

Modellbaserad diagnos på turbojetmotorn RM12

Examensarbete utfört i Fordonssystem
vid Tekniska Högskolan i Linköping
av

Marcus Klein och Fredrik Östling

Reg nr: LiTH-ISY-EX-1980

Modellbaserad diagnos på turbojetmotorn RM12

Examensarbete utfört i Fordonssystem
vid Tekniska Högskolan i Linköping
av

Marcus Klein och Fredrik Östling

Reg nr: LiTH-ISK-EX-1980

Handledare: **Erik Frisk , Fordonssystem LiTH**
Peder Johnson , Volvo Aero Corp.

Examinator: **Lars Nielsen , Fordonssystem LiTH**

Linköping, 26 augusti 1998.



Avdelning, Institution
Division, Department

Department of Electrical Engineering
Division of Vehicular Systems

Datum:
Date:

1998-08-26

Språk
Language

- Svenska/Swedish
 Engelska/English

Rapporttyp
Report category

- Licentiatavhandling
 Examensarbete
 C-uppsats
 D-uppsats
 Övrig rapport

ISBN

ISRN

Serietitel och serienummer ISSN
Title of series, numbering

LiTH-ISY-EX-1980

URL för elektronisk version

<http://www.vehicular.isy.liu.se/>

Titel: Modellbaserad diagnos på turbojetmotorn RM12.
Title: Model Based Diagnosis Applied to Turbo-Jet Engine RM12.

Författare:

Author: Marcus Klein och Fredrik Östling

Sammanfattning

Abstract

This master thesis is a principal study of how model based diagnosis with dynamic models can be applied to the turbo-jet engine RM12. A model based diagnosis system can be used to detect failures automatically. Automatic fault detection is of interest since it simplifies the reparation and increases the safety. A principal study of model based diagnosis applied to RM12 has therefore been performed. The study was performed by linearizing a nonlinear engine model of RM12 in an operating point utilizing a perturbation technique. Then the linear model was reduced with two different methods, reduction via balanced realization and reduction with physical interpretation of states using modal analysis. The diagnosis system was designed from the reduced models with a method called the Minimal Polynomial Basis method. Simulation of faults were performed with the linearization of the engine model. The principal study shows that model based diagnosis with dynamic models applied to RM12 isolates small sensor and actuator faults. The study also shows that a small and well conditioned design model gives the best performing diagnosis system.

Nyckelord

Keywords model based diagnosis, minimal polynomial basis, turbo-jet engine

Sammanfattning

Detta examensarbete är en principstudie av modellbaserad diagnos med dynamiska modeller applicerad på JAS 39 Gripens turbojetmotor RM12. Ett modellbaserat diagnosystem kan användas för att detektera fel automatiskt. Automatisk feldetektering är av intresse då det underlättar reparationsarbetet och ökar säkerheten. En principstudie av modellbaserad diagnos på RM12 har därför utförts. Studien utfördes genom att en olinjär motormodell av RM12 linjäriserades i en arbetspunkt mha störningsteknik. Därefter reducerades den linjära modellen med två olika metoder, reducering med balanserade tillstånd och reducering med fysikalisk innebörd av tillstånden. Diagnossystemet designades utgående från dessa reducerade modeller med minimala polynombasmetoden. Simulering av fel gjordes med en linjärisering av motormodellen. Principstudien visar att modellbaserad diagnos med dynamiska modeller kan isolera små sensor- och ställdonsfel i RM12. Studien visar även att en liten och välkonditionerad designmodell ger det bästa diagnosystemet.

Nyckelord: modellbaserad diagnos, minimal polynombas, turbojetmotor

Förord

Examensarbetet har utförts på uppdrag av Volvo Aero Corporation (VAC) i samarbete med Fordonssystem på Linköpings tekniska högskola (LiTH). Vi vill tacka VAC och Fordonssystem, samt Mattias Nyberg (LiTH) och Roger Persson (VAC) som initierade examensarbetet. Vårt största tack ger vi till våra handledare Erik Frisk (LiTH) och Peder Johnson (VAC) som har varit djupt engagerade och har stöttat oss under arbetets gång. Även Mattias Nyberg har visat stort intresse och givit goda råd. Personalen på Fordonssystem tackas för en trivsamt arbetsmiljö och roliga stunder i fikarummet.

Notation

Modellspecifik notation

Namn	Beskrivning
A8	Area, utloppsmunstycke
CVG	Kompressorns variabla geometri (Compressor Variable Geometry)
EBK	Efterbrännkammare
FVG	Fläktens variabla geometri (Fan Variable Geometry)
GI	Marktomgång PLA 17 grader (Ground Idle)
M0	Machtal
NH	Varvtal hos högtrycksrotor
NL	Varvtal hos lågtrycksrotor
PLA	Gaspådrag (Power Level Angle) kan varieras mellan 17 och 130 grader.
PS0	Statiska trycket för omgivningen
PS3	Statiskt tryck hos högtrycksturbinens utlopp
PT0	Totala trycket för omgivningen
PT5	Totala trycket för lågtryckturbinens utlopp
RM12	Turbojetmotorers modellnamn (Reaktionsmotor 12)
SLS	Havsnivå (Sea Level Static)
TT0	Omgivningens temperatur
TT2.5	Kompressorns inloppstemperatur
TT3	Kompressorns utloppstemperatur
TT5	Lågtryckturbinens utloppstemperatur
W21	Luftflödet efter fläkten
WBL3	Kylluftflödet till högtrycksrotor
WFM	Bränsleflöde i brännkammare
WFR	Bränsleflöde i efterbrännkammare

Övrig notation

Namn	Beskrivning
$\mathcal{N}_L(M)$	Vänster nollrum till M där M är godtycklig matris.
A^*	Adjungerad matris till A , dvs A är transponerad och komplexkonjugerad
$G_{ab}(s)$	Överföringsfunktion $G(s)$ från a till b
$\ f\ _2$	\mathcal{H}_2 -normen.
x	Tillstånd
\dot{x}	Tidsderivatan av x
s	Laplacevariablen
u	Insignal
y	Utsignal
f	Fel
d, v	Störningar
r	Residual
J_{th}	Tröskel
rps	varv per sekund

Innehåll

1	Inledning	1
1.1	Rapportens upplägg	2
2	Allmänt om RM12	3
2.1	Turbojetmotorn RM12	3
2.2	Simuleringsmodell av RM12	4
3	Modellbaserad diagnos	7
3.1	Varför modellbaserad diagnos?	9
3.2	Fel och modellering av fel	10
3.3	Residualstruktur och felisolering	11
3.4	Residualgenerering	14
3.4.1	Generell residualgenerering	14
3.4.2	Chow-Willsky	15
3.4.3	Minimal polynombas	18
3.5	Residualevaluering	20
4	Linjärisering	23
4.1	Allmän linjärisering	23
4.2	Linjärisering av RM12 med störningsteknik	24
4.3	Val av linjäriseringspunkter för RM12	25
4.4	Utvärdering av linjärisering	26
4.5	Sammanfattning	27
5	Reducering	29
5.1	Allmänt om reducering och verifiering	29
5.2	Fysikalisk reducering	31
5.3	Balanserad reducering	32
5.4	Sammanfattning	33
6	Design	39
6.1	Modeller för diagnos	39
6.1.1	Modellering av fel	40
6.1.2	Simulering av fel	40

6.2	Test av designmetoder	40
6.2.1	Val av residualstruktur	41
6.2.2	Chow-Willsky	41
6.2.3	Minimala polynombasmetoden	42
6.3	Sammanfattning	47
7	Slutsatser och rekommendationer	49
7.1	Slutsatser	49
7.2	Rekommendationer till fortsatt arbete	50
A	Kompletterande teori	55
A.1	Tillståndsreducering av modell	55
A.2	Singulära värden	56
A.3	Reducering med balanserad tillståndsrealisering	56
A.4	Cayley-Hamiltons sats	57
B	Modellspecifikt för RM12	59
B.1	Skalningsmatriser	59
B.2	Modeller	59
B.2.1	Modell 1	60
B.2.2	Modell 2	61
B.2.3	Modell 3	62
C	Residualer	65
C.1	Residualer för modell 1	66
C.2	Residualer för modell 2	68
C.3	Residualer för modell 3	70
C.3.1	Simulering 1	70
C.3.2	Simulering 2	72

Kapitel 1

Inledning

Syftet med det här examensarbetet är att undersöka om modellbaserad diagnos med dynamisk modell kan användas för RM12. Arbetet är utfört på uppdrag av Volvo Aero Corporation (VAC) i samarbete med Fordonssystem, LiTH.

Hos VAC används idag gränskontroller av mätvärden för att upptäcka fel. Detta sker i flygplanet under drift. Gränskontrollerna baseras på en statisk modell av motorn och om någon av dessa gränser överskrids sätts en felflagga. På marken efter ett flygpass analyseras sedan trender av mätvärden på lång och kort sikt för att även kunna upptäcka långsamma förändringar i motorn. De felflaggor som har satts genererar sedan ett felmeddelande som leder till undersökning av motorn. Undersökningen görs idag till stor del manuellt av en flygtekniker som med hjälp av ett binärt sökträd kan isolera felet. Motorn måste alltså undersökas fysiskt. Ett framtida felsökningssystem kan utgå från en dynamisk motormodell, där isolering av fel sker automatiskt. Att automatiskt kunna isolera fel underlättar för flygteknikerna att hålla motorn i drift.

Eftersom ett komponentfel i en flygmotor kan leda till att flygplanet störtar är det mycket viktigt att kunna detektera eller förutsäga fel. Kan fel detekteras på ett tidigt stadium kan allvarliga incidenter förhindras och mycket pengar sparas. I dag byts en viss komponent ut efter en så kort flygtid att risken för haveri pga. utslitning är så gott som obefintlig. Byta komponenter så ofta är mycket kostsamt. Flygtiden för en komponent skulle kunna förlängas, med bibehållen säkerhet, med ett system som diagnostiserade komponenten och talade om när det är dags att byta ut den. Förhoppningen att utveckla ett sådant system ligger till grund för denna studie av modellbaserad diagnos.

Syftet med examensarbetet kan sammanfattas som:

- Undersöka om modellbaserad diagnos med dynamisk modell kan användas för feldetektering i RM12.

Målen kan sammanfattas som:

- Utföra principstudie av modellbaserad diagnos applicerad på RM12.
- Öka kunskapen hos VAC om modellbaserad diagnos.
- Utredda för- och nackdelar med modellbaserad diagnos.

1.1 Rapportens upplägg

Rapporten inleds med en kort översikt av RM12 och beskrivning av motormodellen för denna. I kapitel 3 förklaras därefter modellbaserad diagnos och grundläggande begrepp såsom residualer och isolerbarhet. Kapitel 4 beskriver linjärisering av den olinjära motormodellen. Reduceringen av den linjäriserade motormodellen beskrivs sedan i kapitel 5. De tidigare kapitlen ligger till grund för kapitel 6 där design av ett modellbaserat diagnosystem presenteras. Utvärderingen av diagnosystemet finns även i detta kapitel. Slutligen finns slutsatser och rekommendationer i kapitel 7.

Allmänt för rapporten är att teoridelen är blandad med tillämpningarna. I kapitel 4 om linjäriseringar kommer t.ex. först en teoridel som i allmänna ordalag talar om hur linjärisering går till och därefter kommer tillämpningen för vår motormodell samt en utvärdering av hur bra linjäriseringen är. Teorin för kapitel 6 finns dock i kapitel 3. Läsare med kunskaper inom modellbaserad diagnos kan hoppa direkt till kapitel 6, där de huvudsakliga resultaten presenteras.

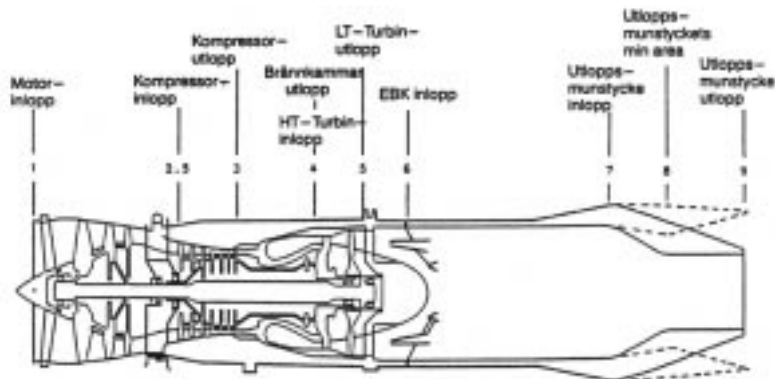
I den här rapporten är en hänvisning till (4.2) och en hänvisning till ekvation 4.2 samma sak. I allmänhet betyder $G_{ab}(s)$ en allmän överföringsfunktion $G(s)$ från a till b . Matriserna A , B , C och D är tillståndsmatriserna för en linjär modell om inte annat uttryckligen står. I de sammanhang ordet 'modell' används och kan missuppfattas kommer motormodell att användas för den modell som används vid simulering av motorn. Designmodell kommer att användas då den modell som diagnosystemet är uppbyggt av åsyftas.

Kapitel 2

Allmänt om RM12

2.1 Turbojetmotorn RM12

Turbojetmotorn RM12, som sitter i JAS 39 Gripen, är utvecklad av Volvo Aero Corporation (VAC) i samarbete med General Electric (GE). RM12 är en anpassning av motorn F404-GE-400 för att passa JAS 39 Gripen och det svenska flygvapnets behov. RM12 är en turbofläktmotor som består av en gasgenerator och en efterbrännkammare (EBK). Själva kärnmotorn består av kompressor, brännkammare och turbiner. Motorns syfte är att producera dragkraft genom att accelerera luft.



Figur 2.1: Förenklad bild av RM12

Detta åstadkomms genom att temperaturen höjs hos luften m.h.a. bränsleförbränning i gasgenerator och efterbrännkammare. För att åstadkomma hög förbränning krävs att bränslet förbränns under högt tryck. Figur 2.1 visar en förenklad bild av RM12. Motorn är numrerad framifrån och bakåt, så att en förkortning som

TT5 betyder totaltemperatur i lågtrycksturbinens utlopp. Luften leds in genom inloppet(1) till en trestegsfläkt(2), där den komprimeras. Ungefär 25% av luften leds förbi kärnmotorn(2.5–5) för att användas som kylmedium för motorns varma delar. Resterande 75% leds vidare till kompressorn(2.5) och komprimeras där i sju steg. Luften förs sedan in i den cirkulära brännkammaren(4) där syret och bränslet förbränns. Gasflödet som bildats vid förbränningen går först in i högtrycksturbinen(4.1), som via en ihålig rotor driver kompressorn. Därefter in i lågtrycksturbinen(5), som i sin tur driver fläkten. I EBK:n(6) kan sedan ytterligare bränsle förbrännas och därmed ökar dragkraften än mer. Gasen förs slutligen ut genom utloppsmunstycket(9) med hög hastighet.

RM12-motorns dragkraft regleras m.h.a. fem styrvariabler. Bränsleflödet i brännkammaren (WFM) och i efterbrännkammaren (WFR) regleras, liksom den minsta utloppsarean (A8). A8-arean påverkar dragkraften via flöde och temperatur hos utloppsgasen. Luftflödet genom motorn varierar genom att vinkeln på de ställbara ledskenorna i fläkten (FVG) och kompressorn (CVG) ändras. I tabell 2.1 finns de fem styrvariablerna uppställda. Flygföraren bestämmer motorns dragkraft med gaspådraget PLA (Power Level Angle). PLA kan varieras från 17 grader vid GI (Ground Idle) till 130 grader vid maximalt tänd efterbrännkammare. Reglersystemet ger utgående från PLA-värdet och inloppsvariablernas värde de fem styrvariablerna till motorn. Ett antal variabler (se tabell 2.2) mäts för att kunna styra motorn,

Styrvariabel	Beskrivning	Enhet
A8	Minsta arean hos utloppsmunstycket	m^2
CVG	Kompressorns ställbara ledskenor	grader
FVG	Fläktens ställbara ledskenor	grader
WFM	Bränsleflöde till brännkammaren	kg/s
WFR	Bränsleflöde till efterbrännkammaren	kg/s

Tabell 2.1: Styrvariabler för RM12

samt för att avgöra motorns kondition. Fel i dessa mätningar kan ge upphov till felaktiga slutsatser. För en fullständig beskrivning av motorn se (VAC 1991).

2.2 Simuleringsmodell av RM12

Volvo Aeros simuleringsmodell av RM12 är olinjär, dynamisk, tidskontinuerlig och framtagen m.h.a. fysikaliska samband. VAC:s erfarenheter är att den efterliknar motorn väl. Den moduluppbyggda modellen är implementerad i SYSTEMBUILD (MATRIXx) och FORTRAN. Den modell som f.n. används (RM12d) har 5 in-signal, 13 utsignal (varav 8 mätsignaler) och 32 tillstånd. De flesta olinjäriteter i modellen existerar som flerdimensionella tabeller. Tidskonstanterna för systemet varierar kraftigt från 0.005 sekunder (temperatur och gasmassa i små kontrollvolymer) till ungefär 6 minuter (metalluppvärmning), vilket innebär att modellen är

Mätvariabel	Beskrivning	Enhet
NH	Varvtal hos högtrycksrotor	rps
NL	Varvtal hos lågtrycksrotor	rps
PT1	Totaltryck vid inloppet	kPa
PT5	Totaltryck vid lågtryckturbinens utlopp	kPa
PS3	Statiskt tryck vid högtrycksturbinens utlopp	kPa
TT1	Total temperatur vid inloppet	K
TT25	Inloppstemperatur vid kompressor	K
TT5	Total temperatur vid lågtryckturbinens utlopp	K

Tabell 2.2: Mätvariabler för RM12

mycket styv. En mer detaljerad beskrivning av simuleringssmodellen finns i (Härefors 1995) och i (Johansson 1997).

Kapitel 3

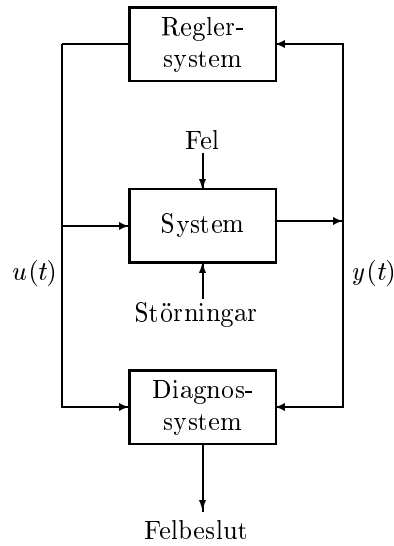
Modellbaserad diagnos

Vad är diagnos? En definition på diagnos, som håller både tekniskt och medicinskt, är att detektera och isolera fel som uppstått m.h.a. de symptom som uppträder när systemet är i drift. Detta görs med observationer och kunskap om vad som är normalt (friskt). Ett diagnossystem ska alltså generera ett felbeslut utgående från lämplig kunskap.

