1

PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Vostra - Your reference

Votre - Your reference

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

**Sliding Mode State Estimation for
Nonlinear Discrete-time Systems:
Applications in Image Sequence Analysis**

by

Victor C. Aitken

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy, Electrical Engineering
Department of Systems and Computer Engineering

Carleton University

Ottawa, Ontario

April, 1995

© Copyright

1995, Victor C. Aitken.

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

THE AUTHOR HAS GRANTED AN
IRREVOCABLE NON-EXCLUSIVE
LICENCE ALLOWING THE NATIONAL
LIBRARY OF CANADA TO
REPRODUCE, LOAN, DISTRIBUTE OR
SELL COPIES OF HIS/HER THESIS BY
ANY MEANS AND IN ANY FORM OR
FORMAT, MAKING THIS THESIS
AVAILABLE TO INTERESTED
PERSONS.

L'AUTEUR A ACCORDE UNE LICENCE
IRREVOCABLE ET NON EXCLUSIVE
PERMETTANT A LA BIBLIOTHEQUE
NATIONALE DU CANADA DE
REPRODUIRE, PRETER, DISTRIBUER
OU VENDRE DES COPIES DE SA
THESE DE QUELQUE MANIERE ET
SOUS QUELQUE FORME QUE CE SOIT
POUR METTRE DES EXEMPLAIRES DE
CETTE THESE A LA DISPOSITION DES
PERSONNE INTERESSEES.

THE AUTHOR RETAINS OWNERSHIP
OF THE COPYRIGHT IN HIS/HER
THESIS. NEITHER THE THESIS NOR
SUBSTANTIAL EXTRACTS FROM IT
MAY BE PRINTED OR OTHERWISE
REPRODUCED WITHOUT HIS/HER
PERMISSION.

L'AUTEUR CONSERVE LA PROPRIETE
DU DROIT D'AUTEUR QUI PROTEGE
SA THESE. NI LA THESE NI DES
EXTRAITS SUBSTANTIELS DE CELLE-
CI NE DOIVENT ETRE IMPRIMES OU
AUTREMENT REPRODUITS SANS SON
AUTORISATION.

ISBN 0-612-02938-7

Canada

Name

VICTOR AITKEN

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

ENGINEERING: General

SUBJECT TERM

0537

SUBJECT CODE

U·M·I

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS

Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0798
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION

General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523

Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS

Language 0679
 General 0289
 Ancient 0290
 Linguistics 0291
 Modern
Literature 0401
 General 0294
 Classical 0295
 Comparative 0297
 Medieval 0298
 Modern 0316
 African 0591
 American 0305
 Asian 0352
 Canadian (English) 0355
 Canadian (French) 0593
 English 0311
 Germanic 0312
 Latin American 0315
 Middle Eastern 0313
 Romance 0314
 Slavic and East European

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy 0422
Religion
 General 0318
 Biblical Studies 0321
 Clergy 0319
 History of 0320
 Philosophy of 0322
Theology 0469

SOCIAL SCIENCES

American Studies 0323
Anthropology
 Archaeology 0324
 Cultural 0326
 Physical 0327
Business Administration
 General 0310
 Accounting 0272
 Banking 0770
 Management 0454
 Marketing 0338
Canadian Studies 0385
Economics
 General 0501
 Agricultural 0503
 Commerce-Business 0505
 Finance 0508
 History 0509
 Labor 0510
 Theory 0511
Folklore 0348
Geography 0366
Gerontology 0351
History
 General 0578

Ancient 0579
Medieval 0581
Modern 0582
Black 0328
African 0331
Asia, Australia and Oceania 0332
Canadian 0334
European 0335
Latin American 0336
Middle Eastern 0333
United States 0337
History of Science 0585
Law 0398
Political Science
 General 0615
 International Law and Relations 0616
 Public Administration 0617
Recreation 0814
Social Work 0452
Sociology
 General 0626
 Criminology and Penology 0627
 Demography 0938
 Ethnic and Racial Studies 0631
 Individual and Family Studies 0628
 Industrial and Labor Relations 0629
 Public and Social Welfare 0630
 Social Structure and Development 0700
 Theory and Methods 0344
Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture
 General 0473
 Agronomy 0285
 Animal Culture and Nutrition 0475
 Animal Pathology 0476
 Food Science and Technology 0359
 Forestry and Wildlife 0478
 Plant Culture 0479
 Plant Pathology 0480
 Plant Physiology 0817
 Range Management 0777
 Wood Technology 0746
Biology
 General 0306
 Anatomy 0287
 Biostatistics 0308
 Botany 0309
 Cell 0379
 Ecology 0329
 Entomology 0353
 Genetics 0369
 Limnology 0793
 Microbiology 0410
 Molecular 0307
 Neuroscience 0317
 Oceanography 0416
 Physiology 0433
 Radiation 0821
 Veterinary Science 0778
 Zoology 0472
Biophysics
 General 0786
 Medical 0760
EARTH SCIENCES
Biogeochemistry 0425
Geochemistry 0996

Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences 0768
Health Sciences
 General 0566
 Audiology 0300
 Chemotherapy 0992
 Dentistry 0567
 Education 0350
 Hospital Management 0769
 Human Development 0758
 Immunology 0982
 Medicine and Surgery 0564
 Mental Health 0347
 Nursing 0569
 Nutrition 0570
 Obstetrics and Gynecology 0380
 Occupational Health and Therapy 0354
 Ophthalmology 0381
 Pathology 0571
 Pharmacology 0419
 Pharmacy 0572
 Physical Therapy 0382
 Public Health 0573
 Radiology 0574
 Recreation 0575

Speech Pathology 0460
Toxicology 0383
Home Economics 0386

PHYSICAL SCIENCES

Pure Sciences
Chemistry
 General 0485
 Agricultural 0749
 Analytical 0486
 Biochemistry 0487
 Inorganic 0488
 Nuclear 0738
 Organic 0490
 Pharmaceutical 0491
 Physical 0494
 Polymer 0495
 Radiation 0754
Mathematics 0405
Physics
 General 0605
 Acoustics 0986
 Astronomy and Astrophysics 0606
 Atmospheric Science 0608
 Atomic 0748
 Electronics and Electricity 0607
 Elementary Particles and High Energy 0798
 Fluid and Plasma 0759
 Molecular 0609
 Nuclear 0610
 Optics 0752
 Radiation 0756
 Solid State 0611
Statistics 0463
Applied Sciences
Applied Mechanics 0346
Computer Science 0984

Engineering
 General 0537
 Aerospace 0538
 Agricultural 0539
 Automotive 0540
 Biomedical 0541
 Chemical 0542
 Civil 0543
 Electronics and Electrical 0544
 Heat and Thermodynamics 0348
 Hydraulic 0545
 Industrial 0546
 Marine 0547
 Materials Science 0794
 Mechanical 0548
 Metallurgy 0743
 Mining 0551
 Nuclear 0552
 Packaging 0549
 Petroleum 0765
 Sanitary and Municipal 0554
 System Science 0790
Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY

General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

The undersigned hereby recommend to
The Faculty of Graduate Studies and Research
acceptance of the thesis,

**Sliding Mode State Estimation for
Nonlinear Discrete-time Systems:
Applications in Image Sequence Analysis**

submitted by

Victor C. Aitken, B.A.Sc., M.Eng.
in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

Chairman, Department of Systems and Computer Engineering

Thesis Supervisor

External Examiner

Carleton University

April, 1995

Abstract

A dynamical system, often implemented numerically, whose purpose is to estimate unmeasurable states based on available measurements of a dynamical process is called an *observer*. Novel methods, based on the theory of dynamical systems with sliding modes, are proposed and analyzed for state estimation of uncertain, nonlinear, discrete-time systems. The proposed discrete-time sliding mode observers retain the fundamental property, familiar in continuous-time sliding mode control literature, of robustness to dynamical modelling errors. Design methodologies are proposed for linear or nonlinear, non-redundant or redundant measurement models. Methods for redundant measurements provide an optimal fusion of the available data during the sliding mode through minimization of a squared error cost measure. New methods to obtain exponential stability of nonlinear discrete-time systems are introduced with applications in stability and robustness analysis of the proposed sliding mode observers. For nonlinear measurements, ideal observer forms are represented by systems of nonlinear integral equations which implicitly define the observer output. Practical implementation methodologies are proposed through development and analysis of local iterations at each measurement event. These local iterations for nonredundant or redundant measurements reduce, respectively, to Newton and modified Gauss-Newton methods for which there is a broad literature base. The resulting iterated sliding mode observers can be applied to a broad class of discrete-time state estimation problems. Applications and performance of the various sliding mode observer forms are demonstrated through simulation results in the problem of estimating motion and structure of manoeuvring objects based on feature position measurements in long multiple-camera image sequences.

Acknowledgements

I would first like to thank my supervisor, Dr. H. M. Schwartz, who has been both encouraging and demanding, and whose high standards have served as an example of academic and moral excellence. I especially wish to thank Dr. Schwartz for his patience during three years of long-distance supervision and his diligence in reviewing hand-written, e-mail, and FAX notes. I would also like to thank Dr. A.U. Sheikh who provided academic counselling and guidance during my on-campus academic program. The students and faculty were a constant source of intellectual stimulation during my time at Carleton.

I gratefully acknowledge financial assistance provided by the Natural Sciences and Engineering Research Council of Canada through their graduate scholarship program, and by the Canadian Department of National Defence. Management and my colleagues at the Defence Research Establishment Suffield, Alberta, have been very supportive of my academic program. In particular, I want to thank Dr. J. Moldon, Dr. R. Stuart, and Mr. C. Laforce for supporting my original proposals for education leave, supporting additional proposals for the Doctoral program, providing ongoing support and advice during the academic program and thesis work, and contributing to a work environment that encourages and fosters academic achievement.

To my wife, Maria, and to our three children, Kathleen, Julie, and Stephanie, thank you for your continuous patience, understanding, support, and encouragement. I promise now to come out of my office.

Contents

1	Introduction	1
1.1	Feature-based Image Analysis	2
1.2	The Observer Problem	6
1.3	The Sliding Mode Concept	7
1.4	Thesis Overview	8
1.5	Thesis Scope and Contributions	11
1.6	Chapter Summary	15
2	Image Sequence Analysis	16
2.1	Introduction	16
2.2	Literature Review	18
2.3	Notation and Geometry	22
2.4	Approximate Models for General Motion	25
2.4.1	Structure Model	25
2.4.2	Measurement Model	28
2.4.3	Translational Motion Model	30
2.4.4	Rotational Motion Models	33
2.5	Obtaining Initial Estimates	46
2.6	EKF Performance With Simulated Imagery	50
2.6.1	Comparison of Cramer Rao Bounds	51

2.6.2	Simulation Results	61
2.7	Chapter Summary	76
3	State Estimation for Nonlinear Systems	78
3.1	Introduction	78
3.2	Definition of the Observer Problem	78
3.3	Nonlinear Observability	80
3.4	Current Methods of State Estimation	83
3.5	Fundamentals of Sliding Mode Theory	86
3.6	Chapter Summary	103
4	A Simple Sliding Mode Observer	104
4.1	Introduction	104
4.2	Form of the Observer	105
4.3	Motivation for Further Study	113
4.4	Chapter Summary	126
5	Exponential Stability of Discrete-time Systems	128
5.1	Introduction	128
5.2	Stability Notions and Definitions	129
5.3	Background and Preliminary Results	132
5.4	Linear Systems	140
5.5	Nonlinear Systems	147
5.6	Exponential Stability of Nonlinear Observers	155
5.7	A Reaching Law Method for Discrete-time Sliding Mode Systems	167
5.8	Chapter Summary	175
6	Sliding Mode Observers: Linear Measurements	177
6.1	Introduction	177
6.2	Notation and Preliminary Results	179

6.3	Linear Measurement Redundancy	182
6.4	Non-redundant Linear Measurement Models	185
6.5	Redundant Linear Measurement Models	193
6.6	Simulation Results for 3D Feature Point Tracking	198
6.7	Chapter Summary	211
7	Sliding Mode Observers: Nonlinear Measurements	212
7.1	Introduction	212
7.2	Nonlinear Measurement Redundancy	214
7.3	Non-redundant Nonlinear Measurement Models	222
7.3.1	Assumptions and Preliminary Results	223
7.3.2	Theoretical Form, Stability, and Properties	229
7.3.3	Implementation With Local Iterations	239
7.4	Redundant Nonlinear Measurement Models	268
7.4.1	Assumptions and Preliminary Results	269
7.4.2	Theoretical Form, Stability, and Properties	272
7.4.3	Implementation With Local Iterations	280
7.5	Simulation Results for Motion and Structure Estimation	286
7.6	Chapter Summary	300
8	Conclusions and Future Research	301
	References	313
A	Parameterizations of Relative Orientation	330
A.1	Fundamental Properties of $\mathbf{I}_E^O$	331
A.2	Euler Angle-Axis Parameterization	333
A.3	Quaternion Parameterization	337
A.4	Roll-Pitch-Yaw Parameterization	341

B	Kalman Filtering Review	345
B.1	Discrete-Time Kalman Filter	346
B.2	Discrete-Time Extended Kalman Filter	350
B.3	Continuous-Discrete Extended Kalman Filter	352
B.4	Local Iterations	354
C	Cramer-Rao Covariance Bounds	356
D	Jacobian of Measurement Models	363
E	Observer Implementations	367

List of Tables

2.1	Simulation parameters for Cases 1 and 2.	66
2.2	Statistical comparison of Angle-Axis and Quaternion algorithms.	73

List of Figures

1.1	Hierarchical structure for motion analysis.	3
2.1	Central projection image formation geometry.	23
2.2	System geometry for general motion analysis.	24
2.3	Fixing the object-centred frame with three feature points.	27
2.4	Object used in simulation studies.	51
2.5	CRLB's of orientation vector with and without resetting.	53
2.6	Cramer Rao bound comparison of position and velocity.	57
2.7	Cramer Rao bound comparison of acceleration and orientation.	58
2.8	Cramer Rao bound comparison of rotational motion.	59
2.9	Cramer Rao bound comparison of structure estimation.	60
2.10	Image subsequence for Case 1.	68
2.11	True and sample noisy image feature point trajectories for Case 1.	69
2.12	Monte Carlo simulation results for Case 1.	70
2.13	Final structure estimates in Monte Carlo simulations for Case 1.	71
2.14	Image subsequence for Case 2.	74
2.15	Monte Carlo simulation results for Case 2.	75
3.1	Conditions for existence of a sliding mode.	88
3.2	Robustness mechanism in continuous-time sliding mode control.	94
3.3	Phase plane diagrams for simple continuous-time sliding mode observer.	97

3.4	Robustness mechanism in continuous-time sliding mode observer.	99
4.1	Phase plane diagram for simple discrete-time sliding mode observer	107
4.2	Phase plane error trajectories for simple discrete-time sliding mode observer.	108
4.3	Example error trajectories of simple discrete-time sliding mode observer for 2D system.	109
4.4	Comparison of discrete-time Kalman filter and discrete-time sliding mode observer.	117
4.5	Tracking feature points in the image plane, Kalman filter vs sliding mode observer.	118
4.6	Tracking 3D feature points, Kalman filter vs sliding mode observer.	119
5.1	Illustration of sets and stability domains for Example 5.4.	154
5.2	Stability of an observer.	158
5.3	Application of the Fundamental Theorem of Integral Calculus in stability analysis of an observer.	160
5.4	Illustration of relationships between the sets X , X_b , X_s , X_k and Ω in sta- bility analysis of an observer.	162
5.5	Observer error bound dependence on matrix W	168
5.6	Possible approaches to the surface with the reaching law method.	174
6.1	Example results of tracking feature point in 3D: nonredundant measurements.	204
6.2	Monte Carlo simulation results for tracking feature point in 3D: nonredun- dant measurements.	205
6.3	Effect of third camera placement for estimation with redundant measure- ments.	206
6.4	Monte Carlo simulation results for tracking feature point in 3D: redundant measurements.	210

7.1	Example results of motion and structure estimation: nonredundant measurements.	292
7.2	Monte Carlo simulation results for translational motion estimation: nonredundant measurements.	293
7.3	Monte Carlo simulation results for rotational motion and structure estimation: nonredundant measurements.	294
7.4	Monte Carlo simulation results for translational motion estimation: redundant measurements.	298
7.5	Monte Carlo simulation results for rotational motion and structure estimation: redundant measurements.	299

Chapter 1

Introduction

In many engineering applications, sensors of various kinds are employed to measure physical properties with conversion into electronic form. These measurements, without further processing, typically yield an incomplete representation of the system under consideration. For example, a sensor which measures a position at each time instant does not, by itself, provide information concerning future positions. In order to predict the future evolution of a dynamic process a more complete representation of the state of the system is required. The state of a dynamical system is a set of physical quantities which, in the absence of external excitation or disturbances, completely determines the evolution of the system [1]. In order to control a dynamic system, many methodologies to generate the control signal require that the complete system state be available. Implementation of the control law, in this case, dictates that the full state of the system be estimated from available measurements.

Over the last two decades, advances in solid state components and computer systems have provided a means to implement control algorithms through digital computation rather than through analog components. In this case, finite clock frequencies of digital systems imply that the algorithm is in fact implemented in a discrete-time form. Available methods, originally developed with continuous-time analysis, are often implemented digitally with the assumption that update rates are sufficient to accurately approximate the continuous-time objectives with a numerical, discrete-time implementation. In many systems, how-

ever, limitations on the output rates of measurement sensors may impose severe restrictions on the accuracy of discrete-time implementations of continuous-time algorithms. For example, scanning sensors such as video, laser, or radar imaging systems may only provide frame rates from 0.2 to 60 Hz. and, in addition, will require processing time to extract required measurements from each frame. The Global Positioning System, which is now seeing broad applications in navigation, also imposes some restrictions with common update rates on the order of 1-2Hz. As a result, more and more emphasis over the last few years has shifted to control system design and analysis with explicit consideration of the discrete-time measurement process.

This thesis focusses on the discrete-time state estimation problem, proposes new approaches based on the sliding mode concept, and also proposes new methods for discrete-time system stability analysis. The theory and estimation algorithms developed in subsequent chapters can be applied to a broad class of systems. The problem of estimating motion and structure of objects from measurements of feature positions in long image sequences is employed in simulation studies of the proposed techniques. Section 1.1 discusses various components of feature-based motion analysis and proposes a possible hierarchical structure for an image analysis system. Applications of the theory developed in subsequent chapters are demonstrated on various components of this hierarchical system. Section 1.2 discusses, from a historical perspective, the state estimation problem, while basic concepts of sliding modes are presented in Section 1.3. These issues are presented in a more concise mathematical framework in Chapter 3. Following these brief introductions to image analysis, observers, and sliding modes, Section 1.4 presents an overview of the research leading to this thesis. Section 1.5 defines more precisely the scope, limitations, and contributions of the thesis work.

1.1 Feature-based Image Analysis

The feature-based image analysis problem is primarily concerned with obtaining estimates of motion and structure of objects within the environment of an imaging system based on measurements of feature positions in the image sequence. One possible hierarchical approach to feature-based motion analysis is shown in Figure 1.1. The system has been di-

Figure 1.1: Hierarchical structure for motion analysis.

vided into eight levels, as indicated on the right, in which higher levels represent increased sophistication in the representation of observed processes. In Level I (bottom row of the Figure), multiple cameras, which need not be stationary, observe a dynamic scene. An estimate of each camera's motion is computed from local sensors and perhaps feedback from higher levels of the system. In Level II, each camera produces an image sequence which must be captured and digitized, and may pass through a preprocessing stage. Feature detection algorithms are then applied in Level III to each frame in order to extract features of interest (points, lines, corners, etc.). Feedback from motion estimation algorithms in higher levels of the system may significantly reduce the computational effort required for feature detection. Output data from Level III includes sets of available features from each image sequence.

Level IV of the system attempts to establish correspondences between features extracted from the image sequences. In a temporal sequence of images generated by any single camera, the process of tracking features from one sample time to the next is referred to as *temporal correspondence*. In the case of feature points, this is essentially a two-dimensional multi-target tracking and data association problem. Again, feedback from motion and structure components of higher levels of the system may guide the two-dimensional tracking, or temporal correspondence problem.

Once feature positions and temporal correspondence have been established for image sequences of each camera, higher levels of the system attempt to fuse information from multiple cameras. Level V focusses on three-dimensional feature tracking and segmentation of feature sets into groups such that each group corresponds to a rigid object. The process of matching features seen in two or more images taken at the same time instant is referred to as *spatial correspondence*. Three dimensional information can be inferred from image plane feature position measurements and spatial correspondence in two or more cameras. As a result, this processing stage leads to a three-dimensional multi-target tracking and data association problem. Initial spatial correspondences must be established through epipolar geometry together with any information available from two-dimensional tracking problems of Level IV. Spatial correspondence is simplified and more robust if more than two cameras are employed.

Level VI comprises motion and structure estimation or three-dimensional object track-

ing. Once at least three noncollinear feature points, for example, are available for a single rigid object from Level V, three dimensional object state estimation may be initiated in order to further refine representations of the observed process. State estimation in Level VI focusses on recovery of translational motion (position, velocity, acceleration, etc.), rotational motion (orientation, angular velocity, etc.) and object structure in terms of the position of features with respect to an object centred frame of reference.

Two additional levels in Figure 1.1 include knowledge refinement in Level VII and environmental interaction in Level VIII. Knowledge refinement refers to the accumulation and refinement of information from all previous levels in order to construct a meaningful representation of the observed processes. This decision-making level might be responsible, for example, for filter decoupling once structure estimates of rigid objects reach a pre-defined confidence level, fusion of rigidly attached objects, maintaining an estimate of relative motion or interaction of multiple objects, or perhaps object recognition and/or classification. The primary objective of most vision systems ultimately is to interact in some way with the environment or at least to form intelligent decisions concerning all observed processes. This is the function assigned to Level VIII and may include, for example, direct control of dynamic systems, path planning and navigation, threat detection with assessment and evasive strategy planning (collision avoidance), or perhaps compilation of surveillance information in security systems. Knowledge based systems will no doubt play a major role in forming solutions to these higher-level problems.

In this system, tasks encountered in Levels IV, V, and VI can be formulated as discrete-time observer problems. A fundamental difficulty in each of these levels is that the true dynamics of the observed objects may be completely unknown in which case approximations must be employed which can lead to significant modelling errors. Many researchers have contributed towards the motion and structure estimation problem of Level VI when measurements are taken either from the image planes, as in Level III, or from backprojection to 3D measurements, as would be available in Level V. The extended Kalman filter has been proposed as a very useful tool for this purpose, and this approach as well as the modelling problem are examined in Chapter 2. The extended Kalman filter, although it appears to work well in some cases, can be highly sensitive to modelling errors which can lead to poor performance or even filter instability. These observations are the primary mo-

tivation for development of state estimation algorithms which are robust in the presence of unmodelled dynamics and parametric uncertainty—an inherent property of the sliding mode method.

1.2 The Observer Problem

A dynamical system, often implemented numerically, whose purpose is to estimate unmeasurable states based on available measurements of some dynamical process is called an *observer*. The observer problem, in very simple terms, can be stated as follows: given an initial estimate, say at time t_0 , of the state of a dynamical system, and given measurements which are related at any time instant to some, but perhaps not all, of the states during the time interval $[t_0, t]$, $t > t_0$, estimate the full state of the system at time t .

A critical property of an observer is that its output, that is the state estimates, converge to the true state with time, in which case the observer is said to be *asymptotic*. The true state of a system and the true dynamics which govern time propagation of the true state are ideal concepts which are seldom, if ever, known precisely in practice. Consequently, the construction of an observer is usually based on a subset of the complete set of true state variables and on an approximate dynamical model. These approximations give rise to modelling errors which influence the observer error. Robustness of the observer to modelling error, or simply robustness to unmodelled dynamics, is a very important issue in observer design. Another problem is that some errors are always present in sensor measurements. It is often the case that the two desirable properties, robustness to unmodelled dynamics and insensitivity to measurement errors, present conflicting requirements in observer design: very simply, observer gain levels must often be kept high to yield robustness to unmodelled dynamics, but low to reduce sensitivity to measurement noise. In this thesis, emphasis is placed primarily on the problem of robustness to unmodelled dynamics, although some effort, particularly through the use of redundant measurements, is devoted to the consideration of effects of measurement errors on the proposed observers.

The term *observer* was apparently [1] first introduced by Luenberger [2, 3, 4] who, following Wiener [5], Kalman [6], and Kalman and Bucy [7], devoted significant effort to state estimation for linear systems. Kalman's work with observers was devoted primarily

towards linear stochastic systems and yields an optimal estimate, in the sense of minimum variance, provided the linear system is known precisely and that disturbances are zero mean Gaussian processes with known covariances. Various extensions of Kalman's technique, most notably extensions to nonlinear systems or systems with time correlated noise processes, lead to alternate forms of the observer as documented in the collection of papers edited by Sorenson [8], and several texts [9]-[13].

Luenberger, [2], treated the continuous-time linear deterministic state estimation problem and is credited [14] for pioneering the notion of reduced-order observers. The number of states in a dynamical system is called the *order* of the system. For a system with n states and m measurements, $m < n$, Luenberger, in his treatment of the linear problem, demonstrated that a dynamical system of order $n - m$ could be used, together with the m measurements, to reconstruct the full state of the system. The reduced-order Luenberger observer for continuous-time linear systems is a sliding mode observer in the sense that the dynamics of the observer error is constrained to lie in an $n - m$ dimensional subspace of the original n dimensional state space [15]. This notion is made clearer in the following section, which introduces the concept of sliding modes in dynamical systems.

1.3 The Sliding Mode Concept

A dynamical system with a *sliding mode* is one in which the states of the system, after some time period, are restricted to lie in some manifold, usually a subspace or hypersurface, in the state space. A sliding mode can appear in systems which have a variable, or discontinuous structure. Prior to 1970, variable structure systems were studied primarily in the Soviet literature with a significant introduction to western literature being provided by Utkin [16]. Variable structure systems are based on differential equations in which the right-hand side, or forcing term, is discontinuous across a submanifold of the state space. A variable structure system is said to be in *sliding mode* when, during a finite time interval, the system state encounters the surface of discontinuity infinitely often. This situation can occur if the trajectories of the differential equations off of the discontinuity surface are directed towards the discontinuity surface. Starting from an initial state off of the surface, the state reaches the surface in finite time and thereafter is constrained to lie on, or slide along, the

discontinuity surface. This sliding action imposes certain constraints on the dynamics of the system and depends largely on the parameters which define the sliding surface.

Filippov [17] developed a solution concept for differential equations of variable structure and studied various properties of their solutions (existence, continuity, uniqueness, etc.). Filippov's solution concept, and in many cases Utkin's [16] method of equivalent control, provide a means to analyze the dynamics of variable structure systems in sliding mode. These sliding dynamics are obtained in the limit as the switching frequency about the discontinuity surface becomes infinite. In practice, therefore, ideal sliding cannot generally be achieved, particularly for the control problem, due to finite bandwidth and saturation limits of the switching mechanism. A common result is control chattering which is highly undesirable in practice since it implies extremely high control activity and may excite high frequency unmodelled dynamics. More recent developments using dynamical controllers or insertion of boundary layers focus on elimination of chattering and are discussed in a more detailed review in Chapter 3. In observer design and implementation, however, chattering issues are less severe since they are related only to numerical precision and update rates of the digital implementation; the observer itself is not responsible for delivering a control signal and hence is not restricted by the physical properties of actuators which influence the plant. The primary advantage of the sliding mode method is that it offers an inherent robustness to parametric uncertainty and unmodelled dynamics.

Extensions to the discrete-time case of the usual continuous-time sliding mode conditions do not necessarily guarantee chattering motions close to the sliding surface [18], and may lead to unstable chattering about the sliding surface [19]. Discrete-time systems, however, can exhibit sliding behavior with continuous systems of difference equations, which is in sharp contrast to the requirement of discontinuities in the continuous-time case. These issues are also considered in more detail in Chapter 3.

1.4 Thesis Overview

Previous research conducted by the candidate [20, 21] focussed on the motion and structure estimation problem from monocular image sequences when the object of interest was constrained to lie on a known planar surface. The iterated extended Kalman filter was em-

ployed for state estimation, and performance assessments resulted from simulation studies with computer-generated imagery and Cramer Rao bound analysis. More recent research [22, 23], which represents initial investigations leading to this thesis, considered extensions of these methods to the case of manoeuvring objects observed in multiple camera image sequences. The focus of this work was on a comparison of the performance of iterated extended Kalman filters based on alternate models of rotational motion including the Euler angle-axis, roll-pitch-yaw, and quaternion parameterizations. A primary observation of this application of the extended Kalman filter was that, although the filters appear to work well in some cases, sensitivity to dynamical modelling errors was a major problem which could lead to filter instability. As a consequence, results obtained from one trajectory say very little about results that could be expected from different trajectories, particularly with only scarce stability theory available in the literature for the extended Kalman filter in the presence of modelling errors.

A search was then initiated for alternate methods which might exhibit the desirable robustness properties. Adaptive extended Kalman filtering, such as adaptive Bayesian multiple model filtering, appeared at first to offer potential advantages in simple simulations. However, theoretical results for these methods are based on precise knowledge of admissible dynamical models and the statistics which govern transitions from one model to another. Further review of the literature on nonlinear state estimation techniques led to the sliding mode approach for continuous-time systems, but methods for the discrete-time nonlinear observer problem are not currently available. A simple approach, using a direct extension of the continuous-time sliding mode observer to the discrete-time problem, was implemented and tested in simulation with comparisons to the Kalman filter. Very encouraging results were obtained, with significant improvements over the Kalman filter in the presence of unmodelled dynamics.

Although this simple approach appears to work well for simple linear systems, robustness properties of the continuous-time counterpart do not carry over to this discrete-time approach and direct extensions to nonlinear systems are not apparent. Following numerous attempts to directly extend the continuous-time methods to the discrete-time problem, an investigation was conducted to determine the form of observer which could not only maintain sliding on the prescribed surface, but also retained the robustness mechanism of

continuous-time sliding mode methods [24]. The results of this investigation led to a new approach, based on a discrete-time reaching law method, which gave promising results in simulations with more general systems. Stability analysis, however, using standard Lyapunov methods for discrete-time systems did not yield sufficient information concerning restrictions on the accuracy of initial estimates nor ultimate bounds on the observer error. New stability results were then obtained [25, 26] using vector norm Lyapunov-like functions for discrete-time nonlinear systems which led to sufficient conditions for exponential stability and quantitative restrictions on initialization as well as explicit bounds on the observer error.

The proposed discrete-time sliding mode observers, even for uncertain nonlinear dynamics, can be realized directly for systems with linear measurement models. Initial results, however, included a type of relative degree constraint [122] which limits applications of the method to systems with non-redundant measurements. An alternate method was then developed for the case of linear redundant measurements which minimizes a quadratic cost function during sliding and yields, in a sense to be defined, an optimal fusion of the redundant measurements. In the case of nonlinear measurement models, however, the proposed *ideal* sliding mode observers for both non-redundant and redundant measurements are represented by an implicit system of nonlinear difference equations. A simple approximation, much like that which is used in the extended Kalman filter, appears to give sufficient accuracy in some cases, however, stability analysis with this simple approximation are non-trivial. Further analytical work suggested that Newton-type local iterations could be used to obtain improved accuracy for more general systems. With these local iterations and the proposed discrete-time stability theory, practical stability results were obtained for the sliding mode observer in the case of nonlinear non-redundant measurement models. For systems with nonlinear redundant measurement models, the proposed observer yields a least squared error estimate during sliding, and local iterations of a modified Gauss-Newton form are employed to approximate the ideal observer. The proposed local iterations provide a practical implementation methodology for sliding mode observers which can be applied to a broad range of state estimation problems.

1.5 Thesis Scope and Contributions

This thesis focusses on the development and stability analysis of sliding mode observers for state estimation of nonlinear discrete-time systems. The general class of systems that is considered are those for which the system dynamics can, in theory, be written as

$$\mathbf{x}(k+1) = \mathbf{f}(k, \mathbf{x}(k)), \quad k = 0, 1, 2, \dots \quad (1.1)$$

where $\mathbf{x} \in \mathbf{X} \subseteq \mathbb{R}^n$, and $\mathbf{f} : \mathbb{N}_0 \times \mathbf{X} \mapsto \mathbf{X}$ is C^p in $\mathbf{x}$, $p \geq 1$, where C^p denotes the space of p -times continuously differentiable functions, and $\mathbb{N}_0 := \{0, 1, 2, 3, \dots\}$.

Practical applications of methods proposed herein do not require specifically that the system dynamics be expressed in discrete-time closed form as shown in (1.1). Quite often the system dynamics may be expressed in continuous-time form, with $\mathbf{x}_c$ representing the continuous-time state vector,

$$\dot{\mathbf{x}}_c = \mathbf{f}_c(t, \mathbf{x}_c(t)), \quad (1.2)$$

in which $\mathbf{f}_c(t, \mathbf{x}_c(t))$ may include a known control signal. In theory, the continuous-time system (1.2) can be integrated over each sample period $[t_k, t_{k+1}]$ to give

$$\mathbf{x}_c(t_{k+1}) = \int_{t_k}^{t_{k+1}} \mathbf{f}_c(t, \mathbf{x}_c(t)) dt + \mathbf{x}_c(t_k), \quad (1.3)$$

which, upon setting $\mathbf{x}(k) := \mathbf{x}_c(t_k)$ and $\mathbf{f}(k, \mathbf{x}(k))$ equal to the right-hand side of (1.3), yields the discrete-time form of (1.1). In practice, numerical integration of (1.3) can be used to propagate the state from one sample time to the next.

A general assumption employed throughout the thesis is that the true system (1.1) is known on $\mathbf{X}$ up to some bounded error in the following sense. A model, $\hat{\mathbf{f}}(k, \mathbf{x}(k))$, of the true dynamics is available such that $\forall \mathbf{x} \in \mathbf{X}$,

$$\mathbf{f}(k, \mathbf{x}) - \hat{\mathbf{f}}(k, \mathbf{x}) = \mathbf{B}(k)\mathbf{d}(k, \mathbf{x}), \quad (1.4)$$

where $\mathbf{B}(k) \in \mathbb{R}^{n \times \ell}$ are known matrices with bounded norm uniformly in k , and $\mathbf{d} : \mathbb{N}_0 \times \mathbf{X} \mapsto \mathbb{R}^\ell$ is bounded over $\mathbf{X}$ uniformly in k . This means that the difference between the true propagation of the system over one sample period from any point in $\mathbf{X}$ and the approximate propagation from that same point is bounded in norm. This property will be referred to as, "bounded propagation error over any sample period." The matrices

$B(k)$ provide a possible mechanism to exploit available knowledge of the structure of modelling errors. It should be noted that bounded modelling error and disturbances in the underlying continuous-time system will result in bounded propagation error over any sample period for the discrete-time representation.

Both linear and nonlinear discrete-time measurement models are considered in subsequent chapters. Generally, for each k , measurements depend only on a subset of the full set of state variables and eventually notation is introduced to single out those states upon which measurements are dependent at each sample time. For brevity in this discussion, however, the measurement model can be written as

$$z(k) = h(x(k)). \quad (1.5)$$

where $z \in \mathbb{R}^m$ and $h : X \mapsto \mathbb{R}^m$. The overall system, with dynamics as in (1.1) and measurements as in (1.5), is assumed to be observable. Fundamental concepts of nonlinear observability are reviewed in Chapter 3. Various additional assumptions are required on the function $h(\cdot)$ in different sections of the thesis, such as continuity, local homeomorphic properties, and bounded norm of the Jacobian and its inverse. These assumptions are stated and discussed in appropriate sections of the thesis.

Essentially all of the analysis given in subsequent chapters is deterministic, with the exception of investigations of the effects of measurement errors on the proposed methods, particularly in the case of redundant measurements. Emphasis instead is placed on robustness to unmodelled dynamics which represents the key advantage of the sliding mode method. In the literature on sliding mode methods, as is discussed further in Chapter 3, it is often assumed that measurements have sufficient accuracy or that measurements have been filtered so that measurement errors can be neglected in the design process. These issues are discussed further as observer development proceeds.

Theoretical results presented in the following chapters are quite general in that they are not specific to any particular state estimation problem. Applications of the proposed methods in different levels of the feature-based image analysis problem provide examples of systems for which the methods are well-suited for state estimation. The primary reason for this is that the problem of tracking unknown objects suffers inherently from various degrees of modelling error. The modelling problem for feature-based image analysis is treated in

Chapter 2 together with a performance assessment of the extended Kalman filter in these applications. Many components of the overall feature-based image analysis problem, for example correspondence and occlusion, remain to be solved and are not treated in this thesis.

The following items represent primary contributions of the thesis work:

1. Comparison of rotational representations, angle-axis, roll-pitch-yaw, and quaternion, in the image analysis problem and assessment of extended Kalman filter performance for the three models;
2. Development of a simple discrete-time sliding mode observer for simple systems;
3. New theoretical results on exponential asymptotic stability of nonlinear discrete-time systems through the use of vector norm Lyapunov functions;
4. Sufficient conditions for exponential stability, and computation of domains of exponential stability, of nonlinear discrete-time observers for system dynamics with bounded propagation error;
5. Development of discrete-time sliding mode observers using a discrete-time reaching law method for nonlinear time-varying system dynamics with nonlinear non-redundant or redundant measurement models together with stability analysis;
6. Development of practical implementation methods of proposed discrete-time sliding mode observers for the case of nonlinear measurement models with stability and simulation analysis;
7. Applications of the proposed sliding mode observer method to image sequence analysis.

The remainder of the thesis is organized as follows:

Chapter 2 treats the modelling problem for six-degree-of-freedom object motion, unknown structure of the rigid object, and measurements of feature positions from multiple camera image sequences. Comparisons of performance of the extended Kalman filter for three representations of rotational motion are presented.

Chapter 3 presents a general overview of the literature on nonlinear observers. It precisely defines the observer problem, introduces the important concept of nonlinear observability, and presents fundamental theory of sliding mode systems. It also briefly outlines current approaches to nonlinear state estimation such as the Extended Kalman filter, linearization methods, and variable structure/sliding mode methods. This review is not restricted to discrete-time systems, but includes the continuous-time case.

Chapter 4 proposes a simple extension of continuous-time sliding mode observer methods to the discrete-time case and presents comparisons of simulation results of this method with the Kalman filtering approach. The difficulties associated with extensions of this method to more general systems, particularly with regards to the stability issue, are discussed.

Chapter 5 presents new exponential stability results for discrete-time systems through the use of vector norm Lyapunov-like functions, and the applications of this theory to observer design, and to the proposed reaching law method for discrete-time sliding mode systems.

Chapter 6 proposes methods for sliding mode observer design for systems with linear measurement models. Concepts of measurement redundancy are treated first, followed by developments of observers for the nonredundant and redundant measurement cases with simulation results.

Chapter 7 proposes methods for sliding mode observer design for systems with nonlinear measurement models. Concepts of nonlinear measurement redundancy are treated first, followed by developments of observers for the nonlinear, nonredundant and redundant measurement cases. In each case, the theoretical form, stability and properties are presented, followed by development of local iterations for observer implementation and simulation results.

Chapter 8 gives the conclusions of the thesis work and provides suggestions for future research.

1.6 Chapter Summary

Basic concepts of the discrete-time state estimation problem, systems with sliding modes, and the feature-based image analysis problem have been introduced in this chapter. An overview of the thesis work and presentation of that work in this document has been given, and primary contributions of the thesis work have been highlighted. The following chapter treats the modelling problem for six-degree-of-freedom object motion, unknown structure of the rigid object, and measurements of feature positions from multiple camera image sequences, together with performance comparisons of extended Kalman filters for three representations of rotational motion.

Chapter 2

Image Sequence Analysis

2.1 Introduction

Construction of an observer begins with the modelling problem which focuses on obtaining, in state space form, mathematical representations for the dynamics and measurements of the system under consideration. This Chapter focusses on the development of motion, structure, and measurement models which cast the feature-based image sequence analysis problem into one of state estimation for a nonlinear, discrete-time dynamical system. In this context, motion parameters include position, orientation, and time derivatives of position and orientation of an object with respect to a specified reference frame. Structure refers to the shape of the observed object in the sense that feature positions are estimated with respect to an object-centred reference frame. These notions are made more precise in Section 2.4. Many existing approaches to motion and structure recovery from long image sequences are somewhat restrictive in that solutions are proposed for specific cases of translational and rotational motion, for example, constant velocity, acceleration, precession, or angular momentum, etc. These techniques have been designed and evaluated on appropriate data generated under these assumptions. However, such precise *a priori* information is seldom available in general object tracking applications, particularly in an unstructured environment.

A fundamental difficulty associated with vision-based motion estimation, in comparison to the more traditional radar-based point tracking problem for example, lies in the recovery of object orientation and rotational motion. Theoretically precise models of rotational

motion may be formulated in terms of the object's inertial parameters, external torques acting on the object, and Euler's equations which represent a coupled system of three nonlinear differential equations that admit closed-form solutions only in very limited cases. In object tracking applications, inertial parameters and external torques are generally not available. Consequently, alternate and often simplified models for rotational motion are employed which naturally results in various degrees of modelling error. Rotational motion is often parameterized with unit quaternions, which leads to further difficulties when enforcing a nonlinear constraint on the quaternion norm. Roll-pitch-yaw parameterizations, on the other hand, have been reported to be neither simple nor well-behaved numerically and have led to overly complicated and computationally demanding implementations. The Euler angle-axis parameterization has appeared in small intersample-angle approximations, but published results in the context of recursive motion and structure estimation are not available. None of these methods have been developed for the case of manoeuvring objects with measurements from multiple-camera systems. These observations are discussed further in the literature review of Section 2.2. Appendix A presents a tutorial review of these parameterizations with the notation introduced in Section 2.3.

This chapter continues with a discussion of image sequence analysis techniques through a review of the literature in Section 2.2 which presents a summary of major contributions to motion analysis that are based on extended Kalman filtering. Section 2.3 establishes notation and geometry which leads to the development of motion, structure, and measurement models in Section 2.4. Six degree-of-freedom motion models for three parameterizations of rotational motion, using roll-pitch-yaw, Euler angle-axis, and quaternion parameters, are developed. Both the roll-pitch-yaw and the Euler angle-axis parameterizations lead to approximate linear dynamic models while the quaternion parameterization results in a nonlinear dynamic model. Strict assumptions are not imposed on the mode of translational or rotational motion except "smoothness" conditions in the sense that parameters may be differentiated with respect to time. Section 2.5 proposes a simple method to obtain initial estimates of position, orientation, and structure. Application and performance assessments of Extended Kalman Filters for each of the three parameterizations of rotational motion is presented in Section 2.6 through Cramer Rao bound analysis and comparisons of simulation results from synthetic imagery.

2.2 Literature Review

Over the last two decades, research in motion detection, analysis, and understanding techniques with applications in computer vision systems for robotics and autonomous vehicles represents an area of vigorous growth. Rosenfeld's annual bibliographical surveys¹ demonstrate the intense interest in this field of research. Motion analysis techniques are often classified according to whether they are formulated through an image flow-based, or feature-based approach. The image flow approach first estimates apparent velocities of intensity patterns in an image sequence before motion analysis begins. In feature-based algorithms, features such as points, lines, contours, corners, etc., are extracted from each image and matched between images in order to estimate relative motion of features in scenes with respect to an observer. This approach implicitly assumes that the object of interest has a set of discernible features which can be detected in the image sequences.

Applications of Kalman filtering to image analysis have been reported by many authors, for example Futrelle and Speckert [27], Gennery [28], Hallman [29], Legters and Young [30], Stuller and Krishnamurthy [31], Wu *et al.* [32], Matthies, Kanade, and Szeliski [33], Chang *et al.* [34], and Iu and Wohn [35, 36] consider diverse vision-based problems to which sequential state estimation has been applied. The work of three other groups, however, provides the primary motivation for investigations of EKF performance in this chapter.

Ayache and Faugeras [37], Ayache and Lustman [38], Randall, Foret, and Ayache [39] (see further references cited therein) recursively build and update three dimensional representations of the environment of a mobile robot. These methods are developed primarily for trinocular (three cameras) image sequences. Positions of world features (points, lines, planes) are recursively estimated from measurements in the image plane through extended Kalman filtering. Geometric relations are inferred between features by computing a generalized Mahalanobis distance and using a χ^2 acceptance test. An estimate of robot motion is also maintained with the Kalman filter. They have applied their technique to real scenes

¹Azriel Rosenfeld's Computer Vision bibliographies from 1984 through 1993 are published in *Computer Vision, Graphics, and Image Processing* and may be obtained through anonymous FTP to [FTP:TELEOS.COM](ftp://TELEOS.COM) in directory [VISION-LIST-ARCHIVE/ROSENFELD-BIBLIOGRAPHIES](ftp://VISION-LIST-ARCHIVE/ROSENFELD-BIBLIOGRAPHIES)

taken from an indoor mobile platform and demonstrate some of the major properties of their algorithm.

Dickmanns and Graefe [40] present a thorough discussion of their approach to dynamic machine vision which employs special hardware [41, 42] and methods for feature extraction and information processing. They have successfully applied their technique to control of an inverted pendulum on a moving cart [43], autonomous vehicle docking problems [41], autonomous road vehicle guidance [41], and vision-based control of an aircraft in a simulated landing approach [44]. More recently, enhancement of road curvature models and extensions of algorithms and hardware to obstacle avoidance in autonomous vehicle guidance applications have been proposed [45, 46, 47].

Broida [48], Broida and Chellappa [49, 50, 51, 53, 52], Broida, Chandrashekhar, and Chellappa [54], and Young and Chellappa [55, 56, 57] have applied extended Kalman filtering techniques to estimation of motion and structural parameters of a rigid object through measurements taken from noisy monocular and stereo image sequences. Early work, [49], considered a special case in which object translational motion is constrained to lie on a plane defined by the optical axis and central scan line of the imaging system. In three dimensional tracking, [50]-[54], rotational motion is modelled with unit quaternions. Several cases have been treated based on *a priori* information concerning the nature of object motion: pure translational motion, [50]; constant translational and rotational velocities, [54, 52]; and constant translational acceleration and rotational motion with constant precession, [55, 56]. Both recursive algorithms, [49, 50, 54, 55, 56], and a batch approach based on a fixed number of frames, [51, 52], have been investigated. Approximate Cramer-Rao covariance bounds in the absence of *a priori* information are used extensively for performance analysis, [53]. Considerable effort has also been devoted to uniqueness issues for the constant acceleration and precession case, [56]. A more detailed review of their recursive methods for monocular systems, [49, 54], is given in previous work by the candidate [20].

Evaluation of proposed estimation schemes for long image sequences using real imagery has been conducted extensively only in the work of Ayache and Faugeras [37], Matthies *et al.* [33], and Dickmans *et al.* [41]. In a more limited sense, Wu *et al.* [32] show results for imagery obtained in a controlled laboratory setting, while Broida [48] and Broida

et al. [54, 52] show qualitative results based on real imagery for which ground truth data is not available and feature points were extracted manually. In most of the above investigations, simulations are performed in which feature detection and correspondence problems are assumed to be solved. In this case, feature positions in each image are obtained from true analytical motion and structural models and an assumed image formation model, and additive noise processes are included to simulate measurement noise.

Application of sequential state estimation techniques to machine vision problems emphasizes temporal continuity conditions for image sequence interpretation. This allows proper definitions of state variables and formulation of systems of differential or difference equations to model motion processes of objects in space and time. In these applications, extended Kalman filters should be considered as nonlinear observers with a particular structure in which the plant, and in some cases the measurement, *covariance* matrices are treated as tuning parameters which are specified by the designer in order to achieve a desired performance over a sufficient range of trajectories.

In planar motion applications, the axis of rotation is fixed and is often assumed to be known [20, 21, 29, 30, 49]. Parameterization in terms of angular velocity then results in linear dynamical models. More general rotational motion is often modelled with unit quaternions [28, 35, 58], [48]-[56]. Dynamical models which govern time propagation of the unit quaternion are parameterized in terms of angular velocity and are nonlinear. State estimation is then subject to a nonlinear constraint which imposes unit length on the four-element quaternion. This constraint is not easily incorporated into the linear structure of the Kalman filter. A common approach, [35], [48]-[56], is to normalize the estimated quaternion following each measurement update of the filter. This outside intervention in the filtering process which, if applied without additional corrective action, has not been analyzed. Impulsive normalization of the estimated quaternion has apparently been attributed (see for example Broida *et al.* [54, p. 646]) to Bar-Itzhack and Oshman [59], but this is **not** what Bar-Itzhack and Oshman propose. In fact, Bar-Itzhack and Oshman [59] observed poor performance with impulsive normalization even when angular velocity measurements are available. They conclude, from a linear analysis, that a normalized quaternion can be extracted for output provided that that portion of the estimated quaternion which is lost due to normalization is propagated in the filter. This latter requirement appears to have been overlooked

in some previous work, for example [54], in which acceptable results may have been due in part to strict dependence of the structural model on assumptions concerning translational and rotational motion.

In this chapter, an approach similar to that employed by Bar-Itzhack and Oshman, [59], is used in that the quaternion estimate is propagated in the filter equations without impulsive normalization. For output purposes only, a unit quaternion as well as object structure vectors are extracted from the state estimate. This problem is discussed further in Section 2.4.4. The use of higher-order time derivatives of angular velocity in a quaternion-based filter has been examined by Iu and Wohn [35] who employ a dynamical model for monocular image sequences similar to Broida *et al.* [54, 52]. However, time propagation in their approach assumes constant angular velocity over the sample period. Methods proposed herein assume instead that the highest time derivative of angular velocity included in the state vector is approximately constant over any sample period. A numerical integration scheme which maintains constant quaternion norm is derived for time propagation of the quaternion estimate.

Quaternions are often selected over roll-pitch-yaw angles due to mathematical singularities of the roll-pitch-yaw parameterization. Wu *et al.*, [32] use roll-pitch-yaw angles and parameterize rotational motion in terms of a constant inter-frame rotation matrix. This approach leads to an extremely complicated and computationally demanding implementation. Their technique has been applied to real scenes generated in a laboratory environment using 14 feature points on an object of known structure which is constrained to lie on a planar surface. Previous research, [20], has demonstrated that very little additional information is provided to the planar motion problem as the number of feature points is increased beyond 4. In the present investigation, both the roll-pitch-yaw and Euler angle-axis filters are developed from approximate linear dynamical models. In this case rotational motion is not parameterized directly in terms of angular velocity and its time derivatives. Instead, angular velocity is computed through a nonlinear function of the state estimates and compared to the true value and that generated by the quaternion filter.

This Section has provided a review of only a small subset of literature available on motion and structure estimation and has focussed on those methods which employ extended Kalman filtering. A more detailed review has been compiled by the candidate in [22]. The

following sections introduce notation and geometry, and develop dynamical and measurement models for motion and structure estimation from multiple camera image sequences.

2.3 Notation and Geometry

Under a rigid object assumption, it is natural to represent feature points on the object as constant vectors in an object-centred reference frame. The perspective projection image formation model requires that feature points be expressed in terms of each camera-centred reference frame, while in many applications it is often preferred that object position, orientation, and motion be described with respect to a stationary earth-fixed reference frame.

Reference frames denoted as

$$\begin{aligned} F_O &\sim \text{Object-fixed,} \\ F_E &\sim \text{Earth-fixed, and} \\ F_{C_j} &\sim j\text{th Camera-fixed, } j = 1, 2, \dots, N_C \\ &\text{for } N_C\text{-ocular imaging systems,} \end{aligned} \quad (2.1)$$

are used in the following development.

A reference frame, F_α , where $\alpha \in \{O, E, C\}$, is defined by a point O_α , the origin of F_α , and three orthonormal vectors $\mathbf{i}_\alpha$, $\mathbf{j}_\alpha$, and $\mathbf{k}_\alpha$ which lie along the x_α , y_α , and z_α axis of F_α , respectively. A vector $\mathbf{r}$ which describes the position of a feature point on the object with respect to O_α and is expressed in F_α is denoted by $\mathbf{r}_\alpha$. The components of $\mathbf{r}_\alpha$ are written as $\mathbf{r}_\alpha = [r_{\alpha,1}, r_{\alpha,2}, r_{\alpha,3}]_\alpha^T$.

A second reference frame, F_β , is related to F_α by a translation $\mathbf{T}_\beta^\alpha$, which represents a vector from O_α to O_β and, when expressed with respect to F_α , is written as $[\mathbf{T}_\beta^\alpha]_\alpha$, and an orthogonal 3×3 matrix, $\mathbf{I}_\beta^\alpha$, which denotes a transformation from the α -basis to the β -basis and represents a rigid rotation of $\mathbb{R}^3$. Appendix A provides a tutorial review of the Euler angle-axis, quaternion, and roll-pitch-yaw parameterizations for basis transformations represented by $\mathbf{I}_\beta^\alpha$. With this notation,

$$[\mathbf{r}_\alpha]_\beta = \mathbf{I}_\beta^\alpha \mathbf{r}_\alpha, \text{ and} \quad (2.2)$$

$$\mathbf{r}_\beta = [\mathbf{r}_\alpha]_\beta + [\mathbf{T}_\beta^\alpha]_\beta. \quad (2.3)$$

Figure 2.1: Central projection image formation geometry. A feature at point p^i projects onto the point (L, u^i, v^i) in the camera-centred frame, F_C , where L is the effective focal length.

If optical distortions of an imaging device are not severe, a perspective projection image formation model can be used to relate the observed location of a feature in a two-dimensional image to the three-dimensional coordinates of that feature in a camera-centred reference frame. Figure 2.1 shows the assumed geometry for the image formation model. The position of the i th feature point, p^i , with respect to a camera-centred reference frame, F_C , is described by the vector $\mathbf{r}_C^i$. A feature at point p^i is projected onto the image plane and appears at point (L, u^i, v^i) in F_C , where L is the effective focal length. The origin of F_C is taken at the centre of projection of the imaging lens. The x_C axis of F_C is aligned with the optical axis, while y_C is parallel to a scan line and positive to the right in the image. The z_C axis is positive down in the image to form a right-handed coordinate system. Neglecting separate horizontal and vertical scale factors, and assuming any distortion effects are negligible, the projected image plane position of feature point p^i is given by

$$u^i = L \left(\frac{r_{C,2}^i}{r_{C,1}^i} \right), \text{ and } v^i = L \left(\frac{r_{C,3}^i}{r_{C,1}^i} \right). \quad (2.4)$$

The central projection transformation defined in (2.4) demonstrates the loss of depth information in monocular imaging systems since the image position (u^i, v^i) is invariant under

Figure 2.2: System geometry for general motion analysis.

any scalar multiple of $\mathbf{r}_i^O$.

Figure 2.2 illustrates a typical physical arrangement for the motion and structure estimation problem considered here. Two or more cameras observe a single rigid object whose motion is both translational and rotational. Image formation and axis orientation for each camera frame are as shown in Figure 2.1. The position, $[\mathbf{T}_{C_j}^E]_E$, and orientation, $\mathbf{I}_{C_j}^E$, of the j th camera frame, F_{C_j} , $j = 1, 2, \dots, N_c$, are known or measured with respect to the stationary earth fixed frame, F_E . Object structure is formulated in terms of positions, $\mathbf{r}_i^O$, of feature points expressed with respect to the object-centred frame, F_O , while object motion, which consists of translational motion of the origin of F_O together with rotational motion of F_O , is described with respect to the stationary earth-fixed frame F_E . As a result, parameterization of object motion and structure is independent of position, orientation, and motion of the imaging systems.

The overall state vector $\mathbf{x} \in \mathbb{R}^N$ includes states corresponding to translational motion,

rotational motion and object structure:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_r \\ \mathbf{x}_s \end{bmatrix}, \quad (2.5)$$

where

$$\begin{aligned} \mathbf{x}_t \in \mathbb{R}^{N_t} &\sim \text{Translational motion state vector,} \\ \mathbf{x}_r \in \mathbb{R}^{N_r} &\sim \text{Rotational motion state vector,} \\ \mathbf{x}_s \in \mathbb{R}^{N_s} &\sim \text{Structure state vector.} \end{aligned} \quad (2.6)$$

The dimensions N_t , N_r , and N_s of these vectors are discussed in the following sections.

2.4 Approximate Models for General Motion

This section develops approximate dynamical models for six-degree-of-freedom manoeuvring object motion, and a perspective projection measurement model for multiple-camera imaging systems. The extended Kalman filtering approach, which is reviewed in Appendix B, requires that dynamical and measurement models be obtained in state space form as

$$\dot{\mathbf{x}}(t) = \mathbf{f}[\mathbf{x}(t)] + \mathbf{G}(t)\mathbf{w}(t). \quad (2.7)$$

$$\mathbf{z}(k) = \mathbf{h}[\mathbf{x}(t_k)] + \mathbf{v}(k). \quad (2.8)$$

where the process noise $\mathbf{w}(t) \sim N(\mathbf{0}, \mathbf{Q}(t))$, which is mapped to the state space through the matrix $\mathbf{G}$, and the measurement noise, $\mathbf{v}(k) \sim N(\mathbf{0}, \mathbf{R}(k))$, with $N(\mathbf{m}, \mathbf{C})$ representing Gaussian processes with mean $\mathbf{m}$ and covariance $\mathbf{C}$, are assumed to be temporally white.

2.4.1 Structure Model

The rigid object is defined by N_f feature points, p^i , $i = 1, 2, \dots, N_f$, whose position vectors, $\mathbf{r}_O^i$, with respect to F_O collectively define the structure set $\mathcal{S} = \{\mathbf{r}_O^1, \mathbf{r}_O^2, \dots, \mathbf{r}_O^{N_f}\}$. Observation of at least three noncollinear feature points by at least two cameras at each

observation event provides sufficient information to fix an object-centred frame on the object. A number of techniques, including [32]-[49], [20, 21], define an object-centred frame whose orientation and position with respect to the object itself is based in part on the assumed motion of the object. These assumptions may lead to unobservability when motion assumptions are violated. This is a property of methods proposed in [32]-[49] which cannot be directly employed in the case of manoeuvring objects for which the true nature of translational and rotational motion is unknown. In this case, the object-centred frame must be independent of object motion and defined on the basis of observed feature points only.

The object centred frame, F_O is fixed on the object by selecting, as will be discussed further in Section 2.5, three noncollinear feature points $\{p^1, p^2, p^3\}$, (three noncollinear image points are necessarily projections of three noncollinear 3D points) and imposing the following constraints as demonstrated in Figure 2.3:

1. In F_O , feature point p^1 lies at a known position $\mathbf{d}_O = [d_{O,1}, d_{O,2}, d_{O,3}]^T$ which is fixed at the time of filter initialization and thereafter remains constant and hence its structure vector is known;
2. One of the coordinate axis of F_O , say the x_O axis, is parallel to the line passing through p^1 and p^2 , in which case the y_O - and z_O -coordinates of the structure vector of p^2 are known to be $d_{O,2}$ and $d_{O,3}$, respectively; and
3. One of the coordinate planes of F_O , say the x_O - y_O plane, is parallel to the plane containing p^1, p^2 , and p^3 , in which case the z_O -coordinate of the structure vector of the third point is known to be $d_{O,3}$.

Selection of the vector $\mathbf{d}_O$ is discussed further in Section 2.5. This method results in one of four possible orientations of F_O with respect to the three points: feature point p^2 in Item 2 could have a positive or negative x_O coordinate, and/or feature point p^3 in Item 3 could have a positive or negative y_O coordinate. These four discrete solutions are related by diagonal basis transformations (rotations of the basis set by 0 or π about a coordinate axis). The three special feature points would most likely be selected, as outlined in Section 2.5, from a larger set of all available feature points, which leads to many possible positions and orientations of F_O with respect to the object. Once the initial ordering of three special points has been

Figure 2.3: Fixing the object-centred frame with three feature points.

specified, depending on the initialization error and observer transients, any one of the four possible orientations of F_O is possible. Obviously, the estimation process is likely to be poorly behaved if images of these three points are very close together relative to measurement noise levels. The important advantage, however, is that F_O is defined independent of object motion.

The structural constraints defined above are enforced by fixing the six known structural coordinates indicated in Items 1, 2, and 3, removing these elements from the structure state vector, and treating them as known parameters. Explicit removal of the six structural coordinates from the structure state vector in the development of dynamic and measurement models below would lead to unnecessarily complicated notation. For brevity, the full structure state vector is indicated by the notation, but it is understood that these six coordinates

are removed from the estimation process. The structure state vector, $\mathbf{x}_s \in \mathbb{R}^{N_s}$, where $N_s = 3N_f$ is written as

$$\mathbf{x}_s = \begin{bmatrix} \mathbf{r}_O^1 \\ \mathbf{r}_O^2 \\ \vdots \\ \mathbf{r}_O^{N_f} \end{bmatrix}. \quad (2.9)$$

For a rigid object, the elements of $\mathbf{x}_s$ are constant in time.

2.4.2 Measurement Model

The measurement model is formed from a set of N_C individual measurement models which have identical form and each corresponds to a single camera. The instantaneous position and orientation of F_O relative to F_E are given by the translation

$$[\mathbf{T}_O^E]_E(t) \equiv \mathbf{p}(t) = [x(t), y(t), z(t)]_E^T. \quad (2.10)$$

and basis transformation $\mathbf{I}_E^O(\zeta)$, respectively, in which elements of the vector ζ depend on the parameterization of rotational motion. The known instantaneous position and orientation of the j th camera frame, F_{C_j} , with respect to F_E is specified by a translation vector $\mathbf{T}_{C_j}^E$ and basis transformation $\mathbf{I}_{C_j}^E$, respectively. The position of the i th feature point with respect to F_{C_j} can be written as

$$\mathbf{r}_{C_j}^i = \mathbf{I}_{C_j}^E \{ \mathbf{I}_E^O \mathbf{r}_O^i + [\mathbf{T}_O^E]_E - [\mathbf{T}_{C_j}^E]_E \}. \quad (2.11)$$

and writing

$$\mathbf{I}_{C_j}^E = [\mathbf{i}_{C_j}]_E, [\mathbf{j}_{C_j}]_E, [\mathbf{k}_{C_j}]_E]^T, \quad (2.12)$$

gives $\mathbf{r}_{C_j}^i$ in component form as

$$\begin{aligned} r_{C_j,1}^i &= [\mathbf{i}_{C_j}]_E^T \{ \mathbf{I}_E^O \mathbf{r}_O^i + [\mathbf{T}_O^E]_E - [\mathbf{T}_{C_j}^E]_E \}, \\ r_{C_j,2}^i &= [\mathbf{j}_{C_j}]_E^T \{ \mathbf{I}_E^O \mathbf{r}_O^i + [\mathbf{T}_O^E]_E - [\mathbf{T}_{C_j}^E]_E \}, \\ r_{C_j,3}^i &= [\mathbf{k}_{C_j}]_E^T \{ \mathbf{I}_E^O \mathbf{r}_O^i + [\mathbf{T}_O^E]_E - [\mathbf{T}_{C_j}^E]_E \}. \end{aligned} \quad (2.13)$$

With N_f observed feature points, the measurement model for the j th camera is given by

$$\mathbf{z}_j(k) = \mathbf{h}_j[\mathbf{x}(k)] + \mathbf{v}_j(k). \quad (2.14)$$

where

$$\mathbf{h}_j[\mathbf{x}(k)] = \begin{bmatrix} \mathbf{h}_j^1[\mathbf{x}(k)] \\ \mathbf{h}_j^2[\mathbf{x}(k)] \\ \vdots \\ \mathbf{h}_j^{N_j}[\mathbf{x}(k)] \end{bmatrix}, \quad (2.15)$$

in which subscripts refer to the camera and superscripts refer to the feature point. The image plane position, $\mathbf{h}_j^i[\mathbf{x}(k)]$, of the i th feature point in the j th camera is computed through perspective projection as

$$\mathbf{h}_j^i[\mathbf{x}(k)] = \begin{bmatrix} h_{j,1}^i[\mathbf{x}(k)] \\ h_{j,2}^i[\mathbf{x}(k)] \end{bmatrix} = L_j \begin{bmatrix} \frac{r_{C_j,2}^i}{r_{C_j,1}^i} \\ \frac{r_{C_j,3}^i}{r_{C_j,1}^i} \end{bmatrix}. \quad (2.16)$$

The extended Kalman filtering approach assumes that measurement noise, $\mathbf{v}_j(k)$, in (2.14) is a discrete-time, zero-mean, white Gaussian process with known covariance, $\mathbf{R}_j(k)$. The proportionality factor L_j in (2.16) represents the effective focal length of the j th camera.

The overall measurement model is written as

$$\mathbf{z}(k) = \mathbf{h}[\mathbf{x}(k)] + \mathbf{v}(k), \quad (2.17)$$

where

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_{N_C} \end{bmatrix}, \quad \mathbf{h} = \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \vdots \\ \mathbf{h}_{N_C} \end{bmatrix}, \quad \text{and } \mathbf{v} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_{N_C} \end{bmatrix}. \quad (2.18)$$

Implementation of the extended Kalman filter, which is reviewed in Appendix B, and estimation of Cramer Rao bounds, which are reviewed in Appendix C, requires the computation of $\mathbf{r}_{C_j}^i$, $\mathbf{I}_E^O(\zeta)$ and $\mathbf{h}[\mathbf{x}]$ as well as the Jacobian of $\mathbf{h}[\mathbf{x}]$,

$$\mathbf{H}[\mathbf{x}] := \frac{\partial \mathbf{h}[\mathbf{x}]}{\partial \mathbf{x}}, \quad (2.19)$$

from the most recent state estimate. Computation of $\mathbf{H}[\mathbf{x}]$ is a straight-forward but tedious exercise in differentiation, however, with algebraic manipulation and simplification, $\mathbf{H}$ can be written in simple form which admits computationally efficient implementation. Equations of the measurement model Jacobian for the three parameterizations of rotational motion are given in Appendix D.

2.4.3 Translational Motion Model

Translational motion in object tracking problems is commonly modelled with a truncated Taylor series expansion in time. In this case the position, $\mathbf{p}(t) := [\mathbf{T}_O^E]_E(t)$, of the object-centred frame over any sample period, $\{t \mid t_k \leq t \leq t_{k+1}\}$, is modelled with an $(N_p - 1)$ th-order polynomial in time,

$$\mathbf{p}(t) = \sum_{j=0}^{N_p-1} \frac{1}{j!} \frac{d^j \mathbf{p}}{dt^j}(t_k) [t - t_k]^j. \quad (2.20)$$

A state vector corresponding to translational motion, $\mathbf{x}_t \in \mathbb{R}^{N_t}$, where $N_t = 3N_p$, is defined as

$$\mathbf{x}_t = \left[\mathbf{p}^T, \dot{\mathbf{p}}^T, \ddot{\mathbf{p}}^T, \dots, \left(\frac{d^{N_p-1} \mathbf{p}}{dt^{N_p-1}} \right)^T \right]^T. \quad (2.21)$$

which propagates in time according to

$$\dot{\mathbf{x}}_t(t) = \mathbf{A}_t \mathbf{x}_t(t) + \mathbf{G}_t \mathbf{w}_t(t). \quad (2.22)$$

where

$$\mathbf{A}_t = \begin{bmatrix} \mathbf{0}_{3(N_p-1) \times 3} & \mathbf{I}_{3(N_p-1)} \\ \mathbf{0}_{3 \times 3(N_p-1)} & \mathbf{0}_{3 \times 3} \end{bmatrix}, \text{ and } \mathbf{G}_t = \begin{bmatrix} \mathbf{0}_{3(N_p-1) \times 3} \\ \mathbf{I}_3 \end{bmatrix}. \quad (2.23)$$

and $\mathbf{w}_t \in \mathbb{R}^3$ is a zero mean, temporally white Gaussian process with covariance $\mathbf{Q}_t$.

In this case, the *manoeuvre variable*, $\mathbf{m}(t)$, defined as

$$\mathbf{m}(t) := \frac{d^{N_p} \mathbf{p}}{dt^{N_p}}(t) = \mathbf{w}_t(t) \quad (2.24)$$

is modelled as being uncorrelated in time. However, if an object is executing a manoeuvre at time t , it is likely to be executing a similar or perhaps even the same manoeuvre at time $t + \tau$ for sufficiently small τ which implies that $\mathbf{m}(t)$ should be modelled as a time-correlated process. One approach, [10, 31, 60] is to model each component of $\mathbf{m}(t) = [m_1(t), m_2(t), m_3(t)]^T$ as a first-order Gauss-Markov process with autocorrelation function $r_i(\tau)$ given by

$$r_i(\tau) := \mathcal{E}\{m_i(t)m_i(t + \tau)\} = \sigma_{t_i}^2 e^{-\alpha_{t_i}|\tau|}, \quad (2.25)$$

where $\mathcal{E}$ denotes expectation, $\sigma_{t_i}^2$ is the variance of the i th component of the manoeuvre variable, and $\alpha_{t_i} \geq 0$ is the reciprocal of the manoeuvre correlation time constant.

The subscripts i on α_i and σ_i will be omitted since, in the absence of alternative information, each component of $\mathbf{m}(t)$ can be treated identically. The Wiener-Kalmogorov whitening procedure can then be used to express the manoeuvre variables $m_i(t)$ in terms of white noise by taking the Laplace transform of $r_i(\tau)$ and writing

$$\begin{aligned} R_i(s) &:= \mathcal{L}\{r_i(\tau)\} \\ &= \frac{-2\alpha_i\sigma_i^2}{(s - \alpha_i)(s + \alpha_i)} \\ &:= H(s)H(-s)W_i(s). \end{aligned} \quad (2.26)$$

The function

$$H(s) := \frac{1}{s + \alpha_i} \quad (2.27)$$

is identified as the Laplace transform of whitening filters for each $m_i(t)$, while

$$W_i(s) := 2\alpha_i\sigma_i^2 \quad (2.28)$$

represents the power spectral density of independent white noise processes $w_i(t)$ that drive identical filters $H(s)$ to produce $m_i(t)$, according to first-order differential equations

$$\dot{m}_i(t) = -\alpha_i m_i(t) + w_i(t). \quad (2.29)$$

With this model for manoeuvring motion, the state vector $\mathbf{x}_t$ in (2.21) is augmented to include $\mathbf{m}(t)$ of (2.24),

$$\mathbf{x}_t = \left[\mathbf{p}^T, \dot{\mathbf{p}}^T, \ddot{\mathbf{p}}^T, \dots, \left(\frac{d^{N_p} \mathbf{p}}{dt^{N_p}} \right)^T \right]^T, \quad (2.30)$$

and propagates in time according to (2.22), but now

$$\mathbf{A}_t = \begin{bmatrix} \mathbf{0}_{3N_p \times 3} & \mathbf{I}_{3N_p} \\ \mathbf{0}_{3 \times 3N_p} & -\alpha_t \mathbf{I}_3 \end{bmatrix}, \text{ and } \mathbf{G}_t = \begin{bmatrix} \mathbf{0}_{3N_p \times 3} \\ \mathbf{I}_3 \end{bmatrix}. \quad (2.31)$$

and $\mathbf{w}_t(t)$ is zero-mean, temporally white Gaussian noise with covariance $\mathbf{Q}_t = 2\alpha_t\sigma_t^2\mathbf{I}_3$.

The linear differential equation in (2.22), with the definitions in (2.31), can be written as a discrete-time difference equation using standard techniques to give

$$\mathbf{x}_t(k+1) = \Phi_t(k)\mathbf{x}_t(k) + \mathbf{w}_t^d(k). \quad (2.32)$$

where

$$\Phi_t(k) = e^{\mathbf{A}_t(t_{k+1}-t_k)}. \quad (2.33)$$

The state transition matrix, Φ_t , can be written in closed form due to the simple form of $\mathbf{A}_t$, [60]. The system is driven by a zero-mean, discrete-time, white noise sequence, $\mathbf{w}_t^d(k)$, with covariance $\mathbf{Q}_t^d$ which may be computed (see [60]) directly from Equation (B.35) of Appendix B, or approximated, assuming the sample period is much smaller than system time constants ($2\alpha T \ll 1$, [60]) as

$$\mathbf{Q}_t^d \approx 2\alpha_t \sigma_t^2 T \mathbf{G}_t \mathbf{G}_t^T. \quad (2.34)$$

where T is the sample period (superscript ' T ' still denotes transposition).

Three parameters, N_p , α_t , and σ_t , must be specified to complete the translational motion model. Extensive simulations with the extended Kalman filter have demonstrated that $N_p = 2$ provides a reasonable balance between accuracy and computational requirements. In fact, because measurements are related directly only to object position, estimation of higher time derivatives ($N_p > 2$) often demonstrated serious lag and overshoot which can lead to filter instability. The parameter α_t is the reciprocal of the manoeuvre correlation time constant. If manoeuvres are expected to occur only over short time periods, α_t should be chosen fairly large (≥ 1), whereas slower expected manoeuvres are modelled with a much smaller α_t ($\approx 1/60$). In the selection of σ_t^2 , Singer [60] proposed a model for the probability distribution of each $m_i(t)$: the maximum and minimum values, $M_{\max}$ and $-M_{\max}$, respectively, each occur with probability $P_{\max}$; $m_i = 0$ occurs with probability P_0 ; and m_i takes on values in the range $(M_{\max}, -M_{\max})$ according to the appropriate uniform distribution. In this case,

$$\sigma_t^2 = \frac{M_{\max}^2}{3} [1 + 4P_{\max} - P_0]. \quad (2.35)$$

Although this discussion may provide some guidance in the selection of key parameters, α_t and σ_t^2 in particular can be treated as tuning parameters in the filtering equations. The first-order Gauss Markov model is also used in the parameterizations of rotational motion developed in the following sections.

2.4.4 Rotational Motion Models

The object's instantaneous orientation with respect to the earth-fixed frame is given by the basis transformation $\mathbf{I}_E^O(\zeta)$, where elements of the vector ζ depend on the parameterization selected for rotational motion. The orthogonal transformation $\mathbf{I}_E^O$ belongs to the three-dimensional rotation group for which there are about eight commonly used parameterizations [61]. Appendix A provides a review of the Euler angle-axis, quaternion, and roll-pitch-yaw parameterizations and equations which govern their temporal behavior.

A state vector, $\mathbf{x}_r \in \mathbb{R}^{N_r}$, corresponding to rotational motion is defined which contains elements of ζ and additional states which describe the temporal behavior of ζ . The dynamical model for rotational motion is written as

$$\dot{\mathbf{x}}_r(t) = \mathbf{f}_r[\mathbf{x}_r(t)] + \mathbf{G}_r \mathbf{w}_r(t), \quad (2.36)$$

where $\mathbf{f}_r[\cdot]$ is a vector-valued function, and the zero-mean Gaussian white noise process $\mathbf{w}_r(t)$ is mapped to the rotational motion state space through the matrix $\mathbf{G}_r$. In the angle-axis and roll-pitch-yaw parameterizations, approximate temporal behavior results in linear dynamic models, however, the quaternion formulation results in a nonlinear model.

In navigation systems, measurements of angular velocity are often provided by rate gyroscopes. Applications in navigation, therefore, commonly parameterize rotational motion directly in terms of angular velocity. The angular velocity of F_O with respect to F_E expressed in F_E , denoted $[\omega_E^O]_E$, is related to $\mathbf{I}_E^O$ and its time derivative according to

$$[\omega_E^O]_E^\# = \dot{\mathbf{I}}_E^O (\mathbf{I}_E^O)^T = \dot{\mathbf{I}}_E^O \mathbf{I}_O^E. \quad (2.37)$$

where $[\omega]^\#$ denotes the matrix cross-product operator corresponding to the vector $\omega = [\omega_1, \omega_2, \omega_3]^T$ and is defined as

$$\omega^\# := \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}. \quad (2.38)$$

When no confusion can result, the angular velocity will be simply written as $\omega \equiv [\omega_E^O]_E$.

Target tracking problems, however, are rarely provided with estimates of angular velocity, but instead must estimate $\omega(t)$ based on angular position information embedded in the

available measurements. In this chapter, only the quaternion parameterization is formulated directly in terms of the angular velocity; dynamical models for angle-axis and roll-pitch-yaw parameterizations employ time derivatives of rotational parameters which, in turn, can be used to compute a suboptimal estimate of angular velocity. Even in the quaternion formulation, an extended Kalman filter must be employed and hence suboptimal performance can be expected.

Angle-Axis Parameterization

An orthogonal basis transformation in $\mathbb{R}^3$ can be defined by an *orientation vector* $\xi_E = [\xi_1, \xi_2, \xi_3]^T_E$ whose direction,

$$\bar{\xi}_E := \frac{\xi_E}{\|\xi_E\|}, \quad (2.39)$$

specifies the axis of the rotation operation relating the two reference frames, and magnitude,

$$\gamma := \|\xi_E\| \quad (2.40)$$

specifies the angle through which the coordinate system is rotated about ξ_E , with the direction of rotation taken by the right-hand rule. Since elements of ξ_E are invariant under this rotation operation, the subscript 'E' is omitted. The basis transformation in this case is computed as

$$\begin{aligned} \mathbf{I}_E^O &= \exp(\xi^\times) \\ &= \mathbf{I}_3 + \sin(\gamma)\bar{\xi}^\times + [1 - \cos(\gamma)](\bar{\xi}^\times)^2, \end{aligned} \quad (2.41)$$

where $\xi^\times$ is the matrix cross product operator, defined in (2.38), associated with ξ .

The axis, $\bar{\xi}$, of $\mathbf{I}_E^O$ is generally distinct from what is commonly called the "axis of rotation" or instantaneous angular velocity².

$$\omega = \dot{\gamma}\bar{\xi} + \sin(\gamma)\dot{\bar{\xi}} + [1 - \cos(\gamma)]\bar{\xi}^\times\dot{\bar{\xi}}, \quad (2.42)$$

which has components on three orthogonal axis $\bar{\xi}$, $\dot{\bar{\xi}}$, and $\bar{\xi}^\times\dot{\bar{\xi}}$. If the axis $\bar{\xi}$ is fixed, (2.42) reduces to the simple expression $\omega = \dot{\gamma}\bar{\xi}$. However, even if the object has constant angular velocity, the axis of $\mathbf{I}_E^O$ need not remain fixed.

²All time derivatives are relative to an observer in the earth-fixed frame, and derivative vectors are expressed in the earth-fixed frame.

Under an assumption of “smooth” object motion, approximate temporal behavior of the orientation vector ξ over any sample period, $\{t|t_k \leq t \leq t_{k+1}\}$, is expressed in terms of a first-order Gauss-Markov process in a manner equivalent to that employed for translational motion in Section 2.4.3. In this case the rotational motion state vector, $\mathbf{x}_r \in \mathbb{R}^{N_r}$, where $N_r = 3(N_\xi + 1)$, is defined as

$$\mathbf{x}_r := \left[\xi^T, \dot{\xi}^T, \dots, \left(\frac{d^{N_\xi} \xi}{dt^{N_\xi}} \right)^T \right]^T, \quad (2.43)$$

and propagates in time according to

$$\dot{\mathbf{x}}_r(t) = \mathbf{A}_r \mathbf{x}_r(t) + \mathbf{G}_r \mathbf{w}_r(t), \quad (2.44)$$

where

$$\mathbf{A}_r = \begin{bmatrix} \mathbf{0}_{3N_\xi \times 3} & \mathbf{I}_{3N_\xi} \\ \mathbf{0}_{3 \times 3N_\xi} & -\alpha_r \mathbf{I}_3 \end{bmatrix}, \quad \text{and} \quad \mathbf{G}_r = \begin{bmatrix} \mathbf{0}_{3N_\xi \times 3} \\ \mathbf{I}_3 \end{bmatrix}, \quad (2.45)$$

and $\mathbf{w}_r(t)$ is zero-mean, temporally white Gaussian noise with covariance $\mathbf{Q}_r = 2\alpha_r \sigma_r^2 \mathbf{I}_3$, and the roles of the reciprocal manoeuvre correlation time constant, α_r , and the variance, σ_r^2 , of the manoeuvre variable are equivalent to that described in the development of the translational motion model.

A primary difficulty with this approach is that the orientation vector ξ must be reset, first to maintain a finite magnitude (rotation angle) when the object continually rotates in one direction, and second to yield decreasing Cramer Rao bounds as is demonstrated in Section 2.6.1, below. A trajectory $\xi(t)$ in the region $\pi < \|\xi\| < 2\pi$ is equivalent, in the sense that it produces the same rotation matrix, to a trajectory $\zeta(t)$ in the region $\|\zeta\| < \pi$ if

$$\zeta(t) = (\|\xi(t)\| - 2\pi)\bar{\xi}(t) = \xi(t) - 2\pi\bar{\xi}(t). \quad (2.46)$$

This transformation represents a reflection through the origin of $\mathbb{R}^3$, followed by a reflection in the plane tangent to the sphere of radius π at the point $-\pi\bar{\xi}$. The problem is that if $\bar{\xi}$ is non-constant, then $\zeta(t)$ may have an infinite number of non-zero time derivatives. Let the orientation vector immediately before the reset be denoted by ξ , and the corresponding orientation vector immediately following the reset be denoted by ζ . An approximate

impulsive reset which limits γ to the range $[0, \pi]$ can be applied according to the algorithm (given only to the first 2 derivatives)

If $(\|\xi\| > \pi)$ then

$$\begin{aligned}
 \gamma &= \|\xi\| \quad \{\text{Preliminary Calculations}\} \\
 \bar{\xi} &= \xi/\gamma \\
 \dot{\gamma} &= \bar{\xi}^T \dot{\xi} \\
 \dot{\bar{\xi}} &= (\mathbf{I}_3 - \bar{\xi} \bar{\xi}^T) \dot{\xi} / \gamma \\
 \ddot{\gamma} &= \bar{\xi}^T \ddot{\xi} + \dot{\bar{\xi}}^T \dot{\xi} / \gamma - (\dot{\gamma} / \gamma)^2 \\
 \ddot{\bar{\xi}} &= (\mathbf{I}_3 - \bar{\xi} \bar{\xi}^T) (\ddot{\xi} / \gamma - \dot{\bar{\xi}} \dot{\gamma} / \gamma^2) - (\dot{\bar{\xi}} \bar{\xi}^T + \bar{\xi} \dot{\bar{\xi}}^T) \dot{\xi} / \gamma \\
 &\vdots \\
 \zeta &= \xi - 2\pi \bar{\xi} \quad \{\text{Reset Procedure}\} \\
 \dot{\zeta} &= \dot{\xi} - 2\pi \dot{\bar{\xi}} \\
 \ddot{\zeta} &= \ddot{\xi} - 2\pi \ddot{\bar{\xi}} \\
 &\vdots
 \end{aligned} \tag{2.47}$$

where, as indicated, the approximate reset can be extended to further time derivatives by differentiating

$$\zeta = (\gamma - 2\pi) \bar{\xi} = \xi - 2\pi \bar{\xi}. \tag{2.48}$$

This model is completely specified by three parameters, N_ξ , α_r , and σ_r^2 , which can be selected based on the discussion of translational motion parameters in Section 2.4.3. Further discussions of the reset and expected performance of the angle-axis filter are deferred to Section 2.6 which compares Cramer Rao bounds and presents simulation results.

Quaternion Parameterization

A quaternion [59, 61, 62, 63] is a four dimensional hypercomplex number³

$$\mathbf{q} = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix} = q_1 i + q_2 j + q_3 k + q_4. \tag{2.49}$$

³Some authors place the scalar quaternion element, q_4 , in the first position while others place the scalar element in the fourth position, as is the case here, when representing the quaternion in vector form.

which is expressed in terms of basis elements consisting of the real number +1, and three imaginary units i, j, k , which satisfy

$$\begin{aligned} i^2 &= j^2 = k^2 = -1, \\ ij &= -ji = k, \\ jk &= -kj = i, \quad \text{and} \\ ki &= -ik = j. \end{aligned} \quad (2.50)$$

If $\mathbf{I}_E^O(\mathbf{q})$ rotates $\mathbb{R}^3$ through an angle γ about the axis $\bar{\xi} = [\bar{\xi}_1, \bar{\xi}_2, \bar{\xi}_3]^T$ to bring F_E into alignment with F_O , the unit quaternion defined by

$$\begin{aligned} \bar{\mathbf{q}} &:= \frac{\mathbf{q}}{\|\mathbf{q}\|} \\ &:= \left[-\bar{\xi}_1 \sin\left(\frac{\gamma}{2}\right), -\bar{\xi}_2 \sin\left(\frac{\gamma}{2}\right), -\bar{\xi}_3 \sin\left(\frac{\gamma}{2}\right), \cos\left(\frac{\gamma}{2}\right) \right]^T \end{aligned} \quad (2.51)$$

gives $\mathbf{I}_E^O(\mathbf{q})$ as

$$\mathbf{I}_E^O(\mathbf{q}) = \frac{1}{\|\mathbf{q}\|^2} \Gamma(\mathbf{q}) = \Gamma(\bar{\mathbf{q}}). \quad (2.52)$$

where

$$\Gamma(\mathbf{q}) := \begin{bmatrix} q_1^2 - q_2^2 - q_3^2 + q_4^2 & 2(q_1q_2 + q_3q_4) & 2(q_1q_3 - q_2q_4) \\ 2(q_1q_2 - q_3q_4) & -q_1^2 + q_2^2 - q_3^2 + q_4^2 & 2(q_2q_3 + q_1q_4) \\ 2(q_1q_3 + q_2q_4) & 2(q_2q_3 - q_1q_4) & -q_1^2 - q_2^2 + q_3^2 + q_4^2 \end{bmatrix}. \quad (2.53)$$

The matrix $\Gamma(\mathbf{q})$ represents combined operations of an orthogonal transformation ($\mathbf{I}_E^O$) and uniform scaling by $\|\mathbf{q}\|^2$. The quaternion parameterization is four-dimensional and is often fixed with a quadratic constraint of the form $\|\mathbf{q}\| = 1$ since $\mathbf{I}_E^O(\mathbf{q})$ depends only on the normalized quaternion $\bar{\mathbf{q}}$. Such constraints are not easily incorporated into the Kalman filtering equations due to the linear structure of the filter.

A common approach, [35], [48]-[56], is to implement the filter based on unit quaternions and employ an impulsive normalization of the estimated quaternion, using standard Kalman filtering notation from Appendix B,

$$\hat{\mathbf{q}}(k|k) \leftarrow \frac{\hat{\mathbf{q}}(k|k)}{\|\hat{\mathbf{q}}(k|k)\|}, \quad (2.54)$$

immediately following each observation event. In the present work, this approach resulted in very poor and often divergent performance of the extended Kalman filter. Tahk and

Spayer [64] propose to incorporate soft or stochastic constraints directly into the measurement model and apply the extended Kalman filter to the augmented system. This approach, although it resulted in an improvement over impulsive normalization, also led to poor and often divergent behavior of the filter. Bar-Itzack and Oshman [59] treat the problem of estimating, in the current notation, a unit quaternion $\bar{\mathbf{q}}$, and hence $\mathbf{I}_E^O(\bar{\mathbf{q}})$, from sequences of observations of corresponding vectors $\{(\mathbf{r}_E^i, \mathbf{r}_C^i); i = 1, 2, \dots, N\}$, and angular velocity $\boldsymbol{\omega}$ over time with an extended Kalman filter. With a first-order approximation, their approach is to propagate the estimated quaternion in the filter without impulsive normalization, which means that the estimated quaternion need not have unit norm. For output purposes only, the estimated quaternion is normalized through division by its Euclidean norm [65].

The dynamical model is based on the assumption that the quaternion has constant norm (not necessarily unity) over any sample period so that

$$\dot{\mathbf{q}}(t) = \|\mathbf{q}(t)\|\dot{\bar{\mathbf{q}}}(t). \quad (2.55)$$

The unit quaternion, $\bar{\mathbf{q}}(t)$ defined in (2.51), propagates in time (see Appendix A) according to the differential equation

$$\dot{\bar{\mathbf{q}}}(t) = \boldsymbol{\Omega}[\boldsymbol{\omega}(t)]\bar{\mathbf{q}}(t), \quad (2.56)$$

where $\boldsymbol{\omega} \equiv [\boldsymbol{\omega}_E^O]_E = [\omega_1, \omega_2, \omega_3]^T$ and

$$\boldsymbol{\Omega}[\boldsymbol{\omega}] := \frac{1}{2} \begin{bmatrix} \boldsymbol{\omega}^\times & -\boldsymbol{\omega} \\ \boldsymbol{\omega}^T & 0 \end{bmatrix}. \quad (2.57)$$

Multiplying both sides of (2.56) by $\|\mathbf{q}\|$ gives, with (2.55),

$$\dot{\mathbf{q}}(t) = \boldsymbol{\Omega}[\boldsymbol{\omega}(t)]\mathbf{q}(t). \quad (2.58)$$

Once again employing the assumption that rotational motion is “smooth”, temporal behavior of $\boldsymbol{\omega}$ over any sample period is approximated by a first-order Gauss-Markov process in a manner equivalent to that employed for translational motion in Section 2.4.3. A state vector, $\mathbf{x}_u$, for the angular velocity and its time derivatives is defined as

$$\mathbf{x}_u := \left[\boldsymbol{\omega}^T, \dot{\boldsymbol{\omega}}^T, \dots, \left(\frac{d^{N_u} \boldsymbol{\omega}}{dt^{N_u}} \right)^T \right]^T. \quad (2.59)$$

which propagates in time according to

$$\dot{\mathbf{x}}_w(t) = \mathbf{A}_w \mathbf{x}_w(t) + \mathbf{G}_w \mathbf{w}_r(t), \quad (2.60)$$

where

$$\mathbf{A}_w = \begin{bmatrix} \mathbf{0}_{3N_w \times 3} & \mathbf{I}_{3N_w} \\ \mathbf{0}_{3 \times 3N_w} & -\alpha_r \mathbf{I}_3 \end{bmatrix}, \text{ and } \mathbf{G}_w = \begin{bmatrix} \mathbf{0}_{3N_w \times 3} \\ \mathbf{I}_3 \end{bmatrix}, \quad (2.61)$$

and $\mathbf{w}_r(t)$ is zero-mean, temporally white Gaussian noise with covariance $\mathbf{Q}_r = 2\alpha_r \sigma_r^2 \mathbf{I}_3$, and the roles of the reciprocal manoeuvre correlation time constant, α_r , and the variance, σ_r^2 , of the manoeuvre variable are equivalent to that described in the development of the translational motion model.

An overall state vector, $\mathbf{x}_r$, corresponding to rotational motion is defined as

$$\mathbf{x}_r := [\mathbf{q}^T, \mathbf{x}_w^T]^T, \quad (2.62)$$

which propagates in time according to the nonlinear differential equation

$$\begin{aligned} \dot{\mathbf{x}}_r(t) &= \mathbf{f}_r[\mathbf{x}_r(t)] + \mathbf{G}_r \mathbf{w}_r(t) \\ &= \begin{bmatrix} \Omega[\omega(t)] & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_w \end{bmatrix} \mathbf{x}_r(t) + \begin{bmatrix} \mathbf{0}_{4 \times 3} \\ \mathbf{G}_w \end{bmatrix} \mathbf{w}_r(t). \end{aligned} \quad (2.63)$$

Time propagation in the extended Kalman filter requires numerical integration of (2.63) with $\mathbf{G}_r \mathbf{w}_r(t) = \mathbf{0}$.

If ω is approximately constant over any sample period, an approximate discrete-time difference equation can be obtained for $\mathbf{x}_r$ of the form

$$\mathbf{x}_r(k+1) = \Phi_r(k) \mathbf{x}_r(k), \quad (2.64)$$

where

$$\Phi_r(k) := \text{diag}[\Phi_q(k), \Phi_w(k)], \quad (2.65)$$

and

$$\Phi_q(k) := \exp(\Omega[\omega(k)]\{t_{k+1} - t_k\}) \quad (2.66)$$

$$\Phi_w(k) := \exp(\mathbf{A}_w\{t_{k+1} - t_k\}). \quad (2.67)$$

A closed-form expression for Φ_q (see Appendix A) is

$$\Phi_q(k) = \cos\left(\|\omega\| \frac{(t_{k+1} - t_k)}{2}\right) \mathbf{I}_4 + \sin\left(\|\omega\| \frac{(t_{k+1} - t_k)}{2}\right) \frac{2}{\|\omega\|} \Omega[\omega]. \quad (2.68)$$

Because $\Omega[\omega]$ is a skew symmetric matrix, $\Phi_q(k)$ is orthogonal⁴ and hence $\|\mathbf{q}(k+1)\| = \|\mathbf{q}(k)\|$ as required.

Numerical integration in (2.64) is based on the assumption that ω is approximately constant over the sample period. This will be essentially true provided either ω is constant, or the sample period is very small. An assumption of small sample periods is often employed in motion analysis, but is very restrictive. Moreover, this approximation does not exploit estimates of time derivatives of ω . Many standard methods for numerical integration of the quaternion differential equation are designed for navigation systems in which measurements of angular velocity are available directly [62]. The following algorithm is similar in form to one analyzed by Branets and Shmyglevski [63]. This approach employs the modelled temporal behavior of ω and the most recent estimates of ω and its time derivatives.

Define a partition

$$\Pi(k) := \{t_k = \tau_0 \leq \tau_1 \leq \dots \leq \tau_{N_\pi} = t_{k+1}\} \quad (2.69)$$

of the k th sample period, and assume for convenience that $\tau_{\ell+1} - \tau_\ell := \tau$ is constant. With the linear model for $\mathbf{x}_u$,

$$\begin{aligned} \mathbf{x}_u(\tau_\ell) &= [\Phi_u(\tau)]^\ell \mathbf{x}_u(t_k) \\ \omega(\tau_\ell) &= \mathbf{C}_u \mathbf{x}_u(\tau_\ell) \end{aligned} \quad (2.70)$$

where

$$\begin{aligned} \Phi_u(\tau) &:= e^{\mathbf{A}_u \tau}, \text{ and} \\ \mathbf{C}_u &:= \begin{bmatrix} 1.0 & 0.0 & 0.0 & \dots & 0 \\ 0.1 & 0.0 & 0.0 & \dots & 0 \\ 0.0 & 0.1 & 0.0 & \dots & 0 \end{bmatrix}. \end{aligned} \quad (2.71)$$

Assuming that the highest time derivative of ω in $\mathbf{x}_u$ is approximately constant over any sample period, and provided τ is chosen small enough, or that N_π is chosen large enough, so

⁴The matrix exponential of a skew symmetric matrix is always orthogonal

that ω is approximately constant over any interval of the partition $\Pi(k)$, time propagation of $\mathbf{q}$ over the complete sample period may be approximated by

$$\mathbf{q}(k+1) = \Phi_q(k)\mathbf{q}(k), \quad (2.72)$$

where

$$\Phi_q(k) := \downarrow \prod_{\ell=0}^{N_\pi-1} \exp \left\{ \Omega \left[\mathbf{C}_w(\Phi_w(\tau))^\ell \mathbf{x}_w(t_k) \right] \tau \right\}, \quad (2.73)$$

in which the notation $\downarrow \prod$ indicates that the index ℓ decreases from left to right in the matrix product, $\Omega[\cdot] : \mathbb{R}^3 \rightarrow \mathbb{R}^{4 \times 4}$ according to (2.57), and (2.68) can be used to evaluate the matrix exponentials. Note that Φ_q in (2.73) is a product of orthogonal matrices and hence is orthogonal, which means that $\|\mathbf{q}(k+1)\| = \|\mathbf{q}(k)\|$.

It should be noted that the continuous-time form, $\Phi_r(t, \lambda)$, of $\Phi_r(k)$ obtained by replacing t_{k+1} with t and t_k by λ in (2.66) or (2.73) and (2.67), is not a "state transition matrix" in the usual sense because it does **not** satisfy the fundamental property

$$\frac{d\Phi_r(t, \lambda)}{dt} = \frac{\partial \mathbf{f}_r[\mathbf{x}_r]}{\partial \mathbf{x}_r} \Phi_r(t, \lambda), \quad (\text{false}). \quad (2.74)$$

As a result, even though Φ_r may be used for approximate time propagation of the state estimate, this matrix cannot be used to propagate the Kalman covariance matrix over time. At any measurement event, the measurements are directly related to the quaternions and not to the angular velocity. The Kalman filter must use the dynamical model to extract information about angular velocity from prediction errors which, in turn, depend on the quaternion estimates. Time propagation of the Kalman covariance, therefore, must maintain the interaction between quaternions and the angular velocity specified in (2.58). The block diagonal matrix Φ_r in (2.65) does not explicitly model the interdependence of $\mathbf{q}$ and $\mathbf{w}$. Approximate time propagation, [35], [48]-[56], of the covariance matrix is performed with an approximate state transition matrix which employs the rotational motion component as

$$\hat{\Phi}_r(k) = \exp\{\mathbf{F}_r(k)[t_{k+1} - t_k]\}, \quad (2.75)$$

where $\mathbf{F}_r(k)$ is the Jacobian of $\mathbf{f}_r[\mathbf{x}_r(k)]$,

$$\begin{aligned} \mathbf{F}_r(k) &:= \frac{\partial \mathbf{f}_r[\mathbf{x}_r]}{\partial \mathbf{x}_r} \\ &:= \begin{bmatrix} \Omega[\omega(k)] & [\dot{\mathbf{q}}_w(k) \ 0] \\ 0 & \mathbf{A}_w \end{bmatrix}, \end{aligned} \quad (2.76)$$

and

$$\dot{\mathbf{q}}_w(k) := \frac{\partial \dot{\mathbf{q}}}{\partial \boldsymbol{\omega}}(k) = \frac{1}{2} \begin{bmatrix} -q_4 & q_3 & -q_2 \\ -q_3 & -q_4 & q_1 \\ q_2 & -q_1 & -q_4 \\ q_1 & q_2 & q_3 \end{bmatrix}. \quad (2.77)$$

Note that $\hat{\Phi}_r(k)$ in (2.75) need not in general maintain constant quaternion norm and hence should not be used in time propagation of the state estimate.

The quaternion filter has been implemented, using standard Kalman filtering notation introduced in Appendix B, with the following approach:

1. The filter is initialized with a unit quaternion, $\|\hat{\mathbf{q}}(0|-1)\| = 1$;
2. The norm of the estimated quaternion is invariant under time propagation, i.e., $\|\hat{\mathbf{q}}(k+1|k)\| = \|\hat{\mathbf{q}}(k|k)\|$, $k = 1, 2, 3, \dots$;
3. The measurement model employs $\Gamma(\hat{\mathbf{q}})$ of (2.53) as the transformation from F_O to F_E instead of the strictly orthogonal matrix $\mathbf{I}_E^O$;
4. For output purposes only, the estimated unit quaternion is taken as

$$\bar{\hat{\mathbf{q}}}(k|k) := \frac{\hat{\mathbf{q}}(k|k)}{\|\hat{\mathbf{q}}(k|k)\|}. \quad (2.78)$$

and, again for output purposes only, structure estimates denoted ${}^*\hat{\mathbf{r}}_O^i(k|k)$ are taken as

$${}^*\hat{\mathbf{r}}_O^i(k|k) := \|\hat{\mathbf{q}}(k|k)\|^2 \hat{\mathbf{r}}_O^i(k|k). \quad (2.79)$$

for $i = 1, 2, \dots, N_f$, where the vectors $\hat{\mathbf{r}}_O^i(k|k)$ are structure estimates generated within the filter;

5. In the k th measurement update, $\hat{\mathbf{d}}_0 := \mathbf{d}_0 / \|\hat{\mathbf{q}}(k|k-1)\|^2$ is used in place of $\mathbf{d}_0$ within the filter
6. The estimation error covariance matrix, $\hat{\mathbf{P}}^*(k|k)$, for the filter output is taken as

$$\hat{\mathbf{P}}^*(k|k) := \beta \hat{\mathbf{P}}(k|k) \beta^T, \quad (2.80)$$

where $\hat{\mathbf{P}}(k|k)$ is the estimated Kalman covariance, and

$$\beta := \text{diag} \left\{ \mathbf{I}_{N_t}, \frac{1}{\|\hat{\mathbf{q}}(k|k)\|} \mathbf{I}_4, \mathbf{I}_{N_w}, \|\hat{\mathbf{q}}(k|k)\|^2 \mathbf{I}_{3N_f} \right\}. \quad (2.81)$$

This approach parallels that proposed by Bar-Itzhack and Oshman [59].

Due to Item 3 above, the measurement model depends on the product of $\Gamma(q)$ and structure vectors $\mathbf{r}_O^i$ internal to the filter which appears in the expression,

$$\mathbf{r}_E^i = \Gamma(\mathbf{q})\mathbf{r}_O^i + [\mathbf{T}_E^O]_E. \quad (2.82)$$

The innovation sequence in the quaternion filter, therefore, depends on the product

$$[\widehat{\mathbf{r}_O^i}]_E := \Gamma(\hat{\mathbf{q}})\hat{\mathbf{r}}_O^i. \quad (2.83)$$

The use of $\Gamma(\hat{\mathbf{q}})$ means that all structure estimates within the filter, $\hat{\mathbf{r}}_O^i$, are scaled by $\|\hat{\mathbf{q}}\|^2$ in the transformation from F_O to F_E . If the quaternion filter works well, $[\widehat{\mathbf{r}_O^i}]_E \rightarrow [\mathbf{r}_O^i]_E$ for $i = 1, 2, \dots, N_f$, even though $\hat{\mathbf{q}}$ may not have unit norm, and $\hat{\mathbf{r}}_O^i$ may not converge to a true structure vector (due to the scaling effect of $\Gamma(\mathbf{q})$). The result of Item 4 is that structural estimates when referenced to the earth-fixed frame within the filter exactly match those extracted from the filter for output purposes; from (2.52) and (2.53),

$$\begin{aligned} [\widehat{\mathbf{r}_O^i}]_E(k|k) &:= \Gamma(\hat{\mathbf{q}}(k|k))\hat{\mathbf{r}}_O^i(k|k) \\ &= \left\{ \frac{1}{\|\hat{\mathbf{q}}\|^2} \Gamma(\hat{\mathbf{q}}(k|k)) \right\} \{ \|\hat{\mathbf{q}}(k|k)\|^2 \hat{\mathbf{r}}_O^i(k|k) \} \\ &= \{ \mathbf{I}_E^O(\tilde{\mathbf{q}}(k|k)) \} \{ {}^* \hat{\mathbf{r}}_O^i(k|k) \} \end{aligned} \quad (2.84)$$

which leads naturally to the extraction, in Item 4 above, of a unit quaternion and appropriate structure estimates.

However, two problems arise from the inability to enforce the constraint $\|\hat{\mathbf{q}}\| = 1$ throughout the measurement update. First, structure estimation errors in the output are computed from

$$\mathbf{r}_O^i - \|\hat{\mathbf{q}}(k|k)\|^2 \hat{\mathbf{r}}_O^i(k|k) \quad (2.85)$$

which demonstrates an extra degree of freedom in the filter that can result in a reduced rate of accumulation of information concerning object structure. This difficulty is demonstrated

through Cramer-Rao bound analysis in Section 2.6.1. The second problem is that structural constraints based on the known vector $\mathbf{d}_O$, which defines the position and orientation of F_O with respect to three observed feature points, are not maintained throughout the measurement update. On entry to the measurement update equations these constraints are satisfied in the sense that

$$\Gamma(\hat{\mathbf{q}}(k|k-1))\hat{\mathbf{d}}_0 = \mathbf{I}_E^O(\hat{\mathbf{q}}(k|k-1))\mathbf{d}_O. \quad (2.86)$$

However, because the quaternion norm is not generally invariant over the measurement update, these constraints may not be precisely satisfied on exit from the measurement update equations; on exit

$$\Gamma(\hat{\mathbf{q}}(k|k))\hat{\mathbf{d}}_0 = \mathbf{I}_E^O(\hat{\mathbf{q}}(k|k))\mathbf{d}_O \{ \|\mathbf{q}(k|k)\|^2 / \|\mathbf{q}(k|k-1)\|^2 \}. \quad (2.87)$$

This inherent scaling of the structural constraints by $\|\mathbf{q}(k|k)\|^2 / \|\mathbf{q}(k|k-1)\|^2$ over the measurement update can effect structure estimates as well as position and translational motion estimates.

The quaternion-based model is completely specified by four parameters, N_u , α_r , σ_r^2 , and N_π . The first three of these can be selected based on the discussion of translational motion parameters in Section 2.4.3. Selection of number of intervals, N_π , in the partition of each sample period should also be based in part on the expected complexity of rotational motion. Investigations leading to the research reported herein indicated that $N_\pi = 5$ with a sample period of $T = 0.1$ seconds gives reasonable accuracy in ideal time propagation of the state at the expense of increased computational complexity.

Roll-pitch-yaw Parameterization

The roll-pitch-yaw (RPY) representation [32, 66] is a particular case of an Euler angle parameterization of the three-dimensional rotation group. The roll, pitch, and yaw angles, denoted ϕ , θ , and ψ , respectively, define an ordered sequence of three plane rotations which can be used to express $\mathbf{I}_E^O$ as

$$\mathbf{I}_E^O = \exp(\psi \mathbf{e}_3^e) \exp(\theta \mathbf{e}_2^e) \exp(\phi \mathbf{e}_1^e). \quad (2.88)$$

where e_i , $i = 1, 2, 3$, are the standard basis vectors of $\mathbb{R}^3$. Expanding the exponentials with (A.13) gives three plane rotations,

$$\begin{aligned} \exp(\phi e_1^\#) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\phi) & -\sin(\phi) \\ 0 & \sin(\phi) & \cos(\phi) \end{bmatrix}, \\ \exp(\theta e_2^\#) &= \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix}, \text{ and} \\ \exp(\psi e_3^\#) &= \begin{bmatrix} \cos(\psi) & -\sin(\psi) & 0 \\ \sin(\psi) & \cos(\psi) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned} \quad (2.89)$$

The angles ϕ , θ , and ψ have positive sense defined by the right hand rule about their respective rotation axis (see Appendix A).

It has been reported, [48, 49, 54], that major difficulties arise with this approach because the sequence of plane rotations are not defined about orthogonal axis. This property results in highly nonlinear dynamical models if rotational motion is formulated directly in terms of angular velocity. The angular velocity $\omega \equiv [\omega_E^O]_E$, is given in terms of time derivatives of roll, pitch, and yaw angles as

$$\omega = J_{RPY}(\phi, \theta, \psi) \begin{bmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\psi} \end{bmatrix}, \quad (2.90)$$

where

$$J_{RPY}(\phi, \theta, \psi) = \begin{bmatrix} \cos(\psi) \cos(\theta) & -\sin(\psi) & 0 \\ \sin(\psi) \cos(\theta) & \cos(\psi) & 0 \\ -\sin(\theta) & 0 & 1 \end{bmatrix}. \quad (2.91)$$

The mathematical singularities, $\theta = \pm\pi/2$, characterize orientations where the determinant of the Jacobian matrix $\det[J_{RPY}] = -\cos(\theta)$ vanishes. A simple calculation shows that at these orientations any nonzero component of angular velocity in $\text{span}\{[\cos(\psi), \sin(\psi), 0]^T\}$ cannot be represented by this parameterization with finite $\dot{\phi}$, $\dot{\theta}$, and $\dot{\psi}$. Computing ω from

the state estimate in this manner can, therefore, lead to significant errors about the singular orientations $\theta = \pm\pi/2$ which represents the primary difficulty with state estimation using this parameterization.

The parameter vector for the roll-pitch-yaw filter is defined as

$$\zeta := [\phi, \theta, \psi]^T. \quad (2.92)$$

Under an assumption of "smooth" object motion, an approximate temporal behavior of ζ over any sample period, $\{t|t_k \leq t \leq t_{k+1}\}$, is expressed in terms of a first-order Gauss-Markov process in a manner equivalent to that employed for translational motion in Section 2.4.3. In this case the rotational motion state vector, $\mathbf{x}_r \in \mathbb{R}^{N_r}$, where $N_r = 3(N_\zeta + 1)$, is defined as

$$\mathbf{x}_r := \left[\zeta^T, \dot{\zeta}^T, \dots, \left(\frac{d^{N_\zeta} \zeta}{dt^{N_\zeta}} \right)^T \right]^T. \quad (2.93)$$

and propagates in time according to

$$\dot{\mathbf{x}}_r(t) = \mathbf{A}_r \mathbf{x}_r(t) + \mathbf{G}_r \mathbf{w}_r(t), \quad (2.94)$$

where

$$\mathbf{A}_r = \begin{bmatrix} \mathbf{0}_{3N_\zeta \times 3} & \mathbf{I}_{3N_\zeta} \\ \mathbf{0}_{3 \times 3N_\zeta} & -\alpha_r \mathbf{I}_3 \end{bmatrix}, \text{ and } \mathbf{G}_r = \begin{bmatrix} \mathbf{0}_{3N_\zeta \times 3} \\ \mathbf{I}_3 \end{bmatrix}. \quad (2.95)$$

and $\mathbf{w}_r(t)$ is zero-mean, temporally white Gaussian noise with covariance $\mathbf{Q}_r = 2\alpha_r \sigma_r^2 \mathbf{I}_3$, and the roles of the reciprocal manoeuvre correlation time constant, α_r , and the variance, σ_r^2 , of the manoeuvre variable are equivalent to that described in the development of the translational motion model.

The roll-pitch-yaw model is completely specified by three parameters, N_ζ , α_r , and σ_r^2 which, again, can be selected based on the discussion of translational motion parameters in Section 2.4.3.

2.5 Obtaining Initial Estimates

Any observer requires an initial estimate, denoted in this Chapter as $\hat{\mathbf{x}}(0|-1)$, of the true state. The extended Kalman filter also requires an estimate of the initial uncertainty,

$P(0|-1)$, in this initial estimate. Broida and Chellappa, for example, have studied a batch approach [51, 53, 54, 48] which employs the Conjugate Gradient descent technique from the IMSL library. Simulation results shown by Broida *et al.*, [54], employ what are called “crude” initial estimates obtained from the first 10 image frames with 75 iterations of the batch technique.

For the purpose of the present investigation, a simple single-frame initialization provides suitable estimates of position, orientation, and structure, while all higher-order motion parameters (time derivatives) are initialized as zero. The initialization algorithm is described as follows:

- Step 1** Obtain an estimate of the 3D positions of three or more feature points with respect to the earth-fixed frame through stereopsis based on measurements in multiple-camera images and assumed correspondence information;
- Step 2** Select three noncollinear feature points from all available 3D points;
- Step 3** Estimate the initial position of the object, form an orthonormal, right-handed triad, F_O , in F_E from the three special 3D points based on constraints detailed in Section 2.4.1 and illustrated in Figure 2.3, and fix the origin, O_O , of F_O with respect to the 3D feature points by specifying the vector $\mathbf{d}$;
- Step 4** Extract a valid set of orientation parameters from the basis vectors of F_O expressed in F_E ;
- Step 5** Express positions of all feature points and $\mathbf{d}$ with respect to F_O .

This initialization is concerned only with obtaining an initial estimate of the state; the initial covariance matrix for the extended Kalman filters is selected as diagonal assuming significant initialization errors (fairly large diagonal entries).

Step 1: Given $N_f \geq 3$ feature points visible in each of $N_C \geq 2$ cameras with spatial correspondence information, the objective is to estimate $\mathbf{r}'_E, i = 1, 2, \dots, N_f$. The position

of the i th feature point in the image plane of the j th camera is written, from (2.16), as

$$\mathbf{z}_j^i := \begin{bmatrix} u_j^i \\ v_j^i \end{bmatrix} = L_j \begin{bmatrix} \frac{r_{C_j,2}^i}{r_{C_j,1}^i} \\ \frac{r_{C_j,3}^i}{r_{C_j,1}^i} \end{bmatrix}. \quad (2.96)$$

With (2.96) and

$$\mathbf{D}_j^i := \begin{bmatrix} u_j^i & -L_j & 0 \\ v_j^i & 0 & -L_j \end{bmatrix}, \quad (2.97)$$

so that $\mathbf{D}_j^i \mathbf{r}_{C_j}^i = 0$ and

$$\mathbf{r}_{C_j}^i = \mathbf{I}_{C_j}^E \mathbf{r}_E^i + [\mathbf{T}_{E^i}^{C_j}]_{C_j}, \quad (2.98)$$

where $\mathbf{I}_{C_j}^E$ and $[\mathbf{T}_{E^i}^{C_j}]_{C_j}$ are known for $j = 1, 2, \dots, N_C$, a system of linear equations can be written in the form

$$\mathbf{D}_j^i \mathbf{I}_{C_j}^E \mathbf{r}_E^i = -\mathbf{D}_j^i [\mathbf{T}_{E^i}^{C_j}]_{C_j}, \quad j = 1, 2, \dots, N_C. \quad (2.99)$$

Forming a matrix $\mathbf{M}^i$ and a vector $\mathbf{N}^i$ defined as

$$\mathbf{M}^i := \begin{bmatrix} \mathbf{D}_1^i \mathbf{I}_{C_1}^E \\ \mathbf{D}_2^i \mathbf{I}_{C_2}^E \\ \vdots \\ \mathbf{D}_{N_C}^i \mathbf{I}_{C_{N_C}}^E \end{bmatrix}, \quad \text{and} \quad \mathbf{N}^i := \begin{bmatrix} \mathbf{D}_1^i [\mathbf{T}_{E^i}^{C_1}]_{C_1} \\ \mathbf{D}_2^i [\mathbf{T}_{E^i}^{C_2}]_{C_2} \\ \vdots \\ \mathbf{D}_{N_C}^i [\mathbf{T}_{E^i}^{C_{N_C}}]_{C_{N_C}} \end{bmatrix}. \quad (2.100)$$

immediately gives the least-mean-squared estimate $\hat{\mathbf{r}}_E^i$ of $\mathbf{r}_E^i$ as

$$\hat{\mathbf{r}}_E^i = \left[(\mathbf{M}^i)^T \mathbf{M}^i \right]^{-1} (\mathbf{M}^i)^T \mathbf{N}^i \quad (2.101)$$

for the i th feature point. This process is repeated for each of the N_f feature points.

Step 2: Many criteria could be specified for selecting the best set of three special non-collinear feature points from all available $\hat{\mathbf{r}}_E^i$. A simple approach is to select those three points for which the area of the triangle (see Figure 2.3) formed by joining the three 3D points is maximized, thus maximizing a measure of the degree of noncollinearity and distance between points. This can be done by selecting 3 indices, $\{i_1, i_2, i_3\} \subseteq \{1, 2, \dots, N_f\}$ such that,

$$\| [\hat{\mathbf{r}}_E^{i_2} - \hat{\mathbf{r}}_E^{i_1}]^T [\hat{\mathbf{r}}_E^{i_3} - \hat{\mathbf{r}}_E^{i_1}] \|^2 \quad (2.102)$$

is maximized. Since the order of selection is unimportant, there are

$$\binom{N_f}{3} := \frac{N_f!}{3!(N_f - 3)!} \quad (2.103)$$

combinations for comparison.

