PRIMENA RETKE REPREZENTACIJE NA MODELIMA GAUSOVIH MEŠAVINA KOJI SE KORISTE ZA AUTOMATSKO PREPOZNAVANJE GOVORA

NIKŠA JAKOVLJEVIĆ

Doktorska disertacija

19.11.2013 – version 2.0

УНИВЕРЗИТЕТ У НОВОМ САДУ ● **ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА**
21000 НОВИ САД, Трг Доситеја Обрадовића 6

КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА

Редни број, **РБР**:		
Идентификациони број, **ИБР**:		
Тип документације, **ТД**:		Монографска документација
Тип записа, **ТЗ**:		Текстуални штампани материјал
Врста рада, **ВР**:		Докторска дисертација
Аутор, **АУ**:		Никша Јаковељевић
Ментор, **МН**:		Проф. др Владо Делић
Наслов рада, **НР**:		Примена ретке репрезентације на моделима Гаусових мешавина који се користе за аутоматско препознавање говора
Језик публикације, **ЈП**:		Српски (латиница)
Језик извода, **ЈИ**:		Српски / енглески
Земља публиковања, **ЗП**:		Србија
Уже географско подручје, **УГП**:		Војводина
Година, **ГО**:		2013
Издавач, **ИЗ**:		Ауторски репринт
Место и адреса, **МА**:		Нови Сад, Трг Доситеја Обрадовића 6
Физички опис рада, **ФО**: (поглавља/страна/ цитата/табела/слика/графика/прилога)		7 поглавља / 100 страна/ 21 слика / 14 табела / 61 референца / 1 прилог
Научна област, **НО**:		Техничко технолошке науке
Научна дисциплина, **НД**:		Телекомуникације и обрада сигнала
Предметна одредница/Кqучне речи, **ПО**:		Препознавање говора, модели Гаусових мешавина, ретка репрезентација
УДК		
Чува се, **ЧУ**:		Библиотека Факултета техничких наука
Важна напомена, **ВН**:		
Извод, **ИЗ**:		У овој дисертацији је представљен модел који апроксимира пуне коваријансне матрице у моделу гаусових мешавина (ГММ) са смањеним бројем параметара и израчунавања који су потребни за израчунавање изгледности. У предложеном моделу инверзне коваријансне матрице су апроксимиране коришћењем ретке репрезентације њихових карактеристичних вектора. Поред самог модела приказан је и алгоритам за естимацију параметара заснован на критеријуму максимизације изгелдности. Експериментални резултати на проблему препознавања говора су показали да предложени модел за исти ниво грешке као ГММ са упуним коваријансним, редукује број параметара за 45%.
Датум прихватања теме, **ДП**:		29.04.2013.
Датум одбране, **ДО**:		
Чланови комисије, **КО**:	Председник:	Др Жељен Трповски (ванредни професор)
	Члан:	Др Слободан Јовичић (редовни професор)
	Члан:	Др Татјана Грбић (доцент)
	Члан:	Др Милан Сечујски (доцент)
	Члан, ментор:	Др Владо Делић (редовни професор)

Потпис ментора

UNIVERSITY OF NOVI SAD ● **FACULTY OF TECHNICAL SCIENCES**
21000 NOVI SAD, Trg Dositeja Obradovića 6

KEY WORDS DOCUMENTATION

Accession number, **ANO**:		
Identification number, **INO**:		
Document type, **DT**:		Monograph documentation
Type of record, **TR**:		Textual printed material
Contents code, **CC**:		PhD thesis
Author, **AU**:		Nikša Jakovljević
Mentor, **MN**:		PhD Vlado Delić
Title, **TI**:		An application of sparse representation in Gaussian mixture models used in speech recognition task
Language of text, **LT**:		Serbian (Latin)
Language of abstract, **LA**:		Serbian / English
Country of publication, **CP**:		Serbija
Locality of publication, **LP**:		Vojvodina
Publication year, **PY**:		2013
Publisher, **PB**:		Author reprint
Publication place, **PP**:		Novi Sad
Physical description, **PD**: (chapters/pages/ref./tables/pictures/graphs/appendixes)		7 chapters / 100 pages / 21 figures / 14 tables / 61 references / 1 appendix
Scientific field, **SF**:		Technical and Technological Sciences
Scientific discipline, **SD**:		Telecommunications and Signal Processign
Subject/Key words, **S/KW**:		Speech recognition, Gaussian mixture models, sparse representation
UC		
Holding data, **HD**:		Library of Faculty of Technical Sciences
Note, **N**:		
Abstract, **AB**:		This thesis proposes a model which approximates full covariance matrices in Gaussian mixture models with a reduced number of parameters and computations required for likelihood evaluations. In the proposed model inverse covariance (precision) matrices are approximated using sparsely represented eigenvectors. A maximum likelihood algorithm for parameter estimation and its practical implementation are presented. Experimental results on a speech recognition task show that while keeping the word error rate close to the one obtained by GMMs with full covariance matrices, the proposed model can reduce the number of parameters by 45%.
Accepted by the Scientific Board on, **ASB**:		29.04.2013.
Defended on, **DE**:		
Defended Board, **DB**:	President:	PhD Željen Trpovski (associate professor)
	Member:	PhD Slobodan Jovičić (professor)
	Member:	PhD Tatjana Grbić (assistant professor)
	Member:	PhD Milan Sečujski (assistant professor)
	Member, Mentor:	PhD Vlado Delić (professor)

Menthor's sign

SAŽETAK

U automatskom prepoznavanju govora dominira statistički pristup koji je zasnovan na skrivenim Markovljevim modelima u kombinaciji sa modelom mešavina Gausovih raspodela. Skup parametara ovog modela čine: inicijalne verovatnoće stanja, verovatnoće prelaza između stanja, kao i težine, srednje vrednosti i kovarijansne matrice Gausovih raspodela. Problem koji je potrebno rešiti jeste kako obezbediti statistički efikasnu estimaciju parametara modela, prvenstveno kovarijansne matrice čiji je broj parametara srazmeran kvadratu dimenzije prostora obeležja kao i dovoljno brzo i tačno izračunavanje verovatnoća emitovanja opservacija. U ovom radu je predložen model koji aproksimira pune kovarijansne matrice smanjujući broj parametara potrebnih za opisivanje modela kao i broj računskih operacija potrebnih za izračunavanje verovatnoća emitovanja. U predloženom modelu inverzna kovarijansna matrica je aproksimirana korišćenjem retke reprezentacije karakterističnih vektora inverznih kovarijansnih matrica, odnosno svaki karakteristični vektor inverzne kovarijansne matrice je predstavljen kao linearna kombinacija malog broja vektora iz skupa vektora čija je kardinalnost nekoliko puta veća od kardinalnosti vektorskog prostora koji je potrebno opisati. Pored samog modela predstavljena je i procedura obuke zasnovana na maksimizaciji izglednosti, kako njene teorijske postavke tako i njena praktična realizacija. Testiranje samog modela, kao i nekoliko alternativnih modela je realizovano na zadatku kontinualnog prepoznavanja govora, na srpskom jeziku, nezavisnom od govornika, sa malim rečnikom (oko 250 reči), nezavisnim od gramatike (red reči je proizvoljan). Testovi su pokazali da model postiže tačnost prepoznavanja koja je približna tačnosti modela sa punim kovarijansnim matricama pri čemu je broj parametara redukovan za 45%. Iako je model formiran za prepoznavanje govora, može se iskoristiti i za druge oblasti u kojima se koriste mešavine Gausovih raspodela, gde je broj komponenata izuzetno veliki (nekoliko desetina hiljada).

ZAHVALNICA

Ova disertacija ne bi bila moguća bez prof. dr Vlade Delića, koji već duži niz godina inicira i vodi projekte iz oblasti govornih tehnologija, kao i cele AlfaNum grupe koji su obezbedili značajnu tehničku i finansijsku podršku mom radu. Posebno bih želeo da se zahvalim dr Marku Janevu i Radovanu Obradoviću, koji su mi ukazali na postojanje zanimljivog sveta retke reprezentacije signala, kao i Borislavu Antiću, dr Čedomiru Stefanoviću, dr Dejanu Vukobratoviću i prof. dr Vojinu Šenku s kojima sam imao zadovoljstvo da kroz nedeljne sastanke savladam estimaciju signala, što mi je puno pomoglo u mom radu. I na kraju želim da se zahvalim mojoj porodici na ljubavi, strpljenju i podršci koju mi pružaju u životu.

SADRŽAJ

1 UVOD 1

2 NAČINI MODELOVANJA KOVARIJANSNIH MATRICA 7
- 2.1 Uvod 7
- 2.2 Dijagonalna aproksimacija kovarijansne matrice 9
- 2.3 Blok dijagonalna aproksimacija kovarijansne matrice 11
- 2.4 Faktorska analiza 12
- 2.5 Aproksimacije inverznih kovarijansnih matrica 14
  - 2.5.1 Delimično povezane kovarijansne matrice – STC 16
  - 2.5.2 Proširena linearna transformacija zasnovana na maksimizaciji izglednosti – EMLLT 18
  - 2.5.3 Gausove mešavine sa inverznim kovarijansnim matricama u ograničenom potprostoru – PCGMM 19
  - 2.5.4 Faktorisane retke inverzne kovarijansne matrice 19
- 2.6 Rezime 20

3 RETKA REPREZENTACIJA 21
- 3.1 Uvod 21
- 3.2 Retko kodovanje 21
  - 3.2.1 Pohlepni algoritmi 22
  - 3.2.2 Relaksacija l_0-pseudo norme 27
  - 3.2.3 Kvalitet dobijenih aproksimacija 33
- 3.3 Formiranje rečnika 34
  - 3.3.1 Formiranje rečnika uz l_0 regularizaciju 34
  - 3.3.2 Formiranje rečnika uz l_1 regularizaciju 38
- 3.4 Određivanje rečnika i faktorizacija matrice 40
- 3.5 Rezime 41

4 RETKA REPREZENTACIJA INVERZNIH KOVARIJANSNIH MATRICA 43
- 4.1 Uvod 43
- 4.2 Opis modela 43
  - 4.2.1 Broj parametara 44
  - 4.2.2 Broj računskih operacija pri izračunavanju logaritma izglednosti 46
- 4.3 Estimacija parametara modela 49
  - 4.3.1 Procedura za etimaciju parametara modela 50
- 4.4 Opravdanost kriterijumske funkcije 52
- 4.5 Rezime 56

5 PREGLED KORIŠĆENIH GOVORNIH BAZA I SOFTVERSKIH ALATA 59
- 5.1 Uvod 59
- 5.2 Opis govornih baza 59
- 5.3 Opis korišćenih softverskih alata 60
  - 5.3.1 Ekstrakcija obeležja 61
  - 5.3.2 Način modelovanja 63
  - 5.3.3 Obuka sistema 68
  - 5.3.4 Dekodovanje 71
- 5.4 Rezime 73

6 REZULTATI 75
6.1 Uvod 75
6.2 (H)LDA 76
6.3 STC 82
6.4 SEGMM 83
6.5 Rezime 86
7 ZAKLJUCAK 89
A DODATAK 91
A.1 Disperzivnost dijagonalnih elemenata i karakterističnih vrednosti kovarijansne matrice 91
A.2 Optimalna estimacija rečnika 92
A.2.1 Gradijentna metoda 92
A.2.2 Lagranžov dualni problem 92

BIBLIOGRAFIJA 95

POPIS SLIKA

Slika 1 NIST evaluacija ASR sistema. Slika je preuzeta sa http://www.itl.nist.gov/iad/mig/publications/ASRhistory/ 2

Slika 2 Sa 'x' su označene opservacije koje predstavljaju realizaciju slučajne promenljive koja ima Gausovu raspodelu sa nultom srednjom vrednošću i kovarijansnom matricom: $\mathbf{\Sigma} = [0.5, 0.4; 0.4, 1]$. Na slici levo je dato poređenje kondicioniranosti estimiranih kovarijansnih matrica u slučaju dijagonalne (isprekidane linije) i pune kovarijansne matrice (pune linije) koje su estimirane na osnovu prikazanih 5 opservacija. Plavom elipsom označene su tačke u kojima stvarna funkcija gustine raspodele ima vrednost 0.02. Na slici desno je ilustrovana greška koja se čini u slučaju dijagonalne aproksimacije kovarijansne matrice. I na ovom grafiku su elipsama označene tačke u kojima estimirane funkcije gustina raspodele imaju vrednost 0.02 i to zelenom bojom u slučaju dijagonalne aproksimacije i crvenom u slučaju da se koristi puna kovarijansna matrica. 10

Slika 3 Uticaj dijagonalne aproksimacije kovarijansne matrice na Gausov klasifikator sa 2 klase. Plava klasa se može modelovati sa Gausovom raspodelom čija je srednja vrednost $\mathbf{\mu}_p = [0, -1]^T$, a varijansa $\mathbf{\Sigma}_p = [1.5, 0.4; 0.4, 1.0]$. Crvena klasa se može modelovati sa Gausovom raspodelom čija je srednja vrednost $\mathbf{\mu}_c = [0, 1]^T$, a varijansa $\mathbf{\Sigma}_c = [1.0, 0.6; 0.6, 1.0]$. Crnom linijom je prikazana granica između klasa. U ovom primeru, uvođenjem dijagonalne aproksimacije greška klasifikacije se povećala sa 14.5% na 18.5%. 10

Slika 4 Aproksimacija Gausove raspodele pomoću GMM kod kojih su kovarijansne matrice aproksimirane dijagonalnim. Na slici levo crvenom elipsom je predstavljena stvarna raspodela, a zelenim elipsama pojedinačne komponente koje čine GMM. Na slici desno predstavljena je raspodela koju daje GMM (ponderisana suma pojedinačnih raspodela). 11

Slika 5 Pseudo kod OMP algoritma. Redni broj iteracije je označen sa i i nalazi se u zagradama u eksponentu promenljive. Skup indeksa iskorišćenih atoma $S^{(i)}$ u indeksu retkog koda $\boldsymbol{\alpha}$ i rečnika $\mathbf{D}$ znači da taj vektor odnosno matrica sadrži samo elemente odnosno kolone sa tim indeksima. 23

Slika 6 Ilustracija OMP algoritma. Plavom bojom (crta-tačka linija) su prikazani svi atomi, crvenom (puna linija) signal, zelenom (puna linija) odstupanje (reziduum). Inicijalno odstupanje je jednako signalu, tako da je ono na slici 6b označeno crvenom bojom. Isprekidanim crvenim linijama su prikazani vektori $\alpha_3 \mathbf{d}_3$ i $\alpha_4 \mathbf{d}_4$, koji u zbiru daju rekonstrukciju signala $\mathbf{x}$. 25

Slika 7 Pseudo kod MP algoritma. Redni broj iteracije je označen sa i i nalazi se u zagradama u eksponentima promenljivih. Treba napomenuti da se vrednost $(\mathbf{d}_k^T \mathbf{r}^{(i-1)})/\|\mathbf{d}_k\|_2^2$ izračunava samo jedanput u jednoj iteraciji algoritma, ali za sve atome koji čine rečnik, kao da u svakoj novoj iteraciji ne mora da dođe do povećanja skupa korišćenih indeksa $S^{(i)}$. 26

Slika 8 Pseudo kod WMP algoritma. Redni broj iteracije je označen sa i i nalazi se u zagradama u eksponentima promenljivih. 27

Slika 9 Rešenje problema $\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_p$ tako da $[\frac{1}{\sqrt{3}}, 1]\boldsymbol{\alpha} = \frac{2}{\sqrt{3}}$, za različitie vrednosti p. 28

Slika 10 Izgled funkcije $|x|^p$ za različite vrednosti p. 29

Slika 11 Pseudo kod FOCUSS algoritma zasnovan na IRLS proceduri. Redni broj iteracije je označen sa i i nalazi se u zagradama u eksponentima promenljivih. Karakter '+' koji se nalazi u eksponentu promenljivih označava pseudo-inverziju, odnosno $\mathbf{X}^+ = \mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}$ 30

Slika 12 Pseudo kod LARS algoritma. Slovo C u eksponentu označava da je u pitanju komplement skupa. Oznaka skupa u indeksu promenljivih $\boldsymbol{\alpha}$ i $\mathbf{D}$ ukazuje da takav vektor odnosno matrica sadrže samo elemente odnosno kolone čiji su indeksi sadržani u tom skupu. Pretpostavka je da se u svakom koraku najviše jedan element, može razmeniti između skupova (S i S^C). Ovde je naveden samo jedan od nekoliko mogućih načina za izbor optimalnog $\boldsymbol{\alpha}$. 32

Slika 13 Pseudo kod MOD algoritma. Redni broj iteracije je označen sa k i nalazi se u zagradama u eksponentima promenljivih. Sa N_o je označen ukupan broj signala. U cilju skraćenja zapisa uzeto je da je $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_o}]$ i $\mathbf{A} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_{N_o}]$. 35

Slika 14 Pseudo kod K-SVD algoritma. U cilju skraćenja zapisa uzeto je da je $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_o}]$, $\mathbf{A} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_{N_o}]$, $\mathbf{A}_{j*}$ j-ta vrsta matrice $\mathbf{A}$ i $\mathbf{A}_{jS}$ j-ta vrsta matrice koja sadrži samo elemente na pozicijama koje su sadržane u skupu izabranih indeksa S. Indeks iteracije je naveden u eksponentu promenljive u okviru zagrada. Primetiti da se vrednost $\mathbf{A}$ izračunava 2 puta u okviru jedne iteracije 37

Slika 15 Pseudo kod za online određivanje rečnika. Redni broj iteracije je označen sa t i naveden je u zagradama u eksponentu promenljive. 40

Slika 16 Zavisnost maksimalnog broja atoma u rečniku K od kardinalnosti retkog koda d, u slučaju 26-dimenzionalnog prostora za različite vrednosti prosečnog broja Gausovih raspodela za koje se izračunava logaritam izglednosti po frejmu M, i različite nivoe redukcije r u odnosu na model sa punim kovarijansnim matricama. Nivo redukcije od r_0% znači da je ukupan broj operacija po jednom frejmu ograničen r_0% od broja operacija koje zahteva model sa punim kovarijansim matricama. 48

Slika 17 Procedura za estimaciju parametara SEGMM. Sa N_{em} je označen broj iteracija EM algoritma, sa N_{sp} broj iteracija za određivanje rečnika. 51

Slika 18 Blok šema ekstraktora obeležja. 62

Slika 19 Pseudo kod procedure kojom se formiraju stanja modela na osnovu inicijalnih modela koje je dala TBC na jednom govorniku i rastojanja između njih. Rezultujući skup HMM stanja je označen sa $\mathcal{C}$. 65

Slika 20 Uporedni prikaz WER-a u zavisnosti od broja izlaznih obeležja za različite vrste ulaznih obeležja. Značenje oznaka na graficima su sledeće: hlda d – varijanta HLDA koja pretpostavlja da su samo diskriminatorna obeležja nekorelisana, hlda – varijanta HLDA koja pretpostavlja da su sva transformisana obeležja nekorelisana, lda – varijanta LDA u kojoj se vrši normalizacija transformacione matrice i ref – WER GMM-a sa dijagonalnim kovarijansnim matricama i obeležjima 12MFCC_E_D. 81

Slika 21 Zavisnost prosečne vrednosti kvadratnog rastojanja (a) i KLD (b) od d za različite vrednosti k 85

POPIS TABELA

Tabela 1 Broj parametara za različite načine aproksimacije kovarijansnih matrica u modelima Gausovih mešavina. 46

Tabela 2 Broj računskih operacija sa pokretnim zarezom za različite načine aproksimacije kovarijansnih matrica u modelima Gausovih mešavina. Broj opseracija je prikazan po fazama: prva faza koja obuhvata izračunavanja sa deljenim parametrima (kolona Deljeni) i druga faza koja podrazumeva izračunavanje logaritma izglednosti za svaku od pojedinačnih Gausovih raspodela (kolona Specifični). 47

Tabela 3 Lista konsonanata i negovornih modela 67

Tabela 4 Lista vokala 68

Tabela 5 Performanse referentnih modela (broj Gausovih raspodela i učestanost grešaka na nivou reči) 75

Tabela 6 Performanse nenormalizovanih LDA sistema za različita ulazna obeležja i različit broj izlaznih (transformisanih) obeležja. 78

Tabela 7 Performanse normalizovanih LDA sistema za različita ulazna obeležja i različit broj izlaznih (transformisanih) obeležja. 79

Tabela 8 Performanse HLDA sistema uz pretpostavku da su samo diskriminativna obeležja nekorelisana za različita ulazna obeležja i različit broj izlaznih (transformisanih) obeležja. 80

Tabela 9 Performanse HLDA sistema uz pretpostavku da su samo diskriminativna obeležja nekorelisana za različita ulazna obeležja i različit broj izlaznih (transformisanih) obeležja. 80

Tabela 10 Performanse STC sistema. 83

Tabela 11 Prosečne vrednosti kvadratnih rastojanja između aproksimirane i stvarne vrednosti karakterističnih vektora kovarijansnih matrica za različite vrednost k i d. 84

Tabela 12 Prosečne vrednosti KLD između Gausovih raspodela sa estimiranom i aporkismiranom kovarijansnom matricom za različite vrednost k i d. 84

Tabela 13 Broj parametara modela za različite vrednost k i d. 86

Tabela 14 Vrednosti WER [%] za različite vrednost k i d. 86

POPIS SKRAĆENICA

ANN Veštačka neuralna mreža (*Artificial Neural Network*)

ASR Automatsko prepoznavanje govora (*Automatic speech recognition*)

CMN Normalizacija srednjom (prosečnom) vrednošću kepstrala (*Cepstral mean normalization*)

DCT Diskretna kosinusna transformacija (*Discrete cosine transformation*)

DFT Diskretna Furijeova transformacija (*Discrete Fourier transformation*)

EM Algoritam očekivanje-maksimizacija (*Expectation maximization*)

EMLLT Proširena linearna transformacija zasnovana na maksimizaciji izglednosti (*Extended maximum likelihood linear transformation*)

FAHMM Skriveni Markovljev model na nivou faktora (*Factor analysed hidden Markov model*)

FOCUSS Algoritam za usmereno rešavanje neodređenog sistema (*Focal underdetermined system solver*)

GMM Model Gausovih mešavina (*Gaussian mixture model*)

HLDA Heteroscedastička linearna diskriminativna analiza (*Heteroscedastic linear discriminative analysis*)

HMM Skriveni Markovljev model (*Hidden Markov model*)

IRLS Iterativna aproksimacija pomoću ponderisanih najmanjih kvadrata (*Iterative reweighted least square*)

KLD Kulbak-Lajbler divergenca (Kullback-Leibler divergence)

LARS Regresija najmanjih uglova (*Least angle regression*)

LASSO Operator najmanjeg apsolutnog sužavanja i selekcije (*Least absolute shrinkage and selection operator*)

LDA Linearna diskriminativna analiza (*Linear discriminant analysis*)

MFCC Mel-frekvencijski kepstralni koeficijent (*Mel-frequency cepstral coefficient*)

MLLT Linearna transformacija zasnovana na maksimizaciji izglednosti (*Maximum likelihood linear transformation*)

MLT Višestruke linearne transformacije (*Multiple linear transforms*)

MOD Metod optimalnih pravaca (*Method of optimal directions*)

MP Traženje poklapanja (*Matching pursuite*)

NIST *National Institute of Standards and Technology*

OMP Traženje poklapanja uz uslov ortogonalnosti (*Orthogonal matching pursuite*)

PCA Rastavljanje na osnovne komponente (*Principal component analysis*)

PCGMM Model Gausovih mešavina sa inverznim kovarijansnim matricama ograničenim u vektorskom potprostoru (*Precision constrained Gaussian mixture model*)

SCGMM Model Gausovih mešavina u ograničenom vektorskom potprostoru (*Subspace constrained Gaussian mixture model*)

SEGMM Model Gausovih mešavina sa inverznim kovarijansnim matricama modelovanim pomoću retke reprezentacije njihovih karakterističnih vektora (Sparse eigenvector Gaussian mixture model)

SLU Automatsko razumevanje govora (*Spoken language understanding*)

SPAM Gausove mešavine u ograničenim vektorskim potprostorima srednjih vrednosti i inverznih kovarijansnih matrica (*Subspace constrained precision and mean*)

SPLICE Deo-po-deo linearna kompenzacija okruženja bazirana na stereo signalu (*Stereo-based picewise linear compenstion for enviroments*)

STC Delimično povezane kovarijansne matrice (*Semi-tied covariances*)

SVD Dekompozicija na singularne vrednosti (*Singular value decomposition*)

TBC Klasterovanje na osnovu stabla (*Tree-based clustering*)

WER Učestanost grešaka na nivou reči (*Word error rate*)

WMP Traženje približnog poklapanja (*Weak matching pursuite*)

MATEMATIČKA NOTACIJA

Prilikom izbora matematičkih oznaka vodilo se računa o tome da oznake u samoj disertaciji budu konzistentne, ali i da budu usklađene sa oznakama u korišćenoj literaturi. Nažalost notacija u literaturi za iste stvari varira u zavisnosti od oblasti (obrada signala, mašinsko učenje), ali i od škole kojoj pripadaju autori. U ovom radu prednost su imale oznake koje se češće pojavljuju u literaturi, kao i one koje su bile primarni izvor za pojedina rešenja.

Da bi čitanje matematičkih izraza bilo jednostavnije matrice su označene podebljanim velikim slovima latiničnog ili grčkog alfabeta (npr. $\mathbf{A}$ i $\mathbf{\Sigma}$), vektori podebljanim malim slovima latiničnog ili grčkog alfabeta (npr. $\boldsymbol{\alpha}$ i $\mathbf{d}$), a skalari malim slovima latiničnog ili grčkog alfabeta. Pojedinačne kolone matrice su označene istim slovom kao matrica ali malim i podebljanim, dok oznaka koja se nalazi na poziciji indeksa označava redni broj kolone (npr. $\mathbf{d}_i$ označava i-tu kolonu matrice $\mathbf{D}$). Pojedinačne vrste matrice označene su na identičan način kao i sama matrica pri čemu se u indeksu nalazi oznaka vrste nakon koje sledi zvezdica (npr. $\mathbf{D}_{i*}$ označava i-tu vrstu matrice $\mathbf{D}$). Pojedinačni elementi matrice su označeni istim slovom kao i matrica pri čemu ono nije zadebljano, a par oznaka u indeksu predstavljaju njegovu poziciju u matrici (npr. D_{ij} predstavlja element matrice $\mathbf{D}$ u i-toj vrsti i j-toj koloni). Pojedinačni element vektora, koji ne predstavlja kolonu ili vrstu matrice je označen istim slovom kao i sam vektor pri čemu ono nije zadebljano, a u indeksu sadrži oznaku njegove pozicije u vektoru (npr. d_i predstavlja i-ti element vektora $\mathbf{d}$). Ukoliko se neka promenljiva izračunava iterativnom procedurom, njena vrednost u tekućoj iteraciji na poziciji eksponenta sadrži redni broj iteracije koji je naveden u običnim zagradama (npr. $\mathbf{d}_i^{(k)}$ označava vrednost i-te kolone matrice $\mathbf{D}$ u k-toj iteraciji). Pošto se karakteristike govora menjaju tokom vremena, ponekad je u opisu potrebno istaći tu zavisnost te će u tim slučajevima naziv promenljive u svom indeksu imati vrednost t (npr. ako $\mathbf{o}$ predstavlja opservaciju u proizvoljnom vremenskom trenutuku, $\mathbf{o}_t$ označava opservaciju u trenutuku t).

U nastavku je data lista svih korišćenih simbola.

$\mathbf{o}$	opservacija odnosno vektor obeležja
$\mathbf{o}_1^T$	sekvenca od T opservacija odnosno $(\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_T)$
D	dimenzionalnost vektora obeležja
M_s	broj komponenata koji čine Gausovu mešavinu pridruženu stanju s
M	ukupan broj Gausovih raspodela koje čine model
w_{sm}	težina m-te komponente Gausove mešavine koja je pridružena stanju s
$\boldsymbol{\mu}_{sm}$	srednja vrednost m-te komponente Gausove mešavine koja je pridružena stanju s
$\mathbf{\Sigma}_{sm}$	kovarijansna matrica m-te komponente Gausove mešavine koja je pridružena stanju s

$\mathbf{P}_{sm}$ inverzna kovarijansna matrica m-te komponente Gausove mešavine koja je pridružena stanju s ($\mathbf{P}_{sm} = \mathbf{\Sigma}_{sm}^{-1}$)

$\boldsymbol{V}$ matrica karakterističnih vrednosti

a_{s_1,s_2} verovatnoća prelaza iz HMM stanja s_1 u HMM stanje s_2

$\mathcal{M}^{hmm}$ skup parametara kojima je opisan skriveni Markovljev model (inicijalne verovatnoće, verovatnoće prelaza kao i težine, srednje vrednosti i kovarijansne matrice komponenti GMM-a)

$\mathbf{D}$ matrica koja sadrži bazne vektore u opštem slučaju odnosno atome u slučaju retke reprezentacije

$\boldsymbol{\alpha}$ redak kod koji sadrži koeficijente koji množe atome

K ukupan broj atoma koji čine rečnik

N_o ukupan broj signala/podataka koji je na raspolaganju za obuku

d maksimalna broj nenultih elemenata u retkom kodu

$\|\cdot\|_p$ l_p-norma ako je $p \geqslant 1$ definisana sa $\|\boldsymbol{x}\|_p = (\sum_i |x_i|^p)^{\frac{1}{p}}$

$\|\cdot\|_0$ l_0-pseudo norma definisana sa $\|\boldsymbol{x}\|_0 = \#\{x_i \neq 0\}$

$\|\cdot\|_F$ Frobeniusova norma koja je za $m \times n$ dimenzionalnu matricu definisana sa $\|\mathbf{X}\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} X_{ij}^2}$

$|\cdot|$ Determinanta ako je promenljiva matrica, odnosno apsolutna vrednost ako je u pitanju skalar

1 UVOD

Govor je osnovni oblik komunikacije između ljudi i kao takav će vrlo verovatno imati centralnu ulogu u budućim interfejsima za komunikaciju sa mašinama. Da bi mašina mogla da razume šta je čovek rekao prvo treba da iz govornog signala izdvoji reči, a potom iz tih reči izvuče informaciju o tome šta je rečeno. Prva faza, preslikavanje akustičkog signala u niz reči se naziva automatsko prepoznavanje govora (ASR, *Automatic Speech Recognition*), a druga faza, preslikavanje reči u značenje automatsko razumevanje govora (SLU, *Spoken Language Understanding*).[1] Za čoveka razumevanje govora čak i u slučaju da je govorni signal oštećen pozadinskom bukom ne predstavlja veći problem, jer pored informacije sadržane u samom govornom signalu, čovek pri razumevanju govora uzima u obzir i kontekst. Nažalost, razumevanje konteksta u kom se govor nalazi uglavnom se zasniva na opštem znanju o svetu i stoga predstavlja jedan od osnovnih problema za formiranje robustnog sistema za automatsko prepoznavanje odnosno razumevanje govora.

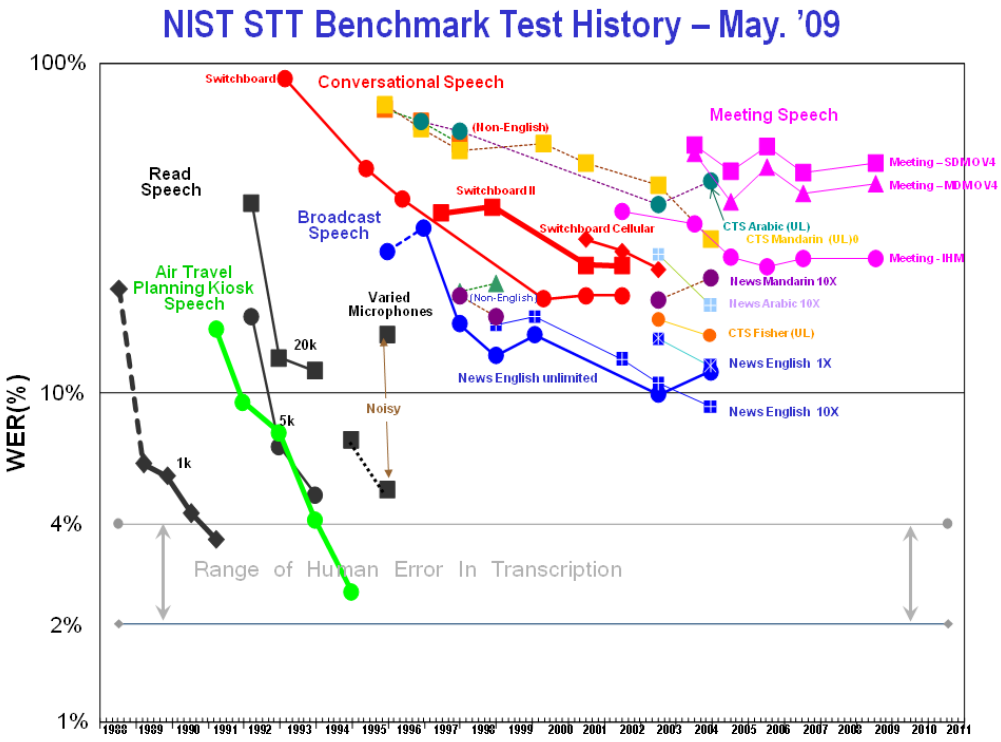
Iako je krajnji cilj razumevanje onoga što je rečeno, zbog složenosti problema koji treba da reši SLU, ASR se obično izdvaja kao posebna celina. Zadatak istraživanja u oblasti ASR-a jeste rešavanje problema brzog i ispravnog prepoznavanja šta je rekao proizvoljan govornik u proizvoljnim akustičkim uslovima na proizvoljnom jeziku. Ovaj krajnji cilj je nekoliko decenija daleko, ali je već sada moguće upotrebiti ASR sisteme u ograničenim i kontrolisanim uslovima, te tako postoje automatizovani pozivni centri, programi za diktiranje, personalni asistenti itd. (Pieraccini, 2012).

Postoji mnogo faktora koji određuju mogućnosti jednog ASR sistema kao što su: veličina rečnika izgovora,[2] prirodnost i tečnost govora, varijabilnost kanala i akustičkog okruženja kao i varijabilnost karakteristika govornika. Na slici 1 je prikazana učestanost greška na nivou reči (WER, *Word error rate*) za različite sisteme tokom godina koji su evaluirani od strane NIST-a (*National Institute of Standards and Technology*), sa koje je moguće sagledati zavisnost pojedinih faktora na tačnost sistema za prepoznavanje. Broj reči koje je potrebno prepoznati u velikoj meri utiče na tačnost ASR sistema pa tako ukoliko je potrebno prepoznati nekoliko reči (npr. sekvencu cifara) tačnost prepoznavanja je gotovo 100%, za sisteme sa malim rečnicima do nekoliko hiljada reči iznad 96%, dok je za sisteme sa velikim rečnicima koji broje nekoliko desetina hiljada reči i više tačnost manja od 90%. Na slici 1 je broj reči koji zahtevaju pojedini zadaci prikazana pored krivih koje prikazuju promenu WER-a tokom godina, dok za zadatke gde to nije navedeno (*Broadcast Speech*, *Conversational Speech* i *Meeting Speech*) podrazumeva se rečnik od nekoliko desetina hiljada reči.

1 U oblasti veštačke inteligencije jednu veliku oblast predstavlja razumevanje prirodnog jezika (NLU *Natural language understanding*), koja ima zadatak da na osnovu teksta zaključi šta je značenje (smisao) teksta. Podoblast koja se bavi govornim jezikom predstavlja SLU, pri čemu je zadatak značajno teži jer reči u tekstu ne moraju da odgovaraju stvarno izgovorenim rečima, a značenje pojedinih reči zavisi i od načina izgovora (nelingvističkih infromacija u govornom signalu).

2 Rečnik izgovora predstavlja listu reči koju sistem očekuje u govornom signalu.

Drugi faktor koji utiče na tačnost sistema za prepoznavanje jeste prirodnost i tečnost govora. Daleko je jednostavnije za jedan ASR sistem da prepoznaje izolovano izgovorene reči, gde su uzastopne reči razdvojene relativno dugom pauzom u govoru, nego kontinualni govor, gde su reči gotovo spojeno izgovarane. U okviru kontinualnog govora razlikuju se tri podvrste: *i*) čitani govor, *ii*) konverzacija čoveka sa mašinom i *iii*) konverzacija čoveka sa čovekom. Jasno je da je najteži oblik za automatsko prepoznavanje konverzacija čoveka sa čovekom pošto je brzina govora velika te ne postoji potpuna i ispravna artikulacija svih glasova (u ovu grupu spadaju *Conversational Speech* i *Meeting Speech* sa slike 1). Pri konverzaciji čoveka sa mašinom čovek se trudi da govori jasnije, artikulisanije i manje spontano (u ovu grupu spada *Air Travel Planning Kiosk Speech* sa slike 1), dok je čitani govor još jednostavniji pošto on podrazumeva i umeren tempo izgovora i glasnoću čime je dodatno smanjena varijabilnost izgovora fonema.



Slika 1: NIST evaluacija ASR sistema. Slika je preuzeta sa http://www.itl.nist.gov/iad/mig/publications/ASRhistory/

Varijabilnost kanala i akustičkog okruženja povećava varijabilnost izgovora glasova što povećava konfuziju između fonema čime se smanjuje tačnost prepoznavanja. Za razliku od čoveka koji ima sposobnost adaptacije na nove uslove, mogućnosti adaptacije sistema za prepoznavanje su ograničene te u situacijama kada postoje značajne rezlike u akustičkim karakteristikama u snimcima koji su korišćeni za obuku i onim koje sistem treba da prepoznaje dolazi do značajne degradacije tačnosti prepoznavanja. Pored varijabilnosti koju unose različiti kanali, propusni opseg kanala može biti uži od širine spektra govornog signala što može dovesti od gubitka informacija, pa tako u slučaju telefonskog kanala koji odseca praktično sve frekvencije iznad 4 kHz, značajan deo energije pojedinih frikativa i afrikata je odstranjen što otežava njihovo pre-

poznavanje. Sa druge strane prisustvo drugih izvora zvuka, značajno menja raspored akustičke energije u spektru maskirajući pojedine foneme dodatno smanjujući razlike između njih, što za rezultat ima smanjenje tačnosti prepoznavanja. Do sada su razvijene mnoge tehnike koje imaju za cilj redukciju ovih varijabilnosti kao što su: normalizacija srednjom vrednošću (CMN, *Cepstral Mean Normalization*) (Saon i Chien, 2012), deo-po-deo linearna kompenzacija okruženja bazirana na stereo signalu (SPLICE, *Stereo-based piecewise linear compensation for enviroments*) (Deng et al., 2001) i ekvalizacija na osnovu kvantila histograma (*Quantile-based Histogram Equalization*)(Hilger, 2004).

Poslednji faktor koji otežava prepoznavanje govora jesu različite individualne karakteristike govornika (fiziološke karakteristike ali i način artikulacije) koje povećavaju varijabilnost pojedinih fonema. Ovo se ogleda u činjenici da ASR sistemi koji su obučeni samo na jednog govornika, tzv. sistemi zavisni od govornika imaju daleko veću tačnost prepoznavanja (blizu 100%) od sistema koji su namenjeni većem broju govornika tzv. sistemima nezavisnim od govornika (Huang i Lee, 1993). Jedan od načina da se ovaj problem prevaziđe jeste pomoću normalizacije vokalnog trakta kojom se transformiše frekvencijska osa tako da odgovara karakteristikama referentnog govornika (Jakovljevič et al., 2009), ili da se parametri modela prilagode datom govorniku pomoću adaptacije na govornika (Gales i Woodland, 1996). Svi prehodno navedeni faktori su razlog, zašto i pored skoro 4 decenije intenzivnog istraživanja u oblastima ASR-a i SLU-a mogućnost prepoznavanja i razumevanja mašine je i dalje daleko manja od one koju ima čovek.

Sam problem ASR-a se može jednostavno definisati kao preslikavanje govornog signala $s(t)$ u odgovarajuću sekvencu reči odnosno:

$$w_1^{N_w} = f(s(t)) \tag{1}$$

gde je $w_1^{N_w}$ sekvenca od N_w reči $(w_1, w_2, \ldots, w_{N_w})$, a $f(\cdot)$ funkcija koja vrši odgovarajuće preslikavanje. Treba primetiti da se prepoznavanje govora definisano na ovaj način može tretirati kao specijalan slučaj prepoznavanja oblika. Standardni pristup prepoznavanju oblika je da se funkcija $f(\cdot)$ dekomponuje na dve od kojih jedna vrši tzv. izdvajanje (ekstrakciju) obeležja iz signala $(g(s(t)))$ i drugu koja vrši klasifikaciju na osnovu izdvojenih obeležja $(h(\cdot))$[3] odnosno:

$$w_1^{N_w} = h(g(s(t))) \tag{2}$$

Pod obeležjima se podrazumevaju karakteristike na osnovu kojih je moguće jednoznačno identifikovati pripadnost klasi, gde se pod klasom u slučaju prepoznavanja govora može podrazumevati fonem. Pošto govorni signal pored lingvističkih informacija nosi i paralingvističke informacije[4] iz govornog signala je neophodno izdvojiti samo karakteristike koje se odnose na to šta je rečeno, odnosno lingvističke informacije. Iz artikulacione fonetike je poznato

3 Ovakav pristup omogućuje da se algoritmi koji se koriste pri klasifikaciji učine relativno nezavisnim od konkretnog problema koji je potrebno rešiti, odnosno jednostavniju primenu postojećih rešenja na nove probleme. Sa druge strane izdvajanje obeležja podrazumeva postojanje znanja koja su specifična za problem odnosno koje su to informacije (obeležja) koja su bitna za klasifikaciju.

4 Pod lingvističkim informacijama u govornom signalu se podrazumevaju informacije o tome šta je rečeno, a pod paralingvističkim informacijama informacije o identitetu govornika (polu, starosti, poreklu, obrazovanju), njegovom emocionalnom i zdravstvenom stanju.

da na osnovu načina artikulacije, mesta tvorbe u slučaju konsonanata, odnosno položaja jezika u slučaju vokala i zvučnosti moguće jednoznačno odrediti sve foneme srpskog jezika, ali ta obeležja je relativno teško pouzdano izdvojiti iz akustičkog oblika govornog signala. S druge strane iz auditorne fonetike poznato je da je informacija o tome šta je rečeno prvenstveno sadržana u obvojnici spektra, stoga je razumno očekivati da bi oblik obvojnice odnosno parametri koji ga opisuju mogli poslužiti kao obeležja pri prepoznavanju oblika. Pošto je potrebno posmatrati spektar govornog signala, odnosno promene spektra tokom vremena, neophodno je izdeliti govorni signal na manje stacionarne segmente. Ovi segmenti se nazivaju frejmovi, a procedura kojom se izdvajaju prozoriranje. Trajanje frejmova ne sme da bude suviše kratko jer bi to vodilo gubitku frekvencijske rezolucije,[5] niti sme da bude suviše dugo jer će obuhvatiti nestacionarne segmente govornog signala, tako da je tipično trajanje frejma između 20 ms do 35 ms. Da bi se izbegla mogućnost da neki događaj ne bude detektovan, što se može desiti ukoliko bi se podelilo između dva susedna frejma tako da ni jedan frejm ne sadržava samo taj događaj, susedni frejmovi se međusobno preklapaju. Tipične vrednosti preklapanja su između 50% i 75% dužine frejma, što odgovara pomerajima frejma od 10 ms do 15 ms, pri čemu se obično ne ide na vrednosti pomeraja frejma koje su manje od 10 ms, jer se time povećava broj frejmova koje je potrebno obraditi, a dobitak po pitanju tačnosti je neznatan. Obeležja koja opisuju obvojnicu spektra i njene promene tokom vremena se izdvajaju za svaki frejm i kombinuju u vektor obeležja i čine jednu opservaciju, tako da funkcija $g(\cdot)$ signal preslikava u sekvencu od T opservacija ($\mathbf{o}_1^T$), odnosno:

$$\mathbf{o}_1^T = g(s(t)) \tag{3}$$

Nakon što se izaberu obeležja potrebno je izabrati i estimirati odgovarajuću funkciju koja će vršiti klasifikaciju na osnovu obeležja, odnosno sekvencu obeležja preslikavati u odgovarajuću sekvencu reči:

$$w_1^N = h\left(\mathbf{o}_1^T\right) \tag{4}$$

Pošto trajanje jednog fonema varira u zavisnosti od mnogo faktora, ovu vremensku varijabilnost je neophodno uzeti u obzir prilikom formiranja klasifikacione funkcije $h(\cdot)$. Jedno od najpodesnijih rešenja za modelovanje vremenske varijabilnosti jesu skriveni Markovljevi modeli (HMM, *Hidden Markov Model*), koji se veoma efikasno mogu ukomponavati sa drugim modelima koji opisuju akustičku varijabilnost klasa kao što su mešavine Gausovih raspodela (GMM, *Gaussian Mixture Model*) ili veštačke neuralne mreže (ANN, *Artificial Neural Networks*).

U ovom radu je za realizaciju klasifikacione funkcije iskoršćen HMM u kombinaciji sa GMM-om, tzv. statistički pristup, pošto predstavlja dominantani uspešan pristup prepoznavanju govora poslednjih nekoliko decenija. Cilj prepoznavanja govora se u slučaju statističkog pristupa može definisati kao prona-

5 Pod frekvencijskom rezolucijom se podrazumeva najmanje rastojanje između učestanosti dve prostoperiodične komponente signala koje je moguće razlikovati, a određena je tipom i širinom (T_w) prozorske funkcije. U slučaju pravougaone prozorske funkcije frekvencijska rezolucija je $\Delta f = \frac{1}{T_w}$, a u slučaju trougaone, Hanove (*Hann*), Hemingove (*Hamming*) i Blekmanove (*Blackman*) prozorske funkcije je $\Delta f = \frac{2}{T_w}$.

laženje najverovatnije sekvence reči $\hat{w}_1^N$ koja odgovara zadatoj ulaznoj sekvenci opservacija $\mathbf{o}_1^T$ odnosno:

$$\hat{w}_1^N = \arg \max_{w_1^{N_1}, N_1} p\left(w_1^{N_1} | \mathbf{o}_1^T\right) \tag{5}$$

gde je $p\left(w_1^N | \mathbf{o}_1^T\right)$, tzv. a posteriori verovatnoća, odnosno uslovna verovatnoća da sekvenci opservacija $\mathbf{o}_1^T$ odgovara sekvenca reči $w_1^{N_1}$. Treba primetiti da je u slučaju statističkog pristupa prepoznavanju govora za klasifikacionu funkciju $h(\cdot)$ izabrana arg max funkcija verovatnoće.

A posteriori verovatnoća sekvence reči koja se koristi u jednačini (5) nije pogodna za izabrani generativni model (HMM u kombinaciji sa GMM-om) stoga se koristi Bejzovo (*Bayes*) pravilo:

$$p\left(w_1^N | \mathbf{o}_1^T\right) = \frac{p\left(\mathbf{o}_1^T | w_1^N\right) p\left(w_1^N\right)}{p\left(\mathbf{o}_1^T\right)} \tag{6}$$

gde je $p\left(\mathbf{o}_1^T | w_1^N\right)$ tzv. izglednost (*likelihood*), odnosno uslovna verovatnoća da sekvenci reči w_1^N odgovara sekvenca opservacija $\mathbf{o}_1^T$, $p\left(w_1^N\right)$ tzv. a priori verovatnoća, odnosno verovatnoća da se pojavi sekvenca reči w_1^N i $p\left(\mathbf{o}_1^T\right)$ verovatnoća da se pojavi sekvenca opservacija $\mathbf{o}_1^T$. Ciljna funkcija data izrazom (5) se stoga može preurediti na sledeći način:

$$\hat{w}_1^N = \arg \max_{w_1^{N_1}, N_1} p\left(\mathbf{o}_1^T | w_1^{N_1}\right) p\left(w_1^{N_1}\right) \tag{7}$$

pri čemu se $p\left(\mathbf{o}_1^T\right)$ koji se nalazi u imeniocu može izostaviti jer ne zavisi od izabrane sekvence reči $w_1^{N_1}$, po kojoj se vrši maksimizacija.