För tekniska applikationer utnyttjas styrsignaler (insignaler) till och mätningar (utsignaler) från det systemet som skall diagnostiseras. Den generella strukturen för ett tekniskt diagnossystem ser ut som i figur 3.1. Insignaler till diagnossystemet är u (insignal till systemet som ska diagnostiseras) och y (mätningar av utsignal från systemet som ska diagnostiseras). Utsignal från diagnossystemet är ett felbeslut som innehåller information om när felet inträffade (detektera fel) och var det inträffade (isolera fel). Att avgöra felets storlek och tidsvariation kallas för felidentifiering. Systemet som diagnostiseras påverkas förutom av fel även av störningar i form av modellosäkerheter och brus, vilket försvårar diagnostiseringen avsevärt.

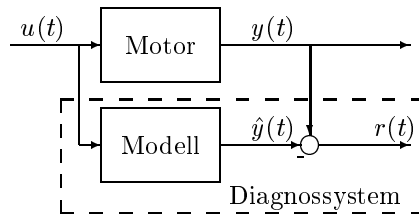
Intressanta områden för tillämpningar på diagnos:

- Kemiska industrier
- Kärnreaktorer
- Gasturbiner
- Autonoma robotar
- Elektriska motorer
- Avgassystem i bilmotorer
- Delsystem i flygplan och helikoptrar, såsom t.ex. reglersystem och navigeringssystem.



Figur 3.1: Generell struktur på ett diagnosystem.

Skälen till att använda diagnos är att öka säkerheten, skydda systemet från att gå sönder, förenkla reparationsarbetet, skydda miljön med minskade utsläpp, samt öka tillgängligheten hos systemet. För att öka prestandan hos ett system kan diagnos användas vid reglering genom att kompensera ut fel hos t.ex. sensorer. Hur designas ett diagnosystem? Genom att utnyttja mätningar från ett system och jämföra dessa med en modell av systemet kan fel detekteras och isoleras, dvs diagnos kan göras. Ett enkelt exempel på detta är figur 3.2 där motorns och modellens utsignal jämförs (y respektive \hat{y}). Om dessa skiljer sig åt har ett fel inträffat. Någon form

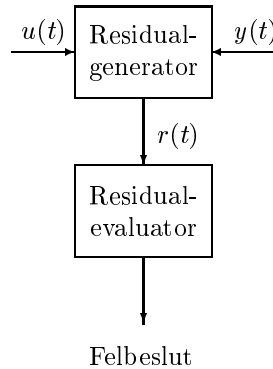


Figur 3.2: Ett enkelt diagnosystem

av redundans måste existera för att avgöra om fel har inträffat. Signalen r som bildats kallas för *residual*. En residual kan definieras som i (Nyberg 1997) :

Definition 3.1 (Residual) *En residual $r(t)$ är en skalär eller vektor som är 0 (eller liten) i det felfria fallet och $\neq 0$ (eller stor) när fel inträffar.*

Residualerna används för att detektera och isolera fel i systemet. De genereras av en residualgenerator, se figur 3.3, som har in- och utsignalerna till systemet (u respektive y) som insignaler. Sedan avgör residualevaluatorn, utgående från residualerna, om ett fel har inträffat, dvs den ger ett felbeslut.



Figur 3.3: Intern struktur på ett diagnosystem.

3.1 Varför modellbaserad diagnos?

En enkel metod för diagnos är att kontrollera att mätdata från systemet håller sig inom ett visst intervall. Om mätdata ligger utanför detta intervall genereras ett felmeddelande. Denna metod kan även tillämpas på trender hos mätdata för att studera långsamma förändringar i systemet och ur dessa förändringar avgöra om något fel inträffat. En annan vanlig metod är användandet av multipla sensorer för att uppnå redundans och därmed kunna avgöra om en sensor har gått sönder. Att använda multipla sensorer kallas *hårdvaruredundans*. Förutom att hårdvaruredundans kräver extra komponenter, som ökar både vikt och kostnad hos systemet, kräver dubbla sensorer även ett modellbaserat diagnossystem för isolering. Trippla sensorer kan däremot isoleras fel med enkel logik.

Vad är då fördelarna med modellbaserad diagnos jämfört med de ovan nämnda metoderna?

- Högre diagnosprestanda jämfört med traditionell gränskontroll. Mindre fel kan isoleras och detektionstiden minskar.
- *Isolering* av fel, dvs felet kan lokaliseras till en viss komponent.
- Diagnosen kan utföras *passivt*, dvs systemet behöver inte exciteras för att isolera fel utan felen kan isoleras vid normal drift i hela området där modellen är giltig.

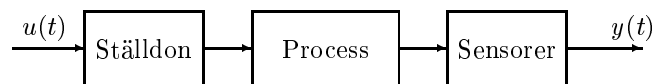
- *Kompensera för störningar*, vilket gör diagnossystemet robustare.

En nackdel med modellbaserad diagnos är att det ställer höga krav på bra överensstämmelse mellan designmodell och system. Eftersom diagnossystemet arbetar utan återkoppling ställer detta högre krav på designmodellen än vid reglering då reglersystemet har möjlighet till återkoppling. Men med en bra designmodell av systemet kan alla de fördelar som modellbaserad diagnos har att erbjuda utnyttjas.

3.2 Fel och modellering av fel

I detta och efterföljande avsnitt kommer kortfattat grunderna till modellbaserad diagnos för linjära system gås igenom. För en djupare genomgång hänvisas till (Nyberg 1997), vilket nedanstående beskrivning av modellbaserad diagnos grundar sig på.

För att använda modellbaserad diagnos måste de fel som ska isoleras modelleras, dvs en modell för hur felen påverkar systemet behövs. Indelningen av felen kan göras efter de delsystem som felen uppträder i (se figur 3.4): *ställdonsfel*, *processfel* och *sensorfel*. För ställdon och sensorer är kortslutning, drift och bias tre av de

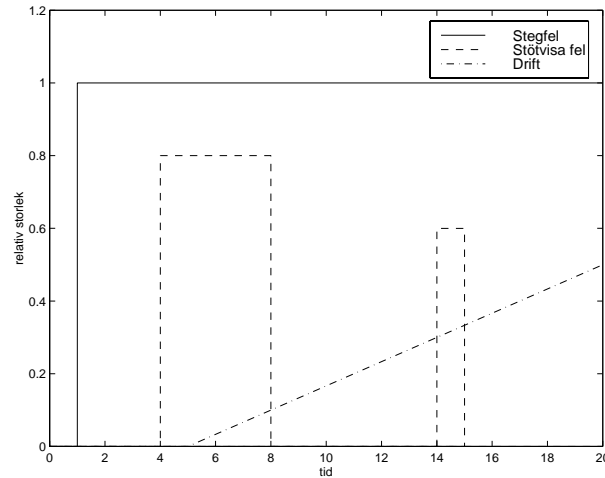


Figur 3.4: Generell struktur på ett system.

vanligaste felen. Hur processfel ska modelleras beror naturligtvis på vilket system det handlar om. Vanliga processfel är ökad friktion, läckage och förslitningar i processen (t.ex. bladfel i en jetmotor). I figur 3.5 åskådliggörs exempel på tre typer av fel:

- *Stegfel*, ex. kortslutning eller något går sönder abrupt.
- *Drift*, ex. åldrande, yttre temperaturpåverkan eller förslitning.
- *Stötvisa fel*, ex. lösa kablar eller glappkontakt.

Modellering av fel måste ske med åtanke på hur de förväntas uppträda. Generellt kan felpåverkan delas upp i tre fall:



Figur 3.5: Tre olika feltyper.

- Additivt fel: $y = Cx + f$
- Multiplikativt fel : $y = Cxf$
- Parameterfel: $y = (C + f)x$

där y är utsignal, x är tillstånd eller insignal, C är överföringsmatrisen från x till y och f är felet. Driftfel kan beskrivas som additivt fel, kortslutning som multiplikativt fel och fysikaliska fel, såsom ökad friktion, som parameterfel.

Av nämnda feltyper är processfel svårast att modellera. Sensorfel kan däremot med fördel modelleras som additiva fel på utsignalen. Ställdonsfel kan också på ett enkelt sätt modelleras som additiva fel på insignalen och får då samma dynamik som insignalerna. Processfelen kräver dock en ingående kunskap om processen och om de fel som ska modelleras.

3.3 Residualstruktur och felisolering

För att isolera flera fel krävs flera olika residualer. För att kunna skilja på felen utgående från residualerna behövs en systematisering av dessa göras. Denna systematisering kallas *residualstruktur* och är en förteckning över vilka residualer, r_i , som är känsliga för de olika felen, f_j (se tabell 3.1). En 1:a på position (i, j) i residualstrukturen innebär att residual r_i är känslig för fel f_j , en 0:a att den är okänslig. Observera att residualstrukturen inte säger något om hur känsliga residualerna är för ett visst fel. Struktur C i tabell 3.1 är ett enkelt exempel på en residualstruktur. Om residual 1 och 3 (r_1 och r_3) är stora, medans residual 2 (r_2) är liten, indikerar struktur C att fel 2 (f_2) ha inträffat. Residualstrukturen används vid design av residualgeneratorerna och vid isolering av fel. I struktur C förutsätts

A	f_1	f_2	f_3	B	f_1	f_2	f_3
r_1	1	1	0	r_1	1	1	0
r_2	1	1	1	r_2	1	0	1
r_3	1	1	1	r_3	1	1	1
C	f_1	f_2	f_3	D	f_1	f_2	f_3
r_1	0	1	1	r_1	1	0	0
r_2	1	0	1	r_2	0	1	0
r_3	1	1	0	r_3	0	0	1

Tabell 3.1: Exempel på residualstrukturer.

att endast ett fel kan inträffa åt gången för att fel ska kunna isoleras. En sådan struktur är i praktiken inte någon större begränsning då oftast ett specifikt fel först uppträder och därefter uppträder följdfelet i resten av processen. Om fler än ett fel uppkommer samtidigt kan det oftast detekteras att något fel har inträffat men ej isoleras. Däremot kan inte fel isoleras. I struktur C blir alla residualer stora om två (eller fler) fel inträffar samtidigt. Slutsatsen att fel har inträffat kan då dras.

Isolering av fel

För att kunna isolera felen utgående från residualstrukturen och residualerna måste felen ha unika kolumner i residualstrukturen. Alla elementen i kolumnerna kan inte heller vara noll samtidigt. Är dessa villkor uppfyllda är residualstrukturen *svagt isolerbar*.

När ett fel inträffar kan det dock hända att en residual som borde ge utslag inte gör detta. I en struktur som enbart är svagt isolerbar betyder detta att ett fel som ej inträffat felaktigt kan isoleras. Om residualstrukturen ser ut som B i tabell 3.1 tror vi felaktigt att fel f_2 inträffar om residual r_2 ej ger utslag då fel f_1 inträffar. För att göra isoleringen mer robust utformas istället residualstrukturen så att en 1:a, vilken som helst, kan bli en 0:a utan att två kolumner blir identiska. Har residualstrukturen utformats på detta sätt benämns den som *starkt isolerbar*. En starkt isolerbar struktur är även svagt isolerbar. Med en starkt isolerbar struktur upptäcks om en residual inte ger tillräckligt utslag istället för att ej avsett fel isoleras, vilket kan ske i en svagt isolerbar struktur. Isolering av ej avsett fel kan undvikas pga att residualerna svarar på ett sätt som ej är förknippat med något av felen. I tabell 3.1 är A icke isolerbar, B svagt isolerbar, C och D starkt isolerbara. En starkt isolerbar residualstruktur är mer robust mot modellosäkerheter och därför önskvärd.

Val av residualstruktur

En residualstruktur kan se ut på en mängd olika sätt. I tabell 3.1 visas fyra varianter av residualstrukturer. Varför är andra strukturer än D i tabell 3.1 intressanta, är

det inte enklast om varje fel är förknippad med en residual, så att när fel i inträffar reagerar endast residual i ? Svaret på detta är: Ja, visst är struktur D enklare men det finns en del problem med en sådan struktur. Residualstrukturen kan vara omöjlig för designmetoden att framställa pga att de många fel som residualen ska vara okänslig för inte kan avkopplas^a från residualen. Avkoppling är svårt att få, eftersom en modell aldrig överensstämmer perfekt med ett riktigt system. Givetvis är det enklare att avkoppla ett fel än flera.

Ett annat problem uppstår när ytterligare ett fel ska adderas till ett redan färdigt diagnosystem. Adderingen av ett fel får som följd att en residual i struktur D inte längre är känslig för endast ett fel, förutsatt att arbetet inte görs om. I struktur C uppkommer inga sådana problem. För att kunna isolera fel under dessa omständigheter är struktur C därför bättre, då struktur C gör diagnosystemet robustare mot modellfel och störningar än D.

Don't care

Modellosäkerheter och störningar kan resultera i att residualerna blir skilda från noll i det felfria fallet vilket gör det svårare att avgöra när fel inträffat. Residualernas känslighet för de olika felen varierar också, vilket visar sig då ett visst fel kan ge ett stort utslag på residualen medans ett annat fel ger ett litet utslag. Detta beror på att förstärkningen inte är lika stor från alla fel till residual. I residualstrukturen innebär en 1:a endast att residualen är känslig för ett visst fel och inte hur känslig den är. För att göra det lättare att isolera fel kan *don't care* eller X i residualstrukturen införas (Nyberg & Nielsen 1997a, Nyberg & Nielsen 1997b), se figur 3.2. *Don't care* i residualstrukturen innebär att ingen hänsyn tas till om det blir en 1:a eller en 0:a, dvs om residualen blir stor eller inte. *Don't care* medför ökad robusthet för modellosäkerheter och störningar, samt minskar sannolikheten för missat felalarm eller falskt felalarm. Svag och stark isolerbarhet fungerar på samma sätt som tidigare, fast nu med X som både 0:a och 1:a. Placeringen av

	f_1	f_2	f_3
r_1	X	1	0
r_2	X	X	1
r_3	1	0	X
r_4	0	1	1

Tabell 3.2: Exempel på en residualstruktur med *don't care*.

don't care kan göras metodiskt (Nyberg 1998). En av dessa metoder ersätter alla utom en 1:a i varje residual med X och sedan skapas nya residualer genom att variera var 1:an bibehålls, se tabell 3.3. Antalet residualer växer då snabbt, men genom att göra på detta sätt ökas robustheten i felbeslutet.

Överföringsfunktionen kan formas från fel till residual efter behag, då residualen endast är känslig för ett fel. Formningen kan i det ideala fallet göras på ett sådant

^aAvkoppling innebär att göra en residual okänslig för ett fel.

sätt att överföringsfunktionen från fel till residual blir 1 (Frisk, Nyberg & Nielsen 1997). En sådan överföringsfunktion kan göra residualevalueringen lättare.

$$\begin{array}{c|ccc} & f_1 & f_2 & f_3 \\ \hline r_1 & 0 & 1 & 1 \\ r_2 & 1 & 0 & 1 \\ r_3 & 1 & 1 & 0 \end{array} \implies \begin{array}{c|ccc} & f_1 & f_2 & f_3 \\ \hline r_{11} & 0 & 1 & X \\ r_{12} & 0 & X & 1 \\ r_{21} & 1 & 0 & X \\ r_{22} & X & 0 & 1 \\ r_{31} & 1 & X & 0 \\ r_{32} & X & 1 & 0 \end{array}$$

Tabell 3.3: Utvidgning av residualstrukturen med *don't care*.

3.4 Residualgenerering

Det finns en mängd olika sätt att designa linjära residualgeneratorer på. *Chow-Willsky* (Chow & Willsky 1984) och diagnosobservatörer är två av de vanligaste designmetoderna. En tredje och ny designmetod kallad *minimala polynombasmetoden* är en annan metod (Frisk & Nyberg 1998b). Sistnämnda metod har egenskapen minimalitet vilket de andra metoderna saknar. Egenskapen minimalitet innebär möjligheten till enkla och lättberäknade residualgeneratorer. Att designmetoden ger en fullständig lösning innebär att metoden hittar alla residualgeneratorer. Det är egenskapen minimalitet som gör minimala polynombasmetoden intressant. De nämnda metoderna designar residualgeneratorer utgående från signalrummet men det finns även andra tillvägagångssätt att skapa residualer, då främst parameterskattningar se (Isermann 1993). Inga metoder utöver de som utgår från signalrummet kommer att tas upp här.

3.4.1 Generell residualgenerering

En generell residualgenerator kan se ut på följande sätt

$$r = H_{yr}(s)y + H_{ur}(s)u = Q(s) \begin{pmatrix} y \\ u \end{pmatrix} \quad (3.1)$$

där u är insignalen till systemet, y är utsignalen och H och Q är överföringsfunktioner. Ekvation 3.1 benämns som residualgeneratorns *beräkningsform* då det är denna form som används vid beräkningen av residualen. Om $G_{uy}(s)$ är överföringsfunktionen från insignal till utsignal, $G_{fy}(s)$ är överföringsfunktionen från fel, f , till utsignal och $G_{dy}(s)$ är överföringsfunktionen från störning, d , till utsignal. Då fås följande utseende på utsignalen:

$$y = G_{uy}(s)u + G_{fy}(s)f + G_{dy}(s)d \quad (3.2)$$

Ersätt nu y i (3.1) med (3.2) och följande ekvation erhålls:

$$\begin{aligned} r &= (H_{yr}(s)G_{uy}(s) + H_{ur}(s))u + H_{yr}(s)G_{fy}(s)f + H_{yr}(s)G_{dy}(s)d = \\ &= Q(s) \begin{bmatrix} G_{uy}(s) & G_{dy}(s) \\ I & 0 \end{bmatrix} \begin{bmatrix} u \\ d \end{bmatrix} + Q(s) \begin{bmatrix} G_{fy}(s) \\ 0 \end{bmatrix} f = 0 \end{aligned} \quad (3.3)$$

Där I är identitetsmatrisen. Enligt definition 3.1 på en residual måste i det felfria fallet r vara noll för alla u . Detta medför att enligt (3.3) måste följande gälla:

$$Q(s) \begin{bmatrix} G_{uy}(s) & G_{dy}(s) \\ I & 0 \end{bmatrix} = Q(s)M(s) = 0 \quad (3.4)$$

Om (3.4) uppfylls beror residualen endast av felet f . Det vill säga om $Q(s)$ ligger i vänster nollrum till $M(s)$. För att fel ska kunna detekteras krävs det även att överföringsfunktionen från varje fel till residual är skild från noll, se (3.5).

$$Q(s) \begin{bmatrix} G_{fy}(s) \\ 0 \end{bmatrix} \neq 0 \quad (3.5)$$

I praktiken existerar det emellertid alltid störningar och modellosäkerheter som inte kan kompenseras bort. Residualen kan då uttryckas i sådana störningar och modellosäkerheter, d , tillsammans med de modellerade felen, f , på följande sätt:

$$r = G_{fr}(s)f + G_{dr}(s)d \quad (3.6)$$

Ovanstående uttryck kallas *intern form*. Ur denna form visas hur ett fel eller en störning påverkar residualen. En välkonstruerad residualgenerator har en hög förstärkning från fel till residual, $G_{fr}(s)$ och en låg förstärkning från störning till residual $G_{dr}(s)$, samtidigt som den fungerar enligt vald residualstruktur.