Step 3: The estimated initial position, $\hat{\mathbf{p}}(0) := [\widehat{\mathbf{T}}_O^E](0)$, of the object is taken as the sample mean of all estimated feature point positions:

$$\hat{\mathbf{p}}(0) = \frac{1}{N_f} \sum_{i=1}^{N_f} \hat{\mathbf{r}}_E^i. \quad (2.104)$$

The three feature points selected in Step 2 are ordered, arbitrarily in the absence of prior information, as $\{\hat{\mathbf{r}}_E^{i_0}, \hat{\mathbf{r}}_E^{i_r}, \hat{\mathbf{r}}_E^{i_\nu}\}$ according to the constraints detailed in Section 2.4.1. The object-centred frame is fixed on the object, and the initial estimates of orientation are provided with

$$[\mathbf{d}_O]_E = \hat{\mathbf{r}}_E^{i_0} - \hat{\mathbf{p}}(0), \quad (2.105)$$

$$[\widehat{\mathbf{i}}_O]_E = \frac{\hat{\mathbf{r}}_E^{i_r} - \hat{\mathbf{r}}_E^{i_0}}{\|\hat{\mathbf{r}}_E^{i_r} - \hat{\mathbf{r}}_E^{i_0}\|}, \quad (2.106)$$

$$[\widehat{\mathbf{j}}_O]_E = \frac{[\mathbf{I}_3 - [\widehat{\mathbf{i}}_O]_E([\widehat{\mathbf{i}}_O]_E)^T]\hat{\mathbf{r}}_E^{i_\nu}}{\|[\mathbf{I}_3 - [\widehat{\mathbf{i}}_O]_E([\widehat{\mathbf{i}}_O]_E)^T]\hat{\mathbf{r}}_E^{i_\nu}\|}, \text{ and} \quad (2.107)$$

$$[\widehat{\mathbf{k}}_O]_E = [\widehat{\mathbf{i}}_O]_E \times [\widehat{\mathbf{j}}_O]_E. \quad (2.108)$$

Step 4: Initial estimates of orientation parameters (angle-axis, quaternion, or roll-pitch-yaw) are extracted from the initial estimate of the basis transformation

$$\widehat{\mathbf{I}}_E^O = [\widehat{\mathbf{i}}_O]_E, [\widehat{\mathbf{j}}_O]_E, [\widehat{\mathbf{k}}_O]_E, \quad (2.109)$$

according to methods outlined in Appendix A.

Step 5: Initial structure estimates with respect to F_O are computed with

$$\mathbf{d}_O = \widehat{\mathbf{I}}_E^O{}^T [\mathbf{d}_O]_E, \text{ and} \quad (2.110)$$

$$\hat{\mathbf{r}}_O^i = \widehat{\mathbf{I}}_E^O{}^T (\hat{\mathbf{r}}_E^i - \hat{\mathbf{p}}(0)), \quad i = 1, 2, \dots, N_f. \quad (2.111)$$

By construction, $\hat{\mathbf{r}}_O^{i_0} = \mathbf{d}_O$, $\hat{\mathbf{r}}_O^{i_r} = d_{O,2}$, $\hat{\mathbf{r}}_O^{i_r} = d_{O,3}$, and $\hat{\mathbf{r}}_O^{i_\nu} = d_{O,3}$, where $\mathbf{d}_O$ is now treated as a known constant vector; these six structural parameters are not estimated by the filters.

2.6 EKF Performance With Simulated Imagery

The previous section introduced motion, structure and measurement models based on three parameterizations of rotational motion. In this section, relative performance of three iterated extended Kalman filters, each based on a single parameterization, is demonstrated through presentation of a subset of results generated from extensive Monte Carlo simulations. One limitation of analysis based on simulated imagery is that the assumed central projection image formation model is exact; that is, the measurement model is based on the central projection transformation, and simulated imagery is generated precisely from this model. This approach, which is common to most simulation studies in image analysis, is only valid if distortion effects in the imaging system are small in comparison to measurement noise levels. Evaluation of motion and structure estimation schemes on simulated imagery, however, provides an efficient means to verify and compare performance of the three filtering algorithms.

The geometry of Figure 2.2 was used to generate binocular image sequences of a single, rigid, translating and rotating block. Figure 2.4 shows the object used for simulations with the convention for labelling of feature points. Features of interest are four, p^3 , p^5 , p^6 , and p^7 , of the eight corners of a rectangular block 20 units wide ($p^6 \rightarrow p^3$) and 10 units in height ($p^6 \rightarrow p^5$) and depth ($p^6 \rightarrow p^7$). Feature detection and both temporal and spatial correspondence information is provided by the imagery generation program, and occlusion is not considered in this analysis.

In all simulations of this section, two cameras are employed with unity focal length. The position and orientation of the left camera (C_1) is defined by a translation $[\mathbf{T}_{C_1}^E]_E = [0, -5, -30]^T_E$ units, and transformation $\mathbf{I}_{C_1}^E$ formed from Euler angles $\psi_{C_1} = 10^\circ$ (yaw or pan), $\theta_{C_1} = -45^\circ$ (pitch or tilt), and $\phi_{C_1} = 0$ (roll about optical axis). The position and orientation of the right camera (C_2) is defined, respectively, by $[\mathbf{T}_{C_2}^E]_E = [0, 5, -30]^T_E$ units, and $\mathbf{I}_{C_2}^E$ formed from Euler angles $\psi_{C_2} = -10^\circ$, $\theta_{C_2} = -45^\circ$, and $\phi_{C_2} = 0$. Measurement error in feature-based motion analysis studies is often taken as uniform digitization noise [54, 56]. Simulations of this section assume moderately high levels of measurement noise in the range ± 0.02 units in the image plane, which corresponds to errors on the order of 4% to 8% of the size of the object in the image when the object is closest to the cameras. Mea-

Figure 2.4: Object used in simulation studies. Feature points of interest are labelled p^3 , p^5 , p^6 , and p^7 .

surement models in the Kalman filters assume Gaussian statistics with standard deviation 0.02 units in both horizontal and vertical components. As a result, the measurement models are not exactly matched to the true statistics of measurement error, as would probably be the case in any implementation.

Section 2.6.1 provides a brief investigation of Cramer Rao bounds, first for the resetting requirement of the angle-axis filter, and second to compare optimal performance for the three filters. Section 2.6.2 presents and discusses simulation results for the three filters in order to highlight the advantages and disadvantages of each parameterization. The true object trajectory employed in Case 1 of Section 2.6.2 is used repeatedly in later Chapters to illustrate performance of proposed sliding mode observers in various components of the image analysis problem.

2.6.1 Comparison of Cramer Rao Bounds

The extended Kalman filter is generally a nonlinear, suboptimal state estimation technique. Performance assessment of such techniques often employ the Cramer-Rao inequality [67] to derive covariance lower bounds. Appendix C reviews the significance and derivation of Cramer-Rao lower bounds (CRLB) for this estimation problem. Generally, the error covari-

ance matrix $S(N)$ of an *unbiased* estimator $\hat{x}(N)$ of the state $x(N)$, where the estimate is based on the prior information in x_0 and the observations $z(0), \dots, z(N)$, is lower-bounded by the inverse of Fisher's information matrix [68, pp. 91-93]. This result, in theory, provides estimation error variance lower bounds for elements of $\hat{x}(N)$. It should be noted, however, that these bounds assume unbiased state estimation; a state dependent bias can result in higher or lower bounds [13] (see Appendix C). As a result, most analysis [49]-[53], [69], [70], proceed under the assumption of unbiased estimation resulting in approximate covariance bounds.

A further assumption in deriving Cramer Rao bounds is that the dynamic model is noise-free. This means that the dynamic model from which the filter is constructed is assumed to exactly match the deterministic temporal behavior of observed processes. For the purposes of satisfying this assumption during investigations of this section, modeling parameters α_t , α_r , σ_t , and σ_r are set to zero, thereby eliminating process noise and reducing the dynamic models to truncated Taylor series expansions, the initialization parameter vector d_0 is set to zero, which means that the origin of the object centred reference frame lies at a feature point, and true object motion is defined such that it can be exactly modelled by any of the three parameterizations of rotational motion. These restrictions obviously place some limitations on the overall scope of evaluations with Cramer Rao bounds, however, comparisons of optimal performance possible under these ideal circumstances provides a general appreciation for accuracy that can be expected for more complex motion investigated in Section 2.6.2.

It was noted in Section 2.4.4 that the orientation vector of the angle-axis filter must be reset, first to maintain a finite magnitude (rotation angle) when the object continually rotates in one direction, and second to yield decreasing Cramer Rao bounds. Figure 2.5 shows Cramer Rao bounds in the absence of prior information ($P_0^{-1} = 0$) for the rotational state vector x_r defined in (2.43) for purely rotational motion with $N_\xi = 1$, i.e. constant first time derivative of ξ . Here, object structure is assumed to be known, and is removed from the state vector. The origin of the object-centred frame, which lies at the block centroid, remains fixed at $[30, 0, 0]_E^T$ units, with rotational motion about an axis parallel to z_E defined by $\xi(0) = [0, 0, 0]_E^T$ and $\dot{\xi}(0) = [0, 0, 1]_E^T$ radians/s. Note that in this case the reset is exact since the axis of the rotation matrix, $\tilde{\xi}$, is constant.

Figure 2.5: CRLB's of orientation vector with and without resetting

The top two graphs of Figure 2.5 are shown without applying the reset of (2.47) while the bottom two graphs are shown with resetting. Without the reset, and for this trajectory, the CRLB of the x - and y -components of ξ increase after 2 seconds. The CRLB of the x - and y -components of $\dot{\xi}$ are approximately an order of magnitude larger than the z -component CRLB at the end of the trajectory. Because, for this trajectory, $\bar{\xi}_1 = \bar{\xi}_2 = 0$, the second and fourth terms in expression (D.10) for $\partial \mathbf{I}_E^O / \partial \xi_j$, with $j = 1, 2$ are zero. As $\gamma := \|\xi\|$ increases, $\partial \mathbf{I}_E^O / \partial \xi_j \rightarrow 0$ for $j = 1, 2$ which implies that the rotation matrix becomes insensitive to changes in $\bar{\xi}_1$ and $\bar{\xi}_2$ which, in turn, leads to increasing CRLB's. The bottom two graphs of Figure 2.5 show CRLB's for ξ and $\dot{\xi}$ with resetting applied as shown in (2.47). In this case, the bounds exhibit a decreasing trend over the trajectory with final values after 20 seconds of approximately 0.005 radians for components of ξ and 0.0005 radians/s for those of $\dot{\xi}$.

Figures 2.6 through 2.9 show Cramer Rao bounds for the three filters in the absence of prior information ($\mathbf{P}_0^{-1} = 0$) during a simple trajectory with constant axis of rotation. Measurements of positions of feature points p^3, p^5, p^6 , and p^7 (see Figure 2.4) are available

throughout the trajectory. The initialization parameter vector $\mathbf{d}_O$ is set to zero, point p^6 is chosen to lie at the origin of the object-centred reference frame, point p^7 lies on the x_O -axis, and point p^3 lies in the x_O - y_O plane. All three coordinates of p^5 are unknown. True motion of the object is defined by constant acceleration translational motion of point p^6 ,

$$\begin{aligned}\mathbf{p}(0) &= [30; -20; 0]_E^T \text{ units,} \\ \dot{\mathbf{p}}(0) &= [2; 2; -2]_E^T \text{ units/s, and} \\ \ddot{\mathbf{p}} &= [-0.4; 0.4; 0.4]_E^T \text{ units/s}^2,\end{aligned}\tag{2.112}$$

and rotational motion about the y_O -axis with constant angular acceleration,

$$\begin{aligned}\text{Angle-Axis -} \\ \xi(0) &= \begin{bmatrix} 0 \\ -\pi/4 \\ 0 \end{bmatrix}, \quad \dot{\xi}(0) = \begin{bmatrix} 0 \\ 0.8\pi \\ 0 \end{bmatrix}, \quad \ddot{\xi} = \begin{bmatrix} 0 \\ -0.16\pi \\ 0 \end{bmatrix}.\end{aligned}\tag{2.113}$$

$$\begin{aligned}\text{Quaternion -} \\ \mathbf{q}(0) &= \begin{bmatrix} 0 \\ \sqrt{2} \\ 0 \\ \sqrt{2} \end{bmatrix}, \quad \boldsymbol{\omega}(0) = \begin{bmatrix} 0 \\ 0.8\pi \\ 0 \end{bmatrix}, \quad \dot{\boldsymbol{\omega}} = \begin{bmatrix} 0 \\ -0.16\pi \\ 0 \end{bmatrix}.\end{aligned}\tag{2.114}$$

$$\begin{aligned}\text{Roll-Pitch-Yaw } (\zeta = [\phi, \theta, \psi]^T) \\ \zeta(0) = \begin{bmatrix} 0 \\ -\pi/4 \\ 0 \end{bmatrix}, \quad \dot{\zeta}(0) = \begin{bmatrix} 0 \\ 0.8\pi \\ 0 \end{bmatrix}, \quad \ddot{\zeta} = \begin{bmatrix} 0 \\ -0.16\pi \\ 0 \end{bmatrix}.\end{aligned}\tag{2.115}$$

The angle-axis filter is reset twice during this trajectory at about 2.1 and 8.2 seconds, and pitch in the roll-pitch-yaw filter passes through the singular point of $\theta = \pi/2$ at about 1 and 9 seconds. True structure parameters in this case (measured in units) are

$$\mathbf{r}_O^5 = \begin{bmatrix} 0 \\ 0 \\ -10 \end{bmatrix}, \quad \mathbf{r}_O^3 = \begin{bmatrix} 0 \\ 20 \\ 0^* \end{bmatrix}, \quad \mathbf{r}_O^1 = \begin{bmatrix} 10 \\ 0^* \\ 0^* \end{bmatrix},\tag{2.116}$$

where $(\cdot)^*$ indicates a known value due to the structural constraints.

Cramer Rao bounds for translational motion estimation, Figure 2.6 and top of Figure 2.7, demonstrate, as might be expected due to identical models for translational motion, that bounds on all three filters are very close for this trajectory, especially for the angle-axis and roll-pitch-yaw filters. Bounds for the quaternion filter lie slightly above those of the other two parameterizations over most of the trajectory. These slightly higher bounds are attributed to the extra degree of freedom in the quaternion filter as previously discussed in Section 2.4.4. Bounds on orientation parameters, shown in the lower part of Figure 2.7, exhibit oscillatory behavior, but a decreasing trend over the trajectory. These oscillations appear to contradict the notion that as more measurements become available, more information should be accumulated about the observed process, which means that one might expect the bounds to decrease monotonically. Appendix C clarifies the meaning of these bounds in the present context and demonstrates why they need not decrease monotonically.

Figure 2.8 compares covariance bounds for estimation of angular velocity and acceleration. Again very similar behavior is demonstrated for the angle-axis and roll-pitch-yaw filters, whereas bounds for the quaternion filter are slightly higher. Figure 2.9 shows the bounds for estimation of the six structural parameters. Again, almost identical behavior is shown for the angle-axis and roll-pitch-yaw filters. However, significantly higher bounds are present for the quaternion filter in the estimation of the z_O -coordinate of p^5 , the y_O -coordinate of p^3 , and the x_O -coordinate of p^7 . Note that these are precisely the non-zero parameters in the structure set shown in (2.116). This investigation was repeated for different combinations of structural parameters and always yielded similar results in that quaternion filter bounds for non-zero structural parameters were significantly above those for the other two filters.

Recall that $\hat{\mathbf{r}}_O^i(k|k)$ denotes structural estimates internal to the quaternion filter whereas ${}^*\hat{\mathbf{r}}_O^i(k|k)$ denotes structure estimates taken as output from the quaternion filter. Intuitively, the accuracy of nonzero structural components depends not only on the internal structure representation, but also on the norm of the quaternion estimate according to

$${}^*\hat{\mathbf{r}}_O^i(k|k) := \|\mathbf{q}(k|k)\|^2 \hat{\mathbf{r}}_O^i(k|k). \quad (2.117)$$

Those structure components which are zero must have a zero internal representation within the filter. Estimation of nonzero structural components, however, is subject to an extra de-

gree of freedom in the quaternion filter due to the inability to enforce unit norm on the estimated quaternion. The Cramer-Rao bounds result from complex calculations involving the Jacobian matrices of dynamical and measurement models (as shown in Appendix C). In addition, there is no mechanism available to incorporate this extraction of structure estimates scaled by the square of the quaternion norm into the calculation of the covariance bounds, which makes an exact analysis of these observations very tedious.

Figure 2.6: Cramer Rao bound comparison of position and velocity.

Figure 2.7: Cramer Rao bound comparison of acceleration and orientation.

Figure 2.8: Cramer Rao bound comparison of rotational motion.

Figure 2.9: Cramer Rao bound comparison of structure estimation.

2.6.2 Simulation Results

Simulation studies have been conducted over a wide range of trajectories, beginning with known structure and simple motion, and progressing to completely unknown structure with more complex motion. Various sets of parameters were employed for dynamical models in the filters during these simulations. A simple process for selecting physically justifiable parameters led to the best results.

Because measurements are related directly only to object position and orientation, estimation of high-order time derivatives often demonstrated serious lag and overshoot which can lead to filter instability. This was particularly true in the estimation of motion parameters whose true temporal behavior contained strong sinusoidal components which represent unmodelled dynamics. Consequently, motion only up to the second time derivative is included in dynamic models by setting $N_t = 2$, $N_\xi = 2$, $N_w = 1$, and $N_\zeta = 2$. Following Singer's approach [60] in Equation (2.35) of Section 2.4.3, the manoeuvre correlation time constant is estimated for the expected class of objects and motion, extreme values are estimated for the manoeuvre variables of translational and rotational motion, and probabilities are assigned to the occurrence of maximum, minimum and zero values of the manoeuvre variable. Results shown in this Section were generated assuming a manoeuvre correlation time constant of 2 seconds which gives $\alpha_t = \alpha_r = 0.5\text{s}^{-1}$. Extreme values of acceleration for translational motion is estimated as $M_{t\max} = 12 \text{ units/s}^2$ which occur with probability $P_{t\max} = 0.1$, and zero translational acceleration occurs with probability $P_{t_0} = 0.1$. These values in Equation (2.35) give

$$\sigma_t = \left\{ \frac{M_{t\max}^2}{3} [1 + 4P_{t\max} - P_{t_0}] \right\}^{\frac{1}{2}} \approx 8. \quad (2.118)$$

The continuous-time white noise process driving the translational motion model then has covariance $\mathbf{Q}_t = 2\alpha_t\sigma_t^2\mathbf{I}_3$, and the approximation of (2.34) is employed to compute the covariance matrix of the corresponding discrete-time process noise sequence. Corresponding parameters for rotational motion are $M_{r\max} = 6\text{rad/s}^2$, $P_{r\max} = 0.1$, and $P_{r_0} = 0.1$, which gives $\sigma_r \approx 4$ and $\mathbf{Q}_r = 2\alpha_r\sigma_r^2\mathbf{I}_3$. Integration of the quaternion differential equation with $N_\pi = 5$ was found to give sufficient accuracy.

Simulation results from two cases are presented:

Case 1 True object motion is defined with constant acceleration of the block centroid, and rotational motion computed from an angle-axis formulation which cannot be exactly matched by any of the three parameterizations of rotational motion employed in the filters. The origin of the object-centred reference frame does not, due to the initialization algorithm described in Section 2.5, lie at the block centroid, and hence translational motion models in each filter are not matched to the trajectory. Simulation results for this case are shown in Figures 2.10 to 2.13. The angle-axis filter gives better overall performance for this trajectory than the quaternion and roll-pitch-yaw filters: the roll-pitch-yaw filter demonstrates poorer performance in estimation of angular velocity, and the quaternion filter exhibits poorer performance in the estimation of position and some structural components. However, all three filters give poor performance in velocity estimation due to strong sinusoidal components, which result from rotational motion, in the trajectory of the origin of the object-centred reference frame.

Case 2 True object motion is defined with constant acceleration of the block centroid, as in Case 1, but rotational motion is computed from a sinusoidal roll-pitch-yaw formulation. As in Case 1, neither translational nor rotational dynamic models are matched to the trajectory. Simulation results are shown in Figures 2.14 and 2.15. The angle-axis and roll-pitch-yaw filters show almost identical performance in this case. Again the quaternion filter exhibits poorer performance in the estimation of position and some structural components. Tabulated statistical results demonstrate superior performance of the angle-axis filter over the quaternion filter at the 95% confidence level for most of the trajectory.

Results from these two Cases are representative of all results obtained from various trajectories investigated during simulation studies. In general, the angle-axis filter gave better performance than the quaternion and roll-pitch-yaw filters. The quaternion filter, apart from its higher computational requirements, generally gave poorer performance in tracking position and estimation of structure. The roll-pitch-yaw filter, which is the simplest from a computational point of view, sometimes showed significant error in angular velocity estimates. The angle-axis filter is only slightly more computationally demanding than

the roll-pitch-yaw filter. All three filters, however, showed poor tracking performance for states with strong sinusoidal components.

With four feature points of interest, the angle-axis and roll-pitch-yaw filters have 24 states, while the quaternion filter has 25 states. Presentation of results for each Case includes a discussion of the true object motion, initialization parameters, true structure (which depends on the initialization parameter vector $\mathbf{d}_O$), and the resultant image sequences. Graphical presentation includes results of 60-run Monte Carlo simulations. In order to duplicate true motion and structure over Monte Carlo simulations, initialization during the first run selects $\mathbf{d}_O$ (the same for all three filters), and subsequent runs employ the same $\mathbf{d}_O$ as the first, but with other parameters being initialized as outlined in Section 2.5.

A convenient measure of orientation estimation accuracy at the k th time step is given by the scaled Frobenius norm

$$C(k) := \frac{1}{2\sqrt{3}} \left(\text{Tr} \{ \hat{\mathbf{I}}_E^O(k) (\hat{\mathbf{I}}_E^O(k))^T \} \right)^{\frac{1}{2}}. \quad (2.119)$$

The Frobenius matrix norm without the normalization factor $1/(2\sqrt{3})$ has been employed in investigations such as [59]. Because $\mathbf{I}_E^O(k)$ and $\hat{\mathbf{I}}_E^O(k)$ are both orthogonal for all three parameterizations, the scale factor employed in (2.119) implies that the smallest interval containing $C(k)$ is $[0, 1]$. Angular velocity is available directly in the quaternion filter, but is computed from state estimates using Equations (2.42) and (2.90) for the angle-axis and roll-pitch-yaw filters, respectively. Monte Carlo simulation results show root-mean-square values of the Euclidean norm of position, velocity, angular velocity, and structure error vectors.

Case 1: True object motion, in this case, is defined with constant acceleration of the block centroid, and rotational motion computed from an angle-axis formulation. For the purpose of describing the true motion of the block, define F_B as the block-centred reference frame used in the imagery generation program. Note that F_B and F_O are generally distinct, but for the purpose of simulation differ only by a translation (so that $\mathbf{I}_B^O \equiv \mathbf{I}_3$). The origin, O_B , of F_B lies at the block centroid, and the coordinate planes of F_B are parallel to faces of the block. Translational motion of O_B was defined in the imagery generation program with

$$[\mathbf{T}_B^E]_E(0) = [30, -20, 0]_E^T \text{ units.}$$

$$\begin{aligned} [\dot{\mathbf{T}}_B^E]_E(0) &= [2, 2, -2]_E^T \text{ units/s, and} \\ [\ddot{\mathbf{T}}_B^E]_E &= [-0.4, 0.4, 0.4]_E^T \text{ units/s}^2. \end{aligned} \quad (2.120)$$

Recall that the initialization routine computes $\mathbf{d}_O$, which defines the position of the origin of the object centred frame, based on observations during the first measurement event. The origin of the object-centred frame employed in the filters, therefore, will not generally coincide with the centroid of the block and, due to rotational motion, will not have the constant acceleration motion defined in (2.120). Specifically, the filters are required to estimate translational motion of F_O whose position and velocity, for example, are given by

$$\mathbf{p}(t) := [\mathbf{T}_O^E]_E(t) = [\mathbf{T}_B^E]_E(t) + \mathbf{I}_E^O \mathbf{I}_O^B \{ \mathbf{r}_B^{i_0} - [\mathbf{d}_O]_B \}, \quad (2.121)$$

$$\dot{\mathbf{p}}(t) = [\dot{\mathbf{T}}_B^E]_E + [\boldsymbol{\omega}_E^O]_E^* \mathbf{I}_E^O \mathbf{I}_O^B \{ \mathbf{r}_B^{i_0} - [\mathbf{d}_O]_B \}, \quad (2.122)$$

where the velocity equation uses (2.37) and the fact that $\mathbf{I}_B^O$ is constant. Higher-order time derivatives of true translational motion are computed in a straight-forward manner by differentiating (2.122).

Rotational motion of the block-centred reference frame is defined in the imagery generation program with an angle-axis parameterization,

$$\mathbf{I}_E^B(t) = \exp(\eta(t) \bar{\mathbf{u}}^z(t)). \quad (2.123)$$

The angle of rotation, $\eta(t)$, has constant second time derivative,

$$\eta(0) = 0 \text{ rad}, \quad \dot{\eta}(0) = 0.4\pi \text{ rad/s, and } \ddot{\eta} = -0.08\pi \text{ rad/s}^2. \quad (2.124)$$

Temporal behavior of $\bar{\mathbf{u}}$, since it is a unit vector, must be rotational with instantaneous angular velocity $\boldsymbol{\omega}_u(t)$ in the earth-fixed frame so that

$$\dot{\bar{\mathbf{u}}}(t) = \boldsymbol{\omega}_u^z(t) \bar{\mathbf{u}}, \quad (2.125)$$

and the initial condition is specified as

$$\bar{\mathbf{u}}(0) = \frac{1}{\sqrt{3}} [1, 1, 1]^T. \quad (2.126)$$

The angular velocity, ω_u , of the axis of the basis transformation (this is not the angular velocity of the object) is defined with constant second time derivative,

$$\omega_u(0) = \begin{bmatrix} -0.2 \\ -0.2 \\ 0.2 \end{bmatrix} \text{ rad}, \quad \dot{\omega}_u(0) = \begin{bmatrix} -0.1 \\ -0.1 \\ 0.1 \end{bmatrix} \text{ rad/s}, \quad \ddot{\omega}_u = \begin{bmatrix} 0.02 \\ 0.02 \\ -0.02 \end{bmatrix} \text{ rad/s}^2. \quad (2.127)$$

Integration of (2.125) is performed numerically using a method similar in form to the quaternion integration scheme described in Section 2.4.4 by defining a partition of each sample period, assuming ω_u is approximately constant over each interval of the partition, and using the closed form expression for the matrix exponential of a 3×3 skew symmetric matrix given in (2.41). As in the quaternion integration scheme, this method ensures that $\bar{u}(t)$ remains a unit vector.

Table 2.1 summarizes simulation parameters for this Case and provides numerical values of a sample initialization denoted by $\hat{x}_{s,t,r}(0)$. The initial filter covariance matrices are taken as diagonal with the square-roots of diagonal entries corresponding to standard deviations, as indicated in Table 2.1, of 5 units for structure states, 5 units for position and its time derivatives, 0.5 for orientation, and 2.0 Rad/s for time derivatives of orientation parameters.

A subsequence of images viewed in the left camera is shown in Figure 2.10. In this Figure, each frame sample number is shown in the lower right: the sequence proceeds across the top row from left to right, back across the second row from right to left, etc. The top of the object first begins to rotate towards the observer as does the right side of the object. The object eventually returns to its original orientation at the end of the trajectory. True and sample noisy feature point trajectories in the image plane of the left camera are shown in Figure 2.11.

Figure 2.12 shows root-mean-square errors from a 60-run Monte Carlo simulation. Very similar performance is demonstrated by all three filters, except for slightly degraded estimation of position and p^5 structure from the quaternion filter. The two distinct peaks in position RMSE at about 3 seconds and 6 seconds occur at that time when the origin of the object-centred frame lies farthest from the cameras. Estimation of object position from stereo cameras is well-known to be degraded as the object moves away from the cameras

Table 2.1: Simulation parameters for Cases 1 and 2. Table entries replaced by 'NC' indicate that quantities have not been computed and are not used in initialization.

Structure: $N_f = 4, \sigma(\tilde{\mathbf{x}}_s) = 5.0, \mathbf{d}_0 = [-3.29, -4.74, -3.06]^T$				
	$\mathbf{r}_O^5$	$\mathbf{r}_O^3$	$\mathbf{r}_O^7$	
$\mathbf{x}_s$	$[-3.28, -4.74, 6.94]$	$[-3.29, 15.26, d_{O,3}]$	$[6.71, d_{O,2}, d_{O,3}]$	
$\tilde{\mathbf{x}}_s(0)$	$[-3.12, -4.80, 6.58]$	$[-4.37, 16.05, d_{O,3}]$	$[9.94, d_{O,2}, d_{O,3}]$	
Translation: $N_t = 2, \alpha_t = 0.5, \sigma_t = 8$				
	$\mathbf{p}$	$\dot{\mathbf{p}}$	$\ddot{\mathbf{p}}$	
$\mathbf{x}_t(0)$	$[28.29, -25.26, -1.94]$	NC	NC	
$\dot{\mathbf{x}}_t(0)$	$[28.39, -25.70, -2.05]$	$[0.0, 0]$	$[0.0, 0]$	
$\sigma(\tilde{\mathbf{x}}_t)$	5.0	5.0	5.0	
Rotation: Angle-Axis, $N_\xi = 2, \alpha_r = 0.5, \sigma_r = 4$				
	ξ	$\dot{\xi}$	$\ddot{\xi}$	
$\mathbf{x}_r(0)$	$[0.0, 0]$	NC	NC	
$\dot{\mathbf{x}}_r(0)$	$[-0.048, -0.047, -0.040]$	$[0.0, 0]$	$[0.0, 0]$	
$\sigma(\tilde{\mathbf{x}}_r)$	0.5	2.0	2.0	
Rotation: Quaternion, $N_u = 1, N_\pi = 5, \alpha_r = 0.5, \sigma_r = 4$				
	$\mathbf{q}$	ω	$\dot{\omega}$	
$\mathbf{x}_r(0)$	$[0.0, 0, 1]$	NC	NC	
$\dot{\mathbf{x}}_r(0)$	$[0.024, 0.023, 0.02, 0.999]$	$[0.0, 0]$	$[0.0, 0]$	
$\sigma(\tilde{\mathbf{x}}_r)$	0.5	2.0	2.0	
Rotation: Roll-Pitch-Yaw, $N_\zeta = 2, \alpha_r = 0.5, \sigma_r = 4$				
$\zeta = [\phi, \theta, \psi]$	ζ	$\dot{\zeta}$	$\ddot{\zeta}$	
$\mathbf{x}_r(0)$	$[0.0, 0]$	NC	NC	
$\dot{\mathbf{x}}_r(0)$	$[-0.047, 0.048, -0.039]$	$[0.0, 0]$	$[0.0, 0]$	
$\sigma(\tilde{\mathbf{x}}_r)$	0.5	2.0	2.0	

which is the primary reason for peak RMS errors at these times. Note the significant oscillations in translational velocity estimation due to lag and some overshoot in filter estimates of sinusoidal components. The amplitude of oscillation appears to decrease during the second half of the trajectory. Angular velocity estimates from the roll-pitch-yaw filter shows distinct peaks in root-mean-squared errors at about 2.5, 4.5, and 6 seconds. The true pitch angle (not shown) for this trajectory behaves like a damped sinusoid with peaks just prior to these times, and has a maximum value of about 0.4π . Although estimation of pitch angle was very accurate, estimation of the time derivative of pitch, which is also a damped sinusoid, shows some lag and overshoot much like those occurring in translational velocity estimation. This behavior would contribute significantly to the peaks present in angular velocity estimation error. The angle-axis filter gives marginally better performance than the other two filters for this simulation. Structure root-mean-square estimation errors decrease rapidly during the first 1 or 2 seconds, and appear to have levelled off at about 0.3 to 0.5 units at the end of the trajectory.

As discussed in Section 2.4.1, there are four possible orientations for the object-centred reference frame with respect to the three special feature points selected at initialization. When structure was initialized with all zeros, for example, convergence of structure estimates was observed in one of the four possible orientations of the object-centred frame, the particular one of the four being dependent on initial filter transients. In simulations with this object and the given measurement noise levels it was found that the simple single-frame initialization of Section 2.5 provided sufficient initial accuracy for structure to converge in the object-centred frame selected at initialization, although obviously this need not always be the case. Figure 2.13 shows the reliable convergence of structure estimates to true structural parameters given in Table 2.1 after 10 seconds (100 measurement events) for each of the Monte-Carlo simulation runs. Again all three filters give similar results, with final estimates from the quaternion filter being slightly noisier than those from the angle-axis and roll-pitch-yaw filters.

Figure 2.10: Image subsequence from the left camera for Case 1. Image frame numbers are shown in the lower left corner of each image. Four feature points, p^1 , p^3 , p^5 , and p^7 , are labelled in each image.

Figure 2.11: True and sample noisy image feature point trajectories in the left camera without occlusion for Case 1.

Figure 2.12: Monte Carlo simulation results for Case 1.

Figure 2.13: Final structure estimates in Monte Carlo simulations for Case 1.

Case 2: This trajectory defines true translational motion with constant acceleration of the block centroid, as in Case 1, but rotational motion is computed from a sinusoidal roll-pitch-yaw formulation. Again, let F_B denote the block-centred reference frame used in the imagery generation program with $\mathbf{I}_B^O \equiv \mathbf{I}_3$. The origin, O_B , of F_B lies at the block centroid, and the coordinate planes of F_B are parallel to faces of the block.

As in Case 1, translational motion of O_B was defined with

$$\begin{aligned} [\mathbf{T}_B^E](0) &= [30, -20, 0]_E^T \text{ units,} \\ [\dot{\mathbf{T}}_B^E](0) &= [2, 2, -2]_E^T \text{ units/s, and} \\ [\ddot{\mathbf{T}}_B^E]_E &= [-0.4, 0.4, 0.4]_E^T \text{ units/s}^2. \end{aligned} \quad (2.128)$$

However, translational motion of the object-centred frame is significantly different from that of Case 1 due to alternate rotational motion of F_B about O_B which is defined with sinusoidal variation of roll and pitch angles and linear yaw angle with time:

$$\begin{aligned} \phi(t) &= a \sin(bt) \\ \theta(t) &= a[1 - \cos(bt)] \end{aligned} \quad (2.129)$$

$$\psi(t) = bt. \quad (2.130)$$

where

$$a = \frac{\pi}{6} \text{ and } b = \frac{2\pi}{10}. \quad (2.131)$$

Initialization parameters in this Case are identical to those for Case 1 shown in Table 2.1. Figure 2.14 shows a subsequence of images from the left camera for this trajectory. The top of the object containing feature points 3 and 7 begins to rotate towards the observer (increasing pitch angle), while the right side containing points 1 and 3 appears to move down in the image due to positive yaw and roll angles. Again, the object returns to its original orientation at the end of the trajectory.

Figure 2.15 shows Monte Carlo simulation results for this trajectory. As in Case 1, the quaternion filter gives poorer performance in estimation of position and some structural components. Velocity root-mean-square errors lie below 2 units/s during the second half of the trajectory and do not show significant oscillations that were present in results of Case 1. Orientation errors lie below 5% after about 2 seconds, with final errors of about 2.5% for

Table 2.2: Time interval test of means for comparison of angle-axis and quaternion algorithms. Test values greater than 1.65 infer superior performance of the angle-axis algorithm over the quaternion algorithm at the 95% confidence level for the indicated time interval of the trajectory of Case 2.

TIME INT.	0-20	20-40	40-60	60-80	80-100
POSITION	4.90	5.13	4.60	4.62	4.21
ORIENTATION	1.40	0.63	0.42	3.62	6.12
ANG. VEL.	5.53	12.57	8.35	15.07	0.70
P5 STRUCTURE	0.54	3.51	6.72	6.66	8.01

all three filters, which is very similar to results of Case 1. Angular velocity errors decrease rapidly during the first 1 second and remain below about 3 rad/s for the remainder of the trajectory. Final root-mean-square errors for structure estimation are on the order of 0.5 to 1 unit.

Bar-Shalom and Fortmann [71] describe in detail a procedure for statistical comparison of algorithm performance that can be applied assuming that estimation errors summed at each time update over the 60 runs exhibit approximately normal statistics. Using this method, Table 2.2 shows the test statistics for which values greater than 1.65 infer superior performance of the angle-axis algorithm over the quaternion algorithm at the 95% confidence level for the indicated time interval of this trajectory. Statistically significant improvements of the angle-axis approach over the quaternion approach are observed in position over all time intervals, in orientation from 60 seconds to the end of the run, in angular velocity from initialization to 80 seconds, and in *P5*-structure from 20 seconds to the end of the run.

Figure 2.14: Image subsequence from the left camera for Case 2. Image frame numbers are shown in the lower left corner of each image. Four feature points, p^1 , p^3 , p^5 , and p^7 , are labelled in each image.

Figure 2.15: Monte Carlo simulation results for Case 2.

2.7 Chapter Summary

This chapter has introduced structure, translational motion, rotational motion, and measurement state space models for manoeuvring objects observed with a multiple-camera imaging system. Both translational and rotational motion are described with first-order Gauss-Markov processes which are completely defined by selection of a small number of parameters: the number of time derivatives (N_t and one of N_x , N_w , or N_z); the reciprocal manoeuvre correlation time constants (α_t and α_r); and the manoeuvre variances (σ_t^2 and σ_r^2). The quaternion filter also requires selection of the number of intervals, N_π , of each sample period for time propagation of the state estimate. A simple single-frame initialization algorithm has been proposed for object position, orientation, and structure.

This Chapter also presented approximate Cramer Rao bounds for a simple trajectory, discussed parameter selection for the dynamical models, and presented a reduced set of simulation results which are representative of performance achieved for a wide range of trajectories. Cramer Rao bounds suggest very similar optimal performance for the three filters over a simple trajectory, and this is supported by the Monte Carlo simulation results for more complex trajectories. One significant advantage of employing dynamic models developed in this chapter is that model parameters can be selected as proposed by Singer [60] based to a large extent on the expected class of objects or motion, rather than purely through heuristic methods.

All three filters demonstrated good performance in tracking position and orientation of manoeuvring objects, however, estimation of translational velocity can be degraded due to filter lag, as in the trajectory of Case 1. Lag in the filters could be reduced to a certain extent by increasing the strength of filter gains (by increasing σ_t), but this may be in conflict with required measurement noise suppression. Angular velocity root-mean-squared errors for Case 1 decrease very rapidly during the first 10 measurement events, and thereafter appear to remain at levels of about 0.3 rad/s (≈ 17 Deg/s). In Case 1, angular velocity estimates from the roll-pitch-yaw filter showed distinct peaks in root-mean-squared errors which are due in part to some lag and overshoot in estimation of time derivatives of a sinusoidal pitch angle. Similar behavior, only on a smaller scale, is seen in Monte Carlo simulation results for Case 2 (see Figure 2.15) with a slight increase in angular velocity root-mean-squared

errors just prior to 5 seconds. Generally, performance of the extended Kalman filters was degraded in the estimation of time derivatives of sinusoidally varying states.

Structure estimation for all three filters showed a sharp decrease in root-mean-squared errors during the first 1 or 2 seconds, with final accuracy after 100 observations of about 0.25 to 0.5 units. Structure estimates for the special feature point (p^7 for these simulations) for which the filters must estimate only a single component showed rapid convergence to within about 0.5 units after only 10 to 20 measurement events, while those for the feature point (p^3 here) in which two components must be estimated required about 50 measurement events.

The primary focus of this chapter was to first cast the general feature-based motion analysis problem into one of nonlinear state estimation. The use of extended Kalman filtering in the estimation of structure and six-degree-of-freedom motion of an object from noisy measurements of feature positions in multiple-camera image sequences was investigated. This work not only demonstrates how previously proposed quaternion-based approaches might be extended to track manoeuvring objects observed in multiple-camera image sequences, but also provides two much simpler alternatives using roll-pitch-yaw and angle axis parameterizations of rotational motion. The angle axis filter based on an approximate linear dynamical model gives performance for all trajectories considered that is at least as good as, and often superior to, the quaternion-based scheme. Simulation results have demonstrated, however, that the extended Kalman filter can result in significant lag in state estimates particularly for those states with high bandwidth trajectories. These observations must be considered together with the following two additional facts: 1) there is no theoretical guarantee of convergence or stability for the extended Kalman filter in the presence of modeling errors; and 2) the extended Kalman filter can be extremely sensitive to dynamical modelling errors. This latter fact is even more pronounced in the generic object tracking problem for which the dynamical model is completely unknown.

Chapter 3

State Estimation for Nonlinear Systems

3.1 Introduction

Chapter 1 included a brief description of the observer problem as one in which the objective is to estimate the full state of a system based on available measurements. The Kalman filtering method, which was the focus of Chapter 2, is one of many proposed approaches for the state estimation problem. Section 3.2 gives a more precise statement of the observer problem considered in this thesis and introduces notation which is common to subsequent chapters. In order for any observer to estimate or reconstruct the state of a system, the dynamical and measurement models, together, must provide sufficient information for the full state to be extracted uniquely from the set of past measurements. Whether or not this is mathematically possible is known as the *observability* of the overall system, and Section 3.3 discusses current approaches to the nonlinear observability problem for discrete-time systems. Section 3.4 presents a brief account of current methods of state estimation for nonlinear systems through a review of the literature. Finally, Section 3.5 reviews fundamental theory of variable structure systems with sliding modes, including current applications and analysis of these systems in control and observation problems.

3.2 Definition of the Observer Problem

The general system under consideration is assumed to be of order n . The system provides m measurements in each sample period. These m measurements depend, at each measurement

event, only on q of the n states with $q \leq n$. For example a system which is modelled with states position, velocity, and acceleration ($n = 3$) may only provide one measurement ($m = 1$) which depends only on position ($q = 1$). In Chapter 6 notation is introduced to single out those states upon which measurements are dependent at each sample time. The present discussion continues with standard notation which is sufficient for presentation of fundamental concepts.

The full state vector is $\mathbf{x}(k) \in \mathbf{X} \subset \mathbb{R}^n$, where $\mathbf{X}$ is a known set. The system dynamics are written as

$$\mathbf{x}(k+1) = \mathbf{f}(k, \mathbf{x}(k)), \quad (3.1)$$

where $\mathbf{f} : \mathbb{N}_0 \times \mathbf{X} \mapsto \mathbf{X}$ and $\mathbb{N}_0 := \{0, 1, 2, \dots\}$. The measurement vector, $\mathbf{z}(k) \in \mathbb{R}^m$, for any time k , is related to the state vector according to the measurement model

$$\mathbf{z}(k) = \mathbf{h}(\mathbf{x}(k)). \quad (3.2)$$

In this thesis, the proposed observers all have the general form

$$\hat{\mathbf{x}}(k+1) = \hat{\mathbf{f}}(k, \hat{\mathbf{x}}(k)) + \mathbf{L}(k)\mathbf{v}(k), \quad (3.3)$$

where $\hat{\mathbf{f}}(k, \cdot)$ is an estimate of the true dynamics, $\mathbf{f}(k, \cdot)$, with $\hat{\mathbf{f}}$ being C^p in $\mathbf{x}$, $p \geq 1$, and C^p denotes the space of p -times continuously differentiable functions. The matrix $\mathbf{L}(k)$ represents observer gains to be determined, and $\mathbf{v}(k)$ depends on system measurements, the observer state $\hat{\mathbf{x}}$, and several user-defined parameters. The dimensions of $\mathbf{L}$ and $\mathbf{v}$ vary among the proposed methods and are discussed in appropriate sections of the thesis.

The observer problem is generally concerned with design of a dynamical system (3.3), such that upon initialization of the observer at time $k = 0$ with an estimate, $\hat{\mathbf{x}}(0)$, of the true state, $\mathbf{x}(0)$, the output of the observer, $\hat{\mathbf{x}}(k)$ for $k = 1, 2, \dots$, provides a “reasonable estimate” of the true state in terms of the properties of the estimation error,

$$\mathbf{e}(k) := \mathbf{x}(k) - \hat{\mathbf{x}}(k). \quad (3.4)$$

The error dynamics are obtained from (3.1) and (3.3) as

$$\begin{aligned} \mathbf{e}(0) &:= \mathbf{x}(0) - \hat{\mathbf{x}}(0), \\ \mathbf{e}(k+1) &= \mathbf{f}(k, \mathbf{x}(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}(k)) - \mathbf{L}(k)\mathbf{v}(k). \end{aligned} \quad (3.5)$$

Mathematical characterizations of what constitutes a “reasonable estimate” take on many forms depending on the theory or applications of interest. In statistical-based observers, such as the Kalman filter, the objective is to obtain unbiased, minimum variance estimation errors, while for stochastic Lyapunov approaches the objective is to obtain unbiased estimation asymptotically. In this thesis, particularly in Chapter 5, emphasis is placed on specification of sufficient conditions under which the observer error is exponentially bounded, including the specification of admissible initialization errors. If, in addition, modelling errors satisfy certain *matching conditions*, which are introduced in Section 3.5, then the observer error is exponentially asymptotically stable. The primary motivation for this approach is that exponential stability not only possesses the fundamental stability requirements, but also provides rates of convergence which are not available, generally, from conditions which guarantee only stability or asymptotic stability.

This section has introduced notation and defined the problem which is to be solved. The following section considers the question of whether or not it is mathematically possible to recover the state under ideal circumstances (no disturbances or modelling errors) from the set of past measurements.

3.3 Nonlinear Observability

The ideas of controllability and observability for linear, time invariant systems have been well understood since the early 1960's [2, 3, 6]. However, it was not until the middle 1970's that similar concepts for nonlinear possibly time varying systems were effectively treated [72, 73]. Hermann and Krener [73] formulate for the nonlinear continuous-time problem concepts of controllability and observability which are analog to those for linear systems through the use of techniques in differential geometry and the Lie Algebra. Essentially, nonlinear controllability and observability is examined by checking the rank of controllability and observability matrices formed by higher-order Lie derivatives and Lie brackets of the system. This work appears to be partially responsible for increased activity in analysis of control and observation of nonlinear systems through differential geometric methods. In observability analysis of discrete-time systems, functional composition replaces the Lie algebraic manipulations used for analysis of continuous-time systems.

In order to simplify the presentation for this section, consider an autonomous discrete-time system written as

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k)) \quad (3.6)$$

$$\mathbf{z}(k) = \mathbf{h}(\mathbf{x}(k)), \quad (3.7)$$

where $\mathbf{x} \in \mathbf{X} \subset \mathbb{R}^n$, $\mathbf{z} \in \mathbb{R}^m$, and $\mathbf{f}(\cdot)$ and $\mathbf{h}(\cdot)$ are smooth functions. With functional composition notation, $\mathbf{f} \circ \mathbf{f}(\mathbf{x}) := \mathbf{f}(\mathbf{f}(\mathbf{x}))$, define

$$\begin{aligned} \mathbf{f}^0(\mathbf{x}) &:= \mathbf{x} \\ \mathbf{f}^i(\mathbf{x}) &= \mathbf{f} \circ \mathbf{f}^{i-1}(\mathbf{x}), \quad i = 1, 2, 3, \dots \end{aligned} \quad (3.8)$$

It is emphasized that the restriction to autonomous systems is not essential and is only done to simplify the presentation of basic ideas and to facilitate the use of functional composition notation. The following development follows Hermann and Krener [73], Nijmeijer [74], Glad [75] and Grizzle and Moraal [76], although there appears to be some variation in terminology between these authors.

For each given initial state $\mathbf{x} \in \mathbf{X}$, the system (3.6) with measurements (3.7) produces an output sequence $\mathbf{O}(\mathbf{x})$ defined by the ordered set

$$\mathbf{O}(\mathbf{x}) := \{\mathbf{h}(\mathbf{x}), \mathbf{h} \circ \mathbf{f}(\mathbf{x}), \mathbf{h} \circ \mathbf{f}^2(\mathbf{x}), \dots, \mathbf{h} \circ \mathbf{f}^k(\mathbf{x}), \dots\}. \quad (3.9)$$

A pair of points, $\mathbf{x}$ and $\bar{\mathbf{x}}$, are said to be *indistinguishable* if their ordered output sequences coincide, i.e. $\mathbf{O}(\mathbf{x}) = \mathbf{O}(\bar{\mathbf{x}})$. For any point $\bar{\mathbf{x}} \in \mathbf{X}$, define the set of points which are indistinguishable from $\bar{\mathbf{x}}$ as $I(\bar{\mathbf{x}})$,

$$I(\bar{\mathbf{x}}) := \{\mathbf{x} \in \mathbf{X} \mid \mathbf{O}(\mathbf{x}) = \mathbf{O}(\bar{\mathbf{x}})\}. \quad (3.10)$$

The system is then said [73, 74] to be *observable at $\bar{\mathbf{x}}$* if $I(\bar{\mathbf{x}}) = \{\bar{\mathbf{x}}\}$ and the system is *observable* if $I(\bar{\mathbf{x}}) = \{\bar{\mathbf{x}}\}$ for every $\bar{\mathbf{x}} \in \mathbf{X}$. Notice that by definition, observability at a point becomes an infinite test since $\mathbf{O}(\cdot)$ is an infinite set. For linear systems, the Cayley-Hamilton theorem reduces this test to one of finite dimension [14].

Several stronger notions of observability for nonlinear systems have been introduced which reduce the above test to one of finite dimension. For some finite positive integer N ,

define the mapping $\mathcal{O}_N : \mathbf{X} \mapsto \mathbb{R}^{Nm}$, called the *observability map*, as

$$\mathcal{O}_N(\mathbf{x}) = \begin{bmatrix} \mathbf{h}(\mathbf{x}) \\ \mathbf{h} \circ \mathbf{f}(\mathbf{x}) \\ \mathbf{h} \circ \mathbf{f}^2(\mathbf{x}) \\ \vdots \\ \mathbf{h} \circ \mathbf{f}^{N-1}(\mathbf{x}) \end{bmatrix}. \quad (3.11)$$

The system is said to be [76] *uniformly observable* in $\mathbf{X}$ if there exists an N such that the mapping $\mathcal{O}_N(\mathbf{x})$ is an injection from $\mathbf{X}$ into $\mathbb{R}^{Nm}$. The terms *strongly observable* or *finite time observable* have also been used [74] when $\mathbf{X} \subset \mathbb{P}^n$ and $N = n$. The system is said [73, 74] to satisfy the *observability rank condition* at $\bar{\mathbf{x}} \in \mathbf{X}$ if the rank of the mapping $\mathcal{O}_N$ (i.e. the rank of its Jacobian matrix) at $\bar{\mathbf{x}}$ equals n , and the system satisfies the observability rank condition if this is true for every $\bar{\mathbf{x}} \in \mathbf{X}$. If the system satisfies the observability rank condition, then for any $\bar{\mathbf{x}} \in \mathbf{X}$, there exists an open neighborhood $\mathcal{U}$ of $\bar{\mathbf{x}}$ such that $\forall \mathbf{x} \in \mathcal{U}$, if $\mathcal{O}_N(\bar{\mathbf{x}}) = \mathcal{O}_N(\mathbf{x})$ then $\bar{\mathbf{x}} = \mathbf{x}$.

The general idea is that if the system is uniformly observable in $\mathbf{X}$, the system of equations

$$\begin{bmatrix} \mathbf{z}(k-N+1) \\ \mathbf{z}(k-N+2) \\ \vdots \\ \mathbf{z}(k) \end{bmatrix} = \mathcal{O}_N(\mathbf{x}(k-N+1)) \quad (3.12)$$

can, in theory, be uniquely solved for the state $\mathbf{x}(k-N+1)$, and the current state can be obtained from

$$\mathbf{x}(k) = \mathbf{f}^{N-1}(\mathbf{x}(k-N+1)). \quad (3.13)$$

The observability rank condition, however, only guarantees this result locally, which places further emphasis on local analysis for nonlinear systems.

The observability map defined in (3.11) plays a major role in methods which seek a transformation of the state which will bring the system into a special form known as *non-linear observer canonical form*. Several such methods are discussed in the literature review of the next section.

2

PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

3.4 Current Methods of State Estimation

Various methods, some of which have been reviewed and compared by Dorling and Zinober [15], Walcott *et al.* [77], and Misawa and Hendrick [78], have been proposed for the observer problem:

1. the Extended Kalman Filter and its variants in the form of Constant Gain Extended Kalman Filters and Statistically Linearized Filters [10, 78] ;
2. observers based directly on Lyapunov methods [79]-[84];
3. observers based on linearization by output injection, pseudolinearization, or transformation to nonlinear observer form [85]-[91];
4. observation by nonlinear map inversion [92, 93]; and
5. sliding mode observers [77, 78],[95]-[108].

All of these methods have been formulated primarily for the continuous-time observer problem; only a small number of publications are available which analyze the discrete-time problem: Nijmeijer [74] and Glad [75] treat discrete-time observability issues; Yaz and Azemi [84] and Tarn and Rasis [82] employ discrete-time Lyapunov theory; Lee and Nam [109] and Chung and Grizzle [110] examine transformations to nonlinear observer form for discrete-time systems; and Grizzle and Moraal [76] investigate nonlinear map inversion for observation of sampled data systems. Discrete-time sliding mode observers have not yet appeared in the literature, other than in the form of reduced-order Luenberger observers for linear systems [111]-[114].

For continuous-time systems, Lyapunov-based methods originally proposed by Thau [79], and further extended for the deterministic case by Kou *et al.* [80], Banks [81], and Xia and Gao [83], and for the stochastic case by Tarn and Rasis [82] and Yaz and Azemi [84], focus on sufficient conditions for observer existence and convergence. Strict conditions are placed on the system nonlinearities such as Lipschitz or linear growth constraints. Implications of convergence for nonlinear systems are, as is expected, usually local in extent, but also depend on accurate characterization of nonlinear dynamics which can lead to poor performance in the presence of modelling errors [78].

Linearization methods, first introduced by Bestle and Zeitz [85], Krener and Isidori [86], and Krener and Respondek [87], employ differential geometric methods to derive a system of coupled partial differential equations which define a nonlinear, diffeomorphic transformation of the state that will bring the nonlinear system into observer canonical form. This approach is applicable to continuous-time systems of the form

$$\begin{aligned}\dot{\mathbf{x}}_c(t) &= \mathbf{f}_c(t, \mathbf{x}_c(t)) \\ \mathbf{z}(t) &= \mathbf{h}(\mathbf{x}_c(t)),\end{aligned}\tag{3.14}$$

for which there exists, at least locally, a diffeomorphic transformation $\mathbf{y}_c(t) = \mathbf{T}(\mathbf{x}_c(t))$ which brings the original system (3.14) to one which is linear with nonlinear output injection:

$$\begin{aligned}\dot{\mathbf{y}}_c(t) &= \mathbf{A}\mathbf{y}_c(t) + \boldsymbol{\phi}(\mathbf{z}(t)) \\ \mathbf{z}(t) &= \mathbf{C}\mathbf{y}_c(t).\end{aligned}\tag{3.15}$$

Observer design is then immediate for this particular form and is constructed from a conventional Luenberger observer (from linear systems theory) plus the nonlinear term $\boldsymbol{\phi}(\mathbf{z}(t))$ which depends only on the measurements. Existence of the transformation is determined from invertibility of the Jacobian of the observability mapping which is formed from higher order Lie derivatives and Lie brackets of the measurement model [73, 77, 87, 88]. The major problem with this approach is the computation of the required transformation through integration of a coupled set of nonlinear partial differential equations. Moreover, existence of the transformation may only be guaranteed locally. In this case, methods of pseudo or extended linearization, proposed by Baumann and Rugh [115], may be appropriate in which linearization is performed about constant operating points and observation employs a type of gain scheduling technique. In principle, these methods require exact knowledge of the system since the design methodology requires successive manipulation of functions representing the dynamic and measurement models, i.e., by taking partial derivatives, and higher order Lie derivatives and Lie brackets of system functions. For discrete-time systems, Lee and Nam [109] obtain necessary and sufficient conditions for the existence of a transformation to discrete-time nonlinear observer canonical form, analog to (3.15), with methods which are similar to the continuous-time counterpart. However, successive manipulation

of system functions again requires essentially exact knowledge of the system. Moreover, Chung and Grizzle [110] have shown that the class of time-sampled systems for which the linearization methodology can be applied for an open set of sampling times is quite small; for example, when the state space is of dimension two, it consists of only those systems which are state equivalent to a linear system. The practical implication is that digital implementations of this methodology would have to be approximate.

The use of observers for nonlinear map inversion has recently been proposed by Nicosia *et al.* [92, 93] and is related to the above linearization methods in that the resultant observer problem can be expressed in nonlinear observer canonical form. Differential geometric methods are employed to derive a "direct extended map" which relates the output and its time derivatives to the state and its time derivatives. The inverse map, if it exists, can be found by solving a coupled set of nonlinear partial differential equations provided time derivatives of the output are available. Alternatively, the extended map can be placed in nonlinear observer canonical form for which an observer can be constructed. The suggested observer design methodology proposes selection of observer gains which result in a generalized Newton-type algorithm. However, just as in the above linearization techniques, successive manipulation of functions representing the dynamic and measurement models requires in principle exact knowledge of the system. Similar approaches for discrete-time systems were proposed by Grizzle and Moraal [76] who considered direct inversion of the observability map and hence the system of equations in (3.12). When direct inversion is not tractable, Newton's algorithm is proposed to iteratively obtain a solution for the system (3.12). These methods are discussed further in Chapter 7 where Newton-type local iterations are proposed and analyzed for implementation of sliding mode observers for systems with nonlinear measurement models.

Sliding mode observers appeared in early work by Hashimoto *et al.* [104, 105], Drakunov [107], and Golembc *et al.* [108] for linear systems, and have been extended to more general nonlinear systems by Walcott and Zak [100, 101], Slotine *et al.* [95] and Misawa [102]. The sliding mode approach potentially offers significant advantages over the previously described methods in terms of robustness in the presence of unmodelled dynamics. In this approach, an observer is constructed which establishes an attractive manifold or surface in state (estimation error) space. Error trajectories must be guaranteed

to reach this surface in finite time and thereafter remain on the surface and “slide” to the origin. Sliding mode observers have also recently appeared in applications to robotic manipulators [96, 97, 98]. All of these methods consider continuous-time systems with linear measurement models. The following Section provides a more detailed theoretical review of systems with sliding modes, their potential applications to controller and observer problems, robustness properties of the sliding mode, and possible extensions to discrete-time systems.

3.5 Fundamentals of Sliding Mode Theory

Over the last decade, significant effort has been devoted to studying sliding mode controllers, and more recently sliding mode observers. The sliding mode control problem, which is the subject of a recent survey by Hung, Gao and Hung [116], is, at the present time, far more developed than the observer problem. The primary advantage of sliding mode methods in control and observation of dynamical systems is their natural robustness to unmodelled dynamics provided sliding conditions are maintained. This section provides an introduction and review of fundamental concepts of variable structure systems in sliding mode. The focus initially is on the control problem, characterization of sufficient conditions for the existence of sliding modes, and a discussion of robustness properties of the sliding mode and problems arising from chattering. This is followed by a review of approaches to sliding mode observers for continuous-time systems, again with emphasis on robustness properties. In particular a dual nature of robustness properties for the controller and observer problems is illustrated and discussed. This Section concludes with a discussion of discrete-time sliding modes and a review of methodologies for analysis of discrete-time sliding mode control systems.

In variable structure control systems, the control process varies, or switches, from one structure to another. A sliding mode arises in such a system if the switching process dictates that system trajectories, after some finite time interval, are restricted to lie on a surface, or submanifold, of state space. Trajectories of such a system generally consist of two parts, representing two modes of the system: the *reaching mode* in which controlled trajectories lie off of the sliding surface, and the *sliding mode* during which trajectories are constrained

to lie on the sliding surface. Design of sliding mode control systems is naturally carried out in two stages: the sliding surface is selected to achieve the desired sliding dynamics, and the switching control law is selected to make the sliding surface attractive.

Following Filippov [17], Slotine and Sastry [117], Fernandez [118], Fernandez and Hendrick [119], and Sira-Ramirez [120, 121, 122], consider the control problem for a continuous-time single input system,

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{b}(\mathbf{x})u(\mathbf{x}), \quad (3.16)$$

where $\mathbf{x} \in \mathbf{X} \subseteq \mathbb{R}^n$, $\mathbf{f}, \mathbf{b} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ are, for convenience, C^∞ vector-valued functions, and the scalar feedback control $u : \mathbb{R}^n \rightarrow \mathbb{R}$. Let $s : \mathbb{R}^n \rightarrow \mathbb{R}$ denote a smooth function with nonzero gradient on $\mathbf{X}$. Define

$$\mathbf{S}_+ := \{\mathbf{x} : s(\mathbf{x}) > 0\}, \quad (3.17)$$

$$\mathbf{S} := \{\mathbf{x} : s(\mathbf{x}) = 0\}, \text{ and} \quad (3.18)$$

$$\mathbf{S}_- := \{\mathbf{x} : s(\mathbf{x}) < 0\}. \quad (3.19)$$

Here, $\mathbf{S}$ is an $(n-1)$ -dimensional submanifold of $\mathbf{X}$ and is oriented such that for any $\mathbf{x}^* \in \mathbf{S}$, the gradient $\partial s / \partial \mathbf{x}$ evaluated at $\mathbf{x}^*$ is directed from $\mathbf{x}^*$ into $\mathbf{S}_+$.

A variable structure control law is obtained by letting the scalar control $u(\mathbf{x})$ take one of two feedback values according to the sign of $s(\mathbf{x})$,

$$u(\mathbf{x}) = \begin{cases} u_+(\mathbf{x}) & \text{for } \mathbf{x} \in \mathbf{S}_+ \\ u_-(\mathbf{x}) & \text{for } \mathbf{x} \in \mathbf{S}_- \end{cases} \quad u_+ \neq u_-. \quad (3.20)$$

In this case $\mathbf{S}$ is called the *sliding manifold* or *switching surface*. In order to establish a *sliding regime* on $\mathbf{S}$, the control law (3.20) must be selected so that the trajectories of the controlled system off the sliding surface are directed towards the sliding surface as illustrated in Figure 3.1.

Let $d\sigma$ denote the one-form corresponding to the gradient, which itself is a row-vector valued function $\partial\sigma/\partial\mathbf{x}$, of the scalar function $\sigma(\mathbf{x})$. The Lie derivative, $L_{\mathbf{f}}(\sigma)$, of σ with respect to $\mathbf{f}$ is the directional derivative defined as

$$L_{\mathbf{f}}(\sigma) := \langle d\sigma, \mathbf{f} \rangle := \frac{\partial\sigma(\mathbf{x})}{\partial\mathbf{x}} \mathbf{f}(\mathbf{x}). \quad (3.21)$$

Figure 3.1: Conditions for existence of a sliding mode in variable structure systems

where $\langle \cdot, \cdot \rangle$ denotes the standard scalar product. The Lie derivative is a scalar field on $\mathbb{R}^n$ which leads naturally to the definition of higher-order Lie derivatives,

$$\begin{aligned} L_f^0(\sigma) &:= \sigma, \\ L_f^k(\sigma) &= L_f(L_f^{k-1}(\sigma)), \quad k = 1, 2, \dots \end{aligned} \quad (3.22)$$

Geometrically, $L_f(\sigma)$ may be interpreted as the rate of change of $\sigma(\mathbf{x})$ along the trajectory $\mathbf{f}(\mathbf{x})$, or equivalently as the component of the vector field $\mathbf{f}$ in the direction of the gradient $d\sigma$. With the Lie derivative notation, temporal dependence of the function $s(\mathbf{x})$, given the system dynamics in (3.16), can be written as

$$\dot{s}(\mathbf{x}(t)) = \frac{\partial s}{\partial \mathbf{x}} \frac{d\mathbf{x}}{dt} := L_{\mathbf{f}+\mathbf{b}u}(s). \quad (3.23)$$

A sliding regime is said [16, 117, 120] to exist locally on S whenever the *local sliding conditions*,

$$\lim_{s \rightarrow 0^+} L_{\mathbf{f}+\mathbf{b}u_+}(s) < 0 \quad \text{and} \quad \lim_{s \rightarrow 0^-} L_{\mathbf{f}+\mathbf{b}u_-}(s) > 0, \quad (3.24)$$

are satisfied. These conditions imply that the rate of change of $s(\mathbf{x})$ measured in the direction of the controlled field is such that a crossing of the surface is guaranteed [120], from each side of the surface, by use of the switching control logic (3.20). Alternatively, (3.24)

demonstrates that the components of the controlled fields $\mathbf{f} + \mathbf{b}u_+$ and $\mathbf{f} + \mathbf{b}u_-$ in the direction of the gradient of $s(\mathbf{x})$ are opposite in sign and hence the controlled fields locally point towards the sliding surface $\mathbf{S}$ as illustrated in Figure 3.1.

A commonly used *global sliding condition* is given by [117]

$$\frac{1}{2} \frac{d}{dt} s^2(\mathbf{x}) \leq -\eta |s(\mathbf{x})| \text{ for } \mathbf{x} \in \mathbb{R}^n - \mathbf{S}, \eta > 0, \quad (3.25)$$

which can be written in the more familiar form

$$\dot{s}(\mathbf{x}) \leq -\eta \operatorname{sgn}(s(\mathbf{x})), \quad (3.26)$$

where $\operatorname{sgn}(\cdot)$ is the signum function which returns $+1$, 0 , or -1 depending on whether its argument is positive, zero, or negative, respectively. It is understood that

$$\frac{1}{2} \frac{d}{dt} s^2(\mathbf{x}) = \begin{cases} sL_{\mathbf{f} + \mathbf{b}u_+}(s) & \text{for } \mathbf{x} \in S_+ \\ sL_{\mathbf{f} + \mathbf{b}u_-}(s) & \text{for } \mathbf{x} \in S_- \end{cases} \quad (3.27)$$

that is, the time derivative of $s^2(\mathbf{x})$ is evaluated along the trajectories of the controlled system. Condition (3.25) renders the surface $\mathbf{S}$ globally attractive and also satisfies the local sliding conditions (3.24).

An alternative approach, known as the *reaching law method*, which can be viewed as an extension of the above global sliding condition, has recently been proposed by Gao and Hung [123]. The reaching law is a differential equation which specifies the dynamics of the switching function $s(\mathbf{x})$ during the reaching mode. Generally, the reaching dynamics are specified in the form

$$\dot{s}(\mathbf{x}) = g(t, s(\mathbf{x})), \quad (3.28)$$

where the function $g(\cdot, \cdot)$ is selected so that $s = 0$ is attained in finite time. Gao and Hung [123] propose the form

$$\dot{s}(\mathbf{x}) = -\beta \operatorname{sgn}(s) - \gamma \phi(s), \quad (3.29)$$

where $\beta > 0$, $\gamma > 0$, and $s\phi(s) > 0$ with $\phi(0) = 0$. Taking $\phi(s) := s$, for example, it can be shown [123, 124] that any trajectory starting at $t = 0$ on the initial value $s(\mathbf{x}(0)) = s_0$ reaches the surface $\mathbf{S}$ in finite time t_f given by

$$t_f = \gamma^{-1} \ln \left(\frac{\gamma}{\beta} s_0 + 1 \right). \quad (3.30)$$

In Chapter 5, Section 5.7, a discrete-time reaching law method, analog to that for continuous-time systems in (3.29), is proposed and analyzed as a design methodology to specify the dynamics of the reaching mode of discrete-time sliding mode observers developed in Chapters 6 and 7. As discussed further in these chapters, the primary difficulty with a reaching law method for the observer problem is precisely that all states are not directly available and hence cannot be incorporated into the reaching law. A reaching law defined in terms of some, but not all of the states cannot completely define the reaching dynamics of the system, and this places further emphasis on stability analysis of the proposed observers during both the reaching and sliding phases.

A natural question is under what conditions is it possible to impose a sliding regime on S through a switching control law of the form (3.20). If $L_b(s) \equiv 0$ in (3.24), then

$$L_{f+bu}(s) = L_f(s), \quad (3.31)$$

and hence changing the control from u_+ to u_- does not have any effect on the sign of $L_{f+bu}(s)$. In this case the switching control (3.20) cannot create a sliding regime on S . Sira-Ramirez [121, 122] introduces the concept of *relative degree* of $s(x)$ with respect to the system (3.16) as the minimum number of times one has to differentiate $s(x)$ with respect to time in order to have the derivative explicitly dependent on $u(x)$. In this sense, the surface function $s(x)$ is said to have relative degree r with respect to $f + bu$ if

$$\begin{aligned} L_b(L_{f+bu}^k(s)) &= 0, \quad k < r - 1, \\ L_b(L_{f+bu}^{r-1}(s)) &\neq 0. \end{aligned} \quad (3.32)$$

Sira-Ramirez [121, 122] proves that a necessary and sufficient condition for the existence of a sliding regime on S , for the switching law of (3.20), is that $s(x)$ have relative degree $r = 1$ with respect to $f + bu$. The following example illustrates one method for derivation of the switching policy (3.20) for control of a class of nonlinear systems.

Example 3.1: Consider the problem of controlling the state x of the single input system (3.16) so that a smooth generally nonlinear scalar function of the state $y(t) = \sigma(x(t))$ tracks a desired smooth trajectory $y_d(t)$ under the assumptions that 1) all states are available for measurement, and 2) the scalar function $\sigma(x)$ has relative degree n with respect to the system (3.16). Following developments for similar problems considered by Fernandez

[118], Fernandez and Hendrick [119], and Sira-Ramirez [120, 121, 122], and due to item 2 above,

$$\frac{d^k y}{dt^k} = L_f^k(\sigma), \text{ for } k = 0, 1, \dots, n-1 \quad (3.33)$$

$$\frac{d^n y}{dt^n} = L_f^n(\sigma) + L_b(L_f^{n-1}(\sigma))u, \quad (3.34)$$

where $L_b(L_f^{n-1}(\sigma)) \neq 0$. The sliding surface S is chosen as a time varying surface $s(x; t) = 0$, where $s(x; t)$ is a linear stable differential operator

$$s(x; t) = \sum_{k=0}^{n-1} \lambda_k \frac{d^k \epsilon}{dt^k}, \quad (3.35)$$

where $\epsilon(t) = y_d(t) - y(t)$, and $\lambda_{n-1} = 1$. Note that $s(x; t)$ has relative degree 1. Assume that the reaching law of (3.28) has been specified. The switching control law is then obtained by substitution of (3.35), (3.33) and (3.34) into (3.28) followed by algebraic manipulation and simplification:

$$u = \frac{-1}{L_b(L_f^{n-1}(\sigma))} \left\{ L_f^n(\sigma) - \dot{y}_d^{(n)} + \sum_{k=0}^{n-2} \beta_k (L_f^{k+1}(\sigma) - \dot{y}_d^{(k+1)}) + g(t, s) \right\}. \quad (3.36)$$

where the notation $y_d^{(\ell)}$ denotes the ℓ th time derivative of y_d . □

Extension of this method to multiple input systems and to the case in which the nonlinear function $\sigma(x)$ is vector valued also follows essentially the methods of Fernandez [118] and Fernandez and Hendrick [119]. These authors, as well as Elmali and Olgac [125], Newman and Souccar [126], and Liao *et al.* [127], also demonstrate a theoretical link between this methodology for sliding mode controllers and that for the so-called input-output linearization problem [73]. The control law in (3.36) is consistent with that proposed by Slotine and Sastry [117] and Slotine [128]. One important drawback of this control law, however, is that it can result in control chattering which may involve extremely high control activity and excite high frequency unmodelled dynamics in the system. Several authors [117, 118, 119, 128, 129, 130, 131] propose approximations with continuous feedback laws through the use of saturation or sigmoid functions instead of the signum function. These methods not only reduce chattering in the control system, but in some cases reduce sensitivity to measurement noise. These possible benefits, however, are at the expense of reduced

robustness to unmodelled dynamics. Robustness properties are obtained by analysis of the *sliding dynamics*, which refers to the dynamics of the controlled system during the sliding mode.

Even if $s(\mathbf{x})$ has relative degree 1 and sliding conditions are satisfied, these conditions do not generally guarantee stable sliding dynamics [119, p. 1030]. Filippov's solution concept [17] or Utkin's [16] method of equivalent control provide a mathematical means to obtain the dynamics of the controlled system during *ideal sliding* which, in theory, requires infinite switching frequency as indicated in (3.20). For $\mathbf{x}^* \in S$ define

$$u_+(\mathbf{x}^*) := \lim_{(\mathbf{x} \in S_+) \rightarrow \mathbf{x}^*} u_+(\mathbf{x}), \text{ and} \quad (3.37)$$

$$u_-(\mathbf{x}^*) := \lim_{(\mathbf{x} \in S_-) \rightarrow \mathbf{x}^*} u_-(\mathbf{x}). \quad (3.38)$$

By Lemma 3 of Filippov [17, page 206], the dynamics on the sliding manifold S can only be a convex combination of the dynamics on either side of S :

$$\dot{\mathbf{x}}^* = \alpha(\mathbf{x}^*) [\mathbf{f}(\mathbf{x}^*) + \mathbf{b}(\mathbf{x}^*)u_+(\mathbf{x}^*)] + (1 - \alpha(\mathbf{x}^*)) [\mathbf{f}(\mathbf{x}^*) + \mathbf{b}(\mathbf{x}^*)u_-(\mathbf{x}^*)], \quad (3.39)$$

where $0 \leq \alpha(\mathbf{x}^*) \leq 1$. The sliding dynamics can be obtained through simplification of (3.39) as

$$\dot{\mathbf{x}}^* = \mathbf{f}(\mathbf{x}^*) + \mathbf{b}(\mathbf{x}^*)u_{eq}(\mathbf{x}^*) \quad (3.40)$$

where the equivalent control [16, 120], $u_{eq}(\mathbf{x}^*)$, is a smooth feedback control law for which S is a local invariant manifold of the controlled system and is given by

$$u_{eq}(\mathbf{x}^*) := \alpha(\mathbf{x}^*)u_+(\mathbf{x}^*) + [1 - \alpha(\mathbf{x}^*)]u_-(\mathbf{x}^*). \quad (3.41)$$

The function $\alpha(\mathbf{x}^*)$ is determined from the *manifold invariance conditions*:

$$s = 0, \quad \text{and} \quad L_{\mathbf{f}+\mathbf{b}u_{eq}}(s) := \frac{\partial s}{\partial \mathbf{x}}(\mathbf{x}^*)\dot{\mathbf{x}}^* = 0. \quad (3.42)$$

which give, from (3.42) and (3.39),

$$\alpha(\mathbf{x}^*) = \left[\frac{L_{\mathbf{f}+\mathbf{b}u_-}(s)}{L_{\mathbf{f}+\mathbf{b}u_-}(s) - L_{\mathbf{f}+\mathbf{b}u_+}(s)} \right]. \quad (3.43)$$

The denominator of (3.43) can be simplified as

$$\begin{aligned} L_{\mathbf{f}+\mathbf{b}u_-}(s) - L_{\mathbf{f}+\mathbf{b}u_+}(s) &= [L_{\mathbf{f}}(s) + L_{\mathbf{b}u_-}(s)] - [L_{\mathbf{f}}(s) + L_{\mathbf{b}u_+}(s)] \\ &= L_{\mathbf{b}}(s)[u_-(\mathbf{x}^*) - u_+(\mathbf{x}^*)]. \end{aligned} \quad (3.44)$$

which is nonzero due to the relative degree 1 constraint and the specification that $u_- \neq u_+$ in (3.20). The equivalent control is then obtained from (3.41) and (3.44) as

$$\begin{aligned}
 u_{eq}(\mathbf{x}^*) &= \alpha(\mathbf{x}^*)u_+(\mathbf{x}^*) + [1 - \alpha(\mathbf{x}^*)]u_-(\mathbf{x}^*) \\
 &= u_-(\mathbf{x}^*) - \alpha(\mathbf{x}^*)[u_-(\mathbf{x}^*) - u_+(\mathbf{x}^*)] \\
 &= \frac{-L_f(s)}{L_b(s)} \\
 &:= -\left(\frac{\partial s}{\partial \mathbf{x}}(\mathbf{x}^*)\mathbf{b}(\mathbf{x}^*)\right)^{-1} \frac{\partial s}{\partial \mathbf{x}}(\mathbf{x}^*)\mathbf{f}(\mathbf{x}^*). \tag{3.45}
 \end{aligned}$$

The ideal sliding dynamics are obtained by substituting (3.45) into (3.40),

$$\dot{\mathbf{x}}^* = \mathbf{F}_c(\mathbf{x}^*)\mathbf{f}(\mathbf{x}^*), \tag{3.46}$$

where

$$\mathbf{F}_c(\mathbf{x}^*) := \left[\mathbf{I}_n - \mathbf{b}(\mathbf{x}^*) \left(\frac{\partial s}{\partial \mathbf{x}}(\mathbf{x}^*)\mathbf{b}(\mathbf{x}^*) \right)^{-1} \frac{\partial s}{\partial \mathbf{x}}(\mathbf{x}^*) \right]. \tag{3.47}$$

Sira-Ramirez [120, Theorem 1] also proves that a necessary and sufficient condition for the existence of a sliding regime on $\mathbf{S}$ is that for $\mathbf{x}^* \in \mathbf{S}$,

$$\min \{u_+(\mathbf{x}^*), u_-(\mathbf{x}^*)\} < u_{eq}(\mathbf{x}^*) = \frac{-L_f(s)}{L_b(s)}(\mathbf{x}^*) < \max \{u_+(\mathbf{x}^*), u_-(\mathbf{x}^*)\}. \tag{3.48}$$

and that [120, Corollary 1] if a sliding regime exists on $\mathbf{S}$, one choice of switching logic which achieves the sliding regime is given by

$$u(\mathbf{x}) = k|u_{eq}(\mathbf{x})| \operatorname{sgn}\{s(\mathbf{x})\}, \tag{3.49}$$

where $k > 1$.

Robustness properties of sliding mode controllers have been analyzed by many authors, for example [117, 118, 119, 120, 122, 128, 131], and are most easily illustrated through geometrical interpretation of the sliding dynamics in (3.46) and (3.47). The matrix $\mathbf{F}_c(\mathbf{x}^*)$ defined in (3.47) satisfies, as can easily be verified by direct calculation, $\forall \mathbf{x}^* \in \mathbf{S}$,

$$\mathbf{F}_c^2(\mathbf{x}^*) = \mathbf{F}_c(\mathbf{x}^*), \tag{3.50}$$

$$\left[\frac{\partial s}{\partial \mathbf{x}}(\mathbf{x}^*) \right] \mathbf{F}_c(\mathbf{x}^*) = \mathbf{0}, \text{ and} \tag{3.51}$$

$$\mathbf{F}_c(\mathbf{x}^*)\mathbf{b}(\mathbf{x}^*) = \mathbf{0}. \tag{3.52}$$

Figure 3.2: The projection operation and matching conditions which form the robustness mechanism in continuous-time sliding mode control.

Hence, for any $x^* \in S$, F_c is a projection matrix taking $\mathbb{R}^n$ into the null space of $\frac{\partial s(x^*)}{\partial x^*}$ along $\text{span}\{b(x^*)\}$. Figure 3.2 illustrates this projection operation and also the matching conditions which form the robustness mechanism in continuous-time sliding mode control. The vector $[\frac{\partial s(x^*)}{\partial x^*}]^T$ is normal to the surface $s(x) = 0$ at $x = x^*$. The null space of $\frac{\partial s(x^*)}{\partial x^*}$ is the tangent plane to the surface at the point $x = x^*$. Trajectories of the controlled system will remain on the surface $s(x) = 0$ provided the motion, $\dot{x}^*$, of the system is tangent to the surface for all $x^* \in S$. Consider, then, the perturbed dynamical system,

$$\dot{x} = f(x) + b(x)u + \xi, \quad (3.53)$$

where ξ represents unmodelled dynamics and/or plant disturbances. Ideally, it would be desirable to specify the feedback control signal $u(x)$ such that the controlled dynamics are independent of the perturbation ξ . Intuitively, since the control effort enters into the dynamics (3.53) through premultiplication by $b(x)$, total rejection of the perturbation can only be accomplished if ξ lies in $\text{span}\{b(x)\}$. This fact can be demonstrated by replacing f with $f + \xi$ in the above developments to find that the equivalent control which satisfies the man-

ifold invariance conditions is

$$u_{eq}(\mathbf{x}^*) = - \left(\frac{\partial s}{\partial \mathbf{x}} \mathbf{b} \right)^{-1} \frac{\partial s}{\partial \mathbf{x}} (\mathbf{f} + \boldsymbol{\xi}). \quad (3.54)$$

and the ideal sliding dynamics become

$$\dot{\mathbf{x}}^* = \mathbf{F}_c(\mathbf{x}^*) (\mathbf{f}(\mathbf{x}^*) + \boldsymbol{\xi}). \quad (3.55)$$

The ideal sliding dynamics, therefore, are invariant to the perturbation $\boldsymbol{\xi}$ (i.e. $\mathbf{F}_c \boldsymbol{\xi} = \mathbf{0}$) if and only if [120] $\boldsymbol{\xi} \in \text{span}\{\mathbf{b}(\mathbf{x}^*)\}$. This condition is often called the *matching condition* under which the sliding mode is said to exhibit a *strong invariance property* with respect to $\boldsymbol{\xi}$. Figure 3.2 illustrates the matching condition and the projection operation. The instantaneous motion formed by the sum $\mathbf{f}(\mathbf{x}^*) + \boldsymbol{\xi}$ may be directed out of the tangent plane to the surface at $\mathbf{x} = \mathbf{x}^*$. Application of the projection $\mathbf{F}_c(\mathbf{x}^*)$ forces $\dot{\mathbf{x}}^*$ to lie in the tangent plane, and the magnitude and direction of $\dot{\mathbf{x}}^*$ are invariant to the disturbance $\boldsymbol{\xi}$ if the matching condition is satisfied. In the general case, the dynamics of the sliding mode are invariant, at each $\mathbf{x}^*$, to the component of any perturbations in $\text{span}\{\mathbf{b}(\mathbf{x}^*)\}$. It is the projection operation performed by $\mathbf{F}_c$ which results in the robustness properties for the controller problem. For the observer problem, a similar projection operation results which, in a sense, is "dual" to that for the controller problem.

An obvious problem in observer design, as opposed to the control problem, is precisely that the full state is not available for measurement and hence a sliding surface definition which depends on all states cannot be employed. The following discussion, which draws from the initial work of Misawa [102], and Slotine, Hendrick and Misawa [95], demonstrates the generation of sliding behavior in a second-order continuous-time system through input switching according to the value of a single component of the state.

Consider a second-order system with state vector $\mathbf{x} = [x_1, x_2]^T$ and simple dynamics

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= 0. \end{aligned} \quad (3.56)$$

This system has phase-plane trajectories as shown in Figure 3.3(a). If component x_1 is measured and employed in a switching term in the x_2 -channel,

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -\ell_2 \text{sgn}(x_1). \end{aligned} \quad (3.57)$$

where $\ell_2 > 0$, the resulting trajectories are as illustrated in Figure 3.3(b), but a sliding regime has not yet been introduced. If, however, switching based on x_1 is also introduced in the x_1 -channel according to

$$\begin{aligned}\dot{x}_1 &= x_2 - \ell_1 \operatorname{sgn}(x_1) \\ \dot{x}_2 &= -\ell_2 \operatorname{sgn}(x_1),\end{aligned}\tag{3.58}$$

where $\ell_1 > 0$, the resultant trajectories, shown in Figure 3.3(c), demonstrate that a sliding regime has been introduced on $x_1 = 0$, for $-\ell_1 < x_2 < \ell_1$. The inclusion of damping terms in the system,

$$\begin{aligned}\dot{x}_1 &= x_2 - k_1 x_1 - \ell_1 \operatorname{sgn}(x_1) \\ \dot{x}_2 &= -k_2 x_1 - \ell_2 \operatorname{sgn}(x_1),\end{aligned}\tag{3.59}$$

results in system trajectories as illustrated in Figure 3.3(d) and does not alter the extent of the sliding regime, but increases the region of direct attraction [102] to the sliding regime.

The above analysis can easily be applied to observer design for systems in companion form.

Example 3.2: Following Misawa [102], and Slotine, Hendrick and Misawa [95], consider the system

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{f}(\mathbf{x}),\tag{3.60}$$

where $\mathbf{x} \in \mathbb{R}^n$ and

$$\mathbf{A} = \begin{bmatrix} \mathbf{0}_{(n-1) \times 1} & \mathbf{I}_{n-1} \\ 0 & \mathbf{0}_{1 \times (n-1)} \end{bmatrix} \text{ and } \mathbf{f}(\mathbf{x}) = \begin{bmatrix} \mathbf{0}_{(n-1) \times 1} \\ f(\mathbf{x}) \end{bmatrix}.\tag{3.61}$$

with the linear scalar measurement model

$$z = \mathbf{C}\mathbf{x},\tag{3.62}$$

where $\mathbf{C} = [1, 0, \dots, 0]$.

The observer is defined as

$$\dot{\hat{\mathbf{x}}} = \mathbf{A}\hat{\mathbf{x}} + \hat{\mathbf{f}}(\hat{\mathbf{x}}) + \mathbf{K}\hat{z} + \mathbf{L}\operatorname{sgn}(\hat{z}),\tag{3.63}$$

Figure 3.3: Phase plane diagrams for simple continuous-time sliding mode observer.

where $\tilde{z} := z - \hat{z} := z - \mathbf{C}\hat{\mathbf{x}} = \mathbf{C}\mathbf{e}$ with $\mathbf{e} := \mathbf{x} - \hat{\mathbf{x}}$, $\tilde{\mathbf{f}}$ has the same form as $\mathbf{f}$ but with an estimate $\hat{f}$ of f , and $\mathbf{K}$ and $\mathbf{L}$ represent vectors of Luenberger and switching gains, respectively. The observer error $\mathbf{e}(t)$ has dynamics

$$\dot{\mathbf{e}}(t) = \mathbf{A}_0 \mathbf{e}(t) + \tilde{\mathbf{f}}(\mathbf{x}, \hat{\mathbf{x}}) - \mathbf{L} \operatorname{sgn}(s(\mathbf{e})), \quad (3.64)$$

where $\mathbf{A}_0 := [\mathbf{A} - \mathbf{K}\mathbf{C}]$, $\tilde{\mathbf{f}}(\mathbf{x}, \hat{\mathbf{x}}) := \mathbf{f}(\mathbf{x}) - \hat{\mathbf{f}}(\hat{\mathbf{x}})$, and $s(\mathbf{e}) := \mathbf{C}\bar{\mathbf{x}}$. Note that the error dynamics in (3.64), for $n = 2$ have the same form as (3.59) apart from the modelling error term.

A sliding regime exists on the surface $s(\mathbf{e}) = 0$ provided $\lim_{s \rightarrow 0} s\dot{s} < 0$. Simple manipulations give, with $\mathbf{e} := [e_1, \dots, e_n]^T$, $\mathbf{K} := [k_1, \dots, k_n]^T$, and $\mathbf{L} := [\ell_1, \dots, \ell_n]^T$,

$$s\dot{s} = \mathbf{e}^T \mathbf{C}^T \mathbf{C} \dot{\mathbf{e}} = e_1 e_2 - e_1^2 k_1 - |e_1| \ell_1. \quad (3.65)$$

Hence a sliding regime exists locally on $s(\mathbf{e}) := e_1 = 0$ for $-\ell_1 < e_2 < \ell_1$.

The dynamics on the sliding regime are determined by Filippov's solution concept: for $\mathbf{e}^* \in \mathbf{S} = \{\mathbf{e} \mid s(\mathbf{e}) = \mathbf{C}\mathbf{e} = 0, -\ell_1 < e_2 < \ell_1\}$, the dynamics of $\mathbf{e}^*$ are given by the convex combination of the dynamics on either side of $\mathbf{S}$,

$$\begin{aligned} \dot{\mathbf{e}}^*(t) &= \alpha(\mathbf{e}^*) [\mathbf{A}_0 \mathbf{e}^* + \tilde{\mathbf{f}} + \mathbf{L}] + (1 - \alpha(\mathbf{e}^*)) [\mathbf{A}_0 \mathbf{e}^* + \tilde{\mathbf{f}} - \mathbf{L}] \\ &= \mathbf{A}_0 \mathbf{e}^* + \tilde{\mathbf{f}} - \mathbf{L} + 2\mathbf{L}\alpha(\mathbf{e}^*). \end{aligned} \quad (3.66)$$

The function $\alpha(\mathbf{e}^*)$ is determined from the manifold invariance conditions $s(\mathbf{e}^*) := \mathbf{C}\mathbf{e}^* = 0$ and $\dot{s}(\mathbf{e}^*) := \mathbf{C}\dot{\mathbf{e}}^* = 0$ which give $\mathbf{A}_0 \mathbf{e}^* = \mathbf{A}\mathbf{e}^*$ and

$$\alpha(\mathbf{e}^*) = -\frac{1}{2} [\mathbf{C}\mathbf{L}]^{-1} \mathbf{C} [\mathbf{A}\mathbf{e}^* + \tilde{\mathbf{f}} - \mathbf{L}]. \quad (3.67)$$

Substituting (3.67) into (3.66) followed by simplification gives

$$\dot{\mathbf{e}}^*(t) = [\mathbf{I}_n - \mathbf{L}(\mathbf{C}\mathbf{L})^{-1} \mathbf{C}] [\mathbf{A}\mathbf{e}^* + \tilde{\mathbf{f}} - \mathbf{L}]. \quad (3.68)$$

Recall from previous discussions that the matrix

$$\mathbf{F}_o := [\mathbf{I}_n - \mathbf{L}(\mathbf{C}\mathbf{L})^{-1} \mathbf{C}] \quad (3.69)$$

represents a projection onto the null space of $\mathbf{C}$ along $\operatorname{span}\{\mathbf{L}\}$. Figure 3.4 illustrates the

Figure 3.4: The projection operation and matching conditions which form the robustness mechanism in a continuous-time sliding mode observer.

projection operation which is seen, through comparison with Figure 3.2, to be consistent with the robustness mechanism of sliding mode control systems. This means that $F_o L = 0$, which gives the final result

$$\dot{e}^*(t) = F_o [Ae^* + \tilde{f}]. \quad (3.70)$$

For the case $n = 2$ with $\tilde{f} = 0$, this result implies that on the sliding regime S , $e_1 = 0$ and e_2 decreases exponentially to zero with time constant ℓ_1/ℓ_2 . The important point demonstrated by this example is that the sliding dynamics, as was the case in the control problem, result from a projection operation which illustrates the robustness properties of the observer. $\square$

It is useful at this point to consider the “dual” nature of these results for the controller and observer problems¹ as illustrated in Figures 3.2 and 3.4. In the control problem, it is often the case that the sliding surface is specified in the form $s(x) = \Gamma x$. In this case the

¹These are not precisely the same duality concepts that are well known in classical controller and observer design methodologies for linear systems.

projection operator for the control problem becomes, from (3.47),

$$\mathbf{F}_c := [\mathbf{I}_n - \mathbf{b}(\Gamma\mathbf{b})^{-1}\Gamma]. \quad (3.71)$$

Comparison of (3.71) and (3.69) suggests the correspondences $\Gamma \sim \mathbf{C}$ and $\mathbf{b} \sim \mathbf{L}$. In the control problem, $\mathbf{b}$ is given while Γ is to be determined, whereas in the observer problem, $\mathbf{C}$ is given and $\mathbf{L}$ is to be determined. In the controller problem, the sliding mode is invariant to disturbances provided it is known *a priori* that the disturbances are in $\text{span}\{\mathbf{b}\}$ —these are the matching conditions which are often assumed. In the observer problem, it is desirable, if possible, to select the observer gains $\mathbf{L}$ so that disturbances lie in $\text{span}\{\mathbf{L}\}$ in which case the sliding mode is invariant to these disturbances. In the control problem, all states are assumed to be available, and Γ is selected so that the zero dynamics given by $\Gamma\mathbf{x} = 0$ are asymptotically stable. In the observer problem, however, $\mathbf{C}$ is specified and it is required that any dynamics present in the equation $\mathbf{C}\mathbf{e} = 0$ be asymptotically stable.

The discrete-time sliding mode observers proposed in Chapters 6 and 7 are developed for systems of the general form (this notation is introduced more formally in Chapter 6)

$$\mathbf{x}_e(k+1) = \mathbf{f}(k, \mathbf{x}_e(k)), \quad (3.72)$$

$$\mathbf{z}(k) = \mathbf{h}(\mathbf{x}(k)), \quad \mathbf{x}(k) := \mathbf{C}\mathbf{x}_e(k), \quad (3.73)$$

where $\mathbf{C} = [\mathbf{I}_q, \mathbf{0}_{q \times (n-q)}]$. Note that, in this representation, the measurements $\mathbf{z}(k)$, for each k , depend only on the first q elements, namely $\mathbf{x}$, of the full state vector $\mathbf{x}_e$. The proposed observers all have the form

$$\hat{\mathbf{x}}_e(k+1) = \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \mathbf{L}(k)\mathbf{v}(k), \quad (3.74)$$

where $\mathbf{L}(k)$ represents the matrix of observer gains to be determined. It is shown in Chapters 6 and 7 that, at least for ideal observer forms, any unmodelled dynamics and disturbances which enter into the observer error dynamics during sliding are premultiplied by a projection matrix of the form

$$\mathbf{F}(k) = [\mathbf{I}_n - \mathbf{L}(k)(\mathbf{C}\mathbf{L}(k))^{-1}\mathbf{C}], \quad (3.75)$$

which takes vectors along $\text{span}\{\mathbf{L}(k)\}$ onto the null space of $\mathbf{C}$. It is this projection operation which, consistent with the above results for the continuous-time controller and ob-

server problems, is responsible for the robustness properties of the proposed observers; the observer error is invariant to any unmodelled dynamics or disturbances in $\text{span}\{\mathbf{L}(k)\}$.

Misawa [102], and Slotine, Hendrick, and Misawa [95] present further analysis of the above approach to observer design for continuous-time systems in terms of robustness to unmodelled dynamics, selection of switching gains, and performance issues in the presence of measurement errors. Applications of these methods for systems in companion form have recently appeared in publications by Canudas de Wit and his colleagues [96, 97, 98, 99], however, this work is explicitly dependent on specific properties of manipulator dynamics. Extensions of sliding mode observer methods to general nonlinear systems (not necessarily in companion form) have been investigated by Misawa [102] and Misawa and Hendrick [78], but the design methodology is somewhat complex and limited [132].

Applications of sliding mode control systems in many cases entails computer implementation at finite update rates. Early extensions to the discrete-time case of the continuous-time conditions for the existence of a sliding regime were proposed by Milosavljevic [133] in the context of linear time invariant systems. This approach, again for linear systems, was refined by Sarpturk *et al.* [19] to obtain sufficient conditions for convergence to the sliding surface. Following this approach, Opitz [134], Furuta [135], Aly and Ali [136], and Chan [137] proposed methods for design and analysis of sliding mode controllers for discrete-time linear time invariant systems. Contributions to nonlinear control problems have also appeared, for example by Sira-Ramirez [18], Drakunov and Utkin [138], Hashimoto and Ishikawa [139], Kaynak and Denker [140], and Baida [141].

Consider a single input system of the form

$$\mathbf{x}(k+1) = \mathbf{F}(\mathbf{x}(k), u(k)), \quad (3.76)$$

where $\mathbf{x} \in \mathbb{R}^n$, and u is the scalar control. The sliding surface is defined as

$$\mathbf{S} := \{\mathbf{x} \mid s_k := s(\mathbf{x}(k)) = 0\}. \quad (3.77)$$

Milosavljevic [133] introduced the term *quasi-sliding regime* for reference to discrete-time systems and defined existence of a quasi-sliding regime locally on $\mathbf{S}$ whenever

$$s_k(s_{k+1} - s_k) < 0. \quad (3.78)$$

This condition can be seen as a direct extension of the continuous-time criterion of the form $s\dot{s} < 0$ by using a forward difference approximation for the derivative $\dot{s}$. An alternate definition has been proposed by Drakunov and Utkin [142] in which a discrete-time sliding mode is said to exist whenever it is possible, once the state gets close enough to the surface, to maintain the state on the surface from the next sample time onward. These definitions are discussed further in Section 4.2.

Condition (3.78) is necessary but not sufficient for the existence of stable quasi-sliding motion [18, 19]. In particular, condition (3.78) does not ensure any convergence of state trajectories to the sliding surface S [18, Ex. 3.1]. A necessary and sufficient condition [19] for the existence of a convergent quasi-sliding regime on S is that

$$|s_{k+1}| < |s_k|. \quad (3.79)$$

Condition (3.79) can be decomposed as

$$[s_{k+1} - s_k] \operatorname{sgn}(s_k) < 0, \quad (3.80)$$

$$[s_{k+1} + s_k] \operatorname{sgn}(s_k) > 0. \quad (3.81)$$

in which (3.80) is called [19] the *sliding condition*, while (3.81) is called the *convergence condition*. Condition (3.79), however, does not necessarily guarantee chattering motions about the sliding surface [18, Ex. 3.2], and it has been noted [18, 19] that, in contrast to the continuous-time case, quasi-sliding regimes can be achieved in discrete-time systems with continuous switching strategies. In Section 4.3, Theorem 4.2, it is shown that continuous switching about the surface is a necessary condition for existence of convergent quasi-sliding motion.

Further developments by Sira-Ramirez [18] for the discrete-time case follow very closely his previous treatment of continuous-time systems [120, 121, 122]. In particular, concepts are introduced for discrete-time relative degree analog to the continuous-time case, and for analysis which provides necessary and/or sufficient conditions for existence and stability issues as well as extensions of the method of equivalent control to discrete-time systems.

3.6 Chapter Summary

This chapter has defined the observer problem which is to be solved, described the class of discrete-time systems considered in this research, gave the form of observers which are developed in subsequent chapters, and provided a review of nonlinear observability concepts and current methods of nonlinear state estimation. The review of fundamental concepts of variable structure systems in sliding mode discussed contributions to both the controller and observer problems for continuous-time systems with emphasis on robustness properties. Apparently, duality concepts between the sliding mode controller and observer problems, as discussed in this Chapter, have not previously appeared in the literature. Extensions of sliding mode methods to discrete-time systems were also introduced and are developed further in Chapter 4.

Chapter 4

A Simple Sliding Mode Observer

4.1 Introduction

The method presented and analyzed in this Chapter is motivated in part by approaches to continuous-time sliding mode observers, and is also due in part to initial investigations in the discrete-time sliding mode control problem [18, 19, 134, 142]. Although some effort has been devoted to the nonlinear controller problem, discrete-time sliding mode observers, particularly for nonlinear systems, are currently not available. In Section 4.2 a simple discrete-time sliding mode observer is developed as a direct application, to the discrete-time case, of continuous-time methods for linear systems proposed in Misawa's dissertation [102]. Misawa's development was previously reviewed in Section 3.5 with a two-dimensional state space and flow diagrams. Similar analysis, presented in Section 4.2, illustrates generation of quasi-sliding motion with discontinuous switching based on available measurements for a two-dimensional linear time invariant discrete-time system. Analysis of applications of this method to higher order linear systems in the absence of modelling errors demonstrate that, under fairly mild assumptions, it is always possible to establish a quasi-sliding regime. Discontinuous switching, however, can give rise to steady-state limit cycles which means that this approach need not yield a convergent quasi-sliding regime.

Motivation for further research is provided through simulation studies and additional analysis presented in Section 4.3. Simulation results demonstrate that the sliding mode method may yield significant performance improvements over the Kalman filtering approach in the presence of unmodelled dynamics, particularly when measurement noise is

not too severe. Although simulation results are encouraging, direct applications of this method to more general nonlinear systems with modelling errors are not apparent. The primary difficulty with these applications of the observer in its present form results from dependence of the sliding surface and convergence criteria on the unknown modelling errors. Moreover, it is shown for the case of linear systems that a convergent quasi-sliding regime cannot exist in the presence of discontinuous switching. This result leads directly to one form of observer for linear systems such that a global sliding regime can be created. This form of the observer is shown to have the dynamics of a conventional reduced-order Luenberger observer for the discrete-time linear system.

4.2 Form of the Observer

In Section 3.5, a two dimensional continuous-time problem was considered in order to demonstrate the generation of sliding behavior through switching according to the value of a single component of the state. Very similar analysis can be used for simple discrete-time systems. Consider a linear time invariant system with scalar measurements

$$\mathbf{x}(k+1) = \Phi \mathbf{x}(k), \quad z(k) = \mathbf{C} \mathbf{x}(k), \quad (4.1)$$

where $\mathbf{x} \in \mathbb{R}^n$, $\Phi \in \mathbb{R}^{n \times n}$ is the state transition matrix, $\mathbf{C} \in \mathbb{R}^{1 \times n}$, and $z \in \mathbb{R}$ is the scalar measurement.

Neglecting modelling error for initial developments, a discrete-time observer is proposed with the form

$$\hat{\mathbf{x}}(k+1) = \Phi \hat{\mathbf{x}}(k) + \mathbf{K} \eta(k+1) + \mathbf{L} \operatorname{sgn}(s(k)), \quad (4.2)$$

where the innovation sequence (prediction error), $\eta(k+1)$, is defined as

$$\eta(k+1) := z(k+1) - \mathbf{C} \Phi \hat{\mathbf{x}}(k), \quad (4.3)$$

and the Luenberger gains, $\mathbf{K} \in \mathbb{R}^n$, and switching gains, $\mathbf{L} \in \mathbb{R}^n$, are to be determined. This form of observer is a direct application of methods previously proposed for continuous-time systems [102]. The switching surface is defined in terms of the observer

error, $\mathbf{e}(k) := \mathbf{x}(k) - \hat{\mathbf{x}}(k)$, as

$$s_k := s(\mathbf{e}(k)) = \mathbf{M}\mathbf{e}(k) = 0, \quad (4.4)$$

where $\mathbf{M} \in \mathbb{R}^{1 \times n}$ must be selected so that s_k can be computed from the available measurements at time $k + 1$. Two possible choices are

$$\mathbf{M} = \mathbf{C} \Rightarrow s_k = z(k) - \mathbf{C}\hat{\mathbf{x}}(k), \quad \text{or} \quad (4.5)$$

$$\mathbf{M} = \mathbf{C}\Phi \Rightarrow s_k = z(k + 1) - \mathbf{C}\Phi\hat{\mathbf{x}}(k). \quad (4.6)$$

With $\mathbf{M}$ selected as in (4.5), the contribution of the switching term in (4.2) which affects the state estimate at time $k + 1$ is based on the error at the previous time step, whereas the choice of $\mathbf{M}$ in (4.6) implies that the switching term used in the computation of $\hat{\mathbf{x}}(k + 1)$ is based on the measurement at time $k + 1$ and the predicted state $\Phi\hat{\mathbf{x}}(k)$.

In the absence of modelling error, the observer error dynamics are

$$\mathbf{e}(k + 1) = \Phi_o \mathbf{e}(k) - \mathbf{L} \operatorname{sgn}(s(k)), \quad (4.7)$$

where

$$\Phi_o := [\mathbf{I}_n - \mathbf{K}\mathbf{C}] \Phi. \quad (4.8)$$

Provided $(\mathbf{C}, \Phi)$ is an observable pair, $\mathbf{K}$ can be selected [143, p. 255] to place the eigenvalues of Φ_o within the unit circle. Generation of quasi-sliding behavior for a two-dimensional system is demonstrated in the following example.

Example 4.1: Consider the constant velocity system (4.1) with

$$\Phi = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{C} = [1, 0], \quad (4.9)$$

where T is the sample period. The observer is defined as in (4.2) with $\mathbf{M} := \mathbf{C}\Phi$, $\mathbf{K} := [k_1, k_2]^T$ satisfying $1 - k_1 - k_2 T > 0$, and $\mathbf{L} := [\ell_1, \ell_2]^T$ satisfying $\mathbf{M}\mathbf{L} > 0$. In this case, with $\mathbf{e}(k) := [e_1(k), e_2(k)]^T$, the switching line is $s_k = \mathbf{C}\Phi\mathbf{e}(k) = e_1(k) + T e_2(k) = 0$. Assuming that the eigenvalues of Φ_o are less than unity in modulus, the motion of the error system, $\delta\mathbf{e}(k) := \mathbf{e}(k + 1) - \mathbf{e}(k)$, in the phase plane is illustrated in Figure 4.1. From the form of the observer error, one can easily identify four regions in Figure 4.1: in

Figure 4.1: Phase plane diagram simple discrete-time sliding mode observer

region R_1 , $\delta \mathbf{e} := \mathbf{e}(k+1) - \mathbf{e}(k)$ lies in the fourth quadrant; in R_2 , $\delta \mathbf{e}$ lies in the third quadrant; in R_3 , the second quadrant; and in R_4 , the first quadrant. Where regions R_2 and R_4 meet along $s(\mathbf{e}) = 0$ between the points A and B , motion of the error system on either side of the sliding surface is seen to be directed towards the sliding surface which is the fundamental requirement to establish a quasi-sliding regime on $s = 0$. Straight-forward calculations show that local to $s_k = 0$, error trajectories satisfy $s_k(s_{k+1} - s_k) < 0$ between the points A' and B' for which

$$(e_1, e_2) = \pm \left(\ell_1 + T\ell_2, -\frac{\ell_1 + T\ell_2}{T} \right). \quad (4.10)$$

Figure 4.2: Phase plane error trajectories for simple discrete-time sliding mode observer.

Figure 4.2 shows the motion of error trajectories for this system from a dense grid of points in the phase plane. In this Figure, the quasi-sliding regime is seen as that region in which the motion of the system on either side of the sliding surface is directed towards the surface. Example error trajectories for the constant velocity system with unity sample period are shown in Figure 4.3. The observer has an initial error of 3 units in position and 12 units/s in velocity. The motion of the error system is initially directed in the fourth quadrant which corresponds to region R_1 in Figure 4.1, then motion is in the third quadrant, crosses the switching surface and moves in the second quadrant while in region R_4 . The lower graph of Figure 4.3 shows an expansion of the sliding surface where the error motion chatters about the surface and eventually reaches a limit cycle passing through the origin. $\square$

Recall that in the continuous-time case reviewed in Section 3.5, the switching surface in

Figure 4.3: Example error trajectories of simple discrete-time sliding mode observer for 2D system.

the observer problem was along the e_2 axis and hence switching was based on the estimation error in the position channel only. This would also be the case if the switching function s_k was defined with $M = C$. In the above example, however, $M := C\Phi$ which means that the switching process depends not only on the position error but also on the velocity estimation error. This choice for the sliding surface, although acceptable performance may be obtained in some simple cases, is not suitable, as discussed later in this chapter, for the case of nonlinear systems.

Before proceeding to consider more general systems than that treated in the above example, it will be useful to define terminology for notions of quasi-sliding motion treated by Sarpturk *et al.* [19] and the sliding motion considered by Drakunov and Utkin [142]. These definitions are also used in the developments of subsequent chapters.

Definition 4.1: (Sarpturk) For a discrete-time system of the form $\mathbf{x}(k+1) = \mathbf{f}(k, \mathbf{x}(k))$, $\mathbf{x}(k) \in \mathbf{X} \subseteq \mathbb{R}^n \forall k \geq 0$, the smooth surface $\mathbf{S} := \{\mathbf{x} \in \mathbf{X} \mid s(\mathbf{x}) = 0\}$, and $s_k := s(\mathbf{x}(k))$,

1. a *quasi-sliding regime* exists locally on $\mathcal{S} \subset \mathbf{S}$ if there exists an open neighbourhood $\mathcal{U}$ of $\mathcal{S}$ such that $\forall \mathbf{x}(k) \in \mathcal{U} - \mathcal{S}$, $s_k(s_{k+1} - s_k) < 0$;
2. a *convergent quasi-sliding regime* exists locally on $\mathcal{S} \subset \mathbf{S}$ if there exists an open neighbourhood $\mathcal{U}$ of $\mathcal{S}$ such that if, for some $k_0 \geq 0$, $\mathbf{x}(k_0) \in \mathcal{U}$, then $\forall k \geq k_0$, whenever $|s_k| > 0$, $|s_{k+1}| < |s_k|$, and for any $k_1 \geq k_0$, if $s_{k_1} = 0$, then $\forall k \geq k_1$, $s_k = 0$. □

Definition 4.2: (Drakunov) For a discrete-time system of the form $\mathbf{x}(k+1) = \mathbf{f}(k, \mathbf{x}(k))$, $\mathbf{x}(k) \in \mathbf{X} \subseteq \mathbb{R}^n \forall k \geq 0$, and the smooth surface $\mathbf{S} := \{\mathbf{x} \in \mathbf{X} \mid s(\mathbf{x}) = 0\}$, and $s_k := s(\mathbf{x}(k))$, a *discrete-time sliding regime* exists locally on $\mathcal{S} \subset \mathbf{S}$ if there exists an open neighbourhood $\mathcal{U}$ of $\mathcal{S}$ such that for some $k_0 \geq 0$, if $\mathbf{x}(k_0) \in \mathcal{U}$, $s(\mathbf{x}(k)) = 0 \forall k > k_0$. □

The primary differences between these definitions are as follows: in a quasi-sliding regime, trajectories on either side of the surface are directed towards the surface, but convergence to the surface is not necessarily guaranteed; a convergent quasi-sliding regime is a quasi-sliding regime in which convergence to the surface is guaranteed, but convergence may only be asymptotic; while a discrete-time sliding regime exists if whenever the state

is close enough to the surface, the state can be restricted to lie on the surface from the next time step onward.

The following theorem illustrates a fundamental property of the linear discrete-time system and simple observer proposed above.

Theorem 4.1: Consider the linear time invariant system (4.1) and observer (4.2). For any given $\Phi \in \mathbb{R}^{n \times n}$, $K \in \mathbb{R}^n$, and nonzero $M \in \mathbb{R}^{1 \times n}$, there exists a vector of switching gains $L \in \mathbb{R}^n$ such that a quasi-sliding regime exists on the surface $S = \{e \mid s(e) := Me = 0\}$.

Proof: Given Φ , K , and M , because $M \in \mathbb{R}^{1 \times n}$ is nonzero, there exists an L such that the scalar inequality

$$ML > 0 \quad (4.11)$$

is satisfied. For any L satisfying (4.11), define the one-parameter family of open sets $\mathcal{U}_\delta$ as

$$\mathcal{U}_\delta := \left\{ e \in \mathbb{R}^n \mid \left| \frac{M[\Phi_o - I_n]e}{ML} \right| < 1 - \delta \right\}, \quad 0 \leq \delta < 1. \quad (4.12)$$

Fix $\epsilon \in (0, 1)$. Set

$$\mathcal{S} = \mathcal{U}_{\delta=\epsilon} \cap S \quad \text{and} \quad \mathcal{U} = \mathcal{U}_{\delta=0}. \quad (4.13)$$

Then $\mathcal{S}$ is nonempty, $\mathcal{U}$ is an open neighbourhood of $\mathcal{S}$, and $\forall e(k) \in \mathcal{U} - \mathcal{S}$,

$$\begin{aligned} \text{sgn}\{s_{k+1} - s_k\} &= \text{sgn}\{M[e(k+1) - e(k)]\} \\ &= \text{sgn}\{M[\Phi_o e(k) - L \text{sgn}(s_k)]\} \\ &= \text{sgn}\{ML\} \text{sgn}\left\{ \frac{M[\Phi_o - I_n]e(k)}{ML} - \text{sgn}(s_k) \right\} \\ &= -\text{sgn}\{s_k\}. \end{aligned} \quad (4.14)$$

Hence $\forall e(k) \in \mathcal{U} - \mathcal{S}$, $s_k(s_{k+1} - s_k) < 0$. $\square$

The set $\mathcal{U}$ in Theorem 4.1 is the region containing the origin between two hyperplanes in $\mathbb{R}^n$. For $e \in \mathbb{R}^2$, these hyperplanes are reduced to lines which are illustrated in Figure 4.1 by the dotted lines emanating from points A' and B' . The size of this region varies directly with the size of the scalar ML ; with all other variables fixed, increasing ML , by increasing the magnitude of elements of L , increases the extent of the quasi-sliding regime. Note

also that the hypothesis of Theorem 4.1 do not require any specific properties of the matrix Φ_o . However, the results of Theorem 4.1 do not give any insight into the stability of the observer. Stability analysis, particularly in the presence of modelling errors, is best performed with methods proposed in Chapter 5, and hence further consideration of stability issues is deferred to the next Chapter.

The fact that this observer does not generally possess a convergent quasi-sliding regime is demonstrated through the existence of a steady state limit cycle.

Example 4.2: If $\text{sgn}\{M[I_n + \Phi_o]^{-1}L\} = 1$, where the inverse exists if all eigenvalues of Φ_o have modulus less than unity, then there exists an $e_\ell \in \mathbb{R}^n$ such that

$$e_\ell = [I_n + \Phi_o]^{-1}L \text{sgn}\{Me_\ell\} \quad (4.15)$$

is consistent. With $e(k) = e_\ell$, the error at the next time step satisfies

$$\begin{aligned} e(k+1) &= \Phi_o e(k) - L \text{sgn}\{Me(k)\} \\ &= -[I_n - \Phi_o[I_n + \Phi_o]^{-1}]L \text{sgn}\{Me(k)\} \\ &= -[I_n + \Phi_o]^{-1}L \text{sgn}\{Me(k)\} \\ &= -e(k), \end{aligned} \quad (4.16)$$

where the second last equality in (4.16) is obtained by applying the Matrix Inversion Lemma,

$$[A + BED]^{-1} = A^{-1} - A^{-1}B[DA^{-1}B + E^{-1}]^{-1}DA^{-1}, \quad (4.17)$$

with $A := I_n$, $B := \Phi_o$, and $E := D := I_n$. Hence if, for some $k_0 \geq 0$, $e(k_0) = e_\ell$, then $e(k+1) = -e(k) \forall k \geq k_0$. There may also exist other limit cycles, possibly with period greater than twice the sample frequency. $\square$

In linear filtering theory for discrete-time systems, placement of poles (in the Z -transform domain) on the negative real axis often results in an undesirable phenomenon called *ringing* in the filter output. A pole placed at $z = -1$ in the Z -plane can lead to undamped ringing which corresponds to a steady-state limit cycle such as that illustrated in the above example. The amplitude of the limit cycle varies directly with the strength of the switching gains L . This suggests a tradeoff between enlarging the quasi-sliding regime by increasing the strength of the switching gains, and reducing the amplitude of ringing in the observer by decreasing the strength of the switching gains.

