



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA STROJNÍHO INŽENÝRSTVÍ

FACULTY OF MECHANICAL ENGINEERING

ÚSTAV MATEMATIKY

INSTITUTE OF MATHEMATICS

AUTONOMNÍ JEDNOKANÁLOVÝ DEINTERLEAVING

AUTONOMOUS SINGLE-CHANNEL DEINTERLEAVING

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Tereza Tomešová

VEDOUCÍ PRÁCE

SUPERVISOR

doc. Mgr. Zuzana Hübnerová, Ph.D.

BRNO 2021

Zadaní diplomové práce

Ústav: Ústav matematiky
Studentka: **Bc. Tereza Tomešová**
Studijní program: Aplikované vědy v inženýrství
Studijní obor: Matematické inženýrství
Vedoucí práce: **doc. Mgr. Zuzana Hübnerová, Ph.D.**
Akademický rok: 2020/21

Ředitel ústavu Vám v souladu se zákonem č.111/1998 o vysokých školách a se Studijním a zkušebním řádem VUT v Brně určuje následující téma diplomové práce:

Autonomní jednokanálový deinterleaving

Stručná charakteristika problematiky úkolu:

Deinterleaving je proces separace impulzů nebo sekvence impulzů z přijatého impulzního vláčku na sekvence impulzů od jednotlivých zdrojů/emiterů. V současné době je ve většině případů deinterleaving prováděn ručně operátorem na základě filtrace RL (radiolokačních) parametrů. V důsledku hustého elektromagnetického prostředí roste potřeba autonomního deinterleavingu v reálném čase pro efektivní identifikaci emiterů v okolí radaru.

Cíle diplomové práce:

Uvedení přehledu přístupů pro deinterleaving sekvence impulzů. Popsání jednotlivých metod a jejich chování vůči zkreslujícím vlivům (šum, chybějící impulzy, odlehlé impulzy). Navržení vhodných metod pro autonomní jednokanálový deinterleaving, implementace vybraných metod a jejich ověření na simulovaných/reálných datech.

Seznam doporučené literatury:

MANICKCHAND, K. Multiple Radar Environment Emission Deinterleaving and PRI Prediction. Cape Town, 2017. Diplomová práce. Department of Electrical Engineering at the University of Cape Town.

MARDIA, H. K. New techniques for the deinterleaving of repetitive sequences. In IEE PROCEEDINGS F - Radar and Signal Processing. 1989, 136(4), 149-154.

LIU, J., MENG, H. a WANG, X. A new pulse deinterleaving algorithm based on Multiple Hypothesis Tracking. In 2009 International Radar Conference "Surveillance for a Safer World" (RADAR 2009). 2009, 1-4. ISSN 1097-5764.

REVILLON, G., MOHAMMAD-DJAFARI, A. a ENDERLI, C. Radar emitters classification and clustering with a scale mixture of normal distributions. IET Radar, Sonar & Navigation. 2019, 13(1), 128-138. ISSN 1751-8784.

Termín odevzdání diplomové práce je stanoven časovým plánem akademického roku 2020/21

V Brně, dne

L. S.

prof. RNDr. Josef Šlapal, CSc.
ředitel ústavu

doc. Ing. Jaroslav Katolický, Ph.D.
děkan fakulty

Abstrakt

Tato práce se zabývá autonomním jednokanálovým deinterleavingem. Autonomní jednokanálový deinterleaving je proces separace přijaté sekvence radiolokačních impulzů od více emiterů na sekvence od jednotlivých emiterů, který je prováděn bez lidské pomoci. Metody využívané pro úlohu deinterleavingu lze rozdělit dle počtu parametrů používaných pro separaci a to na jednoparametrické a víceparametrické metody. Tato práce se zabývá metodami především víceparametrickými. Jako vhodné metody pro autonomní jednokanálový deinterleaving byly vybrány DBSCAN algoritmus a variační bayesovské metody. Vybrané metody byly upraveny pro úlohu deinterleavingu a implementovány v programovacím jazyce Python. Jejich účinnost byla ověřena na simulovaných datech a datech z reálného provozu.

Summary

This thesis deals with an autonomous single-channel deinterleaving. An autonomous single-channel deinterleaving is a separation of the received sequence of impulses from more than one emitter to sequences of impulses from one emitter without a human assistance. Methods used for deinterleaving could be divided into single-parameter and multiple-parameter methods according to the number of parameters used for separation. This thesis primarily deals with multi-parameter methods. As appropriate methods for an autonomous single-channel deinterleaving DBSCAN and variational bayes methods were chosen. Selected methods were adjusted for deinterleaving and implemented in programming language Python. Their efficiency is examined on simulated and real data.

Klíčová slova

Aposteriorní pravděpodobnost, Apriorní pravděpodobnost, DBSCAN, Deinterleaving, Emi-
ter, Gaussův smíšený model, Radiolokační parametry, Sekvence impulzů, Variační baye-
sovské metody.

Keywords

Aposterior probability, DBSCAN, Deinterleaving, Emitter, Gaussian mixture model, Prior
probability, RL parameters, Sequence of impulses, Variational bayes methods.

TOMEŠOVÁ, T. *Autonomní jednokanálový deinterleaving*. Brno: Vysoké učení technické
v Brně, Fakulta strojního inženýrství, 2021. 66 s. Vedoucí doc. Mgr. Zuzana Hübnerová,
Ph.D.

Prohlašuji, že jsem diplomovou práci *Autonomní jednokanálový deinterleaving* vypracovala samostatně pod vedením doc. Mgr. Zuzany Hübnerové, Ph.D. s použitím materiálů uvedených v seznamu literatury.

Tereza Tomešová

Chtěla bych poděkovat vedoucí práce doc. Mgr. Zuzaně Hübnerové, Ph.D. a Ing. Pavlovi Protivovi, Ph.D. za pozitivní a vstřícný přístup při tvorbě této diplomové práce.

Tereza Tomešová

Obsah

Úvod	3
1 Základní terminologie	4
1.1 Pasivní radar	4
1.2 Radiolokační parametry	4
1.2.1 RF parametr	4
1.2.2 PW parametr	5
1.2.3 PA parametr	5
1.2.4 TOA a PRI parametr	5
1.2.5 Modulace parametrů	6
2 Jednparametrické metody	8
2.1 Pulse sorting algoritmus	8
2.2 Sequence search algoritmus	8
2.3 Delta- τ histogram/TOA difference histogram	9
2.4 Cumulative difference (CDIF) histogram	9
2.5 Sequential difference (SDIF) histogram	9
2.6 Integrální transformace (PRI transformace)	10
2.7 Diskrétní Fourierova transformace	12
2.8 Vlnková transformace	12
2.9 Deinterleaving pomocí trackingu	13
3 Víceparametrické metody	15
3.1 Support vector machine (SVM)	15
3.2 K-means algoritmus	16
3.3 Density-based spatial clustering of applications with noise (DBSCAN)	17
3.4 Bayesovské metody	17
3.5 Neuronové sítě	18
3.5.1 Self-organizing neuronové sítě	18
4 Vybrané metody	20
4.1 DBSCAN	20
4.1.1 Vstupní parametry	22
4.1.2 Postup algoritmu	23
4.2 Variační bayesovské metody	23
4.2.1 Směsový model	23
4.2.2 Exponenciální rodina rozdělení	24
4.2.3 Kullback-Leiblerova divergence	25
4.2.4 Mean-field aproximace	26
4.2.5 Gradient dolní meze	26
4.2.6 Laplaceova aproximace	27
4.2.7 Odhad parametrů GMM pro úlohu deinterleavingu	28
4.2.8 Konjugované rodiny apriorních rozdělení	30
4.2.9 Aposteriorní rozdělení parametrů	31
4.2.10 Aposteriorní rozdělení latentních proměnných	34

4.2.11	Variační rozdělení	35
5	Experiment	40
5.1	Hodnocení úspěšnosti deinterleavingu	40
5.1.1	ARI	40
5.1.2	AMI	41
5.1.3	Stejnorodost	41
5.1.4	Úplnost	42
5.2	Simulovaná data	42
5.2.1	Sada dat 1	43
5.2.2	Sada dat 2	44
5.2.3	Sada dat 3	46
5.2.4	Sada dat 4	47
5.2.5	Sada dat 5	49
5.2.6	Sada dat 6	50
5.2.7	Sada dat 7	52
5.2.8	Sada dat 8	53
5.2.9	Sada dat 9	55
5.3	Data z reálného provozu	57
6	Závěr	59
	Seznam použitých zkratk a symbolů	65

Úvod

Postupy pro prozkoumávání vzdušného prostoru, označované jako Electronic Support Measures (ESM), mají za úkol vyhledávat, zachycovat a identifikovat objekty vyzařující radiolokační (RL) signály, nazývané zdroje, přítomné v okolí. V hustém elektromagnetickém prostředí mohou prostředky ESM přijmout sekvence impulzů od více zdrojů, které se překrývají a jsou navíc zatíženy šumem a rušením. To může způsobit nemožnost správné identifikace zdrojů RL signálů. Je tedy nutné přijatou sekvenci impulzů rozdělit na sekvence od jednotlivých zdrojů. Tento proces se nazývá deinterleaving. V dnešní době je deinterleaving prováděn většinou ručně operátory na základě filtrace RL parametrů. V důsledku čím dál hustšího elektromagnetického prostředí roste potřeba autonomního deinterleavingu, tj. deinterleavingu bez lidské pomoci v reálném čase.

Metody využívané pro úlohu autonomního deinterleavingu můžeme rozdělit do dvou skupin. Na metody využívající pouze jeden parametr popisující přijatou sekvenci impulzů a na metody využívající dva a více parametrů.

Jednoparametrické metody deinterleavingu využívají pouze parametr čas příchodu impulzu (TOA). Mezi tyto metody se řadí například tzv. sekvenční algoritmy, metody využívající histogram, transformace nebo Kalmanova filtrace. Tyto zmíněné metody odhadují frekvenci, nebo hodnotu pulse repetition interval (PRI), která popisuje rozdíl času příchodu dvou po sobě následujících impulzů. Pomocí odhadnuté hodnoty PRI, nebo frekvence je poté oddělena odpovídající sekvence od jednoho zdroje. Tyto metody jsou velmi přesné, ale početně náročné. Dalším nedostatkem je jejich klesající přesnost při větším počtu přijatých impulzů, v případě šumu a chybějících impulzů. Metody využívající parametr TOA nedokáží odhadnout komplexnější modulaci PRI (tj. změnu hodnot PRI v čase), jako je například dwell & switch (D&S), wobulated či sliding PRI. Z tohoto důvodu je vhodné pro deinterleaving využít více parametrů a separovat sekvence impulzů na základě podobnosti určitých parametrů namísto odhadu hodnoty PRI. Mezi takovéto metody řadíme především statistické metody (např. shluková analýza) nebo neuronové sítě.

Cílem této práce je vybrat vhodné metody pro úlohu autonomního jednokanálového deinterleavingu, které jsou účinné pro separaci přijaté sekvence impulzů od více zdrojů v reálném provozu, na základě rešerše používaných metod. Dále ověřit jejich účinnost na simulovaných datech nebo datech z reálného provozu.

První kapitola této práce obsahuje základní terminologii popisující RL parametry impulzů a jejich možné modulační. V druhé kapitole je proveden popis jednoparametrických metod využívaných pro úlohu autonomního jednokanálového deinterleavingu a jejich chování vůči zkreslujícím jevům (šum, chybějící impulzy, odlehlé impulzy). Třetí kapitola popisuje metody víceparametrické.

Na základě chování metod vůči zkreslujícím jevům uváděných v literatuře, byly v druhé části práce vybrány dvě metody z druhé a třetí kapitoly pro ověření jejich účinnosti na simulovaných datech a datech z reálného provozu. Čtvrtá kapitola obsahuje popis těchto metod a jejich případnou modifikaci. Jako vhodné metody byly zvoleny Density-based spatial clustering of applications with noise (DBSCAN) a variační bayesovy metody. Poslední kapitola zkoumá účinnost vybraných metod.

1 Základní terminologie

Akronym radar vznikl ze zkrácení anglických slov "Radio detection and ranging", neboli "rádiové rozpoznávání a zaměřování" (cílů). Tento akronym popisuje funkci většiny radarů. Radar je zařízení určené ke sledování cílů, určení jejich polohy nebo jejich identifikaci pomocí elektromagnetických vln. Můžeme je dělit do dvou skupin podle toho, zda vyzařují elektromagnetické vlny nebo nikoliv. Ty, které vyzařují elektromagnetické vlny nazýváme aktivní radary. Radar, který nevyzařuje elektromagnetické vlny, ale pouze je přijímá pak nazýváme pasivní radar [5]. V rámci této diplomové práce se zaměříme na pasivní radary.

1.1 Pasivní radar

Pasivní radar nevyzařuje žádné elektromagnetické vlny a tím je pro svoje okolí těžko zaměřitelný. Pasivní radar pouze přijímá signál vyzařovaný okolními předměty/cíli, nebo odražené vlny od sledovaného cíle vytvořené externími zdroji, neboli emitery. Tyto radary se dále dělí na časoměrné systémy a směroměrné systémy. Směroměrné pasivní radary využívají pro výpočet polohy cíle směrové informace (tj. úhel příchodu impulzu). Naopak časoměrné pasivní radary využívají pro výpočet čas příchodu impulzu [5]. V této diplomové práci se zaměříme na deinterleaving pro pasivní časoměrné radary.

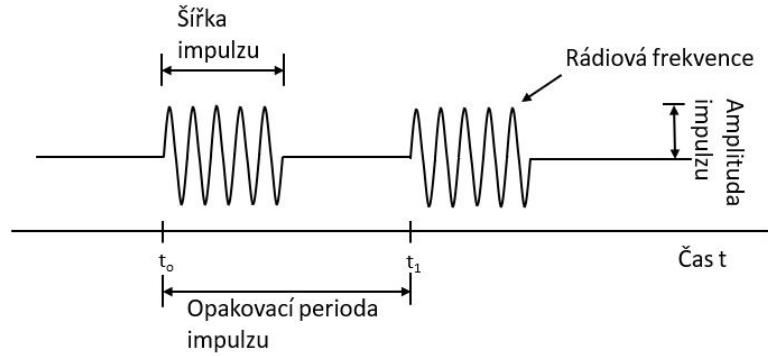
1.2 Radiolokační parametry

Impulzní signály, které jsou vyzařovány různými emitery v okolí, jsou přijímány prostředky ESM. Každý přijatý impulz je popsán parametry, jako je například čas příchodu impulzu (TOA), šířka impulzu (PW), rádiová frekvence (RF), amplituda (PA) a odvozený parametr pulse repetition interval (PRI) atd. (Obrázek 1.1). Každý impulz je reprezentován vektorem parametrů (TOA , PW , RF , PA , PRI ...) a přijatá sekvence impulzů pak maticí pulse descriptor word (***PDW***):

$$\mathbf{PDW} = \begin{pmatrix} TOA_1 & RF_1 & PW_1 & PA_1 & \dots \\ TOA_2 & RF_2 & PW_2 & PA_2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ TOA_n & RF_n & PW_n & PA_n & \dots \end{pmatrix}. \quad (1.1)$$

1.2.1 RF parametr

Frekvence je fyzikální veličina, která udává počet opakování periodického děje za daný časový interval. Tento parametr je udáván v jednotkách Hertz (Hz). Parametr RF je jedním z důležitých parametrů pro deinterleaving přijaté sekvence impulzů díky faktu, že radary, které jsou fyzicky blízko, nemohou operovat na stejné frekvenci. Nevýhoda RF parametru je možnost agility signálu v parametru RF [13]. Agilitu definujeme jako změnu RF parametru v čase viz podkapitola 1.2.5.



Obrázek 1.1: RL parametry signálu

1.2.2 PW parametr

Šířka impulzu (PW), neboli doba trvání impulzu je parametr popisující dobu vyzařování rádiové frekvence. PW parametr je udáván v sekundách (s). Emitery zřídka kdy mění v průběhu vyzařování PW parametr a díky tomu je dalším vhodným parametrem pro úlohu deinterleavingu [13]. PW parametr dělíme na krátký a dlouhý. Řekneme, že impuls má krátkou (resp. dlouhou) šířku, jestliže PW parametr dosahuje délky maximálně několika jednotek μs (resp. řádově desítek μs). Nevýhodou využití PW parametru pro úlohy deinterleavingu je skutečnost, že mnoho emiterů pracuje s krátkou šířkou impulzů a tím mají emitery podobné PW parametry.

1.2.3 PA parametr

Amplituda je veličina popisující změnu v jedné periodě. Většina radarů má mechanicky se pohybující antény, které úzkým svazkem prohledávají prostor. To má za následek, že se parametr PA v čase mění. Pokud je anténa emiteru nasměrována přímo proti přijímači, amplituda přijatého signálu bude mít největší možnou hodnotu. V případě, že je svazek natočený jiným směrem, parametr PA je velmi malý a signál od tohoto emiteru nemusí být zachytitelný [13]. Dále existuje mnoho různých způsobů prohledávání prostoru, např. moderní radary s elektronicky řízeným svazkem se mohou projevit náhodnými změnami PA, jejíž využití pro deinterleaving je pak obtížné.

1.2.4 TOA a PRI parametr

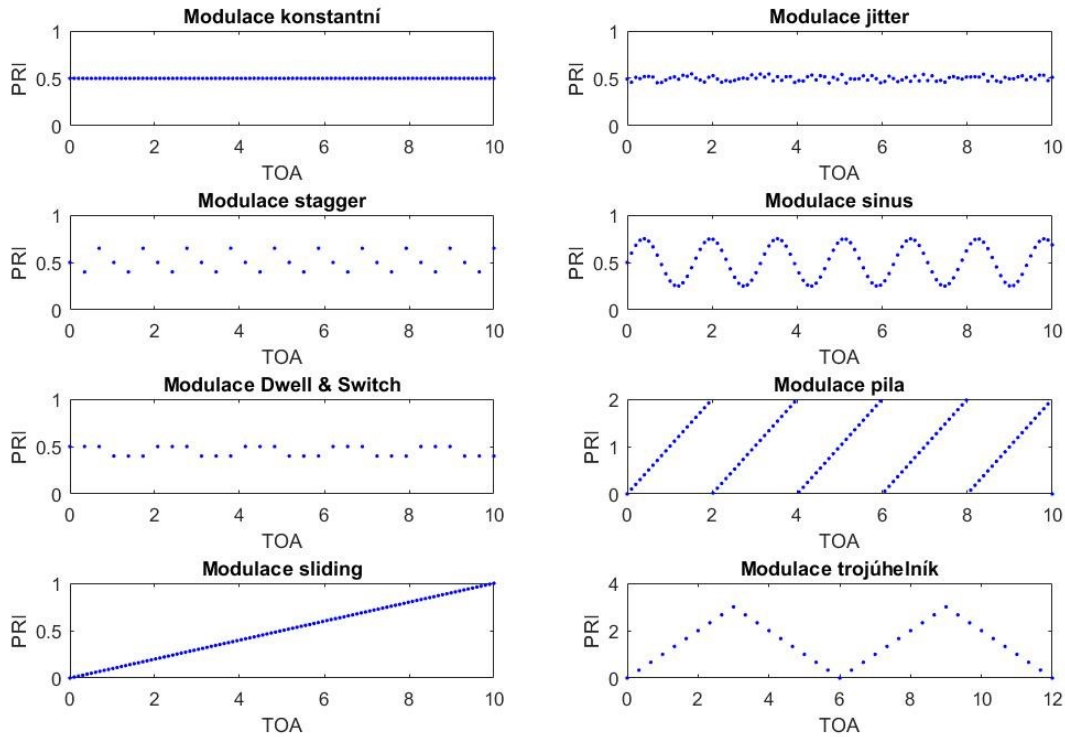
TOA parametr udává čas příchodu impulzu. Z tohoto parametru se odvozuje parametr PRI, který popisuje rozdíl času příchodu dvou po sobě následujících impulzů. PRI parametr je definován následovně:

$$PRI = t_{i+1} - t_i, \quad (1.2)$$

kde t_i je čas příchodu i -tého impulzu v přijaté sekvenci impulzů. PRI je vzhledem k citlivosti na rušení (výpadky a falešné pulzy) parametr problematický, nicméně, vzhledem k tomu, že udává časování radaru, jde o jeden z nejdůležitějších parametrů pro charakterizování radaru.

1.2.5 Modulace parametrů

Emiter může využívat změny parametrů v čase podle různých vzorců nazývaných modulace (Obrázek 1.2). Jedním z možných využití modulace je znesnadnění identifikace emiterů. Parametry s modulací nazýváme agilní, nebo parametry s agilitou.



Obrázek 1.2: PRI modulace

- **Modulace konstantní:** Konstantní modulační typ chápeme jako konstantní průběh určitého pulzního parametru, kde případné odchylky od konstanty nejsou způsobeny záměrně. Pro definici konstantní modulace je nutné tolerované odchylky specifikovat, a to buď intervalem možných odchylek anebo stanovením maximálního rozdílu hodnot pulz od pulzu.
- **Modulace jitter:** Modulaci jitter chápeme jako neperiodický průběh pulzních parametrů, kde se hodnota pulzního parametru mění náhodně pulz od pulzu. Náhodné hodnoty typicky odpovídají rovnoměrnému nebo normálnímu rozdělení. Rozdělení může být spojitě i diskrétní.
- **Modulace stagger:** Modulaci stagger chápeme jako periodický průběh pulzních parametrů, který střídá dva a více úrovně. Změny jsou pulz od pulzu. Hodnoty se mohou v rámci jedné periody opakovat. Pro odlišení modulace stagger od ostatních periodických typů, jako je například sinus nebo pila, je v definici vhodné definovat přípustný minimální rozdíl dvou sousedních hodnot pulz od pulzu. Sinus a pila (viz níže) jsou převážně spojitě průběhy, které mají rozdíl sousedních hodnot oproti staggeru relativně malý.

- **Modulace blokový stagger (D&S):** Modulaci blokový stagger chápeme jako průběh pulzních parametrů, který je po částech konstantní se změnou blok od bloku nebo změnou pulzní skupiny od pulzní skupiny. Podobně jako u konstantního průběhu je nutné specifikovat tolerance konstantnosti. Blokový stagger se může vyskytovat v periodické i neperiodické variantě.
- **Modulace sinus, trojúhelník a pila (wobulated):** Modulaci sinus, trojúhelník a pila obecně chápeme jako periodické průběhy pulzních parametrů, které se mění pulz od pulzu podle odpovídající funkce. Trojúhelník se skládá z části, která je stoupající a klesající. Pila má pouze klesající nebo stoupající průběh.
- **Modulační typ klouzavý (sliding):** Modulační typ klouzavý chápeme jako neperiodický průběh pulzních parametrů, který má rostoucí nebo klesající lineární anebo exponenciální průběh. Některé zdroje uvádí typ klouzavý také jako periodický. V tomto členění je pro periodickou variantu vyhrazeno označení pila.

2 Jednoparametrické metody

V této kapitole je uveden přehled jednoparametrických metod využívaných v dostupné literatuře pro úlohu deinterleavingu. Předpokladem jednoparametrických metod je, že jediný dostupný parametr, popisující přijatou sekvenci impulzů, je čas příchodu impulzů TOA. Mezi tyto algoritmy patří pulse sorting algoritmus, sequence search algoritmus, delta- τ histogram, cumulative difference histogram (CDIF), sequential difference histogram algoritmy (SDIF), transformace a deinterleaving pomocí trackingu. TOA deinterleaving algoritmy odhadují hodnotu PRI přijatých sekvencí impulzů. Pomocí odhadnutých PRI hodnot je přijatá sekvence rozdělena na sekvence impulzů od různých emiterů [25].

2.1 Pulse sorting algoritmus

Pulse sorting algoritmus je jednoduchý algoritmus pro oddělení sekvencí impulzů jednotlivých emiterů z přijaté sekvence. Algoritmus v přijaté sekvenci impulzů hledá tři sousedící impulzy se stejnou hodnotou PRI. Pokud je takováto trojice nalezena, je rozšířena do obou směrů sekvence. Odpovídající impulzy jsou vyjmuty z původní sekvence a jsou prohlášeny za sekvenci impulzů od jednoho emiteru. Postup je opakován, dokud se nenaleznou všechny trojice impulzů se shodným PRI. Dále je postup opakován pro dvojice impulzů.

Tento algoritmus je velmi prostý, ale početně náročný. Pro komplexní prostředí (tj. prostředí husté s agilními signály, nebo s komplexní PRI modulací apod.) není vhodný [25].

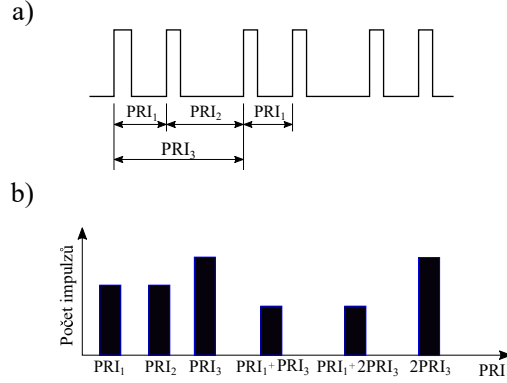
2.2 Sequence search algoritmus

Sequence search algoritmus je velmi přesný a spolehlivý, avšak početně náročný. Algoritmus počítá všechny možné hodnoty PRI a následně hledá sekvence impulzů s jednotlivými PRI v přijaté sekvenci.

V prvním kroku určíme počáteční odhad PRI. Hledáme sekvenci impulzů v přijaté sekvenci, začínající od prvního impulzu, která má PRI rovno počátečnímu odhadu. Pokud takovou sekvenci nenalezneme, hledání začíná od druhého impulzu v řadě se stejným odhadem PRI atd. Pokud nenalezneme žádnou sekvenci pro počáteční odhad PRI, postup je opakován s druhým odhadem PRI. Nalezené sekvence impulzů z přijatého signálu odebereme a testování je opakováno [17], [27].

2.3 Delta- τ histogram/TOA difference histogram

Delta- τ histogram, neboli TOA difference histogram je další jednoduchý algoritmus pro de-interleaving pomocí parametru TOA. Tento algoritmus vypočítá všechny možné hodnoty rozdílů TOA. Jejich četnosti jsou následně vykresleny do grafu (Obrázek 2.1). Hodnoty s nejvyšší četností jsou považovány za PRI jednotlivých zdrojů [25], [27], [41].



Obrázek 2.1: a) přijatá sekvence impulzů b) delta- τ histogram

2.4 Cumulative difference (CDIF) histogram

CDIF je modifikovaný delta- τ histogram. CDIF histogram zobrazuje TOA difference různých řádů. Diferencí n -tého řádu nazveme rozdíl mezi libovolným impulzem a impulzem, který ho předchází o n impulzů.

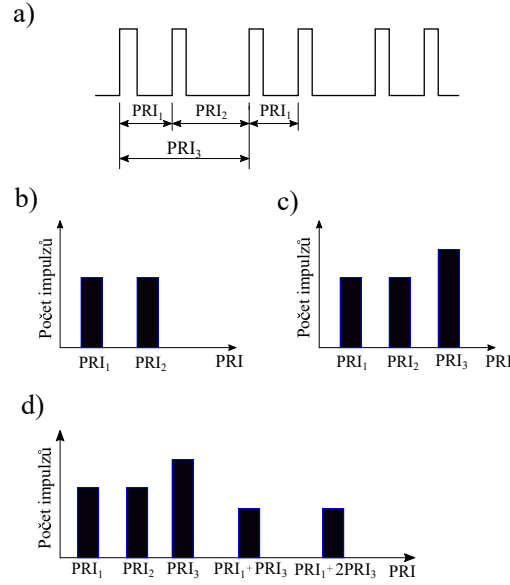
V prvním kroku vykreslíme histogram diferencí TOA prvního řádu. Pokud jakýkoliv sloupec překročí určitou mez, je tato difference TOA považována za PRI. Odpovídající sekvenční pulzy z přijatého impulzního vláčku vyjme a postup je opakován. Pokud žádný sloupec nepřekročí mez, vykreslíme do histogramu difference druhého řádu (Obrázek 2.2). Takto pokračujeme, dokud nevznikne delta- τ histogram, nebo nedojde k separaci přijaté sekvenční impulzů [25], [29].