3.4.2 Chow-Willsky

Bland de många metoder att designa residualgeneratorer för linjära system är en vanlig metod *Chow-Willsky* (Chow & Willsky 1984) eller utvidgning till ULPE^b (Nyberg & Nielsen 1997b). Dessa metoder klarar av att avkoppla ett begränsat antal störningar helt från residualen. *Huvudmålet* för designmetoden är, som tidigare, att hitta en residualgenerator som gör att:

- Överföringsfunktionen från observerade fel till residual är skild från 0.
- Överföringsfunktionen från insignal, störningar och icke-observerade fel blir noll, dvs perfekt avkoppling. (Eller i alla fall liten jämfört med ovanstående överföringsfunktion.)

^bUniversal Linear Parity Equation, vilken även tar hänsyn till att felen kan ha annan dynamik än processen.

För att kunna definiera en residualgenerator har (Chow & Willsky 1984) definierat en paritetsfunktion enligt följande:

Definition 3.2 (Paritetsfunktion) *En paritetsfunktion är en funktion som kan skrivas som*

$$h(u(t), y(t)) = A(\sigma)y(t) + B(\sigma)u(t) \quad (3.7)$$

där $A(\sigma)$ och $B(\sigma)$ är radvektorer med polynom i σ , $u(t)$ och $y(t)$ är systemets insignal resp. utsignalvektorer, samt σ betecknar deriveringsoperatorn p eller tidsförskjutningsoperatorn q . Denna funktion är noll om inga fel existerar.

Definitionsmässigt behöver alltså inte paritetsfunktionen vara skild från noll då ett fel inträffar men praktiskt sett måste den vara det för att vara en bra residualgenerator. En linjär residualgenerator kan uttryckas som (3.8), jämför (3.1).

$$r(u(t), y(t)) = \frac{h(u(t), y(t))}{C(\sigma)} = \frac{A(\sigma)y(t) + B(\sigma)u(t)}{C(\sigma)} \quad (3.8)$$

Polynom $C(\sigma)$ tillför en frihetsgrad i valet av filter och gör detta realiserbart då gradtalet hos C är samma eller större än A och B :s gradtal. Polynomet $C(\sigma)$ kan även fungera som ett lågpasfilter för att avlägsna högfrekvent brus. Chow-Willsky:s metod realiserar på följande sätt. Betrakta systemet (3.9) på tillståndsform

$$\sigma x = Ax + Bu + Ev + Kf \quad (3.9a)$$

$$y = Cx + Du + Jv + Lf \quad (3.9b)$$

där

x : n -dimensionell tillståndsvektor

y : m -dimensionell utsignalvektor

u : k -dimensionell insignalvektor

v : k_d -dimensionell störningsvektor

f : k_f -dimensionell felvektor

A, B, C, D, E, J, K, L : konstanta matriser

Ekvation 3.9 ger

$$\begin{aligned} \sigma y &= C\sigma x + D\sigma u + J\sigma v + L\sigma f = \\ &CAx + CBu + D\sigma u + CEv + CKf + J\sigma v + L\sigma f \end{aligned} \quad (3.10)$$

På samma sätt fås för $\sigma^2 y \dots \sigma^\rho y$, att systemet (3.9) kan skrivas som:

$$Y(t) = Rx(t) + Q_c U(t) + HV(t) + PF(t) \quad (3.11)$$

där

$$Y = \begin{bmatrix} y \\ \sigma y \\ \vdots \\ \sigma^\rho y \end{bmatrix} \quad R = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^\rho \end{bmatrix}$$

$$Q_c = \begin{bmatrix} D & 0 & 0 & \dots \\ CB & D & 0 & \dots \\ \vdots & & \ddots & \\ CA^{\rho-1}B & \dots & CB & D \end{bmatrix} \quad U = \begin{bmatrix} u \\ \sigma u \\ \vdots \\ \sigma^\rho u \end{bmatrix}$$

$$H = \begin{bmatrix} J & 0 & 0 & \dots \\ CE & J & 0 & \dots \\ \vdots & & \ddots & \\ CA^{\rho-1}E & \dots & CE & J \end{bmatrix} \quad V = \begin{bmatrix} v \\ \sigma v \\ \vdots \\ \sigma^\rho v \end{bmatrix}$$

$$P = \begin{bmatrix} L & 0 & 0 & \dots \\ CK & L & 0 & \dots \\ \vdots & & \ddots & \\ CA^{\rho-1}K & \dots & CK & L \end{bmatrix} \quad F = \begin{bmatrix} f \\ \sigma f \\ \vdots \\ \sigma^\rho f \end{bmatrix}$$

Storleken på Y är $(\rho+1)m \times 1$, R är $(\rho+1)m \times n$, Q_c är $(\rho+1)m \times (\rho+1)k$, U är $(\rho+1)k \times 1$, H är $(\rho+1)m \times (\rho+1)k_d$, F är $(\rho+1) \times 1$, P är $(\rho+1)m \times (\rho+1)$, och V är $(\rho+1)k_d \times 1$. Cayley-Hamiltons sats ger att ingen extra information vinnas på att välja ρ större än n , eftersom systemet fullständigt kan uttryckas med de n lägsta ordningstalen (se A.4)^c. Paritetsfunktionen kan nu med en kolonnvektor w av längd $(\rho+1)m$ bildas som (analogt för residualgeneratorn, jämför (3.8))

$$h(u, y) = w^T(Y - Q_c U) = w^T(Rx + HV + PF) \quad (3.12)$$

Den sista likheten följer av (3.11). I (3.12) ses att om residualgeneratorn ska vara noll i det felfria fallet (för alla x och V), så måste w uppfylla

$$w^T [R \ H] = 0 \quad (3.13)$$

Jämför (3.4) i generell residualgenerering där $Q(s)$ söks sådant att det ligger i vänster nollrum till $M(s)$. För att kunna detektera fel krävs även att residualgeneratorn är skild från noll då fel inträffar, vilket implicerar att

$$w^T P_i \neq 0, i = 1, 2, \dots, k_f \quad (3.14)$$

^cSanning med modifikation. Principiellt kan detta dock tolkas så, för ett fullständigt bevis se (Nyberg & Nielsen 1997b).

där P_i är P -matrisen associerad med fel i . Kraven på w är alltså att den ska ligga i vänster nollrum till $[R \ H]$ samtidigt som $w^T P$ har nollskilda komponenter (jämför (3.5)). Om inte w kan väljas sådan att (3.14) är uppfylld för alla fel i innebär det att dessa fel inte är detekterbara. Att fel inte kan detekteras inträffar då för många störningar avkopplas på en gång eller om ett fel kan uttryckas som en linjärkombination av störningar. (I det senare fallet upptäcks felet om $f \gg v$.) *Beräkningsformen* för residualgeneratorn blir (jämför (3.1))

$$h = w^T(Y - Q_c U) \quad (3.15)$$

När w tas fram kan denna (oftast) väljas på en rad olika sätt. Vad som är viktigt att analysera är hur varje fel och dess derivator förstärks för ett visst w . En sådan analys kan göras genom att studera (3.14). Med en given residualstruktur bestäms nollrummet genom att de fel som residualen ska vara okänslig för avkopplas, dvs felen betraktas som störningar. Residualerna kan vara antingen statiska (t.ex. $y - 2u$) eller dynamiska (t.ex. $\dot{y} - \dot{u} - 2u$). Det är dock viktigt att vara försiktig vid val av polynomet $C(\sigma)$ så att inte viktig information filtreras bort.

Om det inte går att avkoppla störningarna helt är det lämpligt att använda sig av något optimeringskriterium. Ett kriterium är att maximera felets (F) inverkan på residualen samtidigt som störningarnas (V) inverkan på residualen minimeras. Detta med något mått^d och under bivillkoret att $w^T R$ är noll.

3.4.3 Minimal polynombas

Att skapa en uppsättning residualgeneratorer är inte speciellt svårt. Det svåra är att hitta de residualgeneratorer som endast är känsliga för fel och att dessutom göra dessa av minimal ordning. *Polynombasmetoden* hittar alla residualgeneratorer i minimal form och av minimal parametrering. Detta gör det möjligt att hitta residualgeneratorer med önskvärda egenskaper.

Polynombaser

Teorin för polynombaser är väletablerad och finns beskriven i flera böcker. En bra referens är (Kailath 1980). I det här avsnittet presenteras kortfattat en del av den teori som gäller för polynombaser. Det mesta av nedanstående är hämtat ur artikeln (Frisk & Nyberg 1998b).

Polynombaser är som namnet antyder en bas bestående av polynom. I detta fall är det laplacepolynom. En polynombas med raderna som basvektorer kan se ut på följande sätt

$$\begin{pmatrix} s^2 + s + 1 & s + 3 & s \\ s + 2 & s + 4 & 1 \end{pmatrix} \quad (3.16)$$

^dT.ex. $J = \max_w \frac{\|w^T P\|_2^2}{\|w^T H\|_2^2}$. Mha *matrix pencils* erhålls ett w som optimerar J (Frisk & Nyberg 1997).

Om en vektor kan uttryckas som en linjärkombination av de andra vektorerna är de beroende och bildar därmed tillsammans inte någon bas. Koefficienterna i linjärkombinationerna behöver inte vara skalärer utan de kan vara polynom. Ekvation 3.17 visar två vektorer som är beroende, däremot är vektorerna i (3.16) (och även i (3.18a)) oberoende eftersom de inte kan uttryckas som linjärkombination av varandra. De är därmed båda två baser. Men eftersom (3.16) och (3.18a) kan uttryckas i varandra är de baser för samma rum.

$$\begin{pmatrix} s+2 & s+4 & 1 \\ (s+1)(s+2) & (s+1)(s+4) & (s+1) \end{pmatrix} \quad (3.17)$$

Graden på en polynomvektor är definierad som det största gradtalet på polynomen i vektorn. För den första vektorn i (3.16) är graden 2 och för den andra är graden 1. Graden på en polynombas kan definieras som:

Definition 3.3 (Graden på en polynombas) Låt raderna i $F(s)$ bilda en bas för vektorrummet \mathcal{F} och μ_i vara radgraden i $F(s)$. Då definieras graden för basen $F(s)$ som $\sum \mu_i$.

Enligt definition 3.3 blir gradtalet 3 för basen i (3.16), samt 5 för basen i (3.18a). En minimal polynombas för \mathcal{F} är en bas som har minimal grad, alltså är det polynombasen i (3.16) som är minimal. Att (3.18a) inte är en minimal polynombas inses då den kan uttryckas i basen (3.16), se (3.18b).

$$\begin{pmatrix} s^2+s+1 & s+3 & s \\ s^3+s^2+2s+2 & s^2+4s+4 & s^2+1 \end{pmatrix} = \quad (3.18a)$$

$$= \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix} \begin{pmatrix} s^2+s+1 & s+3 & s \\ s+2 & s+4 & 1 \end{pmatrix} \quad (3.18b)$$

En annan önskvärd egenskap hos en polynombas är oreducerbarhet. Har en polynombas full rang^e för alla s så är den oreducerbar (Frisk & Nyberg 1998b).

Teorem 3.1 Om basvektorena i $F(s)$ utgör en oreducerbar bas för \mathcal{F} så kan alla polynomvektorer $f(s) \in \mathcal{F}$ skrivas som $f(s) = \phi(s)F(s)$ där $\phi(s)$ är en polynomvektor.

Genom att variera $\phi(s)$ kan alltså hela mängden $F(s)$ tömmas ut.

Utformning av residualgenerator från en minimalpolynombas

En generell residualgenerator kan skrivas på följande sätt se (3.3)

$$r = Q(s) \underbrace{\begin{bmatrix} G_{uy}(s) & G_{dy}(s) \\ I & 0 \end{bmatrix}}_{M(s)} \begin{bmatrix} u \\ d \end{bmatrix} + Q(s) \begin{bmatrix} G_{fy}(s) \\ 0 \end{bmatrix} f \quad (3.19)$$

^eRangen till en matris A kan definieras som dimensionen hos A 's värdeförråd. Den är lika med antalet linjärt oberoende rad- eller kolonnvektorer i A . Om A 's rang är lika med minsta antalet rad- eller kolonnvektorer i A , sägs A ha full rang.

I ovanstående ekvation har systemet delats upp i en del som avkopplas och en del som residualen ska vara känslig för. Om en minimal polynombas $F(s)$ spänner upp vänster nollrum till $M(s)$, $\mathcal{N}_L(M(s))$, och $F(s) \notin \mathcal{N}_L([G_{fy}^T(s) \ 0]^T)$ så kan alla tänkbara $Q(s)$ skapas ur $F(s)$ enligt teorem 3.1. Definitionen på en residualgenerator är att $r(t) = 0$ när $f(t) = 0$, samt att $r(t) \neq 0$ när $f(t) \neq 0$, vilket garanteras om $Q(s)$ ligger i $M(s)$ vänstra nollrum samtidigt som $Q(s)G_{fy}(s)$ har nollskilda komponenter. Uppfylls dessa krav så är $Q(s)$ en residualgenerator. Kärnan i minimala polynombasmetoden är att kunna hitta $F(s)$ och garantera att den är minimal och fullständig.

Valet av den bästa residualgeneratoren ur $F(s)$ är lite svårare. Tidigare har nämnts att en residualgenerator bör förstärka fel och undertrycka brus. När en metod att bedöma dessa egenskaper formulerats används denna metod för att välja ut $Q(s)$ från den minimala polynombasen $F(s)$. Detta urval kan uttryckas som $Q(s) = \phi(s)F(s)$. Matrisen $\phi(s)$ väljer alltså ut den bästa polynomkombinationen av basvektorer ur $F(s)$. Det som återstår är att göra $Q(s)$ realiserbar. För att åstadkomma detta kan ett lågpasfilter D_{LP} av samma eller högre relativt gradtal^f än $\phi(s)F(s)$:s gradtal väljas. Lågpasfiltret är även bra ur brussynpunkt eftersom det kapar höga frekvenser. Jämför med polynom $C(\sigma)$ i avsnitt 3.4.2. $Q(s)$ får nu följande utseende.

$$Q(s) = D_{LP}(s)\phi(s)F(s) \quad (3.20)$$

Nu har residualgeneratoren tagits fram och är färdig att användas i (3.1). Stegen dit kan *sammanfattas* på följande sätt:

1. Ta fram vänster nollrum $F(s)$ till $M(s)$.
2. Välj den "bästa" kombinationen av basvektorerna, $\phi(s)$.
3. Gör systemet realiserbart t.ex. med ett lågpasfilter, $D_{LP}(s)$.

3.5 Residualvaluering

I residualvaluatorn ska, givet r , ett felbeslut tas, dvs om fel inträffar ska felet isoleras. Residualvaluatorn avgör när residualerna är stora och jämför residualerna med residualstrukturen för att kunna isolera felen.

Hur avgörs att en residual är stor? Det enklaste sättet är att sätta en statistisk tröskel J_{th} . Överstiger residualen tröskeln dvs. om $|r(t)| > J_{th}$ så signalerar residualgeneratoren att fel har inträffat. I det ideala fallet, dvs då modellen överensstämmer exakt med verkligheten, vore det möjligt att sätta tröskeln till 0. Detta

^fSkillnaden mellan nämnarens och täljarens gradtal.

är dock orimligt pga modellosäkerheter och störningar. Tröskelsättningen får istället vara en avvägning mellan storleken på felen som ska isoleras och risken för falsklarm. Mer sofistikerade sätt för att avgöra om en residual är stor är t.ex. :

- Fuzzy tröskelsättning (Frank 1994).
- Statistiska test av abrupta förändringar (Basseville & Nikiforov 1993).
- Adaptiv tröskelsättning (Ding & Frank 1991, Höfling & Isermann 1996).

Kapitel 4

Linjärisering

Varför linjäriseras en modell? Teorin för linjära system är mer utvecklad än teorin för olinjära system. Inom teorin för linjär diagnos finns det en mängd systematiska designmetoder för att hitta residualgeneratorer. Så vitt författarna vet, finns ännu inte någon liknande systematisk designmetod för generella olinjära system. De metoder som däremot finns är av *ad hoc* karaktär, vilket kräver ingående kunskap om systemet. Vid linjär diagnos kan fel avkopplas, vilket generellt är mycket svårare i det olinjära fallet. Att använda linjär diagnos är därför mycket enklare. För att kunna göra linjär diagnos krävs en linjär designmodell.

4.1 Allmän linjärisering

En generell tillståndsbeskrivning av ett olinjärt system kan se ut på följande sätt (tidsargumentet för x , y och u utelämnas)

$$\dot{x} = f(x, u) \quad (4.1a)$$

$$y = h(x, u) \quad (4.1b)$$

där

x : tillstånd, n -dimensionell kolonnvektor

y : utsignal, m -dimensionell kolonnvektor

u : insignal, k -dimensionell kolonnvektor

f, h : olinjära funktioner av x och u

Om funktionerna f och h kan skrivas som linjära funktioner av x och u är tillståndsrepresentationen linjär. Linjäriseringen av ett olinjärt system sker oftast

kring en jämviktspunkt (x_0, u_0) för systemet sådant att $f(x_0, u_0) = 0$ (och detta ger $y_0 = h(x_0, u_0)$). Sker inte linjäriseringen kring en jämviktspunkt blir linjäriseringen tidsvariant. Om f och h är differentierbara i en omgivning av jämviktspunkten gäller att systemet (4.1) kan linjäriseras som

$$\Delta \dot{x} = A\Delta x + B\Delta u \quad (4.2a)$$

$$\Delta y = C\Delta x + B\Delta u \quad (4.2b)$$

där

$$A = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}, \quad B = \begin{pmatrix} \frac{\partial f_1}{\partial u_1} & \cdots & \frac{\partial f_1}{\partial u_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial u_1} & \cdots & \frac{\partial f_n}{\partial u_k} \end{pmatrix}$$

$$C = \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_n} \end{pmatrix}, \quad D = \begin{pmatrix} \frac{\partial h_1}{\partial u_1} & \cdots & \frac{\partial h_1}{\partial u_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_m}{\partial u_1} & \cdots & \frac{\partial h_m}{\partial u_k} \end{pmatrix} \quad (4.3)$$

A, B, C och D i (4.3) är partialderivatamatriser till $f(x, u)$ respektive $h(x, u)$ och ska utvärderas i (x_0, u_0) . $\Delta x = x - x_0$, $\Delta y = y - y_0$ samt $\Delta u = u - u_0$ är avvikelserna från jämviktspunkten. Linjäriseringen approximerar systemet i en omgivning av jämviktspunkten.

4.2 Linjärisering av RM12 med störningsteknik

Den olinjära motormodellen av RM12 innehåller olinjäriteter bl.a. i form av tabeller. Dessa tabeller gör att $f(x, u)$ ej är analytisk, med följden att det är omöjligt att hitta jämviktspunkter och att beräkna partialderivatorna i (4.3) analytiskt. Approximativa metoder måste istället användas.

Problemet med att lösa $f(x, u) = 0$ vid linjäriseringen avhjälps genom att låta motormodellen stabilisera sig under 20 sekunder (10 sekunder är tillräckligt enligt (Lind 1998)) med konstanta insignaler. Sedan antas att aktuell arbetspunkt (ett PLA-värde som ger en viss kombination av insignaler) är en jämviktspunkt. Detta antagande kontrolleras genom att störa systemet bort från arbetspunkten och se om det ramlar tillbaka till arbetspunkten. Enligt kapitel 2 finns dynamik med tidskonstanter upp till 6 minuter. Dessa tillstånd påverkar dock motormodellen i mycket liten grad och antas vara oföränderliga.

