vector of unknown variables. For simplicity, let us assume that the vector is time-invariant [1].

Linear errors-in-variables arise when noisy samples of linearly independent and linearly dependent samples $x_i \in \mathbb{R}^m$ and $y_i \in \mathbb{R}$, respectively, are obtained from the linear EIV model (3.1a). We can further assume that the noise ε is additive, and its properties are summed up by $\mathfrak{S}$.

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} \overline{x}_i \\ \overline{y}_i \end{bmatrix} + \varepsilon_i \tag{3.1b}$$

The error-in-variables identification aims to find sought-after parameters θ from noisy data $[x_i^T, y_i]^T$.

Although this definition is clear, one may encounter literature that utilizes the following definition:

$$y = \overline{y}_i + \varepsilon_{y,i}, \tag{3.2}$$

$$\overline{x} = x_i + \varepsilon_{x,i}. \tag{3.3}$$

By this definition, the known input signal is noiseless, while the additive noise acts on the system's input. This type of model is referred to as the Berkson model, named after Joseph Berkson, who first discussed the distinctions between the two types of error models [131]. The confusion between the models may stem from Bayesian analyses, such as [132], which demonstrated that the problems do not need to be treated as distinct under certain conditions. The difference between the models can be studied in the following figure from [133].

[1] As this assumption is unsuitable for PMLSM identification, it will be relaxed in section 4.6 by introducing forgetting factor allowing for identification of slowly varying parameters.

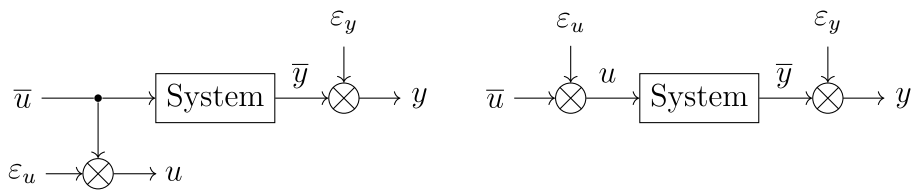


Fig. 3.1: Classical EIV (left), Berkson Model (right).

Our study focused exclusively on the classical Errors-in-Variables (EIV) model, as it is more widely recognized and extensively validated in the literature. The classical EIV model provides a comprehensive framework that ensures our findings' reliability and generalizability.

Using sequence notation introduced in section 2.2, we can define following data sequences:

$$D := \{x_i,\, y_i\}_{i=0}^{N}, \tag{3.4}$$
$$\overline{D} := \{\overline{x}_i,\, \overline{y}_i\}_{i=0}^{N}, \tag{3.5}$$
$$\overline{X} := \{\overline{x}_i\}_{i=0}^{N}, \tag{3.6}$$
$$\overline{Y} := \{\overline{y}_i\}_{i=0}^{N}. \tag{3.7}$$

This notation allows for defining probability density of unknown parameters θ in Bayesian sense, correspondingly with available literature [134, 135, 132] as

$$p(\theta, \overline{X}|D, \mathfrak{S}) \propto \mathcal{L}(\theta, \overline{X}|D, \mathfrak{S})p(\theta)p(\overline{X}). \tag{3.8}$$

In this equation, the $p(\theta, \overline{X}|D)$ is joint posterior distribution for θ and $\overline{X}$, $\mathcal{L}(\theta, \overline{X}|D)$ is corresponding joint likelihood and $p(\theta)$ and $p(\overline{X})$ are prior distributions for θ and $\overline{X}$ respectively.

The joint distributions can be separated into conditionals as

$$p(\theta|D, \overline{X}, \mathfrak{S})p(\overline{X}|D, \mathfrak{S}) \propto \mathcal{L}(\theta|D, \overline{X}, \mathfrak{S})p(\theta)\,\mathcal{L}(\overline{X}|D, \mathfrak{S})p(\overline{X}). \tag{3.9}$$

The target of Bayesian errors-in-variables is to infer a posterior distribution of the unknown parameters θ. However, note that this posterior distribution suffers from a severe disadvantage. That is, the dimension of $p(\overline{X}|D, \mathfrak{S})$ is N and therefore grows in time with the rising number of drawn samples. This is unsuitable for recursive identification, as the computational complexity would grow unboundedly.

While there are applications that would benefit from inference about $\overline{X}$, we are only interested in conditional $p(\theta|D, \overline{X}, \mathfrak{S})$.

The conditional is, however, unobtainable without knowledge of the nuisance pdf $p(\overline{X}|D, \mathfrak{S})$. A widely used remedy for this issue is to replace the nuisance pdf with the expectation for each sample. Using defined sequence notation, we can write

$$p(\overline{X}|D, \mathfrak{S}) \approx \mathbb{E}\left[\overline{X}|D, \mathfrak{S}\right]. \tag{3.10}$$

This method has been popular for a few decades already. For example, this method was disclosed in 1992 as the "most widely used approach" [136]. Refer to [137, 138, 139] for both early and recent examples. Although this approach may seem dated, there are no feasible alternatives, as most authors develop only batch identification algorithms.

Resulting conditional posterior distribution is

$$p\left(\theta|D, \mathfrak{S}, \mathbb{E}\left[\overline{X}|D, \mathfrak{S}\right]\right) \propto \mathcal{L}\left(\theta|D, \mathfrak{S}, \mathbb{E}\left[\overline{X}|D, \mathfrak{S}\right]\right) p(\theta). \tag{3.11}$$

In the literature, distinctions are seldom made between (3.9) and 3.11. To make the difference clear in this text, we will call (3.9) the Bayesian Errors-in-variables (B-EIV) and 3.11 the Bayesian Generalized Total Least Squares (B-GTLS). The connection between the Generalized Total Least Squares will be apparent in the following chapter.

The first goal of this thesis is to propose the Bayesian Generalized Total Least Squares framework, which would allow for solving errors-in-variables problems. As the Generalized Total Least Squares can oftentimes be reduced into the Total Least Squares problem, it is also be beneficial to analyze the special case of errors-in-variables, when noise on all variables is uncorrelated and identical in magnitude. This is the second goal of this thesis, leading to derivation of the Bayesian Total Least Squares.

3.2 Errors-in-Variables Approach to PMLSM Identification

PMLSM model can be easily transformed into linear model (3.1) by rearranging the first two state equations of the PMLSM (1.8), resulting in a set of linear equations

$$u_d = L_d \dot{i_d} - L_q \frac{\pi}{\tau_s} i_q v + R i_d, \tag{3.12}$$

$$u_q = L_d \frac{\pi}{\tau_s} i_d v + L_q \dot{i_q} + R i_q + \Psi_{PM} \frac{\pi}{\tau_s} v. \tag{3.13}$$

The terms of which can be separated into three groups. The first group comprises unknown parameters L_d, L_q, R, and Ψ_{PM}. The second group is known constants π, τ_s. The third group is group of measured terms $\dot{i_d}$, $\dot{i_q}$, i_d, i_q, u_d, u_q and v, known only up to some error. Voltages u_d and u_q can be included in the last group, although they are usually not measured values but rather known voltage input values.

One could argue that there are numerous sources of disturbances caused by voltage generation instrumentation, which could be explained by incorporating uncertainty in the voltages. The input voltages could differ from the true voltages due to, for example, inverter switching noise, inverter controller imperfections, or electromagnetic interference induced in power lines. The evaluation of the magnitude of uncertainty will be left to the statistician. No uncertainty can be simulated by setting arbitrarily low variance.

Defined grouping allows to express the linear equations as

$$\overline{y}_1 = \theta^T \overline{x}_1, \tag{3.14}$$

$$\overline{y}_2 = \theta^T \overline{x}_2, \tag{3.15}$$

where the vector of unknown parameters θ, the two vectors of independent variables $\overline{x}_1, \overline{x}_2$ and the two dependent variables $\overline{y}_1, \overline{y}_2$ are

$$\theta = \begin{bmatrix} L_d \\ L_q \\ Ri_d \\ \Psi_{PM} \end{bmatrix}, \quad \overline{x}_1 = \begin{bmatrix} \dot{i}_d \\ -\frac{\pi}{\tau_s} i_q v \\ i_d \\ 0 \end{bmatrix}, \quad \overline{x}_2 = \begin{bmatrix} \frac{\pi}{\tau_s} i_d v \\ \dot{i}_q \\ i_q \\ \frac{\pi}{\tau_s} v \end{bmatrix}, \quad \overline{y}_1 = u_d, \quad \overline{y}_2 = u_q, \tag{3.16}$$

respectively.

Typically only measured values of voltages u_d, u_q, currents i_d, i_q, and position l are available. The derivatives in (3.16) can be approximated using Euler discretization as follows:

$$\dot{i}_d \approx \frac{i_d(k) - i_d(k-1)}{T_s}, \tag{3.17}$$

$$\dot{i}_q \approx \frac{i_q(k) - i_q(k-1)}{T_s}, \tag{3.18}$$

$$v \approx \frac{l(k) - l(k-1)}{T_s}, \tag{3.19}$$

which inevitably leads to modeling errors. The resulting vectors of synthesized measurements are

$$\overline{x}_1 = \begin{bmatrix} \frac{i_d(k)-i_d(k-1)}{T_s} \\ -\frac{\pi}{\tau_s} i_q \frac{l(k)-l(k-1)}{T_s} \\ i_d \\ 0 \end{bmatrix}, \quad \overline{x}_2 = \begin{bmatrix} \frac{\pi}{\tau_s} i_d \frac{l(k)-l(k-1)}{T_s} \\ \frac{i_q(k)-i_q(k-1)}{T_s} \\ i_q \\ \frac{\pi}{\tau_s} \frac{l(k)-l(k-1)}{T_s} \end{bmatrix}. \tag{3.20}$$

As the information regarding y_1, y_2, x_1, and x_2 is available only subject to measurement and modeling error, the PMLSM identification is now formulated in terms of error-in-variables problem.

Based on the B-GTLS analysis provided in Chapter 4, the posterior distribution for PMLSM is

$$p\left(\theta|D_1, D_2, \mathfrak{S}_1, \mathfrak{S}_2, \mathbb{E}\left[\overline{X_1}|D_1, \mathfrak{S}_1\right], \mathbb{E}\left[\overline{X_2}|D_2, \mathfrak{S}_2\right]\right) \propto \tag{3.21}$$

$$\mathcal{L}\left(\theta|D_1, \mathfrak{S}_1, \mathbb{E}\left[\overline{X_1}|D_1, \mathfrak{S}_1\right]\right) \mathcal{L}\left(\theta|D_2, \mathfrak{S}_2, \mathbb{E}\left[\overline{X_2}|D_2, \mathfrak{S}_2\right]\right) p(\theta) \tag{3.22}$$

The posterior distribution is an unusual function with several challenging features. These features include significant non-linearity, possible multi-modality, and unknown moments. The process of deriving moments is particularly complicated because of denominator terms in the exponent. Due to the multi-modality of the problem, even a maximum a posteriori solution requires advanced and, therefore, resource-demanding computation.

We propose approximating individual likelihoods by the normal distribution in order to find a normal approximation of the posterior distribution.

$$p\left(\theta|D_1, D_2, \mathfrak{S}_1, \mathfrak{S}_2, \mathbb{E}\left[\overline{X_1}|D_1, \mathfrak{S}_1\right], \mathbb{E}\left[\overline{X_2}|D_2, \mathfrak{S}_2\right]\right) \propto \tag{3.23}$$

$$q\left(\theta|D_1, \mathfrak{S}_1, \mathbb{E}\left[\overline{X_1}|D_1, \mathfrak{S}_1\right]\right) q\left(\theta|D_2, \mathfrak{S}_2, \mathbb{E}\left[\overline{X_2}|D_2, \mathfrak{S}_2\right]\right) p(\theta) \tag{3.24}$$

where

$$q\left(\theta|D_j, \mathfrak{S}_j\, \mathbb{E}\left[\overline{X_j}|D_j, \mathfrak{S}_j\right]\right) \approx \mathcal{L}\left(\theta|D_j, \mathfrak{S}_j, \mathbb{E}\left[\overline{X_j}|D_j, \mathfrak{S}_j\right]\right), \quad j \in \{1, 2\}. \tag{3.25}$$

The existing GTLS method could provide an individual point estimate of the likelihood. There is, however, no way to combine those estimates into a single vector of parameters.

The third goal of this thesis is to utilize Bayesian Generalized Total Least Squares in order to obtain an approximation of the PMLSM likelihood functions in real time. For this, a recursive algorithm will be required. As the results provide uncertain information, surrogate likelihoods and prior can be merged into a single surrogate posterior distribution.

4 Bayesian Generalized Total Least Squares

The goal of this chapter is to present a derivation of the functional form of the Bayesian Generalized Total Least Squares (B-GTLS) problem

$$p\left(\theta|D,\mathfrak{S},\mathbb{E}\left[\overline{X}|D,\mathfrak{S}\right]\right) \propto \mathcal{L}\left(\theta|D,\mathfrak{S},\mathbb{E}\left[\overline{X}|D,\mathfrak{S}\right]\right)p(\theta). \tag{4.1}$$

defined in (3.11), which is an approximation of the conditional $p(\theta|D,\mathfrak{S},\overline{X})$ of the Bayesian Errors-in-variables (B-EIV), defined in (3.9) and discussed in the previous chapter.

In order to derive the functional form, noise assumptions need to be stated. Consistently with the aforementioned definition of the GTLS, we can assume the noise ε to be independent, normally distributed with zero mean and covariance matrix Σ

$$\overline{y}_i = \overline{x}_i^T\theta, \tag{4.2a}$$

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} \overline{x}_i \\ \overline{y}_i \end{bmatrix} + \varepsilon_i, \quad \varepsilon \sim \mathcal{N}(\mathbf{0},\Sigma). \tag{4.2b}$$

The noise is described by $\mathfrak{S}$, which in this case denotes knowledge of its distribution, mean, and covariance matrix. As this knowledge is required for all likelihoods and posterior distributions, it would appear in all following equations, cluttering the notation. For improved readability, we decided to omit writing of $\mathfrak{S}$ from further equations. Note, however, that the requirement for knowledge of the noise properties did not vanish.

Due to the complex nature of deriving and formulating the functional expression of the posterior, it is necessary to separate the process into several sections. In the following sections, we will progressively introduce all key elements of the posterior distribution: the hidden variables' expectation $\mathbb{E}\left[\overline{X}|D\right]$, the likelihood function of the dataset $\mathcal{L}\left(\theta|D,\mathbb{E}\left[\overline{X}|D\right]\right)$, and the prior $p(\theta)$. This will result in the posterior $p\left(\theta|D,\mathbb{E}\left[\overline{X}|D\right]\right)$.

Furthermore, we propose using the Laplace method for normal approximation, making the functional form of the resulting surrogate posterior distribution convenient for further implementation in other control and diagnostics methods. As the final step in solving the problem defined in section 3.1, we propose the recursive Bayesian generalized total least squares (B-RGTLS) algorithm. In the last Section of this chapter, we recast PMLSM identification as an errors-in-variables problem and present a new algorithm based on the proposed B-RGTLS, which addresses the problem defined in Section 3.2.

4.1 The Expectation of the Hidden Variables

The probability distribution of the hidden variables $p(\overline{X}|D)$ causes the Bayesian Errors-In-Variables posterior distribution to grow in dimension with each incorporated sample. In order to avoid it and derive the B-GTLS, it needs to be approximated with its expectation

$$q(\overline{X}|D) \approx \mathbb{E}\left[\overline{X}|D\right]. \tag{4.3}$$

With noise assumption from (4.2), the expectation is obtained from the following lemma:

Lemma 1. *Assume the following partition of the covariance matrix*

$$\Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_{yy} \end{bmatrix}. \tag{4.4}$$

Let us further assume that hidden variables' priors are uninformative $p(\overline{X}) \propto 1$. Then, the hidden variables sample expectation is

$$\mathbb{E}\left[\begin{array}{c} \overline{x} \\ \overline{y} \end{array}\middle| d\right] = \begin{bmatrix} x - \dfrac{\Sigma_{xx}\theta x^T\theta - \Sigma_{xx}\theta y - \Sigma_{xy}x^T\theta + \Sigma_{xy}y}{\theta^T\Sigma_{xx}\theta - 2\Sigma_{xy}^T\theta + \Sigma_{yy}} \\ \dfrac{\theta^T\Sigma_{xx}\theta y - \Sigma_{xy}^T\left(\theta y + \theta x^T\theta\right) + \Sigma_{yy}x^T\theta}{\theta^T\Sigma_{xx}\theta - 2\Sigma_{xy}^T\theta + \Sigma_{yy}} \end{bmatrix}, \tag{4.5}$$

where the subscript $_i$ is omitted for improved readability.

Proof. Under the assumption of uninformative normal prior, the probability distribution for the hidden variables associated with the ith sample follows a normal distribution:

$$\begin{aligned} p(\overline{d}|d,\theta) =& (2\pi)^{-\frac{m+1}{2}} \det(\Sigma)^{-\frac{1}{2}} \\ & \exp\left(-\frac{1}{2}\left(\overline{d}-d\right)^T \Sigma^{-1}\left(\overline{d}-d\right)\right). \end{aligned} \tag{4.6}$$

For improved readability, the subscript $_i$ is omitted throughout the whole proof. As for the normal distribution, the modus and the expected value coincide; the following constrained optimization problem defines the expectation

$$\begin{aligned} \mathbb{E}[\overline{d}|d] := \underset{\overline{d}}{\operatorname{argmax}} \quad & -\frac{1}{2}\left(\overline{d}-d\right)^T \Sigma^{-1}\left(\overline{d}-d\right) \\ \text{s.t.} \quad & \overline{d}^T\vartheta = 0. \end{aligned} \tag{4.7}$$

It can be restated using Lagrange multiplier λ as an unconstrained optimization problem

$$\mathbb{E}[\overline{d}|d] := \underset{\overline{d}}{\operatorname{argmax}} \quad F(\overline{d},\lambda) \tag{4.8}$$

$$F(\overline{d},\lambda) = -\frac{1}{2}\left(d-\overline{d}\right)^T \Sigma^{-1}\left(d-\overline{d}\right) - \lambda\overline{d}^T\vartheta, \tag{4.9}$$

which can be resolved by adhering to standard steps of setting the derivations to zero

$$\frac{\partial}{\partial \overline{d}} F(\overline{d}, \lambda) = \Sigma^{-1}(d - \overline{d}) - \lambda\vartheta = 0, \tag{4.10}$$

$$\frac{\partial}{\partial \lambda} F(\overline{d}, \lambda) = -\overline{d}^T \vartheta = 0, \tag{4.11}$$

isolating the $\overline{d}$ from (4.10)

$$\overline{d} = \Sigma(\Sigma^{-1} d - \lambda\vartheta) \tag{4.12}$$

and finding the λ by isolating it from (4.11), while plugging in the $\overline{d}$ from (4.12)

$$\lambda = \frac{d^T \vartheta}{\vartheta^T \Sigma \vartheta}. \tag{4.13}$$

Plugging the λ back into (4.12) yields

$$\mathbb{E}[\overline{d}|d] = d - \Sigma \frac{d^T \vartheta}{\vartheta^T \Sigma \vartheta} \vartheta. \tag{4.14}$$

After rearrangement and substituting the original variables, (4.14) equals (4.5). $\square$

4.2 The Likelihood Function of the Dataset

The functional form of the dataset D likelihood function arises as an approximation of the B-EIV likelihood

$$\mathcal{L}\left(\theta|D, \mathbb{E}\left[\overline{X}|D\right]\right) \approx \mathcal{L}(\theta|D, \overline{X})\, \mathcal{L}(\overline{X}|D) p(\overline{X}). \tag{4.15}$$

Assuming, that the hidden variables prior is uninformative normal $p(\overline{X}) \propto 1$, as stated in lemma 1, the approximative likelihood expectation of a sample is expressed as

$$\mathbb{E}\left[\overline{x}_i|d_i\right] = x - \frac{\Sigma_{xx}\theta x^T\theta - \Sigma_{xx}\theta y - \Sigma_{xy} x^T \theta + \Sigma_{xy} y}{\theta^T \Sigma_{xx} \theta - 2\Sigma_{xy}^T \theta + \Sigma_{yy}}. \tag{4.16}$$

The likelihood is then defined by the following lemma:

Lemma 2. *Let $x_i \sim \mathcal{N}(\mathbb{E}[\overline{x}_i|d_i], \Sigma), \forall i,$ where $\mathbb{E}[\overline{x}_i|d_i]$ is defined in lemma 1 as* (4.16)*. Then, the Likelihood function takes the following functional form:*

$$\mathcal{L}(\theta|D) = (2\pi)^{-\frac{m+1}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \frac{\vartheta^T \left(D^T D\right) \vartheta}{\vartheta^T \Sigma \vartheta}\right). \tag{4.17}$$

Proof. Let the proof begin by establishing the approximate likelihood function as the product of normal distributed samples d_i, correspondingly with definitions in the state-of-the-art Bayesian EIV literature [139].

$$\mathcal{L}\left(\theta|D, \mathbb{E}\left[\overline{X}|D\right]\right) = \prod_{i=1}^{N} \mathcal{N}\left(\mathbb{E}\begin{bmatrix}\mathbb{E}[\overline{x}_i|d_i] \\ \mathbb{E}[\overline{x}_i|d_i]^T\theta\end{bmatrix}, \Sigma\right) \tag{4.18}$$

$$= \prod_{i=1}^{N}(2\pi)^{-\frac{m+1}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp \left(-\frac{1}{2}\left(d_i - \begin{bmatrix}\mathbb{E}[\overline{x}_i|d_i] \\ \mathbb{E}[\overline{x}_i|d_i]^T\theta\end{bmatrix}\right)^T \Sigma^{-1}\left(d_i - \begin{bmatrix}\mathbb{E}[\overline{x}_i|d_i] \\ \mathbb{E}[\overline{x}_i|d_i]^T\theta\end{bmatrix}\right)\right). \tag{4.19}$$

As

$$\mathbb{E}[\overline{x}_i|d_i]^T\theta = \left(x - \frac{\Sigma_{xx}\theta x^T\theta - \Sigma_{xx}\theta y - \Sigma_{xy}x^T\theta + \Sigma_{xy}y}{\theta^T\Sigma_{xx}\theta - 2\Sigma_{xy}^T\theta + \Sigma_{yy}}\right)\theta \tag{4.20}$$

$$= \frac{\theta^T\Sigma_{xx}\theta y - \Sigma_{xy}^T\left(\theta y + \theta x^T\theta\right) + \Sigma_{yy}x^T\theta}{\theta^T\Sigma_{xx}\theta - 2\Sigma_{xy}^T\theta + \Sigma_{yy}} \tag{4.21}$$

$$= \mathbb{E}[\overline{y}_i|d_i], \tag{4.22}$$

we can write

$$\mathcal{L}\left(\theta|D, \mathbb{E}\left[\overline{X}|D\right]\right) \propto \exp\left(-\frac{1}{2}\sum_{i=1}^{N}\left(d_i - \mathbb{E}[\overline{d}_i|d_i]\right)^T \Sigma^{-1}\left(d_i - \mathbb{E}[\overline{d}_i|d_i]\right)\right). \tag{4.23}$$

Plugging in (4.14), simplifying and rearranging, we obtain

$$\mathcal{L}\left(\theta|D, \mathbb{E}\left[\overline{X}|D\right]\right) \propto \exp\left(-\frac{1}{2}\frac{\vartheta^T\left(\sum_{i=1}^{N} d_i d_i^T\right)\vartheta}{\vartheta^T\Sigma\vartheta}\right). \tag{4.24}$$

As the likelihood does not contain any variables other than ϑ and d_i, we can, without loss of generality, write

$$\mathcal{L}\left(\theta|D, \mathbb{E}\left[\overline{X}|D\right]\right) = \mathcal{L}\left(\theta|D\right). \tag{4.25}$$

The acquisition of equation (4.17) necessitates the use of the identity

$$D^T D = \sum_{i=1}^{N} d_i d_i^T, \tag{4.26}$$

which concludes the proof. $\square$

4.3 Establishing Prior Distribution

The prior distribution serves a crucial role in the statistical framework, enabling statisticians to encapsulate pre-existing knowledge about the parameters in question or to articulate uncertainty in the absence of such knowledge. Due to our specific focus on recursive identification, the discourse is confined to the examination of conjugate priors, given their compatibility with Bayesian updating mechanisms.

A natural choice of conjugate prior is

Lemma 3. *Posterior distribution belongs in the same family of distributions when likelihood* (4.17) *is multiplied with*

$$p(\theta) \propto \exp\left(\frac{\vartheta^T \Phi_p \vartheta}{\vartheta^T \Sigma \vartheta}\right), \tag{4.27}$$

where $\Phi_p \preceq 0$ is a symmetric, negative semi-definite matrix representing the statistician's prior knowledge. The covariance Σ is determined by the identified process.

Proof. As the prior is proportional to the likelihood

$$\mathcal{L}(\theta|D) \propto p(\theta), \tag{4.28}$$

the posterior is conjugate. □

The fact that the process predetermines the covariance Σ makes stating the prior belief challenging. This problem will be addressed further in the text by generalizing the problem. However, stating the prior in the shape defined in (4.27) is advantageous. For this case, we propose the following method for constructing Φ_p by sampling the EIV model.

The first step of the procedure is to define the EIV process, assuming prior information is available to the statistician. Next, similarly to (4.26), we can assume

$$\Phi_p = -\frac{1}{2}\sum_{k=1}^{N_p} \phi_i \phi_i^T. \tag{4.29}$$

Let ϕ_i represent the k-th sample from the simulated process. The configuration of ϕ_i mirrors the configuration of d_i. A smaller number of drawn samples N_p suggests a vague prior, whereas a larger number signifies a more informative prior.

Measurements obtained from previous experiments can be utilized in a similar fashion.

When a statistician aims to represent a lack of prior knowledge with an uninformative prior, a uniform distribution over the parameter space $p(\theta) \propto 1$ can be used:

$$\Phi_p = \mathbf{0}. \tag{4.30}$$

4.4 Functional Form of the Posterior Distribution

The shape of the posterior distribution (4.1) is a corollary of lemmas 1, 2 3:

Corollary 1. *Let likelihood and prior be defined by* (4.17) *and* (4.27) *respectively. Then, the functional form of the posterior distribution is*

$$p(\theta|D) \propto \exp\left(\frac{\vartheta^T \Omega \vartheta}{\vartheta^T \Sigma \vartheta}\right), \tag{4.31a}$$

under the following assumption

$$\Omega = -\tfrac{1}{2} D^T D + \Phi_p \tag{4.31b}$$

The posterior distribution is a complex and unfamiliar function characterized by several challenging attributes, such as high non-linearity and unknown moments. The process of deriving these moments is exceptionally intricate, primarily due to the presence of denominator terms within the exponent.

However, it also has a significant advantage: it is fully specified by $\{\Omega, \Sigma\}$, a set of two $(m+1)\times(m+1)$ matrices, regardless of the number of samples N. Moreover, the latter matrix is time-invariant, and the first matrix allows for recursive updates.

Theorem 5. *Assume time-invariant matrix Σ and data matrix $\Omega(k)$ encompassing N samples be given in time k. Assume execution of further measurement, resulting in new data-set $\mathcal{D}$*

$$\mathcal{D} = \left[d_{(N+1)},\ d_{(N+2)},\ \ldots,\ d_{(N+M)}\right]^T \tag{4.32}$$

constructed from M additional samples obtained up to time $k+1$.

The data matrix $\Omega(k+1)$ can be obtained without extrinsic knowledge of original data-set D, by update formula

$$\Omega(k+1) = \Omega(k) - \frac{1}{2}\mathcal{D}^T\mathcal{D}, \tag{4.33}$$

Proof. Using the posterior distribution encompassing data up to time k as prior, the posterior distribution in time $k+1$ can be expressed as

$$p(\theta|D(k+1)) = p(\theta|D(k), \mathcal{D}) = \mathcal{L}(\theta|\mathcal{D})p(\theta|D(k)) = \mathcal{L}(\theta|\mathcal{D})\,\mathcal{L}(\theta|D(k))p(\theta), \tag{4.34}$$

resulting in

$$\Omega(k+1) = -\frac{1}{2}\sum_{i=1}^{N} d_i d_i^T + \Phi_p - \frac{1}{2}\sum_{i=N+1}^{M} d_i d_i^T \tag{4.35}$$

As a direct consequence of (4.26) and (4.31b), (4.35) can be rearranged as (4.33). □

Another noteworthy result stemming from the resulting posterior shape (4.31a) is the presence of the generalized Rayleigh quotient in the exponent. While the derivation of the mean of the posterior distribution is complicated, the maximum a posteriori (MAP) solution can be obtained by solving following constrained optimization problem

$$\begin{aligned} \begin{bmatrix} \theta_{MAP} \\ -1 \end{bmatrix} := \underset{\vartheta}{\operatorname{argmax}} \quad & \frac{\vartheta^T \Omega \vartheta}{\vartheta^T \Sigma \vartheta} \\ \text{s.t.} \quad & \vartheta_{22} = -1, \end{aligned} \tag{4.36}$$

where ϑ_{22} is the last element of ϑ.

As the solution is scale-invariant

$$\frac{c\vartheta^T \Omega c\vartheta}{c\vartheta^T \Sigma c\vartheta} = \frac{\vartheta^T \Omega \vartheta}{\vartheta^T \Sigma \vartheta}, \quad c \in \mathbb{R} \setminus \{0\}, \tag{4.37}$$

the constraints can be temporarily relaxed. The solution of the unconstrained optimization is obtained using generalized singular value decomposition as the generalized singular vector

$$\begin{bmatrix} \vartheta_{12} \\ \vartheta_{22} \end{bmatrix} := \underset{\vartheta}{\operatorname{argmax}} \quad \frac{\vartheta^T \Omega \vartheta}{\vartheta^T \Sigma \vartheta}, \tag{4.38}$$

corresponding to the smallest generalized singular value of the matrix pair (Ω, Σ), as described in chapter 2.3.

The solution is then obtained by scaling such that the last element of the vector is equal to -1:

$$\hat{\theta}_{GTLS} = \frac{-\vartheta_{12}}{\vartheta_{22}}. \tag{4.39}$$

Upon detailed analysis, it is evident that the maximum a posteriori (MAP) solution aligns with the generalized total least squares (GTLS) solution, as elaborated in Chapter 2.8. This equivalence implies that existing algorithms [97] can be effectively utilized to obtain the MAP solution.

In Chapter 3.1, the principal result of this paper is termed Bayesian Generalized Total Least Squares (B-GTLS), despite the underlying issue being an errors-in-variables (EIV) problem. This nomenclature is justified by the uncovered relationship between Bayesian Errors-in-Variables (B-EIV) and Generalized Total Least Squares (GTLS). Therefore, The presented EIV problem analysis can be perceived as a stochastic generalization of the generalized total least squares and all its special cases.

4.5 Gaussian Approximation of the Posterior

Although the posterior distribution is now formally defined, the associated uncertainty presents significant challenges in both manipulation and interpretation. To effectively convey the properties of the posterior distribution $p(\theta|D)$ in an understandable manner, we opted to use the Laplace approximation to generate a surrogate posterior $q(\theta|D)$ that follows a normal distribution. This choice was made because deriving a Kullback-Leibler optimal normal approximation proved to be complex, even for a special case of the posterior, as will be shown in chapter 5.5.

