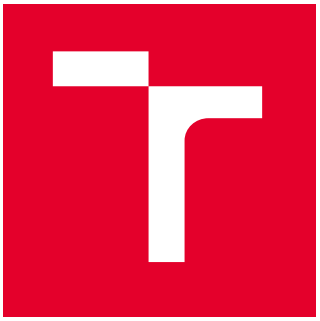


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV AUTOMATIZACE A MĚŘICÍ TECHNIKY

DEPARTMENT OF CONTROL AND INSTRUMENTATION

BAYESOVSKÉ PŘÍSTUPY PRO ADAPTIVNÍ IDENTIFIKACI SYSTÉMŮ

BAYESIAN APPROACHES TO ADAPTIVE SYSTEM IDENTIFICATION

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE — **Bc. Ondřej Skalský**
AUTHOR

VEDOUCÍ PRÁCE — **Ing. Jakub Dokoupil, Ph.D.**
SUPERVISOR

BRNO 2024

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

Diplomová práce

magisterský navazující studijní program **Kybernetika, automatizace a měření**

Ústav automatizace a měřicí techniky

Student: Bc. Ondřej Skalský ***ID:*** 221015

Ročník: 2 ***Akademický rok:*** 2023/24

NÁZEV TÉMATU:

Bayesovské přístupy pro adaptivní identifikaci systémů

POKYNY PRO VYPRACOVÁNÍ:

1. Formulujte úlohu identifikace normálního regresního modelu pomocí bayesovských paradigmat.
2. Pro časově proměnné systémy uvažujte neznalost modelu vývoje parametrů. Absenci modelu kompenzujte aplikací některé z technik statistického rozhodování.
3. Realizujte algoritmizaci pro přepínání mezi modely v rámci konečné banky modelů a algoritmizaci pro datově informované zapomínání.
4. Vlastnosti zvolených přístupů ověřte na numerickém příkladu i na fyzikální soustavě.

DOPORUČENÁ LITERATURA:

V. Peterka, “Bayesian approach to system identification”, in Trends and Progress in System Identification (P. Eykhoff ed.), pp. 239–304. Pergamon Press, Oxford, 1981.

Termín zadání: 5.2.2024 ***Termín odevzdání:*** 15.5.2024

Vedoucí práce: Ing. Jakub Dokoupil, Ph.D.

doc. Ing. Petr Fiedler, Ph.D.
předseda rady studijního programu

UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Práce se zabývá bayesovskou identifikací časově variantních normálních regresních modelů a skýtá celkem čtyři hlavní algoritmy. První dvě uvedené algoritmizace slouží pro průběžnou regularizovanou identifikaci jednoho regresního modelu. Neznámost časového vývoje tohoto modelu je u obou algoritmizací řešena technikou datově informovaného zapomínání. Volba faktoru zapomínání je u první uvedené algoritmizace prováděna pomocí variační bayesovské aproximace. Druhý algoritmus určuje hodnotu faktoru zapomínání statistickým rozhodováním.

Druhá dvojice algoritmů pohlíží na problematiku časového vývoje parametrů jako na sekvenci přepínání normálních regresních modelů. U obou těchto dávkových algoritmizací založených na variační bayesovské aproximaci je iterativně prováděna inference nejen parametrů banky modelů, ale i aktivit těchto modelů po čas experimentu, přičemž skutečný počet modelů tvořících tuto banku je určen automaticky. Rozdíl mezi těmito algoritmy spočívá především v šumových vlastnostech přepínaných modelů.

Všechny čtyři algoritmy jsou testovány na reálné soustavě a v simulacích. Práce je doplněna krátkým úvodem do bayesovského světa, jež vysvětluje užité statistické techniky. Pro úplnost a návaznost na zmíněné algoritmy je popsána i dávková a průběžná identifikace časově invariantního normálního regresního modelu.

KLÍČOVÁ SLOVA

Identifikace, Bayesovská statistika, Pravděpodobnost, Normální regresní model, Časově variantní parametrická identifikace, ARX model, Směs normálních regresních modelů, Bayesovská identifikace směsi normálních regresních modelů, Variační bayesovská aproximace.

ABSTRACT

The paper deals with Bayesian identification of time variant normal regression models and provides four key algorithms. The first two algorithms are designed for the online regularized identification of a single regression model. Absence of the time evolution model is solved in both algorithmizations by a data-informed forgetting technique. The choice of the forgetting factor in the first algorithm is performed using a variational Bayesian approximation. The second algorithm determines the value of the forgetting factor by statistical decision making.

The second pair of algorithms treats the problem of time evolution of parameters as a sequence of switching normal regression models. In both of these offline algorithms which are based on the variational Bayesian approximation, both the inference of the bank of model parameters and the activities of these models over the time of the experiment are iteratively performed. The actual number of models forming this bank is determined automatically. The difference between these algorithms lies mainly in the noise properties of the switched models.

All four algorithms are tested on a real system and in simulations. The paper is complemented by a short introduction to the Bayesian world, which presents the essential statistical techniques that are used. For completeness and continuity with the aforementioned algorithms, the offline and online identification of a time invariant normal regression model is also described.

KEYWORDS

Identification, Bayesian statistics, Probability, Normal regression model, Time variant parametric identification, ARX model, Mixture of normal regression models, Identification of mixture of normal regression models, Variational Bayesian approximation.

Vysázeno pomocí balíčku `thesis` verze 4.09; `https://latex.fekt.vut.cz/`

SKALSKÝ, Ondřej. *Bayesovské přístupy pro adaptivní identifikaci systémů*. Diplomová práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav automatizace a měřicí techniky, 2024. Vedoucí práce: Ing. Jakub Dokoupil, Ph.D.

Prohlášení autora o původnosti díla

Jméno a příjmení autora:	Bc. Ondřej Skalský
VUT ID autora:	221015
Typ práce:	Diplomová práce
Akademický rok:	2023/24
Téma závěrečné práce:	Bayesovské přístupy pro adaptivní identifikaci systémů

Prohlašuji, že svou závěrečnou práci jsem vypracoval samostatně pod vedením vedoucí/ho závěrečné práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené závěrečné práce dále prohlašuji, že v souvislosti s vytvořením této závěrečné práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno ....

podpis autora*

*Autor podepisuje pouze v tištěné verzi.

PODĚKOVÁNÍ

Děkuji vedoucímu diplomové práce, panu Ing. Jakubu Dokoupilovi Ph.D., za odborné vedení, konzultace, trpělivost, motivaci a podnětné návrhy k práci. Děkuji také mé rodině za podporu.

Obsah

Úvod **13**

1 Bayesovský přístup k identifikaci **14**
1.1 Funkce hustoty pravděpodobnosti, pravděpodobnostní funkce 14
1.2 Operace nad náhodnými veličinami 15
1.2.1 Marginalizace 15
1.2.2 Řetězové pravidlo 15
1.2.3 Nezávislost a podmíněná nezávislost 15
1.2.4 Bayesův vzorec 16
1.2.5 Očekávaná hodnota 16
1.3 Vybrané pravděpodobnostní funkce a funkce hustot pravděpodobností 16
1.3.1 Normální rozložení 16
1.3.2 Wishartovo rozložení 17
1.3.3 Zkrácené exponenciální rozložení 18
1.3.4 Dirichletovo rozložení 18
1.3.5 Kategorické rozložení 19
1.3.6 Rovnoměrné rozložení 20
1.4 Volba prioru 20
1.5 Variační bayesovská metoda distribuční aproximace 21
1.5.1 Volba rodiny řešitelných posteriorů 21
1.5.2 Volba optimalizačního funkcionálu 21
1.5.3 Minimalizace optimalizačního funkcionálu 22
1.5.4 Dosažení konjugace pomocí variační bayesovské metody 23

2 Bayesovská identifikace časově invariantního normálního regresního modelu **24**
2.1 Modely s chybou rovnice 24
2.2 Normální regresní model ARX 25
2.3 Bayesovská identifikace časově invariantního normálního regresního modelu 26
2.3.1 Dávková varianta 26
2.3.2 Rekurzivní varianta 28

3 Bayesovská identifikace časově variantního normálního regresního modelu **30**
3.1 Regularizovaná identifikace časově variantního normálního modelu s exponenciálním zapomínáním 30

3.1.1 Časový krok . . . 30
3.1.2 Regularizace . . . 33
3.1.3 Datový krok . . . 34
3.2 Datově informovaná volba koeficientu zapomínání určeným variační bayesovskou aproximací . . . 35
3.2.1 VB aproximace náhodné veličiny Θ_k . . . 35
3.2.2 VB aproximace náhodné veličiny λ_k . . . 36
3.2.3 Algoritmus . . . 39
3.3 Datově informovaná volba koeficientu zapomínání určeným statistickým rozhodováním . . . 40
3.3.1 Volba optimalizačního funkcionálu . . . 40
3.3.2 Volba dodatečné penalizace alternativ predikce . . . 41
3.3.3 Nutné a postačující podmínky optimality . . . 42
3.3.4 Analytické řešení Kullback-Leiblerovy divergence dvojice normálních-Wishartových rozložení . . . 43
3.3.5 Nalezení koeficientu zapomínání . . . 43
3.3.6 Algoritmus . . . 45

4 Bayesovská identifikace banky přepínaných normálních časově invariantních regresních modelů 46
4.1 Bayesovský GM model . . . 46
4.1.1 Užití VB aproximace pro dosažení konjugace . . . 48
4.1.2 Vyjádření věrohodnostní funkce . . . 48
4.2 Identifikace přepínaných ARX modelů zatížených šumem o společné přesnosti . . . 49
4.2.1 Volba prioru . . . 49
4.2.2 Vyjádření VB marginálií . . . 50
4.2.3 Algoritmus . . . 52
4.3 Identifikace přepínaných ARX modelů zatížených šumy o obecně různé přesnosti . . . 53
4.3.1 Volba prioru . . . 53
4.3.2 Vyjádření VB marginálií . . . 53
4.3.3 Algoritmus . . . 56
4.4 Další aproximace pro dosažení kunjugace . . . 56
4.4.1 Důsledky původní aproximace . . . 57
4.4.2 Volba optimalizačního funkcionálu . . . 57

5 Experimenty 58
5.1 Reálná soustava . . . 58

5.1.1 Operátorový přenos 58
5.1.2 Parametry a nastavení měření na reálné soustavě 61
5.2 Testování rekurzivních algoritmů pro online identifikaci časově proměnného ARX modelu 63
5.2.1 Určení koeficientu zapomínání pomocí VB metody 63
5.2.2 Určení koeficientu zapomínání statistickým rozhodováním . . . 66
5.2.3 Vliv regularizace a důsledek při její absenci 70
5.3 Testování dávkových algoritmů pro identifikaci banky ARX modelů . 71
5.3.1 Identifikace přepínaných ARX modelů zatížených šumem o společné přesnosti 71
5.3.2 Identifikace přepínaných ARX modelů zatížených šumy o obecně různé přesnosti 73
5.3.3 Shrnutí obou algoritmů pro identifikaci banky ARX modelů . 75

Závěr **76**

Literatura **77**

Seznam symbolů a zkratek **79**

A Obsah elektronické přílohy **81**

Seznam obrázků

2.1 Obecný model s chybou rovnice [5] 24
5.1 Elektrické schéma reálného dynamického systému [19] 59
5.2 Blokové schéma reálného dynamického systému [19] 61
5.3 Identifikace reálné soustavy algoritmem určujícího koeficient zapomínání VB aproximací při parametrizaci: $\alpha = 0.01$, $b_0 = 0$, $\hat{\lambda}_k^{[0]} = 0.7\,\forall k \in [1, N]$ 64
5.4 Identifikace simulované soustavy zatížené šumem o přesnosti $d = 500$ algoritmem určujícího koeficient zapomínání VB aproximací při parametrizaci: $\alpha = 0.01$, $b_0 = 0$, $\hat{\lambda}_k^{[0]} = 0.7\,\forall k \in [1, N]$ 65
5.5 Identifikace reálné soustavy algoritmem určujícího koeficient zapomínání statistickým rozhodováním - varianta $\gamma \equiv \zeta$ při parametrizaci: $\alpha = 0.2$, $\zeta = 0.15$ 67
5.6 Identifikace simulované soustavy zatížené šumem o přesnosti $d = 100$ algoritmem určujícího koeficient zapomínání statistickým rozhodováním - varianta $\gamma \equiv \zeta$ při parametrizací: $\alpha = 0.2$, $\zeta = 0.2$ 68
5.7 Identifikace simulované soustavy zatížené šumem o přesnosti $d = 100$ algoritmem určujícího koeficient zapomínání statistickým rozhodováním - varianta $\gamma \equiv \rho$ při parametrizaci: $\alpha = 0.2$, $\rho = 0.6$ 69
5.8 Vliv regularizace a důsledek při její absenci 70
5.9 Identifikace reálné soustavy algoritmem pro identifikaci přepínaných ARX modelů zatížených šumy o společné přesnosti $\kappa = 500$ 72
5.10 Identifikace simulované soustavy zatížené šumem o přesnosti $d = 500$ algoritmem pro identifikaci přepínaných ARX modelů zatížených šumy o společné přesnosti 72
5.11 Identifikace reálné soustavy algoritmem pro identifikaci přepínaných ARX modelů zatížených šumy o obecně různé přesnosti 74
5.12 Identifikace simulované soustavy zatížené šumem o přesnosti $d = 500$ algoritmem pro identifikaci přepínaných ARX modelů zatížených šumy o obecně různé přesnosti 74

Seznam algoritmů

1 Algoritmus rekurzivní identifikace s koeficientem zapomínání určeným VB aproximací 39
2 Algoritmus rekurzivní identifikace s koeficientem zapomínání určeným statistickým rozhodováním 45
3 Algoritmus jednokrokové identifikace přepínaných ARX modelů zatížených šumem o společné přesnosti 52
4 Algoritmus jednokrokové identifikace přepínaných ARX modelů zatížených šumy o obecně různé přesnosti 56

Úvod

Identifikace dynamických systémů je neodmyslitelně spjatá s teorií dynamických systémů, s teorií řízení a regulace již od samotného vzniku těchto vědních disciplín. Rozmach adaptivních regulátorů respektujících přirozenou časovou variantnost skutečných systémů si žádá algoritmů schopných v reálném čase identifikovat parametry kýžené soustavy. Identifikace parametrů dynamických časově variantních systémů však nemusí být použita jen pro zkvalitnění regulačního děje. Je možné ji užít i všude tam, kde známost časového vývoje parametrů odráží zkoumaný fyzikální, chemický či jiný děj [1].

Jak průběžná, tak jednorázová identifikace časově proměnných systémů zasahuje do řady oborů a není tak překvapující, že algoritmizací pro tyto účely byla vyvinuta celá řada. Jak tomu ale už tak bývá, ani v této oblasti neexistuje jedna konkrétní algoritmizace představující dokonalé řešení. Jedním z hlavních důvodů je, že možných skutečných vnitřních realizací dynamických systémů, jejich struktur a variant je vlastně nekonečně mnoho.

Akceptujeme-li zanedbání volbou konkrétního modelu, jež zvolíme jako matematickou náhradu skutečného systému, potýkáme se s dalším problémem. Pro identifikaci je totiž zapotřebí dat, jejichž měření je vyjma specifických případů prakticky vždy zatíženo chybou. A to nejen z důvodu omezené přesnosti měřicích přístrojů, ale také vlivem nepřímo měřitelné stochastické složky působící na systém.

Zatímco cílem konvenčních kriteriálních přístupů je navzdory nepřesnému měření co nejlépe odhadovat parametry zvoleného modelu, bayesovský přístup dokáže mimo bodový odhad parametrů přirozeně poskytovat i informaci o nejistotách těchto bodových odhadů, kdy nejistota je vnesena právě prostřednictvím nepřesných dat.

Pro účely této práce je předpokládán normální regresní model. Bayesovskou dávkovou a průběžnou identifikací parametrů tohoto modelu, za předpokladu jeho časové invariantnosti, získáme základní matematický aparát. Ten bude sloužit jako podklad pro formulaci dvou alternativních přístupů ke sledování časového vývoje parametrů identifikovaného systému. Zatímco první přístup bude odrážet časovou varianci identifikované soustavy průběžnou změnou parametrů jednoho normálního regresního modelu, druhý přístup bude na základě datové sady inferovat počet normálních regresních modelů, jejich parametry a retrospektivně určí aktivitu těchto modelů v čase. Pro obě z těchto úloh budou představeny dva algoritmy, jejichž vlastnosti budou demonstrovány na reálné soustavě i v simulaci.

1 Bayesovský přístup k identifikaci

Zatímco přístupy založené na minimalizaci kriteriálních funkcí zachází s parametry jako se souborem pevných, a přesto neznámých hodnot, bayesovská statistika interpretuje parametry jako náhodné veličiny. To, že parametry považujeme za náhodné veličiny jim ve skutečnosti nepřipisuje náhodný charakter. Náhodnost chápeme pouze ve smyslu kvantifikace nejistoty ohledně možných hodnot parametrů, ačkoliv pouze jedna z hodnot je správná. Bayesovský přístup nám jako řešení poskytuje funkce hustot pravděpodobností či pravděpodobnostní funkce namísto bodových odhadů parametrů. Díky tomu problémy konvenčních přístupů jako nestrannost či konfidenční intervaly v bayesovském světě zanikají. Při správné volbě pravděpodobnostního modelu lze tímto přístupem formulovat i některé osvědčené metody jako Kalmanův filtr nebo rekurzivní metodu nejmenších čtverců. Bayesovský přístup využívá pro inferenci konzistentního pravděpodobnostního kalkulu a je tak rovněž konzistentním nástrojem. Lze jej prohlásit za všestranný nástroj, jež nachází uplatnění nejen v identifikaci, ale i v mnoha inženýrských oblastech [2, 3].

Náplní této kapitoly je seznámení se nástroji bayesovského světa, kterých bude v práci využíváno. Dále budou diskutovány konkrétní funkční formy náhodných veličin a provedeny výpočty některých očekávaných hodnot. Na informace shrnuté v této kapitole se budeme v průběhu práce dále odkazovat.

1.1 Funkce hustoty pravděpodobnosti, pravděpodobnostní funkce

Funkce hustoty pravděpodobnosti $f(x)$ spojité náhodné veličiny x může být libovolná funkce splňující podmínky nezápornosti a jednotkového objemu informace.

$$f(x) \geq 0, \forall x \in x^* \tag{1.1}$$

$$\int_{x^*} f(x)\mathrm{d}x = 1 \tag{1.2}$$

Je-li x diskrétní náhodnou veličinou a $f(x)$ pravděpodobnostní funkcí, pak integrál 1.2 přejde v sumu $\sum_{x^*} f(x) = 1$. Značení x^* definuje množinu možných hodnot veličiny x.

I nadále bude $f(\cdot)$ označovat jak funkci hustoty pravděpodobnosti (fhp), tak pravděpodobnostní funkci (pf) a informaci o diskrétním či spojitém charakteru bude nést samotná náhodná veličina. Stejného sjednocení notace lze dosáhnout i pomocí teorie míry, která podle Radonova-Nikodymova teorému pracuje generalizovaně konzistentním způsobem pro diskrétní i spojitá rozdělení [4].

1.2 Operace nad náhodnými veličinami

Pro vysvětlení základních operací nad náhodnými veličinami zavedeme sdruženou funkci hustoty pravděpodobnosti $f(\Theta, \mathcal{D})$. Interpretovatelnost spojitých náhodných veličin $\Theta \equiv \{\theta, d\}, \mathcal{D}$ není v tomto okamžiku nezbytná. Můžeme však již nyní pracovat s myšlenkou, že Θ reprezentuje odhadované parametry a $\mathcal{D}$ jsou měřená data. V souladu s touto myšlenkou a obdobným značením budeme pokračovat i v následujících kapitolách.

Přestože pracujeme v následujících kapitolách se spojitou náhodnou veličinou, jsou popsané operace obecně platné i pro náhodné veličiny diskrétní. Musíme však mít na paměti, že integrace musí být nahrazena sumací, kdykoliv je argument diskrétní.

1.2.1 Marginalizace

Pro vyjádření marginální fhp $f(\theta|\mathcal{D})$ provedeme operaci marginalizace [3, 4]

$$f(\theta|\mathcal{D}) = \int_{d^*} (\theta, d|\mathcal{D}) \mathrm{d}d. \tag{1.3}$$

1.2.2 Řetězové pravidlo

Fhp $f(\theta, d|\mathcal{D})$ lze pomocí řetězového pravidla rozepsat [3, 4]

$$f(\theta, d|\mathcal{D}) = f(\theta|d, \mathcal{D}) f(d|\mathcal{D}). \tag{1.4}$$

1.2.3 Nezávislost a podmíněná nezávislost

Pro vyhnutí se zavádějícímu zápisu vzhledem ke shodě zvolené notace s notací následujících kapitol budeme pracovat s náhodnými veličinami a, b, c. Veličiny a, b jsou nezávislé, pakliže platí

$$\begin{aligned} f(a, b) &= f(a|b) f(b) = f(a) f(b) \\ &= f(b|a) f(a). \end{aligned} \tag{1.5}$$

Veličiny a, b jsou podmínečně nezávislé za předpokladu znalosti c, pokud platí

$$\begin{aligned} f(a, b|c) &= f(a|b, c) f(b|c) = f(a|c) f(b|c) \\ &= f(b|a, c) f(a|c). \end{aligned} \tag{1.6}$$

Podotkněme, že podmínečná nezávislost 1.6 neimplikuje nezávislost 1.5 [3, 4].

1.2.4 Bayesův vzorec

Bayesův vzorec představuje základní stavební kámen celého bayesovského přístupu. Říká, že posterior $f(\Theta|\mathcal{D})$ je možné vyjádřit jako

$$f(\Theta|\mathcal{D}) = \frac{f(\mathcal{D}|\Theta)f(\Theta)}{f(\mathcal{D})} = \frac{f(\mathcal{D}|\Theta)f(\Theta)}{\int_{\Theta^*} f(\mathcal{D}|\Theta)f(\Theta)\mathrm{d}\Theta} \propto f(\mathcal{D}|\Theta)f(\Theta), \tag{1.7}$$

kde $f(\Theta)$ je prior, $f(\mathcal{D}|\Theta)$ je nazývána jako věrohodnostní funkce. Jelikož člen $f(\mathcal{D})$ představuje vzhledem k $f(\Theta|\mathcal{D})$ pouze normalizační konstantu, je možné v 1.7 použít znak přímé úměry $\propto$ pro vyjádření vztahu mezi $f(\Theta|\mathcal{D})$ a $f(\mathcal{D}|\Theta)f(\Theta)$ [3, 4].

1.2.5 Očekávaná hodnota

Očekávaná hodnota obecné vektorové funkce $g(x)$ nad náhodnou veličinou x mající fhp $f(x)$ je vyjádřena zápisem [2, 3]

$$\mathsf{E}_{f(x)}\left[g(x)\right] \equiv \int_{x^*} g(x)f(x)\mathrm{d}x. \tag{1.8}$$

Pro zjednodušení zápisu zavedeme v literatuře často užívané kompaktní značení očekávané hodnoty

$$\widehat{g(x)} \equiv \mathsf{E}_{f(x)}\left[g(x)\right]. \tag{1.9}$$

Fhp náhodné veličiny je u tohoto značení patrno z kontextu za předpokladu provázanosti dané náhodné veličiny s příslušným rozložením [2].

1.3 Vybrané pravděpodobnostní funkce a funkce hustot pravděpodobností

Některé z funkčních forem fhp či pf jsou natolik významné, že mají své jméno. Známá parametrická funkční forma spolu s příslušnými postačujícími statistikami (parametry) definuje funkční předpis fhp či pf příslušné náhodné veličiny. Pro eliminaci nejasností plynoucích z nuancí napříč literaturou definujme parametrizace funkčních forem fhp a pf užitých v této práci. Uveďme rovněž očekávané hodnoty náhodných veličin příslušných rozložení nad konkrétními funkcemi, jejichž výsledků budeme v průběhu práce využívat.

1.3.1 Normální rozložení

Spojitá náhodná veličina $\theta \in \mathbb{R}^p$ má normální rozložení $\mathcal{N}_\theta(\hat{\theta}, P/d)$ s parametry $\hat{\theta} \in \mathbb{R}^p$ a P/d, kde $P \in \mathbb{R}^{p\times p}$ je pozitivně definitní symetrická matice a $d \in \mathbb{R}_{>0}$,

jestliže její fhp je ve tvaru Gaussovy funkce [3]

$$f\left(\theta\middle|\hat{\theta}, P/d\right) = \mathcal{N}_\theta\left(\hat{\theta}, P/d\right) = (2\pi/d)^{-p/2}|P|^{-1/2}\exp\left\{-\frac{d}{2}\left(\theta-\hat{\theta}\right)' P^{-1}\left(\theta-\hat{\theta}\right)\right\}. \tag{1.10}$$

Střední hodnota normálního rozložení je podle 1.8 dána jako

$$\mathsf{E}_{\mathcal{N}_\theta(\hat{\theta},P/d)}\left[\theta\right] = \hat{\theta}, \tag{1.11}$$

očekávaná hodnota součinu $\theta\theta'$, tedy druhý obecný moment, je

$$\mathsf{E}_{\mathcal{N}_\theta(\hat{\theta},P/d)}\left[\theta\theta'\right] = \hat{\theta}\hat{\theta}' + P/d. \tag{1.12}$$

Z 1.12 plyne, že $\widehat{\theta'\theta} = \hat{\theta}'\hat{\theta} + \mathrm{tr}(P)/d$, neboť $\widehat{\theta'\theta} = \mathrm{tr}(\widehat{\theta\theta'})$, kde $\mathrm{tr}(\cdot)$ je stopa matice. Obdobně lze odvodit i výsledek $\widehat{\theta' V \theta'} = \hat{\theta}' V \hat{\theta} + \mathrm{tr}(VP)/d$, kde V je symetrická pozitivně definitní matice. Integrál pro výpočet 1.12 je možné řešit za pomoci Choleského dekompozice $P^{-1} = UU'$ užitím substituce $q = d^{\frac{1}{2}}U'(\theta - \hat{\theta})$ a následným užitím metody per partes pro řešení integrálu součinu polynomu a exponenciální funkce [13]. Řešení 1.11 obdržíme substitucí $q = \theta - \hat{\theta}$.

1.3.2 Wishartovo rozložení

Spojitá náhodná veličina $d \in \mathbb{R}_{>0}$ má jednodimenzionální Wishartovo rozložení $\mathcal{W}_d(\Sigma, \nu)$ s parametry $\Sigma, \nu \in \mathbb{R}_{>0}$, jestliže její fhp je ve tvaru funkce

$$f(d|\Sigma,\nu) = \mathcal{W}_d(\nu,\Sigma) = \frac{(\Sigma/2)^{\nu/2}}{\Gamma(\nu/2)} d^{(\nu-2)/2}\exp\{-\Sigma d/2\}, \tag{1.13}$$

kde $\Gamma(\cdot)$ je gama funkce definována jako [3]

$$\Gamma(z) \equiv \int_0^\infty t^{z-1}\exp\{-t\}\mathrm{d}t. \tag{1.14}$$

Stejný funkční zápis je možné při úpravě parametrizace vyjádřit i pomocí gama $\mathcal{G}(\cdot)$ rozložení, které se v literatuře často užívá. Platí, že $\mathcal{W}_d(\Sigma,\nu) = \mathcal{G}_d(\Sigma/2, \nu/2)$. Střední hodnota Wishartova rozložení je podle 1.8 dána jako

$$\mathsf{E}_{\mathcal{W}_d(\Sigma,\nu)}\left[d\right] = \nu/\Sigma. \tag{1.15}$$

Očekávaná hodnota logaritmu $\ln(d)$ je

$$\mathsf{E}_{\mathcal{W}_d(\Sigma,\nu)}\left[\ln(d)\right] = \Psi\left(\nu/2\right) - \ln\left(\Sigma/2\right), \tag{1.16}$$

kde $\Psi(\cdot)$ je digama funkce definována jako

$$\Psi(z) \equiv \frac{1}{\Gamma(z)}\int_0^\infty t^{z-1}\exp(-t)\ln(t)\mathrm{d}t. \tag{1.17}$$

Řešení integrálů pro obdržení obou expektací 1.15, 1.16 je možné obdržet substitucí $q = d\Sigma/2$ a následným porovnáním s definičními vztahy 1.14, 1.17.

1.3.3 Zkrácené exponenciální rozložení

Spojitá náhodná veličina $\lambda \in [c_l, c_h]$ má omezené exponenciální rozložení $\mathcal{E}_\lambda(b, [c_l, c_h])$ s parametrem $b \in \mathbb{R}$, jestliže její fhp je ve tvaru funkce

$$f(\lambda|b) = \mathcal{E}_\lambda\left(b, [c_l, c_h]\right) = \frac{b}{\exp\{bc_h\} - \exp\{bc_l\}} \exp\{\lambda b\}. \tag{1.18}$$

Střední hodnota veličiny λ je

$$\mathsf{E}_{\mathcal{E}_\lambda(b,[c_l,c_h])}\left[\lambda\right] = \frac{(1 - c_h b)\exp\{c_h b\} - (1 - c_l b)\exp\{c_l b\}}{k\exp\{c_l b\} - \exp\{c_h b\}} \tag{1.19}$$

Výsledek 1.19 obdržíme užitím metody per partes při řešení příslušného integrálu [2].

1.3.4 Dirichletovo rozložení

Spojitá náhodná veličina $\alpha \equiv [\alpha_1, ..., \alpha_M]' \in \Delta_M$ má Dirichletovo rozložení $\mathcal{D}_\alpha(\tau)$ s parametrem $\tau \equiv [\tau_1, ..., \tau_M]'$, jestliže její fhp je ve tvaru funkce

$$f(\alpha|\tau) = \mathcal{D}_\alpha(\tau) = \frac{\Gamma\left(\sum_{i=1}^M \tau_i\right)}{\prod_{i=1}^M \Gamma(\tau_i)} \prod_{i=1}^M \alpha_i^{\tau_i - 1}. \tag{1.20}$$

Množina definující prostor náhodné veličiny $\alpha \in \Delta_M$ je simplex dimenze $M - 1$ definován jako

$$\Delta_M \equiv \left\{ \alpha \middle| \alpha_i \in \mathbb{R}^M_{\geq 0}, \sum_{i=1}^M \alpha = 1 \right\}. \tag{1.21}$$

Zde podotkněme, že simplex může být definován i jako prostor $\alpha_i \in \mathbb{R}^{M-1}_{\geq 0}$ s omezující podmínkou $\sum_{i=1}^{M-1} \alpha_i \leq 1$. Obě definice jsou ekvivalentní, liší se pouze Lebesgueovou mírou [18].