De partiella derivatorna approximeras med en numerisk störningsteknik (Härefors 1995). För varje x_j - och u_j -element läggs en liten störning (Δx_j och Δu_j) på under tiden som de andra elementen hålls konstanta. De partiella derivatorna

räknas därefter ut som f och h :s förändringar, såsom i (4.4).

$$\frac{\partial f_i}{\partial x_j} \approx \frac{f_i(x_j + \Delta x_j) - f_i(x_j - \Delta x_j)}{2\Delta x_j} \quad (4.4a)$$

$$\frac{\partial f_i}{\partial u_j} \approx \frac{f_i(u_j + \Delta u_j) - f_i(u_j - \Delta u_j)}{2\Delta u_j} \quad (4.4b)$$

$$\frac{\partial h_i}{\partial x_j} \approx \frac{h_i(x_j + \Delta x_j) - h_i(x_j - \Delta x_j)}{2\Delta x_j} \quad (4.4c)$$

$$\frac{\partial h_i}{\partial u_j} \approx \frac{h_i(u_j + \Delta u_j) - h_i(u_j - \Delta u_j)}{2\Delta u_j} \quad (4.4d)$$

Här står i -argumentet för det i :te tillståndet och j -argumentet för det störda tillståndet eller insignalen. Valet av störstorlek (Δx_j och Δu_j) påverkar linjäriseringsresultatet mycket. För liten störstorlek kan innebära stora avrundningsfel vid beräkning av linjäriseringen. Om däremot störstorleken väljs för stor så approximerar linjäriseringen systemet sämre i linjäriseringspunkten. En avvägning måste därför ske. Den här tekniken har Härefors (1995) använt i PLA 60 grader. Härefors fann att lämpliga störstorlekar är 3 % av värdet för WFM och A8, 3 grader för FVG och CVG samt 1 % för tillståndens värden. Dessa nivåer antas vara bra för linjäriseringar i alla arbetspunkter (Härefors 1995).

4.3 Val av linjäriseringspunkter för RM12

Eftersom linjäriseringen endast approximerar den olinjära modellen i närheten av linjäriseringspunkten krävs linjärisering i flera arbetspunkter. Med hjälp av parameterstyrning kan sedan linjäriseringarna sammanfogas för att diagnossystemet ska gälla i ett större arbetsområde. Valet av antal linjäriseringspunkter bestäms av hur väl linjäriseringen approximerar den olinjära modellen. Kraven på den linjära modellen avgör givetvis detta. I (Härefors 1995) har linjäriseringar gjorts i 11 arbetspunkter. Dessa punkter sträcker sig från PLA 17 grader till PLA 130 grader i arbetsområdet SLS (Sea Level Static). Åtta av dessa punkter har valts ut, se tabell 4.1. Anledningen till att tre utav punkterna har valts bort är att två utav dem var med tänd EBK och den tredje vid GI (Ground Idle). Dessa områden har valts att inte diagnostiseras, då de uppträder i utkanten av arbetsområdet.

PLA grad	WFM kg/s	A8 m^2	FVG grad	CVG grad	WFR kg/s
28	0.109	0.271	50.0	30.0	0
40	0.189	0.167	50.0	18.4	0
50	0.337	0.151	50.0	11.0	0
60	0.495	0.158	40.3	6.9	0
70	0.661	0.168	28.4	4.3	0
80	0.947	0.168	16.6	0.1	0
90	1.159	0.168	8.4	-1.9	0
103	1.288	0.151	8.4	-3.0	0

Tabell 4.1: Arbetspunkter med respektive insignaler till modellen.

4.4 Utvärdering av linjärisering

Med metoden från avsnitt 4.2 linjäriserades modellen i de åtta valda arbetspunkterna, se tabell 4.1. För att utvärdera den linjära modellens överensstämmelse med den olinjära finns flera metoder att tillgå, bl.a. :

- Stegsvvar
- Frekvensanalys

Stegsvvar och frekvenssegenskaper för ett olinjärt system är amplitudberoende. Amplitudberoendet gör att för stegsvvar måste stegstorlek väljas med omsorg. Stegsvaren för linjäriseringen i PLA 60 grader har undersökts. Stegstorleken på insignalerna valdes till halva "avståndet" till PLA 70 grader, se figur 4.1 (observera logskalan), för att avgöra om linjäriseringarna gjorts tillräckligt tätt. Som exempel på stegsvvar visas enbart NH respektive NL:s uppträdande, se figur 4.2 respektive figur 4.3. Stegsvaren är i dessa figurer för korta, för att några kvalitativa slutsatser ska kunna dras. Vi har dock valt att visa dem för att ge en uppfattning om hur bra linjäriseringen är.

På de båda stegsvaren syns t.ex. att modellen från WFM (bränsleflöde till brännkammare) till NL (lågtrycksturbinens varvtal) är klart bättre än för WFM till NH (högtrycksturbinens varvtal). Det framgår även klart att den linjära modellen inte helt överensstämmer med den olinjära modellen. En möjlig orsak till detta är att systemet avlägsnar sig mer och mer från linjäriseringspunkten då steget läggs på. Som tidigare sagts gäller en linjärisering enbart i närheten av linjäriseringspunkten. Ju längre från denna systemet avlägsnar sig desto sämre stämmer den linjära modellen med den olinjära, vilket visas i figur 4.2.

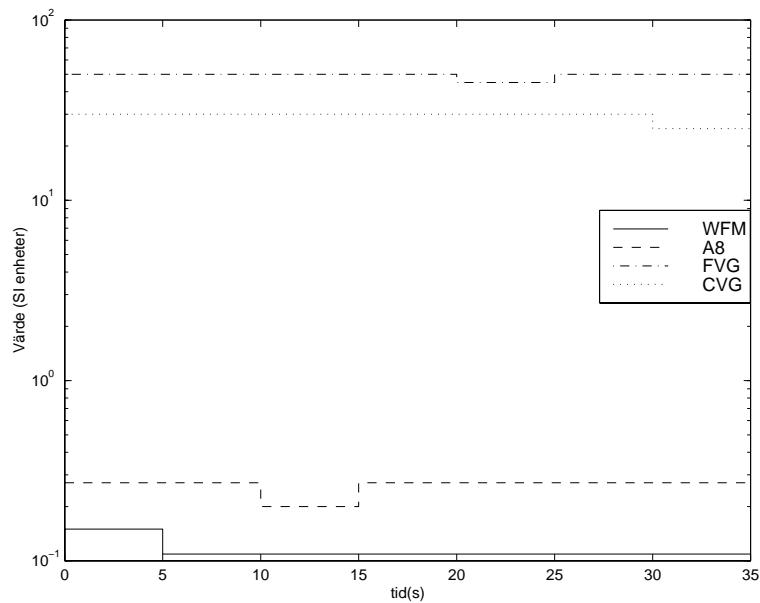
Med frekvensanalys visade Härefors (1995) att modellerna stämmer från 2-3 *rad/s* och uppåt. Då våra linjäriseringar har utförts på samma sätt som Härefors, bör detta gälla även för våra linjäriseringar. Frekvensegenskaperna tog Härefors fram med

en telegrafsignal^a. Denna signal används för att extrahera frekvensegenskaperna ur den olinjära motormodellen. Frekvensegenskaperna för den olinjära motormodellen jämfördes sedan med den linjära motormodellens bodediagram.

4.5 Sammanfattning

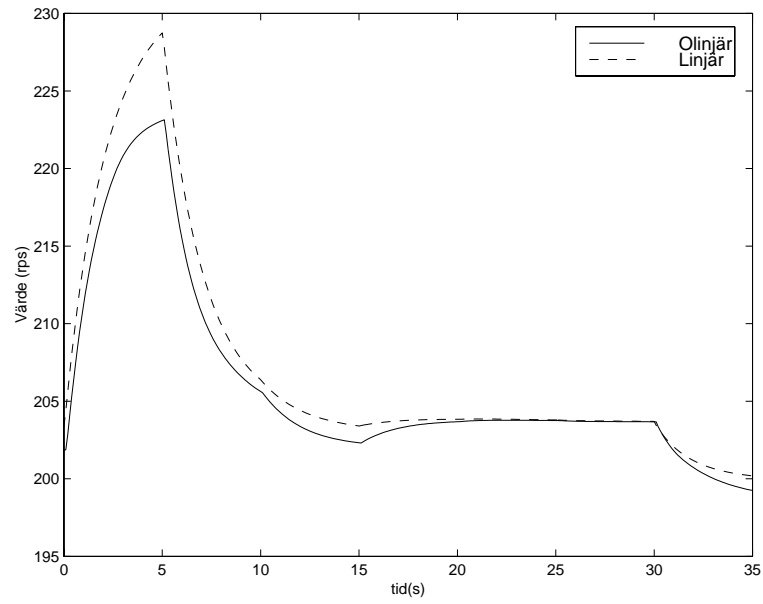
Störningsteknik har använts för att linjärisera motormodellen. Linjäriseringen är bättre för frekvenser över $2\text{-}3\text{ rad/s}$ än under (Härefors 1995). Detta är mycket olyckligt då diagnos kräver bra överensstämmelse för låga frekvenser. Linjäriseringar borde även ha gjorts tätare, då den linjäriserade motormodellen inte överensstämmer halvvägs till nästa linjäriseringspunkt.

Vi kommer i fortsättningen att förutsätta att linjäriseringen blev tillräckligt bra för diagnos. Vid simulering av fel kommer den linjäriserade motormodellen att användas. Alla vidare jämförelser av modeller kommer att göras mot den linjäriserade motormodellen, om det inte uttryckligen står något annat.

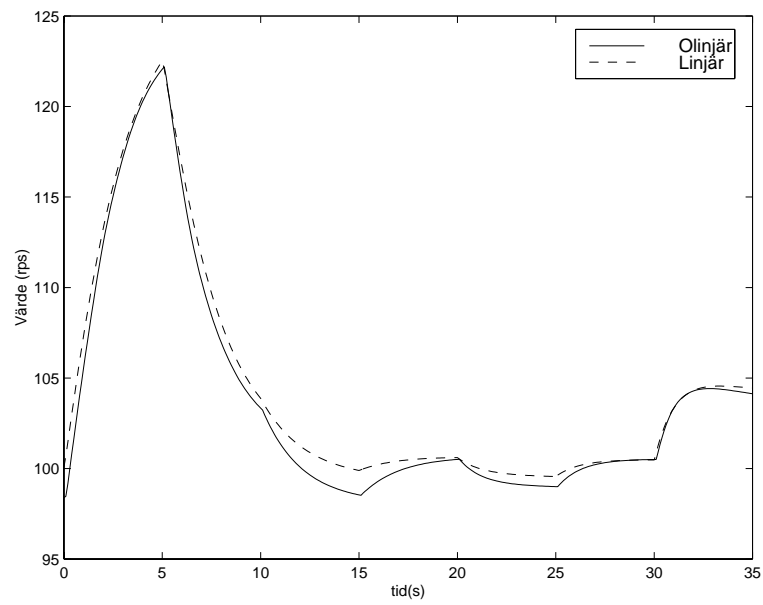


Figur 4.1: Insignaler för stegsvaren.

^aEn signal som slumpmässigt varierar mellan två nivåer och innehåller alla frekvenser.



Figur 4.2: Stegsvär för NH.



Figur 4.3: Stegsvär för NL.

Kapitel 5

Reducering

Varför reduceras en modell? Många fysikaliskt framtagna modeller har en tendens att bli komplexa. En mindre modell som approximerar systemet väl är mycket lättare att arbeta med. Det kan t.ex. vara lämpligt att förenkla modellen vid reglering eller som i detta fall diagnos. Om modellen innehåller stora skillnader i tidskonstanter, dvs är styv, reduceras den dynamik som är ointressant bort. Förutom att den ointressanta dynamiken försvinner blir modellen även bättre konditionerad.

De fel som har valts att detektera är typiskt lågfrekventa ($\leq 10\text{rad/s}$). Därför är det viktigt att den reducerade modellen approximerar systemet väl för låga frekvenser. För högre frekvenser är det däremot inte lika viktigt att redueringen blir bra. Kraven på modellen för diagnos skiljer sig alltså från de krav som ställs vid reglering, där frekvenserna kring skärfrekvensen^a är viktiga för stabilitet.

Vi kommer i kapitel 6 visa att redueringen av motormodellen för RM12 är om inte helt nödvändig så i alla fall mycket viktig då residualgeneratorerna designas. I fallet med Chow-Willsky (avsnitt 3.4.2) fås med en 32:a ordningens modell, en residualgenerator som kan innehålla ett upp till 32:a ordningens filter, vilket gör residualgeneratören beräkningskrävande. En annat motiv till reduering är de numeriska problem som uppstår vid design för stora modeller, vilket kraftigt uppmuntrar till reduering.

5.1 Allmänt om reduering och verifiering

Om den linjära modellen är representerad på tillståndsform kan tillståndsreduering utföras. Reduering med fysikalisk innebörd av tillstånden (hädanefter kallad fysikalisk reduering) och reduering med balanserad realisering är två metoder att välja ut tillstånd för reduering.

Själva redueringen av tillstånden kan göras på flera sätt. Vilket sätt som används beror på vilken tidskonstant som kan förknippas med tillståndet. Oftast är det svårt att förknippa ett tillstånd med endast en tidskonstant, men det går oftast

^aSkärfrekvens är den frekvens där beloppet av överföringsfunktionen är 1.

att relatera grupper av tillstånd till grupper av egenvärden. Vid lösning av (5.1) erhålls en diagonalmatris D med egenvärden och en egenvektormatris V . Matrisen A är den linjära modellens A -matris.

$$AV = VD \quad (5.1)$$

Genom att undersöka egenvektormatrisen V kan grupper av tillstånd relateras till grupper av egenvärden. Om tillståndet är relaterat till en grupp egenvärden med små tidskonstanter (poler långt in i vänster halvplan), dvs de ställer in sig snabbt, kan tillståndets tidsderivata sättas till noll. Tillståndet approximeras som oändligt snabbt. Om tillståndet däremot relateras till en stor tidskonstant, dvs ställer in sig långsamt, approximeras det som oföränderligt inom ändlig tid, dvs konstant. För en matematisk beskrivning se bilaga A.1. I MATLAB utförs dessa reduceringar med kommandot *modred*. Reducering av snabba tillstånd garanterar att DC-förstärkningen är samma för oreducerad och reducerad modell. Reducering av långsamma tillstånd ger ofta bra approximationer för högre frekvenser men DC-förstärkningen kan bli fel. Det är därför lämpligt att främst ta bort snabba tillstånd genom att sätta derivatan till noll, eftersom DC-förstärkningen och låga frekvenser är viktiga för diagnossystemet.

Hur verifieras att reduceringen approximerar systemet väl? Här finns det flera lämpliga metoder, såsom:

- Stegsvvar
- Bodediagram
- Singulära värden

För stegsvvar och bodediagram jämförs alla in- och utsignalkombinationer för den oreducerade och den reducerade modellen. Även de singulära värdena bör jämföras, vilket är ett snabbt sätt att överblicka reduceringen och tar hänsyn till korskopplingar. De singulära värdena till en matris A fås fram som egenvärden till A^*A (se bilaga A.2). Vid jämförelse av singulära värden krävs viss försiktighet då två system med olika egenskaper kan ha samma singulära värden, se nedanstående exempel.

Exempel 5.1

Betrakta en matris A enligt (5.2), som har singulära värden 1 oberoende av θ . Om skillnadsmatrisen $A_s = A(0) - A(\pi/2)$ bildas får den $\sqrt{2}$ som singulära värden.

$$A(\theta) = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \quad (5.2)$$

Här inses att två skilda matriser kan ha samma singulära värden. Detta beror på att egenvektorerna inte är parallella, utan som i det här fallet ortogonala. Det räcker inte med att längden på egenvektorerna är samma, de måste även ha samma riktning för att systemen ska kunna vara lika. Detta är ett nödvändigt krav.

De singulära värdena för två system kan vara lika utan att systemen är det. Jämför med det endimensionella fallet, då det inte räcker med att endast amplitudkurvorna i bodediagrammet överensstämmer. Det krävs även att faskurvorna är lika.

För att bedöma hur lika två modeller G och G_r är kan G_r uttryckas i G och ett relativt fel Δ_{rel} på följande sätt:

$$G_r = G(I + \Delta_{rel}) \quad (5.3)$$

Minimering av Δ_{rel} ger då den bästa överensstämmelsen mellan G och G_r . De metoder som finns för att minimera Δ_{rel} kräver dock att G är inverterbar. Så vitt författarna vet finns inga metoder där inte G^{-1} existerar. Är inte G inverterbar kan istället det absoluta felet Δ_{abs} studeras:

$$G = G_r + \Delta_{abs} \quad (5.4)$$

Det absoluta felet är samma sak som skillnadssystemet för G och G_r . För att bedöma storleken på skillnadssystemet kan de singulära värdena för skillnadssystemet studeras. Det absoluta felet kan dock vara stort trots att det relativa felet är litet. För att kunna dra kvalitativa slutsatser ur skillnadssystemets singulära värden, måste därför systemen G och G_r vara normaliserade. Om systemen G och G_r ej är normaliserade kan man inte vara säker på att olika in- och utsignalkombinationer viktas lika. Vid en väl utförd reducering överensstämmer både singulära värden och överföringsfunktionerna från insignal till utsignal (dvs bodediagrammen).

En viktig fråga är hur mycket en modell ska reduceras. Ett enkelt svar är: så länge som den reducerade modellen fortfarande approximerar det verkliga systemet väl. I vårt fall valdes att god modellöverensstämmelse upp t.o.m. 10 rad/s var viktigt, då de fel vi väljer att detektera fel upp t.o.m. 10 rad/s. Reduceringarna utfördes för linjäriseringen i PLA 60 grader. Reduceringar i de andra arbetspunkterna kan utföras på samma sätt.

5.2 Fysikalisk reducering

För att behålla den fysikaliska tolkningen av tillstånden, dvs. tillstånd med fysikalisk innebörd, krävs en fysikalisk reducering. En anledning till att använda fysikalisk reducering är att det ger möjlighet till att använda tillstånden som styrvariabler vid parameterstyrning. Fysikalisk reducering innebär en ingående analys av tillståndens betydelse för systemet. En sådan analys har gjorts i (Härefors 1995). Här har Härefors delat upp tillstånden i reducerbara och oreducerbara grupper utgående från deras fysikaliska betydelse och från deras tidskonstanter. Reduceringen i (Härefors 1995) gjordes i första hand för att användas vid reglering av RM12. Detta gör, som tidigare påpekats, att den utförda reduceringen inte är idealisk att använda för diagnos. Därför har en reducering med diagnos som ändamål gjorts, utgående från Härefors analys.