Lemma 4. *Let posterior distribution be defined by* (4.31)*. Then, the Laplace approximation is obtained as*

$$p(\theta|D) \approx q(\theta|D) = \mathcal{N}\left(\hat{\theta}, \hat{\Sigma}(\hat{\theta})\right), \tag{4.40a}$$

where $\hat{\theta}$ is the maximum a posteriori point estimate (MAP)

$$\hat{\theta} := \underset{\theta}{\operatorname{argmax}} \ \ln(p(\theta|D_1, D_2)). \tag{4.40b}$$

and

$$\begin{aligned}\hat{\Sigma}^{-1}(\hat{\theta}) = \frac{2\Omega_{11}}{\alpha(\Sigma,\hat{\theta})} &- \frac{4\left(\gamma(\Omega,\Sigma,\hat{\theta})^T - \gamma(\Omega,\Sigma,\hat{\theta})\right)}{\alpha(\Sigma,\hat{\theta})^2} \\ &+ \frac{8\alpha(\Omega,\hat{\theta})\gamma(\Sigma,\Sigma,\hat{\theta})}{\alpha(\Sigma,\hat{\theta})^3} - \frac{2\alpha(\Omega,\hat{\theta})\Sigma_{11}}{\alpha(\Sigma,\hat{\theta})^2}\end{aligned} \tag{4.40c}$$

is hessian of negative log posterior evaluated at $\hat{\theta}$, where

$$\alpha(A, x) = x^T A_{11} x - 2a_{12}^T x + a_{22}, \tag{4.40d}$$

$$\gamma(A, B, x) = A_{11}(xx^T)B_{11} - A_{11}xb_{12}^T - a_{12}x^T B_{11} + a_{12}b_{12}^T, \tag{4.40e}$$

with the following matrix partitioning

$$M = \begin{bmatrix} M_{11} & m_{12} \\ m_{12}^T & m_{22} \end{bmatrix}. \tag{4.40f}$$

Aside from the notable exception of the LS special case, closed-form expressions for MAP estimates are not available, thus requiring the use of numerical methods to solve.

For the general and the special cases, recursive algorithms tailored for solving (4.40b) are readily accessible. The table 4.1 provides an overview of these algorithms.

Special case	Algorithm
LS	RLS [120]
LS-TLS	Recursive RTLS [140]
TLS	RTLS [120]
GTLS	RGLS [120]

Tab. 4.1: Available algorithms for MAP of special cases.

4.6 Recursive Approximate B-GTLS Algorithm

All key elements required for the design of the recursive algorithm are provided in the previous chapters. The posterior distribution and its Laplace approximation are presented, and the connection with the GTLS allows the design of the algorithm based on the existing recursive GTLS algorithm 4.

Incorporation of the prior distribution in Bayesian statistics allows for intrinsic recursive identification by selecting the posterior in step k as prior for step $k+1$.

To avoid approximation error accumulation, the approximate posterior is used only to interpret the result. The original posterior is selected as a prior for the next identification step. This also retains the conjugacy of the inference.

The efficiency of the generalized inverse iteration allows for doing just a single step of iteration for each data sample. The resulting $\hat{\theta}$ of the RGTLS algorithm is then used to obtain Laplace approximation of the covariance matrix using function $\hat{\Sigma}^{-1}(\hat{\theta})$ from (4.40c). The forgetting factor provided by the recursive GTLS algorithm allows for the recursive identification of EIV problems with slowly varying parameters, which enables using this method for PMLSM identification.

This results in the following algorithm

Algorithm 5 Bayesian Recursive Generalized Total Least Squares (B-RGTLS)

Input: D, Σ, $\hat{\theta}_0$, Φ_p^{-1}, λ

Output: $\hat{\theta}$, $\hat{\Sigma}^{-1}$

```
θ̂ ← θ̂_0
P ← Φ_p^{-1}
P^{-1} ← Φ_p
for i ← 1 to N do
    L ← (P d_i)/(λ + d_i^T P d_i)
    P ← (I − L d_i^T) P (1/λ)
    Z ← P Σ [θ̂^T, −1]^T
    θ̂ ← −z_12/z_22
    P^{-1} ← (1/λ) P^{-1} + d_i d_i^T
    Σ̂^{-1} ← Σ̂^{-1}(θ̂)
end for
```

In the Algorithm 5, d_i represents a value from dataset D as specified in the equation (2.27). The parameter $\lambda \in \mathbb{R}$ is forgetting factor, $\hat{\theta}_{GTLS,0}$ is initial value for $\hat{\theta}_{GTLS}$ identical with modus of the prior and $\Phi_p^{-1} \in \mathbb{R}^{m+1\times m+1}$ is the inverse of the prior matrix Φ_p. Calculation of surrogate covariance $\hat{\Sigma}^{-1}$ requires matrix $\Omega = \frac{1}{2}P^{-1}$. As calculating the inverse of P at each step would be inefficient, P^{-1} is obtained recursively.

4.7 Recursive PMLSM Identification Algorithm

In chapter 3.2, the identification problem is presented as two interconnected EIV problems based on the u_d and u_q equations, (3.14) and (3.15) respectively.

After solving each EIV problem separately, Laplace approximations of the likelihoods can be combined with prior into a single posterior approximation.

As described in the chapter 3.2, the extended synthesized measurements $d_{1,i}$ and $d_{2,i}$ based on (3.16) and (3.20) are considered to be normally distributed random variables, constructed from measurements of current i_d, i_q, voltage u_d, u_q and position l. It is safe to assume that the uncertainty of the measurements is independent, identically, and normally distributed.

$$\rho_{\mathrm{PMSM}} = \begin{bmatrix} i_d(k) \\ i_q(k) \\ u_d(k) \\ u_q(k) \\ l(k) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \bar{i}_d(k) \\ \bar{i}_q(k) \\ \overline{u}_d(k) \\ \overline{u}_q(k) \\ \bar{l}(k) \end{bmatrix}, \begin{bmatrix} \sigma_{idid} & \sigma_{idiq} & \sigma_{idud} & \sigma_{iduq} & \sigma_{idl} \\ \sigma_{idiq} & \sigma_{iqiq} & \sigma_{iqud} & \sigma_{iquq} & \sigma_{iql} \\ \sigma_{idud} & \sigma_{iqud} & \sigma_{udud} & \sigma_{uduq} & \sigma_{udl} \\ \sigma_{iduq} & \sigma_{iquq} & \sigma_{uduq} & \sigma_{uquq} & \sigma_{uql} \\ \sigma_{idl} & \sigma_{iql} & \sigma_{udl} & \sigma_{uql} & \sigma_{ll} \end{bmatrix} \right). \tag{4.41}$$

For simplicity of the results, let us assume zero covariance between current, voltage, and position measurements. However, this assumption can be relaxed. In such situation, the same procedure as the one presented holds, but the resulting synthesized noise covariance matrices will be more complicated.

Lets now assemble all random variables into a single vector

$$\rho \sim \mathcal{N}\left(\overline{\rho}, \Sigma_\rho\right), \tag{4.42}$$

where

$$\rho = \begin{bmatrix} i_d(k) \\ i_q(k) \\ u_d(k) \\ u_q(k) \\ l(k) \\ i_d(k-1) \\ i_q(k-1) \\ l(k-1) \end{bmatrix}, \quad \overline{\rho} = \begin{bmatrix} \overline{i}_d(k) \\ \overline{i}_q(k) \\ \overline{u}_d(k) \\ \overline{u}_q(k) \\ \overline{l}(k) \\ \overline{i}_d(k-1) \\ \overline{i}_q(k-1) \\ \overline{l}(k-1) \end{bmatrix}, \tag{4.43}$$

and

$$\Sigma_\rho = \begin{bmatrix} \sigma_{idid} & \sigma_{idiq} & 0 & 0 & 0 & 0 & 0 & 0 \\ \sigma_{idiq} & \sigma_{iqiq} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{udud} & \sigma_{uduq} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{uduq} & \sigma_{uquq} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{ll} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{idid} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{iqiq} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{ll} \end{bmatrix}. \tag{4.44}$$

With this vector of normally distributed random measurements, we can find a normally distributed approximation of the extended measurements.

As the vectors of extended synthesized measurements d_1 and d_2 does not consist solely of linear transformations of the random variables ρ, the probability distributions of d_1 and d_2 are not normal. However, the only non-linear transformation is the product of two normal variables. It is well-known that the distribution of the Gaussian product with a sufficient signal-to-noise ratio is approximately normal. Moreover, the errors-in-variables problem defined in this thesis requires normally distributed measurement noise. For these reasons, the probability distributions of d_1 and d_2 are approximated by a normal distribution. Given the closeness to the normal distribution, this approximation does not lead to significant error.

The approximation is obtained by utilizing the moment-generating function, which allows finding the first and second central moments, leading to a moment-matching approximation of the synthesized measurements. For more information regarding moment-generating functions, the reader is referred to [141].

Based on (3.16) and (3.20), the first extended synthesized measurement vector is

$$d_1 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} \dfrac{i_d(k) - i_d(k-1)}{T_s} \\ -\dfrac{\pi}{\tau_s} i_q \dfrac{l(k) - l(k-1)}{T_s} \\ i_d \\ 0 \\ u_d \end{bmatrix} \sim \mathcal{N}(\mu_1, \Sigma_1). \tag{4.45}$$

As the fourth element of x_1 is zero, it is omitted in the following text, leading to shortened synthesized measurement d_{12}. The zero will be re-included later.

In this case, the moment-generating function is

$$M_{d_{12}}(t) = \mathbb{E}(e^{t^T d_{12}}) = \int_{-\infty}^{\infty} e^{t^T d_{12}} \mathcal{N}\left(\overline{\rho}, \Sigma_\rho\right) d\rho, \tag{4.46}$$

where $t \in \mathbb{R}^4$.

The integral can be solved, for example, as multiple integral

$$M_{d_{12}}(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{t^T d_{12}} \mathcal{N}\left(\overline{\rho}, \Sigma_\rho\right) d\rho_1 \ldots d\rho_8. \tag{4.47}$$

Although calculating this integral is not particularly difficult in this case, the number of variables makes the problem very laborious and space-demanding. The authors were unable to record the resulting integral in a manner that would fit within this work.

The first moment is obtained as

$$\mu_{12} = \mathbb{E}[d_{12}] = \frac{\partial}{\partial t} M_{d_{12}}(t)\Big|_{t=\mathbf{0}} = \begin{bmatrix} \dfrac{\bar{i}_d(k) - \bar{i}_d(k-1)}{T_s} \\ -\dfrac{\pi}{\tau_s} \bar{i}_q \dfrac{\bar{l}(k) - \bar{l}(k-1)}{T_s} \\ \bar{i}_d \\ \overline{u}_d \end{bmatrix} \tag{4.48}$$

The second moment is obtained as Hessian of the MGF:

$$\mathbb{E}[d_{12}d_{12}^T] = \frac{\partial^2}{\partial^2 t} M_{d_{12}}(t)\bigg|_{t=\mathbf{0}} = [M_1, M_2]\,, \tag{4.49}$$

where

$$M_1 = \begin{bmatrix} \frac{2\sigma_{idid}+(\bar{i}_d(k)-\bar{i}_d(k-1))^2}{T_s^2} & \frac{\pi(\bar{l}(k-1)-\bar{l}(k))(\bar{i}_d(k)\bar{i}_q(k)-\bar{i}_d(k-1)\bar{i}_q(k)+\sigma_{idiq})}{T_s^2\tau_s} \\ \frac{\pi(\bar{l}(k-1)-\bar{l}(k))(\bar{i}_d(k)\bar{i}_q(k)-\bar{i}_d(k-1)\bar{i}_q(k)+\sigma_{idiq})}{T_s^2\tau_s} & \frac{\pi^2(\bar{i}_q(k)^2+\sigma_{iqiq})(\bar{l}(k-1)^2-2\bar{l}(k)\bar{l}(k-1)+\bar{l}(k)^2+2\sigma_{ll})}{T_s^2\tau_s^2} \\ \frac{\bar{i}_d(k)^2-\bar{i}_d(k)\bar{i}_d(k-1)+\sigma_{idid}}{T_s} & \frac{\pi(\bar{i}_d(k)\bar{i}_q(k)+\sigma_{idiq})(\bar{l}(k-1)-\bar{l}(k))}{T_s\tau_s} \\ \frac{\overline{u}_d(k)(\bar{i}_d(k)-\bar{i}_d(k-1))}{T_s} & \frac{\bar{i}_q(k)\pi\overline{u}_d(k)(-\bar{l}(k-1)+\bar{l}(k))}{T_s\tau_s} \end{bmatrix} \tag{4.50}$$

and

$$M_2 = \begin{bmatrix} \frac{\bar{i}_d(k)^2-\bar{i}_d(k)\bar{i}_d(k-1)+\sigma_{idid}}{T_s} & \frac{\overline{u}_d(k)(\bar{i}_d(k)-\bar{i}_d(k-1))}{T_s} \\ \frac{\pi(\bar{i}_d(k)\bar{i}_q(k)+\sigma_{idiq})(\bar{l}(k-1)-\bar{l}(k))}{T_s\tau_s} & \frac{\bar{i}_q(k)\pi\overline{u}_d(k)(-\bar{l}(k-1)+\bar{l}(k))}{T_s\tau_s} \\ \bar{i}_d(k)^2+\sigma_{idid} & \bar{i}_d(k)\overline{u}_d(k) \\ \bar{i}_d(k)\overline{u}_d(k) & \overline{u}_d(k)^2+\sigma_{udud} \end{bmatrix}. \tag{4.51}$$

The second central moment can now be calculated as

$$\begin{aligned} &\mathbb{E}[(d_{12}-\mu_{12})(d_{12}-\mu_{12})^T] \\ =&\,\mathbb{E}[d_{12}d_{12}^T - 2d_{12}\mu_{12}^T + \mu_{12}\mu_{12}^T)] \\ =&\,\mathbb{E}[d_{12}d_{12}^T] - 2\,\mathbb{E}[d_{12}]\mu_{12}^T + \mu_{12}\mu_{12}^T \\ =&\,\mathbb{E}[d_{12}d_{12}^T] - \mathbb{E}[d_{12}]\,\mathbb{E}[d_{12}]^T \\ =&\begin{bmatrix} \frac{2\sigma_{idid}}{T_s^2} & \frac{-\pi\sigma_{idiq}(-\bar{l}(k-1)+\bar{l}(k))}{T_s^2\tau_s} & \frac{\sigma_{idid}}{T_s} & 0 \\ \frac{-\pi\sigma_{idiq}(-\bar{l}(k-1)+\bar{l}(k))}{T_s^2\tau_s} & \sigma_{22} & \frac{-\pi\sigma_{idiq}(-\bar{l}(k-1)+\bar{l}(k))}{T_s\tau_s} & 0 \\ \frac{\sigma_{idid}}{T_s} & \frac{-\pi\sigma_{idiq}(-\bar{l}(k-1)+\bar{l}(k))}{T_s\tau_s} & \sigma_{idid} & 0 \\ 0 & 0 & 0 & \sigma_{udud} \end{bmatrix}, \end{aligned} \tag{4.52}$$

where

$$\sigma_{22} = \frac{\pi^2(2\bar{i}_q(k)^2\sigma_{ll} + (\bar{l}(k-1)-\bar{l}(k))^2\sigma_{iqiq} + 2\sigma_{ll}\sigma_{iqiq})}{T_s^2\tau_s^2} \tag{4.53}$$

The values of $\bar{l}(k)-\bar{l}(k-1)$ and $\bar{i}_q(k)$ appear in the second central moment, in order to provide a time-invariant covariance matrix, those need to be approximated with constants $\Delta\bar{l}_0$ and $\bar{i}_{q,0}$ respectively, resulting in

$$\Sigma_{12}=\begin{bmatrix} \frac{2\sigma_{idid}}{T_s^2} & \frac{-\pi\sigma_{idiq}(\Delta\bar{l}_0)}{T_s^2\tau_s} & \frac{\sigma_{idid}}{T_s} & 0 \\ \frac{-\pi\sigma_{idiq}(\Delta\bar{l}_0)}{T_s^2\tau_s} & \dfrac{\pi^2(2\bar{i}_{q,0}^2\sigma_{ll}+(\Delta\bar{l}_0)^2\sigma_{iqiq}+2\sigma_{ll}\sigma_{iqiq})}{T_s^2\tau_s^2} & \frac{-\pi\sigma_{idiq}(\Delta\bar{l}_0)}{T_s\tau_s} & 0 \\ \frac{\sigma_{idid}}{T_s} & \frac{-\pi\sigma_{idiq}(\Delta\bar{l}_0)}{T_s\tau_s} & \sigma_{idid} & 0 \\ 0 & 0 & 0 & \sigma_{udud} \end{bmatrix}, \quad (4.54)$$

The distribution of $d_{12} \sim \mathcal{N}(\mu_{12}, \Sigma_{12})$ is defined by (4.48) and (4.53).

For the second extended synthesized measurement vector

$$d_2 = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} \frac{\pi}{\tau_s} i_d \frac{l(k)-l(k-1)}{T_s} \\ \dfrac{i_q(k)-i_q(k-1)}{T_s} \\ i_q \\ \frac{\pi}{\tau_s}\frac{l(k)-l(k-1)}{T_s} \\ u_q \end{bmatrix} \sim \mathcal{N}(\mu_2, \Sigma_2), \quad (4.55)$$

an analogous procedure can be adopted.

The moment-generating function for the second vector is

$$M_{d_2}(t) = \mathbb{E}(e^{t^T d_2}) = \int_{-\infty}^{\infty} e^{t^T d_2}\, \mathcal{N}(\bar{\rho}, \Sigma_\rho)\, d\rho, \quad (4.56)$$

where $t \in \mathbb{R}^5$.

Once again, the integral can be solved as multiple integral similar to (4.47). However, due to the higher dimensionality of the problem, the resulting function is too large to be included in this thesis.

The first moment is

$$\mu_2 = \mathbb{E}[d_2] = \frac{\partial}{\partial t} M_{d_2}(t)\bigg|_{t=\mathbf{0}} = \begin{bmatrix} \frac{\pi}{\tau_s} \bar{i}_d \frac{\bar{l}(k)-\bar{l}(k-1)}{T_s} \\ \dfrac{\bar{i}_q(k)-\bar{i}_q(k-1)}{T_s} \\ \bar{i}_q \\ \frac{\pi}{\tau_s}\frac{\bar{l}(k)-\bar{l}(k-1)}{T_s} \\ \bar{u}_q \end{bmatrix} \sim \mathcal{N}(\mu_2, \Sigma_2). \quad (4.57)$$

The second moment is obtained as Hessian of the (4.56):

$$\mathbb{E}[d_2 d_2^T] = \frac{\partial^2}{\partial^2 t} M_{d_2}(t)\bigg|_{t=\mathbf{0}} = [M_1, M_2]\,, \tag{4.58}$$

$$M_1 = \begin{bmatrix} \frac{\pi^2(\bar{i}_d(k)^2+\sigma_{idid})(\bar{l}(k-1)^2-2\bar{l}(k)\bar{l}(k-1)+\bar{l}(k)^2+2\sigma_{ll})}{T_s^2\tau_s^2} & \frac{\pi(-\bar{l}(k-1)+\bar{l}(k))(-\bar{i}_d(k)\bar{i}_q(k-1)+\bar{i}_d(k)\bar{i}_q(k)+\sigma_{idiq})}{T_s^2\tau_s} \\ \frac{\pi(-\bar{l}(k-1)+\bar{l}(k))(-\bar{i}_d(k)\bar{i}_q(k-1)+\bar{i}_d(k)\bar{i}_q(k)+\sigma_{idiq})}{T_s^2\tau_s} & \frac{2\sigma_{iqiq}+(-\bar{i}_q(k-1)+\bar{i}_q(k))^2}{T_s^2} \\ \frac{\pi(\bar{i}_d(k)\bar{i}_q(k)+\sigma_{idiq})(-\bar{l}(k-1)+\bar{l}(k))}{T_s\tau_s} & \frac{-\bar{i}_q(k)\bar{i}_q(k-1)+\bar{i}_q(k)^2+\sigma_{iqiq}}{T_s} \\ \frac{\bar{i}_d(k)\pi^2(2\sigma_{ll}+(-\bar{l}(k-1)+\bar{l}(k))^2)}{T_s^2\tau_s^2} & \frac{\pi(-\bar{i}_q(k-1)+\bar{i}_q(k))(-\bar{l}(k-1)+\bar{l}(k))}{T_s^2\tau_s} \\ \frac{\bar{i}_d(k)\pi\bar{u}_q(k)(-\bar{l}(k-1)+\bar{l}(k))}{T_s\tau_s} & \frac{\bar{u}_q(k)(-\bar{i}_q(k-1)+\bar{i}_q(k))}{T_s} \end{bmatrix} \tag{4.59}$$

and

$$M_2 = \begin{bmatrix} \frac{\pi(\bar{i}_d(k)\bar{i}_q(k)+\sigma_{idiq})(-\bar{l}(k-1)+\bar{l}(k))}{T_s\tau_s} & \frac{\bar{i}_d(k)\pi^2(2\sigma_{ll}+(-\bar{l}(k-1)+\bar{l}(k))^2)}{T_s^2\tau_s^2} & \frac{\bar{i}_d(k)\pi\bar{u}_q(k)(-\bar{l}(k-1)+\bar{l}(k))}{T_s\tau_s} \\ \frac{-\bar{i}_q(k)\bar{i}_q(k-1)+\bar{i}_q(k)^2+\sigma_{iqiq}}{T_s} & \frac{\pi(-\bar{i}_q(k-1)+\bar{i}_q(k))(-\bar{l}(k-1)+\bar{l}(k))}{T_s^2\tau_s} & \frac{\bar{u}_q(k)(-\bar{i}_q(k-1)+\bar{i}_q(k))}{T_s} \\ \bar{i}_q(k)^2+\sigma_{iqiq} & \frac{\bar{i}_q(k)\pi(-\bar{l}(k-1)+\bar{l}(k))}{T_s\tau_s} & \bar{i}_q(k)\bar{u}_q(k) \\ \frac{\bar{i}_q(k)\pi(-\bar{l}(k-1)+\bar{l}(k))}{T_s\tau_s} & \frac{\pi^2(\bar{l}(k-1)^2-2\bar{l}(k)\bar{l}(k-1)+\bar{l}(k)^2+2\sigma_{ll})}{T_s^2\tau_s^2} & \frac{\pi\bar{u}_q(k)(-\bar{l}(k-1)+\bar{l}(k))}{T_s\tau_s} \\ \bar{i}_q(k)\bar{u}_q(k) & \frac{\pi\bar{u}_q(k)(-\bar{l}(k-1)+\bar{l}(k))}{T_s\tau_s} & \bar{u}_q(k)^2+\sigma_{uquq} \end{bmatrix} \tag{4.60}$$

The second central moment is obtained from (4.57) and (4.59) as

$$\begin{aligned} &\mathbb{E}[(d_2-\mu_2)(d_2-\mu_2)^T] \\ &= \mathbb{E}[d_2 d_2^T] - \mathbb{E}[d_2]\,\mathbb{E}[d_2]^T \\ &= \begin{bmatrix} \sigma_{11} & \frac{\pi\sigma_{idiq}(-\bar{l}(k-1)+\bar{l}(k))}{T_s^2\tau_s} & \frac{\pi\sigma_{idiq}(-\bar{l}(k-1)+\bar{l}(k))}{T_s\tau_s} & \frac{2i_d\pi^2\sigma_{ll}}{T_s^2\tau_s^2} & 0 \\ \frac{\pi\sigma_{idiq}(-\bar{l}(k-1)+\bar{l}(k))}{T_s^2\tau_s} & \frac{2\sigma_{iqiq}}{T_s^2} & \frac{\sigma_{iqiq}}{T_s} & 0 & 0 \\ \frac{\pi\sigma_{idiq}(-\bar{l}(k-1)+\bar{l}(k))}{T_s\tau_s} & \frac{\sigma_{iqiq}}{T_s} & \sigma_{iqiq} & 0 & 0 \\ \frac{2\bar{i}_d(k)\pi^2\sigma_{ll}}{T_s^2\tau_s^2} & 0 & 0 & \frac{2\pi^2\sigma_{ll}}{T_s^2\tau_s^2} & 0 \\ 0 & 0 & 0 & 0 & \sigma_{uquq} \end{bmatrix} \end{aligned} \tag{4.61}$$

where

$$\sigma_{11} = \frac{\pi^2(2\bar{i}_d(k)^2\sigma_{ll} + (\bar{l}(k-1) - \bar{l}(k))^2\sigma_{idid} + 2\sigma_{ll}\sigma_{idid})}{T_s^2\tau_s^2} \tag{4.62}$$

To ensure a time-invariant covariance matrix, terms $\bar{l}(k)-\bar{l}(k-1$ and $\bar{i}_d(k)$ should be approximated by the constants $\Delta\bar{l}_0$ and $\bar{i}_{d,0}$, respectively. This approximation results in a consistent covariance matrix over time

$$\Sigma_2 = \begin{bmatrix} \frac{\pi^2(2\bar{i}_{d,0}^2\sigma_{ll}+(\Delta\bar{l}_0)^2\sigma_{idid}+2\sigma_{ll}\sigma_{idid})}{T_s^2\tau_s^2} & \frac{\pi\sigma_{idiq}(-\Delta\bar{l}_0)}{T_s^2\tau_s} & \frac{\pi\sigma_{idiq}(-\Delta\bar{l}_0)}{T_s\tau_s} & \frac{2\bar{i}_{d,0}\pi^2\sigma_{ll}}{T_s^2\tau_s^2} & 0 \\ \frac{\pi\sigma_{idiq}(-\Delta\bar{l}_0)}{T_s^2\tau_s} & \frac{2\sigma_{iqiq}}{T_s^2} & \frac{\sigma_{iqiq}}{T_s} & 0 & 0 \\ \frac{\pi\sigma_{idiq}(-\Delta\bar{l}_0)}{T_s\tau_s} & \frac{\sigma_{iqiq}}{T_s} & \sigma_{iqiq} & 0 & 0 \\ \frac{2\bar{i}_{d,0}\pi^2\sigma_{ll}}{T_s^2\tau_s^2} & 0 & 0 & \frac{2\pi^2\sigma_{ll}}{T_s^2\tau_s^2} & 0 \\ 0 & 0 & 0 & 0 & \sigma_{uquq} \end{bmatrix} \tag{4.63}$$

As a result, the distribution of synthesized measurements $d_2 \sim \mathcal{N}(\mu_2, \Sigma_2)$ is fully defined by (4.57) and (4.63).

Recalling from the problem definition, Section 3.2 proposes approximating individual likelihoods by the normal distribution in order to find a normal approximation of the posterior distribution.

$$p\left(\theta|D_1, D_2, \mathbb{E}\left[\overline{X_1}|D_1\right], \mathbb{E}\left[\overline{X_2}|D_2\right]\right) \propto q_1\left(\theta|D_1, \mathbb{E}\left[\overline{X_1}|D_1\right]\right) q_2\left(\theta|D_2, \mathbb{E}\left[\overline{X_2}|D_2\right]\right) p(\theta), \tag{4.64}$$

where

$$q\left(\theta|D_j, \mathbb{E}\left[\overline{X_j}|D_j\right]\right) \approx \mathcal{L}\left(\theta|D_j, \mathbb{E}\left[\overline{X_j}|D_j\right]\right), \quad j \in \{1,2\}. \tag{4.65}$$

As the likelihood is equal to the posterior of the EIV process under the assumption of uninformative prior $p(\theta) \propto 1$

$$\mathcal{L}\left(\theta|D_j, \mathbb{E}\left[\overline{X_j}|D_j\right]\right) = p(\theta_j|D_j), \quad j \in \{1,2\}, \tag{4.66}$$

the surrogate likelihood can be obtained as Laplace surrogate posterior $p(\theta_j|D_j)$

$$q_j\left(\theta|D_j, \mathbb{E}\left[\overline{X_j}|D_j\right]\right) \approx q_j(\theta_j|D_j). \tag{4.67}$$

B-GTLS algorithm can be utilized to provide such surrogate posterior by solving the following EIV processes:

$$\begin{aligned} \overline{y}_{i,1} &= \overline{x}_{i,1}^T\theta_{12} \\ \begin{bmatrix} \overline{x}_{i,1} \\ \overline{y}_{i,1} \end{bmatrix} &+ \varepsilon_{i,1} = \mu_{i,1} + \varepsilon_{i,1}, \quad \varepsilon_1 \sim \mathcal{N}(\mathbf{0}, \Sigma_1), \end{aligned} \tag{4.68}$$

$$\begin{aligned} \overline{y}_{i,2} &= \overline{x}_{i,2}^T\theta \\ \begin{bmatrix} \overline{x}_{i,2} \\ \overline{y}_{i,2} \end{bmatrix} &+ \varepsilon_{i,2} = \mu_{i,2} + \varepsilon_{i,2}, \quad \varepsilon_2 \sim \mathcal{N}(\mathbf{0}, \Sigma_2), \end{aligned} \tag{4.69}$$

where θ_{12} is a subvector of θ, missing its last element due to zero corresponding measurements in the equation.

Both datasets D_1 and D_2 are composed of the same measurements (4.41), we can, without loss of generality, simplify the terminology by writing just dataset D representing the measurements. Solving the problem provides a surrogate likelihoods

$$q_1(\theta_{12}|D) = \mathcal{N}\left(\hat{\theta}_{12}, \hat{\Sigma}_{12}(\hat{\theta}_{12})\right) \tag{4.70}$$

$$q_2(\theta|D) = \mathcal{N}\left(\hat{\theta}_2, \hat{\Sigma}(\hat{\theta}_2)\right) \tag{4.71}$$

respectively, as described in lemma 4.

The zero value omitted in the d_{11} can be re-included into the surrogate posterior as

$$q_1(\theta|D) = \mathcal{N}\left(\hat{\theta}_1, \hat{\Sigma}_1(\hat{\theta}_1)\right), \tag{4.72}$$

where

$$\hat{\theta}_1 = \begin{bmatrix} \hat{\theta}_{12} \\ 0 \end{bmatrix}, \quad \hat{\Sigma}_1(\hat{\theta}_1) = \begin{bmatrix} \hat{\Sigma}_{12}(\hat{\theta}_{12}) & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}. \tag{4.73}$$

Given the surrogate likelihoods (4.72) and (4.71), the surrogate posterior distribution is

$$p\left(\theta|D\right) \approx q\left(\theta|D\right) = q_1(\theta|D)q_2(\theta|D)p(\theta), \tag{4.74}$$

where $p(\theta) = \mathcal{N}\left(\mu_p, \Sigma_p\right)$ is conjugate normal prior distribution of the parameters θ with parameters μ_p and Σ_p selected by the statistician.

As the likelihoods and prior distributions are normal, the posterior distribution is also normal

$$q\left(\theta|D\right) = \mathcal{N}\left(\mu_\theta, \Sigma_\theta\right), \tag{4.75}$$

with

$$\Sigma_\theta^{-1} = \hat{\Sigma}_1(\hat{\theta}_1)^{-1} + \hat{\Sigma}_2(\hat{\theta}_2)^{-1} + \Sigma_p^{-1} \tag{4.76}$$

and

$$\mu_\theta = \Sigma_\theta\left(\hat{\Sigma}(\hat{\theta}_1)^{-1}\hat{\theta}_1 + \hat{\Sigma}(\hat{\theta}_2)^{-1}\hat{\theta}_2 + \Sigma_p^{-1}\mu_p\right). \tag{4.77}$$

This process can be summed up in the single recursive algorithm

Algorithm 6 Bayesian Recursive PMLSM GTLS (PMLSM-RGTLS)

Input: D, Σ_1, Σ_2, Σ_p^{-1}, μ_p λ

Output: $\hat{\mu}_\theta$, $\hat{\Sigma}_\theta^{-1}$

$\hat{\theta}_{12} \leftarrow \mu_{p,21}$

$\hat{\theta}_2 \leftarrow \mu_p$

$P_1 \leftarrow c\mathbf{I}$

$P_2 \leftarrow c\mathbf{I}$

$P_1^{-1} \leftarrow \frac{1}{c}\mathbf{I}$

$P_2^{-1} \leftarrow \frac{1}{c}\mathbf{I}$

for $i \leftarrow 1$ to N **do**

  $\mu_{12} \leftarrow$ evaluate from (4.48)

  $L \leftarrow (P_1\mu_{12})/(\lambda + \mu_{12}^T P_1 \mu_{12})$

  $P_1 \leftarrow (\mathbf{I} - L\mu_{12}^T)P_1\frac{1}{\lambda}$

  $Z \leftarrow P_1\Sigma_1[\hat{\theta}_{12}^T, -1]^T$

  $\hat{\theta}_{12} \leftarrow -z_{12}/z_{22}$

  $P_1^{-1} \leftarrow \frac{1}{\lambda}P_1^{-1} + d_{12,i}d_{12,i}^T$

  $\hat{\Sigma}_{12}^{-1} \leftarrow$ evaluate from (4.53)

  $\hat{\Sigma}_1^{-1} \leftarrow$ evaluate from (4.73)

  $\mu_2 \leftarrow$ evaluate from (4.57)

  $L \leftarrow (P_2\mu_2)/(\lambda + \mu_2^T P_2 \mu_2)$

  $P_2 \leftarrow (\mathbf{I} - L\mu_2^T)P_2\frac{1}{\lambda}$

  $Z \leftarrow P_2\Sigma_2[\hat{\theta}_2^T, -1]^T$

  $\hat{\theta}_2 \leftarrow -z_{12}/z_{22}$

  $P_2^{-1} \leftarrow \frac{1}{\lambda}P_2^{-1} + d_{2,i}d_{2,i}^T$

  $\hat{\Sigma}_2^{-1} \leftarrow$ evaluate from (4.63)

  $\hat{\Sigma}_\theta^{-1} \leftarrow$ evaluate from (4.76)

  $\hat{\mu}_\theta \leftarrow$ evaluate from (4.77)

end for

In Algorithm 6, the D denotes the dataset containing true measurements of d and q voltages, d and q currents and position; Σ_1 and Σ_2 are synthesized measurements noise defined by (4.53) and (4.63), respectively; μ_p and Σ_p^{-1} are prior mean and variance, respectively; λ is forgetting factor and c is a large number.

