

Univerza v Ljubljani
Fakulteta za elektrotehniko

Simon Oblak

**INTERVALNI MEHKI MODELI V PREDIKTIVNEM
VODENJU IN ZAZNAVANJU NAPAK**

DOKTORSKA DISERTACIJA

Mentor: prof. dr. Igor Škrjanc, univ. dipl. inž. el.

Ljubljana, 2008

Zahvala

Moja zahvala je v prvi vrsti namenjena staršema za vso podporo, ki sta mi jo nesebično nudila skozi vsa leta šolanja.

Nadálje se zahvaljujem mentorju Igorju in ostalim sodelavcem Laboratorija za modeliranje, simulacijo in vodenje ter Laboratorija za avtomatizacijo in informatizacijo procesov za izkazano pomoč ter prijetno štiriinpolletno sobivanje.

Glede samega dela pa gre zahvala tudi T. Takagiju in M. Sugenu za zelo uporaben mehki model TS, ki sem ga v delu s pridom uporabljal.

Povzetek

Doktorska disertacija obravnava problematiko zaznavanja napak in prediktivnega vodenja v nelinearnih sistemih z negotovimi parametri. Poseben poudarek je na uporabi intervalnega mehkega modela na omenjenih področjih. Intervalni mehki model ali model INFUMO je vpeljan kot univerzalni aproksimator poljubnega nelinearnega področja ali družine nelinearnih sistemov, s katerim lahko zajamemo vplive neznanih vhodov v sistemu. Ker sta izhoda modela zgornja in spodnja meja območja odzivov, lahko govorimo o robustnem opisu obnašanja sistema. Zaradi optimizacije tipa *min-max* vplivi negotovosti se namreč vedno znotraj omejenega območja. Ime "intervalni modelšledi iz njegove strukture, ker sta meji zapisani z matrikama parametrov, ki tvorita zgornjo in spodnjo mejo intervalov parametrov.

Predstavljena je bila metoda za identifikacijo parametrov splošnega modela INFUMO na podlagi linearnega programiranja. Metoda je bila uporabljena in predstavljena na primerih družine statičnih krivulj, dinamičnega nelinearnega sistema z negotovimi parametri ter modela tipa Wiener-INFUMO, ki je kombinacija dinamičnega linearnega sistema z negotovostmi v parametrih in statične izhodne preslikave.

V zaznavanju napak smo model INFUMO uporabili v obliki, sorodni pasivnim robustnim metodam, za nelinearni proces z negotovimi parametri, katerega izhod v normalnem načinu delovanja (brez prisotnosti napak) opišemo z družino izhodnih funkcij. Za pridobivanje ti. pasu zaupanja smo uporabili nizkopasovni filter, ki predstavlja konsenz med robustnostjo na lažne alarme in hitrostjo zaznave. Pokazali smo, da metoda omogoča tako zaznavanje kot lokalizacijo napak. Na enak na način kot pri prejšnji metodi je bila razvita tudi inačica z visokopasovnim filtrom. Pri obeh primerih smo opravili študijo zaznavnosti, ki pokaže, da mora biti za zaznavo napake njen vpliv v najslabšem primeru enak širini pasu izhodov. Primerjava rezultatov pri uporabi obeh filtrov je dala nekoliko več prednosti metodi z visokopasovnim filtrom, predvsem glede natančnosti označevanja periode napak.

V okviru nelinearnega prediktivnega vodenja smo model INFUMO uporabili pri razvoju metode, sorodne posplošenem prediktivnem vodenju (ang. generalized predictive control, GPC) v zveznem času na podlagi modela Wiener-INFUMO. Metoda je primerna za procese z negotovimi parametri in izrazito statično nelinearnostjo, katerih izhode je moč opisati z družinami funkcij. Ciljni procesi za praktično izvedbo metode so predvsem v kemični, farmacevtski in petrokemični industriji. Izkazalo se je,

da predlagana metoda omogoča tudi kombinacijo sistemov za vodenje in zaznavanje napak.

V sodelovanju z univerzo Istanbul Technical University smo razvili tudi metodo prediktivnega vodenja na podlagi moderne genetske optimizacije Big Bang - Big Crunch (BB-BC). Metoda temelji na ideji inverznega mehkega regulatorja, kjer je v vsakem koraku treba izvesti prilagajanje modela procesu. Z modelom INFUMO smo identificirali pas možnih parametrov procesa in omejili iskalni prostor optimizacijske metode, kar je prineslo močno zmanjšanje računskega bremena v vsakem računskem koraku.

Vsi rezultati so bili pokazani na ilustrativnih primerih, ki kažejo dobre obete za prenos metodologij v praktično uporabo.

Abstract

The doctoral thesis deals with fault detection and predictive control for nonlinear uncertain systems. Special attention was given to the use of the interval fuzzy model (INFUMO) in these areas. INFUMO is a universal approximator of an arbitrary nonlinear area or a nonlinear-function family. We can use it for describing the influences of the unknown system inputs. We refer to a robust description of the system behavior because the model outputs are the upper and the lower limit of the output family. Due to the *min-max* optimization, all the uncertainty influences can be found inside the boundaries. The name *interval model* follows from the model's structure as the parameter matrices form the interval values for each parameter therein.

We presented the identification method for obtaining parameters of a general INFUMO model employing linear programming. The method was then used in the following examples: a static-curve family, a dynamic nonlinear uncertain system, and a Wiener-INFUMO model, which combines a dynamic parameter-uncertain linear system with a static output mapping.

With respect to fault detection we used the INFUMO, in a form that is closely related to the passive robust methods, for a nonlinear uncertain process whose output in a normal operating mode (without faults) can be described by an output-function family. For obtaining the so-called confidence band we used a low-pass filter, which functions as a trade-off between the robustness to false alarms and the detection speed. We have shown that the method enables both fault detection and localization. In the same way we developed a form where a high-pass filter is used. In both cases we conducted a detectability study, which showed that the worst-case fault magnitude should exceed the corresponding width of the confidence band. Comparing the results for both filters puts the high-pass form in advantage over the low-pass one, especially regarding the accuracy in fault-period matching.

The INFUMO was also employed in a nonlinear Wiener-INFUMO-model-based continuous-time predictive control method, which is related to generalized predictive control (GPC). The method is suitable for uncertain processes with an explicit static nonlinearity whose outputs can be described by function families. It was shown that the proposed method also enables the combination of the predictive-control method and on-line fault detection.

In cooperation with Istanbul Technical University we developed a predictive-control method based on a modern genetic optimization method called Big Bang - Big Crunch (BB-BC). The method is founded on the inverse fuzzy controller where model adaptation to the process should be done in each calculation step. The INFUMO was used to identify a band of possible process parameter values and hence to limit the optimization search space, which resulted in a considerable reduction of computational load.

All of the results were demonstrated in illustrative examples, and they indicate a good promise for applying the proposed method in practice.

Vsebina

Seznam slik	xiii
1. Uvod	1
2. Intervalni mehki model	5
2.1 Nelinearni sistem, zapisan v obliki mehkega modela tipa TS	5
2.2 Identifikacija mehkega modela z uporabo norme l_∞	7
2.3 Identifikacija intervalnega mehkega modela	8
2.3.1 Primer: aproksimacija družine krivulj	10
2.4 Model Wiener-INFUMO	13
2.4.1 Primer: modeliranje nelinearnega dinamičnega sistema z modelom Wiener-INFUMO	15
3. Napredne metode zaznavanja napak	19
3.1 Osnovna strategija zaznavanja napak na podlagi modela	20
3.2 Zaznavanje napak na podlagi kvantitativnega modela	21
3.2.1 Ocenjevanje parametrov	21
3.2.2 Pristopi na podlagi observatorja stanj	22
3.2.3 Pristopi na podlagi paritetnih enačb	23
3.2.4 Pristopi na podlagi nelinearnega observatorja	26
3.2.5 Pristopi s pomočjo adaptivnega observatorja	30
3.2.6 Robustno aktivno oblikovanje residualov	33
3.2.7 Pasivno oblikovanje residualov	37
3.3 Zaznavanje napak na podlagi kvalitativnega modela	39

3.3.1	Usmerjeni grafi	40
3.3.2	Drevo napak	42
3.4	Zaznavanje napak na podlagi zgodovine procesa	43
3.4.1	PCA	43
3.4.2	Statistične metode	46
3.4.3	Umetne nevronske mreže	47
3.4.4	Mehki modeli	50
3.5	Lokalizacija napak	52
3.5.1	Boolovo sklepanje	52
3.5.2	Aproksimativno sklepanje	53
4.	Prediktivno vodenje na podlagi mehkih modelov	57
4.1	Osnovni elementi prediktivnega vodenja na podlagi modela	58
4.1.1	Model procesa	60
4.1.2	Kriterijska funkcija	61
4.1.3	Optimizacijski problem	61
4.2	Uporaba mehkega modela v prediktivnem vodenju	62
4.2.1	Prediktivno vodenje na podlagi mehkega modela Takagi-Sugeno .	62
4.2.2	Nekatere druge metode na podlagi mehkega modela	65
4.2.3	Wienerjevo prediktivno vodenje	66
5.	Uporaba modela INFUMO v zaznavanju napak	69
5.1	Opis zasnove sistema	69
5.2	Študija zaznavnosti	72
5.3	Zaznavanje in lokalizacija napak na podlagi nizkopasovnega filtriranja .	74
5.3.1	Zaznavanje napak za model dveh shranjevalnikov	75
5.3.2	Lokalizacija napak za model dveh shranjevalnikov	80
5.4	Zaznavanje in lokalizacija napak na podlagi visokopasovnega filtriranja .	87
5.4.1	Lokalizacija napak za model dveh shranjevalnikov	89

6. Prediktivno vodenje na podlagi modela Wiener-INFUMO v zveznem času	93
6.1 Uvodni pojmi	93
6.1.1 Strategija vodenja	93
6.1.2 Napoved izhoda procesa z modelom Wiener-INFUMO	94
6.2 Izpeljava prediktivnega regulacijskega zakona	97
6.2.1 Obravnava stabilnosti	100
6.3 Kombinacija sistema vodenja in zaznavanja napak	101
6.3.1 Primer: vodenje CTWMPC in zaznavanje napak za pH-proces	101
7. Prediktivno vodenje na podlagi modela INFUMO in optimizacije BB-BC	107
7.1 Optimizacija Big Bang - Big Crunch (BB-BC)	107
7.2 BB-BC v prediktivnem vodenju na podlagi modela INFUMO	108
7.2.1 Primer	110
8. Sklep	115
Dodatek A - Uporaba modela INFUMO v zaznavanju napak (NPF)	117
Dodatek B - Uporaba modela INFUMO v zaznavanju napak (VPF)	127
Dodatek C - Uporaba metode prediktivnega vodenja CTWMPC	135
Literatura	149

Seznam slik

2.1	Množica podatkov in centri rojev s pripadnostnimi funkcijami	11
2.2	Aproksimacija družine krivulj z modelom INFUMO, tip Singleton	12
2.3	Aproksimacija družine krivulj z modelom INFUMO, izhodne funkcije 3. reda	13
2.4	Shema Wienerjevega modela	13
2.5	Shema modela Wiener-INFUMO	14
2.6	Vhodno-izhodne meritve za identifikacijo	15
2.7	Aproksimacija ter preslikava pripadnostnih funkcij za inverzni model INFUMO	16
2.8	Verifikacija identificiranega modela Wiener-INFUMO	18
3.1	Zaznavanje napak na podlagi modela procesa	20
3.2	Paritetne enačbe na podlagi izhodnega pogreška	24
3.3	Metoda paritetnih enačb za model MIMO v prostoru stanj	25
3.4	Usmerjeno (levo) in strukturirano občutljivi residuali (desno)	26
3.5	Uporaba adaptivnega observatorja za oblikovanje residuala	32
3.6	Splošna shema oblikovanja robustnega residuala	34
3.7	Primer usmerjenega grafa za shranjevalnik tekočine	40
3.8	Oblikovanje residuala s pomočjo UNM	48
3.9	Aproksimacija dinamičnega sistema z UNM	49
3.10	Uporaba UNM za vrednotenje residualov	50
3.11	Uporaba mehkega modela za oblikovanje residualov	50
3.12	Uporaba mehkega modela za lokalizacijo napake	51
3.13	Mehka odločitvena funkcija	54

4.1	Osnovni princip prediktivnega vodenja	59
4.2	Shema prediktivnega vodenja	60
4.3	Shema Wienerjevega prediktivnega vodenja	66
5.1	Shema aproksimacije pasu zaupanja z modelom INFUMO	71
5.2	Shema oblikovanja robustnega residuala z modelom INFUMO	72
5.3	Problematika zaznavanja majhnih napak	74
5.4	Zaznavanje napak z nizkopasovnim filtrom, modelom INFUMO in nor- miranim residualom	75
5.5	Proces dveh shranjevalnikov	76
5.6	Meritve izhoda pri danih vhodih: prva, druga in zadnja meritev	77
5.7	Izbira lomne frekvenca filtra	78
5.8	Množica filtriranih podatkov z mejnimi točkami, omejene z modelom INFUMO	79
5.9	Postavitev pripadnostnih funkcij	80
5.10	Vhod in izhod v testnem poskusu	81
5.11	Rezultati zaznavanja napak	81
5.12	Zaznavanje napak z nizkopasovnim filtrom, modelom INFUMO in nor- miranim residualom	83
5.13	Postavitev pripadnostnih funkcij	84
5.14	Množica filtriranih podatkov z mejnimi točkami, omejene z modelom INFUMO	85
5.15	Rezultati zaznavanja napak - residuali 1, 2 in 3	86
5.16	Rezultati lokalizacije napak	86
5.17	Zaznavanje in lokalizacija napak z visokopasovnim filtrom	87
5.18	Množica podatkov (VPF) z mejnimi točkami, omejene z modelom IN- FUMO	89
5.19	Rezultati zaznavanja napak (VPF) - zgoraj: odločitvene funkcije, spodaj: test prehodnega pojava	90
5.20	Rezultati zaznavanja napak (VPF) - funkcije napak	90
5.21	Rezultati lokalizacije napak (VPF)	91

6.1	Izračun kriterijske funkcije	98
6.2	Kombinirani sistem prediktivnega vodenja in zaznavanja napak na podlagi modela Wiener-INFUMO	102
6.3	Shema pH-procesa	102
6.4	Vhodno-izhodni podatki procesa (levo), odziv linearnega modela (desno zgoraj), izhodna preslikava (desno spodaj)	104
6.5	Aproksimacija z modelom Wiener-INFUMO (levo) in verifikacija modela (desno)	105
6.6	Rezultati vodenja pri pojavu dveh napak	105
6.7	Rezultati zaznavanja napak v zaprti zanki	106
7.1	Shema vodenja na podlagi modela INFUMO in optimizacije BB-BC	109
7.2	Interval variabilnega ojačenja za testni proces	110
7.3	Pas meritev in omejitev z modelom INFUMO	111
7.4	BBBC, zaprtozančni preizkus: izhod (zgoraj), vhod (spodaj)	112
7.5	Prilagajanje parametrov med delovanjem	112

1. Uvod

Modeliranje in identifikacija modela procesa sta ključna tako pri načrtovanju sistemov vodenja na podlagi modela kot pri načrtovanju sistemov za zaznavanje in diagnostiko napak. Pri nelinearnem vodenju, kjer na podlagi modela računamo napoved izhoda sistema, mora biti napoved zadosti natančna, poleg tega pa mora model omogočati vpogled v delovanje sistema, da lahko pri načrtovanju vodenja upoštevamo različne delovne pogoje in ustrezno spreminjamo parametre regulatorja. Posebno pozornost pa mora načrtovalec v tem primeru vedno posvetiti tudi odstopanjem modela od realnega procesa, ki bi lahko v neugodnih okoliščinah privedle do destabilizacije sistema vodenja. Pri nelinearnih metodah zaznavanja napak pa neustrezen model povzroča dvoje problemov: lažni alarmi, kjer prihaja do prevelikega vpliva motenj in nemodelirane dinamike na residual, ali prepočasno odzivanje na napake (v najhujših primerih tudi izpuščanje alarmov), če so mere za napako zaradi neustreznega modela nastavljene previsoko.

Največ težav pri modeliranju in identifikaciji povzročajo nelinearni sistemi, ki imajo negotove parametre ali celo negotovo strukturo. Vzrok negotovosti je ponavadi ne-natančno poznavanje fizikalnih zakonitosti znotraj procesa. Klasični pristopi matematičnega modeliranja na podlagi fizikalnih zakonov in ravnovesij takrat odpovejo. Pri uporabi klasičnih metod eksperimentalnega modeliranja pa naletimo na težave že pri pridobivanju identifikacijskih podatkov. Za take sisteme je namreč značilno, da večkratno vzbujanje z istim vzbujevalnim signalom daje neenake odzive, bolj ali manj razpršene po nekem nelinearnem območju. Če za take sisteme načrtujemo vodenje, lahko pričakujemo težave pri zagotavljanju ustrezne kvalitete vodenja na širokem delovnem področju, v primeru večje negotovosti parametrov pa tudi težave s stabilnostjo sistema, kar je v praksi absolutno nesprejemljivo. Pri načrtovanju sistemov za zaznavanje napak pa se negotovosti parametrov kažejo kot motnje in nemodelirana dinamika, v splošnem to pojmujeemo kot neznane vhode v sistem. Največja težava je natančno določiti vplive neznanih vhodov na residual, kar onemogoča uporabo klasičnih metod, kjer morajo biti vplivi motenj in napak znani, pri ostalih metodah pa to vodi v nezanesljivost napovedi napake (lažni alarmi ali slaba občutljivost na napake).

Iz pregleda literature zadnjega obdobja je razvidno, da se v večini primerov za reševanje takih problemov uporabljajo metode robustne analize in robustnega vodenja. Vendar pa je analitična obravnava zelo zapletena že pri najbolj enostavnih sistemih, kar močno otežuje prenos v inženirsko prakso, ker zahteva visoko stopnjo matematičnega znanja. Poleg tega je zelo težko natančno oceniti negotovosti v sistemu (frekvenčna vsebina, lokacije v sistemu), kar še dodatno otežuje analizo. Drugi način določanja negotovosti je uporaba intervalne aritmetike, ki predpostavlja, da se negotovi parametri nahajajo na omejenih intervalih. V literaturi lahko zaznamo kar nekaj možnosti uporabe intervalne aritmetike v sistemih za zaznavanje napak, vendar pa še vedno prevladujejo metode, ki temeljijo na linearnih sistemih. Pojavlja se torej potreba po uvedbi modela, ki bi lahko obravnaval tako linearne kot nelinearne negotove sisteme, zajemal negotovosti parametrov preko intervalne analize in bil vsestransko uporaben - tako v prediktivnem vodenju kot v zaznavanju napak. Poleg tega bi moral biti postopek načrtovanja omenjenih sistemov preprost, intuitiven in računsko nezahteven, da bi bila metoda prenosljiva tudi v prakso.

Intervalni mehki model (INFUMO) je bil razvit v sodelovanju laboratorija LMSV&LAIP s Fakultete za elektrotehniko in argentinsko Universidad Nacional del Sur za potrebe robustnega opisa nelinearnih sistemov z negotovimi parametri. Kmalu se je izkazalo, da identifikacija parametrov takih sistemov daje tudi možnost za opis negotovosti v sistemu. Postopek identifikacije poteka na vhodno-izhodnih podatkih iz delovanja procesa, kar je pomembno predvsem v inženirski praksi. Metoda identifikacije je enostavna in računsko dobro pogojena, ravno tako je jasen in strnjen tudi zapis modela. Na podlagi izkušenj in rezultatov uvodnih raziskav smo ocenili, da je model INFUMO primeren za uporabo na več nivojih, tako v moderni teoriji vodenja in zaznavanja napak kot tudi v zahtevnem inženirskem in tehnološkem nivoju. Pričujoča disertacija je plod dolgoletnih raziskav uporabe modela INFUMO na teh področjih, s posebnim poudarkom na nelinearnih sistemih. Poleg celovitega pregleda naprednih metod zaznavanja napak na podlagi modela procesa ter opisa metod prediktivnega vodenja na podlagi mehkega modela so obravnavane naslednje teme:

- vpeljava intervalnega mehkega modela kot univerzalnega aproksimatorja poljubnega nelinearnega področja ali družine nelinearnih sistemov;
- možnost reševanja problemov identifikacije pri nelinearnih sistemih z negotovimi parametri;
- možnost vpeljave intervalnega mehkega modela v prediktivno vodenje s poudar-

kom na Wienerjevih sistemih z negotovimi parametri, s katerimi opisujemo sisteme z negotovo linearno dinamiko in močno nelinearno izhodno preslikavo;

- predlog novih metod v zaznavanju in diagnostiki napak z vpeljavo intervalnega mehkega modela.

Disertacijo sestavlja osem poglavij. Po uvodnem poglavju podajamo opis modela INFUMO, metode identifikacije intervalnih parametrov in praktične primere uporabe. V tretjem poglavju so opisane glavne metode zaznavanja napak na podlagi modela procesa, njihove dobre in tudi slabe lastnosti. Četrto poglavje se nanaša na metode prediktivnega vodenja na podlagi mehkega modela. Peto poglavje predstavlja uporabo modela INFUMO v zaznavanju napak, opis obeh inačic predobdelave podatkov in primerjavo rezultatov po obeh inačicah. V šestem poglavju je izpeljava metode prediktivnega vodenja na podlagi modela Wiener-INFUMO v zveznem času. V praktičnem primeru obravnavamo kombinirano metodo vodenja in sprotnega zaznavanja napak v zaprtozančnem sistemu, ki je s stališča praktične uporabe tudi najbolj atraktivna. Sedmo poglavje predstavi metodo Big Bang - Big Crunch v prediktivnem vodenju na podlagi modela INFUMO. Osmo poglavje podaja sklepe in glavne doprinose disertacije znanosti.

Poleg omenjenih poglavij ima disertacija tudi tri dodatke, v katerih so predstavljeni praktični rezultati obravnavanih metod v smislu objav člankov v uglednih mednarodnih revijah (indeks SCI) in na uglednih mednarodnih konferencah.

2. Intervalni mehki model

V tem poglavju bomo predstavili intervalni mehki model ali model INFUMO, ki je bil razvit za potrebe robustnega modeliranja in robustne identifikacije negotovih nelinearnih sistemov. Metoda, opisana v [127], je posplošitev pristopa robustne identifikacije s kanoničnimi odsekoma-zveznimi funkcijami [3]. Glavna motivacija za uvedbo takega modela je uporaba na področjih opisa nelinearnih sistemov z negotovimi parametri, kot so npr. nelinearna elektronska vezja [127], Wienerjevih sistemov [128], kompresije podatkov in zaznavanja napak za nelinearne sisteme.

2.1 Nelinearni sistem, zapisan v obliki mehkega modela tipa TS

Tipični model tipa Takagi-Sugeno (TS) [117] podamo v obliki pravil

$$\begin{aligned} \mathbf{R}_j : \text{če } x_{p1} \text{ je } \mathbf{A}_{1,k_1} \text{ in } \dots \text{ in } x_{pq} \text{ je } \mathbf{A}_{q,k_q}, \text{ potem } y = \phi_j(\mathbf{x}), \\ j = 1, \dots, m, k_1 = 1, \dots, f_1, k_2 = 1, \dots, f_2, \dots, k_q = 1, \dots, f_q. \end{aligned} \quad (2.1)$$

Vektor $\mathbf{x}_p = [x_{p1}, x_{p2}, \dots, x_{pq}]^T$ predstavlja vhod oz. spremenljivke pogojne (ang. antecedent) strani, z y pa označujemo izhod modela. Vsak x_{pi} ($i = 1, \dots, q$) iz pogojne strani je povezan s f_i mehкими množicami ($\mathbf{A}_{i,1}, \dots, \mathbf{A}_{i,f_i}$), vsaka mehka množica \mathbf{A}_{i,k_i} ($k_i = 1, \dots, f_i$) pa je zaznamovana z realno funkcijo $\mu_{A_{i,k_i}}(x_{pi}) : \mathbb{R} \rightarrow [0, 1]$, ki pona- zarja pripadnost x_{pi} množici \mathbf{A}_{i,k_i} . Če upoštevamo vse možne kombinacije pravil v 2.1, je celotno število pravil enako $m = f_1 \times f_2 \times \dots \times f_q$. Poleg spremenljivk v \mathbf{x}_p pa so vhodi mehkega sistema tudi elementi vektorja posledične (ang. consequent) strani, $\mathbf{x}^T = [x_1, x_2, \dots, x_n, 1]$. Funkcije $\phi_j(\cdot)$ so v splošnem poljubne gladke funkcije, čeprav se največkrat zaradi enostavnosti uporablja linearne ali afine funkcije.

Sistem iz enačbe (2.1) lahko zapišemo v eksplicitni obliki, če predhodno definiramo presek mehkih množic. Posplošena oblika preseka se imenuje *trikotna norma* (ang. triangular, T-norm). V našem primeru privzamemo algebrajski produkt pripadnostnih funkcij. Izhod modela lahko tako zapišemo kot

$$y = \frac{\sum_{k_1=1}^{f_1} \dots \sum_{k_q=1}^{f_q} \mu_{A_1,k_1}(x_{p1}) \dots \mu_{A_q,k_q}(x_{pq}) \phi_j(\mathbf{x})}{\sum_{k_1=1}^{f_1} \dots \sum_{k_q=1}^{f_q} \mu_{A_1,k_1}(x_{p1}) \dots \mu_{A_q,k_q}(x_{pq})}. \quad (2.2)$$

Pri tej enačbi je treba upoštevati, da se vrednosti indeksa j razlikujejo pri različnih naborih indeksov $k_i, i = 1, \dots, q$, kot je razvidno iz enačbe (2.1). Enačbo (2.2) lahko poenostavimo, če privzamemo, da pripadnostne funkcije tvorijo particijo enote (ang. partition of unity) na celotni vhodni domeni. Pripadnostne funkcije, definirane kot

$$\beta_j(\mathbf{x}_p) = \frac{\mu_{A_1, k_1}(x_{p1}) \cdots \mu_{A_q, k_q}(x_{pq})}{\sum_{k_1=1}^{f_1} \cdots \sum_{k_q=1}^{f_q} \mu_{A_1, k_1}(x_{p1}) \cdots \mu_{A_q, k_q}(x_{pq})}, \quad j = 1, \dots, m, \quad (2.3)$$

na ta način izražajo stopnjo izpolnjenosti posameznega pravila v normalizirani obliki. Enostavno je pokazati, da velja $\sum_{j=1}^m \beta_j(\mathbf{x}_p) = 1$ neodvisno od izbire \mathbf{x}_p , če imenovalec v $\beta_j(\mathbf{x}_p)$ ni enak 0. To lahko dosežemo tako, da razširimo območje pripadnostnih funkcij čez celotno potencialno območje \mathbf{x}_p . Če združimo enačbi (2.1) in (2.2) ter zamenjamo seštevanje po k_i s seštevanjem po j , pridemo do zapisa

$$y = \sum_{j=1}^m \beta_j(\mathbf{x}_p) \phi_j \mathbf{x}. \quad (2.4)$$

Enačba (2.4) kaže, da je izhod mehkega sistema odvisen od vektorja \mathbf{x}_p dimenzije q in vektorja \mathbf{x} dimenzije n . Dimenzija vhodnega prostora pa je le redko enaka $q + n$, ker spremenljivke pogojne strani običajno nastopajo tudi v posledičnem delu. V splošnem lahko zapišemo

$$\max(q, n) \leq d \leq q + n. \quad (2.5)$$

Zato definiramo $\mathbf{z} \in \mathbb{R}^d$ kot vhodni vektor, sestavljen iz elementov \mathbf{x}_p in tistih elementov \mathbf{x} , ki ne nastopajo v \mathbf{x}_p .

Zelo pogosto izhodno funkcijo definiramo kot linearno kombinacijo stanj posledične strani

$$\phi_j(\mathbf{x}) = \boldsymbol{\theta}_j^T \mathbf{x}, \quad j = 1, \dots, m, \quad \boldsymbol{\theta}_j^T = [\theta_{j1}, \dots, \theta_{jn}]^T. \quad (2.6)$$

Če izberemo model tipa TS ničtega reda, je $\phi_j(\mathbf{x}) = \theta_{j0}$, pri modelu prvega reda pa $\phi_j(\mathbf{x}) = \theta_{j0} + \boldsymbol{\theta}_j^T \mathbf{x}$. Oba primera lahko obravnavamo z zapisom v (2.6), če dodamo element 1 v vektor \mathbf{x} in θ_{j0} v vektor $\boldsymbol{\theta}_j^T$. Predpostavimo enoten zapis iz (2.6). Če zapišemo matriko koeficientov celotnega nabora pravil kot $\boldsymbol{\Theta}^T = [\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_m]$ in vektor pripadnosti kot $\boldsymbol{\beta}^T(\mathbf{x}_p) = [\beta_1(\mathbf{x}_p), \dots, \beta_m(\mathbf{x}_p)]$, potem lahko enačbo (2.4) poenostavimo v

$$y = \boldsymbol{\beta}^T(\mathbf{x}_p) \boldsymbol{\Theta} \mathbf{x}. \quad (2.7)$$

Mehki model (2.7) navadno imenujemo afini TS-model. Z njim lahko poljubno natančno aproksimiramo poljubno funkcijo, ki preslika kompaktno množico $\mathbf{C} \subset \mathbb{R}^d$ v množico \mathbb{R} [136]. Z uporabo Stone-Weierstrassovega teorema lahko pokažemo, da je tak model univerzalni aproksimator, ki lahko uniformno aproksimira poljubno zvezno funkcijo, v delu [136] pa so bili podani tudi potrebni pogoji za eksistenco take aproksimacije.

2.2 Identifikacija mehkega modela z uporabo norme l_∞

Parametre mehkega modela lahko ocenimo z identifikacijo, kjer normo l_∞ uporabimo za mero pogreška modela. Predpostavimo N meritev v sistemu in mehki model v obliki 2.1. Sestavimo množico pogojnih vektorjev $\mathbf{X}_p = \{\mathbf{x}_{p1}, \mathbf{x}_{p2}, \dots, \mathbf{x}_{pN}\}$ in množico posledičnih vektorjev $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$, iz katerih oblikujemo vhodno množico $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N\}$. Množica \mathbf{Z} predstavlja meritve vhodov v sistem, zbrane iz kompaktne množice $\mathbf{S} \subset \mathbb{R}^d$. Poleg tega imejmo na voljo množico pripadajočih meritev izhodov sistema $\mathbf{Y} = \{y_1, y_2, \dots, y_N\}$. Meritve zadostijo nelinearni preslikavi

$$y_i = g(\mathbf{z}_i), \quad i = 1, \dots, N. \quad (2.8)$$

Po Stone-Weierstrassovem teoremu za poljubno realno zvezno funkcijo g kompaktne množice $\mathbf{U} \subset \mathbb{R}^d$ in poljuben $\varepsilon > 0$ obstaja mehki model f , za katerega velja

$$\max_{\mathbf{z}_i \in \mathbf{Z}} |f(\mathbf{z}_i) - g(\mathbf{z}_i)| < \varepsilon, \quad \forall i. \quad (2.9)$$

Z drugimi besedami: mehki model iz razreda \mathcal{F}_d , definiran z enačbo (2.7), lahko uporabimo za aproksimacijo poljubne zvezne funkcije s poljubno natančnostjo. po drugi strani pa ne smemo prezreti, da manjše vrednosti ε zahtevajo večjo vrednost m oziroma večje število pravil mehkega sistema. Pri aproksimaciji pogrešek mehkega modela proti merjenim vrednostim izrazimo kot

$$e_i = y_i - f(\mathbf{z}_i), \quad i = 1, \dots, N. \quad (2.10)$$

Za oceno optimalnih parametrov mehkega modela f je treba poiskati minimalno vrednost maksimalnega pogreška

$$\max_{\mathbf{z}_i \in \mathbf{Z}} |y_i - f(\mathbf{z}_i)| \quad (2.11)$$

po celotni vhodni domeni \mathbf{Z} . Ker pa je pri mehkem sistemu (2.7) domena \mathbf{Z} sestavljena iz vektorjev pogojne in posledične strani, minimizacijo enačbe (2.11) izvedemo v dveh korakih. Prvi problem je zmanjševanje pogreška glede na pogojno stran \mathbf{x}_p . To naredimo z optimalno razporeditvijo pripadnostnih funkcij, kar je znan problem s področja mehkih sistemov. Navadno se rešuje s pomočjo algoritmov rojenja (ang. clustering) [6], na tem mestu za primer omenimo samo algoritma *c-means* in *Gustafson-Kessel* (GK). Ko so pripadnostne funkcije razporejene, je struktura modela znana in v drugem koraku z optimizacijsko metodo *min-max* poiščemo parametre, ki zagotovijo minimalni pogrešek po enačbi (2.11):

$$\Theta = \arg \min_{\Theta} \max_{\mathbf{z}_i \in \mathbf{Z}} |y_i - \beta^T(\mathbf{x}_{pi})\Theta \mathbf{x}_i| \quad (2.12)$$

Enačbo (2.12) lahko rešujemo z linearnim programiranjem, o čemer govori naslednja lema.

Lema 2.1. *Optimizacijski problem tipa min-max lahko rešujemo s pomočjo linearnega programiranja kot minimizacijo λ po parametru θ_j , $j = 1, \dots, m$, glede na neenačbi*

$$\begin{aligned} y_i - \sum_{j=1}^m \beta_j(\mathbf{x}_p) \theta_j^T \mathbf{x}_i &\leq \lambda, \quad i = 1, 2, \dots, N, \\ -y_i + \sum_{j=1}^m \beta_j(\mathbf{x}_p) \theta_j^T \mathbf{x}_i &\leq \lambda, \quad i = 1, 2, \dots, N, \quad \lambda \geq 0. \end{aligned} \quad (2.13)$$

Končna vrednost λ je maksimalni pogrešek aproksimacije.

Dokaz. Če definiramo

$$\lambda = \max_{\mathbf{z}_i \in \mathbf{Z}} \left| y_i - \sum_{j=1}^m \beta_j(\mathbf{x}_p) \theta_j^T \mathbf{x}_i \right| \quad (2.14)$$

in upoštevamo, da \mathbf{z}_i vsebuje celotno informacijo kombinacije vektorjev \mathbf{x}_{pi} in \mathbf{x}_i , iz tega neposredno sledi sistem neenačb

$$\left| y_i - \sum_{j=1}^m \beta_j(\mathbf{x}_p) \theta_j^T \mathbf{x}_i \right| \leq \lambda, \quad i = 1, \dots, N, \quad (2.15)$$

ki ga lahko zapišemo kot sistem v (2.13). S tem je dokaz leme 2.1 končan in optimizacijski problem (2.12) lahko predstavimo kot minimizacijo λ glede na (2.13). \square

Idejo aproksimacije lahko predstavimo kot najbolj reprezentativno mehko funkcijo, ki opiše domeno izhodov \mathbf{Y} kot funkcijo vhodov \mathbf{Z} . Druga zelo uporabna lastnost pa je zmanjšanje obsega podatkov, kar pogosto nastopi pri identifikacijskih problemih z obsežnim naborom identifikacijskih podatkov.

2.3 Identifikacija intervalnega mehkega modela

Obravnavali bomo nelinearno funkcijo z negotovimi parametri, ki jo lahko predstavimo kot članico družine funkcij

$$\mathcal{G} = \{g : \mathbf{S} \rightarrow \mathbb{R}^1 | g(\mathbf{z}) = g_{nom}(\mathbf{z}) + \Delta g(\mathbf{z})\}, \quad (2.16)$$

kjer je g_{nom} nominalna funkcija, negotovost Δg pa zadosti $\sup_{\mathbf{z} \in \mathbf{S}} |\Delta g(\mathbf{z})| \leq c$, $c \in \mathbb{R}$. Z drugimi besedami, družina funkcij iz razreda \mathcal{G} je navzgor in navzdol omejena s c .

Predpostavimo, da obstajata natančna zgornja meja \bar{g} in natančna spodnja meja \underline{g} , ki ju za poljuben $\varepsilon > 0$ in vsak \mathbf{z} zapišemo na naslednji način:

$$\bar{g}(\mathbf{z}) \geq \max_{g \in \mathcal{G}} g(\mathbf{z}), \exists g \in \mathcal{G} : \bar{g}(\mathbf{z}) < g(\mathbf{z}) + \varepsilon \quad (2.17)$$

$$\underline{g}(\mathbf{z}) \leq \min_{g \in \mathcal{G}} g(\mathbf{z}), \exists g \in \mathcal{G} : \underline{g}(\mathbf{z}) > g(\mathbf{z}) - \varepsilon \quad (2.18)$$

Ker obravnavamo zvezne funkcije, bi za določitev meja, podanih z enačbama (2.17) in (2.18), potrebovali neskončno veliko podatkov. V našem primeru pa se omejimo na končno množico podatkov, pridobljenih z meritvami v sistemu. Zato predpostavimo množico meritev izhoda $\mathbf{Y} = \{y_1, \dots, y_N\}$ z vhodne domene \mathbf{Z} . Točke y_i predstavimo kot rezultat nelinearne preslikave $y_i = g(\mathbf{z}_i)$, $g \in \mathcal{G}$, $\mathbf{z}_i \in \mathbf{S}$, $i = 1, \dots, N$.

Ideja robustnosti modeliranja negotovega sistema z intervalnim mehkim modelom je v tem, da poiščemo mehki funkciji \underline{f} in \bar{f} , ki aproksimirata natančni meji družine funkcij in tako zadostita pogoju

$$\underline{f}(\mathbf{z}_i) \leq g(\mathbf{z}_i) \leq \bar{f}(\mathbf{z}_i), \quad \forall \mathbf{z}_i \in \mathbf{S}. \quad (2.19)$$

Z drugimi besedami, poljubna funkcija iz razreda \mathcal{G} je vedno znotraj pasu, ki je definiran z zgornjo in s spodnjo mehko funkcijo. Edina zahteva je, da je pas čim ožji glede na dane omejitve.

Zgornjo in spodnjo mehko funkcijo lahko dobimo z rešitvijo naslednjih optimizacijskih problemov:

$$\begin{aligned} \min_{\underline{f}} \max_{\mathbf{z}_i \in \mathbf{Z}} |y_i - \underline{f}(\mathbf{z}_i)|, \quad \text{pri } y_i - \underline{f}(\mathbf{z}_i) \geq 0, \quad \forall i, \\ \min_{\bar{f}} \max_{\mathbf{z}_i \in \mathbf{Z}} |y_i - \bar{f}(\mathbf{z}_i)|, \quad \text{pri } y_i - \bar{f}(\mathbf{z}_i) \leq 0, \quad \forall i. \end{aligned} \quad (2.20)$$

Rešitvi poiščemo z linearnim programiranjem, ker so pri obeh optimizacijskih problemih tako namenski funkciji kot funkciji omejitev linearne funkcije, kot to prikazuje naslednja lema. Najprej definiramo mejni mehki funkciji kot $\underline{f}(\mathbf{z}_i) = \boldsymbol{\beta}^T(\mathbf{x}_p)\underline{\boldsymbol{\Theta}}\mathbf{x}$ in $\bar{f}(\mathbf{z}_i) = \boldsymbol{\beta}^T(\mathbf{x}_p)\bar{\boldsymbol{\Theta}}\mathbf{x}$.

Lema 2.2. *Optimizacijska problema iz enačbe (2.20) lahko rešimo z linearnim programiranjem kot minimizacijo λ_1 in λ_2 pri pogojih*

$$\begin{aligned} y_i - \sum_{j=1}^m \beta_j(\mathbf{x}_p)\underline{\boldsymbol{\theta}}_j^T \mathbf{x}_i \leq \lambda_1, \quad i = 1, 2, \dots, N, \\ -y_i + \sum_{j=1}^m \beta_j(\mathbf{x}_p)\bar{\boldsymbol{\theta}}_j^T \mathbf{x}_i \geq 0, \quad i = 1, 2, \dots, N, \quad \lambda_1 \geq 0. \end{aligned} \quad (2.21)$$

in

$$\begin{aligned} -y_i + \sum_{j=1}^m \beta_j(\mathbf{x}_p) \bar{\boldsymbol{\theta}}_j^T \mathbf{x}_i &\leq \lambda_2, \quad i = 1, 2, \dots, N, \\ y_i - \sum_{j=1}^m \beta_j(\mathbf{x}_p) \underline{\boldsymbol{\theta}}_j^T \mathbf{x}_i &\geq 0, \quad i = 1, 2, \dots, N, \quad \lambda_2 \geq 0. \end{aligned} \quad (2.22)$$

za parametre $\underline{\boldsymbol{\theta}}_j, \bar{\boldsymbol{\theta}}_j, j = 1, \dots, m$. λ_1 in λ_2 sta tako maksimalna pogreška aproksimacije pri obeh funkcijah.

Dokaz. Dokaz lahko neposredno povzamemo iz leme 2.1. □

Model (2.1) lahko zapišemo tudi z intervalnimi parametri v obliki hiperpravokotnikov [127], [2]:

$$\tilde{y} = \sum_{j=1}^m \beta_j(\mathbf{x}_p) \tilde{\boldsymbol{\theta}}_j^T \mathbf{x}_i, \quad \tilde{\boldsymbol{\theta}}_j^T = \left[\underline{\boldsymbol{\theta}}_j \quad \bar{\boldsymbol{\theta}}_j \right]^T. \quad (2.23)$$

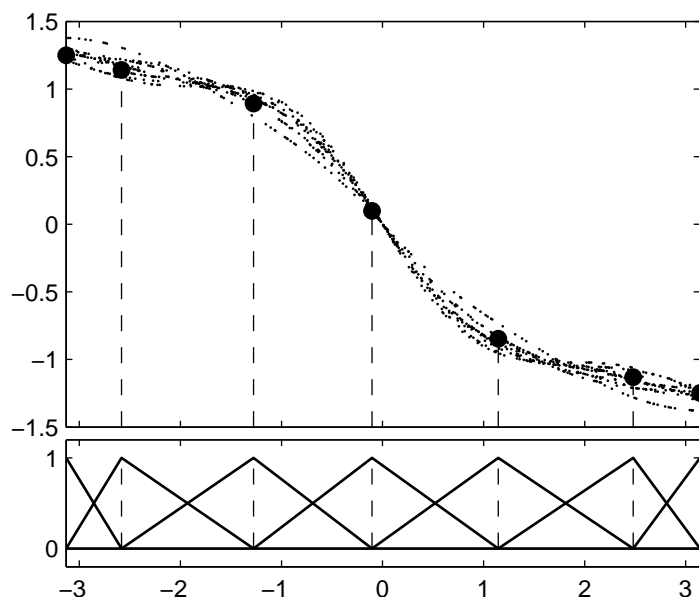
2.3.1 Primer: aproksimacija družine krivulj

V prvem primeru aproksimacije z modelom INFUMO privzemimo družino krivulj \mathcal{G}

$$g(z) = g_0(z) + \Delta g(z) = \arctan(-z) + a_1 \sin(a_2 z), \quad z \in [-\pi, \pi] \quad (2.24)$$

s parametroma $a_1 \in [-0, 15 \ 0, 15]$ in $a_2 \in [1, 75 \ 2, 25]$. Množico podatkov dobimo z desetimi meritvami, pri čemer pri vsaki meritvi parametra a_1 in a_2 zavzameta poljubni vrednosti znotraj danih intervalov. Množica točk je prikazana na zgornjem diagramu slike 2.1.

Najprej uporabimo izhodne funkcije ϕ modela (2.4) tipa Singleton ($\phi_j(\mathbf{x} = \theta_{j0})$). V tem primeru posledični vektor \mathbf{x} ni vhodna spremenljivka mehkega modela, oblika izhoda modela pa je odvisna od oblik pripadnostnih funkcij. Zaradi enostavnosti izberimo trikotne pripadnostne funkcije. Najprej z algoritmom *c-means* poiščemo središča n_c rojev. Izberemo vrednost $n_c = 5$ in razporedimo pripadnostne funkcije, kot to kaže spodnji diagram slike 2.1. V splošnem lahko v naslednjem koraku izberemo poljubne začetne parametre $\underline{\boldsymbol{\Theta}}_0$ in $\bar{\boldsymbol{\Theta}}_0$ ter pristopimo k optimizaciji meja po enačbi (2.20). Po drugi strani pa pri veliki količini podatkov in velikem številu parametrov reševanje optimizacijskega problema zahteva veliko časa. Zato začetne parametre izračunamo tako, da so čim bližje dejanskim parametrom. Najprej za dano porazdelitev pripadnostnih funkcij določimo krivuljo, ki se najbolj prilega dani družini krivulj v smislu najmanjših kvadratov. Parametre izračunamo po metodi najmanjših kvadratov, kot je bilo opisano



Slika 2.1: Množica podatkov in centri rojev s pripadnostnimi funkcijami

v [131]. Nato izvajamo vzporedni premik funkcije proti zgornji meji pasu - parametre spreminjamo toliko časa, dokler vse točke krivulje ne ležijo nad točkami pasu. Enako storimo tudi za spodnjo mejo in na ta način pridobimo začetni vrednosti parametrov intervalnega modela. Optimalne vrednosti dobimo z rešitvijo problema (2.20). V danem primeru so bili položaji centrov rojev (zaradi razširitve aproksimacije na celotno vhodno domeno smo dodali pripadnostni funkciji za krajni vrednosti -3,132 in 3,141) pri

$$\alpha_c = \left[-3,132 \quad -2,584 \quad -1,277 \quad -0,1044 \quad 1,144 \quad 2,48 \quad 3,141 \right], \quad (2.25)$$

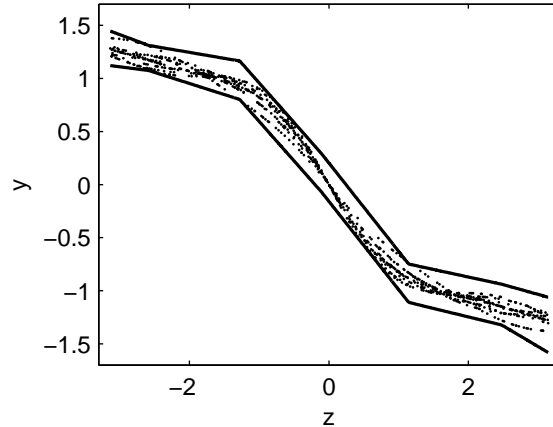
začetne vrednosti intervalnih parametrov

$$\begin{aligned} \bar{\Theta}_0 &= \left[1,491 \quad 1,382 \quad 1,134 \quad 0,340 \quad -0,605 \quad -0,890 \quad -1,007 \right]^T, \\ \underline{\Theta}_0 &= \left[1,008 \quad 0,899 \quad 0,652 \quad -0,142 \quad -1,087 \quad -1,372 \quad -1,490 \right]^T \end{aligned} \quad (2.26)$$

ter končne vrednosti intervalnih parametrov

$$\begin{aligned} \bar{\Theta} &= \left[1,445 \quad 1,311 \quad 1,164 \quad 0,290 \quad -0,749 \quad -0,939 \quad -1,061 \right]^T, \\ \underline{\Theta} &= \left[1,121 \quad 1,075 \quad 0,802 \quad -0,071 \quad -1,110 \quad -1,323 \quad -1,581 \right]^T. \end{aligned} \quad (2.27)$$

Minimalni vrednosti maksimalnih pogreškov sta $\lambda_1 = 0,3609$ in $\lambda_2 = 0,3618$. Slika 2.2 prikazuje meji družine krivulj, podani z dobljenim modelom INFUMO. INFUMO



Slika 2.2: Aproximacija družine krivulj z modelom INFUMO, tip Singleton

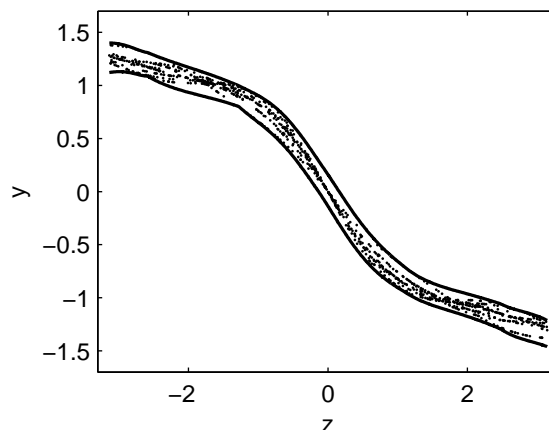
z izhodnimi funkcijami tipa Singleton daje aproksimacijo meja v obliki odsekoma zvezne linearne funkcije, ki je funkcionalno enakovredna intervalu funkcij CPWL (ang. canonical piecewise-linear functions, kanonične odsekoma zvezne funkcije) [60]. Boljše prilaganje funkcij podatkom pa lahko dosežemo, če uporabimo bolj gladke izhodne funkcije. Za primer vzemimo polinome tretje stopnje

$$\phi_j(\mathbf{x}) = \boldsymbol{\theta}_j^T \mathbf{x} = \theta_{j0} + \theta_{j1}x_1 + \theta_{j2}x_2 + \theta_{j3}x_3, \quad (2.28)$$

kjer so parametri posledičnega vektorja $\mathbf{x} = [1, z, z^2, z^3]$. Oblika in porazdelitev pripadnostnih funkcij naj bosta enaki kot v prvem primeru. Po istem postopku kot prej najprej določimo začetne parametre, premaknemo funkciji izven meja in z optimizacijo poiščemo končne parametre. Optimalne vrednosti so

$$\begin{aligned} \overline{\boldsymbol{\Theta}} &= \begin{bmatrix} -0,972 & -0,113 & -0,090 & 0,242 & -0,0722 & 0,025 & 0,126 \\ -8,804 & -0,757 & -0,559 & -0,014 & 0,482 & -0,183 & -1,402 \\ -26,66 & -1,819 & -1,228 & -0,977 & -1,257 & 0,233 & 4,766 \\ -25,60 & -0,273 & 0,153 & 0,159 & 0,181 & -0,886 & -6,254 \end{bmatrix}^T \\ \underline{\boldsymbol{\Theta}} &= \begin{bmatrix} -0,977 & -0,115 & -0,081 & 0,299 & -0,089 & 0,022 & 0,124 \\ -8,773 & -0,729 & -0,552 & -0,011 & 0,523 & -0,175 & -1,390 \\ -26,53 & -1,827 & -1,368 & -1,001 & -1,170 & 0,236 & 4,851 \\ -25,92 & -0,745 & -0,215 & -0,133 & -0,176 & -1,148 & -6,836 \end{bmatrix}^T. \end{aligned} \quad (2.29)$$

Slika 2.3 prikazuje meji družine krivulj v primeru uporabe funkcij tretjega reda. Minimalni vrednosti maksimalnih pogreškov sta $\lambda_1 = 0,2628$ in $\lambda_2 = 0,2584$, kar pomeni 37- in 40-odstotno izboljšanje aproksimacije.

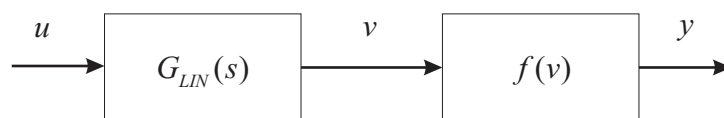


Slika 2.3: Aproximacija družine krivulj z modelom INFUMO, izhodne funkcije 3. reda

2.4 Model Wiener-INFUMO

Wienerjev model uporabljamo predvsem za opis nelinearnosti v procesu, ne da bi pri tem morali uvesti splošne nelinearne operatorje in s tem prispevati k povečanju kompleksnosti obravnavanega problema. Z njim lahko zadovoljivo opisujemo nelinearnosti, ki jih pogosto srečamo pri industrijskih procesih, kot sta destilacija in nevtralizacija pH. Wienerjev model sestoji iz linearnega dinamičnega elementa (L), kateremu v kaskadi sledi nelinearna statična izhodna funkcija (N), medtem ko sta pri sorodnem Hammersteinovem modelu elementa v obratnem vrstnem redu [12], [86]. V originalni izvedbi je linearni del zapisan s prenosnimi funkcijami na podlagi Laguerrovih baznih funkcij, nelinearni element pa s Hermitovimi polinomi. Obstaja pa tudi precej izvedenk, ki praviloma uporabljajo modificirani linearni del [3], na primer Kautzove vrste [121].

Slika 2.4 prikazuje shemo, kjer $G_{LIN}(s)$ označuje prenosno funkcijo linearne dinamike v obliki Laplaceovega transformata, $f(v)$ pa nelinearno preslikavo. Elementa



Slika 2.4: Shema Wienerjevega modela

sta povezana preko vmesne spremenljivke v , ki pa v realnem procesu ni nujno merljiva. Dinamični del ima po definiciji statično ojačenje enako ena. V fazi modeliranja ali identifikacije lahko to v splošnem dosežemo tudi z vzporednim premikom izhodne statične funkcije. Statično karakteristiko lahko opišemo na več načinov, na primer

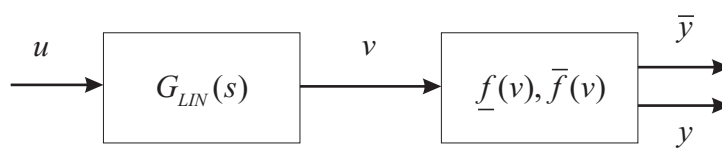
z odsekoma zveznimi linearnimi funkcijami, polinomi, zlepki (ang. spline), mehкими statičnimi funkcijami itd.

Metode identifikacije Wienerjevih modelov, ki jih lahko zasledimo v literaturi, lahko v grobem razdelimo v tri razrede [41]:

- pristop $N-L$, kjer najprej določimo statično karakteristiko iz vhodno-izhodnih podatkov v ustaljenem stanju, nato s preslikavo izhodov preko inverzne statične preslikave rekonstruiramo vrednosti vmesne spremenljivke, dinamični del pa dobimo s pomočjo identifikacije iz vrednosti vhoda in vmesne spremenljivke;
- pristop $L-N$, kjer najprej identificiramo linearni del s korelacijsko tehniko, nato izračunamo vrednosti vmesne spremenljivke in s primerjanjem z vrednostmi izhoda ocenimo statično nelinearnost [133];
- simultani pristop, kjer parametre linearnega in nelinearnega elementa ocenimo istočasno [63],[116].

Najpogostejši in najenostavnejši pristop je $N-L$. Zagotavlja natančen opis statične nelinearnosti in je primeren predvsem za odprtozračno stabilne procese z invertibilno statično karakteristiko. Šibka točka pristopa pa je obravnavanje sistemov z nedoločenimi parametri in motnjami na izhodu procesa. V tem primeru s pomočjo identifikacije ne moremo dobiti nepristranske ocene parametrov.

Z modifikacijo, kot je prikazano na sliki 2.5, Wienerjev model uporabimo za robustno modeliranje takih sistemov v smislu določanja zgornje in spodnje meje pasu odzivov. Opis dinamičnega dela ostaja enak tistemu s slike 2.4, statični del pa je zdaj



Slika 2.5: Shema modela Wiener-INFUMO

v obliki statičnega modela INFUMO, obravnavanega v razdelku 2.3.1. Za identifikacijo uporabimo modificirano strategijo $N-L$:

- na podlagi vhodno-izhodnih meritev identificiramo statični model INFUMO;
- pripadnostne funkcije modela INFUMO preslikamo preko dobljenih statičnih karakteristik in identificiramo parametre zgornje in spodnje inverzne funkcije, s pomočjo katerih rekonstruiramo oceno pasu vmesne spremenljivke;

- iz povprečenja karakteristik z metodo ARX identificiramo parametre linearnega elementa.

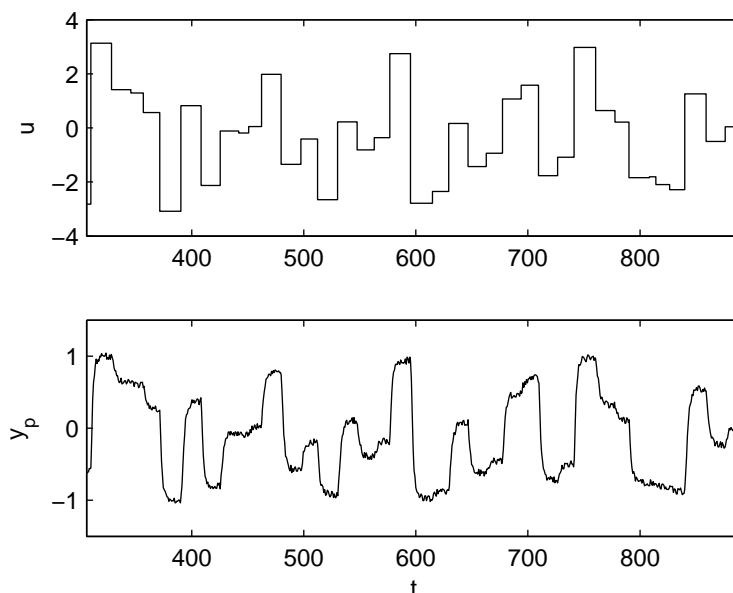
Podrobnosti si bomo ogledali na ilustrativnem primeru.

2.4.1 Primer: modeliranje nelinearnega dinamičnega sistema z modelom Wiener-INFUMO

Denimo, da je obravnavani proces

$$\begin{aligned} v(t) &= \frac{1}{p^2 + \tilde{a}_1 p + \tilde{a}_2} u(t), \\ y(t) &= \arctan(v(t)) + \nu(t), \end{aligned} \quad (2.30)$$

kjer p označuje operator odvajanja, $\tilde{a}_1 \in [4, 75 \ 5, 25]$ in $\tilde{a}_2 \in [1, 9 \ 2, 1]$ sta intervalna parametra, $\nu(t)$ pa je šum meritev izhoda, omejen z $|\nu(t)| < 0.05$. Signale vzorčimo s časom $T_s = 0,2$ s. S poskusi pridobimo vhodno-izhodne podatke, kjer za vhodni signal uporabimo stopničasti signal, kot to kaže slika 2.6. Na sliki je predstavljen samo del meritev; v splošnem lahko opravimo več serij meritev z istim vhodnim signalom ali pa s poljubnimi serijami, edini pogoj je, da zajamemo celotno delovno področje. Iz



Slika 2.6: Vhodno-izhodne meritve za identifikacijo

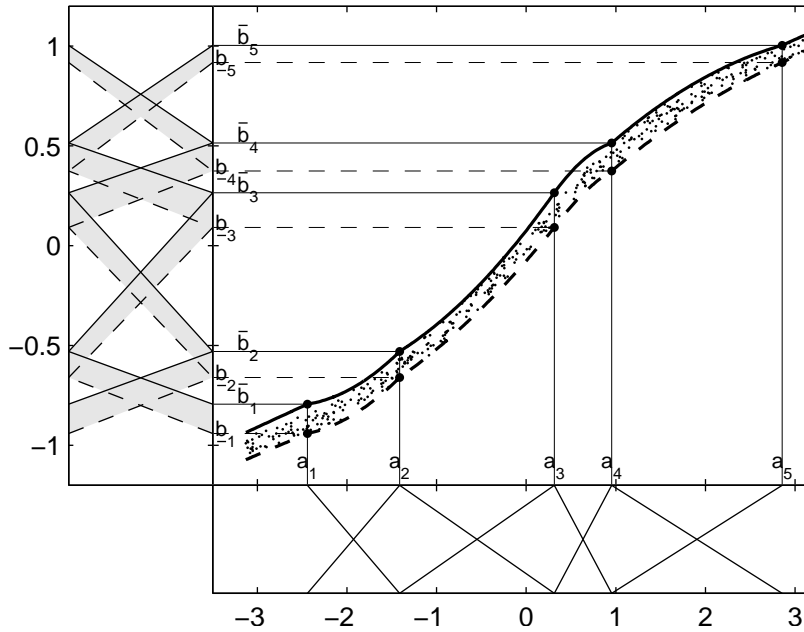
podatkov pridobimo množico točk (u_{si}, y_{si}) v ustaljenem stanju. Prikazana je na sliki

2.7. Množico omejimo z modelom INFUMO

$$\mathbf{R}_j : \text{če } u_s \text{ je } \mathbf{A}_j, \text{ potem } \bar{y}_s = \bar{\theta}_{j,1}u_s + \bar{\theta}_{j,0}, \quad (2.31)$$

$$\underline{y}_s = \underline{\theta}_{j,1}u_s + \underline{\theta}_{j,0}, \quad j = 1, \dots, 5.$$

Z optimizacijskim postopkom, opisanim v razdelku 2.3, izračunamo porazdelitev pripadnostnih funkcij A_j (na sliki 2.7 vidimo položaje vrhov $\mathbf{a} = [a_1, a_2, \dots, a_5]$) in parametre $\{\bar{\theta}_j, \underline{\theta}_j\}$ za izhodni mejni funkciji. Izračunane vrednosti so



Slika 2.7: Aproximacija ter preslikava pripadnostnih funkcij za inverzni model INFUMO

$$\mathbf{a} = \begin{bmatrix} -2,442 & -1,412 & 0,313 & 0,954 & 2,857 \end{bmatrix} \quad (2.32)$$

$$\underline{\Theta} = \begin{bmatrix} \underline{\theta}_1^T \\ \underline{\theta}_2^T \\ \underline{\theta}_3^T \\ \underline{\theta}_4^T \\ \underline{\theta}_5^T \end{bmatrix} = \begin{bmatrix} 0,1909 & -0,4748 \\ 0,3681 & -0,1405 \\ 0,4880 & -0,0621 \\ 0,3170 & 0,0721 \\ 0,2307 & 0,2584 \end{bmatrix}, \quad \bar{\Theta} = \begin{bmatrix} \bar{\theta}_1^T \\ \bar{\theta}_2^T \\ \bar{\theta}_3^T \\ \bar{\theta}_4^T \\ \bar{\theta}_5^T \end{bmatrix} = \begin{bmatrix} 0,2069 & -0,2886 \\ 0,3931 & 0,0253 \\ 0,5702 & 0,0866 \\ 0,3317 & 0,1979 \\ 0,2021 & 0,4264 \end{bmatrix}. \quad (2.33)$$

S tem je nelinearna statična preslikava identificirana, ostane nam še identifikacija linearnega dela G_{LIN} , ki opisuje relacijo med vhom u in vmesno spremenljivko v . Ker je statično ojačenje linearnega dela enako ena, je preslikava iz v_s v y_s enakovredna preslikavi iz u_s v y_s . Zato vrednosti v dobimo z inverzno preslikavo izhoda y preko statične

karakteristike. To naredimo tako, da najprej transformiramo pripadnostne funkcije množic \mathbf{A}_j v pripadnostne funkcije množic $\overline{\mathbf{B}}_j^{inv}$ in $\underline{\mathbf{B}}_j^{inv}$ (glej sliko 2.7). Izhodišča in vrhove pripadnostnih funkcij $\mu_{\mathbf{A}_j}$ preslikamo preko zgornje in spodnje mehke funkcije v pripadajoče vrhove in izhodišča pripadnostnih funkcij $\mu_{\overline{\mathbf{B}}_j^{inv}}$ in $\mu_{\underline{\mathbf{B}}_j^{inv}}$ ter jih povežemo s premicami, tako da spet dobimo trikotno obliko pripadnostnih funkcij. Parametre posledičnega dela preslikave $y \rightarrow v$ dobimo z rešitvijo linearnega problema

$$\begin{aligned} \overline{\mathbf{R}}_j^{inv} : \text{če } y(k) \text{ je } \overline{\mathbf{B}}_j, \text{ potem } \bar{v}(k) &= \bar{\theta}_{j,1}^{inv} y(k) + \bar{\theta}_{j,0}^{inv}, \\ \underline{v}(k) &= \underline{\theta}_{j,1}^{inv} y(k) + \underline{\theta}_{j,0}^{inv}, \quad j = 1, \dots, 5, \end{aligned} \quad (2.34)$$

kjer za identifikacijsko množico uporabimo točke $\{(y_s, u_s)\}$. Dobljene vrednosti so

$$\underline{\Theta}^{inv} = \begin{bmatrix} 4.313 & 0.723 \\ 2.437 & -0.071 \\ 2.954 & -0.073 \\ 2.997 & -0.472 \\ 4.468 & -1.752 \end{bmatrix}, \quad \overline{\Theta}^{inv} = \begin{bmatrix} 4.089 & 1.457 \\ 2.977 & 0.499 \\ 2.905 & 0.054 \\ 2.667 & 0.013 \\ 4.514 & -1.225 \end{bmatrix}. \quad (2.35)$$

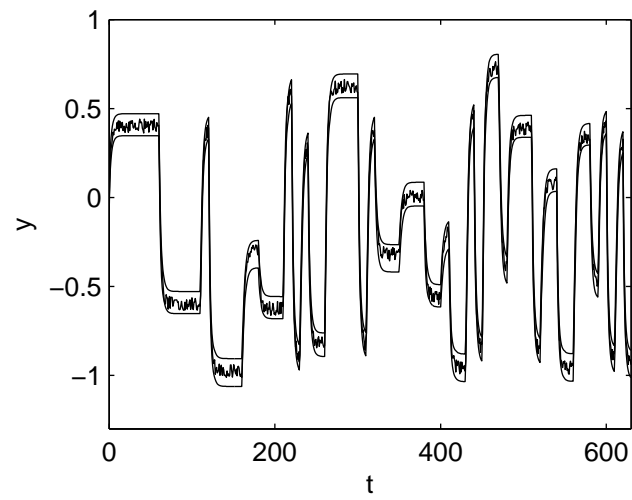
Ko na vhod inverznega modela pripeljemo vrednosti $y(k)$, dobimo vrednosti ocene zgornje in spodnje meje vmesne spremenljivke, \bar{v} in \underline{v} . Potem izračunamo srednjo vrednost intervala kot $\tilde{v} = (\bar{v} + \underline{v})/2$. Prenosno funkcijo linearnega dela $G_{LIN}(q^{-1})$ izračunamo iz množice $\{(u(k), \tilde{v}(k))\}$ po identifikacijski metodi ARX [76] in dobimo

$$G_{LIN}(q^{-1}) = \frac{0.1092q^{-1}}{1 - 0.4631q^{-1} + 0.4293q^{-2}}. \quad (2.36)$$

Celotni model Wiener-INFUMO predstavimo z naslednjo enačbo:

$$\begin{aligned} v(k) &= 0.4631v(k-1) - 0.4293v(k-2) + 0.1092u(k-1), \\ \bar{y}(k) &= \beta^T(v(k))\overline{\Theta}\mathbf{x}(v(k)), \\ \underline{y}(k) &= \beta^T(v(k))\underline{\Theta}\mathbf{x}(v(k)). \end{aligned} \quad (2.37)$$

Slika 2.8 prikazuje rezultat validacije identificiranega modela. Razvidno je, da je omejitvev ustrezna, ker se vse točke verifikacijskega izhodnega signala nahajajo znotraj območja, omejenega z mejnima funkcijama modela.



