



KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT INGENIEURSWETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK
Kasteelpark Arenberg 10, 3001 Leuven (Heverlee)

KALMAN FILTERING TECHNIQUES FOR SYSTEM INVERSION AND DATA ASSIMILATION

Promotor:
Prof.dr.ir. B. De Moor

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de ingenieurswetenschappen

door

Steven GILLIJNS

December 2007



KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT INGENIEURSWETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK
Kasteelpark Arenberg 10, 3001 Leuven (Heverlee)

KALMAN FILTERING TECHNIQUES FOR SYSTEM INVERSION AND DATA ASSIMILATION

Jury:

Prof.dr.ir. P. Van Houtte, voorzitter
Prof.dr.ir. B. De Moor, promotor
Prof.dr.ir. J. Vandewalle
Prof.dr.ir. P. Van Dooren (UCL)
Prof.dr. S. Poedts
Prof.dr.ir. D. Bernstein (University of Michigan)
Prof.dr.ir. J. Willems
Prof.dr.ir. H. Bruyninckx

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de ingenieurswetenschappen
door

Steven GILLIJNS

©Katholieke Universiteit Leuven – Faculteit Ingenieurswetenschappen
Arenbergkasteel, B-3001 Heverlee (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotocopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm or any other means without written permission from the publisher.

ISBN 978-90-5682-893-6

U.D.C. 681.5.015

D/2007/7515/125

Voorwoord

De eindstreep van een marathon bereik je niet zonder aanmoediging en steun. Supporters, medelopers en trainers, allemaal dragen ze op een specifieke manier bij tot het bereiken van de eindstreep. Ze maken van een marathon eerder een feest dan een lange lijdensweg. Ik wil in dit voorwoord dan ook alle personen bedanken die van deze onderzoek-marathon een aangename tijd gemaakt hebben.

Een atleet heeft in de eerste plaats nood aan een coach, een persoon die een aangepast trainingsschema opstelt en bijstuurt indien nodig. Ik wil mijn promotor Bart De Moor danken voor het aanreiken van het interessante onderwerp. Bart z'n enthousiasme en z'n vernieuwende ideeën hebben een stempel gedrukt op dit proefschrift.

Op het eerste verkennend gesprek met Bart, was er sprake dat ik in de loop van mijn doctoraat misschien even aan de universteit van Michigan zou kunnen verblijven voor een project in verband met ruimteweer. Nog geen half jaar later zette ik in Michigan voet aan de grond voor het eerste verblijf uit een reeks van vier. Ik heb Michigan leren kennen onder de meest verscheiden omstandigheden, van barre koude en sneeuwstormen in de winter tot tropische hittegolven en tornado's in de zomer. *I would like to thank Prof. Bernstein for giving me the opportunity to visit his research group. Thanks for your warm hospitality. The many interesting discussions that we had with Harish, Jaganath and Aaron and the joint publications have contributed in considerable measure to this dissertation.*

Specifieke trainers richten zich vanuit hun achtergrond op een deelaspect van de marathon. Ik bedank de juryleden Prof. J. Vandewalle, Prof. P. Van Dooren, Prof. S. Poedts, Prof. D. Bernstein, Prof. J. Willems en Prof. H. Bruyninckx voor hun begeleiding en voor de opbouwende kritiek. Ook dank ik Prof. P. Van Houtte voor het waarnemen van het voorzitterschap.

Een atleet kan zich maar optimaal voorbereiden op een marathon als hij zich niet hoeft te bekommeren over de administratieve kant zoals de inschrijving van de wedstrijd. Bedankt Ida en Ilse voor jullie hulp en raad in de administratieve en financiële zaken.

Mede-atleten zijn cruciaal voor het bereiken van de eindstreep. Ze sporen je aan, geven je een duwtje in de rug of stellen je voor om in groep naar de eindstreep te lopen. Enkele mede-doctoraatsstudenten zijn gedurende deze vier jaren echte vrienden geworden. Ik denk hier in de eerste plaats aan mijn

eilandgenoten Bart, Jeroen en Erik, maar ook aan Tom en Niels. Ook bedank ik de groep waarmee ik op woensdag soms al eens naar de Alma ging en onze leuke groep van nieuwkomers waarmee we het SISTA weekend organiseerden. Verder denk ik met plezier terug aan de korte vakanties die we aan de conferenties in Sydney en San Diego koppelde.

Elke atleet heeft op geregelde tijdstippen nood aan rust en ontspanning. Ik vind de beste ontspanning door te wandelen of te sporten. Ik wil de vele “Anders reizen” genoten danken voor de mooie wandelmomenten die we samen beleefden. Uit deze reizen zijn enkele mooie vriendschappen gegroeid. Verder wil ik ook de vrienden van de Furalopers, HRC en Sport en Vermaak bedanken voor hun interesse in mijn werk en uiteraard ook voor de leuke babbels tijdens en na de trainingen. Ook dank ik mijn vrienden uit de humaniora, Thomas en Hans, voor de gezellige wandelingen, etentjes en babbels.

Vanzelfsprekend richt ik ook een woordje van dank aan mijn fans, aan de personen die mij bestoken met vragen als “en, in vorm vandaag?” of “welke plaats behaalde je?” Katleen & Bernard, Veerle & Hans, bedankt voor jullie steun en voor de vele aangename en gezellige momenten die we met onze hechte familie beleven. Uiteraard kan ik het ook niet nalaten om mijn oogappels Thomas en Kirsten te bedanken. Jullie zijn mijn “grootste” fans.

Tot slot wil ik mijn vaste supporters in de bloemetjes zetten. De supporters die mij op de voet volgen, wedstrijd na wedstrijd. De supporters die delen in mijn vreugde en die mij bijstaan als het eens een dagje iets minder gaat. De supporters die altijd voor mij klaar staan. Bedankt, liefste mama en papa.

Steven Gillijns,
Lewen, November 2007.

Abstract

Since its introduction in 1960, the Kalman filter has gained increasing popularity. It has become the standard technique for estimating the present state of a dynamical system based on a numerical model of that system and a set of observations. This thesis contributes to the popularity of the Kalman filter by addressing the problems of system inversion and data assimilation from the viewpoint of Kalman filtering.

In applications such as fault detection and cryptography, the dynamical system is subject to inputs that are unknown, but yet are of major importance. The problem of estimating the inputs of a dynamical system from observations of that system's outputs, has been termed *system inversion*. In the first part of this thesis, a new inversion procedure based on joint input-state estimation is developed. Conditions are derived under which the poles of the estimator can be assigned and the speed of convergence can thus be tuned. In case of noise, it is shown that the poles can be placed so that, in analogy to the Kalman filter, the estimates of the system state and the system input are optimal in a least-squares sense. Several computational and numerical issues such as reduced order estimation and square-root estimation are addressed. The inversion procedure is employed in four applications.

Due to its high computational cost and its immense storage requirements, the Kalman filter is not directly applicable with the large-scale numerical models that are usually employed in environmental problems such as weather prediction. The challenging problem of assimilating observations in such complex numerical models has been termed *data assimilation*. In the second part of this thesis, data assimilation techniques are developed for nowcasting a space weather event that emulates the topology and the dynamics of the bow shock that is formed when the supersonic solar wind encounters the Earth. A suboptimal Kalman filter is developed that is adapted to the data-sparse environment of space weather. Simulation results on a large-scale model show that the estimates produced by the new suboptimal filter outperform a data-free simulation, even if only a few observations are available.

Korte inhoud

Sinds de introductie in 1960, heeft het Kalman filter enkel aan populariteit gewonnen. Het is momenteel de standaardmethode om de toestand van een dynamisch systeem te schatten op basis van een numeriek model en van metingen van dat systeem. Dit proefschrift draagt bij tot de populariteit van het Kalman filter door de problemen van systeem inversie en data assimilatie te behandelen vanuit het gezichtspunt van Kalman filtering.

In toepassingen zoals foutdetectie en cryptografie is het dynamisch systeem onderhevig aan ongekende ingangen waarvan de waarde van cruciaal belang is. Het probleem dat erin bestaat de ingangen van een systeem te schatten uit kennis van de uitgangen van dat systeem, wordt *systeem inversie* genoemd. In het eerste deel van dit proefschrift, wordt een nieuwe inversie procedure ontwikkeld die gebaseerd is op het gelijktijdig schatten van de ingang en de toestand van een systeem uit kennis van de uitgang. Voorwaarden worden afgeleid waaronder de polen van de schatter geplaatst kunnen worden en de snelheid van convergentie dus geregeld kan worden. In de aanwezigheid van ruis, wordt aangetoond dat de polen zodanig geplaatst kunnen worden dat, in analogie met het Kalman filter, de schattingen optimaal zijn volgens het criterium van de kleinste-kwadraten. Verschillende computationele en numerieke problemen worden aangepakt, zoals een reductie in rekencomplexiteit en een ontwikkeling van numeriek hoogstaande algoritmes. De inversie procedure wordt aangewend in vier toepassingen.

Omwille van de hoge rekencomplexiteit en het extreme geheugenverbruik, is het Kalman filter niet rechtstreeks toepasbaar op de grootschalige modellen die gebruikt worden om onder andere het weer te voorspellen. Het uitdagende probleem om metingen te verwerken in dergelijke grootschalige modellen wordt *data assimilatie* genoemd. In het tweede deel van dit proefschrift worden data assimilatie technieken ontwikkeld voor een toepassing in ruimteweer. De toepassing bestaat erin de topologie en de dynamica van de boegschok te schatten die gevormd wordt als de supersonische zonnwind de aarde passeert. Een suboptimaal Kalman filter wordt ontwikkeld dat geoptimaliseerd is voor de schaarsheid aan metingen in ruimteweer. Simulatieresultaten met een grootschalig model tonen aan dat het suboptimale filter een data-vrije simulatie overtreft, zelfs als er slechts metingen van enkele satellieten beschikbaar zijn.

Glossary

Notation

Variables

a, b, c	Vector variables
A, B, C	Matrix variables
I	Identity matrix of appropriate dimensions

Sets

\mathbb{R}	The set of real numbers
\mathbb{R}^n	The set of n -dimensional real vectors
$\mathbb{R}^{n \times m}$	The set of $n \times m$ real matrices
\mathbb{C}	The set of complex numbers
\mathbb{C}^n	The set of n -dimensional complex vectors
$\{a_1, \dots, a_n\}$	The set consisting of the vectors a_1, \dots, a_n
$\{a_i\}_{i=1}^n$	Shorthand notation for the set $\{a_1, \dots, a_n\}$

Matrix operations

A^T	Transpose of matrix A
A^{-1}	Inverse of matrix A
A^\dagger	Moore-Penrose generalized inverse of matrix A
$A^{(1)}$	One-inverse of matrix A , i.e. any matrix satisfying $AA^{(1)}A = A$
$A^{1/2}$	Square-root of matrix A
$\text{rank}(A)$	Rank of matrix A
$\text{trace}(A)$	Trace of matrix A
$\Lambda(A)$	The set of eigenvalues of A
$\text{diag}(A, B, \dots)$	A (block) diagonal matrix with entries A, B, \dots

Random variables

$\mathbb{E}[a]$	Expected value of the random vector a
\hat{a}	Estimate of the (random) vector a

Norms and optimization

$ x $	Absolute value of the number x
$\ x\ $	Two-norm of vector $x : \sqrt{x^T x}$
$\ x\ _W$	Weighted two-norm of vector $x : \sqrt{x^T W x}$
\min_x	Function minimization over x , optimal function value is returned
$\arg \min_x$	Function minimization over x , optimal value of x is returned

Operators

$\frac{\partial}{\partial x}$	Partial differentiation with respect to x
\times	Vector product
\cdot	Scalar product
∇	Del operator
$:=$	The left hand side is defined as the right hand side
$=:$	The right hand side is defined as the left hand side
\approx	Is approximately equal to
\ll	Is orders of magnitude smaller than
\gg	Is orders of magnitude larger than

Fixed symbols

$x \in \mathbb{R}^n$	System state
$y \in \mathbb{R}^p$	System output
$u \in \mathbb{R}^m$	System input
A, B, C, D, E	System matrices
w, v	Noise vectors
P, Q, R	Covariance matrices

List of abbreviations

CME	Coronal Mass Ejection
LS	Least-Squares
LTI	Linear Time Invariant
MHD	Magnetohydrodynamics
MIMO	Multiple Input / Multiple Output
MSE	Mean Squared Error
MVU	Minimum-Variance Unbiased
NMP	Nonminimum Phase
ODE	Ordinary Differential Equation
PDE	Partial Differential Equation
RLS	Recursive Least-Squares
SISO	Single Input / Single Output

Contents

Voorwoord	i
Abstract	iii
Korte inhoud	v
Glossary	vii
Contents	ix
Nederlandse samenvatting	xiii
1 Introduction	1
1.1 Motivation and objectives	2
1.1.1 Data assimilation	2
1.1.2 System inversion	7
1.2 Chapter-by-chapter overview	9
1.3 Personal contributions	13
1.3.1 System inversion	14
1.3.2 Data assimilation	15
2 The Kalman Filter Revisited	17
2.1 Introduction	17
2.2 Filtering, prediction and smoothing	18
2.3 Recursive estimation for noise-free systems	19
2.3.1 Observability and detectability	19
2.3.2 Asymptotic and deadbeat estimation	21
2.4 The Kalman filter	23
2.4.1 Derivation of the Kalman filter equations	24
2.4.2 Time and measurement update	26
2.5 Information filtering	27
2.5.1 Least-squares state estimation	28
2.5.2 Recursive least-squares filtering	29
2.5.3 Information Kalman filtering	32
2.6 Square-root filtering	33

2.6.1	Square-root covariance filtering	33
2.6.2	Square-root information filtering	35
2.7	The extended Kalman filter	36
2.7.1	Derivation of filter equations	36
2.7.2	Observability	37
2.8	Conclusion	37

I System Inversion 39

3 Inversion of Deterministic Systems 41

3.1	Introduction	41
3.2	Problem formulation	44
3.2.1	Left inversion	44
3.2.2	Right inversion	45
3.2.3	Duality	46
3.3	Left invertibility of state-space systems	46
3.3.1	The invertibility condition of Sain & Massey	48
3.3.2	Left inversion and system zeros	50
3.4	Inversion techniques: state of the art	51
3.4.1	Instantaneous inversion	51
3.4.2	The approach of Sain & Massey	52
3.4.3	Comparison to Silverman's structure algorithm	53
3.5	An estimation approach to system inversion	53
3.5.1	State reconstruction	54
3.5.2	Input reconstruction	56
3.5.3	A general form of an L -delay left inverse	57
3.6	Stable inversion	58
3.6.1	Joint input-state estimation	58
3.6.2	Pole placement	59
3.7	Stable reduced order inversion	64
3.7.1	Reduced order inversion	64
3.7.2	Stable reduced order inversion	66
3.8	Numerical examples	68
3.9	Conclusion	70

4 Inversion of Combined Deterministic-Stochastic Systems 73

4.1	Introduction	73
4.2	Optimal filtering with direct feedthrough	76
4.2.1	State estimation	76
4.2.2	Input estimation	79
4.2.3	Time and measurement update	80
4.2.4	Summary of filter equations	81
4.2.5	Relation to least-squares estimation	82
4.3	Optimal filtering without direct feedthrough	86
4.3.1	State estimation	86
4.3.2	Time and measurement update	88

4.3.3	Input estimation	88
4.3.4	Summary of filter equations	89
4.3.5	Recursive least-squares estimation	90
4.3.6	Square-root information filtering	95
4.3.7	A note on square-root covariance filtering	96
4.4	A general framework	97
4.4.1	State estimation	97
4.4.2	Input estimation	102
4.4.3	Joint input-state estimation	103
4.5	Numerical examples	104
4.6	Conclusion	108
5	Applications of System Inversion	109
5.1	Introduction	109
5.2	Filtering with noisy inputs and outputs	110
5.2.1	Problem formulation	111
5.2.2	Errors-in-variables filtering	112
5.2.3	Filtering with noisy input measurements	112
5.2.4	Summary of filter equations	114
5.2.5	Numerical example	115
5.3	Filtering in the presence of bias	116
5.3.1	Derivation of filter equations	118
5.3.2	Summary of filter equations	120
5.3.3	Numerical example	121
5.4	Model error estimation and model updating	121
5.4.1	Model error estimation	123
5.4.2	Subsystem identification and model updating	125
5.5	Boundary condition estimation	129
5.5.1	Problem formulation	130
5.5.2	Basis function expansion	131
5.5.3	Heat conduction example	131
5.6	Conclusion	133
II	Data Assimilation	135
6	Suboptimal Square-Root Filtering	137
6.1	Introduction	137
6.2	Suboptimal square-root filtering: the idea	140
6.3	Square-root measurement updating	141
6.3.1	Simultaneous processing	142
6.3.2	Sequential processing	142
6.4	Reduced rank filtering	143
6.4.1	The reduced rank square-root filter	144
6.4.2	The reduced rank transform square-root filter	146
6.5	Spatially localized filtering	148
6.5.1	The spatially localized Kalman filter	149

6.5.2	Reduced rank spatially localized filtering	150
6.6	Filter degradation due to a lower rank approximation of the error covariance matrix	157
6.6.1	Error in the covariances	158
6.6.2	Error in the variances and the covariances	159
6.7	Numerical examples	159
6.8	Conclusion	163
7	Space Weather Nowcasting Example	167
7.1	Introduction	167
7.2	Magnetohydrodynamics	169
7.2.1	The ideal MHD equations	170
7.2.2	Computational MHD	171
7.3	MHD shocks	172
7.3.1	Shock topology	172
7.3.2	Two-dimensional MHD flow around a cylinder	173
7.4	Data assimilation for two-dimensional MHD flow around a cylinder	174
7.4.1	Setup of the simulations	175
7.4.2	Known, constant boundary conditions	177
7.4.3	Unknown, time-varying boundary conditions	183
7.5	Conclusion	185
8	Conclusions and directions for further research	189
8.1	Conclusions	189
8.2	Directions for further research	192
8.2.1	System inversion	192
8.2.2	Data assimilation	194
A	Generalized inverses and the matrix inversion lemma	197
A.1	Generalized inverses	197
A.2	The matrix inversion lemma	198
B	Least-squares estimation: deterministic vs. stochastic setting	199
B.1	Deterministic setting	199
B.2	Stochastic setting	200
C	Proofs and derivations	201
C.1	Rank proofs	201
C.2	Proof of Proposition 4.1	204
C.3	Derivation of the Eqs. in Sect. 5.3.2	205
	Bibliography	207
	Scientific curriculum vitae	217
	Publication list	219

Nederlandse samenvatting

Kalman filtering technieken voor systeem inversie en data assimilatie

Hoofdstuk 1: Inleiding

Voortbouwend op de methode van kleinste-kwadraten (KK) schatting, introduceerde Kalman in 1960 [83] een schattingsprocedure die nu het *Kalman filter* genoemd wordt. In wezen is het Kalman filter een recursieve schatter die de interne toestand van een dynamisch systeem schat op basis van een numeriek model voor dat systeem en op basis van kennis van de ingangen (de drijvers) en de uitgangen (de metingen) van dat systeem.

Het Kalman filter werd voor het eerst gebruikt in 1961, toen het de landingsmodule van de Apollo 11 ruimtevaartmissie begeleidde naar het oppervlak van de maan. Al snel vond het Kalman filter toepassingen in andere domeinen, zoals in de chemische industrie, de econometrie, het Global Positioning System en de luchtvaartindustrie. Het aantal wetenschappelijke artikels en boeken dat handelt over het Kalman filter groeit dag na dag. De wetenschappelijke zoekrobot <http://scholar.google.com/> geeft momenteel maar liefst 111000 hits op het trefwoord “Kalman filter”.

Ondanks het enorme succes van het Kalman filter zijn in verschillende toepassingen uitbreidingen of benaderingen van het algoritme nodig. In dit doctoraat behandelen we twee schattingsproblemen die een uitbreiding vereisen. Het eerste schattingsprobleem, genoemd *data assimilatie*, breidt het Kalman filter uit naar grootschalige modellen. Het tweede schattingsprobleem, genoemd *systeem inversie*, behandelt het geval waarbij de ingang van het systeem ongekend is. De motivatie voor de studie van de deze problemen en de persoonlijke bijdragen worden nu meer in detail besproken.

Data assimilatie

Motivatie De motivatie om uitbreidingen van het Kalman filter voor groot-schalige modellen te ontwikkelen, wordt gevoed vanuit een toepassing in *ruimteweer*. De zon stoot een constante stroom van *plasma*, hoog energetische deeltjes, de ruimte in. Dit fenomeen wordt de *zonnewind* genoemd. De zonnewind beweegt zich doorheen de ruimte aan supersonische snelheden. Het gevolg hiervan is dat er een *boegschok* gevormd wordt als de zonnewind een obstakel zoals de aarde tegenkomt, net zoals er een schok gevormd wordt bij een vliegtuig dat door de geluidsmuur gaat. Als gevolg van de geladen deeltjes in het plasma, is de topologie en dynamica van de aardse boegschok echter veel complexer dan deze bij een vliegtuig. Bovendien kan de snelheid van de zonnewind significant wijzigen in een tijdschaal van enkele seconden. Dit gebeurt onder andere wanneer er een *coronale massa ejectie*, een van de meest energetische zonne-uitbarstingen, gesuperponeerd is op de zonnewind. De topologie van de boegschok is dus erg dynamisch.

Gedurende de laatste decennia hebben wetenschappers numerieke modellen ontwikkeld die de dynamica van de aardse boegschok (en talrijke andere fenomenen die gerelateerd zijn aan de zonnewind) beschrijven [27]. Anderzijds zijn er enkele satellieten gelanceerd die het weer in de ruimte waarnemen. Alle ingrediënten (een numeriek model en metingen) voor de toepassing van het Kalman filter zijn bijgevolg aanwezig.

Er zijn echter twee redenen die de toepassing van het Kalman filter belemmeren. De eerste reden is de beperktheid van het Kalman filter tot lineaire modellen. De numerieke modellen in ruimteweer zijn gebaseerd op de stromingswetten van Navier-Stokes en de elektromagnetische wetten van Maxwell en zijn bijgevolg erg niet-lineair. De tweede reden is de rekencomplexiteit en het geheugenverbruik van het Kalman filter. De numerieke modellen in ruimteweer propageren typisch 10^5 à 10^6 variabelen. De rekentijd van het Kalman filter zou dan ongeveer 10^5 à 10^6 keer de tijd voor een simulatie bedragen. Het geheugenverbruik van het Kalman filter, dat voornamelijk bepaald wordt door de opslag van de zogenaamde *foutencovariantiematrix*, zou dan oplopen tot enkele terabytes, dat is ongeveer de totale hoeveelheid informatie in een grote universiteitsbibliotheek.

Het is duidelijk dat een benadering van het Kalman filter nodig is voor dergelijke grootschalige modellen. Gedreven vanuit voornamelijk toepassingen in de voorspelling van het aardse weer en de stroming in oceanen, werden er verschillende suboptimale benaderingen van het Kalman filter voorgesteld in de literatuur [39, 85, 112, 135]. In dit doctoraat bouwen we verder op het *gereduceerde-rang vierkantswortel filter* [135], waarin de benadering gebaseerd is op een optimale lagere-rang benadering van de foutencovariantiematrix. Het doel is om dit algoritme aan te passen en te optimaliseren voor de specifieke omstandigheden in ruimteweer. Deze omstandigheden onderscheiden zich van andere toepassingen door het zeer beperkte aantal metingen aan de ene kant en de enorme afmetingen aan de andere kant. De algoritme moet de schaarsheid aan metingen vertalen in numerieke efficiëntie en moet robuust zijn tegen de

problemen die kunnen optreden als gevolg van het beperkte aantal metingen.

Persoonlijke bijdragen De belangrijkste bijdragen in data assimilatie zijn de aanpassing van het gereduceerde-rang vierkantswortel filter aan de schaarsheid van metingen in ruimteweer enerzijds en de succesvolle toepassing van het resulterende suboptimale filter in een ruimteweer-simulatie anderzijds.

- De schaarsheid van metingen wordt aangepakt door een combinatie van twee technieken. De eerste techniek maakt gebruik van het algoritme van Potter [111] om de schaarsheid aan metingen te vertalen naar numerieke efficiëntie. De tweede techniek is gebaseerd op het *ruimtelijk gelokaliseerd Kalman filter* [9] en heeft als doel enkel de waarden van de variabelen te schatten die effectief gecorreleerd zijn met de metingen. Een belangrijke bijdrage van dit doctoraat is het verweven van beide technieken in het gereduceerde-rang vierkantswortel filter. Het resulterende algoritme is echt geschikt voor grootschalige toepassingen waarin het aantal metingen erg beperkt is.
- Het algoritme wordt succesvol toegepast in grootschalige simulaties (ongeveer 10^5 te schatten variabelen) die de dynamica van de aardse boegschok modelleren onder veranderende condities van de zonnewind. Zowel simulaties met gekende als ongekende randvoorwaarden worden beschouwd. In het laatste geval wordt het filter uitgebreid zodanig dat het de randvoorwaarden mee schat. De simulatieresultaten tonen aan dat het suboptimale filter een significante reductie in de schattingsfout kan leveren, zelfs als er metingen van slechts vier satellieten beschikbaar zijn.

Systeem inversie

Motivatie De motivatie om het probleem van systeem inversie te bestuderen, wordt gevoed vanuit verschillende toepassingen.

- Foutdetectie: Voor bepaalde systemen, zoals vliegtuigen, mechanische robots en chemische installaties, is er een kans op fouten of verstoringen die ernstige verwondingen en schade tot gevolg kunnen hebben. Het detecteren en schatten van dergelijke fouten is bijgevolg van cruciaal belang. Vermits fouten kunnen gemodelleerd worden als ongekende ingangen, komt de schatting ervan neer op het bepalen van de ingang van een systeem uit kennis van de uitgang van het systeem.
- Schatten en verbeteren van modelfouten: Elk model is slechts een benadering van het werkelijke systeem. Onnauwkeurigheden in fysische modellen zijn te wijten aan ongekende dynamica, te grove benaderingen, foute waarden van parameters, . . . Anderzijds zijn er voor de meeste systemen metingen beschikbaar die informatie leveren over de onderliggende dynamica en dus gebruikt kunnen worden om de modelfouten te schatten. Net als verstoringen kunnen modelfouten gezien worden als ongekende ingangen die inwerken op het systeem.

In beide voorbeelden is er een nood om de ingang van een systeem te schatten uit kennis van de uitgang van het systeem. De oplossing van dit probleem komt neer op het ontwikkelen van een schatter die als ingang de uitgang van het systeem heeft en als uitgang de ingang van het systeem. De ingangen en uitgangen van de schatter zijn dus *geïnverteerd* ten opzichte van die van het systeem. Vandaar de naam systeem *inversie*.

De eerste inversie technieken werden ontwikkeld op het einde van de jaren zestig [16, 115, 116]. Bestaande methodes zijn echter beperkt tot het ideale geval van ruis-vrije systemen. Het doel van dit doctoraat is om nieuwe inversie technieken te ontwikkelen die eenvoudig uitbreiden naar systemen met ruis.

Persoonlijke bijdragen In dit doctoraat wordt een nieuwe techniek voor systeem inversie ontwikkeld op basis van schattingstheorie. De inversie techniek wordt eerst uitgewerkt voor ruis-vrije systemen en later uitgebreid naar systemen die onderhevig zijn aan ruis.

- Zoals Sain en Massey [115], beschouwen we in dit doctoraat inverse systemen die bestaan uit een bank van vertragingselementen, gevolgd door een dynamisch systeem. Een belangrijke bijdrage van dit doctoraat is de afleiding van de algemene vorm van zo een dynamisch systeem op basis van schattingstheorie. In het ruis-vrije geval levert het inverse systeem een exacte reconstructie van zowel de ingang als de toestand. Het inverse systeem kan dus beschouwd worden als een gezamenlijke toestands- en ingangsschatter.
- De algemene vorm van de schatter bevat twee parameters die vrij gekozen kunnen worden. We leiden voorwaarden en methodes af om de polen van de schatter te plaatsen door een gepaste keuze van deze parameters. Op die manier kan de convergentiesnelheid van de schatter geregeld worden.
- In de aanwezigheid van ruis, tonen we aan dat de polen zodanig kunnen geplaatst worden dat de schattingen onvertekend zijn en minimale variantie hebben. Bovendien wordt een verband met KK schatting afgeleid.
- Verschillende numerieke problemen worden aangekaart en aangepakt. In het deterministische geval wordt een methode ontwikkeld om de orde van de schatter en dus de rekencomplexiteit te reduceren. In de aanwezigheid van ruis worden er zogenaamde *informatie* en *vierkantswortel* implementaties uitgewerkt. Deze laatste implementaties reduceren de propagatie van numerieke fouten.

De twee laatste bijdragen tonen aan dat verschillende methodes die reeds lang gekend waren in de context van Kalman filtering ook uitbreidbaar zijn naar systeem inversie. Het verband tussen het Kalman filter en KK schatting is immers reeds gekend sinds het einde van de jaren zestig [77, 117]. Ook vierkantswortel en informatie implementaties, die een direct gevolg zijn van

de formulering van het Kalman filter als KK probleem, werden in de context van Kalman filtering reeds uitvoerig bestudeerd [4]. Het regelsysteem dat de landingsmodule van de Apollo 11 ruimtevaartmissie naar het oppervlak van de maan begeleidde, maakte reeds gebruik van een vierkantswortel implementatie.

Hoofdstuk 2: Het Kalman filter herbekeken

Beschouw het lineaire tijdsinvariante systeem

$$\mathcal{S} : \begin{cases} x_{[k+1]} = Ax_{[k]} + Bu_{[k]} + w_{[k]} \\ y_{[k]} = Cx_{[k]} + Du_{[k]} + v_{[k]}, \end{cases}$$

met $x_{[k]} \in \mathbb{R}^n$ de toestandsvector op tijdstip k , $y_{[k]} \in \mathbb{R}^p$ de uitgangsvector op tijdstip k , en $u_{[k]} \in \mathbb{R}^m$ de ingangsvector op tijdstip k . We veronderstellen dat de systeem matrices A, B, C, D , en E evenals de ingang u gekend zijn. De vectoren $w_{[k]} \in \mathbb{R}^l$ en $v_{[k]} \in \mathbb{R}^p$ stellen ongekende ruistermen voor die modelfouten, verstoringen, meetfouten, ... in rekening brengen. Indien het systeem vrij is van ruistermen, spreken we van een deterministisch systeem.

Het *schattingsprobleem* bestaat er in om voor elke k een schatting van $x_{[k]}$ te berekenen uit kennis van de uitgang y tot op tijdstip l . We noteren in het vervolg zo'n schatting als $\hat{x}_{[k|l]}$. Als $l = k$, spreken we over *filteren*, als $l > k$ over *effenen*, en als $l < k$ over *voorspellen*. Filtering is het meest bestudeerde probleem van de drie aangezien dit overeenstemt met schatten in real-time.

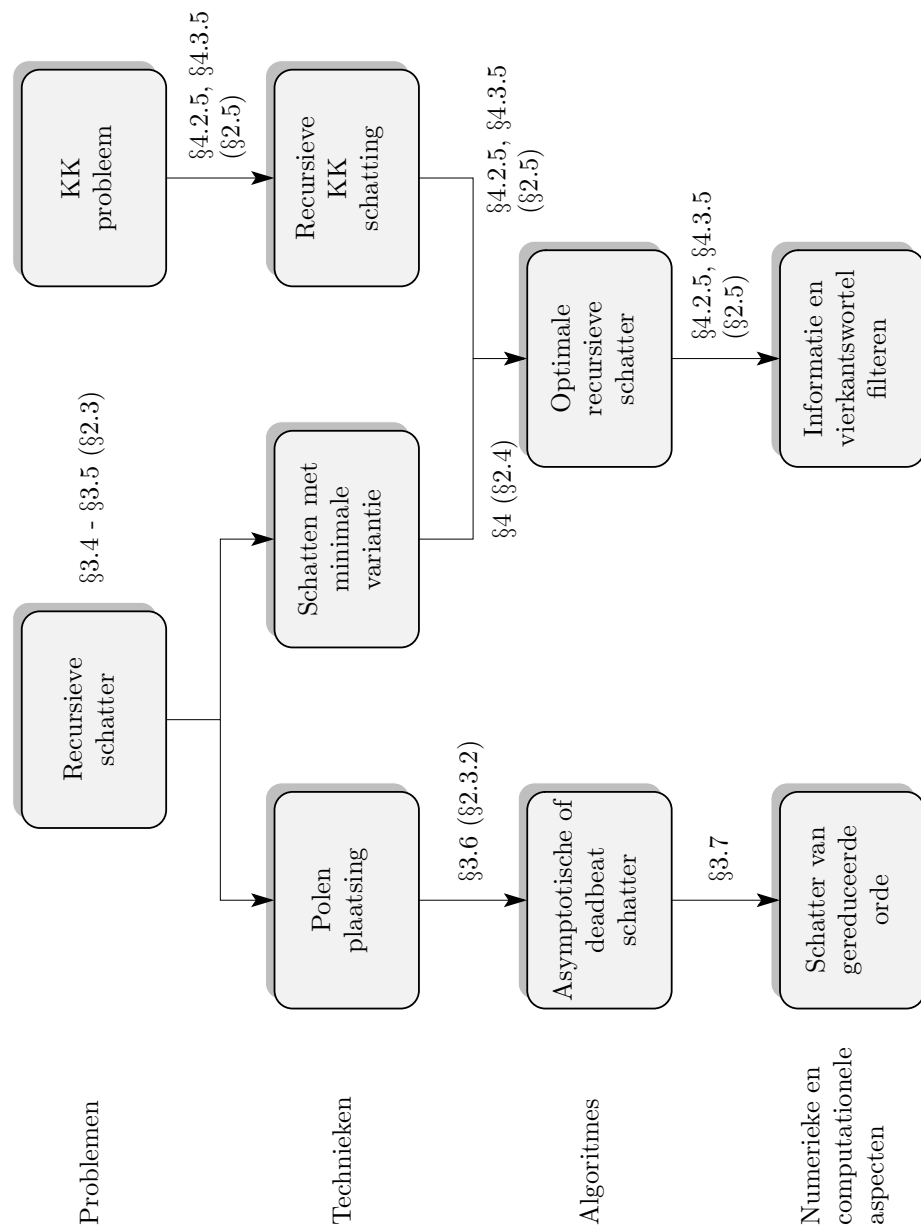
Figuur 0.1 vat een aantal belangrijke concepten en technieken in verband met het schattingsprobleem samen. De pijlen geven de verbanden tussen de concepten en technieken weer. De nummers bij de pijlen duiden de hoofdstukken en paragrafen aan waarin deze problemen bestudeerd worden. De nummers tussen de haakjes slaan op paragrafen in verband met Kalman filtering, de andere nummers op paragrafen in verband met systeem inversie. We beschouwen nu deze concepten meer in detail in de context van Kalman filtering.

Ruis-vrij filteren

Eerst bestuderen we het filteren van ruis-vrije systemen meer in detail. Zoals aangegeven in Figuur 0.1 bestaat een van de meest gebruikte technieken erin om een recursieve toestandsschatter van de vorm

$$\hat{x}_{[k+1|k]} = A\hat{x}_{[k|k-1]} + Bu_{[k]} + K(y_{[k]} - C\hat{x}_{[k|k-1]} - Du_{[k]})$$

te beschouwen en de zogenaamde *winst-matrix* K te bepalen zodanig dat de schattingsfout $x_{[k]} - \hat{x}_{[k|k-1]}$ naar nul convergeert (het *asymptotisch* schattingsprobleem) of exact nul wordt in een eindig aantal stappen (het *deadbeat* schattingsprobleem). Zoals aangegeven in Figuur 0.1, berust de ontwikkeling van een dergelijke schatter op voorwaarden en methodes om de polen van de schatter, of equivalent de eigenwaarden van $A - KC$, te plaatsen. Een asymptotische schatter bestaat als $\{A, C\}$ *detecteerbaar* is, een deadbeat schatter als $\{A, C\}$ *observeerbaar* is [4].



Figuur 0.1: Overzicht van een aantal belangrijke concepten en technieken in schattingsproblemen. De pijlen geven de verbanden tussen de concepten en technieken weer. De nummers bij de pijlen duiden de hoofdstukken en paragrafen aan waarin deze problemen bestudeerd worden. De nummers tussen de haakjes slaan op paragrafen in verband met Kalman filtering, de andere nummers op paragrafen in verband met systeem inversie.

Luenberger [92] toonde aan dat een deel van de toestandsvector $x_{[k]}$ rechtstreeks uit de meting $y_{[k]}$ kan gereconstrueerd worden en ontwikkelde op basis van dit principe een toestandsschatter van gereduceerde orde die dus de rekencomplexiteit beperkt.

Filteren in de aanwezigheid van ruis

Zoals aangegeven in Fig. 0.1, onderscheiden we in de aanwezigheid van ruis twee verschillende methodes. De eerste methode gaat uit van een stochastische veronderstelling over de ruistermen en bestaat erin, zoals in het deterministische geval, te vertrekken van een recursieve schatter en de winst-matrix te bepalen zodat de schattingen onvertekend zijn en minimale variantie hebben. De tweede methode bestaat erin een KK probleem op te stellen en dit recursief op te lossen. In deze methode is geen stochastische veronderstelling over de ruistermen nodig. Beide methodes leveren de Kalman filter vergelijkingen.

Onvertekend filteren met minimale variantie Ondanks de extreem snelle convergentie, is een deadbeat schatter erg gevoelig aan ruis. In de aanwezigheid van ruis, is het eerder aangewezen om de polen van de schatter te plaatsen zodat er een optimale afweging gemaakt wordt tussen de snelheid van convergentie en de gevoeligheid aan ruis. Op dit principe is de afleiding van het Kalman filter gebaseerd.

In de veronderstelling dat de ruistermen w en v ongecorreleerde witte toevalsvariabelen zijn met verwachte waarde nul en gekende covariantie matrices, beschouwde Kalman [83] een recursieve schatter van de vorm

$$\hat{x}_{[k+1|k]} = A\hat{x}_{[k|k-1]} + Bu_{[k]} + K_{[k]}(y_{[k]} - C\hat{x}_{[k|k-1]} - Du_{[k]})$$

en bepaalde de winst-matrix $K_{[k]}$ zodanig dat de *verwachte gekwadrateerde fout* $\mathbb{E}[\|x_{[k+1]} - \hat{x}_{[k+1|k]}\|^2]$ geminimaliseerd werd. De resulterende winst-matrix wordt de *Kalman winst-matrix* genoemd. De berekening van de Kalman winst-matrix vereist dat de zogenaamde *foutencovariantiematrix* $P_{[k|k-1]} := \mathbb{E}[(x_{[k]} - \hat{x}_{[k|k-1]})(x_{[k]} - \hat{x}_{[k|k-1]})^T]$ gepropageerd wordt.

Kleinste-kwadraten filteren Zoals weergegeven in Figuur 0.1 kunnen de Kalman filter vergelijkingen eveneens afgeleid worden door het recursief oplossen van een groot KK probleem. Beschouwen we een KK probleem van de vorm

$$\min_{x_{[0]}, \dots, x_{[k+1]}} \|x_{[0]} - \hat{x}_{[0|-1]}\|_{P_{[0|-1]}^{-1}}^2 + \sum_{i=0}^k \|v_{[i]}\|_{R^{-1}}^2 + \sum_{i=0}^k \|w_{[i]}\|_{Q^{-1}}^2$$

waarin $P_{[0|-1]}$, R en Q gewichtsmatrices zijn en met als beperkingen de systeemvergelijkingen van \mathcal{S} , dan kan er worden aangetoond [77, 99, 127, 136] door recursieve oplossing van dit KK probleem, dat opnieuw de Kalman filter vergelijkingen bekomen worden.

In deze afleiding is geen stochastische veronderstelling van de ruistermen nodig. Bovendien geeft deze afleiding ook aanleiding tot alternatieve formuleringen van de vergelijkingen, zoals een formulering in termen van de *informatiematrix* (de inverse van de foutencovariantiematrix) of een opsplitsing in een zogenaamde *tijdsstap* en *meetstap*, gegeven door

$$\begin{aligned}\hat{x}_{[k|k]} &= \hat{x}_{[k|k-1]} + L_{[k]}(y_{[k]} - C\hat{x}_{[k|k-1]} - Du_{[k]}) \\ \hat{x}_{[k+1|k]} &= A\hat{x}_{[k|k]} + Bu_{[k]},\end{aligned}$$

met $K_{[k]} = AL_{[k]}$. Ook numeriek betrouwbare implementaties die een vierkantswortel van de foutencovariantiematrix of informatiematrix propageren, volgen rechtstreeks uit deze afleiding [4, 82].

Deel I: Systeem Inversie

Er moet een onderscheid gemaakt worden tussen twee varianten van het inversie probleem. Een *linker inverse* van een systeem reconstrueert de ingang die aangelegd werd aan dat systeem uit kennis van de uitgang van dat systeem. Een linker inverse kan dus geïnterpreteerd worden als een schatter. Een *rechter inverse* daarentegen, berekent een ingang zodat de uitgang een gewenste waarde aanneemt. Een rechter inverse kan dus geïnterpreteerd worden als een voorwaartse regelaar. In dit doctoraat wordt voornamelijk het probleem van linker inversie bestudeerd.

De opbouw en samenhang van de belangrijkste resultaten is sterk gerelateerd aan het schema uit Figuur 0.1. In Hoofdstuk 3 beschouwen we de inversie van deterministische systemen. We leiden een recursieve schatter af en bepalen voorwaarden en methodes om de polen te plaatsen. In Hoofdstuk 4 breiden we de methode uit naar systemen met een stochastische component. We leiden een optimale recursieve schatter af op twee verschillende manieren: ten eerste door de polen van de schatter uit Hoofdstuk 3 optimaal te plaatsen en ten tweede aan de hand van KK schatting.

Hoofdstuk 3: Inversie van Deterministische Systemen

De eerste inversie technieken voor deterministische systemen werden ontwikkeld op het einde van de jaren zestig [16, 115, 116]. Al snel werd echter opgemerkt dat deze technieken onstabiele inverses kunnen leveren. De afleiding van stabiele inverses werd eerst aangekaart in [100]. Methodes om de polen te plaatsen, werden eerst bestudeerd in [7].

In dit doctoraat beschouwen we zoals Sain en Massey [115] inverses die bestaan uit een bank van vertragingselementen gevolgd door een dynamisch systeem. De belangrijkste bijdragen van dit doctoraat bestaan uit de afleiding van een algemene vorm van zo een dynamisch systeem, uit de afleiding van methodes en voorwaarden waaronder de polen van het inverse systeem geplaatst kunnen worden en uit een combinatie van polenplaatsing en reductie in orde.

Beschouw opnieuw het systeem \mathcal{S} , maar nu in de veronderstelling dat u ongekend is en dat de ruistermen w en v nul zijn. Een inverse systeem van \mathcal{S} is dan eenvoudig te definiëren op basis van de *transfer functie* $H(z)$ van \mathcal{S} , gegeven door

$$H(z) = C(zI - A)^{-1}B + D,$$

waarbij z een complexe variabele is. Het systeem \mathcal{S} wordt dan *L-vertraagd links inverteerbaar* genoemd als er een systeem bestaat met transfer functie $H_L(z)$ zo dat $H(z)$ links vermenigvuldigd met $H_L(z)$ een vertraging van L stappen levert. In een vergelijking geeft dat

$$H_L(z)H(z) = z^{-L}I_m.$$

Het systeem met transfer functie $H_L(z)$ wordt een *L-vertraagde linker inverse* van \mathcal{S} genoemd. Merk op dat dit inverse systeem inderdaad de ingang van \mathcal{S} reconstrueert met L tijdstappen vertraging.

Sain en Massey [115] toonden aan dat \mathcal{S} *L-vertraagd links inverteerbaar* is als en slechts als

$$\text{rang}(\mathcal{H}_L) - \text{rang}(\mathcal{H}_{L-1}) = m,$$

waarbij

$$\mathcal{H}_L := \begin{bmatrix} D & 0 & 0 & \cdots & 0 \\ CB & D & 0 & \cdots & 0 \\ CAB & CB & D & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{L-1}B & CA^{L-2}B & CA^{L-3}B & \cdots & D \end{bmatrix},$$

voor $L \geq 0$ en $\text{rang}(\mathcal{H}_{-1}) := 0$. Sain en Massey beschouwden linker inverses die bestaan uit twee delen. Het eerste deel is een bank van vertragingselementen die de L vorige waarden van de systeemuitgang opslaat. Op het moment dat we $y_{[k+L]}$ aanleggen aan de bank, geeft deze dus als uitgang $y_{[k:k+L]} := [y_{[k]}^\top y_{[k+1]}^\top \cdots y_{[k+L]}^\top]^\top$. Het tweede deel bestaat uit een dynamisch systeem.

In dit doctoraat beschouwen we linker inverses met dezelfde structuur. Een belangrijke bijdrage is de afleiding van een algemene vorm van het dynamische systeem. Deze algemene vorm wordt gegeven door

$$\mathcal{S}_L^- : \begin{cases} \hat{x}_{[k+1]} = (A - \mathcal{K}_L \mathcal{O}_L) \hat{x}_{[k]} + \mathcal{K}_L y_{[k:k+L]} \\ \hat{u}_{[k]} = -\mathcal{M}_L \mathcal{O}_L \hat{x}_{[k]} + \mathcal{M}_L y_{[k:k+L]}, \end{cases}$$

waarbij $\mathcal{O}_L := [C^\top (CA)^\top \cdots (CA^L)^\top]^\top$ en waarbij \mathcal{K}_L en \mathcal{M}_L gegeven zijn door

$$\begin{aligned} \mathcal{K}_L &= [B \ 0] \mathcal{H}_L^{(1)} + Z_L \Sigma_L \\ \mathcal{M}_L &= [I \ 0] \mathcal{H}_L^{(1)} + U_L \Sigma_L \end{aligned}$$

met $\Sigma_L := I - \mathcal{H}_L \mathcal{H}_L^{(1)}$ en Z_L en U_L matrices die vrij gekozen kunnen worden. Stel dat de initiële toestand $x_{[0]}$ van \mathcal{S} gekend is. Als we dan $\hat{x}_{[0]}$ gelijk nemen

aan $x_{[0]}$, geldt er voor elke keuze van Z_L en U_L dat $\hat{x}_{[k]} = x_{[k]}$ en $\hat{u}_{[k]} = u_{[k]}$ voor alle k .

Als de initiële toestand $x_{[0]}$ van \mathcal{S} niet gekend is (wat meestal het geval is), wensen we dat alle eigenwaarden van $A - \mathcal{K}_L \mathcal{O}_L$ binnen de eenheidscirkel liggen. In dat geval convergeert de schattingsfout immers naar nul. Merk op dat de keuze van Z_L de eigenwaarden van $A - \mathcal{K}_L \mathcal{O}_L$ bepaalt. Een belangrijk resultaat van dit doctoraat is de afleiding van methodes en voorwaarden waaronder Z_L gekozen kan worden zodat de eigenwaarden van $A - \mathcal{K}_L \mathcal{O}_L$ geplaatst zijn op een gewenste positie. Meer bepaald tonen we aan dat de eigenwaarden van $A - \mathcal{K}_L \mathcal{O}_L$ (en dus ook die van het inverse systeem) kunnen geplaatst worden als \mathcal{S} geen onstabiele nullen heeft. Dit resulteert dan in een asymptotische schatter (zie ook Figuur 0.1).

Tot slot wordt er op basis van de theorie van Luenberger [92] een methode ontwikkeld om de orde van \mathcal{S}_L^- te reduceren en tegelijkertijd de polen van de gereduceerde orde schatter te plaatsen. Ook wordt een voorwaarde afgeleid waaronder de ingang $u_{[k]}$ rechtstreeks uit $y_{[k:k+L]}$ berekend kan worden.

Hoofdstuk 4: Inversie van Gecombineerde Deterministische - Stochastische Systemen

In dit hoofdstuk breiden we de inversie procedure van Hoofdstuk 3 uit naar systemen met een stochastische component. In analogie met het Kalman filter, tonen we aan dat we een optimale schatter op twee manieren kunnen afleiden: ten eerste door de polen van de schatter uit Hoofdstuk 3 optimaal te plaatsen en ten tweede aan de hand van KK schatting (zie ook Figuur 0.1). De belangrijkste resultaten van deze twee methodes worden nu besproken.

Onvertkend schatten met minimale variantie

Beschouw opnieuw het systeem \mathcal{S} waarbij de ruistermen w en v stochastisch verondersteld worden. Naar analogie met het Kalman filter, is het idee om een gezamenlijke toestandsschatter en ingangsschatter te ontwikkelen die een optimale afweging maakt tussen de snelheid van convergentie en de gevoeligheid aan ruis. Dit idee wordt in het doctoraat uitgewerkt in verschillende stappen.

In de eerste stap beschouwen we het meest eenvoudige geval, dat is het geval $L = 0$, waarin de toestand en de ingang geschat worden zonder vertraging. We beschouwen een schatter die de vorm van \mathcal{S}_0^- neemt, maar met tijdsvariante winst-matrices $\mathcal{K}_{0[k]}$ en $\mathcal{M}_{0[k]}$ en bepalen deze winst-matrices zodanig dat de verwachte gekwadrateerde fouten $\mathbb{E}[\|x_{[k+1]} - \hat{x}_{[k+1|k]}\|^2]$ en $\mathbb{E}[\|u_{[k]} - \hat{u}_{[k|k]}\|^2]$ geminimaliseerd worden onder beperkingen die onvertkende schattingen leveren. Een belangrijk resultaat is dat we aantonen dat de schatter kan geschreven worden in de volgende vorm,

$$\begin{aligned}\hat{x}_{[k|k]} &= \hat{x}_{[k|k-1]} + L_{[k]}(y_{[k]} - C\hat{x}_{[k|k-1]} - D\hat{u}_{[k|k]}) \\ \hat{x}_{[k+1|k]} &= A\hat{x}_{[k|k]} + B\hat{u}_{[k|k]},\end{aligned}$$

waarbij $\hat{u}_{[k|k]}$ staat voor een optimale schatting van $u_{[k]}$ uit kennis van y tot op tijdstip k en waarbij $AL_{[k]} = \mathcal{K}_{0[k]}$. Deze optimale schatting wordt bekomen uit $y_{[k]} - C\hat{x}_{[k|k-1]}$ met behulp van KK schatting en kan geschreven worden als $\hat{u}_{[k|k]} = \mathcal{M}_{0[k]}(y_{[k]} - C\hat{x}_{[k|k-1]})$. We bekomen dus een schatter waarvan de vorm analoog is aan die van het Kalman filter, met als enige verschil dat de echte waarde van de ingang vervangen is door een optimale schatting.

In een tweede stap beschouwen we het geval $L = 1$ en leiden analoge resultaten af.

In een laatste stap ontwikkelen we een algemeen raamwerk voor willekeurige L dat alle bestaande methodes voor het gezamenlijk schatten van toestanden en ingangen generaliseert. We beschouwen hier een schatter die de vorm van \mathcal{S}_L^- neemt, maar waarbij de vrije parameters $Z_{L[k]}$ en $U_{L[k]}$ nu tijdsvariant zijn en bepalen deze parameters zodanig dat de verwachte gekwadraterde fouten $\mathbb{E}[\|x_{[k+1]} - \hat{x}_{[k+1|k]}\|^2]$ en $\mathbb{E}[\|u_{[k]} - \hat{u}_{[k|k]}\|^2]$ geminimaliseerd worden. Merk op dat deze aanpak verschillend is van diegene die we hierboven voor $L = 0$ en $L = 1$ beschouwd hebben.

Tot slot leiden we in dit hoofdstuk een eenvoudige methode af om een systeem te *ontkoppelen* van ongekeerde ingangen. Het concept van ingangsontkoppeling biedt een rigoureuze aanpak tot het ontwerp van optimale toestandsschatters voor systemen met ongekeerde ingangen [26, 67–70]. Het idee achter ingangsontkoppeling is om de toestandsvergelijking van het systeem te transformeren zodat een equivalente toestandsvergelijking bekomen wordt die ontkoppeld is van de ongekeerde ingang. Door ook de uitgangsvergelijking te ontkoppelen van de ongekeerde ingang, kunnen standaard technieken zoals het Kalman filter toegepast worden om de toestand van het systeem te schatten. Bestaande methodes voor ingangsontkoppeling zijn vrij complex en beperkt tot het geval $L = 0$. In dit doctoraat wordt een eenvoudige procedure ontwikkeld die geldt voor willekeurige L . Het ontkoppelde systeem dat afgeleid wordt in dit doctoraat is van de vorm

$$\begin{aligned} x_{[k+1]} &= (A - \mathcal{K}_L \mathcal{O}_L)x_{[k]} + \mathcal{K}_L y_{[k:k+L]} + \bar{w}_{[k:k+L-1]} \\ \Sigma_L y_{[k:k+L]} &= \Sigma_L \mathcal{O}_L x_{[k]} + \bar{v}_{[k:k+L]}, \end{aligned}$$

waarbij $\bar{w}_{[k:k+L-1]}$ en $\bar{v}_{[k:k+L]}$ ruistermen zijn waarvan de exacte uitdrukking niet van belang is voor deze discussie. Door $y_{[k:k+L]}$ in de toestandsvergelijking te beschouwen als ingang, kunnen nu de standaard technieken zoals het Kalman filter toegepast worden om de toestand te schatten.

Kleinste-kwadraten schatting

We beschouwen hier enkel het geval $L = 0$. Het geval $L = 1$ kan op een analoge manier behandeld worden. Beschouw het KK probleem

$$\min_{\substack{x_{[0]}, \dots, x_{[k+1]} \\ u_{[0]}, \dots, u_{[k]}}} \|x_{[0]} - \hat{x}_{[0|-1]}\|_{P_{[0|-1]}^{-1}}^2 + \sum_{i=0}^k \|v_{[i]}\|_{R^{-1}}^2 + \sum_{i=0}^k \|w_{[i]}\|_{Q^{-1}}^2$$

waarin $P_{[0|-1]}$, R en Q gewichtsmatrices zijn en met als beperkingen de systeemvergelijkingen van \mathcal{S} . Het grote verschil met het KK probleem dat aanleiding geeft tot de Kalman filter vergelijkingen, is dat de ingangen nu ongekend zijn en we bijgevolg ook optimaliseren over de ingangen. In dit doctoraat wordt aangetoond dat het hoger beschouwde KK probleem een oplossing heeft als en slechts als $\text{rang}(D) = m$ en dat de oplossing gegeven wordt door de hoger beschouwde schatter die de vorm heeft van het Kalman filter, maar waar de echte waarde van de ingang vervangen is door een optimale schatting.

Net als bij het Kalman filter, is er in deze afleiding geen stochastische veronderstelling nodig over de ruistermen w en v . Een belangrijk resultaat is ook dat we (voor het geval $L = 1$) op basis van deze afleiding vergelijkingen in informatie vorm hebben ontwikkeld, alsook een numeriek betrouwbare vierkantwortel implementatie.

Hoofdstuk 5: Toepassingen van Systeem Inversie

In dit hoofdstuk worden vier toepassingen van systeem inversie behandeld.

Filteren met ruizige ingangen en uitgangen

De eerste toepassing beschouwt het filtering probleem met ruizige ingangen en uitgangen. Dit probleem werd eerst bestudeerd in [62], waar het *errors-in-variables* filtering genoemd werd. De behandeling in [62] is echter beperkt tot systemen met één ingang en één uitgang en houdt geen verband met het Kalman filter. Het geval met meerdere ingangen en uitgangen is eerst beschouwd in [94], waar er wordt aangetoond dat het probleem vertaald kan worden naar het klassieke Kalman filtering probleem. Een gelijkaardig resultaat werd bekomen in [31, 93].

In dit doctoraat wordt een uitbreiding van het filtering probleem met ruizige ingangen en uitgangen beschouwd. We behandelen het geval waarin er een lineaire combinatie van de ingangsvector gemeten wordt in plaats van de volledige ingangsvector. We tonen aan dat het resulterende probleem kan geherformuleerd worden als een inversie probleem en leiden filter vergelijkingen af waarin de schatting van de ongekende ingang en de toestand verbonden zijn. We tonen aan dat de vergelijkingen equivalent zijn aan deze aan in [31, 93] als de volledige ingangsvector gemeten wordt.

Filteren in de aanwezigheid van vertekening

In verschillende toepassingen is het numerieke model onderhevig aan additieve fouten waarvan de eigenlijke waarde ongekend is, maar waarvan de dynamische vergelijkingen gekend zijn. Dergelijke fouten worden *vertekeningsfouten* genoemd. De meest voorkomende types vertekeningfouten zijn de constante vertekeningfouten, die voortvloeien uit ongekende parameters.

Het probleem van optimaal filteren in de aanwezigheid van vertekening heeft veel aandacht gekregen in het verleden. Een optimale oplossing bestaat erin om de toestandsvector uit te breiden met de vector van ongekende vertekeningfouten en dan beide te schatten met behulp van een Kalman filter. In 1969 stelde Friedland [45] het *twee-traps* filter voor waarin de schatting van de toestand en de vertekening afzonderlijk verloopt en de resultaten slechts op het einde samengevoegd worden. Een zorgvuldige studie van het twee-traps filter kan gevonden worden in [3, 29, 30, 75].

In dit doctoraat leiden we een nieuw filter af door het model dat de dynamica van de vertekeningfout beschrijft, te integreren in de gezamenlijke ingangs- en toestandsschatter die werd afgeleid in Hoofdstuk 4. We tonen aan dat onze aanpak een belangrijke voordeel heeft ten opzichte van het twee-traps filter. Dit voordeel is dat het filter zowel overweg kan met totaal ongekende ingangen als met vertekeningfouten waarvan de dynamica gekend is en tijdens werking kan overschakelen van de ene vorm naar de andere. Zoals aangetoond in een numeriek voorbeeld, is zo'n overschakeling nuttig als de vertekeningfout constant is gedurende een bepaald tijdsinterval en dan plots een abrupte en ongekende sprong maakt.

Schatten en verbeteren van modelfouten

Zowel modellen opgesteld aan de hand van fysische wetten als modellen geïdentificeerd uit data, zijn benaderend. In fysische modellen, zijn fouten onder andere te wijten aan ongemodelleerde dynamica en incorrecte waarden van parameters. In empirische modellen, zijn fouten te wijten aan de keuze van een ongeschikte modelklasse of slechte data. We beschouwen in dit doctoraat een lineair toestandsruimte-model dat is afgeleid op basis van fysische wetten. We veronderstellen dat dit model onderhevig is aan ongemodelleerde dynamica die niet verwaarloosbaar is, zodat er een correctie van het model nodig is.

Fysische modellen met niet verwaarloosbare fouten worden ook beschouwd in [108]. Een methode wordt voorgesteld waarin een correctie-model geplaatst wordt in parallel, serie of terugkoppeling met het bestaande model en technieken worden uitgewerkt om het correctie-model te identificeren op basis van data van de ingangen en de uitgangen van het systeem. Deze methode heeft echter als nadeel dat het correctie-model gewoonlijk complexer is dan het fysische model.

In dit doctoraat is een methode uitgewerkt waarin we rechtstreeks de foutieve dynamica corrigeren in plaats van foutieve dynamica te corrigeren door in te werken op de ingang en uitgang zoals in [108]. Onze methode bestaat uit twee stappen. In de eerste stap modelleren we de fout als een ongekende ingang en schatten we deze gezamenlijk met de toestand door gebruik te maken van de technieken uit Hoofdstuk 4. In de tweede stap identificeren we met de aldus bekomen data de ongekende dynamica met behulp van een deelruimte-identificatie algoritme [105, 106]. We tonen met een numeriek voorbeeld aan dat deze methode correctie modellen levert die van lagere orde zijn dan deze in [108].

Schatten van ongekende randvoorwaarden

De schatting van ongekende randvoorwaarden wordt intensief bestudeerd in inverse warmtegeleidingsproblemen. In [19, 138] wordt er verondersteld dat de functionele vorm van de randvoorwaarde in ruimte en tijd gekend is. De ongekende parameters in de functionele vorm worden dan geschat met behulp van KK schatting. Een uitbreiding naar het gezamenlijk schatten van de randvoorwaarden en de initiële toestand, kan gevonden worden in [74]. Benaderingen waarin een Kalman filter gebruikt wordt, zijn terug te vinden in [101, 121]. De toepasbaarheid van deze methodes is echter beperkt door de veronderstelling dat de functionele vorm van de randvoorwaarde in ruimte en tijd gekend is.

In dit doctoraat wordt een methode ontwikkeld die geen veronderstelling maakt over de functionele vorm van de randvoorwaarde in de tijd. Wat betreft de functionele vorm in de ruimte, wordt er verondersteld dat deze geschreven kan worden als een lineaire combinatie van een beperkt aantal basisfuncties. Het probleem wordt op deze manier herleid tot het schatten van de tijdsafhankelijke coëfficiënten in de lineaire combinatie. Vermits deze coëfficiënten optreden als ongekende ingangen, kunnen we de technieken uit Hoofdstuk 4 gebruiken om ze te schatten. Een voorbeeld dat de warmtegeleiding in een tweedimensionale plaat beschrijft waarvan de voorwaarden op één van de vier randen ongekend zijn, toont de doeltreffendheid van deze techniek aan.

Deel II: Data Assimilatie

Hoofdstuk 6: Suboptimaal Vierkantwortel Filteren

Alhoewel het Kalman filter door zijn eenvoudige recursieve structuur zeer aantrekkelijk is voor data-assimilatie, is het niet rechtstreeks toepasbaar. Toepassing van het Kalman filter wordt voornamelijk belemmerd door de hoge computationele kost en de immense opslagcapaciteit die nodig is om de foutencovariantiematrix te propagieren.

Met het oog op grootschalige schattingsproblemen, werden verschillende benaderingen van het Kalman filter ontwikkeld. We noemen dergelijke benaderingen *suboptimale* filters. Veel gebruikte suboptimale technieken zijn het *ensemble Kalman filter* [13, 38, 39, 71, 85] dat gebaseerd is op een Monte Carlo aanpak, het *gereduceerde-rang vierkantwortel filter* [134, 135] dat gebaseerd is op een optimale lagere-rang benadering van de foutencovariantiematrix, en *variationele data assimilatie* [23, 88], een techniek die gebaseerd is op de KK interpretatie van het Kalman filter en die momenteel de standaard is in het Europees Centrum voor Weersverwachtingen op Middellange Termijn (ECMWF). In dit hoofdstuk beschouwen we echter enkel het gereduceerde-rang vierkantwortel filter.

Gereduceerde-rang vierkantswortel filteren

Het idee achter het gereduceerde-rang vierkantswortel filter is om de foutencovariantiematrix $P \in \mathbb{R}^{n \times n}$ te benaderen als

$$P \approx SS^T,$$

waarbij $S \in \mathbb{R}^{n \times q}$ (met $q \ll n$) zodanig gekozen is dat SS^T een optimale rang q benadering is van P . De Kalman filter vergelijkingen worden dan herschreven in functie van deze vierkantswortel S . Zo een benadering heeft twee voordelen. Ten eerste wordt de rekencomplexiteit en het geheugenverbruik sterk gereduceerd. Ten tweede zorgt de propagatie van dergelijke vierkantswortel ervoor dat de benaderende foutencovariantiematrix altijd positief semi-definiet is.

Het algoritme van het gereduceerde-rang vierkantswortel filter bestaat uit drie stappen [134]: een tijdstap, een reductiestap en een meetstap.

- **Tijdstap:** Gedurende de tijdstap neemt het aantal kolommen van S , of equivalent de rang van de foutencovariantiematrix, toe als gevolg van de covariantiematrix van de procesruis.
- **Reductiestap:** De verhoging in rang tijdens de tijdstap, kan de berekeningstijd snel doen toenemen. Daarom wordt SS^T optimaal benaderd door een matrix van lagere rang. Dit kan op een efficiënte manier door de partiële eigenwaarden ontbinding van SS^T te berekenen uit die van de veel kleinere matrix $S^T S$ [134].
- **Meetstap:** Wat betreft de meetstap zijn er verschillende implementaties mogelijk. We beschouwen twee mogelijke manieren om de metingen te verwerken.
 - Gelijktijdige verwerking: Bij een gelijktijdige verwerking van de metingen, worden alle metingen tesamen verwerkt. Een verwerking van dit type is het meest efficiënt als er veel metingen zijn [5, 13].
 - Opeenvolgende verwerking: Bij een opeenvolgende verwerking van de metingen, worden de metingen na elkaar en apart verwerkt. Een verwerking van dit type is het meest efficiënt als het aantal metingen beperkt is. Zoals aangetoond door Potter [111], kan de verwerking geïmplementeerd worden zonder matrix-matrix vermenigvuldigingen, maar enkel met efficiënte matrix-vector vermenigvuldigingen.

Ruimtelijk gelokaliseerd filteren

Houtekamer en Mitchell [71] merkten op dat de meetstap in het ensemble Kalman filter kan verbeterd worden door enkel de waarde van roosterzellen aan te passen die dicht bij de meetlocatie gelegen zijn. Ze ontdekten dat dit deels te wijten is aan de lagere-rang benadering van de foutencovariantiematrix, die valselijk hoge correlaties introduceert tussen roosterzellen die ruimtelijk verwijderd zijn van elkaar. Sindsdien hebben veel onderzoekers geëxperimenteerd

met technieken die de informatie in de foutencovariantiematrix ruimtelijk lokaliseren [9, 63, 72]. In dit doctoraat bouwen we verder op de techniek in [9], waar (in de veronderstelling dat $D = 0$) een meetstap van de vorm

$$\hat{x}_{[k|k]} = \hat{x}_{[k|k-1]} + \Psi_{[k]} L_{[k]} (y_{[k]} - C_{[k]} \hat{x}_{[k|k-1]})$$

beschouwd wordt. Het verschil met de meetstap in het Kalman filter zit in de aanwezigheid van de *lokalisatiematrix* $\Psi_{[k]}$ die gekozen kan worden zodanig dat het effect van de meting gelokaliseerd wordt in de ruimte.

Gereduceerde-rang vierkantswortel filteren met ruimtelijke lokalisatie

Een belangrijke bijdrage is de ontwikkeling van een gereduceerde-rang vierkantswortel filter dat gebruik maakt van het principe van ruimtelijke lokalisatie. Met de toepassing in ruimteweer (waar zoals eerder besproken het aantal metingen zeer beperkt is) in het achterhoofd, wordt een meetstap beschouwd die de metingen opeenvolgend verwerkt. Dergelijke meetstap heeft buiten efficiëntie nog een bijkomend voordeel, namelijk dat we de lokalisatiematrix verschillend kunnen kiezen voor elk van de metingen.

In dit doctoraat worden twee verschillende implementaties van het resulterende filter uitgewerkt. De eerste implementatie is het meest algemeen, in die zin dat deze steeds geldig is. De tweede implementatie buit een specifieke structuur van de foutencovariantiematrix uit en is daardoor efficiënter dan de eerste implementatie. Er wordt meer bepaald verondersteld dat de correlatie tussen twee roosterzellen waarvoor de ruimtelijke afstand groter is dan een bepaalde drempel, nul is. Deze implementatie is erg efficiënt, temeer omdat ze, net zoals het algoritme van Potter, enkel gebruik maakt van matrix-vector vermenigvuldigingen. Een simulatievoorbeeld op het chaotisch Lorenz model met 40 roosterzellen [90] toont de doeltreffendheid van deze methode aan.

Hoofdstuk 7: Simulatievoorbeeld in ruimteweer

Alhoewel data assimilatie bijna dagelijks uitgevoerd wordt in weerkunde, is de toepassing in ruimteweer zo goed als onbestaande. Dit is enerzijds te wijten aan het feit dat plasma-astrofysica een relatief jong onderzoeksdomein is en anderzijds aan het beperkte aantal metingen in vergelijking met weerkunde. Deze schaarsheid aan metingen maakt data assimilatie voor toepassingen in ruimteweer erg uitdagend en vereist de ontwikkeling van nieuwe technieken die aangepast zijn aan deze situatie. Enkele preliminaire studies die het effect van de schaarsheid aan metingen bestuderen en aanpakken, kunnen gevonden worden in [11, 17, 122].

De belangrijkste bijdrage van dit hoofdstuk bestaat in de toepassing van het in Hoofdstuk 6 ontwikkelde filter op een grootschalig en vrij realistisch model. We beschouwen het boegschok model in [27] en zetten een simulatie op die de geschiktheid van het filter beoordeelt.

Magnetohydrodynamica

De macroscopische stroming van een plasma wordt beschreven door *magnetohydrodynamica* (MHD). De interactie van een plasma met magnetische en elektrische velden zorgt ervoor dat de MHD vergelijkingen complexer zijn dan de hydrodynamische vergelijkingen die de stroming van een neutraal fluïdum beschrijven. De MHD vergelijkingen zijn een combinatie van de Navier-Stokes vergelijkingen en de vergelijkingen van Maxwell. Ze zijn bijgevolg erg niet-lineair.

Analytische oplossingen van de MHD vergelijkingen zijn beperkt tot de meest eenvoudige gevallen en zelfs dan moeten er vaak benaderingen gemaakt worden. Numerieke simulaties daarentegen, laten toe om de meest complexe plasma dynamica te bestuderen. Als gevolg hiervan, zijn er verschillende numerieke codes ontwikkeld. De simulaties in dit doctoraat worden uitgevoerd met de Versatile Advection Code (VAC) [130].

Net als bij hydrodynamische stromingen, kunnen er in MHD stromingen schokken optreden. Daar waar in hydrodynamica slechts één type schok is, laat MHD drie types schokken toe die onderscheiden kunnen worden door de manier waarop ze de magnetische veldlijnen breken. Numerieke simulaties die een tweedimensionale plasma stroming rondheen een cilinder modelleren [27, 28, 119], hebben aangetoond dat de topologie van MHD schokken sterk afhankelijk is van de eigenschappen van de inkomende stroming. Als de stroming gedomineerd wordt door drukeffecten, wordt er één enkel schokfront waargenomen dat gelijkaardig is aan het schokfront in hydrodynamica. Als de stroming daarentegen gedomineerd wordt door magnetische effecten, worden er verschillende opeenvolgende schokfronten waargenomen van een verschillend type.

De simulaties in [27, 28, 119] leveren interessante inzichten in de topologie van de aardse boegschok. In realiteit is de boegschok echter driedimensionaal en heeft de aarde ook een eigen magnetisch veld, wat de topologie nog complexer maakt. Hoe dan ook leveren de simulaties in [27] een eenvoudig tweedimensionaal model dat de belangrijkste karakteristieken van de boegschok beschrijft.

Opzet van de experimenten en resultaten

De simulaties die uitgevoerd worden in dit doctoraat volgen de numerieke opzet in [27]. We beschouwen een cilindrisch rooster dat bestaat uit 124×124 cellen. Elke cel bevat 6 variabelen, wat resulteert in een toestandsvector van dimensie 92256. Alle MHD simulaties worden uitgevoerd met de VAC. De rest van de code is geïmplementeerd in Matlab. Dit leidt helaas tot een constante conversie van data formaten tussen de VAC en Matlab. Een gedeelte van de simulaties werd uitgevoerd op de K.U.Leuven VIC cluster [1].

We maken gebruik van de procedure van *tweelingsexperimenten*, wat wil zeggen dat we eerst ruizige data genereren met behulp van de VAC code en nadien deze data assimileren in een tweede simulatie die start vanuit een foutieve

begintoestand. Er worden twee reeksen van experimenten uitgevoerd. In de eerste reeks veronderstellen we dat de eigenschappen van het inkomende plasma gekend zijn en gaan we het effect na van de rang van de foutencovariantiematrix, het aantal metingen en het type van ruimtelijke lokalisatie. In de tweede reeks experimenten wordt er verondersteld dat de eigenschappen van het inkomende plasma ongekend zijn en bovendien veranderen van druk-gedomineerd naar magnetisch-gedomineerd. Het filter wordt uitgebreid om de randvoorwaarden mee te schatten.

De simulatieresultaten van de eerste reeks experimenten tonen aan dat met slechts 4 metingen een significante reductie van de schattingsfout kan bekomen worden ten opzichte van een data-vrije simulatie. Verder geven de resultaten aan dat ruimtelijke lokalisatie van de metingen een positief effect heeft op de schattingsfout. Uit de resultaten van de tweede reeks experimenten kan besloten worden dat het filter robuust is tegen veranderingen in de randvoorwaarden. De simulaties geven echter aan dat een goede specificatie van de randvoorwaarden cruciaal is.

We besluiten uit deze experimenten dat data assimilatie technieken een groot potentieel hebben in ruimteweer toepassingen. Het valt echter af te wachten hoe de technieken presteren met echte data en met meer realistische modellen die bijvoorbeeld ook het magnetische veld van de aarde in rekening brengen.

Hoofdstuk 8: Besluit

Dit hoofdstuk vat de belangrijkste resultaten van dit proefschrift samen. We beschouwen achtereenvolgens de resultaten in systeem inversie en data assimilatie.

Systeem inversie

- Op basis van schattingstheorie werd een nieuwe inversie procedure afgeleid. De procedure levert in het ruis-vrije geval een exacte reconstructie van zowel de ingang als de toestand van het systeem. Voorwaarden werden afgeleid waaronder de polen van het inverse systeem geplaatst kunnen worden en de snelheid van convergentie dus geregeld kan worden.
- In geval van ruis werd er aangetoond dat de polen zodanig geplaatst kunnen worden dat de schattingen optimaal zijn volgens het criterium van de kleinste-kwadraten.
- Verschillende numerieke problemen werden aangepakt, zoals een reductie in rekencomplexiteit en de ontwikkeling van numeriek hoogstaande implementaties.

Data assimilatie

- Een nieuw suboptimaal Kalman filter werd ontwikkeld dat uitermate geschikt is voor toepassingen waarin slechts een beperkt aantal metingen

beschikbaar is. Het filter vertaalt de schaarsheid aan metingen in numerieke efficiëntie en is robuust tegen de problemen die kunnen optreden als gevolg van het beperkte aantal metingen.

- Het suboptimale filter werd succesvol toegepast in een grootschalige simulatie (ongeveer 10^5 te schatten variabelen) die de dynamica van de boegschok modelleert die gevormd wordt als de supersonische zonnwind de aarde passeert. Simulatieresultaten tonen aan dat het suboptimale filter een significante reductie in de schattingsfout levert, zelfs als er slechts metingen van vier satellieten beschikbaar zijn.

Chapter 1

Introduction

From the earliest times, people have been concerned with estimating unknown quantities. Due to astronomical observations, Egyptian and Chinese astronomers determined the period of alternation of the lunar phases within several minutes of accuracy already in the fifth century before Christ [37].

The first attempt to the development of an estimation theory is due to Galileo Galilei in the beginning of the seventeenth century. He tried to systematize the minimization of various functions of the estimation error [65,80]. Perhaps the most important estimation theory is that of *least-squares* (LS) estimation, connected to such important names as Legendre and Gauss. The breakthrough of LS estimation came in 1801. Based on a limited number of observations, it was the only method that could recover the position of the asteroid Ceres after it had disappeared from sight for almost one year. However, it was reported by Gauss that LS estimation has one major disadvantage: its inability to take the laws governing the dynamics of Ceres into account [117].

In 1960, shortly after the introduction of linear state-space models, Kalman [83] derived a recursive estimation procedure that takes both observations and knowledge of system dynamics into account. It is well established by now that Kalman's estimation procedure, called the *Kalman filter*, recursively solves an LS problem. Since its introduction, the Kalman filter has gained increasing popularity. It was first successfully used in the Ranger, Mariner, and Apollo space missions. In particular, it guided the Apollo 11 lunar module to the Moon's surface in 1961 [21]. Nowadays, the number of applications involving the Kalman filter is almost uncountable. Variants of the Kalman filter are used in airplanes, chemical plants, the Global Positioning System, econometrics, weather prediction, and many other areas.

The popularity of the Kalman filter is due to its simple recursive structure depicted in Fig. 1.1. Consider a dynamical system that is driven by a known input and responds to this input by producing a certain output. The latter is assumed to be measured at regular time instants. In order to estimate the state of such a system, the Kalman filter starts from an initial estimate of the system state and while new measurements become available, it consecutively applies

the following two steps. The first step, the time update, propagates the present state estimate ahead in time using a numerical model of the system. The second step, the measurement update, updates the propagated state estimate based on the newly available measurement.

1.1 Motivation and objectives

Despite its enormous success, extensions or approximations of the Kalman filter are needed in most applications. In this thesis, two estimation problems are considered in which the Kalman filter is not directly applicable.

- The first estimation problem, called *data assimilation*, extends the Kalman filter for use with large-scale numerical models. Motivating examples and objectives for this estimation problem, are given in Sect. 1.1.1.
- The second estimation problem, called *system inversion*, addresses the case in which the system input is unknown. Motivating examples and objectives for this estimation problem, are given in Sect. 1.1.2.

1.1.1 Data assimilation

The recursive two-steps structure of the Kalman filter is very appealing for real-time environmental state estimation problems such as weather nowcasting. Indeed, in such problems both ingredients of the Kalman filter, a numerical models and measurements, are available. The numerical models are usually obtained by discretizing partial differential equations (PDE's) over a huge spatial grid. The resulting number of grid cell ranges from 10^4 in tidal flow forecasting [135] to as much as 10^6 in weather forecasting or ocean circulation prediction [85]. In addition, the numerical models are usually highly nonlinear. The challenging problem of assimilating observations in such large-scale numerical models has been termed *data assimilation*.

Due to complexity of the numerical models, direct application of the Kalman filter in data assimilation is not feasible. For a numerical model consisting of 10^6 grid cells, for example, the Kalman filter would require approximately eight terabytes of computer memory (using eight bytes per number), that is approximately the total amount of information in a large university library. The computational cost would then approximately be the time of 10^6 model simulations. For a simulation of one minute, this comes down to almost two years of computation.