Examination of the algorithm reveals a structure consisting of two parallel B-RGTLS algorithms for parameter identification using synthesized data. The first surrogate likelihood is augmented, and the resulting surrogate likelihoods are combined with the prior distribution to compute the surrogate posterior distribution. This is repeated for each data sample. This process is illustrated in the following figure, referencing the corresponding equations.

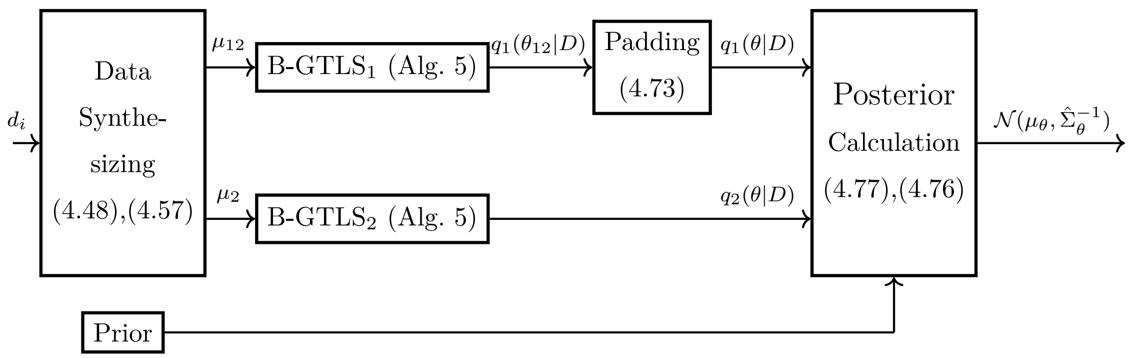


Fig. 4.1: Structure of the PMLSM-RGTLS algorithm.

5 Bayesian Total Least Squares

In the previous chapter, Bayesian generalized total least squares arise from Bayesian analysis of errors-in-variables

$$\overline{y}_i = \overline{x}_i^T \theta, \tag{5.1a}$$

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} \overline{x}_i \\ \overline{y}_i \end{bmatrix} + \varepsilon_i, \quad \varepsilon \sim \mathcal{N}(\mathbf{0}, \Sigma), \tag{5.1b}$$

problem. The approximative posterior distribution is derived

$$p\left(\theta | D, \mathbb{E}\left[\overline{X} | D\right]\right) \propto \mathcal{L}\left(\theta | D, \mathbb{E}\left[\overline{X} | D\right]\right) p(\theta), \tag{5.2}$$

given arbitrary noise covariance matrix Σ. The result can be interpreted as a statistical expansion of the well-known generalized total least squares (GTLS).

As TLS is more popular compared to GTLS, and many GTLS problems can be transformed into TLS problems, the special case of the errors-in-variables with identity noise covariance deserves its chapter. Therefore, in this chapter, we will focus on the same problem as GTLS, but with the assumption of $\Sigma = \sigma_n^2 \mathbf{I}$, stemming from the definition of the total least squares (TLS) introduced in chapter 2.7. As the TLS is a special case of the GTLS, we can expect B-GTLS to reduce into Bayesian expansion of the total least squares (B-TLS).

The goal of this chapter is to explore this special case. It will be shown that the simpler form of the posterior allows for the interpretation of the posterior using directional statistics. This section, along with its findings, serves as an extension of the author's earlier work published in [1].

5.1 Derivation of the Bayesian Total Least Squares

Deriving the B-TLS from the ground up would be repetitive, as the derivation can be analogous to the B-GTLS. Alternative derivations can be studied, for example, in [2, 1]. We will follow the derivation of the B-GTLS from chapter 4 and remark on uniqueness that substitution of $\Sigma = \sigma_n^2 \mathbf{I}$ causes in the next section.

The problem is defined as the identification of parameter θ from a linear model

$$\overline{y}_i = \overline{x}_i^T \theta, \tag{5.3a}$$

based on samples burdened with normal noise with zero mean and unit covariance multiplied with variance σ_n^2

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} \overline{x}_i \\ \overline{y}_i \end{bmatrix} + \varepsilon_i, \quad \varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I}). \tag{5.3b}$$

The result of the identification process is an approximative posterior distribution of the parameter vector

$$p\left(\theta|D, \mathbb{E}\left[\overline{X}|D\right]\right) \propto \mathcal{L}\left(\theta|D, \mathbb{E}\left[\overline{X}|D\right]\right) p(\theta), \tag{5.4}$$

which requires approximation of the hidden variables' probability. The expected value of the probability functions as such approximation

$$p(\overline{X}|D) \approx \mathbb{E}\left[\overline{X}|D\right]. \tag{5.5}$$

From (4.5), the expectation of the hidden variables is

$$\mathbb{E}\left[\begin{array}{c}\overline{x}\\ \overline{y}\end{array}\middle| d\right] = \begin{bmatrix} x - \dfrac{\theta x^T\theta - \theta y}{\theta^T\theta + 1} \\ \dfrac{\theta^T\theta y + x^T\theta}{\theta^T\theta + 1} \end{bmatrix} = d - \frac{d^T\vartheta}{\vartheta^T\vartheta}\vartheta. \tag{5.6}$$

The likelihood of the dataset is from (4.17) equal to

$$\mathcal{L}(\theta|D) = (2\pi\sigma_n^2)^{-\frac{m+1}{2}} \exp\left(-\frac{1}{2\sigma_n^2}\frac{\vartheta^T\left(D^T D\right)\vartheta}{\vartheta^T\vartheta}\right). \tag{5.7}$$

Choosing the prior to be conjugate results in

$$p(\theta) \propto \exp\left(\frac{\vartheta^T\Phi_p\vartheta}{\vartheta^T\vartheta}\right), \tag{5.8}$$

where $\Phi_p \preceq 0$ is a symmetric, negative semi-definite matrix representing prior knowledge of the statistician.

Formulation of the likelihood (5.7) and prior (5.8) leads to the following posterior distribution

$$p(\theta|D) \propto \exp\left(\frac{\vartheta^T\Omega\vartheta}{\vartheta^T\vartheta}\right), \tag{5.9a}$$

where

$$\Omega_j = -\tfrac{1}{2\sigma_n^2}D^T D + \Phi_p. \tag{5.9b}$$

As the resulting posterior is a special case of B-GTLS, all properties of the B-GTLS are carried over, including unknown moments, full specification by $\{\Omega\,\sigma_n^2\}$ and recursive update

$$\Omega(k+1) = \Omega(k) - \frac{1}{2\sigma_n^2}\mathcal{D}^T\mathcal{D}, \tag{5.10}$$

where

$$\mathcal{D} = \left[d_{(N+1)},\, d_{(N+2)},\, \ldots,\, d_{(N+M)}\right]^T. \tag{5.11}$$

The possibility of obtaining a MAP solution using existing GTLS algorithms is also retained. Furthermore, connections with the total least squares will be revealed in the next section.

5.2 Bayesian Total Least Squares properties

This section will focus on new properties emerging from the simplified form of the posterior distribution compared to B-GTLS.

Inspection of the resulting hidden variables expectation reveals, that (5.6) is a well known mathematical expression of the orthogonal projection (2.28). This directly connects the B-TLS with the TLS. The TLS minimizes the perpendicular (orthogonal) distances to the line. From the definition of the TLS provided in chapter 2.7, the minimized distance is the difference between the measured value and its orthogonal projection to the solution, which can be obtained using (5.6).

Compared to the TLS understanding of minimizing orthogonal distances, in this scenario, in order to obtain maximum a posteriori estimate θ_{MAP}, the probability of the orthogonal projections of samples from the dataset D is maximized.

Subsequent connection is apparent in the exponent of the likelihood (5.7) and the posterior functional form (5.9a), which encompasses solely the Rayleigh quotient. By optimizing the Rayleigh Quotient, the probability distribution function is maximized. As a consequence, not only GTLS but also simpler TLS-based algorithms [3, 142] can be utilized to obtain maximum likelihood and maximum a posteriori point estimates.

As the scale-invariant property (4.37) of the likelihood, prior and posterior distributions is also carried over, we can define the following scaling factor

$$c = \frac{1}{\sqrt{\vartheta^T \vartheta}} = \frac{1}{\|\vartheta\|}. \tag{5.12}$$

Utilizing the scaling factor and the following substitution $v = c\vartheta$, the distributions take functional form proportional to the Bingham distribution introduced in chapter 2.5

$$p(\theta) = \mathcal{B}(v|\Phi_p) \propto \exp\left(\frac{\vartheta^T \Phi_p \vartheta}{\vartheta^T \vartheta}\right), \tag{5.13}$$

$$\mathcal{L}(\theta|D) \propto \mathcal{B}\left(v\middle|\frac{-D^T D}{2\sigma_n^2}\right) \propto (2\pi\sigma_n^2)^{-\frac{m+1}{2}} \exp\left(-\frac{1}{2\sigma_n^2}\frac{\vartheta^T \left(D^T D\right)\vartheta}{\vartheta^T \vartheta}\right), \tag{5.14}$$

$$p(\theta|D) = \mathcal{B}(v|\Omega) \propto \exp\left(\frac{\vartheta^T \Omega \vartheta}{\vartheta^T \vartheta}\right). \tag{5.15}$$

The prior, likelihood, and posterior probability distributions, represented by the directional Bingham distribution, provide a clearer understanding of the vector v in the following context: The singular vector $\hat{v}$, which corresponds to the smallest singular value of matrix A, acts as the orthogonal vector to the solution hyperplane. The Bingham distribution $\mathcal{B}(v|A)$ allocates probability values to all potential directions that this orthogonal vector might assume. These possible directions encompass all points within $\mathbb{R}^{m+1}$ constrained by the unit hypersphere $\mathbb{S}^m$, where m denotes the dimension of θ. The vector $\hat{v}$ is identified as the mode of the resulting probability distribution function.

In a manner similar to TLS, the vector $v = [v_{12}^T,\, v_{22}]^T$ can be normalized to derive the unknown vector θ

$$\theta = f(v) = \frac{-v_1 2}{v_{22}}, \tag{5.16}$$

where v_{22} is the last element of vector v.

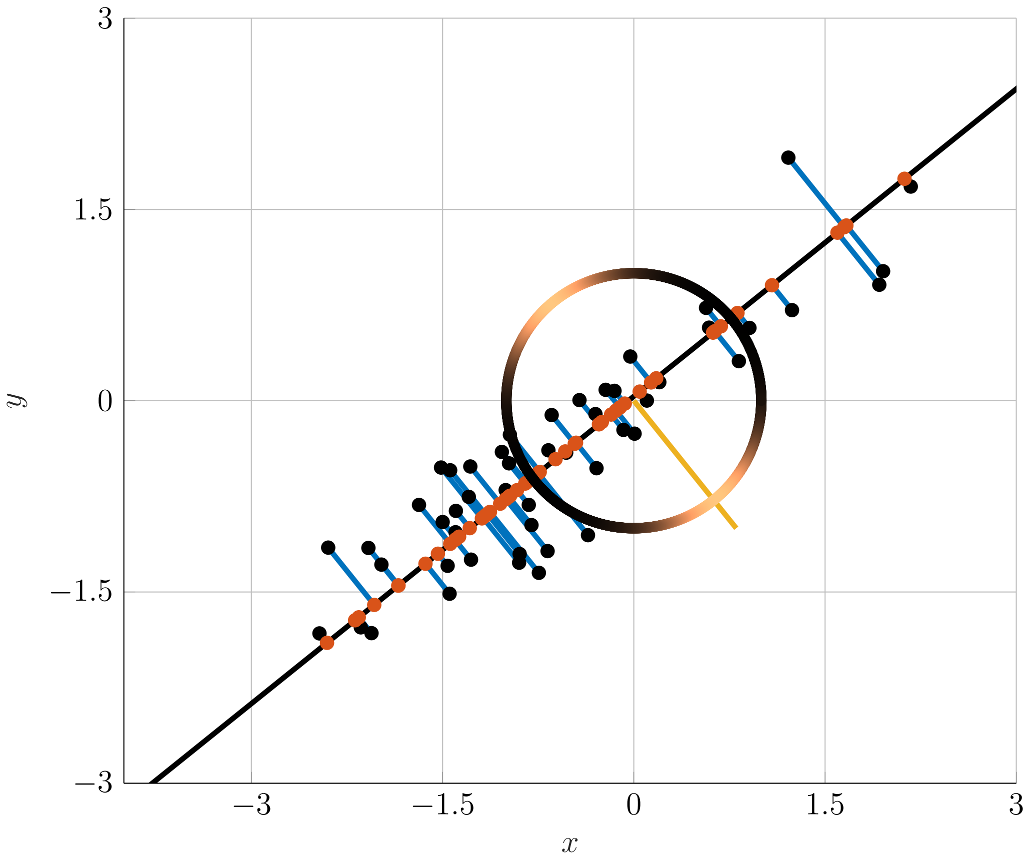


Fig. 5.1: Geometrical interpretation of the Bingham distribution on circle $\mathbb{S}$. The colored circle shows the Bingham distribution for the dataset D (black dots). The lighter the color, the higher the probability. The solution (black line) is defined by its orthogonal vector ϑ (yellow). The scaled orthogonal vector $c\vartheta$ coincides with the mean concentration vector of the Bingham distribution, the MAP solution. The expectation of the nuisance pdf $\mathbb{E}\left[\overline{X}|D\right]$ is displayed as red dots. The connection with TLS is apparent by comparing this figure with Figure 2.2.

It should also be noted that by the antipodal property of the Bingham distribution, there are always two antipodal points, v and $-v$, with the same probability, resulting in a multimodal posterior distribution. It seems as if there are two possible MAP solutions to the problem. However, normalization using (5.16) results in the same solution

$$\theta = f(v) = f(-v). \tag{5.17}$$

Compared to the general case of B-GTLS, the simpler functional form of the special case prior, likelihood, and posterior distribution functions allow for simpler maximum likelihood estimators and maximum a posteriori estimators based on TLS, establishing confidence limits based on knowledge of the Bingham distribution, sampling from the Bingham posterior density, and visual representation and natural representation of the calculation.

5.3 Normal Prior

Previous chapters described the case where the functional form of the prior probability distribution is chosen to be conjugate to the TLS likelihood function. If conjugacy is not a requirement, choosing any probability distribution for the prior is possible, thereby achieving different properties. In the following paragraphs, we will compare the previously chosen conjugate Bingham prior with a normal prior.

The Bingham prior, as introduced in the previous chapters, $p_{\mathcal{B}}(\theta) = \mathcal{B}(v|\Phi_p)$ allows the formulation of the posterior distribution of the transformed parameters of the identified system in the form of the well-known Bingham distribution

$$p_{\mathcal{B}}(v|D) \propto \mathcal{B}(v|\Omega) = \mathcal{B}\left(v \middle| \frac{-D^T D}{2\sigma_n^2} + \Phi_p\right), \tag{5.18}$$

where the subscript denotes posterior with Bingham prior.

Although the Bingham distribution has an analytical form, it is not a closed mathematical expression due to its normalization constant. Therefore, working with this posterior can be challenging. Additionally, the posterior distribution represents the probability distribution of the transformed parameters v instead of the parameters θ, necessitating a back-transformation to interpret the results.

Expressing Bayesian assumptions of transformed parameters in the form of a Bingham prior distribution can also be problematic, as most statisticians are accustomed to working with other, more common distributions. In particular, the normal distribution is widely used.

Using a normal prior $p_{\mathcal{N}}(\theta) = \mathcal{N}(\mu_p, \Sigma_p)$, however, leads to a posterior distribution

$$p_{\mathcal{N}}(\theta|D) \propto \mathcal{L}(\theta|D) p_{\mathcal{N}}(\theta), \tag{5.19}$$

where the subscript denotes posterior with normal prior. The posterior functional form is an unknown probability distribution with an unknown normalization constant. Resultingly, the situation is not improved using the normal prior compared to the Bingham prior.

A closed mathematical expression for the posterior distribution $p(\theta|D)$ does not emerge when using both prior functions. This can be addressed with an appropriate approximation. Similarly to the B-GTLS, the normal distribution is selected as a suitable functional form of the surrogate posterior. The possibility of Laplace approximation for both priors is analyzed. To add to that, variational approximation is also developed.

5.4 Laplace Approximation of the Posterior

As the exact analytical solution for both the normal and the Bingham prior is infeasible, the Laplace approximation is proposed for the posterior distribution in order to provide a tractable and convenient normal approximation of the posterior probability distributions.

Let us begin by examining the case with Bingham prior

Lemma 5. *Let posterior distribution be defined by* (5.18)*. Then, the Laplace approximation is obtained as*

$$p_{\mathcal{B}}(v|D) \approx q_L(\theta|D) = \mathcal{N}\left(\hat{\theta}, \hat{\Sigma}_{\mathcal{B}}(\hat{\theta})\right), \tag{5.20}$$

where $\hat{\theta}$ is the maximum a posteriori point estimate (MAP)

$$\hat{\mu} := \arg\min_{\mu} \quad \frac{[\mu^T, -1]\Omega[\mu^T, -1]^T}{\mu^T\mu + 1}, \tag{5.21}$$

where $\Omega = \frac{-D^TD}{2\sigma_n^2} + \Phi_p$, and the surrogate covariance

$$\hat{\Sigma}_{\mathcal{B}}(\hat{\theta}) = 4\frac{t_1 T_\theta}{t_0^3} - 2\frac{T_\theta \Xi_\theta - \mu \Xi_n^T}{t_0^2} + -2\frac{\Xi_\theta T_\theta - \Xi_n \mu^T}{t_0^2} - \frac{\mathbf{I}t_1}{t_0^2} + \frac{\Xi_\theta}{t_0}, \tag{5.22}$$

is hessian of negative log posterior evaluated at $\hat{\theta}$, where

$$t_0 = \mu^T\mu + 1 \tag{5.23}$$

$$t_1 = [\mu^T - 1]\Xi[\mu^T - 1]^T \tag{5.24}$$

$$T_\theta = \mu\mu^T; \tag{5.25}$$

$\Xi_\theta \in \mathbb{R}^{n\times n}$, $\Xi_n \in \mathbb{R}^{n\times 1}$ *are the submatrices of*

$$\Xi = \begin{bmatrix} \Xi_\theta & \Xi_n \\ \Xi_n^T & \xi_n^2 \end{bmatrix} \tag{5.26}$$

and $\Xi = \Omega^{-1}$.

As in the case of the Bingham prior, the posterior is a B-TLS problem, and the TLS algorithm can be utilized to obtain the MAP solution $\hat{\mu}$.

In the case of a Gaussian prior, it is impossible to express the posterior's mode analytically, and it must be found using an optimization method. The Laplace approximation is fairly similar

Lemma 6. *Let posterior distribution be defined by* (5.19)*. Then, the Laplace approximation is obtained as*

$$p_{\mathcal{N}}(v|D) \approx q_L(\theta|D) = \mathcal{N}\left(\hat{\theta}, \hat{\Sigma}_{\mathcal{N}}(\hat{\theta})\right), \tag{5.27}$$

where $\hat{\theta}$ is the maximum a posteriori point estimate (MAP)

$$\hat{\mu} := \arg\max_{\mu} p_{\mathcal{N}}\left(\begin{bmatrix}\mu \\ -1\end{bmatrix} \middle| D\right), \tag{5.28}$$

and the surrogate covariance is

$$\hat{\Sigma}_{\mathcal{N}}(\hat{\theta}) = \hat{\Sigma}_{\mathcal{B}}(\hat{\theta})\Big|_{\Xi^{-1}=\frac{-D^T D}{2\sigma_n^2}} + \Sigma_p^{-1}, \tag{5.29}$$

where Σ_p is the covariance matrix of the normal prior.

5.5 Variational Approximation of the Posterior Distribution

The variational approximation is a significantly more complex method compared to the Laplace approximation, but it provides an optimal result [143, 144]. There are several popular variational approximation methods, the most well-known being the "mean field variational Bayes approximation" [99, 145]. However, this method cannot be used in this case because the parameter vector θ appears in the denominator of the exponent. Therefore, this chapter will focus exclusively on the variational method known as the "fixed-form variational approximation."

For the reasons described in the earlier chapter, the normal probability distribution is chosen as the functional form of the surrogate variational posterior.

$$q_{\text{VB}}(\theta|\mu, \Sigma) = \mathcal{N}(\theta|\mu, \Sigma) \approx p(\theta|D). \tag{5.30}$$

The optimal parameters are obtained by minimizing the Kullback-Leibler divergence $D_{\text{KL}}(q_{\text{VB}}(\theta|\mu, \Sigma)||\, \mathcal{L}(\theta|D)p(\theta))$.

Minimization of the D_{KL} can be achieved by maximizing the lower bound of the marginal log-likelihood, often referred to as ELBO, which stands for "Evidence Lower Bound" [145]

$$\mathbf{L}\left(q_{\mathrm{VB}}(\theta|\mu, \Sigma)\right) = \left(\ln \frac{\mathcal{L}(\theta|D)p(\theta)}{q_{\mathrm{VB}}(\theta|D)}\right), \tag{5.31}$$

Unfortunately, the integral of the Kullback-Leibler divergence and the integral of ELBO diverge when using the Bingham prior. Therefore, this method can only be utilized for a normal prior.

ELBO can be expressed analytically from the equation (5.31)

$$\begin{aligned} \mathbf{L}\left(q_{\mathrm{VB}}(\theta|z)\right) = \int_{-\infty}^{\infty} q(\theta|z) \ln \mathcal{L}(\theta|D) d\theta + \\ + \int_{-\infty}^{\infty} q(\theta|z) \ln p(\theta) d\theta - \int_{-\infty}^{\infty} q(\theta|z) \ln q(\theta|d) d\theta, \end{aligned} \tag{5.32}$$

where $z = [\mu^T,\ \mathrm{vec}(\Sigma)^T]^T$.

The expression of the second and third integrals is straightforward, but the first integral is complicated. In the case of a single variable, it can be expressed using special functions such as

$$\begin{aligned} \int_{-\infty}^{\infty} q(\theta|z) \ln \mathcal{L}(\theta|D) d\theta = -\frac{1}{2}\left(\frac{\sigma^2 + (\mu - \mu_p)^2}{\sigma_p^2} - \ln\left(\frac{\sigma^2 2\pi e}{e^a}\right) + \right. \\ \left. + \frac{\sqrt{\pi}}{\sqrt{2\sigma^2}}\left[(c-a)\, V\left(\frac{-\mu}{\sqrt{2\sigma^2}}, \frac{1}{\sqrt{2\sigma^2}}\right) - bL\left(\frac{-\mu}{\sqrt{2\sigma^2}}, \frac{1}{\sqrt{2\sigma^2}}\right)\right]\right), \end{aligned} \tag{5.33}$$

where μ and σ mean value and variance respectively of the surrogate distribution, and μ_p and σ_p are the mean and variance respectively of the normal prior. Terms a, b, and c are members of

$$\frac{-D^T D}{2\sigma_n^2} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}. \tag{5.34}$$

Functions $V(\alpha, \beta)$ and $L(\alpha, \beta)$ are the real and imaginary components of the Faddeeva function $w(\gamma)$ [146] or implementation, the Faddeeva function can be computed either using an efficient computational method [147] or a method that guarantees accuracy [148]. Numerical results show that for the purposes of approximation, the first method is sufficient.

An analytical expression of the posterior is not possible, and it must be obtained by numerically solving the optimization problem

$$\begin{aligned} z^* := \arg\min_{z} \quad & -\mathbf{L}\left(q_{\mathrm{VB}}(\theta|z)\right) \\ \text{s.t.} \quad & \Sigma \succ 0, \end{aligned} \tag{5.35}$$

constrained by the requirement for the positive definiteness of the covariance matrix of the surrogate distribution $\Sigma \succ 0$. Numerical optimization is possible, because the function $\mathbf{L}\left(q_{\mathrm{VB}}(\theta|z)\right)$ is differentiable, thanks to the following property of the Faddeeva function

$$\frac{\partial w(\gamma)}{\partial \gamma} = \frac{2i}{\sqrt{\pi}} - 2\gamma w(\gamma). \tag{5.36}$$

5.6 Design of Recursive Identification Approximation Algorithms

The posterior distribution approximation is useful for real-time applications only if recursive identification can still be performed. It is also important that the approximation in the recursive algorithm does not introduce error. Three such algorithms are presented in this chapter.

The most straightforward algorithm provides the Laplace approximation for a Bingham prior. The Laplace approximation requires the mode of the posterior distribution, which is obtained using the RTLS algorithm from [120].

Algorithm 7 Inverse Iteration Laplace RTLS (IIL-RTLS)

Input: D, μ_0, Φ_p
Output: μ, Σ

$\mu \leftarrow \mu_0$
$\Sigma \leftarrow \hat{\Sigma}_{\mathcal{B}}(\mu_0)$
$P \leftarrow \Phi_p^{-1}$ ▷ Initialization
for $k \leftarrow 1$ to N **do**
    $P \leftarrow f(d_i)$ ▷ Sample incorporation
    $V \leftarrow P[\mu^T \ -1]^T$
    $\mu \leftarrow V_{12}/v_{22}$ ▷ Modus calculation
    $\Sigma \leftarrow \hat{\Sigma}_{\mathcal{B}}(\mu)$ ▷ Laplace covariance
end for

In the algorithm 7, v_{22} denotes the last element of $V = [V_{12}^T \, v_{22}]^T$. A new sample is incorporated by

$$f(x_i) = P - \frac{\frac{1}{2\sigma_n^2} P(x_i x_i^T) P}{1 + \frac{1}{2\sigma_n^2}\left(x_i^T P x_i\right)}. \tag{5.37}$$

In the case of a normal prior, it is possible to use iterative convergence of posterior in time in combination with the iterative convergence of the optimization algorithm seeking the mode of $p(\theta|D)$. The proposed algorithms are inspired by the

Levenberg-Marquardt optimization algorithm [149, 150, 151] for both the Laplace and variational approximations.

Algorithm 8 Recursive Laplace TLS (L-RTLS)

Input: D, Σ_p, μ_p
Output: μ, Σ

$\mu \leftarrow \mu_p$
$\Sigma \leftarrow \hat{\Sigma}_{\mathcal{N}}(\mu_p)$
$\lambda \leftarrow \lambda_0$
$\Phi \leftarrow \mathbf{0}$ ▷ Initialization
for $k \leftarrow 1$ to N **do**
    $\Phi \leftarrow \Phi + (2\sigma_n^2)^{-1} d_i d_i^T$ ▷ Sample incorporation
    for $i \leftarrow 1$ to $i_{\max}$ **do**
        $J \leftarrow -p(X|\theta)\, \mathcal{G}(\mu|\mu_p, \Sigma_p)$
        $\hat{\mu} \leftarrow \mu - (H_\mu + \lambda \mathbf{I})^{-1}\, g_\mu$ ▷ Mode estimation update
        if $-p(X|\theta)\, \mathcal{G}(\hat{\mu}|\mu_p, \Sigma_p) > J$ **then** ▷ If $\hat{\mu}$ unsuitable
            $\lambda \leftarrow \min(\lambda\iota,\ \lambda_{\max})$ ▷ Increase λ
        else ▷ If $\hat{\mu}$ is suitable
            $\lambda \leftarrow \max(\lambda/\iota,\ \lambda_{\min})$ ▷ Decrease λ
            $\mu \leftarrow \hat{\mu}$ ▷ Accept mode estimation
            $\Sigma \leftarrow \hat{\Sigma}_{\mathcal{N}}(\mu)$ ▷ Update surrogate covariance
        end if
    end for
end for

Algorithm 9 Recursive VB TLS (VB-RTLS)

Input: D, Σ_p, μ_p

Output: $\hat{z}$

$d \leftarrow [\mu_p^T, \text{vec}(\Sigma_p)^T]^T$
$\lambda \leftarrow \lambda_p$
$\Phi \leftarrow \mathbf{0}$ ▷ Initialization
for $k \leftarrow 1$ to N **do**
  $\Phi \leftarrow \Phi + (2\sigma_n^2)^{-1} d_i d_i^T$ ▷ Sample incorporation
  for $i \leftarrow 1$ to $i_{\max}$ **do**
    $L \leftarrow -\mathbf{L}\left(q_{\text{VB}}(\theta|d)\right)$ ▷ Evaluation of ELBO (5.33)
    $\hat{z} \leftarrow d - \left(H_{\mathbf{L}} + \lambda \mathbf{I}\right)^{-1} g_{\mathbf{L}}$ ▷ Parameter estimation update
    if $-\mathbf{L}\left(q_{\text{VB}}(\theta|\hat{z})\right) > L$ or $\Sigma(\hat{z}) \preceq 0$ **then** ▷ If $\hat{z}$ unsuitable
      $\lambda \leftarrow \min(\lambda\iota,\ \lambda_{\max})$ ▷ Enlarge λ
    else ▷ If $\hat{z}$ is suitable
      $\lambda \leftarrow \max(\lambda/\iota,\ \lambda_{\min})$ ▷ Decrease λ
      $z \leftarrow \hat{z}$ ▷ Accept estimate
    end if
  end for
end for

In conjunction with rejecting unsuitable estimands, the variable damping parameter λ ensures continuous progress in optimization and guarantees the positive definiteness of the surrogate covariance matrix $\Sigma \succ 0$.

The parameter $\lambda > 0$ is the damping factor adopted from the Levenberg-Marquardt optimization algorithm. The scaling parameter $\iota > 1$ controls the magnitude of the change in the scaling parameter. The damping factor is bounded by limits λ_{min} and λ_{max}, which should be set according to the bit width of the data type used for computation.

The learning gradient and Hessian are

$$g_\mu := -\tfrac{\partial}{\partial \theta}\Big(p(X|\theta)p_{\mathcal{G}}(\theta)\Big) \tag{5.38}$$

$$H_\mu := -\tfrac{\partial^2}{\partial \theta^2}\Big(p(X|\theta)p_{\mathcal{G}}(\theta)\Big). \tag{5.39}$$

In the algorithm 9, the learning gradient and Hessian are

$$g_{\mathbf{L}} := -\tfrac{\partial}{\partial d}\Big(\mathbf{L}\left(q_{\text{VB}}(\theta|\hat{z})\right)\Big), \tag{5.40}$$

$$H_{\mathbf{L}} := -\tfrac{\partial^2}{\partial d^2}\Big(\mathbf{L}\left(q_{\text{VB}}(\theta|\hat{z})\right)\Big). \tag{5.41}$$

While the calculation of the gradients and Hessians is straightforward, the author was unable to formulate them in a manner that is not needlessly long for presentation here.

All learning parameters are implemented with the inversion of the covariance matrix, ensuring precision. This parameterization leads to a numerically more stable algorithm.

6 Numerical Analysis

This chapter presents a numerical analysis of the proposed algorithms to demonstrate their correctness and fast convergence to the actual values. We provide thorough simulations, highlighting the performance and accuracy of each algorithm. While the thesis up to this point has been structured to derive the most general concepts first, followed by simpler special and derived cases, the order of this chapter is reversed. We will present the analyses, starting with the most straightforward algorithm and progressing to the most complex one.