A priori verovatnoća sekvence reči $w_1^{N_1}$, ne zavisi od ulazne sekvence opservacija (odatle i naziv a priori) već od samog jezika, stoga se ona često naziva modelom jezika. U slučaju prepoznavanja na malim rečnicima definiše se preko skupa pravila prelaza između reči koji se uobičajeno naziva gramatika, ali u slučaju prepoznavanja na velikim rečnicima koriste se estimirane verovatnoće pojavljivanja pojedinih reči i grupa reči. Više o načinima modelovanja jezika može se pronaći u (Jurafsky et al., 2000).

Izglednost $p\left(\mathbf{o}_1^T | w_1^N\right)$ povezuje foneme sa njihovim akustičkim manifestacijama stoga se naziva akustičkim modelom. Kao što je ranije napomenuto, ovaj akustički model je predstavljen pomoću HMM u kombinaciji sa GMM-om, odnosno:

$$p\left(\mathbf{o}_1^T | w_1^N\right) = \pi_{s_1} b_{s_1}(\mathbf{o}_1) \prod_{i=2}^{T} a_{s_{i-1}, s_i} b_{s_i}(\mathbf{o}_i) \tag{8}$$

gde je π_{s_1} verovatnoća da se u početnom trenutku nađe u stanju s_1, a_{s_{i-1}, s_i} verovatnoća prelaza iz stanja s_{i-1} u stanje s_i, $b_{s_i}(\mathbf{o}_i)$ verovatnoća da je stanje s_i emitovalo opservaciju $\mathbf{o}_i$, pri čemu sekvenca stanja s_1^T odgovara sekvenci reči w_1^N. Treba napomenuti da indeksi koji se nalaze uz odgovarajuća stanja predstavljaju identifikatore trenutka a ne identifikatore stanja, odnosno da gore označena stanja s_i i s_j ne moraju biti nužno različita stanja već da odgovaraju trenucima i i j. Verovatnoća emitovanja $b_{s_i}(\mathbf{o}_i)$ je u ovom modelu opisana pomoću GMM-a odnosno definisana je izrazom:

$$b_{s_i}(\mathbf{o}_i) = \sum_{m=1}^{M_{s_i}} w_{s_i m} \frac{1}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}_{s_i m}|}} e^{-\frac{1}{2}(\mathbf{o}_i - \boldsymbol{\mu}_{s_i m})^T \boldsymbol{\Sigma}_{s_i m}^{-1} (\mathbf{o}_i - \boldsymbol{\mu}_{s_i m})} \tag{9}$$

gde je D broj korišćenih obeležja, $w_{s_i m}$, $\boldsymbol{\mu}_{s_i m}$ i $\boldsymbol{\Sigma}_{s_i m}$ težina, srednja vrednost i kovarijansa m-te komponente GMM-a (Gausove raspodele) respektivno, M_{s_i} ukupan broj Gausovih raspodela pridruženih stanju s_i. Treba primetiti da $b_{s_i}(\mathbf{o}_i)$ po svojoj prirodi ne predstavlja verovatnoću već gustinu verovatnoće.

Pošto stanja HMM-a mogu da se organizuju tako da formiraju trelis strukturu, potraga za najverovatnijom sekvencom reči definisana jednačinom (5) se može efikasno realizovati pomoću Viteribijevog algoritma, potragom za najverovatnijom sekvencom stanja. Da bi se prepoznavanje realizovalo u realnom vremenu, neophodno je obezbediti da proces dekodovanja bude dovoljno brz. Na osnovu prethodno izloženog jasno je da najveći broj računskih operacija pri izračunavanju $p\left(\mathbf{o}_1^T | w_1^N\right)$ odlazi na izračunavanje vrednosti gustina verovatnoća emitovanja, i stoga optimizacijom tog modela bi se mogla obezbediti značajna ušteda.

Pored brzine neophodne pri dekodovanju potrebno je obezbediti i odgovarajuću tačnost modela. Na Fakultetu tehničkih nauka Univerziteta u Novom Sadu istraživanja su išla u pravcu povećanja tačnosti usvajanjem modela u kome su eksplicitno modelovane korelacije između obeležja (GMM sa punim kovarijansnim matricama), a potom ubrzavanjem dekodovanja hijerarhijskim klasterovanjem Gausovih smeša (Janev et al., 2010; Popović et al., 2012). Pristup u ovom radu je bio nešto drugačiji, odnosno išlo se na ubrzavanje dekodovanja smanjenjem broja parametara modela uz neznatan gubitak tačnosti. Pretpostavka na kojoj se zasniva ovaj rad jeste da kovarijansne matrice (tačnije njihove inverzne vrednosti) Gausovih raspodela obrazuju nižedimenzionalne potprostore koje je moguće efikasno predstaviti u prostoru njihovih karakterističnih vektora. Pretpostavka o mogućnosti razapinjanja inverznih kovarijansnih matrica u nekom vektorskom prostoru nije nova, ali ideja da ta predstava može imati retku reprezentaciju, koja je iskorišćena u ovom radu jeste. Detaljan pregled alternativnih metoda za modelovanje Gausovih raspodela u GMM-u je dat u poglavlju 2. Osnovni teorijski principi na kojima se bazira retka reprezentacija kao i algoritmi koji se pri tome koriste, a koji su iskorišćeni u predloženom modelu su dati u poglavlju 3. Detaljan opis predloženog modela je dat u poglavlju 4, dok su rezultati testova praćeni odgovarajućom diskusijom navedeni u poglavlju 6. Detaljan opis metoda kao i govornih resursa koji su korišćeni u testovima su dati u poglavlju 5. Poslednje poglavlje 7 sadrži zaključke ovog istraživanja kao i potencijalne smernice za buduća istraživanja.

2

NAČINI MODELOVANJA KOVARIJANSNIH MATRICA

2.1 UVOD

U ovom poglavlju dat je pregled postojećih aproksimacija kovarijansnih matrica koje se koriste u modelima Gausovih mešavina. Cilj ovih aproksimacija jeste da obezbede:

- tačnu procenu gustine verovatnoće emitovanja,
- robustnu procenu parametara modela i u slučajevima kada nije na raspolaganju celokupna populacija,
- efikasno (brzo) izračunavanje potrebnih emitujućih verovatnoća i
- male memorijske zahteve za smeštanje parametara modela.

Tačna procena gustine verovatnoće emitovanja, znači manje odstupanje modela od stvarnog procesa koji treba da se modeluje, što obično znači veću tačnost prepoznavanja. Većina sistema za prepoznavanje oblika se obučava na instancama koje su pridružene odgovarajućim klasama, ali te instance obično ne obuhvataju sve moguće realizacije, što za posledicu ima degradaciju tačnosti prepoznavanja sistema (Jiang, 2011). U slučaju da je broj parametara modela velik, a broj opservacija koje su na raspolaganju za obuku mali, nije moguće obezbediti pouzdanu procenu parametara. Odnosno, model gotovo savršeno opisuje opservacije na kojima je obučen, ali neviđene opservacije, one koje se ne nalaze u skupu za obuku, opisuje prilično loše. Ova degradacija tačnosti prepoznavanja postaje izraženija, što je skup za obuku manji, i obično samim tim manje reprezentativan. Ovaj problem se prevazilazi povećanjem skupa za obuku, što ponekad i nije tako jednostavno, ili smanjenjem broja parametara koji su potrebni za opis modela. Smanjenje broja parametara modela je bitno i iz praktičnih razloga, jer manji broj parametara obično znači i brže izračunavanje potrebnih verovatnoća emitovanja ali i manje memorijske zahteve za njihovo smeštanje.

Model Gausovih mešavina (GMM *Gaussian mixture model*) predstavlja jedan od najzastupljenijih modela koji se koristi pri statističkom prepoznavanju oblika (Saon i Chien, 2012; Plataniotis i Hatzinakos, 2000). U ovom modelu, gustina raspodele verovatnoće emitovanja opservacije $\mathbf{o}$ modelovana je pomoću ponderisane sume D-dimenzionalnih Gausovih raspodela odnosno:

$$b_s(\mathbf{o}) = \sum_{m=1}^{M_s} w_{sm} \frac{1}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}_{sm}|}} e^{-\frac{1}{2}(\mathbf{o}-\boldsymbol{\mu}_{sm})^T \boldsymbol{\Sigma}_{sm}^{-1}(\mathbf{o}-\boldsymbol{\mu}_{sm})} \tag{10}$$

gde je sa w_{sm}, $\boldsymbol{\mu}_{sm}$ i $\boldsymbol{\Sigma}_{sm}$ označena težina (apriori verovatnoća), srednja vrednost i kovarijansa m-te komponente GMM (Gausove raspodele) koja je pridružena klasi s,[1] a sa M_s ukupan broj Gausovih raspodela koje čine mešavinu.

1 U slučaju sistema za automatsko prepoznavanje govora zasnovanim na skrivenim Markovljevim modelima i GMM-u, klasa odgovara jednom stanju skrivenog Markovljevog modela.

Broj parametara koji opisuje jednu komponentu mešavine (jednu Gausovu raspodelu) je $\frac{1}{2}(D+1)(D+2)$ gde najveći deo predstavljaju elementi kovarijansne matrice $\frac{1}{2}(D+1)D$. [2] Ovako veliki broj parametara znači i relativno velike memorijske zahteve za smeštanje modela, ali i određene probleme pri estimaciji parametara. Ukoliko je broj opservacija manji od dimenzije prostora opservacija D dobijena kovarijansna matrica je singularna, a ako je tek nešto veći od D tada se dobija numerički loše kondicionirana matrica odnosno njena inverzija značajno povećava grešku estimacije (Ledoit i Wolf, 2004). Stoga korišćenje pune kovarijansne matrice, u slučaju konačnog skupa za obuku zbog problema koji se javljaju pri estimaciji parametara ne znači i poboljšanje tačnosti prepoznavanja.

Kao što je prethodno napomenuto drugi problem o kome treba voditi računa jeste brzina izračunavanja. Pošto je potrebno izračunati izglednost sekvence opservacija, a da bi se pritom izbegli problemi sa gubitkom tačnosti koja je posledica množenja vrednosti koje su manje od jedan, umesto vrednosti gustine raspodele verovatnoće emitovanja opservacije posmatra se njen logaritam, odnosno:

$$g_s(\mathbf{o}) = \ln\left(\sum_{m=1}^{M_s} w_{sm} \frac{1}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}_{sm}|}} e^{-\frac{1}{2}(\mathbf{o}-\boldsymbol{\mu}_{sm})^T\boldsymbol{\Sigma}_{sm}^{-1}(\mathbf{o}-\boldsymbol{\mu}_{sm})}\right) \tag{11}$$

Pri izračunavanju gustine raspodele emitovanja opservacija izraz (11) se često aproksimira tako što se umesto sume uzmima maksimum odnosno:

$$g_{s0}(\mathbf{o}) = \ln\left(\max_m w_{sm} \frac{1}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}_{sm}|}} e^{-\frac{1}{2}(\mathbf{o}-\boldsymbol{\mu}_{sm})^T\boldsymbol{\Sigma}_{sm}^{-1}(\mathbf{o}-\boldsymbol{\mu}_{sm})}\right) \tag{12}$$

Gornja aproksimacija sledi iz osobine logaritma $\log(a+b) \approx \log(\max\{a,b\})$ ukoliko važi $a >> b > 0$. Može se pokazati da se stvarna vrednost gustine raspodele emitovanja opservacije nalazi u granicama $g_{s0}(\mathbf{o}) < g_s(\mathbf{o}) \leqslant \ln(M) + g_{s0}(\mathbf{o})$, što znači da je odstupanje manje što je broj komponenata koje čine mešavinu manji. Problemi nastaju ukoliko postoji značajno preklapanje između komponenata jedne mešavine. Ako bi uzeli ekstremni slučaj u kome sve mešavine imaju istu srednju vrednost i varijansu tada bi razlika između stvarne i aproksimirane vrednosti bila $\ln(M)$. (Dognin et al., 2009). Eksperimenti su pokazali da u većini slučajeva aproksimacija maksimum operatorom neznatno utiče na tačnost sistema za prepoznavanje govora.

Izraz (12) je moguće svesti na sledeći oblik:

$$g_{s0}(\mathbf{o}) = \frac{1}{2}\max_m\left\{-\ln(2\pi)^D + \ln\frac{w_{sm}^2}{|\boldsymbol{\Sigma}_{sm}|} - (\mathbf{o}-\boldsymbol{\mu}_{sm})^T\boldsymbol{\Sigma}_{sm}^{-1}(\mathbf{o}-\boldsymbol{\mu}_{sm})\right\} \tag{13}$$

Pošto se član $\ln(2\pi)^D$ pojavljuje kod svih komponenata svih mešavina moguće ga je izostaviti iz računanja. Sa druge strane član $\ln\frac{w_{sm}^2}{|\boldsymbol{\Sigma}_{sm}|}$ se može tretirati kao jedan parametar komponente mešavine i stoga unapred izračunati, čime bi se izbeglo izračunavanje logaritma u izrazu $g_{s0}(\cdot)$. Stoga je jasno da najveći broj operacija odlazi na izračunavanje člana $(\mathbf{o}-\boldsymbol{\mu}_{sm})^T\boldsymbol{\Sigma}_{sm}^{-1}(\mathbf{o}-\boldsymbol{\mu}_{sm})$, što u slučaju punih kovarijansnih matrica znači $D(3D+5)/2$ osnovnih operacija[3]

2 Iako je puna kovarijansna matrica broj D^2 elemenata, zbog simetričnosti samo $\frac{1}{2}D(D+1)$ su jedinstveni. Ukupan broj parametara se dobija kada se broju elemenata kovarijansne matrice doda broj elemenata koji čine vektor srednje vrednosti (D) i težina komponente (1), što u zbiru čini $\frac{1}{2}(D+1)(D+2)$.

3 Osnovne operacije podrazumevaju sabiranje, oduzimanje, množenje i deljenje.

sa pokretnim zarezom (flops ***fl**oating-point **op**eration**s***). Ovaj broj operacija ne uključuje operacije koje su potrebne za izračunavanje inverzne kovarijansne matrice, pošto je inverznu matricu moguće izračunati unapred (u toku procesa obuke) pri čemu se kao parametar modela umesto kovarijansne matrice čuva njena inverzna vrednost. Jedan od načina za prevazilaženje problema memorijske i računske kompleksnosti modela sa punim kovarijansnim matricama jeste dijagonala aproksimacija kovarijansne matrice.

2.2 DIJAGONALNA APROKSIMACIJA KOVARIJANSNE MATRICE

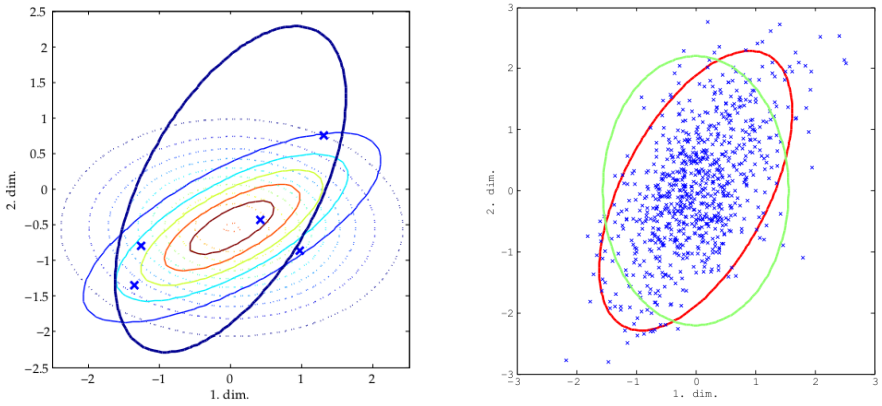
Jedan od najčešćih korišćenih pristupa za prevazilaženje problema velike složenosti koja postoji kod GMM-a jeste aproksimacija kovarijansnih matrica dijagonalnim. Na ovaj način broj parametara po komponenti mešavine se smanjuje na $2D+1$, a računska složenost izračunavanja $(\mathbf{o}-\boldsymbol{\mu}_{sm})^{\top}\boldsymbol{\Sigma}_{sm}^{-1}(\mathbf{o}-\boldsymbol{\mu}_{sm})$ na $4D-1$ flops. Singularna matrica se dobija ukoliko se na glavnoj dijagonali matrice pojavi nula, što se dobija ukoliko je na raspolaganju samo jedna opservacija ili ukoliko je neko obeležje isto u celoj populaciji što je maloverovatno. Pored toga dijagonalna aproksimacija matrice u slučaju malog broja opservacija na osnovu kojih se vrši estimacija je bolje kondicionirana nego estimirana puna kovarijansna matrica, odnosno količnik maksimalne i minimalne varijanse obeležja je uvek manji nego količnik maksimalne i minimalne karakteristične vrednosti[4] (formalni dokaz je naveden u A.1). Ovo svojstvo je ilustrovano na slici 2a, gde je prikazano 5 opservacija i Gausove raspodele koje se dobijaju na osnovu njih u slučaju da je kovarijansna matrica modelovana kao dijagonalna odnosno puna matrica. Kao što se može videti količnik velikog i malog poluprečnika elipse, koji odgovara odnosu varijansi/karakterističnih vrednosti, je veći u slučaju pune kovarijansne matrice.

Uvođenje pretpostavke da je kovarijansna matrica dijagonalna, ekvivalentno je uvođenju pretpostavke da su obeležja međusobno nekorelisana. Ova pretpostavka obično nije tačna, te tako u oblasti prepoznavanja govora pretpostavka o nekorelisanosti mel-frekvencijskih kepstralnih koeficijenata (MFCC *Mel-frequency cepstral coefficient*), koji predstavljaju najčešće korišćena obeležja za prepoznavanje govora[5] ne stoji (Janev et al., 2007), tako da dolazi do degradacije tačnosti prepoznavanja.

Na slici 2b je ilustrovana greška koja je posledica aproksimacije kovarijansne matrice dijagonalnom. Može se primetiti da se dijagonalnom aproksimacijom povećava gustina verovatnoće u oblastima koje ne pripadaju klasi i smanjuje u oblastima koje zaista pripadaju klasi. Treba primetiti da ovo važi ukoliko je na raspolaganju dovoljno veliki broj opservacija, što se vidi i na slici 2a, gde se na primeru dvodimenzionalnog prostora i 5 opservacija manja greška pravi ukoliko se koristi dijagonalna aproksimacija. Kako se ovo odražava na Gausov klasifikator ilustrovano je na slici 3, gde je došlo do povećanja broja pogrešno klasifikovanih opservacija za približno 1/5 uvođenjem dijagonalne aproksimacije. Ovde je data ilustracija za dvodimenzionalni slučaj, a u slučaju velikog broja dimenzija ove razlike su značajnije.

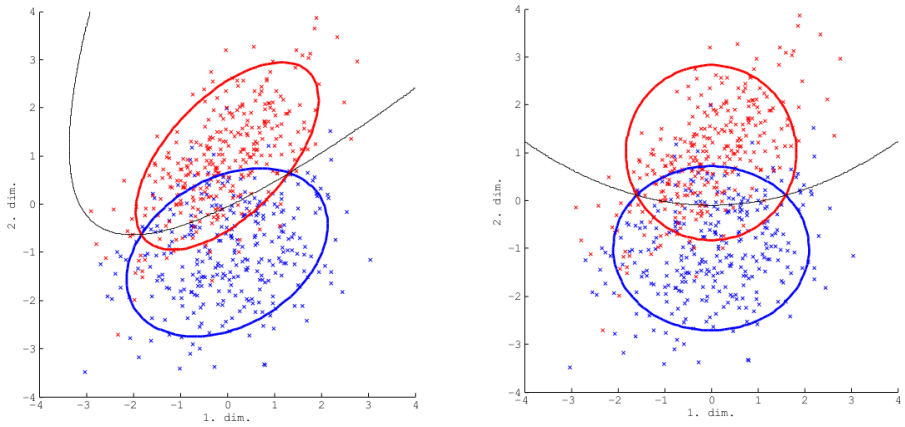
4 Izuzetak je kada su obeležja nekorelisana i kada su varijanse obeležja ujedno i karakteristične vrednosti.

5 Evropski institut za standardizaciju u telekomunikacijama (ETSI) ih je definisao kao standardna obeležja za distribuirano prepoznavanje govora Technical standard ES 201 108, v1.1.3.



(a) Kondiciranost estimiranih matrica (b) Greška aproksimacije

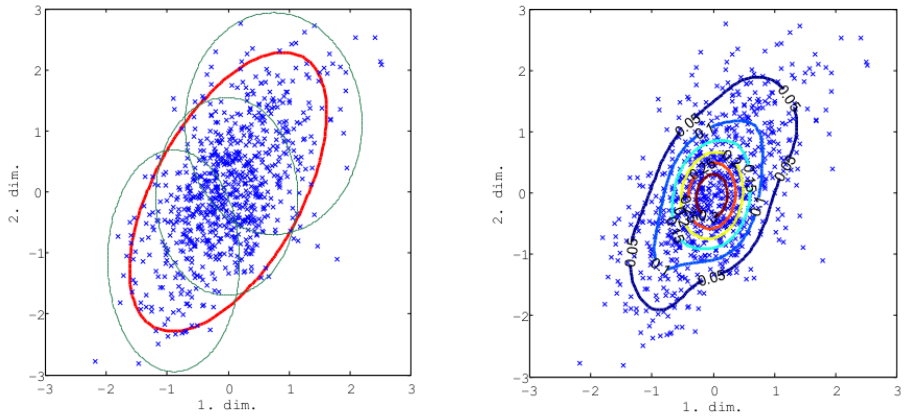
Slika 2: Sa 'x' su označene opservacije koje predstavljaju realizaciju slučajne promenljive koja ima Gausovu raspodelu sa nultom srednjom vrednošću i kovarijansnom matricom: $\mathbf{\Sigma} = [0.5, 0.4; 0.4, 1]$. Na slici levo je dato poređenje kondicioniranosti estimiranih kovarijansnih matrica u slučaju dijagonalne (isprekidane linije) i pune kovarijansne matrice (pune linije) koje su estimirane na osnovu prikazanih 5 opservacija. Plavom elipsom označene su tačke u kojima stvarna funkcija gustine raspodele ima vrednost 0.02. Na slici desno je ilustrovana greška koja se čini u slučaju dijagonalne aproksimacije kovarijansne matrice. I na ovom grafiku su elipsama označene tačke u kojima estimirane funkcije gustina raspodele imaju vrednost 0.02 i to zelenom bojom u slučaju dijagonalne aproksimacije i crvenom u slučaju da se koristi puna kovarijansna matrica.



(a) Puna kovarijansna matrica (b) Dijagonalna kovarijansna matrica

Slika 3: Uticaj dijagonalne aproksimacije kovarijansne matrice na Gausov klasifikator sa 2 klase. Plava klasa se može modelovati sa Gausovom raspodelom čija je srednja vrednost $\mathbf{\mu}_p = [0, -1]^T$, a varijansa $\mathbf{\Sigma}_p = [1.5, 0.4; 0.4, 1.0]$. Crvena klasa se može modelovati sa Gausovom raspodelom čija je srednja vrednost $\mathbf{\mu}_c = [0, 1]^T$, a varijansa $\mathbf{\Sigma}_c = [1.0, 0.6; 0.6, 1.0]$. Crnom linijom je prikazana granica između klasa. U ovom primeru, uvođenjem dijagonalne aproksimacije greška klasifikacije se povećala sa 14.5% na 18.5%.

Jedan od načina da se prevaziđe ovaj problem jeste da se koristi dovoljno veliki broj komponenata GMM-a kojima bi se implicitno modelovale korelacije između obeležja, pošto GMM može praktično modelovati bilo koju raspodelu.



(a) Pojedinačne komponente GMM.

(b) Rezulutujća GMM raspodela

Slika 4: Aproksimacija Gausove raspodele pomoću GMM kod kojih su kovarijansne matrice aproksimirane dijagonalnim. Na slici levo crvenom elipsom je predstavljena stvarna raspodela, a zelenim elipsama pojedinačne komponente koje čine GMM. Na slici desno predstavljena je raspodela koju daje GMM (ponderisana suma pojedinačnih raspodela).

Na slici 4, je data ilustracija smanjenja greške modelovanja ukoliko se umesto jedne Gausove raspodele koristi nekoliko njih. Vidi se da se dobijaju nešto opštiji modeli, pošto opservacije koje se nalaze na rubovima oblasti imaju nešto veće vrednosti gustine verovatnoće kada su modelovani pomoću GMM-a u odnosu na vrednost koju imaju ukoliko su modelovani odgovarajućom Gausovom raspodelom. Ovo može predstavljati problem pri prepoznavanju pošto se ujedno povećava preklapanje između susednih klasa.

Jednostavno uopštenje dijagonalne aproksimacije kovarijansne matrice bi bila blok dijagonalna aproksimacija, kod koje se podrazumeva da su samo pojedini podskupovi obeležja međusobno nekorelisani ili slabo korelisani.

2.3 BLOK DIJAGONALNA APROKSIMACIJA KOVARIJANSNE MATRICE

U slučaju standardnih obeležja koja se koriste pri prepoznavanju govora, a koja obuhvataju MFCC, normalizovanu energiju i njihove prve i druge izvode, uočeno je da postoji slaba korelisanost između statičkih obeležja i njihovih prvih izvoda, kao i između prvih i drugih izvoda. Ovo je posledica regresionog izraza na osnovu kog se izračunava prvi izvod:

$$\Delta c_{t,i} = \frac{\sum_{\theta=1}^{\Theta} \theta \left(c_{t+\theta,i} - c_{t-\theta,i} \right)}{2 \sum_{\theta=1}^{\Theta} \theta^2} \tag{14}$$

jer se pri izračunavanju prvog izvoda i-tog obeležja u trenutku t ($\Delta c_{t,i}$) ne koristi njegova vrednost u trenutku t ($c_{t,i}$).

Struktura blok dijagonalne matrice je data izrazom:

$$\Sigma = \begin{bmatrix} S_0 & 0 & 0 \\ 0 & S_1 & 0 \\ 0 & 0 & S_2 \end{bmatrix} \tag{15}$$

gde su sa $\mathbf{S}_i$ za $i = 0, 1, 2$ označene pune kovarijansne matrice (Vanhoucke i Sankar, 2004). U slučaju standardne pretpostavke o slaboj korelisanosti između statičkih obeležja, njihovih prvih i drugih izvoda, broj parametara po komponenti se svodi na $D(D+9)/6+1$. Pošto inverzna matrica blok dijagonalne matrice takođe blok dijagonalna, računska složenost izračunavanja $(\mathbf{o}-\boldsymbol{\mu}_{sm})^T \boldsymbol{\Sigma}_{sm}^{-1} (\mathbf{o}-\boldsymbol{\mu}_{sm})$ se svodi na $D(D+3)/2$ flops.[6] U poređenju sa dijagonalnom aproksimacijom broj parametara modela, kao i broj operacija pri izračunavanju vrednosti gustine verovatnoće emitovanja opservacije je oko D puta veći, ali i preciznost modela je takođe povećana, pošto su uzete u obzir najznačajnije korelacije koje postoje između obeležja između obeležja. Naravno veći broj parametara podrazumeva i veći broj opservacija koje su potrebne za njihovu pouzdanu estimaciju. Iako je broj parametara modela kao i broj potrebnih operacija za izračunavanje vrednosti gustine verovatnoće u slučaju blok dijagonalne aproksimacije kvadratna funkcija dimenzije prostora kao i u slučaju kada se koriste pune kovarijansne matrice, njihove vrednosti su nekoliko puta manje nego u slučaju punih kovarijansnih matrica.

Ovaj koncept je blizak konceptu koji se koristi u pristupu kod kojeg se ulazni vektor obeležja organizuje u nekoliko podvektora (*streams*), s tom razlikom da je u ovom slučaju broj komponenata koje čine mešavinu isti za svaki podvektor, i pri tome pridruživanje opservacija komponentama mešavina se ne vrši nezavisno po podvektorima.

Prethodno izlaganje je usko vezano za obeležja koja se koristi u automatskom prepoznavanju govora, ali ideja da se ne modeluju korelacije između obeležja za koja unapred možemo pretpostaviti da su slabo korelisana se mogu primeniti u bilo kojem modelu. Polazna pretpostavka da su grupe opservacija međusobno nekorelisane i da ta korelisanost ne zavisi od klase (HMM stanja) koja se modeluje je prilično gruba. Jedan od metoda koja omogućuje relaksaciju pretpostavki o nekorelisanosti pojedinih obeležja jeste faktorska analiza.

2.4 FAKTORSKA ANALIZA

Faktorska analiza[7] je jedan od načina za opisivanje višedimenzionalnih podataka korišćenjem manjeg skupa skrivenih (latentnih) promenljivih. U ovom modelu opservacija $\mathbf{o}$ je predstavljena kao linearna kombinacija faktora na koje je pridodat aditivni Gausov šum $\mathbf{n}$ srednje vrednosti $\boldsymbol{\mu}_n$ i kovarijanse matrice $\boldsymbol{\Sigma}_n$, odnosno:

$$\mathbf{o} = \mathbf{F}\boldsymbol{\beta} + \mathbf{n} \tag{16}$$

gde je sa $\boldsymbol{\beta}$ označena skrivena promenljiva, a sa $\mathbf{F}$ matrica čije su kolone tzv. faktori odnosno nosioci prostora. Pretpostavlja se da je dimenzija prostora opservacija D veća od dimenzije prostora skrivenih promenljivih k ($D > k$), da skrivene promenljive podležu Gausovoj raspodeli nulte srednje vrednosti i jedinične varijanse ($\mathcal{N}(\mathbf{0}, \mathbf{I})$), kao i da su komponente vektora šuma nezavisne odnosno da je matrica $\boldsymbol{\Sigma}_n$ dijagonalna.

6 Ukupan broj operacija sa pokretnim zarezom potreban za izračunavanje logaritma izglednosti jedne Gausove raspodele u ovoj varijanti modela iznosi $D(D+5)/2+2$.

7 Termin je preuzet direktno iz statističke literature na srpskom jeziku, zbog toga nije preveden kao rastavljanje na faktore.

Pošto je linearna kombinacija slučajnih promenljivih sa Gausovim raspodelama takođe slučajna promenljiva sa Gausovom raspodelom, opservacija $\mathbf{o}$ ima Gausovu raspodelu srednje vrednosti $\boldsymbol{\mu}_n$ i kovarijansne matrice $\boldsymbol{\Sigma}_o = \mathbf{F}\mathbf{F}^T + \boldsymbol{\Sigma}_n$, tako da je broj parametara sa kojima je opisana jedna Gausova raspodela $D(k+2)+1$. Jasno je da se ušteda u pogledu broja parametara u odnosu na slučaj sa punom kovarijansnom matricom i istim brojem Gausovih raspodela postiže ukoliko je $k < (D-1)/2$. Broj faktora k je obično broj karakterističnih vrednosti kovarijansne matrice koji je veći od nekog unapred zadatog praga (Saul i Rahim, 1997, 2000).

Prethodno opisan model nije invarijantan na linearnu transformaciju, što znači da je moguće pronaći transformacionu matricu koja će projektovati obeležja u neki drugi prostor kojim bi se sa istim brojem parametara dobilo bolje pokrivanje prostora u smislu maksimalne izglednosti. Da model zavisi od linearne transformacije, lako se pokazuje množenjem opservacije sa proizvoljnom matricom $\mathbf{C}$:

$$\mathbf{C}\mathbf{o} = \mathbf{C}\mathbf{F}\boldsymbol{\beta} + \mathbf{C}\mathbf{n} \tag{17}$$

Jasno je da je u ovom slučaju kovarijansna matrica transformisane šumne komponente $\mathbf{C}\boldsymbol{\Sigma}_n\mathbf{C}^T$ koja za proizvoljnu transformaciju $\mathbf{C}$ ne mora biti dijagonalna, što model pretpostavlja. Stoga je metoda faktorske analize modifikovana tako da je invarijantna na linearnu transformaciju usvajanjem sledeće veze između između komponenata:

$$\mathbf{o} = \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\mu} + \mathbf{T}\mathbf{n} \tag{18}$$

gde je $\mathbf{T}$ transformaciona matrica, ali ovaj put šumna komponenta $\mathbf{n}$ ima nultu srednju vrednost[8], dok je značenje svih ostalih parametara nepromenjeno. Ukoliko bi svaka Gausova raspodela imala sopstvenu vrednost parametara $\{\mathbf{F}, \mathbf{T}, \boldsymbol{\Sigma}_n\}$ tada bi rešenje koje se dobija na osnovu kriterijuma maksimalne izglednosti bilo: $\{\mathbf{0}, \mathbf{V}, \boldsymbol{\Lambda}\}$, gde je sa $\mathbf{V}$ označena matrica karakterističnih vektora, a sa $\boldsymbol{\Lambda}$ matrica odgovarajućih karakterističnih vrednosti za estimiranu kovarijansnu matricu, stoga je neophodno da pojedine klase (HMM stanja, fonemi) međusobno dele parametre. U ovom slučaju opservacija $\mathbf{o}$ ima Gausovu raspodelu sa srednjom vrednošću $\boldsymbol{\mu}$ i kovarijansnom matricom $\mathbf{F}\mathbf{F}^T + \mathbf{T}\boldsymbol{\Sigma}_n\mathbf{T}^T$. Broj parametara sa kojma je opisana jedna Gausova raspodela je $D(D+k+2)+1$, ali stvaran broj parametara na nivou celokupnog modela je daleko manji, pošto Gausove raspodele međusobno dele pojedine parametere. Pošto od načina na koji se dele parametri zavisi ukupan broj parametara modela, bez konkretnog primera nije moguće uporediti ovaj način modelovanja sa drugim (Gopinath et al., 1998).

Uopštenje modela predstavljaju skriveni Markovljevi modeli na nivou faktora (FAHMM *Factor analysed hidden Markov model*) kod kojih je k dimenzionalni vektor stanja $\boldsymbol{\beta}$ generisan od strane skrivenog Markovljevog modela čije su gustine verovatnoća emitovanja modelovane pomoću GMM-a, dok se op-

8 Ovo ograničenje je uvedeno zbog pojednostavljenja notacije i estimacije parametara, jer se parametar $\boldsymbol{\mu}$ može posmatrati i kao $\mathbf{T}\boldsymbol{\mu}_n$.

servacija $\mathbf{o}$ i dalje modeluje kao u jednačini (16), s tom razlikom da je i šumna komponenta $\mathbf{n}$ modelovana pomoću GMM-a, odnosno:

$$\begin{aligned} &\boldsymbol{\beta} \sim \mathcal{M}^{hmm}, \quad \mathcal{M}^{hmm} = \left\{\pi_s, a_{s,s_i}, w_{sm}^{(\beta)}, \boldsymbol{\mu}_{sm}^{(\beta)}, \boldsymbol{\Sigma}_{sm}^{(\beta)}\right\} \\ &\mathbf{o} = \mathbf{F}_s\boldsymbol{\beta} + \mathbf{n} \qquad\qquad \mathbf{n} \sim \textstyle\sum_n w_{sn}^{(o)} \mathcal{N}\left(\boldsymbol{\mu}_{sn}^{(o)}, \boldsymbol{\Sigma}_{sn}^{(o)}\right) \end{aligned} \tag{19}$$

gde je π_s inicijalna verovatnoća HMM stanja s, a_{s,s_i} verovatnoća prelaza iz stanja s u stanje s_i, $w_{sm}^{(\beta)}$, $\boldsymbol{\mu}_{sm}^{(\beta)}$ i $\boldsymbol{\Sigma}_{sm}^{(\beta)}$ težina, srednja vrednost i kovarijansna matrica m-te komponente GMM-a pridružene stanju s respektivno koja emituje skriveni vektor $\mathbf{x}$, a $w_{sn}^{(o)}$, $\boldsymbol{\mu}_{sn}^{(o)}$ i $\boldsymbol{\Sigma}_{sn}^{(o)}$ težina, srednja vrednost i kovarijansna matrica n-te komponente GMM-a šuma $\mathbf{n}$. Da bi se istakla razlika u dimenzionalnosti prostora kojima su opisane promenljive $\boldsymbol{\beta}$ i $\mathbf{n}$ u nijhovim eksponentima je navedena oznaka promenljve koja pripada tom prostoru.

Gustina raspodele verovatnoće da bude generisana opservacija $\mathbf{o}_t$, ako je poznato da se u trenutku t model nalazi u stanju s i da je pri tome vektor stanja $\boldsymbol{\beta}_t$ generisan od komponente m a opservacija od komponente n je Gausova, odnosno:

$$p\left(\mathbf{o}_t|\mathbf{x}_t, q_t = s, m_{st} = m, m_{\beta_t} = n\right) = \mathcal{N}\left(\mathbf{o}_t; \boldsymbol{\mu}_{smn}, \boldsymbol{\Sigma}_{smn}\right) \tag{20}$$

gde je

$$\boldsymbol{\mu}_{smn} = \mathbf{F}_s\boldsymbol{\mu}_{sm}^{(\beta)} + \boldsymbol{\mu}_n^{(o)} \tag{21}$$

$$\boldsymbol{\Sigma}_{smn} = \mathbf{F}_s\boldsymbol{\Sigma}_{sm}^{(\beta)}\mathbf{F}_s^T + \boldsymbol{\Sigma}_n^{(o)} \tag{22}$$

Za razliku od prethodna dva pristupa, gde su modelovane pojedinačne Gausove raspodele, u ovom slučaju je modelovana celokupna Gausova mešavna koja pripada jednom stanju. Ako sa M_β i M_n označimo broj komponenata mešavina kojima su modelovane gustine verovatnoća emitovanja vektora $\boldsymbol{\beta}$ i $\mathbf{n}$ respektivno, broj parametara po stanju modela je $M_\beta(2k+1) + M_n(2D+1) + Dk$. Ovaj model redukuje broj parametara i na taj način obezbeđuje robustniju estimaciju u odnosu na slučaj kada se koriste pune kovarijansne matrice, ali je složenost izračunavanja vrednosti gustine raspodele verovatnoće emitovanja u zadatoj tački nepromenjena, pošto se na osnovu faktorisanog oblika matrice $\boldsymbol{\Sigma}_{smn}$ ne može dobiti faktorisani oblik matrice $\boldsymbol{\Sigma}_{smn}^{-1}$. Ovo je ujedno problem koji postoji kod svih metoda koje se baziraju na faktorskoj analizi (Rosti i Gales, 2004).

Iako primena faktorske analize, poboljšava tačnost modela u odnosu na modele koji koriste dijagonalnu aproksimaciju kovarijansne matrice i kada je broj Gausovih raspodela sličan, nemogućnost brzog izračunavanja funkcije gustine emitovanja je značajan nedostatak. Jedan od načina da se to prevaziđe jeste direktnim modelovanjem inverzne kovarijansne matrice, pošto ona direktno figuriše u izračunavanju izglednosti (videti jednačinu (11)).

2.5 APROKSIMACIJA INVERZNIH KOVARIJANSNIH MATRICA KORIŠĆENJEM VEKTORSKE REPREZENTACIJE

Kao što je navedeno u prethodnom odeljku, smanjenje broja parametara modela ne znači i smanjenje broja računskih operacija koje su potrebne za računanje izglednosti. Direktnim modelovanjem inverzne kovarijansne matrice koje

podrazumeva uvođenje deljenih (zajedničkih) parametara se postiže ubrzanje izračunavanja izglednosti kao i smanjenje broja parametara modela. Pretpostavka je da se inverzna kovarijansna matrica $\mathbf{P}_{sm}$ može predstaviti kao linearna kombinacija simetričnih matrica odnosno:

$$\mathbf{P}_{sm} = \mathbf{\Sigma}_{sm}^{-1} = \sum_{i=1}^{n} p_{smi}\mathbf{B}_i \tag{23}$$

gde je $\mathbf{B}_i$ i-ta simetrična bazna matrica, a p_{smi} težina i-te bazne matrice za predstavu inverzne kovarijansne matrice m-te komponente GMM-a stanja s. U ovom modelu bazne matrice $\mathbf{B}_i$ su deljeni parametri, dok su koeficijenti koji množe te bazne matrice p_{smi} specifični za pojedinačne Gausove raspodele. Broj parametra po jednoj Gausovoj raspodeli direktno zavisi od broja baznih matrica n i iznosi $D + n + 1$. Ukoliko je broj baznih matrica jednak dimenziji prostora D tada je broj parametara po jednoj Gausovoj raspodeli isti kao u slučaju dijagonalne aproksimacije, a ako je jednak $D(D+1)/2$ tada je taj broj isti kao u slučaju pune kovarijansne matrice. Treba napomenuti da je sada ukupan broj parametara koji opisuju model uvećan za broj parametara kojima su opisane simetrične bazne matrice $\mathbf{B}_i$. Bazne matrice su dimenzija $D \times D$, ali su obično manjeg ranga tako da je broj parametara kojim se opisuju sve bazne matrice manji od $nD(D+1)/2$, pošto se bazna matrica nižeg ranga od D može predstaviti u sledećoj formi:

$$\mathbf{B}_i = \sum_{r=1}^{R_i} b_{ir}\mathbf{d}_{ir}\mathbf{d}_{ir}^{T} \tag{24}$$

gde je sa R_i označen rang matrice, $\mathbf{d}_{ir}$ D-dimenzionalni vektor koji obrazuje r-tu simetričnu matricu ranga 1 koja formira i-tu baznu matricu. U ovoj formi broj deljenih parametara iznosi $\sum_{i=1}^{n} R_i(D+1)$, tako da se ušteda u broju parametara u odnosu na varijantu kada je bazna matrica punog ranga postiže ukoliko je $\sum_{i=1}^{n} R_i < nD/2$. Obično su sve bazne matrice istog ranga R i tada je broj deljenih parametara $nR(D+1)$.

Izraz (13) izražen u terminima inverzne kovarijansne matrice dobija oblik:

$$g_{s0}(\mathbf{o}) = \frac{1}{2}\max_{m}\left\{\ln\left(\frac{w_{sm}^2|\mathbf{P}_{sm}|}{(2\pi)^D}\right) - (\mathbf{o}-\boldsymbol{\mu}_{sm})^T\mathbf{P}_{sm}(\mathbf{o}-\boldsymbol{\mu}_{sm})\right\} \tag{25}$$

i može se preurediti u:

$$g_{s0}(\mathbf{o}) = \frac{1}{2}\max_{m}\left\{c_{sm} + \breve{\boldsymbol{\mu}}_{sm}^{T}\mathbf{o} - \mathbf{o}^{T}\mathbf{P}_{sm}\mathbf{o}\right\} \tag{26}$$

gde je

$$\begin{aligned} c_{sm} &= \ln\left(\frac{w_{sm}^2|\mathbf{P}_{sm}|}{(2\pi)^D}\right) - \boldsymbol{\mu}_{sm}^{T}\mathbf{P}_{sm}\boldsymbol{\mu}_{sm} \\ \breve{\boldsymbol{\mu}}_{sm} &= 2\mathbf{P}_{sm}\boldsymbol{\mu}_{sm} \end{aligned}$$

Parametri c_{sm} i $\breve{\boldsymbol{\mu}}_{sm}$ ne zavise od vrednosti trenutne opservacije $\mathbf{o}$ te se mogu unapred izračunati u fazi obuke.

Pošto se inverzna kovarijansna matrica može predstaviti kao linearna kombinacija baznih matrica, kao što je prikazano u jednačini (23) izračunavanje poslednjeg člana u jednačini (26) se može realizovati na sledeći način:

$$\mathbf{o}^{\mathsf{T}}\mathbf{P}_{sm}\mathbf{o} = \sum_{i=1}^{n} p_{smi}\mathbf{o}^{\mathsf{T}}\mathbf{B}_i\mathbf{o} \tag{27}$$

tako da se broj računskih operacija po jednoj Gausovoj raspodeli svodi na $2(D+n)$ flops. Ova ušteda se postiže tako što se prvo izračunaju vrednosti za $\mathbf{o}^{\mathsf{T}}\mathbf{B}_i\mathbf{o}$, pa tek onda pojedinačni logaritmi izglednosti za pojedinačne Gausove raspodele. U najgorem slučaju, kada su bazne matrice reda D, tada je broj dodatnih računanja $n3D(D+1)/2$, što je u situaciji kada je potrebno izračunati logaritam izglednosti za veliki broj Gausovih raspodela značajna ušteda. U slučaju baznih matrica nižeg reda izračunavanje $\mathbf{o}^{\mathsf{T}}\mathbf{B}_i\mathbf{o}$ se može realizovati na sledeći način

$$\mathbf{o}^{\mathsf{T}}\mathbf{B}_i\mathbf{o} = \sum_{r=1}^{R_i} b_{ir}\mathbf{o}^{\mathsf{T}}\mathbf{d}_{ir}\mathbf{d}_{ir}^{\mathsf{T}}\mathbf{o} \tag{28}$$

pri čemu se prvo izračunava skalarni proizvod $\mathbf{o}_t^{\mathsf{T}}\mathbf{b}_{ir}$, a potom se isti koristi kao međurezultat u daljem izračunavanju, čime se broj dodatnih operacija značajno smanjuje i iznosi $(2D+1)\sum_{i=1}^{n} R_i - n$. Može se pokazati da je ovakav način računski efikasniji ukoliko je prosečan rang baznih matrica manji od $\frac{3}{4}D$. Ukoliko se pretpostavi da su sve bazne matrice istog ranga R, tada se broj dodatnih operacija svodi na $n\,(R(2D+1)-1)$ flops.

U literaturi je predstavljeno nekoliko tehnika koje se mogu uklopiti u prethodno opisani pristup, a koje se međusobno razlikuju po broju baznih matrica i njihovom rangu. U ovu grupu spadaju sledeće tehnike: delimično povezane kovarijansne matrice (STC *Semi-tied covariances*) (Gales, 1999), linearna transformacija zasnovana na maksimizaciji izglednosti (MLLT *Maximum likelihood linear transformation*) (Olsen i Gopinath, 2004), proširena linearna transformacija zasnovana na maksimizaciji izglednosti (EMLLT *Extended* MLLT) (Olsen i Gopinath, 2004), hibridna proširena linearna transformacija zasnova na maksimizaciji izglednosti (*Hybrid* EMLLT) (Axelrod et al., 2005), model Gausovih mešavina sa inverznim kovarijasnim matricama ograničenim u vektorskom potprostoru (PCGMM *Precision constrained Gaussian mixture model*) (Axelrod et al., 2005), i njegova uopštena verzija model Gausovih mešavina u ograničenom potprostoru (SCGMM *Subspace constrained Gaussian mixture model*) (Axelrod et al., 2005).

2.5.1 *Delimično povezane kovarijansne matrice – STC*

Ovo je jedan od prvih modela u kome se koristi prethodno opisani pristup modelovanju inverznih kovarijansnih matrica, međutim ideja koja je dovela do njega bila je nešto drugačija. Pošlo se od pretpostavke da se za grupe klasa (HMM stanja) može definisati transformaciona matrica $\mathbf{D}^{(g(s))}$ koja će postojeći prostor obeležja preslikati u novi tako da se klase u tom novom prostoru obeležja mogu opisati pomoću GMM-a sa dijagonalnim kovarijansnim matricama, odnosno:

$$\mathbf{\Sigma}_{sm}^{(diag)} = \mathbf{D}^{(g(s))\mathsf{T}}\mathbf{\Sigma}_{sm}\mathbf{D}^{(g(s))} \tag{29}$$

gde je $\boldsymbol{\Sigma}_{sm}$ puna kovarijansna matrica m-te komponente GMM-a koja opisuje stanje s i koja se dobija na osnovu originalnih opservacija i $\boldsymbol{\Sigma}_{sm}^{(diag)}$ dijagonalna kovarijansna matrica m-te komponente u transformisanom prostoru obeležja. Oznaka u eksponentu $(g(s))$ predstavlja identifikator grupe klasa. Izabrani oblik transformacije kovarijansne matrice ekvivalentan je sledećoj transformaciji obeležja:

$$\mathbf{o}^{(g(s))} = \mathbf{D}^{(g(s))^T}\mathbf{o} \tag{30}$$

gde je $\mathbf{o}^{(g(s))}$ transformisana opservacija, tako da ova transformacija ima uticaj ne samo na kovarijansnu matricu odnosno njenu inverziju već i na srednje vrednosti Gausovih raspodela. Treba naglasiti da u opštem slučaju svaka grupa klasa transformiše opservacije u svoj prostor obeležja, odnosno postojeći prostor obeležja se ne preslikava u jedan novi prostor obeležja već u nekoliko novih prostora obeležja (onoliko koliko ima grupa klasa).