It is clear that approximations of the Kalman filter are needed in data assimilation. Such approximate filters have been termed *suboptimal* Kalman filters. Suboptimal filters have been successfully used in such areas as weather prediction [112], tidal flow forecasting [135], ocean circulation prediction [85], ozone prediction [36], and the estimation of ecosystems [2]. Results show that suboptimal filters at an expense of only a few hundred model simulations

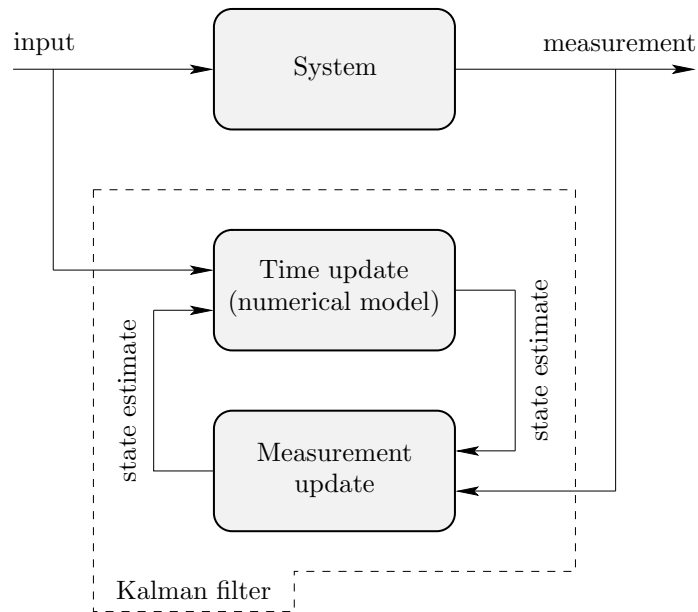


Figure 1.1: *Recursive estimation procedure of the Kalman filter. Consider a dynamical system that is driven by a known input and responds to this input by producing a certain output. The latter is assumed to be measured at regular time instants. In order to estimate the state of such a system, the Kalman filter starts from an initial estimate of the system state and while new measurements become available, it consecutively applies the following two steps. The first step, the time update, propagates the present state estimate ahead in time using a numerical model of the system. The second step, the measurement update, updates the propagated state estimate based on the newly available measurement.*

perform significantly better than a data-free simulation, even for a large-scale numerical model consisting of 10^6 grid cells.

In this thesis, suboptimal Kalman filtering techniques are developed for *space weather* nowcasting. As will be discussed in the next section, the complexity of the numerical models and the sparseness of observations make the development of suboptimal Kalman filters for space weather nowcasting even more challenging than for the applications considered above.

1.1.1.1 Space weather

Around 1930, scientists had discovered that the temperature in the *corona*, the outer atmosphere of the Sun, is approximately one million degrees Celsius. A surprisingly high temperature, knowing that the temperature at the surface

of the Sun is only a few thousand degrees. Due to these extremely high temperatures, all particles in the corona are in the *plasma* state, a state of matter that can be considered as a gas consisting of positively and negatively charged particles. Parker [109] showed that the Sun's corona is so hot and so energetic that the plasma must escape into space with supersonic velocity. He termed this constant stream of plasma emerging from the Sun the *solar wind*.

The speed of the solar wind is not constant. It varies over the position on the Sun (see Fig. 1.2) and over time. The average speed is 500 km/s, but speed varies between 250 km/s in periods of solar minimum to as much as 2500 km/s when the wind is due to a *coronal mass ejection* (see Fig. 1.3) or a *flare*, the most energetic solar eruptions. Due to such eruptions, the speed of the solar wind speed can vary significantly within time scales as short as seconds.

Just like an airplane that breaks the sound barrier, the supersonic solar wind forms a *bow shock* when it encounters an obstacle such as the Earth. Across the shock there is an extremely rapid change in the properties of the solar wind. Like the wind on Earth, the properties of the solar wind can be characterized by a speed, density, and pressure. However, due to the charged plasma particles, the solar wind also drags a magnetic and electric field with it. When passing the Earth, these fields start interacting with the magnetic field of the Earth. As shown in Fig. 1.4, the Earth's magnetic field is compressed at the day-side and expanded at the night-side.

The fluctuations in the speed of the solar wind can strongly perturb the magnetic environment of the Earth. The effects resulting from such perturbations are referred to as *space weather*. Severe space weather can affect human technology. Energy and radiation from flares and CME's can harm astronauts in space, damage sensitive electronics in satellites, disrupt long range communication and navigation systems, create power blackouts and may even have an effect on climate and on biological systems.

In order to observe weather conditions in space, several satellites have been launched. The ACE, SOHO and WIND spacecrafts are located in front of the bow shock. The CLUSTER II, DOUBLE STAR, GEOTAIL and INTERBALL spacecrafts study the solar wind closer to Earth. Some of these spacecrafts cross the Earth's bow shock on a regular basis.

During the last decades, scientists have also developed numerical models that simulate the flow of astrophysical plasmas [61]. The performance of such numerical models has been tested by comparing model simulation to data recorded from spacecrafts. In such simulations, it is usually observed that small errors in the specification of the initial conditions and boundary conditions at the Sun can lead to significant simulation errors at Earth. This motivates the use of data assimilation techniques. A schematic overview of a state estimator for space weather nowcasting, is shown in Fig. 1.5.

1.1.1.2 Challenges and objectives

Two issues make the development of data assimilation techniques for space weather nowcasting even more challenging than for other applications.

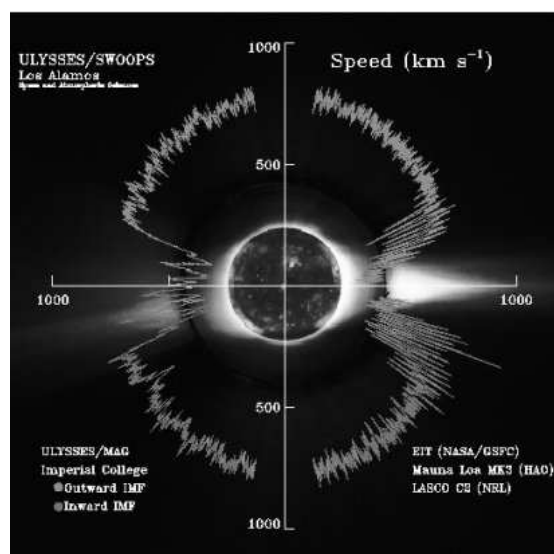


Figure 1.2: Variability in the speed of the solar wind. The speed varies over the position on the Sun and over time. The average speed is 500 km/s, but speed varies between 250 km/s in periods of solar minimum to as much as 2500 km/s when the wind is due to a coronal mass ejection (see Fig. 1.3) or a flare, the most energetic solar eruptions. (Courtesy of SOHO consortium. SOHO is a project of international cooperation between ESA and NASA.)

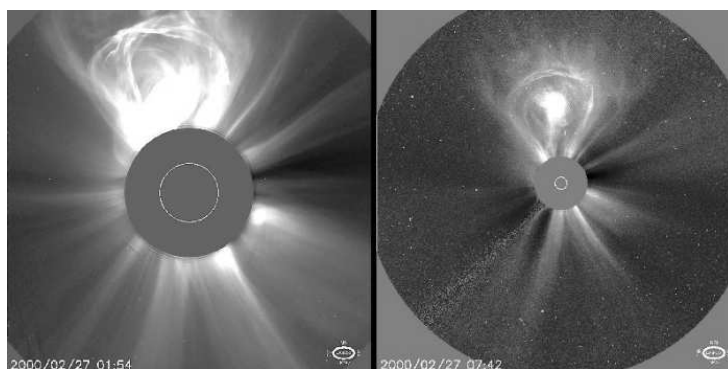


Figure 1.3: A coronal mass ejection (CME), one of the most energetic solar eruptions. The discs in the middle of the figures cover the Sun, so that the coronal mass ejection can be seen more clearly. Left figure: initiation of a CME. Right figure: propagation of the CME through interplanetary space. (Courtesy of SOHO consortium. SOHO is a project of international cooperation between ESA and NASA.)

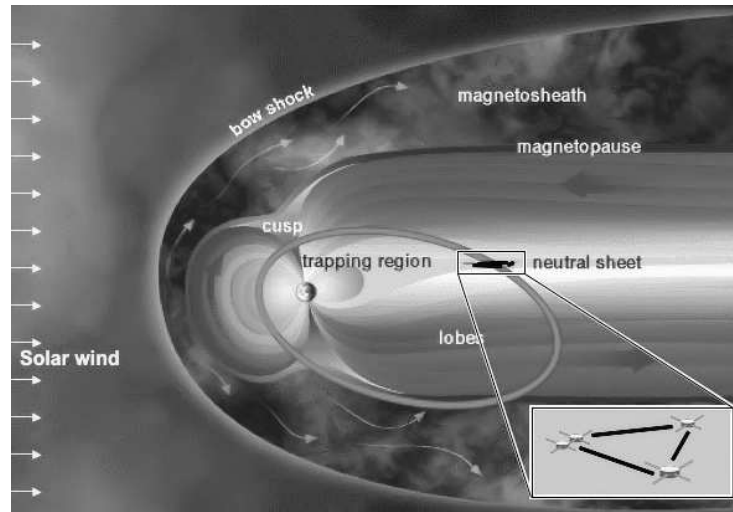


Figure 1.4: *Interaction of the solar wind with the magnetic field of the Earth. Just like an airplane that breaks the sound barrier, the supersonic solar wind forms a bow shock when it encounters an obstacle such as the Earth. Across the shock there is an extremely rapid change in the properties of the solar wind. The Earth's magnetic field is compressed at the day-side and expanded at the night-side due to interaction with the charged plasma particles in the solar wind. The ellipse around the Earth represents the orbit of the four CLUSTER II satellites which observe the weather conditions in space. (Figure taken from <http://clusterlaunch.esa.int/>)*

- The sparseness of observations: Contrary to Earthly weather observation, where almost continuously in the order of 10^5 data points are available, the number of satellites that observe space weather is very limited. In addition, the length scales are enormous, in the order of millions of kilometers.
- The complexity of the numerical models: The presence of charged particles makes plasma behavior more complex than neutral gas behavior. The equations that describe the behavior of a plasma are a combination of the Navier-Stokes equations of fluid dynamics and Maxwell's equations of electromagnetism. They are consequently rich in nonlinearities.

As shown in Fig. 1.5, the objective of this thesis is to develop an advanced suboptimal Kalman filter that is specifically adapted to the data sparse environment of space weather. The filter must translate the sparseness of measurements into numerical efficiency and must be robust against the problems that may arise due to data sparseness. In other words, it must get as much information as possible out of each measurement.

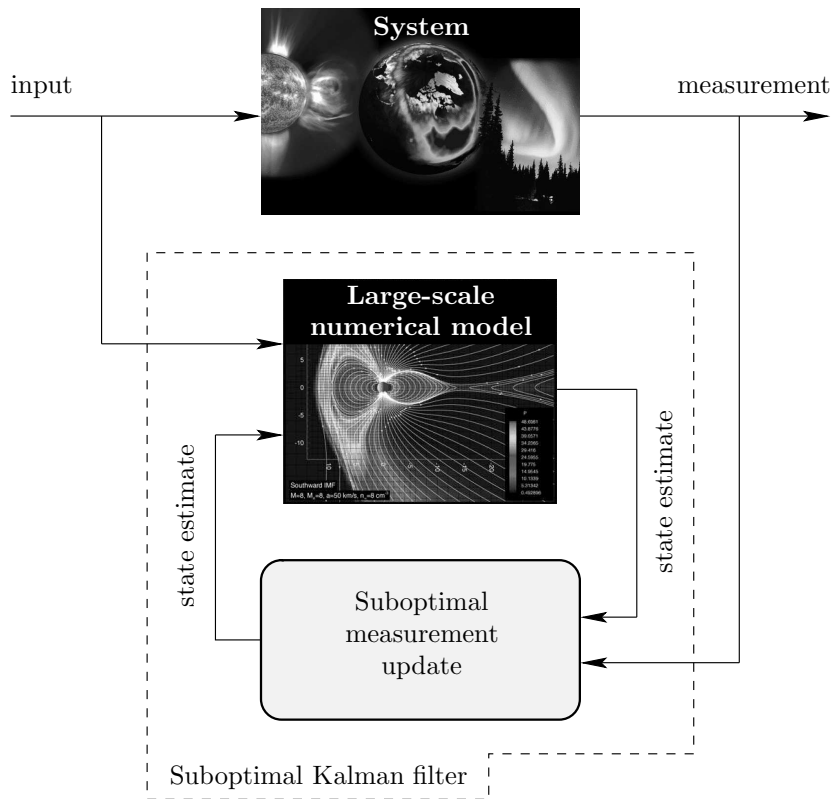


Figure 1.5: Schematic overview of a state estimator for space weather nowcasting. Due to the complexity of the numerical models (typically between 10^5 and 10^6 state variables), direct application of the Kalman filter in space weather nowcasting is not feasible. In this thesis, an advanced suboptimal Kalman filter is developed that is adapted to the data sparse environment of space weather.

1.1.2 System inversion

The Kalman filter is based on the assumption that the system input is known exactly. Such an assumption is valid in control applications, where the input applied to the system is generated by a known control law, but can be too restrictive in other applications. Indeed, in a lot of applications, the system is subject to inputs of which the value is unknown. The second estimation problem considered in this thesis, called *system inversion*, deals with estimating such unknown inputs. Three examples that motivate the study of system inversion are now given.

1.1.2.1 Motivating examples

- **Fault detection:** For certain systems like e.g. airplanes, chemical plants and mechanical robots, there is a potential for faults or disturbances that may cause severe injury or property damage. The detection of faults is therefore a prime concern in such systems. As shown in Fig. 1.6, this thesis considers an actuator fault in an F16 aircraft. More precisely, it will be assumed that the actuator which drives the elevator fails a certain time instant. Actuator faults are typically hard to detect and may have severe consequences. Indeed, the controller that steers the actuator is not aware of the fault and will typically try to correct the movement of the aircraft by steering the faulty actuator. There is thus a need to detect and estimate such faults. Since faults can be modeled as an unknown input, fault estimation can be performed by means of system inversion.
- **Estimation of model errors:** All numerical models are just approximations of the true system. In physical modelling, errors and inaccuracies are due to unknown dynamics, incorrect parameter values, rough approximations, ... However, in most cases also measurements are available that yield information about the underlying system dynamics. This thesis addresses the use of such measurements in the estimation of model errors. Just as faults, model errors can be seen as unknown inputs.
- **Estimation of unknown boundary conditions:** Consider again the space weather example of Fig. 1.4. In order to simulate the effect of space weather on the magnetic field of the Earth, the boundary conditions, i.e. the properties of the incoming solar wind, need to be known. In most environmental simulations, however, the boundary conditions are unknown and thus have to be specified conveniently or have to be estimated.

1.1.2.2 Challenges and objectives

In all of the examples above, there is a need for estimating an unknown input from knowledge of the system output. As shown in Fig. 1.6, the estimation of unknown inputs boils down to the development of an estimator that has as input the output of the system and as output the input the system. The inputs and outputs of the estimator are thus *inverted* in comparison to those of the system. Hence, the name system *inversion*.

The first inversion techniques were introduced at the end of the sixties [16, 115, 116]. However, existing inversion techniques are limited to the ideal situation of noise-free systems. The objective of this thesis is to develop new inversion techniques that extend to systems subject to noise. As shown in Fig. 1.6, the inverse systems considered in this thesis consist of two parts. The first part, the state estimator, yields an estimate of the system state. The second part, the input estimator, uses the state estimate to produce an estimate of the unknown input.

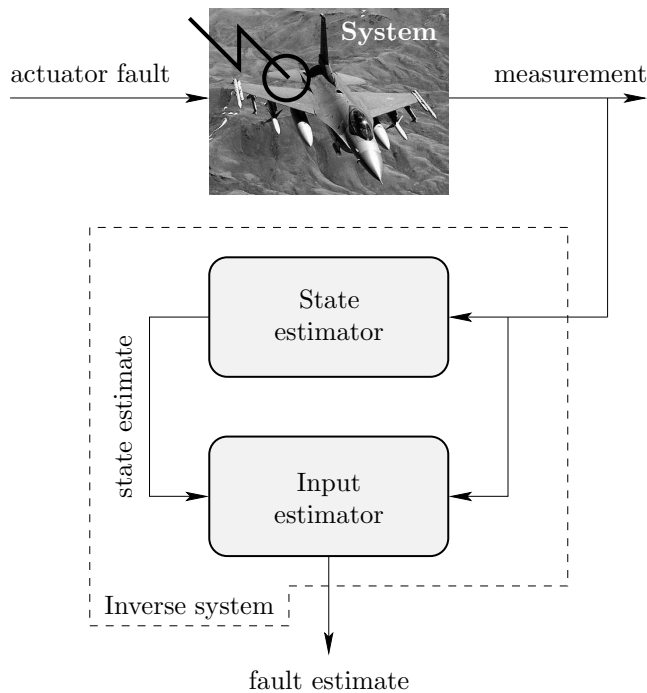


Figure 1.6: Example of the use of system inversion in fault detection. Consider a fault in the actuator that steers the elevator of an F16 aircraft. Actuator faults are typically hard to detect and may have severe consequences. Indeed, the controller that steers the actuator is not aware of the fault and will typically try to correct the aircraft movement by steering the faulty actuator. The detection of actuator faults is therefore a prime concern. Since faults can be modeled as unknown inputs, fault estimation can be performed by means of system inversion. The inverse systems considered in this thesis consist of two parts. The first part, the state estimator, yields an estimate of the system state. The second part, the input estimator, uses the state estimate to produce an estimate of the fault.

1.2 Chapter-by-chapter overview

Figure 1.7 shows the outline of this thesis. The main body of the text is divided in two parts. Part I deals with system inversion, Part II with data assimilation. The two parts stand apart, meaning that e.g. Part II can be read before Part I. Both parts build further on Chapter 2. A chapter-by-chapter overview of this thesis is now given.

Chapter 2 provides a brief introduction to Kalman filtering. After a formulation of the filtering problem, the Kalman filter is introduced as a

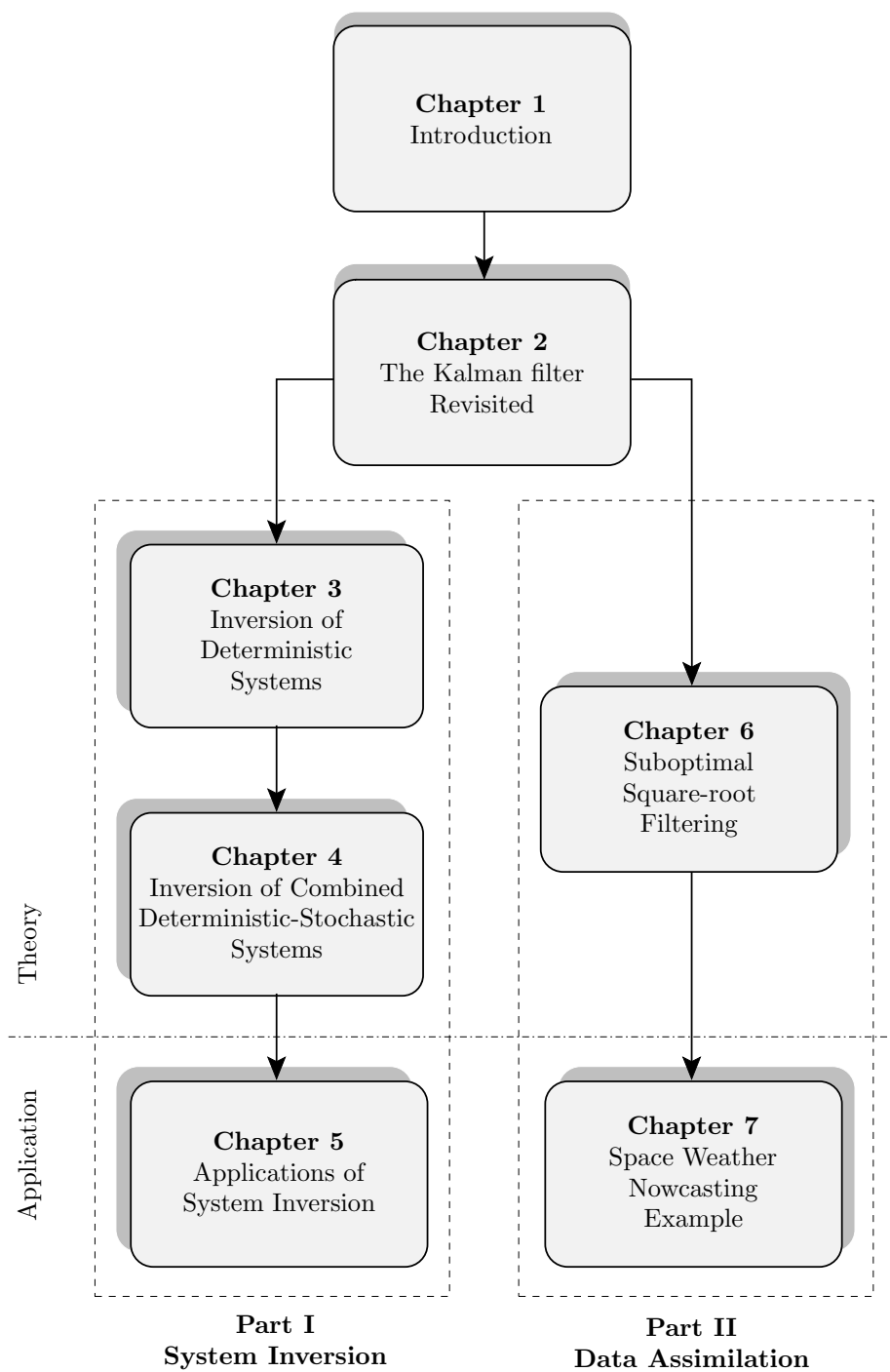


Figure 1.7: Outline of this thesis.

recursive state estimator, optimal in the minimum-variance unbiased (MVU) sense. Its relation to LS estimation is discussed afterwards. Two alternative forms of the traditional equations are considered, the information form and the square-root form, and their numerical advantages are discussed. Finally, the extended Kalman filter, a nonlinear extension of the Kalman filter, is briefly considered.

Part I: System Inversion

The first part of this thesis deals with system inversion. The exposition of this part runs fairly parallel to the exposition in Chapter 2. Figure 1.8 yields an overview of the most important concepts and methods concerning estimation theory considered in this thesis. The arrows denote the relations between the concepts and methods. The numbers denote the chapters and paragraphs in which these relations are studied. Numbers between parentheses refer to paragraphs dealing with Kalman filtering, the other numbers to paragraphs dealing with system inversion.

- **Chapter 3** addresses left inversion of linear discrete-time deterministic systems in state-space form. The problems of left and right inversion are first defined and conditions are derived under which a linear state-space system is left invertible. Next, the state of the art in system inversion is briefly discussed. We mainly consider the inversion approach of Sain and Massey [115] and compare this approach to that of Silverman [116]. Next, a new approach to left inversion based on joint input-state estimation is introduced. Conditions and methods are derived under which the poles of the inverse system can be assigned. Based on the theory of reduced order observers, a technique is developed to simultaneously reduce the order of the inverse system and place its poles. Several numerical examples illustrate the new approach.

Publications related to this chapter: [56].

- **Chapter 4** extends the inversion procedure of Chapter 3 to combined deterministic-stochastic systems, where the aim is to optimally reconstruct the deterministic input from knowledge of the noisy outputs. First, the filtering problem is considered. Filters are developed in which the estimation of the system state and the unknown input are interconnected. An important contribution is the establishment of a relation between the joint input-state estimators and (recursive) LS estimation. Based on this relation, information and square-root information formulas are derived almost instantaneously. Next, a general framework for the one step ahead prediction, the filtering and the smoothing problem is derived that covers both state estimation and joint input-state estimation.

Publications related to this chapter: [49, 52, 55–57].

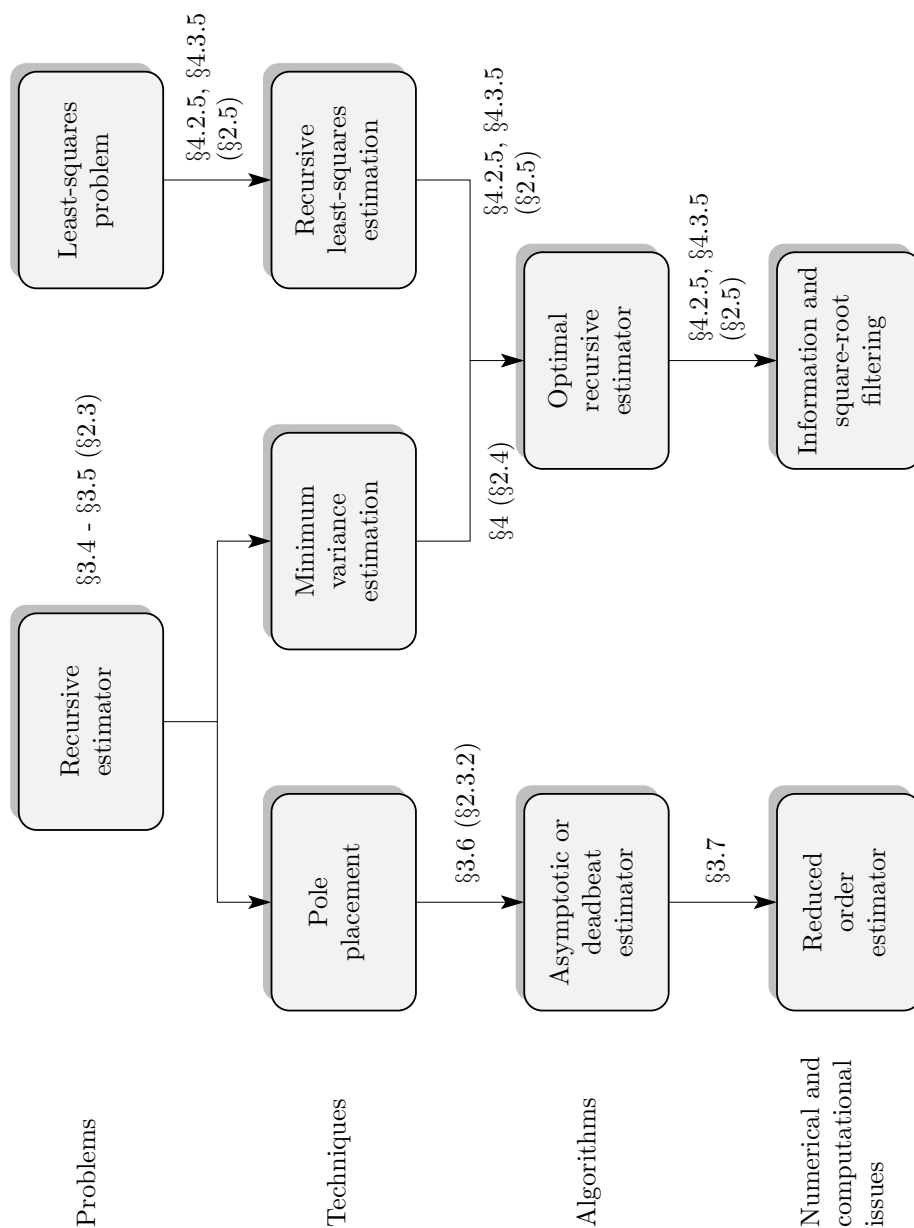


Figure 1.8: Overview of the concepts and techniques in estimation theory considered in this thesis. The arrows denote the relations between the concepts and techniques. The numbers denote the chapters and paragraphs in which these relations are studied. Numbers between parentheses refer to paragraphs dealing with Kalman filtering, the other numbers to paragraphs dealing with system inversion.

- **Chapter 5** considers four applications of system inversion. First, a new solution to the errors-in-variables filtering problem is derived in which the estimation of the system state and the input are interconnected. Next, the problem of filtering in the presence of bias is considered. A suboptimal filter, closely related to the two-stage Kalman filter [45], is developed. The last two applications are more practical. First, model error estimation and dynamic model updating is addressed. An empirical technique is outlined to correct a physical model for unknown dynamics. Finally, an approach to joint state and boundary condition estimation is considered in which the spatial component of the boundary condition is expanded as a linear combination of orthogonal basis functions.

Publications related to this chapter: [50, 51, 53].

Part II: Data Assimilation

- **Chapter 6** addresses the challenging problem of data assimilation. First, a brief overview of the most commonly used suboptimal Kalman filtering techniques is given. Next, the idea of suboptimal square-root filtering is introduced and two procedures are described to process the measurements: sequential processing and simultaneous processing. Two extensions of the reduced rank square-root (RRSQRT) filter [135] are developed in this chapter. The first extension speeds-up the RRSQRT filter by interweaving the so-called reduction step into the measurement update. The second extension addresses the problem of reduced rank spatially localized square-root (RRSLSQRT) filtering, where the objective is to update only the subset of the grid cells that is effectively correlated to the measurements. The performance of the extensions is assessed in two numerical examples.

Publications related to this chapter: [48, 54].

- **Chapter 7** considers the application of data assimilation techniques for nowcasting a space weather event. First, the magnetohydrodynamic (MHD) equations are introduced and the different types of shocks that occur in MHD are discussed. Next, the RRSLSQRT is applied with a large-scale numerical MHD model, consisting of approximately 10^5 state variables, which emulates the dynamics of the bow shock that is formed when the solar wind encounters the Earth. The performance of the RRSLSQRT is investigated for different types of spatial localization and different values of the rank of the approximate error covariance matrix. Simulations with both known constant and unknown time-varying boundary conditions are considered.

Publications related to this chapter: [48].

1.3 Personal contributions

This section summarizes the personal contributions of this thesis.

1.3.1 System inversion

First, we consider the contributions in system inversion. The most important contributions are development of a new inversion procedure for deterministic system on the one hand, and the extension of this procedure to combined deterministic-stochastic systems on the other hand.

- Based on estimation theory, a general form of a time-delay left inverse system is derived in Sect. 3.5. Like the approach of Sain and Massey [115], the left inverses considered in this thesis consist of a bank of delay elements followed by a dynamical system. In this thesis, the most general form of such a dynamical system is derived. This dynamical system reconstructs both the system input and the system state and can thus be considered as a joint input-state estimator. The general form consists of two matrix parameters which can be free chosen. In Sect. 3.6, we derive methods and conditions under which the poles of the estimator can be assigned by the choice of these parameters. These conditions generalize earlier results [7, 100].

Publications related to this topic: [56].

- An important contribution is the extension of the inversion procedure to systems subject to both unknown inputs and noise, considered in Chapter 4. Optimal recursive state estimators for such systems have been extensively studied in literature. An important contribution of this thesis is the extension to joint input-state estimation. In particular, it is shown that the poles of the joint-input state estimator considered above can be assigned so that the estimates are optimal in the minimum-variance unbiased sense.

Publications related to this topic: [49, 57].

- In Sects. 4.2.5 and 4.3.5, the relation between the joint input-state estimator and LS estimation is established. More precisely, it is shown that the joint input-state estimator can be derived by recursively solving an LS problem. It is shown that the equations of the joint input-state estimators can be split into a time update and a measurement update. In particular, in Sect. 4.2, a state estimator is derived in which the time update and the measurement update take the form of that in the Kalman filter, except that the true value of the input is replaced by an optimal estimate.

Publications related to this topic: [52, 55].

- Considering state estimation in the presence of unknown inputs, a new and straightforward procedure to decouple a system from unknown inputs is developed in Sect. 4.4.1.1. The procedure generalizes existing results [68, 70].

Publications related to this topic: [56].

- Several computational and numerical issues are addressed. In Sect. 3.7, a novel procedure is developed to reduce the order of the input estimator and simultaneously place its poles. The problem of information and square-root filtering, which was not yet considered in literature in the context of system inversion, is addressed in Sects. 4.2.5, 4.3.5 and 4.3.6.

Publications related to this topic: [52, 55, 56].

- The inversion procedure is applied in four applications. In Sect. 5.2, a new solution to the errors-in-variables filtering problem is derived. In Sect. 5.3, a new solution to the optimal filtering problem in the presence of bias errors is derived. Section 5.4 outlines a novel procedure for model updating. Finally, in Sect. 5.5, a new approach to the estimation of unknown boundary conditions is considered.

Publications related to this topic: [50, 51, 53].

1.3.2 Data assimilation

The most important contributions in data assimilation are the adaptation of the RRSQRT filter to the sparseness of measurements in space weather on the one hand, and the application of the resulting suboptimal filter in a space weather simulation on the other hand.

- The sparseness of measurements is dealt with by a combination of two techniques. The first technique uses the algorithm of Potter [111] to translate the sparseness of measurements to numerical efficiency. The second technique is based on the spatially localized Kalman filter [9] and addresses the observability problem in data assimilation. An important contribution of this thesis is the incorporation of both techniques in the RRSQRT filter, considered in Sect. 6.5.2. The resulting suboptimal filter, called the reduced rank spatially localized Kalman (RRSLSQRT) filter, is well suited for large-scale applications in which only few measurements are available.

Publications related to this topic: [48, 53, 54].

- In Chapter 7, the RRSLSQRT filter is successfully applied in a large-scale simulation (approximately 10^5 state variables) which emulates the dynamics of the bow shock under various conditions of the solar wind. Simulation results indicate that the suboptimal filter yields a significant reduction in estimation error over a data-free simulation, even if measurements of only 4 satellites are available. However, it remains to be seen how the method performs with more complex and realistic models.

Publications related to this topic: [48].

Chapter 2

The Kalman Filter Revisited

This chapter provides a brief introduction to Kalman filtering. After a formulation of the filtering problem, the Kalman filter is introduced as a recursive state estimator, optimal in the minimum-variance unbiased sense. Its relation to least-squares estimation is discussed afterwards. Two alternative forms of the traditional equations are considered, the information form and the square-root form, and their numerical advantages are discussed. Finally, the extended Kalman filter, a nonlinear extension of the Kalman filter, is briefly considered. This chapter contains no personal contributions.

2.1 Introduction

The number of books and papers dealing with Kalman filtering is almost uncountable. The objective of the present chapter is neither to give a detailed and rigorous overview of literature on the Kalman filter, nor to provide a full theoretical study of the technique, but rather to introduce those ingredients of Kalman filtering that will be used in the remainder of this thesis. For a deeper theoretical treatment, we refer the reader to e.g. [4, 47, 96]. It is assumed throughout this chapter that the reader is familiar with probability theory, stochastic processes, linear state-space models and LS estimation. An introduction to probability theory and stochastic processes in the context of filtering can be found in [77]. For an introduction to linear state-space models, we refer the reader to [81] or [113]. A brief introduction to LS estimation can be found in Appendix B.

Chapter outline

This chapter is outlined as follows. Section 2.2 defines the filtering, prediction and smoothing problems for linear dynamical systems. Next, in Sect. 2.3, the concepts of observability and detectability are introduced and their significance in the design of recursive filters for noise-free systems is discussed. The Kalman filter is introduced in Sect. 2.4 as a recursive filter for linear systems subject to noise, optimal in the MVU sense. Its relation to LS estimation is discussed in Sect. 2.5, where also information formulas for the Kalman filter are derived. Section 2.6 discusses the numerical advantages of square-root filtering. Finally, in Sect. 2.7, the extended Kalman filter, an extension of the Kalman filter to nonlinear systems, is considered.

2.2 Filtering, prediction and smoothing

This section defines the state estimation problem for dynamical systems, and more particularly the filtering, prediction and smoothing problems. State estimation for a dynamical system requires that a numerical model is available which describes the dynamics of the system. We are concerned with linear time-invariant (LTI) discrete-time models described by the state-space equations

$$x_{[k+1]} = Ax_{[k]} + w_{[k]} \quad (2.1a)$$

$$y_{[k]} = Cx_{[k]} + v_{[k]}, \quad (2.1b)$$

where $x_{[k]} \in \mathbb{R}^n$ denotes the *state vector* at the discrete time k and $y_{[k]} \in \mathbb{R}^p$ denotes the *output vector* at time k . The *state equation* (2.1a) usually follows from the physical laws that govern the dynamics of the system, such as e.g. the laws of mechanics, thermodynamics or electricity, or from black box identification [106]. The *output equation* (2.1b) models the relation between the state vector and the actual measurements of the system. The *noise vectors* $w_{[k]} \in \mathbb{R}^n$ and $v_{[k]} \in \mathbb{R}^p$ account for the errors introduced in the modeling procedure and are assumed to be unknown.

Without loss of generality, we assume that the initial time at which the model (2.1) commences equals 0, so that (2.1) holds for $k \geq 0$. We denote the sequence $\{x_{[0]}, x_{[1]}, \dots, x_{[N]}\}$ with $N \geq 0$, by $\{x_{[k]}\}_{k=0}^N$.

Without loss of generality, we assume that there exists an initial state $x_{[0]}$ and realizations of the noise processes $\{w_{[k]}\}_{k=0}^{\infty}$ and $\{v_{[k]}\}_{k=0}^{\infty}$ such that for all $k \geq 0$, $x_{[k]}$ equals the state of the true system at time instant k and $y_{[k]}$ equals the measurement at time instant k . Although a system should actually not be given a mathematical description, we will usually refer to (2.1) as the “system”. The word “system” should in this context be interpreted as a model that we consider to give a perfect mathematical description of the system. In the same context, we refer to $x_{[k]}$ as the system state at time instant k .

We are now in place to define the filtering, prediction and smoothing problems.

Definition 2.1 (Estimation, Smoothing, Filtering, Prediction). *Given a realization of the output process of (2.1), that is, given a sequence of measurements $Y_{[l]} := \{y_{[k]}\}_{k=0}^l$, the state estimation problem consists in computing an estimate of the system state $x_{[k]}$ based on $Y_{[l]}$. If $k < l$, the estimation problem is called a smoothing problem. If $k = l$, it is called a filtering problem. And if $k > l$, it is called a prediction problem.*

If $k = l + 1$, we talk about a *one step ahead prediction* problem. In the remainder, we denote an estimate of $x_{[k]}$ given $Y_{[l]}$ by $\hat{x}_{[k|l]}$.

The difference between filtering, smoothing and prediction is schematically shown in Fig. 2.1. The filtering and the prediction problems are usually employed in real-time operations, where the estimates are based on measurements up to the present time instant. In the smoothing problem, a time delay between the receipt of the last measurement and the production of the estimates is allowed and the measurements that come available during that delay are used in the estimation procedure.

2.3 Recursive estimation for noise-free systems

Let us start by considering the most simple state estimation problem, that is, the estimation problem for the noise-free system

$$x_{[k+1]} = Ax_{[k]} \quad (2.2a)$$

$$y_{[k]} = Cx_{[k]}, \quad (2.2b)$$

where $x_{[k]} \in \mathbb{R}^n$ denotes the system state at time instant k and $y_{[k]} \in \mathbb{R}^p$ denotes the measurement at time k . The system matrices A and C are assumed to be known. The initial state $x_{[0]}$, on the other hand, is assumed to be unknown.

This section is outlined as follows. In Sect. 2.3.1, conditions are derived under which the state sequence $\{x_{[k]}\}_{k=0}^{\infty}$ of system (2.2) can be reconstructed from knowledge of the sequence of measurements $\{y_{[k]}\}_{k=0}^{\infty}$. Next, in Sect. 2.3.2, recursive state estimators for system (2.2) are derived.

2.3.1 Observability and detectability

The determination of conditions under which $\{x_{[k]}\}_{k=0}^{\infty}$ can be reconstructed from knowledge of $\{y_{[k]}\}_{k=0}^{\infty}$, has led to the concepts of observability and detectability.

2.3.1.1 Observability

We define observability both in terms of the system (2.2) and in terms of the pair of matrices $\{A, C\}$.

Definition 2.2. *The system (2.2) is said to be observable if there exists a number $N \geq 0$ such that given $\{y_{[k]}\}_{k=0}^N$, $x_{[.]}$ can be deduced.*

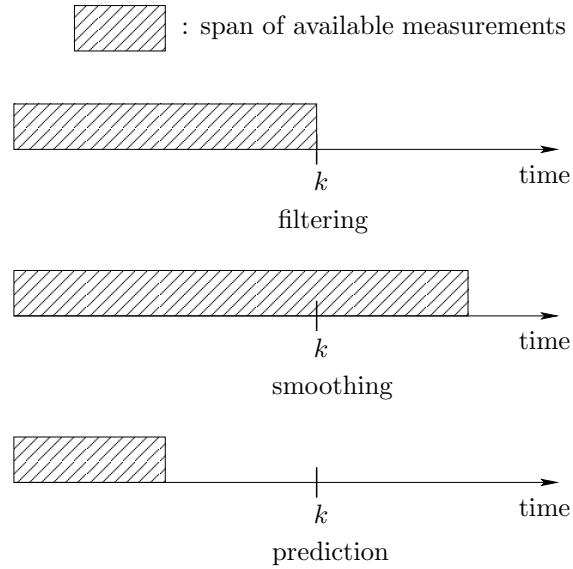


Figure 2.1: *Filtering, smoothing and prediction: the index k denotes the time at which the state vector is to be estimated.*

Let $N \geq 0$ and define

$$\mathcal{O}_N := \begin{bmatrix} C \\ CA \\ \vdots \\ CA^N \end{bmatrix}. \quad (2.3)$$

Then, it is well known that observability of (2.2) can be checked by the rank of the so-called *observability matrix* \mathcal{O}_{n-1} of (2.2).

Theorem 2.1 ([12]). *Rank* $(\mathcal{O}_{n-1}) = n$ if and only if (2.2) is observable.

Proof: Defining $y_{[0:N]} := [y_{[0]}^\top \ y_{[1]}^\top \ \dots \ y_{[N]}^\top]^\top$, it follows from (2.2) that $y_{[0:N]} = \mathcal{O}_N x_{[0]}$. The initial state $x_{[0]}$ can be uniquely determined from the latter equation if and only if $\text{rank}(\mathcal{O}_N) = n$. It follows from the Cayley-Hamilton theorem that $\text{rank}(\mathcal{O}_N) = \text{rank}(\mathcal{O}_{n-1})$ for all $N \geq n-1$, which concludes the proof. ■

Instead of defining observability in terms of the system (2.2), observability is sometimes also defined in terms of the pair of matrices $\{A, C\}$. The reason is that, as we will see in Chapter 3, observability of a matrix pair may have a meaning also if there is no system attached to that matrix pair.

Definition 2.3 ([12]). *Let $\lambda \in \Lambda(A)$, where $\Lambda(A)$ denotes the set of eigenvalues*

of A . Then, λ is said to be an observable mode of $\{A, C\}$ if

$$\text{rank} \left(\begin{bmatrix} \lambda I - A \\ C \end{bmatrix} \right) = n.$$

Otherwise, λ is said to be an unobservable mode of $\{A, C\}$.

Definition 2.4 ([12]). *The pair $\{A, C\}$ is said to be observable if all $\lambda \in \Lambda(A)$ are observable modes of $\{A, C\}$.*

The following proposition yields a relation between observability of the pair $\{A, C\}$ and observability of the system (2.2).

Proposition 2.1 ([12]). *The pair $\{A, C\}$ is observable if and only if the system (2.2) is observable.*

It follows that observability of the pair $\{A, C\}$ can be checked by the rank of \mathcal{O}_{n-1} . Consequently, we also refer to \mathcal{O}_{n-1} as the observability matrix of the pair $\{A, C\}$.

2.3.1.2 Detectability

We define detectability in terms of the matrix pair $\{A, C\}$ and give an interpretation in terms of the system (2.2) afterwards.

Definition 2.5. *The pair $\{A, C\}$ is said to be detectable if all $\lambda \in \Lambda(A)$ with $|\lambda| \geq 1$ are observable modes of $\{A, C\}$.*

Notice that detectability is less strong than observability, that is, if the pair $\{A, C\}$ is observable, it is also detectable, but not vice versa.

In terms of the system (2.2), detectability of the pair $\{A, C\}$ means that, asymptotically, the state vector can be uniquely determined from knowledge of the output, as will be shown in the next section.

2.3.2 Asymptotic and deadbeat estimation

In this section, we consider recursive state estimators for the system (2.2). Such estimators are initialized with an estimate of the initial state of the system. Every time instant a new measurement becomes available, the estimator updates the previous estimate with this measurement according to a pre-specified recursive law. The recursive law is usually designed so that the estimates converge to the actual values.

We consider two types of estimators. In the asymptotic estimation problem, the estimator is designed such that the estimation error converges asymptotically to zero. We will see that detectability of the pair $\{A, C\}$ is necessary for the existence of such an estimator. In the deadbeat estimation problem, the estimator is designed such that the estimation error becomes zero in a finite number of steps. We will see that a necessary condition for the existence of such an estimator is observability of the pair $\{A, C\}$.

2.3.2.1 Asymptotic estimation

Consider a recursive state estimator for the system (2.2) of the form

$$\hat{x}_{[k+1|k]} = A\hat{x}_{[k|k-1]} + K(y_{[k]} - C\hat{x}_{[k|k-1]}), \quad (2.4)$$

initialized with an estimate $\hat{x}_{[0|-1]}$ of the initial state $x_{[0]}$. The term $y_{[k]} - C\hat{x}_{[k|k-1]}$ in (2.4) can be interpreted as the difference between the true and estimated measurement at time instant k , and thus yields information about the error in $\hat{x}_{[k|k-1]}$. The so-called *gain matrix* K is a design parameter. In the asymptotic estimation problem, the gain matrix is determined so that the estimation error $\tilde{x}_{[k|k-1]}$, defined by $\tilde{x}_{[k|k-1]} := x_{[k]} - \hat{x}_{[k|k-1]}$, converges asymptotically to zero for $k \rightarrow \infty$.

We now derive conditions under which such a gain matrix exists. It follows from (2.2) and (2.4) that the estimation error obeys the following recursion,

$$\tilde{x}_{[k+1|k]} = (A - KC)\tilde{x}_{[k|k-1]}. \quad (2.5)$$

Consequently, the estimation error converges asymptotically to zero if the gain matrix K can be chosen so that $|\lambda| < 1$ for all $\lambda \in \Lambda(A - KC)$, i.e. so that all eigenvalues of $A - KC$ lie inside the unit circle. The following theorem yields a relation between the eigenvalues of $A - KC$ and the unobservable modes of the pair $\{A, C\}$.

Theorem 2.2 ([40]). *Let λ be an unobservable mode of the pair $\{A, C\}$. Then $\lambda \in \Lambda(A - KC)$ for all K . In particular, let $\{A, C\}$ have l distinct unobservable modes, then l eigenvalues of $A - KC$ will equal the unobservable modes of $\{A, C\}$, while the other eigenvalues can be assigned by the choice of K .*

The following corollary, which immediately follows from Theorem 2.2, provides conditions under which K can be chosen so that the eigenvalues of $A - KC$ are assigned.

Corollary 2.1.

- (i) *If and only if $\{A, C\}$ is detectable, the gain matrix K can be chosen so that $|\lambda| < 1$ for all $\lambda \in \Lambda(A - KC)$.*
- (ii) *If and only if $\{A, C\}$ is observable, the gain matrix K can be chosen so that all $\lambda \in \Lambda(A - KC)$ can be assigned at any desired location.*

It follows from (2.5) and from Corollary 2.1 that the gain matrix K can be chosen so that $\tilde{x}_{[k|k-1]}$ converges asymptotically to zero for $k \rightarrow \infty$ if and only if $\{A, C\}$ is detectable. The procedure of assigning the eigenvalues of $A - KC$ by the choice of the gain matrix K is called *pole placement*. For pole placement techniques and algorithms, we refer the reader to the specialized literature [40].

2.3.2.2 Deadbeat estimation

The smaller the eigenvalues of $A - KC$, the faster the state estimator (2.4) converges. It follows from Corollary 2.1 that if $\{A, C\}$ is observable, the eigenvalues can be placed at any desired location. In *deadbeat estimation*, the eigenvalues are placed at the origin, yielding convergence to the exact state vector in only a few steps. A deadbeat estimator thus “beats” the error in the initial state estimate to “death” in a finite number of steps.

2.4 The Kalman filter

Deadbeat estimators are attractive because of their extremely rapid convergence. However, it is well-known that deadbeat estimators can be highly sensitive to noise. In case of noise, it is more convenient to place the poles of the estimator so that the estimates satisfy a certain optimality condition. This is formalized in the Kalman filter.

This section assumes knowledge of basic statistical concepts and of stochastic processes. The reader who is not familiar with these concepts is referred to [77].

Consider the LTI discrete-time system

$$x_{[k+1]} = Ax_{[k]} + Bu_{[k]} + Ew_{[k]} \quad (2.6a)$$

$$y_{[k]} = Cx_{[k]} + Du_{[k]} + v_{[k]}, \quad (2.6b)$$

where $x_{[k]} \in \mathbb{R}^n$ denotes the state vector at time instant k , $y_{[k]} \in \mathbb{R}^p$ denotes the measurement at time k , and $u_{[k]} \in \mathbb{R}^m$ denotes the input vector at time k . The system matrices A, B, C, D , and E and the input sequence $\{u_{[k]}\}_{k=0}^{\infty}$ are assumed to be known. For the purpose of pole placement, we assume that the pair $\{A, C\}$ is observable.

We assume that the initial state $x_{[0]}$ is a random variable. The noise processes $\{w_{[k]} \in \mathbb{R}^l\}_{k=0}^{\infty}$ and $\{v_{[k]} \in \mathbb{R}^p\}_{k=0}^{\infty}$ are assumed to be stochastic with the properties given in the following assumption.

Assumption 2.1. *The stochastic noise processes $\{w_{[k]}\}_{k=0}^{\infty}$ and $\{v_{[k]}\}_{k=0}^{\infty}$ are*

- (a) *zero-mean processes with known covariance matrices*
- (b) *stationary processes*
- (c) *mutually uncorrelated processes*
- (d) *white processes, meaning that for any k and l with $k \neq l$, the random vectors $w_{[k]}$ and $w_{[l]}$ are uncorrelated and the random vectors $v_{[k]}$ and $v_{[l]}$ are uncorrelated.*

Notice that Assumptions 2.1 (b) – (d) can be summarized as

$$\mathbb{E} \left\{ \begin{bmatrix} w_{[k]} \\ v_{[k]} \end{bmatrix} \begin{bmatrix} w_{[l]}^{\top} & v_{[l]}^{\top} \end{bmatrix} \right\} = \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \delta_{[k-l]},$$

where Q and R denote the covariance matrices of the noise processes $\{w_{[k]}\}_{k=0}^{\infty}$ and $\{v_{[k]}\}_{k=0}^{\infty}$, respectively and where $\delta_{[k]} := 1$ for $k = 0$ and $\delta_{[k]} := 0$ otherwise. We assume that R is positive definite.

2.4.1 Derivation of the Kalman filter equations

In his original derivation, Kalman used the concept of orthogonal projections [83]. Afterwards, the Kalman filter equations have been rederived using concepts such as maximum likelihood estimation, LS estimation and linear MVU estimation [77]. The relation between the Kalman filter and LS estimation, will be considered in Sect. 2.5. In this section, we consider the most simple derivation, namely that based on MVU estimation.

Kalman considered a recursive state estimator of the form (2.4), except that the gain matrix is now time-varying,

$$\hat{x}_{[k+1|k]} = A\hat{x}_{[k|k-1]} + Bu_{[k]} + K_{[k]}(y_{[k]} - C\hat{x}_{[k|k-1]} - Du_{[k]}). \quad (2.7)$$

It is assumed that an *unbiased* estimate $\hat{x}_{[0|-1]}$ of the initial state $x_{[0]}$ is available, that is, $\mathbb{E}[x_{[0]} - \hat{x}_{[0|-1]}] = 0$. The error $\tilde{x}_{[0|-1]}$ in the initial state estimate $\hat{x}_{[0|-1]}$, defined by $\tilde{x}_{[0|-1]} := x_{[0]} - \hat{x}_{[0|-1]}$, is assumed to be uncorrelated to the noise processes $\{w_{[k]}\}_{k=0}^{\infty}$ and $\{v_{[k]}\}_{k=0}^{\infty}$. Furthermore, it is assumed that the *error covariance matrix* $P_{[0|-1]}$ defined by

$$P_{[0|-1]} := \mathbb{E}[(x_{[0]} - \hat{x}_{[0|-1]})(x_{[0]} - \hat{x}_{[0|-1]})^T],$$

is known.

In the MVU setting of the Kalman filtering problem, the optimal value of the gain matrix $K_{[k]}$ is defined as that value that satisfies the following two conditions:

- (i) The optimal gain matrix $K_{[k]}$ yields an unbiased estimate $\hat{x}_{[k+1|k]}$ of the form (2.7), meaning that $\mathbb{E}[x_{[k+1]} - \hat{x}_{[k+1|k]}] = 0$.
- (ii) The optimal gain matrix $K_{[k]}$ minimizes the mean squared error

$$\mathbb{E}[\|x_{[k+1]} - \hat{x}_{[k+1|k]}\|^2]$$

over all linear unbiased estimates of the form (2.7).

First, we determine the condition that the gain matrix should satisfy in order that the estimator (2.7) is unbiased. Defining the error in $\hat{x}_{[k+1|k]}$ by $\tilde{x}_{[k+1|k]} := x_{[k+1]} - \hat{x}_{[k+1|k]}$, it follows from (2.7) and (2.6) that

$$\tilde{x}_{[k+1|k]} = (A - K_{[k]}C)\tilde{x}_{[k|k-1]} - K_{[k]}v_{[k]} + w_{[k]}. \quad (2.8)$$

Since $\hat{x}_{[0|-1]}$ is assumed to be unbiased, it follows from (2.8) that $\hat{x}_{[k+1|k]}$ is unbiased for all $k \geq 0$ and for all values of the gain matrix $K_{[k]}$.

The quantity $y_{[k]} - C\hat{x}_{[k|k-1]} - Du_{[k]}$ on the basis of which the Kalman filter assimilates the measurements is called the *innovation*. Notice that the innovation can be written as

$$y_{[k]} - C\hat{x}_{[k|k-1]} - Du_{[k]} = C\tilde{x}_{[k|k-1]} + v_{[k]}.$$

It follows from the discussion above that the innovation has expected value zero. The Kalman filter thus assimilates the measurement based on a zero-mean random variable.

Now, we determine the gain matrix that minimizes the mean squared error. It can be shown that minimizing the mean squared error $\mathbb{E}[\|x_{[k+1]} - \hat{x}_{[k+1|k]}\|^2]$ is equivalent to minimizing the trace of the error covariance matrix $P_{[k+1|k]}$, defined by

$$P_{[k+1|k]} := \mathbb{E}[(x_{[k+1]} - \hat{x}_{[k+1|k]})(x_{[k+1]} - \hat{x}_{[k+1|k]})^\top].$$

It follows from (2.8) that the error covariance matrix $P_{[k+1|k]}$ obeys the following recursion,

$$\begin{aligned} P_{[k+1|k]} &= (A - K_{[k]}C)P_{[k|k-1]}(A - K_{[k]}C)^\top + K_{[k]}RK_{[k]}^\top + EQE^\top \\ &= K_{[k]}\tilde{R}_{[k]}K_{[k]}^\top - K_{[k]}CP_{[k|k-1]}A^\top - AP_{[k|k-1]}C^\top K_{[k]}^\top \\ &\quad + AP_{[k|k-1]}A^\top + EQE^\top, \end{aligned} \quad (2.9)$$

where $\tilde{R}_{[k]}$ is defined by

$$\tilde{R}_{[k]} := CP_{[k|k-1]}C^\top + R. \quad (2.10)$$

Notice that $\tilde{R}_{[k]}$ is invertible since R was assumed to be positive definite.

The gain matrix $K_{[k]}$ minimizing the trace of (2.9) is then found by setting the derivative of the trace of (2.9) equal to zero. This yields,

$$K_{[k]} = AP_{[k|k-1]}C^\top \tilde{R}_{[k]}^{-1}. \quad (2.11)$$

Substituting (2.11) in (2.9), yields the following equivalent expressions for the update of the error covariance matrix,

$$\begin{aligned} P_{[k+1|k]} &= AP_{[k|k-1]}A^\top - AP_{[k|k-1]}C^\top \tilde{R}_{[k]}^{-1} CP_{[k|k-1]}A^\top + EQE^\top, \\ &= (A - K_{[k]}C)P_{[k|k-1]}A^\top + EQE^\top. \end{aligned} \quad (2.12)$$

Summarizing, the Kalman filter equations are given by (2.7) and (2.12). Since the latter equations actually yield a recursive procedure to update a one step ahead predicted estimate $\hat{x}_{[k+1|k]}$ and its corresponding error covariance matrix, we refer to these equations as the Kalman filter equations in *prediction form*.

2.4.2 Time and measurement update

Completely analogous the developments in the previous section, it is possible to derive an MVU recursion in terms of a filtered estimate $\hat{x}_{[k|k]}$ and its error covariance matrix $P_{[k|k]}$, defined by

$$P_{[k|k]} := \mathbb{E} [(x_{[k]} - \hat{x}_{[k|k]})(x_{[k]} - \hat{x}_{[k|k]})^T].$$

The resulting equations are called the Kalman filter equations in *filter form*. The equations in filter form are very similar to those in prediction form, see e.g. [4]. Both forms are related through the so-called *measurement update* and *time update*.

Kalman filter

- **Measurement update**

The measurement update expresses the filtered quantities in terms of the one step ahead predicted quantities,

$$\hat{x}_{[k|k]} = \hat{x}_{[k|k-1]} + L_{[k]}(y_{[k]} - C\hat{x}_{[k|k-1]} - Du_{[k]}) \quad (2.13)$$

$$L_{[k]} = P_{[k|k-1]}C^T(CP_{[k|k-1]}C^T + R)^{-1} \quad (2.14)$$

$$P_{[k|k]} = P_{[k|k-1]} - P_{[k|k-1]}C^T(CP_{[k|k-1]}C^T + R)^{-1}CP_{[k|k-1]}. \quad (2.15)$$

Notice that the gain matrix $L_{[k]}$ is related to (2.11) by $K_{[k]} = AL_{[k]}$.

- **Time update**

The time update expresses the one step ahead predicted quantities in terms of the filtered quantities,

$$\hat{x}_{[k+1|k]} = A\hat{x}_{[k|k]} + Bu_{[k]} \quad (2.16)$$

$$P_{[k+1|k]} = AP_{[k|k]}A^T + EQE^T. \quad (2.17)$$

It is easily verified that substituting (2.13)-(2.15) in (2.16)-(2.17), yields the equations in one step ahead prediction form derived in the previous section. Analogously, the equations in filter form can be derived by substituting (2.16)-(2.17) in (2.13)-(2.15).

Due to the split-up in the time update and the measurement update, the Kalman filter can be interpreted as a recursive state estimator that consecutively updates the previously predicted state estimate with the new observation (the measurement update) and then predicts the state at the next time instant based on the model equations (the time update). This is schematically shown in Fig. 2.2.

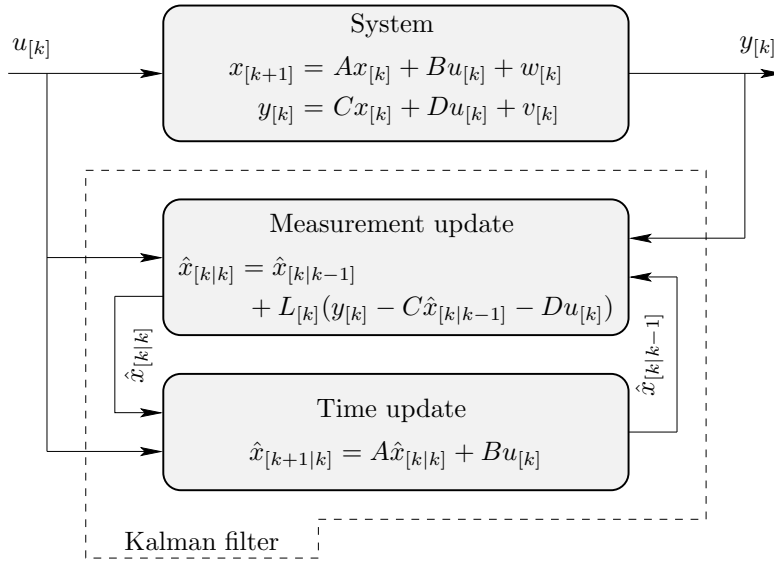


Figure 2.2: Interpretation of the Kalman filter as a recursive state estimator that consecutively updates the previously predicted state estimate with the new observation (measurement update) and then predicts the state at the next time instant based on the model equations (time update).

2.5 Information filtering

Shortly after the introduction of the Kalman filter, alternative implementations of the original formulas appeared. In this section, we consider one such alternative implementation called *information filtering*. Instead of propagating the error covariance matrix, information filters work with its inverse, the *information matrix*. Such an approach is especially useful if no knowledge of the initial state is available ($P_{[0|-1]} = \infty$), since in that case the traditional covariance formulas can not be used.

Information filters were derived by establishing a relation between the Kalman filter and LS estimation. Rigorous derivations of such a relation can be found in [32, 77, 127]. However, the first step towards the establishment of such a relation already dates back to the work of Mowery [99]. Numerically accurate implementations that make use of orthogonal operations can be found in e.g. [97, 107].

This section is outlined as follows. In Sect. 2.5.1, we consider a sequence of growing LS problems that yield smoothed, filtered and one step ahead predicted estimates of the system state and discuss the relation to the Kalman filter. Next, in Sect. 2.5.2, a recursive solution to the sequence of LS problems is derived. Finally, in Sect. 2.5.3, it is shown that by solving this RLS problem analytically, information formulas for the Kalman filter are obtained.

2.5.1 Least-squares state estimation

This section assumes that the reader is familiar with both the stochastic and the deterministic setting of the LS problem. A brief discussion of both settings can be found in Appendix B.

We consider system (2.6) with $E = I$. Contrary to the derivation of the Kalman filter in Sect. 2.4, we do not make any assumption about the properties the initial state $x_{[0]}$ and the noise processes $\{w_{[k]}\}_{k=0}^{\infty}$ and $\{v_{[k]}\}_{k=0}^{\infty}$. For this system, we set-up a sequence of growing LS problems. The LS problem considered at time instant k yields estimates of the state sequence $\{x_{[i]}\}_{i=0}^{k+1}$ based on knowledge of $\{y_{[i]}\}_{i=0}^k$ and $\{u_{[i]}\}_{i=0}^k$. To this aim, the system equations from time instant 0 to time instant k are written into a form that expresses the data (i.e. the known vectors) as a linear combination of the unknowns (i.e. the state sequence) plus noise terms. Appending an equation that summarizes the information in the initial state estimate $\hat{x}_{[0|-1]}$, yields

$$\underbrace{\begin{bmatrix} \hat{x}_{[0|-1]} \\ y_{[0]} - Du_{[0]} \\ -Bu_{[0]} \\ y_{[1]} - Du_{[1]} \\ -Bu_{[1]} \\ \vdots \\ y_{[k]} - Du_{[k]} \\ -Bu_{[k]} \end{bmatrix}}_{\text{data}} = \begin{bmatrix} I \\ C \\ A & -I \\ & C & -I \\ & & \ddots \\ & & & C \\ & & & & A & -I \end{bmatrix} \underbrace{\begin{bmatrix} x_{[0]} \\ x_{[1]} \\ \vdots \\ x_{[k]} \\ x_{[k+1]} \end{bmatrix}}_{\text{unknowns}} + \underbrace{\begin{bmatrix} -\hat{x}_{[0|-1]} \\ v_{[0]} \\ w_{[0]} \\ v_{[1]} \\ w_{[1]} \\ \vdots \\ v_{[k]} \\ w_{[k]} \end{bmatrix}}_{\text{noise}}. \quad (2.18)$$

The LS problem considered at time instant k is then given by

$$\min_{x_{[0], \dots, x_{[k+1]}} \left\| \begin{bmatrix} \hat{x}_{[0|-1]} \\ y_{[0]} - Du_{[0]} \\ -Bu_{[0]} \\ y_{[1]} - Du_{[1]} \\ -Bu_{[1]} \\ \vdots \\ y_{[k]} - Du_{[k]} \\ -Bu_{[k]} \end{bmatrix} - \begin{bmatrix} I \\ C \\ A & -I \\ & C & -I \\ & & \ddots \\ & & & C \\ & & & & A & -I \end{bmatrix} \begin{bmatrix} x_{[0]} \\ x_{[1]} \\ \vdots \\ x_{[k]} \\ x_{[k+1]} \end{bmatrix} \right\|_{W_{[k]}}^2, \quad (2.19)$$

where $W_{[k]}$ denotes the weighting matrix, which can be freely chosen.

The arguments that minimize the LS problem (2.19) consist of smoothed estimates $\hat{x}_{[0|k]}, \hat{x}_{[1|k]}, \dots, \hat{x}_{[k-1|k]}$, a filtered estimate $\hat{x}_{[k|k]}$ and a one step ahead predicted estimate $\hat{x}_{[k+1|k]}$. It has been proved [77, 136] that by choosing $W_{[k]} = \text{diag}(P_{[0|-1]}^{-1}, R^{-1}, Q^{-1}, \dots, Q^{-1})$, where $P_{[0|-1]}, Q$ and R denote matrices that can be freely chosen, the filtered estimates $\hat{x}_{[k|k]}$ that minimize two consecutive LS problems ($k = l$ and $k = l + 1, l \geq 0$) of the form (2.19) obey the Kalman filter recursion in filter form. Similarly, it has been proved that the one step

ahead predicted estimates $\hat{x}_{[k+1|k]}$ that minimize two consecutive LS problems of the form (2.19) obey the Kalman filter recursion in one step ahead prediction form. This is formalized in the following theorem.

Theorem 2.3. *Consider for $k = 0, 1, \dots$ an LS problem of the form (2.19). The arguments $\hat{x}_{[k|k]}$ and $\hat{x}_{[k+1|k]}$ that minimize two consecutive LS problems of this sequence ($k = l$ and $k = l + 1, l \geq 0$) obey the Kalman filter recursion.*

Although proving that $\hat{x}_{[k|k]}$ and $\hat{x}_{[k+1|k]}$ obey the Kalman filter recursion is quite straightforward, deriving the Kalman filter equations based on the LS problem (2.19) is very complicated and has been done only for the most simple cases [77].

Notice, very importantly, that so far we have not imposed any filter structure, nor have we given any interpretation to the initial state $x_{[0]}$ and to the noise processes $\{w_{[k]}\}_{k=0}^{\infty}$ and $\{v_{[k]}\}_{k=0}^{\infty}$, and yet the recursive Kalman filter equations are obtained. This shows first of all that the Kalman filter is optimal in an LS sense also if no interpretation is given to the noise processes and secondly that even though its recursive structure, it yields estimates that are globally optimal in an LS sense. By giving the noise processes and the initial state the stochastic interpretation considered in Sect. 2.4, and by choosing $P_{[0|-1]}$, R and Q as the covariance matrices defined in Sect. 2.4, the LS problem (2.19) can be given the interpretation of an MVU estimator. Consequently, under these assumption about the noise processes, we again find that the Kalman filter is optimal in an MVU sense.

It is easily verified that the regressor matrix in (2.19) always has full column rank. By considering a modified LS problem that uses no information about the initial state, that is by removing the first row of the regressor matrix, however, it turns out that the regressor matrix has full column rank if and only if $\{A, C\}$ is observable. A proof can be found in Lemma C.1 in Appendix C.1.

2.5.2 Recursive least-squares filtering

If one is interested only in the one step ahead predicted estimates $\hat{x}_{[1|0]}, \hat{x}_{[2|1]}, \dots$ or the filtered estimates $\hat{x}_{[0|0]}, \hat{x}_{[1|1]}, \dots$, then solving an LS problem of the form (2.19) for $k = 0, 1, \dots$ can be very time consuming.

In this section, it is shown that $\hat{x}_{[k+1|k]}$ can be computed from $\hat{x}_{[k|k-1]}$ using an RLS procedure. The idea behind the derivation is shown in Fig. 2.3. Consider two consecutive LS problems of the form (2.19), the first one using measurements up to time instant $k - 1$, the second one using measurements up to time instant k . Then, due to the structure in the regressor matrix, $\hat{x}_{[k+1|k]}$ can be computed from $\hat{x}_{[k|k-1]}$, as will now be shown.

For simplicity, we use a stochastic approach. We assume that an estimate $\hat{x}_{[k|k-1]}$ is available with error covariance matrix $P_{[k|k-1]}$ and seek for an LS problem that allows to estimate $x_{[k+1]}$ based on knowledge of $\hat{x}_{[k|k-1]}$ and of the newly available measurement $y_{[k]}$. Considering the equations of (2.18) that

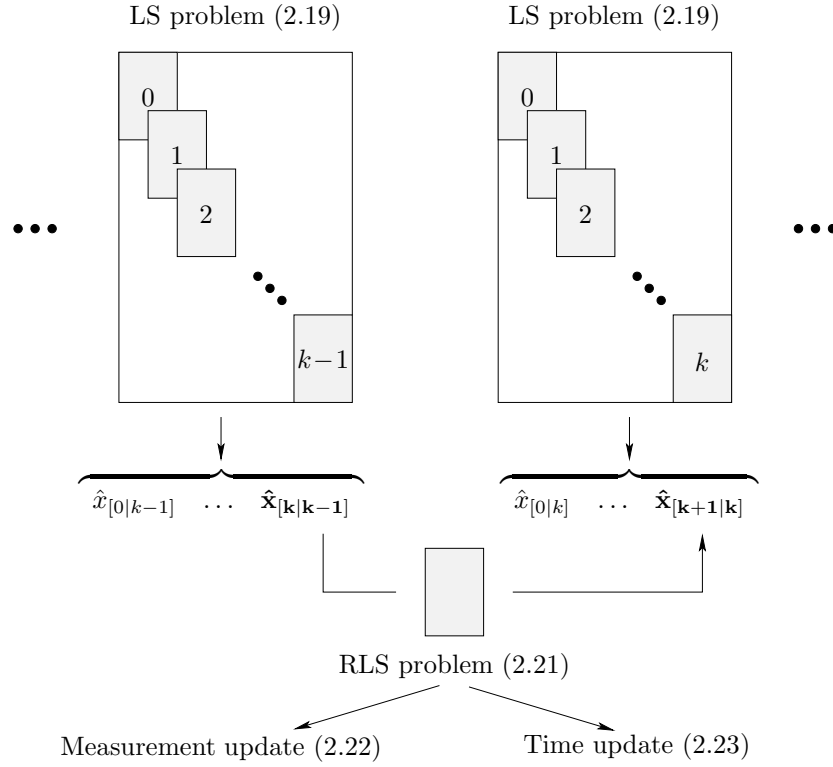


Figure 2.3: Consider two consecutive LS problems of the form (2.19), the first one using measurements up to time instant $k - 1$, the second one using measurements up to time instant k . Then, due to the structure in the regressor matrix, it can be shown that $\hat{x}_{[k+1|k]}$ can be computed from $\hat{x}_{[k|k-1]}$ using the LS problem (2.21). This yields an RLS procedure to propagate a one step ahead predicted estimate, from which the time update and the measurement update follow.

depend on data at time instant k , i.e. the last two equations, and appending an equation that summarizes the information in $\hat{x}_{[k|k-1]}$, yields

$$\begin{bmatrix} \hat{x}_{[k|k-1]} \\ y_{[k]} - Du_{[k]} \\ -Bu_{[k]} \end{bmatrix} = \begin{bmatrix} I & 0 \\ C & 0 \\ A & -I \end{bmatrix} \begin{bmatrix} x_{[k]} \\ x_{[k+1]} \end{bmatrix} + \begin{bmatrix} -\tilde{x}_{[k|k-1]} \\ v_{[k]} \\ w_{[k]} \end{bmatrix}. \quad (2.20)$$

The corresponding LS problem is then given by

$$\min_{x_{[k]}, x_{[k+1]}} \left\| \begin{bmatrix} \hat{x}_{[k|k-1]} \\ y_{[k]} - Du_{[k]} \\ -Bu_{[k]} \end{bmatrix} - \begin{bmatrix} I & 0 \\ C & 0 \\ A & -I \end{bmatrix} \begin{bmatrix} x_{[k]} \\ x_{[k+1]} \end{bmatrix} \right\|_{\tilde{W}_{[k]}}^2, \quad (2.21)$$

where $\bar{W}_{[k]}$ denotes the weighting matrix. We give (2.21) the interpretation of an MVU estimator by choosing $\bar{W}_{[k]} = \text{diag}(P_{[k|k-1]}^{-1}, R^{-1}, Q^{-1})$, where $P_{[k|k-1]}$, R and Q denote the error covariance matrices as defined above. The arguments that minimize (2.21) consist of a filtered estimate $\hat{x}_{[k|k]}$ and a one step ahead predicted estimate $\hat{x}_{[k+1|k]}$. In addition, due to the stochastic assumption, solution of the LS problem (2.21) may provide us with the error covariance matrix $P_{[k+1|k]}$ of $\hat{x}_{[k+1|k]}$. Consequently, (2.21) yields a recursive procedure to propagate a one step ahead predicted state estimate.

Proposition 2.2. *Solution of the LS problem (2.21) yields the Kalman filter equations.*

In the next sections, LS problems for the time update and the measurement update are extracted from (2.21).

2.5.2.1 Measurement update

The measurement update is obtained from (2.21) by extracting the subproblem that depends only on $x_{[k]}$, which yields

$$\min_{x_{[k]}} \left\| \begin{bmatrix} \hat{x}_{[k|k-1]} \\ y_{[k]} - Du_{[k]} \end{bmatrix} - \begin{bmatrix} I \\ C \end{bmatrix} x_{[k]} \right\|_{\bar{W}_{1[k]}}, \quad (2.22)$$

where $\bar{W}_{1[k]}$ denotes the weighting matrix which we choose as $\bar{W}_{1[k]} = \text{diag}(P_{[k|k-1]}^{-1}, R^{-1})$.

Proposition 2.3. *Solution of the LS problem (2.22) yields the measurement update of the Kalman filter.*

2.5.2.2 Time update

For the time update, we extract from (2.20) the equation that depends on $x_{[k+1]}$ and substitute $x_{[k]}$ for its estimate $\hat{x}_{[k|k]}$. This yields,

$$A\hat{x}_{[k|k]} + Bu_{[k]} = x_{[k+1]} - (A\tilde{x}_{[k|k]} + w_{[k]}).$$

The corresponding LS problem with interpretation of an MVU estimator is given by

$$\min_{x_{[k+1]}} \|x_{[k+1]} - A\hat{x}_{[k|k]} - Bu_{[k]}\|_{\bar{W}_{2[k]}}, \quad (2.23)$$

where $\bar{W}_{2[k]}$ denotes the weighting matrix, which we choose as $\bar{W}_{2[k]} = (\mathbb{E}[(A\tilde{x}_{[k|k]} + w_{[k]})(A\tilde{x}_{[k|k]} + w_{[k]})^T])^{-1}$.

Proposition 2.4. *Solution of the LS problem (2.23) yields the time update of the Kalman filter.*

2.5.3 Information Kalman filtering

In this section, we show that application of the Gauss-Markov theorem to the LS problems (2.22) and (2.23) yields information formulas for the time update and measurement update of the Kalman filter. We assume throughout this section that A and Q are nonsingular.

2.5.3.1 Measurement update

It follows from Proposition 2.3 that information formulas for the measurement update of the Kalman filter can be derived by application of the Gauss-Markov theorem (Theorem B.2) to (2.22). This yields,

$$P_{[k|k]}^{-1} = P_{[k|k-1]}^{-1} + C^T R^{-1} C, \quad (2.24)$$

$$P_{[k|k]}^{-1} \hat{x}_{[k|k]} = P_{[k|k-1]}^{-1} \hat{x}_{[k|k-1]} + C^T R^{-1} (y_{[k]} - Du_{[k]}). \quad (2.25)$$

It is easily verified that application of the matrix inversion lemma (Lemma A.2) to (2.24)-(2.25) yields the covariance formulas (2.13)-(2.15), which shows that $P_{[k|k]}^{-1}$ is indeed the information matrix of $\hat{x}_{[k|k]}$.

2.5.3.2 Time update

Information formulas for the time update of the Kalman filter can be derived by application of the Gauss Markov theorem to (2.23), or by application of the matrix inversion lemma to (2.16)-(2.17). In both cases, this yields,

$$P_{[k+1|k]}^{-1} = (I - N_{[k]}) H_{[k]}^{-1} \quad (2.26)$$

$$P_{[k+1|k]}^{-1} \hat{x}_{[k+1|k]} = (I - N_{[k]}) A^{-T} (P_{[k|k]}^{-1} \hat{x}_{[k|k]} + P_{[k|k]}^{-1} A^{-1} B u_{[k]}),$$

where

$$N_{[k]} = H_{[k]}^{-1} (H_{[k]}^{-1} + Q^{-1})^{-1}$$

$$H_{[k]}^{-1} = A^{-T} P_{[k|k]}^{-1} A^{-1}.$$

2.5.3.3 Duality relations

Notice that (2.26), the time update in information form, takes a form which is very similar to (2.15), the measurement update in covariance form. Also, (2.24), the measurement update in information form, is very similar to (2.17), the time update in covariance form. These *duality* relations between the covariance and information formulas are summarized in Table 2.1. The relations will be used in

Sect. 2.6.2 to convert between covariance and information square-root formulas almost immediately.

2.6 Square-root filtering

Already from the origin of the Kalman filter, several numerical problems of the algorithm were reported. Numerical experiments have revealed that, due to the build-up of numerical errors, the error covariance matrix can become non-symmetric. Fitzgerald [43] showed that numerical errors may even lead to *filter divergence*, a phenomenon in which the actual errors diverge out of proportion to the values predicted by the filter.

To prevent loss of symmetry, Potter and Stern [111] introduced the idea of expressing the Kalman filter equations in terms of a square-root, more precisely a Cholesky factor of the error covariance matrix. By propagating such a Cholesky factor, the computed error covariance matrix remains symmetric and positive definite at all times. In addition, due to the numerically stable operations (such as Householder reflections and Givens rotations) that are usually employed in square-root implementations, square-root filters are numerically better conditioned than a direct implementation [133].

This section is outlined as follows. In Sect. 2.6.1, we consider square-root covariance filtering. Next, in Sect. 2.6.2, the problem of square-root information filtering is addressed.

Throughout this section, we consider a system of the form (2.6) with $E = I$, $B = 0$, and $D = 0$. In Sect. 2.6.2, we assume that A and Q are nonsingular. For a matrix X , $X^{1/2}$ denotes the lower triangular Cholesky factor of X .

2.6.1 Square-root covariance filtering

First, we derive a square-root algorithm for the Kalman filter equations in prediction form. Afterwards, we indicate how square-root formulas for the time update and measurement update can be extracted.