6.1 Comparison of Approximation Methods

In chapter 5, two approximations of the posterior distribution were proposed. To recapitulate, the posterior density for normal prior is in the equation (5.19) presented as

$$p(\theta|D)_{\mathcal{N}} \propto \mathcal{L}(\theta|D)p_{\mathcal{N}}(\theta) \propto \exp\left(\frac{\vartheta^T\left(\frac{-D^TD}{2\sigma_n^2}\right)\vartheta}{\vartheta^T\vartheta}\right)\mathcal{N}(\theta|\mu_p, \Sigma_p), \tag{6.1}$$

where D is dataset containing the matrix, σ^2 is the variance of the noise, and μ_p and Σ_p are prior mean vector and covariance matrix.

For this posterior, two ways of obtaining a surrogate posterior were presented. Those are the Laplace approximation $q_L(\theta|D)$ and the variational Bayes approximation $q_{\text{VB}}(\theta|D)$. The presented approximation methods have their advantages and disadvantages. While the Laplace approximation can be used for normal and Bingham priors, the variational approximation only applies to a normal prior. The Laplace approximation is significantly simpler compared to the variational approximation, as it only requires knowledge of the mode of the approximated distribution. However, the variational approximation provides an optimal estimate.

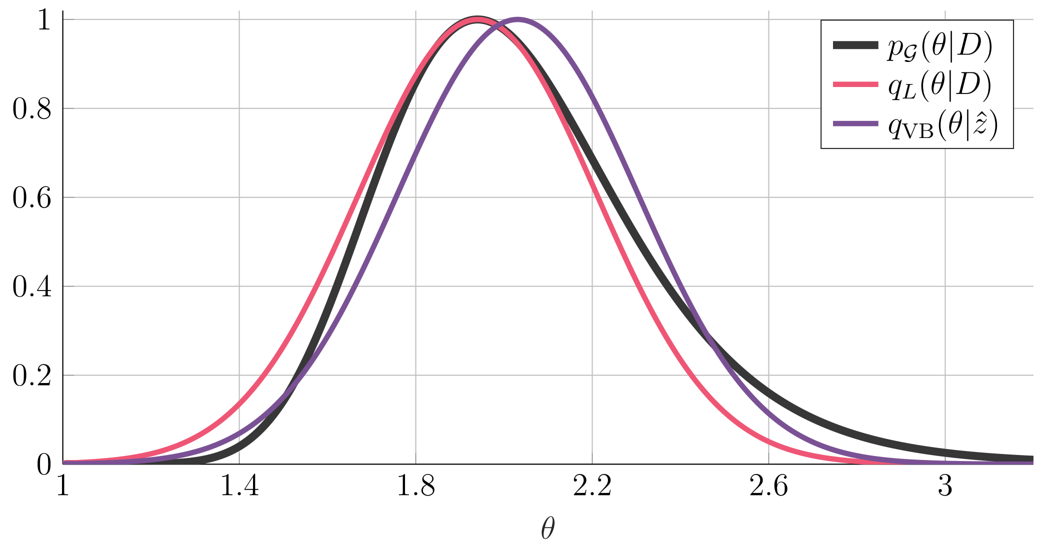


Fig. 6.1: The visual comparison of the posterior $p_{\mathcal{N}}(\theta|D)$, the Laplace surrogate posterior $q_{\mathrm{L}}(\theta|\hat{\mu})$, and VB surrogate posterior $q_{\mathrm{VB}}(\theta|z^*)$ for $N = 50$ samples and noise covariance $0.5\mathbf{I}$, all normalized such that their maxima are equal to one.

Figure 6.1 shows a one-dimensional comparison of the Laplace and variational approximations for a normal prior. It is evident that while the Laplace approximation is valid in the vicinity of the mode of the distribution, the variational approximation accounts for the skewness of the original distribution by shifting the mean of the surrogate distribution.

Similar results can also be observed for B-GTLS and Bayesian PMLSM GTLS presented in chapter 4. To recapitulate, the posterior distributions (4.31) and (3.21) are

$$p(\theta|D) \propto \exp\left(\frac{\vartheta^T\left(-\frac{1}{2}D^TD + \Phi_p\right)\vartheta}{\vartheta^T\Sigma\vartheta}\right) \tag{6.2}$$

and

$$p\left(\theta|D_1, D_2\right) \propto \mathcal{L}\left(\theta|D_1\right)\mathcal{L}\left(\theta|D_2\right)p(\theta), \tag{6.3}$$

respectively, where

$$\mathcal{L}\left(\theta|D_j\right) \propto \exp\left(\frac{\vartheta^T\left(-\frac{1}{2}D_j^TD_j\right)\vartheta}{\vartheta^T\Sigma_j\vartheta}\right) \tag{6.4}$$

The Laplace surrogate posterior for B-GTLS is proposed in Lemma 4 as

$$p(\theta|D) \approx q(\theta|D) = \mathcal{N}\left(\hat{\theta}, \hat{\Sigma}(\hat{\theta})\right). \tag{6.5}$$

Having a surrogate posterior for the B-GTLS problem, the Bayesian PMSM GTLS problem is approximated in (4.74) as a product of likelihood approximates

$$p\left(\theta|D_1, D_2\right) \approx q\left(\theta|D_1, D_2\right) = q_1(\theta|D_1)q_2(\theta|D_2)p(\theta), \tag{6.6}$$

where $q_j(\theta|D_j)$ approximates the likelihood as (6.5) under assumption $\Phi_p = \mathbf{0}$.

In order to provide a visual example of the posterior distributions, likelihoods and their respective surrogates, the PMLSM problem is simplified to two EIV problems

$$0 = \vartheta^T \overline{d}_1, \tag{6.7}$$
$$0 = \vartheta^T \overline{d}_1, \tag{6.8}$$

where the measured datasets D_1 and D_2 consists of measurements

$$d_{1,i} = \overline{d}_{1,i} + \varepsilon_{1,i}, \quad \varepsilon \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 2 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}\right), \tag{6.9}$$
$$d_{2,i} = \overline{d}_{2,i} + \varepsilon_{2,i}, \quad \varepsilon \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 0.5 & 0.002 \\ 0.002 & 0.001 \end{bmatrix}\right). \tag{6.10}$$

Assuming $vartheta = [1.5, \, -1]^T$, $N = 50$ samples of independent variables are drawn from $\mathcal{N}\left(\mathbf{0}, 10\right)$. The resulting posterior distribution and likelihood surrogates can be studied in the following figure.

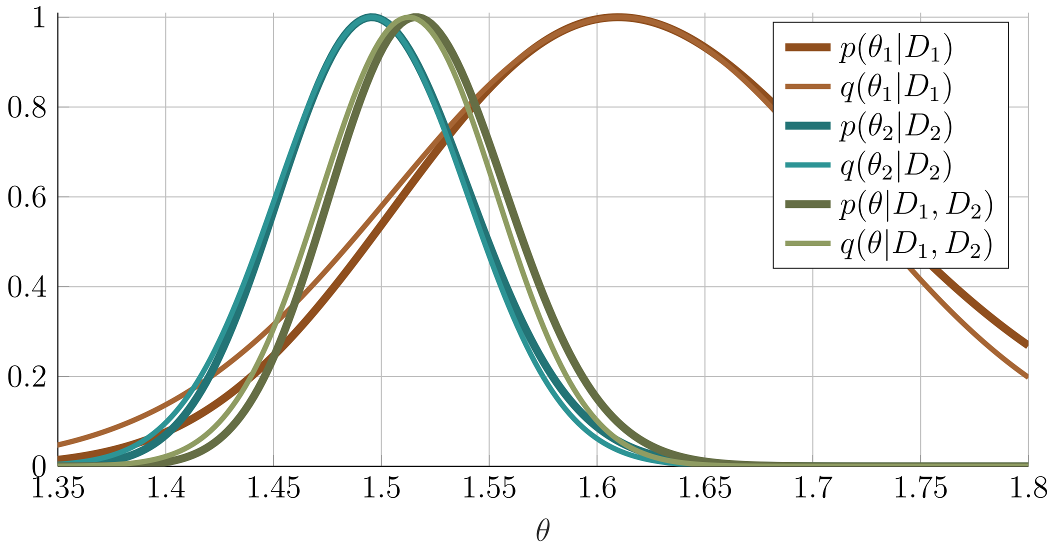


Fig. 6.2: The visual comparison of the posteriors $p_{\mathcal{N}}(\theta_1|D_1)$, $p_{\mathcal{N}}(\theta_2|D_2)$, $p_{\mathcal{N}}(\theta|D_1, D_2)$ and the corresponding Laplace surrogate posteriors $q_{\mathcal{N}}(\theta_1|D_1)$, $q_{\mathcal{N}}(\theta_2|D_2)$, $q_{\mathcal{N}}(\theta|D_1, D_2)$ for $N = 50$ samples, all normalized such that their maxima are equal to one.

In the picture 6.2, we can observe that individual approximations are comparable to the Laplace approximation in the B-TLS case and that the approximation $p(\theta|D_1, D_2) \approx q(\theta|D_1, D_2)$ does successfully recover modus and variance of the true posterior distribution.

6.2 Numerical analysis of B-TLS-based algorithms

The proposed algorithms based on B-TLS are subjected to numerical analysis using MATLAB simulations. The purpose of this analysis is to verify the functionality of the algorithms, particularly their convergence to the results obtained by computationally much more intensive optimization algorithms implemented within MATLAB's optimization toolbox. The comparison is sourced from the author's previously published research in [1].

The presented experiments are conducted with the following parameters.

$$
\begin{aligned}
&\mu_p = 5\\
&\Sigma_p = \sigma_p^2 = 100\\
&\mu_0 = \mu_p\\
&z_0 = [\mu_p, \Sigma_p]^T\\
&\Phi_p = \mathrm{diag}([10, 10])\\
&\lambda_{min} = 10^{-10}\\
&\lambda_{max} = 10^{10}\\
&\lambda_0 = \lambda_{min}\\
&\iota = 2
\end{aligned}
$$

Tab. 6.1: B-TLS numerical analysis simulation parameters.

However, the results are valid for a wide range of suitable parameters. As expected, selecting unreasonable parameters, such as $\iota = 0$, leads to failure.

The simulations were performed with synthetic data generated by sampling from a problem with errors in variables.

$$d_i = [x_i \ \overline{\theta} x_i]^T + \mathcal{N}(\varepsilon|\mathbf{0}, \sigma_n \mathbf{I}), \tag{6.11}$$

where x_i is sampled from $\mathcal{N}(x_i|\mathbf{0}, 1)$ and $\overline{\theta} = 1.8$. This resulted in a dataset D with $N = 500$ samples. Each experiment is repeated a thousand times, and their variance and average convergence values are presented along with the first conducted experiment.

In the case of the Bingham prior, only one algorithm is available, namely Algorithm 7. This algorithm performs recursive identification, and its result $q_{\mathrm{L}}(\theta|\mu)$ is compared with the alternative posterior $q_{\mathrm{L}}(\theta|\mu^*)$, where the mode μ^* is calculated using the TLS algorithm based on singular value decomposition.

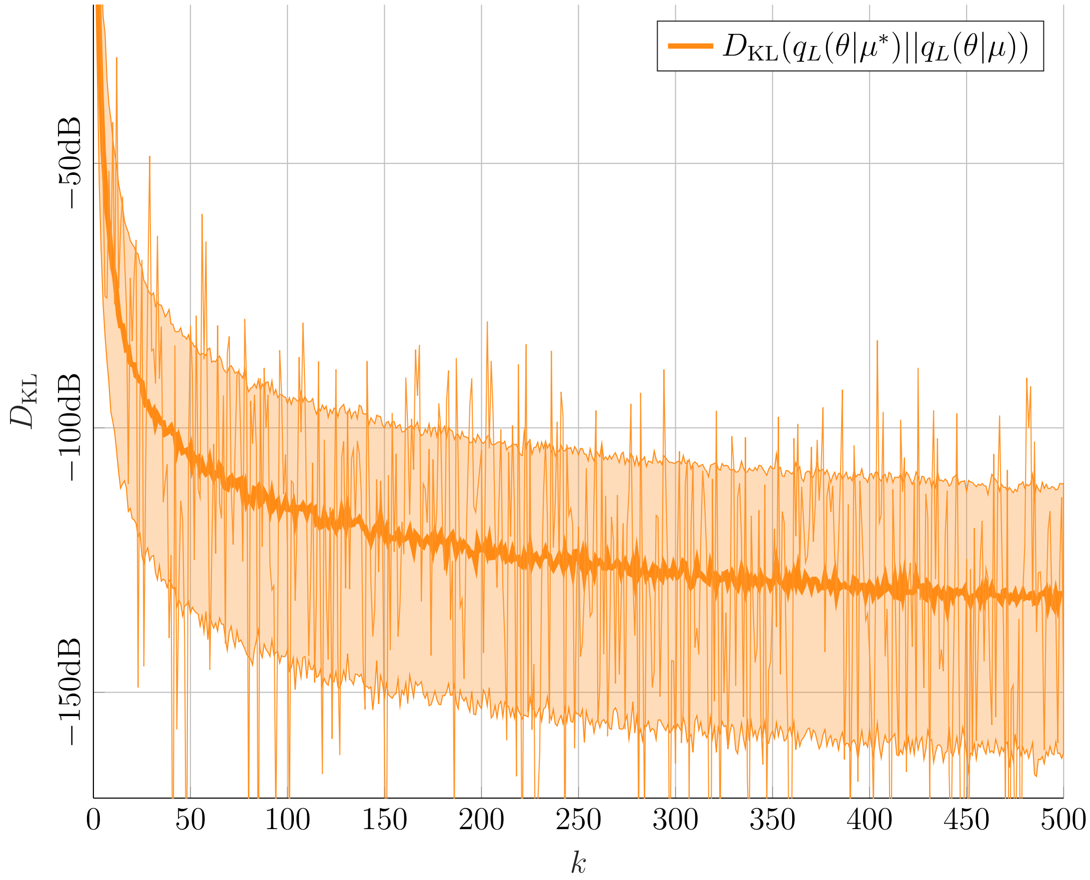


Fig. 6.3: The Kullback-Liebler divergence from the true MAP-centered μ^* Laplace estimate to the estimate calculated by Algorithm 7. The mean and variance over 1,000 runs are represented with the bold line and transparent area, respectively, and the D_{KL} of a single run is denoted by the thin line.

From Figure 6.3, it is evident that despite the poor initial estimate, the result of Algorithm 7 converges to the surrogate posterior $q_{\mathrm{L}}(\theta|\mu^*)$.

In the case of the normal prior, it is possible to use Algorithms 8 and 9. Both algorithms are compared with the optimal variational estimate $q_{\mathrm{VB}}(\theta|z^*)$. This estimate was obtained by a computationally intensive numerical solution of the optimization problem (5.35) at each step.

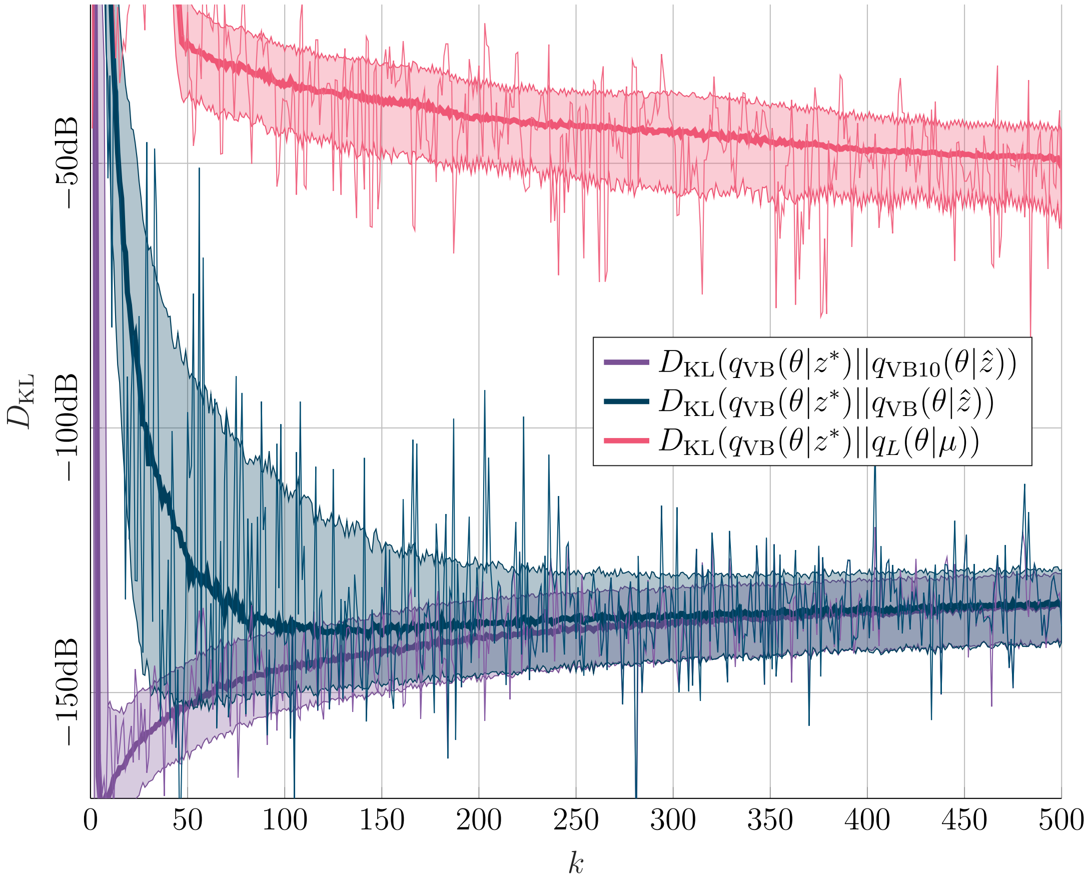


Fig. 6.4: The Kullback-Liebler divergence from the optimal posterior estimate $q_{\mathrm{VB}}(\theta|z^*)$ to the various estimation methods. The mean and variance over 1,000 runs are represented with the bold line and transparent area, respectively; the D_{KL} of a single run is denoted by the thin line.

From Figure 6.4, it is evident that the results of both algorithms are numerically very close. The Laplace approximation is computationally significantly less demanding, but it quickly converges to a suboptimal solution $q_{\mathrm{L}}(\theta|\mu^*) \neq q_{\mathrm{VB}}(\theta|z^*)$. The variational approximation converges more slowly but to an optimal solution.

For both algorithms, it is possible to choose the number of optimization steps performed for each sample using the parameter i_{max}. Simulations show that single optimization step $i_{\mathrm{max}} = 1$ is sufficient for convergence. The progression of $q_{\mathrm{VB10}}(\theta|d)$ indicates that increasing the number of optimization steps, in this case to $i_{\mathrm{max}} = 10$, results in significantly faster convergence. However, this effect is minimal with a sufficient number of samples.

6.3 Numerical analysis of B-GTLS-based algorithms

Both B-GTLS-based algorithms were also evaluated through numerical analysis using MATLAB simulations. This analysis aimed to confirm the effectiveness of the algorithms, specifically their ability to converge to solutions comparable to those produced by more computationally demanding optimization algorithms found in MATLAB's optimization toolbox.

The numerical experiment is executed with the following arbitrarily selected parameters:

$$\mu_p = 5, \quad \Sigma_1 = \begin{bmatrix} 2.0 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}, \quad \text{and} \quad \Sigma_2 = \begin{bmatrix} 0.500 & 0.002 \\ 0.002 & 0.001 \end{bmatrix}. \tag{6.12}$$

However, testing different initial conditions did not cause significant changes in the results.

The simulations were performed with synthetic data generated by sampling from a problem with errors in variables.

$$d_{i,j} = [x_{i,j} \; \overline{\theta} x_{i,j}]^T + \mathcal{N}(\varepsilon|\mathbf{0}, \Sigma_j), \tag{6.13}$$

where x_i is sampled from $\mathcal{N}(x_i|0, 10)$ and $\overline{\theta} = 1.8$. This resulted in two datasets $D_j, \quad j \in \{1, 2\}$ with $N = 500$ samples. In each iteration, surrogate posterior $q(\theta_1|D_1)$ produced by algorithm 4.6 is compared with ideal Laplace approximate $q(\theta_1|\mu_1^*)$ calculated from the true MAP μ_1^* obtained by computationally demanding MATLAB's optimization toolbox calculation. This experiment is repeated a thousand times, and Kullback-Liebler divergence is used to compare the two surrogate distributions. Variance and average convergence values are presented.

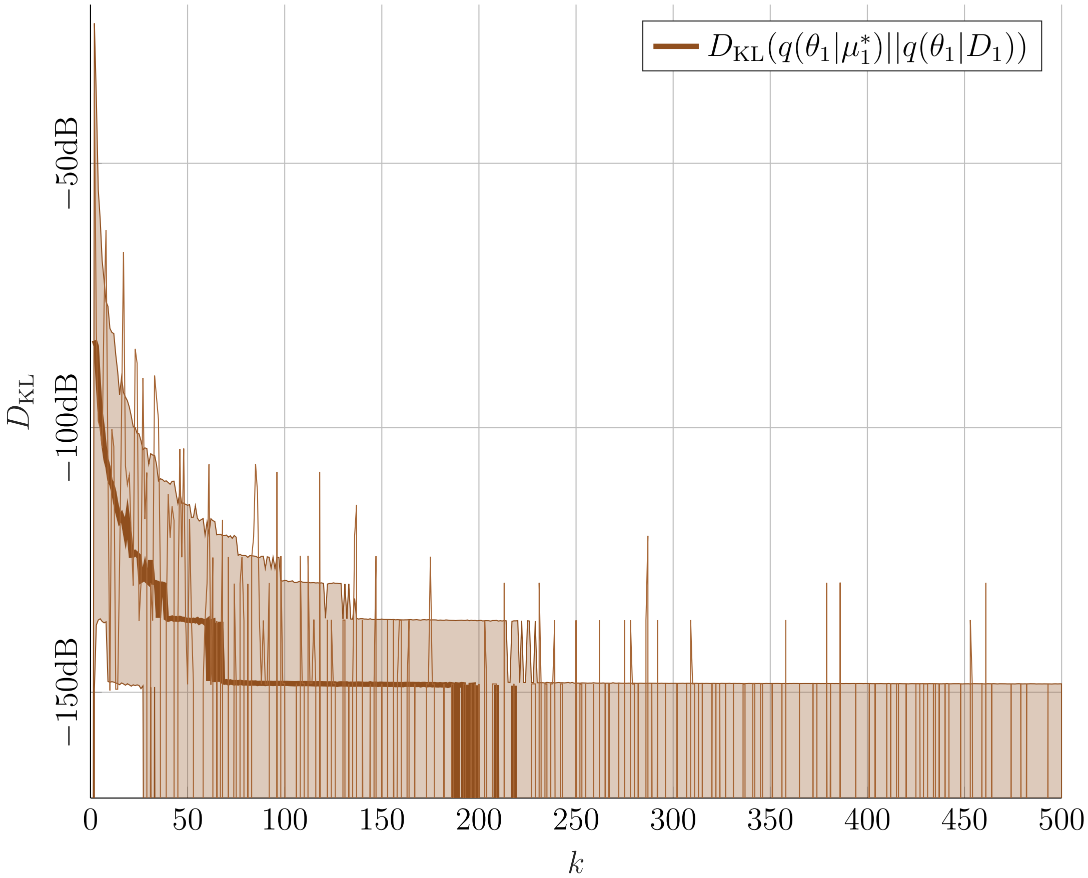


Fig. 6.5: The Kullback-Liebler divergence from the true MAP-centered Laplace estimate to the estimate calculated by Algorithm 4.6. The mean and variance over 1,000 runs are represented with the bold line and transparent area, respectively, and the D_{KL} of a single run is denoted by the thin line.

In figure 6.5, fast convergence of the surrogate posteriors is exhibited. The approximation $q(\theta_1|D_1)$ is after 25 samples so close to the $q(\theta_1|\mu_1^*)$, that small number quantization of D_{KL} is apparent.

The Bayesian PMLSM GTLS algorithm design is based on two ideas. The first one is an approximation of two likelihoods with different noise covariance Σ_j. The second idea is the transformation of true measurements into synthesized measurement vectors. To separate the influence of the transformation, let us look at the convergence of the Kullback-Liebler divergence from the ideal surrogate posterior $q(\theta|\mu^*)$ to surrogate posterior $q(\theta|D) = q(\theta_1|D_1)q(\theta_2|D_2)p(\theta)$ based on (6.6). The ideal surrogate is calculated from the true MAP μ^*, obtained through a computationally demanding optimization process using MATLAB's optimization toolbox. The prior is uninformative $p(\theta) \propto 1$ to simplify the experiment and amplify influence of the approximations.

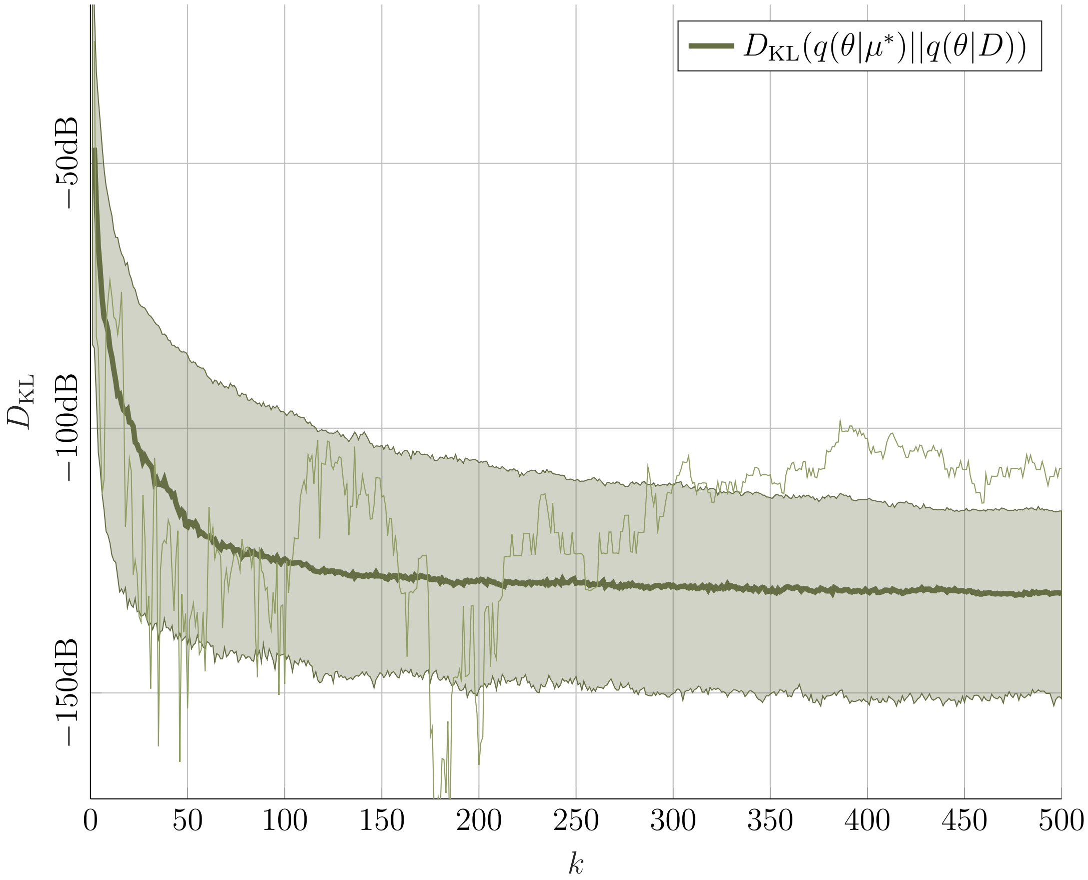


Fig. 6.6: The Kullback-Liebler divergence from the true MAP-centered Laplace estimate to the estimate calculated by (6.6). The mean and variance over 1,000 runs are represented with the bold line and transparent area, respectively, and the D_{KL} of a single run is denoted by the thin line.

As apparent from the figure 6.6, the surrogates converge on average after a few samples.

Furthermore, relative error

$$\delta_x = \left|\frac{x - x^*}{x^*}\right| 100 \tag{6.14}$$

of the mean and variance can be analyzed.

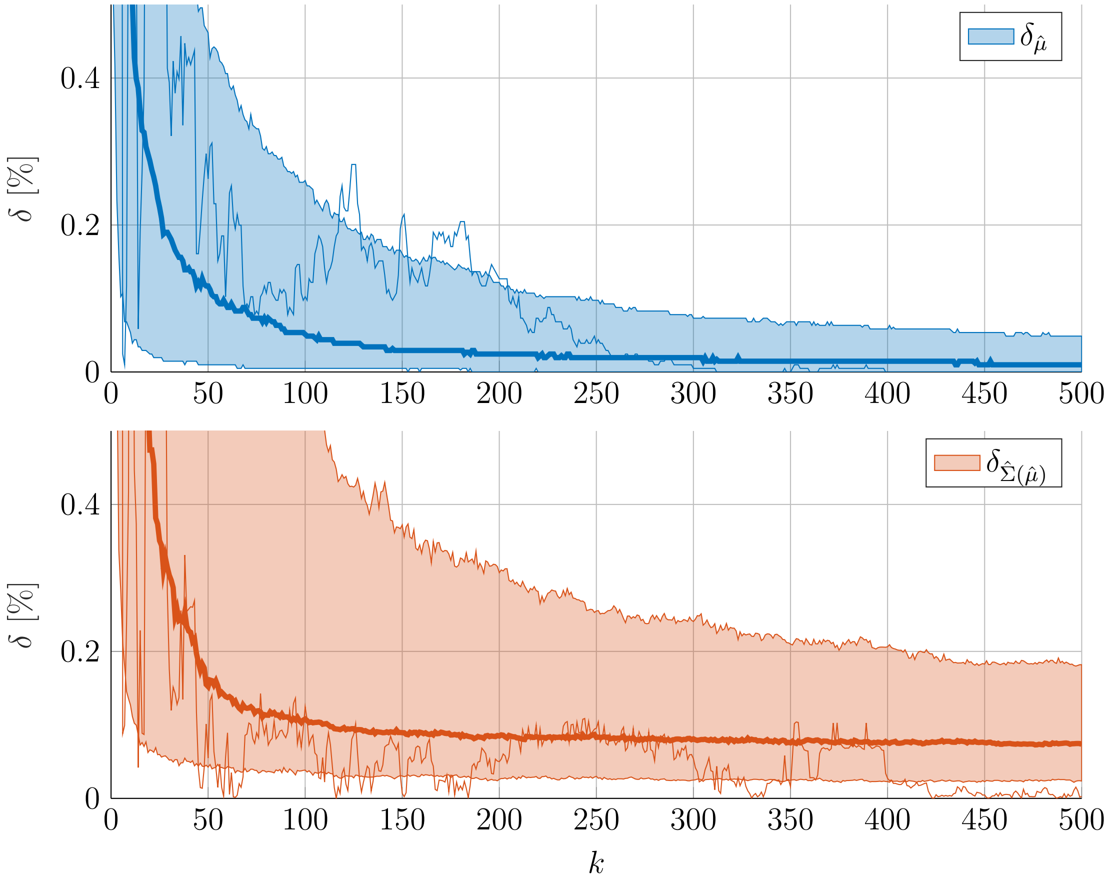


Fig. 6.7: Relative error of the MAP approximate (top) and covariance (bottom) in the example. The mean and variance over 1,000 runs are represented with the bold line and transparent area, respectively, and the D_{KL} of a single run is denoted by the thin line.

Figure 6.7 shows that, on average, the relative error of both the mean and the variance reduces.

6.4 Forgetting Factor

All proposed algorithms allow for the identification of systems with slowly varying parameters due to the incorporation of the forgetting factor λ. This factor is analogous to the forgetting factor in the well-known LS algorithms with forgetting, which are designed to give more weight to recent observations and less to older ones. The forgetting factor, therefore, requires no further explanation.

A numerical demonstration of the forgetting factor is presented on B-RGTLS (Algorithm 4.6). The data is generated with true parameters $\theta = [3,\, 0.5,\, 8]^T$, which are after $K_1 = 5000$ samples switched for $\theta = [-2,\, 3,\, 10]^T$. Additional $K_2 = 5000$ samples are obtained with new parameters. Independent variables are generated by sampling from the multivariate normal distribution $\mathcal{N}(0, 3\mathbf{I})$. All true data are corrupted with additional normal noise with zero mean and covariance

$$\Sigma = \begin{bmatrix} 1 & 0.4 & 0 & 0.2 \\ 0.4 & 1.5 & 0 & 0 \\ 0 & 0 & 0.8 & 0.5 \\ 0.2 & 0 & 0.5 & 2 \end{bmatrix}. \tag{6.15}$$

The experiment is repeated two times, one time with no forgetting $\lambda_1 = 1$ and the second time with $\lambda_2 = 0.999$. The results are presented in the following picture.