4.3 Motivation for Further Study

The previous Section presented the form of a simple discrete-time sliding mode observer and discussed some of its fundamental properties. In this Section, motivation for further research is provided through a comparison of simulation results of this observer with the Kalman filter in the presence of unmodelled dynamics. Difficulties with analysis of observer performance for nonlinear systems and unmodelled dynamics are then highlighted which suggest that applications of this approach to more general systems is not immediate. Necessary conditions for existence of a convergent quasi-sliding regime are given which lead directly to one form of linear system observers for which a global discrete-time sliding regime can be generated. This form of the observer is shown to have the dynamics of a conventional reduced-order Luenberger observer for the discrete-time linear system.

Consider a particle moving in one dimension whose position, $p(t)$, velocity, $\dot{p}(t)$, and acceleration, $\ddot{p}(t)$, are modelled according to the methods developed in Section 2.4.3. The state vector is defined as $\mathbf{x} = [p, \dot{p}, \ddot{p}]^T$ which propagates in time according to

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{G}w(t), \quad (4.18)$$

where

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -\alpha \end{bmatrix}, \text{ and } \mathbf{G} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \quad (4.19)$$

and $w(t)$ is zero-mean, temporally white Gaussian noise with covariance $Q = 2\alpha\sigma^2$. The linear differential equation in (4.18), with the definitions in (4.19), can be written as a discrete-time difference equation using standard techniques to give

$$\mathbf{x}(k+1) = \Phi(k)\mathbf{x}(k) + \mathbf{w}_d(k), \quad (4.20)$$

where $\Phi(k) = \exp\{\mathbf{A}(t_{k+1} - t_k)\}$. The system is driven by a zero-mean, discrete-time, white noise sequence, $\mathbf{w}_d(k)$, with covariance $\mathbf{Q}_d$ which may be approximated, assuming the sample period, T , is much smaller than system time constants, as $\mathbf{Q}_d \approx 2\alpha\sigma^2 T \mathbf{G} \mathbf{G}^T$. Assume that for this system, only position is measured as

$$z(k) = \mathbf{C}\mathbf{x}(k) + v(k), \quad (4.21)$$

where $\mathbf{C} = [1, 0, 0]$, and $v(k)$ is additive zero-mean measurement noise with covariance matrix $\mathbf{R}$.

The discrete-time Kalman filter (see Appendix B) for this linear problem computes the estimate $\hat{\mathbf{x}}(k) \equiv \hat{\mathbf{x}}(k|k)$ from the previous estimate $\hat{\mathbf{x}}(k-1) \equiv \hat{\mathbf{x}}(k-1|k-1)$ according to

$$\hat{\mathbf{x}}(k) = \Phi \hat{\mathbf{x}}(k-1) + \mathbf{K}(k)\eta(k), \quad (4.22)$$

where $\eta(k) := \{z(k) - \mathbf{C}\Phi \hat{\mathbf{x}}(k-1)\}$, and $\mathbf{K}(k)$ are the discrete-time Kalman gains. Because the system is linear time invariant, the Kalman gains eventually reach steady state, $\mathbf{K}(k) \rightarrow \mathbf{K}_{SS}$ as $k \rightarrow \infty$.

The sliding mode observer proposed in Section 4.2 can be constructed as

$$\hat{\mathbf{x}}(k) = \Phi \hat{\mathbf{x}}(k-1) + \mathbf{K}_{SS}\eta(k) + \mathbf{L} \operatorname{sgn}(s_k), \quad (4.23)$$

which employs the steady-state Kalman gains, $\mathbf{K}_{SS}$, for the Luenberger part, and constant switching gains, $\mathbf{L}$, that are to be determined. In this Section, the switching surface is taken as the prediction error,

$$s_k := \{z(k) - \mathbf{C}\Phi \hat{\mathbf{x}}(k-1)\}. \quad (4.24)$$

Figure 4.4 compares simulation results for the Kalman filter and the above sliding mode observer in the absence of measurement noise when the true motion of the particle is $p(t) = \sin(4\pi t)$. Parameters for the Kalman filter were chosen as in Section 2.6.2 as $\alpha = 0.5$, $\sigma = 8$, and $\mathbf{R} = 0.01$, but for this example the sample frequency was chosen very high at 1kHz. The Kalman filter gives significant lag in all states as well as overshoot in position and velocity estimates for this trajectory. The steady-state Kalman gains were computed as $\mathbf{K}_{SS} = [0.02, 0.35, 2.3]^T$, and the switching gains were selected by trial and error methods as $\mathbf{L} = [0.01, 2, 70]^T$. Simulation results for the sliding mode observer show some initial transients during the reaching phase, but attain and remain close to the sliding surface after about 0.25 seconds. The comparisons of estimation errors in the lower three graphs demonstrate the improved performance resulting from the sliding mode method. Note the chattering in velocity and acceleration errors which is characteristic of the sliding motion.

These results are encouraging, however, the sampling frequency in the above simulation is quite high. In the hierarchical approach to image analysis introduced in Section 1.1,

Level IV was concerned with tracking feature points in the image plane. Figure 4.5 shows a comparison of simulation results for the Kalman filter and the simple sliding mode observer when applied to the problem of tracking feature point p^6 (see Figure 2.4) during the trajectory of Case 1 (previous results for object tracking are shown in Figures 2.10-2.13). Independent Kalman filters and sliding mode observers are used in the horizontal and vertical directions, with three states (position, velocity, and acceleration) each, where only position is measured. In this case, the sample frequency is 10Hz, $\beta = 0.5$, $\sigma = 6$, and $\mathbf{R} = 0.01$ for the Kalman filter which gives steady-state gains $\mathbf{K}_{SS} = [0.85, 7.57, 32.65]^T$. The switching gains were again selected through trial and error methods as $\mathbf{L} = [0.0001, 0.001, 0.1]^T$. Results in Figure 4.5 are shown for no measurement noise. The top two graphs show true and estimated trajectories, but differences in performance are not apparent on this scale. Plots of the estimation errors, however, demonstrate that the sliding mode method locks onto and tracks the true trajectory very reliably. An important aspect of tracking feature points is the capability to reliably predict where the feature point will appear in the next image frame. Prediction estimation errors demonstrate improved performance for the sliding mode observer. The lower two graphs show the reconstructed trajectory of the feature point in the image plane.

Level V of the hierarchical system was concerned with tracking feature points in three dimensions from measurements taken in two or more cameras. The single-frame initialization algorithm, presented in Section 2.5, outlined a method to recover three dimensional positions of feature points from two dimensional image plane positions using two or more cameras. Figure 4.6 shows a comparison of simulation results for the Kalman filter and the simple sliding mode observer when applied to the problem of tracking feature point p^6 in three dimensions during the trajectory of Case 1. The 'measurements' for these simulations were obtained from Step 1 of the single-frame initialization algorithm. Independent Kalman filters and sliding mode observers are used in each of the 3D coordinates with three states (position, velocity, and acceleration), where only position is measured. The sample frequency is 10Hz, $\beta = 0.5$, $\sigma = 6$, and $\mathbf{R} = 1$ for the Kalman filter which gives steady-state gains $\mathbf{K}_{SS} = [0.66, 3.56, 9.15]^T$. The switching gains were again selected through trial and error methods as $\mathbf{L} = [0.001, 0.5, 2]^T$. Results in Figure 4.6 are shown for no measurement noise. The top two graphs show the true position and velocity of each of the

three coordinates for this trajectory. Estimation errors are shown in the lower six graphs. In this case, the sliding mode method gives significant performance improvements over the Kalman filter.

Figure 4.4: Comparison of discrete-time Kalman filter and discrete-time sliding mode observer.

Figure 4.5: Tracking feature points in the image plane, Kalman filter vs sliding mode observer.

Figure 4.6: Tracking 3D feature points, Kalman filter vs sliding mode observer.

Although these results are encouraging, there are several problems with analysis of the observer in its present form and extensions to the case of modelling errors and nonlinear systems. The issue of stability is treated in Chapter 5 where new methods are proposed to obtain exponential stability of uncertain discrete-time systems. The remainder of this Chapter focusses on difficulties with extensions of this method to more general systems, and on obtaining a better understanding of properties required by a discrete-time observer in order for the observer error to exhibit convergent quasi-sliding or discrete-time sliding behavior on a prescribed surface.

To facilitate a discussion of difficulties associated with analysis of nonlinear systems with unmodelled dynamics, consider the system

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k)), \quad z(k) = \mathbf{C}\mathbf{x}(k), \quad (4.25)$$

where the scalar measurement model is still assumed to be linear. The observer, corresponding to the same form as in (4.2), is written as

$$\hat{\mathbf{x}}(k+1) = \hat{\mathbf{f}}(\hat{\mathbf{x}}(k)) + \mathbf{K}\eta(k+1) + \mathbf{L}\operatorname{sgn}(s_k), \quad (4.26)$$

where $\hat{\mathbf{f}}(\cdot)$ is an estimate of $\mathbf{f}(\cdot)$, and the innovation sequence (prediction error), $\eta(k)$, is defined as $\eta(k+1) := z(k+1) - \mathbf{C}\hat{\mathbf{f}}(\hat{\mathbf{x}}(k))$. With $\mathbf{e}(k) := \mathbf{x}(k) - \hat{\mathbf{x}}(k)$, the error dynamics are given by

$$\mathbf{e}(k+1) = [\mathbf{I}_n - \mathbf{K}\mathbf{C}] \{ \mathbf{f}(\mathbf{x}(k)) - \hat{\mathbf{f}}(\hat{\mathbf{x}}(k)) \} - \mathbf{L}\operatorname{sgn}(s_k). \quad (4.27)$$

If switching is based on the prediction error,

$$s_k = z(k+1) - \mathbf{C}\hat{\mathbf{f}}[\hat{\mathbf{x}}(k)], \quad (4.28)$$

then s_k can no longer, generally, be written solely as a function of the error $\mathbf{e}(k)$ since

$$s_k = \mathbf{C}\tilde{\mathbf{f}}(\mathbf{x}(k), \hat{\mathbf{x}}(k)), \quad (4.29)$$

where

$$\tilde{\mathbf{f}}(\mathbf{x}(k), \hat{\mathbf{x}}(k)) := \mathbf{f}(\mathbf{x}(k)) - \hat{\mathbf{f}}(\hat{\mathbf{x}}(k)). \quad (4.30)$$

More importantly, the surface $s_k = 0$ depends on the unknown modelling error, which renders generation and analysis of quasi-sliding motion very difficult except in limited cases. If, on the other hand, switching is based on the current error,

$$s_k = z(k) - C\hat{x}(k) = Ce(k), \quad (4.31)$$

then existence of a quasi-sliding regime requires that

$$s_k(s_{k+1} - s_k) = e^T(k)C^T \{C[I_n - KC]\tilde{f}(x(k), \hat{x}(k)) - CL\text{sgn}(s_k)\} < 0. \quad (4.32)$$

in some neighbourhood of the switching surface. If bounds can be placed on the unknown elements of $\tilde{f}$, then in some cases it is possible to select a vector of switching gains, L , to create a quasi-sliding regime. Even if this is possible, however, further analysis is required to demonstrate convergence to within some bounds of the switching surface. Numerous attempts, during the course of this research, to gain useful analytical results for more general systems with this form of the observer and using standard Lyapunov stability methods have not been successful. The primary source of difficulty with this observer lies in the use of a discontinuous switching term.

Returning to the linear single measurement system without modelling errors,

$$x(k+1) = \Phi x(k) \quad z(k) = Cx(k), \quad (4.33)$$

and expressing the observer in the form

$$\hat{x}(k+1) = \Phi\hat{x}(k) + K\eta(k+1) + Lr(e(k)), \quad (4.34)$$

for which the error dynamics satisfy

$$e(k+1) = \Phi_e e(k) - Lr(e(k)). \quad (4.35)$$

a natural question concerns the conditions that must be satisfied by the scalar input $v(e(k))$ in order to generate sliding behavior on a subset $\mathcal{S}$ of the surface $S := \{e \in \mathbb{R}^n \mid s(e) := Me = 0\}$. Of primary concern is whether or not a convergent quasi-sliding regime can exist even locally on $\mathcal{S} \subset S$ when $r(e)$ is discontinuous across $\mathcal{S}$. The following theorem shows that this is not possible.

Theorem 4.2: For the system (4.33) and observer (4.34) with error dynamics (4.35), assume that $\exists L$ such that $ML \neq 0$. Then a convergent quasi-sliding regime exists on a set $S \subset \mathcal{S}$ only if $v(e)$ is continuous across S and for any $e^* \in S$,

$$v(e^*) = (ML)^{-1} M\Phi_o e^*. \quad (4.36)$$

Proof: Define, in accordance with the error dynamics of (4.35), $\forall e \in \mathbb{R}^n$,

$$\Gamma(e) := \Phi_o e - Lv(e). \quad (4.37)$$

Suppose a convergent quasi-sliding regime exists on $S \subset \mathcal{S}$. Then, from Definition 4.1, there exists an open neighbourhood $\mathcal{U}$ of S such that $\forall e \in \mathcal{U} - S$, $|s(\Gamma(e))| < |s(e)|$. As shown by Sarpturk [19], this condition is equivalent to the two inequalities

$$s(e) \{s(\Gamma(e)) - s(e)\} < 0 \quad \text{and} \quad (4.38)$$

$$s(e) \{s(\Gamma(e)) + s(e)\} > 0. \quad (4.39)$$

With straightforward manipulations, these conditions can be reduced, with the assumption¹ that $ML > 0$, to

$$(ML)^{-1} M\Phi_o e - |s(e)| < v(e) < (ML)^{-1} M\Phi_o e + |s(e)|. \quad (4.40)$$

For any point $e^* \in S$, let $\{e_j; j = 0, 1, 2, \dots\} \subset \mathcal{U} - S$ be any sequence contained within $\mathcal{U} - S$ such that $\lim_{j \rightarrow \infty} e_j = e^*$. Then

$$\lim_{j \rightarrow \infty} \{(ML)^{-1} M\Phi_o e_j - |s(e_j)|\} \leq \lim_{j \rightarrow \infty} v(e_j) \leq \lim_{j \rightarrow \infty} \{(ML)^{-1} M\Phi_o e_j + |s(e_j)|\}, \quad (4.41)$$

and continuity of both the left- and right-hand sides, and in particular the continuity of $s(\cdot)$ which gives

$$\lim_{j \rightarrow \infty} |s(e_j)| = |s(e^*)| = 0, \quad (4.42)$$

imply that

$$\lim_{j \rightarrow \infty} v(e_j) = (ML)^{-1} M\Phi_o e^*. \quad (4.43)$$

¹In the case $ML < 0$, the inequalities in (4.40) are reversed, but the result in (4.43) remains unchanged.

The existence of a convergent quasi-sliding regime on S also implies, from Definition 4.1, that $\forall e^* \in S$, $M\Gamma(e^*) = 0$. But then

$$\begin{aligned} M\Gamma(e^*) &= M\Phi_o e - MLv(e^*) = 0 \\ \Rightarrow v(e^*) &= (ML)^{-1} M\Phi_o e^*. \end{aligned} \quad (4.44)$$

Hence, for any point $e^* \in S$ and any sequence $\{e_j\} \subset \mathcal{U} - S$ with $\lim_{j \rightarrow \infty} e_j = e^*$,

$$\lim_{j \rightarrow \infty} v(e_j) = v(e^*), \quad (4.45)$$

which implies (see for example [144, p. 80]) that $v(e)$ is continuous across S . $\square$

An important result of the above Theorem is that for this linear system, the input $v(e)$ during pure sliding must be, for any $e^* \in S$,

$$v(e^*) = (ML)^{-1} M\Phi_o e^*. \quad (4.46)$$

and the error dynamics during pure sliding are given by

$$e^*(k+1) = F\Phi_o e^*(k), \quad (4.47)$$

where

$$F := [I_n - L(ML)^{-1}M]. \quad (4.48)$$

The matrix F represents, as was previously discussed in Section 3.5 in the context of robustness of continuous-time sliding mode systems, a projection along $\text{span}\{L\}$ onto the null space of M . This projection operation is consistent with the robustness mechanism illustrated in Figure 3.2, for the continuous-time control problem, and in Figure 3.4, for the continuous-time observer problem. In fact, with this form for $v(e)$, a global discrete-time sliding regime can be generated on all of S .

Corollary 4.1: For the system (4.33) and observer (4.34) with error dynamics (4.35), if $\exists L$ such that $ML \neq 0$ then the input

$$v(e) = (ML)^{-1} M\Phi_o e \quad (4.49)$$

generates a global discrete-time sliding regime on $S := \{e \in \mathbb{R}^n \mid s(e) := Me = 0\}$.

Proof: Under the stated hypothesis, $\forall e(0) \in \mathbb{R}^n$, and $\forall k \geq 0$,

$$s_{k+1} = Me(k+1) = MF\Phi_e e(k) = 0. \quad (4.50)$$

□

The above results are intuitive in the following sense. Without the term $Lv(e)$, the error system starting from point $e(k)$ at time $k \geq 0$ is taken to point $\Phi_e e(k)$ at time $k+1$. The corrective action, $Lv(e(k))$, operates in $\text{span}\{L\}$. If the error at time $k+1$ is to be driven to the surface, i.e. to the null space of M , by inclusion of the term $Lv(e(k))$, then necessarily the point $\Phi_e e(k)$ must be projected along $\text{span}\{L\}$ into the null space of M .

As previously discussed, M must be selected so that $v(e(k))$ can be computed from the available measurements at time $k+1$ and previous state estimates. Expanding $v(e(k))$ from (4.49) as

$$v(e(k)) = (ML)^{-1} \{Mx(k+1) - MKz(k+1) - M[I_n - KC]\Phi\hat{x}(k)\} \quad (4.51)$$

illustrates that $Mx(k+1)$ must be available. Taking $M := C$ gives $Mx(k+1) = z(k+1)$ and

$$v(e(k)) = (CL)^{-1} [1 - KC] \{z(k+1) - C\Phi\hat{x}(k)\} \quad (4.52)$$

which is a suitable result provided $CL \neq 0$. However, taking $M := C\Phi$ so that the surface depends on the prediction error will require a one-sample delay because in this case $Mx(k+1) = z(k+2)$. The suitability of this selection of M would depend on the application. The consequences of defining the surface in terms of the prediction error in the presence of modelling errors have already been discussed.

In their comparative study of observers for continuous-time linear systems, Dorling and Zinober [15] demonstrated that the sliding dynamics of a sliding mode observer correspond to the dynamics of a reduced-order Luenberger observer. The same is true for the linear discrete-time system considered above when the input $v(e)$ given in (4.49) is employed and $M = C$, as illustrated in the following example.

Example 4.3: Consider the linear discrete-time system $x(k+1) = \Phi x(k)$ with scalar measurements $z(k) = Cx(k)$ where $x \in \mathbb{R}^n$, $z \in \mathbb{R}$, and $C = [1, 0_{1 \times (n-1)}]$ for which the

state vector is partitioned as

$$\mathbf{x}(k) = \begin{bmatrix} x_1(k) \\ \mathbf{x}_2(k) \end{bmatrix} = \begin{bmatrix} z(k) \\ \mathbf{x}_2(k) \end{bmatrix}, \quad (4.53)$$

where $x_1 \in \mathbb{R}$ and $\mathbf{x}_2 \in \mathbb{R}^{n-1}$. The corresponding partition of the state transition matrix is written as

$$\Phi = \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix}. \quad (4.54)$$

Then $x_1(k) := z(k)$ is measured directly and the dynamics of unmeasured states is given by

$$\mathbf{x}_2(k) = \Phi_{22}\mathbf{x}_2(k-1) + \underbrace{\Phi_{21}x_1(k-1)}_{\text{known "input"}}. \quad (4.55)$$

A corresponding measurement model is

$$z(k+1) - \Phi_{11}z(k) = \underbrace{x_1(k+1) - \Phi_{11}x_1(k)}_{\text{known "measurements"}} = \Phi_{12}\mathbf{x}_2(k). \quad (4.56)$$

where the known "measurements" above are available when $z(k+1)$ is available. The reduced-order Luenberger observer then takes the form [143, p. 261]

$$\begin{aligned} \hat{\mathbf{x}}_2(k+1) &= \Phi_{22}\hat{\mathbf{x}}_2(k) + \Phi_{21}x_1(k-1) \\ &\quad + \mathbf{K}_r \{x_1(k+1) - \Phi_{11}x_1(k) - \Phi_{12}\mathbf{x}_2(k)\}, \end{aligned} \quad (4.57)$$

and $\mathbf{K}_r \in \mathbb{R}^{n-1}$ is the vector of reduced-order observer gains. The error dynamics for $\mathbf{e}_2(k) := \mathbf{x}_2(k) - \hat{\mathbf{x}}_2(k)$ are

$$\mathbf{e}_2(k+1) = [\Phi_{22} - \mathbf{K}_r\Phi_{12}]\mathbf{e}_2(k). \quad (4.58)$$

Provided (Φ_{22}, Φ_{12}) is an observable pair, the gain vector $\mathbf{K}_r$ can be selected using Ackermann's synthesis formula.

For the sliding mode observer in (4.34) with

$$v(\mathbf{e}) = (\mathbf{M}\mathbf{L})^{-1}\mathbf{M}\Phi_o\mathbf{e} \quad (4.59)$$

and taking $\mathbf{M} := \mathbf{C}$, the error dynamics are given by

$$\mathbf{e}(k+1) = \mathbf{F}\Phi_o\mathbf{e}(k), \quad (4.60)$$

where $\mathbf{F} := [\mathbf{I}_n - \mathbf{L}(\mathbf{C}\mathbf{L})^{-1}\mathbf{C}]$, and $\Phi_o := [\mathbf{I}_n - \mathbf{K}\mathbf{C}]\Phi$. Define partitions of $\mathbf{K}$ and $\mathbf{L}$ as

$$\mathbf{K} := \begin{bmatrix} k_1 \\ \mathbf{K}_2 \end{bmatrix}, \quad \mathbf{L} := \begin{bmatrix} \ell_1 \\ \mathbf{L}_2 \end{bmatrix}, \quad (4.61)$$

where k_1 and ℓ_1 are scalars, and $\mathbf{K}_2$ and $\mathbf{L}_2$ are $(n-1) \times 1$ column vectors. Partitioning the error vector as $\mathbf{e}(k) := [e_1(k), \mathbf{e}_2^T(k)]^T$, algebraic manipulation and simplification of (4.60) gives the error dynamics as

$$e_1(k+1) = e_1(k) = 0, \quad (4.62)$$

$$\mathbf{e}_2(k+1) = [\Phi_{22} - \mathbf{K}_r \Phi_{12}] \mathbf{e}_2(k), \quad (4.63)$$

where

$$\mathbf{K}_r := \frac{1-k_1}{\ell_1} \mathbf{L}_2 + \mathbf{K}_2. \quad (4.64)$$

Hence the error dynamics during sliding take on the same form as the error dynamics of the reduced-order discrete-time Luenberger observer in (4.58). $\square$

From the above example, the reduced order Luenberger observer for linear systems is identified as one for which the error dynamics exhibit a global discrete-time sliding regime on the surface $\mathbf{C}\mathbf{e} = 0$. Moreover, from the above example and the results of Theorem 4.2, any observer of the form (4.34) for the linear system (4.33) which possesses a convergent quasi-sliding regime on $\mathbf{C}\mathbf{e} = 0$ must have error dynamics during pure sliding that correspond to those of a reduced-order Luenberger observer.

4.4 Chapter Summary

The simple discrete-time sliding mode observer considered in this Chapter can be viewed as a direct application, to the discrete-time case, of continuous-time sliding mode methods for linear systems. Generation of quasi-sliding motion with discontinuous switching based on available measurements can, under fairly mild assumptions, always be achieved with this method for linear systems. Discontinuous switching, however, can give rise to steady-state limit cycles and cannot, at least for the linear system, yield a convergent quasi-sliding regime. Simulation results from the sliding mode method are very encouraging, however,

direct applications of this method to more general nonlinear systems with modelling errors are not apparent. An alternate form of observer for linear systems was derived for which a global discrete-time sliding regime is generated. This form of the observer has sliding dynamics equivalent to conventional reduced-order Luenberger observer for the discrete-time linear system.

A primary objective of Chapters 6 and 7 is to extend the theory of reduced-order observers from the strictly linear case to the case of nonlinear uncertain dynamics (Chapter 6) and nonlinear measurement models (Chapter 7) using the sliding mode concept. The following Chapter focusses first on stability issues for discrete-time, nonlinear, uncertain systems with applications in observer design, and then proposes a reaching law for discrete-time sliding mode system design. These stability results are required repeatedly in analysis of proposed observer forms in Chapters 6 and 7.

Chapter 5

Exponential Stability of Discrete-time Systems

5.1 Introduction

In the previous Chapter, a simple discrete-time sliding mode observer was proposed for linear systems, but the important issue of stability of the observer was not considered. In the observer problem, the existence of a quasi-sliding, convergent quasi-sliding, or even discrete-time sliding regime on a prescribed surface does not necessarily imply stability of the observer. The reason for this is that the sliding surface in the observer problem is often defined in terms of measurements available at a single time instant, and these measurements are related to some but not all states of the system at that time instant. As a result, even if the error trajectories can be maintained on the prescribed surface, there generally exists reduced-order dynamics on the surface that must be checked for stability.

Theory and applications of quadratic form Lyapunov functions in stability analysis are very familiar in the control literature. Piecewise-linear Lyapunov functions [145, 146, 147, 148] have also been shown to have useful properties, particularly in the construction of domains of asymptotic stability [146] or positively invariant sets [147]. More recently, Kiendl *et al.* [149] proposed a unified theory of quadratic and piecewise-linear Lyapunov functions through the use of vector norms, and derived necessary and sufficient conditions for a vector norm to be a Lyapunov function for linear, time invariant systems.

This Chapter presents new results, beyond those in [25, 26], on exponential stability of discrete-time systems, construction of domains of exponential stability, and estimation of exponential bounds on the system state through the use of vector norm Lyapunov, or at least

Lyapunov-like functions. The basis for this approach is to search for a transformation of the state so that the new system, expressed in the transformed coordinates, is contractive in some vector norm. Stability results for the original system can then be inferred from those of the transformed system.

Notation and basic notions of stability for discrete-time systems are introduced in Section 5.2. Section 5.3 presents background material and preliminary results relevant to the theory developed in remaining sections. Building initially on the methods of Kiendl *et al.* [149] and, in a deterministic sense, Tarn and Rasis [82], necessary and sufficient conditions are given in Section 5.4 for exponential stability of linear time invariant and time varying systems, and exponential bounds are derived. For nonlinear systems, sufficient conditions are obtained in Section 5.5 for local exponential asymptotic stability. These conditions also yield a domain of exponential stability and exponential bounds on the state. Section 5.6 demonstrates applications of the theory in observer design for uncertain discrete-time nonlinear systems for which emphasis is placed on sufficient conditions for exponentially bounded observer errors, and estimation of a domain of exponential stability and reasonably tight error bounds. Section 5.7 proposes a reaching law method for design of discrete-time sliding mode observers and applies the stability results to analysis of the reaching dynamics.

5.2 Stability Notions and Definitions

This Section is devoted to establishing notation and several fundamental definitions which are used throughout this Chapter, and to a review of the various notions of stability for discrete-time systems.

For $\mathbf{A} \in \mathbb{C}^{m \times n}$, the set of singular values is $\sigma(\mathbf{A}) := \{\sigma_i(\mathbf{A}), i = 1, 2, \dots, \min(m, n)\}$, and the maximum and minimum singular values of $\mathbf{A}$ are $\bar{\sigma}(\mathbf{A})$ and $\underline{\sigma}(\mathbf{A})$, respectively. When $m = n$, the set of eigenvalues of $\mathbf{A}$ is denoted as $\lambda(\mathbf{A}) := \{\lambda_i(\mathbf{A}), i = 1, 2, \dots, n\}$, and the spectral radius is $\rho(\mathbf{A}) := \max\{|\lambda_i(\mathbf{A})|\}$.

For $\mathbf{y} \in \mathbb{C}^n$, $\mathbf{A} \in \mathbb{C}^{m \times n}$, and $\mathbf{B} \in \mathbb{R}^{n \times q}$, $\|\mathbf{y}\|$ represents a vector norm, while $\|\mathbf{A}\|$

denotes a matrix norm which is both submultiplicative,

$$\|\mathbf{AB}\| \leq \|\mathbf{A}\| \|\mathbf{B}\| \quad (5.1)$$

and compatible with the vector norm,

$$\|\mathbf{Ay}\| \leq \|\mathbf{A}\| \|\mathbf{y}\|. \quad (5.2)$$

The Hölder vector norms are written as

$$\|\mathbf{y}\|_p := \left[\sum_i^n |y_i|^p \right]^{1/p}. \quad (5.3)$$

for $p \geq 1$, the corresponding induced matrix norms are defined as

$$\|\mathbf{A}\|_p := \sup_{\|\mathbf{y}\|_p=1} \|\mathbf{Ay}\|_p. \quad (5.4)$$

and the Frobenius matrix norm is written as

$$\|\mathbf{A}\|_F := \left[\sum_{i,j} |a_{ij}|^2 \right]^{1/2}. \quad (5.5)$$

A thorough treatment of vector and matrix norms is provided by Horn and Johnson [150, Chap. 5]. For square nonsingular matrices $\mathbf{A}$, the condition number with respect to the prescribed matrix norm $\|\cdot\|$ is written as $\kappa(\mathbf{A}) := \|\mathbf{A}\| \|\mathbf{A}^{-1}\|$.

For deterministic analysis the following concepts are based on discrete-time stability properties [151, 152, 153, 154, 155] for a system of the form

$$\mathbf{x}(k+1) = \mathbf{f}(k, \mathbf{x}(k)) \quad (5.6)$$

in which the origin is assumed to be a stationary or fixed point, $\mathbf{f}(k, \mathbf{0}) = \mathbf{0}$, $\forall k \geq 0$. If the fixed point is not at the origin, but at $\mathbf{x}^* \neq \mathbf{0}$, then a change of coordinates $\mathbf{y}(k) = \mathbf{x}(k) - \mathbf{x}^*$ translates the stationary point to the origin according to

$$\begin{aligned} \mathbf{y}(k+1) &= \mathbf{x}(k+1) - \mathbf{x}^* \\ &= \mathbf{f}(k, \mathbf{y}(k) + \mathbf{x}^*) - \mathbf{f}(k, \mathbf{x}^*) \\ &:= \mathbf{g}(k, \mathbf{y}(k)), \end{aligned} \quad (5.7)$$

where $\mathbf{g}(k, \mathbf{0}) \equiv \mathbf{0}$.

Definition 5.1: With respect to the prescribed vector norm $\|\cdot\|$, let $B(y, \delta)$, $\delta > 0$, be the open set, $B(y, \delta) := \{\mathbf{x} \mid \|\mathbf{y} - \mathbf{x}\| < \delta\}$. Define also $B_\delta := B(0, \delta)$. $\square$

Definition 5.2: For system (5.6) the zero state, $\mathbf{x} = 0$, is said to be

1. *Stable* if $\forall \epsilon > 0$ and $\forall k_0 \geq 0$, $\exists \delta(\epsilon, k_0)$ such that

$$\mathbf{x}(k_0) \in B_\delta \Rightarrow \mathbf{x}(k) \in B_\epsilon \quad \forall k \geq k_0;$$

2. *Uniformly stable* if it is stable and δ can be chosen independent of k_0 ;
3. *Attractive* if $\forall k_0 \geq 0$, $\exists \delta(k_0) > 0$ such that $\mathbf{x}(k_0) \in B_\delta \Rightarrow \lim_{k \rightarrow \infty} \mathbf{x}(k) = 0$;
4. *Uniformly attractive* if it is attractive and δ can be chosen independent of k_0 ;
5. *Asymptotically stable* if it is stable and attractive;
6. *Uniformly Asymptotically stable* if it is uniformly stable and uniformly attractive;
7. *Exponentially stable* if $\forall k_0 \geq 0$, $\exists \delta(k_0) > 0$ and constants $M > 0$, $\alpha > 0$ and $\beta \in [0, 1)$, and a vector norm $\|\cdot\|$ such that

$$\mathbf{x}(k_0) \in B_\delta \Rightarrow \|\mathbf{x}(k)\| \leq M + \alpha \|\mathbf{x}(k_0)\| \beta^{k-k_0} \quad \forall k \geq k_0;$$

8. *Exponentially asymptotically stable* if it is exponentially stable and $M = 0$;
9. *Uniformly exponentially stable* if it is exponentially stable and δ can be chosen independent of k_0 ;
10. *Uniformly exponentially asymptotically stable* if it is exponentially asymptotically stable and δ can be chosen independent of k_0 ; $\square$

All of the above concepts are local in the sense that they are dependent on the existence of a scalar $\delta > 0$ such that once $\mathbf{x}(k_0) \in B_\delta$ for some k_0 , the behavior of state can be predicted from time k_0 onwards. If δ can be chosen as infinite, the above properties are modified by the quantifier *global* to obtain, for example, global exponential stability or global exponential asymptotic stability, etc. Global stability results, however, are usually

only available for linear systems. Uniform asymptotic stability and exponential asymptotic stability are in fact equivalent for linear systems [154, Theorem 4.2.2], but this is generally not the case for nonlinear systems. In the nonlinear case, the initial state plays an important role in stability analysis and hence it is imperative to obtain a measure of admissible initial states under which the desired stability properties are guaranteed. The set of initial states for which the system, when initialized within this set, is stable is referred to as a *stability domain* for the system. Appropriate modifiers are also used to identify the nature of stability exhibited by the system; for example a domain of exponential stability, or a domain of exponential asymptotic stability, etc.

Two other definitions are required for the following sections.

Definition 5.3: A continuous, strictly increasing function $a : \mathbb{R}^+ \mapsto \mathbb{R}^+$ with $a(0) = 0$ is called a *class- $\mathcal{K}$ function* and is denoted $a \in \mathcal{K}$. $\square$

Definition 5.4: A function $v : \mathbb{N} \times \mathbb{R}^n \mapsto \mathbb{R}$ is

1. *positive definite* on $\mathbb{R}^n$ if $\forall k \geq 0 \ v(k, \mathbf{0}) = 0$, and there exists a function $a \in \mathcal{K}$, such that for all $k \in \mathbb{N}$, and all $\mathbf{x} \in \mathbb{R}^n$, $v(k, \mathbf{x}) \geq a(\|\mathbf{x}\|)$;
2. *negative definite* on $\mathbb{R}^n$ if $\forall k \geq 0 \ v(k, \mathbf{0}) = 0$, and there exists a function $a \in \mathcal{K}$, such that for all $k \in \mathbb{N}$, and all $\mathbf{x} \in \mathbb{R}^n$, $v(k, \mathbf{x}) \leq -a(\|\mathbf{x}\|)$;
3. *decreascent* if there exists a function $b \in \mathcal{K}$ such that for all $k \in \mathbb{N}$, and all $\mathbf{x} \in \mathbb{R}^n$, $v(k, \mathbf{x}) \leq b(\|\mathbf{x}\|)$. $\square$

5.3 Background and Preliminary Results

For nonlinear discrete-time systems of the form

$$\mathbf{x}(k+1) = \mathbf{f}(k, \mathbf{x}(k)), \quad (5.8)$$

where $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{f} : \mathbb{N} \times \mathbb{R}^n \mapsto \mathbb{R}^n$, $\mathbf{f}$ is C^p in $\mathbf{x}$, $p \geq 1$, and $\mathbf{f}(k, \mathbf{0}) = \mathbf{0}$, the following well-known Lyapunov stability theorem [151, 154, 156, 157] was originated by Hahn [158].

Theorem 5.1: For system (5.8), if there exists a *Lyapunov function*, $V_k := V(k, \mathbf{x}(k))$, $V : \mathbb{N} \times \mathbb{R}^n \mapsto \mathbb{R}$, which is positive definite, radially unbounded, continuous in its second argument, and $\Delta V_k := V_{k+1} - V_k$ is negative definite, then the equilibrium $\mathbf{x} = 0$ is asymptotically stable. If, in addition, V_k is decrescent, stability is uniform in k . If ΔV_k is negative semi-definite, system (5.8) is stable and is asymptotically stable provided ΔV_k does not vanish identically along any solution of (5.8) other than the null solution. $\square$

The above Theorem results in what is commonly known as Lyapunov's second method for discrete-time systems [157]. The primary advantage of this technique is that the solution of the system of difference equations is not required in order to obtain the stability result. The positive definite function $V(k, \mathbf{x})$ is often considered as a generalized perhaps time-varying measure of energy or distance relative to that of the stationary point. Direct application of this Theorem to obtain bounds on systems which are not strictly asymptotically stable, however, often does not lead to useful results as demonstrated in the following example.

Example 5.1: Consider again the linear system $\mathbf{x}(k+1) = \Phi \mathbf{x}(k)$, with scalar measurements $z(k) = \mathbf{C}\mathbf{x}(k)$, for which $(\mathbf{C}, \Phi)$ is an observable pair. The simple sliding mode observer of the previous chapter has error dynamics

$$\mathbf{e}(k+1) = \Phi_o \mathbf{e}(k) - \mathbf{L} \operatorname{sgn}(s_k), \quad (5.9)$$

where $\Phi_o := [\mathbf{I}_n - \mathbf{K}\mathbf{C}]$ and $s_k := \mathbf{M}\mathbf{e}(k)$. Assuming that $\mathbf{K}$ has been chosen [143, p. 255] so that the modulus of all eigenvalues of Φ_o are less than unity, then given any positive definite symmetric $\mathbf{Q}$, there exists [14, p. 179] a unique positive definite symmetric $\mathbf{P}$ such that

$$\mathbf{P} - \Phi_o^T \mathbf{P} \Phi_o = \mathbf{Q}. \quad (5.10)$$

Consider the Lyapunov function candidate

$$V_k := V(\mathbf{e}(k)) = \mathbf{e}^T(k) \mathbf{P} \mathbf{e}(k). \quad (5.11)$$

The first forward differences of V_k along trajectories of the error system satisfy

$$\begin{aligned} \Delta V_k &:= V_{k+1} - V_k \\ &= -\mathbf{e}^T(k) \mathbf{Q} \mathbf{e}(k) - 2 \operatorname{sgn}(s_k) \mathbf{L}^T \mathbf{P} \Phi_o \mathbf{e}(k) + \mathbf{L}^T \mathbf{P} \mathbf{L} \end{aligned} \quad (5.12)$$

For any $\mathbf{y} \in \mathbb{R}^n$, nonsingular symmetric $\mathbf{R} \in \mathbb{R}^{n \times n}$, $\mathbf{v} \in \mathbb{R}^m$, and $\mathbf{B} \in \mathbb{R}^{m \times n}$, the multidimensional analog of completing the square can be written as

$$\begin{aligned} \mathbf{y}^T \mathbf{R} \mathbf{y} + \mathbf{v}^T \mathbf{B} \mathbf{y} &= \left(\mathbf{y} + \frac{\mathbf{R}^{-1} \mathbf{B}^T \mathbf{v}}{2} \right)^T \mathbf{R} \left(\mathbf{y} + \frac{\mathbf{R}^{-1} \mathbf{B}^T \mathbf{v}}{2} \right) \\ &\quad - \frac{\mathbf{v}^T \mathbf{B} \mathbf{R}^{-1} \mathbf{B}^T \mathbf{v}}{4}. \end{aligned} \quad (5.13)$$

Applying this result to (5.12) gives

$$\begin{aligned} \Delta V_k &= - \left(\mathbf{e}(k) + \text{sgn}(s_k) \mathbf{Q}^{-1} \Phi_o^T \mathbf{P} \mathbf{L} \right)^T \mathbf{Q} \left(\mathbf{e}(k) + \text{sgn}(s_k) \mathbf{Q}^{-1} \Phi_o^T \mathbf{P} \mathbf{L} \right) \\ &\quad + \mathbf{L}^T \mathbf{P} \mathbf{L} + \mathbf{L}^T \mathbf{P} \Phi_o \mathbf{Q}^{-1} \Phi_o^T \mathbf{P} \mathbf{L}. \end{aligned} \quad (5.14)$$

The region of error space in which $V_{k+1} - V_k < 0$ is outside of a closed volume bounded by partial surfaces of two hyperellipsoids with centres at $\pm \mathbf{e}_c$, where

$$\mathbf{e}_c := \mathbf{Q}^{-1} \Phi_o^T \mathbf{P} \mathbf{L}, \quad (5.15)$$

and both hyperellipsoids have semi-major axis length ℓ_c , where

$$\ell_c := \left[\frac{\mathbf{L}^T \mathbf{P} \mathbf{L} + \mathbf{L}^T \mathbf{P} \Phi_o \mathbf{Q}^{-1} \Phi_o^T \mathbf{P} \mathbf{L}}{\sigma_{\min}(\mathbf{Q})} \right]^{\frac{1}{2}}. \quad (5.16)$$

Define

$$\begin{aligned} \mathcal{E}^- &:= \{ \mathbf{e} \in \mathbb{R}^n \mid \Delta V(\mathbf{e}) < 0 \} \\ \mathcal{E} &:= \{ \mathbf{e} \in \mathbb{R}^n \mid \Delta V(\mathbf{e}) = 0 \} \\ \mathcal{E}^+ &:= \{ \mathbf{e} \in \mathbb{R}^n \mid \Delta V(\mathbf{e}) > 0 \}. \end{aligned} \quad (5.17)$$

The set $\mathcal{E}$ is said to be *invariant* with respect to the error dynamics (5.9); for any $k_0 \geq 0$, if $\mathbf{e}(k_0) \in \mathcal{E}$, then $\forall k \geq k_0$, $\mathbf{e}(k) \in \mathcal{E}$. If $\text{sgn}\{\mathbf{M}[\mathbf{I}_n + \Phi_o]^{-1} \mathbf{L}\} = 1$, then a straightforward calculation shows that the two points of the steady state limit cycle

$$\mathbf{e}_\ell = \pm [\mathbf{I}_n + \Phi_o]^{-1} \mathbf{L} \quad (5.18)$$

obtained in Example 4.2 of Section 4.2 lie in $\mathcal{E}$. However, convergence of the error state to the set $\mathcal{E}$ in the above example is apparently not implied from the hypothesis of Theorem 5.1. Intuitively, if the initial error lies in $\mathcal{E}^-$, the positive definite decrescent function

$V(e(k))$ must decrease and either approach $\mathcal{E}$ asymptotically from $\mathcal{E}^-$ or, for some $k_0 > 0$, $e(k_0) \in \mathcal{E}^+ \cup \mathcal{E}$. If $e(k_0) \in \mathcal{E}^+$, then $\Delta V(e(k_0)) > 0$ which means that over the next time step the error could migrate outside of $\mathcal{E}^+$ into $\mathcal{E}^-$. Hence, not enough information is available from this example to conclude that the error is ultimately bounded by the set $\mathcal{E}^+ \cup \mathcal{E}$. $\square$

Another alternative which is sometimes employed to obtain ultimate bounds for solutions of nonlinear difference equations is based on a discrete-time analog of La Salle's invariance principle.

Theorem 5.2: For the system (5.8), suppose that $\mathbf{x}(k) \in \mathcal{D} \subset \mathbb{R}^n$ for all $k \geq 0$, and that there exist two real valued functions $V(k, \mathbf{x})$ and $u(\mathbf{x}) \geq 0$ both continuous in $\mathbf{x}$ with $V(k, \mathbf{x})$ bounded below such that

$$V(k+1, \mathbf{f}(k, \mathbf{x}(k))) - V(k, \mathbf{x}(k)) \leq -u(\mathbf{x}(k)) \quad \forall k \geq k_0. \quad (5.19)$$

Then either $\mathbf{x}(k)$ is unbounded or it approaches the set

$$\mathcal{E} := \{\mathbf{x} \in \overline{\mathcal{D}} \mid u(\mathbf{x}) = 0\}. \quad (5.20)$$

where $\overline{\mathcal{D}}$ represents the closure of $\mathcal{D}$. If, in addition, there exist two continuous scalar-valued functions $u(\mathbf{x})$ and $v(\mathbf{x})$ such that

$$u(\mathbf{x}) \leq V(k, \mathbf{x}) \leq v(\mathbf{x}), \quad \forall k \geq k_0. \quad (5.21)$$

then by fixing $\eta > 0$ and defining the two sets

$$\mathcal{D}(\eta) := \{\mathbf{x} \mid u(\mathbf{x}) < \eta\}, \quad \text{and} \quad (5.22)$$

$$\mathcal{D}_1(\eta) := \{\mathbf{x} \mid v(\mathbf{x}) < \eta\}. \quad (5.23)$$

then $\mathcal{D}_1(\eta) \subset \mathcal{D}(\eta)$ and hypothesis (5.19) with $\mathcal{D} \equiv \mathcal{D}(\eta)$ implies that all solutions starting in $\mathcal{D}_1(\eta)$ remain in $\mathcal{D}(\eta)$ and approach $\mathcal{E}$ as $k \rightarrow \infty$.

Proof: See Lakshmikantham and Trigiante, Theorem 4.8.5 and Corollary 4.8.4 [154, p. 109]. $\square$

In Theorem 5.2, the function V is not required to be positive definite as was the case in Theorem 5.1, but V must still be bounded below. Both Theorems require that the sequence $V_k := V(k, \mathbf{x}(k))$ be strictly decreasing and hence convergent since V is bounded from below. Numerous attempts to apply this theorem, even for the scalar case of dynamics considered in Example 5.1, in order to demonstrate convergence to some bounded set containing the origin have been unsuccessful.

Another related method which deals specifically with the case when ΔV_k may be positive in some bounded set containing the origin, but negative outside of this set, is provided by the following Theorem.

Theorem 5.3: For the system (5.8), suppose that for $\mathbf{x} \in \mathcal{D} \subset \mathbb{R}^n$, there exists two continuous real-valued functions, $V(\mathbf{x})$ and $w(\mathbf{x})$, defined on $\mathcal{D}$ such that, $\forall \mathbf{x} \in \mathcal{D}$, $V(\mathbf{x}) \geq 0$ and

$$\Delta V_k := V(f(k, \mathbf{x})) - V(\mathbf{x}) \leq w(\mathbf{x}) \leq a. \quad (5.24)$$

for some constant $a \geq 0$. Let

$$\mathcal{E}^+ := \{\mathbf{x} \in \overline{\mathcal{D}} \mid w(\mathbf{x}) \geq 0\}. \quad (5.25)$$

$$b := \sup_{\{\mathbf{x} \in \mathcal{E}^+\}} \{V(\mathbf{x})\}, \quad (5.26)$$

$$\mathcal{A} := \{\mathbf{x} \in \overline{\mathcal{D}} \mid V(\mathbf{x}) \leq a + b\}. \quad (5.27)$$

Then every solution, which remains in $\mathcal{D}$ and enters $\mathcal{A}$ for $k = k_0 \geq 0$, remains in $\mathcal{A}$ for $k \geq k_0$. If, in addition,

$$\delta := \sup_{\{\mathbf{x} \in \mathcal{D} - \mathcal{A}\}} \{w(\mathbf{x})\} < 0, \quad (5.28)$$

then each solution of (5.8) which remains in $\mathcal{D}$ enters $\mathcal{A}$ in a finite number of steps.

Proof: See Lakshmikantham and Trigiante, Theorem 4.11.3 and Corollary 4.11.2 [154, p. 122]. $\square$

In Theorem 5.3, V is a function of $\mathbf{x}$ only and is positive definite, but the first forward differences of V need not be negative definite along solutions of the difference equation. The set $\mathcal{E}^+$ in Theorem 5.3, which corresponds to that defined in Example 5.1, lies entirely within the set $\mathcal{A}$ if $a > 0$, which means that the bounds provided need not be very tight. Application of Theorem 5.3 to the scalar case of the problem considered in Example 5.1,

with $\Phi_0 = \alpha$, $0 < \alpha < 1$, $L = \ell > 0$, $M = 1$, $P = 1$, and $Q = 1 - \alpha^2$ yields $\mathcal{E}^+ = [-\ell/(1 + \alpha), \ell/(1 + \alpha)]$, $a = \ell^2$, $b = \ell^2/(1 + \alpha)^2$, and the bounds

$$|e| \leq \ell \sqrt{\frac{\alpha^2 + 2}{\alpha^2 + 1}}. \quad (5.29)$$

The primary issue in applications of the above theoretical results appears to be in the selection of a suitable function $V(k, \mathbf{x}(k))$. Contributions of the following Sections treat this issue in part by taking the Lyapunov, or Lyapunov-like function V as a vector norm applied to a possibly time varying nonlinear transformation of the state. In many cases this approach may provide some guidance in the selection of V for stability analysis. Another deficiency, even for the simple system considered in Example 5.1, is that neither Lyapunov's second method nor the approach based on La Salle's invariance principle nor the application of Theorem 5.3 lead to the result that this system in fact exhibits *exponential* convergence to a bounded set containing the origin—a fact that is demonstrated in Section 5.4.

As illustrated in Example 5.1, and this is especially true in the case of uncertain dynamical systems, it may not be possible to specify a Lyapunov function in the strict sense of Theorem 5.1. The following lemma, which has been adapted from Tarn and Rasis [82, Theorem 2], may provide exponential and ultimate bounds on the state through the use of a Lyapunov-like function.

Lemma 5.1: For system (5.8), if there exist functions $V(k, \mathbf{x}(k))$ and $\phi \in \mathcal{K}$, and constants $c > 0$, $M \geq 0$, and $0 < \alpha < 1$ such that, $\forall k \geq 0$, $V(k, \mathbf{0}) = 0$ and

1. $V(k, \mathbf{x}(k)) \geq c\phi(\|\mathbf{x}(k)\|)$, and
2. $V(k + 1, \mathbf{x}(k + 1)) - V(k, \mathbf{x}(k)) \leq M - \alpha V(k, \mathbf{x}(k))$.

then

$$c\phi(\|\mathbf{x}(k)\|) \leq V(k, \mathbf{x}(k)) \leq (1 - \alpha)^k V(0, \mathbf{x}(0)) + M \sum_{i=0}^{k-1} (1 - \alpha)^i. \quad (5.30)$$

and $\limsup_{k \rightarrow \infty} \phi(\|\mathbf{x}(k)\|) \leq M/c\alpha$.

Proof: Repeated applications of Condition 2 gives

$$V(k, \mathbf{x}(k)) \leq M + (1 - \alpha)V(k - 1, \mathbf{x}(k - 1))$$

$$\begin{aligned}
 &\leq M + (1 - \alpha)M + (1 - \alpha)^2 V(k - 2, \mathbf{x}(k - 2)) \\
 &\vdots \\
 &\leq (1 - \alpha)^k V(0, \mathbf{x}(0)) + M \sum_{i=0}^{k-1} (1 - \alpha)^i.
 \end{aligned} \tag{5.31}$$

Then, since $0 < \alpha < 1$, Condition 1 gives

$$\limsup_{k \rightarrow \infty} \phi(\|\mathbf{x}(k)\|) \leq \limsup_{k \rightarrow \infty} \frac{1}{c} V(k, \mathbf{x}(k)) \leq \frac{M}{c} \sum_{i=0}^{\infty} (1 - \alpha)^i = \frac{M}{c\alpha}. \tag{5.32}$$

□

In Lemma 5.1 the function $V_k := V(k, \mathbf{x}(k))$ is positive definite, as was required in Theorem 5.1, but the forward differences of V_k along trajectories of the system are not required to be negative definite. An important result of the Lemma is that explicit bounds are available on the class- $\mathcal{K}$ function $\phi(\mathbf{x})$. However, applications of this result still require the specification of a suitable function $V(k, \mathbf{x}(k))$ which satisfies the hypothesis.

In convergence analysis of iterative numerical methods, Banach's fixed point theorem, or the contraction mapping theorem, often plays a critical role.

Definition 5.5: A mapping $\mathbf{F} : \mathcal{D} \subset \mathbb{R}^n \mapsto \mathbb{R}^n$ is *contractive* on a set $\mathcal{D}_0 \subset \mathcal{D}$ with respect to the vector norm $\|\cdot\|$ if $\exists \beta, 0 \leq \beta < 1$, such that

$$\|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})\| \leq \beta \|\mathbf{x} - \mathbf{y}\| \tag{5.33}$$

for all $\mathbf{x}, \mathbf{y} \in \mathcal{D}_0$.

□

Note that the contractive property is norm dependent, so that a mapping may be contractive in one norm on $\mathbb{R}^n$, but not in another.

Theorem 5.4: Contraction Mapping Theorem. Suppose that $\mathbf{F} : \mathcal{D} \subset \mathbb{R}^n \mapsto \mathbb{R}^n$ maps a closed set $\mathcal{D}_0 \subset \mathcal{D}$ into itself and that $\mathbf{F}$ is a contraction mapping on $\mathcal{D}_0$. Then $\mathbf{F}$ has a unique fixed point $\mathbf{x}^* \in \mathcal{D}_0$, and for any $\mathbf{x}(0) \in \mathcal{D}_0$, the sequence generated by the iteration $\mathbf{x}(k + 1) = \mathbf{F}(\mathbf{x}(k))$ remains in $\mathcal{D}_0$ and converges to $\mathbf{x}^*$.

Proof: See, for example, Ortega and Rheinboldt [159, Theorems 5.1.3 and 12.1.2]. □

In applications of Theorem 5.4 one of the critical steps involves specification of an invariant set $\mathcal{D}_0$ such that $\mathbf{F}[\mathcal{D}_0] \subset \mathcal{D}_0$. In Definition 5.5, if $\mathbf{x}^* = \mathbf{0}$ is the fixed point of the contraction mapping $\mathbf{F}(\mathbf{x})$, then for the system $\mathbf{x}(k+1) = \mathbf{F}(\mathbf{x}(k))$, there exists a β , $0 \leq \beta < 1$, such that

$$\begin{aligned} \|\mathbf{F}(\mathbf{x}(k))\| - \|\mathbf{x}(k)\| &= \|\mathbf{F}(\mathbf{x}(k)) - \mathbf{F}(\mathbf{x}^*)\| - \|\mathbf{x}(k) - \mathbf{x}^*\| \\ &\leq -(1 - \beta)\|\mathbf{x}(k) - \mathbf{x}^*\| < 0, \end{aligned} \quad (5.34)$$

which is the basis for the following result known as the *contraction principle* [151, p. 184] originally stated by Kalman and Bertram [157].

Theorem 5.5: *Kalman and Bertram.* Consider the dynamical system $\mathbf{x}(k+1) = \mathbf{F}(\mathbf{x}(k))$ with $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{F} : \mathbb{R}^n \mapsto \mathbb{R}^n$. If $\mathbf{F}(\mathbf{0}) = \mathbf{0}$ and $\mathbf{F}$ is a contraction on $\mathbb{R}^n$ with respect to a vector norm $\|\cdot\|$, then the origin is globally asymptotically stable and one of its Lyapunov functions is $V(\mathbf{x}) = \|\mathbf{x}\|$.

Proof: See Kalman and Bertram [157, Theorem 4]. □

Unfortunately this result, as it currently stands, is not sufficiently general to be of practical use even for some simple systems.

Example 5.2: The linear system $\mathbf{x}(k+1) = \Phi\mathbf{x}(k)$ is asymptotically stable by the contraction principle if $\|\Phi\|_2 < 1$, but this is a very restrictive condition, for example,

$$\Phi = \begin{bmatrix} 0.2 & 1000 \\ 0 & 0.1 \end{bmatrix} \quad (5.35)$$

is certainly a stability matrix (eigenvalues 0.1 and 0.2), but $\|\Phi\|_2 \approx 1000$. □

In Theorem 5.5 the condition that $\mathbf{F}$ be a contraction in all of $\mathbb{R}^n$ may not be satisfied. A natural question, then, is whether or not there exists a transformation of the state such that the resulting system does represent a contraction, and if so, what does this imply about the original system. This question is partly answered by the following Theorem which employs functional composition notation: $\mathbf{H} \circ \mathbf{G}(\mathbf{x}) := \mathbf{H}(\mathbf{G}(\mathbf{x}))$.

Theorem 5.6: Consider the dynamical system $\mathbf{x}(k+1) = \mathbf{F}(\mathbf{x}(k))$ with $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{F} : \mathbb{R}^n \mapsto \mathbb{R}^n$, and a homeomorphism $\mathbf{T} : \mathbb{R}^n \mapsto \mathbb{R}^n$. Then

1. $F(x)$ has precisely the same number of fixed points in $\mathbb{R}^n$ as $T \circ F \circ T^{-1}(y)$;
2. The sequence of iterates generated by the system $x(k+1) = F(x(k))$ converges if, and only if, the sequence $y(k+1) = T \circ F \circ T^{-1}(y(k))$ converges; and
3. If the compositional mapping $T \circ F \circ T^{-1}$ is a contraction on all of $\mathbb{R}^n$, then $F(x)$ has a unique fixed point $x^* \in \mathbb{R}^n$, and for any $x(0) \in \mathbb{R}^n$, the sequence of iterates generated by the system $x(k+1) = F(x(k))$ converges to x^* .

Proof: See Ortega and Rheinboldt [159, Theorems 12.1.5 and 12.1.6]. $\square$

Techniques developed in the following Sections are related to the contraction principle and the results of applying a state transformation as referred to in Theorem 5.6. Exponential stability and state bounds are provided through applications and extensions of Lemma 5.1.

5.4 Linear Systems

Stability properties of linear autonomous systems of the form

$$x(k+1) = \Phi x(k). \quad (5.36)$$

are most easily determined by examining the eigenvalues of Φ and requiring that the spectral radius $\rho(\Phi) < 1$. For system (5.36), Kiendl *et al.* [149] prove a stability result that is similar to the following Theorem by emphasizing a relationship between vector norm and quadratic form Lyapunov functions. The following Theorem offers an alternate interpretation and relates the existence of a vector norm Lyapunov function to the eigenvalues of Φ and, with the use of Lemma 5.1, obtains the exponential bounds.

Theorem 5.7: The null solution of system (5.36) is asymptotically stable if, and only if, there exists a nonsingular $W \in \mathbb{C}^{n \times n}$ and some matrix norm $\|\cdot\|$ such that

$$\beta := \|W\Phi W^{-1}\| < 1. \quad (5.37)$$

In this case, $\forall x(0) \in \mathbb{R}^n$, $\|x(k)\| \leq \beta^k \kappa(W) \|x(0)\|$.

Proof: The proof is immediate from the following three results:

1. system (5.36) is asymptotically stable if, and only if, $\exists \epsilon_1 > 0$ such that $\rho(\Phi) < 1 - \epsilon_1$ [154, Theorem 4.3.1];
2. given Φ and any matrix norm $\|\cdot\|$, for all nonsingular $W \in \mathbb{C}^{n \times n}$, $\rho(\Phi) \leq \|W\Phi W^{-1}\|$ [150, Theorems 5.6.7 & 5.6.9];
3. $\forall \epsilon_2 > 0$, there exists a $W \in \mathbb{C}^{n \times n}$, nonsingular, and a matrix norm $\|\cdot\|$ such that $\rho(\Phi) \leq \|W\Phi W^{-1}\| \leq \rho(\Phi) + \epsilon_2$ [150, Lemma 5.6.10].

Given Φ , W , and $\beta < 1$, sufficiency follows from Items 1 and 2. For necessity, assume asymptotic stability and hence ϵ_1 of Item 1. Set $\epsilon = \epsilon_2 = \epsilon_1$. Then from Item 3, there exists $W \in \mathbb{C}^{n \times n}$ and a matrix norm $\|\cdot\|$ such that

$$\begin{aligned} \beta &:= \|W\Phi W^{-1}\| \\ &\leq \rho(\Phi) + \epsilon \\ &< 1 - \epsilon + \epsilon = 1. \end{aligned} \tag{5.38}$$

Applying Lemma 5.1 to the Lyapunov function $V_k := \|Wx(k)\|$, which satisfies $\Delta V_k \leq -(1 - \beta)V_k < 0$, with $M = 0$, $\alpha = (1 - \beta)$, $\phi(\|x\|) := \|x\|$, and $c = \|W^{-1}\|^{-1}$ to obtain

$$\|W^{-1}\|^{-1}\|x(k)\| \leq \|Wx(k)\| \leq \beta^k \|W\| \|x(0)\| \tag{5.39}$$

yields the exponential bounds. $\square$

For the time invariant case a matrix W may be computed through the following three methods:

1. If Φ is diagonalizable, $\Phi = E\Lambda E^{-1}$ where $\Lambda := \text{diag}\{\lambda_i(\Phi)\}$, then choosing $W = E^{-1}$ and $\|\cdot\| = \|\cdot\|_2$ gives the infimum $\beta = \rho(\Phi)$. However, if Φ is nearly degenerate then $\kappa(W)$ can be very large which may yield conservative bounds.
2. For any Φ , compute¹ $\Phi = U^*RU$, the Schur decomposition (R upper triangular and U unitary) and set $W = \Gamma U$ where $\Gamma = \text{diag}\{1, \gamma, \gamma^2, \dots, \gamma^{n-1}\}$. Choosing γ sufficiently large can yield β arbitrarily close to $\rho(\Phi)$ [150, Lemma 5.6.10]. However, if $\|\cdot\|$ is a unitarily invariant matrix norm and $\gamma > 1$, $\kappa(W) = \gamma^{n-1}$ which again can yield conservative bounds.

¹Here, $(\cdot)^*$ denotes the conjugate transpose.

3. Choose a positive definite $\mathbf{Q}$ and solve $\mathbf{P} - \Phi^T \mathbf{P} \Phi = \mathbf{Q}$ to obtain a positive definite $\mathbf{P}$. Compute the Cholesky factorization $\mathbf{P} = \mathbf{W}^T \mathbf{W}$. If the system is asymptotically stable, $\beta < 1$, however, β need not be close to its infimum $\rho(\Phi)$ and hence may lead to conservative bounds.

In all of the above methods, if $\|\cdot\|$ is a unitarily invariant matrix norm, then $\mathbf{W}$ can be reduced to upper triangular form with the QR factorization. Further discussion of computing a $\mathbf{W}$ matrix and obtaining tight bounds through numerical minimization is deferred to the example of Section 5.6.

Theorem 5.7 can readily be applied to stability analysis of the simple sliding mode observer proposed in the previous Chapter.

Example 5.3: Consider again the linear system $\mathbf{x}(k+1) = \Phi \mathbf{x}(k)$, with scalar measurements $z(k) = \mathbf{C} \mathbf{x}(k)$, for which $(\mathbf{C}, \Phi)$ is an observable pair. The simple sliding mode observer of the previous chapter has error dynamics

$$\mathbf{e}(k+1) = \Phi_o \mathbf{e}(k) - \mathbf{L} \operatorname{sgn}(s_k), \quad (5.40)$$

where $\Phi_o := [\mathbf{I}_n - \mathbf{K} \mathbf{C}]$ and $s_k := \mathbf{M} \mathbf{e}(k)$. Assuming that $\mathbf{K}$ has been chosen [143, p. 255] so that the modulus of all eigenvalues of Φ_o are less than unity, then from Theorem 5.7 there exist compatible vector and matrix norms and a nonsingular matrix $\mathbf{W}$ such that

$$\beta := \|\mathbf{W} \Phi_o \mathbf{W}^{-1}\| < 1. \quad (5.41)$$

Defining the Lyapunov-like function $V(\mathbf{e}) := \|\mathbf{W} \mathbf{e}\|$, it is found that

$$\Delta V_k \leq \alpha - (1 - \beta) V(\mathbf{e}(k)), \quad (5.42)$$

where $\alpha := \|\mathbf{W} \mathbf{L}\|$. Applying Lemma 5.1 gives

$$\|\mathbf{e}(k)\| \leq \|\mathbf{W}^{-1}\| \left\{ \beta^k \|\mathbf{W} \mathbf{e}(0)\| + \alpha \sum_{i=0}^{k-1} \beta^i \right\}, \quad (5.43)$$

and

$$\limsup_{k \rightarrow \infty} \|\mathbf{e}(k)\| \leq \kappa(\mathbf{W}) \frac{\|\mathbf{L}\|}{1 - \beta}. \quad (5.44)$$

□

The above example demonstrates that choosing $\mathbf{W}$ to minimize β does not necessarily yield the minimum bounds on the error. Instead, tight bounds are provided by this approach if a well conditioned matrix $\mathbf{W}$ is available, i.e. $\kappa(\mathbf{W}) \approx 1$, which also yields a small value of β , i.e. $\beta \approx \rho(\Phi_0)$.

For linear time varying systems of the form

$$\mathbf{x}(k+1) = \Phi(k)\mathbf{x}(k), \quad (5.45)$$

the condition $\rho(\Phi(k)) < 1$ at each sample instant is not sufficient for stability [154, p. 94]. Theorem 5.8 below provides necessary and sufficient conditions for exponential stability, similar to those provided by Theorem 5.7, as well as exponential bounds by appealing to the existence of a quadratic-form Lyapunov function. The following Lemma is required for the proof of necessity in Theorem 5.8.

Lemma 5.2: Define the *fundamental matrix* [154, 151]

$$\Psi(k, k_0) := \begin{cases} \downarrow \prod_{j=k_0}^k \Phi(j), & \text{for } k > k_0 \\ \mathbf{I}_n, & \text{for } k = k_0. \end{cases} \quad (5.46)$$

where, again, the notation $\downarrow \prod$ indicates that the index j decreases from left to right in the matrix product. Suppose the null solution of system (5.45) is uniformly asymptotically stable. Then there exists two numbers $a > 0$ and $0 < \eta < 1$ such that

$$\|\Psi(k, k_0)\|_2 \leq a\eta^{k-k_0}, \quad k \geq k_0; \quad (5.47)$$

Proof: See Lakshmikantham and Trigiante [154, Theorem 4.2.2]. $\square$

Lemma 5.2 in fact shows that uniform asymptotic stability and exponential asymptotic stability are equivalent for linear systems.

Theorem 5.8: The null solution of system (5.45) is uniformly asymptotically stable if, and only if, there exists a sequence of nonsingular matrices $\mathbf{W}_k \in \mathbb{C}^{n \times n}$, and some matrix norm $\|\cdot\|$, with $\|\mathbf{W}_k\|$ and $\|\mathbf{W}_k^{-1}\|$ bounded uniformly in k such that

$$\beta := \sup_{\{k \geq 0\}} \{\beta_k\} := \sup_{\{k \geq 0\}} \|\mathbf{W}_{k+1}\Phi(k)\mathbf{W}_k^{-1}\| < 1. \quad (5.48)$$

In this case, given any initial state $\mathbf{x}(0) \in \mathbb{R}^n$ and defining $w := \sup_{\{k \geq 0\}} \|\mathbf{W}_k^{-1}\|$, the state is exponentially bounded according to

$$\|\mathbf{x}(k)\| \leq \beta^k w \|\mathbf{W}_0 \mathbf{x}(0)\|. \quad (5.49)$$

Proof: (Sufficiency) The positive definite decrescent function $V_k := \|\mathbf{W}_k \mathbf{x}(k)\|$ satisfies

$$\begin{aligned} \Delta V_k &:= V_{k+1} - V_k \\ &= \|\mathbf{W}_{k+1} \Phi(k) \mathbf{x}(k)\| - \|\mathbf{W}_k \mathbf{x}(k)\| \\ &= \|\mathbf{W}_{k+1} \Phi(k) \mathbf{W}_k^{-1} \mathbf{W}_k \mathbf{x}(k)\| - \|\mathbf{W}_k \mathbf{x}(k)\| \\ &\leq -[1 - \beta_k] \|\mathbf{W}_k \mathbf{x}(k)\| \\ &\leq -[1 - \beta] \|\mathbf{W}_k \mathbf{x}(k)\| < 0, \quad \forall k \geq 0. \end{aligned} \quad (5.50)$$

which implies uniform asymptotic stability by Theorem 5.1.

(Necessity) Suppose that system (5.45) is uniformly asymptotically stable. Set the vector norm to be the standard Euclidean norm, $\|\cdot\| = \|\cdot\|_2$, and let the matrix norm be that induced by the Euclidean vector norm, $\|\cdot\| = \|\cdot\|_2$. Let $\bar{\sigma}(\cdot)$ and $\underline{\sigma}(\cdot)$ represent the maximum and minimum singular values of their matrix arguments, respectively. Select any sequence of positive definite symmetric matrices, $\{\mathbf{Q}_k \mid k = 0, 1, \dots\}$, such that

1. $\|\mathbf{Q}_k\|_2 = \bar{\sigma}(\mathbf{Q}_k)$ is bounded uniformly in k ; and
2. $\exists d > 0$ such that upon computing the Cholesky factorization,

$$\mathbf{Q}_k = \mathbf{D}_k^T \mathbf{D}_k, \quad (5.51)$$

the matrices $\mathbf{D}_k$ satisfy, $\forall k \geq 0$,

$$\|\mathbf{D}_k^{-1}\|_2^{-1} = \underline{\sigma}(\mathbf{D}_k) \geq \frac{1}{d}. \quad (5.52)$$

Condition 2, above, bounds the matrices $\mathbf{Q}_k$ away from the set of singular matrices. Consider, as constructed by Willems [151, p. 182], the sequence of positive definite symmetric matrices $\{\mathbf{P}_k \mid k = 0, 1, \dots\}$ defined by

$$\mathbf{P}_k := \mathbf{Q}_k + \Sigma_k. \quad (5.53)$$

where

$$\Sigma_k := \sum_{j=k+1}^{\infty} \Psi^T(j, k) Q_j \Psi(j, k), \quad (5.54)$$

and $\Psi(j, k)$ is the fundamental matrix defined in (5.46). Note that the P_k are well defined; from Lemma 5.2, there exists two numbers $a > 0$ and $0 < \eta < 1$ such that, $\forall k \geq 0$,

$$\begin{aligned} \|P_k\|_2 &\leq a^2 \left[\sup_{j \geq k} \|Q_j\|_2 \right] \sum_{j=k}^{\infty} \eta^{2(j-k)} \\ &= \frac{a^2}{1 - \eta^2} \left[\sup_{j \geq k} \|Q_j\|_2 \right], \end{aligned} \quad (5.55)$$

which means that the sum in (5.54) converges uniformly and $\|P_k\|_2$ is bounded uniformly in k since the Q_k are assumed to be uniformly bounded in k . Moreover, by noting that the fundamental matrix satisfies

$$\Psi(j, k) = \Psi(j, k+1) \Phi(k), \quad \text{for } j \geq k+1. \quad (5.56)$$

the matrices P_k are seen to solve the discrete-time Lyapunov equation,

$$\begin{aligned} P_k - \Phi^T(k) P_{k+1} \Phi(k) &= Q_k + \sum_{j=k+1}^{\infty} \Psi^T(j, k) Q_j \Psi(j, k) \\ &\quad - \Phi^T(k) \left[\sum_{j=k+1}^{\infty} \Psi^T(j, k+1) Q_j \Psi(j, k+1) \right] \Phi^T(k) \\ &= Q_k, \end{aligned} \quad (5.57)$$

in which (5.56) has been used in the last term of the first equality of (5.57) to obtain the result. From the sequence $\{Q_k \mid k = 0, 1, \dots\}$, therefore, one can construct a sequence of positive definite symmetric, uniformly bounded matrices P_k such that $P_k = Q_k + \Sigma_k$ where Σ_k is positive semi-definite symmetric and $\forall x \in \mathbb{R}^n$,

$$x^T [P_k - \Phi^T(k) P_{k+1} \Phi(k)] x = x^T Q_k x. \quad (5.58)$$

Let $P_k = W_k^T W_k$ and $Q_k = D_k^T D_k$ be Cholesky factorizations. Then from the construction of the Cholesky factors [160, Sec. 5.2], $\|W_k\|_2^2 = \|P_k\|_2$ and hence $\|W_k\|_2$ is uniformly bounded,

$$\gamma := \sup_{\{k \geq 0\}} \{\|W_k\|_2\} < \infty. \quad (5.59)$$

From (5.54), the matrices Σ are at least positive semi-definite symmetric and hence $\forall \mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{x}^T \mathbf{P}_k \mathbf{x} = \mathbf{x}^T (\mathbf{Q}_k + \Sigma_k) \mathbf{x} \geq \mathbf{x}^T \mathbf{Q}_k \mathbf{x}. \quad (5.60)$$

In terms of the Cholesky factors and the Euclidean vector norm, (5.60) becomes

$$\|\mathbf{W}_k \mathbf{x}\|_2 \geq \|\mathbf{D}_k \mathbf{x}\|_2, \quad \forall \mathbf{x} \in \mathbb{R}^n. \quad (5.61)$$

In particular, take $\mathbf{x} = \mathbf{x}^*$, where $\mathbf{x}^*$ with $\|\mathbf{x}^*\|_2 = 1$ is the right singular vector of $\mathbf{W}_k$ associated with $\underline{\sigma}(\mathbf{W}_k)$ to get

$$\begin{aligned} \underline{\sigma}(\mathbf{W}_k) &= \|\mathbf{W}_k \mathbf{x}^*\|_2 \\ &\geq \|\mathbf{D}_k \mathbf{x}^*\|_2 \\ &\geq \underline{\sigma}(\mathbf{D}_k) \|\mathbf{x}^*\|_2 = \underline{\sigma}(\mathbf{D}_k). \end{aligned} \quad (5.62)$$

From (5.52) and (5.62), $\underline{\sigma}(\mathbf{W}_k) \geq \underline{\sigma}(\mathbf{D}_k) \geq 1/d$ which implies that $\|\mathbf{W}_k^{-1}\|_2$ is uniformly bounded,

$$\|\mathbf{W}_k^{-1}\|_2 \leq d. \quad (5.63)$$

Combining (5.59) and (5.63) and using (5.62) gives the inequalities

$$0 < \frac{1}{d\gamma} \leq \frac{\underline{\sigma}(\mathbf{D}_k)}{\bar{\sigma}(\mathbf{W}_k)} \leq 1. \quad (5.64)$$

Substitution of the Cholesky factors into (5.58) and using the nonsingularity of the $\mathbf{W}_k$ gives, $\forall \mathbf{x} \in \mathbb{R}^n$,

$$\begin{aligned} \mathbf{x}^T [\mathbf{P}_k - \Phi^T(k) \mathbf{P}_{k+1} \Phi(k)] \mathbf{x} &= \mathbf{x}^T \mathbf{W}_k^T \mathbf{W}_k \mathbf{x} - \mathbf{x}^T \Phi^T(k) \mathbf{W}_{k+1}^T \mathbf{W}_{k+1} \Phi(k) \mathbf{x} \\ &= \mathbf{x}^T \mathbf{W}_k^T \mathbf{W}_k \mathbf{x} - \mathbf{x}^T \mathbf{W}_k^T \mathbf{W}_k^{-T} \Phi^T(k) \mathbf{W}_{k+1}^T \mathbf{W}_{k+1} \Phi(k) \mathbf{W}_k^{-1} \mathbf{W}_k \mathbf{x} \\ &= \mathbf{x}^T \mathbf{D}_k^T \mathbf{D}_k \mathbf{x}, \end{aligned} \quad (5.65)$$

where the identity matrices $(\mathbf{W}_k^{-1} \mathbf{W}_k)$ have been inserted twice in the second equality, and the final equality follows from (5.58). Expressing the above result in terms of the Euclidean norm yields,

$$\|\mathbf{W}_k \mathbf{x}\|_2^2 - \|\mathbf{W}_{k+1} \Phi(k) \mathbf{W}_k^{-1} \mathbf{W}_k \mathbf{x}\|_2^2 = \|\mathbf{D}_k \mathbf{x}\|_2^2. \quad (5.66)$$

Rearranging (5.66) and using the definition of $\|\cdot\|_2$ gives

$$\begin{aligned}
 \beta &:= \sup_{\{k \geq 0\}} \|\mathbf{W}_{k+1} \Phi(k) \mathbf{W}_k^{-1}\|_2 \\
 &= \sup_{\{k \geq 0\}} \sup_{\{\mathbf{x}, \|\mathbf{W}_k \mathbf{x}\|_2 = 1\}} \frac{\|\mathbf{W}_{k+1} \Phi(k) \mathbf{W}_k^{-1} \mathbf{W}_k \mathbf{x}\|_2}{\|\mathbf{W}_k \mathbf{x}\|_2} \\
 &= \sup_{\{k \geq 0\}} \sup_{\{\mathbf{x}, \|\mathbf{W}_k \mathbf{x}\|_2 = 1\}} \left[1 - \frac{\|\mathbf{D}_k \mathbf{x}\|_2^2}{\|\mathbf{W}_k \mathbf{x}\|_2^2} \right]^{\frac{1}{2}} \\
 &\leq \sup_{\{k \geq 0\}} \left[1 - \frac{\underline{\sigma}^2(\mathbf{D}_k)}{\overline{\sigma}^2(\mathbf{W}_k)} \right]^{\frac{1}{2}} < 1
 \end{aligned} \tag{5.67}$$

as required.

If the hypotheses hold, then $V_k = \|\mathbf{W}_k \mathbf{x}(k)\|$ satisfies $\Delta V_k \leq -(1 - \beta)V_k$ and the exponential bounds are obtained from Lemma 5.1 with $M = 0$, $\alpha = (1 - \beta)$, $\phi(\|\mathbf{x}\|) = \|\mathbf{x}\|$, and $c = 1/u$. $\square$

In Theorem 5.8, if the original system (5.45) is transformed according to $\mathbf{y}(k) = \mathbf{W}_k \mathbf{x}(k)$, then $\mathbf{y}(k+1) = \Psi(k) \mathbf{y}(k)$, where $\Psi(k) := \mathbf{W}_{k+1} \Phi(k) \mathbf{W}_k^{-1}$. Given the vector and matrix norms, the condition

$$\sup_{\{k \geq 0\}} \|\mathbf{W}_{k+1} \Phi(k) \mathbf{W}_k^{-1}\| = \sup_{\{k \geq 0\}} \beta_k < 1 \tag{5.68}$$

is satisfied if $\sup_{\{k \geq 0\}} \|\Psi(k)\| < 1$ in which case the transformed system is a contraction on $\mathbb{R}^n$ with respect to the selected vector norm. In this sense, Theorems 5.7 and 5.8 are related, for linear systems, to the contraction principle stated in Theorem 5.5 by way of a state transformation as referred to in Theorem 5.6. However, the problem of obtaining a suitable sequence $\mathbf{W}_k$ is largely dependent on the temporal nature of $\Phi(k)$ and the choice of vector and matrix norms, and remains open.

5.5 Nonlinear Systems

The method of the previous section can be carried over, in a more restricted sense, to nonlinear systems of the form

$$\mathbf{x}(k+1) = \mathbf{f}(k, \mathbf{x}(k)). \tag{5.69}$$

where $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{f} : \mathbb{N} \times \mathbb{R}^n \mapsto \mathbb{R}^n$ is C^p in $\mathbf{x}$, $p \geq 1$, and $\mathbf{f}(k, \mathbf{0}) = \mathbf{0} \forall k$. In order to facilitate the use of functional composition notation in this Section, temporal dependence of the function $\mathbf{f}(k, \mathbf{x})$ is shifted to a subscript by introducing

$$\mathbf{f}_k(\mathbf{x}) := \mathbf{f}(k, \mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n. \quad (5.70)$$

For stability of linear time-invariant systems, treated in Theorem 5.7 of Section 5.4, necessary and sufficient conditions for exponential stability were obtained through the existence of a time-invariant linear nonsingular transformation, $\mathbf{W}$, under which the transformed system represented a contraction, with respect to the prescribed vector norm, in all of $\mathbb{R}^n$. Similarly in the time-varying linear case, treated in Theorem 5.8 of Section 5.4, a time-varying bounded linear nonsingular transformation, $\mathbf{W}_k$, with bounded inverse was required to obtain global exponential stability. In the case of nonlinear time-varying systems, a nonlinear time-varying transformation of the state is considered under which the transformed system, at least locally about the stationary point, represents a contraction with respect to the prescribed vector norm. Equivalent concepts of uniform boundedness are required for this result.

Definition 5.6: Let $\mathbf{T} : \mathcal{D} \subseteq \mathbb{R}^n \mapsto \mathbb{R}^n$ be a continuous mapping from $\mathcal{D}$ into $\mathbb{R}^n$. For any vector norm $\|\cdot\|$ defined on $\mathbb{R}^n$, the *operator norm* of $\mathbf{T}$ is defined as

$$\eta(\mathbf{T}) := \sup_{\{\mathbf{x} \in \mathcal{D}, \mathbf{x} \neq \mathbf{0}\}} \frac{\|\mathbf{T}(\mathbf{x})\|}{\|\mathbf{x}\|} \quad (5.71)$$

□

Intuitively, $\eta(\mathbf{T})$ is the maximum expansion factor resulting from application of the mapping $\mathbf{T}$ to any vector in $\mathcal{D}$. Directly from the definition, the operator norm is compatible with the vector norm in the sense that

$$\|\mathbf{T}(\mathbf{x})\| \leq \eta(\mathbf{T})\|\mathbf{x}\| \quad \forall \mathbf{x} \in \mathcal{D}. \quad (5.72)$$

Another result that is not only used throughout the remainder of this Chapter, but is also used repeatedly in Chapters 6 and 7 is stated in the following Theorem.

Theorem 5.9: *Fundamental Theorem of Integral Calculus for Vector-valued Functions of Several Variables.* Let $F : \mathcal{D} \subseteq \mathbb{R}^n \mapsto \mathbb{R}^n$, for some open convex set $\mathcal{D}$, be C^p , $p \geq 1$, in $\mathbf{x}$. Then, for any $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathcal{D}$,

$$F(\mathbf{x}_1) - F(\mathbf{x}_2) = \left[\int_0^1 \frac{\partial F}{\partial \mathbf{x}}(\omega_\lambda) d\lambda \right] [\mathbf{x}_1 - \mathbf{x}_2], \quad (5.73)$$

where

$$\omega_\lambda := \lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2. \quad (5.74)$$

Proof: See Marsden, [144, p. 199], or Ortega and Rheinboldt, Theorem 3.2.7 [159, p. 71]. $\square$

For system (5.69) the following theorem provides sufficient conditions for exponential asymptotic stability, a domain of exponential stability, and exponential bounds on the state. The proof of this theorem is related in a sense to the approach of Tarn and Rasis [82, Theorem 4], particularly in the use of the fundamental theorem of integral calculus, however, the sufficient conditions are local and less restrictive, and also lead to a stability domain.

Theorem 5.10: Suppose that for system (5.69) there exist an open convex set $\Omega \subseteq \mathbb{R}^n$, with $\mathbf{0} \in \Omega$, compatible vector, $\|\cdot\|$, matrix, $\|\cdot\|$, and operator, $\eta(\cdot)$, norms, and a sequence of continuously differentiable functions $\{\mathbf{T}_k(\mathbf{x}) \mid k = 0, 1, 2, \dots\}$, $\mathbf{T}_k : \mathbb{R}^n \mapsto \mathcal{C} \subseteq \mathbb{C}^n$, each with a continuously differentiable inverse $\mathbf{T}_k^{-1} : \mathcal{C} \mapsto \mathbb{R}^n$, with $\mathbf{T}(\mathbf{0}) = \mathbf{0}$, and $\eta(\mathbf{T}_k)$ and $\eta(\mathbf{T}_k^{-1})$ bounded uniformly in k such that

$$\beta := \sup_{\{k \geq 0\}} \beta_k < 1. \quad (5.75)$$

where

$$\beta_k := \sup_{\mathbf{v} \in \Omega} \left\| \left[\frac{\partial \mathbf{T}_{k+1} \circ \mathbf{f}_k \circ \mathbf{T}_k^{-1}(\mathbf{y})}{\partial \mathbf{y}} \right]_{\mathbf{y}=\mathbf{v}} \right\|. \quad (5.76)$$

In this case, define

$$V(k, \mathbf{x}) := \|\mathbf{T}_k(\mathbf{x})\|. \quad (5.77)$$

$$d_s := \sup_d \{ \mathbf{y} \in \mathbb{R}^n \mid \|\mathbf{y}\| \leq d \} \subset \Omega. \quad (5.78)$$

$$Y_s := \{ \mathbf{y} \in \mathbb{R}^n \mid \|\mathbf{y}\| < d_s \}, \quad (5.79)$$

$$X_k := \{ \mathbf{x} \in \mathbb{R}^n \mid V(k, \mathbf{x}) < d_s \}, \quad \text{and} \quad (5.80)$$

$$\eta_0 := \sup_{\{k \geq 0\}} \eta(\mathbf{T}_k^{-1}). \quad (5.81)$$

Then $V(k, \mathbf{x}(k))$ is a Lyapunov function for the system with initialization in X_0 , and $\exists \beta_s$, $0 \leq \beta_s \leq \beta$ such that if $\mathbf{x}(0) \in X_0$, then, $\forall k > 0$, $\mathbf{x}(k) \in X_k$ and the state is exponentially bounded according to

$$\|\mathbf{x}(k)\| \leq \beta_s^k \eta_0 \|\mathbf{T}_0(\mathbf{x}(0))\|, \quad (5.82)$$

Hence X_0 is a domain of exponential asymptotic stability, and the origin of system (5.69) is locally uniformly exponentially asymptotically stable.

Proof: The function $V(k, \mathbf{x}) := \|\mathbf{T}_k(\mathbf{x})\|$ is positive definite decrescent by assumption since, $\forall \mathbf{x} \in \mathbb{R}^n$,

$$\left[\sup_{k \geq 0} \eta(\mathbf{T}_k^{-1}) \right]^{-1} \|\mathbf{x}\| \leq \|\mathbf{T}_k(\mathbf{x})\| \leq \left[\sup_{k \geq 0} \eta(\mathbf{T}_k) \right] \|\mathbf{x}\|. \quad (5.83)$$

The set Y_s defined in (5.79) is the "largest" set of the form $\{\mathbf{y} \mid \|\mathbf{y}\| < d\}$ which is completely contained within the convex set Ω . Since Ω is open with $\mathbf{0} \in \Omega$, $d_s > 0$, $Y_s \subset \Omega$, and $\mathbf{0} \in Y_s$. The set Y_s is convex from the properties of the vector norm. Define

$$\beta_s := \sup_{\{k \geq 0\}} \sup_{\{\mathbf{v} \in Y_s\}} \left\| \left[\frac{\partial \mathbf{T}_{k+1} \circ \mathbf{f}_k \circ \mathbf{T}_k^{-1}(\mathbf{y})}{\partial \mathbf{y}} \right]_{\mathbf{y}=\mathbf{v}} \right\|, \quad (5.84)$$

which satisfies $0 \leq \beta_s \leq \beta$. Then, for any $k \geq 0$, if $\mathbf{x}(k) \in X_k$, $\mathbf{T}_k(\mathbf{x}(k)) \in Y_s$ and

$$\begin{aligned} \Delta V_k &:= V(k+1, \mathbf{x}(k+1)) - V(k, \mathbf{x}(k)) \\ &= \|\mathbf{T}_{k+1} \circ \mathbf{f}_k(\mathbf{x}(k))\| - \|\mathbf{T}_k(\mathbf{x}(k))\| \\ &= \|\mathbf{T}_{k+1} \circ \mathbf{f}_k \circ \mathbf{T}_k^{-1} \circ \mathbf{T}_k(\mathbf{x}(k))\| - \|\mathbf{T}_k(\mathbf{x}(k))\| \\ &= \left\| \int_0^1 \left\{ \left[\frac{\partial}{\partial \mathbf{y}} \mathbf{T}_{k+1} \circ \mathbf{f}_k \circ \mathbf{T}_k^{-1}(\mathbf{y}) \right]_{\mathbf{y}=\lambda \mathbf{T}_k(\mathbf{x}(k))} \right\} \mathbf{T}_k(\mathbf{x}(k)) d\lambda \right\| - \|\mathbf{T}_k(\mathbf{x}(k))\| \\ &\leq \left\| \int_0^1 \left\{ \left[\frac{\partial}{\partial \mathbf{y}} \mathbf{T}_{k+1} \circ \mathbf{f}_k \circ \mathbf{T}_k^{-1}(\mathbf{y}) \right]_{\mathbf{y}=\lambda \mathbf{T}_k(\mathbf{x}(k))} \right\} d\lambda \right\| \|\mathbf{T}_k(\mathbf{x}(k))\| - \|\mathbf{T}_k(\mathbf{x}(k))\| \\ &\leq \beta_s \|\mathbf{T}_k(\mathbf{x}(k))\| - \|\mathbf{T}_k(\mathbf{x}(k))\| \\ &= -(1 - \beta_s) V(k, \mathbf{x}(k)) < 0. \end{aligned} \quad (5.85)$$

In the above derivation, Theorem 5.9 is applied to obtain the fourth equality, and compatibility of the matrix and vector norms yields the first inequality. The second inequality results from the fact that the norm of the integral is less than or equal to the integral of the norm, the set Y_s is convex with $\mathbf{0} \in Y_s$ and since $\mathbf{x}(k) \in X_k$ the vector $\mathbf{T}_k(\mathbf{x}(k)) \in Y_s$.

and the definition of β_s in (5.84). The final strict inequality in (5.85) follows from the fact that $Y_s \subset \Omega$ and the hypothesis (5.75) which imply that $\beta_s < 1$.

The result in (5.85) shows that if, for any $k \geq 0$, $\mathbf{x}(k) \in X_k$, then $\|\mathbf{T}_k(\mathbf{x}(k))\| < d_s$ and

$$\begin{aligned} \|\mathbf{T}_{k+1}(\mathbf{x}(k+1))\| &\leq \beta_s \|\mathbf{T}_k(\mathbf{x}(k))\| \\ &< \beta_s d_s < d_s. \end{aligned} \quad (5.86)$$

which means that $\mathbf{x}(k+1) \in X_{k+1}$. It follows by induction that if $\mathbf{x}(0) \in X_0$, then $\forall k \geq 0$, $\mathbf{x}(k) \in X_k$ and $\Delta V_k \leq -(1 - \beta_s)V_k < 0$. Hence $V_k := V(k, \mathbf{x}(k)) := \|\mathbf{T}_k(\mathbf{x}(k))\|$ is a Lyapunov function for the system with initialization in X_0 , and the exponential bounds follow from Lemma 5.1 with $M = 0$, $\alpha = (1 - \beta_s)$, $c = \eta_0$, and $\phi(\|\mathbf{x}\|) = \|\mathbf{x}\|$. $\square$

In the above Theorem, X_0 is not necessarily an invariant set of the system, i.e. initialization in X_0 does not necessarily imply that the state remains in X_0 . The sets X_k are the possibly time varying images of the set Y_s under the inverse mappings $\mathbf{T}_k^{-1}$; $X_k = \mathbf{T}_k^{-1}[Y_s]$. The set Y_s is an invariant set of the transformed system $\mathbf{y}(k) := \mathbf{T}_k \mathbf{x}(k)$. The scalar β_s is introduced to possibly yield tighter bounds than that which would result by using β directly. If $\mathbf{x}(0) \in X_0$ is known, then introducing

$$Y_0 := \{\mathbf{y} \in \mathbb{R}^n \mid \|\mathbf{y}\| < \|\mathbf{T}_0(\mathbf{x}(0))\|\} \subset Y_s \subset \Omega. \quad (5.87)$$

Y_0 is convex and substituting β_s in (5.84) with β^* where

$$0 \leq \beta^* := \sup_{\{k \geq 0\}} \sup_{\{\mathbf{v} \in Y_0\}} \left\| \left[\frac{\partial \mathbf{T}_{k+1} \circ \mathbf{f}_k \circ \mathbf{T}_k^{-1}(\mathbf{y})}{\partial \mathbf{y}} \right]_{\mathbf{y}=\mathbf{v}} \right\| \leq \beta_s \leq \beta < 1 \quad (5.88)$$

may possibly yield even tighter bounds.

Applications of Theorem 5.10, however, requires the availability of a sequence of bounded diffeomorphic transformations, $\{\mathbf{T}_k(\mathbf{x}) \mid k = 0, 1, 2, \dots\}$, which satisfy the hypothesis (5.75) in some suitably large convex set Ω . The Jacobian of the composition $\mathbf{T}_{k+1} \circ \mathbf{f}_k \circ \mathbf{T}_k^{-1}$ is computed by means of the chain rule:

$$\left[\frac{\partial \mathbf{T}_{k+1} \circ \mathbf{f}_k \circ \mathbf{T}_k^{-1}(\mathbf{y})}{\partial \mathbf{y}} \right]_{\mathbf{y}=\mathbf{v}} = \left\{ \left[\frac{\partial}{\partial \mathbf{w}_1} \mathbf{T}_{k+1}(\mathbf{w}_1) \right]_{\mathbf{w}_1=\mathbf{f}_k \circ \mathbf{T}_k^{-1}(\mathbf{v})} \right\}^*$$

$$\left\{ \left[\frac{\partial}{\partial \mathbf{w}_2} \mathbf{f}_k(\mathbf{w}_2) \right]_{\mathbf{w}_2 = \mathbf{T}_k^{-1}(\mathbf{v})} \right\}^* \left\{ \left[\frac{\partial}{\partial \mathbf{w}_3} \mathbf{T}_k^{-1}(\mathbf{w}_3) \right]_{\mathbf{w}_3 = \mathbf{v}} \right\}. \quad (5.89)$$

In particular, if the transformations are linear, say

$$\mathbf{T}_k(\mathbf{x}) = \mathbf{W}_k \mathbf{x}, \quad (5.90)$$

where $\mathbf{W}_k \in \mathbb{C}^{n \times n}$, then (5.89) becomes

$$\left[\frac{\partial \mathbf{T}_{k+1} \circ \mathbf{f}_k \circ \mathbf{T}_k^{-1}(\mathbf{y})}{\partial \mathbf{y}} \right]_{\mathbf{y}=\mathbf{v}} = \mathbf{W}_{k+1} \left\{ \left[\frac{\partial}{\partial \mathbf{x}} \mathbf{f}_k(\mathbf{x}) \right]_{\mathbf{x}=\mathbf{W}_k^{-1}\mathbf{v}} \right\} \mathbf{W}_k^{-1}. \quad (5.91)$$

From this representation, an immediate corollary of Theorems 5.10 and 5.8 is the well known result (see for example Willems, Theorem 6.1 [151, p. 183]) of *stability by first approximation*, but with the additional results of exponential asymptotic stability and a means to obtain a domain of exponential asymptotic stability and exponential bounds on the state.

Corollary 5.1: Consider the nonlinear system

$$\mathbf{x}(k+1) = \mathbf{f}_k(\mathbf{x}(k)), \quad (5.92)$$

with $\mathbf{f}_k(\mathbf{0}) = \mathbf{0}$ for all $k \geq 0$ and each $\mathbf{f}_k$ is C^p in $\mathbf{x}$, $p \geq 1$. The equation of first variation of the system about the stationary point $\mathbf{x} = \mathbf{0}$ is

$$\mathbf{z}(k+1) = \mathbf{F}(k)\mathbf{z}(k), \quad (5.93)$$

where

$$\mathbf{F}(k) := \left[\frac{\partial \mathbf{f}_k(\mathbf{x})}{\partial \mathbf{x}} \right]_{\mathbf{x}=\mathbf{0}}. \quad (5.94)$$

If the origin of the linear time varying system (5.93) is uniformly asymptotically stable, then the null solution of the nonlinear system (5.92) is exponentially asymptotically stable.

Proof: If system (5.93) is uniformly asymptotically stable, then from Theorem 5.8 there exists a sequence of nonsingular matrices $\mathbf{W}_k \in \mathbb{C}^{n \times n}$, a vector norm, $\|\cdot\|$, and a compatible matrix norm $\|\cdot\|$, with $\|\mathbf{W}_k\|$ and $\|\mathbf{W}_k^{-1}\|$ uniformly bounded in k such that $\beta := \sup_k \{\beta_k\} < 1$, where

$$\beta_k := \|\mathbf{W}_{k+1}\mathbf{F}(k)\mathbf{W}_k^{-1}\|. \quad (5.95)$$

From the assumed continuity of elements of the Jacobian of $\mathbf{f}_k(\mathbf{x})$, there exists an open convex set $\Omega \in \mathbb{R}^n$ such that the hypothesis of Theorem 5.10 are satisfied with $\mathbf{T}_k(\mathbf{x}) := \mathbf{W}_k \mathbf{x}$. $\square$

The following simple example of a nonlinear system in $\mathbb{R}^2$ illustrates some of the basic steps in applications of this theory.

Example 5.4: Consider $\mathbf{x} = [x_1, x_2]^T$ and the nonlinear system

$$\begin{aligned} x_1(k+1) &= \frac{1}{2}x_1(k) + \frac{1}{2}x_2(k) \\ x_2(k+1) &= ax_1^3(k) + \frac{1}{4}x_2(k). \end{aligned} \quad (5.96)$$

Employing the standard Euclidean vector norm, $\|\cdot\|_2$, and the Frobenius matrix norm, $\|\cdot\|_F$, and defining

$$\mathbf{W} := \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix} \Rightarrow \mathbf{W}^{-1} = \frac{1}{2} \begin{bmatrix} 2 & -2 \\ 0 & 1 \end{bmatrix}. \quad (5.97)$$

a straight forward calculation shows that, with $\mathbf{v} := [v_1, v_2]^T$,

$$\beta := \left\| \mathbf{W} \left\{ \left[\frac{\partial}{\partial \mathbf{x}} \mathbf{f}(\mathbf{x}) \right]_{\mathbf{x}=\mathbf{W}^{-1}\mathbf{v}} \right\} \mathbf{W}^{-1} \right\|_F = \left\| \frac{1}{4} \begin{bmatrix} 2+\alpha & -\alpha \\ \alpha & 1-\alpha \end{bmatrix} \right\|_F \quad (5.98)$$

where $\alpha := 24a(v_1 - v_2)^2$. Hence $\beta < 1$ for $-1.92 < \alpha < 1.42$, which shows that the origin is exponentially asymptotically stable for any finite a . With $a = -0.1$, the set Ω can be taken as

$$\Omega := \{\mathbf{v} \in \mathbb{R}^2 \mid |v_1 - v_2| < 0.89\}. \quad (5.99)$$

Figure 5.1 illustrates the relevant sets employed in this example. Now set

$$d_s := \sup_d [\{\mathbf{y} \in \mathbb{R}^2 \mid \|\mathbf{y}\|_2 \leq d\} \subset \Omega] = \frac{0.89}{\sqrt{2}} = 0.63. \quad (5.100)$$

and then

$$\begin{aligned} V(\mathbf{x}(k)) &:= \|\mathbf{W}\mathbf{x}(k)\|_2 \\ Y_s &:= \{\mathbf{y} \in \mathbb{R}^2 \mid \|\mathbf{y}\|_2 < 0.63\}. \\ X_k &\equiv X := \{\mathbf{x} \in \mathbb{R}^2 \mid V(\mathbf{x}) < 0.63\}. \end{aligned} \quad (5.101)$$

Figure 5.1: Illustration of sets and stability domains for Example 5.4.

and X is a domain of exponential asymptotic stability. If $\|\mathbf{x}(0)\|_2 < 0.2$ then

$$\|\mathbf{W}\mathbf{x}(0)\|_2 \leq \bar{\sigma}(\mathbf{W})\|\mathbf{x}(0)\|_2 \leq d_0 := 0.58. \quad (5.102)$$

with equality for at least one $\mathbf{x}(0)$ satisfying $\|\mathbf{x}(0)\|_2 = 0.2$. Following (5.87) and (5.88), set

$$Y_0 := \left\{ \mathbf{y} \in \mathbb{R}^2 \mid \|\mathbf{y}\|_2 < \sup_{\{\|\mathbf{x}\| \leq 0.2\}} \{\|\mathbf{W}\mathbf{x}\|_2\} = 0.58 \right\}, \quad (5.103)$$

and find that

$$\beta^* := \sup_{\{\mathbf{v} \in Y_0\}} \left\| \mathbf{W} \left[\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \right]_{\mathbf{x}=\mathbf{W}^{-1}\mathbf{v}} \mathbf{W}^{-1} \right\|_F = 0.873. \quad (5.104)$$

State bounds for initialization with $\|\mathbf{x}(0)\|_2 < 0.2$ are then given by

$$\|\mathbf{x}(k)\| \leq (\beta^*)^k \|\mathbf{W}\|_F \|\mathbf{W}^{-1}\|_F \|\mathbf{x}(0)\| \leq 0.9 * (0.88)^k. \quad (5.105)$$

Other choices for the matrix W or matrix and vector norms may yield tighter bounds and/or a larger domain of exponential asymptotic stability. $\square$

5.6 Exponential Stability of Nonlinear Observers

The purpose of this Section is twofold: first, to apply the methods proposed in the previous Sections to obtain stability results for discrete-time nonlinear observers; and second, to include sufficient generalization to facilitate direct applications of the results to discrete-time sliding mode observers proposed in Chapters 6 and 7. As previously discussed, trajectories of sliding mode systems usually consist of a reaching phase followed by the sliding mode. Observers proposed in Chapters 6 and 7 contain a term which is generally nonzero during the reaching phase, but which is zero during the sliding mode. The following Lemma extends the results of Lemma 5.1 to include this additional term.

Lemma 5.3: For the system

$$\mathbf{x}(k+1) = \mathbf{f}(k, \mathbf{x}(k)), \quad (5.106)$$

if there exists functions $V(k, \mathbf{x}(k))$, $\phi \in \mathcal{K}$, and $M_1(k)$, and constants $c > 0$, $M \geq 0$, and $\beta < 1$ such that

1. $V(k, \mathbf{x}(k)) \geq c\phi(\|\mathbf{x}(k)\|)$,
2. $\forall k \geq k_s \geq 0, M_1(k) = 0$,
3. $V(k+1, \mathbf{x}(k+1)) \leq M + M_1(k) + \beta V(k, \mathbf{x}(k))$,

then

$$V(k, \mathbf{x}(k)) \leq M \sum_{i=0}^{k-1} \beta^i + \beta^k V(0, \mathbf{x}(0)) + \sum_{j=0}^{k-1} \beta^j M_1(k-1-j). \quad (5.107)$$

and

$$\limsup_{k \rightarrow \infty} \phi(\|\mathbf{x}(k)\|) \leq \frac{M}{c(1-\beta)}. \quad (5.108)$$

Proof: Repeated applications of Condition 3 gives

$$V(k, \mathbf{x}(k)) \leq M + M_1(k-1) + \beta V(k-1, \mathbf{x}(k-1))$$

$$\begin{aligned}
 &\leq (1 + \beta)M + M_1(k-1) + \beta M_1(k-2) + \beta^2 V(k-2, \mathbf{x}(k-2)) \\
 &\vdots \\
 &\leq M \sum_{i=0}^{k-1} \beta^i + \beta^k V(0, \mathbf{x}(0)) + \sum_{j=0}^{k-1} \beta^j M_1(k-1-j)
 \end{aligned} \tag{5.109}$$

which establishes (5.107). Noting from Condition 2 that

$$M_1(k-1-j) = 0, \quad \text{for } k-j-1 \geq k_s, \tag{5.110}$$

a shift of index in (5.109), $\ell = j - k + k_s \Rightarrow j = \ell + k - k_s$, gives $\forall k > k_s$,

$$\begin{aligned}
 V(k, \mathbf{x}(k)) &\leq M \sum_{i=0}^{k-1} \beta^i + \beta^k V(0, \mathbf{x}(0)) + \sum_{\ell=0}^{k_s-1} \beta^{\ell-k_s+k} M_1(k_s-1-\ell) \\
 &= M \sum_{i=0}^{k-1} \beta^i + \beta^k \left\{ V(0, \mathbf{x}(0)) + \sum_{\ell=0}^{k_s-1} \beta^{\ell-k_s} M_1(k_s-1-\ell) \right\}
 \end{aligned} \tag{5.111}$$

where the term in braces in (5.111) is a finite constant. Hence, with Condition 1,

$$\begin{aligned}
 \limsup_{k \rightarrow \infty} \phi(\|\mathbf{x}(k)\|) &\leq \limsup_{k \rightarrow \infty} \frac{1}{c} V(k, \mathbf{x}(k)) \\
 &\leq \lim_{k \rightarrow \infty} \frac{M}{c} \sum_{i=0}^{k-1} \beta^i \\
 &= \frac{M}{c(1-\beta)},
 \end{aligned} \tag{5.112}$$

as required. $\square$

An important consequence of Lemma 5.3 is that, even with the inclusion of the term $M_1(k)$, the ultimate bounds in (5.112) are identical to those obtained in Lemma 5.1.

In this Section, a nonlinear system is considered with linear measurement model,

$$\mathbf{x}(k+1) = \mathbf{f}(k, \mathbf{x}(k)), \quad \mathbf{z}(k) = \mathbf{C}\mathbf{x}(k) + \boldsymbol{\eta}(k), \tag{5.113}$$

where, $\forall k \geq 0$, $\mathbf{x}(k) \in X \subseteq \mathbb{R}^n$, $\mathbf{f} : \mathbb{N} \times X \mapsto X$, $\mathbf{z} \in \mathbb{R}^m$ contains the system measurements, $\mathbf{C} \in \mathbb{R}^{m \times n}$, and $\boldsymbol{\eta}(k)$ represents bounded measurement errors with

$$\sup_{\{k \geq 0\}} \|\boldsymbol{\eta}(k)\| \leq \eta < \infty. \tag{5.114}$$

The vector $\boldsymbol{\eta}(k)$ could also include bounded uncertainty in the measurement model.

Let an observer for system (5.113) be defined as

$$\hat{\mathbf{x}}(k+1) = \hat{\mathbf{f}}(k, \hat{\mathbf{x}}(k)) + \mathbf{K}(k)[\mathbf{z}(k+1) - \mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}(k)) + \boldsymbol{\xi}(k)], \quad (5.115)$$

where $\hat{\mathbf{f}}$, which is C^p in $\mathbf{x}$, $p \geq 1$, is an estimate of $\mathbf{f}$, and $\mathbf{K}(k) \in \mathbb{R}^{n \times m}$ contains the observer gains which may be time varying. Inclusion of the term $\boldsymbol{\xi}(k)$ in (5.115) is a generalization required for analysis of sliding mode observers proposed in Chapters 6 and 7. The term $\boldsymbol{\xi}(k)$ results from the reaching law method proposed in Section 5.7 and is generally nonzero during the reaching phase, $0 \leq k < k_s$, and zero during the sliding mode, $k \geq k_s$. It is also assumed that $\boldsymbol{\xi}(k)$ is uniformly bounded, i.e.

$$\boldsymbol{\xi}(k) = \mathbf{0}, \quad \forall k \geq k_s \geq 0. \quad (5.116)$$

and

$$\sup_{\{0 \leq k \leq k_s\}} \|\boldsymbol{\xi}(k)\| \leq \xi < \infty. \quad (5.117)$$

Some available knowledge concerning the structure of modelling errors and/or disturbances can be exploited by setting, without loss of generality,

$$\mathbf{f}(k, \mathbf{x}) - \hat{\mathbf{f}}(k, \mathbf{x}) := \mathbf{B}(k)\mathbf{d}(k, \mathbf{x}). \quad (5.118)$$

where, for each $k \geq 0$, $\mathbf{B}(k) \in \mathbb{R}^{n \times q}$ are known matrices, and $\mathbf{d} : \mathbb{N} \times X \mapsto \mathbb{R}^q$. A fundamental assumption employed in this section is that $\|\mathbf{B}(k)\|$ is uniformly bounded and $\|\mathbf{d}(k, \mathbf{x}(k))\|$ is bounded over X uniformly in k . This means that the difference between the true propagation of the system over one sample period from any point in X and the approximate propagation from that same point is bounded in norm. As previously introduced, this property is referred to as, "bounded propagation error over any sample period."

The primary result of this Section, which is presented below in Theorem 5.11 (page 161), provides sufficient conditions under which the observer error is exponentially stable in the sense of Definition 5.2. A domain of exponential stability, exponential state bounds, and ultimate bounds on the observer error are also obtained. Dynamical behavior of the observer error, $\mathbf{e}(k) := \mathbf{x}(k) - \hat{\mathbf{x}}(k)$, is given by

$$\begin{aligned} \mathbf{e}(k+1) &= \mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1) \\ &= \mathbf{f}(k, \mathbf{x}(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}(k)) - \mathbf{K}(k)[\mathbf{z}(k+1) - \mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}(k)) + \boldsymbol{\xi}(k)] \\ &= \mathbf{F}(k) [\mathbf{f}(k, \mathbf{x}(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}(k))] - \mathbf{K}(k) [\boldsymbol{\eta}(k+1) + \boldsymbol{\xi}(k)]. \end{aligned} \quad (5.119)$$

Figure 5.2: Stability of an observer.

where

$$\mathbf{F}(k) := [\mathbf{I}_n - \mathbf{K}(k)\mathbf{C}]. \quad (5.120)$$

The true initial state of the system, $\mathbf{x}(0)$, is unknown, but is perhaps estimated, to obtain $\hat{\mathbf{x}}(0)$, from several measurements taken prior to observer initialization. A fundamental consideration, particularly for nonlinear systems, is the accuracy of initial estimates required to guarantee stability of the observer. Also of primary concern are the selection of the observer gains, $\mathbf{K}(k)$, and the severity of modelling errors and/or disturbances represented in (5.118).

Figure 5.2 illustrates the concept of observer stability considered in this Section. The true state $\mathbf{x}(k)$, for any $k \geq 0$, lies in the set X , the boundary of which is represented by the dark closed curve in Figure 5.2. The true state $\mathbf{x}(k)$ may take on arbitrary values in X . The objective of Theorem 5.11 is to specify sufficient conditions under which a suitable set, Γ_0 ,

can be specified such that if the observer is initialized within Γ_0 , then the observer output remains within the horn-shaped dashed region illustrated in Figure 5.2. This is accomplished by construction of a Lyapunov-like function $V(k, e(k))$ such that, if $\hat{x}(0) \in \Gamma_0$,

$$V(k+1, e(k+1)) \leq M + M_1(k) + \beta V(k, e(k)) \quad \forall k \geq 0, \quad (5.121)$$

with $0 \leq \beta \leq 1$. The results of Lemma 5.3 are then employed to define

$$\Gamma_k := \left\{ \hat{x} \mid V(k, x(k) - \hat{x}) \leq \beta^k V(0, e(0)) + M \sum_{i=0}^{k-1} \beta^i + \sum_{j=0}^{k-1} \beta^j M_1(k-1-j) \right\}. \quad (5.122)$$

and claim, using $M_1(k) = 0 \quad \forall k \geq k_s \geq 0$ which will result from (5.116), that $\forall k > 0$, $\hat{x}(k) \in \Gamma_k$ and

$$\limsup_{k \rightarrow \infty} V(k, x(k) - \hat{x}(k)) \leq \frac{M}{1-\beta}. \quad (5.123)$$

This is the conclusion of Theorem 5.11 stated and proved below.

In Theorem 5.10 of Section 5.5, exponential asymptotic stability of nonlinear discrete-time systems was obtained through the use of a possibly nonlinear time varying transformation of the state, application of the Fundamental Theorem of Integral Calculus (Theorem 5.9, page 148), and the assumption that inequality (5.76) (i.e. $\beta < 1$) was satisfied over all points within some open convex set, Ω , containing the origin. Applications of these fundamental concepts, however, to obtain stability results for the observer considered in this Section require an alternate strategy primarily because stability of the observer is relative to the true trajectory of the system rather than a single stationary point.

In Theorem 5.11 below, a linear time-varying transformation of the state is considered; the possibility of extension to nonlinear diffeomorphic state transformations remains open although initial work towards this goal has led to some difficulties. Another point, which is illustrated in Figure 5.2 and in (5.123), is that the observer output, even after an arbitrarily large time interval, is not necessarily confined to the set X . The Fundamental Theorem of Integral Calculus is again applied, and the hypotheses include an inequality, similar to (5.76) (i.e. $\beta < 1$), which must be satisfied over all points within some open convex set, Ω . This time, however, the set Ω is taken sufficiently large so that, at the very least, it not only contains the set X , but also those points which lie within a *distance* of $M/(1-\beta)$ from the set X , where *distance* is measured in terms of the Lyapunov-like function $V(k, e(k))$. This

Figure 5.3: Application of the Fundamental Theorem of Integral Calculus in stability analysis of an observer.

requirement is illustrated in Figure 5.3. The true state, $x(k)$, could take on values from the boundary of X . However, the conclusions of the Theorem, referred to in (5.123), can only guarantee that in the limit as $k \rightarrow \infty$ the observer output, $\hat{x}(k)$, lies within a *distance* of $d = M/(1 - \beta)$ from $x(k)$. All points along the line segment from $\hat{x}(k)$ to $x(k)$, illustrated in Figure 5.3 by the quantity $w_\lambda(k)$ for $\lambda \in [0, 1]$, must lie within the set Ω in order to obtain the stability result. Another technical complication arises from the possibly time-varying nature of the *distance* measure, namely $V(k, e)$ for $k = 0, 1, 2, \dots$. This is handled with the fundamental definition

$$\{e \in \mathbb{R}^n \mid \exists k_0 \geq 0, V(k_0, e) < d\} = \bigcup_{i=0}^{\infty} \{e \in \mathbb{R}^n \mid V(i, e) < d\}, \quad (5.124)$$

and by recalling that the union of an arbitrary collection of open subsets of $\mathbb{R}^n$ is an open subset of $\mathbb{R}^n$ (see, for example, Marsden, Theorem 2 [144, p. 34]).