2.6.1.1 One step ahead prediction

The basic idea behind square-root filtering is to apply orthogonal transformations to a *pre-array* containing the prior estimates, forming a *post-array* containing the updated estimates. Defining the block matrix

$$\mathcal{M}_{[k]} := \begin{bmatrix} \tilde{R}_{[k]} & CP_{[k|k-1]}A^\top \\ AP_{[k|k-1]}C^\top & AP_{[k|k-1]}A^\top + Q \end{bmatrix},$$

it follows from (2.10) that $\mathcal{M}_{[k]}$ can be decomposed as $\mathcal{M}_{[k]} = \mathcal{C}_{a[k]} \mathcal{C}_{a[k]}^\top$, where

$$\mathcal{C}_{a[k]} := \begin{bmatrix} R^{1/2} & CP_{[k|k-1]}^{1/2} & 0 \\ 0 & AP_{[k|k-1]}^{1/2} & Q^{1/2} \end{bmatrix}.$$

Time update in covariance form	Measurement update in information form
$P_{[k+1 k]}$	$P_{[k k]}^{-1}$
A	C^\top
$P_{[k k]}$	R^{-1}
Q	$P_{[k k-1]}^{-1}$

Measurement update in covariance form	Time update in information form
$P_{[k k]}$	$P_{[k+1 k]}^{-1}$
C	I
$P_{[k k-1]}$	$A^{-\top} P_{[k k]}^{-1} A^{-1}$
R	Q^{-1}

Table 2.1: Duality between the time update and the measurement update in covariance and information form.

Similarly, it follows from (2.9) that $\mathcal{M}_{[k]}$ can be decomposed as $\mathcal{M}_{[k]} = \mathcal{C}_{b[k]} \mathcal{C}_{b[k]}^\top$, where

$$\mathcal{C}_{b[k]} := \begin{bmatrix} \tilde{R}_{[k]}^{1/2} & 0 & 0 \\ K_{[k]} \tilde{R}_{[k]}^{1/2} & P_{[k+1|k]}^{1/2} & 0 \end{bmatrix}.$$

Notice that all matrices in $\mathcal{C}_{a[k]}$ are known prior to the update. Hence, $\mathcal{C}_{a[k]}$ is a pre-array. The block matrix $\mathcal{C}_{b[k]}$, on the other hand, contains matrices that have to be deduced during the update. Hence, $\mathcal{C}_{b[k]}$ is a post-array.

It follows from the discussion above that

$$\mathcal{C}_{a[k]} \mathcal{C}_{a[k]}^\top = \mathcal{C}_{b[k]} \mathcal{C}_{b[k]}^\top.$$

Consequently, there must exist a transformation matrix $\Theta_{[k]}$, with $\Theta_{[k]} \Theta_{[k]}^\top = I$, so that $\mathcal{C}_{a[k]} \Theta_{[k]} = \mathcal{C}_{b[k]}$, that is, so that

$$\begin{bmatrix} R^{1/2} & CP_{[k|k-1]}^{1/2} & 0 \\ 0 & AP_{[k|k-1]}^{1/2} & Q^{1/2} \end{bmatrix} \Theta_{[k]} = \begin{bmatrix} \tilde{R}_{[k]}^{1/2} & 0 & 0 \\ K_{[k]} \tilde{R}_{[k]}^{1/2} & P_{[k+1|k]}^{1/2} & 0 \end{bmatrix}. \quad (2.27)$$

Notice that the post-array in (2.27) is lower triangular. Consequently, (2.27) can be implemented by applying a sequence of orthogonal operations to the pre-array that brings it into the lower triangular form of the post-array. Usually, numerically stable operations like Householder reflections and Givens rotation are used [82]. The algebraic equivalence of (2.27) to the Kalman filter equations in prediction form can be verified by equating inner products on left and right hand side of the equality sign.

2.6.1.2 Time and measurement update

A square-root algorithm for the measurement update and time update can be derived in a similar manner.

For the measurement update, we define the block matrix

$$\mathcal{M}_{1[k]} := \begin{bmatrix} \tilde{R}_{[k]} & CP_{[k|k-1]} \\ P_{[k|k-1]}C^\top & P_{[k|k-1]} \end{bmatrix}.$$

Making a derivation similar to that in the previous section, yields the following array algorithm,

$$\begin{bmatrix} R^{1/2} & CP_{[k|k-1]}^{1/2} \\ 0 & P_{[k|k-1]}^{1/2} \end{bmatrix} \Theta_{1[k]} = \begin{bmatrix} \tilde{R}_{[k]}^{1/2} & 0 \\ L_{[k]} \tilde{R}_{[k]}^{1/2} & P_{[k|k]}^{1/2} \end{bmatrix}, \quad (2.28)$$

where $\Theta_{1[k]}$ denotes a sequence of orthogonal operations that brings the pre-array into the lower triangular form of the post-array.

A square-root algorithm for the time update can be derived in similar manner, yielding

$$\begin{bmatrix} AP_{[k|k]}^{1/2} & Q^{1/2} \end{bmatrix} \Theta_{2[k]} = \begin{bmatrix} P_{[k+1|k]}^{1/2} & 0 \end{bmatrix}, \quad (2.29)$$

where $\Theta_{2[k]}$ denotes a sequence of orthogonal operations that brings the pre-array into the lower triangular form of the post-array.

2.6.2 Square-root information filtering

Square-root algorithms for information filtering can be derived almost immediately from the square-root covariance algorithms by duality relations. Using Table 2.1, it follows from (2.28) that the time update in information form is given by

$$\begin{bmatrix} Q^{-\top/2} & A^{-\top}P_{[k|k]}^{-\top/2} \\ 0 & A^{-\top}P_{[k|k]}^{-\top/2} \\ \hline 0 & \hat{x}_{[k|k]}^\top P_{[k|k]}^{-\top/2} \end{bmatrix} \Theta_{3[k]} = \begin{bmatrix} \star & 0 \\ \star & P_{[k+1|k]}^{-\top/2} \\ \hline \star & x_{[k+1|k]}^\top P_{[k+1|k]}^{-\top/2} \end{bmatrix}, \quad (2.30)$$

where $\Theta_{3[k]}$ denotes a sequence of orthogonal operations that brings the pre-array into the lower triangular form of the post-array and where the “ \star ”-symbols denote matrices that are not important for our discussion. Similarly, using duality relations, it follows from (2.29) that the measurement update in information form is given by

$$\begin{bmatrix} C^\top R^{-\top/2} & P_{[k|k-1]}^{-\top/2} \\ \hline y_{[k]}^\top R^{-\top/2} & \hat{x}_{[k|k-1]}^\top P_{[k|k-1]}^{-\top/2} \end{bmatrix} \Theta_{4[k]} = \begin{bmatrix} P_{[k|k]}^{-\top/2} & 0 \\ \hline \hat{x}_{[k|k]}^\top P_{[k|k]}^{-\top/2} & \star \end{bmatrix}, \quad (2.31)$$

where $\Theta_{4[k]}$ denotes a sequence of orthogonal operations that brings the pre-array into the lower triangular form of the post-array. Notice that we have also included the updates of the state estimates in (2.30) and (2.31), as is traditionally done in the information filters.

2.7 The extended Kalman filter

Until now, we have been concerned with LTI systems. The Kalman filter equations are easily extended to linear time-varying systems, simply by replacing the time invariant system matrices by their time-varying counterparts. In many applications, however, the dynamics of the system are nonlinear. Optimal filtering for nonlinear system is very hard and therefore not feasible in practice. Consequently, a lot of approximate nonlinear filters have been proposed in literature. In this section, we consider the nonlinear filter that is most widely used, the *extended Kalman filter* (EKF).

2.7.1 Derivation of filter equations

We derive the equations using the approach in [4]. Consider a nonlinear discrete-time system governed by

$$x_{[k+1]} = f(x_{[k]}) + Ew_{[k]} \quad (2.32a)$$

$$y_{[k]} = Cx_{[k]} + v_{[k]}, \quad (2.32b)$$

where $f(\cdot)$ is a nonlinear function, $x_{[k]}$ denotes the state vector at time instant k and $y_{[k]}$ denotes the output vector at time k . The initial state $x_{[0]}$ and the noise processes $\{w_{[k]}\}_{k=0}^{\infty}$ and $\{v_{[k]}\}_{k=0}^{\infty}$ can be given the stochastic interpretations considered in Sect. 2.4, however, due to the relation between the Kalman filter and LS estimation, such an interpretation is not necessary.

The approximation in the EKF is based on expanding $f(x_{[k]})$ in Taylor series around the current state estimate and neglecting higher order terms. More precisely, let $\hat{x}_{[k|k]}$ denote the current state estimate, then the approximation is given by

$$f(x_{[k]}) \approx f(\hat{x}_{[k|k]}) + A_{[k]}(x_{[k]} - \hat{x}_{[k|k]}), \quad (2.33)$$

where

$$A_{[k]} := \frac{\partial f}{\partial x}(\hat{x}_{[k|k]}).$$

By substituting (2.33) in (2.32), it follows that the nonlinear system (2.32) can be approximated around the current state estimate by the linear time-varying system

$$x_{[k+1]} = A_{[k]}x_{[k]} + u_{[k]} + Ew_{[k]} \quad (2.34a)$$

$$y_{[k]} = Cx_{[k]} + v_{[k]}, \quad (2.34b)$$

where $u_{[k]} := f(\hat{x}_{[k|k]}) - A_{[k]}\hat{x}_{[k|k]}$.

The equations of the EKF are then defined by the Kalman filter equations for the system (2.34). It is straightforward to show that the measurement update of that Kalman filter is given by (2.13)-(2.15). In the time update, the nonlinear function $f(\cdot)$ shows up. This yields:

Extended Kalman filter

- **Measurement update**

$$\begin{aligned}\hat{x}_{[k|k]} &= \hat{x}_{[k|k-1]} + L_{[k]}(y_{[k]} - C\hat{x}_{[k|k-1]} - Du_{[k]}) \\ L_{[k]} &= P_{[k|k-1]}C^T(CP_{[k|k-1]}C^T + R)^{-1} \\ P_{[k|k]} &= P_{[k|k-1]} - P_{[k|k-1]}C^T(CP_{[k|k-1]}C^T + R)^{-1}CP_{[k|k-1]}\end{aligned}$$

- **Time update**

$$\begin{aligned}\hat{x}_{[k+1|k]} &= f(\hat{x}_{[k|k]}) \\ P_{[k+1|k]} &= A_{[k]}P_{[k|k]}A_{[k]}^T + EQE^T\end{aligned}$$

Notice that even if $P_{[0|-1]}$, R , and Q are given the stochastic interpretations considered in Sect. 2.4, $P_{[k|k]}$ and $P_{[k+1|k]}$ are no longer the error covariance matrices of $\hat{x}_{[k|k]}$ and $\hat{x}_{[k+1|k]}$. They can be considered as approximations to these error covariance matrices.

2.7.2 Observability

Application of the EKF requires that the linear time-varying system (2.34) is observable (or at least detectable). Observability for time-varying systems is usually defined over an interval. More precisely, a time-varying system is said to be observable over the interval $[k_0, k_1]$ if, given y over that interval, we can determine $x_{[k_0]}$. This leads to a time-varying observability matrix that needs now to be of full column rank [113].

2.8 Conclusion

This chapter has provided a brief introduction to Kalman filtering. The Kalman filter was introduced as an extension of asymptotic and deadbeat estimators. It was shown that the Kalman filter equations can be easily derived based on MVU estimation. The relation between the Kalman filter and least-squares estimation was discussed. Two alternative forms of the traditional equations were considered, the information form and the square-root form, and their numerical advantages were discussed. Finally, the extended Kalman filter, a nonlinear extension of the Kalman filter, was briefly considered.

Part I

System Inversion

Chapter 3

Inversion of Deterministic Systems

This chapter addresses left inversion of linear discrete-time deterministic systems in state-space form. The main contribution is the derivation of a general form of a time-delayed left inverse. The general form contains a free matrix parameter which allows to place the poles of the inverse system. It is shown that pole placement is possible if a certain matrix pair is observable. This pair turns out to be observable if the system has no zeros. Based on the theory of reduced order observers, a new technique is developed to simultaneously reduce the order of the inverse system and place its poles. The results of this chapter generalize existing methods for left inversion, and in addition also have direct implications for state estimation in the presence of unknown inputs.

3.1 Introduction

The problem of inverting linear dynamical system has received a lot of attention in the past due to its strong connection to control and estimation theory. System inversion has applications in such areas as fault detection and isolation [128], geophysical estimation [87], simultaneous stabilization of dynamical systems [18] and adaptive tracking control [123].

In this chapter, we will be concerned with left inversion of deterministic systems in state-space form, where the objective is to exactly reconstruct the input applied to the system from knowledge of the system outputs.

A brief overview of the most important accomplishments in the early history of systems inversion is shown in Fig. 3.1. The problem has been intensively studied during the end of the sixties and the beginning of the seventies. The earliest systematic approach to the inversion of deterministic systems is due to Brockett & Mesarovic [15, 16] who considered SISO systems and derived

necessary and sufficient conditions for invertibility as well as an inversion algorithm.

Important contributions in the inversion of MIMO state-space systems were obtained by Sain & Massey [95, 115] and Silverman [116]. Their contributions have given a lot of insight in the structure and the properties of inverse systems. Silverman obtained with his so-called *structure algorithm* the insight that an inverse system can be realized with exactly the same number of differentiators (or delay elements in the discrete-time case) as the original system. Sain & Massey mainly studied left inversion and introduced the concept of the *inherent delay* of a discrete-time system, which is the minimal time delay that needs to be allowed in the reconstruction of the input.

A disadvantage of the inversion procedures by Silverman and Sain & Massey is that the poles of the inverse system can not be tuned. Consequently, the procedures can yield unstable inverses. Unstable inverses are of no harm if the initial state of the system is known exactly. In applications, however, a stable inverse is desired since this introduces robustness against errors in the initial state. Moylan [100] was the first to develop an algorithm that always returns a stable inverse (provided that the original system has no unstable zeros). His algorithm is close to that of Silverman. A very straightforward approach to stable inversion is given by Antsaklis [7]. His treatment is based on feedback control and allows to assign the poles of the inverse system (except those that equal the zeros of the given system). His method is, however, limited to systems with inherent delay zero.

The inversion procedures considered above are all based on time-domain approaches. However, because system inversion basically is an input-output concept, frequency-domain approaches have also received a lot of attention, see e.g. [34, 104, 131, 132] and the references therein.

In case the initial state of the system is unknown, it is more convenient to consider the inverse system as an input estimator. In fact, all time-domain approaches to left inversion considered above not only yield estimates of the system input, but also of the system state. They can thus be considered as joint input-state estimators. Frequency domain approaches lack the capability of state estimation and are therefore only briefly considered in this chapter.

Since the end of the seventies, the problem of left inversion has received only little attention. Most approaches are based on joint input-state estimation and are limited to the one step ahead prediction or the filtering problem [42, 67]. Only few approaches for time-delayed estimation or smoothing have been considered [44]. The problem of assigning the poles of inverses that reconstruct the input with time delays has remained unsolved up to now.

The problem of state estimation for linear deterministic systems with unknown inputs, which is actually closely related to system inversion, has received considerably more attention the last few decades [26, 68, 139]. It is well established by now that state reconstructors exist under less strict conditions than inverse systems. During the last years, research has also shifted towards time-delayed state estimation. The first systematic approach is due to Saberi et al. [114] who handled time delays by state augmentation. Recently, Sundaram

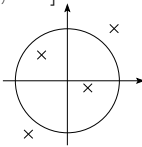
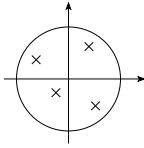
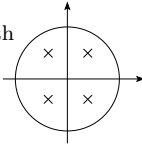
<p>1965</p> <p>SISO systems</p> <p>Brockett & Mesarovic [15, 16]</p> <p>first invertibility criteria and algorithms</p>	<p>1968-1969</p> <p>MIMO systems</p> <p>Sain & Massey [95, 115]</p> <p>inherent delay</p> <p>Silverman [116]</p> <p>structure algorithm</p> 
<p>1977</p> <p>Stable inversion</p> <p>Moylan [100]</p> 	<p>1978</p> <p>Pole placement</p> <p>Antsaklis [7]</p> <p>limited to systems with inherent delay zero</p> 

Figure 3.1: Brief overview of the most important accomplishments in the early history system inversion. The first inversion procedures for SISO systems were introduced by Brockett & Mesarovic in 1965. Important contributions in the inversion of MIMO systems were obtained by Silverman and Sain & Massey in 1969. In 1977, Moylan was the first to derive an algorithm that returns stable inverses. Pole placement was first considered by Antsaklis in 1978.

and Hadjicostis [126] showed that state augmentation is not necessary and developed full order and reduced order observers. They showed that the poles of their observer can be assigned if the system has no zeros.

Personal contributions

The personal contribution of this chapter is the development of a new inversion procedure in Sects. 3.5 to 3.7.

- In Sect. 3.5, a general form of an L -delay left inverse system is derived based on estimation theory. Like in the approach of Sain & Massey, this general form consists of a bank of delay elements followed by a dynamical system. The dynamical system derived in this thesis has the most general form.
- The general form consists of a matrix parameter which can be freely chosen. In Sect. 3.6, it is shown that an appropriate choice allows to place the poles of the inverse system. Conditions are derived under which pole placement is possible.

- In Sect. 3.7, a procedure is developed to reduce the order of the inverse system and simultaneously place its poles based on the theory of reduced order state observers.

Chapter outline

This chapter is outlined as follows. Section 3.2 defines the problems of left and right inversion. In Sect. 3.3, conditions for left invertibility of linear state-space system are derived. Section 3.4.2 briefly describes the inversion approach of Sain & Massey and discusses the advantages/disadvantages over that of Silverman. The new inversion procedure based on estimation theory is developed in Sect. 3.5. Section 3.6 addresses stable inversion. Conditions are derived under which the poles of the inverse system can be assigned. In Sect. 3.7, a procedure is developed to reduce the order of the inverse system. Finally, in Sect. 3.8, three numerical examples are considered.

3.2 Problem formulation

Because system inversion is an input-output concept, it is most easily understood in terms of the transfer function of a system. Consider a LTI discrete-time system \mathcal{S} . Let $U(z)$ and $Y(z)$ denote the z -transforms of the m -dimensional input vector and the p -dimensional output vector of \mathcal{S} , respectively. Then, $Y(z)$ and $U(z)$ are related by

$$Y(z) = H(z)U(z),$$

where the $p \times m$ rational matrix $H(z)$ is called the *transfer function* of \mathcal{S} .

The problem of system inversion deals with deriving a system with input $Y(z)$ and output $U(z)$. A distinction must be made between left inversion and right inversion. As we will see, a left inverse can be interpreted as an input estimator. A right inverse, on the other hand, can be interpreted as a feedforward controller.

3.2.1 Left inversion

Like in [115], we define a left inverse of \mathcal{S} in terms of transfer functions and give an interpretation in the time domain afterwards.

Definition 3.1. *A system is said to be an L -delay left inverse of \mathcal{S} if its transfer function $H_L(z)$ satisfies $H_L(z)H(z) = z^{-L}I$.*

We now give an interpretation in the time domain. Consider the series connection of the system \mathcal{S} with transfer function $H(z)$ and an L -delay left inverse of \mathcal{S} with transfer function $H_L(z)$, as shown in Fig. 3.2. Let $Y_L(z)$ denote the z -transform of the output of the left inverse. Then, it follows that

$$Y_L(z) = H_L(z)H(z)U(z) = z^{-L}U(z). \quad (3.1)$$

Converting (3.1) to the time domain, yields

$$y_L[k] = u_{[k-L]},$$

from which we conclude that an L -delay left inverse of \mathcal{S} reconstructs at its output the input applied to \mathcal{S} with L steps delay.

Definition 3.2. *The system \mathcal{S} is said to be L -delay left invertible if it has an L -delay left inverse.*

Definition 3.3. *The system \mathcal{S} is said to be left invertible if it is L -delay left invertible for some finite nonnegative integer L .*

Notice from Definition 3.1 that a necessary condition for (L -delay) left inversion is that $H(z)$ has full column rank. This implies that $p \geq m$, meaning that the dimension of the output vector must be at least the dimension of the input vector.

3.2.2 Right inversion

Definition 3.4. *A system is said to be an L -delay right inverse of \mathcal{S} if its transfer function $H_R(z)$ satisfies $H(z)H_R(z) = z^{-L}I$.*

We now give an interpretation in the time domain. Consider the series connection of a right inverse of \mathcal{S} with transfer function $H_R(z)$ and \mathcal{S} itself, as shown in Fig. 3.3. Let $U_R(z)$ denote the z -transform of the input of the right inverse. Then, it follows that

$$Y(z) = H(z)H_R(z)U_R(z) = z^{-L}U_R(z). \quad (3.2)$$

Converting (3.2) to the time domain, yields

$$y[k] = u_{R[k-L]},$$

from which we conclude that, given a desired output u_R of \mathcal{S} , an L -delay right inverse of \mathcal{S} computes a signal that when applied to \mathcal{S} yields the desired output with L steps delay.

Definition 3.5. *The system \mathcal{S} is said to be L -delay right invertible if it has an L -delay right inverse.*

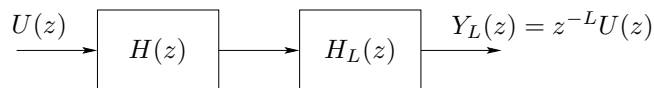


Figure 3.2: Series connection of a system with transfer function $H(z)$ and a left inverse of that system with transfer function $H_L(z)$.

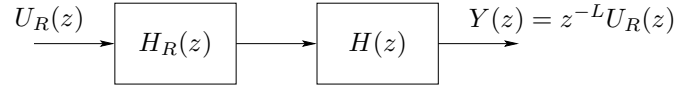


Figure 3.3: Series connection of a right inverse with transfer function $H_R(z)$ and the system itself with transfer function $H(z)$.

Definition 3.6. The system \mathcal{S} is said to be right invertible if it is L -delay right invertible for some finite nonnegative integer L .

Notice from Definition 3.4 that a necessary condition for (L -delay) right inversion is that $H(z)$ has full row rank. This implies that $m \geq p$, meaning that the dimension of the input vector must be at least the dimension of the output vector.

It follows from the discussion above that a system can be both left and right invertible only if it is square ($p = m$). A necessary and sufficient condition for invertibility is then that $H(z)$ is invertible.

3.2.3 Duality

The interpretations and necessary conditions for the existence of left and right inverses derived above are summarized in Fig. 3.4. Notice the duality between left and right inversion. Because the transfer function $H_L(z)$ of an L -delay left inverse of a system with transfer function $H(z)$ satisfies

$$H^T(z)H_L^T(z) = z^{-L}I,$$

it follows that $H_L^T(z)$ is the transfer function of an L -delay right inverse of a system with transfer function $H^T(z)$. Once a left (right) inverse has been found, a right (left) inverse can thus easily be derived based on this duality. This basically means that only one of both inversion problems needs to be studied. Using duality relations, the results translate to the other problem. In the remainder of this chapter, only the problem of left inversion will be considered. We focus on LTI discrete-time systems in state-space form.

3.3 Left invertibility of state-space systems

Consider the LTI discrete-time system

$$\mathcal{S} : \begin{cases} x_{[k+1]} = Ax_{[k]} + Bu_{[k]} \\ y_{[k]} = Cx_{[k]} + Du_{[k]}, \end{cases} \quad (3.3)$$

where $x_{[k]} \in \mathbb{R}^n$ denotes the state vector at time instant k , $u_{[k]} \in \mathbb{R}^m$ denotes the input vector at time k , and $y_{[k]} \in \mathbb{R}^p$ denotes the output vector at time k . The input vector $u_{[k]}$ is assumed to be deterministic and unknown. We assume in

		LEFT INVERSION	RIGHT INVERSION
Interpretation	z-domain	<p>The transfer function $H_L(z)$ of an L-delay left inverse satisfies</p> $H_L(z)H(z) = z^{-L}I_m$	<p>The transfer function $H_R(z)$ of an L-delay right inverse satisfies</p> $H(z)H_R(z) = z^{-L}I_p$
	time domain	<p>The output y_L of an L-delay left inverse reconstructs the system input with L steps delay</p> $y_L[k] = u[k-L]$	<p>Given a desired output signal u_R, an L-delay right inverse computes an input signal that yields the desired output with L steps delay</p> $y[k] = u_R[k-L]$
Necessary condition for existence	z-domain	<p>Necessary conditions for the existence of an L-delay left inverse are</p> $\text{rank}(H(z)) = m, \quad p \geq m$	<p>Necessary conditions for the existence of an L-delay right inverse are</p> $\text{rank}(H(z)) = p, \quad m \geq p$
	time domain (state-space system)	<p>A necessary and sufficient condition for existence of an L-delay left inverse is</p> $\text{rank}(\mathcal{H}_L) - \text{rank}(\mathcal{H}_{L-1}) = m$	<p>A necessary and sufficient condition for existence of an L-delay right inverse is</p> $\text{rank}(\mathcal{H}_L) - \text{rank}(\mathcal{H}_{L-1}) = p$

Figure 3.4: Interpretation and necessary condition for the existence of left and right inverses. Notice the duality between left and right inversion: if $H_L(z)$ is the transfer function of an L -delay left inverse of $H(z)$, then $H_L^T(z)$ is the transfer function of an L -delay right inverse of $H^T(z)$. In this chapter, only left inversion is considered. However, due to the duality, it is expected that most results translate easily to the problem of right inversion.

the remainder of this chapter that $p \geq m$. For clarity of exposition, we consider only time-invariant systems. Most results are, however, easily generalized to time-varying systems.

It follows from the discussion above that a SISO system is always invertible. Let the transfer function be given by

$$H(z) = \frac{f(z)}{g(z)},$$

where $f(z)$ and $g(z)$ are relatively prime polynomials. The roots of $f(z)$ and $g(z)$ are then called the *zeros* and *poles* of the transfer function, respectively. We say that the system has *zeros at infinity* if $\lim_{z \rightarrow \infty} H(z) = 0$. By inverting the system, it is now clear that the poles become zeros and vice versa. This means that a *nonminimum phase* (NMP) system, i.e. a system with unstable zeros, can not have a stable inverse system.

For MIMO state-space systems, the derivation of invertibility conditions and of relations between the poles and zeros of a system and its inverse, are much more involved. This section addresses these problems. In Sect. 3.3.1, conditions on the system matrices A, B, C and D are derived under which the system \mathcal{S} is L -delay left invertible.

3.3.1 The invertibility condition of Sain & Massey

Defining

$$y_{[k:k+L]} := \begin{bmatrix} y_{[k]} \\ y_{[k+1]} \\ \vdots \\ y_{[k+L]} \end{bmatrix}, \quad (3.4)$$

Sain & Massey [115] showed that \mathcal{S} is L -delay left invertible under the condition given in the following proposition.

Proposition 3.1 ([95]). *The system \mathcal{S} is L -delay left invertible if and only if $u_{[k]}$ can be uniquely determined from knowledge of $y_{[k:k+L]}$ and $x_{[k]}$.*

We now derive a condition under which $u_{[k]}$ can be determined from $y_{[k:k+L]}$ and $x_{[k]}$. It is readily checked from (3.3) that

$$y_{[k:k+L]} = \mathcal{O}_L x_{[k]} + \mathcal{H}_L u_{[k:k+L]}, \quad (3.5)$$

where $u_{[k:k+L]}$ is defined similar to $y_{[k:k+L]}$, where \mathcal{O}_L is defined as in (2.3) and where the $p(L+1) \times m(L+1)$ Toeplitz matrix \mathcal{H}_L , defined by

$$\mathcal{H}_L := \begin{bmatrix} D & 0 & 0 & \cdots & 0 \\ CB & D & 0 & \cdots & 0 \\ CAB & CB & D & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{L-1}B & CA^{L-2}B & CA^{L-3}B & \cdots & D \end{bmatrix}, \quad (3.6)$$

contains the Markov parameters D, CB, CAB, \dots

Based on (3.5), Sain & Massey showed that $u_{[k]}$ can be uniquely determined from knowledge of $y_{[k:k+L]}$ and $x_{[k]}$, and thus \mathcal{S} is L -delay left invertible, under the condition given in the following theorem.

Theorem 3.1 ([95]). *System \mathcal{S} is L -delay left invertible if and only if*

$$\text{rank}(\mathcal{H}_L) - \text{rank}(\mathcal{H}_{L-1}) = m, \quad (3.7)$$

where $\text{rank}(\mathcal{H}_{-1}) := 0$.

We give a proof of Theorem 3.1 which differs from that of Sain & Massey in [95]. In the proof, we make use of the following lemma.

Lemma 3.1. *An $m \times p(L+1)$ matrix \mathcal{M}_L satisfying $\mathcal{M}_L \mathcal{H}_L = \check{I}_m$, with $\check{I}_m := [I_m \ 0]$, exists if and only if $\text{rank}(\mathcal{H}_L) - \text{rank}(\mathcal{H}_{L-1}) = m$.*

Proof: The proof assumes knowledge of the solution of linear matrix equations, of which a brief introduction is given in Appendix A.1. It follows from Theorem A.1 that a necessary and sufficient condition for the existence of a matrix \mathcal{M}_L satisfying $\mathcal{M}_L \mathcal{H}_L = \check{I}_m$ is

$$\text{rank} \left(\begin{bmatrix} \check{I}_m \\ \mathcal{H}_L \end{bmatrix} \right) = \text{rank}(\mathcal{H}_L).$$

Noting that

$$\begin{aligned} \text{rank} \left(\begin{bmatrix} \check{I}_m \\ \mathcal{H}_L \end{bmatrix} \right) &= \text{rank} \left(\begin{bmatrix} I_m & 0 \\ 0 & \mathcal{H}_{L-1} \end{bmatrix} \right) \\ &= \text{rank}(\mathcal{H}_{L-1}) + m, \end{aligned}$$

concludes the proof. ■

The proof of Theorem 3.1 can now be given.

Proof: Consider a linear combination of $x_{[k]}$ and $y_{[k:k+L]}$,

$$A_L x_{[k]} + B_L y_{[k:k+L]}, \quad (3.8)$$

where $A_L \in \mathbb{R}^{m \times n}$ and $B_L \in \mathbb{R}^{m \times p(L+1)}$ have to be determined so that $A_L x_{[k]} + B_L y_{[k:k+L]} = u_{[k]}$. Using (3.5), (3.8) is rewritten as

$$(A_L + B_L \mathcal{O}_L) x_{[k]} + B_L \mathcal{H}_L u_{[k:k+L]}. \quad (3.9)$$

Expression (3.9) equals $u_{[k]}$ for all possible $x_{[k]}$ and all possible $u_{[k:k+L]}$ if and only if

$$A_L = -B_L \mathcal{O}_L \quad (3.10)$$

and

$$B_L \mathcal{H}_L = \check{I}_m. \quad (3.11)$$

It follows from Lemma 3.1 that a matrix B_L satisfying (3.11) exists if and only if condition (3.7) obtains, which concludes the proof. ■

Condition (3.7) for L -delay left invertibility is summarized in Fig. 3.7 for $L = 0$ and $L = 1$.

Notice that it follows from (3.7) that if \mathcal{S} has an L_0 -delay left inverse, it has an L -delay left inverse for all $L \geq L_0$. To check whether \mathcal{S} is invertible, one could thus test (3.7) with increasing L until eventually an L is found for which (3.7) obtains. However, the following theorem due to Willsky [137], yields a more convenient way to check for invertibility.

Theorem 3.2 ([137]). *Let q be the dimension of the nullspace of D , then \mathcal{S} is left invertible if and only if*

$$\text{rank}(\mathcal{H}_{n-q+1}) - \text{rank}(\mathcal{H}_{n-q}) = m, \quad (3.12)$$

that is, if \mathcal{S} is invertible, its inherent delay can not exceed $n - q + 1$.

An important concept introduced by Sain & Massey is the inherent delay of a system, which is the minimal delay that needs to be allowed in the reconstruction of the input.

Definition 3.7. *Let \mathcal{S} be left invertible. Then, the least nonnegative integer L for which \mathcal{S} is L -delay left invertible is called the inherent delay of \mathcal{S} .*

3.3.2 Left inversion and system zeros

The inversion of systems with zeros requires special attention. In this section, we first define a transmission zero of \mathcal{S} and then show that the input of a system with such a zero can not be uniquely reconstructed.

Definition 3.8. *A number $\lambda \in \mathbb{C}$ is called a transmission zero of \mathcal{S} if*

$$\text{rank} \left(\begin{bmatrix} A - \lambda I & B \\ C & D \end{bmatrix} \right) < n + \min(m, p).$$

If λ is a transmission zero of \mathcal{S} , there thus exist vectors $x_0 \in \mathbb{C}^n$ and $g \in \mathbb{C}^m$ such that

$$\begin{bmatrix} A - \lambda I & B \\ C & D \end{bmatrix} \begin{bmatrix} x_0 \\ g \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The input

$$u_{z[k]} = \begin{cases} g & \text{for } k = 0 \\ g\lambda^k & \text{for } k = 1, 2, \dots \end{cases}$$

applied to \mathcal{S} with initial condition $x_{[0]} = x_0$, then yields $y_{[k]} = 0$ for $k = 0, 1, 2, \dots$

The physical interpretation of a transmission zero is thus that there exists an input and an initial state for which the system output is zero. An important consequence is that the input of a system with a zero can not be uniquely reconstructed. Indeed, let \mathcal{S} be initialized with x_0 . Let the response of \mathcal{S} to an input signal u_r be given by y_r , then the response of \mathcal{S} to the input signal

$u_r + u_z$ will be the same signal y_r . From knowledge of y_r , we can thus not distinguish whether u_r or $u_r + u_z$ was applied to the input of \mathcal{S} .

If the zero lies inside the unit circle, it follows that $u_{z[k]} \rightarrow 0$ for $k \rightarrow \infty$, meaning that we can asymptotically reconstruct the input signal applied to \mathcal{S} . For an NMP system, however, asymptotic reconstruction is not possible because u_z grows unbounded instead of converging to zero. The input of an NMP system can thus be uniquely reconstructed only if we have the prior knowledge that input does not grow unbounded.

3.4 Inversion techniques: state of the art

This section consider the state of the art in left inversion. In Sect. 3.4.1, we consider a technique for instantaneous inversion. Next, in Sect. 3.4.2, the approach of Sain and Massey is discussed. Finally, in Sect. 3.4.3, the inversion procedure of Sain & Massey is briefly compared to that of Silverman.

3.4.1 Instantaneous inversion

This section considers the problem of constructing an *instantaneous left inverse* of \mathcal{S} , i.e. an inverse that reconstructs the input without delay. It follows from Fig. 3.7 that such an inverse exists if D is full column rank. The idea behind the derivation is to derive an equation that expresses the input in terms of the state and the output, and then substitute this expression into the state equation of the system.

By pre-multiplying the output equation of \mathcal{S} by D^\dagger , it follows that $u_{[k]}$ can be reconstructed from knowledge of $x_{[k]}$ and $y_{[k]}$ as

$$u_{[k]} = -D^\dagger C x_{[k]} + D^\dagger y_{[k]}. \quad (3.13)$$

Substituting (3.13) in the state equation of \mathcal{S} then yields

$$x_{[k+1]} = (A - BD^\dagger C)x_{[k]} + BD^\dagger y_{[k]}. \quad (3.14)$$

Consider now the system with state equation (3.14) and output equation (3.13), that is, the system

$$x_{[k+1]} = (A - BD^\dagger C)x_{[k]} + BD^\dagger y_{[k]} \quad (3.15a)$$

$$u_{[k]} = -D^\dagger C x_{[k]} + D^\dagger y_{[k]}. \quad (3.15b)$$

Choosing the initial state of (3.15) equal to that of \mathcal{S} , it is easily verified that when (3.15) is driven by the output sequence $\{y_{[k]}\}_{k=0}^{\infty}$ of \mathcal{S} , it exactly reconstructs the corresponding input sequence $\{u_{[k]}\}_{k=0}^{\infty}$, and is thus an instantaneous left inverse of \mathcal{S} .

In case D is not full column rank, inversion of \mathcal{S} is much more complicated. The reason is that a certain delay in the reconstruction of the input needs to be allowed.

3.4.2 The approach of Sain & Massey

Sain & Massey derived a very straightforward method for L -delay left inversion. Their approach is similar to that considered in the previous section in the sense that first an expression for the input in terms of the state and the outputs is derived, and then this expression is substituted in the state equation of the system.

In this section, a derivation is considered that differs from that of Sain & Massey [115] in the sense that we directly derive a state-space description of the inverse system, whereas Sain & Massey derived a block diagram from which a state-space description can be deduced.

Our derivation is based on Lemma 3.1. Pre-multiplying left and right hand side of (3.5) by \mathcal{M}_L , shows that $u_{[k]}$ can be reconstructed from knowledge of $x_{[k]}$ and $y_{[k:k+L]}$ as

$$u_{[k]} = -\mathcal{M}_L \mathcal{O}_L x_{[k]} + \mathcal{M}_L y_{[k:k+L]}. \quad (3.16)$$

Substituting (3.16) in the state equation of \mathcal{S} then yields the following dynamical system with input $y_{[k:k+L]}$ and output $u_{[k]}$,

$$x_{[k+1]} = (A - B\mathcal{M}_L\mathcal{O}_L)x_{[k]} + B\mathcal{M}_L y_{[k:k+L]} \quad (3.17a)$$

$$u_{[k]} = -\mathcal{M}_L \mathcal{O}_L x_{[k]} + \mathcal{M}_L y_{[k:k+L]}. \quad (3.17b)$$

Partitioning \mathcal{M}_L as $\mathcal{M}_L = [\mathcal{M}_{L,0} \mathcal{M}_{L,1} \cdots \mathcal{M}_{L,L}]$, where $\mathcal{M}_{L,i} \in \mathbb{R}^{m \times p}$, $i = 0, 1, \dots, L$ it is now straightforward to show that the system depicted in Fig. 3.5 is an L -delay left inverse of \mathcal{S} . It is easily verified that the inverse system shown in Fig. 3.5 is equivalent to that of Fig. 1 in [115].

The inverse system shown in Fig. 3.5 consists of two parts.

- The first part consists of a bank of delay elements with input $y_{[k+L]}$ and output $\mathcal{M}_L y_{[k:k+L]}$. Notice that pL delay elements are needed in order to realize this bank of delay elements. However, as pointed out by Sain & Massey [115], the number of delay elements in the bank can be reduced by solving $\mathcal{M}_L \mathcal{H}_L = \check{I}_m$ so that the $\mathcal{M}_{L,i}$, especially those with small i , have a maximal number of zero columns.
- The second part is the dynamical system (3.17), which has been called the *dynamical portion* by Sain & Massey. Notice that the dynamical portion does not appear in [115] in the state-space form (3.17).

In the remainder of this chapter, we will denote a dynamical portion of an L -delay left inverse of \mathcal{S} by \mathcal{S}_L^- .

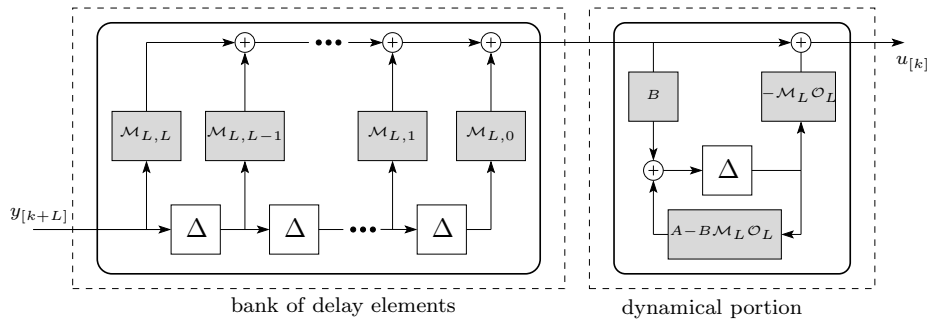


Figure 3.5: Structure of the L -delay left inverse system considered by Sain and Massey.

3.4.3 Comparison to Silverman's structure algorithm

In this section, we briefly compare the inversion procedure of Sain & Massey to Silverman's so-called structure algorithm [116].

In both approaches, the inverse system consists of a bank of delay elements followed by a dynamical portion. While Silverman's algorithm yields inverse systems that can be realized with the same number of delay elements as the original system, the number of delay elements needed to realize the inverse of Sain & Massey can be higher or lower than the order of the original system, depending on the particular system to be inverted. The approach of Sain & Massey may thus require more computer memory. On the other hand, once the inverse system has been computed, the input reconstruction procedure of Sain & Massey is computationally more efficient.

The approach of Sain and Massey allows to derive inverses with arbitrary delay larger than or equal to the inherent delay. Silverman's structure algorithm, on the other hand, is limited to inverses that reconstruct the system input with delay equal to the inherent delay of the system.

3.5 An estimation approach to system inversion

The major disadvantage of the algorithms by Silverman and Sain & Massey is that they can yield unstable inverses. Although some parameters during the design can be freely chosen, there is no systematic way to tune these parameters such that the inverse system is stable. As already discussed, Moylan [100] was the first to derive an algorithm that yields stable inverses. Antsaklis [7] developed a straightforward approach to assign the poles of the inverse system. However, his treatment is limited to systems with inherent delay zero.

In this section, a general form of an L -delay left inverse system is derived based on estimation theory. The inverses considered in this section consist of a bank of delay elements, similar to that of Sain & Massey, followed by a

dynamical portion. This structure of the inverse system is schematically shown in Fig. 3.6. The dynamical portion derived in this chapter has the most general form. This general form consists of a matrix parameter that can be freely chosen. In Sect. 3.6, conditions and methods are derived under which the poles of the dynamical portion can be assigned. These conditions basically extend the results of Antsaklis from $L = 0$ to arbitrary L .

This section is outlined as follows. In Sect. 3.5.1, we consider the problem of state reconstruction in the presence of unknown inputs. The general form of a state reconstructor is derived. Next, in Sect. 3.5.2, a similar derivation is given for input reconstruction. Finally, in Sect. 3.5.3, the state reconstructor and input reconstructor are combined, yielding the dynamical portion of an inverse system.

3.5.1 State reconstruction

First, we turn our attention to the derivation of the state reconstructor. We consider a recursive state reconstructor that computes $x_{[k+1]}$ as a linear combination of $x_{[k]}$ and $y_{[k:k+L]}$. The general form of such a state reconstructor and the condition under which it exists, are given in the following theorem. The theorem assumes knowledge of generalized inverses, of which a brief introduction is given Appendix A.1.

Theorem 3.3. *If and only if*

$$\text{rank}(\mathcal{H}_L) = \text{rank}(\mathcal{H}_{L-1}) + \text{rank}\left(\begin{bmatrix} B \\ D \end{bmatrix}\right), \quad (3.18)$$

$x_{[k+1]}$ can be reconstructed as a linear combination of $x_{[k]}$ and $y_{[k:k+L]}$. The general form of the reconstruction can be written as

$$x_{[k+1]} = (A - \mathcal{K}_L \mathcal{O}_L)x_{[k]} + \mathcal{K}_L y_{[k:k+L]}, \quad (3.19)$$

where \mathcal{K}_L is given by

$$\mathcal{K}_L = \check{B}\mathcal{H}_L^{(1)} + Z_L \Sigma_L, \quad (3.20)$$

with Z_L an $n \times p(L+1)$ arbitrary matrix parameter, $\check{B} := [B \ 0]$ and $\Sigma_L := I - \mathcal{H}_L \mathcal{H}_L^{(1)}$.

In the proof of Theorem 3.3, we make use of the following lemma.

Lemma 3.2. *A matrix \mathcal{K}_L satisfying $\mathcal{K}_L \mathcal{H}_L = \check{B}$ exists if and only if condition (3.18) obtains. Under condition (3.18), the general solution for \mathcal{K}_L is given by (3.20).*

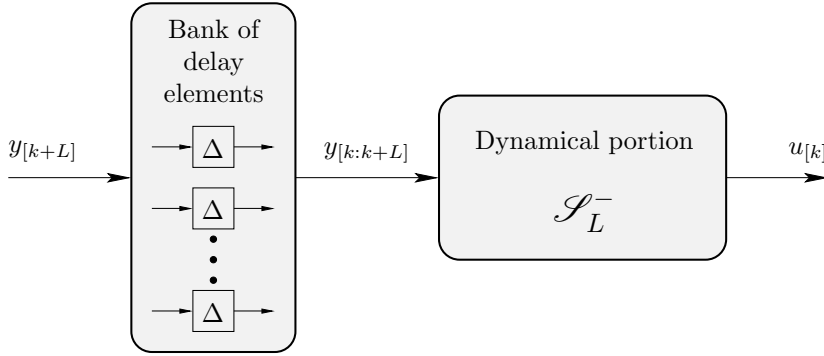


Figure 3.6: Structure of the L -delay left inverses considered in this chapter. The inverses consist of a bank of delay elements followed by a dynamical portion denoted by \mathcal{S}_L^- .

Proof: It follows from Theorem A.1 that a necessary and sufficient condition for the existence of a matrix \mathcal{K}_L satisfying to $\mathcal{K}_L \mathcal{H}_L = \check{B}$ is

$$\text{rank} \left(\begin{bmatrix} \check{B} \\ \mathcal{H}_L \end{bmatrix} \right) = \text{rank}(\mathcal{H}_L).$$

Noting that

$$\begin{aligned} \text{rank} \left(\begin{bmatrix} \check{B} \\ \mathcal{H}_L \end{bmatrix} \right) &= \text{rank} \left(\begin{bmatrix} B & 0 \\ D & 0 \\ \mathcal{O}_{L-1}B & \mathcal{H}_{L-1} \end{bmatrix} \right) \\ &= \text{rank} \left(\begin{bmatrix} B & 0 \\ D & 0 \\ 0 & \mathcal{H}_{L-1} \end{bmatrix} \right) \\ &= \text{rank}(\mathcal{H}_{L-1}) + \text{rank} \left(\begin{bmatrix} B \\ D \end{bmatrix} \right), \end{aligned}$$

concludes the first part of the proof. The second part immediately follows from Theorem A.1. \blacksquare

The proof of Theorem 3.3 can now be given.

Proof: Consider a linear combination of $x_{[k]}$ and $y_{[k:k+L]}$,

$$A_L x_{[k]} + B_L y_{[k:k+L]}, \quad (3.21)$$

where $A_L \in \mathbb{R}^{n \times n}$ and $B_L \in \mathbb{R}^{n \times p(L+1)}$ have to be determined so that $A_L x_{[k]} + B_L y_{[k:k+L]} = x_{[k+1]}$. Using (3.5), (3.21) is rewritten as

$$(A_L + B_L \mathcal{O}_L) x_{[k]} + B_L \mathcal{H}_L u_{[k:k+L]}. \quad (3.22)$$

Expression (3.22) equals $x_{[k+1]}$ for all possible $x_{[k]}$ and all possible $u_{[k:k+L]}$ if and only if

$$A_L + B_L \mathcal{O}_L = A \quad (3.23)$$

and

$$B_L \mathcal{H}_L = \check{B}. \quad (3.24)$$

It follows from Lemma 3.2 that a matrix B_L satisfying (3.24) exists if and only if condition (3.18) obtains, in which case the general form of the solution is given by $B_L = \check{B} \mathcal{H}_L^{(1)} + Z_L \Sigma_L$ with Z_L an $n \times p(L+1)$ arbitrary matrix parameter. Finally, substituting (3.23) and (3.24) in (3.22), yields (3.19). ■

Since the matrix parameter Z_L can be freely chosen, the matrix \mathcal{K}_L in the linear combination (3.19) is in general not unique. It is unique only if \mathcal{H}_L has full row rank, which for an invertible system can occur only if the system is square ($p = m$) and D is invertible. The matrix \mathcal{H}_L is then invertible so that $\Sigma_L = 0$ and the unique matrix \mathcal{K}_L is given by $\mathcal{K}_L = \check{B} \mathcal{H}_L^{-1} = [BD^{-1} \ 0]$.

It follows from the following lemma that Σ_L is not full rank if $\mathcal{H}_L \neq 0$.

Lemma 3.3. $\text{rank}(\Sigma_L) = p(L+1) - \text{rank}(\mathcal{H}_L)$.

Proof: The lemma immediately follows from Lemma A.1. ■

It follows from Lemma 3.3 that no generality is lost by replacing $Z_L \Sigma_L$ in (3.20) by $\bar{Z} \bar{\Sigma}_L$, where the $(p(L+1) - \text{rank}(\mathcal{H}_L)) \times (p(L+1))$ matrix $\bar{\Sigma}_L$ has full row rank and is row equivalent to Σ_L and where the $n \times (p(L+1) - \text{rank}(\mathcal{H}_L))$ matrix \bar{Z} is arbitrary. This fact may be used to reduce the number of computations when implementing the methods described in this chapter. However, for clarity of exposition, this issue is not addressed any further.

Notice that the condition (3.18) for state reconstructability is very similar to (3.7), i.e. to that of left inversion. Furthermore, since $\text{rank}([B^T \ D^T]) \leq m$, it follows that condition (3.7) is stronger than condition (3.18). This means that system invertibility implies state reconstructability, but not vice versa. Also, it implies that $\text{rank}([B^T \ D^T]) = m$ is a necessary condition for invertibility.

3.5.2 Input reconstruction

The general form of an input reconstructor can be derived in a manner completely analogous to that in the previous section. This yields the following theorem, which basically extends the results of Sain & Massey by giving the general form of the matrix \mathcal{M}_L satisfying $\mathcal{M}_L \mathcal{H}_L = \check{I}_m$.

Theorem 3.4. *If and only if \mathcal{S} is L -delay left invertible, that is, if and only if condition (3.7) obtains, $u_{[k]}$ can be reconstructed as a linear combination of $y_{[k:k+L]}$ and $x_{[k]}$. The general form of the reconstruction can be written as (3.16), where \mathcal{M}_L is given by*

$$\mathcal{M}_L = \check{I}_m \mathcal{H}_L^{(1)} + U_L \Sigma_L, \quad (3.25)$$

with U_L an $m \times p(L+1)$ arbitrary matrix parameter.

Proof: The proof is similar to that of Theorem 3.3 and is omitted. Notice that part of the proof has already been given in the proof of Theorem 3.1. ■

Since the matrix parameter U_L can be freely chosen, the matrix \mathcal{M}_L in the linear combination (3.16) is in general not unique. It is unique only if \mathcal{H}_L has full row rank, in which case it is given by $\mathcal{M}_L = \check{I}_m \mathcal{H}_L^{-1} = [D^{-1} \ 0]$.

3.5.3 A general form of an L -delay left inverse

Combining the state reconstructor of Theorem 3.3 with the input reconstructor of Theorem 3.4, yields the most general form of the dynamical portion.

Theorem 3.5. *Let \mathcal{S} be L -delay left invertible, then the system*

$$\mathcal{S}_L^- : \begin{cases} x_{[k+1]} = (A - \mathcal{K}_L \mathcal{O}_L) x_{[k]} + \mathcal{K}_L y_{[k:k+L]} \\ u_{[k]} = -\mathcal{M}_L \mathcal{O}_L x_{[k]} + \mathcal{M}_L y_{[k:k+L]}, \end{cases} \quad (3.26)$$

with \mathcal{K}_L given by (3.20) and \mathcal{M}_L by (3.25) is a dynamical portion of an L -delay left inverse of \mathcal{S} for all possible values of U_L and Z_L .

As already discussed, if \mathcal{H}_L has full rank, the matrix parameters \mathcal{K}_L and \mathcal{M}_L are unique, so that \mathcal{S}_L^- is unique. In addition, no matter what delay L is chosen, \mathcal{S}_L^- always reduces to (3.15) with the generalized inverses replaced by inverses. This uniqueness will have negative consequences for the developments in the remainder of this chapter. For example, tuning the stability of the inverse by placing its poles is not possible for such systems.

As a special case of Theorem 3.5, consider a system \mathcal{S} with full rank D . Making the choices $L = 0, U_L = 0, Z_L = 0$ and taking as $\{1\}$ -inverse the Moore-Penrose generalized inverse, yields the instantaneous left inverse (3.15).

3.5.3.1 State-space form

Theorem 3.5 yields a general form of the dynamical portion of an L -delay left inverse, and thus not of the inverse system itself. Partitioning \mathcal{M}_L and \mathcal{K}_L as $\mathcal{M}_L = [\mathcal{M}_{L,0} \ \mathcal{M}_{L,1} \ \cdots \ \mathcal{M}_{L,L}]$, and $\mathcal{K}_L = [\mathcal{K}_{L,0} \ \mathcal{K}_{L,1} \ \cdots \ \mathcal{K}_{L,L}]$, with

$\mathcal{M}_{L,i} \in \mathbb{R}^{m \times p}$ and $\mathcal{K}_{L,i} \in \mathbb{R}^{n \times p}$, $i = 0, 1, \dots, L$, it is, however, straightforward to show that

$$\begin{aligned} x_{[k+1]} &= \begin{bmatrix} A - \mathcal{K}_L \mathcal{O}_L & \mathcal{K}_{L,0} & \mathcal{K}_{L,1} & \dots & \mathcal{K}_{L,L-1} \\ 0 & I & & & \\ & 0 & I & & \\ & & \ddots & \ddots & \\ & & & 0 & I \\ & & & & 0 \end{bmatrix} x_{[k]} + \begin{bmatrix} \mathcal{M}_{L,L} \\ 0 \\ 0 \\ \vdots \\ 0 \\ I \end{bmatrix} y_{[k+L]} \\ u_{[k]} &= \begin{bmatrix} -\mathcal{M}_L \mathcal{O}_L & \mathcal{M}_{L,0} & \mathcal{M}_{L,1} & \dots & \mathcal{M}_{L,L-1} \end{bmatrix} x_{[k]} + \mathcal{M}_{L,L} y_{[k+L]} \end{aligned}$$

then is a general state-space form of an L -delay left inverse of \mathcal{S}_L^- .

3.5.3.2 Transfer function

Using the state-space form derived above, it is easily verified that the transfer function of the L -delay left inverse system can be written as

$$H_L(z) = \mathcal{M}_{L,L} - \mathcal{M}_L \mathcal{O}_L [zI - (A - \mathcal{K}_L \mathcal{O}_L)]^{-1} [(I + z^{-L} \mathcal{K}_{L,0} + \dots + z^{-1} \mathcal{K}_{L,L-1}) \mathcal{K}_{L,L}].$$

This transfer function has the property that $H_L(z)H(z) = z^{-L}I$, with $H(z)$ the transfer function of \mathcal{S} .

3.6 Stable inversion

So far, we have actually implicitly assumed that the initial state of \mathcal{S} is known. In this section, we consider the initial state to be unknown. It is then more convenient to consider \mathcal{S}_L^- as a joint input-state estimator and to denote its state vector and output vector by $\hat{x}_{[k]}$ and $\hat{u}_{[k]}$, that is, by estimates of $x_{[k]}$ and $u_{[k]}$. In all applications, it is desired that the estimator is stable such that, starting from any arbitrary estimate $\hat{x}_{[0]}$ of the initial state $x_{[0]}$, the estimates $\hat{x}_{[k]}$ and $\hat{u}_{[k]}$ converge to $x_{[k]}$ and $u_{[k]}$ for $k \rightarrow \infty$. This is the problem addressed in the present section.

This section is outlined as follows. In Sect. 3.6.1, we consider the interpretation of \mathcal{S}_L^- as a joint input-state estimator in more detail. Next, in Sect. 3.6.2, we derive conditions under which the poles of \mathcal{S}_L^- can be assigned.

3.6.1 Joint input-state estimation

For convenience of notation, we rewrite (3.26) in this section as

$$\mathcal{S}_L^- : \begin{cases} \hat{x}_{[k+1]} = A\hat{x}_{[k]} + \mathcal{K}_L(y_{[k:k+L]} - \mathcal{O}_L \hat{x}_{[k]}) \\ \hat{u}_{[k]} = \mathcal{M}_L(y_{[k:k+L]} - \mathcal{O}_L \hat{x}_{[k]}). \end{cases} \quad (3.27)$$

We chose the matrices \mathcal{K}_L and \mathcal{M}_L as given in (3.20) and (3.25), respectively, such that if \mathcal{S} is L -delay left invertible and \mathcal{S}_L^- is initialized with $\hat{x}_{[0]} = x_{[0]}$, it holds that $\hat{x}_{[k]} = x_{[k]}$ and $\hat{u}_{[k]} = u_{[k]}$ for all $k \geq 0$.

Notice that according to our notation, we should actually write $\hat{x}_{[k+1|k+L]}$ and $\hat{u}_{[k|k+L]}$. However, for clarity of exposition, we simply write $\hat{x}_{[k]}$ and $\hat{u}_{[k]}$. For $L = 0$, (3.27) yields a one step ahead predicted state estimate $\hat{x}_{[k+1|k]}$ and a filtered input estimate $\hat{u}_{[k|k]}$. For $L = 1$, it yields a filtered state estimate $\hat{x}_{[k+1|k+1]}$ and a smoothed input estimate $\hat{u}_{[k|k+1]}$. And finally, for $L > 1$, it yields smoothed state estimates and input estimates.

The state equation of \mathcal{S}_L^- can be considered as a state estimator for a system with unknown inputs. When initialized with $\hat{x}_{[0]} = x_{[0]}$, it exactly reconstructs the state sequence of \mathcal{S} under condition (3.18), which is less strict than condition (3.7) for system invertibility. Conditions (3.18) and (3.7) under which \mathcal{S}_L^- reconstructs the system state and both the system state and the system input, respectively, are summarized in Fig. 3.7 for different values of L .

The interpretation of \mathcal{S}_L^- as a joint input-state estimator is schematically shown in Fig. 3.8. Notice that the state estimator and the input estimator of \mathcal{S}_L^- exchange information in one direction, i.e. the input estimator uses the estimate $\hat{x}_{[k]}$ produced by the state estimator, but the state estimator does not use the input estimate produced by the input estimator. As will now be shown, (3.27) can be written in a form where the input estimator and state estimator exchange information in both directions, i.e. in a form where the state estimator uses the estimate produced by the input estimator. First, notice that the general forms (3.20) and (3.25) of \mathcal{K}_L and \mathcal{M}_L are related by $\mathcal{K}_L = B\mathcal{M}_L + \mathcal{L}_L\Sigma_L$ where \mathcal{L}_L is an arbitrary matrix. Consequently, (3.27) can be rewritten as

$$\hat{x}_{[k+1]} = A\hat{x}_{[k]} + B\hat{u}_{[k]} + \mathcal{L}_L\Sigma_L(y_{[k:k+L]} - \mathcal{O}_L\hat{x}_{[k]}) \quad (3.28a)$$

$$\hat{u}_{[k]} = \mathcal{M}_L(y_{[k:k+L]} - \mathcal{O}_L\hat{x}_{[k]}). \quad (3.28b)$$

The state estimator (3.28a) and the input estimator (3.28b) exchange information in both directions. Indeed, the state estimator now also uses the input estimate $\hat{u}_{[k]}$ produced by the input estimator. This is schematically shown by the dashed arrow in Fig. 3.8.

3.6.2 Pole placement

In this section, we assume that \mathcal{S}_L^- is initialized with an arbitrary initial state and establish conditions under which the estimation error converges asymptotically to zero. The derivation in this section can be compared to that of the asymptotic and deadbeat estimators considered in Sect. 2.3.2.

Like in Sect. 2.3.2, we start by deriving an expression for the dynamical evolution of the estimation error. Defining the error in $\hat{x}_{[k]}$ by $\tilde{x}_{[k]} := x_{[k]} - \hat{x}_{[k]}$, it follows from (3.27) that (if condition (3.18) obtains),

$$\tilde{x}_{[k+1]} = (A - \mathcal{K}_L\mathcal{O}_L)\tilde{x}_{[k]}. \quad (3.29)$$

STATE ESTIMATION	INPUT ESTIMATION
General condition	
$\text{rank}(\mathcal{H}_L) = \text{rank}(\mathcal{H}_{L-1}) + \text{rank} \left(\begin{bmatrix} B \\ D \end{bmatrix} \right)$	$\text{rank}(\mathcal{H}_L) = \text{rank}(\mathcal{H}_{L-1}) + m$
One step ahead prediction	Instantaneous estimation
$\text{rank}(D) = \text{rank} \left(\begin{bmatrix} B \\ D \end{bmatrix} \right)$	$\text{rank}(D) = m$
Filtering	One step delayed estimation
Direct feedthrough: $D \neq 0$	
$\text{rank} \left(\begin{bmatrix} D & 0 \\ CB & D \end{bmatrix} \right) = \text{rank}(D) + \text{rank} \left(\begin{bmatrix} B \\ D \end{bmatrix} \right)$	$\text{rank} \left(\begin{bmatrix} D & 0 \\ CB & D \end{bmatrix} \right) = \text{rank}(D) + m$
No direct feedthrough: $D = 0$	
$\text{rank}(CB) = \text{rank}(B)$	$\text{rank}(CB) = m$

Figure 3.7: Comparison between (3.7) and (3.18) for $L = 0$ and $L = 1$.

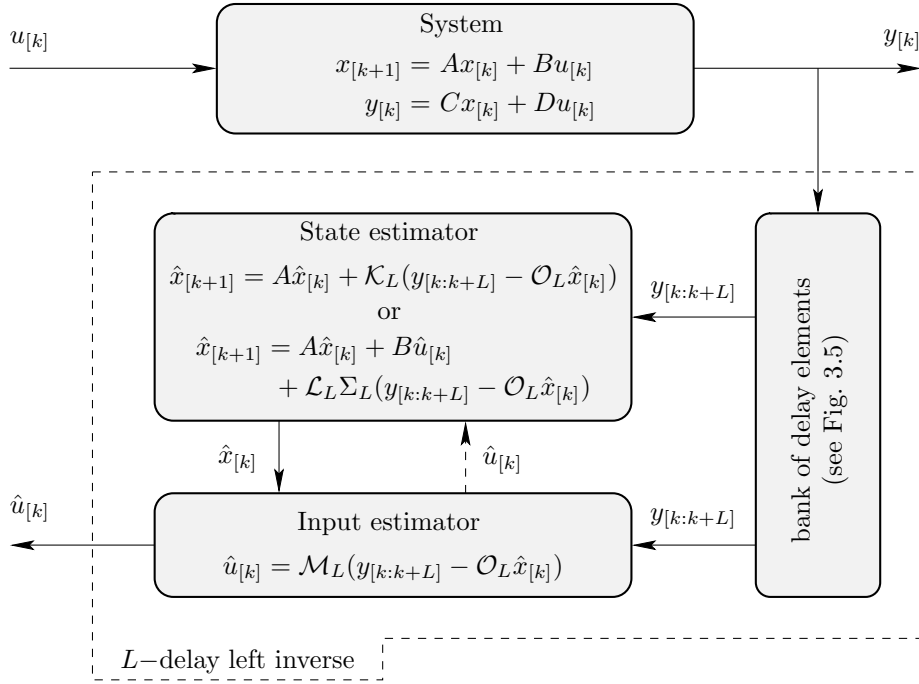


Figure 3.8: Interpretation of the inverse system as a joint input-state estimator

Similarly, defining $\tilde{u}_{[k]} := u_{[k]} - \hat{u}_{[k]}$, it follows from (3.27) that (if \mathcal{S} is L -delay left invertible),

$$\tilde{u}_{[k]} = -\mathcal{M}_L \mathcal{O}_L \tilde{x}_{[k]}. \quad (3.30)$$

Let \mathcal{S} be L -delay left invertible, then it follows from (3.29) and (3.30) that $\tilde{x}_{[k]} \rightarrow 0$ and $\tilde{u}_{[k]} \rightarrow 0$ for $k \rightarrow \infty$ if and only if all eigenvalues of $A - \mathcal{K}_L \mathcal{O}_L$ lie inside the unit circle. Since

$$A - \mathcal{K}_L \mathcal{O}_L = (A - \check{\mathcal{B}}\mathcal{H}_L^{(1)}\mathcal{O}_L) - Z_L(\Sigma_L \mathcal{O}_L), \quad (3.31)$$

the position of the eigenvalues of $A - \mathcal{K}_L \mathcal{O}_L$ can be influenced by the choice of the matrix parameter Z_L . The freedom in the choice of Z_L thus allows to tune the eigenvalues of $A - \mathcal{K}_L \mathcal{O}_L$ (and consequently also the poles of \mathcal{S}_L^-).

Conditions are now derived under which Z_L can be chosen so that all eigenvalues of $A - \mathcal{K}_L \mathcal{O}_L$ lie inside the unit circle. The following corollary immediately follows from Corollary 2.1.

Corollary 3.1.

1. If and only if $\{A - \check{\mathcal{B}}\mathcal{H}_L^{(1)}\mathcal{O}_L, \Sigma_L \mathcal{O}_L\}$ is observable, Z_L can be chosen so that all eigenvalues of $A - \mathcal{K}_L \mathcal{O}_L$ are assigned at any desired location.

2. If and only if $\{A - \check{B}\mathcal{H}_L^{(1)}\mathcal{O}_L, \Sigma_L\mathcal{O}_L\}$ is detectable, Z_L can be chosen so that all eigenvalues of $A - \mathcal{K}_L\mathcal{O}_L$ lie inside the unit circle.

A relation between the unobservable modes of $\{A - \check{B}\mathcal{H}_L^{(1)}\mathcal{O}_L, \Sigma_L\mathcal{O}_L\}$ and the rank of a matrix pencil is given in the following lemma.

Lemma 3.4. *Let condition (3.18) obtain. Then, the unobservable modes $\lambda \in \mathbb{C}$ of $\{A - \check{B}\mathcal{H}_L^{(1)}\mathcal{O}_L, \Sigma_L\mathcal{O}_L\}$ satisfy*

$$\text{rank} \left(\begin{bmatrix} \lambda I - A & B \\ -C & D \end{bmatrix} \right) < n + \text{rank} \left(\begin{bmatrix} B \\ D \end{bmatrix} \right). \quad (3.32)$$

Proof: The unobservable modes of $\{A - \check{B}\mathcal{H}_L^{(1)}\mathcal{O}_L, \Sigma_L\mathcal{O}_L\}$ are those $\lambda \in \mathbb{C}$ for which

$$\text{rank} \left(\begin{bmatrix} \lambda I - (A - \check{B}\mathcal{H}_L^{(1)}\mathcal{O}_L) \\ -\Sigma_L\mathcal{O}_L \end{bmatrix} \right) < n.$$

This implies

$$\text{rank} \left(\begin{bmatrix} \lambda I - (A - \check{B}\mathcal{H}_L^{(1)}\mathcal{O}_L) & \check{B} \\ -\Sigma_L\mathcal{O}_L & \mathcal{H}_L \end{bmatrix} \right) < n + \text{rank} \left(\begin{bmatrix} \check{B} \\ \mathcal{H}_L \end{bmatrix} \right). \quad (3.33)$$

Using the fact that

$$\begin{aligned} & \text{rank} \left(\begin{bmatrix} \lambda I - (A - \check{B}\mathcal{H}_L^{(1)}\mathcal{O}_L) & \check{B} \\ -\Sigma_L\mathcal{O}_L & \mathcal{H}_L \end{bmatrix} \right) \\ &= \text{rank} \left(\begin{bmatrix} \lambda I - A & \check{B} \\ -\mathcal{O}_L & \mathcal{H}_L \end{bmatrix} \begin{bmatrix} I & 0 \\ \mathcal{H}_L^{(1)}\mathcal{O}_L & I \end{bmatrix} \right), \\ &= \text{rank} \left(\begin{bmatrix} \lambda I - A & \check{B} \\ -\mathcal{O}_L & \mathcal{H}_L \end{bmatrix} \right), \end{aligned}$$

(3.33) is rewritten as

$$\text{rank} \left(\begin{bmatrix} \lambda I - A & \check{B} \\ -\mathcal{O}_L & \mathcal{H}_L \end{bmatrix} \right) < n + \text{rank} \left(\begin{bmatrix} \check{B} \\ \mathcal{H}_L \end{bmatrix} \right),$$

or, because (3.18) obtains, as

$$\text{rank} \left(\begin{bmatrix} \lambda I - A & \check{B} \\ -\mathcal{O}_L & \mathcal{H}_L \end{bmatrix} \right) < n + \text{rank}(\mathcal{H}_L). \quad (3.34)$$

On the other hand, it is easily verified that

$$\begin{aligned} \text{rank} \left(\begin{bmatrix} \lambda I - A & \check{B} \\ -\mathcal{O}_L & \mathcal{H}_L \end{bmatrix} \right) &= \text{rank} \left(\begin{bmatrix} \lambda I - A & B & 0 \\ -C & D & 0 \\ -\mathcal{O}_{L-1}A & \mathcal{O}_{L-1}B & \mathcal{H}_{L-1} \end{bmatrix} \right), \\ &\geq \text{rank} \left(\begin{bmatrix} \lambda I - A & B \\ -C & D \end{bmatrix} \right) + \text{rank}(\mathcal{H}_{L-1}). \end{aligned} \quad (3.35)$$

Combining (3.34) and (3.35), yields

$$\text{rank} \left(\begin{bmatrix} \lambda I - A & B \\ -C & D \end{bmatrix} \right) + \text{rank}(\mathcal{H}_{L-1}) < n + \text{rank}(\mathcal{H}_L), \quad (3.36)$$

which, under condition (3.18), can be rewritten as (3.32). ■

Notice, very importantly, that Lemma 3.4 only holds in one direction. That is, all unobservable modes λ of $\{A - \check{B}\mathcal{H}_L^{(1)}\mathcal{O}_L, \Sigma_L\mathcal{O}_L\}$ satisfy (3.32), but not necessarily vice versa. In other words, Lemma 3.4 does not exclude the existence of λ for which (3.32) obtains, but which are observable modes of $\{A - \check{B}\mathcal{H}_L^{(1)}\mathcal{O}_L, \Sigma_L\mathcal{O}_L\}$. We will come back to this further in the text.

We consider the implications of Lemma 3.4 separately for state estimation and input estimation.

3.6.2.1 State estimation

Considering state estimation, we have from Corollary 3.1 and Lemma 3.4 the following theorem.

Theorem 3.6. *Let condition (3.18) obtains. Then, if*

$$\text{rank} \left(\begin{bmatrix} \lambda I - A & B \\ -C & D \end{bmatrix} \right) = n + \text{rank} \left(\begin{bmatrix} B \\ D \end{bmatrix} \right) \quad (3.37)$$

$\forall \lambda \in \mathbb{C}$ with $|\lambda| \geq 1$, the matrix parameter Z_L can be chosen so that the state estimator (i.e. the state equation) of \mathcal{S}_L^- is stable. If (3.37) holds $\forall \lambda \in \mathbb{C}$, the poles of the state estimator can be arbitrary assigned.

Similar results were derived in [68] for disturbance decoupled filtering and in [126] for (reduced order) time delayed state estimation.

3.6.2.2 Input estimation

As will now be shown, for a left invertible system, Lemma 3.4 yields a relation between the unobservable modes of $\{A - \check{B}\mathcal{H}_L^{(1)}\mathcal{O}_L, \Sigma_L\mathcal{O}_L\}$ and the transmission zeros of \mathcal{S} .

Proposition 3.2. *If \mathcal{S} is L -delay left invertible, the unobservable modes of $\{A - \check{B}\mathcal{H}_L^{(1)}\mathcal{O}_L, \Sigma_L\mathcal{O}_L\}$ are transmission zeros of \mathcal{S} .*

Proof: For an L -delay left invertible system, condition (3.32) becomes

$$\text{rank} \left(\begin{bmatrix} \lambda I - A & B \\ -C & D \end{bmatrix} \right) < n + m. \quad (3.38)$$

The λ satisfying (3.38) are transmission zeros of \mathcal{S} . ■

Notice again that Proposition 3.2 only holds in one directions. That is, Proposition 3.2 does not exclude the existence of L for which the zeros of \mathcal{S} are observable modes of $\{A - \check{B}\mathcal{H}_L^{(1)}\mathcal{O}_L, \Sigma_L\mathcal{O}_L\}$.

The following theorem is a direct consequence of Theorems 3.4 and 2.2 and Proposition 3.2.

Theorem 3.7. *Consider an L -delay left invertible system \mathcal{S} . If all the zeros of \mathcal{S} are at infinity, Z_L can be chosen so that all the poles of \mathcal{S}_L^- are arbitrary assigned. If \mathcal{S} has no unstable zeros, Z_L can be chosen so that \mathcal{S}_L^- is stable.*

Now, let λ be an unobservable modes of $\{A - \check{B}\mathcal{H}_L^{(1)}\mathcal{O}_L, \Sigma_L\mathcal{O}_L\}$. Then, it follows from Theorem 2.2 that $\lambda \in \Lambda(A - \check{B}\mathcal{H}_L^{(1)}\mathcal{O}_L - Z_L\Sigma_L\mathcal{O}_L)$ for all Z_L , or equivalently, $\lambda \in \Lambda(A - \mathcal{K}_L\mathcal{O}_L)$ for all Z_L .

We then conclude from Proposition 3.2 that the set of eigenvalues of $A - \mathcal{K}_L\mathcal{O}_L$ contains transmission zeros of \mathcal{S} .

3.7 Stable reduced order inversion

Silverman [116] noticed that the order of the dynamical portion obtained with his structure algorithm can be reduced so that an inverse system is obtained which can be realized with exactly the same number of delay elements as the original system. His approach to reduced order inversion is based on calculating part of the state vector directly from the measurements. An approach to the design of inverses of the lowest possible order is given by Yuan [140] and Emre & Silverman [35]. Their methods are based on the concepts of elementary null sequences and minimal dynamical covers, respectively. In [129], minimality of the inverses of singular systems is addressed.

In this section, the problem of reducing the order of the dynamical portion \mathcal{S}_L^- is addressed. Like the approach of Silverman, the reduction is based on calculating part of the state vector directly from the measurements. The procedure has some similarities to the classical design of reduced order state observers [92].

This section is outlined as follows. In Sect. 3.7.1, we show how part of the state vector can be calculated directly from the measurements and derive the reduced order dynamical portion. Next, in Sect. 3.7.2, pole placement of the reduced order dynamical portion is addressed.

3.7.1 Reduced order inversion

First, notice that (3.5) can be decoupled from $u_{[k:k+L]}$ by pre-multiplying left and right hand side by Σ_L , which yields

$$\Sigma_L y_{[k:k+L]} = \Sigma_L \mathcal{O}_L x_{[k]}. \quad (3.39)$$

Equation (3.39) yields a relation between $x_{[k]}$ and $y_{[k:k+L]}$ which allows to calculate part of $x_{[k]}$ directly from the measurements, as will now be shown. Let the singular value decomposition of $\Sigma_L \mathcal{O}_L$ be given by

$$\Sigma_L \mathcal{O}_L = X \begin{bmatrix} \Xi & 0 \\ 0 & 0 \end{bmatrix} V^\top, \quad (3.40)$$

where $X \in \mathbb{R}^{p(L+1) \times p(L+1)}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal and where $\Xi \in \mathbb{R}^{r_L \times r_L}$ contains the $r_L := \text{rank}(\Sigma_L \mathcal{O}_L)$ singular values on its diagonal. Notice that (3.40) can be rewritten as

$$\Sigma_L \mathcal{O}_L = \tilde{X} \begin{bmatrix} I_{r_L} & 0 \\ 0 & 0 \end{bmatrix} V^\top,$$

where the i -th column of \tilde{X} equals the i -th column of X multiplied by the i -th singular value, $i = 1, 2, \dots, r_L$. Consequently, pre-multiplying (3.39) by \tilde{X}^{-1} , yields

$$\tilde{X}^{-1} \Sigma_L y_{[k:k+L]} = \begin{bmatrix} I_{r_L} & 0 \\ 0 & 0 \end{bmatrix} V^\top x_{[k]}. \quad (3.41)$$

Define $\bar{x}_{[k]} := V^\top x_{[k]}$, and partition $\bar{x}_{[k]}$ as $\bar{x}_{[k]} = [\bar{x}_{1[k]}^\top \bar{x}_{2[k]}^\top]^\top$ with $\bar{x}_{1[k]} \in \mathbb{R}^{r_L}$ and $\bar{x}_{2[k]} \in \mathbb{R}^{n-r_L}$. Also, define $\bar{X} := \tilde{X}^{-1} \Sigma_L$ and partition \bar{X} as $\bar{X} = [\bar{X}_1^\top \bar{X}_2^\top]^\top$ with $\bar{X}_1 \in \mathbb{R}^{r_L \times p(L+1)}$ and $\bar{X}_2 \in \mathbb{R}^{(p(L+1)-r_L) \times p(L+1)}$. Then, it follows from (3.41) that

$$\bar{x}_{1[k]} = \bar{X}_1 y_{[k:k+L]}, \quad (3.42)$$

meaning that $\bar{x}_{1[k]}$ can be computed directly from $y_{[k:k+L]}$.

The remaining part of the derivation is closely related to the design of reduced order observers in [92]. Since $\bar{x}_{1[k]}$ can be computed directly from $y_{[k:k+L]}$, we consider as state equation of the reduced order dynamical portion a dynamic equation for $\bar{x}_{2[k]}$. First, notice that under a similarity transformation with transformation matrix V^\top , the pair $\{A - \mathcal{K}_L \mathcal{O}_L, \mathcal{K}_L\}$ becomes $\{\bar{A}, \bar{B}\}$ with $\bar{A} := V^\top (A - \mathcal{K}_L \mathcal{O}_L) V$ and $\bar{B} := V^\top \mathcal{K}_L$. Consequently, under the similarity transformation, the state equation of \mathcal{S}_L^- becomes

$$\begin{bmatrix} \bar{x}_{1[k+1]} \\ \bar{x}_{2[k+1]} \end{bmatrix} = \begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} \\ \bar{A}_{21} & \bar{A}_{22} \end{bmatrix} \begin{bmatrix} \bar{x}_{1[k]} \\ \bar{x}_{2[k]} \end{bmatrix} + \begin{bmatrix} \bar{B}_1 \\ \bar{B}_2 \end{bmatrix} y_{[k:k+L]}, \quad (3.43)$$

where $\bar{A}_{11} \in \mathbb{R}^{r_L \times r_L}$, $\bar{A}_{12} \in \mathbb{R}^{r_L \times (n-r_L)}$, $\bar{A}_{21} \in \mathbb{R}^{(n-r_L) \times r_L}$, $\bar{A}_{22} \in \mathbb{R}^{(n-r_L) \times (n-r_L)}$, $\bar{B}_1 \in \mathbb{R}^{r_L \times p(L+1)}$, and $\bar{B}_2 \in \mathbb{R}^{(n-r_L) \times p(L+1)}$. The state equation of the reduced order dynamical portion is then easily extracted from (3.43) and (3.42),

$$\bar{x}_{2[k+1]} = \bar{A}_{22} \bar{x}_{2[k]} + (\bar{B}_2 + \bar{A}_{21} \bar{X}_1) y_{[k:k+L]}. \quad (3.44)$$

For the output equation of the reduced dynamical portion, we first rewrite the output equation of \mathcal{S}_L^- as

$$u_{[k]} = -\mathcal{M}_L \mathcal{O}_L V \bar{x}_{[k]} + \mathcal{M}_L y_{[k:k+L]}.$$

Defining $\bar{C} := -\mathcal{M}_L \mathcal{O}_L V$ and partitioning \bar{C} as $\bar{C} = [\bar{C}_1 \ \bar{C}_2]$, with $\bar{C}_1 \in \mathbb{R}^{m \times r_L}$ and $\bar{C}_2 \in \mathbb{R}^{m \times (n-r_L)}$, yields

$$u_{[k]} = \bar{C}_1 \bar{x}_{1[k]} + \bar{C}_2 \bar{x}_{2[k]} + \mathcal{M}_L y_{[k:k+L]}. \quad (3.45)$$

Substituting (3.42) in (3.45), yields the following output equation of the reduced order dynamical portion,

$$u_{[k]} = \bar{C}_2 \bar{x}_{2[k]} + (\mathcal{M}_L + \bar{C}_1 \bar{X}_1) y_{[k:k+L]}. \quad (3.46)$$

Finally, combining (3.46) and (3.44) yields the dynamical portion

$$\bar{x}_{2[k+1]} = \bar{A}_{22} \bar{x}_{2[k]} + (\bar{B}_2 + \bar{A}_{21} \bar{X}_1) y_{[k:k+L]} \quad (3.47a)$$

$$u_{[k]} = \bar{C}_2 \bar{x}_{2[k]} + (\mathcal{M}_L + \bar{C}_2 \bar{X}_1) y_{[k:k+L]}, \quad (3.47b)$$

of order $n - r_L$. Even if all eigenvalues of $A - \mathcal{K}_L \mathcal{O}_L$ lie inside the unit circle, this dynamical portion can, however, be unstable. The problem of assigning the poles of the reduced order dynamical portion is addressed in the next section.