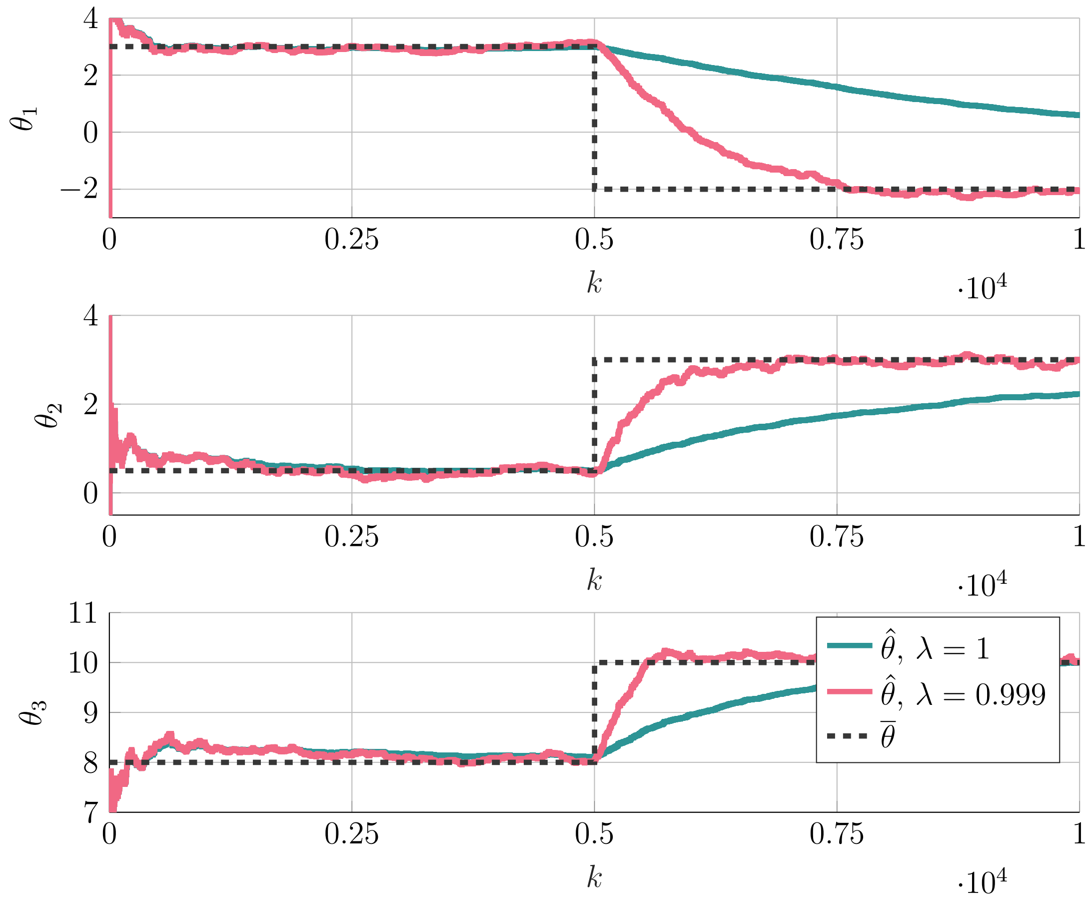


Fig. 6.8: Comparison of identification of system parameters with and without forgetting factor active.

Figure 6.8 demonstrates advantages stemming from forgetting factor utilization in the case of time-variant parameters. After the abrupt change of system parameters, an algorithm with forgetting shows a rapid decrease in the error because older, less relevant data has less influence on the current estimate. In the case without forgetting, the error decreases slowly after the parameter change, as the algorithm considers the entire history of data, making the algorithm less responsive to sudden changes.

7 Examples

Proposed algorithms and frameworks apply to many applications, not excluding systems with large dimensions, slowly variable parameters, zero variance measurements, under-determined systems, numerically unstable cases, or applications requiring real-time implementation. This chapter provides a handful of examples using the derived frameworks to highlight the contributions of methods proposed in this thesis. Although the primary research is focused on the application of PMLSM identification, the examples also show the application of PMSM and a total capacity of the battery cell identification.

7.1 Quality of Identification using B-TLS

One of the main advantages of the B-TLS compared to B-GTLS is the simpler posterior distribution in the form of the Bingham distribution. This allows, for example, improved posterior approximations, graphical demonstration of parameter uncertainty, and sampling from the parameter density function. The latter of those will be demonstrated in this simple simulation example.

The selection of proper identification inputs for identifying PMSM parameters is demonstrated on example previously presented in author's conference paper [2]. To keep the example simple, not all parameters are to be identified. |Identification of the full PMLSM model is the goal of another examples. While the number of parameters for B-TLS is not limited, a small number of identified parameters allows for a graphical demonstration of the parameter posterior distribution. Utilizing discretized model from [41] or [152]. Assuming simplifying conditions of zero mechanical angular velocity $\omega_m = 0$ and zero direct current $i_d = 0$, the equation for $i_q(k+1)$ becomes linear difference equation

$$i_q(k+1) = \theta_1 i_q(k) + \theta_2 i_d(k) + \theta_3 \omega_m + \theta_4 u_q(k) = \theta_1 i_q(k) + \theta_4 u_q(k). \tag{7.1}$$

Linear equations are identifiable using the B-TLS algorithm.

From the definition of the identified model, it is evident that all variables are subject to noise caused by either the measurement $i_q(k)$ or the signal generation u_q. Let us further assume that the signals are scaled such that the noise of each variable has identical and known precision λ_n.

The extended vector of estimands is in the form $\vartheta = [\theta_4, \theta_1, -1]^T$.

The advantage of using Bayesian inference compared to maximum likelihood methods is obtaining information about the identification quality. The quality of identification directly depends on the choice of the identification signal, in this case, the input voltage u_q. This demonstration aims to use the derived Bayesian inference to estimate the parameters θ_4 and θ_1 from (7.1) using two input signals and subsequently compare the identification quality in both cases.

The chosen input signals are a pseudorandom binary sequence (PRBS) and a pseudorandom Gaussian sequence (PRGS). Segments of the simulated signals are shown in Figure 7.1.

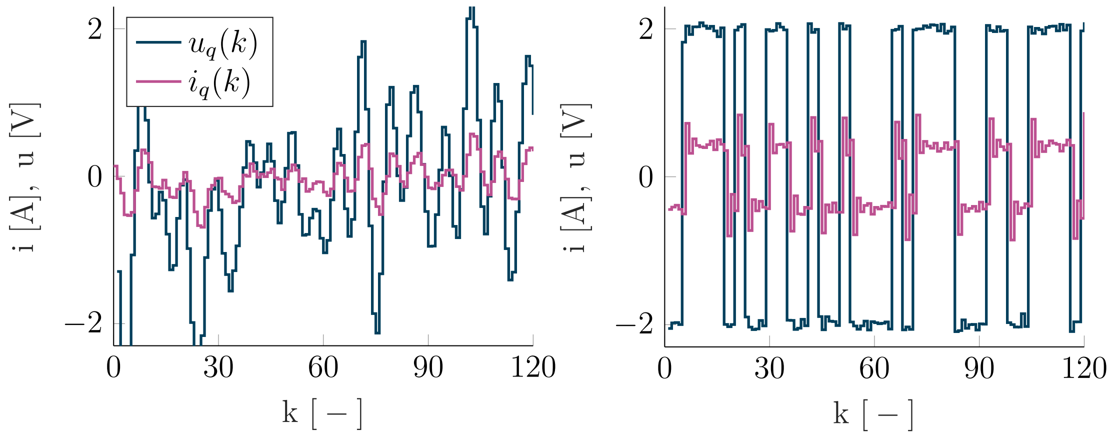


Fig. 7.1: Comparison of simulated data for PRGS (left) and PRBS (right).

The simulation allows for a priori knowledge of the correct parameters

$$\overline{\theta} = [0.3 \; -0.4] \, . \tag{7.2}$$

Assuming uninformative prior $p(\theta) \propto 1$, both input signals result in the maximum a posteriori solution with comparable deviations from $\overline{\theta}$ after a sample size of $N = 1000$.

$$\begin{aligned} \hat{\theta}_{MAP,\mathrm{PRGS}} &= [0.2988 \; -0.39589]^T \, , \\ \hat{\theta}_{MAP,\mathrm{PRBS}} &= [0.29921 \; -0.39926]^T \, . \end{aligned} \tag{7.3}$$

A small deviation in both cases could lead to the misleading conclusion that the identification quality is comparable for both signals. However, this conclusion would be incorrect, as is evident from the posterior probability distribution

$$p(\theta|D) \propto \mathcal{B}(v|\Gamma\Lambda\Gamma^T), \tag{7.4}$$

with the following concentration matrices for the signals PRBS and PRGS.

$$\begin{aligned} \Lambda_{\mathrm{PRGS}} &\doteq diag\,(1092.0,\, 10.0,\, 0)\, , \\ \Lambda_{\mathrm{PRBS}} &\doteq diag\,(4330.5,\, 168.6,\, 0)\, , \end{aligned} \tag{7.5}$$

and the vectors of the mean direction axes

$$\gamma_{3,\mathrm{PRGS}} = [-0.2678,\, 0.3557,\, 0.8954]^T\,,$$
$$\gamma_{3,\mathrm{PRBS}} = [-0.2690,\, 0.3664,\, 0.8907]^T\,. \tag{7.6}$$

As follows from the construction of the prior probability distribution, the vectors of the mean direction axes results, after scaling $\hat{\theta}_{\mathrm{MAP}} = -\frac{[\gamma_{3,1},\, \gamma_{3,2}]^T}{\gamma_{3,3}}$ from Theorem 2, identical to the maximum a posteriori solution (7.3).

The high values of concentration parameters Λ_{PRGS} compared to Λ_{PRBS} indicate greater uncertainty. Therefore, the PRGS signal achieves lower identification quality compared to PRBS. This fact is also evident from the visualization of results in Figure 7.2, where the more distant contours indicate greater uncertainty in the case of PRGS compared to PRBS. The lower identification quality is also apparent from the greater variability of parameters sampled from the posterior distribution of PRGS compared to PRBS.

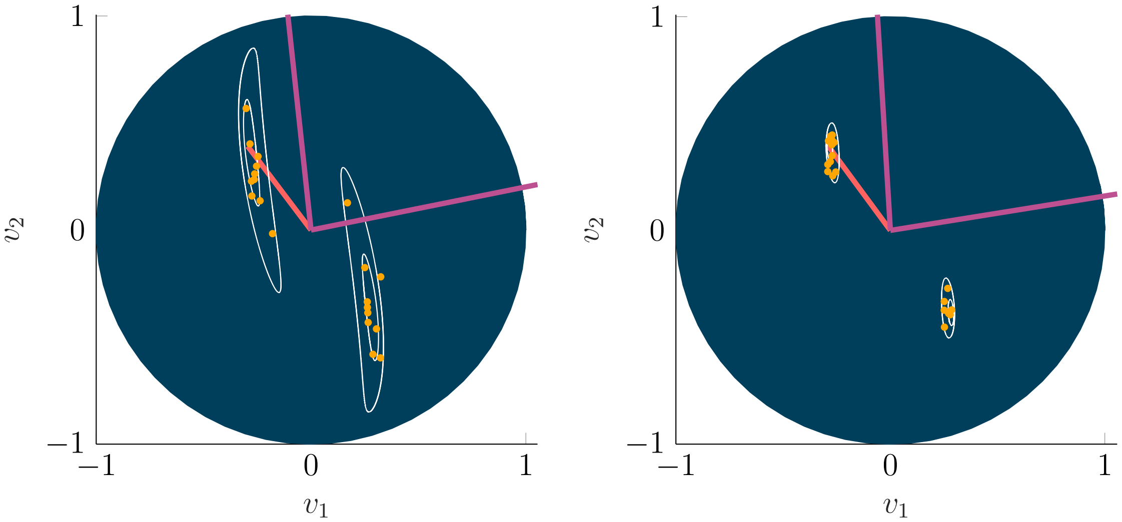


Fig. 7.2: Orthographic projection of the contour plot of the unnormalized posterior probability distribution for identification signals PRGS (left) and PRBS (right). The mean direction axis is marked in red, and the concentration axes are marked in purple. Sampled parameters are shown in yellow. The better identification quality in the case of PRBS is evident from the lower variability of the sampled parameters and the tighter contours.

This result can be further presented by taking 20 parameter samples $\hat{\theta}$ from the posterior probability distribution. Step transition characteristics are plotted for these parameters. The higher variance of parameter samples in the case of

PRGS results in a higher variance of the step transition characteristics, as shown in Figure 7.3.

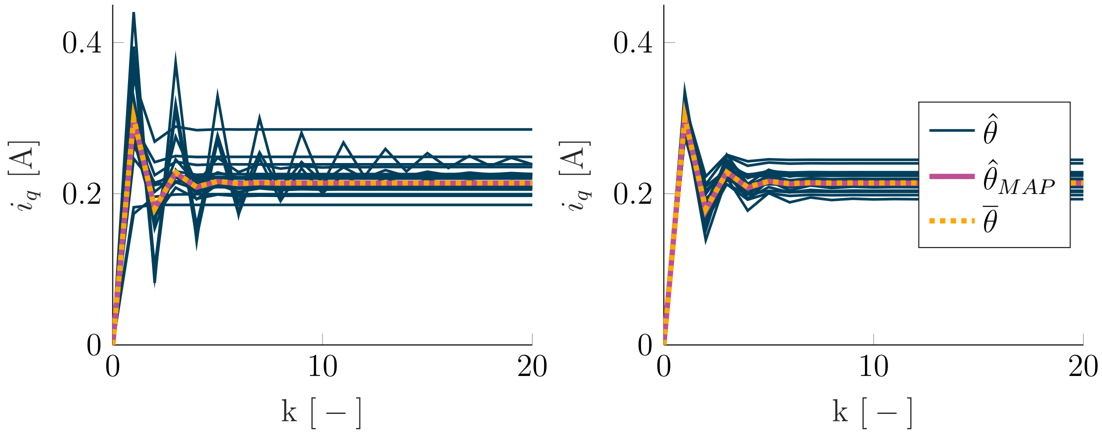


Fig. 7.3: Step transition characteristics of the simulated system using the true parameters $\overline{\theta}$, maximum a posteriori parameters $\hat{\theta}_{MAP}$, and parameters $\hat{\theta}$ sampled from the posterior probability distribution for PRBS signals (left) and PRGS signals (right). The better identification quality for PRBS is evident from the lower variability of the characteristics.

7.2 Quality of Identification using B-TLS in practice

Despite the simulation demonstrating the method's functionality, it is appropriate to test it in practice. This has already been published in author's previous research [2]. The simulated example from Section 7.1 is a potential practical application. Although the identified system is a rotary motor, given the similarities in the models described in Section 1.3, the same method applies to PMLSM.

The TGT2-0032-30-24 motor was subjected to the voltage from Chapter 7.1 with a sampling period of $T_s = 10\,ms$, and the current values were recorded using an NI cRIO-9038. The conditions $\omega_m = 0$ and $i_d = 0$ were ensured by locking the rotor and controlling the i_d current using a PID controller implemented in the cRIO-9038. Figure 7.2 shows a segment of the measured values.

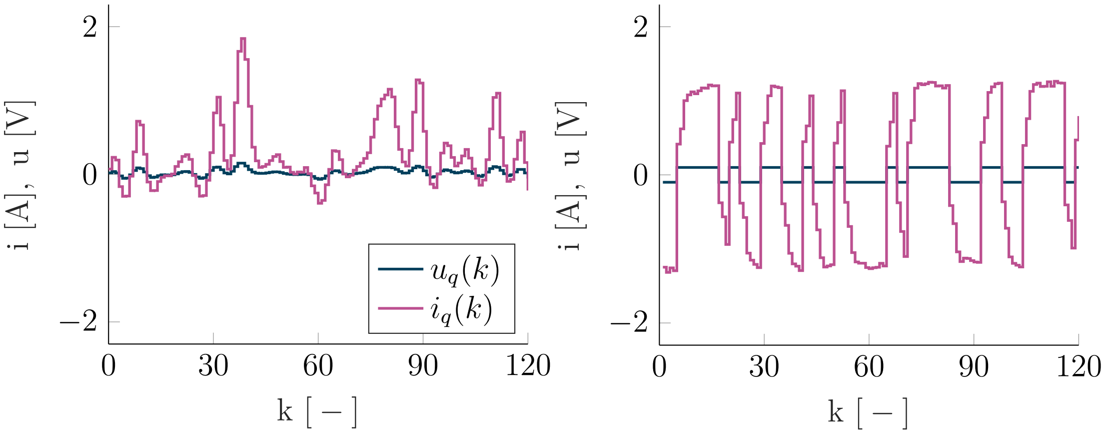


Fig. 7.4: Comparison of measured data for PRGS (left) and PRBS (right).

In this case, the real constants $\overline{\theta}$ cannot be known, so the convergence of the maximum a posteriori method to the correct value cannot be proven. However, by definition, it is identical to TLS, whose convergence is proven. What can be verified through practical measurement is the conclusion that the PRGS signal achieves lower identification quality than PRBS.

Another obstacle compared to the simulation is the unknown precision, or variance, of the process noise λ_n. However, this can be estimated from the measured data as $\overline{\lambda}_n \approx 16 \cdot 10^3$.

It is now possible to proceed identically to Chapter 7.1. The posterior probability distribution has the following concentration matrices:

$$\begin{aligned} \Lambda_{\text{PRGS}} &\doteq diag\,(-2.8390,\ -0.2159,\ 0)\,10^6, \\ \Lambda_{\text{PRBS}} &\doteq diag\,(-2.0399,\ -0.2454,\ 0)\,10^6, \end{aligned} \tag{7.7}$$

and the vectors of the mean direction axes

$$\begin{aligned} \gamma_{3,\text{PRGS}} &= [-0.9557,\ 0.0899,\ 0.0228]^T\,, \\ \gamma_{3,\text{PRBS}} &= [-0.9968,\ 0.0776,\ 0.0197]^T\,. \end{aligned} \tag{7.8}$$

The higher values of concentration parameters Λ_{PRGS} compared to Λ_{PRBS} indicate more significant uncertainty, thus confirming the simulation's conclusion. The results can also be compared graphically in the following Figure.

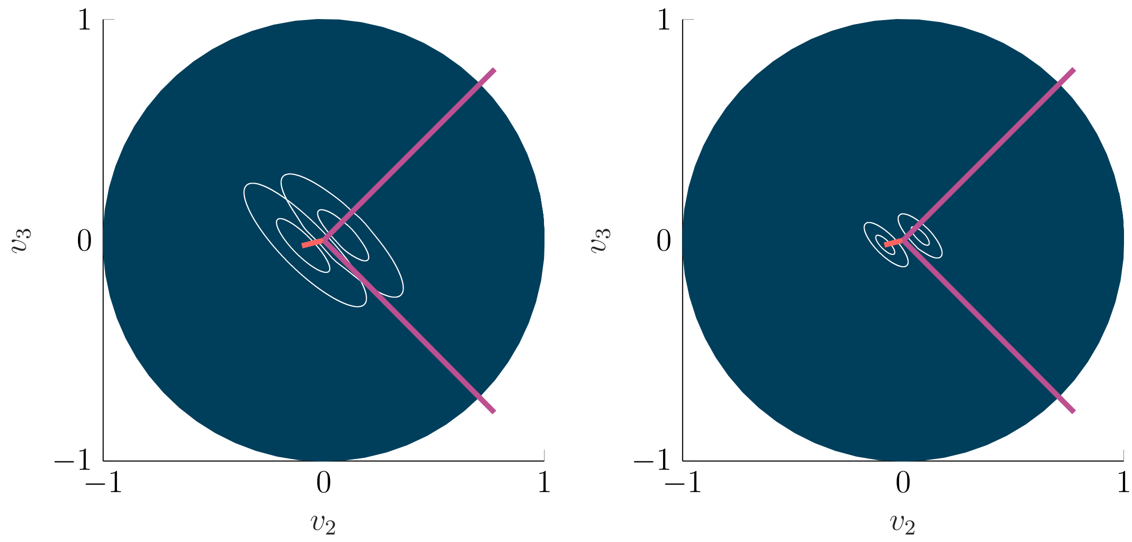


Fig. 7.5: Orthographic projection of the contour plot of the unnormalized posterior probability distribution of measurements with PRGS signals (left) and PRBS signals (right). The mean direction axis is marked in red, and the concentration axes are marked in purple. The better identification quality in the case of PRBS is evident from the lower variability of the sampled parameters and the tighter contours.

The practical example follows theoretical results from the simulation, proving applicability of the method. This concludes the example.

7.3 Approximate Recursive Identification with Confidence Intervals

In some cases, it is crucial to identify not only a point estimate of a parameter but also its confidence interval. In the Bayesian approach, this can be interpreted as the probability that the parameter lies within a given interval of the posterior distribution.

The reliability of the result can be important for several reasons. In previous chapters, it was used to evaluate the quality of identification. This approach can be extended to assess suitable identification signals to minimize confidence intervals.

Another reason might be applications where safety is the main priority. In such cases, it is necessary to consider not only the most probable parameter estimate. Often, confidence intervals are used to define the region where the actual parameter is likely to be found.

Such a case is presented in this simple experiment, previously published by autor's journal paper [1], where the previously mentioned algorithms 8 and 9 are used to identify the total capacity of a battery cell. An overview of other applications using confidence intervals can be found in the book [153].

The results of the proposed algorithms are compared with those from [154], where more information about the simulated problem can also be found.

The total capacity of the battery cell θ, measured in ampere-hours, appears in a linear model.

$$q = \Delta z \theta, \tag{7.9}$$

where $q = \int_{t_1}^{t_2} \frac{\eta i(\tau)}{3600}, d\tau$ is the measurement of accumulated ampere-hours, and $\Delta z = \frac{z(t_2)-z(t_1)}{100}$ is the measurement of the difference in the state of charge of the battery cell. Here, $z(t)$ is the percentage state of charge of the battery cell at time t, η is a dimensionless efficiency factor, and $i(t)$ denotes the current through the battery cell at time t.

Since both expressions in (7.9) needed for identification are measured quantities q and Δz, this case is a suitable example of a problem with errors in variables. A common assumption is that the noise is independent, identically distributed Gaussian noise with zero mean and proportional variances σ_q and σ_z. Due to proportionality, the input signals can be scaled so that the total least squares method is an optimal estimator of the total capacity of the battery cell θ.

Synthetic data for the problem with errors in variables are generated with the actual total capacity of the battery cell $\overline{\theta} = 10$ and noise variances $\sigma_q = \sigma_z = 0.5$. The parameters of the algorithms are the same as those given in Chapter 6.2. All algorithms are initialized with virtual correct measurements $\Delta z = 1$, $q = \overline{\theta}$.

For comparison, the problem is also identified using the least squares (LS) method and the total least squares (TLS) method, where the confidence interval is estimated by inverting the Fisher information. Motivation and details about these methods are available in [154]. The least squares method ignores measurement error in Δz.

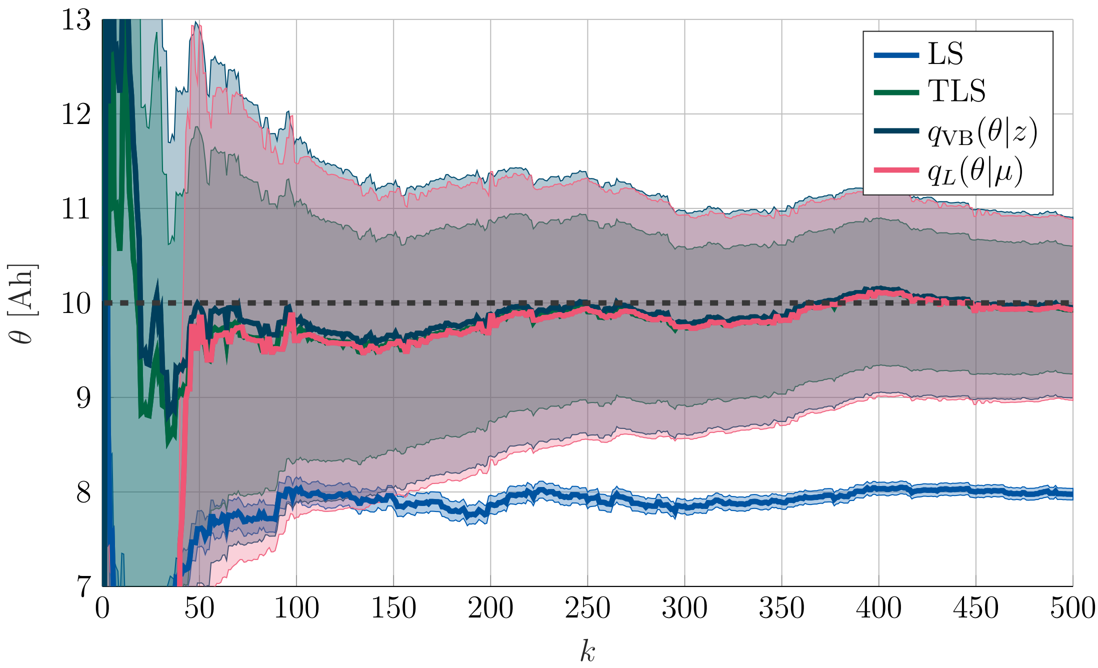


Fig. 7.6: The battery cell total capacity estimation results for the different algorithms. The mean and 3σ interval approximations are represented by the bold line and transparent area, respectively.

The results of recursive identification are shown in Figure 7.6. It is evident that the least squares method provides a biased parameter estimate and a poor estimate of the confidence interval due to ignoring the measurement error in Δz.

The average value estimates of all other algorithms are practically identical. The variational method $q_{\mathrm{VB}}(\theta|D)$ provides a slightly biased MAP estimate, which allows for an optimal estimate of the confidence interval.

The confidence interval estimate using the Laplace approximation of the posterior $q_{\mathrm{L}}(\theta|\mu)$ is very similar to the optimal confidence interval estimate, rendering the VB approximation needlessly complicated for cases with sufficient number of samples.

The confidence interval provided by the improved TLS algorithm from [154] is significantly underestimated due to the simplified derivation using the Fisher information matrix.

Therefore, the proposed algorithms provide an accurate estimate of both the mean value and the confidence interval compared to other available methods.

7.4 Recursive PMLSM Identification Simulation

After the derivation of the B-GTLS method, the application of the proposed algorithm for recursive identification of PMLSM drive is the main contribution of this thesis. The formulation of PMLSM identification as an EIV problem is presented in chapter 3.2, and the formulation for B-GTLS is presented in chapter 4.7. To recapitulate, the identification is divided into two EIV problems

$$0 = \vartheta^T \mu_1, \tag{7.10}$$

$$0 = \vartheta^T \mu_2, \tag{7.11}$$

where the extended vector of unknown parameters ϑ and the two vectors of independent synthesized measurement variables μ_1, μ_2 are

$$\vartheta = \begin{bmatrix} L_d \\ L_q \\ Ri_d \\ \Psi_{PM} \\ -1 \end{bmatrix}, \quad \mu_1 = \begin{bmatrix} \dfrac{\bar{i}_d(k) - \bar{i}_d(k-1)}{T_s} \\ -\dfrac{\pi}{\tau_s}\bar{i}_q \dfrac{\bar{l}(k) - \bar{l}(k-1)}{T_s} \\ \bar{i}_d \\ \bar{u}_d \end{bmatrix}, \quad \mu_2 = \begin{bmatrix} \frac{\pi}{\tau_s}\bar{i}_d \frac{\bar{l}(k)-\bar{l}(k-1)}{T_s} \\ \dfrac{\bar{i}_q(k) - \bar{i}_q(k-1)}{T_s} \\ \bar{i}_q \\ \frac{\pi}{\tau_s} \frac{\bar{l}(k)-\bar{l}(k-1)}{T_s} \\ \bar{u}_q \end{bmatrix} \sim \mathcal{N}(\mu_2, \Sigma_2), \tag{7.12}$$

respectively.

Assuming that the measurements are independently, identically, and normally distributed

$$\rho = \begin{bmatrix} i_d(k) \\ i_q(k) \\ u_d(k) \\ u_q(k) \\ l(k) \\ i_d(k-1) \\ i_q(k-1) \\ l(k-1) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \bar{i}_d(k) \\ \bar{i}_q(k) \\ \bar{u}_d(k) \\ \bar{u}_q(k) \\ \bar{l}(k) \\ \bar{i}_d(k-1) \\ \bar{i}_q(k-1) \\ \bar{l}(k-1) \end{bmatrix}, \begin{bmatrix} \sigma_{idid} & \sigma_{idiq} & 0 & 0 & 0 & 0 & 0 & 0 \\ \sigma_{idiq} & \sigma_{iqiq} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{udud} & \sigma_{uduq} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{uduq} & \sigma_{uquq} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{ll} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{idid} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{iqiq} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{ll} \end{bmatrix} \right), \tag{7.13}$$

the synthesized measurements' probability distribution expressing their respective uncertainties are approximated as

$$d_1 = \begin{bmatrix} \dfrac{i_d(k) - i_d(k-1)}{T_s} \\ -\dfrac{\pi}{\tau_s} i_q \dfrac{l(k) - l(k-1)}{T_s} \\ i_d \\ u_d \end{bmatrix} \sim \mathcal{N}(\mu_1, \Sigma_1) \tag{7.14}$$

and

$$d_2 = \begin{bmatrix} \frac{\pi}{\tau_s} i_d \frac{l(k)-l(k-1)}{T_s} \\ \frac{i_q(k) - i_q(k-1)}{T_s} \\ i_q \\ \frac{\pi}{\tau_s} \frac{l(k)-l(k-1)}{T_s} \\ u_q \end{bmatrix} \sim \mathcal{N}(\mu_2, \Sigma_2), , \tag{7.15}$$

where

$$\Sigma_1 = \begin{bmatrix} \frac{2\sigma_{idid}}{T_s^2} & \frac{-\pi\sigma_{idiq}(\Delta\bar{l}_0)}{T_s^2\tau_s} & \frac{\sigma_{idid}}{T_s} & 0 & 0 \\ \frac{-\pi\sigma_{idiq}(\Delta\bar{l}_0)}{T_s^2\tau_s} & \frac{\pi^2(2\bar{i}_{q,0}^2\sigma_{ll} + (\Delta\bar{l}_0)^2\sigma_{iqiq} + 2\sigma_{ll}\sigma_{iqiq})}{T_s^2\tau_s^2} & \frac{-\pi\sigma_{idiq}(\Delta\bar{l}_0)}{T_s\tau_s} & 0 & 0 \\ \frac{\sigma_{idid}}{T_s} & \frac{-\pi\sigma_{idiq}(\Delta\bar{l}_0)}{T_s\tau_s} & \sigma_{idid} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{udud} \end{bmatrix}, \tag{7.16}$$

$$\Sigma_2 = \begin{bmatrix} \frac{\pi^2(2\bar{i}_{d,0}^2\sigma_{ll} + (\Delta\bar{l}_0)^2\sigma_{idid} + 2\sigma_{ll}\sigma_{idid})}{T_s^2\tau_s^2} & \frac{\pi\sigma_{idiq}(-\Delta\bar{l}_0)}{T_s^2\tau_s} & \frac{\pi\sigma_{idiq}(-\Delta\bar{l}_0)}{T_s\tau_s} & \frac{2\bar{i}_{d,0}\pi^2\sigma_{ll}}{T_s^2\tau_s^2} & 0 \\ \frac{\pi\sigma_{idiq}(-\Delta\bar{l}_0)}{T_s^2\tau_s} & \frac{2\sigma_{iqiq}}{T_s^2} & \frac{\sigma_{iqiq}}{T_s} & 0 & 0 \\ \frac{\pi\sigma_{idiq}(-\Delta\bar{l}_0)}{T_s\tau_s} & \frac{\sigma_{iqiq}}{T_s} & \sigma_{iqiq} & 0 & 0 \\ \frac{2\bar{i}_{d,0}\pi^2\sigma_{ll}}{T_s^2\tau_s^2} & 0 & 0 & \frac{2\pi^2\sigma_{ll}}{T_s^2\tau_s^2} & 0 \\ 0 & 0 & 0 & 0 & \sigma_{uquq} \end{bmatrix}. \tag{7.17}$$

In the covariance matrices, the parameters $\Delta\bar{l}_0$, $\bar{i}_{d,0}$, and $\bar{i}_{q,0}$ are to be stated by the statistician. In this case, they are approximated from data as

$$\Delta\bar{l}_0 = \frac{1}{N}\sum_{k=1}^{N} \bar{l}(k-1) - \bar{l}(k), \tag{7.18}$$

$$\bar{i}_{d,0} = \frac{1}{N}\sum_{k=0}^{N} \bar{i}_d(k), \tag{7.19}$$

$$\bar{i}_{q,0} = \frac{1}{N}\sum_{k=0}^{N} \bar{i}_q(k). \tag{7.20}$$

The B-GTLS algorithm is utilized to obtain two approximations of the likelihood

$$q_1(\theta_{12}|D) = \mathcal{N}\left(\hat{\theta}_{12}, \hat{\Sigma}_{12}(\hat{\theta}_{12})\right) \tag{7.21}$$

$$q_2(\theta|D) = \mathcal{N}\left(\hat{\theta}_2, \hat{\Sigma}(\hat{\theta}_2)\right), \tag{7.22}$$

where the first surrogate likelihood omits the Ψ_{PM} parameter due to no utilization of the parameter in its linear equation. After padding of the first surrogate, the likelihoods and prior can be combined into a single posterior surrogate distribution

$$q(\theta|D) = \mathcal{N}(\mu_\theta, \Sigma_\theta). \tag{7.23}$$

The continuous model of the PMLSM drive based on equations (1.8) is simulated in MATLAB using the ode45 function with parameters presented in Table 7.1.