2.5 Sequential difference (SDIF) histogram

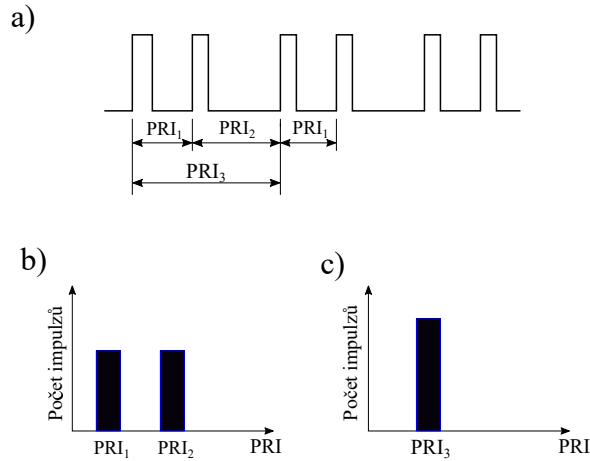
SDIF je modifikovaný CDIF algoritmus. Histogramy diferencí různých řádů jsou vykreslovány jednotlivě (Obrázek 2.3). Díky tomu má algoritmus menší nároky na výpočet než u CDIF algoritmu [23], [25], [29].

Výše zmíněné algoritmy jsou vhodné pro identifikaci konstantní modulace PRI. Dále je možné rozeznat modulace stagger či dwell & switch PRI, jelikož je možné jejich zobecnění na konstantní modulaci PRI. Avšak uvedené algoritmy nejsou schopny jednotlivé PRI přiřadit k jednomu emiteru. Například u modulace stagger PRI se dvěma úrovněmi algoritmus identifikuje dvě konstantní modulace PRI. Pro komplexní typy signálů a prostředí nejsou tyto algoritmy účinné i přes jejich vysokou přesnost. Časová složitost výše uvedených algoritmů je $O(N^2)$, kde N je počet přijatých impulzů. Navíc jsou tyto algoritmy citlivé na chybějící impulzy.

2.6. INTEGRÁLNÍ TRANSFORMACE (PRI TRANSFORMACE)



Obrázek 2.2: CDIF: a) přijatá sekvenční impulsů b) difference prvního řádu c) difference druhého řádu d) difference třetího řádu



Obrázek 2.3: SDIF: a) přijatá sekvenční impulsů b) difference prvního řádu c) difference druhého řádu

2.6 Integrální transformace (PRI transformace)

Nechť t_n , $n = 0, \dots, N-1$ je čas příchodu n -tého impulsu a N je počet přijatých impulsů. Za předpokladu, že TOA je jediný známý parametr impulsu, může být sekvenční impulsů popsána vztahem

$$g(t) = \sum_{n=0}^{N-1} \delta(t - t_n), \quad (2.1)$$

kde N je celkový počet pulzů. Uvažujme následující transformaci funkce $g(t)$

$$D(\tau) = \sum_{n=1}^{N-1} \sum_{m=0}^{n-1} \delta(\tau - t_n + t_m) \exp \left\{ \frac{2\pi i t_n}{t_n - t_m} \right\}. \quad (2.2)$$

2. JEDNOPARAMETRICKÉ METODY

Mějme $[\tau_{min}, \tau_{max}]$ rozmezí PRI hodnot. Rozdělíme-li tento interval na K intervalů, nazývaných PRI biny, dostaneme šířku jednotlivých binů

$$b = \frac{(\tau_{max} - \tau_{min})}{K} \quad (2.3)$$

a jejich střed

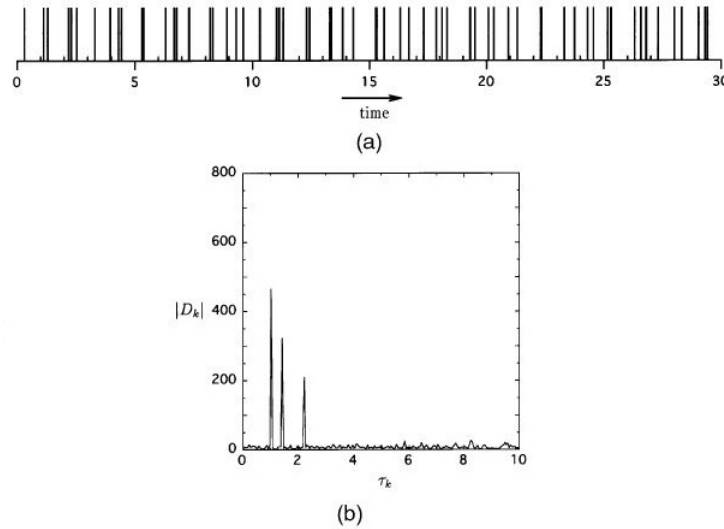
$$\tau_k = (k - \frac{1}{2})b + \tau_{min}. \quad (2.4)$$

Hodnoty PRI pak mohou být určeny jako PRI spektrum $|D_k|$ (Obrázek 2.4), kde

$$D_k = \int_{\tau_k - \frac{b}{2}}^{\tau_k + \frac{b}{2}} D(\tau) d\tau = \sum_{\tau_k - \frac{b}{2} < t_n - t_m < \tau_k + \frac{b}{2}} \exp \left\{ \frac{2\pi i t_n}{t_n - t_m} \right\}. \quad (2.5)$$

Výše popsaná transformace identifikuje konstantní modulaci PRI, avšak v případě jitter modulační PRI spektrum není dostatečně velké pro identifikaci [19], [26], [31], [42]. Je pro to vhodné výše definovanou PRI transformaci modifikovat a to posunem časových počátků a překrytím PRI binů [19], [31].

Modifikovaná PRI transformace detekuje modulaci konstantní a jitter PRI, avšak nedokáže rozeznat modulaci stagger PRI. Pro tento případ je možné metodu spojit s SDIF algoritmem [23], [26].

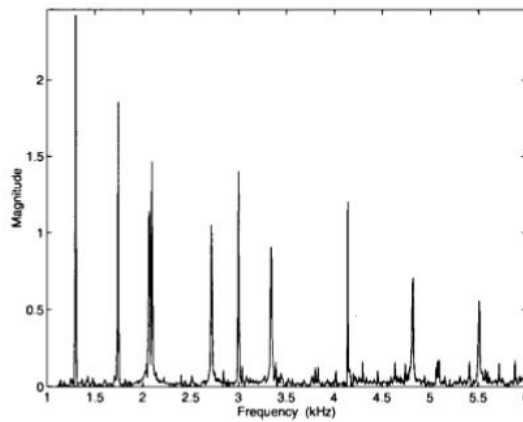


Obrázek 2.4: a) přijatá sekvence impulzů od tří zdrojů s hodnotami PRI 1, $\sqrt{2}$, $\sqrt{5}$ b) PRI spektrum přijaté sekvence impulzů [31].

2.7 Diskrétní Fourierova transformace

Pomocí diskrétní Fourierovy transformace (DFT) pro deinterleaving je odhadnut počet přijatých sekvencí impulzů od jednoho zdroje a jejich frekvence, kdy jediná dostupná informace je čas příchodu jednotlivých impulzů t_n , $n = 0, \dots, N$, kde N je počet přijatých impulzů.

V prvním kroku spočteme $x(n) = \exp \left\{ \frac{2\pi i t_n}{t_{max}} \right\}$, $n = 0, \dots, N-1$, kde $t_0 = 0$ a i je imaginární jednotka. Signál $x(n)$ můžeme chápat jako interval $[t_0, t_{N-1}]$, jehož délka je normalizovaná na hodnotu přibližně 2π a následně je tento normalizovaný interval ovinut na jednotkovou kružnici. V druhém kroku je vypočtena DFT signálu $x(n)$, pro určení počtu sekvencí a jejich frekvencí. Počet emiterů pak určíme, jako počet M frekvencí s největší amplitudou (Obrázek 2.5) [32].



Obrázek 2.5: Závislost frekvence-amplitudy pro signál splňující podmínky (i), (ii), (iii), [32].

2.8 Vlnková transformace

Mějme signál popsán vztahem (2.1) a mateřskou vlnku ve tvaru

$$\psi(t) = M^{1/2} \cdot \chi \left(\frac{t}{M} \right) e^{2\pi i t}, \quad (2.6)$$

kde $\chi(t)$ je obdélníkové okno jednotkové délky a M je libovolný parametr. Vlnková transformace $D(T, t)$ je pak definovaná jako

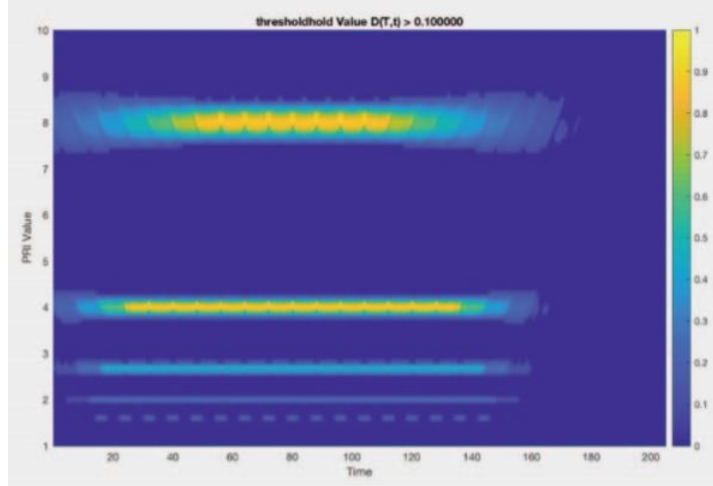
$$D(T, t) = \frac{T}{M} \left| \int x(t') \frac{1}{\sqrt{T}} \cdot \psi^* \left(\frac{t' - t}{T} \right) dt' \right|^2. \quad (2.7)$$

Jestliže hodnota funkce $D(T, t)$ překročí určitou mez, prohlásíme hodnotu T za PRI jednoho emiteru [4], [10].

Použitím vlnkové transformace z rovnice (2.7), dostaneme spektrální energii, kterou je možné znázornit graficky pomocí škálogramu. Obrázek 2.6 zobrazuje škálogram tří přijatých sekvencí. Ze škálogramu je možné identifikovat sekvence $x_1(t)$, $x_2(t)$, $x_3(t)$ a jejich

hodnoty PRI. Tato metoda je vhodná pro modulaci PRI typu konstantní, stagger a komplexní (jitter, wobulated atd.) [4], [9], [10]. Dále lze v literatuře najít spojení vlnkové transformace s neuronovými sítěmi. Neuronové sítě jsou použity pro identifikaci již známých emiterů na základě spektrální energie, určené pomocí vlnkové transformace [10].

Autor [4] zkoumá chování algoritmu v případě agility signálu v PRI a RF na simulovaných datech. Ze simulace vyplývá, že vlnková transformace je účinná pro identifikaci takto agilních signálů, avšak není zkoumán vliv zkreslujících vlivů, jako jsou chybějící nebo odlehlé impulzy.



Obrázek 2.6: Škálogram spektrální energie $D(T, t)$ tří zdrojů [10].

2.9 Deinterleaving pomocí trackingu

Dalším přístupem deinterleavingu přijaté sekvence impulzů je nahlížení na problém jako na sledování cílů neboli tracking. Při trackingu je cíl popsán stavovými veličinami a to pozicí cíle a jeho rychlostí. Pokud nahlížíme na emiter jako na cíl, stavovým veličinám pak odpovídá čas příchodu impulzu TOA a PRI jednotlivých emiterů. Tyto stavové veličiny, pak mohou být odhadovány pomocí Kalmanova filtru a následně porovnány s přijatými impulzy. Výběr impulzů, které nejpravděpodobněji patří ke stejnému emiteru, je prováděn například multiple hypothesis tracking algorithmem [21], nebo pomocí forward dynamic programming (FDP) či probabilistic teacher [30]. Časová složitost Kalmanovy filtrace je řádu $O(N \log N)$.

Změnu stavových veličin cílu můžeme popsat pomocí Markovova řetězce [21], [22]. Model sekvence impulzů definujeme jako

$$\mathbf{x}_{k+1} = \Phi \mathbf{x}_k + \begin{pmatrix} q_P(k) \\ q_T(k) \end{pmatrix}, \quad (2.8)$$

kde $\mathbf{x}_k^T = (\text{PRI}(k) \text{ TOA}(k))$ je dvoudimenzionální stavový vektor, jehož veličiny chceme odhadnout. Přechodová matice Φ je rovna

$$\Phi = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad (2.9)$$

2.9. DEINTERLEAVING POMOCÍ TRACKINGU

a $(q_P(k) \ q_T(k))^T$ je bílý šum s kovarianční maticí \mathbf{Q} a s nulovými středními hodnotami. Dále je měřený čas příchodu y_k zatížen šumem, tedy

$$y_k = \mathbf{\Psi} \mathbf{x}_k + v_k, \quad \mathbf{\Psi} = \begin{pmatrix} 0 & 1 \end{pmatrix}, \quad (2.10)$$

kde v_k je bílý šum s kovarianční maticí \mathbf{R} a s nulovými středními hodnotami.

Pro odhad stavového vektoru pro model (2.8) jsou užity Kalmanovy rovnice ve tvaru

$$\begin{aligned} \mathbf{x}(k | k) &= \mathbf{x}(k | k-1) + \mathbf{K}(k) e_k, \\ \mathbf{x}(k+1 | k) &= \mathbf{\Phi} \mathbf{x}(k | k), \end{aligned} \quad (2.11)$$

kde $\mathbf{K}(k)$ je Kalmanův zisk [21], [22], [30].

3 Víceparametrické metody

V této kapitole je uveden přehled víceparametrických metod využívaných v dostupné literatuře pro úlohu deinterleavingu. Výše popsané metody využívají pro deinterleaving pouze jeden parametr popisující impuls a to čas příchodu TOA. V dnešní době jsou přijaté impulzy komplexní a elektromagnetické prostředí je hustší. Z tohoto důvodu je výhodnější využívat pro deinterleaving více parametrů popisujících přijaté impulzy. Jeden z těchto přístupů jsou statistické metody, ty přiřazují prvky do shluků podle podobnosti určitých parametrů. Mezi ně patří support vector machine, shluková analýza nebo shlukování na základě hustoty. Dále lze využít neuronové sítě. V této kapitole jsou tyto metody popsány pro úlohu deinterleavingu.

3.1 Support vector machine (SVM)

SVM je algoritmus minimalizující strukturální risk. Problém minimalizace strukturálního risku je převeden na maximalizaci vzdálenosti rozdělovací nadroviny klasifikátoru k bodům z trénovací množiny, tato vzdálenost se nazývá okraj. Vhodným vyjádřením tohoto kritéria se vektory trénovací množiny vyskytují pouze ve formě skalárních součinů, což umožňuje použití jádrových funkcí (kernel function) k zobecnění rozdělovací nadroviny na nelineární nadplochu (nelineární klasifikátor) [18].

Nechť $\mathbf{V} \subseteq \mathbb{R}^M$, kde M je počet parametrů popisující impuls a $\mathbf{V} = \{\mathbf{v}_i\}$, $i = 1, \dots, N$, kde N je počet přijatých impulzů. Pro nalezení rozdělovací nadroviny je použita nelineární transformace Φ z prostoru \mathbf{V} do prostoru vyšší dimenze. Hledáme nejmenší ohraničenou sféru s poloměrem R a středem v bodě a , tedy

$$\|\Phi(\mathbf{v}_i) - a\|^2 \leq R^2 + \xi_i, \quad \xi_i \geq 0, \quad (3.1)$$

kde ξ_i je relaxační proměnná pro neseparabilní případ. Pro řešení tohoto problému zavedeme Lagrangian

$$L = R^2 - \sum_i (R^2 + \xi_i - \|\Phi(\mathbf{v}_i) - a\|^2) \beta_i - \sum_i \xi_i \mu_i + C \sum_i \xi_i, \quad (3.2)$$

kde $\beta_i, \mu_i \geq 0$ jsou lagrangerovy multiplikátory a C je konstanta.

Jestliže

1. $\beta_i = C$, bod $\Phi(\mathbf{v}_i)$ leží vně sféry s poloměrem R . Tyto body se nazývají odlehlé a leží mimo shluk,
2. $0 < \beta_i < C$, bod $\Phi(\mathbf{v}_i)$ leží na hranici sféry. Tyto body se nazývají support vectors a leží na hranici shluku,
3. ostatní body leží uvnitř sféry a tedy uvnitř shluku.

SVM algoritmus je časově složitý podobně jako algoritmy z kapitoly 2, tedy $O(N^2)$ [15], [38], [43]. Za účelem snížení časové složitosti navrhuje autor [38] rozdělit přijaté impulzy náhodně do m podmnožin. Dále na jednotlivé podmnožiny aplikovat SVM algoritmus a nakonec jednotlivé shluky spojit. Tím se sníží časová složitost na $mO((N/m)^2)$. V literatuře není diskutováno chování metody vůči zkreslujícím jevům (chybějící, odlehlé impulzy atd.).

3.2 K-means algoritmus

K-means algoritmus hledá pro danou množinu $\mathbf{X} = \{x_1, \dots, x_n\}$ vektory $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k$, kde $k < n$ tak, aby střední kvadratická odchylka prvků množiny \mathbf{X} od vektorů $\boldsymbol{\mu}_i$ byla minimální. Algoritmus hledá takové vektory, ke kterým je euklidovská vzdálenost dat co nejmenší.

Pro množinu dat $\mathbf{V} = \{\mathbf{v}_i\}$, $i = 1, \dots, N$, kde \mathbf{v}_i je vektor parametrů popisující přijatý impuls a N je počet přijatých impulsů, algoritmus hledá rozdělení $\mathbf{P}_k = \{\mathbf{C}_1, \dots, \mathbf{C}_k\}$ množiny \mathbf{V} , tak aby minimalizovalo funkci

$$f(\mathbf{P}_k) = \sum_{i=1}^k \sum_{\mathbf{v}_l \in \mathbf{C}_i} d(\mathbf{v}_l, \mathbf{m}_i), \quad (3.3)$$

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{\mathbf{v}_l \in \mathbf{C}_i} \mathbf{v}_l, \quad (3.4)$$

kde \mathbf{m}_i je pozice i -tého středu shluku a n_i je počet prvků ve shluku \mathbf{C}_i .

K-means algoritmus iterativně opakuje dva kroky:

1. krok: všechna data \mathbf{v}_i jsou přiřazena k jednotlivým shlukům tak, aby se minimalizovala euklidovská vzdálenost d ,
2. krok: vypočítají se nové hodnoty středů shluků jako střední hodnoty jednotlivých množin \mathbf{C}_i .

Kroky 1 a 2 se opakují, dokud se alespoň jeden vektor z množiny \mathbf{X} klasifikuje do jiné třídy, než byl klasifikován v předcházejícím kroku.

Časová složitost tohoto algoritmu je efektivní díky své komplexnosti, $O(nkl)$, kde n je počet impulsů, k je počet shluků a l je počet iterací potřebných pro konvergenci algoritmu. Dále je algoritmus funkční pouze pro konvexní shluky, v případě nekonvexnosti rozdělí algoritmus jeden shluk do více shluků.

K-means algoritmus je velmi citlivý na počáteční volbu počtu shluků k a jejich středů. Řešením tohoto problému je spojení algoritmů SVM a K-means. V prvním kroku je rozdělena malá část dat do k shluků pomocí metody SVM, zde je třeba určit uživatelem parametry q, C . Dále jsou určeny středy jednotlivých shluků a tyto parametry jsou použity pro inicializaci K-means metody [15], [18].

V literatuře není diskutováno chování metody vůči zkreslujícím jevům (chybějící, odlehlé impulsy atd.).

3.3 Density-based spatial clustering of applications with noise (DBSCAN)

DBSCAN je algoritmus založený na shlukování podle hustoty. Tyto algoritmy definují shluky jako oblasti s vysokou hustotou, které jsou rozděleny oblastí s nižší hustotou. Hlavní výhody těchto algoritmů jsou, že není potřeba znát předem počet shluků a schopnost identifikovat shluky libovolných tvarů [20].

DBSCAN algoritmus definuje shluky jako maximální množinu prvků s dostatečně vysokou hustotou. Klasický DBSCAN algoritmus nechává volbu dvou parametrů na uživateli a to $MinPts$ a ϵ . Parametr $MinPts$ určuje minimální počet sousedících bodů pro určení tzv. jádrového bodu a ϵ určuje maximální vzdálenost dvou bodů z jednoho shluku [24].

Algoritmus hledá body ležící v ϵ -okolí náhodně vybraného bodu \mathbf{p} . Pokud je počet těchto bodů větší jak parametr $MinPts$, bod je označen jako jádrový bod a je přidán do shluku C_1 . Následně jsou vyšetřovány body v jeho okolí, zda jsou jádrovými prvky. Pokud ano jsou přidány do stejného shluku. Pokud žádný bod v ϵ -okolí není jádrovým bodem, vyšetřuje se další nenavštívený a náhodně vybraný bod. Takto se postupuje dokud nejsou všechny body navštíveny.

Časová náročnost DBSCAN algoritmu závisí na způsobu implementace. Průměrná časová náročnost je $O(N \log(N))$, kde N je počet vstupních bodů. V nejhorším případě je časová náročnost $O(N^2)$ [1].

3.4 Bayesovské metody

Na sekvenci impulzů od více emiterů můžeme pohlížet jako na náhodný výběr $\mathbf{x}_1, \dots, \mathbf{x}_N$ ze směsi K vícerozměrných normálních rozdělání s různými parametry $\boldsymbol{\theta}_i$. Na úlohu deinterleavingu pak pohlížíme jako na úlohu nalezení parametrů rozdělání $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K$. V bayesovské statistice chápeme parametry pravděpodobnostního rozdělání jako náhodné veličiny. Při odhadu těchto parametrů se využívá předběžné znalosti o těchto parametrech a informací ze vstupních dat.

Nechť X_1, \dots, X_N je náhodný výběr z rozdělání, které má hustotu $p(\mathbf{x}|\boldsymbol{\theta})$ vzhledem k σ -konečné míře ν , přičemž $\boldsymbol{\theta} \in \Omega$ je nějaký parametr. Budeme předpokládat, že Ω je neprázdná borelovská množina. Nechť $\boldsymbol{\theta}$ má apriorní hustotu $p(\boldsymbol{\theta})$ vzhledem k σ -konečné míře λ . Nechť se realizoval takový vektor $\mathbf{X} = \mathbf{x} = (x_1, \dots, x_N)^T$, že

$$0 < \int_{\Omega} p(x_1|\boldsymbol{\theta}) \dots p(x_N|\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\lambda(\boldsymbol{\theta}) < \infty. \quad (3.5)$$

Pak podle Bayesovy věty má aposteriorní hustota $p(\boldsymbol{\theta}|\mathbf{x})$ tvar

$$p(\boldsymbol{\theta}|\mathbf{x}) = c p(x_1|\boldsymbol{\theta}) \dots p(x_n|\boldsymbol{\theta}) p(\boldsymbol{\theta}), \quad (3.6)$$

kde $c > 0$ je konstanta volená tak, aby platilo

$$\int_{\Omega} p(\boldsymbol{\theta}|\mathbf{x}) d\lambda(\boldsymbol{\theta}) = 1. \quad (3.7)$$

Apriorní pravděpodobnost dává informace o parametru $\boldsymbol{\theta}$ ještě před získáním realizace \mathbf{X} . Vztah (3.6) nazýváme Bayesovský model. Jestliže apriorní a aposteriorní pravděpodobnost patří do stejné třídy rozdělání pak jsou tato rozdělání konjugovaná a patří do stejného konjugovaného systému rozdělání, [2].

3.5 Neuronové sítě

Dalším přístupem pro deinterleaving pomocí více parametrů jsou neuronové sítě (NN). NN lze využít pro klasifikaci vstupních dat na základě vzájemné podobnosti či pro predikci vývoje určité veličiny. Neuronové sítě lze rozlišit podle topologie propojení neuronů na dopředné a rekurentní sítě. Pro úlohu deinterleavingu byly navrženy self-organizing neuronové sítě, rekurentní nebo rekurzivní neuronové sítě.

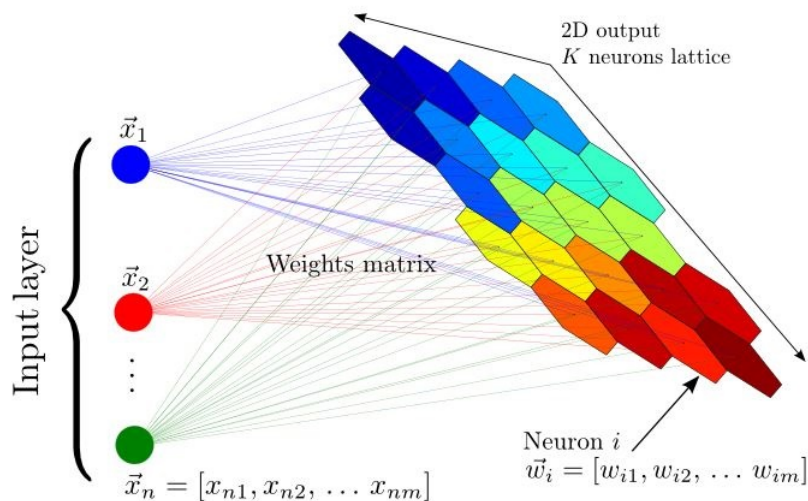
3.5.1 Self-organizing neuronové sítě

Self-organizing NN je označována jako kompetitivní síť založená na učení bez učitele. Tato NN je využívána v mnoha oblastech pro rozdělení prvků do kategorií na základě vzorů. Mezi tyto NN se řadí Self-organizing map (SOM), neboli Kohonenova neuronová síť a Fuzzy ART. Časová náročnost těchto algoritmů je $O(MN)$, kde M je počet impulzů a N je počet neuronů [13], [14].

Self-organizing map

SOM se skládá z jedné vrstvy neuronů tzv. kompetitivní vrstvy, kde její uspořádání je ve většině případů dvoudimenzionální a obecně neexistuje omezení v počtu neuronů. Vstupy sítě jsou propojeny se všemi neurony a jsou ohodnoceny odpovídajícími vahami. Sousedící neurony navíc mají mezi sebou také vazby (Obrázek 3.1).

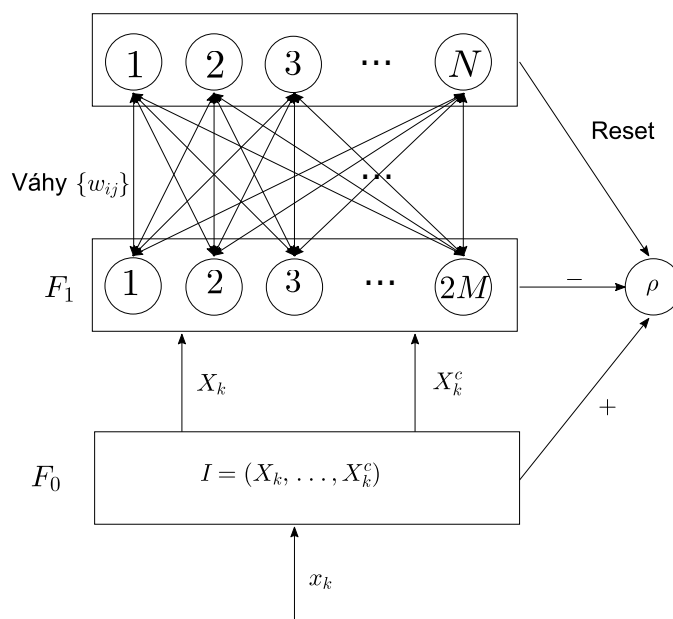
SOM je učena iterativně. V čase t je na vstup přiveden prvek \mathbf{x} z množiny vstupních dat, kde jejich počet je n . Hledá se takový neuron \mathbf{w}_i jehož vzdálenost od prvku \mathbf{x} je minimální [13], [14].



Obrázek 3.1: SOM síť [14].