3.7.2 Stable reduced order inversion

The approach of assigning the poles of the reduced order dynamical portion is closely related to that in [92]. Notice from (3.43) that (3.44) can be written as

$$\begin{aligned} \bar{x}_{2[k+1]} &= \bar{A}_{22} \bar{x}_{2[k]} + \bar{A}_{21} \bar{x}_{1[k]} + \bar{B}_2 y_{[k:k+L]} \\ &\quad + N(\bar{x}_{1[k+1]} - \bar{A}_{11} \bar{x}_{1[k]} - \bar{A}_{12} \bar{x}_{2[k]} - \bar{B}_1 y_{[k:k+L]}) \end{aligned} \quad (3.48)$$

$$\begin{aligned} &= (\bar{A}_{22} - N \bar{A}_{12}) \bar{x}_{2[k]} + \bar{B}_2 y_{[k:k+L]} + \bar{A}_{21} \bar{x}_{1[k]} \\ &\quad + N(\bar{x}_{1[k+1]} - \bar{A}_{11} \bar{x}_{1[k]} - \bar{B}_1 y_{[k:k+L]}), \end{aligned} \quad (3.49)$$

where N is an arbitrary matrix. The freedom in the choice of N will be used to assign the poles of the reduced order dynamical portion. The term $\bar{x}_{1[k+1]}$ in (3.49) can be eliminated by defining $s_{[k]} := \bar{x}_{2[k]} - N \bar{x}_{1[k]}$. Substituting the latter equation in (3.49) yields together with (3.42) the following state equation,

$$s_{[k+1]} = (\bar{A}_{22} - N \bar{A}_{12}) s_{[k]} + \tilde{B} y_{[k:k+L]}, \quad (3.50)$$

where $\tilde{B} := \bar{B}_2 - N \bar{B}_1 + \bar{A}_{21} \bar{X}_1 - N \bar{A}_{11} \bar{X}_1 + \bar{A}_{22} N \bar{X}_1 - N \bar{A}_{12} N \bar{X}_1$.

For the output equation, we first rewrite (3.45) as

$$u_{[k]} = \bar{C}_1 \bar{x}_{1[k]} + \bar{C}_2 (s_{[k]} + N \bar{x}_{1[k]}) + \mathcal{M}_L y_{[k:k+L]}. \quad (3.51)$$

Substituting (3.42) in (3.51), yields the following output equation of the reduced order dynamical portion,

$$u_{[k]} = \bar{C}_2 s_{[k]} + \tilde{C} y_{[k:k+L]}, \quad (3.52)$$

where $\tilde{C} := \mathcal{M}_L + \bar{C}_1 \bar{X}_1 + \bar{C}_2 N \bar{X}_1$.

Finally, combining (3.50) and (3.52) yields the reduced order dynamical portion

$$s_{[k+1]} = (\bar{A}_{22} - N\bar{A}_{12})s_{[k]} + \tilde{B}y_{[k:k+L]} \quad (3.53a)$$

$$u_{[k]} = \tilde{C}_2 s_{[k]} + \tilde{C}y_{[k:k+L]}, \quad (3.53b)$$

of order $n - r_L$. Notice that $x_{[k]}$ can be reconstructed from $s_{[k]}$ and $y_{[k:k+L]}$ as

$$x_{[k]} = V \begin{bmatrix} \bar{X}_1 y_{[k:k+L]} \\ s_{[k]} + N\bar{X}_1 y_{[k:k+L]} \end{bmatrix}.$$

Conditions are now derived under which the poles of (3.53) can be assigned by the choice of the matrix parameter N . The derivation is based on the following lemma.

Lemma 3.5. *If $\{A - \check{B}\mathcal{H}_L^{(1)}\mathcal{O}_L, \Sigma_L\mathcal{O}_L\}$ is observable, then so is $\{\bar{A}_{22}, \bar{A}_{12}\}$.*

Proof: First, notice that $\{A - \check{B}\mathcal{H}_L^{(1)}\mathcal{O}_L, \Sigma_L\mathcal{O}_L\}$ is observable if and only if $\{A - \mathcal{K}_L\mathcal{O}_L, \Sigma_L\mathcal{O}_L\}$ is observable. The remainder of the proof follows by considering the system consisting of (3.19) and (3.39),

$$x_{[k+1]} = (A - \mathcal{K}_L\mathcal{O}_L)x_{[k]} + \mathcal{K}_L y_{[k:k+L]} \quad (3.54a)$$

$$\Sigma_L y_{[k:k+L]} = \Sigma_L \mathcal{O}_L x_{[k]}, \quad (3.54b)$$

and proving that this system is observable if and only if $\{\bar{A}_{22}, \bar{A}_{12}\}$ is observable. The proof is very similar to that in [92] and is hence omitted. ■

The following theorem is a direct consequence of Lemma 3.5 and Lemma 3.2.

Theorem 3.8. *Let \mathcal{S} be an n -th order L -delay left invertible system with all its zeros at infinity. Then, there exists an $(n - \text{rank}(\Sigma_L\mathcal{O}_L))$ -th order dynamical portion of the form (3.53) whose poles can be assigned by the choice of N .*

It follows from Theorem 3.8 that if $\text{rank}(\Sigma_L\mathcal{O}_L) = n$, there exists a dynamical portion of order 0. Indeed, if $\text{rank}(\Sigma_L\mathcal{O}_L) = n$, it follows from (3.39) that $x_{[k]}$ can be reconstructed from $y_{[k:k+L]}$ as

$$x_{[k]} = (\Sigma_L\mathcal{O}_L)^\dagger \Sigma_L y_{[k:k+L]}. \quad (3.55)$$

Substituting (3.55) in the output equation of (3.26), yields

$$u_{[k]} = \mathcal{M}_L(I - \mathcal{O}_L(\Sigma_L\mathcal{O}_L)^\dagger \Sigma_L)y_{[k:k+L]}, \quad (3.56)$$

which is the desired 0-th order dynamical portion.

3.8 Numerical examples

We consider three numerical examples. The first example deals with instantaneous inversion, the second example with time-delayed inversion and the third example with stable inversion of an NMP system.

Example 3.1. Instantaneous inversion

Consider the minimal LTI system

$$x_{[k+1]} = \underbrace{\begin{bmatrix} 0.6 & -0.45 \\ 0 & 0.3 \end{bmatrix}}_A x_{[k]} + \underbrace{\begin{bmatrix} 1 \\ -1 \end{bmatrix}}_B u_{[k]} \quad (3.57a)$$

$$y_{[k]} = \underbrace{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}}_C x_{[k]} + \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_D u_{[k]}. \quad (3.57b)$$

The system (3.57) has poles at 0.6 and 0.3 and is thus stable. Since D is full column rank, (3.57) has an instantaneous left inverse. The instantaneous left inverse (3.15), given by

$$\begin{aligned} x_{[k+1]} &= \begin{bmatrix} 0.3 & -1.45 \\ 0 & 1.3 \end{bmatrix} x_{[k]} + \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} y_{[k]} \\ u_{[k]} &= \begin{bmatrix} 0 & -1 \end{bmatrix} x_{[k]} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} y_{[k]}, \end{aligned}$$

however, has a pole at 1.3 and is thus unstable.

However, since all zeros of (3.57) lie at infinity, it follows from Theorem 3.7 that there exists a stable instantaneous left inverse of which the poles can be arbitrary placed. Suppose that the following complex conjugate poles are desired: $0.7 \pm 0.6i$. These poles can be assigned by calculating the matrix Z_0 so that the eigenvalues of $(A - BD^\dagger C) - Z_0(I - DD^\dagger)C$ are at the desired locations, which can be achieved using pole placement. Together with the choice $U_0 = 0$, this yields the instantaneous full order inverse system

$$x_{[k+1]} = \begin{bmatrix} -0.86 & -2.91 \\ 0.96 & 2.26 \end{bmatrix} x_{[k]} + \begin{bmatrix} 1.46 & 1 \\ -0.96 & -1 \end{bmatrix} y_{[k]}, \quad (3.59a)$$

$$u_{[k]} = \begin{bmatrix} 0 & -1 \end{bmatrix} x_{[k]} + \begin{bmatrix} 0 & 1 \end{bmatrix} y_{[k]}. \quad (3.59b)$$

In Figure 3.9, the inverse system (3.59) is simulated starting from an arbitrary initial state, but with inputs equal to the outputs of (3.57). Since the inverse system is stable, the output of (3.59) converges to the input of (3.57). Convergence is rather slow because the poles of (3.59) have been chosen close to the unit circle.

Notice that the observability and controllability matrix of (3.59) have full rank. However, there exist left inverses of lower order than (3.59). Indeed, since $\text{rank}(\Sigma_0 \mathcal{O}_0) = \text{rank}((I - DD^\dagger)C) = 1$, it follows from Theorem 3.8 that there

exists an instantaneous left inverse of order 1 of which the poles can be arbitrary assigned. Suppose now that a first order left inverse with the *deadbeat* property is desired, that is, starting from an arbitrary initial state, the inverse system should exactly reconstructs the state vector and the input vector of (3.57) from time instant $k = 1$ on. This is achieved by placing the pole of the reduced order inverse at 0, which yields

$$\begin{aligned}\bar{x}_{[k+1]} &= 0\bar{x}_{[k]} + \begin{bmatrix} 1.4708 & -1.4142 \end{bmatrix} y_{[k]} \\ u_{[k]} &= -0.7071\bar{x}_{[k]} + \begin{bmatrix} 1.7333 & 1.0000 \end{bmatrix} y_{[k]}.\end{aligned}$$

Notice that in order to realize this inverse system, only 1 delay element is needed. Using Silverman's structure algorithm, at least 2 delay elements would be needed.

Now, we allow a delay of one step in the reconstruction. Because $\text{rank}(\Sigma_1 \mathcal{O}_1) = 2$, it follows from Theorem 3.8 that there exists a 1–delay left inverse without dynamical portion. Applying (3.56) with $U_1 = 0$, yields

$$u_{[k]} = \begin{bmatrix} -0.8 & 1 & 1.333 & 0 \end{bmatrix} y_{[k:k+1]}.$$

This result indicates that the rank of $\Sigma_L \mathcal{O}_L$ increases with L and thus, that by increasing L , an inverse without dynamical portion may be found. \square

Example 3.2. Time-delayed inversion

Consider again the system (3.57), but now with $D = 0$. This minor change increases the inherent delay of the system from 0 to 1. Since for the resulting system $\text{rank}(\Sigma_1 \mathcal{O}_1) = 2$, it follows that there exists a 1–delay left inverse without dynamical portion. Applying (3.56) with $U_1 = 0$, yields

$$u_{[k]} = \begin{bmatrix} -0.070 & 0.212 & -0.117 & -1 \end{bmatrix} y_{[k:k+1]}.$$

\square

Example 3.3. Inversion of a nonminimum phase system

Consider the observable and controllable LTI SISO system

$$x_{[k+1]} = \begin{bmatrix} 0.5 & 0.25 \\ 0.25 & -0.6 \end{bmatrix} x_{[k]} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u_{[k]} \quad (3.60a)$$

$$y_{[k]} = \begin{bmatrix} -1 & 0.1 \end{bmatrix} x_{[k]}. \quad (3.60b)$$

The system (3.60) has a transmission zero at 3 and is thus NMP. Its inherent delay equals 1, however the pair $\{A - \check{B}\mathcal{H}_1^{(1)}\mathcal{O}_1, \Sigma_1 \mathcal{O}_1\}$ is not detectable so that there does not exist a stable full order 1–delay left inverse. Since the unobservable modes of the pair are zeros of the system, one of the poles of the inverse system will equal the zero of the system, so that the 1–delay left inverse system is unstable.

Figure 3.10 plots the condition number (i.e. the ratio of the maximal singular value to the minimal singular value) of $\Sigma_L \mathcal{O}_L$ as function of L . We observe that

for small L , the condition number is larger than the inverse of machine precision, indicating that $\Sigma_L \mathcal{O}_L$ is singular. The ratio decreases with increasing L and becomes constant from $L = 30$. For $L = 30$, $\Sigma_L \mathcal{O}_L$ is clearly nonsingular, which means that we can reconstruct $u_{[k]}$ from knowledge of $y_{[k:k+30]}$ using (3.56).

Because we are dealing with an NMP system, the input can however not be exactly reconstructed. Indeed, there exists a number g and a vector $x_0 \in \mathbb{R}^2$ so that $[x_0^T \ g]^T$ lies in the null space of

$$\begin{bmatrix} A - 3I & B \\ C & D \end{bmatrix}.$$

A basis for the null space is given by $[-0.0269268 \ -0.26928 \ -0.96268]^T$. Consequently, the input signals u_r and $u_r + u_z$, with u_r arbitrary and

$$u_{z[k]} = \begin{cases} -0.96268 & \text{for } k = 0 \\ -0.96268 \times 3^k & \text{for } k = 1, 2, \dots \end{cases}$$

yield exactly the same output when (3.60) is initialized with $x_{[0]} = [-0.0269268 - 0.26928]^T$. Since u_z grows unbounded, unique input reconstruction is possible only if prior knowledge is available that the input applied to the system is bounded. \square

3.9 Conclusion

A new procedure for left inversion of linear discrete-time systems in state-space form was introduced. The procedure is most closely related to that of Sain and Massey [115], in the sense that inverse systems with a similar structure are considered, that is, inverse systems consisting of a bank of delay elements, followed by a dynamical system. An important contribution is the derivation of the most general form of such a dynamical system.

Conditions were derived under which the poles of the inverse system can be assigned. It was shown that pole placement is possible if a certain matrix pair is observable. This pair turns out to be observable if the original system has no zeros.

Based on the theory of reduced order observers, a technique was developed to simultaneously reduce the order of the inverse system and place its poles. A condition was derived under which an inverse system without dynamical portion exists. Further research should investigate the relation between the order of the inverse system and the delay of reconstruction in more detail, e.g. by how much can the order of the inverse system be decreased given a certain allowable increase in the delay of reconstruction?

The results of this chapter not only have direct implications for joint input-state estimation, but also for optimal state estimation in the presence of unknown inputs. A condition was derived under which the state vector of a system with unknown inputs can be reconstructed from knowledge of the system

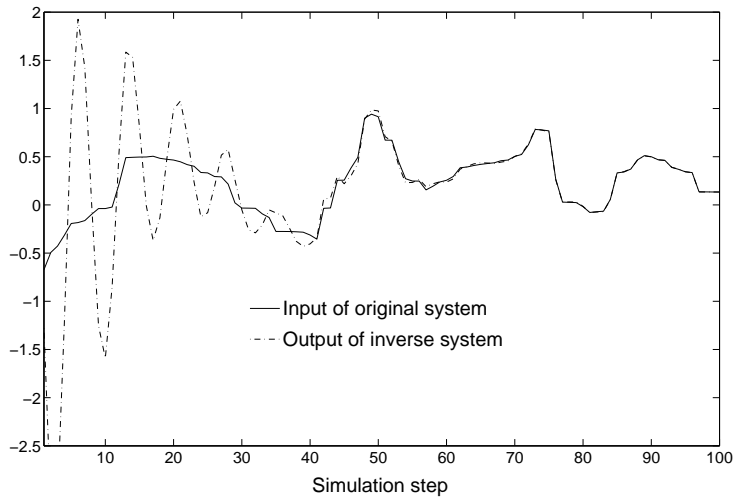


Figure 3.9: Starting from an arbitrary initial state, the output of the left inverse (3.59) converges to the input of the original system (3.57).

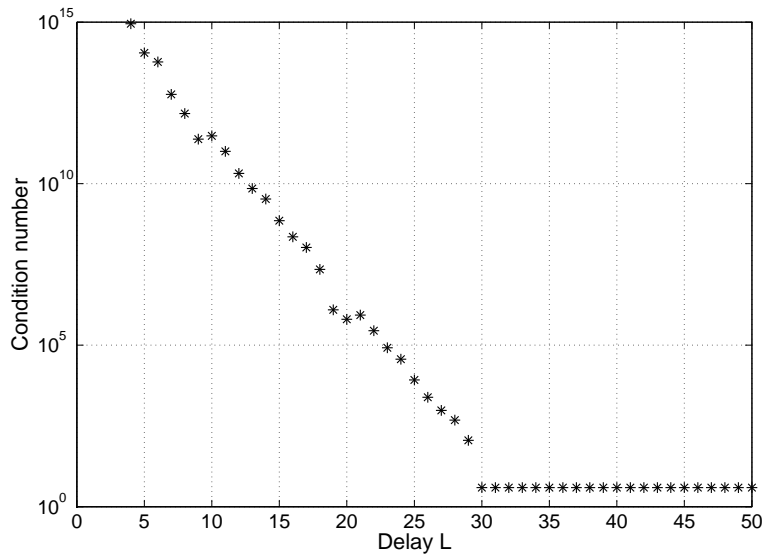


Figure 3.10: Condition number of $\Sigma_L \mathcal{O}_L$ as function of L for the NMP system (3.60). It is found that $\Sigma_L \mathcal{O}_L$ has full rank for $L \geq 30$, so that we can invert the NMP system (3.60) using (3.56). However, because we are dealing with an NMP system, the input can not be uniquely reconstructed.

outputs. In particular, it was shown that input reconstructability implies state reconstructability, but not vice versa.

Although only left inversion has been studied, it is expected that most results translate easily to the dual problem of right inversion. Also, it is expected that most results translate to continuous-time systems.

Chapter 4

Inversion of Combined Deterministic-Stochastic Systems

This chapter extends the inversion procedure of Chapter 3 to combined deterministic-stochastic systems, where the objective is to optimally reconstruct the deterministic inputs from knowledge of the noisy outputs. Optimal recursive state estimators for such systems have been extensively studied in literature. One of the main contributions of this chapter is the extension to joint input-state estimation. Another important contribution is the establishment of a relation between the joint input-state estimators and least-squares estimation. Based on this relation, information and square-root information formulas are derived almost instantaneously. In a final contribution, a unified framework for state estimation and joint input-state estimation in the context of both filtering and smoothing is established.

4.1 Introduction

This chapter extends the inversion procedure developed in Chapter 3 to combined deterministic-stochastic systems. With inverting such a system, we mean estimation of the unknown deterministic input based on the measurements. The step from the previous chapter to this chapter can be compared to that from asymptotic or deadbeat estimation to the Kalman filter. Indeed, due to the noise it is not possible to exactly or asymptotically reconstruct the input of a combined systems. Hence, similar the Kalman filtering problem, it is most convenient to determine the estimates that satisfy a certain optimality criterion.

The choice of the optimality criterion may reflect prior knowledge about the

unknown input. For example, if it is known that the amplitude of the input is small, one may consider minimizing the norm of the input. Another type of prior knowledge may for example be that the unknown input is constant in time. The estimation problem, referred to as optimal estimation in the presence of constant bias, can then be reduced to a standard state estimation problem. This problem will be discussed in more detail in Chapter 5.

In many applications, like e.g. in fault detection or model error estimation, however, no prior knowledge about the unknown inputs is available. As shown in Fig. 4.1, the earliest approach to the estimation of completely unknown inputs in combined deterministic-stochastic systems is due to Glover [58], who considered the best way of estimating the unknown input as a linear combination of past outputs. His treatment, however, is limited to a special class of systems. Unknown inputs are also estimated in [73], however, no proof of optimality is given.

In contrast to the inversion of combined systems, which has received only little attention in the past, the problem of optimal state estimation, and in particular optimal filtering, for combined systems with unknown inputs has received a lot of attention. Numerous methods to deal with unknown inputs in optimal filtering can be found in literature. The earliest approach is due to Kitanidis [87], who considered an optimal filter for a system without direct feedthrough of the unknown input to the output. His approach consists in parameterizing the filter equations using a gain matrix and then calculating the optimal value of the gain matrix by minimizing the trace of the error covariance matrix under an unbiasedness condition. Stability and convergence conditions of the Kitanidis filter were developed in [24]. The approach of Kitanidis has been extended to systems with direct feedthrough of the input to the output in e.g. [25]. Another straightforward method is given by Hou et al. [68, 69]. Their approach consists in first transforming the system into another system that is decoupled from the unknown inputs, but whose state sequence equals that of the original system. Standard filtering techniques, like e.g. the Kalman filter can then be applied to estimate the state of the unknown input decoupled system. Other approaches to the filtering problem consist in transforming the estimation problem into a filtering problem for a descriptor system [102] or in the use of sliding mode observers [33].

It is important to notice that the filters referred to above, which are all based on stochastic assumptions about the noise, yield unbiased estimates under the conditions given in Fig. 3.7. This indicates that for combined systems unbiased estimates can be obtained under the same conditions that allow exact reconstruction in the deterministic case.

Only recently, research has also shifted towards time-delayed state estimation or smoothing [78, 124, 125]. As can be anticipated from Fig. 3.7, smoothing simplifies the unbiasedness conditions of the state estimators. As shown by Sundaram and Hadjicostis [125], a consequence of time-delayed estimation is that the noise processes become correlated with the estimation error, which complicates the state estimation problem.

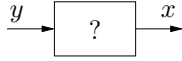
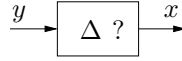
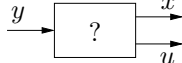
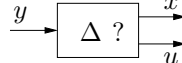
	Filtering	Smoothing
State estimation	Kitanidis, 1987 [87]  §4.2, §4.3 & §4.4	Sundaram et al., 2005 [124]  §4.4
Input estimation/ Joint input-state estimation	Glover, 1969 [58]  §4.2, §4.3 & §4.4	 §4.4

Figure 4.1: *Schematic of the history in optimal estimation in the presence of unknown inputs. Although the first contribution, due to Glover, addressed input estimation, the joint input-state estimation problem has received only little attention in the past. The optimal state estimation problem, in contrast, has received a lot of attention. The paragraphs in which the estimation problems are studied, are also listed.*

Personal contributions

Although the optimal state estimation problem for systems with unknown inputs has received a lot of attention in the past, the input estimation problem has received only little attention. The most important contribution of this chapter is the development of optimal input and joint-input state estimators.

- In Sect. 4.2, we derive a joint input-state filter for a system with direct feedthrough. The filter takes the form of the Kalman filter, except that the true value of the input is replaced by an optimal estimate. In addition, a relation to LS estimation is established.
- In Sect. 4.3, a similar analysis is given for the filtering problem of a system without direct feedthrough. We show that the optimal filter derived by Kitanidis [87] implicitly estimates the unknown input and derive a joint input-state filter based on this observation. A relation to LS estimation is established, and a square-root information algorithm is derived.
- In Sect. 4.4, we derive a general framework for optimal filtering, one step ahead prediction and smoothing in the context of both state estimation in the presence of unknown inputs and joint input-state estimation. A new and straightforward procedure for unknown input decoupled state estimation is derived.

Chapter outline

Section 4.2 starts by considering the most simple estimation problem, i.e. that for a system with direct feedthrough. In Sect. 4.3, a similar analysis is given for a system without direct feedthrough. Finally, in Sect. 4.4, the general framework is developed.

4.2 Optimal filtering with direct feedthrough

As discussed in the introduction, the optimal state estimation problem for a combined system with unknown input has first been considered by Kitanidis [87]. He addressed the filtering problem for a system without direct feedthrough of the unknown input to the output and derived a recursive filter, optimal in the MVU sense. In this section, we consider a similar derivation for the more simple problem of filtering with direct feedthrough. Important results of the section are the derivation of a method to estimate the unknown input and the establishment of a relation to LS estimation.

Consider the LTI discrete-time system

$$x_{[k+1]} = Ax_{[k]} + Bu_{[k]} + w_{[k]} \quad (4.1a)$$

$$y_{[k]} = Cx_{[k]} + Du_{[k]} + v_{[k]}, \quad (4.1b)$$

where $x_{[k]} \in \mathbb{R}^n$ denotes the state vector at time instant k , $u_{[k]} \in \mathbb{R}^m$ denotes an unknown deterministic input vector at time k , and $y_{[k]} \in \mathbb{R}^p$ denotes the vector of measurements at time k . The noise processes $\{w_{[k]} \in \mathbb{R}^n\}_{k=0}^{\infty}$ and $\{v_{[k]} \in \mathbb{R}^p\}_{k=0}^{\infty}$ are assumed to be stochastic with the properties given in Assumption 2.1. We define $Q := \mathbb{E}[w_{[k]}w_{[k]}^T]$ and $R := \mathbb{E}[v_{[k]}v_{[k]}^T]$ and assume that R is positive definite. In addition, we assume that the initial state $x_{[0]}$ is a random variable.

Throughout this section, it will be assumed that $\text{rank}(D) = m$, such that the deterministic system corresponding to (4.1) is instantaneously left invertible. We will see that under this condition also unbiased estimates of $u_{[k]}$ and $x_{[k]}$ can be obtained based on knowledge of the measurements up to time instant k .

This section is outlined as follows. In Sect. 4.2.1, we start by considering the state estimation problem for a system with direct feedthrough. Next, in Sect. 4.2.2, the optimal input estimation problem is addressed. In Sect. 4.2.3, the equations are split into a time update and a measurement update. The resulting equations are summarized in the form of a joint input-state filter in Sect. 4.2.4. Finally, in Sect. 4.2.5, the relation to LS estimation is established.

4.2.1 State estimation

We consider a recursive one step ahead predictor of the form

$$\hat{x}_{[k+1|k]} = A\hat{x}_{[k|k-1]} + \mathcal{K}_{[k]}(y_{[k]} - C\hat{x}_{[k|k-1]}), \quad (4.2)$$

where the gain matrix $\mathcal{K}_{[k]}$ is a design parameter. Notice that (4.2) takes the same form as that of the Kalman filter for a system without external inputs. We assume that an unbiased estimate $\hat{x}_{[0|-1]}$ of the initial state $x_{[0]}$ is available with covariance matrix $P_{[0|-1]}$. The error in the estimate $\hat{x}_{[0|-1]}$ is assumed to be uncorrelated to the noise processes $\{v_{[k]}\}_{k=0}^{\infty}$ and $\{w_{[k]}\}_{k=0}^{\infty}$.

Similar to the optimality condition considered in the derivation of the Kalman filter, we define the optimal value of $\mathcal{K}_{[k]}$ as the value that minimizes the mean squared error $\mathbb{E}[\|x_{[k+1]} - \hat{x}_{[k+1|k]}\|^2]$ over all linear unbiased estimates $\hat{x}_{[k+1|k]}$ of the form (4.2).

First, we determine the condition that $\mathcal{K}_{[k]}$ should satisfy in order that (4.2) is unbiased. It follows from (4.2) and (4.1) that the dynamical evolution of the estimation error $\tilde{x}_{[k+1|k]} := x_{[k+1]} - \hat{x}_{[k+1|k]}$ is governed by

$$\tilde{x}_{[k+1|k]} = (A - \mathcal{K}_{[k]}C)\tilde{x}_{[k|k-1]} + (B - \mathcal{K}_{[k]}D)u_{[k]} - \mathcal{K}_{[k]}v_{[k]} + w_{[k]}. \quad (4.3)$$

Consequently, (4.2) is unbiased for all $k \geq 0$ and all possible $u_{[k]}$ if and only if $\mathcal{K}_{[k]}$ satisfies the unbiasedness condition

$$\mathcal{K}_{[k]}D = B. \quad (4.4)$$

So, in contrast to the Kalman filter, the estimator (4.2) is not unbiased for all values of $\mathcal{K}_{[k]}$.

Now, we determine under all gain matrices $\mathcal{K}_{[k]}$ that satisfy the unbiasedness condition (4.4) the one that minimizes the mean squared error $\mathbb{E}[\|x_{[k+1]} - \hat{x}_{[k+1|k]}\|^2]$, which is the one that minimizes the trace of the error covariance matrix $P_{[k+1|k]} := \mathbb{E}[\tilde{x}_{[k+1|k]}\tilde{x}_{[k+1|k]}^T]$. The calculation of the optimal gain matrix thus requires that $P_{[k+1|k]}$ is expressed as function of $\mathcal{K}_{[k]}$. Let $\mathcal{K}_{[k]}$ satisfy (4.4), then it follows from (4.3) that $P_{[k+1|k]}$ obeys the recursion

$$\begin{aligned} P_{[k+1|k]} &= (A - \mathcal{K}_{[k]}C)P_{[k|k-1]}(A - \mathcal{K}_{[k]}C)^T + \mathcal{K}_{[k]}R\mathcal{K}_{[k]}^T + Q, \\ &= \mathcal{K}_{[k]}\tilde{R}_{[k]}\mathcal{K}_{[k]}^T - \mathcal{K}_{[k]}CP_{[k|k-1]}A^T - AP_{[k|k-1]}C^T\mathcal{K}_{[k]}^T \\ &\quad + AP_{[k|k-1]}A^T + Q, \end{aligned} \quad (4.5)$$

where

$$\tilde{R}_{[k]} := CP_{[k|k-1]}C^T + R.$$

The optimal gain matrix $\mathcal{K}_{[k]}$ is then given in the following theorem.

Theorem 4.1. *The gain matrix $\mathcal{K}_{[k]}$ given by*

$$\mathcal{K}_{[k]} = K_{[k]}(I - D\mathcal{M}_{[k]}) + B\mathcal{M}_{[k]}, \quad (4.6)$$

with $K_{[k]} := AP_{[k|k-1]}C^T\tilde{R}_{[k]}^{-1}$ and

$$\mathcal{M}_{[k]} := (D^T\tilde{R}_{[k]}^{-1}D)^{-1}D^T\tilde{R}_{[k]}^{-1}, \quad (4.7)$$

minimizes the trace of (4.5) under the unbiasedness condition (4.4).

Proof: Following the approach of Kitanidis, we solve the minimization problem using Lagrange multipliers. The Lagrangian is given by

$$\text{trace}\{\mathcal{K}_{[k]}\tilde{R}_{[k]}\mathcal{K}_{[k]}^T - 2\mathcal{K}_{[k]}CP_{[k|k-1]}A^T + AP_{[k|k-1]}A^T + Q\} - 2\text{trace}\{(\mathcal{K}_{[k]}D - B)\Lambda_{[k]}^T\}, \quad (4.8)$$

where $\Lambda_{[k]} \in \mathbb{R}^{n \times m}$ denotes the matrix of Lagrange multipliers and where the factor “2” before the second “trace{·}” is introduced for notational convenience. Setting the derivative of (4.8) with respect to $\mathcal{K}_{[k]}$ and $\Lambda_{[k]}$ equal to zero, yields

$$\tilde{R}_{[k]}\mathcal{K}_{[k]}^T - CP_{[k|k-1]}A^T - D\Lambda_{[k]}^T = 0, \quad (4.9)$$

and (4.4), respectively. Together, (4.9) and (4.4) form the linear system of equations

$$\begin{bmatrix} \tilde{R}_{[k]} & -D \\ D^T & 0 \end{bmatrix} \begin{bmatrix} \mathcal{K}_{[k]}^T \\ \Lambda_{[k]}^T \end{bmatrix} = \begin{bmatrix} CP_{[k|k-1]}A^T \\ B^T \end{bmatrix}, \quad (4.10)$$

which has a unique solution if and only if the coefficient matrix is nonsingular. The coefficient matrix is nonsingular since R is assumed to be positive definite and $\text{rank}(D) = m$. Finally, pre-multiplying left and right-hand side of (4.10) by the inverse of the coefficient matrix, yields (4.6). ■

The assumption that $\text{rank}(D) = m$, which we have made in the beginning of this section, has led to a unique solution of (4.10) and thus to a unique gain matrix $\mathcal{K}_{[k]}$ minimizing the trace of (4.5). The assumption that $\text{rank}(D) = m$, is thus sufficient for MVU state estimation. However, as we will see in Sect. 4.4, it is not a necessary condition. It will be shown in that section that the necessary and sufficient condition is (4.4),

$$\text{rank}(D) = \text{rank}\left(\begin{bmatrix} B \\ D \end{bmatrix}\right). \quad (4.11)$$

Notice that (4.4) is the necessary and sufficient condition for exact reconstruction in the deterministic case (see Fig 3.7).

This indicates, very importantly, that MVU estimates in the combined case can be obtained under the conditions that allow for exact reconstruction in the deterministic case.

One can easily derive an expression for the gain matrix without assuming that $\text{rank}(D) = m$. Proceeding as in the proof of Theorem 4.1, it is then found that the solution to (4.10) is not unique, so that the gain matrix $\mathcal{K}_{[k]}$ minimizing the trace of (4.5) is also not unique.

4.2.2 Input estimation

In the deterministic case, the condition $\text{rank}(D) = m$ is necessary and sufficient for instantaneous input reconstruction, see Fig. 3.7. In this section, we show that this condition is also sufficient to obtain an unbiased estimate of the unknown input in the combined case. The procedure is based on estimating the unknown input from the innovation using LS estimation.

Defining the *innovation* by $y_{[k]} - C\hat{x}_{[k|k-1]}$, it follows from (4.1) that

$$y_{[k]} - C\hat{x}_{[k|k-1]} = Du_{[k]} + e_{[k]}, \quad (4.12)$$

where $e_{[k]} := C\tilde{x}_{[k|k-1]} + v_{[k]}$. Notice that, in contrast to the Kalman filter, the innovation is not zero mean. Let $\hat{x}_{[k|k-1]}$ be unbiased, then it follows that $\mathbb{E}[y_{[k]} - C\hat{x}_{[k|k-1]}] = Du_{[k]}$. This indicates that an unbiased estimate of the unknown input $u_{[k]}$ can be obtained from the innovation using LS estimation.

Theorem 4.2. *Let $\hat{x}_{[k|k-1]}$ be unbiased, then*

$$\hat{u}_{[k|k]} = \mathcal{M}_{[k]}(y_{[k]} - C\hat{x}_{[k|k-1]}), \quad (4.13)$$

with $\mathcal{M}_{[k]}$ given by (4.7) is the MVU estimator of $u_{[k]}$ given the innovation $y_{[k]} - C\hat{x}_{[k|k-1]}$. The error covariance matrix $P_{u_{[k|k]}}$ of $\hat{u}_{[k|k]}$, defined by $P_{u_{[k|k]}} := \mathbb{E}[(u_{[k]} - \hat{u}_{[k|k]})(u_{[k]} - \hat{u}_{[k|k]})^\top]$, is given by

$$P_{u_{[k|k]}} = (D^\top \tilde{R}_{[k]}^{-1} D)^{-1}.$$

Proof: Theorem 4.2 immediately follows by applying the Gauss-Markov theorem (Theorem B.2) to (4.12) and noting that $\mathbb{E}[e_{[k]}e_{[k]}^\top] = \tilde{R}_{[k]}$. ■

It will now be shown that the state estimator (4.2) with $\mathcal{K}_{[k]}$ given by (4.6) implicitly estimates the unknown input. Substituting (4.6) in (4.2), yields

$$\begin{aligned} \hat{x}_{[k+1|k]} &= A\hat{x}_{[k|k-1]} + B\mathcal{M}_{[k]}(y_{[k]} - C\hat{x}_{[k|k-1]}) \\ &\quad + K_{[k]}(I - D\mathcal{M}_{[k]})(y_{[k]} - C\hat{x}_{[k|k-1]}), \\ &= A\hat{x}_{[k|k-1]} + B\hat{u}_{[k|k]} + K_{[k]}(y_{[k]} - C\hat{x}_{[k|k-1]} - D\hat{u}_{[k|k]}), \end{aligned} \quad (4.14)$$

where the last step follows from (4.13). Equation (4.14) indeed reveals the optimal input estimate $\hat{u}_{[k|k]}$.

Notice, very importantly, that (4.14) takes the form of the Kalman filter in one step ahead prediction form, except that the true value of the input is replaced by an optimal estimate. Equation (4.14) not only takes a form similar to that of the Kalman filter, but in addition, the expression for the gain matrix $K_{[k]}$ (given in Theorem 4.1) also equals that of the Kalman gain.

4.2.3 Time and measurement update

Like for the Kalman filter, it is possible to split the state estimator into a time update and a measurement update. The time update yields a one step ahead predicted estimate $\hat{x}_{[k+1|k]}$. The measurement update yields a filtered estimate $\hat{x}_{[k|k]}$.

4.2.3.1 Measurement update

We define the measurement update as

$$\hat{x}_{[k|k]} := \hat{x}_{[k|k-1]} + \mathcal{L}_{[k]}(y_{[k]} - C\hat{x}_{[k|k-1]}), \quad (4.15)$$

where the gain matrix $\mathcal{L}_{[k]}$ has to be determined so that $\hat{x}_{[k|k]}$ is unbiased and has minimal variance. It follows from (4.15) and (4.12) that

$$\tilde{x}_{[k|k]} = (I - \mathcal{L}_{[k]}C)\tilde{x}_{[k|k-1]} - \mathcal{L}_{[k]}Du_{[k]} - \mathcal{L}_{[k]}v_{[k]}. \quad (4.16)$$

Assuming that $\tilde{x}_{[k|k-1]}$ is unbiased, it follows that $\tilde{x}_{[k|k]}$ is unbiased for all $u_{[k]}$ if and only if $\mathcal{L}_{[k]}$ satisfies the unbiasedness condition $\mathcal{L}_{[k]}D = 0$. Under the unbiasedness condition, the error covariance matrix $P_{[k|k]}$, defined by $P_{[k|k]} := \mathbb{E}[\tilde{x}_{[k|k]}\tilde{x}_{[k|k]}^\top]$, is given by

$$P_{[k|k]} = (I - \mathcal{L}_{[k]}C)P_{[k|k-1]}(I - \mathcal{L}_{[k]}C)^\top + \mathcal{L}_{[k]}R_{[k]}\mathcal{L}_{[k]}^\top. \quad (4.17)$$

The optimal gain matrix $\mathcal{L}_{[k]}$ is then given in the following theorem.

Theorem 4.3. *The gain matrix $\mathcal{L}_{[k]}$ minimizing the trace of (4.17) under the unbiasedness condition $\mathcal{L}_{[k]}D = 0$ is given by*

$$\mathcal{L}_{[k]} = L_{[k]}(I - D\mathcal{M}_{[k]}), \quad (4.18)$$

where $\mathcal{M}_{[k]}$ is given by (4.7) and $L_{[k]} := P_{[k|k-1]}C^\top\tilde{R}_{[k]}^{-1}$.

Proof: The proof is very similar to that of Theorem 4.1 and can be found in [57]. ■

Notice that for $\mathcal{L}_{[k]}$ given by (4.18), (4.15) can be rewritten in a form that reveals the optimal estimate of the unknown input,

$$\hat{x}_{[k|k]} = \hat{x}_{[k|k-1]} + L_{[k]}(y_{[k]} - C\hat{x}_{[k|k-1]} - D\hat{u}_{[k|k]}). \quad (4.19)$$

Also, it can be shown that the expression for $P_{[k|k]}$ can be written as a function of $P_{u[k|k]}$ as

$$P_{[k|k]} = P_{[k|k-1]} - L_{[k]}(\tilde{R}_{[k]} - DP_{u[k|k]}D^\top)L_{[k]}^\top. \quad (4.20)$$

4.2.3.2 Time update

It follows from (4.19) and (4.14) that the time update is then given by

$$\hat{x}_{[k+1|k]} = A\hat{x}_{[k|k]} + B\hat{u}_{[k|k]}. \quad (4.21)$$

Consequently, the error covariance matrix $P_{[k+1|k]}$ of $\hat{x}_{[k+1|k]}$ can be written in terms of that of $\hat{x}_{[k|k]}$ as

$$P_{[k+1|k]} = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} P_{[k|k]} & P_{xu[k|k]} \\ P_{ux[k|k]} & P_{u[k|k]} \end{bmatrix} \begin{bmatrix} A^\top \\ B^\top \end{bmatrix} + Q,$$

where $P_{xu[k|k]} = P_{ux[k|k]}^\top := \mathbb{E}[\tilde{x}_{[k|k]}\tilde{u}_{[k|k]}^\top]$. An expression for $P_{xu[k|k]}$ is now derived. It follows from (4.13) and (4.12) that $\tilde{u}_{[k|k]} := u_{[k]} - \hat{u}_{[k|k]}$ is given by

$$\tilde{u}_{[k|k]} = (I - \mathcal{M}_{[k]}D)u_{[k]} - \mathcal{M}_{[k]}e_{[k]} = -\mathcal{M}_{[k]}e_{[k]}. \quad (4.22)$$

Using (4.16) and (4.22), it follows that

$$P_{xu[k|k]} = -L_{[k]}DP_{u[k|k]}.$$

4.2.4 Summary of filter equations

The filter equations derived above can be written in three steps: the estimation of the unknown input, the measurement update and the time update. These steps are given by:

Joint input-state estimation

- **Input estimation:**

$$\tilde{R}_{[k]} = CP_{[k|k-1]}C^\top + R \quad (4.23)$$

$$\mathcal{M}_{[k]} = (D^\top \tilde{R}_{[k]}^{-1} D)^{-1} D^\top \tilde{R}_{[k]}^{-1} \quad (4.24)$$

$$\hat{u}_{[k|k]} = \mathcal{M}_{[k]}(y_{[k]} - C\hat{x}_{[k|k-1]}) \quad (4.25)$$

$$P_{u[k|k]} = (D^\top \tilde{R}_{[k]}^{-1} D)^{-1} \quad (4.26)$$

- **Measurement update:**

$$L_{[k]} = P_{[k|k-1]}C^\top \tilde{R}_{[k]}^{-1} \quad (4.27)$$

$$\hat{x}_{[k|k]} = \hat{x}_{[k|k-1]} + L_{[k]}(y_{[k]} - C\hat{x}_{[k|k-1]} - D\hat{u}_{[k|k]}) \quad (4.28)$$

$$P_{[k|k]} = P_{[k|k-1]} - L_{[k]}(\tilde{R}_{[k]} - DP_{u[k|k]}D^\top)L_{[k]}^\top \quad (4.29)$$

$$P_{xu[k|k]} = P_{ux[k|k]}^\top = -L_{[k]}DP_{u[k|k]} \quad (4.30)$$

• **Time update:**

$$\hat{x}_{[k+1|k]} = A\hat{x}_{[k|k]} + B\hat{u}_{[k|k]} \quad (4.31)$$

$$P_{[k+1|k]} = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} P_{[k|k]} & P_{xu[k|k]} \\ P_{ux[k|k]} & P_u[k|k] \end{bmatrix} \begin{bmatrix} A^\top \\ B^\top \end{bmatrix} + Q \quad (4.32)$$

As already discussed, the time update and the measurement update take the form of the Kalman filter, except that the true value of the input is replaced by an optimal estimate. Also, notice that in case $D = 0$ and $B = 0$, the Kalman filter equations for a system without deterministic inputs are obtained.

A block diagram of the joint input-state estimator summarized above is given in Fig. 4.2.

4.2.5 Relation to least-squares estimation

In this section, we establish the relation between the filter derived above and LS estimation. In Sect. 4.2.5.1, we set-up a sequence of growing LS problems that yield smoothed, filtered and one step ahead predicted estimates of the system state and the unknown input. Next, in Sect. 4.2.5.2 an RLS procedure is derived that propagates a one step ahead predicted state estimate and the relation to the filter derived above is established.

4.2.5.1 Least-squares input and state estimation

We consider system (4.1), but contrary to the derivations in the previous sections, we do not make any stochastic assumption about the initial state $x_{[0]}$ and about the noise processes $\{v_{[k]}\}_{k=0}^\infty$ and $\{w_{[k]}\}_{k=0}^\infty$.

The derivation in this section, should be compared to that for the Kalman filter in Sect. 2.5.1. We start by setting-up a sequence of growing LS problems. The LS problem considered at time instant k estimates the state sequence $\{x_{[i]}\}_{i=0}^{k+1}$ and the unknown input sequence $\{u_{[i]}\}_{i=0}^k$ based on knowledge of $\{y_{[i]}\}_{i=0}^k$. To this aim, the equations of the system (4.1) from time instant 0 to time instant k are written into a form that expresses the data (i.e. the known vectors) as a linear combination of the unknowns (i.e. the state sequence and the unknown input sequence) plus noise terms. This yields the following system of equations,

$$\underbrace{\begin{bmatrix} \hat{x}_{[0|-1]} \\ y_{[0]} \\ 0 \\ \vdots \\ y_{[k]} \\ 0 \end{bmatrix}}_{\text{data}} = \begin{bmatrix} I & & & & \\ C & D & & & \\ & A & B & & -I \\ & & & \ddots & \\ & & & & C & D \\ & & & & A & B & -I \end{bmatrix} \underbrace{\begin{bmatrix} x_{[0]} \\ u_{[0]} \\ \vdots \\ x_{[k]} \\ u_{[k]} \\ x_{[k+1]} \end{bmatrix}}_{\text{unknowns}} + \underbrace{\begin{bmatrix} -\tilde{x}_{[0|-1]} \\ v_{[0]} \\ w_{[0]} \\ \vdots \\ v_{[k]} \\ w_{[k]} \end{bmatrix}}_{\text{noise}}. \quad (4.33)$$

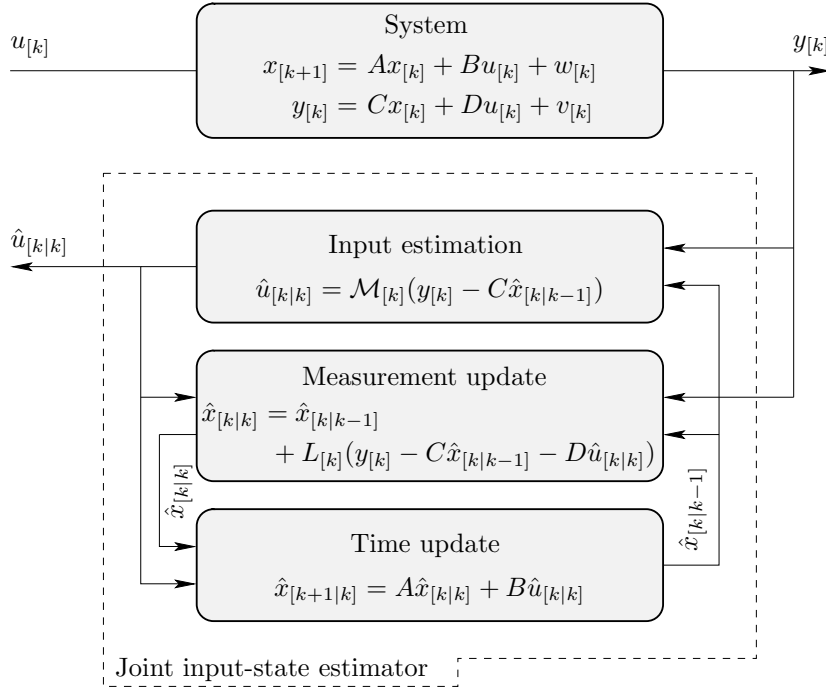


Figure 4.2: Block diagram of the joint input-state estimator summarized in Sect. 4.2.4. Notice that the time update and the measurement update take the form of that in the Kalman filter, except that the true value of the input is replaced by an optimal estimate.

The LS problem corresponding to (4.33), is given by

$$\min_{\substack{x[0], \dots, x[k+1] \\ u[0], \dots, u[k]}} \left\| \begin{bmatrix} \hat{x}[0|-1] \\ y[0] \\ 0 \\ \vdots \\ y[k] \\ 0 \end{bmatrix} - \begin{bmatrix} I & & & & & & \\ C & D & & & & & \\ A & B & -I & & & & \\ & & & \ddots & & & \\ & & & & C & D & \\ & & & & A & B & -I \end{bmatrix} \begin{bmatrix} x[0] \\ u[0] \\ \vdots \\ x[k] \\ u[k] \\ x[k+1] \end{bmatrix} \right\|_{W[k]}^2, \quad (4.34)$$

where $W[k]$ denotes the weighting matrix, which can be freely chosen. Notice that the LS problem (4.34) has $n + (k+1)(n+m)$ unknowns and is built up from $n + (k+1)(n+p)$ equations. Consequently, it is overdetermined for $p \geq m$. The problem has a unique solution if the coefficient matrix has full column rank.

As proven in Lemma C.2, a necessary and sufficient condition for this to hold is that $\text{rank}(D) = m$.

The arguments that minimize the LS problem (4.34) consist of smoothed estimates $\hat{x}_{[0|k]}, \dots, \hat{x}_{[k-1|k]}$ and $\hat{u}_{[0|k]}, \dots, \hat{u}_{[k-1|k]}$, filtered estimates $\hat{x}_{[k|k]}$ and $\hat{u}_{[k|k]}$, and a one step ahead predicted estimate $\hat{x}_{[k+1|k]}$.

Recall from Sect. 2.5.1 that the one step ahead predicted estimate and the filtered estimate obtained as solution of two consecutive LS problems of the form (2.19) obey the Kalman filter equations. Choosing $W_{[k]} = \text{diag}(P_{[0|-1]}^{-1}, R^{-1}, Q^{-1}, \dots, Q^{-1})$, where $P_{[0|-1]}$, Q and R denote matrices that can be freely chosen, the question can now be posed whether the LS estimates $\hat{x}_{[k|k]}$, $\hat{u}_{[k|k]}$ and $\hat{x}_{[k+1|k]}$ obtained as solution of two consecutive LS problems of the form (4.34) ($k = l, k = l + 1$) obey the recursive filter equations derived in the previous sections. Although no proof is given, there is strong belief that this indeed holds.

The LS problem (4.34) can be given the interpretation of an MVU estimator by choosing the weighting matrix $W_{[k]} = \text{diag}(P_{[0|-1]}^{-1}, R^{-1}, Q^{-1}, \dots, Q^{-1})$, where $P_{[0|-1]}$, Q and R denote the error covariance matrices defined above.

4.2.5.2 Recursive LS estimation

We now derive an RLS procedure that propagates a one step ahead predicted state estimate. For simplicity of derivations, we use a stochastic approach. We assume that an estimate $\hat{x}_{[k|k-1]}$ is available with covariance matrix $P_{[k|k-1]}$ and seek for an LS problem that allows to estimate $x_{[k+1]}$ based on $\hat{x}_{[k|k-1]}$ and the newly available measurement $y_{[k]}$. Considering the last two equations of (4.33) and appending an equation that summarizes the information in $\hat{x}_{[k|k-1]}$, yields

$$\begin{bmatrix} \hat{x}_{[k|k-1]} \\ y_{[k]} \\ 0 \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ C & D & 0 \\ A & B & -I \end{bmatrix} \begin{bmatrix} x_{[k]} \\ u_{[k]} \\ x_{[k+1]} \end{bmatrix} + \begin{bmatrix} -\tilde{x}_{[k|k-1]} \\ v_{[k]} \\ w_{[k]} \end{bmatrix}. \quad (4.35)$$

The corresponding LS problem is given by

$$\min_{x_{[k]}, u_{[k]}, x_{[k+1]}} \left\| \begin{bmatrix} \hat{x}_{[k|k-1]} \\ y_{[k]} \\ 0 \end{bmatrix} - \begin{bmatrix} I & 0 & 0 \\ C & D & 0 \\ A & B & -I \end{bmatrix} \begin{bmatrix} x_{[k]} \\ u_{[k]} \\ x_{[k+1]} \end{bmatrix} \right\|_{\bar{W}_{[k]}}^2, \quad (4.36)$$

where $\bar{W}_{[k]}$ denotes the weighting matrix. We give the LS problem (4.36) the interpretation of an MVU estimator by choosing $\bar{W}_{[k]} = \text{diag}(P_{[k|k-1]}^{-1}, R^{-1}, Q^{-1})$, where $P_{[0|-1]}$, Q and R denote the error covariance matrices defined above. Solution of the LS problem yields a one step ahead predicted estimate $\hat{x}_{[k+1|k]}$ and its error covariance matrix which are used to initialize the next step of the RLS procedure.

Like in the previous section, there is strong belief that the LS estimates $\hat{x}_{[k|k]}$, $\hat{u}_{[k|k]}$ and $\hat{x}_{[k+1|k]}$ obtained as solution of (4.36) obey the recursive filter equations summarized in Sect. 4.2.4.

In the next sections, LS problems for the time update and the measurement update are extracted from (4.36) and the relation to the recursive filter equations summarized in Sect. 4.2.4 is established. For simplicity of derivation, we use a stochastic approach.

Measurement update

Similar to the derivation of the Kalman filter considered in Sect. 2.5.2.1, the measurement update is derived from (4.36) by extracting the rows that depend only on $x_{[k]}$ and $u_{[k]}$. This yields,

$$\min_{x_{[k]}, u_{[k]}} \left\| \begin{bmatrix} \hat{x}_{[k|k-1]} \\ y_{[k]} \end{bmatrix} - \begin{bmatrix} I & 0 \\ C & D \end{bmatrix} \begin{bmatrix} x_{[k]} \\ u_{[k]} \end{bmatrix} \right\|_{\bar{W}_{1[k]}}^2, \quad (4.37)$$

where $\bar{W}_{1[k]}$ denotes the weighting which we give a stochastic interpretation by choosing $\bar{W}_{1[k]} = \text{diag}(P_{[k|k-1]}^{-1}, R^{-1})$. Using the Gauss-Markov theorem, it is now straightforward to prove the following proposition.

Proposition 4.1. *Solution of the LS problem (4.37) yields the equations for the measurement update and the estimation of the unknown input considered in Sect. 4.2.4.*

Proof: See Appendix C.2. ■

Time update

For the time update, we extract from (4.36) the equation that depends on $x_{[k+1]}$ and substitute $x_{[k]}$ and $u_{[k]}$ for their LS estimates $\hat{x}_{[k|k]}$ and $\hat{u}_{[k|k]}$ obtained during the measurement update. This yields,

$$A\hat{x}_{[k|k]} + B\hat{u}_{[k|k]} = x_{[k+1]} - (A\tilde{x}_{[k|k]} + B\tilde{u}_{[k]} + w_{[k]}).$$

The corresponding LS problem (with interpretation of an MVU estimator) is given by

$$\min_{x_{[k+1]}} \left\| x_{[k+1]} - A\hat{x}_{[k|k]} - B\hat{u}_{[k|k]} \right\|_{\bar{W}_{2[k]}}, \quad (4.38)$$

where $\bar{W}_{2[k]}$ denotes the weighting matrix which we choose as $\bar{W}_{2[k]} = (\mathbb{E}[(A\tilde{x}_{[k|k]} + B\tilde{u}_{[k]} + w_{[k]})(A\tilde{x}_{[k|k]} + B\tilde{u}_{[k]} + w_{[k]})^\top])^{-1}$.

Proposition 4.2. *Solution of the LS problem (4.38) yields the equations for the time update considered in Sect. 4.2.4.*

Proof: The proof is straightforward and hence omitted. ■

4.3 Optimal filtering without direct feedthrough

In this section, we consider an approach similar to that in the previous section, but now for a system without direct feedthrough of the unknown input to the output. The optimal state estimation problem for such a system has first been considered by Kitanidis [87]. The main contributions of the section are the extension to joint input-state estimation, the establishment of a relation to LS estimation and the derivation of a square-root information algorithm.

Consider the system

$$x_{[k+1]} = Ax_{[k]} + Bu_{[k]} + w_{[k]} \quad (4.39a)$$

$$y_{[k]} = Cx_{[k]} + v_{[k]}, \quad (4.39b)$$

where, as usual, $x_{[k]} \in \mathbb{R}^n$ denotes the state vector at time instant k , $u_{[k]} \in \mathbb{R}^m$ denotes an unknown deterministic input at time k , and $y_{[k]} \in \mathbb{R}^p$ denotes the measurement vector at time k . It is assumed that the initial state $x_{[0]}$ is a random variable. The noise processes $\{w_{[k]} \in \mathbb{R}^n\}_{k=0}^{\infty}$ and $\{v_{[k]} \in \mathbb{R}^p\}_{k=0}^{\infty}$ are assumed to be stochastic with the properties given in Assumption 2.1. We define $Q := \mathbb{E}[w_{[k]}w_{[k]}^T]$ and $R := \mathbb{E}[v_{[k]}v_{[k]}^T]$ and assume that R is positive definite.

Like Kitanidis, we assume throughout this section that $\text{rank}(CB) = m$, so that the deterministic system corresponding to (4.39) is 1–delay left invertible. We will see that under this condition also unbiased estimates of $u_{[k]}$ can be obtained using measurements up to time instant k .

Although the derivations considered in this section are closely related to those considered in the previous section, the optimal state estimation problem for (4.39) is conceptually very different from that of a system with direct feedthrough. The reason is that now the measurement $y_{[k]}$ does not contain any information about $u_{[k]}$. Consequently, no unbiased estimate of $x_{[k+1]}$ can be obtained using measurements up to time instant k . As a result, our discussion will now not start with the derivation of a recursive one step ahead predictor, but with the derivation of a recursive filter.

This section is outlined as follows. In Sect. 4.3.1, we start by considering the state estimation problem. Next, in Sect. 4.3.2, the equations are split into a time update and a measurement update. In Sect. 4.3.3, the optimal input estimation problem is addressed. The resulting equations are summarized in the form of a joint input-state filter in Sect. 4.3.4. In Sect. 4.3.5, the relation to LS estimation is established. Finally, in Sect. 4.3.6, a square-root information algorithm is derived.

4.3.1 State estimation

Following Kitanidis [87], we consider a recursive filter of the form

$$\hat{x}_{[k|k]} = A\hat{x}_{[k-1|k-1]} + \mathcal{L}_{[k]}(y_{[k]} - CA\hat{x}_{[k-1|k-1]}), \quad (4.40)$$

where the gain matrix $\mathcal{L}_{[k]}$ is a design parameter. We assume that an unbiased estimate $\hat{x}_{[0|0]}$ of the initial state $x_{[0]}$ is available with covariance matrix $P_{[0|0]}$. The error in $\hat{x}_{[0|0]}$ is assumed to be uncorrelated to $\{v_{[k]}\}_{k=0}^{\infty}$ and $\{w_{[k]}\}_{k=0}^{\infty}$.

Similar to Sect. 4.2, the optimal value of the gain matrix $\mathcal{L}_{[k]}$ is defined as that value that minimizes the mean squared error $\mathbb{E}[\|x_{[k]} - \hat{x}_{[k|k]}\|^2]$ over all linear unbiased estimates $\hat{x}_{[k|k]}$ of the form (4.40).

First, we determine the condition that $\mathcal{L}_{[k]}$ should satisfy in order that (4.40) is unbiased. It follows from (4.40) that the dynamical evolution of $\tilde{x}_{[k|k]} := x_{[k]} - \hat{x}_{[k|k]}$ is governed by

$$\tilde{x}_{[k|k]} = (I - \mathcal{L}_{[k]}C)(A\tilde{x}_{[k-1|k-1]} + Bu_{[k-1]} + w_{[k-1]}) - \mathcal{L}_{[k]}v_{[k]}. \quad (4.41)$$

Consequently, (4.40) is unbiased for all $k \geq 0$ and all possible $u_{[k-1]}$ if and only if $\mathcal{L}_{[k]}$ satisfies the unbiasedness condition

$$\mathcal{L}_{[k]}CB = B. \quad (4.42)$$

Under the assumption that (4.42) holds, the following recursion for the error covariance matrix $P_{[k|k]}$, defined by $P_{[k|k]} := \mathbb{E}[\tilde{x}_{[k|k]}\tilde{x}_{[k|k]}^\top]$, is easily derived from (4.41),

$$P_{[k|k]} = (I - \mathcal{L}_{[k]}C)(AP_{[k-1|k-1]}A^\top + Q)(I - \mathcal{L}_{[k]}C)^\top + \mathcal{L}_{[k]}R\mathcal{L}_{[k]}^\top. \quad (4.43)$$

Defining

$$\bar{P}_{[k|k-1]} := AP_{[k-1|k-1]}A^\top + Q, \quad (4.44)$$

and $\bar{R}_{[k]} := C\bar{P}_{[k|k-1]}C^\top + R$, (4.43) is rewritten as

$$P_{[k+1|k+1]} = \mathcal{L}_{[k]}\bar{R}_{[k]}\mathcal{L}_{[k]}^\top - \bar{P}_{[k|k-1]}C^\top\mathcal{L}_{[k]}^\top - \mathcal{L}_{[k]}C\bar{P}_{[k|k-1]} + \bar{P}_{[k|k-1]}. \quad (4.45)$$

The gain matrix $\mathcal{L}_{[k]}$ minimizing the trace of (4.45) under the unbiasedness condition (4.42) is then given in the following theorem.

Theorem 4.4. *The gain matrix $\mathcal{L}_{[k]}$ minimizing the trace of (4.43) under the unbiasedness condition (4.42) is given by*

$$\mathcal{L}_{[k]} = \bar{L}_{[k]} + (I - \bar{L}_{[k]}C)B\bar{M}_{[k]}, \quad (4.46)$$

where $\bar{L}_{[k]} := \bar{P}_{[k|k-1]}C^\top\bar{R}_{[k]}^{-1}$ and

$$\bar{M}_{[k]} := (F^\top\bar{R}_{[k]}^{-1}F)^{-1}F^\top\bar{R}_{[k]}^{-1}, \quad (4.47)$$

with $F := CB$.

Proof: The proof is similar to that of Theorem 4.1 and can be found in [87]. ■

The assumption that $\text{rank}(CB) = m$ has led to a unique gain matrix $\mathcal{L}_{[k]}$ minimizing the trace of (4.43). The condition that $\text{rank}(CB) = m$ is thus

sufficient for the existence of an optimal state estimator. However, it will be shown in Sect. 4.4, that it is not necessary. We will show in that section that a necessary and sufficient condition is

$$\text{rank}(CB) = \text{rank}(B), \quad (4.48)$$

which is the necessary and sufficient condition for the existence of a gain matrix $\mathcal{L}_{[k]}$ satisfying (4.42) and which also is the necessary and sufficient condition for exact state reconstruction in the deterministic case (see Fig. 3.7).

4.3.2 Time and measurement update

The equations of the optimal filter derived in the previous section are now split into a time update and a measurement update.

4.3.2.1 Time update

Assume that a filtered estimate $\hat{x}_{[k-1|k-1]}$ and its error covariance matrix $P_{[k-1|k-1]}$ are available. Notice that no unbiased estimate of $x_{[k]}$ can be obtained using measurements up to time instant $k-1$. On the other, an unbiased estimate of $\bar{x}_{[k]} := Ax_{[k-1]} + w_{[k-1]}$ can be obtained. Therefore, we define the time update as

$$\hat{x}_{[k|k-1]} := A\hat{x}_{[k-1|k-1]}. \quad (4.49)$$

It is easily verified that the error covariance matrix $\bar{P}_{[k|k-1]}$ of $\hat{x}_{[k|k-1]}$, defined by $\bar{P}_{[k|k-1]} := \mathbb{E}[\tilde{\tilde{x}}_{[k|k-1]}\tilde{\tilde{x}}_{[k|k-1]}^T]$, with $\tilde{\tilde{x}}_{[k|k-1]} := \bar{x}_{[k]} - \hat{x}_{[k|k-1]}$, is given by (4.44).

4.3.2.2 Measurement update

It follows from (4.40) and (4.49) that the measurement update is given by

$$\hat{x}_{[k|k]} = \hat{x}_{[k|k-1]} + \mathcal{L}_{[k]}(y_{[k]} - C\hat{x}_{[k|k-1]}), \quad (4.50)$$

with $\mathcal{L}_{[k]}$ given by (4.46). An expression for the error covariance matrix $P_{[k|k]}$ of $\hat{x}_{[k|k]}$ in terms of $\bar{P}_{[k|k-1]}$ has already been derived, see (4.43).

4.3.3 Input estimation

Like in the case of direct feedthrough, it can be shown that an unbiased estimate of the unknown input can be determined from the innovation using LS estimation. Defining the *innovation* by $y_{[k]} - C\hat{x}_{[k|k-1]}$, it follows from (4.49) that

$$y_{[k]} - C\hat{x}_{[k|k-1]} = Fu_{[k-1]} + \bar{e}_{[k]}, \quad (4.51)$$

where $\bar{e}_{[k]} := C\tilde{\tilde{x}}_{[k|k-1]} + v_{[k]}$. It is easily verified that the error covariance matrix of $\bar{e}_{[k]}$ is given by $\mathbb{E}[\bar{e}_{[k]}\bar{e}_{[k]}^T] = \bar{R}_{[k]}$. The LS estimate of $u_{[k-1]}$ is then given in the following theorem.

Theorem 4.5. Let $\hat{x}_{[k|k-1]}$ be an unbiased estimate of $\bar{x}_{[k]}$, then

$$\hat{u}_{[k-1|k]} = \bar{\mathcal{M}}_{[k]}(y_{[k]} - C\hat{x}_{[k|k-1]}), \quad (4.52)$$

with $\bar{\mathcal{M}}_{[k]}$ given by (4.47), is the MVU estimator of $u_{[k-1]}$ given the innovation $y_{[k]} - C\hat{x}_{[k|k-1]}$. The error covariance matrix $P_{u_{[k-1|k]}}$, defined by $P_{u_{[k-1|k]}} := \mathbb{E}[(u_{[k-1]} - \hat{u}_{[k-1|k]})(u_{[k-1]} - \hat{u}_{[k-1|k]})^\top]$, is given by

$$P_{u_{[k-1|k]}} = (F^\top \bar{R}_{[k]}^{-1} F)^{-1}.$$

Proof: The proof is similar to that of Theorem 4.2 and can be found in [49]. ■

It can now be shown that the state estimator derived by Kitanidis [87] implicitly estimates the unknown input.

Substituting (4.46) in (4.50) and using (4.52), yields

$$\hat{x}_{[k|k]} = \hat{x}_{[k|k-1]} + \bar{L}_{[k]}(y_{[k]} - C\hat{x}_{[k|k-1]}) + (I - \bar{L}_{[k]}C)B\hat{u}_{[k-1|k]}. \quad (4.53)$$

Equation (4.53) indeed reveals the optimal estimate $\hat{u}_{[k-1|k]}$ of the unknown input $u_{[k-1]}$. It follows from (4.53) that the measurement update can be split in two steps by defining the first step as

$$\hat{\hat{x}}_{[k|k]} := \hat{x}_{[k|k-1]} + B\hat{u}_{[k-1|k]},$$

so that the second step is given by

$$\hat{x}_{[k|k]} = \hat{\hat{x}}_{[k|k]} + \bar{L}_{[k]}(y_{[k]} - C\hat{\hat{x}}_{[k|k]}).$$

Both $\hat{\hat{x}}_{[k|k]}$ and $\hat{x}_{[k|k]}$ are unbiased estimates of $x_{[k]}$. The first step updates $\hat{\hat{x}}_{[k|k-1]}$ with the unbiased estimate $\hat{u}_{[k-1|k]}$ of $u_{[k-1]}$ so that $\hat{\hat{x}}_{[k|k]}$ is unbiased. The second step is similar to the measurement update in the Kalman filter. It minimizes the variance of the state estimate by assimilating the measurement $y_{[k]}$. Expressions for the covariance matrices of $\hat{\hat{x}}_{[k|k]}$ and $\hat{x}_{[k|k]}$ can be found in [49] and are also summarized in the following section.

4.3.4 Summary of filter equations

The filter equations derived above can be split into three steps: the time update, the estimation of the unknown input and the measurement update. These steps are given by:

Joint input-state estimation• **Time update:**

$$\hat{\hat{x}}_{[k|k-1]} = A\hat{\hat{x}}_{[k-1|k-1]} \quad (4.54)$$

$$\bar{\bar{P}}_{[k|k-1]} = A\bar{\bar{P}}_{[k-1|k-1]}A^T + Q \quad (4.55)$$

• **Estimation of unknown input:**

$$\bar{\bar{R}}_{[k]} = C\bar{\bar{P}}_{[k|k-1]}C^T + R$$

$$\bar{\bar{M}}_{[k]} = (F^T\bar{\bar{R}}_{[k]}^{-1}F)^{-1}F^T\bar{\bar{R}}_{[k]}^{-1}$$

$$\hat{\hat{u}}_{[k-1|k]} = \bar{\bar{M}}_{[k]}(y_{[k]} - C\hat{\hat{x}}_{[k|k-1]})$$

$$P_{u[k-1|k]} = (F^T\bar{\bar{R}}_{[k]}^{-1}F)^{-1}$$

• **Measurement update:**

$$\hat{\hat{x}}_{[k|k]} = \hat{\hat{x}}_{[k|k-1]} + B\hat{\hat{u}}_{[k-1|k]}$$

$$\bar{\bar{L}}_{[k]} = \bar{\bar{P}}_{[k|k-1]}C^T\bar{\bar{R}}_{[k]}^{-1}$$

$$\bar{\bar{P}}_{[k|k]} = \bar{\bar{P}}_{[k|k-1]} + B P_{u[k-1|k]} B^T - B P_{u[k-1|k]} F^T \bar{\bar{L}}_{[k]}^T - \bar{\bar{L}}_{[k]} F P_{u[k-1|k]} B^T$$

$$\hat{\hat{x}}_{[k|k]} = \hat{\hat{x}}_{[k|k]} + \bar{\bar{L}}_{[k]}(y_{[k]} - C\hat{\hat{x}}_{[k|k]})$$

$$P_{[k|k]} = \bar{\bar{P}}_{[k|k]} - \bar{\bar{L}}_{[k]}(\bar{\bar{R}}_{[k]} - F P_{u[k-1|k]} F^T)\bar{\bar{L}}_{[k]}^T$$

Notice that for $B = 0$, that is, if the system (4.39) is not subject to unknown inputs, the time update and the measurement update take the form of that of a Kalman filter for a system without external inputs.

A block diagram of the joint input-state estimator summarized above is given in Fig. 4.3.

4.3.5 Recursive least-squares estimation

In this section, we establish the relation between the filter derived above and LS estimation. In Sect. 4.3.5.1, we set-up a sequence of growing LS problems that yield smoothed, filtered and one step ahead predicted estimates of the system state and the unknown input. Next, in Sect. 4.3.5.2 an RLS procedure is derived that propagates a one step ahead predicted state estimate and the relation to the filter derived above is established.

4.3.5.1 Least-squares input and state estimation

We consider system (4.39), but contrary to the derivations in the previous sections, we do not make any stochastic assumption about the initial state $x_{[0]}$

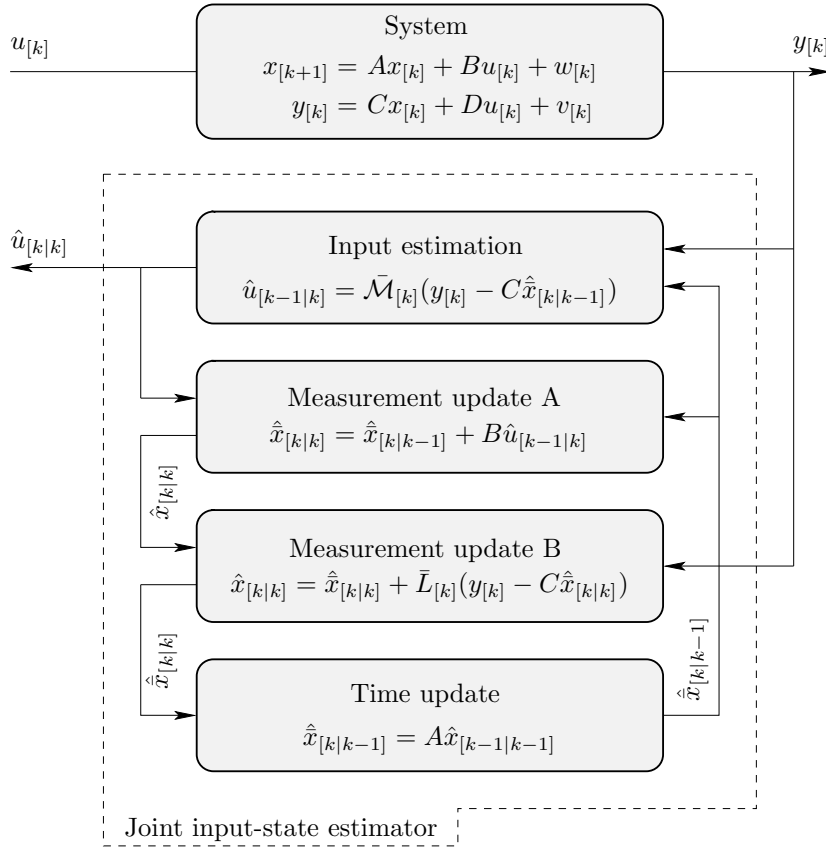


Figure 4.3: Block diagram of the joint input-state estimator summarized in Sect. 4.3.4. Notice that for $B = 0$, that is, if the system (4.39) is not subject to unknown inputs, the time update and the measurement update take the form of that of a Kalman filter for a system without external inputs.

and about the noise processes $\{v_{[k]}\}_{k=0}^{\infty}$ and $\{w_{[k]}\}_{k=0}^{\infty}$.

The derivation in this section should be compared to that for the Kalman filter in Sect. 2.5.1 and to that for the case of direct feedthrough in Sect. 4.2.5.1. The main difference to Sect. 4.2.5.1 is that the input now has to be estimated with one step delay.

Since $y_{[k]}$ does not contain any information about $u_{[k]}$, we consider at time instant k an LS problem that estimates the state sequence $\{x_{[0]}, x_{[1]}, \dots, x_{[k]}, \bar{x}_{[k+1]}\}$ and the unknown input sequence $\{u_{[i]}\}_{i=0}^{k-1}$ based on knowledge of the sequence $\{y_{[i]}\}_{i=0}^k$. To this aim, the equations of the system (4.39) from time instant 0 to time instant k are written into a form that expresses the data (i.e. the known vectors) as a linear combination of the unknowns (i.e. the state sequence and

the unknown input sequence) plus noise terms. This yields,

$$\underbrace{\begin{bmatrix} \hat{x}_{[0|-1]} \\ y_{[0]} \\ 0 \\ y_{[1]} \\ 0 \\ \vdots \\ y_{[k]} \\ 0 \end{bmatrix}}_{\text{data}} = \begin{bmatrix} I \\ C \\ A & B & -I \\ & C & \\ & & A & B & -I \\ & & & \ddots & \\ & & & & C \\ & & & & A & -I \end{bmatrix} \underbrace{\begin{bmatrix} x_{[0]} \\ u_{[0]} \\ x_{[1]} \\ u_{[1]} \\ \vdots \\ u_{[k-1]} \\ x_{[k]} \\ \bar{x}_{[k+1]} \end{bmatrix}}_{\text{unknowns}} + \underbrace{\begin{bmatrix} -\tilde{x}_{[0|-1]} \\ v_{[0]} \\ w_{[0]} \\ v_{[1]} \\ w_{[1]} \\ \vdots \\ v_{[k]} \\ w_{[k]} \end{bmatrix}}_{\text{noise}}. \quad (4.56)$$

The LS problem corresponding to (4.56), is given by

$$\min_{\substack{x_{[0]}, \dots, x_{[k]}, \bar{x}_{[k+1]} \\ u_{[0]}, \dots, u_{[k-1]}}} \left\| \begin{bmatrix} \hat{x}_{[0|-1]} \\ y_{[0]} \\ 0 \\ y_{[1]} \\ 0 \\ \vdots \\ y_{[k]} \\ 0 \end{bmatrix} - \begin{bmatrix} I \\ C \\ A & B & -I \\ & C & \\ & & A & B & -I \\ & & & \ddots & \\ & & & & C \\ & & & & A & -I \end{bmatrix} \begin{bmatrix} x_{[0]} \\ u_{[0]} \\ x_{[1]} \\ u_{[1]} \\ \vdots \\ u_{[k-1]} \\ x_{[k]} \\ \bar{x}_{[k+1]} \end{bmatrix} \right\|_{W_{[k]}}^2, \quad (4.57)$$

where $W_{[k]}$ denotes the weighting matrix, which can be freely chosen. Notice that the LS problem (4.57) has $2n + k(n + m)$ unknowns and is formed on the basis of $n + (k+1)(n+p)$ equations. Consequently, the problem is overdetermined for $p \geq m$. It has a unique solution if the coefficient matrix has full column rank. A necessary and sufficient condition for this to hold is that $\text{rank}(CB) = m$. The proof is similar to that of Lemma C.2, and is hence omitted.

The arguments that minimize the LS problem (4.57) consist of smoothed estimates $\hat{x}_{[0|k]}, \dots, \hat{x}_{[k-1|k]}$ and $\hat{u}_{[0|k]}, \dots, \hat{u}_{[k-1|k]}$, a filtered estimate $\hat{x}_{[k|k]}$ and the one step ahead predicted estimate $\hat{\hat{x}}_{[k+1|k]}$.

Choosing $W_{[k]} = \text{diag}(P_{[0|-1]}^{-1}, R^{-1}, Q^{-1}, \dots, Q^{-1})$, where $P_{[0|-1]}$, Q and R denote matrices that can be freely chosen, the question now poses whether the LS estimates $\hat{x}_{[k|k]}$, $\hat{u}_{[k|k-1]}$ and $\hat{\hat{x}}_{[k|k-1]}$ obtained as solution of two consecutive LS problems of the form (4.57) ($k = l, k = l + 1$) obey the recursive filter equations derived in the previous sections. A strong indication that this holds has been given in [86] and [49], where it is shown that the recursive filter summarized in Sect. 4.3.4 is globally optimal over all linear estimators (also those not constrained to be recursive) in a stochastic LS sense.