Theorem 5.11: Consider system (5.113), the observer (5.115), definitions (5.118) and (5.120), and the constraints on η and ξ in (5.114) and (5.117), respectively. Suppose that for some $\epsilon_2 \geq 0$ (this is required in Chapter 7), $\hat{\mathbf{f}}$ and the observer gains $\mathbf{K}(k)$ can be chosen such that

$$\sup_k \sup_{\mathbf{x} \in X} \|\mathbf{d}(k, \mathbf{x})\| \leq d_0, \quad (5.125)$$

for some known constant d_0 , and there exists a sequence of nonsingular matrices $\mathbf{W}_k \in \mathbb{C}^{n \times n}$, a vector norm, $\|\cdot\|$, and a compatible matrix norm $\|\cdot\|$, with $\|\mathbf{W}_k\|$ and $\|\mathbf{W}_k^{-1}\|$ bounded uniformly in k such that

$$\beta := \sup_k \sup_{\mathbf{v} \in \Omega} \left\| \mathbf{W}_{k+1} \mathbf{F}(k) \left\{ \left[\frac{\partial \hat{\mathbf{f}}(k, \mathbf{x})}{\partial \mathbf{x}} \right]_{\mathbf{x}=\mathbf{v}} \right\} \mathbf{W}_k^{-1} \right\| < 1. \quad (5.126)$$

where, with the definitions

$$V(k, \mathbf{x}) := \|\mathbf{W}_k \mathbf{x}\|. \quad (5.127)$$

$$M := \sup_{\{k \geq 0\}} \|\mathbf{W}_{k+1} \mathbf{F}(k) \mathbf{B}(k)\| d_0 + \sup_{\{k \geq 0\}} \|\mathbf{W}_{k+1} \mathbf{K}(k)\| \eta. \quad (5.128)$$

$$M_1(k) := \|\mathbf{W}_{k+1} \mathbf{K}(k) \xi(k)\|, \quad \text{and} \quad (5.129)$$

$$M_0 := M + \sup_{\{k \geq 0\}} \{M_1(k)\}. \quad (5.130)$$

the open convex set Ω satisfies, for some $\epsilon_1 > \epsilon_2 \geq 0$,

$$X_b := \{\mathbf{x} | \exists \mathbf{u} \in \mathbb{R}^n, \exists k_0 \geq 0, V(k_0, \mathbf{u}) < \frac{M_0}{1-\beta} + \epsilon_1, (\mathbf{x} + \mathbf{u}) \in X\} \subset \Omega. \quad (5.131)$$

In this case define

$$d_s := \sup_d [\{\hat{\mathbf{x}} | \exists \mathbf{e} \in \mathbb{R}^n, \exists k_0 \geq 0, V(k_0, \mathbf{e}) < d + \epsilon_2, \hat{\mathbf{x}} + \mathbf{e} \in X\} \subset \Omega]. \quad (5.132)$$

$$X_s := \{\hat{\mathbf{x}} | \exists \mathbf{e} \in \mathbb{R}^n, \exists k_0 \geq 0, V(k_0, \mathbf{e}) < d_s + \epsilon_2, \hat{\mathbf{x}} + \mathbf{e} \in X\}. \quad (5.133)$$

$$X_k := \{\hat{\mathbf{x}} | V(k, \mathbf{x}(k) - \hat{\mathbf{x}}) < d_s\}, \quad \text{and} \quad (5.134)$$

$$w := \sup_{\{k \geq 0\}} \|\mathbf{W}_k^{-1}\|. \quad (5.135)$$

Then $\exists \beta_s, 0 \leq \beta_s \leq \beta$, such that $\forall \mathbf{x}(0) \in X$, if $\hat{\mathbf{x}}(0) \in X_0$ then $\forall k \geq 0, \hat{\mathbf{x}}(k) \in X_k$ and the observer error is exponentially bounded according to

$$\|\mathbf{e}(k)\| \leq \beta_s^k \|\mathbf{W}_k^{-1}\| \|\mathbf{W}_0 \mathbf{e}(0)\| + \|\mathbf{W}_k^{-1}\| M \sum_{i=0}^{k-1} \beta_s^i + \sum_{j=0}^{k-1} \beta_s^j M_1(k-1-j). \quad (5.136)$$

Figure 5.4: Illustration of relationships between the sets X , X_b , X_s , X_k and Ω in stability analysis of an observer.

and has ultimate bounds given by

$$\limsup_{k \rightarrow \infty} \|e(k)\| \leq Mw/(1 - \beta_s). \quad (5.137)$$

Proof: Parameters ϵ_1 in (5.131), and ϵ_2 in (5.132) and (5.133), are included to facilitate analysis of numerical implementations of the ideal observer form (5.115), such as that proposed in Chapter 7. For some $\epsilon_2 \geq 0$, the hypothesis must be satisfied with $\epsilon_1 > \epsilon_2$. For exact implementation of the observer, parameter ϵ_2 can be set to zero and the stability result holds, as shown below. The use of ϵ_2 in analysis of an approximate implementation of the observer is detailed in results of Chapter 7.

Relationships between the sets X , X_b , X_s , X_k and Ω are illustrated in Figure 5.4. The fact that $X \subset X_b$ follows directly from the definition of X_b in (5.131). The hypothesis (5.131) requires that $X_b \subset \Omega$ where Ω is an open convex set. The set X_b represents the

smallest region that must, under these hypothesis, be contained within the convex set Ω . Because the hypothesis (5.126) may be satisfied over a convex set Ω which is much larger than X_b , a larger stability domain for observer initialization, namely X_0 from (5.134), can be obtained through the definition of d_s in (5.132). The set X_s in (5.133) is the largest set of the same form as X_b in (5.131) which is completely contained within Ω . Inclusion of parameter $\epsilon_1 > \epsilon_2 \geq 0$ in hypothesis (5.131) implies that

$$d_s > M_0/(1 - \beta). \quad (5.138)$$

Hence $X \subset X_b \subset X_s \subseteq \Omega$. In addition, $\forall k \geq 0$, $X_k \subset X_s$. This means that if the elliptical-shaped² set X_k , illustrated in Figure 5.4, is centred on any point $\mathbf{x} \in X$, then X_k remains enclosed within $X_s \subset \Omega$.

The method of proof is by induction: assuming that, for some $k_0 \geq 0$, $\hat{\mathbf{x}}(k_0) \in X_{k_0}$, the objective is to show that $V(k_0 + 1, \mathbf{x}(k_0 + 1) - \hat{\mathbf{x}}(k_0 + 1))$ satisfies the hypothesis of Lemma 5.3 and that this implies that $\hat{\mathbf{x}}(k_0 + 1) \in X_{k_0+1}$.

Define

$$\beta_s := \sup_k \sup_{\mathbf{v} \in X_s} \left\| \mathbf{W}_{k+1} \mathbf{F}(k) \left\{ \left[\frac{\partial \hat{\mathbf{f}}(k, \mathbf{x})}{\partial \mathbf{x}} \right]_{\mathbf{x}=\mathbf{v}} \right\} \mathbf{W}_k^{-1} \right\|, \quad (5.139)$$

which satisfies $0 \leq \beta_s \leq \beta$, and hence

$$d_s > \frac{M_0}{1 - \beta} \geq \frac{M_0}{1 - \beta_s}. \quad (5.140)$$

For all $k \geq 0$, $\mathbf{x}(k) \in X \subset \Omega$. If, for some $k_0 \geq 0$, $\hat{\mathbf{x}}(k_0) \in X_{k_0}$, then the observer error $\mathbf{e}(k_0 + 1)$ satisfies

$$\begin{aligned} \mathbf{e}(k_0 + 1) &= \mathbf{F}(k_0) \{ \mathbf{f}(k_0, \mathbf{x}(k_0)) - \hat{\mathbf{f}}(k_0, \hat{\mathbf{x}}(k_0)) \} - \mathbf{K}(k_0) [\boldsymbol{\eta}(k_0 + 1) + \boldsymbol{\xi}(k_0)] \\ &= \mathbf{F}(k_0) \{ \mathbf{f}(k_0, \mathbf{x}(k_0)) - \hat{\mathbf{f}}(k_0, \mathbf{x}(k_0)) \} \\ &\quad + \mathbf{F}(k_0) \{ \hat{\mathbf{f}}(k_0, \mathbf{x}(k_0)) - \hat{\mathbf{f}}(k_0, \hat{\mathbf{x}}(k_0)) \} \\ &\quad - \mathbf{K}(k_0) \boldsymbol{\eta}(k_0 + 1) - \mathbf{K}(k_0) \boldsymbol{\xi}(k_0) \\ &= \mathbf{F}(k_0) \{ \mathbf{f}(k_0, \mathbf{x}(k_0)) - \hat{\mathbf{f}}(k_0, \mathbf{x}(k_0)) \} \\ &\quad + \mathbf{F}(k_0) \left\{ \int_0^1 \frac{\partial \hat{\mathbf{f}}}{\partial \mathbf{x}}(k_0, \mathbf{w}_\lambda(k_0)) d\lambda \right\} \mathbf{e}(k_0) \\ &\quad - \mathbf{K}(k_0) \boldsymbol{\eta}(k_0 + 1) - \mathbf{K}(k_0) \boldsymbol{\xi}(k_0). \end{aligned}$$

²Level sets of the vector norm need not produce hyperellipsoids: X_k is shown as elliptical for illustration purposes only.

where the second equality follows by adding and subtracting $\hat{\mathbf{f}}(k_0, \mathbf{x}(k_0))$ and the final equality results from application of the Fundamental Theorem of Integral Calculus with

$$\mathbf{w}_\lambda(k_0) := [\lambda \mathbf{x}(k_0) + (1 - \lambda) \hat{\mathbf{x}}(k_0)]. \quad (5.141)$$

Although the set X_s itself need not be convex, both $\mathbf{x}(k_0)$ and $\hat{\mathbf{x}}(k_0)$ are elements of the set X_{k_0} which is convex by the properties of the vector norm. Hence for any $\lambda \in [0, 1]$, $\mathbf{w}_\lambda(k_0) \in X_{k_0} \subset X_s$. Given the sequence of nonsingular matrices $\{\mathbf{W}_k, k = 0, 1, 2, \dots\}$,

$$\begin{aligned} V(k_0 + 1, \mathbf{e}(k_0 + 1)) &:= \|\mathbf{W}_{k_0+1} \mathbf{e}(k_0 + 1)\| \\ &\leq \|\mathbf{W}_{k_0+1} \mathbf{F}(k_0) \mathbf{B}(k_0)\| d_0 + \|\mathbf{W}_{k_0+1} \mathbf{K}(k_0)\| \eta \\ &\quad + \|\mathbf{W}_{k_0+1} \mathbf{K}(k_0) \boldsymbol{\xi}(k_0)\| \\ &\quad + \int_0^1 \left\| \mathbf{W}_{k_0+1} \mathbf{F}(k_0) \frac{\partial \hat{\mathbf{f}}}{\partial \mathbf{x}}(k_0, \mathbf{w}_\lambda(k_0)) \mathbf{W}_{k_0}^{-1} \right\| d\lambda \|\mathbf{W}_{k_0} \mathbf{e}(k_0)\| \\ &\leq M + M_1(k_0) + \beta_s V(k_0, \mathbf{e}(k_0)). \end{aligned} \quad (5.142)$$

Hence if, for some $k_0 \geq 0$, $\hat{\mathbf{x}}(k_0) \in X_{k_0}$, then (5.142) holds. It remains to be shown that (5.142) implies that $\hat{\mathbf{x}}(k_0 + 1) \in X_{k_0+1}$. But $\hat{\mathbf{x}}(k_0) \in X_{k_0}$ implies that $V(k_0, \mathbf{e}(k_0)) < d_s$. Then there are two cases to consider: either Case 1) $M_0/(1 - \beta_s) < V(k_0, \mathbf{e}(k_0)) < d_s$, or Case 2) $0 \leq V(k_0, \mathbf{e}(k_0)) \leq M_0/(1 - \beta_s)$. For Case 1, and using (5.142),

$$\begin{aligned} V(k_0 + 1, \mathbf{e}(k_0 + 1)) - V(k_0, \mathbf{e}(k_0)) &\leq M + M_1(k_0) - (1 - \beta_s) V(k_0, \mathbf{e}(k_0)) \\ &< M + M_1(k_0) - (1 - \beta_s) \left[\frac{M_0}{1 - \beta_s} \right] \\ &= M + M_1(k_0) - M_0 \leq 0. \end{aligned} \quad (5.143)$$

which implies that

$$V(k_0 + 1, \mathbf{e}(k_0 + 1)) < V(k_0, \mathbf{e}(k_0)) < d_s. \quad (5.144)$$

For Case 2, and again using (5.142),

$$\begin{aligned} V(k_0 + 1, \mathbf{e}(k_0 + 1)) &\leq M + M_1(k_0) + \beta_s \left[\frac{M_0}{1 - \beta_s} \right] \\ &\leq M_0 + \beta_s \left[\frac{M_0}{1 - \beta_s} \right] \\ &= \frac{M_0}{1 - \beta_s} < d_s. \end{aligned} \quad (5.145)$$

In either case, $V(k_0 + 1, e(k_0 + 1)) < d_s$ which implies that $\hat{x}(k_0 + 1) \in X_{k_0+1}$. By induction, if $\hat{x}(0) \in X_0$, then $\forall k \geq 0$, $\hat{x}(k) \in X_k$ and (5.142) holds. Hence $V(k, e(k))$ satisfies the hypothesis of Lemma 5.3 and the exponential bounds follow with $c = 1/w$, $\phi(\|e\|) = \|e\|$, M as defined in (5.128), $M_1(k)$ defined as in (5.129), and with “ β ” of Lemma 5.3 taken as $\beta_s < 1$. $\square$

The primary hypothesis (5.126) of Theorem 5.11 is based on properties of the approximation $\hat{f}$ to the true system, on the observer gains $K(k)$, and on the measurement matrix C . The designer has control over specification of both $\hat{f}$ and $K(k)$. The ultimate bounds obtained in Theorem 5.11 are independent of the vector $\xi(k)$ which appears in the observer. This property is much like exponential forgetting of initial conditions, since $\xi(k)$ is assumed to be uniformly bounded and vanishing for all $k \geq k_s \geq 0$.

The stability result provided by Theorem 5.11 does not appear to depend explicitly on an observability requirement; even if the system is not observable, it may still be possible to estimate the state to within some finite bounds. A trivial example results when it is known that $\sup_{k,x} \|f(k, x)\| = \delta < \infty$ in which case an “observer” $\hat{x}(k) \equiv 0$ is exponentially bounded in the sense that $\|e(k)\| \leq \delta$, although this result is of little practical use. It may not be possible, however, to obtain asymptotic stability from methods proposed in Theorem 5.11, which would result in the case $M \equiv 0$, without the observability requirement as illustrated in the following example.

Example 5.5: Consider a linear system $x(k+1) = \Phi x(k)$ with scalar measurements $z(k) = Cx(k)$, where

$$\Phi = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad C = [0, 1]. \quad (5.146)$$

and T is the sample period. This is a constant velocity system in which only velocity is measured and hence the system is not observable. An observer is proposed of the form

$$\hat{x}(k+1) = \Phi \hat{x}(k) + K[z(k+1) - C\Phi \hat{x}(k)], \quad (5.147)$$

where $K = [k_1, k_2]^T$ are the observer gains. In this case $d_0 = 0$, and hence a triplet $(W, K, \|\cdot\|)$ is required such that $\beta := \|W[I_2 - KC]\Phi W^{-1}\| < 1$. But, from Item 2

of the proof of Theorem 5.7 (page 140), this is not possible since the spectral radius, $\rho(\cdot)$, satisfies

$$\rho([I_2 - KC]\Phi) = \rho\left(\begin{bmatrix} 1 & T - k_1 \\ 0 & 1 - k_2 \end{bmatrix}\right) \geq 1. \quad (5.148)$$

Therefore the error equation cannot be shown to be asymptotically stable. $\square$

Example 5.5 considered the possibility of obtaining exponential asymptotic stability in the absence of modelling error. From Theorem 5.11, and assuming no measurement errors ($\eta = 0$), exponential asymptotic stability results if propagation errors over one sample period, (Equation 5.118, repeated)

$$\mathbf{f}(k, \mathbf{x}) - \hat{\mathbf{f}}(k, \mathbf{x}) = \mathbf{B}(k)\mathbf{d}(k, \mathbf{x}), \quad (5.149)$$

lie in the null space of $\mathbf{F}(k)$, where (Equation 5.120, repeated)

$$\mathbf{F}(k) := [\mathbf{I}_n - \mathbf{K}(k)\mathbf{C}], \quad (5.150)$$

since in this case $M = 0$. This is precisely the robustness property that is a primary objective of sliding mode observers and was previously discussed in Section 3.5. Further discussions of this important design objective are included in Chapters 6 and 7.

The following example considers nominally linear systems for which the hypothesis of Theorem 5.11 can be satisfied globally. The objective is to treat a simple two-dimensional case for visualization and to demonstrate the use of numerical methods in the computation of a matrix $\mathbf{W}$ which not only satisfies the hypothesis, but gives reasonably tight error bounds. Further applications of the results of Theorem 5.11 appear in Chapters 6 and 7.

Example 5.6: Consider $\mathbf{x} = [x_1, x_2]^T$ and the continuous-time, nominally constant velocity system which is acted upon by a bounded disturbance, $\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{H}v(t)$, where

$$\mathbf{A} := \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{H} := \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \quad (5.151)$$

and $|v(t)| \leq 1$. Discretization with a sample period of $T = 0.1\text{sec.}$ yields $\mathbf{x}(k+1) = \Phi\mathbf{x}(k) + \xi(k)$, where $\Phi = \exp(\mathbf{A}T)$ and $\xi(k)$ satisfies, $\forall \Gamma \in \mathbb{C}^{2 \times 2}$, $\|\Gamma\xi(k)\| \leq \|\Gamma\mathbf{B}\|$, where $\mathbf{B} = [\int_0^T \exp(\mathbf{A}\mu)d\mu]\mathbf{H}$. Assume the measurements are $z(k) = \mathbf{C}\mathbf{x}(k)$ with

$\mathbf{C} := [1, 0]$, and define an observer as $\hat{\mathbf{x}}(k+1) = \Phi\hat{\mathbf{x}}(k) + \mathbf{K}\{z(k+1) - \mathbf{C}\Phi\hat{\mathbf{x}}(k)\}$. With gains $\mathbf{K} := [0.8, 3]^T$ and $\mathbf{F} := [\mathbf{I}_2 - \mathbf{K}\mathbf{C}]$, $\lambda(\mathbf{F}\Phi) = \{0.4, 0.5\}$. With $\|\cdot\| = \|\cdot\|_2$, $\mathbf{W} = [w_{ij}]$ can be taken as upper triangular with $w_{11} = 1$, and $w_{22} > 0$. From Theorem 5.11, provided $\beta := \|\mathbf{W}\mathbf{F}\Phi\mathbf{W}^{-1}\|_2 < 1$, the ultimate bounds are

$$\limsup_{k \rightarrow \infty} \|\mathbf{e}(k)\|_2 \leq E(w_{12}, w_{22}) := \|\mathbf{W}^{-1}\|_2 \frac{M}{(1 - \beta)}, \quad (5.152)$$

where $M := \|\mathbf{W}\mathbf{F}\mathbf{B}\|_2$.

Figure 5.5a shows the surface $E(w_{12}, w_{22})$. The constraint $\beta < 1$ is satisfied inside the two cup-shaped regions whose contours are shown in Figure 5.5b. Numerical minimization of $E(w_{12}, w_{22})$ using an initial estimate derived from Method 1 of Section 5.4 gives $E_{\min} = 0.24$ with $w_{12} = -0.041$ and $w_{22} = \pm 0.063$. These values give $\beta = 0.61$ which is somewhat larger than $\inf_{\mathbf{W}} \beta = 0.5$. As shown in Figure 5.5c, for gains $\mathbf{K} = [1, 5]^T$ which result in a singular $\mathbf{F}$ and a reduced order Luenberger observer, the minimum $E_{\min} = 0.15$ occurs for $w_{12} = -6.01(10^{-7})$ and $w_{22} = \pm 2.69(10^{-4})$. Numerical minimization of E in this case can be poorly behaved unless the constraint $\beta < 1$ is explicitly considered since small variations in w_{12} or w_{22} from the minimum point can result in migration outside of the constraint. Figure 5.5d shows the observer error $\|\mathbf{e}(k)\|_2$ and the very tight bounds computed from (5.136) for the reduced-order Luenberger observer with $\mathbf{K} = [1, 5]^T$ for the case $v(t) = \cos(t)$, $\mathbf{x}(0) = [0, 1]^T$ and $\hat{\mathbf{x}}(0) = [1, 0]^T$. $\square$

As suggested in the above example, the search for a suitable matrix $\mathbf{W}$ may be simplified through numerical methods. If $\|\cdot\|$ is a unitarily invariant matrix norm, for example, then by setting $w_{11} = 1$, $\mathbf{W}$ can be formed with $N = n(n+1)/2 - 1$ variables for which $E(\mathbf{W})$ will have even symmetry in the remaining $n-1$ diagonal entries of $\mathbf{W}$. Cases in which $\mathbf{W} \in \mathbb{C}^{n \times n}$ can be treated similarly with $2N$ real variables.

5.7 A Reaching Law Method for Discrete-time Sliding Mode Systems

As previously discussed in Section 3.5 (pages 86 and 89), the trajectories of systems with sliding modes generally consist of two parts, representing two modes of the system: the

Figure 5.5: Observer error bound dependence on matrix W : a) bound surface for $K = [0.8, 3]^T$; b) contours for surface in plot a); c) contours of bound surface for $K = [1, 5]^T$; d) observer error (solid) and bounds (dashed) for simulation with $K = [1, 5]^T$. Plots b) and c) show the initial ($\times$) and final ($*$) points for numerical minimization of $E(w_{12}, w_{22})$.

reaching mode in which controlled trajectories lie off of the sliding surface, and the *sliding mode* during which trajectories are constrained to lie on the sliding surface. Design of sliding mode systems, then, naturally is carried out in stages: the sliding surface is selected to achieve the desired sliding dynamics, and parameters which affect the reaching dynamics are selected to make the sliding surface attractive. Gao and Hung [123] recently proposed a *reaching law* method for design of continuous-time sliding mode control systems in which the controller is designed so that the switching function $s(\mathbf{x}(t))$ satisfies a reaching law of the form

$$\dot{s}(\mathbf{x}) = -b \operatorname{sgn}(s) - \gamma \phi(s), \quad (5.153)$$

where $b > 0$, $\gamma > 0$, and $s\phi(s) > 0$ with $\phi(0) = 0$. In this scalar illustration of their approach, the sliding surface is given by $s(\mathbf{x}) = 0$ which is attained, under the reaching law (5.153) with $\phi(s) := s$, in finite time (see Equation 3.30, page 89).

In this Section a reaching law method for creation of sliding modes in discrete-time systems is proposed and analyzed. Although the basic approach could also have applications in discrete-time control problems, the emphasis here is on the observer problem. Given a model of the discrete-time dynamics and a model of the discrete-time measurements, the designer specifies a surface or switching function $s_k := s(\mathbf{x}(k), \hat{\mathbf{x}}(k))$, where $\mathbf{x}(k)$ represents the true system state while $\hat{\mathbf{x}}(k)$ is the state estimate provided by the observer. The sliding surface is defined as $s(\mathbf{x}(k), \hat{\mathbf{x}}(k)) = 0$. The objective of the designer is then to develop an appropriate observer form so that a discrete-time reaching law of the form

$$s_{k+1} = g(k, s_k) \quad (5.154)$$

is satisfied for all $k \geq 0$, where $g(k, s_k)$ is completely specified by the designer. If the reaching law (5.154) is asymptotically stable, then $\lim_{k \rightarrow \infty} s_k = 0$ which implies an asymptotic approach to the surface. If the reaching law (5.154) is satisfied for all $k \geq 0$, then the designer has the option of selecting some time $k_s \geq 0$ and setting $g(k, s_k) = 0$ for all $k \geq k_s$. In this case $s(\mathbf{x}(k), \hat{\mathbf{x}}(k)) = 0$ for all $k \geq k_s$.

In Chapters 6 and 7 observer forms are proposed for various classes of discrete-time systems along with surface functions $s(\mathbf{x}(k), \hat{\mathbf{x}}(k))$ such that the reaching law (5.154) is satisfied for all $k \geq 0$. This property of the proposed observers gives the designer a great

deal of flexibility in specifying the error dynamics during the approach to the sliding surface which is the subject of the remainder of this Section.

A natural extension of the continuous-time approach in (5.153) to the discrete-time case is to replace the signum function with a saturation function

$$\text{sat}(s, \epsilon) := \min \left\{ \frac{|s|}{\epsilon}, 1 \right\} \text{sgn}(s). \quad (5.155)$$

where $\epsilon > 0$. This approach remains consistent with the requirements of Theorem 4.2 (page 122) which states that continuity across the sliding surface is a necessary condition for existence of a convergent quasi-sliding regime. For the vector case, with $\mathbf{s}, \boldsymbol{\epsilon} \in \mathbb{R}^q$, and $\epsilon_i > 0$ for $i = 1, 2, \dots, q$, define

$$\mathbf{Sat}(\mathbf{s}, \boldsymbol{\epsilon}) := [\text{sat}(s_1, \epsilon_1), \text{sat}(s_2, \epsilon_2), \dots, \text{sat}(s_q, \epsilon_q)]^T. \quad (5.156)$$

The proposed reaching law has the form

$$\mathbf{s}_{k+1} = \mathbf{g}(k, \mathbf{s}_k) \quad (5.157)$$

where

$$\mathbf{g}(k, \mathbf{s}_k) := \Lambda_k \mathbf{Sat}(\mathbf{s}_k, \boldsymbol{\epsilon}) + \Gamma_k \mathbf{s}_k. \quad (5.158)$$

The objective now is to specify sufficient conditions on the matrices Λ_k and Γ_k and the vector $\boldsymbol{\epsilon}$ such that the origin of system (5.157) is globally exponentially asymptotically stable.

Definition 5.7: Let $\|\cdot\|_p$ be any vector norm on $\mathbb{R}^q$ which satisfies, for all $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2 \in \mathbb{R}^q$, if $|\xi_{1,i}| \leq |\xi_{2,i}|$ for $i = 1, 2, \dots, q$, then $\|\boldsymbol{\xi}_1\|_p \leq \|\boldsymbol{\xi}_2\|_p$. $\square$

Note that any of the Hölder vector norms,

$$\|\mathbf{y}\|_p := \left[\sum_{i=1}^q |y_i|^p \right]^{1/p}, \quad (5.159)$$

for $p \geq 1$, satisfy the property of Definition 5.7. However, elliptical or derived norms, such as $\|\boldsymbol{\xi}\|_{\mathbf{W}} := \|\mathbf{W}\boldsymbol{\xi}\|_2$, with $\mathbf{W} \in \mathbb{R}^{q \times q}$ nonsingular, need not satisfy this property.

Theorem 5.12: Consider the function, $g : \mathbb{N}_0 \times \mathbb{R}^q \mapsto \mathbb{R}^q$, defined as

$$g(k, \xi) := \Lambda_k \text{Sat}(\xi, \epsilon) + \Gamma_k \xi. \quad (5.160)$$

Let $\|\cdot\|_y$ be a vector norm satisfying Definition 5.7, and let $\|\cdot\|_z$ be a compatible matrix norm. If $\sup_{\{k \geq 0\}} \|\Gamma_k\|_z < 1$, and

$$\min\{\epsilon_i\} > \sup_{\{k \geq 0\}} \left\{ \frac{\|\Lambda_k\|_z}{1 - \|\Gamma_k\|_z} \right\}. \quad (5.161)$$

then

1. $g(k, \xi)$ is a contraction on $\mathbb{R}^q$ with respect to the norm $\|\cdot\|_y$;
2. $\|s_k\|_y \leq \beta^k \|s_0\|_y$, where

$$\beta := \sup_{\{k \geq 0\}} \left\{ \frac{\|\Lambda_k\|_z}{\min\{\epsilon_i\}} + \|\Gamma_k\|_z \right\} < 1; \quad (5.162)$$

3. The origin of system (5.157) (5.158) is globally exponentially asymptotically stable.

Proof: For Statement 1, first note that the scalar saturation function is Lipschitz continuous, i.e. for any $\xi_1, \xi_2, \epsilon \in \mathbb{R}, \epsilon > 0$,

$$|\text{sat}(\xi_1, \epsilon) - \text{sat}(\xi_2, \epsilon)| \leq \frac{1}{\epsilon} |\xi_1 - \xi_2|. \quad (5.163)$$

Hence, using the properties of the vector norm $\|\cdot\|_y$, for any $\xi_1, \xi_2, \epsilon \in \mathbb{R}^q$ with $\epsilon_i > 0$, $i = 1, 2, \dots, q$,

$$\|\text{Sat}(\xi_1, \epsilon) - \text{Sat}(\xi_2, \epsilon)\|_y \leq \left\| \begin{bmatrix} \frac{1}{\epsilon_1} (\xi_{1,1} - \xi_{2,1}) \\ \vdots \\ \frac{1}{\epsilon_q} (\xi_{1,q} - \xi_{2,q}) \end{bmatrix} \right\|_y \leq \frac{1}{\min\{\epsilon_i\}} \|\xi_1 - \xi_2\|_y. \quad (5.164)$$

Then

$$\begin{aligned} \|g(k, \xi_1) - g(k, \xi_2)\|_y &= \|\Lambda_k [\text{Sat}(\xi_1, \epsilon) - \text{Sat}(\xi_2, \epsilon)] + \Gamma_k [\xi_1 - \xi_2]\|_y \\ &\leq \left[\frac{\|\Lambda_k\|_z}{\min\{\epsilon_i\}} + \|\Gamma_k\|_z \right] \|\xi_1 - \xi_2\|_y \\ &\leq \beta \|\xi_1 - \xi_2\|_y. \end{aligned} \quad (5.165)$$

where, by the hypothesis (5.161),

$$\beta := \sup_{\{k \geq 0\}} \left\{ \frac{\|\Lambda_k\|_b}{\min\{\epsilon_1\}} + \|\Gamma_k\|_b \right\} < 1. \quad (5.166)$$

Statement 2 results from application of the above result to system (5.157) (5.158) with $\xi_1 = s$ and $\xi_2 = 0$. Then $g(k, \xi_2) = g(k, 0) = 0$, and

$$\|s_k\|_b = \|g(k-1, s(k-1))\|_b \leq \beta \|s(k-1)\|_b, \quad (5.167)$$

which can be repeatedly applied to give the result $\|s_k\|_b \leq \beta^k \|s_0\|_b$. Since $\beta < 1$ by the hypothesis, Statement 3 then follows. $\square$

If Λ_k and Γ_k are, for all $k \geq 0$, nonzero matrices satisfying the hypothesis of Theorem 5.12, then convergence to the surface is asymptotic and a global convergent quasi-sliding regime exists on $s = 0$ in the sense of Definition 4.1 (Sarıptürk). A global discrete-time sliding regime on $s = 0$, in the sense of Definition 4.2 (Drakunov), can also be created according to the following Corollary which is used repeatedly in Chapters 6 and 7. Recall that here the emphasis is on properties of the reaching law *assuming* that the observer form is such that the surface function s_k under the error dynamics of this particular observer form satisfy the reaching law (5.157). This is the case for observer forms proposed in Chapters 6 and 7.

Corollary 5.2: If the reaching law (5.157) (5.158) satisfies the hypothesis of Theorem 5.12, then a global discrete-time sliding regime, in the sense of Definition 4.2 (Drakunov), can be created on $s = 0$.

Proof: It is required to show that given any s_0 and any tolerance $\epsilon_s > 0$, a design methodology exists such that if, for some $k_s > 0$, $\|s_{k_s}\| < \epsilon_s$, then the surface function s_k is identically zero from time $k_s + 1$ onward. Define β as in (5.162) and

$$k_s := \max \left\{ \left\lceil \frac{\ln(\epsilon_s) - \ln(\|s_0\|_b)}{\ln(\beta)} \right\rceil, 0 \right\}, \quad (5.168)$$

where $\lceil \cdot \rceil$ returns the smallest integer larger than its argument. Theorem 5.12 then guarantees that $\|s_k\|_b \leq \epsilon_s$ for all $k > k_s$. Setting $\Lambda_k = 0$ and $\Gamma_k = 0$ for $k \geq k_s$, ensures that $s_k = 0$ for all $k > k_s$. $\square$

Nonzero matrices Λ_k and Γ_k would be selected in order to influence the transient response of an observer immediately following initialization. If, for all $k \geq 0$, $g(k, s_k) = 0$, then an observer satisfying the reaching law would execute one step to the surface from any initial point. This approach is reasonable provided initialization errors are small. With larger initialization errors, however, a single step to the surface can result in large transients in higher time derivative channels which may be undesirable. This behavior is illustrated in the simulation results of Chapter 7. In the interest of reducing initial transients, a natural approach is to choose Λ_k and Γ_k as non-zero diagonal matrices in which case the system of equations (5.154) are decoupled. The following Example considers the scalar case and provides further insight into selection of the diagonal entries of Λ_k and Γ_k .

Example 5.7: Consider the scalar case of dynamics in (5.154) and (5.158), with $\Lambda_k \equiv \lambda$, $\Gamma_k \equiv \gamma$ and $\epsilon \equiv \epsilon$. In this case, the hypothesis of Theorem 5.12 require that $\gamma \in (-1, 1)$ and $\epsilon > |\lambda|/(1 - |\gamma|)$, and the reaching dynamics are given by

$$s_{k+1} = \lambda \text{sat}(s_k, \epsilon) + \gamma s_k. \quad (5.169)$$

Using the definition of the $\text{sat}(\cdot, \cdot)$ function given in (5.155),

$$\text{sgn}(s_{k+1})\text{sgn}(s_k) = \begin{cases} +1, & \text{sgn}(\gamma) \max(|s_k|, \epsilon) > -\frac{\lambda}{\gamma} \text{sgn}(\lambda) \\ -1, & \text{sgn}(\gamma) \max(|s_k|, \epsilon) < -\frac{\lambda}{\gamma} \text{sgn}(\lambda) \end{cases} \quad (5.170)$$

Possible approaches to the surface, which are illustrated in Figure 5.6, are determined by examination of several cases.

Case 1: If $\gamma \in (-1, 0)$ and either $\lambda < 0$ or both $\lambda > 0$ and

$$\epsilon > \max \left\{ \frac{|\lambda|}{|\gamma|}, \frac{|\lambda|}{1 - |\gamma|} \right\}, \quad (5.171)$$

then $\text{sgn}(s_{k+1})\text{sgn}(s_k) = -1$ for all k which results in the approach shown in Figure 5.6a. This approach to the surface exhibits the typical chattering normally associated with discrete-time sliding mode systems, but could give rise to large initial transients in the observer, since the state must cross the surface in each sample period even if the initial estimate is far from the surface, which would be undesirable.

Case 2: If $\gamma \in (-1, 0)$ and $\lambda > 0$ and

$$\frac{|\lambda|}{|\gamma|} > \epsilon > \frac{|\lambda|}{1 - |\gamma|}. \quad (5.172)$$

Figure 5.6: Possible approaches to the surface with the reaching law method.

then

$$\text{sgn}(s_{k+1})\text{sgn}(s_k) = \begin{cases} +1, & |s_k| < \frac{|\lambda|}{|\gamma|} \\ -1, & |s_k| > \frac{|\lambda|}{|\gamma|} \end{cases} \quad (5.173)$$

and the approach is as shown in Figure 5.6b. Again this approach would likely be undesirable due to the possibility of large initial transients.

Case 3: If $\gamma \in (0, 1)$ and either $\lambda > 0$ or both $\lambda < 0$ and

$$\epsilon > \max \left\{ \frac{|\lambda|}{|\gamma|}, \frac{|\lambda|}{1 - |\gamma|} \right\}, \quad (5.174)$$

then $\text{sgn}(s_{k+1})\text{sgn}(s_k) = 1$ for all k which results in the approach shown in Figure 5.6c.

In this case s_k retains the same sign for all k and asymptotically approaches the surface.

Case 4: If $\gamma \in (0, 1)$ and $\lambda < 0$ and

$$\frac{|\lambda|}{|\gamma|} > \epsilon > \frac{|\lambda|}{1 - |\gamma|}, \quad (5.175)$$

then

$$\text{sgn}(s_{k+1})\text{sgn}(s_k) = \begin{cases} +1, & |s_k| > \frac{|\lambda|}{|\gamma|} \\ -1, & |s_k| < \frac{|\lambda|}{|\gamma|} \end{cases}, \quad (5.176)$$

and the approach is as shown in Figure 5.6d. In this case, s_k retains the same sign until $|s_k| < |\lambda|/|\gamma|$, and thereafter the state crosses the surface in each sample period.

For the problem of reducing initial observer transients, selection of parameters which give either the approach of Case 3, shown in Figure 5.6c, or the approach of Case 4, shown in Figure 5.6d, is recommended. $\square$

5.8 Chapter Summary

This Chapter presented a novel approach to obtain exponential stability of discrete-time systems, stability domains, and explicit bounds on the state. The proposed approach is based on the specification of suitable vector and matrix norms, and a state transformation under which the transformed system represents, at least locally, a contraction with respect to the prescribed norms. This approach together with the use of vector norm Lyapunov, or Lyapunov-like functions in stability analysis can not only demonstrate exponential stability, but may also provide practical information in terms of a domain of exponential stability and exponential bounds on the state.

Applications of this approach in stability analysis of discrete-time observers are of particular interest in that some measure of initialization errors under which the observer error is exponentially bounded can be determined. The availability of such information, however, is dependent on the ability to specify suitable state transformations which, with respect to the chosen vector and matrix norms, not only satisfy the sufficient conditions for stability, but also give a sufficiently large stability domain and tight enough bounds to be of practical use. Further analysis is required to determine if numerical methods, such as the numerical minimization employed in Example 2, can be applied in the search for state transformations with the required properties for more general time-varying and/or nonlinear systems. A promising approach may be to begin with stability by first approximation (Corollary 5.1) followed by expansion of stability domains through numerical searches for appropriate state transformations.

The reaching law method for design of discrete-time sliding mode systems described in this Chapter is used throughout the following two Chapters which develop several forms of sliding mode observers for various classes of systems. The error dynamics for these ob-

servers can be expressed in a form to which the stability results of this Chapter can be applied. Stability properties of these observers are obtained under the assumption that the true system and parameters specified for the observer satisfy the hypothesis of Theorem 5.11.

Chapter 6

Sliding Mode Observers for Systems With Linear Measurement Models

6.1 Introduction

In Chapter 4, a simple sliding mode observer was presented which is a direct extension and application of methods proposed for continuous-time systems to the discrete-time case. For linear systems in the absence of modelling errors, Theorem 4.1 demonstrated that generation of a quasi-sliding regime with this method is always possible, and application of Theorem 5.7 in Example 5.3 resulted in exponential and ultimate bounds on the observer error. Simulation results of simple systems, shown in Section 4.3, suggest that this approach can yield good performance in the presence of modelling errors. Further analysis presented in Section 4.3, however, illustrated some of the primary difficulties associated with this form of the observer: the quasi-sliding regime is only of finite extent (Theorem 4.1); due to discontinuous switching, the quasi-sliding regime cannot be convergent (Example 4.2 and Theorem 4.2); and that extensions to the case of nonlinear systems with unmodelled dynamics is not immediate. Corollary 4.1 gave the form of an observer for the linear system which creates a global discrete-time sliding regime, in the sense of Definition 4.2 (Drakunov), and also results, analytically, in the desired robustness properties by means of a projection of the error into the sliding surface. For linear time invariant dynamics, this approach results in sliding dynamics which are equivalent (Example 4.3) to those obtained with a conventional reduced-order Luenberger observer. Methods proposed in this Chapter can be viewed as

reduced-order nonlinear observers with design methodologies based on sliding mode concepts.

In this Chapter, sliding mode observers are proposed for systems with nonlinear dynamics which are known up to bounded propagation error over a single sample period, and for which measurements are linear with respect to the state of the system. Both systems with nonredundant and redundant measurement models are considered. In the case of redundant measurements, the proposed approach minimizes a quadratic cost function during the sliding motion. The proposed observers possess the following properties:

1. they are continuous across the sliding surface thereby satisfying the necessary condition of Theorem 4.2 for convergent quasi-sliding motion;
2. a global convergent quasi-sliding regime is generated in error space, independent of the modelling error;
3. the reaching dynamics can be specified through the use of the reaching law method presented in Section 5.7;
4. they have a form which allows stability analysis, in the presence of measurement errors, to proceed under the hypothesis of Theorem 5.11;
5. error dynamics during sliding result from a projection operation which means that, in the ideal case of no measurement errors, invariance to unmodelled dynamics can be achieved provided certain *matching conditions* are satisfied;
6. redundant measurements are optimally fused during sliding to minimize a quadratic cost measure.

Section 6.2 introduces notation and preliminary results required for subsequent developments. Section 6.3 defines the concept of linear measurement redundancy and discusses the implications of redundant measurements in the design of sliding mode observers. The proposed observer for nonredundant measurements is presented in Section 6.4, while that for redundant measurements is treated in Section 6.5. Simulation results for 3D feature point tracking are presented and discussed in Section 6.6.

6.2 Notation and Preliminary Results

The system under consideration is assumed to be of order n . The system provides m , $m < n$, measurements in each sample period. These m measurements depend, at each measurement event, only on q of the n states with $q \leq n$. For example a system which is modelled with states position, velocity, and acceleration ($n = 3$) may only provide one measurement ($m = 1$) which depends only on position ($q = 1$). The following notation is introduced to single out those states upon which measurements are dependent at each sample time.

The full, or *extended*, state vector is $\mathbf{x}_e(k) \in \mathbf{X}_e \subseteq \mathbb{R}^n$. The system dynamics are written as

$$\mathbf{x}_e(k+1) = \mathbf{f}(k, \mathbf{x}_e(k)), \quad (6.1)$$

where $\mathbf{f} : \mathbb{N}_0 \times \mathbf{X}_e \mapsto \mathbf{X}_e$ and $\mathbb{N}_0 := \{0, 1, 2, \dots\}$.

The error-free measurement vector, $\mathbf{z}(k) \in \mathbb{R}^m$, for any time k , is related to the extended state vector according to the usual form of measurement model

$$\mathbf{z}(k) = \mathbf{h}_e(\mathbf{x}_e(k)), \quad (6.2)$$

but this representation suggests that measurements, for each k , may depend on all elements of the extended state vector. Instead, developments of this and the following Chapter employ, without loss of generality, a measurement model in the form

$$\mathbf{z}(k) = \mathbf{h}(\mathbf{x}(k)), \quad \mathbf{x}(k) := \mathbf{C}\mathbf{x}_e(k), \quad (6.3)$$

where $\mathbf{h} : \mathbf{X} \mapsto \mathbb{R}^m$, $\mathbf{X} = \mathbf{C}\mathbf{X}_e \subseteq \mathbb{R}^q$, and $\mathbf{C} \in \mathbb{R}^{q \times n}$ can be written, with perhaps some requirement for reordering elements of the extended state vector, as

$$\mathbf{C} = [\mathbf{I}_q, \mathbf{0}_{q \times (n-q)}], \quad (6.4)$$

in which $\mathbf{I}_i$ denotes the identity matrix of dimension i , and $\mathbf{0}_{i \times j}$ denotes an i by j block of zeros. Without reordering the elements of the extended state vector, the matrix $\mathbf{C}$ can be taken as some column-wise permutation of the $q \times n$ matrix on the right of (6.4). The vector $\mathbf{x}$ is referred to as the *measured state*, while $\mathbf{x}_e$ is referred to as the *extended state*.

PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

$$\begin{aligned}
&= \mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \\
&\quad - \mathbf{L}(k) [\mathbf{H}_e \mathbf{L}(k)]^{-1} \{ \mathbf{z}_f(k+1) - \mathbf{H}_e \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \} \\
&= \mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) - \mathbf{L}(k) [\mathbf{H}_e \mathbf{L}(k)]^{-1} \\
&\quad \cdot \{ \mathbf{z}(k+1) - \mathbf{z}(k+1) + \mathbf{z}_f(k+1) - \mathbf{H}_e \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \} \\
&= \mathbf{F}(k) \{ \mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \} \\
&\quad - \mathbf{L}(k) [\mathbf{H}_e \mathbf{L}(k)]^{-1} \{ \mathbf{z}_f(k+1) - \mathbf{z}(k+1) + \boldsymbol{\zeta}(k+1) \} \\
&= \mathbf{F}(k) \{ \mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \} + [\mathbf{I}_n - \mathbf{F}(k)] \mathbf{e}_e(k+1). \quad (6.70)
\end{aligned}$$

Rearranging (6.70) gives the desired result

$$\mathbf{F}(k) \mathbf{e}_e(k+1) = \mathbf{F}(k) \{ \mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \}, \quad (6.71)$$

where $\mathbf{F}(k)$, as defined in (6.52), represents a projection along the column space of $\mathbf{L}(k)$ into the null space of $\mathbf{C}$. $\square$

This approach does not retain completely the robustness property of the unfiltered observer. From (6.71), if the matching condition $\exists \boldsymbol{\Xi}(k) \in \mathbb{R}^{m \times q}$ such that $\mathbf{B}(k) = \mathbf{L}(k) \boldsymbol{\Xi}(k)$ is satisfied, then $\mathbf{F}(k) \mathbf{e}_e(k+1)$ is invariant to the modelling error. Complete error dynamics, given by the combined systems (6.69) and (6.70), are generally influenced by modelling errors which enter through (6.69). Stability of the filtered observer, at this time, remains open. An alternate method to reduce sensitivity to measurement errors is, if possible, to employ redundant measurements as developed in the following Section.

6.5 Redundant Linear Measurement Models

In this Section, the true dynamics are again given by

$$\mathbf{x}_e(k+1) = \mathbf{f}(k, \mathbf{x}_e(k)), \quad (6.72)$$

where $\mathbf{x}_e$ is the extended state vector and $\mathbf{f} : \mathbb{N}_0 \times \mathbf{X}_e \mapsto \mathbf{X}_e$. The redundant linear measurement model is written as

$$\begin{aligned}
\mathbf{z}(k) &= \mathbf{H}_e \mathbf{x}_e(k) + \boldsymbol{\zeta}(k) \\
&= \mathbf{H} \mathbf{C} \mathbf{x}_e(k) + \boldsymbol{\zeta}(k) \\
&= \mathbf{H} \mathbf{x}(k) + \boldsymbol{\zeta}(k), \quad (6.73)
\end{aligned}$$

to $q = 4$ elements of the extended state vector. The measurement model, with $m = 6$, can be written as in (6.6) with

$$\mathbf{z}(k) = \mathbf{h}(\mathbf{x}(k)) = \mathbf{H}\mathbf{x}(k), \quad \mathbf{x}(k) = \mathbf{C}\mathbf{x}_e(k), \quad (6.10)$$

where $\mathbf{H} \in \mathbb{R}^{6 \times 4}$ and $\mathbf{C} \in \mathbb{R}^{4 \times 6}$ are given by

$$\mathbf{H} := \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{C} := \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \quad (6.11)$$

and hence

$$\mathbf{x}(k) := \mathbf{C}\mathbf{x}_e(k) = [p_x, \dot{p}_x, p_y, \dot{p}_y]^T. \quad (6.12)$$

A reordering of the elements of the extended state vector would give $\mathbf{C}$ in the form of (6.4). $\square$

In Example 6.1, one observation which is very important in subsequent analysis is that $\mathbf{H}_e := \mathbf{H}\mathbf{C}$ is rank deficient since it has linearly dependent rows, but both $\mathbf{H}$ and $\mathbf{C}$ have full rank since $\mathbf{H}$ has linearly independent columns and $\mathbf{C}$ has linearly independent rows. This will generally be the case for the matrix $\mathbf{C}$ due to its special structure formed through some column-wise permutation of the $q \times n$ matrix on the right of (6.4).

Specification of sliding surfaces and design of the sliding mode observers proposed in this Chapter require alternate strategies depending on the rank of the matrix $\mathbf{H}_e$. Consider the case in which the surface is defined as

$$\mathbf{s}_k := \mathbf{z}(k) - \mathbf{H}_e \hat{\mathbf{x}}_e(k) = \mathbf{H}_e \mathbf{e}_e(k), \quad (6.13)$$

where $\hat{\mathbf{x}}_e(k)$ is the estimated state at time k , and

$$\mathbf{e}_e(k) := \mathbf{x}_e(k) - \hat{\mathbf{x}}_e(k), \quad (6.14)$$

is the observer error. The primary objective of the sliding mode approach is to achieve robustness to unmodelled dynamics by arranging, as previously discussed in Sections 3.5,

4.3, and 5.6, for unmodelled dynamics and disturbances to influence the sliding dynamics through premultiplication by a projection matrix (see (3.69), page 98) of the form

$$\mathbf{F}(k) := [\mathbf{I}_n - \mathbf{L}(k) (\mathbf{H}_e \mathbf{L}(k))^{-1} \mathbf{H}_e], \quad (6.15)$$

where $\mathbf{L}(k)$ are the observer gains. The matrix $\mathbf{F}(k)$, for each k , represents a projection along the column space of $\mathbf{L}(k)$ into the null space of $\mathbf{H}_e$. In order to achieve this objective, gain matrices $\mathbf{L}(k) \in \mathbb{R}^{n \times m}$ must exist such that the product $\mathbf{H}_e \mathbf{L}(k)$ is nonsingular.

Theorem 6.1: For $\mathbf{H}_e \in \mathbb{R}^{m \times n}$, there exists a matrix $\mathbf{L} \in \mathbb{R}^{n \times m}$ such that the product $\mathbf{H}_e \mathbf{L}$ is nonsingular if, and only if, $\text{rank}\{\mathbf{H}_e\} = m$.

Proof: The existence of such a matrix $\mathbf{L}$ is equivalent to the existence of a right inverse of $\mathbf{H}_e$. That is, if $\exists \mathbf{L}$ such that $\mathbf{H}_e \mathbf{L} = \mathbf{M}$ where $\mathbf{M} \in \mathbb{R}^{m \times m}$ is nonsingular, then $\mathbf{B} := \mathbf{L} \mathbf{M}^{-1}$ is a right inverse of $\mathbf{H}_e$, and, conversely, if $\mathbf{B}$ is a right inverse of $\mathbf{H}_e$, then $\mathbf{H}_e \mathbf{L}$ is nonsingular with $\mathbf{L} = \mathbf{B}$. A necessary and sufficient condition for the existence of a right inverse of $\mathbf{H}_e$ is that $\mathbf{H}_e$ have linearly independent rows, i.e. $\text{rank}\{\mathbf{H}_e\} = m$ (see, for example, Perlis, Theorem 3-7 [161, p. 58]). $\square$

6.3 Linear Measurement Redundancy

The purpose of this Section is to quantify what is meant by a redundant linear measurement model. Measurement redundancy refers to the case in which more measurements are available than the minimum required to uniquely reconstruct the extended state of the system under ideal conditions (no disturbances, unmodelled dynamics, or measurement noise). Example 6.1 considered a system which has *redundant measurements*; the system provides two measurements of position in each of the x - and y -components. If $\text{rank}\{\mathbf{H}_e\} = r < m$, then the ideal (no noise) measurements for all time lie in the r -dimensional column space of $\mathbf{H}_e$. If the system is uniformly observable with N measurement events (see Section 3.3, page 82), then the observability map defined in (3.11) is injective from $\mathbf{X}_e$ into $\mathbb{R}^{Nr} \subset \mathbb{R}^{Nm}$. Intuitively, this means that the measurement model can be reduced to an r -dimensional “measurement” vector for each k , and still retain uniform observability. It must be emphasized that the objective is not to propose methods to remove

redundant measurements, but rather to define clearly what is meant by the term *redundant measurements* and to provide a means to determine whether or not the measurement model is redundant so that the proper form of sliding mode observer can be selected.

The following definition is proposed to characterize redundant linear measurement models.

Definition 6.1: For system (6.1) with linear measurement model (6.2), which is assumed to be uniformly observable in X_e with N measurement events, the measurements are said to be *redundant* if $\exists D \in \mathbb{R}^{r \times m}$ with $r < m$ such that system (6.1) with the reduced measurement model

$$z'(k) = Dz(k) = DH_e x_e(k) \in \mathbb{R}^r \quad (6.16)$$

is uniformly observable in X_e with N measurement events. $\square$

Theorem 6.2: Consider a system and measurements of the form

$$x_e(k+1) = f(x_e(k)), \quad z(k) = H_e x_e(k), \quad (6.17)$$

where $x_e \in X_e$, and $H_e \in \mathbb{R}^{m \times n}$. Assume the system is uniformly observable (see page 82) in X_e with N measurement events. If $\text{rank}\{H_e\} = r < m$, then the system has redundant measurements.

Proof: If the system with original measurement model (6.17) is uniformly observable in X_e with N measurement events, then $\forall x_e^* \in X_e$,

$$\text{rank} \left\{ \left[\frac{\partial \mathcal{O}_N(x_e)}{\partial x_e} \right]_{x_e=x_e^*} \right\} = n, \quad (6.18)$$

where $\mathcal{O}_N : X_e \mapsto \mathbb{R}^{Nm}$ is the observability map, from (3.11),

$$\mathcal{O}_N(x_e) = \begin{bmatrix} H_e x_e \\ H_e f(x_e) \\ H_e f \circ f(x_e) \\ \vdots \\ H_e f^{N-1}(x_e) \end{bmatrix}. \quad (6.19)$$

Let the set $\{\theta_i \in \mathbb{R}^m \mid i = 1, 2, \dots, r\}$ be a basis for the column space of $\mathbf{H}_r$. Define

$$\Theta := [\theta_1, \theta_2, \dots, \theta_r] \in \mathbb{R}^{m \times r}. \quad (6.20)$$

$$\mathbf{D} := [\Theta^T \Theta]^{-1} \Theta^T, \quad (6.21)$$

where the inverse $[\Theta^T \Theta]^{-1}$ exists since Θ , by construction, has linearly independent columns. Define also a matrix Γ such that

$$\mathbf{H}_r = \Theta \Gamma. \quad (6.22)$$

in which case

$$\mathbf{D} \mathbf{H}_r = \Gamma. \quad (6.23)$$

Consider the dynamics in system (6.17) with the reduced-order measurement model

$$\mathbf{z}'(k) := \mathbf{D} \mathbf{H}_r \mathbf{x}_r(k) = \Gamma \mathbf{x}_r(k). \quad (6.24)$$

The corresponding observability map is

$$\mathcal{O}'_N(\mathbf{x}_r) = \begin{bmatrix} \Gamma \mathbf{x}_r \\ \Gamma \mathbf{f}(\mathbf{x}_r) \\ \Gamma \mathbf{f} \circ \mathbf{f}_r(\mathbf{x}_r) \\ \vdots \\ \Gamma \mathbf{f}^{N-1}(\mathbf{x}_r) \end{bmatrix}. \quad (6.25)$$

It is required to show that the Jacobian matrix of the observability map in (6.25) has rank n for all $\mathbf{x}_r \in \mathbf{X}_r$. This Jacobian matrix is of dimension $Nr \times n$ and hence, $\forall \mathbf{x}_r^* \in \mathbf{X}_r$,

$$\text{rank} \left\{ \left[\frac{\partial \mathcal{O}'_N(\mathbf{x}_r)}{\partial \mathbf{x}_r} \right]_{\mathbf{x}_r = \mathbf{x}_r^*} \right\} \leq n. \quad (6.26)$$

By construction,

$$\mathcal{O}_N(\mathbf{x}_r) = \text{diag}[\underbrace{\Theta, \Theta, \dots, \Theta}_{N \text{ copies}}] \mathcal{O}'_N(\mathbf{x}_r). \quad (6.27)$$

Using (6.18), inequality (6.26), and the fact that, for matrices $\mathbf{A}$ and $\mathbf{B}$ of suitable dimension,

$$\text{rank}\{\mathbf{AB}\} \leq \min(\text{rank}\{\mathbf{A}\}, \text{rank}\{\mathbf{B}\}). \quad (6.28)$$

$$\forall \mathbf{x}_e^* \in \mathbf{X}_e,$$

$$\begin{aligned} n &= \text{rank} \left\{ \left[\frac{\partial \mathcal{O}_N(\mathbf{x}_e)}{\partial \mathbf{x}_e} \right]_{\mathbf{x}_e = \mathbf{x}_e^*} \right\} \\ &\leq \text{rank} \left\{ \left[\frac{\partial \mathcal{O}'_N(\mathbf{x}_e)}{\partial \mathbf{x}_e} \right]_{\mathbf{x}_e = \mathbf{x}_e^*} \right\} \leq n. \end{aligned} \quad (6.29)$$

Hence equality must hold on both sides of (6.29) which completes the proof. $\square$

In Theorem 6.2, one choice for the matrix Θ is available from the singular value decomposition [160, p. 16] of $\mathbf{H}_e$,

$$\mathbf{H}_e = \mathbf{Q}_1 \Sigma \mathbf{Q}_2^T, \quad (6.30)$$

with

$$\begin{aligned} \mathbf{Q}_1 &:= [\mathbf{q}_1^{(1)}, \mathbf{q}_2^{(1)}, \dots, \mathbf{q}_m^{(1)}] \in \mathbb{R}^{m \times m} \sim \text{orthogonal} \\ \mathbf{Q}_2 &:= [\mathbf{q}_1^{(2)}, \mathbf{q}_2^{(2)}, \dots, \mathbf{q}_n^{(2)}] \in \mathbb{R}^{n \times n} \sim \text{orthogonal} \end{aligned} \quad (6.31)$$

$$\Sigma := \text{diag} \{ \sigma_1, \dots, \sigma_r, 0, \dots, 0 \} \in \mathbb{R}^{m \times n}, \quad (6.32)$$

and $\sigma_1, \dots, \sigma_r$ are the nonzero singular values of $\mathbf{H}_e$. Define

$$\Theta := \tilde{\mathbf{Q}}_1 := [\mathbf{q}_1^{(1)}, \mathbf{q}_2^{(1)}, \dots, \mathbf{q}_r^{(1)}] \in \mathbb{R}^{m \times r} \quad (6.33)$$

$$\tilde{\mathbf{Q}}_2 := [\mathbf{q}_1^{(2)}, \mathbf{q}_2^{(2)}, \dots, \mathbf{q}_r^{(2)}] \in \mathbb{R}^{n \times r} \quad (6.34)$$

$$\tilde{\Sigma} := \text{diag} \{ \sigma_1, \dots, \sigma_r \} \in \mathbb{R}^{r \times r}. \quad (6.35)$$

Then

$$\mathbf{H}_e = \tilde{\mathbf{Q}}_1 \tilde{\Sigma} \tilde{\mathbf{Q}}_2^T = \Theta \Gamma \quad (6.36)$$

where $\Gamma := \tilde{\Sigma} \tilde{\mathbf{Q}}_2^T$, and the columns of $\Theta := \tilde{\mathbf{Q}}_1$ form an orthonormal basis for the column space of $\mathbf{H}_e$. In this case, $\Theta^T \Theta = \mathbf{I}_r$ and the matrix $\mathbf{D}$ of Definition 6.1 can be taken as $\mathbf{D} = \Theta^T$.

6.4 Non-redundant Linear Measurement Models

In this Section, the true dynamics are given by

$$\mathbf{x}_e(k+1) = \mathbf{f}(k, \mathbf{x}_e(k)), \quad (6.37)$$

where $\mathbf{x}_e$ is the extended state vector and $\mathbf{f} : \mathbb{N}_0 \times \mathbf{X}_e \mapsto \mathbf{X}_e$. The non-redundant linear measurement model is written as

$$\begin{aligned} \mathbf{z}(k) &= \mathbf{H}_e \mathbf{x}_e(k) + \boldsymbol{\zeta}(k) \\ &= \mathbf{H} \mathbf{C} \mathbf{x}_e(k) + \boldsymbol{\zeta}(k) \\ &= \mathbf{H} \mathbf{x}(k) + \boldsymbol{\zeta}(k). \end{aligned} \quad (6.38)$$

where $\mathbf{H}_e \in \mathbb{R}^{m \times n}$ has $\text{rank}\{\mathbf{H}_e\} = m$, $\mathbf{H} \in \mathbb{R}^{m \times q}$, $\mathbf{C} \in \mathbb{R}^{q \times n}$ has the form shown in (6.4), $\mathbf{H}_e := \mathbf{H}\mathbf{C}$, and $\mathbf{x}(k) := \mathbf{C}\mathbf{x}_e(k)$. Measurement errors $\boldsymbol{\zeta}(k)$ are assumed, just as in Chapter 5, to be uniformly bounded with

$$\sup_{\{k \geq 0\}} \|\boldsymbol{\zeta}(k)\| \leq \zeta < \infty. \quad (6.39)$$

In the case of non-redundant measurements $\mathbf{H}_e$ has linearly independent rows, i.e. $\text{rank}\{\mathbf{H}_e\} = m$, and Theorem 6.1 guarantees that gain matrices $\mathbf{L}(k)$ exist such that the projection operator

$$\begin{aligned} \mathbf{F}(k) &:= [\mathbf{I}_n - \mathbf{L}(k) (\mathbf{H}_e \mathbf{L}(k))^{-1} \mathbf{H}_e] \\ &= [\mathbf{I}_n - \mathbf{L}(k) (\mathbf{H} \mathbf{C} \mathbf{L}(k))^{-1} \mathbf{H} \mathbf{C}] \end{aligned} \quad (6.40)$$

is well defined, and represents a projection along the column space of $\mathbf{L}(k)$ into the null space of $\mathbf{H}_e$. In this nonredundant case, the measurement model $\mathbf{z}(k) = \mathbf{H}\mathbf{x}(k)$, $\mathbf{H} \in \mathbb{R}^{m \times q}$, has the same or fewer equations than unknowns since

$$\begin{aligned} m = \text{rank}\{\mathbf{H}_e\} &= \text{rank}\{\mathbf{H}\mathbf{C}\} \leq \min(\text{rank}\{\mathbf{H}\}, \text{rank}\{\mathbf{C}\}) \\ &\leq \min(\min(m, q), m) = \min(q, m) \end{aligned} \quad (6.41)$$

results in a contradiction unless $q \geq m$. Moreover,

$$m = \text{rank}\{\mathbf{H}_e\} \leq \text{rank}\{\mathbf{H}\} \leq m \quad (6.42)$$

implies that $\text{rank}\{\mathbf{H}\} = m$. If $m = q$, then $\mathbf{H}$ must be nonsingular in which case (6.40) reduces to

$$\mathbf{F}(k) = [\mathbf{I}_n - \mathbf{L}(k) (\mathbf{C} \mathbf{L}(k))^{-1} \mathbf{C}] \quad (6.43)$$

which represents a projection along the column space of $\mathbf{L}(k)$ into the null space of $\mathbf{C}$. During sliding in the ideal case of no measurement noise,

$$\mathbf{s}_k = \mathbf{H}_e \mathbf{e}_e(k) = \mathbf{H} \mathbf{C} \mathbf{e}_e(k) = 0 \quad (6.44)$$

implies that

$$\mathbf{C} \mathbf{e}_e(k) = \mathbf{e}(k) := \mathbf{x}(k) - \hat{\mathbf{x}}(k) = 0, \quad (6.45)$$

which means that the observer tracks the measured state exactly. The condition $q = m$, that is the measurement model has the same number of equations as unknowns, is not generally a necessary condition for observability.

Example 6.2: Consider

$$\begin{aligned} \mathbf{x}_e(k+1) &= \Phi \mathbf{x}_e(k) = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \mathbf{x}_e(k) \\ z(k) &= \mathbf{H}_e \mathbf{x}_e(k) = \begin{bmatrix} 1 & 1 \end{bmatrix} \mathbf{x}_e(k). \end{aligned} \quad (6.46)$$

This is a constant velocity system with T being the sample period. Here, $n = 2$, $m = 1$, and $q = 2 = n$. The observability matrix for this linear time-invariant system is

$$\mathcal{O} = \begin{bmatrix} \mathbf{H}_e \\ \mathbf{H}_e \Phi \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & T+1 \end{bmatrix}, \quad (6.47)$$

which has $\text{rank}\{\mathcal{O}\} = 2$ for $T > 0$, and hence the system is observable, although the observability matrix will be poorly conditioned for inversion if the sample period T is very small. $\square$

If $q > m$, then during sliding $\mathbf{e}(k)$ from (6.45) lies in the $(q - m)$ -dimensional null space of $\mathbf{H}$ which places further emphasis on stability analysis of the zero dynamics, i.e. the requirement that any dynamics present in the expression $\mathbf{H} \mathbf{e}(k) = 0$ are stable.

The following Theorem gives the proposed observer form and details its fundamental properties.

Theorem 6.3: For the system (6.37) with measurement model (6.38) consider the observer

$$\hat{\mathbf{x}}_e(k+1) = \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \mathbf{L}(k) \mathbf{v}(k), \quad (6.48)$$

where $\hat{\mathbf{f}} : \mathbb{N}_0 \times \mathbb{R}^n \mapsto \mathbb{R}^n$ is C^p in $\hat{\mathbf{x}}_e$, $p \geq 1$, and is an estimate of the true dynamics $\mathbf{f}$, $\mathbf{L}(k) \in \mathbb{R}^{n \times m}$ are the observer gains and $\mathbf{H}_e \mathbf{L}(k)$ is nonsingular for each k . Assume the following definitions:

$$\mathbf{v}(k) := (\mathbf{H}_e \mathbf{L}(k))^{-1} \{ \mathbf{z}(k+1) - \mathbf{H}_e \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) - \mathbf{g}(k, \mathbf{s}_k) \}, \quad (6.49)$$

$$\mathbf{s}_k := \mathbf{z}(k) - \mathbf{H}_e \hat{\mathbf{x}}_e(k), \quad (6.50)$$

$$\mathbf{g}(k, \mathbf{s}_k) := \mathbf{\Lambda}_k \text{Sat}(\mathbf{s}_k, \epsilon) + \mathbf{\Gamma}_k \mathbf{s}_k, \quad (6.51)$$

$$\mathbf{F}(k) := [\mathbf{I}_n - \mathbf{L}(k) (\mathbf{H}_e \mathbf{L}(k))^{-1} \mathbf{H}_e]. \quad (6.52)$$

$$\mathbf{B}(k) \mathbf{d}(k, \mathbf{y}_e) := \mathbf{f}(k, \mathbf{y}_e) - \hat{\mathbf{f}}(k, \mathbf{y}_e), \quad \forall \mathbf{y}_e \in \mathbf{X}_e. \quad (6.53)$$

The following statements hold:

1. *Reaching Law*: If $\mathbf{\Lambda}_k$, $\mathbf{\Gamma}_k$ and ϵ satisfy the hypothesis of Theorem 5.12, then a global convergent quasi-sliding regime exists on $\mathbf{s}_k = \mathbf{0}$ independent of the modelling error, and pure sliding on $\mathbf{s}_k = \mathbf{0}$ can be maintained for all $k > k_s \geq 0$ by setting $\mathbf{\Lambda}_k = \mathbf{0}$ and $\mathbf{\Gamma}_k = \mathbf{0}$ for all $k \geq k_s$ (see Corollary 5.2);
2. *Stability of Reaching and Sliding Dynamics*: With Statement 1 above, if $\mathbf{B}(k) \in \mathbb{R}^{n \times p}$, $\mathbf{d}(k, \mathbf{x}_e) \in \mathbb{R}^{p \times 1}$, $\mathbf{F}(k)$, $\hat{\mathbf{f}}$ and satisfy the hypothesis of Theorem 5.11, where the parameters $\boldsymbol{\xi}(k)$ and $\boldsymbol{\eta}(k)$ of Theorem 5.11 are taken as

$$\boldsymbol{\xi}(k) := -\mathbf{g}(k, \mathbf{s}_k), \quad \text{and} \quad (6.54)$$

$$\boldsymbol{\eta}(k) = \boldsymbol{\zeta}(k), \quad (6.55)$$

(here, parameter ϵ_2 of Theorem 5.11 can be set to zero), then the error dynamics are locally exponentially stable;

3. *Robustness of Sliding Mode*: With Statements 1 and 2 above, if, in addition, $\exists \boldsymbol{\Xi}(k) \in \mathbb{R}^{m \times q}$ such that $\mathbf{B}(k) = \mathbf{L}(k) \boldsymbol{\Xi}(k)$, then the sliding dynamics are invariant to dynamical modelling errors/disturbances. In the ideal case of no measurement errors ($\boldsymbol{\eta}(k) \equiv \mathbf{0}$), the sliding dynamics are locally exponentially asymptotically stable.

Proof: For Statement 1, it is required to show that under the above definitions, $\mathbf{s}_{k+1} = \mathbf{g}(k, \mathbf{s}_k)$, from which the result follows by Theorem 5.12. But from (6.50),

$$\mathbf{s}_{k+1} := \mathbf{z}(k+1) - \mathbf{H}_e \hat{\mathbf{x}}_e(k+1)$$

$$\begin{aligned}
&= \mathbf{z}(k+1) - \mathbf{H}_e \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) - \mathbf{H}_e \mathbf{L}(k) \mathbf{v}(k) \\
&= \left\{ \mathbf{z}(k+1) - \mathbf{H}_e \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \right\} - \left\{ \mathbf{z}(k+1) - \mathbf{H}_e \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \right\} + \mathbf{g}(k, \mathbf{s}_k) \\
&= \mathbf{g}(k, \mathbf{s}_k).
\end{aligned} \tag{6.56}$$

For Statement 2, it is required to show that the observer can be placed in the Luenberger form of (5.115, page 157). But the observer can be written as

$$\hat{\mathbf{x}}_e(k+1) = \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \mathbf{K}(k) [\mathbf{z}(k+1) - \mathbf{H}_e \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \boldsymbol{\xi}(k)], \tag{6.57}$$

with

$$\mathbf{K}(k) := \mathbf{L}(k) [\mathbf{H}_e \mathbf{L}(k)]^{-1}, \text{ and} \tag{6.58}$$

$$\boldsymbol{\xi}(k) := -\mathbf{g}(k, \mathbf{s}_k). \tag{6.59}$$

where $\boldsymbol{\xi}(k)$ is uniformly bounded by Theorem 5.12, and vanishing from Statement 1 and application of Corollary 5.2 for all $k \geq k_s \geq 0$.

For Statement 3, the error dynamics of the observer in (6.48) can be written, using $\mathbf{K}(k)$ defined in (6.58), as

$$\begin{aligned}
\mathbf{e}_e(k+1) &:= \mathbf{x}_e(k+1) - \hat{\mathbf{x}}_e(k+1) \\
&= \mathbf{F}(k) \left\{ \mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \right\} \\
&\quad + \mathbf{K}(k) \mathbf{g}(k, \mathbf{s}_k) - \mathbf{K}(k) \boldsymbol{\zeta}(k+1).
\end{aligned} \tag{6.60}$$

During sliding, $\mathbf{g}(k, \mathbf{s}_k) = \mathbf{0}$ and the error dynamics become, from (6.60) and using (6.53),

$$\begin{aligned}
\mathbf{e}_e(k+1) &= \mathbf{F}(k) \left\{ \mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \right\} - \mathbf{K}(k) \boldsymbol{\zeta}(k+1) \\
&= \mathbf{F}(k) \left\{ \mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \mathbf{x}_e(k)) \right\} + \mathbf{F}(k) \left\{ \hat{\mathbf{f}}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \right\} \\
&\quad - \mathbf{K}(k) \boldsymbol{\zeta}(k+1) \\
&= \mathbf{F}(k) \mathbf{B}(k) \mathbf{d}(k, \mathbf{x}_e(k)) + \mathbf{F}(k) \left\{ \hat{\mathbf{f}}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \right\} \\
&\quad - \mathbf{K}(k) \boldsymbol{\zeta}(k+1) \\
&= \mathbf{F}(k) \mathbf{L}(k) \boldsymbol{\Xi}(k) \mathbf{d}(k, \mathbf{x}_e(k)) + \mathbf{F}(k) \left\{ \hat{\mathbf{f}}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \right\} \\
&\quad - \mathbf{K}(k) \boldsymbol{\zeta}(k+1) \\
&= \mathbf{F}(k) \left\{ \hat{\mathbf{f}}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \right\} - \mathbf{K}(k) \boldsymbol{\zeta}(k+1),
\end{aligned} \tag{6.61}$$

where the second equality is obtained by adding and subtracting $F(k)\hat{f}(k, x_e(k))$, as was done in the proof of Theorem 5.11, and the final equality results since $F(k)L(k) \equiv 0$. This means that the sliding dynamics are invariant to dynamical modelling errors/disturbances, and application of Theorem 5.11 gives exponential stability. In the ideal case of no measurement errors ($\zeta(k) \equiv 0$), application of Theorem 5.11 results in exponential asymptotic stability. $\square$

Item 1 of Theorem 6.3 shows that the proposed observer satisfies the reaching law

$$s_{k+1} = g(k, s_k). \quad (6.62)$$

where, neglecting measurement error for the purpose of this discussion,

$$s_k := s(e(k)) = H_e e(k) = H e(k). \quad (6.63)$$

Stability of the reaching law, therefore, governs temporal behavior of the error, $e(k)$, in the estimate of the measured state vector only; the reaching law says nothing about the behavior, during the reaching phase, of errors in estimation of elements of the extended state vector which do not appear in the measured state vector. That is, the reaching law does not completely define error trajectories in the estimate of states $(I_n - C)x_e(k)$. This difficulty appears in the observer problem since the available information, the system measurements, are generally related to some but not all of the states of the system at each measurement event. This is in sharp contrast to the controller problem in which all states are assumed to be available; in this case a reaching law method may completely define the reaching dynamics of the entire state of the system.

It is often the case in the literature on sliding mode control, that the stability problem is treated in two distinct steps: 1) stability of the reaching dynamics which establish attractiveness of the sliding surface; and 2) stability of the sliding mode in which the system state is restricted to the sliding surface. A similar approach could have been used here in analysis of the proposed observer. With this approach, however, a fundamental difficulty occurs because of the local stability properties of the sliding mode. For illustration, consider an initial estimate $\hat{x}_e(0)$ which lies off of the sliding surface. Because the observer form is such that the reaching law is satisfied, the designer can specify a time $k_s \geq 0$ such

that pure sliding can be maintained for all $k > k_s$. Analysis of stability of the sliding mode could then be based on the value of the state estimate $\hat{\mathbf{x}}_e(k_s + 1)$ at time $k_s + 1$. The analysis would state that, "provided $\hat{\mathbf{x}}_e(k_s + 1)$ lies in a local stability domain of the sliding dynamics, then the sliding dynamics are stable." In this case the stability result is incomplete in the sense that the analysis has not specified admissible initial observer errors at time $k = 0$ which guarantee that the observer error, after passing through the reaching phase, will lie within a stability domain of the sliding mode at time $k = k_s + 1$. Item 2 of Theorem 6.3 does **not** produce this incomplete result, but instead treats stability of the total observer error, $\mathbf{e}_e(k)$, during both the reaching and sliding phases by appealing to the hypothesis of Theorem 5.11. In this way a stability domain for initialization of the observer at time $k = 0$ can, at least in principle, be computed such that if the observer is initialized in this stability domain, then the stability of the observer error, through both the reaching and sliding phases, is guaranteed.

Theorem 6.3 includes, through a deterministic treatment, the effect of bounded measurement errors on observer performance. Note that

$$\begin{aligned} \mathbf{s}_k &:= \mathbf{z}(k) - \mathbf{H}_e \hat{\mathbf{x}}_e(k) \\ &= \mathbf{H}_e \mathbf{e}_e(k) + \boldsymbol{\zeta}(k), \end{aligned} \quad (6.64)$$

and that Statement 1 of Theorem 6.3 implies that during the sliding motion, $\mathbf{s}_k = 0$ which in turn means that the state estimate tracks the measurement error. One method to reduce the effects of measurement errors, which has been suggested by Slotine *et al.* [95] for the continuous-time observer problem, corresponds to replacing $\mathbf{z}(k)$ in the above filter with $\mathbf{z}_f(k)$ obtained by passing the measurements through a low-pass filter. For example, consider the use of a low-pass FIR filter with

$$\mathbf{z}_f(k) := \sum_{i=0}^p \mathbf{A}_i \mathbf{z}(k-i), \quad (6.65)$$

where the coefficients matrices $\mathbf{A}_i \in \mathbb{R}^{m \times m}$ have been specified. The *filtered observer* is written as

$$\hat{\mathbf{x}}_e(k+1) = \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \mathbf{L}(k) \hat{\mathbf{v}}(k). \quad (6.66)$$

where

$$\hat{\mathbf{v}}(k) := (\mathbf{H}_e \mathbf{L}(k))^{-1} \left\{ \mathbf{z}_f(k+1) - \mathbf{H}_e \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) - \mathbf{g}(k, \mathbf{s}_k) \right\}.$$

$$\begin{aligned} \mathbf{s}_k &:= \mathbf{z}_f(k) - \mathbf{H}_e \hat{\mathbf{x}}_e(k), \\ \mathbf{g}(k, \mathbf{s}_k) &:= \mathbf{\Lambda}_k \mathbf{Sat}(\mathbf{s}_k, \epsilon) + \mathbf{\Gamma}_k \mathbf{s}_k, \end{aligned} \quad (6.67)$$

and has the following properties.

Theorem 6.4: For system (6.37) with measurement model (6.38) consider the observer (6.66) with definitions in (6.67). The following statements hold:

1. *Reaching Law:* If $\mathbf{\Lambda}_k$, $\mathbf{\Gamma}_k$ and ϵ satisfy the hypothesis of Theorem 5.12, then a global convergent quasi-sliding regime exists on $\mathbf{s}_k = \mathbf{0}$ independent of the modelling error, and pure sliding on $\mathbf{s}_k = \mathbf{0}$ can be maintained for all $k > k_s \geq 0$ by setting $\mathbf{\Lambda}_k = \mathbf{0}$ and $\mathbf{\Gamma}_k = \mathbf{0}$ for all $k \geq k_s$ according to Corollary 5.2;
2. *Relationship to Unfiltered Observer:* With Statement 1, the projection of the error dynamics during pure sliding along the column space of $\mathbf{L}(k)$ into the null space of $\mathbf{C}$ is independent of the coefficient matrices $\mathbf{A}_i$ and is identical to that for the unfiltered observer considered in Theorem 6.3.

Proof: For Statement 1, it is required to show that $\mathbf{s}_{k+1} = \mathbf{g}(k, \mathbf{s}_k)$, from which the result follows by Theorem 5.12. But

$$\begin{aligned} \mathbf{s}_{k+1} &:= \mathbf{z}_f(k+1) - \mathbf{H}_e \hat{\mathbf{x}}_e(k+1) \\ &= \left\{ \mathbf{z}_f(k+1) - \mathbf{H}_e \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \right\} - \left\{ \mathbf{z}_f(k+1) - \mathbf{H}_e \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \right\} + \mathbf{g}(k, \mathbf{s}_k) \\ &= \mathbf{g}(k, \mathbf{s}_k). \end{aligned} \quad (6.68)$$

For Statement 2, first recall that pure sliding initiated at time k_s , gives, for all $k > k_s$, $\mathbf{s}_k = \mathbf{0} \Rightarrow \mathbf{z}_f(k) = \mathbf{H}_e \hat{\mathbf{x}}_e(k)$ and hence, $\forall k > k_s$,

$$\begin{aligned} \mathbf{H}_e \mathbf{e}_e(k+1) &:= \mathbf{H}_e \mathbf{x}_e(k+1) - \mathbf{H}_e \hat{\mathbf{x}}_e(k+1) \\ &= \mathbf{H}_e \mathbf{x}_e(k+1) - \mathbf{z}_f(k+1) \\ &= -\left\{ \mathbf{z}_f(k+1) - \mathbf{z}(k+1) + \boldsymbol{\zeta}(k+1) \right\}. \end{aligned} \quad (6.69)$$

With (6.69), the error dynamics for $k > k_s$ can be written as

$$\mathbf{e}_e(k+1) := \mathbf{x}_e(k+1) - \hat{\mathbf{x}}_e(k+1)$$

$$\begin{aligned}
&= \mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \\
&\quad - \mathbf{L}(k) [\mathbf{H}_e \mathbf{L}(k)]^{-1} \{ \mathbf{z}_f(k+1) - \mathbf{H}_e \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \} \\
&= \mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) - \mathbf{L}(k) [\mathbf{H}_e \mathbf{L}(k)]^{-1} \\
&\quad \cdot \{ \mathbf{z}(k+1) - \mathbf{z}(k+1) + \mathbf{z}_f(k+1) - \mathbf{H}_e \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \} \\
&= \mathbf{F}(k) \{ \mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \} \\
&\quad - \mathbf{L}(k) [\mathbf{H}_e \mathbf{L}(k)]^{-1} \{ \mathbf{z}_f(k+1) - \mathbf{z}(k+1) + \boldsymbol{\zeta}(k+1) \} \\
&= \mathbf{F}(k) \{ \mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \} + [\mathbf{I}_n - \mathbf{F}(k)] \mathbf{e}_e(k+1). \quad (6.70)
\end{aligned}$$

Rearranging (6.70) gives the desired result

$$\mathbf{F}(k) \mathbf{e}_e(k+1) = \mathbf{F}(k) \{ \mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \}, \quad (6.71)$$

where $\mathbf{F}(k)$, as defined in (6.52), represents a projection along the column space of $\mathbf{L}(k)$ into the null space of $\mathbf{C}$. $\square$

This approach does not retain completely the robustness property of the unfiltered observer. From (6.71), if the matching condition $\exists \boldsymbol{\Xi}(k) \in \mathbb{R}^{m \times q}$ such that $\mathbf{B}(k) = \mathbf{L}(k) \boldsymbol{\Xi}(k)$ is satisfied, then $\mathbf{F}(k) \mathbf{e}_e(k+1)$ is invariant to the modelling error. Complete error dynamics, given by the combined systems (6.69) and (6.70), are generally influenced by modelling errors which enter through (6.69). Stability of the filtered observer, at this time, remains open. An alternate method to reduce sensitivity to measurement errors is, if possible, to employ redundant measurements as developed in the following Section.

6.5 Redundant Linear Measurement Models

In this Section, the true dynamics are again given by

$$\mathbf{x}_e(k+1) = \mathbf{f}(k, \mathbf{x}_e(k)), \quad (6.72)$$

where $\mathbf{x}_e$ is the extended state vector and $\mathbf{f} : \mathbb{N}_0 \times \mathbf{X}_e \mapsto \mathbf{X}_e$. The redundant linear measurement model is written as

$$\begin{aligned}
\mathbf{z}(k) &= \mathbf{H}_e \mathbf{x}_e(k) + \boldsymbol{\zeta}(k) \\
&= \mathbf{H} \mathbf{C} \mathbf{x}_e(k) + \boldsymbol{\zeta}(k) \\
&= \mathbf{H} \mathbf{x}(k) + \boldsymbol{\zeta}(k), \quad (6.73)
\end{aligned}$$

where $\mathbf{H}_e \in \mathbb{R}^{m \times n}$ has $\text{rank}\{\mathbf{H}_e\} = r < m$, $\mathbf{H} \in \mathbb{R}^{m \times q}$, $\mathbf{C} \in \mathbb{R}^{q \times n}$ has the form shown in (6.4), $\mathbf{H}_e := \mathbf{H}\mathbf{C}$, and $\mathbf{x}(k) := \mathbf{C}\mathbf{x}_e(k)$. Again, a simple treatment of the effects of measurement errors is given assuming that $\zeta(k)$ is uniformly bounded,

$$\sup_{\{k \geq 0\}} \|\zeta(k)\| \leq \zeta < \infty. \quad (6.74)$$

For redundant measurements, the matrix $\mathbf{H}_e = \mathbf{H}\mathbf{C}$ has rank less than m and hence a right inverse of $\mathbf{H}_e$ does not exist (Theorem 6.1). Of primary concern is that the projection operator $\mathbf{F}(k)$ defined in (6.15) cannot be attained. Consider instead, a least squared error approach based on minimization of a quadratic cost function defined on $\mathbb{R}^q$ as

$$C(\hat{\mathbf{x}}(k)) := [\mathbf{z}(k) - \mathbf{H}\hat{\mathbf{x}}(k)]^T \mathbf{P}_k [\mathbf{z}(k) - \mathbf{H}\hat{\mathbf{x}}(k)], \quad (6.75)$$

where $\mathbf{P}_k \in \mathbb{R}^{m \times m}$ is positive definite symmetric. The sliding surface is defined as

$$\mathbf{s}_k := -\frac{1}{2} \left(\frac{\partial C(\hat{\mathbf{x}}(k))}{\partial \hat{\mathbf{x}}(k)} \right)^T = \mathbf{H}^T \mathbf{P}_k [\mathbf{z}(k) - \mathbf{H}\hat{\mathbf{x}}(k)] = \mathbf{0}. \quad (6.76)$$

in which case (6.75) is minimized during the sliding motion.

The form of observer, given in Theorem 6.5 below, that is required to achieve sliding for the redundant measurement case demands the existence of gain matrices $\mathbf{L}(k)$ such that the product $\mathbf{H}^T \mathbf{P}_{k+1} \mathbf{H} \mathbf{C} \mathbf{L}(k)$ is nonsingular for each k . This requirement arises through formulation of a projection operator

$$\mathbf{F}(k) := \left[\mathbf{I}_n - \mathbf{L}(k) \left(\mathbf{H}^T \mathbf{P}_{k+1} \mathbf{H} \mathbf{C} \mathbf{L}(k) \right)^{-1} \mathbf{H}^T \mathbf{P}_{k+1} \mathbf{H} \mathbf{C} \right]. \quad (6.77)$$

which takes vectors in $\mathbb{R}^n$ along the column space of $\mathbf{L}(k)$ into the null space of $\mathbf{H}^T \mathbf{P}_{k+1} \mathbf{H} \mathbf{C}$. However, the matrix $\mathbf{C}$ always has a right inverse since it has linearly independent rows due to its special form defined as some column-wise permutation of the $q \times n$ matrix on the right of (6.4). Hence the existence of gain matrices $\mathbf{L}(k)$ can be guaranteed by restricting the matrix $\mathbf{H}$ to have linearly independent columns, i.e. $\text{rank}\{\mathbf{H}\} = q$, in which case the product $\mathbf{H}^T \mathbf{P}_{k+1} \mathbf{H}$ is nonsingular (see for example Strang [164, p. 109]). The projection operator defined in (6.77) then reduces to

$$\mathbf{F}(k) := \left[\mathbf{I}_n - \mathbf{L}(k) (\mathbf{C} \mathbf{L}(k))^{-1} \mathbf{C} \right], \quad (6.78)$$

which takes vectors in $\mathbb{R}^n$ along the column space of $L(k)$ into the null space of C .

The following Theorem gives the proposed observer form and details its fundamental properties.

Theorem 6.5: For the system (6.72) with measurement i.e. (6.73) consider the observer

$$\hat{\mathbf{x}}_e(k+1) = \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + L(k)v(k), \quad (6.79)$$

where $\hat{\mathbf{f}} : \mathbb{N}_0 \times \mathbb{R}^n \mapsto \mathbb{R}^n$ is C^p in $\hat{\mathbf{x}}_e$, $p \geq 1$, and is an estimate of the true dynamics $\mathbf{f}$, $L(k) \in \mathbb{R}^{n \times q}$ are the observer gains, $\mathbf{H}_e := \mathbf{H}\mathbf{C}$ where $\mathbf{H} \in \mathbb{R}^{m \times q}$ has $\text{rank}\{\mathbf{H}\} = q$, and $\mathbf{C}L(k)$ is nonsingular for each k . Assume the following definitions for some positive definite symmetric $\mathbf{P}_k \in \mathbb{R}^{m \times m}$:

$$v(k) := \left[\mathbf{H}^T \mathbf{P}_{k+1} \mathbf{H} \mathbf{C} L(k) \right]^{-1} \cdot \left\{ \mathbf{H}^T \mathbf{P}_{k+1} [\mathbf{z}(k+1) - \mathbf{H}_e \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k))] - \mathbf{g}(k, \mathbf{s}_k) \right\}, \quad (6.80)$$

$$\mathbf{s}_k := \mathbf{H}^T \mathbf{P}_k [\mathbf{z}(k) - \mathbf{H}_e \hat{\mathbf{x}}_e(k)], \quad (6.81)$$

$$\mathbf{g}(k, \mathbf{s}_k) := \Lambda_k \text{Sat}(\mathbf{s}_k, \epsilon) + \Gamma_k \mathbf{s}_k, \quad (6.82)$$

$$\mathbf{F}(k) := [\mathbf{I}_n - L(k)(\mathbf{C}L(k))^{-1}\mathbf{C}], \quad (6.83)$$

$$\mathbf{B}(k)\mathbf{d}(k, \mathbf{y}_e) := \mathbf{f}(k, \mathbf{y}_e) - \hat{\mathbf{f}}(k, \mathbf{y}_e), \quad \forall \mathbf{y}_e \in \mathbf{X}_e. \quad (6.84)$$

The following statements hold:

1. *Reaching Law:* If Λ_k, Γ_k and ϵ satisfy the hypothesis of Theorem 5.12, then a global convergent quasi-sliding regime exists on $\mathbf{s}_k = \mathbf{0}$ independent of the modelling error, and pure sliding on $\mathbf{s}_k = \mathbf{0}$ can be maintained for all $k > k_s > 0$ by setting $\Lambda_k = \mathbf{0}$ and $\Gamma_k = \mathbf{0}$ for all $k \geq k_s$ (see Corollary 5.2);
2. *Quadratic Optimization:* With Statement 1, during the sliding motion, $k > k_s$, the estimated measured state, $\hat{\mathbf{x}}(k)$, minimizes the quadratic cost function

$$C(\hat{\mathbf{x}}(k)) := [\mathbf{z}(k) - \mathbf{H}\hat{\mathbf{x}}(k)]^T \mathbf{P}_k [\mathbf{z}(k) - \mathbf{H}\hat{\mathbf{x}}(k)], \quad (6.85)$$

where $\mathbf{P}_k \in \mathbb{R}^{m \times m}$ is user-specified, and positive definite symmetric.

3. *Stability of Reaching and Sliding Dynamics:* With Statement 1 above, if $\mathbf{B}(k) \in \mathbb{R}^{n \times p}$, $\mathbf{d}(k, \mathbf{x}_e) \in \mathbb{R}^{p \times 1}$, $\mathbf{F}(k)$, $\hat{\mathbf{f}}$ satisfy the hypothesis of Theorem 5.11, where the parameters $\xi(k)$ and $\eta(k)$ of Theorem 5.11 are taken as

$$\xi(k) := -[\mathbf{H}^T \mathbf{P}_{k+1} \mathbf{H}]^{-1} \mathbf{H}^T \mathbf{P}_{k+1} \mathbf{g}(k, \mathbf{s}_k), \quad \text{and} \quad (6.86)$$

$$\eta(k) = [\mathbf{H}^T \mathbf{P}_{k+1} \mathbf{H}]^{-1} \mathbf{H}^T \mathbf{P}_{k+1} \zeta(k), \quad (6.87)$$

(here, parameter ϵ_2 of Theorem 5.11 can be set to zero), then the error dynamics are locally exponentially stable;

4. *Robustness of Sliding Mode:* With Statements 1 and 3 above, if, in addition, $\exists \Xi(k) \in \mathbb{R}^{m \times p}$ such that $\mathbf{B}(k) = \mathbf{L}(k)\Xi(k)$, then the sliding dynamics are invariant to the dynamical modelling errors/disturbances. In the ideal case of no measurement errors ($\zeta(k) \equiv 0$) the sliding dynamics are locally exponentially asymptotically stable.

Proof: For Statement 1, it is required to show that under the above definitions, $\mathbf{s}_{k+1} = \mathbf{g}(k, \mathbf{s}_k)$, from which the result follows by Theorem 5.12. But from (6.81),

$$\begin{aligned} \mathbf{s}_{k+1} &:= \mathbf{H}^T \mathbf{P}_{k+1} [\mathbf{z}(k+1) - \mathbf{H}_e \hat{\mathbf{x}}_e(k+1)] \\ &= \mathbf{H}^T \mathbf{P}_{k+1} \mathbf{z}(k+1) - \mathbf{H}^T \mathbf{P}_{k+1} \mathbf{H} \mathbf{C} \left\{ \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \mathbf{L}(k) \mathbf{v}(k) \right\} \\ &= \mathbf{H}^T \mathbf{P}_{k+1} \mathbf{z}(k+1) - \mathbf{H}^T \mathbf{P}_{k+1} \mathbf{H} \mathbf{C} \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \\ &\quad - \mathbf{H}^T \mathbf{P}_{k+1} \mathbf{z}(k+1) + \mathbf{H}^T \mathbf{P}_{k+1} \mathbf{H} \mathbf{C} \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \mathbf{g}(k, \mathbf{s}_k) \\ &= \mathbf{g}(k, \mathbf{s}_k). \end{aligned} \quad (6.88)$$

Statement 2 follows immediately from the definition of $\mathbf{s}_k$ in (6.81) and $C(\hat{\mathbf{x}})$ in (6.85), which satisfy

$$\mathbf{s}_k := -\frac{1}{2} \left(\frac{\partial C(\hat{\mathbf{x}}(k))}{\partial \hat{\mathbf{x}}(k)} \right)^T. \quad (6.89)$$

and, from Statement 1, the fact that $\mathbf{s}_k = 0$ for all $k > k_s$ and hence (6.85) is minimized during the sliding motion.

For Statement 3, it is required to show that the observer can be placed in the Luenberger form considered in Theorem 5.11 and stated in (5.115, page 157). But the observer can be written, using the nonsingularity of the product $\mathbf{H}^T \mathbf{P}_{k+1} \mathbf{H}$, as

$$\hat{\mathbf{x}}_e(k+1) = \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \mathbf{K}(k)[\mathbf{z}'(k+1) - \mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \xi(k)]. \quad (6.90)$$

where

$$\begin{aligned} \mathbf{z}'(k+1) &:= \mathbf{C}\mathbf{x}_e(k+1) + \boldsymbol{\eta}(k+1), \\ \mathbf{K}(k) &:= \mathbf{L}(k) [\mathbf{C}\mathbf{L}(k)]^{-1}, \end{aligned} \quad (6.91)$$

$$\boldsymbol{\xi}(k) := -[\mathbf{H}^T \mathbf{P}_{k+1} \mathbf{H}]^{-1} \mathbf{H}^T \mathbf{P}_{k+1} \mathbf{g}(k, \mathbf{s}_k), \quad \text{and} \quad (6.92)$$

$$\boldsymbol{\eta}(k) := [\mathbf{H}^T \mathbf{P}_{k+1} \mathbf{H}]^{-1} \mathbf{H}^T \mathbf{P}_{k+1} \boldsymbol{\zeta}(k), \quad (6.93)$$

in which $\boldsymbol{\xi}(k)$ is uniformly bounded by Theorem 5.12, and vanishing from Statement 1 with application of Corollary 5.2 for all $k \geq k_s \geq 0$.

For Statement 4, the sliding dynamics, with $\mathbf{g}(k, \mathbf{s}_k) = \mathbf{0}$, of the observer in (6.79) can be written, using $\mathbf{K}(k)$ defined in (6.91), as

$$\begin{aligned} \mathbf{e}_e(k+1) &:= \mathbf{x}_e(k+1) - \hat{\mathbf{x}}_e(k+1) \\ &= \mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) - \mathbf{L}(k)\mathbf{v}(k) \\ &= \mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \\ &\quad - \mathbf{L}(k) [\mathbf{H}^T \mathbf{P}_{k+1} \mathbf{H} \mathbf{C} \mathbf{L}(k)]^{-1} \mathbf{H}^T \mathbf{P}_{k+1} \mathbf{z}(k+1) \\ &\quad - \mathbf{L}(k) [\mathbf{H}^T \mathbf{P}_{k+1} \mathbf{H} \mathbf{C} \mathbf{L}(k)]^{-1} \mathbf{H}^T \mathbf{P}_{k+1} \mathbf{H} \mathbf{C} \hat{\mathbf{f}}(k, \mathbf{x}_e(k)) \\ &= \mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \\ &\quad - \mathbf{L}(k) [\mathbf{H}^T \mathbf{P}_{k+1} \mathbf{H} \mathbf{C} \mathbf{L}(k)]^{-1} \mathbf{H}^T \mathbf{P}_{k+1} \{\mathbf{H} \mathbf{C} \mathbf{x}_e(k+1) + \boldsymbol{\zeta}(k+1)\} \\ &\quad - \mathbf{L}(k) [\mathbf{H}^T \mathbf{P}_{k+1} \mathbf{H} \mathbf{C} \mathbf{L}(k)]^{-1} \mathbf{H}^T \mathbf{P}_{k+1} \mathbf{H} \mathbf{C} \hat{\mathbf{f}}(k, \mathbf{x}_e(k)) \\ &= \mathbf{F}(k) \{\mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k))\} - \mathbf{K}(k) \boldsymbol{\eta}(k+1), \end{aligned} \quad (6.94)$$

where $\boldsymbol{\eta}(k)$ is defined in (6.93). With (6.84), the result in (6.94) becomes

$$\begin{aligned} \mathbf{e}_e(k+1) &= \mathbf{F}(k) \{\mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k))\} - \mathbf{K}(k) \boldsymbol{\eta}(k+1) \\ &= \mathbf{F}(k) \{\mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \mathbf{x}_e(k))\} \\ &\quad + \mathbf{F}(k) \{\hat{\mathbf{f}}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k))\} - \mathbf{K}(k) \boldsymbol{\eta}(k+1) \\ &= \mathbf{F}(k) \mathbf{B}(k) \mathbf{d}(k, \mathbf{x}_e(k)) \\ &\quad + \mathbf{F}(k) \{\hat{\mathbf{f}}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k))\} - \mathbf{K}(k) \boldsymbol{\eta}(k+1) \\ &= \mathbf{F}(k) \mathbf{L}(k) \boldsymbol{\Xi}(k) \mathbf{d}(k, \mathbf{x}_e(k)) \\ &\quad + \mathbf{F}(k) \{\hat{\mathbf{f}}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k))\} - \mathbf{K}(k) \boldsymbol{\eta}(k+1) \\ &= \mathbf{F}(k) \{\hat{\mathbf{f}}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k))\} - \mathbf{K}(k) \boldsymbol{\eta}(k+1), \end{aligned} \quad (6.95)$$

where the second equality is obtained by adding and subtracting $\mathbf{F}(k)\hat{\mathbf{f}}(k, \mathbf{x}_e(k))$, and the final equality results since $\mathbf{F}(k)\mathbf{L}(k) \equiv \mathbf{0}$. This means that the sliding dynamics are invariant to dynamical modelling errors/disturbances, and application of Theorem 5.11 gives exponential stability. In the ideal case of no measurement errors ($\zeta(k) \equiv \mathbf{0}$) application of Theorem 5.11 results in exponential asymptotic stability. $\square$

The results of Theorem 6.5 again provide a simple treatment of the effects of measurement errors on observer stability. The weighted mean-square estimation in Statement 2, however, allows some statistical interpretation of the approach. In the case where measurement errors, $\zeta(k)$, are zero mean Gaussian with covariance matrix $\mathbf{R}_k := \mathcal{E}\{\zeta(k)\zeta^T(k)\}$, and with $\mathbf{P}_k := \mathbf{R}_k^{-1}$, the observer output during sliding satisfies

$$\mathbf{e}(k) := \mathbf{x}(k) - \hat{\mathbf{x}}(k) = [\mathbf{H}^T \mathbf{R}_k^{-1} \mathbf{H}]^{-1} \mathbf{H}^T \mathbf{R}_k^{-1} \zeta(k). \quad (6.96)$$

and hence $\mathbf{e}(k)$ is zero mean Gaussian with covariance

$$\mathcal{E}\{\mathbf{e}(k)\mathbf{e}^T(k)\} = [\mathbf{H}^T \mathbf{R}_k^{-1} \mathbf{H}]^{-1}. \quad (6.97)$$

The estimated measured state, $\hat{\mathbf{x}}(k)$, in this case is the maximum likelihood estimate [10, p. 103] in the sense that $\hat{\mathbf{x}}(k)$ is that value which maximizes the probability of the measurements $\mathbf{z}(k)$ that actually occurred, taking into account the assumed statistical properties of $\mathbf{z}(k)$. However, in this approach there is no statistical model assumed for $\mathbf{x}(k)$. Another pathological complication is that Gaussian statistics for $\zeta(k)$ do not strictly satisfy the assumption of bounded measurement errors required in the stability proof since arbitrarily large values of a Gaussian random variable can occur with nonzero probability.

6.6 Simulation Results for 3D Feature Point Tracking

Level V of the hierarchical image analysis system proposed in Figure 1.1 (page 3) is concerned with estimation of the position and motion of object features based on solutions to tracking and correspondence problems in Level IV. Results for the 3D feature point tracking problem were previously shown in applications of the simple sliding mode observer of Chapter 4. In this Section, example simulation results are shown from applications of

the sliding mode observers proposed in this Chapter, for both nonredundant and redundant measurements, in the problem of estimating the position and motion of feature point p^6 (see Figure 2.4, page 51) during the object trajectory described in Case 1 of Section 2.6.2.

Simulation results for this Chapter are presented in two Cases:

Case 1: Nonredundant Measurements. Measurements from binocular stereo cameras are backprojected into 3D to give *derived* linear measurements which depend on the feature point position. A linear Gauss-Markov model for translational motion is assumed and state estimation is provided by the nonredundant sliding mode observer. A practical method of estimating observer error bounds based on the theoretical results of this Chapter is proposed and compared to sample simulation results for the noise-free case. Monte-Carlo simulation results illustrate expected performance in the presence of measurement errors.

Case 2: Redundant Measurements. A third camera is added to the configuration of Case 1. Three sets of derived redundant measurements are provided through backprojection of measurements taken in any two of the three image planes. Monte-Carlo simulation results illustrate the expected performance of the redundant sliding mode observer and the relative improvements over the nonredundant Case 1.

Case 1: Nonredundant Measurements. In this Case, the true object motion is identical to that described in Case 1 of Section 2.6.2, with image plane feature positions and correspondence information being provided by the imagery generation program. Two cameras are employed with unity focal length. The position and orientation of the left camera (C_1) are defined by a translation $[\mathbf{T}_{C_1}^E]_E = [0, -5, -30]^T_E$ units, and transformation $\mathbf{I}_{C_1}^E$ formed from Euler angles $\psi_{C_1} = 10^\circ$ (yaw or pan), $\theta_{C_1} = -45^\circ$ (pitch or tilt), and $\phi_{C_1} = 0$ (roll about optical axis). The position and orientation of the right camera (C_2) are defined, respectively, by $[\mathbf{T}_{C_2}^E]_E = [0, 5, -30]^T_E$ units, and $\mathbf{I}_{C_2}^E$ formed from Euler angles $\psi_{C_2} = -10^\circ$, $\theta_{C_2} = -45^\circ$, and $\phi_{C_2} = 0$. Since some filtering is expected from the 2D feature point tracking systems of Level IV, simulations for this Chapter assume a modest level of uniformly distributed measurement errors, on the order of 2.5% of the object image size, in the image plane. Derived 3D measurements were then obtained over the entire trajectory using Step 1 of the single-frame initialization algorithm of Section 2.5.

Independent nonredundant sliding mode observes are used in each of the 3D coordinate axis (X, Y, Z) with states, position, velocity and acceleration. The extended state vector, for each of the three observers, is then $\mathbf{x}_e := [p, v, a]$, where $p \sim$ position, $v \sim$ velocity, and $a \sim$ acceleration. A Gauss-Markov acceleration model, identical in form to that described in Section 2.4.3, is employed to approximate the true translational motion of the feature point. In this case, the approximate dynamical model uses

$$\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) := \hat{\Phi} \hat{\mathbf{x}}_e(k), \quad (6.98)$$

where, with the sample period $T := 0.1\text{sec.}$, $\hat{\Phi} = \exp(\hat{\mathbf{A}}T)$ where

$$\hat{\mathbf{A}} := \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -\alpha_t \end{bmatrix}. \quad (6.99)$$

and α_t is the *reciprocal manoeuvre correlation time constant* for translational motion. The three measurement models, for each of the three observers, all have the same form:

$$\mathbf{z}(k) = \mathbf{h}(\hat{\mathbf{x}}_e(k)) + \zeta(k) = \mathbf{C}\hat{\mathbf{x}}_e(k) + \zeta(k) := \hat{\mathbf{x}}(k) + \zeta(k), \quad (6.100)$$

where $\mathbf{C} := [1, 0, 0]$, and $\zeta(k)$ represents measurement errors.

Each of the observers is constructed as

$$\hat{\mathbf{x}}_e(k+1) = \hat{\Phi} \hat{\mathbf{x}}_e(k) + \mathbf{L}v(k), \quad (6.101)$$

where

$$v(k) := (\mathbf{C}\mathbf{L}(k))^{-1} \{z(k+1) - \mathbf{C}\hat{\Phi}\hat{\mathbf{x}}_e(k) - \mathbf{g}(s_k)\}. \quad (6.102)$$

$$s_k := z(k) - \mathbf{C}\hat{\mathbf{x}}_e(k). \quad (6.103)$$

$$\mathbf{g}(s_k) := \mathbf{\Lambda}\text{Sat}(s_k, \epsilon) + \mathbf{\Gamma}s_k. \quad (6.104)$$

In this case, $z, s, \mathbf{g}, \mathbf{\Gamma}, \mathbf{\Lambda}$, and ϵ are scalars. Parameter values were selected from simulation studies as

$$\alpha_t := 0.$$

$$\mathbf{\Lambda} := -0.01.$$

$$\begin{aligned}
\Gamma &:= 0.01, \\
\epsilon &:= 0.012, \\
\mathbf{L} &:= [1, 8, 25]^T
\end{aligned} \tag{6.105}$$

for which the reaching law satisfies the hypothesis of Theorem 5.12 and also the conditions of Case 4 of Example 5.7 (page 173), which means that the scalar s_k retains the same sign until $|s_k| < |\Lambda|/|\Gamma| = 1$, and thereafter the state crosses the surface in each sample period.

Figure 6.1 shows sample results over a single run for this trajectory in the absence of measurement errors, but with artificially induced initialization errors in position estimates of 5 units. Velocity and acceleration are initialized as zero. Position errors, shown in plot a), during initial transients illustrate the familiar chattering motion commonly associated with discrete-time sliding mode systems, however, the magnitude of chattering decays exponentially according to the stability of the reaching law. Velocity errors, shown in plot b), contain fairly large initial transients due to the artificially high initialization errors, followed by small amplitude oscillations over the remainder of the trajectory. These oscillations result from rotational motion of the object and lag in the velocity channel of the observer. The switching functions s_k , shown in plot c), have entered the *chattering region* (crossing the surface in each sample period) following the first measurement event and thereafter decrease exponentially.

A simple approach can be used to apply the theory of this Chapter and Chapter 5 to obtain approximate bounds on the observer error, much like the approach illustrated in Example 5.6 (page 166). Neglecting measurement errors, the true system *jerk* (time derivative of acceleration) is assumed to be bounded according to

$$|\dot{a}(t)| \leq \mu < \infty. \tag{6.106}$$

The true continuous-time dynamics of the feature point can then be written with extended state vector $\mathbf{x}_e = [p, v, a]^T$ as

$$\dot{\mathbf{x}}_e(t) = \mathbf{A}\mathbf{x}_e(t) + \mathbf{B}\eta(t). \tag{6.107}$$

where $\mathbf{A}$ is a companion form matrix,

$$\mathbf{A} := \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad (6.108)$$

and $\mathbf{B} := [0, 0, 1]^T$, and from (6.106), $|\eta(t)| \leq \mu$. Discretization then yields $\mathbf{x}_e(k+1) = \Phi \mathbf{x}_e(k) + \xi(k)$, where $\Phi = \exp(\mathbf{A}T)$ and $\xi(k)$ satisfies, $\forall \mathbf{Q} \in \mathbb{C}^{3 \times 3}$, $\|\mathbf{Q}\xi(k)\| \leq \|\mathbf{Q}\mathbf{B}_d\|\mu$, where $\mathbf{B}_d := [\int_0^T \exp(\mathbf{A}\theta) d\theta] \mathbf{B}$. With

$$\mathbf{F} := [\mathbf{I}_3 - \mathbf{L}(\mathbf{C}\mathbf{L})^{-1}\mathbf{C}], \quad (6.109)$$

and a matrix $\mathbf{W} \in \mathbb{C}^{3 \times 3}$ to be specified, and defining $\mathbf{Q} := \mathbf{W}\mathbf{F}$, the scalar M of Theorem 5.11 satisfies

$$M := \sup_{\{k \geq 0\}} \|\mathbf{Q}\xi(k)\| \leq \|\mathbf{W}\mathbf{F}\mathbf{B}_d\|\mu. \quad (6.110)$$

Using the Euclidean vector norm, $\|\cdot\| = \|\cdot\|_2$, and the induced matrix norm, $\|\cdot\| = \|\cdot\|_2$, $\mathbf{W} = [w_{ij}]$ can be taken as upper triangular with $w_{11} = 1$, and $w_{ii} > 0$. Noting that here $\Phi = \hat{\Phi}$, Theorem 5.11 states that, provided $\beta := \|\mathbf{W}\mathbf{F}\Phi\mathbf{W}^{-1}\|_2 < 1$, the ultimate bounds are

$$\lim_{k \rightarrow \infty} \|\mathbf{e}_e(k)\|_2 \leq E(\mathbf{W}) := \|\mathbf{W}^{-1}\|_2 \frac{M}{(1-\beta)}. \quad (6.111)$$

where $\mathbf{e}_e(k) := \mathbf{x}_e(k) - \hat{\mathbf{x}}_e(k)$. For this trajectory, the bound μ is estimated as $\mu \leq 26$ and the vector $\mathbf{B}_d = [0.000167, 0.005, 0.1]^T$. Numerical minimization of $E(\mathbf{W})$ using an initial estimate derived from Method 1 of Section 5.4 gives $E_{\min} \approx 9$ with $w_{12} = 0.0053$, $w_{13} = -0.0006$, $w_{22} = 0.032$, $w_{23} = -0.0036$, and $w_{33} = 0.0031$. Plot d) of Figure 6.1 shows the Euclidean norm of observer errors for the extended state over this trajectory in comparison to the reasonably tight although approximate bounds provided by this simple procedure.

Figure 6.2 shows the true position and velocity of x -, y -, and z -components of the feature point in plots a) and b), and root-mean-squared errors (RMSE) and mean errors (ME) from a 60-run Monte Carlo simulation in the presence of measurement errors. Mean errors in position estimation (plots c, e, and g) indicate negligible bias, and RMS errors remain under 0.2 units over the trajectory. The distinct peaks seen in RMS errors at about 3s

and 6.5s are due largely to the fact that this feature point lies farthest from the cameras at these times (see frames 31 and 63 of Figure 2.10, page 68) which leads to larger derived measurement errors over these portions of the trajectory. Mean errors for velocity estimation (plots d, f, and h) indicate a sinusoidal bias of amplitude less than 1 unit/s which results primarily from observer lag in the velocity channel. RMS velocity errors remain within 2 units/s over the trajectory.

Figure 6.1: Example simulation results of sliding mode observer for tracking 3D feature point through a linear non-redundant measurement model for the ideal case of no measurement errors. Plots show a) position and b) velocity errors for each of the three observers. c) values of switching functions (s_k), and d) Euclidean norm of observer errors in comparison to approximate bounds.

Figure 6.2: Monte Carlo simulation results of sliding mode observer for tracking 3D feature point through a linear non-redundant measurement model.

Figure 6.3: Effect of third camera placement for estimation with redundant measurements.

Case 2: Redundant Measurements. Two primary motivations for the use of multiple base-line stereo (more than two cameras) is to achieve improved robustness in the correspondence problem, as discussed in Chapter 2, and to reduce sensitivity to measurement errors through fusion of redundant measurements. In order to obtain redundant linear measurements in this Case, a third camera is added to the stereo pair employed in Case 1. Three sets of derived 3D measurements are obtained by taking corresponding feature positions from any two of the three cameras and using Step 1 of the single-frame initialization algorithm of Section 2.5. The redundant sliding mode observer developed in this Chapter is then applied to estimate the position and motion of the single feature point p^6 over the same trajectory as considered in Case 1.

State estimation through fusion of redundant measurements can lead to improved results over state estimation based only on a subset of available measurements provided appropriate weighting, through the matrices denoted by $\mathbf{P}_k$ in this chapter, is employed. This concept, in the context of the current problem, is illustrated in Figure 6.3. Cameras C_1 and C_2 represent the stereo pair employed in Case 1, with a baseline $b_{12} = 10$ units. The length of the baseline, to a large extent, dictates error levels in estimating 3D positions from corresponding feature positions in the stereo image pairs. If a third camera is placed at position C_c (shown as the dashed outline), then the three baselines are $b_{12} = 10$ units, and

$b_{1c} = b_{2c} = 5$ units. In this configuration, derived 3D measurements from pairs (C_1, C_c) and (C_2, C_c) have standard deviations almost twice as large as the derived measurements from pair (C_1, C_2) . For example, for a stationary point at position $[45, 0, 0]_E^T$ units (approximately the largest depth achieved by p^6 over the trajectory), considering only depth measurements (X-component), and letting σ_X^{ij} represent the standard deviation of measurement errors resulting from the use of camera pair (C_i, C_j) , $i, j = 1, c, 2$, simulations suggest that $\sigma_X^{12} \approx 0.22$ units, while $\sigma_X^{1c} \approx \sigma_X^{2c} \approx 0.48$ units. Following the development of Equation (6.97) (page 198) with

$$\begin{aligned} \mathbf{R}_X &:= \text{diag}\{(\sigma_X^{1c})^2, (\sigma_X^{2c})^2, (\sigma_X^{12})^2\}, \quad \text{and} \\ \mathbf{P}_X &:= \mathbf{R}_X^{-1}, \end{aligned} \quad (6.112)$$

the error covariance in X -position estimation following application of the redundant sliding mode observer is expected to be, with $\mathbf{H} := [1, 1, 1]^T$ and $\mathbf{e} := p_X - \hat{p}_X$,

$$\mathcal{E}\{\mathbf{e}(k)\mathbf{e}^T(k)\} = [\mathbf{H}^T \mathbf{R}_X^{-1} \mathbf{H}]^{-1} \approx (0.185)^2. \quad (6.113)$$

Instead, if the third camera is placed at location C_0 shown in Figure 6.3, then approximately $\sigma_X^{01} \approx \sigma_X^{02} \approx \sigma_X^{12} \approx 0.22$ units and (6.113) yields an expected standard deviation of $0.22/\sqrt{3} \approx 0.12$ units (when the feature point is at position $[45, 0, 0]_E^T$ units). Figure 6.4 shows root-mean-squared errors (RMSE) and mean errors (ME) from a 60-run Monte Carlo simulation for the three-camera triad (C_1, C_2, C_0) . The trajectory of the object is identical to that for Case 1, as is the imaging geometry for cameras C_1 and C_2 . The third camera, C_0 , has position $[\mathbf{T}_{C_0}^E]_E = [0, 0, -21.4]_E^T$ units and orientation formed from Euler angles $\psi_{C_0} = 0^\circ$, $\theta_{C_0} = -35^\circ$, and $\phi_{C_0} = 0$.

