Modeling and robust estimation of a series of point sources in multidimensional imaging

Gribl Koščević, Anita

University of Zagreb

FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING

Anita Gribl Koščević

MODELING AND ROBUST ESTIMATION OF A SERIES OF POINT SOURCES IN MULTIDIMENSIONAL IMAGING

DOCTORAL THESIS

Zagreb, 2023

University of Zagreb

FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING

Anita Gribl Koščević

MODELING AND ROBUST ESTIMATION OF A SERIES OF POINT SOURCES IN MULTIDIMENSIONAL IMAGING

DOCTORAL THESIS

Supervisor: Professor Davor Petrinović, PhD

Zagreb, 2023

Sveučilište u Zagrebu
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

Anita Gribl Koščević

MODELIRANJE I ROBUSNA ESTIMACIJA NIZA TOČKASTIH IZVORA U VIŠEDIMENZIONALNOM OSLIKAVANJU

DOKTORSKI RAD

Mentor: Prof. dr. sc. Davor Petrinović

Zagreb, 2023.

This doctoral thesis was completed at the University of Zagreb Faculty of Electrical Engineering and Computing, on Department of Electronic Systems and Information Processing.

Supervisor: Professor Davor Petrinović, PhD

The thesis has 158 pages.

Thesis No.: _____

About the Supervisor

Davor Petrinović was born on Feb. 24, 1965 in Zagreb. He graduated, received Master of Science degree and doctoral degree from the Faculty of Electrical Engineering and Computing (FEEC), University of Zagreb, in 1988, 1996 and 1999 respectively.

From Oct. 1988 Davor Petrinović has been employed at the FEEC. In 2016, he was elected to the academic title of Full Professor. He was a Fulbright Postdoctoral Scholar at University of California, Santa Barbara, USA for one year in 2000/01, and a Visiting Researcher at School of Electrical Engineering, KTH, Stockholm, for one year in 2005/06. From 2010/2011 to 2013/2014 he was serving as Vice dean for academic affairs of FEEC and from 2014/2015 Head of the University Undergraduate Study Programme in Military Engineering. Professor Davor Petrinović's scientific interests are focused on speech and signal processing, in particular in the area of speech and audio modeling, processing and coding. He published more than 40 papers in journals and conference proceedings. He led two scientific projects funded by the Ministry of Science, Education and Sports of the Republic of Croatia, and participated as a collaborator in four scientific projects, one technological project, one information technology implementation project, and served as a coordinator of one HKO-ESF project. He participated in two international scientific projects during his stay in the USA and in Sweden, on the EU FP7 project ACROSS, and is a member of the Scientific center of excellence for Data science and cooperative systems within the Datacross project.

O mentoru

Davor Petrinović je rođen 24.02.1965. u Zagrebu. Diplomirao je 1988. godine, magistrirao 1996. godine, a doktorirao 1999. na Sveučilištu u Zagrebu, Fakultetu elektrotehnike i računarstva (FER).

Od listopada 1988. godine zaposlen je na FER-u. 2016. godine je izabran za redovitog profesora u trajnom zvanju. Kao Fulbrightov postdoktorski stipendist proveo je jednu godinu (2000/2001.) na University of California, Santa Barbara, USA. Boravio je i godinu dana (2005/2006.) na School of Electrical Engineering, KTH, Stockholm, kao gostujući istraživač. U periodu od akademske godine 2010/2011. do 2013/2014. obavlja dužnost prodekana za nastavu na FER-u, a od 2014./2015. voditelja sveučilišnog preddiplomskog studija Vojno inženjerstvo. Područje istraživanja Davora Petrinovića je fokusirano na obradu govora i signala, posebice u području audio modeliranja i modeliranja, obrade i kodiranja govora. Autor je više od 40 znanstvenih radova objavljenih u časopisima ili zbornicima radova znanstvenih skupova. Vodio je dva znanstvena projekta financirana od strane Ministarstva znanosti, obrazovanja i športa Republike Hrvatske, a sudjelovao je kao suradnik na četiri znanstvena projekta, jednom

tehnološkom projektu, jednom projektu primjene informacijskih tehnologija, te kao koordinator jednog HKO-ESF projekta. Sudjelovao je na dva međunarodna znanstvena projekta za vrijeme boravka u SAD i u Švedskoj, na EU FP7 projektu ACROSS, te je član znanstvenog centra izvrsnosti za Znanost o podatcima i kooperativne sustave u sklopu projekta Datacross.