Střední hodnota i-tého prvku náhodné veličiny Dirichletova rozložení je dána jako

$$\mathsf{E}_{\mathcal{D}_\alpha(\tau)}\left[\alpha_i\right] = \frac{\tau_i}{\sum_{i=1}^M \tau_i}, \tag{1.22}$$

očekávaná hodnota logaritmu $\ln(\alpha_i)$ i-tého prvku je

$$\mathsf{E}_{\mathcal{D}_\alpha(\tau)}\left[\ln(\alpha_i)\right] = \Psi(\tau_i) - \Psi\left(\sum_{i=1}^M \tau_i\right). \tag{1.23}$$

Pro získání expektací 1.22 a 1.23 byla nejprve provedena marginalizace fhp 1.20 přes prvky α_j, kde $j \in [1, M], j \neq i$, neb řešení těchto expektací nad $f(\alpha|\tau)$ se redukuje na řešení expektací nad marginální hustotou $f(\alpha_i|\tau)$.

$$f(\alpha_i|\tau) = \frac{\Gamma\left(\sum_{j=1}^{M} \tau_j\right)}{\Gamma\left(\tau_i\right)\Gamma\left(\sum_{j=1, j\neq i}^{M} \tau_j\right)} \left(\alpha_i\right)^{\tau_i - 1} \left(1 - \alpha_i\right)^{\sum_{j=1, j\neq i}^{M} \tau_j - 1} \tag{1.24}$$

Obecné odvození výsledku 1.24 je příliš zdlouhavé. Odvození ukažme na příkladu pro $M = 4$, kdy hledáme $f(\alpha_1|\tau)$. Při postupu řešení marginalizačních integrálů 1.25 nejprve za α_4 dosadíme $\alpha_4 = \alpha_1 + \alpha_2 + \alpha_3$ a následně provedeme substituci $\alpha_3 = z\left(1 - \alpha_1 - \alpha_2\right)$. Algebraickými úpravami obdržíme integrál, jehož řešení je patrné z definice beta funkce 1.26 uvedené níže.

$$\begin{aligned}
&\alpha_1^{\tau_1 - 1} \int_0^{1-\alpha_1} \alpha_2^{\tau_2 - 1} \int_0^{1-\alpha_1-\alpha_2} \alpha_3^{\tau_3 - 1} \alpha_4^{\tau_4 - 1} \mathrm{d}\alpha_3 \mathrm{d}\alpha_2 \\
&= \alpha_1^{\tau_1 - 1} \int_0^{1-\alpha_1} \alpha_2^{\tau_2 - 1} \left(1 - \alpha_1 - \alpha_2\right)^{\tau_3 + \tau_4 - 1} \int_0^1 z^{\tau_3 - 1} \left(1 - z\right)^{\tau_4 - 1} \mathrm{d}z \mathrm{d}\alpha_2 \\
&= \mathrm{B}(\tau_3, \tau_4) \alpha_1^{\tau_1 - 1} \int_0^{1-\alpha_1} \alpha_2^{\tau_2 - 1} \left(1 - \alpha_1 - \alpha_2\right)^{\tau_3 + \tau_4 - 1} \mathrm{d}\alpha_2
\end{aligned} \tag{1.25}$$

$$\mathrm{B}(z_1, z_2) \equiv \int_0^1 t^{z_1 - 1} (1 - t)^{z_2 - 1} \mathrm{d}z = \frac{\Gamma(z_1)\Gamma(z_2)}{\Gamma(z_1 + z_2)}. \tag{1.26}$$

Opakovaným zaváděním takovéto lineární substituce, která transformuje horní integrační mez na hodnotu 1 a dolní zachovává na hodnotě 0, a následným porovnáním s 1.26, bychom vyřešili všechny marginalizační integrály. Generalizací tohoto postupu obdržíme obecný vztah 1.24.

Jak bychom přirozeně předpokládali, marginální hustota 1.24 nabývá tvaru beta rozložení s parametry τ_i a $\sum_{j=1, j\neq i}^{M} \tau_j$, neboť Dirichletovo rozložení je vícerozměrným zobecněním beta rozložení.

Při vycházení z marginální hustoty 1.24 spočívá nalezení řešení 1.22 ve vyjádření normalizační konstanty pro dvourozměrné Dirichletovo rozložení (respektive beta rozložení) s parametry $\tau_i + 1$ a $\sum_{j=1, j\neq i}^{M} \tau_j$ za využití vlastnosti $\Gamma(z + 1) = z\Gamma(z)$.

Pro obdržení řešení 1.23 bylo použito vyjádření 1.27

$$\ln\left(\alpha_i\right) \alpha_i^{\tau_i - 1} = \frac{\mathrm{d}}{\mathrm{d}\alpha_i} \left(\alpha_i\right)^{\tau_i - 1} \tag{1.27}$$

a derivace logaritmu gama funkce byla popsána jako

$$\frac{\mathrm{d}}{\mathrm{d}z} \ln\left(\Gamma(z)\right) = \Psi(z). \tag{1.28}$$

1.3.5 Kategorické rozložení

Spojitá náhodná veličina $z \equiv [z_1, ..., z_M]' \in \Delta_M$ má kategorické rozložení $\mathcal{C}_z\left(\alpha\right)$ s parametrem $\alpha \equiv [\alpha_1, ..., \alpha_M]' \in \Delta_M$, jestliže její fhp je ve tvaru funkce

$$f(z|\alpha) = \mathcal{C}_z(\alpha) = \frac{1}{\prod_{i=1}^{M} \Gamma\left(z_i\right)} \prod_{i=1}^{M} \alpha_i^{z_i}. \tag{1.29}$$

Střední hodnota kategorického rozložení je určena vektorem α [2].

$$\mathsf{E}_{\mathcal{C}_z(\alpha)}[z] = \alpha \tag{1.30}$$

Přestože funkční forma kategorického a Dirichletova rozložení je podobná, jak bylo uvedeno v [17], marginalizace kategorického rozložení již nevede na kategorické rozložení. Výsledek 1.30 tak byl bez analytického důkazu převzat z multinomiálního rozložení uvedeném v [2], které je zobecněním kategorického rozložení.

Pakliže však uvážíme původ kategorického rozložení, kdy $\sum_{i=1}^{M} z_i$ představuje celkový počet pokusů a α_i popisuje pravděpodobnost, s jakou nastane i-tý jev z celkem M kategorií, pak můžeme vydedukovat, že očekávaný počet výsledků $\hat{z}$ bude přímo úměrný pravděpodobnostem příslušných kategorií α. Pro rovnost pak zřejmě platí $\sum_{i=1}^{M} z_i = \sum_{i=1}^{M} \hat{z}_i$.

1.3.6 Rovnoměrné rozložení

Jedná se o takové rozložení spojité či diskrétní náhodné veličiny $x \in x^*$, kde platí, že $f(x) =$ konstanta. Můžeme se setkat s případy, kdy x^* je jednostranně či oboustranně neohraničeno. V takových případech rovnoměrné rozložení nemůže splňovat obě podmínky 1.1 a 1.2 současně a toto rozložení je tedy neřádné [4].

1.4 Volba prioru

Volba prioru $f(\Theta)$, jež se vyskytuje v Bayesově vzorci 1.7, neodráží jen apriorní znalosti statistika, ale svou funkcionální formou spolu s formou $f(\mathcal{D}|\Theta)$ determinují tvar posterioru $f(\Theta|\mathcal{D})$. Je-li forma prioru $f(\Theta)$ a posterioru $f(\Theta|\mathcal{D})$ shodná, hovoříme o konjungovaném prioru. Konjugace je přínosná zejména u rekurzivních algoritmizací, neboť výpočetní složitost zůstává v průběhu experimentu neměnná a funkcionální aktualizace zastaralého posterioru je převedena na algebraickou, jež představuje výpočet nových postačujících statistik. Docílení konjugace je však mnohdy kýžené i u jednokrokových algoritmizací.

Kromě zajištění konjugace volbou funkcionální formy $f(\Theta)$ může statistik prostřednictvím prioru pro špatně datově podmíněné případy suplementovat data $\mathcal{D}$. To je nazýváno jako regularizace priorem. Naopak je možné neinformativním priorem vyjádřit absenci apriorní znalosti o parametrech a nechat o posterioru $f(\Theta|\mathcal{D})$ rozhodnout věrohodnostní funkci $l(\Theta|\mathcal{D}) \equiv f(\mathcal{D}|\Theta)$ (*"Let the data speak for themselves."* [4]). Volbou prioru můžeme také zavést omezení reflektující reálné podmínky, jako například pozitivita. Z 1.7 plyne, že pokud prior $f(\Theta)$ bude pro některá $\Theta \in \Theta^*$ nula, pak i posterior $f(\Theta|\mathcal{D})$ bude pro tyto hodnoty nulový [2].

1.5 Variační bayesovská metoda distribuční aproximace

Bayesovské operace vedoucí k formulaci funkčních vyjádření posteriorních distribucí, jako jsou aplikace Bayesova pravidla, marginalizace, normalizace či vyjadřování momentů, mají analytická řešení jen pro omezený počet distribučních forem. Při neexistenci analytického řešení může být použita numerická integrace, je však výpočetně náročná, zejména pro více dimenzí.

Problém je možné obejít užitím analyticky řešitelné aproximace $\tilde{f}(\Theta|\mathcal{D})$ skutečného posterioru $f(\Theta|\mathcal{D})$. Pro zajištění řešitelnosti je potřeba nejprve zvolit podmnožinu řešitelných distribucí $\breve{f}(\Theta|\mathcal{D}) \in \breve{f}^*(\Theta|\mathcal{D}) \subset f^*(\Theta|\mathcal{D})$, kde $f^*(\Theta|\mathcal{D})$ je množina všech posteriorů. Distribuce $\breve{f}(\Theta|\mathcal{D})$ představují potencionální kandidáty na aproximaci, z nichž je vybrána taková funkční forma, která minimalizuje ztrátový funkcionál. Tu pak označíme jako $\tilde{f}(\Theta|\mathcal{D})$.

Existují různé volby $\breve{f}^*(\Theta|\mathcal{D})$ množiny analyticky řešitelných rozložení a ztrátových funkcionálů, jejichž optimalizací obdržíme různé $\tilde{f}(\Theta|\mathcal{D})$. V následujících odstavcích této kapitoly se však budeme zabývat aproximací pouze dle variační bayesovské metody, zkráceně VB aproximací [2].

1.5.1 Volba rodiny řešitelných posteriorů

Nechť $\Theta \equiv \{\Theta'_1, \Theta'_2, ..., \Theta'_q\} = \{\theta_1, \theta_2, ..., \theta_Q\} \in \mathbb{R}^Q$, kde každý z členů Θ_i zapouzdřuje jednu nebo více náhodných veličin θ_i, což implikuje $Q \geq q$. VB metoda aproximace množinu $\breve{f}^*(\Theta|\mathcal{D})$ definuje jako produkt podmínečně nezávislých maginálií

$$\breve{f}^*(\Theta|\mathcal{D}) \equiv \left\{ f(\Theta|\mathcal{D}) \middle| \breve{f}(\Theta|\mathcal{D}) = \prod_{i=1}^{q} \breve{f}(\Theta_i|\mathcal{D}) \right\}. \tag{1.31}$$

Cílem je volit q co nejmenší možné pro zajištění maximální korelace skutečného posterioru s aproximací za zachování analytické řešitelnosti [2].

1.5.2 Volba optimalizačního funkcionálu

Za předpokladu znalosti $\breve{f}^*(\Theta|\mathcal{D})$ je nalezení $\tilde{f}(\Theta|\mathcal{D})$ řešením optimalizačního problému

$$\tilde{f}(\Theta|\mathcal{D}) \equiv \arg\min_{\breve{f}(\cdot)} \Delta\Big(\breve{f}(\Theta|\mathcal{D}) \Big\| f(\Theta|\mathcal{D})\Big), \tag{1.32}$$

kdy hledáme $\breve{f}(\Theta|\mathcal{D}) \in \breve{f}^*(\Theta|\mathcal{D})$ minimalizující konvexní funkcionál $\Delta(\breve{f}(\cdot)\|f(\cdot))$. Volba tohoto funkcionálu provádějící zobrazení $\breve{f}^*(\Theta|\mathcal{D}) \times f^*(\Theta|\mathcal{D}) \to \mathbb{R}$ s unikátním minimem pro $\breve{f}(\Theta|\mathcal{D}) = f(\Theta|\mathcal{D})$ předurčuje vlastnosti aproximace $\tilde{f}(\Theta|\mathcal{D})$.

VB aproximace používá logaritmický ztrátový funkcionál, jež vede na Kullback-Leiblerovu divergenci (KLD), známou také jako relativní entropii

$$KL\Big(\breve{f}(\Theta|\mathcal{D})\Big\|f(\Theta|\mathcal{D})\Big) \equiv \int_{\Theta^*} \breve{f}(\Theta|\mathcal{D}) \ln\left(\frac{\breve{f}(\Theta|\mathcal{D})}{f(\Theta|\mathcal{D})}\right) \mathrm{d}\Theta. \tag{1.33}$$

Volba logaritmického ztrátového funkcionálu plyne z důkazu [14] o dosažení maximální extrakce informace z dat za použití ztrátové logaritmické funkce. Z definice 1.33 plynou vlastnosti tohoto funkcionálu, jako je jeho striktní konvexnost dokazatelná pomocí Jensenovy nerovnosti, nezápornost s unikátním minimem $KL(\breve{f}(\cdot)\|f(\cdot)) = 0$ pouze pro $\breve{f}(\cdot) = f(\cdot)$, neplatnost trojúhelníkové nerovnosti plynoucí z obecné nerovnosti $KL(\breve{f}(\cdot)\|f(\cdot)) \neq KL(f(\cdot)\|\breve{f}(\cdot))$ a $KL(\breve{f}(\cdot)\|f(\cdot)) = \infty$ při existenci takové množiny $\Theta_c^* \subset \Theta^*$, kdy $\int_{\Theta_c^*} \breve{f}(\Theta|\mathcal{D})\mathrm{d}\Theta > \int_{\Theta_c^*} f(\Theta|\mathcal{D})\mathrm{d}\Theta = 0$ [2].

1.5.3 Minimalizace optimalizačního funkcionálu

Za předpokladu platnosti 1.31 lze dokázat [2], že VB marginálie $\tilde{f}(\Theta_i|\mathcal{D})$ aproximace posterioru $\tilde{f}(\Theta|\mathcal{D}) = \prod_{i=1}^{q} \tilde{f}(\Theta_i|\mathcal{D})$ získaného jako

$$\tilde{f}(\Theta|\mathcal{D}) \equiv \arg\min_{\breve{f}(\cdot)} KL\Big(\breve{f}(\Theta|\mathcal{D})\Big\|f(\Theta|\mathcal{D})\Big) \tag{1.34}$$

lze vyjádřit až na normalizační konstantu pomocí

$$\tilde{f}(\Theta_i|\mathcal{D}) \propto \exp\left\{\mathsf{E}_{\tilde{f}(\Theta_{/i}|\mathcal{D})}\left[\ln\left(f(\Theta,\mathcal{D})\right)\right]\right\}, \tag{1.35}$$

kde $\Theta_{/i}$ je doplněk k Θ_i vzhledem k Θ a $\tilde{f}(\Theta_{/i}|\mathcal{D}) = \prod_{j=1,j\neq i}^{q} \tilde{f}(\Theta_j|\mathcal{D})$.

Jednokrokové řešení

Z 1.35 je patrné, že vyjádření standardních forem marginální $\tilde{f}(\Theta_i|\mathcal{D})$ jsou závislá na expektacích $f(\Theta|\mathcal{D})$ nad $\tilde{f}(\Theta_{/i}|\mathcal{D})$. Z toho plyne, že postačující statistiky VB marginálií $\tilde{f}(\Theta_i|\mathcal{D}), i \in [1,q]$ jsou vzájemně závislé. Pakliže soustava rovnic daných expektací má analytické řešení, lze tyto závislosti odstranit a postačující statistiky marginálních distribucí $\tilde{f}(\Theta_i|\mathcal{D})$ nalézt v jediném kroku výpočtu [2].

Iterativní řešení

Daleko častější je však situace, kdy rovnice jsou vzájemně transcendentální. Řešení je v tomto případě hledáno iterativním algoritmem známým pod názvem iterativní variační bayesovský algoritmus, zkráceně IVB algoritmus.

VB marginální hustota $\tilde{f}^{[r]}(\Theta|\mathcal{D})$ v r-tém kroku IVB je $\forall i \in [1,q]$ počítána jako expektace $f(\Theta|\mathcal{D})$ nad produktem zbývajících nejnovějších VB marginálií.

$$\tilde{f}^{[r]}(\Theta_i|\mathcal{D}) \propto \exp\left\{\mathsf{E}_{\left\{\prod_{j=1}^{i-1} \tilde{f}^{[r]}(\Theta_j|\mathcal{D}) \prod_{j=i+1}^{q} \tilde{f}^{[r-1]}(\Theta_j|\mathcal{D})\right\}}\left[\ln\left(f(\Theta,\mathcal{D})\right)\right]\right\} \tag{1.36}$$

Jak si ukážeme, inicializace $\tilde{f}^{[0]}(\Theta_i|\mathcal{D}), i \in [2, q]$ hraje roli nejen v rychlosti konvergence algoritmu, ale také předurčuje vzdálenost výsledné aproximace od globálního optima.

Jako ukazatel konvergence můžeme vyjádřit v každém kroku IVB algoritmu hodnotu funkcionálu 1.33 a na základě něj se rozhodovat o ukončení IVB. V rekurzivních algoritmizacích však bývá počet kroků IVB algoritmu mnohdy pevně stanoven pro zajištění konstantních časových výpočetních nároků.

Pakliže je možné úpravami vyjádřit některé z marginálií $\tilde{f}(\Theta|\mathcal{D})$ nezávisle na momentech náhodných veličin $\Theta_{/i}$, je nasnadě výpočet jejich postačujících statistik provést analyticky a IVB algoritmem počítat pouze zbylé marginální distribuce [2].

1.5.4 Dosažení konjugace pomocí variační bayesovské metody

Zajištění konjugace je nezbytné nejen pro rekurzivní algoritmy, ale obecně pro algoritmy, u kterých požadujeme nezávislost posteriorní funkční formy na počtu měření. Pro případ analytické řešitelnosti a nekonjugované formy posterioru $f(\Theta|\mathcal{D})$ je možné pomocí skryté (latentní) náhodné veličiny z rozšířit paramertický model $f(\mathcal{D}|\Theta) \rightarrow f(\mathcal{D}|\Theta, z)$ a získat konjungovaný rozšířený posterior $f(\Theta, z|\mathcal{D})$, jehož marginalizací přes z obdržíme původní $f(\Theta|\mathcal{D})$.

Namísto této analytické marginalizace, která je mnohdy neřešitelná, však můžeme použít VB metodu aproximace rozšířeného posterioru $f(\Theta, z|\mathcal{D})$. Množinu možných kandidátů této aproximace volíme jako

$$\breve{f}^*(\Theta, z|\mathcal{D}) \equiv \left\{ f(\Theta, z|\mathcal{D}) \middle| \breve{f}(\Theta, z|\mathcal{D}) = \breve{f}(z|\mathcal{D}) \prod_{i=1}^{q} \breve{f}(\Theta_i|\mathcal{D}) \right\}, \tag{1.37}$$

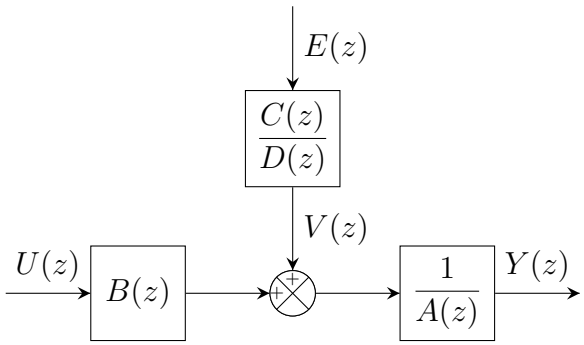
zbývající procedura výpočtu VB marginálií je obdobná k postupu popsanému výše, přičemž k náhodné veličině z přistupujeme stejně jako k Θ_i. Ze získané VB aproximace $\tilde{f}(\Theta, z|\mathcal{D}) = \tilde{f}(\Theta|\mathcal{D})\tilde{f}(z|\mathcal{D})$ použijeme VB marginálie $\tilde{f}(\Theta|\mathcal{D}) = \prod_{i=1}^{q} \tilde{f}(\Theta_i|\mathcal{D})$ pro aproximaci původního posterioru $f(\Theta|\mathcal{D}) \approx \tilde{f}(\Theta|\mathcal{D})$. VB marginálie $\tilde{f}(z|\mathcal{D})$ se dále nepropaguje [2].

2 Bayesovská identifikace časově invariantního normálního regresního modelu

V této kapitole formulujeme pomocí Bayesovských paradigmat jednokrokovou a rekurzivní variantu identifikace normálního regresního modelu.

2.1 Modely s chybou rovnice

Normální regresní model, představující stochastický model ARX (autoregresní model s exogonním výstupem), spadá do rodiny lineárních modelů s chybou rovnice. Modely této rodiny lze obecně popsat schématem 2.1.



Obr. 2.1: Obecný model s chybou rovnice [5]

Přenos systému na obrázku 2.1 je

$$Y(z) = \begin{bmatrix} \frac{B(z)}{A(z)} & \frac{C(z)}{D(z)A(z)} \end{bmatrix} \begin{bmatrix} U(z) \\ E(z) \end{bmatrix} \tag{2.1}$$

kde polynomy $A(z), B(z), C(z), D(z)$ jsou tvořeny ve formátu

$$A(z) \equiv 1 + a_1 z^{-1} + ... + a_{n_a} z^{-n_a}, \tag{2.2a}$$

$$B(z) \equiv b_0 + b_1 z^{-1} + ... + b_{n_b} z^{-n_b}, \tag{2.2b}$$

$$C(z) \equiv 1 + c_1 z^{-1} + ... + c_{n_c} z^{-n_c}, \tag{2.2c}$$

$$D(z) \equiv 1 + d_1 z^{-1} + ... + d_{n_d} z^{-n_d}, \tag{2.2d}$$

kde $n_a, n_b, n_c, n_d \in \mathbb{N}$ jsou stupně polynomů. Bez újmy na obecnosti uvádíme polynom $B(z)$ včetně nultého členu b_0, který však u reálných dynamických systémů zaniká. Absence tohoto členu odráží skutečnost, že vstup v kroku k se na výstupu projeví nejdříve v následujícím kroku $k+1$. Toto odstranění přímé vazby mezi vstupem a výstupem je implikováno konečným rozsahem přenášených frekvencí všech reálných dynamických systémů [5].

Neměřitelné poruše působící na model přisuzujeme vlasnosti bílého šumu popsaného normálním rozložením

$$e_k \sim \mathcal{N}_{e_k}(0, 1/d) \tag{2.3}$$

o nulové střední hodnotě a neznáme varianci d^{-1}. Obraz tohoto signálu v z oblasti označujeme $E(z)$. Obdobně $Y(z)$ je obrazem výstupu y_k, $U(z)$ obrazem vstupu u_k a $V(z)$ obrazem barevného šumu v_k obarveného stochastickou částí $C(z)/D(z)$ [5].

Diferenční rovnici tohoto modelu popisují rovnice 2.4, 2.5.

$$y_k = \sum_{i=1}^{n_b} b_i u_{k-i} - \sum_{i=1}^{n_a} a_i y_{k-i} + v_k \tag{2.4}$$

$$v_k = e_k + \sum_{i=1}^{n_c} c_i e_{k-i} - \sum_{i=1}^{n_d} d_i v_{k-i} \tag{2.5}$$

Při uvažování $n_a, n_b, n_c, n_d \in \mathbb{N}_{>0}$ nazýváme model na obrázku 2.1 ARARMAX. Z této zkratky první dvě dvojice písmen AR (*Auto-Regression*) odpovídají autoregresivní deterministické části $1/A(z)$ a stochastické části $1/D(z)$. Písmena MA (*Moving Average*) odkazují na přítomnost čitatelového polynomu $C(z)$ ve stochastické části modelu. Exogenitu vstupu vyjadřuje ve zkratce písmeno X (*eXogonnous*) [5, 6].

2.2 Normální regresní model ARX

Struktura ARARMAX je mnohdy nadbytečně složitá. Algoritmy pro identifikaci koeficientů všech polynomiálních funkcí $A(z), B(z), C(z), D(z)$ skýtají různá úskalí a omezení. Zjednodušení dosáhneme položením některých z polynomů stochastické části rovným jedné. Získáme tak nové struktury jako ARMAX ($n_a, n_b, n_c \in \mathbb{N}_{>0}$, $n_d = 0$), ARARX ($n_a, n_b, n_d \in \mathbb{N}_{>0}$, $n_c = 0$) či ARX ($n_a, n_b \in \mathbb{N}_{>0}, n_c = n_d = 0$).

Pro ARX model se rovnice 2.5 redukuje na $v_k = e_k$. Šum vstupující do deterministické části zůstává bílý a diferenční rovnice pro ARX model bude dána vztahem

$$y_k = \sum_{i=1}^{n_b} b_i u_{k-i} - \sum_{i=1}^{n_a} a_i y_{k-i} + e_k, \tag{2.6}$$

jehož deterministickou část lze vektorem parametrů $\theta \equiv [b_1, ...b_{n_b}, a_1, ...a_{n_a}]' \in \mathbb{R}^p$, $p = n_a + n_b$ a regresním vektorem $h_k \equiv [u_{k-1}, ..., u_{k-n_b}, -y_{k-1}, ..., -y_{k-n_a}]' \in \mathbb{R}^p$ vyjádřit kompaktně [6]

$$y_k = h_k'\theta + e_k. \tag{2.7}$$

2.3 Bayesovská identifikace časově invariantního normálního regresního modelu

Podle věty o transformaci náhodné veličiny [3] můžeme na základě 2.3, 2.7 vyjádřit y_k funkcí hustoty pravděpodobnosti

$$f(y_k|h_k,\theta,d) = \mathcal{N}_{y_k}\left(h_k'\theta, 1/d\right) \tag{2.8}$$

podmíněnou znalostí měřených dat h_k a parametry θ, d. Předpokládejme znalost vstupně-výstupních datových párů od kroku $1-n$ do kroku $k \in \mathbb{N}$ datové množiny $\mathcal{D}_{1-n}^{k} \equiv \{y_i, u_i\}_{i=1-n}^{k}$, kde $n = \max(n_a, n_b)$ [8].

Volba této indexace formálně zajišťuje $\forall k \geq 1$ dostupnost datové informace pro vyjádření datově informovaného posterioru $f(\theta, d|\mathcal{D}_{1-n}^{k})$. Vzhledem ke znalosti dat pouze od kroku $1-n$ nemáme úplnou datovou informaci pro podmínění $f(y_k|h_k,\theta,d)$ již v kroku $k = 0$. Pro nultý krok je tak posterior datově neinformovaný a je určen apriorní znalostí $f(\theta, d|\mathcal{D}_{1-n}^{0}) \equiv f(\theta, d)$.

2.3.1 Dávková varianta

Cílem je vyjádřit fhp hledaných veličin θ a d na základě historie známých vstupů a výstupů, jež jsou uchovávány v $\mathcal{D}_{1-n}^{k}$. Aplikací Bayesova vzorce získáme

$$f(\theta, d|\mathcal{D}_{1-n}^{k}) \propto f(\mathcal{D}_{1-n}^{k}|\theta, d) f(\theta, d). \tag{2.9}$$

Fhp $f(\mathcal{D}_{1-n}^{k}|\theta, d)$ lze pomocí řetězového pravidla rozepsat jako

$$f(\mathcal{D}_{1-n}^{k}|\theta, d) = \prod_{i=1}^{k} \left\{ f(y_i|u_i, \mathcal{D}_{1-n}^{i-1}, \theta, d) f(u_i|\mathcal{D}_{1-n}^{i-1}, \theta, d) \right\} f(\mathcal{D}_{1-n}^{0}|\theta, d). \tag{2.10}$$

S odvoláním na přirozenou podmínku řízení [1]

$$f(u_i|\mathcal{D}_{1-n}^{i-1}, \theta, d) = f(u_i|\mathcal{D}_{1-n}^{i-1}) \tag{2.11}$$

a aplikací Bayesova vzorce můžeme vztah 2.9 upravit na

$$\begin{aligned} f(\theta, d|\mathcal{D}_{1-n}^{k}) &\propto \prod_{i=1}^{k} \left\{ f(y_i|u_i, \mathcal{D}_{1-n}^{i-1}, \theta, d) \right\} f(\mathcal{D}_{1-n}^{0}|\theta, d) f(\theta, d) \\ &= \prod_{i=1}^{k} \left\{ f(y_i|h_i, \theta, d) \right\} f(\mathcal{D}_{1-n}^{0}|\theta, d) f(\theta, d) \\ &\propto \prod_{i=1}^{k} \left\{ f(y_i|h_i, \theta, d) \right\} f(\theta, d). \end{aligned} \tag{2.12}$$

[1] Přirozená podmínka řízení říká, že vstup systému je podmíněný pouze daty, nikoliv parametry systému, tedy $f(u_k|\Theta, \mathcal{D}_{1-n}^{k-1}) = f(u_k|\mathcal{D}_{1-n}^{k-1})$. Tato podmínka plyne z předpokladu kauzality systému a je mimo evidentních případů identifikace obecně platná pro systémy, kde $f(\theta, d|u_k, \mathcal{D}_{1-n}^{k-1}) = f(\theta, d|\mathcal{D}_{1-n}^{k-1})$[3, 4].