Sampling Period	$T_s = 50$	$[\mu S]$
Number of Samples	$N = 10001$	$[-]$
Direct Inductance	$L_d = 405$	$[\mu H]$
Quadrature Inductance	$L_q = 665$	$[\mu H]$
Stator Resistance	$R = 380$	$[m\Omega]$
Permanent Magnet Flux	$\Psi_{PM} = 25.94$	$[mWb]$
Pole Pitch	$\tau_s = 30$	$[mm]$
Load Force	$F_L = 0$	$[N]$
Damping Coefficient	$B = 1$	$[-]$
Mass of the Mover	$M = 1.98$	$[Kg]$

Tab. 7.1: PMLSM simulation parameters.

The independent variables, voltages u_d and u_q, are sampled from normal distribution

$$\begin{bmatrix} \overline{u}_d \\ \overline{u}_q \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, 10\mathbf{I}). \tag{7.24}$$

The voltages change with the sample period of $T_u = 17.24\,ms$, resulting in $N_u = 30$ samples.

In order to demonstrate the ability of the B-GTLS algorithm to handle even extreme noise conditions, we decided to design additive normal noise with an exorbitant covariance matrix, resulting in the following noisy measurements

$$\rho_{\text{PMSM}} \sim \mathcal{N}\left(\begin{bmatrix} \bar{i}_d(k) \\ \bar{i}_q(k) \\ \overline{u}_d(k) \\ \overline{u}_q(k) \\ \bar{l}(k) \end{bmatrix}, \begin{bmatrix} 12 & 5 & 0 & 0 & 0 \\ 5 & 12 & 0 & 0 & 0 \\ 0 & 0 & 0.05 & 0.03 & 0 \\ 0 & 0 & 0.03 & 0.05 & 0 \\ 0 & 0 & 0 & 0 & 0.0001 \end{bmatrix}\right). \tag{7.25}$$

Figure 7.7 shows the resulting noise free output of simulation and simulated measurements. The figure's first $0.5\,s$ shows simulated data and corresponding noisy measurements used for the identification.

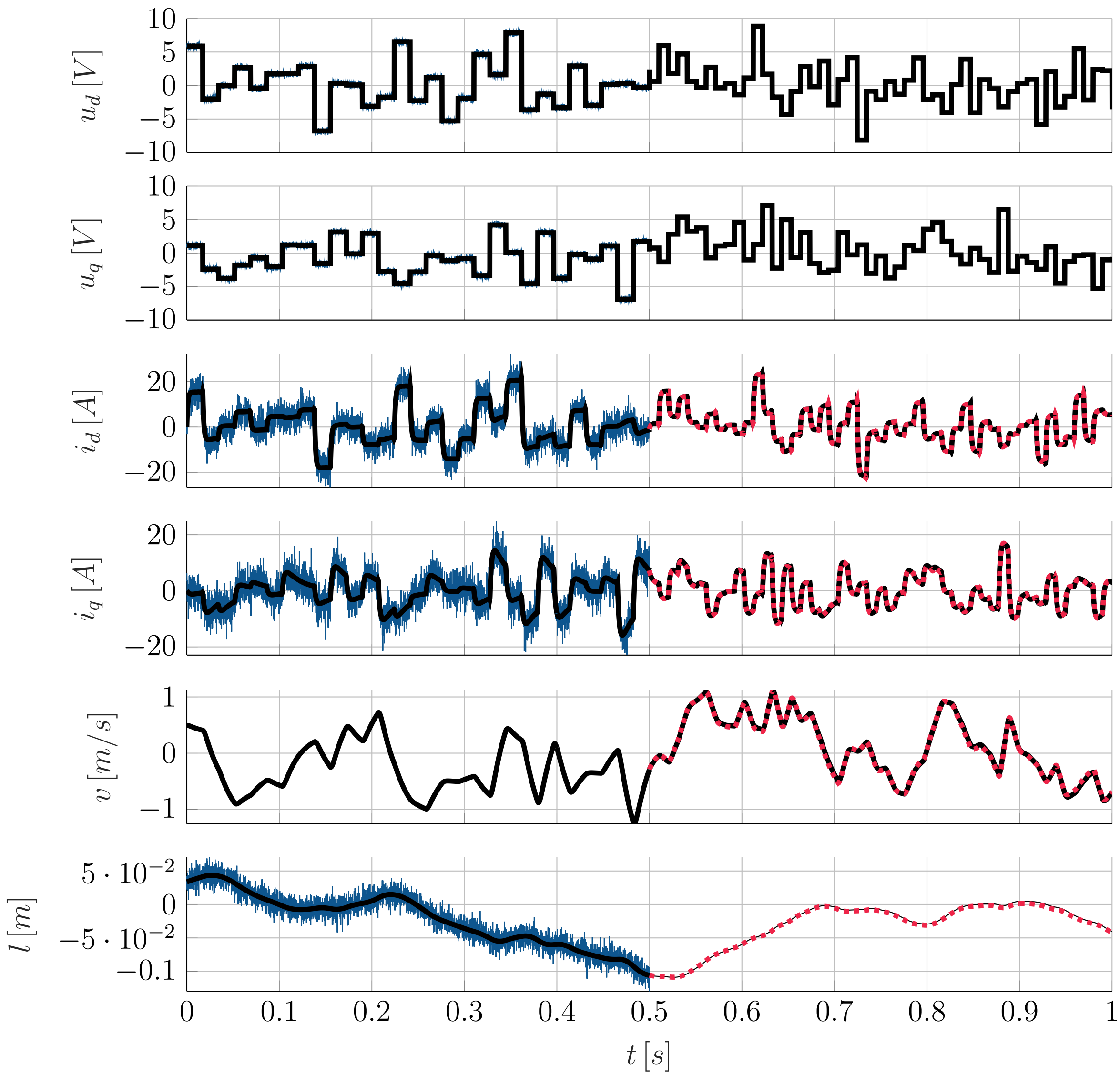


Fig. 7.7: Comparison of identification and validation data (black) with measurements (blue) and simulation with identified parameters (red), respectively.

The prior distribution is selected as

$$p(\theta) = \mathcal{N}\left(1.5\,\overline{\theta}, \mathrm{diag}(10^{-8},\, 10^{-8},\, 10^{-2},\, 10^{-4})\right), \tag{7.26}$$

where $\overline{\theta} = [L_d,\, L_q,\, R,\, \Psi_{PM}]^T$.

The measurements and prior distribution are provided to the PMLSM-RGTLS algorithm (Algorithm 6), resulting in the following surrogate posterior distributions and surrogate likelihoods.

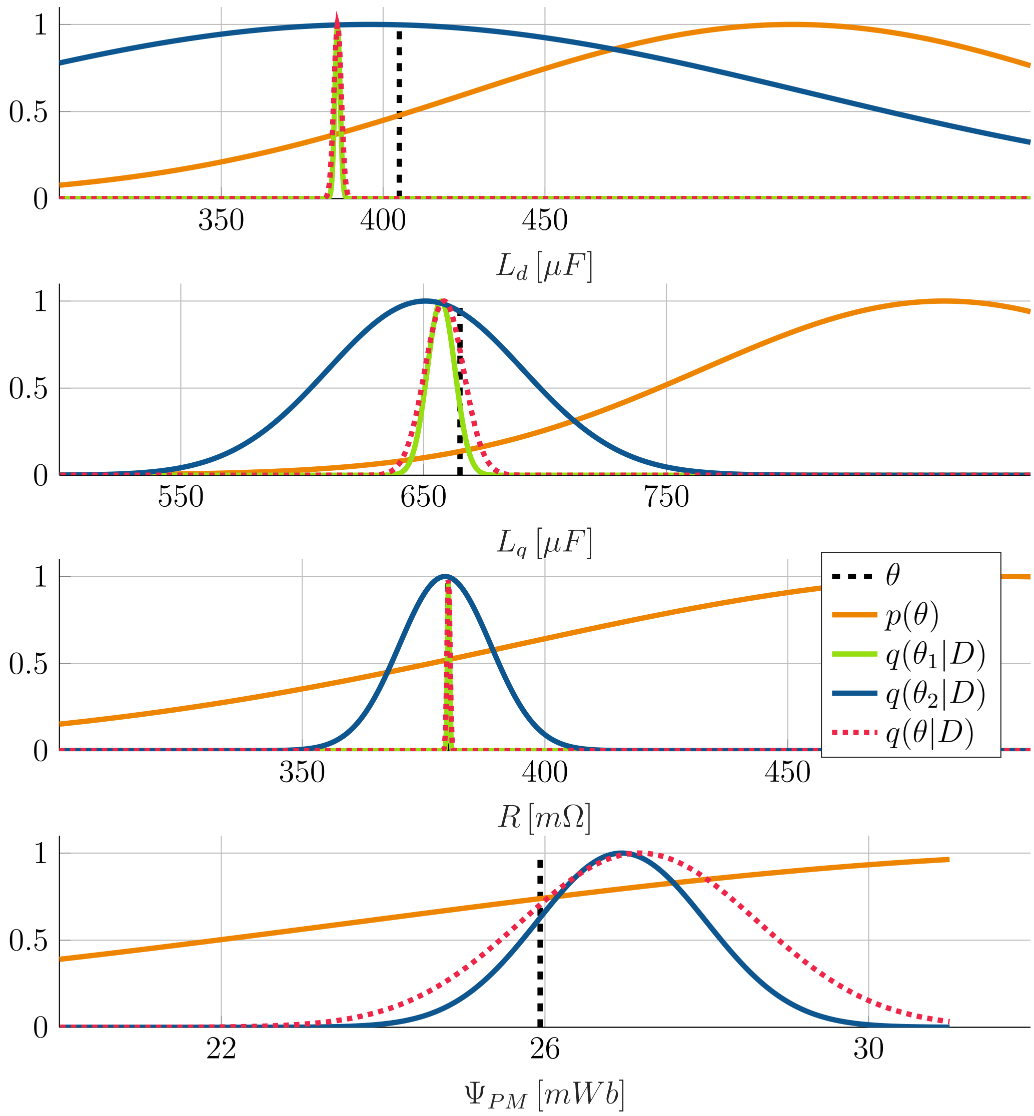


Fig. 7.8: Demonstration of scaled surrogate posterior distributions of the PMLSM parameters obtained from PMLSM-RGTLS algorithm.

Figure 7.8 shows that the surrogate likelihoods from individual equations provide both information about the sought-after parameter vector θ, except for the first surrogate posterior missing information about Ψ_{PM}, as was designed. The resulting posterior combines information from the prior and surrogate likelihoods. While the noise conditions are extreme, the parameters are close to the actual values.

Utilizing the modus of the surrogate posterior, the MAP estimate $\hat{\theta}$, simulation with new input data is done. This time with voltage period $T_u = 10.2\,ms$, resulting in voltage $N_u = 50$ samples. The simulation using MAP estimate of θ can be visually compared with simulation using true parameters $\overline{\theta}$ in the second half of Figure 7.7. It is apparent from the figure that the identified parameters sufficiently approximate the true parameters, as there is no visible difference between the simulations.

As the identification is recursive, the evolution of the MAP estimate can also be plotted.

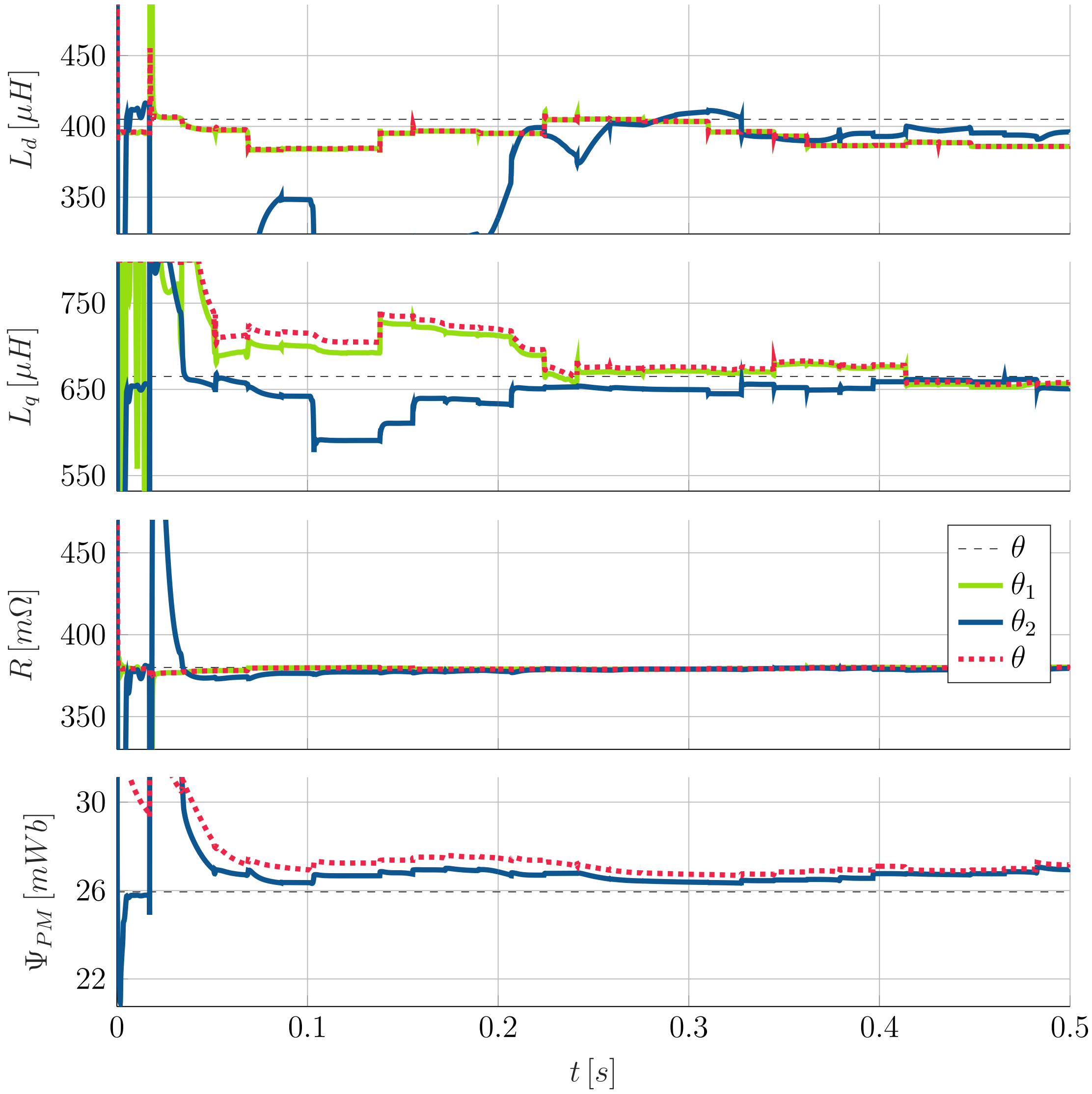


Fig. 7.9: Evolution of the MAP parameter estimates provided by the PMLSM-RGTLS algorithm.

Figure 7.9 shows that even with strong, correlated noise, the estimation of the PMLSM parameters converges to the actual parameters with a growing number of samples.

The reader can observe that uncertainty plays a significant role in the process of combining estimates from individual equations. In the L_d plot of Fig 7.9, the posterior closely follows θ_1 (green line) due to its lower uncertainty compared to θ_2 (blue line). This could never be achieved without the uncertainty estimate. The respective uncertainties are compared in Fig 7.8.

This numerical example demonstrates successful application of the PMLSM-RGTLS algorithm for real-time identification of PMLSM parameters. However, it is important to validate these findings with real data, which is addressed in the following example.

7.5 Recursive PMLSM Identification

The theoretical experiment proves the viability of the proposed PMLSM-RGTLS algorithm, and a practical experiment is the next logical step. In this experiment, we aim to demonstrate the practical implementation of the proposed algorithm. We will evaluate its performance by applying it to a test scenario similar to that in the theoretical case.

In this study, we used hardware provided by the Central European Institute of Technology through the Research and Innovation Centre on Advanced Industrial Production. The purpose of the innovation center is to deploy multiple testbeds to provide a testing environment that closely mimics real industrial conditions.

Fig. 7.10: Photo of used linear dynamometer.

One of the testbeds is a dynamometer for an industrial linear actuator, pictured in Figure 7.10, which allows accurate evaluation of the algorithm's capabilities under realistic conditions. This dynamometer consists of two linear PMLSM primary mover drives on a $5\,m$ long track; one drive contains the tested motor, which can be controlled by a National Instruments CompactRIO NI-cRIO-9038, while the other motor provides either position or force feedback to the tested motor, establishing safety boundaries or simulating load force. The drives are permanently connected and must be operated simultaneously. Currently, only one feedback mode can be used at a time.

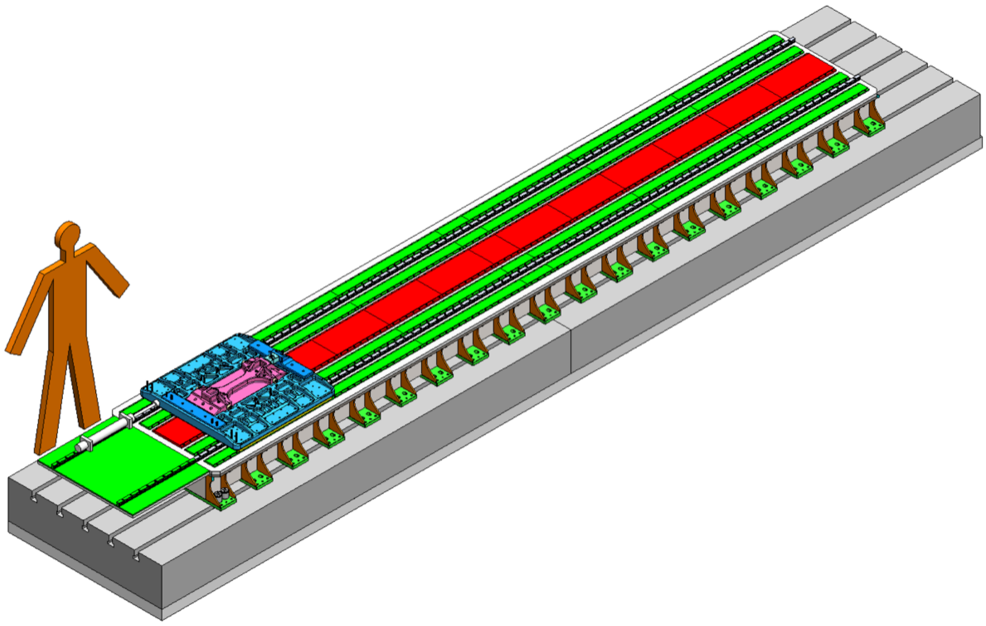


Fig. 7.11: Schematic view on used linear dynamometer, the tested and feedback motors are highlighted in purple and blue, respectively.

The inverter of the tested motor is the Hybrid Kit for HybridPACK™1 by Infineon, which consists of the Driver Board 6ED100HP1-FA (a six-channel IGBT driver) and the HybridPack™1 power module. The logic board from the pack is replaced with the mentioned CompactRIO.

While this solution provides complete control over the tested motor, it also has the disadvantage of high nonlinearity in the voltage output of the IGBT power module at low voltages. Mitigation of this effect necessitates the use of higher voltage.

The experiment design is identical to the theoretical case, with additional preventive measures. The main difference is that, in the simulation, the position can freely deviate from the starting position. In the practical case, the mover should not move close to the end of the base. This cannot be guaranteed in the load force feedback mode of the dynamometer; therefore, position feedback is utilized, allowing sufficient voltage to mitigate the nonlinearity of the IGBT power module. While the goal of the dynamometer in this feedback mode is to hold the position of both motors, the control is not ideal, and the tested motor's mover still moves. This effect can be modeled as substantial friction. Since the algorithm does not require knowledge of the load force, mass of the mover, or friction, this preventive measure does not pose a problem.

The tested motor, together with the entire dynamometer track, is a prototype machine. Its electrical parameters are known only from finite-element method (FEM) simulations. They may therefore differ from actual parameters. From the FEM simulations, the values of inductances are dependent on the state of the motor. The mean over all possible states is selected as the point estimate. The resistance is accurate for a temperature of $20°C$. All parameters used in the experiment and its validation are summarized in Table 7.2.

Sampling Period	$T_s = 50$	$[\mu S]$
Number of Samples	$N = 303173$	$[-]$
Direct Inductance	$L_d = 405$	$[\mu H]$
Quadrature Inductance	$L_q = 665$	$[\mu H]$
Stator Resistance	$R = 380$	$[m\Omega]$
Permanent Magnet Flux	$\Psi_{PM} = 25.94$	$[mWb]$
Pole Pitch	$\tau_s = 32$	$[mm]$

Tab. 7.2: Tested PMLSM drive parameters based on construction parameters and FEM simulations.

The independent variables are generated analogously to the previous experiment, with a sampling period of $T_u = 10ms$. Executing the experiment for approximately $15.1s$ results in $N = 303173$ data samples. The data is split into two parts: the identification part (the first $8s$ of the data) and the verification part (the remaining data).

From the data, measurement uncertainty is empirically estimated as

$$\rho_{\text{PMSM}} \sim \mathcal{N}\left(\begin{bmatrix} \bar{i}_d(k) \\ \bar{i}_q(k) \\ \bar{u}_d(k) \\ \bar{u}_q(k) \\ \bar{l}(k) \end{bmatrix}, \begin{bmatrix} 0.3162 & 0.3700 & 0 & 0 & 0 \\ 0.3700 & 0.3162 & 0 & 0 & 0 \\ 0 & 0 & 0.1802 & 0 & 0 \\ 0 & 0 & 0 & 0.1802 & 0 \\ 0 & 0 & 0 & 0 & 8.4853 \cdot 10^{-6} \end{bmatrix}\right). \tag{7.27}$$

The σ_{uduq} parameter does not appear in the noise covariance matrices and is therefore estimated as 0. Comparing the noise covariance matrix with the previous example (7.25), current and position uncertainty is considerably lower. This is expected, as the noise was exaggerated in the theoretical case. Voltage noise, on the other hand, is more significant. While in the theoretical case, only the switching noise of the inverter is considered the primary source of uncertainty, in the practical case, the noise from the voltage DC-DC converter used as the power source is the prominent source of voltage noise.

Details of the measured data with approximated σ intervals are presented in Figure 7.12.

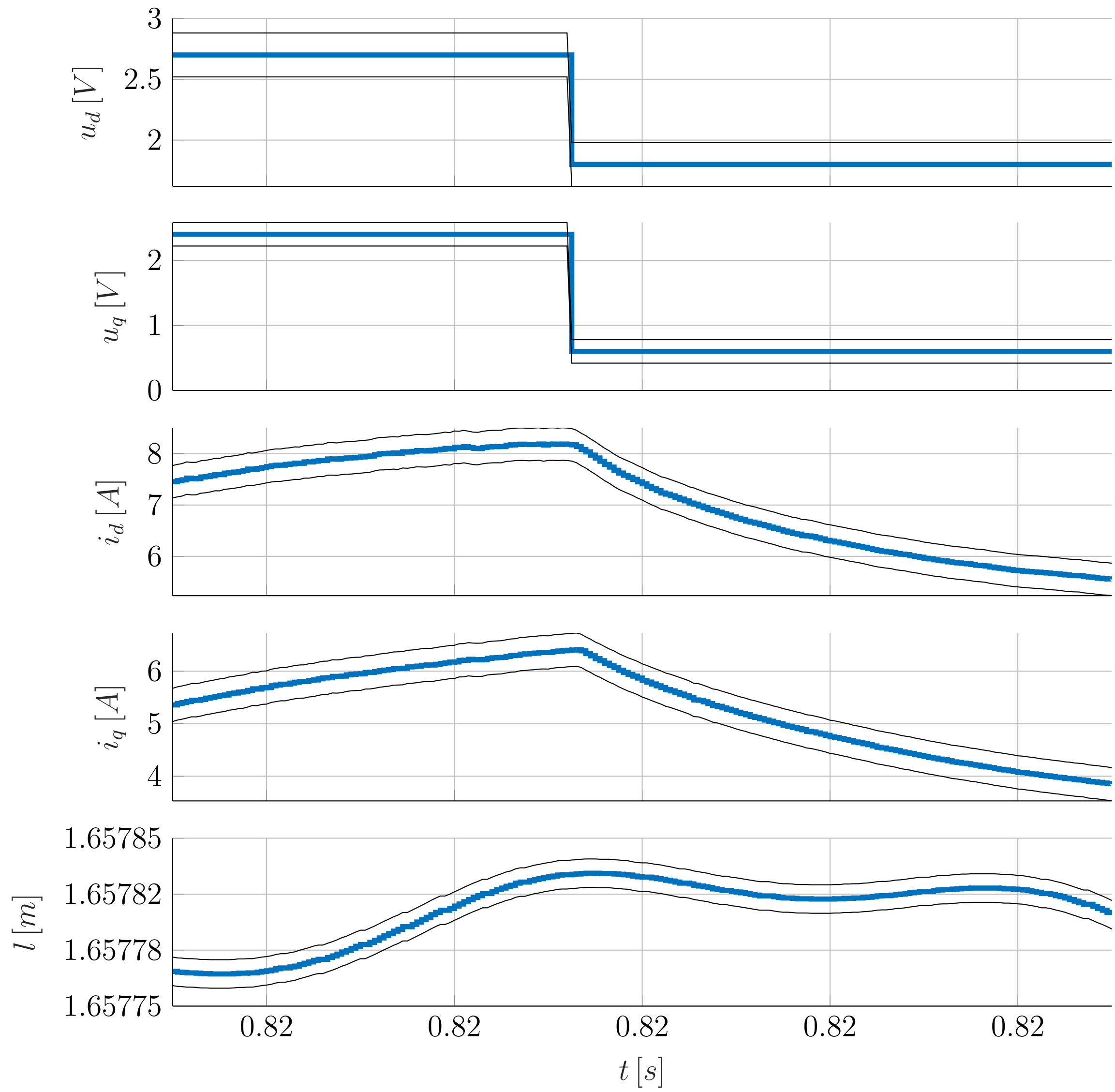


Fig. 7.12: Detail of experiment data with σ intervals.

The prior distribution is selected

$$p(\theta) = \mathcal{N}\left(\overline{\theta}, \mathrm{diag}(10^4,\, 10^4,\, 10^2,\, 1)\right), \tag{7.28}$$

where $\overline{\theta} = [L_d,\, L_q,\, R,\, \Psi_{PM}]^T$ is the original estimate of the parameters, assembled from FEM estimations in Table 7.2.

The PMLSM-RGTLS algorithm (Algorithm 6) receives the synthesized measurements built from the identification part (the first $8s$) of the measured data and the prior distribution as inputs. This process yields the corresponding evolution of the surrogate posterior MAP and surrogate likelihood MAPs depicted in Figure 7.13.

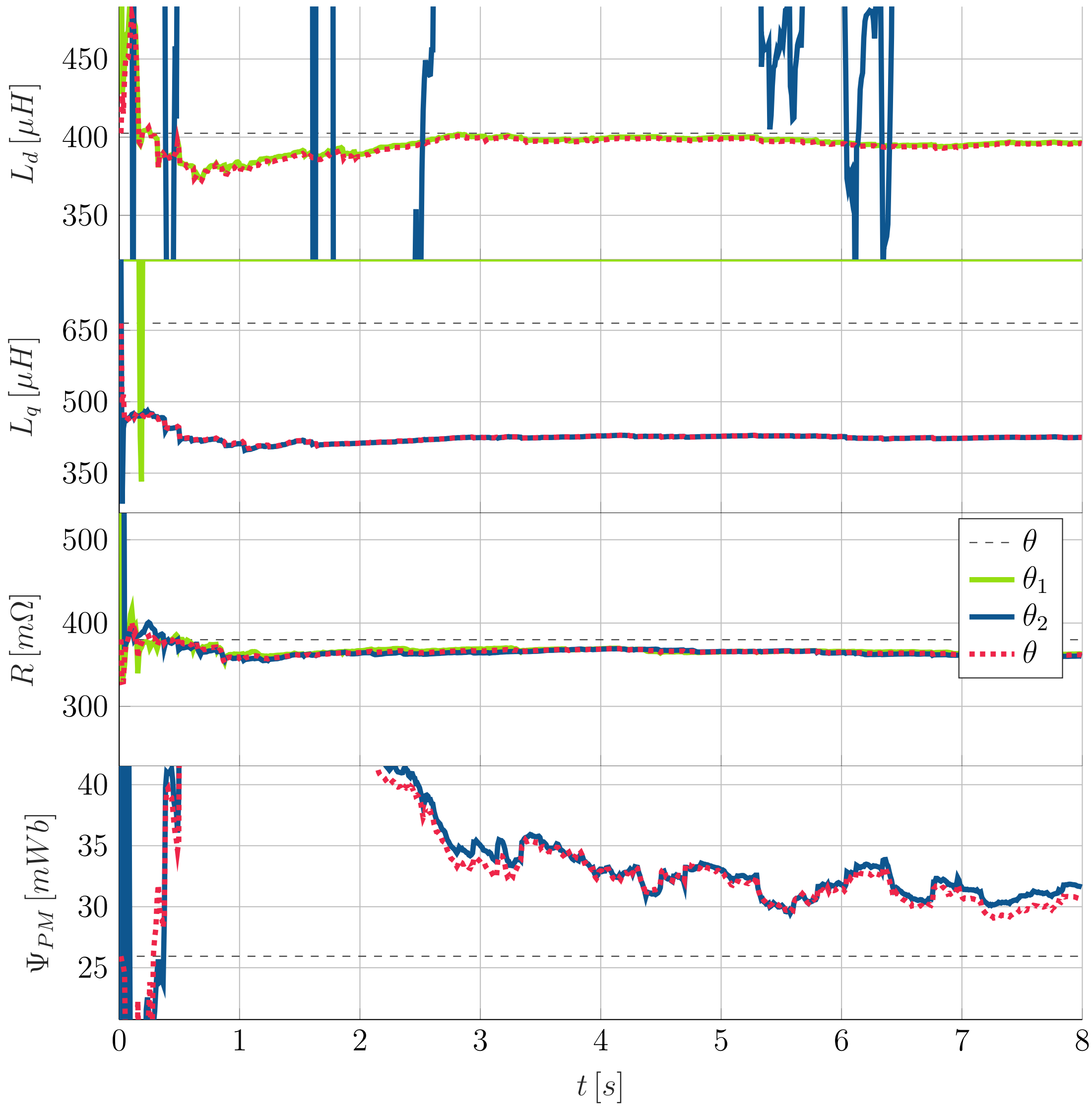


Fig. 7.13: Evolution of the MAP parameter estimates provided by the PMLSM-RGTLS algorithm.