U akademskoj godini 2013/2014 bio je član Vijeća tehničkog područja, Sveučilišta u Zagrebu, a aktivan je član niza državnih i sveučilišnih povjerenstava i radnih skupina vezanih uz upravljanje kvalitetom. Urednik je područja za obradu signala i računalne inteligencije časopisa Automatika – Journal for Control, Measurement, Electronics, Computing and Communications. Davor Petrinović bio je član programskog odbora pet međunarodnih konferencija.

Preface

I would like to express my sincere gratitude to my thesis supervisor Prof. Davor Petrinović for his guidance and continuous support during my doctoral studies, as well as for his help and contribution to the scientific papers included in this thesis. He is an outstanding mentor to whom I am truly grateful for the effort and time spent with me, teaching me and solving problems together.

A special thanks goes to all great doctoral and postdoctoral researchers I met and had the privilege to work with, share tough and joyful times.

I also wish to express my special thanks to my parents, brothers, family and friends for their support and encouragement during my doctoral studies and years of education.

Most importantly, thanks to my husband Karlo and my son Roko, who are my greatest motivation and endless support.

Thank you!

Abstract

This research encompasses modeling and robust parameter estimation of a series of point light sources in multidimensional imaging. In 2D imaging, the point sources are blurred and extended in the image plane due to diffraction, optical aberrations, distortion, and other degradations that occur during the imaging. According to the central limit theorem, the joint action of all these degradations can be approximated by 2D Gaussian profiles. The first part of the research deals with detecting stellar objects in astronomical images through precise estimation of 2D Gaussian profile parameters where the main issues are overlapping adjacent objects and noise. The weighted least squares method with Huber weights is proposed to precisely estimate the overlapping Gaussian profiles even above the resolution limit. The method is based on robust statistics of error dispersion and does not require prior knowledge of the profile's parameters or noise statistics. The method's robustness is analyzed for different noise sources, and the optimal width of the estimation input domain is derived using differential entropy.

In the second part of the research, the estimation problem of Gaussian profile parameters is generalized to the multidimensional case. Parameters of the profile can be estimated either in the least squares or maximum likelihood sense. Independently of a chosen approach, the estimation of optimal parameters requires iterative techniques for nonlinear optimization, which depend on the initial guess for parameters and can get stuck in local minima. To circumvent the drawbacks of the conventionally used iterative methods and to speed up and simplify the estimation procedure, a two-step method for parameter estimation of the multidimensional Gaussian profile in the domain of the exponential function argument is developed. The method alternatively combines analytical and iterative approaches in parameter estimation to find the optimal solution in the least squares sense. In the 2D case, the theory of resultants is applied to calculate all 2D Gaussian profile parameters analytically.

The third research part deals with the image reconstruction of an unknown 2D process in low-dose positron emission tomography (PET). Reduction of the radiotracer dose causes the reduction of PET input data, so-called lines of response (LORs), a sparse sinogram, and low spatial and temporal image reconstruction. Thus, a novel approach for 2D PET data acquisition is proposed, which uses the intersections of LORs to boost the sinogram and enable reconstruction of the underlying 2D process. The underlying 2D image is reconstructed from analytical relations between the spatial statistical properties of intersections of LORs and properties of the input process.

Keywords: astronomical imaging, modeling and detection of stellar objects, multidimensional Gaussian profile, parameter estimation, iteratively reweighted least squares method, robust statistics, low-dose 2D PET imaging, 2D PET image reconstruction

Prošireni sažetak

Modeliranje i robusna estimacija niza točkastih izvora u višedimenzionalnom oslikavanju