Från den gruppindelning av tillstånden som Härefors presenterade reducerades modellen med ett tillstånd i taget. Innan tillståndsreduceringen påbörjades minskades antalet insignaler till fyra och antalet utsignaler till åtta (se bilaga B.2). Detta då de borttagna signalerna inte påverkade modellen. Därefter skalades modellen enligt bilaga B.1 för att skillnadssystemets singulära värden skulle kunna jämföras. Tillståndsreduceringens påverkan på modellen analyserades med bodediagram, singulära värden för modellerna och skillnadssystemets singulära värden. Efter reducereing av nio tillstånd (för detaljer se B.2.2) ser de singulära värdena för de oskalade modellerna ut enligt figur 5.1. Här överlappar de singulära värdena för den oreducerade och den reducerade modellen med varandra upp t.o.m. 40 rad/s . För det oskalade skillnadssystemet erhöles singulära värden enligt figur 5.2. Här visas att modellerna överensstämmer väl för låga frekvenser pga. de låga singulära värdena för skillnadssystemet, men för 10 rad/s och uppåt överensstämmer modellerna inte lika bra (maximala singulära värdet är större än 0 dB). Anledningen till tvetydigheten grundar sig på det som tidigare nämnts om tolkningen av singulära värden. Inspektion av bodediagrammen för alla in- och utsignalkombinationer, visar att reducereingen i värsta fall är bra för frekvenser upp till 40 rad/s . I figur 5.3 respektive figur 5.4 visas enbart bodediagrammen för WFM som insignal och NH respektive NL som utsignal, då dessa är representativa för de övriga in- och utsignalkombinationerna. Bodediagrammen överensstämmer väl för WFM till NH respektive NL upp till 100 rad/s . En tvetydighet finns dock. Enligt bodediagrammen och de singulära värdena är reduktionen bra upp till 40 rad/s . Skillnadssystemets singulära värden visar däremot att modellerna endast har god överensstämmelse upp till 10 rad/s . Vi har valt den mest pessimistiska av dessa för att avgöra modellöverensstämmelsen. Detta medför att modellen ej kan reduceras ytterligare då vårt krav var god överensstämmelse mellan modellerna för frekvenser under 10 rad/s .

5.3 Balanserad reducereing

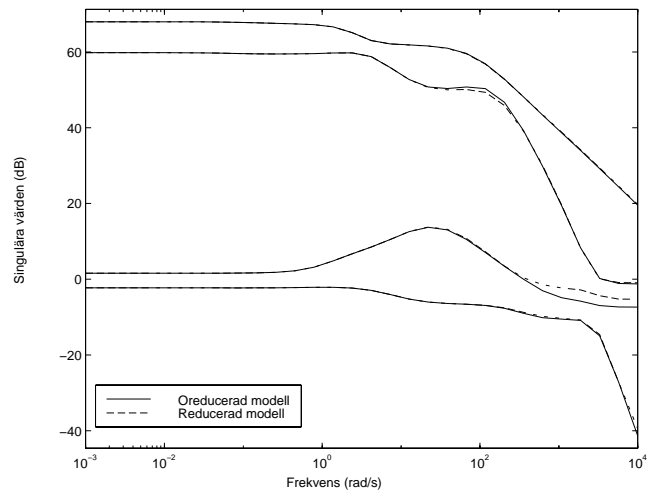
Eftersom den fysikaliska reducereingen inte kunde reducera vår modell till mindre än 23 tillstånd, har även en annan teknik använts. Tekniken avser att hitta en balanserad realisering (se bilaga A.3) av systemet och utifrån denna reducera systemet. En nackdel med denna metod är att den fysikaliska tolkningen av tillstånden försvinner i samband med att modellen balanseras. Den balanserade realiseringen åstadkoms med en transformation som gör att observerbarhets- och styrbarhetsgramianerna blir lika. För en förklaring av styrbarhets- och observerbarhetsgramian se bilaga A.3. Transformationen görs för att kunna avgöra hur styrbart och observerbart ett tillstånd är. Genom att ta bort de tillstånd som är minst styr- och observerbara reduceras modellen. I MATLAB finns balanseringen implementerad som kommandot *balreal*. Innan tillståndsreduceringen påbörjades minskades antalet insignaler till fyra och antalet utsignaler till åtta (se bilaga B.2). Detta då de borttagna signalerna inte påverkade modellen. Därefter skalades modellen enligt bilaga B.1 för att skillnadssystemets singulära värden skulle kunna

jämföras. Först efter skalningen balanserades tillstånden. Därefter togs de tillstånd med minst observer- och styrbarhet bort. Genom att göra på detta sätt erhöles en modell med 12 tillstånd (bilaga B.2.3). Detta är dock en sanning med modifikation, eftersom två tillstånd som borde ha tagits bort, enligt gramianen, behölls då de fanns påverka systemet mycket. Anledningen till att dessa tillstånd bibehölls är att systemet ej är helt normaliserat, vilket krävs för att gramianerna skall vara rättvisande vid balansering. Vid den balanserade reduceringen studerades de singulära värdena för skillnadssystemet efter varje tillstånd som togs bort. För 12 tillstånd ser de singulära värdena för de oskalade modellerna ut enligt figur 5.5 där modellerna ser ut att stämma upp till 100 rad/s . Det oskalade skillnadssystemets singulära värden indikerar att systemen börjar skilja sig åt redan vid 10 rad/s (maximala singulära värdet är 0 dB), se figur 5.6. Inspektion av bodediagrammen för alla in- och utsignalkombinationer visar att reduceringen är bra för frekvenser upp till drygt 40 rad/s . I figur 5.7 respektive figur 5.8 visas enbart bodediagrammen med WFM som insignal och NH respektive NL som utsignal.

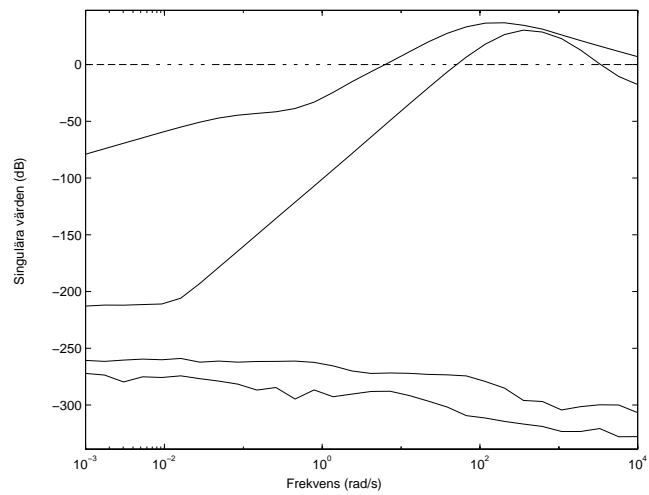
Ytterligare reducering av modellen medför för stora singulära värden i skillnadssystemet.

5.4 Sammanfattning

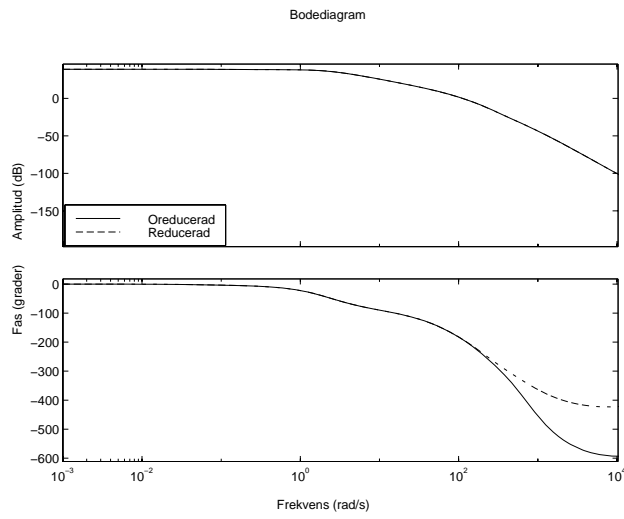
Två metoder har använts för att reducera motormodellen i PLA 60 grader. Den ena utgår från fysikalisk förståelse och kan vara lämplig då fysikalisk tolkning av tillstånden krävs. Krävs inte detta är den andra metoden, balanserad reducering, mycket bättre. Den ger klart färre tillstånd än en fysikaliskt reducerad modell för samma avvikelse från den oreducerade motormodellen. Med fysikalisk reducering erhöles en designmodell med 23 tillstånd och med balanserad reducering erhöles en designmodell med 12 tillstånd.



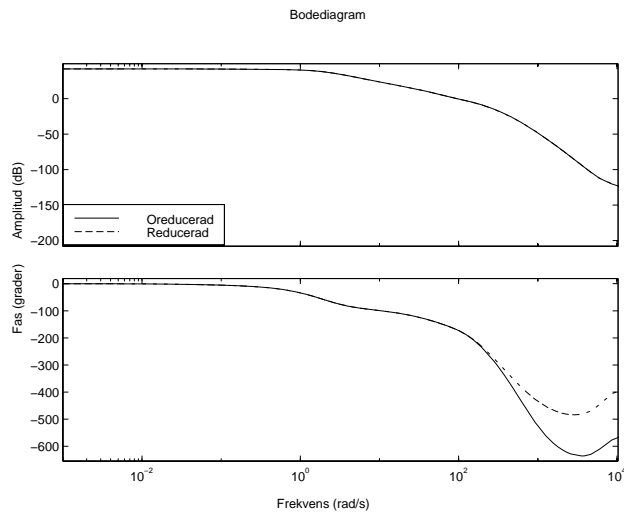
Figur 5.1: Singulära värden för 32 och 23 tillstånd (fysikalisk reducering).



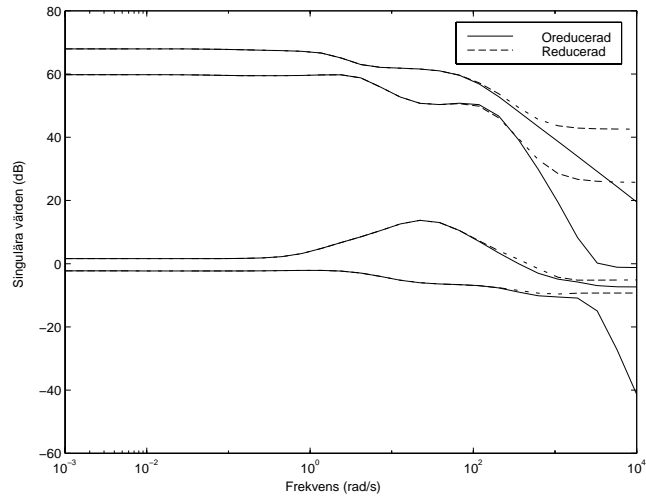
Figur 5.2: Skillnadssystemets singulära värden för 32 och 23 tillstånd (fysikalisk reducering).



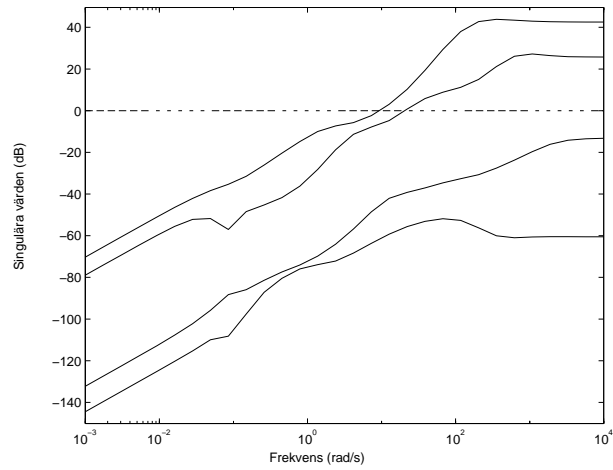
Figur 5.3: Bodediagram från WFM till NH (fysikalisk reducering).



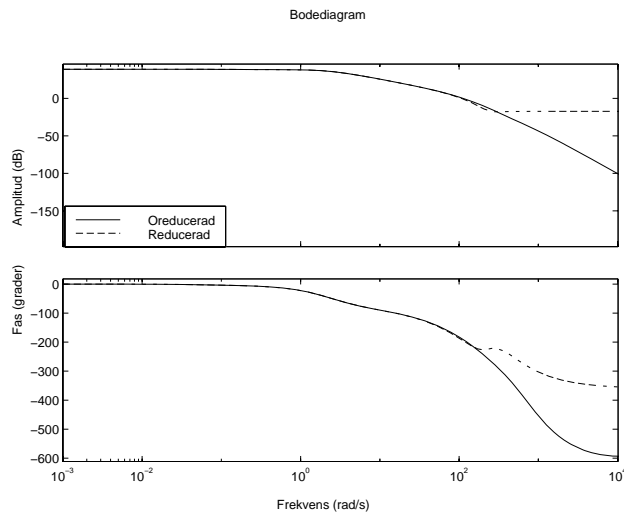
Figur 5.4: Bodediagram från WFM till NL (fysikalisk reducering).



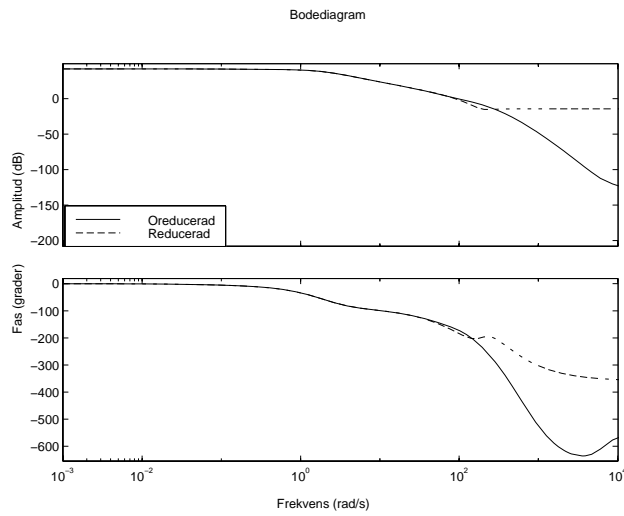
Figur 5.5: Singulära värden för 32 och 12 tillstånd (balanserad reducering).



Figur 5.6: Skillnadssystemets singulära värden för 32 och 12 tillstånd (balanserad reducering).



Figur 5.7: Bodediagram från WFM till NH (balanserad reducering).



Figur 5.8: Bodediagram från WFM till NL (balanserad reducering).

Kapitel 6

Design

I detta kapitel skall uppbyggnaden av vårt diagnosystem presenteras. Ett litet modellbaserat diagnosystem har framställts baserat på tre linjära dynamiska modeller. De tre modellerna är framtagna från den olinjära motormodellen av RM12. Den olinjära motormodellen linjäriserades med störningsteknik i åtta arbetspunkter. En av arbetspunkterna valdes ut för en principdesign av diagnosystem, denna arbetspunkt var PLA 60 grader (SLS) (kapitel 4). Den linjära motormodellen som då erhöles reducerades till två olika modeller (kapitel 5). Även en tredje modell användes vid design av residualgenerator. Den tredje modellen erhöles från VAC (se bilaga B.2.1). Dessa tre modeller benämner vi designmodeller. Därefter valdes en residualstruktur och utgående från denna designades residualgeneratorer med två metoder, Chow-Willskys metod och minimala polynombasmetoden (avsnitt 3.4). Diagnosystemet utvärderades sedan genom att fel i ställdon och sensorer simulerades. Simuleringen gjordes mot den linjära oreducerade motormodellen, då linjäriseringen ej blev så bra att simulering av fel med den olinjära motormodellen var genomförbar.

6.1 Modeller för diagnos

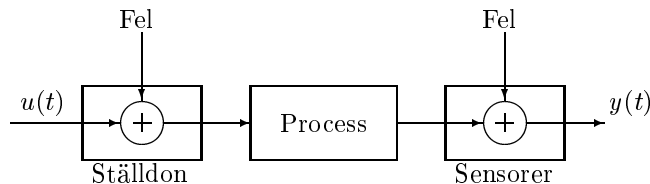
Tre diagnosystem har framställts i arbetspunkten PLA 60 grader (SLS), då denna ligger mitt i arbetsområdet. För att framställa dessa diagnosystem har tre linjäriserade och reducerade modeller använts. Den första designmodellen, här kallad modell 1, har Lind (1998) bistått med, se bilaga B.2.1. Denna modell har endast sju tillstånd och användes ursprungligen till att estimeras temperaturen i högtrycksturbinen. Den andra designmodellen (modell 2) är en linjäriserad och fysikaliskt reducerad variant av RM12:s olinjära motormodell. Skillnaden mellan modellerna är att vid reduceringen till modell 2 har större tonvikt lagts på överensstämmelse för låga frekvenser. Denna tonvikt medförde att modell 2 endast kunde reduceras till 23 tillstånd (se avsnitt 5.2). Den tredje designmodellen (modell 3) har däremot reducerats genom balansering (se avsnitt 5.3) och detta resulterade i en modell

med 12 tillstånd. Balanserad reducering medför dock att den fysikaliska tolkningen av tillstånden för modell 3 försvinner.

6.1.1 Modellering av fel

Som tidigare visats i avsnitt 3.2 kan fel modelleras på olika sätt. Diagnossystemet har begränsats till att inkludera sensor- och ställdonsfel pga den kunskap som krävs vid modellering av processfel och den tid det tar. Ett exempel är att modellera bladfel som periodiska störningar av tryckförändringar vid bladen och sedan frekvensanalysera trycket (Aretakis, Mathioudakis & Dedoussis 1997). Mönsterigenkänning används för att skilja på de olika typer av bladfel som modellerats. Ett sådant system skulle kunna bli ett komplement till ett sådant modellbaserat diagnos som vi konstruerar.

Ställdons- och sensorfelen har modellerats som additiva fel på insignal respektive utsignal enligt figur 6.1, eftersom det är enklast och mest generellt. Då detta är en principstudie begränsas felen till fyra, två ställdonsfel (WFM och CVG) och två sensorfel (NL och NH). Dynamiken i ställdonen och sensorerna har ej inkluderats.



Figur 6.1: Modellerade fel.

6.1.2 Simulering av fel

För att testa diagnossystemet simuleras felen. Ställdonsfel och sensorfel kan, pga sin additiva modellering, enkelt simuleras genom att signaler adderas till in- respektive utsignalerna. Simulering av fel sker mot den linjära motormodellen.

6.2 Test av designmetoder

De två designmetoderna Chow-Willsky och minimala polynombaser som tidigare presenterats i avsnitt 3.4.2 respektive 3.4.3 har testats. Valet av Chow-Willsky är grundat på att det är en av de vanligast förekommande metoderna. När det visade sig att Chow-Willskys metod inte klarade av att designa en residualgenerator användes minimala polynombasmetoden.

6.2.1 Val av residualstruktur

Residualstrukturen har, för de fyra olika felen, valts enligt tabell 6.1. Först bestämdes att en struktur enligt C (se tabell 3.1 sid 12) skulle användas. Denna struktur är som tidigare sagts robustare mot modellosäkerheter än vad struktur D är. Utgående från C infördes don't care i strukturen för att delvis kunna forma överföringsfunktionen G_{fr} , men även för att underlätta tröskelsättning vid residualevaluering. Med don't care blir residualstrukturen robustare.