Pro zajištění konjugace budeme volit prior $f(\theta, d)$ z rodiny dynamických exponenciálních rozložení. Parametry ARX modelu θ budeme modelovat normálním rozložením a přesnost šumu d Wishartovým rozložením.

$$f(\theta, d) = \mathcal{N}_\theta\left(\hat{\theta}_0, P_0/d\right)\mathcal{W}_d\left(\Sigma_0, \nu_0\right) \tag{2.13}$$

Nyní jsme již schopni vyjádřit všechny pravostranné fhp. Posterior $f(\theta, d|\mathcal{D}_{1-n}^k)$ bude mít funkcionální formu normálního-Wishartova ($\mathcal{NW}$) rozložení.

$$\begin{aligned} f(\theta, d|\mathcal{D}_{1-n}^k) &= \mathcal{N}_\theta\left(\hat{\theta}_k, P_k/d\right)\mathcal{W}_d\left(\Sigma_k, \nu_k\right) \\ &\propto \prod_{i=1}^{k}\left\{\mathcal{N}_{y_k}\left(h_k'\theta_k, 1/d\right)\right\}\mathcal{N}_\theta\left(\hat{\theta}_0, P_0/d\right)\mathcal{W}_d\left(\Sigma_0, \nu_0\right) \end{aligned} \tag{2.14}$$

Zavedením regresní matice $H_k \equiv [h_1, h_2, ..., h_k]'$ a vektoru výstupů $Y_k \equiv [y_1, y_2, ..., y_k]'$ můžeme zjednodušit zápis v 2.14 na

$$\prod_{i=1}^{k}\left\{\mathcal{N}\left(h_k'\theta_k, 1/d\right)\right\} = (2\pi/d)^{-k/2}\exp\left\{-\frac{d}{2}\left(H_k\theta - Y_k\right)'\left(H_k\theta - Y_k\right)\right\}. \tag{2.15}$$

Dosazením 2.15 do 2.14 a porovnáním pravé a levé strany

$$\begin{aligned} &\mathcal{N}_\theta\left(\hat{\theta}_k, P_k/d\right)\mathcal{W}_d\left(\Sigma_k, \nu_k\right) \\ &\quad\propto\ d^{p/2}\exp\left\{-\frac{d}{2}\left(\theta - \hat{\theta}_k\right)'P_k^{-1}\left(\theta - \hat{\theta}_k\right)\right\}d^{(\nu_k-2)/2}\exp\left\{-d\Sigma_k/2\right\} \\ &\quad=\ d^{k/2}\exp\left\{-\frac{d}{2}\left(H_k\theta - Y_k\right)'\left(H_k\theta - Y_k\right)\right\} \\ &\quad\times d^{p/2}\exp\left\{-\frac{d}{2}\left(\theta - \hat{\theta}_0\right)'P_0^{-1}\left(\theta - \hat{\theta}_0\right)\right\}d^{(\nu_0-2)/2}\exp\left\{-d\Sigma_0/2\right\} \end{aligned} \tag{2.16}$$

získáme postačující statistiky posterioru. Je patrné, že $\nu_k = \nu_0 + k$. Doplněním na úplný čtverec výrazů v exponentech o přirozeném základu e v 2.16 získáme zbylé postačující statistiky posterioru [6].

$$\begin{bmatrix} P_k^{-1} & -P_k^{-1}\hat{\theta}_k \\ -\hat{\theta}_k'P_k^{-1} & \hat{\theta}_k'P_k^{-1}\hat{\theta}_k + \Sigma_k \end{bmatrix} = \begin{bmatrix} P_0^{-1} + H_k'H_k & -P_0^{-1}\hat{\theta}_0 - H_k'Y_k \\ -\hat{\theta}_0'P_0^{-1} - Y_k'H_k & \hat{\theta}_0'P_0^{-1}\hat{\theta}_0 + Y_k'Y_k + \Sigma_0 \end{bmatrix} \tag{2.17}$$

Z rovnice 2.17 plyne, že $P_k = (P_0^{-1} + H_k'H_k)^{-1}$, $\hat{\theta}_k = P_k^{-1}(\hat{\theta}_0'P_0^{-1} + Y_k'H_k)$, $\Sigma_k = \Sigma_0 + Y_k'Y_k + \hat{\theta}_0'P_0^{-1}\hat{\theta}_0 - \hat{\theta}_k'P_k^{-1}\hat{\theta}_k$ a $\nu_k = \nu_0 + k$.

Můžeme si povšimnout, že při volbě neinformativního prioru $f(\theta, d)$ bude kovarianční matice $P_k = H_k'H_k$ a střední hodnota parametrů $\hat{\theta}_k = (H_k'H_k)^{-1}Y_k'H_k$, což je výsledek dávkové metody nejmenších čtverců. Zde vidíme důkaz možnosti formulace v praxi osvědčeného algoritmu pomocí bayesovského přístupu, jak bylo předesláno v úvodu kapitoly 1 [6].

2.3.2 Rekurzivní varianta

Cílem je nalézt funkcionální rekurzi pro aktualizaci fhp hledaných veličin θ a d v okamžiku dostupnosti nového měření $\{y_k, u_k\}$. Vyjdeme ze vztahu 2.9, jehož pravou stranu můžeme užitím řetězového pravidla a Bayesova vzorce rozepsat a upravit [3].

$$\begin{aligned} f(\theta, d|\mathcal{D}_{1-n}^{k}) &\propto f(y_k, u_k|\mathcal{D}_{1-n}^{k-1}, \theta, d) f(\mathcal{D}_{1-n}^{k-1}|\theta, d) f(\theta, d) \\ &\propto f(y_k, u_k|\mathcal{D}_{1-n}^{k-1}, \theta, d) f(\theta, d|\mathcal{D}_{1-n}^{k-1}) \end{aligned} \tag{2.18}$$

Další aplikací řetězového pravidla a využitím zmíněné přirozené podmínky řízení dále zjednodušíme zápis 2.18 na

$$\begin{aligned} f(\theta, d|\mathcal{D}_{1-n}^{k}) &\propto f(y_k|u_k, \mathcal{D}_{1-n}^{k-1}, \theta, d) f(\theta, d|\mathcal{D}_{1-n}^{k-1})) \\ &= f(y_k|h_k, \theta, d) f(\theta, d|\mathcal{D}_{1-n}^{k-1}). \end{aligned} \tag{2.19}$$

Jak plyne z 2.19, aktualizace zastaralého posterioru $f(\theta, d|\mathcal{D}_{1-n}^{k-1})$ je prováděna novým měřením, jež představuje fhp $f(y_k|h_k, \theta, d)$. Je zřejmé, že tato funkcionální rekurze je inicializována priorem $f(\theta, d|\mathcal{D}_{1-n}^{0}) \equiv f(\theta, d)$. V souladu s nerekurzivní variantou volíme konjungovaný prior z rodiny dynamických exponenciálních rozložení dle 2.13. Posterior tak bude mít rovněž funkcionální formu $\mathcal{NW}$ rozložení [3].

$$f(\theta, d|\mathcal{D}_{1-n}^{k}) = \mathcal{N}_\theta\left(\hat{\theta}_k, P_k/d\right) \mathcal{W}_d\left(\Sigma_k, \nu_k\right) \tag{2.20}$$

Stejně jako u dávkové varianty nahraďme funkcionální rekurzi algebraickým vyjádřením posteriorních postačujících statistik. Z algebraického vyjádření 2.21 funkcionální formy 2.20 plyne, že $\nu_k = \nu_{k-1} + 1$.

$$\begin{aligned} &\mathcal{N}_\theta\left(\hat{\theta}_k, P_k/d\right)\mathcal{W}_d\left(\Sigma_k, \nu_k\right) \\ &\quad\propto\ d^{p/2} \exp\left\{ -\frac{d}{2}\left(\theta - \hat{\theta}_k\right)' P_k^{-1}\left(\theta - \hat{\theta}_k\right)\right\} d^{(\nu_k - 2)/2} \exp\left\{ -d\Sigma_k/2\right\} \\ &\quad=\ d^{1/2} \exp\left\{ -\frac{d}{2}\left(h_k'\theta - y_k\right)^2\right\} \\ &\qquad\times d^{p/2} \exp\left\{ -\frac{d}{2}\left(\theta - \hat{\theta}_{k-1}\right)' P_{k-1}^{-1}\left(\theta - \hat{\theta}_{k-1}\right)\right\} d^{(\nu_{k-1} - 2)/2} \exp\left\{ -d\Sigma_{k-1}/2\right\} \end{aligned} \tag{2.21}$$

Vyjádření zbylých postačujících statistik provedeme opět doplněním na úplný čtverec výrazů v exponentech o základu e.

$$\begin{bmatrix} P_k^{-1} & -P_k^{-1}\hat{\theta}_k \\ -\hat{\theta}_k' P_k^{-1} & \hat{\theta}_k' P_k^{-1}\hat{\theta}_k + \Sigma_k \end{bmatrix} = \begin{bmatrix} P_{k-1}^{-1} + h_k h_k' & -P_{k-1}^{-1}\hat{\theta}_{k-1} - h_k y_k \\ -\hat{\theta}_{k-1}' P_{k-1}^{-1} - h_k' y_k & \hat{\theta}_{k-1}' P_{k-1}^{-1}\hat{\theta}_{k-1} + y_k^2 + \Sigma_{k-1} \end{bmatrix} \tag{2.22}$$

Inverzi matice $(P_{k-1}^{-1} + h_k h_k')$ nutnou pro výpočet P_k můžeme užitím Woodburovy identity 2.23 [6, 7]

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1} \tag{2.23}$$

a zavedením substitucí $A = P_{k-1}, U = h_k, C = I_p, V = h'_k$ zjednodušit na

$$P_k = P_{k-1} - \frac{P_{k-1}h_k h'_k P_{k-1}}{1 + h'_k P_{k-1} h_k}. \tag{2.24}$$

Vidíme, že toto vyjádření nahrazuje inverzi matice výpočetně levnější operací - skalárním dělením. Zavedením vektoru K_k dle 2.25a, který lze ekvivalentě vyjádřit jako $K_k = P_k h_k$, můžeme zápis 2.23 matice P_k zjednodušit na $P_k = (I_p - K_k h'_k)P_{k-1}$, kde I_p je jednotková matice o rozměru $p \times p$. Tento tvar je dále možné užitím Josephovy formy převést na kvadratickou formu 2.25c zajišťující symetričnost a pozitivní definitnost P_k [6].

$$K_k \equiv P_{k-1}h_k / \left(1 + h'_k P_{k-1} h_k\right), \tag{2.25a}$$

$$\hat{e}_k \equiv y_k - h'_k \hat{\theta}_{k-1}, \tag{2.25b}$$

$$P_k = (I_p - K_k h'_k)P_{k-1}(I_p - K_k h'_k)' + K_k K'_k \tag{2.25c}$$

$$\hat{\theta}_k = \hat{\theta}_{k-1} + K_k \hat{e}_k \tag{2.25d}$$

$$\Sigma_k = \Sigma_{k-1} + \hat{e}_k^2/(1 + h'_k P_{k-1} h_k), \tag{2.25e}$$

$$\nu_k = \nu_{k-1} + 1 \tag{2.25f}$$

Statistika $\hat{\theta}_k$ je pak dána 2.25d, kde $\hat{e}_k$ představuje chybu predikce modelu. Vyjádřením $\Sigma_k = \Sigma_{k-1} - \hat{\theta}'_k P_k^{-1} \hat{\theta}_k + \hat{\theta}'_{k-1} P_{k-1}^{-1} \hat{\theta}_{k-1} + y_k^2$, dosazením za $P_k, \hat{\theta}_k, \hat{e}_k$, užitím výše uvedených rovností a úpravami lze dosáhnout 2.25e [8].

3 Bayesovská identifikace časově variantního normálního regresního modelu

Rekurzivní algoritmizace popsaná v kapitole 2 předpokládá, že skutečná hodnota identifikovaných statistik Θ se s časem nemění. V průběhu identifikace tímto algoritmem je navyšována důvěra v odhady $\hat{\theta}$ a $\hat{d}$ a s každým dalším krokem se tak snižuje schopnost adaptace algoritmu na potenciální změnu systému. Algoritmus je tím pádem pro průběžnou identifikaci časově variantních systémů nevhodný.

V praxi se ale často potýkáme se systémy, jejichž parametry $\Theta_k \equiv \{\theta_k, d_k\}$ jsou časově variantní a jejichž časový vývoj chceme průběžně sledovat. Neznámost časového vývoje parametrů nám však neumožňuje získat exaktní řešení časového kroku. Jedním z řešení tohoto problému je sestavení algoritmu, který v průběhu identifikace bude akumulovanou informaci uměle snižovat tak, aby byl algoritmus schopen adaptace na změny. Tato technika je nazývána zapomínání [8].

3.1 Regularizovaná identifikace časově variantního normálního modelu s exponenciálním zapomínáním

Algoritmů se zapomínáním pro rekurzivní identifikaci je celá řada, jejichž hlavním rozdílem je strategie při určování koeficientu zapomínání. Je pochopitelné, že triviální techniky, jako je například konstantní exponenciální zapomínání, mají celou řadu omezení. Představme nyní algoritmus pro regularizovanou estimaci parametrů normálního regresního modelu a diskuzi určování koeficientu zapomínání dvojicí datově informovaných technik si popišme až v kapitolách 3.2 a 3.3.

Tento algoritmus nehledě na volbu koeficientu zapomínání je možné rozdělit na časový a datový krok, jako je tomu například u Kalmanova filtru [3].

3.1.1 Časový krok

Časovým krokem rozumíme predikci estimovaných parametrů platných pro čas k na základě znalosti měřených dat jen do $k-1$ a zastaralého posterioru, který má v našem případě tvar

$$\begin{aligned} f\left(\Theta_{k-1}|\mathcal{D}_{1-n}^{k-1}, \Pi_{k-1}\right) = \mathcal{N}_{\theta_{k-1}}\left(\hat{\theta}_{k-1|k-1}, P_{k-1|k-1}/d_{k-1}\right) \\ \times \mathcal{W}_{d_{k-1}}\left(\Sigma_{k-1|k-1}, \nu_{k-1|k-1}\right). \end{aligned} \tag{3.1}$$

Uskutečňuje tedy aktualizaci $f(\Theta_{k-1}|\mathcal{D}_{1-n}^{k-1}, \Pi_{k-1}) \rightarrow f(\Theta_k|\mathcal{D}_{1-n}^{k-1}, \Pi_{k-1})$. Podmíněnost výrazem Π_{k-1} odkazuje na regularizační statistiky, které budou diskutovány v kapitole 3.1.2.

Model časového vývoje parametrů $f(\Theta_k|\Theta_{k-1}, \mathcal{D}_{1-n}^{k-1}, \Pi_{k-1})$ prostřednictvím 3.2 přirozeně umožňuje získat časovou predikci. Pokud známe tuto fhp a marginalizace 3.2 je řešitelná, časový krok představuje triviální problém.

$$f\left(\Theta_k|\mathcal{D}_{1-n}^{k-1}, \Pi_{k-1}\right) = \int_{\Theta^*} f\left(\Theta_k|\Theta_{k-1}, \mathcal{D}_{1-n}^{k-1}, \Pi_{k-1}\right) f\left(\Theta_{k-1}|\mathcal{D}_{1-n}^{k-1}, \Pi_{k-1}\right) \mathrm{d}\Theta_{k-1} \tag{3.2}$$

Při absenci informace o časovém vývoji $f(\Theta_k|\Theta_{k-1}, \mathcal{D}_{1-n}^{k-1}, \Pi_{k-1})$ nemá 3.2 řešení a pro časový vývoj parametrů je zapotřebí použít jinou techniku [8].

Alternativy časové predikce

Předem podotkněme, že po zbytek této kapitoly 3.1.1 bude pro účely zpřehlednění značení na několika místech opomenuta podmíněnost náhodné veličiny Θ_k členy $\mathcal{D}_{1-n}^{k-1}$, Π_{k-1}.

Absence exaktního řešení časového kroku je možné nahradit kombinováním dvou alternativních hypotéz $H \in \{1, 2\}$ a získat odhad co nejlépe vystihující skutečný vývoj parametrů. Pro docílení konjugace přiřazujeme těmto hypotézám formy

$$f_1(\Theta_k) \equiv \mathcal{N}_{\theta_k}\left(\hat{\theta}_{k-1|k-1} P_{k-1|k-1}/d_k\right) \mathcal{W}_{d_k}\left(\Sigma_{k-1|k-1}\nu_{k-1|k-1}\right), \tag{3.3a}$$

$$f_2(\Theta_k) \equiv \mathcal{N}_{\theta_k}\left(\hat{\theta}_{k-1|k-1}, P_{k-1|k-1}/(d_k\alpha)\right) \mathcal{W}_{d_k}\left(\alpha\Sigma_{k-1|k-1}, \alpha\nu_{k-1|k-1}\right). \tag{3.3b}$$

První hypotéza $f_1(\Theta_k)$ má shodné postačující statistiky se zastaralým posteriorem 3.1 a předpokládá invarianci parametrů. Hypotéza $f_2(\Theta_k)$ představuje maximální možné zapomenutí informace uchovávané ve statistikách $P_{k-1|k-1}, \Sigma_{k-1|k-1}, \nu_{k-1|k-1}$ zploštěním zastaralého posterioru volitelným koeficientem $\alpha \in (0, 1)$. Pravděpodobnosti příslušných hypotéz budou popsány vektorem vah $\varphi \equiv [\varphi_1, \varphi_2]'$ pomocí $f(H|\varphi)$ s vlastností $f(H = i|\varphi) = \varphi_i$ pro $i \in \{1, 2\}$. Z pravděpodobnostního charakteru φ zřejmě platí $\varphi_i \geq 0$ a $\varphi_1 + \varphi_2 = 1$ [8, 10].

Výběr optimálního kompromisu

Pro určení optimální volby $\tilde{f}(\Theta_k)$, kombinující obě alternativy $f_1(\Theta_k)$, $f_2(\Theta_k)$ do jediné hustoty, zavedeme konvexní ztrátový funkcionál [8]

$$\begin{aligned} J\left(\breve{f}(\Theta_k), \varphi\right) &\equiv \mathsf{E}_{f(H|\varphi)}\left[KL\left(\breve{f}(\Theta_k)\middle\| f_H(\Theta_k)\right)\right] \\ &= \sum_{i=1}^{2} \varphi_i \left[KL\left(\breve{f}(\Theta_k)\middle\| f_i(\Theta_k)\right)\right]. \end{aligned} \tag{3.4}$$

Značení $\breve{f}(\Theta_k)$ má v souladu s kapitolou 1.5 význam distribuce představující potenciálního kandidáta na aproximaci exaktního řešení 3.2.

Úpravami funkcionálu $J(\breve{f}(\Theta_k), \varphi)$ obdržíme

$$\begin{aligned} J\Big(\breve{f}(\Theta_k), \varphi\Big) &= \int_{\Theta^*} \breve{f}(\Theta_k) \left(\varphi_1 \ln\left(\frac{\breve{f}(\Theta_k)}{f_1(\Theta_k)}\right) + \varphi_2 \ln\left(\frac{\breve{f}(\Theta_k)}{f_2(\Theta_k)}\right)\right) \mathrm{d}\Theta_k \\ &= KL\left(\breve{f}(\Theta_k) \middle\| \frac{f_1^{\varphi_1}(\Theta_k) f_2^{\varphi_2}(\Theta_k)}{\int_{\Theta^*} f_1^{\varphi_1}(\Theta_k) f_2^{\varphi_2}(\Theta_k) \mathrm{d}\Theta_k}\right) \\ &\quad - \ln\left(\int_{\Theta^*} f_1^{\varphi_1}(\Theta_k) f_2^{\varphi_2}(\Theta_k) \mathrm{d}\Theta_k\right). \end{aligned} \tag{3.5}$$

Vidíme, že optimum vzhledem k volbě $\breve{f}(\Theta_k)$ minimalizující $J(\breve{f}(\Theta_k), \varphi)$ je dosaženo vektorem φ váženým geometrickým průměrem 3.6 hypotéz $f_1(\Theta_k)$ a $f_2(\Theta_k)$.

$$\tilde{f}\left(\Theta_k | \mathcal{D}_{1-n}^{k-1}, \Pi_{k-1}, \varphi\right) = \frac{f_1^{\varphi_1}(\Theta_k) f_2^{\varphi_2}(\Theta_k)}{\int_{\Theta^*} f_1^{\varphi_1}(\Theta_k) f_2^{\varphi_2}(\Theta_k) \mathrm{d}\Theta_k} \tag{3.6}$$

Koeficient zapomínání

Užitím tvaru optima 3.6 a formulací hypotéz 3.3a, 3.3b je možné zavedením

$$\lambda_k \equiv \varphi_1(1-\alpha) + \alpha \tag{3.7}$$

vyjádřit $\tilde{f}(\Theta_k)$ ekvivalentně i jako

$$\begin{aligned} \tilde{f}\left(\Theta_k | \mathcal{D}_{1-n}^{k-1}, \Pi_{k-1}, \lambda_k\right) &= \frac{f_1^{\lambda_k}(\Theta_k) d_k^{(p-2)(1-\lambda_k)/2}}{\int_{\Theta^*} f_1^{\lambda_k}(\Theta_k) d_k^{(p-2)(1-\lambda_k)/2} \mathrm{d}\Theta_k} \\ &= \mathcal{N}_{\theta_k}\left(\hat{\theta}_{k|k-1}, P_{k|k-1}/d_k\right) \mathcal{W}_{d_k}\left(\Sigma_{k|k-1}, \nu_{k|k-1}\right). \end{aligned} \tag{3.8}$$

Je patrné, že parametrem $\lambda_k \in [\alpha, 1]$ zvaným koeficient zapomínání je zplošťována hustota $f_1(\Theta_k)$, jejíž postačující statistiky jsou shodné se zastaralým posteriorem. Algebraická vyjádření postačujících statistik po provedení zapomenutí odhadem $\hat{\lambda}_k$ jsou

$$P_{k|k-1} = P_{k-1|k-1}/\lambda_{k|k-1}, \tag{3.9a}$$
$$V_{k|k-1} = V_{k-1|k-1}\lambda_{k|k-1}, \tag{3.9b}$$
$$\hat{\theta}_{k|k-1} = \hat{\theta}_{k-1|k-1}, \tag{3.9c}$$
$$\Sigma_{k|k-1} = \Sigma_{k-1|k-1}\lambda_{k|k-1}, \tag{3.9d}$$
$$\nu_{k|k-1} = \nu_{k-1|k-1}\lambda_{k|k-1}, \tag{3.9e}$$
$$\Xi_{k|k-1} = \Xi_{k-1|k-1}\lambda_{k|k-1}. \tag{3.9f}$$

Jelikož budeme často pracovat s inverzí matice $P_{k|l}$, zavádíme $V_{k|l} \equiv P_{k|l}^{-1}$, kde $l \in \{k-1, k\}$, a to nejen pro přehlednější značení, ale také pro snížení počtu počítaných inverzí maticí nutných pro provedení jednoho kroku algoritmu [8].

Je zřejmé, že obecné momenty prvního řádu náhodných veličin θ, d jsou časovým krokem nezměněny. Zapomenutí 3.9f regularizační matice bude diskutováno záhy v následující kapitole 3.1.2.

Vzhledem k neznalosti časového vývoje parametrů nedokážeme určit exaktní hodnotu koeficientu zapomínání λ_k. Obě dále uvedené algoritmizace nahrazují znalost λ_k odhadem $\hat{\lambda}_k$. V konečném důsledku to tedy znamená, že namísto neznámé λ_k použijeme pro aktualizaci posterioru $\hat{\lambda}_k$ a všechny výrazy λ_k v kapitole 3.1 tímto odhadem nahradíme. Tato formální náhrada však zde zakomponována není, neboť algoritmus popsaný v kapitole 3.2 pracuje s λ_k jako s náhodnou veličinou a touto přímou náhradou λ_k bodovým odhadem $\hat{\lambda}_k$ by se náhodnost ze zápisů v této kapitole 3.1 vytratila.

3.1.2 Regularizace

Při nedostatečně pestré datové informaci v delším časovém horizontu se může matice $P_{k|k-1}$ přiblížit k singulární matici natolik, že algoritmus začne být kvůli omezené numerické přesnosti výpočtů nestabilní. Regularizace je zvláště užitečná u vestavěných systémů, kde je důraz často kladen na nízkou cenu, z čehož plynou omezené výpočetní možnosti.

Výraz 3.10 představující regularizaci časové predikce $\tilde{f}(\Theta_k|\mathcal{D}_{1-n}^{k-1}, \Pi_{k-1}, \lambda_k)$ je možné odvodit opakovaným užitím Bayesova vzorce za předpokladu podmíněné nezávislosti parametrického modelu na regularizačních členech $f(\mathcal{D}_{1-n}^{k-1}|\Theta_k, \Pi_k) = f(\mathcal{D}_{1-n}^{k-1}|\Theta_k,)$ [10].

$$f\left(\Theta_k\middle|\mathcal{D}_{1-n}^{k-1}, \Pi_k, \lambda_k\right) \propto \frac{\tilde{f}\left(\Theta_k\middle|\mathcal{D}_{1-n}^{k-1}, \Pi_{k-1}, \lambda_k\right) f\left(\Theta_k\middle|\Pi_k\right)}{f\left(\Theta_k\middle|\Pi_{k-1}\right)} \tag{3.10}$$

Z důvodu přehlednosti již pišme výslednou regularizovanou statistiku v 3.10 bez znaku vlnovky, tedy jen $f(\cdot)$.

Pro konjugaci s posteriorem volíme regularizující $f(\Theta_k|\Pi_k)$ jako $\mathcal{NW}$ rozložení 3.11a, kde $f(\theta_k|d_k, \Pi_k) = \mathcal{N}_{\theta_k}(\hat{\theta}_{k|k-1}, \Xi_{k|k}^{-1}/d_k)$ a $f(d_k|\Pi_k) \equiv f(d_k) = \mathcal{W}_{d_k}(\Sigma_{k|0}, \nu_{k|0})$ [8].

$$f(\Theta_k|\Pi_k) = \mathcal{N}_{\theta_k}\left(\hat{\theta}_{k|k-1}, \Xi_{k|k}^{-1}/d_k\right)\mathcal{W}_{d_k}\left(\Sigma_{k|0}, \nu_{k|0}\right) \tag{3.11a}$$

$$f(\Theta_k|\Pi_{k-1}) = \mathcal{N}_{\theta_k}\left(\hat{\theta}_{k-1|k-2}, \Xi_{k|k-1}^{-1}/d_k\right)\mathcal{W}_{d_k}\left(\Sigma_{k|0}, \nu_{k|0}\right) \tag{3.11b}$$

Algebraickým vyjádřením pravé strany 3.10 pomocí 3.8, 3.11a, 3.11b a doplněním

na úplný čtverec

$$\begin{bmatrix} P^{-1}_{c;k|k-1} & -P^{-1}_{c;k|k-1}\hat{\theta}_{c;k|k-1} \\ -\hat{\theta}'_{c;k|k-1}P^{-1}_{c;k|k-1} & \hat{\theta}'_{c;k|k-1}P^{-1}_{c;k|k-1}\hat{\theta}_{c;k|k-1} + \Sigma_{c;k|k-1} \end{bmatrix}$$
$$= \begin{bmatrix} P^{-1}_{k|k-1} + \Xi_{k|k} - \Xi_{k|k-1} & \underbrace{-P^{-1}_{k|k-1}\hat{\theta}_{k|k-1} - \Xi_{k|k}\hat{\theta}_{k|k-1} + \Xi_{k|k-1}\hat{\theta}_{k-1|k-2}}_{U_\theta} \\ U'_\theta & U_\Sigma \end{bmatrix} \tag{3.12}$$
$$U_\Sigma = \hat{\theta}'_{k|k-1}P^{-1}_{k|k-1}\hat{\theta}_{k|k-1} + \Sigma_{k|k-1} + \hat{\theta}'_{k|k-1}\Xi_{k|k}\hat{\theta}_{k|k-1} - \hat{\theta}'_{k-1|k-2}\Xi_{k|k-1}\hat{\theta}_{k-1|k-2}$$

získáme postačující statistiky

$$P^{-1}_{c;k|k-1} = P^{-1}_{k|k-1} + \Xi_{k|k} - \Xi_{k|k-1}, \tag{3.13a}$$
$$V_{c;k|k-1} = V_{k|k-1} + \Xi_{k|k} - \Xi_{k|k-1}, \tag{3.13b}$$
$$\varepsilon_k \equiv \hat{\theta}_{k|k-1} - \hat{\theta}_{k-1|k-2}, \tag{3.13c}$$
$$\hat{\theta}_{c;k|k-1} = \hat{\theta}_{k|k-1} - P_{c;k|k-1}\Xi_{k|k-1}\varepsilon_k, \tag{3.13d}$$
$$\Sigma_{c;k|k-1} = \Sigma_{k|k-1} - \varepsilon'_k\left(\Xi_{k|k-1} + \Xi_{k|k-1}P_{c;k|k-1}\Xi_{k|k-1}\right)\varepsilon_k, \tag{3.13e}$$

regularizované fhp $f(\Theta_k|\mathcal{D}^{k-1}_{1-n}, \Pi_k, \lambda_k) = \mathcal{N}_{\theta_k}(\theta_{c;k|k-1}, P_{c;k|k-1}/d_k)\mathcal{W}_{d_k}(\Sigma_{c;k|k-1}, \nu_{k|k-1})$.

Provedením časového kroku je škálováním zastaralého posterioru koeficientem λ_k snížena i hodnota regularizační statistiky $\Xi_{k-1|k-1}$, jak uvádí 3.9f. Pro zajištění konzistentního vlivu regularizace při běhu algoritmu je pozůstatek regularizační matice po časovém kroku $\Xi_{k|k-1} = \lambda_k\Xi_{k-1|k-1}$ od matice $V_{k|k-1}$ odečtena a regularizace je realizována novou maticí $\Xi_{k|k}$, jak popisuje 3.13b. Tento algoritmus tak zamezuje neomezenému nárůstu vlastních čísel matice $P_{k|k}$, respektive poklesu u matice $V_{k|k}$ a zajišťuje regulárnost a tedy invertibilitu těchto matic. Při absenci dodatečné informace je externě dodávaná regularizační matice $\Xi_{k|k}$ často volena jako konstantní diagonální matice [8].

Podotkněme, že regularizaci je možné provádět nikoliv po časovém, ale po datovém kroku. Regularizování po datovém kroku by však ztížilo určování koeficientu zapomínání λ_k popsaného v kapitole 3.3 a vedlo by na iterativní řešení, jak je popsáno v [16].

3.1.3 Datový krok

Datový krok začleňuje informaci měřené hodnoty y_k do regularizované časové predikce. Za užití Bayesova vzorce, zmíněné podmínky řízení a nezávislosti dat na regularizačních členech lze formulovat datový krok

$$f\left(\Theta_k\middle|\mathcal{D}^k_{1-n}, \Pi_k, \lambda_k\right) \propto f\left(y_k\middle|h_k, \Theta_k\right) f\left(\Theta_k\middle|\mathcal{D}^{k-1}_{1-n}, \Pi_k, \lambda_k\right). \tag{3.14}$$

Využitím algebraického vyjádření fhp $f(y_k|h_k, \Theta_k)$ vztahem 2.8 můžeme snadno provést datový krok 3.14. Problém vyjádření postačujících statistik posterioru je identický k řešení popsanému v kapitole 2.3.2, a proto zde uvedeme pouze výsledky datového kroku [8]

$$K_k \equiv P_{c;k|k-1}h_k/(1 + h_k'P_{c;k|k-1}h_k), \tag{3.15a}$$

$$\hat{e}_{c;k|k-1} \equiv y_k - h_k'\hat{\theta}_{c;k|k-1}, \tag{3.15b}$$

$$P_{k|k} = (I_p - K_k h_k')P_{c;k|k-1}(I_p - K_k h_k')' + K_k K_k', \tag{3.15c}$$

$$V_{k|k} = V_{c;k|k-1} + h_k h_k', \tag{3.15d}$$

$$\hat{\theta}_{k|k} = \hat{\theta}_{c;k|k-1} + K_k\hat{e}_{c;k|k-1}, \tag{3.15e}$$

$$\Sigma_{k|k} = \Sigma_{c;k|k-1} + \hat{e}^2_{c;k|k-1}/(1 + h_k'P_{c;k|k-1}h_k), \tag{3.15f}$$

$$\nu_{k|k} = \nu_{k|k-1} + 1. \tag{3.15g}$$

3.2 Datově informovaná volba koeficientu zapomínání určeným variační bayesovskou aproximací

Tato algoritmizace inspirovaná [15] spočívá v uchopení koeficientu zapomínání λ_k jako náhodné veličiny. Pomocí prioru $f(\lambda_k)$ obdržíme $f(\Theta_k, \lambda_k|\mathcal{D}^k_{1-n}, \Pi_k)$. Neřešitelnou analytickou marginalizaci pro zisk posterioru nepodmíněným koeficientem zapomínání aproximujeme pomocí VB metody, pro níž je zapotřebí vyjádřit sdruženou hustotu $f(\Theta_k, \lambda_k|\mathcal{D}^k_{1-n}, \Pi_k)$. Závislost normalizační konstanty hustoty regularizované časové predikce $f(\Theta_k|\lambda_k, \mathcal{D}^{k-1}_{1-n}, \Pi_k)$ na koeficientu zapomínání λ_k způsobuje, že VB marginálie $\tilde{f}(\lambda_k|\mathcal{D}^{k-1}_{1-n}, \Pi_k)$ nemá standardní formu. Tato normalizační konstanta je aproximována exponenciální funkcí a hustota faktoru zapomínání tak získává formu zkráceného exponenciálního rozložení.