Broj transformacionih matrica zavisi od izabranog načina deljenja parametara. U radu u kom je predložena ova metoda (Gales, 1999), transformacione matrice se dele između kontekstno zavisnih modela istog fonema, ali moguće su i neke druge varijante. U slučaju da sva stanja svih modela imaju istu transformacionu matricu, tada se ovaj model svodi na heteroscedastičku linearnu diskriminativnu analizu (HLDA) u kojoj ne postoji redukcija dimenzionalnosti prostora obeležja i za koju važi pretpostavka da su transformisana obeležja međusobno nekorelisana. Pošto se ne vrši redukcija dimenzionalnosti prostora obeležja, što se uglavnom podrazumeva kad se koristi HLDA, da bi se to istaklo ovakav pristup se naziva još i linearna transformacija zasnovana na maksimizaciji izglednosti (MLLT).

Na osnovu jednačine (29) nije moguće uočiti direktnu vezu ovog modela sa opštim modelom predstavljenim u prethodnom odeljku čija je osnovna ideja sažeta u jednačinama (23) i (24), stoga ju je potrebno transformisati u pogodniji oblik. Izvođenjem izraza za inverznu punu kovarijansnu matricu na osnovu jednačine (29) dobija se:

$$\mathbf{P}_{sm} = \mathbf{D}^{(g(s))}\left(\boldsymbol{\Sigma}_{sm}^{(diag)}\right)^{-1}\mathbf{D}^{(g(s))^T} = \sum_{i=1}^{D}\frac{1}{\Sigma_{smi}^{diag}}\mathbf{d}_i^{(g(s))}\mathbf{d}_i^{(g(s))^T} \tag{31}$$

gde je sa $\mathbf{d}_i^{(g(s))}$ označena i-ta kolona matrice $\mathbf{D}^{(g(s))}$, a sa Σ_{smi}^{diag} i-ti dijagonalni element matrice Σ_{sm}^{diag}. Kao što se može videti na osnovu poslednjeg člana jednakosti (31), inverzna kovarijansna matrica predstavljena je kao linearna kombinacija D matrica ranga jedan ($\mathbf{d}_i^{(g(s))}\mathbf{d}_i^{(g(s))^T}$), odnosno kolone matrice $\mathbf{d}^{(g(s))}$ predstavljaju bazne vektore. Treba primetiti da je broj vektora koji razapinju prostor inverznih kovarijansnih matrica jednak proizvodu dimenzionalnosti prostora obeležja D i broja grupa klasa, pri čemu se za predstavu jedne inverzne kovarijansne matrice koristi uvek samo D baznih vektora. Bilo koja puna inverzna kovarijansna matrica se može predstaviti u $D(D+1)/2$ dimenzionalnom vektorskom prostoru, pomoću trivijalnih jediničnih vektora stoga se postavilo pitanje opravdanosti baze sa više od $(D+1)/2$ vektora. Odgovor na ovo pitanje je proizveo tzv. EMLLT model.

2.5.2 *Proširena linearna transformacija zasnovana na maksimizaciji izglednosti – EMLLT*

Ova metoda predstavlja uopštenje MLLT metode u smislu da je broj baznih matrica, koje su predstavljene kao proizvod kolona transformacione matrice ($\mathbf{d}_i \mathbf{d}_i^T$), povećan sa D na n, odnosno inverzna kovarijansna matrica je predstavljena kao:

$$\mathbf{P}_{sm} = \mathbf{D}\mathbf{\Lambda}_{sm}\mathbf{D}T = \sum_{i=1}^{n} \lambda_{smi} \mathbf{d}_i \mathbf{d}_i^T \tag{32}$$

gde je $\mathbf{D}$ transformaciona matrica, $\mathbf{\Lambda}_{sm}$ dijagonalna težinska matrica i λ_{smi} i-ti element na glavnoj dijagonali matrice $\mathbf{\Lambda}$. Treba primetiti da je matrica $\mathbf{D}$ ujedno i matrica koja sadrži sve bazne vektore, za razliku od matrice $\mathbf{D}^{(g(s))}$ koja je sadržala samo D baznih vektora koji su bili vezani za grupu klasa $g(s)$.

Kao što je ranije već napomenuto, inverzna kovarijansna matrica ima tačno $D(D+1)/2$ parametara, odakle sledi da je maksimalan potreban broj baznih matrica ranga 1 $D(D+1)/2$. Stoga u zavisnosti od broja n zavisi i složenost modela koja može da ide od složenosti dijagonalnih kovarijasnih matrica za $n = D$ do složenosti punih kovarijansnih matrica za $n = D(D+1)/2$.

Iako na prvi pogled ovaj pristup predstavlja trivijalno uopštenje MLLT, postoji nekoliko bitnih razlika u slučaju kada je $n > D$. Inverzna kovarijansna matrica treba da bude pozitivno definitna, što u slučaju MLLT znači da sve težine baznih matrica treba da budu pozitivne, dok ovo ograničenje u slučaju EMLLT ne postoji, odnosno koeficijenti kojima se množe matrice mogu da budu negativni. Ovo se može objasniti prirodom težine bazne matrice λ_{smjj}. Projekcija inverzne kovarijansne matrice na j-tu kolonu transformacione matrice data je izrazom:

$$\begin{aligned} p_j &= \mathbf{d}_j^T \mathbf{P}_{sm} \mathbf{d}_j && (33)\\ &= \mathbf{d}_j^T \left(\sum_{i=1}^{n} \lambda_{smi} \mathbf{d}_i \mathbf{d}_i^T \right) \mathbf{d}_j && (34)\\ &= \sum_{i=1}^{n} \lambda_{smi} (\mathbf{d}_j^T \mathbf{d}_i)^2 && (35) \end{aligned}$$

tako da je $p_j = \lambda_{smj}$ ako su sve kolone međusobno ortogonalne, što u slučaju kada postoji više od D D-dimenzionalnih vektora nije ispunjeno. Težinu bazne matrice λ_{smi} treba shvatiti kao doprinos pojedinačne bazne matrice inverznoj kovarijansnoj matrici, pri čemu taj doprinos može da bude kako pozitivan tako i negativan. Treba napomenuti da p_j treba da bude uvek pozitivno, da bi se obezbedila pozitivna definitnost inverzne kovarijansne matrice. Pored toga prilikom traženja transformacione matrice $\mathbf{D}$ treba izbegavati redundantne parametre[9] čime se eliminišu moguća degenerativna rešenja za transformacionu matricu $\mathbf{D}$, što nije postojalo u slučaju MLLT.

Specijalni slučaj EMLLT jeste tzv. procedura višestrukih linearnih transformacija (MLT *Multiple linear transforms*) (Goel i Gopinath, 2001) za koju važi

9 Kolona matrice $\mathbf{D}$ ($\mathbf{d}_j$) je redundantna ukoliko se matrica $\mathbf{d}_j \mathbf{d}j^T$ može predstaviti kao linearna kombinacija matrica oblika $\{\mathbf{d}_i \mathbf{d}i^T\}_{i \neq j}$.

da je $n < D(D+1)/2$ i da za reprezentaciju jedne matrice koristi isključivo D baznih vektora odnosno:

$$\mathbf{P}_{sm} = \sum_{i=1}^{D} \lambda_{smi} \mathbf{d}_{f(s,i)} \mathbf{d}_{f(s,i)}^{T} \tag{36}$$

gde je sa $f(s,i)$ označena funkcija koja za svako stanje s i redni broj vektora u sumi i određuje odgovarajući indeks kolone u trasformacionoj matrici $\mathbf{D}$.

Pošto je inverzna kovarijansna matrica punog ranga, postavilo se pitanje opravdanosti korišćenja isključivo baznih matrica ranga 1 što je rezultovalo PCGMM-om.

2.5.3 *Gausove mešavine sa inverznim kovarijansnim matricama u ograničenom potprostoru – PCGMM*

Ova metoda predstavlja uopštenje EMLLT u smislu da bazna matrica može da bude simetrična matrica punog ranga, a ne isključivo ranga 1. Ova ideja je iskorišćena i u okviru modela mešavina inverznih kovarijansnih matrica (MIC *Mixture of Inverse Covariances*) (Vanhoucke i Sankar, 2004), s tom razlikom da je algoritam estimacije parametara nešto drugačiji. Ukoliko je rang baznih matrica manji od D, tada se ovaj specijalni slučaj PCGMM naziva Hibridna EMLLT (Axelrod et al., 2005). Pošto ovi modeli predstavljaju uopštenje svih prethodno nabrojanih pristupa, matematički izrazi kojima se definiše ovaj model su već navedeni u uvodnom delu odeljka 2.5, te neće biti ponavljani.

Ideja povezivanja parametara GMM-a nije ograničena samo na (inverzne) kovarijansne matrice, pošto se sličan pristup može primeniti i na srednje vrednosti. Uopštenje PCGMM-a jesu Gausove mešavine u ograničenom vektorskom potprostoru (SCGMM *Subspace constrained Gaussian mixture model*) i Gausove mešavine u ograničenim vektorskim potprostorima srednjih vrednosti i inverznih kovarijansnih matrica (SPAM *Subspace constrained precision and mean*). Razlika između ova dva modela, je u načinu tretiranja prostora srednjih vrednosti i inverznih kovarijansnih matrica, gde u slučaju SCGMM prostor srednjih vrednosti i inverznih kovarijansnih matrica su objedinjeni (združeni), a u slučaju SPAM nezavisni. Tačnost ovih modela je bliska tačnosti modela koji se dobijaju pomoću punih kovarijansnih matrica. Treba napomenuti da tačnost pojedinih modela zavisi i od baze na kojoj su vršeni testovi (Axelrod et al., 2005; Varjokallio i Kurimo, 2007), tako da nije moguće dati ocenu o tome koji je model od ova dva bolji.

2.5.4 *Faktorisane retke inverzne kovarijansne matrice*

Ovaj model, opisan u (Bilmes, 2000), polazi od pretpostavke da je inverzna kovarijansna matrica retka, odnosno da su mnogi njeni elementi jednaki nuli. Nule u inverznoj kovarijansnoj matrici nalaze se na mestima koja odgovaraju obeležjima koja su međusobno uslovno nezavisna[10], odnosno ukoliko je element inverzne kovarijansne matrice u i-toj vrsti i j-toj koloni jednak nula tada

10 Dve slučajne promenljive x_i i x_j su uslovno nezavisne ako se uslovna verovatnoća pojave promenljive x_i/x_j za date sve ostale slučajne promenljive ne menja ako je poznata konkretna realizacija promenljive x_j/x_i.

su i-to i j-to obeležje uslovno nezavisni. Usled grešaka koje su posledica procena kovarijansne matrice, mnogi elementi inverzne kovarijansne matrice se razlikuju od nule iako odgovaraju uslovno nezavisnim obeležjima, stoga ih je potrebno postaviti na nulu.

Slična ideja sa postavljanjem pojedinih elemenata inverzne kovarijansne matrice na nulu je iskorišćena i kod ranije opisane blok-dijagonalne aproksimacije kovarijansne matrice, s tom razlikom da se pretpostavka o nezavisnosti pojedinih obeležja zasnivala na načinu njihovog izračunavanja i bila je ista za sve Gausove raspodele koje čine model. Postupak pronalaženja elemenata inverzne kovarijansne matrice koje treba postaviti na nulu uz maksimizaciju izglednosti je nelinearan, stoga se problem pojednostavljuje tako što se umesto inverzne kovarijanse matrice posmatra njena reprezentacija zasnovana na LDL dekompoziciji, odnosno:

$$\mathbf{P}_{sm} = \mathbf{U}_g^T \mathbf{D}_{sm} \mathbf{U}_g \tag{37}$$

gde je $\mathbf{D}_{sm}$ dijagonalna matrica, a $\mathbf{U}_g$ gornja trougaona matrica čiji su svi elementi na glavnoj dijagonali jednaki 1. Pošto za vandijagonalne elemente inverzne kovarijanse matrice važi $P_{ij} = D_{ii}U_{ij} + \sum_{k=1}^{i-1} D_{k,k}U_{k,i}U_{k,j}$ za $i < j$, gde su sa P_{ij}, D_{ij} i U_{ij} označeni elementi koji se nalaze u i-toj vrsti i j-toj koloni matrica $\mathbf{P}_{sm}$, $\mathbf{D}_{sm}$ i $\mathbf{U}_g$ respektivno, izjednačavanje elementa U_{ij} sa nulom vrednost elementa P_{ij} se približava nuli, što za posledicu ima mogućnost da se gornji problem svede na linearni, a koji podrazumeva pronalaženje elemenata iznad glavne dijagonale matrice $\mathbf{U}_g$ koje treba izjednačiti sa nulom. Ušteda u broju parametara se postiže tako što se matrica $\mathbf{U}_g$ deli između više Gausovih raspodela, a svaka Gausova raspodela ima svoju specifičnu matricu $\mathbf{D}_{sm}$.

Treba napomenuti da se ovaj model može vrlo lepo uklopiti u teoriju probabilističkih grafičkih modela, ali to prevazilazi obim ovog rada.

2.6 REZIME

U ovom poglavlju je objašnjena potreba za alternativnom reprezentacijom punih kovarijansnih matrica u GMM. Sve opisane aproksimacije kovarijansnih matrica po broju potrebnih parametara i računskoj složenosti se nalaze između dijagonalne aproksimacije, koja je najjednostavnija i najefikasnija, i modela pune kovarijansne matrice, koja je najsloženija i računski najzahtevnija, ali i najtačnija. Ušteda u broju parametara, kao i smanjenje računske složenosti se postiže uvođenjem parametara koji se dele između nekoliko Gausovih raspodela. Na ovaj nači se postižu robustniji modeli, ali dolazi do gubitka tačnosti. Svi navedeni modeli se uglavnom koriste u sistemima za automatsko prepoznavanje govora, ali se mogu proširiti i na druge sisteme za prepoznavanje oblika.

3

RETKA REPREZENTACIJA

3.1 UVOD

U ovom poglavlju je dat pregled osnovnih teorijskih principa na kojima se zasniva retka reprezentacija, a koji su iskorišćeni u ovom radu za aproksimaciju inverznih kovarijansnih matrica u modelima Gausovih raspodela.

Retka reprezentacija[1] podrazumeva predstavu podataka/signala pomoću linearne kombinacije malog broja "tipičnih" podataka/signala, koji se nazivaju atomi, a koji su dobijeni na osnovu raspoloživih podataka/signala. Nadalje u tekstu će biti korišćen samo termin signal umesto podatak/signal, ali sve što će biti navedeno važi kako za signale tako i za podatke. Već u samoj definiciji pojma retke reprezentacije uočavaju se dva problema koja je potrebno rešiti:

i) kako pronaći linearnu kombinaciju atoma tako da odstupanje od stvarne vrednosti bude minimalno i da se pri tome iskoristi što manji broj atoma,

ii) kako pronaći odgovarajuće atome.

Procedura rešavanja prvog problema se naziva retko kodovanje ili retka dekompozicija, a drugog formiranje rečnika[2] pošto se skup svih atoma naziva rečnikom.[3]

3.2 RETKO KODOVANJE

Retko kodovanje je postupak izračunavanja koeficijenata vektora $\boldsymbol{\alpha}$, tzv. retkog koda, preko kojih je moguće predstaviti signal $\mathbf{x}$ pomoću atoma rečnika $\mathbf{D}$, a podrazumeva rešavanje sledećeg problema:

$$(P_0^{\varepsilon}): \quad \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \text{ tako da: } \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2 \leqslant \varepsilon \tag{38}$$

gde je sa $\|\cdot\|_0$ označena l_0-pseudo norma i $\|\cdot\|_p$ l_p-norma. Pored ovog oblika, u literaturi (Aharon et al., 2006; Elad, 2010) se može sresti nešto stroža varijanta, gde nije dopušteno odstupanje reprezentacije signala ($\mathbf{D}\boldsymbol{\alpha}$) od samog signala, a koja podrazumeva rešavanje sledećeg problema:

$$(P_0): \quad \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \text{ tako da: } \mathbf{x} = \mathbf{D}\boldsymbol{\alpha} \tag{39}$$

Ova varijanta postavke problema je privlačna za analizu, pošto se može pokazati da ukoliko je broj nenultih elemenata vektora $\boldsymbol{\alpha}$ manji od polovine naj-

1 U literaturi se pored termina retka reprezentacija mogu sresti i termini retka aproksimacija i retka dekompozicija.

2 Originalni engleski termin je "dictionary learning", ali doslovan prevod "učenje rečnika" deluje neprikladno pošto reč učenje ne podrazumeva proceduru izbora osnovnih oblika čijim je kombinovanjem moguće generisati sve preostale "reči". Alternativni termin estimacija odnosno procena ne deluje adekvatno pošto u stručnoj literaturi na engleskom jeziku njemu odgovaraju drugi termini.

3 U literaturi se za termin rečnik koriste i termini kodna knjiga i alfabet.

manjeg broja kolona rečnika $\mathbf{D}$ koje su linearno zavisne,[4] tada je to rešenje jedinstveno i "najređe"[5] moguće (Elad, 2010). Ovo je posebno interesantno svojstvo, pošto ukoliko se dobije rešenje koje zadovoljava gornji kriterijum, to rešenje je ujedno i globalno rešenje, iako ciljna funkcija (l_0-pseudo norma) koja se minimizuje nije konveksna.

Ukoliko se formulacije problema (P_0^ε) i (P_0) primene na isti signal sa istim zadatim rečnikom, rešenje koje se dobija na osnovu (P_0^ε) je jednako retko ili ređe od rešenja koje se dobija na osnovu (P_0), što je bilo i očekivano pošto su uslovi relaksirani. Sa druge strane, rešenje koje se dobija na osnovu (P_0^ε) nije jedinstveno, pri čemu nejedinstvenost ne podrazumeva samo varijacije u vrednostima nenultih elemenata vektora $\boldsymbol{\alpha}$, već i onih elemenata koji su jednaki nuli uz ograničenje da je broj elemenata koji su različiti od nule fiksan.[6] Svojstvo od interesa za analizu rešenja u slučaju (P_0^ε) jeste njegova stabilnost, odnosno koliko dobijeno rešenje odstupa od idealnog. Intuitivno je jasno da što je manje ε to je odstupanje dobijenog rešenja od stvarnog rešenja manje. Pored toga vrednost odstupanja u velikoj meri zavisi i od sličnosti atoma koji obrazuju rečnik, kao i od retkosti stvarnog rešenja. Odnosno, što su atomi sličniji to je i mogućnost odstupanja dobijenog rešenja od stvarnog veća, jer je moguće lako zameniti slične atome, dok što je stvarno rešenje manje retko (ima više nenultih elemenata) moguće ga je predstaviti na više različitih načina (Elad, 2010).

Još jedna, nešto prirodnija, interpretacija problema (P_0^ε) jeste uklanjanje aditivnog šuma. Polazi se od pretpostavke da se signal $\mathbf{x}$ može predstaviti sa $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}_0 + \mathbf{n}$ gde je $\boldsymbol{\alpha}_0$ redak kod, a $\mathbf{n}$ aditivni šum konačne energije ($\|\mathbf{n}\|_2^2 = \varepsilon^2$). Potrebno je proceniti vrednost $\boldsymbol{\alpha}_0$ i na taj način dobiti signal bez šuma ($\mathbf{D}\boldsymbol{\alpha}_0$).

Pronalaženje retke reprezentacije predstavlja NP-težak problem bez obzira na to koja se formulacija problema koristi, (P_0^ε) ili (P_0), pošto podrazumeva formiranje svih mogućih podskupova atoma (kolona rečnika), čija je kardinalnost jednaka broju nenultih elemenata retkog koda. Stoga se umesto detaljne pretrage svih mogućih kombinacija atoma, pristupa pronalaženju suboptimalnog rešenja. Svi algoritmi za pronalaženje suboptimalnog rešenja se mogu podeliti na 2 velike grupe: pohlepne algoritme i algoritme sa relaksiranim uslovom retkosti.

3.2.1 *Pohlepni algoritmi*

Ideja na kojoj se zasniva ova grupa algoritama jeste da se u svakom koraku donosi odluka koja je lokalno optimalna, pri čemu se zanemaruje opšta situacija.[7] Ovakvi algoritmi su jednostavni kako za dizajn tako i za implementaciju, i pružaju mogućnost rešavanja nekih NP-složenih problema. Iako u nekim slučajevima dovode do globalno optimalnog rešenja, u pri rešavanju mnogih problema su se pokazali kao neefikasni. Pored toga, za svaki pohlepni algoritam

4 U literturi ovaj broj se naziva spark. Može se uočiti sličnost definicije sa rangom matrice, koji predstavlja najveći broj kolona matrice koje su linearno nezavisne, gde je reč najveći zamenjena sa najmanji, a reč nezavisne sa rečju zavisne. Treba napomenuti da se do sparka teže dolazi, pošto podrazumeva kombinatornu pretragu svih mogućih podskupova kolona matrice.

5 Pod najređim rešenjem se podrazumeva rešenje sa najmanjim brojem nenultih elemenata.

6 Ukoliko bi broj nenultih elemenata bio promenljiv tada vektori $\boldsymbol{\alpha}$ čiji je broj veći od minimalnog dobijenog ne predstavljaju rešenja optimizacionog problema.

7 Ovakvo "kratkovido" odlučivanje je karakteristično za pohlepne ljude, što je i poslužilo kao osnov za naziv ove grupe algoritama.

Parametri: Data je matrica $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_K,]$, vektor $\mathbf{x}$ i dozvoljeno odstupanje ε.

1: Inicijalizovati promenljive:
$i = 0$, $\boldsymbol{\alpha}^{(0)} = \mathbf{0}$, $\mathbf{r}^{(0)} = \mathbf{x} - \mathbf{D}\boldsymbol{\alpha}^{(0)} = \mathbf{x}$ i $S = \emptyset$.
2: **Ponavljaj sledeće:**
3: $i = i + 1$.
4: $e(k) = \left\| \mathbf{d}_k \frac{\mathbf{d}_k^T \mathbf{r}^{(i-1)}}{\|\mathbf{d}_k\|_2^2} - \mathbf{r}^{(i-1)} \right\|_2^2$ za svako $k \in \{1, 2, \ldots, K\} \setminus S^{(i-1)}$.
5: $S^{(i)} = S^{(i-1)} \cup \{k^{(i)}\}$ gde je $k^{(i)} = \arg\min_{k \in \{1,\ldots,K\} \setminus S^{(i-1)}} e(k)$.
6: $\boldsymbol{\alpha}_{S^{(i)}}^{(i)} = \left(\mathbf{D}_{S^{(i)}}^T \mathbf{D}_{S^{(i)}} \right)^{-1} \mathbf{D}_{S^{(i)}}^T \mathbf{x}$.
7: $\mathbf{r}^{(i)} = \mathbf{x} - \mathbf{D}\boldsymbol{\alpha}^{(i)}$.
8: **dok važi** $\|\mathbf{r}^{(i)}\|_2 \leqslant \varepsilon$.
9: **Rezultat:** $\boldsymbol{\alpha}^{(i)}$.

Slika 5: Pseudo kod OMP algoritma. Redni broj iteracije je označen sa i i nalazi se u zagradama u eksponentu promenljive. Skup indeksa iskorišćenih atoma $S^{(i)}$ u indeksu retkog koda $\boldsymbol{\alpha}$ i rečnika $\mathbf{D}$ znači da taj vektor odnosno matrica sadrži samo elemente odnosno kolone sa tim indeksima.

je potrebno obezbediti dokaz njegove korektnosti. Algoritmi koju pripadaju ovoj grupi, a koriste se za određivanje retke reprezentacije signala su: traženje poklapanja uz uslov ortogonalnosti[8] (OMP *Orthogonal matching pursuite*), traženje poklapanja (MP *Matching pursuite*) i traženje približnog poklapanja (WMP *Weak matching pursuite*).[9] Ono što je zajedničko za ove algoritme jeste način izbora atoma koji će se koristiti za reprezentaciju konkretnog signala koji se vrši sekvencijalno.

3.2.1.1 *Traženje poklapanja uz uslov ortogonalnosti – OMP*

Formalni opis algoritma u obliku pseudo koda je dat na slici 5. Inicijalno ni jedan atom (kolona rečnika) nije iskorišćen za reprezentaciju signala $\mathbf{x}$, stoga je vektor $\boldsymbol{\alpha}$ jednak nula vektoru, skup izabranih indeksa S je prazan, a vektor greške koja se pri tome pravi (reziduum) $\mathbf{r}$ jednak je samom signalu (korak 1 na slici 5). Algoritam pretrage započinje izračunavanjem minimalnih kvadrata rastojanja od reziduuma $\mathbf{r}$ do vektora koji su kolinearani sa atomima $\mathbf{d}_k$ rečnika $\mathbf{D}$ ($\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_K]$) a koji još nisu iskoršćeni za reprezentaciju signala, odnosno:

$$e(k) = \min_{z_k} \left\| z_k \mathbf{d}_k - \mathbf{r}^{(i-1)} \right\|_2^2 \tag{40}$$

gde je sa i označen redni broj iteracije. Može se pokazati da se minimalno rastojanje dobija ukoliko je $z_k = (\mathbf{d}_k^T \mathbf{r}^{(i-1)})/(\|\mathbf{d}_k\|_2^2)$. U sledećem koraku se skup atoma koji se koristi za reprezentaciju signala proširuje dodavanjem atoma koji je rezultirao najmanjim $e(k)$, odnosno atomom koji zaklapa najmanji ugao sa

8 Ograničenje podrazumeva da je odstupanje rekonstruisanog signala od stvarnog signala ortogonalno na rekonstruisani signal. Ovi detalji su izostavljeni iz naziva algoritma, što je učinjeno i u izvornom engleskom nazivu algoritma.

9 U ovom radu su preuzeti nazivi koji se koriste u obradi signala. U teoriji aproksimacije, nazivi za ove algoritme su ortogonalni pohlepni algoritam (*Orhtogonal greedy algorithm*), čisti pohlepni algoritam (*Pure greedy algorithm*) i slabi pohlepni algoritam (*Weak greedy algorithm*) respektivno.

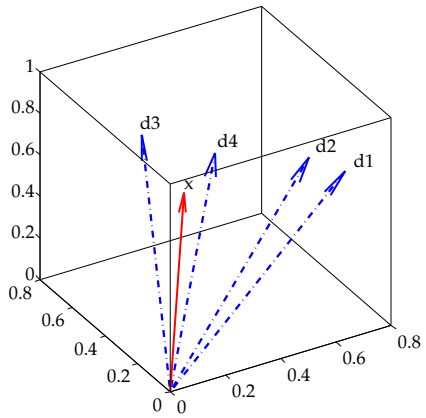
trenutnim reziduumom. Ovo proširenje se formalno realizuje dodavanjem indeksa novoizabranog atoma skupu do tada iskorišćenih indeksa $S^{(i-1)}$ (koraci 4 i 5 na slici 5). Treba primetiti da se izbor atoma svodi na izbor onog atoma čiji je normalizovan vektorski proizvod sa reziduumom najveći. Nakon proširenja skupa izabranih atoma pristupa se izračunavanju vrednosti koeficijenata kojima je potrebno pomnožiti izabrane atome tako da kvadratno odstupanje stvarne vrednosti signala od njegove reprezentacije dato sa $\|\mathbf{x} - \mathbf{D}_{S^{(i)}} \boldsymbol{\alpha}^{(i)}_{S^{(i)}}\|_2^2$ bude minimalno, što se svodi na izračunavanje sledećeg izraza:

$$\boldsymbol{\alpha}^{(i)}_{S^{(i)}} = \left(\mathbf{D}^T_{S^{(i)}} \mathbf{D}_{S^{(i)}}\right)^{-1} \mathbf{D}^T_{S^{(i)}} \mathbf{x}. \tag{41}$$

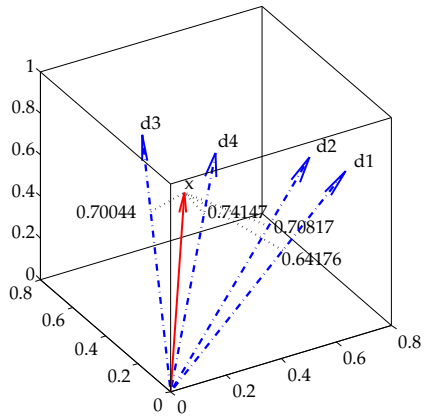
Indeksi koji se nalaze uz rečnik i retki kod ukazuju da oni sadrže samo kolone rečnika odnosno vrednosti koeficijenata čiji su indeksi sadržani u $S^{(i)}$ (korak 6 na slici 5). [10] Treba obratiti pažnju da ovaj korak podrazumeva izračunavanje inverzne kovarijansne matrice dimenzija $i \times i$ koja za veliko i može da bude računski zahtevna. Jedan od načina da se ovo delimično ubrza jeste korišćenjem vrednosti inverzne matrice iz prethodne iteracije, čija je detaljna procedura opisana u (Elad, 2010). Nakon ovoga se vrši izračunavanje novog odstupanja i ukoliko je ono manje od unapred zadatog praga ε procedura se prekida (koraci 7 i 8 na slici 5). Nakon toga se u sledećoj iteraciji po istom principu dodaje novi atom i izračunava novi retki kod, sve dok odstupanje ne postane dovoljno malo (manje od nekog unapred zadatog praga).

Na slici 6 su na jednom jednostavnom primeru ilustrovani pojedini koraci OMP algoritma. Izabrani rečnik sadrži 4 atoma koji su prikazani plavom bojom, dok je signal kojeg je potrebno razložiti prikazan crvenom bojom. Inicijalno svi elementi vektora $\boldsymbol{\alpha}$ su jednaki nuli, tako da je inicijalno odstupanje jednako samom signalu koji treba razložiti. U prvoj iteraciji se vrši projekcija signala na sve atome i za prvi atom koji će se iskoristiti u dekompoziciji signala bira se onaj sa najvećom vrednošću normalizovane projekcije, što je u ovom primeru $\mathbf{d}_4$, jer je vrednost normalizovanog koeficijenta projekcije maksimalna i iznosi 0.74 (videti 6b). Nakon toga se za izabrani atom određuje vrednost koeficijenta kojim ga je potrebno pomnožiti i koji iznosi 0.74. Potom se izračunava novi reziduum, koji predstavlja razliku signala $\mathbf{x}$ i trenutne rekonstrukcije $0.74\mathbf{d}_4$, a koja je na slici 6 prikazana zelenom bojom. Pošto je dobijeno odstupanje veće od dozvoljenog traži se sledeći vektor koji će se iskoristiti za dekompoziciju signala. Ponovo se vrši projekcija reziduuma na vektore iz rečnika koji još nisu iskorišćeni za dekompoziciju ($\{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3\}$) i bira se onaj vektor koji je najbliži odstupanju, što je u ovom slučaju $\mathbf{d}_3$. Za izabrana dva atoma ($\mathbf{d}_3$ i $\mathbf{d}_4$) se izračunavaju vrednosti koeficijenata koji u ovom primeru iznose 0.31 i 0.50 respektivno. Za ove koeficijente rekonstruisani signal $0.31\mathbf{d}_3 + 0.50\mathbf{d}_4$ se u potpunosti poklapa sa signalom $\mathbf{x}$ (videti 6f), te je odstupanje jednako nuli stoga se procedura pretrage završava.

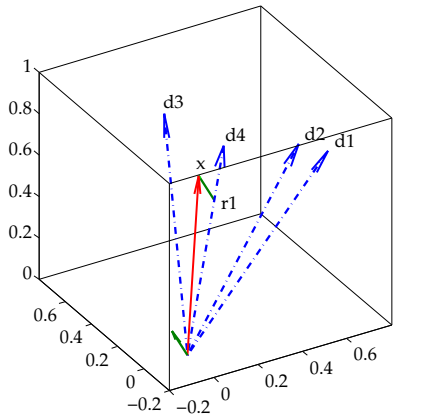
10 Do izraza (41) dolazi se jednostavnim diferenciranjem ciljne funkcije $\|\mathbf{x} - \mathbf{D}_{S^{(i)}} \boldsymbol{\alpha}^{(i)}_{S^{(i)}}\|_2^2$ po svim elementima vektora $\boldsymbol{\alpha}^{(i)}_{S^{(i)}}$ i njenim izjednačavanjem sa nulom. Rezultat je sledeća jednakost $\mathbf{D}^T_{S^{(i)}} \left(\mathbf{D}_{S^{(i)}} \boldsymbol{\alpha}^{(i)}_{S^{(i)}} - \mathbf{x}\right) = \mathbf{0}$. koja se može preurediti u sledeću $-\mathbf{D}^T_{S^{(i)}} \mathbf{r}^{(i)} = \mathbf{0}$, odakle se vidi da je novi reziduum $\mathbf{r}^{(i)}$ ortogonalan na sve atome koji su iskorišćeni za reprezentaciju signala $\mathbf{x}$. Ova osobina je ujedno i odredila ime ovog algoritma.



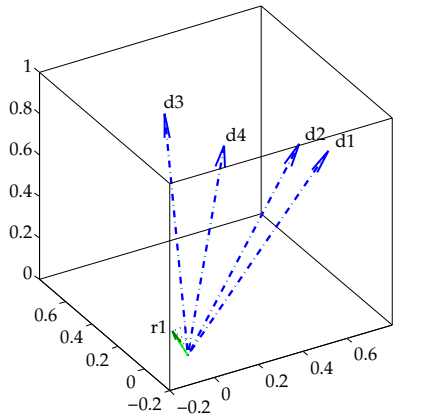
(a) Signala $\mathbf{x}$ i atomi u 3D prostoru.



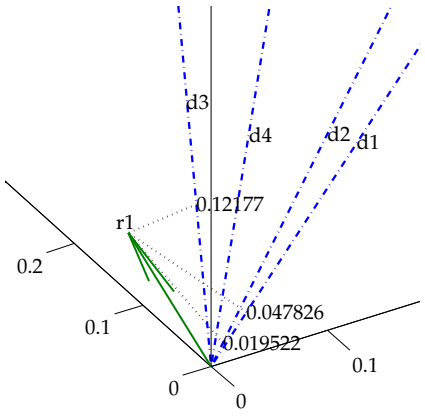
(b) Projekcije odstupanja u 1. iteraciji za $S^{(0)} = \emptyset$ i $\boldsymbol{\alpha}^{(0)} = [0, 0, 0, 0]$.



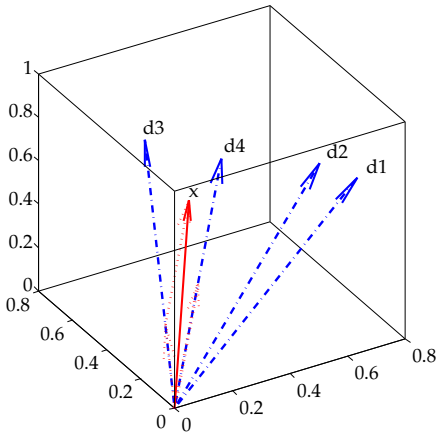
(c) Odstupanje u 1. iteraciji za $S^{(1)} = \{4\}$ i $\boldsymbol{\alpha}^{(1)} = [0, 0, 0, 0.74]$



(d) Projekcije odstupanja u 2. iteraciji za $S^{(1)} = \{4\}$ i $\boldsymbol{\alpha}^{(1)} = [0, 0, 0, 0.74]$



(e) Uvećan segment 6d



(f) Rekonstrukcija signala ($S^{(2)} = \{3, 4\}$ i $\boldsymbol{\alpha}^{(2)} = [0, 0, 0.31, 0.50]$)

Slika 6: Ilustracija OMP algoritma. Plavom bojom (crta-tačka linija) su prikazani svi atomi, crvenom (puna linija) signal, zelenom (puna linija) odstupanje (reziduum). Inicijalno odstupanje je jednako signalu, tako da je ono na slici 6b označeno crvenom bojom. Isprekidanim crvenim linijama su prikazani vektori $\alpha_3 \mathbf{d}_3$ i $\alpha_4 \mathbf{d}_4$, koji u zbiru daju rekonstrukciju signala $\mathbf{x}$.

Parametri: Data je matrica $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_K]$, vektor $\mathbf{x}$ i dozvoljeno odstupanje ε.

1: Inicijalizovati promenljive: $i = 0$, $\boldsymbol{\alpha}^{(0)} = \mathbf{0}$, $\mathbf{r}^{(0)} = \mathbf{x} - \mathbf{D}\boldsymbol{\alpha}^{(0)} = \mathbf{x}$ i $S = \emptyset$.
2: **Ponavljaj sledeće:**
3: $\quad i = i + 1$.
4: $\quad e(k) = \left\| \frac{\mathbf{d}_k^T \mathbf{r}^{(i-1)}}{\|\mathbf{d}_k\|_2^2} \mathbf{d}_k - \mathbf{r}^{(i-1)} \right\|_2^2$ za svako $k \in \{1, 2, \ldots, K\}$.
5: $\quad S^{(i)} = S^{(i-1)} \cup \{k^{(i)}\}$ gde je $k^{(i)} = \arg\min_{k \in \{1,2,\ldots,K\}} e(k)$.
6: $\quad \alpha_{k^{(i)}} = \alpha_{k^{(i)}} + \frac{\mathbf{d}_{k^{(i)}}^T \mathbf{r}^{(i-1)}}{\|\mathbf{d}_{k^{(i)}}\|_2^2}$.
7: $\quad \mathbf{r}^{(i)} = \mathbf{r}^{(i-1)} - \frac{\mathbf{d}_{k^{(i)}}^T \mathbf{r}^{(i-1)}}{\|\mathbf{d}_{k^{(i)}}\|_2^2} \mathbf{d}_{k^{(i)}}$.
8: **dok važi** $\|\mathbf{r}^{(i)}\|_2 \leqslant \varepsilon$.
9: **Rezultat:** $\boldsymbol{\alpha}^{(i)}$.

Slika 7: Pseudo kod MP algoritma. Redni broj iteracije je označen sa i i nalazi se u zagradama u eksponentima promenljivih. Treba napomenuti da se vrednost $(\mathbf{d}_k^T \mathbf{r}^{(i-1)})/\|\mathbf{d}_k\|_2^2$ izračunava samo jedanput u jednoj iteraciji algoritma, ali za sve atome koji čine rečnik, kao da u svakoj novoj iteraciji ne mora da dođe do povećanja skupa korišćenih indeksa $S^{(i)}$.

3.2.1.2 *Traženje poklapanja – MP*

Algoritam traženja poklapanja je vrlo sličan prethodno opisanom OMP algoritmu, ali postoje bitne razlike koje ga čine računski jednostavnijim, ali nažalost i manje tačnim. Formalni opis algoritma u obliku pseudo koda je dat na slici 7. Kao što se iz priloženog može videti, osnovna razlika je u načinu izračunavanju koeficijenata (korak 6 na slici 7), pošto se u svakoj iteraciji menja vrednost samo jednog koeficijenta, i to koeficijenta onog atoma koji zaklapa najmanji ugao sa trenutnim reziduumom (onog sa najvećim $(\mathbf{d}_k^T \mathbf{r})/\|\mathbf{d}_k\|_2^2$). Na ovaj način je izbegnuta potreba za izračunavanjem inverzne matrice $(\mathbf{D}_{S^{(i)}}^T \mathbf{D}_{S^{(i)}})^{-1}$, kao i potreba za reestimacijom svih preostalih nenultih koeficijenata vektora $\boldsymbol{\alpha}$ što omogućava i pojednostavljeno izračunavanje reziduuma tako što se od trenutne vrednosti oduzima projekcija trenutnog reziduuma na izabrani atom $\mathbf{d}_k$ odnosno:

$$\mathbf{r}^{(novo)} = \mathbf{r}^{(tren.)} - \frac{\mathbf{d}_k^T \mathbf{r}^{(tren.)}}{\|\mathbf{d}_k\|_2^2} \mathbf{d}_k$$

Cena koja je plaćena za ova ubrzanja jeste da se greška $e(k)$ u svakoj iteraciji mora izračunavati za sve atome (kako neiskorišćene tako i već iskorišćene za reprezentaciju signala), i udređenoj meri gubitak tačnosti.

3.2.1.3 *Traženje približnog poklapanja – WMP*

Prethodni opisani MP algoritam pretrage se može dodatno ubrzati tako što se umesto atoma koji zaklapa najmanji mogući ugao sa trenutnim reziduumom izabere prvi atom koji sa reziduumom zaklapa ugao koji je manji od nekog unapred zadatog praga. Formalni opis algoritma u obliku pseudo koda je dat

Parametri: Data je matrica $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_K]$, vektor $\mathbf{x}$, dozvoljeno odstupanje ε i prag t.

1: Inicijalizovati promenljive: $i = 0$, $\boldsymbol{\alpha}^{(0)} = \mathbf{0}$, $\mathbf{r}^{(0)} = \mathbf{x} - \mathbf{D}\boldsymbol{\alpha}^{(0)} = \mathbf{x}$ i $S = \emptyset$.
2: **Ponavljaj sledeće:**
3: $i = i + 1$.
4: $e(k) = \left\| \frac{\mathbf{d}_k^T \mathbf{r}^{(i-1)}}{\|\mathbf{d}_k\|_2^2} \mathbf{d}_k - \mathbf{r}^{(i-1)} \right\|_2^2$ za svako $k \in \{1, \ldots, k_t^{(i)}\}$ gde je $k_t^{(i)}$ najmanje k za koje važi $\mathbf{d}_k^T \mathbf{r}^{(i-1)} / \|\mathbf{d}_k\|_2 \geqslant t \|\mathbf{r}^{(i-1)}\|_2$.
5: $S^{(i)} = S^{(i-1)} \cup \{k^{(i)}\}$ gde je $k^{(i)} = k_t^{(i)}$ ukoliko postoji $k_t^{(i)}$ odnosno $k^{(i)} = \arg\min_{k \in \{1,2,\ldots,K\}} e(k)$ ukoliko $k_t^{(i)}$ ne postoji.
6: $\alpha_{k^{(i)}} = \alpha_{k^{(i)}} + \frac{\mathbf{d}_{k^{(i)}}^T \mathbf{r}^{(i-1)}}{\|\mathbf{d}_{k^{(i)}}\|_2^2}$.
7: $\mathbf{r}^{(i)} = \mathbf{r}^{(i-1)} - \frac{\mathbf{d}_{k^{(i)}}^T \mathbf{r}^{(i-1)}}{\|\mathbf{d}_{k^{(i)}}\|_2^2} \mathbf{d}_{k^{(i)}}$.
8: **dok važi** $\|\mathbf{r}^{(i)}\|_2 \leqslant \varepsilon$.
9: **Rezultat:** $\boldsymbol{\alpha}^{(i)}$.

Slika 8: Pseudo kod WMP algoritma. Redni broj iteracije je označen sa i i nalazi se u zagradama u eksponentima promenljivih.

na slici 8. Ideja za ubrzanje procedure se zasniva na Koši-Švarcovoj (*Cauchy-Schwarz*) nejednakosti,[11] odnosno činjenici da je:

$$\frac{(\mathbf{d}_k^T \mathbf{r}^{(i-1)})^2}{\|\mathbf{d}_k\|_2^2} \leqslant \|\mathbf{r}^{(i-1)}\|_2^2$$

što važi za svaki atom, pa i onaj sa minimalnom mogućom greškom $e(k)$. Pošto važi ekvivalencija između problema minimizacije $e(k)$ i maksimizacije normalizovanog unutrašnjeg proizvoda reziduuma i atoma pretraga za lokalno optimalnim atomom se može ubrzati izborom atoma za koji je ispunjen uslov:

$$\mathbf{d}_{k_t^{(i)}}^T \mathbf{r}^{(i-1)} / \|\mathbf{d}_{k_t^{(i)}}\|_2 \geqslant t \|\mathbf{r}^{(i-1)}\|_2$$

pri čemu je t neki broj iz intervala $(0, 1]$. Pomoću vrednosti t moguće je kontrolisati nivo greške, koji će se praviti pri aproksimaciji. Što je t bliže jedinici to je greška koja se pravi pogrešnim izborom atoma manja, a u slučaju da je $t = 1$ WMP se svodi na MP algoritam.