4.3.5.2 Recursive LS estimation

We now derive an RLS procedure that propagates a one step ahead predicted estimate of $\bar{x}_{[k]}$. For simplicity of derivation, we use a stochastic approach. We assume that an estimate $\hat{\bar{x}}_{[k|k-1]}$ is available with covariance matrix $\bar{P}_{[k|k-1]}$ and seek for an LS problem that yields an estimate of $\bar{x}_{[k+1]}$ based on $\hat{\bar{x}}_{[k|k-1]}$ and on the newly available measurement $y_{[k]}$. Considering the last two equations of (4.56) and appending an equation that summarizes the information in $\hat{\bar{x}}_{[k|k-1]}$, yields

$$\begin{bmatrix} \hat{\bar{x}}_{[k|k-1]} \\ y_{[k]} \\ 0 \end{bmatrix} = \begin{bmatrix} I & -B & 0 \\ C & 0 & 0 \\ A & 0 & -I \end{bmatrix} \begin{bmatrix} x_{[k]} \\ u_{[k-1]} \\ \bar{x}_{[k+1]} \end{bmatrix} + \begin{bmatrix} -\tilde{\bar{x}}_{[k|k-1]} \\ v_{[k]} \\ w_{[k]} \end{bmatrix}. \quad (4.58)$$

The corresponding LS problem is given by

$$\min_{x_{[k]}, u_{[k-1]}, \bar{x}_{[k+1]}} \left\| \begin{bmatrix} \hat{\bar{x}}_{[k|k-1]} \\ y_{[k]} \\ 0 \end{bmatrix} - \begin{bmatrix} I & -B & 0 \\ C & 0 & 0 \\ A & 0 & -I \end{bmatrix} \begin{bmatrix} x_{[k]} \\ u_{[k-1]} \\ \bar{x}_{[k+1]} \end{bmatrix} \right\|_{\bar{W}_{[k]}}^2, \quad (4.59)$$

where $\bar{W}_{[k]}$ denotes the weighting matrix. We give the LS problem (4.59) the interpretation of an MVU estimator by choosing $\bar{W}_{[k]} = \text{diag}(\bar{P}_{[k|k-1]}^{-1}, R^{-1}, Q^{-1})$, where $\bar{P}_{[0|-1]}$, Q and R denote the error covariance matrices as defined above. Solution of the LS problem (4.59) yields a one step ahead predicted estimate $\hat{\bar{x}}_{[k+1|k]}$ and its covariance matrix, which can be used to initialize the next step of the RLS procedure.

In the next sections, LS problems for the time update and the measurement update are extracted from (4.59) and the relation to the recursive filter equations summarized in Sect. 4.3.4 is established. For simplicity of derivation, we use a stochastic approach.

Measurement update

The measurement update is derived from (4.59) by extracting the rows that depend only on $x_{[k]}$ and $u_{[k-1]}$. This yields the LS problem,

$$\min_{x_{[k]}, u_{[k-1]}} \left\| \begin{bmatrix} \hat{\bar{x}}_{[k|k-1]} \\ y_{[k]} \end{bmatrix} - \begin{bmatrix} I & -B \\ C & 0 \end{bmatrix} \begin{bmatrix} x_{[k]} \\ u_{[k-1]} \end{bmatrix} \right\|_{\bar{W}_{1[k]}}^2, \quad (4.60)$$

which we give the interpretation of an MVU estimator by choosing $\bar{W}_{1[k]} = \text{diag}(\bar{P}_{[k|k-1]}^{-1}, R^{-1})$. Using the Gauss-Markov theorem, it is now straightforward to prove the following proposition.

Proposition 4.3. *Solution of the LS problem (4.60) yields the equations for the measurement update and the estimation of the unknown input considered in Sect. 4.3.4.*

Proof: The proof is similar to that of Proposition 4.1 and hence omitted. It can be found in [52, 55]. ■

The derivation in [52, 55] yields information formulas for the measurement update and the estimation of the unknown input. As will now be shown, there is a duality between the information formulas and the Kalman filter equations. First, we summarize the information formulas in [52, 55]:

- Estimation of unknown input:

$$P_{u[k-1|k]}^{-1} = F^T R^{-1} F - \check{L}_{u[k]} C^T R^{-1} F \quad (4.61)$$

$$P_{u[k-1|k]}^{-1} \hat{u}_{[k-1|k]} = F^T R^{-1} y_{[k]} - \check{L}_{u[k]} (C^T R^{-1} y_{[k]} + \bar{P}_{[k|k-1]}^{-1} \hat{x}_{[k|k-1]}), \quad (4.62)$$

$$\text{with } \check{L}_{u[k]} := F^T R^{-1} C (\bar{P}_{[k|k-1]}^{-1} + C^T R^{-1} C)^{-1}.$$

- Measurement update:

$$P_{[k|k]}^{-1} = \bar{P}_{[k|k-1]}^{-1} + C^T R^{-1} C - \check{L}_{[k]} B^T \bar{P}_{[k|k-1]} \quad (4.63)$$

$$P_{[k|k]}^{-1} \hat{x}_{[k|k]} = \bar{P}_{[k|k-1]}^{-1} \hat{x}_{[k|k-1]} + C^T R^{-1} y_{[k]} - \check{L}_{[k]} B^T \bar{P}_{[k|k-1]}^{-1} \hat{x}_{[k|k-1]}, \quad (4.64)$$

$$\text{with } \check{L}_{[k]} := \bar{P}_{[k|k-1]}^{-1} B (B^T \bar{P}_{[k|k-1]}^{-1} B)^{-1}.$$

The duality between (4.61), (4.63) and the recursion (2.12) for $P_{[k|k-1]}$ in the Kalman filter is given in Table 4.1. It will be used in Sect. 4.3.6 to derive a square-root information algorithm almost instantaneously.

Time update

For the time update, we extract from (4.59) the equation that depends on $\bar{x}_{[k+1]}$ and substitute $x_{[k]}$ for its LS estimates $\hat{x}_{[k|k]}$. This yields,

$$A \hat{x}_{[k|k]} = \bar{x}_{[k+1]} - (A \tilde{x}_{[k|k]} + w_{[k]}).$$

The corresponding LS problem with interpretation of MVU estimator is given by

$$\min_{\bar{x}_{[k+1]}} \left\| \bar{x}_{[k+1]} - A \hat{x}_{[k|k]} \right\|_{\bar{W}_{2[k]}}, \quad (4.65)$$

where $\bar{W}_{2[k]} := (\mathbb{E}[(A \tilde{x}_{[k|k]} + w_{[k]})(A \tilde{x}_{[k|k]} + w_{[k]})^T])^{-1}$. The following proposition follows immediately from the equivalence of (4.65) to the time update in the Kalman filter.

Kalman filter, Eq. (2.12)	Eq. (4.63)	Eq. (4.61)
$P_{[k k-1]}$	$\bar{P}_{[k k-1]}^{-1}$	R^{-1}
A	I	F^\top
R	0	$\bar{P}_{[k k-1]}^{-1}$
C	B^\top	C^\top
E	C^\top	0
Q	R^{-1}	0

Table 4.1: Duality between the recursion (2.12) for $P_{[k|k-1]}$ in the Kalman filter, the measurement update (4.63) and the estimation of the unknown input (4.61).

Proposition 4.4. *Solution of the LS problem (4.65) yields the equations for the time update considered in Sect. 4.3.4.*

4.3.6 Square-root information filtering

In this section, we use the duality relations established in Table 4.1 to derive a square-root information algorithm. Like the square-root implementations of the Kalman filter, the algorithm applies orthogonal transformations to triangularize a pre-array, which contains the prior estimates, forming a post-array which contains the updated estimates.

We assume throughout this section that A and Q are nonsingular. For a matrix X , $X^{1/2}$ denotes the lower triangular Cholesky factor of X .

4.3.6.1 Time update

Since the time update (4.54)-(4.55) takes the form of that in the Kalman filter, a square-root information algorithm can be implemented as in (2.30),

$$\begin{bmatrix} Q^{-\top/2} & A^{-\top} P_{[k|k]}^{-\top/2} \\ 0 & A^{-\top} P_{[k|k]}^{-\top/2} \\ \hline 0 & \hat{x}_{[k|k]}^\top P_{[k|k]}^{-\top/2} \end{bmatrix} \Theta_{1,k} = \begin{bmatrix} \star & 0 \\ \star & \bar{P}_{[k+1|k]}^{-\top/2} \\ \hline \star & x_{[k+1|k]}^\top \bar{P}_{[k+1|k]}^{-\top/2} \end{bmatrix},$$

where the “ \star ”-symbols denote measurements that are not important for our discussion and where $\Theta_{1,k}$ denotes an orthogonal transformation matrix that brings the pre-array into the lower triangular form of the post-array.

4.3.6.2 Measurement update

Based on Table 4.1 and on the square-root covariance algorithm (2.27) for the Kalman filter, we obtain the following update,

$$\left[\begin{array}{ccc} 0 & B^\top \bar{P}_{[k|k-1]}^{-\top/2} & 0 \\ 0 & \bar{P}_{[k|k-1]}^{-\top/2} & C^\top R^{-\top/2} \\ \hline 0 & \hat{x}_{[k|k-1]}^\top \bar{P}_{[k|k-1]}^{-\top/2} & y_{[k]}^\top R^{-\top/2} \end{array} \right] \Theta_{2,k} = \left[\begin{array}{ccc} \star & 0 & 0 \\ \star & P_{[k|k]}^{-\top/2} & 0 \\ \hline \star & \hat{x}_{[k|k]}^\top P_{[k|k]}^{-\top/2} & \star \end{array} \right], \quad (4.66)$$

where $\Theta_{2,k}$ denotes an orthogonal transformation matrix that brings the pre-array into the lower triangular form of the post-array. The algebraic equivalence of (4.66) to (4.63) and (4.64) can be verified by equating inner products of corresponding block rows of the post- and pre-array.

4.3.6.3 Input estimation

Using the duality given in Table 4.1, we obtain the following array algorithm for the estimation of the unknown input,

$$\left[\begin{array}{cc} \bar{P}_{[k|k-1]}^{-\top/2} & C^\top R^{-\top/2} \\ 0 & F^\top R^{-\top/2} \\ \hline \hat{x}_{[k|k-1]}^\top \bar{P}_{[k|k-1]}^{-\top/2} & y_{[k]}^\top R^{-\top/2} \end{array} \right] \Theta_{3,k} = \left[\begin{array}{ccc} \star & 0 & 0 \\ \star & P_{u[k-1|k]}^{-\top/2} & 0 \\ \hline \star & \hat{u}_{[k-1|k]}^\top P_{u[k-1|k]}^{-\top/2} & \star \end{array} \right], \quad (4.67)$$

where $\Theta_{3,k}$ denotes an orthogonal transformation matrix that brings the pre-array into the lower triangular form of the post-array. The algebraic equivalence of (4.67) to (4.61) and (4.62) can be verified by equating inner products of corresponding block rows of the post- and pre-array.

4.3.7 A note on square-root covariance filtering

A standard approach to convert between square-root covariance and square-root information implementations is to augment the post- and pre-array with extra rows and columns such that they become nonsingular and then invert both of them [98]. However, this procedure can not be carried out for the post- and pre-arrays in (4.66) due to the zero-matrix in the upper-left entry of the pre-array. This indicates that square-root covariance filtering in the presence of unknown inputs is not possible.

A second indication for this fact is now given. It follows from Table 4.1 that the dual of deriving a square-root covariance algorithm for the measurement update is deriving a square-root information algorithm for the Kalman filter equations of a system with perfect measurements. The latter problem is, however, unsolved.

4.4 A general framework

So far, we have considered only the filtering and one step ahead prediction problems. In this section, a general framework for one step ahead prediction, filtering and smoothing in the context of both state estimation and joint input-state estimation is established.

We consider the LTI discrete-time system

$$x_{[k+1]} = Ax_{[k]} + Bu_{[k]} + w_{[k]} \quad (4.68a)$$

$$y_{[k]} = Cx_{[k]} + Du_{[k]} + v_{[k]}, \quad (4.68b)$$

where, as usual, $x_{[k]} \in \mathbb{R}^n$ denotes the state vector at time instant k , $u_{[k]} \in \mathbb{R}^m$ denotes an unknown deterministic input vector at time k , and $y_{[k]} \in \mathbb{R}^p$ denotes the measurement vector at time k . The initial state $x_{[0]}$ is assumed to be a random variable. The noise processes $\{w_{[k]} \in \mathbb{R}^n\}_{k=0}^{\infty}$ and $\{v_{[k]} \in \mathbb{R}^p\}_{k=0}^{\infty}$ are assumed to be stochastic with the properties given in Assumption 2.1. We define $Q := \mathbb{E}[w_{[k]} w_{[k]}^T]$ and $R := \mathbb{E}[v_{[k]} v_{[k]}^T]$ and assume that R is positive definite.

The derivations in this section can be considered as extensions of the inversion procedure developed in Chapter 3. More precisely, we consider an estimator for (4.68) of the form (3.27) and show that unbiased estimates of the system state and the unknown input are obtained under the conditions derived in Fig. 3.7. Also, we show how to compute the matrix parameters Z_L and U_L so that the estimates of the system state and the unknown input are MVU.

This section is outlined as follows. In Sect. 4.4.1, we consider the problem of optimal state estimation in the presence of unknown inputs. Next, in Sect. 4.4.2, the input estimation problem is addressed. And finally, in Sect. 4.4.3, an estimator is developed in which the estimation of the system state and the unknown input are interconnected.

4.4.1 State estimation

We consider two approaches to optimal state estimation. In the first approach, we design a new and straightforward method for unknown input decoupled state estimation. In the second approach, a state estimator similar to that in Chapter 3 is considered and the gain matrix is determined so that the estimate of the system state has minimal variance.

4.4.1.1 Unknown input decoupled state estimation

The concept of unknown input decoupling yields a rigorous and straightforward approach to the design of optimal filters for systems with unknown inputs [26, 67–70]. The idea behind unknown input decoupling is to transform the state equation of the system into an equivalent state equation that is decoupled from the unknown input. By also deriving an output equation that is decoupled from the unknown input, standard filtering techniques, like e.g. the Kalman filter, can be employed to estimate the system state.

In this section, a new and very straightforward derivation of an unknown input decoupled system is given. In contrast to existing techniques, which are limited to the filtering and one step ahead prediction case, the derivation in this section yields a general framework for one step ahead prediction, filtering and smoothing.

Before deriving the unknown input decoupled system, some notations have to be introduced. We recursively define

$$\begin{aligned} \mathcal{N}_0 &:= 0, & \mathcal{N}_1 &:= \begin{bmatrix} 0 \\ C \end{bmatrix}, \\ \mathcal{N}_k &:= \begin{bmatrix} 0 & 0 \\ \mathcal{O}_{k-1} & \mathcal{N}_{k-1} \end{bmatrix}, & k &\geq 2. \end{aligned}$$

Furthermore, we define $y_{[k:k+L]} := [y_{[k]}^\top \ y_{[k+1]}^\top \ \cdots \ y_{[k+L]}^\top]^\top$, and use similar definitions for $u_{[k:k+L]}$, $v_{[k:k+L]}$ and $w_{[k:k+L-1]}$.

The response of (4.68) over $L + 1$ consecutive time units is then given by

$$y_{[k:k+L]} = \mathcal{O}_L x_{[k]} + \mathcal{H}_L u_{[k:k+L]} + \mathcal{N}_L w_{[k:k+L-1]} + v_{[k:k+L]}, \quad (4.69)$$

where \mathcal{O}_L denotes the extended observability matrix as defined in (2.3) and where \mathcal{H}_L , as defined in (3.6), contains the Markov parameters.

The derivation of the unknown input decoupled system is then based on the following lemma.

Lemma 4.1. *If condition (3.18) obtains, $Bu_{[k]}$ can be expressed as*

$$Bu_{[k]} = \mathcal{K}_L (y_{[k:k+L]} - \mathcal{O}_L x_{[k]}) - \mathcal{K}_L \mathcal{N}_L w_{[k:k+L-1]} - \mathcal{K}_L v_{[k:k+L]}, \quad (4.70)$$

where the general form of \mathcal{K}_L is given by (3.20).

Proof: Using (4.69), (4.70) is rewritten as $(\check{B} - \mathcal{K}_L \mathcal{H}_L)u_{[k:k+L]} = 0$. It follows from Lemma 3.2 that the equation $\check{B} = \mathcal{K}_L \mathcal{H}_L$ has a unique solution if and only if condition (3.18) obtains. ■

Substituting (4.70) in (4.68), yields the state equation of the unknown input decoupled system,

$$x_{[k+1]} = (A - \mathcal{K}_L \mathcal{O}_L)x_{[k]} + \mathcal{K}_L y_{[k:k+L]} + (\check{J}_n - \mathcal{K}_L \mathcal{N}_L)w_{[k:k+L-1]} - \mathcal{K}_L v_{[k:k+L]}.$$

This state equation is decoupled from the unknown input, but yet is equivalent to the state equation of (4.68) if condition (3.18) obtains.

An output equation that is decoupled from $u_{[k]}$ is obtained by pre-multiplying (4.69) by Σ_L , which yields

$$\Sigma_L y_{[k:k+L]} = \Sigma_L \mathcal{O}_L x_{[k]} + \Sigma_L \mathcal{N}_L w_{[k:k+L-1]} + \Sigma_L v_{[k:k+L]}. \quad (4.71)$$

Since Σ_L does not have full rank, the $p(L + 1)$ linear equations of (4.71) are linearly dependent. Therefore, Σ_L in (4.71) may also be replaced by $\bar{\Sigma}_L := \beta_L \Sigma_L$, where the $p(L + 1) \times p(L + 1) - \text{rank}(\mathcal{H}_L)$ matrix β_L is chosen so that $\bar{\Sigma}_L$ has full row rank.

Summarizing, the unknown input decoupled system is given by

$$x_{[k+1]} = (A - \mathcal{K}_L \mathcal{O}_L)x_{[k]} + \mathcal{K}_L y_{[k:k+L]} + \bar{w}_{[k:k+L-1]} \quad (4.72a)$$

$$\bar{\Sigma}_L y_{[k:k+L]} = \bar{\Sigma}_L \mathcal{O}_L x_{[k]} + \bar{v}_{[k:k+L]}, \quad (4.72b)$$

with $\bar{w}_{[k:k+L-1]} := (\check{I}_n - \mathcal{K}_L \mathcal{N}_L)w_{[k:k+L-1]} - \mathcal{K}_L v_{[k:k+L]}$ and $\bar{v}_{[k:k+L]} := \bar{\Sigma}_L \mathcal{N}_L w_{[k:k+L-1]} + \bar{\Sigma}_L v_{[k:k+L]}$. Notice that $\bar{w}_{[k:k+L-1]}$ is correlated to $\bar{v}_{[k:k+L]}$.

By treating $\bar{\Sigma}_L y_{[k:k+L]}$ in (4.72b) as output and $y_{[k:k+L]}$ in (4.72a) as input, standard estimation techniques like the Kalman filter can be employed to obtain optimal estimates of $x_{[k]}$. The optimal state estimation problem for the system (4.68) has thus been transformed into a standard Kalman filtering problem.

In order to place the poles of the state estimator, the pair $\{A - \mathcal{K}_L \mathcal{O}_L, \bar{\Sigma}_L \mathcal{O}_L\}$ must be observable. It is straightforward to prove that this pair is observable if and only if $\{A - \check{B} \mathcal{H}_L^{(1)} \mathcal{O}_L, \bar{\Sigma}_L \mathcal{O}_L\}$ is observable. A sufficient condition for the latter pair to be observable was given in Theorem 3.6.

4.4.1.2 MVU state estimation

Consider the state estimator of Theorem 3.3. In this section, it is first shown that this state estimator yields unbiased estimates of the state vector of (4.68). Next, the matrix parameter Z_L in the design of the state estimator is chosen so that the estimate of the system state has minimal variance.

Consider the state estimator of Theorem 3.3, but with time-varying gain matrix,

$$\hat{x}_{[k+1]} = A\hat{x}_{[k]} + \mathcal{K}_{L[k]}(y_{[k:k+L]} - \mathcal{O}_L \hat{x}_{[k]}). \quad (4.73)$$

Notice that in accordance to our notation we should actually write $\hat{x}_{[k:k+L-1]}$ instead of $\hat{x}_{[k]}$. However, for conciseness of equations, we will use $\hat{x}_{[k]}$ in the remainder of this chapter. It is assumed that an unbiased estimate $\hat{x}_{[0]}$ of $x_{[0]}$ is available with covariance matrix $P_{[0]}$. The error in $\hat{x}_{[0]}$ is assumed to be uncorrelated to $\{w_{[k]}\}$ and $\{v_{[k]}\}$.

The optimal value of the gain matrix $\mathcal{K}_{L[k]}$ is defined as that value that minimizes the mean squared error $\mathbb{E}[\|\hat{x}_{[k+1]} - x_{[k+1]}\|^2]$ over all linear unbiased estimates $\hat{x}_{[k+1]}$ of the form (4.73).

First, we derive a condition that $\mathcal{K}_{L[k]}$ should satisfy in order that (4.73) is unbiased. Using (4.68) and (4.73), we obtain the following expression for $\tilde{x}_{[k+1]} := x_{[k+1]} - \hat{x}_{[k+1]}$,

$$\tilde{x}_{[k+1]} = A\tilde{x}_{[k]} + Bu_{[k]} + w_{[k]} - \mathcal{K}_{L[k]}\tilde{y}_{[k:k+L]}, \quad (4.74)$$

where $\tilde{y}_{[k:k+L]} := y_{[k:k+L]} - \mathcal{O}_L \hat{x}_{[k]}$. It follows from (4.68) that

$$\tilde{y}_{[k:k+L]} = \mathcal{O}_L \tilde{x}_{[k]} + \mathcal{H}_L u_{[k:k+L]} + \mathcal{N}_L w_{[k:k+L-1]} + v_{[k:k+L]}. \quad (4.75)$$

Substituting (4.75) in (4.74), yields

$$\begin{aligned}\tilde{x}_{[k+1]} &= (A - \mathcal{K}_{L[k]}\mathcal{O}_L)\tilde{x}_{[k]} + (\check{B} - \mathcal{K}_{L[k]}\mathcal{H}_L)u_{[k:k+L]} \\ &\quad + (\check{I}_n - \mathcal{K}_{L[k]}\mathcal{N}_L)w_{[k:k+L-1]} - \mathcal{K}_{L[k]}v_{[k:k+L]}.\end{aligned}\quad (4.76)$$

We conclude from (4.76) that the estimator (4.73) is unbiased for all possible $u_{[k:k+L]}$ if and only if $\mathcal{K}_{L[k]}$ satisfies $\mathcal{K}_{L[k]}\mathcal{H}_L = \check{B}$. It follows from Lemma 3.2 that a solution to the latter equation exists if and only if condition (3.18) obtains, in which case the general solution is given by

$$\mathcal{K}_{L[k]} = \check{B}\mathcal{H}_L^{(1)} + Z_{L[k]}\Sigma_L, \quad (4.77)$$

where $Z_{L[k]}$ is an arbitrary matrix. This yields the following lemma.

Lemma 4.2. *For $\mathcal{K}_{L[k]}$ given by (4.77), the estimator (4.73) is unbiased if and only if condition (3.18) obtains.*

Now, we determine the value of $Z_{L[k]}$ that minimizes the mean squared error $\mathbb{E}[\|\hat{x}_{[k+1]} - x_{[k+1]}\|^2]$, or equivalently the trace of the error covariance matrix $P_{[k+1]} := \mathbb{E}[\tilde{x}_{[k+1]}\tilde{x}_{[k+1]}^\top]$. It follows from (4.76) that $P_{[k+1]}$ obeys the following recursion,

$$P_{[k+1]} = \mathcal{K}_{L[k]}\bar{R}_{[k]}\mathcal{K}_{L[k]}^\top - \mathcal{K}_{L[k]}\bar{S}_{[k]}^\top - \bar{S}_{[k]}\mathcal{K}_{L[k]}^\top + \bar{T}_{[k]}, \quad (4.78)$$

where

$$\begin{aligned}\bar{R}_{[k]} &:= \mathbb{E}[(\tilde{y}_{[k:k+L]} - \mathcal{H}_L u_{[k:k+L]})(\tilde{y}_{[k:k+L]} - \mathcal{H}_L u_{[k:k+L]})^\top], \\ &= \mathcal{O}_L P_{[k]}\mathcal{O}_L^\top + R_{[k:k+L]} + \mathcal{N}_L Q_{[k:k+L-1]}\mathcal{N}_L^\top \\ &\quad + \mathcal{O}_L P_{xw[k]}\mathcal{N}_L^\top + \mathcal{N}_L P_{wx[k]}\mathcal{O}_L^\top + \mathcal{O}_L P_{xv[k]} + P_{vx[k]}\mathcal{O}_L^\top, \\ \bar{S}_{[k]} &:= AP_{[k]}\mathcal{O}_L^\top + \check{I}_n Q_{[k:k+L-1]}\mathcal{N}_L^\top + \check{I}_n P_{wx[k]}\mathcal{O}_L^\top + AP_{xw[k]}\mathcal{N}_L^\top + AP_{xv[k]}, \\ \bar{T}_{[k]} &:= AP_{[k]}A^\top + Q + AP_{xw[k]}\check{I}_n^\top + \check{I}_n P_{wx[k]}A^\top,\end{aligned}$$

with

$$\begin{aligned}R_{[k:k+L]} &:= \mathbb{E}[v_{[k:k+L]}v_{[k:k+L]}^\top], \\ Q_{[k:k+L-1]} &:= \mathbb{E}[w_{[k:k+L-1]}w_{[k:k+L-1]}^\top], \\ P_{xw[k]} &= (P_{wx[k]})^\top := \mathbb{E}[\tilde{x}_{[k]}w_{[k:k+L-1]}^\top], \\ P_{xv[k]} &= (P_{vx[k]})^\top := \mathbb{E}[\tilde{x}_{[k]}v_{[k:k+L]}^\top].\end{aligned}$$

Closed form expressions for $R_{[k:k+L]}$ and $Q_{[k:k+L-1]}$ are obtained from (4.68), $R_{[k:k+L]} = \text{diag}_0^{L+1}(R, R, \dots, R)$ and $Q_{[k:k+L-1]} = \text{diag}_0^L(Q, Q, \dots, Q)$, where $\text{diag}_j^i(\cdot)$ denotes a matrix with the i entries between the brackets on the j -th diagonal above the main diagonal and zeros elsewhere. Using (4.76), we obtain the following closed form expressions for $P_{xw[k]}$ and $P_{xv[k]}$,

$$P_{xw[k]} = (\check{I}_n - \mathcal{K}_{L[k-1]}\mathcal{N}_L)\mathbb{E}[w_{[k-1:k+L-2]}w_{[k:k+L-1]}^\top]$$

$$\begin{aligned}
& + (A - \mathcal{K}_{L[k-1]}\mathcal{O}_L)\mathbb{E}[\tilde{x}_{[k-1]}w_{[k:k+L-1]}^\top], \\
& = \sum_{i=1}^{\min\{k, L-1\}} \left(\prod_{j=1}^{i-1} (A - \mathcal{K}_{L[k-j]}\mathcal{O}_L) \right) (\check{I}_n - \mathcal{K}_{L[k-i]}\mathcal{N}_L) \text{diag}_{-i}^L(Q, \dots, Q),
\end{aligned}$$

and

$$\begin{aligned}
P_{xv[k]} & = -\mathcal{K}_{L[k-1]}\mathbb{E}[v_{[k-1:k+L-1]}v_{[k:k+L]}^\top] - (A - \mathcal{K}_{L[k-1]}\mathcal{O}_L)\mathbb{E}[\tilde{x}_{[k-1]}v_{[k:k+L]}^\top], \\
& = -\sum_{i=1}^{\min\{k, L\}} \left(\prod_{j=1}^{i-1} (A - \mathcal{K}_{L[k-j]}\mathcal{O}_L) \right) \mathcal{K}_{L[k-i]} \text{diag}_{-i}^{L+1}(R, \dots, R),
\end{aligned}$$

where $\sum_i^j(\cdot) := 0$ for $i > j$ and $\prod_i^j(\cdot) := I$ for $i > j$.

Substituting (4.77) in (4.78), yields

$$\begin{aligned}
P_{[k+1]} & = Z_{L[k]}\Sigma_L\bar{R}_{[k]}\Sigma_L^\top Z_{L[k]}^\top + \bar{T}_{[k]} + \check{B}\mathcal{H}_L^{(1)}\bar{R}_{[k]}(\check{B}\mathcal{H}_L^{(1)})^\top \\
& \quad - Z_{L[k]}\Sigma_L\check{S}_{[k]}^\top - \check{S}_{[k]}\Sigma_L^\top Z_{L[k]}^\top - \check{B}\mathcal{H}_L^{(1)}\bar{S}_{[k]}^\top - \bar{S}_{[k]}(\check{B}\mathcal{H}_L^{(1)})^\top, \quad (4.79)
\end{aligned}$$

where $\check{S}_{[k]} := \bar{S}_{[k]} - \check{B}\mathcal{H}_L^{(1)}\bar{R}_{[k]}$. Uniqueness of the gain matrix $Z_{L[k]}$ minimizing the trace of (4.79) requires invertibility of $\Sigma_L\bar{R}_{[k]}\Sigma_L^\top$. However, it follows from Lemma 3.3 that Σ_L does not have full rank if $\mathcal{H}_L \neq 0$.

Based on the rank of $\Sigma_L\bar{R}_{[k]}\Sigma_L^\top$, we consider three cases.

- Case 1: $\text{rank}(\Sigma_L\bar{R}_{[k]}\Sigma_L^\top) = 0$

Notice that this occurs e.g. when \mathcal{H}_L has full row rank. The matrix $Z_{L[k]}$ minimizing the trace of (4.79) is then given by $Z_{L[k]} = 0$, so that $\mathcal{K}_{L[k]}$ becomes time-invariant,

$$\mathcal{K}_{L[k]} = \check{B}\mathcal{H}_L^{-1}. \quad (4.80)$$

Substituting (4.80) in (4.79), yields

$$P_{[k+1]} = \bar{T}_{[k]} + \check{S}_{[k]}\bar{R}_{[k]}^{-1}\check{S}_{[k]}^\top - \bar{S}_{[k]}\bar{R}_{[k]}^{-1}\bar{S}_{[k]}^\top.$$

This corresponds to the finding in Chapter 3 that the state estimator of Theorem 3.3 is unique if \mathcal{H}_L has full row rank.

- Case 2: $0 < \text{rank}(\Sigma_L\bar{R}_{[k]}\Sigma_L^\top) < p(L+1)$

The optimal gain matrix is not unique. One of the matrices $Z_{L[k]}$ minimizing the trace of (4.79) is obtained by the use of the Moore Penrose generalized inverse,

$$Z_{L[k]} = \check{S}_{[k]}\Sigma_L^\top(\Sigma_L\bar{R}_{[k]}\Sigma_L^\top)^\dagger. \quad (4.81)$$

Substituting (4.81) in (4.77) and (4.79), yields

$$\mathcal{K}_{L[k]} = \check{B}\mathcal{H}_L^{(1)} + \check{S}_{[k]}\Sigma_L^\top(\Sigma_L\bar{R}_{[k]}\Sigma_L^\top)^\dagger\Sigma_L, \quad (4.82)$$

and

$$P_{[k+1]} = \bar{T}_{[k]} - \bar{S}_{[k]}\bar{R}_{[k]}^{-1}\bar{S}_{[k]}^\top + \check{S}_{[k]}(\bar{R}_{[k]}^{-1} - \Sigma_L^\top(\Sigma_L\bar{R}_{[k]}\Sigma_L^\top)^\dagger\Sigma_L)\check{S}_{[k]}^\top. \quad (4.83)$$

- Case 3: $\text{rank}(\Sigma_L \bar{R}_{[k]} \Sigma_L^\top) = p(L+1)$

This can occur only if (4.68) is unaffected by unknown inputs, that is, if $B = 0$ and $D = 0$. Since $\Sigma_L \bar{R}_{[k]} \Sigma_L^\top$ is invertible, the optimal gain matrix is unique and the generalized inverse of $\Sigma_L \bar{R}_{[k]} \Sigma_L^\top$ in (4.81) can be replaced by an inverse. Furthermore, it is straightforward to verify that we obtain the Kalman filter equations in filter form for $L = 0$ and in prediction form for $L = 1$. For $L > 1$, we obtain fixed-lag smoothing formulas.

4.4.2 Input estimation

In this section, we address the problem of optimal input estimation. We consider the input estimator of Theorem 3.4 and determine the gain matrix U_L so that the estimate of the input is unbiased and has minimal variance.

Consider the input estimator of Theorem 3.4, but now with time varying gain matrix,

$$\hat{u}_{[k]} = \mathcal{M}_{L[k]}(y_{[k:k+L]} - \mathcal{O}_L \hat{x}_{[k]}), \quad (4.84)$$

where $\hat{x}_{[k]}$ denotes the optimal estimate of the system state obtained with (4.73).

First, we determine the condition that $\mathcal{M}_{L[k]}$ should satisfy in order that $\hat{u}_{[k]}$ is unbiased. It follows from (4.84) that $\tilde{u}_{[k]} := u_{[k]} - \hat{u}_{[k]}$, is given by

$$\tilde{u}_{[k]} = (\check{I}_m - \mathcal{M}_{L[k]} \mathcal{H}_L) u_{[k:k+L]} - \mathcal{M}_{L[k]} (\mathcal{O}_L \tilde{x}_{[k]} + \mathcal{N}_L w_{[k:k+L-1]} + v_{[k:k+L]}). \quad (4.85)$$

The estimator (4.84) is thus unbiased for all possible $u_{[k:k+L]}$ if and only if $\mathcal{M}_{L[k]}$ satisfies $\mathcal{M}_{L[k]} \mathcal{H}_L = \check{I}_m$. We know from Lemma 3.1 that a solution to the latter equation exists if and only if condition (3.7) obtains, that is, if and only if the deterministic system corresponding to (4.68) is L -delay left invertible. The general solution is then given by

$$\mathcal{M}_{L[k]} = \check{I}_m \mathcal{H}_L^{(1)} + U_{L[k]} \Sigma_L, \quad (4.86)$$

where $U_{L[k]}$ is an arbitrary matrix. This yields the following lemma.

Lemma 4.3. *For $\mathcal{M}_{L[k]}$ given by (4.86), the estimator (4.84) is unbiased if and only if condition (3.7) obtains.*

Now, we calculate the value of $U_{L[k]}$ that minimizes the trace of the error covariance matrix $P_{u[k]} := \mathbb{E}[\tilde{u}_{[k]} \tilde{u}_{[k]}^\top]$. It follows from (4.85) that $P_{u[k]}$ is given by

$$P_{u[k]} = \mathcal{M}_{L[k]} \bar{R}_{[k]} \mathcal{M}_{L[k]}^\top. \quad (4.87)$$

Substituting (4.86) in (4.87), yields

$$P_{u[k]} = U_{L[k]} \Sigma_L \bar{R}_{[k]} \Sigma_L^\top U_{L[k]}^\top + U_{L[k]} \Sigma_L \bar{R}_{[k]} (\check{I}_m \mathcal{H}_L^{(1)})^\top + \check{I}_m \mathcal{H}_L^{(1)} \bar{R}_{[k]} \Sigma_L^\top U_{L[k]}^\top + \check{I}_m \mathcal{H}_L^{(1)} \bar{R}_{[k]} (\check{I}_m \mathcal{H}_L^{(1)})^\top. \quad (4.88)$$

Based on the rank of $\Sigma_L \bar{R}_{[k]} \Sigma_L^\top$, we distinguish two cases. Notice that $\text{rank}(\Sigma_L \bar{R}_{[k]} \Sigma_L^\top) = p(L+1)$ is not possible for a system for which (3.7) obtains. Therefore, this case is not considered.

- Case 1: $\text{rank}(\Sigma_L \bar{R}_{[k]} \Sigma_L^\top) = 0$
The matrix $U_{L[k]}$ minimizing the trace of (4.88) is given by $U_{L[k]} = 0$, so that $\mathcal{M}_{L[k]}$ becomes time-invariant,

$$\mathcal{M}_{L[k]} = \check{I}_m \mathcal{H}_L^{-1}. \quad (4.89)$$

Substituting (4.89) in (4.88), yields

$$P_{u[k]} = \check{I}_m \mathcal{H}_L^{-1} \bar{R}_{[k]} (\check{I}_m \mathcal{H}_L^{-1})^\top.$$

- Case 2: $0 < \text{rank}(\Sigma_L \bar{R}_{[k]} \Sigma_L^\top) < p(L+1)$
The matrix $U_{L[k]}$ minimizing the trace of (4.88) is not unique. One of the optimal gain matrices is given by

$$U_{L[k]} = -\check{I}_m \mathcal{H}_L^{(1)} \bar{R}_{[k]} \Sigma_L^\top (\Sigma_L \bar{R}_{[k]} \Sigma_L^\top)^\dagger. \quad (4.90)$$

Substituting (4.90) in (4.86) and (4.88) yields

$$\mathcal{M}_{L[k]} = \check{I}_m \mathcal{H}_L^{(1)} \bar{R}_{[k]}^{-1} \tilde{R}_{[k]}, \quad (4.91)$$

and

$$P_{u[k]} = \check{I}_m \mathcal{H}_L^{(1)} \tilde{R}_{[k]} (\check{I}_m \mathcal{H}_L^{(1)})^\top, \quad (4.92)$$

respectively.

4.4.3 Joint input-state estimation

Combining the state estimator (4.73) and the input estimator (4.84) derived in the previous section, yields a joint input-state estimator. Notice that the resulting estimator is time-varying, and can at every time instant be considered as an L -delay left inverse of the deterministic system corresponding to (4.68). Furthermore, if the covariance matrices $P_{[k]}$ and $P_{u[k]}$ converge for $k \rightarrow \infty$, also $Z_{L[k]}$, $\mathcal{K}_{L[k]}$, $U_{L[k]}$ and $\mathcal{M}_{L[k]}$ will converge. The joint input-state estimator then converges to the dynamical portion (3.26) with $Z_L = Z_{L[\infty]}$ and $U_L = U_{L[\infty]}$.

The state estimator (4.73) and input estimator (4.84) exchange information in one direction: the input estimator uses information from the state estimator. In this section, we show that if condition (3.7) obtains, the state estimator implicitly estimates the unknown input. This will yield a joint input-state estimator in which the state estimator and input estimator exchange information in both directions.

Based on the rank of $\Sigma_L \bar{R}_{[k]} \Sigma_L^\top$, we distinguish two cases.

- Case 1: $\text{rank}(\Sigma_L \bar{R}_{[k]} \Sigma_L^\top) = 0$

Using the fact that $\check{B} = B\check{I}_m$, it follows from (4.89) that (4.80) can be written as $\mathcal{K}_{L[k]} = B\mathcal{M}_{L[k]}$. The state estimator (4.73) can thus be written as $\hat{x}_{[k+1]} = A\hat{x}_{[k]} + B\mathcal{M}_{L[k]}\tilde{y}_{[k:k+L]} = A\hat{x}_{[k]} + B\hat{u}_{[k]}$, where the last equality follows from the assumption that condition (3.7) obtains. We conclude that the state estimator implicitly estimates the unknown input.

- Case 2: $0 < \text{rank}(\Sigma_L \bar{R}_{[k]} \Sigma_L^\top) < p(L+1)$

It follows from (4.91) that (4.82) can be rewritten as

$$\begin{aligned} \mathcal{K}_{L[k]} &= \check{B}\mathcal{H}_L^{(1)}\bar{R}_{[k]}^{-1}\tilde{R}_{[k]} + \bar{S}_{[k]}\Sigma_L^\top(\Sigma_L\bar{R}_{[k]}\Sigma_L^\top)^\dagger\Sigma_L \\ &= B\mathcal{M}_{L[k]} + \bar{S}_{[k]}\Sigma_L^\top(\Sigma_L\bar{R}_{[k]}\Sigma_L^\top)^\dagger\Sigma_L. \end{aligned}$$

The state estimator (4.73) can thus be written as

$$\begin{aligned} \hat{x}_{[k+1]} &= A\hat{x}_{[k]} + [B\mathcal{M}_{L[k]} + \bar{S}_{[k]}\Sigma_L^\top(\Sigma_L\bar{R}_{[k]}\Sigma_L^\top)^\dagger\Sigma_L]\tilde{y}_{[k:k+L]} \\ &= A\hat{x}_{[k]} + B\hat{u}_{[k]} + \bar{S}_{[k]}\Sigma_L^\top(\Sigma_L\bar{R}_{[k]}\Sigma_L^\top)^\dagger\Sigma_L\tilde{y}_{[k:k+L]}, \end{aligned} \quad (4.93)$$

where the last equality follows from the assumption that condition (3.7) obtains. We conclude that the state estimator implicitly estimates the unknown input. It follows from (4.83) and (4.92) that $P_{[k+1]}$ can be written in function of $P_{u[k]}$ as

$$\begin{aligned} P_{[k+1]} &= \bar{T}_{[k]} + BP_{u[k]}B^\top - \bar{S}_{[k]}\bar{R}_{[k]}^{-1}\bar{S}_{[k]}^\top - \check{B}\mathcal{H}_L^{(1)}\tilde{R}_{[k]}(\check{B}\mathcal{H}_L^{(1)})^\top \\ &\quad + (\bar{S}_{[k]}\bar{R}_{[k]}^{-1} - \check{B}\mathcal{H}_L^{(1)})\tilde{R}_{[k]}(\bar{S}_{[k]}\bar{R}_{[k]}^{-1} - \check{B}\mathcal{H}_L^{(1)})^\top. \end{aligned}$$

Combining (4.93) with (4.84) yields a joint input-state estimator in which the estimators exchange information in both directions.

4.5 Numerical examples

We consider two numerical examples. The first example deals with optimal filtering, the second one with optimal smoothing.

Example 4.1. Fault reconstruction in an F16 aircraft

This example addresses reconstruction of actuator and sensor faults in an F16 aircraft. We assume that the class and the location of the fault is known, and address the problem of fault estimation. Consider the following linearized F16 longitudinal model [120],

$$\begin{aligned} \frac{dx}{dt}(t) &= \begin{bmatrix} -0.0193 & 8.82 & -32.2 & -0.48 \\ -0.000254 & -1.02 & 0 & 0.91 \\ 0 & 0 & 0 & 1 \\ 0 & 0.82 & 0 & -1.08 \end{bmatrix} x(t) + \begin{bmatrix} 0.17 \\ -0.00215 \\ 0 \\ -0.18 \end{bmatrix} u_e(t) \\ y(t) &= \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \end{bmatrix} x(t), \end{aligned}$$

where $x = [V \ \alpha \ \theta \ q]^T$ with V the velocity (ft/s), α the angle of attack (rad), θ the pitch angle (rad), and q the pitch rate (rad/s). The control input u_e is the elevator angle deflection (deg). The outputs are the pitch angle θ and the flight path angle $\theta - \alpha$.

The linearized model is discretized in time using a first order hold method with time step 0.1 s, resulting in an LTI discrete-time state-space model. We assume that the dynamics of the true system can be written as the discrete-time model plus normally distributed random white process noise with variance $Q = 10^{-8}I$ and measurement noise with variance $R = 10^{-7}I$, meaning that the pitch angle and the flight path angle can be measured accurately up to one tenth of a degree.

In a first experiment, an actuator fault is simulated. It is assumed that the actuator that steers the elevator to the desired position fails at time step 50. The desired elevator angle and the actual value due to the failure of the actuator are shown in Fig. 4.4. In order to deal with actuator faults, we consider the elevator angle deflection as an unknown input. Notice that this unknown input only enters the state equation so that the true system can be written as (4.39). We now apply the filter summarized in Sect. 4.3.4 to simultaneously estimate the elevator angle deflection and the system state. The estimate of the elevator angle deflection is shown in Fig. 4.4. An estimate of the actuator fault can be obtained by subtracting the desired elevator deflection from the estimated one. Fig. 4.5 shows true and estimated values of the pitch rate. For the purpose of comparison, the Kalman filter estimate and the pitch rate that would be obtained if no offset on the actuator were present, are also shown. As can be seen, the joint input-state estimator follows the true trajectory very closely. The Kalman filter, on the other hand, diverges at the time the actuator fault occurs because it gives too much weight to the model equations.

In a second experiment, a sensor fault is simulated. More precisely, it is assumed that the flight path measurement is subject to an unknown disturbance. The objective is to estimate this disturbance based on the measurement of the pitch angle. Notice that the disturbance can be modeled as an unknown input that enters the output equation. We thus apply the filter summarized in Sect. 4.2.4 to simultaneously estimate the system state and the sensor error. We assume that $Q = 10^{-4}I$ and $R = 10^{-4}I$. The true and estimated value of the sensor errors are compared in Fig. 4.6. The estimator reconstructs the sensor error with high precision. \square

Example 4.2. Optimal smoothing

In a second example, we investigate the benefit of smoothing on the estimation accuracy. Consider again the system of Example 3.1, but now subject to normally distributed random white noise processes $\{v_{[k]}\}_{k=0}^{\infty}$ and $\{w_{[k]}\}_{k=0}^{\infty}$ with covariance matrices $Q = 10^{-3}I$ and $R = 10^{-2}I$, respectively. The initial state of the system equals $x_{[0]} = 0$. It is assumed that an unbiased estimate $\hat{x}_{[0]}$ is available with covariance matrix $P_{[0]} = 10^{-2}I$.

We compare the estimation accuracy of the joint input-state estimator developed in Sect. 4.4.3 for $L = 0$ (filtering) and $L = 5$ (smoothing). It is found

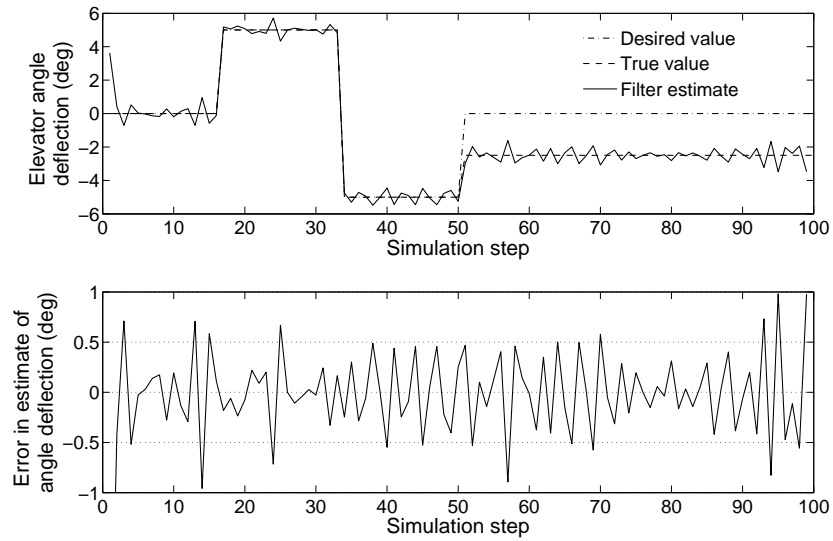


Figure 4.4: Actuator fault detection in a linearized F16 model. Top figure: comparison between desired value, true value and estimated value of the elevator angle deflection. Bottom figure: error in the estimate of the elevator angle deflection.

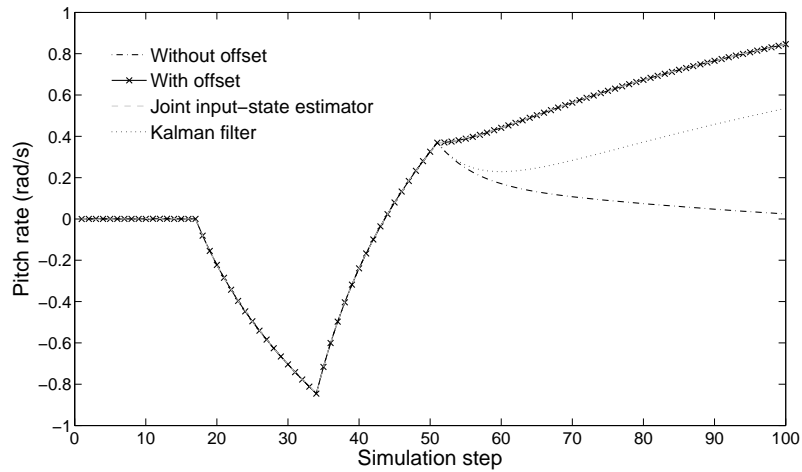


Figure 4.5: Actuator fault detection in a linearized F16 model. The joint input-state estimator follows the true trajectory of the pitch rate very closely. For the purpose of comparison, the Kalman filter estimate and pitch rate that would be obtained if no offset on the actuator were present, are also shown.

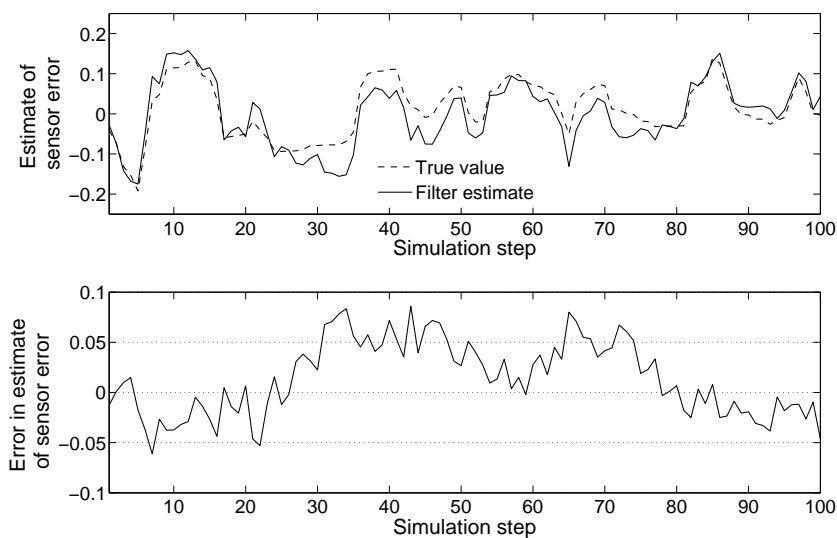


Figure 4.6: *Sensor fault detection in a linearized F16 model. Top figure: comparison between true and estimated value of sensor error. Bottom figure: error in the estimate of the sensor error.*

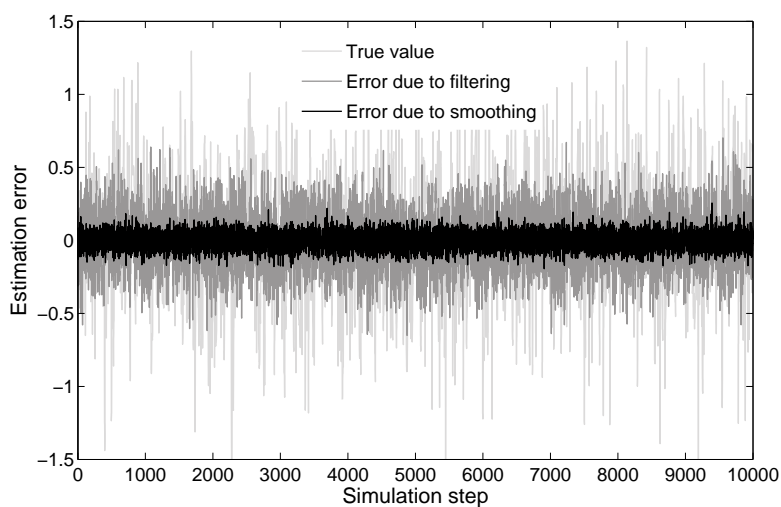


Figure 4.7: *Effect of smoothing on the estimation accuracy in Example 4.2. The errors in the smoothed ($L = 5$) and the filtered ($L = 0$) estimates of the system input are compared. The true value of the unknown input is also shown. Smoothing clearly increases estimation accuracy.*

that the error covariance matrices of the state estimator and input estimator converge as $k \rightarrow \infty$. For $L = 0$, they converge to

$$P_{[\infty]} \approx \begin{bmatrix} 0.1644 & -0.0986 \\ -0.0986 & 0.0627 \end{bmatrix}, \quad P_{u[\infty]} \approx 0.0406.$$

For $L = 5$, we obtain

$$P_{[\infty]} \approx \begin{bmatrix} 0.0078 & -0.0038 \\ -0.0038 & 0.0038 \end{bmatrix}, \quad P_{u[\infty]} \approx 0.0035.$$

These results indicate that the smoothed estimates are more accurate than the filtered ones. The errors in the smoothed and filtered estimates of the system input are compared in Fig. 4.7. This figure confirms that estimation accuracy is highest in the case of smoothing. \square

4.6 Conclusion

This chapter has studied the problems of state estimation and joint input-state estimation for combined deterministic-stochastic systems. Existing techniques were mainly concerned with the state estimation problem and in particular with filtering.

In a first contribution, the problem of joint input-state filtering was addressed. Based on MVU estimation, filter equations were derived in which the estimation of the system state and the unknown input are interconnected. Next, the relation to least-squares estimation was established. More precisely, it was shown that the filtered state and input estimates can be obtained as the solution of a least-squares problem that takes the same form as that of the Kalman filter, except that the inputs are now unknown and thus have to be estimated. Solution of the least-squares problem has provided information formulas for joint input-state filtering. By establishing duality relations to the Kalman filter equations, square-root information formulas were derived almost immediately. Finally, it was shown that square-root covariance filtering in the presence of unknown inputs is not possible.

In a second contribution, the one step ahead prediction, filtering and smoothing problems for both state estimation and joint input-state estimation were put in a general framework. This framework generalizes all previous results. The derivation of the joint input-state estimator is closely related to that of the inverse system in Chapter 3. More precisely, the joint input-state estimator can at every time instant be interpreted as the inverse of the corresponding deterministic system.

Numerical examples have indicated that the covariance matrices of the joint input-state estimators converge. Further research should investigate convergence conditions and properties of the corresponding difference equation.

Chapter 5

Applications of System Inversion

This chapter considers four applications of system inversion. First, filtering with noisy inputs and outputs is addressed. Recursive filter equations are derived in which the estimation of the system state and the input are interconnected. Next, the problem of filtering in the presence of bias is considered. A suboptimal filter, closely related to the two-stage Kalman filter [45], is developed. The last two applications are more practical. First, model error estimation and dynamic model updating is addressed. An empirical technique is outlined to correct a physical model for unknown dynamics. Next, an approach to joint state and boundary condition estimation is considered in which the spatial component of the boundary condition is expanded as a linear combination of orthogonal basis functions.

5.1 Introduction

One application of system inversion, namely fault detection, has already been addressed in Example 4.1. In this chapter, four other applications are considered.

Personal contributions

The personal contribution of this chapter is the application of the inversion procedure developed in Chapters 3 and 4 in four concrete problems.

- In Sect. 5.2, a new solution to the errors-in-variables filtering problem is derived in which the estimation of the system state and the unknown input are interconnected. The solution is shown to be algebraically equivalent to existing solutions.

- In Sect. 5.3, a new solution to the optimal filtering problem in the presence of bias errors is derived. A suboptimal filter, closely related to the two-stage Kalman filter [45] is developed. The major difference is that the new filter can be used also if the equation governing the dynamical evolution of the bias error is unknown.
- In Sect. 5.4, model error estimation and dynamic model updating is addressed. An empirical technique is outlined to update a non-satisfactory accurate physical state-space model. The technique consists in first estimating the model error and then identifying an empirical correction model based on the estimated data.
- In Sect. 5.5, a new approach to the estimation of unknown boundary conditions is considered in which the temporal component of the boundary is assumed to be unknown and the spatial form is expanded as a linear combination of orthogonal basis functions.

Chapter outline

This chapter consists of four sections. Each section is devoted to one application. Section 5.2 addresses filtering with noisy input and output measurements. Next, in Sect. 5.3, the problem of optimal filtering in the presence of bias errors is considered. Section 5.4 deals with model error estimation and dynamic model updating. Finally, in Sect. 5.5, the problem of joint state and boundary condition estimation is addressed.

5.2 Filtering with noisy inputs and outputs

Closely related to system inversion is the *noisy input-output* filtering problem, in which the system input is known up to an additive noise term. The noisy input-output filtering problem is first considered in [62], where it is called *errors-in-variables* filtering. The treatment of [62] is, however, limited to SISO systems and is not linked to the classical Kalman filter. The MIMO case is first addressed in [94], where it is shown that the problem can be translated into a standard Kalman filtering problem. A similar result is obtained in [31, 93].

In this section, we address an extension of the noisy input-output filtering problem. We consider the case where a linear combination of the input vector is measured instead of the entire input vector. We show that the resulting filtering problem can be reformulated as that considered in Sect. 4.3 and derive filter equations in which the estimation of the system state and the unknown input are interconnected. As a special case, the filter provides a new solution to the errors-in-variables filtering problem which is shown to be algebraically equivalent to the filters in [31, 93].

5.2.1 Problem formulation

Consider the set-up in Fig. 5.1. This set-up consists of a discrete-time LTI system and two sensors. The input $u_{[k]}$ of the LTI system is assumed to be unknown. One of the sensor measures a noisy linear combination $z_{2[k]}$ of the system input $u_{[k]}$. The other sensor measures a noisy linear combination $z_{1[k]}$ of the system output $y_{[k]}$.

The equations of the system and the sensors are assumed to be given by:

- **System:**

The dynamics of the LTI system are assumed to be governed by

$$x_{[k+1]} = Ax_{[k]} + Bu_{[k]} + w_{[k]} \quad (5.1)$$

$$y_{[k]} = Cx_{[k]} + Du_{[k]}, \quad (5.2)$$

where $x_{[k]} \in \mathbb{R}^n$ denotes the state vector at time instant k , $u_{[k]} \in \mathbb{R}^m$ denotes the unknown input vector at time k , and $y_{[k]}$ denotes the output vector at time k . The noise process $\{w_{[k]}\}_{k=0}^{\infty}$ is assumed to be a zero-mean white stationary stochastic process with covariance matrix Q . The initial state $x_{[0]}$ is assumed to be unknown and random.

- **Sensor 1:**

It is assumed that sensor 1 measures the following linear combination of the output $y_{[k]}$ of (5.1),

$$z_{1[k]} = S_1 y_{[k]} + v_{1[k]},$$

where $z_{1[k]} \in \mathbb{R}^{p_1}$ denotes the measurement at time instant k and where the noise process $\{v_{1[k]}\}_{k=0}^{\infty}$ is assumed to be zero-mean stationary white and uncorrelated to $\{w_{[k]}\}_{k=0}^{\infty}$. We define $\mathbb{E}[v_{1[k]} v_{1[k]}^T] =: R_1$.

- **Sensor 2:**

It is assumed that sensor 2 measures the following linear combination of the input $u_{[k]}$ of (5.1),

$$z_{2[k]} = S_2 u_{[k]} + v_{2[k]},$$

where $z_{2[k]} \in \mathbb{R}^{p_2}$ denotes the measurement at time instant k and where the noise process $\{v_{2[k]}\}_{k=0}^{\infty}$ is assumed to be zero-mean stationary white. We define $\mathbb{E}[v_{2[k]} v_{2[k]}^T] =: R_2$. The noise processes $\{v_{2[k]}\}_{k=0}^{\infty}$ is assumed to be uncorrelated to $\{w_{[k]}\}_{k=0}^{\infty}$ and $\{v_{1[k]}\}_{k=0}^{\infty}$.

The interconnection of the system and both sensors yields the LTI system

$$x_{[k+1]} = Ax_{[k]} + Bu_{[k]} \quad (5.3a)$$

$$\underbrace{\begin{bmatrix} z_{1[k]} \\ z_{2[k]} \end{bmatrix}}_{z_{[k]}} = \underbrace{\begin{bmatrix} \bar{C}_1 \\ 0 \end{bmatrix}}_{\bar{C}} x_{[k]} + \underbrace{\begin{bmatrix} \bar{D}_1 \\ \bar{D}_2 \end{bmatrix}}_{\bar{D}} u_{[k]} + \underbrace{\begin{bmatrix} v_{1[k]} \\ v_{2[k]} \end{bmatrix}}_{v_{[k]}}, \quad (5.3b)$$

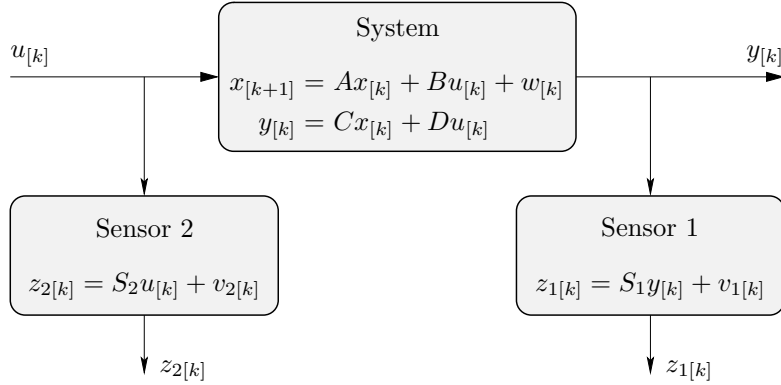


Figure 5.1: *Set-up of the filtering problem with noisy input and output measurements.*

where $\bar{C}_1 := S_1C$, $\bar{D}_1 := S_1D$, $\bar{D}_2 := S_2$ and where we have eliminated the output $y[k]$ of the LTI system (5.1). We define $R := \text{diag}(R_1, R_2)$.

The objective of this section is to design an optimal filter that recursively estimates both the system state $x[k]$ and the unknown input $u[k]$ from knowledge of the sequences of measurements $\{z_{1[i]}\}_{i=0}^k$ and $\{z_{2[i]}\}_{i=0}^k$. It is assumed that the pair $\{A, C\}$ is observable and that an unbiased estimate $\hat{x}_{[0|-1]}$ of the initial state $x_{[0]}$ is available with covariance matrix $P_{[0|-1]}$. The error in the initial estimate $\hat{x}_{[0|-1]}$ is assumed to be uncorrelated to $\{v_{[k]}\}_{k=0}^{\infty}$ and $\{w_{[k]}\}_{k=0}^{\infty}$.

5.2.2 Errors-in-variables filtering

Note that the errors-in-variables filtering problem considered in [31, 93] is obtained for $\bar{D}_2 = S_2 = I$, that is, when sensor 2 measures the entire input vector. Substituting $u[k] = z_{2[k]} - v_{2[k]}$ in (5.3), then yields the following system which is decoupled from the unknown input,

$$\begin{aligned} x_{[k+1]} &= Ax_{[k]} + Bz_{2[k]} + \tilde{w}_{[k]} \\ z_{1[k]} &= \bar{C}_1x_{[k]} + \bar{D}_1z_{2[k]} + \tilde{v}_{1[k]}, \end{aligned}$$

where $\tilde{w}_{[k]} := -Bv_{2[k]}$ is correlated to $\tilde{v}_{1[k]} := v_{1[k]} - \bar{D}_1v_{2[k]}$. The optimal filtering problem for the system (5.3) has thus been transformed into a standard Kalman filtering problem for a system with correlated noise processes. The solution of the resulting Kalman filtering problem, together with equations for the optimal estimate of the input can be found in e.g. [93].

5.2.3 Filtering with noisy input measurements

In this section, we extend the errors-in-variables filtering problem to the case where a linear combination of the input vector is measured instead of the entire

input vector, that is, we do no longer assume that $\bar{D}_2 = I$.

It is readily observed that the filtering problem formulated in Sect. 5.2.1 is a special case of the joint input-state filtering problem addressed in Sect. 4.2. Based on the equations summarized in Sect. 4.2.4, we now derive explicit filter formulas by exploiting the specific structure of the output equation of (5.3). Like in Sect. 4.2.4, the equations are split into a time update, a measurement update and a step in which the unknown input is estimated.

We assume that $\text{rank}(D) = m$, which, as shown in Sect. 4.2, is a necessary and sufficient condition for MVU estimation of the unknown input. We use the same notations as in Sect. 4.2.

5.2.3.1 Time update

The time update is given by (4.31)-(4.32),

$$\begin{aligned}\hat{x}_{[k+1|k]} &= A\hat{x}_{[k|k]} + B\hat{u}_{[k|k]} \\ P_{[k+1|k]} &= \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} P_{[k|k]} & P_{xu[k|k]} \\ P_{ux[k|k]} & P_{u[k|k]} \end{bmatrix} \begin{bmatrix} A^\top \\ B^\top \end{bmatrix} + Q.\end{aligned}$$

5.2.3.2 Measurement update

For a reformulation of the measurement update (4.27)-(4.30), we first note from (5.3) that

$$z_{[k]} - \bar{C}\hat{x}_{[k|k-1]} = \begin{bmatrix} z_{1[k]} - \bar{C}_1\hat{x}_{[k|k-1]} \\ z_{2[k]} \end{bmatrix}. \quad (5.4)$$

Furthermore, it follows from (5.3) that (4.23) can now be written as $\tilde{R}_{[k]} := \text{diag}(\tilde{R}_{1[k]}, R_2)$, with $\tilde{R}_{1[k]} = \bar{C}_1 P_{[k|k-1]} \bar{C}_1^\top + R_1$. Substituting the latter expression for $\tilde{R}_{[k]}$ in (4.27), yields

$$L_{[k]} = [L_{1[k]} \quad 0], \quad (5.5)$$

where $L_{1[k]} \in \mathbb{R}^{n \times p_1}$ is given by $L_{1[k]} = P_{[k|k-1]} \bar{C}_1^\top \tilde{R}_{1[k]}^{-1}$. Finally, substituting (5.4) and (5.5) in (4.27)-(4.30), yields

$$\begin{aligned}\hat{x}_{[k|k]} &= \hat{x}_{[k|k-1]} + L_{1[k]}(z_{1[k]} - \bar{C}_1\hat{x}_{[k|k-1]} - \bar{D}_1\hat{u}_{[k|k]}) \\ P_{[k|k]} &= P_{[k|k-1]} - L_{1[k]}(\tilde{R}_{1[k]} - \bar{D}_1 P_{u[k|k]} \bar{D}_1^\top) L_{1[k]}^\top \\ P_{xu[k|k]} &= P_{ux[k|k]}^\top = -L_{1[k]} \bar{D}_1 P_{u[k|k]}.\end{aligned} \quad (5.6)$$

5.2.3.3 Estimation of unknown input

By substituting (5.4) in (4.24)-(4.26), we obtain the following equations for the estimation of the input,

$$\hat{u}_{[k|k]} = P_{u[k|k]} \left[\bar{D}_1^\top \tilde{R}_{1[k]}^{-1} (z_{1[k]} - \bar{C}_1\hat{x}_{[k|k-1]}) + \bar{D}_2^\top R_2^{-1} z_{2[k]} \right] \quad (5.7)$$

$$P_{u[k|k]} = (\bar{D}_1^T \tilde{R}_{1[k]}^{-1} \bar{D}_1 + \bar{D}_2^T R_2^{-1} \bar{D}_2)^{-1}. \quad (5.8)$$

In case of errors-in-variables filtering ($\bar{D}_2 = I$), equations (5.7) and (5.8) can be written into a more convenient form, as will now be shown. Substituting $\bar{D}_2 = I$ in (5.8) and applying the matrix inversion lemma, yields

$$P_{u[k|k]} = (I - L_{u[k]} \bar{D}_1) R_2, \quad (5.9)$$

where the gain matrix $L_{u[k]}$ is defined by

$$L_{u[k]} := R_2 \bar{D}_1^T (\bar{D}_1 R_2 \bar{D}_1^T + \tilde{R}_{1[k]})^{-1}.$$

Furthermore, substituting (5.9) in (5.7), yields the following estimate of the input,

$$\hat{u}_{[k|k]} = z_{2[k]} + L_{u[k]} (z_{1[k]} - \bar{C}_1 \hat{x}_{[k|k-1]} - \bar{D}_1 z_{2[k]}).$$

5.2.4 Summary of filter equations

The filter equations derived above can be split into three steps: the estimation of the unknown input, the measurement update and the time update. These steps are given by:

Filtering with noisy inputs and outputs

- Estimation of unknown input:

$$\begin{aligned} \tilde{z}_{1[k]} &= z_{1[k]} - \bar{C}_1 \hat{x}_{[k|k-1]} \\ \tilde{R}_{1[k]} &= \bar{C}_1 P_{[k|k-1]} \bar{C}_1^T + R_1 \end{aligned}$$

– General case

$$\begin{aligned} \hat{u}_{[k|k]} &= P_{u[k|k]} (\bar{D}_1^T \tilde{R}_{1[k]}^{-1} \tilde{z}_{1[k]} + \bar{D}_2^T R_2^{-1} z_{2[k]}) \\ P_{u[k|k]} &= (\bar{D}_1^T \tilde{R}_{1[k]}^{-1} \bar{D}_1 + \bar{D}_2^T R_2^{-1} \bar{D}_2)^{-1} \end{aligned}$$

– Errors-in-variables filtering ($\bar{D}_2 = I$)

$$\begin{aligned} \hat{u}_{[k|k]} &= z_{2[k]} + L_{u[k]} (\tilde{z}_{1[k]} - \bar{D}_1 z_{2[k]}) \\ L_{u[k]} &= R_2 \bar{D}_1^T (\bar{D}_1 R_2 \bar{D}_1^T + \tilde{R}_{1[k]})^{-1} \\ P_{u[k|k]} &= (I - L_{u[k]} \bar{D}_1) R_2 \end{aligned}$$

- **Measurement update:**

$$\begin{aligned}
 L_{1[k]} &= P_{[k|k-1]} \bar{C}_1^\top \tilde{R}_{1[k]}^{-1} \\
 \hat{x}_{[k|k]} &= \hat{x}_{[k|k-1]} + L_{1[k]} (\tilde{z}_{1[k]} - \bar{D}_1 \hat{u}_{[k|k]}) \\
 P_{[k|k]} &= P_{[k|k-1]} - L_{1[k]} (\tilde{R}_{1[k]} - \bar{D}_1 P_{u[k|k]} \bar{D}_1^\top) L_{1[k]}^\top \\
 P_{xu[k|k]} &= P_{ux[k|k]}^\top = -L_{1[k]} \bar{D}_1 P_{u[k|k]}
 \end{aligned}$$

- **Time update:**

$$\begin{aligned}
 \hat{x}_{[k+1|k]} &= A \hat{x}_{[k|k]} + B \hat{u}_{[k|k]} \\
 P_{[k+1|k]} &= \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} P_{[k|k]} & P_{xu[k|k]} \\ P_{ux[k|k]} & P_{u[k|k]} \end{bmatrix} \begin{bmatrix} A^\top \\ B^\top \end{bmatrix}
 \end{aligned}$$

A block diagram of the errors-in-variables filter summarized above is given in Fig. 5.2.

In contrast to the results in [31,93], the equations for the errors-in-variables filter derived in this section are written in a form in which the state estimator and the input estimator exchange information in both directions, that is, in contrast to existing results, the state estimator is written in a form that reveals optimal estimates of the unknown input. By eliminating the first step, i.e. the estimation of the unknown input, it can be shown that the errors-in-variables filter derived above, is algebraically equivalent to the filters in [31,93]. A proof of equivalence can be found in [51].

5.2.5 Numerical example

Example 5.1.

Consider the system described by

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right] = \left[\begin{array}{ccc|cc} 0 & 1 & 0 & 1 & 0 \\ -0.3 & 0.4 & -0.2 & 0 & 0.5 \\ -0.1 & 0.2 & 0.4 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{array} \right],$$

with initial state $x_{[0]} = 0$. The sensors are assumed to be described by $\bar{S}_1 = I$ and $\bar{S}_2 = [1 \ 0]$. Note that the errors-in-variables filters of [31,93] can not be applied to this system because $\bar{D}_2 = \bar{S}_2 \neq I$. The noise processes are assumed to be characterized by $R_1 = 0.025I$ and $R_2 = 0.025$. Fig. 5.3 compares the true and estimated values of the second component of the unknown input vector. Simulation results were obtained with $\hat{x}_{[0|-1]} = 0$ and $P_{[0|-1]} = 10^{-2}I$. \square

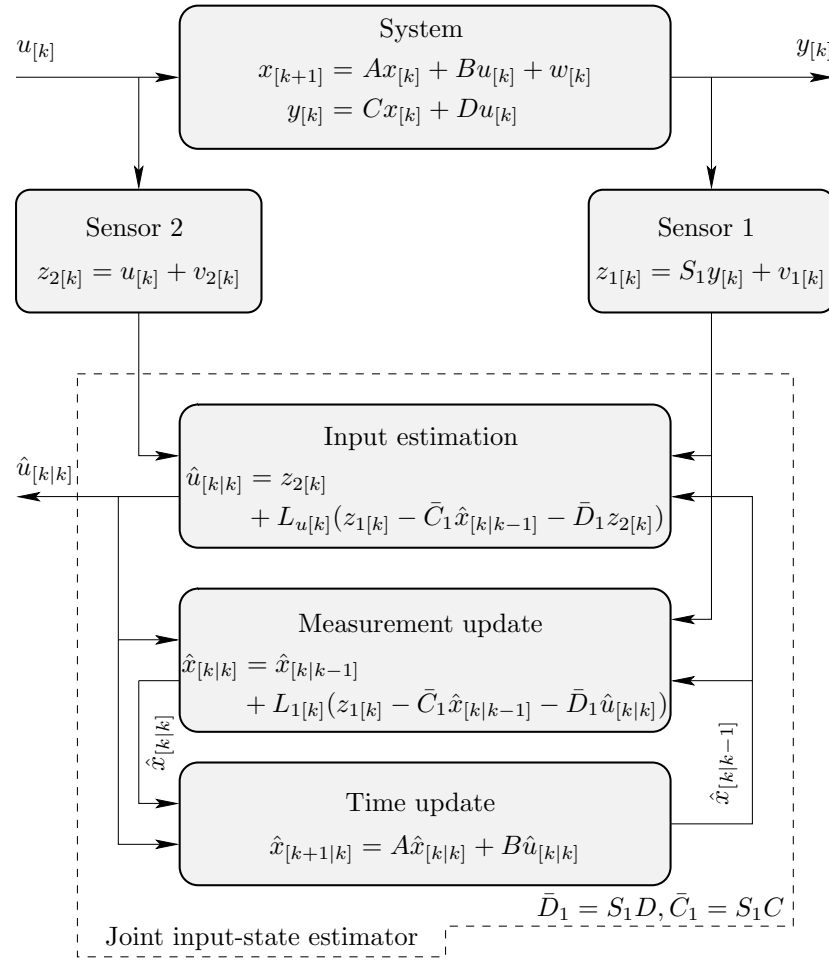


Figure 5.2: Block diagram for the errors-in-variables filter summarized in Sect. 5.2.4.

5.3 Filtering in the presence of bias

In many applications, the numerical model is subject to an additive error of which the actual value is unknown, but the equations governing its dynamics are known. Such errors are called *bias errors*. The most common type of bias errors, is the constant bias error, in which the error obeys $u_{[k+1]} = u_{[k]}$.

The problem of optimal filtering in the presence of bias errors has received a lot of attention in the past. The optimal solution of the problem consists in augmenting the state vector with the vector $u_{[k]}$ of bias errors and then estimating both of them using the Kalman filter. This procedure is called

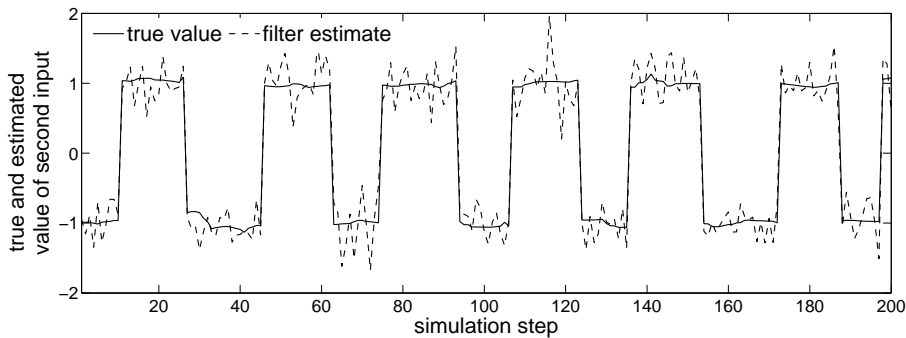


Figure 5.3: Comparison between true and estimated value of the input in Example 5.1.

augmented state filtering. To reduce the computational cost of the augmented state filter, Friedland [45] proposed the *two-stage filter*, in which the estimation of the system state and the bias error are separated. As the name suggests, the two-stage filter consists of two stages. The first stage yields a bias-unaware estimate of the system state. The second stage consists in estimating the bias error. Both stages are computed in parallel (with little exchange of information) and their results are merged afterwards, yielding bias-aware state estimates. A detailed treatment of two-stage filtering can be found in e.g. [3, 29, 30, 75].

An extension of Friedland's algorithm to bias models with a stochastic component, i.e. bias models of the form $u_{[k+1]} = u_{[k]} + \mu_{[k]}$ with $\mu_{[k]}$ a zero-mean random vector, were first considered in [75]. It has become common practice to use such a bias model in case the dynamics of the bias error are unknown. The variance of $\mu_{[k]}$ is then a design parameter that should be carefully chosen.

Dee and Da Silva [29,30] further decreased computational costs by deriving a suboptimal variant of the two-stage filter. In their approach, the state estimator and the bias estimator are interconnected. More precisely, a feedback from the bias estimator to the state estimator is introduced, making the state estimator no longer bias-blind.

In this section, we consider a system without direct feedthrough of the input to the output and derive a bias filter by incorporating a bias model of the form $u_{[k+1]} = u_{[k]} + \mu_{[k]}$ into the joint input-state estimator of Sect. 4.3. In contrast to existing techniques, the filter considered in the section estimates the bias with one step delay, that is, it estimates $u_{[k-1]}$ based on measurements up to time instant k . This delay in estimation introduces a major advantage over the augmentation method. It allows to switch to the joint input-state estimator of Sect. 4.3 and back during operation. As will be show in Example 5.2, such a switching regime is especially useful if e.g. the bias is constant for some time interval and then suddenly undergoes an abrupt and unknown change.