Istraživanje obuhvaća modeliranje i robusnu procjenu parametara niza točkastih izvora svjetlosti iz snimljenih slika. Slike točkastih objekata trebali bi biti rijetki signali s vrijednostima različitim od nule, koji odgovaraju Diracovim Delta impulsima, samo na očekivanim pozicijama projekcija 3D točkastih izvora svjetlosti u 2D slici. Međutim, slike točkastih izvora svjetlosti su zamućene i proširene u ravnini slike. Čak i u slučaju idealnog optičkog sustava ograničenog samo difrakcijom svjetlosti na leći, slike točkastih objekata prate oblik Airy diska. Osim difrakcije, postoje mnoge druge degradacije koje se javljaju pri oslikavanju kao što su optičke aberacije, distorzije, vignetting, neželjeni pomaci sustava oslikavanja, električna svojstva i šum senzora te spektralni odziv optike i senzora. Različite degradacije mogu se opisati pojedinačnim funkcijama širenja točkastog izvora čije se zajedničko djelovanje opisuje konvolucijom u prostornoj domeni, odnosno umnoškom njihovih spektralnih odziva u frekvencijskoj domeni. Ukupna konvolucija pojedinačnih funkcija širenja se prema centralnom graničnom teoremu može aproksimirati 2D Gaussovim profilom. Također, u umnošku spektralnih odziva pojedinačnih funkcija širenja, 2D Gaussova funkcija najčešće određuje prostornu graničnu frekvenciju rezolucije optičkog sustava. Stoga je precizna procjena parametara Gaussovih profila iznimno važna za identifikaciju i praćenje takvih točkastih izvora svjetlosti. Primjene se mogu pronaći u astronomskim slikama za detekciju zvjezdanih objekata ili u mikroskopiji i medicinskom oslikavanju za praćenje pojedinačnih čestica.

Prvi dio istraživanja bavi se procjenom parametara višedimenzionalnog Gaussovog profila iz zašumljenih mjerenja. Estimaciju parametara 2D Gaussovih profila u astronomskim slikama otežavaju šum i susjedni objekti koji se preklapaju. Glavni problem kod preklopljenih susjednih profila je određivanje pripadnosti piksela pojedinom objektu. Uobičajeno korištene metode za prostornu segmentaciju pretpostavljaju postojanje lokalnog minimuma između preklopljenih profila i prestaju raditi kad bliski točkasti izvori prelaze granicu razlučivosti definiranoj Sparrowim kriterijem. Nekoliko je pristupa u estimaciji parametara višedimenzionalnog Gaussovog profila iz zašumljenih ulaznih podataka ovisno o zahtijevanoj preciznosti, o prethodnom znanju o parametrima i statistici šuma. Dva su pristupa estimaciji s obzirom na domenu estimacije: estimacija u domeni vrijednosti i estimacija u domeni argumenta eksponencijalne funkcije. Najjednostavnija analitička metoda u domeni vrijednosti je metoda momenata gdje prvi moment daje očekivanu poziciju središta profila, dok drugi moment daje očekivane varijance profila. Nedostaci ove metode su osjetljivost na šum i odabranu ulaznu estimacijsku regiju s pristranošću podestimiranju varijanci profila, ali je zbog brzine prikladna za rad u stvarnom vremenu. Druga grupa metoda za estimaciju u domeni vrijednosti su iterativne metode koje pronalaze

optimalne parametre ili u smislu najmanjih kvadrata ili u smislu najveće izglednosti. Oba pristupa zahtijevaju rješavanje prezadanog sustava nelinearnih jednadžbi koristeći iterativne optimizacijske tehnike. Uobičajeno korištena metoda najmanjih kvadrata računalno je skupa, ne garantira konvergenciju, a rezultat optimizacije je jako ovisan o odabiru početnih parametara. S druge strane, traženje optimalnih parametara u smislu najveće izglednosti zahtjeva veliki broj mjerenja i prethodno znanje o distribuciji šuma, ali daje nepristrane procjene minimalne varijance. Drugi pristup estimaciji parametara Gaussovog profila je estimacija u domeni argumenta eksponencijalne funkcije logaritmiranjem ulaznih mjerenja i modela. U slučaju 1D Gaussovog profila, estimacija parametara profila u log domeni se svodi na prilagođavanje parabole ulaznim mjerenjima. Međutim, u slučaju višedimenzionalnog profila, broj parametara kvadratno raste, pa posljedično, i složenost estimacije. Štoviše, za više dimenzije parametri profila su čak i u domeni argumenta nelinearno vezani. Prvi dio istraživanja rješava problem estimacije preklopljenih 2D Gaussovih profila kako bi se omogućila identifikacije slabijih i gusto raspoređenih zvjezdanih objekata u astronomskim slikama. U nastavku istraživanja rješava se poopćeni problem estimacije parametara višedimenzionalnog Gaussovog profila s ciljem pronalaska analitičkog rješenja za parametre, te ubrzanjem i pojednostavljenjem postupka estimacije.