V následujících dvou kapitolách 3.2.1 a 3.2.2 tuto motivační myšlenku realizujeme.

3.2.1 VB aproximace náhodné veličiny Θ_k

Ze zápisu regularizované časové predikce 3.10 a sdružené hustoty

$$f\Big(\Theta_k, \lambda_k, \mathcal{D}^k_{1-n}, \Pi_k\Big) = f\Big(\Theta_k|\lambda_k, \mathcal{D}^{k-1}_{1-n}, \Pi_k\Big)f(y_k|\Theta_k, h_k)f(\mathcal{D}^{k-1}_{1-n})f(\Pi_k) \tag{3.16}$$

je patrné, že $\tilde{f}(\Theta_k|\mathcal{D}_{1-n}^k, \Pi_k)$ bude dána jako

$$\begin{aligned}\tilde{f}\left(\Theta_k\middle|\mathcal{D}_{1-n}^k, \Pi_k\right) &\propto \exp\left\{\mathsf{E}_{\tilde{f}(\lambda_k|\mathcal{D}_{1-n}^k,\Pi_k)}\left[\ln\left(f(\Theta_k|\lambda_k, \mathcal{D}_{1-n}^{k-1}, \Pi_k)\right)\right]\right\} f(y_k|\Theta_k, h_k) \\ &\propto \exp\left\{\mathsf{E}_{\tilde{f}(\lambda_k|\mathcal{D}_{1-n}^k,\Pi_k)}\Big[\ln\left(\tilde{f}(\Theta_k|\lambda_k, \mathcal{D}_{1-n}^{k-1}, \Pi_{k-1})\right)\right. \\ &\quad \left.- \ln\left(f(\Theta_k|\lambda_k, \Pi_{k-1})\right)\Big]\right\} f(\Theta_k|\Pi_k) f(y_k|\Theta_k, h_k).\end{aligned} \tag{3.17}$$

Algebraická vyjádření logaritmů hustot $\tilde{f}(\Theta_k|\lambda_k, \mathcal{D}_{1-n}^{k-1}, \Pi_{k-1})$ a $f(\Theta_k|\lambda_k, \Pi_{k-1})$ jsou vyjma výrazů, jež pro Θ_k představují normalizační konstantu, na λ_k závislá lineárně. Tato závislost je patrná z 3.19 a 3.20a - 3.20c. Pro provedení expektace v 3.17 tak bude zapotřebí počítat jen první obecný moment $\hat{\lambda}_k$. Vidíme, že $\tilde{f}(\Theta_k|\mathcal{D}_{1-n}^k, \Pi_k)$ nabývá $\mathcal{NW}$ funkční formy, jejíž postačující statistiky obdržíme doplněním na úplný čtverec. Mnohem snazší je ale využít vztahů pro provedení časového kroku, regularizace a datového kroku dle 3.9a - 3.9f, 3.13a - 3.13e a 3.15a - 3.15g, kde namísto náhodné veličiny λ_k dosadíme její střední hodnotu $\hat{\lambda}_k$.

3.2.2 VB aproximace náhodné veličiny λ_k

VB marginálii faktoru zapomínání λ_k obdržíme na základě 3.16 dle

$$\tilde{f}\left(\lambda_k\middle|\mathcal{D}_{1-n}^k, \Pi_k\right) \propto \exp\left\{\mathsf{E}_{\tilde{f}(\Theta_k|\mathcal{D}_{1-n}^k,\Pi_k)}\left[\ln\left(f(\Theta_k|\lambda_k, \mathcal{D}_{1-n}^{k-1}, \Pi_k)\right)\right]\right\} f(\lambda_k). \tag{3.18}$$

Pro řešení expektace v 3.18 proveďme algebraické vyjádření hustoty

$$\begin{aligned}f(\Theta_k|\lambda_k, \mathcal{D}_{1-n}^{k-1}, \Pi_k) &= (2\pi/d_k)^{-\frac{p}{2}} \mathrm{Z}_{c;k|k-1}(\lambda_k) d_k^{\frac{\lambda_k \nu_{k-1|k-1}-2}{2}} \\ &\quad \times \exp\left\{-\frac{d_k}{2}\begin{bmatrix}\theta_k' & 1\end{bmatrix}\left[\lambda_k U_1 - \lambda_k U_2 + U_3\right]\begin{bmatrix}\theta_k' & 1\end{bmatrix}' - \Sigma_{k|0}\right\},\end{aligned} \tag{3.19}$$

kde členy U_1, U_2, U_3 jsou matice tvořené postačujícími statistikami

$$U_1 = \begin{bmatrix} P_{k-1|k-1}^{-1} & P_{k-1|k-1}^{-1}\hat{\theta}_{k-1|k-1} \\ \hat{\theta}_{k-1|k-1}' P_{k-1|k-1}^{-1} & \hat{\theta}_{k-1|k-1}' P_{k-1|k-1}^{-1}\hat{\theta}_{k-1|k-1} + \Sigma_{k-1|k-1}\end{bmatrix} \tag{3.20a}$$

$$U_2 = \begin{bmatrix} \Xi_{k-1|k-1} & \Xi_{k-1|k-1}\hat{\theta}_{k-1|k-2} \\ \hat{\theta}_{k-1|k-2}' \Xi_{k-1|k-1} & \hat{\theta}_{k-1|k-2}' \Xi_{k-1|k-1}\hat{\theta}_{k-1|k-2}\end{bmatrix} \tag{3.20b}$$

$$U_3 = \begin{bmatrix} \Xi_{k|k} & \Xi_{k|k}\hat{\theta}_{k|k-1} \\ \hat{\theta}_{k|k-1}' \Xi_{k|k} & \hat{\theta}_{k|k-1}' \Xi_{k|k}\hat{\theta}_{k|k-1} + \Sigma_{k|0}\end{bmatrix} \tag{3.20c}$$

a $\mathrm{Z}_{c;k|k-1}(\lambda_k)$ je část normalizační konstanty pro $f(\Theta_k|\lambda_k, \mathcal{D}_{1-n}^{k-1}, \Pi_k)$ závislá na λ_k.

$$\begin{aligned}\mathrm{Z}_{c;k|k-1}(\lambda_k) &= \left|V_{c;k|k-1}\right|^{\frac{1}{2}} \left(\Sigma_{c;k|k-1}/2\right)^{\frac{\nu_{k|k-1}}{2}} \left(\Gamma(\nu_{k|k-1}/2)\right)^{-1} \\ &= \left|\lambda_k \left(V_{k-1|k-1} - \Xi_{k-1|k-1}\right) + \Xi_{k|k}\right|^{\frac{1}{2}} \left(\Gamma(\lambda_k \nu_{k-1|k-1}/2)\right)^{-1} \\ &\times \left(\frac{\lambda_k}{2}\left(\Sigma_{k-1|k-1} - \epsilon_k' U_4 \epsilon_k\right)\right)^{\frac{\lambda_k \nu_{k-1|k-1}}{2}}\end{aligned} \tag{3.21}$$

kde

$$U_4 = \Xi_{k-1|k-1}\left(I_p + \left(\lambda_k\left(V_{k-1|k-1} - \Xi_{k-1|k-1}\right) + \Xi_{k|k}\right)^{-1} \lambda_k \Xi_{k-1|k-1}\right). \tag{3.22}$$

Vzhledem k tvaru $\mathrm{Z}_{c;k|k-1}(\lambda_k)$ je patrné, že exaktní řešení aproximace 3.18 nevede na rozložení standardní formy. Zavedeme tedy, obdobně jako v [15], aproximaci této normalizační konstanty

$$\mathrm{Z}_{c;k|k-1}(\lambda_k) \approx \tilde{\mathrm{Z}}_{c;k|k-1}(\lambda_k) \equiv \exp\left\{g_1 + \lambda_k g_2\right\}, \tag{3.23}$$

která se bude shodovat se skutečnou $\mathrm{Z}_{c;k|k-1}(\lambda_k)$ v krajních bodech definičního oboru λ_k. Řešením soustavy dvou rovnic $\mathrm{Z}_{c;k|k-1}(\lambda_k = i) = \tilde{\mathrm{Z}}_{c;k|k-1}(\lambda_k = i)$ pro $i \in \{\alpha, 1\}$ obdržíme

$$g_1 = \frac{\ln\left(\mathrm{Z}_{c;k|k-1}\left(\lambda_k = a\right)\right) - \alpha \ln\left(\mathrm{Z}_{c;k|k-1}\left(\lambda_k = 1\right)\right)}{1-\alpha} \tag{3.24a}$$

$$g_2 = \frac{\ln\left(\mathrm{Z}_{c;k|k-1}\left(\lambda_k = 1\right)\right) - \ln\left(\mathrm{Z}_{c;k|k-1}\left(\lambda_k = a\right)\right)}{1-\alpha} \tag{3.24b}$$

Při užití aproximace $\tilde{\mathrm{Z}}_{c;k|k-1}(\lambda_k)$ se λ_k vyskytuje v algebraickém vyjádření 3.19 pouze v exponentech, a to v první mocnině. Provedením logaritmu $f(\Theta_k|\lambda_k, \mathcal{D}_{1-n}^{k-1}, \Pi_k)$ tak obdržíme výraz lineárně závislý na λ_k. VB marginálie $\tilde{f}(\lambda_k|\mathcal{D}_{1-n}^{k}, \Pi_k) = \mathcal{E}_{\lambda_k}(b_k, [\alpha, 1])$ tím pádem nabývá zkráceného exponenciálního rozložení. Pro účely konjugace volíme prior $f(\lambda_k) = \mathcal{E}_{\lambda_k}(b_0, [\alpha, 1])$ shodné funkční formy.

Z algebraického vyjádření 3.19 a funkční formy $\tilde{f}(\lambda_k|\mathcal{D}_{1-n}^{k}, \Pi_k)$ je patrné, že řešení 3.18 se redukuje na vyjádření statistiky b_k, neb $\tilde{f}(\lambda_k|\mathcal{D}_{1-n}^{k}, \Pi_k) \propto \exp\{b_k \lambda_k\}$[15].

$$\begin{aligned}b_k &= \mathsf{E}_{\tilde{f}(\Theta_k|\mathcal{D}_{1-n}^{k}, \Pi_k)}\left[-\frac{d_k}{2}\begin{bmatrix}\theta_k \\ 1\end{bmatrix}' [U_1 - U_2] \begin{bmatrix}\theta_k \\ 1\end{bmatrix} + \frac{\nu_{k-1|k-1}}{2}\ln(d_k)\right] + g_2 + b_0 \\ &= \frac{1}{2}\mathrm{tr}\left((\Xi_{k-1|k-1} - P_{k-1|k-1}^{-1})P_{k|k}\right) - \frac{1}{2}\frac{\nu_{k|k}}{\Sigma_{k|k}}\begin{bmatrix}\hat{\theta}_{k|k} \\ 1\end{bmatrix}' [U_1 - U_2] \begin{bmatrix}\hat{\theta}_{k|k} \\ 1\end{bmatrix} \\ &+ \frac{1}{2}\nu_{k-1|k-1}\left(\Psi(\nu_{k|k}/2) - \ln\left(\Sigma_{k|k}/2\right)\right) + g_2 + b_0\end{aligned} \tag{3.25}$$

Nyní již můžeme vyjádřit střední hodnotu $\hat{\lambda}_k$

$$\hat{\lambda}_k = \frac{\exp(b_k)(1-b_k) - \exp(\alpha b_k)(1-\alpha b_k)}{b_k(\exp(\alpha b_k) - \exp(b_k))}. \tag{3.26}$$

Vyjádření vzájemně nezávislých postačujících statistik náhodných veličin λ_k a Θ by bez další aproximace vedlo na sadu transcendentních rovnic. Možného zjednodušení bychom mohli dosáhnout aplikováním Stirlingovy aproximace gama a digama funkce v 3.21 a 3.25. Existence řešení vedoucí na jednokrokové vyjádření $\hat{\lambda}_k$ však vyšetřována nebyla. Pro určení $\hat{\lambda}_k$ tak použijeme IVB altoritmus popsaný v kapitole 1.5.3 [15].

3.2.3 Algoritmus

Algoritmus 1: Algoritmus rekurzivní identifikace s koeficientem zapomínání určeným VB aproximací

Inicializace:

Nastav dolní limit faktoru zapomínání $\alpha \in (0, 1)$.

Nastav maximální počet iterací IVB algoritmu r_{max} a koeficient pro jeho předčasné ukončení q.

Proveď n měření pro naplnění h_1.

Inicializuj hodnoty apriorních statistik $\hat{\theta}_{0|0}, P_{0|0}, \Sigma_{0|0}, \nu_{0|0}, \Xi_{0|0}$.

Proveď přiřazení $\hat{\theta}_{0|-1} = \hat{\theta}_{0|0}$ a výpočet $V_{0|0} = P_{0|0}^{-1}$.

Online učení:	**Rovnice**
while $k \leftarrow 1, N$ **do**	
Sestav matice U_1, U_2.	▷ 3.20a, 3.20b
Dodej $\Xi_{k\|k}$.	
Inicializuj IVB algoritmus: $\hat{\lambda}_k^{[0]} = \hat{\lambda}_0$.	
for $r \leftarrow 1, r_{max}$ **do**	
Časový krok:	
Vypočti: $P_{k\|k-1}^{[r]}$, $V_{k\|k-1}^{[r]}$, $\hat{\theta}_{k\|k-1}^{[r]}$, $\nu_{k\|k-1}^{[r]}$, $\Sigma_{k\|k-1}^{[r]}$, $\Xi_{k\|k-1}^{[r]}$.	▷ 3.9a - 3.9f
Regularizace:	
Vypočti: $P_{c;k\|k-1}^{[r]}$, $V_{c;k\|k-1}^{[r]}$, $\hat{\theta}_{c;k\|k-1}^{[r]}$, $\Sigma_{c;k\|k-1}^{[r]}$.	▷ 3.13a - 3.13e
Datový krok:	
Vypočti: $P_{k\|k}^{[r]}$, $V_{k\|k}^{[r]}$, $\hat{\theta}_{k\|k}^{[r]}$, $\nu_{k\|k}^{[r]}$, $\Sigma_{k\|k}^{[r]}$, $\Xi_{k\|k}^{[r]}$.	▷ 3.15a - 3.15g
Vypočti: $g_2^{[r]}$ a $b_k^{[r]}$.	▷ 3.21, 3.24a
Vypočti: $\hat{\lambda}_k^{[r]}$.	▷ 3.26
if $\|\hat{\lambda}_k^{[r]} - \hat{\lambda}_k^{[r-1]}\| < q$ **then**	
break for loop	
end	
end	
Aktualizuj regresor: $\{h_k, u_k, y_k\} \rightarrow h_{k+1}$.	
end	

3.3 Datově informovaná volba koeficientu zapomínání určeným statistickým rozhodováním

Faktor zapomínání je dle této algoritmizace [8] také určen jako optimum konvexního optimalizačního problému postaveném na KLD. Oproti řešení v kapitole 3.2 však λ_k zůstává v podmíněnosti náhodné veličiny Θ_k. Na λ_k tak můžeme nahlížet jako na parametr, nikoliv jako na náhodnou veličinu, jehož nejlepší odhad $\hat{\lambda}_k$ získáme řešením optimalizačního problému. Ztrátový funkcionál pro určení tohoto odhadu je pro zamezení úplné modulace na zastaralý posterior oproti 3.4 doplněn o další členy. Pro jeho optimalizaci je potřeba vzhledem k omezenému definičnímu oboru vektoru φ nutno užít Karush-Kuhn-Tuckerových (KKT) optimalizačních podmínek.

Předem podotkněme, že po zbytek této kapitoly 3.3 pro přehlednost notace budeme opět opomíjet podmíněnost náhodné veličiny Θ_k.

3.3.1 Volba optimalizačního funkcionálu

Z 3.5 si můžeme povšimnout, že člen nezávislý na neznámé $\breve{f}(\Theta_k)$ představuje minimum $\min_{\breve{f}(\Theta_k)} J(\breve{f}(\Theta_k), \varphi) = -\ln\left(\int_{\Theta^*} f_1^{\varphi_1}(\Theta_k) f_2^{\varphi_2}(\Theta_k) \mathrm{d}\Theta_k\right)$ tohoto ztrátového funkcionálu. Vyjádřením KLD

$$KL\left(\breve{f}(\Theta_k) \middle\| \tilde{f}(\Theta_k)\right) = J\left(\breve{f}(\Theta_k), \varphi\right) - \min_{\breve{f}(\Theta_k)} J(\breve{f}(\Theta_k), \varphi) \tag{3.27}$$

ze vztahu 3.5 můžeme pozorovat, že přesnost aproximace $\tilde{f}(\Theta_k)$, kterou popisuje zvolená KLD, je tím vyšší, čím větší hodnoty minimum v 3.27 dosahuje. Cílem je tedy volit takové $\hat{\varphi}$, které maximalizuje $\min_{\breve{f}(\Theta_k)} J(\breve{f}(\Theta_k), \varphi)$. Je však patrné, že optimum $\hat{\varphi} = \arg\max_{\varphi} \min_{\breve{f}(\Theta_k)} J(f(\Theta_k), \varphi)$ není datově informované. Takový výsledek je očividně nežádoucí. Funkcionál pro nalezení optimálního koeficientu zapomínání $\hat{\lambda}_k$ je tedy nutno formulovat jinak.

Nová volba optimalizačního funkcionálu

Zavádíme nový ztrátový funkcionál 3.28, který alternativy predikcí popsané vektorem φ dodatečně penalizuje.

$$\begin{aligned} J\left(\breve{f}(\Theta_k), \varphi\right) &\equiv \mathsf{E}_{f(H|\varphi)}\left[KL\left(\breve{f}(\Theta_k) \middle\| f_H(\Theta_k)\right) - \varrho_{H,\gamma}\right] \\ &= \sum_{i=1}^{2} \varphi_i \left[KL\left(\breve{f}(\Theta_k) \middle\| f_i(\Theta_k)\right) - \varrho_{i,\gamma}\right] \end{aligned} \tag{3.28}$$

Optimální hodnota $\tilde{f}(\Theta_k)$ bude opět ve formě geometrického váženého průměru 3.6. Záporně vzaté minimum 3.29 funkcionálu 3.28 však bude nyní závislé na členech

$\varrho_{H,\gamma}$, jejichž volba bude popsána v kapitole 3.3.2 [8].

$$\begin{aligned}\omega(\varphi) &\equiv -\min_{f(\Theta_k)} J(f(\Theta_k),\varphi)\\ &= \ln\left(\int_{\Theta^*} f_1^{\varphi_1}(\Theta_k) f_2^{\varphi_2}(\Theta_k)\mathrm{d}\Theta_{k+1}\right) + \varphi_1\varrho_{1,\gamma} + \varphi_2\varrho_{2,\gamma}.\end{aligned} \tag{3.29}$$

Volba optima minimalizací 3.29 ekvivalentní maximalizaci $\min_{f(\Theta_k)} J(f(\Theta_k),\varphi)$ vzhledem k φ již nyní bude zahrnovat datovou informaci, a to díky φ. Vzhledem k tvaru 3.29 optimalizace spočívá v hledání $\hat{\varphi}$ namísto $\hat{\lambda}_k$ jako

$$\hat{\varphi} = \arg\min_{\varphi} \omega(\varphi). \tag{3.30}$$

Jak ale plyne z 3.7, tyto optimalizační problémy jsou však vzájemně ekvivalentní.

3.3.2 Volba dodatečné penalizace alternativ predikce

Datová zpětná vazba je do optimalizačního funkcionálu zavedena prostřednictvím $\varrho_{H,\gamma} \in \{\varrho_{1,\gamma}, \varrho_{2,\gamma}\}$ realizovaného ve dvou variantách $\gamma \in \{\zeta, \varrho\}$. Myšlenka spočívá v posouzení úspěšnosti alternativ predikcí $f_1(\Theta_{k-1})$, $f_2(\Theta_{k-1})$ v předchozím kroku $k-1$ vzhledem k posterioru $f(\Theta_{k-1}|\mathcal{D}_{1-n}^{k-1}, \Pi_{k-1})$ tohoto kroku. Pro vyjádření rozdílností těchto fhp je opět užito KLD. Pokud bychom za $\varrho_{1,\gamma}, \varrho_{2,\gamma}$ dosadili

$$KL\Big(f\left(\Theta_{k-1}|\mathcal{D}_{1-n}^{k-1}, \Pi_{k-1}\right)\Big\|f_1\left(\Theta_{k-1}\right)\Big), \tag{3.31a}$$

$$KL\Big(f\left(\Theta_{k-1}|\mathcal{D}_{1-n}^{k-1}, \Pi_{k-1}\right)\Big\|f_2\left(\Theta_{k-1}\right)\Big), \tag{3.31b}$$

byl by výsledek optimalizace $\omega(\varphi)$ příliš citlivý na šum a docházelo by k nesprávným detekcím změny parametrů soustavy. Varianty ζ a ρ představují dva různé způsoby robustifikace 3.31a, 3.31b potlačující toto chování [8].

Robustifikace vůči šumu $\gamma \equiv \zeta$

První varianta $\gamma \equiv \zeta$ uměle snižuje odhadovanou přesnost d_k

$$f_\zeta(\Theta_{k-1}) \equiv \mathcal{N}_{\theta_{k-1}}\left(\hat{\theta}_{k-1|k-1}, P_{k-1|k-1}/(d_{k-1}\zeta)\right) f\left(d_{k-1}|\mathcal{D}_{1-n}^{k-1}, \Pi_{k-1}\right) \tag{3.32a}$$

$$f_{1,\zeta}(\Theta_{k-1}) \equiv \mathcal{N}_{\theta_{k-1}}\left(\hat{\theta}_{k-2|k-2}, P_{k-2|k-2}/(d_{k-1}\zeta)\right) f_1\left(d_{k-1}\right) \tag{3.32b}$$

$$f_{2,\zeta}(\Theta_{k-1}) \equiv \mathcal{N}_{\theta_{k-1}}\left(\hat{\theta}_{k-2|k-2}, P_{k-2|k-2}/(d_{k-1}\zeta\alpha)\right) f_2\left(d_{k-1}\right) \tag{3.32c}$$

parametrem $\zeta \in (0,1]$ a redukuje tak reakce na šum.

$$\varrho_{1,\zeta} \equiv KL\Big(f_\zeta(\Theta_{k-1})\Big\|f_{1,\zeta}(\Theta_{k-1})\Big) \tag{3.33a}$$

$$\varrho_{2,\zeta} \equiv KL\Big(f_\zeta(\Theta_{k-1})\Big\|f_{2,\zeta}(\Theta_{k-1})\Big) \tag{3.33b}$$

Robustifikace vůči šumu $\gamma \equiv \rho$

U druhé varianty $\gamma \equiv \rho$ je robustifikace proti šumu provedena snižováním penalty pro první hypotézu parametrem $\varrho \in (\alpha, 1]$.

$$\varrho_{1,\rho} \equiv \rho KL\Big(f\left(\Theta_{k-1}\Big|\mathcal{D}_{1-n}^{k-1}, \Pi_{k-1}\right)\Big\|f_1(\Theta_{k-1})\Big) \tag{3.34a}$$

$$\varrho_{2,\rho} \equiv KL\Big(f\left(\Theta_{k-1}|\mathcal{D}_{1-n}^{k-1}, \Pi_{k-1}\right)\Big\|f_2(\Theta_{k-1})\Big) \tag{3.34b}$$

Obě tyto varianty jsou převzaty z [8].

3.3.3 Nutné a postačující podmínky optimality

KKT optimalizační podmínky, které je nutno kvůli omezení definičního oboru φ užít, poskytují vzhledem ke striktní konvexnosti $\omega(\varphi)$ nutné a zároveň postačující podmínky optimality. Za užití omezení plynoucích z pravděpodobnostního charakteru φ definujeme vazební podmínky

$$\varphi_1 + \varphi_2 = 1 \rightarrow g_1(\varphi) \equiv \varphi_1 + \varphi_2 - 1 = 0 \tag{3.35a}$$

$$\varphi_1 \geq 0 \rightarrow g_2(\varphi) \equiv -\varphi_1 \leq 0 \tag{3.35b}$$

$$\varphi_2 \geq 0 \rightarrow g_3(\varphi) \equiv -\varphi_2 \leq 0 \tag{3.35c}$$

a sestavíme Lagrangeovu funkci $\mathcal{L}(\varphi, \eta)$

$$\mathcal{L}(\varphi, \eta) \equiv -\omega(\varphi) - \sum_{i=1}^{3} \eta_i g_i(\varphi), \tag{3.36}$$

kde $\eta \equiv \{\eta_1, \eta_2, \eta_3\}$ je množina Lagrangeových násobitelů $\eta_1 \in \mathbb{R}_{>0}$, $\eta_2, \eta_3 \in \mathbb{R}_{\geq 0}$. Provedením derivací $\mathcal{L}(\varphi, \eta)$ podle všech proměnných a položením rovnosti nule získáme sadu rovnic, jejichž řešení poskytuje $\hat{\varphi}$.

$$\mathrm{I}: \int_{\Theta^*} \tilde{f}(\Theta_k) \ln\left(\tilde{f}(\Theta_k)\right) \mathrm{d}\Theta_k - KL\left(\tilde{f}(\Theta_k)\Big\|f_1(\Theta_k)\right) + \varrho_{1,\gamma} - \eta_1 + \eta_2 = 0 \tag{3.37a}$$

$$\mathrm{II}: \int_{\Theta^*} \tilde{f}(\Theta_k) \ln\left(\hat{f}(\Theta_k)\right) \mathrm{d}\Theta_k - KL\left(\tilde{f}(\Theta_k)\Big\|f_2(\Theta_k)\right) + \varrho_{2,\gamma} - \eta_1 + \eta_3 = 0 \tag{3.37b}$$

$$\mathrm{III}: \eta_1(\hat{\varphi}_1 + \hat{\varphi}_2 - 1) = 0 \tag{3.37c}$$

$$\mathrm{VI}: -\eta_2\hat{\varphi}_1 = 0 \tag{3.37d}$$

$$\mathrm{V}: -\eta_3\hat{\varphi}_2 = 0 \tag{3.37e}$$

Kladnost násobitele η_1 odkazuje na skutečnost, že podmínka 3.35a je omezující vždy. Podmínky 3.35b a 3.35c mohou být neomezující ($\eta_2 = \eta_3 = 0 \rightarrow \hat{\varphi}_i \in (0,1), i \in \{1,2\}$), nebo mohou být exkluzivně omezující ($\eta_2 = 0$, $\eta_3 > 0$

$\rightarrow \hat{\varphi} = [1,0]'$ či $\eta_2 > 0$, $\eta_3 = 0 \rightarrow \hat{\varphi} = [0,1]'$). Výsledky nezbytných a postačujících KKT podmínek pro optimální volbu φ jsou pak dány pro $\eta_2 = \eta_3 = 0$ vztahem 3.38a, pro $\eta_2 = 0$, $\eta_3 > 0$ vztahem 3.38b a pro $\eta_2 > 0$, $\eta_3 = 0$ vztahem 3.38c [8, 11].

$$KL\left(\tilde{f}(\Theta_k)\middle\|f_1(\Theta_k)\right) - \varrho_{1,\gamma} = KL\left(\tilde{f}(\Theta_k)\middle\|f_2(\Theta_k)\right) - \varrho_{2,\gamma} \tag{3.38a}$$

$$KL\left(\tilde{f}(\Theta_k)\middle\|f_1(\Theta_k)\right) - \varrho_{1,\gamma} \geq KL\left(\tilde{f}(\Theta_k)\middle\|f_2(\Theta_k)\right) - \varrho_{2,\gamma} \tag{3.38b}$$

$$KL\left(\tilde{f}(\Theta_k)\middle\|f_1(\Theta_k)\right) - \varrho_{1,\gamma} \leq KL\left(\tilde{f}(\Theta_k)\middle\|f_2(\Theta_k)\right) - \varrho_{2,\gamma} \tag{3.38c}$$

3.3.4 Analytické řešení Kullback-Leiblerovy divergence dvojice normálních-Wishartových rozložení

Z 3.38a - 3.38c plyne, že pro získání $\hat{\varphi}$ je potřeba najít obecné analytické řešení KLD dvojice $\mathcal{NW}$ rozložení. Pro řešení itegrálu, jež KLD představuje, je možno užít výsledků expektací uvedených v kapitolách 1.3.1 a 1.3.2, kdy nejprve provádíme integraci podle θ, poté podle d.

$$\begin{aligned} &KL\Big(\mathcal{N}_\theta(\hat{\theta}_1, (V_1 d)^{-1})\mathcal{W}_d(\Sigma_1, \nu_1)\Big\|\mathcal{N}_\theta(\hat{\theta}_2, (V_2 d)^{-1})\mathcal{W}_d(\Sigma_2, \nu_2)\Big) \\ &= \mathsf{E}_{\mathcal{N}_\theta(\hat{\theta}_1,(V_1 d)^{-1})\mathcal{W}_d(\Sigma_1,\nu_1)}\left[\frac{d}{2}\left(\theta'\left(V_2 - V_1\right)\theta - 2\left(\hat{\theta}_2' V_2 - \hat{\theta}_1' V_1\right)\theta\right.\right. \\ &\left.\left. -\hat{\theta}_1' V_1 \hat{\theta}_1 + \hat{\theta}_2' V_2 \hat{\theta}_2 - (\Sigma_1 - \Sigma_2)\right) - \frac{1}{2}\left(\nu_2 - \nu_1\right)\ln\left(d\right)\right] + \ln\left(\frac{\iota_1}{\iota_2}\right) \end{aligned} \tag{3.39}$$

Člen $\iota_i \equiv |V_i|^{1/2}(2\pi)^{-p/2}(\Sigma_i/2)^{\nu_i/2}\Gamma^{-1}(\nu_i/2),\ i \in \{1,2\}$ v 3.39 zapouzdřuje normalizační konstantu pro i-té $\mathcal{NW}$ rozložení.