Independent redundant sliding mode observers are used in each of the 3D coordinate axis (X, Y, Z) with states, position, velocity and acceleration. A Gauss-Markov acceleration model, identical to that used in Case 1, is employed to approximate the true translational motion of the feature point. The three measurement models, for each of the three observers, all have the same form. Let $z_{ij}(k)$ represent the derived position measurements from camera pair (C_i, C_j) , $i, j = 0, 1, 2$. The measurement models can then be written as

$$\begin{aligned} \mathbf{z}(k) &= [z_{01}, z_{02}, z_{12}]^T \\ &= \mathbf{H}\mathbf{C}\mathbf{x}_e(k) + \boldsymbol{\zeta}(k), \end{aligned} \quad (6.114)$$

where $\mathbf{H} := [1, 1, 1]^T$, $\mathbf{C} := [1, 0, 0]$, and $\zeta(k)$ represents the vector of measurement errors.

Each of the observers is constructed as

$$\hat{\mathbf{x}}_e(k+1) = \hat{\Phi} \hat{\mathbf{x}}_e(k) + \mathbf{L} \mathbf{v}(k), \quad (6.115)$$

where

$$\begin{aligned} \mathbf{v}(k) &:= [\mathbf{H}^T \mathbf{P} \mathbf{H} \mathbf{C} \mathbf{L}]^{-1} \\ &\quad \cdot \{ \mathbf{H}^T \mathbf{P} [\mathbf{z}(k+1) - \mathbf{H} \mathbf{C} \hat{\Phi} \hat{\mathbf{x}}_e(k)] - \mathbf{g}(\mathbf{s}_k) \}, \end{aligned} \quad (6.116)$$

$$\mathbf{s}_k := \mathbf{H}^T \mathbf{P} [\mathbf{z}(k) - \mathbf{H} \mathbf{C} \hat{\mathbf{x}}_e(k)], \quad (6.117)$$

$$\mathbf{g}(\mathbf{s}_k) := \Lambda \text{Sat}(\mathbf{s}_k, \epsilon) + \Gamma \mathbf{s}_k. \quad (6.118)$$

$$(6.119)$$

In this case, $\mathbf{s}$, $\mathbf{g}$, Γ , Λ , and ϵ are scalars. Parameter values are identical to those employed in Case 1, namely

$$\begin{aligned} \alpha_t &:= 0, \\ \Lambda &:= -0.01, \\ \Gamma &:= 0.01, \\ \epsilon &:= 0.012, \\ \mathbf{L} &:= [1.8, 25]^T \end{aligned} \quad (6.120)$$

for which the reaching law satisfies the hypothesis of Theorem 5.12. The conditions of Case 4 of Example 5.7 (page 173) are also satisfied which means that the scalar $\mathbf{s}_k$ retains the same sign until $|\mathbf{s}_k| < |\Lambda|/|\Gamma| = 1$, and thereafter the state crosses the surface in each sample period. The weighting matrix $\mathbf{P}$ is set to the identity matrix since redundant measurements within each observer have approximately the same error levels due to the three-camera geometry.

Figure 6.4 shows the true position and velocity of x -, y -, and z -components of the feature point in plots a) and b), and root-mean-squared errors (RMSE) and mean errors (ME) from a 60-run Monte Carlo simulation in the presence of measurement errors. Mean errors in position estimation (plots c, e, and g) indicate negligible bias, and RMS errors

remain under 0.13 units over the trajectory. These results agree very well with the predicted position estimation variance of 0.12 units when the feature point is furthest from the cameras (at approximately 3 and 6.5 seconds). Mean errors for velocity estimation (plots d, f, and h) again indicate a sinusoidal bias of amplitude less than 1 unit/s and are almost identical to those for the nonredundant Case. RMS velocity errors, however, show a slight improvement over those for the nonredundant case and remain within 2 units/s over the trajectory.

Figure 6.4: Monte Carlo simulation results of sliding mode observer for tracking 3D feature point through a linear redundant measurement model.

6.7 Chapter Summary

Sliding mode observers proposed and analyzed in this chapter are specific to that class of state estimation problems which are governed by possibly nonlinear time-varying dynamics, known up to some bounded propagation error over any sample period, and which provide measurements that are linearly related to elements of the extended state vector. Existence of the projection operation which provides the robustness mechanism requires alternate observer forms depending on whether the measurement model is minimal or redundant. Concepts of linear measurement redundancy were proposed which result in a simple rank test to determine which form of sliding mode observer is required. Observer forms, and stability and robustness properties have been detailed for both nonredundant and redundant measurements.

The following chapter extends these methods to the case of nonlinear measurement models. Again, measurement redundancy must be examined to select the appropriate form of sliding mode observer. Extensions of linear measurement redundancy to the nonlinear case are developed which again results in a rank test to select the appropriate observer form. *Ideal* observer forms, however, resulting from direct extensions of methods from the linear measurement case do not intuitively suggest practical implementation strategies, but are well-suited to stability analysis with techniques proposed in Chapter 5. Alternate ideal forms are proposed in the following chapter which do lead to practical implementation methods through local iterations at each measurement event.

Chapter 7

Sliding Mode Observers for Systems With Nonlinear Measurement Models

7.1 Introduction

The previous Chapter proposed sliding mode observers for discrete-time systems with nonlinear time-varying dynamics and linear nonredundant or redundant measurement models. In this Chapter, methods for sliding mode observer design are proposed for systems again with nonlinear time-varying dynamics,

$$\mathbf{x}_e(k+1) = \mathbf{f}(k, \mathbf{x}_e(k)), \quad (7.1)$$

where $\mathbf{x}_e(k) \in \mathbf{X}_e \subseteq \mathbb{R}^n$ is the *extended* state vector and $\mathbf{f} : \mathbb{N}_0 \times \mathbf{X}_e \mapsto \mathbf{X}_e$, but which have nonlinear measurement models of the form

$$\mathbf{z}(k) = \mathbf{h}_t(\mathbf{x}_e(k)) = \mathbf{h}(\mathbf{x}(k)). \quad (7.2)$$

where, following the same notation as in Chapter 6, the *measured state*, $\mathbf{x}$, is related to the extended state, $\mathbf{x}_e$, by

$$\mathbf{x}(k) := \mathbf{C}\mathbf{x}_e(k), \quad \mathbf{C} = [\mathbf{I}_q, \mathbf{0}_{q \times (n-q)}]. \quad (7.3)$$

and $\mathbf{h} : \mathbf{X} \mapsto \mathbb{R}^m$, $\mathbf{X} = \mathbf{C}\mathbf{X}_e \subseteq \mathbb{R}^q$, is C^p , $p \geq 1$.

The function of an observer, in the simplest terms, includes inversion of the measured state to measurements map, $\mathbf{h}(\mathbf{x})$, together with estimation of the extended state, $\mathbf{x}_e$, subject

to the dynamics in (7.1). The use of observers for continuous-time nonlinear map inversion was recently proposed by Nicosia *et al.* [92, 93]. In their approach, differential geometric methods are employed to derive a “direct extended map” which relates the output, $z(t)$, and its time derivatives to the measured state, $x(t)$, and its time derivatives, which results in a dynamical model in a companion form. The inverse of the direct extended map, if it exists, can be found by solving a coupled set of nonlinear partial differential equations provided time derivatives of the output are available. Alternatively, the extended map can be placed in nonlinear observer canonical form for which an observer can be constructed. They propose selection of observer gains which results in an observer structure that corresponds to a generalized Newton algorithm. A similar approach for discrete-time systems has been proposed by Grizzle and Moraal [76] who considered direct inversion of the observability map (from (3.12), page 82)

$$\begin{bmatrix} z(k - N + 1) \\ z(k - N + 2) \\ \vdots \\ z(k) \end{bmatrix} = \mathcal{O}_N(x(k - N + 1)). \quad (7.4)$$

where

$$\mathcal{O}_N(x) = \begin{bmatrix} h(x) \\ h \circ f(x) \\ h \circ f^2(x) \\ \vdots \\ h \circ f^{N-1}(x) \end{bmatrix}. \quad (7.5)$$

and $N = n$. When direct inversion is not tractable, Newton’s algorithm is proposed to iteratively obtain, through off-line computation, solutions for system (7.4) at each point of a dense grid in the image of the observability map. The observer is then constructed by producing an approximate solution to the observability map through an interpolation/extrapolation scheme based upon solutions at the grid points. The important point is that the use of Newton’s method for practical implementation of observers for nonlinear systems has previously been accepted as a viable approach.

In this Chapter, observer design for nonlinear measurements proceeds in a manner similar to linearization, but using instead the Fundamental Theorem of Integral Calculus (The-

orem 5.9, page 148). Section 7.3.1 reviews this Theorem and presents notation and additional preliminary results employed throughout the remainder of this Chapter. Again, design of the sliding mode observers proposed in this Chapter require different strategies depending on whether or not the measurement model is redundant. Concepts of nonlinear measurement redundancy, proposed in Section 7.2, are a direct extension of those proposed for linear measurement models and draw on some basic results from differential geometry.

The *ideal* observers for the nonredundant or redundant cases take on very similar forms as those proposed in Sections 6.4 and 6.5, respectively. In the case of linear measurement models these ideal forms can be realized directly. In the nonlinear case, however, ideal forms for the observers represent systems of nonlinear integral equations which implicitly define the observer output. Implementation may proceed with a simple approximation, much like that used in the Extended Kalman Filter, which can yield good performance in some cases. More general results including stability analysis are obtained by introducing local iterations at each observation event, much like those employed in the Iterated Extended Kalman Filter [94], which yield Newton-type iterations for nonredundant measurements, or Gauss-Newton-type iterations for redundant measurements. Sections 7.3 and 7.4 present, for the nonredundant and redundant cases respectively, the ideal observer form, stability and robustness properties, followed by implementation with local iterations. Complete analysis of effects of measurement errors on observers proposed in this Chapter are not yet available. Discussions of the effects of measurement errors are included which suggest possible approaches for future work. Stability analysis, which yield sufficient conditions for the numerical implementation to approximate the ideal observer with arbitrary accuracy, is given for the iterated nonredundant observer. Just as in Chapter 6, observer design for redundant measurements focusses on minimization of a nonlinear cost function. Simulation results for of the proposed observers when applied to components of the image analysis problem are given in Section 7.5.

7.2 Nonlinear Measurement Redundancy

The purpose of this Section is to quantify what is meant by a redundant nonlinear measurement model. Just as in Chapter 6, measurement redundancy refers to the case in which

more measurements are available than the minimum required to uniquely reconstruct the extended state of the system under ideal conditions (no disturbances, unmodelled dynamics, or measurement noise). Again, the objective is not to propose methods to remove redundant measurements, but rather to define clearly what is meant by the term *redundant measurements* and to provide a means to determine whether or not the measurement model is redundant so that the proper form of sliding mode observer can be selected.

In Chapter 6, rank deficiency in the linear state-to-measurement map was sufficient to specify a single reduced-order measurement model for the whole state space and still retain uniform observability. In the case of general nonlinear measurement models, a similar approach can be taken, however the analysis is local rather than global. The following proposed definition extends the notions of Definition 6.1 to nonlinear measurement models.

Definition 7.1: For system (7.1) with measurement model (7.2), which is assumed to be uniformly observable in X_e with N measurement events, the measurements are said to be *redundant* if at the k th measurement event, $k = 0, 1, 2, 3, \dots$, there exists a function $\gamma_k : \mathcal{B}^{(k)} \mapsto \mathbb{R}^r$, $r < m$, defined and continuously differentiable on some open neighbourhood $\mathcal{B}^{(k)} \subset \mathbb{R}^m$ of $h_e(x_e(k))$, such that system (7.1) with a reduced order measurement model

$$z'(k) = \gamma_k \circ h_e(x_e(k)) \in \mathbb{R}^r, \quad (7.6)$$

is uniformly observable in X_e with N measurement events. $\square$

Definition 7.1 captures the intuitive notion that if the extended state can be uniquely recovered from the measurements in N measurement events, but there are more measurements than the minimum required to uniquely reconstruct the extended state of the system, then it should be possible, at each measurement event, to apply some possibly nonlinear transformation to the available m measurements in order to yield $r < m$ *derived* measurements such that the extended state can still be uniquely recovered from these derived measurements in N measurement events. The functions $\gamma_k(\cdot)$ must be defined on some open set containing the measurements $h_e(x_e(k))$ at time k , and must be continuously differentiable on this open set so that the Jacobian of the reduced-order measurement model can be computed. The collection of functions from which each $\gamma_k(\cdot)$ is selected need not be finite, nor even countable. For practical considerations, it is useful to establish conditions under which $\gamma_k(\cdot)$ can be selected, at time k , from a finite collection of functions.

A sufficient condition for the measurement model to be redundant is formulated in terms of the rank of the state-to-measurement map in (7.2), i.e. the rank of the Jacobian matrix of $\mathbf{h}_e(\mathbf{x}_e)$, through the use of the following fundamental result from differential geometry.

Theorem 7.1: Rank Theorem. Let $\mathcal{A}_0 \subset \mathbb{R}^n$, $\mathcal{B}_0 \subset \mathbb{R}^m$ be open sets, $\mathbf{F} : \mathcal{A}_0 \mapsto \mathcal{B}_0$ be a C^p , $p \geq 1$, mapping, and suppose that $\mathbf{F}$ has constant rank r on $\mathcal{A}_0$. Then for any $\mathbf{x} \in \mathcal{A}_0$, there exist open sets $\mathcal{A} \subset \mathcal{A}_0$ and $\mathcal{B} \subset \mathcal{B}_0$, with $\mathbf{x} \in \mathcal{A}$ and $\mathbf{F}(\mathbf{x}) \in \mathcal{B}$, and there exist C^p diffeomorphisms $\mathbf{G} : \mathcal{A} \mapsto \mathcal{U} \subset \mathbb{R}^n$ and $\mathbf{H} : \mathcal{B} \mapsto \mathcal{V} \subset \mathbb{R}^m$, $\mathcal{U}$ and $\mathcal{V}$ open, such that

$$\mathbf{H} \circ \mathbf{F} \circ \mathbf{G}^{-1}(\mathcal{U}) \subset \mathcal{V}, \quad (7.7)$$

and such that this map has, for any $\mathbf{y} \in \mathcal{U}$, the simple form

$$\mathbf{H} \circ \mathbf{F} \circ \mathbf{G}^{-1}(\mathbf{y}) = \mathbf{P}\mathbf{y}. \quad (7.8)$$

where

$$\mathbf{P} := \begin{bmatrix} \mathbf{I}_r & \mathbf{0}_{r \times n-r} \\ \mathbf{0}_{m-r \times r} & \mathbf{0}_{m-r \times n-r} \end{bmatrix}. \quad (7.9)$$

Proof: See Boothby [163, Theorem 7.1]. $\square$

Theorem 7.1 states that a mapping of constant rank r behaves locally like a projection of $\mathbb{R}^n = \mathbb{R}^r \times \mathbb{R}^{n-r}$ to $\mathbb{R}^r$, followed by injection of $\mathbb{R}^r$ onto $\mathbb{R}^r \times \{\mathbf{0}\} \subset \mathbb{R}^r \times \mathbb{R}^{m-r} = \mathbb{R}^m$. The result of Theorem 7.1 is local in the sense that functions $\mathbf{G}$ and $\mathbf{H}$, which give the simple result in (7.8), are defined only locally about $\mathbf{x}$ and $\mathbf{F}(\mathbf{x})$, respectively. It is interesting also to note the similarity between this result for nonlinear systems and the singular value decomposition for linear systems: if $\mathbf{F}(\mathbf{x}) = \mathbf{A}\mathbf{x}$, $\mathbf{A} \in \mathbb{R}^{m \times n}$, $m < n$, with $\text{rank}\{\mathbf{A}\} = r < m$, and

$$\mathbf{A} = \mathbf{Q}_1 \mathbf{\Sigma} \mathbf{Q}_2^T \quad (7.10)$$

is the singular value decomposition of $\mathbf{A}$, then $\forall \mathbf{y} \in \mathbb{R}^n$,

$$\mathbf{\Sigma}^+ \mathbf{Q}_1^T \mathbf{A} \mathbf{Q}_2 \mathbf{y} = \mathbf{P} \mathbf{y} \quad (7.11)$$

where

$$\mathbf{\Sigma} = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_r, 0, \dots, 0\} \in \mathbb{R}^{m \times n}, \quad (7.12)$$

$$\mathbf{\Sigma}^+ := \text{diag}\{\sigma_1^{-1}, \sigma_2^{-1}, \dots, \sigma_r^{-1}, 0, \dots\} \in \mathbb{R}^{n \times m}, \quad (7.13)$$

so that $\mathbf{\Sigma}^+$ is the pseudoinverse of $\mathbf{\Sigma}$, and $\mathbf{P} = \mathbf{\Sigma}^+ \mathbf{\Sigma}$ is as defined in (7.9).

Definition 7.2: Let $G : \mathcal{D} \subseteq \mathbb{R}^n \mapsto \mathbb{R}^m$, for some open convex set $\mathcal{D}$, be C^p , $p \geq 1$. For any $\xi \in \mathcal{D}$ define $J_G(\xi)$ to be the Jacobian matrix of G evaluated at ξ ,

$$J_G(\xi) := \left[\frac{\partial G(\mathbf{x})}{\partial \mathbf{x}} \right]_{\mathbf{x}=\xi}. \quad (7.14)$$

For $n = m$ and $J_G(\xi)$ nonsingular, the inverse of this matrix is denoted $J_G^{-1}(\xi)$; this does not represent the inverse of the function $J_G : \mathcal{D} \mapsto \mathbb{R}^{m \times n}$. $\square$

The following Theorem provides sufficient conditions for a nonlinear measurement model to be redundant and under which the $\gamma_k(\cdot)$ of Definition 7.1 can be selected from a finite collection of functions.

Theorem 7.2: Consider a system of the form

$$\mathbf{x}_e(k+1) = \mathbf{f}(\mathbf{x}_e(k)), \quad (7.15)$$

with measurement model

$$\mathbf{z}(k) = \mathbf{h}_e(\mathbf{x}_e(k)), \quad (7.16)$$

where $\mathbf{x}_e \in X_e$, X_e is a compact subset of $\mathbb{R}^n$, $\mathbf{f} : X_e \mapsto X_e$ is continuously differentiable, and $\mathbf{h}_e : \mathcal{A}_0 \subset \mathbb{R}^n \mapsto \mathcal{B}_0 \subset \mathbb{R}^m$, is C^p , $p \geq 1$, with $\mathcal{A}_0$ and $\mathcal{B}_0$ open sets and $X_e \subset \mathcal{A}_0$. Assume the system is uniformly observable (see page 82) in X_e with N measurement events. If, $\forall \xi \in \mathcal{A}_0$, $\text{rank}\{J_{\mathbf{h}_e}(\xi)\} = r < m$, then the system has redundant measurements and, for each k , $\gamma_k(\cdot)$ of Definition 7.1 can be selected from a finite collection of functions.

Proof: The hypothesis state that $\mathbf{h}_e(\cdot)$ is defined and C^p , $p \geq 1$, on some open set $\mathcal{A}_0$ which contains the compact (closed, bounded) set X_e , and that $\forall \xi \in \mathcal{A}_0$, $\text{rank}\{J_{\mathbf{h}_e}(\xi)\} = r < m$. Under these conditions, for each $\mathbf{x}_e \in X_e$, Theorem 7.1 guarantees the existence of an open set $\mathcal{A}^{(\mathbf{x}_e)} \subset \mathcal{A}_0$ and associated open sets $\mathcal{B}^{(\mathbf{x}_e)} \subset \mathcal{B}_0$, $\mathcal{U}^{(\mathbf{x}_e)} \subset \mathbb{R}^n$, and $\mathcal{V}^{(\mathbf{x}_e)} \subset \mathbb{R}^m$, and associated C^p diffeomorphic functions $\mathbf{H}^{(\mathbf{x}_e)}$ and $\mathbf{G}^{(\mathbf{x}_e)}$ (following the notation of Theorem 7.1) such that $\forall \xi \in \mathcal{A}^{(\mathbf{x}_e)}$, upon setting $\mathbf{y} = \mathbf{G}^{(\mathbf{x}_e)}(\xi)$,

$$\mathbf{H}^{(\mathbf{x}_e)} \circ \mathbf{h}_e(\xi) = \mathbf{P} \mathbf{G}^{(\mathbf{x}_e)}(\xi), \quad (7.17)$$

where $\mathbf{P}$ is defined in (7.9). This results in a denumerable collection of open sets

$$\mathcal{C} := \{\mathcal{A}^{(\mathbf{x}_e)} \mid \mathbf{x}_e \in X_e\} \quad (7.18)$$

for which $\mathbf{x}_e \in \mathcal{A}^{(\mathbf{x}_e)}$. This collection, $\mathcal{C}$, of open sets satisfies

$$\mathbf{X}_e \subset \bigcup_{\{\mathbf{x}_e \in \mathbf{X}_e\}} \mathcal{A}^{(\mathbf{x}_e)}, \quad (7.19)$$

and hence $\mathcal{C}$ is an open cover of the compact set $\mathbf{X}_e$. The Heine-Borel Theorem (see, for example, Marsden [144, Sec. 3.1, Theorem 1]) states that, because $\mathbf{X}_e$ is compact, every open cover of $\mathbf{X}_e$ contains a finite subcover which covers $\mathbf{X}_e$. Hence there exists a finite integer constant $M > 0$ and a finite collection, $\tilde{\mathcal{C}}$, of open sets taken from $\mathcal{C}$,

$$\tilde{\mathcal{C}} := \{\mathcal{A}^{(j)} \mid j = 1, 2, \dots, M, \mathcal{A}^{(j)} \in \mathcal{C}\} \quad (7.20)$$

such that

$$\mathbf{X}_e \subset \bigcup_{j \in \{1, 2, \dots, M\}} \mathcal{A}^{(j)}. \quad (7.21)$$

Associate with each $\mathcal{A}^{(j)}$, the corresponding open sets $\mathcal{B}^{(j)} \subset \mathcal{B}_0$, $\mathcal{U}^{(j)} \subset \mathbb{R}^n$, and $\mathcal{V}^{(j)} \subset \mathbb{R}^m$, and C^p diffeomorphic functions $\mathbf{H}^{(j)}$ and $\mathbf{G}^{(j)}$.

At the k th measurement event, the extended state $\mathbf{x}_e(k) \in \mathcal{A}^{(j)}$ for some $j \in \{1, 2, \dots, M\}$. Since $\mathbf{x}_e(k)$ could lie in more than one of the sets $\mathcal{A}^{(j)}$, a simple selection strategy, since there are now only a finite number of sets $\mathcal{A}^{(j)}$, is to choose the minimum index, $j_k \in \{1, 2, \dots, M\}$, for which $\mathbf{x}_e(k) \in \mathcal{A}^{(j_k)}$. The functions $\mathbf{H}^{(j_k)}$ and $\mathbf{G}^{(j_k)}$ are those associated with $\mathcal{A}^{(j_k)}$, and also note that $\mathbf{x}_e(k) \in \mathcal{A}^{(j_k)}$ implies that $\mathbf{h}_e(\mathbf{x}_e(k)) \in \mathcal{B}^{(j_k)}$.

Consider, then, the reduced-order measurement model

$$\mathbf{z}'(k) = \gamma_k \circ \mathbf{h}_e(\mathbf{x}_e(k)). \quad (7.22)$$

where the functions $\gamma_k : \mathcal{B}^{(j_k)} \mapsto \mathbb{R}^r$ are selected at the k th measurement event according to

$$\begin{aligned} \gamma_k(\eta) &:= \mathbf{CH}^{(j_k)}(\eta), \quad \forall \eta \in \mathcal{B}^{(j_k)}, \\ j_k &:= \min \left\{ \ell \mid \ell \in \{1, 2, \dots, M\}, \mathbf{x}_e(k) \in \mathcal{A}^{(\ell)} \right\}, \end{aligned} \quad (7.23)$$

and

$$\mathbf{C} := [\mathbf{I}_r, \mathbf{0}_{r \times n-r}]. \quad (7.24)$$

Note that γ_k is a C^p mapping from $\mathcal{B}^{(j_k)}$ into $\mathcal{CV}^{(j_k)}$, and that

$$\begin{aligned}\gamma_k \circ \mathbf{h}_e(\mathbf{x}_e(k)) &= \mathbf{CH}^{(j_k)} \circ \mathbf{h}_e(\mathbf{x}_e(k)) \\ &= \mathbf{CPG}^{(j_k)}(\mathbf{x}_e(k)),\end{aligned}\tag{7.25}$$

where, from the definitions of $\mathbf{C}$ and $\mathbf{P}$, $\mathbf{CP} = \mathbf{C}$.

It is now required to show that system (7.15) with the reduced-order measurement model (7.22) is uniformly observable in N measurement events. But $\forall \mathbf{x}_e \in \mathbf{X}_e$, since $\mathbf{f} : \mathbf{X}_e \mapsto \mathbf{X}_e$ implies that the k th composition of $\mathbf{f}$ with itself satisfies $\mathbf{f}^k : \mathbf{X}_e \mapsto \mathbf{X}_e$, there exists a set of indices $\{j_\ell \in \{1, 2, \dots, M\} \mid \ell = 1, 2, \dots, N-1\}$ such that the observability mapping, $\mathcal{O}'(\mathbf{x}_e)$, for system (7.15) with reduced-order measurements (7.22) can be written, using (7.25), as

$$\mathcal{O}'(\mathbf{x}_e) = \text{diag} \left\{ \underbrace{\mathbf{C}, \mathbf{C}, \dots, \mathbf{C}}_{N \text{ copies}} \right\} \mathcal{O}''(\mathbf{x}_e),\tag{7.26}$$

where

$$\mathcal{O}''(\mathbf{x}_e) := \begin{bmatrix} \mathbf{PG}^{(j_0)}(\mathbf{x}_e) \\ \mathbf{PG}^{(j_1)} \circ \mathbf{f}(\mathbf{x}_e) \\ \mathbf{PG}^{(j_2)} \circ \mathbf{f}^2(\mathbf{x}_e) \\ \vdots \\ \mathbf{PG}^{(j_{N-1})} \circ \mathbf{f}^{N-1}(\mathbf{x}_e) \end{bmatrix}.\tag{7.27}$$

Let $\mathbf{J}_{\mathcal{O}'}(\mathbf{x}_e)$ and $\mathbf{J}_{\mathcal{O}''}(\mathbf{x}_e)$ be the Jacobian matrices of $\mathcal{O}'(\mathbf{x}_e)$ and $\mathcal{O}''(\mathbf{x}_e)$, respectively. Then, from the definition of $\mathbf{C}$ in (7.24), $\mathbf{J}_{\mathcal{O}'}(\mathbf{x}_e)$ differs from $\mathbf{J}_{\mathcal{O}''}(\mathbf{x}_e)$ only in that rows numbered

$$\begin{aligned}r+1, r+2, \dots, m; m+r+1, m+r+2, \dots, 2m; \dots; \\ (N-1)m+r+1, (N-1)m+r+2, \dots, Nm,\end{aligned}\tag{7.28}$$

which are identically zero in $\mathbf{J}_{\mathcal{O}''}(\mathbf{x}_e)$, are removed to form $\mathbf{J}_{\mathcal{O}'}(\mathbf{x}_e)$. Since removal of zero rows does not affect the rank of a matrix,

$$\text{rank} \{ \mathbf{J}_{\mathcal{O}'}(\mathbf{x}_e) \} = \text{rank} \{ \mathbf{J}_{\mathcal{O}''}(\mathbf{x}_e) \}.\tag{7.29}$$

The problem is thus reduced to showing that $\text{rank} \{ \mathbf{J}_{\mathcal{O}''}(\mathbf{x}_e) \} = n$.

System (7.15) with original measurement model (7.16) is assumed to be uniformly observable in N measurement events. The observability mapping, $\mathcal{O}(\mathbf{x}_e)$, of the original system can be written as

$$\begin{aligned} \mathcal{O}(\mathbf{x}_e) &:= \begin{bmatrix} \mathbf{h}_e(\mathbf{x}_e) \\ \mathbf{h}_e \circ \mathbf{f}(\mathbf{x}_e) \\ \mathbf{h}_e \circ \mathbf{f}^2(\mathbf{x}_e) \\ \vdots \\ \mathbf{h}_e \circ \mathbf{f}^{N-1}(\mathbf{x}_e) \end{bmatrix} \\ &= \begin{bmatrix} (\mathbf{H}^{(j_0)})^{-1} \circ (\mathbf{P}\mathbf{G}^{(j_0)}) (\mathbf{x}_e) \\ (\mathbf{H}^{(j_1)})^{-1} \circ (\mathbf{P}\mathbf{G}^{(j_1)}) \circ \mathbf{f}(\mathbf{x}_e) \\ (\mathbf{H}^{(j_2)})^{-1} \circ (\mathbf{P}\mathbf{G}^{(j_2)}) \circ \mathbf{f}^2(\mathbf{x}_e) \\ \vdots \\ (\mathbf{H}^{(j_{N-1})})^{-1} \circ (\mathbf{P}\mathbf{G}^{(j_{N-1})}) \circ \mathbf{f}^{N-1}(\mathbf{x}_e) \end{bmatrix}. \end{aligned} \quad (7.30)$$

The Jacobian, $\mathbf{J}_{\mathcal{O}}(\mathbf{x}_e)$, of $\mathcal{O}(\mathbf{x}_e)$, which is assumed to be of rank n for all $\mathbf{x}_e \in \mathbf{X}_e$, can be written, using the chain rule, as

$$\mathbf{J}_{\mathcal{O}}(\mathbf{x}_e) = \text{diag} \left\{ \left. \frac{\partial (\mathbf{H}^{(j_\ell)})^{-1}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} \right|_{\boldsymbol{\xi}=\boldsymbol{\xi}_\ell(\mathbf{x}_e)} : \ell = 0, 1, \dots, N-1 \right\} \mathbf{J}_{\mathcal{O}''}(\mathbf{x}_e). \quad (7.31)$$

where

$$\boldsymbol{\xi}_\ell(\mathbf{x}_e) := (\mathbf{P}\mathbf{G}^{(j_\ell)}) \circ \mathbf{f}^\ell(\mathbf{x}_e). \quad (7.32)$$

But then

$$\mathbf{J}_{\mathcal{O}''}(\mathbf{x}_e) \in \mathbb{R}^{Nm \times n} \Rightarrow \text{rank} \{ \mathbf{J}_{\mathcal{O}''}(\mathbf{x}_e) \} \leq n. \quad (7.33)$$

and

$$n = \text{rank} \{ \mathbf{J}_{\mathcal{O}}(\mathbf{x}_e) \} \leq \text{rank} \{ \mathbf{J}_{\mathcal{O}''}(\mathbf{x}_e) \} \leq n. \quad (7.34)$$

Hence equality must hold on both sides of (7.34) which completes the proof. $\square$

The above Theorem used the hypothesis that $\mathbf{X}_e$ is compact in order to obtain a finite collection of functions from which the $\gamma_k(\cdot)$ could be selected. The following Example illustrates a natural redundancy in stereo imaging systems, and also demonstrates that $\mathbf{X}_e$ compact is not a necessary condition for the existence of a finite collection of such functions $\gamma_k(\cdot)$.

Example 7.1: Consider the problem of estimating the position, velocity, and perhaps higher time derivatives, of a single feature point from feature position measurements in binocular stereo image sequences. Using the notation and geometry established in Section 2.3, assume that the earth-fixed frame, F_E , is coincident with one of the camera frames, F_{C_1} , that the second camera frame, F_{C_2} , is related to the first by a simple translation along the y_E axis by b units (the baseline), and that the translational motion of the feature point satisfies a Gauss-Markov acceleration model as described in Section 2.4.3. With $\mathbf{p}(k)$ denoting the position of the feature point with respect to F_E at time k , the extended state vector

$$\mathbf{x}_e(k) := [\mathbf{p}(k), \dot{\mathbf{p}}(k), \ddot{\mathbf{p}}(k)]^T, \quad (7.35)$$

propagates in time, neglecting disturbances, according to linear dynamics

$$\mathbf{x}_e(k+1) = \Phi \mathbf{x}_e(k), \quad (7.36)$$

where Φ is the state transition matrix defined in (2.33).

The measurement model from Section 2.4.2, assuming unity focal length, depends only on the position, $\mathbf{p} = [p_1, p_2, p_3]^T$, of the feature point according to

$$\begin{aligned} \mathbf{z}(k) = \mathbf{h}_e(\mathbf{x}_e(k)) &= \mathbf{h}(\mathbf{x}(k)) \\ &= \frac{1}{p_1} [p_2 + b, p_3, p_2, p_3]^T, \end{aligned} \quad (7.37)$$

where $\mathbf{x} := [\mathbf{I}_3, \mathbf{0}_{3 \times 6}] \mathbf{x}_e = \mathbf{p}$ is the measured state. Note that, generally, $\text{rank}\{\mathbf{J}_{\mathbf{h}_e}(\mathbf{x}_e)\} = \text{rank}\{\mathbf{J}_{\mathbf{h}}(\mathbf{x})\}$. The noncompact region of interest could be any open set in $\mathbb{R}^9$ which lies in front of the cameras, $p_1 > 0$, in which case

$$\mathbf{J}_{\mathbf{h}}(\mathbf{p}) = \frac{1}{p_1^2} \begin{bmatrix} -(p_2 + b) & p_1 & 0 \\ -p_3 & 0 & p_1 \\ -p_2 & p_1 & 0 \\ -p_3 & 0 & p_1 \end{bmatrix} \quad (7.38)$$

satisfies $\text{rank}\{\mathbf{J}_{\mathbf{h}}(\mathbf{p})\} = 3 < m = 4$ in the region of interest and hence the system has redundant measurements. For this example, it is easily seen from (7.38) that the reduced set of derived measurements

$$\mathbf{z}'(k) := \gamma_k(\mathbf{z}(k)), \quad (7.39)$$

where

$$\gamma_k(\mathbf{z}(k)) \equiv [\mathbf{I}_3, \mathbf{0}_{3 \times 1}] \mathbf{z}(k), \quad (7.40)$$

which is obtained by ignoring the vertical position of the feature point in the second camera, still retains uniform observability. $\square$

7.3 Non-redundant Nonlinear Measurement Models

In this Section, the true dynamics are given by

$$\mathbf{x}_e(k+1) = \mathbf{f}(k, \mathbf{x}_e(k)), \quad (7.41)$$

where $\mathbf{x}_e \in \mathbf{X}_e \subset \mathbb{R}^n$ is the extended state vector and $\mathbf{f} : \mathbb{N}_0 \times \mathbf{X}_e \mapsto \mathbf{X}_e$. The measurement model is written as

$$\mathbf{z}(k) = \mathbf{h}_e(\mathbf{x}_e(k)) = \mathbf{h}(\mathbf{C}\mathbf{x}_e(k)) = \mathbf{h}(\mathbf{x}(k)), \quad (7.42)$$

where $\mathbf{h}_e : \mathbf{X}_e \mapsto \mathbb{R}^m$ is C^d , $d > 1$, $\mathbf{h} : \mathbf{C}\mathbf{X}_e \subseteq \mathbb{R}^q \mapsto \mathbb{R}^m$, $\mathbf{x}(k) := \mathbf{C}\mathbf{x}_e(k)$, and $\mathbf{C} \in \mathbb{R}^{q \times n}$ has the form

$$\mathbf{C} := [\mathbf{I}_q, \mathbf{0}_{q \times (n-q)}]. \quad (7.43)$$

Note that measurement errors are not included in the formal analysis at this time.

The Jacobian matrices $\mathbf{J}_{\mathbf{h}_e}(\mathbf{x}_e)$ and $\mathbf{J}_{\mathbf{h}}(\mathbf{x})$ satisfy

$$\mathbf{J}_{\mathbf{h}_e}(\mathbf{x}_e) = \mathbf{J}_{\mathbf{h}}(\mathbf{x})\mathbf{C}. \quad (7.44)$$

For non-redundant measurements, $\text{rank}\{\mathbf{J}_{\mathbf{h}_e}\} \equiv m \leq q < n$, and hence

$$\text{rank}\{\mathbf{J}_{\mathbf{h}_e}(\mathbf{x}_e)\} = \text{rank}\{\mathbf{J}_{\mathbf{h}}(\mathbf{x})\} = m. \quad (7.45)$$

In the case $m < q$, even if $\mathbf{J}_{\mathbf{h}}(\mathbf{x})$ has rank m uniformly, the Implicit Function Theorem (see the discussion by Ortega and Rheinboldt ([159, p. 129])) shows that there are infinitely many solutions, $\hat{\mathbf{x}}$, to the system of equations

$$\mathbf{s}(\mathbf{x}, \hat{\mathbf{x}}) := \mathbf{h}(\mathbf{x}) - \mathbf{h}(\hat{\mathbf{x}}) = \mathbf{0}. \quad (7.46)$$

Solutions of system (7.46) correspond to *zero dynamics* which must be checked for stability to ensure that, under (7.46), $\lim_{k \rightarrow \infty} \{\mathbf{x}(k) - \hat{\mathbf{x}}(k)\} = \mathbf{0}$. It may be possible, in the case

of $m < q$, to augment the measurement model with $q - m$ additional equations as constraints, and then continue the following design process on the augmented measurement model. For example, when using quaternions, $\mathbf{q}$, to model rotational motion, as was done in Chapter 2, the measurement model could be augmented with the constraint $\|\mathbf{q}\|_2^2 = 1$. For the remainder of this Section, it is assumed that $m = q$.

7.3.1 Assumptions and Preliminary Results

Proposed developments of sliding mode observers for nonlinear measurement models proceed very much like those of Chapter 6 through the use of the Fundamental Theorem of Integral Calculus (Theorem 5.9, page 148). This Section establishes some notation and presents preliminary results towards resolving a primary assumption required in the remaining theoretical developments of this Chapter.

For notational convenience, the following definitions are introduced. Recall from Definition 7.2, that for $\mathbf{G} : \mathcal{D} \subseteq \mathbb{R}^n \mapsto \mathbb{R}^m$, for some open convex set $\mathcal{D}$, being C^p , $p \geq 1$, then for any $\mathbf{y} \in \mathcal{D}$, $\mathbf{J}_{\mathbf{G}}(\mathbf{y})$ denotes the Jacobian matrix of $\mathbf{G}$ evaluated at $\mathbf{y}$,

$$\mathbf{J}_{\mathbf{G}}(\mathbf{y}) := \left[\frac{\partial \mathbf{G}(\mathbf{x})}{\partial \mathbf{x}} \right]_{\mathbf{x}=\mathbf{y}}. \quad (7.47)$$

For $n = m$ and $\mathbf{J}_{\mathbf{G}}(\mathbf{y})$ nonsingular, the inverse of this matrix is denoted $\mathbf{J}_{\mathbf{G}}^{-1}(\mathbf{y})$.

Definition 7.3: Let $\mathbf{G} : \mathcal{D} \subseteq \mathbb{R}^n \mapsto \mathbb{R}^m$, for some open convex set $\mathcal{D}$, be C^p , $p \geq 1$. For any $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathcal{D}$, define

$$\mathcal{I}_{\mathbf{G}}[\mathbf{x}_1, \mathbf{x}_2] := \int_0^1 \mathbf{J}_{\mathbf{G}}[\boldsymbol{\omega}_{\lambda}(\mathbf{x}_1, \mathbf{x}_2)] d\lambda, \quad (7.48)$$

where

$$\boldsymbol{\omega}_{\lambda}(\mathbf{x}_1, \mathbf{x}_2) := \lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2. \quad (7.49)$$

□

With Definitions 7.2 and 7.3, the Fundamental Theorem of Integral Calculus states that

$$\mathbf{G}(\mathbf{x}_1) - \mathbf{G}(\mathbf{x}_2) = \mathcal{I}_{\mathbf{G}}[\mathbf{x}_1, \mathbf{x}_2] [\mathbf{x}_1 - \mathbf{x}_2]. \quad (7.50)$$

The following technical Lemma presents results which are employed in subsequent Sections.

Lemma 7.1: Let $G : \mathcal{D} \subseteq \mathbb{R}^n \mapsto \mathbb{R}^m$, for some open convex set $\mathcal{D}$, be C^p , $p \geq 1$. Then $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$,

1. $\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2] = \mathcal{I}_G[\mathbf{x}_2, \mathbf{x}_1]$, and

2. For any sequence of points $\{\mathbf{y}_i \in \mathcal{D} \mid i = 0, 1, \dots, N\}$ with $\mathbf{y}_0 := \mathbf{x}_2$ and $\mathbf{y}_N := \mathbf{x}_1$,

$$\begin{aligned} \mathbf{G}(\mathbf{x}_1) - \mathbf{G}(\mathbf{x}_2) &= \mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2] \left[\sum_{i=0}^{N-1} (\mathbf{y}_{i+1} - \mathbf{y}_i) \right] \\ &= \sum_{i=0}^{N-1} \mathcal{I}_G[\mathbf{y}_{i+1}, \mathbf{y}_i] [\mathbf{y}_{i+1} - \mathbf{y}_i]. \end{aligned} \quad (7.51)$$

Proof: Item 1 follows directly with the substitution $\lambda = 1 - \mu$ in the integral of Definition 7.3. Item 2 follows from the use of telescopic summations in (7.50),

$$\mathbf{x}_1 - \mathbf{x}_2 = \sum_{i=0}^{N-1} (\mathbf{y}_{i+1} - \mathbf{y}_i), \text{ and} \quad (7.52)$$

$$\mathbf{G}(\mathbf{x}_1) - \mathbf{G}(\mathbf{x}_2) = \sum_{i=0}^{N-1} [\mathbf{G}(\mathbf{y}_{i+1}) - \mathbf{G}(\mathbf{y}_i)]. \quad (7.53)$$

□

A primary assumption employed in theoretical developments of Section 7.3 is that $\mathcal{I}_h[\mathbf{x}_1, \mathbf{x}_2]$, where $\mathbf{h} : \mathcal{A}_0 \subseteq \mathbb{R}^m \mapsto \mathcal{B}_0 \subseteq \mathbb{R}^m$ is the measured state to measurements map from (7.2), is nonsingular for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{A}_0$, and $\mathcal{A}_0$ is a suitably large open convex set. More specifically, this assumption is denoted

A1: $\mathbf{h} : \mathcal{A}_0 \subseteq \mathbb{R}^m \mapsto \mathcal{B}_0 \subseteq \mathbb{R}^m$, is C^p , $p \geq 1$, is 1-to-1 from $\mathcal{A}_0$ onto $\mathcal{B}_0$, and

1. $\forall \mathbf{x} \in \mathcal{A}_0$, $\mathbf{J}_h(\mathbf{x})$ is nonsingular (local homeomorphic property),
2. $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{A} \subseteq \mathcal{A}_0$, $\mathcal{I}_h(\mathbf{x}_1, \mathbf{x}_2)$ is nonsingular and

$$\|\mathcal{I}_h^{-1}(\mathbf{x}_1, \mathbf{x}_2)\| \leq L_{h^{-1}} < \infty. \quad (7.54)$$

In Chapter 3 the notation L_u was used to denote the Lie derivative of a vector-valued function $\mathbf{u}$. Throughout the remainder of the present chapter, the notation L_u is used to denote a Lipschitz constant associated with a vector-valued function $\mathbf{u}$. If $\mathbf{h}$ has an inverse

and if Assumption A1 above holds, then $L_{h^{-1}}$ is a Lipschitz constant for h^{-1} . That is, with $y_1 := h(x_1)$ and $y_2 := h(x_2)$, then

$$\mathcal{I}_h[x_1, x_2](x_1 - x_2) = y_1 - y_2 \quad (7.55)$$

$$\Rightarrow x_1 - x_2 = \mathcal{I}_h^{-1}[x_1, x_2](y_1 - y_2)$$

$$\Rightarrow \|x_1 - x_2\| \leq \|\mathcal{I}_h^{-1}[x_1, x_2]\| \|y_1 - y_2\|$$

$$\Rightarrow \|h^{-1}(y_1) - h^{-1}(y_2)\| \leq L_{h^{-1}} \|y_1 - y_2\|, \quad (7.56)$$

and hence $h^{-1}(\cdot)$ is Lipschitz continuous with a Lipschitz constant $L_{h^{-1}}$. So if Assumption A1 holds, then $h^{-1}(\cdot)$ is Lipschitz continuous over the set $h[\mathcal{A}]$. The converse statement, however, remains open at this time.

Applications of Assumption A1 in theoretical results which follow can in fact be substituted, as will be discussed in the next subsection, with Assumption A1':

A1': $h : \mathcal{A}_0 \subseteq \mathbb{R}^m \mapsto \mathcal{B}_0 \subseteq \mathbb{R}^m$, is C^p , $p \geq 1$, is 1-to-1 from $\mathcal{A}_0$ onto $\mathcal{B}_0$, and

1. $\forall x \in \mathcal{A}_0$, $J_h(x)$ is nonsingular (local homeomorphic property).
2. $h^{-1} : \mathcal{B}_0 \mapsto \mathcal{A}_0$ is Lipschitz continuous, i.e., $\forall y_1, y_2 \in \mathcal{B}_0$,

$$\|h^{-1}(y_1) - h^{-1}(y_2)\| \leq L_{h^{-1}} \|y_1 - y_2\|. \quad (7.57)$$

It is emphasized that the claim being made is **not** that Assumptions A1 and A1' are equivalent, since this is not yet clear, but rather that in the specific applications of Assumption A1 which follow, A1 can be replaced with A1' without altering the results.

Assumption A1 is employed in theoretical analysis of the ideal observer forms; proposed methods for implementation require only Item 1 of A1, namely the local homeomorphic property of $h(\cdot)$. A primary issue which is yet to be resolved concerns the extent of the set $\mathcal{A}$ in which the matrix inverse $\mathcal{I}_h^{-1}$ must be uniformly bounded in norm. If $h(\cdot)$ is one-to-one, then

$$h(x_1) - h(x_2) = \mathcal{I}_h[x_1, x_2][x_1 - x_2] = 0 \quad (7.58)$$

if, and only if, $x_1 = x_2$. This case leads to the conclusion, when h is 1-to-1, that $[x_1 - x_2]$ cannot lie in the null space of $\mathcal{I}_h[x_1, x_2]$, but does not appear to necessarily imply the non-

singularity of $\mathcal{I}_h[\mathbf{x}_1, \mathbf{x}_2]$. At this time, the following results are offered¹ which provide sufficient conditions for nonsingularity of $\mathcal{I}_h[\mathbf{x}_1, \mathbf{x}_2]$ and lead up to a region of nonsingularity. The first method results from application of the Mean Value Theorem for Integrals (see [144, p 174]).

Theorem 7.3: Let $\mathbf{G} : \mathcal{D} \subseteq \mathbb{R}^m \mapsto \mathbb{R}^m$, for some open convex set $\mathcal{D}$, be C^p , $p \geq 1$. Denote the components of $\mathbf{G}(\cdot)$ as $G_i(\cdot)$, $i = 1, 2, \dots, m$. If for every set of m vectors $\{\mathbf{y}_i \in \mathcal{D} \mid i = 1, 2, \dots, m\}$ the matrix

$$\begin{bmatrix} \frac{\partial G_1}{\partial \mathbf{x}}(\mathbf{y}_1) \\ \frac{\partial G_2}{\partial \mathbf{x}}(\mathbf{y}_2) \\ \vdots \\ \frac{\partial G_m}{\partial \mathbf{x}}(\mathbf{y}_m) \end{bmatrix} \quad (7.59)$$

is nonsingular, then $\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2]$ is nonsingular $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$.

Proof: Since $\mathbf{G}(\cdot)$ is continuously differentiable, the Mean Value Theorem for Integrals states that there exists a set of m scalars, $\{\lambda_i \mid i = 1, 2, \dots, m, \lambda_i \in [0, 1]\}$, such that, $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$,

$$\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2] = \begin{bmatrix} \frac{\partial G_1}{\partial \mathbf{x}}(\mathbf{w}_1) \\ \frac{\partial G_2}{\partial \mathbf{x}}(\mathbf{w}_2) \\ \vdots \\ \frac{\partial G_m}{\partial \mathbf{x}}(\mathbf{w}_n) \end{bmatrix}. \quad (7.60)$$

where the $\mathbf{w}_i \in \mathcal{D}$ are defined as $\mathbf{w}_i := \lambda_i \mathbf{x}_1 + (1 - \lambda_i) \mathbf{x}_2$. Setting $\mathbf{y}_i := \mathbf{w}_i$ completes the proof. $\square$

The conditions of Theorem 7.3, however, may be very difficult to verify in practice. More useful results are based on the following Lemma.

Lemma 7.2: Let $\mathbf{A} \in \mathbb{R}^{m \times m}$ and suppose there exists a matrix norm $\|\cdot\|$ such that $\|\mathbf{A} - \mathbf{I}_m\| < \alpha < 1$. Then $\mathbf{A}$ is nonsingular and

$$\|\mathbf{A}^{-1}\| \leq \frac{1}{1 - \alpha}. \quad (7.61)$$

Proof: See, for example, Blum [162, Theorem 3.5.1 and Corollary 2]. $\square$

¹Helpful suggestions from Dr. R. Israel, Dept. of Mathematics, University of British Columbia, and Dr. E. Yee, Defence Research Establishment Suffield, Alberta, which led to the proof of Theorems 7.5 are gratefully acknowledged.

The following Theorem shows that, if $\mathbf{J}_G(\mathbf{x})$ is nonsingular at each $\mathbf{x} \in \mathcal{D}$, then $\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2]$ is nonsingular provided $\mathbf{x}_1$ is "close enough" to $\mathbf{x}_2$.

Theorem 7.4: Let $\mathbf{G} : \mathcal{D} \subseteq \mathbb{R}^m \mapsto \mathbb{R}^m$, for some open convex set $\mathcal{D}$, be C^p , $p \geq 1$. If $\mathbf{J}_G(\mathbf{x})$ is nonsingular at each $\mathbf{x} \in \mathcal{D}$, then $\forall \mathbf{x}_2 \in \mathcal{D}$, $\exists \epsilon > 0$ such that $\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2]$ is nonsingular for all $\mathbf{x}_1 \in \mathcal{D}$ satisfying $\|\mathbf{x}_1 - \mathbf{x}_2\|_2 < \epsilon$.

Proof: Fix $\mathbf{x}_2 \in \mathcal{D}$. Define the continuous mapping $d : \mathcal{D} \mapsto \mathbb{R}$ as

$$d(\mathbf{x}) = \sup_{\lambda \in [0,1]} \left\| [\mathbf{J}_G(\mathbf{w}_\lambda(\mathbf{x}))] \mathbf{J}_G^{-1}(\mathbf{x}_2) - \mathbf{I}_m \right\|_2, \quad (7.62)$$

where

$$\mathbf{w}_\lambda(\mathbf{x}) := \lambda \mathbf{x} + (1 - \lambda) \mathbf{x}_2. \quad (7.63)$$

Continuity of $d(\mathbf{x})$ and the fact that $d(\mathbf{x}_2) = 0$ implies that there exists an open ball, $B(\mathbf{x}_2, \epsilon)$ for some $\epsilon > 0$ such that $\forall \mathbf{x} \in B(\mathbf{x}_2, \epsilon) \cap \mathcal{D}$, $d(\mathbf{x}) < 1$. But then, for any $\mathbf{x}_1 \in B(\mathbf{x}_2, \epsilon) \cap \mathcal{D}$,

$$\begin{aligned} \left\| \mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2] \mathbf{J}_G^{-1}(\mathbf{x}_2) - \mathbf{I}_m \right\|_2 &= \left\| \int_0^1 \left\{ \mathbf{J}_G(\mathbf{w}_\lambda(\mathbf{x}_1)) \mathbf{J}_G^{-1}(\mathbf{x}_2) - \mathbf{I}_m \right\} d\lambda \right\|_2 \\ &\leq \int_0^1 \left\| \mathbf{J}_G(\mathbf{w}_\lambda(\mathbf{x}_1)) \mathbf{J}_G^{-1}(\mathbf{x}_2) - \mathbf{I}_m \right\|_2 d\lambda \\ &\leq d(\mathbf{x}_1) < 1. \end{aligned} \quad (7.64)$$

Hence, by Lemma 7.2, the product $\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2] \mathbf{J}_G^{-1}(\mathbf{x}_2)$ is nonsingular which, in turn, implies the nonsingularity of $\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2]$. $\square$

Theorem 7.4, however, does not provide any information concerning the size of the region, i.e. the magnitude of ϵ , in which $\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2]$ is nonsingular. Another method, and perhaps the most useful, is provided by the following Theorem which relies on more information concerning the function $\mathbf{G}(\cdot)$, but gives a region of nonsingularity of $\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2]$ as well as bounds on the norm of inverse.

Theorem 7.5: Let $\mathbf{G} : \mathcal{D} \subseteq \mathbb{R}^m \mapsto \mathbb{R}^m$, for some open convex set $\mathcal{D}$, be C^p , $p \geq 1$. Let $\|\cdot\|$ and $\|\cdot\|$ be any compatible vector and matrix norms, respectively. Suppose that

1. $\forall \mathbf{x}_2 \in \mathcal{D}$, $\mathbf{J}_G(\mathbf{x}_2)$ is nonsingular and

$$\left\| \mathbf{J}_G^{-1}(\mathbf{x}_2) \right\| < a, \quad (7.65)$$

2. $\mathbf{J}_G(\mathbf{x}_2)$ is Hölder continuous on $\mathcal{D}$, i.e. $\exists c, r$ such that $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$,

$$\|\mathbf{J}_G(\mathbf{x}_1) - \mathbf{J}_G(\mathbf{x}_2)\| \leq c \|\mathbf{x}_1 - \mathbf{x}_2\|^r. \quad (7.66)$$

Then for any $\mathbf{x}_2 \in \mathcal{D}$, $\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2]$ is nonsingular for all $\mathbf{x}_1 \in \mathcal{D}$ satisfying $\|\mathbf{x}_1 - \mathbf{x}_2\| < \delta$, where

$$\delta < \left[\frac{r+1}{ac} \right]^{1/r}. \quad (7.67)$$

Moreover,

$$\|\mathcal{I}_G^{-1}[\mathbf{x}_1, \mathbf{x}_2]\| \leq \frac{a}{1 - \frac{ac\delta^r}{r+1}}. \quad (7.68)$$

Proof: Fix $\mathbf{x}_2 \in \mathcal{D}$. Then for any $\mathbf{x}_1 \in B(\mathbf{x}_2, \delta) \cap \mathcal{D}$,

$$\begin{aligned} \|\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2] \mathbf{J}_G^{-1}(\mathbf{x}_2) - \mathbf{I}_m\| &= \|[\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2] - \mathbf{J}_G(\mathbf{x}_2)] \mathbf{J}_G^{-1}(\mathbf{x}_2)\| \\ &\leq a \int_0^1 \|\mathbf{J}_G(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) - \mathbf{J}_G(\mathbf{x}_2)\| d\lambda \\ &\leq ac \int_0^1 \|\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 - \mathbf{x}_2\|^r d\lambda \\ &= ac \int_0^1 \lambda^r d\lambda \|\mathbf{x}_1 - \mathbf{x}_2\|^r \\ &\leq \frac{ac\delta^r}{1+r} < 1. \end{aligned} \quad (7.69)$$

Hence, by Lemma 7.2, the product $\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2] \mathbf{J}_G^{-1}(\mathbf{x}_2)$ is nonsingular which, in turn, implies the nonsingularity of $\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2]$. Also from Lemma 7.2,

$$\|[\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2] \mathbf{J}_G^{-1}(\mathbf{x}_2)]^{-1}\| \leq \frac{1}{1 - \frac{ac\delta^r}{r+1}}. \quad (7.70)$$

and the bounds in (7.68) are obtained with

$$\begin{aligned} \|\mathcal{I}_G^{-1}[\mathbf{x}_1, \mathbf{x}_2]\| &= \|\mathbf{J}_G^{-1}(\mathbf{x}_2) [\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2] \mathbf{J}_G^{-1}(\mathbf{x}_2)]^{-1}\| \\ &\leq \frac{a}{1 - \frac{ac\delta^r}{r+1}}. \end{aligned} \quad (7.71)$$

□

The use of $\mathcal{I}_h$ in initial developments, rather than $\mathbf{J}_h$ and approximate linearization, allows an exact development of ideal forms with equality throughout. Moreover, this approach provides intuitive insight into appropriate linearization methods, implementation

with local iterations, and analysis of this implementation. In the case in which $\mathcal{I}_h[\mathbf{x}_1, \mathbf{x}_2]$ is singular for some points $(\mathbf{x}_1, \mathbf{x}_2)$ it may be possible to augment the analysis by using the pseudoinverse or perhaps by writing

$$\tilde{\mathcal{I}}_h[\mathbf{x}_1, \mathbf{x}_2] := \mathcal{I}_h[\mathbf{x}_1, \mathbf{x}_2] + \mathbf{E}[\mathbf{x}_1, \mathbf{x}_2] \quad (7.72)$$

where $\mathbf{E} : \mathbb{R}^m \times \mathbb{R}^m \mapsto \mathbb{R}^{m \times m}$ has arbitrarily small matrix norm and $\tilde{\mathcal{I}}_h[\mathbf{x}_1, \mathbf{x}_2]$ is nonsingular. This is always possible because the set of nonsingular matrices is dense in $\mathbb{R}^{m \times m}$. These alternatives are not investigated in this Thesis, but should be considered in future work in order to fully assess the limitations imposed by this assumption. The implications of Assumption A1 and replacement of A1 with A1' are discussed further in the following subsections.

7.3.2 Theoretical Form, Stability, and Properties

The observers proposed in this Section have the form

$$\hat{\mathbf{x}}_e(k+1) = \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \mathbf{L}(k)\mathbf{v}(k). \quad (7.73)$$

where $\hat{\mathbf{f}}$ is an estimate of $\mathbf{f}$, and the observer gain matrices, $\mathbf{L}(k)$, and the function $\mathbf{v}(k)$ are to be determined. The sliding surface is defined, with $\mathbf{x} = \mathbf{C}\mathbf{x}_e$ and $\hat{\mathbf{x}} = \mathbf{C}\hat{\mathbf{x}}_e$, as

$$\begin{aligned} \mathbf{s}_k := \mathbf{s}(\mathbf{x}(k), \hat{\mathbf{x}}(k)) &:= \mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}(k)) \\ &= \mathbf{h}(\mathbf{x}(k)) - \mathbf{h}(\hat{\mathbf{x}}(k)) \\ &= \mathcal{I}_h[\mathbf{x}(k), \hat{\mathbf{x}}(k)]\mathbf{e}(k), \end{aligned} \quad (7.74)$$

where

$$\begin{aligned} \mathbf{e}(k) &:= \mathbf{x}(k) - \hat{\mathbf{x}}(k) \\ &= \mathbf{C}[\mathbf{x}_e(k) - \hat{\mathbf{x}}_e(k)] \\ &:= \mathbf{C}\mathbf{e}_e(k). \end{aligned} \quad (7.75)$$

With the assumption that $\mathcal{I}_h[\mathbf{x}_1, \mathbf{x}_2]$ is nonsingular, and with the special form of $\mathbf{C}$, Theorem 6.1 guarantees that gain matrices $\mathbf{L}(k)$ exist such that the projection operator

$$\begin{aligned} \mathbf{F}(k) &:= [\mathbf{I}_n - \mathbf{L}(k)(\mathcal{I}_h[\mathbf{x}(k), \hat{\mathbf{x}}(k)]\mathbf{C}\mathbf{L}(k))^{-1}\mathcal{I}_h[\mathbf{x}(k), \hat{\mathbf{x}}(k)]\mathbf{C}] \\ &= [\mathbf{I}_n - \mathbf{L}(k)(\mathbf{C}\mathbf{L}(k))^{-1}\mathbf{C}] \end{aligned} \quad (7.76)$$

is well defined, and represents a projection along the column space of $\mathbf{L}(k)$ into the null space of $\mathbf{C}$. This result is the primary link with methods of Chapter 6, even though direct applications of those methods leads to a nonintuitive observer form.

Consider the following two proposed observer forms:

$$O_i: \quad \hat{\mathbf{x}}_e^{(i)}(k+1) = \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(i)}(k)) + \mathbf{L}^{(i)}(k) \mathbf{v}^{(i)}(k), \quad i = 1, 2, \quad (7.77)$$

where

$$\mathbf{v}^{(1)}(k) := \left[\mathcal{I}_h[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)] \mathbf{C} \mathbf{L}^{(1)}(k) \right]^{-1} \quad (7.78)$$

$$\cdot \left\{ \mathcal{I}_h[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)] [\mathbf{x}(k+1) - \hat{\mathbf{x}}_p^{(1)}(k+1)] - \mathbf{g}(k, \mathbf{s}_k^{(1)}) \right\} :$$

$$\mathbf{v}^{(2)}(k) := \left[\mathcal{I}_h[\hat{\mathbf{x}}_p^{(2)}(k+1), \hat{\mathbf{x}}^{(2)}(k+1)] \mathbf{C} \mathbf{L}^{(2)}(k) \right]^{-1} \quad (7.79)$$

$$\cdot \left\{ \mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}_p^{(2)}(k+1)) - \mathbf{g}(k, \mathbf{s}_k^{(2)}) \right\} :$$

$$\hat{\mathbf{x}}_p^{(i)}(k+1) := \mathbf{C} \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(i)}(k)), \quad i = 1, 2; \quad (7.80)$$

$$\mathbf{s}_k^{(i)} := \mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}^{(i)}(k)), \quad i = 1, 2; \quad (7.81)$$

$$\mathbf{g}(k, \mathbf{s}_k^{(i)}) := \mathbf{\Gamma}_k \mathbf{Sat}(\mathbf{s}_k^{(i)}, \epsilon) + \mathbf{\Lambda}_k \mathbf{s}_k^{(i)}, \quad i = 1, 2, \quad \mathbf{\Gamma}_k, \mathbf{\Lambda}_k \in \mathbb{R}^{m \times m}. \quad (7.82)$$

These two observers represent ideal forms in the sense that, as will be shown, under certain conditions both possess a global convergent quasi-sliding regime, respectively, on $\mathbf{s}_k^{(i)} = \mathbf{0}$, $i = 1, 2$. However, the observer outputs, $\hat{\mathbf{x}}_e^{(i)}(k+1)$, are represented by implicit systems of nonlinear integral difference equations.

Observer O_1 results from the use of the Fundamental Theorem of Integral Calculus and direct application of the methods proposed in Chapter 6 for linear measurement models. In O_1 , the measurements $\mathbf{z}(k+1)$ do not explicitly appear in the observer equations, and the true measured state, $\mathbf{x}(k+1)$, which is not directly available, appears three times in the expression for $\mathbf{v}^{(1)}(k)$. This form neither lends itself easily to an intuitive methodology for implementation, nor to theoretical analysis of possible approximations. This form, however, is well suited for stability analysis under the hypothesis of Theorem 5.11.

Observer O_2 again results from the use of the Fundamental Theorem of Integral Calculus, but is developed and analyzed with a slightly alternate approach than that followed in Chapter 6. In O_2 , the vector $\mathbf{v}^{(2)}(k)$ depends explicitly on the measurements and does not contain the true measured state $\mathbf{x}(k+1)$ (except via the measurements). The vector

$\mathbf{v}^{(2)}(k)$ also depends on the predicted measured state, $\hat{\mathbf{x}}_p^{(2)}(k+1)$, which can be computed as in (7.80) just following the k th measurement update. The implicit definition of $\mathbf{v}^{(2)}(k)$ is demonstrated by noting that

$$\hat{\mathbf{x}}_p^{(2)}(k+1) - \hat{\mathbf{x}}^{(2)}(k+1) = -\mathbf{CL}^{(2)}(k)\mathbf{v}^{(2)}(k), \quad (7.83)$$

and writing

$$\begin{aligned} \mathcal{I}_h[\hat{\mathbf{x}}_p^{(2)}(k+1), \hat{\mathbf{x}}^{(2)}(k+1)] &:= \int_0^1 \mathbf{J}_h \left(\lambda \hat{\mathbf{x}}_p^{(2)}(k+1) + (1-\lambda) \hat{\mathbf{x}}^{(2)}(k+1) \right) d\lambda \\ &= \int_0^1 \mathbf{J}_h \left(\hat{\mathbf{x}}^{(2)}(k+1) + \lambda \left[\hat{\mathbf{x}}_p^{(2)}(k+1) - \hat{\mathbf{x}}^{(2)}(k+1) \right] \right) d\lambda \\ &= \int_0^1 \mathbf{J}_h \left(\mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(2)}(k+1)) + (1-\lambda) \left[\mathbf{CL}^{(2)}(k)\mathbf{v}^{(2)}(k) \right] \right) d\lambda. \end{aligned} \quad (7.84)$$

The following Theorem details the fundamental properties of these observers including the fact that, under mild assumptions, these two observers are equivalent.

Theorem 7.6: For the system (7.41) with measurement model (7.42) consider the observers O_1 in (7.77)-(7.78), and O_2 in (7.77)-(7.79). Assume that

1. $\hat{\mathbf{x}}_e^{(1)}(0) = \hat{\mathbf{x}}_e^{(2)}(0)$,
2. $\mathbf{L}(k) := \mathbf{L}^{(1)}(k) \equiv \mathbf{L}^{(2)}(k)$,
3. $\hat{\mathbf{f}}(k, \cdot)$ is 1-to-1,
4. Assumption A1 (page 224) holds or Assumption A1' (page 225) holds.

Define

$$\mathbf{F}(k) := \left[\mathbf{I}_n - \mathbf{L}(k) (\mathbf{CL}(k))^{-1} \mathbf{C} \right], \quad (7.85)$$

$$\mathbf{B}(k)\mathbf{d}(k, \mathbf{y}_e) := \mathbf{f}(k, \mathbf{y}_e) - \hat{\mathbf{f}}(k, \mathbf{y}_e), \quad \forall \mathbf{y}_e \in \mathbf{X}_e. \quad (7.86)$$

The following statements hold:

1. **Reaching Law:** If Λ_k , Γ_k and ϵ satisfy the hypothesis of Theorem 5.12, then global convergent quasi-sliding regimes exists on $\mathbf{s}_k^{(1)} = \mathbf{0}$ and $\mathbf{s}_k^{(2)} = \mathbf{0}$ independent of the dynamical modelling error, and pure sliding can be attained for all $k > k_s > 0$ by setting $\Lambda_k = \mathbf{0}$ and $\Gamma_k = \mathbf{0}$ for all $k \geq k_s$ (see Corollary 5.2);

2. *Equivalence*: $\forall k \geq 0, \hat{\mathbf{x}}_e^{(1)}(k) = \hat{\mathbf{x}}_e^{(2)}(k)$;
3. *Stability of Reaching and Sliding Dynamics*: With Statements 1 and 2 above, if $\mathbf{B}(k) \in \mathbb{R}^{n \times p}$, $\mathbf{d}(k, \mathbf{x}_e) \in \mathbb{R}^{p \times 1}$, $\mathbf{F}(k)$, and $\hat{\mathbf{f}}$ satisfy the hypothesis of Theorem 5.11, where the parameters ϵ_2 , $\xi(k)$ and $\eta(k)$ of Theorem 5.11 are taken as

$$\epsilon_2 = 0, \quad (7.87)$$

$$\xi(k) := -L_{\mathbf{h}^{-1}} \mathbf{g}(k, \mathbf{s}_k). \quad \text{and} \quad (7.88)$$

$$\eta(k) = 0. \quad (7.89)$$

then the error dynamics for O_1 and O_2 are locally exponentially stable;

4. *Robustness of Sliding Mode*: With Statements 1 and 3 above, if, in addition, $\exists \Xi(k) \in \mathbb{R}^{m \times p}$ such that $\mathbf{B}(k) = \mathbf{L}(k)\Xi(k)$, then the sliding dynamics of O_1 and O_2 are invariant to dynamical modelling errors/disturbances in which case the sliding dynamics are locally exponentially asymptotically stable.

Proof: For Statement 1, it is required to show that under the above definitions, $\mathbf{s}_{k+1}^{(i)} = \mathbf{g}(k, \mathbf{s}_k^{(i)})$, $i = 1, 2$, from which the result follows by Theorem 5.1. For O_1 ,

$$\begin{aligned} \mathbf{s}_{k+1}^{(1)} &:= \mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}^{(1)}(k+1)) \\ &= \mathcal{I}_{\mathbf{h}}[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)] [\mathbf{x}(k+1) - \hat{\mathbf{x}}^{(1)}(k+1)] \\ &= \mathcal{I}_{\mathbf{h}}[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)] [\mathbf{x}(k+1) - \mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(1)}(k))] \\ &\quad - \mathcal{I}_{\mathbf{h}}[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)] \mathbf{C}\mathbf{L}^{(1)}(k)\mathbf{v}^{(1)}(k) \\ &= \mathcal{I}_{\mathbf{h}}[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)] [\mathbf{x}(k+1) - \mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(1)}(k))] \\ &\quad - \mathcal{I}_{\mathbf{h}}[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)] [\mathbf{x}(k+1) - \mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(1)}(k))] \\ &\quad + \mathbf{g}(k, \mathbf{s}_k^{(1)}). \\ &= \mathbf{g}(k, \mathbf{s}_k^{(1)}). \end{aligned} \quad (7.90)$$

For O_2 ,

$$\begin{aligned} \mathbf{s}_{k+1}^{(2)} &:= \mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}^{(2)}(k+1)) \\ &= \mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}_p^{(2)}(k+1)) + \mathbf{h}(\hat{\mathbf{x}}_p^{(2)}(k+1)) - \mathbf{h}(\hat{\mathbf{x}}^{(2)}(k+1)) \end{aligned}$$

$$\begin{aligned}
&= \mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}_p^{(2)}(k+1)) \\
&\quad + \mathcal{I}_h[\hat{\mathbf{x}}_p^{(2)}(k+1), \hat{\mathbf{x}}^{(2)}(k+1)] [\hat{\mathbf{x}}_p^{(2)}(k+1) - \hat{\mathbf{x}}^{(2)}(k+1)] \\
&= \mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}_p^{(2)}(k+1)) \\
&\quad - \mathcal{I}_h[\hat{\mathbf{x}}_p^{(2)}(k+1), \hat{\mathbf{x}}^{(2)}(k+1)] \mathbf{CL}^{(2)}(k) \mathbf{v}^{(2)}(k) \\
&= \left\{ \mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}_p^{(2)}(k+1)) \right\} \\
&\quad - \left\{ \mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}_p^{(2)}(k+1)) \right\} + \mathbf{g}(k, \mathbf{s}_k^{(2)}) \\
&= \mathbf{g}(k, \mathbf{s}_k^{(2)}). \tag{7.91}
\end{aligned}$$

The proof of Statement 2 is by induction. Suppose, for some $k_0 \geq 0$, $\hat{\mathbf{x}}_e^{(1)}(k_0) = \hat{\mathbf{x}}_e^{(2)}(k_0)$. Then $\mathbf{s}_{k_0}^{(1)} = \mathbf{s}_{k_0}^{(2)}$, and from (7.90) and (7.91),

$$\mathbf{s}_{k_0+1}^{(1)} = \mathbf{g}(k_0, \mathbf{s}_{k_0}^{(1)}) = \mathbf{g}(k_0, \mathbf{s}_{k_0}^{(2)}) = \mathbf{s}_{k_0+1}^{(2)}. \tag{7.92}$$

Using (7.92) and the hypothesis that $\mathbf{h}(\cdot)$ is 1-to-1 gives

$$\begin{aligned}
\mathbf{0} &= \mathbf{s}_{k_0+1}^{(1)} - \mathbf{s}_{k_0+1}^{(2)} \\
&= \mathbf{h}(\hat{\mathbf{x}}^{(2)}(k_0+1)) - \mathbf{h}(\hat{\mathbf{x}}^{(1)}(k_0+1)) \\
\Rightarrow \hat{\mathbf{x}}^{(1)}(k_0+1) &= \hat{\mathbf{x}}^{(2)}(k_0+1). \tag{7.93}
\end{aligned}$$

Also, under the induction hypothesis, and assuming that $\hat{\mathbf{f}}(k, \cdot)$ is 1-to-1, the predicted measured states satisfy

$$\hat{\mathbf{x}}_p^{(1)}(k_0+1) := \mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(1)}(k_0)) = \mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(2)}(k_0)) := \hat{\mathbf{x}}_p^{(2)}(k_0+1). \tag{7.94}$$

With $\mathbf{L}(k_0) := \mathbf{L}^{(1)}(k_0) = \mathbf{L}^{(2)}(k_0)$ and using

$$\mathbf{CL}(k_0)\mathbf{v}^{(1)}(k_0) = \hat{\mathbf{x}}^{(1)}(k_0+1) - \hat{\mathbf{x}}_p^{(1)}(k_0+1), \tag{7.95}$$

consider

$$\begin{aligned}
&\mathcal{I}_h[\hat{\mathbf{x}}_p^{(1)}(k_0+1), \hat{\mathbf{x}}^{(1)}(k_0+1)] \mathbf{CL}(k_0)\mathbf{v}^{(1)}(k_0) \\
&= \mathcal{I}_h[\hat{\mathbf{x}}_p^{(1)}(k_0+1), \hat{\mathbf{x}}^{(1)}(k_0+1)] [\hat{\mathbf{x}}^{(1)}(k_0+1) - \hat{\mathbf{x}}_p^{(1)}(k_0+1)] \\
&= \mathbf{h}(\hat{\mathbf{x}}^{(1)}(k_0+1)) - \mathbf{h}(\hat{\mathbf{x}}_p^{(1)}(k_0+1)) \\
&= \mathbf{h}(\hat{\mathbf{x}}^{(1)}(k_0+1)) - \mathbf{h}(\mathbf{x}(k_0+1)) + \mathbf{h}(\mathbf{x}(k_0+1)) - \mathbf{h}(\hat{\mathbf{x}}_p^{(1)}(k_0+1)) \\
&= -\mathbf{s}_{k_0+1}^{(1)} + \mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}_p^{(1)}(k_0+1)). \tag{7.96}
\end{aligned}$$

Substituting the results of (7.92), (7.93), and (7.94) into (7.96) and rearranging gives

$$\begin{aligned} \mathbf{v}^{(1)}(k_0) &= \left[\mathcal{I}_h \left(\hat{\mathbf{x}}_p^{(2)}(k_0 + 1), \hat{\mathbf{x}}^{(2)}(k_0 + 1) \right) \mathbf{CL}(k_0) \right]^{-1} \\ &\quad \cdot \left\{ \mathbf{z}(k + 1) - \mathbf{h}(\hat{\mathbf{x}}_p^{(2)}(k_0 + 1)) - \mathbf{g}(k_0, \mathbf{s}_{k_0}^{(2)}) \right\} \\ &= \mathbf{v}^{(2)}(k). \end{aligned} \quad (7.97)$$

Finally, with the induction hypothesis and (7.97), and again using the assumption that $\hat{\mathbf{f}}(k, \cdot)$ is 1-to-1,

$$\begin{aligned} \hat{\mathbf{x}}_e^{(1)}(k_0 + 1) &:= \hat{\mathbf{f}}(k_0, \hat{\mathbf{x}}_e^{(1)}(k_0)) + \mathbf{L}(k_0) \mathbf{v}^{(1)}(k_0) \\ &= \hat{\mathbf{f}}(k_0, \hat{\mathbf{x}}_e^{(2)}(k_0)) + \mathbf{L}(k_0) \mathbf{v}^{(2)}(k_0) \\ &:= \hat{\mathbf{x}}_e^{(2)}(k_0 + 1). \end{aligned} \quad (7.98)$$

Hence, given $\hat{\mathbf{x}}_e^{(1)}(0) = \hat{\mathbf{x}}_e^{(2)}(0)$, it follows that $\forall k \geq 0$, $\hat{\mathbf{x}}_e^{(1)}(k) = \hat{\mathbf{x}}_e^{(2)}(k)$.

For Statement 3, and with the equivalence result of Statement 2, it is sufficient to show that observer O_1 can be placed in a Luenberger form, i.e. (5.115), page 157, satisfying the hypothesis of Theorem 5.11. But observer O_1 can be written as

$$\hat{\mathbf{x}}_e^{(1)}(k + 1) = \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(1)}(k)) + \mathbf{K}(k) \left[\hat{\mathbf{z}}(k + 1) - \mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(1)}(k)) + \boldsymbol{\gamma}(k) \right]. \quad (7.99)$$

where

$$\hat{\mathbf{z}}(k + 1) := \mathbf{C}\mathbf{x}_e(k + 1). \quad (7.100)$$

$$\mathbf{K}(k) := \mathbf{L}(k) [\mathbf{CL}(k)]^{-1}, \quad \text{and} \quad (7.101)$$

$$\boldsymbol{\gamma}(k) := -\mathcal{I}_h^{-1}[\mathbf{x}(k + 1), \hat{\mathbf{x}}^{(1)}(k + 1)]\mathbf{g}(k, \mathbf{s}_k^{(1)}). \quad (7.102)$$

This is the observer form considered in Theorem 5.11 with the linear "measurement model" in (7.100), and the result follows with hypothesis (7.88) provided it is shown that $\|\boldsymbol{\gamma}(k)\| \leq \|\boldsymbol{\xi}(k)\|$. But under Assumption A1,

$$\begin{aligned} \|\boldsymbol{\gamma}(k)\| &:= \|\mathcal{I}_h^{-1}[\mathbf{x}(k + 1), \hat{\mathbf{x}}^{(1)}(k + 1)]\mathbf{g}(k, \mathbf{s}_k^{(1)})\| \\ &\leq L_{\mathbf{h}^{-1}} \|\mathbf{g}(k, \mathbf{s}_k^{(1)})\| := \|\boldsymbol{\xi}(k)\|. \end{aligned} \quad (7.103)$$

or from Assumption A1' and the result of Statement 1,

$$\|\boldsymbol{\gamma}(k)\| := \|\mathcal{I}_h^{-1}[\mathbf{x}(k + 1), \hat{\mathbf{x}}^{(1)}(k + 1)]\mathbf{g}(k, \mathbf{s}_k^{(1)})\|$$

$$\begin{aligned}
&= \|\mathcal{I}_h^{-1}[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)]\mathbf{s}_{k+1}^{(1)}\| \\
&= \|\mathcal{I}_h^{-1}[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)]\{\mathbf{h}(\mathbf{x}(k+1)) - \mathbf{h}(\hat{\mathbf{x}}^{(1)}(k+1))\}\| \\
&= \|\mathbf{x}(k+1) - \hat{\mathbf{x}}^{(1)}(k+1)\| \\
&\leq L_{h^{-1}}\|\mathbf{h}(\mathbf{x}(k+1)) - \mathbf{h}(\hat{\mathbf{x}}^{(1)}(k+1))\| \\
&= L_{h^{-1}}\|\mathbf{s}_{k+1}^{(1)}\| \\
&= L_{h^{-1}}\|\mathbf{g}(k, \mathbf{s}_k^{(1)})\| := \|\boldsymbol{\xi}(k)\|.
\end{aligned} \tag{7.104}$$

Note that if $k_s = 0$, then neither Item 2 of Assumption A1 nor Item 2 of Assumption A1' is required for stability since in that case $\mathbf{g}(k, \mathbf{s}_k) \equiv \mathbf{0}$, but that Item 1 of these Assumptions is still required since in this case, $\forall k \geq 0$, $\mathcal{I}_h[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)] = \mathbf{J}_h(\hat{\mathbf{x}}^{(1)}(k+1))$.

For Statement 4, $\mathbf{g}(k, \mathbf{s}_k) \equiv \mathbf{0}$ for all $k \geq k_s$, and the 1-to-1 property of $\mathbf{h}(\cdot)$ gives $\mathbf{x}(k+1) = \hat{\mathbf{x}}^{(i)}(k+1)$, $i = 1, 2$. This in turn implies that $\forall k \geq k_s$,

$$\mathcal{I}_h[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)] = \mathbf{J}_h(\hat{\mathbf{x}}^{(1)}(k+1)), \tag{7.105}$$

and

$$\mathcal{I}_h[\hat{\mathbf{x}}_p^{(2)}(k+1), \hat{\mathbf{x}}^{(2)}(k+1)] = \mathcal{I}_h[\hat{\mathbf{x}}_p^{(2)}(k+1), \mathbf{x}(k+1)]. \tag{7.106}$$

With (7.105) and (7.106),

$$\mathbf{v}^{(i)}(k) = [\mathbf{CL}(k)]^{-1} \{\mathbf{x}(k+1) - \hat{\mathbf{x}}_p^{(i)}(k+1)\}, \tag{7.107}$$

The error dynamics during sliding can then be written, with $\mathbf{e}_e^{(i)}(k) := \mathbf{x}_e(k) - \hat{\mathbf{x}}_e^{(i)}(k)$ and using (7.99), as

$$\begin{aligned}
\mathbf{e}_e^{(i)}(k+1) &:= \mathbf{x}_e(k+1) - \hat{\mathbf{x}}_e^{(i)}(k+1) \\
&= \mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(i)}(k)) - \mathbf{L}(k)\mathbf{v}^{(i)}(k) \\
&= \mathbf{F}(k) \{\mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(i)}(k))\} \\
&= \mathbf{F}(k) \{\mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \mathbf{x}_e(k))\} + \mathbf{F}(k) \{\hat{\mathbf{f}}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(i)}(k))\} \\
&= \mathbf{F}(k)\mathbf{B}(k)\mathbf{d}(k, \mathbf{x}_e(k)) + \mathbf{F}(k) \{\hat{\mathbf{f}}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(i)}(k))\} \\
&= \mathbf{F}(k)\mathbf{L}(k)\boldsymbol{\Xi}(k)\mathbf{d}(k, \mathbf{x}_e(k)) + \mathbf{F}(k) \{\hat{\mathbf{f}}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(i)}(k))\} \\
&= \mathbf{F}(k) \{\hat{\mathbf{f}}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(i)}(k))\},
\end{aligned} \tag{7.108}$$

where the fourth equality is obtained by adding and subtracting $F(k)\hat{f}(k, x_e(k))$, and the final equality results since $F(k)L(k) \equiv 0$. Hence the sliding dynamics are invariant to dynamical modelling errors/disturbances, and application of Theorem 5.11 gives exponential asymptotic stability. $\square$

Statement 3 of Theorem 7.6 treats the local exponential stability of both the reaching and sliding dynamics. Assumptions **A1** or **A1'** are invoked to obtain bounds on the term

$$\begin{aligned} \mathcal{I}_h^{-1}[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)]\mathbf{g}(k, \mathbf{s}_k^{(1)}) = \\ \mathcal{I}_h^{-1}[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)]\{\Gamma_k \text{Sat}(\mathbf{s}_k^{(1)}, \epsilon) + \Lambda_k \mathbf{s}_k^{(1)}\}. \end{aligned} \quad (7.109)$$

which appears through simplification of $\mathbf{v}^{(1)}(k)$ (7.78) in observer O_1 . These bounds are required during the reaching phase for $0 \leq k \leq k_s$. The designer has full control over parameters Γ_k and Λ_k of the reaching law which govern the rate of approach to the sliding surface where $\mathbf{x}(k+1) = \hat{\mathbf{x}}^{(1)}(k+1)$. One possible method available to the designer is to select the elements of Γ and Λ such that for all $k > 0$, $\|\mathbf{x}(k+1) - \hat{\mathbf{x}}^{(1)}(k+1)\| < \delta$ where δ is defined in hypothesis (7.67) of Theorem 7.5. In this case, Assumption **A1** would be guaranteed under the hypothesis of Theorem 7.5.

A primary limitation of Theorem 7.6 is that measurement errors have not been considered. Note that the derivation (7.91) leading to the reaching law for observer O_2 is independent of measurement errors in the sense that (7.91) holds for any measurement sequence $\mathbf{z}(k)$. Hence a global convergent quasi-sliding regime exists on $\mathbf{s}_k^{(2)} = \mathbf{0}$ even in the presence of measurement errors. However, even though the surface $\mathbf{s}_k^{(2)} = \mathbf{0}$ can be achieved independent of measurement errors, this does not mean that the state estimate is independent of those errors; if $\mathbf{z}(k) = \mathbf{h}(\mathbf{x}(k)) + \zeta(k)$, where $\zeta(k)$ represents measurement errors, then during sliding $\mathbf{s}_k^{(2)} = \mathbf{0} = \mathbf{h}(\mathbf{x}(k)) - \mathbf{h}(\hat{\mathbf{x}}(k)) + \zeta(k)$. Further work is required to analyze observer stability and robustness properties in the case of severe measurement errors.

The following example illustrates some of the properties of observer O_2 , answers a rather subtle question concerning the form of O_2 , and demonstrates the interaction between the sliding and reaching dynamics during the reaching phase (i.e. before $\mathbf{s}_k^{(2)} = \mathbf{0}$).

Example 7.2: Consider observer O_2 in (7.77)-(7.79). With the equivalence statement of Theorem 6.3, and for brevity in this example, the superscript “(2)” will be dropped. Sub-

stituting, from (7.91), the relationship

$$\mathbf{s}_{k+1} := \mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}(k+1)) = \mathbf{g}(k, \mathbf{s}_k) \quad (7.110)$$

into the expression for $\mathbf{v}(k)$ in (7.79) means that the observer can be simplified, assuming the nonsingularity of $\mathcal{I}_h[\cdot, \cdot]$, and expressed as

$$\begin{aligned} \hat{\mathbf{x}}_e(k+1) &= \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \mathbf{L}(k) \{ \mathcal{I}_h[\hat{\mathbf{x}}_p(k+1), \hat{\mathbf{x}}(k+1)] \mathbf{CL}(k) \}^{-1} \\ &\quad \cdot \{ \mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}_p(k+1)) - \mathbf{z}(k+1) + \mathbf{h}(\hat{\mathbf{x}}(k+1)) \} \\ &= \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \mathbf{L}(k) [\mathbf{CL}(k)]^{-1} [\hat{\mathbf{x}}(k+1) - \hat{\mathbf{x}}_p(k+1)]. \end{aligned} \quad (7.111)$$

The observer dynamics in (7.111) *appear* to be an implicit system of difference equations for $\hat{\mathbf{x}}_e(k+1)$ in terms of $\hat{\mathbf{x}}_e(k)$, *independent* of the measurements $\mathbf{z}(k+1)$ and the true extended state, $\mathbf{x}_e$, of the system. Continuing from (7.111), however, with $\hat{\mathbf{x}}(k+1) := \mathbf{C}\hat{\mathbf{x}}_e(k+1)$ and $\hat{\mathbf{x}}_p(k+1) := \mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k))$, gives

$$\hat{\mathbf{x}}_e(k+1) = \mathbf{F}(k)\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \mathbf{L}(k) [\mathbf{CL}(k)]^{-1} \hat{\mathbf{x}}_e(k+1) \quad (7.112)$$

which can be written as

$$\mathbf{F}(k)\hat{\mathbf{x}}_e(k+1) = \mathbf{F}(k)\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \quad (7.113)$$

where

$$\mathbf{F}(k) := [\mathbf{I}_n - \mathbf{L}(k) [\mathbf{CL}(k)]^{-1} \mathbf{C}] \quad (7.114)$$

is a singular matrix (the null space of $\mathbf{F}(k)$ is the column space of $\mathbf{L}(k)$). Hence (7.111), or equivalently (7.113), without (7.110) is an incomplete representation of the observer dynamics; Equation (7.111), or equivalently (7.113), represents only the component of the observer dynamics which lies in the null space of $\mathbf{C}$. The complete dynamics can be expressed as the coupled system, from (7.113) and (7.110),

$$\mathbf{h}(\hat{\mathbf{x}}(k+1)) = \mathbf{z}(k+1) - \mathbf{g}(k, \mathbf{s}_k). \quad (7.115)$$

$$\mathbf{F}(k)\hat{\mathbf{x}}_e(k+1) = \mathbf{F}(k)\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \quad (7.116)$$

The coupling between the reaching and sliding dynamics before the onset of pure sliding (i.e. before $\mathbf{s}_k = 0$) is illustrated more clearly by introducing

$$\hat{\mathbf{x}}_e := \hat{\mathbf{x}}_e^r + \hat{\mathbf{x}}_e^n, \quad (7.117)$$

where $\hat{\mathbf{x}}_e^r$ is the component of $\hat{\mathbf{x}}_e$ in the row space of $\mathbf{C}$, and $\hat{\mathbf{x}}_e^n$ is the component of $\hat{\mathbf{x}}_e$ in the null space of $\mathbf{C}$. The null space of $\mathbf{C}$ is the sliding surface in n -dimensional error space, and the row space is orthogonal to the null space. Then $\hat{\mathbf{x}} := \mathbf{C}\hat{\mathbf{x}}_e = \mathbf{C}\hat{\mathbf{x}}_e^r$ and

$$\mathbf{F}(k)\hat{\mathbf{x}}_e(k) = \hat{\mathbf{x}}_e^n(k) + \mathbf{F}(k)\hat{\mathbf{x}}_e^r(k). \quad (7.118)$$

Hence the coupled system (7.116) and (7.115) can be written as

$$\mathbf{h}(\mathbf{C}\hat{\mathbf{x}}_e^r(k+1)) = \mathbf{z}(k+1) - \mathbf{g}(k, \mathbf{s}_k). \quad (7.119)$$

$$\hat{\mathbf{x}}_e^n(k+1) = \mathbf{F}(k) [\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) - \hat{\mathbf{x}}_e^r(k+1)]. \quad (7.120)$$

Equation (7.119) represents the reaching law which specifies the observer dynamics in the row space of $\mathbf{C}$. The output, $\hat{\mathbf{x}}_e^r(k+1)$, of (7.119) then appears as a driving term in (7.120) which specifies the component of observer output that lies in the null space of $\mathbf{C}$. During sliding, with $\mathbf{g}(k, \mathbf{s}_k) \equiv \mathbf{0}$, the *zero dynamics* are governed by the constraint

$$\mathbf{h}(\mathbf{C}\hat{\mathbf{x}}_e^r(k+1)) = \mathbf{z}(k+1). \quad (7.121)$$

which must represent stable dynamics. □

The representation of O_2 in (7.119) and (7.120) suggests a possible methodology for implementation of the observer. Given the extended state estimate $\hat{\mathbf{x}}_e(k)$ and the measurements $\mathbf{z}(k+1)$, $\mathbf{s}_k$ can be computed and equation (7.119) could first be solved, perhaps numerically, for $\hat{\mathbf{x}}_e^r(k+1)$ which is then used in (7.120) to directly compute $\hat{\mathbf{x}}_e^n(k+1)$. The full observer output is then computed as

$$\hat{\mathbf{x}}_e(k+1) = \hat{\mathbf{x}}_e^r(k+1) + \hat{\mathbf{x}}_e^n(k+1). \quad (7.122)$$

This is essentially the methodology proposed in the following Section through introduction of local iterations at each measurement event. During the reaching phase, however, errors propagate in the numerical implementation over time due to the recursive nature of (7.119). The proposed local iterations operate on the observer extended state vector, $\hat{\mathbf{x}}_e$, and are constructed in order to facilitate analysis of the effect of numerical errors, particularly on the stability of the iterated observer.

7.3.3 Implementation With Local Iterations

In this Section, an approximate implementation of observer O_2 from (7.77) and (7.79) is proposed and analyzed. The primary difficulty in implementing observer O_2 is that the desired output, $\hat{\mathbf{x}}_e(k+1)$, of the observer is represented by an implicit system of nonlinear integral equations. In observer O_2 , as opposed to the representation in O_1 , the implicit representation results from the term $\mathcal{I}_h[\hat{\mathbf{x}}_p^{(2)}(k+1), \hat{\mathbf{x}}^{(2)}(k+1)]$ which appears on the left of (7.79), page 230, (see also (7.84), page 231). This term contains the quantity, $\hat{\mathbf{x}}^{(2)}(k+1) := \mathbf{C}\hat{\mathbf{x}}_e(k+1)$ which is not yet available, but is that quantity which an implementation of the observer is required to compute at time $k+1$. Numerical implementation methods proposed in this section focus on approximations to the term $\mathcal{I}_h[\hat{\mathbf{x}}_p^{(2)}(k+1), \hat{\mathbf{x}}^{(2)}(k+1)]$. The primary objective is to propose a practical implementation methodology for the observer and obtain stability results for the proposed implementation.

The following developments begin with a simple approximation, equivalent to that used in the Extended Kalman Filter (see Appendix B), which may lead to suitable results in some cases. This simple approximation does not readily lead to complete stability analysis of the implementation. This simple method does, however, lead to a generalized Newton approach, much like that proposed for continuous-time nonlinear map inversion by Nicosia *et al* [92, 93], and also to the formulation and analysis of observer implementation through local iterations. The proposed local iterations can be reduced to Newton's method for the estimation of a value for measured state, $\hat{\mathbf{x}}(k)$, which satisfies the reaching law. Significant effort, however, is required to obtain stability results for the full error $\mathbf{e}_e(k) := \mathbf{x}_e(k) - \hat{\mathbf{x}}_e(k)$ in this numerical implementation during both the reaching and sliding phases.

For brevity in the remainder of this section, the superscript "(2)" will be dropped. For the system (7.41, repeated)

$$\mathbf{x}_e(k+1) = \mathbf{f}(k, \mathbf{x}_e(k)), \quad \mathbf{x}(k) := \mathbf{C}\mathbf{x}_e(k), \quad (7.123)$$

with measurements (7.42, repeated)

$$\mathbf{z}(k) = \mathbf{h}(\mathbf{x}(k)), \quad (7.124)$$

observer O_2 has the form

$$\hat{\mathbf{x}}_e(k+1) = \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \mathbf{L}(k)\mathbf{v}(k), \quad (7.125)$$

where

$$\mathbf{v}(k) := [\mathcal{I}_h[\hat{\mathbf{x}}_p(k+1), \hat{\mathbf{x}}(k+1)]\mathbf{CL}(k)]^{-1} \quad (7.126)$$

$$\cdot \{\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}_p(k+1)) - \mathbf{g}(k, \mathbf{s}_k)\};$$

$$\hat{\mathbf{x}}_p(k+1) := \mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)), \quad (7.127)$$

$$\mathbf{s}_k := \mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}(k)), \quad (7.128)$$

$$\mathbf{g}(k, \mathbf{s}_k) := \mathbf{\Gamma}_k \mathbf{Sat}(\mathbf{s}_k, \epsilon) + \mathbf{\Lambda}_k \mathbf{s}_k. \quad (7.129)$$

The first approach for implementation results from expanding $\mathcal{I}_h[\hat{\mathbf{x}}_p(k+1), \hat{\mathbf{x}}(k+1)]$ as

$$\mathcal{I}_h[\hat{\mathbf{x}}_p(k+1), \hat{\mathbf{x}}(k+1)] := \int_0^1 \mathbf{J}_h(\mathbf{w}_\lambda(k+1)) d\lambda, \quad (7.130)$$

where

$$\begin{aligned} \mathbf{w}_\lambda(k+1) &:= \lambda \hat{\mathbf{x}}_p(k+1) + (1-\lambda) \hat{\mathbf{x}}(k+1) \\ &= \hat{\mathbf{x}}(k+1) + \lambda [\hat{\mathbf{x}}_p(k+1) - \hat{\mathbf{x}}(k+1)] \\ &= \mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \mathbf{CL}(k)\mathbf{v}(k) + \lambda [-\mathbf{CL}(k)\mathbf{v}(k)] \\ &= \hat{\mathbf{x}}_p(k+1) + (1-\lambda)\mathbf{CL}(k)\mathbf{v}(k). \end{aligned} \quad (7.131)$$

In implementations, a natural approach, which is essentially the same as that used in the extended Kalman filter, is to assume that $\|\mathbf{CL}(k)\mathbf{v}(k)\|$ is small and write

$$\mathcal{I}_h[\hat{\mathbf{x}}_p(k+1), \hat{\mathbf{x}}(k+1)] \approx \mathbf{J}_h(\hat{\mathbf{x}}_p(k+1)) := \frac{\partial \mathbf{h}}{\partial \mathbf{x}}[\mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k))]. \quad (7.132)$$

This approximation leads to the observer

$$\hat{\mathbf{x}}_e(k+1) = \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \mathbf{L}(k)\tilde{\mathbf{v}}(k), \quad (7.133)$$

$$\begin{aligned} \tilde{\mathbf{v}}(k) &:= [\mathbf{J}_h(\hat{\mathbf{x}}_p(k+1))\mathbf{CL}(k)]^{-1} \\ &\quad \cdot \{\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}_p(k+1)) - \mathbf{g}(k, \mathbf{s}_k)\}, \end{aligned} \quad (7.134)$$

$$\hat{\mathbf{x}}_p(k+1) := \mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)), \quad (7.135)$$

$$\mathbf{s}_k := \mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}(k)), \quad (7.136)$$

and $\mathbf{g}(k, \mathbf{s}_k)$ is defined by the reaching law in (7.129).

This observer represents a generalized Newton method in the sense that one step of Newton's method is applied at each measurement update. This is illustrated more clearly by using the nonsingularity of $[\mathbf{CL}(k)]$, writing

$$\begin{aligned} \mathbf{C}\hat{\mathbf{x}}_e(k+1) &:= \hat{\mathbf{x}}(k+1) = \\ &\hat{\mathbf{x}}_p(k+1) + \mathbf{J}_h^{-1}(\hat{\mathbf{x}}_p(k+1)) \{ \mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}_p(k+1)) - \mathbf{g}(k, \mathbf{s}_k) \}, \end{aligned} \quad (7.137)$$

and noting that

$$-\mathbf{J}_h(\mathbf{y}) = \left. \frac{\partial}{\partial \hat{\mathbf{x}}} \{ \mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}) - \mathbf{g}(k, \mathbf{s}_k) \} \right|_{\hat{\mathbf{x}}=\mathbf{y}}. \quad (7.138)$$

Hence (7.137) represents one Newton step towards the computation of a zero of reaching law $\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}) - \mathbf{g}(k, \mathbf{s}_k) = \mathbf{0}$ starting from the initial estimate $\hat{\mathbf{x}}_p(k+1)$. For a detailed treatment of Newton's method and several of its variants, see Ortega and Rheinboldt [159] or Blum [162].

This approximate observer no longer satisfies the desired reaching law, but approximates the reaching dynamics up to linear terms of a Taylor series expansion of the function $\mathbf{h}(\cdot)$. This result is based on a characteristic property of Newton's method.

Lemma 7.3: The observer (7.133) satisfies

$$\mathbf{s}_{k+1} = \mathbf{g}(k, \mathbf{s}_k) - \mathbf{r}(k+1), \quad (7.139)$$

where $\mathbf{r}(k+1)$ is the remainder after truncation to linear terms in the Taylor series expansion of $\mathbf{h}(\mathbf{x})$ about $\hat{\mathbf{x}}_p(k+1)$ in the computation of $\mathbf{h}(\hat{\mathbf{x}}(k+1))$.

Proof: Consider the Taylor series expansion with remainder

$$\mathbf{h}(\hat{\mathbf{x}}(k+1)) = \mathbf{h}(\hat{\mathbf{x}}_p(k+1)) + \mathbf{J}_h(\hat{\mathbf{x}}_p(k+1)) [\hat{\mathbf{x}}(k+1) - \hat{\mathbf{x}}_p(k+1)] + \mathbf{r}_{k+1}. \quad (7.140)$$

Also, using (7.133) and recalling that

$$\hat{\mathbf{x}}(k+1) - \hat{\mathbf{x}}_p(k+1) = \mathbf{CL}(k)\hat{\mathbf{v}}(k), \quad (7.141)$$

gives

$$\begin{aligned} &\mathbf{J}_h(\hat{\mathbf{x}}_p(k+1)) [\hat{\mathbf{x}}(k+1) - \hat{\mathbf{x}}_p(k+1)] \\ &= \mathbf{J}_h(\hat{\mathbf{x}}_p(k+1)) \mathbf{CL}(k) \hat{\mathbf{v}}(k) \\ &= \mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}_p(k+1)) - \mathbf{g}(k, \mathbf{s}_k). \end{aligned} \quad (7.142)$$

Using (7.140) and (7.142),

$$\begin{aligned}
 \mathbf{s}_{k+1} &:= \mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}(k+1)) \\
 &= \{\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}_p(k+1))\} + \{\mathbf{h}(\hat{\mathbf{x}}_p(k+1)) - \mathbf{h}(\hat{\mathbf{x}}(k+1))\} \\
 &= \{\mathbf{g}(k, \mathbf{s}_k) + \mathbf{J}_h(\hat{\mathbf{x}}_p(k+1))[\hat{\mathbf{x}}(k+1) - \hat{\mathbf{x}}_p(k+1)]\} \\
 &\quad + \{-\mathbf{J}_h(\hat{\mathbf{x}}_p(k+1))[\hat{\mathbf{x}}(k+1) - \hat{\mathbf{x}}_p(k+1)] - \mathbf{r}_{k+1}\} \\
 &= \mathbf{g}(k, \mathbf{s}_k) - \mathbf{r}_{k+1}.
 \end{aligned} \tag{7.143}$$

□

This observer provides a state estimate $\hat{\mathbf{x}}_e(k)$ at time k . When measurements are available at time $k+1$, the new estimate is

$$\hat{\mathbf{x}}_e(k+1) = \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \mathbf{L}(k)\mathbf{v}(k). \tag{7.144}$$

With

$$\mathbf{s}_{k+1} = \mathbf{z}(k+1) - \mathbf{h}(\mathbf{C}\hat{\mathbf{x}}_e(k)), \tag{7.145}$$

the remainder

$$\mathbf{r}_{k+1} := \mathbf{g}(k, \mathbf{s}_k) - \mathbf{s}_{k+1} \tag{7.146}$$

can be computed. This result provides a means to monitor the progress or behavior of the simple approximation since $\mathbf{r}_{k+1}$ at each time step can be computed to determine if it decreases to within acceptable levels. However, analysis which yields theoretical bounds on this approximate observer are not immediate. It is natural, then, to ask if this value of $\mathbf{r}_{k+1}$ could be used to improve the state estimate and thereby reduce the effect of the approximation. One method is to employ $\mathbf{r}_{k+1}$ in local Newton-type iterations at each measurement event, much like those used in the Iterated Extended Kalman Filter [94].

Suppose at time k an estimate $\hat{\mathbf{x}}_e(k)$ is available. Set

$$\mathbf{s}_k := \mathbf{z}(k) - \mathbf{h}(\mathbf{C}\hat{\mathbf{x}}_e(k)). \tag{7.147}$$

$$\mathbf{g}(k, \mathbf{s}_k) = \mathbf{\Gamma}_k \mathbf{Sat}(\mathbf{s}_k, \epsilon) + \mathbf{\Lambda}_k \mathbf{s}_k, \text{ and} \tag{7.148}$$

$$\hat{\mathbf{x}}_e^{(0)}(k+1) := \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)). \tag{7.149}$$

In this notation, $\hat{\mathbf{x}}_e^{(0)}(k+1)$ is the predicted extended state based on the estimate $\hat{\mathbf{x}}_e(k)$ at time k . With the availability of measurements $\mathbf{z}(k+1)$, the proposed local iterations take the form, for k fixed, $N_{k+1} \geq 1$, and with $j = 0, 1, \dots, N_{k+1} - 1$,

$$\begin{aligned} \mathbf{s}_{k+1}^{(j)} &:= \mathbf{z}(k+1) - \mathbf{h}(\mathbf{C}\hat{\mathbf{x}}_e^{(j)}(k+1)), \\ \mathbf{r}_{k+1}^{(j)} &= \mathbf{g}(k, \mathbf{s}_k) - \mathbf{s}_{k+1}^{(j)}, \\ \mathbf{v}^{(j)}(k) &= \left[\mathbf{J}_h(\mathbf{C}\hat{\mathbf{x}}_e^{(j)}(k+1)) \mathbf{C} \mathbf{L}^{(j)}(k) \right]^{-1} \mathbf{r}_{k+1}^{(j)}, \\ \hat{\mathbf{x}}_e^{(j+1)}(k+1) &:= \hat{\mathbf{x}}_e^{(j)}(k+1) - \mathbf{L}^{(j)}(k) \mathbf{v}^{(j)}(k). \end{aligned} \quad (7.150)$$

The observer output at time $k+1$ is then taken as

$$\hat{\mathbf{x}}_e(k+1) := \hat{\mathbf{x}}_e^{(N_{k+1})}(k+1). \quad (7.151)$$

Alternatively, the iterations in (7.150) can be expressed as

$$\hat{\mathbf{x}}_e^{(j+1)}(k+1) = \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \sum_{\ell=0}^j \mathbf{L}^{(\ell)}(k) \mathbf{v}^{(\ell)}(k), \quad j = 0, 1, \dots, N_{k+1} - 1, \quad (7.152)$$

where

$$\begin{aligned} \mathbf{v}^{(\ell)}(k) &:= \left[\mathbf{J}_h(\hat{\mathbf{x}}^{(\ell)}(k+1)) \mathbf{C} \mathbf{L}^{(\ell)}(k) \right]^{-1} \\ &\quad \left\{ \mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}^{(\ell)}(k+1)) - \mathbf{g}(k, \mathbf{s}_k) \right\}, \end{aligned} \quad (7.153)$$

and $\hat{\mathbf{x}}^{(\ell)}(k+1) := \mathbf{C}\hat{\mathbf{x}}_e^{(\ell)}(k+1)$. In this form, it is easily seen, by multiplying (7.152) by $\mathbf{C}$ on the left to get

$$\begin{aligned} \hat{\mathbf{x}}^{(j+1)}(k+1) &= \hat{\mathbf{x}}^{(0)}(k+1) \\ &\quad + \sum_{\ell=0}^j \mathbf{J}_h^{-1}(\hat{\mathbf{x}}^{(\ell)}(k+1)) \left[\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}^{(\ell)}(k+1)) - \mathbf{g}(k, \mathbf{s}_k) \right], \end{aligned} \quad (7.154)$$

that the local iterations represent Newton's method to obtain a zero of the reaching law $\{\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}(k+1)) - \mathbf{g}(k, \mathbf{s}_k)\} = \mathbf{0}$.

The following Lemma is a generalization of the previous result in Lemma 7.3 for the iterated observer.

Lemma 7.4: The vectors $\mathbf{r}_{k+1}^{(j)}$ in (7.150) satisfy, for $j = 1, 2, \dots$,

$$\begin{aligned} \mathbf{r}_{k+1}^{(j)} &= \mathbf{h}(\hat{\mathbf{x}}^{(j)}(k+1)) - \mathbf{h}(\hat{\mathbf{x}}^{(j-1)}(k+1)) \\ &\quad - \mathbf{J}_h(\hat{\mathbf{x}}^{(j-1)}(k+1)) \left[\hat{\mathbf{x}}^{(j)}(k+1) - \hat{\mathbf{x}}^{(j-1)}(k+1) \right]. \end{aligned} \quad (7.155)$$

Proof: By direct computation, and using the relationships in (7.150),

$$\begin{aligned}
 \mathbf{r}_{k+1}^{(j)} &= \mathbf{g}(k, \mathbf{s}_k) - \mathbf{s}_{k+1}^{(j)} \\
 &= \mathbf{g}(k, \mathbf{s}_k) - \mathbf{z}(k+1) + \mathbf{h}(\hat{\mathbf{x}}^{(j)}(k+1)) \\
 &\quad + \mathbf{h}(\hat{\mathbf{x}}^{(j-1)}(k+1)) - \mathbf{h}(\hat{\mathbf{x}}^{(j-1)}(k+1)) \\
 &= -\left\{ \mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}^{(j-1)}(k+1)) - \mathbf{g}(k, \mathbf{s}_k) \right\} \\
 &\quad + \mathbf{h}(\hat{\mathbf{x}}^{(j)}(k+1)) - \mathbf{h}(\hat{\mathbf{x}}^{(j-1)}(k+1)) \\
 &= -\mathbf{J}_h(\hat{\mathbf{x}}^{(j-1)}(k+1)) \left[\mathbf{C}\mathbf{L}^{(j-1)}(k+1)\mathbf{v}^{(j-1)}(k+1) \right] \\
 &\quad + \mathbf{h}(\hat{\mathbf{x}}^{(j)}(k+1)) - \mathbf{h}(\hat{\mathbf{x}}^{(j-1)}(k+1)) \\
 &= -\mathbf{J}_h(\hat{\mathbf{x}}^{(j-1)}(k+1)) \left[\hat{\mathbf{x}}^{(j)}(k+1) - \hat{\mathbf{x}}^{(j-1)}(k+1) \right] \\
 &\quad + \mathbf{h}(\hat{\mathbf{x}}^{(j)}(k+1)) - \mathbf{h}(\hat{\mathbf{x}}^{(j-1)}(k+1)). \tag{7.156}
 \end{aligned}$$

□

Hence the vectors $\mathbf{r}_{k+1}^{(j)}$ represent the remainder, after truncation to linear terms, of the Taylor series expansion of $\mathbf{h}(\mathbf{x})$ about $\hat{\mathbf{x}}^{(j-1)}(k+1)$ in the computation of $\mathbf{h}(\hat{\mathbf{x}}^{(j)}(k+1))$. Moreover, $\mathbf{r}_{k+1}^{(j)}$ can be computed to monitor the progress of the iterated observer.

The primary objective of the remainder of this Section is to specify sufficient conditions for stability of the iterated observer. The proposed approach first assumes stability, through Theorem 6.3, of the ideal observer forms. The problem is thus reduced to specifying sufficient conditions for the iterated observer output to lie within an arbitrarily small tolerance of the ideal observer output. In this case the output of the iterated observer can be made arbitrarily close to the ideal observer output and hence stability and robustness properties of the ideal form carry over, to within an arbitrarily small tolerance, to the iterated observer. More specifically, let

$$\hat{\mathbf{x}}_e^*(k) \sim \text{Ideal output from observer } O_2. \tag{7.157}$$

The objective of the remainder of this Section is to specify sufficient conditions such that for any given tolerance $\epsilon_{\mathbf{x}_e} > 0$ there exists, for each $k \geq 0$, an integer $N_k \geq 0$ such that

$$\|\hat{\mathbf{x}}_e^*(k) - \hat{\mathbf{x}}_e(k)\| := \|\hat{\mathbf{x}}_e^*(k) - \hat{\mathbf{x}}_e^{(N_k)}(k)\| < \epsilon_{\mathbf{x}_e}. \tag{7.158}$$

There are at least three significant difficulties in obtaining this result:

Convergence of Local Iterations : The proposed local iterations will generally have a finite region of attraction, and this region must be of sufficient extent at each measurement update;

Error Propagation : During the reaching phase errors from previous measurement updates propagate, as illustrated in (7.119) of Example 7.2, which means that the local iterations, even if they converge, need not converge at each measurement update precisely to the ideal observer output; and

Assumptions : Several assumptions employed to obtain the desired result are taken from hypothesis of Theorems 5.11 and 5.12. Several other hypothesis have not previously been introduced, but some of these are closely related to those already discussed in Theorems 5.11 and 5.12, and Section 7.3.1. Some of these assumptions, such as Assumption A1 (page 224) investigated in Section 7.3.1, may be overly restrictive and require further analysis in future work.

Each of the above items are now discussed in more detail in order to highlight the fundamental difficulties in obtaining the desired result, and to emphasize limitations of the analysis methods available at this time. Future work must address these current limitations.

Convergence of Local Iterations. As indicated in (7.154), the local iterations, at time $k + 1$, can be reduced to Newton iterations for computation of a zero of the expression

$$\mathbf{G}(\hat{\mathbf{x}}(k + 1)) := \mathbf{z}(k + 1) - \mathbf{h}(\hat{\mathbf{x}}(k + 1)) - \mathbf{g}(k, \mathbf{s}_k). \quad (7.159)$$

which satisfies

$$\mathbf{J}_{\mathbf{G}}(\mathbf{x}) = -\mathbf{J}_{\mathbf{h}}(\mathbf{x}). \quad (7.160)$$

Definition 7.4: Point and Region of Attraction. Let $\Theta : \mathcal{D} \subset \mathbb{R}^m \mapsto \mathbb{R}^m$, and consider the iterative process

$$\mathbf{y}^{(j+1)} = \Theta(\mathbf{y}^{(j)}). \quad (7.161)$$

Then $\mathbf{y}_a$ is a *point of attraction* of the iteration (7.161) if there is an open neighbourhood $\mathcal{D}_0$ of $\mathbf{y}_a$ such that $\mathcal{D}_0 \in \mathcal{D}$ and $\forall \mathbf{y}^{(0)} \in \mathcal{D}_0$, the iterates $\{\mathbf{y}^{(j)}\}$ all lie in $\mathcal{D}_0$ and converge to $\mathbf{y}_a$. The largest such region, $\mathcal{D}_0$, for any $\mathbf{y}_a$ is called the *region of attraction* of $\mathbf{y}_a$ and is denoted $\mathcal{R}(\mathbf{y}_a)$. $\square$

Several results which govern convergence of Newton's method are now reviewed.

Theorem 7.7: Newton Attraction Theorem Assume that $\mathbf{G} : \mathcal{D} \subset \mathbb{R}^m \mapsto \mathbb{R}^m$ is continuously differentiable on an open neighbourhood $\mathcal{D}_0 \subset \mathcal{D}$ of a point $\mathbf{y}^* \in \mathcal{D}$ for which $\mathbf{G}(\mathbf{y}^*) = \mathbf{0}$, and that $\mathbf{J}_{\mathbf{G}}(\mathbf{y}^*)$ is nonsingular. Then $\mathbf{y}^*$ is a point of attraction of the Newton iteration

$$\mathbf{y}^{(i+1)} = \mathbf{y}^{(i)} - \mathbf{J}_{\mathbf{G}}^{-1}(\mathbf{y}^{(i)})\mathbf{G}(\mathbf{y}^{(i)}). \quad (7.162)$$

Proof: See Ortega and Rheinboldt [159, Theorem 10.2.2]. $\square$

Theorem 7.7 concludes that if $\mathbf{y}^*$ is a zero of $\mathbf{G}$ and $\mathbf{J}_{\mathbf{G}}(\mathbf{y}^*)$ is nonsingular, then there exists a region of attraction for the iterations (7.162), but no information is given about the extent of this region. In order to obtain additional results concerning the region of attraction, further information about the function $\mathbf{G}$ is required. There are a very large number of published results on convergence of Newton's method, from which the following three Theorems give an appreciation for the nature of sufficient conditions employed. Let $\mathcal{B}(\mathbf{y}_0, \epsilon)$ denote the open ball centred at $\mathbf{y}_0$ with radius ϵ ,

$$\mathcal{B}(\mathbf{y}_0, \epsilon) := \{\mathbf{y} \mid \|\mathbf{y} - \mathbf{y}_0\| < \epsilon\}, \quad (7.163)$$

and let $\bar{\mathcal{B}}(\mathbf{y}_0, \epsilon)$ denote the closure of $\mathcal{B}(\mathbf{y}_0, \epsilon)$.