5.3.1 Derivation of filter equations

We consider a system without direct feedthrough of the unknown input to the output, i.e. a system of the form

$$x_{[k+1]} = Ax_{[k]} + Bu_{[k]} + w_{[k]} \quad (5.10a)$$

$$y_{[k]} = Cx_{[k]} + v_{[k]}, \quad (5.10b)$$

where $x_{[k]} \in \mathbb{R}^n$ denotes the state vector at time instant k , $u_{[k]} \in \mathbb{R}^m$ denotes the bias error at time k , and $y_{[k]} \in \mathbb{R}^p$ denotes the measurement at time k . The initial state $x_{[0]}$ is assumed to be a random variable. The noise processes $\{w_{[k]} \in \mathbb{R}^n\}_{k=0}^{\infty}$ and $\{v_{[k]} \in \mathbb{R}^p\}_{k=0}^{\infty}$ are assumed to be stochastic with the properties given in Assumption 2.1. We define $Q := \mathbb{E}[w_{[k]}w_{[k]}^T]$ and $R := \mathbb{E}[v_{[k]}v_{[k]}^T]$ and assume that R is positive definite.

We assume that the dynamical evolution of the bias error is governed by $u_{[k+1]} = u_{[k]} + \mu_{[k]}$ where $\{\mu_{[k]}\}_{k=0}^{\infty}$ is a zero-mean stationary white noise process, assumed to be uncorrelated to $\{w_{[k]}\}_{k=0}^{\infty}$ and $\{v_{[k]}\}_{k=0}^{\infty}$. We define $Q_u := \mathbb{E}[\mu_{[k]}\mu_{[k]}^T]$.

We assume that unbiased estimates $\hat{x}_{[0|0]}$ and $\hat{u}_{[0|0]}$ of the initial state and the bias vector are available with covariance matrices $P_{[0|0]}$ and $P_{u[0|0]}$. The errors in $\hat{x}_{[0|0]}$ and $\hat{u}_{[0|0]}$ are assumed to be uncorrelated to $\{w_{[k]}\}_{k=0}^{\infty}$, $\{v_{[k]}\}$ and $\{\mu_{[k]}\}_{k=0}^{\infty}$.

Like the joint input-state estimator of Sect. 4.3, we consider a filter that consist of three steps: a time update, a step in which the bias is estimated and a measurement update. These three steps are now addressed.

5.3.1.1 Time update

Assume that knowledge of the measurements up to time instant $k-1$ has provided us with estimate $\hat{x}_{[k-1|k-1]}$ and $\hat{u}_{[k-2|k-1]}$. Like in Sect. 4.3, we define $\bar{x}_{[k]} := Ax_{[k-1]} + w_{[k-1]}$, and consider a time update in which $\bar{x}_{[k]}$ is estimated instead of $x_{[k]}$. More precisely, we assume that $\hat{x}_{[k-1|k-1]}$ is unbiased and has covariance matrix $P_{[k-1|k-1]}$ and we consider a time update of the form

$$\hat{\bar{x}}_{[k|k-1]} = A\hat{x}_{[k-1|k-1]}. \quad (5.11)$$

As shown in Sect. 4.3, the update of the error covariance matrix is then given by

$$\bar{P}_{[k|k-1]} = AP_{[k-1|k-1]}A^T + Q,$$

where $\bar{P}_{[k|k-1]}$ denotes the error covariance matrix of $\hat{\bar{x}}_{[k|k-1]}$. Also, we assume that $\hat{u}_{[k-2|k-1]}$ is unbiased and has error covariance matrix $P_{u[k-2|k-1]}$ and we consider a time update of the bias estimate of the form

$$\hat{u}_{[k-1|k-1]} = \hat{u}_{[k-2|k-1]}.$$

It is easily verified that the update of the error covariance matrix is given by

$$P_{u[k-1|k-1]} = P_{u[k-2|k-1]} + Q_u,$$

where $P_{u[k-1|k-1]}$ denotes the error covariance matrix of $\hat{u}_{[k-1|k-1]}$. Notice that the estimation of the system state runs one time step ahead of that of the bias vector.

5.3.1.2 Estimation of bias

In contrast to the joint input-state estimation problem of Sect. 4.3, where only one source of information about $u_{[k-1]}$ is available, namely the innovation $y_{[k]} - C\hat{x}_{[k|k-1]}$, here two sources of information are available, namely the innovation and $\hat{u}_{[k-1|k-1]}$. It follows from (4.51) that the LS problem obtained by combining the information about $u_{[k-1]}$ contained in the innovation and in $\hat{u}_{[k-1|k-1]}$, is given by

$$\hat{u}_{[k-1|k]} = \arg \min_{u_{[k-1]}} \left\| \begin{bmatrix} y_{[k]} - C\hat{x}_{[k|k-1]} \\ \hat{u}_{[k-1|k-1]} \end{bmatrix} - \begin{bmatrix} F \\ I \end{bmatrix} u_{[k-1]} \right\|_{W_{[k]}}^2, \quad (5.12)$$

where $F := CB$ and where $W_{[k]}$ denotes the weighting matrix of the LS problem, which we choose as $W_{[k]} = \text{diag}(\bar{R}_{[k]}^{-1}, P_{u[k-1|k-1]}^{-1})$, where $\bar{R}_{[k]} := \mathbb{E}[\bar{e}_{[k]}\bar{e}_{[k]}^T]$ with $\bar{e}_{[k]} := y_{[k]} - C\hat{x}_{[k|k-1]} - Fu_{[k-1]}$. It follows from (5.10) and (5.11) that

$$\bar{e}_{[k]} = C\tilde{x}_{[k|k-1]} + v_{[k]},$$

with $\tilde{x}_{[k|k-1]} := x_{[k]} - \hat{x}_{[k|k-1]}$. Consequently, $\bar{R}_{[k]}$ is given by $\bar{R}_{[k]} = C\bar{P}_{[k|k-1]}C^T + R$. Note that choosing the weighting matrix $W_{[k]}$ as given above does in general not yield the MVU estimate of $u_{[k-1]}$. The reason is that $\bar{e}_{[k]}$ is correlated to the error in $\hat{u}_{[k-1|k-1]}$. Although the choice of the diagonal weighting matrix is suboptimal from an MVU point of view, we proceed with it for conciseness of equations.

The solution to (5.12) can then be written as

$$\hat{u}_{[k-1|k]} = \hat{u}_{[k-1|k-1]} + K_{u[k]}(y_{[k]} - C\hat{x}_{[k|k-1]} - F\hat{u}_{[k-1|k-1]}), \quad (5.13)$$

where $K_{u[k]}$ is given by

$$K_{u[k]} = P_{u[k-1|k-1]}F^T(FP_{u[k-1|k-1]}F^T + \bar{R}_{[k]})^{-1}.$$

The error covariance matrix $P_{u[k-1|k]}$ of $\hat{u}_{[k-1|k]}$ is given by

$$P_{u[k-1|k]} = (I - K_{u[k]}F)P_{u[k-1|k-1]}.$$

5.3.1.3 Measurement update

We consider a measurement update that is similar to that of the joint input-state estimator of Sect. 4.3. It consists of two steps,

$$\hat{x}_{[k|k]} = \hat{x}_{[k|k-1]} + B\hat{u}_{[k-1|k]} \quad (5.14)$$

$$\hat{x}_{[k|k]} = \hat{x}_{[k|k]} + \bar{L}_{[k]}(y_{[k]} - C\hat{x}_{[k|k]}). \quad (5.15)$$

It is straightforward to show that both $\hat{x}_{[k|k]}$ and $\hat{x}_{[k|k]}$ are unbiased estimates of $x_{[k]}$. The calculation of the gain matrix $\bar{L}_{[k]}$ that minimizes the variance of $\hat{x}_{[k|k]}$ is straightforward, but quite involved and is therefore given in Appendix C.3. The resulting equations are summarized in the next section.

5.3.2 Summary of filter equations

The filter equations can be split into three steps: the time update, the estimation of the bias vector and the measurement update. These steps are given by:

Filtering in the presence of bias

- **Time update:**

$$\begin{aligned} \hat{x}_{[k|k-1]} &= A\hat{x}_{[k-1|k-1]} \\ \bar{P}_{[k|k-1]} &= AP_{[k-1|k-1]}A^T + Q \\ \hat{u}_{[k-1|k-1]} &= \hat{u}_{[k-2|k-1]} \\ P_{u[k-1|k-1]} &= P_{u[k-2|k-1]} + Q_u \end{aligned}$$

- **Bias estimation:**

$$\begin{aligned} \hat{u}_{[k-1|k]} &= \hat{u}_{[k-1|k-1]} + K_{u[k]}(y_{[k]} - C\hat{x}_{[k|k-1]} - F\hat{u}_{[k-1|k-1]}) \\ K_{u[k]} &= P_{u[k-1|k-1]}F^T(FP_{u[k-1|k-1]}F^T + \bar{R}_{[k]})^{-1} \\ \bar{R}_{[k]} &= C\bar{P}_{[k|k-1]}C^T + R \\ P_{u[k-1|k]} &= (I - K_{u[k]}F)P_{u[k-1|k-1]} \end{aligned}$$

- **Measurement update:**

$$\begin{aligned} \hat{x}_{[k|k]} &= \hat{x}_{[k|k-1]} + B\hat{u}_{[k-1|k]} \\ \bar{P}_{[k|k]} &= (I - BK_{u[k]}C)(\bar{P}_{[k|k-1]} + BP_{u[k-1|k]}B^T)(I - BK_{u[k]}C)^T \\ &\quad + BK_{u[k]}RK_{u[k]}^TB^T \\ \hat{x}_{[k|k]} &= \hat{x}_{[k|k]} + \bar{L}_{[k]}(y_{[k]} - C\hat{x}_{[k|k]}) \\ \bar{L}_{[k]} &= \bar{P}_{[k|k-1]}C^T\bar{R}_{[k]}^{-1} \\ P_{[k|k]} &= (I - \bar{L}_{[k]}C)\bar{P}_{[k|k]}(I - \bar{L}_{[k]}C)^T + \bar{L}_{[k]}R\bar{L}_{[k]}^T \\ &\quad + (I - \bar{L}_{[k]}C)BK_{u[k]}R\bar{L}_{[k]}^T + \bar{L}_{[k]}RK_{u[k]}^TB^T(I - \bar{L}_{[k]}C)^T \end{aligned}$$

The structure of the filter equations summarized above is very similar to that of the suboptimal filter proposed by Dee and Da Silva [30]. The main difference is that we estimate the bias with one step delay. Also, the equations are very

similar to those of the joint input-state estimator summarized in Sect. 4.3.4. In fact, one can switch from the bias filter to the joint input-state estimator of Sect. 4.3.4 and vice versa during operation. As will be shown in the next example, such a switching is especially useful if e.g. the bias is constant for some time interval and then suddenly undergoes an abrupt and unknown change.

5.3.3 Numerical example

Example 5.2. Fault reconstruction in an F16 aircraft

Consider again the actuator fault in the linearized F16 model of Example 4.1. In this example, we model the elevator angle deflection as $u_{[k+1]} = u_{[k]} + \mu_{[k]}$, where the zero-mean random vector $\mu_{[k]}$ represents the uncertainty or error in the model. The value of Q_u , is a design parameter. Fig. 5.4 compares the estimates of the elevator angle deflection for $Q_u = 10^{-4}$, $Q_u = 10^{-2}$ and $Q_u = 10^6$. As expected, smaller values of Q_u yield more accurate estimates of the deflection in regions where the latter is constant. On the other hand, larger values of Q_u follow the fast changes in the elevator angle deflection more closely. This suggest to adaptively update Q_u according to the rate of change of the bias estimate.

Fig. 5.4 should be compared to Fig. 4.5 in which no model for the bias was used. Comparison of both figures indicates that the bias filter for $Q_u = 10^6$ performs almost identical to the unknown input filter used in Fig. 4.5. In fact, it is found that the estimates of the bias filter converge to those of the joint input-state estimator of Sect. 4.3.4 for $Q_u \rightarrow \infty$. Indeed, $Q_u = \infty$ means that the uncertainty in the bias model is infinite. Therefore, the bias estimate produced during the time update will be completely neglected during the measurement update. This suggests that one should switch to the joint input-state estimator of Sect. 4.3.4 if the bias undergoes an unknown and rapid change.

Finally, we investigate the error made by neglecting the correlation between $\bar{e}_{[k]}$ and the error in $\hat{u}_{[k-1|k-1]}$. To this aim, the estimates of the filter derived above are compared to those of a filter in which the correlation is not neglected. The equations of the latter filter are involved and are not derived here. The difference in the estimates of the angle deflection, averaged over 100 steps, equals 0.00022, showing that the error made by neglecting the correlation is almost negligible. \square

5.4 Model error estimation and model updating

Models induced from physical laws and models identified from data are always approximate. In this section, we consider the case where model accuracy is not satisfactory, so that a correction or update of the model is needed. In physical models, inaccuracies can be due to unmodeled dynamics or incorrect parameter values. In empirical models, inaccuracies can be due to an inappropriate choice of the model class or to bad data quality.

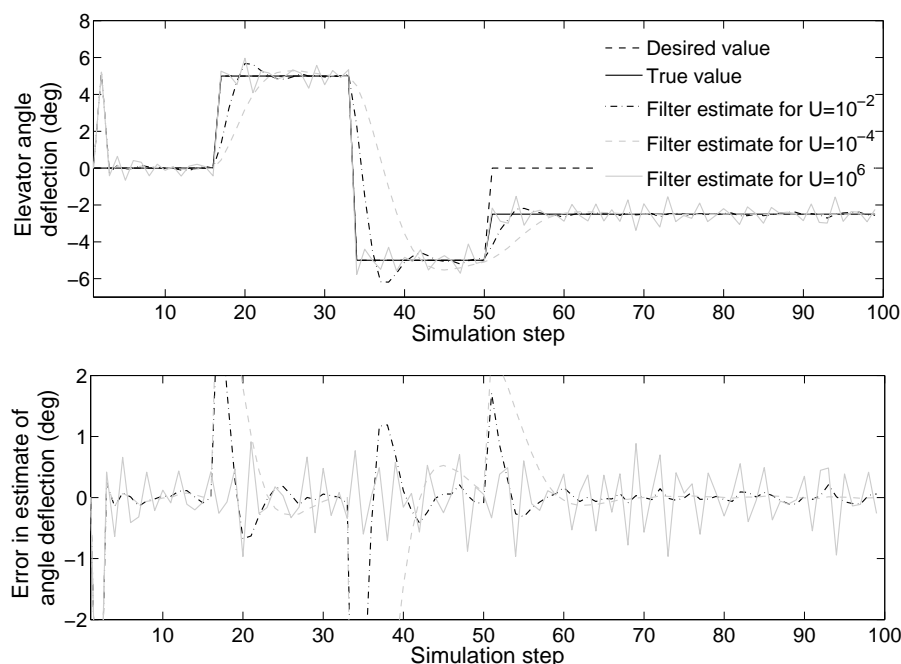


Figure 5.4: Actuator fault detection in a linearized F16 model. Top figure: comparison between desired value, true value and estimated value of the elevator angle deflection. Bottom figure: error in the estimate of the elevator angle deflection.

We consider the case where the known dynamics of the system have been incorporated into a linear state-space model. However, this physical model is assumed to be subject to non-negligible unmodeled dynamics. The objective of this section is to correct or update the physical model using empirical modeling techniques, while the states keep their physical meaning.

Physical models with non-negligible errors are also considered in [110]. A method is outlined for adaptively updating a nonlinear state observer for a system subject to unmodeled dynamics and incorrect parameter values. However, the aim of [110] is not to correct the model, but rather to design the observer such that it compensates model errors. In [79], a nonlinear model representation consisting of an interpolation of several physical models, that are valid within certain operation regimes, is considered. In operational points where the physical models are not satisfactory accurate, they are integrated with empirical models to compensate for unmodeled dynamics. Finally, in [108] linear state space models are updated by adding a correction model, called a *delta* in [108], in parallel, cascade or feedback with the initial model. However, this method yields delta models that are generally of higher order than the initial model.

This section is outlined as follows. In Sect. 5.4.1, we consider the problem of estimating the model error. Next, in Sect. 5.4.2, a technique is developed to correct or update the dynamics of a model that is not satisfactory accurate.

5.4.1 Model error estimation

Consider a set of linear ODE's representing a physical model. Introducing measurements that are linearly dependent on the physical variables, the model can be written in state-space form as

$$\frac{dx}{dt}(t) = A(t)x(t) + B_c(t)u_c(t) + w(t) \quad (5.16a)$$

$$y(t) = C(t)x(t) + v(t), \quad (5.16b)$$

where the $x(t) \in \mathbb{R}^n$ denotes the state vector at time t , $u_c(t) \in \mathbb{R}^{m_c}$ the control input vector at time t and the $y(t) \in \mathbb{R}^p$ the vector of measurements at time t . It is assumed that these vectors have a physical meaning. The vectors $w(t)$ and $v(t)$ denote noise terms. For simulation on a computer, the continuous-time model (5.16) is usually discretized in time, resulting in a LTI discrete-time model of the form

$$x_{[k+1]} = Ax_{[k]} + B_c u_{c[k]} + w_{[k]} \quad (5.17a)$$

$$y_{[k]} = Cx_{[k]} + v_{[k]}, \quad (5.17b)$$

where $x_{[k]} \in \mathbb{R}^n$ denotes the state vector at the discrete time instant k , $u_{c[k]} \in \mathbb{R}^{m_c}$ denotes the input vector at time k , and $y_{[k]} \in \mathbb{R}^p$ denotes the vector of measurements at time k . We assume that the noise processes $\{w_{[k]} \in \mathbb{R}^n\}$ and the measurement noise $\{v_{[k]} \in \mathbb{R}^p\}$ have the properties given in Assumption 2.1. For linear ODE's, a substantial amount of discretization methods are available. While some methods preserve the physical meaning of the state, other methods lack this property. We consider discretization methods that preserve the physical meaning of the state, such that $x_{[k]} \simeq x(kT_s)$, with T_s the sampling time.

For physical models, the measurements are usually direct observations of a state variable or well-known linear combinations of only a few state variables. Consequently, the output equation (5.17b) is assumed to be very accurate. On the other hand, we assume that the state equation (5.17a) is subject to incorrect parameters and/or unmodeled dynamics. To compensate for these model errors, we add a correction term $u_{[k]} \in \mathbb{R}^m$ to (5.17a), resulting in the LTI discrete-time model

$$x_{[k+1]} = Ax_{[k]} + B_c u_{c[k]} + Bu_{[k]} + w_{[k]} \quad (5.18a)$$

$$y_{[k]} = Cx_{[k]} + v_{[k]}, \quad (5.18b)$$

where the matrix B is assumed to be known, or chosen appropriately. We will refer to $u_{[k]}$ as the *model error vector*.

The objective of this section is the estimate the model error vector from knowledge of the measurements up to time instant k . Note that $u_{[k]}$ enters

(5.18) like an unknown input. Consequently, the joint input-state estimators of Sect. 4.3 can be employed.

Example 5.3. Tape drive modeling example

Consider errors in a tape drive model. The tape drive comprises one tape, two drive wheels and two DC-motors. The DC-motors drive the wheels and are independently controllable by voltage sources V_1 and V_2 . The armature circuit of DC-motor j ($j = 1, 2$) is expressed as

$$L_{a,j} \frac{dI_j(t)}{dt} + R_{a,j} I_j(t) + K_{e,j} \omega_j(t) = V_j(t), \quad (5.19)$$

where L_a is the armature inductance, R_a is the armature resistance, K_e is the electrical constant of the motor, I is the current and ω is the rotational speed of the drive wheel. The position p of the drive wheels is related to ω by the radius r ,

$$\frac{dp_j(t)}{dt} = r_j \omega_j(t). \quad (5.20)$$

The equation for the rotational speed of the drive wheels is given by

$$J_j \frac{d\omega_j(t)}{dt} = -T(t)r_j - \beta\omega_j(t) + K_{t_j} I_j(t), \quad (5.21)$$

where J is the inertia of the drive wheel and motor, β is the rotational friction of the drive wheel and motor, K_t is the torque constant of the motor and T is the tape tension. This tension is given by

$$T(t) = \frac{K}{2} \Delta p(t) + \frac{D}{2} \left(\frac{d\Delta p(t)}{dt} \right), \quad (5.22)$$

where K is the spring constant of the tape, D is the damping in the tape-stretch motion and $\Delta p(t) = p_2(t) - p_1(t)$. The parameter values used in the example, are given in Table 5.1.

Now, assume that the dynamics describing the tension in the tape are not known and thus omitted in the modeling procedure. Hence, the physical model is given by (5.19)-(5.20) and

$$J_j \frac{d\omega_j(t)}{dt} = -\beta\omega_j(t) + K_{t_j} I_j(t). \quad (5.23)$$

A continuous-time state space model is obtained by defining the state vector $x(t) := [I_1(t) \ I_2(t) \ p_1(t) \ p_2(t) \ \omega_1(t) \ \omega_2(t)]^T$ and the control input vector $u_c(t) := [V_1(t) \ V_2(t)]^T$. It is assumed that measurements of all state variables are available. The continuous-time model is discretized in time using the zero order hold method with sampling time $T_s = 10^{-4}s$. To account for the unknown tension, we add a term $Bu_{[k]}$ to the state equation of the discretized model, where

$$B = \left[\begin{array}{cccccc} 0 & 0 & 0 & 0 & -\frac{r_1}{J_1} & -\frac{r_2}{J_2} \end{array} \right]^T$$

Table 5.1: Parameter values used in the tape drive example

Parameter	Value	Unit
L_a	10^{-3}	H
R_a	1	Ω
K_e	$3 \cdot 10^{-2}$	V.s
r	$25 \cdot 10^{-3}$	m
J	$5 \cdot 10^{-5}$	kg.m ²
β	$5 \cdot 10^{-2}$	kg.m ² .s ⁻¹
K_t	$3 \cdot 10^{-2}$	N.m.A ⁻¹
K	$2 \cdot 10^4$	N.m ⁻¹
D	10	N.m ⁻¹ .s ⁻¹

is chosen so that $u_{[k]} = T(kT_s)$. The noise processes $\{w_{[k]}\}_{k=0}^{\infty}$ and $\{v_{[k]}\}_{k=0}^{\infty}$ are assumed to have the properties given in Assumption 2.1. Their covariance matrices are assumed to be given by $Q = R = \text{diag}(10^{-6}, 10^{-6}, 10^{-8}, 10^{-8}, 10^{-4}, 10^{-4})$.

We now apply the joint input state estimator summarized in Sect. 4.3.4 to simultaneously estimate the system state and the tape tension. The “measurements” used in the simulation were obtained by simulating a discrete-time model in which the dynamics describing the tension in the tape are not omitted. The true value (i.e. the value obtained in the simulation just described) and the estimated value of the tape tension are shown in Fig. 5.5 together with 95% confidence intervals. The confidence intervals are calculated from the error covariance matrix $P_{u_{[k]}}$. The *procentual estimation error* (EE), defined by

$$\text{EE} := \frac{100}{p} \sum_{i=1}^p \left[\sqrt{\frac{\sum_{k=1}^N ((y_{[k]})_i - (\hat{y}_{[k|k]})_i)^2}{\sum_{k=1}^N ((y_{[k]})_i)^2}} \right] \%, \quad (5.24)$$

with $\hat{y}_{[k|k]} = C\hat{x}_{[k|k]}$, equals 0.2. For comparison, the EE of the Kalman filter equals 29.5. \square

5.4.2 Subsystem identification and model updating

So far, we have considered the problem of estimating model errors. The objective of this section is to correct or update the dynamics of a model that is not satisfactory accurate. We consider the case where the model error is arising from an unknown linear *subsystem*.

5.4.2.1 Problem formulation

We make the assumption that the true system can be written as (5.18) with $u_{[k]}$ arising from an unknown LTI subsystem driven by $u_{c[k]}$ and $x_{[k]}$,

$$z_{[k+1]} = A_d z_{[k]} + B_{d_u} u_{c[k]} + B_{d_x} x_{[k]} + \omega_{[k]} \quad (5.25a)$$

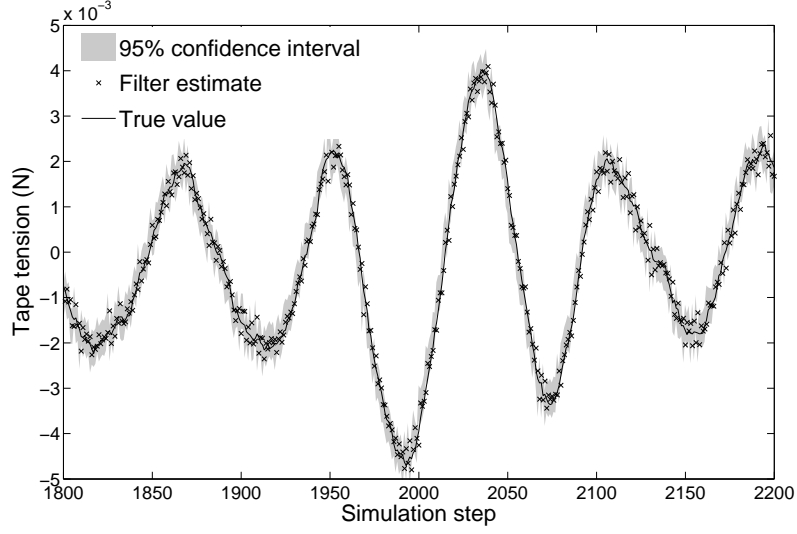


Figure 5.5: *Tape drive modeling example. Comparison between true and estimated value of unknown tape tension.*

$$u_{[k]} = C_d z_{[k]} + \nu_{[k]}, \quad (5.25b)$$

where $z_{[k]} \in \mathbb{R}^{n_d}$ denotes the state vector and where the noise processes $\{\omega_{[k]} \in \mathbb{R}^{n_d}\}_{k=0}^{\infty}$ and $\{\nu_{[k]} \in \mathbb{R}^{m_d}\}_{k=0}^{\infty}$ are assumed to have the properties given in Assumption 2.1. We define $Q_\omega := \mathbb{E}[\omega_{[k]}\omega_{[k]}^\top]$ and $R_\nu := \mathbb{E}[\nu_{[k]}\nu_{[k]}^\top]$. The objective of the next sections is to update the initial model (5.17) in case of unmodeled dynamics of the form (5.25).

5.4.2.2 State of the art

The *parallel delta-augmentation* method of [108] updates an inaccurate initial model by identifying a *delta-model* with input $u_{c[k]}$ and with output an additive correction term for the output of the initial model. Let the initial model be given by (5.17) and let the identified delta-model be given by

$$z_{[k+1]} = \hat{A}_\Delta z_{[k]} + \hat{B}_\Delta u_{c[k]} + \varpi_{[k]} \quad (5.26a)$$

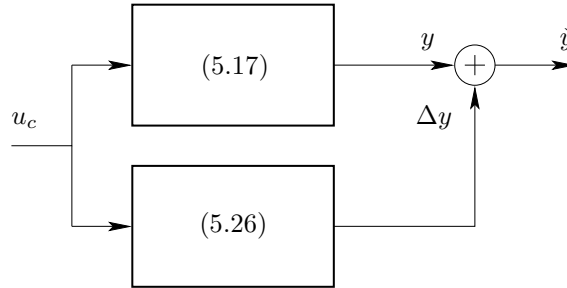
$$\Delta y_{[k]} = \hat{C}_\Delta z_{[k]} + v_{[k]}, \quad (5.26b)$$

then the updated model with corrected output $\check{y}_{[k]}$ is given by

$$\begin{bmatrix} x_{[k+1]} \\ z_{[k+1]} \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & \hat{A}_\Delta \end{bmatrix} \begin{bmatrix} x_{[k]} \\ z_{[k]} \end{bmatrix} + \begin{bmatrix} B_c \\ \hat{B}_\Delta \end{bmatrix} u_{c[k]} + \begin{bmatrix} w_{[k]} \\ \varpi_{[k]} \end{bmatrix} \quad (5.27a)$$

$$\check{y}_{[k]} = \begin{bmatrix} C & \hat{C}_\Delta \end{bmatrix} \begin{bmatrix} x_{[k]} \\ z_{[k]} \end{bmatrix} + v_{[k]} + v_{[k]}. \quad (5.27b)$$

a) Parallel delta-augmentation



b) Dynamic model updating

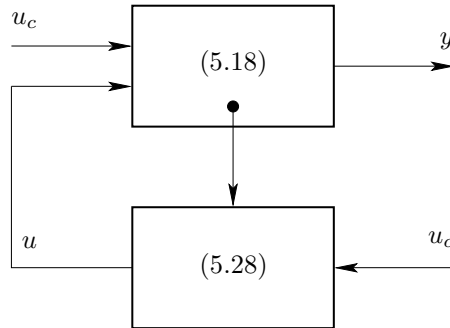


Figure 5.6: Comparison between the parallel delta augmentation method of Sect. 5.4.2.2 (a) and the dynamic model updating technique of Sect. 5.4.2.3 (b). Note that dynamic model updating introduces feedback.

A schematic of parallel-delta augmentation is shown in Fig. 5.6a.

In case the unknown subsystem takes the form (5.25), it is easy to show that the best possible delta-model has order $n + n_d$. In cases where n is large, while n_d is relatively small, this may yield a delta-model that is of much higher order than the unknown subsystem (5.25). Furthermore, this method does only correct the output equation, but not the erroneous state equations of the initial model. This is a major disadvantage if the states have a physical meaning that is of importance for e.g. control.

5.4.2.3 Dynamic model updating

To overcome the problems encountered with the method of [108], we consider the problem of identifying a correction model of the form (5.25). The updated model then has the structure shown in Fig. 5.6b.

Notice that the optimal filter summarized in Sect. 4.3.4 yields estimates

$\hat{x}_{[k|k]}$ of the inputs $x_{[k]}$ and estimates $\hat{u}_{[k|k]}$ of the outputs $u_{[k]}$ of the unknown subsystem (5.25). More precisely, after running the filter from time instant $k = 0$ to $k = N + 1$, the following two data-sets are obtained,

$$\begin{aligned}\mathcal{X}_N &= \{\hat{x}_{[j|j]}\}_{j=0}^N, \\ \mathcal{U}_N &= \{\hat{u}_{[j]}\}_{j=0}^N.\end{aligned}$$

Hence, noisy datasets of the inputs and outputs of the unknown subsystem (5.25) are available. A LTI correction model approximating the dynamics of the subsystem can be identified from these data-sets by a combined deterministic-stochastic *subspace identification* algorithm [106]. Subspace identification is an empirical identification technique that yields a linear state space model from a set of input-output data only. The major advantage of subspace identification algorithms over the classical prediction error methods is the absence of non-linear parametric optimization problems. Subspace identification algorithms are non-iterative, and thus never get stuck in local minima or never suffer from convergence problems. They always produce a result, which is often surprisingly good for practical data. The algorithm is based on geometric and algebraic operations like projections and singular value decompositions, for which efficient and stable numerical implementations have been developed. The algorithm returns the identified system matrices and the covariance matrices of the noise. Hence, the identified *correction model* takes the form

$$z_{[k+1]} = \hat{A}_d z_{[k]} + \hat{B}_{d_u} u_{[k]} + \hat{B}_{d_x} \hat{x}_{[k]} + \hat{\omega}_{[k]} \quad (5.28a)$$

$$\hat{u}_{[k|k]} = \hat{C}_d z_{[k]} + \hat{\nu}_{[k]}, \quad (5.28b)$$

where $z_{[k]} \in \mathbb{R}^{\hat{n}_d}$ denotes the state vector and where the noise processes $\{\omega_{[k]} \in \mathbb{R}^{\hat{n}_d}\}_{k=0}^{\infty}$ and $\{\nu_{[k]} \in \mathbb{R}^m\}_{k=0}^{\infty}$ have the properties given in Assumption 2.1 and have covariance matrices $\hat{Q}_{\hat{\omega}}$ and $\hat{R}_{\hat{\nu}}$, respectively. In practical applications where a lot of data is available, it is to be expected that \hat{n}_d is close to n_d . In cases where the true system is high order, while the error is relatively low order, this may yield a considerable storage and computational saving over the parallel delta-augmentation method.

The initial model (5.17) is then augmented with the identified correction model (5.28), resulting in the model

$$\begin{bmatrix} x_{[k+1]} \\ z_{[k+1]} \end{bmatrix} = \begin{bmatrix} A & B\hat{C}_d \\ \hat{B}_{d_x} & \hat{A}_d \end{bmatrix} \begin{bmatrix} x_{[k]} \\ z_{[k]} \end{bmatrix} + \begin{bmatrix} B_c \\ \hat{B}_{d_u} \end{bmatrix} u_{c[k]} + \begin{bmatrix} w_{[k]} + B\hat{\nu}_{[k]} \\ \hat{\omega}_{[k]} \end{bmatrix} \quad (5.29a)$$

$$\begin{bmatrix} y_{[k]} \\ \hat{u}_{[k|k]} \end{bmatrix} = \begin{bmatrix} C & 0 \\ 0 & \hat{C}_d \end{bmatrix} \begin{bmatrix} x_{[k]} \\ z_{[k]} \end{bmatrix} + \begin{bmatrix} v_{[k]} \\ \hat{\nu}_{[k]} \end{bmatrix}, \quad (5.29b)$$

of order $n + \hat{n}_d$. In contrast to the parallel delta-augmentation method, this kind of model updating directly corrects errors in the state equation of the initial model. The difference between both methods is also noticeable in the

interconnection between the dynamics of the initial model and the correction model. This interconnection can be seen in the “ A ” matrix of the augmented model (5.29a), which is dense, in contrast to (5.27a), where it is block-diagonal.

Example 5.4. Tape drive modeling example

Consider again the tape drive example. Using the data sets \mathcal{U}_N and \mathcal{X}_N , we identify an unknown subsystem of the form (5.25) with the N4SID subspace identification algorithm [105]. The order of the identified model is determined by the N4SID algorithm and equals 1. The nominal model is then augmented with the identified correction model, resulting in an updated model of order 7. For comparison purpose, we also identified a correction model using the parallel delta-augmentation method. The delta model has order 6, resulting in an updated model of order 12. This confirms the theoretical result that the parallel delta-augmentation method yields updated models which are generally of higher order than the dynamic model updating technique.

The updated models are validated in two different ways. Firstly, the initial model and the updated models are simulated using validation inputs and the outputs are compared to the measurements of the true system. Table 5.2 compares the simulation error (SE) and the one step ahead prediction error (PE) for all models. The SE is defined by (5.24) with $\hat{y}_{[k|k]}$ replaced by the simulated model output. The PE is defined by (5.24) with $\hat{y}_{[k|k]}$ replaced by $\hat{y}_{[k|k-1]} = C\hat{x}_{[k|k-1]}$, where $\hat{x}_{[k|k-1]}$ is obtained with a Kalman filter. Secondly, the dynamically updated model is validated by computing the autocorrelation of the one step ahead prediction residuals $y_{[k]} - C\hat{x}_{[k|k-1]}$. Optimally, the residuals are uncorrelated in time. Fig. 5.7 shows the autocorrelation (with 99% confidence intervals) for the initial model (top) and the updated model (bottom). For the initial model, the correlation between the current and future residuals falls out the confidence region, indicating that the residuals are strongly correlated. For the updated model, correlation is much smaller. \square

5.5 Boundary condition estimation

The estimation of unknown boundary conditions has been intensively studied in inverse heat conduction problems. In [19,138] it is assumed that the initial state and the functional form in space and time of the boundary condition are known.

Table 5.2: Comparison between simulation error (SE) and prediction (PE) of initial and updated models.

Model	Order	SE (%)	PE (%)
Initial model	6	38.6	26.4
Dynamic updating	7	5.2	0.3
Parallel updating	12	6.7	4.5

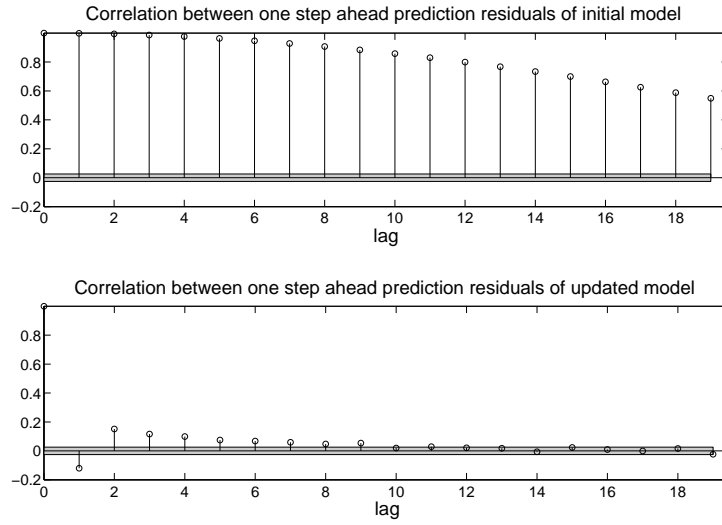


Figure 5.7: *Tape drive modeling example; autocorrelation of the prediction residuals $y_{[k]} - C\hat{x}_{[k|k-1]}$ (with 99% confidence intervals, indicated by the grey rectangles) for the initial model (top) and the updated model (bottom).*

The unknown parameters in the functional form are then estimated using LS estimation. An extension to simultaneous boundary condition and initial state estimation, can be found in [74]. Approaches using an augmented state Kalman filter are developed in [101, 121]. The applicability of all of these methods is, however, limited by the assumption that the functional form of the boundary condition in space and time is known.

In this section, we address the problem of estimating time and space varying boundary conditions. In contrast to existing techniques, we make no assumption about the functional form in time. However, it is assumed the functional form in space can be written as a linear combination of basis functions.

This section is outlined as follows. The problem is formulated in more detail in Sect. 5.5.1. In Sect. 5.5.2, the expansion of the boundary condition as a linear combination of basis functions is addressed. Finally, in Sect. 5.5.3, a numerical example is considered.

5.5.1 Problem formulation

Consider a set of linear PDE's with (partially) unknown boundary conditions. After discretization in space and time, this yields an LTI discrete-time model that can be written into the state-space form

$$x_{[k+1]} = Ax_{[k]} + Bu_{[k]} + w_{[k]} \quad (5.30a)$$

$$y_{[k]} = Cx_{[k]} + v_{[k]}, \quad (5.30b)$$

where $x_{[k]} \in \mathbb{R}^n$ denotes the state vector at time instant k , $y_{[k]} \in \mathbb{R}^p$ denotes the measurement at time k , and $u_{[k]} \in \mathbb{R}^m$ denotes the influence of the unknown boundary conditions at time k . The initial state $x_{[0]}$ is assumed to be a random variable. The noise processes $\{w_{[k]}\}_{k=0}^{\infty}$ and $\{v_{[k]}\}_{k=0}^{\infty}$ are assumed to have the properties given in Assumption 2.1. We define $Q := \mathbb{E}[w_{[k]}w_{[k]}^T]$ and $R := \mathbb{E}[v_{[k]}v_{[k]}^T]$. The problem considered in the next sections is that of simultaneously estimating the system state $x_{[k]}$ and the unknown boundary condition $u_{[k]}$.

5.5.2 Basis function expansion

It follows from the theory in Chapters 3 and 4 that a necessary condition for reconstructing the unknown boundary condition using system inversion techniques is that $p \geq m$. In many practical applications, however, this condition is too restrictive. As will now be shown, the condition can be relaxed by expanding the unknown boundary condition as a linear combination of basis functions. More precisely, we assume that $u_{[k]}$ can be written as a linear combination of N , with $N \ll m$, prescribed basis vectors $\phi_i \in \mathbb{R}^m$, $i = 1 \dots N$. That is,

$$u_{[k]} = \sum_{i=1}^N a_{i[k]} \phi_i. \quad (5.31)$$

Defining the vector of coefficients $a_{[k]} \in \mathbb{R}^N$ by $a_{[k]} := [a_{1[k]} \ a_{2[k]} \ \dots \ a_{N[k]}]^T$, and defining the matrix $\Phi := [\phi_1 \ \phi_2 \ \dots \ \phi_N]$, (5.31) is rewritten as $u_{[k]} = \Phi a_{[k]}$. By substituting the latter equation in (5.30a), the problem of estimating the unknown boundary condition $u_{[k]}$ is transformed to that of estimating the vector of coefficients $a_{[k]}$. A necessary condition for reconstruction of $a_{[k]}$ is that $p \geq N$. Since $N \ll m$, this condition is less strong than $p \geq m$.

5.5.3 Heat conduction example

As shown in Figure 5.8, we consider heat conduction in a plate with dimensions $L_x \times L_y$. The plate is heated from below by a flame. At three boundaries, the temperature is fixed at 300 K. The temperature of the fourth boundary is to be estimated.

Heat conduction in the two-dimensional plate is governed by the PDE

$$\frac{\partial T}{\partial t} = \alpha \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right) + u(x, y, t), \quad (5.32)$$

where $T(x, y, t)$ denotes the temperature at position (x, y) and time instant t , $u(x, y, t)$ denotes the influence of an external heat source and α denotes the thermal diffusivity, which is material dependent. The dimension of the plate is $L_x = 1\text{m}$ by $L_y = 2\text{m}$, the thermal diffusivity is $\alpha = 10^{-4} \text{m}^2/\text{s}$ and the external heat input is assumed to be given by

$$u(x, y, t) = \frac{1}{2} e^{-\left(\frac{(x-L_x/2)^2}{2\sigma^2} + \frac{(y-L_y/2)^2}{2\sigma^2} \right)}$$

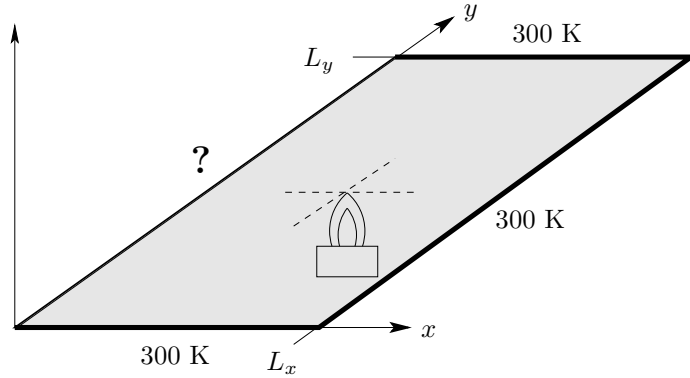


Figure 5.8: Setup of the heat conduction example. A plate with dimensions $L_x \times L_y$ is considered. The plate is heated from below by a flame. At three boundaries, the temperature is fixed at 300 K. The temperature of the fourth boundary is to be estimated.

with $\sigma = 10^{-1}$, which represents the influence of a flame centered under the middle of the plate. The boundary condition at $x = 0$ is unknown. The other boundary conditions are given by

$$T(L_x, y, t) = T(x, 0, t) = T(x, L_y, t) = 300\text{K}.$$

The initial condition is given by $T(x, y, 0) = 300\text{K}$.

The PDE (5.32) is discretized in space and time using finite differences with $\Delta x = \Delta y = 0.1\text{m}$ and $\Delta t = 2\text{s}$, resulting a linear discrete-time state space model of order $n = 200$. Process noise with variance 10^{-6} is introduced. The matrix B is chosen so that $u_{[k]} \in \mathbb{R}^{20}$ represents the unknown boundary condition at $x = 0$. It is assumed that $p = 14$ measurements are available. The covariance matrix of the measurement noise is $R = 10^{-3}I_p$.

In a first experiment, we set up a simple problem in order to test the performance of the approach. We use the method of twin-experiments. First, we simulate the discretized model and add process noise and measurement noise. The boundary condition at $x = 0$ is chosen as a linear combination of the first 4 Chebyshev polynomials. Next, we use the joint input-state estimators of Chapter 4 to simultaneously estimate the system state and the boundary condition at $x = 0$. By expanding the boundary condition as a linear combination of the first 4 Chebyshev polynomials, the problem boils down to the joint estimation of the system state and the coefficients in the expansion. Note that in order to apply the filter of Sect. 4.3, we must have $\text{rank}(CB\Phi) = N = 4$. This condition implies that values of at least 4 boundary states should be incorporated into the measurements. We consider the measurement locations indicated by the stars in Fig. 5.9(a). The latter figure shows the estimation error after 250 steps. The estimation error is clearly largest at the boundary

$x = 0$. In Fig. 5.10, the estimation error of the joint input-state estimator is compared to that of a Kalman filter (where the boundary condition is assumed to be known). Note that both estimators have approximately the same speed of convergence.

The condition that values of the boundary states should be incorporated into the measurements can be relaxed by considering the smoother of Sect. 4.4. We now consider the measurement locations indicated by the stars in Fig. 5.9(b). It turns out that the corresponding system is 3–delay left invertible. The estimation error of a smoother with $L = 3$ is shown in Fig. 5.9(b). Note that the estimation error is largest at the boundary $x = 0$. Also, note that the estimation error is larger than in part (a) of the figure.

5.6 Conclusion

Four applications of system inversion were considered.

The first application extends the Kalman filtering problem to the case where the system input is unknown, but a noisy linear combination of the inputs is available. Based on the joint input-state estimator developed in Sect. 4.2, filter equations were developed in which the estimation of the system state and the unknown input are interconnected. As a special case, the filter provides a new solution to the errors-in-variables filtering problem, which is shown to be algebraically equivalent to existing techniques [31, 93].

The second application has considered the optimal filtering problem for systems subject to bias errors. By incorporating the bias model into the joint input-state estimator of Sect. 4.3, a suboptimal estimator was developed in which the bias is estimated with one step delay. It was shown in a simulation example that such an approach is especially useful if the bias error is constant for a certain period of time and then suddenly undergoes an abrupt and unknown change.

The third application has dealt with model error estimation and dynamic model updating. An empirical technique was outlined to update a non-satisfactory accurate physical state space model. The technique consists in first estimating the model error and then identifying an empirical correction model based on the estimated data. It was shown in a numerical example that this procedure yields updated models which are more usually of much lower order than existing techniques.

The last application has addressed the problem of joint state and boundary condition estimation. In contrast to existing methods, boundary conditions that vary in space and time were considered and no assumption was made about the time evolution. Concerning the variation in space, it was assumed that boundary condition can be expanded as a linear combination of a few basis functions. A simulation example has shown that the estimator converges as fast as a Kalman filter in which the boundary conditions are assumed to be known. The major drawback of the method is that a lot of measurements close to the boundary are needed.

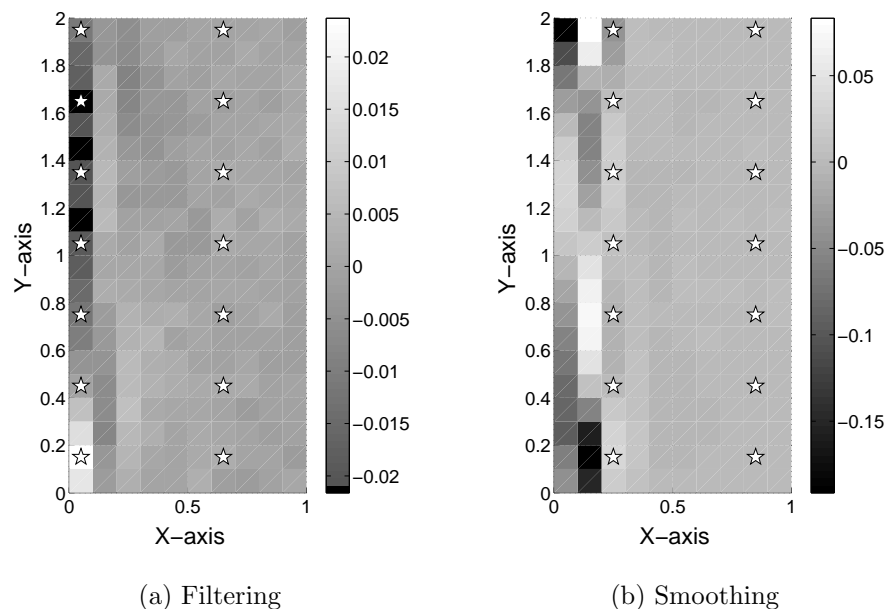


Figure 5.9: Heat conduction example: estimation error after 250 simulation steps. The stars denote the locations where measurements are taken. (a) Results for the filter of Sect. 4.3. (b) Results for the smoother of Sect. 4.4.

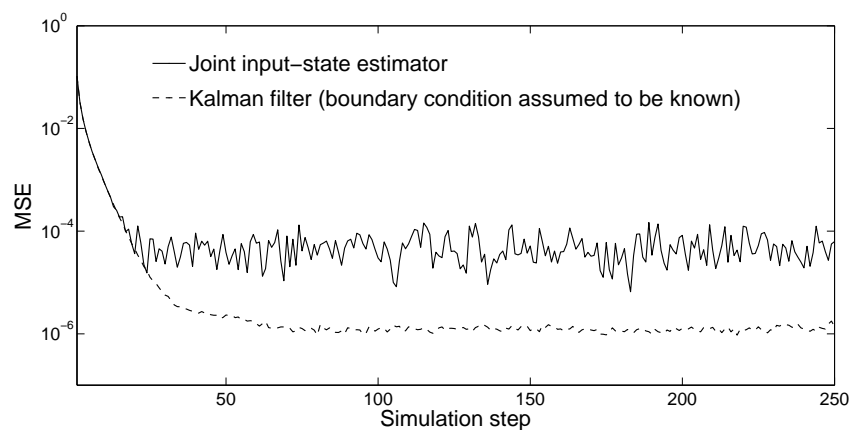


Figure 5.10: Heat conduction example: comparison between the MSE of the Kalman filter (where the boundary condition is assumed to be known) and the joint input-state estimator.

Part II

Data Assimilation

Chapter 6

Suboptimal Square-Root Filtering

This chapter addresses the challenging problem of data assimilation, which is concerned with assimilating observations into large-scale numerical models. After a brief overview of the most commonly used suboptimal Kalman filtering techniques, two extensions of the reduced rank square-root filter [135] are developed. The first extension speeds-up the RRSQRT filter by interweaving the so-called reduction step into the measurement update. The second extension addresses the problem of reduced rank spatially localized square-root filtering. The resulting algorithm is extremely efficient if only few measurements are available.

6.1 Introduction

Although the Kalman filter may seem very appealing for data assimilation because of its simple recursive structure, it is not directly applicable. The application of the Kalman filter is hampered by its high computational cost and its immense storage requirements needed to propagate the error covariance matrix.

The update of the error covariance matrix in the Kalman filter requires $\mathcal{O}(n^3)$ flops, where n is the dimension of the state vector. The number of memory elements needed to store the covariance matrix depends on n as $\mathcal{O}(n^2)$. As a result, the Kalman filter is feasible on today's computers until $n = 10^4$.

The numerical models used in data assimilation, however, usually have a much higher state dimension. Numerical models in data assimilation are mostly based on physical laws, usually PDE's. For accurate simulation on a computer, these PDE's are discretized over a huge spatial grid. The number of cells in the grid determines the dimension n of the state vector. This dimension is usually chosen so that a single simulation can be performed in a reasonable amount

of time. In contrast to Kalman filtering, simulating typically takes only $\mathcal{O}(n)$ flops. Consequently, the state dimension ranges from $n = 10^4$ in tidal flow forecasting [135] to as much as $n = 10^7$ in space weather forecasting [61].

State estimation for such high dimensional systems thus requires approximations of the Kalman filter algorithm. In the remainder of this section, we give a short historical overview of suboptimal filtering techniques for data assimilation.

The earliest systematic approaches to data assimilation were called *objective analyses*. The objective methods used simple interpolation techniques. Before, scientists used *subjective analyses*, i.e. their expertise, to assimilate observations in numerical model predictions. A simplified form of the Kalman filter, called *optimal interpolation*, made its introduction in data assimilation already in 1963 [46]. Since then, researchers have been experimenting with all kinds of approximations of the Kalman filter algorithm. We refer to such approximate filters as *suboptimal Kalman filters*. A substantial amount of suboptimal Kalman filters have been proposed in literature.

Variational data assimilation [23, 88] is an approach that is based on the LS interpretation of the Kalman filter. In this approach, the squared difference between the real observations and their simulated counterparts is minimized. This requires the solution of a huge optimization problem. The development of efficient iterative solvers have made this approach the standard at the European Center for Medium-Range Weather Forecasts. Despite its success, variational assimilation is based on the assumption that the numerical model is perfect.

Boggs et al. [14] proposed a suboptimal approximation of the Kalman filter equations based on a banded approximation of the error covariance matrix. This approximation is motivated by the fact that correlations in environmental applications have only a limited spatial range. However, creating artificial zeros in the error covariance can lead to the occurrence of negative eigenvalues. This problem is tackled in [14] by adopting a square-root decomposition of the error covariance matrix.

A disadvantage of a banded approximation of the error covariance matrix is that it may sometimes discard large eigenvalues. It is well known that this may lead to filter divergence. To prevent divergence as much as possible, one can make an optimal lower rank approximation of the error covariance matrix. This idea is worked out in the *partial eigendecomposition Kalman filter* [22] where, as the name suggests, the lower rank approximation is based on an eigenvalue decomposition. Efficiency is increased by storing and propagating the error covariance matrix relative to the space spanned by its leading eigenvectors.

Verlaan and Heemink [134] extended the approach of [22] by expressing the equations in square-root form. Their suboptimal filter, which is called the *reduced rank square-root (RRSQRT) filter*, thus combines the efficiency of the partial eigendecomposition filter with the numerical advantages of square-root filtering. In addition, the RRSQRT filter is algebraically equivalent to the Kalman filter if the rank of the error covariance matrix is chosen equal to the state dimension. For nonlinear systems, efficient extensions based on the EKF have been developed.

The *ensemble Kalman filter* (EnKF) [13, 38, 39, 71, 85] approximates the

error covariance matrix based on a Monte Carlo approach. The starting point of the EnKF is an ensemble of state estimates that tries to capture the probability density function of the initial state. The ensemble of estimates is then propagated through the nonlinear model and the error covariance matrix of the actual state is approximated from the ensemble of estimates. Although one would expect that an enormous amount of ensemble members is needed, literature suggests that only a few hundreds are sufficient, even for a model with 10^6 grid cells. The major advantage of the EnKF over other techniques is that it is relatively simple to implement and well suited for highly nonlinear models. The EnKF, however, introduces sampling errors due to the low number of ensemble members.

Personal contributions

The main contribution of this chapter is the development of two extensions of the RRSQRT filter.

- The first extension, considered in Sect. 6.4.2, speeds-up the RRSQRT filter by eliminating the so-called *reduction step*. However, the resulting filter is more approximate than the RRSQRT filter in the sense that it underestimates the error covariance matrix more.
- The second extension, considered in Sect. 6.5.2, combines ideas from the RRSQRT filter with those of spatially localized filtering [9]. Two variants of the extension are developed. The first variant is equivalent to the spatially localized Kalman filter if no lower rank approximation of the error covariance matrix is made. The second variant is based on the assumption that correlation between grid cells drops to zero within a distance of a relatively low number of grid cells. Although being more approximate, this variant turns out to be extremely efficient, especially if only few measurements are available.

Chapter outline

This chapter is outlined as follows. Section 6.2 introduces the basic ideas behind suboptimal square-root filtering. Section 6.3 briefly summarizes some existing square-root measurement updates for large-scale systems. In Sect. 6.4, the RRSQRT filter is discussed and a more efficient variation, the reduced rank *transform* square-root filter is introduced. Next, in Sect. 6.5, we address the problem of spatially localized filtering. We derive a reduced rank spatially localized filter that is extremely efficient if only few measurements are available. In Sect. 6.6, filter degradation due to a lower rank approximation of the error covariance matrix is addressed. Finally, in Sect. 6.7, two numerical examples are considered.

6.2 Suboptimal square-root filtering: the idea

As already discussed, the idea of square-root filtering has been introduced to avoid numerical problems in a direct implementation of the Kalman filter equations. As will now be discussed, most suboptimal filters also employ a square-root formulation, but mainly for a different reason, namely to increase computational efficiency.

Consider an error covariance matrix $P_{[k|k-1]} \in \mathbb{R}^{n \times n}$ of rank q with $q \ll n$. Then, $P_{[k|k-1]}$ can be decomposed as

$$P_{[k|k-1]} = S_{[k|k-1]} S_{[k|k-1]}^T,$$

where $S_{[k|k-1]}$ is $n \times q$. Consequently, the memory consuming matrix $P_{[k|k-1]}$ can be constructed from its much smaller square-root $S_{[k|k-1]}$. The idea behind suboptimal square-root filtering is to propagate $S_{[k|k-1]}$ instead of $P_{[k|k-1]}$ and to ensure that $S_{[k|k-1]}$ has rank q for all k . As we will see, the latter requires some approximations to be made, such as e.g. an optimal lower rank approximation.

In the remainder of this chapter, the symbols $P_{[k|k-1]}$ and $S_{[k|k-1]}$ will denote approximations to the true error covariance matrix and its square-root, respectively.

Notice that the conventional square-root filters would decompose $P_{[k|k-1]}$ as

$$P_{[k|k-1]} = P_{[k|k-1]}^{1/2} P_{[k|k-1]}^{\top/2},$$

where $P_{[k|k-1]}^{1/2}$ denotes an $n \times n$ Cholesky factor of $P_{[k|k-1]}$. Consequently, conventional square-root filters do not decrease computational load or storage requirements, in contrast. Figure 6.1 compares suboptimal square-root filtering to the conventional approach. In practical applications, q is chosen in the order of 10^2 while n can be in the order of 10^6 or 10^7 . Suboptimal square-root filtering then yields a huge decrease in computation times and storage requirements over the conventional Kalman filter.

We consider in this chapter linear time-varying systems of the form

$$x_{[k+1]} = A_{[k]}x_{[k]} + B_{[k]}u_{[k]} + E_{[k]}w_{[k]}, \quad (6.1a)$$

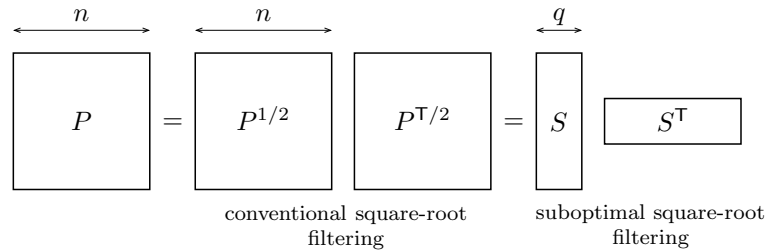


Figure 6.1: *Conventional versus suboptimal square-root filtering. It is assumed that P has rank q with $q \ll n$.*

$$y_{[k]} = C_{[k]}x_{[k]} + v_{[k]}, \quad (6.1b)$$

where $x_{[k]} \in \mathbb{R}^n$ denotes the state vector at time instant k , $u_{[k]} \in \mathbb{R}^m$ denotes a known deterministic input at time k and $y_{[k]} \in \mathbb{R}^p$ denotes the measurement at time k . The initial state $x_{[0]}$ is assumed to be a random variable. The noise processes $\{w_{[k]} \in \mathbb{R}^l\}_{k=0}^{\infty}$ and $\{v_{[k]} \in \mathbb{R}^p\}_{k=0}^{\infty}$ are assumed to be zero-mean with

$$\mathbb{E} \left\{ \begin{bmatrix} w_{[k]} \\ v_{[k]} \end{bmatrix} \begin{bmatrix} w_{[j]}^\top & v_{[j]}^\top \end{bmatrix} \right\} = \begin{bmatrix} Q_{[k]} & 0 \\ 0 & R_{[k]} \end{bmatrix} \delta_{[k-j]},$$

where $\delta_{[k]} := 1$ for $k = 0$ and $\delta_{[k]} := 0$ otherwise.

The equations of suboptimal square-root filters are usually split into a time-update and a measurement update. The measurement update is addressed in the next section.

6.3 Square-root measurement updating

Consider the measurement update of $P_{[k|k-1]}$ in the Kalman filter,

$$P_{[k|k]} = P_{[k|k-1]} - P_{[k|k-1]}C_{[k]}^\top (C_{[k]}P_{[k|k-1]}C_{[k]}^\top + R_{[k]})^{-1}C_{[k]}P_{[k|k-1]}, \quad (6.2)$$

and assume that $P_{[k|k-1]}$ has rank q . Then, Potter [111] showed that (6.2) can be written in terms of $S_{[k|k-1]}$ as

$$P_{[k|k]} = S_{[k|k-1]} \left[I - V_{[k]}(V_{[k]}^\top V_{[k]} + R_{[k]})^{-1}V_{[k]}^\top \right] S_{[k|k-1]}^\top, \quad (6.3)$$

where the $q \times p$ matrix $V_{[k]}$ is defined as $V_{[k]} := (C_{[k]}S_{[k|k-1]})^\top$.

It turns out that the measurement update (6.2) does not increase the rank of the error covariance matrix. Consequently, $P_{[k|k]}$ has rank q and may be decomposed as $P_{[k|k]} = S_{[k|k]}S_{[k|k]}^\top$ with $S_{[k|k]}$ an $n \times q$ matrix.

As shown by Potter, the measurement update can be rewritten in a form that computes $S_{[k|k]}$ directly in terms of $S_{[k|k-1]}$. For convenience of notation, we define the square matrix $T_{[k]} \in \mathbb{R}^{q \times q}$ as

$$T_{[k]} := I - V_{[k]}\tilde{R}_{[k]}^{-1}V_{[k]}^\top, \quad (6.4)$$

where $\tilde{R}_{[k]} := V_{[k]}^\top V_{[k]} + R_{[k]}$. Decomposing $T_{[k]}$ as

$$T_{[k]} = G_{[k]}G_{[k]}^\top,$$

with $G_{[k]} \in \mathbb{R}^{q \times q}$, it follows from (6.3) that $S_{[k|k]}$ can be computed as

$$S_{[k|k]} = S_{[k|k-1]}G_{[k]}. \quad (6.5)$$

This procedure yields a very convenient manner to update the error covariance matrix. Indeed, it shows that the actual error covariance matrix $P_{[k|k]}$ never needs to be computed.

Potter's algorithm has been extended in a variety of ways. We discuss two extensions. Section 6.3.1 considers *simultaneous processing*, which is most efficient when the number of measurements is high. Section 6.3.2 discusses *sequential processing*, which is most efficient when there are only few measurements.

6.3.1 Simultaneous processing

A direct implementation of the procedure above can be very inefficient if the number of measurements p is much larger than q . Indeed, the inversion of $\tilde{R}_{[k]}$ then takes up a lot of the computation time. This inversion is avoided in the measurement update of the ensemble transform Kalman filter (ETKF) [13] and the ensemble adjustment Kalman filter (EAKF) [5] by the use of the matrix inversion lemma. We discuss here the update of the ETKF. That of the EAKF is very similar.

Using the matrix inversion lemma, Bishop et al. [13] observed that (6.4) can be rewritten as

$$T_{[k]} = (I + V_{[k]}R_{[k]}^{-1}V_{[k]}^T)^{-1}.$$

Now, define the square matrix $W_{[k]} \in \mathbb{R}^{q \times q}$ as

$$W_{[k]} := V_{[k]}R_{[k]}^{-1}V_{[k]}^T.$$

Let the eigenvalue decomposition of $W_{[k]}$ be given by

$$W_{[k]} = U_{[k]}\Lambda_{[k]}U_{[k]}^T, \quad (6.6)$$

where $\Lambda_{[k]}$ contains the eigenvalues, ordered from large to small. Then, it is straightforward to show that

$$T_{[k]} = U_{[k]}(I + \Lambda_{[k]})^{-1}U_{[k]}^T \quad (6.7)$$

is the eigenvalue decomposition of $T_{[k]}$. Consequently, $G_{[k]}$ can be obtained as

$$G_{[k]} = U_{[k]}(I + \Lambda_{[k]})^{-\tau/2}.$$

Notice that the algorithm mainly works with small matrices. For example, the eigenvalue decomposition (6.6) is performed on the small $q \times q$ matrix $W_{[k]}$. The algorithm still requires the inversion of one large $p \times p$ matrix, namely that of $R_{[k]}$. However, the inverse of $R_{[k]}$ is mostly easy to compute since $R_{[k]}$ is usually diagonal or at least structured.

6.3.2 Sequential processing

The update in the previous section is performed simultaneously on all measurements. In this section, we consider an update that processes measurements sequentially, i.e. one after another.

Assume for the moment that $p = 1$. Then, Potter [111] showed that the measurement update can be implemented with matrix-vector multiplications only, that is, without any matrix-matrix multiplication. This yields a significant decrease in computational complexity.

For $p = 1$, it follows that $R_{[k]}$ and $\tilde{R}_{[k]}$ are scalars and $V_{[k]}$ is a vector. For convenience of notation, we define

$$\begin{aligned}\sigma_{[k]} &:= R_{[k]} \\ v_{[k]} &:= V_{[k]} \\ \alpha_{[k]} &:= \tilde{R}_{[k]}^{-1} = \frac{1}{v_{[k]}^T v_{[k]} + \sigma_{[k]}}.\end{aligned}$$

Potter observed that if a scalar $\gamma_{[k]}$ can be found so that

$$\begin{aligned}T_{[k]} &= (I - \alpha_{[k]} v_{[k]} v_{[k]}^T) \\ &= (I - \gamma_{[k]} \alpha_{[k]} v_{[k]} v_{[k]}^T)^2,\end{aligned}\tag{6.8}$$

then the update (6.5) can be written as

$$S_{[k|k]} = S_{[k|k-1]} (I - \gamma_{[k]} \alpha_{[k]} v_{[k]} v_{[k]}^T).\tag{6.9}$$

Solving (6.8) for $\gamma_{[k]}$, yields

$$\gamma_{[k]} = \frac{1}{1 + \sqrt{\alpha_{[k]} \sigma_{[k]}}}.\tag{6.10}$$

Furthermore, it follows from (2.14) that Kalman gain $l_{[k]} \in \mathbb{R}^n$ can be obtained as

$$l_{[k]} = \alpha_{[k]} S_{[k|k-1]} v_{[k]}.\tag{6.11}$$

Consequently, (6.9) can be rewritten as

$$S_{[k|k]} = S_{[k|k-1]} - \gamma_{[k]} l_{[k]} v_{[k]}^T.\tag{6.12}$$

Notice that the calculation of the gain matrix (2.14) and the update of error covariance square-root (6.9) indeed require only matrix-vector multiplications.

For $p > 1$, the measurements can be processed sequentially, that is, one after another [111]. It is assumed here that the noise on the measurements is uncorrelated. In case of correlation, the measurement vector must first be multiplied by $R_{[k]}^{-1/2}$. The algorithm thus requires $R_{[k]}^{-1/2}$ to be computed. However, the latter matrix is easy to compute since $R_{[k]}$ is usually diagonal or at least structured.

6.4 Reduced rank filtering

So far, we have been concerned only with the measurement update. In this section, we consider also the time update. We address the problem of

propagating a reduced rank approximation of the error covariance matrix and show that approximations are generally needed in order to preserve the rank during the time update.

This section is outlined as follows. In Sect. 6.4.1, we discuss the RRSQRT filter. Next, in Sect. 6.4.2, an extension of the RRSQRT filter is proposed that speeds up the algorithm, but is more approximate.

6.4.1 The reduced rank square-root filter

The RRSQRT filter [134] is a square-root algorithm based on an optimal lower rank approximation of the error covariance matrix. Although there exist nonlinear extension of the algorithm [66,135], we will mainly focus on the linear case. A nonlinear extension based on the EKF will be briefly summarized. For linear systems, the RRSQRT filter has the interesting property that it is algebraically equivalent to the Kalman filter if the rank of the error covariance matrix equals the dimension of the state vector.

The algorithm of the RRSQRT for such a system consists of three steps: the time update, the reduction step and the measurement update. These steps are now addressed.

6.4.1.1 Time update

It follows from (2.16)-(2.17) that the time update can be written in square-root form as

$$\hat{x}_{[k+1|k]} = A_{[k]}\hat{x}_{[k|k]} + B_{[k]}u_{[k]}, \quad (6.13)$$

$$S_{[k+1|k]} = \begin{bmatrix} A_{[k]}S_{[k|k]} & E_{[k]}Q_{[k]}^{1/2} \end{bmatrix}. \quad (6.14)$$

Notice that the number of columns in the square-root of the error covariance matrix grows from q to $q+l$. If this number of columns is not reduced, computation times will quickly blow up.

If the process noise is negligible, speed-up can be obtained by assuming $Q_{[k]} = 0$. The update of the error covariance square root and the computation of the Kalman gain can then efficiently be implemented by using the QR-decomposition. This leads to the *singular square-root Kalman filter* [10].

6.4.1.2 Reduction step

The augmentation of the rank during the time update, can quickly blow up computation times. Therefore, the number of columns in $S_{[k+1|k]}$ is reduced from $q+l$ back to q by truncating the approximate error covariance matrix

$$P_{[k+1|k]} := S_{[k+1|k]}S_{[k+1|k]}^T \in \mathbb{R}^{n \times n}$$

after the q largest eigenvalues and corresponding eigenvectors. It turns out that the eigenvalue decomposition of $P_{[k+1|k]}$ can be computed without forming

the matrix $P_{[k+1|k]}$. Indeed, the eigenvalue decomposition of $P_{[k+1|k]}$ can be computed from the one of the much smaller matrix

$$S_{[k+1|k]}^T S_{[k+1|k]} \in \mathbb{R}^{(q+l) \times (q+l)},$$

as will now be shown. Let the eigenvalue decomposition of $S_{[k+1|k]}^T S_{[k+1|k]}$ be given by

$$S_{[k+1|k]}^T S_{[k+1|k]} = X_{[k]} \Omega_{[k]} X_{[k]}^T,$$

then it is straightforward to show that

$$P_{[k+1|k]} = (S_{[k+1|k]} X_{[k]} \Omega_{[k]}^{-1/2}) \Omega_{[k]} (S_{[k+1|k]} X_{[k]} \Omega_{[k]}^{-1/2})^T$$

is the reduced eigenvalue decomposition of $P_{[k+1|k]}$. And thus,

$$[S_{[k+1|k]} X_{[k]}]_{(:,1:q)},$$

where $A_{(:,1:q)}$ denotes the matrix formed from A by retaining only its first q columns, is a square-root of the optimal rank- q approximation of $P_{[k+1|k]}$. Since $q, l \ll n$ this procedure is much faster than first forming $P_{[k+1|k]}$ and then applying an eigenvalue decomposition directly on $P_{[k+1|k]}$.

6.4.1.3 Measurement update

The RRSQRT filter was proposed in [134] with the sequential update of Potter. However, any square-root formulation of the Kalman filter measurement update can in principle be used.

6.4.1.4 Extension to nonlinear systems

In case of a nonlinear system of the form,

$$\begin{aligned} x_{[k+1]} &= f(x_{[k]}, u_{[k]}, k) + E_{[k]} w_{[k]} \\ y_{[k]} &= C_{[k]} x_{[k]} + v_{[k]}, \end{aligned}$$

where $w_{[k]}$ and $v_{[k]}$ have the usual properties, one can use an EKF-like approach to deal with the nonlinearity. This results in a time-update of the error covariance square-root of the form

$$S_{[k+1|k]} = \begin{bmatrix} \frac{\partial f}{\partial x}(\hat{x}_{[k|k]}, u_{[k]}) S_{[k|k]} & E_{[k]} Q_{[k]}^{1/2} \end{bmatrix}.$$

However, computing the Jacobian $\frac{\partial f}{\partial x}$ can be very time consuming. A more simple and efficient approach consists in approximating

$$\left(\frac{\partial f}{\partial x}(\hat{x}_{[k|k]}, u_{[k]}) \right) S_{[k|k]}$$

using finite differences [135]. The resulting time update then requires $q + 1$ evaluations of the nonlinear model $f(\cdot)$.

6.4.1.5 Summary of filter equations

This section summarizes the equations of the RRSQRT filter for a nonlinear system. The equations consist of three steps: a measurement update, a time update and a reduction step. For the measurement update, we employ the sequential processing technique of Sect. 6.3.2.

Reduced rank square-root filter (RRSQRT)

- Time update

$$\begin{aligned}\hat{x}_{[k+1|k]} &= f(\hat{x}_{[k|k]}, u_{[k]}) \\ S_{[k+1|k]} &= \begin{bmatrix} \frac{\partial f}{\partial x}(\hat{x}_{[k|k]}, u_{[k]}) S_{[k|k]} & E_{[k]} Q_{[k]}^{1/2} \end{bmatrix}\end{aligned}$$

- Reduction step

$$\begin{aligned}S_{[k+1|k]}^T S_{[k+1|k]} &= X_{[k]} \Omega_{[k]} X_{[k]}^T \\ S_{[k+1|k]} &\leftarrow [S_{[k+1|k]} X_{[k]}]_{(:,1:q)}\end{aligned}$$

- Measurement update

– update of state estimate:

$$\begin{aligned}\hat{x}_{[k|k]} &= \hat{x}_{[k|k-1]} + L_{[k]}(y_{[k]} - C_{[k]}\hat{x}_{[k|k-1]}) \\ L_{[k]} &= \alpha_{[k]} S_{[k|k-1]} v_{[k]} \\ \alpha_{[k]} &= \frac{1}{v_{[k]}^T v_{[k]} + \sigma_{[k]}} \\ v_{[k]} &= (c_{[k]} S_{[k|k-1]})^T\end{aligned}$$

– Update of error covariance matrix:

$$\begin{aligned}S_{[k|k]} &= S_{[k|k-1]} - \gamma_{[k]} L_{[k]} v_{[k]}^T \\ \gamma_{[k]} &= \frac{1}{1 + \sqrt{\alpha_{[k]} \sigma_{[k]}}}\end{aligned}$$

6.4.2 The reduced rank transform square-root filter

The SVD-based reduction step in the RRSQRT filter can be very costly. It has been reported that the reduction step is in some cases even the most time consuming step of the RRSQRT filter. This motivates research to speed-up

the reduction step. In this section, we propose a variant of the RRSQRT filter in which the reduction is interweaved in the measurement update. We take as measurement update a variant of that in the ETKF and therefore call the resulting filter the *reduced rank transform square-root* (RRTSQRT) filter.

6.4.2.1 The algorithm

The idea behind the RRTSQRT filter is very simple. Instead of approximating the error covariance matrix, we make a lower rank approximation of another matrix in the simultaneous update of the ETKF. It turns out that approximating $T_{[k]}$ is most convenient for a combined reduction and update.

Let $S_{[k|k-1]}$ be an $n \times (q+l)$ error covariance square root obtained after the time update of the RRSQRT filter. Suppose that we want to reduce the number of columns back to q during the measurement update. It follows from (6.7) that the optimal rank q approximation of the $(q+l) \times (q+l)$ matrix $T_{[k]}$ is given by

$$\tilde{T}_{[k]} = \tilde{U}_{[k]}(I + \tilde{\Lambda}_{[k]})^{-1}\tilde{U}_{[k]}^T,$$

where $\tilde{\Lambda}_{[k]} := [\Lambda_{[k]}]_{(:,l+1:q+l)}$ and $\tilde{U}_{[k]} := [U_{[k]}]_{(:,l+1:q+l)}$. Consequently,

$$S_{[k|k]} = S_{[k|k-1]}\tilde{U}_{[k]}(I + \tilde{\Lambda}_{[k]})^{-T/2}$$

performs simultaneously a measurement update and a reduction based on an optimal rank q approximation of $T_{[k]}$. Notice that in order to make the optimal rank q approximation of $T_{[k]}$, the smallest q eigenvalues of $W_{[k]}$ need to be retained.

6.4.2.2 Properties

Since the reduction step is eliminated, it should come as no surprise that the RRTSQRT filter is computationally more efficient than the RRSQRT filter. The computational complexity of the RRTSQRT filter is compared to that of the RRSQRT filter and that of the Kalman filter in Table 6.1. It is assumed here that $n > p \gg q$. Also, it is assumed that $A_{[k]}$ is sparse, so that the covariance update in the Kalman filter takes $\mathcal{O}(n^2)$ flops instead of $\mathcal{O}(n^3)$. It follows from Table 6.1 that the computational saving of the RRTSQRT filter over the RRSQRT filter is roughly that of one reduction step.

	KF	RRSQRT	RRTSQRT
time update	$\mathcal{O}(n^2)$	$\mathcal{O}(nq)$	$\mathcal{O}(nq)$
measurement update	$\mathcal{O}(n^2p)$	$\mathcal{O}(nq(q+r))$	$\mathcal{O}(nq(q+r))$
reduction step	–	$\mathcal{O}(n(q+r)^2)$	–

Table 6.1: Comparison between the complexity of the Kalman filter (KF), the RRSQRT and the RRTSQRT.

The SVD based reduction step of the RRSQRT leads to an underestimation of the trace of the error covariance matrix, which is equivalent to an underestimation of the total variance. Due to the optimality of the lower rank approximation, the norm of the truncated part of the error covariance matrix is minimal and equals the first eigenvalue of the error covariance matrix that has been ignored. In the RRTSQRT filter, the norm of the truncated part is always larger than or equal to the first eigenvalue that has been ignored. The RRTSQRT filter thus gives less weight to the measurements and is therefore more vulnerable to filter divergence. It will turn out in a numerical example that underestimation in the RRTSQRT filter can be quite large, making the filter very sensitive to divergence.

6.5 Spatially localized filtering

In environmental problems, the correlation between grid cells drops relatively quickly with distance. When assimilating a single observation with the Kalman filter, this means that the value of a lot of cells will almost not change. As shown in Fig. 6.2, the idea behind spatially localized Kalman filtering is to update only those cells of which the actual correlation to the measurement lies above a certain threshold.

Houtekamer and Mitchell [71] even noted that the measurement update in the EnKF can be improved by excluding observations greatly distant from the grid point to be updated. They found out that this is due to the approximation of the error covariance matrix, which may cause spuriously large correlations between greatly separated grid points. A lot of researchers have since been experimenting with techniques that localize the covariance information. In [63, 72], the estimated covariances are multiplied element by element with a distance dependent correlation function that drops to zero beyond some prespecified distance.

A more theoretical treatment is given in the *spatially localized Kalman filter* (SLKF) of Barrero et al. [9], where the gain matrix is constrained a priori to update a specified subset of the states. The optimal gain matrix is then determined by a procedure similar to that of the Kalman filter in Sect. 2.4.1.

Constraining the states that are updated is also motivated by the observability problem that occurs in data assimilation. For example, in some applications, the number of available measurements may be too small to determine the entire state vector. Constraining the update to the observable part of the state vector can be beneficial in such cases.

This section is outlined as follows. In Sect. 6.5.1, we discuss the SLKF. Next, in Sect. 6.5.2, a reduced rank version of the SLKF is developed that processes measurements sequentially.