Slika 2.8: Verifikacija identificiranega modela Wiener-INFUMO

3. Napredne metode zaznavanja napak

V zadnjih desetletjih konkurenca na trgih močno povečuje zahteve po končni kakovosti proizvodov. Prilagajanje zahtevam trga narekuje opuščanje ročnega dela v proizvodnji in posledično vse večjo avtomatizacijo proizvodnih sistemov. Vendar pa tudi sama avtomatizacija proizvodnega postopka še ni zagotovilo za brezhibnost končnega izdelka. Proizvodna oprema, merilni in izvršni členi so podvrženi številnim vplivom, ki zmanjšujejo njihovo funkcionalnost. Po splošnem konsenzu Tehničnega komiteja SAFEPROCESS [57] nezadovoljivo delovanje delov sistema delimo na

- **napake** (ang. faults) ali nedovoljene odmike vsaj ene karakteristične vrednosti dela sistema od standardne ali dovoljene vrednosti; napake predstavljajo poslabšano funkcionalnost opreme, celoten sistem pa zaradi njih navadno ne odpove;
- **odpovedi** (ang. failures) ali trajne odpovedi dela sistema, da opravi svojo funkcijo pod predpisanimi pogoji; odpovedi pomenijo porušeno funkcionalnost opreme in nemalokrat neposredno ogrozijo delovanje celotnega sistema.

Učinki napak in odpovedi so lahko povezani z velikanskimi proizvodnimi stroški, navadno zaradi prekinitve proizvodnje (npr. valjarna, če pride do okvare pogonskega dela valjev) ali zmanjšanja kvalitete serije izdelkov (npr. v proizvodnji zdravil lahko zaradi napake senzorja v enem od kemijskih reaktorjev pride do izmeta velikanskih serij). Po drugi strani pa odpovedi lahko pomenijo veliko nevarnost za ljudi in okolje. Kot primer navedimo okvare na izvršnih členih v letalu, ki lahko posledično privedejo do strmoglavljenja, in napake v delovanju podsklopov jedrskega reaktorja, ki lahko privedejo do odpovedi ključnih delov in eksplozije reaktorja.

V industriji se odkrivanje neželenih (nevarnih) stanj izvaja s pomočjo nadzornega sistema, ki je navadno del celotnega sistema vodenja. Ločimo tri glavne oblike nadzor-
nih sistemov:

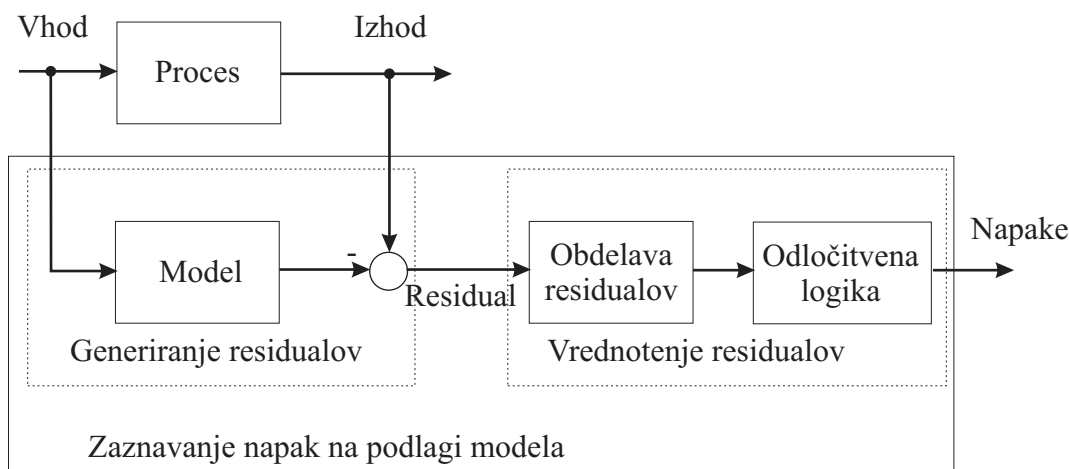
- **kontrola procesnih veličin** (ang. variable monitoring) - merljive veličine testiramo glede na podane tolerance, glede na kršenje tolerančnih meja pa generiramo alarme;

- **avtomatska zaščita** (ang. automatic protection) - v primeru nevarnih procesnih stanj, ki jih odkrivamo s kontrolo procesnih veličin, se avtomatsko začnejo predvideni ukrepi (sprememba vhodov, rekonfiguracija sistema, zaustavitev sistema idr.);
- **avtomatski nadzor z diagnostiko napak** (ang. supervision with fault diagnosis) - na podlagi meritev v procesu nadzorni sistem poda karakteristične veličine, izračuna simptome, diagnosticira napake in poda odločitve za ustrezne ukrepe.

Prvi dve obliki štejemo med *klasične metode*, tretjo pa med *napredne metode* zaznavanja napak. V tem delu bomo podali pregled in kratek opis nekaterih najbolj znanih in/ali uporabljanih naprednih metod zaznavanja in lokalizacije napak na podlagi modela procesa.

3.1 Osnovna strategija zaznavanja napak na podlagi modela

Strategijo zaznavanja napak na podlagi modela procesa predstavlja slika 3.1. Model



Slika 3.1: Zaznavanje napak na podlagi modela procesa

procesa služi za kvalitativen ali kvantitativen opis dinamičnega in statičnega obnašanja procesa v normalnem načinu delovanja (brez prisotnosti napak). Če ga vzбудimo z istim vhodom kot proces, dobimo oceno merjenih procesnih spremenljivk. S primerjavo ustreznih merjenih in ocenjenih vrednosti izračunamo ti. *residual*. Navadno ga označujemo s spremenljivko $r(t)$. Označimo signal napake $f(t)$. V idealnih pogojih, ko niso prisotni neznani vhodi v sistem in lahko proces natančno modeliramo, je logika

zaznavanja napak enostavna:

$$\begin{aligned} r(t) \neq 0 &\rightarrow f(t) = 1 \\ r(t) = 0 &\rightarrow f(t) = 0. \end{aligned} \quad (3.1)$$

Vendar pa v realni situaciji ne moremo natančno modelirati procesa, tako da residual vedno vsebuje tudi vplive negotovosti modela in neznanih vhodov v proces. Da bi ločili te vplive od vplivov napak, moramo uvesti dodatno procesiranje residualov.

Metode zaznavanja in lokalizacije napak na podlagi modela procesa klasificiramo na podlagi uporabljenega modela oziroma načina oblikovanja residualov in metode procesiranja residualov. Na podlagi tega bomo metode, opisane v disertaciji, razdelili v tri razrede:

- metode na podlagi kvantitativnega modela;
- metode na podlagi kvalitativnega modela;
- metode na podlagi zgodovine procesa.

3.2 Zaznavanje napak na podlagi kvantitativnega modela

3.2.1 Ocenjevanje parametrov

Uporabo ocenjevanja parametrov v zaznavanju napak je predstavil Isermann s sodelavci [54]. Model procesa zapišemo kot

$$(1 + a_1 z^{-1} + \dots + a_n z^{-n})y(k) = (b_1 z^{-1} + \dots + b_n z^{-n})u(k) \quad (3.2)$$

ali v regresijski obliki

$$y(k) = \boldsymbol{\psi}^T(k)\boldsymbol{\theta}, \quad (3.3)$$

$$\begin{aligned} \boldsymbol{\psi}^T &= [-y(k-1), \dots, -y(k-n), u(k-1), \dots, u(k-m)], \\ \boldsymbol{\theta} &= [a_1, \dots, a_n, b_1, \dots, b_m]^T \end{aligned} \quad (3.4)$$

kjer je informacija o napaki vsebovana v vektorju $\boldsymbol{\theta}$. Oceno parametrov izračunamo po metodi najmanjših kvadratov v rekurzivni obliki

$$\begin{aligned} \hat{\boldsymbol{\theta}}(k+1) &= \hat{\boldsymbol{\theta}}(k) + \gamma(k) \left[y(k+1) - \boldsymbol{\psi}^T(k+1)\hat{\boldsymbol{\theta}}(k) \right], \\ \gamma(k) &= \frac{1}{1 + \boldsymbol{\psi}^T(k+1)\mathbf{P}(k)\boldsymbol{\psi}(k+1)} \mathbf{P}(k)\boldsymbol{\psi}^T(k+1) \\ \mathbf{P}(k+1) &= [\mathbf{I} - \gamma(k)\boldsymbol{\psi}^T(k+1)] \mathbf{P}(k). \end{aligned} \quad (3.5)$$

V splošnem so procesni parametri neposredno povezani s fizikalnimi parametri p (npr. togost, faktor dušenja, upornost...) preko nelinearnih algebrajskih enačb

$$\boldsymbol{\theta} = f(p). \quad (3.6)$$

Napaka velikokrat povzroči spremembo parametrov procesa, npr. zamašitev cevi spremeni koeficient prevodnosti ventila. Če obstaja inverzna relacija

$$p = f^{-1}(\boldsymbol{\theta}), \quad (3.7)$$

lahko s pomočjo ocene parametrov izračunamo spremembe fizikalnih parametrov v procesu. Te spremembe pa so večinoma neposredno povezane z napakami v procesu. Ocenjevanje parametrov tako omogoča diagnostiko napak, ker lahko napake ločimo med seboj po tem, na katerem parametru ugotovimo odstopanje od nominalne vrednosti. Po drugi strani pa za samo zaznavanje napak poznavanje fizikalnih parametrov ni nujno, dovolj je poznati samo vrednosti parametrov procesa v normalnem načinu delovanja.

Slabosti omenjenega pristopa sta predvsem dve:

- relacija f med procesnimi in fizikalnimi parametri velikokrat ni neposredna, njen inverz pa lahko v nekaterih primerih obstaja samo za modele nizkega reda;
- sprotno ocenjevanje parametrov potrebuje stalno vzbujanje procesa, kar lahko povzroči težave pri delovanju v stacionarnih delovnih točkah.

3.2.2 Pristopi na podlagi observatorja stanj

Filter za zaznavanje napak

Za začetek metod na podlagi observatorja štejem pristop s **filtrom za zaznavanje napak** (ang. fault-detection filter), ki sta ga opisala Beard in Jones v letih 1971 in 1973 [32]. Model procesa zapišemo v obliki

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{E}_f\mathbf{f}(t) + \mathbf{E}_d\mathbf{d}(t) \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{F}_f\mathbf{f}(t) + \mathbf{F}_d\mathbf{d}(t), \end{aligned} \quad (3.8)$$

kjer je \mathbf{f} vektor napak, \mathbf{d} vektor neznanih vhodov v sistem, ki pomeni motnje ali pa strukturirane negotovosti modela, \mathbf{E}_f , \mathbf{E}_d , \mathbf{F}_f in \mathbf{F}_d pa ustrezne prenosne funkcije. Ključni element pristopa je observator polnega ranga

$$\frac{d\hat{\mathbf{x}}}{dt}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{L}(\mathbf{y}(t) - \mathbf{C}\hat{\mathbf{x}}(t)), \quad (3.9)$$

s pomočjo katerega oblikujemo residual

$$\mathbf{r}(t) = \mathbf{V}(\mathbf{y}(t) - \mathbf{C}\hat{\mathbf{x}}(t)). \quad (3.10)$$

Dinamika residuala je podana z enačbama

$$\begin{aligned} \dot{\mathbf{e}}(t) &= (\mathbf{A} - \mathbf{LC})\mathbf{e}(t) + (\mathbf{E}_f - \mathbf{LF}_f)\mathbf{f}(t) + (\mathbf{E}_d - \mathbf{LF}_d)\mathbf{d}(t) \\ \mathbf{r}(t) &= \mathbf{VCe}(t) + \mathbf{VF}_f\mathbf{f}(t) + \mathbf{VF}_d\mathbf{d}(t). \end{aligned} \quad (3.11)$$

Načrtovalska parametra sta matrika ojačenja \mathbf{L} in matrika \mathbf{V} , s katero oblikujemo residual.

Diagnostični observator

Imejmo sistem iz enačbe (3.8). Diagnostični observator (ang. diagnostic observer, DO) je generator residuala, ki ga podaja enačba

$$\begin{aligned} \dot{\mathbf{z}}(t) &= \mathbf{Gz}(t) + \mathbf{Hu}(t) + \mathbf{Ly}(t), \\ \mathbf{r}(t) &= \mathbf{Wz}(t) + \mathbf{Vy}(t) + \mathbf{Qu}(t), \end{aligned} \quad (3.12)$$

kjer je $\mathbf{z}(t) \in \mathbb{R}^s$ vektor dimenzije s , ki pomeni red observatorja, matrike \mathbf{G} , \mathbf{H} , \mathbf{L} , \mathbf{Q} , \mathbf{V} in \mathbf{W} so ustreznih dimenzij in skupaj z matriko $\mathbf{T} \in \mathbb{R}^{s \times n}$ zadostijo ti. Luenbergerjevimi pogojem:

$$\begin{aligned} \mathbf{TA} - \mathbf{GT} &= \mathbf{LC}, & \mathbf{H} &= \mathbf{TB} - \mathbf{LD}, \\ \mathbf{VC} - \mathbf{WT} &= \mathbf{0}, & \mathbf{Q} &= -\mathbf{VD}. \end{aligned} \quad (3.13)$$

Dinamiko generatorja residuala opišemo z

$$\begin{aligned} \dot{\mathbf{e}}(t) &= \mathbf{Ge}(t) + (\mathbf{TE}_f - \mathbf{LF}_f)\mathbf{f}(t) + (\mathbf{TE}_d - \mathbf{LF}_d)\mathbf{d}(t), \\ \mathbf{r}(t) &= \mathbf{We}(t) + \mathbf{VF}_f\mathbf{f}(t) + \mathbf{VF}_d\mathbf{d}(t). \end{aligned} \quad (3.14)$$

Iz primerjave s filtrom za zaznavanje napak (3.11) je očitno, da lahko pri pristopu DO uporabimo reducirani red observatorja, kar je zaželeno in zelo uporabno pri praktičnih izvedbah. Načrtovalski parametri so matrike \mathbf{G} , \mathbf{L} , \mathbf{T} , \mathbf{V} in \mathbf{W} , kar nam daje veliko prostosti pri načrtovanju, po drugi strani pa naredi postopek načrtovanja bolj zahteven in zapleten.

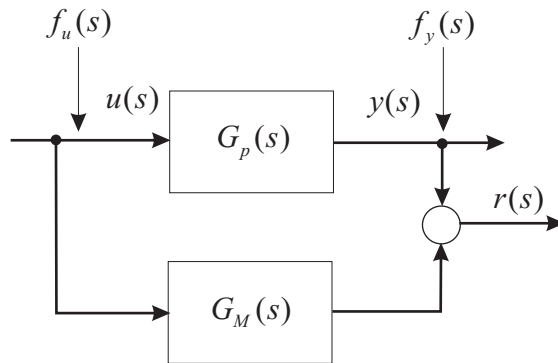
3.2.3 Pristopi na podlagi paritetnih enačb

Paritetne enačbe so preurejene in transformirane izvedenke zapisov modelov v prostoru stanj ali vhodno-izhodni obliki [44]. Bistvo pristopa je preverjanje paritete ali konsistentnosti (ang. parity, consistency) modelov procesov z merjenimi izhodi in znanimi

procesnimi vhodi. Dinamične paritetne relacije so bile prvič predstavljene v delu [134]. Osnovno in najenostavnejšo izvedbo [56] prikazuje slika 3.2. Model $G_M(s)$ vežemo vzporedno s procesom $G_P(s)$ in računamo izhodni pogrešek po enačbi

$$r(s) = (G_P(s) - G_M(s)) u(s). \quad (3.15)$$

Če je proces znan, z izenačenjem modela in procesa ter upoštevanjem $y(s) =$



Slika 3.2: Paritetne enačbe na podlagi izhodnega pogreška

$G_P(s) (u(s) + f_u(s)) + f_y(s)$ dobimo paritetno enačbo

$$r(s) = G_P(s)f_u(s) + f_y(s). \quad (3.16)$$

Časovni potek residuala $r(s)$ se razlikuje za napake na vhodu ali izhodu procesa. Z dodatnim filtriranjem

$$r_f(s) = G_F(s)r(s) \quad (3.17)$$

lahko oblikujemo residual, tako da dosežemo želene lastnosti zaznavanja različnih napak. Slabost tega pristopa je, da lahko za procese SISO generiramo samo en residual in je zato ločevanje med napakami relativno zahtevno.

Pri drugem pristopu pa izhajamo iz zapisa modela v prostoru stanj, obravnavamo pa lahko sisteme SISO in MIMO. Predpostavljamo, da so odvodi vhodov in izhodov merljivi ali pa jih dobimo s pomočjo filtriranja [58]. Imejmo sistem MIMO

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \quad (3.18)$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t). \quad (3.19)$$

Če enačbo (3.19) p -krat odvajamo po času in v odvode vstavljamo (3.18), sistem zapišemo v strnjeni matrični obliki

$$\begin{bmatrix} \mathbf{y}(t) \\ \dot{\mathbf{y}}(t) \\ \ddot{\mathbf{y}}(t) \\ \vdots \\ \mathbf{y}^{[p]}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA}^2 \\ \vdots \\ \mathbf{CA}^p \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 & \cdots & \cdots & 0 \\ \mathbf{CB} & 0 & \cdots & 0 \\ \mathbf{CAB} & \mathbf{CB} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{CA}^{p-1}\mathbf{B} & \mathbf{CA}^{p-2}\mathbf{B} & \cdots & \mathbf{CB} \end{bmatrix} \begin{bmatrix} \mathbf{u}(t) \\ \dot{\mathbf{u}}(t) \\ \vdots \\ \mathbf{u}^{[p]}(t) \end{bmatrix} \quad (3.20)$$

$$\mathbf{Y}_F(t) = \mathbf{T}\mathbf{x}(t) + \mathbf{Q}\mathbf{U}_F(t). \quad (3.21)$$

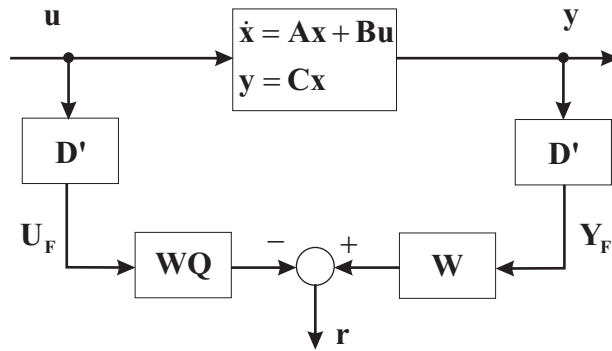
Če hočemo odstraniti vpliv nemerljivih stanj $\mathbf{x}(t)$ na residual, pomnožimo (3.21) s projekcijsko matriko \mathbf{W} , ki jo izberemo tako, da velja

$$\mathbf{W}\mathbf{T} = \mathbf{0}. \quad (3.22)$$

Vrstice matrike \mathbf{W} morajo biti ortogonalne. Tako dobimo

$$\mathbf{r}(t) = \mathbf{W}\mathbf{Y}_F(t) - \mathbf{W}\mathbf{Q}\mathbf{U}_F(t). \quad (3.23)$$

Dobljena shema je prikazana na sliki 3.3. Šibka točka tega pristopa je računanje odvo-



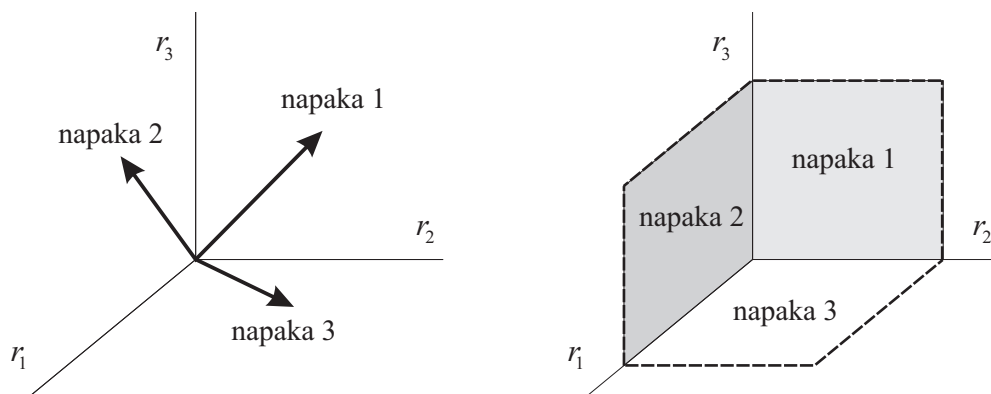
Slika 3.3: Metoda paritetnih enačb za model MIMO v prostoru stanj

dov vhodnega in izhodnega signala, ki lahko vnaša precejšnje motnje v residual.

Z ustreznim načrtovanjem matrike \mathbf{W} lahko oblikujemo množico razširjenih residualov tako, da so

- diagonalno občutljivi - i -ta napaka odkloni samo i -ti residual;
- usmerjeno občutljivi na napake (slika 3.4 levo), kjer i -ta napaka povzroči odklon residuala v smeri danega vektorja v prostoru residualov;

- strukturirano občutljivi na napake (slika 3.4 desno), kjer je i -ti residual neobčutljiv na i -to napako, to pa pomeni, da je pri nastopu i -te napake vrednost i -tega residuala enaka nič, ostali residuali pa morajo biti neničelni.



Slika 3.4: Usmerjeno (levo) in strukturirano občutljivi residuali (desno)

Metode paritetnih enačb so razmeroma enostavno razumljive, po drugi strani pa moramo omeniti tudi njihove slabe strani:

- izbira transformacijske matrike W ni enoumna [42],[32];
- število napak, ki se jih da narediti neobčutljive za določeni residual iz nabora residualov, je manjše od števila izhodov sistema [42];
- želenega odziva residualov in vedno mogoče doseči zaradi motenj, občutljivosti modela procesa na spremembe parametrov itd.

3.2.4 Pristopi na podlagi nelinearnega observatorja

V tradicionalni pristopih zaznavanja napak za nelinearne sisteme vedno uporabimo eno od tehnik linearizacije, ki prevede nelinearni problem v načrtovanje na podlagi linearnega modela. V inženirski praksi je taka rešitev splošno utemeljena in posledično tudi zelo razširjena. Vendar pa je stopnja nelinearnosti nekaterih sistemov tako visoka, da lahko pogoške linearizacije sistem napačno interpretira in prihaja do lažnih alarmov ali pa, ob zvišanju toleranc, do premajhne občutljivosti na napake. Razširitev tradicionalnih pristopov zaznavanja napak na nelinearne modele procesov je potekalo predvsem v osemdesetih in devetdesetih letih prejšnjega stoletja, ko so nova dognanja na področju linearnega modeliranja in simulacije ter razvoj digitalnih računalnikov dali zagon nelinearnim metodam računalniškega vodenja.

Metode na podlagi nelinearnega observatorja so podrobno predstavljene v delih [4] in [33]. Razdelimo jih v dva razreda:

- metode na podlagi linearizacije preko povratne vezave iz stanj procesa (ang. linearization via state feedback, LSF);
- metode na podlagi teorije stabilizacije nelinearnih sistemov (ang. theory of nonlinear-system stabilization, NSS)

V nadaljevanju bomo na kratko opisali obe metodi.

Metoda na podlagi teorije LSF

Imejmo poseben tip nelinearnega dinamičnega sistema, ki ga opisuje enačba

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}(\mathbf{u}(t), \mathbf{y}(t)) + \mathbf{E}_f(\mathbf{x}(t), \mathbf{u}(t), \mathbf{f}(t)) + \mathbf{E}_d\mathbf{d}(t), \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{F}_f(\mathbf{x}(t), \mathbf{u}(t), \mathbf{f}(t)) + \mathbf{F}_d\mathbf{d}(t),\end{aligned}\tag{3.24}$$

kjer so \mathbf{x} stanja, \mathbf{u} vhodi in \mathbf{y} izhodi procesa, \mathbf{f} je vektor napak in \mathbf{d} vektor neznanih vhodov v sistem. $\mathbf{B}(\mathbf{u}(t), \mathbf{y}(t))$, $\mathbf{E}_f(\mathbf{x}(t), \mathbf{u}(t), \mathbf{f}(t))$ in \mathbf{E}_d so vektorska polja, \mathbf{A} , \mathbf{C} , \mathbf{E}_d in \mathbf{F}_d pa matrike ustreznih dimenzij s poznanimi elementi. Wünnenberg [135] je za oblikovanje residualov predlagal sledeči dinamični sistem:

$$\begin{aligned}\frac{d\hat{\mathbf{x}}(t)}{dt} &= \mathbf{A}\hat{\mathbf{x}}(t) + \mathbf{B}(\mathbf{u}(t), \mathbf{y}(t)) + \mathbf{L}(\mathbf{y}(t) - \mathbf{C}\hat{\mathbf{x}}(t)), \\ \mathbf{r}(t) &= \mathbf{V}(\mathbf{y}(t) - \mathbf{C}\hat{\mathbf{x}}(t)),\end{aligned}\tag{3.25}$$

z enačbo dinamike pogreška in residualom

$$\begin{aligned}\frac{d\hat{\mathbf{x}}(t)}{dt} &= \mathbf{A}\hat{\mathbf{x}}(t) + \mathbf{B}(\mathbf{u}(t), \mathbf{y}(t)) + \mathbf{L}(\mathbf{y}(t) - \mathbf{C}\hat{\mathbf{x}}(t)), \\ \mathbf{r}(t) &= \mathbf{V}(\mathbf{y}(t) - \mathbf{C}\hat{\mathbf{x}}(t)).\end{aligned}\tag{3.26}$$

Sistem v (3.25) imenujemo generator residuala z linearno dinamiko pogreška, ker je v odsotnosti napak dinamika v (3.26) linearna.

Zdaj pa vzemimo nelinearni sistem v splošnejši obliki

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}(\mathbf{x}(t), \mathbf{u}(t)) + \mathbf{E}_d(\mathbf{x}(t))\mathbf{d}(t) + \mathbf{E}_f(\mathbf{x}(t), \mathbf{u}(t), \mathbf{f}(t)), \\ \mathbf{y}(t) &= \mathbf{C}(\mathbf{x}(t), \mathbf{u}(t)) + \mathbf{F}_d(\mathbf{x}(t))\mathbf{d}(t) + \mathbf{F}_f(\mathbf{x}(t), \mathbf{u}(t), \mathbf{f}(t))\end{aligned}\tag{3.27}$$

s poznanimi vektorskimi polji $\mathbf{C}(\mathbf{x}(t), \mathbf{u}(t))$, $\mathbf{E}_d(\mathbf{x}(t))$ in $\mathbf{F}_d(\mathbf{x}(t))$. Generator residuala z linearno dinamiko zgradimo v dveh korakih:

1. sistem v (3.27) preoblikujemo v (3.24);
2. načrtamo generator residuala po (3.26).

Predpostavimo, da transformacijo stanj \mathbf{x} označimo z $\mathbf{z}(t) = \mathbf{T}(\mathbf{x}(t))$. Enačba (3.27) tako postane

$$\dot{\mathbf{z}}(t) = \frac{\partial \mathbf{T}(\mathbf{x}(t))}{\partial \mathbf{x}(t)} (\mathbf{A}(\mathbf{x}(t), \mathbf{u}(t)) + \mathbf{E}_d(\mathbf{x}(t)) \mathbf{d}(t) + \mathbf{E}_f(\mathbf{x}(t), \mathbf{u}(t), \mathbf{f}(t))). \quad (3.28)$$

Če lahko izberemo $\mathbf{T}(\mathbf{x}(t))$ in $\mathbf{r}(\mathbf{T}(\mathbf{x}(t)), \mathbf{C}(\mathbf{x}(t), \mathbf{u}(t)))$ tako, da velja

$$\begin{aligned} \frac{\partial \mathbf{T}(\mathbf{x}(t))}{\partial \mathbf{x}(t)} (\mathbf{A}(\mathbf{x}(t), \mathbf{u}(t))) &= \mathbf{G}\mathbf{T}(\mathbf{x}(t)) + \mathbf{B}(\mathbf{x}(t), \mathbf{u}(t)), \\ \frac{\partial \mathbf{T}(\mathbf{x}(t))}{\partial \mathbf{x}(t)} \mathbf{E}_d(\mathbf{x}(t)) \mathbf{d}(t) &= \mathbf{0}, \\ \frac{\partial \mathbf{T}(\mathbf{x}(t))}{\partial \mathbf{x}(t)} (\mathbf{E}_f(\mathbf{x}(t), \mathbf{u}(t), \mathbf{f}(t))) &\neq \mathbf{0}, \\ \mathbf{r}(\mathbf{T}(\mathbf{x}(t)), \mathbf{C}(\mathbf{x}(t), \mathbf{u}(t))) &= \mathbf{0}, \end{aligned} \quad (3.29)$$

potem dinamični sistem

$$\begin{aligned} \frac{d\hat{\mathbf{z}}(t)}{dt} &= \mathbf{G}\hat{\mathbf{z}}(t) + \mathbf{B}(\mathbf{u}(t), \mathbf{y}(t)) \\ \mathbf{r}(t) &= r(\hat{\mathbf{z}}(t), \mathbf{y}(t)) \end{aligned} \quad (3.30)$$

rezultira v generatorju residuala, čigar dinamika je linearna in podana z enačbo

$$\begin{aligned} \dot{\mathbf{e}}(t) &= \mathbf{G}\mathbf{e}(t) = \frac{\partial \mathbf{T}(\mathbf{x}(t))}{\partial \mathbf{x}(t)} (\mathbf{E}_f(\mathbf{x}(t), \mathbf{u}(t), \mathbf{f}(t))), \\ \mathbf{r}(t) &= \mathbf{r}(\mathbf{T}(\mathbf{x}(t)) + \mathbf{e}(t), \mathbf{C}(\mathbf{x}(t), \mathbf{u}(t))). \end{aligned} \quad (3.31)$$

Enačbo (3.31) lahko interpretiramo kot Luenbergerjeve pogoje iz razdelka 3.2.2. V delu [110] so bili predstavljeni pogoji za eksistenco metode, pokazano pa je bilo tudi, da so v splošnem zelo restriktivni in jim je moč zadostiti le v nekaterih posebnih primerih.

Metode na podlagi teorije stabilizacije NSS

Prva razširitev teorije zaznavanja napak na podlagi observatorja na nelinearne sisteme je delo [48], v katerem sta avtorja obravnavala sistem

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A}(\mathbf{x}(t), \mathbf{u}(t), \mathbf{f}_c(t)) \\ \mathbf{y}(t) &= \mathbf{C}(\mathbf{x}(t), \mathbf{u}(t), \mathbf{f}_s(t)) \end{aligned} \quad (3.32)$$

kjer \mathbf{f}_c in \mathbf{f}_s označujeta vektorja napak komponent in senzorjev kot odmike od nominalnih vrednosti v \mathbf{f}_{c0} in \mathbf{f}_{s0} . Avtorja predlagata za zaznavanje napak sistem

$$\begin{aligned}\frac{d\hat{\mathbf{x}}(t)}{dt} &= \mathbf{A}(\mathbf{x}(t), \mathbf{u}(t), \mathbf{f}_{c0}(t)) + \mathbf{L}(\hat{\mathbf{x}}(t), \mathbf{u}(t))(\mathbf{y}(t) - \hat{\mathbf{y}}(t)) \\ \mathbf{y}(t) &= \mathbf{C}(\mathbf{x}(t), \mathbf{u}(t), \mathbf{f}_{s0}(t)) \\ \mathbf{r}(t) &= \mathbf{y}(t) - \hat{\mathbf{y}}(t)\end{aligned}\quad (3.33)$$

Ob uvedbi $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$ dobimo

$$\begin{aligned}\dot{\mathbf{e}}(t) &= \left. \frac{\partial \mathbf{A}(\hat{\mathbf{x}}(t), \mathbf{u}(t), \mathbf{f}_{c0})}{\partial \mathbf{x}(t)} \right|_{\mathbf{x}=\hat{\mathbf{x}}} \mathbf{e}(t) - \mathbf{L}(\hat{\mathbf{x}}(t), \mathbf{u}(t)) \left. \frac{\partial \mathbf{C}(\hat{\mathbf{x}}(t), \mathbf{u}(t), \mathbf{f}_{s0})}{\partial \mathbf{x}(t)} \right|_{\mathbf{x}=\hat{\mathbf{x}}} \mathbf{e}(t) + O_1(\mathbf{e}) \\ \mathbf{r}(t) &= \left. \frac{\partial \mathbf{C}(\hat{\mathbf{x}}(t), \mathbf{u}(t), \mathbf{f}_{s0})}{\partial \mathbf{x}(t)} \right|_{\mathbf{x}=\hat{\mathbf{x}}} \mathbf{e}(t) + O_2(\mathbf{e})\end{aligned}\quad (3.34)$$

kjer $O_1(\mathbf{e})$ in $O_2(\mathbf{e})$ predstavljata člene drugega reda in višjih redov glede na \mathbf{e} . Osnovna ideja pristopa se skriva v načrtovanju matrike $\mathbf{L}(\hat{\mathbf{x}}, \mathbf{u})$, ki zagotovi stabilno ravnotežno točko $\mathbf{e} = \mathbf{0}$, če zanemarimo člene v $O_1(\mathbf{e})$ in $O_2(\mathbf{e})$. Iz stabilnostne teorije pa je znano, da take matrike $\mathbf{L}(\hat{\mathbf{x}}, \mathbf{u})$ ni moč najti, če nimamo nobenih podatkov o nelinearnostih v $\mathbf{A}(\mathbf{x}, \mathbf{u}, \mathbf{f}_c)$ in $\mathbf{C}(\mathbf{x}, \mathbf{u}, \mathbf{f}_s)$. Do sedaj se je pojavilo nekaj del, kjer so bile izpeljane splošne rešitve za določene posebne oblike nelinearnih sistemov (npr. [30]).

V zadnjem času pa se bolj uveljavljajo izpeljanke, kjer na podlagi teorije nelinearnih observatorjev iščemo sistematičen razvoj residuala za posamezne razrede nelinearnih sistemov. Tako je bila na primer v delu [4] prikazana metoda načrtovanja generatorja residuala, kjer se nelinearni observator oblikuje na podlagi Lipschitzovega pogoja. Oglejmo bistvene točke pristopa, kot je bilo opisano v [109]. Obravnavamo model procesa

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \bar{\mathbf{A}}(\mathbf{x}(t), \mathbf{u}(t)) + \mathbf{E}_d(\mathbf{x}(t), \mathbf{u}(t)) + \mathbf{E}_f(\mathbf{x}(t), \mathbf{u}(t))\mathbf{f}(t), \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t),\end{aligned}\quad (3.35)$$

ki zadosti naslednjima predpostavkama:

1. linearni del sistema, tj. par \mathbf{C}, \mathbf{A} je spoznaven;
2. nelinearnost $\bar{\mathbf{A}}(\mathbf{x}(t), \mathbf{u}(t))$ zadosti Lipschitzovemu pogoju, ki pravi, da obstaja pozitivna konstanta ε , tako da je

$$\|\bar{\mathbf{A}}(\mathbf{x}(t_1), \mathbf{u}(t_1)) - \bar{\mathbf{A}}(\mathbf{x}(t_2), \mathbf{u}(t_2))\| \leq \varepsilon \|\mathbf{x}(t_1) - \mathbf{x}(t_2)\|. \quad (3.36)$$

Za sistem enačb

$$\begin{aligned}\dot{\mathbf{z}}(t) &= \mathbf{N}\mathbf{z}(t) + \mathbf{L}\mathbf{y}(t) + \mathbf{G}\mathbf{u}(t) + \mathbf{T}\bar{\mathbf{A}}(\hat{\mathbf{x}}(t), \mathbf{u}(t)) - \mathbf{P}^{-1}\mathbf{C}(\hat{\mathbf{y}}(t) - \mathbf{y}(t)), \\ \hat{\mathbf{y}}(t) &= \mathbf{C}\hat{\mathbf{x}}(t), \quad \hat{\mathbf{x}}(t) = \mathbf{z}(t) - \mathbf{K}\mathbf{y}(t)\end{aligned}\quad (3.37)$$

je bilo pokazano, da je stabilen in da omogoči oblikovanje residuala

$$\mathbf{r}(t) = \mathbf{C}\mathbf{z}(t) - (\mathbf{C}\mathbf{K} - \mathbf{I})\mathbf{y}(t), \quad (3.38)$$

kjer morajo matrice \mathbf{G} , \mathbf{K} , \mathbf{L} , \mathbf{N} in \mathbf{T} zadostiti pogojem

$$\begin{aligned} \mathbf{N} - \mathbf{P}^{-1}\mathbf{C}^T\mathbf{C} \text{ je stabilen, } \mathbf{G} &= \mathbf{T}\mathbf{B}, \\ \mathbf{N}\mathbf{T} - \mathbf{T}\mathbf{A} &= -\mathbf{L}\mathbf{C}, \quad \mathbf{T} + \mathbf{K}\mathbf{C} = \mathbf{I}. \end{aligned} \quad (3.39)$$

Matrika \mathbf{P} , ki je odvisna od parametrov α in κ , je rešitev Ljapunove enačbe

$$\mathbf{N}^T\mathbf{P} + \mathbf{P}\mathbf{N} - \alpha\mathbf{C}^T\mathbf{C} + \kappa\mathbf{P} = 0, \quad (3.40)$$

parametra α in κ pa izračunamo tako, da je \mathbf{P} pozitivno definitna in da Lipschitzova konstanta ε zadosti neenakosti

$$\varepsilon \leq \frac{\lambda_{\min}((2 - \alpha)\mathbf{C}^T\mathbf{C} + \kappa\mathbf{P})}{2\sigma_{\max}\mathbf{P}\mathbf{T}}. \quad (3.41)$$

Dinamika sistema v enačbah (3.37) in (3.38) je

$$\begin{aligned} \dot{\mathbf{e}}(t) &= (\mathbf{N} - \mathbf{P}^{-1}\mathbf{C}^T)\mathbf{C}\mathbf{e}(t) + \mathbf{T}(\bar{\mathbf{A}}(\mathbf{x}(t), \mathbf{u}(t)) - \bar{\mathbf{A}}(\hat{\mathbf{x}}(t), \mathbf{u}(t))) - \\ &\quad - \mathbf{T}\mathbf{E}_d(\mathbf{x}(t), \mathbf{u}(t))\mathbf{d}(t) - \mathbf{T}\mathbf{E}_f(\mathbf{x}(t), \mathbf{u}(t))\mathbf{f}(t), \end{aligned} \quad (3.42)$$

$$\mathbf{r}(t) = \mathbf{C}\mathbf{e}(t). \quad (3.43)$$

Očitno je, da $\mathbf{r}(t)$ vsebuje informacijo o napaki, zato ga lahko uporabimo kot residual za zaznavanje napak.

3.2.5 Pristopi s pomočjo adaptivnega observatorja

Ena od večjih prednosti zaznavanja napak na podlagi observatorja je enostavno zagotavljanje robustnosti na vplive negotovosti modela. Ker pa je povečanje robustnosti povezano z zmanjšanjem občutljivosti zaznavanja napak z dolgimi časovnimi konstantami, obstaja tveganje, da počasi razvijajočih se napak ne bomo mogli odkriti [7]. Konec osemdesetih in v začetku devetdesetih letih prejšnjega stoletja so strokovnjaki s področja zaznavanja napak v izogib tem težavam začeli razvijati metode, ki v observatorske metode integrirajo metode ocenjevanja parametrov. Oceno stanj in počasi spreminjajočih se parametrov dobimo s pomočjo tako imenovanega adaptivnega observatorja. Pričakovati bi bilo, da bo generator residuala, osnovan na adaptivnem observatorju, poleg zgodnje zaznave nenadnih (hitrih) napak omogočil tudi zaznavo

počasi razvijajočih se napak. Poleg tega je motivacija za uvedbo adaptivnih observatorskih metod tudi povečanje robustnosti na negotovosti v procesu zaradi sprotnega posodabljanja modela.

Bastin in Gevers [9] sta prva predstavila koncept adaptivnega observatorja, ki je razmeroma enostaven za načrtovanje in implementacijo ter ni uporaben samo za linearne, pač pa tudi za razred nelinearnih sistemov. Osnovna ideja njenega pristopa je, da lahko določene vplive nelinearnosti obravnavamo kot neznane parametre, ki jih sprotno ocenjujemo.

Predpostavimo nelinearni sistem

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{a}(\mathbf{x}(t)) + \mathbf{q}_0(\mathbf{x}(t), \mathbf{u}(t)) + \mathbf{Q}(\mathbf{x}(t), \mathbf{u}(t)) \boldsymbol{\theta}(t) + \mathbf{H}(\mathbf{x}(t), \mathbf{u}(t)) \mathbf{f} + \mathbf{g}(t), \\ \mathbf{y}(t) &= \mathbf{c}(\mathbf{x}(t)),\end{aligned}\quad (3.44)$$

kjer so \mathbf{x} , \mathbf{y} in \mathbf{u} vektorji stanj procesa, merljivih izhodov procesa in vhodov, \mathbf{a} , $\mathbf{Q}(\mathbf{x}(t), \mathbf{u}(t))$ in $\mathbf{H}(\mathbf{x}(t), \mathbf{u}(t))$ pa so poznane gladke nelinearne funkcije. $\boldsymbol{\theta}(t)$ je vektor, ki predstavlja nepoznane časovno spremenljive procesne parametre, počasi spreminjajoče se napake ali del nelinearnosti sistema. Naj bo njegov odvod omejen z

$$\left| \dot{\boldsymbol{\theta}}(t) \right| \leq M, \quad 0 < M \ll \infty. \quad (3.45)$$

Elemente vektorja $\boldsymbol{\theta}$ se ocenjuje sprotno. Vektor \mathbf{f} predstavlja nenadne spremembe (napake) v sistemu, ki jih želimo zaznati v čim krajšem času.

Iz tega zapisa lahko neposredno izpeljemo splošne izraze za računanje adaptivnega generatorja residuala:

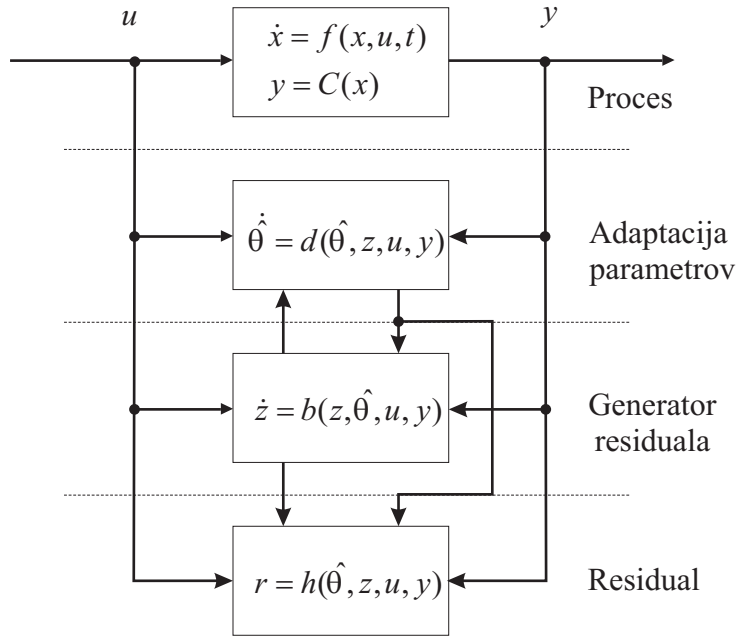
$$\begin{aligned}\dot{\hat{\mathbf{z}}}(t) &= \mathbf{b}(\hat{\mathbf{z}}(t), \hat{\boldsymbol{\theta}}(t), \mathbf{u}(t), \mathbf{y}(t)), \\ \mathbf{r}(t) &= \mathbf{h}(\hat{\mathbf{z}}(t), \hat{\boldsymbol{\theta}}(t), \mathbf{u}(t), \mathbf{y}(t))\end{aligned}\quad (3.46)$$

za generator residuala in

$$\dot{\hat{\boldsymbol{\theta}}}(t) = \mathbf{d}(\hat{\mathbf{z}}(t), \hat{\boldsymbol{\theta}}(t), \mathbf{u}(t), \mathbf{y}(t)) \quad (3.47)$$

Kako dobimo residual, je shematsko prikazano na sliki 3.5. Bastin in Gevers sta pokazala, da se splošni nelinearni sistem (3.44) z ustreznimi transformacijami da prevesti v ti. adaptivno observatorsko kanonično obliko (ang. adaptive observer canonical form, AOCF)

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{R}\mathbf{x}(t) + \boldsymbol{\Omega}(\boldsymbol{\omega}(t)) \boldsymbol{\theta}(t) + \mathbf{g}(t), \\ \mathbf{y}(t) &= \mathbf{c}\mathbf{x}(t)\end{aligned}\quad (3.48)$$



Slika 3.5: Uporaba adaptivnega observatorja za oblikovanje residuala

kjer je $\omega(t)$ vektor *znanih* nelinearnih funkcij vhoda in izhoda procesa, $\Omega(\omega(t))$ je matrika uteženih funkcij $\omega(t)$, \mathbf{R} pa je konstantna matrika, ki ji lahko poljubno določimo lastne vrednosti in s tem vplivamo na dinamiko observatorja. Za tak sistem načrtamo adaptivni observator

$$\dot{\hat{\mathbf{x}}}(t) = \mathbf{R}\hat{\mathbf{x}}(t) + \Omega(\omega(t))\hat{\boldsymbol{\theta}}(t) + \mathbf{g}(t) + \begin{bmatrix} c_1\tilde{\mathbf{y}}(t) \\ \mathbf{V}(t)\dot{\hat{\boldsymbol{\theta}}}(t) \end{bmatrix}, \quad (3.49)$$

$$\hat{\mathbf{y}}(t) = \mathbf{x}_1(t), \quad \tilde{\mathbf{y}}(t) \triangleq \mathbf{y}(t) - \hat{\mathbf{y}}(t),$$

kjer je c_1 poljubna konstanta, pri čemer adaptacijo parametrov določa enačba

$$\dot{\hat{\boldsymbol{\theta}}}(t) = \mathbf{\Gamma}\varphi(t)\tilde{\mathbf{y}}(t), \quad (3.50)$$

kjer je $\mathbf{\Gamma}$ poljubna (največkrat diagonalna) pozitivno definitna matrika. $\mathbf{V}(t)$ in $\varphi(t)$ sta matrika in vektor, ki rešita enačbo pomožnega filtra

$$\begin{aligned} \dot{\mathbf{V}}(t) &= \mathbf{F}\mathbf{V}(t) + \bar{\Omega}(\omega(t)), \\ \varphi(t) &= \mathbf{V}^T(t)\mathbf{k} + \Omega_1(\omega(t)), \quad \Omega \triangleq \begin{bmatrix} \Omega_1 \\ \bar{\Omega} \end{bmatrix}, \end{aligned} \quad (3.51)$$

\mathbf{F} in \mathbf{k} pa sta podmatriki matrike \mathbf{R} . Avtorja sta pokazala, da je pri taki obliki sistema adaptivni observator stabilen in identifikabilen v smislu ocenjevanja parametrov $\hat{\boldsymbol{\theta}}$.

Generator residuala enostavno izvedemo po enačbi

$$\mathbf{r}(t) = \mathbf{y}(t) - \mathbf{c}\dot{\mathbf{x}}(t). \quad (3.52)$$

Na ta način nenadne napake \mathbf{f} zaznamo z vrednotenjem residuala, počasi razvijajoče spremembe pa zaznamo s sprotnim ocenjevanjem parametrov $\hat{\boldsymbol{\theta}}$. Nekaj let kasneje sta Ding in Frank [22] na podoben način izpeljala enačbe residualnega generatorja za reducirani red observatorja ter razširila rezultate na sisteme MIMO.

Druga pomembna metodologija v sklopu adaptivnih observatorjev pa je osnovana na naboru estimatorjev napak s pomočjo ti. metodologije učenja (ang. learning methodology) [123],[119],[92],[138],[139]. Sistem vsebuje $N + 1$ nelinearnih adaptivnih observatorjev, kjer je N število modeliranih napak, ki jih želimo zaznati. Prvi služi za zaznavanje napak, ostali pa za lokalizacijo napak. V normalnem načinu delovanja sistema je aktiven samo prvi observator. Ko pride do napake in z vrednotenjem prvega residuala zaznamo pojav napake, se aktivirajo preostali observatorji, ki s prilagajanjem parametrov ocenjujejo vsak svoj model napake (vsak od njih je v povezavi z določeno napako). Na ta način dobimo ne samo lokalizacijo napake, pač pa tudi oceno njene amplitude in dinamike. Če sistem napake ne prepozna, lahko novi model doda v nabor estimatorjev. V delu [137] so avtorji predlagali razširjeni sistem, ki na podlagi informacije iz estimatorjev prilagodi parametre adaptivnega regulatorja, tako da stanja in izhodi sistema ostanejo znotraj predpisanih meja. Glavna težava tega pristopa pri prenosu v prakso je potreba po modeliranju napak in zahtevna matematična analiza.

3.2.6 Robustno aktivno oblikovanje residualov

Predpostavimo model procesa

$$y(s) = G_u(s)u(s) + G_d(s)d(s) + G_f(s)f(s), \quad (3.53)$$

kjer so $G_u(s)$, $G_d(s)$ in $G_f(s)$ poznane prenosne funkcije med vhom, neznanimi vhodi in napakami ter izhodom procesa, $f(s)$ napake, $d(s)$ pa neznan vhodi, ki so omejeni z $\|d\|_2 \leq \Delta_d$. Že iz najenostavnejše izvedbe residuala

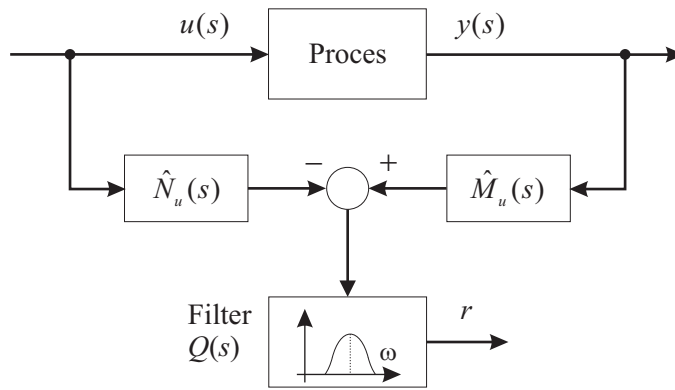
$$r(s) = y(s) - G_u(s)u(s) = G_d(s)d(s) + G_f(s)f(s) \quad (3.54)$$

je očitno, da pri neznanem vhomu $d(s) \neq 0$ v odsotnosti napak $f(s) = 0$ vrednost residuala ne bo enaka nič. Obstaja sicer veliko metod, kjer se z različnimi tehnikami doseže popolno ločitev residuala (tj. vplivov napak) od vplivov neznanih vhomov, npr. z geometrijskim pristopom [81] s filtriranjem residuala v frekvenčnem prostoru [135],[25], v paritetnem prostoru [44] itd., ki pa jih ne bomo podrobneje obravnavali.

Bolj pomembne so metode optimizacije generatorja residuala v frekvenčnem in časovnem prostoru, ki so se razvile v poznih osemdesetih in predvsem devetdesetih letih prejšnjega stoletja. Cilj je načrtati filter, ki minimizira vpliv neznanih vhodov na residual glede na vpliv napak, tako da zaradi vplivov negotovosti ni treba povečevati pragov residuala in tako zmanjševati občutljivost na napake. Enačbo (3.54) zapišemo v faktorizirani obliki [21],[20]

$$\begin{aligned} r(s) &= Q(s)\hat{M}(s)(y(s) - G_u(s)u(s)) = Q(s)\hat{M}(s)y(s) - \hat{N}_u(s)u(s) \\ &= Q(s)\hat{M}(s)(G_d(s)d(s) + G_f(s)f(s)), \end{aligned} \quad (3.55)$$

kjer sta $\hat{M}(s)$ in $\hat{N}(s)$ faktorja, za katera velja $\hat{M}(s)^{-1}\hat{N}(s) = G_u(s)$, $Q(s)$ pa označuje prenosno funkcijo filtra. Shema je prikazana na sliki 3.6 Vrednotenje residuala izvedemo



Slika 3.6: Splošna shema oblikovanja robustnega residuala

z 2-normo in premičnim oknom, kar za časovni prostor zapišemo kot

$$\|r\|_e = \|r\|_{2,\tau} = \left(\int_{t_1}^{t_2} r^T(t)r(t)dt \right)^{1/2}, \quad \tau = t_2 - t_1, \quad (3.56)$$

za frekvenčni prostor pa

$$\|r\|_e = \|r\|_{2,\Phi} = \left(\frac{1}{2\pi} \int_{\omega_1}^{\omega_2} r^T(-j\omega)r(j\omega)d\omega \right)^{1/2}, \quad \Phi = \omega_2 - \omega_1. \quad (3.57)$$

Za odločitev o napaki moramo izbrati vrednost praga (ang. threshold). Ker je najbolj običajna odločitvena logika

$$\begin{aligned} \|r\|_e &\geq J_p \Rightarrow \text{napaka } (f = 1) \text{ in alarm,} \\ \|r\|_e &< J_p \Rightarrow \text{ni napake } (f = 0), \end{aligned} \quad (3.58)$$

definiramo prag kot

$$J_p = \sup_{d, f=0} \|r\|_e, \quad (3.59)$$

kar v povezavi z enačbo (3.55) da

$$J_p = \sup_d \left\| Q(s) \hat{M}_u(s) G_d(s) d(s) \right\|_e = \left\| Q(s) \hat{M}_u(s) G_d(s) \right\|_\infty \Delta_d. \quad (3.60)$$

Napaka pa ima lahko zaradi negotovosti v modelu različen vpliv na residual in posredno tudi na odločitveno funkcijo. Zato lahko rečemo, da je napaka zaznavna, če velja

$$\inf_{\Delta y} \|r\|_e \geq J_p, \quad (3.61)$$

tj. napako lahko zaznamo, če povzroči, odločitvena funkcija preseže vrednost residuala za vse možne negotovosti v sistemu. Tako pridemo do pojma *množica zaznavnih napak*, ki jo definiramo kot

$$S_f := \left\{ f : \inf_{\Delta y} \|r\|_e \geq J_p \right\}. \quad (3.62)$$

Velikost množice S_f je odvisna od strukture $Q(s)$, funkcije $\|\cdot\|$ in praga. Sistem za zaznavanje napak je tako optimalen, kadar velikost množice S_f doseže maksimum.

Predlaganih je bilo več rešitev optimizacije parametrov $Q(s)$ v časovnem prostoru. Lou in soavtorji [77] so razvili algoritem, osnovan na dekompoziciji singularnih vrednosti v paritetnem prostoru, ki pa, roko na srce, ni enostaven za uporabo. Frank in Wünnenberg [34],[135] pa sta predstavila metodo, ki maksimizira normo občutljivosti residuala na napake glede na neznane vhode, tako da optimizacijski problem prevede na posplošeni problem tipa lastna vrednost - lastni vektor (ang. generalized eigenvector-eigenvalue problem).

Bolj razširjene so rešitve v frekvenčnem prostoru, ki ponuja več manevrskega prostora pri specifikaciji napak in močnejše metode analize robustnosti, kot sta na primer teoriji H_∞ in H_2 . V delih [21],[89],[100],[36],[20],[47] so avtorji predlagali različne pristope, kjer je problem robustnega zaznavanja za strukturirane in nestrukturirane negotovosti rešen na podlagi splošnega zapisa residuala v (3.54). Poglejmo si kratek opis skupnih značilnosti teh pristopov.

Predpostavimo, da želimo maksimizirati kriterijsko funkcijo

$$J = \frac{\|\partial r / \partial f\|}{\|\partial r / \partial d\|}, \quad (3.63)$$

pri čemer $\|\cdot\|$ označuje normo občutljivosti na napake f in neznane vhode d . Cilj načrtovanja je najti tak generator residuala, da postane J maksimalen. Pri zapisu

(3.55) lahko (3.63) izrazimo tudi kot

$$J = \frac{\left\| Q(s)\hat{M}_u(s)G_f(s) \right\|}{\left\| Q(s)\hat{M}_u(s)G_d(s) \right\|}. \quad (3.64)$$

Če za $\|\cdot\|$ uporabimo normo L_2 , označimo $G_1(s) = \hat{M}_u(s)G_d(s)$ in $G_2(s) = \hat{M}_u(s)G_f(s)$ ter reduciramo $Q(s)$ v vektor, ki ga označimo s $q(s)$, potem $J \xrightarrow{Q(s)} \max$ v posplošenem zapisu postane

$$J = \frac{\|q(s)G_2(s)\|_2}{\|q(s)G_1(s)\|_2} \xrightarrow{q(s)} \max. \quad (3.65)$$

Maksimizacija nas pripelje do enačbe

$$V(j\omega) [G_2(j\omega)G_1^T(-j\omega) - \lambda(\omega)G_1(j\omega)G_2^T(-j\omega)] = 0, \quad (3.66)$$

kjer $V(j\omega)$ označuje lastni vektor, $\lambda(\omega)$ pa lastno vrednost. To je posplošeni problem tipa lastna vrednost - lastni vektor. Rešitev problema v frekvenčnem prostoru je podana z [31]

$$q_{opt}(s) = q_f(s)V(s), \quad J_{opt} = \sup_{\omega} (\lambda(\omega)). \quad (3.67)$$

Lastni vektor $V(s)$ določa izbiro optimalnega generatorja residuala, maksimalna lastna vrednost $\lambda(\omega_0)$ je vrednost kriterijske funkcije pri optimalni frekvenci ω_0 , $q_f(s)$ pa predstavlja frekvenčni pas, ki izloči vrednost $r_L(j\omega_0)$ pri frekvenci ω_0 . To pomeni, da filter izbere samo tisti del frekvenčnega spektra residuala, ki da najboljši kompromis med občutljivostjo na napake in občutljivostjo na neznane vhode.