Residual	WFM	CVG	NH	NL
r_{11}	0	1	X	X
r_{12}	0	X	1	X
r_{13}	0	X	X	1
r_{21}	1	0	X	X
r_{22}	X	0	1	X
r_{23}	X	0	X	1
r_{31}	1	X	0	X
r_{32}	X	1	0	X
r_{33}	X	X	0	1
r_{41}	1	X	X	0
r_{42}	X	1	X	0
r_{43}	X	X	1	0

Tabell 6.1: Residualstruktur för diagnossystemet.

6.2.2 Chow-Willsky

modell 1 användes för design av residualgenerator med Chow-Willskys metod. Tabell 6.1 visar den residualstruktur som användes. Problem uppstod dock vid genereringen av basen w^T för vänster nollrum till $[R \ H]$, se (3.13) sid 17. Orsaken till problemet är att systemet är mycket styvt. Vid beräkningen av Q_c , H och P i (3.11), där bl.a. $A^{\rho-1}$ finns, uppstår stora potensdifferenser mellan matriselementen, vilket är anledningen till de numeriska svårigheterna. Ett sätt att undkomma svårigheten är att välja ρ mindre än antalet tillstånd ($\rho < n$), men då riskeras att viktig information förloras och att den bästa residualgeneratoren inte hittas.

Simulering av fel utfördes trots detta och gav exponentiellt växande residualer, även i det felfria fallet. Vid beräkning av nollrum användes funktionen *null* i MATLAB 5.2. För att bedöma om nollrummet verkligen var ett nollrum, multiplicerades basen w^T med $[R \ H]$ och produkten blev större än den borde bli. MATLAB kan alltså inte beräkna det korrekta nollrummet till $[R \ H]$ för vårt problem, vilket troligen beror på finit ordlängd (64 bitar). Den relativa flyttalsprecisionen är $2.2 \cdot 10^{-16}$ för den MATLAB-version som använts. Nollrumsproblemet kan troligen avhjälpas med ett annat program, t.ex. MATHEMATICA där godtycklig noggrannhet kan väljas. Detta har dock inte undersökts.

En residualgenerator kan alltså inte designas i MATLAB med Chow-Willskys metod om systemet är för styvt och stort. Design med Chow-Willskys metod utgående från modell 2 och 3 är därför omöjligt i vårt fall.

6.2.3 Minimala polynombasmetoden

Eftersom design inte lyckats med Chow-Willskys metod undersöktes minimala polynombasmetoden. Denna designmetod har som tidigare nämnts i avsnitt 3.4.3 egenskapen minimalitet. Polynombasberäkningarna utfördes mha. en toolbox i MATLAB 5.2 med namnet *The Polynomial Toolbox for Matlab* från (Henrion, Krafer, Kwakernaak, Pejchov, Sebek & Strijbos 1997). I denna toolbox finns flera algoritmer för beräkning av polynombaser. Algoritmen som använts är baserad på *polynomial echelon form*. Det är den numeriskt stabilare av dessa och snabbast. För en mer detaljerad beskrivning se (Frisk & Nyberg 1998a).

Basen, $F(s)$, som beräknats med ovanstående algoritm är både minimal, fullständig och har minimal parametrering. Tre viktiga egenskaper, som tidigare nämnts i avsnitt 3.4.3. För att välja ut den bästa residualgeneratorm ur $F(s)$, har vi valt att maximera förhållandet (6.1) vid frekvensen noll.

$$\max_Q \frac{\|G_{fr}(0)\|_2}{\|G_{dr}(0)\|_2} \quad (6.1)$$

$$Q(s) = D_{LP}(s)\phi(s)F(s) \quad (6.2)$$

Detta förhållande valdes att maximeras då det är svårt att optimera över flera frekvenser, samtidigt som det är viktigast att upptäcka konstanta och lågfrekventa fel. Koefficienterna i en linjärkombination $\phi(s)$ kan som tidigare sagts vara polynom. Detta gör koefficienterna svårare att handskas med. I brist på metoder att välja $\phi(s)$ har valet av $Q(s)$ begränsats till att utgöra en av basvektorerna i $F(s)$ (dvs. $\phi(s)$ innehåller endast en etta och resten nollor). Även gradtalet av basvektorerna i $F(s)$ har minimerats, då $\phi(s)$ har valts. För att realisera residualgeneratorm $Q(s)$ krävs ett filter med tillräckligt högt relativt gradtal. Detta filter benämns $D_{LP}(s)$, se (6.2).

Nu har residualgeneratorm (6.2) realiserats och den bästa basvektorm i $F(s)$ kan väljas ut genom (6.1). $D_{LP}(s)$ kan sedan väljas på en rad olika sätt, t.ex. som ett lågpasfilter. Ett sätt är att forma $G_{fr}(s)$ i (3.6) till 1 för alla s . Detta kan göras för de fel som har 1:a i residualstrukturen, se tabell 6.1. Överföringsfunktionen G_{fr} kan ses som skalär eftersom det avkopplade felet räknas som en störning och de fel residualen har X för ignoreras. Rent teoretiskt kan detta åstadkommas genom att välja $D_{LP}(s) = (\phi(s)F_y(s)G_{fy}(s))^{-1}$ (jämför (3.19)), där $F(s) = [F_y(s) F_u(s)]$. Formning av G_{fr} har ej lyckats då $D_{LP}(s)$ får väldigt hög ordning.

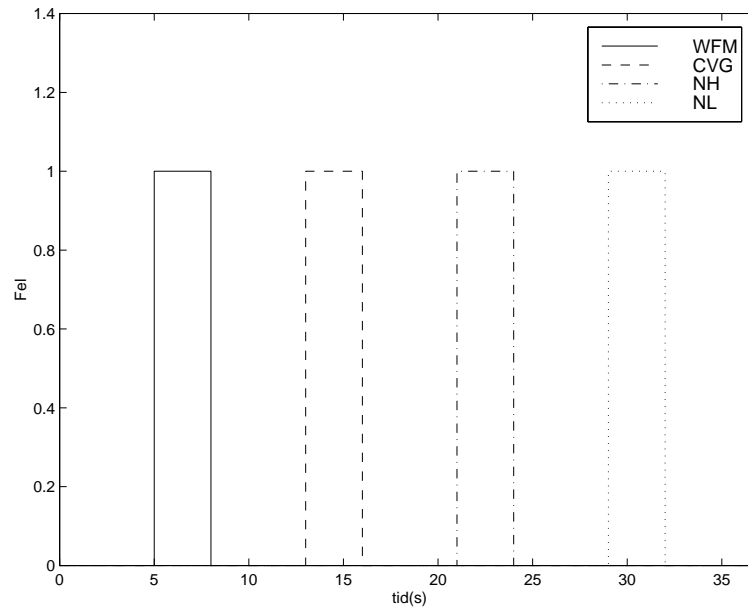
En framgångsrikare metod var att släppa på kraven och forma G_{fr} till 1 enbart vid DC-nivån, vilket realiserades genom att välja $D_{LP}(s)$ till ett lågpasfilter med lämplig DC-förstärkning. $D_{LP}(s)$:s ordningstal kunde nu väljas klart lägre för att realisera residualgeneratorm.

Design mot tre olika modeller

Som tidigare nämnts är designmodellen av systemet en mycket viktig del i modellbaserad diagnos. Därför har tre olika modeller legat till grund för design. Design mot de olika modellerna och den efterföljande simuleringen av fel kommer att beskrivas nedan. Då linjäriseringen av den olinjära motormodellen inte blev bra kommer den linjäriserade oreducerade motormodellen att användas som modell av motorn vid simuleringarna.

Modell 1

Med minimala polynombasmetoden togs en residualgenerator fram för modell 1. Ett lågpasfilter med relativt gradtal 3 och skärfrekvensen 10 rad/s användes för att realisera residualgeneratorn detta då $\phi(s)F(s)$ hade gradtal ≤ 3 . Därefter simulerades felen i sensorerna och ställdonen enligt figur 6.2, som schematiskt visar när felen inträffar. Felens tidsmässiga utsträckning har valts till 3 sekunder, vilket är en kort tid. Om de simulerade felen var aktiva en längre tid skulle säkerligen mindre fel kunna detekteras. Tabell 6.2 visar de felstorlekar som användes. I bilaga C.1



Figur 6.2: Schematisk bild av de simulerade felen.

visas de tolv residualerna. Residualstrukturen såg ut som i tabell 6.1, där residual r_{ij} har avkopplat fel i . Ett X betyder som tidigare nämnts don't care, dvs det spelar ingen roll om residualen blir stor eller liten. Vid evaluering valdes statistiska trösklar J_{ij} sådana att tröskeln var strax större än r_{ij} 's värde då fel i är aktivt.

Fel	Felstorlek	Relativt fel (%)
WFM	0.1 kg/s	20
CVG	1.5 grader	21
NH	30 rps	12
NL	10 rps	6

Tabell 6.2: De simulerade felens värde för modell 1.

Med denna enkla tröskelsättningsregel erhöles felbeslut från residualevaluatorn enligt tabell 6.3. "Är" är det felbeslut som residualevaluatorn för modell 1 gav. "Bör" är det felbeslut som borde generats enligt residualstrukturen. Ur tabellen framgår att diagnossystemet lyckas med att isolera fel i CVG, NH och NL. Däremot lyckas diagnossystemet inte detektera fel i WFM. Valet att inte gå vidare med modell 1 gjordes, då den inte klarar av att isolera alla fel.

Residual	WFM		CVG		NH		NL	
	är	bör	är	bör	är	bör	är	bör
r_{11}	0	0	1	1	1	X	1	X
r_{12}	0	0	1	X	1	1	1	X
r_{13}	0	0	1	X	1	X	1	1
r_{21}	0	1	0	0	1	X	1	X
r_{22}	0	X	0	0	1	1	1	X
r_{23}	0	X	0	0	1	X	1	1
r_{31}	0	1	1	X	0	0	1	X
r_{32}	0	X	1	1	0	0	1	X
r_{33}	0	X	1	X	0	0	1	1
r_{41}	0	1	1	X	1	X	0	0
r_{42}	0	X	1	1	1	X	0	0
r_{43}	0	X	1	X	1	1	0	0

Tabell 6.3: Utvärdering av modell 1.

Modell 2

Med minimala polynombasmetoden togs en residualgenerator fram för modell 2. Ett LP-filter med relativt gradtal 3 och skärfrekvensen 10 rad/s användes för att realisera residualgeneratorn detta då $\phi(s)F(s)$ hade gradtal ≤ 3 . Samma felstorlekar som för modell 1 användes vid felsimuleringen, se tabell 6.2. I bilaga C.2 visas de tolv residualerna. Här visas att residualerna inte lyckas med avkopplingen för fel i WFM, CVG eller NH. Däremot lyckades residualerna avkoppla NL bra. Problem med avkoppling uppstår även för mindre och för större felstorlekar. Vad

detta beror på är inte klart. Bodediagrammen i figur 5.3 och i figur 5.4 visar att överföringsfunktionerna från WFM till NH respektive NL överensstämmer ungefär lika bra. Designmodellen är inte bättre för NL än den är för NH. En trolig orsak till den dåliga avkopplingen är att korskopplingarna mellan WFM, CVG och NH är stora, då de alla på något sätt är relaterade till högtrycksturbinen. Samma residualstruktur och metod att sätta statistiska trösklar som i modell 1 användes. "Är" är det felbeslut som residualevaluatorn för modell 2 gav. "Bör" är det felbeslut som borde generats enligt residualstrukturen. Tabell 6.4 visar att residualgenera-

Residual	WFM	WFM	CVG	CVG	NH	NH	NL	NL
	är	bör	är	bör	är	bör	är	bör
r_{11}	0	0	0	1	0	X	0	X
r_{12}	0	0	0	X	0	1	0	X
r_{13}	0	0	0	X	0	X	1	1
r_{21}	1	1	0	0	1	X	1	X
r_{22}	1	X	0	0	1	1	1	X
r_{23}	1	X	0	0	1	X	1	1
r_{31}	1	1	0	X	0	0	1	X
r_{32}	1	X	0	1	0	0	1	X
r_{33}	1	X	0	X	0	0	1	1
r_{41}	0	1	1	X	1	X	0	0
r_{42}	0	X	1	1	1	X	0	0
r_{43}	0	X	1	X	1	1	0	0

Tabell 6.4: Utvärdering av modell 2.

torn inte fungerar enligt vald residualstruktur. Den klarar däremot av att isolera fel i NL. Med lämpligt vald logik i residualevaluatorn kan fel detekteras då WFM, CVG eller NH fallerar. Modell 2 klarar alltså av att detektera fel i WFM, vilket inte modell 1 gjorde. Däremot lyckades modell 1 isolera felen i CVG och NH, vilka modell 2 endast detekterar. Skillnaden beror på att residualgeneratorm designad med modell 2 inte lyckas avkoppla fel lika bra. Beslutet att gå vidare till modell 3 togs.

Modell 3

Med minimala polynombasmetoden togs en residualgenerator fram för modell 3. Ett lågpasfilter med relativt gradtal 3 och skärfrekvensen 10 rad/s användes för att realisera residualgeneratorm detta då $\phi(s)F(s)$ hade gradtal ≤ 3 . Samma felstorlekar som för modell 1 användes vid felsimuleringen, se tabell 6.2. Förutom CVG:s felstorlek som fick ökas till 2 grader, då denna var svår att isolera.

I bilaga C.3.1 visas de tolv residualerna. Här syns att avkopplingen lyckas väl för alla fel. Samma tröskelsättningsregel som för modell 2 användes och gav föl-

jande felbeslut från residualevaluatorn, se tabell 6.5. Residualgeneratorn fungerar

Residual	WFM	WFM	CVG	CVG	NH	NH	NL	NL
	är	bör	är	bör	är	bör	är	bör
r_{11}	0	0	1	1	1	X	1	X
r_{12}	0	0	1	X	1	1	1	X
r_{13}	0	0	1	X	1	X	1	1
r_{21}	1	1	0	0	1	X	1	X
r_{22}	1	X	0	0	1	1	1	X
r_{23}	1	X	0	0	1	X	1	1
r_{31}	1	1	1	X	0	0	1	X
r_{32}	1	X	1	1	0	0	1	X
r_{33}	1	X	1	X	0	0	1	1
r_{41}	1	1	1	X	1	X	0	0
r_{42}	1	X	1	1	1	X	0	0
r_{43}	1	X	1	X	1	1	0	0

Tabell 6.5: Utvärdering av modell 3.

enligt residualstrukturen. Den klarar av att isolera alla felen. Även simuleringar av mindre felstorlekar gjordes. Tabell 6.6 visar de minsta felstorlekar som testades och den relativa felstorleken. I kolumnen "VAC idag" visas hur stora relativa fel som feldetekteringssystemet kan detektera idag. Dessa fel kan detekteras i hela arbetsområdet, vilket inte vårt diagnosystem klarar av. Däremot klarar vårt diagnosystem av att isolera mycket kortare fel än VAC:s nuvarande system, som grundar sig på medelvärdesbildning under en viss tidsperiod. Simuleringarna av

Fel	Felstorlek	Relativt fel (%)	VAC idag (%)
WFM	0.05 <i>kg/s</i>	10	i.u.
CVG	2 <i>grader</i>	29	i.u.
NH	5 <i>rps</i>	2	8
NL	1.5 <i>rps</i>	< 1	4

Tabell 6.6: De simulerade felens värde för modell 3.

residualerna visas i bilaga C.3.2. Felbeslutet från residualevaluatorn blev som i det tidigare fallet för modell 3, se tabell 6.5. Felen lyckades alltså även i detta fallet isoleras.

6.3 Sammanfattning

Design av residualgeneratorer med två olika metoder har utförts. Chow-Willskys metod visade sig ha stora numeriska problem redan för en modell med sju tillstånd. Dessa problem uppträdde inte i lika stor utsträckning vid användandet av minimala polynombasmetoden. Dock visade det sig att enbart residualgeneratoren baserad på designmodell 3 med 12 tillstånd gav önskvärd isolering av felen. Att residualgeneratoren baserad på den fysikaliska designmodellen med 23 tillstånd inte blev bättre än den baserad på sju tillstånd, beror på att den helt enkelt är för stor och/eller dåligt konditionerad. Problem uppstår då designalgoritmen ska ta fram nollrummet till en så stor modell, vilket medför att en modell förutom att ha god överensstämmelse med systemet även bör ha få tillstånd och/eller vara välkonditionerad.

Kapitel 7

Slutsatser och rekommendationer

7.1 Slutsatser

Syftet med examensarbetet var att undersöka om modellbaserad diagnos kan användas för feldetektering av Volvos RM12-motor. En principstudie av linjär modellbaserad diagnos applicerad på RM12 har därför utförts. Studien visar att ställdonsfel och sensorfel kan isoleras. Vid simuleringar av sensorfel isoleras fel av mindre storlek än vad dagens feldetekteringssystem detekterar i motorn. En felstorlek på 1% av sensorvärdet isoleras i NL (lågtrycksturbinens varvtalsensor) och 2 % i NH (högtrycksturbinens varvtalsensor), vilket är en fjärdedel av vad som kan detekteras i dag. För ställdonen kunde en felstorlek på 2 grader i CVG (kompressorns variabla geometri) och 0.05 kg/s i WFM (bränsleflöde i brännkammare) isoleras. Dessa resultat grundar sig på simuleringar mot den linjäriserade motormodellen.

Utgående från den olinjära motormodellen av RM12 har en linjärisering med störningsteknik i PLA 60 grader tagits fram. Den linjära motormodellen reducerades därefter med två olika metoder, reducering med fysikalisk innebörd av tillstånden samt reducering med balanserade tillstånd. De reducerade modellerna kallas designmodeller. Vid design av residualgenerator visade det sig att den balanserade reduceringen gav residualgeneratoren goda egenskaper. Det var residualgeneratoren baserad på denna designmodell som gav bäst resultat vid simulering av fel mot den linjära motormodellen.

De för- och nackdelar med modellbaserad diagnos som vi stött på kan sammanfattas som:

Fördelar:

- Små fel kan isoleras
- Ingen extra hårdvara behövs
- Isolering av fel kan ske både i drift och efteråt
- Isolering av fel sker passivt

Nackdelar:

- Kräver bra designmodell
- Modellering av fel kan vara svårt

7.2 Rekommendationer till fortsatt arbete

Nedanstående punkter har vi numrerat i den ordning de kan vara lämpliga att utreda. De tre första punkterna är klart viktigast, då dessa måste genomföras för att ett väl fungerande diagnosystem ska kunna skapas. Först då kan den sista punkten genomföras.

1. Linjärisering av den olinjära motormodellen är en viktig del och kan förbättras genom att de lågfrekventa egenskaperna bibehålls. En bättre överensstämmelse mellan designmodell och den olinjära motormodellen i det för diagnos viktiga lågfrekvensområdet fås då. En sådan förbättring skulle innebära att den olinjära motormodellen kan användas vid felsimulering. Simulering av fel skulle då bli mer realistisk. Det är dock oklart hur en sådan linjärisering skall gå till.
2. Parameterstyrning kan användas för att sammanföra diagnosystem i olika arbetspunkter. En sådan sammanföring skulle innebära att diagnos skulle kunna ställas i ett större arbetsområde. En enkel metod för parameterstyrning är att abrupt byta diagnosystem mellan arbetspunkter. Om en sådan parameterstyrning inte blir tillräckligt bra bör parameterstyrning med någon slags interpolering mellan diagnosystemen i olika arbetspunkter göras.
3. För att få ett väl fungerande diagnosystem bör även processfel inkluderas. De allvarligaste felen som kan inträffa i en flygmotor är just processfel. Viktiga processfel att detektera är bladfel samt nedsmutsning och förslitning av motorn. Processfel modelleras med fördel som parameterförändringar då denna metod beskriver processfel bra.