Drugi dio istraživanja bavi se rekonstrukcijom nepoznatog 2D procesa u pozitronskoj emisijskoj tomografiji (PET) iz smanjene doze radioaktivnog sredstva. Takav se proces sastoji od točkastih izvora koji su povučeni iz neke nepoznate 2D distribucije čiji se parametri pokušavaju estimirati iz smanjenog broja mjerenja uz analizu moguće primjene rezultata prethodno navedenih istraživanja. Kako bi smanjili rizike od izloženosti zračenju tijekom PET snimanja, mnogi istraživači pokušavaju rekonstruirati sliku nepoznatog 2D procesa koristeći najmanju moguću dozu radioaktivne tvari i pritom postići istu kvalitetu rekonstruirane slike kao i sa standardnom dozom. Glavni izazov kod PET snimanja s niskim dozama je nizak omjer signala i šuma (SNR), i posljedično, niska kvaliteta rekonstruiranih slika. Za postizanje kvalitete slika standardne doze unosom niže doze kod PET-a potrebno je poboljšati ili performanse mjernog sklopovlja ili tehnike obrade signala. Performanse mjerne instrumentacije mogu se poboljšati povećanjem osjetljivosti PET senzora čime se povećava postotak detektiranih događaja i smanjuje prigušenje i raspršenje. S druge strane, postoje mnoga programska rješenja koja se temelje na otklanjanju šuma iz slike prije ili poslije rekonstrukcije ili tehnike koje kombiniraju ulazne podatke dobivene iz niskih doza PET-a s ulaznim podacima iz drugih modaliteta snimanja kao što su CT ili MRI. U novije vrijeme postaju sve popularnije metode koje koriste strojno učenje i metode dubokog učenja koje koriste slike dobivene s niskim i standardnim dozama za treniranje modela koji mogu predvidjeti slike standardnih doza iz slika niskih doza. Međutim, računalna složenost takvih metoda je znatno veća od složenosti uobičajeno korištene analitičke metode filtrirane unazadne projekcije.

Izvorni znanstveni doprinosi disertacije su

- •Robusna metoda za estimaciju parametara preklopljenih Gaussovih profila uz prisustvo šuma
- •Metoda za estimaciju parametara višedimenzionalnog Gaussovog profila u domeni eksponencijalne funkcije
- •Analiti čki opis odnosa prostorne statistike ulaznog procesa i virtualnih sjecišta linija odziva u pozitronskoj emisijskoj tomografiji

Ova se disertacija sastoji od šest poglavlja.

U prvom poglavlju daje se uvod u problematiku te se opisuje struktura rada.

U drugom poglavlju opisuje se generiranje sintetičke astronomske slike koja uključuje sve očekivane degradacije tipičnog sustava oslikavanja, a čije se zajedničko djelovanje aproksimira 2D Gaussovim profilima. Inicijalni se parametri profila estimiraju koristeći konvencionalnu proceduru za obradu niza astronomskih slika. Sintetička slika je dobivena projekcijom točkastih objekata iz zvjezdanih kataloga i njihovom konvolucijom s 2D Gaussovim profilima s inicijalno estimiranim parametrima. Eksperimentalni rezultati pokazuju da je tako dobivena sintetička slika bliska stvarnoj, no potrebna je daljnja manipulacija parametrima primjenom objektivnih mjera sličnosti sa stvarnim slikama kako bi se dobili precizniji rezultati. Takva sintetička slika predstavlja temeljnu referentnu sliku koja se može koristiti za daljnja istraživanja pod uvjetom da su parametri sustava precizno određeni.