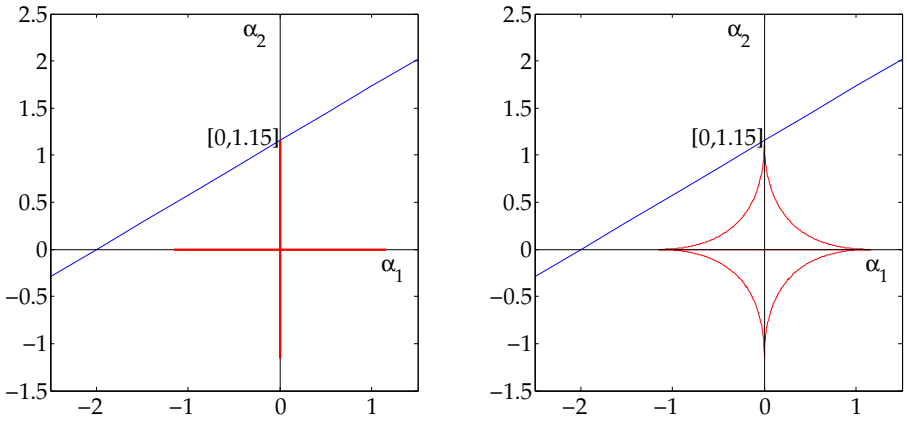
3.2.2 *Relaksacija l_0-pseudo norme*

Izbor l_0-pseudo norme kao mere retkosti rešenja je prilično intuitivan, jer ona predstavlja broj nenultih elemenata u nekom vektoru. Nažalost l_0-pseudo norma nije kontinualna funkcija, stoga se zamenjuje kontinualnim ili čak glatkim aproksimacijama, tako da su problemi koje treba rešiti umesto (P_0^ε) i (P_0) sledeći:

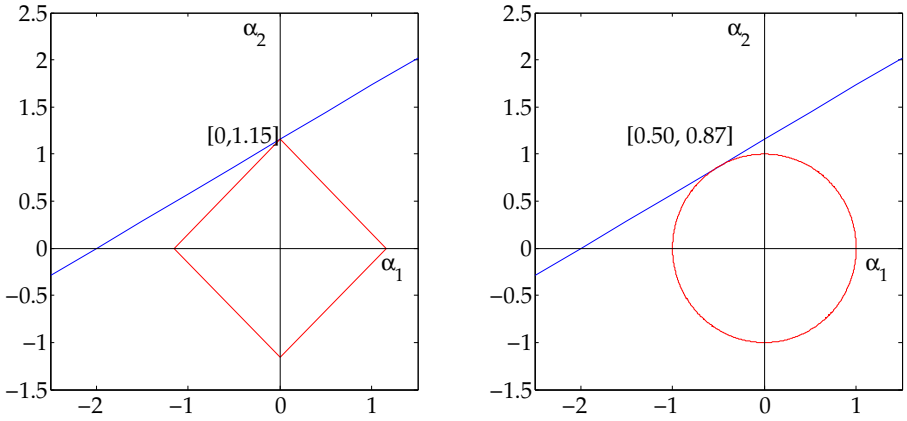
$$(P_p^\varepsilon): \quad \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_p^p \text{ tako da } \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2 \leqslant \varepsilon \tag{42}$$

$$(P_p): \quad \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_p^p \text{ tako da } \mathbf{D}\boldsymbol{\alpha} = \mathbf{x} \tag{43}$$

11 Za proizvoljni par vektora $\mathbf{x}$ i $\mathbf{y}$ važi sledeća nejednakost $(\mathbf{x}^T \mathbf{y})^2 \leqslant \mathbf{x}^T \mathbf{x} \cdot \mathbf{y}^T \mathbf{y}$



(a) $p = 0.05$ (b) $p = 0.5$



(c) $p = 1$ (d) $p = 2$

Slika 9: Rešenje problema $\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_p$ tako da $[\frac{1}{\sqrt{3}}, 1]\boldsymbol{\alpha} = \frac{2}{\sqrt{3}}$, za različitie vrednosti p.

pri čemu je $p \in (0, 1]$.[12]

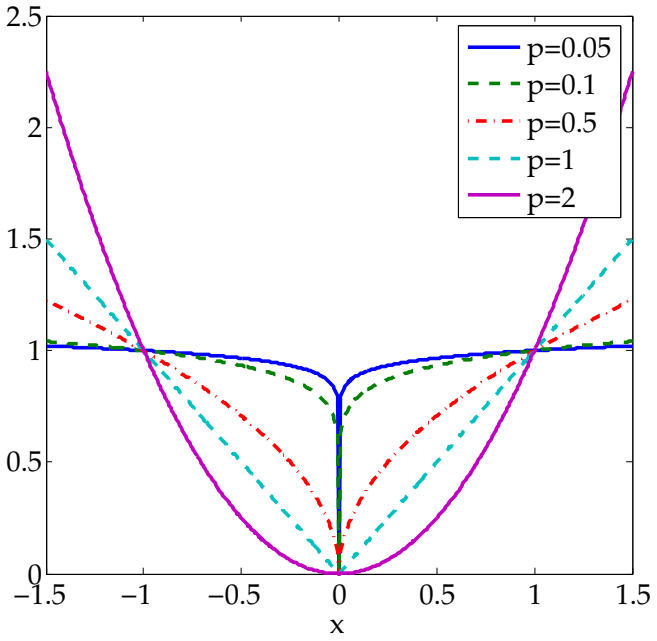
Jedan od načina da se pokaže da se smanjivanjem vrednosti p favorizuju retka rešenja jeste jednostavnim poređenjem normi. Neka vektor $\boldsymbol{\alpha}$ ima a nenultih elemenata i neka mu je l_p-norma jednaka 1. Potrebno je naći vektor sa najmanjom l_q-normom, pri čemu je $q < p$, što se svodi na rešavanje sledećeg optimizacionog problema:

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_q^q \text{ tako da } \|\boldsymbol{\alpha}\|_p^p = 1 \tag{44}$$

Bez gubitka opštosti, a u cilju pojednostavljenja zapisa, uvodi se pretpostavka da su nenulti elementi na početku vektora i da su svi pozitivni tako da odgovarajući Lagranžijan ima sledeći oblik:

$$\mathcal{L}(\boldsymbol{\alpha}) = \|\boldsymbol{\alpha}\|_q^q + \lambda\left(\|\boldsymbol{\alpha}\|_p^p - 1\right) = -\lambda + \sum_{k=1}^{a}\left(|\alpha_k|^q + \lambda|\alpha_k|^p\right) \tag{45}$$

12 Ukoliko je $0 < p < 1$, $\|\boldsymbol{\alpha}\|_p$ nije norma pošto ne važi jednakost trougla. Sa druge strane za l_0-pseudo normu važi jednakost trougla, ali ne važi homogenost ($\|c\boldsymbol{\alpha}\|_0 = \|\boldsymbol{\alpha}\|_0 \neq c\|\boldsymbol{\alpha}\|_0$).



Slika 10: Izgled funkcije $|x|^p$ za različite vrednosti p.

Pošto je funkcija separabilna, odnosno svaki element vektora je moguće nezavisno tretirati od drugih, optimum se dobija za $\alpha_k^{p-q} = \frac{-q}{\lambda p} = const$[13] za svako $k = 1, \ldots, a$. Na osnovu prethodnog i činjenice da je $\|\boldsymbol{\alpha}\|_p^p = 1$ sledi da je $\alpha_k = a^{-1/p}$, tako da je l_q-norma data sa $\|\boldsymbol{\alpha}\|_q^q = a^{1-q/p}$. Pošto je $q < p$, sledi da se najmanja l_q-norma dobija za $a = 1$, odnosno u slučaju da vektor $\boldsymbol{\alpha}$ ima samo jedan nenulti element.

Drugi način kojim se može pokazati da se smanjivanjem p favorizuju retka rešenja je geometrijski. Rešavanje problema (P_p) se može interpretirati kao postepeno širenje l_p-lopte[14] sa centrom u koordinatnom početku, sve dok lopta ne dodirne hiperravan definisanu jednačinom $\mathbf{D}\boldsymbol{\alpha} = \mathbf{x}$. Ukoliko je $p < 1$ l_p-lopta nije konveksna, tako da se favorizuju rešenja koja su bliža "uglovima" lopte, što vodi retkim rešenjima. Slično važi i za l_1-loptu, koja je konveksna, ali favorizuje retka rešenja. Na slici 9 je ilustrovan izgled l_p-lopti za trivijalan slučaj jednodimenzionalnog prostora (matrica $\mathbf{D}$ ima samo jednu vrstu). Kao što se iz priloženog može videti što je p bliže nuli l_p-loptu favorizuje rešenja koja su bliža dijagonalama, za razliku od njih l_2-lopta favorizuje rešenje koje je bliže koordinatnom početku. Sa druge strane, umesto otvorene lopte može da se posmatra i sama funkcija $|x|^p$, koja je za jednodimenzionalni slučaj ilustrovana na slici 10. Može se uočiti da kako p teži nuli tako $|x|^p$ teži indikatorskoj funkciji koja je jednaka 0 za $x = 0$ i jednaka 1 za $x \neq 0$, odnosno l_p-norma teži l_0-pseudo normi.

13 Do ovog izraza se dolazi jednostavnim diferenciranjem Lagranžijana datog jednačinom (45) po svakom koeficijentu α_k.

14 Pod l_p-loptom poluprečnika r sa centrom u tački $\mathbf{x}_0$ se podrazumevaju sve tačke za koje važi $\|\mathbf{x} - \mathbf{x}_0\|_p^p \leqslant r$.

Parametri: Data je matrica $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_K]$, vektor $\mathbf{x}$ i dozvoljeno odstupanje ε.

1: Inicijalizovati promenljive: $i = 0$, $\boldsymbol{\alpha}^{(0)} = \mathbf{1}$ i $\boldsymbol{W}^{(0)} = \mathrm{diag}(\boldsymbol{\alpha}^{(0)}) = \mathbf{I}$
2: **Ponavljaj sledeće:**
3: $\quad i = i + 1$.
4: $\quad \boldsymbol{\alpha}^{(i)} = \left(\boldsymbol{W}^{(i-1)}\right)^2 \mathbf{D}^T \left(\mathbf{D}\left(\boldsymbol{W}^{(i-1)}\right)^2 \mathbf{D}^T\right)^+ \mathbf{x}$.
5: $\quad W_{jj}^{(i)} = |\alpha_j^{(i)}|^{1-p/2}$ za svako $j \in \{1, \ldots, K\}$
6: **dok važi** $\|\boldsymbol{\alpha}^{(i)} - \boldsymbol{\alpha}^{(i-1)}\|_2 \leqslant \varepsilon$.
7: **Rezultat:** $\boldsymbol{\alpha}^{(i)}$.

Slika 11: Pseudo kod FOCUSS algoritma zasnovan na IRLS proceduri. Redni broj iteracije je označen sa i i nalazi se u zagradama u eksponentima promenljivih. Karakter '+' koji se nalazi u eksponentu promenljivih označava pseudo-inverziju, odnosno $\mathbf{X}^+ = \mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}$

3.2.2.1 *FOCUSS algoritam*

Primer algoritma gde je l_0-pseudo norma zamenjena sa l_p-normom pri čemu je $p \in (0, 1]$ jeste algoritam za usmereno rešavanje neodređenog sistema jednačina (FOCUSS **foc**al **u**ndetermined **s**ystem **s**olver) (Gorodnitsky i Rao, 1997). U ovom algoritmu se l_p-norma za neko fiksno p iz intervala $(0, 1]$, aproksimirana ponderisanom l_2-normom,[15] tako da se za njeno rešavanje koristi iterativna aproksimacija pomoću ponderisanih najmanjih kvadrata (IRLS *Iterative reweighted least square*).

Formalna procedura za FOCUSS algoritam je u obliku pseudo koda prikazana na slici 11. Inicijalno su vrednosti svih koefijenata kao i težina jednake 1 (korak 1 na slici 11). Iterativna procedura je relativno jednostavna i sastoji se od dva koraka. Prvi korak (korak 4 na slici 11) podrazumeva izračunavanje koeficijenata korišćenjem jednakosti:

$$\boldsymbol{\alpha} = (\boldsymbol{W})^2 \mathbf{D}^T \left(\mathbf{D}(\boldsymbol{W})^2 \mathbf{D}^T\right)^+ \mathbf{x} \tag{46}$$

gde je sa $\boldsymbol{W}$ označena dijagonalna težinska matrica, a sa znakom '+' u eksponentu pseudo-inverzija matrice. [16] Do izraza (46) se dolazi jednostavno, diferenciranjem Lagranžijana

$$\mathcal{L}(\boldsymbol{\alpha}) = \|\boldsymbol{W}'\boldsymbol{\alpha}\|_2^2 + \lambda^T(\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}) \tag{47}$$

po $\boldsymbol{\alpha}$ i izjednačavanjem dobijenog prvog izvoda sa nulom uz korišćenje činjenice da je $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$. U izrazu (47) sa $\boldsymbol{W}'$ je označena pseudo-inverzija dijagonalne matrice koja je definisana na sledeći način $W'_{j,j} = 1/W_{j,j}$ ukoliko je $W_{j,j} \neq 0$ odnosno $W'_{j,j} = 0$ ako je $W_{j,j} = 0$. Drugi korak (korak 5 na slici 11) podrazumeva preračun težina komponenata (vrednosti matrice $\boldsymbol{W}$ na glavnoj dijagonali). Težina svake komponente je srazmerna njenoj apsolutnoj vrednosti,

15 Veza između l_2-norme i l_p-norme je sledeća: Neka je $\boldsymbol{\alpha}$ tekuće aproksimativno rešenje i $\mathbf{A} = \mathrm{diag}(|\boldsymbol{\alpha}|^q)$. Uz pretpostavku da je $\mathbf{A}$ invertibilna matrica tada važi $\|\mathbf{A}^{-1}\boldsymbol{\alpha}\|_2^2 = \|\boldsymbol{\alpha}\|_{2-2q}^{2-2q}$. Ukoliko se izabere $q = 1 - p/2$ dobija se da je $\|\mathbf{A}^{-1}\boldsymbol{\alpha}\|_2^2 = \|\boldsymbol{\alpha}\|_p^p$. Pošto matrica $\mathbf{A}$ sadrži i nulte elemente na glavnoj dijagonali, koristi se pseudo-inverzija, koja podrazumeva inverziju elemenata koji su različiti od nule, a elementi koji su jednaki nuli ostaju nepromenjeni.

16 Pseudo-inverzija matrice $\mathbf{X}$ definisana je sledećim izrazom: $\mathbf{X}^+ = \mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1}$.

odnosno $W_{j,j} = |\alpha_j|^{1-p/2}$. Procedura se ponavlja sve dok razlika vrednosti koeficijenata u dve uzastopne iteracije ne postane dovoljno mala. Ovaj algoritam se pokazao efikasnim u praksi, odnosno lokalni minimum do kog je algoritam doveo ujedno je i globalni minimum, ali okolnosti pod kojima se to dešava su uglavnom nepoznate (Elad, 2010; Burdge et al., 2010). Treba napomenuti da se prikazana IRLS procedura uz neznatne modifikacije može primeniti i u slučaju aproksimacije sa l_1-normom uz uslov $\|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2 < \varepsilon$, što je detaljno opisano u (Elad, 2010).

3.2.2.2 *LASSO problem*

U cilju pojednostavljenja optimizacione funkcije najčešće se l_0-pseudo norma zamenjuje sa l_1-normom, te se problem svodi na sledeći:

$$(P_1^{\varepsilon}): \quad \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \text{ tako da } \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2 \leqslant \varepsilon \tag{48}$$

Ovaj problem je "ekvivalentan" problemima:

$$(P_{1'}^{\varepsilon}): \quad \min_{\boldsymbol{\alpha}} \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 \text{ tako da } \|\boldsymbol{\alpha}\|_1 \leqslant d \tag{49}$$

i

$$(L_{1'}): \quad \min_{\boldsymbol{\alpha}} \left(\frac{1}{2}\|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1 \right), \tag{50}$$

pošto za svako ε postoje vrednosti d i λ takve da su rešenja sva tri problema identična. Isto važi i za druga dva parametra, odnosno za svaku vrednost nekog od druga dva parametra (d i λ) postoje vrednosti preostalih parametara tako da su rešenja sva tri problema identična. Nažalost, u praksi ova ekvivalencija i ne stoji pošto nije poznata veza između parametara ε, λ i d (Mairal, 2010). Problem ($L_{1'}$) je u oblasti mašinskog učenja poznat pod nazivom LASSO što je skraćeno od *least absolulte shrinkage and selection operator*, a u oblasti obrade signala pod imenom pretraživanje baze (*Basis pursuit*). Jedna od veoma uspešnih tehnika za rešavanje ovog problema je regresija najmanjih uglova (LARS *Least angle regression*) čija je procedura u obliku pseudo koda prikazana na slici 12.

Skup sub-gradijenta[17] ciljne funkcije definisane jednačinom (50) čine vektori sledećeg oblika:

$$\left\{ \mathbf{D}^T (\mathbf{D}\boldsymbol{\alpha} - \mathbf{x}) + \lambda z \right\} \ \forall z_i = \begin{cases} 1 & x_i > 0 \\ [-1, 1] & x_i = 0 \\ -1 & x_i < 0 \end{cases} \tag{51}$$

Pošto za bilo koju konveksnu funkciju važi da se globalni minimum nalazi u tački čiji skup sub-gradijenata sadrži vektor $\mathbf{0}$, rešenje problema ($L_{1'}$) podrazumeva rešavanje sistema jednačina datog izrazom:

$$\mathbf{D}^T (\mathbf{D}\boldsymbol{\alpha} - \mathbf{x}) + \lambda z = \mathbf{0} \tag{52}$$

17 Skup sub-gradijenta predstavlja uopštenje gradijenta za funkcije koje nisu glatke. Za funkciju $f(\mathbf{x})$ u tački $\mathbf{x}_0$ definiše se kao skup svih mogućih vektora $\{v\}$ za koje važi $f(\mathbf{x}) - f(\mathbf{x}_0) \geqslant \mathbf{v}^T(\mathbf{x} - \mathbf{x}_0)$ za dovoljno malu okolinu $\|\mathbf{x} - \mathbf{x}_0\|_2 \leqslant \delta$. Odgovarajuća geometrijska interpretacija sub-gradijenta je da predstavlja skup svih tangentnih ravni koje prolaze kroz tačku $\mathbf{x}_0$ i ograničavaju konveksnu funkciju $f(\mathbf{x})$ sa donje strane u tački $\mathbf{x}_0$.

Parametri: Data je matrica $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_3, \ldots, \mathbf{d}_K]$, vektor $\mathbf{x}$ i maksimalan broj nenultih koeficijenata d.

1: Inicijalizovati promenljive: $\lambda = \|\mathbf{D}^T\mathbf{x}\|_\infty$, $\boldsymbol{\alpha}(\lambda) = \mathbf{0}$ i $S = \{k \in \{1, 2, \ldots, K\}; \mathbf{d}_j^T\mathbf{x} = \lambda\}$
2: **Ponavljaj sledeće:**
3: Smanjiti λ.
4: $\boldsymbol{\alpha}_S(\lambda) = \left(\mathbf{D}_S^T\mathbf{D}_S\right)^{-1}\left(\mathbf{D}_S^T\mathbf{x} - \lambda\,\mathrm{sign}(\boldsymbol{\alpha}_S(\lambda))\right)$ i $\boldsymbol{\alpha}_{S^C}(\lambda) = \mathbf{0}$.
5: **Ako** $k \in S^C \wedge \mathbf{d}_k^T(\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}) = \lambda$ **onda**
$S = S \cup \{k\}$ i $S^C = S^C \setminus \{k\}$
6: **ili ako** $k \in S^C \wedge \alpha_k = 0$ **onda**
$S^C = S^C \cup \{k\}$ i $S = S \setminus \{k\}$
7: **Kraj ako**
8: **dok važi** $\lambda > 0$.
9: **Rezultat:** $\boldsymbol{\alpha}(\lambda)$ za koje važi: $\|\boldsymbol{\alpha}(\lambda)\|_0 \leqslant d$ i $\min_\lambda \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}(\lambda)\|_2^2$.

Slika 12: Pseudo kod LARS algoritma. Slovo C u eksponentu označava da je u pitanju komplement skupa. Oznaka skupa u indeksu promenljivih $\boldsymbol{\alpha}$ i $\mathbf{D}$ ukazuje da takav vektor odnosno matrica sadrže samo elemente odnosno kolone čiji su indeksi sadržani u tom skupu. Pretpostavka je da se u svakom koraku najviše jedan element, može razmeniti između skupova (S i S^C). Ovde je naveden samo jedan od nekoliko mogućih načina za izbor optimalnog $\boldsymbol{\alpha}$.

i po $\boldsymbol{\alpha}$ i po z. Treba primetiti da rešenje koje se dobija za $\boldsymbol{\alpha}$ i z zavisi od vrednosti parametra λ. Ukoliko se pusti da $\lambda \to \infty$, dominantan član pri minimizaciji ($L_{1'}$) će biti $\|\boldsymbol{\alpha}\|_1$, koji je minimalan ukoliko je $\boldsymbol{\alpha} = \mathbf{0}$. Pošto je $\boldsymbol{\alpha} = \mathbf{0}$ vrednosti elemenata vektora z su iz intervala $[-1, 1]$, a do njih se dolazi na osnovu jednačine

$$z = \frac{1}{\lambda}\mathbf{D}^T\mathbf{x} \tag{53}$$

koja sledi iz jednačine (52). Rešenje $\boldsymbol{\alpha} = \mathbf{0}$ je optimalno sve dok je $\lambda \geqslant \|\mathbf{D}^T\mathbf{x}\|_\infty$[18] pošto su svi elementi vektora z iz opsega $[-1, 1]$. Bez gubitka opštosti može se pretpostaviti da za k-tu kolonu matrice važi $\mathbf{d}_k^T\mathbf{x} = \|\mathbf{D}^T\mathbf{x}\|_\infty$, tada ukoliko vrednost λ postane neznatno manja od $\|\mathbf{D}^T\mathbf{x}\|_\infty$ apsolutna vrednost z_k na osnovu izraza (53) postaje veća od 1 što nije dozvoljena vrednost. Vrednost z_k u tom slučaju treba da bude $\mathrm{sign}(\alpha_k)$, odnosno α_k treba da se razlikuje od 0. Pošto je samo α_k različito od nula jednačina (52) se svodi na

$$\mathbf{D}^T\left(\alpha_k\mathbf{d}_k - \mathbf{x}\right) + \lambda z = \mathbf{0} \tag{54}$$

odnosno nova vrednost za α_k se izračunava na osnovu izraza:

$$\alpha_k = \frac{\mathbf{d}_k^T\mathbf{x} - \lambda\mathrm{sign}(\alpha_k)}{\mathbf{d}_k^T\mathbf{d}_k} \tag{55}$$

Pošto je $\lambda < \mathbf{d}_k^T\mathbf{x}$ sledi da je $\mathrm{sign}(\alpha_k) = \mathrm{sign}(\mathbf{d}_k^T\mathbf{x})$. Sve preostale vrednosti elemenata vektora z se dobijaju na osnovu

$$z = \frac{1}{\lambda}\mathbf{D}^T\left(\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\right) \tag{56}$$

18 Važi sledeće: $\|\mathbf{x}\|_\infty = \max_i |x_i|$.

što sledi direktno iz jednačine (52).

Daljim smanjivanjem vrednosti λ apsolutne vrednosti elemenata vektora z i α_k rastu. Zbog porasta apsolutne vrednosti elemenata z za neko $m \neq k$ vrednost može izaći iz opsega $[-1, 1]$, te je treba zameniti sa $\text{sign}(\alpha_m)$. Ukoliko sa S označimo skup koji sadrži indekse elemenata vektora $\boldsymbol{\alpha}$ koji su različiti od nule, a sa $\mathbf{y}_S$ i $\mathbf{Y}_S$ podvektore i podmatrice vektora $\mathbf{y}$ i matrice $\mathbf{Y}$ koji sadrže samo elemente odnosno kolone čiji su indeksi sadržani u S, jednačina na osnovu koje se vrši izračunavanje nenultih elemenata vektora $\boldsymbol{\alpha}$ je sledeća:

$$\boldsymbol{\alpha}_S = \left(\mathbf{D}_S^T \mathbf{D}_S\right)^{-1} \left(\mathbf{D}_S^T \mathbf{x} - \lambda \text{sign}(\boldsymbol{\alpha}_S)\right) \tag{57}$$

Način izračunavanja vrednosti elemenata vektora z zavisi od toga da li je indeks elementa u skupu S, te $z_S = \text{sign}(\boldsymbol{\alpha}_S)$ odnosno $z_{S^C} = \mathbf{D}_{S^C}^T(\mathbf{x} - \mathbf{D}\boldsymbol{\alpha})/\lambda$. Dalje smanjivanje λ uzrokuje proširenje skupa S, ali zbog načina izračunavanja $\boldsymbol{\alpha}_S$ (videti jednačinu (57)) pojedini elementi vektora $\boldsymbol{\alpha}_S$ mogu postati nula. U tom slučaju potrebno je taj element izbaciti iz skupa S i odgovarajuću vrednost elementa z izračunati na osnovu jednačine (56).

Procedura pretrage se završava kada se ispitaju rešenja za sve vrednosti λ. Pošto λ uzima vrednosti iz skupa $(0, \|\mathbf{D}^T\mathbf{x}\|_\infty)$, što obuhvata beskonačno mnogo različitih vrednosti, u praksi se vrednost λ menja u koracima, a ne kontinualno, da bi se obezbedila traktabilnost algoritma. Kao rešenje se bira onaj vektor $\boldsymbol{\alpha}$ čiji je broj nenultih elemenata manji ili jednak specificiranom maksimalnom broju nenultih elemenata i to za ono λ za koje se postiže najmanje kvadratno odstupanje stvarne vrednosti od rekonstruisane vrednosti. Ovo nije jedini mogući način izbora "optimalnih" koeficijenata $\boldsymbol{\alpha}$.

Iako to prethodno nije eksplicitno navedeno, treba primetiti da LARS procedura pretpostavlja sledeća dva uslova:

- da za neku vrednost λ skup S se ili ne menja ili ako se menja tada se menja samo za jedan element koji se ili dodaje ili uklanja;
- da je matrica $\mathbf{D}_S^T\mathbf{D}_S$ uvek invertibilna, odnosno da je preslikavanje $\lambda \to \boldsymbol{\alpha}(\lambda)$ jedinstveno.

U praktičnim primerima nezadovoljenje prvog uslova je posledica konačne preciznosti izračunavanja, što dovodi do neuspešne realizacije algoritma, ali su takve situacije malo verovatne (Mairal, 2010). Invertibilnost matrice $\mathbf{D}_S^T\mathbf{D}_S$ se obezbeđuje dodavanjem člana kojim se ograničava "energija" retke reprezentacije $\frac{\gamma}{2}\|\boldsymbol{\alpha}\|_2^2$, čime se menja matrica koju je potrebno invertovati, a koja je oblika $\mathbf{D}_S^T\mathbf{D}_S + \gamma\mathbf{I}$ i invertibilna je već i za male vrednosti γ.

3.2.3 *Kvalitet dobijenih aproksimacija*

Ni jedan od prethodno opisanih algoritama ne predstavlja rešenje problema (P_0^ε), već njegovu aproksimaciju, tako da je opravdano postaviti pitanje koliko su one dobre. Test na veštački generisanim podacima u vektorskom prostoru dimenzije 30, sa rečnikom koji sadrži 50 atoma, čiji su rezultati prikazani u (Elad, 2010) je pokazao sledeće:

- Kardinalnost[19] rešenja dobijenog pomoću OMP je bliža stvarnoj nego onog što se dobija pomoću LARS.

19 Pod kardinalonošću se podrazumeva broj atoma koji se koriste za reprezentaciju signala.

- Verovatnoća pogrešnog izbora atoma[20] u slučaju kada je kardinalnost manja od 8 je značajno manja ako se koristi OMP algoritam nego LARS algoritam. U slučaju da je kardinalnost 3 ili manja OMP skoro da ne greši pri izboru atoma, dok u slučaju da je kardinalnost veća od 10 verovatnoća pogrešnog izbora je preko 45% za oba algoritma.
- Kvadratno odstupanje rekonstruisanog signala od originalnog je za kardnalnost manju od 6 značajno manje u slučaju OMP algoritma u odnosu na slučaj LARS algoritma, dok je za kardinalnost veću od 8 situacija obrnuta.

Odavde sledi da OMP ima smisla koristiti u slučaju da je kardinalnost 6 ili manja, u svim ostalim slučajevima je bolje koristiti LARS. Interesantno je da za veliku kardinalnost ni jedan od algoritama (ni OMP ni LARS) u značajnom broju slučajeva ne izabere odgovarajuće atome, pri čemu ovaj pogrešan izbor u slučaju LARS ne utiče u značajnoj meri na srednje kvadratno odstupanje. Što se tiče računske složenosti LARS iako na prvi pogled deluje nešto komplikovaniji od OMP, pošto podrazumeva izračunavanje za svako λ, zbog pretrage nad skupom diskretnih vrednosti, računska složenost je približno ista.

3.3 FORMIRANJE REČNIKA

U prethodnom odeljku je predstavljeno nekoliko algoritama za određivanje retke reprezentacije u slučajevima kada je rečnik unapred dat odnosno poznat. U ovom odeljku će biti predstavljeni osnovni algoritmi kojima se formira rečnik na osnovu raspoloživih signala. Ovi algoritmi se kao i algoritmi za retko kodovanje mogu podeliti na dve velike grupe u zavisnosti od toga da li se vrši minimizacija l_0-pseudo norme ili l_1-norme.

3.3.1 *Formiranje rečnika uz l_0 regularizaciju*

Problem koji treba rešiti pri formiranju rečnika uz l_0 regularizaciju je sledeći:

$$\min_{\mathbf{D}\in\mathbb{R}^{D\times K},\mathbf{A}\in\mathbb{R}^{K\times N_o}} \frac{1}{N_o}\sum_{i=1}^{N_o}\frac{1}{2}\|\boldsymbol{x}_i-\mathbf{D}\boldsymbol{\alpha}_i\|_2^2 \text{ tako da } \|\boldsymbol{\alpha}_i\|_0 \leqslant d, \forall i \in 1,2,\ldots,N_o. \tag{58}$$

gde je N_o ukupan broj instanci signala, d maksimalan broj nenultih elemenata u retkom kodu, $\mathbf{A}$ matrica svih retkih kodova odnosno $\mathbf{A} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_{N_o}]$, dok su značenja svih ostalih parametara ista kao i do sada. Jedan od prvih metoda namenjenih rešavanju problema koji je definisan jednačinom (58) jeste metod optimalnih pravaca (MOD method of optimal directions) (Engan et al., 1999). Formalni opis MOD algoritma je u formi pseudo koda dat na slici 13.

Osnovna ideja na kojoj se zasniva MOD algoritam jeste da se združena minimizacija po $\mathbf{D}$ i $\mathbf{A}$ razdvoji na pojedinačne minimizacije samo po $\mathbf{D}$ odnosno samo po $\mathbf{A}$ pri čemu se u toku minimizacije po jednom od ova dva parametra

20 Pogrešan atom je onaj atom koji nije iskorišćeni za generisanje signala, a algoritam ga je izabrao za reprezentaciju signala.

Parametri: Dat je skup signala $\{\mathbf{x}_i\}_{i=1}^{N_o}$, broj nenultih koeficijenata koji će biti iskorišćen za retku reprezentaciju d i maksimalno dozvoljeno odstupanje ε.

1: Inicijalizovati promenljive $k = 0$ i $\mathbf{D}^{(0)}$.
2: **Ponavljaj sledeće:**
3: $k = k + 1$.
4: Za svaki signal $\mathbf{x}_i$ odrediti koeficijente $\boldsymbol{\alpha}_i^{(k)}$ koji zadovoljavaju sledeće:

$$\boldsymbol{\alpha}_i^{(k)} = \arg\min_{\boldsymbol{\alpha}} \left\| \mathbf{x}_i - \mathbf{D}^{(k-1)}\boldsymbol{\alpha} \right\|_2^2 \text{ tako da } \|\boldsymbol{\alpha}_i^{(k)}\|_0 \leqslant d.$$

5: $\mathbf{D}^{(k)} = \mathbf{X}\mathbf{A}^{(k)} \left(\mathbf{A}^{(k)}\mathbf{A}^{(k)T}\right)^{-1}$.
6: $\mathbf{d}_i^{(k)} = \mathbf{d}_i^{(k)} \sqrt{\mathbf{d}_i^{(k)T}\mathbf{d}_i^{(k)}}$ za svaku kolonu matrice $\mathbf{D}^{(k)}$.
7: **dok važi** $\sum_{i=1}^{N_o} \|\mathbf{x}_i - \mathbf{D}^{(k)}\boldsymbol{\alpha}_i^{(k)}\|_2^2 < \varepsilon$.
8: **Rezultat:** $\mathbf{D}^{(k)}$

Slika 13: Pseudo kod MOD algoritma. Redni broj iteracije je označen sa k i nalazi se u zagradama u eksponentima promenljivih. Sa N_o je označen ukupan broj signala. U cilju skraćenja zapisa uzeto je da je $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_o}]$ i $\mathbf{A} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_{N_o}]$.

podrazumeva da je drugi parametar nepromenljiv. Ovo je urađeno da bi se problem minimizacije učinio traktabilinim. Algoritam započinje sa inicijalizacijom rečnika na slučaj, tako što se nekoliko signala iz skupa koji je na raspolaganju bira za atome ili tako što se vrednost svakog elementa atoma postavlja nezavisno. Nakon inicijalizacije rečnika potrebno je normalizovati atome, tako da l_2-norma svakog od njih bude jednaka 1. Normalizacija se vrši da bi se izbegla situacija u kojoj zbog rasta l_2-norme atoma, vrednost njemu pridruženog koeficijenta teži ka nuli, i koja kao posledica konačne preciznosti predstave broja na računaru može biti izjednačena sa nulom. Za tako formiran rečnik $\mathbf{D}$ za svaki od signala koji je na raspolaganju $\mathbf{x}_i$ izračunava se retka reprezentacija (korak 4 na slici 13). Sam algoritam ne specificira metodu kojom se izračunava retki kod za svaki od signala, a pošto je u pitanju l_0-pseudo norma, može se primenti bilo koja od pohlepnih metoda (u originalnom radu (Engan et al., 1999) je iskorišćen OMP). Nakon toga se izračunavaju vrednosti za atome (korak 5 na slici 13) na osnovu sledećeg obrasca:

$$\mathbf{D} = \mathbf{X}\mathbf{A}^T \left(\mathbf{A}\mathbf{A}^T\right)^{-1} \tag{59}$$

koji sledi direktno na osnovu (58).[21] Treba primetiti da član $\mathbf{A}^T \left(\mathbf{A}\mathbf{A}^T\right)^{-1}$ predstavlja pseudo inverziju. Ovako dobijeni rečnik se potom skalira, tako da svaki od atoma ima jediničnu l_2-normu. Procedura u kojoj se naizmenično računa

21 Pošto je $\mathbf{A}$ poznato, rečnik se dobija rešavanjem sledećeg problema:

$$\begin{aligned} \mathbf{D} &= \min_{\mathbf{D}_0} \tfrac{1}{N_o} \textstyle\sum_{i=1}^{N_o} \tfrac{1}{2}\|\mathbf{x}_i - \mathbf{D}_0\boldsymbol{\alpha}_i\|_2^2 \\ &= \min_{\mathbf{D}_0} \tfrac{1}{2N_o} \operatorname{trag}\left((\mathbf{X} - \mathbf{D}_0\mathbf{A})^T(\mathbf{X} - \mathbf{D}_0\mathbf{A})\right). \end{aligned}$$

Diferenciranjem ciljen funkcije po $\mathbf{D}_0$ i izjednačavanjem dobijenog prvog izvoda sa nula vektorom dobija se: $(\mathbf{X} - \mathbf{D}\mathbf{A})\mathbf{A}^T = 0$, odakle direktno sledi izraz za $\mathbf{D}$.

retka reprezentacija i atomi rečnika se nastavlja sve dok srednja kvadratna greška ne padne ispod nekog unapred definisanog praga.

Drugi način za formiranje rečnika uz uslov l_0 regularizacije jeste pomoću K-SVD algoritma (Aharon et al., 2006). Ono što je novo u ovom algoritmu jeste da se atomi preračunavaju pojedinačno. Ciljna funkcija data sa $\frac{1}{N_o}\sum_{i=1}^{N_o}\frac{1}{2}\|\boldsymbol{x}_i - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2$, koja se može predstaviti kao $\frac{1}{2N_o}\|\mathbf{X}-\mathbf{D}\boldsymbol{A}\|_F^2$, gde je $\|\cdot\|_F$ Frobeniusova norma,[22] može se preurediti tako da se izdvoji uticaj pojedinačnog atoma na sledeći način:

$$\|\mathbf{X}-\mathbf{D}\boldsymbol{A}\|_F^2 = \|\mathbf{X}-\sum_{\substack{j=1\\ j\neq m}}^{K}\mathbf{d}_j\boldsymbol{A}_{j*}-\mathbf{d}_m\boldsymbol{A}_{m*}\|_F^2 \tag{60}$$

gde je sa $\boldsymbol{A}_{j*}$ označena j-ta vrsta matrice $\boldsymbol{A}$. Matrica koja sadrži odstupanja stvarnog signala od rekonstruisanog ukoliko se izostavi atom $\mathbf{d}_m$ je data sa:

$$\mathbf{E}^{(m)} = \mathbf{X}-\sum_{\substack{j=1\\ j\neq m}}^{K}\mathbf{d}_j\boldsymbol{A}_{j*} \tag{61}$$

Optimalno $\mathbf{d}_m$ i $\boldsymbol{A}_{m*}$ koje minimizuje jednačinu (60) je aproksimacija ranga 1 matrice $\mathbf{E}^{(m)}$, a može se dobiti na osnovu dekompozicije matrice $\mathbf{E}^{(m)}$ na singularne vrednosti, kao par vektora koji se množi sa najvećom singularnom vrednošću.[23] Ovakvo rešenje obično ima veliki broj nenultih elemenata za vektor $\boldsymbol{A}_{m*}$, tako da uglavnom dovodi do povećanja kardinalnosti retkih kodova $\boldsymbol{\alpha}_i$. Da bi se ovo izbeglo, kolone koje odgovaraju odstupanjima signala na čiju rekonstrukciju ne utiče posmatratni atom $\mathbf{d}_m$ (indeksi kolona odgovaraju indeksima elementima $\boldsymbol{A}_{m*}$ koji su jednaki nuli) se izbacuju iz matrica $\mathbf{E}^{(m)}$ i $\mathbf{X}$. Naravno i elementi vektora $\boldsymbol{A}_{m*}$ koji su jednaki nuli se izbacuju. Ove redukovane verzije matrice $\mathbf{E}^{(m)}$ i vektora $\boldsymbol{A}_{m*}$ će nadalje u tekstu biti označene sa $\mathbf{E}_S^{(m)}$ i $\boldsymbol{A}_{mS}$ respektivno, gde je S skup koji sadrži indekse vektora $\boldsymbol{A}_{m*}$ koji su različiti od nule. Jasno je da ova redukcija za posledicu ima da se menjaju isključivo vrednosti koeficijenata koji su već različite od nule. Zbog toga prvi korak u K-SVD algoritmu jeste da se za dati rečnik odrede koeficijenti rešavanjem problema (P_0^ε), što se obično realizuje pomoću OMP algoritma. Primenom SVD-a na matricu $\mathbf{E}_S^{(m)}$ estimiraju se vrednosti i za atom $\mathbf{d}_m$ i njemu pridružene koeficijente $\boldsymbol{A}_{mS}$ na sledeći način:

$$\mathbf{d}_m = \mathbf{u}_1 \tag{62}$$

$$\boldsymbol{A}_{mS} = \Sigma_{11}\boldsymbol{v}_1^T \tag{63}$$

gde je Σ_{11} najveća singularna vrednost matrice $\mathbf{E}_S^{(m)}$, a $\mathbf{u}_1$ i $\boldsymbol{v}_1$ prve kolone matrica $\mathbf{U}$ i $\mathbf{V}$ respektivno, a koje se dobijaju na osnovu SVD-a, pri čemu važi $\mathbf{E}_S^{(m)} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$. Treba napomenuti da ponovna estimacija vrednosti koeficijenata u $\boldsymbol{A}_{mS}$ značajno ubrzava konvergenciju K-SVD algoritma. K-SVD algoritam u formi pseudo koda je naveden na slici 14.

22 Frobeniusova norma je za $m \times n$ dimenzionalnu matricu $\mathbf{X}$ definisana sa $\|\mathbf{X}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}X_{i,j}^2}$.

23 U nastavku teksta, procedura dekompozicije matrice na singularne vrednosti biće označavana sa SVD (*Singular value decomposition*)

Parametri: Dat je skup signala $\{\mathbf{x}_i\}_{i=1}^{N_o}$, broj nenultih koeficijenata koji će biti iskorišćen za retku reprezentaciju d i maksimalno dozvoljeno odstupanje ε.

1: Inicijalizovati promenljive $k = 0$ i $\mathbf{D}$.
2: **Ponavljaj sledeće:**
3: $k = k + 1$.
4: Za svaki signal $\mathbf{x}_i$ odrediti koeficijente $\boldsymbol{\alpha}_i^{(k)}$ koji zadovoljavaju sledeće:

$$\boldsymbol{\alpha}_i^{(k)} = \arg\min_{\boldsymbol{\alpha}} \left\| \mathbf{x}_i - \mathbf{D}^{(k-1)}\boldsymbol{\alpha} \right\|_2^2 \text{ tako da } \|\boldsymbol{\alpha}_i^{(k)}\|_0 \leqslant d$$

5: **Za** $m = 1$ **do** p **radi sledeće:**
6:

$$\begin{aligned} \mathbf{E}^{(m)} &= \mathbf{X} - \sum_{j=m+1}^{p} \mathbf{d}_j^{(k-1)} \mathbf{A}_{j,*}^{(k)} - \sum_{j=1}^{m-1} \mathbf{d}_j^{(k)} \mathbf{A}_{j,*}^{(k)} \\ S_m &= \{j | A_{m,j}^{(k)} \neq 0\} \end{aligned}$$

7: Izračunati SVD matrice $\mathbf{E}_{S_m}^{(m)}$ ($\mathbf{E}_{S_m}^{(m)} = \mathbf{U}^{(m)}\boldsymbol{\Sigma}^{(m)}\mathbf{V}^{(m)T}$).

$$\begin{aligned} \mathbf{d}_m^{(k)} &= \mathbf{u}_1^{(m)} \\ \mathbf{A}_{m,S}^{(k)} &= \Sigma_{1,1}^{(m)} \boldsymbol{v}_1^{(m)T} \end{aligned}$$

8: **Kraj Za**
9: **dok važi** $\sum_{i=1}^{n} \|\mathbf{x}_i - \mathbf{D}^{(k)}\boldsymbol{\alpha}_i^{(k)}\|_2^2 < \varepsilon$.
10: **Rezultat:** $\mathbf{D}$

Slika 14: Pseudo kod K-SVD algoritma. U cilju skraćenja zapisa uzeto je da je $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_o}]$, $\mathbf{A} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_{N_o}]$, $\mathbf{A}_{j*}$ j-ta vrsta matrice $\mathbf{A}$ i $\mathbf{A}_{jS}$ j-ta vrsta matrice koja sadrži samo elemente na pozicijama koje su sadržane u skupu izabranih indeksa S. Indeks iteracije je naveden u eksponentu promenljive u okviru zagrada. Primetiti da se vrednost $\mathbf{A}$ izračunava 2 puta u okviru jedne iteracije

Umesto primenom SVD algoritma, do vrednosti za $\mathbf{d}_m$ i $\mathbf{A}_{mS}$ je moguće doći i primenom jednostavne numeričke procedure, koja podrazumeva nekoliko iteracija u kojima se jedna od vrednosti fiksira, a druga izračunava primenom sledećih formula:

$$\mathbf{d}_m = \frac{\mathbf{E}_S^{(m)} \mathbf{A}_{mS}^T}{\mathbf{A}_{mS}\mathbf{A}_{mS}^T} \tag{64}$$

$$\mathbf{A}_{mS} = \frac{\mathbf{E}_S^{(m)} \mathbf{A}_{mS}^T}{\mathbf{d}_m^T \mathbf{d}_m} \tag{65}$$

Interesantno je da u slučaju da se izabere da je kardinalnost retke reprezentacije 1 (vektor $\boldsymbol{\alpha}_i$ ima isključivo jedan nenulti element) i vrednost koeficijenta ograniči na 0 ili 1, K-SVD algoritam se svodi na K-means algoritam. Za razliku od K-means algoritma gde se u svakoj iteraciji izračunavaju srednje vrednosti (*means*) za svaki od potskupova, kod K-SVD algoritma se u svakoj iteraciji pri-

menjuje SVD na svaku od K-različitih podmatrica. Ovo svojstvo je poslužilo kao inspiracija za ime algoritma. Pored toga ova sličnost sa K-means algoritmom omogućila je da se neke tehnike koje se koriste kod K-means algoritma direktno primene kod K-SVD algoritma, kao što je procedura postepenog povećanja broja atoma koja se pokazala izuzetno korisnom pri dekompoziciji signala.

Opisani K-SVD algoritam se dodatno može unaprediti ukoliko se uvede korekcioni korak koji podrazumeva zamenu atoma koji se veoma retko koristi ili koji se malo razlikuje od nekog drugog atoma sa signalom koji je najlošije reprezentovan. Treba napomenuti da je ovaj korekcioni korak moguće primeniti i u slučaju MOD algoritma.

Performanse K-SVD algoritma su približno slične kao i MOD algoritma. Interesantno je da se u eksperimentima rečnici koji se dobiju primenom ova dva algoritma na istom skupu preklapaju oko 15%, iako su im prosečna odstupanja približno ista (Elad, 2010). Velika mana ova dva algoritma je nemogućnost garantovanja dostizanja čak i lokalnog minimuma, pošto algoritam kao rezultat može da vrati vrednosti parametara koje odgovaraju sedalnim tačkama.

3.3.2 *Formiranje rečnika uz* l_1 *regularizaciju*

Problem koji treba rešiti pri formiranju rečnika uz l_1 regularizaciju je sledeći:

$$\min_{\mathbf{D}\in\mathbb{R}^{D\times K},\mathbf{A}\in\mathbb{R}^{K\times N_o}} \frac{1}{N_o}\sum_{i=1}^{N_o}\frac{1}{2}\|\boldsymbol{x}_i-\mathbf{D}\boldsymbol{\alpha}_i\|_2^2 \text{ tako da } \|\boldsymbol{\alpha}_i\|_1 \leqslant \delta, \forall i \in 1,2,\ldots,N_o. \tag{66}$$

gde je δ maksimalna l_1 norma. Za rešavanje problema definisanog jednačinom (66) obično se korisit Lagranžijan, što se svodi na oblik:

$$\min_{\mathbf{D}\in\mathbb{R}^{D\times K},\mathbf{A}\in\mathbb{R}^{K\times N_o}} \frac{1}{N_o}\sum_{i=1}^{N_o}\frac{1}{2}\|\boldsymbol{x}_i-\mathbf{D}\boldsymbol{\alpha}_i\|_2^2+\lambda\|\boldsymbol{\alpha}_i\|_1. \tag{67}$$

Za razliku od varijante kada se koristi l_0 regularizacija, gornji problem je konveksan i po $\mathbf{D}$ i $\mathbf{A}$, ali nije združeno konveksan po $(\mathbf{D}, \mathbf{A})$, stoga se optimizacija i u ovom slučaju vrši tako što se jedna od ove dve promenljive fiksira dok se druga estimira. Za estimaciju $\mathbf{A}$ može da se koristi bilo koja prethodno opisana procedura za retko kodovanje uz l_1 ograničenje, pri čemu se najčešće koristi LARS. Optimalna estimacija rečnika $\mathbf{D}$ se može realizovati ili pomoću gradijentne metode kao što je urađeno u (Olshausen i Field, 1996) ili pomoću Njutnove (*Newton*) metode u varijanti sa Langranžovim dualnim problemom (Lee et al., 2006). Druga metoda je nešto efikasnija, pošto se uvođenjem dualnog problema smanjio broj parametara koji treba optimizovati u samom procesu. Izrazi koji su neophodni za određivanje rečnika pomoću ove dve metode, kao i način kako se do njih dolazi su navedeni u prilogu A.2.

Prethodno opisani algoritam podrazumevaju da se pri iterativnom određivanju rečnika u svakom koraku koriste svi raspoloživi signali. Ove procedure postaju prilično računski komplikovane ukoliko je broj signala koji su na raspolaganju veliki. Jedan od načina za prevazilaženje ovog problema jeste pomoću

tzv. određivanja rečnika u letu (*Online dictionary learning*) (Mairal et al., 2009). Osnovna ideja je da se adaptacija parametara vrši za svaki novi signal ili grupu signala. Može se pokazati da u slučaju beskonačno mnogo signala ovakvo rešenje konvergira stacionarnoj tački (Mairal et al., 2009).

Kao i kod prethodnih algoritama za određivanje rečnika i kod ovog algoritma određivanje koeficijenata retke reprezentacije i rečnika se vrši odvojeno. Algoritam započinje inicijalizacijom rečnika na slučaj, što može da bude u varijanti da svaki elemenat rečnika bude inicijalizovan na slučaj ili da se na slučaj izabere nekoliko signala iz skupa za obuku. Iz istih razloga kao i u prethodnim algoritmima vrši se normalizacija atoma tako da im l_2-norma bude jedinična. Pored toga pomoćne matrice $\mathbf{B}$ i $\mathbf{C}$ se inicijalizuju nula matricama. Potom se uzima prvi signal koji je na raspolaganju i za njega se određuje retka reprezentacija pomoću LARS algoritma. Nakon toga se koriguju vrednosti pomoćnih matrica $\mathbf{B}$ i $\mathbf{C}$, koje se potom koriste za korekciju vrednosti atoma, pomoću sledećih jednakosti:

$$\mathbf{u}_i = \frac{1}{C_{i,i}}(\mathbf{b}_i - \mathbf{D}\mathbf{c}_i) + \mathbf{d}_i \tag{68}$$

$$\mathbf{d}_i = \frac{1}{\max(\|\mathbf{u}_i\|_2, 1)}\mathbf{u}_i \tag{69}$$

za svako $i = 1, \ldots, k$, pri čemu su $\mathbf{b}_i$ i $\mathbf{c}_i$ i-te kolone matrica $\mathbf{B}$ i $\mathbf{C}$ respektivno. Ova procedura sa ponavlja sa svakim novim signalom. Formalan opis algoritma u formi pseudo koda je dat na slici 15.