Kasnejše študije so pokazale, da lahko uporaba drugačnih oblik kriterijske funkcije močno vpliva na učinkovitost zaznavanja napak. Primer je

$$J = \frac{\|Q(s)G_2(s)\|_-}{\|Q(s)G_1(s)\|_\infty} \xrightarrow{Q(s)} \max, \quad (3.68)$$

kjer $\|Q(s)G_2(s)\|_- = \inf_{f \neq 0} \|Q(s)G_2(s)f(s)\|$ pomeni minimalen vpliv napake na residual. Praktičen pomen te izbire se kaže v tem, da uporablja koncept minimalne zaznavne napake (ang. minimum detectable fault) f_{min} , ki je definirana kot

$$\|f_{min}\|_e = \inf \{ \|f\|_e : f \in S_f \}. \quad (3.69)$$

Za sisteme (3.55) dobimo

$$\inf_{\Delta y} \|r\|_e = \inf_d \left\| Q\hat{M}_u(G_f f + G_d d) \right\|_e = \left\| Q\hat{M}_u G_f f \right\|_e - \sup_d \left\| Q\hat{M}_u G_d d \right\|_e. \quad (3.70)$$

Če prag označimo z $J_p = \sup_d \left\| Q\hat{M}_u G_d d \right\|_e$, dobimo $\left\| Q\hat{M}_u G_f f \right\|_e \geq 2J_p$ in končno

$$\|f_{min}\|_e = \frac{2J_p}{\inf_{\|f\|_e=1} \left\| Q\hat{M}_u G_f f \right\|}. \quad (3.71)$$

Ding in soavtorji so v delih [24] in [23] podali tudi postopek optimalnega izračuna minimalne zaznavne napake, ki je ekvivalenten rešitvi naslednjega optimizacijskega problema:

$$\inf_{Q(s)} \|f_{\min}\|_e = \inf_{Q(s)} \frac{2 \max_{\omega=\varepsilon} \bar{\sigma} \left(\delta_d(\omega) Q(j\omega) \hat{M}_u(j\omega) G_d(j\omega) \right)}{\min_{\omega=\varepsilon} \underline{\sigma} \left(Q(j\omega) \hat{M}_u(j\omega) G_f(j\omega) \right)}, \quad (3.72)$$

kjer je $\underline{\sigma}(G(j\omega))$ minimalna singularna vrednost prenosne funkcije $G(j\omega)$.

3.2.7 Pasivno oblikovanje residualov

Kot smo videli v razdelku 3.2.6, pri aktivnih metodah robustnost residualov dosežemo s filtri, ki zmanjšajo občutljivost residuala na vplive negotovosti in povečajo občutljivost na eno ali več napak. Vendar pa je, kot je opisano v [42], v primeru modelov z negotovimi parametri popolna ločitev residualov od negotovosti omejena s številom merjenih izhodov in možna samo za omejeno število parametrov. Alternativno ponuja strategija *pasivnega oblikovanja residuala*, ki zagotovi robustnost sistema zaznavanja napak v fazi odločanja [97]. Vplive negotovosti parametrov v tem primeru "prepustimo" v residual, tako da oblikujejo interval možnih vrednosti residuala, robustnost pa dosežemo z uporabo adaptivnega praga. Osnova za adaptacijo praga pa je uporaba modela z negotovimi parametri, ki so podani s pripadajočimi intervali. Puig in sodelavci so v delu [98] predstavili inačico, kjer se za model uporabi nelinearni intervalni observator. Tak tip observatorja v vsakem trenutku določi interval, ki omejuje negotovosti stanj procesa. Poglejmo si kratko ilustracijo metode.

Imejmo nelinearni dinamični sistem v diskretnem času

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{g}(\mathbf{x}(k), \mathbf{u}(k), \mathbf{d}(k), \mathbf{f}(k), \boldsymbol{\theta}), \\ \mathbf{y}(k) &= \mathbf{h}(\mathbf{x}(k), \mathbf{u}(k), \mathbf{n}(k), \mathbf{f}(k), \boldsymbol{\theta}), \end{aligned} \quad (3.73)$$

kjer je \mathbf{x} vektor stanj procesa, \mathbf{y} vektor izhodov, \mathbf{g} in \mathbf{h} sta nelinearni funkciji, $\mathbf{d}(k)$ je aditivna motnja, $\mathbf{n}(k)$ šum meritve, $\mathbf{f}(k)$ vektor napak, $\boldsymbol{\theta} \in \Theta$ pa je vektor negotovih parametrov, omejenih s hiperpravokotnikom $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^p | \underline{\boldsymbol{\theta}} \leq \boldsymbol{\theta} \leq \bar{\boldsymbol{\theta}}\}$. Za oceno stanj uporabimo nelinearni intervalni observator

$$\begin{aligned} \hat{\mathbf{x}}(k+1) &= \mathbf{g}(\hat{\mathbf{x}}(k), \mathbf{u}(k), \boldsymbol{\theta}) + \mathbf{K}(\mathbf{y}(k) - \hat{\mathbf{y}}(k)) \\ \hat{\mathbf{y}}(k) &= \mathbf{h}(\hat{\mathbf{x}}(k), \mathbf{u}(k), \boldsymbol{\theta}) \end{aligned} \quad (3.74)$$

kjer sta $\hat{\mathbf{x}}$ in $\hat{\mathbf{y}}$ oceni stanj in izhodov procesa, \mathbf{K} pa je observatorjevo ojačenje, ki zagotavlja njegovo stabilnost za vsak $\boldsymbol{\theta} \in \Theta$. Ker je v splošnem zelo težko izračunati

natančno oceno stanj glede na negotovosti parametrov, raje računamo interval ocene kot najmanjši interval, ki zavzema vsa dosegljiva stanja:

$$\begin{aligned}\hat{\mathbf{X}}(k) &= [\underline{\hat{\mathbf{x}}}(k), \bar{\hat{\mathbf{x}}}(k)], \\ \hat{x}_i(k) &= \min\{\hat{x}_i(k) : \hat{\mathbf{x}}(k) \in \hat{\mathbf{X}}(k)\}, \\ \bar{x}_i(k) &= \max\{\bar{x}_i(k) : \bar{\hat{\mathbf{x}}}(k) \in \hat{\mathbf{X}}(k)\},\end{aligned}\tag{3.75}$$

kjer $\hat{x}_i(k)$ označuje posamezno komponento vektorja stanj. Minimizacija in maksimizacija v enačbi (3.75) za časovni trenutek k je odvisna od predhodnih vrednosti trajektorije izhodov $\mathbf{y}(k-1)$, $\mathbf{y}(k-2)$, ..., $\mathbf{y}(1)$, $\mathbf{y}(0)$, torej

$$\begin{aligned}\underline{\hat{\mathbf{x}}}(k) &= \min \hat{\mathbf{x}}(k) \\ \bar{\hat{\mathbf{x}}}(k) &= \max \hat{\mathbf{x}}(k) \quad \text{glede na} \\ \hat{\mathbf{x}}(k) &= \mathbf{g}(\hat{\mathbf{x}}(k-1), \mathbf{u}(k-1), \theta) + \mathbf{K}(\mathbf{y}(k-1) - \hat{\mathbf{y}}(k-1)), \\ \hat{\mathbf{x}}(k-1) &= \mathbf{g}(\hat{\mathbf{x}}(k-2), \mathbf{u}(k-2), \theta) + \mathbf{K}(\mathbf{y}(k-2) - \hat{\mathbf{y}}(k-2)), \\ &\dots \\ \hat{\mathbf{x}}(1) &= \mathbf{g}(\hat{\mathbf{x}}(0), \mathbf{u}(0), \theta) + \mathbf{K}(\mathbf{y}(0) - \hat{\mathbf{y}}(0)), \\ \hat{\mathbf{x}}(j) &\in [\underline{\hat{\mathbf{x}}}(j), \bar{\hat{\mathbf{x}}}(j)] \quad \text{za } j = 0, \dots, k, \quad \theta \in [\underline{\theta}, \bar{\theta}]\end{aligned}\tag{3.76}$$

Optimizacijski problem v (3.76) je v splošnem nelinearen in nekonveksen, poleg tega pa se z večanjem števila meritev tudi povečuje računska kompleksnost. Druga slaba stran tega pristopa je, da ne obstaja splošna metoda za računanje stabilnega observatorja za nelinearen intervalni sistem. Zato Puig in sodelavci predlagajo poenostavitev algoritma, kjer se v vsakem koraku enačbo (3.74) linearizira okoli središča intervala $\mathbf{x}_c(k) = 1/2 \cdot (\underline{\hat{\mathbf{x}}}(k) + \bar{\hat{\mathbf{x}}}(k))$, za izračun optimizacije pa se vzame samo (končno število) L zadnjih meritev izhoda in ocen stanj. Na ta način se uvede linearen observator s spremenljivimi parametri

$$\hat{\mathbf{x}}(k+1) = \mathbf{A}(\hat{\mathbf{x}}_e(k), \theta) \hat{\mathbf{x}}(k) + \mathbf{B}(\hat{\mathbf{x}}_e(k), \theta) \mathbf{u}_n(k)\tag{3.77}$$

okoli trajektorije $\hat{\mathbf{x}}_e(k)$, kjer sta \mathbf{A} in \mathbf{B} matriki pripadajočih parcialnih odvodov nelinearne funkcije \mathbf{g} . Pokazano je bilo, da je

$$\|\mathbf{A}(\mathbf{x}_{k-L}, \theta) \mathbf{A}(\mathbf{x}_{k-L+1}, \theta) \dots \mathbf{A}(\mathbf{x}_k, \theta)\|_\infty < 1\tag{3.78}$$

zadosten pogoj za stabilnost takega observatorja. Ko izračunamo trajektorijo ocen intervalov stanj $[\underline{\hat{\mathbf{x}}}(k), \bar{\hat{\mathbf{x}}}(k)]$ in izhodov modela $[\underline{\hat{\mathbf{y}}}(k), \bar{\hat{\mathbf{y}}}(k)]$, residual preoblikujemo v intervalno obliko. Napaka se ni zgodila, če residual zadosti pogoju

$$\mathbf{r}(k) = \mathbf{y}(k) - \hat{\mathbf{y}}_c(k) \in [-\Delta\hat{\mathbf{y}}(k), \Delta\hat{\mathbf{y}}(k)] = [\underline{\mathbf{r}}(k), \bar{\mathbf{r}}(k)],\tag{3.79}$$

kjer je $\hat{y}_c(k)$ srednja vrednost ocenjenega intervala in $\Delta\hat{y}(k) = 1/2(\overline{\hat{y}}(k) - \underline{\hat{y}}(k))$ njen radij.

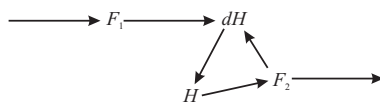
Slaba stran tega pristopa je predvsem v tem, da moramo za učinkovit izračun intervala izhodov poznati intervale negotovosti parametrov, po drugi strani pa observator s prilagajanjem ocene izhoda trenutni vrednosti merjenega izhoda delno zabiše vplive napak na izhod procesa.

3.3 Zaznavanje napak na podlagi kvalitativnega modela

Odkrivanje napak je, nasprotno od simulacije, povezano s sklepanjem o funkcionalnosti sistema glede na obnašanje, za to pa je potrebno sklepanje o vzročno-posledičnih povezavah znotraj sistema. Pri neposrednem pristopu uporabljamo hevristične informacije v obliki proizvodnih pravil. Sklepanje o vzrokih za določene posledice je tako izvedeno neposredno iz sledenja pravilom. Če pa uporabljamo pristop na podlagi fundamentalnih fizikalnih modelov, matematični opis sistema združimo z opazovanjem (meritvami) sistema. V tem primeru je sklepanje o napakah odkrivanje funkcionalnih sprememb znotraj sistema, ki so se odrazile v napačnem delovanju ali odpovedi sistema [124]. V drugem pristopu se za potrebe sklepanja zelo intenzivno uporablja kvalitativne modele.

Prvi poskus oblikovanja baze znanja, ki bi omogočala odločanje v nekem formalnem okvirju, so bili ekspertni sistemi. Ekspertni sistem je računalniški program, ki posnema kognitivne procese odločanja človeškega strokovnjaka na nekem področju. Sestavlja ga obsežna baza pravil tipa če-potem in mehanizem sklepanja, ki po bazi išče najverjetnejši vzrok za dane posledice. Obsežnost baze je ena glavnih težav ekspertnih sistemov, saj v velikih sistemih in pri velikem številu predpostavljenih napak število pravil hitro naraste čez vse razumne meje. Po drugi strani pa znanje, zajeto v pravilih, navadno ne odraža fizikalnega ozadja sistema, zato se pojavijo težave pri odločanju, kadar pride do nepredvidene situacije, ki ni zajeta v bazi.

Kasneje se je pojavila potreba po uvedbi orodij, ki bi modelirala dani sistem v kvalitativnem smislu, zajela vzročno-posledično strukturo sistema na bolj poglobljen način, kot je to pri ekspertnih sistemih, po drugi strani pa ne bi bila toga in omejena kot npr. numerična simulacija. Orodja bi omogočala kvalitativno predstavitev znanj in uvedbo metod sklepanja na podlagi teh znanj. V nadaljevanju si bomo ogledali dve najbolj znani izmed metod, usmerjene grafe in drevesa napak.



Slika 3.7: Primer usmerjenega grafa za shranjevalnik tekočine

3.3.1 Usmerjeni grafi

Usmerjeni graf ali digraf (ang. directed graph, digraph) je graf vozlišč, povezanih s krožnimi loki (ang. arc), pri predznačenem usmerjenem grafu (ang. signed directed graph, SDG) pa vsak lok označimo tudi z ustreznim predznakom. Krožni loki vodijo od vozlišč, ki pomenijo "vzroke", do vozlišč, ki pomenijo "posledice". Vsako vozlišče predstavlja odmik ustrezne spremenljivke od svojega ustaljenega stanja. Grafi SDG so veliko kompaktnjši od odločitvenih tabel ali modelov s končnim številom stanj. Za primer pogledjmo usmerjeni graf shranjevalnika tekočine, ki je predstavljen na sliki 3.7. F_1 je vhodni pretok, F_2 je izhodni pretok, H pa višina tekočine v shranjevalniku. Enačbi sistema sta

$$\begin{aligned} F_1 - F_2 &= \frac{dH}{dt}, \\ F_2 &= \frac{H}{R}, \end{aligned} \quad (3.80)$$

kjer je R hidravlična upornost izhodnega pretoka. Slika 3.7 lahko v besedi predstavimo na naslednji način: zunanji vpliv povzroči spremembo pretoka F_1 , kar povzroči spremembo nivoja tekočine (dH in H), to povzroči spremembo izhodnega pretoka, ki potem v povratni zanki ponovno vpliva na nivo tekočine. Znaki za smeri povezav predstavljajo smer vplivanja. V splošnem so lahko povezave odvisne od dogodkov, tj. odnos med dvema dogodkoma ali spremenljivkama je lahko odvisen od drugih (zunanjih) dogodkov v sistemu.

V tipičnem grafu SDG obstajajo trije tipi vozlišč:

- izvori povezav (iz njih povezave samo izhajajo), ki predstavljajo osnovne spremenljivke, katere se spreminjajo neodvisno od drugih spremenljivk;
- vhodno-izhodna vozlišča, ki jih najpogosteje imenujemo procesne spremenljivke;
- ponôri povezav (v njih se povezave samo stekajo), ki jih imenujemo izhodne spremenljivke in ki nimajo vpliva na ostale spremenljivke.

Poleg tega omenimo še dve definiciji. Močno povezana komponenta (ang. strongly connected component, SCC) je podmnožica usmerjenega grafa, če lahko vsako vozlišče

te podmnožice dosežemo iz kateregakoli od preostalih vozlišč iste podmnožice. Maksimalno močno povezana komponenta (ang. maximal strongly connected component, MSCC) je vozlišče ali SCC, v katerega ne vodi nobena vhodna povezava.

V nadaljevanju bomo navedli in opisali nekaj najpomembnejših metod uporabe usmerjenih grafov v zaznavanju napak. V začetnem delu iz leta 1979 [53] so avtorji predlagali ti. vzročno-posledične grafe (ang. cause-effect (CE) graph), ki se jih izpelje iz grafa SDG na podlagi matematičnega modela ali realnih podatkov. Graf CE je sestavljen samo iz "nenormalnih" vozlišč in konsistentnih povezav, ki potencialno pojasnjujejo lokalno napredovanje napake in s tem opazovan simptom ali vzorec. "Normalna" vozlišča so izvzeta iz obravnave, ker ne določajo poti od senzorskih vozlišč do vozlišč z napakami. Usmerjenost vozlišč v SDG tvori določen vzorec. Ker v tipičnem procesu vse procesne spremenljivke navadno niso merljive, tudi vsa vozlišča grafa niso poznana, takrat govorimo o delnem vzorcu (ang. partial pattern). Kadar nekatera vozlišča kažejo na nenavadnost, se za odkrivanje napak uporabi graf CE z delnim vzorcem (graf kvazi-CE). Usmerjenost nemerjenih vozlišč se računa sekvenčno, tako da se vsaka komponenta MSCC predpostavi kot potencialno vozlišče napake. Ko se kombinatorično iskanje usmerjenosti vozlišč izteče, se konča tudi diagnostično sklepanje.

Umeda in sodelavci [120] so pokazali, kako razviti graf SDG iz diferencialnih algebrskih enačb procesa. V tem delu so imele povezave med vozlišči dinamično spremenljivo ojačenje, da se je lahko učinkovito obravnavalo nelinearnosti v sistemu. Kokawa in sodelavci [70] pa so uporabili delno sistemsko dinamiko (npr. čas napredovanja napake od ene spremenljivke do druge), statistične informacije o verjetnosti napak na opremi in usmerjene grafe za predstavitev mreže prenašanja napak in posledično identifikacijo mesta pojava napake. Določevanje usmerjenosti povezav ni potrebno, je pa metoda omejena na sisteme brez povratnih povezav.

Kramer in sodelavci [71], [90] so združili simulacijska orodja (simulacijska drevesa, kvalitativno statično simulacijo (QSIM)) in grafe SDM v sistematičnih metodah na podlagi avtomatskega iskanja in tvorjenja pravil. Cilj je pretvorba informacije iz simulacijskih dreves preko dreves SDG v nabor logičnih pravil. Predvsem delo [90] se osredotoča na sprotno izločanje neustreznih logičnih povezav od vzrokov do napak, do katerih prihaja zaradi kvalitativne narave enačb.

V zadnjem času pa se večina del s področja zaznavanja napak na podlagi usmerjenih grafov osredotoča na izboljšanje ločljivosti napak. Tako na primer Vedam in Venkatasubramanian [122] uporabita bazo znanj, ki je sestavljena iz podatkov o procesnih omejitvah, urniku vzdrževanja itd., za izboljšanje ločljivosti grafa SDG. Na podlagi

tega predlagana metoda zagotovi boljše ločevanje različnih napak. Močno pa se v ta namen uveljavlja tudi uporaba mehkih modelov. Shih in Lee [111],[112] sta podala izčrpen pregled, kako s kombinacijo principov teorije mehkih množic in grafov SDG izločamo napačne odločitve sistema za zaznavanje napak. V svojih delih sta uporabljala predvsem strategijo, kjer se dinamične poti skozi graf pretvarjajo v dinamične mehke relacije. Prehod na "mehki" graf pa omogoča eliminacijo napačnih odločitev, ki so predvsem posledica kompenzacij v sistemu (več neposrednih poti iz osnovnega do zelenega vozlišča z izključujočimi učinki) in povratnih zank. Druga možnost uporabe mehke logike pa je uvedba mehkega sklepanja, s čimer lahko natančneje določimo pot skozi graf od vzroka do posledice.

3.3.2 Drevo napak

Drevesa napak se uporabljajo pri analizi zanesljivosti in varnosti delovanja sistema. Prvi so jih razvili v laboratorijih Bell Telephone Laboratories v letu 1961 [124]. Drevo napak je logično drevo, ki ima v praksi dva strukturna tipa [62]. Prvi je ti. drevo tipa napaka-simptom (fault-symptom tree) [55], kjer so napake na vrhu drevesa, ostali dogodki pa tvorijo veje in liste drevesa. Dogodki v vozliščih (ang. nodes) so opisani s kvalitativnimi spremenljivkami in medsebojno (tipično) povezani z relacijami IN/ALI (ang. AND/OR), lahko pa tudi drugimi. Drugi tip je odločitveno drevo [13], ki začne pri simptomih in se razveja v napake v drevesnih listih.

Analiza drevesa napak poteka v splošnem v štirih korakih [38]:

- **definicija sistema**, kjer se predpostavlja, da pred gradnjo drevesa podrobno spoznamo sistem;
- **gradnja drevesa napak**, ki vključuje vprašanja o vzrokih za pojav obravnavanih napak, odgovori na vprašanja pa dajo informacijo o povezavah med napakami in primarnimi dogodki (ang. primary events), ki jih ni več mogoče razvejati;
- **kvalitativno ovrednotenje**, kjer iščemo preseke napak, ki so potrebni in zadostni pogoji za pojav odpovedi sistema;
- **kvantitativno ovrednotenje**, kjer določimo verjetnosti pojava primarnih dogodkov in računamo skupno verjetnost odpovedi pri danih pogojih.

Drevesa napak predstavljajo računsko orodje, ki omogoča uporabo logike za analizo napačnega delovanja sistema. V primerjavi z ostalimi metodami, kot npr. usmerjeni

grafi, se tu kot privlačna lastnost izpostavlja možnost uporabe različnih logičnih tipov vozlišč (OR, AND, XOR). Slednje nam omogoča zmanjševanje verjetnosti napačnih odločitev in pa zgoščeno predstavitev sistema. Največja težava dreves napak je ravno v tem, da je gradnja drevesa podvržena napakam v različnih fazah gradnje. Pri tradicionalnem pristopu drevo gradimo in spreminjamo s pomočjo tehnik pridobivanja znanja, kjer je največja težava vrednotenje dobljenega drevesa. Gradnja se začne z vprašanjem, kaj bi lahko povzročilo določen dogodek v vrhu drevesa. Možni odgovori na to vprašanje dajo dogodke, ki so z glavnim dogodkom povezane z logičnimi povezavami ali vejami. Drevo se z nadaljnjimi vprašanji in odgovori razveja, dokler ne pridemo do primarnih dogodkov. Hitro lahko uvidimo, da je kvaliteta takega drevesa odvisna predvsem od poznavanja modela, na podlagi katerega načrtovalec zgradi model na miselnem nivoju (vprašanja, odgovori). Zato velja, da če uporabimo samo hevristično znanje, ne moremo dokazati kompletnosti baze znanj. V zadnjem času se zato uporabljajo metode avtomatske gradnje dreves, kjer s kvalitativnimi modeli dosežemo bolj "oprijemljivo" predstavitev relacij med elementi drevesa. Delo [62] omenja hierarhične strukturne modele (ang. hierarchical structural models, [13]) in predznačene usmerjene grafe (ang. signed directed graphs). Dobra stran teh pristopov je, da simptome za napake predstavijo na kvalitativen način, slaba stran pa je potreba po pretvarjanju kvantitativnih veličin (meritev v sistemu) v kvalitativen opis.

3.4 Zaznavanje napak na podlagi zgodovine procesa

V nasprotju z metodami zaznavanja napak, ki temeljijo na modelu procesa in zahtevajo predhodno poznavanje procesa, je pri metodah na podlagi zgodovine procesa dovolj samo, da je na voljo dovolj velika količina podatkov o delovanju procesa v preteklem času. Obstaja več načinov, kako to množico podatkov transformiramo v obliko, primerno za zaznavanje napak, s skupnim imenom pa jih imenujemo iskanje značilnosti (ang. feature extraction). Metode se po tipu iskanja delijo na statistične in nestatistične. Od prvih si bomo pogledali metodo PCA, od drugih pa mehke modele in umetne nevronske mreže.

3.4.1 PCA

Analiza glavnih komponent (ang. principal component analysis, PCA) je linearna metoda, ki omogoča ortogonalno dekompozicijo kovariančne matrike \mathbf{X}_{cov} procesnih spremenljivk v smereh z maksimalno razpršenostjo podatkov. Glavna naloga metode PCA

je najti zapis sistema, ki je manjših dimenzij kot originalni zapis, a vseeno zadovoljivo natančno opiše glavne trende, ki jih je moč najti iz vzorčenih podatkov iz sistema. Zanimivo je, da je metodo PCA prvi uporabil Sylvester leta 1889 v psihologiji [91].

Matrika podatkov \mathbf{X} v svojih vrsticah vsebuje vzorčene vrednosti procesnih spremenljivk v določenih časovnih trenutkih, stolpci pa vzorčene časovne poteke posameznih procesnih spremenljivk, ki jih pred tem normiramo (srednja vrednost nič in varianca ena). Metoda PCA preslika prostor vhodnih podatkov \mathbf{X} v podprostor ortogonalnih podatkov, ki je definiran z množico lastnih vektorjev variančno-kovariančne matrike podatkov

$$\mathbf{X}_{cov} = \mathbf{X}^T \mathbf{X} = \mathbf{P} \mathbf{D} \mathbf{P}^T \quad (3.81)$$

Matrika preslikave \mathbf{P} je matrika lastnih vektorjev, ki jih uredimo po padajočih lastnih vrednostih (varianca vhodnih podatkov) v diagonalni matriki \mathbf{D} . Ortogonalne spremenljivke, imenovane zadetki, so torej notranji produkt merjenih spremenljivk in matrike preslikave, kar opišemo z enačbo

$$\mathbf{T} = \mathbf{X} \mathbf{P}, \quad (3.82)$$

kjer \mathbf{T} pomeni matriko zadetkov. Matrika glavnih komponent \mathbf{P} je statistični model procesa, zadetki pa so psevdospremenljivke, ki nimajo realnega ozadja, saj so kombinacije stolpcev matrike \mathbf{X} [59].

Če sestavimo matriko \mathbf{P}_K tako, da vzamemo le nekaj prvih stolpcev oziroma lastnih vektorjev matrike \mathbf{P} , ki pomenijo večino variance oziroma informacije vhodnih podatkov matrike \mathbf{X} , dobimo aproksimirano matriko podatkov

$$\mathbf{T}_a = \mathbf{X} \mathbf{P}_K, \mathbf{X}_a = \mathbf{T}_a \mathbf{P}_K^T. \quad (3.83)$$

Matriko podatkov lahko zapišemo z reduciranim prostorom in ostankom

$$\mathbf{X} = \mathbf{T}_a \mathbf{P}_K^T + \mathbf{E}_X, \quad (3.84)$$

kjer indeks K označuje prvih k izbranih stolpcev \mathbf{P} , imenovanih glavne komponente prostora zadetkov, \mathbf{E}_X pa označuje matriko ostankov oziroma residualov, imenovanih tudi prostor šuma. Ta pomeni vse odkrite linearne relacije med stolpci vhodne matrike \mathbf{X} . Ravno ta zadnja lastnost analize glavnih komponent je najpomembnejša, saj omogoča eliminacijo kolinearnih podatkov [69].

Za analizo dinamičnih sistemov je treba metodo PCA razširiti. Da se v prostoru šuma \mathbf{E}_X pojavijo tudi dinamične linearne relacije, ki opisujejo obravnavani proces, moramo vhodni matriki podatkov dodati še ustrezno število časovno zamaknjenih vhodnih

podatkov:

$$\mathbf{X}(l) = [\mathbf{X}_k, \mathbf{X}_{k-1}, \dots, \mathbf{X}_{k-l},] \quad (3.85)$$

Poiskati moramo torej linearne relacije med zamaknjenimi vhodnimi podatki \mathbf{X}_{k-i} . S primernim številom l (red sistema) zagotovimo, da se tako statične kot tudi dinamične komponente pojavijo v prostoru šuma. Za nelinearne sisteme lahko vrednost l preseže dejanski red sistema, saj tako dosežemo boljše prileganje statističnega modela izmerjenim podatkom opazovanega sistema. Postopek določanja l je podrobneje opisan v [69].

Za sprotni nadzor procesa je treba iz matrike podatkov izračunati skupno statistično oceno ali mero, ki jo nato opazujemo ter tudi grafično prikažemo. Skupna statistična ocena, ki definira odstopanje zadetkov opazovanega sistema od referenčnih zadetkov, se imenuje Mahalanobisova razdalja. Zapišemo jo z enačbo

$$h_i = (\mathbf{x}_i - \mathbf{m})\mathbf{S}^{-1}(\mathbf{x}_i - \mathbf{m})^T \quad (3.86)$$

kjer je \mathbf{x}_i i -ta vrstica podatkov v matriki podatkov \mathbf{X} , \mathbf{m} je vektor srednjih vrednosti referenčnih podatkov, N je število referenčnih podatkov in \mathbf{S} je kovariančna matrika

$$\mathbf{S} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i - \mathbf{m})^T (\mathbf{x}_i - \mathbf{m}) = \mathbf{PDP}^T. \quad (3.87)$$

Statistika h_i ima centralno χ^2 -porazdelitev z N prostostnih stopenj [125]. Ker je prostor zadetkov osrediščen in skaliran na varianco ena, se Mahalanobisova mera v prostoru zadetkov poenostavi v $h_i = \mathbf{t}_i \mathbf{D}^{-1} \mathbf{t}_i^T = \mathbf{t}_i \mathbf{t}_i^T$, kjer \mathbf{t}_i pomeni i -to vrstico zadetkov v matriki \mathbf{T} . Mera pomeni glavnino variance zadetkov.

Kadar je \mathbf{S}^{-1} nepoznana ali slabo poznana, uporabimo statistiko, ki se imenuje Hotellingova T^2 mera. Definiramo jo z enačbo

$$d_i = (\mathbf{x}_i - \mathbf{m}) (\mathbf{PDP}^T)^{-1} (\mathbf{x}_i - \mathbf{m})^T \frac{I(I-R)}{R(I^2-I)}, \quad (3.88)$$

kjer je R število glavnih komponent referenčnega modela in I število šarž modeliranih v referenčnem modelu. V prostoru zadetkov se mera poenostavi v

$$d_i = \mathbf{t}_i \mathbf{t}_i^T \frac{I(I-R)}{R(I^2-I)}. \quad (3.89)$$

Hotellingova in Mahalanobisova mera se pri poznani kovariančni matriki razlikujeta le v konstantnem faktorju.

Na opisani način se lahko metoda PCA uporabi za ugotavljanje (ne)pravilnosti delovanja procesa. Na podlagi statističnega nominalnega modela iz zgodovine procesa

lahko vsako novo meritev projiciramo v prostor zadetkov in primerjamo s statističnimi mejami iz nominalnega modela. Z uporabo statističnih mer je napake možno zaznati, ni pa jih mogoče lokalizirati. Statistična mera je omejena na formiranje zgolj enega residuala, kar je sicer največja hiba klasične metode PCA. To pa učinkovito odpravi metoda parcialna PCA (ang. partial PCA, PPCA), ki je osnovana na ideji paritetnih relacij in preko oblikovanja večjega števila residualov in deljenja kovariančne matrike na dva dela omogoča tudi lokalizacijo napak [91]. Gertler in McAvoy sta leta 1997 [43] pokazala (teoretično) neposredno povezavo med residuali na podlagi paritetnih enačb in metode PCA, vendar je ta primerjava v praksi vprašljiva zaradi vstopa motenj na različnih mestih.

Uporaba metode PCA v zaznavanju napak je trdno usidrana tako v strokovni literaturi kot tudi v inženirski praksi, naj omenim samo knjigo [8], tuja dela [85] (šaržni procesi), [26] (nelinearna PCA), [27] (sekvenčna rekonstrukcija), [74] (rekurzivna PCA) ter domača dela [69], [91] in [132] (adaptivna PCA).

3.4.2 Statistične metode

V realnih procesih imamo vedno opravka z naključnimi motnjami. Nasprotno od determinističnih sistemov pri naključnih sistemih ne moremo natančno napovedati obnašanja v prihodnosti, zato moramo preiti v okvir verjetnostne obravnave sistemov. Za sistem v normalnem načinu obnašanja velja, da imajo meritve vedno določeno verjetnostno porazdelitev. Ob nastopu napak se te porazdelitve do določene mere spremenijo. Na primer, v parametričnem zapisu ima vsak parameter procesa v normalnem načinu delovanja določeno srednjo vrednost in standardno deviacijo. Ob pojavu napake se na ustreznem parametru spremeni ali srednja vrednost ali standardna deviacija ali pa oboje hkrati. Zato lahko zaznavanje napak v tem okviru predstavimo kot zaznavanje sprememb statističnih parametrov, kot je bilo to prikazano v delu [8].

Bayesov način klasifikacije je osnovan na računanju verjetnosti, da se v sistemu zgodi določen dogodek. Prvi način uporabe v odločanju pri lokalizaciji napak je ti. Bayesov test hipoteze (ang. Bayesian hypothesis test). Klasifikator na podlagi Gaussove funkcije gostote porazdelitve za problem dveh razredov je v tem primeru

$$\begin{aligned} d_1 &= (\mathbf{y} - \mathbf{m}_1)^T \mathbf{S}_1^{-1} (\mathbf{y} - \mathbf{m}_1), \\ d_2 &= (\mathbf{y} - \mathbf{m}_2)^T \mathbf{S}_2^{-1} (\mathbf{y} - \mathbf{m}_2). \end{aligned} \tag{3.90}$$

Diskriminanta h preslika odločitveni prostor v prostor razredov (če privzamemo, da sta

verjetnosti za oba razreda enaki):

$$\begin{aligned}
 h &= d_1 - d_2 \\
 \delta &= \log \left(\frac{|\mathbf{S}_2|}{|\mathbf{S}_1|} \right), \text{ prag klasifikatorja oz. Bayesov faktor,} \\
 h < \delta, x &\text{ pripada razredu 1,} \\
 h > \delta, x &\text{ pripada razredu 2.}
 \end{aligned} \tag{3.91}$$

Enaka ideja se lahko razširi na problem klasifikacije v n razredov. Bayesov klasifikator je optimalni klasifikator (zagotavlja minimalni pogrešek, tudi če se razredi prekrivajo) za razrede z Gaussovo porazdelitvijo in če je informacija o porazdelitvi na voljo. Opoznimo, da je mera (3.90) v normirani obliki. Za $\mathbf{S}_1 = \mathbf{S}_2 = \mathbf{I}$ se Gaussov klasifikator spremeni v klasifikator na podlagi evklidske distance.

Pri statističnih metodah v stohastičnem okviru obravnave se vedno poraja vprašanje o robustnosti metod na motnje in nemodelirano dinamiko. Pogosta rešitev je modeliranje vplivov neznanih vhodov v sistem s statističnimi parametri. Da bi se izognili težavam zaradi prevelikega števila procesnih parametrov, mora biti statistična razširitev modela čim manjša. V delu [45] je bilo pokazano, da se vplivi neznanih vhodov lahko zadovoljivo opišejo z dvema parametroma, ki ju lahko identificiramo iz procesnih podatkov. Kwon s sodelavci [73] je predlagal metodo zaznavanja napak, ki na statističen način ocenjuje in vrednoti vplive motenj, nemodelirane dinamike in linearizacije s pomočjo parametrične identifikacije. Čeprav doseže povečanje robustnosti, je slaba stran metode, da mora biti proces zaradi parametrične identifikacije stalno vzburjan. Izboljšava je bila podana v delu [61], kjer avtorja za statistični model vpliva motenj predlagata statistični test, podoben testu GLR (ang. generalized likelihood ratio) [8]. S pomočjo primerjave utežene vsote residualov s konstantnim pragom lahko zaznavamo tudi počasi razvijajoče se napake na senzorjih ob dobršem zmanjšanju lažnih alarmov.

3.4.3 Umetne nevronske mreže

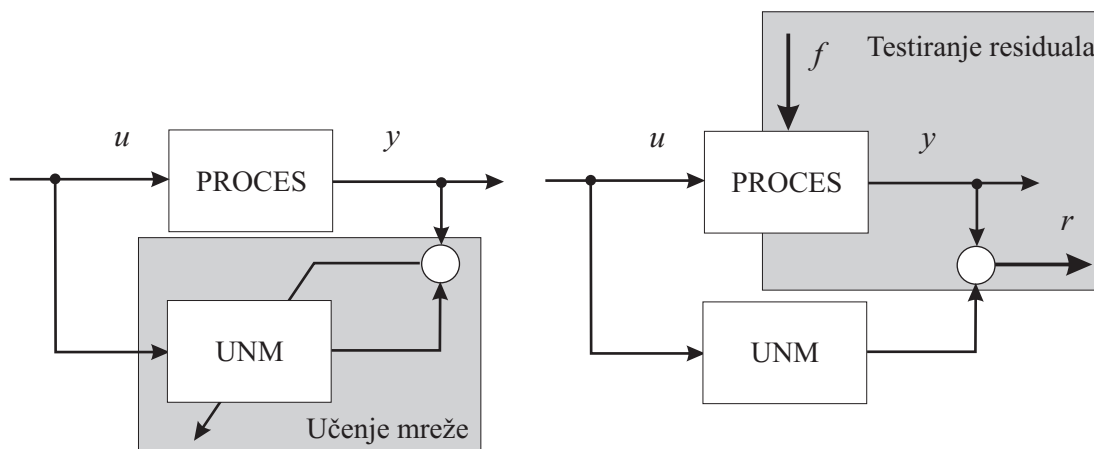
Umetne nevronske mreže ali UNM (ang. artificial neural networks, ANN) so se v zaznavanju napak pojavile v poznih osemdesetih letih prejšnjega stoletja. V devetdesetih je sledila široka uporaba v zaznavanju in lokalizaciji napak za počasi spreminjajoče se kompleksne sisteme, kjer je matematični model slabo poznan ali pa celo ni na voljo [50], [35].

Umetna nevronska mreža je sestavljena iz nevronov, enostavnih procesnih elementov, ki se aktivirajo, kadar njihovi vhodi presežejo določen aktivacijski prag. Nevroni

so nato razporejeni v nivoje in tvorijo urejeno mrežo povezav, skozi katero signali na vhodih prehajajo na svoji poti proti izhodom. Izbira prenosnih funkcij za vsak posamezni nevron in s tem povezanih uteži prispeva k celotnemu nelinearnemu obnašanju mreže, ki tako postane univerzalni aproksimator nelinearnosti v modelih procesa. Med fazo učenja s pomočjo optimizacije računamo vrednosti določenega nabora parametrov mreže, tako da dobimo optimalno aproksimacijo želenega obnašanja sistema, ki je predstavljeno z naborom vhodno-izhodnih vzorčenih podatkov iz sistema. Za zaznavanje napak faza učenja poteka na podatkih, pridobljenih v normalnem načinu delovanja procesa (brez prisotnosti napak), če pa je možno, pa tudi pri pojavu določenih poznanih napak.

Že ob prvi uporabi nevronske mreže v diagnostiki napak [51], [115] se je pokazal velik potencial predlaganega pristopa v teoretičnem in praktičnem smislu, še posebej zaradi uporabe meritev iz procesa. Po drugi strani pa v praktičnem smislu mnogo problemov, še posebej glede aproksimacije dinamike procesa in oblikovanja residualov, ostaja nerešenih in bodo predmet obravnave bodočih raziskav.

Uporabo UNM v diagnostiki napak razdelimo na dva dela: **oblikovanje residualov** in **vrednotenje residualov**. Pri oblikovanju residualov gre dejansko za pridobivanje analitičnega modela procesa, ki opisuje delovanje procesa v normalnem načinu delovanja [35]. Najprej moramo pridobiti učno množico vhodno-izhodnih podatkov, neposredno iz procesa ali pa ob pomoči realističnega simulacijskega modela. Zadnje je še posebej uporabno, kadar želimo obravnavati pojave napak v sistemu v fazi testiranja residuala. Nato sledi učenje mreže, oblikovanje in testiranje residuala, kot je to prikazano na shemah na sliki 3.8. Po koncu učenja in testiranja je mreža pripravljena za sprotno računanje residuala.

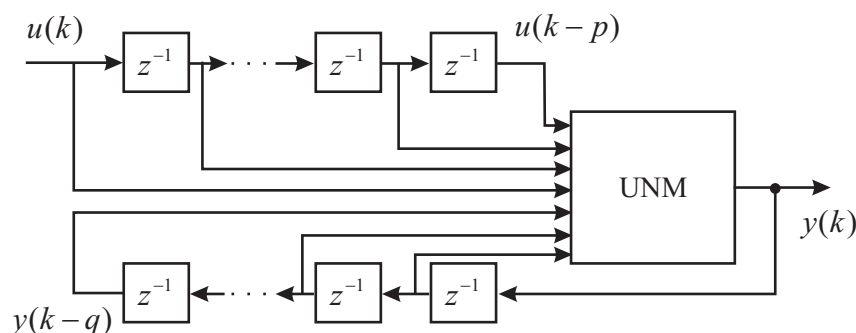


Slika 3.8: Oblikovanje residuala s pomočjo UNM

Če UNM uporabljamo kot dinamični model procesa, strukturo mreže določa izbrana struktura modela (vhodno-izhodna ali prostor stanj), ki najbolje opiše dinamiko procesa. Pri diagnostiki napak se je za boljšo strukturo izkazala vhodno-izhodna oblika, ker za predstavitev splošne nelinearne funkcije procesa q -tega reda

$$y(k) = g(y(k-1), \dots, y(k-q), u(k), \dots, u(k-p)) \quad (3.92)$$

potrebujemo samo eno nevronske mreže. Zaradi tega dejstva se izboljšajo konvergenčne in aproksimacijske lastnosti mreže, poleg tega pa je neposredno uporabna tudi za aproksimacijo sistemov s hitro dinamiko, kot so npr. motorji [35]. Shema strukture je predstavljena na sliki 3.9. Praktični problemi pri tej izbiri strukture pa so predvsem

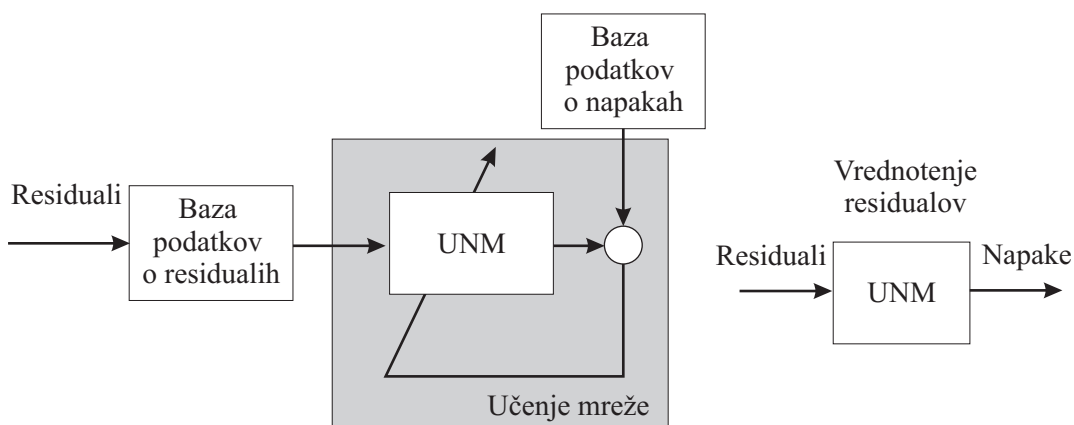


Slika 3.9: Aproksimacija dinamičnega sistema z UNM

- izbira zakasnitev vhoda in izhoda pri učenju mreže (dinamika procesa);
- izbira vhodov, če imamo več izhodov - vse napake ne delujejo na vse izhode;
- nezadostna količina podatkov za zadovoljivo aproksimacijo dinamike procesa - dinamična in statična odstopanja, lažni alarmi, slaba občutljivost glede na delovno točko

Omenjene probleme navadno rešujemo s poskušanjem (ang. trial-and-error), ker bi moralo v drugem primeru biti na voljo dovolj *á priori* znanja o procesu.

Pri vrednotenju residualov na vhode UNM pripeljemo residuele, ki jih generira model procesa na podlagi druge UNM, lahko pa tudi po katerikoli od že omenjenih metod v tem poglavju. Učenje UNM za vrednotenje poteka na enak način kot pri fazi modeliranja procesa, vendar moramo tu imeti na voljo podatke o residualih (vhodi) in odločitvah o napakah (izhodi). Po koncu učenja je UNM na voljo za sprotno ocenjevanje, ali je prišlo do napake, in morebitno lokalizacijo vzroka za napako. Shema vrednotenja je predstavljena na sliki 3.10.

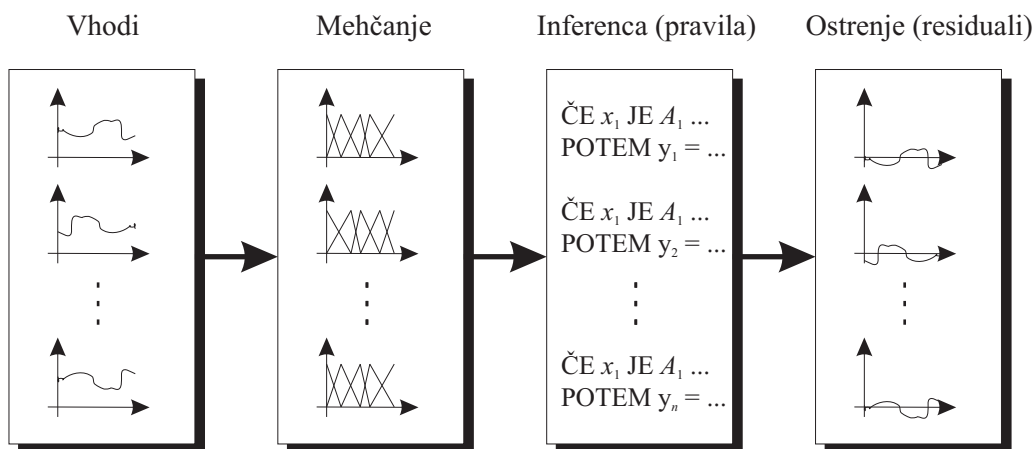


Slika 3.10: Uporaba UNM za vrednotenje residualov

Največja težava pri uporabi UNM v zaznavanju napak je pomanjkanje dokazov o stabilnosti in robustnosti. Zato dobršen del trenutnih raziskav poteka v smeri sprotnih (ang. on-line) aproksimatorjev in stabilnih algoritmov učenja s pomočjo adaptivnih tehnik. Kot primer navedimo dela Polycarpoua in sodelavcev [93], [119]. Robustnost metod UNM pa je bila obravnavana tudi v domači literaturi [68].

3.4.4 Mehki modeli

Enako kot pri umetnih nevronskih mrežah se mehki model največkrat uporablja kot univerzalni aproksimator dinamike procesa. Na podlagi izhodov mehkega modela lahko oblikujemo residuele in na ta način izvedemo zaznavanje napak za nelinearne sisteme. Slika 3.11 prikazuje faze pri računanju residualov na podlagi mehke logike.



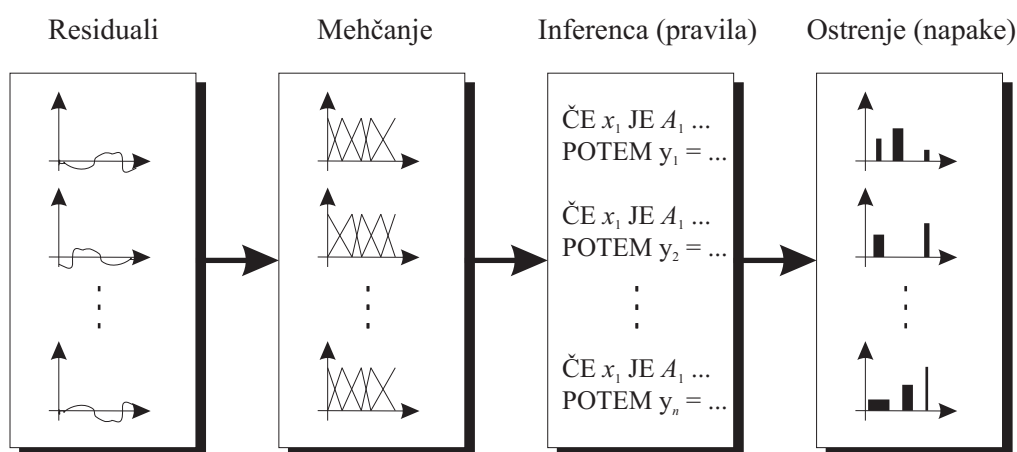
Slika 3.11: Uporaba mehkega modela za oblikovanje residualov

Tako kot pri metodah z UNM (slika 3.8) zaznavanje napak razdelimo na

- optimizacijo pravil in pripadnostnih funkcij modela na podlagi podatkov;
- sprotno računanje residualov ob uporabi izhodov modela.

Poglavitna razlika med pristopoma pa je, da pri mehkih sistemih lahko bolj transparentno predstavimo relacije med vhodi in izhodi v smislu lokalnih linearnih modelov (model tipa Takagi-Sugeno). Eden vidnejših pristopov v tem okvirju je zaznavanje napak na podlagi mehkega funkcionalnega modela (MFM, ang. fuzzy functional model). MFM aproksimira nelinearni dinamiko procesa z odsekoma linearnimi modeli [84], [37]. Vsak lokalni model je veljaven samo v določeni regiji vhodnega prostora. Izhod modela se izračuna kot utežena vsota prispevkov vseh lokalnih modelov. Model je tako ekvivalenten mehkeemu modelu Takagi-Sugeno. Mehki funkcionalni observator se načrta po večmodelnem pristopu, kjer vhodni vektorji v vsak podmodel vsebujejo različne nabore trenutnih in preteklih vrednosti procesnih vhodov in izhodov. Residualne uporabimo za računanje množice strukturiranih simptomov, tako da je vsak residual odvisen od različnega nabora vhodov v proces in procesnih komponent. Vzorec, kjer določeni residuali presežejo svoje mejne vrednosti, nekateri pa ne, nam pove lokacijo napake v sistemu.

Pri metodah na podlagi mehkih modelov pa je pomembno tudi mehko sklepanje o napaki z analizo residualov. Dejansko gre pri vrednotenju residualov za mehanizem logičnega odločanja oziroma za problem klasifikacije. Shemo prikazuje slika 3.12. Residualom najprej priredimo pripadnostne funkcije glede na njihove mejne vrednosti.



Slika 3.12: Uporaba mehkega modela za lokalizacijo napake

Na ta način izboljšamo občutljivost residualov. Pri uporabi ostrih vrednosti se pojavi

težava, če je vrednost residuala blizu mejne vrednosti, saj že mala sprememba povzroči veliko spremembo odločitvene funkcije. Zato moramo v tipičnem primeru dvigniti vrednosti meja, kar zmanjša občutljivost. Pri mehčanem residualu za to ni potrebe, ker je pri majhnih spremembah residuala možnost za napačno odločitev veliko manjša. Pri inferenci povežemo vzroke (spremembe residualov) s posledicami (napake), tako da napišemo množico pravil če-potem. Pri ostrenju pa glede na vrednosti residualov preko enega od znanih algoritmov ostrenja izračunamo številske vrednosti, ki nam tipično povejo kode napake, ki so se zgodile.

3.5 Lokalizacija napak

Bistvo lokalizacije je ugotoviti vzrok, da se je zaznana napaka zgodila, in lokacijo vzroka v sistemu. Lokalizacija napak poteka na podlagi vrednotenja simptomov in povezovanja le-teh z upoštevanjem medsebojnih odvisnosti s pomočjo logičnega sklepanja. Lokalizacijske sposobnosti diagnostičnih sistemov so v veliki meri odvisne od razpoložljivosti meritev iz procesa in nivojem natančnosti pri vrednotenju simptomov. S povečanjem natančnosti lahko znatno izboljšamo ločljivost napak, in to brez potrebe po dodatnih merilnih točkah v sistemu. Vendar pa večja natančnost zahteva tudi več truda pri modeliranju procesa.

Ogledali si bomo tri metode sklepanja o napakah, Boolovo sklepanje in dve metodi aproksimativnega sklepanja. Ker se Boolovo sklepanje zaradi svojih neugodnih lastnosti uporablja samo pri teoretičnem načrtovanju sistemov za lokalizacijo napak, bomo večji poudarek dali aproksimativnim metodam. Glavna razlika med slednjimi je v načinu integriranja dodatnega znanja o procesu. V delih [103] in [102] Rakar omenja nekatere napredne pristope, kot npr. mehka logika, konfirmacijska teorija in njeni izpeljanki DMP (ang. Diagnostic Model Processor) in DMA (Deep Model Algorithm) ter metoda prenosa zaupanja (ang. Transferable Belief Model, TBM). Nekateri od njih tudi omogočajo obravnavanje nekonsistence v podatkih, ki se potem odraža kot mera za zanesljivost in verodostojnost diagnostičnih rezultatov.

3.5.1 Boolovo sklepanje

Boolovo sklepanje je najosnovnejša metoda lokalizacije napak. Potreben pogoj za lokalizacijo je, da obstajajo neposredne in enolične povezave med odločitvenimi funkcijami in napakami. Če je ta pogoj izpolnjen, residuele strukturiramo tako, da določena napaka povzroči značilen odziv residualov, ki je definiran z enoumno kodo

$F_j = [f_{1j}, f_{2j}, \dots, f_{kj}]$. Na podlagi kod zgradimo tabelo kod napak ali ti. *incidenčno matriko*

$$\Lambda = \begin{bmatrix} f_{11} & \cdots & f_{1m} \\ \vdots & \ddots & \vdots \\ f_{k1} & \cdots & f_{km} \end{bmatrix}, f_{ij} = \begin{cases} 1, & \text{če } j\text{-ta napaka vpliva na } i\text{-ti residual,} \\ 0, & \text{če } j\text{-ta napaka ne vpliva na } i\text{-ti residual.} \end{cases} \quad (3.93)$$

Binarni nizi v stolpcih matrike Λ so kombinacije odločitev, ki pomenijo pripadajoče kode napak. Glede na stolpce tudi ločimo

- neločljive množice, kjer obstajajo enake kode v dveh ali več stolpcih;
- slabo ločljive množice, kjer so stolpci med seboj različni, vendar s spremembo ene same vrednosti kode dobimo kodo drugega stolpca;
- strogo ločljive množice, kjer je razlika med vsemi stolpci najmanj v dveh kodah.

Za lokalizacijo v sprotne delovanju oblikujemo odločitvene vektorje in jih primerjamo s stolpci incidenčne matrike.

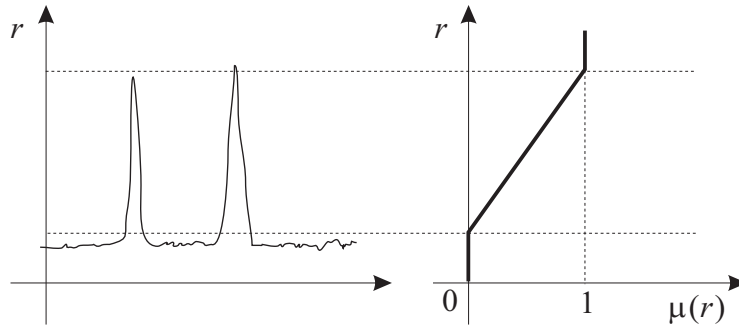
Glavna slabost je v tem, da lahko pri vrednotenju simptoma privzamemo le dve vrednosti: prisoten (logična 1) ali ne prisoten (logična 0). Zato lahko, kot smo omenili že v razdelku 3.4.4, že majhne spremembe v signalih zaradi šuma, odstopanj modela in drugih motenj vodijo do velikih sprememb v končni diagnozi. Ta pojav imenujemo diagnostična nestabilnost [102]. Diagnostični nestabilnosti se izognemo s preoblikovanjem odločitvene funkcije v zvezno ali odsekoma zvezno funkcijo.

3.5.2 Aproksimativno sklepanje

Z uporabo aproksimativnih metod sklepanja simptome ovrednotimo z merami prisotnosti v obliki števila iz intervala $[0, 1]$. Sklepanje torej temelji na mehkih odločitvenih funkcijah. Slika 3.13 prikazuje, kako mala sprememba residuala preko zvezne odločitvene funkcije povzroči samo malo spremembo mere za odločitev.

Na tak način dajo majhne spremembe v signalih tudi majhne spremembe v diagnostičnih rezultatih. Bistvo aproksimativnih metod sklepanja je, da nam aproksimativno sklepanje da urejeni seznam osumljenih napak s pripisanimi merami zaupanja, na podlagi katerih nato pridemo do robustnejše končne odločitve o dejanski napaki.

V nadaljevanju bomo na kratko opisali metodo TBM. Podrobnejša primerjava metod aproksimativnega sklepanja je v delu [102].



Slika 3.13: Mehka odločitvena funkcija

Metoda prenosa zaupanja (TBM)

Metoda prenosa zaupanja (ang. transferable belief model, TBM) je zasnovana kot razširitev Dempster-Shaferjeve matematične teorije dejstev [102]. Njena posebnost je vpeljava t.i. modela odprtega sveta [114]. V nasprotju z drugimi pristopi, poleg mogočih in nemogočih dogodkov, predvideva tudi obstoj neznanih dogodkov. Nastop neznane napake tako ne povzroči zavajajočih diagnostičnih rezultatov.

Sklepanje poteka v dveh korakih. Najprej določimo uteži osnovnega zaupanja m za podmnožici

$$A_i = \left\{ \bigvee_{\lambda_{ij} \neq 0} f_j \right\}, B_i = \left\{ \bigvee_{\lambda_{ij} = 0} f_j \vee f_0 \right\}, i = 1, 2, \dots, K, j = 1, 2, \dots, M \quad (3.94)$$

ki sta med seboj komplementarni ($m(B_i) = 1 - m(A_i)$). Mere osnovnega zaupanja določimo kot gladke funkcije residualov. Osnovna oblika je sigmoidna funkcija, definirana kot

$$m_i(A_i) = \frac{1}{1 + \frac{1-a}{a} \left(\frac{\tau_i}{r_i} \right)^{2\gamma}}, \quad (3.95)$$

kjer je a zaupanje, prirejeno mejni vrednosti τ_i , in γ prilagodljiva stopnja glajenja. V naslednjem koraku določimo mere zaupanja v posamezne napake in stanje brez napak z uporabo nenormiranega Dempsterjevega pravila kombiniranja [101]. V danem okviru, ko so residuali edini vir dejstev, se pravilo poenostavi v

$$m(f_j) = \prod_{\substack{i=1 \\ f_j \in A_i}}^K m_i(A_i) \prod_{\substack{i=1 \\ f_j \in B_i}}^K m_i(B_i). \quad (3.96)$$

Preostanek zaupanja pripišemo prazni množici:

$$m(\phi) = SC = 1 - \sum_{j=0}^M m(f_j) \quad (3.97)$$

in ga imenujemo moč konflikta (ang. strength of conflict, SC). To mero si lahko razlagamo kot nepripisano zaupanje, ki je posledica napake modela, šuma ali neznane (nepredvidene) napake. Služi nam lahko kot mera zanesljivosti diagnostičnih rezultatov, kar je pomembna prednost te teorije. Diagnostične rezultate na koncu predstavimo kot urejeni seznam mogočih napak glede na dane mere zaupanja. Pomembno je poudariti, da stanja napak niso nujno posamezne napake, temveč so lahko tudi množice več napak. Z morebitnimi novimi dejstvi, ki omogočajo boljše ločevanje napak, lahko tako pride do prenosa zaupanja na bolj natančno stanje napake. Od tod tudi ime metoda prenosa zaupanja.

4. Prediktivno vodenje na podlagi mehkih modelov

Prediktivno vodenje na podlagi modela (ang. model-based predictive control, MBPC) je v zadnjih štirih desetletjih postalo pomembna veja raziskav v teoriji vodenja in se je, v nasprotju z nekaterimi drugimi metodami, uveljavilo tudi v praktični uporabi v industrijskem okolju [105]. V delu [99] je bilo denimo navedeno 2200 industrijskih aplikacij predvsem na področjih procesnega vodenja (rafinerije, petrokemična in kemična industrija), avtomobilske, vojaške industrije, metalurgije ipd. Poglavitne lastnosti, ki dajejo metodam MBPC prednost pred ostalimi v smislu industrijskih aplikacij [65], so:

- primernost za vodenje procesov z zahtevno dinamiko, mrtvim časom in neminimalno fazo;
- sposobnost obravnavanja uni- in multivariabilnih sistemov;
- predkompenzacija (ang. feedforward) merljivih motenj;
- upoštevanje omejitev reguliranih in reguliranih veličin ter njihovih gradientov;
- veliko prostostnih stopenj pri načrtovanju;
- ni eksplicitnega odvajanja in integriranja;
- razumljivost tudi za neeksperte in s tem povezan kompromis med kvaliteto vodenja ter enostavnostjo izvedbe.

Po drugi strani pa ne smemo niti mimo dveh največjih omejitev:

- računsko zahtevnost metod, ki povzročajo probleme pri procesih s hitro dinamiko;
- potreba po zelo natančnem dinamičnem modelu procesa; nekateri viri navajajo podatek, da gradnja modela predstavlja kar 80% vsega časa, potrebnega za razvoj aplikacije vodenja [94], [65].

V začetni fazi razvoja, tj. v osemdesetih letih prejšnjega stoletja, so metode v grobem izhajale iz dveh pristopov:

- vodenje z dinamično matriko (ang. dynamic matrix control, DMC, [18]), ki temelji na neparametričnem modelu in je bilo razvito predvsem zaradi potreb naftne in kemične industrije;
- posplošeno prediktivno vodenje (ang. generalized predictive control, GPC, [17]), ki temelji na parametričnem modelu in je nadgradnja posplošenega minimalno-variančnega regulatorja.