Použití exaktního řešení 3.39 pro vyjádření $\hat{\lambda}_k$ z 3.38a - 3.38c by vedlo na transcendentální rovnice vzhledem k $\hat{\lambda}_k$, a to kvůli funkcím $\Gamma(\cdot)$ a $\Psi(\cdot)$. Použitím Stirlingových asymptotických aproximací $\Gamma(z) \simeq \exp\{-z\}z^{z-1/2}(2\pi)^{1/2}$ a $\Psi(z) \simeq \ln(z) - 1/(2z)$ těchto funkcí obdržíme vyjádření [8]

$$\begin{aligned} &KL\left(\mathcal{N}_\theta(\hat{\theta}_1, V_1^{-1}/d)\mathcal{W}_d(\Sigma_1, \nu_1)\middle\|\mathcal{N}_\theta(\hat{\theta}_2, V_2^{-1}/d)\mathcal{W}_d(\Sigma_2, \nu_2)\right) \\ &\simeq \frac{1}{2}\ln\left(\frac{|V_1|}{|V_2|}\right) + \frac{\nu_2}{2}\ln\left(\frac{\Sigma_1}{\Sigma_2}\right) + \frac{1}{2}\ln\left(\frac{(\nu_2/2)^{(\nu_2-1)}}{(\nu_1/2)^{(\nu_1-1)}}\right) \\ &+ \frac{\nu_1}{2}\left(\hat{\theta}_1 - \hat{\theta}_2\right)' V_2\left(\hat{\theta}_1 - \hat{\theta}_2\right)\Sigma_1{}^{-1} - \frac{\nu_1}{2}\left(\Sigma_1 - \Sigma_2\right)\Sigma_1{}^{-1} \\ &+ \frac{1}{2}\mathrm{tr}\left(V_1^{-1}V_2\right) + \frac{\nu_1 - \nu_2}{2}\left(\ln\left(\frac{\nu_1}{2}\right) - \frac{1}{\nu_1} + 1\right) - \frac{p}{2}. \end{aligned} \tag{3.40}$$

3.3.5 Nalezení koeficientu zapomínání

Dosadíme-li do vztahu 3.38a funkční formy jednotlivých členů a použijeme 3.40, můžeme algebraicky vyjádřit hodnotu $\hat{\lambda}_k$, a to pro obě realizace ζ, ρ. Úpravy vedoucí

k výsledným hodnotám vyžadují pouze snadné matematické operace, jsou však poměrně pracné. [8].

Řešení odpovídající rovnosti 3.38a představují

$$\gamma \equiv \zeta : \frac{1}{\hat{\lambda}_k} = \frac{X_\zeta}{\hat{\lambda}_{k-1}\,(p+1)}, \tag{3.41a}$$

$$\gamma \equiv \rho : \frac{1}{\hat{\lambda}_k} = \frac{(\rho-\alpha)\,X_{\rho 1}}{\hat{\lambda}_{k-1}\,(1-\alpha)\,(p+1)} + \frac{(1-\rho)\,X_{\rho 2}}{(1-\alpha)\,(p+1)} \tag{3.41b}$$

pro příslušné varianty γ. Výrazy $X_\zeta, X_{\rho 1}, X_{\rho 2}$ jsou dány jako

$$\begin{aligned} X_\zeta &= \operatorname{tr}\left(P_{k-1|k-1}V_{k-1|k-2}\right) + \nu_{k-1|k-2}\ln\left(\hat{d}_{k-1|k-2}/\hat{d}_{k-1|k-1}\right) \\ &\quad + \hat{d}_{k-1|k-1}\Sigma_{k-1|k-2} + \hat{d}_{k-1|k-1}\varepsilon_k'\zeta V_{k-1|k-2}\varepsilon_k + \nu_{k-1|k-2}/\nu_{k-1|k-1} - \nu_{k-1|k-2}, \end{aligned} \tag{3.42a}$$

$$\begin{aligned} X_{\rho 1} &= \operatorname{tr}\left(P_{k-1|k-1}V_{k-1|k-2}\right) + \nu_{k-1|k-2}\ln\left(\hat{d}_{k-1|k-2}/\hat{d}_{k-1|k-1}\right) \\ &\quad + \hat{d}_{k-1|k-1}\Sigma_{k-1|k-2} + \hat{d}_{k-1|k-1}\varepsilon_k' V_{k-1|k-2}\varepsilon_k + \nu_{k-1|k-2}/\nu_{k-1|k-1} - \nu_{k-1|k-2}, \end{aligned} \tag{3.42b}$$

$$X_{\rho 2} = \ln\left(\frac{\left|V_{k-1|k-2}\right|\nu_{k-1|k-2}}{\left|V_{k-1|k-1}\right|\nu_{k-1|k-1}}\right) + (p+1)\left(1 - \ln\left(\hat{\lambda}_{k-1}\right)\right), \tag{3.42c}$$

kde $\hat{d}_{k|k} = \nu_{k|k}/\Sigma_{k|k}$ a $\varepsilon_k \equiv \hat{\theta}_{k-1|k-1} - \hat{\theta}_{k-1|k-2}$, přičemž bylo použito zastaralého koeficientu zapomínání $\hat{\lambda}_{k-1}$ pro provedení časového kroku dle 3.9a - 3.9f [8].

Řešení plynoucí z nerovnic 3.38b a 3.38c užitím 3.40 není klíčové. Výsledné nerovnice by představovaly podmínky, jejichž platnost by referovala na aktivitu jednoho z omezení 3.35b, 3.35c. Pokud by některé z těchto omezení bylo aktivní, pak $\hat{\lambda}_k$ by vyšlo dle 3.41a, respektive dle 3.41b, mimo rozsah $(\alpha, 1)$. Pouhým přenastavením $\hat{\lambda}_k$ na nejbližší hodnotu tohoto rozsahu však získáme validní výsledek respektující aktivitu příslušného omezení.

3.3.6 Algoritmus

Algoritmus 2: Algoritmus rekurzivní identifikace s koeficientem zapomínání určeným statistickým rozhodováním

Inicializace:

Nastav dolní limit faktoru zapomínání $\alpha \in (0, 1)$ a parametr $\zeta \in (0, 1]$ nebo $\rho \in (\alpha, 1]$ podle zvolené varianty.

Proveď n měření pro naplnění regresoru h_1.

Inicializuj hodnoty apriorních statistik $\hat{\theta}_{0|0}, P_{0|0}, \Sigma_{0|0}, \nu_{0|0}, \Xi_{0|0}$.

Proveď přiřazení $\hat{\theta}_{0|-1} = \hat{\theta}_{0|0}$ a výpočet $V_{0|0} = P_{0|0}^{-1}$.

Online učení: **Rovnice:**

while $k \leftarrow 1, N$ **do**

Určení $\hat{\lambda}_k$:

if $k = 1$ **then**

$\hat{\lambda}_k = 1$

else

Vypočti X_ζ nebo $X_{\rho 1}, X_{\rho 2}$ ▷ 3.42a - 3.42c

Vypočti $\hat{\lambda}_k$ ▷ 3.41a - 3.41b

if $\hat{\lambda}_k > 1$ **then**

$\hat{\lambda}_k = 1$

else if $\hat{\lambda}_k < \alpha$ **then**

$\hat{\lambda}_k = \alpha$

end

end

Časový krok:

Vypočti: $P_{k|k-1}$, $V_{k|k-1}$, $\hat{\theta}_{k|k-1}$, $\nu_{k|k-1}$, $\Sigma_{k|k-1}$, $\Xi_{k|k-1}$. ▷ 3.9a - 3.9f

Regularizace:

Dodej $\Xi_{k|k}$.

Vypočti: $P_{c;k|k-1}$, $V_{c;k|k-1}$, $\hat{\theta}_{c;k|k-1}$, $\Sigma_{c;k|k-1}$. ▷ 3.13a - 3.13e

Datový krok:

Vypočti: $P_{k|k}$, $V_{k|k}$, $\hat{\theta}_{k|k}$, $\nu_{k|k}$, $\Sigma_{k|k}$, $\Xi_{k|k}$. ▷ 3.15a - 3.15g

Aktualizuj $h_k \rightarrow h_{k+1}$.

end

4 Bayesovská identifikace banky přepínaných normálních časově invariantních regresních modelů

U některých typů systémů dochází ke změně parametrů skokem a často existuje konečný počet modelů, které popisují chování mezi skoky. Jako příklad mohou posloužit chemické procesy, kdy jednotlivé operace během krátkého časového horizontu změní svou strukturu spjatou s dynamickým chováním systému. Ta je pak až do další operace považována za neměnnou. Identifikací banky modelů, popisující chování takového systému, získáme nejen popis dynamických částí, ale také informaci o aktivitě jednotlivých modelů v čase. Pro další pozorování je pak na základě získaných statistik možné posoudit časový vývoj přepínání modelů a vyhodnocovat tak aktuální pracovní stav [1].

Tento koncept, známý také jako "*Gaussian Mixture*" (GM), tedy předpokládá, že pozorovanou veličinu y_k v daném časovém kroku k generuje jeden z M normálních modelů, v našem případě ARX modelů. Problém identifikace při apriorní znalosti příslušností z pozorovaných dat k daným modelům pro jednotlivá měření je redukován na identifikaci ARX modelů popsanou v kapitole 2.3. My však budeme předpokládat neznalost identit z, která je ale pro inferenci parametrů ARX modelů nezbytná. Segmentace dat spočívající v odhadování těchto identit z spolu s inferencí banky modelů ARX je náplní následujících podkapitol [7].

Nadále budeme indexem i odkazovat na i-tý model, kde $i \in [1, M]$ a indexem k na k-tý časový krok, kde $k \in [1, N]$. N je počet měřených dat, jež jsme schopni zcela datově podmínit regresory h_k a M představuje horní limit počtu identifikovaných modelů. Skutečný počet modelů dokáží algoritmy inferovat samy.

V následujících kapitolách 4.2 a 4.3 představíme dvě algoritmizace identifikace GM modelu, obě inspirované [1]. Jak již bylo zmíněno, volba prioru hraje podstatnou roli ve způsobu inference a řešitelnosti posterioru. Právě z různých voleb priorů budou pramenit vlastnosti a rozdílnosti těchto algoritmů. Nejprve však diskutujme bayesovský GM model společný pro oba algoritmy.

4.1 Bayesovský GM model

Diferenční rovnice GM modelu 4.1 sestává z lineární kombinace výstupů ARX modelů popsaných v kapitolách 2.2, 2.3. Modelům přiřazujeme shodnou strukturu deterministické části s různými parametry θ_i, s příslušným regresorem h_k a odlišnou

přesnost d_i šumu $e_k^{(i)} \sim \mathcal{N}(0, 1/d_i)$.

$$y_k = \sum_{i=1}^{M} z_{k,i} \left(h_k' \theta_i + e_{k,i} \right) \tag{4.1}$$

Obě algoritmizace uvedené v kapitolách 4.2 a 4.3 pracují s předpokladem exkluzivity generování y_k danými modely, kterou vyjadřuje množina možných hodnot $z_k \in \{\epsilon_M(1), \ldots, \epsilon_M(M)\}$. Vektor $\epsilon_M(i) \equiv [\mathbf{0}_{i-1}', 1, \mathbf{0}_{M-i+1}']' \in \mathbb{R}^M$ je takový vektor, jehož i-tý prvek je 1 a zbylé prvky jsou 0. Na základě tohoto předpokladu považujeme zápisy 4.1 a 4.2 za ekvivalentní [2].

$$y_k = \prod_{i=1}^{M} \left(h_k' \theta_i + e_{k,i} \right)^{z_{k,i}} \tag{4.2}$$

Identifikace dále popsanými algoritmy je založena na parametrickém modelu 4.2 s příslušnou hustotou

$$f(y_k | h_k, \theta, d, z_k) = \prod_{i=1}^{M} \mathcal{N}_{y_k} \left(h_k{}' \theta_i, 1/d_i \right)^{z_{k,i}}, \tag{4.3}$$

kde $\theta \equiv \{\theta_1, ..., \theta_M\}$ a $d \equiv \{d_1, ..., d_M\}$. Jedná se tedy o identifikaci modelu daného váženým geometrickým průměrem výstupu modelů, nikoliv váženým aritmetickým průměrem, jak bylo původně uvedeno v 4.1. Neekvivalence 4.1 a 4.2 pro obecný vektor $z_k \in \Delta_M$ představuje nemalé omezení těchto algoritmů. Podotkněme, že rovnost levé a pravé strany v 4.3 platí pouze za zmíněného předpokladu kladeného na vektor $z_{k,i}$.

Na přechod k modelu 4.3 od původního modelu

$$f(y_k | h_k, \theta, d, z_k) = \sum_{i=1}^{M} z_{k,i} \mathcal{N}_{y_k} \left(h_k{}' \theta_i, 1/d_i \right) \tag{4.4}$$

se můžeme dívat i jako na výsledek optimalizace

$$\tilde{f}(y_k) = \arg\min_{\breve{f}(y_k)} \mathsf{E}_{f(z_k)} \left[KL\Big(\breve{f}(y_k) \Big\| f(y_k) \Big) \right], \tag{4.5}$$

která je bezeztrátová právě za předpokladu $z_k \in \{\epsilon_M(1), \ldots, \epsilon_M(M)\}$. S touto volbou ztrátového funkcionálu i výsledkem v podobě váženého geometrického průměru jsme se setkali již v kapitole 3.1.1, odvození optima zde tedy provádět nebudeme. Podotkněme, že podmíněnost náhodné veličiny y_k znalostí y_k, θ, d, z_k byla v 4.5 pro přehlednost notace opomenuta. Vzhledem k zavedení předpokladu o definičním oboru z_k nebudeme na 4.3 nahlížet jako na aproximaci, ale jako na ekvivalentní vyjádření 4.4. Znak $\tilde{f}$ užívaný pro označení aproximace tak pro hustotu popisující y_k uvádět nebudeme.

4.1.1 Užití VB aproximace pro dosažení konjugace

Jelikož předpokládáme neznalost z_k, je nasnadě za pomoci prioru $f(z_k|\alpha)$ vyjádřit sdruženou hustotu $f(y_k, z_k|h_k, \theta, d, \alpha)$ a pomocí marginalizace získat $f(y_k|h_k, \theta, d, \alpha)$, která za předpokladu kategorické distribuce $f(z_k|\alpha) = \mathcal{C}_{z_k}(\alpha)$ má tvar

$$f(y_k|h_k, \theta, d, \alpha) = \sum_{i=1}^{M} \alpha_i \mathcal{N}_{y_k}\left(h_k{}'\theta_i, 1/d_i\right). \tag{4.6}$$

Tvar 4.6 je v literatuře označován jako kýžený bayesovksý GM model, kde α_i je i-tý prvek vektoru $\alpha \in \Delta_M$, jež popisuje pravděpodobnost aktivity jednotlivých modelů.

Tento model, narozdíl od 4.3, nespadá do rodiny dynamických exponenciálních rozložení. Suma ve výrazu 4.6 zaručuje exponenciální nárůst složitosti posterioru s počtem datových párů N. Přestože se nebudeme zabývat rekurzivním algoritmem, je zaručení konjugace klíčové, neboť požadujeme, aby posteriorní statistiky vypovídaly o předem definovaném, avšak apriorně ne nutně známém skutečném počtu modelů. Není žádoucí, aby byl počet těchto modelů závislý na počtu měření, natož abychom pro každý datový pár k získávali další sadu modelů. Tuto situaci lze řešit užitím VB aproximace v souladu s postupem uvedeným v kapitole 1.5.4. Namísto nekonjugovaného pozorovacího modelu 4.6 použijeme za zmíněného omezení pro z_k konjugovaný model 4.3, kde z_k představuje latentní náhodnou veličinu. Posterior nezávislý na této náhodné veličině obdržíme ve formě VB marginálií [2].

4.1.2 Vyjádření věrohodnostní funkce

Pro obdržení sdružené hustoty potřebné pro VB metodu potřebujeme vyjádřit $f(\mathcal{D}_{1-n}^{N}|\theta, d, z)$. Vycházejme z parametrického modelu 4.3. Užitím předpokladu platnosti přirozené podmínky řízení $f(u_k|\mathcal{D}_{1-n}^{k-1}, \theta, d, z) = f(u_k|\mathcal{D}_{1-n}^{k-1})$ a řetězového pravidla, obdobně jako v kapitole 2.3, obdržíme

$$f\left(\mathcal{D}_{1-n}^{N}\middle|\theta, d, z, z_{1-n}^{0}\right) = \prod_{k=1}^{N} \left\{f(y_k|h_k, \theta, d, z) f(u_k|\mathcal{D}_{1-n}^{k-1})\right\} f(\mathcal{D}_{1-n}^{0}|\theta, d, z_{1-n}^{0}). \tag{4.7}$$

Člen z_{1-n}^{0} doplňuje identitu výstupů y_k pro kroky $k \in [1-n, 0]$, což je část experimentu sloužící pouze k naplnění regresorů. Pro tyto kroky $k \in [1-n, 0]$ nelze výstup datově zcela podmínit a neúplnou informaci tak považujeme za neinformativní $f(\theta, d, z_{1-n}^{0}|\mathcal{D}_{1-n}^{0}) = f(\theta, d, z_{1-n}^{0}) \Rightarrow f(\mathcal{D}_{1-n}^{0}|\theta, d, z_{1-n}^{0}) = f(\mathcal{D}_{1-n}^{0})$. Podmíněnost veličinou z_{1-n}^{0} se z rovnice 4.7 vytrácí a výraz se zjednoduší na

$$f(\mathcal{D}_{1-n}^{N}|\theta, d, z) = \prod_{k=1}^{N} \left\{f(y_k|h_k, \theta, d, z) f(u_k|\mathcal{D}_{1-n}^{k-1})\right\} f(\mathcal{D}_{1-n}^{0}). \tag{4.8}$$

4.2 Identifikace přepínaných ARX modelů zatížených šumem o společné přesnosti

Zatímco v následující kapitole 4.3 budeme volit prior náhodných veličin θ_i shodně s [1], nyní pracujme s priorem voleným v souladu s kapitolou 2.3. Při volbě konjungovaného prioru $f(\theta_i|d_i) = \mathcal{N}_{\theta_i}(\hat{\theta}_{0,i}, P_{0,i}/d_i)$ konverguje výpočet posteriorní aproximace k nechtěným výsledkům. Některým z modelů je při výpočtu IVB algoritmem, na který výpočet VB marginálií povede, vlivem velkého rozdílu predikce výstupu $\hat{y}_{k,i}$ od skutečné hodnoty y_k přiřazena nízká odhadovaná přesnost šumu $\hat{d}_i$ a chyba těchto odhadů tak není dostatečně penalizována. To způsobuje nesprávné přiřazení identit z_k daných výstupů y_k k těmto modelům. Na základě těchto nesprávných odhadů identit se tyto modely v dalším kroku IVB algoritmu učí na datech, jež jim ale ve skutečnosti nepřísluší. Výsledkem je, že marginální statistiky počítané IVB algoritmem nekonvergují ke správným hodnotám.

Pakliže bychom akceptovali degradaci modelu 4.3 zavedením předpokladu, že všechny modely jsou zatíženy shodně velkým bílým šumem $e \sim \mathcal{N}(0, 1/d)$, zbavíme se zmíněné nechtěné vlastnosti algoritmu. Za tohoto předpokladu obdržíme parametrický model

$$f(y_k|\mathcal{D}_{1-n}^{k-1}, \theta, d, z_k) = \prod_{i=1}^{M} \mathcal{N}_{y_k}\left({h_k}'\theta_i, 1/d\right)^{z_{k,i}}. \tag{4.9}$$

Předpoklad 4.9 je hlavním rozdílem mezi touto algoritmizací a tou popsané v kapitole 4.3.

4.2.1 Volba prioru

Pro zajištění konjugace k parametrickému modelu 4.9 volíme pro θ, d normálně-Wishartův prior. Prior k náhodné veličině z_k modelované kategorickým rozložením 4.10c bude pro zajištění konjugace pocházet z Dirichletova rozložení 4.10d. Parametrizace těchto rozložení popisují rovnice 4.10a - 4.10d.

$$f(\theta|d) = \prod_{i=1}^{M} \mathcal{N}_{\theta_i}\left(\hat{\theta}_{0,i}, P_{0,i}/d\right), \tag{4.10a}$$

$$f(d) = \mathcal{W}_d\left(\Sigma_0, \nu_0\right) \tag{4.10b}$$

$$f(z|\alpha) = \prod_{k=1}^{N} f(z_k|\alpha) = \prod_{k=1}^{N} \mathcal{C}_{z_k}\left(\alpha\right) \tag{4.10c}$$

$$f(\alpha) = \mathcal{D}_\alpha\left(\tau_0\right) \tag{4.10d}$$

4.2.2 Vyjádření VB marginálií

Za pomoci 4.8, 4.9 a 4.10a - 4.10d jsme nyní až na $\prod_{k=1}^{M} f(u_k|\mathcal{D}_{1-n}^{k-1})$ schopni algebraicky vyjádřit všechny členy tvořící sdruženou hustotu

$$\begin{aligned} f(\theta, d, z, \alpha, \mathcal{D}_{1-n}^{N}) = \prod_{k=1}^{M} \left\{ f(y_k|h_k, \theta, d, z_k) f(u_k|\mathcal{D}_{1-n}^{k-1}) \right\} \\ \times f(\theta|d) f(d) f(z|\alpha) f(\alpha). \end{aligned} \tag{4.11}$$

Neznalost algebraického vyjádření zmíněné hustoty popisující u_k pro nás však není klíčová, neboť vzhledem k $\Theta \equiv \{\theta, d, \alpha\}$, z představuje pouze konstantu.

Prostor $\breve{f}^*(\Theta, z|\mathcal{D}_{1-n}^{N})$ možných kandidátů na aproximaci budeme definovat jako

$$\breve{f}^*(\Theta, z|\mathcal{D}_{1-n}^{N}) \equiv \left\{ f(\Theta, z|\mathcal{D}_{1-n}^{N}) \middle| f(\Theta, z|\mathcal{D}_{1-n}^{N}) = f(\theta, d|\mathcal{D}_{1-n}^{N}) f(\alpha|\mathcal{D}_{1-n}^{N}) f(z|\mathcal{D}_{1-n}^{N}) \right\}. \tag{4.12}$$

Důvod volby 4.12 bude patrný z algebraických vyjádření jednotlivých VB marginalií uvedených v následujících podkapitolách.

VB pro θ, d

$$\begin{aligned} \tilde{f}(\theta, d|\mathcal{D}_{1-n}^{N}) &\propto \exp\left\{ \mathsf{E}_{\tilde{f}(z|\mathcal{D}_{1-n}^{N})} \left[\ln\left(\prod_{k=1}^{M} f(y_k|h_k, \theta, d, z_k) \right) \right] \right\} f(\theta, d) \\ &\propto \prod_{i=1}^{M} \exp\left\{ -\frac{d}{2} \begin{bmatrix} \theta_i \\ 1 \end{bmatrix}' \begin{bmatrix} P_{0,i}^{-1} + \sum_{k=1}^{N} \hat{z}_{k,i} h_k h_k' & -P_{0,i}^{-1}\hat{\theta}_0 - \sum_{k=1}^{N} \hat{z}_{k,i} h_k y_k \\ -\hat{\theta}_0' P_{0,i}^{-1} - \sum_{k=1}^{N} \hat{z}_{k,i} h_k' y_k & \hat{\theta}_0' P_{0,i}^{-1}\hat{\theta}_0 + \sum_{k=1}^{N} \hat{z}_{k,i} y_k^2 \end{bmatrix} \begin{bmatrix} \theta_i \\ 1 \end{bmatrix} \right\} \\ &\times d^{\frac{Np}{2}} d^{\frac{\nu_0 + N - 2 + Np}{2}} \exp\left\{ -\Sigma_0 d/2 \right\} \end{aligned} \tag{4.13}$$

Vidíme, že $\tilde{f}(\theta, d|\mathcal{D}_{1-n}^{N})$ zachovává shodný funkční tvar jako prior $f(\theta, d)$, tedy produkt normálních rozložení a Wishartova rozložení

$$\tilde{f}(\theta, d|\mathcal{D}_{1-n}^{N}) = \prod_{i=1}^{M} \left\{ \mathcal{N}_{\theta_i}\left(\hat{\theta}_i, P_i/d\right) \right\} \mathcal{W}_d\left(\nu, \Sigma\right) \tag{4.14}$$

s parametry

$$P_i = \left(P_{0,i}^{-1} + \sum_{k=1}^{N} \hat{z}_{k,i} h_k h_k' \right)^{-1}, \tag{4.15a}$$

$$\hat{\theta}_i = P_i \left(P_{0,i}^{-1}\hat{\theta}_0 + \sum_{k=1}^{N} \hat{z}_{k,i} h_k y_k \right), \tag{4.15b}$$

$$\nu = \nu_0 + N, \tag{4.15c}$$

$$\Sigma = \Sigma_0 + \sum_{i=1}^{M} \left(\hat{\theta}_0' P_{0,i}^{-1}\hat{\theta}_0 - \hat{\theta}_i' P_i^{-1}\hat{\theta}_i + \sum_{k=1}^{N} \hat{z}_{k,i} y_k^2 \right). \tag{4.15d}$$

Z níže uvedené VB marginálie pro z můžeme v 4.15a, 4.15b a 4.15d nahradit expektace postačujícími statistikami $\hat{z}_{k,i} = \gamma_{k,i}$.

VB pro z

$$\begin{aligned}\tilde{f}(z|\mathcal{D}_{1-n}^N) \propto \exp\Bigg\{&\mathsf{E}_{\tilde{f}(\theta,d|\mathcal{D}_{1-n}^N)}\left[\ln\left(\prod_{k=1}^{M} f(y_k|h_k,\theta,d,z_k)\right)\right]\\ &+\mathsf{E}_{\tilde{f}(\alpha|\mathcal{D}_{1-n}^N)}\left[\ln\left(f(z|\alpha)\right)\right]\Bigg\}\propto\prod_{i=1}^{M}\prod_{k=1}^{N}\exp\left\{a_{k,i}\right\}^{z_{k,i}}\end{aligned} \tag{4.16}$$

kde člen $a_{k,i}$ zapouzdřující expektace

$$\begin{aligned}a_{k,i} =&\frac{1}{2}\mathsf{E}_{\tilde{f}(\theta,d|\mathcal{D}_{1-n}^N)}\left[\ln(d)-d(h_k'\theta_i-y_k)^2\right]+\widehat{\ln(\alpha_i)}.\\ =&b_{k,i}+\frac{1}{2}\Psi\left(\frac{\nu}{2}\right)-\frac{1}{2}\ln\left(\frac{\Sigma}{2}\right)-\Psi\left(\sum_{i=1}^{M}\tau_i\right)\end{aligned} \tag{4.17}$$

můžeme vyjádřit užitím postačujících statistik ostatních VB marginálií, jak uvádí 4.17, přičemž

$$b_{k,i}=-\frac{1}{2}\frac{\nu}{\Sigma}\left(h_k'P_ih_k+\left(h_k'\hat{\theta}_i-y_k\right)^2\right)+\Psi\left(\tau_i\right). \tag{4.18}$$

Je patrné, že z bude opět produktem $\tilde{f}(z_k|\mathcal{D}_{1-n}^N)$, jež pochází z kategorických rozložení

$$\tilde{f}(z|\mathcal{D}_{1-n}^N)=\prod_{k=1}^{N}\tilde{f}(z_k|\mathcal{D}_{1-n}^N)=\prod_{k=1}^{N}\mathcal{C}_{z_k}\left(\gamma_k\right). \tag{4.19}$$

Abychom však splnili podmínky $\gamma_k\in\Delta_M\ \forall k$ kladené na postačující statistiky, je nutné výrazy umocněné na $z_{k,i}$ normalizovat.

$$\gamma_k=\frac{\exp\left\{a_{k,i}\right\}}{\sum_{i=1}^{M}\exp\left\{a_{k,i}\right\}}=\frac{\exp\left\{b_{k,i}\right\}}{\sum_{i=1}^{M}\exp\left\{b_{k,i}\right\}} \tag{4.20}$$

Krátce odběhněme od teoretické roviny k implementační a poznamenejme, že při numerickém hledání γ_k dle 4.20 může vlivem omezené numerické přesnosti dojít k nestabilitě ve výpočtu. Pokud jsou hodnoty $b_{k,i}\gtrsim 100$ řádově v kladných stovkách, pak se výraz 4.20 numericky blíží k tvaru $\frac{\infty}{\infty}$ a naopak pro $b_{k,i}\lesssim -100$ k výrazu $\frac{0}{0}$. Uvedené přibližné meze nestability se samozřejmě odvíjí od použité výpočetní přesnosti.

Vynásobení čitatele i jmenovatele 4.20 výrazem $\exp\{m_k\}$ nezmění jeho hodnotu, kdy volbou

$$m_k\equiv-\max_i(b_{k,i}) \tag{4.21}$$

docílíme, že jmenovatel výrazu

$$\gamma_k=\frac{\exp\left\{b_{k,i}+m_k\right\}}{\sum_{i=1}^{M}\exp\left\{b_{k,i}+m_k\right\}} \tag{4.22}$$

bude vždy v intervalu $[1,M]$ a jeho čitatel v $[0,1]$. Shodnou implementační úpravu je potřeba provést i pro algoritmus popsaný v následující kapitole 4.3.

VB pro α

$$\begin{aligned}\tilde{f}(\alpha|\mathcal{D}_{1-n}^{N}) &\propto \exp\left\{\mathsf{E}_{\tilde{f}(z|\mathcal{D}_{1-n}^{N})}\left[\ln\left(f(z|\alpha)\right)\right]\right\} f(\alpha) \\ &\propto \prod_{i=1}^{M} \alpha_i^{\tau_{0,i}-1+\sum_{k=1}^{N}\hat{z}_{k,i}}\end{aligned} \tag{4.23}$$

VB marginálie náhodné veličiny α bude rovněž nabývat stejné funkční formy jako její prior $f(\alpha)$. Postačující statistiky tohoto Dirichletova rozložení

$$\tilde{f}(\alpha|\mathcal{D}_{1-n}^{N}) = \mathcal{D}_{\alpha}(\tau) \tag{4.24}$$

jsou dány dle

$$\tau = \tau_0 + \sum_{k=1}^{N} \hat{z}_{k,i}, \tag{4.25}$$

kde expektaci nahradíme $\hat{z}_{k,i} = \gamma_{k,i}$.

4.2.3 Algoritmus

Algoritmus 3: Algoritmus jednokrokové identifikace přepínaných ARX modelů zatížených šumem o společné přesnosti

Inicializace:

1 Nastav horní limit počtu ARX modelů M.
2 Nastav maximální počet iterací IVB algoritmu r_{max} a koeficient pro jeho předčasné ukončení q.
3 Proveď $n + N$ měření pro naplnění regresorů h_k a zisk y_k $\forall k \in [1, N]$.
4 $\forall i \in [1, M]$ inicializuj hodnoty apriorních statistik $\hat{\theta}_{0,i}, P_{0,i}, \Sigma_0, \nu_0, \tau_0$.
5 $\forall i \in [1, M]$ inicializuj hodnoty statistik IVB algoritmu $\hat{\theta}_i^{[0]}$, $P_i^{[0]}$, $\Sigma^{[0]}$, $\nu^{[0]}$, $\tau^{[0]}$ a $\forall k \in [1, N]$ $\gamma_{k,i}^{[0]} = 1/M$

Offline učení: **Rovnice:**

for $r \leftarrow 1, r_{max}$ **do**
6 $\quad \forall k \in [1, N]$ vypočti $b_{k,i}^{[r]}$ a $m_k^{[r]}$. ▷ 4.18, 4.21
7 $\quad \forall k \in [1, N]$ vypočti $\gamma_k^{[r]}$. ▷ 4.22
8 $\quad$ Vypočti $\tau^{[r]}$. ▷ 4.25
9 $\quad \forall i \in [1, M]$ vypočti $P_i^{[r]}, \hat{\theta}_i^{[r]}, \nu^{[r]}, \Sigma^{[r]}$. ▷ 4.15a - 4.15d
10 $\quad$ **if** $\sum_{k=1}^{N}\sum_{i=1}^{M} |\gamma_{k,i}^{[r-1]} - \gamma_{k,i}^{[r]}| < q$ **then**
11 $\quad\quad$ **break for loop**
$\quad$ **end**
end

4.3 Identifikace přepínaných ARX modelů zatížených šumy o obecně různé přesnosti

Nyní představíme algoritmus, který nemusí být omezen předpokladem jednotností šumu pro všechny modely a bude tak skutečně pracovat s 4.3. Odlišná volba prioru oproti kapitole 4.2 zajistí odlišný výpočet postačujících statistik potřebných pro určení identit výstupů z_k a nedojde již k nechtěnému nárůstu odhadované variance šumu pro jeden z modelů a následné přiřazení nepříslušících dat tomuto modelu [8].