4. Vid design av residualgenerator kan urvalskriteriet förbättras, för att en bättre residualgeneratör ska kunna designas. En sådan residualgenerator skulle kunna hitta fel av mindre storlek. Därför är linjärkombinationer av basvektorer med polynom som koefficienter intressant att utreda, då en större mängd av möjliga residualgeneratorer undersöks.
5. Residualvärderingen kan som tidigare nämnts baseras på mer avancerade metoder för att säkrare kunna avgöra när ett fel inträffar. En metod för detta är adaptiva trösklar. Adaptiva trösklar kan kompensera bort åldrande i motorn, vilket innebär att diagnossystemet kan användas under motorns hela livslängd. För att detektera driftfel kan ett medelvärdesfönster bildas som upptäcker om medelnivån för en residual har förändrats mycket den senaste tidsperioden. Detta skulle även kunna göras med ett lågpasfilter med nollställe nära origo i residualgeneratören.
6. För att detektera åldrande i motorn krävs en adaptiv designmodell. Då motorn åldras kommer designmodellen att stämma allt sämre med den riktiga processen, vilket innebär att residualerna kan ge utslag utan att ett fel har inträffat. Falsklarm kan undvikas om designmodellen och därmed residualgeneratören görs adaptiv. Designmodellen skulle då efterlikna den riktiga processen även när den åldras.

Litteraturförteckning

- Aretakis, N., Mathioudakis, K. & Dedoussis, V. (1997), Derivation of signatures for faults in gas turbine compressor blading, *in* 'Symposium on Fault Detection, Supervision and Safety for Technical Processes', Vol. 1, IFAC, pp. 77–82.
- Basseville, M. & Nikiforov, I. (1993), *Detection of abrupt changes*, PTR Prentice-Hall Inc. ISBN 0-13-126780-9.
- Chow, E. & Willsky, A. (1984), 'Analytical redundancy and the design of robust failure detection systems', *IEEE Trans. on Automatic Control* **29**(7), 603–614.
- Ding, X. & Frank, P. (1991), Frequency domain approach and threshold selector for robust model-based fault detection and isolation, *in* 'Fault Detection, Supervision and Safety for Technical Processes', IFAC, pp. 271–276.
- Frank, P. (1994), Application of fuzzy logic to process supervision and fault diagnosis, *in* 'Fault Detection, Supervision and Safety for Technical Processes', IFAC, Espoo, Finland, pp. 507–514.
- Frisk, E. & Nyberg, M. (1997), 'Fault detection and isolation in dynamic systems'. Kurskompendium LiTH.
- Frisk, E. & Nyberg, M. (1998a), A description of the minimal polynomial basis approach to linear residual generation, Technical report, ISY, Linköping, Sweden.
- Frisk, E. & Nyberg, M. (1998b), A minimal polynomial basis solution to residual generation for linear systems, Submitted to IFAC 99.
- Frisk, E., Nyberg, M. & Nielsen, L. (1997), FDI with adaptive residual generation applied to a DC-servo, *in* 'Symposium on Fault Detection, Supervision and Safety for Technical Processes', Vol. 1, IFAC, pp. 438–443.
- Glad, T. & Ljung, L. (1989), *Reglerteknik, Grundläggande teori*, 2:a utgåvan, Studentlitteratur. ISBN 91-44-17892-1.
- Glad, T. & Ljung, L. (1997), *Reglerteori, flervariabla och olinjära metoder*, Studentlitteratur. ISBN 91-44-00472-9.

- Hackman, P. (1991), 'Kossa h – komplexa vektorrum'. Kurskompendium LiTH.
- Henrion, D., Krafer, F., Kwakernaak, H., Pejchov, S., Sebek, M. & Strijbos, R. (1997), 'The polynomial toolbox for matlab', <http://www.math.utwente.nl/polbox/>.
- Höfling, T. & Isermann, R. (1996), 'Fault detection based on adaptive parity equations and single-parameter tracking', *Control Eng. Practice* **10**(4), 1361–1369.
- Härefors, M. (1995), 'Multivariable control design for a jet engine', Licentiatavhandling, Chalmers Tekniska Högskola. ISBN 91-7197-150-5.
- Isermann, R. (1993), 'Fault diagnosis of machines via parameter estimation and knowledge processing - tutorial paper', *Automatica* **29**(4), 815–835.
- Johansson, M. (1997), 'Modellering i MATRIXx, effektivisering av RM12 jetmotor-modell', Examensarbete, Tekniska Högskolan i Linköping. LiTH-ISY-EX-1813.
- Kailath, T. (1980), *Linear Systems*, Prentice-Hall Inc. ISBN 0-13-536961-4.
- Lind, K. (1998), Dynamic real-time estimation of high pressure turbine inlet temperature in the RM12 jet engine, Technical report, Volvo Aero Corporation.
- Nyberg, M. (1997), 'Model based diagnosis with application to automotive engines', Licentiatavhandling, Linköpings Universitet. ISBN 91-7219-006-X.
- Nyberg, M. (1998), Increasing fault diagnosis performance by separating the incidence structure from the decision structure, Submitted to IFAC 99.
- Nyberg, M. & Nielsen, L. (1997a), 'Model based diagnosis for the air intake system of the SI-engine', *SAE Paper*.
- Nyberg, M. & Nielsen, L. (1997b), Parity functions as universal residual generators and tool for fault detectability analysis, in 'Proc. IEEE Conf. on Decision and Control', San Diego, California.
- VAC (1991), *Motor RM12, systemkurs*, Volvo Aero Corp.
- Zhou, K., Doyle, J. & Glover, K. (1996), *Robust and optimal control*, Prentice Hall. ISBN 0-13-456567-3.

Bilaga A

Kompletterande teori

A.1 Tillståndsreducering av modell

Tillstånden i ett linjärt system kan delas upp i grupper, se (A.1a). Vid reducering av tillstånden kan de tillstånd som ska behållas placeras i en grupp x_1 . De tillstånd som ska reduceras placeras sedan i x_2 respektive x_3 beroende på om tidskonstanterna är *snabba* respektive *långsamma*. Hela systemet kan då skrivas som

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad (\text{A.1a})$$

$$\dot{x} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} x + \begin{pmatrix} B_1 \\ B_2 \\ B_3 \end{pmatrix} u \quad (\text{A.1b})$$

$$y = (C_1 \ C_2 \ C_3)x + Du \quad (\text{A.1c})$$

De långsamma tillstånden i x_3 sätts till noll i (A.1). De snabba tillstånden i x_2 antas vara så snabba (jämfört med det intressanta frekvensområdet) att de betraktas som oändligt snabba, dvs $\dot{x}_2 = 0$. Det reducerade systemet blir då

$$\dot{x}_r = (A_{11} - A_{12}A_{22}^{-1}A_{21})x_r + (B_1 - A_{12}A_{22}^{-1}B_2)u \quad (\text{A.2a})$$

$$y = (C_1 - C_2A_{22}^{-1}A_{21})x_r + (D - C_2A_{22}^{-1}B_2)u \quad (\text{A.2b})$$

under förutsättning att A_{22}^{-1} existerar.

A.2 Singulära värden

Ett envariabelt linjärt systems egenskaper kan tolkas med ett *bodediagram*. Motsvarande tolkning kan även göras för ett flervariabelt linjärt system. Detta görs genom att tolka bodediagrammen för var och en av de olika insignal/utsignal-kombinationerna. För att analysera alla egenskaper hos flervariabla linjära system räcker det inte med bodediagram. Korskopplingar mellan de olika delsystemen är något som bodediagrammen missar. Istället kan något kallat singulära värden användas och de definieras enligt följande:

Definition A.1 (Singulära värden (Glad & Ljung 1997)) *Givet en matris G . Dess singulära värden σ_i definieras som $\sigma_i = \sqrt{\lambda_i}$, där λ_i är egenvärdena till G^*G^a . Det största singulära värdet till G betecknas $\bar{\sigma}(G)$ och det minsta med $\underline{\sigma}(G)$.*

För ett stabilt system gäller att

$$Y(iw) = G(iw)U(iw) \quad (\text{A.3})$$

där $U(iw)$ och $Y(iw)$ är in- respektive utsignalernas fouriertransformer. Då gäller att

$$\underline{\sigma}(G(iw)) \leq \frac{|Y(iw)|}{|U(iw)|} \leq \bar{\sigma}(G(iw)) \quad (\text{A.4})$$

Systemets förstärkning för frekvens w ligger mellan minsta och största singulära värdet. Den verkliga förstärkningen beror på vilken riktning insignalen $U(iw)$ har. Om $U(iw)$'s riktning är parallell med den största egenvektorn till $G^*(iw)G(iw)$ så är förstärkningen $\bar{\sigma}(G(iw))$. Utförligare beskrivning av singulära värden finns i (Glad & Ljung 1997).

A.3 Reducering med balanserad tillståndsrealisering

Vid reducering av en linjär modell är det förstas viktigt att behålla de moder som påverkar modellen mest. Ett sätt att åstadkomma detta är balanserad tillståndsrealisering ((Zhou, Doyle & Glover 1996) kapitel 7).

Ur en ickeminimal tillståndsrepresentation, kan de tillstånd som har noll som diagonalelement i styrbarhets- (A.5a) och observerbarhetsgramianerna (A.5b) tas bort. Styrbarhetsgramianen beskriver hur mycket de olika komponenterna av x påverkas av insignalen och observerbarhetsgramianen beskriver hur mycket de olika komponenterna av x påverkas av utsignalen.

$$S = \int_0^{\infty} e^{At} B B^T e^{A^T t} dt \quad (\text{A.5a})$$

^a G^* : adjungerad matris, dvs G är transponerad och dess komponenter är komplexkonjugerade

$$O = \int_0^{\infty} e^{At} C C^T e^{A^T t} dt \quad (\text{A.5b})$$

Om S och O inte har några diagonalelement som är noll måste balanserade matriser tas till, dvs en transformation $\hat{x} = Tx$ måste hittas sådan att \hat{S} och \hat{O} blir likadana ($\hat{S} = \hat{O} = \Sigma$). Matrisen Σ i sin tur är en diagonalmatris med $\sigma_1, \dots, \sigma_n$ ^b som diagonalelement. Diagonalelementet σ_k beskriver den ”relativa styrbarheten och observerbarheten“ hos systemet för tillstånd \hat{x}_k . Dessa diagonalelement kallas Hankels singulära värden och har egenskapen $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$. Antag att för ett visst r gäller att $\sigma_r \gg \sigma_{r+1}$. Tillstånd som svarar mot $\sigma_{r+1}, \dots, \sigma_n$ är därför mindre styr- och observerbara än de som svarar mot $\sigma_1, \dots, \sigma_r$. Tillstånden $\hat{x}_{r+1}, \dots, \hat{x}_n$ kan då tas bort utan att det påverkar systemet alltför mycket. En sådan reducering minimerar största skillnaden mellan den reducerade modellen och den oreducerade modellen ($\min \|\Delta_{abs}\|_{\infty}$, jämför ekvation 5.4). Ovanstående resonemang kräver att systemet är stabilt.

Metoden förutsätter att tillståndsvariablerna är normaliserade så att de varierar över ett lika stort intervall.

A.4 Cayley-Hamiltons sats

Cayley-Hamiltons sats säger att varje kvadratisk matris A (av dimension n) satisfierar sin egen karakteristiska ekvation (A.6), dvs om

$$\det(\lambda I - A) = \lambda^n + a_1 \lambda^{n-1} + \dots + a_{n-1} \lambda + a_n \quad (\text{A.6})$$

så gäller

$$A^n + a_1 A^{n-1} + \dots + a_{n-1} A + a_n I = 0 \quad (\text{A.7})$$

Ovanstående innebär att varje potens av A , högre eller lika med n kan uttryckas som en linjärkombination av A^k , $k = 0, 1, \dots, n-1$. Ytterligare resonemang kring satsen finns i (Glad & Ljung 1989) och beviset av satsen finns i (Hackman 1991).

^b n är antalet tillstånd

Bilaga B

Modellspecifikt för RM12

B.1 Skalningsmatriser

Skalningar på motormodellen utfördes endast på insignaler och utsignaler. Vid den fysikaliska reduceringen skulle även tillstånden kunna skalas. Dock kan det inte göras för balanserad reducering, eftersom det inte går att skala tillbaka efter reducering har skett. Därför valdes att använda samma skalning för båda reduceringsmetoderna. Skalningen utfördes enligt (B.1)

$$A_{skal} = A \quad (\text{B.1a})$$

$$B_{skal} = BT_u \quad (\text{B.1b})$$

$$C_{skal} = T_y^{-1}C \quad (\text{B.1c})$$

$$D_{skal} = T_y^{-1}DT_u \quad (\text{B.1d})$$

Skalningsfaktorerna i diagonalmatriserna T_u och T_y återfinns i tabell B.1 och i tabell B.2.

Dessa valdes som största skillnad för respektive variabel för alla arbetspunkter.

B.2 Modeller

Den olinjära motormodellen RM12d har 5 insignaler, 32 tillstånd och 13 utsignaler. Utgående från denna har alla designmodeller linjäriserats i PLA 60 grader(SLS). Därefter gjordes en reducering av antalet in- och utsignaler. Insignalen som har tagits bort är WFR (bränslet till efterbrännkammaren). Den togs bort eftersom

Insignal	Skalningsfaktor
WFM	1.1790
A8	0.1200
FVG	41.6000
CVG	33.0000

Tabell B.1: Skalningsfaktorer i matrisen T_u .

Utsignal	Skalningsfaktor
NH	73.876
NL	110.75
PS3C	2213.8
PT25	294.4
TT25	141.5
TT3	344.61
W21	48.854
TT5	539.85

Tabell B.2: Skalningsfaktorer i matrisen T_y .

efterbrännkammaren är släckt i hela det arbetsområde som diagnostiseras. Utsignaler som finns i den olinjära motormodellen men som ej är med vid designen är PT0, TT0, WBL3, M0 och PS0. Dessa utsignaler påverkas ej av modellen (kan lätt ses i tillståndsmatriserna C och D för modellen). Gemensamt för modell 1 till 3 är att de har insignaler enligt tabell B.2 och utsignaler enligt tabell B.4. Däremot skiljer sig antalet tillstånd då individuella reduceringar av antalet tillstånd har gjorts. Innan reduceringen hade motormodellen de tillstånd som återfinns i tabell B.6.

Insignal	Beskrivning	Enhet
WFM	Bränsleflöde till brännkammaren	kg/s
FVG	Fläktens ställbara ledskenor	grad
CVG	Kompressorns ställbara ledskenor	grad
A8	Minsta arean hos utloppsmunstycket	m^2

Tabell B.3: Insignaler till modellen som används vid diagnos.

B.2.1 Modell 1

Designmodell 1 har reducerats på samma sätt som i (Lind 1998). Modellreduktion har gjorts fysikaliskt och i första hand för att estimeras temperaturen i högtrycksturbinen. Modellen reducerades ner till *sju* tillstånd, vilka finns i tabell B.5. De singulära värdena för modell 1 ser ut att överensstämma bra med den oreducerade

Utsignal	Beskrivning	Enhet
NH	Varvtal hos högtrycksrotor	rps
NL	Varvtal hos lågtrycksrotor	rps
PT5	Totala trycket för lågtrycksturbinens utlopp	kPa
PS3	Statiskt tryck hos högtrycksturbinens utlopp	kPa
TT25	Kompressorns inloppstemperatur	K
TT3	Kompressorns utloppstemperatur	K
W21	Luftflödet efter fläkten	kg/s
TT5	Lågtrycksturbinens utloppstemperatur	K

Tabell B.4: Utsignaler till modellen som används vid diagnos.

Tillstånd	Beskrivning
MW31	Gasmassa i volym 31
TT31	Gastemperatur i volym 31
MW39	Gasmassa i volym 39
TT39	Gastemperatur i volym 39
WFMC	Förbränningsdynamik, WFM lagfilter
NH	Varvtal hos högtrycksrotor
NL	Varvtal hos lågtrycksrotor

Tabell B.5: Tillstånd i modell 1.

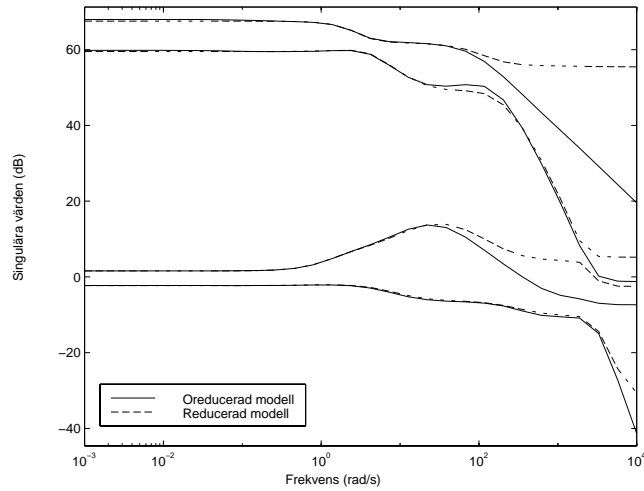
modellen, se figur B.1. Men skenet kan bedra, vilket inses då skillnadssystemets singulära värden beskådas i figur B.2. Om bodediagrammen för alla in- och utsignaler betraktas, skiljer sig dessa kraftigt mellan modell 1 och den oreducerade modellen, för vissa kombinationer. Ett exempel på detta är bodediagrammet från A8 till NH, se figur B.3.

B.2.2 Modell 2

Designmodell 2 har reducerats enligt samma princip som den första, men här har större vikt lagts vid lägre frekvenser. De höga kraven vi ställde på modellen gjorde att inte lika många tillstånd kunde tas bort. Efter reduceringen hade modellen 23 tillstånd (för detaljer se 5.2). De grupper av tillstånd som togs bort var:

1. Snabb dynamik TT15, DY2, FTT558, PLW61, W21FIL, PT0FIL
2. Långsam dynamik T5R, T5S, TM205

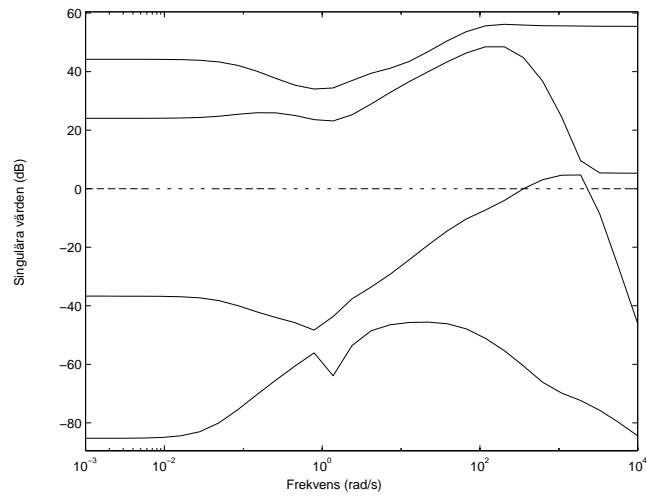
Tillstånden med snabb dynamik har antagits vara oändligt snabba och de med långsam dynamik har antagits vara oföränderliga under ändlig tid. De togs bort med metoden i avsnitt A.1.