Fuzzy ART

Fuzzy ART je neuronová síť, která je založena na kompetitivním učení bez učitele. Skládá se ze tří vrstev. První vrstva F_0 neobsahuje žádné neurony a normalizuje vstupní data. Její výstup \mathbf{I} vstupuje do druhé vrstvy F_1 s počtem neuronů $2M$, která určí příslušnost vstupu s neuronem v poslední vrstvě F_2 s počtem neuronů N , kde M je dimenze vstupu a N je počet výstupních kategorií (Obrázek 3.2). Fuzzy ART následně vybere ten neuron, který má nejvyšší příslušnost a označí ho jako vítězný a jeho váha \mathbf{w}_j je upravena [3], [14], [16].



Obrázek 3.2: Struktura Fuzzy ART.

V práci [13] je provedeno testování SOM a FuzzyArt na simulovaných datech s agilitou v PRI a PW. Ze simulací vyplývá, že uvedená metoda je účinná i pro agilní impulzy. V literatuře není diskutováno chování metody vůči zkreslujícím jevům (chybějící, odlehlé impulzy atd.).

4 Vybrané metody

Na základě provedené rešerše literatury, zaměřující se na deinterleaving sekvenční impulzů, byly vybrány dvě metody pro implementaci a ověření jejich účinnosti. Byly zvoleny metody řadící se mezi víceparametrické, které jsou vhodnější pro sekvenční s komplexnější PRI modulací a metody vyžadující co nejmenší zásah uživatelem (tj. volbu vstupních parametrů). Dalším důležitým kritériem bylo automatické určení počtu emiterů, jelikož v reálném provozu není tato informace známa. Na základě těchto kritérií byly vybrány metody DBSCAN a variační bayesovské (VB) metody. Jednokanálový autonomní deinterleaving je pak proveden pomocí dvou parametrů RF a PW.

4.1 DBSCAN

DBSCAN algoritmus definuje shluky jako maximální množinu prvků s dostatečně vysokou hustotou, jejíž prvky jsou spojené vzhledem k hustotě (density-connected).

Pro použití DBSCAN algoritmu je nutné definovat následující pojmy [11]. Uvažujme libovolné konstanty ϵ , $MinPts \in \mathbb{R}^+$ a množinu bodů D .

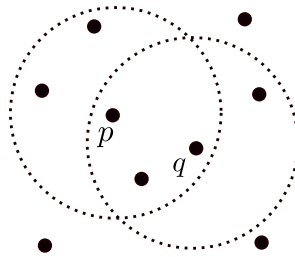
Definice 4.1.1 (ϵ -okolí bodu). ϵ -okolí bodu $p \in D$ je množina bodů $q \in D$, pro které platí $d(p, q) \leq \epsilon$, kde $d(p, q)$ je vzdálenost mezi bodem p a q .

Poznámka. Vzdálenost $d(\cdot, \cdot)$ lze uvažovat jako libovolnou vzdálenost, například manhattanovskou, euklidovskou, maximovou atd. V této práci uvažujeme euklidovskou vzdálenost.

Poznámka. Řekneme, že body p, q jsou sousedící, pokud platí nerovnost $d(p, q) \leq \epsilon$.

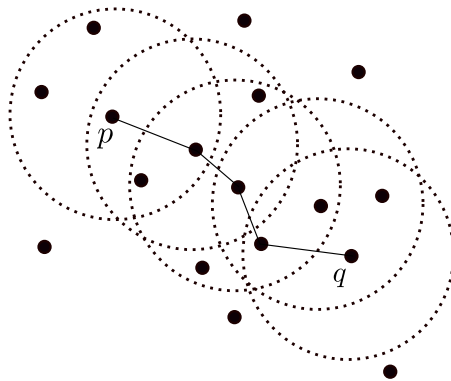
Definice 4.1.2 (Jádrový bod). Bod p nazveme jádrovým bodem, jestliže počet bodů ležících v ϵ -okolí bodu p je větší jak $MinPts$.

Definice 4.1.3 (Přímo dosažitelný bod vzhledem k hustotě (directly density-reachable point)). Řekneme, že bod q je přímo dosažitelný vzhledem k hustotě od bodu p , jestliže p je jádrovým bodem a bod q leží v ϵ -okolí bodu p .

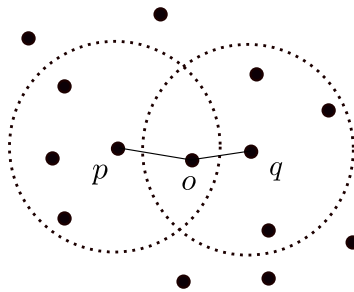


Obrázek 4.1: Bod q je přímo dosažitelný vzhledem k hustotě od bodu p , pro $MinPts = 3$.

Definice 4.1.4 (Dosažitelný bod vzhledem k hustotě (density-reachable point)). Řekneme, že bod q je dosažitelný vzhledem k hustotě od bodu p , jestliže existuje řetězec bodů p_1, \dots, p_n , kde $p_1 = p, p_n = q$ takový, že bod p_{i+1} je přímo dosažitelný vzhledem k hustotě od bodu p_i .

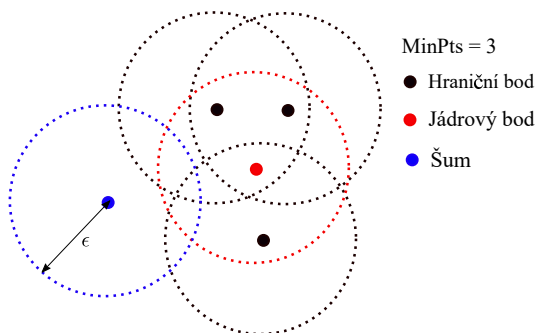
Obrázek 4.2: Bod q je dosažitelný vzhledem k hustotě od bodu p .

Definice 4.1.5 (Spojené body vzhledem k hustotě (density-connected points)). Body p, q nazveme spojenými body vzhledem k hustotě, jestliže existuje bod o takový, že oba body p a q jsou od bodu o dosažitelné vzhledem k hustotě.

Obrázek 4.3: Body q a p spojené vzhledem k hustotě.

Definice 4.1.6 (Hraniční bod). Nejádrový bod p nazveme hraničním bodem, jestliže leží v ϵ -okolí libovolného jádrového bodu.

Definice 4.1.7 (Šum). Bod, který není jádrovým ani hraničním bodem, nazveme šumem.

Obrázek 4.4: Typy bodů pro $MinPts = 3$.

Definice 4.1.8 (Shluk). Mějme množinu bodů D . Podmnožinu $C \subseteq D$ nazveme shlukem, jestliže $\forall p, q \in C$ platí tyto dvě podmínky:

- (i) jestliže $p \in C$ a q je dosažitelný vzhledem k hustotě z bodu p , pak $q \in C$,
- (ii) jestliže $p, q \in C$ pak body p a q jsou spojené vzhledem k hustotě.

4.1. DBSCAN

4.1.1 Vstupní parametry

Hlavní výhodou DBSCAN algoritmu je, že nevyžaduje znát dopředu počet shluků jako například u k-means algoritmu a také schopnost identifikovat shluky libovolných tvarů. Jediné vstupní parametry, které je nutné určit uživatelem jsou konstanty ϵ a $MinPts$ [1].

Parametr ϵ určuje maximální vzdálenost dvou bodů, patřících do jednoho shluku. Parametr $MinPts$ je minimální počet sousedících bodů pro definování jádrového bodu. Hodnoty těchto dvou parametrů je možné určit z rozložení vstupních dat. DBSCAN algoritmus je velmi citlivý na volbu těchto vstupních parametrů [24].

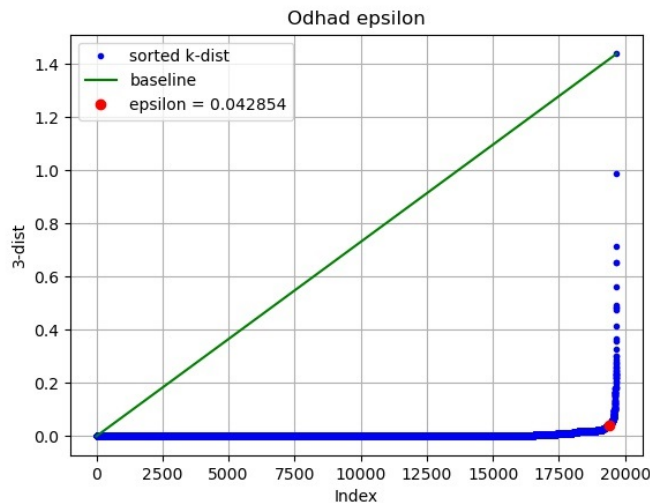
Parametr $MinPts$

Parametr $MinPts$ určuje minimální počet sousedících bodů pro určení jádrového bodu. Obecně se hodnota $MinPts$ určuje pomocí dimenze množiny vstupních dat \mathbf{D} , kde musí platit $MinPts \geq 3$. Obecně volíme $MinPts = 2 \dim(\mathbf{D})$, avšak pro data s šumem či data velkých rozměrů je vhodná větší hodnota parametru $MinPts$ [37].

Parametr ϵ

Parametr ϵ určuje maximální vzdálenost dvou bodů, patřících do jednoho shluku. Jestliže určíme hodnotu parametru příliš malou, velká část dat nebude rozdělena do shluků. Tyto body budou prohlášeny za šum, jelikož nesplní podmínky hustoty. Naopak pokud hodnota ϵ bude příliš velká, jednotlivé shluky budou spojeny do jednoho velkého shluku.

Optimální hodnotu ϵ lze odhadnout z grafu k -nejbližších sousedů. Graf k -nejbližších sousedů je tvořen ze vzestupně seřazených hodnot k -dist, kde k -dist je vzdálenost mezi k -tým nejbližším bodem od bodu \mathbf{p}_i pro $i = 0, 1, \dots, N$. ϵ je pak hodnota, v níž se objeví tzv ohyb (Obrázek 4.5). Hodnotu k volíme dle vztahu $k = MinPts - 1$ [37].



Obrázek 4.5: Vzdálenosti třetích nejbližších sousedů seřazené vzestupně.

4.1.2 Postup algoritmu

Jednotlivé kroky algoritmu, při daných parametrech $\epsilon, MinPts$, jsou následující [28]. Pro $i = 0$ označíme všechny body jako nenavštívené.

1. Vybere se libovolný bod \mathbf{p} z množiny vstupních dat \mathbf{D} , který ještě nebyl navštíven.
2. Jsou určeny sousedící body, které leží v ϵ -okolí bodu \mathbf{p} . Jestliže počet sousedících bodů je menší jak hodnota $MinPts$, pak je vybraný bod \mathbf{p} označen jako šum a pokračuje se krokem 4. Jestliže je počet sousedících bodů větší nebo roven hodnotě $MinPts$, pak je vybraný bod \mathbf{p} označen jako jádrový bod. Vytvoří se shluk \mathbf{C}_i , kde $\mathbf{p} \in \mathbf{C}_i$.
3. Body ležící v okolí bodu \mathbf{p} , také náleží shluku \mathbf{C}_i . Pro všechny sousedící body jádrového bodu \mathbf{p} je opakován krok 2, dokud existují sousedící body bodů patřících do shluku \mathbf{C}_i .
4. Vybere se nový nenavštívený bod z množiny \mathbf{D} a $i = i + 1$.
5. Opakují se kroky 1 – 4, dokud nejsou všechny body navštíveny.

4.2 Variační bayesovské metody

Jak již bylo zmíněno v podkapitole 3.4 máme Bayesovský model

$$p(\boldsymbol{\theta}|\mathbf{x}) = cp(x_1|\boldsymbol{\theta}) \dots p(x_n|\boldsymbol{\theta})p(\boldsymbol{\theta}), \quad (4.1)$$

pro realizaci $\mathbf{X} = \mathbf{x} = (x_1, \dots, x_N)$, kde $c > 0$ je konstanta volená tak, aby platilo

$$\int_{\Omega} p(\boldsymbol{\theta}|\mathbf{x}) d\lambda(\boldsymbol{\theta}) = 1. \quad (4.2)$$

Ve většině případů je velmi obtížné určit hustotu aposteriorní pravděpodobnosti $p(\boldsymbol{\theta}|\mathbf{x})$ a je nutné danou hustotu vhodně aproximovat. K tomu jsou využívány variační metody či metoda Monte Carlo. V této práci se budeme zabývat variačními metodami.

Myšlenkou variačních metod je vytvořit vhodnou třídu rozdělení neznámých parametrů a z ní vybrat rozdělení, které je v jistém smyslu nejpodobnější hustotě aposteriorního rozdělení $p(\boldsymbol{\theta}|\mathbf{x})$. Podobnost rozdělení je měřena pomocí Kullback-Leiblerovy divergence [6], [7].

4.2.1 Směsový model

Definice 4.2.1. Hustota $p(\mathbf{x})$ náhodného vektoru \mathbf{X} se nazývá hustota pravděpodobnosti K směsových komponent, jestliže

$$p(\mathbf{x}|\boldsymbol{\Theta}) = \sum_{k=1}^K a_k m_k(\mathbf{x}|\boldsymbol{\Theta}_k), \quad (4.3)$$

kde m_k jsou hustoty komponent pocházející ze stejné rodiny rozdělení s lišícími se parametry $\boldsymbol{\Theta}_k$ a a_k jsou váhy směsi pro která platí $a_k > 0$, $\sum_{k=1}^K a_k = 1$. Vektor $\boldsymbol{\Theta}$ obsahuje všechny parametry směsového modelu $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}_1, \dots, \boldsymbol{\Theta}_K, a_1, \dots, a_K\}$.

4.2. VARIČNÍ BAYESOVSKÉ METODY

Příklad 4.2.1. Uvažujeme-li směs K J -rozměrných normálních rozdělání, pak hustota $p(\mathbf{x})$ náhodného vektoru $\mathbf{X} \in \mathbb{R}^J$ je rovna

$$p(\mathbf{x}) = \sum_{k=1}^K a_k \mathcal{N}_J(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad (4.4)$$

kde $\mathcal{N}_J(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ je hustota k -tého J -rozměrného normálního rozdělání s vektorem středních hodnot $\boldsymbol{\mu}_k$ a kovarianční maticí $\boldsymbol{\Sigma}_k$

$$\mathcal{N}_J(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{\sqrt{(2\pi)^J |\boldsymbol{\Sigma}_k|}} \exp \left(-\frac{(\mathbf{x} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)}{2} \right). \quad (4.5)$$

Směsový model (4.4) nazýváme Gaussův smíšený model (GMM).

4.2.2 Exponenciální rodina rozdělání

Mnoho pravděpodobnostních rozdělání patří do široké třídy rozdělání nazývané exponenciální rodina, kde členové této rodiny mají mnoho výhodných vlastností [2], [6].

Definice 4.2.2. Mějme pravděpodobnostní prostor $(\Omega, \mathcal{A}, P_{\Lambda})$ a na něm definovaný náhodný vektor X_1, \dots, X_N . Označme jeho pozorování $\mathbf{x} = (x_1, \dots, x_N)$, hustotu rozdělání p_{Λ} nazveme hustotou exponenciálního typu, jestliže jí lze zapsat ve tvaru

$$p_{\Lambda}(\mathbf{x}) = p(\mathbf{x} | \Lambda) = h(\mathbf{x}) \exp \left(\sum_{n=1}^N \eta_n^T(\Lambda) T_n(\mathbf{x}) - a(\Lambda) \right), \quad (4.6)$$

kde $\eta_n(\Lambda)$, $T_n(\mathbf{x})$, $a(\Lambda)$ jsou reálné funkce. Funkci $T_n(\mathbf{x})$ nazýváme postačující statistikou a systém $\{P_{\Lambda}\}$ exponenciální rodinou.

Poznámka. Vztah (4.6) lze zkráceně zapsat pomocí skalárního součinu

$$p(\mathbf{x} | \Lambda) = h(\mathbf{x}) \exp (\boldsymbol{\eta}^T(\Lambda) T(\mathbf{x}) - a(\Lambda)). \quad (4.7)$$

Poznámka. Hustota pravděpodobnosti přes celý obor hodnot náhodné veličiny \mathbf{X} musí být normovaná, tedy:

$$\int h(\mathbf{x}) \exp (\boldsymbol{\eta}^T(\Lambda) T(\mathbf{x}) - a(\Lambda)) d\mathbf{x} = 1. \quad (4.8)$$

Tím je funkce $a(\Lambda)$ jednoznačně určena vztahem

$$a(\Lambda) = \ln \left(\int h(\mathbf{x}) \exp (\boldsymbol{\eta}^T(\Lambda) T(\mathbf{x})) d\mathbf{x} \right). \quad (4.9)$$

Definice 4.2.3. Mějme hustotu exponenciálního typu (4.6). Hustotu

$$p(\mathbf{x} | \boldsymbol{\lambda}) = h(\mathbf{x}) \exp (\boldsymbol{\lambda}^T T(\mathbf{x}) - a(\boldsymbol{\lambda})), \quad (4.10)$$

pak nazýváme přirozenou parametrizací a parametr $\boldsymbol{\lambda}$ přirozeným parametrem.

Poznámka. Uvažujeme-li přirozenou parametrizaci, z gradientu obou stran vztahu (4.8) podle přirozeného parametru $\boldsymbol{\lambda}$ dostaneme

$$\nabla_{\boldsymbol{\lambda}} a(\boldsymbol{\lambda}) = E[T(\mathbf{x})]. \quad (4.11)$$

4.2.3 Kullback-Leiblerova divergence

Kullback-Leiblerova divergence měří odlišnost dvou pravděpodobnostních rozdělení. V případě variačních metod máme hustotu rozdělení $p(\boldsymbol{\Theta}|\mathbf{x})$, kterou aproximujeme hustotou $q(\boldsymbol{\Theta})$. Míru jejich odlišnosti lze vyjádřit pomocí Kullback-Lieblerovy divergence [6].

Definice 4.2.4. Mějme náhodnou veličinu \mathbf{X} a dvě pravděpodobnostní hustoty $p(\mathbf{x})$ a $q(\mathbf{x})$. Pak Kullback-Leiblerovu divergenci $KL(p(\mathbf{x})||q(\mathbf{x}))$ definujeme jako

$$\begin{aligned} KL(p(\mathbf{x})||q(\mathbf{x})) &= - \int p(\mathbf{x}) \ln(q(\mathbf{x})) d\mathbf{x} - \left(- \int p(\mathbf{x}) \ln(p(\mathbf{x})) d\mathbf{x} \right) = \\ &= \int p(\mathbf{x}) \ln \left(\frac{p(\mathbf{x})}{q(\mathbf{x})} \right) d\mathbf{x} = E_p \left[\ln \left(\frac{p(\mathbf{x})}{q(\mathbf{x})} \right) \right] = \\ &= E_p [\ln p(\mathbf{x})] - E_p [\ln q(\mathbf{x})]. \end{aligned} \quad (4.12)$$

Poznámka. Poznamenejme, že Kullback-Leiblerova divergence je nezáporná a není symetrická, tedy

$$\begin{aligned} KL(p(\mathbf{x})||q(\mathbf{x})) &\geq 0, \\ KL(p(\mathbf{x})||q(\mathbf{x})) &\neq KL(q(\mathbf{x})||p(\mathbf{x})). \end{aligned}$$

Předpokládejme, že máme Bayesovský model (3.6), kde všechny parametry tohoto modelu mají dané apriorní rozdělení s hustotou $p(\boldsymbol{\Theta})$. Dále tento model kromě neznámých parametrů $\boldsymbol{\Theta}$ obsahuje také latentní (skryté) proměnné $\mathbf{z} = (z_1, \dots, z_n)$. Množinu všech neznámých hodnot označíme $\mathbf{Z} = \{\boldsymbol{\Theta}, \mathbf{z}\}$ a množinu pozorovaných dat označíme $\mathbf{x} = (x_1, \dots, x_n)$. Náš model specifikuje sdruženou pravděpodobnost $p(\mathbf{x}, \mathbf{Z})$ a naším cílem je najít aproximaci aposteriorní hustoty $p(\mathbf{Z}|\mathbf{x})$.

Uvažujeme-li hustotu pravděpodobnosti $q(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\phi})$ s parametry $\boldsymbol{\lambda}, \boldsymbol{\phi}$, pak platí

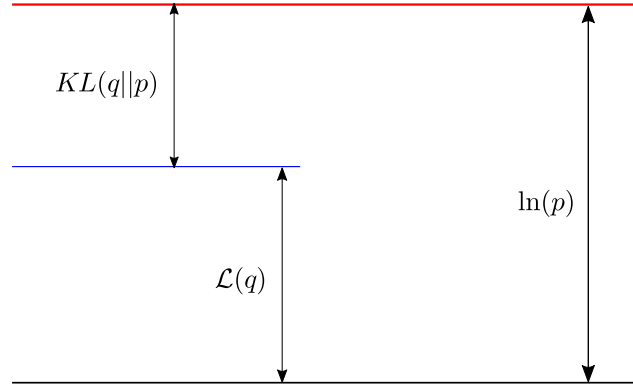
$$\ln p(\mathbf{x}) = \mathcal{L}(q(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\phi})) + KL(q(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\phi})||p(\mathbf{Z}|\mathbf{x})), \quad (4.13)$$

kde definujeme

$$\begin{aligned} \mathcal{L}(q(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\phi})) &= \int q(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\phi}) \ln \left(\frac{p(\mathbf{x}, \mathbf{Z})}{q(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\phi})} \right) d\mathbf{Z} = \\ &= E_q [\ln p(\mathbf{x}, \mathbf{Z})] - E_q [\ln q(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\phi})], \end{aligned} \quad (4.14)$$

$$\begin{aligned} KL(q(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\phi})||p(\mathbf{Z}|\mathbf{x})) &= - \int q(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\phi}) \ln \left(\frac{p(\mathbf{Z}|\mathbf{x})}{q(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\phi})} \right) d\mathbf{Z} = \\ &= E_q [\ln q(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\phi})] - E_q [\ln p(\mathbf{Z}|\mathbf{x})]. \end{aligned} \quad (4.15)$$

Funkcionál $\mathcal{L}(q(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\phi}))$ nazýváme dolní mez. Ilustrace dekompozice (4.13) je na obrázku 4.6.



Obrázek 4.6: Ilustrace dekompozice (4.13).

4.2.4 Mean-field aproximace

Předpokládejme, že prvky neznámých hodnot \mathbf{Z} lze rozdělit do M disjunktních, vzájemně nezávislých množin $\mathbf{Z}_i, i = 1, \dots, M$. Pak hustotu q lze faktorizovat vzhledem k těmto množinám

$$q(\mathbf{Z}) = \prod_{i=1}^M q_i(\mathbf{Z}_i). \quad (4.16)$$

Tento přístup se nazývá mean-field teorie.

Variační metody využívají mean-field aproximaci. Aposteriorní rozdělení s hustotou $p(\mathbf{Z}|\mathbf{x})$ aproximuje tzv. variačním rozdělením s hustotou pravděpodobnosti $q(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\phi})$, které splňuje mean-field teorii [6].

Hustotu variačního rozdělení $q(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\phi})$ splňující (4.16) lze zapsat ve tvaru hustoty exponenciálního typu, tedy

$$q(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\phi}) = q(\boldsymbol{\Theta}|\boldsymbol{\lambda})q(\mathbf{z}|\boldsymbol{\phi}), \quad (4.17)$$

$$q(\boldsymbol{\Theta}|\boldsymbol{\lambda}) = h_p(\boldsymbol{\Theta}) \exp(\boldsymbol{\lambda}^T T_p(\boldsymbol{\Theta}) - a_p(\boldsymbol{\lambda})), \quad (4.18)$$

$$q(\mathbf{z}|\boldsymbol{\phi}) = h_l(\mathbf{z}) \exp(\boldsymbol{\phi}^T T_l(\mathbf{z}) - a_l(\boldsymbol{\phi})), \quad (4.19)$$

Za předpokladu, že hustota variačního rozdělení $q(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\phi})$ aproximuje rozdělení $p(\mathbf{Z}|\mathbf{x})$ a tyto dvě rozdělení patří do exponenciální rodiny, pak vyplývá, že funkce h_p, T_p a h_l, T_l těchto dvou hustot jsou ekvivalentní.

4.2.5 Gradient dolní meze

Úkolem variačních metod je aproximovat hustotu aposteriorního rozdělení $p(\mathbf{Z}|\mathbf{x})$. Hledáme tedy hustotu variačního rozdělení $q(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\phi})$, tak aby $KL(q(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\phi})||p(\mathbf{Z}|\mathbf{x}))$ byla minimální. Úlohu minimalizace Kullback-Lieblerovy divergence lze převést na maximalizaci dolní meze $\mathcal{L}(q(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\phi}))$. Pro maximalizaci dolní meze, hledáme maximum funkce $\mathcal{L}(q(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\phi}))$. Maximalizaci provedeme ve dvou krocích. Nejdříve zafixujeme parametr $\boldsymbol{\lambda}$ a určíme maximum funkce $\mathcal{L}(q(\mathbf{z}|\boldsymbol{\phi}))$. V druhém kroku poté zafixujeme parametr $\boldsymbol{\phi}$ a určíme maximum funkce $\mathcal{L}(q(\boldsymbol{\Theta}|\boldsymbol{\lambda}))$.

Platí

$$\mathcal{L}(q(\boldsymbol{\Theta}|\boldsymbol{\lambda})) = E_q[\ln p(\mathbf{x}, \boldsymbol{\Theta})] - E_q[\ln q(\boldsymbol{\Theta}|\boldsymbol{\lambda})] + k_1, \quad (4.20)$$

$$\mathcal{L}(q(\mathbf{z}|\boldsymbol{\phi})) = E_q[\ln p(\mathbf{x}, \mathbf{z})] - E_q[\ln q(\mathbf{z}|\boldsymbol{\phi})] + k_2 \quad (4.21)$$

kde k_1 resp. k_2 je konstanta obsahující členy, které nezávisí na parametru $\boldsymbol{\lambda}$ resp. na parametru $\boldsymbol{\phi}$.

Vyjádříme-li hustoty pravděpodobností ve tvaru hustot exponenciálního typu a využijeme vztahu (4.11), po úpravách dostaneme

$$\mathcal{L}(q(\boldsymbol{\Theta}|\boldsymbol{\lambda})) = E_q[\eta_p(\mathbf{x}, \mathbf{z})^T] \nabla_{\boldsymbol{\lambda}} a_p(\boldsymbol{\lambda}) - \boldsymbol{\lambda}^T \nabla_{\boldsymbol{\lambda}} a_p(\boldsymbol{\lambda}) + a_p(\boldsymbol{\lambda}) + k_1, \quad (4.22)$$

$$\mathcal{L}(q(\mathbf{z}|\boldsymbol{\phi})) = E_q[\eta_l(\mathbf{x}, \boldsymbol{\Theta})^T] \nabla_{\boldsymbol{\phi}} a_l(\boldsymbol{\phi}) - \boldsymbol{\phi}^T \nabla_{\boldsymbol{\phi}} a_l(\boldsymbol{\phi}) + a_l(\boldsymbol{\phi}) + k_2. \quad (4.23)$$

Lokální maxima funkce (4.22), (4.23) nalezneme výpočtem gradientu a jeho položením nule.

$$\nabla_{\boldsymbol{\lambda}} \mathcal{L}(q(\boldsymbol{\Theta}|\boldsymbol{\lambda})) = \nabla_{\boldsymbol{\lambda}}^2 a_p(\boldsymbol{\lambda}) (E_q[\eta_p(\mathbf{x}, \mathbf{z})] - \boldsymbol{\lambda}), \quad (4.24)$$

$$\nabla_{\boldsymbol{\lambda}} \mathcal{L}(q(\boldsymbol{\Theta}|\boldsymbol{\lambda})) = 0 \iff \boldsymbol{\lambda} = E_q[\eta_p(\mathbf{x}, \mathbf{z})] \quad (4.25)$$

$$\nabla_{\boldsymbol{\phi}} \mathcal{L}(q(\mathbf{z}|\boldsymbol{\phi})) = \nabla_{\boldsymbol{\phi}}^2 a_l(\boldsymbol{\phi}) (E_q[\eta_l(\mathbf{x}, \boldsymbol{\Theta})] - \boldsymbol{\phi}), \quad (4.26)$$

$$\nabla_{\boldsymbol{\phi}} \mathcal{L}(q(\mathbf{z}|\boldsymbol{\phi})) = 0 \iff \boldsymbol{\phi} = E_q[\eta_l(\mathbf{x}, \boldsymbol{\Theta})] \quad (4.27)$$

Parametry $\boldsymbol{\lambda}$ určíme jako střední hodnoty vzhledem k variačnímu rozdělení s hustotou $q(\boldsymbol{\Theta}|\boldsymbol{\lambda})$ přirozeného parametru posteriorního rozdělení. A parametry $\boldsymbol{\phi}$ určíme jako střední hodnoty vzhledem k variačnímu rozdělení s hustotou $q(\mathbf{z}|\boldsymbol{\phi})$ přirozeného parametru posteriorního rozdělení.