Pomoćne matrice $\mathbf{B}$ i $\mathbf{C}$ predstavljaju proizvod do tog trenutka t iskorišćenih signala sa njihovim retkim kodovima, odnosno proizvod retkih kodova:

$$\mathbf{B} = \mathbf{X}^{(t)}\mathbf{A}^{(t)T} = \sum_{i=1}^{t} \mathbf{x}_t \boldsymbol{\alpha}_t^T \tag{70}$$

$$\mathbf{C} = \mathbf{A}^{(t)}\mathbf{A}^{(t)T} = \sum_{i=1}^{t} \boldsymbol{\alpha}_t \boldsymbol{\alpha}_t^T \tag{71}$$

gde je sa $\mathbf{X}^{(t)}$ označena matrica $\mathbf{X}^{(t)} = [\mathbf{x}_1, \ldots, \mathbf{x}_t]$ sa $\mathbf{A}^{(t)}$ matrica $\mathbf{A}^{(t)} = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_t]$. Ove matrice figurišu direktno u ciljnoj funkciji za određivanje rečnika u slučaju kada je poznata matrica retkih kodova $\mathbf{A}$:

$$\frac{1}{2}\sum_{i=1}^{t} \|\mathbf{x}_i - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2 = \frac{1}{2}\mathrm{trag}\left(\left(\mathbf{X}^{(t)} - \mathbf{D}\mathbf{A}^{(t)}\right)^T\left(\mathbf{X}^{(t)} - \mathbf{D}\mathbf{A}^{(t)}\right)\right) \tag{72}$$

$$= \frac{1}{2}\mathrm{trag}\left(\mathbf{D}^T\mathbf{D}\mathbf{A}^{(t)}\mathbf{A}^{(t)T}\right) - \mathrm{trag}\left(\mathbf{D}^T\mathbf{X}^{(t)}\mathbf{A}^{(t)T}\right) + \frac{1}{2}\mathrm{trag}\left(\mathbf{X}^{(t)T}\mathbf{X}^{(t)}\right) \tag{73}$$

$$= \frac{1}{2}\mathrm{trag}\left(\mathbf{D}^T\mathbf{D}\mathbf{C}\right) - \mathrm{trag}\left(\mathbf{D}^T\mathbf{B}\right) + \frac{1}{2}\mathrm{trag}\left(\mathbf{X}^{(t)T}\mathbf{X}^{(t)}\right) \tag{74}$$

a samim tim i u gradijentu ove funkcije koji je jednak: $\mathbf{D}\mathbf{C} - \mathbf{B}$, a koji se koristi za korekciju rečnika u jednačini (68).

Algoritam koji je naveden na slici 15 se odnosi na teorijsku varijantu u kojoj postoji beskonačno mnogo signala, tako da je umesto konkretnih signala

Parametri: Data je gustina raspodele signala $p(\mathbf{x})$, maksimalan broj opservacija/iteracija T i λ regularizacioni parametar.

1: Inicijalizovati promenljive: $\mathbf{B}^{(0)} = \mathbf{0}$, $\mathbf{C}^{(0)} = \mathbf{0}$ i $\mathbf{D}^{(0)}$.
2: **Za** $t = 1$ **do** T **radi sledeće:**
3: Generisati $\mathbf{x}_t$ na osnovu gustine raspodele $p(\mathbf{x})$.
4: Za signal $\mathbf{x}_t$ odrediti redak kod koji zadovoljava sledeće:

$$\boldsymbol{\alpha}_t = \arg\min_{\boldsymbol{\alpha}} \frac{1}{2} \left\| \mathbf{x}_t - \mathbf{D}^{(t-1)}\boldsymbol{\alpha} \right\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1.$$

5: $\mathbf{B}^{(t)} = \mathbf{B}^{(t-1)} + \mathbf{x}_t \boldsymbol{\alpha}_t^T$.
6: $\mathbf{C}^{(t)} = \mathbf{C}^{(t-1)} + \boldsymbol{\alpha}_t \boldsymbol{\alpha}_t^T$.
7: **Za** $i = 1$ **do** K **radi sledeće:**

$$\begin{aligned} \mathbf{u}_i &= \frac{1}{C_{i,i}} \left(\mathbf{b}_i - \sum_{j=1}^{i-1} C_{j,i} \mathbf{d}_j^{(t)} - \sum_{j=i}^{k} C_{j,i} \mathbf{d}_j^{(t-1)} \right) + \mathbf{d}_i^{(t-1)}; \\ \mathbf{d}_i^{(t)} &= \frac{1}{\max\left(\|\mathbf{u}_i\|_2, 1\right)} \mathbf{u}_i; \end{aligned}$$

8: **Kraj Za**
9: **Kraj Za**
10: **Rezultat:** $\mathbf{D}^{(T)}$

Slika 15: Pseudo kod za online određivanje rečnika. Redni broj iteracije je označen sa t i naveden je u zagradama u eksponentu promenljive.

zadata funkcija gustine raspodele, a korekcija se vrši za svaki od signala. U praksi su na raspolaganju skupovi signala koji imaju mnogo elemenata, ali ipak konačno mnogo, stoga se signali uzimaju po nekoliko puta, pri čemu nijhov redosled varira. Pored toga da bi se ubrzala brzina konvergencije algoritma, umesto samo jednog signala pri korekciji rečnika koristi se nekoliko signala. Ovaj algoritam se može koristiti i u slučaju da se retko kodovanje vrši nekom od metoda koje direktno minimizuju l_0-pseudo normu, ali se u tom slučaju ne može dokazati konvergencija rešenja, što važi i za MOD i K-SVD algoritam.

3.4 ODREĐIVANJE REČNIKA I FAKTORIZACIJA MATRICE

Problem određivanja rečnika se može tretirati kao problem faktorizacije matrice $\mathbf{X}$ na matrice $\mathbf{A}$ i $\mathbf{D}$ uz ograničenje $l'(\mathbf{A}) = \sum_i l(\boldsymbol{\alpha}_i)$, gde $l(\boldsymbol{\alpha})$ zamenjuje odgovarajući regularizator (l_0-pseudo norma ili l_1-norma):

$$\min_{\mathbf{D},\mathbf{A}} \frac{1}{2} \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 + \lambda l'(\mathbf{A}), \tag{75}$$

stoga bi ga trebalo uporediti sa nekim drugim metodama faktorizacije matrice.

Jedna on najčešće korišćenih metoda za faktorizaciju matrice koja se koristi u oblasti analize podataka jeste rastavljanje na osnovne komponente (PCA *Principal component analysis*). Problem koji rešava PCA je:

$$\min_{\mathbf{D},\mathbf{A}} \frac{1}{2}\|\mathbf{X}-\mathbf{D}\mathbf{A}\|_F^2 \text{ tako da } \mathbf{D}^T\mathbf{D}=\mathbf{I} \text{ i } \mathbf{A}\mathbf{A}^T \text{ je dijagonalna matrica} \quad (76)$$

Jedan od standardnih pristupa za rešavanje ovog problema je pomoću SVD-a.

Vektorska kvantizacije (ili klasterovanje) se takođe mogu posmatrati kao problem faktorizacije matrice, odnosno:

$$\min_{\mathbf{D},\mathbf{A}} \frac{1}{2}\|\mathbf{X}-\mathbf{D}\mathbf{A}\|_F^2 \text{ tako da } \|\alpha_i\|_2=1 \text{ i } A_{j,i}\in\{0,1\} \text{ za } i=1,\ldots,N_o. \quad (77)$$

gde je $\mathbf{D}$ matrica koja sadrži centroide, a $\mathbf{A}$ indikatorska matrica, koja upućuje kom vektoru je dodeljen koji centrodi. Kao što je već ranije napomenuto, vektorska kvantizacija se može posmatrati kao poseban slučaj K-SVD ukoliko se kardinalnost retkog koda ograniči na 1.

Vektorska kvantizacija sa mekom dodelom (Seo i Obermayer, 2003) je varijanta u kojoj se vektor prikazuje kao linarna kombinacija centroida, tako da je suma odgovarajućih nenegativnih težina jednaka 1 odnosno:

$$\min_{\mathbf{D},\mathbf{A}} \frac{1}{2}\|\mathbf{X}-\mathbf{D}\mathbf{A}\|_F^2 \text{ tako da } \|\alpha_i\|_2=1 \text{ i } A_{j,i}\geqslant 0 \text{ za } i=1,\ldots,N_o. \quad (78)$$

Nenegativna faktorizacija matrice (Seung i Lee, 2001) je varijanta rastavljanja matrice na dve matrice čiji su svi elementi nenegativni odnosno:

$$\min_{\mathbf{D},\mathbf{A}} \frac{1}{2}\|\mathbf{X}-\mathbf{D}\mathbf{A}\|_F^2 \text{ tako da } D_{l,j}\geqslant 0 \text{ i } A_{j,i}\geqslant 0 \quad (79)$$

Primena nenegativne faktorizacije matrice u obradi slike, na primeru lica daje retka rešenja (Seung i Lee, 2001).

3.5 REZIME

U ovom poglavlju su izložene osnovne ideje, kao i algoritimi koji se koriste u oblasti retke reprezentacije. Definisana su dva osnovna problema koja je potrebno rešiti: kako pronaći redak kod i odgovarajući rečnik potreban za reprezentaciju. Za svaki od problema predstavljeno je nekoliko najznačajnijih rešenja sa njihovim prednostima i manama. Naravno ovo poglavlje nije obuhvatilo sve varijacije algoritama, već samo one najvažnije, za koje sam smatrao da su neophodne za razumevanje ideje aproksimacije punih kovarijansnih matrica pomoću retke reprezentacije njihovih karakterističnih vrednosti i koji su najčešće referencirani u stručnoj literaturi.

Svi opisani algoritmi i korišćena literatura je vezana za digitalnu obradu slike, gde se retka reprezentacija koristi za uklanjanje šuma, uklanjanje zamućena, kompresiju itd.. Postoje i trendovi da se retka reprezentacija primeni i u drugim oblastima, npr. za obradu audio signala, ali istraživanja pokazuju da bi za meru rastojanja između signala trebalo koristiti neku drugu metriku umesto Euklidove. U ovom radu će biti formalno pokazano da se postojeći algoritmi zasnovani na minimizaciji Euklidske metrike mogu uspešno primeniti za aproksimaciju punih kovarijansnih matrica.

4

MODELOVANJE INVERZNIH KOVARIJANSNIH MATRICA POMOĆU RETKE REPREZENTACIJE NJIHOVIH KARAKTERISTIČNIH VEKTORA

4.1 UVOD

U ovom poglavlju je predstavljen model aproksimacije punih kovarijansnih matrica modela Gausovih mešavina pomoću retke reprezentacije njihovih karakterističnih vektora. Ova aproksimacija ima za cilj prevazilaženje sledećih problema:

- smanjivanje broja parametara potrebnih za reprezentaciju modela,
- smanjivanje računske složenosti izračunavanja logaritma izglednosti,
- obezbeđivanje tačnosti modela.

Pored samog opisa modela, data je i analiza broja parametara, broja potrebnih računskih operacija neophodnih za izračunavanje logaritma izglednosti, kao i način estimacije parametara na osnovu kriterijuma maksimalne izglednosti.

4.2 OPIS MODELA

Pošto je cilj ovog modela, pored smanjenja broja parametara i ubrzanje izračunavanja logaritma izglednosti Gausijana, a na osnovu izloženog u poglavlju 2, jasno je da je neophodno direktno aproksimirati inverznu kovarijansnu matricu. Inverzna kovarijansna matrica $\boldsymbol{\Sigma}_{sm}^{-1}$, koja opisuje m-tu komponentu mešavine pridruženu stanju s može se predstaviti na sledeći način:

$$\boldsymbol{\Sigma}_{sm}^{-1} = \sum_{i=1}^{D} \lambda_{smi} \boldsymbol{v}_{smi} \boldsymbol{v}_{smi}^{T} \tag{80}$$

gde je D dimenzionalnost prostora, a λ_{smi} i $\boldsymbol{v}_{smi}$ i-ta karakteristična vrednost i i-ti karakteristični vektor matrice $\boldsymbol{\Sigma}_{sm}^{-1}$ respektivno. Sve karakteristične vrednosti i svi karakteristični vektori su realni pošto je matrica simetrična i pozitivno definitna, čime se prostor pretrage za retkom reprezentacijom karakterističnog vektora ograničava na $\mathbb{R}^D$, što je jedan od preduslova za korišćenje algoritama koji su opisani u poglavlju 3. Treba napomenuti da kovarijansna matrica i njena inverzija imaju iste karakteristične vektore, dok su im karakteristične vrednosti recipročne, što će biti iskorišćeno kasnije pri formalnom dokazu opravdanosti izbora ciljne funkcije.

Ovaj model polazi od pretpostavke da svaki karakteristični vektor $\boldsymbol{v}_{smi}$ ima retku predstavu odnosno:

$$\boldsymbol{v}_{smi} = \mathbf{D}\boldsymbol{\alpha}_{smi} \tag{81}$$

gde je $\mathbf{D} \in \mathbb{R}^{D \times K}$ rečnik, a $\boldsymbol{\alpha}_i \in \mathbb{R}^K$ redak kod vektora $\boldsymbol{v}_{smi}$. Pri tome predloženi model podrazumeva da je rečnik parametar koji se deli između svih

Gausovih raspodela koje čine model, dok su retki vektori specifični za svaku od Gausovih raspodela. Po uzoru na jednačinu (26), logaritam izglednosti jedne Gausove raspodele za opservaciju $\mathbf{o}$ može se napisati na sledeći način:

$$l_{sm}(\mathbf{o}) = c_{sm} + \bar{\mu}_{sm}^{T}\mathbf{o} - \frac{1}{2}\sum_{i=1}^{D}\lambda_{smi}\left(\mathbf{o}^{T}\mathbf{D}\boldsymbol{\alpha}_{smi}\right)^{2} \tag{82}$$

gde je:

$$c_{sm} = -\frac{D}{2}\ln(2\pi) + \frac{1}{2}\ln\left(|\boldsymbol{\Sigma}_{sm}^{-1}|\right) - \frac{1}{2}\boldsymbol{\mu}_{sm}^{T}\boldsymbol{\Sigma}_{sm}^{-1}\boldsymbol{\mu}_{sm} \tag{83}$$

$$\bar{\mu}_{sm} = \boldsymbol{\Sigma}_{sm}^{-1}\boldsymbol{\mu}_{sm} \tag{84}$$

pri čemu c_{sm} i $\bar{\mu}_{sm}$ ne zavise od $\mathbf{o}$ te se mogu unapred izračunati i čuvati kao parametri Gausove raspodele. Naravno u slučaju GMM svaka od Gausovih raspodela je pomnožena odgovarajućom težinom, čiji se logaritam može dodati novom parametru c_{sm}. Ukoliko sa d označimo kardinalnost retke reprezentacije $\boldsymbol{\alpha}_{smi}$, tada je ukupan broj parametara sa kojim je opisana jedna Gausova raspodela jednak $D(d+2)+1$.[1]

4.2.1 *Broj parametara*

Ukoliko se predloženi model uporedi sa najčešće korišćenom varjantom GMM-a, GMM-om sa dijagonalnim kovarijansnim matricama, gde je svaka Gausova raspodela opisana samo sa po $2D+1$ parametrom, uočava se da je za isti ukupan broj Gausovih raspodela M broj parametara predloženog modela značajno veći. Pošto predloženi model implicitno podrazumeva modelovanje korelacija između obeležja, a kod GMM-a sa dijagonalnim kovarijansnim matricama to ne postoji već se to postiže povećanjem broja Gausovih raspodela, razlika u broju parametara između ova dva modela u slučaju sistema za automatsko prepoznavanje govora nije u toj meri izražena.

Kao što je već napomenuto u poglavlju 2 varijanta GMM-a sa punim kovarijansnim matricama u slučaju dovoljne količine podataka za obuku daje najbolje moguće rezultate u odnosu na sve druge varijante GMM-a (Axelrod et al., 2005), te se često koristi kao referentni model. Cilj je da alternativni model ima manju složenost od njega, uz neznatan gubitak tačnosti, s tim da u slučajevima kada nema dovoljno podataka za obuku alternativni model može dati bolju tačnost zbog robustnije estimacije parametara. Direktno poređenje broja prametara ova dva modela nije tako jednostavno pošto predloženi model ima i $K \cdot D$ deljenih parametara, pored onih kojima su opisani pojedinačni Gausiani. Pošto oba modela modeluju korelacije koje postoje između obeležja očekivani broj Gausovih raspodela je približno sličan tako da je složenost predloženog modela prvenstveno određena brojem atoma K, kao i kardinalnošću retke reprezentacije d. U slučaju velikog broja mešavina (nekoliko desetina hiljada) i relativno velike dimenzionalnosti prostora obeležja može se smatrati da je predloženi model manje složenosti ukoliko važi $d < (D-1)/2$.

1 Do ovog broja se dolazi relativno jednostavno: postoji D karakterističnih vrednosti λ_{smi}, D retkih kodova $\boldsymbol{\alpha}_{smi}$ svaki sa po d nenultih elemenata, D elemenata vektora $\bar{\mu}_{sm}$ i 1 koeficijent c_{sm}. Potrebno je zapamtiti i indekse koji odgovaraju nenultim elementima retkog koda, što je moguće realizovati u formi binarnog vektora dužine K, što neće značajno uticati na ukupan broj parametara, te se može izostaviti iz razmatranja.

U predloženom modelu inverzna kovarijansna matrica se može predstaviti kao:

$$\boldsymbol{\Sigma}_{sm}^{-1} = \sum_{i=1}^{D} \lambda_{smi} \mathbf{D} \boldsymbol{\alpha}_{smi} \boldsymbol{\alpha}_{smi}^{T} \mathbf{D}^{T} \tag{85}$$

$$= \sum_{i=1}^{D} \lambda_{smi} \sum_{j=1}^{d} \alpha_{smi,f_{smi}(j)} \mathbf{d}_{f_{smi}(j)} \sum_{k=1}^{d} \alpha_{i,f_{smi}(k)} \mathbf{d}_{f_{smi}(k)}^{T} \tag{86}$$

$$= \sum_{i=1}^{D} \sum_{j=1}^{d} \sum_{k=1}^{d} \lambda_{smi} \alpha_{i,f_{smi}(j)} \alpha_{i,f_{smi}(k)} \mathbf{d}_{f_{smi}(j)} \mathbf{d}_{f_{smi}(k)}^{T} \tag{87}$$

gde je $f_{smi}(j)$ funkcija koja vraća indekse nenultih elemenata retkog vektora $\boldsymbol{\alpha}_{smi}$ i $\mathbf{d}_i$ i-ta kolona matrice $\mathbf{D}$, što odgovara reprezentaciji kovarijansne matrice u EMLLT modelu. U zavisnosti od toga da li se za predstavu karakterističnih vektora jedne matrice koriste isti atomi ili ne, broj koeficijenata kojim je predstavljena matrica varira od Dd (ako su svi karakteristični vektori predstavljeni preko istih atoma) do $\min\{Dd^2, K\}$. Treba primetiti da faktorisana varijanta predstave inverzne kovarijansne matrice zahteva uvek isti broj parametara $D(d+1)$ nezavisno od toga koji se atomi koriste za reprezentaciju karakterističnih vektora i taj broj je blizak minimalnom potrebnom broju za razvijenu formu koja je data jednačinom (87).

Kao što je navedeno u poglavlju 2, specijalni slučaj EMLLT jeste MLT, gde se za reprezentaciju svake matrice koristi isključivo podskup od D vektora iz unapred izabrane baze. Predloženi model se svodi na MLT ukoliko se kardinalnost retkih vektora ograniči na 1, te se umesto retke reprezentacije karakterističnih vektora vrši nijhova kvantizacija. Treba napomenuti da bi u tom slučaju broj parametara kojim je opisana jedna Gausova raspodela GMM-a bio $2D+1$, kao u slučaju dijagonalnog modela.[2]

Uopštenje EMLLT modela je PCGMM, gde je inverzna kovarijansna matrica predstavljena kao linearna kombinacija matrica proizvoljnog ranga umesto ranga 1 što je bio slučaj kod EMLLT modela. Ovaj model se pokazao boljim od EMLLT te je i ovde naveden radi poređenja broja potrebnih parametara. U eksperimentima koji su prezentovani u radu (Axelrod et al., 2005), broj parametara n koji se koristio za reprezentaciju inverzne kovarijansne matrice u varijantama koje su davale tačnost prepoznavanja blisku tačnosti koja je postignuta pomoću GMM-a sa punim kovarijansnim matricama, je bila od 4D do 8D. Takva složenost se u predloženom modelu postiže ukoliko se za kardinalnost retkog vektora d uzmu vrednosti iz intervala $[2, 6]$.

Radi bolje preglednosti u tabeli 1 su navedene vrednosti broja parametara za različite načine aproksimacije kovarijansnim matrica u modelima Gausovih mešavina. Značenja oznaka promenjljivih su kao u dosadašnjem tekstu odnosno: D – dimenzionalnost prostora obeležja, M – ukupan broj Gausovih raspodela u modelu, n – broj vektora/matrica koji se koriste za reprezentaciju inverzne kovarijansne matrice u slučaju EMLLT/PCGMM, K – broj atoma i d – kardinalnost retkog vektora. Pošto su u radu za skraćene nazive modela korišćeni engleski akronimi, to je urađeno i za model koje je predložen u ovom radu čiji je akronim SEGMM od *sparse eigenvector* GMM.

2 Ukupan broj parametara je naravno nešto veći, pošto je potrebno zapamtiti deljene parametre, ali i indekse atoma iz rečnika koje moguće sačuvati u formi binarnih vektora.

Tabela 1: Broj parametara za različite načine aproksimacije kovarijansnih matrica u modelima Gausovih mešavina.

Model	Broj parametara
Diag.	$M(2D+1)$
Pune	$M\left(D(D+3)/2+1\right)$
EMLLT	$nD + M(D+n+1)$
PCGMM	$nD(D+1) + M(D+n+1)$
SEGMM	$KD + M\left(D(d+2)+1\right)$

4.2.2 *Broj računskih operacija pri izračunavanju logaritma izglednosti*

Kao što su parametri koji opisuju model podeljeni u dve grupe: deljeni i specifični za Gausovu raspodelu, tako je i izračunavanje logaritma izglednosti podeljeno u dve faze. Prvo se vrši izračunavanje sa deljenim parametrima, pošto njih koriste sve Gausove raspodele koje čine model. Ono podrazumeva množenje tekuće opservacije $\mathbf{o}$ sa rečnikom $\mathbf{D}$ što zahteva $K(2D-1)$ osnovnih operacija sa pokretnim zarezom. Potom se za svaku od mogućih[3] Gausovih raspodela vrši izračunavanje logaritma izglednosti na osnovu jednačine (82). Najveći broj računskih operacija odlazi na izračunavanje člana:

$$\sum_{i=1}^{D} \lambda_{smi} \left(\sum_{j=1}^{d} (\mathbf{o}^T \mathbf{d}_{f_{smi}(j)}) \alpha_{smi, f_{smi}(j)} \right)^2 \tag{88}$$

što iznosi $2D(d+1)-1$ flops, [4] dok je ukupan broj operacija sa pokretnim zarezom za izračunavanje (82) $2D(d+2)+1$ flops.

U tabeli 2 je naveden broj potrebnih operacija sa pokretnim zarezom za predloženi model kao i za referentne pristupe modelovanju kovarijansnih matrica radi jednostavnijeg poređenja. Detaljana analiza za svaki od navedenih modela je napravljena u okviru poglavlja 2, stoga ovde neće biti ponovljena. Pošto uspešnije aproksimacije punih kovarijansnih matrica podrazumevaju uvođenje deljenih parametara, operacije koje su potrebne za izračunavanje logaritma izglednosti su podeljene na operacije sa deljenim parametrima (njihov broj je naveden u koloni "Deljeni") i operacije sa parametrima koji su specifični za svaku od raspodela (njihovi broj je naveden u koloni "Specifični"). Razlog zašto su nazivi kolona u tabeli 2 izabrani na osnovu kategorije parametara koji u njima učestvuju, a ne po fazama izračunavanja logaritma izglednosti jeste činjenica da se logaritam izglednosti u modelima sa dijagonalnim i punim kovarijansnim matricama izračunava odjednom, tako da bi ove oznake uvele određen nivo konfuzije.

Kao što se iz priloženog može videti odnos broja operacija koji zahtevaju pojedini modeli je sličan odnosu broja parametara koji se koriste za opisiva-

3 Da bi se ubrzao proces prepoznavanja vrši se odsecanje (pruning) manje verovatnih putanja u trelisu, odnosno odbacivanje pojedinih HMM stanja, te se u praktičnim aplikacijama vrlo retko javlja potreba za izračunavanjem svih Gausovih raspodela koje čine model.

4 Član $\mathbf{o}^T \mathbf{d}_i$ je skalar koji je izračunat u fazi izračunavanja sa deljenim parametrima.

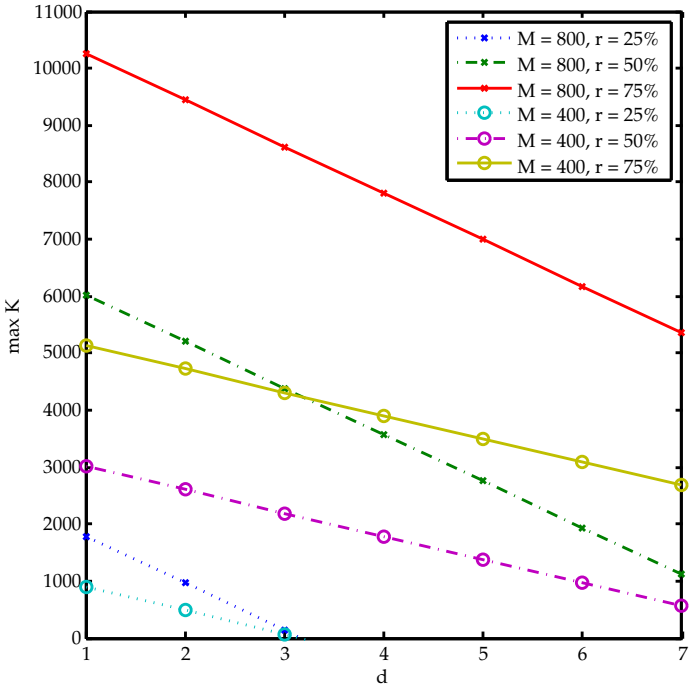
Tabela 2: Broj računskih operacija sa pokretnim zarezom za različite načine aproksimacije kovarijansnih matrica u modelima Gausovih mešavina. Broj opseracija je prikazan po fazama: prva faza koja obuhvata izračunavanja sa deljenim parametrima (kolona Deljeni) i druga faza koja podrazumeva izračunavanje logaritma izglednosti za svaku od pojedinačnih Gausovih raspodela (kolona Specifični).

Model	Deljeni	Secifični
Diag.	0	$4D+1$
Pune	0	$D(3D+5)/2+2$
EMLLT	$n(2D-1)$	$2(D+n)$
PCGMM	$3nD(D+1)/2$	$2(D+n)$
SEGMM	$K(2D-1)$	$2D(d+2)+1$

nje pojedinih modela (videti tabelu 1). Odnosno, najmanji broj izračunavanja po jednoj Gausovoj raspodeli zahteva model sa dijagonalnom aproksimacijom punih kovarijansnih matrica. Ovaj broj operacija po jednoj Gausovoj raspodeli se u slučaju EMLLT i PCGMM modela postiže ukoliko se uzme da je $n = D$,[5] ali naravno ne smeju se zaboraviti potrebna izračunavanja sa deljenim parametrima. Sa druge strane, predloženi model (SEGMM) ni za jednu vrednost parametra d ne može da postigne ovaj broj računskih operacija po jednoj Gausovoj raspodeli. Kao što je već napomenuto modeli EMLLT i PCGMM postižu performanse bliske performansama modela sa punim kovarijansnim matricama (što je i cilj kojem se i teži u alternativnom modelovanju GMM-a) ukoliko je n reda $4D$ ili $8D$ što se dobija za d jedanko 3 odnosno 7. Što se tiče odnosa broja računskih operacija potrebnih za izračunavanje deljenih parametara i on zavisi od vrednosti za K i n. Ukoliko se predloženi model uporedi sa EMLLT može se videti da se isti broj izračunavanja postiže ukoliko je $K = n$. Obično se za K uzimaju vrednosti koje su za nekoliko redova veličine veće od dimenzionalnosti prostora koji je potrebno opisati stoga predloženi model zahteva značajno veći broj operacija sa deljenim parametrima od onog koji zahteva EMLLT u slučaju tačnosti prepoznavanja koju postižu modeli sa punim kovarijansnim matricama. Ukoliko predloženi model poredimo sa PCGMM tada se isti broj izračunavanja sa deljenim parametrima postiže ukoliko je $K \approx 1.5nD$, što je daleko veće od onoga što je bilo potrebno za eksperimente prikazane u ovomo radu.

U cilju ubrzanja procesa dekodovanja obično se vrši odsecanje malo verovatnih hipoteza, tako da broj Gausovih raspodela za koje je potrebno izračunati logaritam izglednosti varira tokom vremena, a samim tim i broj potrebnih računskih operacija sa pokretnim zarezom. Ove varijacije zavise od velikog broja faktora kao što su: odstupanje akustičkih karakteristika test sekvence od akustičkih uslova u sekvencama koje su korišćene za obuku modela, fonetski sadržaj test sekvence, skup reči koje se koriste u test skupu i sl., što dodatno otežava analizu računske složenosti modela u opštem slučaju. Ukoliko je cilj alternativnog modela povećanje brzine izračunavanja, odsecanje u

5 Podsećanja radi EMLLT se za $n = D$ se svodi na tzv. MLLT.



Slika 16: Zavisnost maksimalnog broja atoma u rečniku K od kardinalnosti retkog koda d, u slučaju 26-dimenzionalnog prostora za različite vrednosti prosečnog broja Gausovih raspodela za koje se izračunava logaritam izglednosti po frejmu M, i različite nivoe redukcije r u odnosu na model sa punim kovarijansnim matricama. Nivo redukcije od r_0% znači da je ukupan broj operacija po jednom frejmu ograničen r_0% od broja operacija koje zahteva model sa punim kovarijansim matricama.

značajnoj meri utiče na broj i organizaciju deljenih parametara. Za ilustraciju uticaja odsecanja na maksimalan broj deljenih parametara, odnosno atoma, u predloženom modelu može da posluži slika 16. Na slici 16 je ilustrovana zavisnost maksimalnog broja atoma K od kardinalnosti retkog koda d, za različite vrednosti prosečnog broja Gausovih raspodela po opservaciji za koje se izračunava logaritam izglednosti M i stepene redukcije broja operacija r. Stepen redukcije se odnosi na procenat operacija koje se izvršavaju u predloženom modelu u odnosu na broj operacija koje zahteva model sa punim kovarijansnim matricama i istim brojem Gausovih raspodela koje je potrebno izračunati. Kao što se iz priloženog može videti sa porastom kardinalnosti retkog koda uz nepromenjen broj operacija sa pokretnim zarezom, broj atoma koji čine rečnik se smanjuje. Ova promena je izraženija ukoliko je broj Gausovih raspodela za koje je potrebno izračunati logaritam izglednosti veći. Ukoliko se zahteva značajna redukcija broja operacija, npr. 4 puta ($r = 25\%$), tada za nešto veće vrednosti kardinalnosti retke reprezentacije ona nije moguća. Naravno u ovim slučajevima je upitna i tačnost reprezentacije ovako redukovanog modela, odnosno pri estimaciji parametara modela potrebno je voditi računa i o tačnosti modela, a ne samo o računskoj i memorijskoj složenosti modela. Stoga drugi problem koji je potrebno rešiti jeste kako estimirati parametere modela.

4.3 ESTIMACIJA PARAMETARA MODELA

Pri estimaciji parametara cilj je pronaći retku reprezentaciju karakterističnih vektora kovarijansnih matrica, tako da razlika između verodostojnosti Gausovih raspodela sa estimiranim i aproksimiranim kovarijansnim matricama bude minimalna. Jasno je da će ta razlika biti mala ukoliko je razlika između te dve raspodele mala. Prirodna mera rastojanja između dve raspodele jeste Kulbak-Lajbler divergenca (KLD *Kullback-Leibler divergence*) (Kotz i Johnson, 1993), te se ona nameće kao logičan izbor za ciljnu funkciju. Može se pokazati da se minimizacija KLD-a između dve Gausove raspodele koje imaju istu srednju vrednost i iste karakteristične vrednosti kovarijansnih matrica svodi na minimizaciju euklidskog rastojanja između karakterističnih vektora koji odgovaraju istim karakterističnim vrednostima, što je formalno pokzano u odeljku 4.4. Ova tvrdnja važi i za inverzne kovarijansne matrice, pošto su karakteristični vektori kovarijansne matrice i njene inverzne vrednosti isti. Iako se predloženi model zasniva na aproksimaciji inverznih kovarijansnih matrica, formalni dokaz je naveden za kovarijansnu matricu pošto je u toj varijanti nešto jednostavniji.

Estimacija parametara kojima su opisane inverzne kovarijansne matrice u predloženom modelu se svodi na problem pronaženja odgovarajućeg rečnika kao i odgovarajućih retkih kodova za retku reprezentaciju karakterističnih vektora kovarijansnih matrica. Ovaj problem se može formalno zapisati na sledeći način:

$$\min_{\mathbf{D}\in\mathcal{C},\{\boldsymbol{\alpha}_{mi}\}} \sum_{s=1}^{S}\sum_{m=1}^{M_s}\sum_{i=1}^{D}\frac{1}{2}\|\boldsymbol{\nu}_{smi}-\mathbf{D}\boldsymbol{\alpha}_{smi}\|_2^2 \text{ tako da: } \|\boldsymbol{\alpha}_{smi}\|_0 \leqslant d \tag{89}$$

gde je $\mathcal{C}$ konveksni skup matrica u $\mathbb{R}^{D\times K}$ takvih da je l_2-norma njihovih kolona manja ili jedanaka 1, S ukupan broj stanja, M_s ukupan broj komponenti mešavine kojom je opisano stanje s, dok su sve ostale oznake iste kao i ranije. Kao što je napomenuto u poglavlju 3, eksperimenti na veštački generisanim podacima koji su navedeni u (Elad, 2010) su pokazali da se u slučaju male kardinalnosti retkih kodova (ako je manja od 6) nešto bolji rezultati u smislu odstupanja aproksimirane vrednosti od stvarne se dobijaju ukoliko se kao regularizator koristi l_0-pseudo norma. Ovo svojstvo je bilo osnovni motiv za izbor l_0-pseudo norme kao regularizatora u ovom radu. Nažalost, u slučaju kada se koristi ovaj regularizator, koji nije konveksan, nije moguće pokazati konvergenciju rezultata (Mairal et al., 2010).

U slučaju velikog broja Gausovih raspodela i relativno visokodimenzionalnog prostora obeležja broj vektora ($D\sum_{s=1}^{S} M_s$) koji je na raspolaganju za određivanje rečnika je prilično velik, tako da standardne procedure za određivanje rečnika koje u jednoj iteraciji uzimaju u obzir sve vektore zbog konačnih resursa nisu pogodne, stoga je za ove potrebe iskorišćena procedura određivanja rečnika u letu (Mairal et al., 2009). Osnovni opis ove procedure naveden je u poglavlju 3. Za realizaciju ove procedure iskorišćena je varijanta koja je implementirana u okviru javno-dostupne biblioteke SPAMS (*SPArse Modeling Software*), a koju je moguće preuzeti sa adrese: `http://spams-devel.gforge.inria.fr`. Procedura se relativno jednostavno uklapa u standardnu proceduru estimacije parametara HMM-GMM-a baziranu na principu maksimizacije verodostojnosti što je detaljno opisano u narednom odeljku.

4.3.1 *Procedura za etimaciju parametara modela*

U prethodnom tekstu akcenat je stavljen na ono što je novo u odnosu na standardni HMM-GMM, odnosno na način modelovanja emitujućih gustina raspodela preko nove organizacije GMM-a, tako da nisu razmatrani parametri kojima se opisuje vremenska varijabilnost govora. Da bi se obuhvatili svi parametri koji opisuju govor prethodno opisani skup parametara treba proširiti verovatnoćama prelaza između HMM stanja kao i inicijalnim verovatnoćama HMM stanja. Verovatnoća prelaza iz stanja s_1 u stanje s_2 će biti označena sa $a_{s_1 s_2}$, dok će inicijalna verovatnoća stanja s biti označena sa π_s.

Procedura obuke započinje estimacijom parametara HMM-GMM modela koja je zasnovana na kriterijumu maksimalne izglednosti, primenom algoritma očekivanje-maksimizacija (EM *Expectation-Maximization*). Treba napomenuti da ovaj inicijalni GMM koristi pune kovarijansne matrice. Broj Gausovih raspodela (komponenti mešavina) po stanju se određuje na osnovu jednostavne validacione procedure, koja podrazumeva podelu skupa za obuku na nekoliko podskupova iste veličine od kojih se jedan bira za procenu prosečnog logaritma izglednosti rezultujućeg modela (tzv. validacioni podskup) dok se preostali podskupovi koriste za estimaciju parametara modela (obuku). Procedura započinje sa GMM-om koji ima samo jednu komponentu, pri čemu se broj komponenti postepeno povećava sve dok prosečan logaritam izglednosti na validacionom skupu ne počne da opada. Prosečan logaritam izglednosti na validacionom skupu se izračunava kao aritmetička sredina prosečnih logaritama izglednosti dobijenih sa svim validacionim podskupovima. Treba primetiti da se za dati broj komponenti po mešavini svaki od formiranih podskupova koristi za validaciju, čime je postignuto da je praktično celokupan skup za obuku iskorišćen za validaciju. U slučaju velikog broja opservacija u skupu za obuku, a u cilju ubrzanja procedure procene broja komponenti po mešavini, nije neophodno uključiti sve podskupove za validaciju. Iako ovaj pristup praktično smanjuje broj paralelnih estimacija parametara GMM-a, cena koja se pri tome plaća jeste lošija procena prosečnog logaritma izglednosti na validacionom skupu (veće odstupanje od stvarne vrednosti), što zahteva relaksaciju kriterijum prekida procedure, tako što se procedura neće zaustaviti za broj klastera za koji je opala prosečna izglednost na validacionom skupu, već kada započne trend pada. Broj mešavina koji se odredi u toku ove procedure se u narednim koracima više ne menja.

Nakon što se za svako HMM stanje odredi broj komponenata GMM-a koje opisuju gustinu raspodele verovatnoće emitovanja stanja, pristupa se estimaciji parametara modela pomoću standardnih izraza za EM algoritam:

$$\pi_s = \frac{\gamma_1(s)}{\sum_{s_i=1}^{S} \gamma_1(s_i)} \tag{90}$$

$$a_{s_i,s_k} = \frac{\sum_{t=2}^{T} \xi_t(s_i, s_k)}{\sum_{s_j=1}^{S} \sum_{t=2}^{T} \xi_t(s_i, s_j)} \tag{91}$$

$$w_{sm} = \frac{\sum_{t=1}^{T} \gamma_t(s, m)}{\sum_{j=1}^{M_s} \sum_{t=1}^{T} \gamma_t(s, j)} = \frac{\sum_{t=1}^{T} \gamma_t(s, m)}{\sum_{t=1}^{T} \gamma_t(s)} \tag{92}$$

$$\boldsymbol{\mu}_{sm} = \frac{\sum_{t=1}^{T} \gamma_t(s,m)\mathbf{o}_t}{\sum_{t=1}^{T} \gamma_t(s,m)} \tag{93}$$

$$\boldsymbol{\Sigma}_{sm} = \frac{\sum_{t=1}^{T} \gamma_t(s,m)\,(\mathbf{o}_t - \boldsymbol{\mu}_{sm})\,(\mathbf{o}_t - \boldsymbol{\mu}_{sm})^T}{\sum_{t=1}^{T} \gamma_t(s,m)} \tag{94}$$

gde je S ukupan broj HMM stanja u modelu, $\gamma_t(s)$ verovatnoća da se model našao u HMM stanju s u trenutku t ako je poznato da je generisana sekvenca opservacija $\mathbf{O} = [\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_T]$, $\gamma_t(s,m)$ verovatnoća da se model našao u HMM stanju s i u komponenti mešavine m u trenutku t ako je poznato da je generisana sekvenca opservacija $\mathbf{O}$, $\xi_t(s_1, s_2)$ verovatnoća da se model u trenutku t našao u HMM stanju s_2 i da je u prethodnom trenutku bio u HMM stanju s_1 ako je poznato da je generisana sekvenca opservacija $\mathbf{O}$, dok su značenja svih ostalih promenljivih nepromenjena. Ova estimacija parametara inicijalnog modela zahteva nekoliko iteracija EM algoritma.

Parametri: Dat je skup opservacija $\{\mathbf{o}_t\}$.

1: Estimirati parametre HMM-a (π_s,a_{s_1,s_2}) i GMM-a sa punim kovarijansnim matricama ($w_{sm}, \boldsymbol{\mu}_{ms}, \boldsymbol{\Sigma}_{ms}$) primenom standardne EM procedure obuke.
2: **Za** EM iteraciju $k = 1$ **do** N_{em} **radi sledeće:**
3: Izračunati karakteristične vektore $\boldsymbol{\nu}_{smi}$ i karakteristične vrednosti λ_{smi}^{-1} za svaku kovarijansnu matricu $\boldsymbol{\Sigma}_{sm}$.
4: **Ako** $k = 1$ **onda**
5: Inicijalizovati rečnik $\mathbf{D}$.
6: **Kraj ako**
7: **Za** iteraciju $j = 1$ **do** N_{sp} **radi sledeće:**
8: Izračunati retke kodove za svaki karakteristični vektor $\boldsymbol{\alpha}_{smi}$.
9: Za nove retke kodove odrediti novi rečnik $\mathbf{D}$.
10: **Kraj Za**
11: Izračunati retke kodove za svaki karakteristični vektor $\boldsymbol{\alpha}_{smi}$.
12: **Ako** $k \neq N_{em}$ **onda**
13: Izračunati verovatnoće unapred i unazad.
14: Izračunati inicijalne verovatnoće stanja π_s, verovatnoće prelaza između stanja a_{s_1,s_2}, težine komponenti mešavina w_{sm}, kao i njihove srednje vrednosti $\boldsymbol{\mu}_{sm}$ i kovarijansne matrice $\boldsymbol{\Sigma}_{sm}$.
15: **Kraj ako**
16: **Kraj Za**

Slika 17: Procedura za estimaciju parametara SEGMM. Sa N_{em} je označen broj iteracija EM algoritma, sa N_{sp} broj iteracija za određivanje rečnika.

Na slici 17 je u formi pseudo koda prikazana procedura obuke SEGMM-a. Pošto se predloženi model zasniva na ideji retke reprezentacije karakterističnih vektora, prvi korak u estimaciji parametara modela predstavlja izračunavanje karakterističnih vektora $\boldsymbol{\nu}_{smi}$ i karakterističnih vrednosti λ_{smi}^{-1} svih kovarijansnih matrica koje čine model (korak 3 na slici 17). Inicijalizacija vektora se vrši izborom grupa karakterističnih vektora, pri čemu te grupe obrazuju karakteristični vektori jedne kovarijansne matrice. Motivacija za ovakav izbor inicijalnih atoma rečnika leži u činjenici da su karakteristični vektori jedne kovarijansne matrice međusobno ortogonalni i da je na ovaj način moguće pokriti više različitih pravaca u prostoru karakterističnih vektora, i time donekle ubrzati proce-

duru određivanja rečnika. Treba istaći da se inicijalizacija rečnika (korak 5 na slici 17) realizuje samo na početku, kada ne postoji prethodno određeni rečnik.

Nakon izračunavanja karakterističnih vektora svih kovarijansnih matrica i formiranja inicijalnog rečnika pristupa se određivanju retkih kodova i atoma (koraci 7-10 na slici 17). Za ove potrebe kao što je već navedeno iskorišćena je metoda učenja rečnika u letu uz l_0 regularizator, koja je detaljno opisana u poglavju 3. Umesto osnovne varijante ovog algoritma, koja podrazumeva određivanje rečnika vektor po vektor, izabrana je varijansa u kojoj se preračun novog rečnika vrši uzimanjem grupa vektora, da bi se povećala brzina konvergencije (Mairal et al., 2009). Nakon N_{sp} iteracija koje su imale za cilj određivanje rečnika, vrši se ponovno izračunavanje retkih kodova za svaki karakteristični vektor (korak 11 na slici 17). Da bi se upotpunio SEGMM pored rečnika i retkih kodova potrebno je težine (w_{sm}) i srednje vrednosti ($\boldsymbol{\mu}_{sm}$) zameniti odgovarajućim konstantama c_{sm} i projekcijama srednjih vrednosti $\boldsymbol{\mu}_{sm}^{-}$ datih jednačinama (83) i (84) respektivno.

Nastavak modifikovane EM procedure podrazumeva poravnanje modela sa odgovarajućim opservacijama i izračunavanje verovatnoća unapred (*forward*) i unazad (*backward*) koje se koriste dalje za izračunavanje vrednosti $\gamma_t(s)$ i $\xi_t(s_1, s_2)$ (korak 13 na slici 17). Ovaj modifikovani E (očekivanje) korak se razlikuje od standardnog E koraka EM algoritma za HMM-GMM opisanog u (Huang et al., 2001), po tome što se za izračunavanje gustina verovatnoća emitovana koristi SEGMM. Sa druge strane M (maksimizacija) korak podrazumeva izračunavanje parametara standardnog HMM-GMM primenom jednačina (90–94) (korak 14 na slici 17). Naravno prethodna dva koraka (koraci 13 i 14 na slici 17) nisu neophodna ukoliko se procedura obuke završava. Ukoliko to nije slučaj prethodno opisana procedura se nastavlja od faze u kojoj se izračunavaju karakteristični vektori i vrednosti svih estimiranih punih kovarijansnih matrica (korak 3 na slici 17).

Kod sistema koji su predstavljeni u ovom radu realizovane su samo dve iteracije EM algoritma ($N_{em} = 2$) pri čemu druga iteracije nije rezultovala poboljšanjem performansi prepoznavača. Sa druge strane broj iteracija za estimaciju rečnika je bio relativno veliki ($N_{sp} = 240$).

4.4 OPRAVDANOST KRITERIJUMSKE FUNKCIJE

U ovom delu je formalno pokazano da se minimizovanje KLD-a između dve D-dimenzionalne Gausove raspodele koje imaju iste srednje vrednosti, i kovarijansne matrice sa istim karakterističnim vrednostima može svesti na minimizaciju euklidskog rastojanja između karakterističnih vektora koji odgovaraju istim karakterističnim vrednostima. U cilju skraćenja notacije iz oznaka za pune kovarijansne matrice, karakteristične vektore i karakteristične vrednosti biće izostavljeni indeksi koji označavaju pripadnost komponente određenom stanju i mešavini. Stoga će u daljem tekstu originalna puna kovarijansna matrica biti označena sa $\boldsymbol{\Sigma}_o$, njeni karakteristični vektori sa $\boldsymbol{\nu}_{oi}$, njene karakteristične vrednosti λ_i^{-1} njena aproksimaciona matrica sa $\boldsymbol{\Sigma}_a$, a karakteristični vektori aproksimacione matrice sa $\boldsymbol{\nu}_{ai}$.