Similar to the previous example, in the L_d plot shown in Figure 7.13, the posterior closely follows θ_1 (green line) due to its lower uncertainty compared to θ_2 (blue line), which has a higher uncertainty. Similarly, in the L_q plot, the posterior follows θ_2 (blue line).

From Figure 7.13, it is apparent that the L_d and R parameter MAPs converge close to the parameters estimated by the authors of the tested drive, while the

MAPs of the L_q and Ψ_{PM} parameters seem to converge to different values. This discrepancy does not necessarily indicate a fault in the identification process. To determine which parameter approximation is better, verification is performed.

The differences and relative errors of individual parameters can be compared in the Table 7.3.

	$L_d\,[\mu H]$	$L_q\,[\mu H]$	$R\,[m\Omega]$	$\Psi_{PM}\,[mWb]$
Original parameters ($\overline{\theta}$)	405.00	665.00	380.00	25.94
MAP approximation ($\hat{\theta}_{MAP}$)	391.42	425.04	361.44	30.43
Relative error ($\frac{\overline{\theta}-\hat{\theta}_{MAP}}{\overline{\theta}}100\,[\%]$)	3.35	36.08	4.88	-17.33

Tab. 7.3: Comparison of the original parameters and parameters obtained from PMLSM-RGTLS algorithm.

To verify the results, simulations with the estimated parameters are executed and compared with the verification part of the measured data. Analogous to the previous chapter, simulations are conducted using Matlab's ode45 solver to solve the PMLSM state equations (1.8). Since the mechanical parameters (such as the joint mass of the two movers, damping coefficient, and load force in the case of the position feedback mode) are unknown, the speed of the mover v is estimated from the data rather than from the simulation. However, only the position is measured. The Euler approximation Δl can be utilized, but the resulting signal is noisy. For filtration, an angle tracking observer (ATO) improved with a first-order Butterworth filter from [155] is used. The Laplace transform of ATO's transfer function is

$$G(s) = \frac{3P^2 s(s+\frac{P}{3})}{(s+P)^3}, \tag{7.29}$$

where s is the Laplace variable, and P is ATO's tuning parameter, which is set to 1700. The resulting signals are depicted in Figure 7.14.

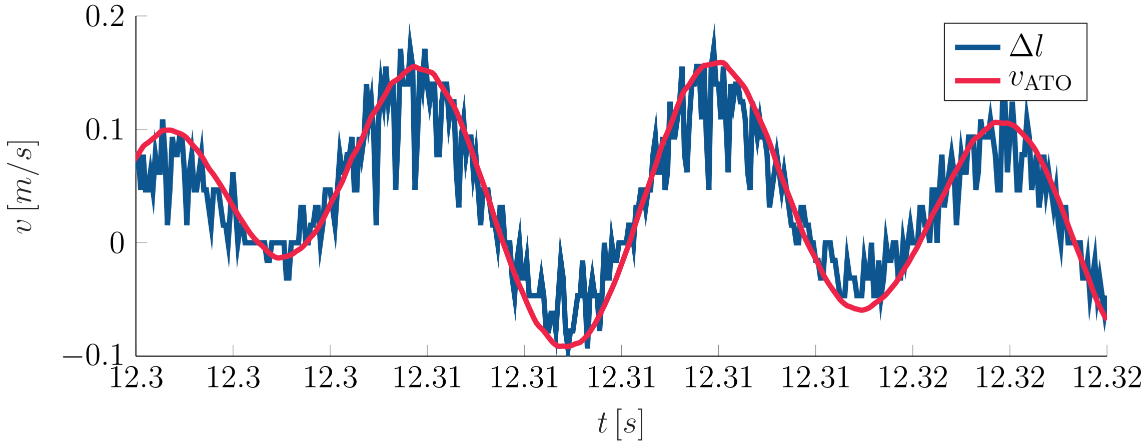


Fig. 7.14: Comparison of euler Δl and ATO filtered euler v_{ATO} mover speed approximations.

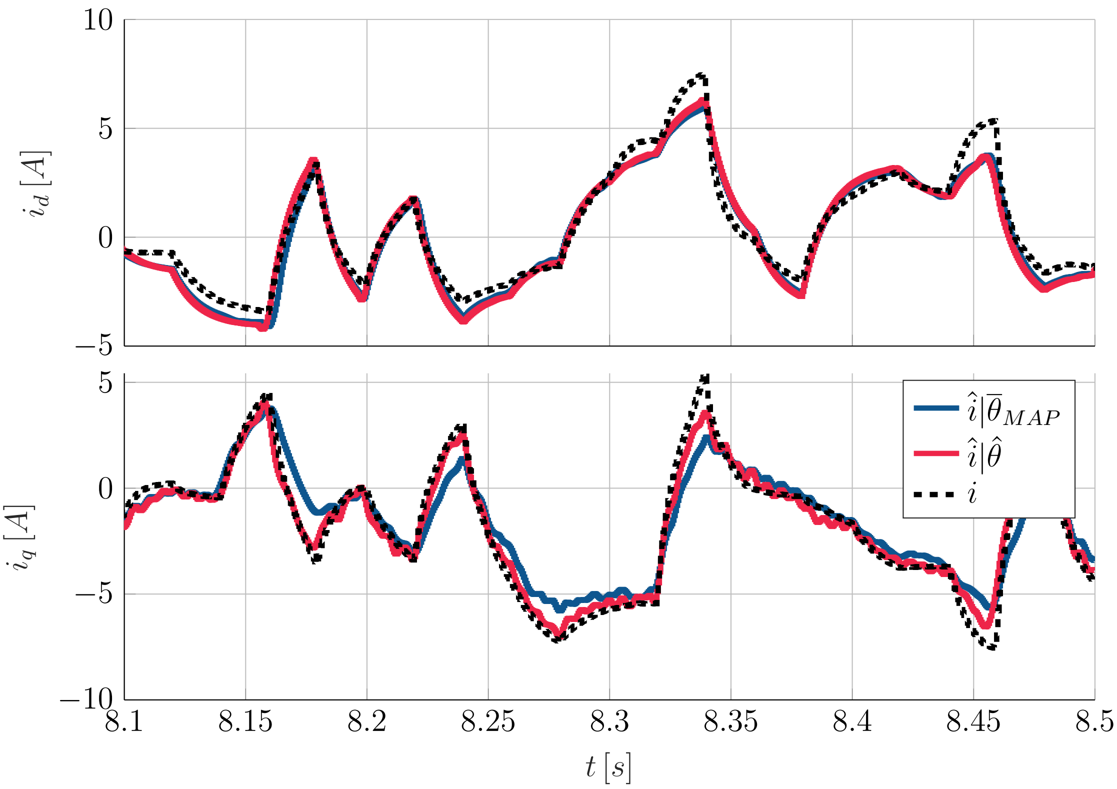
The simulation, using the measured voltages, estimated speed v_{ATO}, and parameter estimates $\overline{\theta}$ and $\hat{\theta}_{MAP}$, results in $\hat{i}_d$ and $\hat{i}_q$ current waveforms. These are compared with the measured current waveforms i. The mean squared errors for the individual parameter estimates are:

$$MSE_{\overline{\theta}} = \sum_{k=1}^{N_v} (i_d(k) - \hat{i}_d(k)|\overline{\theta})^2 + (i_q(k) - \hat{i}_q(k)|\overline{\theta})^2 = 1.2245, \tag{7.30}$$

$$MSE_{\hat{\theta}_{MAP}} = \sum_{k=1}^{N_v} (i_d(k) - \hat{i}_d(k)|\hat{\theta}_{MAP})^2 + (i_q(k) - \hat{i}_q(k)|\hat{\theta}_{MAP})^2 = 0.7481, \tag{7.31}$$

where N_v is number of verification samples.

The fact that $MSE_{\overline{\theta}}$ is considerably larger than $MSE_{\hat{\theta}_{MAP}}$ shows that the identification process improved the estimates. This can also be observed visually in Figure 7.5.



The identification is successful, demonstrating that the Bayesian Generalized Total Least Squares is useful, and that the derived PMLSM-RGTLS algorithm is both effective and applicable under real industrial conditions. The PMLSM-RGTLS algorithm is designed so that existing GTLS algorithms can be utilized for real-time implementation. By selecting an appropriate forgetting factor, this algorithm can accurately track parameters in real time. This concludes the experiment and validates the approach.

Conclusion

The main goal of this thesis is an analysis of the errors-in-variables problem from the Bayesian point of view with a primary goal of proposing a new, EIV-driven view on PMLSM parameter identification, opening entryway for the derivation of novel diagnostic and control algorithms. The main advantage of the Bayesian framework is the incorporation of prior knowledge, resulting in inherent sequential learning. Probabilistic interpretation allows for expressing uncertainty of not only resulting parameters but also uncertainty of measurements or lack thereof. Stochastic interpretation of results also allows for the integration of multiple data sources. All mentioned properties are utilized in this thesis. Additionally, employment of the Bayesian framework allows for the extension of the results with existing Bayesian methods, such as Bayesian decision-making, robustness-improving methods, and various regularization methods. This, however, does not exclude expansions of the results by utilizing non-statistical methods, as maximum a posteriori (MAP) or mean estimate can always be utilized as a point estimate.

The thesis provides three main contributions:

I. The main contribution is the proposed Bayesian generalized total least squares (B-GTLS) framework, allowing for the identification of EIV problems with the arbitrary noise covariance matrix. It is shown that under classical EIV assumptions, the MAP solution is equal to the solution provided by the well-known generalized total least squares algorithm. This implies that existing, numerically stable, and efficient algorithms can be used to obtain the MAP solution. The posterior distribution is an unknown distribution function with unfavorable properties. An analytical equation providing a normal approximation of the posterior distribution is derived such that the only required information is the noise covariance and the MAP solution, which can be provided by the existing algorithms. The B-GTLS framework can, therefore, be perceived as a statistical extension of the GTLS framework for solving EIV problems. The mentioned properties of the B-GTLS allow for the derivation of algorithms suitable for problems with extreme noises, large dimensions, and multiple sources and, due to support of existing GTLS algorithms, also problems with, E.g., slowly varying parameters, zero variance measurements, under-determined systems and numerically unstable cases while achieving the possibility of implementation in realtime hardware due to efficient computation and guaranteed numerical stability. It is also shown that B-GTLS problems with non-singular noise covariance can be solved by scaling the data such that the noise covariance is the identity matrix. The simpler B-TLS problem can be solved to obtain a solution, which can be re-scaled back to the B-GTLS framework. This motivates the second main contribution.

II. A special case of B-GTLS with an scaled identity covariance matrix is analyzed, resulting in the Bayesian total least squares (B-TLS) framework. Its simpler form allows for additional possibilities stemming from the transformation of the posterior distribution into the Bingham distribution. This allows, for example, a graphical representation of the results, sampling from the posterior distribution, and the derivation of variation approximation of both likelihood and posterior distribution. All other properties are carried over from the B-GTLS framework, meaning that existing algorithms can also be utilized in this case. This leads to additional advantages compared to B-GTLS, stemming from the bigger popularity of the TLS problem and a more comprehensive selection of available algorithms.

III. The B-GTLS framework is utilized for the design of online, recursive algorithm for identification of PMLSM and PMSM parameters. This method does not require knowledge of the mover mass, load force, and friction coefficient, nor does it require locking of the rotor or unique design of identification inputs, as is customary for many identification algorithms. As rapid measurements of the PM(L)SM are usually available, model error can be neglected and the measurement error is dominant. This leads to re-casting the identification problem as an EIV problem, which natively supports assigning uncertainty also to the inputs, in this case, direct and quadrature voltages, which can be altered due to disturbances of voltage generation. The PMLSM model is transformed into a linear EIV problem with synthesized measurements by the moment matching technique. The algorithm is validated on both simulation and practical experiment. The simulation experiment shows, that the algorithm converges to the actual parameters even in case of substantial noise. The practical experiment confirms simulation results and proves applicability in real industrial conditions.

As an unexpected finding, the shape of the EIV posterior distribution under the B-TLS formulation recovers previously unrecognized connection with directional analysis. This connection allows for the utilization of findings from directional statistics to provide graphical and geometrical representation, sampling from the posterior distribution, and a window for further possible research.

For the B-GTLS and B-TLS, five algorithms are derived in total. Those are the I. B-RGTLS, the algorithm based on recursive GTLS that provides statistical results for the general EIV problem; II. the PMLSM-RGTLS algorithm builds upon B-RGTLS, and allows for recursive online identification of the PM(L)SM parameters; III. the IIL-RTLS algorithm, providing posterior approximation for the EIV problem with identity covariance assuming prior in form of Bingham distribution; IV. the Recursive Laplace TLS algorithm, providing posterior approximation for the EIV problem with identity covariance assuming prior in form of normal distribution and V. Recursive VB TLS, providing variational posterior approximation for problems with identity covariance and two measurements. All designed algorithms are numerically analyzed, showing fast convergence.

Utilization of the algorithms is further demonstrated by both numerical and practical experiments, illustrating various ways of posterior parameter uncertainty utilization.

Existing recursive identification methods supporting errors-in-variables problems are almost nonexistent, as the recent research of the errors-in-variables methods usually analyzes only batch identification. The closest work is [156], which formulates a Bayesian view on the maximum likelihood identification of problems with correlated errors. However, the error covariance is approximated with the Fisher information matrix, which can be imprecise for a smaller number of samples, as was demonstrated by an example.

While the work provides numerous novel views and algorithms, there is still room for further development, improvements, and extensions. Those are mainly the derivation of improved posterior approximations, implementation of more sophisticated forgetting methods, or extension of the algorithms by decision algorithm, allowing for storing only essential parts of data. Another possible improvement is to regularize the algorithm to provide sparse approximates, store only sparse matrices, or adding support for identification of some parts of the noise covariance matrix. Furthermore, specialized algorithms can be generated for the identification of various other models, which could, by their structure, provide additional constraints and information about the parameters.

The application of the algorithms and findings presented in this study extends naturally to other domains, such as diagnostics, control systems, model comparison, mitigation of adverse effects, and decision-making processes.

In conclusion, this doctoral thesis has presented a comprehensive study of the errors-in-variables and, thanks to appropriate approximations of problematic components, results in a practical framework that extends existing GTLS and TLS methods by providing accurate normal posterior approximations entailing posterior uncertainty. The practicality of the research is confirmed by the formulation of recursive algorithms, which, by expanding on existing algorithms, preserves the advantages of years behind the development of TLS and GTLS methods. Examples demonstrate the extensive application possibilities opened by proposed algorithms.

PMLSM identification problem is reformulated into an errors-in-variables problem, and the B-RGTLS algorithm is utilized for real-time PMLSM identification. B-RGTLS's ability allows for online adaptation of identified parameters and enables derivation of novel diagnostic and control methods that incorporate parameter uncertainty.

Author's publications

1. FRIML, Dominik; VÁCLAVEK, Pavel. Recursive Variational Inference for Total Least-Squares. *IEEE Control Systems Letters*. 2023, vol. 7, pp. 2839–2844.
2. FRIML, Dominik; VÁCLAVEK, Pavel. Bayesian inference of total least-squares with known precision. In: *2022 IEEE 61st Conference on Decision and Control (CDC)*. 2022, pp. 203–208.
3. FRIML, Dominik; KOZUBÍK, Michal; VÁCLAVEK, Pavel. On Improving TLS Identification Results Using Nuisance Variables with Application on PMSM. In: *IECON 2021–47th Annual Conference of the IEEE Industrial Electronics Society*. 2021, pp. 1–6.
4. FRIML, Dominik. Derivation and practical comparison of recursive LS and TLS system identification methods. In: 2021, 404 – 409.
5. KOZUBIK, Michal; FRIML, Dominik. Differential evolution based nonlinear model predictive speed control of pmsm implemented on gpu. In: *2021 IEEE 30th International Symposium on Industrial Electronics (ISIE)*. 2021, pp. 1–6.

Bibliography

6. BOLDEA, I.; NASAR, S. A. Linear electric actuators and generators. *IEEE Transactions on Energy Conversion*. 1999, vol. 14, no. 3, pp. 712–717. ISSN 08858969. Available from DOI: 10.1109/60.790940. Publisher: IEEE.
7. EGAMI, Tadashi; TSUCHIYA, Takeshi. Disturbance Suppression Control with Preview Action of Linear DC Brushless Motor. *IEEE Transactions on Industrial Electronics*. 1995, vol. 42, no. 5, pp. 494–500. ISSN 15579948. Available from DOI: 10.1109/41.464612.
8. SANADA, Masayuki; MORIMOTO, Shigeo; TAKEDA, Yoji. Interior permanent magnet linear synchronous motor for high performance drives. *Conference Record - IAS Annual Meeting (IEEE Industry Applications Society)*. 1996, vol. 1, pp. 15–20. ISSN 01972618. Available from DOI: 10.1109/IAS.1996.556991. Publisher: IEEE.
9. TAM, Michael SW; CHEUNG, Norbert C. A robust fully-digital drive for linear permanent magnet synchronous motor. In: *2004 1st International Conference on Power Electronics Systems and Applications Proceedings*. 2004, pp. 188–193. ISBN 962-367-434-1.
10. TAM, M. S.W.; CHEUNG, N. C. A high speed, high precision linear drive system for manufacturing automation. *Conference Proceedings - IEEE Applied Power Electronics Conference and Exposition - APEC*. 2001, vol. 1, pp. 440–444. Available from DOI: 10.1109/APEC.2001.911684.
11. GIERAS, Jacek F. *Permanent Magnet Motor Technology Design and Applications*. CRC Press, 2010. ISBN 978-1-4200-6440-7. Available from DOI: 10.1201/9780429292736. arXiv: 1011.1669v3 Publication Title: Taylor & Francis Group,LLC ISSN: 1098-6596.
12. TROUT, S. R. Material selection of permanent magnets, considering thermal properties correctly. *Proceedings of the Electrical/Electronics Insulation Conference*. 2001, pp. 365–370. Available from DOI: 10.1109/EEIC.2001.965683.
13. GIERAS, Jacek F.; PIECH, Zbigniew J.; TOMCZUK, Bronisław Zbigniew. *Linear synchronous motors : transportation and automation systems*. CRC Press, 2012. ISBN 978-1-138-07205-3.
14. ISHII, Toshiaki. Elevators for skyscrapers. *IEEE Spectrum*. 1994, vol. 31, no. 9, pp. 42–46. ISSN 00189235. Available from DOI: 10.1109/6.309960.
15. S. MARKON, H. KITA, AND H. KISE. *Control of Traffic Systems in Buildings*. Springer London, 2006. Available from DOI: 10.1007/1-84628-449-x. Publication Title: Control of Traffic Systems in Buildings.

16. JONES, Willie D. How to build a mile-high skyscraper. *IEEE Spectrum*. 2007, vol. 44, no. 6, pp. 52–53. ISSN 00189235. Available from DOI: 10.1109/MSPEC.2007.369270.

17. KRÄMER, Christian; KUGI, Andreas; KEMMETMÜLLER, Wolfgang. Optimal force control of a permanent magnet linear synchronous motor based on a magnetic equivalent circuit model. *Control Engineering Practice*. 2022, vol. 122, p. 105076. ISSN 0967-0661. Available from DOI: 10.1016/J.CONENGPRAC.2022.105076. Publisher: Pergamon.

18. GIERAS, J F; PIECH, Z J; TOMCZUK, B. *Linear Synchronous Motors: Transportation and Automation Systems, Second Edition*. CRC Press, 2016. ISBN 978-1-4398-4222-5. Series Title: Electric Power Engineering Series.

19. GAN, Wai Chuen; QIU, Li. Design and analysis of a plug-in robust compensator: An application to indirect-field-oriented-control induction machine drives. *IEEE Transactions on Industrial Electronics*. 2003, vol. 50, no. 2, pp. 272–282. ISSN 02780046. Available from DOI: 10.1109/TIE.2003.809393.

20. DOYLE, John. Robust and optimal control. *Proceedings of the IEEE Conference on Decision and Control*. 1996, vol. 2, pp. 1595–1598. ISSN 01912216. Available from DOI: 10.1016/s0005-1098(97)00132-5. Publisher: IEEE ISBN: 0-7803-3590-2.

21. ZHOU Kemin and Doyle, John Comstock. *Essentials of robust control*. Vol. 104. Prentice hall Upper Saddle River, NJ, 1998. ISBN 978-0-13-525833-0.

22. ELMAN, Jeffrey L. Finding Structure in Time. *Cognitive Science*. 1990, vol. 14, no. 2, pp. 179–211. ISSN 1551-6709. Available from DOI: 10.1207/S15516709COG1402_1. Publisher: John Wiley & Sons, Ltd.

23. KHAN, Muhammad Yousaf Ali; KOO, Ja Yoon. Neural network based diagnosis of partial discharge defects patterns at XLPE cable under DC stress. *Electrical Engineering*. 2017, vol. 99, no. 1, pp. 119–132. ISSN 14320487. Available from DOI: 10.1007/S00202-016-0395-0/TABLES/4. Publisher: Springer Verlag.

24. ASLAN, Yılmaz; YAĞAN, Yunus Emre. Artificial neural-network-based fault location for power distribution lines using the frequency spectra of fault data. *Electrical Engineering*. 2017, vol. 99, no. 1, pp. 301–311. ISSN 14320487. Available from DOI: 10.1007/S00202-016-0428-8/TABLES/7. Publisher: Springer Verlag.

25. CAMPOS, Paul J.; CORIA, Luis N.; TRUJILLO, Leonardo. Nonlinear speed sensorless control of a surface-mounted PMSM based on a Thau observer. *Electrical Engineering*. 2018, vol. 100, no. 1, pp. 177–193. ISSN 14320487. Available from DOI: 10.1007/S00202-016-0491-1/TABLES/3. Publisher: Springer Verlag.

26. HARROUZ, Abdelkader; BECHERI, Houcine; COLAK, Ilhami; KAYISLI, Korhan. Backstepping control of a separately excited DC motor. *Electrical Engineering*. 2018, vol. 100, no. 3, pp. 1393–1403. ISSN 14320487. Available from DOI: 10.1007/S00202-017-0592-5/TABLES/1. Publisher: Springer Verlag.

27. LIN, Chih Hong; CHEN, An Cheng. Adaptive backstepping control for SynRel motor drive using FNN uncertainty observer. *IECON Proceedings (Industrial Electronics Conference)*. 2007, pp. 433–438. Available from DOI: 10.1109/IECON.2007.4459914. ISBN: 1424407834.

28. BRDYŚ, Mietek A.; KULAWSKI, Grzegorz J. Dynamic neural controllers for induction motor. *IEEE Transactions on Neural Networks*. 1999, vol. 10, no. 2, pp. 340–355. ISSN 10459227. Available from DOI: 10.1109/72.750564.

29. LIN, Chih Hong. Linear permanent magnet synchronous motor drive system using AAENNB Control system with error compensation controller and CPSO. *Electrical Engineering*. 2020, vol. 102, no. 3, pp. 1311–1325. ISSN 14320487. Available from DOI: 10.1007/S00202-020-00953-4. Publisher: Springer.

30. SU, Wei Te; LIAW, Chang Ming. Adaptive positioning control for a LPMSM drive based on adapted inverse model and robust disturbance observer. *IEEE Transactions on Power Electronics*. 2006, vol. 21, no. 2, pp. 505–517. ISSN 08858993. Available from DOI: 10.1109/TPEL.2005.869729.

31. SUN, Xiaodong; DIAO, Kaikai; LEI, Gang; GUO, Youguang; ZHU, Jianguo. Direct Torque Control Based on a Fast Modeling Method for a Segmented-Rotor Switched Reluctance Motor in HEV Application. *IEEE Journal of Emerging and Selected Topics in Power Electronics*. 2021, vol. 9, no. 1, pp. 232–241. ISSN 21686785. Available from DOI: 10.1109/JESTPE.2019.2950085. Publisher: Institute of Electrical and Electronics Engineers Inc.

32. SUN, Xiaodong; CAO, Junhao; LEI, Gang; GUO, Youguang; ZHU, Jianguo. Speed sensorless control for permanent magnet synchronous motors based on finite position set. *IEEE Transactions on Industrial Electronics*. 2020, vol. 67, no. 7, pp. 6089–6100. ISSN 15579948. Available from DOI: 10.1109/TIE.2019.2947875. Publisher: Institute of Electrical and Electronics Engineers Inc.

33. KANELLAKOPOULOS, I.; KOKOTOVIC, P. V.; MORSE, A. S. Systematic design of adaptive controllers for feedback linearizable systems. *Proceedings of the American Control Conference*. 1991, vol. 1, pp. 649–654. ISSN 07431619. Available from DOI: 10.23919/ACC.1991.4791451. Publisher: Publ by American Automatic Control Council ISBN: 0879425652.

34. BARTOLINI, Giorgio; FERRARA, Antonella; GIACOMINI, Luisa; USAI, Elio. Properties of a combined adaptive/second-order sliding mode control algorithm for some classes of uncertain nonlinear systems. *IEEE Transactions on Automatic Control*. 2000, vol. 45, no. 7, pp. 1334–1341. ISSN 00189286. Available from DOI: 10.1109/9.867041. Publisher: IEEE.

35. CHOW, Tommy W.S.; FANG, Yong. A recurrent neural-network-based real-time learning control strategy applying to nonlinear systems with unknown dynamics. *IEEE Transactions on Industrial Electronics*. 1998, vol. 45, no. 1, pp. 151–161. ISSN 02780046. Available from DOI: 10.1109/41.661316.

36. LI, Xiao Dong; HO, John K.L.; CHOW, Tommy W.S. Approximation of Dynamical Time-Variant Systems by Continuous-Time Recurrent Neural Networks. *IEEE Transactions on Circuits and Systems II: Express Briefs*. 2005, vol. 52, no. 10, pp. 656–660. ISSN 15583791. Available from DOI: 10.1109/TCSII.2005.852006.

37. YANG, Xi Yun; XU, Da Ping; HAN, Xiao Juan; ZHOU, Hai Ning. Predictive functional control with modified Elman neural network for reheated steam temperature. *2005 International Conference on Machine Learning and Cybernetics, ICMLC 2005*. 2005, pp. 4699–4703. Available from DOI: 10.1109/ICMLC.2005.1527768. ISBN: 078039092X.

38. KALU, Obinna O.; MADUEME, T. C. Application of artificial neural network (ANN) to enhance power systems protection: a case of the Nigerian 330 kV transmission line. *Electrical Engineering*. 2018, vol. 100, no. 3, pp. 1467–1479. ISSN 14320487. Available from DOI: 10.1007/S00202-017-0599-Y/FIGURES/13. Publisher: Springer Verlag.

39. BLAIFI, S.; MOULAHOUM, S.; COLAK, I.; MERROUCHE, W. Monitoring and enhanced dynamic modeling of battery by genetic algorithm using LabVIEW applied in photovoltaic system. *Electrical Engineering*. 2018, vol. 100, no. 2, pp. 1021–1038. ISSN 14320487. Available from DOI: 10.1007/S00202-017-0567-6/FIGURES/24. Publisher: Springer Verlag.

40. SUN, Xiaodong; HU, Changchang; LEI, Gang; GUO, Youguang; ZHU, Jianguo. State feedback control for a PM hub motor based on gray Wolf optimization algorithm. *IEEE Transactions on Power Electronics*. 2020, vol. 35, no.

1, pp. 1136–1146. ISSN 19410107. Available from DOI: 10.1109/TPEL.2019.2923726. Publisher: Institute of Electrical and Electronics Engineers Inc.

41. KOZUBIK, Michal; VACLAVEK, Pavel. Speed Control of PMSM with Finite Control Set Model Predictive Control Using General-purpose Computing on GPU. In: *IECON Proceedings (Industrial Electronics Conference)*. IEEE Computer Society, 2020, vol. 2020-Octob, pp. 379–383. ISBN 978-1-72815-414-5. Available from DOI: 10.1109/IECON43393.2020.9254381.

42. KENNEDY, J.; EBERHART, R. *Particle swarm optimization*. Vol. 4. IEEE, 1995. ISBN 0-7803-2768-3. Available from DOI: 10.1109/ICNN.1995.488968. Publication Title: Proceedings of ICNN'95 - International Conference on Neural Networks.

43. TAMILSELVI, S.; BASKAR, S.; SIVAKUMAR, T.; ANANDAPADMANABAN, L. Evolutionary algorithm-based design optimization for right choice of transformer conductor material and stepped core. *Electrical Engineering*. 2019, vol. 101, no. 1, pp. 259–277. ISSN 14320487. Available from DOI: 10.1007/S00202-019-00771-3/FIGURES/8. Publisher: Springer Verlag.

44. KRÄMER, Christian; KUGI, Andreas; KEMMETMÜLLER, Wolfgang. Modeling of a permanent magnet linear synchronous motor using magnetic equivalent circuits. *Mechatronics*. 2021, vol. 76, p. 102558. ISSN 0957-4158. Available from DOI: 10.1016/J.MECHATRONICS.2021.102558. Publisher: Pergamon.

45. KIM, J. K.; JOO, S. W.; HAHN, S. C.; HONG, J. P.; KANG, D. H.; KOO, D. H. Static characteristics of linear BLDC motor using equivalent magnetic circuit and finite element method. *IEEE Transactions on Magnetics*. 2004, vol. 40, no. 2 II, pp. 742–745. ISSN 00189464. Available from DOI: 10.1109/TMAG.2004.825033.

46. VAEZ-ZADEH, Sadegh; ISFAHANI, A. Hassanpour. Multiobjective design optimization of air-core linear permanent-magnet synchronous motors for improved thrust and low magnet consumption. *IEEE Transactions on Magnetics*. 2006, vol. 42, no. 3, pp. 446–452. ISSN 00189464. Available from DOI: 10.1109/TMAG.2005.863084.

47. KANO, Yoshiaki; KOSAKA, Takashi; MATSUI, Nobuyuki. A simple nonlinear magnetic analysis for axial-flux permanent-magnet machines. *IEEE Transactions on Industrial Electronics*. 2010, vol. 57, no. 6, pp. 2124–2133. ISSN 02780046. Available from DOI: 10.1109/TIE.2009.2034685.

48. LIM, Ki Chae; WOO, Joon Keun; KANG, Gyu Hong; HONG, Jung Pyo; KIM, Gyu Tak. Detent force minimization techniques in permanent magnet linear synchronous motors. *IEEE Transactions on Magnetics*. 2002, vol. 38, no. 2 I, pp. 1157–1160. ISSN 00189464. Available from DOI: 10.1109/20.996296.

49. VAEZ-ZADEH, S.; ISFAHANI, A. Hassanpour. Enhanced modeling of linear permanent-magnet synchronous motors. *IEEE Transactions on Magnetics*. 2007, vol. 43, no. 1, pp. 33–39. ISSN 00189464. Available from DOI: 10.1109/TMAG.2006.886970.

50. KANG, Gyu Hong; HONG, Jung Pyo; KIM, Gyu Tak. A novel design of an air-core type permanent magnet linear brushless motor by space harmonics field analysis. *IEEE Transactions on Magnetics*. 2001, vol. 37, no. 5 I, pp. 3732–3736. ISSN 00189464. Available from DOI: 10.1109/20.952701. Publisher: Institute of Electrical and Electronics Engineers Inc.

51. PILLAY, P; KRISHNAN, R. Modeling, simulation, and analysis of permanent-magnet motor drives. Part I: The permanent-magnet synchronous motor drive. *IEEE Transactions on Industry Applications*. 1989, vol. 25, no. 2, pp. 265–273.

52. YAHIAOUI, Maamar; KECHICH, Abderrahmane; KHALIL BOUSERHANE, Ismail. Adaptive Sliding Mode Control of PMLSM Drive. *International Journal of Power Electronics and Drive System (IJPEDS)*. 2017, vol. 8, no. 2, pp. 639–646. ISSN 2088-8694. Available from DOI: 10.11591/ijpeds.v8i2.pp639-646.

53. ENJETI, P; LINDSAY, J F; RASHID, M H. Stability and dynamic performance of variable speed permanent magnet synchronous motors. In: *Proc. IECON*. 1985, pp. 749–754.