4.2.6 Laplaceova aproximace

Laplaceova aproximace je metoda pro nahrazení hustoty pravděpodobnosti s určitými vlastnostmi hustotou normálního rozdělení [6], [33], kterou využijeme níže.

Mějme funkci $g(x)$, která má spojité derivace do řádu 2, s maximem v bodě x_0 , tedy $g'(x_0) = 0$. Chceme vypočítat

$$\int_a^b g(x) dx. \quad (4.28)$$

Nechť $h(x) = \ln g(x)$, pak

$$\int_a^b g(x) dx = \int_a^b \exp(h(x)) dx. \quad (4.29)$$

Využijeme-li Taylorova rozvoje se středem v x_0 , pak

$$\int_a^b \exp(h(x)) dx \approx \int_a^b \exp \left(h(x_0) + h'(x_0)(x - x_0) + \frac{1}{2} h''(x_0)(x - x_0)^2 \right) dx. \quad (4.30)$$

Jelikož člen $h(x_0)$ nezávisí na proměnné x a $h'(x_0) = 0$, pak

$$\int_a^b \exp(h(x)) dx \approx \exp(h(x_0)) \int_a^b \exp \left(-\frac{1}{2} \frac{(x - x_0)^2}{h''(x_0)^{-1}} \right) dx. \quad (4.31)$$

4.2. VARIČNÍ BAYESOVSKÉ METODY

Dostali jsme hustotu normálního rozdělení se střední hodnotou x_0 a rozptylem $-h''(x_0)^{-1}$. V případě, že $a = -\infty, b = \infty$ aproximace (4.31) je ve tvaru

$$\int_a^b \exp(h(x)) dx \approx \exp(h(x_0)) \sqrt{\frac{2\pi}{-h''(x_0)}}. \quad (4.32)$$

4.2.7 Odhad parametrů GMM pro úlohu deinterleavingu

Standardní GMM

Mějme N přijatých impulzů $\mathbf{X}_1, \dots, \mathbf{X}_N$, kde impulz \mathbf{X}_n je popsán J parametry, tedy $\mathbf{X}_n = (X_1, \dots, X_J)$. Předpokládáme, že tyto impulzy jsou vzájemně nezávislé a generovány K emitery. Parametry impulzů všech emiterů mají J -rozměrné normální rozdělení. Máme tedy GMM

$$p(\mathbf{x}_n | \Theta) = \sum_{k=1}^K a_k \mathcal{N}_J(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad (4.33)$$

kde K je známý počet emiterů, $\boldsymbol{\mu}_k$ a $\boldsymbol{\Sigma}_k$ jsou střední hodnota a kovarianční matice k -tého emiteru, respektive k -tého J -rozměrného normálního rozdělení a a_k je parametr vah.

Naším úkolem je určit parametry GMM, tedy kovarianční matice $\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_K$, vektory středních hodnot $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K$ a vektor parametrů vah $\mathbf{a} = (a_1 \dots a_K)$ splňující dvě vlastnosti: $a_k > 0$, $\sum_{k=1}^K a_k = 1$. Chceme určit parametry $\Theta = \{\mathbf{a}, \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_K\}$.

GMM může obsahovat latentní proměnnou $\mathbf{z} \in \mathbb{R}^N$, která popisuje k jakému komponentu patří proměnná \mathbf{x}_n . Tedy $z_n \in \{1, \dots, K\}$. Pak je sdružená hustota pravděpodobnosti ve tvaru

$$p(\mathbf{x}_n, z_n | \Theta) = \sum_{z_n \in \{1, \dots, K\}} p(\mathbf{x}_n | z_n, \Theta) p(z_n | \Theta), \quad (4.34)$$

$$p(\mathbf{x}_n | z_n, \Theta) = \prod_{k=1}^K \mathcal{N}_J(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{\delta_{z_n}^k}, \quad (4.35)$$

$$p(z_n | \Theta) = \prod_{k=1}^K a_k^{\delta_{z_n}^k}. \quad (4.36)$$

Sdružená pravděpodobnost je pak ve tvaru

$$p(\mathbf{x}_1, \dots, \mathbf{x}_N, \mathbf{z} | \Theta) = \prod_{n=1}^N \prod_{k=1}^K [a_k \mathcal{N}_J(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)]^{\delta_{z_n}^k}. \quad (4.37)$$

Odlehlé hodnoty

Hlavním nedostatkem standardního GMM (4.37), jenž je uváděn v dostupné literatuře, je citlivost na odlehlé hodnoty, díky nimž dochází k špatnému odhadu parametrů modelu. Tento nedostatek je možné vyřešit zavedením nové latentní proměnné, která mění kovarianční matici Σ_k každého komponentu modelu pro každé \mathbf{x}_n [34].

Odlehlé hodnoty definujeme pomocí latentní proměnné $\mathbf{u} = (u_1, \dots, u_N)$. Hustoty pravděpodobnosti $p(\mathbf{x}_n|u_n, z_n, \Theta)$ a $p(\mathbf{x}_n, u_n|z_n, \Theta)$ určíme následovně

$$p(\mathbf{x}_n|u_n, z_n, \Theta) = \prod_{k=1}^K \mathcal{N}_J(\mathbf{x}_n|\boldsymbol{\mu}_k, u_n^{-1}\Sigma_k)^{\delta_{z_n}^k}, \quad (4.38)$$

$$p(\mathbf{x}_n, u_n|z_n, \Theta) = \prod_{k=1}^K \mathcal{N}_J[(\mathbf{x}_n|\boldsymbol{\mu}_k, u_n^{-1}\Sigma_k)p_k(u_n)]^{\delta_{z_n}^k}, \quad (4.39)$$

kde $\Theta = \{\alpha_k, \beta_k, \boldsymbol{\mu}_k, \Sigma_k\}_{k=1, \dots, K}$ a $p_k(u_n)$ je hustota apriorního rozdělení latentní proměnné u_n v závislosti na $z_n = k$, která má gamma rozdělení s parametry α_k, β_k . Tedy

$$p_k(u_n) = p(u_n|z = k) = \mathcal{G}(u_n|\alpha_k, \beta_k). \quad (4.40)$$

Pro účely deinterleavingu chápeme odlehlé hodnoty jako odlehlé impulzy způsobené chybou měření, nebo jako impulzy patřící k jednomu emiteru s agilitou v parametrech.

Autor [34] dále rozšiřuje model o chybějící hodnoty. Tato úprava je použita z důvodu využívání parametru PRI pro úlohu deinterleavingu. Jak bylo zmíněno v kapitole 1, PRI parametr není kvůli citlivosti na rušení (výpadky a falešné pulzy) vhodný pro úlohu deinterleavingu. V této práci je použit nový model modifikovaný pouze o odlehlé hodnoty a je tedy nutné vlastní odvození vztahů pro odhad parametrů modelu.

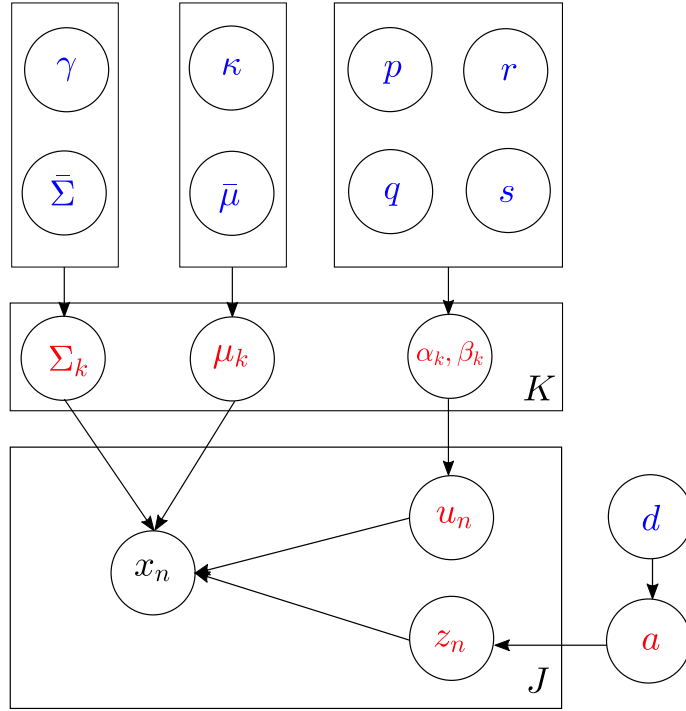
Výsledný GMM model

Kombinací (4.37), (4.39) dostaneme výslednou sdruženou pravděpodobnost

$$p(\mathbf{x}_1, \dots, \mathbf{x}_N, \mathbf{H}|\Theta) = \prod_{n=1}^N \prod_{k=1}^K [a_k \mathcal{N}_J(\mathbf{x}_n|\boldsymbol{\mu}_k, u_n^{-1}\Sigma_k) \mathcal{G}(u_n|\alpha_k, \beta_k)]^{\delta_{z_n}^k}, \quad (4.41)$$

kde $\mathbf{H} = (\mathbf{u}, \mathbf{z})$ jsou latentní proměnné a $\Theta = \{\mathbf{a}, \alpha_k, \beta_k, \boldsymbol{\mu}_k, \Sigma_k\}_{k=1, \dots, K}$ jsou neznáme parametry. Vztahy pro výpočet parametrů výsledného GMM modelu nejsou známy a je tedy nutné jejich odvození, viz níže. Grafická reprezentace výsledného modelu je znázorněna na obrázku 4.7.

4.2. VARIČNÍ BAYESOVSKÉ METODY



Obrázek 4.7: Grafická reprezentace výsledného modelu

4.2.8 Konjugované rodiny apriorních rozděléní

V Bayesovské statistice říkáme, že apriorní a aposteriorní rozděléní jsou konjugované, jestliže patří do stejné rodiny rozděléní. V našem případě předpokládáme, že apriorní i aposteriorní rozděléní patří do exponenciální rodiny a tedy konjugované jsou. Konjugované apriorní rozděléní latentních proměnných \mathbf{H} a neznámých parametrů Θ jsou známy dle [12] a jsou uvedeny v tabulce 4.1 a jejich pravděpodobnostní hustoty v tabulce 4.2.

Rozděléní s $p(\mathbf{x}_1, \dots, \mathbf{x}_N \Theta)$	Parametry rozděléní	Konjugované apriorní rozděléní	Parametry konjugovaného apriorního rozděléní
$\mathcal{N}_J(\boldsymbol{\mu}_k, u_n^{-1} \boldsymbol{\Sigma}_k)$	$\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k$	$\mathcal{N}_J \mathcal{W}^{-1}(\bar{\boldsymbol{\mu}}_k, \kappa^{-1} \boldsymbol{\Sigma}_k, \gamma, \bar{\boldsymbol{\Sigma}})^1$	$\bar{\boldsymbol{\mu}}_k, \kappa^{-1} \boldsymbol{\Sigma}_k, \gamma, \bar{\boldsymbol{\Sigma}}$
$\mathcal{CAT}(\mathbf{a})$	$\mathbf{a} = (a_1, \dots, a_K)$	$\mathcal{D}(d_1, \dots, d_K)$	d_1, \dots, d_K
$\mathcal{G}(\alpha_k, \beta_k)$	α_k, β_k	$\mathcal{R}(p, q, r, s)^2$	p, q, r, s

Tabulka 4.1: Konjugované apriorní rozděléní

¹Hustota pravděpodobnosti J –rozměrného normálního-inverzního Wishartova rozděléní $\mathcal{N}_J \mathcal{W}^{-1}$ je součin hustot vícerozměrného normálního rozděléní a inverzního Wishartova rozděléní.

²Hustota rozděléní $\mathcal{R}(p, q, r, s)$ je specifikována níže.

Rozdělení	Hustota rozdělení
$\mathbf{x}_n \sim \mathcal{N}_J(\boldsymbol{\mu}_k, u_n^{-1} \boldsymbol{\Sigma}_k)$	$p(\mathbf{x}_n \boldsymbol{\mu}_k, u_n^{-1} \boldsymbol{\Sigma}_k) = \frac{1}{\sqrt{(2\pi)^J \boldsymbol{\Sigma}_k }} \exp \left(-\frac{(\mathbf{x}_n - \boldsymbol{\mu}_k)^T u_n \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_k)}{2} \right)$
$z_n \sim \mathcal{CAT}(\mathbf{a})$	$p(z_n \mathbf{a}) = \prod_{k=1}^K a_k^{\delta_{z_n}^k}$
$\mathbf{a} \sim \mathcal{D}(d_1, \dots, d_K)$	$p(\mathbf{a} d_1, \dots, d_K) = \frac{\prod_{k=1}^K a_k^{d_k-1}}{\mathcal{B}(d_1, \dots, d_K)}$
$\boldsymbol{\mu}_k \sim \mathcal{N}_J(\bar{\boldsymbol{\mu}}_k, \kappa^{-1} \boldsymbol{\Sigma}_k)$	$p(\boldsymbol{\mu}_k \bar{\boldsymbol{\mu}}_k, \kappa^{-1} \boldsymbol{\Sigma}_k) = \frac{1}{\sqrt{(2\pi)^J \boldsymbol{\Sigma}_k }} \exp \left(-\frac{(\boldsymbol{\mu}_k - \bar{\boldsymbol{\mu}}_k)^T \kappa \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{\mu}_k - \bar{\boldsymbol{\mu}}_k)}{2} \right)$
$\boldsymbol{\Sigma}_k \sim \mathcal{W}^{-1}(\gamma, \bar{\boldsymbol{\Sigma}})$	$p(\boldsymbol{\Sigma}_k \gamma, \bar{\boldsymbol{\Sigma}}) = \frac{ \bar{\boldsymbol{\Sigma}} ^{\frac{\gamma}{2}}}{2^{\frac{\gamma J}{2}} \Gamma_J(\frac{\gamma}{2})} \boldsymbol{\Sigma}_k ^{-\frac{\gamma+J+1}{2}} \exp \left(-\frac{\text{tr}(\bar{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}_k^{-1})}{2} \right)$
$u_n \sim \mathcal{G}(\alpha_k, \beta_k)$	$p(u_n \alpha_k, \beta_k) = \frac{\beta_k^{\alpha_k} u_n^{\alpha_k-1}}{\Gamma(\alpha_k)} \exp(-\beta_k u_n)$
$\alpha_k, \beta_k \sim \mathcal{R}(p, q, r, s)$	$p(\alpha_k, \beta_k p, q, r, s) = \frac{p^{\alpha_k-1}}{\Gamma(\alpha_k)^r \beta_k^{-\alpha_k s}} \exp(-\beta_k q) _{\alpha_k, \beta_k > 0}$

Tabulka 4.2: Hustoty vybraných rozdělení

Konjugovanou apriorní hustotu pro α_k, β_k lze zapsat jako

$$p(\alpha_k, \beta_k | p, q, r, s) = p(\alpha_k | p, q, r, s) p(\beta_k | \alpha_k, q, s),$$

kde

$$\begin{aligned} p(\beta_k | \alpha_k, q, s) &= \mathcal{G}(\beta_k | s\alpha_k + 1, q), \\ p(\alpha_k | p, q, r, s) &= \frac{1}{M_0} \frac{p^{\alpha_k-1} \Gamma(s\alpha_k + 1)}{q^{s\alpha_k+1} \Gamma(\alpha_k)^r}, \\ M_0 &= \int \frac{p^{\alpha_k-1} \Gamma(s\alpha_k + 1)}{q^{s\alpha_k+1} \Gamma(\alpha_k)^r} d\alpha_k. \end{aligned}$$

Normalizační konstanta M_0 se určí pomocí metody založené na Laplaceově aproximaci dle [39] viz podkapitola 4.2.6.

4.2.9 Aposteriorní rozdělení parametrů

Dále indexem N_k značíme nový parametr k -té komponenty modelu (4.41). Pro aposteriorní rozdělení neznámých parametrů $\boldsymbol{\Theta}$ platí

$$\begin{aligned} p(\boldsymbol{\Theta} | \mathbf{X}, \mathbf{H}) &= p(\mathbf{a}) \prod_{n=1}^N p(z_n | \mathbf{a}) \prod_{k=1}^K \left[p(\alpha_k, \beta_k) p(\boldsymbol{\mu}_k) p(\boldsymbol{\Sigma}_k) \prod_{n=1}^N p(u_n | \alpha_k, \beta_k) p(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right], \\ p(\mathbf{a}) \prod_{n=1}^N p(z_n | \mathbf{a}) &= \frac{\prod_{k=1}^K a_k^{d_k-1}}{\mathcal{B}(d_1, \dots, d_K)} \prod_{k=1}^K a_k^{\sum_{n=1}^N \delta_{z_n}^k} \propto \prod_{k=1}^K a_k^{d_k + \sum_{n=1}^N \delta_{z_n}^k - 1}, \end{aligned}$$

tj.

$$\mathbf{a} | \mathbf{X}, \mathbf{H} \sim \mathcal{D}(d_{N_1}, \dots, d_{N_K}),$$

kde

$$d_{N_k} = d_k + \sum_{n=1}^N \delta_{z_n}^k.$$

4.2. VARIČNÍ BAYESOVSKÉ METODY

$$\begin{aligned}
& \prod_{k=1}^K \left[p(\alpha_k, \beta_k) \prod_{n=1}^N p(u_n | \alpha_k, \beta_k) \right] = \\
& = \prod_{k=1}^K \frac{p^{\alpha_k-1}}{\Gamma(\alpha_k)^r \beta_k^{-\alpha_k s}} \exp(-\beta_k q) \frac{\beta_k^{\sum_{n=1}^N \delta_{z_n}^k \alpha_k} u_n^{\sum_{n=1}^N \delta_{z_n}^k (\alpha_k-1)}}{\Gamma(\alpha_k)^{\sum_{n=1}^N \delta_{z_n}^k}} \exp \left(- \sum_{n=1}^N \delta_{z_n}^k \beta_k u_n \right) \propto \\
& \propto \frac{\left(p u_n^{\sum_{n=1}^N \delta_{z_n}^k} \right)^{(\alpha_k-1)}}{\Gamma(\alpha_k)^{r+\sum_{n=1}^N \delta_{z_n}^k} \beta_k^{-\alpha_k (s+\sum_{n=1}^N \delta_{z_n}^k)}} \exp \left(-\beta_k \left(q + \sum_{n=1}^N \delta_{z_n}^k u_n \right) \right),
\end{aligned}$$

tj.

$$\alpha_k, \beta_k | \mathbf{X}, \mathbf{H} \sim \mathcal{R}(p_{N_k}, q_{N_k}, r_{N_k}, s_{N_k}),$$

kde

$$\begin{aligned}
p_{N_k} &= p u_n^{\sum_{n=1}^N \delta_{z_n}^k}, \\
q_{N_k} &= q + \sum_{n=1}^N \delta_{z_n}^k u_n, \\
r_{N_k} &= r + \sum_{n=1}^N \delta_{z_n}^k, \\
s_{N_k} &= s + \sum_{n=1}^N \delta_{z_n}^k.
\end{aligned}$$

$$\begin{aligned}
& \prod_{k=1}^K \left[p(\boldsymbol{\mu}_k) p(\boldsymbol{\Sigma}_k) \prod_{n=1}^N p(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right] = \\
& = \prod_{k=1}^K \frac{1}{\sqrt{(2\pi)^J |\boldsymbol{\kappa}^{-1} \boldsymbol{\Sigma}_k|}} \exp \left(- \frac{(\boldsymbol{\mu}_k - \bar{\boldsymbol{\mu}}_k)^T \boldsymbol{\kappa} \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{\mu}_k - \bar{\boldsymbol{\mu}}_k)}{2} \right) \times \\
& \times \frac{|\bar{\boldsymbol{\Sigma}}|^{\frac{\gamma}{2}}}{2^{\frac{\gamma J}{2}} \Gamma_J(\frac{\gamma}{2})} |\boldsymbol{\Sigma}_k|^{-\frac{\gamma+J+1}{2}} \exp \left(- \frac{\text{tr}(\bar{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}_k^{-1})}{2} \right) \times \\
& \times \left(\frac{1}{\sqrt{(2\pi)^J |u_n^{-1} \boldsymbol{\Sigma}_k|}} \right)^{\sum_{n=1}^N \delta_{z_n}^k} \exp \left(- \sum_{n=1}^N \delta_{z_n}^k \frac{(\mathbf{x}_n - \boldsymbol{\mu}_k)^T u_n \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_k)}{2} \right) \propto \\
& \propto |\boldsymbol{\Sigma}_k|^{-1/2} |\boldsymbol{\Sigma}_k|^{-\frac{\gamma+\sum_{n=1}^N \delta_{z_n}^k + J+1}{2}} \exp \left(-\frac{1}{2} ((\boldsymbol{\mu}_k - \bar{\boldsymbol{\mu}}_k)^T \boldsymbol{\kappa} \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{\mu}_k - \bar{\boldsymbol{\mu}}_k) + \right. \\
& \left. + \sum_{n=1}^N \delta_{z_n}^k u_n (\mathbf{x}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_k) + \text{tr}(\bar{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}_k^{-1})) \right).
\end{aligned}$$

Pro úpravu této hustoty aposteriorního rozdělení využijeme následující vztahy [12]:

- Suma rozdílů od střední hodnoty:

$$\sum_{n=1}^N (\mathbf{x}_n - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_n - \boldsymbol{\mu}) = N(\boldsymbol{\mu} - \bar{\mathbf{x}})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \bar{\mathbf{x}}) + \sum_{n=1}^N (\mathbf{x}_n - \bar{\mathbf{x}})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_n - \bar{\mathbf{x}}),$$

$$\text{kde } \bar{\mathbf{x}} = \sum_{n=1}^N \mathbf{x}_n / N,$$

- Součet dvou kvadratických forem:

$$\begin{aligned} (\mathbf{y} - \mathbf{x})^T \mathbf{A}(\mathbf{y} - \mathbf{x}) + (\mathbf{x} - \mathbf{z})^T \mathbf{B}(\mathbf{x} - \mathbf{z}) &= \\ &= (\mathbf{x} - \mathbf{c})^T (\mathbf{A} + \mathbf{B})(\mathbf{x} - \mathbf{c}) + (\mathbf{y} - \mathbf{z})^T (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} (\mathbf{y} - \mathbf{z}), \end{aligned}$$

$$\text{kde } \mathbf{c} = (\mathbf{A} + \mathbf{B})^{-1} (\mathbf{A}\mathbf{y} + \mathbf{B}\mathbf{z}),$$

- Stopa skaláru a cyklická vlastnost stopy:

$$(\mathbf{x} - \mathbf{y})^T \mathbf{A}(\mathbf{x} - \mathbf{y}) = \text{tr} [((\mathbf{x} - \mathbf{y})(\mathbf{x} - \mathbf{y})^T) \mathbf{A}].$$

S využitím vztahů výše, lze exponent $E = -\frac{1}{2}((\boldsymbol{\mu}_k - \bar{\boldsymbol{\mu}}_k)^T \kappa \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{\mu}_k - \bar{\boldsymbol{\mu}}_k) + \sum_{n=1}^N \delta_{z_n}^k u_n (\mathbf{x}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_k) + \text{tr}(\bar{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}_k^{-1}))$ přepsat do tvaru

$$E = -\frac{1}{2} ((\boldsymbol{\mu}_k - \boldsymbol{\mu}_{N_k})^T \boldsymbol{\Sigma}_{N_k}^{-1} (\boldsymbol{\mu}_k - \boldsymbol{\mu}_{N_k}) + \text{tr}(\bar{\boldsymbol{\Sigma}}_{N_k} \boldsymbol{\Sigma}_k^{-1})),$$

tj.

$$\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k | \mathbf{X}, \mathbf{H} \sim \mathcal{N}(\boldsymbol{\mu}_{N_k}, \boldsymbol{\Sigma}_{N_k}) \otimes \mathcal{W}^{-1}(\gamma_{N_k}, \bar{\boldsymbol{\Sigma}}_{N_k}),$$

kde

$$\boldsymbol{\Sigma}_{N_k} = \boldsymbol{\Sigma}_k \left(\kappa + \sum_{n=1}^N \delta_{z_n}^k u_n \right)^{-1} \Rightarrow \kappa_{N_k} = \kappa + \sum_{n=1}^N \delta_{z_n}^k u_n,$$

$$\boldsymbol{\mu}_{N_k} = \boldsymbol{\Sigma}_{N_k} \left(\boldsymbol{\Sigma}_k^{-1} \left(\bar{\boldsymbol{\mu}}_k \kappa + \sum_{n=1}^N \delta_{z_n}^k u_n \mathbf{x}_n \right) \right) = \frac{\bar{\boldsymbol{\mu}}_k \kappa + \sum_{n=1}^N \delta_{z_n}^k u_n \mathbf{x}_n}{\kappa + \sum_{n=1}^N \delta_{z_n}^k u_n},$$

$$\gamma_{N_k} = \gamma + \sum_{n=1}^N \delta_{z_n}^k,$$

$$\bar{\boldsymbol{\Sigma}}_{N_k} = \bar{\boldsymbol{\Sigma}} + \sum_{n=1}^N \delta_{z_n}^k u_n (\mathbf{x}_n - \bar{\mathbf{x}}_k)(\mathbf{x}_n - \bar{\mathbf{x}}_k)^T + \frac{\kappa \sum_{n=1}^N \delta_{z_n}^k u_n}{\kappa + \sum_{n=1}^N \delta_{z_n}^k u_n} (\bar{\boldsymbol{\mu}}_k - \bar{\mathbf{x}}_k)(\bar{\boldsymbol{\mu}}_k - \bar{\mathbf{x}}_k)^T,$$

$$\bar{\mathbf{x}}_k = \frac{\sum_{n=1}^N \delta_{z_n}^k \mathbf{x}_n}{\sum_{n=1}^N \delta_{z_n}^k}.$$

4.2. VARIČNÍ BAYESOVSKÉ METODY

Tedy parametr Θ má aposteriorní rozdělení

$$\Theta \sim \mathcal{D}(d_{N_1}, \dots, d_{N_K}) \otimes \mathcal{R}(p_{N_k}, q_{N_k}, r_{N_k}, s_{N_k}) \otimes \mathcal{N}(\mu_{N_k}, \Sigma_{N_k}) \otimes \mathcal{W}^{-1}(\gamma_{N_k}, \bar{\Sigma}_{N_k})$$

a patří do exponenciální rodiny. Pak funkce η_p v hustotě exponenciálního typu (4.6) je ve tvaru

$$\begin{aligned} \eta_p = & (d_{N_1} + 1, \dots, d_{N_K} + 1, \ln p_{N_1}, \dots, \ln p_{N_K}, -q_{N_1}, \dots, -q_{N_K}, -r_{N_1}, \dots, -r_{N_K}, \\ & s_{N_1}, \dots, s_{N_K}, -\frac{\Sigma_{N_1}^{-1}}{2}, \dots, -\frac{\Sigma_{N_K}^{-1}}{2}, -\frac{\Sigma_{N_1}^{-1}\mu_{N_1}}{2}, \dots, -\frac{\Sigma_{N_K}^{-1}\mu_{N_K}}{2}, \\ & -\frac{\gamma_{N_1} + J + 1}{2}, \dots, -\frac{\gamma_{N_K} + J + 1}{2}, \bar{\Sigma}_{N_1}, \dots, \bar{\Sigma}_{N_K}). \end{aligned}$$