Definicioni izraz za Kulbak-Lajbler divergencu gustine raspodele $q(x)$ do gustine raspodele $p(x)$ je:

$$D_{KL}(p||q) = \int_{-\infty}^{\infty} p(x) \ln\left(\frac{p(x)}{q(x)}\right) dx \tag{95}$$

Može se pokazati da u slučaju dve Gausove raspodele $\mathcal{N}_1$ i $\mathcal{N}_2$ opisane parametrima $(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ i $(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ KLD svodi na sledeći izraz:

$$D_{KL}(\mathcal{N}_1||\mathcal{N}_2) = \frac{1}{2}\left((\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}_2^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) - \ln\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_2|} + \operatorname{trag}\left(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1\right) - D\right) \tag{96}$$

U slučaju kada dve Gausove raspodele imaju istu srednju vrednost prvi član se eliminiše, tako da se izraz (96) za problem od interesa svodi na:

$$D_{KL}(\mathcal{N}_o||\mathcal{N}_a) = \frac{1}{2}\left(\ln\frac{|\boldsymbol{\Sigma}_a|}{|\boldsymbol{\Sigma}_o|} + \operatorname{trag}\left(\boldsymbol{\Sigma}_a^{-1}\boldsymbol{\Sigma}_o\right) - D\right) \tag{97}$$

Pošto za bilo koje $a, b \in \mathbb{R}$ važi $a + b \leqslant |a| + |b|$[6] KLD Gausove raspodele sa aproksimiranom kovarijansnom matricom do Gausove raspodele sa originalnom kovarijansnom matricom se može ograničiti sa gornje strane na sledeći način:

$$D_{KL}(\mathcal{N}_o||\mathcal{N}_a) \leqslant \frac{1}{2}\left|\ln\frac{|\boldsymbol{\Sigma}_o|}{|\boldsymbol{\Sigma}_a|}\right| + \frac{1}{2}\left|\operatorname{trag}(\boldsymbol{\Sigma}_a^{-1}\boldsymbol{\Sigma}_o) - D\right| \tag{98}$$

Da bi se gornja granica prikazala kao funkcija odstupanja aproksimirane karakteristične vrednosti od originalne biće iskorišćene sledeće činjenice:

- marica $\boldsymbol{\Sigma}_o$ se može dekomponovati na sledeći način $\boldsymbol{\Sigma}_o = \boldsymbol{V}_o \boldsymbol{\Lambda}^{-1} \boldsymbol{V}_o^T$, gde je $\boldsymbol{V}_o$ matrica koja sadrži karakteristične vektore matrice $\boldsymbol{\Sigma}_o$ odnosno $\boldsymbol{V}_o = [\boldsymbol{v}_{o1}, \boldsymbol{v}_{o2}, \ldots, \boldsymbol{v}_{oD}]$, a $\boldsymbol{\Lambda}$ dijagonalna matrica koja sadrži karakteristične vrednosti matrice $\boldsymbol{\Sigma}_o^{-1}$ odnosno $\Lambda_{ii} = \lambda_i$[7] za $i = 1, 2 \ldots, D$,
- matrica $\boldsymbol{\Sigma}_a$ se može prikazati na sledeći način $\boldsymbol{\Sigma}_a = (\boldsymbol{V}_o + \boldsymbol{E})\boldsymbol{\Lambda}^{-1}(\boldsymbol{V}_o + \boldsymbol{E})^T$ gde je sa $\boldsymbol{E}$ označena matrica grešaka aproksimacije čije su kolone $\boldsymbol{e}_i = \boldsymbol{v}_{ai} - \boldsymbol{v}_{oi}$ za $i = 1, 2 \ldots, D$.

U cilju bolje preglednosti u daljem tekstu svaki od članova izraza (98) će se razmatrati zasebno.

Prvi član izraza (98) može se transformisati korišćenjem osobina determinante i činjenice da je apsolutna vrednost determinantne matrice $\boldsymbol{V}_o$ jednaka 1, na sledeći način:

6 Oznaka za determinantu matrice i apsolutnu vrednost je identična ($|\cdot|$) ali na osnovu oznake promenljive je jasno koja je od ove dve operacije u pitanju pošto su matrice označene velikim masnim slovima, a skalari normalnim slovima.

7 Da bi se izbeglo uvođenje novih oznaka, karakteristične vrednosti kovarijansnih matrica izražene su preko karakterističnih vrednosti odgovarajućih inverznih matrica pošto su one mnogo ranije uvedene.

$$\left|\ln\frac{|\boldsymbol{\Sigma}_o|}{|\boldsymbol{\Sigma}_a|}\right| = \left|\ln\frac{\left|(\mathbf{V}_o+\mathbf{E})\boldsymbol{\Lambda}^{-1}(\mathbf{V}_o+\mathbf{E})^T\right|}{\left|\mathbf{V}_o\boldsymbol{\Lambda}^{-1}\mathbf{V}_o^T\right|}\right| \tag{99}$$

$$= \left|\ln\frac{|\mathbf{V}_o+\mathbf{E}|\,|\boldsymbol{\Lambda}|^{-1}\,|\mathbf{V}_o+\mathbf{E}|}{|\mathbf{V}_o|\,|\boldsymbol{\Lambda}|^{-1}\,|\mathbf{V}_o|}\right| \tag{100}$$

$$= \left|\ln|\mathbf{V}_o+\mathbf{E}|^2\right| \tag{101}$$

Teorema u teoriji preturbacija matrica (Ipsen i Rehman, 2008) tvrdi sledeće:

$$|\,|\mathbf{V}_o|-|\mathbf{V}_o+\mathbf{E}|\,| \leqslant s_{D-1}\|\mathbf{E}\|_2+O(\|\mathbf{E}\|_2^2) \tag{102}$$

gde je $\|\cdot\|_2$ 2-norma matrice,[8] i s_{D-1} je $D-1$-va elementarna simetrična funkcije singularnih vrednosti matrice $\mathbf{V}_o$.[9] Pošto je $s_{D-1} \leqslant D\sigma_1\sigma_2\cdots\sigma_{D-1}$ (Ipsen i Rehman, 2008) i pošto su singularne vrednosti matrice $\mathbf{V}_o$ koja je unitarna jednake 1 (Golub i van Van Loan, 1996) sledi da je $s_{D-1} \leqslant D$, odnosno jednačina (102) dobija sledeći oblik:

$$|\,|\mathbf{V}_o|-|\mathbf{V}_o+\mathbf{E}|\,| \leqslant D\|\mathbf{E}\|_2+O(\|\mathbf{E}\|_2^2) \tag{103}$$

Pošto je $\mathbf{V}_o$ unitarna tada je njena determinanta $|\mathbf{V}_o|$ jednaka ili 1 ili -1 (Golub i van Van Loan, 1996). Ukoliko je $|\mathbf{V}_o|=1$ tada na osnovu (103) sledi:

$$1-\left(D\|\mathbf{E}\|_2+O(\|\mathbf{E}\|_2^2)\right) \leqslant |\mathbf{V}_o+\mathbf{E}| \leqslant 1+\left(D\|\mathbf{E}\|_2+O(\|\mathbf{E}\|_2^2)\right) \tag{104}$$

odnosno ukoliko je $|\mathbf{V}_o|=-1$ onda:

$$-\left(1+D\|\mathbf{E}\|_2+O(\|\mathbf{E}\|_2^2)\right) \leqslant |\mathbf{V}_o+\mathbf{E}| \leqslant -\left(1-D\|\mathbf{E}\|_2-O(\|\mathbf{E}\|_2^2)\right) \tag{105}$$

Na osnovu nejednakosti (104) i (105) i činjenice da je $D\|\mathbf{E}\|_2+O(\|\mathbf{E}\|_2^2)$ nenegativno sledi:

$$|\,|\mathbf{V}_o+\mathbf{E}|\,| \leqslant 1+\left(D\|\mathbf{E}\|_2+O(\|\mathbf{E}\|_2^2)\right) \tag{106}$$

Stoga prvi član u izrazu (98) se može ograničiti sa gornje strane na sledeći način:

$$\left|\ln|\mathbf{V}_o+\mathbf{E}|^2\right| = |2\ln|\,|\mathbf{V}_o+\mathbf{E}|\,|| \leqslant \left|2\ln\left(1+D\|\mathbf{E}\|_2+O(\|\mathbf{E}\|_2^2)\right)\right| \tag{107}$$

odnosno koristeći poznatu nejednakost $\ln(1+x) \leqslant x$ za svako $x \in \mathbb{R}^+$

$$\left|\ln|\mathbf{V}_o+\mathbf{E}|^2\right| \leqslant 2D\|\mathbf{E}\|_2+2O(\|\mathbf{E}\|_2^2) \tag{108}$$

Cilj je pokazati da je gornja granica greške aproksimacije funkcija l_2-norme razlike odgovarajućih karakterističnih vektora (koji predstavljaju kolone matrice

8 Za proizvoljnu matricu $\mathbf{A}$ 2-norma se definiše pomoću količnika l_2-normi vektora $\mathbf{A}x$ i x na sledeći način: $\|\mathbf{A}\|_2 = \sup_{x\neq 0}\frac{\|\mathbf{A}x\|_2}{\|x\|_2}$.

9 Za kvadratnu matricu $\mathbf{A}$ dimenzija $D\times D$ čije su singularne vrednosti $\sigma_1 \geqslant \ldots \geqslant \sigma_D$, k-ta elementarna simetrična funkcija simetričnih vrednosti definisana je sa:

$\mathbf{E}$) potrebno je 2-normu matrice $\mathbf{E}$ prevesti u Frobenijusovu normu.[10] Ubacivanjem sledeće nejednakosti (Golub i van Van Loan, 1996):

$$\|\mathbf{E}\|_2 \leqslant \|\mathbf{E}\|_F = \sqrt{\sum_{i=1}^{D}\sum_{j=1}^{D} E_{ij}^2} = \sqrt{\sum_{i=1}^{D} \|\boldsymbol{e}_i\|_2^2} \tag{109}$$

u izraz (108) dobija se:

$$\left|\ln|\mathbf{V}_o + \mathbf{E}|^2\right| \leqslant 2D\sqrt{\sum_{i=1}^{D}\|\boldsymbol{e}_i\|_2^2} + 2O(\sum_{i=1}^{D}\|\boldsymbol{e}_i\|_2^2). \tag{110}$$

čime je završeno ograničavanje prvog člana izraza (98).

Kovarijansne matrice koje figurišu u drugom članu izraza (98) se takođe mogu predstaviti pomoću proizvoda odgovarajućih karakterističnih vektora i vrednosti odnosno:

$$\left|\text{trag}(\boldsymbol{\Sigma}_a^{-1}\boldsymbol{\Sigma}_o) - D\right| = \left|\text{trag}\left(\left(\mathbf{V}_o^T + \mathbf{E}^T\right)^{-1}\boldsymbol{\Lambda}\left(\mathbf{V}_o + \mathbf{E}\right)^{-1}\mathbf{V}_o\boldsymbol{\Lambda}^{-1}\mathbf{V}_o^T\right) - D\right| \tag{111}$$

Ukoliko se iskoristi činjenica da je matrica $\mathbf{V}_o$ unitarna[11] za dovoljno male vrednosti $\mathbf{E}$ može se pokazati da važi $(\mathbf{V}_o + \mathbf{E})(\mathbf{V}_o + \mathbf{E})^T \approx \mathbf{I}$ odnosno da je $\left(\mathbf{V}_o^T + \mathbf{E}^T\right)^{-1} \approx (\mathbf{V}_o + \mathbf{E})$ i $(\mathbf{V}_o + \mathbf{E})^{-1} \approx (\mathbf{V}_o + \mathbf{E})^T$, te izraz (111) dobija sledeći oblik:

$$\left|\text{trag}(\boldsymbol{\Sigma}_a^{-1}\boldsymbol{\Sigma}_o) - D\right| \approx \left|\text{trag}\left(\left(\mathbf{V}_o + \mathbf{E}\right)\boldsymbol{\Lambda}\left(\mathbf{V}_o + \mathbf{E}\right)^T\mathbf{V}_o\boldsymbol{\Lambda}^{-1}\mathbf{V}_o^T\right) - D\right| \tag{112}$$

Primenom osobina cikličnosti traga, kao i činjenice da je $\mathbf{V}_o$ unitarna matrica trag sa desne strane u izrazu (112) se može dodatno uprostiti na sledeći način:

$$\begin{aligned}
&\text{trag}\left(\left(\mathbf{V}_o + \mathbf{E}\right)\boldsymbol{\Lambda}\left(\mathbf{V}_o + \mathbf{E}\right)^T\mathbf{V}_o\boldsymbol{\Lambda}^{-1}\mathbf{V}_o^T\right) \\
&\quad= \text{trag}\left(\boldsymbol{\Lambda}\left(\mathbf{V}_o + \mathbf{E}\right)^T\mathbf{V}_o\boldsymbol{\Lambda}^{-1}\mathbf{V}_o^T\left(\mathbf{V}_o + \mathbf{E}\right)\right) \\
&\quad= \text{trag}\left(\boldsymbol{\Lambda}\left(\mathbf{I} + \mathbf{E}^T\mathbf{V}_o\right)\boldsymbol{\Lambda}^{-1}\left(\mathbf{I} + \mathbf{V}_o^T\mathbf{E}\right)\right) \\
&\quad= \sum_{i=1}^{D}\sum_{j=1 j\neq i}^{D}\frac{\lambda_i}{\lambda_j}\left(\boldsymbol{e}_i^T\boldsymbol{v}_j\right)^2\sum_{i=1}^{D}\left(1 + \boldsymbol{e}_i^T\boldsymbol{v}_i\right)^2 \\
&\quad= \sum_{i=1}^{D}\sum_{j=1}^{D}\frac{\lambda_i}{\lambda_j}\left(\boldsymbol{e}_i^T\boldsymbol{v}_j\right)^2 + 2\sum_{i=1}^{D}\boldsymbol{e}_i^T\boldsymbol{v}_i + D
\end{aligned} \tag{113}$$

čime se dobija:

$$\left|\text{trag}(\boldsymbol{\Sigma}_a^{-1}\boldsymbol{\Sigma}_o) - D\right| \approx \left|\sum_{i=1}^{D}\sum_{j=1}^{D}\frac{\lambda_i}{\lambda_j}\left(\boldsymbol{e}_i^T\boldsymbol{v}_j\right)^2 + 2\sum_{i=1}^{D}\boldsymbol{e}_i^T\boldsymbol{v}_i\right| \tag{114}$$

10 Frobeniusova norma koja je za $m \times n$ dimenzionalnu matricu definisana sa $\|\mathbf{X}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}X_{ij}^2}$.

11 Proizvoljna matrica $\mathbf{A}$ je unitarna ukoliko važi $\mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A} = \mathbf{I}$, odnosno $\mathbf{A}^{-1} = \mathbf{A}^*$ gde je * konjugovano transponovanje.

Pošto je vrednost skalarnog (unutrašnjeg) proizvoda proizvoljnog vektora $\boldsymbol{x}$ i jediničnog vektora $\boldsymbol{u}$ uvek manja ili jednaka od l_2-norme vektora $\boldsymbol{x}$, sledi da je $\boldsymbol{e}_i^T \boldsymbol{v}_j \leqslant \|\boldsymbol{e}_i\|_2$ odnosno:

$$\left|\operatorname{trag}(\boldsymbol{\Sigma}_a^{-1}\boldsymbol{\Sigma}_o) - D\right| \leqslant \left|\sum_{i=1}^{D}\sum_{j=1}^{D}\frac{\lambda_i}{\lambda_j}\|\boldsymbol{e}_i\|_2^2 + 2\sum_{i=1}^{D}\|\boldsymbol{e}_i\|_2\right| \tag{115}$$

Na osnovu jednačina (98), (110) i (115) sledi da je gornja granica KLD direktna funkcija euklidskog rastojanja između odgovarajućih karakterističnih vektora odnosno:

$$\begin{aligned} D_{KL}(\mathcal{N}_o\|\mathcal{N}_a) \leqslant \quad & D\sqrt{\textstyle\sum_{i=1}^{D}\|\boldsymbol{e}_i\|_2^2} + O(\textstyle\sum_{i=1}^{D}\|\boldsymbol{e}_i\|_2^2) + \\ & + \tfrac{1}{2}\textstyle\sum_{i=1}^{D}\sum_{j=1}^{D}\frac{\lambda_i}{\lambda_j}\|\boldsymbol{e}_i\|_2^2 + \sum_{i=1}^{D}\|\boldsymbol{e}_i\|_2 \end{aligned} \tag{116}$$

Treba napomenuti da su iz gornjeg izraza izostavljene apsolutne vrednosti pošto su norme vektora nenegativne, kao i karakteristične vrednosti kovarijansnih matrica jer su pozitivno definitne.

Kao što se iz priloženog može videti, smanjivanjem euklidskog rastojanja između originalnog karakterističnog vektora i njegove aproksimacije (koja može da bude dobijena pomoću retke reprezentacije), smanjuje se razlika između originalne Gausove respodele i njene aproksimacije što za posledicu ima i manju grešku prilikom izračunavanja izglednosti pojedinačne opservacije. Iz priloženog se može videti da u slučaju loše kondicioniranih kovarijansnih matrica količnik karakterističnih vrednosti λ_i/λ_j može postati dominantan faktor u izrazu (116) i u slučajevima kada su razlike između originalnog i aproksimiranog karakterističnog vektora male. Stoga da bi predloženi model funkcionisao neophodno je obezbediti da kovarijansne matrice budu dobro kondicionirane, odnosno da se za estimaciju parametara svake Gausove raspodele obezbedi dovoljan broj opservacija.

4.5 REZIME

U ovom poglavlju je izložen detaljan pregled modela koji se zasniva na aproksimaciji inverznih kovarijansnih matrica korišćenjem retke reprezentacije njihovih karakterističnih vektora. Pored samog formalnog opisa modela data je i analiza njegovih prednosti kao i mana u odnosu na postojeće varijante GMM-a sa aspekta broja potrebnih parametara za opis modela i broja računskih operacija za izračunavanje izglednosti. Kao i u slučaju većine drugih naprednijih pristupa modelovanju zasnovanih na GMM-u broj parametara i računskih operacija je nekoliko puta veći od broja koji je potreban u slučaju GMM-a sa dijagonalnim aproksimacijama kovarijansnih matrica i značajno manji od broja koji je potreban za GMM-a sa punim kovarijansnim matricama. U odnosu na načine modelovanja koji postižu tačnost blisku tačnosti GMM-a sa punim kovarijansnim matricama, broj parametara kao i broj izračunavanja je približan, ali to zavisi i od izbora konkretnih parametara modela što će biti detaljnije obrađeno u eksperimentalnom delu ovog rada. Iako je broj parametara redukovan u odnosu na varijantu GMM-a sa punim kovarijansnim matricama, za uspešnu estimaciju parametara neophodno je obezbediti dobro kondicioniranu kovarijansnu matricu, odnosno dovoljan broj opservacija.

Pošto je minimizacija euklidskog rastojanja između odgovarajućih[12] karakterističnih vektora kovarijansnih matrica isto što i minimizacija KLD-a između Gausovih raspodela koje opisuju, obuka postojećeg modela se prirodno uklapa u obuku na principu maksimizacije izglednosti. Iako je sam model opisan sa značajno manjim brojem parametara nego model sa punim kovarijansnim matricama, procedura obuke je značajno duža nego u slučaju obuke GMM-a sa punim kovarijansnim matricama, pošto je pored izračunavanja punih kovarijansnih matrica potrebno u iterativnoj proceduri odrediti rečnik i retke kodove za sve karakteristične vektore. S aspekta performansi modela, produženje procedure obuke nije kritičan faktor jer nije neophodno da se procedura obuke realizuje u realnom vremenu, ali je vrlo bitna redukcija broja operacija pri izračunavanju logaritma izglednosti opservacije.

12 Odgovarajućih u smislu da odgovaraju istim karakterističnim vrednostima.

5

PREGLED KORIŠĆENIH GOVORNIH BAZA I SOFTVERSKIH ALATA

5.1 UVOD

Ovaj odeljak daje pregled govornih baza i softverskih alata koji su korišćeni za obuku i testiranje analiziranih sistema za prepoznavanje govora. Za potrebe ovog rada su iskorišćene dve govorne baze koje se već duži niz godina u gotovo neizmenjenom obliku koriste za obuku i testiranje sistema za prepoznavanje govora na srpskom jeziku. Prva govorna baza, nosi oznaku SpeechDat II, snimljena je na Fakultetu tehničkih nauka u Novom Sadu i sadrži veći deo materijala koji se koristi za potrebe obuke i testiranja sistema za prepoznavanje govora. Druga baza nosi oznaku S70W100s120T je po obimu znatno manja, preuzeta je sa Elektrotehničkog fakulteta u Beogradu i naknadnom obradom prilagođena telefonskom kanalu. Za realizaciju eksperimenata u ovom radu korišćeni su softverski alati koji su razvijeni na Fakultetu tehničkih nauka, ali i neki javno dostupni softverski paketi, namenjeni optimizaciji ciljnih funkcija kao što je SPAMS.

5.2 OPIS GOVORNIH BAZA

Govorna baza SpeechDat II je snimljena na Fakultetu tehničkih nauka u Novom Sadu za potrebe istraživanja i razvoja govornih aplikacija kojima bi se pristupalo preko javne telefonske mreže. Svi audio fajlovi sadrže snimke govornih signala koji su prošli kroz javnu telefonsku mrežu i snimljeni su u PCM formatu sa po 16 bita po odmerku i učestanošću odabiranja 8 kHz. Govorna baza je podeljena na dva disjunktna podskupa od kojih se jedan koristi isključivo za obuku sistema i drugi koji se koristi za testiranje. Deo baze koji je namenjen obuci sistema sačinjavaju snimci 513 govornika (266 muških i 247 ženskih) ukupnog trajanja 28.5 sati od čega 12 sati čini govor, a 16.5 sati tišina i oštećeni govorni segmenti. Prateće transkripcije audio fajlova su na nivou fonema, pri čemu su granice između fonema "ručno" pregledane i po potrebi korigovane. Kao i druge baze snimljene po SpeechDat(E) standardu i ovu bazu čine snimci sa izolovano izgovorenim rečima (komande za upravljanje računarom, kretanje kroz menije, gradovi, horoskopski znaci, lična imena i prezimena), izolovano izgovorenim ciframa i sekvencama cifara, sintagme koje predstavljaju novčane iznose, datume kao i rečenice opšte sadržine. Deo baze koji je namenjen testiranju sačinjavaju snimci 184 govornika (107 muških i 77 ženskih) ukupnog trajanja od oko 60 minuta od čega je oko 32 minuta govor, a 28 minuta tišina i šum. Pošto se u testovima ne koristi blok za automatsku detekciju govorne aktivnosti, celokupan materijal se koristi pri prepoznavnaju, tako da za razliku od baze za obuku ova podela na govorne i negovorne segmente nije u toj meri bitna, kao u slučaju baze koja se koristi za obuku.

Govorna baza koja nosi oznaku S70W100s120 je snimljena u studijskim uslovima na Elektrotehničkom fakultetu u Beogradu 80-ih godina prošlog veka. Inicijalni zvučni zapis je bio analogni, koji je naknadno u prostorijama Radio Novog Sada prebačen u digitalni PCM format sa po 16 bita po odmerku i sa učestanošću odabiranja od 22.05 kHz. Za potrebe istraživanja i razvoja sistema za prepoznavanje govora telefonskog kvaliteta formirana je nova baza koja nosi oznaku S70W100s120T, koja je dobijena propuštanjem signala kroz FIR filtre koji su simulirali telefonske kanale, smanjivanjem učestanosti odabiranja na 8 kHz i dodavanjem određenog nivoa Gausovog šuma. Bazu S70W100s120T sačinjavaju snimici 180 govornika (109 muškaraca i 71 žena) ukupnog trajanja 6.4 sata od čega 3.7 sati čini govor i 2.7 sati tišina. Snimci sadrže izolovano izgovorene reči (uglavnom vojne komande i termine), cifre od 0 do 9, kao i rečenice opšte tematike. Prilikom formiranja baze vođeno je računa da baza bude fonetski izbalansirana, tako da je frekvencija pojavljivanja fonema koji se retko sreću u srpskom jeziku nešto veća od njihove frekvencije u prirodnom jeziku. Kao i u slučaju SpeecDat II i u ovoj bazi prateće transkripcije audio fajlova su na nivou fonema, pri čemu su granice između fonema "ručno" pregledane i po potrebi korigovane.

5.3 OPIS KORIŠĆENIH SOFTVERSKIH ALATA

Osnovni skup softverskih alata koji su korišćeni za obuku i testiranje sistema za automatsko prepoznavanje govora su razvijeni na Fakultetu tehničkih nauka Univerziteta u Novom Sadu. Ovi alati su realizovani u programskom jeziku C++ korišćenjem Microsoftovog razvojnog okruženja Visual Studio. Kod je organizovan u nekoliko biblioteka u zavisnosti od zadataka koje pojedine C++ klase odnosno funkcije (procedure) vrše, te razlikujemo sledeće:

- **slib** – biblioteka namenjena obradi signala
- **sslib** – biblioteka za konverziju ASCII tekstualnih fajlova u odgovarajuće sisteme za obradu signala kombinujući elemente slib bibilioteke.
- **csrlib** – biblioteka namenjena obuci i testiranju sistema za prepoznavanje govora
- **an_misc** – biblioteka opšte namene, namenjena obradi i parsiranju teksta, streamingu podataka, radu sa fajlovima i sl..

Sve modifikacije postojećih funkcija kao i nove funkcije koje su bile neophodne za realizaciju sistema opisanih u ovom radu autor ovog rada je samostalno implementirao i testirao.

Estimacija transformacionih matrica za slučaj HLDA realizovana je u programskom paketu Matlab korišćenjem koda koji je naveden u (Kumar, 1997). Za potrebe rada skup funkcija je proširen i varijantom transformacije koja pretpostavlja nezavisnost diskriminativnih transformisanih obeležja i zavisnosti nediskriminativnih transformisanih obeležja (Jakovljević et al., 2013). Za potrebe pronalaženja retke reprezentacije karakterističnih vektora kovarijansnih matrica, iskorišćena je javno dostupna Matlabova bibilioteka alata za optimizaciju SPAMS koju je moguće preuzeti sa `http://spams-devel.gforge.inria.fr`. Razlog zašto je izabrana ova implementacija za optimizaciju jeste da je u okviru

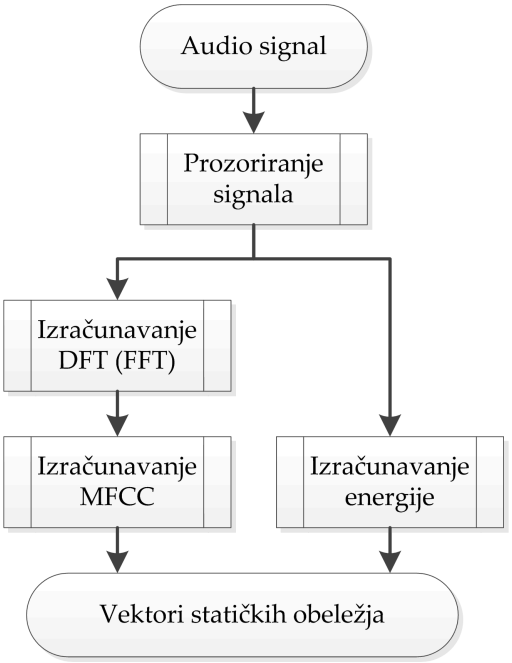
nje podržan algoritam određivanja rečnika u letu, koji je pogodan u situacijama kada je na raspolaganju izuzetno veliki broj vektora, što je bio slučaj u ovom radu.

5.3.1 *Ekstrakcija obeležja*

Modul za ekstrakciju obeležja je realizovan kombinovanjem i povezivanjem elemenata slib biblioteke. Blok šema realizacije ekstraktora je data na slici 18. Na ulazu ekstraktora se nalazi digitalizovan audio signal (16 bita po odmerku, učestanost odabiranja 8 kHz), koji se prozorira Hemingovom (*Hamming*) prozorskom funkcijom čija je širina 30 ms (240 odbiraka), što se vrši na svakih 10 ms (80 odbiraka). Za svaki prozorirani segment signala (u nastavku teksta će biti korišćen termin frejm) se izdvaja po jedan vektor obeležja. Da bi se izračunali mel-frekvencijski kepstralni koeficijenti (MFCC *Mel-frequency cepstral coefficient*) neophodno je prvo izračunati amplitudski spektar korišćenjem diskretne Furijeove transformacije (DFT). Pri izračunavanju MFCC-a delimično se oponaša proces koji se dešava na bazilarnoj membrani uha, odnosno spektar signala se deli na podopsege koji simuliraju kritične opsege, zatim se za svaki od podopsega procenjuje energija koja se potom logaritmuje (Morgan et al., 2004). Svrha logaritmovanja je smanjenje dinamičkog opsega amplituda pojedinih komponenata u spektru, te se ponekad logaritam zamenjuje funkcijom koja vrši kvadratno ili kubno korenovanje. Treba napomenuti da je logaritmovanje podesnije, pošto omogućava eliminisanje uticaja konvolutivnog šuma (kanala) jednostavnim oduzimanjem, što se koristi u postupku normalizacije srednjom (prosečnom) vrednošću kepstrala (CMN *Cepstral mean normalization*).[1] Zbog preklapanja koje postoji između spektralnih podopsega ovako izračunati logaritmi energija su u velikoj meri korelisani stoga se primenjuje diskretna kosinusna transformacija (DCT) koja ih u određenoj meri dekoreliše. Pošto je informacija o tome šta je rečeno sadržana u obvojnici spektra, DCT koeficijenti koji odgovaraju sinusoidama na višim učestanostima se odbacuju kao suvišni. Još jedan od razloga zašto se vrši odbacivanje ovih koeficijenata jeste redukcija dimenzionalnosti prostora obeležja (Chiu i Stern, 2008), pošto u sistemima namenjenim prepoznavanju oblika obeležja koja ne doprinose diskriminaciji između klasa nažalost najčešće povećavaju konfuziju između klasa (Jiang, 2011).

Pošto su svi sistemi koji su razmatrani u ovom radu namenjeni prepoznavanju govora telefonskog kvaliteta, što podrazumeva da je spektar signala ograničen na 4 kHz kao i da su komponente signala na niskim i visokim učestanostima značajno oslabljene, nakon izračunavanja DFT spektra ovi oštećeni segmenti, odnosno komponente signala koje se nalaze na učestanostima ispod

1 Ideja na kojoj se zasniva CMN je relativno jednostavna i podrazumeva korišćenje činjenice da konvoluciji u vremenskom domenu odgovara množenje u frekvencijskom domenu. Neka je $x(n)$ govorni signal, $h(n)$ impulsni odziv kanala i $y(n)$ govorni signal na izlazu iz kanala, tada važi $y(n) = x(n) * h(n)$ odnosno $Y(\omega) = X(\omega)H(\omega)$. Pošto se pri ekstrakciji obeležja posmatra samo amplitudski spektar (vrednost energije na pojedinim učestanosnima, ali ne i njena faza) tada se logaritmovanjem dobija: $\ln|Y(\omega)| = \ln|X(\omega)H(\omega)| = \ln|X(\omega)| + \ln|H(\omega)|$ odakle se jasno vidi da se do logaritma amplitudskog spektra originalnog govornog signala može doći tako što se od $\ln|Y(\omega)|$ oduzme $\ln|H(\omega)|$. Uprosečivanjem vrednosti $\ln|Y(\omega)|$ se dolazi do grube procene $\ln|H(\omega)|$ ali i dela govornog signala koji je sporopromenljiv odnosno ne nosi informaciju o tome šta je rečeno.



Slika 18: Blok šema ekstraktora obeležja.

50 Hz i iznad 3.8 kHz se zanemaruju prilikom izračunavanja MFCC obeležja. Preostali frekvencijski opseg (od 50 Hz do 3.8 kHz) se deli na 22 podopsega, koji su raspoređeni ekvidistantno na mel skali tako da je preklapanje između susednih podopsega 50%. Treba napomenuti da u ovoj realizaciji pojasni filtri koji dele spektar na podopsege imaju standardni oblik frekvencijskih karakteristika (na mel skali obrazuju jednokrake trouglove). Logaritmi energija ova 22 podopsega se primenom DCT transformišu u 13 MFCC-a, pri čemu se nulti koeficijent koji predstavlja grubu procenu energije frejma odbacuje (Huang et al., 2001). Umesto njega vektoru obeležja se dodaje energija koja je izračunata na standardan način kao suma kvadrata odmeraka koji pripadaju datom frejmu. Da bi se eliminisale varijacije energije koje su posledica različite glasnoće različitih govornika i promene glasnoće jednog govornika tokom vremena vrši se normalizacija energije (Zhu i O'Shaughnessy, 2005). Prvi korak pri normalizaciji energije jeste redukcija njene dinamike tako što se umesto energije posmatra njen četvrti koren, nakon čega se traže položaji maksimuma i same maksimalne vrednosti u prozorima širine 250ms. Normalizacija energije se vrši tako što vrednost četvrtog korena energije deli sa procenjenim lokalnim maksimumom koji se dobija kao linearna interpolacija njemu najbližih lokalnih maksimuma. Više detalja o prednostima ovakvog načina normalizacije energije moguće je naći u (Jakovljević et al., 2008).

Izdvajanje dinamičkih beležja (delta i delta-delta obeležja) se vrši naknadno pomoću odgovarajućih funkcija iz csrlib biblioteke. Funkcija koja se koristi za izračunavanje delta obeležja koristi regresioni obrazac:

$$\Delta c_{t,i} = \frac{\sum_{\theta=1}^{\Theta} \theta \left(c_{t+\theta,i} - c_{t-\theta,i} \right)}{2 \sum_{\theta=1}^{\Theta} \theta^2} \tag{117}$$

gde je $c_{t,i}$ i-to obeležje u trenutku t, a 2Θ broj susednih frejmova sa leve i desne strane posmatranog frejma koje je potrebno uzeti pri izračunavanju. Delta-

delta obeležja se dobijaju ukoliko se jednačina (117) primeni na prethodno izračunata delta obeležja. U eksperimentima koji su realizovani za potrebe ovog rada, pri izračunavanju delta obeležja parametar Θ je bio 2, a za delta-delta obeležja 1.

Da bi se veštački proširio skup za obuku i na taj način poboljšala efikasnost estimiranih parametara modela, prilikom obuke obeležja su izdvajana na svakih 5 ms. Ova promena nema uticaja na vrednosti statičkih obeležja, ali značajno utiče na dinamička, odnosno neophodno je uvesti odgovarajuću korekciju. Ovaj problem je prevaziđen tretiranjem dinamičkih obeležja kao izvoda kontinualne funkcije po vremenu odnosno:

$$\Delta c_{t,i} = \frac{dc_i(t)}{dt} = \frac{1}{T_s} \frac{\sum_{\theta=1}^{\Theta_{T_s}} \theta \left(c_{t+\theta,i} - c_{t-\theta,i} \right)}{2 \sum_{\theta=1}^{\Theta_{T_s}} \theta^2} \tag{118}$$

gde je sa T_s označeno vreme između dva susedna frejma, pri čemu se i vrednost Θ_{T_s} menja tako da se pri računanju dinamičkih obeležja uzima prozor istog ili približno istog trajanja. Pošto je usvojeno da se obeležja za vrednost pomeraja frejma od 10 ms računaju i dalje kao u izrazu (117), izraz sa desne strane jednačine (118) se dodatno množi sa 10 ms. Više detalja u vezi sa korišćenjem različitih vrednosti pomeraja frejma u procesu obuke i testiranja se može naći u (Pekar et al., 2010).

Iako se statističko modelovanje akustičkih karakteristika govora realizuje pomoću GMM-a koje uzimaju u obzir rasipanje (varijanse) pojedinih obeležja, ali samo na lokalnom nivou, sva obeležja se skaliraju sa njihovim standardnim devijacijama koje su procenjene na celokupnom skupu za obuku. Na ovaj način se ujednačava uticaj korišćenih obeležja i prevazilazi mogući problem da obeležja koja imaju veće rasipanje pošto imaju veće vrednosti budu bitnija za proces prepoznavanja (Theodoridis i Koutroumbas, 2006). Standardni način na koji se ovo prevazilazi u automatskom prepoznavanju govora je pomoću tzv. lifterovanja, množenja koeficijenata podignutom sinusnom funkcijom (Young et al., 2009).

Pored informacije o tome kako izgleda obvojnica spektra u datom trenutku koju nose statička obeležja za prepoznavanje govora bitan je i kontekst odnosno oblik obvojnice spektra u susednim frejmovima (Morgan et al., 2004). Potreba za poznavanjem konteksta i mogućnost redukcije dimenzionalnosti prostora obeležja koju pruža (H)LDA je poslužila kao motiv za formiranje tzv. konkatenativnih obeležja koja se dobijaju spajanjem (konkatenacijom) statičkih obeležja iz susednih frejmova. Treba napomenuti da se i kod standardnih obeležja informacija o kontekstu uključuje preko dinamičkih obeležja, što je bio jedan od motiva za njihovo korišćenje (Morgan et al., 2004).[2] Broj uzastopnih frejmova koji se koriste formiranje konkatenativnih obeležja varira od 5 do 11 čime je obuhvaćen interval od oko 50 ms do 110 ms.

5.3.2 *Način modelovanja*

Osnovna jedinica modelovanja je fonem zavisan od konteksta tzv. trifon. Izborom trifona za jedinicu modelovanja smanjuju se razlike unutar jednog mo-

2 Drugi razlog koji se znatno češće navodi kao obrazloženje za uvođenje dinamičkih obeležja jeste postizanje uslovne nezavisnosti između uzastopnih obeležja koju zahteva HMM (Gales i Young, 2008).

dela koje nastaju kao posledica koartikulacija u govoru. Biranjem većih jedinica modelovanja kao što su slogovi ili reči uticaj koartikulacije na varijabilnost modela bi se dodatno smanjio, ali bi se i značajno smanjila mogućnost proširivosti skupa reči koje bi sistem mogao da prepoznaje. Pored toga ovo povećanje jedinice modelovanja bez proširenja skupa za obuku bi smanjilo statističku efikasnost procene parametara. Problem obezbeđivanja statistički efikasne procene parametara se javlja i pri izboru trifona kao jedinice modelovanja, pošto se broj modela povećava, a količina podataka ostaje nepromenjena tako da pojedina HMM stanja dobiju nedovoljan broj vektora obeležja za adekvatnu estimaciju parametara Gausovih raspodela. Problem nedovoljnog broja parametara se delimično može prevazići spajanjem akustički i fonetski sličnih stanja. Standardni način za spajanje[3] sličnih stanja je pomoću klasterovanja na osnovu stabla (TBC *Tree-based clustering*) (Young et al., 1994).

Kao što je poznato standardna TBC procedura vrši spajanja stanja na osnovu opservacija koje su namenjene obuci sistema. Pošto skup za obuku sadrži iskaze više različitih govornika koji su snimani preko različitih telfonskih kanala razumno je očekivati da je rasipanje unutar klasa veće nego u slučaju da snimci sadrže iskaze jednog govornika snimljenog preko jednog kanala. Ova varijabilnost može da izazove približavanje fonetski i akustički različitih klasa, a samim tim i njihovo spajanje tokom TBC procedure. Da bi izbegli ovakve potencijalne greške u okviru ovog rada TBC procedura se primenjuje samo na opservacijama jednog govornika koji je sniman preko istog kanala. Pri TBC-u svako potencijalno stanje je opisano pomoću jedne Gausove raspodele, tako da je za svaku od njih potrebno obezbediti dovoljan broj opservacija za estimaciju parametara raspodele, odnosno potrebna je relativno velika, fonetski balansirana govorna baza. Gore navedene zahteve obično ispunjavaju govorne baza namenjene formiranju sistema za sintezu govora na osnovu teksta, te je za potrebe ovog rada iskorišćena jedna takva baza namenjena sintezi govora na srpskom jeziku, TTSlsSnezana (čiji je detaljan opis naveden u (Delić et al., 2013)).

Rezultat TBC procedure pokrenute na bazi koja sadrži snimke samo jednog govornika jeste skup potencijalnih stanja koji odgovaraju listovima stabla ($\mathcal{P}$) kojima se pridružuje matrica njihovih međusobnih rastojanja ($\mathbf{D}$). Kao mera rastojanja izabranja je simetrična KLD koja je definisana sa:

$$D_{KL_{sym}}(p\|q) = \frac{1}{2}\left(D_{KL}(p\|q) + D_{KL}(q\|p)\right) \tag{119}$$

gde su sa p i q označene gustine raspodela kojima su opisana potencijalna stanja. Pošto se pretpostavlja da je svako potencijalno stanje opisano pomoću Gausove raspodele simetrična KLD dobija sledeći oblik:

$$\begin{aligned} D_{KL_{sym}}(s_1\|s_2) \;=\; & \frac{1}{4}\Big((\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)^T\left(\boldsymbol{\Sigma}_2^{-1}+\boldsymbol{\Sigma}_1^{-1}\right)(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)+ \\ & + \operatorname{trag}\left(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2+\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1\right)-2D\Big) \end{aligned} \tag{120}$$

gde su $\boldsymbol{\mu}_i$ i $\boldsymbol{\Sigma}_i$ srednja vrednosti i kovarijansa Gausove raspodele koje su pridružene potencijalnim stanjima s_i (za $i = 1, 2$).

Pošto se govorna baza za formiranje potencijalnih stanja i govorna baza koja se koristi za obuku sistema razlikuju, pri TBC-u se favorizuje pravljenje što

3 U ovom kontekstu se pod spojenim stanjima smatraju ona stanja koja dele istu funkciju gustine raspodele emitovanja.

Parametri: Dat je skup svih potencijalnih HMM stanja $\mathcal{P}$, matrica rastojanja između stanja **D**, minimalni broj opservacija za građenje stanja n_{min}, maksimalni broj opservacija za građenje stanja n_{max} i maksimalna KLD K_{max}.

1: Inicijalizovati $\mathcal{C} = \emptyset$
2: **Za svako** potencijalno stanje p **iz** $\mathcal{P}$ **radi sledeće:**
3: $\mathcal{S}_p = p$
4: Odrediti n_p – broj opservacija koje pripadaju trifonima iz $\mathcal{S}_p$.
5: **Dok je** $n_p < n_{max}$ **radi sledeće:**
6: $\mathcal{S}_p = \mathcal{S}_p \cup s$ pri čemu $s = \arg\min_{q \in \mathcal{P} \setminus \mathcal{S}_p} D_{pq}$.
7: Odredi novo n_p
8: **Ako** $D_{ps} \geqslant K_{max}$ **onda**
9: **Ako** $n_{min} \leqslant n_p < n_{max}$ **onda**
10: $\mathcal{C} = \mathcal{C} \cup p$, pri čemu za obuku p treba korisiti sve instance koje su navedene u $\mathcal{S}_p$.
11: **Kraj ako**
12: Izađi iz petlje.
13: **Kraj ako**
14: **Kraj petlje**
15: **Ako** $n_p \geqslant n_{max}$ **onda**
16: $\mathcal{C} = \mathcal{C} \cup p$, pri čemu za obuku p treba korisiti sve instance koje su navedene u $\mathcal{S}_p$,
17: **Kraj ako**
18: **Kraj Za**
19: Ukloniti duplikate iz $\mathcal{C}$ – stanja koja se obučavaju na istim instancama.
20: **Rezultat:** $\mathcal{C}$.

Slika 19: Pseudo kod procedure kojom se formiraju stanja modela na osnovu inicijalnih modela koje je dala TBC na jednom govorniku i rastojanja između njih. Rezultujući skup HMM stanja je označen sa $\mathcal{C}$.

većeg broja listova odnosno potencijalnih stanja. Da bi se izbegla mogućnost da ovaj veliki broj potencijalnih stanja generiše nepotrebno veliki broj stanja konačnog modela, definišu se tri parametra: *i*) minimalni broj opservacija po stanju, *ii*) maksimalni broj opservacija po stanju i *iii*) maksimalna vrednost KLD-a po stanju. Na ovaj način se postiže kontrola rasipanja u okviru jednog stanja koje se treba izgraditi (preko maksimalne KLD), kao i nepotrebno povećanje rasipanja za stanja čije su opservacije značajno zastupljena u bazi (preko maksimalnog broja opservacija po stanju). Parametar kojim se ograničava minimalan broj opservacija se koristi za sprečavanje formiranja stanja kod kojih je rasipanje jednako zadatoj maksimalnoj KLD dok je broj opservacija nedovoljan za statistički efikasnu estimaciju parametara. Vrednosti za ova tri parametra se određuju heuristički i u ovom radu su iznosile 600, 4000 i 7 za minimalan broj opservacija po stanju, maksimalan broj opservacija po stanju i KLD respektivno.

Procedura kojom se na osnovu potencijalnih stanja ($\mathcal{P}$) koju daje TBC i rastojanja između potencijalnih stanja određuje skup stanja koji će se praviti ($\mathcal{C}$) je u formi pseudo koda prikazana na slici 19. Koje će od potencijalnih stanja biti izabrano kao HMM stanje modela prvenstveno zavisi od broja opserva-

cija u bazi za obuku koje su joj na raspolaganju. Ukoliko neka od klasa nema dovoljan broj opservacija onda joj se pridružuju opservacije iz klasa koje su joj bliske sve dok se ne sakupi dovoljan broj opservacija koje su potrebne za izgradnju modela, ili dok se ne prekorači maksimalno dozvoljeno rastojanje.

U cilju smanjenja varijabilnosti unutar jednog modela i pojednostavljenja načina obuke, umesto standardnih 30 fonema srpskog jezika modeluje se njihov prošireni skup koji je formiran:

- razlikovanjem naglašenih i nenaglašenih vokala,
- uvođenjem glasa šva,
- razdvajanjem okluzija i eksplozija kod ploziva i
- razdvajanjem okluzije i frikcije kod afrikata.

Podela na naglašene i nenaglašene vokale nije izvršena na osnovu lingvističke definicije, koja podrazumeva da je naglašen vokal onaj koji se nalazi u naglašenom slogu reči, već na osnovu energije i trajanja vokala i to ukoliko je vokal duži od 65 ms i ako mu je energija veća od 85% vrednosti prosečne energije tog vokala u bazi onda je naglašen u suprotnom je nenaglašen. Glas šva je uveden da bi se modelovao zvučni deo vibranta r kada se nalazi u kontekstu konsonanata, ali i neutralni vokal koji se pojavljuje prilikom izgovora pojedinačnih konsonanata. Razdvajanje okluzija i eksplozija kod ploziva, odnosno okluzija i frikcija kod afrikata je motivisana željom da se ubrza konvergencija inicijalnih modela. Da bi se izbeglo uvođenje novih termina i pojednostavilo izlaganje, u nastavku teksta izraz fonem će biti korišćen za elemente ovog proširenog skupa iako njegovi pojedini elementi to nisu. Lista fonema sa njihovim oznakama i kratkim opisima je data u tabelama 3 i 4.

Pored modela glasova postoje modeli za tišinu, buku i neartikulisane segmente. Model za buku ne opisuje sve tipove akustičkih smetnji već samo onih koje su impulsnog karaktera kao što su pucketanje, praskovi i sl., dok su preostale vrste smetnji opisane modelom tišine. Model neartikulisanih segmenata, kako što i samo ime kaže opisuje govorne segmente koji su neadekvatno artikulisani, odnosno izgovoreni fonem je do te mere oštećen da ga nije moguće identifikovati samo na osnovu akustičke informacije. Treba napomenuti da u labelama postoji i klasa oštećenih fonema, ali za razliku od neartikulisanih identitet oštećenog fonema se lako određuje. Opservacije koje pripadaju oštećenim fonemima se ne mogu koristiti za obuku odgovarajućih modela pošto predstavljaju atipične predstavnike klasa, ali se same labele mogu uzimati kao ispravan kontekst pošto je kod govornika postojala jasna namera da ih izgovori odnosno vokalni trakt se kretao u ispravnom smeru. Model za neartikulisane segmente se može lako uklopiti u bilo koji govorni segment stoga se ovaj model izostavlja u procesu prepoznavanja. Za modele tišine i buke se ne formiraju kontekstno zavisni modeli, pošto je cilj prepoznavanja govora odrediti šta je rečeno a ne tačne granice između fonema odnosno govornih i negovornih segmenata. Treba napomenuti da se informacija o početku govorne aktivnosti uzima u obzir pri formiranju trifona, pošto je tišina jedan od predviđenih konteksta.