U trećem se poglavlju opisuje metoda za preciznu estimaciju preklopljenih 2D Gaussovih profila kako bi se riješio problem identifikacije gusto raspoređenih točkastih izvora u astronomskim slikama. Predložena metoda je iterativna težinska metoda najmanjih kvadrata (IRWLS) za koju nije potrebno definirati ni broj, a niti središte preklopljenih komponenti. Metoda redom estimira i uklanja pojedine komponente na temelju lokalnih ekstrema, a zasniva se na robusnoj statistici rezidualne pogreške te u nekim slučajevima radi čak i iznad granice razlučivosti. Huberove težine su korištene za razdvajanje valjanih uzoraka od uzoraka koji pripadaju susjednim objektima. Takve se težine ažuriraju u svakoj iteraciji na temelju robusne analize rezidualne pogreške iz prethodne iteracije. Robusne metrike u analizi pogreške koriste se u definiranju mekog praga koji razdvaja valjane uzorke od iznimaka. Na valjane uzorke se primjenjuje L_2 norma, dok se na iznimke primjenjuje L_1 norma. Točnost metode i estimiranih parametara verificira se numerički koristeći Monte Carlo simulaciju, a rezultati nadmašuju konvencionalnu metodu najmanjih kvadrata u estimaciji preklopljenih objekata na prethodno sintetički dobivenim slikama. Pokazuje se da metoda postiže do 25 dB veću točnost u usporedbi s konvencionalnom metodom najmanjih kvadrata za određeni skup razlika u magnitudama profila i udaljenostima njihovih središta. Štoviše, 80% ukupne točnosti metoda postiže u samo pet iteracija. U nastavku se analizira utjecaj odabranog praga za razdvajanje valjanih uzoraka i iznimaka na točnost estimacije. Analizira se robusnost predložene IRWLS metode s Huberovim težinama na aditivni, multiplikativni Poissonov šum i šum s teškim repovima koji se susreću u

astronomskim slikama. Rezultati metode najmanjih kvadrata u estimaciji jednog zašumljenog 2D Gaussovog profila uspoređuju se s predloženom IRWLS metodom za sve tri distribucije šuma i različite omjere signala i šuma. Kao mjera točnosti koristi se ukupna pogreška modeliranja u decibelima definirana omjerom sume kvadrata razlike estimiranog i stvarnog modela i sume kvadrata stvarnog modela za sve uzorke koji su unutar odabrane evaluacijske regije. Odgovarajućim odabirom praga koji se koristi u generiranju Huberovih težina postiže se visoka asimptotska učinkovitost IRWLS metode čak i u slučaju aditivnog Gaussovog šuma za koji je rješenje najmanjih kvadrata optimalno. Za slučaj Poissonovog šuma, IRWLS metoda s težinama jednakim recipročnim varijancama uzoraka ima do 3 dB veći dobitak modeliranja u usporedbi s IRWLS metodom s Huberovim težinama i metodom najmanjih kvadrata budući da su takve težine optimalne u smislu maksimalne izglednosti za ovu vrstu šuma. Budući da se točnost estimacije čak i uobičajene metode najmanjih kvadrata mijenja ovisno o odabranoj širini ulazne estimacijske domene oko pretpostavljenog maksimuma profila, analitički se izvodi optimalna širina koja maksimizira točnost metode analizom informativnosti ulaznih podataka. Optimalno rješenje kao i predikcija smanjenja točnosti estimacije za druge neoptimalne širine se verificira Monte Carlo simulacijama.

U četvrtom se poglavlju opisuje predložena metoda za estimaciju parametara višedimenzionalnog Gaussovog profila u domeni argumenta eksponencijalne funkcije. Predložena se metoda sastoji od dva koraka gdje se iterativno i naizmjenično estimiraju parametri matrice kovarijance i linearni faktor skale za danu poziciju središta profila, a zatim se središte profila ažurira koristeći estimiranu matricu kovarijance iz prethodnog koraka. Glavna prednost metode je analitičko rješenje parametara matrice kovarijance i linearne skale za početno zadano središte profila proizvoljne dimenzije. Točnost metode u estimaciji parametara iz podataka kontaminiranih aditivnim šumom analizira se numerički Monte Carlo simulacijama gdje se kao mjera točnosti koristi ukupna pogreška modeliranja, a rezultati se uspoređuju s konvencionalnom metodom najmanjih kvadrata, metodom momenata i metodom najveće izglednosti u domeni vrijednosti. Za slučaj 2D Gaussovog profila, predlaže se metoda za analitičku estimaciju pozicije centroida za danu matricu kovarijance profila primjenom teorije rezultanti za rješavanje sustava dva nelinearna polinoma trećeg reda u dvije varijable. Kod estimacije višedimenzionalnog Gaussovog profila u domeni argumenta eksponencijalne funkcije, eksperimentalni rezultati pokazuju da u slučaju visokih SNR-ova predložena metoda postiže istu točnost kao i metoda najmanjih kvadrata u domeni vrijednosti Gaussovog profila. Ipak, predložena metoda u logaritamskoj domeni konvergira brže, posebno za Gaussove profile većih dimenzija. Za mnoge praktične primjene gdje je početno središte već dovoljno točno određeno, dovoljna je samo jedna iteracija predloženog algoritma za određivanje svih preostalih parametara modela korištenjem analitičkog postupka bez potrebe za daljnjim iteracijama. Eksperimenti su pokazali da su u slučaju estimacije 3D Gaussovog profila iz podataka zašumljenih aditivnim