Theorem 7.8: Newton-Kantorovich Theorem: Assume that $\mathbf{G} : \mathcal{D} \subset \mathbb{R}^m \mapsto \mathbb{R}^m$ is continuously differentiable on an open convex set $\mathcal{D}_0 \subset \mathcal{D}$ and that

$$\|\mathbf{J}_{\mathbf{G}}(\mathbf{y}_1) - \mathbf{J}_{\mathbf{G}}(\mathbf{y}_2)\| \leq \gamma \|\mathbf{y}_1 - \mathbf{y}_2\|, \quad \forall \mathbf{y}_1, \mathbf{y}_2 \in \mathcal{D}_0. \quad (7.164)$$

Suppose that there exists a $\mathbf{y}^{(0)} \in \mathcal{D}_0$ such that $\|\mathbf{J}_{\mathbf{G}}^{-1}(\mathbf{y}^{(0)})\| \leq \beta$ and $\|\mathbf{J}_{\mathbf{G}}^{-1}(\mathbf{y}^{(0)})\mathbf{G}(\mathbf{y}^{(0)})\| \leq \eta$ where $\alpha := \beta\gamma\eta \leq \frac{1}{2}$. Set

$$t_1 := (\beta\gamma)^{-1}[1 - (1 - 2\alpha)^{\frac{1}{2}}] \quad (7.165)$$

$$t_2 := (\beta\gamma)^{-1}[1 + (1 - 2\alpha)^{\frac{1}{2}}], \quad (7.166)$$

and assume that $\bar{\mathcal{B}}(\mathbf{y}^{(0)}, t_1) \subset \mathcal{D}_0$. Then the Newton iterates

$$\mathbf{y}^{(i+1)} = \mathbf{y}^{(i)} - \mathbf{J}_{\mathbf{G}}^{-1}(\mathbf{y}^{(i)})\mathbf{G}(\mathbf{y}^{(i)}) \quad (7.167)$$

are well-defined, remain in $\bar{B}(y^{(0)}, t_2)$, and converge to a solution y^* of $G(y) = 0$ which is unique in $\bar{B}(y^{(0)}, t_2) \cap \mathcal{D}_0$. Moreover,

$$\|y^* - y^{(i)}\| \leq (2^i \beta \gamma)^{-1} (2\alpha)^{2^i}. \quad (7.168)$$

Proof: See Ortega and Rheinboldt [159, Theorem 12.6.2]. $\square$

Theorem 7.9: Newton-Mysovskii Theorem: Assume that $G : \mathcal{D} \subset \mathbb{R}^m \mapsto \mathbb{R}^m$ is continuously differentiable on an open convex set $\mathcal{D}_0 \subset \mathcal{D}$ and that

$$\|J_G(y_1) - J_G(y_2)\| \leq \gamma \|y_1 - y_2\|, \quad \forall y_1, y_2 \in \mathcal{D}_0. \quad (7.169)$$

Suppose that $\forall y \in \mathcal{D}_0$, $\|J_G^{-1}(y)\| \leq \beta$ and that $y^{(0)} \in \mathcal{D}_0$ is such that $\|J_G^{-1}(y^{(0)})G(y^{(0)})\| \leq \eta$ where $\alpha := \beta\gamma\eta/2 \leq 1$. Assume further that $\bar{B}(y_0, r_0) \subset \mathcal{D}_0$ where

$$r_0 := \eta \sum_{j=0}^{\infty} \alpha^{2^j-1}. \quad (7.170)$$

Then the Newton iterates

$$y^{(i+1)} = y^{(i)} - J_G^{-1}(y^{(i)})G(y^{(i)}) \quad (7.171)$$

are well-defined, remain in $\bar{B}(y^{(0)}, r_0)$, and converge to a solution y^* of $G(y) = 0$. Moreover

$$\|y^* - y^{(i)}\| \leq \frac{\eta \alpha^{2^i-1}}{1 - \alpha^{2^i}}. \quad (7.172)$$

Proof: See Ortega and Rheinboldt [159, Theorem 12.4.6]. $\square$

Theorem 7.10: Let $G : \mathcal{D} \subset \mathbb{R}^m \mapsto \mathbb{R}^m$ be such that $G''(y) := [\partial J_G / \partial y](y)$ exists in some open set $\mathcal{D}_0 \subset \mathcal{D}$. Suppose $y^{(0)} \in \mathcal{D}_0$ is such that $J_G^{-1}(y^{(0)})$ exists, and in some closed ball $\bar{B}(y^{(0)}, r_0)$ the following conditions hold for all $y \in \bar{B}(y^{(0)}, r_0)$ and some constant $0 < L < 1$:

1. $\|J_G^{-1}(y^{(0)})\| \|G''(y)\| \leq \frac{L}{r_0}$,
2. $\|J_G^{-1}(y)G(y)\| \leq \frac{(1-L)}{r_0}$, and
3. $\|J_G^{-1}(y)\| \|G''(y)\| \leq \frac{L}{r_0}$.

Then the sequence of Newton iterates

$$\mathbf{y}^{(i+1)} = \mathbf{y}^{(i)} - \mathbf{J}_{\mathbf{G}}^{-1}(\mathbf{y}^{(i)})\mathbf{G}(\mathbf{y}^{(i)}) \quad (7.173)$$

starting with $\mathbf{y}^{(0)}$ converges to a unique $\mathbf{y}^* \in \bar{\mathcal{B}}(\mathbf{y}^{(0)}, r_0)$ for which $\mathbf{G}(\mathbf{y}^*) = \mathbf{0}$. Furthermore, the convergence is quadratic in the sense that there exists a constant c such that

$$\|\mathbf{y}^* - \mathbf{y}^{(i+1)}\| \leq c\|\mathbf{y}^* - \mathbf{y}^{(i)}\|^2. \quad (7.174)$$

Proof: See Blum [162, Theorem 5.4.1]. $\square$

Theorems 7.8, 7.9, and 7.10, in essence, are based on very similar hypothesis which require that

1. $\mathbf{J}_{\mathbf{G}}$ be Lipschitz continuous (Condition 1 of Theorem 7.10 can also lead to Lipschitz continuity of $\mathbf{J}_{\mathbf{G}}$).
2. $\|\mathbf{J}_{\mathbf{G}}^{-1}\|$ be bounded, and
3. $\|\mathbf{J}_{\mathbf{G}}^{-1}\mathbf{G}\|$ be bounded.

It is interesting first to note that Items 1 and 2 above are essentially the same hypothesis used in Theorem 7.5 (page 227) to obtain a region of nonsingularity for $\mathcal{I}_{\mathbf{h}}[\cdot, \cdot]$. Theorems 7.8, 7.9, and 7.10 all are concerned with an initial iterate $\mathbf{y}^{(0)}$ and a region $\bar{\mathcal{B}}(\mathbf{y}^{(0)}, r)$ such that all iterates remain in $\bar{\mathcal{B}}(\mathbf{y}^{(0)}, r)$ and converge to a unique point $\mathbf{y}^* \in \bar{\mathcal{B}}(\mathbf{y}^{(0)}, r)$ which satisfies $\mathbf{G}(\mathbf{y}^*) = \mathbf{0}$. The method of proofs differ in the three Theorems, but are essentially structured to obtain a contraction mapping in $\bar{\mathcal{B}}(\mathbf{y}^{(0)}, r)$ from which the result follows from the Contraction Mapping Theorem (Theorem 5.4, page 5.4). This procedure is best illustrated by defining the function

$$\Theta(\mathbf{y}) := \mathbf{y} - \mathbf{J}_{\mathbf{G}}^{-1}(\mathbf{y})\mathbf{G}(\mathbf{y}). \quad (7.175)$$

which satisfies (see Blum [162, p. 198]), $\forall \boldsymbol{\eta} \in \mathbb{R}^m$,

$$[\mathbf{J}_{\Theta}(\mathbf{y})]\boldsymbol{\eta} = \mathbf{J}_{\mathbf{G}}^{-1}(\mathbf{y}) \{[\mathbf{G}''(\mathbf{y})]\boldsymbol{\eta}\} \mathbf{J}_{\mathbf{G}}^{-1}(\mathbf{y})\mathbf{G}(\mathbf{y}). \quad (7.176)$$

Note that the expressions $\mathbf{G}'' : \mathbb{R}^m \mapsto \mathbb{R}^{m \times m \times m}$ and $\{[\mathbf{G}''(\mathbf{y})]\boldsymbol{\eta}\} : \mathbb{R}^m \times \mathbb{R}^m \mapsto \mathbb{R}^{m \times m}$ are evaluated according to (D.9) of Appendix D. The function $\Theta(\mathbf{y})$ will then be contractive

if

$$\begin{aligned}\|J_{\theta}(y)\| &= \|J_G^{-1}(y)G'(y)J_G^{-1}(y)G(y)\| \\ &\leq \|J_G^{-1}(y)\| \|G''(y)\| \|J_G^{-1}(y)G(y)\| < 1.\end{aligned}\quad (7.177)$$

In the above, bounds on $\|G''(y)\|$ can be replaced by Lipschitz conditions on $J_G(y)$.

From the above discussion, it is apparent that practical verification of hypothesis which guarantee convergence of the local iterations could be very tedious. Some work has been devoted during the course of this research to investigation of alternatives and extensions to Newton's method such as variable step length algorithms [159, Section 8.3] and continuation methods [159, Section 10.4]. These alternatives are not included in this thesis, but appear to offer significant advantages in increasing the region of attraction and improving convergence rates, and should be considered in detail in future work.

The stability proof offered in Theorem 7.12 below states that if the region of convergence at each measurement event contains the open ball centred at the point of attraction with radius $\|CL(k)v(k)\|$, then, provided several additional conditions are satisfied, the stability result follows. This radius, $\|CL(k)v(k)\|$, is dependent on stability of the ideal observer O_2 , and bounds on this quantity are provided in Lemma 7.5 below by appealing to the hypothesis of Theorem 5.11. Newton stability results available in the literature, such as Theorems 7.8, 7.9, and 7.10 reviewed above, could in principle be used to verify that this constraint is satisfied.

Error Propagation. Another primary difficulty in obtaining the result in (7.158) is that during the reaching phase, the local iterations, even if they converge, need not converge at each measurement update to the ideal observer output. During the reaching phase errors from previous measurement updates propagate through the local iterations over time as illustrated in (7.119) of Example 7.2. This concept is now discussed in more detail.

Using superscript "*" to denote quantities associated with the ideal observer,

$$s_k^* := z(k) - h(\hat{x}^*(k)), \quad (7.178)$$

and

$$s_k^{(\ell)} := z(k) - h(\hat{x}^{(\ell)}(k)), \quad (7.179)$$

and hence, unless $\hat{\mathbf{x}}^*(k) = \hat{\mathbf{x}}^{(\ell)}(k)$, $\mathbf{s}_k^* \neq \mathbf{s}_k^{(\ell)}$. The ideal output, $\hat{\mathbf{x}}^*(k+1)$, from observer $\mathcal{O}_2$ satisfies the ideal reaching law

$$\mathbf{s}_{k+1}^* := \mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}^*(k+1)) = \mathbf{g}(k, \mathbf{s}_k^*). \quad (7.180)$$

Define the sequence $\{\hat{\mathbf{x}}^b(j), j = 0, 1, 2, \dots\}$ which satisfies

$$\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}^b(k+1)) = \mathbf{g}(k, \mathbf{s}_k^{(N_k)}), \quad (7.181)$$

where $\mathbf{s}_k^{(N_k)}$ has been computed after performing N_k local iterations at time k ,

$$\mathbf{s}_k^{(N_k)} := \mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}^{(N_k)}(k)). \quad (7.182)$$

In this case, $\mathbf{s}_k^{(N_k)}$ depends on all past approximations which means that during the reaching phase, $\hat{\mathbf{x}}^*(k) \neq \hat{\mathbf{x}}^b(k)$ in general. The iterates $\hat{\mathbf{x}}^{(\ell)}(k)$, if they converge, converge to $\hat{\mathbf{x}}^b(k)$, not necessarily to $\hat{\mathbf{x}}^*(k)$, and this represents a primary difficulty in the stability proof. Another possible limitation is that *existence* of the sequence $\{\hat{\mathbf{x}}^b(j), j = 0, 1, 2, \dots\}$ which satisfies (7.181) is assumed, which may require that $\mathbf{h}(\cdot)$ be an "onto" mapping.

Assumptions. Several assumptions employed to obtain the desired result are taken from hypothesis of Theorems 5.11 and 5.12. Several other hypothesis have not previously been introduced, but some of these are closely related to those already discussed in Theorems 5.11 and 5.12, and Section 7.3.1. First, the matrices $\mathbf{L}^{(\ell)}(k)$ in the iterated observer are defined as being related to the matrices $\mathbf{L}(k)$ in the ideal observer form according to

$$\mathbf{L}^{(\ell)}(k) := \mathbf{L}(k)\mathbf{Q}^{(\ell)}(k). \quad (7.183)$$

where the product $\mathbf{CL}(k)$ is nonsingular and $\mathbf{Q}^{(\ell)}(k) \in \mathbb{R}^{q \times q}$ is nonsingular for all $\ell = 0, 1, 2, \dots$ and $k = 0, 1, 2, \dots$. In this case

$$\mathbf{L}(k)[\mathbf{CL}(k)]^{-1} = \mathbf{L}^{(\ell)}(k)[\mathbf{CL}^{(\ell)}(k)]^{-1}. \quad (7.184)$$

With the definitions

$$\mathbf{F}(k) := [\mathbf{I}_n - \mathbf{L}(k)[\mathbf{CL}(k)]^{-1}\mathbf{C}], \quad (7.185)$$

$$\mathbf{B}(k, \mathbf{d}(k, \mathbf{y})) := \mathbf{f}(k, \mathbf{y}) - \hat{\mathbf{f}}(k, \mathbf{y}), \quad \forall \mathbf{y} \in \mathbf{X}_e, \quad (7.186)$$

the following properties of the system and observer parameters are assumed.

A1 : (repeated from page 224), $h : \mathcal{A}_0 \subseteq \mathbb{R}^m \mapsto \mathcal{B}_0 \subseteq \mathbb{R}^m$, is C^p , $p \geq 1$, is 1-to-1 from $\mathcal{A}_0$ onto $\mathcal{B}_0$, and

1. $\forall \mathbf{x} \in \mathcal{A}_0$, $J_h(\mathbf{x})$ is nonsingular (local homeomorphic property),
2. $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{A} \subseteq \mathcal{A}_0$, $\mathcal{I}_h(\mathbf{x}_1, \mathbf{x}_2)$ is nonsingular and

$$\|\mathcal{I}_h^{-1}(\mathbf{x}_1, \mathbf{x}_2)\| \leq L_{h^{-1}} < \infty. \quad (7.187)$$

A2 : Parameters Γ_k , Λ_k , and ϵ of the reaching law (7.129) satisfy the hypothesis of Theorem 5.12, page 171. Hence with

$$L_g(k) := \left\{ \frac{\|\Lambda_k\|_p}{\min\{\epsilon_i\}} + \|\Gamma_k\|_p \right\}, \quad (7.188)$$

where the vector norm $\|\cdot\|_p$ is defined in Definition 5.7,

$$\bar{L}_g := \sup_{\{k \geq 0\}} L_g(k) < 1. \quad (7.189)$$

Application of Corollary 5.2 provides a means to specify $k_s \geq 0$ such that during pure sliding, $L_g(k) = 0$ for all $k \geq k_s$.

A3 : $\mathbf{B}(k) \in \mathbb{R}^{n \times p}$, $\mathbf{d}(k, \mathbf{x}_e^*) \in \mathbb{R}^{p \times 1}$, $\mathbf{F}(k)$, and $\hat{\mathbf{f}}$ satisfy the hypothesis of Theorem 5.11, page 161, where the parameters ϵ_2 , $\xi(k)$ and $\eta(k)$ of Theorem 5.11 satisfy

$$\epsilon_2 > 0, \quad (7.190)$$

$$\xi(k) := -L_{h^{-1}} \mathbf{g}(k, \mathbf{s}_k^*), \quad \text{and} \quad (7.191)$$

$$\eta(k) = 0, \quad (7.192)$$

With Ω and the sequence $\mathbf{W}_k$ as defined in Theorem 5.11, define also

$$\lambda := \sup_{k \geq 0, 0 \leq j \leq k} \left\{ \|\mathbf{W}_k^{-1}\| \|\mathbf{W}_j \mathbf{L}(j-1) [\mathbf{C} \mathbf{L}(j-1)]^{-1}\| \right\}, \quad (7.193)$$

and a constant $L_{\hat{\mathbf{f}}}$ such that

$$\sup_{\{k \geq 0\}} \sup_{\{\mathbf{v} \in \Omega\}} \left\| \mathbf{C} \frac{\partial \hat{\mathbf{f}}}{\partial \hat{\mathbf{x}}_e^*}(k, \mathbf{v}) \right\| \leq L_{\hat{\mathbf{f}}} < \infty. \quad (7.194)$$

Assumption A1 has already been discussed in Section 7.3.1 and in the proof of Theorem 7.6. Again, this assumption can be substituted with Assumption A1', page 225, in the theoretical results which follow. Assumption A2 is not restrictive since the designer has control over specification of parameters Γ_k , Λ_k , and ϵ . Dependency on the specific norms used is eliminated through the following fundamental result.

Theorem 7.11: Norm Equivalence Let $\|\cdot\|$ and $\|\cdot\|'$ be any two norms on $\mathbb{R}^n$. Then there exists two constants, $c_2 \geq c_1 > 0$, such that $\forall \xi \in \mathbb{R}^n$,

$$c_1 \|\xi\| \leq \|\xi\|' \leq c_2 \|\xi\|. \quad (7.195)$$

Proof: See Horn and Johnson [150, Corollary 5.4.5]. □

Assumption A3 is based on stability, under the hypothesis of Theorem 5.11, of the ideal observer O_2 as detailed in Theorem 7.6. Existence of constants λ and L_f , defined in (7.193) and (7.194), respectively, is obviously guaranteed under these hypothesis.

The following three Lemmas establish results which are fundamental to the stability proof of the iterated observer.

Definition 7.5: For any given norm $\|\cdot\|$, define c_1^2 and c_2^2 from the Norm Equivalence Theorem 7.11 as

$$\|\cdot\| \leq \frac{1}{c_1^2} \|\cdot\|, \leq \frac{c_2^2}{c_1^2} \|\cdot\|. \quad (7.196)$$

where $\|\cdot\|_p$ is defined in Definition 5.7. □

The following Lemma establishes bounds on $\|\mathbf{CL}(k)\mathbf{v}(k)\|$ in the ideal observer O_2 . This result is provided primarily to illustrate a mechanism for obtaining these bounds since, in the stability proof, it is assumed that the region of convergence at each measurement event contains the open ball centred at the point of attraction with radius $\|\mathbf{CL}(k)\mathbf{v}(k)\|$.

Lemma 7.5: Under assumptions A1, A2, and A3, the ideal observer O_2 with initialization in the stability domain $\mathbf{X}_0$ of Theorem 5.11 (in the notation of that Theorem) satisfies

$$\|\mathbf{CL}(k)\mathbf{v}(k)\| \leq \|\mathbf{CB}(k)\|d_0 + L_f \|\mathbf{W}_k^{-1}\| \frac{M}{1 - \beta_s} + u(k) \|\mathbf{e}_\epsilon^*(0)\|. \quad (7.197)$$

where $\beta_s < 1$ and the sequence $\{\mathbf{W}_k \mid k = 0, 1, \dots\}$ satisfy the hypothesis of Theorem 5.11, and

$$M := \sup_{\{k \geq 0\}} \|\mathbf{W}_{k+1} \mathbf{F}(k) \mathbf{B}(k)\| d_0, \quad (7.198)$$

$$\mathbf{F}(k) := \{\mathbf{I}_n - \mathbf{L}(k) [\mathbf{C} \mathbf{L}(k)]^{-1} \mathbf{C}\}, \quad (7.199)$$

$$\alpha := \sup_{\{k \geq 0\}} \|\mathbf{W}_{k+1} \mathbf{L}(k) [\mathbf{C} \mathbf{L}(k)]^{-1}\|, \quad \text{and} \quad (7.200)$$

$$\begin{aligned} u(k) := & L_{\hat{\mathbf{f}}} \beta_s^k \|\mathbf{W}_k^{-1}\| \|\mathbf{W}_0\| \\ & + \frac{c_2^b}{c_1^b} \alpha L_{\mathbf{h}^{-1}}^2 L_{\hat{\mathbf{f}}} \left\{ \sum_{j=0}^{k-1} \beta_s^j \left[\prod_{i=0}^{k-j-1} L_{\mathbf{g}}(i) \right] \right\} \\ & + L_{\mathbf{h}^{-1}}^2 \frac{c_2^b}{c_1^b} \left[\prod_{i=0}^k L_{\mathbf{g}}(i) \right]. \end{aligned} \quad (7.201)$$

Proof: From the equivalence statement of Theorem 7.6, which showed that $\forall k \geq 0$, $\mathbf{v}^{(1)}(k) = \mathbf{v}^{(2)}(k)$, it is sufficient to show the result for $\mathbf{v}(k) \equiv \mathbf{v}^{(1)}(k)$. Multiplying $\mathbf{v}^{(1)}(k)$ defined in (7.78), page(230), on the left by $\mathbf{C} \mathbf{L}(k)$, dropping the superscript $(\cdot)^{(1)}$, and using Assumptions A1 and A2,

$$\begin{aligned} \|\mathbf{C} \mathbf{L}(k) \mathbf{v}(k)\| &= \|\mathbf{x}(k+1) - \mathbf{C} \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^*(k)) - \mathbf{I}_{\mathbf{h}}^{-1} [\mathbf{x}(k+1), \hat{\mathbf{x}}^*(k+1)] \mathbf{g}(k, \mathbf{s}_k^*)\| \\ &\leq \|\mathbf{C} \{\mathbf{x}_e(k+1) - \hat{\mathbf{f}}(k, \mathbf{x}_e^*(k))\}\| + L_{\mathbf{h}^{-1}} \|\mathbf{g}(k, \mathbf{s}_k^*)\| \\ &\leq \|\mathbf{C} \{\mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \mathbf{x}_e(k))\} + \mathbf{C} \{\hat{\mathbf{f}}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^*(k))\}\| \\ &\quad + L_{\mathbf{h}^{-1}} \|\mathbf{g}(k, \mathbf{s}_k^*)\| \\ &\leq \|\mathbf{C} \mathbf{B}(k)\| d_0 + L_{\hat{\mathbf{f}}} \|\mathbf{e}_e^*(k)\| + L_{\mathbf{h}^{-1}} \|\mathbf{g}(k, \mathbf{s}_k^*)\|, \end{aligned} \quad (7.202)$$

where

$$\mathbf{e}_e^*(k) := \mathbf{x}_e(k) - \hat{\mathbf{x}}_e^*(k). \quad (7.203)$$

Note that the fundamental Theorem of Integral Calculus has been used in the expression $\mathbf{C} \{\hat{\mathbf{f}}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^*(k))\}$ of second last inequality in (7.202), followed by the definition of constant $L_{\hat{\mathbf{f}}}$ in Assumption A3 to obtain the final inequality.

From Theorem 5.11 and with initialization in the stability domain $\mathbf{X}_0$, Equation 5.136 (page 161) gives the error bounds

$$\|\mathbf{e}_e^*(k)\| \leq \beta_s^k \|\mathbf{W}_k^{-1}\| \|\mathbf{W}_0 \mathbf{e}_e^*(0)\| + \|\mathbf{W}_k^{-1}\| M \sum_{i=0}^{k-1} \beta_s^i$$

$$+ \sum_{j=0}^{k-1} \beta_s^j M_1(k-1-j), \quad (7.204)$$

where, in the notation of Theorem 5.11, $\mathbf{K}(k) := \mathbf{L}(k)[\mathbf{CL}(k)]^{-1}$, and using Assumption A3,

$$\begin{aligned} M_1(k) &:= \|\mathbf{W}_{k+1}\mathbf{K}(k)\xi(k)\| \\ &\leq L_{h^{-1}}\|\mathbf{W}_{k+1}\mathbf{L}(k)[\mathbf{CL}(k)]^{-1}\mathbf{g}(k, \mathbf{s}_k^*)\| \\ &\leq \alpha L_{h^{-1}}\|\mathbf{g}(k, \mathbf{s}_k^*)\|. \end{aligned} \quad (7.205)$$

Also with Assumptions A1 and A2,

$$\begin{aligned} \|\mathbf{g}(k, \mathbf{s}_k^*)\| &\leq \frac{1}{c_1^2}\|\mathbf{g}(k, \mathbf{s}_k^*)\|, \\ &\leq \frac{1}{c_1^2}L_{\mathbf{g}}(k)\|\mathbf{s}_k^*\| \\ &= \frac{1}{c_1^2}L_{\mathbf{g}}(k)\|\mathbf{g}(k-1, \mathbf{s}_{k-1}^*)\| \\ &\vdots \\ &\leq \frac{c_2^2}{c_1^2}\left[\prod_{i=0}^k L_{\mathbf{g}}(i)\right]\|\mathbf{s}_0^*\| \\ &\leq \frac{c_2^2}{c_1^2}L_{h^{-1}}\left[\prod_{i=0}^k L_{\mathbf{g}}(i)\right]\|\mathbf{e}_e^*(0)\|. \end{aligned} \quad (7.206)$$

The last term of (7.204) can then be bounded by

$$\begin{aligned} \sum_{j=0}^{k-1} \beta_s^j M_1(k-1-j) &\leq \alpha L_{h^{-1}} \sum_{j=0}^{k-1} \beta_s^j \|\mathbf{g}(k-1-j, \mathbf{s}_{k-1-j}^*)\| \\ &\leq \frac{c_2^2}{c_1^2} \alpha L_{h^{-1}}^2 \left\{ \sum_{j=0}^{k-1} \beta_s^j \left[\prod_{i=0}^{k-1-j} L_{\mathbf{g}}(i) \right] \right\} \|\mathbf{e}_e^*(0)\|. \end{aligned} \quad (7.207)$$

Substitution of (7.206), (7.205), (7.204) and (7.207) into (7.202) gives

$$\begin{aligned} \|\mathbf{CL}(k)\mathbf{v}(k)\| &\leq \|\mathbf{CB}(k)\|d_0 + L_{\hat{\mathbf{f}}}\|\mathbf{W}_k^{-1}\|\frac{M}{1-\beta_s} \\ &\quad + L_{\hat{\mathbf{f}}}\beta_s^k\|\mathbf{W}_k^{-1}\|\|\mathbf{W}_0\|\|\mathbf{e}_e^*(0)\| \\ &\quad + \frac{c_2^2}{c_1^2}\alpha L_{h^{-1}}^2 L_{\hat{\mathbf{f}}} \left\{ \sum_{j=0}^{k-1} \beta_s^j \left[\prod_{i=0}^{k-1-j} L_{\mathbf{g}}(i) \right] \right\} \|\mathbf{e}_e^*(0)\| \\ &\quad + L_{h^{-1}}^2 \frac{c_2^2}{c_1^2} \left[\prod_{i=0}^k L_{\mathbf{g}}(i) \right] \|\mathbf{e}_e^*(0)\|. \end{aligned} \quad (7.208)$$

The desired result follows with the definition of $u(k)$ in (7.201). Note that under Assumption A2, $L_g(i) = 0$ for all $i > k_s$. In this case, if $k = k_s + p$, $p = 0, 1, 2, \dots$, then $\prod_{i=0}^{k-1-j} L_g(i) = 0$ for $j = 0, 1, 2, \dots, p-1$ and hence the summation in the second last term of (7.208) can be written as

$$\sum_{j=0}^{k-1} \beta_s^j \left[\prod_{i=0}^{k-1-j} L_g(i) \right] = \sum_{j=p}^{k_s+p-1} \beta_s^j \left[\prod_{i=0}^{k_s+p-1-j} L_g(i) \right], \quad (7.209)$$

and using the substitution $\ell = j - p$ in the summation index gives

$$\begin{aligned} \sum_{j=0}^{k-1} \beta_s^j \left[\prod_{i=0}^{k-1-j} L_g(i) \right] &= \sum_{\ell=0}^{k_s-1} \beta_s^{\ell+p} \left[\prod_{i=0}^{k_s-1-\ell} L_g(i) \right] \\ &= \beta_s^p \left\{ \sum_{\ell=0}^{k_s-1} \beta_s^{\ell} \left[\prod_{i=0}^{k_s-1-\ell} L_g(i) \right] \right\}, \end{aligned} \quad (7.210)$$

in which the term in braces is a constant sum of k_s finite terms. Note also that the last term in the expression for $u(k)$ in (7.201) is zero for $k \geq k_s$. $\square$

The proof of stability for the iterated observer, given below in Theorem 7.12, employs strong induction as the method of proof. In the strong induction proof, the desired result is first shown to hold for time $k = 0$ (induction initialization). The desired result is then *assumed* to hold for all $k \in \{0, 1, 2, \dots, k_0 - 1\}$ (strong induction hypothesis). Finally, based on the strong induction hypothesis, it is required to demonstrate that the desired result holds for all $k \in \{0, 1, 2, \dots, k_0 - 1, k_0\}$ (induction step). In this case, it follows that the desired result holds for all $k \in \mathbb{N}_0 := \{0, 1, 2, \dots\}$.

The following Lemma establishes two key results which are required in the induction step of Theorem 7.12. Suppose that for some given $\epsilon'_{x_e} > 0$, N_k local iterations have been performed at each measurement update $k \in \{0, 1, 2, \dots, k_0 - 1\}$ so that the distance from the ideal observer output to that of the iterated observer output for each k is less than ϵ'_{x_e} , i.e.

$$\|\hat{x}_e^*(k) - \hat{x}_e^{(N_k)}(k)\| < \epsilon'_{x_e}. \quad (7.211)$$

An assumption of this form is part of the strong induction hypothesis used in Theorem 7.12 below. The induction step of this theorem then requires that bounds be obtained on the distance from the ideal observer output to that of the iterated observer output for the j th iteration of the observer during the measurement update for time $k = k_0$. A simple form

for these bounds is provided in Conclusion C1 of Lemma 7.6 below. Note that for the purposes of Lemma 7.6, it is not assumed that the local iterations are strictly *convergent* at each past measurement update nor even at the present measurement update. It is only assumed that the local iterations through all past measurement updates yield solutions $\hat{\mathbf{x}}_e(k)$ that are within $\epsilon'_{\mathbf{x}_e}$ of the ideal observer output $\hat{\mathbf{x}}_e^*(k)$, where $\epsilon'_{\mathbf{x}_e}$ is some given positive tolerance. Conclusion C1 of Lemma 7.6 below draws extensively from the discrete-time stability methods of Chapter 5 and, in particular, the hypothesis and proof mechanism of Theorem 5.11.

The second result, Conclusion C2, established in Lemma 7.6 below provides a useful form for the distance from the initial iterate, $\hat{\mathbf{x}}^{(0)}(k_0)$, of the local iterations to the solution, $\hat{\mathbf{x}}^*(k_0)$, of the expression

$$\mathbf{z}(k_0) - \mathbf{h}(\hat{\mathbf{x}}^*(k_0)) = \mathbf{g}(k_0 - 1, \mathbf{s}_{k_0-1}^{(N_{k_0-1})}) \quad (7.212)$$

which represents the reaching law. The local iterations at time k_0 will converge if the initial iterate $\hat{\mathbf{x}}^{(0)}(k_0)$ lies within a region of attraction of the point $\hat{\mathbf{x}}^*(k_0)$. One of the tasks of the stability proof in Theorem 7.12 below is to show that the region of attraction at the k th measurement update need not be larger than $\|\mathbf{CL}(k-1)\mathbf{v}(k-1)\| + \epsilon_r$, where ϵ_r is any arbitrarily small positive tolerance. Conclusion C2 of Lemma 7.6 provides the mechanism for this step of the strong induction proof. Again, Conclusion C2 draws extensively from the discrete-time stability methods of Chapter 5 and, in particular, the hypothesis and proof mechanism of Theorem 5.11.

Lemma 7.6: Assume identical initialization for the ideal and iterated observers, i.e. $\hat{\mathbf{x}}_e^*(0) = \hat{\mathbf{x}}_e^{(N_0)}$, with $N_0 := 0$, and that Assumptions A1, A2, and A3 hold. In particular, assume that

H1: the ideal observer satisfies the hypothesis of Theorem 5.11 for some $\epsilon_2 > 0$, and

H2: for some $\epsilon'_{\mathbf{x}_e} > 0$ satisfying

$$\epsilon'_{\mathbf{x}_e} < \frac{\epsilon_2}{\sup_{\{k \geq 0\}} \|\mathbf{W}_k\|}. \quad (7.213)$$

and for each $k = 0, 1, \dots, k_0 - 1$, N_k local iterations are performed which yield

$$\|\hat{\mathbf{x}}_e^*(k) - \hat{\mathbf{x}}_e^{(N_k)}(k)\| < \epsilon'_{\mathbf{x}_e} \quad \forall k = 0, 1, \dots, k_0 - 1. \quad (7.214)$$

Then

C1: The ideal, $\hat{\mathbf{x}}_e^*(k_0)$, and iterated, $\hat{\mathbf{x}}_e^{(j)}(k_0)$, extended states at time $k = k_0$ following the j th iteration satisfy

$$\begin{aligned} \|\hat{\mathbf{x}}_e^*(k_0) - \hat{\mathbf{x}}_e^{(j)}(k_0)\| &\leq \lambda \|\hat{\mathbf{x}}^*(k_0) - \hat{\mathbf{x}}^{(j)}(k_0)\| \\ &\quad + \lambda \sum_{\ell=1}^{k_0-1} \beta^\ell \|\hat{\mathbf{x}}^*(k_0 - \ell) - \hat{\mathbf{x}}^{(N_{k_0-\ell})}(k_0 - \ell)\|, \end{aligned} \quad (7.215)$$

where $\beta < 1$ is as defined in Theorem 5.11, and λ is the finite constant defined in (7.193) as

$$\lambda := \sup_{k, j \leq k} \left\{ \|\mathbf{W}_k^{-1}\| \|\mathbf{W}_j \mathbf{L}(j) [\mathbf{CL}(j)]^{-1}\| \right\}. \quad (7.216)$$

C2: Recall that the sequence $\hat{\mathbf{x}}^b(k)$ defined in (7.181), page 250, satisfies

$$\mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}^b(k)) = \mathbf{g}(k-1, \mathbf{s}_{k-1}^{(N_{k-1})}). \quad (7.217)$$

With the initial iterate $\hat{\mathbf{x}}^{(0)}(k)$ defined as

$$\hat{\mathbf{x}}^{(0)}(k) := \mathbf{C}\hat{\mathbf{f}}(k-1, \hat{\mathbf{x}}_e^{(N_{k-1})}(k-1)), \quad (7.218)$$

Hypothesis **H1** and **H2** imply that

$$\begin{aligned} \|\hat{\mathbf{x}}^b(k_0) - \hat{\mathbf{x}}^{(0)}(k_0)\| &\leq \|\hat{\mathbf{x}}^b(k_0) - \hat{\mathbf{x}}^*(k_0)\| + L_{\hat{\mathbf{f}}} \|\hat{\mathbf{x}}_e^*(k_0-1) - \hat{\mathbf{x}}_e^{(N_{k_0-1})}(k_0-1)\| \\ &\quad + \|\mathbf{CL}(k_0-1)\mathbf{v}(k_0-1)\|. \end{aligned} \quad (7.219)$$

Proof: The ideal observer O_2 satisfies

$$\mathbf{x}_e^*(k) = \hat{\mathbf{f}}(k-1, \hat{\mathbf{x}}_e^*(k-1)) + \mathbf{L}(k-1)\mathbf{v}(k-1), \quad (7.220)$$

and hence

$$\hat{\mathbf{x}}^*(k) - \mathbf{C}\hat{\mathbf{f}}(k-1, \hat{\mathbf{x}}_e^*(k-1)) = \mathbf{CL}(k-1)\mathbf{v}(k-1). \quad (7.221)$$

Multiplying both sides by $\mathbf{L}(k-1)[\mathbf{CL}(k-1)]^{-1}$ and taking the right-hand side to the left gives

$$\mathbf{L}(k-1)\mathbf{v}(k-1) = \mathbf{L}(k-1)[\mathbf{CL}(k-1)]^{-1} \left\{ \hat{\mathbf{x}}^*(k) - \mathbf{C}\hat{\mathbf{f}}(k-1, \hat{\mathbf{x}}_e^*(k-1)) \right\}. \quad (7.222)$$

Similarly, the iterated observer satisfies

$$\mathbf{x}_e^{(j)}(k) = \hat{\mathbf{f}}(k-1, \hat{\mathbf{x}}_e^{(N_{k-1})}(k-1)) + \sum_{\ell=0}^{j-1} \mathbf{L}^{(\ell)}(k-1) \mathbf{v}^{(\ell)}(k-1), \quad (7.223)$$

and hence

$$\hat{\mathbf{x}}^{(j)}(k) - \mathbf{C}\hat{\mathbf{f}}(k-1, \hat{\mathbf{x}}_e^{(N_{k-1})}(k-1)) = \sum_{\ell=0}^{j-1} \mathbf{C}\mathbf{L}^{(\ell)}(k-1) \mathbf{v}^{(\ell)}(k-1). \quad (7.224)$$

Recall from (7.183), page 250, that $\mathbf{L}(k)$ and $\mathbf{L}^{(\ell)}(k)$ are related so that

$$\mathbf{L}(k)[\mathbf{C}\mathbf{L}(k)]^{-1} = \mathbf{L}^{(\ell)}(k)[\mathbf{C}\mathbf{L}^{(\ell)}(k)]^{-1}. \quad (7.225)$$

Multiplying the left-hand side of (7.224) on the left by $\mathbf{L}(k)[\mathbf{C}\mathbf{L}(k)]^{-1}$ leads to the simplification

$$\begin{aligned} & \mathbf{L}(k)[\mathbf{C}\mathbf{L}(k)]^{-1} \sum_{\ell=0}^{j-1} \mathbf{C}\mathbf{L}^{(\ell)}(k-1) \mathbf{v}^{(\ell)}(k-1) \\ &= \sum_{\ell=0}^{j-1} \{\mathbf{L}(k)[\mathbf{C}\mathbf{L}(k)]^{-1}\} \mathbf{C}\mathbf{L}^{(\ell)}(k-1) \mathbf{v}^{(\ell)}(k-1) \\ &= \sum_{\ell=0}^{j-1} \{\mathbf{L}^{(\ell)}(k)[\mathbf{C}\mathbf{L}^{(\ell)}(k)]^{-1}\} \mathbf{C}\mathbf{L}^{(\ell)}(k-1) \mathbf{v}^{(\ell)}(k-1) \\ &= \sum_{\ell=0}^{j-1} \mathbf{L}^{(\ell)} \mathbf{v}^{(\ell)}(k-1). \end{aligned} \quad (7.226)$$

Hence multiplying both sides of (7.224) by $\mathbf{L}(k)[\mathbf{C}\mathbf{L}(k)]^{-1}$ and bringing the right-hand side to the left gives

$$\begin{aligned} & \sum_{\ell=0}^{j-1} \mathbf{L}^{(\ell)}(k-1) \mathbf{v}^{(\ell)}(k-1) = \\ & \mathbf{L}(k-1)[\mathbf{C}\mathbf{L}(k-1)]^{-1} \left\{ \hat{\mathbf{x}}^{(j)}(k) - \mathbf{C}\hat{\mathbf{f}}(k-1, \hat{\mathbf{x}}_e^{(N_{k-1})}(k-1)) \right\}. \end{aligned} \quad (7.227)$$

Substituting (7.222) into (7.220), and (7.227) into (7.223) and subtracting the resulting expressions gives, for any $k = 1, 2, \dots, k_0$, and any $j = 1, 2, \dots, N_k$,

$$\begin{aligned} \hat{\mathbf{x}}_e^*(k) - \hat{\mathbf{x}}_e^{(j)}(k) &= \hat{\mathbf{f}}(k-1, \hat{\mathbf{x}}_e^*(k-1)) - \hat{\mathbf{f}}(k-1, \hat{\mathbf{x}}_e^{(N_{k-1})}(k-1)) \\ &+ \mathbf{L}(k-1)[\mathbf{C}\mathbf{L}(k-1)]^{-1} \left\{ \hat{\mathbf{x}}^*(k) - \mathbf{C}\hat{\mathbf{f}}(k-1, \hat{\mathbf{x}}_e^*(k-1)) \right\} \\ &- \mathbf{L}(k-1)[\mathbf{C}\mathbf{L}(k-1)]^{-1} \left\{ \hat{\mathbf{x}}^{(j)}(k) - \mathbf{C}\hat{\mathbf{f}}(k-1, \hat{\mathbf{x}}_e^{(N_{k-1})}(k-1)) \right\} \end{aligned}$$

$$\begin{aligned}
&= [\mathbf{I}_n - \mathbf{L}(k-1)[\mathbf{CL}(k-1)]^{-1}\mathbf{C}] \hat{\mathbf{f}}(k-1, \hat{\mathbf{x}}_e^*(k-1)) \\
&\quad - [\mathbf{I}_n - \mathbf{L}(k-1)[\mathbf{CL}(k-1)]^{-1}\mathbf{C}] \hat{\mathbf{f}}(k-1, \hat{\mathbf{x}}_e^{(N_{k-1})}(k-1)) \\
&\quad + \mathbf{L}(k-1)[\mathbf{CL}(k-1)]^{-1} \{\hat{\mathbf{x}}^*(k) - \hat{\mathbf{x}}^{(j)}(k)\} \\
&= \mathbf{F}(k-1) \{\hat{\mathbf{f}}(k-1, \hat{\mathbf{x}}_e^*(k-1)) - \hat{\mathbf{f}}(k-1, \hat{\mathbf{x}}_e^{(N_{k-1})}(k-1))\} \\
&\quad + \mathbf{L}(k-1)[\mathbf{CL}(k-1)]^{-1} \{\hat{\mathbf{x}}^*(k) - \hat{\mathbf{x}}^{(j)}(k)\}.
\end{aligned} \tag{7.228}$$

Define

$$\hat{V}_k := \begin{cases} \|\mathbf{W}_k [\hat{\mathbf{x}}_e^*(k) - \hat{\mathbf{x}}_e^{(j)}(k)]\|, & \text{for } k = k_0 \\ \|\mathbf{W}_k [\hat{\mathbf{x}}_e^*(k) - \hat{\mathbf{x}}_e^{(N_k)}(k)]\|, & \text{for } k < k_0. \end{cases} \tag{7.229}$$

and

$$\sigma_k := \begin{cases} \|\mathbf{W}_k \mathbf{L}(k-1)[\mathbf{CL}(k-1)]^{-1}\| \|\hat{\mathbf{x}}^*(k) - \hat{\mathbf{x}}^{(j)}(k)\|, & \text{for } k = k_0 \\ \|\mathbf{W}_k \mathbf{L}(k-1)[\mathbf{CL}(k-1)]^{-1}\| \|\hat{\mathbf{x}}^*(k) - \hat{\mathbf{x}}^{(N_k)}(k)\|, & \text{for } k < k_0 \end{cases} \tag{7.230}$$

Using the Fundamental Theorem of Integral Calculus in (7.228) gives, for any $k = 1, 2, \dots, k_0$,

$$\hat{V}_k \leq \int_0^1 \|\mathbf{W}_k \mathbf{F}(k-1) \mathbf{J}_{\hat{\mathbf{f}}}(k-1, \mathbf{w}_\mu(k-1)) \mathbf{W}_{k-1}^{-1}\| d\mu \hat{V}_{k-1} + \sigma_k. \tag{7.231}$$

where

$$\mathbf{w}_\mu(k-1) := \mu \hat{\mathbf{x}}_e^*(k-1) + (1-\mu) \hat{\mathbf{x}}_e^{(N_{k-1})}(k-1). \tag{7.232}$$

It is now required to show that for any $k \in \{1, 2, \dots, k_0 - 1\}$, $\mathbf{w}_\mu(k-1) \in \Omega$, where the open, convex set Ω is as defined in Theorem 5.11. In this case (7.231) can be relaxed further, for any $k = 1, 2, \dots, k_0$, as

$$\hat{V}_k \leq \beta \hat{V}_{k-1} + \sigma_k. \tag{7.233}$$

The mechanism for this step is provided by the parameter ϵ_2 included in Theorem 5.11. This step is completed by first noting that stability of the ideal observer under the hypothesis of Theorem 5.11 with $\epsilon_2 > 0$ implies that for all k , $\hat{\mathbf{x}}_e^*(k-1) \in X_{k-1}$ (see Equation 5.134, page 161) and hence

$$\|\mathbf{W}_{k-1}[\mathbf{x}_e(k-1) - \hat{\mathbf{x}}_e^*(k-1)]\| < d_s \tag{7.234}$$

holds for any $k \in \{1, 2, \dots, k_0 - 1\}$, where d_s is defined in (5.132) of Theorem 5.11.

With (7.234) and Hypothesis H2, for any $k \in \{1, 2, \dots, k_0 - 1\}$

$$\begin{aligned}
 \|\mathbf{W}_{k-1}[\mathbf{x}_e(k-1) - \hat{\mathbf{x}}_e^{(N_{k-1})}(k-1)]\| &\leq \|\mathbf{W}_{k-1}[\mathbf{x}_e(k-1) - \hat{\mathbf{x}}_e^*(k-1)]\| \\
 &\quad + \|\mathbf{W}_{k-1}[\hat{\mathbf{x}}_e^*(k-1) - \hat{\mathbf{x}}_e^{(N_{k-1})}(k-1)]\| \\
 &< d_s + \epsilon'_{\mathbf{x}_e} \sup_{\{k>0\}} \|\mathbf{W}_{k-1}\| \\
 &< d_s + \epsilon_2.
 \end{aligned} \tag{7.235}$$

Now recall from (5.133), page 161, of Theorem 5.11 that the set $X_s \in \Omega$ is defined (altering the notation slightly for the present context) as

$$X_s := \{\hat{\mathbf{x}}_e | \exists \mathbf{u} \in \mathbb{R}^n, \exists k \geq 0, V(k, \mathbf{u}) < d_s + \epsilon_2, \hat{\mathbf{x}}_e + \mathbf{e} \in X_s\}. \tag{7.236}$$

where $V(k, \mathbf{u}) := \|\mathbf{W}_k \mathbf{u}\|$. But then (7.235) implies that, for any $k \in \{1, 2, \dots, k_0 - 1\}$, $\hat{\mathbf{x}}_e^{(N_{k-1})}(k-1) \in X_s$ due to the available choice of $\mathbf{u} = \mathbf{x}_e(k-1) - \hat{\mathbf{x}}_e^{(N_{k-1})}(k-1)$, where $\mathbf{x}_e(k-1)$ is the true extended state and, therefore, $\hat{\mathbf{x}}_e^{(N_{k-1})}(k-1) + \mathbf{u} = \mathbf{x}_e(k-1) \in X_s$.

Hence, for any $k \in \{1, 2, \dots, k_0 - 1\}$, both $\hat{\mathbf{x}}_e^*(k-1)$ and $\hat{\mathbf{x}}_e^{(N_{k-1})}(k-1)$ are elements of the set X_s of Theorem 5.11. Although the set X_s itself need not be convex, $X_s \subseteq \Omega$ and hence both $\hat{\mathbf{x}}_e^*(k-1)$ and $\hat{\mathbf{x}}_e^{(N_{k-1})}(k-1)$ are elements of the convex set Ω . From this it follows that $\mathbf{w}_\mu(k-1) \in \Omega$ and hence, for any $k = 1, 2, \dots, k_0$, (7.233) holds, i.e.

$$\tilde{\mathbf{V}}_k \leq \beta \tilde{\mathbf{V}}_{k-1} + \sigma_k. \tag{7.237}$$

Now, following the proof of Lemma 5.1, page 137, for $k = k_0$,

$$\begin{aligned}
 \tilde{\mathbf{V}}_{k_0} &\leq \beta^{k_0} \tilde{\mathbf{V}}_0 + \sum_{\ell=0}^{k_0-1} \beta^\ell \sigma_{k_0-\ell} \\
 &= \sigma_{k_0} + \sum_{\ell=1}^{k_0-1} \beta^\ell \sigma_{k_0-\ell},
 \end{aligned} \tag{7.238}$$

where the last inequality results since $\tilde{\mathbf{V}}_0 = 0$ by the hypothesis. The desired result in C1 is then obtained, using the definition of λ in (7.216), as

$$\begin{aligned}
 \|\hat{\mathbf{x}}_e^*(k_0) - \hat{\mathbf{x}}_e^{(j)}(k_0)\| &\leq \|\mathbf{W}_{k_0}^{-1}\| \tilde{\mathbf{V}}_{k_0} \\
 &\leq \|\mathbf{W}_{k_0}^{-1}\| \|\mathbf{W}_{k_0} \mathbf{L}(k_0-1) [\mathbf{CL}(k_0-1)]^{-1}\| \|\hat{\mathbf{x}}^*(k_0) - \hat{\mathbf{x}}^{(j)}(k_0)\| \\
 &\quad + \|\mathbf{W}_{k_0}^{-1}\| \sum_{\ell=1}^{k_0-1} \left\{ \beta^\ell \|\mathbf{W}_{k_0-\ell} \mathbf{L}(k_0-\ell-1) [\mathbf{CL}(k_0-\ell-1)]^{-1}\| \right. \\
 &\quad \cdot \|\hat{\mathbf{x}}^*(k_0-\ell) - \hat{\mathbf{x}}^{(N_{k_0-\ell})}(k_0-\ell)\| \left. \right\}
 \end{aligned}$$

$$\begin{aligned} &\leq \lambda \|\hat{\mathbf{x}}^*(k_0) - \hat{\mathbf{x}}^{(j)}(k_0)\| \\ &\quad + \lambda \sum_{\ell=1}^{k_0-1} \beta^\ell \|\hat{\mathbf{x}}^*(k_0 - \ell) - \hat{\mathbf{x}}^{(N_{k_0-\ell})}(k_0 - \ell)\|. \end{aligned} \quad (7.239)$$

The result of C2 is derived by writing

$$\begin{aligned} \|\hat{\mathbf{x}}^b(k_0) - \hat{\mathbf{x}}^{(0)}(k_0)\| &\leq \|\hat{\mathbf{x}}^b(k_0) - \hat{\mathbf{x}}^*(k_0)\| + \|\hat{\mathbf{x}}^*(k_0) - \hat{\mathbf{x}}^{(0)}(k_0)\| \\ &\leq \|\hat{\mathbf{x}}^b(k_0) - \hat{\mathbf{x}}^*(k_0)\| \\ &\quad + \|\mathbf{C}\hat{\mathbf{f}}(k_0 - 1, \hat{\mathbf{x}}_e^*(k_0 - 1)) - \mathbf{C}\hat{\mathbf{f}}(k_0 - 1, \hat{\mathbf{x}}_e^{(N_{k_0-1})}(k_0 - 1))\| \\ &\quad + \|\mathbf{C}\mathbf{L}(k_0 - 1)\mathbf{v}(k_0 - 1)\| \\ &\leq \|\hat{\mathbf{x}}^b(k_0) - \hat{\mathbf{x}}^*(k_0)\| \\ &\quad + \int_0^1 \|\mathbf{C}\mathbf{J}_{\hat{\mathbf{f}}}(k_0 - 1, \mathbf{w}_\mu(k_0 - 1))\| d\mu \\ &\quad + \|\mathbf{C}\mathbf{L}(k_0 - 1)\mathbf{v}(k_0 - 1)\|, \end{aligned} \quad (7.240)$$

where $\mathbf{w}_\mu(k_0 - 1)$ is as defined in (7.232). However, it has already been shown that $\mathbf{w}_\mu(k_0 - 1) \in \Omega$ and hence, with $L_{\hat{\mathbf{f}}}$ defined in (7.194) of Assumption A3 as

$$\sup_{\{k \geq 0\}} \sup_{\{\mathbf{v} \in \Omega\}} \left\| \mathbf{C} \frac{\partial \hat{\mathbf{f}}}{\partial \hat{\mathbf{x}}_e}(k, \mathbf{v}) \right\| \leq L_{\hat{\mathbf{f}}} < \infty. \quad (7.241)$$

the result in (7.240) can be further approximated as

$$\begin{aligned} \|\hat{\mathbf{x}}^b(k_0) - \hat{\mathbf{x}}^{(0)}(k_0)\| &\leq \|\hat{\mathbf{x}}^b(k_0) - \hat{\mathbf{x}}^*(k_0)\| \\ &\quad + L_{\hat{\mathbf{f}}} \|\hat{\mathbf{x}}_e^*(k_0 - 1) - \hat{\mathbf{x}}_e^{(N_{k_0-1})}(k_0 - 1)\| \\ &\quad + \|\mathbf{C}\mathbf{L}(k_0 - 1)\mathbf{v}(k_0 - 1)\|. \end{aligned} \quad (7.242)$$

□

The proof of stability for the iterated observer, given below in Theorem 7.12, also draws from the following collection of technical results.

Lemma 7.7:

R1: Define $\mathbf{s}_0^{(N_0)} := \mathbf{s}_0^*$. For each $k = 1, 2, \dots, k_0 - 1$, assume that N_k iterations have been performed at the k th time step with

$$\mathbf{s}_k^* = \mathbf{g}(k - 1, \mathbf{s}_{k-1}^*) \quad (7.243)$$

$$\mathbf{s}_k^{(N_k)} = \mathbf{g}(k - 1, \mathbf{s}_{k-1}^{(N_{k-1})}) - \mathbf{r}_k^{(N_k)}. \quad (7.244)$$

Given also that the j th iteration at time step $k = k_0$ satisfies

$$\mathbf{s}_{k_0}^{(j)} = \mathbf{g}(k_0 - 1, \mathbf{s}_{k_0-1}^{(N_{k_0-1})}) - \mathbf{r}_k^{(j)} \quad (7.245)$$

then given any norm $\|\cdot\|$, and taking c_1^* and c_2^* from Definition 7.5, under Assumption A2,

$$\begin{aligned} \|\mathbf{g}(k_0, \mathbf{s}_{k_0}^{(j)}) - \mathbf{g}(k_0, \mathbf{s}_{k_0}^*)\| &\leq \frac{c_2^*}{c_1^*} L_{\mathbf{g}}(k_0) \|\mathbf{r}_{k_0}^{(j)}\| \\ &+ \frac{c_2^*}{c_1^*} \sum_{\ell=1}^{k_0-1} \left[\prod_{i=0}^{\ell} L_{\mathbf{g}}(k_0 - i) \right] \|\mathbf{r}_{k_0-\ell}^{(N_{k_0-\ell})}\|. \end{aligned} \quad (7.246)$$

R2: With Assumption A1 and the sequence $\hat{\mathbf{x}}^*(k)$ as defined in (7.181),

$$\|\hat{\mathbf{x}}^*(k+1) - \hat{\mathbf{x}}^*(k+1)\| \leq L_{\mathbf{h}^{-1}} \|\mathbf{g}(k, \mathbf{s}_k^{(N_k)}) - \mathbf{g}(k, \mathbf{s}_k^*)\|. \quad (7.247)$$

R3: With the sequence $\hat{\mathbf{x}}^*(k)$ as defined in (7.181),

$$\mathbf{r}_k^{(j)} = \mathbf{h}(\hat{\mathbf{x}}^{(j)}(k)) - \mathbf{h}(\hat{\mathbf{x}}^*(k)). \quad (7.248)$$

R4: From Assumption A2, since $\bar{L}_{\mathbf{g}} < 1$,

$$\sum_{\ell=0}^{k-1} \left[\prod_{i=0}^{\ell} L_{\mathbf{g}}(k-i) \right] \leq \sum_{\ell=0}^{\infty} (\bar{L}_{\mathbf{g}})^{\ell} = \frac{1}{1 - \bar{L}_{\mathbf{g}}}. \quad (7.249)$$

Proof: For Result R1,

$$\begin{aligned} \|\mathbf{g}(k_0, \mathbf{s}_{k_0}^{(j)}) - \mathbf{g}(k_0, \mathbf{s}_{k_0}^*)\| &\leq \frac{1}{c_1^*} L_{\mathbf{g}}(k_0) \|\mathbf{s}_{k_0}^{(j)} - \mathbf{s}_{k_0}^*\|, \\ &= \frac{1}{c_1^*} L_{\mathbf{g}}(k_0) \|\mathbf{g}(k_0 - 1, \mathbf{s}_{k_0-1}^{(N_{k_0-1})}) - \mathbf{g}(k_0 - 1, \mathbf{s}_{k_0-1}^*) - \mathbf{r}_{k_0}^{(j)}\|, \\ &\leq \frac{1}{c_1^*} L_{\mathbf{g}}(k_0) \|\mathbf{r}_{k_0}^{(j)}\| + \frac{1}{c_1^*} L_{\mathbf{g}}(k_0) L_{\mathbf{g}}(k_0 - 1) \|\mathbf{s}_{k_0-1}^{(N_{k_0-1})} - \mathbf{s}_{k_0-1}^*\|, \\ &\vdots \\ &\leq \frac{1}{c_1^*} L_{\mathbf{g}}(k_0) \|\mathbf{r}_{k_0}^{(j)}\|, \\ &\quad + \frac{1}{c_1^*} \sum_{\ell=1}^{k_0-1} \left[\prod_{i=0}^{\ell} L_{\mathbf{g}}(k_0 - i) \right] \|\mathbf{r}_{k_0-\ell}^{(N_{k_0-\ell})}\|, \\ &\quad + \frac{1}{c_1^*} \left[\prod_{i=0}^{k_0} L_{\mathbf{g}}(k_0 - i) \right] \|\mathbf{s}_0^{N_0} - \mathbf{s}_0^*\|. \end{aligned} \quad (7.250)$$

where the last term in (7.250) is zero by the hypothesis. Using c_2^b from Definition 7.5 gives the desired result.

For Result R2, recall that the sequence $\hat{\mathbf{x}}^b(k)$ satisfies, from (7.181) page 250,

$$\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}^b(k+1)) = \mathbf{g}(k, \mathbf{s}_k^{(N_k)}), \quad (7.251)$$

whereas $\hat{\mathbf{x}}^*(k)$ satisfies

$$\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}^*(k+1)) = \mathbf{g}(k, \mathbf{s}_k^*). \quad (7.252)$$

Subtracting (7.251) from (7.252), using the Fundamental Theorem of Integral Calculus, and applying Assumption A1 gives

$$\begin{aligned} & \|\hat{\mathbf{x}}^*(k+1) - \hat{\mathbf{x}}^b(k+1)\| \\ &= \left\| \mathcal{I}_{\mathbf{h}}^{-1}[\hat{\mathbf{x}}^*(k+1), \hat{\mathbf{x}}^b(k+1)] \left\{ \mathbf{h}(\hat{\mathbf{x}}^*(k+1)) - \mathbf{h}(\hat{\mathbf{x}}^b(k+1)) \right\} \right\| \\ &\leq L_{\mathbf{h}^{-1}} \|\mathbf{h}(\hat{\mathbf{x}}^*(k+1)) - \mathbf{h}(\hat{\mathbf{x}}^b(k+1))\| \\ &= L_{\mathbf{h}^{-1}} \|\mathbf{g}(k, \mathbf{s}_k^{(N_k)}) - \mathbf{g}(k, \mathbf{s}_k^*)\|. \end{aligned} \quad (7.253)$$

For Result R3, from the defining equations of the iterated observer (7.150), page 243,

$$\mathbf{r}^{(j)}(k) := \mathbf{h}(\hat{\mathbf{x}}^{(j)}(k)) - \mathbf{z}(k) + \mathbf{g}(k-1, \mathbf{s}_{k-1}^{(N_{k-1})}), \quad (7.254)$$

and from the definition of the sequence $\hat{\mathbf{x}}^b(k)$, from (7.181) page 250,

$$\mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}^b(k)) = \mathbf{g}(k-1, \mathbf{s}_{k-1}^{(N_{k-1})}). \quad (7.255)$$

Subtracting (7.254) from (7.255) gives

$$\mathbf{r}_k^{(j)} = \mathbf{h}(\hat{\mathbf{x}}^{(j)}(k)) - \mathbf{h}(\hat{\mathbf{x}}^b(k)). \quad (7.256)$$

Result R4 is immediate from summation of a geometric series. $\square$

The following Theorem provides the stability result for the iterated observer under Assumptions A1, A2, and A3, and drawing from the results of Lemmas 7.6 and 7.7. A fundamental hypothesis of this theorem is that the region of attraction, denoted by $\mathcal{R}(\mathbf{x}^b(k))$, at the k th measurement update contains the closed ball, $\bar{\mathcal{B}}(\hat{\mathbf{x}}^b(k), r_c(k))$, centred on the solution, $\hat{\mathbf{x}}^b(k)$, to the reaching law with radius $r_c(k) := \|\mathbf{CL}(k-1)\mathbf{v}(k-1)\| + \epsilon_r$, where ϵ_r is any arbitrarily small positive tolerance.

Theorem 7.12: Assume that Assumptions A1, A2, and A3 hold. Also assume the following initialization at time $k = 0$:

$$\hat{\mathbf{x}}_e^{(N_0)}(0) := \hat{\mathbf{x}}_e(0) = \hat{\mathbf{x}}_e^*(0), \quad (7.257)$$

$$\hat{\mathbf{x}}^b(0) := \hat{\mathbf{x}}^{(0)}(0) := \mathbf{C}\hat{\mathbf{x}}_e(0), \quad (7.258)$$

$$\mathbf{r}_0^{(0)} := \mathbf{0}, \quad (7.259)$$

$$\mathcal{R}(\hat{\mathbf{x}}^b(0)) := \mathbb{R}^m, \quad (7.260)$$

$$\mathbf{L}(-1) := \mathbf{0}, \quad \mathbf{v}(-1) := \mathbf{0}. \quad (7.261)$$

Then $\forall \epsilon_{\mathbf{x}_e} > 0$, and $\forall \epsilon_c > 0$, given that the region of attraction $\mathcal{R}(\hat{\mathbf{x}}^*(k))$ for each $k = 0, 1, 2, \dots$ satisfies

$$\bar{\mathcal{B}}(\hat{\mathbf{x}}^*(k), r_c(k)) \subset \mathcal{R}(\hat{\mathbf{x}}^*(k)), \quad (7.262)$$

where

$$r_c(k) := \|\mathbf{C}\mathbf{L}(k-1)\mathbf{v}(k-1)\| + \epsilon_c, \quad (7.263)$$

there exists a sequence of positive integers

$$\mathcal{N} := \{N_j \mid j = 0, 1, 2, \dots\}. \quad (7.264)$$

where each N_j can be fixed at time $k = j$ and thereafter held constant, such that

$$\|\hat{\mathbf{x}}_e^*(k) - \hat{\mathbf{x}}_e^{(N_k)}(k)\| < \epsilon_{\mathbf{x}_e}, \quad k = 0, 1, 2, \dots \quad (7.265)$$

Proof: The proof is by strong induction. Fix $\epsilon_{\mathbf{x}_e} > 0$ and $\epsilon_c > 0$. Define

$$\mathcal{N}_k := \{N_j \mid j = 0, 1, \dots, k\}. \quad (7.266)$$

With $\epsilon_2 > 0$ and the sequence of matrices $\mathbf{W}_k$ satisfying the hypothesis of Theorem 5.11, set

$$0 < \epsilon'_{\mathbf{x}_e} < \min \left\{ \epsilon_{\mathbf{x}_e}, \frac{\epsilon_c}{2L_{\hat{\mathbf{f}}}}, \frac{\epsilon_2}{\sup_{\{k \geq 0\}} \|\mathbf{W}_k\|} \right\}. \quad (7.267)$$

$$0 < \epsilon_{b^*} < \min \left\{ \frac{1-\beta}{\lambda} \epsilon'_{\mathbf{x}_e}, \frac{\epsilon_c}{2} \right\}. \quad (7.268)$$

$$\epsilon_b := \frac{1-\beta}{\lambda} \epsilon'_{\mathbf{x}_e} - \epsilon_{b^*}. \quad (7.269)$$

$$0 < \epsilon_r < \min \left\{ \frac{\epsilon_b}{L_{\mathbf{h}^{-1}}}, \frac{c_1^2(1-\bar{L}_{\mathbf{g}})}{c_2^2 L_{\mathbf{h}^{-1}}} \epsilon_{b^*} \right\}. \quad (7.270)$$

Note in (7.267) that $\epsilon'_{\mathbf{x}_e}$ satisfies

$$\epsilon_2 > \epsilon'_{\mathbf{x}_e} \sup_{\{k \geq 0\}} \|\mathbf{W}_k\|, \quad (7.271)$$

which is required in Hypothesis **H1** of Lemma 7.6, and also that $\epsilon_b > 0$ due to the definition of ϵ_{b^*} in (7.268). The proof in fact shows that $\|\hat{\mathbf{x}}_e^*(k) - \hat{\mathbf{x}}_e^{(N_k)}(k)\| < \epsilon'_{\mathbf{x}_e} < \epsilon_{\mathbf{x}_e}$.

Induction Initialization: At time $k = 0$ from the initialization assumptions, set

$$\mathcal{N}_0 := \{N_0 = 0\} \quad (7.272)$$

to obtain

$$\|\mathbf{r}_0^{(N_0)}\| = 0 < \epsilon_r, \quad (7.273)$$

$$\|\dot{\mathbf{x}}^{(N_0)}(0) - \dot{\mathbf{x}}^b(0)\| = 0 < \epsilon_b, \quad (7.274)$$

$$\|\hat{\mathbf{x}}_e^*(0) - \hat{\mathbf{x}}_e^{(N_0)}(0)\| = 0 < \epsilon'_{\mathbf{x}_e} < \epsilon_{\mathbf{x}_e}. \quad (7.275)$$

With **A1**, and Results **R1** and **R2** of Lemma 7.7,

$$\begin{aligned} \|\hat{\mathbf{x}}^*(1) - \hat{\mathbf{x}}^b(1)\| &\leq L_{\mathbf{h}^{-1}} \|\mathbf{g}(0, \mathbf{s}_0^{(N_0)}) - \mathbf{g}(0, \mathbf{s}_0^*)\| \\ &\leq \frac{c_2^b}{c_1^b} L_{\mathbf{g}}(0) \|\mathbf{r}_0^{(0)}\| \\ &= 0 < \epsilon_{b^*}. \end{aligned} \quad (7.276)$$

Finally, with Conclusion **C2** of Lemma 7.6,

$$\begin{aligned} \|\hat{\mathbf{x}}^{(0)}(1) - \hat{\mathbf{x}}^b(1)\| &\leq \|\hat{\mathbf{x}}^b(1) - \hat{\mathbf{x}}^*(1)\| + L_{\hat{\mathbf{f}}} \|\hat{\mathbf{x}}_e^*(0) - \hat{\mathbf{x}}_e^{(0)}(0)\| + \|\mathbf{CL}(0)\mathbf{v}(0)\| \\ &= \|\mathbf{CL}(0)\mathbf{v}(0)\| < r_c(1). \end{aligned} \quad (7.277)$$

Strong Induction Hypothesis: Assume that, for some $k_0 > 0$,

$$\mathcal{N}_{k_0-1} := \{N_j \mid j = 0, 1, 2, \dots, k_0 - 1\} \quad (7.278)$$

have been selected such that

$$\|\mathbf{r}_k^{(N_k)}\| < \epsilon_r, \quad \forall k = 0, 1, \dots, k_0 - 1, \quad (7.279)$$

$$\|\hat{\mathbf{x}}^{(N_k)}(k) - \hat{\mathbf{x}}^b(k)\| < \epsilon_b, \quad \forall k = 0, 1, \dots, k_0 - 1. \quad (7.280)$$

$$\|\hat{\mathbf{x}}_e^*(k) - \hat{\mathbf{x}}_e^{(N_k)}(k)\| < \epsilon'_{\mathbf{x}_e}, \quad \forall k = 0, 1, \dots, k_0 - 1. \quad (7.281)$$

$$\|\hat{\mathbf{x}}^b(k) - \hat{\mathbf{x}}^*(k)\| < \epsilon_{b^*}, \quad \forall k = 0, 1, \dots, k_0, \quad (7.282)$$

$$\|\hat{\mathbf{x}}^{(0)}(k) - \hat{\mathbf{x}}^b(k)\| < r_c(k), \quad \forall k = 0, 1, \dots, k_0, \quad (7.283)$$

Note that (7.279) through (7.283) are satisfied for $k_0 = 1$ by the Induction Initialization.

Induction Step: It is required to show that the strong induction hypothesis (7.279) through (7.283) are sufficient to produce $N_{k_0} > 0$, thereby extending the set $\mathcal{N}_{k_0-1}$ to $\mathcal{N}_{k_0}$, such that (7.279) through (7.283) are extended to time $k_0 + 1$. The induction step proceeds as follows:

1. From (7.283), the initial iterate $\hat{\mathbf{x}}^{(0)}(k_0)$ lies in the region of attraction for $\hat{\mathbf{x}}^*(k_0)$ and hence

$$\lim_{j \rightarrow \infty} \hat{\mathbf{x}}^{(j)}(k_0) = \hat{\mathbf{x}}^*(k_0). \quad (7.284)$$

2. From the continuity of $\mathbf{h}$, Result R3 of Lemma 7.7, and using (7.284),

$$\lim_{j \rightarrow \infty} \mathbf{r}_{k_0}^{(j)} = \lim_{j \rightarrow \infty} \{ \mathbf{h}(\hat{\mathbf{x}}^{(j)}(k_0)) - \mathbf{h}(\hat{\mathbf{x}}^*(k_0)) \} = 0. \quad (7.285)$$

Moreover, $\mathbf{r}_{k_0}^{(j)}$ can be computed from (7.150), page 243, at each iteration. Choose N_{k_0} , for which existence is guaranteed by (7.285), such that

$$\|\mathbf{r}_{k_0}^{(N_{k_0})}\| < \epsilon_r. \quad (7.286)$$

3. With Assumption A1, Result R3 of Lemma 7.7, results (7.285) and (7.286) above, and with the definition of ϵ_b in (7.270),

$$\begin{aligned} \|\hat{\mathbf{x}}^{(N_{k_0})}(k_0) - \hat{\mathbf{x}}^*(k_0)\| &\leq L_{\mathbf{h}^{-1}} \|\mathbf{r}_{k_0}^{(N_{k_0})}\| \\ &\leq L_{\mathbf{h}^{-1}} \epsilon_r \\ &< L_{\mathbf{h}^{-1}} \frac{\epsilon_b}{L_{\mathbf{h}^{-1}}} = \epsilon_b. \end{aligned} \quad (7.287)$$

4. From Conclusion C1 of Lemma 7.6, using (7.280), (7.282), and (7.287), and with the definitions of ϵ_b in (7.269) and $\epsilon'_{\mathbf{x}_e}$ in (7.267),

$$\begin{aligned} \|\hat{\mathbf{x}}_e^*(k_0) - \hat{\mathbf{x}}_e^{(N_{k_0})}(k_0)\| &\leq \lambda \|\hat{\mathbf{x}}^*(k_0) - \hat{\mathbf{x}}^{(N_{k_0})}(k_0)\| \\ &\quad + \lambda \sum_{\ell=1}^{k_0-1} \beta^\ell \|\hat{\mathbf{x}}^*(k_0 - \ell) - \hat{\mathbf{x}}^{(N_{k_0}-\ell)}(k_0 - \ell)\| \\ &\leq \lambda \sum_{\ell=0}^{k_0-1} \beta^\ell \|\hat{\mathbf{x}}^*(k_0 - \ell) - \hat{\mathbf{x}}^*(k_0 - \ell)\| \end{aligned}$$

$$\begin{aligned}
& + \lambda \sum_{\ell=0}^{k_0-1} \beta^\ell + \|\hat{\mathbf{x}}^b(k_0 - \ell) - \hat{\mathbf{x}}^{(N_{k_0-\ell})}(k_0 - \ell)\| \\
& < \frac{\lambda}{1-\beta} \{\epsilon_{b^*} + \epsilon_b\} \\
& = \frac{\lambda}{1-\beta} \left\{ \epsilon_{b^*} + \frac{1-\beta}{\lambda} \epsilon'_{\mathbf{x}_e} - \epsilon_{b^*} \right\} \\
& = \epsilon'_{\mathbf{x}_e} < \epsilon_{\mathbf{x}_e}.
\end{aligned} \tag{7.288}$$

5. Starting with Results **R1**, **R2**, and **R4** of Lemma 7.7(7.247), and using the definition of ϵ_r in (7.270), and the results in (7.279) and (7.286),

$$\begin{aligned}
\|\hat{\mathbf{x}}^b(k_0 + 1) - \hat{\mathbf{x}}^*(k_0 + 1)\| & \leq L_{\mathbf{h}^{-1}} \|\mathbf{g}(k_0, \mathbf{s}_{k_0}^{(N_{k_0})}) - \mathbf{g}(k_0, \mathbf{s}_{k_0}^*)\| \\
& \leq L_{\mathbf{h}^{-1}} \frac{c_2^b}{c_1^b} \sum_{\ell=0}^{k_0-1} \left[\prod_{i=0}^{\ell} L_{\mathbf{g}}(k_0 - i) \right] \|\mathbf{r}_{k_0-\ell}^{(N_{k_0-\ell})}\| \\
& \leq \frac{L_{\mathbf{h}^{-1}} c_2^b}{c_1^b (1 - \bar{L}_{\mathbf{g}})} \epsilon_r \\
& < \left[\frac{L_{\mathbf{h}^{-1}} c_2^b}{c_1^b (1 - \bar{L}_{\mathbf{g}})} \right] \left[\frac{c_1^b (1 - \bar{L}_{\mathbf{g}})}{L_{\mathbf{h}^{-1}} c_2^b} \epsilon_{b^*} \right] = \epsilon_{b^*}.
\end{aligned} \tag{7.289}$$

6. Beginning with Conclusion **C2** of Lemma 7.6, using (7.288), and with the definitions of $\epsilon'_{\mathbf{x}_e}$ in (7.267) and ϵ_{b^*} in (7.268), and noting that Step 4 above has extended Hypothesis **H1** of Lemma 7.6 to time $k_0 + 1$,

$$\begin{aligned}
\|\hat{\mathbf{x}}^b(k_0 + 1) - \hat{\mathbf{x}}^{(0)}(k_0 + 1)\| & \leq \|\hat{\mathbf{x}}^b(k_0 + 1) - \hat{\mathbf{x}}^*(k_0 + 1)\| \\
& \quad + L_{\hat{\mathbf{f}}} \|\hat{\mathbf{x}}^*(k_0) - \hat{\mathbf{x}}^{(N_{k_0})}(k_0)\| + \|\mathbf{CL}(k_0)\mathbf{v}(k_0)\| \\
& \leq \epsilon_{b^*} + L_{\hat{\mathbf{f}}} \epsilon'_{\mathbf{x}_e} + \|\mathbf{CL}(k_0)\mathbf{v}(k_0)\| \\
& < \frac{\epsilon_c}{2} + L_{\hat{\mathbf{f}}} \frac{\epsilon_c}{2L_{\hat{\mathbf{f}}}} + \|\mathbf{CL}(k_0)\mathbf{v}(k_0)\| \\
& = \epsilon_c + \|\mathbf{CL}(k_0)\mathbf{v}(k_0)\| := r_c(k_0 + 1).
\end{aligned} \tag{7.290}$$

7. Set

$$\mathcal{N}_{k_0} := \{N_j \mid j = 0, 1, 2, \dots, k_0\}. \tag{7.291}$$

The results in (7.286) through (7.291) extend the strong induction hypothesis to time $k_0 + 1$ which completes the proof. $\square$

The primary result of the above theorem is that, provided the region of convergence of the local iterations is sufficiently large according to hypothesis (7.262) and (7.263), then it is possible to monitor the quantity $\|\mathbf{r}_k^{(j)}\|$ and terminate the local iterations for $j = N_k$ when $\|\mathbf{r}_k^{(j)}\| < \epsilon_r$ with the result that $\|\hat{\mathbf{x}}_e^*(k) - \hat{\mathbf{x}}_e^{(N_k)}(k)\| < \epsilon_{\mathbf{x}_e}$. This implies sequential processing of the measurements at each time update without having to return to past measurement events in order to improve the state estimate through additional iterations. Further discussion of implementation issues are included with simulation results in Section 7.5.

7.4 Redundant Nonlinear Measurement Models

In Section 7.3 the measurements were assumed to be nonredundant in the sense of definitions proposed in Section 7.2. In that case the Jacobian matrix, $\mathbf{J}_h$, of the measurement model, $\mathbf{h}(\mathbf{x})$, had linearly independent rows, and hence a matrix $\mathbf{L}$ can be found such that the product $\mathbf{J}_h \mathbf{C} \mathbf{L}$ is nonsingular. In this Section the restriction that measurements be nonredundant is removed in order to consider sliding mode observer design for the case in which the Jacobian matrix $\mathbf{J}_h$ has linearly dependent rows which occurs when redundant measurements are provided by the system.

The true dynamics are again given by

$$\mathbf{x}_e(k+1) = \mathbf{f}(k, \mathbf{x}_e(k)), \quad (7.292)$$

where $\mathbf{x}_e$ is the extended state vector and $\mathbf{f} : \mathbb{N}_0 \times \mathbf{X}_e \mapsto \mathbf{X}_e$. The redundant measurement model is written as

$$\mathbf{z}(k) = \mathbf{h}_e(\mathbf{x}_e(k)) = \mathbf{h}(\mathbf{C}\mathbf{x}_e(k)) = \mathbf{h}(\mathbf{x}(k)), \quad (7.293)$$

where $\mathbf{h} : \mathbb{R}^q \mapsto \mathbb{R}^m$ satisfies $\text{rank}\{\mathbf{J}_h(\mathbf{x})\} = r$, with $r = q < m$, and $\mathbf{C} \in \mathbb{R}^{q \times n}$ again has the form

$$\mathbf{C} := [\mathbf{I}_q, \mathbf{0}_{q \times (n-q)}]. \quad (7.294)$$

The proposed approach for redundant measurements follows very closely those given in Section 6.5 for linear redundant measurements coupled with those of Section 7.3 for nonlinear non-redundant measurement models. The primary objective of this Section is to

develop the proposed method, illustrate theoretical forms and properties of *ideal* observers, introduce and discuss implementation methods through local iterations, and highlight current limitations of analysis methods available at this time.

The Jacobian matrices $\mathbf{J}_{h_e}(\mathbf{x}_e)$ and $\mathbf{J}_h(\mathbf{x})$ satisfy

$$\mathbf{J}_{h_e}(\mathbf{x}_e) = \mathbf{J}_h(\mathbf{x})\mathbf{C}. \quad (7.295)$$

In this Section, the non-redundant measurement model is assumed to satisfy $\text{rank}\{\mathbf{J}_{h_e}\} \equiv q < m < n$, and hence

$$\text{rank}\{\mathbf{J}_{h_e}(\mathbf{x}_e)\} = \text{rank}\{\mathbf{J}_h(\mathbf{x})\} = q. \quad (7.296)$$

This implies that $\mathbf{J}_h(\mathbf{x})$ has linearly independent columns, but has linearly dependent rows. As a result, for any $\mathbf{x}$, there does not exist, from Theorem 6.1, a matrix $\mathbf{L}$ such that the product $\mathbf{J}_h(\mathbf{x})\mathbf{C}\mathbf{L}$ is nonsingular.

Instead, again consider the cost function defined on $\mathbb{R}^q$ as

$$C(\hat{\mathbf{x}}(k)) := [\mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}(k))]^T \mathbf{P}_k [\mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}(k))]. \quad (7.297)$$

where $\mathbf{P}_k \in \mathbb{R}^{m \times m}$ is positive definite symmetric. The sliding surface is defined as

$$\begin{aligned} \mathbf{s}_k &:= -\frac{1}{2} \left[\frac{\partial C}{\partial \hat{\mathbf{x}}} \right]^T \\ &= \mathbf{J}_h^T(\hat{\mathbf{x}}(k)) \mathbf{P}_k [\mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}(k))] = \mathbf{0}, \end{aligned} \quad (7.298)$$

in which case, during the sliding motion, $\hat{\mathbf{x}}(k)$ is a least-squares solution of the problem $\mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}(k)) = \mathbf{0}$ (see Ortega and Rheinboldt [159, Sec. 4.1] for a detailed treatment of the nonlinear least-squares concept).

7.4.1 Assumptions and Preliminary Results

Developments of proposed observers for the redundant measurement case proceed very much like those of Chapter 6 through the use of the Fundamental Theorem of Integral Calculus (Theorem 5.9, page 148). This Section focusses on two primary assumptions and establishes some preliminary results.

In Section 7.3.1, Assumption A1 (page 224) was introduced concerning nonsingularity of the matrix integral $\mathcal{I}_h[\mathbf{x}_1, \mathbf{x}_2]$. In the present Section, $\mathbf{J}_h : \mathbb{R}^q \mapsto \mathbb{R}^{m \times q}$, $q < m$, is assumed to be uniformly of rank q , and $\mathcal{I}_h : \mathbb{R}^q \times \mathbb{R}^q \mapsto \mathbb{R}^{m \times q}$. The following assumption extends the notions of Assumption A1 to the redundant measurement case.

A4 : $\mathbf{h} : \mathcal{A}_0 \subseteq \mathbb{R}^q \mapsto \mathcal{B}_0 \subseteq \mathbb{R}^m$, is C^p , $p \geq 1$, from $\mathcal{A}_0$ into $\mathcal{B}_0$, and

1. $\forall \mathbf{x} \in \mathcal{A}_0$, $\text{rank}\{\mathbf{J}_h(\mathbf{x})\} = q$, i.e. $\mathbf{J}_h$ has linearly independent columns,
2. $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{A} \subseteq \mathcal{A}_0$, and with $\mathbf{P} \in \mathbb{R}^{m \times m}$ positive definite symmetric, the product $\mathbf{J}_h^T(\mathbf{x}_2)\mathbf{P}\mathcal{I}_h(\mathbf{x}_1, \mathbf{x}_2)$ is nonsingular and

$$\| [\mathbf{J}_h^T(\mathbf{x}_2)\mathbf{P}\mathcal{I}_h(\mathbf{x}_1, \mathbf{x}_2)]^{-1} \| \leq L < \infty. \quad (7.299)$$

It is also assumed, at this time, that the sets $\mathcal{A}$, $\mathcal{A}_0$, and $\mathcal{B}_0$ are sufficiently large. Future analysis with the sets of Theorem 5.11 may lead to a more precise characterization of the above sets. Assumption A4 is employed in theoretical analysis of the ideal observer forms; proposed methods for implementation require Item 1 of A4.

Item 1 of Assumption A4 implies that $\forall \mathbf{x} \in \mathcal{A}_0$, the product $\mathbf{J}_h^T(\mathbf{x}_2)\mathbf{P}\mathbf{J}_h(\mathbf{x}_2)$ is nonsingular. Item 2 of Assumption A4 is a sufficient condition for the zero dynamics given by

$$\begin{aligned} \mathbf{0} = \mathbf{s}(k, \hat{\mathbf{x}}(k)) &:= \mathbf{J}_h^T(\hat{\mathbf{x}}(k))\mathbf{P}_k[\mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}(k))] \\ &= \mathbf{J}_h^T(\mathbf{x}(k))\mathbf{P}_k\mathcal{I}_h[\mathbf{x}(k), \hat{\mathbf{x}}(k)]\mathbf{e}(k) \end{aligned} \quad (7.300)$$

to have only a trivial solution $\mathbf{e}(k) := \mathbf{x}(k) - \hat{\mathbf{x}}(k) = \mathbf{0}$. If Assumption A4 is not satisfied, then the zero dynamics would have to be checked for stability.

Results similar to those shown in Theorems 7.3 and 7.4 can also be stated for sufficiency of Item 2 of A4, as illustrated in the following Theorem which follows the development of Theorem 7.5.

Theorem 7.13: Let $\mathbf{G} : \mathcal{D} \subseteq \mathbb{R}^q \mapsto \mathbb{R}^m$, for some open convex set $\mathcal{D}$, be C^p , $p \geq 1$. Let $\|\cdot\|$ and $\|\cdot\|$ be any compatible vector and matrix norms, respectively. Suppose that $\mathbf{P} \in \mathbb{R}^{m \times m}$ is positive definite symmetric and that

1. $\forall \mathbf{x}_2 \in \mathcal{D}$, $\mathbf{J}_G^T(\mathbf{x}_2)\mathbf{P}\mathbf{J}_G(\mathbf{x}_2)$ is nonsingular and

$$\left\| \left[\mathbf{J}_G^T(\mathbf{x}_2)\mathbf{P}\mathbf{J}_G(\mathbf{x}_2) \right]^{-1} \right\| < a. \quad (7.301)$$

2. $\sup_{\mathbf{x}_2 \in \mathcal{D}} \{ \|\mathbf{J}_G(\mathbf{x}_2)\mathbf{P}\| \} \leq b < \infty$,

3. $\mathbf{J}_G(\mathbf{x}_2)$ is Hölder continuous on $\mathcal{D}$, i.e. $\exists c, r$ such that $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$,

$$\|\mathbf{J}_G(\mathbf{x}_1) - \mathbf{J}_G(\mathbf{x}_2)\| \leq c\|\mathbf{x}_1 - \mathbf{x}_2\|^r. \quad (7.302)$$

Then for any $\mathbf{x}_2 \in \mathcal{D}$, the product $\mathbf{J}_G^T(\mathbf{x}_2)\mathbf{P}\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2]$ is nonsingular for all $\mathbf{x}_1 \in \mathcal{D}$ satisfying $\|\mathbf{x}_1 - \mathbf{x}_2\| < \delta$, where

$$\delta < \left[\frac{r+1}{abc} \right]^{1/r}. \quad (7.303)$$

Moreover,

$$\left\| \left[\mathbf{J}_G^T(\mathbf{x}_2)\mathbf{P}\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2] \right]^{-1} \right\| \leq \frac{a}{1 - \frac{abc\delta^r}{r+1}}. \quad (7.304)$$

Proof: Fix $\mathbf{x}_2 \in \mathcal{D}$. Define, for notational convenience

$$\mathbf{N}(\mathbf{x}_2) := \mathbf{J}_G^T(\mathbf{x}_2)\mathbf{P} \quad (7.305)$$

$$\mathbf{M}(\mathbf{x}_2) := \mathbf{J}_G^T(\mathbf{x}_2)\mathbf{P}\mathbf{J}_G(\mathbf{x}_2). \quad (7.306)$$

The proof follows that of Theorem 7.5. For any $\mathbf{x}_1 \in B(\mathbf{x}_2, \delta) \cap \mathcal{D}$,

$$\begin{aligned} & \left\| \mathbf{N}(\mathbf{x}_2)\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2]\mathbf{M}^{-1}(\mathbf{x}_2) - \mathbf{I}_m \right\| \\ &= \left\| [\mathbf{N}(\mathbf{x}_2)\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2] - \mathbf{M}(\mathbf{x}_2)] \mathbf{M}^{-1}(\mathbf{x}_2) \right\| \\ &\leq a \int_0^1 \left\| \mathbf{N}(\mathbf{x}_2) \{ \mathbf{J}_G(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) - \mathbf{J}_G(\mathbf{x}_2) \} \right\| d\lambda \\ &\leq abc \int_0^1 \left\| \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 - \mathbf{x}_2 \right\|^r d\lambda \\ &= abc \int_0^1 \lambda^r d\lambda \|\mathbf{x}_1 - \mathbf{x}_2\|^r \\ &\leq \frac{abc\delta^r}{1+r} < 1. \end{aligned} \quad (7.307)$$

Hence, by Lemma 7.2, the product $\mathbf{N}(\mathbf{x}_2)\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2]\mathbf{M}^{-1}(\mathbf{x}_2)$ is nonsingular which, in turn, implies the nonsingularity of $\mathbf{N}(\mathbf{x}_2)\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2]$. Also from Lemma 7.2,

$$\left\| \left[\mathbf{N}(\mathbf{x}_2)\mathcal{I}_G[\mathbf{x}_1, \mathbf{x}_2] \right]^{-1} \right\| \leq \frac{1}{1 - \frac{abc\delta^r}{r+1}}. \quad (7.308)$$

and the bounds in (7.304) are obtained with

$$\begin{aligned} \left\| [N(x_2) \mathcal{I}_G[x_1, x_2]]^{-1} \right\| &= \left\| M^{-1}(x_2) [N(x_2) \mathcal{I}_G[x_1, x_2] M^{-1}(x_2)]^{-1} \right\| \\ &\leq \frac{a}{1 - \frac{abcd^r}{r+1}}. \end{aligned} \quad (7.309)$$

□

In Section 7.3, the measured state to measurements map $h(\cdot)$ was assumed to be 1-to-1 which led, in Theorem 7.6, to equivalence of the two ideal observer forms. In Section 7.4.2 below, two ideal observer forms are presented for the redundant measurement model case. Assumption A4 is sufficient to obtain equivalence and stability of the ideal forms during the sliding mode. At this time, the following additional assumption is employed to obtain equivalence and stability of both the reaching and sliding modes of the two observers.

A5 : The function $s : N_0 \times \mathcal{A}_0 \mapsto \mathcal{C}_0 \subset \mathbb{R}^q$ defined as

$$s(k, \hat{x}) := J_h^T(\hat{x}) P_k [z(k) - h(\hat{x})] \quad (7.310)$$

for each fixed $k \geq 0$ is 1-to-1 from $\mathcal{A}_0$ onto $\mathcal{C}_0$.

A fundamental consequence of Assumption A5 and the use of the discrete-time reaching law method is that given any initial point $\hat{x}(0)$, and given the sequence of measurements $\{z(k) \mid k = 0, 1, \dots, k_s\}$, where pure sliding is initiated at time $k = k_s + 1$, there exists one, and only one, path $\{\hat{x}(k) \mid k = 0, 1, \dots, k_s\}$ during the reaching phase. If Assumption A5 is not satisfied, then there could exist multiple paths satisfying the reaching law from the initial point to the sliding surface. Even in this case, however, Assumption A4 guarantees that upon reaching the sliding surface, $e(k) = 0$ for all $k > k_s$ and stability can be obtained provided $e_s(k_s)$ lies within a stability domain of the sliding dynamics. Further consequences of these assumptions are discussed in the following section which details the observer forms, stability, and properties.

7.4.2 Theoretical Form, Stability, and Properties

For system (7.292) with measurement model (7.293), the proposed observers again have the form

$$\hat{x}_e(k+1) = f(k, \hat{x}_e(k)) + L(k)v(k). \quad (7.311)$$

where $\hat{\mathbf{f}}$ is an estimate of $\mathbf{f}$, and the observer gain matrices, $\mathbf{L}(k)$, and the function $\mathbf{v}(k)$ are to be determined. The sliding surface is defined, with $\mathbf{x} = \mathbf{C}\mathbf{x}_r$ and $\hat{\mathbf{x}} = \mathbf{C}\hat{\mathbf{x}}_r$, as

$$\begin{aligned} \mathbf{s}_k &:= \mathbf{s}(\mathbf{x}(k), \hat{\mathbf{x}}(k)) &:= \mathbf{J}_h^T(\hat{\mathbf{x}}(k))\mathbf{P}_k[\mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}(k))] \\ &= \mathbf{J}_h^T(\hat{\mathbf{x}}(k))\mathbf{P}_k[\mathbf{h}(\mathbf{x}(k)) - \mathbf{h}(\hat{\mathbf{x}}(k))] \\ &= \mathbf{J}_h^T(\hat{\mathbf{x}}(k))\mathbf{P}_k\mathcal{I}_h[\mathbf{x}(k), \hat{\mathbf{x}}(k)]\mathbf{e}(k). \end{aligned} \quad (7.312)$$

where

$$\begin{aligned} \mathbf{e}(k) &:= \mathbf{x}(k) - \hat{\mathbf{x}}(k) \\ &= \mathbf{C}[\mathbf{x}_r(k) - \hat{\mathbf{x}}_r(k)] \\ &:= \mathbf{C}\mathbf{e}_r(k). \end{aligned} \quad (7.313)$$

With Assumption A4, Theorem 6.1 guarantees that gain matrices $\mathbf{L}(k)$ exist such that the projection operator

$$\begin{aligned} \mathbf{F}(k) &:= [\mathbf{I}_n - \mathbf{L}(k)(\mathbf{J}_h(\hat{\mathbf{x}}(k))\mathbf{P}_k\mathcal{I}_h[\mathbf{x}(k), \hat{\mathbf{x}}(k)]\mathbf{C}\mathbf{L}(k))^{-1}\mathbf{J}_h(\hat{\mathbf{x}}(k))\mathbf{P}_k\mathcal{I}_h[\mathbf{x}_1, \mathbf{x}_2]\mathbf{C}] \\ &= [\mathbf{I}_n - \mathbf{L}(k)(\mathbf{C}\mathbf{L}(k))^{-1}\mathbf{C}] \end{aligned} \quad (7.314)$$

is well defined, and represents a projection along the column space of $\mathbf{L}(k)$ into the null space of $\mathbf{C}$. This result is the primary link with methods of Chapter 6 and allows analysis to follow the same approach as employed in Section 7.3.

Consider the following two proposed observer forms:

$$\mathcal{O}_1: \quad \hat{\mathbf{x}}_r(k+1) = \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_r(k)) + \mathbf{L}^{(1)}(k)\mathbf{v}^{(1)}(k), \quad (7.315)$$

where

$$\begin{aligned} \mathbf{v}^{(1)}(k) &:= \left\{ \mathbf{J}_h^T(\hat{\mathbf{x}}^{(1)}(k+1))\mathbf{P}_{k+1}\mathcal{I}_h[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)]\mathbf{C}\mathbf{L}(k) \right\}^{-1} \\ &\quad \cdot \left\{ \begin{array}{l} \mathbf{J}_h^T(\hat{\mathbf{x}}^{(1)}(k+1))\mathbf{P}_{k+1}\mathcal{I}_h[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)] \\ \cdot [\mathbf{x}(k+1) - \hat{\mathbf{x}}_p^{(1)}(k+1)] \\ - \mathbf{g}(k, \mathbf{s}_k^{(1)}) \end{array} \right\} \end{aligned} \quad (7.316)$$

$$\begin{aligned} \mathbf{v}^{(2)}(k) &:= \left\{ \mathbf{J}_h^T(\hat{\mathbf{x}}^{(2)}(k+1))\mathbf{P}_{k+1}\mathcal{I}_h[\hat{\mathbf{x}}_p^{(2)}(k+1), \hat{\mathbf{x}}^{(2)}(k+1)]\mathbf{C}\mathbf{L}(k) \right\}^{-1} \\ &\quad \cdot \left\{ \mathbf{J}_h^T(\hat{\mathbf{x}}^{(2)}(k+1))\mathbf{P}_{k+1}[\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}_p^{(2)}(k+1))] - \mathbf{g}(k, \mathbf{s}_k^{(2)}) \right\} \end{aligned} \quad (7.317)$$

$$\hat{\mathbf{x}}_p^{(i)}(k+1) := \mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_c^{(i)}(k)), \quad i = 1, 2. \quad (7.318)$$

$$\mathbf{g}(k, \mathbf{s}_k^{(i)}) := \mathbf{\Gamma}_k \text{Sat}(\mathbf{s}_k^{(i)}, \epsilon) + \mathbf{\Lambda}_k \mathbf{s}_k^{(i)} \quad i = 1, 2. \quad (7.319)$$

$$\mathbf{s}_k^{(i)} := \mathbf{J}_h^T(\hat{\mathbf{x}}^{(i)}(k)) \mathbf{P}_k [\mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}^{(i)}(k))]. \quad (7.320)$$

These two observers are direct extensions of those proposed in Section 7.3 and again represent ideal forms in the sense that, as will be shown, under certain conditions both possess a global convergent quasi-sliding regime, respectively, on $\mathbf{s}_k^{(i)} = \mathbf{0}$, $i = 1, 2$. The observer outputs, $\hat{\mathbf{x}}_c^{(i)}(k+1)$, are again represented by implicit systems of nonlinear integral difference equations.

Observer (O_1) results from the use of the Fundamental Theorem of Integral Calculus and direct application of the methods proposed in Chapter 6 for linear redundant measurement models. In O_1 , the measurements $\mathbf{z}(k+1)$ do not explicitly appear in the observer equations, and the true measured state, $\mathbf{x}(k+1)$, which is not directly available, appears three times in the expression for $\mathbf{v}^{(1)}(k)$. This form neither lends itself easily to an intuitive methodology for implementation, nor to theoretical analysis of possible approximations. This form, however, is well suited for stability analysis under the hypothesis of Theorem 5.11.

Observer (O_2) again results from the use of the Fundamental Theorem of Integral Calculus, but is developed and analyzed with a slightly alternate approach from than that followed in Chapter 6. In O_2 , the vector $\mathbf{v}^{(2)}(k)$ depends explicitly on the measurements and does not contain the true measured state $\mathbf{x}(k+1)$ (except via the measurements). The vector $\mathbf{v}^{(2)}(k)$ also depends on the predicted measured state, $\mathbf{x}_p^{(2)}(k+1)$, which can be computed as in (7.318) just following the k th measurement update.

The following Theorem details fundamental properties of these observers including the fact that, under mild assumptions, these two observers are equivalent.

Theorem 7.14: For the system (7.292) with measurement model (7.293) consider the observers O_1 in (7.315)-(7.316), and O_2 in (7.315)-(7.317). Assume that

1. $\hat{\mathbf{x}}_c^{(1)}(0) = \hat{\mathbf{x}}_c^{(2)}(0)$,
2. $\mathbf{L}(k) := \mathbf{L}^{(1)}(k) \equiv \mathbf{L}^{(2)}(k)$,
3. $\hat{\mathbf{f}}(k, \cdot)$ is 1-to-1,

4. Assumptions A4 (page 270) and A5 (page 272) hold.

Define

$$\mathbf{F}(k) := [\mathbf{I}_n - \mathbf{L}(k)(\mathbf{C}\mathbf{L}(k))^{-1}\mathbf{C}], \quad (7.321)$$

$$\mathbf{B}(k)\mathbf{d}(k, \mathbf{y}_e) := \mathbf{f}(k, \mathbf{y}_e) - \hat{\mathbf{f}}(k, \mathbf{y}_e), \quad \forall \mathbf{y}_e \in \mathbf{X}_e. \quad (7.322)$$

The following statements hold:

1. *Reaching Law*: If Λ_k , Γ_k and ϵ satisfy the hypothesis of Theorem 5.12, then global convergent quasi-sliding regimes exists on $\mathbf{s}_k^{(1)} = \mathbf{0}$ and $\mathbf{s}_k^{(2)} = \mathbf{0}$ independent of the dynamical modelling error, and pure sliding can be attained for all $k > k_s > 0$ by setting $\Lambda_k = \mathbf{0}$ and $\Gamma_k = \mathbf{0}$ for all $k \geq k_s$ according to Corollary 5.2;

2. *Equivalence*: $\forall k \geq 0$, $\hat{\mathbf{x}}_e^{(1)}(k) = \hat{\mathbf{x}}_e^{(2)}(k)$;

3. *Least Mean Square Optimization*: With Statement 1, during the sliding motion, $k > k_s$, the estimated measured state, $\hat{\mathbf{x}}(k)$, is a least-squares solution of the cost function

$$C(\hat{\mathbf{x}}(k)) := [\mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}(k))]^T \mathbf{P}_k [\mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}(k))], \quad (7.323)$$

where $\mathbf{P}_k \in \mathbb{R}^{m \times m}$ is user-specified, and positive definite symmetric.

4. *Stability of Reaching and Sliding Dynamics*: With Statements 1 and 2 above, if $\mathbf{B}(k) \in \mathbb{R}^{n \times p}$, $\mathbf{d}(k, \mathbf{x}_e) \in \mathbb{R}^{p \times 1}$, $\mathbf{F}(k)$, and $\hat{\mathbf{f}}$ satisfy the hypothesis of Theorem 5.11, where the parameters ϵ_2 , $\xi(k)$ and $\eta(k)$ of Theorem 5.11 are taken, with L as defined in Assumption A4, as

$$\epsilon_2 = 0, \quad (7.324)$$

$$\xi(k) := -L\mathbf{g}(k, \mathbf{s}_k), \quad \text{and} \quad (7.325)$$

$$\eta(k) = 0, \quad (7.326)$$

then the error dynamics for O_1 and O_2 are locally exponentially stable;

5. *Robustness of Sliding Mode*: With Statements 1 and 4 above, if, in addition, $\exists \Xi(k) \in \mathbb{R}^{m \times p}$ such that $\mathbf{B}(k) = \mathbf{L}(k)\Xi(k)$, then the sliding dynamics of O_1 and O_2 are invariant to dynamical modelling errors/disturbances in which case the sliding dynamics are locally exponentially asymptotically stable.

Proof: For Statement 1, it is required to show that under the above definitions, $s_{k+1}^{(i)} = \omega(k, s_k^{(i)})$, $i = 1, 2$, from which the result follows by Theorem 5.12. For $\mathcal{O}_1$,

$$\begin{aligned} s_{k+1}^{(1)} &:= \mathbf{J}_h^T(\hat{\mathbf{x}}^{(1)}(k+1))\mathbf{P}_{k+1} [\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}^{(1)}(k+1))] \\ &= \mathbf{J}_h^T(\hat{\mathbf{x}}^{(1)}(k))\mathbf{P}_k \mathcal{I}_h[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)]\mathbf{C} [\mathbf{x}_e(k+1) - \hat{\mathbf{x}}_e^{(1)}(k+1)] \\ &= \mathbf{J}_h^T(\hat{\mathbf{x}}^{(1)}(k))\mathbf{P}_k \mathcal{I}_h[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)]\mathbf{C} [\mathbf{f}(k, \mathbf{x}_e(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(1)}(k))] \\ &\quad - \mathbf{J}_h^T(\hat{\mathbf{x}}^{(1)}(k))\mathbf{P}_k \mathcal{I}_h[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)]\mathbf{C}\mathbf{L}^{(1)}(k)\mathbf{v}^{(1)}(k) \\ &= \mathbf{g}(k, s_k^{(1)}). \end{aligned} \quad (7.327)$$

For $\mathcal{O}_2$,

$$\begin{aligned} s_{k+1}^{(2)} &:= \mathbf{J}_h^T(\hat{\mathbf{x}}^{(2)}(k+1))\mathbf{P}_{k+1} [\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}^{(2)}(k+1))] \\ &= \mathbf{J}_h^T(\hat{\mathbf{x}}^{(2)}(k+1))\mathbf{P}_{k+1} [\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}_p^{(2)}(k+1))] \\ &\quad + \mathbf{J}_h^T(\hat{\mathbf{x}}^{(2)}(k+1))\mathbf{P}_{k+1} [\mathbf{h}(\hat{\mathbf{x}}_p^{(2)}(k+1)) - \mathbf{h}(\mathbf{x}^{(2)}(k+1))] \\ &= \mathbf{J}_h^T(\hat{\mathbf{x}}^{(2)}(k+1))\mathbf{P}_{k+1} [\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}_p^{(2)}(k+1))] \\ &\quad + \mathbf{J}_h^T(\hat{\mathbf{x}}^{(2)}(k+1))\mathbf{P}_{k+1} \mathcal{I}_h[\hat{\mathbf{x}}_p^{(2)}(k+1), \hat{\mathbf{x}}^{(2)}(k+1)] \\ &\quad \cdot [\hat{\mathbf{x}}_p^{(2)}(k+1) - \hat{\mathbf{x}}^{(2)}(k+1)] \\ &= \mathbf{J}_h^T(\hat{\mathbf{x}}^{(2)}(k+1))\mathbf{P}_{k+1} [\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}_p^{(2)}(k+1))] \\ &\quad - \mathbf{J}_h^T(\hat{\mathbf{x}}^{(2)}(k+1))\mathbf{P}_{k+1} \mathcal{I}_h[\hat{\mathbf{x}}_p^{(2)}(k+1), \hat{\mathbf{x}}^{(2)}(k+1)]\mathbf{C}\mathbf{L}^{(2)}(k)\mathbf{v}^{(2)}(k) \\ &= \mathbf{g}(k, s_k^{(2)}). \end{aligned} \quad (7.328)$$

The proof of Statement 2 is by induction. Suppose, for some $k_0 \geq 0$, $\mathbf{x}_e^{(1)}(k_0) = \hat{\mathbf{x}}_e^{(2)}(k_0)$. Then $s_{k_0}^{(1)} = s_{k_0}^{(2)}$, and from (7.327) and (7.328),

$$s_{k_0+1}^{(1)} = \mathbf{g}(k_0, s_{k_0}^{(1)}) = \mathbf{g}(k_0, s_{k_0}^{(2)}) = s_{k_0+1}^{(2)}. \quad (7.329)$$

Using (7.329) and Assumption A5 gives

$$\begin{aligned} \mathbf{0} &= s_{k_0+1}^{(1)} - s_{k_0+1}^{(2)} \\ \Rightarrow \hat{\mathbf{x}}^{(1)}(k_0+1) &= \hat{\mathbf{x}}^{(2)}(k_0+1). \end{aligned} \quad (7.330)$$

Also, under the induction hypothesis, and assuming that $\hat{\mathbf{f}}(k, \cdot)$ is 1-to-1, the predicted measured states satisfy

$$\hat{\mathbf{x}}_p^{(1)}(k_0+1) := \mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(1)}(k_0)) = \mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(2)}(k_0)) := \hat{\mathbf{x}}_p^{(2)}(k_0+1). \quad (7.331)$$

4.3, and 5.6, for unmodelled dynamics and disturbances to influence the sliding dynamics through premultiplication by a projection matrix (see (3.69), page 98) of the form

$$\mathbf{F}(k) := [\mathbf{I}_n - \mathbf{L}(k) (\mathbf{H}_e \mathbf{L}(k))^{-1} \mathbf{H}_e], \quad (6.15)$$

where $\mathbf{L}(k)$ are the observer gains. The matrix $\mathbf{F}(k)$, for each k , represents a projection along the column space of $\mathbf{L}(k)$ into the null space of $\mathbf{H}_e$. In order to achieve this objective, gain matrices $\mathbf{L}(k) \in \mathbb{R}^{n \times m}$ must exist such that the product $\mathbf{H}_e \mathbf{L}(k)$ is nonsingular.

Theorem 6.1: For $\mathbf{H}_e \in \mathbb{R}^{m \times n}$, there exists a matrix $\mathbf{L} \in \mathbb{R}^{n \times m}$ such that the product $\mathbf{H}_e \mathbf{L}$ is nonsingular if, and only if, $\text{rank}\{\mathbf{H}_e\} = m$.

Proof: The existence of such a matrix $\mathbf{L}$ is equivalent to the existence of a right inverse of $\mathbf{H}_e$. That is, if $\exists \mathbf{L}$ such that $\mathbf{H}_e \mathbf{L} = \mathbf{M}$ where $\mathbf{M} \in \mathbb{R}^{m \times m}$ is nonsingular, then $\mathbf{B} := \mathbf{L} \mathbf{M}^{-1}$ is a right inverse of $\mathbf{H}_e$, and, conversely, if $\mathbf{B}$ is a right inverse of $\mathbf{H}_e$, then $\mathbf{H}_e \mathbf{L}$ is nonsingular with $\mathbf{L} = \mathbf{B}$. A necessary and sufficient condition for the existence of a right inverse of $\mathbf{H}_e$ is that $\mathbf{H}_e$ have linearly independent rows, i.e. $\text{rank}\{\mathbf{H}_e\} = m$ (see, for example, Perlis, Theorem 3-7 [161, p. 58]). $\square$

6.3 Linear Measurement Redundancy

The purpose of this Section is to quantify what is meant by a redundant linear measurement model. Measurement redundancy refers to the case in which more measurements are available than the minimum required to uniquely reconstruct the extended state of the system under ideal conditions (no disturbances, unmodelled dynamics, or measurement noise). Example 6.1 considered a system which has *redundant measurements*; the system provides two measurements of position in each of the x - and y -components. If $\text{rank}\{\mathbf{H}_e\} = r < m$, then the ideal (no noise) measurements for all time lie in the r -dimensional column space of $\mathbf{H}_e$. If the system is uniformly observable with N measurement events (see Section 3.3, page 82), then the observability map defined in (3.11) is injective from $\mathbf{X}_e$ into $\mathbb{R}^{Nr} \subset \mathbb{R}^{Nm}$. Intuitively, this means that the measurement model can be reduced to an r -dimensional “measurement” vector for each k , and still retain uniform observability. It must be emphasized that the objective is not to propose methods to remove

4

PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

With $\mathbf{L}(k_0) := \mathbf{L}^{(1)}(k_0) = \mathbf{L}^{(2)}(k_0)$ and using

$$\mathbf{CL}(k_0)\mathbf{v}^{(1)}(k_0) = \hat{\mathbf{x}}^{(1)}(k_0 + 1) - \hat{\mathbf{x}}_p^{(1)}(k_0 + 1), \quad (7.332)$$

consider

$$\begin{aligned} & \mathbf{J}_h^T(\hat{\mathbf{x}}^{(1)}(k_0 + 1))\mathbf{P}_{k_0-1}\mathcal{I}_h[\hat{\mathbf{x}}_p^{(1)}(k_0 + 1), \hat{\mathbf{x}}^{(1)}(k_0 + 1)]\mathbf{CL}(k_0)\mathbf{v}^{(1)}(k_0) \\ &= \mathbf{J}_h^T(\hat{\mathbf{x}}^{(1)}(k_0 + 1))\mathbf{P}_{k_0-1}\mathcal{I}_h[\hat{\mathbf{x}}_p^{(1)}(k_0 + 1), \hat{\mathbf{x}}^{(1)}(k_0 + 1)] \\ & \quad \cdot [\hat{\mathbf{x}}^{(1)}(k_0 + 1) - \hat{\mathbf{x}}_p^{(1)}(k_0 + 1)] \\ &= \mathbf{J}_h^T(\hat{\mathbf{x}}^{(1)}(k_0 + 1))\mathbf{P}_{k_0-1} \left\{ \mathbf{h}(\hat{\mathbf{x}}^{(1)}(k_0 + 1)) - \mathbf{h}(\hat{\mathbf{x}}_p^{(1)}(k_0 + 1)) \right\} \\ &= \mathbf{J}_h^T(\hat{\mathbf{x}}^{(1)}(k_0 + 1))\mathbf{P}_{k_0-1} \left\{ \mathbf{h}(\hat{\mathbf{x}}^{(1)}(k_0 + 1)) - \mathbf{h}(\mathbf{x}(k_0 + 1)) \right\} \\ & \quad + \mathbf{J}_h^T(\hat{\mathbf{x}}^{(1)}(k_0 + 1))\mathbf{P}_{k_0-1} \left\{ \mathbf{h}(\mathbf{x}(k_0 + 1)) - \mathbf{h}(\hat{\mathbf{x}}_p^{(1)}(k_0 + 1)) \right\} \\ &= -\mathbf{s}_{k_0-1}^{(1)} + \mathbf{J}_h^T(\hat{\mathbf{x}}^{(1)}(k_0 + 1))\mathbf{P}_{k_0-1} \left\{ \mathbf{z}(k + 1) - \mathbf{h}(\hat{\mathbf{x}}_p^{(1)}(k_0 + 1)) \right\}. \quad (7.333) \end{aligned}$$

Substituting the results of (7.329), (7.330), and (7.331) into (7.333) and rearranging gives

$$\begin{aligned} \mathbf{v}^{(1)}(k_0) &= \left[\mathbf{J}_h^T(\hat{\mathbf{x}}^{(2)}(k_0 + 1))\mathbf{P}_{k_0-1}\mathcal{I}_h(\hat{\mathbf{x}}_p^{(2)}(k_0 + 1), \hat{\mathbf{x}}^{(2)}(k_0 + 1))\mathbf{CL}(k_0) \right]^{-1} \\ & \quad \cdot \left\{ \mathbf{J}_h^T(\hat{\mathbf{x}}^{(2)}(k_0 + 1))\mathbf{P}_{k_0-1} \left[\mathbf{z}(k + 1) - \mathbf{h}(\hat{\mathbf{x}}_p^{(2)}(k_0 + 1)) \right] - \mathbf{g}(k_0, \mathbf{s}_{k_0}^{(2)}) \right\} \\ &= \mathbf{v}^{(2)}(k). \quad (7.334) \end{aligned}$$

Finally, with the induction hypothesis and (7.334), and again using the assumption that $\hat{\mathbf{f}}(k, \cdot)$ is 1-to-1,

$$\begin{aligned} \hat{\mathbf{x}}_e^{(1)}(k_0 + 1) &:= \hat{\mathbf{f}}(k_0, \mathbf{x}_e^{(1)}(k_0)) + \mathbf{L}(k_0)\mathbf{v}^{(1)}(k_0) \\ &= \hat{\mathbf{f}}(k_0, \hat{\mathbf{x}}_e^{(2)}(k_0)) + \mathbf{L}(k_0)\mathbf{v}^{(2)}(k_0) \\ &:= \hat{\mathbf{x}}_e^{(2)}(k_0 + 1). \quad (7.335) \end{aligned}$$

Hence, if $\hat{\mathbf{x}}_e^{(1)}(0) = \hat{\mathbf{x}}_e^{(2)}(0)$, then $\forall k \geq 0$, $\hat{\mathbf{x}}_e^{(1)}(k) = \hat{\mathbf{x}}_e^{(2)}(k)$.

Statement 3 follows immediately from the definitions of $\mathbf{s}_k^{(1)}$ in (7.320) and the cost function $C(\hat{\mathbf{x}})$ in (7.323), which satisfy

$$\mathbf{s}_k := -\frac{1}{2} \left(\frac{\partial C(\hat{\mathbf{x}}(k))}{\partial \hat{\mathbf{x}}(k)} \right)^T, \quad (7.336)$$

and, from Statement 1, the fact that $\mathbf{s}_k^{(1)} = \mathbf{0}$ for all $k > k_s$ and hence $\hat{\mathbf{x}}(k)$ is a least-squares solution of (7.323) during the sliding motion.