4.3.1 Volba prioru

Prior $f(\theta|\beta)$ a k němu konjungovaný prior $f(\beta)$ jsou voleny shodně jako v [1]. Náhodným veličinám d, z, α jsou přiřazeny apriorní formy shodně jako v kapitole 4.2.

$$f(\theta|\beta) = \prod_{i=1}^{M} \mathcal{N}_{\theta_i}\left(\hat{\theta}_{0,i}, \beta_i I_p\right) \tag{4.26a}$$

$$f(\beta) = \prod_{i=1}^{M} \mathcal{W}_{\beta_i}\left(\xi_{0,i}, \upsilon_{0,i}\right) \tag{4.26b}$$

$$f(d) = \prod_{i=1}^{M} \mathcal{W}_{d_i}\left(\Sigma_{0,i}, \nu_{0,i}\right) \tag{4.26c}$$

$$f(z|\alpha) = \prod_{k=1}^{N} f(z_k|\alpha) = \prod_{k=1}^{N} \mathcal{C}_{z_k}\left(\alpha\right) \tag{4.26d}$$

$$f(\alpha) = \mathcal{D}_{\alpha}\left(\tau_0\right) \tag{4.26e}$$

4.3.2 Vyjádření VB marginálií

Sdruženou hustotu můžeme až na $\prod_{k=1}^{M} f(u_k|\mathcal{D}_{1-n}^{k-1})$, jež představuje pro inferované náhodné veličiny $\Theta \equiv \{\theta, \beta, d, \alpha\}$, z konstantu, vyjádřit 4.8, 4.3 a 4.26a - 4.26d nyní algebraicky.

$$\begin{aligned} f(\theta, \beta, d, z, \alpha, \mathcal{D}_{1-n}^{N}) = \prod_{k=1}^{M} \left\{ f(y_k|h_k, \theta, d, z_k) f(u_k|\mathcal{D}_{1-n}^{k-1}) \right\} \\ \times f(\theta|\beta) f(\beta) f(d) f(z|\alpha) f(\alpha). \end{aligned} \tag{4.27}$$

Množinu $\breve{f}^*(\Theta, z|\mathcal{D}_{1-n}^{N})$, představující potenciální kandidáty na aproximaci, budeme definovat jako

$$\begin{aligned} \breve{f}^*(\Theta, z|\mathcal{D}_{1-n}^{N}) \equiv \Big\{ f(\Theta, z|\mathcal{D}_{1-n}^{N}) \Big| f(\Theta, z|\mathcal{D}_{1-n}^{N}) = f(\theta|\mathcal{D}_{1-n}^{N}) f(\beta|\mathcal{D}_{1-n}^{N}) \\ \times f(d|\mathcal{D}_{1-n}^{N}) f(\alpha|\mathcal{D}_{1-n}^{N}) f(z|\mathcal{D}_{1-n}^{N}) \Big\}. \end{aligned} \tag{4.28}$$

Tato volba bude opět patrná z algebraických vyjádření VB margináliích uvedených v následujících podkapitolách.

VB pro θ

$$\begin{aligned}\tilde{f}(\theta|\mathcal{D}_{1-n}^N) &\propto \exp\left\{\mathsf{E}_{\tilde{f}(z|\mathcal{D}_{1-n}^N)}\left[\ln\left(\prod_{k=1}^M f(y_k|h_k,\theta,d,z_k)\right)\right]+\mathsf{E}_{\tilde{f}(\beta|\mathcal{D}_{1-n}^N)}\left[\ln\left(f(\theta|\beta)\right)\right]\right\}\\ &\propto \prod_{i=1}^M \exp\left\{-\frac{1}{2}\begin{bmatrix}\theta_i\\1\end{bmatrix}'\begin{bmatrix}\hat{\beta}_i I_p+\sum_{k=1}^N \hat{z}_{k,i}\hat{d}_i h_k h_k' & -\hat{\beta}_i\hat{\theta}_{0,i}-\sum_{k=1}^N \hat{z}_{k,i}\hat{d}_i h_k y_k\\ -\hat{\beta}_i\hat{\theta}_{0,i}'-\sum_{k=1}^N \hat{z}_{k,i}\hat{d}_i h_k' y_k & \hat{\beta}_i\hat{\theta}_{0,i}'\hat{\theta}_{0,i}+\sum_{k=1}^N \hat{z}_{k,i}\hat{d}_i {y_k}^2\end{bmatrix}\begin{bmatrix}\theta_i\\1\end{bmatrix}\right\}\end{aligned} \tag{4.29}$$

Vidíme, že VB magrinální hustota pro θ bude nabývat opět produktu normálních rozložení. Můžeme si povšimnout, že z matice v 4.35 není možné vytknout ani $\hat{d}_i$ a ani $\hat{\beta}_i$, což poukazuje na nutnost separace θ do samostatné VB marginální hustoty [8].

$$\tilde{f}(\theta|\mathcal{D}_{1-n}^N)=\prod_{i=1}^M \mathcal{N}_{\theta_i}\left(\hat{\theta}_i,P_i\right) \tag{4.30}$$

Postačující statistiky tohoto rozložení jsou

$$P_i=\left(\hat{\beta}_i I_p+\sum_{k=1}^N \hat{z}_{k,i}\hat{d}_i h_k h_k'\right)^{-1} \tag{4.31a}$$

$$\hat{\theta}_i=P_i\left(\hat{\beta}_i\hat{\theta}_{0,i}+\sum_{k=1}^N \hat{z}_{k,i}\hat{d}_i h_k y_k\right), \tag{4.31b}$$

kde $\hat{z}_{k,i}=\gamma_{k,i}$, $\hat{d}_i=\nu_i/\Sigma_i$ a $\hat{\beta}_i=\upsilon_i/\xi_i$.

VB pro β

$$\begin{aligned}\tilde{f}(\beta|\mathcal{D}_{1-n}^N) &\propto \exp\left\{\mathsf{E}_{\tilde{f}(\theta|\mathcal{D}_{1-n}^N)}\left[\ln\left(f(\theta|\beta)\right)\right]\right\}f(\beta)\\ &\propto \prod_{i=1}^M \beta_i^{\frac{\upsilon_{0,i}-2+p}{2}}\exp\left\{-\frac{\beta_i}{2}\mathsf{E}_{\tilde{f}(\theta|\mathcal{D}_{1-n}^N)}\left[\left(\theta_i-\hat{\theta}_{0,i}\right)'\left(\theta_i-\hat{\theta}_{0,i}\right)\right]\right\}\end{aligned} \tag{4.32}$$

VB marginální hustota nabývá formy produktu Wishartových rozložení

$$\tilde{f}(\beta|\mathcal{D}_{1-n}^N)=\prod_{i=1}^M \mathcal{W}_{\beta_i}\left(\xi_i,\upsilon_i\right) \tag{4.33}$$

s postačujícími statistikami

$$\xi_i=\xi_{0,i}\ +\widehat{\theta_i'\theta_i}-2\hat{\theta}_{0,i}'\hat{\theta}_i+\hat{\theta}_{0,i}'\hat{\theta}_{0,i}, \tag{4.34a}$$

$$\upsilon_i=\upsilon_{0,i}+p. \tag{4.34b}$$

Druhý obecný moment θ_i vyjádříme jako $\widehat{\theta_i'\theta_i}=\hat{\theta}_i'\hat{\theta}_i+\mathrm{tr}(P_i)$.

VB pro d

$$\begin{aligned}\tilde{f}(d|\mathcal{D}_{1-n}^N) &\propto \exp\left\{\mathsf{E}_{\tilde{f}(\theta|\mathcal{D}_{1-n}^N)\tilde{f}(z|\mathcal{D}_{1-n}^N)}\left[\ln\left(\prod_{k=1}^M f(y_k|h_k,\theta,d,z_k)\right)\right]\right\} f(d) \\ &\propto \prod_{i=1}^M d_i^{\frac{\nu_{0,i}-2+\hat{z}_{k,i}}{2}} \exp\left\{-\frac{d_i}{2}\left(\Sigma_{0,i}+\hat{z}_{k,i}\mathsf{E}_{\tilde{f}(\theta|\mathcal{D}_{1-n}^N)}\left[(h_k'\theta_i-y_k)^2\right]\right)\right\}\end{aligned} \tag{4.35}$$

Rovněž i VB hustota $\tilde{f}(d|\mathcal{D}_{1-n}^N)$ zachovává své rozložení vzhledem k prioru $f(d)$ a pochází z produktu Wishartových rozložení

$$\tilde{f}(d|\mathcal{D}_{1-n}^N) = \prod_{i=1}^M \mathcal{W}_{d_i}(\Sigma_i,\nu_i) \tag{4.36}$$

popsaném postačujícími statistikami

$$\Sigma_i = \Sigma_{0,i} + \sum_{k=1}^N \hat{z}_{k,i}\left(h_k'\,\widehat{\theta_i\theta_i'}h_k - 2h_k'\hat{\theta}_i y_k + y_k^2\right), \tag{4.37a}$$

$$\nu_i = \nu_{0,i} + \sum_{k=1}^N \hat{z}_{k,i}. \tag{4.37b}$$

Momenty můžeme opět nahradit $\hat{z}_{k,i} = \gamma_{k,i}, \widehat{\theta_i\theta_i'} = \hat{\theta}_i\hat{\theta}_i' + P_i$.

VB pro z, α

VB marginálie $\tilde{f}(z|\mathcal{D}_{1-n}^N)$ a $\tilde{f}(\alpha|\mathcal{D}_{1-n}^N)$ budou vyjádřeny shodně jako v kapitole 4.2, přičemž při výpočtu členů $a_{k,i}$ budeme v 4.17 namísto d uvažovat d_i a nahradíme pak statistiky Σ, ν statistikami Σ_i a ν_i [8].

Výraz $a_{k,i}$

$$\begin{aligned}a_{k,i} =& \frac{1}{2}\mathsf{E}_{\tilde{f}(\theta,d|\mathcal{D}_{1-n}^N)}\left[\ln(d_i) - d_i(h_k'\theta_i - y_k)^2\right] + \widehat{\ln(\alpha_i)}. \\ =& b_{k,i} - \Psi\left(\sum_{i=1}^M \tau_i\right)\end{aligned} \tag{4.38}$$

opět vyjádříme pomocí $b_{k,i}$

$$b_{k,i} = \frac{1}{2}\Psi\left(\frac{\nu_i}{2}\right) - \frac{1}{2}\ln\left(\frac{\Sigma_i}{2}\right) - \frac{1}{2}\frac{\nu_i}{\Sigma_i}\left(h_k'P_ih_k + \left(h_k'\hat{\theta}_i - y_k\right)^2\right) + \Psi(\tau_i). \tag{4.39}$$

Postačující statistiky γ_k obdržíme shodně jako v kapitole 4.2 vztahy 4.21 a 4.22, kdy však dosazujeme $b_{k,i}$ dle 4.39. Užitím tohoto výsledku pak získáme i statistiku τ opět pomocí již uvedeného vztahu 4.25.

4.3.3 Algoritmus

Algoritmus 4: Algoritmus jednokrokové identifikace přepínaných ARX modelů zatížených šumy o obecně různé přesnosti

Inicializace:

Nastav horní limit počtu ARX modelů M.

Nastav maximální počet iterací IVB algoritmu r_{max} a koeficient pro jeho předčasné ukončení q.

Proveď $n + N$ měření pro naplnění regresorů h_k a zisk y_k $\forall k \in [1, N]$.

$\forall i \in [1, M]$ inicializuj hodnoty apriorních statistik $\hat{\theta}_{0,i}$, $P_{0,i}$, $\xi_{0,i}$, $\upsilon_{0,i}$, $\Sigma_{0,i}$, $\nu_{0,i}$, τ_0.

$\forall i \in [1, M]$ inicializuj hodnoty statistik IVB algoritmu $\hat{\theta}_i^{[0]}$, $P_i^{[0]}$, $\xi_i^{[0]}$, $\upsilon_i^{[0]}$, $\Sigma_i^{[0]}$, $\nu_i^{[0]}$, $\tau^{[0]}$.

Offline učení:	**Rovnice:**
for $r \leftarrow 1, r_{max}$ **do**	
$\forall k \in [1, N]$ vypočti $b_{k,i}^{[r]}$ a $m_k^{[r]}$.	▷ 4.39, 4.21
$\forall k \in [1, N]$ vypočti $\gamma_k^{[r]}$.	▷ 4.22
Vypočti $\tau^{[r]}$.	▷ 4.25
$\forall i \in [1, M]$ vypočti $P_i^{[r]}, \hat{\theta}_i^{[r]}, \nu_i^{[r]}, \Sigma_i^{[r]}$.	▷ 4.15a - 4.15d
$\forall i \in [1, M]$ vypočti $\xi_i^{[r]}, \upsilon_i^{[r]}$.	▷ 4.34a, 4.34b
if $\sum_{k=1}^{N} \sum_{i=1}^{M} \lvert\gamma_{k,i}^{[r-1]} - \gamma_{k,i}^{[r]}\rvert < q$ **then**	
break for loop	
end	
end	

4.4 Další aproximace pro dosažení kunjugace

Vzhledem k užité aproximaci geometrickým váženým průměrem, jež je ekvivalentní původnímu řešenému problému jen za omezujícího předpokladu kladeného na exkluzivitu generování výstupů y_k, bylo nasnadě nalézt odlišný způsob aproximace. Pro dosažení kýžené konjugace je potřeba opět namísto 4.4 užít aproximaci z dynamické exponenciální rodiny. Přestože optimum $\tilde{f}(y_k)$ níže voleného funkcionálu pro obdržení unimodální hustoty popisující y_k není vzhledem k θ, d konjugativní, uvádíme zde alespoň základní myšlenku a důsledek, který vyšetřovaná aproximace má.

Předem podotkněme, že pro přehlednost notace bude místy opět opomíjena podmíněnost náhodné veličiny y_k.

4.4.1 Důsledky původní aproximace

Nechtěné chování pozorované při volbě prioru $f(\theta_i|d_i) = \mathcal{N}_{\theta_i}(\hat{\theta}_{0,i}, P_{0,i}/d_i)$ u prvního z algoritmů diskutovaném v kapitole 4.2 plyne z tvaru funkcionálu $\tilde{f}(y_k) = \arg\min_{\breve{f}(y_k)} \mathsf{E}_{f(z_k)}[KL(\breve{f}(y_k)\|f(y_k))]$, jehož minimalizací obdržíme konjungovaný model 4.3 z původního 4.4. Toto nežádoucí chování můžeme přisuzovat faktu plynoucího z KLD, že minimalizovaný funkcionál nabývá velmi vysokých hodnot pro takové $\breve{f}(y_k)$, pro které platí $\int_{y_c^*} \breve{f}(y_k)\mathrm{d}y_k \gg \int_{y_c^*} f(y_k)\mathrm{d}y_k$, kde $y_c^* \subset y^* = \mathbb{R}$ a je roven nekonečnu pro $\int_{y_c^*} \breve{f}(y_k)\mathrm{d}y_k > \int_{y_c^*} f(y_k)\mathrm{d}y_k = 0$. Tento důsledek pak zajistí, že odhadovaná hodnota přesnosti šumu $\hat{d}_i$ výstupu jednoho z modelů je oproti skutečné hodnotě menší a jako optimální aproximace GM se jeví jediný ARX model s podhodnocenou přesností šumu.

4.4.2 Volba optimalizačního funkcionálu

Změnou pořadí argumentů v KLD ztrátového funkcionálu 4.5 se tohoto nutného důsledku v aproximaci $\tilde{f}(y_k)$ zbavíme. Aproximace GM jedinou hustotou bude nyní dána optimalizací

$$\tilde{f}(y_k) = \arg\min_{\breve{f}(y_k)} \mathsf{E}_{f(z_k)} \left[KL\Big(f(y_k)\Big\|\breve{f}(y_k)\Big)\right]. \tag{4.40}$$

Lze dokázat, že pro normální rozložení se řešení optimalizačního problému 4.40 redukuje na techniku ekvivalence momentů ("*Moment matching*") [7]. Z rovnosti

$$\mathsf{E}_{\sum_{i=1}^{M} z_{k,i}\mathcal{N}_{y_k}(h_k'\theta_i, d_i)} \left[[y_k^2, y_k]'\right] = \mathsf{E}_{\mathcal{N}_{y_k}(\mu_{k,A}, d_{k,A})} \left[[y_k^2, y_k]'\right] \tag{4.41}$$

plyne, že postačující statistiky $\mu_{k,A}, d_{k,A}$ optima $\tilde{f}(y_k|h_k, \theta, d, z) = \mathcal{N}_{y_k}(\mu_{k,A}, d_{k,A})$ budou ve tvaru

$$\mu_{k,A} = \sum_{i=1}^{M} z_{k,i} h_k' \theta_i, \tag{4.42a}$$

$$d_{k,A} = \left(\sum_{i=1}^{M} z_{k,i} \left((h_k'\theta_i)^2 + \frac{1}{d_i} \right) - \left(\sum_{i=1}^{M} z_{k,i} h_k' \theta_i \right)^2 \right)^{-1}. \tag{4.42b}$$

Je patrné, že výsledek logaritmu tohoto nového rozložení již nebude separovatelný v parametrech vzhledem θ a d a VB marginálie tak nebudou nabývat standardních funkčních forem.

Marginální hustoty ve standardních formách bychom mohli obdržet užitím dalšího z nástrojů bayesovského světa, jako je například fixní forma (*Fixed form*) VB aproximace. Vyšetření tohoto řešení však není součástí této práce.

5 Experimenty

Abychom mohli posoudit funkčnosti uvedených algoritmů, pozorovat jejich chování a poukázat na nectnosti, je nezbytné provést sadu experimentů. Pro vskutku věrohodné srovnání algoritmů by bylo potřeba provést experimentů celou řadu. Rovněž aspektů, kritérií či nepříznivých vlivů, které bychom chtěli vyšetřovat, existuje nepřeberné množství. Navzdory této bohaté nabídce bylo zapotřebí zvolit takové podmínky experimentů a posuzovat takové vlastnosti algoritmů, které poskytnou patřičný vhled do jejich chování a jejich schopností, zatímco si diskuze získaných výsledků zachová svou kompaktnost.

Měření na reálné soustavě bude komplementováno simulacemi na analytickém modelu reálné soustavy. Simulace poslouží nejen pro porovnání s výsledky dosaženými na reálné soustavě, ale především jimi otestujeme robustnost algoritmů na různé velikosti přesností šumů, jejichž velikosti bychom na reálné soustavě ovlivňovali stěží.

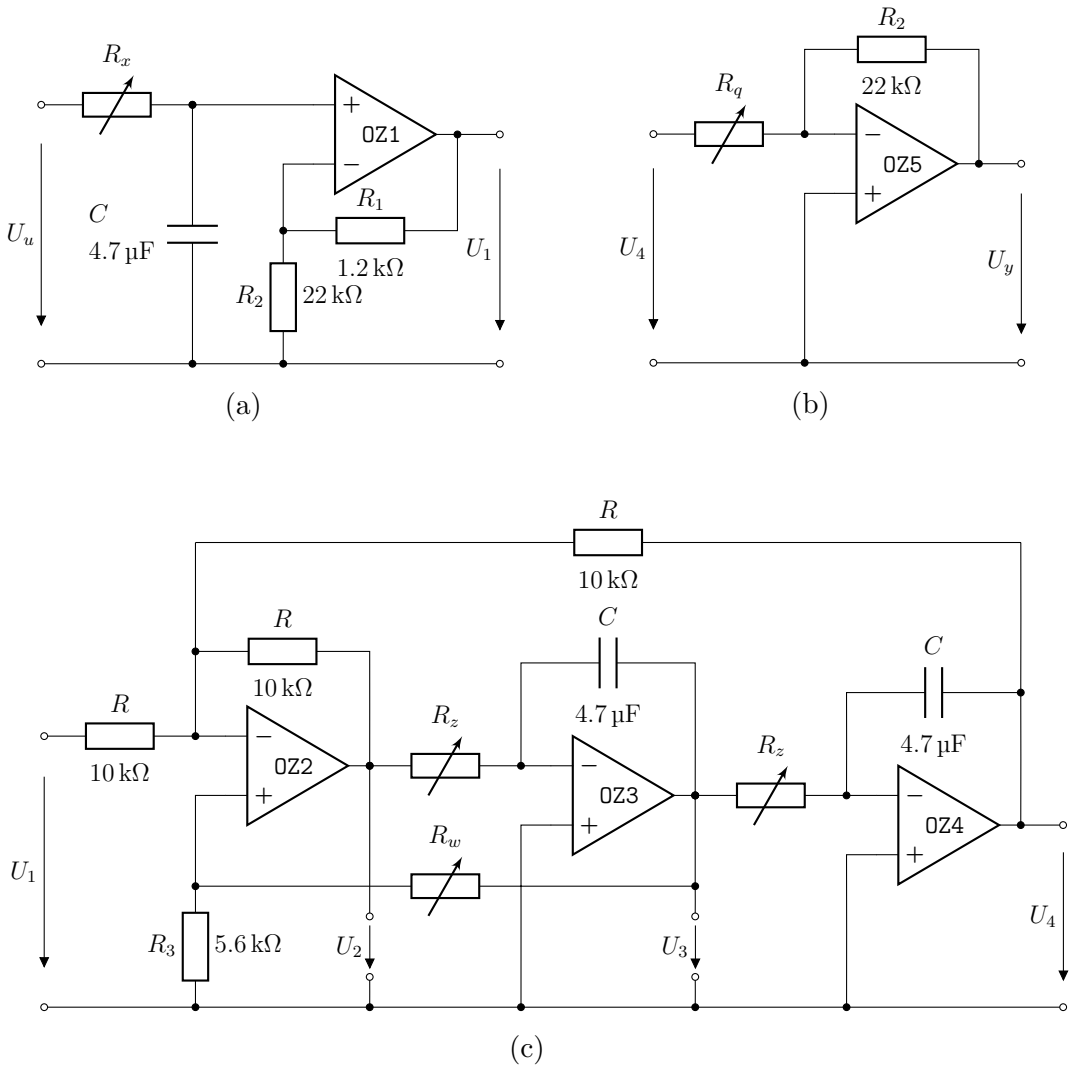
5.1 Reálná soustava

Reálnou soustavou, jež sloužila pro testování algoritmů, byl elektrický systém tvořený sadou kondenzátorů, rezistorů a operačních zesilovačů (OZ) doplněných o další komponenty pro zajištění funkčnosti, které ale z hlediska dynamického chování systému nejsou podstatné. Elektrické schéma tohoto systému je zakresleno na obrázku 5.1. Vstupem systému je napětí U_u, výstupem je napětí U_y. Stavová napětí U_1, U_3 a U_4 spolu s napětím U_2 vzhledem ke zvolenému vstupně-výstupnímu popisu pro naše účely zpracovávány nejsou. Jejich vyznačení nám však poslouží pro účely analytického odvození operátorového přenosu celé soustavy. Pro dosažení proměnného dynamického chování jsou v zapojení i rezistory s proměnnými odpory. Změna jejich odporů je prováděna skokově otočnými přepínači, konkrétní použité hodnoty budou zmíněny až při popisu experimentů v kapitole 5.1.2.

Ke značení užitém ve schématu 5.1 podotkněme, že navzdory zaběhlé konvenci byla součástkám se shodnými nominálními hodnotami přiřazena shodná označení. Toto značení bylo zvoleno pro zjednodušení při vyjadřování přenosu soustavy.

5.1.1 Operátorový přenos

Pro odvození operátorového přenosu uvažujme ideálnost všech prvků v níže uvedeném schématu. Předpokládané nekonečné zesílení OZ zajistí při jejich zpětnovazebním zapojení, že diferenční napětí $U_{\mathrm{OZ}_+} - U_{\mathrm{OZ}_-}$ mezi neinvertujícím a invertujícím vstupem bude nulové. Nekonečná vstupní impedance OZ pak způsobí, že do jejich



Obr. 5.1: Elektrické schéma reálného dynamického systému [19]

vstupů neteče žádný proud. Jejich nulová výstupní impedance nám dále zajišťuje, že výstupní napětí OZ bude ovlivněno pouze jeho vstupy a jednotlivé stupně zapojení jsou tak impedančně odděleny. Tyto důsledky předpokladu ideálnosti OZ nám zjednoduší následující odvození.

Napětími U_{OZ_+} a U_{OZ_-} označujeme velikosti napěťových potenciálů mezi neinvertujícím, respektive invertujícím, vstupem OZ vůči společné zemi, ke které jsou vztahována všechna napětí.

První část přenosu - Setrvačný článek prvního řádu

Vstupní napětí U_u je děleno impedančním děličem tvořeným R_x a C. Na výstupu OZ1 je tedy tvořeno takové napětí, které zajistí, že diferenční napětí na vstupu OZ1

je nulové. Laplaceův obraz výstupního napětí tohoto členu tedy bude

$$U_1(s) = \frac{1}{1 + R_x C s} \frac{R_1 + R_2}{R_2} U_u(s), \tag{5.1}$$

kde s je Laplaceův operátor.

Druhá část přenosu - Setrvačný článek druhého řádu

Pro určení napětí na výstupu OZ2 nejprve spočítejme napětí na jeho vstupech

$$U_{OZ2_+}(s) = \frac{R_3}{R_w + R_3} U_3(s). \tag{5.2}$$

Proudy vyvolané rozdílovými napětími $U_1 - U_{OZ2_-}$ a $U_4 - U_{OZ2_-}$ způsobí průtokem přes R ve zpětné vazbě OZ2 úbytek napětí $U_{OZ2_+} - U_2$. Za předpokladu $U_{OZ2_+} = U_{OZ2_-}$ můžeme snadno určit obraz $U_2(s)$

$$U_2(s) = 3U_{OZ2_+}(s) - U_1(s) - U_4(s) \tag{5.3}$$

Je patrné, že napětí na vstupech OZ3 bude nulové. Proud vyvolaný napětím U_2 tekoucím přes R_z způsobí na impedanci kondenzátoru C ve zpětné vazbě OZ3 úbytek napětí, jehož záporně vzatá hodnota bude rovna napětí U_3.

$$U_3(s) = -\frac{1}{R_z} \frac{1}{Cs} U_2(s) \tag{5.4}$$

Shodnou úvahou pak můžeme odvodit i

$$U_4(s) = -\frac{1}{R_z} \frac{1}{Cs} U_3(s). \tag{5.5}$$

Substituováním za obrazy napětí $U_{OZ2_+}(s), U_2(s)$ a $U_3(s)$ do 5.3 můžeme užitím 5.2, 5.4 a 5.5 vyjádřit operátorový přenos celého zapojení na obrázku 5.1c.

$$U_4(s) = \frac{-1}{C^2 R_z^2 s^2 + \frac{3R_3}{R_w + R_3} R_z C s + 1} U_1(s) \tag{5.6}$$

Podotkněme, že zda póly přenosu 5.6 budou komplexně sdružené a tento článek bude kmitavý nastavuje R_w.

Třetí část přenosu - Statické zesílení

Z obrázku 5.1b je patrné, že vztah mezi U_4 a U_y je realizován pouze proporční vazbou. V operátorové oblasti tedy bude výstup soustavy dán jako

$$U_y(s) = -\frac{R_2}{R_q} U_4(s). \tag{5.7}$$

Výsledný operátorový přenos

Užitím rovnic 5.1 a 5.3 - 5.6 můžeme vyjádřit operátorový přenos celé soustavy. Při zavedení substitucí

$$k_1 = \frac{R_1}{R_2} + 1 \tag{5.8a}$$

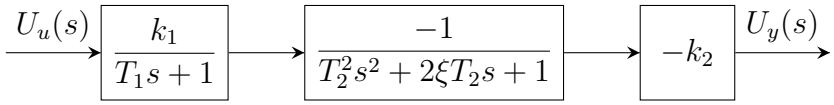
$$T_1 = R_x C \tag{5.8b}$$

$$T_2 = R_z C \tag{5.8c}$$

$$\xi = \frac{3}{2} \frac{R_3}{R_w + R_3} \tag{5.8d}$$

$$k_2 = \frac{R_2}{R_q} \tag{5.8e}$$

je možné vyjádřit vstupně-výstupní vlastnosti elektrického systému v přehledném obecném blokovém tvaru 5.2.



Obr. 5.2: Blokové schéma reálného dynamického systému [19]

Diskretizace spojitého systému

Vstupní napětí U_u je nastavováno a výstupní napětí U_y je měřeno pomocí PLC. PLC pracuje v definovaných časových okamžicích daných periodou T_s s hodnotami těchto napětí pouze v diskrétních časových krocích kT_s, kde $k \in \mathbb{N}$. Hodnotu vstupu v těchto krocích označíme jako u_k a výstupu y_k. Abychom mohli predikovat chování na základě sestaveného analytického popisu spojité soustavy v diskrétních okamžicích, provedeme diskretizaci tohoto přenosu tvarovačem nultého řádu

$$G_{tv}(s) = \frac{1 - \exp\{-sT_s\}}{s}. \tag{5.9}$$

Vstupně-výstupní popis diskretizovaného systému popsaného v z oblasti tak bude záviset nejen na nastavení proměnných rezistorů, ale také na vzorkovací periodě T_s. Analytické vyjádření tohoto přenosu uvádět nebudeme. Numericky získané diskrétní ekvivalenty konkrétních přenosů budou pro zvolená nastavení proměnných rezistorů a periody vzorkování uvedeny číselně.

5.1.2 Parametry a nastavení měření na reálné soustavě

Jak bylo zmíněno, z prakticky nepřeberného množství kombinací konfigurací a podmínek prováděných experimentů je nasnadě vybrat takovou kompaktní podmnožinu, na které budou vlastnosti a schopnosti algoritmů dostatečně patrné.

Zvolená banka přepínaných ARX modelů

Pro naše účely bude testování na reálné soustavě spočívat v identifikaci diskrétního ekvivalentu spojité elektrické soustavy s proměnlivými dynamickými vlastnostmi. Změny v této soustavě budou realizovány přepínáním hodnoty rezistoru R_z. Na tomto typu experimentu budeme moci vyzkoušet jak schopnosti rekurzivních algoritmů pro online identifikaci, tak otestovat schopnosti dvojice algoritmů pro dávkovou identifikaci přepínaných ARX modelů.