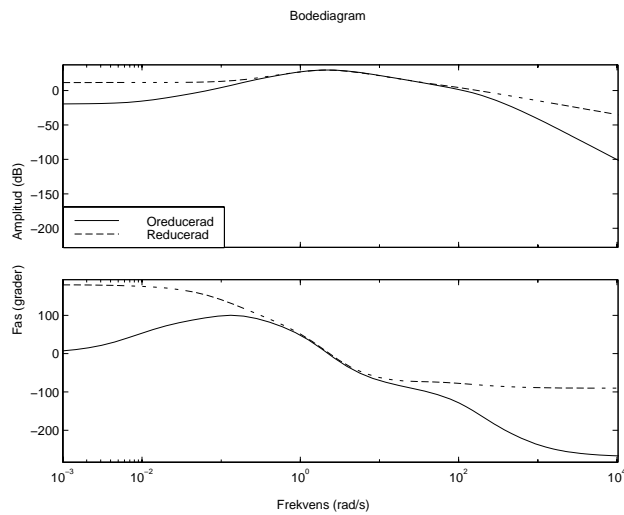
Figur B.1: Singulära värden för 32 och 7 tillstånd (fysikalisk reducering).

B.2.3 Modell 3

Designmodell 3 har reducerats med balanserad reduktion, se avsnitt 5.3. Reducering har i det här fallet renderat i en modell med 12 tillstånd. Då balansering av modellen utfördes som ett delsteg i reduceringen, försvann den fysikaliska tolkningen av tillstånden. De tillstånd som har tagits bort har antagits vara oändligt snabba. Tillstånden togs bort ett i taget. Det visade sig att efter balanseringen kunde de balanserade tillstånden 13 och 17 ej tas bort då de hade för stor inverkan på systemet. Orsaken till att detta inträffade är att skalningen inte normaliserar modellen perfekt.



Figur B.2: Skillnadssystemets singulära värden för 32 och 7 tillstånd (fysikalisk reducering).



Figur B.3: Bodediagram från A8 till NH för modell 1.

Tillstånd	Beskrivning
MW15	Gasmassa i volym 15
TT15	Gastemperatur i volym 15
TM205	Metalltemperatur i fläktstation
BPRFIL	Fläktkanalsfilter
MW31	Gasmassa i volym 31
TT31	Gastemperatur i volym 31
TM27	Metalltemperatur i kompressorstation
MW39	Gasmassa i volym 39
TT39	Gastemperatur i volym 39
TM34	Metalltemperatur i brännkammaren
DY2	Derivata av Y2
Y2	Del i bränsledynamik
WFMC	Förbränningsdynamik, WFM lagfilter
TT5D	Gastemperatur i station 5
PL56	Tryckförändringsfilter i station 56
MW45	Gasmassa i volym 45
TT45	Gastemperatur i volym 45
T5R	Metalltemperatur i lågtryckturbinsrotorn
T5S	Metalltemperatur i lågtryckturbinshåljet
FTT558	Lågtryckturbins utloppstemperatur
TM47	Metalltemperatur i lågtryckturbinen
TM41	Metalltemperatur högtryckturbinstation
PLW61	Tryckförändringsfilter i station 61
PL61	Tryckförändringsfilter i station 61
MW7	Gasmassa i volym 7
TT7	Gastemperatur i volym 7
NH	Varvtal hos högtrycksrotor
NL	Varvtal hos lågtrycksrotor
W158	Fläktkanalsluftflöde
PRL15	Tryckförändringsfilter i station 15
W21FIL	Luftflödesfilter i fläkten
PT0FIL	Omgivningstrycksfilter

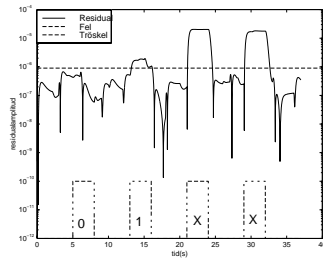
Tabell B.6: Tillstånd i oreducerad modell.

Bilaga C

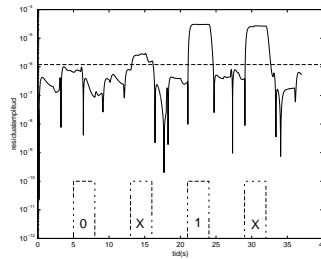
Residualer

Denna bilaga innehåller residualerna för de olika simuleringarna. Residualerna för en viss simulering finns på ett uppslag. Observera att residualernas amplitudskala är logaritmisk. Under varje residual visas när felen inträffar och vilka av dessa fel som residualen ska vara känslig för enligt residualstrukturen i tabell 6.1.

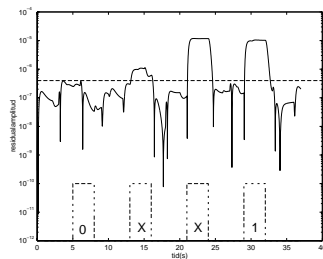
C.1 Residualer för modell 1



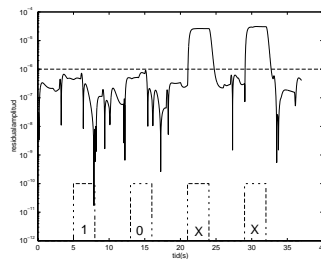
(a) Residual 11



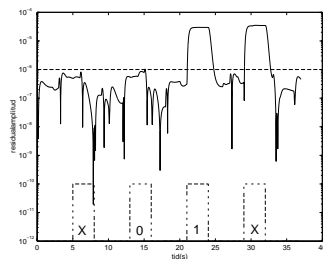
(b) Residual 12



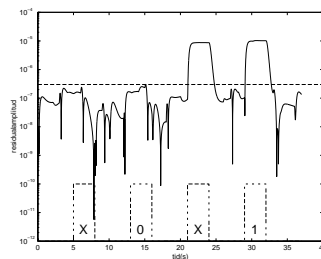
(c) Residual 13



(d) Residual 21

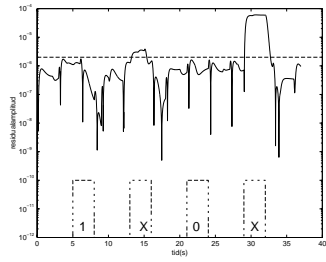


(e) Residual 22

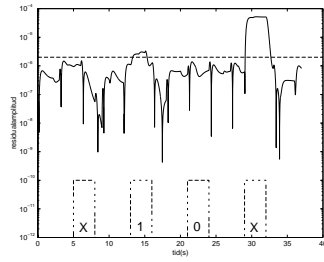


(f) Residual 23

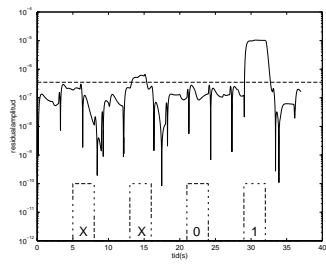
Figur C.1: Residualer för modell 1 (11 - 23).



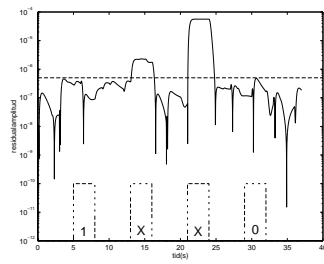
(a) Residual 31



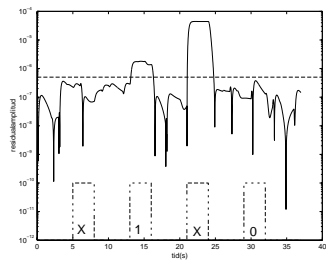
(b) Residual 32



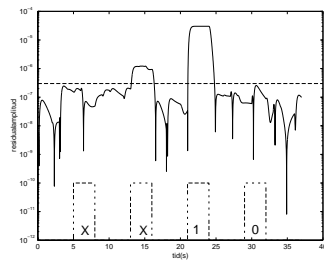
(c) Residual 33



(d) Residual 41



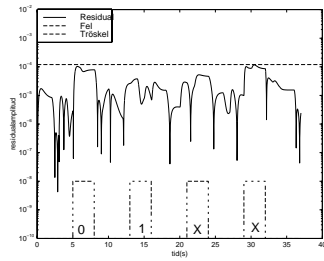
(e) Residual 42



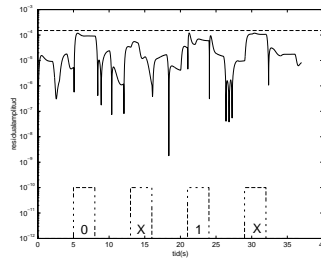
(f) Residual 43

Figur C.2: Residualer för modell 1 (31 - 43).

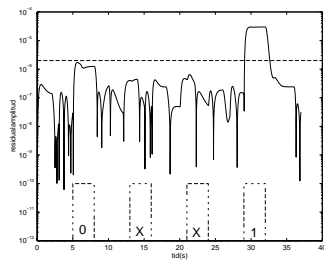
C.2 Residualer för modell 2



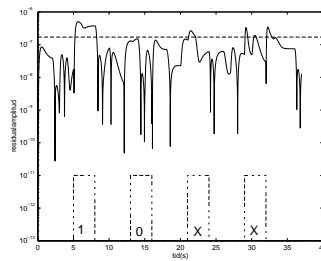
(a) Residual 11



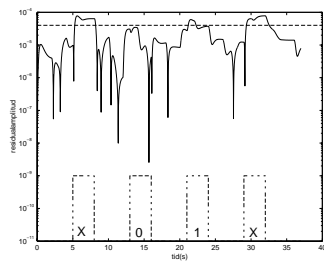
(b) Residual 12



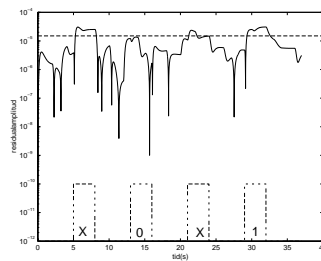
(c) Residual 13



(d) Residual 21

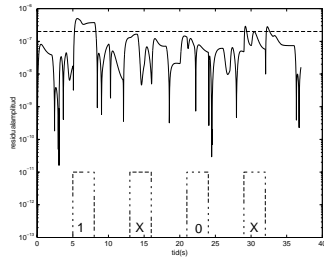


(e) Residual 22

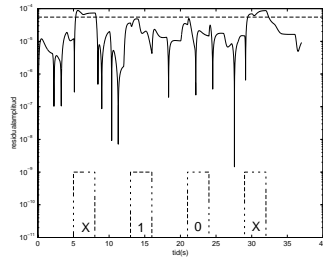


(f) Residual 23

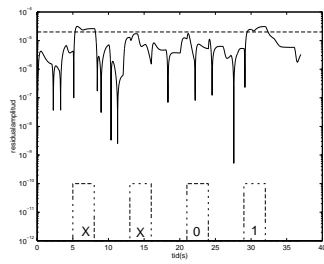
Figur C.3: Residualer för modell 2 (11 - 23).



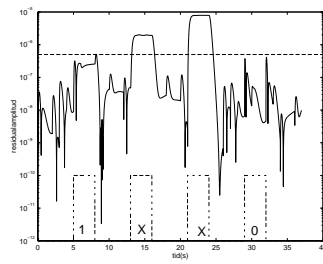
(a) Residual 31



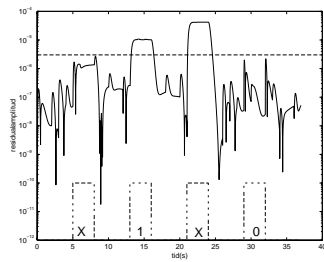
(b) Residual 32



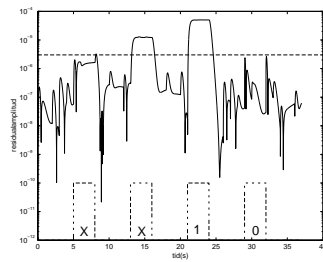
(c) Residual 33



(d) Residual 41



(e) Residual 42

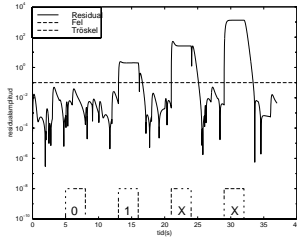


(f) Residual 43

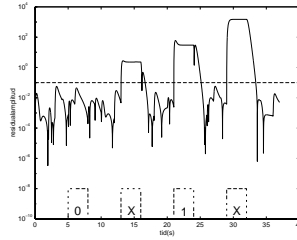
Figur C.4: Residualer för modell 2 (31 - 43).

C.3 Residualer för modell 3

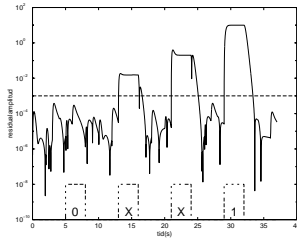
C.3.1 Simulering 1



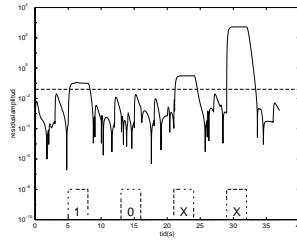
(a) Residual 11



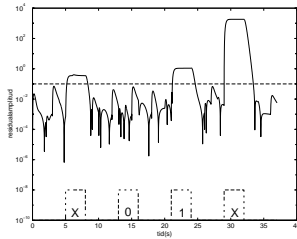
(b) Residual 12



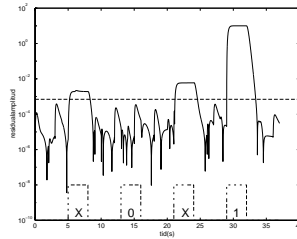
(c) Residual 13



(d) Residual 21

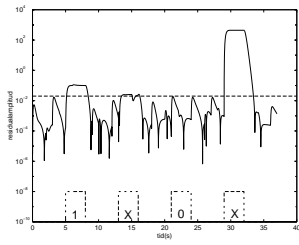


(e) Residual 22

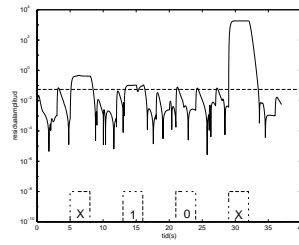


(f) Residual 23

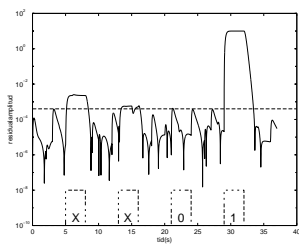
Figur C.5: Residualer för modell 3, simulering 1 (11 - 23).



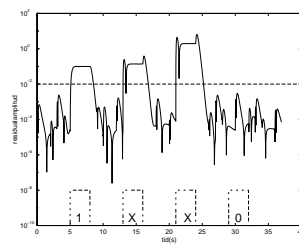
(a) Residual 31



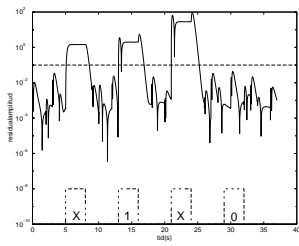
(b) Residual 32



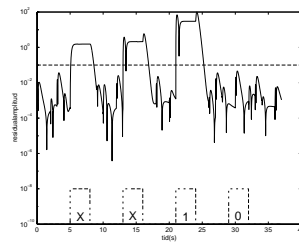
(c) Residual 33



(d) Residual 41



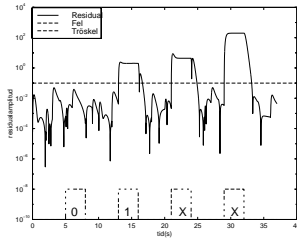
(e) Residual 42



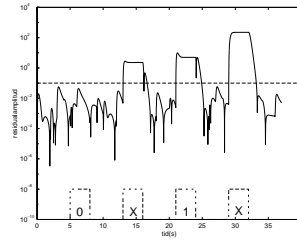
(f) Residual 43

Figur C.6: Residualer för modell 3, simulering 1 (31 - 43).

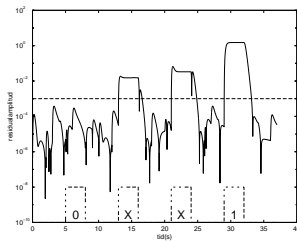
C.3.2 Simulering 2



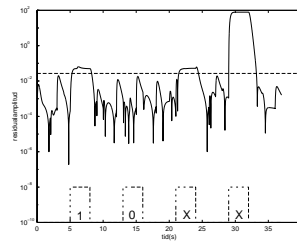
(a) Residual 11



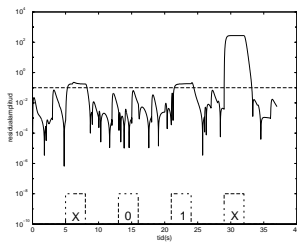
(b) Residual 12



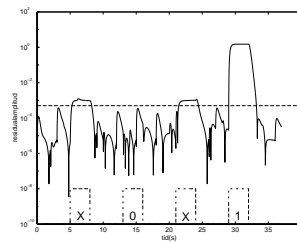
(c) Residual 13



(d) Residual 21

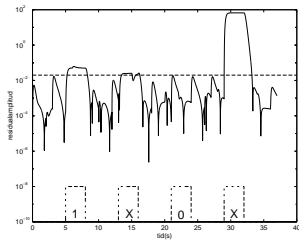


(e) Residual 22

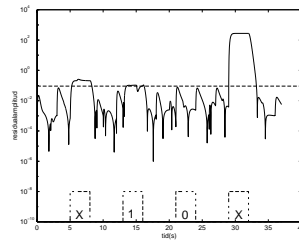


(f) Residual 23

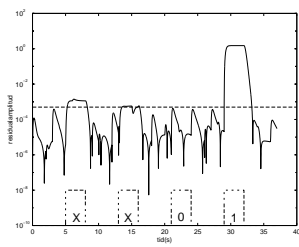
Figur C.7: Residualer för modell 3, simulering 2 (11 - 23).



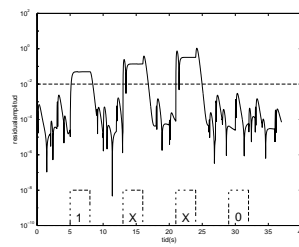
(a) Residual 31



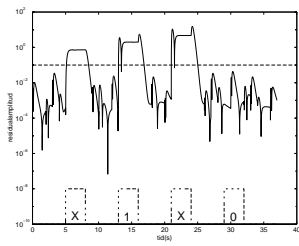
(b) Residual 32



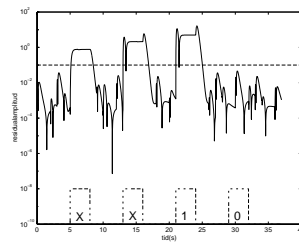
(c) Residual 33



(d) Residual 41



(e) Residual 42



(f) Residual 43

Figur C.8: Residualer för modell 3, simulering 2 (31 - 43).