For Statement 4, and with the equivalence result of Statement 2, it is sufficient to show that observer O_1 can be placed in a Luenberger form, i.e. (5.115), page 157, satisfying the hypothesis of Theorem 5.11. But observer O_1 can be written as

$$\hat{\mathbf{x}}_e^{(1)}(k+1) = \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(1)}(k)) + \mathbf{K}(k) [\bar{\mathbf{z}}(k+1) - \mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e^{(1)}(k)) + \boldsymbol{\gamma}(k)], \quad (7.337)$$

where

$$\bar{\mathbf{z}}(k+1) := \mathbf{C}\mathbf{x}_e(k+1). \quad (7.338)$$

$$\mathbf{K}(k) := \mathbf{L}(k) [\mathbf{C}\mathbf{L}(k)]^{-1}, \quad \text{and} \quad (7.339)$$

$$\boldsymbol{\gamma}(k) := -\mathbf{J}_h^T(\hat{\mathbf{x}}^{(1)}(k+1))\mathbf{P}_{k+1}\mathcal{I}_h^{-1}[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)]\mathbf{g}(k, \mathbf{s}_k^{(1)}). \quad (7.340)$$

This is the observer form considered in Theorem 5.11 with the linear "measurement model" in (7.338), and the result follows with hypothesis (7.325) provided it is shown that $\|\boldsymbol{\gamma}(k)\| \leq \|\boldsymbol{\xi}(k)\|$. But under Assumption A4,

$$\begin{aligned} \|\boldsymbol{\gamma}(k)\| &:= \|\mathbf{J}_h^T(\hat{\mathbf{x}}^{(1)}(k+1))\mathbf{P}_{k+1}\mathcal{I}_h^{-1}[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)]\mathbf{g}(k, \mathbf{s}_k^{(1)})\| \\ &\leq L\|\mathbf{g}(k, \mathbf{s}_k^{(1)})\| := \|\boldsymbol{\xi}(k)\|. \end{aligned} \quad (7.341)$$

Note that if $k_s = 0$, then a value for L defined in Item 2 of Assumption A4 is not required since in that case $\mathbf{g}(k, \mathbf{s}_k) \equiv \mathbf{0}$. Item 1 of Assumption A4, however, is still required since in this case, $\forall k \geq 0$, $\mathbf{J}_h^T(\hat{\mathbf{x}}^{(1)}(k+1))\mathbf{P}_{k+1}\mathcal{I}_h[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)] = \mathbf{J}_h^T(\hat{\mathbf{x}}^{(1)}(k+1))\mathbf{P}_{k+1}\mathbf{J}_h(\hat{\mathbf{x}}^{(1)}(k+1))$.

For Statement 5, $\mathbf{g}(k, \mathbf{s}_k) \equiv \mathbf{0}$ for all $k \geq k_s$, and Assumption A4 gives $\mathbf{x}(k+1) = \hat{\mathbf{x}}^{(i)}(k+1)$, $i = 1, 2$. This in turn implies that $\forall k \geq k_s$,

$$\mathbf{J}_h^T(\hat{\mathbf{x}}^{(1)}(k+1))\mathbf{P}_{k+1}\mathcal{I}_h[\mathbf{x}(k+1), \hat{\mathbf{x}}^{(1)}(k+1)] = \mathbf{J}_h^T(\hat{\mathbf{x}}^{(1)}(k+1))\mathbf{P}_{k+1}\mathbf{J}_h(\hat{\mathbf{x}}^{(1)}(k+1)). \quad (7.342)$$

and

$$\begin{aligned} \mathbf{J}_h^T(\hat{\mathbf{x}}^{(1)}(k+1))\mathbf{P}_{k+1}\mathcal{I}_h[\hat{\mathbf{x}}_p^{(2)}(k+1), \hat{\mathbf{x}}^{(2)}(k+1)] &= \\ \mathbf{J}_h^T(\hat{\mathbf{x}}^{(1)}(k+1))\mathbf{P}_{k+1}\mathcal{I}_h[\hat{\mathbf{x}}_p^{(2)}(k+1), \mathbf{x}(k+1)]. \end{aligned} \quad (7.343)$$

With (7.342) and (7.343),

$$\mathbf{v}^{(i)}(k) = [\mathbf{C}\mathbf{L}(k)]^{-1} \{\mathbf{x}(k+1) - \hat{\mathbf{x}}_p^{(i)}(k+1)\}. \quad (7.344)$$

The error dynamics during sliding can then be written, with $e_e^{(i)}(k) := x_e(k) - \hat{x}_e^{(i)}(k)$ and using (7.337), as

$$\begin{aligned}
 e_e^{(i)}(k+1) &:= x_e(k+1) - \hat{x}_e^{(i)}(k+1) \\
 &= f(k, x_e(k)) - \hat{f}(k, \hat{x}_e^{(i)}(k)) - L(k)v^{(i)}(k) \\
 &= F(k) \{f(k, x_e(k)) - \hat{f}(k, \hat{x}_e^{(i)}(k))\} \\
 &= F(k) \{f(k, x_e(k)) - \hat{f}(k, x_e(k))\} + F(k) \{\hat{f}(k, x_e(k)) - \hat{f}(k, \hat{x}_e^{(i)}(k))\} \\
 &= F(k)B(k)d(k, x_e(k)) + F(k) \{\hat{f}(k, x_e(k)) - \hat{f}(k, \hat{x}_e^{(i)}(k))\} \\
 &= F(k)L(k)\Xi(k)d(k, x_e(k)) + F(k) \{\hat{f}(k, x_e(k)) - \hat{f}(k, \hat{x}_e^{(i)}(k))\} \\
 &= F(k) \{\hat{f}(k, x_e(k)) - \hat{f}(k, \hat{x}_e^{(i)}(k))\}.
 \end{aligned} \tag{7.345}$$

where the fourth equality is obtained by adding and subtracting $F(k)\hat{f}(k, x_e(k))$, and the final equality results since $F(k)L(k) \equiv 0$. Hence the sliding dynamics are invariant to dynamical modelling errors/disturbances, and application of Theorem 5.11 gives exponential asymptotic stability. $\square$

Statement 4 of Theorem 7.14 treats the local exponential stability of both the reaching and sliding dynamics. Assumptions A4 and A5 are invoked to obtain equivalence of O_1 and O_2 and bounds on the term

$$\begin{aligned}
 J_h^T(\hat{x}(k+1))P_{k+1}I_h^{-1}[x(k+1), \hat{x}^{(1)}(k+1)]g(k, s_k^{(1)}) &= \\
 J_h^T(\hat{x}(k+1))P_{k+1}I_h^{-1}[x(k+1), \hat{x}^{(1)}(k+1)]\{\Gamma_k \text{Sat}(s_k^{(1)}, \epsilon) + \Lambda_k s_k^{(1)}\}.
 \end{aligned} \tag{7.346}$$

which appears through simplification of $v^{(1)}(k)$ (7.316) in observer O_1 . These bounds are required during the reaching phase for $0 \leq k \leq k_s$. The designer has full control over parameters Γ_k and Λ_k of the reaching law which govern the rate of approach to the sliding surface where, under Assumption A5, $x(k+1) = \hat{x}^{(1)}(k+1)$. One possible method, in principle, is to select the elements of Γ and Λ such that for all $k \geq 0$, $\|x(k+1) - \hat{x}^{(1)}(k+1)\| < \delta$ where δ is defined in hypothesis (7.303) of Theorem 7.13. In this case, Assumption A4 would be guaranteed under the hypothesis of Theorem 7.13.

A primary limitation of Theorem 7.14 is that measurement errors have not been considered. Note that the derivation (7.328) leading to the reaching law for observer O_2 is independent of measurement errors. Hence a global convergent quasi-sliding regime exists

on $s_k^{(2)} = 0$ even in the presence of measurement errors. However, further work is required to analyze observer stability and robustness properties in this case.

7.4.3 Implementation With Local Iterations

In this Section, an approximate implementation of observer O_2 from (7.315) and (7.317) is proposed. For brevity, the superscript “(2)” will be dropped.

For the system (7.292, repeated)

$$\mathbf{x}_e(k+1) = \mathbf{f}(k, \mathbf{x}_e(k)), \quad \mathbf{x}(k) := \mathbf{C}\mathbf{x}_e(k). \quad (7.347)$$

with measurements (7.293, repeated)

$$\mathbf{z}(k) = \mathbf{h}(\mathbf{x}(k)). \quad (7.348)$$

observer O_2 has the form

$$\hat{\mathbf{x}}_e(k+1) = \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \mathbf{L}(k)\mathbf{v}(k), \quad (7.349)$$

where

$$\begin{aligned} \mathbf{v}(k) &:= \left\{ \mathbf{J}_h^T(\hat{\mathbf{x}}(k+1))\mathbf{P}_{k+1}\mathcal{I}_h[\hat{\mathbf{x}}_p(k+1), \hat{\mathbf{x}}(k+1)]\mathbf{C}\mathbf{L}(k) \right\}^{-1} \\ &\quad \cdot \left\{ \mathbf{J}_h^T(\hat{\mathbf{x}}(k+1))\mathbf{P}_{k+1}[\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}_p(k+1))] - \mathbf{g}(k, \mathbf{s}_k) \right\} \end{aligned} \quad (7.350)$$

$$\hat{\mathbf{x}}_p(k+1) := \mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)), \quad (7.351)$$

$$\mathbf{s}_k := \mathbf{J}_h^T(\hat{\mathbf{x}}(k))\mathbf{P}_k[\mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}(k))]. \quad (7.352)$$

$$\mathbf{g}(k, \mathbf{s}_k) := \mathbf{\Gamma}_k\mathbf{Sat}(\mathbf{s}_k, \epsilon) + \mathbf{\Lambda}_k\mathbf{s}_k. \quad (7.353)$$

The first approach for implementation follows that of Section 7.3.3 with the approximation

$$\mathcal{I}_h[\hat{\mathbf{x}}_p(k+1), \hat{\mathbf{x}}(k+1)] \approx \mathbf{J}_h(\hat{\mathbf{x}}_p(k+1)) := \frac{\partial \mathbf{h}}{\partial \mathbf{x}}[\mathbf{C}\hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k))]. \quad (7.354)$$

which leads to the observer

$$\hat{\mathbf{x}}_e(k+1) = \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \mathbf{L}(k)\tilde{\mathbf{v}}(k), \quad (7.355)$$

$$\tilde{\mathbf{v}}(k) = \left[\mathbf{J}_h^T(\hat{\mathbf{x}}_p(k+1))\mathbf{P}_{k+1}\mathbf{J}_h(\hat{\mathbf{x}}_p(k+1))\mathbf{C}\mathbf{L}(k) \right]^{-1} \quad (7.356)$$

$$\begin{aligned} & \cdot \left\{ \mathbf{J}_h^T(\hat{\mathbf{x}}_p(k+1)) \mathbf{P}_{k+1} [\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}_p(k+1))] - \mathbf{g}(k, \mathbf{s}_k) \right\} \\ \hat{\mathbf{x}}_p(k+1) &:= \mathbf{C} \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)), \end{aligned} \quad (7.357)$$

$$\mathbf{s}_k := \mathbf{J}_h^T(\hat{\mathbf{x}}(k)) \mathbf{P}_k \mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}(k)). \quad (7.358)$$

and $\mathbf{g}(k, \mathbf{s}_k)$ is defined by the reaching law in (7.353). In this observer, as will be shown below, one step of a modified Gauss-Newton method is applied at each measurement update.

Local modified Gauss-Newton iterations are proposed as follows. Suppose at time k an estimate $\hat{\mathbf{x}}_e(k)$ is available. Set

$$\mathbf{s}_k := \mathbf{J}_h^T(\mathbf{C} \hat{\mathbf{x}}_e(k)) \mathbf{P}_k [\mathbf{z}(k) - \mathbf{h}(\mathbf{C} \hat{\mathbf{x}}_e(k))]. \quad (7.359)$$

$$\mathbf{g}(k, \mathbf{s}_k) = \mathbf{\Gamma}_k \mathbf{Sat}(\mathbf{s}_k, \epsilon) + \mathbf{\Lambda}_k \mathbf{s}_k. \quad (7.360)$$

$$\hat{\mathbf{x}}_e^{(0)}(k+1) := \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)). \quad (7.361)$$

In this notation, $\hat{\mathbf{x}}_e^{(0)}(k+1)$ is the predicted extended state based on the estimate $\hat{\mathbf{x}}_e(k)$ at time k . With the availability of measurement $\mathbf{z}(k+1)$, the proposed local iterations take the form, for k fixed, $N_{k+1} \geq 1$, and with $j = 0, 1, \dots, N_{k+1} - 1$,

$$\mathbf{s}_{k+1}^{(j)} = \mathbf{J}_h^T(\mathbf{C} \hat{\mathbf{x}}_e^{(j)}(k+1)) \mathbf{P}_{k+1} [\mathbf{z}(k+1) - \mathbf{h}(\mathbf{C} \hat{\mathbf{x}}_e^{(j)}(k+1))]. \quad (7.362)$$

$$\mathbf{w}_{k+1}^{(j)} = \mathbf{g}(k, \mathbf{s}_k) - \mathbf{s}_{k+1}^{(j)}. \quad (7.363)$$

$$\begin{aligned} \mathbf{v}^{(j)}(k) &= \left[\mathbf{J}_h^T(\mathbf{C} \hat{\mathbf{x}}_e^{(j)}(k+1)) \mathbf{P}_{k+1} \mathbf{J}_h(\mathbf{C} \hat{\mathbf{x}}_e^{(j)}(k+1)) \mathbf{C} \mathbf{L}(k) \right]^{-1} \\ &\quad \cdot \mathbf{w}_{k+1}^{(j)} \end{aligned} \quad (7.364)$$

$$\hat{\mathbf{x}}_e^{(j+1)}(k+1) = \hat{\mathbf{x}}_e^{(j)}(k+1) - \mathbf{L}^{(j)}(k) \mathbf{v}^{(j)}(k).$$

The observer output at time $k+1$ is then taken as

$$\hat{\mathbf{x}}_e(k+1) := \hat{\mathbf{x}}_e^{(N_{k+1})}(k+1). \quad (7.365)$$

Alternatively, the iterations in (7.363) can be expressed as

$$\hat{\mathbf{x}}_e^{(j+1)}(k+1) = \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) + \sum_{\ell=0}^j \mathbf{L}^{(\ell)}(k) \mathbf{v}^{(\ell)}(k), \quad j = 0, 1, \dots, N_{k+1} - 1. \quad (7.366)$$

where

$$\begin{aligned} \mathbf{v}^{(\ell)}(k) &:= \left[\mathbf{J}_h^T(\hat{\mathbf{x}}^{(\ell)}(k+1)) \mathbf{P}_{k+1} \mathbf{J}_h(\hat{\mathbf{x}}^{(\ell)}(k+1)) \mathbf{C} \mathbf{L}^{(\ell)}(k) \right]^{-1} \\ &\quad \cdot \left\{ \mathbf{J}_h^T(\hat{\mathbf{x}}^{(\ell)}(k+1)) \mathbf{P}_{k+1} [\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}^{(\ell)}(k+1))] - \mathbf{g}(k, \mathbf{s}_k) \right\}. \end{aligned} \quad (7.367)$$

and $\hat{\mathbf{x}}^{(l)}(k+1) := \mathbf{C}\hat{\mathbf{x}}_e^{(l)}(k+1)$.

The iterated observer approximates the reaching law according to the following Lemma.

Lemma 7.8: Define $\mathbf{r}_{k+1}^{(j)} \in \mathbb{R}^m$ and $\mathbf{U}_{k+1}^{(j)} \in \mathbb{R}^{m \times q}$ as

$$\begin{aligned} \mathbf{r}_{k+1}^{(j)} &:= \mathbf{h}(\hat{\mathbf{x}}^{(j)}(k+1)) - \mathbf{h}(\hat{\mathbf{x}}^{(j-1)}(k+1)) \\ &\quad - \mathbf{J}_h(\hat{\mathbf{x}}^{(j-1)}(k+1)) [\hat{\mathbf{x}}^{(j)}(k+1) - \hat{\mathbf{x}}^{(j-1)}(k+1)] \end{aligned} \quad (7.368)$$

$$\begin{aligned} &= \mathbf{h}(\hat{\mathbf{x}}^{(j)}(k+1)) - \mathbf{h}(\hat{\mathbf{x}}^{(j-1)}(k+1)) \\ &\quad + \mathbf{J}_h(\hat{\mathbf{x}}^{(j-1)}(k+1)) \mathbf{CL}(k) \mathbf{v}^{(j-1)}(k). \end{aligned} \quad (7.369)$$

$$\mathbf{U}_{k+1}^{(j)} := \mathbf{J}_h(\hat{\mathbf{x}}^{(j)}(k+1)) - \mathbf{J}_h(\hat{\mathbf{x}}^{(j-1)}(k+1)). \quad (7.370)$$

Following the j th iteration for the measurement update at time $k+1$, (7.363) gives the approximate reaching law

$$\mathbf{s}_{k+1}^{(j)} = \mathbf{g}(k, \mathbf{s}_k) - \mathbf{w}_{k+1}^{(j)}. \quad (7.371)$$

where $\mathbf{w}_{k+1}^{(j)}$ satisfies

$$\mathbf{w}_{k+1}^{(j)} = \mathbf{J}_h^T(\hat{\mathbf{x}}^{(j-1)}(k+1)) \mathbf{P}_{k+1} \mathbf{r}_{k+1}^{(j)} - (\mathbf{U}_{k+1}^{(j)})^T \mathbf{P}_{k+1} [\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}^{(j)}(k+1))]. \quad (7.372)$$

Proof: By direct computation,

$$\begin{aligned} \mathbf{s}_{k+1}^{(j)} &= \mathbf{J}_h^T(\hat{\mathbf{x}}^{(j)}(k+1)) \mathbf{P}_{k+1} [\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}^{(j)}(k+1))] \\ &= [\mathbf{J}_h^T(\hat{\mathbf{x}}^{(j-1)}(k+1)) + (\mathbf{U}_{k+1}^{(j)})^T] \mathbf{P}_{k+1} [\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}^{(j)}(k+1))] \\ &= \mathbf{J}_h^T(\hat{\mathbf{x}}^{(j-1)}(k+1)) \mathbf{P}_{k+1} [\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}^{(j-1)}(k+1))] \\ &\quad - \mathbf{J}_h^T(\hat{\mathbf{x}}^{(j-1)}(k+1)) \mathbf{P}_{k+1} [\mathbf{h}(\hat{\mathbf{x}}^{(j)}(k+1)) - \mathbf{h}(\hat{\mathbf{x}}^{(j-1)}(k+1))] \\ &\quad + (\mathbf{U}_{k+1}^{(j)})^T \mathbf{P}_{k+1} [\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}^{(j)}(k+1))] \\ &= \mathbf{s}_{k+1}^{(j-1)} + \mathbf{w}_{k+1}^{j-1} \\ &\quad - \mathbf{J}_h^T(\hat{\mathbf{x}}^{(j-1)}(k+1)) \mathbf{P}_{k+1} \mathbf{r}_{k+1}^{(j)} \\ &\quad + (\mathbf{U}_{k+1}^{(j)})^T \mathbf{P}_{k+1} [\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}^{(j)}(k+1))] \\ &= \mathbf{g}(k, \mathbf{s}_k) - \mathbf{w}_{k+1}^{(j)}. \end{aligned} \quad (7.373)$$

In the above, (7.370) was used to obtain the second equality, (7.362) was used in the first expression of the third equality, (7.368) and (7.364) were used in the second expression of the third equality, and (7.363) was used in the fourth equality to obtain the result. $\square$

The proposed local iterations represent a Gauss-Newton method suitably modified to incorporate the reaching law. The following example, which draws from Ortega and Rheinboldt [159, pp. 267-269] reviews the standard Gauss-Newton approach.

Example 7.3: Suppose that $\mathbf{F} : \mathcal{D} \subset \mathbb{R}^q \mapsto \mathbb{R}^m$ and $\gamma : \mathbb{R}^m \mapsto \mathbb{R}^+$. Define the cost function $C : \mathcal{D} \subset \mathbb{R}^q \mapsto \mathbb{R}^+$ as

$$C(\mathbf{x}) = \gamma(\mathbf{F}(\mathbf{x})), \quad \mathbf{x} \in \mathcal{D}. \quad (7.374)$$

For example, if $\gamma(\boldsymbol{\xi})$ is the quadratic $\boldsymbol{\xi}^T \boldsymbol{\xi}/2$, then a minimizer of

$$C(\mathbf{x}) = \frac{1}{2} \mathbf{F}^T(\mathbf{x}) \mathbf{F}(\mathbf{x}) \quad (7.375)$$

is a least-squares solution of the equation $\mathbf{F}(\mathbf{x}) = \mathbf{0}$. Assume that $\mathbf{F}(\cdot)$ is differentiable on $\mathcal{D}$ and that $\hat{\mathbf{x}}^{(j)}$ is an approximate minimizer of (7.374). A natural approach to the computation of the next approximation $\hat{\mathbf{x}}^{(j+1)}$ is to minimize the functional

$$C^{(j)}(\mathbf{x}) := \gamma \left(\mathbf{F}(\hat{\mathbf{x}}^{(j)}) + \mathbf{J}_{\mathbf{F}}(\hat{\mathbf{x}}^{(j)}) [\mathbf{x} - \hat{\mathbf{x}}^{(j)}] \right) \quad (7.376)$$

obtained by linearizing $\mathbf{F}$ about $\hat{\mathbf{x}}^{(j)}$. Critical points, $\mathbf{x}^*$, of (7.376) satisfy

$$\mathbf{J}_{C^{(j)}}(\mathbf{x}^*) = \mathbf{0}^T. \quad (7.377)$$

The mechanism of Gauss-Newton's method is to compute an approximate minimizer of (7.376) with a one-step Newton iteration starting from $\hat{\mathbf{x}}^{(j)}$. Note that

$$\mathbf{J}_{C^{(j)}}(\mathbf{x}) = \mathbf{J}_{\gamma} \left(\mathbf{F}(\hat{\mathbf{x}}^{(j)}) + \mathbf{J}_{\mathbf{F}}(\hat{\mathbf{x}}^{(j)}) [\mathbf{x} - \hat{\mathbf{x}}^{(j)}] \right) \mathbf{J}_{\mathbf{F}}(\hat{\mathbf{x}}^{(j)}), \quad (7.378)$$

and that the Hessian, $\mathbf{H}_{C^{(j)}}(\mathbf{x})$, of $C^{(j)}(\mathbf{x})$ is

$$\mathbf{H}_{C^{(j)}}(\mathbf{x}) = \mathbf{J}_{\mathbf{F}}^T(\mathbf{x}^{(j)}) \mathbf{H}_{\gamma} \left(\mathbf{F}(\hat{\mathbf{x}}^{(j)}) + \mathbf{J}_{\mathbf{F}}(\hat{\mathbf{x}}^{(j)}) [\mathbf{x} - \hat{\mathbf{x}}^{(j)}] \right) \mathbf{J}_{\mathbf{F}}(\hat{\mathbf{x}}^{(j)}), \quad (7.379)$$

where $\mathbf{H}_{\gamma}$ is the Hessian of $\gamma(\cdot)$. Application of one-step of Newton's algorithm then gives the generalized Gauss-Newton method,

$$\begin{aligned} \mathbf{x}^{(j+1)} &= \mathbf{x}^{(j)} - \left[\mathbf{H}_{C^{(j)}}(\mathbf{x}^{(j)}) \right]^{-1} \left[\mathbf{J}_{C^{(j)}}(\mathbf{x}^{(j)}) \right]^T \\ &= \mathbf{x}^{(j)} - \left[\mathbf{J}_{\mathbf{F}}^T(\mathbf{x}^{(j)}) \mathbf{H}_{\gamma}(\mathbf{x}^{(j)}) \mathbf{J}_{\mathbf{F}}(\mathbf{x}^{(j)}) \right]^{-1} \mathbf{J}_{\mathbf{F}}^T(\mathbf{x}^{(j)}) \left[\mathbf{J}_{C^{(j)}}(\mathbf{x}^{(j)}) \right]^T. \end{aligned} \quad (7.380)$$

For example, with

$$\gamma(\xi) = \frac{1}{2} \xi^T \mathbf{P} \xi, \quad (7.381)$$

where $\mathbf{P} \in \mathbb{R}^{m \times m}$ is positive definite symmetric, then

$$\mathbf{J}_\gamma(\xi) = \xi^T \mathbf{P} \quad \text{and} \quad \mathbf{H}_C \equiv \mathbf{P}, \quad (7.382)$$

and the resulting iterations are

$$\mathbf{x}^{(j+1)} = \mathbf{x}^{(j)} - [\mathbf{J}_F^T(\mathbf{x}^{(j)}) \mathbf{P} \mathbf{J}_F(\mathbf{x}^{(j)})]^{-1} \mathbf{J}_F^T(\mathbf{x}^{(j)}) \mathbf{P} \mathbf{F}(\mathbf{x}^{(j)}). \quad (7.383)$$

□

In the notation of observer O_2 , set

$$\mathbf{F}(\hat{\mathbf{x}}(k)) := \mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}(k)), \quad (7.384)$$

then with $\gamma(\cdot)$ as defined in (7.381),

$$C(\mathbf{x}) := \gamma(\mathbf{F}(\hat{\mathbf{x}}(k))). \quad (7.385)$$

which gives $C(\cdot)$ as the original cost function in (7.323). However, the single Newton step in (7.380) attempts to achieve

$$\mathbf{J}_C^T(\hat{\mathbf{x}}(k)) = -\mathbf{J}_h^T(\hat{\mathbf{x}}(k)) \mathbf{P}_k [\mathbf{z}(k) - \mathbf{h}(\hat{\mathbf{x}}(k))] = \mathbf{0}. \quad (7.386)$$

Recognizing that

$$\mathbf{s}_k := -\mathbf{J}_C^T(\hat{\mathbf{x}}(k)), \quad (7.387)$$

the single Newton step is modified to incorporate the desired reaching law

$$\mathbf{s}_k = \mathbf{g}(k-1, \mathbf{s}_{k-1}), \quad (7.388)$$

by treating $\mathbf{g}(k-1, \mathbf{s}_{k-1})$ as a constant in the iterations and using the one-step Newton method towards computing a zero of

$$\mathbf{J}_C^T(\hat{\mathbf{x}}(k)) + \mathbf{g}(k-1, \mathbf{s}_{k-1}) = \mathbf{0}. \quad (7.389)$$

This modification requires that (7.377) be replaced with (7.389). The modified Gauss-Newton iterations then become, with $\mathbf{J}_F(\hat{\mathbf{x}}) = -\mathbf{J}_h(\hat{\mathbf{x}})$, and using (7.381), (7.382), (7.384), and (7.389),

$$\begin{aligned} \hat{\mathbf{x}}^{(j+1)}(k+1) = & \hat{\mathbf{x}}^{(j)}(k+1) + \left[\mathbf{J}_h^T(\hat{\mathbf{x}}^{(j)}(k+1)) \mathbf{P}_{k+1} \mathbf{J}_h(\hat{\mathbf{x}}^{(j)}(k+1)) \right]^{-1} \\ & \cdot \left\{ \mathbf{J}_h^T(\hat{\mathbf{x}}^{(j)}(k+1)) \mathbf{P}_{k+1} \left[\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}^{(j)}(k+1)) \right] \right\} \end{aligned} \quad (7.390)$$

which is precisely the proposed iterated form of observer O_2 .

At the present time, stability results for the iterated observer are not available. Preliminary work towards this goal has indicated that, just as in Section 7.3.3, the primary difficulties include the following items: 1) the proposed local iterations will generally have a finite region of attraction, and this region must be of sufficient extent at each measurement update; and 2) during the reaching phase residual errors from previous measurement updates propagate which means that the local iterations, even if they converge, need not converge at each measurement update precisely to the ideal observer output. Performance of the iterated observer, however, can be monitored through the value of $\mathbf{w}_{k+1}^{(j)}$ in (7.363) since this quantity determines, from Lemma 7.8, how well the iterated observer approximates the ideal reaching law. Further effort is required to obtain a formal stability result, and should also consider alternatives to the standard Gauss-Newton approach, such as the use of continuation methods and step-length algorithms, which may extend the region of attraction significantly.

7.5 Simulation Results for Motion and Structure Estimation

Level IV of the hierarchical image analysis system proposed in Figure 1.1 (page 3) is concerned with estimation of object motion and structure based on solutions to 3D tracking and correspondence problems for feature point positions in Level V. Results for the 3D feature point tracking problem were previously shown in applications of the simple sliding mode observer of Chapter 4 and in applications of the proposed sliding mode observers of Chapter 6. In this Section, example simulation results are shown from applications of the sliding mode observers proposed in this Chapter, for both nonredundant and redundant measurements, in the problem of estimating motion and structure of a translating and rotating block (see Figure 2.4, page 51), during the object trajectory described in Case 1 of Section 2.6.2.

Image plane positions of four feature points on the object (p^3, p^5, p^6, p^7) are provided with correspondence information for multiple cameras from the imagery generation program. Motion and structure models for the filters are constructed from methods proposed in Chapter 2 with rotational motion being represented by the angle-axis parameterization. The same ordering of *special* feature points to fix the object-centred reference frame on the object, and the same Gauss-Markov models for translational and rotational motion developed in Chapter 2 are used to approximate the true object motion. The imaging geometry is identical to that used in Section 6.6 with two cameras (C_1 , and C_2 of Figure 6.3, page 206) for nonredundant measurements, and three cameras (C_1 , C_0 , and C_2) for the redundant case. A modest level of uniformly distributed measurement errors, on the order of 2.5% of the object image size, in the image plane are assumed. Derived 3D measurements were then obtained over the entire trajectory using Step 1 of the single-frame initialization algorithm of Section 2.5.

With four feature points of interest, using the angle-axis parameterization, and retaining first and second time derivatives of position, $\mathbf{p}(t) := [\mathbf{T}_O^E]_E(t)$, and orientation, $\boldsymbol{\xi}(t)$ (in the notation of Chapter 2), the extended state vector, $\mathbf{x}_e$, has 24 states arranged as

$$\mathbf{x}_e := [\mathbf{p}^T, \boldsymbol{\xi}^T, \mathbf{x}_s^T, \dot{\mathbf{p}}^T, \dot{\boldsymbol{\xi}}^T, \ddot{\mathbf{p}}^T, \ddot{\boldsymbol{\xi}}^T]^T, \quad (7.391)$$

where $\mathbf{x}_s$ denotes the constant six-element structure state vector to be estimated².

Simulation results for this Chapter are presented in two Cases:

Case 1: Nonredundant Nonlinear Measurements. Measurements from binocular stereo cameras are backprojected into 3D to give derived measurements representing the positions of the four feature points. The nonredundant iterated sliding mode observer proposed in Section 7.3 is employed to estimate position, velocity and acceleration of the origin of the object-centred reference frame, orientation and rotational motion of this frame, and position vectors of object features with respect to the object-centred frame. Example results for no measurement noise illustrate ideal performance for this observer, and Monte-Carlo simulation results represent expected performance in the presence of modest measurement errors.

Case 2: Redundant Nonlinear Measurements. A third camera is added to the configuration of Case 1 using the same camera geometry as employed in Section 6.6. Three sets of derived redundant measurements are provided through backprojection of measurements taken in any two of the three image planes. The redundant iterated sliding mode observer proposed in Section 7.4 is employed to estimate translational and rotational motion and object structure. Monte-Carlo simulation results illustrate expected performance of the redundant sliding mode observer and the relative improvements over the nonredundant Case 1.

Case 1: Nonredundant Nonlinear Measurements. In this Case, the measured state, $\mathbf{x} \in \mathbb{R}^{12}$, is defined by $\mathbf{x}(k) := \mathbf{C}\mathbf{x}_c(k)$, where $\mathbf{C} \in \mathbb{R}^{12 \times 24}$ is given by

$$\mathbf{C} := [\mathbf{I}_{12}, \mathbf{0}_{12 \times 12}]. \quad (7.392)$$

The measurement model is expressed, dropping the measurement noise term, as

$$\mathbf{z}(k) = \mathbf{h}(\mathbf{x}(k)) = \begin{bmatrix} \mathbf{h}^1(\mathbf{x}(k)) \\ \mathbf{h}^2(\mathbf{x}(k)) \\ \mathbf{h}^3(\mathbf{x}(k)) \\ \mathbf{h}^4(\mathbf{x}(k)) \end{bmatrix}. \quad (7.393)$$

²Recall that six elements from the total twelve structural parameters are held constant at known values satisfying the structural constraints.

where $\mathbf{h}^i(\mathbf{x}(k))$, $i = 1, 2, 3, 4$, denotes the 3D position measurements of the i th feature point,

$$\mathbf{h}^i(\mathbf{x}(k)) := \mathbf{I}_E^O(\boldsymbol{\xi}(k))\mathbf{r}_O^i + \mathbf{p}(k), \quad (7.394)$$

where $\mathbf{r}_O^i$, $i = 1, 2, 3, 4$, are the constant structure vectors representing positions of the i th feature point with respect to the object-centred reference frame. The Jacobian, $\mathbf{J}_h(k)$, of $\mathbf{h}(\mathbf{x}(k))$ can be computed using methods outlined in Appendix D.

Assuming that an estimate $\hat{\mathbf{x}}_e(k)$ is available, the iterated nonredundant sliding mode observer, at time step $k + 1$, takes the form

Initialization:

$$\begin{aligned} \mathbf{s}_k &:= \mathbf{z}(k) - \mathbf{h}(\mathbf{C}\hat{\mathbf{x}}_e(k)), \\ \mathbf{g}(\mathbf{s}_k) &= \mathbf{\Gamma} \text{Sat}(\mathbf{s}_k, \epsilon) + \mathbf{\Lambda}\mathbf{s}_k, \\ \hat{\mathbf{x}}_e^{(0)}(k+1) &:= \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)) \end{aligned} \quad (7.395)$$

Local Iterations: With measurements $\mathbf{z}(k+1)$ available and for k fixed, $N_{k+1} \geq 1$, and with $j = 0, 1, \dots, N_{k+1} - 1$,

$$\begin{aligned} \mathbf{s}_{k+1}^{(j)} &:= \mathbf{z}(k+1) - \mathbf{h}(\mathbf{C}\hat{\mathbf{x}}_e^{(j)}(k+1)), \\ \mathbf{r}_{k+1}^{(j)} &= \mathbf{g}(\mathbf{s}_k) - \mathbf{s}_{k+1}^{(j)}, \\ \mathbf{v}^{(j)}(k) &= [\mathbf{J}_h(\mathbf{C}\hat{\mathbf{x}}_e^{(j)}(k+1))\mathbf{C}\mathbf{L}^{(j)}(k)]^{-1} \mathbf{r}_{k+1}^{(j)}, \\ \hat{\mathbf{x}}_e^{(j+1)}(k+1) &= \hat{\mathbf{x}}_e^{(j)}(k+1) - \mathbf{L}^{(j)}(k)\mathbf{v}^{(j)}(k) \end{aligned} \quad (7.396)$$

Output:

$$\hat{\mathbf{x}}_e(k+1) := \hat{\mathbf{x}}_e^{(N_{k+1})}(k+1). \quad (7.397)$$

In this case, $\mathbf{s}_k \in \mathbb{R}^{12}$, $\mathbf{g} : \mathbb{R}^{12} \mapsto \mathbb{R}^{12}$, $\mathbf{v}^{(j)}(k) \in \mathbb{R}^{12}$, and $\mathbf{L} \in \mathbb{R}^{24 \times 12}$. Appendix E includes a MATLAB³ script file implementation of the iterated nonredundant observer.

The gain matrices $\mathbf{L}^{(j)}(k)$ are specified by defining another constant matrix $\Xi \in \mathbb{R}^{24 \times 12}$ satisfying

$$\mathbf{C}\Xi = \mathbf{I}_{12}, \quad (7.398)$$

³MATLAB is a trademark of The MathWorks, Inc.

and setting

$$\mathbf{L}^{(j)}(k) = \Xi \mathbf{J}_h^{-1} \left(\mathbf{C} \hat{\mathbf{x}}_e^{(j)}(k+1) \right). \quad (7.399)$$

In this case,

$$\mathbf{v}^{(j)}(k) := \left[\mathbf{J}_h \left(\mathbf{C} \hat{\mathbf{x}}_e^{(j)}(k+1) \right) \mathbf{C} \mathbf{L}^{(j)}(k) \right]^{-1} \mathbf{r}_{k+1}^{(j)} \quad (7.400)$$

$$= \mathbf{r}_{k+1}^{(j)}, \quad (7.401)$$

and

$$\hat{\mathbf{x}}_e^{(j+1)}(k+1) = \hat{\mathbf{x}}_e^{(j)}(k+1) - \Xi \mathbf{J}_h^{-1} \left(\mathbf{C} \hat{\mathbf{x}}_e^{(j)}(k+1) \right) \mathbf{r}_{k+1}^{(j)}. \quad (7.402)$$

Note that an efficient implementation need not compute the matrix inverse $\mathbf{J}_h^{-1}$ in each iteration since the system of linear equations

$$\mathbf{J}_h \left(\mathbf{C} \hat{\mathbf{x}}_e^{(j)}(k+1) \right) \mathbf{b}^{(j)}(k+1) = \mathbf{r}_{k+1}^{(j)} \quad (7.403)$$

could first be solved for the unknown vector $\mathbf{b}^{(j)}(k+1)$ (by Gaussian elimination, for example), and $\hat{\mathbf{x}}_e^{(j+1)}(k+1)$ can then be computed as

$$\hat{\mathbf{x}}_e^{(j+1)}(k+1) = \hat{\mathbf{x}}_e^{(j)}(k+1) - \Xi \mathbf{b}^{(j)}(k+1). \quad (7.404)$$

For simulations shown in this Chapter, elements of Ξ are chosen from desired gain levels in the first and second time derivative channels. This notion is illustrated by considering first-order approximations during sliding ($\mathbf{g} \equiv \mathbf{0}$).

$$\begin{aligned} \mathbf{r}_{k+1}^{(j)}|_{\mathbf{g} \equiv \mathbf{0}} &= -\mathbf{s}_{k+1}^{(j)} \\ &= -[\mathbf{z}(k+1) - \mathbf{h}(\hat{\mathbf{x}}^{(j)}(k+1))] \\ &\approx -\mathbf{J}_h(\hat{\mathbf{x}}^{(j)}(k+1))[\mathbf{x}(k+1) - \hat{\mathbf{x}}^{(j)}(k+1)], \end{aligned} \quad (7.405)$$

$$\Rightarrow \hat{\mathbf{x}}_e^{(j+1)}(k+1) \approx \hat{\mathbf{x}}_e^{(j)}(k+1) + \Xi [\mathbf{x}(k+1) - \hat{\mathbf{x}}^{(j)}(k+1)]. \quad (7.406)$$

Let $\alpha^{(s)}$ be the gain level in structure channels, and let

$$\begin{aligned} \boldsymbol{\alpha}^{(t)} &:= [\alpha_1^{(t)}, \alpha_2^{(t)}, \alpha_3^{(t)}]^T \\ \boldsymbol{\alpha}^{(r)} &:= [\alpha_1^{(r)}, \alpha_2^{(r)}, \alpha_3^{(r)}]^T, \end{aligned} \quad (7.407)$$

be the gain levels for translational (t) and rotational (r) channels. The matrix Ξ is then defined as

$$\Xi := \begin{bmatrix} \alpha_1^{(t)} \mathbf{I}_3 & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 6} \\ \mathbf{0}_{3 \times 3} & \alpha_1^{(r)} \mathbf{I}_3 & \mathbf{0}_{3 \times 6} \\ \mathbf{0}_{6 \times 3} & \mathbf{0}_{6 \times 3} & \alpha^{(s)} \mathbf{I}_6 \\ \alpha_2^{(t)} \mathbf{I}_3 & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 6} \\ \mathbf{0}_{3 \times 3} & \alpha_2^{(r)} \mathbf{I}_3 & \mathbf{0}_{3 \times 6} \\ \alpha_3^{(t)} \mathbf{I}_3 & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 6} \\ \mathbf{0}_{3 \times 3} & \alpha_3^{(r)} \mathbf{I}_3 & \mathbf{0}_{3 \times 6} \end{bmatrix}. \quad (7.408)$$

The constraint $\mathbf{C}\Xi = \mathbf{I}_m$ requires that $\alpha_1^{(t)} = \alpha_1^{(r)} = \alpha^{(s)} = 1$. Simulations of this Chapter employ

$$\alpha^{(s)} = 1 \quad (7.409)$$

$$\alpha^{(t)} = \alpha^{(r)} = [1, 8, 25]^T. \quad (7.410)$$

Local iterations are performed at each measurement update with termination whenever $N_k = N_{\max} := 10$ or $\|\mathbf{r}_{k+1}^{N_k}\|_2 < \epsilon_r := 0.001$.

Figure 7.1 shows sample simulation results for this observer in the absence of measurement errors, with $\mathbf{g} \equiv \mathbf{0}$, and with artificially high initial errors in position (~ 5 units), orientation ($\sim 5^\circ$), and structure (~ 3 units), and all temporal derivatives initialized as zero. These results provide an example of executing a single step to the surface. Convergence of the local iterations yield very accurate position estimates with small oscillations in velocity errors due to filter lag. Note the large transients in velocity estimation due to the artificially high initial errors in position and the single step to the surface. The convergence index for orientation estimation is again the scaled Frobenius norm, with $\tilde{\mathbf{I}}_E^O(k) := \mathbf{I}_E^O(k) - \hat{\mathbf{I}}_E^O(k)$,

$$C(k) := \frac{1}{2\sqrt{3}} \left(\text{Tr}\{\tilde{\mathbf{I}}_E^O(k)(\tilde{\mathbf{I}}_E^O(k))^T\} \right)^{\frac{1}{2}}, \quad (7.411)$$

which has a maximum possible value of 1. The convergence index, which remains within 0.002% of its maximum, indicates good performance of the locally iterated observer in orientation ($\hat{\xi}$) estimates. The Euclidean norm of angular velocity errors, which are computed using the available estimates $\hat{\xi}$ and $\hat{\dot{\xi}}$, remain within about 0.05 rad/s following initial transients. These transients are again quite large due to artificially high initial errors in orientation and the single step to the surface. Structure estimates are also well behaved, since the

relative error remains below 0.005% of true structural parameters. Over this trajectory, the maximum number of local iterations performed at any measurement update was 4, with 1 or 2 iterations being typical following initial transients.

Figures 7.2 and 7.3 show root-mean-squared errors (RMSE) and mean errors (ME) resulting from a 60-run Monte Carlo simulation with measurement errors added to image plane measurements. Translational motion estimation, shown in Figure 7.2, indicates very small bias levels (ME's) over the trajectory, with RMS errors for position remaining below 0.2 units while those for velocity estimates are under 2 units/s. Recall that for this trajectory the origin of the object-centred frame lies farthest from the cameras at about 3 seconds and 6 seconds. At these positions, errors in derived measurements obtained from back-projection of image plane measurements are larger than those present over other portions of the trajectory. The distinct peaks in RMS errors at these times are attributed to these higher state-dependent error levels in the derived measurements. Estimation of object orientation and rotational motion, shown in Figure 7.3, show good performance with the RMS convergence index remaining below 1.5% and RMS angular velocity error peaks, again at approximately 3 and 6 seconds, remaining below 0.35 rad/s. Structure estimation exhibits very small bias (ME's) with RMS errors on the order of 2.5% over the trajectory.

Figure 7.1: Example simulation results of sliding mode observer for motion and structure estimation through a nonlinear, non-redundant measurement model.

Figure 7.2: Monte Carlo simulation results of sliding mode observer for motion and structure estimation through a nonlinear, non-redundant measurement model; translational motion.

Figure 7.3: Monte Carlo simulation results of sliding mode observer for motion and structure estimation through a nonlinear, non-redundant measurement model; rotational motion and structure.

Case 2: Redundant Nonlinear Measurements. As discussed in Section 6.6, two primary motivations for the use of multiple base-line stereo (more than two cameras) are to achieve improved robustness in the correspondence problem and to reduce sensitivity to measurement errors through fusion of redundant measurements. In order to obtain redundant measurements in this Case, a third camera is added to the stereo pair employed in Case 1 as was done in Case 2 of Section 6.6. Three sets of derived 3D measurements for each of the four feature points of interest are obtained over the entire trajectory by taking corresponding feature positions from any two of the three cameras and using Step 1 of the single-frame initialization algorithm of Section 2.5. The third camera is placed at location C_0 shown in Figure 6.3, in which case similar error levels are present in each of the 3 sets of derived measurements.

Figures 7.4 and 7.5 show root-mean-squared errors (RMSE) and mean errors (ME) in estimation of translational and rotational motion and object structure over a 60-run Monte Carlo simulation for the three-camera triad (C_1, C_2, C_0) with imaging geometry identical to that used in Case 2 of Section 6.6. The true trajectory of the object is identical to that for Case 1, and approximate dynamical models for translational and rotational motion are again formulated in terms of Gauss-Markov acceleration models as was done in Case 1 above.

Let $\mathbf{z}_{ij}(k)$ represent the derived position measurements from camera pair (C_i, C_j) , $i, j = 0, 1, 2$. The redundant set of derived measurements, $\mathbf{z}_r(k) \in \mathbb{R}^{36}$, can be modelled, neglecting measurement errors, as

$$\begin{aligned}\mathbf{z}_r(k) &= [\mathbf{z}_{01}, \mathbf{z}_{02}, \mathbf{z}_{12}]^T \\ &= \mathbf{h}_r(\mathbf{x}(k)),\end{aligned}\tag{7.412}$$

where

$$\mathbf{h}_r(\mathbf{x}(k)) := \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \mathbf{z}(k)\tag{7.413}$$

and $\mathbf{z}(k)$ are the nonredundant measurements defined in Equations (7.393) and (7.394). Recall that the measured state $\mathbf{x}(k)$ is related to the extended state $\mathbf{x}_e(k)$ by $\mathbf{x}(k) = \mathbf{C}\mathbf{x}_e(k)$ where, as in Case 1 above, $\mathbf{C} := [\mathbf{I}_{12}, \mathbf{0}_{12 \times 12}]$.

Assuming that an estimate $\hat{\mathbf{x}}_e(k)$ is available, and that the weighting matrices $\mathbf{P}_k \equiv \mathbf{P}$

have been specified, the iterated redundant sliding mode observer, at time step $k + 1$, takes the form

Initialization:

$$\begin{aligned} \mathbf{s}_k &:= \mathbf{J}_{\mathbf{h}_r}^T (\mathbf{C}\hat{\mathbf{x}}_e(k)) \mathbf{P} [\mathbf{z}(k) - \mathbf{h}_r(\mathbf{C}\hat{\mathbf{x}}_e(k))], \\ \mathbf{g}(\mathbf{s}_k) &= \mathbf{\Gamma} \mathbf{Sat}(\mathbf{s}_k, \epsilon) + \mathbf{\Lambda} \mathbf{s}_k, \\ \hat{\mathbf{x}}_e^{(0)}(k+1) &:= \hat{\mathbf{f}}(k, \hat{\mathbf{x}}_e(k)). \end{aligned} \quad (7.414)$$

Local Iterations: With redundant measurements $\mathbf{z}_r(k+1)$ available and for k fixed, $N_{k+1} \geq 1$, and with $j = 0, 1, \dots, N_{k+1} - 1$,

$$\begin{aligned} \mathbf{s}_{k+1}^{(j)} &= \mathbf{J}_{\mathbf{h}_r}^T (\mathbf{C}\hat{\mathbf{x}}_e^{(j)}(k+1)) \mathbf{P} [\mathbf{z}(k+1) - \mathbf{h}_r(\mathbf{C}\hat{\mathbf{x}}_e^{(j)}(k+1))], \\ \mathbf{w}_{k+1}^{(j)} &= \mathbf{g}(\mathbf{s}_k) - \mathbf{s}_{k+1}^{(j)}, \\ \mathbf{v}^{(j)}(k) &= [\mathbf{J}_{\mathbf{h}_r}^T (\mathbf{C}\hat{\mathbf{x}}_e^{(j)}(k+1)) \mathbf{P} \mathbf{J}_{\mathbf{h}_r} (\mathbf{C}\hat{\mathbf{x}}_e^{(j)}(k+1)) \mathbf{C} \mathbf{L}(k)]^{-1} \\ &\quad \cdot \mathbf{w}_{k+1}^{(j)}, \\ \hat{\mathbf{x}}_e^{(j+1)}(k+1) &= \hat{\mathbf{x}}_e^{(j)}(k+1) - \mathbf{L}^{(j)}(k) \mathbf{v}^{(j)}(k). \end{aligned} \quad (7.415)$$

Output:

$$\hat{\mathbf{x}}_e(k+1) := \hat{\mathbf{x}}_e^{(N_{k+1})}(k+1). \quad (7.416)$$

In this case, $\mathbf{s}_k \in \mathbb{R}^{12}$, $\mathbf{g} : \mathbb{R}^{12} \mapsto \mathbb{R}^{12}$, $\mathbf{v}^{(j)}(k) \in \mathbb{R}^{12}$, and $\mathbf{L} \in \mathbb{R}^{24 \times 12}$. The weighting matrix $\mathbf{P} \in \mathbb{R}^{36 \times 36}$ is taken as the identity matrix since derived measurements for the given camera geometry have approximately the same error levels. The termination criteria for local iterations is equivalent to that used in the nonredundant Case 1 (with $\|\mathbf{w}_{k+1}^{(j)}\|_2$ in above iterations replacing $\|\mathbf{r}_{k+1}^{(j)}\|_2$ employed in Case 1). Appendix E includes a MATLAB script file implementation of the iterated redundant observer.

The gain matrices $\mathbf{L}^{(j)}(k)$ are selected in an equivalent manner as that described for the nonredundant Case 1. A matrix $\mathbf{\Xi}$ is defined satisfying

$$\mathbf{C}\mathbf{\Xi} = \mathbf{I}_{12}, \quad (7.417)$$

and the gain matrices $\mathbf{L}^{(j)}(k)$ are set as

$$\mathbf{L}^{(j)}(k) = \mathbf{\Xi} [\mathbf{J}_{\mathbf{h}_r}^T (\mathbf{C}\hat{\mathbf{x}}_e^{(j)}(k+1)) \mathbf{P} \mathbf{J}_{\mathbf{h}_r} (\mathbf{C}\hat{\mathbf{x}}_e^{(j)}(k+1))]^{-1}. \quad (7.418)$$

In this case,

$$\mathbf{v}^{(j)}(k) = \mathbf{w}_{k+1}^{(j)}. \quad (7.419)$$

and an approximate first-order analysis during sliding yields

$$\hat{\mathbf{x}}_e^{(j+1)}(k+1) \approx \hat{\mathbf{x}}_e^{(j)}(k+1) + \Xi [\mathbf{x}(k+1) - \hat{\mathbf{x}}^{(j)}(k+1)], \quad (7.420)$$

which corresponds to the result of Equation (7.406). Simulations of this Section use this gain selection strategy with Ξ taken as shown in Equation (7.408).

Figures 7.4 and 7.5 show root-mean-squared errors (RMSE) and mean errors (ME) resulting from a 60-run Monte Carlo simulation with measurement errors added to image plane measurements. With the three sets of redundant measurements, and with similar error levels in each redundant set, it is expected that RMS error levels from the iterated redundant observer be reduced by a factor of approximately $1/\sqrt{3}$ over those for the nonredundant case. Results shown in Figures 7.4 and 7.5 approximately support this predicted performance improvement, keeping in mind that the local iterations are now of a Gauss-Newton form. Over this trajectory, the maximum number of local iterations performed at any measurement update was 5, with 1 or 2 iterations again being typical following initial transients. Translational motion estimation, shown in Figure 7.4, indicates very small bias levels (ME's) over the trajectory, with RMS errors for position remaining below 0.18 units while those for velocity estimates are under 1.8 units/s. Again, distinct peaks are seen in RMS errors at approximately 3 and 6 seconds when the origin of the object-centred frame lies farthest from the cameras. Estimation of object orientation and rotational motion, shown in Figure 7.3, show good performance with the RMS convergence index remaining below 1% and RMS angular velocity error peaks remaining below 0.25 rad/s. Structure estimation exhibits very small bias (ME's) with RMS errors on the order of 2% over the trajectory.

Figure 7.4: Monte Carlo simulation results of sliding mode observer for motion and structure estimation through a nonlinear, redundant measurement model; translational motion.

Figure 7.5: Monte Carlo simulation results of sliding mode observer for motion and structure estimation through a nonlinear, redundant measurement model; rotational motion and structure.

7.6 Chapter Summary

Sliding mode observers proposed and analyzed in this chapter are specific to that class of state estimation problems which are governed by possibly nonlinear time-varying dynamics, known up to some bounded propagation error over any sample period, and which provide measurements that may be nonlinear with respect to elements of the extended state vector. Existence of the projection operation which provides the robustness mechanism of the sliding mode requires alternate observer forms depending on whether the measurement model is minimal or redundant. Concepts of nonlinear measurement redundancy were proposed which result in a rank test to determine which form of sliding mode observer is required.

The *ideal* observers for nonredundant or redundant measurement cases take on very similar forms as those proposed for the case of linear measurements treated in Sections 6.4 and 6.5, respectively. In the case of linear measurement models these ideal forms can be realized directly. In the nonlinear case, however, ideal forms for the observers represent systems of nonlinear integral equations which implicitly define the observer output. A simple implementation method, using an approximation similar to that employed in the Extended Kalman Filter, can yield good performance in some cases. More reliable performance is obtained through introduction of local iterations at each observation event, similar to those employed in the Iterated Extended Kalman Filter. These local iterations yield Newton-type iterations for nonredundant measurements and Gauss-Newton-type iterations for redundant measurements. The ideal observer forms, together with stability and robustness properties were given for both nonredundant and redundant observers, followed by proposed implementation methods with local iterations. Stability analysis, which yields sufficient conditions for the numerical implementation to approximate the ideal observer with arbitrary accuracy, was given for the iterated nonredundant observer. Further analysis is required to obtain stability for the iterated redundant observer and to obtain more precise quantitative results for observer performance in the presence of measurement errors.

Chapter 8

Conclusions and Future Research

Research presented in this thesis focussed on design, analysis, and implementation of observers, or state estimators, for dynamical systems in which measurements are available at discrete time instants. Novel approaches, based on the sliding mode concept, were proposed for observer design. New methods for discrete-time stability analysis have been proposed and applied to obtain sufficient conditions for stability of the proposed state estimation techniques. The theory and state estimation algorithms developed in the thesis can be applied to a broad class of systems. Simulation studies of the proposed methods were provided to assess observer performance in various levels of a hierarchical approach to obtain motion and structure of objects based on measurements of feature positions in long multiple-camera image sequences. This chapter summarizes the thesis work, draws conclusions from previous chapters, and discusses limitations of the proposed methods and several promising areas for future research.

Design methodologies for observer implementation require first that mathematical models for the system dynamics and available measurements be obtained in state space form. The objective of observer design is then to employ the available measurements, together with the derived models for dynamics and measurements, to estimate the full state of the system. Generally, measurements available at the k th sample time are directly related to the values of some, but not all, of the states of the system. This property of the measurement model is explicitly considered by defining the full or *extended* state, $\mathbf{x}_e \in \mathbb{R}^n$, of the system and extracting from these states only those elements, $\mathbf{x} \in \mathbb{R}^q$, $q < n$, upon which the m measurements, $m \leq n$, provided by the system are dependent at each sample time.

The vector $\mathbf{x}$ is referred to as the *measured* state and is related to the extended state according to $\mathbf{x} = \mathbf{C}\mathbf{x}_e$, where $\mathbf{C} := [\mathbf{I}_q, \mathbf{0}_{q \times (n-q)}]$. Perhaps the most common situation is when the extended state, $\mathbf{x}_e$, consists of the measured state $\mathbf{x}$ and its first N time derivatives, for some integer N .

Proposed observer design and analysis methodologies assume that the true dynamical and measurement models, in theory, can be expressed as

$$\begin{aligned} \mathbf{x}_e(k+1) &= \mathbf{f}(k, \mathbf{x}_e(k)) \quad \sim \text{Dynamics} \\ \mathbf{z}(k) &= \mathbf{h}(\mathbf{x}(k)) \quad \sim \text{Measurements,} \end{aligned} \quad (8.1)$$

where $\mathbf{x}_e \in \mathbf{X}_e \subseteq \mathbb{R}^n$, and $\mathbf{f} : \mathbb{N}_0 \times \mathbf{X}_e \mapsto \mathbf{X}_e$ is C^p in $\mathbf{x}_e$, $p \geq 1$.

Applications and analysis of the proposed observers require that a model, $\hat{\mathbf{f}} : \mathbb{N}_0 \times \mathbb{R}^n \mapsto \mathbb{R}^n$, of the true dynamics is available such that $\forall \mathbf{x}_e \in \mathbf{X}_e$,

$$\mathbf{f}(k, \mathbf{x}_e) - \hat{\mathbf{f}}(k, \mathbf{x}_e) = \mathbf{B}(k)\mathbf{d}(k, \mathbf{x}_e), \quad (8.2)$$

where $\mathbf{B}(k) \in \mathbb{R}^{n \times \ell}$ are known matrices with bounded norm uniformly in k , and $\mathbf{d} : \mathbb{N}_0 \times \mathbf{X}_e \mapsto \mathbb{R}^\ell$ is bounded over $\mathbf{X}_e$ uniformly in k . This means that the difference between the true time propagation of the system over one sample period from any point $\mathbf{x}_e \in \mathbf{X}_e$ and the approximate propagation using the model $\hat{\mathbf{f}}$ from that same point $\mathbf{x}_e$ is bounded in norm. This property of the system model is referred to as, “bounded propagation error over any sample period.” The matrices $\mathbf{B}(k)$ provide a mechanism to exploit available knowledge of the structure of this propagation error.

Practical applications of the proposed methods do not require specifically that the system dynamics be expressed in discrete-time closed form as shown in (8.1); implementation requires only that the estimated extended state, $\hat{\mathbf{x}}_e(k)$, can be propagated to the $(k+1)$ st sample time, and that this time propagation method satisfies the propagation error criterion. Numerical integration of a model of the underlying continuous-time dynamics is, therefore, a reasonable approach for implementation. Systems represented by continuous-time dynamics and discrete-time measurements are commonly called *sampled data* systems. Analysis and implementation of extensions of the proposed methods to sampled data systems is a promising area for future research that could extend significantly the class of systems to which the proposed methods can be applied.

Construction of an observer begins with the modelling problem which focusses on obtaining, in state space form, mathematical representations for the dynamics and measurements of the system under consideration. Proposed theoretical results in preceding chapters are quite general in that they are not specific to any particular state estimation problem. Applications of the various observer design and stability methods were presented in different levels of the proposed hierarchical feature-based image analysis system.

The modelling problem for feature-based motion and structure estimation of a manoeuvring object observed with a multiple-camera imaging system was treated in Chapter 2. Both translational and rotational motion were described with first-order Gauss-Markov processes which are completely defined by selection of a small number of parameters. This work led [22, 23] to the development of motion, structure, and measurement models for three parameterizations of rotational motion, Euler angle-axis, quaternion, and roll-pitch-yaw, and a statistical performance assessment and comparison of the extended Kalman filter for the three models. The work not only demonstrates how previously proposed quaternion-based approaches might be extended to track manoeuvring objects observed in multiple-camera image sequences, but also provides two much simpler alternatives using roll-pitch-yaw and angle axis parameterizations of rotational motion. The angle axis filter based on an approximate linear dynamical model gives performance for all trajectories considered that is at least as good as, and often superior to, the quaternion-based scheme.

A primary observation of this application of the extended Kalman filter was that, although the filters appear to work well in some cases, sensitivity to dynamical modelling errors was a major problem which could lead to filter instability. As a consequence, results obtained from one trajectory say very little about results that could be expected from different trajectories, particularly with only scarce stability theory available in the literature for the extended Kalman filter in the presence of modelling errors. This observation provided the initial motivation for development of state estimation methods which can exhibit the desirable robustness properties.

Robustness of an observer to modelling error, or simply robustness to unmodelled dynamics, is a very important issue in observer design. Another problem is that some errors are always present in sensor measurements. It is often the case that the two desirable properties, robustness to unmodelled dynamics and insensitivity to measurement errors, present

conflicting requirements in observer design; very simply, observer gain levels must often be kept high to yield robustness to unmodelled dynamics, but low to reduce sensitivity to measurement noise. In this thesis, emphasis has been placed primarily on the problem of robustness to unmodelled dynamics, although some effort, particularly through the use of redundant measurements, was devoted to the consideration of effects of measurement errors on the proposed observers.

As reviewed in Chapter 3, a dynamical system with a *sliding mode* is one in which the states of the system, after some time period, are restricted to lie in some manifold, usually a subspace or hypersurface, in the state space. Starting from an initial state off of the surface, the state reaches the surface in finite time and thereafter is constrained to lie on, or slide along, the surface. This sliding action imposes certain constraints on the dynamics of the system during the sliding mode. Extensions to the discrete-time case of the usual continuous-time sliding mode conditions do not necessarily guarantee chattering motions close to the sliding surface, and may lead to unstable chattering about the sliding surface. Discrete-time systems, however, can exhibit sliding behavior with continuous systems of difference equations, which is in sharp contrast to the requirement of discontinuities in the continuous-time case. Although some effort has been devoted in the literature to the nonlinear discrete-time sliding mode controller problem, discrete-time sliding mode observers, particularly for nonlinear systems, have not previously appeared.

The primary advantage of the sliding mode method for continuous-time systems is that it offers an inherent robustness to parametric uncertainty and unmodelled dynamics. In these systems, the robustness mechanism, as discussed in Chapter 3, results from a projection of the perturbed dynamics onto the sliding surface through application of the control law. A similar, or "dual" projection mechanism occurs in the continuous-time sliding mode observer approach in that perturbed observer error dynamics are projected onto the sliding surface. Direct extensions of continuous-time methods to the discrete-time case, however, do not generally retain these robustness properties. Chapter 4 illustrated the basic properties of a discontinuous, discrete-time implementation of an available continuous-time method and led to a simple observer form for linear systems. This observer was constructed from a steady-state discrete-time Kalman filter with an additional switching term. Although simulation results for this observer are encouraging, direct applications of the method to more

general nonlinear systems with modelling errors are not apparent. The primary difficulty results from dependence of the sliding surface and convergence criteria on the unknown modelling errors and the fact that, as shown in Chapter 4, a convergent quasi-sliding regime cannot exist in the presence of discontinuous switching.

Further analysis conducted in Chapter 4 led [24] to a structure for discrete-time observers which, for linear systems, not only maintains sliding on the prescribed surface, but also retains the robustness mechanism of continuous-time sliding mode methods. This form of observer was shown to have sliding dynamics equivalent to a conventional reduced-order Luenberger observer for the discrete-time linear system. A primary objective of Chapters 6 and 7 was to extend the theory of reduced-order observers from the strictly linear case, treated in Chapter 4, to the case of nonlinear uncertain dynamics (Chapter 6) and nonlinear measurement models (Chapter 7) using the sliding mode concept.

In the observer problem, the existence of a quasi-sliding, convergent quasi-sliding, or even discrete-time sliding regime on a prescribed surface does not necessarily imply stability of the observer. The reason for this is that the sliding surface in the observer problem is defined in terms of measurements available at a single time instant, and these measurements are related to some but not all states of the system at that time instant. As a result, even if the error trajectories can be maintained on the prescribed surface, there generally exists reduced-order dynamics on the surface that must be checked for stability. Stability analysis of the proposed observers, however, using standard Lyapunov methods for discrete-time systems did not yield sufficient information concerning restrictions on the accuracy of initial estimates nor ultimate bounds on the observer error.

Chapter 5 presented new exponential stability results for discrete-time systems through the use of vector norm Lyapunov-like functions, and applications of this theory to observer design. Results for nonlinear systems and observer stability proposed in Chapter 5 include significant generalizations over those published by the candidate in References [25, 26]. Theoretical results of Chapter 5 include necessary and sufficient conditions for exponential stability of linear time invariant and time varying systems, and sufficient conditions for local exponential asymptotic stability of nonlinear systems. These conditions also yield a domain of exponential stability and exponential bounds on the state. Extensions of the theory to obtain stability of a class of observers for uncertain discrete-time nonlinear sys-

tems were developed in which emphasis is placed on sufficient conditions for exponentially bounded observer errors, and estimation of a domain of exponential stability and reasonably tight error bounds. The class of observers to which these stability methods can be applied includes the ideal observer forms proposed in Chapters 6 and 7. Sufficient generalization was also included in developments of the stability theory so that the methods could also be applied to analysis of observer implementation, detailed in Chapter 7, with local iterations.

The basis for the proposed stability approach is to search for suitable vector and matrix norms, and a state transformation under which the transformed system represents, at least locally, a contraction with respect to the prescribed norms. Stability results for the original system can then be inferred from those of the transformed system. This approach, together with the use of vector norm Lyapunov, or Lyapunov-like functions, can not only demonstrate exponential stability, but also provide a domain of exponential stability and exponential bounds. Applications of this approach in stability analysis of discrete-time observers are of particular interest in that some measure of admissible initialization errors, under which the observer error is exponentially bounded, can, in principle, be determined. The availability of such information, however, is dependent on the ability to specify suitable state transformations which, with respect to the chosen vector and matrix norms, not only satisfy the sufficient conditions for stability, but also give a sufficiently large stability domain and tight enough bounds to be of practical use.

One area for future work with the stability approach is to determine if numerical methods, such as numerical minimization, can be applied in a search for state transformations with the required properties for more general time-varying and/or nonlinear systems. A promising approach may be to begin with stability by first approximation (Corollary 5.1) which yields time-varying linear systems along a given trajectory. The proposed stability conditions are both necessary and sufficient for linear time varying systems. In this way, it may be possible to obtain appropriate linear state transformations locally over the state space. Significant effort, however, may be required to determine how these local transformations might be combined, or scheduled, in order to infer stability of the original system.

A second area for future work could focus on relaxing the convexity requirement in the proposed methods for stability of nonlinear systems (Theorem 5.10) with the objective of enlarging the class of systems to which the theory can be applied and the stability domain.

Recall that in Theorem 5.10 (page 149), Ω denotes the set in which the stability hypothesis must hold and represents a stability domain for the system. The set Ω , in the current approach, contains the origin and is assumed to be convex. This convexity assumption leads to the desired result through direct use of the fundamental theorem of integral calculus. In Theorem 5.10, however, application of the hypothesis requires only that all points along the line connecting the origin to the state be contained within Ω . Hence, Ω could be taken as any set which is *star-like* with respect to the origin¹ provided the hypothesis hold within this set. A further generalization may also be possible. Lemma 7.1 illustrates a type of path independence property of the fundamental theorem of integral calculus. Beginning with extensions of Lemma 7.1, preliminary work suggests that it may be possible to replace convexity of Ω , or even the star-like alternative, with a path connected property, but retaining $0 \subseteq \Omega$. If successful, this approach may lead to a means to replace the convexity requirement of Ω in observer stability (Theorem 5.11) with a path connected property which, in turn, may reduce the minimum admissible size of Ω . Reducing the minimum admissible size of Ω for the observer stability problem may lead to significantly tighter bounds on observer errors in applications to a broader class of systems.

Chapter 5 also proposed a reaching law method for design of discrete-time sliding mode observers. This method is a direct extension to the discrete-time case of the reaching law method previously proposed by Gao and Hung [123] for continuous-time sliding mode control problems. It must be emphasized, however, that because not all states are directly available in the observer problem, this reaching law design method generally does not completely define the reaching dynamics of the *extended* observer error $e_e(k)$. This property of the observer problem places significant emphasis on stability analysis during both the reaching and sliding phases.

It was previously mentioned in Section 5.7 (see page 169), that the discrete-time reaching law methods proposed here for the observer problem could also have applications in the discrete-time sliding mode control problem. Indeed, many of the concepts developed in this thesis, including the stability theory, carry over to a certain degree to the discrete-time sliding mode control problem. This represents a very promising area for future work.

¹ All points on the line connecting any point in Ω to the origin also lie in Ω , see for example [145].

For example, denote the (full) state of a system by $y(k) \in \mathbb{R}^n$ and assume that all states are available for use in the control law. Let the true dynamics of the system, with scalar input $u(k)$, be given by

$$y(k+1) = f(k, y(k)) + Du(k). \quad (8.3)$$

Define the desired sliding surface as $s_k := s(y(k)) = 0$ as

$$s_k = \Gamma y(k), \quad (8.4)$$

where $\Gamma \in \mathbb{R}^{1 \times n}$ has been selected so that ΓD is nonsingular (a relative degree constraint). Let the desired reaching law be given by

$$s_{k+1} = g(k, s_k), \quad (8.5)$$

where $g(k, s_k)$ is a scalar version of the reaching law dynamics defined and analyzed in Section 5.7. If the true dynamics f are available, then the control law

$$u_k := -[\Gamma D]^{-1} \{ \Gamma f(k, y(k)) - g(k, s_k) \} \quad (8.6)$$

satisfies the reaching law and leads to pure sliding, say for $k > k_s$, if $g(k, s_k) = 0$ for all $k \geq k_s$.

Now assume that only a model, $\hat{f}$, of the true dynamics are available, and that control system design is based on this model. Let $\hat{y}(k)$ be the resulting controlled state

$$\hat{y}(k+1) = \hat{f}(k, \hat{y}(k)) + D\hat{u}(k) \quad (8.7)$$

under the control law

$$\hat{u}(k) := -[\Gamma D]^{-1} \{ \Gamma \hat{f}(k, \hat{y}(k)) - g(k, \hat{s}_k) \}, \quad (8.8)$$

where $\hat{s}_k := \Gamma \hat{y}(k)$. This system satisfies the reaching law

$$\hat{s}_{k+1} = g(k, \hat{s}_k). \quad (8.9)$$

In this case, the error, $e(k) := y(k) - \hat{y}(k)$, due to the use of the control law based on the system model has sliding dynamics (i.e. with $g(k, s_k) \equiv 0$) given by

$$e(k+1) := y(k+1) - \hat{y}(k+1)$$

$$\begin{aligned}
&= [\mathbf{f}(k, \mathbf{y}(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{y}}(k))] - \mathbf{D}[\hat{\mathbf{u}}(k) - \mathbf{u}(k)] \\
&= [\mathbf{f}(k, \mathbf{y}(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{y}}(k))] \\
&\quad - \mathbf{D}[\mathbf{D}\mathbf{D}]^{-1}\mathbf{D}[\mathbf{f}(k, \mathbf{y}(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{y}}(k))] \\
&= \mathbf{F}[\mathbf{f}(k, \mathbf{y}(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{y}}(k))], \tag{8.10}
\end{aligned}$$

where

$$\mathbf{F} := [\mathbf{I}_n - \mathbf{D}[\mathbf{D}\mathbf{D}]^{-1}\mathbf{D}] \tag{8.11}$$

represents a projection along the column space of $\mathbf{D}$ onto the null space of $\mathbf{D}$. The sliding surface is the null space of $\mathbf{D}$. The dynamics in (8.10) have a form which is “dual” to the error dynamics of sliding mode observers proposed in the thesis (see also the discussions on “duality” in the continuous-time sliding mode observer and controller problems, page 99).

In order to illustrate the “dual” matching condition for the robustness property, suppose that for all ξ in the state space

$$\mathbf{f}(k, \xi) - \hat{\mathbf{f}}(k, \xi) = \mathbf{B}\mathbf{d}(k, \xi). \tag{8.12}$$

and that $\mathbf{B}$ and $\mathbf{d}(k, \xi)$ satisfy the bounded propagation error criteria. Then adding and subtracting $\hat{\mathbf{f}}(k, \mathbf{y}(k))$ in (8.10) gives

$$\mathbf{e}(k+1) = \mathbf{F}\mathbf{B}\mathbf{d}(k, \mathbf{y}(k)) + \mathbf{F}[\hat{\mathbf{f}}(k, \mathbf{y}(k)) - \hat{\mathbf{f}}(k, \hat{\mathbf{y}}(k))]. \tag{8.13}$$

This is precisely the form of error dynamics treated repeatedly in the thesis; if $\mathbf{B} = \mathbf{D}\mathbf{\Xi}$ for some $\mathbf{\Xi} \in \mathbb{R}^{1 \times \ell}$ (i.e. the column space of $\mathbf{B}$ lies in the column space of $\mathbf{D}$), then $\mathbf{F}\mathbf{B} = \mathbf{0}$ and the error during sliding is invariant to the modelling error. Moreover preliminary work suggests that the discrete-time stability methods developed for the observer problem in Chapter 5 need only slight modifications to treat this form of the discrete-time sliding mode control problem. A more complete treatment of the discrete-time sliding mode controller problem using the methods developed for the observer problem in this thesis is a promising area for future research.

Sliding mode observers proposed and analyzed in Chapter 6 are specific to that class of state estimation problems which are governed by possibly nonlinear time-varying dynamics, known up to some bounded propagation error over any sample period, and which

provide measurements that are linearly related to elements of the extended state vector. Existence of the projection operation which provides the robustness mechanism requires alternate observer forms depending on whether the measurement model is minimal or redundant. Concepts of linear measurement redundancy were proposed which result in a simple rank test to determine which form of sliding mode observer is required. Observer forms, and stability and robustness properties were detailed for both nonredundant and redundant measurements. These observer forms can be implemented directly due to the linear measurement model. In the case of redundant measurements, the proposed approach minimizes a quadratic cost function during the sliding motion and hence the method provides an optimal fusion of the redundant measurements. Expected performance of the proposed observers was illustrated through Monte Carlo simulation results obtained in the estimation of position and motion of 3D feature points in Level V of the hierarchical motion analysis system.

Stability and robustness properties of the sliding mode observer proposed in Chapter 6 require the existence of possibly time-varying matrices $L(k)$ which not only satisfy certain criteria for stability (Theorem 6.3, Part 2), but also satisfy a certain *matching condition* (Theorem 6.3, Part 3). This matching condition is similar in form to that required in continuous-time robust observer design methods [77, 100, 101, 102], particularly for linear systems with unknown inputs [165, 166]. Conditions for existence of matrices $L(k)$ satisfying both the stability and robustness conditions developed herein remain open at the present time. In many linear continuous-time sliding mode methods (see reviews by Fernandez [118] and Misawa [102]), combined stability and robustness is shown to be possible provided a certain transfer function can be made strictly positive real, followed by application of the Popov-Kalman-Yakubovich (Positive Real) Lemma. This approach, again for continuous-time systems, has also been applied to observer design for nonlinear systems, provided the nonlinearities also satisfy similar matching conditions [77, 100, 101, 102]. One possible avenue to obtain existence conditions for the matrices $L(k)$, at least for linear systems, is through careful examination of discrete-time strictly positive real transfer functions and the corresponding Discrete-time Positive Real Lemma (see, for example, [167, Section 2.6]). Initial work towards this goal, however, has led to some difficulties due to the distinct differences in form between the Positive Real Lemma for continuous-time systems

and that for discrete-time systems.

In Chapter 7, observer design for nonlinear measurements proceeded in a manner similar to linearization, but using instead the Fundamental Theorem of Integral Calculus. Again, design of sliding mode observers for the nonlinear measurement case require alternate strategies depending on whether the measurement model is minimal or redundant. Concepts of nonlinear measurement redundancy were proposed which are a direct extension of the linear case and draw from some basic results from differential geometry. For systems with nonlinear redundant measurement models, the objective of observer design is to produce a least squared error estimate during the sliding motion.

The *ideal* observer forms for nonlinear nonredundant or redundant measurements are very similar to those proposed for linear measurement models. In the nonlinear case, however, ideal observer forms are represented by systems of nonlinear integral equations which implicitly define the observer output. A simple implementation method, using an approximation similar to that employed in the Extended Kalman Filter, can yield good performance in some cases. More reliable performance is obtained through introduction of local iterations at each observation event, similar to those employed in the Iterated Extended Kalman Filter. These local iterations yield Newton-type iterations for nonredundant measurements and Gauss-Newton-type iterations for redundant measurements. The ideal observer forms, together with stability and robustness properties were given for both nonredundant and redundant observers, followed by proposed implementation methods with local iterations. Stability analysis, which yield sufficient conditions for the numerical implementation to approximate the ideal observer with arbitrary accuracy, was given for the iterated nonredundant observer. The proposed local iterations provide a practical implementation methodology for sliding mode observers which can be applied to a broad range of state estimation problems.

Further analysis is required to obtain stability for the iterated redundant observer and to obtain more precise quantitative results for observer performance in the presence of measurement errors. Another area of future work, towards which some effort has already been devoted, is investigation of alternatives and extensions to Newton and Gauss-Newton methods, such as variable step length algorithms [159, Section 8.3] and continuation methods [159, Section 10.4]. Initial analysis of these approaches for the iterated observers have in-

licated that proposed observer forms are well suited to inclusion of the additional terms required. These methods also appear to offer significant advantages in increasing the region of attraction and improving convergence rates of the local iterations.

Applications of the proposed iterated sliding mode observers were developed for estimating motion and structure of objects visible in multiple-camera image sequences based on feature position measurements. Expected performance of the proposed observers was illustrated through Monte Carlo simulation results. Several primary components of the overall feature-based image analysis problem, for example correspondence and occlusion, remain to be solved. Further development of the proposed observers, not only in more complete implementations of the vision-based motion analysis problem, but also to other systems such as robotics, unmanned vehicle control, and remote sensor deployment represent primary research areas of particular interest to the candidate.

References

- [1] B. Friedland, *Control system design: an introduction to state-space methods*, McGraw-Hill: New York, 1986.
- [2] D.G. Luenberger, "Observing the state of a linear system," *IEEE Transactions on Military Electronics*, Vol. MIL-8, April, 1964, pp 74-80.
- [3] D.G. Luenberger, "Observers for multivariable systems," *IEEE Transactions on Automatic Control*, Vol. AC-11, No. 2, April, 1966, pp 190-197.
- [4] D.G. Luenberger, "An introduction to observers," *IEEE Transactions on Automatic Control*, Vol. AC-16, No. 6, Dec. 1971, pp 596-602.
- [5] N. Wiener, *Extrapolation, interpolation and smoothing of stationary time series, with engineering applications*, Technology Press and Wiley, New York, 1949.
- [6] R.E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME Journal of Basic Engineering*, Vol. 82D, No. 1, March 1960, pp 35-45.
- [7] R.E. Kalman, R.S. Bucy, "New results in linear filtering and prediction theory," *Transactions of the ASME Journal of Basic Engineering*, Vol. 83D, No. 1, March 1961, pp 95-108.
- [8] H.W. Sorenson, ed., *Kalman filtering: Theory and practice*, IEEE Press, New York, N.Y. 1985.
- [9] A.H. Jazwinski, *Stochastic Processes and Filtering Theory*, New York: Academic, 1970.

- [10] A. Gelb, *Applied Optimal Estimation*, Cambridge MA: M.I.T. Press, 1974.
- [11] B.D.O. Anderson, J.B. Moore, *Optimal Filtering*, Englewood Cliffs, NJ: Prentice-Hall, 1979.
- [12] P.S. Maybeck, *Stochastic Models, Estimation, and Control*, New York: Academic, Vol. 1, 1979, and Vol. 2, 1982.
- [13] A.V. Balakrishnan, *Kalman Filtering Theory*, Optimization Software Inc., Publications Division, New York, 1984.
- [14] T. Kailath, *Linear Systems*, Prentice-Hall, Englewood Cliffs, N.J., 1980.
- [15] C.M. Dorling, A.S.I. Zinober, "A comparative study of the sensitivity of observers," in *Proceedings of the IASTED Symposium on Applied Control and Identification*, Copenhagen, 1983, pp 6.32-6.38.
- [16] V.I. Utkin, "Variable structure systems with sliding modes," *IEEE Transactions on Automatic Control*, Vol. AC-22, No. 2, April 1977, pp 212-222.
- [17] A.F. Filippov, "Differential equations with discontinuous right-hand side," *American Math Society Translations*, Series 2, Vol. 42, 1964, pp 199-231.
- [18] H. Sira-Ramirez, "Non-linear discrete variable structure systems in quasi-sliding mode," *International Journal of Control*, Vol. 54, No. 5, 1991, pp 1171-1187.
- [19] S.Z. Sarpturk, Y. Istefanopulos, O. Kaynak, "On the stability of discrete-time sliding mode control systems," *IEEE Transactions on Automatic Control*, Vol. AC-32, No. 10, Oct. 1987, pp 930-932.
- [20] V.C. Aitken, "Motion and structure estimation in noisy monocular image sequences: specialization to planar motion," Master's Thesis, Carleton University, Ottawa, Ontario, 1991.
- [21] V.C. Aitken, H.M. Schwartz, "Motion and structure from planar motion in monocular image sequences," in *Proceedings of the Third Conference on Military Robotic Applications, MRV91*, Medicine Hat, Alberta, Sept. 9-12, 1991, pp 177-184.

- [22] V.C. Aitken, "Motion and structure estimation of manoeuvring objects in multiple-camera image sequences," Suffield Report No. SR-577, Defence Research Establishment Suffield, Medicine Hat, Alberta, Canada, 1992.
- [23] V.C. Aitken and H.M. Schwartz, "A comparison of rotational representations in structure and motion estimation for manoeuvring objects," *IEEE Transactions on Image Processing*, to appear, April 1995.
- [24] V.C. Aitken and H.M. Schwartz, "Towards robust discrete-time sliding mode observers," in *Proceedings, 1995 American Control Conference*, Seattle, Washington, June 21-23, to appear, 1995.
- [25] V.C. Aitken and H.M. Schwartz, "On the exponential stability of discrete-time systems with applications in observer design," *IEEE Transactions on Automatic Control*, Vol. 39, No. 9, Sept. 1994, pp 1959-1962.
- [26] V.C. Aitken and H.M. Schwartz, "On the exponential stability of discrete-time systems with applications in observer design," in *Proceedings, 1994 American Control Conference*, Baltimore, Maryland, June 29-July 1, 1994 pp. 2196-2201.
- [27] R.P. Futrelle, G.C. Speckert, "Extraction of motion data by interactive image processing," in *Proceedings, IEEE Conference on Pattern Recognition and Image Processing*, Chicago, Illinois, May 31-June 2, 1978, pp 405-408.
- [28] D.B. Gennery, "Tracking known three-dimensional objects," in *Proceedings, National Conference on Artificial Intelligence*, AAAI-82, Pittsburgh, Pennsylvania, Aug. 18-20, 1982, pp 13-17.
- [29] J. Hallam, "Resolving observer motion by object tracking," in *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, Karlsruhe, West Germany, Aug. 8-12, 1983, pp 792-797.
- [30] G.R. Legters, T.Y. Young, "A mathematical model for computer image tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-4, No. 6, Nov. 1982, pp 583-594.

- [31] J. Stuller, G. Krishnamurthy, "Kalman filter formulation of low-level television image motion estimation," *Computer Vision, Graphics, and Image Processing*, Vol. 21, 1983, pp 169-204.
- [32] J.J. Wu, R.E. Rink, T.M. Caelli, V.G. Gourishankar, "Recovery of the 3-D location and motion of a rigid object through camera image (an extended Kalman filter approach)," *International Journal of Computer Vision*, Vol. 3, 1988, pp 373-394.
- [33] L. Matties, T. Kanade, R. Szeliski, "Kalman filter-based algorithms for estimating depth from image sequences," *International Journal of Computer Vision*, Vol. 3, 1989, pp 209-236.
- [34] K. Chang, H. Lee, C. Chung, "Adaptive tracking system for a manoeuvring target using images with correlated noises," *International Journal of Systems Science*, Vol. 20, No. 5, 1989, pp 839-857.
- [35] S.-L. IU, K. Wohn, "Estimation of general rigid body motion from a sequence of images," in *Proceedings of the IEEE 10th International Conference on Pattern Recognition*, Vol. 1, Atlantic City, NJ, USA, June 16-21, 1990, pp 217-219.
- [36] S.-L. IU, K. Wohn, "Recovery of 3D motion of a single particle," *Pattern Recognition*, Vol. 24, No. 3, 1991, pp 241-252.
- [37] N. Ayache, O. Faugeras, "Maintaining representations of the environment of a mobile robot," *IEEE Transactions on Robotics and Automation*, Vol. 5, No. 6, Dec. 1989, pp 804-819.
- [38] N. Ayache, F. Lustman, "Trinocular stereo vision for robotics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 1, Jan. 1991, pp 73-85.
- [39] G. Randall, S. Foret, N. Ayache, "Final steps towards real time trinocular stereo vision," in *Proceedings of the First European Conference on Computer Vision, ECCV-90*, Antibes, France, April 23-27, 1990, pp 601-603.

- [40] E.D. Dickmanns, V. Graefe, "Dynamic monocular machine vision," *International Journal of Machine Vision and Applications*, Springer-Verlag, New York, Vol. 1, No. 4, 1988, pp 223-240.
- [41] E.D. Dickmanns, V. Graefe, "Applications of dynamic monocular machine vision," *International Journal of Machine Vision and Applications*, Springer-Verlag, New York, Vol. 1, No. 4, 1988, pp 241-261.
- [42] V. Graefe, "A pre-processor for the real-time interpretation of dynamic scenes," in T.S. Huang (Ed.), *Image Sequence Processing and Dynamic Scene Analysis*, Springer-Verlag, Berlin, 1983, pp 519-531.
- [43] H.G. Meissner, E.D. Dickmanns, "Control of an unstable plant by computer vision," in T.S. Huang (Ed.), *Image Sequence Processing and Dynamic Scene Analysis*, Springer-Verlag, Berlin, 1983, pp 532-548.
- [44] R. Schell, E.D. Dickmanns, "Autonomous automatic landing through computer vision," in *AGARD Conference Proceedings No. 455*, Guidance and Control Panel 48th Symposium, Instituto da Defesa Nacional in Lisbon, Portugal, May 9-12, 1989, pp 24:1-24:9.
- [45] E.D. Dickmanns, B. Mysliwitz, T. Christians, "An integrated spatio-temporal approach to automatic guidance of autonomous vehicles," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 20, No. 6, Nov./Dec. 1990, pp 1273-1284.
- [46] E.D. Dickmanns, T. Christians, "Relative 3D-state estimation for autonomous visual guidance of road vehicles," *Robotics and Autonomous Systems*, Vol. 7, 1991, pp 113-123.
- [47] E.D. Dickmanns, B.D. Mysliwetz, "Recursive 3-D road and relative ego-state recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 2, Feb. 1992, pp 199-213.

- [48] T.J. Broida, "Estimating the kinematics and structure of a moving object from a sequence of images," Ph.D. dissertation, University of Southern California, Los Angeles, 1987.
- [49] T. Broida, R. Chellappa, "Estimation of object motion parameters from noisy images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-8, No. 1, Jan. 1986, pp 90-99.
- [50] T. Broida, R. Chellappa, "Kinematics and structure of a rigid object from a sequence of noisy images," in *Proceedings, Workshop on Motion: Representation and Analysis*, Charleston, S.C., May 7-9, 1986, pp 95-100.
- [51] T. Broida, R. Chellappa, "Kinematics of a rigid object from a sequence of noisy images: a batch approach," in *Proceedings, Computer Vision and Pattern Recognition*, Miami Beach, Fla., June 22-26, 1986, pp 176-182.
- [52] T.J. Broida, R. Chellappa, "Estimating the kinematics and structure of a rigid object from a sequence of monocular images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 6, 1991, pp 497-513.
- [53] T. Broida, R. Chellappa, "Performance bounds for estimating three dimensional motion parameters from a sequence of noisy images," *Journal of the Optical Society of America*, Vol. 6, No. 6, June 1989, pp 879-889.
- [54] T. Broida, S. Chandrashekhar, R. Chellappa, "Recursive 3-D motion estimation from a monocular image sequence," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 26, No. 4, July 1990, pp 639-656.
- [55] G.J. Young, R. Chellappa, "3-D motion estimation using a sequence of noisy stereo images," in *Proceedings, Computer Vision and Pattern Recognition*, Ann Arbor, Michigan, June 5-9, 1988, pp 710-716.
- [56] G.J. Young, R. Chellappa, "3-D motion estimation using a sequence of noisy stereo images: models, estimation, and uniqueness results," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-12, No. 8, Aug. 1990, pp 735-759.

- [57] G.S. Young, R. Chellappa, T.H. Wu, "Monocular motion estimation using a long sequence of noisy images," in *Proceedings, IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toronto, Ontario, May 14-17, 1991, pp 2437-2440.
- [58] B.P.K. Horn, "Closed form solutions of absolute orientation using unit quaternions," *Journal of the Optical Society of America, A*, Vol. 4, No. 4, Apr. 1987, pp 629-642.
- [59] I.Y. Bar-Itzhack, Y. Oshman, "Attitude determination from vector observations: quaternion estimation," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-21, No. 1, Jan. 1985, pp 128-135.
- [60] R.A. Singer, "Estimating optimal tracking filter performance for manned manoeuvring targets," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-6, No. 4, July, 1970, pp 473-483.
- [61] J. Stuelpnagel, "On the parameterization of the three-dimensional rotation group," *SIAM Review*, Vol. 6, No. 4, Oct. 1964, pp 422-430.
- [62] P.C. Hughes, *Spacecraft Attitude Dynamics*, New York: John Wiley and Sons, 1986.
- [63] V.N. Branets and I.P. Shmyglevski, *Primeneniye Kvaternionov V Zadachakh Orientatsii Tverdogo Tela*, (Application of Quaternions To Rigid Body Rotation Problems), Nauka Press, Moscow, 1973, NASA technical translation NASA TT F-15,414, Translated for National Aeronautics and Space Administration under contract No. NASw 2483, by SCITRAN, P.O. Box 5456, Santa Barbera, California, 93108.
- [64] M. Tahk, J.L. Spayer, "Target tracking problems subject to kinematic constraints," *IEEE Transactions on Automatic Control*, Vol. 35, No. 3, March, 1990, pp 324-326.
- [65] I.Y. Bar-Itzhack, "Optimum normalization of a computed quaternion of rotation," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-7, Mar. 1971, pp 401-402.

- [66] M.H. Ang, V.D. Tourassis, "Singularities of Euler and roll-pitch-yaw representations," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-23, No. 3, May 1987, pp 317-324.
- [67] C.R. Rao, *Linear Statistical Inference and Its Applications, Second Edition*, New York: John Wiley and Sons, 1973.
- [68] K.V. Bury, *Statistical Models in Applied Science*, New York: John Wiley and Sons, 1975.
- [69] J.H. Taylor, "The Cramer-Rao estimation error lower bound computation for deterministic nonlinear systems," *IEEE Transactions on Automatic Control*, Vol. AC-24, No.2, April 1979, pp 343-344.
- [70] C. Chang, J. Tabaczynski, "Application of state estimation to target tracking," *IEEE Transactions on Automatic Control*, Vol. AC-29, No. 2, Feb. 1984, pp 98-109.
- [71] Y. Bar-Shalom, T.E. Fortmann, *Tracking and data association*, Academic Press, London, 1988, Appendix B and pp. 145-148.
- [72] S.R. Kou, D.L. Elliot, T.J. Tarn, "Observability of nonlinear systems," *Information and Control*, Vol. 22, 1973, pp 89-99.
- [73] R. Hermann, A.J. Krener, "Nonlinear controllability and observability," *IEEE Transactions on Automatic Control*, Vol. 22, No. 5, Oct. 1977, pp 728-740.
- [74] H. Nijmeijer, "Observability of autonomous discrete-time nonlinear systems: a geometric approach," *International Journal of Control*, Vol. 36, No. 5, 1982, pp 867-874.
- [75] S.T. Glad, "Observability and nonlinear dead beat observers," in *Proceedings of the 28th IEEE Conference on Decision and Control*, Tampa, FL, 1989, pp 90-95.
- [76] J.W. Grizzle, P.E. Moral, "On observers for smooth nonlinear digital systems," in A. Bensoussan and J. Lions (Eds.), *Proceedings of the 9th International Conference on Analysis and Optimization of Systems*, 1990, pp 401-410.

- [77] B.L. Walcott, M.J. Corless, S.H. Zak, "Comparative study of nonlinear state observation techniques," *International Journal of Control*, Vol. 45, No. 6, 1987, pp 2109-2132.
- [78] E.A. Misawa, J.K. Hendrick, "Nonlinear observers—A state-of-the-art survey," *Transactions of the ASME Journal of Dynamic Systems, Measurement, and Control*, Vol. 111, Sept. 1989, pp 344-352.
- [79] F.E. Thau, "Observing the state of nonlinear dynamic systems," *International Journal of Control*, Vol. 17, No. 3, 1973, pp 471-479.
- [80] S.R. Kou, D.L. Elliot, T.J. Tarn, "Exponential observers for nonlinear dynamic systems," *Information and Control*, Vol. 29, 1975, pp 204-216.
- [81] S.P. Banks, "A note on nonlinear observers," *International Journal of Control*, Vol. 34, No. 1, 1981, pp 185-190.
- [82] T.J. Tarn, Y. Rasis, "Observers for nonlinear stochastic systems," *IEEE Transactions on Automatic Control*, Vol. 21, No. 4, Aug. 1976.
- [83] X.-H. Xia, W.-B. Gao, "On exponential observers for nonlinear systems," *Systems and Control Letters*, Vol. 11, 1988, pp 319-325.
- [84] E. Yaz, A. Azemi, "Lyapunov-based nonlinear observer design for stochastic systems," in *Proceedings of the 29th IEEE Conference on Decision and Control*, Honolulu, HI, USA, Dec. 5-7, 1990 pp. 218-219.
- [85] D. Bestle, M. Zeitz, "Canonical form observer design for nonlinear time-variable systems," *International Journal of Control*, Vol. 38, No. 2, 1983, pp 419-431.
- [86] A.J. Krener, A. Isidori, "Linearization by output injection and non-linear observers," *Systems and Control Letters*, Vol. 3, 1983, pp 47-52.
- [87] A.J. Krener, W. Respondek, "Nonlinear observers with linearizable error dynamics," *SIAM Journal of Control and Optimization*, Vol. 23, No. 2, March 1985, pp 197-216.

- [88] X.-H. Xia, W.-B. Gao, "Nonlinear observer design by observer error linearization," *SIAM Journal of Control and Optimization*, Vol. 27, No. 1, Jan. 1989, pp 199-216.
- [89] S. Nicosia, P. Tomei, A. Tornambe, "A nonlinear observer for elastic robots," *IEEE Journal of Robotics and Automation*, Vol. 4, No. 1, Feb. 1988, pp 45-52.
- [90] J.P. Gauthier, H. Hammouri, S. Othman, "A simple observer for nonlinear systems applications to bioreactors," *IEEE Transactions on Automatic Control*, Vol. 37, No. 6, June 1992, pp 875-880.
- [91] M. Zeitz, "The extended Luenberger observer for nonlinear systems," *Systems and Control Letters*, Vol. 9, 1987, pp 149-156.
- [92] S. Nicosia, A. Tornambe, P. Valigi, "Use of observers for the inversion of nonlinear maps," *Systems and Control Letters*, Vol. 16, 1991, pp 447-455.
- [93] S. Nicosia, A. Tornambe, P. Vagili, "A solution to the generalized problem of nonlinear map inversion," *Systems and Control Letters*, Vol. 17, 1991, pp 383-394.
- [94] B.M. Bell, F.W. Cathey, "The iterated Kalman filter update as a Gauss-Newton Method," *IEEE Transactions on Automatic Control*, Vol. 38, No. 2, Feb. 1993, pp 294-297.
- [95] J.-J. Slotine, J.K. Hendrick, E.A. Misawa, "On sliding observers for nonlinear systems," *Transactions of the ASME Journal of Dynamic Systems, Measurement, and Control*, Vol. 109, Sept. 1987, pp 245-252.
- [96] C. Canudas de Wit, J.-J.E. Slotine, "Sliding observers for robot manipulators," *Automatica*, Vol. 27, No. 5, 1991, pp 859-864.
- [97] C. Canudas de Wit, N. Fixot, "Robust control via estimated state feedback," *IEEE Transactions on Automatic Control*, Vol. 36, No. 12, Dec. 1991, pp 1497-1501.
- [98] C. Canudas de Wit, N. Fixot, K.J. Astrom, "Trajectory tracking of nonlinear estimated state feedback," *IEEE Transactions on Robotics and Automation*, Vol. 8, No. 1, Feb. 1992, pp 138-144.

- [99] C. Canudas De Wit, N. Fixot, "Adaptive control of robotic manipulators via velocity estimated feedback," *IEEE Transactions on Automatic Control*, Vol. 37, No. 8, Aug. 1992, pp 1234-1237.
- [100] B.L. Walcott, S.H. Zak, "Observation of dynamical systems in the presence of bounded nonlinearities/uncertainties," in *Proceedings of the 25th IEEE Conference on Decision and Control*, Athans, Greece, Dec. 1986, pp 961-966.
- [101] B.L. Walcott, S.H. Zak, "State observation of nonlinear uncertain dynamical systems," *IEEE Transactions on Automatic Control*, Vol. AC-32, No. 2, Feb. 1987, pp 166-170.
- [102] E.A. Misawa, "Nonlinear state estimation using sliding mode observers," Ph.D. thesis, Massachusetts Institute of Technology, February, 1988.
- [103] Z.R. Novakovic, B. Nemec, "A solution of the inverse kinematics problem using the sliding mode," *IEEE Transactions on Robotics and Automation*, Vol. 6, No. 2, April 1990, pp 247-252.
- [104] H. Hashimoto, V.I. Utkin, J. Xu, H. Suzuki, F. Harishima, "VSS observer for linear time varying systems," in *Proceedings, IEEE Conference on Industrial Electronics, IECON'90*, 1990, pp 34-39.
- [105] H. Hashimoto, J. Xu, F. Harishima, "VSS observer for motion control systems," in *Proceedings of the IEEE International Conference on Control and Applications, ICCON '89*, Jerusalem, April 3-6, 1989, pp 1-2.
- [106] S. Darkunov, U. Ozguner, D. Schoenwald, "Decentralized sliding mode observers for interconnected nonlinear systems," in *Proceedings of the IEEE International Workshop on Variable Structure and Lyapunov Control of Uncertain Dynamical Systems*, Sheffield, UK, Sept. 7-9, 1992.
- [107] S.V. Drakunov, "An adaptive quasioptimal filter with discontinuous parameters," *Automation and Remote Control*, Vol. 44, No. 9, 1983, pp 1167-1175.

- [108] B.Z. Golembo, S.V. Emel'yanov, V.I. Utkin, A.M. Shadladze, "Applications of piecewise-continuous dynamic systems to filtering problems", *Automation and Remote Control*, Vol. 37, No. 3, 1976, pp 369-377.
- [109] W. Lee, K. Nam, "Observer design for autonomous discrete-time nonlinear systems," *Systems and Control Letters*, Vol. 17, 1991, pp. 49-58.
- [110] S.-T. Chung, J.W. Grizzle, "Observer error linearization for sampled data systems," in *Proceedings of the 28th IEEE Conference on Decision and Control*, Tampa, FL, 1989, pp 90-95.
- [111] E. Tse, M. Athans, "Optimal minimal-order observer-estimator for discrete linear time-varying systems," *IEEE Transactions on Automatic Control*, Vol. AC-15, No. 4, Aug. 1970, pp 416-426.
- [112] E. Tse, "Observer-estimators for discrete-time systems," *IEEE Transactions on Automatic Control*, Vol. AC-18, No. 1, Feb. 1973, pp 10-16.
- [113] C.T. Leondes, L.M. Novak, "Optimal minimal-order observers for discrete-time systems—A unified theory," *Automatica*, Vol. 8, July 1972, pp 379-387.
- [114] B. Priel, E. Soroka, U. Shaked, "The design of optimal reduced-order stochastic observers for discrete-time linear systems," *IEEE Transactions on Automatic Control*, Vol. 36, No. 12, Dec. 1991, pp 1502-1509.
- [115] W. Baumann, W. Rugh, "Feedback control of nonlinear systems by extended linearization", *IEEE Transactions on Automatic Control*, Vol. AC-31, No. 1, Jan. 1986, pp 40-46.
- [116] J.Y. Hung, W. Gao, J.C. Hung, "Variable structure control: a survey," *IEEE Transactions on Industrial Electronics*, Vol. 40, No. 1, 1993, pp 2-22.
- [117] J.-J.E. Slotine, S.S. Sastry, "Tracking control of non-linear systems using sliding surfaces, with applications to robotic manipulators", *International Journal of Control*, Vol. 38, No. 2, 1983, pp 465-492.

- [118] B. Fernandez R., "Generalized sliding mode control for MIMO nonlinear systems," Ph.D. thesis, Massachusetts Institute of Technology, 1988.
- [119] B.R. Fernandez, J.K. Hendrick, "Control of multivariable nonlinear systems by sliding mode method," *International Journal of Control*, Vol. 46, No. 3, 1987, pp 1019-1040.
- [120] H. Sira-Ramirez, "Differential geometric methods in variable structure control," *International Journal of Control*, Vol. 48, No. 4, Oct. 1988, pp 1359-1390.
- [121] H. Sira-Ramirez, "Nonlinear variable structure systems in sliding mode: the general case," *IEEE Transactions on Automatic Control*, Vol. 34, No. 11, Nov. 1989, pp 1186-1188.
- [122] H. Sira-Ramirez, "Sliding regimes in general nonlinear systems: a relative degree approach," *International Journal of Control*, Vol. 50, No. 4, Oct. 1989, pp 1487-1506.
- [123] W. Gao, J.C. Hung, "Variable structure control of nonlinear systems: a new approach," *IEEE Transactions on Industrial Electronics*, Vol. 40, No. 1, Feb. 1993, pp 45-55.
- [124] H. Sira-Ramirez, "A dynamical variable structure control strategy in output tracking problems," *IEEE Transactions on Automatic Control*, Vol. 38, No. 4, April, 1993, pp 615-620.
- [125] H. Elmali, N. Olgac, "Robust output tracking control of nonlinear MIMO systems via sliding mode technique," *Automatica*, Vol. 28, No. 1, 1992, pp. 145-151.
- [126] W.S. Newman, K. Souccar, "Robust, near time-optimal control of nonlinear second-order systems: theory and experiments," *Transactions of the ASME. Journal of Dynamic Systems, Measurement and Control*, Vol. 113, No. 3, Sept. 1991, pp. 363-370.
- [127] T.-L. Liao, L.-C. Fu, C.-F. Hsu, "Output tracking control of nonlinear systems with mismatched uncertainties," *Systems and Control Letters*, Vol. 18, 1992, pp 39-47.

- [128] J.-J.E. Slotine, "Sliding controller design for nonlinear systems," *International Journal of Control*, Vol. 40, No. 2, 1984, pp 421-434.
- [129] J.A. Burton, A.S.I. Zinober, "Continuous approximation of variable structure control," *International Journal of Systems Science*, Vol. 17, No. 6, 1986, pp 875-885.
- [130] B. Fernandez, "Robust feedback linearization through sliding mode control," in *Proceedings of the 29th IEEE Conference on Decision and Control*, Honolulu, HI, USA, Dec. 5-7, 1990, pp. 3398-3399.
- [131] H. Sira-Ramirez, S. Ahmad, M. Zribi, "Dynamical feedback control of robotic manipulators with joint flexibility," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 22, No. 4, July/Aug. 1992, pp 736-747.
- [132] H.B. Shin, K.Y. Cho, M.J. Youn, "Variable structure observer of mismatched uncertain dynamical systems," *Electronics Letters*, Vol. 27, No. 20, Sept. 26th, 1991, pp 1798-1800.
- [133] C. Milosavljevic, "General conditions for the existence of a quasi-sliding mode on the switching hyperplane in discrete variable structure systems," *Automation and Remote Control*, Vol. 46, 1985, pp 307-314.
- [134] H.P. Opitz, "Robustness properties of discrete-variable structure controllers," *International Journal of Control*, Vol. 43, No. 3, 1986, pp 1003-1014.
- [135] K. Furuta, "Sliding mode control of a discrete system," *Systems and Control Letters*, Vol. 14, 1990, pp 145-152.
- [136] G.M. Aly, W.G. Ali, "Digital design of variable structure control systems," *International Journal of Systems Science*, Vol. 21, No. 8, 1990, pp 1709-1720.
- [137] C.Y. Chan, "Servo-systems with discrete variable structure control," *Systems and Control Letters*, Vol. 17, 1991, pp 321-325.
- [138] S.V. Drakunov, V.I. Utkin, "Sliding mode control in dynamic systems," *International Journal of Control*, Vol. 55, No. 4, 1992, pp 1029-1037.

- [139] H. Hashimoto, Y. Ishikawa, "Robust digital sliding mode control applied to motion control systems," in *Proceedings of the Japan/USA Symposium on Flexible Automation*, San Francisco, July, 13-15, 1992, pp 25-30.
- [140] O. Kaynak, A. Denker, "Discrete-time sliding mode control in the presence of system uncertainty," *International Journal of Control*, Vol. 57, No. 5, May 1993, pp 1177-1190.
- [141] S.V. Baida, "Unit sliding mode control in continuous and discrete-time systems," *International Journal of Control*, Vol. 57, No. 5, May 1993, pp 1125-1132.
- [142] S.V. Drakunov, V.I. Utkin, "On discrete-time sliding modes," in *Proceedings of the IFAC Symposium on Nonlinear Control Systems (NOLCOS)*, Capri, Italy, 1989, pp 484-489.
- [143] G.F. Franklin, J.D. Powell, M.L. Workman, *Digital Control of Dynamic Systems*, Addison-Wesley, New York, 1990.
- [144] J.E. Marsden, *Elementary Classical Analysis*, W.H. Freeman and Co. New York, 1974.
- [145] H.H. Rosenbrock, "A method of investigating stability," in *Proc. I.F.A.C. Congress*, Basle, 1963, pp. 590-594.
- [146] S. Weissenberger, "Piecewise-quadratic and piecewise-linear Lyapunov functions for discontinuous systems," *Int. J. Contr.*, Vol. 10, No. 2, 1969, pp. 171-180.
- [147] G. Bitsoris, "Positively invariant polyhedral sets of discrete-time linear systems," *Int. J. Contr.*, Vol. 47, No. 6, 1988, pp. 1713-1726.
- [148] A.P. Molchanov, "Lyapunov functions for nonlinear discrete-time control systems," *Automation and Remote Control*, Vol. 48, 1987, pp. 728-736.
- [149] H. Kiendl, J. Adamy, and P. Stelzner, "Vector norms as Lyapunov functions for linear systems," *IEEE Trans. Automat. Contr.*, Vol. 37, No. 6, June 1992, pp. 839-842.

- [150] R.A. Horn, and C.R. Johnson, *Matrix Analysis*, Cambridge: Cambridge University Press, 1991.
- [151] J.L. Willems, *Stability Theory of Dynamical Systems*, New York: Wiley, 1970.
- [152] J.P. La Salle, *Stability of Dynamical Systems*, Philadelphia: SIAM, 1976.
- [153] J. Ačkermann, *Sampled Data Control Systems*, (Translated from German), Berlin: Springer Verlag, 1985.
- [154] V. Lakshmikantham, and D. Trigiante, *Theory of Difference Equations, Numerical Methods and Applications*, New York: Academic, 1988.
- [155] E.D. Sontag, *Mathematical control theory: deterministic finite dimensional systems*, Springer, New York, 1990.
- [156] L. Weiss, and L. Lam, "Stability of non-linear discrete-time systems," *Int. J. Contr.*, Vol. 17, No. 3, pp. 465-470.
- [157] R.E. Kalman, J.E. Bertram, "Control system analysis and design via the second method of Liapunov, Part II, discrete systems," *Trans. ASME J. Basic Eng.*, Vol. 82, 1960, pp. 394-400.
- [158] W. Hahn, *Stability of Motion*, Berlin: Springer Verlag, 1967.
- [159] J.M Ortega, W.C. Rheinboldt, *Iterative solution of nonlinear equations in several variables*, Academic Press Inc., 1970.
- [160] G.H. Golub, C.F. Van Loan, *Matrix computations*, John Hopkins, Baltimore MA, 1983.
- [161] S. Perlis, *Theory of matrices*, Dover Publications Inc., New York, 1991.
- [162] E.K. Blum, *Numerical analysis and computation: theory and practice*, Addison-Wesley, Reading, Mass., 1972.
- [163] W.M. Boothby, *An introduction to differentiable manifolds and Riemannian geometry*, Academic Press, New York, 1975.

- [164] G. Strang, *Linear algebra and its applications*, 2nd Edition, Academic Press, New York, 1980.
- [165] S.P. Bhattacharyya, "Observer design for linear systems with unknown inputs," *IEEE Transactions on Automatic Control*, Vol. AC-23, No. 3, June, 1978, pp 483-484.
- [166] X. Ding, P.M. Frank, L. Guo, "Nonlinear observer design via an extended observer canonical form," *Systems and Control Letters*, Vol. 15, 1990, pp 313-332.
- [167] B.D.O Anderson *et al.*, *Stability of adaptive systems*, MIT Press, Cambridge MA, 1986.
- [168] Y.L. Gu, "An exploration of orientation representation by Lie algebra for robotic applications," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 20, No. 1, Jan. 1990, pp 243-248.
- [169] A. Edwards, "The state of strapdown inertial guidance and navigation," *Journal of The Institute of Navigation*, Vol. 18, No. 4, Winter, 1971-72, pp 386- 391.
- [170] B. Etkin, *Dynamics of Atmospheric Flight*, New York: John Wiley and Sons, 1972.

Appendix A

Parameterizations of Relative Orientation

This Appendix describes three parameterizations for change-of-basis transformations denoted $\mathbf{I}_E^O$ such that a vector $\mathbf{r}_O$ in an object-centred reference frame F_O is transformed to $[\mathbf{r}_O]_E$, its representation in an earth-fixed frame F_E , through the expression

$$[\mathbf{r}_O]_E = \mathbf{I}_E^O \mathbf{r}_O. \quad (\text{A.1})$$

The object-centred frame, F_O , is defined by a point O_O , the origin of F_O , and a right-handed, ordered, orthonormal basis set $\{\mathbf{i}_O, \mathbf{j}_O, \mathbf{k}_O\}$. Equivalent notation also applies to the earth-fixed frame F_E . In order to focus on relative orientation only, it is assumed throughout this Appendix that O_O and O_E are coincident. In this case,

$$[\mathbf{r}_O]_E \equiv \mathbf{r}_E \quad (\text{A.2})$$

since the translation vector from F_E to F_O is zero. The orthogonal transformation $\mathbf{I}_E^O$ belongs to the three-dimensional rotation group, denoted here by $\mathcal{R}_3$, for which there are about eight commonly used representations [61]. This Appendix provides a tutorial on three such representations: the Euler angle-axis, quaternion (sometimes called Euler parameters, or Rodriguez-Hamilton parameters), and roll-pitch-yaw (sometimes called Euler angle) parameterizations. The material presented in this Appendix draws primarily from references [61, 62, 63, 168, 169, 170].

The problem of parameterizing the group $\mathcal{R}_3$ has been of interest since 1776, when Euler first showed that this group is itself a three-dimensional manifold. Euler observed that the general displacement of a rigid body with one fixed point is a rotation about an

axis through that point (Euler's theorem). In the present context, Euler's theorem states that F_O is related to F_E by a rigid rotation of $\mathbb{R}^3$ about an axis through their common origin. Since $\mathcal{R}_3$ is known to be three-dimensional, the rotational degrees-of-freedom in $\mathbb{R}^3$ are at most three. However, it is topologically impossible to have a global and nonsingular three-dimensional parameterization of $\mathcal{R}_3$, [61]. In fact, Hopf showed in 1940 that the minimum number of parameters required to represent $\mathcal{R}_3$ in a one-to-one global manner is five, (see [61]). The quaternion representation, a four-dimensional parameterization of $\mathcal{R}_3$, results in a two-to-one mapping of $\mathbb{R}^4$ onto $\mathcal{R}_3$, and is often considered sufficient for most practical purposes.

It is important to appreciate a rather subtle distinction between two interpretations of operations performed by elements of $\mathcal{R}_3$. One interpretation treats an element $\mathbf{R} \in \mathcal{R}_3$ as a *rotation operator* which relates two vectors, say $\mathbf{x}$ and $\mathbf{y}$ with $\|\mathbf{x}\| = \|\mathbf{y}\|$, in the same bases according to $\mathbf{y} = \mathbf{R}\mathbf{x}$. In this case, $\mathbf{R}$ rotates the vector $\mathbf{x}$ into a new vector $\mathbf{y}$. The second interpretation treats an element $\mathbf{I}_\alpha^3 \in \mathcal{R}_3$ as an orthogonal (or identity, hence the use of 'I') *change-of-basis transformation* which operates on a vector $\mathbf{r}_\beta$ expressed with respect to F_β to obtain a representation for the same vector, $\mathbf{r}_\alpha$, expressed with respect to F_α according to $\mathbf{r}_\alpha = \mathbf{I}_\alpha^3 \mathbf{r}_\beta$. The main point is that vectors $\mathbf{r}_\beta$ and $\mathbf{r}_\alpha$ represent the same physical entity, while the vectors $\mathbf{x}$ and $\mathbf{y}$ in general do not. It is often the case that "rotation" is used interchangeably to imply either interpretation. This appendix deals exclusively with the "change-of-bases transformation" interpretation.

This review first introduces fundamental properties of rotation matrices in Section A.1, treats the Euler angle-axis parameterization in Section A.2, quaternions in Section A.3, and the roll-pitch-yaw representation in Section A.4.

A.1 Fundamental Properties of $\mathbf{I}_E^O$

The following list states some fundamental and well-known properties of the matrix $\mathbf{I}_E^O$ representing the change-of-bases transformation from F_O to F_E .

Property 1 The matrix $\mathbf{I}_E^O$ is orthogonal with determinant +1.

Property 2 The matrix $\mathbf{I}_E^O$ has eigenvalues $\{+1, e^{\pm j\gamma}\}$. Euler's axis of rotation of $\mathbf{I}_E^O$ is parallel to the (normalized) eigenvector, $\bar{\xi}$, corresponding to the eigenvalue $+1$, while the angle of rotation, γ , about $\bar{\xi}$ is measured with positive sense from F_E to F_O according to the right-hand-rule about $\bar{\xi}$.

Property 3 The matrix $\mathbf{I}_E^O$ can be written as

$$\begin{aligned}\mathbf{I}_E^O &= \begin{bmatrix} [\mathbf{i}_O]_E & [\mathbf{j}_O]_E & [\mathbf{k}_O]_E \end{bmatrix} \\ &= \begin{bmatrix} [\mathbf{i}_E]_O & [\mathbf{j}_E]_O & [\mathbf{k}_E]_O \end{bmatrix}^T, \end{aligned} \quad (\text{A.3})$$

where superscript ' T ' denotes transposition.