Svi trifoni koji predstavljaju kontekstno zavisne varijante istog fonema imaju isti broj HMM stanja. Broj stanja po fonemu je srazmeran trajanju fonema u

Tabela 3: Lista konsonanata i negovornih modela

Oznaka fonema	Opis		Broj stanja
_	tišina	negovorni model	1
int	buka	negovorni model	2
unk	nedefinisano	nedovoljno artikulisan govor	1
F	fonem f	frikativ/labio-dentalni/bezvučni	4
H	fonem h	frikativ/velarni/bezvučni	4
S	fonem s	frikativ/dentalni/bezvučni	4
SH	fonem š	frikativ/alveolarni/bezvučni	4
V	fonem v	frikativ/labio-dentalni/zvučni	3
Z	fonem z	frikativ/dentalni/zvučni	4
ZH	fonem ž	frikativ/alveolarni/zvučni	4
L	fonem l	likvid/dentalni/zvučni	3
LJ	fonem lj	likvid/palatalni/zvučni	3
M	fonem m	nazal/bilibijalni/zvučni	4
N	fonem n	nazal/dentalni/zvučni	4
NJ	fonem nj	nazal/palatalni/zvučni	4
Bo	okluzija b	ploziv/bilibijalni/zvučni	3
Be	eksplozija b	ploziv/bilibijalni/zvučni	1
Po	okluzija p	ploziv/bilibijalni/bezvučni	3
Pe	eksplozija p	ploziv/bilibijalni/bezvučni	1
Do	okluzija d	ploziv/dentalni/zvučni	3
De	eksplozija d	ploziv/dentalni/zvučni	1
To	okluzija t	ploziv/dentalni/bezvučni	3
Te	eksplozija t	ploziv/dentalni/bezvučni	1
Go	okluzija g	ploziv/velarni/zvučni	3
Ge	eksplozija g	ploziv/velarni/zvučni	1
Ko	okluzija k	ploziv/velarni/bezvučni	3
Ke	eksplozija k	ploziv/velarni/bezvučni	1
CCo	okluzija ć	afrikat/postdentalni/bezvučni	2
CCe	frikativni deo ć	afrikat/postdentalni/bezvučni	2
DJo	okluzija đ	afrikat/postdentalni/zvučni	2
DJe	frikativni deo đ	afrikat/postdentalni/zvučni	2
CHo	okluzija č	afrikat/alveolarni/bezvučni	2
CHe	frikativni deo č	afrikat/alveolarni/bezvučni	2
DZo	okluzija dž	afrikat/alveolarni/zvučni	2
DZe	frikativni deo dž	afrikat/alveolarni/zvučni	2
Co	okluzija c	afrikat/dentalni/bezvučni	2
Ce	frikativni deo c	afrikat/dentalni/bezvučni	2
J	fonem j	poluvokal/palatalni/zvučni	3
R	šumni deo r	vibrant/alveolarni/bezvučni	2

Tabela 4: Lista vokala

Oznaka fonema	**Opis**		**Broj stanja**
A	fonem a	vokal/srednji/donji	3
As	fonem a - "naglašen"	vokal/srednji/donji	6
E	fonem e	vokal/prednji/srednji	3
Es	fonem e - "naglašen"	vokal/prednji/srednji	6
I	fonem i	vokal/prednji/gornji	3
Is	fonem i - "naglašen"	vokal/prednji/gornji	6
O	fonem o	vokal/zadnji/srednji	3
Os	fonem o - "naglašen"	vokal/zadnji/srednji	6
U	fonem u	vokal/zadnji/gornji	3
Us	fonem u - "naglašen"	vokal/zadnji/gornji	6
Y	fonem šva	vokal/srednji/srednji	2

bazi za obuku, a vrednosti se kreću od jednog HMM stanja za modele eksplozija kod ploziva do šest za modele naglašenih vokala. Broj stanja za svaki od fonema je naveden u tabelama 3 i 4.

Pošto se broj komponenata GMM-a određuje pomoću validacione procedure, čiji je detaljan opis dat u odeljku 4.3.1, broj komponenata varira od stanja do stanja. Zbog načina na koji se vrši povezivanje stanja (ograničenja rasipanja) prosečan broj komponenti se ne menja značajno u zavisnosti od pripadnosti fonemu. Sa druge strane broj komponenata zavisi od vrste izabrane aproksimacije GMM-a, što je ujedno i razlog zašto ni prosečan broj komponenata po stanju nije naveden u okviru tabela 3 i 4.

5.3.3 *Obuka sistema*

U ovom radu je analizirano nekoliko varijanata GMM-a, koje se razlikuju po parametrima modela, tako da se za estimaciju parametara modela koriste i različite procedure obuka. Iako svaka varijanta ima svoje specifičnosti veliki broj koraka je zajednički. Procedura obuke za GMM sa dijagonalnim i punim kovarijansnim matricama je praktično identična, a jedina razlika je način estimacije kovarijansne matrice gde u slučaju dijagonalne aproksimacije svi vandijagonalni elementi se postavljaju na nulu.[4] Pri obuci na raspolaganju su audio signali u parametrizovanom obliku kao i odgovarajuće granice između fonema koje su ručno pregledane i po potrebi korigovane, stoga prvi korak u obuci započinje ravnomernom preraspodelom opservacija trifona između njemu pripadajućih HMM stanja. Pri ovoj preraspodeli opservacija nije dozvoljeno narušavanje vremenskog sleda, odnosno opservacije koje se nalaze na početku sekvence opservacija pridružene jednoj instanci trifona ne mogu

4 Naravno u programskoj implementaciji računaju se samo dijagonalni elementi odnosno vektor varijansi umesto celokupne kovarijansne matrice.

da se dodele krajnjem HMM stanju umesto početnom i obrnuto. Nakon preraspodele opservacija po stanjima vrši se povezivanje stanja pomoću algoritma koji je opisan u prethodnom poglavlju (videti pseudo kod na slici 19). Kada se oforme odgovarajuća HMM stanja pristupa se određivanju broja komponenata mešavine pomoću krosvalidacije koja je opisana u odeljku 4.3.1, nakon koje sledi estimacija parametara modela koja obuhvata kako estimaciju parametara GMM tako i estimaciju histograma trajanja HMM stanja.[5] Iako se pri obuci estimira histogram trajanja HMM stanja pri dekodovanju u ovom radu su korišćene standardne verovatnoće prelaza koje se izračunavanju prilikom učitavanja modela na osnovu postojećih histograma. Estimacija parametara Gausovih raspodela se vrši pomoću modifikovanog k-means algoritma, koji se razlikuje od standardnog po načinu izračunavanja rastojanja opservacija do tekućih klastera (Janev et al., 2007), gde se pri računanju rastojanja opservacije do klastera u obzir uzima i rasipanje pojedinih obeležja u klasteru.[6] Nakon nekoliko iteracija k-means algoritma, kada parametri (srednje vrednosti i varijanse) iskonvergiraju realizuju se dva koraka EM-algoritma isključivo sa opservacijama koje su pridružene datom stanju. Prethodno opisani izbor jedinica modelovanja i njihovih sturktura omogućio je da gore opisani jednostavni algoritmi za estimaciju parametara HMM-GMM modela rezultuju zadovoljavajućim nivoom tačnosti ASR sistema. Prethodno opisana procedura je u celosti implementirana u okviru odgovarajućih funkcija csrlib biblioteke.

U odnosu na prethodno opisanu proceduru obuke, procedura obuke modela koji koriste (H)LDA podrazumeva dodatni korak u kojem se vrši estimacija transformacione matrice. Kao što je poznato, transformaciona matrica koja se dobija kao rezultat (H)LDA transformiše prostor obeležja tako da su nova (transformisana) obeležja dekorelisana, a rastojanja između različitih klasa maksimizovana pri čemu je rasipanje unutar klase ostalo nepromenjeno (Bishop, 2006), stoga se opravdano postavlja pitanje šta izabrati kao klase pri (H)LDA estimaciji. Ukoliko se kao (H)LDA klase izaberu klase koje se značajno poklapaju tada ni rezultujuća transformaciona matrica neće dati željene rezultate, jer takve klase i nije moguće razdvojiti. Pošto klase koje se koriste pri (H)LDA proceduri ne moraju da se poklapaju sa klasama koje se koriste pri prepoznavanju govora, za potrebe ovog rada su realizovani testovi sa različitim skupovima klasa, a više detalja o izabranim klasama će biti dato u poglavlju u kome su navedeni rezultati 6. Nakon estimacije transformacione matrice, postojeći vektori obeležja se jednostavnim množenjem sa estimiranom matricom transformišu u nove vektore obeležja, koji se potom koriste za estimaciju parametara modela. Estimacija parametara modela se na dalje vrši kao u slučaju prethodno opisanih GMM-a sa dijagonalnim kovarijansnim matricama, ali sa novim (transformisanim) vektorima obeležja. Procedura za estimaciju transformacione matrice

5 Standardni parametri kojim se opisuje sekvencijalni HMM jesu verovatnoće prelaza (verovatnoće ostanka u stanju i izlaska iz stanja), ali da bi se omogućilo preciznije modelovanje trajanja trifona odnosno reči prilikom dekodovanja koristi se histogram. Korišćenjem vrednosti koje su navedene u histogramu umesto verovatnoća prelaza pri dekodovanju se povećava broj logičkih stanja što za posledicu ima značajno sporiju proceduru dekodovanja.

6 U standardnom k-means algoritmu rastojanje opservacije do centroida (srednje vrednosti klastera) se izračunava kao euklidsko rastojanje, a u modifikovanoj varijanti kao negativna vrednost logaritma funkcije gustine raspodele uz pretpostavku da je raspodela Gausova sa uzoračkom srednjom vrednošću i varijansnom. Pošto su inicijalne vrednosti kovarijansi Gausovih raspodela jedinične matrice prva iteracija je identična u obe verzije algoritma.

u slučaju LDA varijante realizovana je u okviru odgovarajućih funkcija csrlib biblioteke, a svodi se na jednostavno rešavanje sistema linearnih jednačina. Sa druge strane, za HLDA varijantu ne postoji rešenje u zatvorenoj formi, te je za estimaciju transformacione matrice neophodno primeniti optimizacionu proceduru. Pošto programski paket Matlab sadrži veliki broj gotovih algoritama, on se nametnuo kao znatno pogodnije rešenje za implementaciju procedure estimacije HLDA transformacione matrice. Kao što je već ranije napomenuto u ovom radu je izabrana procedura koja je opisana u (Kumar, 1997).

Druga varijanta koja je analizirana u ovom radu jeste STC varijanta GMM modela, gde više HMM stanja deli istu transformacionu matricu. Za ove potrebe u okviru csrlib biblioteke implementirana je delimično modifikovana varijanta procedure koja je opisana u (Gales, 1999), a koja podrazumeva iterativno rešavanje sistema linearnih jednačina. Modifikacija podrazumeva povećanje broja transformacionih matrica dok se ne postigne željena greška aproksimacije izražena preko KLD-a, dok je sama procedura estimacije kolona transformacione matrice nepromenjena. Procedura obuke se u velikoj meri poklapa sa procedurom obuke GMM-a sa punim kovarijansnim matricama, pri čemu se nakon estimacije punih kovarijansnih matrica vrši estimacija transformacione matrice za grupe stanja. Nakon što se pronađe odgovarajuća transformaciona matrica za neku grupu stanja g, Gausovim raspodelama koje opisuju tu grupu stanja modifikuju se srednje vrednosti i kovarijansne matrice na sledeći način:

$$\boldsymbol{\mu}_{tsm} = \mathbf{D}^{(g)\mathsf{T}}\boldsymbol{\mu}_{sm} \tag{121}$$

$$\boldsymbol{\Sigma}^{(diag)}_{tsm} = \mathbf{D}^{(g)\mathsf{T}}\boldsymbol{\Sigma}_{sm}\mathbf{D}^{(g)} \tag{122}$$

gde je $\boldsymbol{\mu}_{tsm}$ srednja vrednost u transformisanom skupu obeležja koja odgovara srednjoj vrednosti m-te komponente GMM-a stanja s, $\boldsymbol{\Sigma}^{(diag)}_{tsm}$ dijagonalna kovarijansna matrica u transformisanom prostoru obeležja koja odgovara originalnoj kovarijansnoj matrici $\boldsymbol{\Sigma}_{sm}$ m-te komponente mešavine stanja s, a $\mathbf{D}^{(g)}$ odgovarajuća transformaciona matrica. Proizvod $\mathbf{D}^{(g)\mathsf{T}}\boldsymbol{\Sigma}_{sm}\mathbf{D}^{(g)}$ obično nije dijagonalna matrica, tako da je potrebno sve vandijagonalne elemente postaviti na nulu, što se praktično svodi da se matrica $\boldsymbol{\Sigma}^{(diag)}_{tsm}$ čuva kao vektor koji sadrži samo elemente na glavnoj dijagonali.

Kao i procedura obuke u slučaju STC varijante GMM-a procedura u varijanti SEGMM nastavlja se na standardnu proceduru obuke modela sa punim kovarijansnim matricama, s tom razlikom da se posle estimacije punih kovarijansnih matrica vrši njihova dekompozicija na karkateristične vrednosti i vektore. Nakon što se estimiraju ovi karakteristični vektori traži se rečnik i odgovarajuća retka reprezentacija za svaki od njih, za šta je u ovom radu korišćen algoritam određivanja rečnika u letu koji je implementiran u SPAMS optimizacionom paketu (Mairal et al., 2009). Za potrebe ovog rada, veličina podskupa (*batch*) koji se uzima u jednoj iteraciji je 30000, a broj iteracija je 3000, što je ekvivalentno varijanti da je uzet celokupan skup karakterističnih vektora i da je vršeno oko 240 iteracija. Za l_2-norme normu ciljanog odstupanje aproksimirane vrednosti od stvarne izabrana je vrednost od 10^{-3} dok je veličina rečnika i kardinalnost retke reprezentacija varirana u skladu sa potrebama eksperimenta (videti poglavlje 6). Pronalaženjem parametara retke reprezentacije (rečnika i retkih kodova), modifikovanjem težina mešavina i njihovih srednjih vrednosti po uzoru na jednačine (83) i (84) se završava procedura obuke sistema.

Treba primetiti da su prethodno opisane procedure obuka eliminisale mogućnost varijacija u procenama prerformansi sistema koje bi bile posledica različitih modela, različitih stanja ili različitih poravnanja, odnosno da je jedina razlika koja postoji između ispitivanih sistema izabrana struktura GMM-a.

5.3.4 *Dekodovanje*

Dekodovanje (prepoznavanje) podrazumeva traženje sekvence reči koja je najverovatnije generisala zadatu sekvencu opservacija. U statističkom modelu zasnovanom na HMM-u, reči su modelovane kao sekvence HMM stanja koja su pridružena odgovarajućim fonemima zavisnim od konteksta, tako da se problem prepoznavanja svodi na traženje optimalne putanje kroz trelis koji obrazuju HMM stanja. Struktura trelisa se definiše zadavanjem gramatike[7] i rečnika izgovora.[8] U ovom radu je izabrana tzv. gramatika nezavisna od konteksta u kojoj je dozvoljen prelazak iz svake reči u svaku direktno ili preko tišine i/ili buke. Broj reči u test skupu je oko 150, pri čemu za pojedine reči postoji po nekoliko varijanata izgovora, tako da je stvaran broj različitih putanja nešto veći (oko 200). Izabrana je ovakva struktura pošto su svi testovi imali za cilj procenu kvaliteta akustičkih modela sistema za prepoznavanje, a ne stvarnih mogućnosti sistema koje bi se dobile uvođenjem odgovarajućeg modela jezika i odgovarajuće restriktivne gramatike.

Postoji nekoliko algoritama koji se koriste za traženje optimalne putanje kroz trelis, a predstavljaju varijacije Viterbijevog ili A-star stek (*A* stack*) algoritma (Huang et al., 2001). U okviru ovog rada korišćen je Viterbijev algoritam koji je implementiran u okviru odgovarajućih funkcija csrlib biblioteke. Kao što je poznato Viterbijev algoritam se sastoji iz dva ključna koraka: *i*) propagacije u napred pri kojoj se vrši izračunavanje izglednosti pojedinih hipoteza i odbacivanje manje izglednih i *ii*) propagacije u nazad u kojoj se vrši određivanje najizglednije sekvence stanja (dekodovanje). Dekodovanje se vrši ako je primljena celokupna sekvenca opservacija koju je potrebno prepoznati ili ako je primljeno V opservacija. Potreba za dekodovanjem nakon primljenih V opservacija je posledica memorijskih ograničenja koje postavlja hardver i ovom radu on iznosi 1000 što odgovara segmentu signala trajanja 10 s. U slučaju da je primljena celokupna sekvenca opservacija, dekodovanje započinje iz HMM stanja koje ima najveću akumulisanu izglednost i koje je završno stanje u reči,[9] buci ili tišini. Sa druge strane ako je primljeno V opservacija umesto celokupne sekvence opservacija, povratak (ali ne i samo dekodovanje) započinje iz HMM stanja koje ima najveću moguću izglednost, dok dekodovanje započinje nakon D koraka unazad, odnosno samo za preostalih $V - D$ opservacija se određuje odgovarajuća sekvenca stanja. Nakon toga, propagacija u napred se nastavlja iz stanja koje je bilo u pobedničkoj sekvenci na D koraka od kraja, da bi se obezbedila naprekidnost dekodovane sekvence na nivou celokupnog audio fajla koji je potrebno prepoznati. U ovom radu D iznosi 100 opservacija što odgovara segmentu signala trajanja 1 s.

7 Gramatika je skup pravila koja definišu moguće prelaze između reči.
8 Rečnik izgovora je skupa pravila koji definišu preslikavanje reči u sekvence odgovarajućih modela.
9 Završno stanje u reči je završno stanje poslednjeg trifona u nekoj reči.

Da bi se uskladili doprinosi verovatnoća prelaza između stanja i vrednosti gustina raspodela emitujućih verovatnoća stanja uveden je težinski faktor (Huang et al., 2001) kojim se množe logaritmi verovatnoća prelaza i u ovom radu on iznosi 5. Sam matematički okvir na kojem se zasniva procedura dekodovanja ne podrazumeva ponderisanje pojedinačnih faktora koji ga čine (verovatnoće prelaza i emitovanja), ali dodavanjem različitih težina pojedinim faktorima se smanjuje broj grešaka sistema, stoga se ovaj faktor određuje heuristički.

Korišćeni softver omogućava ubrzavanje procesa prepoznavanja odbacivanjem HMM stanja koja su dobila malu vrednost akumulisane izglednosti primenom tzv. potkresivanja (pruning), ali pošto je prvenstveni cilj ovih testova bilo ispitivanje kvaliteta akustičkih modela ova opcija nije korišćena. Na ovakav način se izbegavaju greške koje unosi sam algoritam dekodovanja koje su posledica odbacivanja ispravnih sekvenci u toku dekodovanja. Pri korišćenju sistema u praktičnim aplikacijama da bi se obezbedio rad u realnom vremenu neophodno je vršiti odsecanje manjeverovatnih putanja. Parametri koji definišu odsecanje, maksimalan broj aktivnih stanja kao i opseg vrednosti akumulisanih izglednosti (od maksimalne do minimalne) se određuje heuristički, tako da se postigne brzina prepoznavanja što bliža ciljnoj uz što manju degradaciju tačnosti sistema za prepoznavanje.

Za različite varijante GMM-a, pošto se razlikuju po parametrima kojima su opisani, bilo je potrebno obezbediti posebne funkcije za izračunavanje vrednosti gustina raspodele verovatnoće emitovanja HMM stanja. Za slučaj GMM-a sa punim kovarijansnim matricama ova vrednost se izračunava na osnovu izraza (10) dok se u slučaju GMM-a sa dijagonalnim kovarijansnim matricama svodi na:

$$b_s(\mathbf{o}) = \sum_{m=1}^{M_s} w_{sm} \frac{1}{\sqrt{(2\pi)^D \prod_{k=1}^{D} \Sigma_{smkk}}} e^{-\frac{1}{2}\sum_{k=1}^{D} \frac{(o_k - \mu_{smk})^2}{\Sigma_{smkk}}} \tag{123}$$

gde je sa D označena dimenzionalnost prostora obeležja, o_k vrednost k-tog obeležja opservacije $\mathbf{o}$, μ_{smk} srednja vrednost za k-to obeležje m-te komponente stanja s, Σ_{smkk} varijansa za k-to obeležje m-te komponente stanja s.

U slučaju (H)LDA, transformacionu matricu dele sva stanja tako da ista matrica množi sve vektore obeležja te se ova operacija može izmestiti u blok za estimaciju obeležja. Formalno gledano množenje vektora obeležja transformacionom matricom (odnosno $\mathbf{y} = \mathbf{D}\mathbf{o}$) predstavlja linearnu transformaciju slučajnih promenljivih tako da je vrednost gustine raspodele emitovanja stanja s data izrazom:

$$b_s(\mathbf{o}) = \sum_{m=1}^{M_s} w_{ysm} \frac{\mathrm{abs}|\mathbf{D}|}{\sqrt{(2\pi)^D \prod_{k=1}^{D_y} \Sigma_{ysmkk}}} e^{-\frac{1}{2}\sum_{k=1}^{D_y} \frac{(y_k - \mu_{ysmk})^2}{\Sigma_{ysmkk}}} \tag{124}$$

gde su svi parametri kao ranije, a oznaka y u indeksu parametra ima za cilj da ukaže da su parametri estimirani u transformisanom prostoru obeležja. Pošto apsolutna vrednost determinante transformacione matrice množi sve emitujuće verodostojnosti bez obzira na stanje, relativni odnos akumuliranih verodostojnosti za pojedine hipoteze ostaje nepromenjen pa tako i rezultat dekodovanja, stoga ga je moguće izostaviti, što je u ovoj implementaciji i učinjeno.

Izraz za izračunavanje vrednosti gustine raspodele emitovanja stanja u slučaju STC se svodi na (124), s tom razlikom da transformacione matrice i trans-

formisani vektori obeležja nisu isti za sva stanja već za grupe stanja, odnosno:

$$b_s(\mathbf{o}) = \sum_{m=1}^{M_s} w_{ysm} \frac{\text{abs}|\mathbf{D}^{(g)}|}{\sqrt{(2\pi)^D \prod_{k=1}^{D_y^{(g)}} \Sigma_{ysmkk}^{(g)}}} e^{-\frac{1}{2}\sum_{k=1}^{D_y^{(g)}} \frac{(y_k^{(g)} - \mu_{ysmk}^{(g)})^2}{\Sigma_{ysmkk}^{(g)}}} \tag{125}$$

gde oznaka g u eksponentu promenljivih treba da ukaže da se vrednosti transformacionih parametara menjaju u zavisnosti od grupe stanja, dok su značenja ostalih oznaka nepromenjena. Pošto transformaciona matrica nije ista za sva stanja, nije moguće izostaviti množenje sa $\text{abs}|\mathbf{D}^{(g)}|$ osim ukoliko se ne uvede ograničenje da je apsolutna vrednost determinante transformacione matrice jednaka 1. Da bi se optimizovao broj potrebnih računskih operacija, množenje odgovarajućim transformacionim matricama se vrši samo za u tom trenutku aktivna stanja, stoga transformaciju obeležja nije moguće preneti u blok za estimaciju obeležja.

Izračunavanje vrednosti gustine raspodele emitovanja stanja za SEGMM varijantu GMM-a je definisan jednačinama (82) i (88) i detaljno je opisano u odeljku 4, te ovde neće biti posebno ponovo navođeno.

Celokupni dekoder je realizovan u okviru odgovarajućih funkcija csrlib biblioteke. Potrebne modifikacije dekodera za (H)LDA, STC i SEGMM je samostalno realizovao autor ovog rada. Iako implementirani dekoder podržava i osnovu varijantu (datu izrazima (10), (82), (88) i (123–125)) i brzu varijantu u kojoj operator suma zamenjena operatorom max, pri čemu je u testovima korišćena isključivo druga varijanta jer je brža i u ranijim eksperimentima nije bilo razlike u prepoznatim sekvencama reči.

5.4 REZIME

U ovom poglavju je dat pregled resursa (govornih baza i softverskih alata) koji su korišćeni za potrebe izrade ovog rada u cilju obezbeđenja ponovljivosti rezultata. Iako su opisani algoritmi nezavisni od jezika, svi testovi su realizovani isključivo na srpskim govornim bazama zbog nedostupnosti resursa za druge jezike usled finansijskih ograničenja. Osnovni skup alata koji je korišćen za realizaciju eksperimenata u okviru ovog rada je rezultat višegodišnjeg razvoja sistema za prepoznavanje govora na srpskom jeziku na Fakultetu tehničkih nauka. Za složene optimizacione procedure iskorišćeni su javno dostupni gotovi Matlabovi alati (Kumar, 1997; Mairal et al., 2010). Sve modifikacije osnovnih procedura i funkcija potrebnih za implementaciju (H)LDA, STC i SEGMM je autor ovog rada samostalno realizovao. Izbor modela, njihova struktura kao i načini za prevazilaženje problema usled malog broja opservacija po modelu nisu posebno obrazloženi jer predstavljaju rezultat višegodišnjih istraživanja na ovom polju realizovanih na Fakultetu tehničkih nauka.

6 REZULTATI

6.1 UVOD

Ovaj odeljak daje sveobuhvatan pregled realizovnih eksperimenata koji su imali za cilj da uporede pojedine varijante GMM-a sa aspekta tačnosti prepoznavanja izraženog preko učestanosti grešaka na nivou reči (WER word error rate) i broja parametara koji su potrebni za opisivanje modela. Treba naglasiti da svi analizirani sistemi imaju isti skup HMM stanja i da je poravnanje opservacija i stanja uvek isto bez obzira na to koja je varijanta vektora obeležja ili GMM-a u pitanju (više detalja o načinu kako je to realizovano dato je u poglavlju 5).

Kao što je napomenuto u poglavlju 2, načešće korišćena varijanta GMM-a jeste GMM sa dijagonalnim kovarijansnim matricama, zbog svoje male računske složenosti prilikom izračunavanja verovatnoća emitovanja i robustne estimacije parametara u slučaju malog broja opservacija za obuku. Pošto ovaj model predstavlja polaznu tačku u većini sistema za prepoznavanje oblika zasnovanim na statističkom pristupu, on se nameće kao logičan izbor za referentni model. Ovaj referentni model je realizovan u dve varijante u zavisnosti od obeležja koja se koriste. Prva varijanta predstavlja standardno korišćen skup obeležja za prepoznavanje govora koji podrazumeva: 12 MFCC-ova, normalizovanu energiju i njihove prve i druge izvode (u nastavku ovaj skup obeležja nosiće oznaku 12MFCC_E_D_A[1]). U ranijim eksperimentima (Janev et al., 2007; Delić et al., 2010) na govornoj bazi koja je korišćena i u okviru ovog rada nešto bolja tačnost prepoznavanja za istu složenost modela je postignuta ukoliko se koriste obeležja koja obuhvataju: 12 MFCC-ova, normalizovanu energiju i samo njihove prve izvode (12MFCC_E_D), što se može objasniti relativno skromnom veličinom baze tako da delta-delta obeležja unose više šuma nego korisinih informacija. Slično ponašanje je dobijeno i u eksperimentima koji su realizovani za potrebe ovog rada gde je broj Gausiana po mešavini određen na osnovu validacionog skupa (videti rezultate u tabeli 5).

Tabela 5: Performanse referentnih modela (broj Gausovih raspodela i učestanost grešaka na nivou reči)

Varijanta	Obeležja	# Gausiana	# Parametara	WER[%]
Diag.	12MFCC_E_D	52940	2.81M	4.04
Diag.	12MFCC_E_D_A	61460	4.86M	4.73
Pune	12MFCC_E_D	14560	5.50M	2.16
Pune	12MFCC_E_D_A	13990	11.47M	2.25

Pored GMM-a sa dijagonalnim kovarijansnim matricama razumno je kao referentni model uzeti i varijantu GMM-a sa punim kovarijansnim matricama,

1 Preuzete su oznake koje se koriste u HTK, gde E označava energiju, D delta obeležja i A (akceleracijska) delta-delta obeležja.

pošto je to varijanta koja najpreciznije modeluje korelacije koje postoje između pojedinih obeležja. Kao što je već navedeno u poglavlju 2 da bi model sa punim kovarijansnim matricama bio uspešan neophodno je obezbediti dovoljan broj opservacija za statističku efikasnu estimaciju parametara, što je u ovom radu i urađeno postavljanjem minimalnog broja opservacija po Gausovoj raspodeli na 350. I ova varijanta GMM-a je realizovana na dva skupa obeležja (12MFCC_E_D_A i 12MFCC_E_D) pri čemu se sa aspekta tačnosti prepoznavanja i ukupnog broja parametara boljom pokazala varijanta sa 12MFCC_E_D (videti tabelu 5).

U tabeli 5 je dat uporedni prikaz performansi prethodno pomenuta 4 modela. Kao što se iz priloženog može videti daleko manji broj grešaka (skoro duplo manje) prave sistemi sa punim kovarijansnim matricama, ali je i broj parametara koje je potrebno estimirati skoro duplo veći za isti skup obeležja,[2] iako je broj Gausovih raspodela nekoliko puta manji u slučaju punih kovarijansnih matrica. Kao što je napomenuto u poglavlju 2, alternativni modeli GMM-a treba da obezbede memorijsku složenost i tačnost prepoznavanja koja je između one koje imaju GMM-i sa dijagonalnim i punim kovarijansnim matricama.

U nastavku ovog rada biće izloženi rezultati koji su dobijeni za SEGMM, ali i za nekoliko drugih varijanata GMM-a koje su opisane u poglavlju 2 kao što su (H)LDA i STC. Deo ovde navedenih rezultata je već objavljen u (Jakovljević et al., 2012, 2013), a deo je u procesu evaluacije za publikovanje u časopisu i na konferenciji.

6.2 (H)LDA

Kao što je već napomenuto u poglavlju 2, motivacija za uvođenje (H)LDA je bila nešto drugačija, odnosno cilj je bio pronaći linearnu transformaciju (transformacionu matricu) koja prostor obeležja preslikava u novi prostor obeležja u kome je moguće efikasnije razdvojiti klase i u kome su obeležja nekorelisana.[3] Pošto (H)LDA pruža mogućnost redukcije prostora obeležja, u okviru ovog rada analizirane su varijante sa različitim ulaznim obeležjima koje su navedene u literaturi (Haeb-Umbach i Ney, 1992; Kumar i Andreou, 1998; Westphal, 2004; Morgan, 2012) kao što su: 12MFCC_E_D_A i vektori koji se dobijaju konkatenacijom nekoliko (od 3 do 11) uzastopnih vektora obeležja koji sadrže 12MFCC_E.[4] Za (H)LDA klase izabrana su HMM stanja trifona pošto u literaturi oni predstavljaju najčešći izbor, a u pilot testovima su dali i najbolje rezultate (Jakovljević et al., 2012).

2 Za izračunavanje broja parametara na osnovu broja Gausovih raspodela iskorišćeni su izrazi dati u tabeli 1.

3 U slučaju LDA navedene pretpostavke o prostoru transformisanih obeležja su direktno vidljive iz same ciljne funkcije. Sa druge strane u slučaju HLDA, to nije vidljivo iz ciljne funkcije koja podrazumeva maksimizaciju izglednosti, ali pošto se varijanta HLDA koja podrazumeva nekorelisanost obeležja uvođenjem dodatne pretpostavke da sve klase imaju istu matricu rasipanja svodi na LDA, može se zaključiti da je efekat koji se postiže skoro identičan (Kumar i Andreou, 1998).

4 Ideja za uvođenjem konkatenativnih obeležja polazi od pretpostavke da su delta i delta-delta obeležja samo jedan vid aproksimacije vrednosti obeležja u neposrednoj okolini trenutka posmatranja te da originalna statička obeležja nose više informacija.

U originalnom LDA algoritmu (Bishop, 2006) kolone transformacione matrice se dobijaju kao karakteristični vektori matrice $W^{-1}T$, gde je T matrica ukupnog rasipanja, a W matrica prosečnog rasipanja unutar klase. Matrice rasipanja su definisane sa:

$$T = \frac{1}{N}\sum_{i=1}^{N}(o_i - \mu)(o_i - \mu)^T \quad (126)$$

$$W = \frac{1}{N}\sum_{c}\sum_{i\in c}(o_i - \mu_c)(o_i - \mu_c)^T \quad (127)$$

gde je o_i i-ta opservacija, μ srednja vrednost svih opservacija, μ_c srednja vrednost opservacija koje pripadaju klasi c i N ukupan broj opservacija. Pošto iterativni algoritam za HLDA (Kumar i Andreou, 1998) predviđa da inicijalna transformaciona matrica bude ona koja se dobija pomoću LDA ali normalizovana tako da joj determinanta bude jednaka 1, da bi se mogli uporediti rezultati koje daje LDA i HLDA realizovani su i eksperimenti sa normalizovanom varijantom transformacione matrice. Na dalje u tekstu ove dve varijante LDA će nositi oznaku nenormalizovana i normalizovana.

U tabeli 6 su prikazane performanse sistema baziranih na nenormalizovanoj (originalnoj) varijanti LDA, za različita ulazna obeležja i za različit broj izlaznih obeležja. Kao što se iz priloženog može videti performanse modela koji koriste LDA su daleko bliže performansama GMM-a sa dijagonalnim kovarijansnim matricama nego punim, što je donekle bilo za očekivati jer je malo verovatno da je moguće dekorelisati obeležja u svim klasama korišćenjem jedne transformacione matrice. Ukoliko se posmatra tačnost sistema, može se uočiti da ona u velikoj meri zavisi od vrste ulaznih obeležja, pa tako sistemi koji kao ulazna obeležja koriste konkatenativna obeležja nastala spajanjem od 3 do 5 uzastopnih opservacija nisu dala nikakvo poboljšanje u odnosu na referentne modele sa dijagonalnim kovarijansnim matricama za razliku od npr. sistema koji koriste konkatenativna obeležja nastala spajanjem 7 uzastopnih opservacija. Na osnovu priloženog se može videti da korišćenje 3 uzastopne opservacije koje sadrže samo statička obeležja (12MFCC_E) dovodi do gubitka bitnih diskriminativnih informacija u odnosu na druge varijante konkatenativnih obeležja kao i GMM-a sa dijagonalnim kovarijansnim matricama čija su obeležja 12MFCC_E_D i 12MFCC_E_D_A. Pored toga konstantan pad WER sa padom dimenzionalnosti prostora izlaznih obeležja je posledica činjenice da su obeležja u 3 uzastopne opservacije u značajnoj meri korelisana. Ono što donekle iznenađuje jeste da sistem koji koristi 5 uzastopnih frejmova (što je ujedno i broj frejmova koji se uzima u obzir prilikom računanja delta obeležja u varijanti 12MFCC_E_D) u slučaju kada je broj izlaznih obeležja 26 odnosno 39 ima manju tačnost prepoznavanja od odgovarajućih referentnih GMM-a sa dijagonalnim kovarijansnim matricama, odakle sledi da su mogućnosti LDA za dekorelaciju obeležja prilično skromne (lošije nego da se dekorelacija susednih frejmova vrši diferenciranjem obeležja pomoću regresionog obrasca). Najveća tačnost je postignuta u varijanti u kojoj se ulazna obeležja formiraju od 7 uzastopnih opservacija i ukoliko je broj izlaznih obeležja 35 odnosno 32. Dalje povećanje dimenzionalnosti ulaznog vektora dodavanjem novih susednih opservacija dovodi do degradacije tačnosti prepoznavanja. Interesantno je da ovakav trend ne postoji u varijanti sa normalizovanim transformacionim

Tabela 6: Performanse nenormalizovanih LDA sistema za različita ulazna obeležja i različit broj izlaznih (transformisanih) obeležja.

Ulazni vektor	# Izlaznih obeležja	# Gausiana	# Parametara	WER [%]
3×12MFCC_E	39	57260	4.53M	7.17
3×12MFCC_E	35	57119	4.06M	6.24
3×12MFCC_E	32	57135	3.71M	5.55
3×12MFCC_E	26	54463	2.89M	5.11
5×12MFCC_E	39	60009	4.74M	5.12
5×12MFCC_E	35	59952	4.26M	4.39
5×12MFCC_E	32	59228	3.85M	4.22
5×12MFCC_E	26	57048	3.03M	4.35
7×12MFCC_E	39	60922	4.82M	4.41
7×12MFCC_E	35	60586	4.30M	3.76
7×12MFCC_E	32	60154	3.91M	3.78
7×12MFCC_E	26	58723	3.11M	4.14
9×12MFCC_E	39	61153	4.83M	4.62
9×12MFCC_E	35	60840	4.32M	3.88
9×12MFCC_E	32	60669	3.95M	4.10
9×12MFCC_E	26	59434	3.15M	4.15
11×12MFCC_E	39	61283	4.85M	5.23
11×12MFCC_E	35	61185	4.35M	3.97
11×12MFCC_E	32	61047	3.97M	3.98
11×12MFCC_E	26	59818	3.17M	4.25
12MFCC_E_D_A	39	57260	4.55M	4.26
12MFCC_E_D_A	35	57119	4.04M	3.95
12MFCC_E_D_A	32	57135	3.65M	3.63
12MFCC_E_D_A	26	54463	2.86M	4.22

matricama (videti tabelu 7) iako skaliranje kolona matrice koje se radi pri normaliziciji transformacione matrice ne menja ciljnu funkciju (Bishop, 2006). Ove varijacije se mogu objasniti drugačijom preraspodelom opservacija između inicijalnih klastera odnosno komponenti Gausovih mešavina, jer pošto su obeležja drugačija razumno je očekivati da su i rastojanja između opservacija drugačija. Tačnosti prepoznavanja sistema koji koriste vektore obeležja koji nastaju spajanjem uzastopnih 9 i 11 opservacija su slične, što znači da dodatne 2 opservacije ne sadrže bitne diskriminativne informacije. Bez obzira na varijantu ulaznih obeležja najlošija tačnost prepoznavanja se postiže ukoliko je dimenzionalnost izlaznog prostora 39, što dodatno potvrđuje ranije iznesenu tezu (kod referentnih sistema) da je broj opservacija koje su na raspolaganju suviše mali da bi na zadovoljavajući način popunio 39-dimenzionalni prostor obeležja. Izbacivanje manje informativnih dimenzija (kojima odgovara manja karakteristična vrednost) dodatno povećava tačnost prepoznavanja jer se uklanja i šum koji ta

Tabela 7: Performanse normalizovanih LDA sistema za različita ulazna obeležja i različit broj izlaznih (transformisanih) obeležja.

Ulazni vektor	# Izlaznih obeležja	# Gausiana	# Parametara	WER [%]
3×12MFCC_E	39	56841	4.49M	7.25
3×12MFCC_E	35	56899	4.04M	6.14
3×12MFCC_E	32	56649	3.68M	5.57
3×12MFCC_E	26	54507	2.89M	5.72
5×12MFCC_E	39	60441	4.78M	5.32
5×12MFCC_E	35	59854	4.25M	4.75
5×12MFCC_E	32	59268	3.85M	4.32
5×12MFCC_E	26	56986	3.02M	4.30
7×12MFCC_E	39	60704	4.80M	4.38
7×12MFCC_E	35	60594	4.31M	3.91
7×12MFCC_E	32	60093	3.91M	3.90
7×12MFCC_E	26	58888	3.12M	3.93
9×12MFCC_E	39	61097	4.83M	4.45
9×12MFCC_E	35	60789	4.32M	3.83
9×12MFCC_E	32	60633	3.94M	3.83
9×12MFCC_E	26	59308	3.15M	4.10
11×12MFCC_E	39	61330	4.85M	4.87
11×12MFCC_E	35	61128	4.34M	3.80
11×12MFCC_E	32	61032	3.97M	3.83
11×12MFCC_E	26	59750	3.17M	4.20
12MFCC_E_D_A	39	57163	4.52M	4.03
12MFCC_E_D_A	35	56623	4.02M	3.81
12MFCC_E_D_A	32	55774	3.63M	4.04
12MFCC_E_D_A	26	52773	2.80M	4.11

obeležja unose. Na osnovu rezultata priloženih u tabeli 6 može se zaključiti da je broj obeležja koji nosi diskriminativne informacije između 32 i 35 i to ne zavisi od broja ulaznih obeležja koji varira od 39 do 143.

Interesantno je da varijante LDA koje kao ulazna obeležja koriste standardna obeležja (12MFCC_E_D_A) imaju manji WER od referentnog dijagonalnog modela sa istim ulaznim obeležjima. U ovom slučaju jedna ulazna opservacija nosi informaciju o kontekstu širine 11 frejmova pošto sadrži delta-delta koeficijente što ukazuje da degradacija performansi do koje dolazi kada se povećava broj sukcesivnih opservacija nije posledica širine konteksta već dimenzionalnosti ulaznog vektora, odnosno loše estimacije pojedinačnih matrica rasipanja. Treba primetiti da ni jedna varijanta LDA kod koje je broj izlaznih obeležja 26 nije dala manji WER od referentnog dijagonalnog modela koji koristi 26-dimenzionalne vektore 12MFCC_E_D. Pošto postoje sistemi bazirani na LDA koji imaju manji WER nego prethodno pomenuti referentni sistem, može se

Tabela 8: Performanse HLDA sistema uz pretpostavku da su samo diskriminativna obeležja nekorelisana za različita ulazna obeležja i različit broj izlaznih (transformisanih) obeležja.

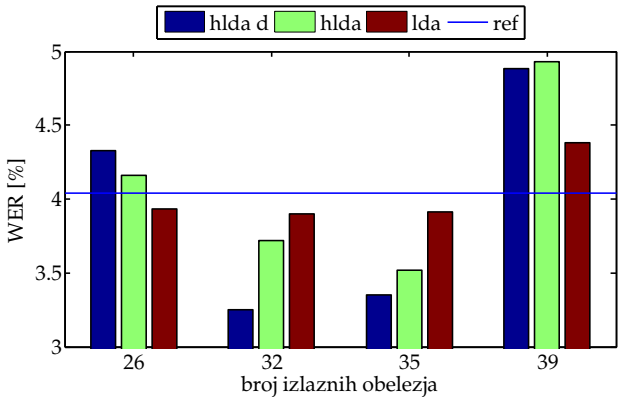
Ulazni vektor	# Izlaznih obeležja	# Gausiana	# Parametara	WER [%]
7×12MFCC_E	39	59025	4.67M	4.88
7×12MFCC_E	35	58175	4.13M	3.35
7×12MFCC_E	32	56784	3.69M	3.25
7×12MFCC_E	26	53356	2.83M	4.33
9×12MFCC_E	39	59751	4.72M	3.85
9×12MFCC_E	35	58492	4.16M	3.63
9×12MFCC_E	32	58860	3.83M	3.89
9×12MFCC_E	26	56076	2.97M	4.28
12MFCC_E_D_A	39	55082	4.35M	3.95
12MFCC_E_D_A	35	53648	3.81M	3.67
12MFCC_E_D_A	32	51778	3.37M	3.41
12MFCC_E_D_A	26	49166	2.61M	3.85

zaključiti da transformisana (izlazna) obeležja sadrže više diskriminativnih karakteristika, ali da prvih 26 sa njavećim karakterističnim vrednostima nisu dovoljne za diskriminaciju.

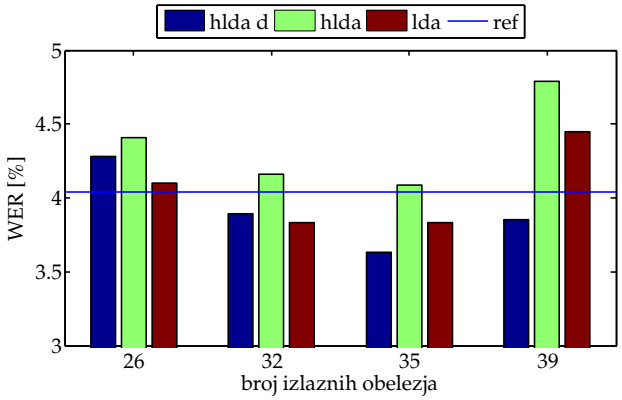
Slično ponašanje koje postoji kod nenormalizovane (originalne) LDA se uočava i za normalizovanu varijantu LDA, čije su performanse navedene u 7.

Tabela 9: Performanse HLDA sistema uz pretpostavku da su samo diskriminativna obeležja nekorelisana za različita ulazna obeležja i različit broj izlaznih (transformisanih) obeležja.

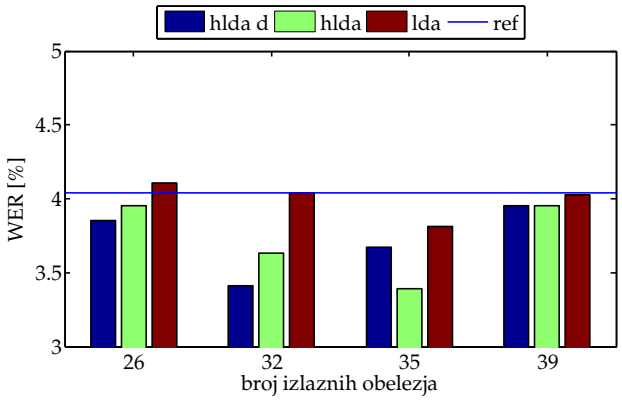
Ulazni vektor	# Izlaznih obeležja	# Gausiana	# Parametara	WER [%]
7×12MFCC_E	39	58381	4.62M	4.93
7×12MFCC_E	35	58092	4.13M	3.52
7×12MFCC_E	32	56651	3.69M	3.72
7×12MFCC_E	26	53356	2.83M	4.16
9×12MFCC_E	39	61439	4.86M	4.79
9×12MFCC_E	35	60438	4.30M	4.09
9×12MFCC_E	32	57873	3.77M	4.16
9×12MFCC_E	26	58459	3.10M	4.41
12MFCC_E_D_A	39	55082	4.35M	3.95
12MFCC_E_D_A	35	53805	3.82M	3.39
12MFCC_E_D_A	32	52138	3.39M	3.63
12MFCC_E_D_A	26	48790	2.59M	3.95



(a) Ulazni vektor 7×12MFCC_E



(b) Ulazna vektor 9×12MFCC_E



(c) Ulazni vektor 12MFCC_E_D_A

Slika 20: Uporedni prikaz WER-a u zavisnosti od broja izlaznih obeležja za različite vrste ulaznih obeležja. Značenje oznaka na graficima su sledeće: hlda d – varijanta HLDA koja pretpostavlja da su samo diskriminatorna obeležja nekorelisana, hlda – varijanta HLDA koja pretpostavlja da su sva transformisana obeležja nekorelisana, lda – varijanta LDA u kojoj se vrši normalizacija transformacione matrice i ref – WER GMM-a sa dijagonalnim kovarijansnim matricama i obeležjima 12MFCC_E_D.

U većini slučajeva normalizovana varijanta LDA rezultovala je nešto manjim WER-om u odnosu na odgovarajuću normalizovanu varijantu, ali razlika koja postoji između ova dva sistema po pitanju tačnosti i broja parametara je neznatna.

Pošto su ovi eksperimenti kao prvenstveni cilj imali procenu tačnosti sistema za prepoznavanje govora baziranu na (H)LDA, eksperimenti sa HLDA su ograničeni samo na nekoliko varijanata ulaznih obeležja koje su dale najniži WER sa LDA modelima. Performanse analiziranih sistema su navedene u tabelama 8 i 9, koje se međusobno razlikuju po tome da li ciljna funkcija pretpostavlja da su samo diskriminativna obeležja nekorelisana ili pak sva transformisana obeležja.[5] Kao što se iz priloženog može videti za različite skupove ulaznih obeležja najveće vrednosti WER-a se uglavnom dobijaju ukoliko je broj izlaznih obeležja 39. Ovo bi se ponovo moglo objasniti nedovoljnim brojem opservacija u skupu za obuku potrebnim za efikasnu estimaciju parametara u 39-dimenzionalnom prostoru, da ne postoji značajno odstupanje u varijanti sa konkatenativnim ulaznim vektorima sačinjenim od 9 uzastopnih opservacija i HLDA koja pretpostavlja da su samo diskriminativna obeležja nekorelisana.

Kao i u slučaju LDA, bez obzira na vrstu ulaznih obeležja najveća tačnost prepoznavanja se dobija ukoliko je broj izlaznih obeležja 35 ili 32. Za ovaj broj izlaznih obeležja relaksiranje ograničenja u vezi sa korelisanošću transformisanih obeležja je u slučaju konkatenativnih ulaznih obeležja rezultirao smanjenjem WER-a, što nije slučaj ukoliko su ulazna obeležja 12MFCC_E_D_A (videti sliku 20). Ovakvo ponašanje je vrlo verovatno posledica broja ulaznih obeležja, a ne njihove prirode, jer u posmatranim slučajevima konkatenativnih obeležja se vrši odbacivanje barem 52 odnosno 78 obeležja, dok u slučaju da su ulazna obeležja 12MFCC_E_D_A maksimalno se odbacuje svega 13 obeležja. Interesantno je da primena HLDA, odnosno relaksacija uslova po pitanju rasipanja unutar klasa (da rasipanje ne mora biti isto za sve klase), ne vodi nužno smanjenju WER-a (npr. ulazni vektor $7\times$12MFCC_E_D a broj izlaznih obeležja 26 i 39) što je takođe ilustrovano na slici 20.