Morari v delu [83] za področje kemične in petrokemične industrije navaja, da je v praktičnih izvedbah največkrat uporabljena ena od izvedenk metode DMC. Razlog je v tem, da GPC ni primeren za multivariabilne procese z omejitvami. Obe metodi sta opisani in kritično ovrednoteni v [65]. V zadnjem času pa so bolj uveljavljene formulacije prediktivnega vodenja v prostoru stanj, ker omogočajo obravnavo optimalnega sledenja stanj procesa. Tipičen primer industrijske izvedbe je metoda prediktivnega funkcionalnega vodenja (ang. predictive functional control, PFC) [105], [131].

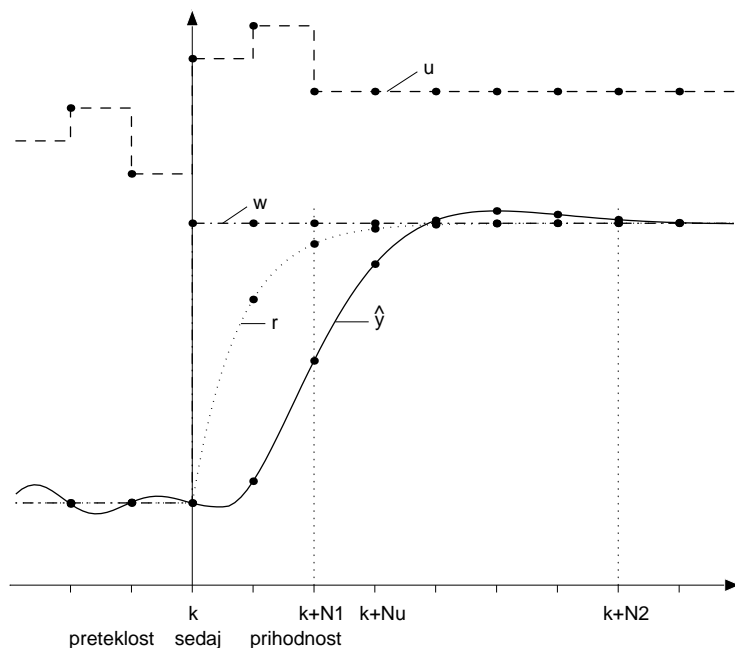
4.1 Osnovni elementi prediktivnega vodenja na podlagi modela

Prediktivno vodenje na podlagi modela je osnovano na treh ključnih konceptih [40]:

- eksplicitna uporaba modela za napoved izhoda procesa v prihodnosti;
- izračun sekvence prihodnjih regulirnih signalov preko minimizacije dane kriterijske funkcije;
- uporaba strategije pomičnega horizonta; prvi regulirni signal v sekvenci se uporabi kot trenutni vhod v proces, horizont napovedi se premakne za en korak naprej in izračun sekvence se ponovi.

Glavni koraki metodologije, ki so orisani na sliki 4.1, so naslednji [65]:

- Ob vsakem diskretnem časovnem trenutku k izračunamo **napoved izhodnega signala procesa** $y(k + j)$ za velik korak opazovanja v prihodnosti ($j = N_1, \dots, N_2$). Napovedane vrednosti označimo kot $\hat{y}(k + j|k)$ in predstavljajo j -koračno napoved modela, vrednosti N_1 in N_2 pa označujeta **spodnjo** in **zgor-njo vrednost predikcijskega horizonta**. Spodnja in zgornja vrednost predikcijskega horizonta izhodnega signala določata **horizont ujemanja** (ang. *coincidence horizon*), znotraj katerega želimo doseči ujemanje izhodnega signala s predpisanim obnašanjem. Napoved izhodnega signala izračunamo s pomočjo



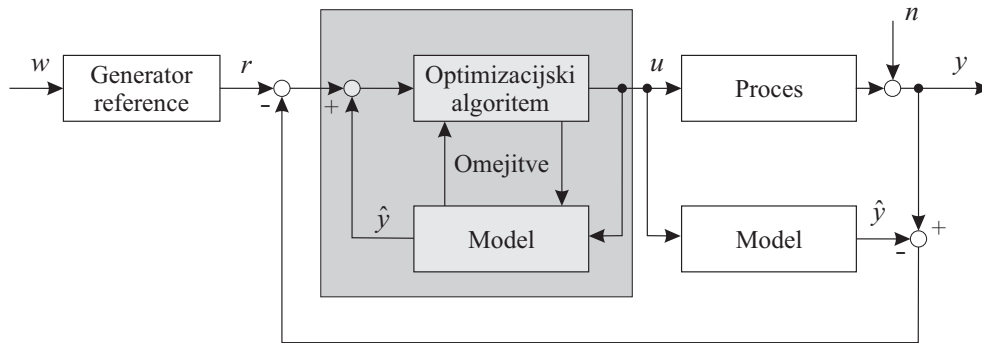
Slika 4.1: Osnovni princip prediktivnega vodenja

modela procesa. Napovedane vrednosti pa so seveda odvisne tudi od regulirnega scenarija v prihodnosti $u(k+j|k), j=0, \dots, N_u$, ki ga nameravamo uporabiti od trenutka k naprej.

- Definiramo želeno **modelno-referenčno trajektorijo** (ang. *reference trajectory*) $r(k+j|k), j=N_1, \dots, N_2$, s katero opišemo, kako naj se odziv procesa giblje od trenutne vrednosti $y(k)$ do predpisane referenčne vrednosti $w(k)$ (ang. *setpoint trajectory*).
- Sekvenco regulirnih signalov za prihodnost ($u(k+j|k), j=0, \dots, N_u$) izračunamo z **minimizacijo** ustrezne **kriterijske funkcije** tako, da minimiziramo napovedano napako med $r(k+j|k)$ in $\hat{y}(k+j|k), j=N_1, \dots, N_2$.
- Za vodenje uporabimo samo prvi element $u(k|k)$ **optimalnega** vektorja regulirnih signalov $u(k+j|k), j=0, \dots, N_u$.

V naslednjem diskretnem trenutku imamo na voljo nov merjeni izhod procesa $y(k+1)$ in celoten postopek, opisan v teh štirih točkah, se ponovi. Ta princip se imenuje **strategija pomičnega horizonta** (ang. *receding horizon strategy*).

Osnovno shemo zaprtozančnega sistema vodenja si lahko ogledamo na sliki 4.2.



Slika 4.2: Shema prediktivnega vodenja

Pri opisani strategiji lahko izluščimo naslednje pomembne elemente prediktivnega vodenja:

- napoved oziroma predikcija na osnovi modela procesa;
- oblikovanje (računanje) modelno-referenčne trajektorije;
- strukturiranje regulirne akcije;
- algoritimizirano računanje najboljšega regulirnega scenarija.

Za vsakega od naštetih elementov lahko najdemo različne rešitve in po tem se posamezne metode prediktivnega vodenja razlikujejo med seboj.

4.1.1 Model procesa

Naloga modela je natančno napovedati izhode in/ali stanja procesa za določen časovni horizont v prihodnosti, pri tem pa ni toliko pomembno, kakšen tip modela uporabimo - v obliki črne, sive ali bele škatle [113]. Prediktivno vodenje je ena redkih strategij vodenja, ki jo lahko neposredno uporabimo za nelinearne sisteme. Model v shemi na sliki 4.2 je lahko linearen ali nelinearen; s časovno nespremenljivimi parametri, s spremenljivim ojačenjem ali v celoti spremenljivimi parametri; v obliki prenosne funkcije, konvolucijski obliki ali v prostoru stanj; umetna nevronska mreža, mehki model, Wienerjev ali Hammersteinov model ipd. Poleg tega nekatere metode, kot je prediktivno vodenje z notranjim modelom (ang. internal model control, IMC, Morari:89), ločijo model procesa na dva aditivna podmodela:

- model procesa, ki opisuje relacije med merjenimi vhodi in merjenimi izhodi procesa;

- model motenj, s katerim opisujemo prispevek merjenega izhoda procesa, ki ni zajet v modelu procesa.

4.1.2 Kriterijska funkcija

Sekvenco prihodnjih regulirnih signalov $u(k+j|k)$, $j = 0, \dots, N_u$ dobimo s sprotno minimizacijo dane kriterijske funkcije, ki definira optimalno sledenje napovedi izhoda referenčni trajektoriji (podani z referenčnimi točkami ali krivuljami). Navadno se uporablja razlika med izhodom in referenčno trajektorijo, kombinirano vsoto regulirnih akcij. Splošna matematična formulacija nelinearnega prediktivnega vodenja je torej iskanje minimuma kvadratične kriterijske funkcije

$$J = \sum_{j=N_1}^{N_2} \|\hat{y}(k+j|k) - r(k+j|k)\|^2 + \beta \sum_{j=0}^{N_u-1} \|\Delta u(k+j|k)\|^2, \quad (4.1)$$

ob predpostavki, da so vse spremembe regulirnega signala enake 0 za $j \geq N_u$. Prvi člen predstavlja odstopanje med napovedjo izhoda procesa in oblikovano referenčno trajektorijo, medtem ko drugi člen, utežen s faktorjem $\beta > 0$, pomeni zmanjšanje regulirne akcije. Drugi člen lahko predstavimo tudi s samim regulirnim signalom $u(k+j)$, odvisno od problema in zasnove sistema. Z utežnim faktorjem dosežemo kompromis med načelno nasprotujočima si členoma. Začetna vrednost horizonta N_1 mora biti različna od nič za procese z relativnim redom $\rho > 1$, za fazno-neminimalne procese in procese z zakasnitvijo, ker v tem primeru vhodni signal v časovnih trenutkih $k < N_1$ nima neposrednega vpliva na izhod.

Pogosto iščemo minimum kriterijske funkcije (4.1) ob upoštevanju naslednjih omejitev vhoda in izhoda:

- omejitev izhodnega signala procesa $\underline{y} \leq y(k+j|k) \leq \bar{y}$, $j = N_1, \dots, N_2$;
- omejitev vhodnega signala procesa $\underline{u} \leq u(k+j|k) \leq \bar{u}$, $j = 1, \dots, N_u$;
- omejitev spremembe vhodnega signala $\underline{\Delta u} \leq \Delta u(k+j|k) \leq \bar{\Delta u}$, $j = 0, \dots, N_u-1$.

Ostale omejitve, kot npr. omejitve procesnih stanj za modele v prostoru stanj, lahko podamo na enak način.

4.1.3 Optimizacijski problem

Kombinacija modela in kriterijske funkcije definira optimizacijski problem. Iz literature [40],[14],[83],[49] je dobro znano dejstvo, da pri reševanju enačbe (4.1) dobimo

- eksplicitno analitično rešitev, če imamo linearen časovno nespremenljiv model in nimamo omejitev;
- rešitev kvadratičnega optimizacijskega problema (ang. quadratic programming, QP), če so prisotne omejitve; problem QP je v splošnem dobro numerično pogojen;
- rešitev nekonveksnega optimizacijskega problema, če uporabljamo nelinearni model.

Nekonveksna optimizacija je računsko zelo potratna, metode, kot npr. Nelder-Meadova metoda ali sekvenčno kvadratično programiranje, pa so podvržene tudi k temu, da lahko obstanejo v lokalnih minimumih kriterijske funkcije. To pa v primeru hitrih sistemov, kjer moramo poiskati globalni minimum enačbe (4.1) v zelo kratkem času, otežuje izvedbo v realnem času.

4.2 Uporaba mehkega modela v prediktivnem vodenju

Prediktivno vodenje na podlagi mehkega modela (ang. fuzzy-model-based predictive control, FMBPC) se je tako kot ostale metode nelinearnega prediktivnega vodenja začelo razvijati v osemdesetih letih prejšnjega stoletja. Prva omemba FMBPC je delo Grahama in Newella [46], v katerem sta avtorja uporabila mehki relacijski model, vgrajen v prediktivno shemo vodenja. Kasneje so se začele uveljavljati tudi druge oblike mehkih modelov, po drugi strani pa se je spreminjal tudi cilj uporabe mehke logike v sami strategiji prediktivnega vodenja. Zaradi različnih struj razvoja je danes oznaka FMBPC v literaturi sinonim za tri med seboj dokaj različne pristope:

- mehki model uporabimo kot prediktor v klasični shemi MBPC, predstavljeni na sliki 4.2 [75],[1];
- kriterijska funkcija in/ali omejitve so zapisane z mehкими relacijami [67],[19];
- optimizator, vključujoč strategijo vodenja, je zasnovan na mehkih pravilih [78].

Daleč najbolj razširjena je seveda prva oblika, ki izkorišča mehki model kot prediktor odziva nelinearnega dinamičnega sistema.

4.2.1 Prediktivno vodenje na podlagi mehkega modela Takagi-Sugeno

Kot smo že omenili, naletimo pri optimizaciji na podlagi nelinearnega modela na nekonveksen optimizacijski problem, za katerega v splošnem ni na voljo zanesljivih in dovolj

hitrih postopkov. V izogib težavam s sprotno optimizacijo se je pri prediktivnem vodenju na podlagi modela Takagi-Sugeno pojavilo mnogo poskusov poenostavljanja samega algoritma nelinearnega MPC. V splošnem jih lahko razdelimo na

- eksplicitne večmodelne algoritme v obliki Takagi-Sugeno [5], [118];
- algoritme na podlagi sprotne linearizacije modela Takagi-Sugeno [106], [1].

Eksplicitni večmodelni algoritmi

Načrtovanje algoritma, ki na eksplicitni način priredi strategijo linearnega MPC za nelinearni mehki model procesa, lahko razdelimo v tri faze [118]:

1. Najprej načrtamo model TS, kjer so posledični deli lokalni linearni modeli dinamike procesa v odgovarjajočih lokalnih regijah. Vhodne in izhodne signale je treba izbrati tako, da zadovoljivo pokrijejo čim večje območje delovanja procesa. Nato definiramo spremenljivke, ki nastopajo v pogojnem (ang. antecedent) delu pravil in za vsako regijo identificiramo parametre posledičnega dela.
2. Za vsak linearni podmodel načrtamo parametre linearnega lokalnega regulatorja MPC, ki deluje po predpisanih zahtevah v regiji pripadajočega podmodela. Nato sestavimo mehki regulator TS MPC, ki je mehki sistem TS z enakim pogojnim delom, kot ga ima model procesa, ter funkcijami posledičnega dela s parametri eksplicitnih lokalnih linearnih regulatorjev iz prejšnje točke.
3. Opravimo analizo dobljenega mehkega TS algoritma. Če ne zadosti zahtevam, se vrnemo na prejšnjo točko, popravimo načrtovanje parametrov regulatorjev ali lokalnih modelov in ponovimo analizo.

S predstavljeno strategijo dobimo zelo enostavno implementacijo nelinearnega regulatorja, če sta strukturi mehkih lokalnih modelov in mehkega regulatorja enaki (če imata enako število pravil in parametrov). V tem primeru se spreminjajo samo parametri ojačenja regulatorjev. Izračunamo jih kot uteženo vsoto ojačenj lokalnih regulatorjev.

Podoben postopek je bil opisan v delu [126], kjer po metodi prediktivnega funkcionalnega vodenja načrtamo parametre regulatorja za model procesa tipa TS v prostoru stanj. Izračunani regulirni signal je seštevek prispevkov lokalnih mehkih regulatorjev. Zlivanje prispevkov poteka preko enakih pripadnostnih funkcij, kot jih ima model procesa, tako da lahko regulirni zakon predstavimo z m paralelno razstavljenih lokalnih regulatorjev za vsakega od m regij mehkega modela.

Šibka točka eksplicitnega pristopa je, da dobimo regulirni zakon vedno v obliki brez omejitev, tako da moramo omejitve regulirnega signala vedno zagotoviti, ko je ta že izračunan, po drugi strani pa pri računanju vhoda v proces ne moremo upoštevati omejitev izhoda.

Algoritmi na podlagi sprotne linearizacije

Najosnovnejši način, kako dobiti suboptimalno rešitev problema nelinearne optimizacije, je linearizacija modela procesa v vsakem časovnem trenutku in uporaba linearnega MPC za trenutne parametre lineariziranega modela [80], [118]. Napoved trajektorije izhoda procesa je tu sestavljena iz dveh delov: prostega odziva, ki je odvisen od preteklih vrednosti vhodov in izhodov, in vsiljenega odziva, ki je odvisna od trenutnega in prihodnjih regulirnih signalov. Iz računskega vidika se odziva razlikujeta v tem, da se prvi računa v vsakem računskem koraku samo enkrat (in služi za računanje konstantnih parametrov v okviru enega optimizacijskega teka), medtem ko drugega ovrednotimo večkrat znotraj enega optimizacijskega okvirja, zato njegove lastnosti določajo tudi lastnosti celotne optimizacije. Če torej prosti odziv računamo na podlagi nelinearnega modela, vsiljeni pa na podlagi lineariziranega modela, je optimizacijski problem še vedno kvadratičen (QP) in ga lahko rešimo razmeroma enostavno in hitro. Še bolj pa se poenostavi, če predpostavimo enako strukturo vseh lokalnih modelov, ki sestavljajo mehki nelinearni model TS.

Predpostavimo mehki model v obliki pravil

$$\mathbf{R}_j : \text{če } x_1 \text{ je } \mathbf{A}_j \text{ in } \dots \text{ in } x_2 \text{ je } \mathbf{B}_i, \text{ potem } y(k+1) = a_i y(k) + b_i u(k) + r_i, \quad i = 1, \dots, m. \quad (4.2)$$

V obliki globalnega linearnega modela ga zapišemo tako, da njegov izhod izračunamo kot izhod linearnega modela s časovno spremenljivimi parametri

$$y(k+1) = \tilde{\mathbf{a}}(k)y(k) + \tilde{\mathbf{b}}(k)u(k) + \tilde{\mathbf{r}}(k), \quad (4.3)$$

kjer so $\tilde{\mathbf{a}}(k)$, $\tilde{\mathbf{b}}(k)$ in $\tilde{\mathbf{r}}(k)$ seštevki prispevkov parametrov vseh podmodelov. Parametre izračunamo v vsakem koraku tako, da najprej izračunamo funkcije normaliziranih stopenj izpolnjenosti posameznih pravil $\boldsymbol{\beta}^T = [\beta_1, \dots, \beta_m]$

$$\beta_j(k) = \frac{\mu_{A_{j,1}}(y(k)) \cdot \mu_{B_{j,1}}(u(k))}{\sum_{i=1}^m \mu_{A_{i,1}}(y(k)) \cdot \mu_{B_{i,1}}(u(k))}, \quad j = 1, \dots, m, \quad (4.4)$$

ki jih nato pomnožimo s parametri lokalnih modelov, da dobimo mehčane parametre

$$\begin{aligned}\tilde{\mathbf{a}}(k) &= \boldsymbol{\beta}^T [a_1, a_2, \dots, a_n]^T, \\ \tilde{\mathbf{b}}(k) &= \boldsymbol{\beta}^T [b_1, b_2, \dots, b_n]^T, \\ \tilde{\mathbf{r}}(k) &= \boldsymbol{\beta}^T [r_1, r_2, \dots, r_n]^T.\end{aligned}\tag{4.5}$$

Ker v vsakem računskem koraku izračunamo nove parametre dinamike procesa, ki velja v okolici trenutne delovne točke, lahko to predstavimo kot sprotno linearizacijo sistema. Trenutni parametri, ki so dejansko časovno spremenljivi, veljajo v enem računskem horizontu kot nespremenljivi, zato lahko tak mehki model uporabimo kot linearni prediktor in zanj načrtamo vodenje na podlagi ene od metod linearnega prediktivnega vodenja. Največkrat se uporabljata metodi DMC in GPC, v zadnjem času pa tudi druge metode. Abonyi in sodelavci [1] predlagajo metodo, kjer pri linearizaciji zaradi močno poenostavljene strukture mehkega modela parametre lineariziranega modela izračuna po enostavni vsoti produktov odvodov pripadnostnih funkcij in parametrov lokalnih modelov, tako da lahko v vsakem računskem koraku na enostaven način izračuna analitično rešitev kvadratičnega optimizacijskega problema. Roubos in sodelavci [106] pa so primerjali MPC na podlagi sprotno linearizacije modela TS metodo Branch & bound, ki v vsakem koraku izračuna globalno optimalno rešitev nelinearnega optimizacijskega problema, a za ceno zahteve po večji računski moči.

4.2.2 Nekatere druge metode na podlagi mehkega modela

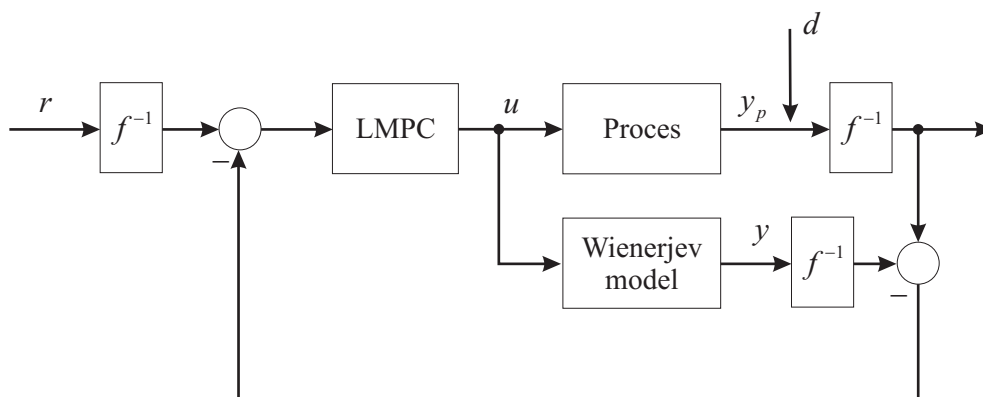
Ker se ostale metode od omenjenih razlikujejo le po podrobnostih, bomo, brez podrobnejšega opisa, omenili samo nekatere. Postlethwaite [94] je predstavil sistem s klasičnim prediktivnim regulatorjem na podlagi mehkega relacijskega modela, ki sledi pionirsko delo Grahama in Newella [46]. Huang in sodelavci [52] so uporabili mehki konvolucijski model in dvonivojski iterativni proces optimizacije, kjer v prvi fazi na spodnjem nivoju iščejo minimume pogreškov napovedi v vsakem od lokalnih sistemov, na zgornjem pa koordinirajo podsisteme, da dosežejo končen minimum pogreška. Zhang je za model TS v prostoru stanj uporabil optimizacijo *min-max* na podlagi odsekoma zveznih Lyapunovih funkcij. Za zapis v prostoru stanj sta elegantno elegantno rešitev predlagala Škrjanc in Matko [130], [131], kjer na podlagi strategije prediktivnega funkcionalnega vodenja [104] izračunata eksplicitno analitično rešitev za globalni linearni model. V delu [66] je Kavšek-Biassizo s sodelavci predstavila modifikacijo metode DMC za nelinearne sisteme, ki so modelirani z mehkim modelom.

V zadnjem času se v mehkem prediktivnem vodenju vse bolj uveljavljajo evlucijska optimizacija in hibridni pristopi. Omenimo samo deli Sarimveisa in Bafasa [108], ki

sta za model TS uvedla genetsko optimizacijo, ter Karerja in sodelavcev [64], ki so predlagali razširitev metode MPC na mehke hibridne modele.

4.2.3 Wienerjevo prediktivno vodenje

Wienerjevo prediktivno vodenje (ang. Wiener-model predictive control, WMPC) je poseben primer nelinearnega vodenja, kjer za napoved uporabimo Wienerjev model. V delih [86] in [87] je bila predstavljena danes najbolj razširjena metoda WMPC, katere shemo si lahko ogledamo na sliki 4.3. Za začetek uporabe Wienerjevih modelov v prediktivnem vodenju sicer šteje delo [95]. Iz podatkov najprej identificiramo nelinearno statično karakteristiko modela $f(v)$, kjer je v vmesna spremenljivka med linearnim in nelinearnim delom, nato izračunamo njen inverz f^{-1} in ga uporabimo za preslikavo reference ter izhoda. Na ta način namesto izhoda modela v povratno zanko vodimo oceno vmesne spremenljivke, ki je z vhomom v proces povezana preko linearnega dinamičnega dela. Vodenje tako prevedemo na linearni problem in zato lahko uporabimo linearni prediktivni regulator (LMPC) na podlagi sprotne optimizacije kvadratične kriterijske funkcije. V [87] je bila nelinearna statična karakteristika predstavljena z zleпки



Slika 4.3: Shema Wienerjevega prediktivnega vodenja

tretjega reda (odsekoma zvezni polinomi tretje stopnje), metoda pa v splošnem dopušča uporabo katerekoli invertibilne funkcije, torej tudi mehke modele. Le-ti so zaradi svoje fleksibilne strukture zelo primerni za opisovanje modelov z izrazito nelinearnostjo, kot npr. proces nevtralizacije pH. Dobra stran te metode je, da učinkovito odpravi nelinearnost v sistemu, tako da lahko s pomočjo optimizacije QP-problema izvedemo realno aplikacijo vodenja z omejitvami. Slaba stran pa je, da se lahko motnje in pogoški modeliranja dodatno povečajo pri preslikavi preko inverzne statične karakteristike, poleg tega pa linearne omejitve zaradi preslikave postanejo nelinearne in originalnega pristopa

LMPC spet ni mogoče uporabiti.

Izmed ostalih metod, ki ne uporabljajo eksplicitne preslikave preko inverzne statične karakteristike, omenimo dve metodi nelinearnega prediktivnega vodenja na podlagi Wienerjevega modela:

- Wienerjevo nelinearno prediktivno vodenje s sprotno linearizacijo po statični krivulji, ki prevede problem vodenja na LMPC s spremenljivim ojačenjem [41];
- nelinearno prediktivno vodenje na podlagi Wienerjevih odsekoma linearnih funkcij, pri katerih uporabimo LMPC in kompenzacijo spremenljivega odsekoma konstantnega ojačenja [79].

5. Uporaba modela INFUMO v zaznavanju napak

V tem poglavju bomo obravnavali možnosti vpeljave modela INFUMO v zaznavanje napak za nelinearne vhodno-izhodne procese z negotovimi parametri. Z negotovostmi parametrov navadno opisujemo vplive ti. neznanih vhodov v sistem - motenj, nemodelirane dinamike, merilnih šumov, neznanih bremen na vhodu procesa idr. Značilnost takih sistemov je, da večkratno vzbujanje z istim vzbujevalnim signalom daje neenake odzive, bolj ali manj razpršene po nekem nelinearnem območju. Težave pri uporabi klasičnih metod zaznavanja napak nastanejo takrat, kadar vplivov negotovosti ne moremo opisati z matematičnimi modeli, zato ne moremo izračunati parametrov za oblikovanje residuala. Pri metodah, ki uporabljajo robustne residue, pa moramo povečati pragove residualov, zato prihaja ali do nezanesljivosti napovedi napake ali pa pojavljanja lažnih alarmov.

Kot smo videli v poglavju 2., lahko s pomočjo modela INFUMO natančno opišemo tolerančni pas izhodov pri znanih vhodih v sistem. Predpostavka, iz katere izhajamo, je, da lahko procesne izhode v normalnem načinu delovanja, tj. v odsotnosti napak, obravnavamo kot nelinearne družine funkcij. Glavna ideja predlaganega pristopa je identificirati parametra modela INFUMO iz vhodno-izhodnih podatkov, pridobljenih ob normalnem delovanju procesa in obdelanih z določenimi metodami obdelave podatkov, nato pa model uporabiti v robustnem pasivnem zaznavanju napak. V nadaljevanju si bomo ogledali dve metodi, razviti v sklopu uporabe intervalnega mehkega modeliranja v zaznavanju in lokalizaciji napak.

5.1 Opis zasnove sistema

Naj bo nelinearni proces tipa SISO podan v splošni obliki na naslednji način:

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \sigma(\mathbf{x}, u, t) + \eta(\mathbf{x}, u, t) + \phi(\mathbf{x}, u, t) \\ y(t) &= \gamma(\mathbf{x}, u, t) + \rho(\mathbf{x}, u, t)\end{aligned}\tag{5.1}$$

kjer je $\mathbf{x} \in \mathbb{R}^n$ vektor stanj procesa, $u \in \mathbb{R}$ je vhod v proces, $y \in \mathbb{R}$ označuje izhod, $\eta : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}^n$ je negotovost stanj procesa, $\rho : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}$ je negotovost

izhoda, $\phi : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}^n$ označuje funkcijo napak, $\sigma : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}^n$ in $\gamma : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}$ pa sta nelinearni funkciji stanj, vhoda in časa. Naj bo merljiv samo izhod procesa. Opis negotovosti strnimo v dve predpostavki.

Predpostavka 5.1. *Negotovosti modela, ki jih predstavljata funkciji η in ρ v (5.1), sta neznani nestrukturirani nelinearni funkciji spremenljivk x , u in t , vendar omejeni z znanima funkcionaloma [138]*

$$\begin{aligned} |\eta(\mathbf{x}, u, t)| &\leq \bar{\eta}(y, u, t), \quad |\rho(\mathbf{x}, u, t)| \leq \bar{\rho}(y, u, t), \\ \forall(\mathbf{x}, y, u) &\in \mathcal{X} \times \mathcal{Y} \times \mathcal{U}, \quad \forall t \geq 0, \end{aligned} \quad (5.2)$$

kjer sta funkciji $\bar{\eta}(y, u, t)$ in $\bar{\rho}(y, u, t)$ poznani in uniformno omejeni. $\mathcal{X} \subset \mathbb{R}^n$ je domena stanj, $\mathcal{U} \subset \mathbb{R}$ in $\mathcal{Y} \subset \mathbb{R}$ pa sta kompaktni množici dosegljivih vhodov in izhodov.

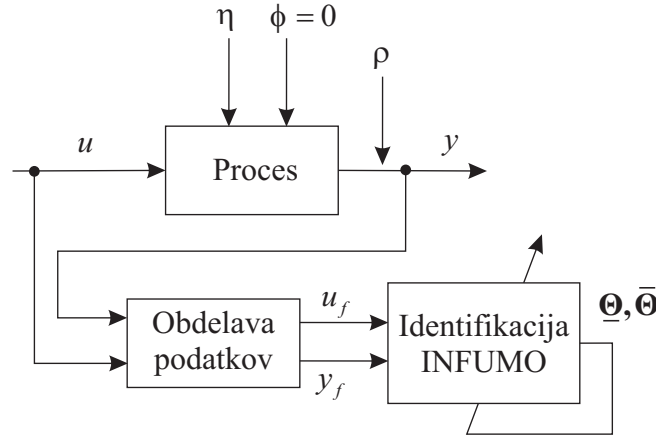
Poudarimo, da sta funkciji $\bar{\eta}$ in $\bar{\rho}$ odvisni od izhoda y medtem ko sta η in ρ funkciji vektorja stanj \mathbf{x} . To je posledica dejstva, da stanja procesa niso direktno merljiva in da lahko govorimo o omejitvi funkcij samo na podlagi podatkov, ki jih pridobimo iz izhoda procesa.

Predpostavka 5.2. *Izhod procesa $y(t)$ je, ko velja $\phi(x, u, t) = 0$, omejen z naslednjim intervalom:*

$$y(t) \in [\underline{y}(t), \bar{y}(t)] \subset \mathcal{Y}. \quad (5.3)$$

Predpostavka 5.1 določa, da so možne negotovosti modeliranja nestrukturirane in omejene z neko določeno konstanto ali funkcijo, predpostavka 5.2 pa zagotavlja, da se izhod procesa vedno nahaja v nekem intervalu, če niso prisotne napake v delovanju procesa. Interval $[\underline{y}, \bar{y}]$ lahko imenujemo pas zaupanja izhoda procesa, ker se izhod v normalnem delovanju vedno nahaja znotraj pasu z verjetnostjo enako 1.

Glavna težava je natančno določanje omenjenega intervala. V literaturi zasledimo nekaj pristopov, ki so večinoma osnovani na modeliranju nedoločenosti. V [29] so avtorji uporabili mrežo modelov, v katerih so negotovosti modelirali z intervalnimi parametri, izhod modela pa računali s pomočjo intervalne aritmetike. Za določanje meja intervalov so predlagali empirični ali numerični pristop. Pri prvem je treba poznati fizikalne lastnosti negotovosti parametrov, s pomočjo katerih potem prilagajamo vrednosti modela, da so izhodi modela konsistentni z množico podatkov, pridobljenih v normalnem delovanju procesa. Numerični pristop pa zahteva optimizacijo intervalov v skladu z najožjim intervalom izhoda, ki še ustreza podatkom normalnega delovanja. Šibka točka pristopa je v uporabi intervalne aritmetike, ki daje preveč konzervativne ocene kombinacij intervalov. Sainz in sodelavci [107] so zato predlagali uporabo modeliranja



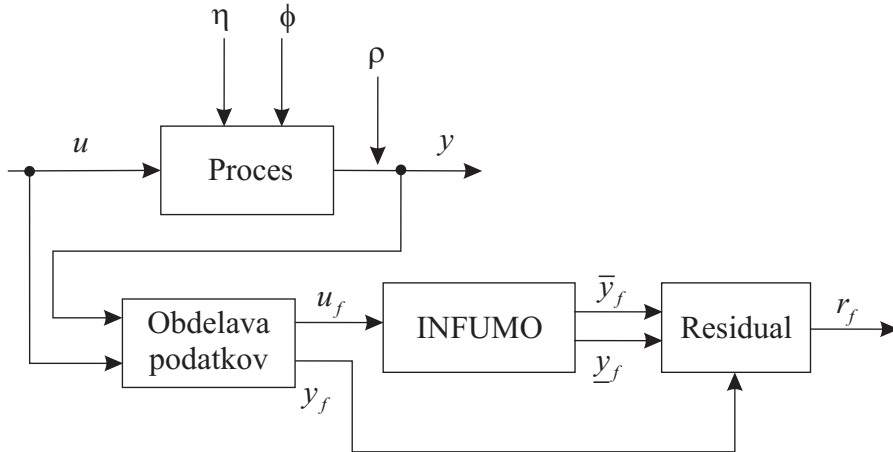
Slika 5.1: Shema aproksimacije pasu zaupanja z modelom INFUMO

z modalnimi intervali, ki zožijo interval zaupanja, pridobljen z intervalno aritmetiko. Za oceno intervala pa mora biti podan matematični model procesa, kjer so negotovosti parametrov podane s pripadajočimi tolerancami.

V našem primeru meja pasu zaupanja zaradi nepoznavanja vplivov negotovosti ne moremo določiti analitično. Zato moramo s poskusi pridobiti vhodno-izhodne podatke v normalnem načinu delovanja. Z obdelavo podatkov določimo vhodno-izhodno množico podatkov, ki jo aproksimiramo s pomočjo modela INFUMO. Na ta način dobimo funkcionalni opis pasu zaupanja, ki ga nato uporabimo za računanje residuala. Označimo z $u_f(t)$ in $y_f(t)$ vhod in izhod procesa, katera smo dobili po obdelavi podatkov. Slika 5.1 prikazuje strategijo pridobivanja pasu zaupanja. Dano množico podatkov $\{(u_{fi}, y_{fi})\}$, na katero vplivajo nedoločenosti v ρ , $|\rho| \in \bar{\rho}$ in η , $|\eta| \in \bar{\eta}$, omejimo z zgornjo in spodnjo funkcijo modela INFUMO. Postopek identifikacije, opisan v razdelku 2.3, nam omogoči aproksimacijo najožjega intervala, ki zajema vse točke množice podatkov. Pas zaupanja je tako podan z intervalom matrik parametrov $[\underline{\Theta}, \bar{\Theta}]$. Model INFUMO nato vežemo vzporedno procesu in uporabimo za računanje zgornje in spodnje meje odziva $y_f(t)$ pri danem vhodu $u_f(t)$. Meji $[\underline{y}_f, \bar{y}_f]$ lahko predstavimo tudi s širino in srednjo vrednostjo intervala, kar bomo uporabili pri zapisu residuala in oceni zaznavnosti napak:

$$\begin{aligned} y_f(t) &\in [\tilde{y}_f(t) - \Delta y_f(t), \tilde{y}_f(t) + \Delta y_f(t)], \\ \tilde{y}_f(t) &= \frac{1}{2}(\bar{y}_f + \underline{y}_f), \quad \Delta y_f(t) = \frac{1}{2}(\bar{y}_f - \underline{y}_f) \end{aligned} \quad (5.4)$$

V predlagani metodi bomo uporabili pasivni pristop oblikovanja residuala, robustnega na negotovosti. Shema prikazuje slika 5.2. Trenutni vhod v proces, katerega moramo predhodno obdelati po istem postopku kot pri pridobivanju podatkov, uporabimo za vhod v model INFUMO in izračun meja, trenutni obdelani izhod procesa pa primerjamo



Slika 5.2: Shema oblikovanja robustnega residuala z modelom INFUMO

s srednjo vrednostjo intervala zaupanja $\tilde{y}_f(t)$. Do napake ni prišlo, če residual $r_f(t)$ zadosti naslednjemu pogoju:

$$r_f(t) = y_f(t) - \tilde{y}_f(t) \in [-\Delta y_f(t), \Delta y_f(t)] = [r_f(t), \bar{r}_f(t)] \quad (5.5)$$

Razlika med danim residualom in residualom, predlaganim v npr. [98], je v tem, da tu upoštevamo obdelavo podatkov iz faze identifikacije modela INFUMO tudi pri zaznavanju napak. Ker gre, kot bomo videli v nadaljevanju, največkrat za filtriranje podatkov, lahko razliko razložimo kot poseben primer aktivnega pristopa oblikovanja robustnega generatorja residuala, opisanega v [25] in [33]:

$$r(t) = Q(p)(y(t) - \hat{y}(t)), \quad (5.6)$$

kjer je $\hat{y}(t)$ ocena izhoda procesa, $Q(p)$ pa filter, ki povečuje robustnost residuala na neznane vhode v proces, načrtovanje pa je zasnovano na poznavanju prenosnih funkcij napak in negotovosti. V našem primeru obdelava podatkov omogoči pridobivanje pasu zaupanja, po drugi strani pa omogoči prenos vplivov negotovosti v pas in s tem v residual.

5.2 Študija zaznavnosti

Kot je bilo nakazano že v [98], je test residuala (5.5) ekvivalenten limitnem testu $y_f(t) \in [\underline{y}_f, \bar{y}_f]$. Ker sta meji odvisni od nemodeliranih vplivov negotovosti v procesu, tj. neznanih vhodov in motenj, je slabost pristopa, da ne zazna poljubno majhnih napak. Za študijo zaznavnosti iščemo torej skrajni primer, ko napaka ϕ ni zaznana. Povzema ga naslednji teorem.

Teorem 5.1. Imenujmo $y_{\phi f}(t)$ doprinos napake $\phi(x, u, t)$ k izhodu procesa po obdelavi podatkov, razliko med mejnima funkcijama modela INFUMO pa označimo $\bar{y}_f(t) - \underline{y}_f(t) = \lambda(t)$. Napaka bo zagotovo zaznana, če bo absolutna vrednost njenega doprinosa večja od širine pasu zaupanja, tj.

$$|y_{\phi f}(t)| > \lambda(t). \quad (5.7)$$

Dokaz. Za potrebe analize zaznavnosti zapišimo izhod procesa po obdelavi podatkov kot

$$y_f(t) = y_{0f}(t) + \rho_f(t) + y_{\phi f}(t), \quad (5.8)$$

kjer $y_{0f}(t)$, $\rho_f(t)$ in $y_{\phi f}(t)$ označujejo prispevke nominalnega izhoda, negotovosti in napak. Nominalni izhod bi na izhodu procesa dobili, če ne bi bilo negotovosti in napak. Ker sta meji pasu zaupanja $\bar{y}_f(t)$ in $\underline{y}_f(t)$ dobljeni s poskusi v normalnem načinu delovanja ($\phi(t) = 0$), lahko zapišemo naslednji neenakosti:

$$\begin{aligned} y_{0f}(t) + \rho_f(t) - \underline{y}_f(t) &\geq 0, \\ \bar{y}_f(t) - y_{0f}(t) - \rho_f(t) &\geq 0. \end{aligned} \quad (5.9)$$

Če prištejemo doprinos napak $y_{\phi f}(t)$ na obeh straneh neenačbe (5.9), dobimo

$$\begin{aligned} y_{0f}(t) + y_{\phi f}(t) + \rho_f(t) - \underline{y}_f(t) &\geq y_{\phi f}(t) \Rightarrow y_{\phi f}(t) \leq y_f(t) - \underline{y}_f(t), \\ \bar{y}_f(t) - y_{0f}(t) - \rho_f(t) - y_{\phi f}(t) &\geq -y_{\phi f}(t) \Rightarrow \bar{y}_f(t) - y_f(t) \geq -y_{\phi f}(t). \end{aligned} \quad (5.10)$$

Sledi, da je $y_{\phi f}(t)$ omejen z

$$y_f(t) - \bar{y}_f(t) \leq y_{\phi f}(t) \leq y_f(t) - \underline{y}_f(t). \quad (5.11)$$

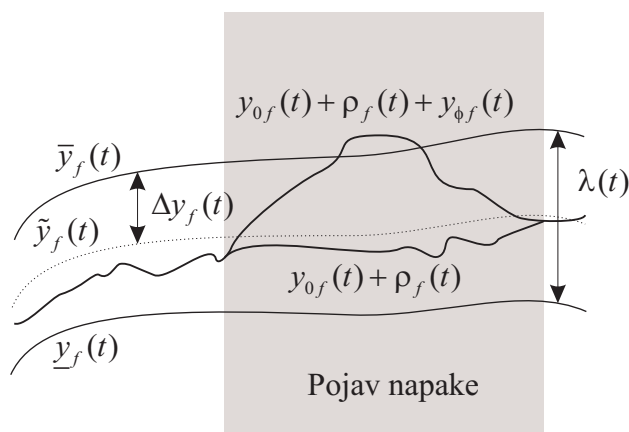
Kadar ni zaznana nobena napaka, tj. $\underline{y}_f(t) \leq y_f(t) \leq \bar{y}_f(t)$, lahko neenačbo (5.10) razširimo v

$$\begin{aligned} y_{\phi f}(t) \leq y_f(t) - \underline{y}_f(t) &= \left(\bar{y}_f(t) - \underline{y}_f(t) \right) + \left(y_f(t) - \bar{y}_f(t) \right) \\ &= \lambda(t) + \left(y_f(t) - \bar{y}_f(t) \right) \leq \lambda(t) + 0 = \lambda(t), \\ -y_{\phi f}(t) \leq \bar{y}_f(t) - y_f(t) &= \left(\bar{y}_f(t) - \underline{y}_f(t) \right) + \left(\underline{y}_f(t) - y_f(t) \right) \\ &= \lambda(t) + \left(\underline{y}_f(t) - y_f(t) \right) \leq \lambda(t) + 0 = \lambda(t). \end{aligned} \quad (5.12)$$

Oba izraza v (5.12) lahko združimo v

$$|y_{\phi f}(t)| \leq \lambda(t). \quad (5.13)$$

Kadar napaka ni zaznana, je zadoščeno neenačbi (5.13). Iz tega sledi, da v primeru, ko velja $|y_{\phi f}(t)| > \lambda(t)$, napako zanesljivo zaznamo. \square



Slika 5.3: Problematika zaznavanja majhnih napak

Pojasnimmo dobljeni rezultat na sliki 5.3. Meji pasu zaupanja se gibata v skladu s trenutno delovno točko sistema, v odsotnosti napak pa je $y_f(t) = y_{0f}(t) + \rho_f(t)$ vedno znotraj pasu. Ob pojavu napake se vrednost y_f spremeni, vendar ni nujno, da zapusti pas zaupanja - v skrajnem primeru, ko je signal $y_{0f}(t) + \rho_f(t)$ na eni od meja, mora doprinos napake biti večji od širine intervala v tistem trenutku.

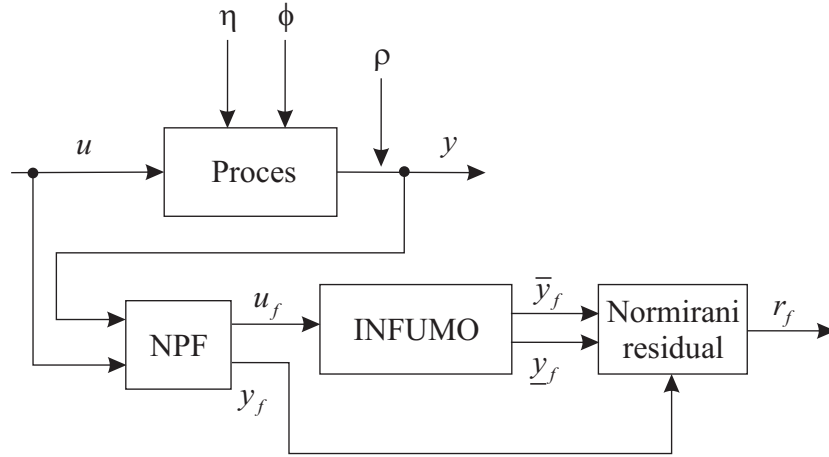
5.3 Zaznavanje in lokalizacija napak na podlagi nizkopasovnega filtriranja

Obdelava podatkov, ki jih dobimo s poskusi na procesu v normalnem načinu delovanja, je ključni del predlaganega sistema za zaznavanje napak. V prvem pristopu uporabimo filtriranje vhoda in izhoda z nizkopasovnim filtrom. Cilj filtriranja je odrezati visokofrekvenčni del frekvenčnega spektra signalov. Na ta način aproksimiramo nelinearno statično vhodno-izhodno karakteristiko procesa, ki je ob odsotnosti napak že pas zaupanja, poleg tega pa v karakteristično območje filter prepusti določen del vplivov negotovosti v procesu, odvisno od karakteristike filtra.

Slika 5.4 prikazuje shemo zaznavanja napak. Filtriranje z nizkopasovnim filtrom omogoča zvezno računanje residuala, vendar pa ga iz enačbe (5.5) ne dobimo neposredno, ker filtrirani izhod primerjamo s srednjo vrednostjo intervala \tilde{y}_f . Zato uvedemo normirani residual na način, podoben računanju distančne funkcije v delu [29]:

$$r_f(t) = \begin{cases} \frac{y_f(t) - \tilde{y}_f(t)}{\underline{y}_f(t) - \tilde{y}_f(t)}, & y_f(t) < \tilde{y}_f(t), \\ \frac{y_f(t) - \tilde{y}_f(t)}{\tilde{y}_f(t) - \underline{y}_f(t)}, & y_f(t) > \tilde{y}_f(t). \end{cases} \quad (5.14)$$

Zdaj za zaznavanje napak neposredno testiramo residual $r_f(t)$. Odločitveno funkcijo



Slika 5.4: Zaznavanje napak z nizkopasovnim filtrom, modelom INFUMO in normiranim residualom

$f(t)$ definiramo kot

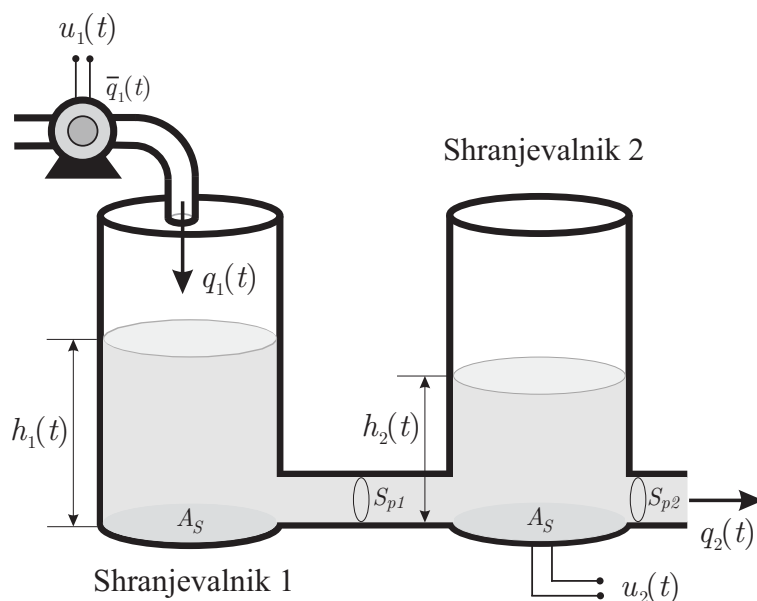
$$f(t) = \begin{cases} 1, & r_f(t) > 1, \\ 0, & r_f(t) \leq 1. \end{cases} \quad (5.15)$$

Če residual preseže vrednost ena, je filtrirani izhod procesa zapustil pas zaupanja, zato lahko z verjetnostjo ena trdimo, da se je v procesu zgodila napaka in sistem javi alarm.

5.3.1 Zaznavanje napak za model dveh shranjevalnikov

Model dveh shranjevalnikov tekočine, povezanih preko ventila, je standardni primer v študijah nelinearnih metod zaznavanja napak. V primeru bomo uporabili model, opisan v delu [137]. Prikazuje ga slika 5.5. Imejmo dva identična cilindrična shranjevalnika s prečnim presekom $A_s = 0.0154 \text{ m}^2$. Prečni presek povezovalnega ventila in izhodnega ventila naj bo $S_{p1} = S_{p2} = 3.6 \cdot 10^{-5} \text{ m}^2$, nivoja tekočine v shranjevalnikih 1 in 2 pa označimo h_1 in h_2 . Vhodni pretok tekočine v shranjevalnik 1 dovaja električna črpalka, označimo ga z $q_1(t)$, izhodni pretok iz shranjevalnika 2 pa $q_2(t)$. Ob upoštevanju ravnovesja mas v shranjevalnikih in Torricellijevega zakona, ki pravi, da je hitrost izhodnega toka tekočine iz shranjevalnika sorazmerna kvadratnemu korenu nivoja tekočine, dobimo naslednje enačbe:

$$\begin{aligned} \dot{h}_1 &= \frac{1}{A_s} \left(-K_{p1} \text{sign}(h_1 - h_2) \sqrt{2g|h_1 - h_2|} + q_1 \right), \\ \dot{h}_2 &= \frac{1}{A_s} \left(K_{p1} \text{sign}(h_1 - h_2) \sqrt{2g|h_1 - h_2|} - K_{p2} \sqrt{2gh_2} \right), \end{aligned} \quad (5.16)$$



Slika 5.5: Proces dveh shranjevalnikov

kjer s $K_{p1} = a_1 S_{p1}$ in $K_{p2} = a_2 S_{p2}$ označimo konstanti izhodnega pretoka, g pa je gravitacijski pospešek. Zaradi enostavnejšega računanja privzemimo $a_1 = a_2 = 1$.

Da bi se bolj približali dejanskim razmeram v industriji, spremenimo zasnovo sistema na naslednji način. Vhod v proces naj bo napetost na električni črpalki, ki proizvede pretok

$$\bar{q}_1(t) = K_u(1 + \nu_1(t))u_1(t), \quad (5.17)$$

kjer s K_u označimo konstanto pretvorbe napetost-pretok, z $\nu_1(t)$ pa negotovost pretvorbe. Edini merljivi signal je napetost tlačnega senzorja, ki pretvarja nivo tekočine v shranjevalniku 2 ($h_2(t)$) v napetostni signal $u_2(t)$ po enačbi

$$y(t) = u_2(t) = K_h(1 + \nu_2(t))h_2(t), \quad (5.18)$$

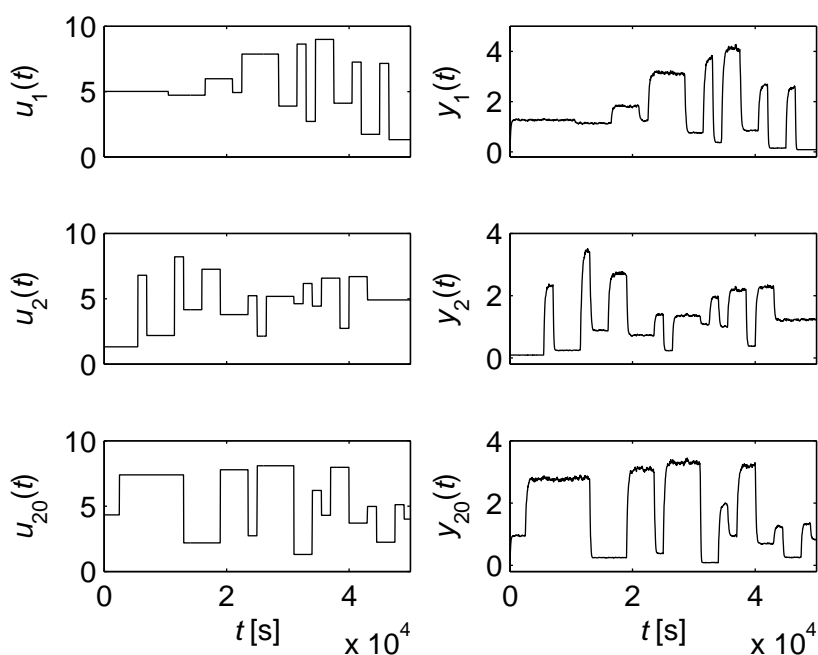
kjer je K_h konstanta pretvorbe nivo-napetost, $\nu_2(t)$ pa negotovost te pretvorbe. Privzemimo, da sta negotovosti omejeni, kot je določeno v predpostavki ass:omej. Vrednosti konstant sta $K_u = 8.8 \cdot 10^{-6} \text{ m}^3/\text{Vs}$ in $K_h = 16.667 \text{ V/m}$, zgornji meji negotovosti pa $\bar{\nu}_1 = \bar{\nu}_2 = 0.03$. Pri merjenju privzamemo komunikacijo z osebnim računalnikom preko pretvornika AD-DA s časom vzorčenja $T_s = 10 \text{ s}$.

Obravnavali bomo nabor standardnih napak v sistemu shranjevalnikov:

- **napaka na črpalki** - privzemimo faktor napake $K_f \in [0, 1]$, s katerim opišemo napako vhodnega pretoka kot $q_1(t) = q_{1,0}(t) + (1 - K_f)q_{1,0}(t)$, kjer je $q_{1,0}(t)$ nominalni pretok;

- **puščanje v shranjevalniku 1** - predpostavimo okroglo odprtino neznanega polmera r_1 , ki povzroči dodatni iztok $q_{f1} = a_1\pi(r_1)^2\sqrt{2gh_1}$;
- **puščanje v shranjevalniku 2** - enako kot pri shranjevalniku tu predpostavimo polmer r_2 in iztok $q_{f2} = a_2\pi(r_2)^2\sqrt{2gh_2}$.

Najprej se posvetimo pridobivanju podatkov za oblikovanje pasu zaupanja. Izvedemo 20 eksperimentov s stopničastim signalom na vходу procesa. Omenimo naj, da ni nujno, da je vhodni signal vedno enak; lahko določimo vsakič drugačen časovni potek. To dejstvo odgovarja stanju v industriji, kjer imamo navadno za dani proces na voljo veliko različnih meritev. Paziti je treba le, da množica meritev zaobjame vse delovne pogoje normalnega delovanja procesa. V našem primeru smo to simulirali tako, da so pri vsaki meritvi negotovi parametri zavzeli poljubne vrednosti iz svojih intervalov, določenih z omejitvami pripadajočih negotovosti, vhodni signali pa so bili izbrani tako, da smo pokrili večino delovnega območja. Slika 5.6 prikazuje rezultate eksperimentov; zaradi omejitve prostora smo vključili samo prvo, drugo in zadnjo meritev. Meritve



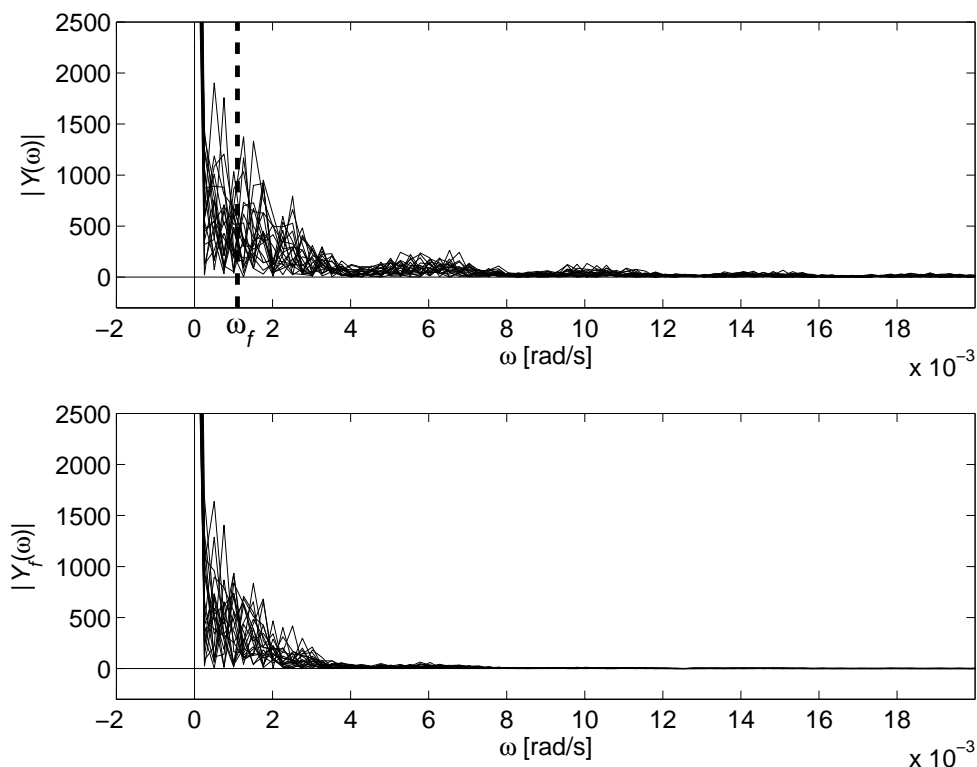
Slika 5.6: Meritve izhoda pri danih vhodih: prva, druga in zadnja meritev

vhoda in izhoda filtriramo z nizkopasovnim filtrom prvega reda

$$Q(p) = \frac{1}{T_f p + 1}. \quad (5.19)$$

V splošnem lahko izberemo poljubni nizkopasovni filter, tudi višjega reda, npr. Butterworthov filter. V praksi se je izkazalo, da spreminjanje oblike filtra ne prinese bi-

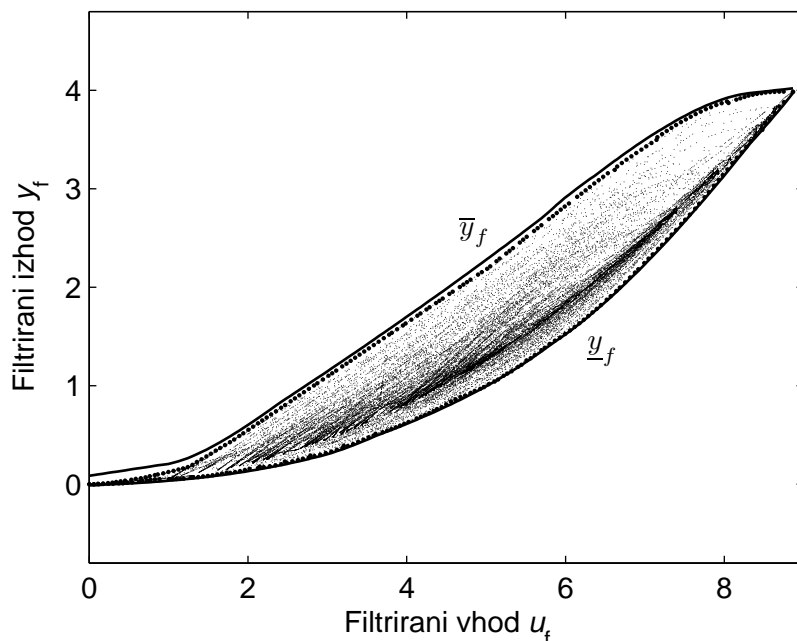
stvene razlike pri oblikovanju pasu zaupanja. Glavna naloga filtra je izločanje visokofrekvenčnega spektra, zato je izbira filtra (5.19) najenostavnejša in najbolj transparentna. Časovna konstanta T_f v tem primeru predstavlja kompromis med hitrostjo zaznavanja napak in občutljivostjo sistema na lažne alarme. Določimo jo lahko s pomočjo Fourierove analize signalov, kot je prikazano na sliki 5.7. Lomno frekvenco ω_f postavimo



Slika 5.7: Izbira lomne frekvence filtra

tako, da filter prepusti večji del zanimive frekvenčne vsebine signalov, tj. frekvenčno območje, kjer je absolutna vrednost Fourierjevega transformata visoka glede na ostalo območje. Časovno konstanto potem neposredno izračunamo iz vrednosti izbrane lomne frekvence, $T_f = 476$ s.