4.2.10 Aposteriorní rozdělení latentních proměnných

Dále indexem N_{nk} značíme nový parametr k -té komponenty modelu (4.41) pro n -té pozorování. Pro aposteriorní rozdělení latentních proměnných \mathbf{H} platí

$$p(\mathbf{H}|\mathbf{X}, \Theta) = \prod_{n=1}^N p(z_n|\mathbf{a}) \prod_{k=1}^K p(u_n|\alpha_k, \beta_k) p(\mathbf{x}_n|u_n, \Theta).$$

$$\begin{aligned} & \prod_{n=1}^N \prod_{k=1}^K p(u_n|\alpha_k, \beta_k) p(\mathbf{x}_n|u_n, \Theta) = \\ & = \prod_{k=1}^K \frac{\beta_k^{\sum_{n=1}^N \delta_{z_n}^k \alpha_k} u_n^{\sum_{n=1}^N \delta_{z_n}^k (\alpha_k - 1)}}{\Gamma(\alpha_k)^{\sum_{n=1}^N \delta_{z_n}^k}} \exp\left(-\sum_{n=1}^N \delta_{z_n}^k \beta_k u_n\right) \times \\ & \times \left(\frac{1}{\sqrt{(2\pi)^J |\Sigma_k^{-1}|}}\right)^{\sum_{n=1}^N \delta_{z_n}^k} \exp\left(-\sum_{n=1}^N \delta_{z_n}^k u_n \frac{(\mathbf{x}_n - \mu_k)^T \Sigma_k^{-1} (\mathbf{x}_n - \mu_k)}{2}\right) \propto \\ & \propto u_n^{\sum_{n=1}^N \delta_{z_n}^k (\alpha_k + \frac{J}{2} - 1)} \exp\left(-\sum_{n=1}^N \delta_{z_n}^k u_n \left(\beta_k + \frac{(\mathbf{x}_n - \mu_k)^T \Sigma_k^{-1} (\mathbf{x}_n - \mu_k)}{2}\right)\right), \end{aligned}$$

tj.

$$u_n \sim \Gamma(\alpha_{N_{nk}}, \beta_{N_{nk}}),$$

kde

$$\begin{aligned} \alpha_{N_{nk}} &= \alpha_k + \frac{J}{2}, \\ \beta_{N_{nk}} &= \beta_k + \frac{(\mathbf{x}_n - \mu_k)^T \Sigma_k^{-1} (\mathbf{x}_n - \mu_k)}{2}. \end{aligned}$$

Tedy

$$p(\mathbf{H}|\mathbf{X}, \Theta) = \prod_{n=1}^N \prod_{k=1}^K \left(\frac{a_k}{\sqrt{(2\pi)^J |\Sigma_k|}} \frac{\beta_k^{\alpha_k} \Gamma(\alpha_{N_{nk}})}{\Gamma(\alpha_k) \beta_{N_{nk}}^{\alpha_{N_{nk}}}} \frac{\beta_{N_{nk}}^{\alpha_{N_{nk}}} u_n^{\alpha_{N_{nk}} - 1}}{\Gamma(\alpha_{N_{nk}})} \exp(-\beta_{N_{nk}} u_n) \right)^{\delta_{z_n}^k}.$$

Označíme-li

$$\rho_{nk} = \frac{a_k}{\sqrt{(2\pi)^J |\Sigma_k|}} \frac{\beta_k^{\alpha_k} \Gamma(\alpha_{N_{nk}})}{\Gamma(\alpha_k) \beta_{N_{nk}}^{\alpha_{N_{nk}}}},$$

pak normovaná hustota pravděpodobnosti $p(z_n = k | \mathbf{x}_n, u_n, \Theta)$ je rovna

$$p(z_n = k | \mathbf{x}_n, u_n, \Theta) = \frac{\rho_{nk}}{\sum_{m=1}^K \rho_{nm}} = \theta_{nk}.$$

Tedy latentní proměnné \mathbf{H} mají aposteriorní rozdělení

$$\mathbf{H} \sim \mathcal{CAT}(\theta_n) \otimes \mathcal{G}(\alpha_{N_{nk}}, \beta_{N_{nk}})$$

a patří do exponenciální rodiny. Pak funkce η_l v hustotě exponenciálního typu (4.6) je ve tvaru

$$\eta_l = (\alpha_{N_{n1}} - 1, \dots, \alpha_{N_{nK}} - 1, -\beta_{N_{n1}}, \dots, -\beta_{N_{nK}}, \ln \rho_{n1}, \dots, \ln \rho_{nK}).$$

4.2.11 Variační rozdělení

Hustotu variačního rozdělení $q(\Theta | \lambda)$ pro vektor neznámých parametrů Θ volíme ve stejném tvaru jako hustotu aposteriorního rozdělení $p(\Theta | \mathbf{x}_1, \dots, \mathbf{x}_N, \mathbf{H})$. Tedy

$$q(\Theta | \lambda) = h_p(\Theta) \exp(\lambda^T T_p(\Theta) - a_p(\lambda)),$$

kde λ^T je přirozený parametr ve vztahu (4.18) roven

$$\begin{aligned} \lambda^T = & (d_{N_1}^* + 1, \dots, d_{N_K}^* + 1, \ln p_{N_1}^*, \dots, \ln p_{N_K}^*, -q_{N_1}^*, \dots, -q_{N_K}^*, -r_{N_1}^*, \dots, -r_{N_K}^*, s_{N_1}^*, \\ & \dots, s_{N_K}^*, -\frac{\Sigma_{N_1}^{*-1}}{2}, \dots, -\frac{\Sigma_{N_K}^{*-1}}{2}, -\frac{\Sigma_{N_1}^{*-1} \mu_{N_1}^*}{2}, \dots, -\frac{\Sigma_{N_K}^{*-1} \mu_{N_K}^*}{2}, \\ & -\frac{\gamma_{N_1}^* + J + 1}{2}, \dots, -\frac{\gamma_{N_K}^* + J + 1}{2}, \bar{\Sigma}_{N_1}^*, \dots, \bar{\Sigma}_{N_K}^*). \end{aligned}$$

Hustotu variačního rozdělení $q(\mathbf{H} | \phi_n)$ pro vektor latentních proměnných \mathbf{H} volíme ve stejném tvaru jako hustotu aposteriorního rozdělení $p(\mathbf{H} | \mathbf{X}, \Theta)$. Tedy

$$q(\mathbf{H} | \phi_n) = h_l(\mathbf{H}) \exp(\phi_n^T T_l(\mathbf{H}) - a_l(\phi_n)),$$

kde ϕ_n^T je přirozený parametr ve vztahu (4.19) roven

$$\phi_n^T = (\alpha_{N_{n1}}^* - 1, \dots, \alpha_{N_{nK}}^* - 1, -\beta_{N_{n1}}^*, \dots, -\beta_{N_{nK}}^*, \ln \rho_{n1}^*, \dots, \ln \rho_{nK}^*).$$

Dle (4.25) a (4.27) vypočítáme přirozené parametry λ, ϕ_n jako střední hodnoty funkcí η_p, η_l hustot aposteriorních rozdělení $p(\Theta | \mathbf{X}, \mathbf{H}), p(\mathbf{H} | \mathbf{X}, \Theta)$:

$$\begin{aligned} \lambda &= E_q[\eta_p], \\ \phi_n &= E_q[\eta_l]. \end{aligned}$$

Výpočet přirozených parametrů provádíme iteračně. Hodnotu $\phi_n^{[i+1]}$ v $i + 1$ iteraci získáme pomocí předchozí iterace $\lambda^{[i]}$. Hodnotu $\lambda^{[i+1]}$ získáme z $\phi_n^{[i+1]}$. Postupujeme podle algoritmu 1.

4.2. VARIČNÍ BAYESOVSKÉ METODY

Data: Sekvence impulzů popsána PDW maticí \mathbf{X}

Input: Inicializace $\boldsymbol{\lambda}^{[1]}$

while *dokud není splněna zastavovací podmínka* **do**

for $n=1:N$ **do**
 $\phi_n^{[i]} = E_{q^{[i-1]}}[\eta_l]$
 end
 $\lambda^{[i]} = E_{q^{[i]}}[\eta_p]$

end

Algoritmus 1: Algoritmus pro update přirozených parametrů $\boldsymbol{\lambda}, \phi_n$

Vypočteme požadované střední hodnoty:

$$\begin{aligned}
 E_q[d_{N_k} + 1] &= E_q \left[d_k + \sum_{n=1}^N \delta_{z_n}^k + 1 \right] = d_k + \sum_{n=1}^N E_q [\delta_{z_n}^k] + 1, \\
 E_q[\ln(p_{N_k})] &= E_q \left[\ln(p u_n^{\sum_{n=1}^N \delta_{z_n}^k}) \right] = \ln(p) + \sum_{n=1}^N E_q [\delta_{z_n}^k] E_q [\ln u_n], \\
 E_q[-q_{N_k}] &= E_q \left[-q - \sum_{n=1}^N \delta_{z_n}^k u_n \right] = -q - \sum_{n=1}^N E_q [\delta_{z_n}^k] E_q [u_n], \\
 E_q[-r_{N_k}] &= E_q \left[-r - \sum_{n=1}^N \delta_{z_n}^k \right] = -r - \sum_{n=1}^N E_q [\delta_{z_n}^k], \\
 E_q[s_{N_k}] &= E_q \left[s + \sum_{n=1}^N \delta_{z_n}^k \right] = s + \sum_{n=1}^N E_q [\delta_{z_n}^k], \\
 E_q \left[-\frac{\Sigma_{N_k}^{-1}}{2} \right] &= E_q \left[-\frac{1}{2} \Sigma_k^{-1} \left(\kappa + \sum_{n=1}^N \delta_{z_n}^k u_n \right) \right] = -\frac{1}{2} \Sigma_k^{-1} \left(\kappa + \sum_{n=1}^N E_q [\delta_{z_n}^k] E_q [u_n] \right), \\
 E_q \left[-\frac{\Sigma_{N_k}^{-1} \boldsymbol{\mu}_{N_k}}{2} \right] &= E_q \left[-\frac{1}{2} \Sigma_k^{-1} \left(\bar{\boldsymbol{\mu}}_k \kappa + \sum_{n=1}^N \delta_{z_n}^k u_n \mathbf{x}_n \right) \right] = \\
 &= -\frac{1}{2} \Sigma_k^{-1} \left(\bar{\boldsymbol{\mu}}_k \kappa + \sum_{n=1}^N E_q [\delta_{z_n}^k] E_q [u_n] \mathbf{x}_n \right), \\
 E_q \left[-\frac{\gamma_{N_k} + J + 1}{2} \right] &= E_q \left[-\frac{\gamma + \sum_{n=1}^N \delta_{z_n}^k + J + 1}{2} \right] = -\frac{\gamma + \sum_{n=1}^N E_q [\delta_{z_n}^k] + J + 1}{2}, \\
 E_q [\bar{\Sigma}_{N_k}] &= E_q \left[\bar{\Sigma} + \sum_{n=1}^N \delta_{z_n}^k u_n (\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T + \frac{\kappa \sum_{n=1}^N \delta_{z_n}^k u_n}{\kappa + \sum_{n=1}^N \delta_{z_n}^k u_n} (\bar{\boldsymbol{\mu}}_k - \bar{\mathbf{x}})(\bar{\boldsymbol{\mu}}_k - \bar{\mathbf{x}})^T \right] = \\
 &= \bar{\Sigma} + \sum_{n=1}^N E_q [\delta_{z_n}^k] E_q [u_n] (\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T + \frac{\kappa \sum_{n=1}^N E_q [\delta_{z_n}^k] E_q [u_n]}{\kappa + \sum_{n=1}^N E_q [\delta_{z_n}^k] E_q [u_n]} (\bar{\boldsymbol{\mu}}_k - \bar{\mathbf{x}})(\bar{\boldsymbol{\mu}}_k - \bar{\mathbf{x}})^T, \\
 E_q[\alpha_{N_k} - 1] &= E_q[\alpha_k] + \frac{J}{2} - 1, \\
 E_q[-\beta_{N_k}] &= -E_q[\beta_k] - E_q \left[\frac{(\mathbf{x}_n - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_k)}{2} \right]
 \end{aligned}$$

$$\begin{aligned}
E_q[\ln(\rho_{nk})] &= E_q[\ln(a_k) + \alpha_{N_{nk}} \ln(\beta_{N_{nk}}) - \ln(\Gamma(\alpha_{N_{nk}})) + \alpha_k \ln(\beta_k) - \ln(\Gamma(\alpha_k)) - \\
&\quad - \frac{J}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_k|] = E_q[\ln a_k] + \alpha_{N_{nk}}^* \ln(\beta_{N_{nk}}^*) - \ln(\Gamma(\alpha_{N_{nk}}^*)) + E_q[\alpha_k] E_q[\ln \beta_k] \\
&\quad - E_q[\ln(\Gamma(\alpha_k))] - \frac{J}{2} \ln(2\pi) - \frac{1}{2} E_q[\ln |\Sigma_k|].
\end{aligned}$$

Nakonec nahradíme přirozené parametry λ běžnými parametry $(d_{Nk}^*, p_{Nk}^*, q_{Nk}^*, r_{Nk}^*, s_{Nk}^*, \Sigma_{Nk}^*, \mu_{Nk}^*, \gamma_{Nk}^*, \bar{\Sigma}_{Nk}^*)$ a přirozené parametry ϕ_n běžnými parametry $(\alpha_{N_{nk}}^*, \beta_{N_{nk}}^*, \rho_{nk}^*)$:

$$\begin{aligned}
d_{Nk}^* &= d_k + \sum_{n=1}^N E_q[\delta_{z_n}^k] \\
p_{Nk}^* &= p \exp \left(\sum_{n=1}^N E_q[\delta_{z_n}^k] E_q[\ln u_n] \right) \\
q_{Nk}^* &= q + \sum_{n=1}^N E_q[\delta_{z_n}^k] E_q[u_n] \\
r_{Nk}^* &= r + \sum_{n=1}^N E_q[\delta_{z_n}^k] \\
s_{Nk}^* &= s + \sum_{n=1}^N E_q[\delta_{z_n}^k] \\
\Sigma_{Nk}^* &= \Sigma_k \left(\kappa + \sum_{n=1}^N E_q[\delta_{z_n}^k] E_q[u_n] \right)^{-1} \Rightarrow \kappa_{Nk}^* = \kappa + \sum_{n=1}^N E_q[\delta_{z_n}^k] E_q[u_n] \\
\mu_{Nk}^* &= \frac{\bar{\mu}_k \kappa + \sum_{n=1}^N E_q[\delta_{z_n}^k] E_q[u_n] \mathbf{x}_n}{\kappa + \sum_{n=1}^N E_q[\delta_{z_n}^k] E_q[u_n]} = \frac{\bar{\mu}_k \kappa + \sum_{n=1}^N E_q[\delta_{z_n}^k] E_q[u_n] \mathbf{x}_n}{\kappa_{Nk}^*} \\
\gamma_{Nk}^* &= \gamma + \sum_{n=1}^N E_q[\delta_{z_n}^k] \\
\bar{\Sigma}_{Nk}^* &= \bar{\Sigma} + \sum_{n=1}^N E_q[\delta_{z_n}^k] E_q[u_n] (\mathbf{x}_n - \bar{\mathbf{x}}_k)(\mathbf{x}_n - \bar{\mathbf{x}}_k)^T + \frac{\kappa \sum_{n=1}^N E_q[\delta_{z_n}^k] E_q[u_n]}{\kappa + \sum_{n=1}^N E_q[\delta_{z_n}^k] E_q[u_n]} \\
&\quad (\bar{\mu}_k - \bar{\mathbf{x}}_k)(\bar{\mu}_k - \bar{\mathbf{x}}_k)^T \\
\alpha_{N_{nk}}^* &= E_q[\alpha_k] + \frac{J}{2} \\
\beta_{N_{nk}}^* &= E_q[\beta_k] + \frac{1}{2} E_q[(\mathbf{x}_n - \mu_k)^T \Sigma_k^{-1} (\mathbf{x}_n - \mu_k)] \\
\rho_{nk}^* &= \exp(E_q[\ln a_k] - \alpha_{N_{nk}}^* \ln(\beta_{N_{nk}}^*) + \ln(\Gamma(\alpha_{N_{nk}}^*)) + E_q[\alpha_k] E_q[\ln \beta_k] - E_q[\ln(\Gamma(\alpha_k))] - \\
&\quad - \frac{J}{2} \ln(2\pi) - \frac{1}{2} E_q[\ln |\Sigma_k|]).
\end{aligned}$$

4.2. VARIČNÍ BAYESOVSKÉ METODY

Střední hodnoty objevující se ve variační aproximaci určíme z vlastností odpovídajících rozdělení [34]. Z vlastností kategoričkého rozdělení dostaneme pro $k = 1, \dots, K$

$$E_q [\delta_{z_n}^k] = \theta_{nk}^*, \quad \sum_{k=1}^K \theta_{nk}^* = 1. \quad (4.42)$$

Z vlastností Dirichletova rozdělení dostaneme pro $k = 1, \dots, K$

$$E_q [\ln a_k] = \psi(d_{N_k}^*) - \psi\left(\sum_{k=1}^K d_{N_k}^*\right). \quad (4.43)$$

Z vlastností gamma rozdělení dostaneme $k = 1, \dots, K$

$$E_q [u_n] = \frac{\alpha_{N_{nk}}^*}{\beta_{N_{nk}}^*}, \quad (4.44)$$

$$E_q [\ln u_n] = \psi(\alpha_{N_{nk}}^*) - \ln \beta_{N_{nk}}^*. \quad (4.45)$$

Z vlastností inverzního Wishartova rozdělení dostaneme pro $k = 1, \dots, K$

$$E_q [\ln |\Sigma_k|] = \left(\psi_J \left(\frac{\gamma_{N_k}^*}{2} \right) + J \ln(2) + \ln |\bar{\Sigma}_{N_k}^*| \right), \quad (4.46)$$

$$E_q [\Sigma_k^{-1}] = \gamma_{N_k}^* \bar{\Sigma}_{N_k}^{*-1}. \quad (4.47)$$

Z vlastností J -rozměrného normálního rozdělení a inverzního Wishartova rozdělení lze odvodit pro $k = 1, \dots, K$

$$E_q [(\mathbf{x}_n - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_k)] = (\mathbf{x}_n - \boldsymbol{\mu}_{N_k}^*)^T \gamma_{N_k}^* \bar{\Sigma}_{N_k}^{*-1} (\mathbf{x}_n - \boldsymbol{\mu}_{N_k}^*) + \frac{J}{\kappa_{N_k}^*}. \quad (4.48)$$

Střední hodnotu parametru β_k lze určit z aposteriorního rozdělení v závislosti na parametru α_k , tedy pro $k = 1, \dots, K$

$$E_q [\beta_k] = \frac{s_{N_k}^* \mathbb{E}_q [\alpha_k] + 1}{q_{N_k}^*}, \quad (4.49)$$

$$E_q [\ln \beta_k] = \mathbb{E}_q [\psi(s_{N_k} \alpha_k + 1)] - \ln q_{N_k}^*. \quad (4.50)$$

Nicméně střední hodnoty závisující na parametru α_k jsou analyticky neřešitelné a je nutná jejich aproximace pomocí metody založené na Laplacově aproximaci dle [39] viz podkapitola 4.2.6. Tedy aproximujeme integrály pro $k = 1, \dots, K$

$$E_q [\alpha_k] = \int \alpha_k p(\alpha_k | p_{N_k}^*, q_{N_k}^*, r_{N_k}^*, s_{N_k}^*) d\alpha_k, \quad (4.51)$$

$$E_q [\psi(s_{N_k} \alpha_k + 1)] = \int \psi(s_{N_k}^* \alpha_k + 1) p(\alpha_k | p_{N_k}^*, q_{N_k}^*, r_{N_k}^*, s_{N_k}^*) d\alpha_k, \quad (4.52)$$

$$E_q [\ln(\Gamma(\alpha_k))] = \int \ln(\Gamma(\alpha_k)) p(\alpha_k | p_{N_k}^*, q_{N_k}^*, r_{N_k}^*, s_{N_k}^*) d\alpha_k. \quad (4.53)$$

Výsledný algoritmus pro určený parametrů GMM pro deinterleaving a latentních proměnných:

Data: Sekvence impulzů popsána PDW maticí \mathbf{X}
Input: $K, d, p, q, r, s, \kappa, \bar{\boldsymbol{\mu}}_k, \gamma, \bar{\boldsymbol{\Sigma}}_k, k = 1, \dots, K$
Output: Latentní proměnná \mathbf{z}
while *dokud není splněna zastavovací podmínka* **do**
 | update latentních proměnných $\alpha_{N_{nk}}, \beta_{N_{nk}}, \rho_{nk}, \theta_{nk}$
 | update neznámých parametrů $d_{Nk}, p_{Nk}, q_{Nk}, r_{Nk}, s_{Nk}, \boldsymbol{\mu}_{Nk}, \boldsymbol{\Sigma}_{Nk}, \gamma_{Nk}, \bar{\boldsymbol{\Sigma}}_{Nk}$
end
 $z_n = \arg \max_{k=1, \dots, K} \theta_{nk}$
return \mathbf{z}

Algoritmus 2: Algoritmus pro deinterleaving pomocí GMM

VB metoda pro model (4.41) je schopna automaticky určit počet shluků, za předpokladu, že vstupní parametr K je větší jak skutečný počet shluků. Zadáním většího počtu shluků metoda přiřadí automaticky některým ze shluků váhu a_k blízko nule a tím dojde k automatickému výběru počtu shluků. Vstupní parametr K lze určit například z histogramu parametrů, kde se odhad počtu shluků vezme jako počet lokálních maxim.

5 Experiment

Ověření vhodnosti metod z kapitoly 4 pro úlohu autonomního jednokanálového deinterleavingu bylo provedeno na simulovaných a datech z reálného provozu. Simulovaná data byla vygenerována v softwaru Matlab. Implementace metod byla provedena v programovacím jazyce Python.

5.1 Hodnocení úspěšnosti deinterleavingu

U simulovaných dat je známo správné roztrídění signálů k jednotlivým emiterům (dále budeme nazývat třídy). Pro hodnocení úspěšnosti deinterleavingu lze využít metriky pro hodnocení úspěšnosti shlukování jako je adjusted rand index (ARI), adjusted mutual information (AMI), stejnorodost (homogeneity) a úplnost (completeness).

5.1.1 ARI

Rand index (RI) měří podobnost dvou rozdělení dat poměrem shodně zařazených dvojic a všech možných dvojic dat. ARI je pak RI upraveno vůči náhodě. Pro definování RI si zavedeme následující značení. Označme známé rozdělení dat symbolem \mathbf{C} a rozdělení dat získané shlukováním symbolem \mathbf{P} , pak

- N_{11} je počet dvojic patřících do stejného shluku v \mathbf{C} i v \mathbf{P} ,
- N_{00} je počet dvojic patřících do různých shluků v \mathbf{C} i v \mathbf{P} ,
- N_{10} je počet dvojic patřících do stejného shluku v \mathbf{C} ale do různých shluků v \mathbf{P} ,
- N_{01} je počet dvojic patřících do různých shluků v \mathbf{C} ale do stejných shluků v \mathbf{P} .

N_{11} a N_{00} představují počet shodně zařazených dvojic a N_{01} a N_{10} pak počet neshodně zařazených dvojic. Rand index je pak roven

$$RI = \frac{N_{00} + N_{11}}{N_{00} + N_{11} + N_{01} + N_{10}} = \frac{N_{00} + N_{11}}{\binom{N}{2}} \in \langle 0, 1 \rangle, \quad (5.1)$$

kde N je počet dat.

Při shlukování, kde počet shluků je relativně velký vůči počtu dat dostaneme vysokou hodnotu RI i přesto, že daný výsledek není úspěšný. V takovémto případě dostaneme vysokou hodnotu N_{00} . Z tohoto důvodu upravíme RI vůči náhodě. Vezmeme v potaz pravděpodobnost, že prvky jsou přiřazeny shodně ve smyslu N_{11} , N_{00} náhodou. Hodnota ARI je pak rovna

$$ARI = \frac{2(N_{00}N_{11} - N_{01}N_{10})}{(N_{00} + N_{01})(N_{01} + N_{11}) + (N_{00} + N_{10})(N_{10} + N_{11})} \in \langle -1, 1 \rangle \quad (5.2)$$

Pokud $ARI = 0$, pak je shlukování provedeno náhodně, zatímco $ARI = \pm 1$ značí perfektní shodu. Obecně lze vztah (5.2) zapsat pomocí střední hodnoty [40]

$$ARI = \frac{RI - E[RI]}{\max(RI) - E[RI]}. \quad (5.3)$$

Dále poznamenejme případ, kdy počet prvků v jednotlivých třídách je značně rozdílný. V takovémto případě může dojít k nízké hodnotě *ARI* i přesto, že je vyhodnocen špatně pouze jeden shluk. To může nastat, pokud špatně vyhodnocený shluk má velký počet prvků oproti ostatním. Tento případ je velmi častý pro úlohu deinterleavingu, kdy každý z emiterů vyzařuje signál různou dobu a s různými hodnotami PRI. To způsobí že počet impulzů od jednotlivých emiterů je nevyrovnaný. Proto je vhodné využít pro hodnocení úspěšnosti deinterleavingu i jiné metriky.

5.1.2 AMI

Mutual information (MI) měří shodu dvou rozdělení dat. Adjusted mutual information (AMI) pak upravuje, ve stejném smyslu jako *ARI*, výpočet *MI* vůči náhodě. Mějme dvě rozdělení dat \mathbf{C}, \mathbf{P} , kde $|\mathbf{C}_i|$ označuje počet prvků ve shluku \mathbf{C}_i . *MI* je určeno vztahem

$$MI = \sum_{i=1}^{|\mathbf{C}|} \sum_{j=1}^{|\mathbf{P}|} \frac{|\mathbf{C}_i \cap \mathbf{P}_j|}{N} \log \frac{N|\mathbf{C}_i \cap \mathbf{P}_j|}{|\mathbf{C}_i||\mathbf{P}_j|}, \quad (5.4)$$

kde N je celkový počet dat a $|\mathbf{C}|$ resp. $|\mathbf{P}|$ je počet shluků v rozdělení \mathbf{C} resp. \mathbf{P} . AMI je pak obecně rovno

$$AMI = \frac{MI - E[MI]}{\max(MI) - E[MI]}. \quad (5.5)$$

Detailní vyjádření AMI lze nalézt v [40].

5.1.3 Stejnorodost

Řekneme, že výsledek shlukování je stejnorodý, jestliže každý ze shluků obsahuje pouze prvky z jedné třídy. Stejnorodost tedy popisuje v jaké míře došlo ke spojení různých tříd do jednoho shluku. Stejnorodost je určena vztahem

$$Homogeneity = 1 - \frac{H(\mathbf{C}|\mathbf{P})}{H(\mathbf{C})}, \quad (5.6)$$

kde

$$H(\mathbf{C}|\mathbf{P}) = \sum_{j=1}^{|\mathbf{P}|} \sum_{i=1}^{|\mathbf{C}|} \frac{|\mathbf{C}_i \cap \mathbf{P}_j|}{N} \log \frac{|\mathbf{C}_i \cap \mathbf{P}_j|}{|\mathbf{C}_i|}, \quad (5.7)$$

$$H(\mathbf{C}) = - \sum_{i=1}^{|\mathbf{C}|} \frac{|\mathbf{C}_i|}{N} \log \frac{|\mathbf{C}_i|}{N}. \quad (5.8)$$

Jestliže $Homogeneity = 1$, pak při shlukování nedošlo ke spojení více tříd do jednoho shluku [35].

5.2. SIMULOVANÁ DATA

5.1.4 Úplnost

Řekneme, že výsledek shlukování je úplný, jestliže všechny prvky příslušející jedné třídě jsou zařazeny do stejného shluku. Úplnost tedy popisuje v jaké míře došlo k rozdělení prvků, patřících do jedné třídy, do různých shluků. Jedná se tedy o komplement ke stejnorodosti. Úplnost je určena vztahem

$$Completeness = 1 - \frac{H(\mathbf{P}|\mathbf{C})}{H(\mathbf{P})}, \quad (5.9)$$

kde

$$H(\mathbf{P}|\mathbf{C}) = \sum_{i=1}^{|\mathbf{C}|} \sum_{j=1}^{|\mathbf{P}|} \frac{|\mathbf{C}_i \cap \mathbf{P}_j|}{N} \log \frac{|\mathbf{C}_i \cap \mathbf{P}_j|}{|\mathbf{P}_j|}, \quad (5.10)$$

$$H(\mathbf{P}) = - \sum_{j=1}^{|\mathbf{P}|} \frac{|\mathbf{P}_j|}{N} \log \frac{|\mathbf{P}_j|}{N}. \quad (5.11)$$

Jestliže $Completeness = 1$, pak při shlukování nedošlo k rozdělení žádné z tříd do více shluků [35].