Hodnota rezistoru R_z byla přepínána z hodnoty 340 kΩ na 170 kΩ, pak na 630 kΩ a znovu na 340 kΩ, a to v časech 250 s, 500 s a 750 s. Ostatní z volitelných rezistorů zůstaly po čas experimentu neměnné s nastavenými hodnotami $R_x = 161.2\,\mathrm{k\Omega}$, $R_w = 23.2\,\mathrm{k\Omega}$ a $R_q = 17.2\,\mathrm{k\Omega}$. Počátek experimentu volíme s ohledem na stupně čitatelových a jmenovatelových polynomů diskrétních ekvivalentů $n_a = n_b = 3$ v čase $-2T_s$. Tím formálně zajistíme, že v kroku $k = 0$ budou postačující statistiky rekurentních algoritmů dány jejich apriorním nastavením a první datově informované výsledky obdržíme v kroku $k = 1$, kdy bude regresor h_k naplněn výhradně experimentálně získanými daty. Musíme však mít na paměti, že systém je potřeba začít budit nenulovým vstupem ještě před časem $-2T_s$. Perioda vzorkování byla vzhledem k dynamice systému volena $T_s = 1\,s$.

Regresní vektor a vektor parametrů deterministické části diferenční rovnice $y_k = h_k'\theta_k + e_k$ odpovídající diskrétní náhradě analytického popisu spojitého systému bude ve tvaru $h_k = [u_{k-1}, u_{k-2}, u_{k-3}, -y_{k-1}, -y_{k-2}, -y_{k-3}]'$ a $\theta_k = [b_1, b_2, b_3, a_1, a_2, a_3]'$. Časový vývoj proměnného parametru θ_k popisuje 5.10.

$$\theta_k \approx \begin{cases} [0.077,\ 0.204,\ 0.033,\ -1.644,\ 1.062,\ -0.185]'\ k \in [-2, 250] \cup [751, 1000] \\ [0.264,\ 0.604,\ 0.096,\ -0.774,\ 0.617,\ -0.129]'\ k \in [251, 500] \\ [0.024,\ 0.067,\ 0.011,\ -1.986,\ 1.280,\ -0.219]'\ k \in (501, 750] \end{cases} \tag{5.10}$$

Znak přibližné rovnosti v 5.10 plyne především z nepřesnosti analytického modelu soustavy způsobené desetiprocentní výrobní tolerancí kondenzátorů, které jsou v reálné soustavě použity.

Na této reálné časově variantní soustavě budou testovány všechny algoritmy. Simulace budou prováděny na modelu, jež popisuje 5.10. Pro jeho parametry θ_k však budeme uvažovat, že skutečné hodnoty kondenzátorů a rezistorů se shodují s jejich jmenovitými hodnotami.

Volba vstupního signálu

Pro garanci dostatečné datové informace ve vstupně-výstupních datových párech o dynamice systému, kterou chceme identifikovat, musíme volit vstupní signál o do-

statečném stupni perzistence. Praktické omezení konečnou přesností analogově digitálních (AD) a digitálně analogových (DA) převodníků však vnáší na volbu budicího signálu další požadavky.

Abychom eliminovali vlivy nelinearit typu necitlivosti způsobenými kvantizačními chybami převodníků, volíme vstupní signál ve tvaru pseudonáhodné binární posloupnosti, zkráceně PRBS (*PseudoRandom Binary Sequence*) nabývající hodnot $-2.5\,\mathrm{V}$ a $2.5\,\mathrm{V}$. Při volbě většího rozkmitu PRBS by výstup soustavy již překračoval omezený rozsahu AD převodníku $-10\,\mathrm{V}$ až $10\,\mathrm{V}$ vstupní analogové karty PLC.

Situaci bychom mohli dále idealizovat nastavením rezistoru R_q na vyšší hodnotu, v důsledku čehož bychom mohli nastavit větší rozsah budicího signálu a snížit relativní chybu ve velikosti vstupního signálu. Je ale jasné, že chyba DA převodníku ovlivní pouze přesnost získaného statického zesílení systému, nikoliv jeho dynamické vlastnosti. Volbu R_q tedy ponecháváme na původní hodnotě.

Velikost maximální kvantizační chyby je přibližně $0.3\,\mathrm{mV}$. Takt použitého PRBS generátoru tvořeného 9-bitovým posuvným registrem byl volen $1\,\mathrm{s}$, přičemž jeho běh je synchronní vzhledem k měření vstupů, nastavování výstupů a běhu rekurzivních algoritmů.

5.2 Testování rekurzivních algoritmů pro online identifikaci časově proměnného ARX modelu

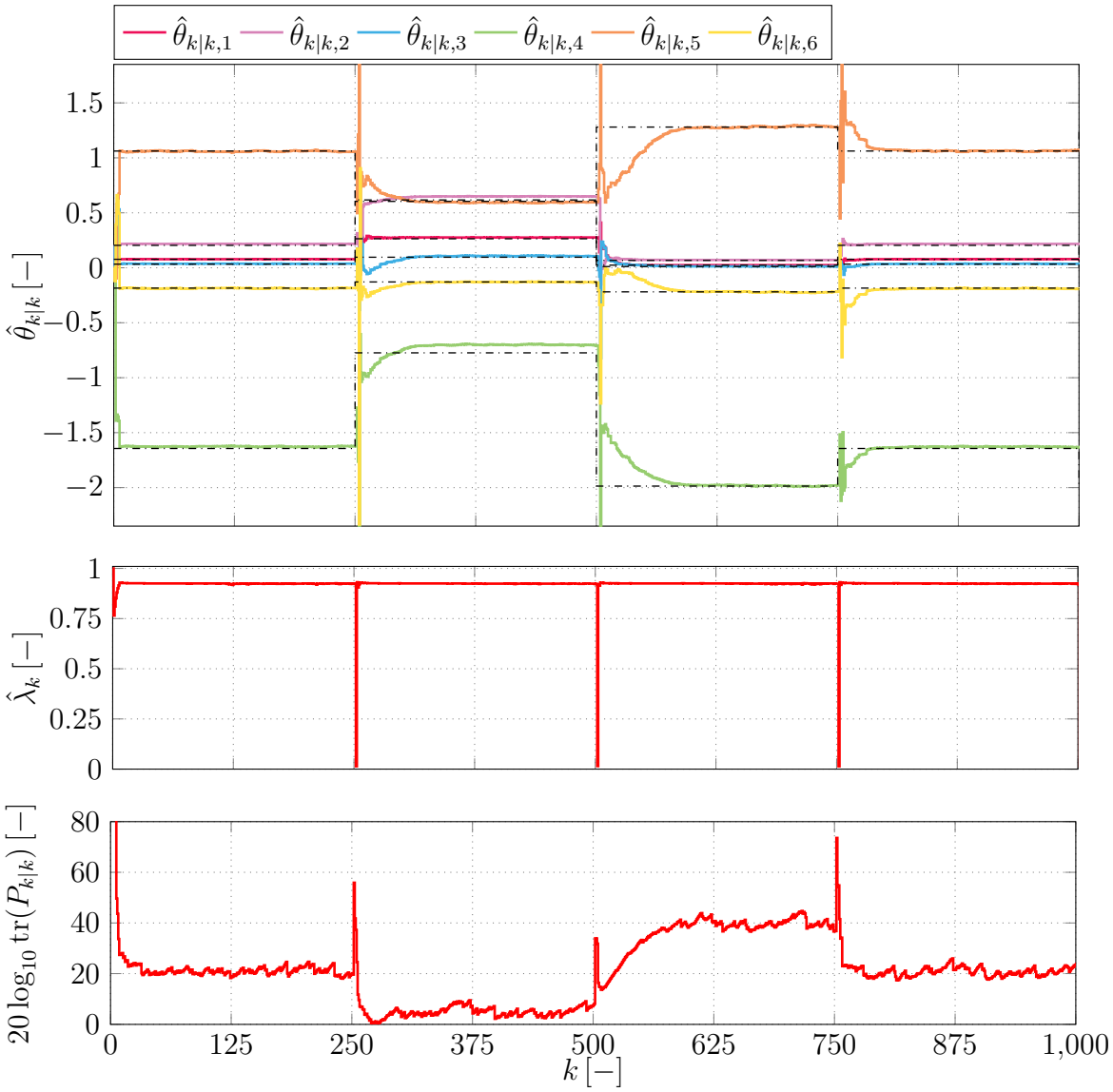
Pro všechny rekurzivní algoritmy volíme shodné apriorní nastavení $P_{0|0} = 10^6 I_p$, $\hat{\theta}_{0|0} = \mathbf{0}_p$, $\Sigma_{0|0} = 0.1$, $\nu_{0|0} = 0.1$. Externě dodávaná regularizační matice byla volena jako konstantní matice $\Xi_{k|k} = 10^{-6} I_p \; \forall k \in [1, N]$. Úrovně PRBS byly pro účely simulace voleny $-1\,\mathrm{V}$ a $1\,\mathrm{V}$, aby bylo srovnání vůči nastaveným přesnostem šumů d patrnější.

Předem podotkněme, že černé čerchované průběhy v grafech 5.3 - 5.7 odpovídající parametrům 5.10 se pro měření na reálné soustavě více či méně odlišují od skutečných hodnot. Původ této odchylky byl diskutován v kapitole 5.1.2. Pro simulace však tyto průběhy již odpovídají identifikovaným hodnotám parametrů modelu.

5.2.1 Určení koeficientu zapomínání pomocí VB metody

Pro reálnou soustavu i pro simulaci bylo užito stejného nastavení $b_0 = 0$ a $\alpha = 0.01$. Maximální počet iterací IVB algoritmu byl nastaven $r_{max} = 10$ a ukončovací podmínka pro jeho předčasné ukončení byla dána parametrem $q = 10^{-4}$. IVB algoritmus byl inicializován nastavením $\hat{\lambda}_k^{[0]} = 0.7 \, \forall k \in [1, N]$. Pro simulaci byla nastavená přesnost šumu $d = 500$.

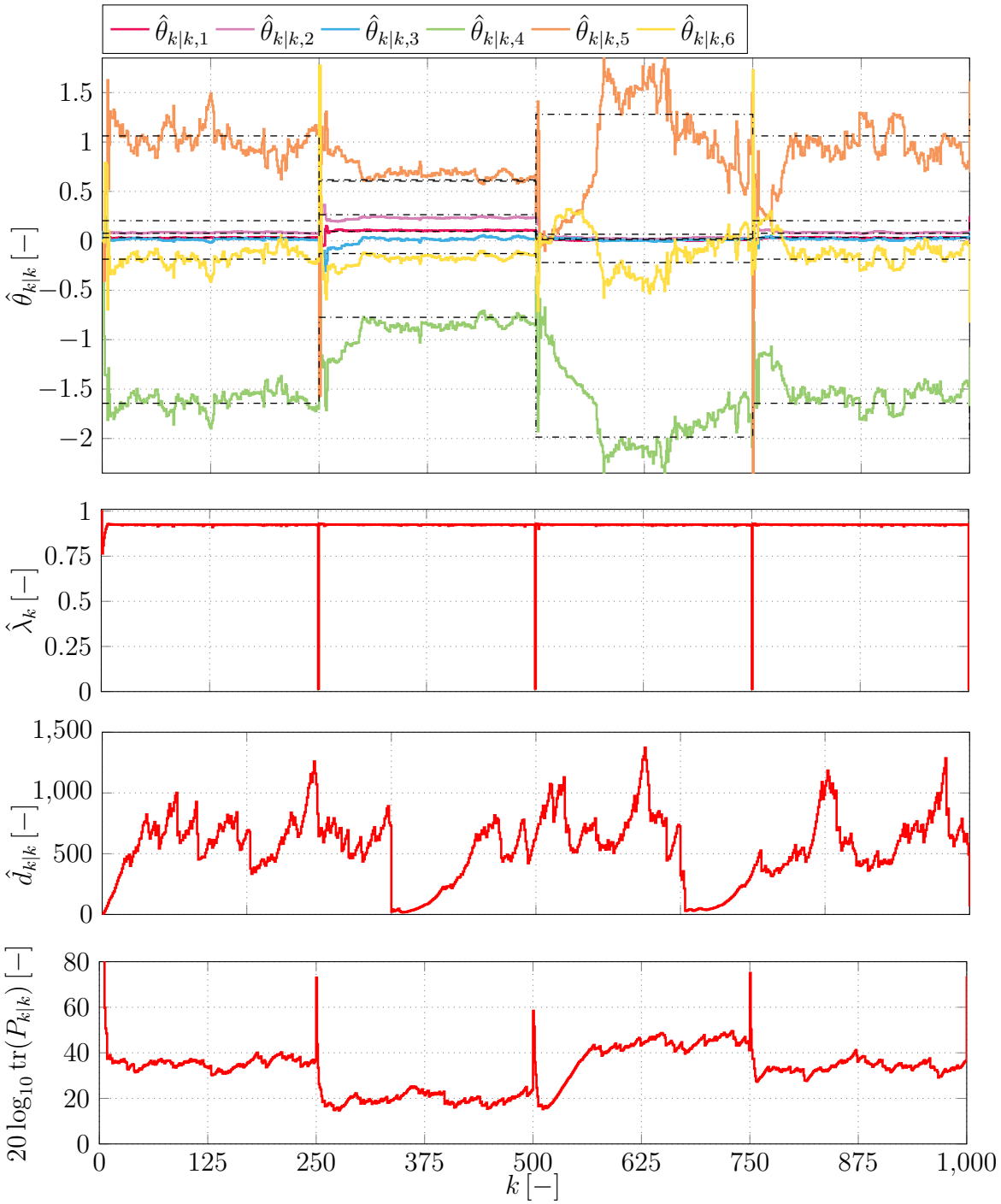
Průměrný počet iterací IVB algoritmu byl pro obě měření 4. U identifikace reálné soustavy algoritmus IVB splnil ukončovací podmínku za nejvíce 8 kroků, a to jednou během identifikace. Pro simulaci byl maximální počet iterací 10, který nastal také jen jednou.



Obr. 5.3: Identifikace reálné soustavy algoritmem určujícího koeficient zapomínání VB aproximací při parametrizaci: $\alpha = 0.01$, $b_0 = 0$, $\hat{\lambda}_k^{[0]} = 0.7 \, \forall k \in [1, N]$

Z obou obrázků 5.3, 5.4 je patrný, pro tento algoritmus charakteristický, průběh koeficientu zapomínání $\hat{\lambda}_k$. Vidíme, že mimo počátečního nastavení v $k = 0$ nedosahuje tento koeficient hodnoty 1. To je dáno tím, že $\hat{\lambda}_k$ je počítán jako první obecný moment náhodné veličiny $\lambda_k \in [\alpha, 1]$ pocházející z omezeného exponenciálního rozložení. Algoritmus tak s každým krokem zapomíná informaci uchovávanou v postačujících statistikách $\mathcal{NW}$ rozložení. Důsledkem je, že po určité době se přesnosti

odhadů $\hat{\theta}_{k|k}$ a $\hat{d}_{k|k}$ již nezvyšují, přestože nejsou rozpoznávány změny v soustavě. Z grafu 5.4 rovněž vidíme, že skutečná přesnost šumu d je v důsledku průběhu $\hat{\lambda}_k$ nadhodnocená.



Obr. 5.4: Identifikace simulované soustavy zatížené šumem o přesnosti $d = 500$ algoritmem určujícího koeficient zapomínání VB aproximací při parametrizaci: $\alpha = 0.01$, $b_0 = 0$, $\hat{\lambda}_k^{[0]} = 0.7 \, \forall k \in [1, N]$

Při poklesu pod hodnotu přesnosti šumu $d = 500$ u simulované soustavy tento algoritmus začíná selhávat při detekci změny v soustavě, což je způsobeno právě v důsledku nedostatečné důvěry v odhady $\hat{\theta}_{k|k}$ a $\hat{d}_{k|k}$.

Navýšením b_0 pro významný nárůst koeficientu $\hat{\lambda}_k$ mezi okamžiky změn v soustavě si však příliš nepomůžeme, protože navýšením postačující statistiky b_k o hodnotu b_0 se významně navýší odhad $\hat{\lambda}_k$ i pro okamžiky změn v soustavě, jak plyne z 3.25 a 3.26.

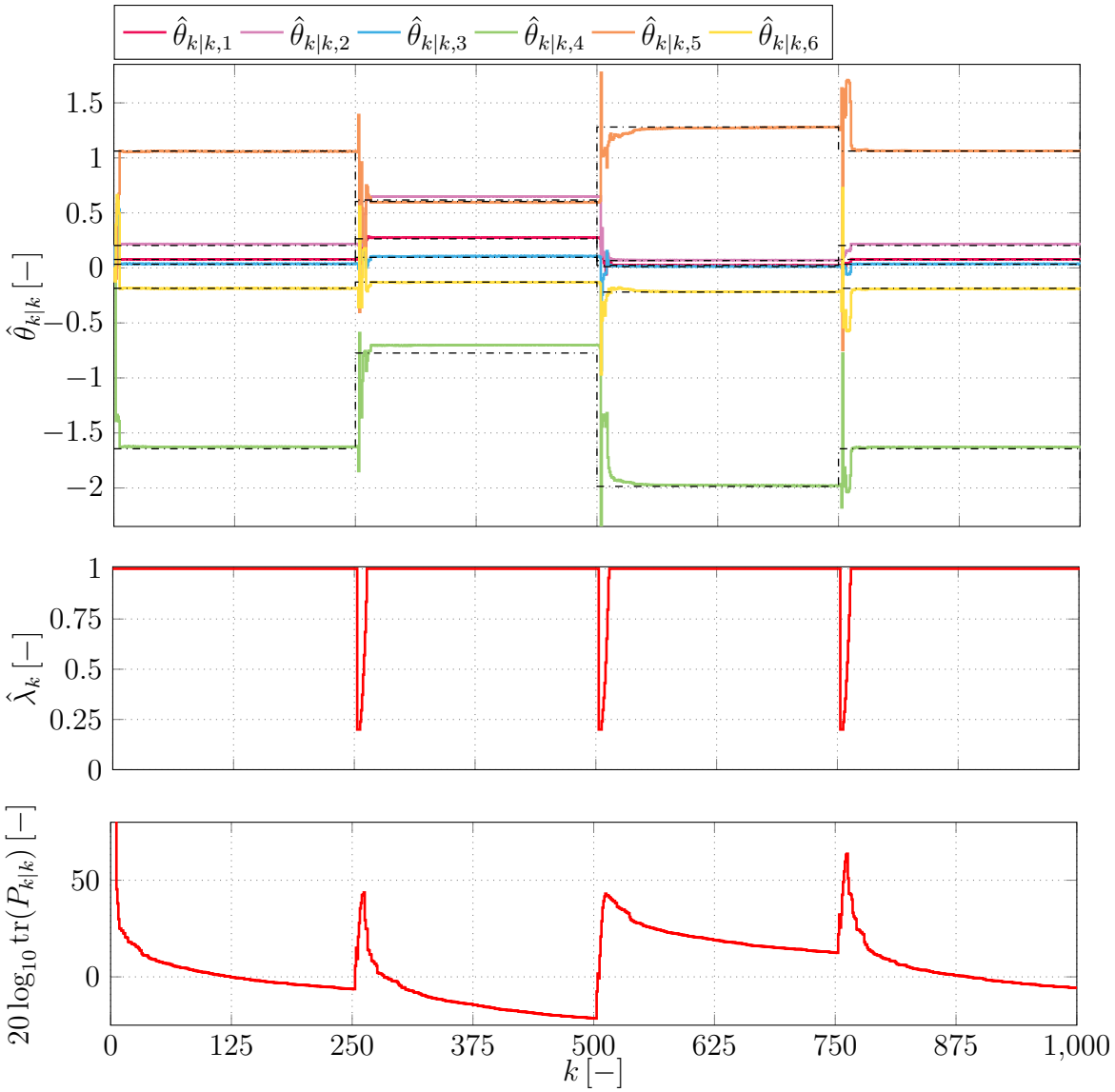
Za zmínku také stojí, že koeficient $\hat{\lambda}_k$ se po detekci změny v soustavě provázené jeho poklesem navrací již v dalším kroku zpátky na prakticky neměnnou hodnotu, která se pro obě ukázky 5.3, 5.4 pohybuje okolo hodnoty 0.92. Právě z tohoto chování plynula relativně malá hodnota volby parametru α.

5.2.2 Určení koeficientu zapomínání statistickým rozhodováním

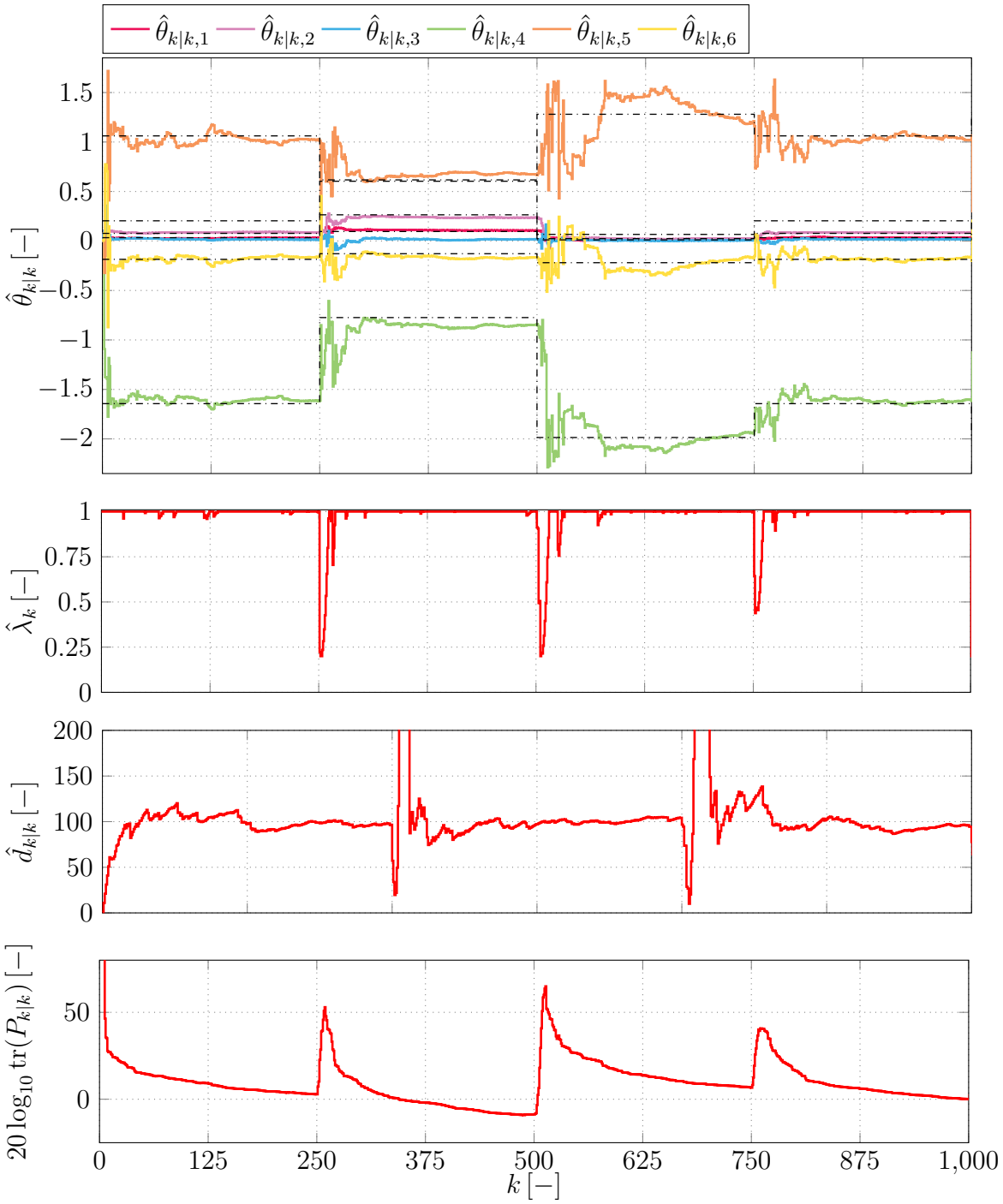
Pro reálnou soustavu byly voleny nastavení $\zeta = 0.15$ a $\rho = 0.8$. Výsledky variant $\gamma \equiv \zeta$ i $\gamma \equiv \rho$ byly při tomto nastavení velmi podobné a průběhy časových vývojů koeficientu zapomínání $\hat{\lambda}_{k|k}$ a tedy i ostatních statistik se lišily jen nepatrně. Proto pro reálnou soustavu uvádíme pouze výsledky $\gamma \equiv \zeta$. U identifikace simulované soustavy zatížené šumem s přesností $d = 1000$ vykazovala tato dvojice algoritmů při volbě $\zeta = 0.2$ a $\rho = 0.6$ odlišné výsledky, uvádíme je tedy pro obě varianty.

Charakteristický průběh koeficientu $\hat{\lambda}_k$ určovaného statistickým rozhodováním je patrný z grafu 5.5. Vidíme, že mezi okamžiky detekcí změn nabývá koeficient $\hat{\lambda}_k$ hodnoty 1, díky čemuž důvěra v odhady $\hat{\theta}_{k|k}$ a $\hat{d}_{k|k}$ roste až do další detekce změny. Důsledkem průběhu $\hat{\lambda}_k$ konverguje $\hat{d}_{k|k}$, jak plyne z grafů 5.6 a 5.7.

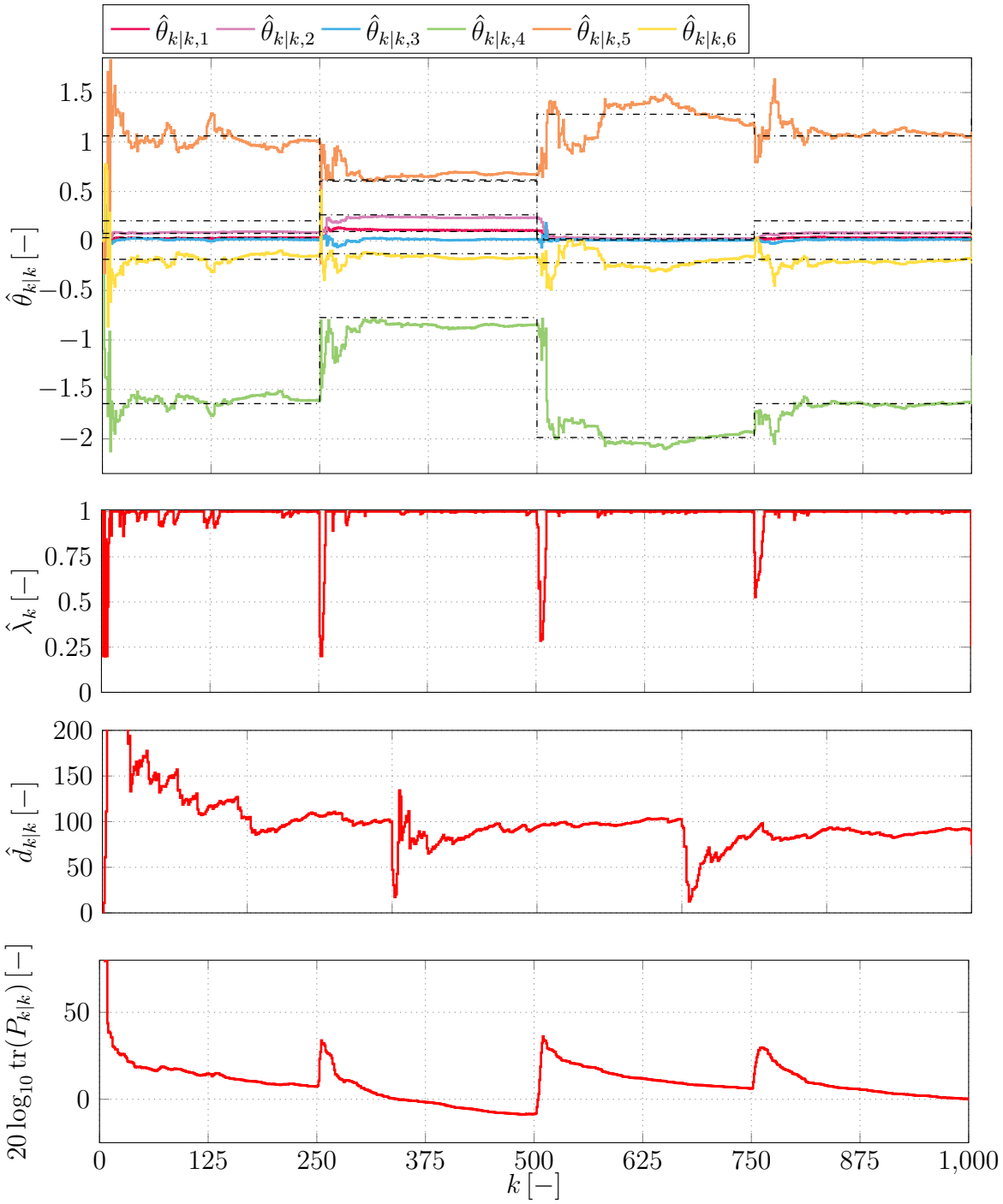
Na obrázcích 5.6 a 5.7 můžeme vidět, že při přesnosti šumu $d = 100$ obě varianty algoritmů již místy reagují na šum. Snížením parametrů ζ či ρ podle užité varianty sice částečně eliminujeme tyto nechtěné reakce, zhoršíme ale citlivosti obou algoritmů na detekci změny. Příčina tohoto chování je objasněna v kapitole 3.3.2. Rovněž hodnota parametru α byla volena jako kompromis. Volbou vyšší hodnoty bychom sice snížili citlivost na šum, zastaralá informace by však po detekci změny hrála významnější roli v novém posterioru a zhoršila by se tak rychlost adaptace.