Po filtriranju dobimo množico vhodno-izhodnih filtriranih podatkov, ki je zaradi obsega meritev zelo velika. Prikazana je s točkami na sliki ???. Da bi se izognili težavam s konvergenco optimizacijske metode, najprej določimo mejne točke. Vhodni prostor razdelimo na ozke podprostore, v katerih določimo maksimum in minimum. Širino podprostorov določa področje z največjo gostoto podatkov. Med mejnimi točkami potegnemo daljice in preverimo, če ni katera od preostalih točk izven prostora, omejenega z daljicami. Na ta način smo prostor omejili s 302 mejnima točkama, množica mejnih točk pa je na sliki ??? poudarjena. Za aproksimacijo smo uporabili model INFUMO,



Slika 5.8: Množica filtriranih podatkov z mejnimi točkami, omejene z modelom INFUMO

zapisan s pravili

$$R_j : \text{če } u_f \text{ je } A_j, \text{ potem } \bar{y}_f = \bar{\theta}_{j,0} + \bar{\theta}_{j,1}u_f, \quad (5.20)$$

$$\underline{y}_f = \underline{\theta}_{j,0} + \underline{\theta}_{j,1}u_f, \quad j = 1, \dots, m.$$

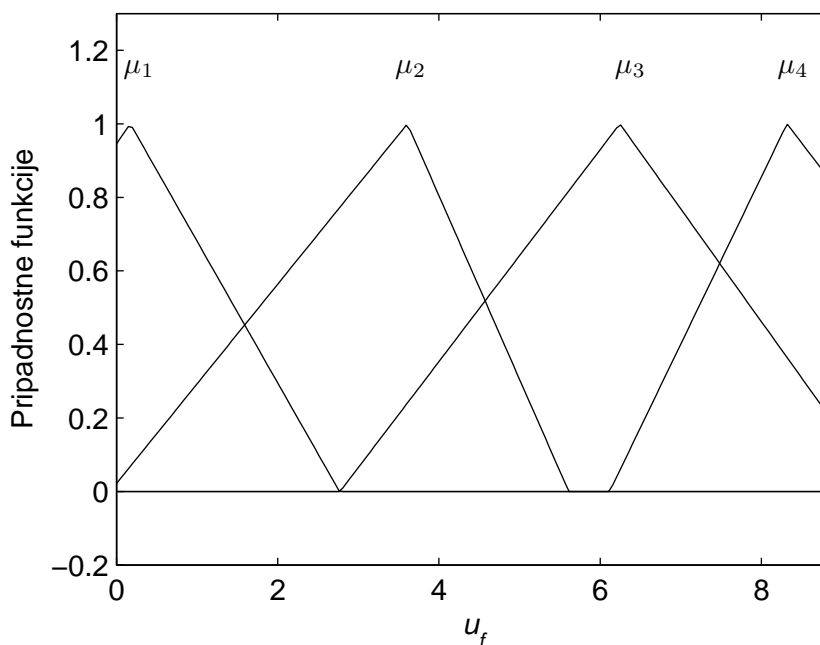
Uporabili smo štiri trikotne pripadnostne funkcije, njihovo število pa smo določili glede na obliko področja, ki ga aproksimiramo. Razporeditev pripadnostnih funkcij, prikazano na sliki 5.9, smo izračunali po metodi *grid partitioning*, opisani v [6]. Parametre modela INFUMO smo optimizirali po postopku iz leme 2.2 in dobili

$$\underline{\Theta} = \begin{bmatrix} 0,22 & 0,0060 \\ 0,3764 & -0,7846 \\ 0,5773 & -1,94 \\ 0,8948 & -3,774 \end{bmatrix}, \quad \bar{\Theta} = \begin{bmatrix} 0,1226 & 0,0855 \\ 0,4491 & -0,2499 \\ 0,4767 & -0,01455 \\ 0,0882 & 3,238 \end{bmatrix}. \quad (5.21)$$

Mejni funkciji sta prikazani na sliki 5.8. Očitno je, da dosežemo zadovoljiv nivo aproksimacije; maksimalna pogreška sta $\lambda_1 = 0,1061$ in $\lambda_2 = 0,0371$.

V testnem poskusu smo simulirali odziv sistema pri treh napakah:

- napaka na črpalki - 20-odstotna sprememba napetosti ($K_f = 0,2$) se zgodi v periodi $t_{akt} = 8000 - 15000$ s;



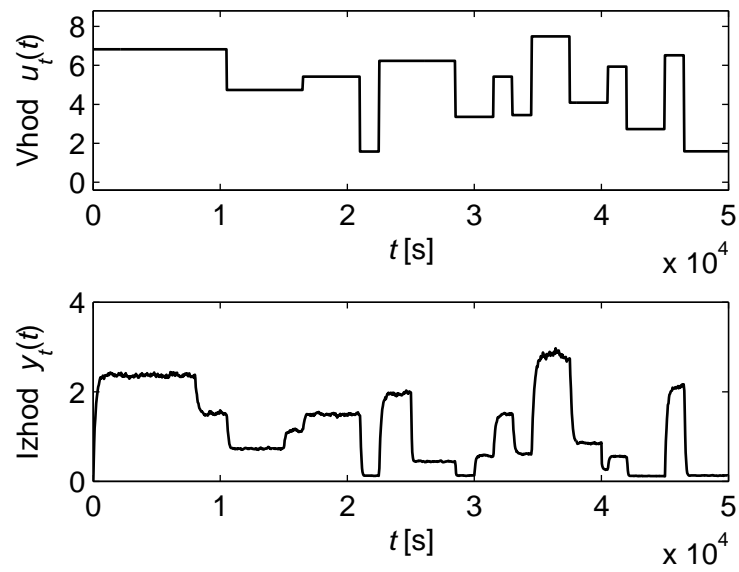
Slika 5.9: Postavitev pripadnostnih funkcij

- puščanje v shranjevalniku 1 - iztok iz luknje s polmerom $r_1 = 3 \cdot 10^{-3}$ v periodi $t_{izt1} = 25000 - 30000$ s;
- puščanje v shranjevalniku 2 - iztok iz luknje s polmerom $r_2 = 3 \cdot 10^{-3}$ v periodi $t_{izt2} = 40000 - 45000$ s.

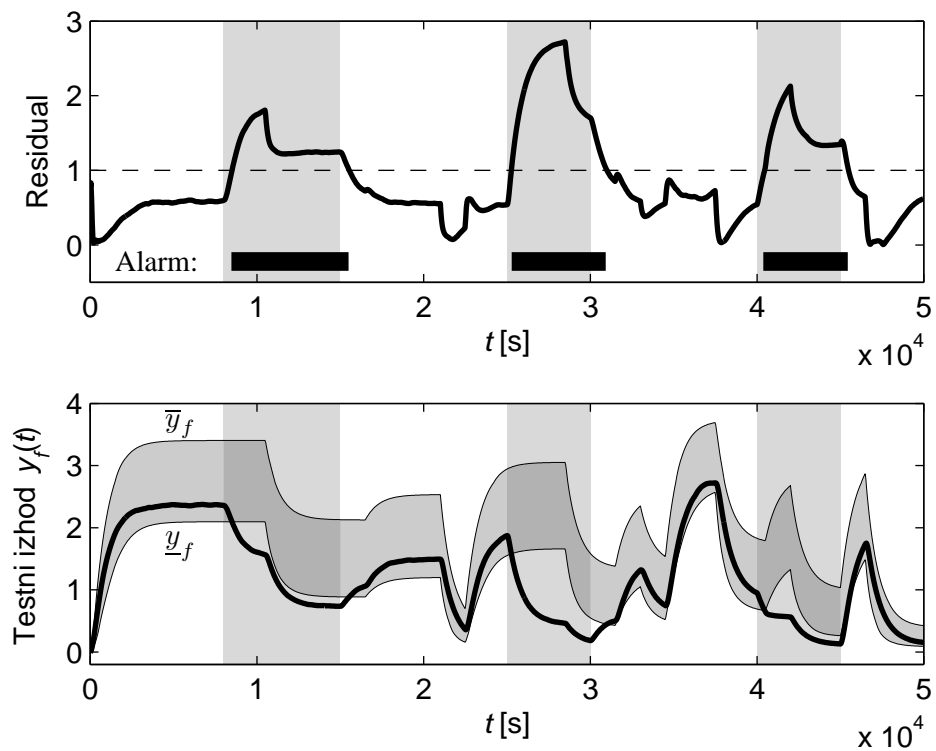
V poskusu smo uporabili signal vhoda, ki je s pripadajočim izhodom prikazan na sliki 5.10. Rezultate kaže slika 5.11. Zgornji diagram prikazuje normalizirani residual, ki preseže vrednost ena samo v obdobjih pojavov napak (senčeno področje). Na spodnjem diagramu iste slike pa vidimo časovne poteke filtriranega izhoda procesa ter meja pasu zaupanja, ki ju določa model INFUMO. Signal "Alarm" kaže, kje sistem javi napako. Časi zaznave (od pojava napake do zaznave) so 467,7 s, 288,0 s in 380,2 s.

5.3.2 Lokalizacija napak za model dveh shranjevalnikov

Sistem zaznavanja napak v razdelku 5.3.1 je sposoben odkriti vse napake, ni pa sposoben ločevati med njimi. Za lokalizacijo napak je treba v sistemu dodati karakteristične veličine, ki reagirajo različno na različne tipe napak. To seveda v splošnem pomeni razširitev sistema - več meritev, dodajanje strojne opreme (novi senzorji) ali modifikacija programske opreme. Enačba (5.1) je treba prirediti za sistem tipa SIMO (en vhod,



Slika 5.10: Vhod in izhod v testnem poskusu



Slika 5.11: Rezultati zaznavanja napak

več izhodov):

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \sigma(\mathbf{x}, u, t) + \eta(\mathbf{x}, u, t) + \phi(\mathbf{x}, u, t), \\ \mathbf{y}(t) &= \gamma(\mathbf{x}, u, t) + \rho(\mathbf{x}, u, t),\end{aligned}\tag{5.22}$$

kjer $y \in \mathbb{R}^p$ označuje p izhodov, $\rho : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}^p$ je funkcija negotovosti izhodov, $\gamma : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}^p$ pa je nelinearna funkcija izhodov. Spremeni se tudi predpostavka 5.2:

Predpostavka 5.3. *Funkcije izhoda $y_i(t)$, $i = 1, \dots, p$ so, ko velja $\phi(\mathbf{x}, u, t) = 0$, omejene z intervali*

$$y_i(t) \in [\underline{y}_i(t), \bar{y}_i(t)] \subset \mathcal{Y}.\tag{5.23}$$

Ob odsotnosti napak v procesu torej vsi izhodi zavzamejo vrednosti znotraj pripadajočih intervalov $[\underline{y}_i, \bar{y}_i]$.

V predlagani metodi je razširitev neposredna - na podlagi meritev izhodov razširjenega sistema in nizkopasovnega filtriranja oblikujemo intervale zaupanja, identificiramo pripadajoče modele INFUMO in jih priključimo paralelno procesu. Vsak model INFUMO za pripadajoči procesni izhod $y_{fi}(t)$ podaja mejni funkciji $[\underline{y}_{fi}(t), \bar{y}_{fi}(t)]$ in na podlagi teh signalov izračunamo p normiranih residualov po enačbi

$$r_{fi}(t) = \begin{cases} \frac{y_{fi}(t) - \bar{y}_{fi}(t)}{\bar{y}_{fi}(t) - \underline{y}_{fi}(t)}, & y_{fi}(t) < \bar{y}_{fi}(t), \\ \frac{y_{fi}(t) - \underline{y}_{fi}(t)}{\bar{y}_{fi}(t) - \underline{y}_{fi}(t)}, & y_{fi}(t) > \bar{y}_{fi}(t). \end{cases}\tag{5.24}$$

Zaznavanje napak je v tem primeru preverjanje p residualov, ki pogojuje p odločitvenih funkcij

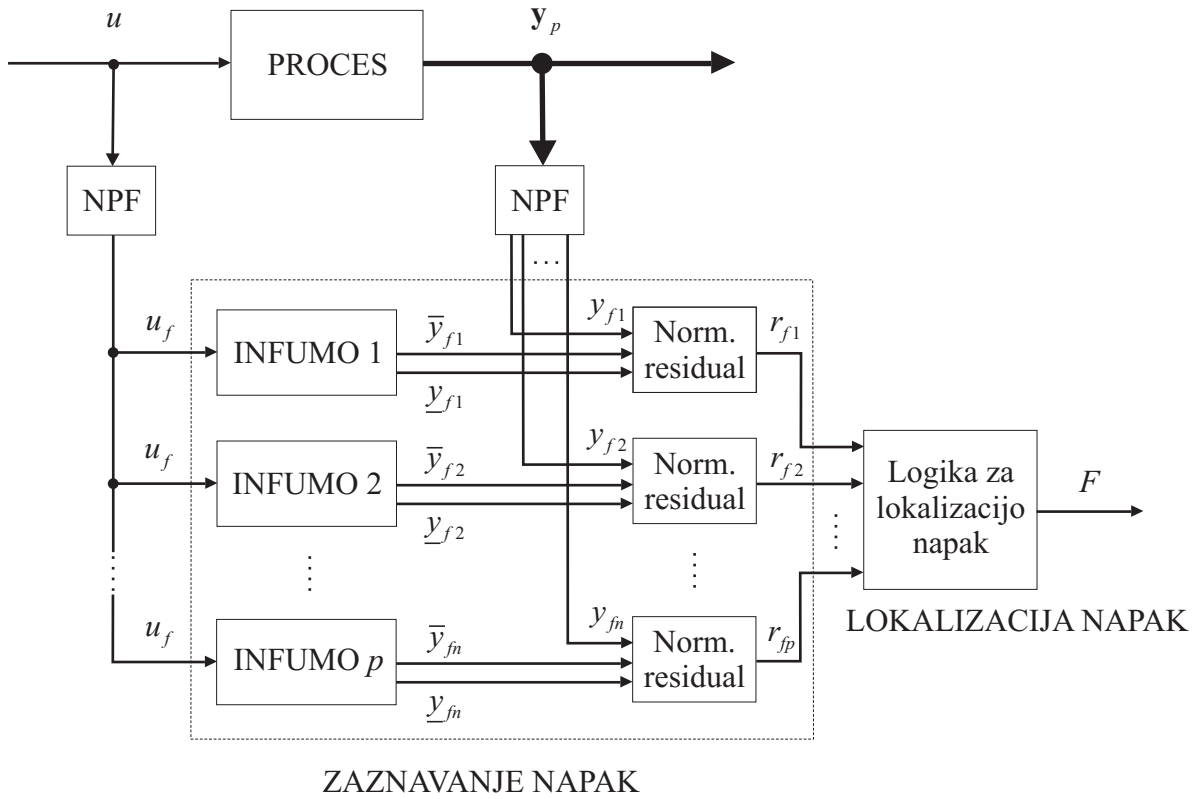
$$f_i(t) = \begin{cases} 1, & r_{fi}(t) > 1, \\ 0, & r_{fi}(t) \leq 1. \end{cases}\tag{5.25}$$

Koda napake je p -mestno binarno število, ki ga navadno spremenimo v desetiški zapis $F = f_1 \cdot 2^0 + f_2 \cdot 2^1 + \dots + f_p \cdot 2^{p-1}$.

Shemo zaznavanja in lokalizacije napak prikazuje slika 5.12.

Imejmo enak sistem shranjevalnikov (5.16) kot v razdelku 5.3.1, dodajmo pa merilca nivoja v prvem shranjevalniku in pretoka med shranjevalnikoma. Za vhod velja enačba (5.17), enačba izhoda pa se spremeni v

$$\begin{aligned}y_1(t) &= K_{h1}(1 + \nu_2(t))h_1(t), \\ y_2(t) &= K_{h2}(1 + \nu_3(t))h_2(t), \\ y_3(t) &= K_q(1 + \nu_4(t))q_{12}(t),\end{aligned}\tag{5.26}$$

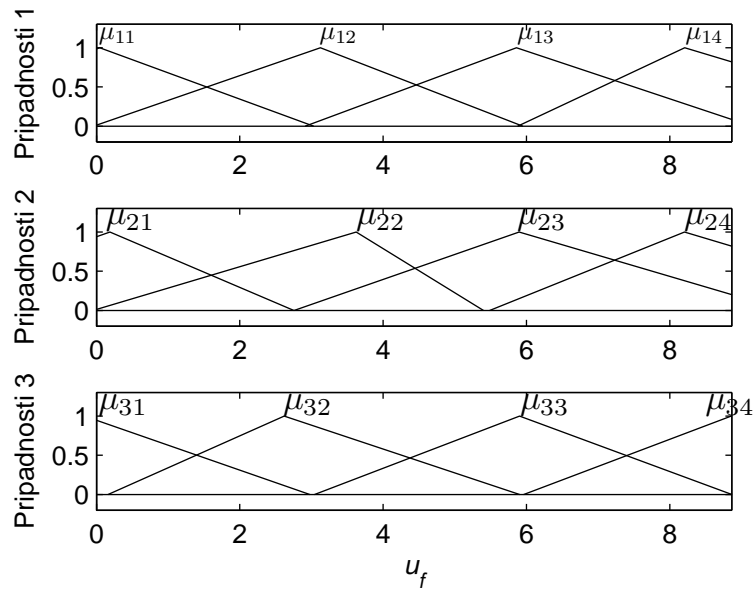


Slika 5.12: Zaznavanje napak z nizkopasovnim filtrom, modelom INFUMO in normiranim residualom

kjer sta $y_1(t)$ in $y_2(t)$ napetosti na senzorjih nivoja, K_{h1} in K_{h2} konstanti pretvorbe nivo-napetost, $\nu_2(t)$ in $\nu_3(t)$ negotovosti pretvorb nivoja, $q_{12}(t)$ pretok tekočine v povezovalnem ventilu, $y_3(t)$ napetost merilca pretoka, K_q konstanta pretvorbe pretok-napetost in $\nu_4(t)$ negotovost pretvorbe pretoka. Vrednosti konstant naj bodo $K_u = 8.8 \cdot 10^{-6} \text{ m}^3/\text{Vs}$, $K_q = 1.1364 \cdot 10^5 \text{ Vs/m}^3$, $K_{h1} = K_{h2} = 16.667 \text{ V/m}$ ter $\bar{\nu}_1 = \bar{\nu}_2 = \bar{\nu}_3 = \bar{\nu}_4 = 0.03$.

Obravnavali bomo pet napak:

- **puščanje v shranjevalniku 1** - predpostavimo okroglo odprtino neznanega polmera r_1 , ki povzroči dodatni iztok $q_{f1} = a_1 \pi (r_1)^2 \sqrt{2gh_1}$;
- **puščanje v shranjevalniku 2** - enako kot pri shranjevalniku tu predpostavimo polmer r_2 in iztok $q_{f2} = a_2 \pi (r_2)^2 \sqrt{2gh_2}$;
- **napaka v senzorju nivoja 1** - predpostavimo multiplikativno napako, ki spremeni nominalno napetost $y_{1,0}$ v $y_1(t) = y_{1,0}(t) + (1 - K_{s1})y_{1,0}(t)$, kjer s $K_{1off} \in [0, 1]$ označimo velikost napake;

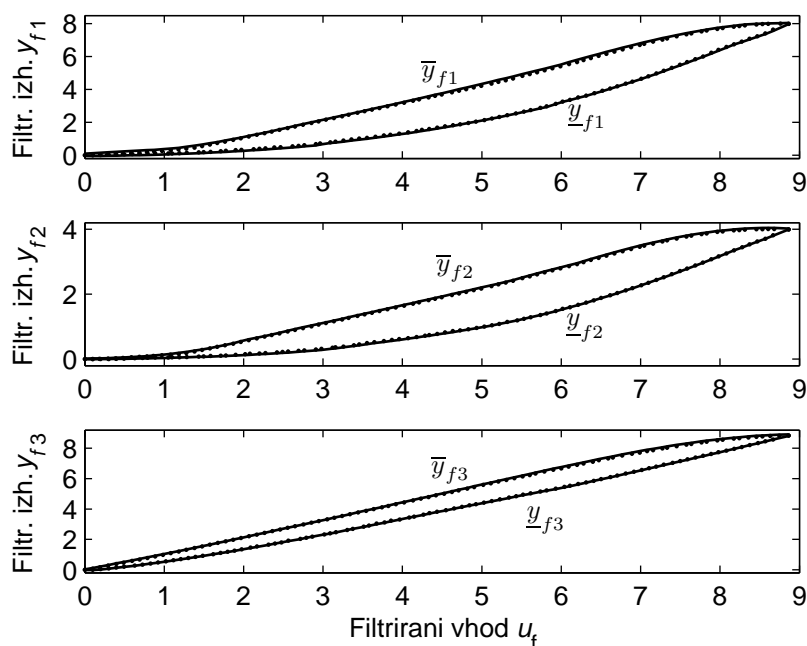


Slika 5.13: Postavitve pripadnostnih funkcij

- **napaka v senzorju nivoja 2** - enako kot pri senzorju nivoja 1 opišemo dejansko napetost senzorja z $y_2(t) = y_{2,0}(t) + (1 - K_{s2})y_{2,0}(t)$;
- **napaka v senzorju pretoka** - enako kot prej opišemo dejansko napetost senzorja z $y_3(t) = y_{3,0}(t) + (1 - K_{s3})y_{3,0}(t)$.

Ponovimo 20 serij meritev (signali $u(t)$ na levi strani slike 5.6), privzamemo enako obliko modela INFUMO (5.20) in nizkopasovnega filtra (5.19) z enako časovno konstanto $T_f = 476$ s. Izvedemo redukcijo množic podatkov. Po enakem postopku izračunamo tudi razporeditve štirih pripadnostnih funkcij za vsak izhod, ki jih prikazuje slika 5.13. Z identifikacijo modelov INFUMO omejimo filtrirane množice vhodno-izhodnih podatkov, kot to prikazuje slika 5.14. V testnem poskusu simulirajmo iztok iz luknje polmera $r_1 = 2 \cdot 10^{-3}$ m v shranjevalniku 1 ob $t_{izt1} = 10000 - 15000$ s, iztok iz luknje polmera $r_2 = 2 \cdot 10^{-3}$ m v shranjevalniku 2 ob $t_{izt2} = 20000 - 25000$ s, 20-odstotno napako v senzorju nivoja 1 ($K_{s1} = 0.2$) ob $t_{s1} = 30000 - 35000$ s, 20-odstotno napako v senzorju nivoja 2 ($K_{s2} = 0.2$) ob $t_{s2} = 40000 - 45000$ s in 20-odstotno napako v senzorju pretoka 3 ($K_{s3} = 0.2$) ob $t_{s3} = 50000 - 55000$ s. Glede na odločitvene funkcije f_i residualov r_{fi} s sheme 5.12, zgradimo incidenčno matriko, kot je podano v tabeli 5.1. Matrika je ločljiva, vendar zaradi prvih dveh stolpcev ni strogo ločljiva. Za to bi potrebovali dodatne senzorje.

Rezultati testnega poskusa so predstavljeni na slikah 5.15 in 5.16. S slike 5.15 je

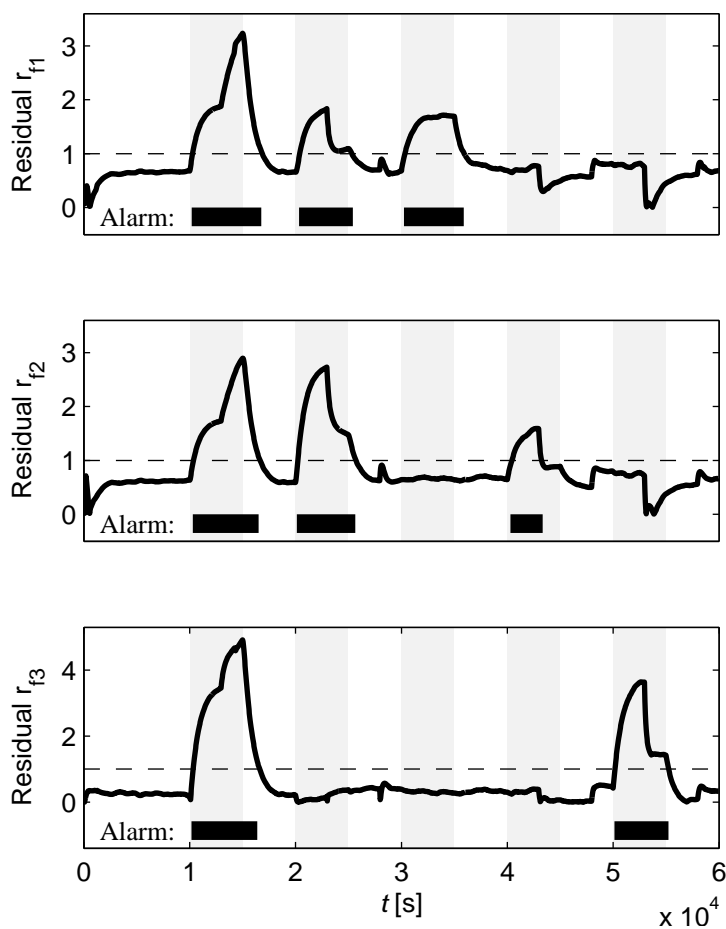


Slika 5.14: Množica filtriranih podatkov z mejnimi točkami, omejene z modelom IN-FUMO

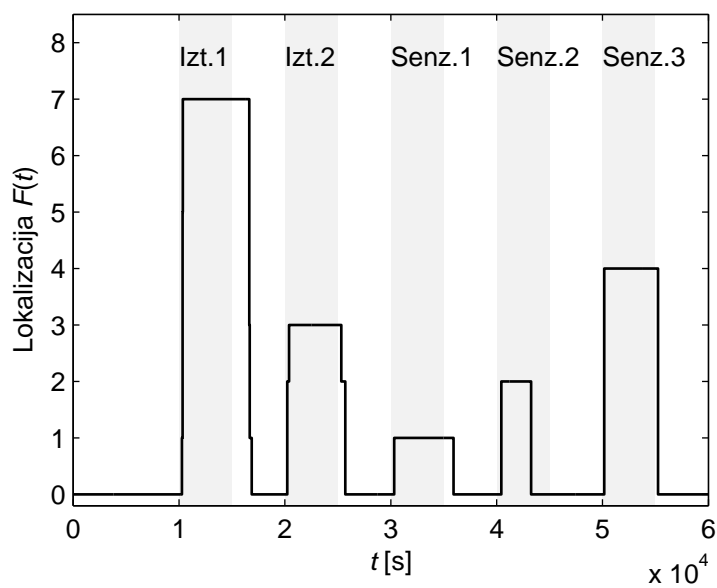
Tabela 5.1: Incidenčna matrika za dan nabor napak

Napaka	Iztok 1	Iztok 2	Senzor 1	Senzor 2	Senzor 3
f_1	1	1	1	0	0
f_2	1	1	0	1	0
f_3	1	0	0	0	1
f	7	3	1	2	4

razvidno, da sistem za zaznavanje napak odkrije vseh pet napak v ustreznih časovnih obdobjih. Poleg tega so vidne tudi kombinacije, če se osredotočimo na alarme vseh treh residualov v posameznih obdobjih napak. Do lažnih alarmov ni prišlo, prav tako ni bilo spregledanih alarmov. Časi zaznave so podani v tabeli 5.2. Vsi časi so v sekundah. Del sistema, ki skrbi za lokalizacijo napak, potem preko odločitvenih funkcij in incidenčne matrike izračuna kode napak, ki so predstavljene na sliki 5.16. Sistem je robusten na negotovosti v procesu z izjemo kratkih period ob nastopih in koncu napak, ko lahko zaradi različnih širin pasov zaupanja pride do kratkotrajne napačne lokalizacije.



Slika 5.15: Rezultati zaznavanja napak - residuali 1, 2 in 3



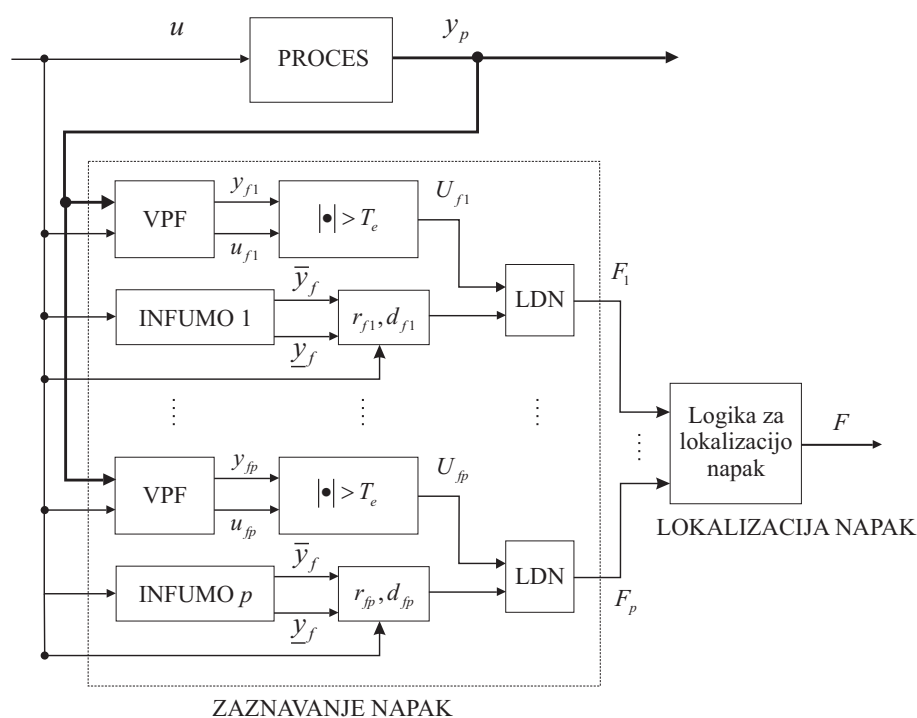
Slika 5.16: Rezultati lokalizacije napak

Tabela 5.2: Časi zaznave

Napaka	Iztok 1	Iztok 2	Senzor 1	Senzor 2	Senzor 3
r_{f1}	290	410	340	x	x
r_{f2}	370	230	x	450	x
r_{f3}	340	x	x	x	180

5.4 Zaznavanje in lokalizacija napak na podlagi visokopasovnega filtriranja

V drugem pristopu za obdelavo podatkov uporabimo visokopasovni filter. Oglejmo si shemo zaznavanja in lokalizacije napak na sliki 5.17. Blok "VPF" predstavlja visoko-



Slika 5.17: Zaznavanje in lokalizacija napak z visokopasovnim filtrom

pasovni filter, katerega uporabimo za določanje ustaljenega stanja procesa. Glede na strategijo z nizkopasovnim filtrom je razlika v tem, da pri pridobivanju podatkov za pas zaupanja tu ne uporabimo filtrirane podatke, ampak dejanske meritve, vzorčene v ustaljenem stanju. Pri obdelavi podatkov na ta način izločimo vse podatke, pridobljene v prehodnih pojavih, seveda pa to pomeni, da pri zaznavanju napak ne moremo jamčiti

za točnost zaznave, ko je proces v prehodnem pojavu.

Za oceno ustaljenega stanja uporabimo filtriranje vhoda, izhoda ali pa obeh, kot je to prikazano na sliki 5.17. Privzemimo visokopasovni filter prvega reda s prenosno funkcijo

$$Q(p) = \frac{p}{p + k_f}. \quad (5.27)$$

Visoka vrednost filtriranega signala nakazuje prehodni pojav, ko pa se proces bliža ustaljenemu stanju, gre vrednost proti nič. Vrednost k_f določa hitrost odziva filtra - enotina stopnica na vhodu filtra bo dala izhod $y_f(t) = e^{-k_f t}$. Test ustaljenega stanja predstavimo z enačbo

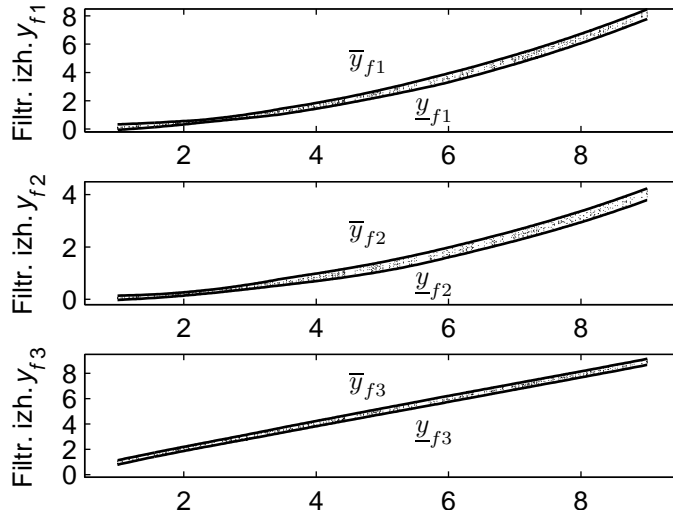
$$U_f = \begin{cases} 1, & (y_f(t) > T_e) \wedge (u_f(t) > T_e), \\ 0, & (y_f(t) \leq T_e) \vee (u_f(t) \leq T_e), \end{cases} \quad (5.28)$$

kjer je T_e konstanta, ki označuje mejo med prehodnim pojavom in ustaljenim stanjem. Navadno T_e in k_f določimo s poskušanjem glede na eksperimentalne podatke procesa. Pri pridobivanju podatkov v pas zaupanja uvrstimo vzorce vhoda in izhoda, kjer je vrednost spremenljivke U_f enaka nič.

Pri filtriranju merjene izhodne veličine navadno nastopijo težave z merilnim šumom, ki v ustaljenem stanju povzroča nagle in relativno velike spremembe filtriranega signala. Zato moramo vrednost meje T_e postaviti tako visoko, da je filtrirani izhod v ustaljenem stanju ne presega. Visokopasovnemu filtru pa lahko dodamo tudi nizkopasovni filter z nizko časovno konstanto, ki služi izključno za filtriranje motenj. Kombinacija filtriranja vhoda in izhoda pa odpravlja težavo, ki se pojavi pri hitrih spremembah vhoda. Ker se izhod odziva počasneje, je vrednost njegovega odvoda lahko še pod mejo T_e , medtem ko gre vhod že proti novi delovni točki. V pas zaupanja bi tako prišle nedovoljene točke zunaj nominalnega intervala (ang. outliers). Hkratno preverjanje odvoda vhoda in izhoda to prepreči.

Meje pasu zaupanja modeliramo z modelom INFUMO. Če imamo več izhodov, kot je prikazano na sliki 5.17, za vsak izhod računamo spremenljivko U_{fi} in oblikujemo model INFUMO. Pri zaznavanju napak računamo normirani residual po enačbi (5.14), kjer namesto filtriranega izhoda y_{fi} uporabimo kar dejanski izhod y_i . Ko je proces v prehodnem pojavu, ocena napake ni verodostojna in je ne smemo upoštevati, kar pomeni nezvezno odločitveno funkcijo d_f . Zato moramo residual r_{fi} in funkcijo d_{fi} kombinirati z vrednostmi pripadajoče ocene ustaljenega stanja U_{fi} . S predlagano logiko določanja napake (blok "LDN") dobimo zvezno funkcijo d_{fi} .

Najprej določimo časovne trenutke, v katerih se zgodijo spremembe $U_f = 0 \rightarrow 1$



Slika 5.18: Množica podatkov (VPF) z mejnimi točkami, omejene z modelom INFUMO

(začetek prehodnega pojava) in $U_f = 1 \rightarrow 0$ (konec prehodnega pojava). Označimo i -ti prehod $0 \rightarrow 1$ s t_i^\uparrow in nasprotni prehod s t_i^\downarrow . Funkcijo napake določimo po naslednji enačbi:

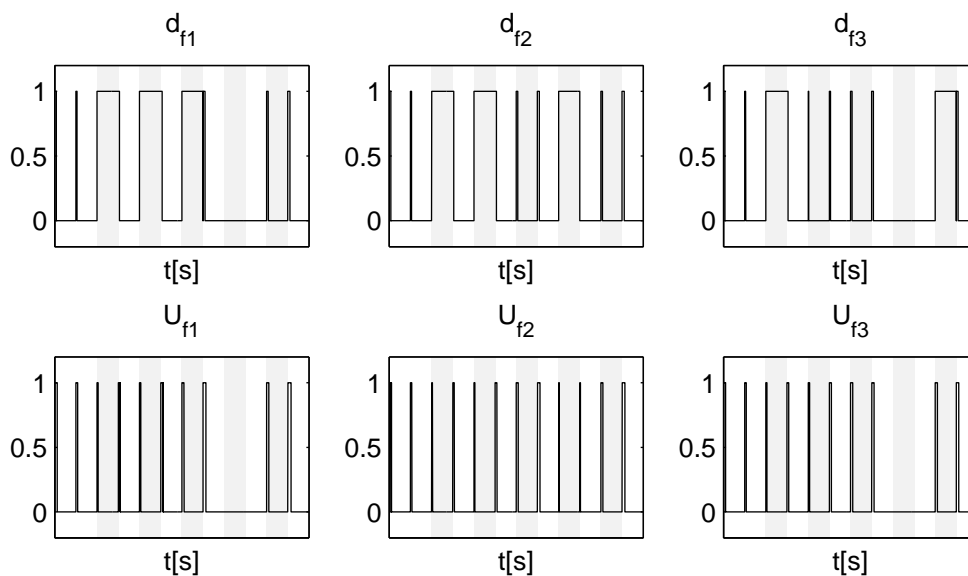
$$F_i(t) = \begin{cases} d_{fi}(t_i^\uparrow), & t_i^\uparrow \leq t < t_i^\downarrow, \quad i = 1, 2, \dots \\ d_{fi}(t), & t_{i-1}^\downarrow \leq t < t_i^\uparrow, \quad i = 1, 2, \dots \end{cases}, \quad (5.29)$$

kjer s $t_0^\uparrow = t_0$ označimo začetni čas. Na ta način v prehodnem pojavu zadržimo zadnje vrednosti d_{fi} iz ustaljenega stanja. Čeprav tako v zaznavo napak vnesemo zakasnitev $t_i^\downarrow - t_i^\uparrow$, predlagana strategija pripomore k odpravljanju lažnih alarmov.

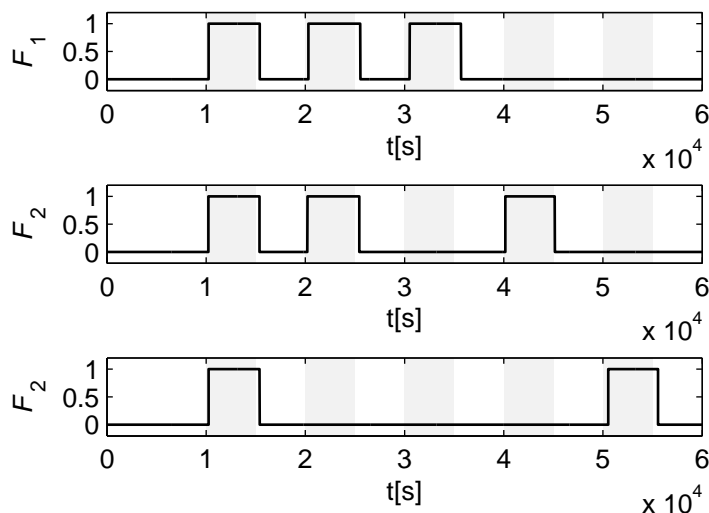
Lokalizacijo napak na enak način kot v razdelku 5.3.2 izvedemo z računanjem kode napake in incidenčno matriko.

5.4.1 Lokalizacija napak za model dveh shranjevalnikov

Oglejmo si predlagano metodo na primeru dveh shranjevalnikov. Uporabimo enake meritve kot v razdelku 5.3.2. Filtriramo s filtrom prvega reda 5.27, kjer določimo $k_f = 0,1$. Zaradi izhodnih motenj dodamo še prefilter s prenosno funkcijo $G_{pf}(s) = 1/(50s + 1)$. Za mejo prehodnega pojava določimo $T_e = 0,01$. Uporabimo tudi enake oblike modelov INFUMO. Slika 5.18 prikazuje nove pasove zaupanja in aproksimacije meja. Uporabimo tudi enak testni signal in predpostavimo enake napake v sistemu. Slika 5.19 prikazuje odločitvene funkcije d_{fi} in pripadajoče vrednosti spremenljivk U_{fi} . Vidimo, da se pri odločitvenih funkcijah poleg napak pojavljajo tudi "lažni" alarmi, ki zaznamujejo obdobja prehodnega pojava in sovpadajo z vrednostmi $U_{fi} = 1$. Brez



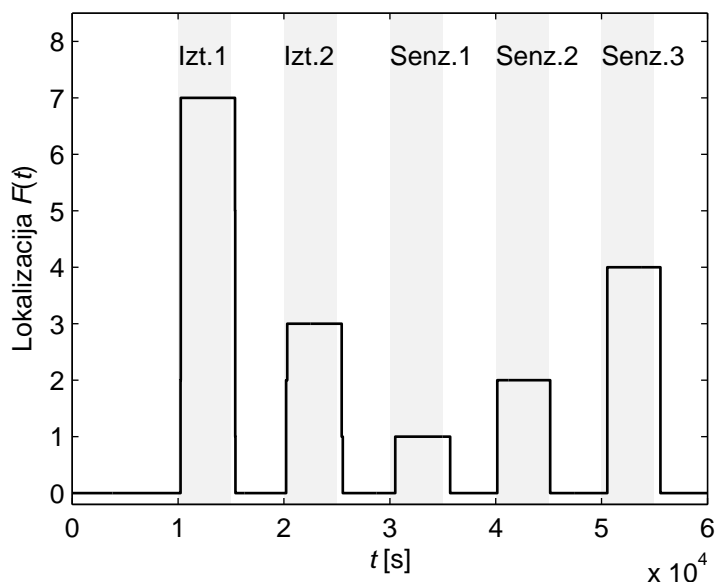
Slika 5.19: Rezultati zaznavanja napak (VPF) - zgoraj: odločitvene funkcije, spodaj: test prehodnega pojava



Slika 5.20: Rezultati zaznavanja napak (VPF) - funkcije napak

logike za določanje napak bi sistem v teh trenutkih javil alarm. Po uporabi logike pa se oblikujejo funkcije napak F_i , kot jih prikazuje slika 5.20. Rezultati lokalizacije (časovni potek kod napak) so prikazani na sliki 5.21, v tabeli 5.3 pa so zbrani časi zaznav napak. Če rezultate primerjamo z rezultati s slike 5.16, lahko povzamemo naslednje ugotovitve:

- oba sistema pravilno lokalizirata vse napake;



Slika 5.21: Rezultati lokalizacije napak (VPF)

Tabela 5.3: Časi zaznave (VPF)

Napaka	Iztok 1	Iztok 2	Senzor 1	Senzor 2	Senzor 3
r_{f1}	270	320	530	x	x
r_{f2}	240	220	x	170	x
r_{f3}	270	x	x	x	550

- sistem z VPF načeloma zagotavlja hitrejšo zaznavo napak - šest od osmih napak je odkril prej kot sistem z NPF;
- sistem z VPF zagotavlja bolj natančno oceno trajanja napak, kar se vidi v usklajenosti signalov napak in senčenih obdobj (trajanje napak).

Zadnjo alinejo potrjuje tabela 5.4, ki primerja zaznave časov prenehanja napak. Očitno je, da je pri metodi z VPF zakasnitev predvidljiva (med 15 in 70 vzorci), medtem ko pri metodi z NPF preseže 180 vzorcev ali pa tudi 170 vzorcev prehitro oceni konec napake.

Tabela 5.4: Primerjava časov zaznave prenehanja napake

Napaka	$r_{f1,NPF}$	$r_{f1,VPF}$	$r_{f2,NPF}$	$r_{f2,VPF}$	$r_{f3,NPF}$	$r_{3,VPF}$
Iztok 1	1860	410	1170	380	1130	400
Iztok 2	320	550	710	450	x	x
Senzor 1	940	700	x	x	x	x
Senzor 2	x	x	-1750	150	x	x
Senzor 3	x	x	x	x	250	560

6. Prediktivno vodenje na podlagi modela Wiener-INFUMO v zveznem času

6.1 Uvodni pojmi

6.1.1 Strategija vodenja

Vzemimo nelinearni časovno zvezni proces

$$\begin{aligned}\dot{\mathbf{x}}_p(t) &= f(\mathbf{x}_p(t), u(t)) \\ y_p(t) &= g(\mathbf{x}_p(t)) \\ y_p(0) &= y_{p0}, \quad y_p(T) = y_{pT},\end{aligned}\tag{6.1}$$

kjer sta $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ in $g : \mathbb{R}^n \rightarrow \mathbb{R}$ gladki funkciji, $\mathbf{x}_p \in \mathbb{R}^n$ je vektor n spremenljivk stanj, $u \in \mathbb{R}$ je vhod, $y_p \in \mathbb{R}$ pa izhod procesa. Optimalno vodenje takega sistema je v splošnem definirano kot

$$\min_{u \in \mathcal{U}} J(\mathbf{x}_p, y_p, t) = \min_{u \in \mathcal{U}} \int_0^T F(\mathbf{x}_p, y_p, t) dt,\tag{6.2}$$

kjer je $F \in \mathbb{R}^+$ kriterijska funkcija, ki zadosti kriteriju optimalnosti, \mathcal{U} pa predstavlja množico dopustnih vhodov v proces (regulirnih signalov). Povedano z drugimi besedami, cilj je načrtati regulator, ki asimptotično stabilizira zaprtozančni sistem, tako da izhod procesa $y_p(t)$ optimalno sledi predpisani referenčni trajektoriji $y_r(t)$ glede na dani performančni indeks J (ang. performance index). Treba pa se je zavedati, da problem (6.2) pri nelinearnih časovno zveznih sistemih vodi v reševanje Hamilton-Jacobi-Bellmanovih parcialnih diferencialnih enačb. V delih Bertsekasa [10] in Chena s sodelavci [16] je navedeno, da rešitev sistema enačb HJB navadno računamo numerično. Tak pristop pa je po pravilu računsko zelo zahteven in ni primeren za praktične izvedbe vodenja. Kot alternativo uporabimo koncept vodenja na podlagi premičnega horizonta (ang. moving-horizon control), kot je npr. predlagano v delih [82], [17] in [16]. Glavna ideja je računanje optimalnega zaporedja regulirnih signalov, ki v danem časovnem trenutku minimizirajo dani performančni indeks, ki vsebuje napovedi odprtozančnega odziva modela procesa in referenčne trajektorije za določen horizont. Začetni pogoji

so referenca, izhod modela in izhod procesa v danem trenutku t . Dobljena rešitev je optimalna glede na dani odprtozančno kriterijsko funkcijo. Ko uporabimo prvega iz zaporedja izračunanih optimalnih signalov za vhod, premaknemo horizont v naslednji časovni trenutek in ponovimo računanje.

6.1.2 Napoved izhoda procesa z modelom Wiener-INFUMO

Če je obravnavani sistem nelinearen in ima hkrati še negotove parametre in celo negotovo strukturo, to navadno predstavlja resen problem za modeliranje in vodenje na podlagi modela. Videli smo, da lahko nekatere oblike nelinearnih sistemov učinkovito obravnavamo z modelom Wiener-INFUMO. Njegova posebna struktura pa omogoča tudi zelo elegantno izpeljavo kompaktnega zapisa napovedi izhoda modela, kar omogoča neposredno uporabo v prediktivnem vodenju na podlagi modela procesa.

Zapišimo proces z enačbama

$$\begin{aligned}\dot{\mathbf{x}}_p(t) &= f(\mathbf{x}_p(t), u(t), \mathbf{q}), \\ y_p(t) &= g(\mathbf{x}_p(t), \mathbf{q}),\end{aligned}\tag{6.3}$$

kjer s \mathbf{q} označimo negotovosti parametrov v obliki hiperpravokotnika $q_i \in [\underline{q}_i, \bar{q}_i]$. Predpostavimo, da je izhod procesa moč opisati z intervalom

$$y_p(t) \in [g(\mathbf{x}_p(t), \mathbf{q}), \bar{g}(\mathbf{x}_p(t), \mathbf{q})].\tag{6.4}$$

Privzeli bomo Wienerjev model z dinamičnim delom, zapisanim v prostoru stanj, ki mu sledi statični model INFUMO

$$\begin{aligned}\dot{\mathbf{x}}(t) &= A\mathbf{x}(t) + Bu(t), \\ v(t) &= C\mathbf{x}(t), \\ \bar{y}(t) &= \bar{h}(v(t)), \\ \underline{y}(t) &= \underline{h}(v(t)),\end{aligned}\tag{6.5}$$

kjer so $A \in \mathbb{R}^n \times \mathbb{R}^n$, $B \in \mathbb{R}^n$ in $C \in \mathbb{R}^n$ matrike zapisa v prostoru stanj, $\bar{h}, \underline{h}: \mathbb{R} \rightarrow \mathbb{R}$ zaznamujeta zgornjo in spodnjo mejno funkcijo statičnega modela INFUMO, $\bar{y}, \underline{y} \in \mathbb{R}$ pa sta izhoda procesa, ki oblikujeta interval dopustnih izhodov. Spremenljivka $v(t) \in \mathbb{R}$ predstavlja vmesno spremenljivko Wienerjevega modela, ki v procesu ni nujno merljiva oziroma včasih celo nima jasne fizikalne podlage. Opomnimo naj tudi, da funkciji \underline{h}, \bar{h} nista enaki \underline{g}, \bar{g} iz (6.4), ker v splošnem statični del modela Wiener-INFUMO zaobjame tudi vplive nelinearnosti stanj na izhod procesa.

Ker obravnavamo proces z modelom, kjer lahko v vsakem trenutku z gotovostjo določimo le interval izhodov, bomo cilj vodenja prilagodili na ta način, da naj bo referenca vedno znotraj intervala izhodov, ki ga določa model. To zagotovimo tako, da za izhod modela uporabimo srednjo vrednost intervala $\tilde{y}(t) = \frac{1}{2}(\bar{y}(t) + \underline{y}(t))$. Če upoštevamo zapis modela INFUMO

$$\begin{aligned}\bar{y}(t) &= \bar{h}(v(t)) = \boldsymbol{\beta}(v(t))\bar{\boldsymbol{\Theta}}\mathbf{x}_c(v(t)), \\ \underline{y}(t) &= \underline{h}(v(t)) = \boldsymbol{\beta}(v(t))\underline{\boldsymbol{\Theta}}\mathbf{x}_c(v(t)),\end{aligned}\tag{6.6}$$

lahko izhod modela podamo kot

$$\tilde{y}(t) = \tilde{h}(t) = \beta(v(t))\tilde{\boldsymbol{\Theta}}\mathbf{x}_c(v(t)), \quad \tilde{\boldsymbol{\Theta}} = \frac{1}{2}(\bar{\boldsymbol{\Theta}} + \underline{\boldsymbol{\Theta}}).\tag{6.7}$$

Prediktivno vodenje na podlagi modela v splošnem zahteva izračun napovedi izhoda modela za dani horizont v prihodnosti. V zvezni časovni domeni in strategiji premičnega horizonta napoved izhoda ob času $t + \tau$, izračunano ob času t , izrazimo z razširitvijo v Taylorjevo vrsto

$$\tilde{y}(t + \tau|t) = \tilde{y}(t) + \tau \frac{d\tilde{y}(t)}{dt} + \frac{\tau^2}{2!} \frac{d^2\tilde{y}(t)}{dt^2} + \dots + \frac{\tau^i}{i!} \frac{d^i\tilde{y}(t)}{dt^i} + \dots\tag{6.8}$$

Napoved aproksimiramo tako, da upoštevamo prvih n_y členov vrste, ki jih zapišemo v obliki skalarnega produkta

$$\tilde{y}(t + \tau|t) = \boldsymbol{\Gamma}^T(\tau)\mathbf{Y}(t).\tag{6.9}$$

Vektorja $\boldsymbol{\Gamma}^T \in \mathbb{R}^{n_y+1}$ in $\mathbf{Y} \in \mathbb{R}^{n_y+1}$ sta podana na naslednji način:

$$\boldsymbol{\Gamma}^T(\tau) = \left[1 \quad \tau \dots \frac{\tau^i}{i!} \dots \frac{\tau^{n_y}}{n_y!} \right],\tag{6.10}$$

$$\mathbf{Y}(t) = \left[\tilde{y}(t) \quad \tilde{y}^{[1]}(t) \dots \tilde{y}^{[i]}(t) \dots \tilde{y}^{[n_y]}(t) \right]^T,\tag{6.11}$$

kjer je n_y red izhoda, $\tilde{y}^{[i]}(t)$ pa zaznamuje i -ti odvod $\tilde{y}(t)$ po času t . Da bi napoved zapisali v kompaktni obliki, moramo izpeljati izraz za splošni odvod izhoda modela. Pri izpeljavi bomo opustili označevanje odvisnosti vektorjev $\boldsymbol{\beta}$ in \mathbf{x}_c od $v(t)$. Prvi odvod izhoda je

$$\frac{d\tilde{y}(t)}{dt} = \frac{d\tilde{h}(v)}{dv} \cdot \dot{v}(t),\tag{6.12}$$

drugi odvod pa postane

$$\frac{d^2\tilde{y}(t)}{dt^2} = \frac{d^2\tilde{h}(v)}{dv^2} \cdot (\dot{v}(t))^2 + \frac{d\tilde{h}(v)}{dv} \cdot \ddot{v}(t).\tag{6.13}$$

Ker drugi odvod izhodne funkcije po v (prvi člen desne strani enačbe (6.13)) izrazimo z

$$\frac{d^2 \tilde{h}(v)}{dv^2} = \frac{d^2 \boldsymbol{\beta}^T}{dv^2} \tilde{\boldsymbol{\Theta}} \mathbf{x}_c + 2 \frac{d \boldsymbol{\beta}^T}{dv} \tilde{\boldsymbol{\Theta}} \frac{d \mathbf{x}_c}{dv} + \boldsymbol{\beta}^T \tilde{\boldsymbol{\Theta}} \frac{d^2 \mathbf{x}_c}{dv^2}, \quad (6.14)$$

je očitno naslednje: tudi če vzamemo trikotne pripadnostne funkcije in linearne izhodne funkcije, ki povzročijo $\frac{d^i \boldsymbol{\beta}^T}{dv^i}$, $\frac{d^i \mathbf{x}_c}{dv^i} = 0$ za $i \geq 2$, da se v izraze za višje odvode prenese odvod vmesnega člena $2 \frac{d \boldsymbol{\beta}^T}{dv} \tilde{\boldsymbol{\Theta}} \frac{d \mathbf{x}_c}{dv}$ in analitična obravnava postane prekompleksna. Zato predpostavimo, da gredo členi $(\dot{v}(t))^k$, $k \geq 2$ proti 0. V delu [88] je bilo na praktičnem primeru pokazano, da taka poenostavitev ne prinese zmanjšanja kvalitete aproksimacije. Izraz za i -ti odvod $\tilde{y}(t)$ se tako poenostavi v

$$\begin{aligned} \tilde{y}^{[i]}(t) &= \frac{d \tilde{h}}{dv} \cdot \frac{d^i v}{dt^i} = \\ &= \left(\frac{d \boldsymbol{\beta}^T}{dv} \tilde{\boldsymbol{\Theta}} \mathbf{x}_c + \boldsymbol{\beta}^T \tilde{\boldsymbol{\Theta}} \frac{d \mathbf{x}_c}{dv} \right) [\mathbf{C} \mathbf{A}^i x(t) + [\mathbf{C} \mathbf{A}^{i-1} \mathbf{B} \dots \mathbf{C} \mathbf{B}] \mathbf{U}(t)], \end{aligned} \quad (6.15)$$

kjer je $\mathbf{U}(t)$ vektor odvodov vhodnega signala

$$\mathbf{U}(t) = \left[u(t) \ u^{[1]}(t) \dots u^{[i]}(t) \right]^T. \quad (6.16)$$

Definirajmo red vhoda na način, kot je predstavljen v definiciji 6.1.

Definicija 6.1. Red vhoda v prediktivnem vodenju v zveznem času je n_u , če velja naslednje: $u^{[n_u]}(t + \tau) \neq 0$, $\forall \tau \in [0, T]$ in $u^{[i]}(t + \tau) = 0$, $\forall i > n_u$, $\tau \in [0, T]$, kjer z $u^{[n_u]}(t + \tau)$ izrazimo n_u -ti odvod napovedi vhoda $u(t + \tau)$ po času τ . Red vhoda določa množico \mathcal{U} kot domeno dopustnih optimalnih vhodnih funkcij po strategiji premičnega horizonta in s tem določa omejitve signala $u(t + \tau)$.

Opomba 6.1. V predlagani metodi sta red izhoda n_y in red vhoda n_u dva načrtovalska parametra. Avtorji v nekaterih delih podajajo določene omejitve glede parametra n_y [16]. Če z ρ označimo relativni red procesa, mora biti n_y vsaj reda $n_u + \rho$, če želimo, da n_u -ti odvod vhodnega signala nastopi v napovedi izhoda procesa; povedano drugače - $n_y \geq n_u + \rho$. To je posledica dejstva, da n_u definira najvišji odvod vhodnega signala, ki vpliva na proces v predikcijskem horizontu.

Vektor odvodov vhoda $\mathbf{U}(t)$ reda n_u tako postane

$$\mathbf{U}(t) = \left[u(t) \ u^{[1]}(t) \dots u^{[n_u]}(t) \right]^T. \quad (6.17)$$

Označimo s

$$q_f(v) = \left(\frac{d \boldsymbol{\beta}^T}{dv} \tilde{\boldsymbol{\Theta}} \mathbf{x}_c + \boldsymbol{\beta}^T \tilde{\boldsymbol{\Theta}} \frac{d \mathbf{x}_c}{dv} \right) \quad (6.18)$$

smerni odvod statične izhodne karakteristike, odvisen od trenutne vrednosti vmesne spremenljivke $v(t)$. Vstavimo prvih n_y odvodov (6.15) v enačbo (6.8), preuredimo jo v obliko iz enačbe (6.9) in zapišimo napoved izhoda modela $\tilde{y}(t + \tau|t)$ ob času t kot

$$\tilde{y}(t + \tau|t) = \mathbf{\Gamma}^T [\mathbf{P}\tilde{y}(t) + q_f(v)\mathbf{K}_q x(t) + q_f(v)\mathbf{K}_h \mathbf{U}(t)], \quad (6.19)$$

kjer so matrice $\mathbf{P} \in \mathbb{R}^{n_y+1}$, $\mathbf{K}_q \in \mathbb{R}^{n_y+1} \times \mathbb{R}^n$ in $\mathbf{K}_h \in \mathbb{R}^{n_y+1} \times \mathbb{R}^{n_u+1}$ definirane kot

$$\mathbf{P} = [1 \ 0 \ \dots \ 0]^T, \quad (6.20)$$

$$\mathbf{K}_q = [\mathbf{0} \ (\mathbf{CA})^T \ (\mathbf{CA}^2)^T \ \dots \ (\mathbf{CA}^{n_y})^T]^T \quad (6.21)$$

in

$$\mathbf{K}_h = \begin{bmatrix} 0 & \dots & \dots & 0 \\ \mathbf{CB} & 0 & \dots & 0 \\ \mathbf{CAB} & \mathbf{CB} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{CA}^{n_y-1}\mathbf{B} & \mathbf{CA}^{n_y-2}\mathbf{B} & \dots & \mathbf{CA}^{n_y-1-n_u}\mathbf{B} \end{bmatrix}. \quad (6.22)$$

Vrednosti matrik \mathbf{K}_q in \mathbf{K}_h lahko izračunamo vnaprej, saj sta odvisni samo od matrik linearne dinamike modela. V vsakem računskem koraku je treba izračunati le vrednost parametra $q_f(v) \in \mathbb{R}$.

6.2 Izpeljava prediktivnega regulacijskega zakona

Kriterijska funkcija za izpeljavo prediktivnega regulirnega zakona navadno vsebuje razliko med napovedjo referenčne trajektorije in izhoda modela

$$J = \int_0^T \|y_r(t + \tau) - y(t + \tau|t)\|^2 d\tau. \quad (6.23)$$

Razliko med izhodoma modela in procesa se potem kompenzira z uvedbo modela motenj. V našem primeru želimo v izpeljavo eksplicitno vnesti merjeni izhod procesa $y_p(t)$. Podobna zasnova je pri razvoju prediktivnega funkcionalnega regulatorja [131], [129]; razlikuje se v tem, da v diskretnem času lahko horizont napovedi skrčimo na eno samo točko v prihodnosti, kjer naj bosta referenčna trajektorija in izhod procesa enaka. V zveznem času je treba upoštevati integral pogreška na celotnem horizontu, zato kriterijsko funkcijo ustrezno prilagodimo. Predpostavimo, da je referenčna dinamika podana v prostoru stanj z \mathbf{A}_r , \mathbf{B}_r in \mathbf{C}_r . V trenutku t (začetku horizonta) imamo na voljo referenco $w(t)$, meritev izhoda $y_p(t)$ in spremenljivke modela $x(t)$, $v(t)$ ter $\tilde{y}(t)$. Izhod

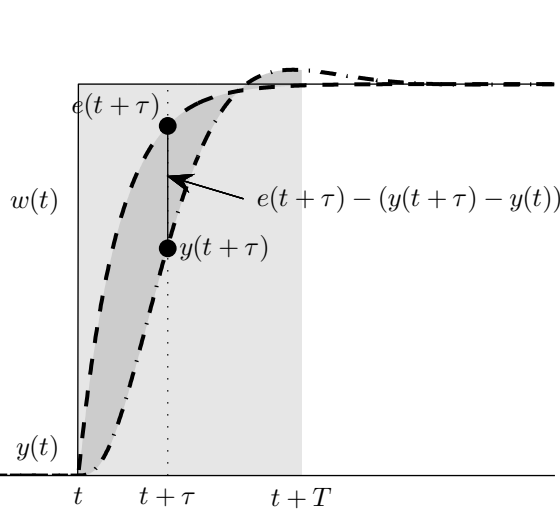
procesa sledi referenci, če predpostavimo, da gre pogrešek proti nič; to zagotovimo s tem, da določimo napoved pogreška z naslednjim izrazom:

$$e(t + \tau) = \mathbf{\Gamma}^T \mathbf{r} (w(t) - y_p(t)). \quad (6.24)$$

Vektor $\mathbf{r} = [0, \mathbf{C}_r \mathbf{B}_r, \mathbf{C}_r \mathbf{A}_r \mathbf{B}_r, \dots, \mathbf{C}_r \mathbf{A}_r^{n_y - 1} \mathbf{B}_r]^T$ je vektor parametrov Markova, ki določa impulzni odziv referenčnega modela v prostoru stanj. Pomnožen z vektorjem $\mathbf{\Gamma}^T$ določa, da pogrešek $w(t) - y_p(t)$ v horizontu eksponentno upada proti nič. Če se $w(t)$ v horizontu ne spreminja, to pomeni, da napovemo, da se bo izhod procesa eksponentno približeval $w(t + \tau) = w(t)$. Cilj vodenja je zdaj poiskati tak vhod $u(t)$, da se bo prirast izhoda modela $\tilde{y}(t + \tau) - \tilde{y}(t)$ v horizontu čimbolj približal napovedi referenčnega signala $e(t + \tau)$, ali drugače, da bo pogrešek

$$\epsilon(t, \tau) = e(t + \tau) - (\tilde{y}(t + \tau|t) - \tilde{y}(t)). \quad (6.25)$$

minimalen na časovnem intervalu $\tau \in [0, T]$. Strategijo predstavlja slika 6.1. Regulirni



Slika 6.1: Izračun kriterijske funkcije

zakon dobimo z minimizacijo kvadratične kriterijske funkcije

$$J = \int_0^T \epsilon(t, \tau)^T \epsilon(t, \tau) d\tau. \quad (6.26)$$

Če v (6.26) vstavimo napoved izhoda modela (6.19), dobimo

$$J(\mathbf{U}, v) = \int_0^T \left((w - y_p)^T \mathbf{r}^T - \mathbf{U}^T \mathbf{K}_h^T q_f(v)^T - x^T \mathbf{K}_q^T q_f(v)^T \right) \mathbf{\Gamma} \times \mathbf{\Gamma}^T (\mathbf{r} (w - y_p) - q_f(v) \mathbf{K}_h \mathbf{U} - q_f(v) \mathbf{K}_q \mathbf{x}) d\tau \quad (6.27)$$

Omenimo, da sta člena $\mathbf{\Gamma}^T \mathbf{P} \tilde{y}(t)$ iz (6.19) in $\tilde{y}(t)$ iz (6.25) enaka, zato izhod modela ob času t ne nastopa več v kriterijski funkciji.