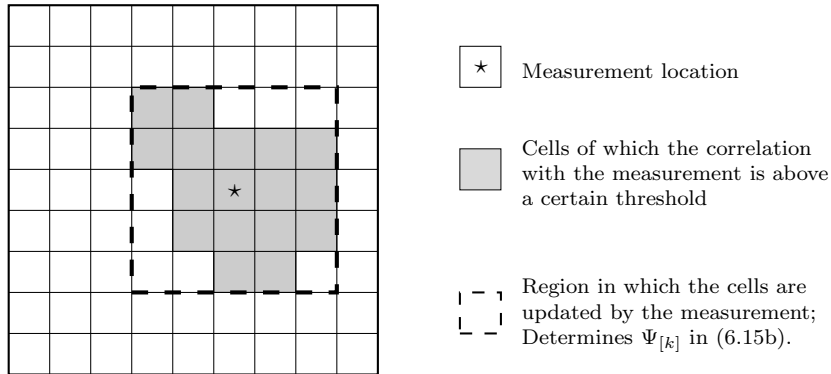


Figure 6.2: *The idea behind spatially localized Kalman filtering. In environmental problems, the correlation between grid cells drops usually very quickly with distance. When assimilating a single observation with the Kalman filter, this means that the value of a lot of cells will almost not change. The idea is to update only those cells of which the actual correlation to the measurement lies above a certain threshold.*

6.5.1 The spatially localized Kalman filter

Barrero et al [9] attached to the system (6.1) a recursive state estimator of the form

$$\hat{x}_{[k+1|k]} = A_{[k]}\hat{x}_{[k|k]} + B_{[k]}u_{[k]} \quad (6.15a)$$

$$\hat{x}_{[k|k]} = \hat{x}_{[k|k-1]} + \Psi_{[k]}L_{[k]}(y_{[k]} - C_{[k]}\hat{x}_{[k|k-1]}), \quad (6.15b)$$

where $\Psi_{[k]} \in \mathbb{R}^{n \times n_p}$ and $L_{[k]} \in \mathbb{R}^{n_p \times p}$. The nontraditional feature of (6.15) is the presence of the matrix $\Psi_{[k]}$, which equals the identity matrix in the classical Kalman filter case. Here, $\Psi_{[k]}$ constrains the state estimator so that only the states in the range of $\Psi_{[k]}$ are updated by the measurements. We assume that $\Psi_{[k]}$ has full column rank. For example, $\Psi_{[k]}$ can have the form

$$\Psi_{[k]} = \begin{bmatrix} I_{n_p} \\ 0 \end{bmatrix},$$

denoting that we want to update only the first n_p components of the state vector with the measurements. For an unobservable system, one can choose $\Psi_{[k]}$ so that only the observable part of the state vector is updated by the measurements.

It is shown in [9] that the optimal value of the gain matrix $\tilde{L}_{[k]}$ can be derived in a manner that is very similar to the derivation of the Kalman gain considered in Sect. 2.4.1. We now summarize the filter equations in a time update and a measurement update.

Spatially localized Kalman filter (SLKF)

- **Measurement update**

Due to the presence of the matrix $\Psi_{[k]}$, the measurement update is more complex than that of the Kalman filter. The update is given by

$$\hat{x}_{[k|k]} = \hat{x}_{[k|k-1]} + \Pi_{[k]} L_{[k]} (y_{[k]} - C_{[k]} \hat{x}_{[k|k-1]}) \quad (6.16)$$

$$\Pi_{[k]} = \Psi_{[k]} (\Psi_{[k]}^T \Psi_{[k]})^{-1} \Psi_{[k]}^T \quad (6.17)$$

$$L_{[k]} = P_{[k|k-1]} C_{[k]}^T \tilde{R}_{[k]}^{-1}$$

$$\tilde{R}_{[k]} = C_{[k]} P_{[k|k-1]} C_{[k]}^T + R_{[k]}$$

$$P_{[k|k]} = P_{[k|k-1]} - P_{[k|k-1]} C_{[k]}^T \tilde{R}_{[k]}^{-1} C_{[k]} P_{[k|k-1]} \\ + \Pi_{\perp[k]} P_{[k|k-1]} C_{[k]}^T \tilde{R}_{[k]}^{-1} C_{[k]} P_{[k|k-1]} \Pi_{\perp[k]}^T \quad (6.18)$$

$$\Pi_{\perp[k]} = I - \Pi_{[k]}. \quad (6.19)$$

- **Time update**

The time update takes the form of the update in the Kalman filter and is given by,

$$\hat{x}_{[k+1|k]} = A_{[k]} x_{[k]} + B_{[k]} u_{[k]}$$

$$P_{[k+1|k]} = A_{[k]} P_{[k|k]} A_{[k]}^T + E_{[k]} Q_{[k]} E_{[k]}^T.$$

Notice that the square matrix $\Pi_{[k]} \in \mathbb{R}^{n \times n}$ is an orthogonal projector, that is, $\Pi_{[k]}^2 = \Pi_{[k]}$ and $\Pi_{[k]}^T = \Pi_{[k]}$.

6.5.2 Reduced rank spatially localized filtering

Due to the complexity of the measurement update, the SLKF is computationally more demanding than the Kalman filter. Hence, for use in data assimilation, the algorithm needs to be approximated. Barrero et al. [9] proposed a suboptimal spatially localized filter based on the principles of the EnKF.

In this section, we develop a reduced rank version of the spatially localized Kalman filter that processes measurements sequentially. We consider two variants. The first variant, considered in Sect. 6.5.2.1 is most general in the sense that it holds for any $\Psi_{[k]}$ and any $P_{[k|k-1]}$. The second variant, considered in Sect. 6.5.2.2, exploits the specific structure of $\Psi_{[k]}$ and $P_{[k|k-1]}$ to further reduce the computation times. We address the structure that is obtained if the correlation between grid cells drop to zero within a distance of a few grid cells. The resulting algorithm is extremely efficient, especially if there are only few measurements.

6.5.2.1 General form

We propose an algorithm that uses the same principles as the RRSQRT filter and thus makes an optimal lower rank approximation of the error covariance matrix. We consider sequential processing of measurements.

Derivation of filter equations

We consider only the measurement update. The time update can be implemented as in (6.13)-(6.14). Assume for the moment that $p = 1$ and define for convenience of notation the row vector $c_{[k]} := C_{[k]}$. Notice that (6.18) can be rewritten as

$$P_{[k|k]} = P_{1[k|k]} + P_{2[k|k]},$$

where $P_{1[k|k]}$ and $P_{2[k|k]}$ are defined by

$$P_{1[k|k]} := P_{[k|k-1]} - \alpha_{[k]} P_{[k|k-1]} c_{[k]}^T c_{[k]} P_{[k|k-1]}, \quad (6.20)$$

and

$$P_{2[k|k]} := \alpha_{[k]} \Pi_{\perp[k]} P_{[k|k-1]} c_{[k]}^T c_{[k]} P_{[k|k-1]} \Pi_{\perp[k]}^T, \quad (6.21)$$

respectively, with $P_{[k|k-1]} = S_{[k|k-1]} S_{[k|k-1]}^T$. Consequently, $S_{[k|k]}$ can be obtained as

$$S_{[k|k]} = \begin{bmatrix} S_{1[k|k]} & S_{2[k|k]} \end{bmatrix},$$

where $S_{1[k|k]}$ and $S_{2[k|k]}$ obey $S_{1[k|k]} S_{1[k|k]}^T = P_{1[k|k]}$ and $S_{2[k|k]} S_{2[k|k]}^T = P_{2[k|k]}$.

The problem has thus reduced to finding such $S_{1[k|k]}$ and $S_{2[k|k]}$. Since (6.20) takes the form of the update in the Kalman filter, it immediately follows from (6.12) that $S_{1[k|k]}$ can be computed as

$$S_{1[k|k]} = S_{[k|k-1]} - \gamma_{[k]} l_{[k]} v_{[k]}^T,$$

where $\gamma_{[k]}$ is given by (6.10) and $L_{[k]}$ by (6.11). An expression for $S_{2[k|k]}$ follows from (6.21),

$$\begin{aligned} S_{2[k|k]} &= \sqrt{\alpha_{[k]}} \Pi_{\perp[k]} P_{[k|k-1]} c_{[k]}^T \\ &= \frac{1}{\sqrt{\alpha_{[k]}}} \Pi_{\perp[k]} l_{[k]}. \end{aligned}$$

Notice that the number of columns in the error covariance square-root grows with one during the update. Consequently, a reduction step is needed in order to confine the computational complexity.

Summary of filter equations

The filter equations consist of three steps: a measurement update, a time update and a reduction step. These steps are given by:

Reduced rank spatially localized Kalman filter (RRSLSQRT) – general form

- **Measurement update**

We give an update for $p = 1$. For $p > 1$, measurements can be processed sequentially.

– Update of state estimate:

$$\begin{aligned}\hat{x}_{[k|k]} &= \hat{x}_{[k|k-1]} + \Pi_{[k]} L_{[k]} (y_{[k]} - c_{[k]} \hat{x}_{[k|k-1]}) \\ \Pi_{[k]} &= \Psi_{[k]} (\Psi_{[k]}^T \Psi_{[k]})^{-1} \Psi_{[k]}^T \\ L_{[k]} &= \alpha_{[k]} S_{[k|k-1]} v_{[k]} \\ \alpha_{[k]} &= \frac{1}{v_{[k]}^T v_{[k]} + \sigma_{[k]}} \\ v_{[k]} &= (c_{[k]} S_{[k|k-1]})^T.\end{aligned}\tag{6.22}$$

– Update of error covariance matrix:

$$S_{[k|k]} = \begin{bmatrix} S_{1[k|k]} & S_{2[k|k]} \end{bmatrix}$$

with

$$\begin{aligned}S_{1[k|k]} &= S_{[k|k-1]} - \gamma_{[k]} L_{[k]} v_{[k]}^T \\ \gamma_{[k]} &= \frac{1}{1 + \sqrt{\alpha_{[k]} \sigma_{[k]}}}\end{aligned}$$

and

$$\begin{aligned}S_{2[k|k]} &= \frac{1}{\sqrt{\alpha_{[k]}}} \Pi_{\perp[k]} L_{[k]} \\ \Pi_{\perp[k]} &= I - \Pi_{[k]}.\end{aligned}$$

- **Time update**

The time update takes the form of the update in the RRSQRT filter and is given by,

$$\begin{aligned}\hat{x}_{[k+1|k]} &= A_{[k]} \hat{x}_{[k|k]} + B_{[k]} u_{[k]}, \\ S_{[k+1|k]} &= \begin{bmatrix} A_{[k]} S_{[k|k]} & E_{[k]} Q_{[k]}^{1/2} \end{bmatrix}.\end{aligned}$$

- **Reduction step**

The reduction step consists of an eigenvalue decomposition and a truncation,

$$\begin{aligned} S_{[k+1|k]}^\top S_{[k+1|k]} &= X_{[k]} \Omega_{[k]} X_{[k]}^\top \\ S_{[k+1|k]} &\leftarrow [S_{[k+1|k]} X_{[k]}]_{(:,1:q)}. \end{aligned}$$

Notice that one can actually choose a different $\Psi_{[k]}$ for each of the p measurements, meaning that one can determine the components of the state estimate that are updated individually for each measurement.

Except for the calculation of $\Pi_{[k]}$, the measurement update can be implemented with matrix-vector products only. This yields a huge decrease in computational complexity. Furthermore, in practice $\Pi_{[k]}$ is not computed based on (6.22), but is mostly chosen as a diagonal matrix, see Example 6.2. This further reduces computation times.

6.5.2.2 Efficient form

In this section, we consider yet a further reduction in computational complexity. This reduction is based on the assumption that correlation between grid cells decreases to zero within a distance of only a few grid cells.

Derivation of filter equations

Consider a PDE that has been discretized over a one-dimensional grid with n cells. Results are easily generalized to higher dimensional problems. Assume that $c_{[k]}$ takes the form $c_{[k]} = [0 \dots 0 \ 1 \ 0 \dots 0]$, where the 1 is in the i -th column. That is, the value of the i -th grid cell is measured. Furthermore, assume that correlation between grid cell i and its neighboring cells drops to zero in a distance of $j > 0$ grid cells. Then, only the grid cells $i - j, i - j + 1, \dots, i, \dots, i + j - 1, i + j$ will be updated by the measurement of grid cell i . Consequently, we may choose the matrix $\Psi_{[k]}$ as

$$\Psi_{[k]} = \begin{bmatrix} 0 & & \\ & I_{2j+1} & \\ & & 0 \end{bmatrix},$$

where the identity matrix goes from row $i - j$ to row $i + j$. For this choice of $\Psi_{[k]}$, it follows from (6.17) that $\Pi_{[k]}$ takes the form

$$\Pi_{[k]} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & I_{2j+1} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

By exploiting this particular structure, we will now reduce the computations needed for the update of the error covariance matrix. First, note that (6.18) can be rewritten for $p = 1$ as

$$\begin{aligned} P_{[k|k]} &= P_{[k|k-1]} + \alpha_{[k]} \Pi_{[k]} P_{[k|k-1]} c_{[k]}^T c_{[k]} P_{[k|k-1]} \Pi_{[k]}^T \\ &\quad - \alpha_{[k]} \Pi_{[k]} P_{[k|k-1]} c_{[k]}^T c_{[k]} P_{[k|k-1]} - \alpha_{[k]} P_{[k|k-1]} c_{[k]}^T c_{[k]} P_{[k|k-1]} \Pi_{[k]}^T. \end{aligned} \quad (6.23)$$

The structure and the sum of the last three terms is schematically shown in Fig. 6.3. The shaded squares denote the nonzero entries in the matrices. The white areas are zero due to the structure of $\Pi_{[k]}$. The gray areas are zero because the correlation between grid cell i and its neighboring grid cells is assumed to drop to zero in a distance of j cells. It follows from the sum in Fig. 6.3 that (6.23) reduces to

$$P_{[k|k]} = P_{[k|k-1]} - \alpha_{[k]} \Pi_{[k]} P_{[k|k-1]} c_{[k]}^T c_{[k]} P_{[k|k-1]} \Pi_{[k]}^T. \quad (6.24)$$

Notice that (6.24) updates only the area in the error covariance matrix represented by the shaded square. The update of this shaded square can be summarized as

$$\bar{\Pi}_{[k]} P_{[k|k]} \bar{\Pi}_{[k]}^T = \bar{\Pi}_{[k]} P_{[k|k-1]} \bar{\Pi}_{[k]}^T - \alpha_{[k]} \bar{\Pi}_{[k]} P_{[k|k-1]} c_{[k]}^T c_{[k]} P_{[k|k-1]} \bar{\Pi}_{[k]}^T,$$

where $\bar{\Pi}_{[k]} := [0 \ I_{2j+1} \ 0]$.

We now derive the efficient square-root update. Defining

$$\begin{aligned} \bar{P}_{[k|k]} &:= \bar{\Pi}_{[k]} P_{[k|k]} \bar{\Pi}_{[k]}^T \\ \bar{P}_{[k|k-1]} &:= \bar{\Pi}_{[k]} P_{[k|k-1]} \bar{\Pi}_{[k]}^T \\ \bar{S}_{[k|k-1]} &:= \bar{\Pi}_{[k]} S_{[k|k-1]} \\ \bar{c}_{[k]} &:= c_{[k]} \bar{\Pi}_{[k]}^T, \end{aligned}$$

it follows that

$$\begin{aligned} \bar{P}_{[k|k]} &= \bar{P}_{[k|k-1]} - \bar{\alpha}_{[k]} \bar{P}_{[k|k-1]} \bar{c}_{[k]}^T \bar{c}_{[k]} \bar{P}_{[k|k-1]} \\ &= \bar{S}_{[k|k-1]} (I - \bar{\alpha}_{[k]} \bar{v}_{[k]} \bar{v}_{[k]}^T) \bar{S}_{[k|k-1]}^T, \end{aligned} \quad (6.25)$$

where $\bar{v}_{[k]} := (\bar{c}_{[k]} \bar{S}_{[k|k-1]})^T$ and $\bar{\alpha}_{[k]} := 1/(\bar{v}_{[k]}^T \bar{v}_{[k]} + \sigma_{[k]})$. Equation (6.25) follows from the fact that $c_{[k]}^T = \bar{\Pi}_{[k]}^T \bar{\Pi}_{[k]} c_{[k]}^T$. Using a procedure similar to that of Sect. 6.3.2, we finally obtain the following square-root update,

$$\begin{aligned} \bar{S}_{[k|k]} &= \bar{S}_{[k|k-1]} (I - \bar{\gamma}_{[k]} \bar{\alpha}_{[k]} \bar{v}_{[k]} \bar{v}_{[k]}^T) \\ &= \bar{S}_{[k|k-1]} - \bar{\gamma}_{[k]} \bar{L}_{[k]} \bar{v}_{[k]}^T, \end{aligned}$$

where $\bar{\gamma}_{[k]} = 1/(1 + \sqrt{\bar{\alpha}_{[k]} \sigma_{[k]}})$ and

$$\bar{L}_{[k]} = \bar{\alpha}_{[k]} \bar{S}_{[k|k-1]} \bar{v}_{[k]}.$$

The filter equations are summarized in the following section.

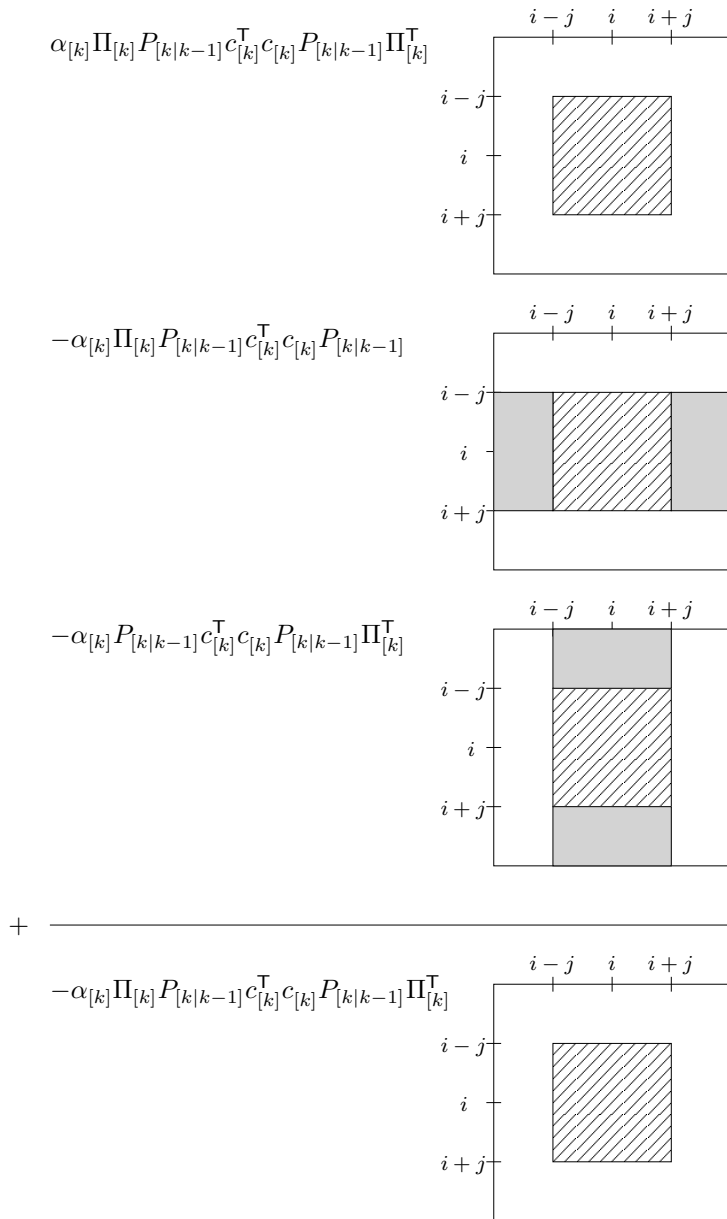


Figure 6.3: Structure and sum of the last three terms in (6.23). The shaded squares denote the nonzero entries in the matrices. The white areas are zero due to the choice of $\Pi_{[k]}$. The gray areas are zero because the correlation between grid cell i and its neighboring grid cells is assumed to drop to zero in a distance of j cells.

Summary of filter equations

The filter equations consist of three steps: a measurement update, a time update and a reduction step. These steps are given by:

Reduced rank spatially localized Kalman filter (RRSLSQRT) – efficient form

- **Measurement update**

– Initialization:

$$\begin{aligned}\hat{x}_{[k|k]} &\leftarrow \hat{x}_{[k|k-1]} \\ S_{[k|k]} &\leftarrow S_{[k|k-1]}\end{aligned}$$

– Select appropriate rows or columns:

$$\begin{aligned}\bar{S}_{[k|k-1]} &= \bar{\Pi}_{[k]} S_{[k|k-1]} \\ \hat{\bar{x}}_{[k|k-1]} &= \bar{\Pi}_{[k]} \hat{x}_{[k|k-1]} \\ \bar{c}_{[k]} &= c_{[k]} \bar{\Pi}_{[k]}^T\end{aligned}$$

– Update of state estimate:

$$\begin{aligned}\bar{\Pi}_{[k]} \hat{x}_{[k|k]} &\leftarrow \hat{\bar{x}}_{[k|k-1]} + \bar{L}_{[k]} (y_{[k]} - \bar{c}_{[k]} \hat{\bar{x}}_{[k|k-1]}) \\ \bar{L}_{[k]} &= \bar{\alpha}_{[k]} \bar{S}_{[k|k-1]} \bar{v}_{[k]} \\ \bar{\alpha}_{[k]} &= \frac{1}{\bar{v}_{[k]}^T \bar{v}_{[k]} + \sigma_{[k]}} \\ \bar{v}_{[k]} &= (\bar{c}_{[k]} \bar{S}_{[k|k-1]})^T.\end{aligned}$$

– Update of error covariance matrix:

$$\begin{aligned}\bar{\Pi}_{[k]} S_{[k|k]} &\leftarrow \bar{S}_{[k|k-1]} - \bar{\gamma}_{[k]} \bar{L}_{[k]} \bar{v}_{[k]}^T \\ \bar{\gamma}_{[k]} &= \frac{1}{1 + \sqrt{\bar{\alpha}_{[k]} \sigma_{[k]}}}\end{aligned}$$

- **Time update**

The time update takes the form of the update in the RRSQRT filter and is given by,

$$\begin{aligned}\hat{x}_{[k+1|k]} &= A_{[k]} \hat{x}_{[k|k]} + B_{[k]} u_{[k]}, \\ S_{[k+1|k]} &= \begin{bmatrix} A_{[k]} S_{[k|k]} & E_{[k]} Q_{[k]}^{1/2} \end{bmatrix}.\end{aligned}$$

• **Reduction step**

The reduction step consists of an eigenvalue decomposition and a truncation,

$$\begin{aligned} S_{[k+1|k]}^\top S_{[k+1|k]} &= X_{[k]} \Omega_{[k]} X_{[k]}^\top \\ S_{[k+1|k]} &\leftarrow [S_{[k+1|k]} X_{[k]}]_{(:,1:q)}. \end{aligned}$$

In practical applications, it is to be expected that correlation drops almost to zero over a number of grid cells that is much smaller than the total number of cells in the grid. Consequently, the measurement update will perform operations on matrices of which the size is much smaller than the total number of grid cells. This yields a huge decrease in computation times over the equations of the general algorithm.

6.6 Filter degradation due to a lower rank approximation of the error covariance matrix

As already discussed, the SVD based reduction step in the reduced rank filter described above leads to an underestimation of the total variance. In this section, another consequence of a lower rank approximation will be discussed. It will be shown that a lower rank approximation may lead to an increase in the actual variance of some grid cells although the filter thinks that the variance has decreased.

The analysis in this section considers the effect of deterministic error in the covariance matrix (for example due to an SVD based reduction) on the actual variance of the estimates. A similar analysis has been used by Houtekamer and Mitchell [71] for random errors. The derivation in this section allows to determine the grid cells of which the actual variance will increase by assimilating a particular measurement due a lower rank approximation of the error covariance matrix.

For conciseness of equations, we use the following notations: $\text{Cov}(x_1, x_2)$ denotes the covariance matrix of $x_1 \in \mathbb{R}$ and $x_2 \in \mathbb{R}$, $\text{Var}(x)$ denotes the variance of the random variable $x \in \mathbb{R}$.

Consider two random variables $x_1 \in \mathbb{R}$ and $x_2 \in \mathbb{R}$ that have to be estimated based on a measurement y_1 of x_1 , a priori estimates \hat{x}_1 and \hat{x}_2 , and the covariance matrix

$$\text{Cov}(x_1, x_2) =: \begin{bmatrix} \sigma_1^2 & c_{12} \\ c_{12} & \sigma_2^2 \end{bmatrix}.$$

Assuming that the measurement error is uncorrelated to the errors in \hat{x}_1 and

\hat{x}_2 , the MVU estimates $\hat{x}_{1|y}$ and $\hat{x}_{2|y}$ of x_1 and x_2 , are given by

$$\begin{aligned}\hat{x}_{1|y} &= \hat{x}_1 + \frac{\sigma_1^2}{\sigma^2 + \sigma_1^2}(y_1 - \hat{x}_1) \\ \hat{x}_{2|y} &= \hat{x}_2 + \frac{c_{12}}{\sigma^2 + \sigma_1^2}(y_1 - \hat{x}_1),\end{aligned}\quad (6.26)$$

and their variances by

$$\text{Var}(\hat{x}_{1|y}) = \sigma_1^2 \left(1 - \frac{\sigma_1^2}{\sigma^2 + \sigma_1^2}\right) \quad (6.27)$$

$$\text{Var}(\hat{x}_{2|y}) = \sigma_2^2 - \frac{c_{12}^2}{\sigma^2 + \sigma_1^2}, \quad (6.28)$$

where σ^2 denotes the variance of the measurement error. These equations are not more than the Kalman filter equations in scalar form. It follows from (6.27)-(6.28) that the variances reduces by assimilating the observation (or stays constant for $c_{12} = 0$). As will be shown in the next section, this no longer necessarily holds if there is an error in the covariance c_{12} .

6.6.1 Error in the covariances

Consider the approximate covariance matrix

$$\begin{bmatrix} \sigma_1^2 & c_{12} + e_{12} \\ c_{12} + e_{12} & \sigma_2^2 \end{bmatrix},$$

with an error e_{12} in the covariance. It follows from (6.26) that the estimate of x_2 now becomes

$$\hat{x}_{2|y} = \hat{x}_2 + \frac{c_{12} + e_{12}}{\sigma^2 + \sigma_1^2}(y_1 - \hat{x}_1).$$

Its variance is given by

$$\begin{aligned}\text{Var}(\hat{x}_{2|y}) &= \mathbb{E}[(x_{2[k]} - \hat{x}_{2|y})(x_{2[k]} - \hat{x}_{2|y})^T] \\ &= \sigma_2^2 + \frac{(c_{12} + e_{12})^2}{\sigma^2 + \sigma_1^2} - 2 \frac{(c_{12} + e_{12})c_{12}}{\sigma^2 + \sigma_1^2} \\ &= \sigma_2^2 - \frac{c_{12}^2}{\sigma^2 + \sigma_1^2} \left(1 - \left(\frac{e_{12}}{c_{12}}\right)^2\right).\end{aligned}$$

If $|e_{12}/c_{12}| > 1$, it follows that $\text{Var}(\hat{x}_{2|y}) > \sigma_2^2$, meaning that the estimate degrades by assimilating the measurement y_1 .

6.6.2 Error in the variances and the covariances

Now, consider the approximate error covariance matrix

$$\begin{bmatrix} \sigma_1^2 + e_1 & c_{12} + e_{12} \\ c_{12} + e_{12} & \sigma_2^2 + e_2 \end{bmatrix}.$$

in which both the variances and covariances have an error. The estimate of x_2 now becomes

$$\hat{x}_{2|y} = \hat{x}_2 + \frac{c_{12} + e_{12}}{\sigma^2 + (\sigma_1^2 + e_1)}(y_1 - \hat{x}_1).$$

This estimate has variance

$$\begin{aligned} \text{Var}(\hat{x}_{2|y}) &= \sigma_2^2 + \frac{(c_{12} + e_{12})^2(\sigma_1^2 + \sigma^2)}{(\sigma^2 + (\sigma_1^2 + e_1))^2} - 2 \frac{(c_{12} + e_{12})c_{12}}{\sigma^2 + (\sigma_1^2 + e_1)} \\ &= \sigma_2^2 + \frac{(e_{12}^2 - c_{12}^2)(\sigma^2 + \sigma_1^2) - 2c_{12}e_1(c_{12} + e_{12})}{(\sigma^2 + (\sigma_1^2 + e_1))^2}. \end{aligned}$$

We conclude that the estimate degrades if

$$e_{12}^2(\sigma^2 + \sigma_1^2) > c_{12}^2(\sigma^2 + \sigma_1^2) + 2c_{12}e_1(c_{12} + e_{12}).$$

The estimate of x_1 becomes

$$\hat{x}_{1|y} = \hat{x}_1 + \frac{\sigma_1^2 + e_1}{\sigma^2 + (\sigma_1^2 + e_1)}(y_1 - \hat{x}_1).$$

This estimate has variance

$$\begin{aligned} \text{Var}(\hat{x}_{1|y}) &= \sigma_1^2 + \frac{(\sigma_1^2 + e_1)^2(\sigma_1^2 + \sigma^2)}{(\sigma^2 + (\sigma_1^2 + e_1))^2} - 2 \frac{(\sigma_1^2 + e_1)\sigma_1^2}{\sigma^2 + (\sigma_1^2 + e_1)} \\ &= \sigma_1^2 + \frac{e_1^2\sigma^2 - \sigma_1^6 - \sigma_1^4\sigma^2 - e_1^2\sigma_1^2 - 2e_1\sigma_1^4}{(\sigma^2 + (\sigma_1^2 + e_1))^2}. \end{aligned}$$

We conclude that the estimate degrades if

$$e_1^2\sigma^2 > e_1^2\sigma_1^2 + \sigma_1^6 + \sigma_1^4\sigma^2 + 2e_1\sigma_1^4.$$

6.7 Numerical examples

Two numerical examples are considered. The first example compares the RRSQRT filter and RRTSQR filter on a linear heat conduction example. The second example applies the RRSLSQRT filter to a chaotic nonlinear system.

Example 6.1. Linear heat conduction

Consider heat transfer in an infinitely thin ring, governed by the PDE

$$\frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2},$$

where $T(x, t)$ denotes the temperature at position x and time t and $\alpha = 2 \cdot 10^{-3}$ m²/s denotes the thermal diffusivity. Using a central difference method, the PDE is discretized over a grid with 100 cells. Discretization in time is achieved with a forward difference method with time step 10^{-2} s.

We assume that the discretized model is corrupted by noise that affects only a few cells, so that the matrix $E_{[k]} \in \mathbb{R}^{n \times l}$ has only few columns. We choose $l = 5$ and take the noise term $w_{[k]}$ as a zero-mean white process with $Q_{[k]} = 10^{-4}I$. It is assumed that measurements of every third grid cell are available. The noise term $v_{[k]}$ is chosen as a zero-mean white process with $R_{[k]} = 10^{-2}I$.

We choose $q = 25$ and compare the assimilation result of the RRTSQR filter to that of the RRSQR filter and the Kalman filter. Fig. 6.4 shows the MSE in the state estimates. It can be seen that the RRSQR filter converges almost as fast as the Kalman filter. The RRTSQR filter, on the other hand, converges much slower. This is due to the underestimation of the error covariance matrix. For comparison, during the first simulation step the reduction of the RRSQR filter retained 99.68% of the total variance, whereas that of the RRTSQR filter retained only 86.52%. The underestimation not only yields slower convergence, but also makes the filter more sensitive to divergence. The reason is that the filter gives too less weight to the measurements.

To reduce filter divergence in the EnKF, Anderson [6] introduced the idea of *covariance inflation*. Covariance inflation is a heuristic technique where the error covariance matrix (or a square-root of it) is multiplied by a factor $\kappa > 1$ that increases the total variance artificially. We choose the inflation factor so that the RRTSQR filter retains the same amount of the total variance as the RRSQR filter during the first simulation step. This yields an inflation factor of $\kappa = 1.075$. As can be seen in Fig. 6.4, covariance inflation strongly increases the speed of convergence. \square

Example 6.2. Nonlinear Lorenz model

In a second example, we consider the nonlinear Lorenz [90] model. This model mimics the propagation of an unspecified meteorological quantity along a latitude circle. It exhibits chaotic behavior. The equations are governed by

$$\frac{dx_i}{dt}(t) = (x_{i+1}(t) - x_{i-2}(t))x_{i-1}(t) - x_i(t) + f,$$

where the index $i = 1, \dots, n$ is cyclic so that $x_{i-n}(t) = x_{i+n}(t) = x_i(t)$. The symbol $x_i(t)$ denotes the meteorological quantity at the i -th grid point at time t . We choose $n = 40$ and $f = 8$ and discretize the model using a fourth-order Runge-Kutta scheme with a sampling time of 0.05s. Figure 6.5 illustrates the chaotic behavior of the system. This figure shows two trajectories of x_{20} . The initial state $x(0)$ of the second simulation differs only slightly from that of the first simulation. However, due to the chaotic behavior, the trajectories diverge very quickly.

We assume that the discretized model is corrupted by noise that affects only a few cells, so that the matrix $E_{[k]} \in \mathbb{R}^{n \times l}$ has only few columns. We choose

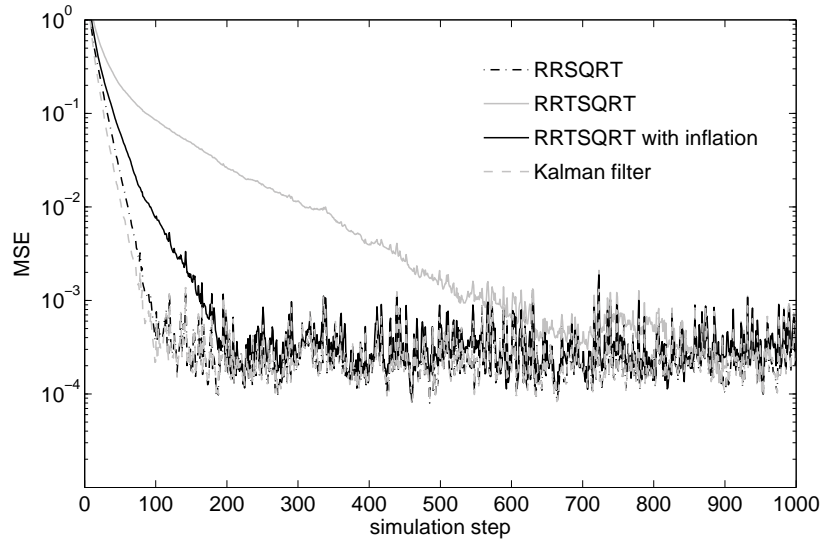


Figure 6.4: Comparison between the MSE of the RRSQRT, RRTSQRT, RRTSQRT with covariance inflation, and the Kalman filter on a linear heat transfer problem in an infinitely thin ring.

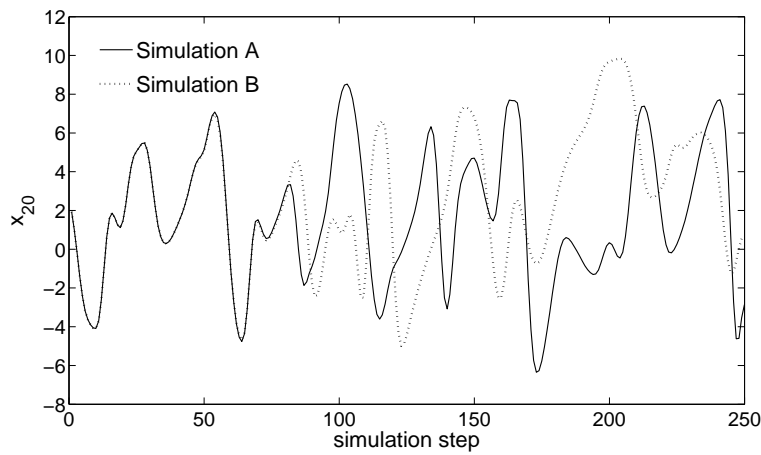


Figure 6.5: Illustration of the chaotic behavior of the Lorenz [90] model. The initial state of the second simulation differs only slightly from that of the first simulation. However, due to the chaotic behavior, the trajectories diverge very quickly.

$l = 10$ and take the disturbance $w_{[k]}$ itself as a zero-mean white process with $Q_{[k]} = 10^{-4}I$. It is assumed that measurements of the following six grid cells are available: 3, 10, 17, 24, 31 and 38. The measurement noise $v_{[k]} \in \mathbb{R}^6$ is chosen as a zero-mean white process with $R_{[k]} = 10^{-4}I$.

We use the procedure of *twin experiments*. This means that we first simulate the discretized model from an arbitrary initial state. We call the resulting trajectory the “true” trajectory. Then, we artificially create measurements based on the “true” trajectory. Finally, we add a perturbation to the initial state and let the filter recursively estimate the state trajectory based on the artificially created measurements.

In a first experiment, we study how the value of q affects the filter performance. We consider two criteria. The first criterion is the percentage of the total variance that is retained by the lower rank approximation. To this aim, we first apply the RRSQRT filter with a value of q that is assumed to give the correct picture of the error covariance matrix. We take $q = 200$ and consider the error covariance matrix after 100 simulation steps. Fig. 6.6 shows the percentage of the variance that is retained when making an optimal lower rank approximation of this error covariance matrix. A rank 10 approximation retains approximately 90%, a rank 15 approximation 95% and a rank 20 approximation 97.5%. As a second criterion, we consider “correctness” of the measurement update. With a “correct” update of a grid cell, we mean that the variance of that grid cell decreases by assimilating observations. Fig. 6.7 compares the “correctness” of the update for different values of q . A black square denotes a grid cell of which the variance increases by assimilating the particular measurement. The grid cell at which the measurement was taken, is denoted by a star. For a rank 5 approximation, almost half of the grid cells degrade by assimilating measurements. For $q = 20$, less than 20% of the grid cells degrade. In addition, for the latter value of q , mostly grid cells that are distant from the measurement point degrade. It turns out that correlation between the measurement and such grid points is very small, so that filter degradation will be only minor.

The discussion above indicates that estimation accuracy for $q = 20$ will be only slightly worse than for $q = 200$. On the other hand, the RRSQRT filter is far more efficient for $q = 20$ than for $q = 200$. A simulation example indeed confirms these findings. The state estimates for $q = 20$ and $q = 200$ are almost indistinguishable.

In a second experiment, we address the problem of spatially localized filtering. Fig. 6.8 compares the correlation matrices computed from an approximate error covariance matrix with $q = 200$ and a rank 20 approximation of that covariance matrix. The values show in the figure are actually the \log_{10} of the computed correlations. Notice that correlation drops almost to zero in a distance of approximately 10 grid cells. This motivates the use of the RRSLSQRT filter. Simulation results show that the general form and the efficient form of the RRSLSQRT filter perform almost equal. In fact, simulation results with $q = 20$ are almost indistinguishable from the RRSQRT filter with $q = 200$. Table 6.2 compares the processing time of the measurement update for various filter algorithms. Notice that the efficient form of the RRSLSQRT

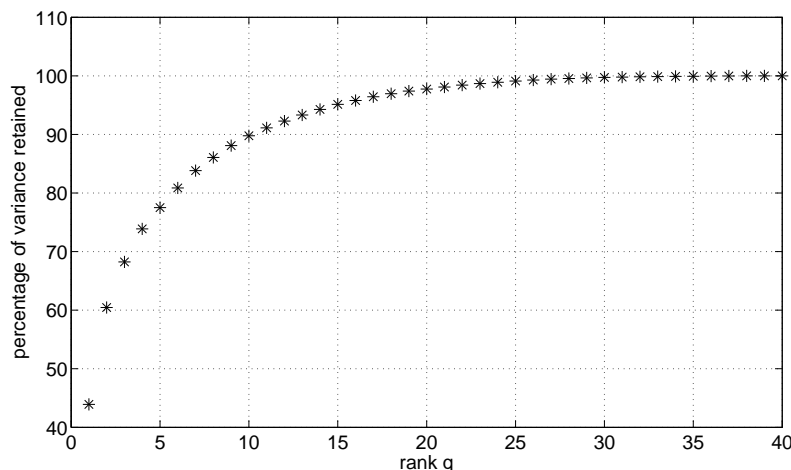


Figure 6.6: Percentage of the variance that is retained when making an optimal lower rank approximation of the error covariance matrix obtained with an RRSQRT filter with $q = 200$.

filter outperforms the general form. However, the efficient form is slower than the sequential update of the traditional RRSQRT filter. This is due to the overhead of selecting the rows of the error covariance square-root that need to be updated. It is, however, expected that the RRSLSQRT will also outperform the RRSQRT for larger values of n . The row selection will then turn into a computational advantage rather than into a disadvantage.

In a third experiment, we study the effect of decreasing the rank of the error covariance matrix below 20. Fig. 6.9 compares the MSE of the RRSQRT filter and the RRSLSQRT filter (efficient form) for $q = 15$. Notice that the RRSQRT filter blows up around simulation step 500. The RRSLSQRT filter, on the other, still performs satisfactory for $q = 10$. Simulations indicate that the efficient form is more accurate than the general form for such low values of q . \square

6.8 Conclusion

Two extensions of the RRSQRT filter were considered in this chapter.

The first extension, the reduced rank transform square-root (RRTSQRT) filter, speeds up the RRSQRT filter by interweaving the reduction step into the measurement update. The time saving is approximately that of the reduction step in the RRSQRT filter. The major drawback of the method is that the error covariance matrix is more underestimated, making the filter more sensitive to divergence. A simulation example has indicated that underestimation of the error covariance matrix really is an issue in the RRTSQRT.

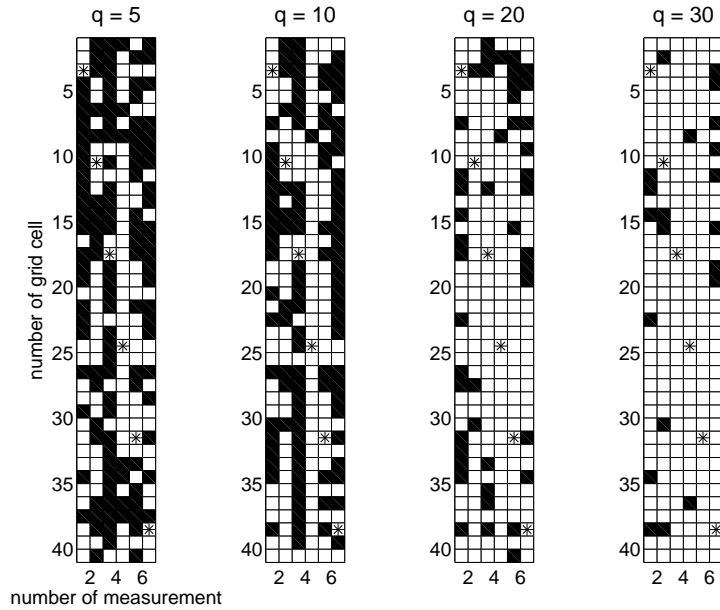


Figure 6.7: “Correctness” of the measurement update for different values of q . A black square denotes a grid cell of which the actual variance increases by assimilating the particular measurement, although the filter thinks that the variance has decreased. The grid cell at which the measurement was taken, is denoted by a star.

The second extension has addressed the problem of reduced rank spatially localized square-root (RRSLSQRT) filtering. Sequential updating of measurements was considered in which the grid cells that are updated can be defined for each of the measurements individually. Two versions of the algorithm were derived. The first form is equivalent to the spatially localized Kalman filter [9] if the rank of the error covariance matrix is chosen equal to the order of the system. The second form approximates the equations of the first from by assuming that correlation between widely separated grid cells equals zero. A simulation example on a nonlinear chaotic model has illustrated the advantages of spatially localized filtering. The spatially localized filter seems to be more robust to filter divergence than the traditional RRSQRT filter. The efficient form was in this example least sensitive to filter divergence.

algorithm	absolute time	relative time
RRSQRT ($q = 200$)	1.449 ms	1.57
RRSQRT ($q = 20$)	0.924 ms	1
RRSLSQRT (general form, $q = 20$)	2.781 ms	3.01
RRSLSQRT (efficient form, $q = 20$)	1.097 ms	1.19

Table 6.2: Processing time of the measurement update in Example 6.2.

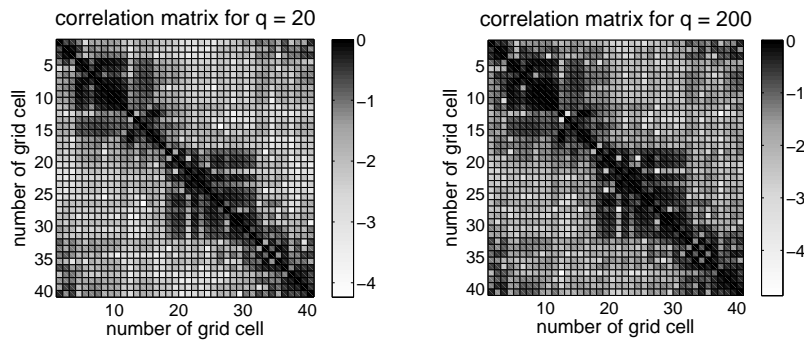


Figure 6.8: Comparison between the correlation matrices computed from an approximate error covariance matrix with $q = 200$ and a rank 20 approximation of that covariance matrix. The values show actually are the \log_{10} of the computed correlations. Notice that correlation drops quickly with distance. This motivates the use of the RRSLSQRT filter.

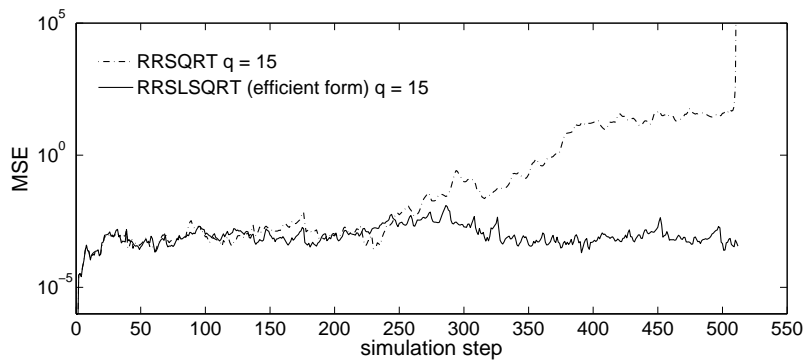


Figure 6.9: Comparison between the MSE of the RRSQRT filter and the RRSLSQRT filter (efficient form) for $q = 15$. Notice that the RRSQRT filter blows up around simulation step 500.

Chapter 7

Space Weather Nowcasting Example

This chapter considers the application of data assimilation techniques for nowcasting a space weather event. The RRSLSQRT developed in Chapter 6 is applied with a large-scale model, consisting of approximately 10^5 state variables. The model emulates the dynamics of the bow shock that is formed when the solar wind encounters the Earth. The performance of the RRSLSQRT is investigated for different types of spatial localization and different values of the rank of the approximate error covariance matrix. Simulations with both known constant and unknown time-varying boundary conditions are considered. It is found that even with only a few measurements, the RRSLSQRT yields a significant reduction in estimation error over a data-free simulation. Results also indicate that spatial localization of the measurements has a positive effect on estimation accuracy.

7.1 Introduction

Although data assimilation is performed almost routinely in meteorology, it is largely unknown in the space physics community. This is due to the fact that space physics is a relatively young research field in which the number of observations is quite small, orders of magnitude smaller than in meteorological applications. As the research and insight in space dynamics increases, it is to be expected that more and more satellites will be launched for the observation of space weather phenomena. However, the amount of data can almost impossibly reach that of Earthly weather observations. This sparseness of observations, together with the enormous length scales, makes data assimilation for space weather applications very challenging and requires the development of new techniques that are adapted to this situation.

A first step towards data assimilation for space weather nowcasting has been set by Chandrasekar et al. [17]. They derived a linearized model for plasma flow in a two-dimensional channel and compared the performance of the Kalman filter and the RRSQRT filter. In [122], sparseness of observations is addressed in a relatively simple nonlinear model. It is found that for a model with 100 grid cells, a systematic reduction in estimation error can be achieved with as few as 2 observations. A large-scale assimilation problem has first been considered by Barrero et al. [11]. They considered a simulation with approximately 10^4 state variables which emulates a solar storm interacting with the Earth's magnetosphere. The performance of the EnKF was compared for different numbers of observations. An extension to spatially localized filtering was addressed in [8], where it was shown that the localization introduces robustness against filter divergence in case of very few observations.

Personal contributions

This chapter addresses the use of the RRSLSQRT for estimating the topology of the bow shock that is formed when the solar wind encounters the Earth. We consider the complex bow shock model of [27] and set-up a large-scale data assimilation experiment. The grid considered in the assimilation experiment consists of approximately 10^5 cells, that is only a factor 10 smaller than the most realistic weather assimilation problems which are run on the most advanced supercomputers on Earth. The performance of the RRSLSQRT will be assessed in two series of experiments. In the first series, it is assumed that the boundary conditions, i.e. the properties of the incoming solar wind, are constant and known. The ability of the RRSLSQRT to reconstruct the actual plasma flow, starting from an incorrect initial estimate, will be studied under various conditions. For example, the influence of the type of spatial localization and the rank of the approximate error covariance matrix will be studied. In the second series of experiments, the boundary conditions are assumed to be time-varying and unknown. The RRSLSQRT is extended to estimate the boundary condition and its stability and performance under such conditions is assessed. It is found that even with only a few measurements, the RRSLSQRT yields a significant reduction in estimation error over a data-free simulation. Results also indicate that spatial localization of the measurements has a positive effect on estimation accuracy.

Chapter outline

This chapter is outlined as follows. Section 7.2 introduces the magnetohydrodynamic equations that govern the flow of plasma and thus form the basis of all simulation concerning the solar wind. In Sect. 7.3, the different types of shocks than occur in magnetohydrodynamics are discussed and their topology is considered. Finally, in Sect. 7.4, the application of the RRSLSQRT in the estimation the bow shock topology is considered.

7.2 Magnetohydrodynamics

Almost all matter on Earth is either in the solid, the liquid or the gas state. On the other hand, most of the matter in space is in the plasma state. A *plasma* can be considered as a gas consisting of positively and negatively charged particles. It is electrically neutral overall, but the presence of charged particles means that a plasma can support electric currents and that it can interact with electric and magnetic fields. This makes plasma behavior more complex and varied than neutral gas behavior. Plasma is therefore considered to be a distinct state of matter.

The behavior of a plasma can be modeled at three different levels. In the lowest level, the individual movement of particles is considered. For large plasma system, however, such a description is computationally not feasible. In the second level, the average behavior of the particles is described based on kinetic phenomena. Finally, in *magnetohydrodynamics* (MHD), the highest level, plasma is considered as a fluid. MHD thus yields a macroscopic description of a plasma.

The interaction of a plasma with magnetic and electric fields makes the MHD equations more complex than the equations that govern the flow of an electrically neutral fluid. The set of equations which describe MHD are a combination of the Navier-Stokes equations of fluid dynamics and Maxwell's equations of electromagnetism. The MHD equations are thus rich in nonlinearities.

In this chapter, the most simple form of MHD, *ideal* MHD, is considered. Ideal MHD basically assumes that the fluid has so little resistivity that it can be treated as a perfect conductor. The derivation of the ideal MHD equations is based on several assumptions and approximations of which the validity needs to be assessed in every application. The ideal MHD equations are applicable if the following conditions are satisfied [60]:

- The plasma is strongly collisional. The time scale of collisions is shorter than the other characteristic times in the system.
- The resistivity due to collisions is small, meaning that the magnetic diffusion times must be longer than any time scale of interest.
- The length scales and time scales under consideration are much longer than the characteristic length scales and time scales of the individual particles.

For the large-scale application considered in this chapter, the last two conditions are almost certainly satisfied. The first condition, on the other hand, is probably not satisfied. However, it turns out that in collisionless MHD, the part of the dynamics that is inaccurately described by MHD does not matter [60], which validates the use of the ideal MHD equations in this chapter. The ideal MHD equations are now discussed.

7.2.1 The ideal MHD equations

The ideal MHD equations can be written in several different forms. In Sect. 7.2.1.1, we write the equations in a traditional form that relates to the Navier-Stokes and Maxwell's equations. In Sect. 7.2.1.2, the equations are reformulated in terms of conservation laws. A rigorous derivation of the equations and a discussion concerning the conditions of their validity can be found in e.g. [59, 60],

7.2.1.1 Traditional form

At a given point in space and time, the state of a plasma fluid can be described by eight quantities: the density ρ , the pressure p , the three components of the velocity vector v and the three components of the magnetic field vector B . The dynamical evolution of these variables is governed by eight equations.

- The *mass continuity* equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0$$

describes the conservation of mass in the plasma flow.

- The *momentum* equation

$$\rho \frac{\partial v}{\partial t} + \rho(v \cdot \nabla)v = -\nabla p + J \times B \quad (7.1)$$

provides the next three equations. The *current density* J is given by Ampère's law,

$$J = \frac{1}{\mu_0} \nabla \times B,$$

where μ_0 denotes the *permeability of free space*.

- The *induction* equation

$$\frac{\partial B}{\partial t} = \nabla \times (v \times B) \quad (7.2)$$

provides three more equations. Notice that (7.2) derives from Faraday's law and from Ohm's law of induction,

$$E = -v \times B,$$

where E is the *electric field*.

- The final equation is the *pressure* equation,

$$\frac{\partial p}{\partial t} + (v \cdot \nabla)p + \gamma p \nabla \cdot v = 0,$$

where ideal gas behavior is assumed and where the *adiabatic index* $\gamma = 5/3$.

Because magnetic field lines are always closed, the MHD equations have to be supplemented with the divergence free condition

$$\nabla \cdot B = 0,$$

which can be seen as a constraint to the MHD equation.

7.2.1.2 Conservative form

The ideal MHD equations summarized above can be reformulated in terms of conservation laws by introducing the *total energy* e , defined by

$$e := \frac{p}{\gamma - 1} + \rho \frac{v \cdot v}{2} + \frac{B \cdot B}{2}.$$

The ideal MHD equations can then be written in conservative form as

$$\frac{\partial}{\partial t} \begin{bmatrix} \rho \\ \rho v \\ B \\ e \end{bmatrix} + \nabla \cdot \begin{bmatrix} \rho v \\ \rho v v + I(p + B \cdot B/2) - BB \\ vB - Bv \\ (e + p + B \cdot B/2)v - (v \cdot B)B \end{bmatrix} = 0.$$

The conservative form is used most often in computational MHD.

7.2.2 Computational MHD

Analytical solutions of the MHD equations are limited to the most simple cases, and even then a lot of approximations are usually made. Numerical simulations provide an effective manner to study the most complex plasma dynamics. Of course, numerical simulations also have their limitations, but in most cases they yield results that are more than satisfactory. MHD simulations have given new and interesting insights in space weather phenomena such as the propagation of the solar wind or of a coronal mass ejection.

The MHD simulations considered in the remainder of this chapter are performed with the Versatile Advection Code (VAC) [130]. The VAC is a general code developed for solving MHD and hydrodynamical problems on parallel computers. It allows the user to solve a hyperbolic system of PDE's with a variety of modern numerical schemes and provides methods to numerically maintain the $\nabla \cdot B = 0$ condition.

7.3 MHD shocks

When an airplane flies at supersonic speeds, a bow shock is formed in front of it. Upstream from the shock, the flow is supersonic, and downstream the flow is subsonic. Similarly, the supersonic solar wind induces a bow shock when it encounters the Earth. As we will see in this section, due to the magnetic field that is dragged by the solar wind, the topology of the latter bow shock can be much more complicated than that formed by an airplane.

Several MHD models have been developed that describe the topology of the Earth's bowshock (see e.g. [119] and the references therein) or the bowshock of other planets like Saturn [64] or Jupiter [89].

This section is outlined as follows. In Sect. 7.3.1, the topology of MHD shocks is discussed. Next, in Sect. 7.3.2, we consider the numerical technique in [27], which deals with two-dimensional MHD flow around a perfectly conduction cylinder.

7.3.1 Shock topology

While the equations of neutral gas dynamics only allow for one type of wave, the sound wave, MHD allows for three different types of waves: the *Alfvén* wave and the fast and slow *magnetosonic* wave. The wave speeds depend strongly on the angle between the direction of wave propagation and the direction of the local magnetic field. MHD waves are thus *anisotropic*.

As a result, there exist three different types of MHD shocks. As shown in Fig. 7.1, they can be distinguished by considering the way in which they refract the magnetic field lines.

- In *slow* MHD shocks (Fig. 7.1a), the magnetic field is refracted towards the shock normal such that the upstream angle between the shock normal and the magnetic field θ_1 is larger than the downstream angle θ_2 .

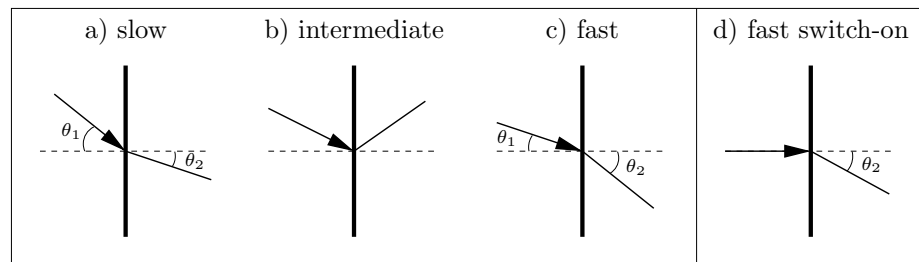


Figure 7.1: *MHD shock types. The thick vertical line represents the shock front, the dotted line is the shock normal. The arrowed lines denote magnetic field lines that are refracted through the shock surface. (From [119]).*

- In *intermediate* MHD shocks (Fig. 7.1b), the magnetic field line is reflected by the shock normal.
- In *fast* MHD shocks (Fig. 7.1c), the magnetic field is refracted away from the shock normal such that the upstream angle between the shock normal and the magnetic field θ_1 is smaller than the downstream angle θ_2 .

Fast *switch-on* MHD shocks (Fig. 7.1d) are characterized by an upstream magnetic field that is perpendicular to the shock front, but downstream there exists an angle θ_2 between the shock normal and the magnetic field. Switch-on shocks arise when the magnetic field is strong so that magnetic forces dominate over thermal pressure. This occurs if the following two conditions are satisfied [84].

1. First, the *plasma* β , which is defined as the ratio of the thermal pressure over the magnetic pressure,

$$\beta := \frac{2p}{\|B\|^2},$$

must satisfy $\beta < 2/\gamma$.

2. Secondly, the incoming plasma velocity must lie between the fast MHD wave speed and roughly twice this speed. This is equivalent to the condition

$$1 < M_{Ax} < \sqrt{\frac{\gamma(1-\beta)+1}{\gamma-1}},$$

where M_{Ax} denotes the Alfvénic Mach number in the upstream direction, defined by $M_{Ax} := |v_x|/c_{Ax}$, with v_x and c_{Ax} the plasma velocity and the Alfvén speed in the upstream direction, respectively. The Alfvén speed is given by $c_{Ax} = |B_x|/\sqrt{\rho}$.

An upstream flow for which switch-on shocks occur is called a *magnetically dominated* flow. If a switch-on shock does not occur, the flow is referred to as *pressure dominated* [27].

7.3.2 Two-dimensional MHD flow around a cylinder

The shock conditions above can be derived from the Rankine-Hugoniot jump equations across a shock [84, 118]. Analytical solutions of the Rankine-Hugoniot equations are mostly impossible to obtain, so that numerical simulations are necessary to study the shock topology.

De Sterck et al. [27, 28, 119] considered numerical simulations of the bow shock that is formed when plasma flows around a perfectly conducting cylinder at supersonic speed. It was observed that, as long as the upstream flow is pressure dominated, MHD shocks exhibit a topology similar to hydrodynamic

shocks. As shown in Fig. 7.2a, a bow shock with a single front is formed that is entirely of the fast type. However, numerical simulations have revealed that MHD bow shocks exhibit an entirely different topology when the flow is magnetically dominated. As shown in Figs. 7.2b and 7.3, several consecutive shock fronts of various shock types are formed among which intermediate shocks and fast switch-on shocks.

The simulation results presented in Figs. 7.2 and 7.3 were generated using the VAC on a cylindrical 124×124 grid (taking the *ghost* cells for the boundary conditions into account). Top-bottom symmetry is exploited so that only the upper part of Figs. 7.2a and 7.2b is actually simulated. Part of the simulation grid is shown in Fig. 7.3. Notice that the grid is stretched so that resolution is highest in the region of the consecutive shock fronts. The parameters of the incoming plasma flow (which are assumed to be uniform over the boundary) are chosen so that $\beta = 0.4$. In Fig. 7.2a, the Alfvénic Mach number is chosen as $M_{Ax} = 2$. In Fig. 7.2b, it is chosen as $M_{Ax} = 1.5$, so that both conditions for a switch-on shock are satisfied.

The numerical simulation considered above provides a simple two-dimensional model for the bow shock that is formed when the solar wind encounters the Earth. In reality, the Earth is of course not a perfectly conducting cylinder, but rather a sphere that produces its own magnetic field which is compressed by the solar wind at the day-side and expanded at the night-side. Consequently, the real bow shock is more complex than in the simulation above. However, the simulations of De Sterck et al. are considered to give a quite accurate picture of the features involved in the topology of the Earth's bow shock. Observations from satellites indicate that the solar wind at the Earth is approximately 8% of the time in the switch-on regime.

7.4 Data assimilation for two-dimensional MHD flow around a cylinder

In this section, we consider data assimilation for 2D MHD flow around a perfectly conducting cylinder. Using the numerical technique of De Sterck et al., we set up a large scale simulation with the RRSLSQRT filter developed in Sect. 6.5.2. The performance of the RRSLSQRT filter is studied under various conditions.

This section is outlined as follows. Section 7.4.1 describes the setup of the simulations. Next, in Sect. 7.4.2, we consider data assimilation under known, constant boundary conditions and study the ability of the RRSLSQRT to reconstruct the true state, starting from an incorrect initial estimate. Next, in Sect. 7.4.3, it is assumed that the boundary condition is unknown and in addition also time varying. More precisely, we consider the case where the incoming plasma flow changes from pressure dominated to magnetically dominated. The RRSLSQRT will be extended such that it simultaneously estimates the boundary condition and the system state and its stability under

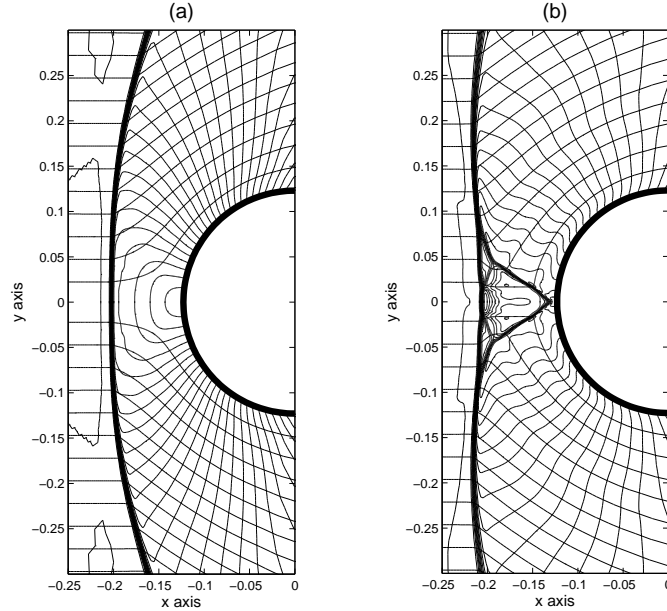


Figure 7.2: Bow shock topology in 2D MHD flow around a perfectly conducting cylinder. Density contours (pilled up in the shocks) and magnetic field lines (coming in horizontally from the left) are shown. (a) Pressure dominated flow: $M_{Ax} = 2, \beta = 0.4$. (b) Magnetically dominated flow: $M_{Ax} = 1.5, \beta = 0.4$. (From [27]).

such conditions will be assessed.

7.4.1 Setup of the simulations

The setup of the numerical simulations is schematically shown in Fig. 7.4. All MHD simulations are performed with the VAC. The equations of the RRSLSQRT filter, on the other hand, are implemented in Matlab. This requires a constant conversion and exchange of data between Matlab and VAC, which unfortunately leads to a high computational overload.

The numerical grid used in the VAC is chosen as described in the previous section, i.e. a stretched cylindrical grid consisting of 124×124 cells. Each cell contains six variables (one for the density ρ , two for the momentum density ρv , two for the magnetic field B and one for the energy e), resulting in a state dimension of $n = 92256$. The actual values of the variables used during the simulations do not have any physical meaning. The nominal values of the incoming variables, i.e. the variables at the left boundary, are chosen in the order of one.

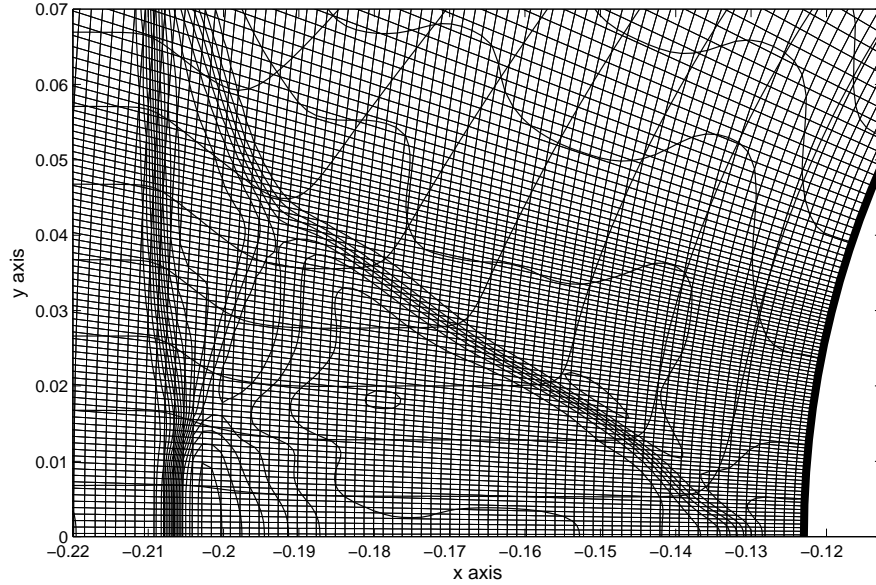


Figure 7.3: *Detail of the stretched simulation grid. Density contours and magnetic field lines of the magnetically dominated flow presented in Fig. 7.2b are shown. Several consecutive shock fronts of various shock types are formed among which intermediate shocks and fast switch-on shocks.*

In order to study the performance of the RRSLSQRT filter, we use the procedure of twin experiments. This means that we first artificially create “measurements” by simulating the plasma flow using the VAC. During the simulation, random noise sampled from a normal distribution with mean zero and variance 10^{-8} is added to the state variables every $5 \cdot 10^{-4}$ s. The process noise covariance matrix is chosen of rank 250. Furthermore, random noise sampled from a normal distribution with mean zero and variance 10^{-6} is added to the measurements. It is assumed that measurements are available every $5 \cdot 10^{-4}$ s.

In the second step of the twin procedure, the RRSLSQRT filter is initialized with an estimate $\hat{x}_{[0|-1]}$ of the initial state $x_{[0]}$ that was used in the simulation and with a square-root $S_{[0|-1]} \in \mathbb{R}^{n \times q}$ that is obtained from an ensemble of initial state estimates. The RRSLSQRT filter is then employed to recursively estimate the actual state, i.e. the state vector simulated during the first step of the twin procedure, based on the measurements generated in the first step.

The measurement update of the RRSLSQRT filter is implemented entirely in Matlab on a single processor. The nonlinear nature of the numerical model is dealt with during the time update using the extension of the RRSQRT filter considered in Sect. 6.4.1.4. Such an extension requires during each time update

$q + 1$ simulations with the VAC. The simulations exchange no information and can therefore be implemented in parallel. The simulation results for the highest values of q presented in the next sections were obtained using parallel computations on the K.U.Leuven VIC cluster [1], which at the time of the simulations consisted of 876 processors and had a peak performance of approximately 4 teraflops/s. For the lowest values of q , however, it was found that the conversion of data formats between Matlab and VAC leads to such a high computational overload that sequential simulations on a single processor are almost as fast as a parallel computation.

The dash-dot line from the boundary condition to the time update in Fig. 7.4 denotes that in some simulations the left boundary condition, i.e. the properties of the incoming plasma flow, will be assumed to be known, while in others it will be assumed to be unknown and thus estimated by the filter.

7.4.2 Known, constant boundary conditions

In a first series of experiments, we consider the boundary conditions at the left hand side of the grid, i.e. the properties of the incoming plasma flow, to be known and constant. The ability of the RRSLSQRT to converge to the actual state, starting from an arbitrary initial estimate, will be tested under various conditions. The main objective is to assess the influence of the number of measurements, the type of spatial localization and the rank q of the approximate error covariance matrix on the speed of convergence.

The plasma β and the Alfvénic Mach number of the incoming plasma flow are chosen as in Fig. 7.2b, that is, so that the flow is magnetically dominated. The initial state estimate $\hat{x}_{[0|-1]}$, which is chosen exactly the same in all experiments, is constructed by adding a random perturbation to the flow pattern shown in Fig. 7.2b. The error in the initial state estimate can be seen in Fig. 7.5 for $k = 1$.

In a first experiment, we carry out simulations without spatial localization. The RRSLSQRT filter then reduces to the RRSQRT filter. The error in the density estimates of the RRSQRT filter as function of q and of the simulation step k is shown in Fig. 7.5. It was assumed that measurements are available at 8 locations. These locations are indicated by the crosses. Density contours of the actual flow are also shown. Notice that the estimation error decreases with k , which means that the filter is converging. Also, notice that for $k = 1000$, the estimation error is largest in the region of the consecutive shock fronts.

In a second experiment, simulations with spatial localization are performed. The localization patterns that determine the matrices $\Pi_{[k]}$ in the RRSLSQRT ($\Pi_{[k]}$ is chosen differently for each of the 8 measurements) are shown in Fig. 7.7. Locations where measurements are available are denoted by stars. The box around each of the stars denotes the region over which this particular measurement has an influence on the grid cells. The color intensity of a grid cell denotes the number of measurements that influence this particular grid cell. The MSE in the estimates of the RRSLSQRT is shown in Fig. 7.6 as function of the type of spatial localization and the rank approximation q of the error covariance

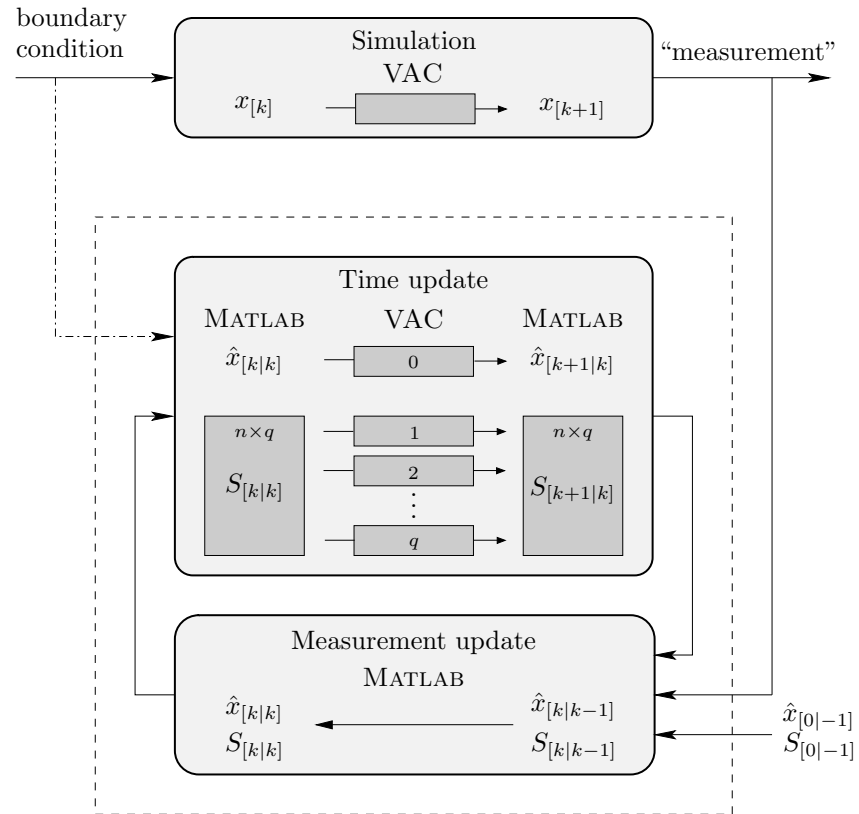


Figure 7.4: Setup of the numerical simulations with the RRSLSQRT. The “measurements” are generated by simulating the plasma flow with the VAC and adding noise. The measurement update is implemented in Matlab on a single processor. The time update requires $q + 1$ simulations with the VAC code. The simulations are implemented in parallel on the K.U.Leuven VIC cluster [1]. The dash-dot line from the boundary condition to the time update denotes that in some simulations the properties of the incoming plasma flow will be assumed unknown.

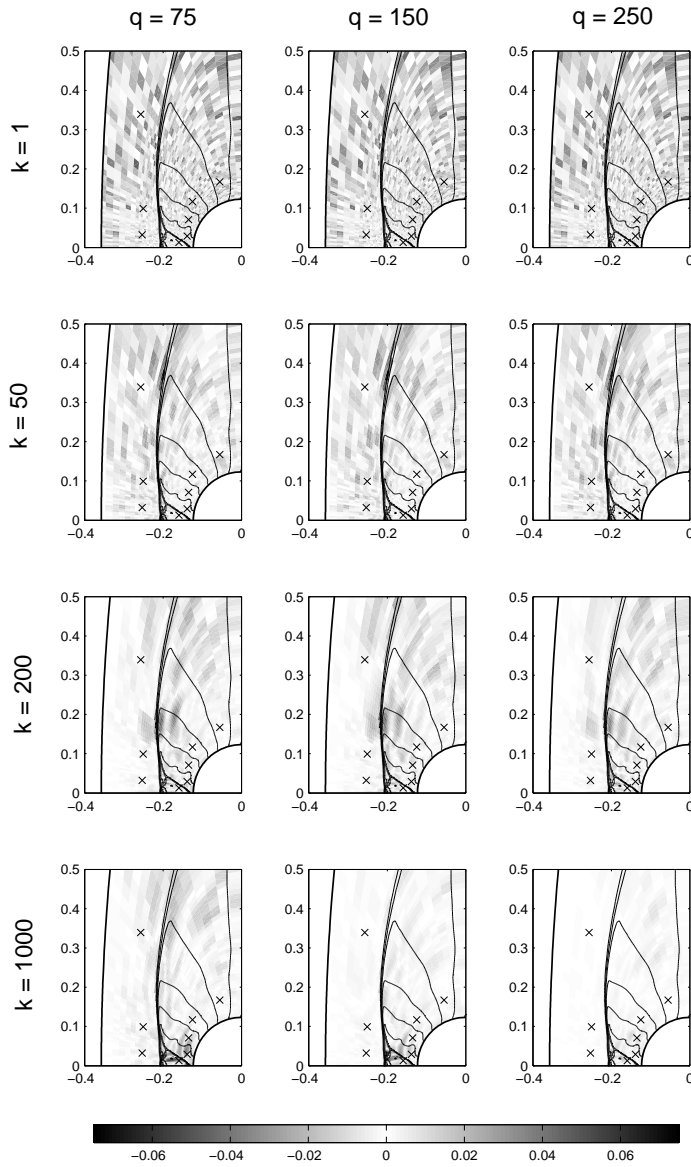


Figure 7.5: Error in the density estimates of the RRSQRT filter as function of the rank q of the error covariance matrix and of the simulation step k . The locations where measurements are available, are denoted by the crosses. Density contours of the actual flow, which is magnetically dominated, are shown. Notice that the estimation error for $k = 1000$ is largest in the region of the consecutive shock fronts.

matrix. As expected, the MSE decreases with increasing q . Spatial localization of the measurements clearly has a positive effect on estimation accuracy. Notice that even for q as low as 75, the RRSLSQRT filter with localization of type B performs more than satisfactory. It clearly outperforms the RRSQRT filter with $q = 250$.

In a third experiment, simulations with a reduced number of 4 measurements are carried out. The localization patterns used in the experiment are shown in Fig. 7.8. Figure 7.9 shows the MSE in the estimates as function of the type of spatial localization and the rank q of the approximate error covariance matrix. The results are very similar to those in the experiments with 8 measurements (Fig. 7.6). In particular, it is found that the filters without spatial localization perform almost identical to a data free simulation. The results for spatial localization, however, clearly outperform a data free simulation. As in Fig. 7.6, localization of type B is more accurate than type A.

Both the experiments with 8 and 4 measurements suggest that localization of type B is to be preferred over type A, meaning that the regions over which the measurements influence the grid cells should be taken quite small. Evidence of this is obtained by considering the correlation information contained in the approximate error covariance matrix of the RRSLSQRT. Figure 7.10 plots the absolute value of the correlation between the density at the grid cell indicated by the star (at which a measurement was assumed to be taken during the simulations) and all other cells computed from the approximate error covariance matrix of the RRSLSQRT for $q = 250$ and localization of type A. As expected, correlation drops relatively quickly with distance. However, even for grid points greatly distant, correlation is certainly not exactly zero. Notice that the correlations at the left of the bow shock and in the upper right corner are rather noisy. These correlations are probably spurious and due to the approximation. In Example 6.2, it was shown that such spurious correlations may increase the variance of the corresponding grid cell. This may explain the superior performance of localization type B (which confines the left boundary of the localization region closer to the bow shock) over type A. Also, notice that the correlation inside the region of the consecutive shocks is significantly different from zero and not noisy. This may indicate that extending the localization region in that direction may further increase estimation accuracy.

Table 7.1 compares the relative processing times for the measurement update in the RRSLSQRT as function of the type of localization and the rank q of the approximate error covariance matrix. Results for the simulations with 8 measurements are presented. Contrary to Table 6.2 where the results of a relatively small scale example were presented, the RRSQRT is now most time consuming. The procedure of updating only a selected number of rows has now turned into a computational advantage.