5.2 Simulovaná data

Ověření účinnosti metod z kapitoly 4 pro autonomní jednokanálový deinterleaving bylo provedeno na devíti sadách simulovaných dat, kde každá sada obsahovala sto scénářů. Parametry emiterů pro jednotlivé scénáře byly generovány náhodně v softwaru Matlab se zvoleným rozmezím, s případnou modulací parametrů a s odlišným počtem vyzařovaných impulzů. Dále popisujeme pouze parametry RF a PW, jelikož deinterleaving je prováděn právě pomocí těchto dvou parametrů. Parametry RF a PW byly generovány v rozmezí: $\text{RF} \in \langle 9000, 9250 \rangle \text{ MHz}$, $\text{PW} \in \langle 0.5, 20 \rangle \mu\text{s}$.

Možné modulační v RF parametru uvažujeme konstantní, jitter, stagger a D&S a v PW parametru konstantní a stagger. Jejich definice jsou analogické s modulacemi PRI v podkapitole 1.2.4. Zdvihy jednotlivých modulací byly voleny tak, aby odpovídaly reálným hodnotám. Dále byly namodelovány rušivé vlivy a to výpadky impulzů, falešné impulzy a chyby měření. Výpadky impulzu vznikají vlivem poklesu výkonu signálu pod detekční práh přijímače v důsledku prohledávání prostoru anténím svazkem. Falešné impulzy vznikají vlivem externího rušení a vlivem šumu přijímače. Modelování chyb měření parametru v přijímači bylo provedeno gaussovským šumem, jehož rozptyl je závislý na amplitudě (respektive výkonu) impulzu PA.

Vstupní parametry DBSCAN metody byly voleny na základě podkapitoly 4.1, tedy $MinPts = 2 \dim(\mathbf{X})$, kde \mathbf{X} je množina dat a parametr ϵ byl odhadnut z grafu k -nejbližších sousedů, pro $k = MinPts - 1$.

Inicializační parametry pro VB metodu byly zvoleny následovně

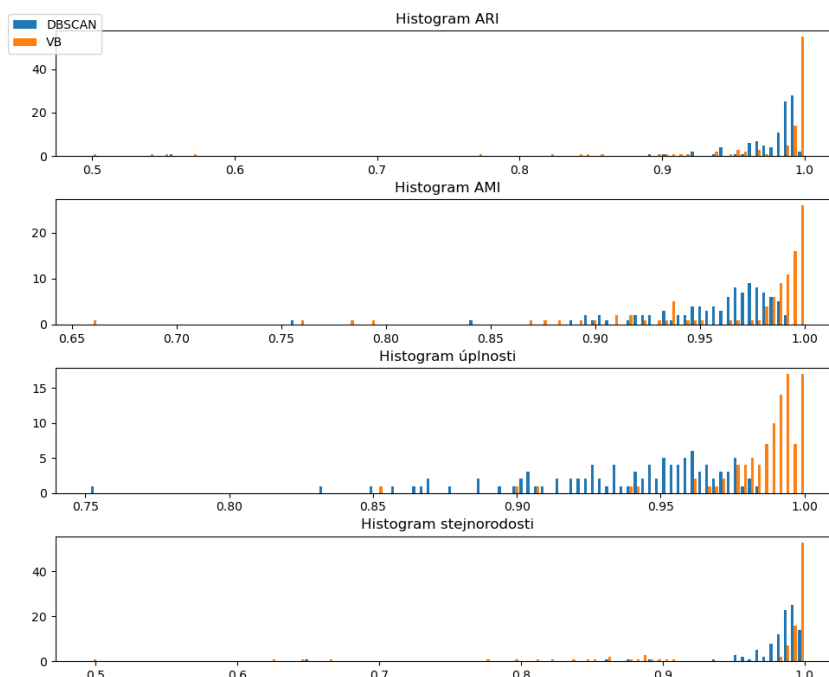
$$p, q, r, s = 0, 2; \quad \kappa = 1; \quad \gamma = 1; \quad d = 1/K. \quad (5.12)$$

Počet shluků K , byl určen jako počet lokálních maxim v histogramu parametru RF. Z důvodu citlivosti metody na volbu inicializačních parametrů $\bar{\mu}_k, \bar{\Sigma}_k$, byly parametry $\bar{\mu}_k$ určeny pomocí k-means algoritmu a parametry $\bar{\Sigma}_k$ jako diagonální matice obsahující rozptyly jednotlivých parametrů využívaných pro deinterleaving pro $k = 1, \dots, K$.

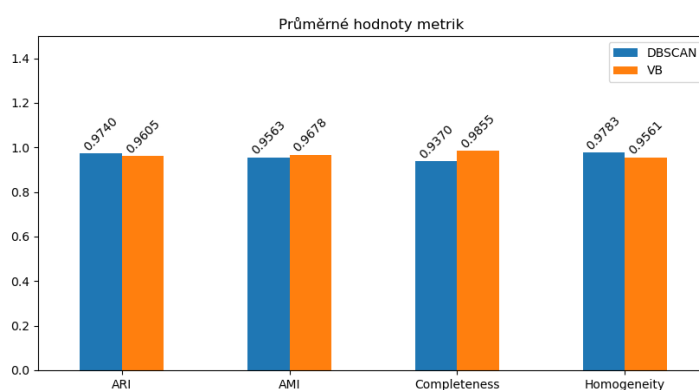
5.2.1 Sada dat 1

Scénáře první sady dat obsahují signály od čtyř emiterů, jejichž parametry RF a PW jsou konstantní po celou dobu vyzařování. Každý scénář obsahuje dva emitery s krátkou šířkou pulzu a dva emitery s dlouhou šířkou impulsu.

Deinterleaving první sady dat byl proveden pomocí DBSCAN a VB metod pro všechny scénáře. Vyhodnocení úspěšnosti bylo provedeno pomocí metrik z podkapitoly 5.1. Histogramy hodnot metrik pro obě metody jsou zobrazeny na obrázku 5.1 a jejich průměrné hodnoty na obrázku 5.2. Odhadnutý počet emiterů pomocí metod DBSCAN a VB pro jednotlivé scénáře je na obrázku 5.3.

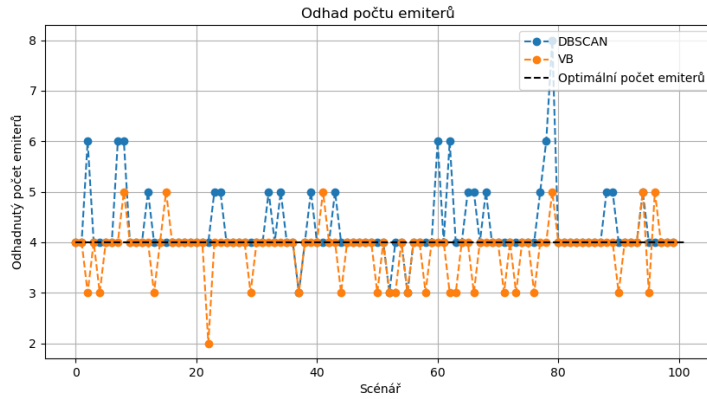


Obrázek 5.1: Histogramy metrik pro hodnocení úspěšnosti deinterleavingu sady dat 1



Obrázek 5.2: Průměrné hodnoty metrik pro sadu dat 1

5.2. SIMULOVANÁ DATA



Obrázek 5.3: Odhad počtu emiterů pro všechny scénáře sady dat 1

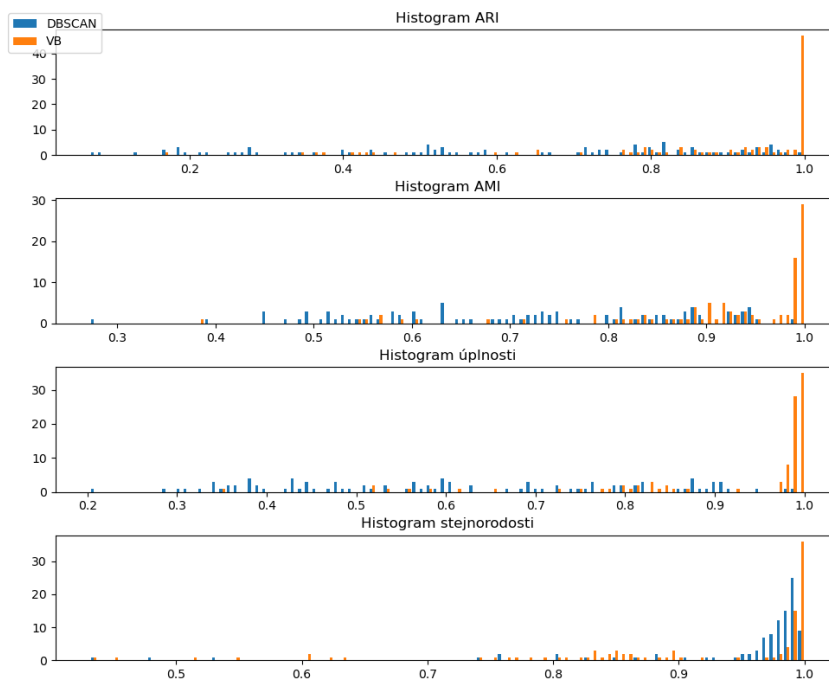
DBSCAN algoritmus identifikuje body, které nemají ve svém ϵ -okolí dostatečné množství bodů, jako šum. Jelikož data jsou zatížená gaussovským šumem, dochází k odlehlým hodnotám, které již DBSCAN algoritmus vyhodnotí jako šum. Tento fakt se promítne do hodnot metrik ARI, AMI a úplnosti, avšak pro účely deinterleavingu je tato vlastnost žádoucí. Pokud zohledníme tento fakt, lze z obrázků 5.1, 5.2, 5.3 usoudit, že metody pro data s konstantními parametry RF a PW mají stejnou účinnost.

Mohou se vyskytnout případy, kdy účinnost jedné z metod převažuje nad tou druhou. Například pokud dva emiteri mají podobné parametry RF a PW, metoda VB tyto impulzy přiřadí k jednomu emiteru. Naopak pokud šum emiteru má velký rozptyl, DBSCAN rozdělí impulzy do dvou nebo více emiterů. Projevy těchto jevů můžeme pozorovat na obrázku 5.3.

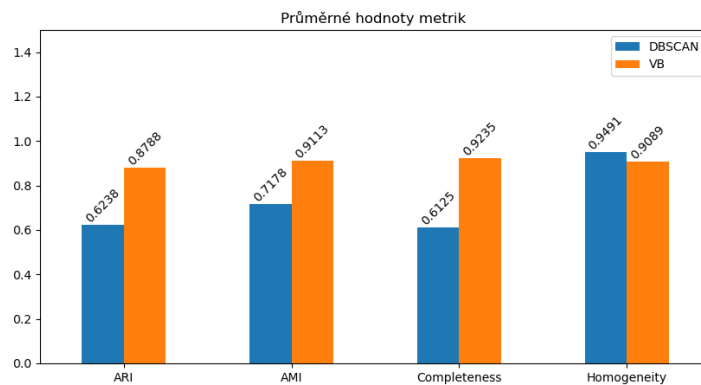
5.2.2 Sada dat 2

Scénáře druhé sady dat obsahují čtyři emiteri, které mají agilní RF parametr s modulací jitter a konstantní PW parametr. Každý scénář obsahuje dva emiteri s krátkou šířkou pulzu a dva emiteri s dlouhou šířkou impulzu.

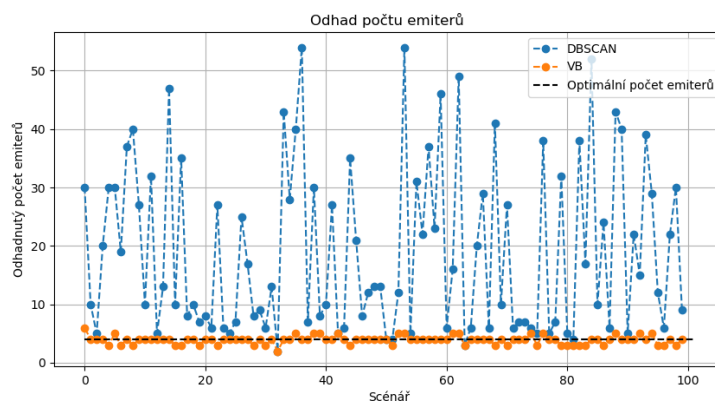
Deinterleaving druhé sady dat byl proveden pomocí DBSCAN a VB metod pro všechny scénáře. Vyhodnocení úspěšnosti bylo provedeno pomocí metrik z podkapitoly 5.1. Histogramy hodnot metrik pro obě metody jsou zobrazeny na obrázku 5.4 a jejich průměrné hodnoty na obrázku 5.5. Odhadnutý počet emiterů pomocí metod DBSCAN a VB pro jednotlivé scénáře je na obrázku 5.6.



Obrázek 5.4: Histogramy metrik pro hodnocení úspěšnosti deinterleavingu sady dat 2



Obrázek 5.5: Odhad počtu emiterů pro všechny scénáře sady dat 2



Obrázek 5.6: Odhad počtu emiterů pro všechny scénáře sady dat 2

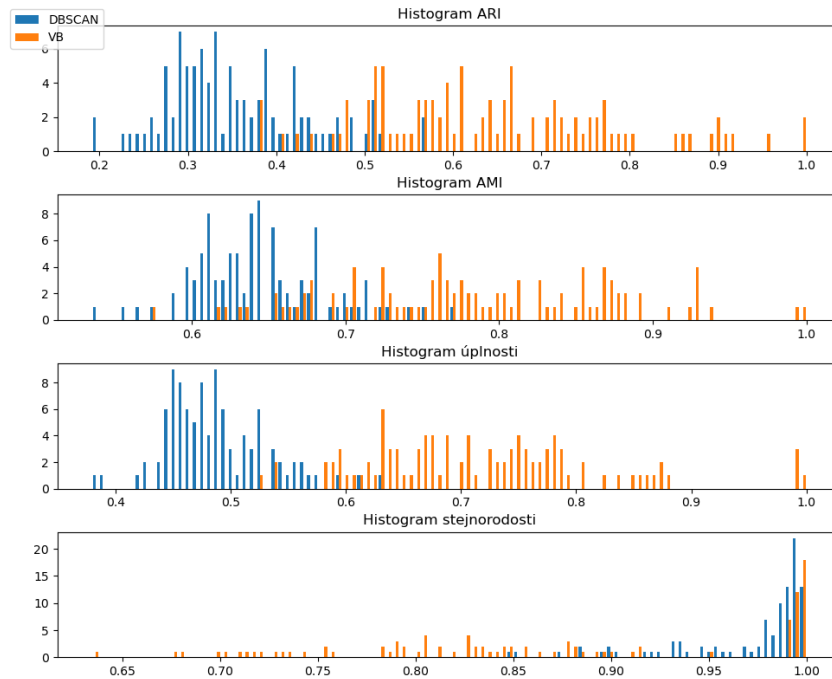
5.2. SIMULOVANÁ DATA

V případě jitter modulace parametru RF dochází u DBSCAN algoritmu k rozdělení impulzů z jednoho emiteru do více shluků (Obrázek 5.6). To nastává v případě velkého rozptylu modulace jitter, kdy DBSCAN z důvodů nízké hustoty rozdělí jeden shluk do více shluků. V takovémto případě dochází k velkému nadhodnocení počtu shluků. VB metoda je pro tento případ mnohem účinnější, jelikož dokáže emiteru s jitter modulací ve většině případů úspěšně identifikovat.

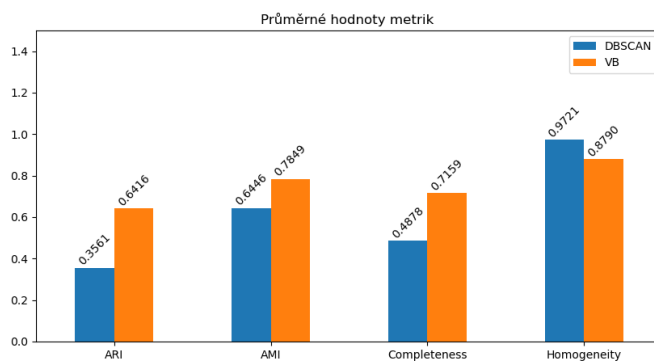
5.2.3 Sada dat 3

Scénáře třetí sady dat obsahují čtyři emiteru, které mají modulaci stagger RF parametru a konstantní PW parametr. Každý scénář obsahuje dva emiteru s krátkou šířkou pulzu a dva emiteru s dlouhou šířkou impulzu.

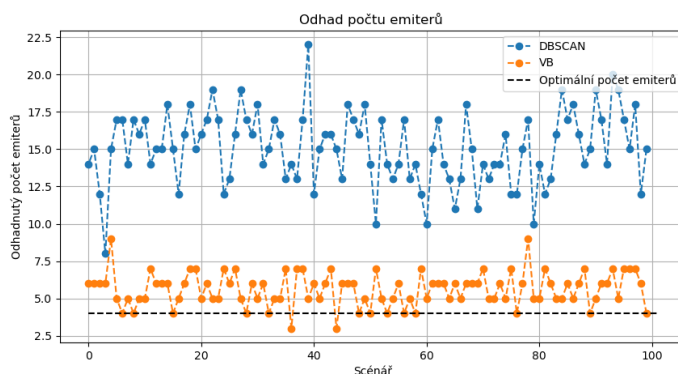
Deinterleaving třetí sady dat byl proveden pomocí DBSCAN a VB metod pro všechny scénáře. Vyhodnocení úspěšnosti bylo provedeno pomocí metrik z podkapitoly 5.1. Histogramy hodnot metrik pro obě metody jsou zobrazeny na obrázku 5.7 a jejich průměrné hodnoty na obrázku 5.8. Odhadnutý počet emiterů pomocí metod DBSCAN a VB pro jednotlivé scénáře je na obrázku 5.9.



Obrázek 5.7: Histogramy metrik pro hodnocení úspěšnosti deinterleavingu sady dat 3



Obrázek 5.8: Odhad počtu emiterů pro všechny scénáře sady dat 3



Obrázek 5.9: Odhad počtu emiterů pro všechny scénáře sady dat 3

DBSCAN definuje shluky jako body, jenž mají dostatečnou hustotu a jsou spojené vzhledem k hustotě. Tuto vlastnost emiterů s modulací stagger nesplňují. RF parametr s modulací stagger má několik úrovní RF, které se impuls od impulsu mění. Tím se vytvoří několik úrovní RF pro jeden emiter. Účinnost DBSCAN algoritmu je nízká kvůli vyhodnocování jednotlivých úrovní stagger modulace jako samostatných emiterů. Dochází tedy k rozdělení shluků, což je patrné z hodnot úplnosti (Obrázek 5.7) a z odhadu počtu emiterů (Obrázek 5.9).

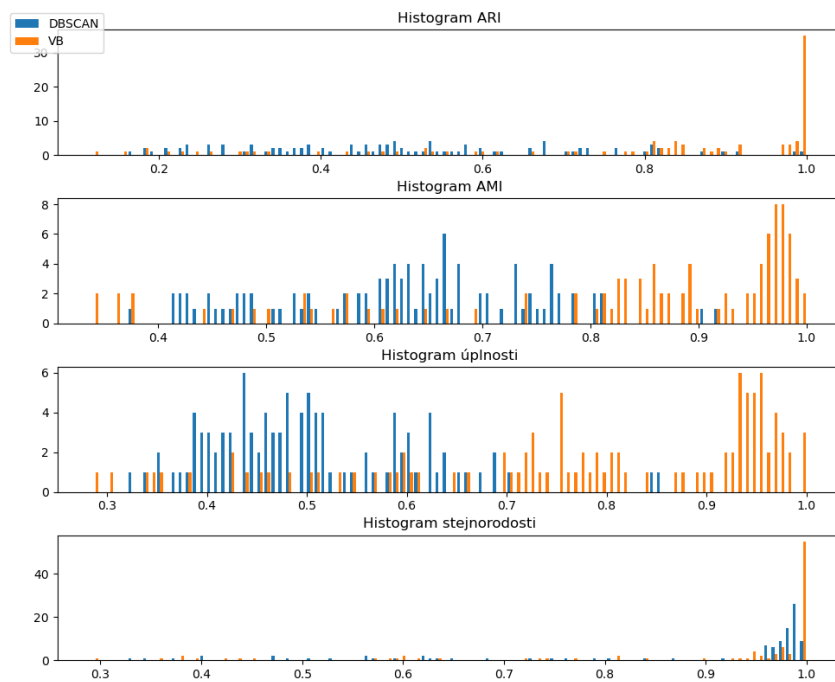
VB metoda dokáže v některých případech jednotlivé úrovně správně zařadit k jednomu emiteru a to díky definování odlehlých hodnot v modelu (4.41). Avšak pokud je změna parametru příliš velká, dochází také k rozdělení úrovní stagger modulace. Tento fakt je patrný jak z histogramu, tak z průměrných hodnot ARI a AMI.

5.2.4 Sada dat 4

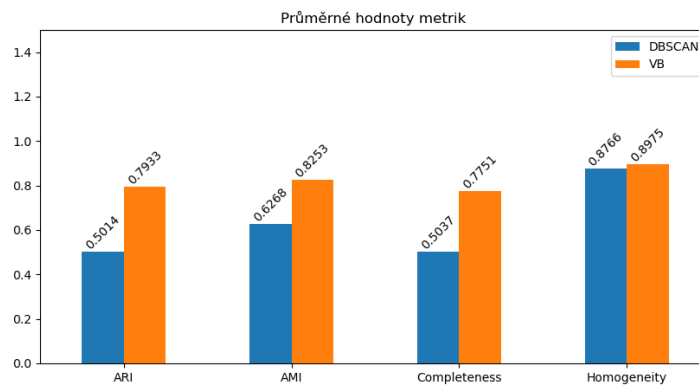
Scénáře čtvrté sady dat obsahují čtyři emiterů, které mají agilní RF parametr s modulací D&S a konstantní PW parametr. Každý scénář obsahuje dva emiterů s krátkou šířkou pulzu a dva emiterů s dlouhou šířkou impulsu.

Deinterleaving čtvrté sady dat byl proveden pomocí DBSCAN a VB metod pro všechny scénáře. Vyhodnocení úspěšnosti bylo provedeno pomocí metrik z podkapitoly 5.1. Histogramy hodnot metrik pro obě metody jsou zobrazeny na obrázku 5.10 a jejich průměrné hodnoty na obrázku 5.11. Odhadnutý počet emiterů pomocí metod DBSCAN a VB pro jednotlivé scénáře je na obrázku 5.12.

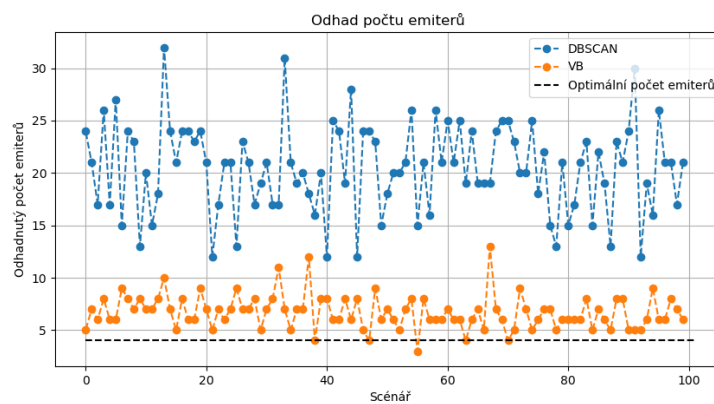
5.2. SIMULOVANÁ DATA



Obrázek 5.10: Histogramy metrik pro hodnocení úspěšnosti deinterleavingu sady dat 4



Obrázek 5.11: Odhad počtu emiterů pro všechny scénáře sady dat 4



Obrázek 5.12: Odhad počtu emiterů pro všechny scénáře sady dat 4

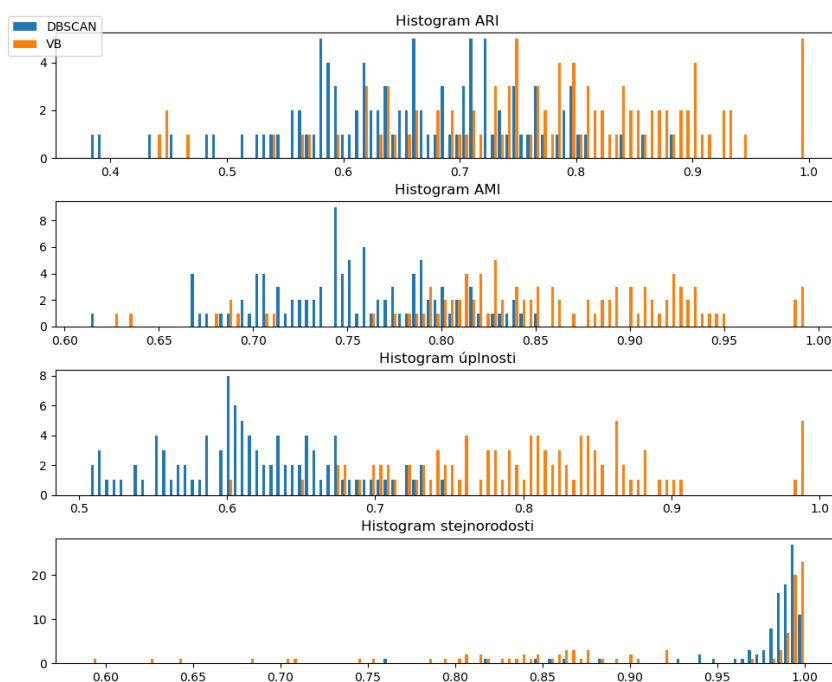
Podobně jako u stagger modulace, viz podkapitola 5.2.3, bloková změna D&S modulace vytvoří několik úrovní parametru RF. DBSCAN není schopen tyto úrovně vyhodnotit jako jeden shluk. Dochází tedy stejně jako u stagger modulace k rozdělení impulzů jednoho emiteru do více shluků. To je patrné z hodnot úplnosti (Obrázek 5.10) a z odhadu počtu emiterů (Obrázek 5.12).

VB metoda v některých případech tuto blokovou změnu dokáže identifikovat a to díky definování odlehklých hodnot v modelu (4.41). Avšak pokud je změna parametru příliš velká, dochází také k rozdělení úrovní D&S modulace. Tento fakt je patrný jak z histogramu, tak z průměrných hodnot ARI a AMI.

5.2.5 Sada dat 5

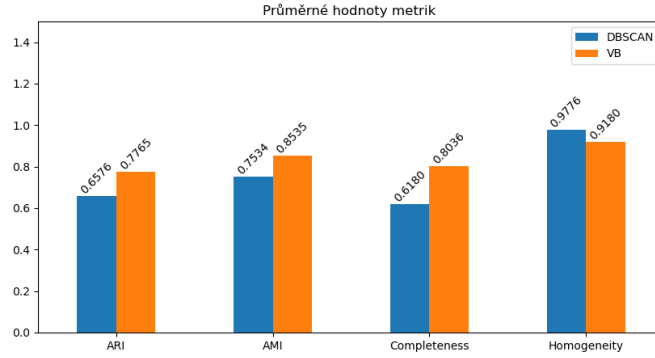
Scénáře páté sady dat obsahují čtyři emitory, kde dva emitory mají konstantní RF i PW parametr a dva emitory mají agilní RF parametr s modulací stagger a konstantní PW parametr.