6.3 STC

Ideja da je sa samo jednom transformacionom matricom moguće transformisati obeležja tako da ta obeležja budu nekorelisana u svakoj klasi je malo verovatna (Gales, 1999), ali se može uopštiti tako da se definiše po jedna transformaciona matrica za grupe stanja na čemu se zasniva STC model. Za razliku od (H)LDA, STC model ne podrazumeva redukciju dimenzionalnosti prostora obeležja, stoga su u ovom radu obeležja ograničena na 12MFCC_E_D za koje su referentni sistemi dali najmanju vrednost WER-a (videti tabelu 5). Prva varijanta je napravljena po uzoru na sisteme koji su predstavljeni u (Gales, 1999), a koja podrazumeva da se jedna matrica deli između svih Gausovih raspo-

5 HLDA pretpostavlja da se transformisana obeležja mogu podeliti na diskriminativna (ona koja nose informaciju o pripadnosti klasi) i nediskriminativna. U slučaju GMM-a pretpostavka da pojedina obeležja ne nose informaciju o pripadnosti klasi je ekvivalentna pretpostavci da su srednje vrednosti i kovarijanse raspodela klasa za ova obeležja iste u svim klasama, tako da se mogu izostaviti iz modela, te izlazna obeležja obuhvataju samo diskriminativna obeležja. Sa druge strane pretpostavka o nekorelisanosti pojedinih obeležja modifikuje ciljnu funkciju (videti (Jakovljević et al., 2013)).

dela koje fromiraju GMM-ove stanja trifona izvedenih iz istog monofona uz dodatno ograničenje da su ta stanja na istim pozicijama u modelu. Pošto procedura obuke podrazumeva određivanje optimalnog broja komponenata po stanju pomoću krosvalidacije, za razliku od testova navedenih u radu (Gales, 1999) gde je broj Gausovih raspodela postepeno povećavan i vršeno je poređenje GMM sa dijagonalnim kovarijansnim matricama i STC sa istim brojem Gausiana, u ovom radu je formiran samo jedan model sa "optimalnim" brojem Gausiana. Ovaj model će u nastavku teksta biti označen sa STCRef. Druga varijanta, koja nosi oznaku STC(2.0)/STC(1.0), predstavlja nešto precizniju aproksimaciju od prethodne, pošto se svaka grupa Gausovih raspodela formirana u prethodnoj varijanti dodatno deli sve dok maksimalna KLD između Gausovih raspodela sa estimiranom i aproksimiranom kovarijansnom matricom ne bude jednaka 3.0/1.5. Performanse ova tri sistema su prikazana u tabeli 10

Tabela 10: Performanse STC sistema.

Oznaka	# Gausiana	# Transformacionih matrica	# Parametara	WER [%]
STCRef	14654	145	0.87M	5.92
STC(2.0)	14654	854	1.35M	4.40
STC(1.0)	14654	2357	2.37M	3.65

Kao što se iz priloženog može videti sa porastom broja parametara koji opisuju model raste i tačnost prepoznavanja. Ukoliko se ovi rezultati uporede sa referentnim modelima koji koriste isti skup obeležja (12MFCC_E_D) može se uočiti da je broj parametara u slučaju STCRef skoro 3 puta manji od GMM-a sa dijagonalnom kovarijansnom matricom, ali je i WER značajno veći (relativno povećanje je oko 50%). Modeli sa većim broj parametara su dali bolje rezultate, jer je odstupanje od modela sa punim kovarijansnim matricama manje, ali i dalje broj grešaka znatno veći nego u slučaju GMM-a sa punim kovarijansnim matricama. Ukoliko se ovaj model uporedi sa HLDA modelima uočava se da je moguće postići sličnu tačnost uz smanjenje broja parametara od nekih 30%.

6.4 SEGMM

Kao što je napomenuto u poglavlju 4 SEGMM varijanta GMM-a polazi od pretpostavke da karakteristični vektori obrazuju potprostore u D dimenzionalnom vektorskom prostoru, odnosno da je za njih moguće pronaći retku reprezentaciju. Eksperimenti koji su ovde prikazani imali su za cilj da ispitaju mogućnost retke reprezentacije karakterističnih vektora kao i određivanje odgovarajuće kardinalnosti retkih vektora i rečnika. Broj atoma (k) je biran tako da se poklopi sa brojem karakterističnih vektora u slučaju 10, 20, 30 i 40 kovarijansnih matrica, dok je kardinalnost retkih vektora (d) postepeno povećavana od 3 do 7. Pošto inicijalni eksperimenti sa kardinalnošću retkih vektora jednakoj 3 nisu dali visoku tačnost prepoznavanja (lošiju od dijagonalnih modela za pojedine kombinacije rečnika) varijante sa nižom kardinalnošću nisu ispitivane. Sa druge strane vrednost kardinalnosti veće od 7 nisu bile od interesa zbog povećanja računske i memorijske složenosti modela (videti analizu u poglavlju 4).

Tabela 11: Prosečne vrednosti kvadratnih rastojanja između aproksimirane i stvarne vrednosti karakterističnih vektora kovarijansnih matrica za različite vrednost k i d.

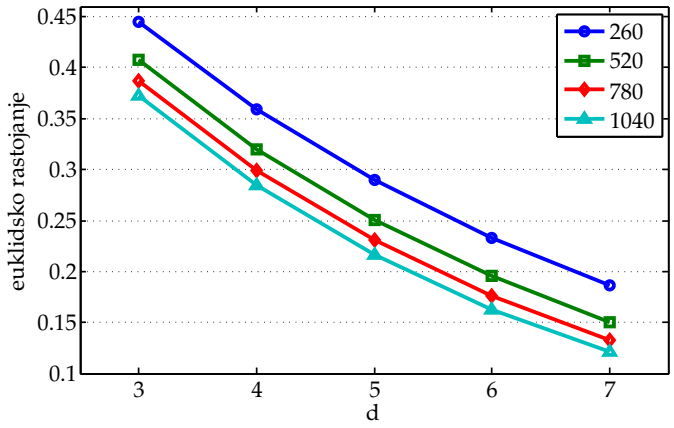
		k			
		260	520	780	1040
d	3	0.44	0.41	0.39	0.37
	4	0.36	0.32	0.30	0.28
	5	0.29	0.25	0.23	0.22
	6	0.23	0.20	0.18	0.16
	7	0.19	0.15	0.13	0.12

Tabela 12: Prosečne vrednosti KLD između Gausovih raspodela sa estimiranom i aporkismiranom kovarijansnom matricom za različite vrednost k i d.

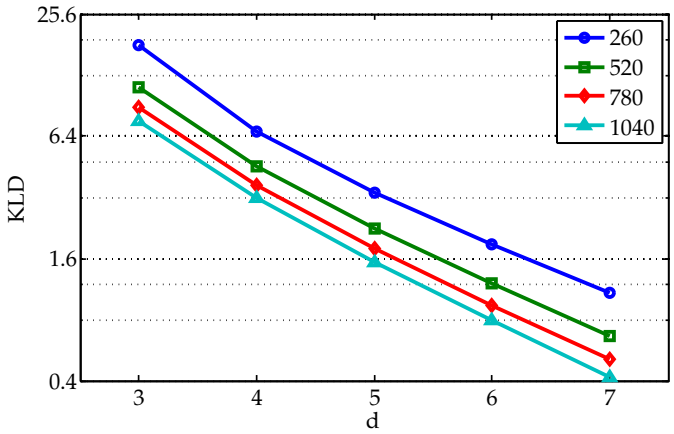
		k			
		260	520	780	1040
d	3	18.03	11.09	8.89	7.65
	4	6.77	4.54	3.67	3.17
	5	3.39	2.25	1.80	1.54
	6	1.87	1.20	0.94	0.79
	7	1.08	0.67	0.51	0.42

Pored samih vrednosti WER-a, pri obuci analizirano je i srednje kvadratno odstupanje aproksimirane vrednosti od stvarne vrednosti karakterisitčnog vektora, kao i KLD-ovi između Gausovih raspodela sa estimiranom i aproksimiranom kovarijansnom matricom. Dobijene vrednosti su prikazane u tabelama 11 i 12, kao i na slici 21. Kao što se na osnovu priloženog može videti povećanjem veličine rečnika i broja nenultih elemenata u retkom vektoru smanjuje se kako kvadratno rastojanje između stvarne i aproksimirane vrednosti karakterističnog vektora, ali i KLD između raspodela, što je i bilo očekivano (videti odeljak 4.4). Interesantan rezultat se može uočiti na slici 21 gde linearno smanjenje prosečnog kvadratnog odstupanja (e_{ED}) rezultuje eksponencijalnim smanjenjem prosečnog KLD-a (e_{KLD}), koje se može opisati jednačinom $e_{KLD} = e^{11 e_{ED} - 2.6}$.[6] Ono što je bitno jeste da eksperimentalni rezultati potvrđuju delimično očekivanu hipotezu da smanjenje kvadratnog rastojanja između aproksimirane i stvarne vrednosti karakterističnog vektora smanjuje i odstupanje Gausove raspodele sa aproksimiranom i estimiranom kovarijansnom matricom.

6 Do navedene formule se došlo na osnovu eksponencijalne aproksimacije metodom najmanjih kvadrata, a ne na osnovu nekih teorijskih predviđanja.



(a)



(b)

Slika 21: Zavisnost prosečne vrednosti kvadratnog rastojanja (a) i KLD (b) od d za različite vrednosti k

Zavisnost WER-a od kardinalnosti retkog vektora i broja atoma u rečniku je data u tabeli 14. Kao što se iz priloženog može videti da WER uglavnom ima trend smanjenja sa povećanjem broja atoma koji čine rečnik i sa povećanjem kardinalnosti retkih vektora, što je i očekivano jer se razlika između modela smanjuje. Treba primetiti da manja vrednost prosečnog KLD-a ne znači nužno i veću tačnost prepoznavanja (npr: varijante sistema za $d = 7$ i $k = 520$ ili $k = 780$ imaju manji prosečni KLD od sistema sa $d = 6$ i $k = 1024$, ali i prva dva sistema imaju i veći WER). Ovo se delimično može objasniti činjenicom da ciljna funkcija koja obezbeđuje što bolje pokrivanje prostora obeležja koje zahvata model ne odgovara stvarnom zadatku modela, a to je diskriminacija između klasa, te postoji mogućnost da se povećalo i preklapanje između klasa odnosno konfuzija. Treba primetiti da je SEGMM za $k = 1024$ i $d = 6$ i $d = 7$ rezultovao manjim brojem grešaka nego odgovarajući referentni GMM sa punim kovarijansnim matricama. Ovaj boljitak bi se mogao objasniti činjenicom da se redukcijom broja parametara dobio model koji je generalniji od

Tabela 13: Broj parametara modela za različite vrednost k i d.

		k			
		260	520	780	1040
d	3	1.91M	1.92M	1.93M	1.93M
	4	2.29M	2.30M	2.31M	2.31M
	5	2.67M	2.68M	2.68M	2.69M
	6	3.05M	3.06M	3.06M	3.07M
	7	3.43M	3.44M	3.44M	3.45M

Tabela 14: Vrednosti WER [%] za različite vrednost k i d.

		k			
		260	520	780	1040
d	3	5.60	4.79	4.54	4.08
	4	4.43	3.33	3.02	3.55
	5	3.19	3.19	2.63	2.31
	6	2.98	2.45	2.31	1.92
	7	2.31	2.45	2.27	1.95

referentnog modela. Ako se uporedi broj parametara koji je potreban za opisivanje modela, može se uočiti da je manji od onog koji je potreban u slučaju GMM-a sa punim kovarijansnim matricama i onog koji se dobija na osnovu (H)LDA, ali i dalje veći od onog koji je potreban za GMM sa dijagonalnim kovarijansnim matricama. Treba primetiti da broj parametara neznatno raste sa porastom broja atoma koji čine rečnik, ali ne treba smetnuti sa uma da veličina rečnika u značajnoj meri utiče na broj računskih operacija potrebnih za izračunavanje vrednosti gustine emitovanja stanja i da pri tome ne zavisi od broja aktivnih stanja.

6.5 REZIME

U ovom odeljku je dat prikaz performansi koji su dobijeni za nekoliko različitih varijanata GMM-a. Sa aspekta tačnosti prepoznavanja, kao najbolji su se pokazali SEGMM i GMM sa punim kovarijansnim matricama, pri čemu je SEGMM koji predstavlja aproksimaciju GMM-a sa punim kovarijansnim matricama imao bolju tačnost, što se može objasniti boljom estimacijom usled manjeg broja parametara. SEGMM se po broju parametara našao u grupi sa (H)LDA modelom, ali daleko iznad STC i GMM-a sa dijagonalnim kovarijansnim matricama. Treba napomenuti da STC iako predstavlja uopštenje (H)LDA pristupa, nije rezultovao značajno manjom greškom prepoznavanja (naprotiv

greška je i porasla) što se može objasniti malom složenošću modela koja je manja i od GMM-a sa dijagonalnim kovarijansnim matricama. Drugi razlog bi mogao biti u činjenici da (H)LDA koristi obeležja koja u obzir uzimaju nešto širi kontekst zbog korišćenja delta-delta obeležja, što nije slučaj za STC.

7

ZAKLJUCAK

U ovom radu je predstavljen novi način aproksimacije inverznih kovarijansnih matrica u Gausovim mešavinama koji se zasniva na retkoj reprezentaciji njihovih karakterističnih vektora. Cilj ove aproksimacije jeste: *i*) smanjenje broja parametara potrebnih za reprezentaciju modela, *ii*) smanjenje računske složenosti izračunavanja izglednosti i *iii*) obezbeđivanje dovoljne tačnosti modela. Pored samog opisa modela u radu je dat i detaljan opis procedure obuke modela, kao i formalni dokaz da se predložena obuka uklapa u standardnu obuku zasnovanu na principu maksimizacije izglednosti. Urađeno je poređenje predloženog modela sa već postojećim načinima aproksimacije (inverznih) kovarijansnih matrica, kao što su dijagonalan, MLLT (koja je ovde tretirana kao specijalni slučaj HLDA), STC i PCGMM, sa aspekta broja parametara potrebnih za opis modela kao i broja računskih operacija potrebnih za izračunavanje izglednosti. Pored memorijske i računske kompleksnosti modela analizirana je i tačnost prepoznavanja, za šta su korišćene govorne baze SpeechDat II i S70W100s120T na srpskom jeziku. Pošto su testovi imali za cilj proveru tačnosti akustičkih modela koji su opisani pomoć različitih varijanata GMM-a, svi testovi su realizovani korišćenjem gramatike nezavisne od konteksta. Računska kompleksnost je iskazana samo u opštim brojevima, pošto prilikom testiranja broj Gausovih mešavina za koje treba izračunati vrednost zavisi od broja aktivnih stanja koji se vremenom menja (od opservacije do opservacije) i broja reči koje je potrebno prepoznati što je u ovom primeru oko 200 reči što je relativno malo za adekvatnu procenu. Što se tiče tačnosti za odgovarajući broj parametara predloženi model aproksimacije je postigao tačnost prepoznavanja koja je bila u nivou tačnosti GMM-a sa punim kovarijansnim matricama uz redukciju broja parametara od nekih 40%. U pojedinim slučajevima predloženi model je imao i nešto bolju tačnost od GMM-a sa punim kovarijansnim matricama što se može objasniti manjim brojem parametara koji se estimiraju i samim tim robustnijom estimacijom. U odnosu na dijagonanlu aproksimaciju predloženi model za dovoljno veliki rečnik i približno isti broj parametara postiže znatno veću tačnost. Sličan odnos postoji i za preostale alternativne modele (HLDA i STC), pri čemu je razlika u tačnosti nešto manja. Treba napomenuti da su eksperimenti sa (H)LDA u kojima se vrši redukcija obeležja dali nešto veću tačnost od odgovarajućih STC modela što ukazuje da se dodatno povećanje tačnosti može dobiti korišćenjem i delta-delta obeležja uz obaveznu redukciju dimenzionalnosti.

Procedura obuke predloženog modela je značajno produžena u odnosu na standardnu proceduru obuke, pošto uključuje i proceduru pronalaženja rečnika i odgovarajućih retkih reprezentacija za svaki od karakterističnih vektora. Trajanje obuke nije kritičan faktor pošto obuku nije potrebno realizovati u realnom vremenu za razliku od prepoznavanja (dekovanja), ali bi u nekom daljem istraživanju bilo interesantno razmotriti i mogućnosti ubrzanja procedure obuke. Pri obuci predloženog metoda potrebno je estimirati uzoračku kovarijansu, koja treba da bude dobro kondicionirana, stoga iako je broj parametara

modela značajno redukovan neophodno je obezbediti dovoljan broj opservacija po komponenti Gausove mešavine. Budući pravac istraživanja bi trebao obuhvatiti analizu ponašanja modela u situaciji kada je broj opservacija suviše mali. Iako je sa aspekta brzine daleko značajnije aproksimirati kovarijansne matrice, sličan princip retke reprezentacije bi se mogao primeniti i na srednje vrednosti, što bi moglo dovesti do povećanja tačnosti prepoznavanja uz neznatno povećanje brzine dekodovanja. U slučaju kada je na raspolaganju velika količina podataka za obuku diskriminativni trening dovodi do povećanja tačnosti prepoznavanja modela, tako da bi bilo interesantno predložen model uklopiti u obuku u kojoj se maksimizuje među informacija ili minimizuje greška na nivou fonema odnosno reči. Predloženi model je testiran na sistemu za prepoznavanje govora, ali zbog široke primene GMM modela jedan od budućih pravaca bi mogla biti primena predloženog modela u nekim drugim oblastima prepoznavanja oblika koje koriste GMM sa velikim brojem Gausovih raspodela.

A

DODATAK

A.1 DISPERZIVNOST DIJAGONALNIH ELEMENATA I KARAKTERISTIČNIH VREDNOSTI KOVARIJANSNE MATRICE

U ovom delu je naveden formalni dokaz da je u slučaju simetrične matrice disperzivnost njenih karakterističnih vrednosti veća od disperzivnosti njenih dijagonalnih elemenata.

Naka je $\mathbf{A}$ proizvoljna simetrična matrica dimenzija $d \times d$, i neka su λ_i njene karakteristične vrednosti, tada važi $\sum_{i=1}^{d} A_{i,i} = \sum_{i=1}^{d} \lambda_i$, stoga su prosečne vrednosti dijagonalnih elemenata i karakterističnih vrednosti jednake. U nastavku će prosečna vrednost dijagonalnih elemenata biti označena sa $\bar{\lambda}$.

$$\sum_{i=1}^{d} \left(A_{i,i} - \bar{\lambda}\right)^2 \leqslant \sum_{i=1}^{d} \left(A_{i,i} - \bar{\lambda}\right)^2 + \sum_{i=1}^{d} \sum_{\substack{j=1 \\ i \neq j}}^{d} A_{i,j}^2 = \operatorname{tr}\left\{\left(\mathbf{A} - \bar{\lambda}\mathbf{I}\right)^T \left(\mathbf{A} - \bar{\lambda}\mathbf{I}\right)\right\}$$

Cilj je povezati gornje jednačine sa karakterističnim vrednostima matrice $\mathbf{A}$ do kojih se dolazi množenjem sa matricama karakterističnih vektora $\mathbf{Q}$ za koju važi: $\mathbf{Q}^{-1} = \mathbf{Q}^T$ i $|\mathbf{Q}| = 1$.

$$\begin{aligned}
\operatorname{tr} & \left\{\left(\mathbf{Q}^T\mathbf{A}\mathbf{Q} - \bar{\lambda}\mathbf{I}\right)^T \left(\mathbf{Q}^T\mathbf{A}\mathbf{Q} - \bar{\lambda}\mathbf{I}\right)\right\} = \\
= & \operatorname{tr}\left(\mathbf{Q}^T\mathbf{A}^T\mathbf{Q}\mathbf{Q}^T\mathbf{A}\mathbf{Q} - \bar{\lambda}\mathbf{Q}^T\mathbf{A}\mathbf{Q} - \bar{\lambda}\mathbf{Q}^T\mathbf{A}^T\mathbf{Q} + \bar{\lambda}^2\mathbf{I}\right) \\
= & \operatorname{tr}\left(\mathbf{Q}^T\mathbf{A}^T\mathbf{A}\mathbf{Q}\right) - \operatorname{tr}\left(\bar{\lambda}\mathbf{Q}^T\mathbf{A}\mathbf{Q}\right) - \operatorname{tr}\left(\bar{\lambda}\mathbf{Q}^T\mathbf{A}^T\mathbf{Q}\right) + \operatorname{tr}\left(\bar{\lambda}^2\mathbf{I}\right) \\
= & \operatorname{tr}\left(\mathbf{A}^T\mathbf{A}\mathbf{Q}\mathbf{Q}^T\right) - \operatorname{tr}\left(\bar{\lambda}\mathbf{A}\mathbf{Q}\mathbf{Q}^T\right) - \operatorname{tr}\left(\bar{\lambda}\mathbf{A}^T\mathbf{Q}\mathbf{Q}^T\right) + \operatorname{tr}\left(\bar{\lambda}^2\mathbf{I}\right) \\
= & \operatorname{tr}\left(\mathbf{A}^T\mathbf{A}\right) - \operatorname{tr}\left(\bar{\lambda}\mathbf{A}\right) - \operatorname{tr}\left(\bar{\lambda}\mathbf{A}^T\right) + \operatorname{tr}\left(\bar{\lambda}^2\mathbf{I}\right) \\
= & \operatorname{tr}\left(\mathbf{A}^T\mathbf{A} - \bar{\lambda}\mathbf{A} - \bar{\lambda}\mathbf{A}^T + \bar{\lambda}^2\mathbf{I}\right) \\
= & \operatorname{tr}\left\{\left(\mathbf{A} - \bar{\lambda}\mathbf{I}\right)^T \left(\mathbf{A} - \bar{\lambda}\mathbf{I}\right)\right\}
\end{aligned}$$

Treba primetiti da gornja jednakost važi i za bilo koju drugu matricu koja je ortogonalna i čija je determinanta jednaka jedan.

Odavde sledi:

$$\sum_{i=1}^{d} \left(A_{i,i} - \bar{\lambda}\right)^2 \leqslant \operatorname{tr}\left\{\left(\mathbf{Q}^T\mathbf{A}\mathbf{Q} - \bar{\lambda}\mathbf{I}\right)^T \left(\mathbf{Q}^T\mathbf{A}\mathbf{Q} - \bar{\lambda}\mathbf{I}\right)\right\} = \sum_{i=1}^{d} \left(\lambda_i - \bar{\lambda}\right)^2$$

odnosno disperzija dijagonalnih elemenata matrice manja je od disperzije karakterističnih vrednosti.

A.2 OPTIMALNA ESTIMACIJA REČNIKA

A.2.1 *Gradijentna metoda*

Ukoliko su određene vrednosti koeficijenata $\mathbf{A}$ tada se problem pronalaženja vrednosti atoma svodi na minimizaciju sledeće funkcije:

$$f(\mathbf{D}) = \frac{1}{2}\sum_{i=1}^{n} \|x_i - \mathbf{D}\alpha_i\|_2^2 = \frac{1}{2}\text{trag}\left((\mathbf{X} - \mathbf{D}\mathbf{A})^T (\mathbf{X} - \mathbf{D}\mathbf{A})\right)$$

gde $\mathbf{X}$ predstavlja matricu svih signala koji su na raspolaganju $\mathbf{X} = [x_1, x_2, \ldots, x_n]$, a matrica $\mathbf{A}$ predstavlja matricu odgovarajućih retkih vektora $\mathbf{A} = [\alpha_1, \alpha_2, \ldots, \alpha_n]$. Pošto važi sledeće jednakosti:

$$\begin{aligned} \frac{\partial \text{trag}(\mathbf{B}^T\mathbf{B})}{\partial \mathbf{B}} &= 2\mathbf{B} \\ \frac{\partial \mathbf{B}\mathbf{C}}{\partial \mathbf{B}} &= \mathbf{C}^T \end{aligned}$$

gde su $\mathbf{B}$ i $\mathbf{C}$ proizvoljne matrice, diferenciranjem funkcije $f(\mathbf{D})$ dobija se:

$$\nabla f(\mathbf{D}) = -(\mathbf{X} - \mathbf{D}\mathbf{A})\,\mathbf{A}^T$$

Nova vrednost rečnika $\mathbf{D}_n$ se dobija na osnovu stare $\mathbf{D}_o$ na sledeći način:

$$\mathbf{D}_n = \mathbf{D}_o + \eta\,(\mathbf{X} - \mathbf{D}_o\mathbf{A})\,\mathbf{A}^T$$

gde je sa η označen koeficijent brzine učenja. Procedura se ponavlja sve dok razlika između odgovarajućih starih i novih atoma ne postane dovoljno mala $\|\mathbf{D}_n - \mathbf{D}_o\|_F^2$.

A.2.2 *Lagranžov dualni problem*

U ovom slučaju je problem minimizacije kvadratnog odstupanja dodatno proširen ograničenjem da l_2 norma atoma bude manja ili jednaka 1, odnosno problem koji je potrebno rešiti je sledeći:

$$\min_{\mathbf{D}} = \text{trag}\left((\mathbf{X} - \mathbf{D}\mathbf{A})^T (\mathbf{X} - \mathbf{D}\mathbf{A})\right) \text{ tako da } \|\mathbf{d}_i\|_2^2 \leqslant 1, \forall i = 1, 2, \ldots, k.$$

Ovaj problem se efikasno rešava korišćenjem Lagranžovog duala. Prvo se formira Lagranžijan:

$$\mathcal{L}(\mathbf{D}, \mathbf{\Lambda}) = \text{trag}\left((\mathbf{X} - \mathbf{D}\mathbf{A})^T (\mathbf{X} - \mathbf{D}\mathbf{A})\right) + \text{trag}\left(\mathbf{\Lambda}\left(\mathbf{D}^T\mathbf{D} - \mathbf{I}\right)\right)$$

gde je $\mathbf{\Lambda}$ dijagonalna matrica koja na glavnoj dijagonali sadrži Lagranžove koeficijente. Diferencijranjem $\mathcal{L}(\mathbf{D}, \mathbf{\Lambda})$ pod $\mathbf{D}$ i izjednačavanjem sa nula matricom dobija se optimalna vrednost $\mathbf{D}$ u funkciji parametra $\mathbf{\Lambda}$ odnosno:

$$-2\,(\mathbf{X} - \mathbf{D}\mathbf{A})\,\mathbf{A}^T + 2\mathbf{D}\mathbf{\Lambda} = 0$$

odnosno:

$$\mathbf{D} = \mathbf{X}\mathbf{A}^T\left(\mathbf{A}\mathbf{A}^T + \mathbf{\Lambda}\right)^{-1}.$$

Preostaje pronaći odgovarajuće vrednosti $\mathbf{\Lambda}$. Ako sa $f_0(\mathbf{D})$, označimo ciljnu funkciju, odnosno $f_0(\mathbf{D}) = \operatorname{trag}\left((\mathbf{X}-\mathbf{DA})^T(\mathbf{X}-\mathbf{DA})\right)$ tada važi $f_0(\mathbf{D}) \geqslant \mathcal{L}(\mathbf{D},\mathbf{\Lambda})$ ukoliko su vrednosti Lagranžovih multiplikatora pozitivne. Pored toga ako sa $g(\mathbf{\Lambda})$ označimo funkciju $\min_{\mathbf{D}} \mathcal{L}(\mathbf{D},\mathbf{\Lambda})$ tada važi sledeće:

$$g(\mathbf{\Lambda}) \leqslant f_0(\mathbf{D})$$

odnosno najbolja granica se dobija ukoliko se nađe maksimum funkcije $g(\mathbf{\Lambda})$.

U ovom primeru $g(\mathbf{\Lambda})$ je data izrazom

$$\begin{aligned} g(\mathbf{\Lambda}) &= \operatorname{trag}\left(\left(\mathbf{X}-\mathbf{XA}^T\left(\mathbf{AA}^T+\mathbf{\Lambda}\right)^{-1}\mathbf{A}\right)^T\left(\mathbf{X}-\mathbf{XA}^T\left(\mathbf{AA}^T+\mathbf{\Lambda}\right)^{-1}\mathbf{A}\right)\right) \\ &\quad + \operatorname{trag}\left(\mathbf{\Lambda}\left(\left(\mathbf{XA}^T\left(\mathbf{AA}^T+\mathbf{\Lambda}\right)^{-1}\right)^T\mathbf{XA}^T\left(\mathbf{AA}^T+\mathbf{\Lambda}\right)^{-1}-\mathbf{I}\right)\right) \\ &= \operatorname{trag}\left(\mathbf{X}^T\mathbf{X}\right)-\operatorname{trag}\left(\mathbf{A}^T\left(\mathbf{AA}^T+\mathbf{\Lambda}\right)^{-1T}\mathbf{AX}^T\mathbf{X}\right) \\ &\quad - \operatorname{trag}\left(\mathbf{X}^T\mathbf{XA}^T\left(\mathbf{AA}^T+\mathbf{\Lambda}\right)^{-1}\mathbf{A}\right) \\ &\quad + \operatorname{trag}\left(\mathbf{A}^T\left(\mathbf{AA}^T+\mathbf{\Lambda}\right)^{-1T}\mathbf{AX}^T\mathbf{XA}^T\left(\mathbf{AA}^T+\mathbf{\Lambda}\right)^{-1}\mathbf{A}\right) \\ &\quad + \operatorname{trag}\left(\mathbf{\Lambda}\left(\mathbf{AA}^T+\mathbf{\Lambda}\right)^{-1T}\mathbf{AX}^T\mathbf{XA}^T\left(\mathbf{AA}^T+\mathbf{\Lambda}\right)^{-1}\right)-\operatorname{trag}(\mathbf{\Lambda}) \end{aligned}$$

koji se primenom osobina traga, i činjenice da je matrica $(\mathbf{AA}^T+\mathbf{\Lambda})$ simetrična svodi na:

$$\begin{aligned} g(\mathbf{\Lambda}) &= \operatorname{trag}\left(\mathbf{X}^T\mathbf{X}\right)-2\operatorname{trag}\left(\mathbf{A}^T\left(\mathbf{AA}^T+\mathbf{\Lambda}\right)^{-1}\mathbf{AX}^T\mathbf{X}\right) \\ &\quad + \operatorname{trag}\left(\left(\mathbf{AA}^T+\mathbf{\Lambda}\right)^{-1}\mathbf{AX}^T\mathbf{XA}^T\left(\mathbf{AA}^T+\mathbf{\Lambda}\right)^{-1}\left(\mathbf{AA}^T+\mathbf{\Lambda}\right)\right) \\ &\quad - \operatorname{trag}(\mathbf{\Lambda}) \end{aligned}$$

odnosno:

$$g(\mathbf{\Lambda}) = \operatorname{trag}\left(\mathbf{X}^T\mathbf{X}-\mathbf{A}^T\left(\mathbf{AA}^T+\mathbf{\Lambda}\right)^{-1}\mathbf{AX}^T\mathbf{X}-\mathbf{\Lambda}\right)$$

Dobijenu funkciju $g(\mathbf{\Lambda})$ je moguće optimizovati pomoću Njutnove metode, za šta je potrebno odrediti vrednosti gradijenta i Hesijana (Hessian) funkcije $g(\mathbf{\Lambda})$.

Korišćenjem osobine:

$$\frac{\partial \operatorname{trag}\left(\mathbf{B}^{-1}\mathbf{C}\right)}{\partial \mathbf{B}} = -\mathbf{B}^{-1T}\mathbf{CB}^{-1T}$$

dobija se:

$$\frac{\partial g(\mathbf{\Lambda})}{\partial \mathbf{\Lambda}} = \left(\mathbf{AA}^T+\mathbf{\Lambda}\right)^{-1}\mathbf{AX}^T\mathbf{XA}^T\left(\mathbf{AA}^T+\mathbf{\Lambda}\right)^{-1}-\mathbf{I}$$

Pošto je matrica $\boldsymbol{\Lambda}$ dijagonalna, odnosno interesuju nas samo elementi na glavnoj dijagonali gornji izraz se može preurediti u sledeći:

$$\frac{\partial g(\boldsymbol{\Lambda})}{\partial \lambda_i} = \|\mathbf{X}\mathbf{A}^T \left(\mathbf{A}\mathbf{A}^T + \boldsymbol{\Lambda}\right)^{-1} \boldsymbol{e}_i\|_2^2 - 1$$

gde je $\boldsymbol{e}_i$ vektor čiji je samo i-ti element jednak 1, a svi ostali elementi jednaki 0, tzv. i-ti jedinični vektor.

Diferencijranjem $\partial g(\boldsymbol{\Lambda})/\partial \lambda_i$ po λ_j se dobijaju vrednosti elemenata Hesijana:

$$\frac{\partial^2 g(\boldsymbol{\Lambda})}{\partial \lambda_i \partial \lambda_j} = -2 \left[\left(\mathbf{A}\mathbf{A}^T + \boldsymbol{\Lambda}\right)^{-1} \mathbf{A}\mathbf{X}^T\mathbf{X}\mathbf{A}^T \left(\mathbf{A}\mathbf{A}^T + \boldsymbol{\Lambda}\right)^{-1}\right]_{i,j} \left[\left(\mathbf{A}\mathbf{A}^T + \boldsymbol{\Lambda}\right)^{-1}\right]_{i,j}$$

što je bilo neophodno za Njutnovu metodu optimizacije.

Prednost dualne metode je u smanjenju broja promenljivih koje učestvuju u optimizaciji, umesto svih elemenata rečnika optimizuju se samo Langražovi koeficijenti čiji je broj jednak broju atoma u rečniku.

BIBLIOGRAFIJA

M. Aharon, M. Elad, i A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, str. 4311–4322, 2006. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1710377.

S. Axelrod, V. Goel, R. A. Gopinath, P. A. Olsen, i K. Visweswariah. Subspace constrained gaussian mixture models for speech recognition. *Speech and Audio Processing, IEEE Transactions on*, vol. 13, no. 6, str. 1144–1160, 2005. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1518915.

J. A. Bilmes. Factored sparse inverse covariance matrices. U *Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on*, vol. 2, str. II1009–II1012. IEEE, 2000. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=859133.

C. M. Bishop. *Pattern recognition and machine learning*, vol. 1. springer New York, 2006.

B. Burdge, K. Kreutz-Delgado, i J. Murray. A unified focuss framework for learning sparse dictionaries and non-squared error. U *Signals, Systems and Computers (ASILOMAR), 2010 Conference Record of the Forty Fourth Asilomar Conference on*, str. 2037–2041. IEEE, 2010. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5757905.

Y.-H. B. Chiu i R. M. Stern. Analysis of physiologically-motivated signal processing for robust speech recognition. U *INTERSPEECH*, str. 1000–1003, 2008.

V. Delić, M. Sečujski, N. Jakovljević, M. Janev, R. Obradović, i D. Pekar. *Speech Technologies for Serbian and Kindred South Slavic Languages*, chapter 9, str. 141–164. SCIYO, 2010. URL http://www.intechopen.com/books/advances-in-speech-recognition/speech-technologies-for-serbian-and-kindred-south-slavic-languages.

V. Delić, M. Sečujski, N. Jakovljević, D. Pekar, D. Mišković, B. Popović, S. Ostrogonac, M. Bojanić, i D. Knežević. Speech and language resources within speech recognition and synthesis systems for serbian and kindred south slavic languages. U M. Železný, I. Habernal, i A. Ronzhin, urednici, *Speech and Computer*, vol. 8113 of *Lecture Notes in Computer Science*, str. 319–326. Springer International Publishing, 2013. ISBN 978-3-319-01930-7. doi: 10.1007/978-3-319-01931-4_42. URL http://dx.doi.org/10.1007/978-3-319-01931-4_42. 15th International Conference, SPECOM 2013, Pilsen, Czech Republic, September 1-5, 2013. Proceedings.

L. Deng, A. Acero, L. Jiang, J. Droppo, i X. Huang. High-performance robust speech recognition using stereo training data. U *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, vol. 1, str. 301–304 vol.1, 2001. doi: 10.1109/ICASSP.2001.940827.

P. L. Dognin, V. Goel, J. R. Hershey, i P. A. Olsen. A fast, accurate approximation to log likelihood of gaussian mixture models. U *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, str. 3817–3820. IEEE, 2009. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4960459.

M. Elad. *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer, 2010.

K. Engan, S. O. Aase, i J. Husoy. Frame based signal compression using method of optimal directions (mod). U *Circuits and Systems, 1999. ISCAS'99. Proceedings of the 1999 IEEE International Symposium on*, vol. 4, str. 1–4. IEEE, 1999. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=779928.

M. Gales i S. Young. The application of hidden markov models in speech recognition. *Foundations and Trends in Signal Processing*, vol. 1, no. 3, str. 195–304, 2008. URL http://dl.acm.org/citation.cfm?id=1373537.

M. J. Gales. Semi-tied covariance matrices for hidden markov models. *Speech and Audio Processing, IEEE Transactions on*, vol. 7, no. 3, str. 272–281, May 1999. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=759034.

M. J. Gales i P. Woodland. Mean and variance adaptation within the mllr framework. *Computer Speech & Language*, vol. 10, no. 4, str. 249–264, 1996. URL http://www.sciencedirect.com/science/article/pii/S0885230896900133.

N. K. Goel i R. A. Gopinath. Multiple linear transforms. U *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, vol. 1, str. 481–484. IEEE, 2001. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=940872.

G. H. Golub i C. F. van Van Loan. *Matrix Computations (Johns Hopkins Studies in Mathematical Sciences)(3rd Edition)*. Johns Hopkins University Press, 1996. ISBN 0801854148.

R. A. Gopinath, B. Ramabhadran, i S. Dharanipragada. Factor analysis invariant to linear transformations of data. U *Proceedings International Conference on Speech and Language Processing*, str. 397–400, 1998.

I. F. Gorodnitsky i B. D. Rao. Sparse signal reconstruction from limited data using focuss: A re-weighted minimum norm algorithm. *Signal Processing, IEEE Transactions on*, vol. 45, no. 3, str. 600–616, 1997. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=558475.

R. Haeb-Umbach i H. Ney. Linear discriminant analysis for improved large vocabulary continuous speech recognition. U *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*, vol. 1, str. 13–16. IEEE, 1992. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=225984.

F. E. Hilger. *Quantile based histogram equalization for noise robust speech recognition*. PhD thesis, Rheinisch-Westfalischen Technischen Hochschule Aachen, 2004. URL http://darwin.bth.rwth-aachen.de/opus/volltexte/2005/1050/.

X. Huang i K. F. Lee. On speaker-independent, speaker-dependent, and speaker-adaptive speech recognition. *Speech and Audio Processing, IEEE Transactions on*, vol. 1, no. 2, str. 150–157, 1993. ISSN 1063-6676. doi: 10.1109/89.222875.

X. Huang, A. Acero, i H.-W. Hon. *Spoken language processing: a guide to theory, algorithm, and system development*. Prentice Hall PTR New Jersey, Upper Saddle River, New Jersey, 2001.

I. C. F. Ipsen i R. Rehman. Perturbation bounds for determinants and characteristic polynomials. *SIAM J. Matrix Analysis Applications*, vol. 30, no. 2, str. 762–776, 2008. doi: http://dx.doi.org/10.1137/070704770.

N. Jakovljević, M. Janev, D. Pekar, i D. Mišković. Energy normalization in automatic speech recognition. U *Text, Speech and Dialogue*, str. 341–347. Springer, 2008. URL http://link.springer.com/chapter/10.1007/978-3-540-87391-4_44.

N. Jakovljevič, M. Sečujski, i V. Delić. Vocal tract length normalization strategy based on maximum likelihood criterion. U *EUROCON 2009*, str. 399–402. IEEE, 2009. doi: 10.1109/EURCON.2009.5167662. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5167662.

N. Jakovljević, D. Mišković, M. Janev, D. Knežević, i T. Grbić. Primena linearne diskriminativne analize u prepoznavanju govora. U *Zbornik radova 9. konferencija Digitalna obrada govora i slike (DOGS)*, str. 40–43, 2012. ISBN 978-86-7892-439-2.

N. Jakovljević, D. M. Mišković, M. Janev, M. Sečujski, i V. Delić. Comparison of linear discriminant analysis approaches in automatic speech recognition. *Electronics and Electrical Engineering*, vol. 19, no. 7, str. 76–79, 2013. doi: http://dx.doi.org/10.5755/j01.eee.19.7.5167. URL http://www.eejournal.ktu.lt/index.php/elt/article/view/5167.

M. Janev, N. Jakovljević, i D. Pekar. Poredenje sistema za prepoznavanje govora na srpskom jeziku baziranih na punim i dijagonalnim kovarijansnim matricama. U *Telecommunications Forum (TELFOR), 2007 15th*, str. 342–345, 2007.

M. Janev, D. Pekar, N. Jakovljević, i V. Delić. Eigenvalues driven gaussian selection in continuous speech recognition using hmms with full covariance matrices. *Applied Intelligence*, vol. 33, no. 2, str. 107–116, 2010. doi: 10.1007/s10489-008-0152-9. URL http://link.springer.com/article/10.1007/s10489-008-0152-9.

X. Jiang. Linear subspace learning-based dimensionality reduction. *Signal Processing Magazine, IEEE*, vol. 28, no. 2, str. 16–26, 2011. ISSN 1053-5888. doi: 10.1109/MSP.2010.939041.

D. Jurafsky, J. H. Martin, A. Kehler, K. Vander Linden, i N. Ward. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Prentice-Hall Inc., eglewood clifs, new jersey edition, 2000.

S. Kotz i N. L. Johnson, urednici. *Breakthroughs in Statistics: Volume 1: Foundations and Basic Theory*, vol. 1 of *Springer Series in Statistics*. Springer, corrected edition, 1993.

N. Kumar. *Investigation of Silicon Auditory Models and Generalization of Linear Discriminant Analysis for Improved Speech Recognition*. PhD thesis, Johns Hopkins University, 1997. URL old-site.clsp.jhu.edu/~kumar/thesis.ps.

N. Kumar i A. Andreou. Heteroscedastic discriminant analysis and reduced rank hmms for improved speech recognition. *Speech Communication*, vol. 26, no. 4, str. 283–297, December 1998. doi: 10.1016/S0167-6393(98)00061-2.

O. Ledoit i M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, vol. 88, no. 2, str. 365–411, 2004. URL http://www.sciencedirect.com/science/article/pii/S0047259X03000964.

H. Lee, A. Battle, R. Raina, i A. Ng. Efficient sparse coding algorithms. U *Advances in neural information processing systems*, str. 801–808, 2006.

J. Mairal. *Sparse coding for machine learning, image processing and computer vision*. PhD thesis, École normale supérieure de Cachan-ENS Cachan, 2010. URL http://tel.archives-ouvertes.fr/tel-00595312/.

J. Mairal, F. Bach, J. Ponce, i G. Sapiro. Online dictionary learning for sparse coding. U *Proceedings of the 26th Annual International Conference on Machine Learning*, str. 689–696. ACM, 2009. URL http://dl.acm.org/citation.cfm?id=1553463.

J. Mairal, F. Bach, J. Ponce, i G. Sapiro. Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, vol. 11, str. 19–60, 2010.

N. Morgan. Deep and wide: Multiple layers in automatic speech recognition. *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 1, str. 7–13, 2012. doi: 10.1109/TASL.2011.2116010.

N. Morgan, H. Bourlard, i H. Hermansky. *Automatic Speech Recognition: An Auditory Perspective*, chapter 6, str. 309–338. Number 18 in Springer Handbook of Auditory Research. Springer, 2004. ISBN 9781441918314.

P. A. Olsen i R. A. Gopinath. Modeling inverse covariance matrices by basis expansion. *Speech and Audio Processing, IEEE Transactions on*, vol. 12, no. 1, str. 37–46, 2004. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1261270.

B. A. Olshausen i D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, vol. 381, no. 6583, str. 607–609, 1996.

D. Pekar, N. Jakovljević, M. Janev, D. Misković, i V. Delić. On the use of higher frame rate in the training phase of asr. U *14th WSEAS Int. Conf. on Latest Trends on Comp*, str. 127–130. WSEAS, 2010. URL link:http://www.wseas.us/e-library/conferences/2010/Corfu/COMPUTERS/COMPUTERS1-17.pdf.

R. Pieraccini. *The Voice in the Machine: Building Computers That Understand Speech*. The MIT Press, 2012. ISBN 0262016850.

K. Plataniotis i D. Hatzinakos. *Gaussian Mixtures and their Applications to Signal Processing*, chapter 3, str. 3.1–3.32. CRC Press, 1 edition, December 2000.

B. Popović, M. Janev, D. Pekar, N. Jakovljević, M. Gnjatović, M. Sečujski, i V. Delić. A novel split-and-merge algorithm for hierarchical clustering of gaussian mixture models. *Applied Intelligence*, vol. 37, no. 3, str. 377–389, 2012. doi: 10.1007/s10489-011-0333-9. URL http://link.springer.com/article/10.1007/s10489-011-0333-9.

A. Rosti i M. Gales. Factor analysed hidden markov models for speech recognition. *Computer Speech & Language*, vol. 18, no. 2, str. 181–200, 2004. URL http://www.sciencedirect.com/science/article/pii/S0885230803000536.

G. Saon i J.-T. Chien. Large-vocabulary continuous speech recognition systems: A look at some recent advances. *Signal Processing Magazine, IEEE*, vol. 29, no. 6, str. 18–33, 2012. ISSN 1053-5888. doi: 10.1109/MSP.2012.2197156.

K. Saul, Lawrence i G. Rahim, Mazin. Modeling acoustic correlations by factor analysis. U *Neural Information Processing Systems*, 1997.

L. K. Saul i M. G. Rahim. Maximum likelihood and minimum classification error factor analysis for automatic speech recognition. *Speech and Audio Processing, IEEE Transactions on*, vol. 8, no. 2, str. 115–125, 2000. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=824696.

S. Seo i K. Obermayer. Soft learning vector quantization. *Neural computation*, vol. 15, no. 7, str. 1589–1604, 2003. URL http://www.mitpressjournals.org/doi/abs/10.1162/089976603321891819.

D. Seung i L. Lee. Algorithms for non-negative matrix factorization. *Advances in neural information processing systems*, vol. 13, str. 556–562, 2001.

S. Theodoridis i K. Koutroumbas. *Pattern Recognition, Third Edition*. Academic Press, 3 edition, 2006. ISBN 0123695317.

V. Vanhoucke i A. Sankar. Mixtures of inverse covariances. *Speech and Audio Processing, IEEE Transactions on*, vol. 12, no. 3, str. 250 – 264, may 2004. ISSN 1063-6676. doi: 10.1109/TSA.2004.825675.

M. Varjokallio i M. Kurimo. Comparison of subspace methods for gaussian mixture models in speech recognition. U *INTERSPEECH*, str. 2121–2124, 2007.

M. Westphal. Tc-star recognition baseline results. Technical report, IBM, 2004. URL http://www.tcstar.org/documents/deliverable/deliverable_updated14april05/D6.pdf.

S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. A. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, i P. Woodland. *The HTK book*. Cambridge University Engineering Department, 3.4 edition, March 2009.

S. J. Young, J. Odell, i P. C. Woodland. Tree-based state tying for high accuracy acoustic modelling. U *Proceedings of the workshop on Human Language Technology*, str. 307–312. Association for Computational Linguistics, 1994. URL http://dl.acm.org/citation.cfm?id=1075885.

W. Zhu i D. O'Shaughnessy. Log-energy dynamic range normalizaton for robust speech recognition. U *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, vol. 1, str. 245–248, 2005. doi: 10.1109/ICASSP.2005.1415096.