Gaussovim šumom, matrica kovarijance i linearna skala procijenjene jedan ili čak dva reda veličine brže korištenjem analitičkog rješenja predložene metode nego korištenjem iterativne metode najmanjih kvadrata u domeni vrijednosti. U isto vrijeme, za visoke SNR-ove ($\geq$ 40dB) predložena metoda postiže sumjerljivu ukupnu pogrešku modeliranja bez prethodnog znanja o parametrima ili statistici šuma, pod uvjetom da je pozicija središta poznata. Štoviše, eksperimenti su pokazali da predložena metoda konvergira u najviše četiri iteracije, dok postupci numeričke optimizacije u domeni vrijednosti zahtijevaju znatno veći broj iteracija od predložene metode za pronalaženje svih parametara modela. Zbog svoje brzine, metoda se može koristiti u aplikacijama u stvarnom vremenu. Ipak, metoda najmanjih kvadrata u domeni vrijednosti točnija je od predložene metode u slučaju niskih SNR-ova ($\leq$ 20 dB) i širokih estimacijskih regija. Gubitak preciznosti predložene metode dominantno je uzrokovan malim vrijednostima profila na tako širokim regijama koji mogu unijeti značajne pogreške u log domenu jer aproksimacija eksponencijalne funkcije s prvim članom Taylorovog niza nije dovoljna.

Peto poglavlje opisuje predloženu metodu za rekonstrukciju 2D PET slike iz smanjene doze ubrizgane radioaktivne tvari. Kod 2D PET-a ulazni podaci su koincidentni događaji koji se generiraju u uzorcima nepoznatog 2D procesa, a putuju prema detektorima u suprotnim smjerovima duž tzv. linija odziva. Njihov je broj proporcionalan aktivnosti ubrizgane radioaktivne tvari, a te se linije odziva koriste u rekonstrukciji nepoznatog 2D procesa, dok je pozicija uzorka procesa koji je generirao pojedinu liniju u stvarnosti nepoznata. Za točnu rekonstrukciju većina dosadašnjih metoda zahtijeva izuzetno veliki broj detektiranih koincidentnih događaja što znači i veliku dozu ubrizgane radioaktivne tvari za praćenje. U ovom se istraživanju na drugi način pristupa procesu prikupljanja podataka kako bi se smanjila potreba za količinom ubrizgane radioaktivne tvari koristeći sjecišta linija odziva čiji je broj proporcionalan kvadratu broja uzoraka procesa, tj. kvadratu linija odziva. Pri razvoju metode koristi se činjenica da su prostorna statistička svojstva sjecišta linija odziva povezana s prostornim statističkim svojstvima nepoznatog 2D procesa koji se estimira. Virtualne linije odziva se generiraju pod slučajnim kutevima na mjestima sjecišta te se koriste u generiranju sinograma i u rekonstrukciji konvencionalnom metodom filtrirane unazadne projekcije (FBP algoritam) kako bi se postigla veća prostorna i vremenska razlučivost rekonstrukcije. U eksperimentima se provodi Monte Carlo simulacija gdje se generiraju i uspoređuju rekonstrukcije iz sinograma dobivenih iz malog broja stvarnih linija odziva, rekonstrukcije dobivene iz idealnog sinograma te rekonstrukcije dobivene iz virtualnih linija odziva. Eksperimentalni rezultati pokazuju da se sjecišta linija odziva mogu uspješno koristiti za rekonstrukciju nepoznatog 2D procesa čak i u slučaju vrlo malog broja uzoraka procesa koji je nedovoljan za predstavljanje te nepoznate 2D distribucije. Iako je na prvi pogled rekonstruirana slika daleko od stvarnog procesa, usrednjavanjem samo 50 takvih rekonstrukcija dobiva se stabilna slika male varijance reda veličine 10^{-7}. Rekonstruirana slika primarno ima neželjenu pristranost u odnosu na izvorni proces zbog razlike u matricama kovarijanci procesa