Deinterleaving páté sady dat byl proveden pomocí DBSCAN a VB metod pro všechny scénáře. Vyhodnocení úspěšnosti bylo provedeno pomocí metrik z podkapitoly 5.1. Histogramy hodnot metrik pro obě metody jsou zobrazeny na obrázku 5.13 a jejich průměrné hodnoty na obrázku 5.14. Odhadnutý počet emiterů pomocí metod DBSCAN a VB pro jednotlivé scénáře je na obrázku 5.15.

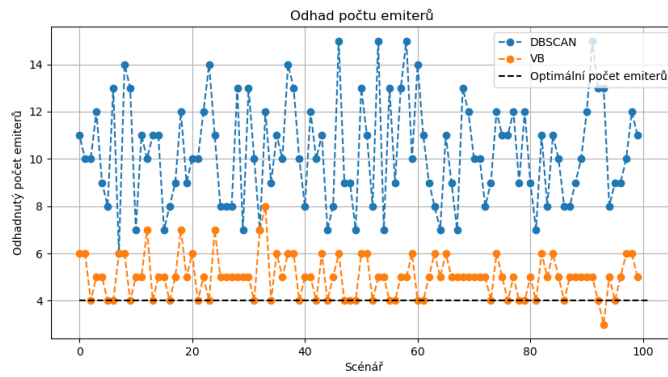


Obrázek 5.13: Histogramy metrik pro hodnocení úspěšnosti deinterleavingu sady dat 5

5.2. SIMULOVANÁ DATA



Obrázek 5.14: Odhad počtu emiterů pro všechny scénáře sady dat 5



Obrázek 5.15: Odhad počtu emiterů pro všechny scénáře sady dat 5

Jak již bylo zmíněno výše DBSCAN metoda nedokáže správně identifikovat emitery se stagger modulací. DBSCAN identifikuje každou úroveň stagger modulace jako samostatný shluk. To lze opět vidět na účinnost algoritmu (Obrázek 5.13, 5.14) a na odhadu počtu emiterů (Obrázek 5.15).

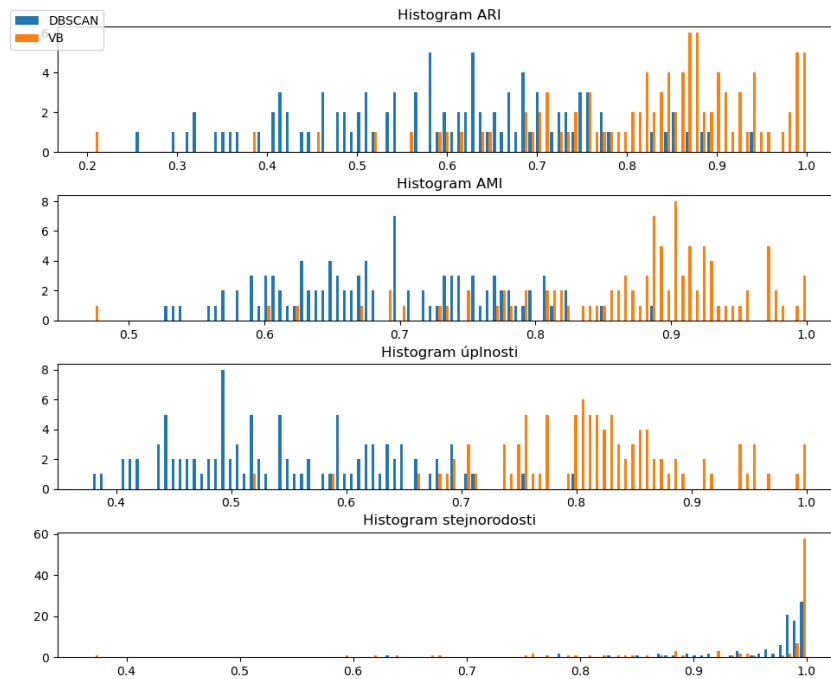
Ani metoda VB není v tomto případě příliš účinná, jelikož dochází často ke spojování úrovní stagger modulace od různých emiterů. V některých případech stejně jako DBSCAN i metoda VB rozděluje jednotlivé úrovně. Dalším nedostatkem VB metody je spojování emiterů s konstantním RF s emiterem se stagger modulací RF parametru.

5.2.6 Sada dat 6

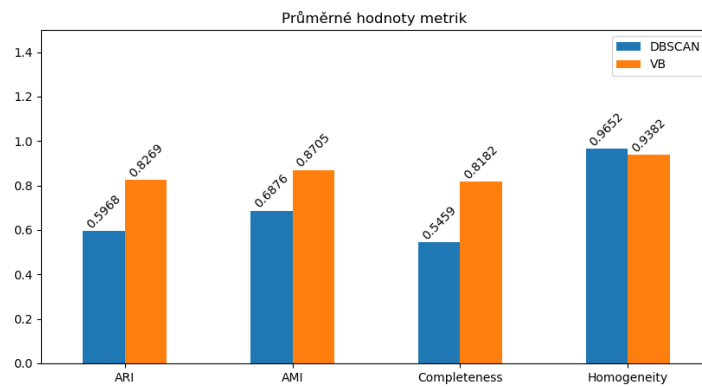
Scénáře šesté sady dat obsahují čtyři emitery, kde všechny emitery mají dlouhou šířku impulzů. Parametry PW pro sadu dat 6 byly generovány v rozmezí $\langle 5, 20 \rangle \mu s$. Všechny emitery mají konstantní PW parametr a každý z nich má agilní RF parametr. Je zde zastoupena modulace konstantní, jitter, stagger a D&S.

Dlouhá šířka impulzu způsobuje zúžení PW pásma. Díky tomu dochází ke ztížení deinterleavingu, jelikož PW parametry emiterů jsou blízkých hodnot.

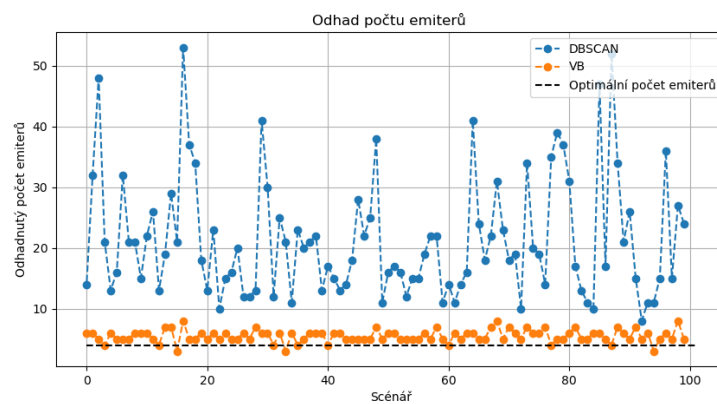
Deinterleaving šesté sady dat byl proveden pomocí DBSCAN a VB metod pro všechny scénáře. Vyhodnocení úspěšnosti bylo provedeno pomocí metrik z podkapitoly 5.1. Histogramy hodnot metrik pro obě metody jsou zobrazeny na obrázku 5.16 a jejich průměrné hodnoty na obrázku 5.17. Odhadnutý počet emiterů pomocí metod DBSCAN a VB pro jednotlivé scénáře je na obrázku 5.18.



Obrázek 5.16: Histogramy metrik pro hodnocení úspěšnosti deinterleavingu sady dat 6



Obrázek 5.17: Odhad počtu emiterů pro všechny scénáře sady dat 6



Obrázek 5.18: Odhad počtu emiterů pro všechny scénáře sady dat 6

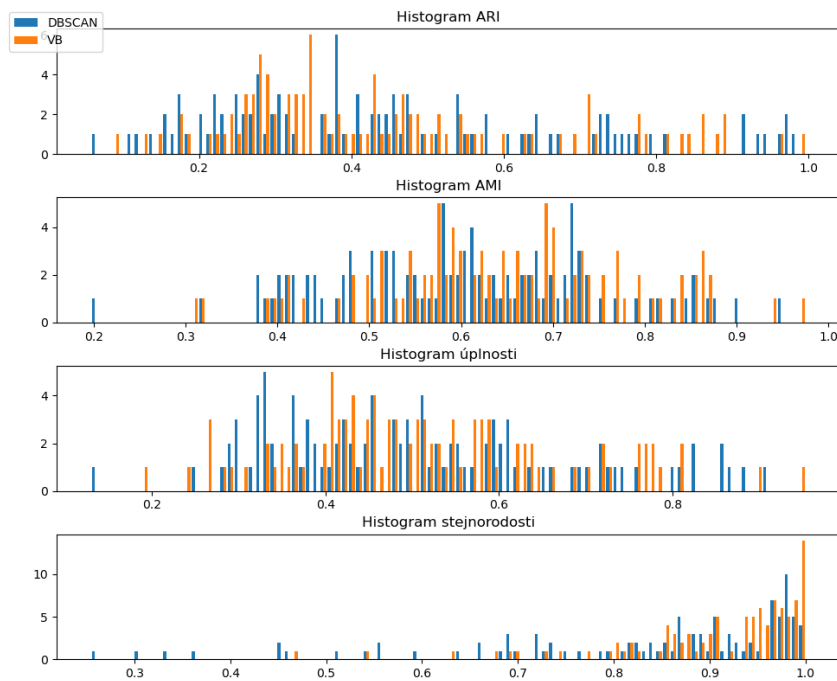
5.2. SIMULOVANÁ DATA

DBSCAN algoritmus vykazuje nízkou účinnost z důvodu rozdělení emiterů s modulací stagger, D&S a jitter (viz předchozí výsledky). Tento fakt můžeme pozorovat na histogramu úplnosti a odhadu počtu emiterů. Z histogramu úplnosti je patrné, že VB metoda také rozděluje impulzy do různých shluků avšak v menší míře než DBSCAN (obrázky 5.16, 5.18).

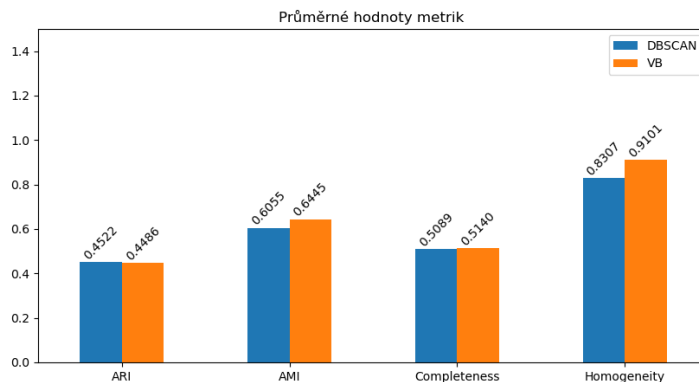
5.2.7 Sada dat 7

Scénáře sedmé sady dat obsahují čtyři emitery, kde všechny emitery mají krátkou šířku impulzů. Všechny emitery mají konstantní PW parametr a každý z nich má agilní RF parametr. Je zde zastoupena modulace konstatní, jitter, stagger a D&S.

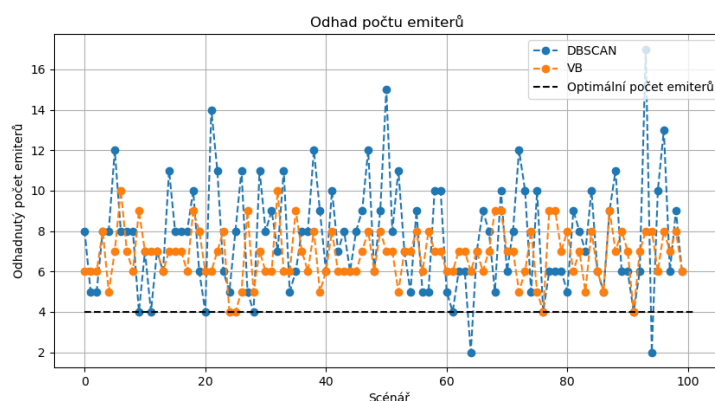
Deinterleaving sedmé sady dat byl proveden pomocí DBSCAN a VB metod pro všechny scénáře. Vyhodnocení úspěšnosti bylo provedeno pomocí metrik z podkapitoly 5.1. Histogramy hodnot metrik pro obě metody jsou zobrazeny na obrázku 5.19 a jejich průměrné hodnoty na obrázku 5.20. Odhadnutý počet emiterů pomocí metod DBSCAN a VB pro jednotlivé scénáře je na obrázku 5.21.



Obrázek 5.19: Histogramy metrik pro hodnocení úspěšnosti deinterleavingu sady dat 7



Obrázek 5.20: Odhad počtu emiterů pro všechny scénáře sady dat 7



Obrázek 5.21: Odhad počtu emiterů pro všechny scénáře sady dat 7

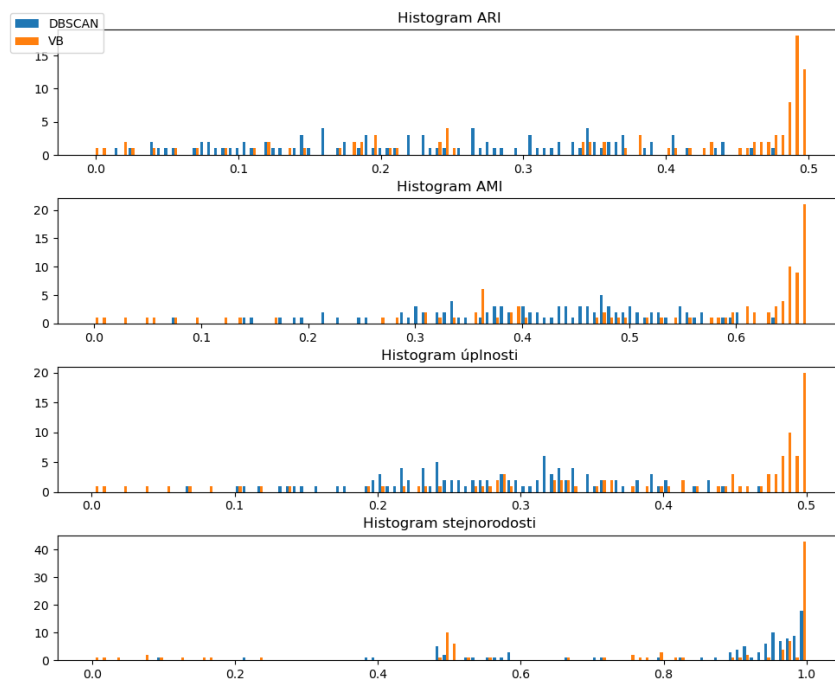
Krátká šířka impulzů způsobí ještě větší zúžení PW pásma, než u dlouhé šířky impulzů. Díky tomu dojde ke ztížení deinterleavingu, jelikož PW parametry emiterů jsou velmi blízkých hodnot. Ze stejného důvodu jako u sady dat 6, vykazují metody nízkou účinnost. Z obrázku 5.21 je patrné, že rozdělují impulzy z jednoho emiteru do více emiterů. To lze pozorovat z výsledků na obrázcích 5.19 a 5.20, kde průměrné hodnoty ARI jsou menší jak 0.5.

5.2.8 Sada dat 8

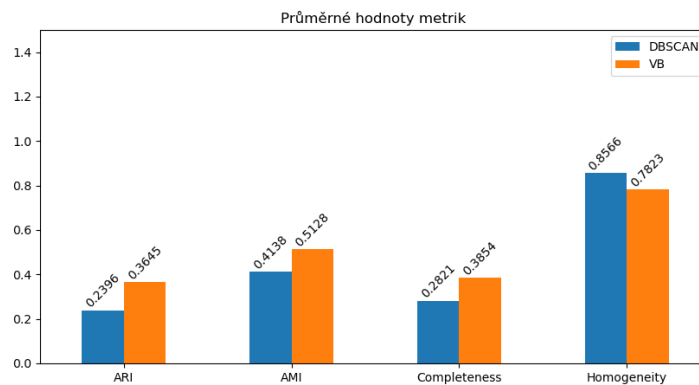
Scénáře osmé sady dat obsahují dva emitery, kde oba emitery mají agilní RF parametr s modulací jitter a agilní PW parametr s modulací stagger.

Deinterleaving osmé sady dat byl proveden pomocí DBSCAN a VB metod pro všechny scénáře. Vyhodnocení úspěšnosti bylo provedeno pomocí metrik z podkapitoly 5.1. Histogramy hodnot metrik pro obě metody jsou zobrazeny na obrázku 5.22 a jejich průměrné hodnoty na obrázku 5.23. Odhadnutý počet emiterů pomocí metod DBSCAN a VB pro jednotlivé scénáře je na obrázku 5.24.

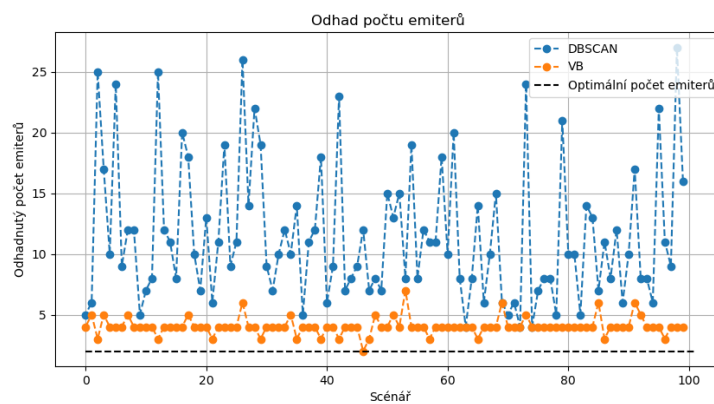
5.2. SIMULOVANÁ DATA



Obrázek 5.22: Histogramy metrik pro hodnocení úspěšnosti deinterleavingu sady dat 8



Obrázek 5.23: Odhad počtu emiterů pro všechny scénáře sady dat 8



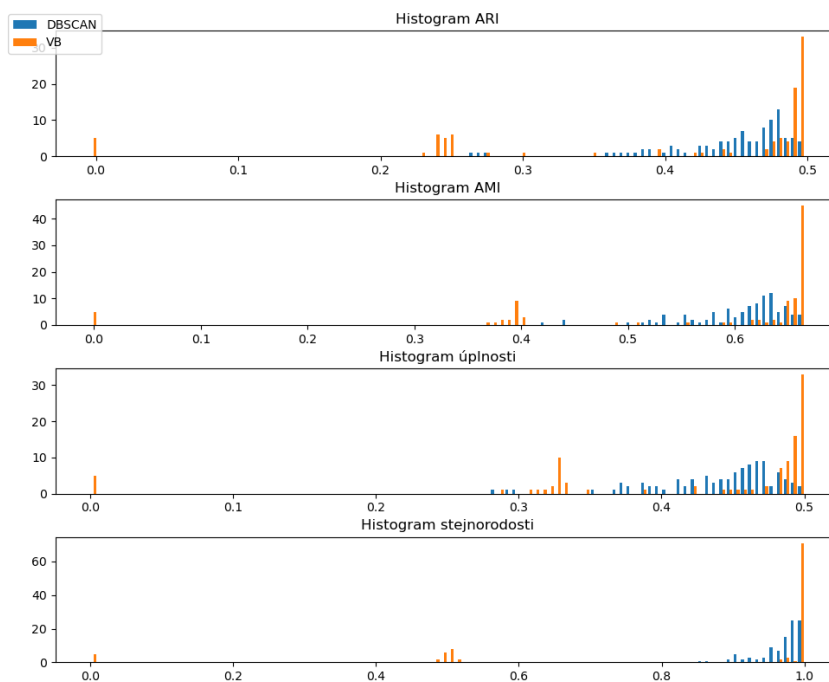
Obrázek 5.24: Odhad počtu emiterů pro všechny scénáře sady dat 8

Tato sada dat je velmi obtížná pro úlohu deinterleavingu pomocí obou metod, což lze pozorovat z hodnot ARI a AMI. VB metoda vykazuje lepší účinnost, jelikož je účinnější pro rozpoznání emiterů s modulací jitter (viz 5.2.2). Avšak identifikace modulace stagger PW parametru je velmi obtížná pro obě metody. Tento fakt lze pozorovat z obrázku 5.24, kde VB metoda ve většině případů odhadne čtyři emiterů. Každou úroveň stagger modulace identifikuje jako jeden emiter. DBSCAN odhaduje velký počet emiterů, což je způsobeno rozdělováním impulzů s jitter modulací RF parametru.

5.2.9 Sada dat 9

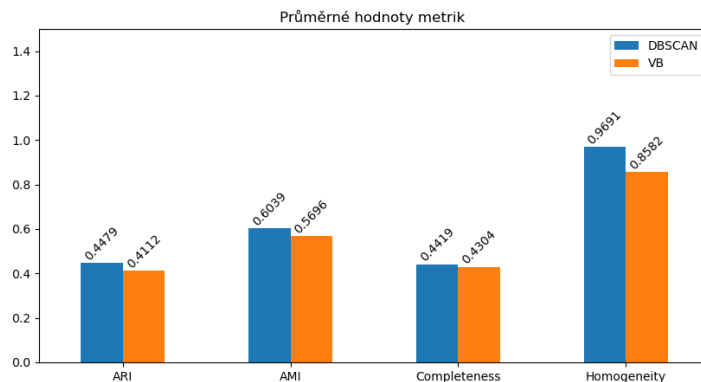
Scénáře deváté sady dat obsahují dva emiterů, kde oba emiterů mají modulaci stagger jak v RF, tak i PW parametru.

Deinterleaving deváté sady dat byl proveden pomocí DBSCAN a VB metod pro všechny scénáře. Vyhodnocení úspěšnosti bylo provedeno pomocí metrik z podkapitoly 5.1. Histogramy hodnot metrik pro obě metody jsou zobrazeny na obrázku 5.25 a jejich průměrné hodnoty na obrázku 5.26. Odhadnutý počet emiterů pomocí metod DBSCAN a VB pro jednotlivé scénáře je na obrázku 5.27.

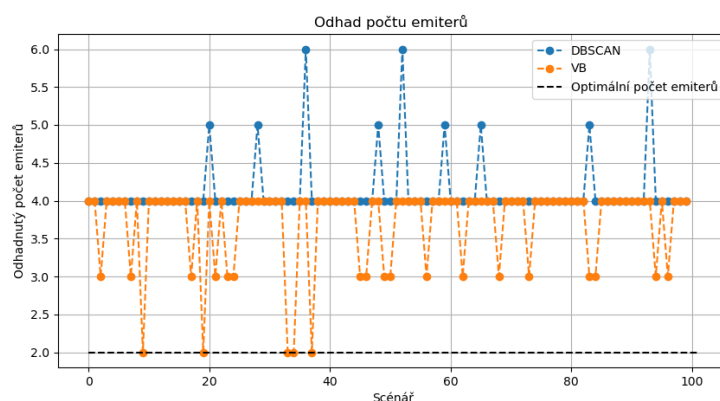


Obrázek 5.25: Histogramy metrik pro hodnocení úspěšnosti deinterleavingu sady dat 9

5.2. SIMULOVANÁ DATA



Obrázek 5.26: Odhad počtu emiterů pro všechny scénáře sady dat 9



Obrázek 5.27: Odhad počtu emiterů pro všechny scénáře sady dat 9

Deinterrelaving dvou emiterů, jenž mají stagger modulaci jak v RF tak i v PW parametru, je velmi obtížný. Obě metody vykazují hodnoty ARI menší jak 0.5. Jak je vidět z obrázku 5.27 a z histogramu úplnosti 5.25 obě metody ve většině případů odhadují čtyři emitory. To znamená, že každou úroveň stagger modulace vyhodnotí jako samostatný shluk. V několika málo případech VB metoda dokáže odhadnout dva nebo tři emitory. To znamená, že metoda spojí dvě úrovně stagger modulace do jednoho emiteru. Z hodnot ARI a AMI je však patrné, že spojení je nesprávné (dojde ke spojení dvou úrovní stagger od různých emiterů).

5.3 Data z reálného provozu

Účinnost vybraných metod byla ověřena na datech z reálného provozu. U těchto dat není známo správné rozřídění impulzů k jednotlivým emiterům a tedy nelze využít metriky použité pro hodnocení účinnosti u simulovaných dat. V literatuře se pro vyhodnocení úspěšnosti shlukování při neznámém rozdělení dat využívá například Silhouette score [36] nebo Davies-Bouldin Index [8]. Tyto metriky porovnávají průměrnou podobnost prvků v jednom shluku a jejich rozdílnost k prvkům z jiných shluků. Pro úlohu deinterleavingu je tento přístup nevhodný. Impulzy jsou vyzařovány emitery, které ve většině případů pracují na úzkém PW nebo RF pásmu a také se vyskytují emitery s agilními parametry. Z tohoto důvodu jsou si shluky ve smyslu Silhouette score nebo Davies-Bouldin Index podobné (úzké PW nebo RF pásmo) nebo naopak jsou prvky z jednoho shluku odlišné (agilita). Tudíž je účinnost nízká a pro hodnocení deinterleavingu nic neříkající.

Hodnocení účinnosti deinterleavingu na datech z reálného provozu je prováděno subjektivním hodnocením. V této diplomové práci byl zvolen přístup vizuální kontroly všech parametrů popisující impulzy v závislosti na čase.

V některých případech lze identifikovat emiter ze závislosti parametru PA na čase, kde je zřetelný 'oblouk', způsobený zřejmě mechanickým pohybem antény emiteru. Dále je možné identifikovat emiter s konstantním RF či PW parametrem. Případnou agilitu v RF nebo PW, lze pak ověřit výpočtem odvozeného parametru PRI pro jednotlivé shluky. Tyto hodnoty by měli mít jednu z PRI modulací zmíněných v podkapitole 1.2.4.

Z průběhu RF, PW a PA parametrů dat z reálného provozu v závislosti na čase (Obrázek 5.28) lze usoudit, že se zde vykytují tyto emitery:

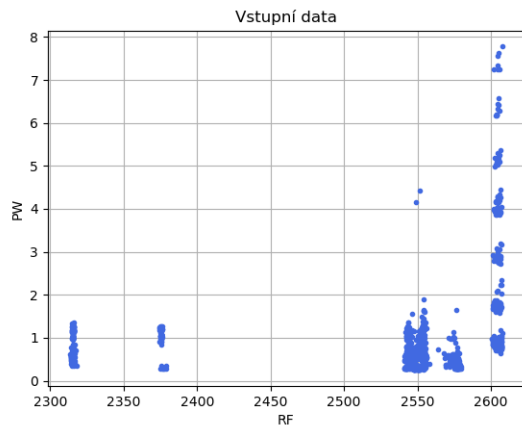
- emiter 1: RF ≈ 2550 MHz, PW $\approx 0.6\mu s$
- emiter 2: RF ≈ 2575 MHz, PW $\approx 0.4\mu s$
- emiter 3: RF ≈ 2375 MHz, PW $\approx 1.1\mu s$
- emiter 4: RF ≈ 2315 MHz, PW $\approx 0.9\mu s$
- emiter 5: RF ≈ 2605 MHz, se stagger nebo D&S modulací PW: PW $\approx 0.8 - 7.8\mu s$

Poznamenejme, že se jedná pouze o odhady parametrů signálu od emiterů z obrázku 5.28. Například je možné, že emitery 3 a 4 jsou jeden emiter se stagger nebo D&S modulací RF. Stejně tak emiter 5 může být sedm samostatných emiterů.

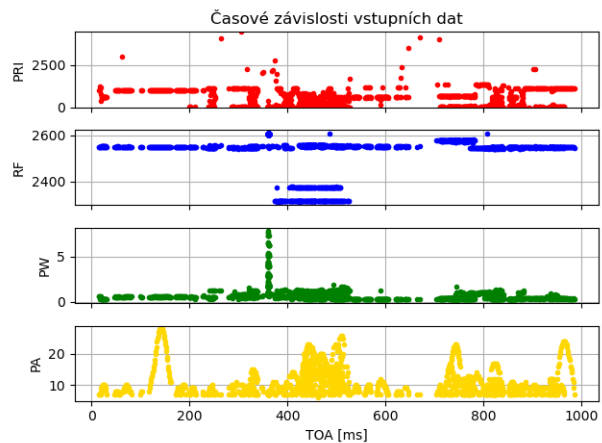
Deinterleaving byl proveden na datech zobrazených na obrázku 5.28. Vizualizace dat po deinterleavingu je na obrázcích 5.29 a 5.30, kde každý ze shluků je vykreslen jinou barvou a šum je vykreslen šedě. Jak můžeme vidět výsledky deinterleavingu pomocí vybraných metod jsou odlišné. Jak bylo popsáno u testování na simulovaných datech, obecně DBSCAN shluky rozděluje, zatímco VB metoda má tendenci shluky spojovat.