Property 4 The angular velocity of F_O with respect to F_E expressed in F_E , denoted $[\omega_E^O]_E$, can be computed by noting that

$$\begin{aligned}\frac{d[\mathbf{i}_O]_E}{dt} &= [\omega_E^O]_E \times [\mathbf{i}_O]_E, \\ \frac{d[\mathbf{j}_O]_E}{dt} &= [\omega_E^O]_E \times [\mathbf{j}_O]_E, \\ \frac{d[\mathbf{k}_O]_E}{dt} &= [\omega_E^O]_E \times [\mathbf{k}_O]_E, \end{aligned} \quad (\text{A.4})$$

which gives

$$\dot{\mathbf{I}}_E^O = [\omega_E^O]_E^\# \mathbf{I}_E^O, \quad (\text{A.5})$$

where $[\cdot]^\#$ is the matrix cross-product operator defined in terms of components of a vector $\xi = [\bar{\xi}_1, \bar{\xi}_2, \bar{\xi}_3]^T$ as

$$\xi^\# := \begin{bmatrix} 0 & -\xi_3 & \xi_2 \\ \xi_3 & 0 & -\xi_1 \\ -\xi_2 & \xi_1 & 0 \end{bmatrix}. \quad (\text{A.6})$$

Rearranging (A.5) gives

$$[\omega_E^O]_E^\# = \dot{\mathbf{I}}_E^O (\mathbf{I}_E^O)^T = \dot{\mathbf{I}}_E^O \mathbf{I}_O^E. \quad (\text{A.7})$$

The matrix (tensor) $[\omega_E^O]_E^\#$ is basis-dependent and is transformed to F_O according to

$$\begin{aligned}[\omega_E^O]_O^\# &= \mathbf{I}_O^E [\omega_E^O]_E^\# \mathbf{I}_E^O \\ &= \mathbf{I}_O^E \dot{\mathbf{I}}_E^O \mathbf{I}_O^E \mathbf{I}_E^O \\ &= \mathbf{I}_O^E \dot{\mathbf{I}}_E^O \end{aligned} \quad (\text{A.8})$$

Property 1 follows from the fact that both F_O and F_E are right-handed orthonormal triads. Property 2 then follows immediately from Property 1. Property 3 is derived by considering identities of the form $[i_O]_E = I_E^O[i_O]_O$ and noting that $[i_O]_O = [1, 0, 0]^T$. Property 4 is fundamental in derivations of relationships between angular velocity and the selected parameter set and their time derivatives.

A.2 Euler Angle-Axis Parameterization

According to Euler's theorem, a rigid rotation of $\mathbb{R}^3$ is completely defined by a unit vector $\bar{\xi}$ which specifies the axis of the basis transformation operator I_E^O , and an angle γ through which the coordinate system is rotated about $\bar{\xi}$. This is a four-parameter representation with one quadratic constraint. Note that $\bar{\xi}$ is invariant under the basis transformation and hence the subscript specifying the frame to which $\bar{\xi}$ is referenced has been dropped.

Given an angle γ and axis $\bar{\xi}$ the corresponding matrix I_E^O can be determined by the following argument. Suppose at $t = 0$ reference frames F_O and F_E are aligned so that $I_E^O(t = 0) = I_3$, the 3×3 identity matrix. At time $t = 1$, the orientation of reference frame F_O relative to F_E is given by a rotation about $\bar{\xi}$ in F_E through an angle γ . For the purposes of deriving $I_E^O(t = 1)$, the motion during this time interval can be modelled with constant angular velocity $[\omega_E^O]_E = \gamma \bar{\xi}$. Integration of (A.5) from $t = 0$ to $t = 1$ with initial condition $I_E^O(t = 0) = I_3$ immediately gives the matrix exponential representation

$$I_E^O(t = 1) = \exp(\gamma \bar{\xi}^\sharp). \quad (\text{A.9})$$

or, with

$$\xi := \gamma \bar{\xi}, \quad (\text{A.10})$$

the basis transformation in terms of ξ is

$$I_E^O(\xi) = \exp(\xi^\sharp). \quad (\text{A.11})$$

Equation (A.11) is a three-dimensional parameterization in which ξ is often called the *orientation vector*. This representation has been used extensively by Faugeras [37], for example. The matrix cross product operator, $(\cdot)^\sharp$, yields a 3×3 skew symmetric matrix. As

suggested by (A.9), every 3×3 rotation matrix can be expressed as an exponential function of some 3×3 skew symmetric matrix, which has led to elegant studies in terms of Lie algebra as an approach to resolving the problem of rotation and orientation representations, [168].

By applying the well-known Cayley-Hamilton theorem to the characteristic equation of the skew symmetric matrix $\bar{\xi}^\sharp$, and using the fact that $\bar{\xi}$ is a unit vector, the following important identity is obtained:

$$(\bar{\xi}^\sharp)^3 + \bar{\xi}^\sharp = 0. \quad (\text{A.12})$$

A result of (A.12) is that the highest order of any power series of $\bar{\xi}^\sharp$ is two. Consequently, (A.9) can be written as

$$\begin{aligned} \mathbf{I}_E^O(\gamma, \bar{\xi}) &= \sum_{j=0}^{\infty} \frac{1}{j!} (\gamma \bar{\xi}^\sharp)^j \\ &= \mathbf{I}_3 + \sin(\gamma) \bar{\xi}^\sharp + [1 - \cos(\gamma)] (\bar{\xi}^\sharp)^2. \end{aligned} \quad (\text{A.13})$$

With the identity

$$(\bar{\xi}^\sharp)^2 \equiv \bar{\xi} \bar{\xi}^T - \mathbf{I}_3, \quad (\text{A.14})$$

the result in Equation (A.13) can also be written as

$$\mathbf{I}_E^O(\gamma, \bar{\xi}) = \cos(\gamma) \mathbf{I}_3 + \sin(\gamma) \bar{\xi}^\sharp + [1 - \cos(\gamma)] \bar{\xi} \bar{\xi}^T. \quad (\text{A.15})$$

The angle of rotation, γ , can be obtained from $\mathbf{I}_E^O$ either by using Eigen analysis with Property 2 of Section A.1 and recalling that the trace of a matrix equals the sum of its eigenvalues, or by noting that in (A.13) $\bar{\xi}^\sharp$ is skew symmetric and¹ $\text{Tr}\{(\bar{\xi}^\sharp)^2\} = -2$. In either case,

$$\cos(\gamma) = \frac{1}{2} (\text{Tr}\{\mathbf{I}_E^O\} - 1). \quad (\text{A.16})$$

In addition, the axis $\bar{\xi}$ may be obtained from (A.13) with

$$\bar{\xi}^\sharp = \frac{1}{2 \sin(\gamma)} (\mathbf{I}_E^O - (\mathbf{I}_E^O)^T). \quad (\text{A.17})$$

Because a rotation about $\bar{\xi}$ through an angle γ is equivalent to a rotation about $-\bar{\xi}$ through an angle $-\gamma$, the angle of rotation may be restricted to the interval $0 \leq \gamma \leq \pi$ for unique

¹The notation $\text{Tr}\{\mathbf{A}\}$ denotes the trace of the matrix $\mathbf{A}$.

inversion of the cosine in (A.16). In this case, $\bar{\xi}$ computed from (A.17) together with γ define the transformation I_E^O . However, numerical difficulties may arise when I_E^O is symmetric ($\gamma = 0, \pi$) for which the expression in (A.17) is of indeterminate form (0/0) indicating the mathematical singularities of this parameterization; when $\gamma = 0$ the axis $\bar{\xi}$ is undefined and when $\gamma = \pi$ the sign of $\bar{\xi}$ is arbitrary.

It is important to note the difference between the axis, $\bar{\xi}$, of the basis transformation operator, I_E^O , and what is commonly called the "axis of rotation" or instantaneous angular velocity of the object $\omega(t) := [\omega_E^O]_E(t)$. In the following development, all time derivatives are with respect to an observer in the earth-fixed frame, and, although the subscript 'E' is omitted, $\dot{\bar{\xi}}$ is expressed in F_E . Several identities are required in the derivation of ω , and are also used to write $\dot{\gamma}$ and $\dot{\bar{\xi}}$ in terms of ω :

1. Because $\bar{\xi}$ is by definition a unit vector,

$$\bar{\xi}^T \bar{\xi} = 1, \text{ and} \quad (\text{A.18})$$

$$\dot{\bar{\xi}}^T \bar{\xi} = \bar{\xi}^T \dot{\bar{\xi}} = 0. \quad (\text{A.19})$$

which means that $\bar{\xi}$ is always orthogonal to its time derivative.

2. For any vector $\mathbf{a}$, $\dot{\bar{\xi}} \times (\bar{\xi} \times \mathbf{a})$ is collinear with $\bar{\xi}$ and hence

$$\bar{\xi} \times (\dot{\bar{\xi}} \times (\bar{\xi} \times \mathbf{a})) = 0. \quad (\text{A.20})$$

In terms of matrix cross product operators, (A.20) is written as

$$\bar{\xi}^\# (\dot{\bar{\xi}})^\# \bar{\xi}^\# = 0. \quad (\text{A.21})$$

3. The set $\{\bar{\xi}, \dot{\bar{\xi}}, \bar{\xi}^\# \dot{\bar{\xi}}\}$ forms a right-handed orthogonal triad and hence

$$\bar{\xi}^T \bar{\xi}^\# \dot{\bar{\xi}} = 0, \text{ and} \quad (\text{A.22})$$

$$\bar{\xi}^\# \bar{\xi}^\# \dot{\bar{\xi}} = -\dot{\bar{\xi}}. \quad (\text{A.23})$$

Similar identities can be applied to various combinations of the three elements of this triad.

4. Two standard results from vector analysis give, for any $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$,

$$\begin{aligned}\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) &= \mathbf{a}^T \mathbf{c} \mathbf{b} - \mathbf{a}^T \mathbf{b} \mathbf{c}, \text{ and} \\ (\mathbf{a} \times \mathbf{b}) \times \mathbf{c} &= (\mathbf{b} \mathbf{a}^T - \mathbf{a} \mathbf{b}^T) \mathbf{c}.\end{aligned}\quad (\text{A.24})$$

Substitution of (A.13) into (A.7) followed by repeated applications of identities (A.12) and (A.21) results in

$$\boldsymbol{\omega}^\# = \dot{\gamma} \bar{\boldsymbol{\xi}}^\# + \sin(\gamma) \dot{\bar{\boldsymbol{\xi}}}^\# + (1 - \cos(\gamma)) (\bar{\boldsymbol{\xi}}^\# \dot{\bar{\boldsymbol{\xi}}}^\# - \dot{\bar{\boldsymbol{\xi}}}^\# \bar{\boldsymbol{\xi}}^\#). \quad (\text{A.25})$$

All terms in (A.25) are cross product operators. With identities (A.24) and using the obvious result that $\bar{\boldsymbol{\xi}}^\# \bar{\boldsymbol{\xi}} = \mathbf{0}$, the following result holds for any vector $\mathbf{c} \in \mathbb{R}^3$:

$$\begin{aligned}(\bar{\boldsymbol{\xi}}^\# \dot{\bar{\boldsymbol{\xi}}}^\# - \dot{\bar{\boldsymbol{\xi}}}^\# \bar{\boldsymbol{\xi}}^\#) \mathbf{c} &= \bar{\boldsymbol{\xi}} \times (\dot{\bar{\boldsymbol{\xi}}} \times \mathbf{c}) - \dot{\bar{\boldsymbol{\xi}}} \times (\bar{\boldsymbol{\xi}} \times \mathbf{c}) \\ &= (\dot{\bar{\boldsymbol{\xi}}} \bar{\boldsymbol{\xi}}^T - \bar{\boldsymbol{\xi}} \dot{\bar{\boldsymbol{\xi}}}^T) \mathbf{c} \\ &= (\bar{\boldsymbol{\xi}} \times \dot{\bar{\boldsymbol{\xi}}}) \times \mathbf{c}.\end{aligned}\quad (\text{A.26})$$

which implies that

$$(\bar{\boldsymbol{\xi}}^\# \dot{\bar{\boldsymbol{\xi}}}^\# - \dot{\bar{\boldsymbol{\xi}}}^\# \bar{\boldsymbol{\xi}}^\#) = (\bar{\boldsymbol{\xi}}^\# \dot{\bar{\boldsymbol{\xi}}})^\# . \quad (\text{A.27})$$

This result allows the cross product operators of (A.25) to be removed which gives,

$$\boldsymbol{\omega} = \dot{\gamma} \bar{\boldsymbol{\xi}} + \sin(\gamma) \dot{\bar{\boldsymbol{\xi}}} + (1 - \cos(\gamma)) \bar{\boldsymbol{\xi}}^\# \dot{\bar{\boldsymbol{\xi}}}^\# . \quad (\text{A.28})$$

Equation (A.28) demonstrates that the angular velocity has components on three orthogonal axis $\bar{\boldsymbol{\xi}}$, $\dot{\bar{\boldsymbol{\xi}}}$, and $\bar{\boldsymbol{\xi}}^\# \dot{\bar{\boldsymbol{\xi}}}^\#$.

Multiplying both sides of (A.28) on the left by $\bar{\boldsymbol{\xi}}^T$ and using identities (A.22) and its counterparts immediately gives

$$\dot{\gamma} = \bar{\boldsymbol{\xi}}^T \boldsymbol{\omega}. \quad (\text{A.29})$$

Multiplying both sides of (A.28) on the left by $\bar{\boldsymbol{\xi}}^\#$ and using identity (A.23) gives

$$\bar{\boldsymbol{\xi}}^\# \boldsymbol{\omega} = -(1 - \cos(\gamma)) \dot{\bar{\boldsymbol{\xi}}}^\# + \sin(\gamma) \bar{\boldsymbol{\xi}}^\# \dot{\bar{\boldsymbol{\xi}}}^\# . \quad (\text{A.30})$$

Multiplying both sides of (A.30) on the left by $\bar{\boldsymbol{\xi}}^\#$ and again using identity (A.23) gives

$$\bar{\boldsymbol{\xi}}^\# \bar{\boldsymbol{\xi}}^\# \boldsymbol{\omega} = -\sin(\gamma) \dot{\bar{\boldsymbol{\xi}}}^\# - (1 - \cos(\gamma)) \bar{\boldsymbol{\xi}}^\# \dot{\bar{\boldsymbol{\xi}}}^\# . \quad (\text{A.31})$$

Equations (A.30) and (A.31) can be treated as two equations in the two unknowns $\dot{\xi}$ and $\bar{\xi}^{\#}\dot{\xi}$. Solving (A.30) for the quantity $\bar{\xi}^{\#}\dot{\xi}$ and substituting the result into (A.30), followed by algebraic manipulation and the use of standard trigonometric identities gives

$$\dot{\xi} = -\frac{1}{2} \left\{ \bar{\xi}^{\#} + \cot\left(\frac{\gamma}{2}\right) (\bar{\xi}^{\#})^2 \right\} \omega. \quad (\text{A.32})$$

Equations (A.29) and (A.32) together define the relationship between the time derivative of the orientation vector and angular velocity,

$$\dot{\xi} = \left[\bar{\xi} \bar{\xi}^T - \frac{1}{2} \xi^z - \frac{\|\xi\|}{2} \cot\left(\frac{\|\xi\|}{2}\right) (\bar{\xi}^z)^2 \right] \omega. \quad (\text{A.33})$$

The complexity of this expression clearly indicates the difficulty of implementing an extended Kalman filter with the angle-axis parameterization if temporal behavior is modelled directly in terms of angular velocity.

A.3 Quaternion Parameterization

In order to extend the operations of three-dimensional vector algebra to include multiplication and division, Hamilton (1843) introduced an algebra for four-dimensional numbers, or quaternions. The objective of this section is to provide an introduction to quaternion calculus with emphasis on their use in representing change-of-basis transformations. Results of the previous section on the Euler angle-axis parameterization will play a major role in the following derivations. Branets and Shmyglevski [63] present a very detailed treatment of the application of quaternions to rigid body rotation problems and provide an elegant interpretation of rotation in terms of spherical geometry. Hughes [62] treats several parameterizations, including quaternions, in applications to attitude dynamics.

A quaternion is a hypercomplex number which is expressed in terms of basis elements consisting of the real number +1, and three imaginary units i, j, k , which satisfy

$$\begin{aligned} i^2 &= j^2 = k^2 = -1, \\ ij &= -ji = k, \\ jk &= -kj = i, \quad \text{and} \\ ki &= -ik = j. \end{aligned} \quad (\text{A.34})$$

A quaternion, $\mathbf{q}$ can be written² as

$$\mathbf{q} = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix} = q_1 i + q_2 j + q_3 k + q_4. \quad (\text{A.35})$$

Quaternions can be viewed as containing the real numbers $[0, 0, 0, a]^T$ with the real unit 1, the complex numbers $[b, 0, 0, a]^T$ with the two units 1 and i , and the vectors $[b, c, d, 0]^T$ in three-dimensional space. Some additional notation is required for developments of this section. Both quaternions and vectors are represented by bold face lower-case letters. As before, matrices are represented by bold upper-case letters (with the exception of matrix cross product operators). The vector part of a quaternion $\mathbf{q}$ is written as $\text{vect}(\mathbf{q}) = [q_1, q_2, q_3]^T$ while the real or scalar part of $\mathbf{q}$ is denoted by $\text{scal}(\mathbf{q}) = q_4$. With this notation,

$$\mathbf{q} = \text{vect}(\mathbf{q}) + \text{scal}(\mathbf{q}). \quad (\text{A.36})$$

A quaternion formed from a vector $\mathbf{r} \in \mathbb{R}^3$ by adding a zero scalar part will be denoted by $\text{quat}(\mathbf{r})$.

Addition of quaternions and their multiplication by a scalar (which can also be viewed as a quaternion) are performed in the same way as in an ordinary vector space. Quaternion multiplication, which is denoted here with the symbol ' $\circ$ ', is defined in terms of the products of basis elements shown in (A.34). Quaternion multiplication is associative, and distributive with respect to addition, but it is noncommutative which is evident from the last three rows of (A.34), i.e. $ij \neq ji$ for example. For any quaternion $\mathbf{q}$, the conjugate, $\mathbf{q}^*$, of $\mathbf{q}$ is defined as

$$\mathbf{q}^* := -\text{vect}(\mathbf{q}) + \text{scal}(\mathbf{q}), \quad (\text{A.37})$$

and the norm is defined as the Euclidean length of the 4-vector representing $\mathbf{q}$:

$$\|\mathbf{q}\| := \sqrt{q_1^2 + q_2^2 + q_3^2 + q_4^2}. \quad (\text{A.38})$$

$$= \sqrt{\mathbf{q} \circ \mathbf{q}^*}, \quad (\text{A.39})$$

²Some authors place the scalar quaternion element, q_4 , in the first position while others place the scalar element in the fourth position, as is the case here, when representing the quaternion in vector form.

where the second equation above follows from the rules of multiplication given in (A.34).
 Provided $\|\mathbf{q}\| \neq 0$, the inverse of $\mathbf{q}$ is

$$\mathbf{q}^{-1} := \frac{\mathbf{q}^*}{\|\mathbf{q}\|^2}, \quad (\text{A.40})$$

and satisfies $\mathbf{q} \circ \mathbf{q}^{-1} = \mathbf{q}^{-1} \circ \mathbf{q} = 1$. For any quaternion $\mathbf{q}$, the corresponding unit quaternion $\bar{\mathbf{q}}$ is given by

$$\bar{\mathbf{q}} := \frac{\mathbf{q}}{\|\mathbf{q}\|}, \quad (\text{A.41})$$

and note that $\bar{\mathbf{q}}^{-1} = \bar{\mathbf{q}}^*$. The set of unit quaternions can be placed in a two-to-one correspondence with elements of the three dimensional rotation matrices.

In Equation (A.13), the transformation from F_O to F_E was written as

$$\mathbf{I}_E^O = \mathbf{I}_3 + \sin(\gamma) \bar{\boldsymbol{\xi}}^z + [1 - \cos(\gamma)] (\bar{\boldsymbol{\xi}}^z)^2. \quad (\text{A.42})$$

Using the familiar double-angle formulas

$$\sin(\gamma) \equiv 2 \cos\left(\frac{\gamma}{2}\right) \sin\left(\frac{\gamma}{2}\right), \text{ and} \quad (\text{A.43})$$

$$\cos(\gamma) \equiv \cos^2\left(\frac{\gamma}{2}\right) - \sin^2\left(\frac{\gamma}{2}\right) \quad (\text{A.44})$$

$$= 1 - 2 \sin^2\left(\frac{\gamma}{2}\right) \quad (\text{A.45})$$

in (A.42) gives

$$\mathbf{I}_E^O = \mathbf{I}_3 - 2 \cos\left(\frac{\gamma}{2}\right) \left[-\sin\left(\frac{\gamma}{2}\right) \bar{\boldsymbol{\xi}}\right]^z + 2 \left(\left[-\sin\left(\frac{\gamma}{2}\right) \bar{\boldsymbol{\xi}}\right]^z\right)^2. \quad (\text{A.46})$$

The set of four parameters $[\boldsymbol{\eta}^T, \varepsilon]^T$, where

$$\begin{aligned} \boldsymbol{\eta} &:= -\sin\left(\frac{\gamma}{2}\right) \bar{\boldsymbol{\xi}} \\ \varepsilon &:= \cos\left(\frac{\gamma}{2}\right) \end{aligned} \quad (\text{A.47})$$

are called *Euler Parameters* [62] and can be identified with the unit quaternion

$$\bar{\mathbf{q}} = [\boldsymbol{\eta}^T, \varepsilon]^T. \quad (\text{A.48})$$

With

$$\mathbf{q} = \|\mathbf{q}\|(\boldsymbol{\eta} + \varepsilon), \quad (\text{A.49})$$

substitution of (A.47) into (A.46) gives the following matrix representation for $\mathbf{I}_E^O$ in terms of the elements of a quaternion $\mathbf{q}$:

$$\mathbf{I}_E^O(\mathbf{q}) = \frac{1}{\|\mathbf{q}\|^2} \begin{bmatrix} q_1^2 - q_2^2 - q_3^2 + q_4^2 & 2(q_1q_2 + q_3q_4) & 2(q_1q_3 - q_2q_4) \\ 2(q_1q_2 - q_3q_4) & -q_1^2 + q_2^2 - q_3^2 + q_4^2 & 2(q_2q_3 + q_1q_4) \\ 2(q_1q_3 + q_2q_4) & 2(q_2q_3 - q_1q_4) & -q_1^2 - q_2^2 + q_3^2 + q_4^2 \end{bmatrix}. \quad (\text{A.50})$$

It can be verified by direct computation, using the multiplication rules in (A.34), that for any vector $\mathbf{r}$, the representation of $\mathbf{r}$ in F_E can be obtained from its representation in F_O according to

$$\begin{aligned} \mathbf{r}_E &= \mathbf{I}_E^O(\mathbf{q})\mathbf{r}_O \\ &= \text{vect} \{ \mathbf{q}^{-1} \circ \text{quat}(\mathbf{r}_O) \circ \mathbf{q} \} \\ &= \text{vect} \{ \bar{\mathbf{q}}^* \circ \text{quat}(\mathbf{r}_O) \circ \bar{\mathbf{q}} \} \end{aligned} \quad (\text{A.51})$$

This property of quaternion calculus is extremely useful when multiple consecutive transformations (or rotations) must be performed as in the roll-pitch-yaw formulation which will be demonstrated in Section A.4.

Temporal behavior of the unit quaternion $\bar{\mathbf{q}}(t)$ can be obtained through differentiation of $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$ in (A.47) and the use of results in Equations (A.29) and (A.32) of Section A.2. With some algebraic manipulation, one easily obtains

$$\dot{\boldsymbol{\eta}} = -\frac{1}{2} (\boldsymbol{\eta}^\varepsilon + \boldsymbol{\varepsilon} \mathbf{I}_3) \boldsymbol{\omega}, \text{ and} \quad (\text{A.52})$$

$$\dot{\boldsymbol{\varepsilon}} = -\frac{1}{2} \boldsymbol{\eta}^T \boldsymbol{\omega}, \quad (\text{A.53})$$

where $\boldsymbol{\omega} \equiv [\boldsymbol{\omega}_E^O]_E = [w_1, w_2, w_3]^T$ is, as before, the angular velocity vector of F_O with respect to F_E expressed in F_E . Two equivalent expressions for $\dot{\bar{\mathbf{q}}}(t)$ are then given by using (A.48) in (A.52) and (A.53):

$$\dot{\bar{\mathbf{q}}} = \frac{\partial \bar{\mathbf{q}}}{\partial \boldsymbol{\omega}} \boldsymbol{\omega}, \text{ and} \quad (\text{A.54})$$

$$\dot{\bar{\mathbf{q}}} = \boldsymbol{\Omega}[\boldsymbol{\omega}(t)] \bar{\mathbf{q}}(t), \quad (\text{A.55})$$

where

$$\frac{\partial \dot{\bar{\mathbf{q}}}}{\partial \boldsymbol{\omega}} = -\frac{1}{2} \begin{bmatrix} \bar{q}_4 & -\bar{q}_3 & \bar{q}_2 \\ \bar{q}_3 & \bar{q}_4 & -\bar{q}_1 \\ -\bar{q}_2 & \bar{q}_1 & \bar{q}_4 \\ \bar{q}_1 & \bar{q}_2 & \bar{q}_3 \end{bmatrix}, \text{ and} \quad (\text{A.56})$$

$$\boldsymbol{\Omega}[\boldsymbol{\omega}] = \frac{1}{2} \begin{bmatrix} 0 & -w_3 & w_2 & -w_1 \\ w_3 & 0 & -w_1 & -w_2 \\ -w_2 & w_1 & 0 & -w_3 \\ w_1 & w_2 & w_3 & 0 \end{bmatrix}. \quad (\text{A.57})$$

Both (A.56) and (A.57) are required in the extended Kalman filtering equations.

With constant angular velocity, the solution of (A.55) with initial conditions given as $\bar{\mathbf{q}}(t_0)$ is simply

$$\bar{\mathbf{q}}(t) = \exp [\boldsymbol{\Omega} \cdot (t - t_0)] \bar{\mathbf{q}}(t_0). \quad (\text{A.58})$$

Note that $\boldsymbol{\Omega}$ is skew symmetric and, since the matrix exponential of a skew symmetric is always orthogonal, time propagation of the quaternion in (A.58) with unit length is maintained. It can be verified by direct computation that

$$\boldsymbol{\Omega}^2 = -\frac{1}{4} \|\boldsymbol{\omega}\|^2 \mathbf{I}_4. \quad (\text{A.59})$$

which implies that the highest power of $\boldsymbol{\Omega}$ in a power series is one. Expansion of (A.58) followed by repeated application of (A.59) gives

$$\bar{\mathbf{q}}(t) = \left[\cos (\|\boldsymbol{\omega}\|(t - t_0)/2) \mathbf{I}_4 + \frac{2}{\|\boldsymbol{\omega}\|} \sin (\|\boldsymbol{\omega}\|(t - t_0)/2) \boldsymbol{\Omega} \right] \bar{\mathbf{q}}(t_0). \quad (\text{A.60})$$

Equation (A.60) is only valid when $\boldsymbol{\omega}$ is constant. Closed form solutions for $\bar{\mathbf{q}}(t)$ are also available [56] for rotation with constant precession.

A.4 Roll-Pitch-Yaw Parameterization

The roll-pitch-yaw representation [32, 66, 170] is a particular case of an Euler angle parameterization of the three-dimensional rotation group. The roll, pitch, and yaw angles, denoted

ϕ , θ , and ψ , respectively, define an ordered sequence of three plane rotations which can be used to express the basis transformation $\mathbf{I}_E^O$ as

$$\mathbf{I}_E^O = \exp(\psi \mathbf{e}_3^\#) \exp(\theta \mathbf{e}_2^\#) \exp(\phi \mathbf{e}_1^\#), \quad (\text{A.61})$$

where $\mathbf{e}_i$, $i = 1, 2, 3$, are the standard basis vectors of $\mathbb{R}^3$. Expanding the exponentials with (A.13) gives the three plane rotations,

$$\begin{aligned} \exp(\phi \mathbf{e}_1^\#) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\phi) & -\sin(\phi) \\ 0 & \sin(\phi) & \cos(\phi) \end{bmatrix}, \\ \exp(\theta \mathbf{e}_2^\#) &= \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix}, \text{ and} \\ \exp(\psi \mathbf{e}_3^\#) &= \begin{bmatrix} \cos(\psi) & -\sin(\psi) & 0 \\ \sin(\psi) & \cos(\psi) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned} \quad (\text{A.62})$$

The angles ϕ , θ , and ψ have positive sense defined by the right hand rule about their respective rotation axis. Specifically, the rotation from F_E to F_O is carried out by the following sequence:

1. rotate about $\mathbf{k}_E$ by the yaw angle ψ ;
2. rotate about the axis $[-\sin(\psi), \cos(\psi), 0]_E^T$ by the pitch angle θ ; and
3. rotate about the axis $[\cos(\psi) \cos(\theta), \sin(\psi) \cos(\theta), -\sin(\theta)]_E^T$ by the roll angle ϕ .

In order to avoid ambiguities in these angles their ranges are often limited to

$$\begin{aligned} -\pi &\leq \psi < \pi, \\ -\frac{\pi}{2} &\leq \theta \leq \frac{\pi}{2}, \text{ and} \\ -\pi &\leq \phi < \pi. \end{aligned} \quad (\text{A.63})$$

The angular velocity of F_O with respect to F_E expressed in F_E , $\omega \equiv [\omega_E^O]_E$, is given in terms of the time derivatives of roll, pitch, and yaw angles as

$$[\omega_E^O]_E = \mathbf{J}_{RPY} \begin{bmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\psi} \end{bmatrix}, \quad (\text{A.64})$$

where

$$\mathbf{J}_{RPY} = \begin{bmatrix} \cos(\psi) \cos(\theta) & -\sin(\psi) & 0 \\ \sin(\psi) \cos(\theta) & \cos(\psi) & 0 \\ -\sin(\theta) & 0 & 1 \end{bmatrix}. \quad (\text{A.65})$$

contains the three rotational axis (from items 1, 2, and 3 above) as its columns (which follows from the vector addition property of angular velocity). The mathematical singularities, $\theta = \pm\pi/2$, characterize the orientations where the determinant of the Jacobian matrix $\det[\mathbf{J}_{RPY}] = -\cos(\theta)$ vanishes. At $\theta = \pi/2$, the roll and yaw angles are indistinguishable in the sense that the rotation matrix can be written as

$$\mathbf{I}_E^O(\theta = \frac{\pi}{2}) = \begin{bmatrix} 0 & \sin(\phi - \psi) & \cos(\phi - \psi) \\ 0 & \cos(\phi - \psi) & -\sin(\phi - \psi) \\ 1 & 0 & 0 \end{bmatrix}. \quad (\text{A.66})$$

and hence depends only on $\theta = \pi/2$ and the difference $\phi - \psi$.

The sequence of roll, pitch and yaw transformations can be represented in terms of corresponding unit quaternions,

$$\bar{\mathbf{q}}_\phi := \begin{bmatrix} -\sin(\phi/2) \\ 0 \\ 0 \\ \cos(\phi/2) \end{bmatrix}, \quad \bar{\mathbf{q}}_\theta := \begin{bmatrix} 0 \\ -\sin(\theta/2) \\ 0 \\ \cos(\theta/2) \end{bmatrix}, \quad \bar{\mathbf{q}}_\psi := \begin{bmatrix} 0 \\ 0 \\ -\sin(\psi/2) \\ \cos(\psi/2) \end{bmatrix}. \quad (\text{A.67})$$

according to

$$\begin{aligned} \mathbf{r}_E &= \mathbf{I}_E^O \mathbf{r}_O \\ &= \exp(\psi \mathbf{e}_3^\star) \left[\exp(\theta \mathbf{e}_2^\star) \left(\exp(\phi \mathbf{e}_1^\star) \mathbf{r}_O \right) \right] \\ &= \text{vect} \left\{ \bar{\mathbf{q}}_\psi^\star \circ \bar{\mathbf{q}}_\theta^\star \circ \bar{\mathbf{q}}_\phi^\star \circ \text{quat}(\mathbf{r}_O) \circ \bar{\mathbf{q}}_\phi \circ \bar{\mathbf{q}}_\theta \circ \bar{\mathbf{q}}_\psi \right\}. \end{aligned} \quad (\text{A.68})$$

A single equivalent quaternion representing the basis transformation can be then defined as

$$\bar{\mathbf{q}}_{\text{equiv}} := \bar{\mathbf{q}}_{\phi} \circ \bar{\mathbf{q}}_{\theta} \circ \bar{\mathbf{q}}_{\psi}. \quad (\text{A.69})$$

Noting that multiplication and conjugation rules for quaternions imply that

$$\bar{\mathbf{q}}_{\text{equiv}}^* = \bar{\mathbf{q}}_{\psi}^* \circ \bar{\mathbf{q}}_{\theta}^* \circ \bar{\mathbf{q}}_{\phi}^* \quad (\text{A.70})$$

the transformation operation can be written as

$$\mathbf{I}_E^O \mathbf{r}_O = \text{vect} \left\{ \bar{\mathbf{q}}_{\text{equiv}}^* \circ \text{quat}(\mathbf{r}_O) \circ \bar{\mathbf{q}}_{\text{equiv}} \right\}. \quad (\text{A.71})$$

The Euler axis-angle representation corresponding to a particular set of roll, pitch and yaw angles can then easily be extracted from the equivalent quaternion.

By forming the product of the three plane rotations in (A.61) and defining the elements of $\mathbf{I}_E^O$ as ℓ_{ij} , the following algorithm can be used to extract a valid set of roll (ϕ), pitch (θ), and yaw (ψ) angles:

$$\begin{aligned} & \text{Check for singularity with } \epsilon \ll 1 \\ & \text{if } ((|\ell_{11}| < \epsilon) \ \& \ (|\ell_{21}| < \epsilon)), \\ & \quad \phi = 0.0; \\ & \quad \psi = -\text{atan2}(\ell_{12}, \ell_{22}); \\ & \quad \text{if } (\ell_{31} > 0.0), \\ & \quad \quad \theta = \pi/2; \\ & \quad \text{else} \\ & \quad \quad \theta = -\pi/2; \\ & \text{else} \\ & \quad \phi = \text{atan2}(\ell_{32}, \ell_{33}); \\ & \quad \theta = \text{atan} \left(\frac{-\ell_{31}}{\sqrt{\ell_{11}^2 + \ell_{21}^2}} \right); \\ & \quad \psi = \text{atan2}(\ell_{21}, \ell_{11}); \\ & \text{end.} \end{aligned} \quad (\text{A.72})$$

Functions $\text{atan2}(y, x)$ and $\text{atan}(x)$ return, respectively, the four-quadrant arctangent of y/x and the two-quadrant arctangent of x , and are available in most programming languages. Note that in the singular orientations ($\theta = \pm\pi$), the roll angle is arbitrarily set to zero.

Appendix B

Kalman Filtering Review

This appendix provides a review and discussion of the linear discrete-time Kalman filter, the discrete-time extended Kalman filter, and the continuous-discrete Kalman filter. Local iterations which lead to the iterated extended Kalman filter are also introduced. Applications of Kalman filtering to trajectory estimation problems has received widespread attention since the introduction of Kalman's sequential state estimation technique [6]. Chang and Tabaczynski [70] give a thorough discussion of many important aspects of Kalman filtering approaches for target tracking problems. In general, the trajectory estimation problem is one of nonlinear estimation for which the optimal (conditional mean) nonlinear estimator cannot be realized with a finite-dimensional implementation. Furthermore, the mathematical model representing the target trajectory is seldom exact. Consequently, most practical implementations of nonlinear tracking filters are suboptimal.

The Extended Kalman Filter (EKF) is a suboptimal approach which applies linear Kalman filtering theory to a linear approximation of the nonlinear problem. A rigorous development of filtering equations requires the use of stochastic integral and differential equations and mathematical expectation conditioned on σ -algebras generated by the observation process, and formulates solutions through Ito calculus (see for example [9]). In applications, however, the somewhat more intuitive approaches using "formal" manipulations of white noise processes and expectation conditioned on sets of observations are often preferred [10, 70, 59]. This approach will also be taken in the following review of the EKF formulation. The following subsections draw primarily from [9]-[13].

In many cases, the target tracking problem can be formulated in terms of discrete-time

system (plant) and measurement models. In the case of image-based tracking, measurements are taken from image frames which arrive at discrete instants in time. The measurement model for this class of problems, therefore, is always discrete. In addition, the perspective projection image formation model results in a nonlinear measurement model. If the system model is continuous in time, it can often be discretized, or the "continuous-discrete" [9] form of the filter can be employed.

Notation and filtering equations for the discrete-time linear state estimation problem are reviewed in Section B.1, with extensions to approximate filters for the nonlinear discrete-time problem in Section B.2 and the continuous-discrete form of extended Kalman filtering in Section B.3. Finally, Section B.4 describes local iterations which enhance filter performance in the presence of nonlinear measurement models.

B.1 Discrete-Time Kalman Filter

Consider a linear discrete-time dynamical system model given by

$$\mathbf{x}(k+1) = \Phi(k)\mathbf{x}(k) + \mathbf{B}(k)\mathbf{u}(k) + \mathbf{G}(k)\mathbf{w}(k), \quad (\text{B.1})$$

where $k \sim$ an integer valued sample index, time $t = kT$ where T is the sample period, and

$$\begin{aligned} \mathbf{x} \in \mathbb{R}^n &\sim \text{vector of state variables,} \\ \mathbf{u} \in \mathbb{R}^q &\sim \text{deterministic input vector,} \\ \mathbf{w} \in \mathbb{R}^p &\sim \text{zero mean, white, Gaussian, system noise} \\ &\quad \text{process with covariance matrix } \mathbf{Q}(\mathbf{k}), \\ \Phi \in \mathbb{R}^{n \times n} &\sim \text{state transition matrix,} \\ \mathbf{B} \in \mathbb{R}^{n \times q} &\sim \text{input matrix mapping the input vector} \\ &\quad \mathbf{u} \text{ into the state space, and} \\ \mathbf{G} \in \mathbb{R}^{n \times p} &\sim \text{matrix mapping the noise process} \\ &\quad \mathbf{w} \text{ into the state space.} \end{aligned} \quad (\text{B.2})$$

The linear, discrete-time measurement model is given by

$$\mathbf{z}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{v}(k), \quad (\text{B.3})$$

where

$$\begin{aligned}
 \mathbf{z} \in \mathbb{R}^m &\sim \text{vector of observations or measurements,} \\
 \mathbf{H} \in \mathbb{R}^{n \times m} &\sim \text{measurement matrix relating state variables} \\
 &\quad \text{to measurement variables, and} \\
 \mathbf{v} \in \mathbb{R}^m &\sim \text{zero mean, white, Gaussian, measurement noise} \\
 &\quad \text{process with covariance matrix } \mathbf{R}(k).
 \end{aligned} \tag{B.4}$$

With definitions (B.2) and (B.4), the system and measurement noise processes satisfy

$$\begin{aligned}
 \mathcal{E}\{\mathbf{w}(k)\} &= \mathbf{0} & \mathcal{E}\{\mathbf{w}(k)\mathbf{w}^T(j)\} &= \mathbf{Q}(k)\delta_{kj} \\
 \mathcal{E}\{\mathbf{v}(k)\} &= \mathbf{0} & \mathcal{E}\{\mathbf{v}(k)\mathbf{v}^T(j)\} &= \mathbf{R}(k)\delta_{kj}.
 \end{aligned} \tag{B.5}$$

where $\mathcal{E}\{\cdot\}$ denotes expectation and δ_{jk} is the Kronecker delta product. In the present analysis, $\mathbf{w}$ and $\mathbf{v}$ are also assumed to be uncorrelated, i.e.,

$$\mathcal{E}\{\mathbf{w}(k)\mathbf{v}^T(j)\} = \mathbf{0}. \tag{B.6}$$

for all integers j and k .

The expectation of the state $\mathbf{x}$ at time step k conditioned on all measurements up to time step j is denoted by

$$\hat{\mathbf{x}}(k|j) := \mathcal{E}\{\mathbf{x}(k)|Z_j\} \tag{B.7}$$

where $Z_j = \{\mathbf{z}(i), i = 1, 2, \dots, j\}$ represents the set of all measurements up to and including the j th time step. It is precisely the conditional mean in (B.7), usually with $j = k$, which the optimal filter is required to estimate. In the case of linear system and measurement models, assuming exact mathematical models and characterization of the noise processes, the Kalman filter provides a recursive “formula” for this optimal estimate. It is well known that in this case the Kalman filter provides an unbiased, minimum variance, consistent estimate $\hat{\mathbf{x}}(k|k)$, which is also the maximum likelihood estimate of the true state $\mathbf{x}(k)$.

The error between the filter state and the true state is given by

$$\tilde{\mathbf{x}}(k|j) := \mathbf{x}(k) - \hat{\mathbf{x}}(k|j), \tag{B.8}$$

with covariance matrix

$$\mathbf{P}(k|j) := \mathcal{E}\{\tilde{\mathbf{x}}(k|j)\tilde{\mathbf{x}}^T(k|j)\}. \tag{B.9}$$

The development of the filtering equations assumes that the true initial state of the system $\mathbf{x}(0) \sim N(\bar{\mathbf{x}}_0, \mathbf{P}_0)$ ¹. Throughout this work, it is assumed that an initial estimate, $\hat{\mathbf{x}}(0|-1)$, of the mean of the true initial state, $\mathcal{E}\{\mathbf{x}(0)\}$, is available just prior to taking initial measurements, $\mathbf{z}(0)$. In theory, then, the recursive filter is initialized with

$$\begin{aligned}\hat{\mathbf{x}}(0|-1) &= \mathcal{E}\{\mathbf{x}(0)\} := \bar{\mathbf{x}}_0, \text{ and} \\ \mathbf{P}(0|-1) &= \mathcal{E}\{\tilde{\mathbf{x}}(0|-1)\tilde{\mathbf{x}}^T(0|-1)\} := \mathbf{P}_0.\end{aligned}\quad (\text{B.10})$$

In practice, however, the filter is often initialized with an unbiased estimator of $\bar{\mathbf{x}}_0$ for which some *a priori* statistical information is known of the form

$$\hat{\mathbf{x}}(0|-1) \sim N(\bar{\mathbf{x}}_0, \mathbf{P}_0). \quad (\text{B.11})$$

After $k-1$ measurements have been taken, the recursive equations of the linear Kalman filter are summarized as follows:

1. State and covariance prediction (or time update) -

$$\begin{aligned}\hat{\mathbf{x}}(k|k-1) &= \Phi(k-1)\hat{\mathbf{x}}(k-1|k-1) + \mathbf{B}(k-1)\mathbf{u}(k-1), \text{ and} \\ \mathbf{P}(k|k-1) &= \Phi(k-1)\mathbf{P}(k-1|k-1)\Phi^T(k-1) \\ &\quad + \mathbf{G}(k-1)\mathbf{Q}(k-1)\mathbf{G}^T(k-1).\end{aligned}\quad (\text{B.12})$$

2. Measurement update computed following the k th measurement -

$$\begin{aligned}\mathbf{K}(k) &= \mathbf{P}(k|k-1)\mathbf{H}^T(k) \\ &\quad \cdot [\mathbf{H}(k)\mathbf{P}(k|k-1)\mathbf{H}^T(k) + \mathbf{R}(k)]^{-1}, \\ \hat{\mathbf{x}}(k|k) &= \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k)[\mathbf{z}(k) - \mathbf{H}(k)\hat{\mathbf{x}}(k|k-1)], \text{ and} \\ \mathbf{P}(k|k) &= [\mathbf{I} - \mathbf{K}(k)\mathbf{H}(k)]\mathbf{P}(k|k-1).\end{aligned}\quad (\text{B.13})$$

An equivalent form, known as the "information form", of the measurement update stage of the filter, which is often used in place of (B.13), is given by

$$\begin{aligned}\mathbf{P}^{-1}(k|k) &= \mathbf{P}^{-1}(k|k-1) + \mathbf{H}^T(k)\mathbf{R}^{-1}(k)\mathbf{H}(k), \\ \mathbf{K}(k) &= \mathbf{P}(k|k)\mathbf{H}^T(k)\mathbf{R}^{-1}(k), \text{ and} \\ \hat{\mathbf{x}}(k|k) &= \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k)[\mathbf{z}(k) - \mathbf{H}(k)\hat{\mathbf{x}}(k|k-1)].\end{aligned}\quad (\text{B.14})$$

¹The notation $N(\mathbf{m}, \mathbf{C})$ denotes a normally distributed random variable with mean $\mathbf{m}$ and covariance $\mathbf{C}$.

In any implementation of the above filtering equations, the following parameters must be specified in addition to the measurement sequence: 1) an initial state estimate $\hat{\mathbf{x}}(0|-1)$; 2) an estimate of the initial error covariance $\mathbf{P}_0$; 3) the system or plant noise covariances $\mathbf{Q}(k)$; and 4) the measurement or observation noise covariances $\mathbf{R}(k)$. The initial state estimate and its error covariance matrix often result from a batch estimation technique applied to the first few image frames [51, 54, 32]. In simulations, the measurement noise covariance is known, however, in applications to real imagery, determination of the actual measurement noise is difficult since this involves modeling of various sources of error in the imaging system and feature detection algorithms. Specification of the system noise covariance $\mathbf{Q}(k)$ is seldom exact in object tracking problems since unmodeled dynamics resulting from manoeuvring object trajectories are difficult to quantify. As a result, $\mathbf{Q}(k)$ together with the matrix $\mathbf{G}(k)$ are often treated as "tuning" parameters which are selected to give satisfactory convergent behavior of the filter over a wide range of applications.

In Equation (B.12) the previous state estimate $\hat{\mathbf{x}}(k-1|k-1)$ is propagated in time between measurements according to the noise-free dynamic system model to obtain a state estimate $\hat{\mathbf{x}}(k|k-1)$ just prior to a measurement event. Once measurements are available, the state estimate is corrected in (B.13) by an amount proportional to the current error $[\mathbf{z}(k) - \mathbf{H}(k)\hat{\mathbf{x}}(k|k-1)]$ called the innovation. The proportionality factor $\mathbf{K}(k)$ is called the Kalman gain which, from (B.14), is proportional to the uncertainty in the state estimate $\mathbf{P}(k|k)$ and inversely proportional to the measurement uncertainty $\mathbf{R}(k)$. The system noise covariance $\mathbf{Q}(k)$ which is mapped to the state space through $\mathbf{G}(k)$ adds a positive definite matrix to the filter error covariance matrix during each time update (B.12) which reflects a "diffusion" of confidence in the state estimate over the time interval between measurements. In the absence of system noise ($\mathbf{Q} \equiv \mathbf{0}$) the error covariance and gain matrices may decrease monotonically in the presence of measurement data, depending on the state transition matrix through the similarity transformation in (B.12), i.e., consider (B.14) for the scalar case with $\mathbf{H} \equiv 1$. In this case, the filter eventually neglects further measurements and operates solely on the system model which, with $\mathbf{Q} \equiv \mathbf{0}$, is assumed to be exact.

B.2 Discrete-Time Extended Kalman Filter

The problem of estimating the state of a nonlinear system can be formulated as an extension of mean square estimation to dynamic systems. Formally, the method for deriving the fundamental equations of the optimal filter is quite similar to that used to design the optimal linear filter (Kalman filter) for linear systems with Gaussian noise processes. However, in the nonlinear case, the explicit analytical solution cannot, in general, be found and hence approximate numerical solutions are often considered.

In the case of linear systems, all of the variables and processes involved (states, noises, measurements, estimates, and errors) are usually assumed to be normally distributed. Therefore, mean values and covariance matrices are "sufficient statistics." In the nonlinear case, on the other hand, the state, observation, and estimation processes are not, in general, normally distributed even if the input noise sources are. As a result, the entire probability distribution may be required for optimal filter design. In addition, the filter performance and parameters are often dependent on the filter estimate so that it is even more difficult to obtain bounds for filter performance.

Consider the nonlinear system model given by

$$\mathbf{x}(k+1) = \mathbf{f}[\mathbf{x}(k)] + \mathbf{B}(k)\mathbf{u}(k) + \mathbf{g}[\mathbf{x}(k)]\mathbf{w}(k), \quad (\text{B.15})$$

and the nonlinear measurement model

$$\mathbf{z}(k) = \mathbf{h}[\mathbf{x}(k)] + \mathbf{v}(k), \quad (\text{B.16})$$

where $\mathbf{f} : R^n \mapsto R^n$, $\mathbf{g} : R^n \mapsto (R^{n \times p})$, and $\mathbf{h} : R^n \mapsto R^m$ are, in general, nonlinear functions of the state vector $\mathbf{x}(k)$. The mapping of the input vector $\mathbf{u}$ to the state space may also be nonlinear, but this case will not be considered here and, in any event, does not alter the filter structure significantly.

The Extended Kalman Filter (EKF) is based on the concept that if the nonlinearities in $\mathbf{f}$ and $\mathbf{h}$ are sufficiently smooth, these functions can be expanded in a Taylor series and approximated by linear terms. Furthermore, if the filter works properly, it will provide a reasonable estimate of the conditional mean $\mathcal{E}\{\mathbf{x}(k)|Z_k\}$ which is what the optimal filter is required to estimate. For this reason, the Taylor series expansion of nonlinear terms is for-

culated about the most recent state estimate. The primary consequence of this "feedback" of the estimate into the model is that the EKF is a nonlinear estimation technique.

With the definitions

$$\Phi(k) := \left. \frac{\partial \mathbf{f}[\mathbf{x}]}{\partial \mathbf{x}} \right|_{\mathbf{x}=\hat{\mathbf{x}}(k|k)}, \quad (\text{B.17})$$

$$\mathbf{G}(k) := \mathbf{g}[\hat{\mathbf{x}}(k|k)], \quad \text{and} \quad (\text{B.18})$$

$$\mathbf{H}(k) := \left. \frac{\partial \mathbf{h}[\mathbf{x}]}{\partial \mathbf{x}} \right|_{\mathbf{x}=\hat{\mathbf{x}}(k|k-1)}, \quad (\text{B.19})$$

and neglecting all terms beyond first order in the Taylor series expansions, approximate linear system and measurement models can be written, respectively, as

$$\mathbf{x}(k+1) = \Phi(k)\mathbf{x}(k) + \hat{\mathbf{u}}(k) + \mathbf{G}(k)\mathbf{w}(k), \quad \text{and} \quad (\text{B.20})$$

$$\mathbf{z}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{y}(k) + \mathbf{v}(k), \quad (\text{B.21})$$

where

$$\hat{\mathbf{u}}(k) := \mathbf{f}[\hat{\mathbf{x}}(k|k)] - \Phi(k)\hat{\mathbf{x}}(k|k) + \mathbf{B}(k)\mathbf{u}(k), \quad \text{and} \quad (\text{B.22})$$

$$\mathbf{y}(k) := \mathbf{h}[\hat{\mathbf{x}}(k|k-1)] - \mathbf{H}(k)\hat{\mathbf{x}}(k|k-1). \quad (\text{B.23})$$

The EKF equations are formulated, essentially, as the standard Kalman filtering equations for the linear system given by (B.17)-(B.23).

After $k-1$ measurements have been taken, the recursive equations of the extended Kalman filter are summarized as follows:

1. State and covariance prediction (or time update) -

$$\hat{\mathbf{x}}(k|k-1) = \mathbf{f}[\hat{\mathbf{x}}(k-1|k-1)] + \mathbf{B}(k)\mathbf{u}(k), \quad \text{and} \quad (\text{B.24})$$

$$\begin{aligned} \hat{\mathbf{P}}(k|k-1) &= \Phi(k-1)\hat{\mathbf{P}}(k-1|k-1)\Phi^T(k-1) \\ &\quad + \mathbf{G}(k-1)\mathbf{Q}(k-1)\mathbf{G}^T(k-1). \end{aligned}$$

2. Measurement update computed following the k th measurement -

$$\begin{aligned} \mathbf{K}(k) &= \hat{\mathbf{P}}(k|k-1)\mathbf{H}^T(k) \\ &\quad \cdot [\mathbf{H}(k)\hat{\mathbf{P}}(k|k-1)\mathbf{H}^T(k) + \mathbf{R}(k)]^{-1}. \end{aligned} \quad (\text{B.25})$$

$$\hat{\mathbf{x}}(k|k) = \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k) \{ \mathbf{z}(k) - \mathbf{h}[\hat{\mathbf{x}}(k|k-1)] \}, \quad \text{and}$$

$$\hat{\mathbf{P}}(k|k) = [\mathbf{I} - \mathbf{K}(k)\mathbf{H}(k)]\hat{\mathbf{P}}(k|k-1).$$

The accent on $\hat{\mathbf{P}}$ has been used to emphasize the fact that the covariance estimate is dependent on the filter estimate.

B.3 Continuous-Discrete Extended Kalman Filter

In Section B.2, both the system and measurement models were assumed to be available in discrete-time form. In general, however, the system model is often developed in continuous-time form, while measurements are taken at discrete instants in time.

Consider a continuous-time system which is described by the stochastic differential equation

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}, t) + \mathbf{G}(t)\mathbf{w}(t), \quad t \geq 0, \quad \mathbf{x}(0) \sim N(\bar{\mathbf{x}}_0, \mathbf{P}_0). \quad (\text{B.26})$$

where $\mathbf{w}(t)$ is a zero-mean white Gaussian noise process with

$$\mathcal{E} \{ \mathbf{w}(t)\mathbf{w}^T(\tau) \} = \mathbf{Q}(t)\delta(t - \tau), \quad (\text{B.27})$$

and $\delta(t - \tau)$ is the Dirac-delta function. The system model may also depend on a deterministic input vector $\mathbf{u}(t)$. To simplify the notation, this dependence is not shown explicitly here and, in any event, does not alter the filter structure significantly. As in the discrete-time case of Section B.2, the nonlinear system model of (B.26) is linearized approximately through a first-order Taylor series expansion about a nominal trajectory $\hat{\mathbf{x}}(t)$ which satisfies

$$\dot{\hat{\mathbf{x}}}(t) = \mathbf{f}(\hat{\mathbf{x}}, t), \quad t \geq 0. \quad (\text{B.28})$$

The state perturbation is defined as

$$\tilde{\mathbf{x}}(t) := \mathbf{x}(t) - \hat{\mathbf{x}}(t), \quad (\text{B.29})$$

which satisfies the stochastic differential equation

$$\begin{aligned} \dot{\tilde{\mathbf{x}}}(t) &= \mathbf{f}(\mathbf{x}, t) - \mathbf{f}(\hat{\mathbf{x}}, t) + \mathbf{G}(t)\mathbf{w}(t) \\ &\approx \mathbf{F}[t; \hat{\mathbf{x}}]\tilde{\mathbf{x}}(t) + \mathbf{G}(t)\mathbf{w}(t), \end{aligned} \quad (\text{B.30})$$

with the initial condition $\tilde{\mathbf{x}}(0) \sim N(\tilde{\mathbf{x}}_0 - \hat{\mathbf{x}}(0), \mathbf{P}_0)$ and where

$$\mathbf{F}[t; \hat{\mathbf{x}}] := \left. \frac{\partial \mathbf{f}(\mathbf{x}, t)}{\partial \mathbf{x}} \right|_{\mathbf{x}=\hat{\mathbf{x}}} \quad (\text{B.31})$$

It should be noted that $\mathbf{F}$ is evaluated at particular values of $\hat{\mathbf{x}}(t)$ and hence is a function of t only. Specifically, as in the discrete-time case of Section B.2, the nominal trajectory is taken as the most recent state estimate which, for the linearized system model, is given by

$$\hat{\mathbf{x}}(t_k) \equiv \hat{\mathbf{x}}(k|k). \quad (\text{B.32})$$

In principle, equation (B.30) can be written in discrete-time form as

$$\tilde{\mathbf{x}}(k+1) = \hat{\Phi}(t_{k+1}, t_k) \tilde{\mathbf{x}}(k) + \mathbf{w}_d(k+1). \quad (\text{B.33})$$

where $\hat{\Phi}(t_{k+1}, t_k)$ is the state transition matrix which satisfies [9]

$$\begin{aligned} \frac{d\hat{\Phi}(t, \tau)}{dt} &= \mathbf{F}[t; \mathbf{x}(t)] \hat{\Phi}(t, \tau), \\ \hat{\Phi}(\tau, \tau) &= \mathbf{I}_n, \quad \forall \tau, \\ \hat{\Phi}(t, \tau) \hat{\Phi}(\tau, \zeta) &= \hat{\Phi}(t, \zeta), \quad \forall t, \tau, \zeta. \end{aligned} \quad (\text{B.34})$$

and $\mathbf{w}_d(k+1)$ is a zero-mean, white Gaussian sequence, $\mathbf{w}_d(k+1) \sim N(0, \mathbf{Q}_d(k+1))$, where

$$\mathbf{Q}_d(k+1) = \int_{t_k}^{t_{k+1}} \dot{\Phi}(t_{k+1}, \tau) \mathbf{G}(\tau) \mathbf{Q}(\tau) \mathbf{G}^T(\tau) \dot{\Phi}^T(t_{k+1}, \tau) d\tau. \quad (\text{B.35})$$

If $\mathbf{F}$ is approximately constant over the interval $(t_k, t_{k+1}]$, where $t_{k+1} - t_k = T_k$, the state transition matrix for each sample interval can be estimated as

$$\hat{\Phi}(t_{k+1}, t_k) = \exp(\mathbf{F}T_k). \quad (\text{B.36})$$

In general, however, the computation of the state transition matrix is often difficult and, therefore, replaced by numerical integration of the Riccati covariance equation [9]

$$\begin{aligned} \dot{\hat{\mathbf{P}}}(t|t_k) &= \mathbf{F}[t; \hat{\mathbf{x}}(t|t_k)] \hat{\mathbf{P}}(t|t_k) \\ &\quad + \hat{\mathbf{P}}(t|t_k) \mathbf{F}^T[t; \hat{\mathbf{x}}(t|t_k)] + \mathbf{G}(t) \mathbf{Q}(t) \mathbf{G}^T(t). \end{aligned} \quad (\text{B.37})$$

The nonlinear, discrete-time measurement model is again given by (B.16) and is treated in a manner identical to that used in Section B.2. The extended Kalman filter is then formulated based on the approximate linear perturbation equation (B.30), or its discrete form (B.33), and the linear approximation of the measurement model (B.21). However, in this case, the state prediction is computed by integration of the nonlinear noise-free system differential equation. After $k - 1$ measurements have been taken, the recursive equations of the continuous-discrete extended Kalman filter are summarized as follows:

1. State and covariance prediction (or time update) - Either

$$\begin{aligned}\hat{\mathbf{x}}(k|k-1) &= \hat{\mathbf{x}}(k-1|k-1) + \int_{t_{k-1}}^{t_k} \mathbf{f}[\hat{\mathbf{x}}(t|t_{k-1}), t] dt \text{ and} \\ \hat{\mathbf{P}}(k|k-1) &= \hat{\mathbf{P}}(k-1|k-1) + \int_{t_{k-1}}^{t_k} \hat{\mathbf{P}}[t; \hat{\mathbf{x}}(t|t_{k-1})] dt, \quad (\text{B.38})\end{aligned}$$

where $\hat{\mathbf{P}}[t; \hat{\mathbf{x}}(t|t_{k-1})]$ is defined in (B.37), or equivalently,

$$\begin{aligned}\hat{\mathbf{x}}(k|k-1) &= \hat{\Phi}(k-1)\hat{\mathbf{x}}(k-1|k-1) \\ \hat{\mathbf{P}}(k|k-1) &= \hat{\Phi}(k-1)\hat{\mathbf{P}}(k-1|k-1)\hat{\Phi}^T(k-1) + \mathbf{Q}_d(k).\end{aligned} \quad (\text{B.39})$$

2. Measurement update computed following the k th measurement -

$$\begin{aligned}\mathbf{K}(k) &= \hat{\mathbf{P}}(k|k-1)\mathbf{H}^T(k) \\ &\quad \cdot [\mathbf{H}(k)\hat{\mathbf{P}}(k|k-1)\mathbf{H}^T(k) + \mathbf{R}(k)]^{-1}, \quad (\text{B.40}) \\ \hat{\mathbf{x}}(k|k) &= \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k) \{ \mathbf{z}(k) - \mathbf{h}[\hat{\mathbf{x}}(k|k-1)] \}, \text{ and} \\ \hat{\mathbf{P}}(k|k) &= [\mathbf{I} - \mathbf{K}(k)\mathbf{H}(k)]\hat{\mathbf{P}}(k|k-1).\end{aligned}$$

Again, the accent on $\hat{\mathbf{P}}$ has been used to emphasize the fact that the covariance estimate is dependent on the filter estimate.

B.4 Local Iterations

The Iterated Extended Kalman Filter (IEKF) is designed to reduce the effect of measurement nonlinearities on the performance of the linear filter applied to a linearized nonlinear system [9, p. 279]. Between observations the conditional mean and covariance matrix

propagate according to the equations of (B.24). Following the time update of the state and covariance from time step $k - 1$ to time step k , and with the k th measurement available, the iteration is initialized with

$$\hat{\mathbf{x}}(k|k)_0 = \hat{\mathbf{x}}(k|k - 1). \quad (\text{B.41})$$

Iterations are carried out on the measurement update equations in (B.25) according to the following algorithm:

1. Linearize the measurement equations about the current iterate -

$$\mathbf{H}(k)_n = \left. \frac{\partial \mathbf{h}[\mathbf{x}]}{\partial \mathbf{x}} \right|_{\mathbf{x}=\hat{\mathbf{x}}(k|k)_n}. \quad (\text{B.42})$$

2. Compute the Kalman gain -

$$\begin{aligned} \mathbf{K}(k)_n &= \hat{\mathbf{P}}(k|k - 1)\mathbf{H}^T(k)_n \\ &\cdot [\mathbf{H}(k)_n\hat{\mathbf{P}}(k|k - 1)\mathbf{H}^T(k)_n + \mathbf{R}(k)]^{-1}. \end{aligned} \quad (\text{B.43})$$

3. Compute the measurement update -

$$\begin{aligned} \hat{\mathbf{x}}(k|k)_{n+1} &= \hat{\mathbf{x}}(k|k - 1) + \mathbf{K}(k)_n \{ \mathbf{z}(k) - \mathbf{h}[\hat{\mathbf{x}}(k|k)_n] \\ &\quad - \mathbf{H}(k)_n [\hat{\mathbf{x}}(k|k - 1) - \hat{\mathbf{x}}(k|k)_n] \}. \end{aligned} \quad (\text{B.44})$$

Iteration is terminated when there is no significant change in consecutive iterates. Then the last iterate, $\hat{\mathbf{x}}(k|k)_N$ say, is taken for the estimate $\hat{\mathbf{x}}(k|k)$ and the covariance matrix is computed based on the linearization about $\hat{\mathbf{x}}(k|k)_N$,

$$\hat{\mathbf{P}}(k|k) = [\mathbf{I} - \mathbf{K}(k)_N\mathbf{H}(k)_N] \hat{\mathbf{P}}(k|k - 1). \quad (\text{B.45})$$

The iterative procedure described by (B.42)-(B.44) can be interpreted as a modified Newton-Raphson search for the conditional mode of the posterior density of the state estimate given the measurement at time step k , assuming the prior density is Gaussian [9, pp. 349-351]. The mode of the posterior density is then used for the conditional mean. In non-linear problems, these local iterations can be expected to produce a biased estimate since the mode is not, in general, equivalent to the mean. However, as the error variance becomes small, so does the bias in this estimate.

Appendix C

Cramer-Rao Covariance Bounds

The extended Kalman filter is generally a suboptimal state estimation technique. A natural consideration in any estimation problem is how well the proposed technique performs when compared to the optimal performance possible. A powerful result which provides a performance assessment of parameter estimation techniques is the Cramer-Rao inequality [67]. The following derivation draws from methods outlined in [9, 67, 68, 69, 70].

Consider a nonlinear discrete-time estimation problem with noise-free system model

$$\mathbf{x}(k+1) = \mathbf{f}[\mathbf{x}(k)] \quad (\text{C.1})$$

and measurement model

$$\mathbf{z}(k) = \mathbf{h}[\mathbf{x}(k)] + \mathbf{v}(k), \quad (\text{C.2})$$

where $\mathbf{v}(k) \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ and is temporally white. Assume further that some unbiased *a priori* statistical information about $\mathbf{x}(0)$ is available of the form

$$\mathbf{x}_0 \sim \mathcal{N}(\mathbf{x}(0), \mathbf{P}_0), \quad (\text{C.3})$$

where $\mathbf{x}_0$ and $\mathbf{v}(k)$ are independent. The objective is to determine how well $\mathbf{x}(N)$, $N \geq 0$, can be estimated given the sequence of observations Z_N^* , where

$$Z_N^* := \{\mathbf{x}_0, \mathbf{z}(0), \mathbf{z}(1), \dots, \mathbf{z}(N)\}. \quad (\text{C.4})$$

Under the stated assumptions of independence and normal distributions, the joint conditional probability density function of Z_N^* given $\mathbf{x}(N)$ can be written, from (C.2) and (C.3), as

$$p_N := p_{Z_N^*|\mathbf{x}(N)} = \frac{1}{A} e^{-\frac{1}{2}\Psi[Z_N^*|\mathbf{x}(N)]}, \quad (\text{C.5})$$

where¹

$$A := (2\pi)^{n/2} |\mathbf{P}_0|^{1/2} \prod_{k=0}^N [(2\pi)^{m/2} |\mathbf{R}(k)|^{1/2}], \quad (\text{C.6})$$

$$\begin{aligned} \Psi[Z_N^*; \mathbf{x}(N)] &:= (\mathbf{x}_0 - \bar{\mathbf{x}}_0)^T \mathbf{P}_0^{-1} (\mathbf{x}_0 - \bar{\mathbf{x}}_0) \\ &\quad + \sum_{k=0}^N [\mathbf{z}(k) - \bar{\mathbf{z}}(k)]^T \mathbf{R}^{-1}(k) [\mathbf{z}(k) - \bar{\mathbf{z}}(k)], \end{aligned} \quad (\text{C.7})$$

and

$$\bar{\mathbf{x}}_0 := \mathcal{E}\{\mathbf{x}_0\} = \mathbf{x}(0), \text{ and} \quad (\text{C.8})$$

$$\bar{\mathbf{z}}(k) := \mathcal{E}\{\mathbf{z}(k)\} = \mathbf{h}[\mathbf{x}(k)]. \quad (\text{C.9})$$

The Cramer-Rao lower bound for the error covariance matrix $\mathbf{S}(N)$ of an unbiased estimator $\hat{\mathbf{x}}(N)$ of the state $\mathbf{x}(N)$, where the estimate is based on the observation sequence Z_N^* , is given by

$$\mathbf{S}(N) := \mathcal{E}\{[\hat{\mathbf{x}}(N) - \mathbf{x}(N)][\hat{\mathbf{x}}(N) - \mathbf{x}(N)]^T\} \geq \mathbf{J}^{-1}(N), \quad (\text{C.10})$$

where $\mathbf{J}(N)$ is Fisher's information matrix. The matrix inequality $\mathbf{S} \geq \mathbf{J}^{-1}$ is equivalent to stating that $(\mathbf{S} - \mathbf{J}^{-1})$ is positive semi-definite. In particular, since the diagonal elements of a positive semi-definite matrix are non-negative, the diagonal elements of $\mathbf{J}^{-1}(N)$ provide the estimation error variance lower bounds for the corresponding elements of $\hat{\mathbf{x}}(N)$. In the linear system and measurement case, (B.1) and (B.3), the discrete Kalman filter given by (B.12) and (B.13) provides the minimum variance unbiased estimate. If the system noise $\mathbf{w}(k)$ in (B.1) is zero, then for the linear case

$$\mathbf{S}(N) = \mathbf{P}(N|N) = \mathbf{J}^{-1}(N). \quad (\text{C.11})$$

In the nonlinear EKF formulation, $\mathbf{S}(N)$ and $\hat{\mathbf{P}}(N|N)$ are not, in general, equivalent.

Fisher's information measure on $\mathbf{x}(N)$ contained in Z_N^* is given by two equivalent representations [68, pp. 91-93]:

$$\mathbf{J}(N) = \mathcal{E} \left\{ \left(\frac{\partial \ln p_N}{\partial \mathbf{x}(N)} \right)^T \left(\frac{\partial \ln p_N}{\partial \mathbf{x}(N)} \right) \middle| \mathbf{x}(N) \right\}, \text{ or} \quad (\text{C.12})$$

$$\mathbf{J}(N) = -\mathcal{E} \left\{ \left(\frac{\partial^2 \ln p_N}{\partial \mathbf{x}(N)^2} \right) \middle| \mathbf{x}(N) \right\}. \quad (\text{C.13})$$

¹The notation $|\mathbf{P}|$ denotes the determinant of $\mathbf{P}$.

where the gradient $\partial \ln p_N / \partial \mathbf{x}(N)$ is a row vector. The inequality (C.10) with $\mathbf{J}(N)$ expressed as in (C.12) is termed the "Cramer-Rao inequality."

Taking the logarithm of p_N in (C.5) yields

$$\ln p_N = -\ln A - \frac{1}{2} \Psi[Z_N^*; \mathbf{x}(N)]. \quad (\text{C.14})$$

The gradient of (C.14) is given by the row vector

$$\begin{aligned} \frac{\partial \ln p_N}{\partial \mathbf{x}(N)} &= (\mathbf{x}_0 - \bar{\mathbf{x}}_0)^T \mathbf{P}_0^{-1} \frac{\partial \bar{\mathbf{x}}_0}{\partial \mathbf{x}(N)} \\ &\quad + \sum_{k=0}^N [\mathbf{z}(k) - \bar{\mathbf{z}}(k)]^T \mathbf{R}^{-1}(k) \frac{\partial \bar{\mathbf{z}}(k)}{\partial \mathbf{x}(N)}. \end{aligned} \quad (\text{C.15})$$

To evaluate (C.12), consider the four terms of the outer product of (C.15):

$$\begin{aligned} \left(\frac{\partial p_N}{\partial \mathbf{x}(N)} \right)^T \left(\frac{\partial p_N}{\partial \mathbf{x}(N)} \right) &= \\ &\left(\frac{\partial \bar{\mathbf{x}}_0}{\partial \mathbf{x}(N)} \right)^T \mathbf{P}_0^{-1} (\mathbf{x}_0 - \bar{\mathbf{x}}_0) (\mathbf{x}_0 - \bar{\mathbf{x}}_0)^T \mathbf{P}_0^{-1} \left(\frac{\partial \bar{\mathbf{x}}_0}{\partial \mathbf{x}(N)} \right) \\ &+ \sum_{k=0}^N \left(\frac{\partial \bar{\mathbf{x}}_0}{\partial \mathbf{x}(N)} \right)^T \mathbf{P}_0^{-1} (\mathbf{x}_0 - \bar{\mathbf{x}}_0) [\mathbf{z}(k) - \bar{\mathbf{z}}(k)]^T \mathbf{R}^{-1}(k) \left(\frac{\partial \bar{\mathbf{z}}(k)}{\partial \mathbf{x}(N)} \right) \\ &+ \sum_{k=0}^N \left(\frac{\partial \bar{\mathbf{z}}(k)}{\partial \mathbf{x}(N)} \right)^T \mathbf{R}^{-1}(k) [\mathbf{z}(k) - \bar{\mathbf{z}}(k)] (\mathbf{x}_0 - \bar{\mathbf{x}}_0)^T \mathbf{P}_0^{-1} \left(\frac{\partial \bar{\mathbf{x}}_0}{\partial \mathbf{x}(N)} \right) \\ &+ \left\{ \sum_{k=0}^N \left(\frac{\partial \bar{\mathbf{z}}(k)}{\partial \mathbf{x}(N)} \right)^T \mathbf{R}^{-1}(k) [\mathbf{z}(k) - \bar{\mathbf{z}}(k)] \right\} \\ &\quad \cdot \left\{ \sum_{j=0}^N [\mathbf{z}(j) - \bar{\mathbf{z}}(j)]^T \mathbf{R}^{-1}(j) \left(\frac{\partial \bar{\mathbf{z}}(j)}{\partial \mathbf{x}(N)} \right) \right\}. \end{aligned} \quad (\text{C.16})$$

The conditional expectation in (C.12) can be computed from (C.16) with the following observations:

1. The Jacobian matrices $\partial \bar{\mathbf{x}}_0 / \partial \mathbf{x}(N)$ and $\partial \bar{\mathbf{z}}(k) / \partial \mathbf{x}(N)$ are deterministic;
2. From (C.2), (C.8), and (C.9), the random variables $(\mathbf{x}_0 - \bar{\mathbf{x}}_0)$ and $[\mathbf{z}(k) - \bar{\mathbf{z}}(k)]$, $k = 0, 1, \dots, N$, are independent with zero mean and hence the second and third terms of (C.16) do not contribute to the expectation and the product of summations in the last term of (C.16) reduces to a single summation of products with $k = j$;

3. By definitions (C.2) and (C.3),

$$\begin{aligned}\mathcal{E} \left\{ (\mathbf{x}_0 - \bar{\mathbf{x}}_0)(\mathbf{x}_0 - \bar{\mathbf{x}}_0)^T \right\} &= \mathbf{P}_0 \text{ and} \\ \mathcal{E} \left\{ [\mathbf{z}(k) - \bar{\mathbf{z}}(k)][\mathbf{z}(k) - \bar{\mathbf{z}}(k)]^T \right\} &= \mathbf{R}(k).\end{aligned}$$

Furthermore, from the definitions given in (C.8) and (C.9),

$$\frac{\partial \bar{\mathbf{x}}_0}{\partial \mathbf{x}(N)} = \frac{\partial \mathbf{x}(0)}{\partial \mathbf{x}(N)} \quad (\text{C.17})$$

and

$$\begin{aligned}\frac{\partial \bar{\mathbf{z}}(k)}{\partial \mathbf{x}(N)} &= \frac{\partial \mathbf{h}[\mathbf{x}(k)]}{\partial \mathbf{x}(N)} \\ &= \mathbf{H}_t(k) \frac{\partial \mathbf{x}(k)}{\partial \mathbf{x}(N)},\end{aligned} \quad (\text{C.18})$$

where

$$\mathbf{H}_t(k) := \left. \frac{\partial \mathbf{h}[\mathbf{x}]}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}(k)}. \quad (\text{C.19})$$

The subscript "t" on $\mathbf{H}_t$ is used to emphasize that (C.19) is evaluated along the true trajectory $\mathbf{x}(k)$ rather than along the estimated trajectory $\hat{\mathbf{x}}(k)$ as in the usual EKF realization.

Performing the expectation of (C.16) in accordance with (C.12) and substituting (C.17), (C.18), and (C.19) yields Fisher's information matrix:

$$\begin{aligned}\mathbf{J}(N) &= \left(\frac{\partial \mathbf{x}(0)}{\partial \mathbf{x}(N)} \right)^T \mathbf{P}_0^{-1} \left(\frac{\partial \mathbf{x}(0)}{\partial \mathbf{x}(N)} \right) \\ &\quad + \sum_{k=0}^N \left(\frac{\partial \mathbf{x}(k)}{\partial \mathbf{x}(N)} \right)^T \mathbf{H}_t^T(k) \mathbf{R}^{-1}(k) \mathbf{H}_t(k) \left(\frac{\partial \mathbf{x}(k)}{\partial \mathbf{x}(N)} \right).\end{aligned} \quad (\text{C.20})$$

Evaluation of the alternate representation for Fisher's information matrix in (C.13) is most easily accomplished through component-wise differentiation of (C.14). Consider a vector $\xi \in R^n$ and a scalar-valued function $\mathcal{F}(\xi)$ given by

$$\mathcal{F}(\xi) = \frac{1}{2} \xi^T \Gamma \xi. \quad (\text{C.21})$$

where $\Gamma \in R^{n \times n}$ is positive definite symmetric. Equation (C.21) represents the general form of $\ln p_N$ in (C.14) which must be differentiated. Provided the components of ξ admit

second derivatives with respect to components of a vector $\vartheta \in R^n$,

$$\begin{aligned} \left[\frac{\partial^2 \mathcal{F}(\xi)}{\partial \vartheta^2} \right]_{i,j} &= \frac{\partial^2 \mathcal{F}(\xi)}{\partial \vartheta_i \partial \vartheta_j} \\ &= \sum_{\ell,k} \left[\frac{\partial \xi_\ell}{\partial \vartheta_i} \Gamma_{\ell,k} \frac{\partial \xi_k}{\partial \vartheta_j} + \xi_\ell \Gamma_{\ell,k} \frac{\partial^2 \xi_k}{\partial \vartheta_i \partial \vartheta_j} \right] \\ &= \left[\left(\frac{\partial \xi}{\partial \vartheta} \right)^T \Gamma \left(\frac{\partial \xi}{\partial \vartheta} \right) \right]_{i,j} + \sum_{\ell,k} \left[\xi_\ell \Gamma_{\ell,k} \frac{\partial^2 \xi_k}{\partial \vartheta_i \partial \vartheta_j} \right]. \end{aligned} \quad (\text{C.22})$$

Taking the expectation of (C.22) in accordance with (C.13) in a sequence of steps with $\vartheta = \mathbf{x}(N)$ and first with $\xi = (\mathbf{x}_0 - \bar{\mathbf{x}}_0)$ and then with $\xi = [\mathbf{z}(k) - \bar{\mathbf{z}}(k)]$, $k = 0, 1, 2, \dots, N$, which implies that in all steps ξ has zero mean and deterministic first and second partial derivatives, eliminates the second term on the right of (C.22) giving, with (C.17), (C.18), and (C.19), the previous result (C.20) as expected.

Finally, from (C.1)

$$\frac{\partial \mathbf{x}(k+1)}{\partial \mathbf{x}(N)} = \Phi_t(k) \frac{\partial \mathbf{x}(k)}{\partial \mathbf{x}(N)}, \quad \text{for } 0 \leq k \leq N-1, \quad (\text{C.23})$$

where

$$\Phi_t(k) := \left. \frac{\partial \mathbf{f}[\mathbf{x}]}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}(k)}, \quad \text{for } 0 \leq k \leq N-1 \quad (\text{C.24})$$

is evaluated along the true trajectory as shown. In particular, with $\mathbf{I}_n$ representing the $n \times n$ identity matrix,

$$\frac{\partial \mathbf{x}(N)}{\partial \mathbf{x}(N)} = \mathbf{I}_n = \left[\downarrow \prod_{i=N-1}^k \Phi_t(i) \right] \frac{\partial \mathbf{x}(k)}{\partial \mathbf{x}(N)}, \quad (\text{C.25})$$

where the notation " $\downarrow \prod$ " indicates that the index i decreases from left to right in the matrix product. The Jacobian matrices in (C.20) are then computed as

$$\frac{\partial \mathbf{x}(k)}{\partial \mathbf{x}(N)} = \prod_{i=k}^{N-1} \Phi_t^{-1}(i). \quad (\text{C.26})$$

Substitution of (C.26) into (C.20) gives a closed-form expression for Fisher's information matrix for the system and measurement models defined in (C.1) and (C.2):

$$\begin{aligned} \mathbf{J}(N) &= \left(\prod_{i=0}^{N-1} \Phi_t^{-1}(i) \right)^T \mathbf{P}_0^{-1} \left(\prod_{i=0}^{N-1} \Phi_t^{-1}(i) \right) \\ &\quad + \sum_{k=0}^N \left(\prod_{i=k}^{N-1} \Phi_t^{-1}(i) \right)^T \mathbf{H}_t^T(k) \mathbf{R}^{-1}(k) \mathbf{H}_t(k) \left(\prod_{i=k}^{N-1} \Phi_t^{-1}(i) \right). \end{aligned} \quad (\text{C.27})$$

A more useful form for practical computation is a recursive formulation which follows from (C.27) by inspection:

1. Initialization –

$$\mathbf{J}(0) = \mathbf{P}_0^{-1} + \mathbf{H}_t^T(0)\mathbf{R}^{-1}(0)\mathbf{H}_t(0); \quad (\text{C.28})$$

2. Recursive evaluation for $k \geq 0$ –

$$\mathbf{J}(k+1) = [\Phi_t^{-1}(k)]^T \mathbf{J}(k) [\Phi_t^{-1}(k)] + \mathbf{H}_t^T(k)\mathbf{R}^{-1}(k)\mathbf{H}_t(k). \quad (\text{C.29})$$

If the information matrix $\mathbf{J}(k)$ remains singular in the absence of prior information, i.e. $\mathbf{P}_0^{-1} = 0$, then the system has an unobservable subspace. The converse is not generally true, that is the bounds may decrease even though the system is unobservable. Observability of nonlinear systems, although often difficult to establish, can be investigated locally through methods from differential geometry [73].

The above derivation has been given for discrete-time, nonlinear system and measurement models. Furthermore, this derivation assumes an unbiased estimation procedure. A more general result for an estimator with state dependent bias $\mathbf{b}(\mathbf{x})$ is given by [13]

$$\mathbf{S}(N) \geq \left(\mathbf{I}_n + \frac{\partial \mathbf{b}(\mathbf{x})}{\partial \mathbf{x}} \right) \mathbf{J}^{-1} \left(\mathbf{I}_n + \frac{\partial \mathbf{b}(\mathbf{x})}{\partial \mathbf{x}} \right)^T, \quad (\text{C.30})$$

where $\mathbf{S}(N)$ is defined in (C.10) and the Jacobian of $\mathbf{b}$ is evaluated along the true trajectory. However, the bias $\mathbf{b}(\mathbf{x})$ is generally not available in analytical form. Unfortunately, a state dependent bias can result in higher or lower bounds in comparison to $\mathbf{J}^{-1}$: for example, consider $\mathbf{b}(\mathbf{x}) = -\mathbf{x}/2$ which gives $\mathbf{S} \geq \mathbf{J}^{-1}/4$, while $\mathbf{b}(\mathbf{x}) = \mathbf{x}$ gives $\mathbf{S} \geq 4\mathbf{J}^{-1}$. As a result, most analysis proceed under the assumption of unbiased estimation resulting in approximate covariance bounds.

Broida and Chellappa [48, 49, 51, 53], Young and Chellappa [56], Taylor [69], and Chang [70] all employ Cramer-Rao analysis for filter performance evaluation. Broida and Chellappa, and Young and Chellappa do not consider *a priori* information. Taylor does consider information provided by the initial state estimate and treats the continuous-time system case. In the continuous-discrete filtering techniques of Section B.3, results in [69] indicate that approximate covariance bounds can be computed as in (C.28) and (C.29), but with $\Phi_t(k)$ replaced with $\hat{\Phi}_t(k, k-1)$ defined in (B.34).

The Cramer Rao bounds shown in Section 2.6.2 do not decrease monotonically. This seems to contradict the notion that as more measurements become available, more and more information should be accumulated about the observed process and, as a result, the CRLB's would be expected to decrease. Fisher's information matrix defined in (C.12) provides a measure of information contained in Z_N^* concerning the state $\mathbf{x}(N)$. As time, N , increases, the measurement set Z_N^* accumulates more measurements but, in addition, the state $\mathbf{x}(N)$ may vary with time. As discussed by Young and Chellappa [56], the noise-free dynamical model implies a one-to-one deterministic relationship between $\mathbf{x}(0)$ and $\mathbf{x}(N)$ so that

$$p_{Z_N^*|\mathbf{x}(N)} = p_{Z_N^*|\mathbf{x}(0)}, \quad (\text{C.31})$$

and

$$\frac{\partial p_{Z_N^*|\mathbf{x}(N)}}{\partial \mathbf{x}(N)} = \left[\frac{\partial p_{Z_N^*|\mathbf{x}(0)}}{\partial \mathbf{x}(0)} \right] \frac{\partial \mathbf{x}(0)}{\partial \mathbf{x}(N)}. \quad (\text{C.32})$$

Inversion of Fisher's information matrix then gives

$$\mathbf{J}^{-1}(N) = \left(\frac{\partial \mathbf{x}(0)}{\partial \mathbf{x}(N)} \right)^{-1} \mathbf{K}^{-1}(N) \left(\frac{\partial \mathbf{x}(0)}{\partial \mathbf{x}(N)} \right)^{-T}, \quad (\text{C.33})$$

where $(\cdot)^{-T}$ denotes the inverse transpose and

$$\begin{aligned} \mathbf{K}^{-1}(N) &:= \left[\mathcal{E} \left\{ \left(\frac{\partial \ln p_{Z_N^*|\mathbf{x}(0)}}{\partial \mathbf{x}(N)} \right)^T \left(\frac{\partial \ln p_{Z_N^*|\mathbf{x}(0)}}{\partial \mathbf{x}(N)} \right) \middle| \mathbf{x}(N) \right\} \right]^{-1} \\ &= \left[\mathbf{P}_0^{-1} + \sum_{n=0}^N \left(\frac{\partial \mathbf{x}(0)}{\partial \mathbf{x}(N)} \right)^T \mathbf{H}_t^T(n) \mathbf{R}^{-1}(n) \mathbf{H}_t(n) \left(\frac{\partial \mathbf{x}(0)}{\partial \mathbf{x}(N)} \right) \right]^{-1}. \end{aligned} \quad (\text{C.34})$$

Cramer Rao bounds for estimation of the initial state $\mathbf{x}(0)$ from the measurement set Z_N^* are the diagonal elements of $\mathbf{K}^{-1}(N)$, and these decrease monotonically. The bounds on $\mathbf{x}(N)$, however, depend on the time-varying congruent transformation of $\mathbf{K}^{-1}(N)$ in (C.34) and need not decrease monotonically.

Appendix D

Jacobian of Measurement Models

Implementation of the extended Kalman filter and estimation of Cramer Rao bounds requires the computation of $\mathbf{r}_{C_j}^t$, $\mathbf{I}_E^Q(\boldsymbol{\zeta})$ and $\mathbf{h}[\mathbf{x}]$, which are all components of the measurement model, as well as the Jacobian of $\mathbf{h}[\mathbf{x}]$,

$$\mathbf{H}[\mathbf{x}] := \frac{\partial \mathbf{h}[\mathbf{x}]}{\partial \mathbf{x}}, \quad (\text{D.1})$$

from the most recent state estimate. Computation of $\mathbf{H}[\mathbf{x}]$ for the three parameterizations of rotational motion is a tedious exercise in differentiation, but, with algebraic manipulation and simplification, can be written in simple form which admits computationally efficient implementation.

The Jacobian matrix, $\mathbf{H} \in R^{2 \times N_c \times N}$, can be written as

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \vdots \\ \mathbf{H}_{N_c} \end{bmatrix} \quad (\text{D.2})$$

where $\mathbf{H}_j$ is the approximate linear measurement matrix corresponding to the j th camera. Each $\mathbf{H}_j$ can be further partitioned according to the feature point of interest as

$$\mathbf{H}_j = \begin{bmatrix} \mathbf{H}_j^1 \\ \mathbf{H}_j^2 \\ \vdots \\ \mathbf{H}_j^{N_f} \end{bmatrix}. \quad (\text{D.3})$$

$$= \begin{bmatrix} \mathbf{P}_j^1 & \mathbf{0} & \mathbf{W}_j^1 & \mathbf{0} & \mathbf{S}_j^1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{P}_j^2 & \mathbf{0} & \mathbf{W}_j^2 & \mathbf{0} & \mathbf{0} & \mathbf{S}_j^2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_j^{N_f} & \mathbf{0} & \mathbf{W}_j^{N_f} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{S}_j^{N_f} \end{bmatrix}, \quad (\text{D.4})$$

where each $\mathbf{H}_j^i \in \mathbb{R}^{2 \times N}$, $i = 1, 2, \dots, N_f$, $j = 1, 2, \dots, N_C$ is defined as

$$\mathbf{H}_j^i := \frac{\partial \mathbf{h}_j^i[\mathbf{x}]}{\partial \mathbf{x}}. \quad (\text{D.5})$$

The submatrices $\mathbf{P}_j^i \in \mathbb{R}^{2 \times 3}$, $\mathbf{W}_j^i \in \mathbb{R}^{2 \times (3 \text{ or } 4)}$, depending on the parameterization of rotational motion, and $\mathbf{S}_j^i \in \mathbb{R}^{2 \times 3}$ can be written, after some algebraic manipulation, as follows:

$$\begin{aligned} \mathbf{P}_j^i &:= \frac{\partial \mathbf{h}_j^i[\mathbf{x}]}{\partial \mathbf{p}} \\ &= \frac{L_j}{(r_{C_j,1}^i)^2} \begin{bmatrix} -(\mathbf{r}_{C_j}^i)^T \mathbf{k}_E^* \\ (\mathbf{r}_{C_j}^i)^T \mathbf{j}_E^* \end{bmatrix} \mathbf{I}_{C_j}^E, \end{aligned} \quad (\text{D.6})$$

$$\begin{aligned} \mathbf{W}_j^i &:= \frac{\partial \mathbf{h}_j^i[\mathbf{x}]}{\partial \boldsymbol{\zeta}} \\ &= \mathbf{P}_j^i \frac{\partial \mathbf{I}_E^O(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}} \mathbf{r}_{O}^i. \end{aligned} \quad (\text{D.7})$$

$$\begin{aligned} \mathbf{S}_j^i &:= \frac{\partial \mathbf{h}_j^i[\mathbf{x}]}{\partial \mathbf{r}_O^i} \\ &= \mathbf{P}_j^i \mathbf{I}_E^O. \end{aligned} \quad (\text{D.8})$$

All entries of $\mathbf{H}$ other than those specified above are zero. In the above equations, $\mathbf{j}_E = [0, 1, 0]^T$, $\mathbf{k}_E = [0, 0, 1]^T$, and $\mathbf{j}_E^*$ and $\mathbf{k}_E^*$ are the matrix cross product operators associated with $\mathbf{j}_E$ and $\mathbf{k}_E$, respectively. In Equation (D.7), the notation $\partial \mathbf{I}_E^O / \partial \boldsymbol{\zeta}$ denotes a three-dimensional entity which premultiplies a vector $\boldsymbol{\vartheta} \in \mathbb{R}^3$ to yield a matrix according to the rule

$$\left[\frac{\partial \mathbf{I}_E^O(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}} \boldsymbol{\vartheta} \right]_{ij} = \sum_{k=1}^3 \frac{\partial [\mathbf{I}_E^O]_{ik}}{\partial \zeta_j} \vartheta_k. \quad (\text{D.9})$$

The general form of the measurement model is common to all parameterizations of rotational motion, except that different parameterizations will yield different expressions for $\partial \mathbf{I}_E^O(\boldsymbol{\zeta}) / \partial \boldsymbol{\zeta}$. These expressions are now stated for each parameterization.

Angle-Axis Parameterization: The measurement model Jacobian can be computed with

$$\frac{\partial \mathbf{I}_E^O(\boldsymbol{\xi})}{\partial \xi_j} = \frac{\{1 - \cos(\gamma)\}}{\gamma} \mathbf{B}_j + \cos(\gamma) \bar{\xi}_j \bar{\boldsymbol{\xi}}^\# + \frac{\sin(\gamma)}{\gamma} \Gamma_j + \sin(\gamma) \bar{\xi}_j (\bar{\boldsymbol{\xi}}^\#)^2, \quad (\text{D.10})$$

where

$$\gamma := \|\boldsymbol{\xi}\|, \quad (\text{D.11})$$

$$\bar{\boldsymbol{\xi}} := \frac{\boldsymbol{\xi}}{\|\boldsymbol{\xi}\|}. \quad (\text{D.12})$$

$$\Gamma_j := \mathbf{e}_j^\# - \bar{\xi}_j \bar{\boldsymbol{\xi}}^\#, \quad (\text{D.13})$$

$$\mathbf{B}_j := \Gamma_j \bar{\boldsymbol{\xi}}^\# + \bar{\boldsymbol{\xi}}^\# \Gamma_j. \quad (\text{D.14})$$

and $\bar{\xi}_j$ denotes the j th component of the unit vector $\bar{\boldsymbol{\xi}}$, and $\mathbf{e}_j$ denotes the j th standard basis vector of $\mathbb{R}^3$. For very small γ , it is apparent from (D.10)-(D.14) that

$$\mathbf{I}_E^O(\boldsymbol{\xi}) \sim \mathbf{I}_3 + \boldsymbol{\xi}^\#, \text{ and} \quad (\text{D.15})$$

$$\frac{\partial \mathbf{I}_E^O(\boldsymbol{\xi})}{\partial \xi_j} \sim \frac{\gamma}{2} \mathbf{B} + \mathbf{e}_j^\#. \quad (\text{D.16})$$

This information can be used to eliminate numerical difficulties encountered in applications when (D.10) must be computed with $\gamma = \|\boldsymbol{\xi}\| \sim 0$.

Quaternion Parameterization: The Jacobian of the measurement model is computed with

$$\frac{\partial \mathbf{I}_E^O(\mathbf{q})}{\partial q_1} = 2 \begin{bmatrix} q_1 & q_2 & q_3 \\ q_2 & -q_1 & q_4 \\ q_3 & -q_4 & -q_1 \end{bmatrix}. \quad (\text{D.17})$$

$$\frac{\partial \mathbf{I}_E^O(\mathbf{q})}{\partial q_2} = 2 \begin{bmatrix} -q_2 & q_1 & -q_4 \\ q_1 & q_2 & q_3 \\ q_4 & q_3 & -q_2 \end{bmatrix}. \quad (\text{D.18})$$

$$\frac{\partial \mathbf{I}_E^O(\mathbf{q})}{\partial q_3} = 2 \begin{bmatrix} -q_3 & q_4 & q_1 \\ -q_4 & -q_3 & q_2 \\ q_1 & q_2 & q_3 \end{bmatrix}, \text{ and} \quad (\text{D.19})$$

$$\frac{\partial \mathbf{I}_E^O(\mathbf{q})}{\partial q_4} = 2 \begin{bmatrix} q_4 & q_3 & -q_2 \\ -q_3 & q_4 & q_1 \\ q_2 & -q_1 & q_4 \end{bmatrix}. \quad (\text{D.20})$$

Roll-Pitch-Yaw Parameterization: The measurement model Jacobian is computed with

$$\frac{\partial \mathbf{I}_E^O(\zeta)}{\partial \phi} = \mathbf{I}_E^O \mathbf{e}_1^\#, \quad (\text{D.21})$$

$$\frac{\partial \mathbf{I}_E^O(\zeta)}{\partial \theta} = \exp(\psi \mathbf{e}_3^\#) \mathbf{e}_2^\# \exp(\theta \mathbf{e}_2^\#) \exp(\phi \mathbf{e}_1^\#), \text{ and} \quad (\text{D.22})$$

$$\frac{\partial \mathbf{I}_E^O(\zeta)}{\partial \psi} = \mathbf{e}_3^\# \mathbf{I}_E^O, \quad (\text{D.23})$$

where, again, the $\mathbf{e}_j$, $j = 1, 2, 3$, are the standard basis vectors of $\mathbb{R}^3$.

Appendix E

Iterated Nonlinear Observer Implementations

This Appendix provides example MATLAB script implementations of the nonredundant and redundant sliding mode observers proposed in Chapter 7 for the problem of estimating translational and rotational motion and structure of an object from sets of 3D measurements of positions of four feature points on an object.

Simulation results of Section 7.5 for the nonredundant iterated sliding mode observer were obtained with the following MATLAB script.

```
***** Function: nrsmo *****
function [xk,xk1,Ik,wk,sk] = nrsmo(T,z,x,do,s);
%
%Function [xk,xk1,Ik,wk,sk] = nrsmo(T,z,x,do,s)
% Purpose: One measurement update of nonredundant sliding mode
%           observer with local iterations, plus time propagation
%           to next measurement event.
% Input Parameters:
%   T ~ sample period
%   z ~ 3D measurements
%   x ~ time propagated state estimate
%   do ~ posn of riO wrt Co in Fo
%   s ~ value of switching function
% Output Parameters:
%   xk - state estimate after measurement
%   xk1 - time propagated state estimate
```

```

%      Ik - estimated rotation matrix Ioe after measurement
%      wk - angular velocity estimate after measurement
%      sk - value of switching function after measurement update
%
%*****
%
% Constant definitions
twopi = 2*pi;
e1_sh = [0,0,0;0,0,-1;0,1,0]; % Cross-product matrices
e2_sh = [0,0,1;0,0,0;-1,0,0];
e3_sh = [0,-1,0;1,0,0;0,0,0];
%
% Parameters
%
num_pts = 4;          % Number of feature points
max_it = 10;          % Max number of local iterations
threshold = 0.001; % Threshold level for local iterations
%
% Define motion models
%
% x = [x,y,z,xi_x,xi_y,xi_z,r^4,r^3(xy),r^2(x)]';
% x_{e} = [x',dx,dy,dz,dxi_x,dxi_y,dxi_z,...
%          d2x,d2y,d2z,d2xi_x,d2xi_y,d2xi_z];
%
alf_cor_t = 1/2; % Reciprocal manoeuvre correlation time const
alf_cor_r = 1/2;
%
A = [zeros(3,12),eye(3),zeros(3,9);
     zeros(3,15),eye(3),zeros(3,6);
     zeros(6,24);
     zeros(3,18),eye(3),zeros(3,3);
     zeros(3,21),eye(3);
     zeros(3,18),-alf_cor_t*eye(3),zeros(3,3);

```



```

        zeros(3,21),-alf_cor_r*eye(3)];
%
Phi = expm(A*T);    % State transition matrix
C = [eye(12),zeros(12,12)];
%
% Reaching Law Parameters
%
Gamma = -0.005*eye(12);
Lambda=  0.025*eye(12);
sat_eps=10*diag(abs(Gamma))./(ones(12,1)-diag(abs(Lambda)));
%
% Observer Gain Parameters
%
sigma=  [1;8;8;25;25];
Chi = [sigma(1)*eye(12);
       sigma(2)*eye(3),zeros(3,9);
       zeros(3,3),sigma(3)*eye(3),zeros(3,6);
       sigma(4)*eye(3),zeros(3,9);
       zeros(3,3),sigma(5)*eye(3),zeros(3,6)];
%
% Start Observer Equations
%
xk = x;                % Copy received time propagated state estimate
%
gs = Gamma*sat(s,sat_eps) + Lambda*s; %reaching law update
%
xm = C*xk;             % Get measured state
                        % Collect estimated structure
ro = [xm(7:9),[xm(10:11);do(3)],[xm(12);do(2:3)],do];
%
xi = xm(4:6);          % Compute rotation matrix Ioe
Ioe = exp2rot(xi);
                        % Rotate and translate structure from Fo to Fe

```

```

re = [Ioe,xm(1:3)]*[ro;ones(1,num_pts)];
%
% Compute h(xm)
%
h = [re(:,1);re(:,2);re(:,3);re(:,4)];
%
% Initialize for local iterations
%
sk = z-h;
rk = gs - sk;
%
% Start local iterations
%
i=0;
while (i<=max_it), % Must pass into loop at least once
%
%   Compute Jacobian of measurement model
%
    alf = norm(xi);
    xi_bar = xi/alf;
    xi_bar_sh = sharp(xi_bar);

    % Compute dIoe/d(xi_1)
    gamma = (e1_sh - xi_bar_sh*xi_bar(1))/alf;
    beta = gamma*xi_bar_sh + xi_bar_sh*gamma;
    dI_dx1 = beta + cos(alf)*(xi_bar(1)*xi_bar_sh - beta)...
        + sin(alf)*(gamma + xi_bar(1)*xi_bar_sh*xi_bar_sh);

    % Compute dIoe/d(xi_2)
    gamma = (e2_sh - xi_bar_sh*xi_bar(2))/alf;
    beta = gamma*xi_bar_sh + xi_bar_sh*gamma;
    dI_dx2 = beta + cos(alf)*(xi_bar(2)*xi_bar_sh - beta)...
        + sin(alf)*(gamma + xi_bar(2)*xi_bar_sh*xi_bar_sh);

```

```

% Compute dIoe/d(xi_3)
gamma = (e3_sh - xi_bar_sh*xi_bar(3))/alf;
beta = gamma*xi_bar_sh + xi_bar_sh*gamma;
dI_dxi3 = beta + cos(alf)*(xi_bar(3)*xi_bar_sh - beta)...
          + sin(alf)*(gamma + xi_bar(3)*xi_bar_sh*xi_bar_sh);

% u~unknown, xy~xy unknown, x~x unknown, do~wrt do measurement
dhdxi_u = [dI_dxi1*ro(:,1),dI_dxi2*ro(:,1),...
           dI_dxi3*ro(:,1)];
dhdxi_xy = [dI_dxi1*ro(:,2),dI_dxi2*ro(:,2),...
            dI_dxi3*ro(:,2)];
dhdxi_x = [dI_dxi1*ro(:,3),dI_dxi2*ro(:,3),...
            dI_dxi3*ro(:,3)];
dhdxi_do = [dI_dxi1*ro(:,4),dI_dxi2*ro(:,4),...
            dI_dxi3*ro(:,4)];

% H is the desired Jacobian
H = [eye(3),dhdxi_u,Ioe,zeros(3,3);
     eye(3),dhdxi_xy,zeros(3,3),Ioe(:,1:2),zeros(3,1);
     eye(3),dhdxi_x,zeros(3,5),Ioe(:,1);
     eye(3),dhdxi_do,zeros(3,6)];

%Compute gains
L = Chi*inv(H);

% Perform one local iteration
xk = xk - L*rk;

% Compute resulting sk and rk to check for early termination
% ...this must be done anyway to return updated values

xm = C*xk;          % Get new measured state estimate

```

```

        % Collect estimated structure
        ro = [xm(7:9), [xm(10:11);do(3)], [xm(12);do(2:3)],do];

        xi = xm(4:6);    % Compute rotation matrix Ioe
        Ioe = exp2rot(xi);

        % Rotate and translate structure from Fo to Fe
        re = [Ioe,xm(1:3)]*[ro;ones(1,num_pts)];

        % Compute h(xr)
        h = [re(:,1);re(:,2);re(:,3);re(:,4)];
        sk = z-h;        % New value of switching function
        rk = gs - sk;    % New value of rk

        % Check for loop termination
        if (norm(rk) < threshold),
            break;        % Break out of While loop
        end;

        i = i+1;        % Increment loop counter
    end; %While not_done ...local iterations
%
% Now return required parameters and perform time propagation
%
% Compute estimated rotation matrix Ioe
%
Ik = exp2rot(xk(4:6));
%
% Compute estimated angular velocity
%
xi = xk(4:6);
alf = norm(xi);
xi_bar = xi/alf;
xi_dot = xk(16:18);
alf_dot = xi_bar'*xi_dot;
xi_bar_dot = (1/alf)*(eye(3)-xi_bar*xi_bar')*xi_dot;

```

```

wk = alf_dot*xi_bar+sin(alf)*xi_bar_dot+...
      (1-cos(alf))*sharp(xi_bar)*xi_bar_dot;

%
% Check for reset of angle for angle-axis parameterization
%
if (alf > pi)
    xk(4:6) = xi - twopi*xi_bar;
    xk(16:18) = xi_dot - twopi*xi_bar_dot;
    xi_ddot = xk(22:24);
    alf_ddot = xi_bar'*xi_ddot + xi_dot'*xi_dot/alf - ...
              (alf_dot/alf)^2;
    xi_bar_ddot = (eye(3)-xi_bar*xi_bar')*...
              (xi_ddot/alf-xi_dot*alf_dot/alf^2)-...
              (xi_bar_dot*xi_bar' + xi_bar*xi_bar_dot')*xi_dot/alf;
    xk(22:24) = xi_ddot - twopi*xi_bar_ddot;
end; %If
%
% Perform time propagation of state estimate
%
xk1 = Phi*xk;
%
%***** END FUNCTION nrsmo *****

```

5 of/de 5

PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

Simulation results of Section 7.5 for the redundant iterated sliding mode observer were obtained with the following MATLAB script.

```
***** Function: rsmo *****
function [xk,xk1,Ik,wk,sk] = rsmo(T,z,x,do,s);
%
%Function [xk,xk1,Ik,Ik1,wk,gK,gs,sk] = rsmo(T,z,x,do,s)
% Purpose: One measurement update of redundant sliding mode
%           observer with local iterations, plus time propagation
%           to next measurement event.
% Input Parameters:
%     T ~ sample period
%     z ~ 3D measurements
%     x ~ time propagated state estimate
%     do ~ posn of ri0 wrt Oo in Fo
%     s ~ value of switching function
% Output Parameters:
%     xk - state estimate after measurement
%     xk1 - time propagated state estimate
%     Ik - estimated rotation matrix Ioe after measurement
%     wk - angular velocity estimate after measurement
%     sk - value of switching function after measurement update
%*****
%
% Constant definitions
twopi = 2*pi;
e1_sh = [0,0,0;0,0,-1;0,1,0]; % Cross-product matrices
e2_sh = [0,0,1;0,0,0;-1,0,0];
e3_sh = [0,-1,0;1,0,0;0,0,0];
%
% Parameters
%
num_pts = 4;           % Number of feature points
```

```

max_it = 10;          % Max number of local iterations
threshold = 0.001; % Threshold level for local iterations
%
% Define motion models
%
% x = [x,y,z,xi_x,xi_y,xi_z,r^4,r^3(xy),r^2(x)]';
% x_{e} = [x',dx,dy,dz,dxi_x,dxi_y,dxi_z,...
%          d2x,d2y,d2z,d2xi_x,d2xi_y,d2xi_z];
%
alf_cor_t = 1/2;      % Reciprocal manoeuvre correlation time const
alf_cor_r = 1/2;
%
A = [zeros(3,12),eye(3),zeros(3,9);
     zeros(3,15),eye(3),zeros(3,6);
     zeros(6,24);
     zeros(3,18),eye(3),zeros(3,3);
     zeros(3,21),eye(3);
     zeros(3,18),-alf_cor_t*eye(3),zeros(3,3);
     zeros(3,21),-alf_cor_r*eye(3)];
%
Phi = expm(A*T);      % State transition matrix
C = [eye(12),zeros(12,12)];
%
% Reaching Law Parameters
%
Gamma = -0.005*eye(12);
Lambda= 0.025*eye(12);
sat_eps=10*diag(abs(Gamma))./(ones(12,1)-diag(abs(Lambda)));
P = eye(36);          % Measurement weighting matrix
%
% Observer Gain Parameters
%
sigma= [1;8;8;25;25];

```



```

Chi = [sigma(1)*eye(12);
       sigma(2)*eye(3),zeros(3,9);
       zeros(3,3),sigma(3)*eye(3),zeros(3,6);
       sigma(4)*eye(3),zeros(3,9);
       zeros(3,3),sigma(5)*eye(3),zeros(3,6)];

%
% Start Observer Equations
%
xk = x;           % Copy received time propagated state estimate
%
gs = Gamma*sat(s,sat_eps) + Lambda*s; %reaching law update
%
xm = C*xk;        % Get measured state
                   % Collect estimated structure
ro = [xm(7:9),[xm(10:11);do(3)],[xm(12);do(2:3)],do];
%
xi = xk(4:6);     % Compute rotation matrix Ioe
Ioe = exp2rot(xi);
                   % Rotate and translate structure from Fo to Fe
re = [Ioe,xm(1:3)]*[ro;ones(1,num_pts)];
%
% Compute hr(xm)
%
hi = [re(:,1);re(:,2);re(:,3);re(:,4)];
hr = [hi;hi;hi];
%
% Initialize for local iterations...must compute Jacobian
%
alf = norm(xi);
xi_bar = xi/alf;
xi_bar_sh = sharp(xi_bar);

% Compute dIoe/d(xi_1)

```

```

gamma = (e1_sh - xi_bar_sh*xi_bar(1))/alf;
beta = gamma*xi_bar_sh + xi_bar_sh*gamma;
dI_dxi1 = beta + cos(alf)*(xi_bar(1)*xi_bar_sh - beta)...
          + sin(alf)*(gamma + xi_bar(1)*xi_bar_sh*xi_bar_sh);

% Compute dIoe/d(xi_2)
gamma = (e2_sh - xi_bar_sh*xi_bar(2))/alf;
beta = gamma*xi_bar_sh + xi_bar_sh*gamma;
dI_dxi2 = beta + cos(alf)*(xi_bar(2)*xi_bar_sh - beta)...
          + sin(alf)*(gamma + xi_bar(2)*xi_bar_sh*xi_bar_sh);

% Compute dIoe/d(xi_3)
gamma = (e3_sh - xi_bar_sh*xi_bar(3))/alf;
beta = gamma*xi_bar_sh + xi_bar_sh*gamma;
dI_dxi3 = beta + cos(alf)*(xi_bar(3)*xi_bar_sh - beta)...
          + sin(alf)*(gamma + xi_bar(3)*xi_bar_sh*xi_bar_sh);

% u~unknown, xy~xy unknown, x~x unknown, do~wrt do measurement
dhdxi_u = [dI_dxi1*ro(:,1),dI_dxi2*ro(:,1),...
           dI_dxi3*ro(:,1)];
dhdxi_xy = [dI_dxi1*ro(:,2),dI_dxi2*ro(:,2),...
            dI_dxi3*ro(:,2)];
dhdxi_x = [dI_dxi1*ro(:,3),dI_dxi2*ro(:,3),...
            dI_dxi3*ro(:,3)];
dhdxi_do = [dI_dxi1*ro(:,4),dI_dxi2*ro(:,4),...
            dI_dxi3*ro(:,4)];

Hi = [eye(3),dhdxi_u,Ioe,zeros(3,3);
      eye(3),dhdxi_xy,zeros(3,3),Ioe(:,1:2),zeros(3,1);
      eye(3),dhdxi_x,zeros(3,5),Ioe(:,1);
      eye(3),dhdxi_do,zeros(3,6)];

% H is the desired Jacobian

```

```

H = [Hi;Hi;Hi];

sk = H'*P*(z-h);
wk = gs - sk;
%
% Start local iterations
%
i=0;
while (i<=max_it), % Must pass into loop at least once

    %Compute gains
    L = Chi*inv(H'*P*H);

    % Perform one local iteration
    xk = xk - L*wk;

    % Compute resulting sk and wk to check for early termination
    % ...this must be done anyway to return updated values

    xm = C*xk;          % Get measured state
                        % Collect estimated structure
    ro = [xm(7:9), [xm(10:11);do(3)], [xm(12);do(2:3)],do];
    %
    xi = xm(4:6);        % Compute rotation matrix Ioe
    Ioe = exp2rot(xi);

                        % Rotate and translate structure from Fc to Fe
    re = [Ioe,xm(1:3)]*[ro;ones(1,num_pts)];
    %
    % Compute hr(xm)
    %
    hi = [re(:,1);re(:,2);re(:,3);re(:,4)];
    hr = [hi;hi;hi];
    %

```

```

% Compute Jacobian
%
alf = norm(xi);
xi_bar = xi/alf;
xi_bar_sh = sharp(xi_bar);

% Compute dIoe/d(xi_1)
gamma = (e1_sh - xi_bar_sh*xi_bar(1))/alf;
beta = gamma*xi_bar_sh + xi_bar_sh*gamma;
dI_dxi1 = beta + cos(alf)*(xi_bar(1)*xi_bar_sh - beta)...
          + sin(alf)*(gamma + xi_bar(1)*xi_bar_sh*xi_bar_sh);

% Compute dIoe/d(xi_2)
gamma = (e2_sh - xi_bar_sh*xi_bar(2))/alf;
beta = gamma*xi_bar_sh + xi_bar_sh*gamma;
dI_dxi2 = beta + cos(alf)*(xi_bar(2)*xi_bar_sh - beta)...
          + sin(alf)*(gamma + xi_bar(2)*xi_bar_sh*xi_bar_sh);

% Compute dIoe/d(xi_3)
gamma = (e3_sh - xi_bar_sh*xi_bar(3))/alf;
beta = gamma*xi_bar_sh + xi_bar_sh*gamma;
dI_dxi3 = beta + cos(alf)*(xi_bar(3)*xi_bar_sh - beta)...
          + sin(alf)*(gamma + xi_bar(3)*xi_bar_sh*xi_bar_sh);

% u~unknown xy~xy unknown, x~x unknown, do~wrt do measurement
dhdxi_u = [dI_dxi1*ro(:,1),dI_dxi2*ro(:,1),...
           dI_dxi3*ro(:,1)];
dhdxi_xy = [dI_dxi1*ro(:,2),dI_dxi2*ro(:,2),...
           dI_dxi3*ro(:,2)];
dhdxi_x = [dI_dxi1*ro(:,3),dI_dxi2*ro(:,3),...
           dI_dxi3*ro(:,3)];
dhdxi_do = [dI_dxi1*ro(:,4),dI_dxi2*ro(:,4),...
           dI_dxi3*ro(:,4)];

```

```

Hi = [eye(3),dhdxi_u,Ioe,zeros(3,3);
      eye(3),dhdxi_xy,zeros(3,3),Ioe(:,1:2),zeros(3,1);
      eye(3),dhdxi_x,zeros(3,5),Ioe(:,1);
      eye(3),dhdxi_do,zeros(3,6)];

% H is the desired Jacobian
H = [Hi;Hi;Hi];

sk = H'*P*(z-h); % New value of sk
wk = gs - .sk;   % New value of w
                % Check for loop termination
if (norm(wk) < threshold),
    break;       % Break out of While loop
end;

i = i+1;        % Increment loop counter
end; %While not_done ...local iterations
%
% Now return required parameters and perform time propagation
%
% Compute estimated rotation matrix Ioe
%
Ik = exp2rot(xk(4:6));
%
% Compute estimated angular velocity
%
xi = xk(4:6);
alf = norm(xi);
xi_bar = xi/alf;
xi_dot = xk(16:18);
alf_dot = xi_bar'*xi_dot;
xi_bar_dot = (1/alf)*(eye(3)-xi_bar*xi_bar')*xi_dot;

```

```

wk = alf_dot*xi_bar+sin(alf)*xi_bar_dot+...
      (1-cos(alf))*sharp(xi_bar)*xi_bar_dot;

%
% Check for reset of angle for angle-axis parameterization
%
if (alf > pi)
    xk(4:6) = xi - twopi*xi_bar;
    xk(16:18) = xi_dot - twopi*xi_bar_dot;
    xi_ddot = xk(22:24);
    alf_ddot = xi_bar'*xi_ddot + xi_dot'*xi_dot/alf - ...
               (alf_dot/alf)^2;
    xi_bar_ddot = (eye(3)-xi_bar*xi_bar')*...
                  (xi_ddot/alf-xi_dot*alf_dot/alf^2)-...
                  (xi_bar_dot*xi_bar' + xi_bar*xi_bar_dot')*xi_dot/alf;
    xk(22:24) = xi_ddot - twopi*xi_bar_ddot;
end; %If

%
% Perform time propagation of state estimate
%
xk1 = Phi*xk;

%
%***** END FUNCTION rsmo *****

```

END

1 7 - 1 / 1 - 9 5

FIN