i sjecišta. Ova bi se pristranost u odnosu na ulaznu funkciju gustoće mogla kompenzirati nelinearnom dekonvolucijom poznavanjem analitičkog odnosa prostornih distribucija procesa i sjecišta. Stoga je u nastavku istraživanja funkcija gustoće vjerojatnosti (PDF) sjecišta linija odziva izvedena za par točkastih izvora u prostornoj domeni i domeni sinograma. Pokušalo se izvesti očekivanje PDF-a sjecišta za sve parove točkastih izvora iz nepoznatog 2D ulaznog procesa s ciljem primjene algoritama dekonvolucije ili maksimizacije očekivanja za rekonstrukciju nepoznatog procesa. Uvodi se regularizacija početnih kuteva između linija odziva koje se sijeku kako bi se ograničilo raspršenje sjecišta i omogućila rekonstrukcija ulaznog 2D procesa. Regularizacija osigurava konačne prve i druge momente distribucije sjecišta, omogućujući također izračun srednje vrijednosti i matrice kovarijance ulaznog procesa. U slučaju numerički izračunatog ukupnog PDF-a sjecišta osnovnog 2D Gaussovog procesa, distribucija njegovih sjecišta je aproksimirana mješavinom generalizirane normalne distribucije bliske Laplaceovoj distribuciji i normalne distribucije osnovnog procesa.

U šestom zaključnom poglavlju navedeni su glavni rezultati istraživanja te su dane smjernice za daljnja istraživanja.

Keywords: astronomske slike, modeliranje i detekcija zvjezdanih objekata, višedimenzionalni Gaussov profil, estimacija parametara, iterativna težinska metoda najmanjih kvadrata, robusna statistika, 2D PET oslikavanje s niskom dozom radioaktivnog sredstva, rekonstrukcija slike kod 2D PET-a

Contents

Chapter 1

Introduction

Modeling and robust estimation of point light sources are essential in many image processing fields where images are sparse signals of point sources, such as astronomical imaging, microscopy, and medical imaging. Even in an ideal diffraction-limited imaging system, the image of a point light source is an Airy disk. Apart from diffraction, many optical aberrations, distortions, and other image degradations have to be modeled and identified for precise point source detection and tracking. The first part of the research deals with detecting overlapping stellar components from astronomical images. In the second part of the research, the method for generalized multidimensional Gaussian profile estimation in the least squares sense is proposed, which transforms the estimation problem to the argument domain of the exponential function. The third part of the research deals with image reconstruction in low-dose 2D PET imaging. Instead of counting the so-called lines of response (LORs), the intersections of these LORs are used to increase the amount of the input PET data. A detailed analysis of the statistical properties of intersections of LORs is performed, and they are related to the statistical properties of the underlying 2D process. The aim is to reconstruct the underlying 2D process from the reduced amount of radiotracer by employing the quadratically larger number of intersections than the number of LORs.

This thesis is organized as follows. The second chapter describes the main properties of astrophotography and gives a few examples of star catalogs that can be used for synthetic astronomical image generation. Linear camera calibration, which represents the projection of the stellar object from a star catalog to the image plane, and the nonlinear conventional lens distortion model are given. The contributions of individual image degradations are modeled by their own point spread functions, which are convolved to obtain the system point spread function, approximated by the 2D Gaussian profile. The conventional astronomical stack processing pipeline for the estimation of initial 2D Gaussian model parameters of stellar objects is explained. Finally, such estimated initial model parameters are used for synthetic image sequence generations, and the synthesis results are compared with captured images for the same

part of the sky.

The third chapter considers the issue of overlapping adjacent object detection in astronomical images. Conventional segmentation methods fail if overlapping objects exceed the resolution criteria. In the case of overlapping 2D Gaussian components, the Sparrow resolution criterion is derived. The iteratively reweighted least squares method is used to solve overlapping components and precisely estimate their parameters. Thereby, robust statistics of residual errors and Huber weights are applied. The matching pursuit algorithm based on the method of local extrema is used for iterative estimation and removal of 2D Gaussian components to reduce residual energy in the image. The robustness of the method to the different noise types is considered. Also, the optimal width of the estimation input domain is derived by introducing the concept of differential entropy.