DBSCAN metoda rozdělila emiter 5 na sedm samostatných sluků. Stejně tak oddělila emitery 3 a 4. VB metoda naopak emitery 3 a 4 vyhodnotila jako jeden emiter. Také šest úrovní emiteru 5 vyhodnotila jako samostatný shluk. Dále si můžeme všimnout, že VB metoda určila navíc oproti DBSCAN další emiter (fialový). Z průběhu PA lze soudit, že emiter 1 opravdu obsahuje další emitery a měl by být rozdělen. Jedna z vlastností DBSCAN algoritmu je neschopnost oddělit shluky, které jsou spojené vzhledem k hustotě po daná ϵ -okolí. Z tohoto důvodu není DBSCAN schopen rozdělit emiter 1 na více emiterů.

5.3. DATA Z REÁLNÉHO PROVOZU

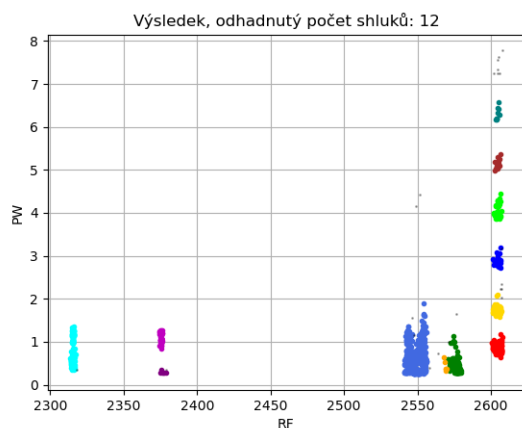


(a) Vstupní data v rovině RF a PW

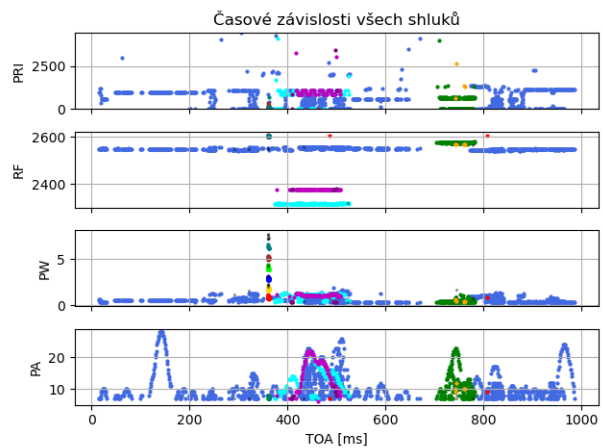


(b) Časová závislost vstupních dat

Obrázek 5.28: Vstupní data

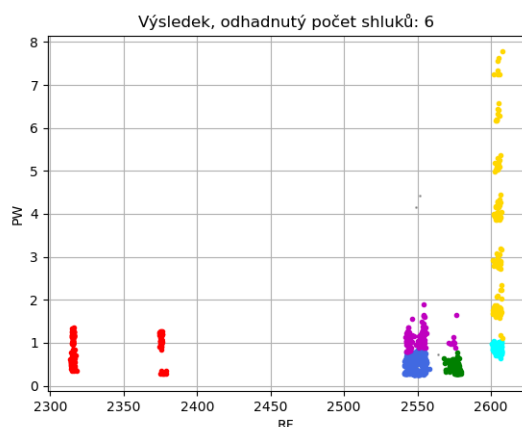


(a) Data v rovině RF a PW

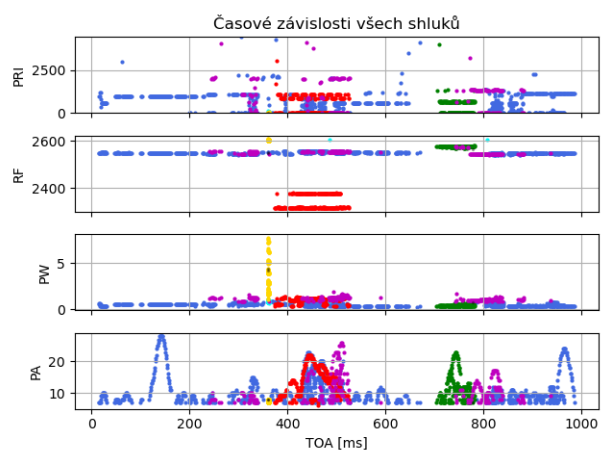


(b) Časová závislost dat

Obrázek 5.29: Vizualizace dat po deinterleavingu pomocí DBSCAN metody



(a) Data v rovině RF a PW



(b) Časová závislost dat

Obrázek 5.30: Vizualizace dat po deinterleavingu pomocí VB metody

6 Závěr

První část této diplomové práce se zabývala řešením používaných metod pro autonomní jednokanálový deinterleaving. Metody autonomního deinterleavingu můžeme rozdělit do dvou skupin. Na metody využívající pouze jeden parametr popisující přijatou sekvenci impulzů a na metody využívající dva a více parametrů.

Jednoparametrické metody deinterleavingu využívají pouze parametr čas příchodu impulzu (TOA). Mezi tyto metody se řadí například tzv. sekvenční algoritmy, metody využívající histogram, transformace nebo Kalmanova filtrace. Tyto zmíněné metody odhadují frekvenci, nebo hodnotu pulse repetition interval (PRI). Pomocí odhadnuté hodnoty PRI je poté separována odpovídající sekvence od jednoho zdroje. Tyto metody jsou velmi přesné, ale početně náročné. Dalším nedostatkem je jejich klesající přesnost při větším počtu přijatých impulzů, v případě šumu a chybějících impulzů. Metody využívající parametr TOA nedokáží odhadnout komplexnější modulaci PRI, jako je například D&S, wobulated či sliding PRI.

Z tohoto důvodu je vhodné pro deinterleaving využít více parametrů a separovat sekvence impulzů na základě podobnosti určitých parametrů namísto odhadu hodnoty PRI. Mezi takovéto metody řadíme především statistické metody (např. shluková analýza) nebo neuronové sítě.

Vzhledem k chování popsaných metod vůči zkreslujícím jevům (šum, chybějící impulzy, odlehlé impulzy), byly v druhé části práce vybrány dvě metody pro ověření jejich účinnosti na simulovaných datech a datech z reálného provozu. Jako vhodné metody byly zvoleny DBSCAN a VB metody.

DBSCAN je algoritmus založený na shlukování podle hustoty. Tyto algoritmy definují shluky jako oblasti s vysokou hustotou, které jsou rozděleny oblastí s nižší hustotou. Hlavní výhody těchto algoritmů je, že není potřeba znát předem počet shluků a jejich schopnost identifikovat shluky libovolných tvarů [20]. DBSCAN algoritmus definuje shluky jako maximální množinu prvků s dostatečně vysokou hustotou. Klasický DBSCAN algoritmus nechává volbu dvou parametrů na uživateli a to *MinPts* a ϵ . Avšak tyto dva parametry lze odhadnout z rozložení vstupních dat. Výhody tohoto algoritmu jsou schopnost automaticky určit počet shluků (emiterů) a identifikovat šum. Šum můžeme chápat jako impulzy, které nepocházejí z žádného emiteru (tzv. falešné impulzy), nebo impulzy, které pocházejí z emiteru, ale bylo přijato pouze malé množství impulzů, tudíž nejsou vyhodnoceny jako sekvence impulzů.

Při bayesovských metodách jsou parametry pravděpodobnostního rozdělení θ chápány jako náhodné veličiny. Při odhadu těchto parametrů se využívá předběžné znalosti o těchto parametrech a informace ze vstupních dat. Bayesovské metody vycházejí z Bayesovy věty, resp. ze vztahu mezi apriorní a aposteriorní pravděpodobností. Na sekvenci impulzů od více emiterů můžeme pohlížet jako na náhodný výběr X_1, \dots, X_N ze směsi K J -rozměrných normálních rozdělení s různými parametry $\theta_i, i = 1, \dots, K$ [2]. Na úlohu deinterleavingu pak pohlížíme jako na úlohu nalezení aposteriorní pravděpodobnosti $p(\theta_1, \dots, \theta_K | \mathbf{x}_1, \dots, \mathbf{x}_N)$. Nalezení aposteriorní pravděpodobnosti lze provést pomocí variačních metod nebo pomocí metody Monte Carlo. V této práci jsme se zabývali variačními metodami. Variační metody aproximují aposteriorní rozdělení s hustotou $p(\theta_1, \dots, \theta_K | \mathbf{x}_1, \dots, \mathbf{x}_N)$ variačním rozdělením s hustotou $q(\theta_1, \dots, \theta_K)$. Hledáme takové variační rozdělení, které je aposteriornímu v jistém smyslu co nejpodobnější. Podobnost dvou hustot rozdělení je pak měřena pomocí Kullback-Lieblerovy konvergence. Úloha

určení aposteriorní pravděpodobnosti je pak optimalizační úloha pro nalezení minima Kullback-Lieblerovy konvergence. Minimalizace Kullback-Lieblerovy konvergence je ekvivalentní s maximalizací tzv. dolní meze.

V této práci byl využit pro úlohu deinterleavingu GMM, který navíc zahrnoval odlehlé hodnoty, reprezentující nepřesnosti měření a možnou agilitu v parametrech. Iterační vztahy pro výpočet parametrů variačního rozdělení výsledného GMM (4.41) nejsou v literatuře uváděny. Odvození těchto vztahů bylo provedeno v podkapitole 4.2.

VB metoda je schopna automaticky určit počet shluků, za předpokladu, že vstupní parametr K je větší jak skutečný počet shluků. Zadáním většího počtu shluků metoda přiřadí automaticky některým ze shluků váhu a_k blízko nule a tím dojde k automatickému výběru počtu shluků. VB metoda na rozdíl od DBSCAN nedokáže identifikovat šum. Lze pouze zadat podmínku, že výsledný emiter musí mít minimální počet impulzů. Pokud tomu tak není jsou tyto impulzy vyhodnoceny jako šum.

Autoři pojednávající o metodách využívaných pro úlohu deinterleavingu provádějí mnohdy simulace pouze na ideálních datech, které neobsahují žádné zkreslující jevy a obsahují pouze parametry s konstantní modulací. Tato práce zkoumá účinnost vybraných metod na simulovaných datech, které obsahují rušivé vlivy a tedy odpovídají reálným hodnotám.

Účinnost vybraných metod byla testována na simulovaných datech a datech z reálného provozu. Pro hodnocení účinnosti deinterleavingu byly zvoleny metriky pro hodnocení účinnosti shlukování. Jako vhodné metriky byly vybrány ARI, AMI, stejnorodost a úplnost. Deinterleaving byl proveden pomocí DBSCAN a VB metody, které byly implementované v programovacím jazyce Python na devíti sadách dat, kde každá z nich obsahovala sto náhodně vygenerovaných scénářů. Scénáře obsahovaly data od čtyř nebo dvou emiterů. Jednotlivé sady dat se lišily agilitou neboli modulací RF a PW parametru.

Z provedeného experimentu je patrné, že vybrané metody mají podobnou účinnost pro emitery s konstantními parametry. Dále z experimentu vyplývá, že VB metoda je pro určité případy vhodnější, především pro výskyt emiterů s agilními parametry.

DBSCAN definuje shluky jako množiny s dostatečnou hustotou, jejichž prvky jsou spojeny vzhledem k hustotě. Tato definice shluků znemožňuje rozpoznání emiterů s agilitou D&S, stagger a jitter s velký rozptylem. V tomto případě se jako lepší jeví VB metoda. Model VB použitý v této diplomové práci zahrnuje odlehlé impulzy, které můžeme chápat jako impulzy od emiteru s agilitou v parametrech. Z tohoto důvodu je VB metoda účinnější. Avšak i tato metoda má pro data s agilitou svá omezení. Pokud má modulace příliš velký zdvih ani VB metoda nedokáže tuto modulaci identifikovat. Další faktorem ovlivňujícím účinnost deinterleavingu pomocí vybraných metod je počet emiterů. Čím větší je počet emiterů tím se deinterleaving ztíží, v důsledku podobnosti parametrů.

Literatura

- [1] ALDOSSARY, Mohammad. De-interleaving of Radar Pulses for EW Receivers with an ELINT Application. Cape Town, 2017. Disertační práce. Department of Electrical Engineering at the University of Cape Town.
- [2] ANDĚL, Jiří. *Základy matematické statistiky*. Vyd. 3. Praha: Matfyzpress, 2011. ISBN 978-80-7378-162-0.
- [3] ATA'A, A.W. a S.N. ABDULLAH. *Deinterleaving of radar signals and PRF identification algorithms*. 2007, 1(5), 340-347. DOI: 10.1049/iet-rsn:20070037. ISSN 17518784. Dostupné také z: <https://digital-library.theiet.org/content/journals/10.1049/iet-rsn-20070037>
- [4] ASLAN, Mehmet Kadir. *Emitter identification techniques in electronic warfare*. 2006. Disertační práce. Graduate School of Natural and Applied Sciences of middle east technical university.
- [5] BEZOUŠEK, Pavel a Pavel ŠEDIVÝ. *Radarová technika*. Praha: Vydavatelství ČVUT, 2004. ISBN 80-010-3036-9.
- [6] BISHOP, Christopher M. *Pattern recognition and machine learning*. [New York]: Springer, 2006. Information science and statistics. ISBN 03-873-1073-8.
- [7] BLEI, David M., Alp KUCUKELBIR a Jon D. MCAULIFFE. *Variational Inference: A Review for Statisticians*. Journal of the American Statistical Association. 2017, 112(518), 859-877. DOI: 10.1080/01621459.2017.1285773. ISSN 0162-1459. Dostupné také z: <https://www.tandfonline.com/doi/full/10.1080/01621459.2017.1285773>
- [8] DAVIES, David L. a Donald W. BOULDIN. *A Cluster Separation Measure*. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1979, PAMI-1(2), 224-227. ISSN 0162-8828. Dostupné z: doi:10.1109/TPAMI.1979.4766909
- [9] DRISCOLL, D.E. a S.D. HOWARD. *The detection of radar pulse sequences by means of a continuous wavelet transform*. 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258). IEEE, 1999, 1999, 1389-1392 vol.3. DOI: 10.1109/ICASSP.1999.756240. ISBN 0-7803-5041-3. Dostupné také z: <http://ieeexplore.ieee.org/document/756240/>
- [10] ERDOGAN, Alex a Kiran GEORGE. *Deinterleaving Radar Pulse Train Using Neural Networks*. 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC). IEEE, 2019, 2019, 141-147. DOI: 10.1109/CSE/EUC.2019.00036. ISBN 978-1-7281-1664-8. Dostupné také z: <https://ieeexplore.ieee.org/document/8919517/>
- [11] ESTER, Martin, Hans-Peter KRIEGEL, Jörg SANDER a Xiaowei XU. *A density-based algorithm for discovering clusters in large spatial databases with noise*. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96). AAAI Press, 1996, 226-231.

LITERATURA

- [12] FINK, Daniel. *A Compendium of Conjugate Priors*. Bozeman, 1997. Montana State Univeristy.
- [13] GENÇOL, Kenan. *New Methods for radar emitter identification*. 2015. Disertační práce. Anadolu University.
- [14] GENÇOL, Kenan, Ali KARA a Nuray AT. *Improvements on deinterleaving of radar pulses in dynamically varying signal environments*. Digital Signal Processing. 2017, 69, 86-93. DOI: 10.1016/j.dsp.2017.06.010. ISSN 10512004. Dostupné také z: <https://linkinghub.elsevier.com/retrieve/pii/S1051200417301240>
- [15] GUO, Qiang, Xingzhou ZHANG a Zheng LI. *SVC&K-Means and Type-Entropy Based DeInterleaving/Recognition System of Radar Pulses*. 2006 IEEE International Conference on Information Acquisition. IEEE, 2006, 2006, 742-747. DOI: 10.1109/ICIA.2006.305821. ISBN 1-4244-0528-9. Dostupné také z: <http://ieeexplore.ieee.org/document/4097754/>
- [16] JIANG, Wen, Xiongjun FU a Jiayun CHANG. *Improved de-interleaving algorithm of radar pulses based on dual fuzzy vigilance ART*. Journal of Systems Engineering and Electronics. 2020, 31(2), 303-311. DOI: 10.23919/JSEE.2020.000008. ISSN 1004-4132. Dostupné také z: <https://ieeexplore.ieee.org/document/9082309/>
- [17] KOCAMIS, Mehmet Burak, Hakan ABACI, Safak Bilgi AKDEMIR, Sertan VARMA a Alper YILDIRIM. *Deinterleaving for radar warning receivers with missed pulse consideration*. 2016 European Radar Conference (EuRAD). London, 2016, 225-228.
- [18] KREIDL, Marcel a Radislav ŠMÍD. *Technická diagnostika: senzory, metody, analýza signálu*. Praha: BEN - technická literatura, 2006. Senzory neelektrických veličin. ISBN 80-730-0158-6.
- [19] LIU, Yuhao, Yu CHEN a Songlin SUN. *A radar signal sorting algorithm based on PRI*. 2019 19th International Symposium on Communications and Information Technologies (ISCIT). IEEE, 2019, 2019, 144-149. DOI: 10.1109/ISCIT.2019.8905239. ISBN 978-1-7281-5009-3. Dostupné také z: <https://ieeexplore.ieee.org/document/8905239/>
- [20] LIU, Ying a Hongyuan CUI. *A novel self-adaptive density-based clustering algorithm for radar signal sorting*. IET International Radar Conference 2015. Institution of Engineering and Technology, 2015, 2015, , 7.-7. ISBN 978-1-78561-038-7. Dostupné z: doi:10.1049/cp.2015.1183
- [21] LIU, Jingyao, Huadong MENG, Yimin LIU a Xiqin WANG. *Deinterleaving pulse trains in unconventional circumstances using multiple hypothesis tracking algorithm*. Signal Processing. 2010, 90(8), 2581-2593. DOI: 10.1016/j.sigpro.2010.03.004. ISSN 01651684. Dostupné také z: <https://linkinghub.elsevier.com/retrieve/pii/S0165168410000897>
- [22] LIU, Jingyao, Huadong MENG a Xiqin WANG. *A new pulse deinterleaving algorithm based on Multiple Hypothesis Tracking*. 2009 International Radar Conference "Surveillance for a Safer World" (RADAR 2009). 2009, 1-4. ISSN 1097-5764.

- [23] LIU, Yanchao a Qunying ZHANG. *Improved method for deinterleaving radar signals and estimating PRI values*. 2018, 12(5), 506-514. DOI: 10.1049/iet-rsn.2017.0516. ISSN 1751-8784. Dostupné také z: <https://digital-library.theiet.org/content/journals/10.1049/iet-rsn.2017.0516>
- [24] MAHMOD, Shad. *Deinterleaving pulse trains with DBSCAN and FART*. 2019. Diplomová práce. Uppsala Universitet.
- [25] MANICKCHAND, Kaveer. *Multiple Radar Environment Emission Deinterleaving and PRI Prediction*. Cape Town, 2017. Diplomová práce. Department of Electrical Engineering at the University of Cape Town.
- [26] MAO, Yan, Jun HAN, Guohua GUO a Xu QING. *An Improved Algorithm of PRI Transform*. 2009 WRI Global Congress on Intelligent Systems. IEEE, 2009, 145-149. DOI: 10.1109/GCIS.2009.313. ISBN 978-0-7695-3571-5. Dostupné také z: <http://ieeexplore.ieee.org/document/5209181/>
- [27] MARDIA, H.K. *New techniques for the deinterleaving off repetitive sequences*. IEE PROCEEDINGS. 1989, 136(4), 149-154.
- [28] *dbscan: Density-based spatial clustering of applications with noise (DBSCAN)*. The MathWorks, Inc [online]. [cit. 2020-10-19]. Dostupné z: <https://www.mathworks.com/help/stats/dbscan.html>
- [29] MILOJEVIC, D.J. a B.M. POPOVIC. *Improved algorithm for the deinterleaving of radar pulses*. IEE Proceedings F Radar and Signal Processing. 1992, 139(1). DOI: 10.1049/ip-f-2.1992.0012. ISSN 0956375X. Dostupné také z: <https://digital-library.theiet.org/content/journals/10.1049/ip-f-2.1992.0012>
- [30] MOORE, John B. a Vikram KRISHNAMURTHY. *Deinterleaving pulse trains using discrete-time stochastic dynamic-linear models*. IEEE Transactions on Signal Processing. 1994, 42(11), 3092-3103. DOI: 10.1109/78.330369. ISSN 1053587X. Dostupné také z: <http://ieeexplore.ieee.org/document/330369/>
- [31] NISHIGUCHI, Ken'ichi a Masaaki KOBAYASHI. *Improved algorithm for estimating pulse repetition intervals*. IEEE Transactions on Aerospace and Electronic Systems. 2000, 36(2), 407-421. DOI: 10.1109/78.845217. ISSN 00189251. Dostupné také z: <http://ieeexplore.ieee.org/document/845217/>
- [32] ORSI, Robert J., John B. MOORE a Robert E. MAHONY. *Spectrum estimation of interleaved pulse trains*. IEEE Transactions on Signal Processing. 1999, 47(6), 1646-1653. DOI: 10.1109/78.765135. ISSN 1053587X. Dostupné také z: <http://ieeexplore.ieee.org/document/765135/>
- [33] PENG, Roger D. *Advanced Statistical Computing* [online]. 2021 [cit. 2021-5-8]. Dostupné z: <https://bookdown.org/rdpeng/advstatcomp/>
- [34] REVILLON, Guillaume, Ali MOHAMMAD-DJAFARI a Cyrille ENDERLI. *Radar emitters classification and clustering with a scale mixture of normal distributions*. 2019, 13(1), 128-138. DOI: 10.1049/iet-rsn.2018.5202. ISSN 1751-8784. Dostupné také z: <https://digital-library.theiet.org/content/journals/10.1049/iet-rsn.2018.5202>

- [35] ROSENBERG, Andrew a Julia HIRSCHBERG. *V-Measure: A conditional entropy-based external cluster evaluation measure*. Department of Computer Science Columbia University. New York, 2007, , 410-420.
- [36] ROUSSEEUW, Peter J. *Silhouettes: A graphical aid to the interpretation and validation of cluster analysis*. Journal of Computational and Applied Mathematics. 1987, 53-65. ISSN 0377-0427.
- [37] SANDER, Jörg, Martin ESTER, Hans-Peter KRIEGEL a Xiaowei XU. *Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and its Applications*. Data Mining and Knowledge Discovery. 1998, 2(2), 169-194. ISSN 13845810. Dostupné z: doi:10.1023/A:1009745219419
- [38] SHENG, Yang, Changbo HOU a Weijian SI. *Extract pulse clustering in radar signal sorting*. 2017 International Applied Computational Electromagnetics Society Symposium - Italy (ACES). IEEE, 2017, 2017, 1-2. DOI: 10.23919/ROPACES.2017.7916027. ISBN 978-0-9960078-3-2. Dostupné také z: <http://ieeexplore.ieee.org/document/7916027/>
- [39] TIERNEY, Luke a Joseph B. KADANE. *Accurate approximations for posterior moments and marginal densities*. Journal of the American statistical association. 1986, 393(81), 82-86.
- [40] VINH, Nguyen Xuan, Julien EPPS a James BAILEY. *Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance*. Journal of Machine Learning Research. 2010, 2010(11), 2838-2854.
- [41] WILEY, Richard G. PRI Analysis and Deinterleaving. Slideshare [online]. Research Associates of Syracuse, 2013 [cit. 2020-10-02]. Dostupné z: <https://www.slideshare.net/RohdeSchwarzNA/pri-and-deinterleaving-aoc-course-2013-2-per-page>
- [42] XI, Yin, Xiongjun WU, Yingchun WU a Lei DENG. *A Fast and Real-time PRI Transform Algorithm for Deinterleaving Large PRI Jitter Signals*. 2018 37th Chinese Control Conference (CCC). IEEE, 2018, 2018, 4465-4469. DOI: 10.23919/ChiCC.2018.8483497. ISBN 978-988-15639-5-8. Dostupné také z: <https://ieeexplore.ieee.org/document/8483497/>
- [43] WANG, Zhanling, Dengfu ZHANG, Duyan BI a Shiqiang WANG. *Multiple-Parameter Radar Signal Sorting Using Support Vector Clustering and Similarity Entropy Index*. Circuits, Systems, and Signal Processing. 2014, 33(6), 1985-1996. DOI: 10.1007/s00034-013-9721-1. ISSN 0278-081X. Dostupné také z: <http://link.springer.com/10.1007/s00034-013-9721-1>

Seznam použitých zkratek a symbolů

Seznam použitých zkratek

AMI	Adjusted mutual information
ARI	Adjusted rand index
CDIF	Cumulative difference histogram
D&S	Dwell and switch (blokový stagger)
DBSCAN	Density-based spatial clustering of applications with noise
DFT	Diskrétní Fourierova transformace
ESM	Electronic Support Measures
FDP	Forward dynamic programming
GMM	Gaussův smíšený model
MI	Mutual information
NN	Neuronové síť
PA	Amplituda
PRI	Pulse repetition interval
PW	Šířka impulzu
RF	Rádiová frekvence
RI	Rand Index
RL	Radiolokační
SDIF	Sequential difference histogram
SOM	Self-organizing map
SVM	Support vector machine
TOA	Čas příchodu signálu
VB	Variační Bayes

Seznam použitých symbolů

∇	Gradient
\approx	Je přibližně rovno
\sim	Má rozdělení

SEZNAM POUŽITÝCH ZKRATEK A SYMBOLŮ

\propto	Rovnost až na multiplikační konstantu
\otimes	Součin nezávislých rozdělení, tj. výsledná hustota je rovna součinu těchto hustot
$\mathcal{B}(\cdot, \cdot)$	Beta funkce
$\psi(\cdot)$	Digamma funkce
$\psi_J(\cdot)$	J -rozměrná digamma funkce definovaná vztahem $\psi_J(a) = \psi(\sum_{i=1}^J a + \frac{1-i}{2})$
$\Gamma(\cdot)$	Gamma funkce
$\Gamma_J(\cdot)$	J -rozměrná gamma funkce
$\delta(\cdot)$	Diracovo delta
δ_i^j	Kroneckerovo delta
$\ \cdot\ $	Euklidovská norma
$\mathcal{L}(p(\mathbf{x}))$	Dolní mez
$KL(p(\mathbf{x}) q(\mathbf{x}))$	Kullback-Lieblerova divergence
$\text{tr}(\cdot)$	Stopa matice
$E_q[\cdot]$	Střední hodnota vzhledem k hustotě q
$d(\cdot, \cdot)$	Vzdálenost dvou bodů (euklidovská, manhattanovská atd.)
$O(\cdot)$	Časová náročnost
$\mathcal{D}(d_1, \dots, d_K)$	Dirichletovo rozdělení s parametry d_1, \dots, d_K
$\mathcal{G}(\alpha, \beta)$	Gamma rozdělení s parametry α, β
$\mathcal{W}^{-1}(\gamma, \bar{\Sigma})$	Inverzní Wishartovo rozdělení s parametry $\gamma, \bar{\Sigma}$
$\mathcal{N}_J(\boldsymbol{\mu}, \Sigma)$	J -rozměrné normální rozdělení s parametry $\boldsymbol{\mu}, \Sigma$
$\mathcal{N}_J\mathcal{W}^{-1}(\boldsymbol{\mu}, \Sigma, \gamma, \bar{\Sigma})$	J -rozměrné normální-inverzní Wishartovo rozdělení s parametry $\boldsymbol{\mu}, \Sigma, \gamma, \bar{\Sigma}$
$\mathcal{CAT}(\mathbf{a})$	Kategorické rozdělení s parametrem \mathbf{a}
$\mathcal{G}(\mathbf{x} \alpha, \beta)$	Hustota gamma rozdělení s parametry α, β
$\mathcal{N}_j(\mathbf{x} \boldsymbol{\mu}, \Sigma)$	Hustota J -rozměrného normální rozdělení s parametry $\boldsymbol{\mu}, \Sigma$