Minimizacijo kriterijske funkcije izvedemo na standarden način z odvajanjem izraza po \mathbf{U} in izenačenjem z nič,

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{U}} = -2q_f(v) \mathbf{K}_h^T \left[\int_0^T \mathbf{\Gamma} \mathbf{\Gamma}^T \mathbf{r}(w - y_p) d\tau - \right. \\ \left. - \int_0^T (\mathbf{\Gamma} \mathbf{\Gamma}^T q_f(v) \mathbf{K}_h \mathbf{U} + \mathbf{\Gamma} \mathbf{\Gamma}^T q_f(v) \mathbf{K}_q \mathbf{x}) d\tau \right] = 0. \end{aligned} \quad (6.28)$$

Ker so spremenljivke $q_f(v)$, w , y_p in \mathbf{x} ovrednotene samo na začetku horizonta, niso odvisne od τ ; tako vidimo, da je edino vektor $\mathbf{\Gamma}$ odvisen od τ . Definirajmo matriko $\bar{\mathbf{\Gamma}} \in \mathbb{R}^{n_y+1} \times \mathbb{R}^{n_y+1}$ kot integral produkta

$$\bar{\mathbf{\Gamma}} = \int_0^T \mathbf{\Gamma} \mathbf{\Gamma}^T d\tau. \quad (6.29)$$

Ker je splošni člen simetrične matrike $\mathbf{\Gamma} \mathbf{\Gamma}^T$ enak $T^{i-1+j-1}/((i-1)!(j-1)!)$, enačbo (6.29) zapišemo kot

$$\bar{\mathbf{\Gamma}} = \begin{bmatrix} \gamma_{(1,1)} & \cdots & \gamma_{(1,n_y+1)} \\ \vdots & \ddots & \vdots \\ \gamma_{(n_y+1,1)} & \cdots & \gamma_{(n_y+1,n_y+1)} \end{bmatrix}, \quad (6.30)$$

kjer velja

$$\gamma_{(i,j)} = \frac{1}{(i+j-1)(i-1)!(j-1)!} T^{i+j-1} \quad (6.31)$$

za vsak $i, j = 1, \dots, n_y + 1$. Enačbo (6.28) preoblikujemo v

$$\frac{\partial J}{\partial \mathbf{U}} = -2q_f(v) \mathbf{K}_h^T \bar{\mathbf{\Gamma}} [\mathbf{r}(w - y_p) - q_f(v) \mathbf{K}_h \mathbf{U} - q_f(v) \mathbf{K}_q \mathbf{x}] = 0 \quad (6.32)$$

in iz njega izrazimo vektor odvodov vhodnega signala

$$\mathbf{U} = q_f(v)^{-1} (\mathbf{K}_h^T \bar{\mathbf{\Gamma}} \mathbf{K}_h)^{-1} \mathbf{K}_h^T \bar{\mathbf{\Gamma}} [\mathbf{r}(w - y_p) - q_f(v) \mathbf{K}_q \mathbf{x}]. \quad (6.33)$$

Vektor \mathbf{U} nosi informacijo o amplitudi in obliki optimalnega vhodnega signala, ki na intervalu $[t, t + T]$ minimizira kriterijsko funkcijo 6.26. Vendar pa lahko uporabimo samo $u(t)$, torej prvi element vektorja. Zato vzamemo prvo vrstico matrike $(\mathbf{K}_h^T \bar{\mathbf{\Gamma}} \mathbf{K}_h)^{-1} \mathbf{K}_h^T \bar{\mathbf{\Gamma}} \in \mathbb{R}^{n_u+1} \times \mathbb{R}^{n_y+1}$ in jo označimo s $\boldsymbol{\kappa}^T$. Regulirni zakon je zdaj podan z naslednjo enačbo:

$$u(t) = \frac{1}{q_f(v)} \boldsymbol{\kappa}^T [\mathbf{r}(w - y_p) - q_f(v) \mathbf{K}_q \mathbf{x}]. \quad (6.34)$$

Izraz (6.34) lahko razčlenimo na

- spremenljivo ojačenje $\frac{1}{q_f(v)}$, ki kompenzira nelinearnost na izhodu y_p ;
- fiksno ojačenje regulatorja κ^T ;
- implicitni referenčni model v \mathbf{r} , ki določa potek izhodnega pogoška;
- korekcija iz stanj modela \mathbf{x} preko spoznavnostne matrike \mathbf{K}_q .

Predlagani regulirni zakon je zelo primeren za praktično uporabo, ker je večina računskega bremena na sprotne računanju $q(v)$. Vrednosti κ , \mathbf{r} in \mathbf{K}_q lahko izračunamo vnaprej.

6.2.1 Obravnava stabilnosti

Če se osredotočimo na člen $\epsilon(t, \tau)$ v kriterijski funkciji (6.26), je jasno, da je kritična točka izračuna regulirnega signala razlika med $e(t + \tau)$ in $\tilde{y}(t + \tau|t) - \tilde{y}(t)$ v (6.25). Teoretično je možno, da gresta oba člena proti neskončnosti, razlika ostaja blizu nič, tako da je dosežen minimum in izračun u se ne spremeni. Zato moramo poleg domene dopustnih vhodnih signalov definirati tudi domeno dopustnih izhodov.

Rešitev enačbe (6.33) nam zagotovi optimalne vrednosti vhodnega signala in njegovih odvodov glede na odprtozančno kriterijsko funkcijo. Ključni del enačbe je matrika $(\mathbf{K}_h^T \bar{\Gamma} \mathbf{K}_h)^{-1}$. Da bi dobili enolično rešitev, mora eksistirati njen inverz. Vendar pa tega ne moremo zagotoviti za poljubno kombinacijo izbir načrtovalskih parametrov n_y in n_u . Poleg tega je tudi možno, da so produkti $\mathbf{C} \mathbf{A}^\zeta \mathbf{B}$, kjer je ζ odvisna od n_y in n_u , enaki nič. To je odvisno od relativnega reda procesa. Eksistenco rešitve (6.34) podaja teorem 6.1.

Teorem 6.1. *Optimalna analitična rešitev (6.34) odprtozančne kriterijske funkcije prediktivnega vodenja na podlagi modela Wiener-INFUMO eksistira za vsak neničelni horizont napovedi T tedaj in le tedaj, kadar je razlika med redoma izhoda in vhoda večja ali enaka relativnemu redu modela procesa, tj. $n_y - n_u \geq \rho$.*

Dokaz. Relativni red ρ linearnega dela modela vpliva na produkte $\mathbf{C} \mathbf{A}^\zeta \mathbf{B}$ tako, da so enaki nič za $\zeta < \rho - 1$. Produkt $\mathbf{K}_h^T \bar{\Gamma} \mathbf{K}_h$ je matrika dimenzije $(n_u + 1) \times (n_u + 1)$. Po definiciji je invertibilna le tedaj, ko je polnega ranga, tj. ima n_u linearno neodvisnih in od nič različnih stolpcev. Vzemimo primer, ko je $n_y - n_u < \rho$. Zaradi diagonalne strukture matrike \mathbf{K}_h je zadnji element $(n_u + 1)$ -ga stolpca enak nič, ker velja

$$\zeta = n_y - n_u - 1 < \rho - 1. \quad (6.35)$$

Ker je pri zadnjem elementu vrednost ζ največja, so vsi ostali elementi zadnjega stolpca enaki nič in matrika \mathbf{K}_h ima vsaj en ničelni stolpec. Ker je $\bar{\mathbf{T}}$ simetrična matrika, bo produkt $\mathbf{K}_h^T \bar{\mathbf{T}} \mathbf{K}_h$ singularen in rešitev (6.34) ne bo eksistirala. \square

6.3 Kombinacija sistema vodenja in zaznavanja napak

Zelo ugodna posledica uporabe modela Wiener-INFUMO je, da imamo v vsakem trenutku na voljo interval dopustnih vrednosti izhoda modela pri danem vходу v proces. Ker z identifikacijo parametrov modela zajamemo tudi vplive negotovosti parametrov procesa, lahko o intervalu govorimo kot o pasu zaupanja tudi v zaprtozančnem sistemu. Meji intervala lahko torej uporabimo za zaznavanje napak v zaprtozančnem delovanju sistema.

Normirani residual ovrednotimo za izhod procesa y_p in interval $[\underline{y}(t), \bar{y}(t)] = [\tilde{y}(t) - \Delta y(t), \tilde{y}(t) + \Delta y(t)]$ po enačbi

$$r(t) = \begin{cases} \frac{y_p(t) - \tilde{y}(t)}{\underline{y}(t) - \tilde{y}(t)}, & y_p(t) < \tilde{y}(t), \\ \frac{y_p(t) - \tilde{y}(t)}{\bar{y}(t) - \tilde{y}(t)}, & y_p(t) > \tilde{y}(t). \end{cases} \quad (6.36)$$

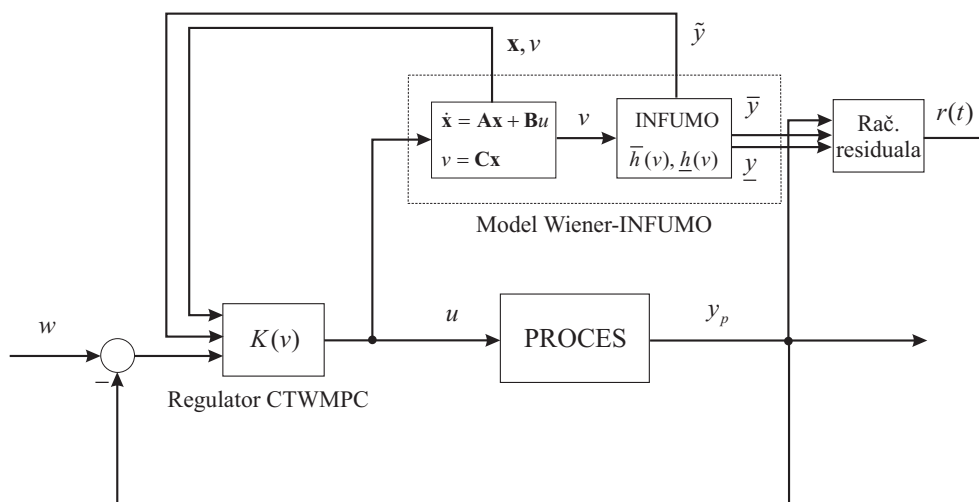
Postopek je podoben tistemu iz del [96] in [98], kjer za ocenjevanje dopustnih meja izhoda skrbi intervalni observator. Model Wiener-INFUMO vidimo kot generator adaptivnega residuala, ki prilagaja položaj intervala zaupanja glede na vrednost vmesne spremenljivke $v(t)$, ki je preko linearne dinamike modela povezana z vhomom v proces $u(t)$. Klasični pomislek pri zaprtozančnih sistemih zaznavanja napak na podlagi modela - regulator s svojim regulacijskim delovanjem zakrije vplive napak - tu ne pride do izraza, ker ob nastopu napake prilagajanje vhodnega signala premakne interval zaupanja iz nominalne lege, medtem ko izhod procesa ostane na referenčni vrednosti. Zato za zaznavnost velja enak kriterij kot v razdelku 5.2.

Shema na sliki 6.2 predstavlja celoten sistem vodenja in zaznavanja napak.

6.3.1 Primer: vodenje CTWMPC in zaznavanje napak za pH-proces

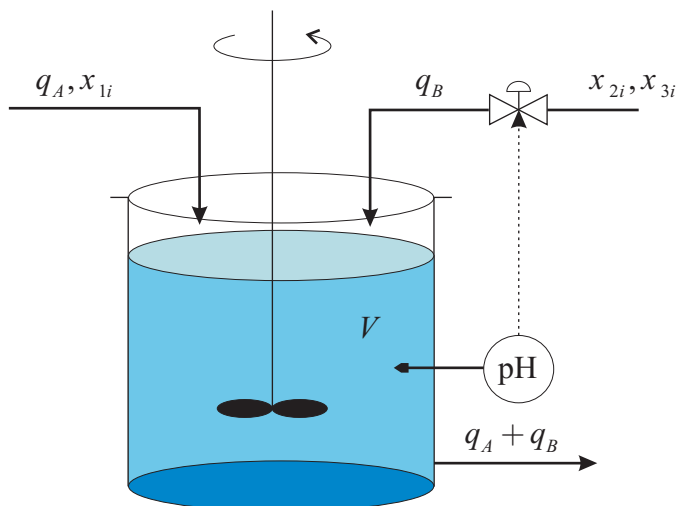
Matematični model procesa nevtralizatorja pH povzemamo iz del [39] in [11]. Obravnava reakcijo nevtralizacije močne kisline (HA) s kombinacijo močne baze (BOH) in pufra (BX). Reakcija poteka v reaktorju s stalnim mešanjem (ang. continuous stirred tank reactor, CSTR) konstantne prostornine V .

Slika 6.3 prikazuje shemo reaktorja. Kislo raztopino s koncentracijo $x_{1i}(t)$ in konstantnim volumskim pretokom q_A nevtralizira bazična raztopina s pretokom $q_B(t)$, v



Slika 6.2: Kombinirani sistem prediktivnega vodenja in zaznavanja napak na podlagi modela Wiener-INFUMO

kateri sta baza (koncentracija x_{2i}) in pufer (koncentracija x_{3i}). Zaradi hitre reakcije med kislino in bazo se kemijsko ravnotežje vzpostavi praktično v trenutku. Če pa predpostavimo, da so kislina, baza in pufer dovolj močni, potem sledi popolna disociacija vseh treh sestavin. Dinamični model dobimo z upoštevanjem ravnovesij električnega



Slika 6.3: Shema pH-procesa

naboja in mase. Predpostavimo spremenljivke stanj

$$\begin{aligned}
 x_1 &= [A^-] \quad (\text{kislina}); \\
 x_2 &= [B^+] \quad (\text{baza}); \\
 x_3 &= [X^-] \quad (\text{pufer}).
 \end{aligned}
 \tag{6.37}$$

Matematični model zapišemo na naslednji način:

$$\begin{aligned}\dot{x}_1(t) &= \frac{1}{\theta} \cdot (x_{1i} - x_1(t)) - \frac{1}{\theta} \cdot x_1(t) \cdot u(t), \\ \dot{x}_2(t) &= -\frac{1}{\theta} \cdot x_2(t) + \frac{1}{\theta} \cdot (x_{2i} - x_2(t)) u(t), \\ \dot{x}_3(t) &= -\frac{1}{\theta} \cdot x_3(t) + \frac{1}{\theta} \cdot (x_{3i} - x_3(t)) u(t),\end{aligned}\quad (6.38)$$

$$g(\mathbf{x}, \xi) = \xi + x_2 + x_3 - x_1 - \frac{K_w}{\xi} - \frac{x_3}{1 + \frac{K_x \xi}{K_w}} = 0, \quad (6.39)$$

kjer so $\mathbf{x} = [x_1, x_2, x_3]^T$, $\xi = 10^{-pH}$, $\theta = V/q_A$ in $u(t) = 1/q_A \cdot q_B(t)$. K_x in K_w sta disociacijski konstanti pufru in vode. Parametri modela so $x_{1i} = 0.0012$ mol HCl/L, $x_{2i} = 0.0020$ mol NaOH/L, $x_{3i} = 0.0025$ mol NaHCO₃/L, $K_x = 10^{-7}$ mol/L, $K_w = 10^{-14}$ mol²/L², $q_A = 1$ L/min in $V = 2.5$ L. Enačba (6.39) je standardni implicitni zapis relacije med stanji procesa in številom pH. Zapišemo jo lahko tudi v polinomski obliki

$$g(\mathbf{x}, \xi) = \xi^3 + (K_w/K_x + x_2 + x_3 - x_1)\xi^2 + (x_2 - x_1 + K_x)\xi - K_w^2/K_x = 0. \quad (6.40)$$

Model Wiener-INFUMO identificiramo v dveh korakih: najprej lineariziramo funkciji f in g iz (6.3) okoli dane delovne točke, nato pa izvedemo poskuse, pri katerih primerjamo odzive lineariziranega modela in nelinearnega modela. Zaradi negotovosti parametrov nelinearnega modela bomo dobili pas izhodnih nelinearnih karakteristik.

Linearizacija (6.38)-(6.39) okoli $x_{1,s}$, $x_{2,s}$, $x_{3,s}$ in u_s je podana z enačbami

$$\begin{aligned}\mathbf{A} &= \begin{bmatrix} -\frac{1}{\theta}(1 + u_s) & 0 & 0 \\ 0 & -\frac{1}{\theta}(1 + u_s) & 0 \\ 0 & 0 & -\frac{1}{\theta}(1 + u_s) \end{bmatrix}, \\ \mathbf{B} &= \begin{bmatrix} -\frac{1}{\theta}x_{1,s} \\ \frac{1}{\theta}(x_{2i} - x_{2,s}) \\ \frac{1}{\theta}(x_{3i} - x_{3,s}) \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \frac{\partial \eta}{\partial x_1} & \frac{\partial \eta}{\partial x_2} & \frac{\partial \eta}{\partial x_3} \end{bmatrix},\end{aligned}\quad (6.41)$$

kjer je

$$\frac{\partial \eta}{\partial x_k} = \frac{\partial g(\mathbf{x})/\partial x_k}{\xi \ln(10) \partial g(\mathbf{x})/\partial \xi}, \quad k = 1, 2, 3. \quad (6.42)$$

Iz polinomske enačbe pH (6.39) izračunamo naslednje izraze:

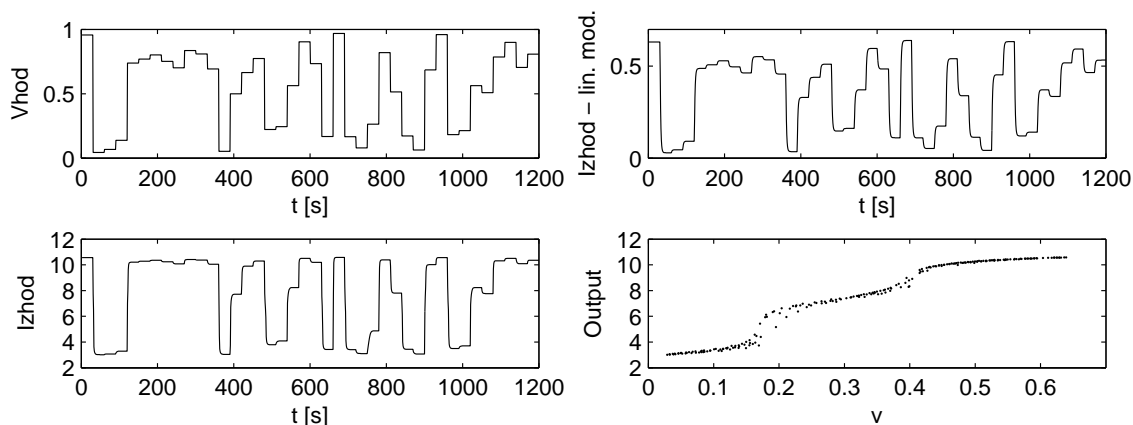
$$\begin{aligned}\frac{\partial g(\mathbf{x})}{\partial \xi} &= 3K_x \xi^2 + 2[K_w + (x_3 + x_2 - x_1)K_x]\xi + (x_2 - x_1 - K_x)K_w, \\ \frac{\partial g(\mathbf{x})}{\partial x_1} &= -K_x \xi^2 - K_w \xi, \quad \frac{\partial g(\mathbf{x})}{\partial x_2} = -K_x \xi^2 + K_w \xi, \quad \frac{\partial g(\mathbf{x})}{\partial x_3} = K_x \xi^2.\end{aligned}\quad (6.43)$$

Linearizacija okoli $\text{pH} = 7$ ($u_s = 0,369$, $x_{1,s} = 8,77 \cdot 10^{-4}$, $x_{2,s} = 5,39 \cdot 10^{-4}$ in $x_{3,s} = 6,74 \cdot 10^{-4}$) nam da

$$\mathbf{A} = \begin{bmatrix} -0,548 & 0 & 0 \\ 0 & -0,548 & 0 \\ 0 & 0 & -0,548 \end{bmatrix}, \quad \mathbf{B} = 10^{-3} \begin{bmatrix} -0.351 \\ 0.584 \\ 0.661 \end{bmatrix}, \quad (6.44)$$

$$\mathbf{C} = \begin{bmatrix} -463,43 & -462,39 & 462,91 \end{bmatrix}.$$

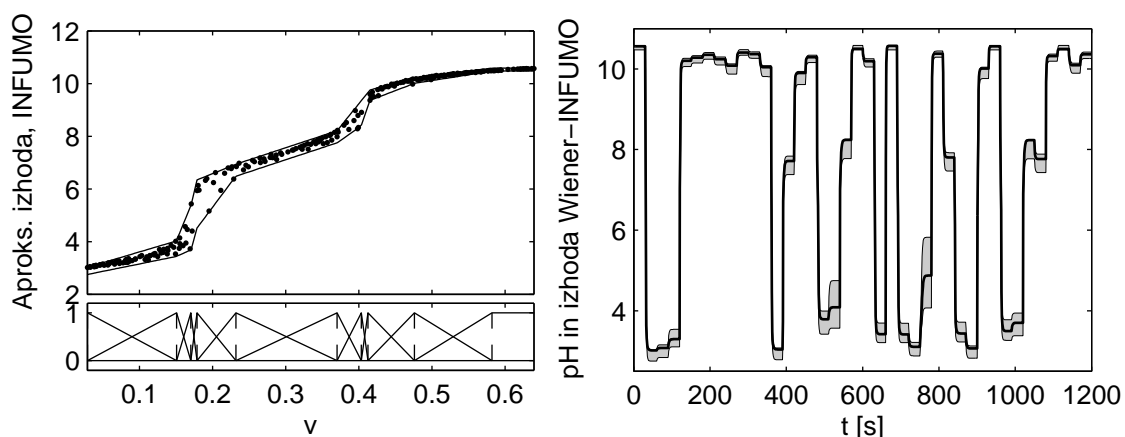
Na sliki 6.4 so predstavljeni signali identifikacijskega poskusa: vhodno-izhodni podatki procesa (levo), odziv linearnega modela (desno zgoraj) in dobljena aproksimacija izhodne preslikave (desno spodaj). Na zadnjem diagramu je opazna neenakomerna širina intervala izhodne preslikave. Razlog za to je v negotovosti dinamike sistema, ki je odvisna od delovne točke.



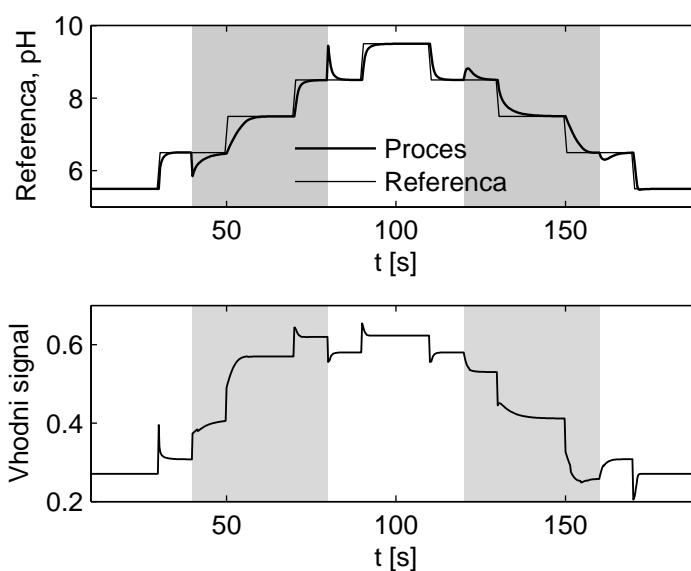
Slika 6.4: Vhodno-izhodni podatki procesa (levo), odziv linearnega modela (desno zgoraj), izhodna preslikava (desno spodaj)

Za model INFUMO smo uporabili statični model T-S z izhodnimi funkcijami tipa Singleton in desetimi trikotnimi pripadnostnimi funkcijami. Slika 6.5 prikazuje identificirano aproksimacijo s porazdelitvijo pripadnostnih funkcij (levi diagram) ter verifikacijo modela (desni diagram).

V zaprtozančnem preizkusu smo uporabili parametre regulatorja $n_u = 1$, $n_y = 4$ in $T = 0.5$ s, referenčni model pa je bil podan s parametri $\mathbf{A}_r = -0.5$, $\mathbf{B}_r = 1$ in $\mathbf{C}_r = 0.5$. Predpostavili smo dve napaki: 10-odstotno odstopanje vrednosti meritve pH-senzorja ob času $t_{f1} = 40 - 80$ s in spremembo pretoka q_A , ki povzroči napako na vhodnem signalu $\Delta u = 0.05$ ob $t_{f2} = 120 - 160$ s. Slika 6.6 kaže sledenje izhoda procesa referenci pri pojavu dveh napak (zgoraj) ter pripadajoči vhodni signal (spodaj). Vidimo, da



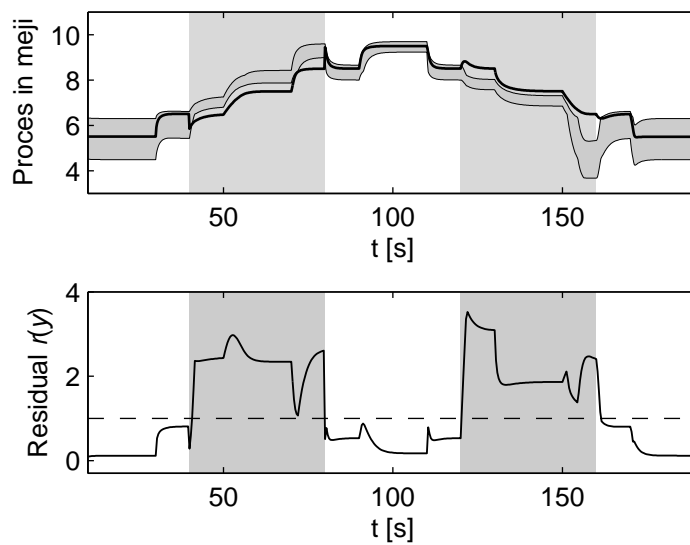
Slika 6.5: Aproksimacija z modelom Wiener-INFUMO (levo) in verifikacija modela (desno)



Slika 6.6: Rezultati vodenja pri pojavu dveh napak

sistem uspešno sledi referenci kljub pojavu napak na izhodu (prvo senčeno območje) in vходу procesa (drugo senčeno območje).

Slika 6.7 pa prikazuje rezultate zaznavanja napak. Zgornji diagram kaže izhod procesa v povezavi z mejama modela Wiener-INFUMO, na spodnjem diagramu pa je prikazan residual. Z zgornjega diagrama je razvidna strategija zaznavanja napak s prilagajanjem meje pasu zaupanja. Pri pojavu napake se zaradi drugačnega regulirnega signala interval premakne stran od reference, s tem pa izhod procesa zapusti dani interval. Čas zaznave je odvisen od širine intervala, dinamike regulirnega signala in



Slika 6.7: Rezultati zaznavanja napak v zaprti zanki

dinamike modela.

7. Prediktivno vodenje na podlagi modela INFUMO in optimizacije BB-BC

V tem poglavju bomo predstavili novo metodo prediktivnega vodenja na podlagi modernega optimizacijskega postopka Big Bang - Big Crunch (BB-BC) in modela INFUMO.

7.1 Optimizacija Big Bang - Big Crunch (BB-BC)

V tem delu bomo podali opis metode BB-BC, ki sta jo leta 2006 predstavila O. Erol in I. Eksin. Metoda temelji na osnovnih (danes veljavnih) principih evolucije vesolja - veliki pok (ang. big bang) in veliko skrčenje (ang. big crunch). V delu [28] sta avtorja pokazala, da je BB-BC učinkovitejša od osnovnega genetskega algoritma (GA) in nekaterih naprednejših genetskih metod optimizacije.

V evolucijskih principih je *naključnost* predstavljena kot disipacija energije, *konvergenca k lokalnem/globalnem optimumu* pa kot gravitacijska privlačnost. Ker disipacija energije iz nekega reda povzroči nered, je naključnost pri metodi BB-BC uporabljena kot prehod med trenutno kovergirano rešitvijo (red) in tvorbo novih kandidatov (nered) za rešitev optimizacijskega problema.

Prva faza optimizacije je ti. faza Big Bang. Začetna populacija osebkov ali kandidatov za rešitev optimizacijskega problema je naključno porazdeljena po celotnem prostoru iskanja. V tem je metoda BB-BC podobna ostalim evolucijskim algoritmom. Fazi Big Bang sledi faza Big Crunch, kjer postopek škrčenja "izračuna" masno središče" populacije po dani kriterijski funkciji in tako skrči prostor iskanja. V novih fazah Big Bang osebke naključno porazdelimo okrog izračunanih masnih središč. Postopek je povzet v naslednjem algoritmu.

1. (*Big Bang*) Generiramo začetno populacijo, tako da N osebkov naključno porazdelimo po celotnem prostoru iskanja rešitve;
2. za celotno populacijo izračunamo vrednosti kriterijskih funkcij;

3. (*Big Crunch*) Izračunamo masno središče populacije glede na vrednosti kriterijskih funkcij;
4. (*Big Bang*) Okoli središča naključno porazdelimo nove osebke, tako da prištevamo naključna števila, normirana s številom iteracije; z naraščanjem iteracije radij porazdelitve upada;
5. če je dosežen ustavitveni pogoj, prenehamo, drugače se vrnemo v korak 2.

Denimo, da osebke označimo $x_i \in \mathcal{X}$, vrednosti pripadajočih kriterijskih funkcij pa f_i . Masno središče v koraku 3 izračunamo po enačbi

$$\tilde{x} = \frac{\sum_{i=1}^N \frac{1}{f_i} x_i}{\sum_{i=1}^N \frac{1}{f_i}}, \quad (7.1)$$

V novi generaciji oblikujemo populacijo po enačbi $x_i^{novi} = \tilde{x} + \sigma$, kjer je σ naključno število z normalno porazdelitvijo preko omejenega prostora iskanja. Označimo s k zaporedno število generacije. Število σ izračunamo po enačbi

$$\sigma = \frac{r\alpha(\bar{x} - \underline{x})}{k} \quad (7.2)$$

kjer je r naključno število, α je parameter, s katerim omejimo prostor iskanja, \bar{x} in \underline{x} pa sta zgornja meja in spodnja meja prostora. Nove osebke torej porazdelimo po enačbi

$$x_i^{novi} = \tilde{x} + \frac{r\alpha(\bar{x} - \underline{x})}{k}. \quad (7.3)$$

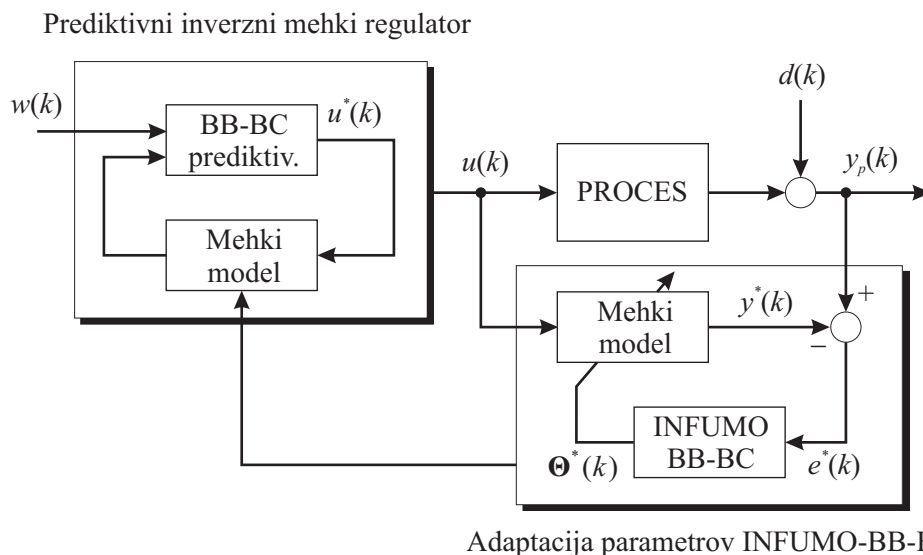
7.2 BB-BC v prediktivnem vodenju na podlagi modela INFUMO

Dosedanja uporaba metode BB-BC je omejena na optimizacijo v konstrukciji lahkih ogrodij [15] in optimizacijo parametrov mehkega modela pri vodenju z notranjim modelom (ang. internal model control, IMC) na podlagi inverznega mehkega modela [72]. Pri slednjem je glavna naloga BB-BC, da v čim krajšem času privede do parametrov mehkega modela tipa Singleton, ki v enokoračni napovedi minimizira razliko med izhodom modela in procesa, nato pa se analitično izračuna inverz dobljenega modela za prihodnji regulirni signal. Kljub dobri rezultatom opazimo predvsem naslednji šibki točki omenjenega pristopa:

- enokoračna napoved modela za izračun regulirnega signala;

- začetni prostor iskanja za BB-BC je vedno celoten prostor rešitev .

Imejmo proces z negotovimi parametri, katerega je moč opisati z družino izhodnih funkcij. Prvo slabost bomo poskušali odpraviti z uvedbo večkoračne napovedi in strategije premičnega horizonta, drugo pa z uporabo modela INFUMO. Ideja pristopa, ki ga bomo predlagali v tem razdelku, je predstavljena na sliki 7.1. Iz eksperimentalnih



Slika 7.1: Shema vodenja na podlagi modela INFUMO in optimizacije BB-BC

podatkov identificiramo model INFUMO, ki ga predstavlja interval parametrov v matrikah $\underline{\Theta}$ in $\overline{\Theta}$. Na podlagi meritve procesa $y_p(k)$, zadnjega vhoda v proces $u(k-1)$ in izhoda modela $y(k)$ s pomočjo optimizacije BB-BC izračunamo parametre internega mehkega modela, podane v matriki $\Theta^*(k)$, tako da minimiziramo pogrešek $e(k)$. Parametri v $\Theta^*(k)$ so v vsakem časovnem trenutku in za vsake delovne pogoje omejeni z istoležnimi parametri matrik $\underline{\Theta}$ in $\overline{\Theta}$. Glavna prednost tega pristopa je, da je začetni interval parametrov za metodo BB-BC vedno omejen s tema matrikama, to pa pohitri konvergenco rešitve v vsakem koraku prilagajanja parametrov.

Referenčni signal označimo z $w(k)$. Izračunane parametre iz koraka k posredujemo bloku "Prediktivni inverzni mehki regulator", v katerem po metodi BB-BC iščemo tak signal $u(k)^*$, ki minimizira kriterijsko funkcijo

$$J(k) = \sum_{i=1}^H (y_r(k+i) - y(k+i))^2, \quad (7.4)$$

kjer z $y_r(k+i)$ označimo napoved izhoda referenčnega modela $y_r(k+1) = a_r y_r(k) + (1 - a_r)w(k)$ za i vzorcev v prihodnosti, a_r je časovna konstanta referenčne dinamike

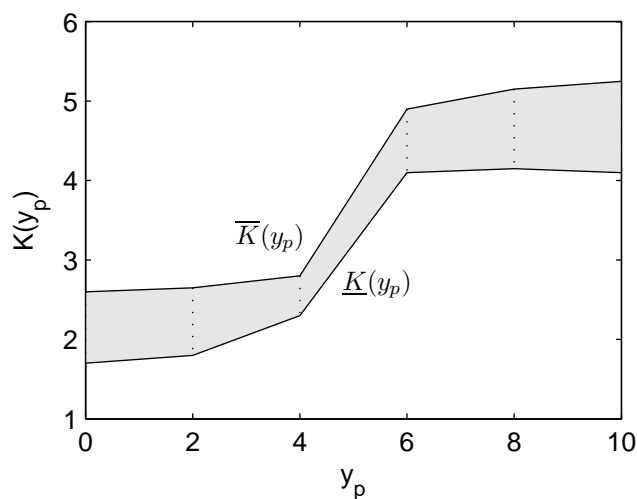
prvega reda v diskretni obliki, $y(k+i)$ pa je napoved izhoda mehkega modela kot odziv na $u^*(k)$. Tak pristop odpravlja negotovosti meritev $y_p(k)$, poleg tega pa kompenzira negotovosti v parametrih procesa s prilagajanjem modela pred izračunom kriterijske funkcije, ki minimizira razliko med napovedjo modela in referenčne trajektorije.

7.2.1 Primer

Predlagano metodo bomo preizkusili na modelu procesa

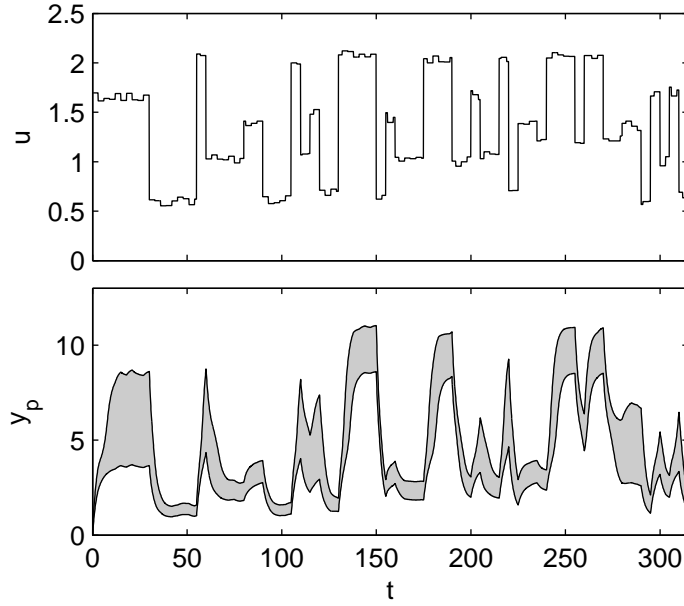
$$\dot{y}_p(t) = -\frac{1}{2}y_p(t) + \frac{1}{2}K(y_p(t))u(t - T_d), \quad (7.5)$$

kjer je $T_d = 0.5$ s časovna zakasnitev, $K(y_p(t))$ pa variabilno ojačenje, odvisno od izhoda procesa, za katerega poznamo samo zgornjo in spodnjo mejo, kot je to prikazano na sliki 7.2. Vhodni signal $u(t)$ je omejen z intervalom $[0 \ 2]$. Predpostavimo, da lahko vhodna statična krivulja procesa zavzame kakršnokoli obliko funkcije znotraj podanih intervalov. V primeru smo zaradi enostavnosti definirali preslikavo $y \rightarrow K(y)$ v šestih točkah ($y_p = \{0, 2, 4, 6, 8, 10\}$) in v vsakem teku procesa izračunali naključne vrednosti preslikave v teh šestih točkah znotraj danih intervalov ($y_{p,i} \rightarrow K(y_{p,i}) \in [\underline{K}(y_{p,i}), \overline{K}(y_{p,i})]$), ostale vrednosti $K(y_p)$, $y_p \neq y_{p,i}$ pa linearno interpolirali. Za identifikacijo smo opr-



Slika 7.2: Interval variabilnega ojačenja za testni proces

vili poskuse s stopničastim signalom $u(t)$, vhodno-izhodne podatke pa vzorčili s časom vzorčenja $T_s = 0,1$ s. Ker je dinamika procesa prvega reda, sta zgornja in spodnja meja izhoda vedno odziva pri zgornji in spodnji meji ojačenja, $\overline{K}(y_{p,i})$ in $\underline{K}(y_{p,i})$. Slika 7.3 predstavlja pas meritev (senčeno področje). Na podlagi podatkov identificiramo



Slika 7.3: Pas meritev in omejitev z modelom INFUMO

parametre dinamičnega modela INFUMO

$$\mathbf{R}_j : \text{ če } y(k) \text{ je } \mathbf{A}_j, \text{ potem } \bar{y}(k+1) = \bar{\theta}_{j,2}\bar{y}(k) + \bar{\theta}_{j,1}u(k) + \bar{\theta}_{j,0}, \quad (7.6)$$

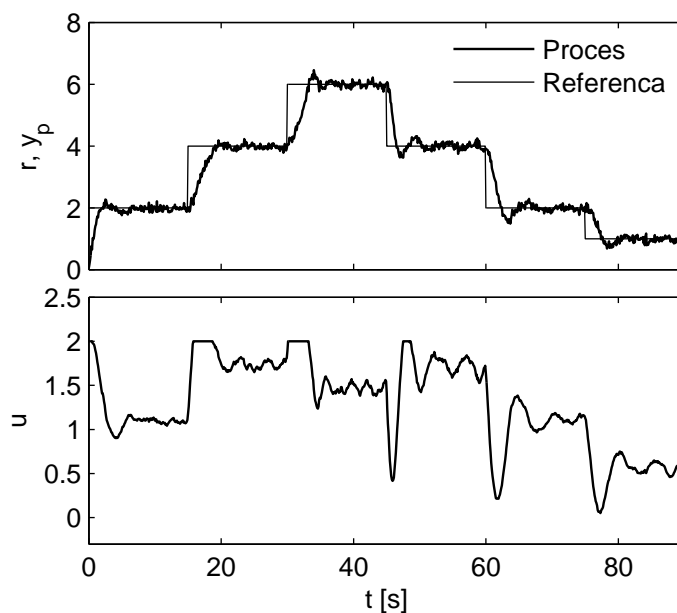
$$\underline{y}(k+1) = \underline{\theta}_{j,2}\underline{y}(k) + \underline{\theta}_{j,1}u(k) + \underline{\theta}_{j,0}, \quad j = 1, \dots, 5.$$

Parametre podajata matriki

$$\bar{\Theta} = \begin{bmatrix} 0.955 & 0.124 & 0.010 \\ 1.013 & 0.194 & -0.295 \\ 0.964 & 0.242 & -0.079 \\ 0.956 & 0.250 & -0.028 \\ 0.955 & 0.250 & -0.020 \end{bmatrix}, \quad \underline{\Theta} = \begin{bmatrix} 0.970 & 0.093 & -0.046 \\ 0.969 & 0.110 & -0.064 \\ 0.990 & 0.089 & -0.141 \\ 0.990 & 0.166 & -0.227 \\ 0.959 & 0.218 & -0.110 \end{bmatrix}, \quad (7.7)$$

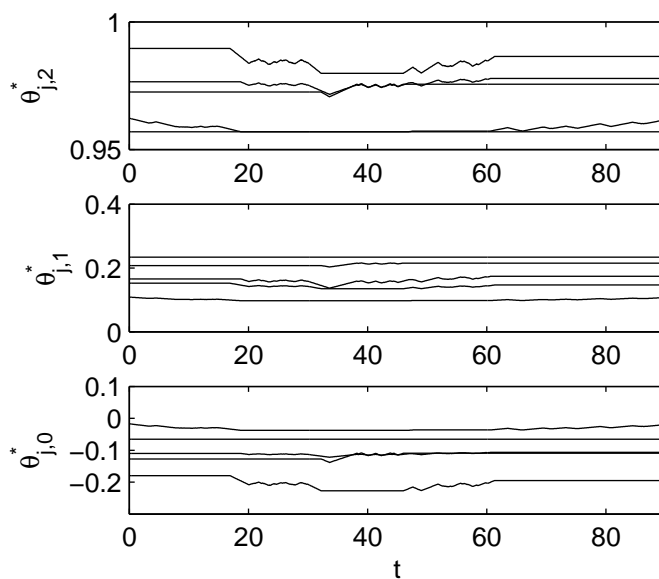
vrhovi pripadnostnih funkcij so pri $\alpha = [3.02 \quad 3.79 \quad 4.83 \quad 7.42 \quad 7.80]$, mejni funkciji pa sta prikazani na sliki 7.3. Maksimalna pogreška aproksimacije sta $\lambda_1 = 0,047$ in $\lambda_2 = 0,070$.

V zaprtizančnem poskusu smo predpostavili sledenje stopničastemu referenčnemu signalu, procesu pa smo dodali normalno porazdeljeni šum z varianco enako $5 \cdot 10^{-4}$. Za horizont napovedi smo določili vrednost $H = 5$. Parametra referenčnega modela sta bila $a_r = 0,9048$ in $b_r = 0,0952$. Končna toleranca kriterijske funkcije je bila 0,001. Predpostavili smo 40 osebkov v populaciji in maksimalno število 20 iteracij na računski korak. Rezultati vodenja so prikazani na sliki 7.4. Na zgornjem diagramu vidimo, da iz-



Slika 7.4: BBBC, zaprtzoančni preizkus: izhod (zgoraj), vhod (spodaj)

hod procesa uspešno sledi referenci kljub velikim spremembam delovnih pogojev zaradi nelinearnega ojačenja ter prisotni šuma. Spodnji diagram kaže na to, da mora sistem v prehodnem pojavu kompenzirati spremembe ojačenja z močnimi spremembami vhodnega signala. Na sliki 7.5 pa je prikazano prilagajanje parametrov modela INFUMO med izvajanjem poskusa. Poskus smo ponovili po metodi BB-BC-IMC, opisani v [72],



Slika 7.5: Prilagajanje parametrov med delovanjem

kjer je začetna porazdelitev parametrov mehkega modela v celoti naključna. Izmerili in primerjali smo čas, ki je potreben za izvedbo poskusov na osebem računalniku s procesorjem Intel Pentium Dual CoreTM 1,46 MHz, 2 GB pomnilnika ter programskim okoljem Matlab-Simulink 7.01. Pri metodi BB-BC-IMC je čas simulacije znašal 42,67 s, pri predlagani metodi INFUMO-BB-BC pa smo dosegli čas 28,91 s. To pomeni 47,6-odstotno pohiترitev na račun omejitve prostora rešitev v vsakem računskem koraku s pomočjo modela INFUMO.

8. Sklep

Glavne izvirne prispevke k znanosti lahko povzamemo v naslednjih točkah:

- intervalni mehki model ali model INFUMO je bil vpeljan kot univerzalni aproksimator poljubnega nelinearnega področja ali družine nelinearnih sistemov, predstavljena je bila metoda za identifikacijo parametrov modela INFUMO na podlagi linearnega programiranja;
- predstavljena je bila identifikacija nelinearnega procesa z negotovimi parametri na podlagi dinamičnega modela INFUMO ter modela tipa Wiener-INFUMO;
- model INFUMO je bil uporabljen v zaznavanju napak za nelinearni proces z negotovimi parametri, katerega izhod v normalnem načinu delovanja (brez prisotnosti napak) opišemo z družino izhodnih funkcij; za pridobivanje ti. pasu zaupanja smo uporabili nizkopasovni filter, metoda pa tako omogoča zaznavanje in lokalizacijo napak;
- na enak na način kot pri prejšnji metodi je bila razvita tudi inačica z visokopasovnim filtrom; za oba primera je bila opravljena analiza zaznavnosti, primerjava rezultatov pa daje nekoliko več prednosti slednji metodi;
- razvita je bila metoda nelinearnega prediktivnega vodenja v zveznem času na podlagi modela Wiener-INFUMO, ki je primerna za procese z negotovimi parametri in izrazito statično nelinearnostjo, katerih izhode opišemo z družinami funkcij; metoda omogoča tudi kombinacijo sistemov za vodenje in zaznavanje napak;
- razvita je bila metoda prediktivnega vodenja na podlagi moderne genetske optimizacije Big Bang - Big Crunch (BB-BC); metoda temelji na ideji inverznega mehkega regulatorja, kjer je v vsakem koraku treba izvesti prilagajanje modela procesu; z modelom INFUMO identificiramo pas možnih parametrov procesa in omejimo iskalni prostor optimizacijske metode, kar močno zmanjša računsko breme v vsakem koraku.

Dodatek A

Uporaba modela INFUMO v zaznavanju napak (NPF)

Available online at www.sciencedirect.com

Engineering Applications of Artificial Intelligence 20 (2007) 503–510

 Engineering Applications of
**ARTIFICIAL
 INTELLIGENCE**
www.elsevier.com/locate/engappai

Brief paper

Fault detection for nonlinear systems with uncertain parameters based on the interval fuzzy model

Simon Oblak*, Igor Škrjanc, Sašo Blažič

Faculty of Electrical Engineering, Tržaška 25, 1000 Ljubljana, Slovenija

Received 24 June 2005; received in revised form 1 August 2006; accepted 2 August 2006
Available online 2 October 2006

Abstract

In the paper an application of the interval fuzzy model (INFUMO) in fault detection for nonlinear systems with uncertain interval-type parameters is presented. A confidence band for the input–output data, obtained in the normal operating conditions of a system, is approximated using a fuzzy model with interval parameters. The approximation is based on linear programming using l_∞ -norm as a measure of the modelling error. Applying low-pass filtering when obtaining the confidence band makes it possible to use arbitrary sets of identification input signals. An application of the INFUMO in a fault-detection system for a two-tank system is presented to demonstrate the benefits of the proposed method.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Fuzzy model; Interval model; Fault detection; Interval-type uncertainties

1. Introduction

Automated supervision and fault diagnosis are important features in design of efficient and reliable production systems. Fault detection (FD) problem for linear systems has been extensively studied over the past three decades, and a lot of powerful methods have been developed. For a more thorough review, the reader is referred to survey papers by Frank (1990), Isermann (1997), and Patton and Chen (1997). However, the use of linear approaches is limited if the system to be monitored is strongly nonlinear. Since many industrial systems are nonlinear in nature, the development of nonlinear FD methods plays a significant role in practical applications. Recently, observer-based approaches (Bastin and Gevers, 1988; Chen et al., 1996; Hammouri et al., 1999), adaptive-residual-threshold approach (Frank and Ding, 1997), fuzzy-model-based (Ballé and Fuessel, 2000; Amann et al., 2001) and neural-network-based methods (Vemuri and Polycarpou, 2004; Klančar et al., 2002) have been proposed. For more information on the use of artificial intelligence in fault

detection the reader is referred to survey papers by Alcorta García and Frank (1997), Patton and Chen (1996), and Frank and Köppen-Seliger (1997). Most of the above-mentioned methods are based on a *decoupling* framework, where the modelling uncertainty and all possible faults can be decoupled through an appropriate coordinate transformation and residual generation technique. However, the modelling uncertainty is often unstructured, which makes it difficult to achieve exact decoupling between faults and modelling errors. In addition, some problems taking into consideration the input–output representation of systems as well as the design of the corresponding nonlinear observers are still open.

This paper presents a fault-detection scheme for nonlinear input–output systems with unstructured interval-type uncertainties. A fault-detection method using nonlinear adaptive fault estimators for dealing with the same system type was presented by Zhang et al. (2001). The presented approach is based on the use of an interval fuzzy model (INFUMO). Interval arithmetics has often been used in system identification, see e.g. the work by Jaulin and Walter (1993) and Malan et al. (1997), and fuzzy systems are usually applied as a nonlinear function estimators (Ying and Chen, 1997). As was introduced

*Corresponding author. Tel.: +38614768278; fax: +38614264631.
 E-mail address: Simon.Oblak@fe.uni-lj.si (S. Oblak).

and shown by Škrjanc et al. (2005a), by applying only one Takagi–Sugeno-type (Takagi and Sugeno, 1985) fuzzy model with interval parameters, one is able to approximate the upper and lower boundaries of the domain of functions that result from an uncertain system. When system parameters vary in a certain tolerance band, it is advantageous to define a confidence band over a finite set of input and output measurements in which the effects of unknown process inputs are already included. If the outer bounds of the set, obtained during fault-free operating conditions, are determined, one is able to use them as the bounds of the admissible output area, and thus being able to create a simple and intuitive fault-detection system. The main idea of the proposed approach is to use the INFUMO as an admissible-filtered-output estimator in a fault-detection system. By calculating the normalized distance of the filtered system output from the boundary model outputs, a residual function is obtained.

The paper is organized in the following way. In Section 2 the system description, the main idea of interval fuzzy model identification using l_∞ -norm, the residual formation with a fault-diagnostic scenario, and detectability issues are described. Section 3 presents the two-tank system with interval-type uncertain parameters, and the application to fault detection with data pre-processing and low-pass filtering is introduced. In the final part, some outlines of future work are given.

2. Using the fuzzy interval model in fault detection

2.1. System description

Let the nonlinear system be given in a general form as follows:

$$\begin{aligned} \dot{x}(t) &= \sigma(x, u, t) + \eta(x, u, t) + \phi(x, u, t), \\ y(t) &= \gamma(x, u, t) + \rho(x, u, t), \end{aligned} \quad (1)$$

where $x \in \mathbb{R}^n$ is the state vector of the system, $u \in \mathbb{R}$ is the system input, $y \in \mathbb{R}$ denotes the system output, $\eta: \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}^n$ is the state disturbance, $\rho: \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}$ is the output disturbance, $\phi: \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}^n$ denotes the fault function, and $\sigma: \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}^n$ and $\gamma: \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}$ are the nonlinear functions of the state vector, input and time, respectively. Let us assume that only the system outputs can be directly measured. Throughout the paper the following assumptions will be made:

Assumption 1. The modelling uncertainties, represented by η and ρ in (1), are unstructured unknown nonlinear functions of x , u and t , but bounded by some known functionals (Zhang et al., 2001), i.e.,

$$\begin{aligned} |\eta(x, u, t)| &\leq \bar{\eta}(y, u, t), \quad |\rho(x, u, t)| \leq \bar{\rho}(y, u, t) \\ \forall (x, y, u) &\in \mathcal{X} \times \mathcal{Y} \times \mathcal{U}, \quad \forall t \geq 0, \end{aligned} \quad (2)$$

where the bounding functions $\bar{\eta}(y, u, t)$ and $\bar{\rho}(y, u, t)$ are known and uniformly bounded. $\mathcal{X} \subset \mathbb{R}^n$ is some compact domain of interest, and $\mathcal{U} \subset \mathbb{R}$ and $\mathcal{Y} \subset \mathbb{R}$ are the compact sets of admissible inputs and outputs, respectively.

Note that $\bar{\eta}$ and $\bar{\rho}$ are the functions of the output y whereas η and ρ are the functions of the state vector x . This is due to the fact that the states are not directly measured.

Assumption 2. The output functions $y(t)$, when $\phi(x, u, t) = 0$, are bounded by the following interval:

$$y(t) \in [\underline{y}(t), \bar{y}(t)] \subset \mathcal{Y}. \quad (3)$$

Assumption 1 characterizes the possible modelling uncertainties as unstructured but bounded by some constant or function, and Assumption 2 guarantees that in the absence of faults the bounds of the interval can be determined. As a consequence, a confidence band of outputs guarantees that a process output exhibiting normal behaviour is found in the interval $[\underline{y}, \bar{y}]$. However, due to the unknown effect of the actual disturbance functions the exact bounds cannot be defined analytically. In the paper by Fagarasan et al. (2004) two approaches to determine the boundary functions have been proposed: empirical and numerical. In the former approach, physical knowledge of the uncertainties is used to adjust its values in the model. The latter approach consists of using a constrained linear optimization technique to minimize the model precision objective function $J = (1/N) \sum_{i=1}^N (\bar{y}_i - \underline{y}_i)$, where \bar{y}_i and \underline{y}_i are the maximal and minimal output values that result from different uncertainty values. The proposed approach, introducing the interval fuzzy model, is qualitatively different because the boundary responses will be obtained by a fuzzy function approximation of the bounds of a set of filtered input–output data that already comprises the effects of disturbances.

As was shown by Ding et al. (1994) and Frank and Ding (1997), a residual generator can be designed by

$$r(t) = Q(p)(y(t) - \hat{y}(t)), \quad (4)$$

with $\hat{y}(t)$ as a process-output estimation and $Q(p)$ is a filter which enhances the residual robustness to unknown process inputs and is to be designed according to the demands of the application. Since in this case only the input and output data are available and the transfer functions from the faults and disturbances are not known, the above-mentioned approach is altered accordingly. The main idea of the presented approach is to filter both the input and output data through a low-pass filter in order to get a static input–output mapping approximation. Since the main role of the filter is to cut out the high-frequency spectre of the data set, the design procedure is further simplified by using a first-order filter. The data pairs $u_f(i)$ and $y_f(i)$, obtained in the normal operating conditions, i.e., in the absence of faults, are gathered in a set of the so-called admissible input–output data. Robustness to the effects of disturbances can be seen in the fact that the latter

are already included in the set. If the set bounds are modelled as a function of the system input, they can be used online for residual formation

$$\begin{aligned} r_{f1}(t) &= \bar{y}_f(t) - Q(p)y(t) = \bar{f}(u_f, t) - Q(p)y(t), \\ r_{f2}(t) &= Q(p)y(t) - \underline{y}_f(t) = Q(p)y(t) - \underline{f}(u_f, t), \end{aligned} \quad (5)$$

where

$$Q(p) = \frac{1}{T_f p + 1} \quad (6)$$

and T_f denotes the time constant of the filter $Q(p)$. Filtering both the input and output signals before calculation of the residuals further decreases the residual sensitivity to false alarms, resulting from the disturbances.

2.2. Interval fuzzy model

The interval fuzzy model was introduced by Škrjanc et al. (2005a) as a means of robust system identification, and it was also used in the work by Škrjanc et al. (2005b) to provide a functional description of a static nonlinear area approximation.

In this paper the modelling of $\bar{f}(u_f, t)$ and $\underline{f}(u_f, t)$ in (5) will be carried out using an interval fuzzy model. Note that for the confidence-band identification a set of either open-loop or closed-loop experiments in the absence of faults have to be conducted, and the online fault detection based on the obtained INFUMO can also be realized in both systems.

The procedure of obtaining the exact lower and upper bounds for the family of functions and assigning the INFUMO to them will be briefly reviewed. Let $C \subset \mathbb{R}$ be a compact set and $\mathcal{G} = \{g(z) : C \rightarrow \mathbb{R}\}$ be a class of nonlinear functions. Let us assume that there exist the exact upper bound \bar{g} and the exact lower bound \underline{g} that satisfy the following conditions for an arbitrary $\varepsilon > 0$ and for each $z \in C$:

$$\bar{g}(z) \geq \max_{g \in \mathcal{G}} g(z), \quad \exists g \in \mathcal{G} : \bar{g}(z) < g(z) + \varepsilon, \quad (7)$$

$$\underline{g}(z) \leq \min_{g \in \mathcal{G}} g(z), \quad \exists g \in \mathcal{G} : \underline{g}(z) > g(z) - \varepsilon. \quad (8)$$

Obtaining the bounds in Eqs. (7) and (8) would require an infinite amount of data; however, in this case we are limited to the finite set of measured filtered output values $Y = \{y_{f1}, y_{f2}, \dots, y_{fN}\}$ and the finite set of filtered input data $Z = \{z_1, z_2, \dots, z_N\} = \{u_{f1}, u_{f2}, \dots, u_{fN}\}$:

$$y_{fi} = g(z_i), \quad g \in \mathcal{G}, \quad z_i \in C \subset \mathbb{R}, \quad y_i \in \mathbb{R}, \quad i = 1, \dots, N. \quad (9)$$

The exact upper and lower boundary functions will be approximated by fuzzy functions, and the Stone–Weierstrass theorem (Ying and Chen, 1997) is extended to prove that there exist fuzzy systems \bar{f} and \underline{f} such that the bounds of an arbitrary nonlinear area can be approximated with

any precision, i.e.,

$$\begin{aligned} 0 < \bar{f}(z_i) - g(z_i) < \varepsilon \quad \forall i, \\ -\varepsilon < \underline{f}(z_i) - g(z_i) < 0 \quad \forall i. \end{aligned} \quad (10)$$

In this paper an INFUMO TS-type model in affine form with one antecedent variable will be assumed. It can be given as a set of rules

$$\begin{aligned} R_j : \text{if } x_p \text{ is } A_j \text{ then } \bar{f}(z_i) = \bar{y}_f = \bar{\theta}_j^T x_c, \quad j = 1, \dots, m, \\ \underline{f}(z_i) = \underline{y}_f = \underline{\theta}_j^T x_c, \quad j = 1, \dots, m. \end{aligned} \quad (11)$$

The antecedent variable $x_p = u_f \in \mathbb{R}$ denotes the input or variable in premise, and variables $\bar{y}_f, \underline{y}_f \in \mathbb{R}$ are the outputs of the interval fuzzy model that provide the upper and lower boundary function. The antecedent variable is connected with m fuzzy sets A_j , and each fuzzy set A_j ($j = 1, \dots, m$) is associated with a real-valued function $\mu_{A_j}(x_p) : \mathbb{R} \rightarrow [0, 1]$, that produces a membership grade of the variable x_p with respect to the fuzzy set A_j . The consequent vector is denoted $x_c^T = [u_f \ 1] \in \mathbb{R}^2$. As the output functions are in affine form, 1 was added to the vector x_c . The system outputs are linear combinations of the consequent states, and $\bar{\theta}_j, \underline{\theta}_j \in \mathbb{R}^2$ are vectors of fuzzy parameters. The system in Eq. (11) can be described in closed form

$$\begin{aligned} \bar{y}_f &= \beta^T(x_p) \bar{\Theta} x_c, \\ \underline{y}_f &= \beta^T(x_p) \underline{\Theta} x_c, \end{aligned} \quad (12)$$

where $\bar{\Theta}^T = [\bar{\theta}_1, \dots, \bar{\theta}_m]$ and $\underline{\Theta}^T = [\underline{\theta}_1, \dots, \underline{\theta}_m]$ denote the upper and lower coefficient matrices for the complete set of rules, and $\beta^T(x_p) = [\beta_1(x_p), \dots, \beta_m(x_p)]$ is a vector of normalized membership functions with elements that indicate the degree of fulfilment of the respective rule. Functions $\beta_j(x_p)$ can be defined as

$$\beta_j(x_p) = \frac{\mu_{A_j}(x_p)}{\sum_{j=1}^m \mu_{A_j}(x_p)}, \quad j = 1, \dots, m, \quad (13)$$

if the partition of unity is assumed. Note that $\beta_j(x_p)$ and x_c are equal in both the upper and lower boundary-function calculations in (12).