54. PILLAY, Pragasen; KRISHNAN, Ramu. Control characteristics and speed controller design for a high performance permanent magnet synchronous motor drive. In: *1987 IEEE Power Electronics Specialists Conference*. IEEE, 1987, pp. 598–606.

55. HARNEFORS, Lennart; NEE, Hans Peter. Model-based current control of ac machines using the internal model control method. *IEEE Transactions on Industry Applications*. 1998, vol. 34, no. 1, pp. 133–141. ISSN 00939994. Available from DOI: 10.1109/28.658735.

56. LABORDA, Diego F.; DIAZ REIGOSA, David; FERNANDEZ, Daniel; SASAKI, Kensuke; KATO, Takashi; BRIZ, Fernando. Enhanced Torque Estimation in Variable Leakage Flux PMSM Combining High and Low Frequency Signal Injection. In: *ECCE 2020 - IEEE Energy Conversion Congress and Exposition*.

2020, pp. 1764–1771. ISBN 978-1-72815-826-6. Available from DOI: 10.1109/ECCE44975.2020.9235869.

57. PERUTKA, Karel. *MATLAB for Engineers Applications in Control, Electrical Engineering, IT and Robotics*. In Tech, 2011. ISBN 978-953-307-914-1.
58. NEBORAK, I. *Modelování a simulace elektrických regulovaných pohonů*. Vysoká škola báňská - Technická univerzita Ostrava, 2002.
59. SETHUPATHI, P.; SENTHILNATHAN, N. Comparative analysis of line-start permanent magnet synchronous motor and squirrel cage induction motor under customary power quality indices. *Electrical Engineering*. 2020, vol. 102, no. 3, pp. 1339–1349. ISSN 14320487. Available from DOI: 10.1007/s00202-020-00955-2. Publisher: Springer.
60. ISFAHANI, Arash Hassanpour; VAEZ-ZADEH, Sadegh; AZIZUR RAHMAN, M. Using modular poles for shape optimization of flux density distribution in permanent-magnet machines. *IEEE Transactions on Magnetics*. 2008, vol. 44, no. 8, pp. 2009–2015. ISSN 00189464. Available from DOI: 10.1109/TMAG.2008.922781.
61. DE LA BARRIERE, O.; HLIOUI, S.; BEN AHMED, H.; GABSI, M.; LOBUE, M. Three-dimensional analytical modeling of a permanent-magnet linear actuator with circular magnets. *IEEE Transactions on Magnetics*. 2010, vol. 46, no. 9, pp. 3608–3616. ISSN 00189464. Available from DOI: 10.1109/TMAG.2010.2045507.
62. TAVANA, Nariman Roshandel; SHOULAIE, Abbas. Analysis and design of magnetic pole shape in linear permanent-magnet machine. *IEEE Transactions on Magnetics*. 2010, vol. 46, no. 4, pp. 1000–1006. ISSN 00189464. Available from DOI: 10.1109/TMAG.2009.2037951.
63. ISFAHANI, A. H.; VAEZ-ZADEH, S.; RAHMAN, M. A. Performance improvement of permanent magnet machines by modular poles. *IET Electric Power Applications*. 2009, vol. 3, no. 4, pp. 343–351. ISSN 17518660. Available from DOI: 10.1049/iet-epa.2007.0391.
64. ZHU, Z. Q.; HOWE, David; BOLTE, Ekkehard; ACKERMANN, Bernd. Instantaneous Magnetic Field Distribution in Brushless Permanent Magnet DC Motors, Part I: Open-Circuit Field. *IEEE Transactions on Magnetics*. 1993, vol. 29, no. 1, pp. 124–135. ISSN 19410069. Available from DOI: 10.1109/20.195557.

65. ZHU, Z. Q.; HOWE, D. Instantaneous magnetic field distribution in permanent magnet brushless dc motors, Part IV: Magnetic field on load. *IEEE Transactions on Magnetics*. 1993, vol. 29, no. 1, pp. 152–158. ISSN 19410069. Available from DOI: 10.1109/20.195560.

66. ZHU, Z. Q.; HOWE, D. Instantaneous magnetic field distribution in brushless permanent magnet dc motors, Part III: Effect of stator slotting. *IEEE Transactions on Magnetics*. 1993, vol. 29, no. 1, pp. 143–151. ISSN 19410069. Available from DOI: 10.1109/20.195559.

67. ZHU, Z. Q.; HOWE, David. Instantaneous Magnetic Field Distribution in Brushless Permanent Magnet DC Motors, Part II: Armature-Reaction Field. *IEEE Transactions on Magnetics*. 1993, vol. 29, no. 1, pp. 136–142. ISSN 19410069. Available from DOI: 10.1109/20.195558.

68. LIU, Z. J.; LI, J. T. Analytical solution of air-gap field in permanent-magnet motors taking into account the effect of pole transition over slots. *IEEE Transactions on Magnetics*. 2007, vol. 43, no. 10, pp. 3872–3883. ISSN 00189464. Available from DOI: 10.1109/TMAG.2007.903417.

69. PLATEN, M.; HENNEBERGER, G. Examination of leakage and end effects in a linear synchronous motor for vertical transportation by means of finite element computation. *IEEE Transactions on Magnetics*. 2001, vol. 37, no. 5 I, pp. 3640–3643. ISSN 00189464. Available from DOI: 10.1109/20.952680.

70. URRESTY, Julio Csar; RIBA, Jordi Roger; ROMERAL, Lus; GARCIA, Antonio. A simple 2-D finite-element geometry for analyzing surface-mounted synchronous machines with skewed rotor magnets. *IEEE Transactions on Magnetics*. 2010, vol. 46, no. 11, pp. 3948–3954. ISSN 00189464. Available from DOI: 10.1109/TMAG.2010.2062529.

71. ONAT, Ahmet; KAZAN, Ender; TAKAHASHI, Norio; MIYAGI, Daisuke; KOMATSU, Yasuhiro; MARKON, Sandor. Design and implementation of a linear motor for multicar elevators. *IEEE/ASME Transactions on Mechatronics*. 2010, vol. 15, no. 5, pp. 685–693. ISSN 10834435. Available from DOI: 10.1109/TMECH.2009.2031815.

72. LI, A. Liyi; MA, B. Mingna; CHEN, C. Qingquan. Detent force analysis in permanent magnet linear synchronous motor considering longitudinal end effects. In: *ICEMS 2012 - Proceedings: 15th International Conference on Electrical Machines and Systems*. 2012. ISBN 978-4-88686-077-4. Available from DOI: 10.11142/jicems.2013.2.1.9. ISSN: 2234-6902.

73. CUI, Jiefan; WANG, Chengyuan; YANG, Junyou; LIU, Lifeng. Analysis of direct thrust force control for permanent magnet linear synchronous motor. *Proceedings of the World Congress on Intelligent Control and Automation (WCICA)*. 2004, vol. 5, pp. 4418–4421. Available from DOI: `10.1109/WCICA.2004.1342349`.

74. ZHU, Yu Wu; CHO, Yun Hyun. Thrust ripples suppression of permanent magnet linear synchronous motor. In: *IEEE Transactions on Magnetics*. 2007, vol. 43, pp. 2537–2539. Available from DOI: `10.1109/TMAG.2007.893308`. Issue: 6 ISSN: 00189464.

75. LIN, Faa Jeng; SHYU, Kuo Kai; LIN, Chih Hong. Incremental motion control of linear synchronous motor. *IEEE Transactions on Aerospace and Electronic Systems*. 2002, vol. 38, no. 3, pp. 1011–1022. ISSN 00189251. Available from DOI: `10.1109/TAES.2002.1039417`.

76. GUO, Qingding; GUO, Wei; ZHOU, Yue; WANG, Limei. Preview feedforward compensation of permanent magnet linear synchronous motor servo system implemented with Adaline. *International Workshop on Advanced Motion Control, AMC*. 2000, pp. 576–579. Available from DOI: `10.1109/AMC.2000.862942`. Publisher: IEEE.

77. DENG, Z.; NASAR, S. A.; BOLDEA, I. Forces and parameters of permanent magnet linear synchronous machines. *IEEE Transactions on Magnetics*. 1987, vol. 23, no. 1, pp. 305–309. ISSN 19410069. Available from DOI: `10.1109/TMAG.1987.1064775`.

78. ZHANG, Xi; PAN, Junmin. Nonlinear Robust Sliding Mode Control for PM Linear Synchronous Motors. In: *2006 CES/IEEE 5th International Power Electronics and Motion Control Conference*. Institute of Electrical and Electronics Engineers (IEEE), 2009, pp. 1–5. Available from DOI: `10.1109/ipemc.2006.4777988`.

79. ZHU, Yu Wu; KOO, Dae Hyun; CHO, Yun Hyun. Detent force minimization of permanent magnet linear synchronous motor by means of two different methods. In: *IEEE Transactions on Magnetics*. 2008, vol. 44, pp. 4345–4348. Available from DOI: `10.1109/TMAG.2008.2001320`. Issue: 11 PART 2 ISSN: 00189464.

80. KAZAN, Ender; ONAT, Ahmet. Modeling of air core permanent-magnet linear motors with a simplified nonlinear magnetic analysis. *IEEE Transactions on Magnetics*. 2011, vol. 47, no. 6 PART 2, pp. 1753–1762. ISSN 00189464. Available from DOI: `10.1109/TMAG.2011.2111375`.

81. WANG, Mingyi; LI, Liyi; PAN, Donghua. Detent force compensation for PMLSM systems based on structural design and control method combination. *IEEE Transactions on Industrial Electronics*. 2015, vol. 62, no. 11, pp. 6845–6854. ISSN 02780046. Available from DOI: 10.1109/TIE.2015.2443096. Publisher: Institute of Electrical and Electronics Engineers Inc.

82. PÉREZ, J. N.H.; HERNANDEZ, O. S.; CAPORAL, R. M.; MAGDALENO, J. De J.R.; BARRETO, H. P. Parameter identification of a permanent magnet synchronous machine based on current decay test and particle swarm optimization. *IEEE Latin America Transactions*. 2013, vol. 11, no. 5, pp. 1176–1181. ISSN 15480992. Available from DOI: 10.1109/TLA.2013.6684392.

83. HORNING, S.; KEYHANI, A.; KAMWA, I. On-line evaluation of a round rotor synchronous machine parameter set estimated from standstill time-domain data. *IEEE Transactions on Energy Conversion*. 1997, vol. 12, no. 4, pp. 289–296. ISSN 08858969. Available from DOI: 10.1109/60.638863.

84. ZHANG, Jinhui; RADUN, Arthur V. A new method to measure the switched reluctance motor's flux. *IEEE Transactions on Industry Applications*. 2006, vol. 42, no. 5, pp. 1171–1176. ISSN 00939994. Available from DOI: 10.1109/TIA.2006.880876.

85. KILTHAU, A.; PACAS, J. M. Parameter-measurement and control of the synchronous reluctance machine including cross saturation. *Conference Record - IAS Annual Meeting (IEEE Industry Applications Society)*. 2001, vol. 4, pp. 2302–2309. ISSN 01972618. Available from DOI: 10.1109/IAS.2001.955945.

86. SELLSCHOPP, F. S.; ARJONA L., M. A. A tool for extracting synchronous machines parameters from the dc flux decay test. *Computers & Electrical Engineering*. 2005, vol. 31, no. 1, pp. 56–68. ISSN 0045-7906. Available from DOI: 10.1016/J.COMPELECENG.2004.10.001. Publisher: Pergamon.

87. GAO, Yuting; QU, Ronghai; LIU, Yang. An improved AC standstill method for inductance measurement of interior permanent magnet synchronous motors. *2013 International Conference on Electrical Machines and Systems, ICEMS 2013*. 2013, pp. 927–931. Available from DOI: 10.1109/ICEMS.2013.6754404. Publisher: IEEE Computer Society ISBN: 9781479914470.

88. CAVAGNINO, A.; LAZZARI, M.; PROFUMO, F.; TENCONI, A. Axial flux interior PM synchronous motor: parameters identification and steady-state performance measurements. *Conference Record - IAS Annual Meeting (IEEE Industry Applications Society)*. 1999, vol. 4, pp. 2552–2559. ISSN 01972618. Available from DOI: 10.1109/IAS.1999.799199.

89. BALDA, J. C.; FAIRBAIRN, R. E.; HARLEY, R. G.; RODGERSON, J. L.; EITELBERG, E. Measurement of synchronous machine parameters by a modified frequency response method - part II: Measured results. *IEEE Transactions on Energy Conversion*. 1987, vol. EC-2, no. 4, pp. 652–657. ISSN 15580059. Available from DOI: 10.1109/TEC.1987.4765905.

90. TOUHAMI, O.; GUESBAOUI, H.; IUNG, C. Synchronous machine parameter identification by a multi-time scale technique. *Conference Record - IAS Annual Meeting (IEEE Industry Applications Society)*. 1992, vol. 1992-Janua, pp. 178–184. ISSN 01972618. Available from DOI: 10.1109/IAS.1992.244297. Publisher: Institute of Electrical and Electronics Engineers Inc. ISBN: 078030635X.

91. AHN, Hanwoong; PARK, · Hyunjong; KIM, Changhyun; LEE, Hyungwoo. A Review of State-of-the-art Techniques for PMSM Parameter Identification. *Journal of Electrical Engineering & Technology*. 2020, vol. 15, no. 3, pp. 1177–1187. Available from DOI: 10.1007/s42835-020-00398-6. ISBN: 0123456789.

92. KUMAR, Rajesh; GUPTA, R. A.; BANSAL, Ajay Kr. Identification and control of PMSM using artificial neural network. *IEEE International Symposium on Industrial Electronics*. 2007, pp. 30–35. Available from DOI: 10.1109/ISIE.2007.4374567. ISBN: 1424407559.

93. LIU, Kan; ZHANG, Qiao; CHEN, Jintao; ZHU, Z. Q.; ZHANG, Jing. Online multiparameter estimation of nonsalient-pole PM synchronous machines with temperature variation tracking. *IEEE Transactions on Industrial Electronics*. 2011, vol. 58, no. 5, pp. 1776–1788. ISSN 02780046. Available from DOI: 10.1109/TIE.2010.2054055.

94. LIU, Kan; ZHU, Z. Q. Online estimation of the rotor flux linkage and voltage-source inverter nonlinearity in permanent magnet synchronous machine drives. *IEEE Transactions on Power Electronics*. 2014, vol. 29, no. 1, pp. 418–427. ISSN 08858993. Available from DOI: 10.1109/TPEL.2013.2252024.

95. LIU, Kan; ZHU, Z. Q.; STONE, David A. Parameter estimation for condition monitoring of PMSM stator winding and rotor permanent magnets. *IEEE Transactions on Industrial Electronics*. 2013, vol. 60, no. 12, pp. 5902–5913. ISSN 02780046. Available from DOI: 10.1109/TIE.2013.2238874. Publisher: Institute of Electrical and Electronics Engineers Inc.

96. NEUGEBAUER, Henrik. *Parameter identification of a permanent magnet synchronous motor*. 2012. Master's Thesis. Chalmers University of Technology.

97. HUFFEL, Sabine Van; VANDEWALLE, Joos. Analysis and Properties of the Generalized Total Least Squares Problem $AX \approx B$ When Some or All Columns in A are Subject to Error. *SIAM Journal on Matrix Analysis and Applications*. 1989, vol. 10, no. 3, pp. 294–315. ISSN 0895-4798. Available from DOI: 10.1137/0610023. Publisher: Society for Industrial & Applied Mathematics (SIAM).

98. GOLUB, Gene H.; LOAN, Charles F. Van. *Matrix Computations*. Johns Hopkins University Press, 2013. ISBN 978-1-4214-0794-4 978-1-4214-0859-0. Available from DOI: 10.56021/9781421407944.

99. BISHOP, Christopher M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006. ISBN 0-387-31073-8.

100. ROSSI, Peter E.; ALLENBY, Greg M.; MCCULLOCH, Robert. Bayesian Statistics and Marketing. *Bayesian Statistics and Marketing*. 2006, pp. 1–348. Available from DOI: 10.1002/0470863692. Publisher: wiley ISBN: 9780470863695.

101. BINGHAM, Christopher. An Antipodally Symmetric Distribution on the Sphere. *The Annals of Statistics*. 1974, vol. 2, no. 6, pp. 1201–1225. ISSN 00905364. Publisher: Institute of Mathematical Statistics.

102. STRELITZ, Richard A. Computing the normalization factor in bingham distributions. *http://dx.doi.org/10.1080/00949658908811176*. 2007, vol. 32, no. 4, pp. 193–200. ISSN 15635163. Available from DOI: 10.1080/00949658908811176. Publisher: Gordon and Breach Science Publishers.

103. ONSTOTT, T. C. Application of the Bingham distribution function in palaeomagnetic studies. *Journal of Geophysical Research*. 1980, vol. 85, no. B3, pp. 1500–1510. ISSN 01480227. Available from DOI: 10.1029/JB085iB03p01500. Publisher: John Wiley & Sons, Ltd.

104. GOH, S. L.; CHEN, M.; POPOVIĆ, D. H.; AIHARA, K.; OBRADOVIC, D.; MANDIC, D. P. Complex-valued forecasting of wind profile. *Renewable Energy*. 2006, vol. 31, no. 11, pp. 1733–1750. ISSN 09601481. Available from DOI: 10.1016/j.renene.2005.07.006. Publisher: Pergamon.

105. ZHANG, Xi; PAN, Junmin. Nonlinear Robust Sliding Mode Control for PM Linear Synchronous Motors. In: Institute of Electrical and Electronics Engineers (IEEE), 2009, pp. 1–5. Available from DOI: 10.1109/ipemc.2006.4777988.

106. BINGHAM, Melissa A.; LOGRASSO, Barbara K.; LAABS, Francis C. A statistical analysis of the variation in measured crystal orientations obtained through electron backscatter diffraction. *Ultramicroscopy*. 2010, vol. 110, no. 10, pp. 1312–1319. ISSN 03043991. Available from DOI: `10.1016/j.ultramic.2010.06.001`. Publisher: North-Holland.

107. GILITSCHENSKI, Igor; KURZ, Gerhard; JULIER, Simon J.; HANEBECK, Uwe D. Unscented Orientation Estimation Based on the Bingham Distribution. *IEEE Transactions on Automatic Control*. 2016, vol. 61, no. 1, pp. 172–177. ISSN 00189286. Available from DOI: `10.1109/TAC.2015.2423831`. arXiv: 1311.5796 Publisher: Institute of Electrical and Electronics Engineers Inc.

108. KUME, Alfred; WALKER, Stephen G. Sampling from compositional and directional distributions. *Statistics and Computing*. 2006, vol. 16, no. 3, pp. 261–265. ISSN 09603174. Available from DOI: `10.1007/s11222-006-8077-9`. Publisher: Springer.

109. MARDIA, Kanti V.; JUPP, Peter E. Directional Statistics. *Directional Statistics*. 2008, pp. 1–432. Available from DOI: `10.1002/9780470316979`. Publisher: John Wiley and Sons Ltd. ISBN: 9780470316979.

110. HILLEN, Thomas; PAINTER, Kevin J.; SWAN, Amanda C.; MURTHA, Albert D. Moments of von Mises and Fisher distributions and applications. *Mathematical Biosciences and Engineering*. 2017, vol. 14, no. 3, pp. 673–694. ISSN 15510018. Available from DOI: `10.3934/mbe.2017038`. Publisher: Mathematical Biosciences & Engineering.

111. LUO, Yixiang; XU, Jie; ZHANG, Pingwen. A Fast Algorithm for the Moments of Bingham Distribution. *Journal of Scientific Computing*. 2018, vol. 75, no. 3, pp. 1337–1350. ISSN 08857474. Available from DOI: `10.1007/s10915-017-0589-2`. arXiv: 1612.01015 Publisher: Springer New York LLC.

112. LOVE, J. J. Paleomagnetic Principles and Practice. *Eos, Transactions American Geophysical Union*. 2000, vol. 81, no. 16, pp. 172–172. Available from DOI: `10.1029/00eo00122`. Publisher: Kluwer Academic Publishers ISBN: 0306481286.

113. MØLLER, J.; PETTITT, A. N.; REEVES, R.; BERTHELSEN, K. K. Miscellanea an efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika*. 2006, vol. 93, no. 2, pp. 451–458. ISSN 00063444. Available from DOI: `10.1093/biomet/93.2.451`. Publisher: Oxford Academic.

114. MURRAY, Iain; GHAHRAMANI, Zoubin; MACKAY, David J.C. MCMC for doubly-intractable distributions. In: *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence, UAI 2006*. 2006, pp. 359–366. ISBN 0-9749039-2-2. arXiv: 1206.6848.

115. KUME, A.; WALKER, S. G. On the Bingham distribution with large dimension. *Journal of Multivariate Analysis*. 2014, vol. 124, pp. 345–352. ISSN 0047259X. Available from DOI: 10.1016/j.jmva.2013.10.023. Publisher: Academic Press.

116. FALLAIZE, Christopher J.; KYPRAIOS, Theodore. Exact Bayesian inference for the Bingham distribution. *Statistics and Computing*. 2016, vol. 26, no. 1-2, pp. 349–360. ISSN 15731375. Available from DOI: 10.1007/s11222-014-9508-7. arXiv: 1401.2894 Publisher: Springer New York LLC.

117. TSIONAS, Mike G. Note on posterior inference for the Bingham distribution. *Communications in Statistics - Theory and Methods*. 2018, vol. 47, no. 12, pp. 3022–3028. ISSN 1532415X. Available from DOI: 10.1080/03610926.2017.1346805. Publisher: Taylor & Francis.

118. GREEN, Peter J. Reversible jump Markov chain monte carlo computation and Bayesian model determination. *Biometrika*. 1995, vol. 82, no. 4, pp. 711–732. ISSN 00063444. Available from DOI: 10.1093/biomet/82.4.711. Publisher: Oxford Academic.

119. KURZ, Gerhard; GILITSCHENSKI, Igor; PFAFF, Florian; DRUDE, Lukas; HANEBECK, Uwe D; HAEB-UMBACH, Reinhold; SIEGWART, Roland Y. Directional Statistics and Filtering Using {libDirectional}. *Journal of Statistical Software*. 2019, vol. 89, no. 4, pp. 1–31. Available from DOI: 10.18637/jss.v089.i04.

120. RHODE, Stephan; BLEIMUND, Felix; GAUTERIN, Frank. Recursive generalized total least squares with noise covariance estimation. In: *IFAC Proceedings Volumes (IFAC-PapersOnline)*. Elsevier, 2014, vol. 19, pp. 4637–4643. ISBN 978-3-902823-62-5. Available from DOI: 10.3182/20140824-6-za-1003.01568. Issue: 3 ISSN: 14746670.

121. LJUNG, Lennart. *System identification*. Univ., 1995.

122. VAN HUFFEL, Sabine. The Generalized Total Least Squares Problem : Formulation, Algorithm and Properties. In: *Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms*. Springer, Berlin, Heidelberg, 1991, pp. 651–660. Available from DOI: 10.1007/978-3-642-75536-1_54.

123. MARKOVSKY, Ivan; VAN HUFFEL, Sabine. Overview of total least-squares methods. *Signal Processing*. 2007, vol. 87, no. 10, pp. 2283–2302. ISSN 01651684. Available from DOI: 10.1016/j.sigpro.2007.04.004. Publisher: Elsevier.

124. GOLUB, Gene H.; LOAN, Charles F. van. An Analysis of the Total Least Squares Problem. *SIAM Journal on Numerical Analysis*. 1980, vol. 17, no. 6, pp. 883–893. ISSN 0036-1429. Available from DOI: 10.1137/0717073. Publisher: Society for Industrial and Applied Mathematics.

125. SCHUERMANS, M.; MARKOVSKY, I.; WENTZELL, Peter D.; VAN HUFFEL, S. On the equivalence between total least squares and maximum likelihood PCA. In: *Analytica Chimica Acta*. Elsevier, 2005, vol. 544, pp. 254–267. Available from DOI: 10.1016/j.aca.2004.12.059. Issue: 1-2 SPEC. ISS. ISSN: 00032670.

126. MARKOVSKY, Ivan; VAN HUFFEL, Sabine. Overview of total least-squares methods. *Signal Processing*. 2007, vol. 87, pp. 2283–2302. Available from DOI: 10.1016/j.sigpro.2007.04.004.

127. YAN, Shijian; FAN, Jinyan. The solution set of the mixed ls-tls problem. *International Journal of Computer Mathematics*. 2001, vol. 77, no. 4, pp. 545–561. ISSN 0020-7160, ISSN 1029-0265. Available from DOI: 10.1080/00207160108805084.

128. GLESER, Leon Jay. *Calculation and simulation in errors-in-variables regression problems*. Purdue University. Department of Statistics, 1978.

129. GLESER, Leon Jay. Estimation in a Multivariate "Errors in Variables" Regression Model: Large Sample Results. *The Annals of Statistics*. 1981, vol. 9, no. 1, pp. 24–44. ISSN 0090-5364. Publisher: Institute of Mathematical Statistics.

130. HIGHAM, Nicholas J. Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and its Applications*. 1988, vol. 103, pp. 103–118. ISSN 0024-3795. Available from DOI: 10.1016/0024-3795(88)90223-6.

131. BERKSON, Joseph. Are there Two Regressions? *Journal of the American Statistical Association*. 1950, vol. 45, no. 250, pp. 164–180. ISSN 0162-1459, ISSN 1537-274X. Available from DOI: 10.1080/01621459.1950.10483349.

132. DELLAPORTAS, Petros; STEPHENS, David A. Bayesian Analysis of Errors-in-Variables Regression Models. *Biometrics*. 1995, vol. 51, no. 3, pp. 1085–1095. ISSN 0006-341X. Available from DOI: 10.2307/2533007. Publisher: [Wiley, International Biometric Society].

133. SÖDERSTRÖM, Torsten. Errors-in-variables methods in system identification. *Automatica*. 2007, vol. 43, no. 6, pp. 939–958. ISSN 00051098. Available from DOI: `10.1016/j.automatica.2006.11.025`. Publisher: Pergamon.

134. ROZLIMAN, Nur Aainaa; IBRAHIM, Adriana Irawati Nur; YUNUS, Rossita Mohammad. Bayesian approach to errors-in-variables in regression models. In: *AIP Conference Proceedings*. AIP Publishing LLC AIP Publishing, 2017, vol. 1842, p. 030018. ISBN 978-0-7354-1512-6. Available from DOI: `10.1063/1.4982856`. Issue: 1 ISSN: 15517616.

135. MALLICK, Bani K.; GELFAND, Alan E. Semiparametric errors-in-variables models A Bayesian approach. *Journal of Statistical Planning and Inference*. 1996, vol. 52, no. 3, pp. 307–321. ISSN 0378-3758. Available from DOI: `10.1016/0378-3758(95)00139-5`.

136. CARROLL, R. J. Approaches to Estimation with Errors in Predictors. In: FAHRMEIR, Ludwig; FRANCIS, Brian; GILCHRIST, Robert; TUTZ, Gerhard (eds.). *Advances in GLIM and Statistical Modelling*. New York, NY: Springer, 1992, pp. 40–47. Lecture Notes in Statistics. ISBN 978-1-4612-2952-0. Available from DOI: `10.1007/978-1-4612-2952-0_7`.

137. FULLER, Wayne A. Measurement error models. 1987, p. 440. Publisher: Wiley ISBN: 0470317337.

138. CARROLL, Raymond J.; STEFANSKI, Leonard A. Approximate Quasi-likelihood Estimation in Models with Surrogate Predictors. *Journal of the American Statistical Association*. 1990, vol. 85, no. 411, pp. 652–663. ISSN 0162-1459. Available from DOI: `10.1080/01621459.1990.10474925`. Publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1990.10474925.

139. FANG, Xing; LI, Bofeng; ALKHATIB, Hamza; ZENG, Wenxian; YAO, Yibin. Bayesian inference for the Errors-In-Variables model. *Studia Geophysica et Geodaetica*. 2017, vol. 61, no. 1, pp. 35–52. ISSN 15731626. Available from DOI: `10.1007/s11200-015-6107-9`.

140. RHODE, Stephan; USEVICH, Konstantin; MARKOVSKY, Ivan; GAUTERIN, Frank. A recursive restricted total least-squares algorithm. *IEEE Transactions on Signal Processing*. 2014, vol. 62, no. 21, pp. 5652–5662. ISSN 1053587X. Available from DOI: `10.1109/TSP.2014.2350959`.

141. CASELLA, George; BERGER, Roger. *Statistical inference*. CRC Press, 2024.

142. KONG, Xiangyu; FENG, Dazheng. *Efficient Online Learning Algorithms for Total Least Square Problems*. Springer, 2024.

143. FOX, Charles W.; ROBERTS, Stephen J. A tutorial on variational Bayesian inference. *Artificial Intelligence Review 2011 38:2*. 2011, vol. 38, no. 2, pp. 85–95. ISSN 1573-7462. Available from DOI: 10.1007/S10462-011-9236-8. Publisher: Springer.

144. GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. Approximate Inference. In: *Deep Learning*. MIT Press, 2016, pp. 629–650. Section: 19.

145. PENNY, W; KIEBEL, S; FRISTON, K. *Variational bayes*. Elsevier, London, 2006.

146. NAVAS-PALENCIA, Guillermo. Extending error function and related functions to complex arguments. 2016.

147. POPPE, G. P.M.; WIJERS, C. M.J. More Efficient Computation of the Complex Error Function. *ACM Transactions on Mathematical Software (TOMS)*. 1990, vol. 16, no. 1, pp. 38–46. ISSN 15577295. Available from DOI: 10.1145/77626.77629.

148. AZAH, Mohammad A.L.; CHANDLER-WILDE, Simon N. Computation of the complex error function using modified Trapezoidal rules. *SIAM Journal on Numerical Analysis*. 2021, vol. 59, no. 5, pp. 2346–2367. ISSN 00361429. Available from DOI: 10.1137/20M1373037. arXiv: 2010.05659.

149. MARQUARDT, Donald W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*. 1963, vol. 11, no. 2, pp. 431–441. ISSN 0368-4245. Available from DOI: 10.1137/0111030. Publisher: Society for Industrial and Applied Mathematics.

150. MORÉ, Jorge J. The Levenberg-Marquardt algorithm: Implementation and theory. In: *Numerical Analysis: Proceedings of the Biennial Conference Held at Dundee*. Springer, Berlin, Heidelberg, 1978, pp. 105–116. Available from DOI: 10.1007/bfb0067700.

151. BELLAVIA, Stefania; GRATTON, Serge; RICCIETTI, Elisa. A Levenberg–Marquardt method for large nonlinear least-squares problems with dynamic accuracy in functions and gradients. *Numerische Mathematik*. 2018, vol. 140, pp. 791–825. Publisher: Springer.

152. LI, Xinyue; KENNEL, Ralph. Comparison of state-of-the-art estimators for electrical parameter identification of PMSM. In: *Proceedings - PRECEDE 2019: 2019 IEEE International Symposium on Predictive Control of Electrical Drives and Power Electronics*. Institute of Electrical and Electronics Engineers Inc., 2019. ISBN 978-1-5386-9414-5. Available from DOI: 10.1109/PRECEDE.2019.8753197.

153. VAN HUFFEL, Sabine; LEMMERLING, Philippe. *Total least squares and errors-in-variables modeling: analysis, algorithms and applications*. Springer Science & Business Media, 2013.

154. PLETT, Gregory L. Recursive approximate weighted total least squares estimation of battery cell total capacity. *Journal of Power Sources*. 2011, vol. 196, no. 4, pp. 2319–2331. ISSN 0378-7753. Available from DOI: 10.1016/J.JPOWSOUR.2010.09.048. Publisher: Elsevier.

155. BARTIK, Ondrej; BUCHTA, Ludek. Angle tracking observer for the velocity estimation with the filtered control variable. *IFAC-PapersOnLine*. 2019, vol. 52, no. 27, pp. 496–501. ISSN 2405-8963. Available from DOI: 10.1016/j.ifacol.2019.12.712.

156. CRASSIDIS, John L.; CHENG, Yang. Maximum Likelihood Analysis of the Total Least Squares Problem with Correlated Errors. *https://doi.org/10.2514/1.G003815*. 2019, vol. 42, no. 6, pp. 1204–1217. Available from DOI: 10.2514/1.G003815. Publisher: American Institute of Aeronautics and Astronautics.