Obr. 5.5: Identifikace reálné soustavy algoritmem určujícího koeficient zapomínání statistickým rozhodováním - varianta $\gamma \equiv \zeta$ při parametrizaci: $\alpha = 0.2$, $\zeta = 0.15$



Obr. 5.6: Identifikace simulované soustavy zatížené šumem o přesnosti $d = 100$ algoritmem určujícího koeficient zapomínání statistickým rozhodováním - varianta $\gamma \equiv \zeta$ při parametrizací: $\alpha = 0.2$, $\zeta = 0.2$

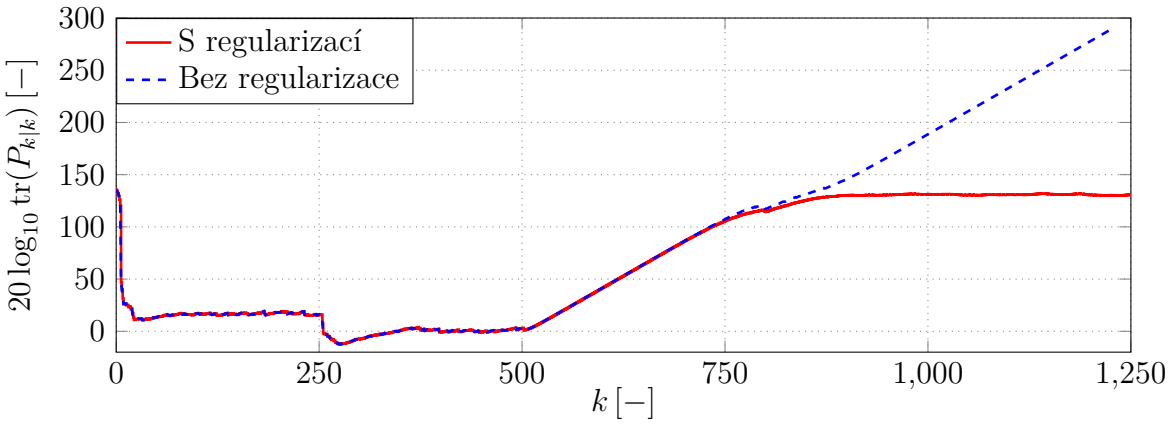


Obr. 5.7: Identifikace simulované soustavy zatížené šumem o přesnosti $d = 100$ algoritmem určujícího koeficient zapomínání statistickým rozhodováním - varianta $\gamma \equiv \rho$ při parametrizaci: $\alpha = 0.2$, $\rho = 0.6$

5.2.3 Vliv regularizace a důsledek při její absenci

Jak bylo zmíněno, regularizace zajišťuje stabilitu uvedených rekurzivních algoritmů se zapomínáním. Její význam a vliv demonstrujme na případu konstantního zapomínání s faktorem $\hat{\lambda}_k = 0.95 \; \forall k \in [1, N]$. Získané poznatky jsou ale analogicky aplikovatelné pro libovolnou formu zapomínání, kde není garantováno, že koeficient zapomínání bude roven jedné pro datově nedostatečně informovaná měření.

Na reálné soustavě bylo prováděno měření, které až do kroku $k = 500$ odpovídá tomu uvedenému výše. V kroku $k = 501$ však přepnutí neproběhlo a vstupní signál PRBS byl nahrazen konstantní hodnotou $u_k = 2.5\,\mathrm{V} \; \forall k > 500$.



Obr. 5.8: Vliv regularizace a důsledek při její absenci

U konstantního zapomínání se exponenciálně snižuje vliv datové informace s jejím stářím. Od kroku 501 se tedy začíná kvůli nedostatečně perzistentnímu vstupnímu signálu vytrácet informace o dynamickém chování systému. Exponenciální nárůst stopy matice $P_{k|k}$ odpovídá exponenciálnímu růstu jednoho z jeho vlastních čísel. Při užití regularizace se však tento růst zamezí a matice zůstává regulární. Naopak, bez regularizace tento růst pokračuje. Konec modré křivky na obrázku 5.8 představuje okamžik, kdy algoritmus bez užití regularizace numericky kolabuje. Podotkněme, že výpočet z měřených dat byl prováděn na PC, nikoliv na PLC.

Algoritmus, jež určuje koeficient zapomínání pomocí VB metody, kolabuje bez užití regularizace ještě dříve. Důvodem je, že při konstantním vstupu i výstupu klesá velikost $\Sigma_{k|k}$ exponenciálně k nule a normalizační konstanta pro Wishartovu část přestává být numericky spočitatelná. Je tedy patrné, že užití regularizace je pro algoritmy rekurzivní identifikace s technikou zapomínání nezbytné.

5.3 Testování dávkových algoritmů pro identifikaci banky ARX modelů

Srovnání algoritmů pro identifikaci banky ARX modelů je prováděno rovněž na reálné a simulované soustavě. Úrovně PRBS byly pro účely simulace opět voleny $-1\,\mathrm{V}$ a $1\,\mathrm{V}$. Změna parametrů v čase byla prováděna dle 5.10, stejně jako při testování rekurzivních algoritmů identifikace. Tuto volbu jsme provedli záměrně, abychom věrněji vystihovali reálnou situaci, neboť jen stěží bychom mohli očekávat, že apriorně neznámá identifikovaná banka modelů bude rovnoměrně zastoupená jednotlivými modely. Je však samozřejmé, že oproti soudělnému zastoupení dílčích modelů je tento typ identifikace obecně obtížnější.

Jak bylo uvedeno, při znalosti příslušností datových párů jednotlivým modelům představuje identifikace směsi modelů triviální problém. A proto právě získání odhadů těchto identit bude cílem našeho pozorování. Společným kritériem tak bude shoda skutečných identit s těmi identifikovanými, kterou určíme jako

$$\kappa = \frac{1}{N}\sum_{i=1}^{M}\sum_{k=1}^{N}\hat{z}_{k,i}z_{k,i}, \tag{5.11}$$

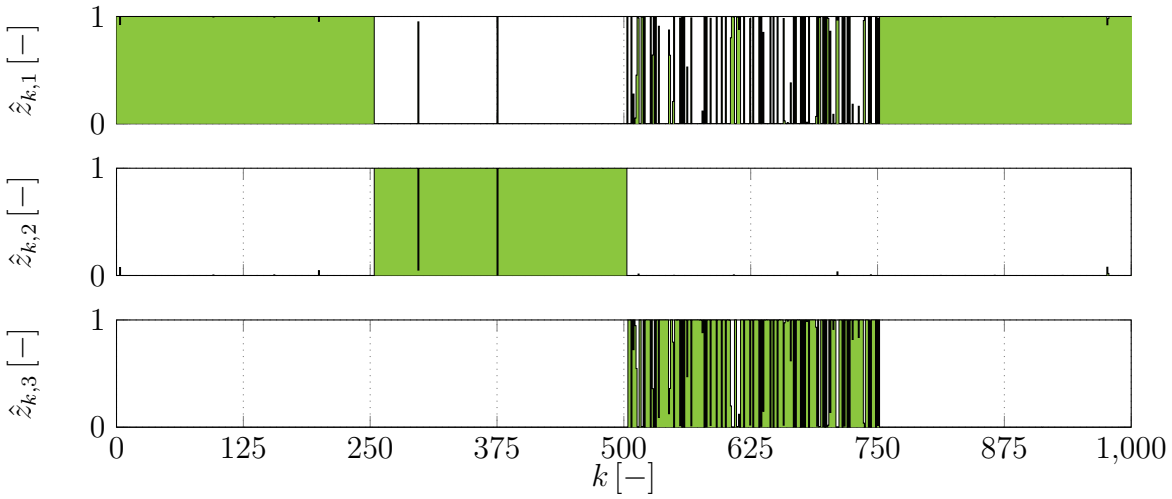
kde $z_{k,i}$ představuje skutečné identity dat.

Maximální počet IVB algoritmů byl nastaven na $r_{max} = 1000$, abychom zajistili, že všechny algoritmy budou ukončeny splněním podmínky pro předčasné ukončení IVB algoritmu, která byla nastavena $q = 10^{-4}$. Maximální počet modelů pro identifikaci byl volen $M = 6$. Modely, jejichž odhadovaná aktivita dosahovala zastoupení menší než 1 %, nejsou ve výsledcích zahrnuty.

5.3.1 Identifikace přepínaných ARX modelů zatížených šumem o společné přesnosti

Apriorní nastavení bylo voleno $\theta_{0,i} = \mathbf{0}_p$, $P_{0,i} = 10^6 I_p\ \forall i \in [1, M]$, $\Sigma_0 = 10^{-3}$, $\nu_0 = 10^{-3}$, $\tau_0 = 10^{-3}$. a inicializace IVB algoritmu byla provedena statistikami $P_{0,i}^{[0]} = 10^6 I_p\ \forall i \in [1, M]$, $\Sigma_0^{[0]} = 1$, $\nu_0^{[0]} = 10^{-5}$, $\tau_0^{[0]} = 1$. Hodnoty vektorů $\theta_{0,i}^{[0]}$ byly $\forall i \in [1, M]$ inicializovány náhodnými čísly v rozsahu $[-2.5, 2.5]$ pocházejících z rovnoměrného rozložení. Toto nastavení bylo společné pro identifikaci reálné i simulované soustavy. Simulovaný model byl zatížen šumem o přesnosti $d = 500$.

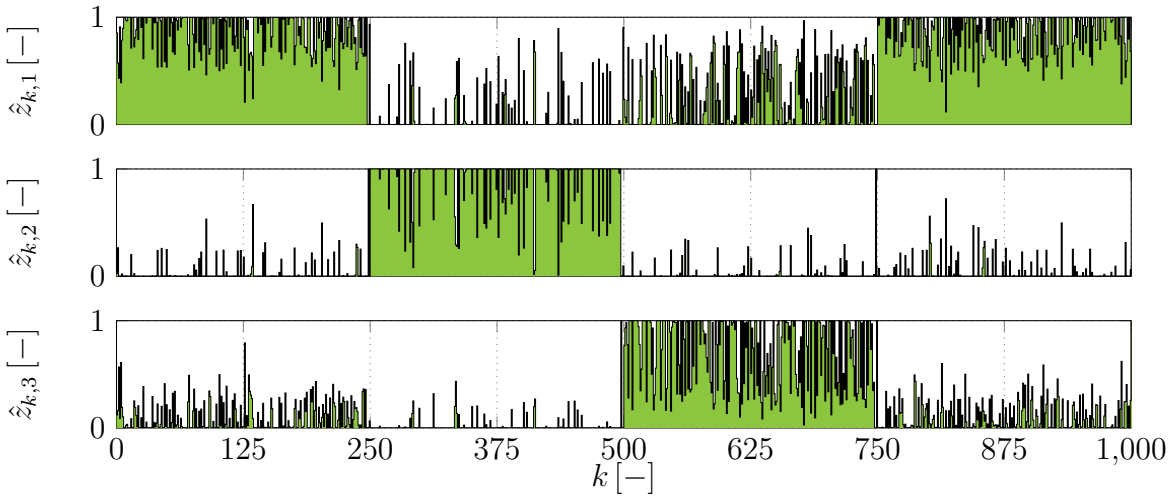
Jak je patrné z obrázku 5.9, nesymetrické zastoupení modely způsobuje, že i na prakticky deterministickém příkladu reálné soustavy se identity dat neidentifikují zcela správně. Shoda identifikovaných dat byla $\kappa = 0.92$, pro případ simulace $\kappa = 0.83$.



Obr. 5.9: Identifikace reálné soustavy algoritmem pro identifikaci přepínaných ARX modelů zatížených šumy o společné přesnosti $\kappa = 500$

Identifikované odhady parametrů $\hat{\theta}$ jsou pro identifikaci reálné soustavy dány 5.12.

$$\begin{aligned} \hat{\theta}_1 &= [0.079,\ 0.212,\ 0.022,\ -1.694,\ 1.162,\ -0.240] \\ \hat{\theta}_2 &= [0.275,\ 0.648,\ 0.105,\ -0.703,\ 0.596,\ -0.132] \\ \hat{\theta}_3 &= [0.024,\ 0.073,\ 0.020,\ -1.887,\ 1.107,\ -0.132] \end{aligned} \tag{5.12}$$



Obr. 5.10: Identifikace simulované soustavy zatížené šumem o přesnosti $d = 500$ algoritmem pro identifikaci přepínaných ARX modelů zatížených šumy o společné přesnosti

Identifikované odhady parametrů $\hat{\theta}$ jsou pro identifikaci simulované soustavy dány 5.13.

$$\begin{aligned}\hat{\theta}_1 &= [0.071,\ 0.202,\ 0.025,\ -1.669,\ 1.100,\ -0.205]\\ \hat{\theta}_2 &= [0.267,\ 0.532,\ -0.068,\ -1.052,\ 0.836,\ -0.274]\\ \hat{\theta}_3 &= [0.027,\ 0.057,\ 0.009,\ -2.040,\ 1.406,\ -0.294]\\ \hat{d} &= 483\end{aligned} \tag{5.13}$$

Závislost na velikosti vstupně-výstupních datových párů

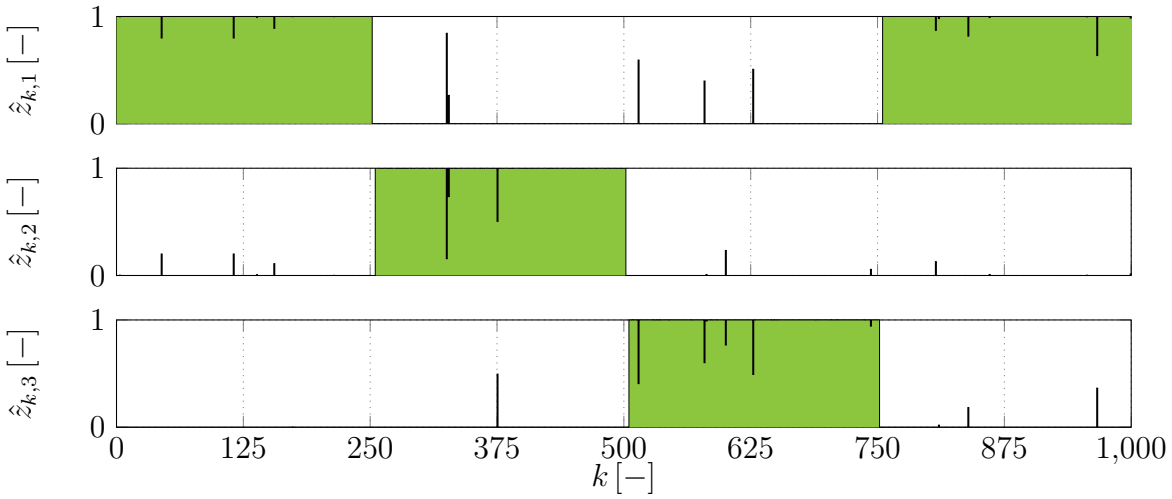
Před přechodem k popisu dalších výsledků upozorněme na poněkud závažný nedostatek tohoto algoritmu. Bylo vypozorováno, že vlastnosti algoritmu závisí i na velikosti vstupně-výstupních dat. Snahy o eliminaci tohoto chování různými volbami apriorních statistik i statistik inicializačních IVB se s úspěchem nesetkaly. Je tedy třeba vyzdvihnout, že pro simulaci byla vstupní a výstupní posloupnost dat dodatečně škálována hodnotou 5. Podotkněme, že výstup již byl zatížen šumem, přičemž budicí vstupní signál nabýval skutečně hodnot $-1\,\mathrm{V}$ a $1\,\mathrm{V}$. Při ponechání původních hodnot algoritmus detekoval oproti skutečnosti menší počet modelů.

5.3.2 Identifikace přepínaných ARX modelů zatížených šumy o obecně různé přesnosti

Apriorní nastavení bylo voleno $\theta_{0,i} = \mathbf{0}_p$, $\Sigma_{0,i} = 10^{-3}$, $\nu_{0,i} = 10^{-3}$, $\tau_{0,i} = 10^{-3}$. $\xi_{0,i} = 10^{-3}$, $\upsilon_{0,i} = 10^{-3}$ $\forall i \in [1, N]$ a inicializace IVB algoritmu byla dána statistikami $\Sigma_{0,i}^{[0]} = 1$, $\nu_{0,i}^{[0]} = 10^{-3}$, $\tau_{0,i}^{[0]} = N/M$, $\xi_{0,i}^{[0]} = 10^{-3}$, $\upsilon_{0,i}^{[0]} = 10^{-3}$ $\forall i \in [1, N]$. Hodnoty vektorů $\theta_{0,i}^{[0]}$ byly inicializovány náhodnými čísly v rozsahu $[-2.5, 2.5]$ pocházejících z rovnoměrného rozložení. Shodné nastavení je společné pro identifikaci reálné i simulované soustavy.

Abychom otestovali schopnosti identifikovat modely s různými šumy, byla v simulaci prvnímu modelu nastavena přesnost šumu $d_1 = 1000$ a zbylým dvěma $d_2 = d_3 = 500$.

Pro identifikaci reálné soustavy byla dosažna shoda $\kappa = 0.98$. Tento výsledek je poměrně jasně patrný i z grafu 5.11.

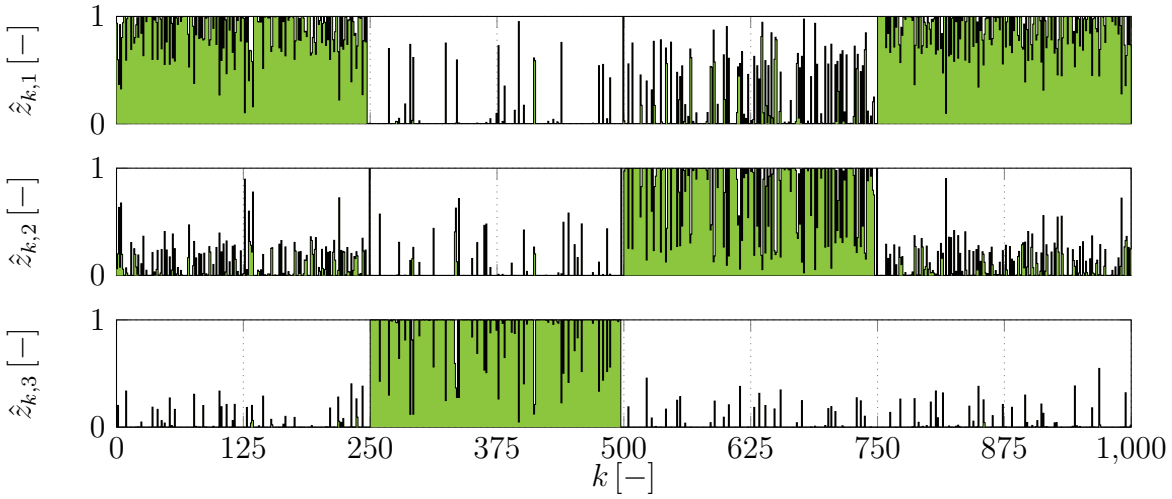


Obr. 5.11: Identifikace reálné soustavy algoritmem pro identifikaci přepínaných ARX modelů zatížených šumy o obecně různé přesnosti

Odhady parametrů $\hat{\theta}$ pro identifikaci simulované soustavy uvádí 5.14.

$$\begin{aligned}\hat{\theta}_1 &= [0.078,\ 0.217,\ 0.037,\ -1.627,\ 1.061,\ -0.188]\\ \hat{\theta}_2 &= [0.276,\ 0.650,\ 0.107,\ -0.701,\ 0.595,\ -0.131]\\ \hat{\theta}_3 &= [0.024,\ 0.071,\ 0.013,\ -1.981,\ 1.281,\ -0.220]\end{aligned} \tag{5.14}$$

Při zatížení simulovaného modelu šumy $d_1 = 1000$ a $d_2 = d_3 = 500$ byla obdržena shoda identifikovaných identit se skutečnými $\kappa = 0.876$.



Obr. 5.12: Identifikace simulované soustavy zatížené šumem o přesnosti $d = 500$ algoritmem pro identifikaci přepínaných ARX modelů zatížených šumy o obecně různé přesnosti

Odhady parametrů $\hat{\theta}$ a přesností šumů $\hat{d}$ pro identifikaci simulované soustavy uvádí 5.15.

$$\begin{aligned}
\hat{\theta}_1 &= [0.073,\ 0.207,\ 0.039,\ -1.610,\ 1.008,\ -0.157] \\
\hat{\theta}_2 &= [0.260,\ 0.609,\ 0.110,\ -0.757,\ 0.608,\ -0.120] \\
\hat{\theta}_3 &= [0.025,\ 0.058,\ 0.011,\ -2.010,\ 1.361,\ -0.277] \\
\hat{d}_1 &= 910 \qquad\qquad \hat{d}_2 = 478 \qquad\qquad \hat{d}_3 = 455
\end{aligned} \tag{5.15}$$

Jak vidíme, hodnoty šumů jsou opět podhodnoceny, nicméně je patrné, že identifikovaná přesnost šumu prvního modelu je přibližně dvakrát vyšší, než u zbývajících, což odpovídá nastavení.

5.3.3 Shrnutí obou algoritmů pro identifikaci banky ARX modelů

Pomineme-li zmíněný nedostatek algoritmu pro identifikaci banky ARX modelů zatíženými šumy o společné přesnosti, můžeme diskutovat společný neduh, který se vyskytuje u obou algoritmů. Zejména z grafů 5.10 a 5.15 pro simulovanou směs přepínaných ARX modelů je patrné, že i přes dominanci jednoho z modelů v určitých časových oknech odraženou v odhadech $\hat{z}_k$ pozorujeme, že některá data z tohoto časového okna jsou přiřazena v daný okamžik jinému, minoritně zastoupenému modelu. V důsledku tohoto chování není inference parametrů tak přesná.

Dále nezastírejme, že algoritmus pro identifikaci banky ARX modelů zatížené šumem se společnou přesností, je poměrně citlivý na nastavení apriorních statistik i na inicializaci IVB algoritmu. Opakovatelná konvergence tohoto algoritmu k obdobným výsledkům, jak uvádí grafy 5.9 a 5.10, bylo docílen pouze za vybrané konfigurace nastavení.

Závěr

První kapitola 1 diplomové práce skýtá vhled do problematiky, poskytuje potřebný matematický aparát a koncentruje potřebné informace, ze kterých je v průběhu práce vycházeno. V následující kapitole 2 byl vypracován první bod zadání, který spočíval v bayesovské formulaci algoritmu identifikace normálního časově invariantního regresního modelu, a to v dávkové i průběžné variantě.

Na tomto základu bylo následně stavěno při formulaci dvojice algoritmů pro průběžnou identifikaci časově invariantního systému. Oba uvedené rekurzivní algoritmy řešily neznalost exaktního časového vývoje technikou zapomínání. Pro zaručení numerické stability těchto algoritmů byla prováděna regularizace [8]. Algoritmus [15], u kterého byla volba koeficientu určována VB aproximací, nedosahoval u provedených měření takové odolnosti na šum, jako algoritmus [8], jež určoval koeficient zapomínání statistickým rozhodováním. Toto chování jsme přisuzovali faktu, že koeficient zapomínání určovaný tímto algoritmem [15] nikdy nedosahoval hodnoty 1. Vylepšení tohoto neduhu je jedním z námětů na další studium a výzkum.

V kapitole 4 byla uvedena dvojice algoritmů pro identifikaci banky ARX modelů. Jak bylo diskutováno, algoritmus pro identifikaci směsi ARX modelů zatížených společným šumem vykazoval nežádoucí závislost výpočtu posteriorních statistik na škálování datové informace. U obou z algoritmů pro dávkovou inferenci směsi ARX modelů bylo pozorováno nežádoucí chování, kdy identita dat nebyla pro některé úseky přiřazována exkluzivně jen jednomu z modelů. Tyto nectnosti je nasnadě studovat hlouběji a odvodit takové řešení, které jim bude předcházet.

Všechny čtyři algoritmy pro pro inferenci časově variantních regresních modelů byly testovány na fyzické soustavě tvořené elektrickým systémem a v simulaci. Výsledky jsou zhodnoceny v kapitole 5. Bylo rovněž poukázáno na důležitost provádění regularizačního kroku u algoritmů s faktorem zapomínání.

Přestože bychom mohli o patřičném splnění zadání této práce vést diskuze, ten nejdůležitější, přesto nepsaný cíl, byl dle mého názoru splněn. A tím bylo proniknout alespoň na kraj bayesovského světa a získat tak nový vhled do celé řady technických disciplín.

Literatura

[1] LU, Y., S. KHATIBISEPEHR a B. HUANG. Regularized Estimation with Variable Exponential Forgetting. IEEE [online]. 53rd IEEE Conference on Decision and Control, 2014 [cit. 2023-12-22]. Dostupné z: doi:10.1109/CDC.2014.7039777

[2] ŠMÍDL, V. a A. QUINN. The Variational Bayes Method in Signal Processing. New York: Springer, 2006. ISBN 978-3-642-06690-0. [cit. 2023-12-22].

[3] DOKOUPIL, J. Bayesovský přístup k identifikaci a filtraci. In: Optimalizace regulátorů [online]. [cit. 2023-12-23].

[4] PETERKA V. Bayesian approach to system identification. In: *Trends and Progress in System Identification* (P. Eykhoff ed.). Oxford, 1981. s. 239–304. [cit. 2023-12-22].

[5] BLAHA, P. Modely lineárních systémů. In: Modelování a identifikace [online]. [cit. 2023-12-25]-.

[6] DOKOUPIL, Jakub. Identifikace metodou nejmenších čtverců. In: Optimalizace regulátorů [online]. [cit. 2023-12-27].

[7] BISHOP, Christopher M. Pattern recognition and machine learning [online]. [New York]: Springer, c2006 [cit. 2023-12-30]. Information science and statistics. ISBN 03-873-1073-8.

[8] DOKOUPIL, J. a P. VÁCLAVEK. Regularized Estimation with Variable Exponential Forgetting [online]. 59th IEEE Conference on Decision and Control (CDC), 2020 [cit. 2023-12-22]. Dostupné z: doi:10.1109/CDC42340.2020.9304385

[9] BLAHA, P. Rekurzivní MNČ. In: Modelování a identifikace [online]. [cit. 2023-12-25]-.

[10] DOKOUPIL Jakub, Alina VODA a Pavel VÁCLAVEK. Regularized Extended Estimation With Stabilized Exponential Forgetting. IEEE Transactions on Automatic Control [online]. 2017, 62(12), 6513-6520 [cit. 2024-01-04]. ISSN 0018-9286. Dostupné z: doi:10.1109/TAC.2017.2656379

[11] Examples for optimization subject to inequality constraints, Kuhn-Tucker. Youtube [online]. 27.02.2007 [cit. 2011-05-25]. Dostupné z: https://www.youtube.com/watch?v=TqN-8fxYUYY. Kanál Mathematics for Economists.

[12] BOYD, Stephen a Lieven VANDENBERGHE. Convex Optimization [online]. New York: Cambridge University Press, 2004 [cit. 2024-02-09]. Dostupné z: doi:978-0-521-83378-3

[13] CHOLESKY, André Louis. Note Sur Une Méthode de Résolution des équations Normales Provenant de L'Application de la MéThode des Moindres Carrés a un Système D'équations Linéaires en Nombre Inférieur a Celui des Inconnues. — Application de la Méthode a la Résolution D'un Système Defini D'éQuations LinéAires. Bull. Géodésique [online]. 1924, 2 [cit. 2024-04-12]. Dostupné z: doi:10.1007/BF03031308

[14] BERNARDO, José-Miguel. [online]. 1979 [cit. 2024-04-15]. Dostupné z: doi:https://doi.org/10.1214/aos/1176344689

[15] V. ŠMÍDL, Anthony Quinn. Bayesian estimation of non-stationary AR model parameters via an unknown forgetting factor. 3rd IEEE Signal Processing Education Workshop. 2004 IEEE 11th Digital Signal Processing Workshop, 2004 [online]. IEEE, 2004, 221-225 [cit. 2024-04-17]. ISBN 0-7803-8434-2. Dostupné z: doi:10.1109/DSPWS.2004.1437946

[16] DOKOUPIL Jakub a Pavel VÁCLAVEK. Data-Driven Stabilized Forgetting Design Using the Geometric Mean of Normal Probability Densities. 2018 IEEE Conference on Decision and Control (CDC) [online]. IEEE, 2018, 68(5), 1403-1408 [cit. 2024-04-18]. ISBN 978-1-5386-1395-5. ISSN 0018-9286. Dostupné z: doi:10.1109/CDC.2018.8619117

[17] GORDON-RODRIGUEZ, Elliott; LOAIZA-GANEM, Gabriel a CUNNINGHAM, John P. The continuous categorical: a novel simplex-valued exponential family. Online. 2020. Dostupné z: https://doi.org/10.48550/arXiv.2002.08563. [cit. 2024-04-29].

[18] GORDON-RODRIGUEZ, Elliott; LOAIZA-GANEM, Gabriel; POTAPCZYNSKI, Andres a CUNNINGHAM, John P. On the Normalizing Constant of the Continuous Categorical Distribution. Online. 2022. Dostupné z: https://doi.org/https://doi.org/10.48550/arXiv.2204.13290. [cit. 2024-05-10].

[19] DOKOUPIL, Jakub. Model analogové soustavy. In: Optimalizace regulátorů [online]. [cit. 2024-5-12].

Seznam symbolů a zkratek

fhp	Funkce hustoty pravděpodobnosti.
pf	Pravděpodobnostní funkce.
KLD	Kullbackova–Leiblerova divergence.
ARX	Autoregresní model s exogonním vstupem.
VB	Variační Bayesovský.
IVB	Iterativní Variační Bayesovský.
KKT	Karush-Kuhn-Tucker.
GM	Směs normálních regresních modelů.
OZ	Operační zesilovač.
$\mathcal{NW}$	Normální-Wishartovo rozložení.
$\equiv$	Rovnost dle definice.
$\propto$	Přímá úměra (rovnost až na normalizační konstantu).
$\lvert\cdot\rvert$	Determinant matice.
$\mathrm{tr}(\cdot)$	Stopa matice.
I_p	Jednotková matice o rozměrech $p \times p$.
x'	Transpozice vektoru x
x^*	Definiční obor veličiny x.
$x_{l\|m}$	Hodnota veličiny x v kroku l počítaná v kroku m.
$\mathbf{1}_n$	$\mathbf{1}_n \equiv [1, ..., 1]' \in \mathbb{R}^n$
$\mathbf{0}_n$	$\mathbf{0}_n \equiv [0, ..., 0]' \in \mathbb{R}^n$
$\epsilon_M(i)$	$\epsilon_M(i) \equiv [\mathbf{0}'_{i-1}, 1, \mathbf{0}'_{M-i+1}]' \in \mathbb{R}^M$
Δ_M	Simplex v prostoru $M-1$
$f(\cdot)$	Funkce hustoty pravděpodobnosti či pravděpodobnostní funkce.
$\breve{f}(\cdot)$	Fhp nebo pf představující kandidáta na aproximaci.

$\tilde{f}(\cdot)$	Fhp nebo pf představující optimální volbu aproximaci.
$\mathsf{E}_{\{f(x)\}}[g(x)]$	Očekávaná hodnota obecné funkce $g(x)$ nad náhodnou veličinou x mající fhp (pf) $f(x)$.
$\widehat{g(x)}$	Zkrácená notace pro $\mathsf{E}_{\{f(x)\}}[g(x)]$.
$\mathcal{N}_\theta(\hat{\theta}, P/d)$	Normální rozložení náhodné veličiny θ o střední hodnotě $\hat{\theta}$ a varianci P/d.
$\mathcal{W}_d(\Sigma, \nu)$	Wishartovo rolożení náhodné veličiny d s parametry Σ, ν.
$\mathcal{D}_\alpha(\tau)$	Dirichletovo rolożení náhodné veličiny α s parametrem τ.
$\mathcal{C}_z(\gamma)$	Kategorické rolożení náhodné veličiny z s parametrem γ.
$\Gamma(\cdot)$	Gama funkce.
$\Psi(\cdot)$	Digama funkce.

A Obsah elektronické přílohy

Elektronická příloha obsahuje naprogramované algoritmy doplněné o instrukce a naměřená data pro snadné zkoumání jejich vlastností. Algoritmy byly implementovány v programu Matlab, verze 2022a.

/ ..... Kořenový adresář přiloženého archivu
- Algoritmy průběžné identifikace .....
- Algoritmy dávkové identifikace .....