The membership functions $\mu_{A_j}(x_p)$ and the fuzzy parameters in $\bar{\Theta}, \underline{\Theta}$ of the proposed INFUMO have to be determined for optimal approximation of the bounds. While the proper arrangement and shape of the membership functions will not be considered—the mentioned can be approached by different clustering methods (Babuška, 1998)—for the optimal fuzzy parameters the minimization of the maximum modelling error ε in Eq. (10) over the whole input set Z is performed. This implies the min–max optimization method, and l_∞ -norm is used as the modelling error measure, yielding

$$\begin{aligned} \min_{\bar{\Theta}} \max_{z_i \in Z} |y_i - \bar{f}(z_i)| \quad \text{s.t. } y_i - \bar{f}(z_i) \geq 0, \\ \min_{\underline{\Theta}} \max_{z_i \in Z} |y_i - \underline{f}(z_i)| \quad \text{s.t. } y_i - \underline{f}(z_i) \leq 0, \end{aligned} \quad (14)$$

where $f(z_i) = \beta^T \underline{Q} x_c(z_i)$ and $\bar{f}(z_i) = \beta^T \bar{\Theta} x_c(z_i)$. Note that the data are obtained by sampling different functions from \mathcal{G} with arbitrary values of z_i . The solutions to both problems can be found by linear programming, because both problems can be viewed as linear programming problems. This brings simplicity to the realization of the optimizing process. However, large data sets and a large number of parameters will still pose a threat to optimization convergence. In the first case the problems are approached with data-reduction methods, and in the latter case, on the other hand, solutions to reduce the number of parameters have to be found.

2.3. Formation of the proposed fault-detection system

When the parameter identification is done, the INFUMO is connected to the process in parallel to get online estimations of the boundary outputs. In terms of fault detection, the decision function should consist of verifying that each measurement belongs to the corresponding confidence band. In order to provide normalized quantitative information about the proximity of the measurements to the closest interval bound, distances were used, as presented by Fagarasan et al. (2004). If a filtered output value $y_f(t)$ belongs to an interval $[\underline{y}_f(t), \bar{y}_f(t)]$, and if $\tilde{y}_f(t)$ denotes the mean interval value

$$\tilde{y}_f(t) = \frac{\bar{y}_f(t) + \underline{y}_f(t)}{2} \tag{15}$$

the proposed distance is defined in the following way:

$$\begin{aligned} \text{if } y_f(t) < \tilde{y}_f(t), \quad d(y_f) &= \frac{y_f(t) - \tilde{y}_f(t)}{\underline{y}_f(t) - \tilde{y}_f(t)}, \\ \text{if } y_f(t) > \tilde{y}_f(t), \quad d(y_f) &= \frac{y_f(t) - \tilde{y}_f(t)}{\bar{y}_f(t) - \tilde{y}_f(t)}. \end{aligned} \tag{16}$$

The distance in (16) is zero when the measurement is equal to \tilde{y}_f , and approaches the value 1 if the measurement is close to one of the interval bounds. A fault is signalled every time $d(y_f)$ exceeds the value 1. Fig. 1 gives a schematic representation of the proposed fault-detection system. The filter $Q(p)$ is represented by a block denoted LPF (low pass filter), and the distance is calculated in the DIST block.

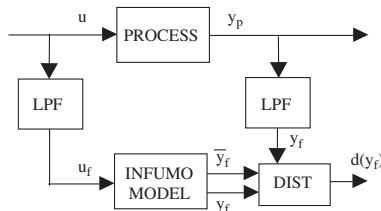


Fig. 1. Fault-detection system using static INFUMO model.

2.3.1. Detectability properties

To study the detectability properties of the proposed method the worst-case scenario will be considered. In this sense the following theorem guarantees detectability of a fault ϕ :

Theorem 1. Let the contribution of a fault $\phi(x, u, t)$ to the filtered process output be denoted $y_{\phi f}(t)$, and the difference between the boundary functions $\bar{y}_f(t) - \underline{y}_f(t) = \lambda(t)$. A fault will certainly be detected if the absolute value of the former is greater than the width of the output confidence band, i.e.

$$|y_{\phi f}(t)| > \lambda(t). \tag{17}$$

Proof. For the purpose of fault-detectability analysis, the output function in Eq. (1) will be rewritten to

$$y_f(t) = Q(p) \cdot y(t) = y_{0f}(t) + \rho_f(t) + y_{\phi f}(t), \tag{18}$$

where $y_{0f}(t)$, $\rho_f(t)$ and $y_{\phi f}(t)$ denote the filtered contributions of the “undisturbed” output, disturbances and fault, respectively. Since the upper and lower confidence-band bounds $\bar{y}_f(t)$ and $\underline{y}_f(t)$ were obtained in the experiments where no faults were present, the following inequalities hold:

$$\begin{aligned} y_{0f}(t) + \rho_f(t) - \underline{y}_f(t) &\geq 0, \\ \bar{y}_f(t) - y_{0f}(t) - \rho_f(t) &\geq 0. \end{aligned} \tag{19}$$

By adding the fault contribution $y_{\phi f}(t)$ to both sides of (19), the following can be written:

$$\begin{aligned} y_{0f}(t) + y_{\phi f}(t) + \rho_f(t) - \underline{y}_f(t) &\geq y_{\phi f}(t) \Rightarrow y_{\phi f}(t) \leq y_f(t) - \underline{y}_f(t), \\ \bar{y}_f(t) - y_{0f}(t) - \rho_f(t) - y_{\phi f}(t) &\geq -y_{\phi f}(t) \Rightarrow \bar{y}_f(t) - y_f(t) \\ &\geq -y_{\phi f}(t). \end{aligned} \tag{20}$$

Hence, $y_{\phi f}(t)$ is bounded by $y_f(t) - \bar{y}_f(t) \leq y_{\phi f}(t) \leq y_f(t) - \underline{y}_f(t)$. When no fault is detected, i.e., $\underline{y}_f(t) \leq y_f(t) \leq \bar{y}_f(t)$, inequality (20) can be further extended to

$$\begin{aligned} y_{\phi f}(t) \leq y_f(t) - \underline{y}_f(t) &= (\bar{y}_f(t) - \underline{y}_f(t)) + (y_f(t) - \bar{y}_f(t)) \\ &= \lambda(t) + (y_f(t) - \bar{y}_f(t)) \leq \lambda(t) + 0 \\ &= \lambda(t), \end{aligned}$$

$$\begin{aligned} -y_{\phi f}(t) \leq \bar{y}_f(t) - y_f(t) &= (\bar{y}_f(t) - \underline{y}_f(t)) + (y_f(t) - \bar{y}_f(t)) \\ &= \lambda(t) + (y_f(t) - \bar{y}_f(t)) \leq \lambda(t) + 0 \\ &= \lambda(t). \end{aligned} \tag{21}$$

Both inequalities in (21) can be united to

$$|y_{\phi f}(t)| \leq \lambda(t). \tag{22}$$

When a fault is not detected, inequality (22) is satisfied, thus implying that in the case when $|y_{\phi f}(t)| > \lambda(t)$ the fault is certainly detected.

Remark 1. If the steady-state gain of the filter is $Q(s)|_{s=0} = 1$, and if

$$\int_{1/T}^{\infty} |Y_{\phi}(\omega)|^2 d\omega \ll \int_0^{1/T} |Y_{\phi}(\omega)|^2 d\omega \quad (23)$$

holds for any $y_{\phi}(t)$, i.e., the dominant part of spectral power density is at low frequencies with respect to the LPF bandwidth $1/T$, then $y_{\phi f}(t) \doteq y_{\phi}(t)$ and the condition in (22) can be extended to $|y_{\phi}(t)| \leq \lambda'(t)$ where $\lambda'(t) \doteq \lambda(t)$.

3. Simulation example

In this section the benefits of the proposed method will be illustrated by a simulation example. A well-known benchmark problem will be considered. It deals with a laboratory plant using two tanks with fluid flow, as was described by Zhang et al. (2004). The two cylindrical tanks are identical, with a cross section $A_s = 0.0154 \text{ m}^2$. The cross section of the connection pipe and the outlet pipe is $S_{p1} = S_{p2} = 3.6 \times 10^{-5} \text{ m}^2$, and the liquid levels in the two tanks are denoted h_1 and h_2 , respectively. The plant-setup scheme is presented in Fig. 2. The supplying flow rates coming from an electric pump to tank 1 are denoted $q_1(t)$, and there is an outflow from tank 2 denoted $q_2(t)$. Using the mass balance equations and Toricelli's rule, the following equations are obtained:

$$\begin{aligned} \dot{h}_1 &= \frac{1}{A_s} (-K_{p1} \text{sign}(h_1 - h_2) \sqrt{2g|h_1 - h_2|} + q_1), \\ \dot{h}_2 &= \frac{1}{A_s} (K_{p1} \text{sign}(h_1 - h_2) \sqrt{2g|h_1 - h_2|} - K_{p2} \sqrt{2gh_2}), \end{aligned} \quad (24)$$

where $K_{p1} = a_1 S_{p1}$ and $K_{p2} = a_2 S_{p2}$ denote the outflow constants, and g is the gravity acceleration. Let $a_1 = a_2 = 1$ for the sake of simplicity.

To get an input–output system that is similar to industrial processes the model will be modified in the following way. The input to the system is the electric-pump

voltage $u_1(t)$ that produces the inlet flow

$$\bar{q}_1(t) = K_u(1 + v_1(t))u_1(t), \quad (25)$$

where K_u is the voltage-to-flow-conversion constant, and $v_1(t)$ denotes the inaccuracy of the conversion. The only measurable output signal is the voltage of the pressure sensor, converting the fluid level $h_2(t)$ in tank 2 into the output voltage $u_2(t)$ according to the following equation:

$$u_2(t) = K_h(1 + v_2(t))h_2(t), \quad (26)$$

where K_h is the height-to-voltage-conversion constant, and $v_2(t)$ denotes the inaccuracy of the conversion. The values of the constants are $K_u = 8.8 \times 10^{-6} \text{ m}^3/\text{Vs}$ and $K_h = 16.667 \text{ V/m}$, and the upper bounds of the inaccuracies are $\bar{v}_1 = \bar{v}_2 = 0.03$.

The set of faults under consideration will follow the examples presented in Zhang et al. (2002, 2004). It will consist of the following faults:

- *Actuator fault in the pump:* A simple multiplicative actuator fault is assumed by letting the actual inlet flow be described by $q_1(t) = \bar{q}_1(t) + (1 - K_f)\bar{q}_1(t)$, where $\bar{q}_1(t)$ is the flow in the non-fault case, and $K_f \in [0, 1]$ is the fault constant.
- *Leakage in tank 1:* The leak is assumed to be circular in shape and of unknown radius r_1 . As a consequence, the outflow rate of the unknown-size leak is $q_{f1} = a_1\pi(r_1)^2\sqrt{2gh_1}$.
- *Leakage in tank 2:* Analogously to the case of leakage in tank 1, the outflow rate is $q_{f2} = a_2\pi(r_2)^2\sqrt{2gh_2}$.

With reference to the given INFUMO identification procedure, a confidence band of input–output data must be defined. This band will also include all unexpected output deviations due to parameter uncertainties. A set of 20 experiments was carried out. The inputs and associated output signals are shown in Fig. 3. For the sake of brevity, only the first, the second, and the last data sets are presented. One of the major benefits of the interval fuzzy model identification, shown in Fig. 3, is that the input signals can be arbitrary.

According to Eq. (4), the input and output signals are subjected to low-pass filtering (LPF). The structure of the LPF was chosen as a simple first-order system, represented by the transfer function in Eq. (6). The optimal design of the LPF time constant T_f was not considered in this study. The filter can also be seen as a means of trade-off between the fault-detecting speed and robustness to false alarms. The cut-off frequency ω_f was chosen according to the absolute values of the Fourier transforms of the output signals, as presented in Fig. 4. Hence, the filter time constant was defined as $T_f = 1/\omega_f = 476 \text{ s}$. The lower diagram of Fig. 4 demonstrates the absolute values of the Fourier transforms of the filtered outputs. This way a compact set of measurements that represents steady-state system behaviour is obtained. It can be seen as an approximation of a static input–output mapping area.

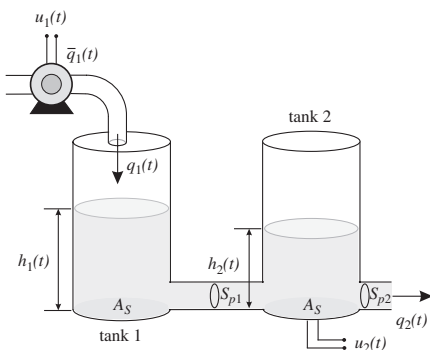


Fig. 2. Two-tank laboratory plant.

508

S. Oblak et al. / Engineering Applications of Artificial Intelligence 20 (2007) 503–510

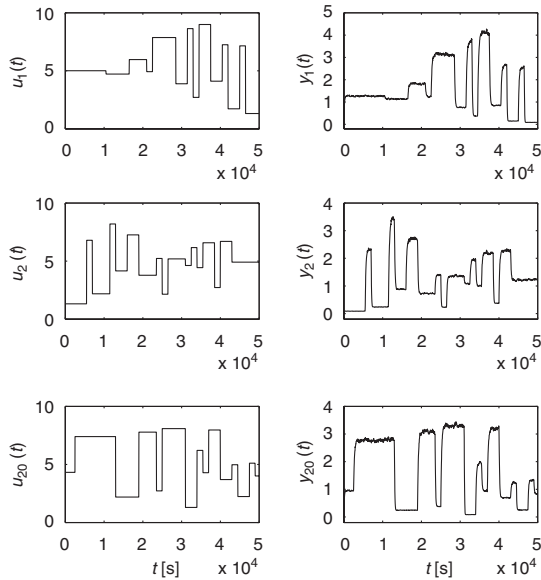


Fig. 3. Inputs and outputs: the first, the second, and the last experiment.

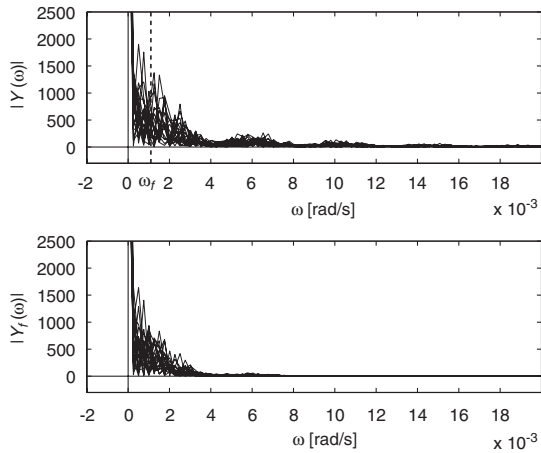


Fig. 4. The choice of the filter cut-off frequency.

The identification experiments resulted in a huge set of data. To avoid the problems with optimization convergence, data reduction is performed by determining the boundary points. Firstly, the range of input measurements is divided into equidistant subspaces. The length of the step is chosen according to the subspace with the highest density of data. In each subspace the extremal points are determined. The input–output data is presented in Fig. 5, and the resulting set of 302 boundary points is emphasized. These data were used as the training data set for the INFUMO identification. A static INFUMO can be

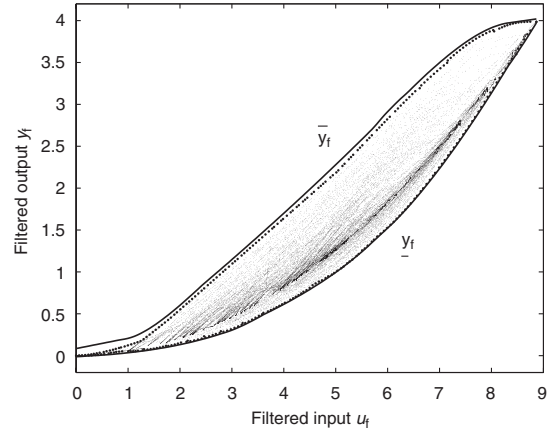


Fig. 5. Set of filtered input–output data with boundary points and boundary INFUMO functions.

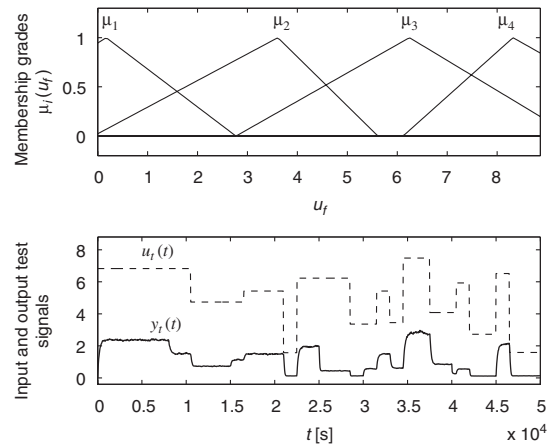


Fig. 6. Membership-function arrangement and test experiment signals.

employed. This brings additional reduction of the number of fuzzy parameters to be optimized. The membership functions of the INFUMO antecedent variables were of triangular shape and arranged using grid partitioning (Babuška, 1998). According to the data-area shape, it was sufficient to use four fuzzy subsets for the upper and lower fuzzy functions.

The parameters were optimized using the proposed INFUMO optimization algorithm in Eq. (14). The resulting boundary functions can be seen in Fig. 5. It is evident that the min–max optimization gave satisfactory results in approximating the given area.

To realize a fault-detection system, INFUMO is connected to the process in parallel, as shown in Fig. 1. In the test experiment a 20% actuator fault ($K_f = 0.2$) is assumed to occur in time period $t_{act} = 8000–15000$ s, a

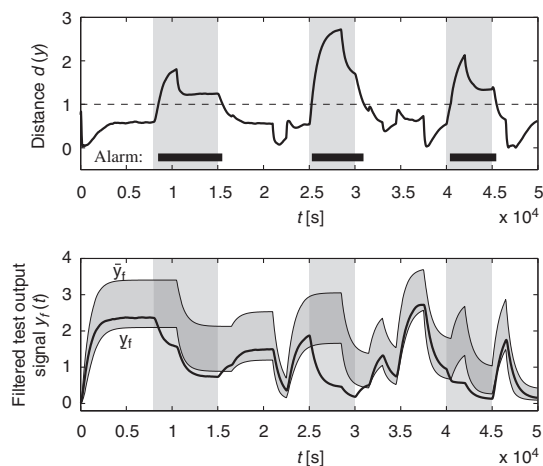


Fig. 7. Results of the fault detection system.

leakage of $r_1 = 3 \times 10^{-3}$ m in tank 1 at $t_{leak1} = 25\,000$ – $30\,000$ s, and a leakage of $r_2 = 3 \times 10^{-3}$ m in tank 2 at $t_{leak2} = 40\,000$ – $45\,000$ s, respectively. The input test signal and the corresponding process output signal are presented in the lower plot of Fig. 6. The results of the test run can be seen in Fig. 7. The first diagram demonstrates the calculation of the distance function, and in the second diagram time-dependent courses of the filtered process output y_f and the INFUMO boundary functions $\bar{y}_f, \underline{y}_f$ are shown. It is evident that the proposed FD system successfully tracks the filtered output crossing of the permitted band, denoting the expected normal system behaviour. In the shaded areas, i.e., where the faults from the fault set occurred, fault is declared with reasonably small time delays, depending on the time constant of the proposed low-pass filter: 467.7, 288.0, and 380.2 s, respectively.

4. Conclusion

A novel approach of the fault-detection system design for nonlinear input–output systems was presented. The interval fuzzy model (INFUMO), which is suitable for robust identification of nonlinear functions, was applied to the residual generation and decision stages. The benefit in fault detection is to be able to directly model a family of fault-free system responses from the confidence band, which already includes the effects of uncertainties, based only on the input–output data. Employing the low-pass filtering to obtain the input–output data set, relatively simple fuzzy functions are sufficient to provide a fairly efficient fault-detection system.

The filter served as a trade-off between the speed of the FD-system and robustness to false alarms; however, investigating the performance resulting from different choices of filter structure, the extension of the presented

INFUMO-based FD method to fault isolation, and investigating possible extensions to frequency-based methods and fault-tolerant control deserves further attention. In addition, the results of the simulated example demonstrate the quality of performance coupled with simplicity of application, which is very important from the application point of view.

Acknowledgements

This work was supported by the Ministry of Higher Education, Science and Technology of Republic of Slovenia.

References

- Alcorta García, E., Frank, P.M., 1997. Deterministic nonlinear observer-based approaches to fault diagnosis: a survey. *Control Engineering Practice* 5 (5), 663–670.
- Amann, P., Perronne, J.M., Gissinger, G.L., Frank, P.M., 2001. Identification of fuzzy relational models for fault detection. *Control Engineering Practice* 9, 555–562.
- Babuška, R., 1998. *Fuzzy Modeling for Control*. Kluwer Academic Publishers, Boston, USA.
- Ballé, P., Fuessel, D., 2000. Closed-loop fault diagnosis based on a nonlinear process model and automatic fuzzy rule generation. *Engineering Applications of Artificial Intelligence* 13 (6), 695–704.
- Bastin, G., Gevers, M.R., 1988. Stable adaptive observers for nonlinear time-varying systems. *IEEE Transactions on Automatic Control* 33, 650–658.
- Chen, J., Patton, R.J., Zhang, H.-Y., 1996. Design of unknown input observers and robust fault detection filters. *International Journal of Control* 63 (1), 85–105.
- Ding, X., Guo, L., Frank, P.M., 1994. Parameterization of linear observers and its application to observer design. *IEEE Transactions on Automatic Control* 39 (8), 1648–1652.
- Fagarasan, I., Ploix, S., Gentil, S., 2004. Causal fault detection and isolation based on a set-membership approach. *Automatica* 40, 2099–2110.
- Frank, P.M., 1990. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy—a survey and some new results. *Automatica* 26, 459–474.
- Frank, P.M., Ding, X., 1997. Survey of robust residual generation and evaluation methods in observer-based fault detection systems. *Journal of Process Control* 7 (6), 403–424.
- Frank, P.M., Köppen-Seliger, B., 1997. New developments using ai in fault diagnosis. *Engineering Applications of Artificial Intelligence* 10 (1), 3–14.
- Hammouri, H., Kinnaert, M., El Yaagoubi, E., 1999. Observer-based approach to fault detection and isolation for nonlinear systems. *IEEE Transactions on Automatic Control* 44 (10), 1879–1884.
- Isermann, R., 1997. Supervision, fault-detection and fault-diagnosis methods—an introduction. *Control Engineering Practice* 5 (5), 639–652.
- Jaulin, L., Walter, E., 1993. Set inversion via interval analysis for nonlinear bounded-error estimation. *Automatica* 29 (4), 1053–1064.
- Klančar, G., Juričić, D., Karba, R., 2002. Robust fault detection based on compensation of the modelling error. *International Journal of Systems Science* 33 (2), 97–105.
- Malan, S., Milanese, M., Taragna, M., 1997. Robust analysis and design of control systems using interval arithmetic. *Automatica* 33 (7), 1363–1372.
- Patton, R.J., Chen, J., 1996. Neural networks in fault diagnosis of nonlinear dynamic systems. *Engineering Simulation* 13 (6), 905–924.

- Patton, R.J., Chen, J., 1997. Observer-based fault detection and isolation: Robustness and applications. *Control Engineering Practice* 5 (5), 671–682.
- Škrjanc, I., Blažič, S., Agamennoni, O., 2005a. Identification of dynamical systems with a robust interval model. *Automatica* 41 (2), 327–332.
- Škrjanc, I., Blažič, S., Agamennoni, O., 2005b. Interval fuzzy modelling applied to wiener models with uncertainties. *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics* 35 (5), 1092–1095.
- Takagi, T., Sugeno, M., 1985. Fuzzy identification of systems and its applications to modelling and control. *IEEE Transactions on Systems, Man, and Cybernetics* 15, 116–132.
- Vemuri, A.T., Polycarpou, M.M., 2004. A methodology for fault diagnosis in robotic systems using neural networks. *Robotica* 22, 419–438.
- Ying, H., Chen, G., 1997. Necessary conditions for some typical fuzzy systems as universal approximators. *Automatica* 33, 1333–1338.
- Zhang, X., Polycarpou, M.M., Parisini, T., 2001. Robust fault isolation for a class of non-linear input–output systems. *International Journal of Control* 74 (13), 1295–1310.
- Zhang, X., Polycarpou, M.M., Parisini, T., 2002. A robust detection and isolation scheme for abrupt and incipient faults in nonlinear systems. *IEEE Transactions on Automatic Control* 47 (4), 576–593.
- Zhang, X., Parisini, T., Polycarpou, M.M., 2004. Adaptive fault-tolerant control of nonlinear uncertain systems: An information-based diagnostic approach. *IEEE Transactions on Automatic Control* 49 (8), 1259–1274.

Dodatek B

Uporaba modela INFUMO v zaznavanju napak (VPF)

A new fault-detection system for nonlinear systems based on an interval fuzzy model

Simon Oblak, Igor Škrjanc, and Sašo Blažič

Abstract—The paper proposes a new fault-detection scheme for nonlinear input-output systems with uncertain interval-type parameters. Using a high-pass filter for detecting the system's steady-state conditions, one is able to build a confidence band of data for the system in normal operating mode. The band can be seen as an approximation of the system's nonlinear input-output mapping including the effects of uncertainties. By approximating the boundary functions of the band using an interval fuzzy model (INFUMO), we are able to build a simple and effective fault-detection system based on the boundary-crossing test and logical relations. An example of a fault-detection system for a pH neutralization process is presented to demonstrate the benefits of the proposed method.

I. INTRODUCTION

Fault detection (FD) problem for linear systems has been extensively studied over the past three decades, and a lot of powerful methods have been developed, see e.g. [5], [9], and [11]. However, the use of linear approaches is limited if the system to be monitored is strongly nonlinear. The development of nonlinear FD methods plays a significant role in industrial applications because many industrial systems are nonlinear in nature. Tackling FD in nonlinear systems, different methods, employing e.g. nonlinear observers [7] or adaptive observers [16], have been presented in recent years. Another important scope of investigation is the use of artificial intelligence, for instance fuzzy systems or artificial neural networks [1], [10], [7]. Most of the above-mentioned methods are based on a *decoupling* framework, where the modelling uncertainty and all possible faults can be decoupled through an appropriate coordinate transformation and residual generation technique. However, the modelling uncertainty is often unstructured, which makes it difficult to achieve exact decoupling between faults and modelling errors. In addition, some problems taking into consideration the input-output representation of systems as well as the design of the corresponding nonlinear observers are still open.

This paper presents a fault-detection scheme for nonlinear input-output systems with unstructured interval-type uncertainties. The presented approach is based on the use of an interval fuzzy model (INFUMO). Fuzzy systems are often applied as nonlinear-function estimators [15], and the proposed approach employs a special form of a Takagi-Sugeno-type [12] fuzzy model. As was introduced and shown in

[13], by applying a TS fuzzy model with interval parameters, one is able to approximate the upper and lower boundaries of the domain of functions that result from an uncertain system. When system parameters vary in a certain tolerance band, it is advantageous to define a confidence band over a finite set of input and output measurements in which the effects of unknown process inputs are already included. If the outer bounds of the set, obtained during fault-free operating conditions, are determined, one is able to use them as the bounds of the admissible output area, and thus being able to create a simple and intuitive fault-detection system. The main idea of the proposed approach is to use the INFUMO as an admissible-filtered-output estimator in a fault-detection system. A high-pass filter governs the steady-state estimation and gathering the confidence data, and a set of logical rules and distance calculation is used to form a smooth fault function from the residual.

The paper is organized in the following way. In Section 2 the system description, the main idea of interval fuzzy model identification, the residual formation with a fault-diagnostic scenario, and detectability issues are described. Section 3 presents the pH neutralization process, and the application to fault detection with high-pass filtering is introduced. In the final part, some outlines of future work are given.

II. USING THE FUZZY INTERVAL MODEL IN FAULT DETECTION

A. System description

Let the nonlinear system be given in a general form as follows:

$$\begin{aligned}\dot{x}(t) &= \sigma(x, u, t) + \eta(x, u, t) + \phi(x, u, t) \\ y(t) &= \gamma(x, u, t) + \rho(x, u, t)\end{aligned}\quad (1)$$

where $x \in \mathbb{R}^n$ is the state vector of the system, $u \in \mathbb{R}$ is the system input, $y \in \mathbb{R}$ denotes the system output, $\eta: \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}^n$ is the state disturbance, $\rho: \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}$ is the output disturbance, $\phi: \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}^n$ denotes the fault function, and $\sigma: \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}^n$ and $\gamma: \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^+ \rightarrow \mathbb{R}$ are the nonlinear functions of the state vector, input and time, respectively. Let us assume that only the system outputs can be directly measured. Throughout the paper the following assumptions will be made:

Assumption 1: The modelling uncertainties, represented by η and ρ in (1), are unstructured unknown nonlinear functions of x , u and t , but bounded by some known functionals [16], i.e.,

$$\begin{aligned}|\eta(x, u, t)| &\leq \bar{\eta}(y, u, t), \quad |\rho(x, u, t)| \leq \bar{\rho}(y, u, t), \\ \forall(x, y, u) &\in \mathcal{X} \times \mathcal{Y} \times \mathcal{U}, \quad \forall t \geq 0\end{aligned}\quad (2)$$

This work was supported by the Ministry of science, higher education and technology of Republic of Slovenia

S. Oblak, I. Škrjanc, and S. Blažič are with Faculty of Electrical Engineering, University of Ljubljana, Tržaška 25, SI-1000 Ljubljana, Slovenia
simon.oblak@fe.uni-lj.si

where the bounding functions $\bar{\eta}(y, u, t)$ and $\bar{\rho}(y, u, t)$ are known and uniformly bounded. $\mathcal{X} \subset \mathbb{R}^n$ is some compact domain of interest, and $\mathcal{U} \subset \mathbb{R}$ and $\mathcal{Y} \subset \mathbb{R}$ are the compact sets of admissible inputs and outputs, respectively.

Note that $\bar{\eta}$ and $\bar{\rho}$ are the functions of the output y whereas η and ρ are the functions of the state vector x . This is due to the fact that the states are not directly measured.

Assumption 2: The output functions $y(t)$, when $\phi(x, u, t) = 0$, are bounded by the following interval:

$$y(t) \in [\underline{y}(t), \bar{y}(t)] \subset \mathcal{Y} \quad (3)$$

Assumption 1 characterizes the possible modelling uncertainties as unstructured but bounded by some constant or function, and Assumption 2 guarantees that in the absence of faults the bounds of the interval can be determined. As a consequence, a confidence band of outputs guarantees that a process output exhibiting normal behaviour is found in the interval $[\underline{y}, \bar{y}]$. However, due to the unknown effect of the actual disturbance functions the exact bounds cannot be defined analytically. In the proposed approach, introducing the interval fuzzy model, the boundary responses will be obtained by a fuzzy function approximation of the bounds of a set of filtered input-output data that already comprises the effects of disturbances.

As was shown in [3] and [6], a residual generator can be designed by

$$r(t) = Q(p)(y(t) - \hat{y}(t)), \quad (4)$$

with $\hat{y}(t)$ as a process-output estimation and $Q(p)$ is a filter which enhances the residual robustness to unknown process inputs and is to be designed according to the demands of the application. Since in this case only the input and output data are available and the transfer functions from the faults and disturbances are not known, the above-mentioned approach is altered accordingly. The main idea of the presented approach is to detect the process steady-state by using a high-pass filter. Let us assume that the filter is given in the following form:

$$Q(p) = \frac{p}{p + k_f}. \quad (5)$$

If we use a piece-wise constant input signal and subject it to filtering by $Q(p)$, the obtained signal will be exponentially decreasing. When its absolute value drops below a preassigned value, the process is approximately in steady-state; thus, in this period, given that no faults have occurred, we gather the input and output data $u_f(i)$ and $y_f(i)$. Such data, obtained in normal operating conditions, correspond to a set of the so-called admissible input-output data. Robustness to the effects of disturbances can be seen in the fact that the latter are already included in the set. If the set bounds are modelled as a function of the system input, they can be used online for residual formation:

$$\begin{aligned} r_{f1}(t) &= \bar{y}_f(t) - y(t) = \bar{f}(u_f, t) - y(t), \\ r_{f2}(t) &= y(t) - \underline{y}_f(t) = y(t) - \underline{f}(u_f, t). \end{aligned} \quad (6)$$

B. Interval fuzzy model

The interval fuzzy model was introduced in [13] as a means of robust system identification, and it was also used in [14] to provide a functional description of a static nonlinear area approximation.

In this paper the interval fuzzy model will be used for the modelling of $\bar{f}(u_f, t)$ and $\underline{f}(u_f, t)$ in (6). Note that for the confidence-band identification either an open-loop or a closed-loop experiment in the absence of faults has to be conducted, and the online fault detection based on the obtained INFUMO can also be realized in both systems.

The procedure of obtaining the exact lower and upper bounds for the family of functions and assigning the INFUMO to them will be briefly reviewed. Let $C \subset \mathbb{R}$ be a compact set and $\mathcal{G} = \{g(z) : C \rightarrow \mathbb{R}\}$ be a class of nonlinear functions. Let us assume that there exist the exact upper bound \bar{g} and the exact lower bound \underline{g} that satisfy the following conditions for an arbitrary $\varepsilon > 0$ and for each $z \in C$:

$$\bar{g}(z) \geq \max_{g \in \mathcal{G}} g(z), \exists g \in \mathcal{G} : \bar{g}(z) < g(z) + \varepsilon \quad (7)$$

$$\underline{g}(z) \leq \min_{g \in \mathcal{G}} g(z), \exists g \in \mathcal{G} : \underline{g}(z) > g(z) + \varepsilon \quad (8)$$

Obtaining the bounds in Eqs. (7) and (8) would require an infinite amount of data; however, in this case we are limited to the finite set of measured filtered output values $Y = \{y_{f1}, y_{f2}, \dots, y_{fN}\}$ and the finite set of filtered input data $Z = \{z_1, z_2, \dots, z_N\} = \{u_{f1}, u_{f2}, \dots, u_{fN}\}$:

$$y_{fi} = g(z_i), g \in \mathcal{G}, z_i \in C \subset \mathbb{R}, y_i \in \mathbb{R}, i = 1, \dots, N \quad (9)$$

The exact upper and lower boundary functions will be approximated by fuzzy functions, and the Stone-Weierstrass theorem [15] is extended to prove that there exist fuzzy systems \bar{f} and \underline{f} such that the bounds of an arbitrary nonlinear area can be approximated with any precision, i.e.,

$$\begin{aligned} 0 &< \bar{f}(z_i) - g(z_i) < \varepsilon, \quad \forall i, \\ -\varepsilon &< \underline{f}(z_i) - g(z_i) < 0, \quad \forall i. \end{aligned} \quad (10)$$

In this paper an INFUMO TS-type model with one antecedent variable and singleton consequent part will be assumed. It can be given as a set of rules

$$\begin{aligned} R_j : \text{if } x_p \text{ is } A_j \text{ then } \bar{f}(z_i) &= \bar{y}_f = \bar{\theta}_j^T, \quad j = 1, \dots, m \\ \underline{f}(z_i) &= \underline{y}_f = \underline{\theta}_j^T, \quad j = 1, \dots, m \end{aligned} \quad (11)$$

The antecedent variable $x_p = u_f \in \mathbb{R}$ denotes the input or variable in premise, and variables $\bar{y}_f, \underline{y}_f \in \mathbb{R}$ are the outputs of the interval fuzzy model that provide the upper and lower boundary function. The antecedent variable is connected with m fuzzy sets A_j , and each fuzzy set A_j ($j = 1, \dots, m$) is associated with a real-valued function $\mu_{A_j}(x_p) : \mathbb{R} \rightarrow [0, 1]$, that produces a membership grade of the variable x_p with respect to the fuzzy set A_j . The system outputs are linear combinations of the consequent states, and $\bar{\theta}_j, \underline{\theta}_j \in \mathbb{R}^2$ are

vectors of fuzzy parameters. The system in Eq. (11) can be described in closed form

$$\begin{aligned}\bar{y}_f &= \beta^T(x_p)\bar{\Theta}, \\ \underline{y}_f &= \beta^T(x_p)\underline{\Theta},\end{aligned}\quad (12)$$

where $\bar{\Theta}^T = [\bar{\theta}_1, \dots, \bar{\theta}_m]$ and $\underline{\Theta}^T = [\underline{\theta}_1, \dots, \underline{\theta}_m]$ denote the upper and lower coefficient matrices for the complete set of rules, and $\beta^T(x_p) = [\beta_1(x_p), \dots, \beta_m(x_p)]$ is a vector of normalized membership functions with elements that indicate the degree of fulfilment of the respective rule. Functions $\beta_j(x_p)$ can be defined as

$$\beta_j(x_p) = \frac{\mu_{A_j}(x_p)}{\sum_{j=1}^m \mu_{A_j}(x_p)}, \quad j = 1, \dots, m, \quad (13)$$

if the partition of unity is assumed. Note that $\beta_j(x_p)$ are equal in both the upper and lower boundary-function calculations in (12).

The membership functions $\mu_{A_j}(x_p)$ and the fuzzy parameters in $\bar{\Theta}$, $\underline{\Theta}$ of the proposed INFUMO have to be determined for optimal approximation of the bounds. While the proper arrangement and shape of the membership functions will not be considered—the mentioned can be approached by different clustering methods [2]—for the optimal fuzzy parameters the minimization of the maximum modelling error ε in Eq. (10) over the whole input set Z is performed. This implies the *min-max* optimization method, and l_∞ -norm is used as the modelling error measure, yielding

$$\begin{aligned}\min_{\bar{\Theta}} \max_{z_i \in Z} |y_i - \underline{f}(z_i)| \quad \text{s.t. } y_i - \underline{f}(z_i) &\geq 0, \\ \min_{\underline{\Theta}} \max_{z_i \in Z} |y_i - \bar{f}(z_i)| \quad \text{s.t. } y_i - \bar{f}(z_i) &\leq 0,\end{aligned}\quad (14)$$

where $\underline{f}(z_i) = \beta^T \underline{\Theta}$ and $\bar{f}(z_i) = \beta^T \bar{\Theta}$. Note that the data are obtained by sampling different functions from \mathcal{G} with arbitrary values of z_i . The solutions to both problems can be found by linear programming, because both problems can be viewed as linear programming problems. This brings simplicity to the realization of the optimizing process. However, large data sets and a large number of parameters will still pose a threat to optimization convergence. In the first case the problems are approached with data-reduction methods, and in the latter case, on the other hand, solutions to reduce the number of parameters have to be found.

C. Formation of the proposed fault-detection system

When the parameter identification is done, the INFUMO is connected to the process in parallel to get online estimations of the boundary outputs. Fig. 1 gives a schematic representation of the proposed fault-detection system. The filter $Q(p)$ is represented by a block denoted $s/s+k_f$ (high-pass filter).

In terms of fault detection, the decision function should consist of verifying that each measurement belongs to the corresponding confidence band. In order to provide normalized quantitative information about the proximity of the measurements to the closest interval bound, distances were used, as presented by [4]. If a filtered output value $y_f(t)$

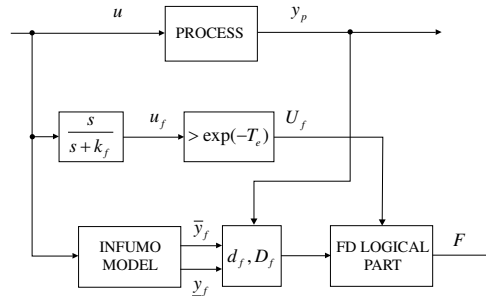


Fig. 1. Fault-detection system using a static INFUMO and high-pass filtering

belongs to an interval $[\underline{y}_f(t), \bar{y}_f(t)]$, and if $\tilde{y}_f(t)$ denotes the mean interval value

$$\tilde{y}_f(t) = \frac{\bar{y}_f(t) + \underline{y}_f(t)}{2} \quad (15)$$

the proposed distance is defined in the following way:

$$\begin{aligned}\text{if } y_f(t) < \tilde{y}_f(t), \quad d_f(y_f) &= \frac{y_f(t) - \tilde{y}_f(t)}{\underline{y}_f(t) - \tilde{y}_f(t)} \\ \text{if } y_f(t) > \tilde{y}_f(t), \quad d_f(y_f) &= \frac{y_f(t) - \tilde{y}_f(t)}{\bar{y}_f(t) - \tilde{y}_f(t)}\end{aligned}\quad (16)$$

Note that by "filtered output" $y_f(t)$ we mean the actual process output $y_p(t)$ in estimated steady-state conditions, where the estimation is governed by $Q(p)$. The distance in (16) is zero when the measurement is equal to \tilde{y}_f , and approaches the value 1 if the measurement is close to one of the interval bounds. Faults can then be characterized using the following binary distance function D_f :

$$D_f = \begin{cases} 1, & d_f(y_f) > 1 \\ 0, & d_f(y_f) \leq 1 \end{cases} \quad (17)$$

However, due to the fact that the confidence data was gathered in approximation of the process steady-state, declaration of a fault is not certain during the transient. In other words, filtered input signal should be taken into account also during the FD procedure. Therefore, let us introduce a binary signal U_f as

$$U_f = \begin{cases} 1, & u_f(t) > \exp(-T_e) \\ 0, & u_f(t) \leq \exp(-T_e) \end{cases} \quad (18)$$

where T_e denotes the multiple of the filter time constant. The value of T_e determines the approximation of steady-state conditions. For instance, choosing $T_e = 5$, a filtered input signal will be considered to be in steady-state after reaching 99.33% of its actual steady-state value. We can also see the choice of T_e as a trade-off between the width of the confidence band (robustness to false alarms) and better detectability (detection speed).

WeA11.5

When $U_f = 1$, the system is in transient and D_f should not be considered for fault detection. Therefore, we should determine the samples in which U_f increases and decreases its value. Let us denote the i th time instant, in which U_f changes the value from zero to one, by t_i^\uparrow , and the reverse change by t_i^\downarrow . The logical fault function that gives a smooth fault signal during the transients is defined as

$$F(t) = \begin{cases} D_f(t_i^\uparrow), & t_i^\uparrow \leq t < t_i^\downarrow, \quad i = 1, 2, \dots \\ D_f(t), & t_{i-1}^\downarrow \leq t < t_i^\uparrow, \quad i = 1, 2, \dots \end{cases} \quad (19)$$

with $t_0^\uparrow = t_0$ as the starting time of the experiment. This way the value of D_f is held during the transient periods. While it is true that the latter introduces a delay of $t_i^\downarrow - t_i^\uparrow$ to detection time, it helps eliminating false alarms substantially.

D. Detectability properties

To study the detectability properties of the proposed method the worst-case scenario will be considered. In this sense the following theorem guarantees detectability of a fault ϕ :

Theorem 1: Let the contribution of a fault $\phi(x, u, t)$ to the filtered process output be denoted $y_{\phi f}(t)$, and the difference between the boundary functions $\bar{y}_f(t) - \underline{y}_f(t) = \lambda(t)$. A fault will certainly be detected if the absolute value of the former is greater than the width of the output confidence band, i.e.

$$|y_{\phi f}(t)| > \lambda(t). \quad (20)$$

Proof. For the purpose of fault-detectability analysis, the output function in Eq. (1) will be rewritten to

$$y_f(t) = y_{0f}(t) + \rho_f(t) + y_{\phi f}(t), \quad (21)$$

where $y_{0f}(t)$, $\rho_f(t)$ and $y_{\phi f}(t)$ denote the filtered contributions of the "undisturbed" output, disturbances and fault, respectively. Since the upper and lower confidence-band bounds $\bar{y}_f(t)$ and $\underline{y}_f(t)$ were obtained in the experiments where no faults were present, the following inequalities hold:

$$\begin{aligned} y_{0f}(t) + \rho_f(t) - \underline{y}_f(t) &\geq 0 \\ \bar{y}_f(t) - y_{0f}(t) - \rho_f(t) &\geq 0. \end{aligned} \quad (22)$$

From this point forward we will omit the notation t for the sake of clarity. By adding the fault contribution $y_{\phi f}$ to both sides of (22), the following can be written:

$$\begin{aligned} y_{0f} + y_{\phi f} + \rho_f - \underline{y}_f &\geq y_{\phi f} \Rightarrow y_{\phi f} \leq y_f - \underline{y}_f \\ \bar{y}_f - y_{0f} - \rho_f - y_{\phi f} &\geq -y_{\phi f} \Rightarrow \bar{y}_f - y_f \geq -y_{\phi f} \end{aligned} \quad (23)$$

Hence, $y_{\phi f}$ is bounded by $y_f - \bar{y}_f \leq y_{\phi f} \leq y_f - \underline{y}_f$. When no fault is detected, i.e., $\underline{y}_f \leq y_f \leq \bar{y}_f$, inequality (23) can be further extended to

$$\begin{aligned} y_{\phi f} \leq y_f - \underline{y}_f &= (\bar{y}_f - \underline{y}_f) + (y_f - \bar{y}_f) \\ &= \lambda + (y_f - \bar{y}_f) \leq \lambda + 0 = \lambda \\ -y_{\phi f} \leq \bar{y}_f - y_f &= (\bar{y}_f - \underline{y}_f) + (\underline{y}_f - y_f) \\ &= \lambda + (\underline{y}_f - y_f) \leq \lambda + 0 = \lambda \end{aligned} \quad (24)$$

Both inequalities in (24) can be united to

$$|y_{\phi f}| \leq \lambda. \quad (25)$$

When a fault is not detected, inequality (25) is satisfied, thus implying that in the case when $|y_{\phi f}| > \lambda$ the fault is certainly detected. \square

III. SIMULATION EXAMPLE

The proposed method was tested on a mathematical model of a pH neutralization process, adopted from the work by [8]. The example consists of a neutralization reaction between a strong acid (HCl) and a strong base (NaOH) in the presence of a buffer agent (NaHCO₃). The neutralization takes place in a continuous stirred tank reactor (CSTR) with a constant volume V .

Figure 2 shows the scheme of the continuous pH neutralization process. An acidic solution with volumetric flow $q_A(t)$ of a composition $x_{1i}(t)$ is neutralized using an alkaline solution with volumetric flow $q_B(t)$ of known composition consisting base x_{2i} and buffer agent x_{3i} . Due to the high reaction rates of the acid-base neutralization, chemical equilibrium conditions are instantaneously achieved. Moreover, under the assumption that the acid, base and buffer are strong enough, the total dissociation the three compounds takes place.

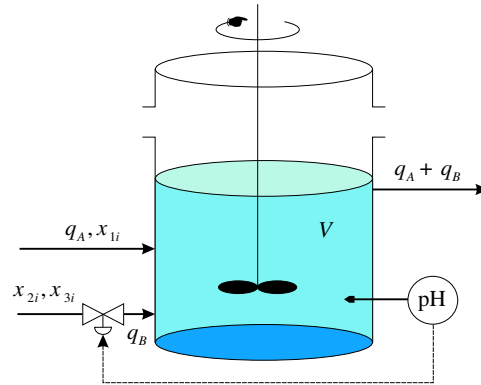


Fig. 2. pH-neutralization-process scheme

The dynamic behaviour of this process is given by the following equations:

$$\begin{aligned} \dot{x}_1 &= \frac{1}{\theta} \cdot (x_{1i} - x_1) - \frac{1}{\theta} \cdot x_1 \cdot u, \\ \dot{x}_2 &= -\frac{1}{\theta} \cdot x_2 + \frac{1}{\theta} \cdot (x_{2i} - x_2) u, \\ \dot{x}_3 &= -\frac{1}{\theta} \cdot x_3 + \frac{1}{\theta} \cdot (x_{3i} - x_3) u, \end{aligned} \quad (26)$$

$$h(x, \xi) = \xi^3 + (K_w/K_x + x_2 + x_3 - x_1)\xi^2 + (x_2 - x_1 + K_x)\xi - K_w^2/K_x = 0 \quad (27)$$

where $\xi = 10^{-pH}$, $\theta = V/q_A$ and $u = q_B/q_A$. K_w and K_x are the dissociation constants of the buffer and water, respectively. The parameters of the system represented by (26)-(27) are $x_{1i} = 0.0012$ mol HCl/L, $x_{2i} = 0.0020$ mol NaOH/L, $x_{3i} = 0.0025$ mol NaHCO₃/L, $K_x = 10^{-7}$ mol/L, $K_w = 10^{-14}$ mol²/L² and $V = 2.5$ L. The output $pH(t) = -\log_{10} \xi(t)$ is assumed to be corrupted with noise of maximal amplitude $|\bar{p}| = 0.1$.

With reference to the given INFUMO identification procedure, an experiment using a piece-wise constant input signal was carried out. The input and associated output signals are shown in Fig. 3.

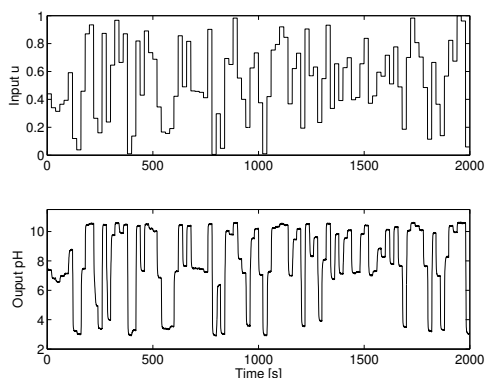


Fig. 3. Input and the corresponding output signal

The parameters of the filter $Q(p)$ were chosen to be $k_f = 1$ and $T_e = 7$, i.e., the process was assumed to be in steady-state after seven filter time constants $1/k_f$. The data was gathered in a filtered set (u_f, y_f) , which is shown in the upper diagram of Fig. 4. In the same diagram one can see the resulting INFUMO boundary functions, and in the lower diagram of the figure the corresponding membership functions are presented. The number of the membership functions was chosen to be 10 and the positions of the apexes were calculated using a c-means clustering [2]. The values of the apex positions and the INFUMO parameters are:

$$\begin{aligned} \alpha &= [0.001 \ 0.19 \ 0.23 \ 0.26 \ 0.33 \ 0.51 \ 0.55 \ 0.61 \ 0.68 \ 0.90] \\ \bar{\theta} &= [3.07 \ 3.63 \ 4.18 \ 6.10 \ 6.99 \ 7.82 \ 8.17 \ 9.44 \ 10.22 \ 10.65] \\ \underline{\theta} &= [2.78 \ 3.32 \ 3.55 \ 4.62 \ 6.70 \ 7.63 \ 7.70 \ 9.05 \ 9.89 \ 10.453] \end{aligned} \quad (28)$$

It is evident that the approximation on the given boundaries is satisfactory. This way a functional description of a compact set of measurements, representing system's steady-state behaviour, is obtained. It can also be seen as an approximation of a static input-output mapping area, in this case the pH titration curve.

To realize a fault-detection system, the INFUMO is connected to the process in parallel, as shown in Fig. 1. In the test experiment a 10% fault in the pH sensor is assumed

WeA11.5

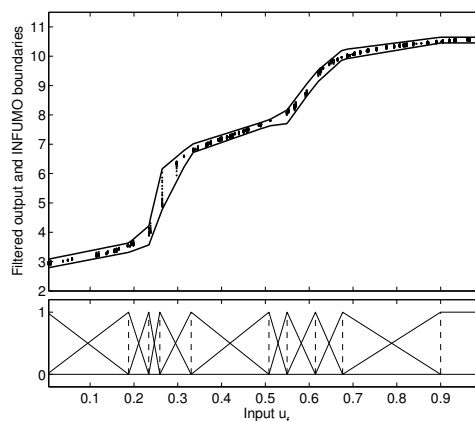


Fig. 4. Set of filtered input-output data with boundary INFUMO functions (upper), corresponding membership functions (lower)

to occur in time period $t_{fs} = 200 - 400$ s. The input test signal and the corresponding process output signal are presented in the lower diagram of Fig. 5. Also notice the shaded area denoting the period of the fault occurrence. The

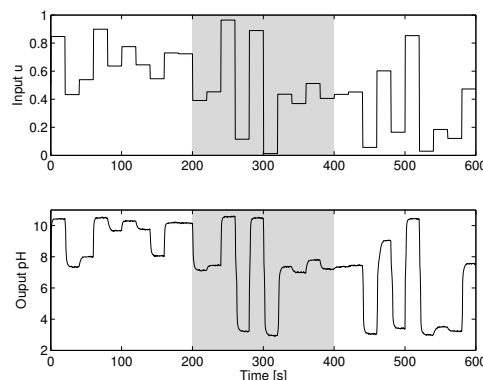


Fig. 5. Test experiment signals containing effects of faults

results of the test run can be seen in Figs. 6 and 7. The first diagram demonstrates the distance-calculation, and in the second diagram time-dependent courses of the process output y_p and the INFUMO boundary functions $\bar{y}_f, \underline{y}_f$ are shown. In Fig. 7 one can notice high values of the distance d_f in the period when the fault has not occurred. However, those values coincide with the periods when the value of the filtered input signal is above the pre-set value, i.e., the systems is in transient. Fig. 8 presents the binary signals $D_f(t)$ and $U_f(t)$ during the experiment. Finally, the fault signal $F(t)$ is calculated and presented in Fig. 9. The dots

WeA11.5

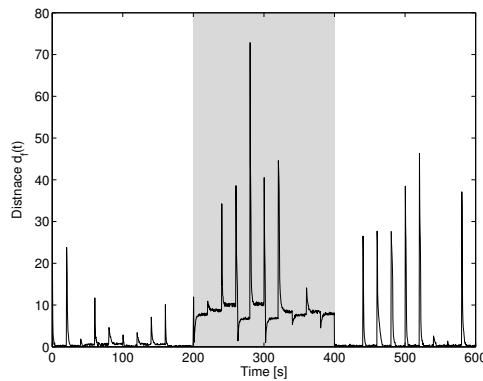


Fig. 6. Results of the fault detection system - distance

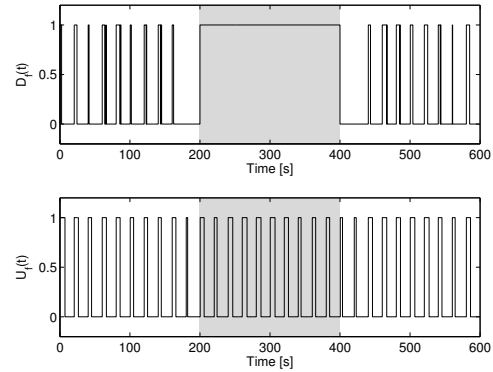


Fig. 8. Logical part of the FD system

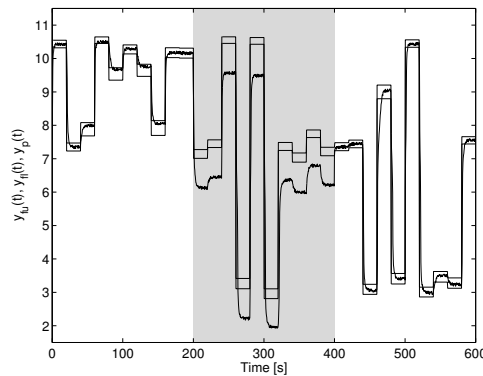


Fig. 7. Results of the fault detection system - INFUMO

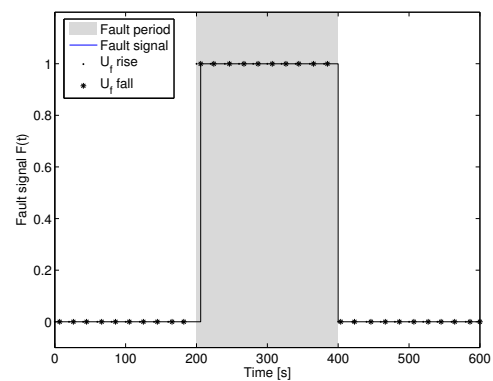


Fig. 9. Fault signal

and stars in the diagram denote the risings and drops of the binary filter output U_f (time instants t_i^+ and t_i^-). One can see that the fault signal only changes its values at the filter-output drop. It is evident that the proposed FD system successfully tracks the occurrence of the fault. The fault was declared with reasonably small time delays, depending mainly on the filter time constant. This is due to the fact that the proposed method tightens the bounds of the confidence band to minimum, and no further delay is introduced on account of the band width.

IV. CONCLUSION

A novel approach of the fault-detection system design for nonlinear input-output systems was presented. The interval fuzzy model (INFUMO) was applied to the residual generation and decision stages. The benefit in fault detection is to be able to directly model a family of fault-free system responses from the confidence band, which already includes the effects of uncertainties, based only on the input-output

data. High-pass filtering was employed to obtain the input-output data set. The choice of filter constant served as a trade-off between the speed of the FD-system and robustness to false alarms. As we are dealing with a confidence data set, simple and intuitive calculations and logical rules are sufficient to form an effective fault-detection system. The results of the simulated example demonstrate the quality of performance coupled with simplicity of application, which is very important from the application point of view. However, some issues deserve further attention:

- Investigating the performance resulting from different choices of filter structure;
- The extension of the presented INFUMO-based FD method to fault isolation;
- Investigating possible extensions to frequency-based methods and fault-tolerant control;
- Performance of the proposed system, when exposed to slowly varying changes, incipient faults, and outliers.

WeA11.5

REFERENCES

- [1] E. Alcorta García and P. M. Frank, "Deterministic nonlinear observer-based approaches to fault diagnosis: A survey," *Control Engineering Practice*, vol. 5, no. 5, pp. 663–670, 1997.
- [2] R. Babuška, *Fuzzy Modeling for Control*. Boston, USA: Kluwer Academic Publishers, 1998.
- [3] X. Ding, L. Guo, and P. M. Frank, "Parameterization of linear observers and its application to observer design," *IEEE Transactions on Automatic Control*, vol. 39, no. 8, pp. 1648–1652, 1994.
- [4] I. Fagarasan, S. Ploix, and S. Gentil, "Causal fault detection and isolation based on a set-membership approach," *Automatica*, vol. 40, pp. 2099–2110, 2004.
- [5] P. M. Frank, "Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy - a survey and some new results," *Automatica*, vol. 26, pp. 459–474, 1990.
- [6] P. M. Frank and X. Ding, "Survey of robust residual generation and evaluation methods in observer-based fault detection systems," *J. Proc. Control*, vol. 7, no. 6, pp. 403–424, 1997.
- [7] P. M. Frank and B. Köppen-Seliger, "New developments using ai in fault diagnosis," *Engineering Applications of Artificial Intelligence*, vol. 10, no. 1, pp. 3–14, 1997.
- [8] O. Galán, J. A. Romagnoli, and A. Palazoglu, "Robust h_∞ control of nonlinear plants based on multi-linear models: an application to a bench-scale ph neutralization reactor," *Chemical Engineering Science*, vol. 55, pp. 4435–4450, 2000.
- [9] R. Isermann, "Supervision, fault-detection and fault-diagnosis methods - an introduction," *Control Engineering Practice*, vol. 5, no. 5, pp. 639–652, 1997.
- [10] R. J. Patton and J. Chen, "Neural networks in fault diagnosis of nonlinear dynamic systems," *Engineering Simulation*, vol. 13, no. 6, pp. 905–924, 1996.
- [11] —, "Observer-based fault detection and isolation: Robustness and applications," *Control Engineering Practice*, vol. 5, no. 5, pp. 671–682, 1997.
- [12] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modelling and control," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 15, pp. 116–132, 1985.
- [13] I. Škrjanc, S. Blažič, and O. Agamennoni, "Identification of dynamical systems with a robust interval model," *Automatica*, vol. 41, no. 2, pp. 327–332, 2005.
- [14] —, "Interval fuzzy modelling applied to wiener models with uncertainties," *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, vol. 35, no. 5, pp. 1092–1095, 2005.
- [15] H. Ying and G. Chen, "Necessary conditions for some typical fuzzy systems as universal approximators," *Automatica*, vol. 33, pp. 1333–1338, 1997.
- [16] X. Zhang, M. M. Polycarpou, and T. Parisini, "Robust fault isolation for a class of non-linear input-output systems," *International Journal of Control*, vol. 74, no. 13, pp. 1295–1310, 2001.

Dodatek C

Uporaba metode prediktivnega vodenja CTWMPC

J Intell Robot Syst (2006) 47: 125–137
DOI 10.1007/s10846-006-9075-z

A Comparison of Fuzzy and CPWL Approximations in the Continuous-time Nonlinear Model-predictive Control of Time-delayed Wiener-type Systems

Simon Oblak · Igor Škrjanc

Received: 24 May 2006 / Accepted: 19 August 2006 /
Published online: 30 September 2006
© Springer Science + Business Media B.V. 2006

Abstract This paper deals with a novel method of continuous-time model-predictive control for nonlinear time-delayed systems. The problems relating to time delays are solved by incorporating the Smith-predictor scheme in a control-law derivation. A nonlinear-mapping approximation, employing either continuous piece-wise linear functions or a fuzzy system, is also an integral part of the control scheme, and thus removes the need for output-function invertibility. An illustrative experiment is conducted to compare the control quality in both approaches when tackling a time-delayed Wiener-type system control.

Key words continuous systems · fuzzy systems · nonlinear predictive control · piece-wise linear functions · time-delayed systems · Wiener-type model

1 Introduction

Wiener-type systems are a special class of nonlinear systems that are mainly used for modelling nonlinear processes encountered in process industries, for example, pH neutralization processes [3, 16] and distillation processes [17]. The large number of Wiener-system-based predictive-control methods developed in recent years [9, 12, 16] indicates the interest in this field in the control community. An additional difficulty is that some of the systems also incorporate pure time delays [21] as a consequence of non-ideal situations such as non-ideal mixing and transportation time delays. This fact means that some of the “classical” approaches are unable to reach a satisfactory level of control quality.