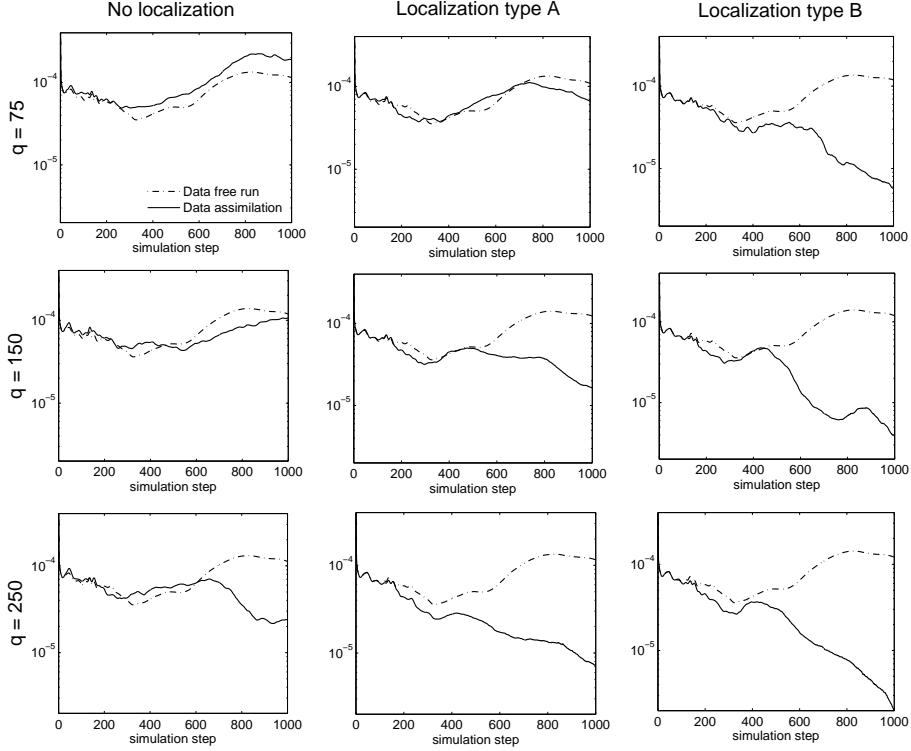


Figure 7.6: MSE in the estimates of the RRSLSQRT for the simulations with 8 measurements. The MSE is plotted as function of the type of spatial localization and the rank q of the approximate error covariance matrix. As expected, the MSE decreases with increasing q . Spatial localization of the measurements clearly has a positive effect on estimation accuracy.

Algorithm	q		
	75	150	250
RRSQRT	5.67	11.33	18.96
RRSLSQRT Loc. A	1.89	4.41	10.44
RRSLSQRT Loc. B	1	2.22	4.74

Table 7.1: Relative processing times for the measurement update in the RRSLSQRT. The actual processing time for $q = 75$ and localization of type B was 2.7 s. For comparison, the time update under the same circumstances took approximately 115 s.

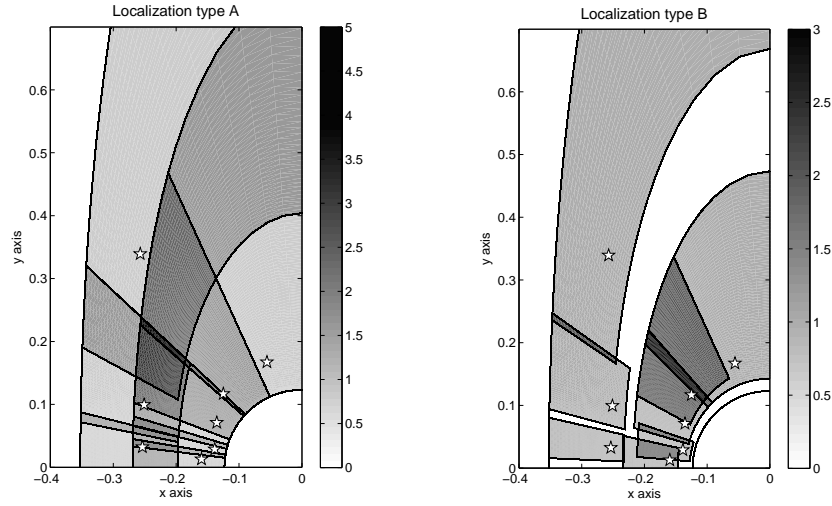


Figure 7.7: Localization patterns used in the simulations with 8 measurements. The pattern determines the matrix $\Pi_{[k]}$ in the RRSLSQRT, which is chosen differently for each of the 8 measurements. Locations where measurements are available are denoted by stars. The box around each of the stars denotes the region over which this particular measurement has an influence on the grid cells. The color intensity of a grid cell denotes the number of measurements that influence this particular grid cell.

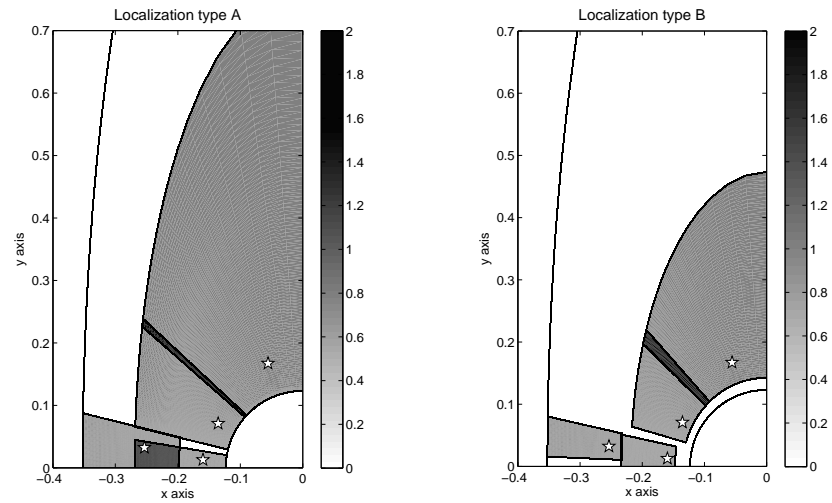


Figure 7.8: Localization patterns for the simulations with 4 measurements.

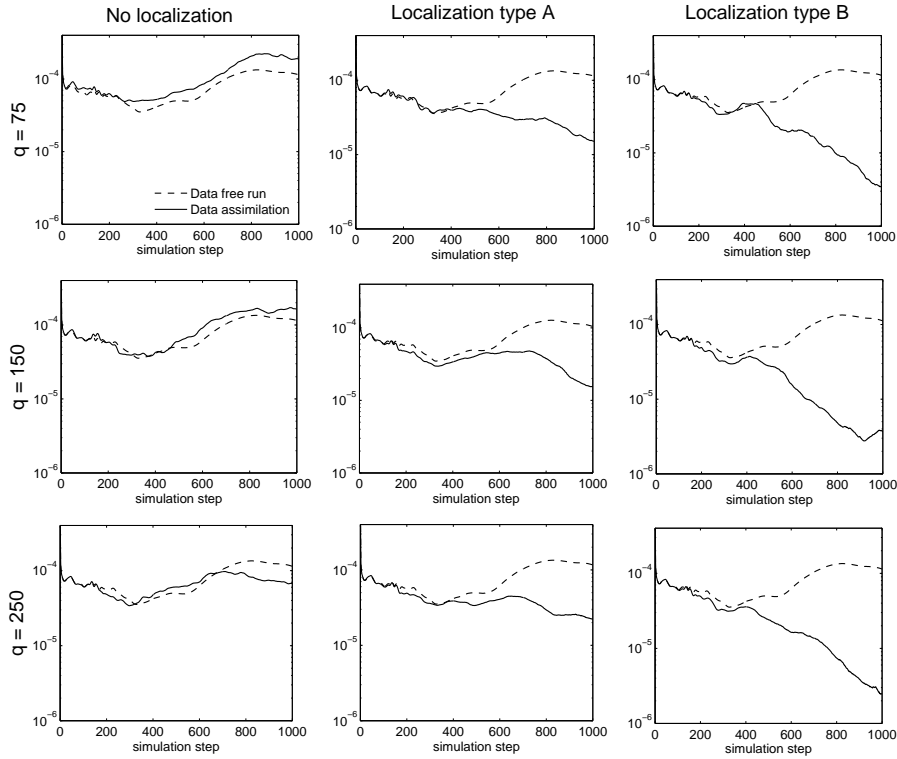


Figure 7.9: MSE in the estimates of the RRSLSQRT for the simulations with 4 measurements. The MSE is plotted as function of the type of spatial localization and the rank q of the approximate error covariance matrix.

7.4.3 Unknown, time-varying boundary conditions

In a second series of experiments, we consider the properties of the incoming plasma flow to be unknown and in addition also time-varying. More precisely, experiments are conducted in which the flow changes from pressure dominated to magnetically dominated. The RRSLSQRT is extended to estimate the unknown boundary condition based on the measurements at the left hand side of the bow shock, which observe changes in the boundary condition with a relatively short time delay. The extension is based on the use of a separate Kalman filter which estimates the unknown boundary condition based on the dynamical model $u_{[k+1]} = u_{[k]} + \eta_{[k]}$, where $u_{[k]}$ denotes the boundary condition at the discrete time instant k and $\eta_{[k]}$ denotes a zero-mean random vector. The variance of $\eta_{[k]}$ is a design parameter. A low variance yields optimal performance if the boundary condition is constant in time. A high variance, on the other hand, yields more noisy estimates for a constant boundary condition, but can better

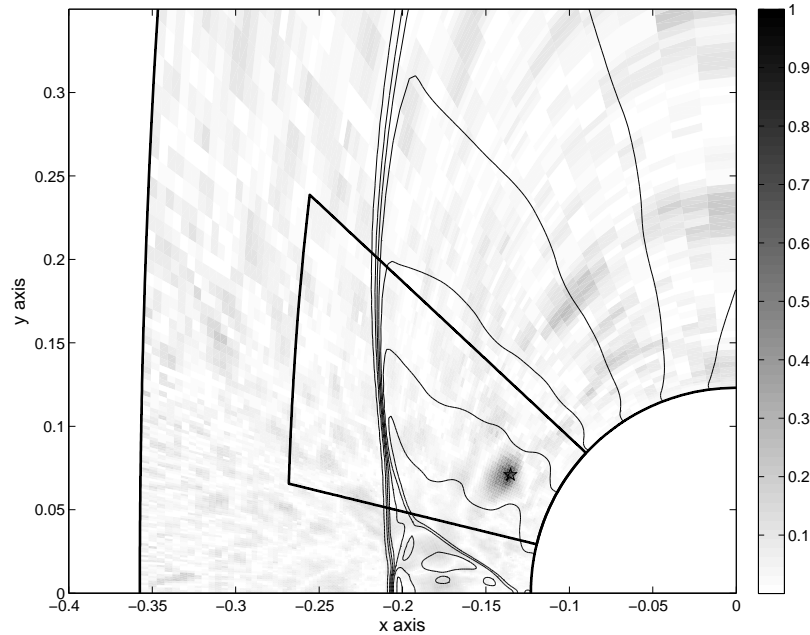


Figure 7.10: Absolute value of the correlation between the density at the grid cell indicated by the star (at which a measurement was assumed to be taken during the simulations) and all other cells. The correlation is computed from the approximate error covariance matrix of the RRSLSQRT for $q = 250$ and localization of type A. The thick lines mark the region in which the grid cells are updated by the measurement.

track a time-varying boundary condition. The estimates obtained with the Kalman filter are then employed as boundary condition in the VAC.

Figure 7.11 compares the true and estimated value of the momentum density of the plasma flow coming in at the left boundary. As expected, the change in the boundary condition is reconstructed only with a certain time delay. Notice the overshoot in the estimate around simulation step 350.

The time evolution of the MSE in the state estimates is compared in Fig. 7.12 for a data free simulation, an RRSLSQRT where the boundary condition is estimated from the three measurements at the left of the bow shock and an RRSLSQRT where the boundary condition is known up to a small additive noise term. The results of the RRSLSQRT were obtained for $q = 250$ and covariance localization type B. Although the RRSLSQRT extended with boundary condition estimation performs significantly better than a data free simulation, the estimation error increases very rapidly between simulation step 200 and 400, which is due to the delay and overshoot in the estimate of the boundary condition. The RRSLSQRT in which the boundary condition is known

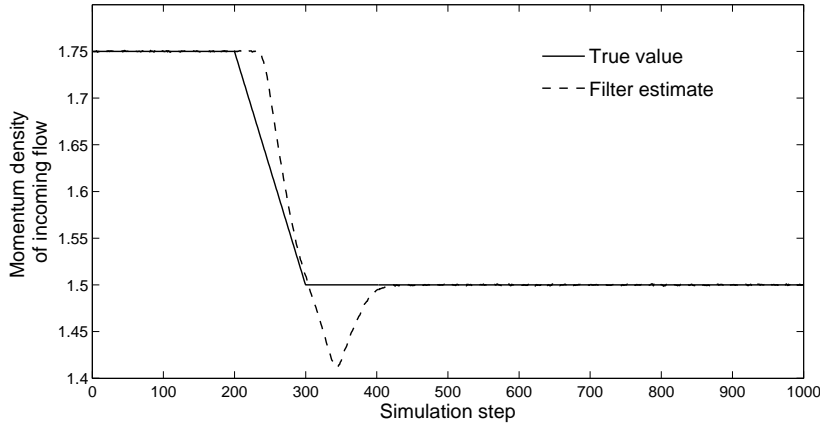


Figure 7.11: Comparison between true and estimated value of the momentum density of the plasma flow coming in at the left boundary. The estimates are obtained using the three measurements at the left of the bow shock.

up to an additive noise term, on the other hand, is relatively insensitive to the change in the properties of the incoming plasma flow and performs orders of magnitude better. This result indicates that the simulations are quite sensitive to the specification of the boundary condition and that estimation accuracy can be significantly increased if accurate measurements or estimates of the incoming plasma flow are available.

Figure 7.13 shows the estimation error of the RRSLQRT ($q = 250$) extended with boundary condition estimation as function of time and of the type of spatial localization. Density contours of the plasma flow are shown. Notice that at time instant $k = 300$, the error is highest left of the shock front, at $k = 500$ right of the front, and at $k = 700$ in the region where the consecutive shock fronts have just been formed.

7.5 Conclusion

This chapter has considered the use of the RRSLQRT filter, developed in Chapter 6, for data assimilation in a space weather application. Based on the numerical results presented in [27], a simulation was set-up that emulates the topology of the bow shock that is formed when the solar wind encounters the Earth. The numerical model consists of approximately 10^5 grid cells. Two series of experiments were conducted.

In the first series, the boundary condition was assumed to be known and constant and the performance of the RRSLQRT was investigated under various conditions. It was found that even for as few as 4 measurements, the

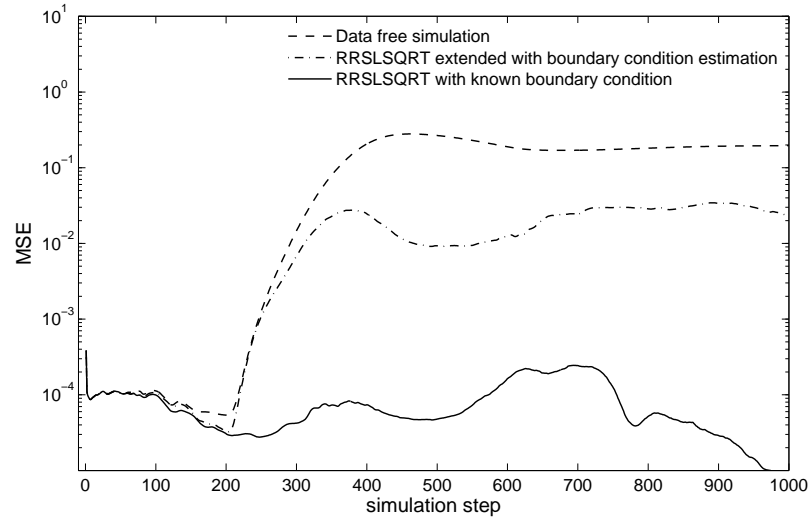


Figure 7.12: Comparison in the MSE between a data free simulation, an RRSLSQRT where the boundary condition is estimated from the three measurements at the left of the bow shock and an RRSLSQRT where the boundary condition is known up to a small additive noise term.

RRSLSQRT may yield a significant reduction in estimation error over a data-free simulation. Results also indicate that spatial localization of measurements has a positive effect on estimation accuracy.

In the second series, the boundary condition was assumed to be unknown and in addition also time-varying. More precisely, experiments were conducted in which the flow changes from pressure dominated to magnetically dominated and in which the bow shock consequently undergoes a complete change in topology. The RRSLSQRT was extended to simultaneously estimate the time-varying boundary condition and the system state. It was found that the extended RRSLSQRT is robust against changes in the boundary condition. Results indicate that estimation accuracy is strongly dependent on the quality of the boundary condition estimates.

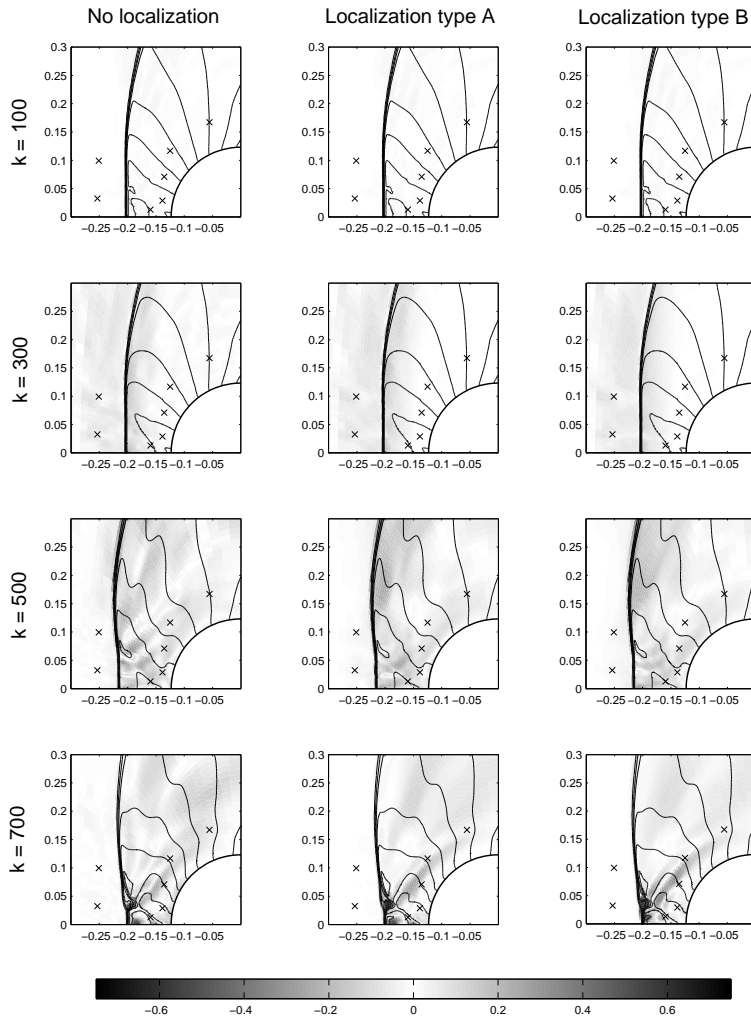


Figure 7.13: Estimation error of the RRSLSQRT ($q = 250$) extended with boundary condition estimation as function of time and of the type of spatial localization. Density contours of the actual plasma flow, which changes from pressure dominated to magnetically dominated, are also shown.

Chapter 8

Conclusions and directions for further research

This chapter summarizes the most important results obtained in this thesis and suggests some directions for further research.

8.1 Conclusions

The first part of this thesis has addressed the problem of inverting linear dynamical systems. In Chapter 3, a new and straightforward inversion procedure for deterministic system was introduced. This procedure was extended in Chapter 4 to combined deterministic-stochastic systems. Finally, in Chapter 5, four applications of system inversion were considered. The main conclusions of these three chapters are now discussed.

- **Chapter 3**

Based on estimation theory, a new procedure for left inversion of linear deterministic systems was introduced. Like the approach of Sain and Massey [115], the left inverses considered in this thesis consist of a bank of delay elements followed by a dynamical system. In this thesis, the most general form of such a dynamical system was derived. This dynamical system reconstructs both the system input and the system state and can thus be considered as a joint input-state estimator. The general form consists of two matrix parameters which can be free chosen. It was shown that the poles of the inverse system can be assigned by tuning one of these parameters if a certain matrix pair is observable. This pair turns out to be observable if the system has no zeros. Based on the theory of reduced order observers, a new technique was developed to simultaneously reduce the order of the inverse system and place its poles. The results of this chapter not only generalize existing methods for left inversion, but also have direct implications for optimal state estimation in the presence

of unknown inputs. It is expected that most results translate easily to continuous-time systems and to the dual problem of right inversion.

- **Chapter 4**

The inversion procedure developed in Chapter 3 was extended to combined deterministic-stochastic systems by determining the two matrix parameters so that the estimates of the system state and the input are optimal in the minimum-variance unbiased sense. Although optimal state estimation in the presence of both unknown inputs and noise has received a lot of attention in the past, this is the first extension to joint-input state estimation. The estimator yields a general framework for the one step ahead prediction, the filtering and the smoothing problem. Another important contribution is the establishment of a relation between the joint input-state estimator and least-squares estimation. Although not explicitly proven, this relation suggests that the joint input-state estimator is optimal in a the least-squares sense. Based on the relation, information and square-root information formulas were derived almost instantaneously. Finally, it was shown that square-root covariance filtering in the presence of unknown inputs is not possible. A numerical example has indicated that the error covariance matrices of the joint input-state estimator converge. Further research should establish convergence conditions.

- **Chapter 5**

Four applications of system inversion were considered. First, the problem of optimal filtering with noisy input and output measurements was addressed. Recursive filter equations were derived in which the estimation of the system state and the input are interconnected. As a special case, the filter provides a new solution to the errors-in-variables filtering problem, which was shown to be algebraically equivalent to existing techniques. Next, the problem of filtering in the presence of bias errors was considered. A suboptimal filter, closely related to the two-stage Kalman filter [45], was developed. The major difference is that the new filter can be used also if the equation governing the dynamical evolution of the bias error is unknown. A simulation example has shown that such an approach is especially useful if the bias error is constant for a certain period of time and then suddenly undergoes an abrupt and unknown change. The last two applications are more practical. First, model error estimation and dynamic model updating was addressed. An empirical technique was outlined to update a non-satisfactory accurate physical state-space model. The technique consists in first estimating the model error and then identifying an empirical correction model based on the estimated data. Finally, an approach to joint state and boundary condition estimation was considered in which the temporal component of the boundary is completely unknown and the spatial form is expanded as a linear combination of orthogonal basis functions. It was shown that such an expansion strongly simplifies the existence conditions of the joint input-state estimator. Simulation results on a linear heat conduction model

indicate that measurements in the close neighborhood of the boundary are needed in order to accurately reconstruct the unknown boundary condition.

The second part of this thesis has addressed the problem of suboptimal square-root filtering for large-scale numerical models obtained by discretizing partial differential equations over a huge spatial grid. In Chapter 6, a spatially localized variant of the reduced rank square-root (RRSQRT) filter [135] was developed that is extremely efficient if only few measurements are available. In Chapter 7, this variant was used in a simulation to study the effectiveness of data assimilation for the estimation of the bow shock that is formed when the solar wind encounters the Earth. The main conclusions of both chapters are now discussed.

- **Chapter 6**

Two extensions of the RRSQRT filter were developed. The first extension speeds up the RRSQRT filter by interweaving the so-called reduction step into the measurement update. The reduction in time is approximately that of the reduction step. However, a numerical example has shown that the resulting filter is more vulnerable to divergence than the RRSQRT because the error covariance matrix is more underestimated. The second extension has addressed the problem of reduced rank spatially localized square-root (RRSLSQRT) filtering, where the objective is to update only a subset of the grid cells by the measurement. The development of such an extension is motivated by the lower rank approximation which, as was shown in a theoretical study, may cause an increase in the actual variance of grid cells that are greatly distant from the cell at which the measurement is taken. Two variants of the RRSLSQRT were developed. The first variant is equivalent to the spatially localized Kalman filter [9] if no lower rank approximation of the error covariance matrix is made. The second variant is based on the assumption that correlation between grid cells drops to zero within a distance of a relatively low number of grid cells. Although being more approximate, this variant turns out to be extremely efficient, especially if only few measurements are available. A simulation example on a nonlinear chaotic model has indicated that the RRSLSQRT is more robust to filter divergence than the traditional RRSQRT filter, especially if the error covariance matrix is approximated by one of very low rank.

- **Chapter 7**

The RRSLSQRT is applied with a large-scale model, consisting of approximately 10^5 state variables, which emulates the dynamics of the bow shock that is formed when the solar wind encounters the Earth. Two series of simulations were performed. The first series assumes that boundary conditions are known and investigates the performance of the RRSLSQRT under various conditions. It was found that even for as few as 4 measurements, the RRSLSQRT may yield a significant reduction in estimation error over a data-free simulation. Results also indicate that

spatial localization of measurements has a positive effect on estimation accuracy. In the second series, the boundary condition was assumed to be unknown and in addition also time-varying. More precisely, experiments were conducted in which the bow shock undergoes a complete change in topology. The RRSLSQRT was extended to simultaneously estimate the time-varying boundary condition and the system state. It was found that the extended RRSLSQRT is robust against changes in the boundary condition. Results indicate that estimation accuracy is strongly dependent on the quality of the boundary condition estimates.

8.2 Directions for further research

In this section, some directions for further research are given. In Sect. 8.2.1, we consider directions concerning system inversion. Directions concerning data assimilation are considered in Sect. 8.2.2.

8.2.1 System inversion

A lot of open questions remain in the inversion of linear systems remain. The main problems are now briefly discussed.

- Due to the duality between left and right inversion, it is expected that most results derived in this thesis for left inversion translate easily to right inversion. It can, however, be interesting to study the problem of right inversion from the viewpoint of estimation theory. To set the idea, consider the LTI discrete-time system

$$x_{[k+1]} = Ax_{[k]} + Bu_{[k]} \quad (8.1a)$$

$$y_{[k]} = Cx_{[k]} + Du_{[k]}, \quad (8.1b)$$

with $x_{[k]} \in \mathbb{R}^n$ the state vector, $y_{[k]} \in \mathbb{R}^p$ the output vector and $u_{[k]} \in \mathbb{R}^m$ the vector of inputs. Assume that the system

$$\bar{x}_{[k+1]} = \bar{A}\bar{x}_{[k]} + \bar{B}\bar{y}_{[k]} \quad (8.2a)$$

$$u_{[k]} = \bar{C}\bar{x}_{[k]} + D\bar{y}_{[k]} \quad (8.2b)$$

is an instantaneous right inverse of (8.1). Assuming that $\bar{x}_{[k]} = x_{[k]}$, it is sufficient to require that $y_{[k]} = \bar{y}_{[k]}$ and $x_{[k+1]} = \bar{x}_{[k+1]}$.

- It follows from (8.1) and (8.2) that

$$y_{[k]} - \bar{y}_{[k]} = (C + D\bar{C})x_{[k]} + (D\bar{D} - I)\bar{y}_{[k]}.$$

Consequently, $y_{[k]} = \bar{y}_{[k]}$ if

$$\begin{cases} C + D\bar{C} = 0 \\ I - D\bar{D} = 0. \end{cases}$$

Solutions to the latter equations exist if $\text{rank}(D) = p$, in which case the general solutions are given by

$$\begin{aligned}\bar{C} &= D^{(1)}C + (I - D^{(1)}D)Z \\ \bar{D} &= D^{(1)} + (I - D^{(1)}D)U,\end{aligned}$$

where Z and U are matrix parameters that can be freely chosen.

– Furthermore, it follows from (8.1) and (8.2) that

$$x_{[k+1]} - \bar{x}_{[k+1]} = (C + D\bar{C})x_{[k]} + (D\bar{D} - I)\bar{y}_{[k]}.$$

Consequently, $x_{[k+1]} = \bar{x}_{[k+1]}$ if

$$\begin{cases} A + B\bar{C} - \bar{A} = 0 \\ B\bar{D} - \bar{B} = 0, \end{cases}$$

which yields

$$\begin{aligned}\bar{A} &= A + B\bar{C} \\ \bar{B} &= B\bar{D}.\end{aligned}$$

The procedure above yields a general form of an instantaneous right inverse of (8.1). An important research question is whether this technique can be extended to the concept of L -delay right inversion.

- Theorem 3.5 provides a general form of an L -delay left inverse of a system. However, how general is this form exactly? In other words, which subset of the set of all possible L -delay left inverses does it yield?
- The numerical examples in Sect. 4.5 indicate that the joint input-state estimators developed in Chapter 4 converge. Further research should investigate conditions under which convergence occurs. How do the algebraic equations from which we can compute the steady-state gain matrices look like? Are they, like in the case of Kalman filtering, algebraic Riccati equations? If not, what are the properties of the equations and how do we solve them?
- A necessary condition for left invertibility is that the system has more inputs than outputs. If this is not the case, the system input can not be uniquely reconstructed from knowledge of the output. The least-squares problem (4.34), for example, is then underdetermined. An underdetermined LS problem of the form

$$\begin{aligned}\min \quad & \|v\|^2 \\ \text{s.t.} \quad & y = Ax + v,\end{aligned}$$

where y is given and x has to be deduced, has infinitely many solutions. However, by considering the problem

$$\begin{aligned}\min \quad & \|v\|^2 + \|x\|^2 \\ \text{s.t.} \quad & y = Ax + v,\end{aligned}$$

that is, by imposing that the norm of the solution must be minimal, the LS solution becomes again unique. In case the system of which the input has to be estimated has more inputs than outputs, it can thus be convenient to consider a minimum-norm extension of (4.34). An interesting research problem is then whether this minimum-norm solution can be computed recursively.

- Is there a notion of ill-conditioning in system inversion, just as there is a notion of ill-conditioning in the inversion of a matrix? To set the idea, let A be a square $n \times n$ matrix that is ill-conditioned with respect to inversion. Special attention must then be paid to the numerical solution of an equation $y = Ax$ where y is given and x has to be deduced. The notion of ill-conditioning in system inversion is most easily understood in terms of the transfer function $H(z)$. Since $Y(z) = H(z)U(z)$, with $U(z)$ and $Y(z)$ the z -transforms of the system input and output, system inversion basically comes down to inverting the transfer function $H(z)$ (assuming the latter is square). Can we define a condition number in terms of the transfer function that tells us how the conditioning of certain inversion problem is?

8.2.2 Data assimilation

Although the results of Chapter 7 are very promising, we are still far from a real-life implementation in which the RRSLSQRT, or another data assimilation technique, is used to estimate the topology of the Earth's bow shock in real-time. Such an implementation requires further research in which the following problems are addressed.

- The numerical model used in the experiments is still relatively simple. First of all, the model is two-dimensional, whereas the actual bowshock is three-dimensional. The extension to three-dimensional simulations will probably require the development of advanced data assimilation techniques in which parallel computation and code optimization are carefully studied. Secondly, the model neglects several important phenomena such as the influence of the Earth's magnetic field.
- The simulations considered in this thesis assume that the measurements are taken at fixed locations. Simulations should be performed in which the actual orbits of the satellites are taken into account. In addition, simulations should be considered in which the time interval between the availability of two consecutive measurements is realistic.
- The performance and stability of data assimilation techniques should be tested under more various conditions. For example, the influence of the magnitude of the measurement noise and process noise on the stability must be assessed.

-
- Placing a satellite in an orbit is very costly. The orbit of a new satellite should thus be carefully chosen. This motivates the development of methods that determine the optimal location for a supplementary satellite. A preliminary study in the context of Earthly weather models can be found in e.g. [91].
 - The data assimilation techniques developed in this thesis are aimed to estimate the effects of the solar wind close to the Earth. Considering space weather forecasts, it can be more interesting to estimate the initial conditions at the origin of the solar wind, i.e. at the Sun. These estimates may then serve as the initial conditions of a space weather forecast.
 - Finally, simulations with real measurements observed by satellites should be considered.

Appendix A

Generalized inverses and the matrix inversion lemma

A.1 Generalized inverses

Definition A.1. Let $A \in \mathbb{R}^{m \times n}$. Then, $A^{(1)} \in \mathbb{R}^{n \times m}$ is said to be a $\{1\}$ -inverse of A if $AA^{(1)}A = A$.

Definition A.2. Let $A \in \mathbb{R}^{m \times n}$. Then, the Moore-Penrose generalized inverse $A^\dagger \in \mathbb{R}^{n \times m}$ of A is the unique matrix satisfying $AA^\dagger A = A$, $A^\dagger AA^\dagger = A^\dagger$, $(AA^\dagger)^\top = AA^\dagger$, $(A^\dagger A)^\top = A^\dagger A$.

Generalized inverses play a fundamental role in the solution of linear matrix equations. Consider the equation $Y = CX$, where $Y \in \mathbb{R}^{p \times m}$ and $C \in \mathbb{R}^{p \times n}$ are known matrices, and $X \in \mathbb{R}^{n \times m}$ has to be deduced. The following theorem characterizes the solutions in terms of a $\{1\}$ -inverse of C .

Theorem A.1. Let $Y \in \mathbb{R}^{p \times m}$ and $C \in \mathbb{R}^{p \times n}$ be known. Then, there exists a matrix $X \in \mathbb{R}^{n \times m}$ satisfying $Y = CX$ if and only if $\text{rank } C = \text{rank } \begin{bmatrix} C & Y \end{bmatrix}$. The general solution for X is given by $X = C^{(1)}Y + (I - C^{(1)}C)Z$, where $Z \in \mathbb{R}^{n \times m}$ is an arbitrary matrix.

Lemma A.1. Let $A \in \mathbb{R}^{n \times m}$, then $\text{rank } (I_n - AA^{(1)}) = n - \text{rank } A$.

Proof: First, note that $AA^{(1)}$ is idempotent. Consequently, the rank of $I_n - AA^{(1)}$ is given by

$$\begin{aligned} \text{rank } (I_n - AA^{(1)}) &= n - \text{rank } (AA^{(1)}), \\ &= n - \text{rank } A, \end{aligned}$$

see e.g. [12]. ■

A.2 The matrix inversion lemma

Lemma A.2 (Matrix inversion lemma). *Let $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{m \times n}$ and $D \in \mathbb{R}^{m \times m}$ be real matrices. If A , $D - CA^{-1}B$, and D are nonsingular, then $A - BD^{-1}C$ is nonsingular, and*

$$(A - BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}.$$

The matrix inversion lemma is used in this thesis for two purposes.

- It can be used to reduce the number of computations when inverting an expression of the form $A - BD^{-1}C$. Suppose that $n \gg m$ and that the inverse of A is easy to compute, e.g. A is diagonal. Then, the matrix inversion lemma provides a very efficient manner to calculate the inverse of $A - BD^{-1}C$.
- It can be used to convert between covariance and information filter formulas.

The following formula provides a manner to invert a 2×2 block matrix based on the matrix inversion lemma,

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & 0 \\ 0 & (D - CA^{-1}B)^{-1} \end{bmatrix} \begin{bmatrix} I & -BD^{-1} \\ -CA^{-1} & I \end{bmatrix}. \quad (\text{A.1})$$

Indeed, the diagonal entries of the first matrix on the right hand side of the equality sign can be computed using the matrix inversion lemma.

Appendix B

Least-squares estimation: deterministic vs. stochastic setting

In many applications, one is interested in determining an unknown vector $x \in \mathbb{R}^n$ based on a vector of measurements $y \in \mathbb{R}^p$ and a model that relates x to y . In linear LS estimation, the model is assumed to have the form $y = Cx + v$, where $C \in \mathbb{R}^{p \times n}$ is chosen appropriately and where the *noise vector* $v \in \mathbb{R}^p$ denotes the mismatch between the model and the measurement.

Depending on the nature of the noise vector v , two settings of the LS problem are usually considered: the deterministic setting and the stochastic setting.

B.1 Deterministic setting

In the deterministic setting of LS estimation, the noise vector v is assumed to be deterministic. This setting of the LS problem has been studied by both Legendre and Gauss. It deals with determining an estimate $\hat{x}_{\text{LS}} \in \mathbb{R}^n$ that minimizes the weighted sum of squares of the residuals $\|y - Cx\|_W^2$, i.e.

$$\hat{x}_{\text{LS}} = \arg \min_x \|y - Cx\|_W^2, \quad (\text{B.1})$$

where the weighting matrix $W \in \mathbb{R}^{p \times p}$ represents the degree of confidence in the individual measurements.

The solution to the LS problem (B.1) is given in the following theorem.

Theorem B.1. *Consider the equation $y = Cx + v$, where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^p$ and $v \in \mathbb{R}^p$ are deterministic. Let $\text{rank } C = n$ and let the weighting matrix W be positive definite. Then, the LS estimate (B.1) is unique and given by*

$$\hat{x}_{\text{LS}} = (C^T W C)^{-1} C^T W y. \quad (\text{B.2})$$

Proof: First, note that (B.1) can be rewritten as

$$\hat{x}_{\text{LS}} = \arg \min_x x^\top (C^\top W C) x - 2x^\top C^\top W y + y^\top W y.$$

Setting the derivative of the objective function with respect to x equal to zero, yields $C^\top W y = (C^\top W C) \hat{x}_{\text{LS}}$. The latter equation has a unique solution if and only if $C^\top W C$ is non-singular, that is, if and only if C has full column rank and W is positive definite. The unique solution is then given by (B.2). ■

B.2 Stochastic setting

Usually, however, a stochastic assumption is made regarding the noise vector v . Consider again the model $y = Cx + v$, where x is still deterministic, but v is assumed to be a zero-mean random variable, i.e. $\mathbb{E}[v] = 0$, with covariance matrix $R := \mathbb{E}[vv^\top]$. The objective of the stochastic LS problem is to determine an estimate \hat{x}_{LS} of x that satisfies the following two conditions: (i) The estimate \hat{x}_{LS} is *unbiased*, meaning that $\mathbb{E}[x - \hat{x}_{\text{LS}}] = 0$, and (ii) the estimate \hat{x}_{LS} minimizes the *mean squared error* $\mathbb{E}[\|x - \hat{x}_{\text{LS}}\|^2]$ over all linear unbiased estimates. It is straightforward to show that the second condition is equivalent to minimization of the trace of the *error covariance matrix* P_{LS} defined by

$$P_{\text{LS}} := \mathbb{E}[(x - \hat{x}_{\text{LS}})(x - \hat{x}_{\text{LS}})^\top].$$

The estimate \hat{x}_{LS} referred to above is called the *best linear unbiased estimate* or the *minimum-variance unbiased* (MVU) estimate of x . It is characterized by the Gauss-Markov theorem.

Theorem B.2 (Gauss-Markov theorem). *Consider the equation $y = Cx + v$, where $x \in \mathbb{R}^n$ is deterministic and $y \in \mathbb{R}^p$ and $v \in \mathbb{R}^p$ are random vectors. It is assumed that v is zero-mean and has covariance matrix R . Let $\text{rank } C = n$ and let R be non-singular, then*

$$\hat{x}_{\text{LS}} = (C^\top R^{-1} C)^{-1} C^\top R^{-1} y \tag{B.3}$$

is the unique best linear unbiased estimator of x . The covariance matrix P_{LS} of \hat{x}_{LS} , is given by $P_{\text{LS}} = (C^\top R^{-1} C)^{-1}$.

Proof: See e.g. [82]. ■

Note that by making the choice $W = R^{-1}$, (B.2) reduces to (B.3). This yields the interesting insight that an LS problem of the form (B.1) can be given a stochastic interpretation by weighting the norm by the covariance matrix of the random vector between the $\|\cdot\|$ signs.

The inverse of the covariance matrix, P_{LS}^{-1} , is called the *information matrix* of \hat{x}_{LS} . Note that it follows from Theorem B.2 that $P_{\text{LS}}^{-1} \hat{x}_{\text{LS}} = C^\top R^{-1} y$. This kind of expression in which the LS estimate \hat{x}_{LS} is pre-multiplied by its information matrix, is said to be the LS solution in information form.

Appendix C

Proofs and derivations

C.1 Rank proofs

Lemma C.1. *Let $A \in \mathbb{R}^{n \times n}$ and $C \in \mathbb{R}^{p \times n}$. Then, the matrix*

$$\mathcal{A}_N := \begin{bmatrix} C & & & & & \\ A & -I & & & & \\ & C & & & & \\ & A & -I & & & \\ & & & \ddots & & \\ & & & & C & \\ & & & & A & -I \end{bmatrix} \begin{matrix} \} 1 \\ \} 2 \\ \vdots \\ \} N \end{matrix}$$

with $N \geq n$, has full column rank if and only if $\{A, C\}$ is observable.

Proof: The idea behind the proof is to apply column and row operations that preserve the rank, but transform the matrix so that an extended observability matrix of $\{A, C\}$ appears in the first column. Without loss of generality, we assume that N is even. The proof consists of two steps.

- In the first step, we apply a sequence of column operations that preserve the rank. This sequence is determined by

$$\underbrace{\begin{bmatrix} I & & & & & \\ A & I & & & & \\ & A & I & & & \\ & & & \ddots & & \\ & & & & A & I \end{bmatrix}}_{\mathcal{U}_{1,N}}, \underbrace{\begin{bmatrix} I & & & & & \\ 0 & I & & & & \\ A^2 & 0 & I & & & \\ & & & \ddots & & \\ & & & & A^2 & 0 & I \end{bmatrix}}_{\mathcal{U}_{2,N}}, \dots, \underbrace{\begin{bmatrix} I & & & & & \\ 0 & I & & & & \\ 0 & 0 & I & & & \\ \vdots & \vdots & & \ddots & & \\ A^N & 0 & 0 & \dots & I \end{bmatrix}}_{\mathcal{U}_{N,N}}.$$

It then follows that

$$\text{rank}(\mathcal{A}_N) = \text{rank}(\mathcal{A}_N \mathcal{U}_{1,N} \mathcal{U}_{2,N} \cdots \mathcal{U}_{N,N})$$

$$= \text{rank} \left(\underbrace{\begin{bmatrix} C & & & & \\ 0 & -I & & & \\ CA & C & & & \\ 0 & 0 & -I & & \\ \vdots & \vdots & \ddots & & \\ CA^{N-1} & CA^{N-2} & \cdots & C & \\ 0 & 0 & \cdots & 0 & -I \end{bmatrix}}_{\bar{\mathcal{A}}_N} \right).$$

- In the second step, we apply row operations by left multiplication with

$$\underbrace{\begin{bmatrix} I & & & & \\ 0 & I & & & \\ 0 & C & I & & \\ 0 & 0 & 0 & I & \\ 0 & CA & 0 & C & I \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ 0 & CA^{N-2} & 0 & CA^{N-2} & \cdots & C & I \\ 0 & 0 & 0 & 0 & \cdots & 0 & I \end{bmatrix}}_{\mathcal{V}_N}.$$

It then follows that

$$\begin{aligned} \text{rank}(\bar{\mathcal{A}}_N) &= \text{rank}(\mathcal{V}_N \bar{\mathcal{A}}_N) \\ &= \text{rank} \left(\begin{bmatrix} C & & & & \\ 0 & -I & & & \\ CA & 0 & & & \\ 0 & 0 & -I & & \\ CA^2 & 0 & 0 & & \\ \vdots & \vdots & \vdots & \ddots & \\ CA^{N-1} & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & -I \end{bmatrix} \right). \end{aligned}$$

The latter matrix has full column rank if and only if \mathcal{O}_{N-1} has full column rank. Finally, it follows from Theorem 2.1 and Proposition 2.1 that \mathcal{O}_{N-1} , with $N \geq n$, has full column rank if and only if $\{A, C\}$ is observable.

■

Lemma C.2. *The matrix*

$$\mathcal{B}_N := \begin{bmatrix} I & & & & & \\ C & D & 0 & & & \\ A & B & -I & & & \\ & & & C & D & 0 \\ & & & A & B & -I \\ & & & & \ddots & \\ & & & & & C & D & 0 \\ & & & & & A & B & -I \end{bmatrix} \begin{matrix} \} 1 \\ \\ \\ \} 2 \\ \\ \vdots \\ \\ \} N \end{matrix}$$

has full column rank if and only if D has full column rank.

Proof: The proof is similar to that of Lemma C.1. The idea is to apply column and row operations that preserve the rank, but transform the matrix. We will see that by a convenient choice of column and row operations, the matrix \mathcal{H}_N now turns up in the transformed matrix.

We give a short outline of the proof. Without loss of generality, we assume that N is even. The proof consists of two steps.

- In the first step, we apply a sequence of column operations that preserve the rank. This sequence is determined by

$$\begin{bmatrix} I & & & & & \\ 0 & I & & & & \\ A & B & I & & & \\ & & & \ddots & & \\ & & & & I & \\ & & & & 0 & I \\ & & & & A & B & I \end{bmatrix}, \begin{bmatrix} I & & & & & \\ 0 & I & & & & \\ 0 & 0 & I & & & \\ 0 & 0 & 0 & I & & \\ A^2 & AB & 0 & 0 & I & \\ & & & & & \ddots \end{bmatrix}, \dots$$

- In the second step, we apply row operations by left multiplication with

$$\begin{bmatrix} I & & & & & \\ C & I & & & & \\ 0 & 0 & I & & & \\ CA & 0 & C & I & & \\ \vdots & \vdots & \vdots & \ddots & & \\ CA^{N-2} & 0 & CA^{N-2} & \dots & C & I \\ 0 & 0 & 0 & \dots & 0 & I \end{bmatrix}.$$

We then find that \mathcal{B}_N has full column rank if and only if \mathcal{H}_N (defined by (3.6)) has full column rank. A necessary and sufficient condition for this to hold is that D has full column rank. \blacksquare

C.2 Proof of Proposition 4.1

Proof: It follows from the Gauss-Markov theorem (Theorem B.2) that the solution to (4.37) is given by

$$\begin{bmatrix} \hat{x}_{[k|k]} \\ \hat{u}_{[k|k]} \end{bmatrix} = \begin{bmatrix} P_{[k|k-1]}^{-1} + C^T R^{-1} C & C^T R^{-1} D \\ D^T R^{-1} C & D^T R^{-1} D \end{bmatrix}^{-1} \begin{bmatrix} P_{[k|k-1]}^{-1} & C^T R^{-1} \\ 0 & D^T R^{-1} \end{bmatrix} \begin{bmatrix} \hat{x}_{[k|k-1]} \\ y_{[k]} \end{bmatrix}, \quad (\text{C.1})$$

where the block matrix on the right hand side of the equality sign can be identified as the error covariance matrix of $[\hat{x}_{[k|k]}^T \hat{u}_{[k|k]}^T]^T$. It follows from (A.1) that the error covariance matrix is given by

$$\begin{bmatrix} P_{[k|k-1]}^{-1} + C^T R^{-1} C & C^T R^{-1} D \\ D^T R^{-1} C & D^T R^{-1} D \end{bmatrix}^{-1} = \begin{bmatrix} P_{[k|k]} & 0 \\ 0 & P_{u[k|k]} \end{bmatrix} \begin{bmatrix} I & -D^T R^{-1} C (P_{[k|k-1]}^{-1} + C^T R^{-1} C)^{-1} \\ -C^T R^{-1} D (D^T R^{-1} D)^{-1} & I \end{bmatrix}, \quad (\text{C.2})$$

where the inverses of $P_{[k|k]}$ and $P_{u[k|k]}$ are given by

$$P_{[k|k]}^{-1} = P_{[k|k-1]}^{-1} + C^T R^{-1} C - C^T R^{-1} D (D^T R^{-1} D)^{-1} D^T R^{-1} C, \quad (\text{C.3})$$

and

$$P_{u[k|k]}^{-1} = D^T R^{-1} D - D^T R^{-1} C (P_{[k|k-1]}^{-1} + C^T R^{-1} C)^{-1} C^T R^{-1} D, \quad (\text{C.4})$$

respectively. Note that, as suggested by the notation, $P_{[k|k]}$ and $P_{u[k|k]}$ can be identified as the error covariance matrices of $\hat{x}_{[k|k]}$ and $\hat{u}_{[k|k]}$, respectively. Consequently, (C.3) and (C.4) yield equations for the information matrices. Substituting (C.2) in (C.1), yields the following equation for the estimate of the system state and the unknown input in information form,

$$P_{[k|k]}^{-1} \hat{x}_{[k|k]} = P_{[k|k-1]}^{-1} \hat{x}_{[k|k-1]} + C^T R^{-1} y_{[k]} - C^T R^{-1} D (D^T R^{-1} D)^{-1} D^T R^{-1} y_{[k]},$$

and

$$P_{u[k|k]}^{-1} \hat{u}_{[k|k]} = D^T R^{-1} y_{[k]} - D^T R^{-1} C (P_{[k|k-1]}^{-1} + C^T R^{-1} C)^{-1} (P_{[k|k-1]}^{-1} \hat{x}_{[k|k-1]} + C^T R^{-1} y_{[k]}),$$

respectively. Applying the matrix inversion lemma to the information formulas yields, after some calculation, the equations for the measurement update and the estimation of the unknown input considered in Sect. 4.2.4. ■

C.3 Derivation of the Eqs. in Sect. 5.3.2

First, we derive an equation for $\bar{P}_{[k|k]} := \mathbb{E}[\tilde{\tilde{x}}_{[k|k]}\tilde{\tilde{x}}_{[k|k]}^\top]$. It follows from (5.14) that

$$\begin{aligned}\tilde{\tilde{x}}_{[k|k]} &= Ax_{[k-1]} + Bu_{[k-1]} - \hat{\tilde{x}}_{[k|k-1]} + B\hat{u}_{[k-1|k]}, \\ &= \tilde{\tilde{x}}_{[k|k-1]} + B\tilde{u}_{[k-1|k]}.\end{aligned}\quad (\text{C.5})$$

An equation for $\tilde{u}_{[k-1|k]}$ is obtained from (5.13), which yields

$$\tilde{u}_{[k-1|k]} = (I - K_{u[k]}F)\tilde{u}_{[k-1]} - K_{u[k]}(C\tilde{\tilde{x}}_{[k|k-1]} + v_{[k]}), \quad (\text{C.6})$$

where $\tilde{u}_{[k-1]} := u_{[k-1]} - \hat{u}_{[k-1]}$. Substituting (C.6) in (C.5), yields

$$\tilde{\tilde{x}}_{[k|k]} = (I - BK_{u[k]}C)(\tilde{\tilde{x}}_{[k|k-1]} + B\tilde{u}_{[k-1]}) - BK_{u[k]}v_{[k]}.$$

Consequently,

$$\begin{aligned}\bar{P}_{[k|k]} &= (I - BK_{u[k]}C)(\bar{P}_{[k|k-1]} + BP_{u[k-1|k]}B^\top)(I - BK_{u[k]}C)^\top \\ &\quad + BK_{u[k]}RK_{u[k]}^\top B^\top.\end{aligned}$$

Now, we derive an expression for $P_{[k|k]}$ in terms of $\bar{L}_{[k]}$. It follows from (5.15) that

$$\tilde{x}_{[k|k]} = (I - \bar{L}_{[k]}C)\tilde{\tilde{x}}_{[k|k]} - \bar{L}_{[k]}v_{[k]}.$$

Consequently,

$$\begin{aligned}P_{[k|k]} &= (I - \bar{L}_{[k]}C)\bar{P}_{[k|k]}(I - \bar{L}_{[k]}C)^\top + \bar{L}_{[k]}R\bar{L}_{[k]}^\top \\ &\quad + (I - \bar{L}_{[k]}C)BK_{u[k]}R\bar{L}_{[k]}^\top + \bar{L}_{[k]}RK_{u[k]}^\top B^\top(I - \bar{L}_{[k]}C)^\top.\end{aligned}$$

After some calculation, it is found that the gain matrix $\bar{L}_{[k]}$ minimizing the trace of $P_{[k|k]}$ can be written as

$$\begin{aligned}\bar{L}_{[k]} &= \bar{P}_{[k|k-1]}C^\top [(I - FK_{u[k]})(\bar{R}_{[k]} + FP_{u[k-1|k-1]}F^\top)]^{-1} \\ &= \bar{P}_{[k|k-1]}C^\top \bar{R}_{[k]}^{-1}.\end{aligned}$$

Bibliography

- [1] VIC - High-performance Linux Cluster. <http://ludit.kuleuven.be/hpc/>.
- [2] J.I. Allen, M. Eknes, and G. Evensen. An ensemble Kalman filter with a complex marine ecosystem model: hindcasting phytoplankton in the Cretan sea. *Annales Geophysicae*, 20:113, 2002.
- [3] A.T. Alouani, P. Xia, T.R. Rice, and W.D. Blair. On the optimality of two-stage state estimation in the presence of random bias. *IEEE Transaction on Automatic Control*, 38(8):1279–1283, 1993.
- [4] B.D.O. Anderson and J.B. Moore. *Optimal filtering*. Prentice-Hall, 1979.
- [5] J.L. Anderson. An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, 129(12):2884–2903, Dec 2001.
- [6] J.L. Anderson and S.L. Anderson. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127:2741–2758, 1999.
- [7] P.J. Antsaklis. Stable proper n th-order inverses. *IEEE Transactions on Automatic Control*, 23:1104–1106, 1978.
- [8] O. Barrero. *Data assimilation in magnetohydrodynamics systems using Kalman filtering*. PhD thesis, Katholieke Universiteit Leuven, Belgium, 2005.
- [9] O. Barrero, D.S. Bernstein, and B. De Moor. Spatially localized Kalman filtering for data assimilation. In *Proceedings of the American Control Conference*, pages 3468–3473, 2005.
- [10] O. Barrero and B. De Moor. A singular square root algorithm for large scale systems. In *Proceedings 15th IASTED International Conference on Modelling and Simulation*, 2004.
- [11] O. Barrero, B. De Moor, and D.S. Bernstein. Data assimilation for magnetohydrodynamics systems. *Journal of Computational and Applied Mathematics*, 189:242–259, 2006.
- [12] D.S. Bernstein. *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*. Princeton University Press, Princeton, New Jersey, 2005.

- [13] C.H. Bishop, B. Etherton, and S.J. Majundar. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly Weather Review*, 129:420–436, 2001.
- [14] D. Boggs, M. Ghil, and C. Keppenne. A stabilized sparse-matrix U-D square-root implementation of a large-state extended Kalman filter. In *Second International Symposium on the Assimilation of Observations in Meteorology and Oceanography*, pages 219–224, 1995.
- [15] R.W. Brockett. Poles, zeros and feedback: state-space representations. *IEEE Transactions on Automatic Control*, 10:129–135, 1965.
- [16] R.W. Brockett and M.D. Mesarovic. The reproducibility of multivariable control systems. *Journal of Mathematical Analysis and Applications*, 11:548–563, 1965.
- [17] J. Chandrasekar, O. Barrero, A. Ridley, D.S. Bernstein, and B. De Moor. State estimation for linearized MHD flow. In *Proceedings of the 43th IEEE Conference on Decision and Control*, volume 3, pages 2584–2589, 2004.
- [18] H.-B. Chen, J.H. Chow, M.A. Kale, and K.D. Minto. Simultaneous stabilization using stable system inversion. *Automatica*, 31(2):531–542, 1995.
- [19] H.T. Chen, S.Y. Lin, H.R. Wang, and L.C. Fang. Estimation of two-sided boundary conditions for two-dimensional inverse heat conduction problems. *International Journal of Heat Mass Transfer*, 45:15–43, 2002.
- [20] J. Chen and R.J. Patton. *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Kluwer Academic Publishers, London, 1999.
- [21] B. Cipra. Engineers look to Kalman filtering for guidance. *SIAM News*, 26(5):757–764, 1993.
- [22] S.E. Cohn and R. Todling. Approximate Kalman filters for unstable dynamics. In *Second International Symposium on the Assimilation of Observations in Meteorology and Oceanography*, pages 241–246, 1995.
- [23] P. Courtier and O. Talagrand. Variational assimilation of meteorological observations with the direct and adjoint shallow water equations. *Tellus*, 42A:531–549, 1990.
- [24] M. Darouach and M. Zasadzinski. Unbiased minimum variance estimation for systems with unknown exogenous inputs. *Automatica*, 33(4):717–719, 1997.
- [25] M. Darouach, M. Zasadzinski, and M. Boutayeb. Extension of minimum variance estimation for systems with unknown inputs. *Automatica*, 39:867–876, 2003.
- [26] M. Darouach, M. Zasadzinski, and S.J. Xu. Full-order observers for linear systems with unknown inputs. *IEEE Transactions on Automatic Control*, 39(3):606–609, 1994.

-
- [27] H. De Sterck, B.C. Low, and S. Poedts. Complex magnetohydrodynamic bow shock topology in field-aligned low- β flow around a perfectly conducting cylinder. *Physics of Plasmas*, 5(11):4015–4027, 1998.
- [28] H. De Sterck and S. Poedts. Overcompressive shocks and compound shocks in 2D and 3D magnetohydrodynamic flows. In *Proceedings of the 8th International Conference on Hyperbolic Problems*, 2000.
- [29] D.P. Dee and A.M. Da Silva. Data assimilation in the presence of forecast bias. *Quarterly Journal of the Royal Meteorological Society*, 117:269–295, 1998.
- [30] D.P. Dee and R. Todling. Data assimilation in the presence of forecast bias: the GEOS moisture analysis. *Monthly Weather Review*, 128:3268–3282, 2000.
- [31] R. Diversi, R. Guidorzi, and U. Soverini. Kalman filtering in symmetrical noise environments. In *Proceedings of the 11th IEEE Mediterranean Conference on Control and Automation*, June 2003.
- [32] D.B. Duncan and S.D. Horn. Linear dynamic recursive estimating from the viewpoint of regression analysis. *Journal of the American Statistical Association*, 67:815–821, 1972.
- [33] C. Edwards and C.P. Tan. A comparison of sliding mode and unknown input observers for fault reconstruction. *European Journal of control*, 12:245–260, 2006.
- [34] E. Emre and Ö. Hüseyin. Invertibility criteria for linear multivariable systems. *IEEE Transactions on Automatic Control*, 19(5):609–610, 1974.
- [35] E. Emre and L.M. Siverman. Minimal dynamic inverses for linear systems with arbitrary initial states. *IEEE Transactions on Automatic Control*, 21(5):766–769, 1976.
- [36] H.J. Eskes, P.F.J. van Velthoven, and H.M. Kelder. Global ozone forecasting based on ERS-2 GOME observations. *Atmospheric Chemistry and Physics*, 2:271–278, 2002.
- [37] J. Evans. *The History and Practice of Ancient Astronomy*. Oxford University Press, 1998.
- [38] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal Geophysical Research*, 99(C5):10143–10162, 1994.
- [39] G. Evensen and P.J. van Leeuwen. Assimilation of GEOSAT altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model. *Monthly Weather Review*, 124:85–96, 1996.
- [40] F.W. Fairman. *Linear Control Theory: The State Space Approach*. John Wiley and Sons, 1998.

- [41] U. Feldmann, M. Hasler, and W. Schwarz. Communication by chaotic signals: The inverse system approach. *International Journal of Circuit Theory and Applications*, 24(5):551–579, 1996.
- [42] T. Fernando and H. Trinh. Design of reduced order state/unknown input observers: a descriptor system approach. In *Proceedings of the IEEE International Conference on Control Applications*, October 2006.
- [43] R. Fitzgerald. Divergence of the Kalman filter. *IEEE Transactions on Automatic Control*, 16(6):736–747, 1971.
- [44] T. Floquet and J.-P. Barbot. State and unknown input estimation for linear discrete-time systems. *Automatica*, 42:1883–1889, 2006.
- [45] B. Friedland. Treatment of bias in recursive filtering. *IEEE Transactions on Automatic Control*, 14:359–367, 1969.
- [46] L.S. Gandin. *Objective analysis of meteorological fields*. Gidrometeoizdat, Leningrad. [Translated from Russian, Israel Program for Scientific Translation, Jerusalem], 1965.
- [47] A. Gelb. *Applied optimal estimation*. MIT press, 1974.
- [48] S. Gillijns, D.S. Bernstein, and B. De Moor. The reduced rank transform square root filter for data assimilation. In *Proceedings of the 14th IFAC Symposium on System Identification*, April 2006.
- [49] S. Gillijns and B. De Moor. Unbiased minimum-variance input and state estimation for linear discrete-time systems. *Automatica*, 43(1):111–116, 2007.
- [50] S. Gillijns and B. De Moor. Data-based subsystem identification for dynamic model updating. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 3303–3308, 2006.
- [51] S. Gillijns and B. De Moor. Linear recursive filtering with noisy input and output measurements. Internal report, KULeuven, 2006.
- [52] S. Gillijns and B. De Moor. Information, covariance and square-root filtering in the presence of unknown inputs. In *Proceedings of the European Control Conference*, pages 2213–2217, 2007.
- [53] S. Gillijns and B. De Moor. Joint state and boundary condition estimation in linear data assimilation using basis function expansion. In *Proceedings of the 26th IASTED International Conference on Modelling, Identification, and Control*, pages 458–463, 2007.
- [54] S. Gillijns and B. De Moor. Model error estimation in ensemble data assimilation. *Nonlinear Processes in Geophysics*, 14(1):59–71, 2007.
- [55] S. Gillijns and B. De Moor. Recursive least-squares estimation for systems with unknown inputs. Accepted for publication in *IEEE Transactions on Automatic Control*, KULeuven, 2007.

-
- [56] S. Gillijns and B. De Moor. System inversion with application to filtering and smoothing in the presence of unknown inputs. Submitted for publication, KULeuven, 2007.
- [57] S. Gillijns and B. De Moor. Unbiased minimum-variance input and state estimation for linear discrete-time systems with direct feedthrough. *Automatica*, 43(5):934–937, 2007.
- [58] J. Glover. The linear estimation of completely unknown signals. *IEEE Transactions on Automatic Control*, 14(6):766–767, 1969.
- [59] H. Goedbloed and S. Poedts. *Principles of Magnetohydrodynamics; with Applications to Laboratory and Astrophysical Plasmas*. Cambridge University Press, 2004.
- [60] M. Goossens. *An introduction to plasma astrophysics and magnetohydrodynamics*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003.
- [61] C.P.T. Groth, D.L. DeZeeuw, T.I. Gombosi, and K.G. Powell. Global three-dimensional MHD simulation of a space weather event: CME formation, interplanetary propagation and interaction with the magnetosphere. *Journal of Geophysical Research*, 105:25053–25078, 2000.
- [62] R. Guidorzi, R. Diversi, and U. Soverini. Optimal errors-in-variables filtering. *Automatica*, 39(2):281–289, 2003.
- [63] T.M. Hamill, J.S. Whitaker, and C. Snyder. Distance-dependent filtering of background-error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, 129:2776–2790, 2001.
- [64] K.C. Hansen, T.I. Gombosi, D.L. DeZeeuw, C.P.T. Groth, and K.G. Powell. A 3D global MHD simulation of Saturn’s magnetosphere. *Advances in Space Research*, 26(10):1681–1690, 2000.
- [65] H.L. Harter. The method of least squares and some alternatives - Part I. *International Statistical Review*, 42(2):147–174, 1974.
- [66] A.W. Heemink, M. Verlaan, and A.J. Segers. Variance reduced ensemble Kalman filtering. *Monthly Weather Review*, 129:1718–1728, 2001.
- [67] M. Hou and P.C. Müller. Design of observers for linear systems with unknown inputs. *IEEE Transactions on Automatic Control*, 37(6):871–874, 1992.
- [68] M. Hou and P.C. Müller. Disturbance decoupled observer design: A unified viewpoint. *IEEE Transactions on Automatic Control*, 39(6):1338–1341, 1994.
- [69] M. Hou and J. Patton. Input observability and input reconstruction. *Automatica*, 34(6):789–794, 1998.

- [70] M. Hou and R.J. Patton. Optimal filtering for systems with unknown inputs. *IEEE Transactions on Automatic Control*, 43(3):445–449, 1998.
- [71] P.L. Houtekamer and H.L. Mitchell. Data assimilation using an ensemble Kalman filtering technique. *Monthly Weather Review*, 126:796–811, 1998.
- [72] P.L. Houtekamer and H.L. Mitchell. A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129:123–137, 2001.
- [73] C.S. Hsieh. Robust two-stage Kalman filters for systems with unknown inputs. *IEEE Transactions on Automatic Control*, 45(12):2374–2378, 2000.
- [74] P. Hsu, Y. Yang, and C. Chen. Simultaneously estimating the initial and boundary conditions in a two-dimensional hollow cylinder. *International Journal of Heat Mass Transfer*, 41(1):219–227, 1998.
- [75] M.B. Ignani. Separate-bias Kalman estimator with bias state noise. *IEEE Transactions on Automatic Control*, 35(3):338–341, 1990.
- [76] S. Jakubek and H.P. Jörgl. Sensor-fault-diagnosis using inverse dynamic systems. In *Proceedings of the American Control Conference*, pages 2131–2136, 2001.
- [77] A.H. Jazwinski. *Stochastic processes and filtering theory*. Academic Press, New York, 1970.
- [78] J. Jin and M.-J. Tahk. Time-delayed state estimator for linear systems with unknown inputs. *International Journal of Control, Automation, and Systems*, 3(1):117–121, 2005.
- [79] T.A. Johansen and B.A. Foss. Representing and learning unmodeled dynamics with neural network memories. In *Proceedings of the American Control Conference*, pages 3037–3043, Chicago, USA, 1992.
- [80] T. Kailath. A view of three decades of linear filtering theory. *IEEE Transactions on Information Theory*, 20(2):146–181, 1974.
- [81] T. Kailath. *Linear Systems*. Prentice Hall, Englewood Cliffs, New Jersey, 1980.
- [82] T. Kailath, A.H. Sayed, and B. Hassibi. *Linear Estimation*. Prentice Hall, Upper Saddle River, New Jersey, 2000.
- [83] R.E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering*, 82(Series D):35–45, 1960.
- [84] C.F. Kennel. MHD intermediate shock discontinuities. Part 1. Rankine-Hugoniot conditions. *Journal of Plasma Physics*, 42:299–319, 1989.

-
- [85] C.L. Keppenne and M.M. Rienecker. Initial testing of a massively parallel ensemble Kalman filter with the poseidon isopycnal ocean general circulation model. *Monthly Weather Review*, 130:2951–2965, 2002.
- [86] W.S. Kerwin and J.L. Prince. On the optimality of recursive unbiased state estimation with unknown inputs. *Automatica*, 36:1381–1383, 2000.
- [87] P.K. Kitaniadis. Unbiased minimum-variance linear state estimation. *Automatica*, 23(6):775–778, 1987.
- [88] F.-X. Le Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus*, 38A:97–110, 1986.
- [89] Y. Liu, A.F. Nagy, K. Kabin, M.R. Combi, D.L. De Zeeuw, T.I. Gombosi, and K.G. Powell. Two species, 3D, MHD simulation of Europa’s interaction with Jupiter’s magnetosphere. *Geophysical Research Letters*, 27:1791, 2000.
- [90] E.N. Lorenz. Predictability: A problem partly solved. In *Proceedings ECMWF Seminar on Predictability*, volume I, pages 1–18, 1996.
- [91] E.N. Lorenz and K.A. Emanuel. Optimal sites for supplementary weather observations: Simulation with a small model. *Journal of the Atmospheric Sciences*, 55:399–414, 1998.
- [92] D.G. Luenberger. An introduction to observers. *IEEE Transactions on Automatic Control*, 16(6):596–602, 1971.
- [93] I. Markovsky and B. De Moor. Linear dynamic filtering with noisy input and output. *Automatica*, 41(1):167–171, 2005.
- [94] I. Markovsky and B. De Moor. Linear dynamic filtering with noisy input and output. In *Proceedings of the 13th IFAC symposium on System Identification*, pages 1749–1754, 2003.
- [95] J.L. Massey and M.K. Sain. Inverses of linear sequential circuits. *IEEE Transactions on Automatic Control*, 17(4):330–337, 1968.
- [96] P.S. Maybeck. *Stochastic models, estimation, and control*. Academic Press, London, 1979.
- [97] M. Moonen. Implementing the square-root information kalman filter on a jacobi-type systolic array. *Journal of VLSI Signal Processing Systems*, 8(3):283–291, 1994.
- [98] M. Morf and T. Kailath. Square-root algorithms for least-squares estimation. *IEEE Transactions on Automatic Control*, 20:487–497, 1975.
- [99] V.O. Mowery. Least squares recursive differential-correction estimation in nonlinear problems. *IEEE Transactions on Automatic Control*, 10:399–407, 1965.

-
- [100] P.J. Moylan. Stable inversion of linear systems. *IEEE Transactions on Automatic Control*, 22(1):74–78, 1977.
- [101] K.M. Neaupane and M. Sugimoto. An inverse boundary value problem using the extended Kalman filter. *ScienceAsia*, 29:121–126, 2003.
- [102] R. Nikoukhah, S.L. Campbell, and F. Delebecque. Observer design for general linear time-invariant systems. *Automatica*, 34(5):575–583, 1998.
- [103] A. Oksasoglu and T. Akgul. A linear inverse system approach in the context of chaotic communications. *IEEE Transactions Circuits and Systems I*, 44:75–79, 1997.
- [104] P.A. Orner. Construction of inverse systems. *IEEE Transactions on Automatic Control*, 17(1):151–153, 1972.
- [105] P. Van Overschee and B. De Moor. N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30(1):75–93, 1994.
- [106] P. Van Overschee and B. De Moor. *Subspace Identification for Linear Systems: Theory, Implementation, Applications*. Kluwer Academic Publishers, 1996.
- [107] C.C. Paige and M.A. Saunders. Least squares estimation of discrete linear dynamic systems from using orthogonal transformations. *Siam Journal on Numerical Analysis*, 14(2):180–193, 1977.
- [108] H. Palanthandalam-Madapusi, E.L. Renk, and D.S. Bernstein. Data-based model refinement for linear and Hammerstein systems using subspace identification and adaptive disturbance rejection. In *Proceedings of IEEE Conference on Control Applications*, August 2005.
- [109] E.N. Parker. Dynamics of the interplanetary gas and magnetic fields. *Astrophysical Journal*, 128:664–676, 1958.
- [110] A.G. Parlos, S.K. Menon, and A.F. Atiya. An adaptive state filtering algorithm for systems with partially known dynamics. *Journal of Dynamic Systems, Measurement and Control*, 124:364–374, 2002.
- [111] J.E. Potter and R.G. Stern. Statistical filtering of space navigation measurements. In *Proceedings of AIAA Guidance and Control Conference*, 1963.
- [112] R.H. Reichle, D.B. McLaughlin, and D. Entekhabi. Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, 130:103–114, 2002.
- [113] H.H. Rosenbrock. *State-space and multivariable theory*. Nelson, London, 1970.

- [114] A. Saberi, A.A. Stoorvogel, and P. Sannuti. Exact, almost and optimal input decoupled (delayed) observers. *International Journal of Control*, 73(7):552–581, 2000.
- [115] M.K. Sain and J.L. Massey. Invertibility of linear time-invariant dynamical systems. *IEEE Transactions on Automatic Control*, 14(2):141–149, 1969.
- [116] L.M. Silverman. Inversion of multivariable linear systems. *IEEE Transactions on Automatic Control*, 14(3):270–276, 1969.
- [117] H.W. Sorenson. Least-squares estimation: From Gauss to Kalman. *IEEE Spectrum*, 7:63–68, 1970.
- [118] R.S. Steinolfson and A.J. Hundhausen. MHD intermediate shocks in coronal mass ejections. *Journal of Geophysical Research*, 95(A5):6389–6401, 1990.
- [119] H. De Sterck. *Numerical simulation and analysis of magnetically dominated MHD bow shock flows with applications in space physics*. PhD thesis, Katholieke Universiteit Leuven, Belgium, and National Center for Atmospheric Research, Colorado, USA, 1999.
- [120] B.L. Stevens and F.L. Lewis. *Aircraft Control and Simulation*. John Wiley and Sons, 1992.
- [121] K. Suma and M. Kawahara. Estimation of boundary conditions for ground temperature control using Kalman filter and finite element method. *International Journal for Numerical Methods in Fluids*, 31:261–274, 1999.
- [122] Z. Sun, A. Tangborn, and W. Kuang. Data assimilation in a sparsely observed one-dimensional modeled MHD system. *Nonlinear Processes in Geophysics*, 14:181–192, 2007.
- [123] Z. Sun and T.-C. Tsao. Adaptive tracking control by system inversion. In *Proceedings of American Control Conference*, pages 29–33, 1999.
- [124] S. Sundaram and C.N. Hadjicostis. Comments on “Time-delayed state estimator for linear systems with unknown inputs”. *International Journal of Control, Automation, and Systems*, 3(4):646–647, 2005.
- [125] S. Sundaram and C.N. Hadjicostis. Optimal state estimators for linear systems with unknown inputs. In *Proceedings of 45th IEEE Conference on Decision and Control*, December 2006.
- [126] S. Sundaram and C.N. Hadjicostis. Delayed observers for linear systems with unknown inputs. *IEEE Transactions on Automatic Control*, 52(2):334–339, 2007.
- [127] P. Swerling. Modern state estimation methods from the viewpoint of the method of least squares. *IEEE Transactions on Automatic Control*, 16(6):707–719, 1971.

- [128] F. Szigeti, C.E. Vera, J. Bokor, and A. Edelmayer. Inversion based fault detection and isolation. In *Proceedings of 40th IEEE Conference on Decision and Control*, 2001.
- [129] S. Tan and J. Vandewalle. Inversion of singular systems. *IEEE Transactions on Circuits and systems*, 35(5):583–587, 1988.
- [130] G. Tóth. General code for modeling MHD flows on parallel computers: Versatile Advection Code. *Astrophysical Letters and Communications*, 34:245–258, 1996.
- [131] P. Van Dooren, P. Dewilde, and J. Vandewalle. On the determination of the Smith-Macmillan form of a rational matrix from its laurent expansion. *IEEE Transactions on Circuits and systems*, 26(3):180–189, 1979.
- [132] J. Vandewalle and J. Van Daele. On the computation of the inverse of a rational matrix via expansions. *International Journal of Control*, 31(1):95–107, 1980.
- [133] M. Verhaegen and P. Van Dooren. Numerical aspects of different Kalman filter implementations. *IEEE Transactions on Automatic Control*, 31(10):907–917, 1986.
- [134] M. Verlaan and A.W. Heemink. Reduced rank square-root filters for large-scale data assimilation problems. In *Second International Symposium on the Assimilation of Observations in Meteorology and Oceanography*, pages 247–252, 1995.
- [135] M. Verlaan and A.W. Heemink. Tidal flow forecasting using reduced rank square root filters. *Stochastic Hydrology and Hydraulics*, 11:349–368, 1997.
- [136] J.C. Willems. Deterministic least squares filtering. *Journal of Econometrics*, 118:341–373, 2004.
- [137] A.S. Willsky. On the invertibility of linear systems. *IEEE Transactions on Automatic Control*, 19(3):272–274, 1974.
- [138] C. Yang and C. Chen. Inverse estimation of the boundary condition in three-dimensional heat conduction. *Journal of Physics D: Applied Physics*, 30:2209–2216, 1997.
- [139] F. Yang and R.W. Wilde. Observers for linear systems with unknown inputs. *IEEE Transactions on Automatic Control*, 33(7):677–681, 1988.
- [140] F.-M. Yuan. Minimal dimension inverses of linear sequential circuits. *IEEE Transactions on Automatic Control*, 20(1):42–52, 1975.

Scientific curriculum vitae

Steven Gillijns was born on May 9, 1980, in Leuven, Belgium. He went to the Sint-Jozefscollege in Sint-Pieters-Woluwe, with majors science and mathematics. In July 2003, he received the degree of *Burgerlijk Werktuigkundig-Elektrotechnisch Ingenieur* (Electrical Engineer), option Data Mining and Automation, at the Katholieke Universiteit Leuven. In October 2003, he started pursuing a Ph.D. on the subject of Kalman filtering in the SCD research group of the Department of Electrical Engineering (ESAT), under the supervision of Prof.dr.ir Bart De Moor.

Publication list

Journal Papers

- S. Gillijns and B. De Moor. Unbiased minimum-variance input and state estimation for linear discrete-time systems. *Automatica*, 43(1), 111–116, 2007.
- S. Gillijns and B. De Moor. Model error estimation in ensemble data assimilation. *Nonlinear Processes in Geophysics*, 14(1), 59–71, 2007.
- S. Gillijns and B. De Moor. Unbiased minimum-variance input and state estimation for linear discrete-time systems with direct feedthrough. *Automatica*, 43(5), 934–937, 2007.
- S. Gillijns and B. De Moor. Recursive least-squares estimation for systems with unknown inputs. To be published in *IEEE Transactions on Automatic Control*.
- S. Gillijns and B. De Moor. System inversion with application to filtering and smoothing in the presence of unknown inputs. Submitted for publication.

International Conference Papers

- H.J. Palanthandalam-Madapusi, S. Gillijns, A.J. Ridley, D.S. Bernstein, and B. De Moor. Electric potential estimation with line-of-sight measurements using basis function optimization. In *Proceedings of the 43rd IEEE Conference on Decision and Control (CDC 2004)*, Paradise Island, The Bahamas, pages 3625–3630, 2004.
- S. Gillijns, D.S. Bernstein, and B. De Moor. The reduced rank transform square root filter for data assimilation. In *Proceedings of the 14th IFAC Symposium on System Identification (SYSID 2006)*, Newcastle, Australia, pages 1252–1257, 2006.
- S. Gillijns, O. Barrero Mendoza, J. Chandrasekar, B. De Moor, D.S. Bernstein and A.J. Ridley. What Is the ensemble Kalman filter and how well

- does it work? In *Proceedings of the 2006 American Control Conference (ACC 2006)*, Minneapolis, USA, pages 4448–4453, 2006.
- H.J. Palanthandalam-Madapusi, S. Gillijns, B. De Moor, and D.S. Bernstein. Subsystem identification for nonlinear model updating. In *Proceedings of the 2006 American Control Conference (ACC 2006)*, Minneapolis, USA, pages 3056–3061, 2006.
 - S. Gillijns and B. De Moor. Data-based subsystem identification for dynamic model updating In *Proceedings of the 45th IEEE Conference on Decision and Control (CDC 2006)*, San Diego, USA, pages 3303–3308, 2006.
 - S. Gillijns and B. De Moor. Joint state and boundary condition estimation in linear data assimilation using basis function expansion In *Proceedings of the 26th IASTED International Conference on Modelling, Identification, and Control (MIC 2007)*, Innsbruck, Austria, pages 458–463, 2007.
 - S. Gillijns, N. Haverbeke, and B. De Moor. Information, covariance and square-root filtering in the presence of unknown inputs. In *Proceedings of the European Control Conference (ECC 2007)*, Kos, Greece, pages 2213 – 2217, 2007.

Internal Reports

- S. Gillijns and B. De Moor. Linear recursive filtering with noisy input and output measurements. Internal Report 06-166, ESAT-SISTA, K.U. Leuven (Leuven, Belgium), 2006.