Nonlinear-model predictive-control (NMPC) methods were initially developed as an extension of extremely successful model-predictive control (MPC) methods [8] based on linear models. The key motivation was to employ more accurate

S. Oblak (✉) · I. Škrjanc
Faculty of Electrical Engineering, University of Ljubljana, Tržaška 25, 1001 Ljubljana, Slovenia
e-mail: simon.oblak@fe.uni-lj.si

nonlinear models in process prediction and optimization, and thus achieve better control quality for highly nonlinear processes and moderately nonlinear processes with large operating regimes. For the state of the art of NMPC methods the reader is referred to the papers by Morari and Lee [15] and Henson [10]. The majority of nonlinear predictive methods are based on a discrete-time representation; however, this can result in several shortcomings in terms of relying on a system approximation and inaccurate system intersample behaviour [13]. Moreover, a description of the system in a continuous-time domain is much more natural, especially in the case of the Wiener-type systems. The first attempt to resolve the issues of continuous-time predictive control was made in the 1990s by Demircioğlu and Gawthrop [6]. The discrete-time-based method of generalized predictive control (GPC) [5] was first of all reformulated into the continuous-time domain for the SISO [6] and MIMO [7] systems, and then extended to nonlinear systems [4]. Our proposed method relates to these works, and also attempts to resolve the following issues:

1. Some discrete-time NMPC methods (e.g., [2, 19]) suffer from having to solve an on-line non-convex optimization problem, which is in general computationally expensive and can lead to solutions with local minima.
2. Since the core of all model-based predictive methods is explicit-model-based optimal open-loop control [4], the model's accuracy plays a very important role. For Wiener models in particular, an inadequate nonlinear-mapping approximation can adversely affect the control quality in some operating regions.
3. When dealing with time-delayed models, in the work by Demircioğlu and Gawthrop [6] the process-model order is augmented by incorporating the Padé approximation of the time delay into the model. However, a model-order increase can lead to the bigger computational burden of a control algorithm and is, in general, avoided if possible.

The proposed approach in this paper tackles the control of time-delayed Wiener-type systems using continuous-time nonlinear model-based predictive control. The problem of pure time delays is approached by estimating the auxiliary undelayed process output and including it in the designated cost function. Thus, the original idea of the Smith predictor [20] is directly involved in the control law, and the model order does not need to be augmented. The receding-horizon strategy was combined with a cost function that, by adopting the ideas from predictive functional control [22], minimizes the difference between the future-output-prediction error and the model-prediction error. In this way the control law is closed-form optimal, and on-line optimization is not needed. The output nonlinearity of the Wiener model is approximated by using two different approaches: continuous piece-wise linear (CPWL) functions [11] and a static Takagi–Sugeno-type fuzzy system (FS) [1, 24]. The FS approximation is based on the method of robust Wiener-model identification presented by Škrjanc et al. [23]. Unlike the methods that invert the static nonlinearity and transform the control problem to a linear one [12, 16], here the CPWL and FS approximations are a part of the calculation of the output prediction. This in fact raises the key questions of this paper – how do the approximations function in an analytical prediction of the model output and what are the effects on the closed-loop control quality?

The outline of the paper is as follows. In Section 2 the CPWL and FS functions are introduced. In Section 3 the model-output predictions for both cases are formulated

in the continuous-time domain, and the nonlinear predictive control law is derived. Section 4 gives a comparison of the closed-loop-control results for both cases using a simple and illustrative example. Section 5 presents the conclusions.

2 Problem Formulation

Let us assume a nonlinear time-delayed continuous-time system

$$\begin{aligned}\dot{x}_p(t) &= f(x_p(t), u(t)) \\ y_p(t) &= g(x_p(t - T_d))\end{aligned}\quad (1)$$

where $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ and $g: \mathbb{R}^n \rightarrow \mathbb{R}$ are smooth functions, $x_p \in \mathbb{R}^n$ is a vector of n state variables, T_d denotes the time delay, $u \in \mathbb{R}$ is a process input and $y_p \in \mathbb{R}$ is a process output. The process input is bounded by $u_l \leq u(t) \leq u_u$. An optimal-control problem is in general defined as the design of a controller that asymptotically stabilizes a closed-loop system in such a way that the process output, $y_p(t)$, optimally follows a prescribed reference trajectory, $y_r(t)$, according to the given performance index. The solution of the classical optimal-control problem is difficult, and in this paper it is avoided by use of the moving-horizon control concept [4, 5, 14]. Furthermore, the system's nonlinearity presents an additional difficulty in terms of system modelling and control. This problem can be successfully solved by using a Wiener-type system that has a special structure that facilitates its application to model-based predictive control. The Wiener time-delayed system has the structure of a dynamic linear block

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t) \\ v(t) &= Cx(t - T_d)\end{aligned}\quad (2)$$

where $A \in \mathbb{R}^n \times \mathbb{R}^n$, $B \in \mathbb{R}^n$ and $C \in \mathbb{R}^n$. The variable $v(t) \in \mathbb{R}$ represents the intermediate variable that when connected in series with a static nonlinearity forms the model output

$$y(t) = h(v(t)), \quad (3)$$

where $h: \mathbb{R} \rightarrow \mathbb{R}$ denotes the static nonlinear mapping and $y \in \mathbb{R}$ is the process-model output. Furthermore, we assume the so-called undelayed linear system, the output of which forms the auxiliary model output containing no time delays:

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t) \\ \bar{v}(t) &= Cx(t) \\ \bar{y}(t) &= h(\bar{v}(t))\end{aligned}\quad (4)$$

The static nonlinearities in this paper are approximated by using CPWL functions [11] and fuzzy systems [1]. The stress will be on investigating the differences in the performance of closed-loop continuous-time model predictive control when the approximations are used individually in the model-output prediction.

2.1 CPWL Approximation

The process-model output using the CPWL approximation is defined as

$$y_{mp}(t) = \hat{h}_{mp}(v(t)) = \Theta^T \Lambda(v(t)), \tag{5}$$

where $\Theta^T \in \mathbb{R}^{\sigma+1}$ and $\Lambda \in \mathbb{R}^{\sigma+1}$. Using the CPWL approximation, any nonlinear function h can be uniquely represented by the segmentation of its input domain. Let us consider the segmentation into σ segments by the parameters α_i , with $\alpha_0 \leq \alpha_1 \leq \dots \leq \alpha_\sigma$. In addition, the elements of the basis functions can be expressed as

$$\Lambda(v) = \begin{bmatrix} 1 \\ \frac{1}{2}(v - \alpha_0 + |v - \alpha_0|) \\ \vdots \\ \frac{1}{2}(v - \alpha_{\sigma-1} + |v - \alpha_{\sigma-1}|) \end{bmatrix} \tag{6}$$

and the vector of the parameters is defined as

$$\Theta^T = [\theta_0, \theta_1, \dots, \theta_\sigma]. \tag{7}$$

The locations of the segments are chosen by clustering algorithms [18], and the vector of the parameters can be calculated using common least-square algorithms.

2.2 Fuzzy-system Approximation

A fuzzy TS-type system in affine form with one antecedent variable and two consequent parameters is assumed. It can be given as a set of rules in the form

$$\mathbf{R}_j : \text{if } x_p \text{ is } \mathbf{A}_j, \text{ then } y_{mf} = \theta_{j,0} + \theta_{j,1}x_{c1}, \tag{8}$$

where $j = 1, \dots, m$ is the number of fuzzy rules. The variable x_p denotes the input or variable in premise, and the variable y is the output of the model. The antecedent variable is connected with m fuzzy sets \mathbf{A}_j , and each fuzzy set \mathbf{A}_j ($j = 1, \dots, m$) is associated with a real-valued function $\mu_{A_j}(x_p) : \mathbb{R} \rightarrow [0, 1]$, that produces a membership grade of the variable x_p with respect to the fuzzy set \mathbf{A}_j . The consequent vector is denoted $x_c^T = [1, x_{c1}]$, and it implicitly represents an additional input to the fuzzy system. The system output is a linear combination of the consequent states. The system in Eq. 8 can be described in closed form

$$y_{mf} = \beta^T(x_p)\Theta_f x_c, \tag{9}$$

where the membership vector $\beta^T(x_p) = [\beta_1(x_p), \dots, \beta_m(x_p)]$ is composed of normalized degrees of fulfilment

$$\beta_j(x_p) = \frac{\mu_{A_j}(x_p)}{\sum_{j=1}^m \mu_{A_j}(x_p)}, \quad j = 1, \dots, m, \tag{10}$$

and the matrix of fuzzy-model parameters

$$\Theta_f^T = [\theta_1 \dots \theta_m] \tag{11}$$

is composed of vectors of parameters in individual fuzzy domains:

$$\theta_j^T = [\theta_{j,0} \ \theta_{j,1}], \quad j = 1, \dots, m \tag{12}$$

In this particular case the products of the parameter vectors and the consequent vectors, $\theta_j x_c$, form affine output functions. It is obvious that $\sum_{j=1}^m \beta_j(x_p) = 1$ irrespective of x_p as long as the denominator of $\beta_j(x_p)$ in Eq. 10 is not equal to zero (which can easily be prevented by stretching the membership functions over the whole potential area of x_p).

Using the intermediate variable $v(t)$ as the antecedent variable x_p , the nonlinear output mapping can be written in closed form as

$$y_{mf}(t) = \hat{h}_{mf}(v(t)) = \beta^T(v(t))\Theta_f x_c(v(t)), \quad (13)$$

where $\beta^T \in \mathbb{R}^m$, $\Theta_f \in \mathbb{R}^m \times \mathbb{R}^2$ and $x_c \in \mathbb{R}^2$.

3 Nonlinear Model-predictive Control of Wiener-type Time-delayed Systems

In general the objective of a model-predictive control law is to drive the predicted future output of a system as close as possible to the future reference, subject to the input constraints. In the continuous-time framework this implies that the predictions of the reference and the process output must be either known or estimated. Let us define the reference model by the triple in state-space as A_r , B_r and C_r and denote the reference signal as $w(t)$. In the moving time frame the model-output prediction at time τ can be approximated by a truncated Maclaurin series expansion

$$y(t + \tau|t) = \Gamma^T(\tau)Y(t) \quad (14)$$

where the vectors Γ and Y are given by

$$\Gamma(\tau) = \left[1 \quad \tau \quad \dots \quad \frac{\tau^i}{i!} \quad \dots \quad \frac{\tau^{n_y}}{n_y!} \right]^T, \quad (15)$$

$$Y(t) = \left[y(t) \quad y^{[1]}(t) \quad \dots \quad y^{[i]}(t) \quad \dots \quad y^{[n_y]}(t) \right]^T, \quad (16)$$

with $Y \in \mathbb{R}^{n_y}$, n_y is the output order, and $y^{[i]}(t)$ stands for the i th derivative of $y(t)$ with respect to t . Analogously, the reference-model output prediction can be defined as

$$y_r(t + \tau|t) = \Gamma^T(\tau) \cdot r \cdot w(t), \quad (17)$$

where the vector of the Markov parameters $r \in \mathbb{R}^{n_y+1}$ is defined as

$$r = \left[0 \quad C_r B_r \quad C_r A_r B_r \quad \dots \quad C_r A_r^{n_y-1} B_r \right]^T. \quad (18)$$

Based on the idea of the Smith predictor, when dealing with systems containing pure time delays the future reference must be compared to the undelayed process output. The future control error should decrease according to the dynamics defined by the reference model

$$e(t + \tau) = \Gamma^T r (w(t) - \bar{y}_p(t)), \quad (19)$$

where $\bar{y}_p(t)$ is the output of the estimated undelayed process output. Since the undelayed process output is not available, it has to be estimated from the actual process output and the process model. We assume that the difference between the

actual and the undelayed process outputs is equal to the difference between the delayed and the undelayed process-model outputs:

$$y_p(t) - \bar{y}_p(t) = y(t) - \bar{y}(t). \tag{20}$$

In this sense the undelayed process output \bar{y}_p can be replaced by

$$\bar{y}_p(t) = y_p(t) - y(t) + \bar{y}(t). \tag{21}$$

The idea of the proposed continuous-time MPC, referring to the predictive functional control derivation [22], is based on a minimization of the difference between the future control error and the difference between the predicted model output at time horizon τ , $\tau \in [0, T]$ and the current undelayed-model output:

$$\epsilon(t, \tau) = e(t + \tau) - (\bar{y}(t + \tau|t) - \bar{y}(t)) \tag{22}$$

The control law will be obtained by minimizing the cost function

$$V = \int_0^T \|\epsilon(t, \tau)\|^2 d\tau = \int_0^T \|e(t+\tau) - (\bar{y}(t+\tau|t) - \bar{y}(t))\|^2 d\tau. \tag{23}$$

Let us first investigate the model-output prediction (14) in the CPWL approximation case. The i th derivative of $y(t)$ is defined as

$$y_{mp}^{[i]}(t) = \Theta^T \frac{d\Lambda(v)}{dv} CA^i x(t) + \Theta^T \frac{d\Lambda(v)}{dv} [CA^{i-1} B \dots CB] U(t), \tag{24}$$

where $U(t)$ stands for

$$U(t) = [u(t) \ u^{[1]}(t) \ \dots \ u^{[l]}(t)]^T \tag{25}$$

and where

$$\frac{d\Lambda(v)}{dv} = \begin{bmatrix} 0 \\ \frac{1}{2} (1 + \text{sign}(v - \alpha_0)) \\ \vdots \\ \frac{1}{2} (1 + \text{sign}(v - \alpha_{\sigma-1})) \end{bmatrix}. \tag{26}$$

Because all of the higher derivatives of the CPWL mapping with respect to v are equal to 0 ($\frac{d^2\Lambda(v)}{dv^2} = \dots = \frac{d^n\Lambda(v)}{dv^n} = 0$), all of the higher powers of $\dot{v}(t)$ are canceled as well. This is, however, not the case when using the FS approximation. Differentiating Eq. 13 with respect to time, the first two derivatives will be as follows:

$$\dot{y}_{mf} = \frac{d\hat{h}_{mf}(v)}{dv} \cdot \dot{v}(t) \tag{27}$$

$$\ddot{y}_{mf} = \frac{d^2\hat{h}_{mf}(v)}{dv^2} \cdot (\dot{v}(t))^2 + \frac{d\hat{h}_{mf}(v)}{dv} \cdot \ddot{v}(t). \tag{28}$$

Since the first term on the right-hand side in Eq. 28 can be written as

$$\frac{d^2\hat{h}_{mf}(v)}{dv^2} = \frac{d^2\beta^T}{dv^2} \Theta_f x_c + 2 \frac{d\beta^T}{dv} \Theta_f \frac{dx_c}{dv} + \beta^T \Theta_f \frac{d^2x_c}{dv^2}, \tag{29}$$

it is obvious that, even with the assumption that the second and higher derivatives of β and x_c are equal to 0, the first term cannot be canceled, and hence the analytical definition of the output prediction is too complex. For this reason, all the terms $(\dot{v}(t))^k$, $k \geq 2$ are assumed to be 0, and the prediction problem is reformulated to be very similar to the form in the CPWL case:

$$y_{mf}^{[i]}(t) = \frac{d\hat{h}_{mf}(v)}{dv} \cdot \frac{d^i v}{dt^i} = \frac{d\hat{h}_{mf}}{dv} (CA^i x(t) + [CA^{i-1} B \dots CB] U(t)), \quad (30)$$

where

$$\frac{d\hat{h}_{mf}}{dv} = \frac{d\beta^T}{dv} \Theta_f x_c + \beta^T \Theta_f \frac{dx_c}{dv}. \quad (31)$$

Let us define the control order as follows.

Definition 1 The control order in the continuous-time predictive control is said to be n_u if the following is valid: $u^{[n_u]}(t + \tau) \neq 0$, $\forall \tau \in [0, T]$ and $u^{[i]}(t + \tau) = 0$, $\forall i > n_u$, $\tau \in [0, T]$ where $u^{[n_u]}(t + \tau)$ stands for n_u th derivative of $u(t + \tau)$ with respect to τ . The control order defines the allowable set, \mathcal{U} , of the optimal control input in the receding horizon frame, and hence imposes the constraints on $u(t + \tau)$.

Remark 1 In this paper the output order n_y and the control order n_u are two design parameters. However, there are some limitations in the choice of n_y . If the relative order of a process is denoted ρ , n_y should be at least of the same order as $n_u + \rho$ if the n_u th derivative of the control signal is to appear in the prediction, i.e., $n_y \geq n_u + \rho$.

The control vector $U(t)$ of the n_u th order is then defined as

$$U(t) = [u(t) \ u^{[1]}(t) \ \dots \ u^{[n_u]}(t)]^T. \quad (32)$$

Combining Eqs. 14, 15 and 16 with Eqs. 24 and 30, the predictions of the model outputs $y_{mp}(t + \tau|t)$ and $y_{mf}(t + \tau|t)$ at time τ are defined as

$$y(t + \tau|t) = \Gamma^T [Py(t) + Q(v)x(t) + H(v)U(t)], \quad (33)$$

where $P \in \mathbb{R}^{n_y+1}$ is

$$P = [1 \ 0 \ \dots \ 0]^T. \quad (34)$$

The matrices $Q \in \mathbb{R}^{n_y+1} \times \mathbb{R}^n$ and $H \in \mathbb{R}^{n_y+1} \times \mathbb{R}^{n_u+1}$ are calculated differently for each of the approximation cases. In the CPWL case we can write $Q(v) = q_p(v)K_q$ and $H(v) = q_p(v)K_h$, and analogously in the FS case $Q(v) = q_f(v)K_q$ and $H(v) = q_f(v)K_h$, where q_p, q_v, K_q and K_h are defined as

$$q_p(v) = \Theta^T \frac{d\Lambda(v)}{dv}, \quad (35)$$

$$q_f(v) = \frac{d\beta^T(v)}{dv} \Theta_f x_c(v) + \beta^T(v) \Theta_f \frac{dx_c(v)}{dv}, \quad (36)$$

$$K_q = [0 \ CA \ CA^2 \ \dots \ CA^{n_y}]^T, \quad (37)$$

and

$$K_h = \begin{bmatrix} 0 & \cdots & \cdots & 0 \\ CB & 0 & \cdots & 0 \\ CAB & CB & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ CA^{n_y-1}B & CA^{n_y-2}B & \cdots & CA^{n_y-1-n_u}B \end{bmatrix}. \tag{38}$$

Given the prediction of the process-model output in Eq. 33, the cost function (23) is

$$V(U, t) = \int_0^T \left((w - \bar{y}_p)^T r^T - U^T H^T - x^T Q^T \right) \Gamma \Gamma^T (r(w - \bar{y}_p) - HU - Qx) dt \tag{39}$$

Notice that, taking into account the calculation in Eq. 33, the product $\Gamma^T P \bar{y}(t)$ is equal to $\bar{y}(t)$, and hence cancels the last term of Eq. 22. The minimization of the cost function results in the continuous-time model-predictive control law

$$\frac{\partial V}{\partial U} = -2H^T \left[\int_0^T \Gamma \Gamma^T r(w - \bar{y}_p) dt - \int_0^T (\Gamma \Gamma^T HU + \Gamma \Gamma^T Qx) dt \right] = 0. \tag{40}$$

Let us define the matrix $\bar{\Gamma} \in \mathbb{R}^{n_y+1} \times \mathbb{R}^{n_y+1}$ as

$$\bar{\Gamma} = \int_0^T \Gamma \Gamma^T dt. \tag{41}$$

Given that the general term of the matrix $\Gamma \Gamma^T$ is $T^{i-1+j-1}/((i-1)!(j-1)!)$, Eq. 41 can be rewritten as

$$\bar{\Gamma} = \begin{bmatrix} \gamma_{(1,1)} & \cdots & \gamma_{(1,n_y+1)} \\ \vdots & \ddots & \vdots \\ \gamma_{(n_y+1,1)} & \cdots & \gamma_{(n_y+1,n_y+1)} \end{bmatrix}, \tag{42}$$

where

$$\gamma_{(i,j)} = \frac{1}{(i+j-1)(i-1)!(j-1)!} T^{i+j-1} \tag{43}$$

for every $i, j = 1, \dots, n_y + 1$. Equation 40 is then reformulated as

$$\frac{\partial V}{\partial U} = -2H^T \bar{\Gamma} [r(w - \bar{y}_p) - HU - Qx] = 0 \tag{44}$$

and, using the substitution from Eq. 21, the control vector becomes

$$U = (H^T \bar{\Gamma} H)^{-1} H^T \bar{\Gamma} [r(w - y_p + y - \bar{y}) - Qx]. \tag{45}$$

At this point we return to the separate notation for the CPWL and FS cases and employ the notation in Eqs. 35, 36, 37 and 38. Let us define the first row of the matrix $(H(v)^T \bar{\Gamma} H(v))^{-1} H(v)^T \bar{\Gamma} \in \mathbb{R}^{n_u+1} \times \mathbb{R}^{n_y+1}$ in the CPWL case as $\kappa_p(v)$ and analogously in the FS case as $\kappa_f(v)$. Now the control law of the nonlinear Wiener-type model-predictive control is given by

$$u(t) = \kappa_p(v)r(w - y_p + y_{mp} - \bar{y}_{mp}) - \kappa_p(v)Q_p(v)x \tag{46}$$

for the CPWL approximation case and

$$u(t) = \kappa_f(v)r(w - y_p + y_{mf} - \bar{y}_{mf}) - \kappa_f(v)Q_f(v)x \tag{47}$$

for the FS case.

4 Simulation Example

The proposed method was tested on a third-order linear time-delayed system

$$G_p(s) = \frac{v(s)}{u(s)} = \frac{1.2}{(s + 0.2)(s + 2)(s + 3)} \cdot e^{-5s} \tag{48}$$

with a static, nonlinear output mapping

$$y_p(t) = 4.2 \cdot (\arctan [10(v(t) - 0.5)] + \pi/2) . \tag{49}$$

The control-design parameters were chosen as follows: $n_u = 1$, $n_y = 4$, $T = 0.5$ s, $A_r = -1/3$, $B_r = 1$, and $C_r = 1/3$. The choice of n_y and T will be discussed later. The process input was assumed to be bounded by the interval $0 \leq u(t) \leq 1$, hence the intermediate-variable range was $[0, 1]$. Even though the input bounds are not considered in the control-law derivation, bounding the controller output implicitly bounds the model-output prediction and the parameters for the control-action calculation. For the CPWL the intermediate-variable range was further divided into five non-equidistantly spread segments. The segment-parameter positions α_i , $i = 1, 2, \dots, 4$ were calculated using a c-means clustering method [1]. The initial parameter α_0 was set to 0 because otherwise the derivative of \hat{h}_{mp} in the interval $[0 \alpha_1]$ would be 0. However, the parameter set is in this way augmented by 1. The optimization of the

Figure 1 Comparison of the CPWL and FS approximations of the output mapping using four clusters.

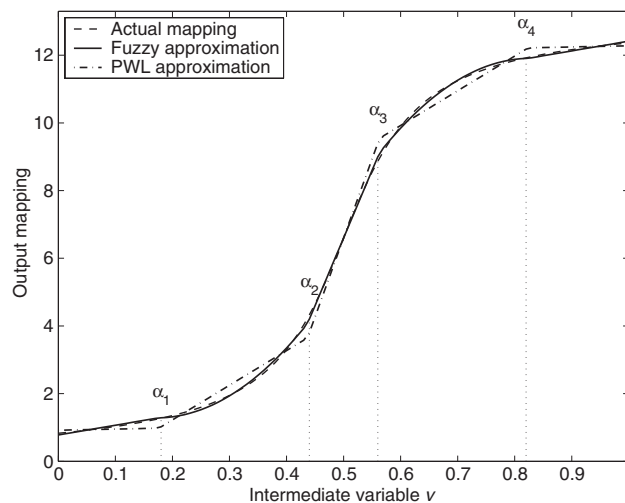
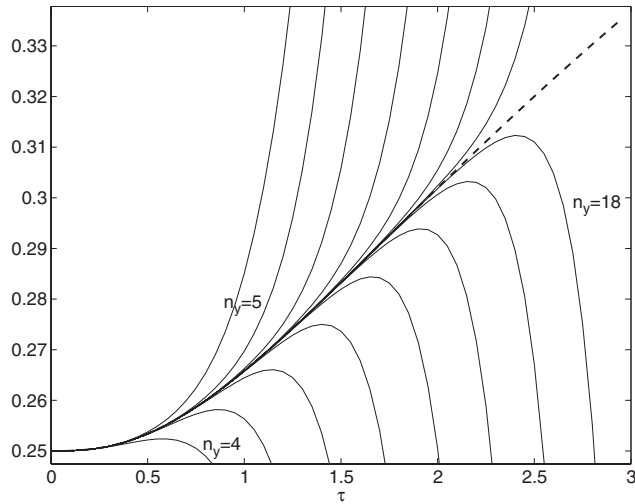


Figure 2 Linear-model output prediction in terms of n_y .



parameters Θ was carried out using linear programming. The resulting parameter vectors are given by

$$\alpha = [0, 0.176, 0.437, 0.563, 0.824],$$

$$\Theta = [0.919, 0.339, 9.895, 36.441, -36.420, -9.923].$$

In the FS case we assumed four triangular membership functions and linear output functions, i.e., $x_c = [1, v(t)]^T$ and $\Theta_f \in \mathbb{R}^4 \times \mathbb{R}^2$. The membership function apexes were at the same positions as in the CPWL case, i.e., $\alpha_i, i = 1, 2, \dots, 4$. Therefore,

Figure 3 Comparison of the CPWL and FS nonlinear-model predictions in terms of n_y .

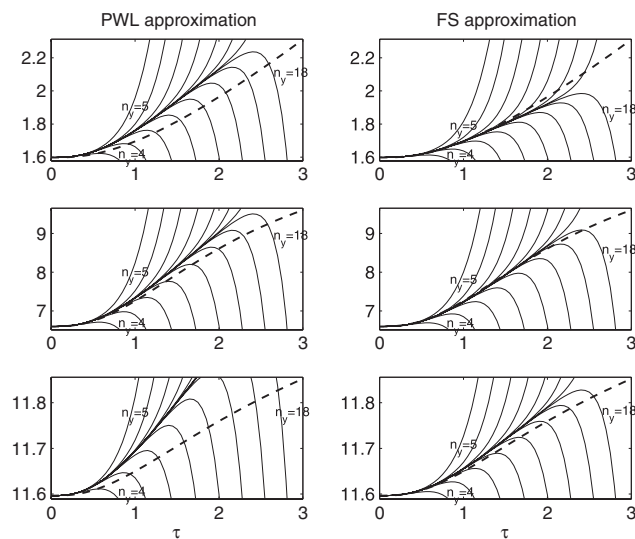
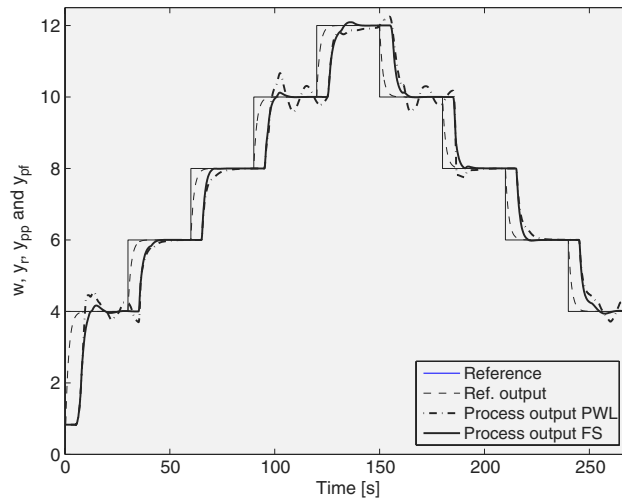


Figure 4 Comparison of the closed-loop-experiment results for both approximation cases.



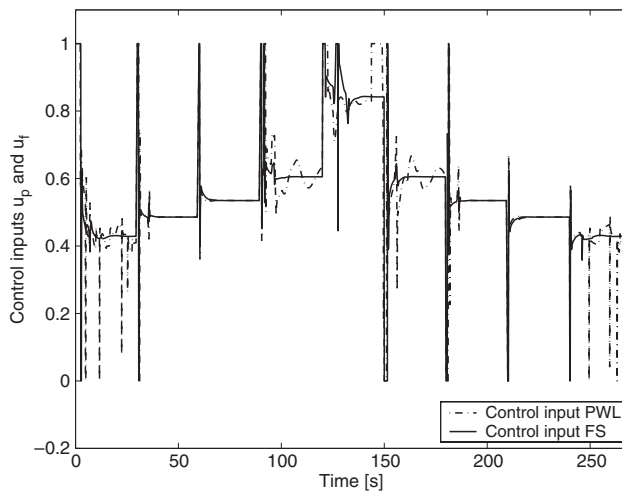
after completing the linear-programming optimization, the fuzzy-parameter matrix yielded

$$\Theta_f^T = \begin{bmatrix} 0.778 & -1.715 & 1.645 & 9.537 \\ 2.877 & 13.241 & 13.276 & 2.880 \end{bmatrix}.$$

The resulting approximations $\hat{h}_{mp}(v)$ and $\hat{h}_{mf}(v)$ are compared to the actual mapping $\hat{h}(v)$ in Figure 1.

One way to compare the effects due to the approximation choice is to investigate the behaviour of the open-loop model predictions in both cases. Figure 2 shows the linear-model output prediction dependent on n_y . It is clear that increasing the value of n_y improves the prediction accuracy, and hence the prediction horizon T can easily be increased. However, in the nonlinear-model case some of the prediction error can be attributed to the approximation accuracy. Figure 3 shows the benefit of

Figure 5 Control signals in the experiment with our proposed method.



choosing the FS approximation rather than the CPWL approximation. In all three regions the FS prediction gives better results, even though in the FS approximation procedure the higher derivatives of v were left out. Nevertheless, in terms of the horizon choice there is little or no change from the linear-model case – the choice of n_y is still dictated by the choice of T . In our case if we wanted to make a fair comparison, the value of T had to be chosen so that it was low enough for the model-output predictions to be fairly accurate. When choosing $T = 0.5$, a fourth-order output prediction ($n_y = 4$) was sufficient.

A closed-loop experiment with a series of step signals as the reference signal was conducted. The results, presented in Figure 4, imply that the process output successfully follows the reference-model output for the whole operating region in both approximation cases, and, at the same time, does not suffer from the process time delay. However, it is clear that in the FS case the overall performance is much better. In the CPWL case the reference tracking depends on the operating point much more than in the opposite case, and in some regions the system output is subject to oscillations. The corresponding input variables are shown in Figure 5. In the CPWL case the input is much more oscillatory. Also notice the impulse-like behaviour of the control signals in the time after the step reference changes – this is because a third-order system was made to closely follow a first-order reference model.

5 Conclusion

A novel method of continuous-time model-predictive control for nonlinear time-delayed systems was presented. In the derivation procedure the method implicitly incorporates two different static nonlinear-mapping approximations – using continuous piece-wise linear functions and a fuzzy system – and the solution for tackling system time delays. It was shown for the case of a third-order system with an arc-tangent output mapping and a considerable pure time delay that the proposed closed-loop system in both cases exhibits quality control; however, the performance in the FS case is clearly better due to a better model-output prediction. The proposed method is thus very appropriate for chemical processes that can be described by a time-delayed Wiener-type system and where high-quality control is desired.

References

1. Babuška, R., Verbrüggen, H.B.: An overview of fuzzy modeling for control. *Control Eng. Pract.* **4**(11), 1593–1606 (1996)
2. Bequette, B.W.: Nonlinear control of chemical processes: a review. *Ind. Eng. Chem. Res.* **30**, 1391–1413 (1991)
3. Biagiola, S.I., Agamennoni, O.E., Figueroa, J.L.: H_∞ control of a Wiener-type system. *Int. J. Control* **77**(6), 572–583 (2004)
4. Chen, W., Ballance, D.J., Gawthrop, P.J.: Optimal control of nonlinear systems: a predictive control approach. *Automatica* **39**, 633–641 (2003)
5. Clarke, D.W., Mohtadi, C., Tuffs, P.S.: Generalized predictive control – Parts 1, Part 2. *Automatica* **24**, 137–160 (1987)
6. Demircioğlu, H., Gawthrop, P.J.: Continuous-time generalized predictive control. *Automatica* **27**(1), 55–74 (1991)

7. Demircioğlu, H., Gawthrop, P.J.: Multivariable continuous-time generalized predictive control (MCGPC). *Automatica* **28**(4), 697–713 (1992)
8. García, C.E., Prett, D.M., Morari, M.: Model predictive control: theory and practice – a survey. *Automatica* **25**(3), 335–348 (1989)
9. Gerškšič, S., Juričič, D., Strmčnik, S., Matko, D.: Wiener model based nonlinear predictive control. *Int. J. Syst. Sci.* **31**(2), 189–202 (2000)
10. Henson, M.A.: Nonlinear model predictive control: current status and future directions. *Comput. Chem. Eng.* **23**, 187–202 (1998)
11. Julián, P., Jordán, M., Desages, A.: Canonical piecewise linear approximation of smooth functions. *IEEE Trans. Circuits Syst.* **45**, 567–571 (1998)
12. Lussón Cervantes, A., Agamennoni, O.E., Figueroa, J.L.: A nonlinear model predictive control system based on Wiener piecewise linear models. *J. Process Control* **13**, 655–666 (2003)
13. Magni, L., Scattolini, R.: Stabilizing model predictive control of nonlinear continuous time systems. *Annu. Rev. Control* **28**, 1–11 (2004)
14. Mayne, D.Q., Michalska, H.: Receding horizon control of nonlinear systems. *IEEE Trans. Automat. Contr.* **35**(7), 814–824 (1990)
15. Morari, M., Lee, J.H.: Model predictive control: past, present and future. *Comput. Chem. Eng.* **23**, 667–682 (1999)
16. Norquay, S.J., Palazoğlu, A., Romagnoli, J.A.: Application of Wiener model predictive control (WMPC) to a pH neutralization experiment. *IEEE Trans. Control Syst. Technol.* **7**(4), 437–445 (1999)
17. Norquay, S.J., Palazoğlu, A., Romagnoli, J.A.: Application of Wiener model predictive control (WMPC) to an industrial C2-splitter. *J. Process Control* **9**, 461–473 (1999)
18. Pal, N.R., Bezdek, J.C.: On cluster validity for the fuzzy c-means model. *IEEE Trans. Fuzzy Syst.* **3**(3), 370–379 (1995)
19. Sentoni, G., Agamennoni, O., Desages, A., Romagnoli, J.: Approximate models for nonlinear process control. *AIChE J.* **42**, 2240–2250 (1996)
20. Smith, O.J.: Controller to overcome dead time. *Instrum. Soc. Am. J.* **6**(2), 28–33 (1959)
21. Sung, S.W., Lee, J.: Modeling and control of Wiener-type processes. *Chem. Eng. Sci.* **59**, 1515–1521 (2004)
22. Škrjanc, I., Matko, D.: Fuzzy predictive functional control in the state space domain. *J. Intell. Robot. Syst.* **31**, 283–297 (2001)
23. Škrjanc, I., Blažič, S., Agamennoni, O.: Interval fuzzy modelling applied to Wiener models with uncertainties. *IEEE Trans. Syst. Man Cybern., Part B, Cybern.* **35**(5), 1092–1095 (2005)
24. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its applications to modelling and control. *IEEE Trans. Syst. Man. Cybernetics* **15**, 116–132 (1985)

Literatura

- [1] J. Abonyi, L. Nagy, F. Szeifert, Fuzzy model-based predictive control by instantaneous linearization, *Fuzzy Sets and Systems*, Vol. 120, str. 109–122, 2001.
- [2] J. Ackermann, *Robust control*, Berlin: Springer, 1993.
- [3] O. E. Agamennoni, I. Škrjanc, M. Lepetić, H. Chiacchiarini, D. Matko, Nonlinear uncertainty model of a magnetic suspension system, *Mathematical and Computer Modelling*, Vol. 40, str. 1075–1087, 2004.
- [4] E. Alcorta García, P. M. Frank, Deterministic nonlinear observer-based approaches to fault diagnosis: A survey, *Control Engineering Practice*, Vol. 5, No. 5, str. 663–670, 1997.
- [5] R. Babuška, J. M. Sousa, H. B. Verbruggen, Predictive control of nonlinear systems based on fuzzy and neural models, *Proceedings of European Control Conference, Karlsruhe, Germany*, 1999.
- [6] Robert Babuška, *Fuzzy Modeling for Control*, Kluwer Academic Publishers, Boston, USA, 1998.
- [7] M. Basseville, Detecting changes in signals and systems - a survey, *Automatica*, Vol. 24, str. 309–326, 1988.
- [8] M. Basseville, T. V. Nikiforov, *Detection of abrupt changes - theory and application*, Prentice Hall. Information and system sciences series, 1993.
- [9] G. Bastin, M. R. Gevers, Stable adaptive observers for nonlinear time-varying systems, *IEEE Trans. Contr.*, Vol. 33, No. 7, str. 650–658, 1988.
- [10] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, Athena Scientific, Belmont, Massachusetts, 1995.
- [11] S. I. Biagiola, O. E. Agamennoni, J. L. Figueroa, H_∞ control of a wiener-type system, *International Journal of Control*, Vol. 77, No. 6, str. 572–583, 2004.

-
- [12] S. A. Billings, S. Y. Fakhouri, Identification of systems containing linear dynamic and static nonlinear elements, *Automatica*, Vol. 18, str. 15–26, 1982.
- [13] I. Bratko, I. Mozetič, N. Lavrač, *KARDIO: A Study in Deep and Qualitative Knowledge for Expert Systems*, MIT Press, Cambridge, MA, 1989.
- [14] E. F. Camacho, C. Bordons, *Model Predictive Control in the Process industry*, Springer, Berlin, 1995.
- [15] C. V. Camp, Design of space trusses using big bang–big crunch optimization, *Journal of Structural Engineering*, Vol. 133, No. 7, str. 999–1008, 2007.
- [16] Wen-Hua Chen, Donald J. Ballance, Peter J. Gawthrop, Optimal control of nonlinear systems: a predictive control approach, *Automatica*, Vol. 39, str. 633–641, 2003.
- [17] D. W. Clarke, C. Mohtadi, P. S. Tuffs, Generalized predictive control - part 1, part 2, *Automatica*, Vol. 24, str. 137–160, 1987.
- [18] C. R. Cutler, B. L. Ramaker, Dynamic matrix control - a computer control algorithm, *Proceedings of the ACC, San Francisco*, 1980.
- [19] J. M. da Costa Sousa, U. Kaymak, Model predictive control using fuzzy decision functions, *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, Vol. 31, No. 1, str. 54–65, 2001.
- [20] S. X. Ding, T. Jeinsch, P. M. Frank, E. L. Ding, A unified approach to the optimization of fault detection systems, *International Journal of Adaptive Control and Signal Processing*, Vol. 14, str. 725–745, 2000.
- [21] X. Ding, P. M. Frank, Fault detection via factorization approach, *Syst. Contr. Lett.*, Vol. 14, str. 431–436, 1990.
- [22] X. Ding, P. M. Frank, On-line fault detection in uncertain systems using adaptive observers, *European J. of Diagnosis and Safety in Automation*, Vol. 3, str. 9–21, 1993.
- [23] X. Ding, L. Guo, P. M. Frank, An approach to residual generator and evaluator design and synthesis, *Proceedings of 12th IFAC World Congress*, Sydney, Australia, 1993.
- [24] X. Ding, L. Guo, P. M. Frank, A frequency domain approach to fault detection of uncertain dynamic systems, *Proceedings of 32nd CDC*, str. 1722–1727, 1993.

-
- [25] Xianchun Ding, Limin Guo, Paul M. Frank, Parameterization of linear observers and its application to observer design, *IEEE Transactions on Automatic Control*, Vol. 39, No. 8, str. 1648–1652, 1994.
- [26] D. Dong, T. J. McAvoy, Batch tracking via nonlinear principal component analysis, *AIChE Journal*, Vol. 42, No. 8, str. 2199–2208, 1996.
- [27] R. Dunia, S. J. Qin, T. F. Edgar, T. J. McAvoy, Identification of faulty sensors using principal component analysis, *AIChE Journal*, Vol. 42, No. 10, str. 2797–2812, 1996.
- [28] O. K. Erol, I. Eksin, A new optimization method: Big bang–big crunch, *Advances in Engineering Software*, Vol. 37, str. 106–111, 2006.
- [29] I. Fagarasan, S. Ploix, S. Gentil, Causal fault detection and isolation based on a set-membership approach, *Automatica*, Vol. 40, str. 2099–2110, 2004.
- [30] P. M. Frank, *Systems fault diagnostics, reliability and related knowledge-based approaches*, Volume 1, poglavje Fault diagnosis in dynamic systems via state estimation – a survey, str. 35–98, Dordrecht: D Reidel, 1986.
- [31] P. M. Frank, Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy - a survey and some new results, *Automatica*, Vol. 26, str. 459–474, 1990.
- [32] P. M. Frank, S. X. Ding, T. Marcu, Model-based fault diagnosis in technical processes, *Transactions of the Institute of Measurement and Control*, Vol. 22, No. 1, str. 55–101, 2000.
- [33] P. M. Frank, X. Ding, Survey of robust residual generation and evaluation methods in observer-based fault detection systems, *J. Proc. Control*, Vol. 7, No. 6, str. 403–424, 1997.
- [34] P. M. Frank, J. Wünnenberg, *Fault Diagnosis in Dynamic Systems*, poglavje Robust fault diagnosis using unknown input observer schemes, str. 46–98, Prentice Hall, 1989.
- [35] Paul M. Frank, Birgit Köppen-Seliger, New developments using ai in fault diagnosis, *Engineering Applications of Artificial Intelligence*, Vol. 10, No. 1, str. 3–14, 1997.

-
- [36] E. Frisk, L. Nielsen, Robust residual generation for diagnosis including a reference model for residual behavior, *Proceedings of the 14th IFAC world congress*, Volume P, str. 55–60, Beijing, China, 1999.
- [37] D. Füssel, D. Ballé, R. Isermann, Closed loop fault diagnosis based on a non-linear process model and automatic fuzzy rule generation, *SAFEPROCESS'97, Kingston upon Hull*, Volume 1, str. 359–364, 1997.
- [38] J. B. Fussell, Fault tree analysis - state of the art, *IEEE Transactions on Reliability*, Vol. 23, No. 1, str. 51–53, 1974.
- [39] Omar Galán, José A. Romagnoli, Ahmet Palazoglu, Robust h_∞ control of non-linear plants based on multi-linear models: an application to a bench-scale ph neutralization reactor, *Chemical Engineering Science*, Vol. 55, str. 4435–4450, 2000.
- [40] C. E. García, D. M. Prett, M. Morari, Model predictive control: theory and practice - a survey, *Automatica*, Vol. 25, No. 3, , 1989.
- [41] Samo Gerškšič, Djani Juričić, Stanko Strmčnik, Drago Matko, Wiener model based nonlinear predictive control, *International Journal of System Science*, Vol. 31, No. 2, str. 189–202, 2000.
- [42] J. Gertler, *Fault detection and diagnosis in engineering systems*, New York: Marcel Dekker, 1998.
- [43] J. Gertler, T. J. McAvoy, Principal component analysis and parity relations - a strong duality, *IFAC SAFEPROCESS'97*, 1997.
- [44] J. Gertler, D. Singer, A new structural framework for parity equation-based failure detection and isolation, *Automatica*, Vol. 26, str. 381–388, 1990.
- [45] G. C. Goodwin, M. Gevers, B. Ninness, Quantifying the error in estimated transfer functions with application to model order selection, *IEEE Transactions on Automatic Control*, Vol. 37, No. 7, str. 913–929, 1992.
- [46] B.P. Graham, R.B. Newell, Fuzzy identification and control of a liquid level rig, *Fuzzy Sets and Systems*, Vol. 26, str. 255–273, 1988.
- [47] F. Hamelin, D. Sauter, Robust fault detection in uncertain dynamic systems, *Automatica*, Vol. 36, str. 1747–1754, 2000.

-
- [48] D. Hengy, P.M. Frank, Component failure detection via nonlinear observer, *Proceedings of the IFAC Workshop on Fault Detection and Safe in Chemical Plants*, str. 153–157, 1986.
- [49] Michael A. Henson, Nonlinear model predictive control: current status and future directions, *Computers and Chemical Engineering*, Vol. 23, str. 187–202, 1998.
- [50] D. M. Himmelblau, R. W. Braker, W. Suewatanakul, Fault classification with the aid of artificial neural networks, *Safeprocess '91*, str. 369–373, Baden-Baden, 1991.
- [51] J. C. Hoskins, K. M. Kaliyur, D. M. Himmelblau, Fault diagnosis in complex chemical plants using artificial neural networks, *AIChE Journal*, Vol. 37, str. 137–142, 1991.
- [52] Y. L. Huang, H. H. Lou, J. P. Gong, T. F. Edgar, Fuzzy model predictive control, *IEEE Transactions on Fuzzy Systems*, Vol. 8, No. 6, str. 665–678, 2000.
- [53] M. Iri, K. Aoki, E. O'Shima, H. Matsuyama, An algorithm for diagnosis of system failures in the chemical process, *Computers and Chemical Engineering*, Vol. 3, No. 1-4, str. 489–493, 1979.
- [54] R. Isermann, Process fault detection based on modelling and estimation methods - a survey, *Automatica*, Vol. 20, No. 4, str. 387–404, 1984.
- [55] R. Isermann, Integration of fault detection and diagnosis methods, *SAFEPROCESS'94, Espoo*, Volume 2, str. 597–612, 1994.
- [56] R. Isermann, Supervision, fault-detection and fault-diagnosis methods - an introduction, *Control Engineering Practice*, Vol. 5, No. 5, str. 639–652, 1997.
- [57] R. Isermann, P. Ballé, Trends in the application of model-based fault detection and diagnosis of technical processes, *Control Engineering Practice*, Vol. 5, No. 5, str. 707–719, 1997.
- [58] R. Isermann, K. H. Lachmann, D. Matko, *Adaptive Control Systems*, Prentice Hall, New York, 1992.
- [59] J. E. Jackson, *A User's Guide To Principal Components*, John Wiley & Sons, New York, 1991.

-
- [60] Pedro Julián, Mario Jordán, Alfredo Desages, Canonical piecewise linear approximation of smooth functions, *IEEE Transactions on Circuits and Systems*, Vol. 45, str. 567–571, 1998.
- [61] D. Juričić, M. Žele, Robust detection of sensor faults by means of a statistical test, *Automatica*, Vol. 38, str. 737–742, 2002.
- [62] D. Juričić, A. Žnidaršič, D. Füssel, A methodology for robust fault detection in dynamic systems, *Engineering Applications of Artificial Intelligence*, Vol. 10, No. 1, str. 15–29, 1997.
- [63] A. Kalafatis, N. Arifin, L. Wang, W. R. Cluett, A new approach to the identification of ph processes based on the wiener model, *Chemical Engineering Science*, Vol. 50, str. 3693–3701, 1995.
- [64] Gorazd Karer, Gašper Mušič, Igor Škrjanc, Borut Zupančič, Hybrid fuzzy model-based predictive control of temperature in a batch reactor, *Computers and Chemical Engineering*, Vol. 31, str. 1552–1564, 2007.
- [65] K. Kavšek-Biasizzo, *Prediktivno vodenje nelinearnih sistemov*, Doktorska disertacija, Univerza v Ljubljani, Fakulteta za elektrotehniko, 1998.
- [66] K. Kavšek-Biasizzo, I. Škrjanc and D. Matko, Fuzzy predictive control of highly nonlinear ph process, *Computers Chem. Engng*, Vol. 21, str. 613–618, 1997.
- [67] U. Kaymak, J.M. Sousa, H.B. Verbruggen, A comparative study of fuzzy and conventional criteria in model-based predictive control, *FUZZ-IEEE, Barcelona, Spain*, Volume 2, str. 907–914, 1997.
- [68] Gregor Klančar, Djani Juričić, Rihard Karba, Robust fault detection based on compensation of the modelling error, *International Journal of Systems Science*, Vol. 33, No. 2, str. 97–105, 2002.
- [69] Gregor Klančar, Igor Škrjanc, Metoda glavnih komponent pri odkrivanju in izolaciji napak: primer hidravličnega procesa in procesa fermentacije, *Elektrotehniški vestnik*, Vol. 69, No. 5, str. 311–316, 2002.
- [70] M. Kokawa, M. Satoshi, S. Shigai, Fault location using digraph and inverse direction search with application, *Automatica*, Vol. 19, No. 6, str. 729–735, 1983.
- [71] M. A. Kramer, B. L. Palowitch, A rule based approach to fault diagnosis using the signed directed graph, *AIChE Journal*, Vol. 33, No. 7, str. 1067–1078, 1987.

-
- [72] T. Kumbasar, E. Yeşil, I. Eksin, M. Güzelkaya, Inverse fuzzy model control with online adaptation via big bang-big crunch optimization, *ISCCSP '08, Malta*, str. 697–702, 2008.
- [73] O.-K. Kwon, G. C. Goodwin, W. H. Kwon, Robust fault detection method accounting for modelling errors in uncertain systems, *Control Engineering Practice*, Vol. 2, No. 5, str. 763–771, 1994.
- [74] W. Li, H. Yue, S. Valle-Cervantes, S. Qin, Recursive pca for adaptive process monitoring, *Journal of Process Control*, Vol. 10, No. 5, str. 471–486, 2000.
- [75] D. A. Linkens, S. Kandiah, Long-range predictive control using fuzzy process models, *Chemical Engineering Research & Design*, Vol. 74, str. 77–88, 1996.
- [76] Lennart Ljung, *System Identification*, Prentice-Hall, 1987.
- [77] X. C. Lou, A. S. Willsky, G. C. Verghese, Optimally robust redundancy relations for failure detection in uncertain systems, *Automatica*, Vol. 22, No. 3, str. 333–344, 1986.
- [78] Y. Z. Lu, M. He, C. W. Xu, Fuzzy modeling and expert optimization control for industrial processes, *IEEE Trans. Control Systems Technol.*, Vol. 5, str. 2–12, 1997.
- [79] Ania Lussón Cervantes, Osvaldo Enrique Agamennoni, José Luis Figueroa, A nonlinear model predictive control system based on wiener piecewise linear models, *Journal of Process Control*, Vol. 13, str. 655–666, 2003.
- [80] J. M. Maciejowski, *Predictive Control with Constraints*, Prentice Hall, Pearson Education Limited, 2002.
- [81] M. A. Massoumnia, A geometric approach to the synthesis of failure detection filters, *IEEE Transactions on Automatic Control*, Vol. 9, str. 839–846, 1986.
- [82] D. Q. Mayne, H. Michalska, Receding horizon control of nonlinear systems, *IEEE Transactions on Automatic Control*, Vol. 35, No. 7, str. 814–824, 1990.
- [83] Manfred Morari, Jay H. Lee, Model predictive control: past, present and future, *Computers and Chemical Engineering*, Vol. 23, str. 667–682, 1999.
- [84] O. Nelles, M. Fischer, Local linear model trees (lolimot) for nonlinear system identification of a cooling blast, *Preprints of the European Congresss on Intelligent Techniques and Soft Computing, Aachen*, Volume 2, str. 1187–1191, 1996.

-
- [85] P. Nomikos, J. MacGregor, Monitoring batch processes using multiway principal component analysis, *AIChE Journal*, Vol. 40, No. 8, str. 1361–1375, 1994.
- [86] S. J. Norquay, A. Palazoğlu, J. A. Romagnoli, Model predictive control based on wiener models, *Chemical Engineering Science*, Vol. 53, No. 1, str. 75–84, 1998.
- [87] S. J. Norquay, A. Palazoğlu, J. A. Romagnoli, Application of wiener model predictive control (wmpc) to a ph neutralization experiment, *IEEE Transactions on Control Systems Technology*, Vol. 7, No. 4, str. 437–445, 1999.
- [88] Simon Oblak, Igor Škrjanc, A comparison of fuzzy and cpwl approximations in the continuous-time nonlinear model-predictive control of time-delayed wiener-type systems, *Journal of intelligent and robotic systems*, Vol. 47, str. 125–137, 2006.
- [89] P. M. Olin, G. Rizzoni, Residual generation for fault detection and isolation in the joint time-frequency domain, *Proceedings of Int. Conference on Fault Diagnosis*, Toulouse, France, 1993.
- [90] O. O. Oyeleye, M. A. Kramer, Qualitative simulation of chemical process systems: steady state analysis, *AIChE Journal*, Vol. 34, No. 9, str. 1441–1454, 1988.
- [91] S. Peršin, *Nadgradnja nadzornih sistemov industrijskih procesov s podsistemi za sprotno odkrivanje napak*, Doktorska disertacija, Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko, 2004.
- [92] M. M. Polycarpou, A. B. Trunov, Learning approach to nonlinear fault diagnosis: Detectability analysis, *IEEE Trans. Automat. Contr.*, Vol. 45, str. 806–812, 2000.
- [93] M. M. Polycarpou, A. T. Vemuri, Learning methodology for failure detection and accommodation, *IEEE Control Systems Intelligence and Learning*, Vol. 15, No. 3, , 1995.
- [94] B. E. Postlethwaite, Building a model-based fuzzy controller, *Fuzzy Sets and Systems*, Vol. 79, str. 3–13, 1996.
- [95] T. Proll, M. N. Karim, Model-predictive ph control using real-time narx approach, *AIChE Journal*, Vol. 40, str. 269–282, 1994.
- [96] V. Puig, J. Quevedo, T. Escobet, S. de las Heras, Passive robust fault detection approaches using interval models, *Proceedings of the 15th IFAC World Congress*, 2002.

-
- [97] V. Puig, J. Quevedo, S. Tornil, Robust fault detection: Active versus passive approaches, *Proceedings of the IFAC SAFEPROCESS, Hungary*, 2000.
- [98] V. Puig, A. Stancu, T. Escobet, F. Nejjari, J. Quevedo, R.J. Patton, Passive robust fault detection using interval observers: Application to the damadics benchmark problem, *Control Engineering Practice*, Vol. 14, str. 621–633, 2006.
- [99] S. J. Qin, T. A. Badgwell, *Chemical process control - V, Assessment and New Directions for Research*, poglavje An Overview of Industrial Predictive Control Technology, Tahoe City, CA, 1996.
- [100] Z. Qiu, J. Gertler, Robust fdi systems and h_∞ -optimization, *Proceedings of the 32nd CDC*, San Antonio, USA, 1993.
- [101] A. Rakar, *Odkrivanje napak v tehničnih sistemih z metodami aproksimativnega sklepanja*, Doktorska disertacija, Univerza v Ljubljani, Fakulteta za elektrotehniko, 2000.
- [102] A. Rakar, D. Juričić, Primerjava metod aproksimativnega sklepanja pri izolaciji napak - simulacijska študija, *Elektrotehniški vestnik*, Vol. 73, No. 1, str. 47–52, 2002.
- [103] A. Rakar, D. Juričić, P. Ballé, Transferable belief model in fault diagnosis, *Engineering Applications of Artificial Intelligence*, Vol. 12, No. 5, str. 555–567, 1999.
- [104] J. Richalet, A. Rault, J. L. Testud, J. Papon, Model predictive heuristic control: Applications to industrial processes, *Automatica*, Vol. 14, str. 413–428, 1978.
- [105] Jacques Richalet, Industrial application of model-based predictive control, *Automatica*, Vol. 29, No. 5, str. 1251–1274, 1993.
- [106] J.A. Roubos, S. Mollov, R. Babuška, H.B. Verbruggen, Fuzzy model-based predictive control using takagi-sugeno models, *International Journal of Approximate Reasoning*, Vol. 22, str. 3–30, 1999.
- [107] M. A. Sainz, J. Armengol, J. Vehí, Fault detection and isolation of the three-tank system using the modal interval analysis, *Journal of Process Control*, Vol. 12, str. 325–338, 2002.
- [108] H. Sarimveis, G. Bafas, Fuzzy model predictive control of non-linear processes using genetic algorithms, *Fuzzy Sets and Systems*, Vol. 139, str. 59–80, 2003.

-
- [109] G. Schreier, J. Ragot, R.J. Patton, P.M. Frank, Observer design for a class of nonlinear systems, *Proceedings of the IFAC Symposium SAFEPROCESS '97*, str. 483–488, 1997.
- [110] R. Seliger, P. M. Frank, Fault diagnosis by disturbance decoupled nonlinear observers, *Proceedings of the 30th IEEE CDC*, Volume 3, 1991.
- [111] R. Shih, L. Lee, Use of fuzzy cause-effect digraph for resolution fault diagnosis for process plants i. fuzzy cause-effect digraph, *Industrial and Engineering Chemistry Research*, Vol. 34, No. 5, str. 1688–1702, 1995.
- [112] R. Shih, L. Lee, Use of fuzzy cause-effect digraph for resolution fault diagnosis for process plants ii. diagnostic algorithm and applications, *Industrial and Engineering Chemistry Research*, Vol. 34, No. 5, str. 1703–1717, 1995.
- [113] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.-Y. Glorennec, H. Hjalmarsson, A. Juditsky, Nonlinear black-box modeling in system identification: A unified overview, *Automatica*, Vol. 31, str. 1691–1724, 1995.
- [114] P. Smets, R. Kennes, The transferable belief model, *Artificial Intelligence*, Vol. 66, str. 191–234, 1999.
- [115] T. Sorsa, H. N. Koivo, Application of artificial neural networks in process fault diagnosis, *Safeprocess '91*, Baden-Baden, 1991.
- [116] Su Whan Sung, Jitae Lee, Modeling and control of wiener-type processes, *Chemical Engineering Science*, Vol. 59, str. 1515–1521, 2004.
- [117] T. Takagi, M. Sugeno, Fuzzy identification of systems and its applications to modelling and control, *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. 15, str. 116–132, 1985.
- [118] P. Tatjewski, M. Ławryńczuk, Soft computing in model-based predictive control, *Int. J. Appl. Math. Comput. Sci.*, Vol. 16, No. 1, str. 7–26, 2006.
- [119] A. B. Trunov, M. M. Polycarpou, Automated fault diagnosis in nonlinear multivariable systems using a learning methodology, *IEEE Trans. Neural Networks*, Vol. 11, str. 91–101, 2000.
- [120] T. Umeda, T. Kuriyama, E. O'Shima, H. Matsuyama, A graphical approach to cause and effect analysis of chemical processing systems, *Chemical Engineering Science*, Vol. 35, No. 12, str. 2379–2388, 1980.

-
- [121] M. A. Vázquez, O. E. Agamennoni, Approximate models for nonlinear dynamical systems and its generalization properties, *Mathematical and Computer Modelling*, Vol. 33, str. 965–986, 2001.
- [122] H. Vedam, V. Venkatasubramanian, Signed digraph based multiple fault diagnosis, *Computers and Chemical Engineering*, Vol. 21, str. S655–S660, 1997.
- [123] A. T. Vemuri, M. M. Polycarpou, Robust nonlinear fault diagnosis in input–output systems, *Int. J. Control*, Vol. 68, str. 343–360, 1997.
- [124] V. Venkatasubramanian, R. Rengaswamy, S. Kavuri, A review of process fault detection and diagnosis part ii: Qualitative models and search strategies, *Computers and Chemical Engineering*, Vol. 27, str. 313–326, 2003.
- [125] V. Venkatasubramanian, R. Rengaswamy, S. Kavuri, K. Yin, A review of process fault detection and diagnosis part iii: Process history based methods, *Computers and Chemical Engineering*, Vol. 27, str. 327–346, 2003.
- [126] Igor Škrjanc, Sašo Blažič, Predictive functional control based on fuzzy model: design and stability study, *Journal of Intelligent and Robotic Systems*, Vol. 43, str. 283–299, 2005.
- [127] Igor Škrjanc, Sašo Blažič, Osvaldo Agamennoni, Identification of dynamical systems with a robust interval model, *Automatica*, Vol. 41, No. 2, str. 327–332, 2005.
- [128] Igor Škrjanc, Sašo Blažič, Osvaldo Agamennoni, Interval fuzzy modelling applied to wiener models with uncertainties, *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, Vol. 35, No. 5, str. 1092–1095, 2005.
- [129] Igor Škrjanc, Sašo Blažič, Simon Oblak, Jacques Richalet, An approach to predictive control of multivariable time-delayed plant: Stability and design issues, *ISA Transactions*, Vol. 43, str. 585–595, 2004.
- [130] Igor Škrjanc, Drago Matko, Predictive functional control based on fuzzy model for heat-exchanger pilot plant, *IEEE Transactions on Fuzzy Systems*, Vol. 8, No. 6, str. 705–712, 2000.
- [131] Igor Škrjanc, Drago Matko, Fuzzy predictive functional control in the state space domain, *Journal of Intelligent and Robotic Systems*, Vol. 31, str. 283–297, 2001.

-
- [132] M. Žele, D. Vrečko, D. Juričić, Zaznavanje nepravilnega delovanja senzorjev v čistilni napravi odpadnih voda z uporabo adaptivne metode glavnih komponent, *Elektrotehniški vestnik*, Vol. 73, No. 1, str. 47–52, 2006.
- [133] D. Westwick, M. Verhaegen, Identifying mimo wiener systems using subspace model identification methods, *Signal Processing*, Vol. 52, str. 235–258, 1996.
- [134] A. S. Willsky, A survey of design methods for failure detection in dynamic systems, *Automatica*, Vol. 12, str. 601–611, 1976.
- [135] J. Wünnenberg, *Observer-based fault detection in dynamic systems*, Doktorska disertacija, Universitaet Duisburg, Fachgebiet Mess und Regelungstechnik, 1990.
- [136] H. Ying, G. Chen, Necessary conditions for some typical fuzzy systems as universal approximators, *Automatica*, Vol. 33, str. 1333–1338, 1997.
- [137] Xiaodong Zhang, Thomas Parisini, Marios M. Polycarpou, Adaptive fault-tolerant control of nonlinear uncertain systems: An information-based diagnostic approach, *IEEE Transactions on Automatic Control*, Vol. 49, No. 8, str. 1259–1274, 2004.
- [138] Xiaodong Zhang, Marios M. Polycarpou, Thomas Parisini, Robust fault isolation for a class of non-linear input-output systems, *International Journal of Control*, Vol. 74, No. 13, str. 1295–1310, 2001.
- [139] Xiaodong Zhang, Marios M. Polycarpou, Thomas Parisini, A robust detection and isolation scheme for abrupt and incipient faults in nonlinear systems, *IEEE Transactions on Automatic Control*, Vol. 47, No. 4, str. 576–593, 2002.

Izjava

Izjavljam, da sem doktorsko disertacijo izdelal samostojno pod vodstvom mentorja prof. dr. Igorja Škrjanca. Izkazano pomoč drugih sodelavcev sem v celoti navedel v zahvali.

Simon Oblak