The fourth chapter deals with parameter estimation of the multidimensional Gaussian profile, where the estimation problem is transformed in the exponential argument domain by taking the logarithm of both input measurements and the model. The objective function that is minimized is the sum of weighted squared residuals with weights selected so that simultaneously the sum of squared residuals in the domain of values is minimized as well. The two-step method is proposed where in the first step, the covariance matrix and linear scale factor are analytically estimated for a given initial centroid by solving the system of linear equations, while in the second step, the centroid is updated through nonlinear optimization given all other Gaussian profile's parameters for arbitrary dimension. For the 2D case, the theory of resultants is used for analytical updating of the centroid as well.

In the fifth chapter, low-dose 2D PET imaging is considered. The method for data acquisition from a reduced number of coincident events using intersections of LORs is proposed. The probability density function (PDF) of intersections of LORs for a pair of point sources is derived in the spatial domain and the sinogram domain. Statistical properties of intersections are related to the statistical properties of the underlying 2D process, which is reconstructed. There was an attempt to derive the expectation of PDF of intersections for all point sources from the arbitrarily underlying process aiming to apply the deconvolution and expectation maximization algorithms for unknown process reconstruction with the discussion of main issues. The blur of intersections in the case of the rotationally symmetric underlying 2D Gaussian process is approximated by the mixture of generalized normal distribution and the distribution of the process.

Finally, the conclusion of overall results is provided, as well as directions for future research.

Chapter 2

Astrophotography

An astronomical image, also called astrophotography, is an image of the night sky that contains astronomical objects such as stars, nebulae, and galaxies. Astronomical objects are usually distant and faint point light sources that emit light in different spectral bands. Therefore, a long exposure time is characteristic for astrophotography to collect as much incident light as possible from such distant objects. Also, specialized optical filters are used in astronomical imaging to pass only specific wavelengths of light through, thus enabling the detection of objects emitting even outside the visible light spectrum.

Stellar objects are point light sources, i.e., Dirac Delta impulses in their spatial domain, but extended and blurred in the image plane due to different degradations that occur during the imaging such as diffraction of light, optical aberrations, distortions, an unwanted residual motion of the imaging system during exposure, sensor properties and noise. Every optical system has many such degradations that need to be identified and compensated for. The identification and compensation of image degradations are usually done as part of complex calibration procedures. Still, it can also be directly done from a single night sky image in which all those degradations are reflected [1]. Astrophotography is ideal for this since it consists of point sources of known positions, spectra, and intensities found in publicly available star catalogs.

Star catalogs contain information on stellar objects' angular positions (declination δ, and right ascension α) in the equatorial coordinate system and their magnitudes in different spectral bands. As a relative measure of the brightness of stars observed from the Earth, the apparent magnitude * is usually used. The most famous star catalog database is VizieR † which contains almost 20000 catalogs that differ by wavelength regimes, missions, and astronomy target objects. The accuracy and completeness of each catalog are described by the faintest apparent magnitude (the highest magnitude number) contained in the catalog and the accuracy of objects' positions. Some examples of star catalogs are the Bright Star Catalogue, the Thycho-2

*unit increase of apparent magnitude implies the reduction in brightness by $\sqrt[5]{100} \approx 2.512$

†VizieR database: `https://vizier.cds.unistra.fr/viz-bin/VizieR`

Catalogue, and the Guide Star Catalogue. The Bright Star Catalogue [2, 3] contains around 9000 brightest stars that can be seen with the naked eye from the Earth. An example of a more extensive catalog is the Tycho-2 Catalogue [4] which contains more than 2.5 million of the brightest stars in the Milky Way along with their positions, proper motions, and magnitudes in two different spectral bands. Finally, the Guide Star Catalog [5, 6] is related to the Hubble Space Telescope, where version GSC-I contains about 20 million stars with an apparent magnitude of 6 up to 15. The GSC-II version has almost a billion stars up to magnitude 21.

Astrophotography is characterized by the geometry of the camera and lens, which are determined as a part of the camera calibration procedure, long exposure time, various optical aberrations, motions of the imaging system during exposure, the lens vignetting, photon sensitivity of the sensor, sensor gain (ISO) and noise, color filter array in front of the sensor, and finally by the portion of the sky enclosed in the frame.

2.1 Imaging system modeling

A precise mathematical model of the imaging system and identification of its parameters enables the generation of synthetic astronomical images. Such a synthetic image should match the image captured by the camera for the same field of view (FOV). Furthermore, such a synthetic image represents the ground truth image or a high-resolution image of the same part of the sky since it can contain a more significant number of extracted and projected stellar objects than the captured image and can be sampled over the grid of the arbitrary density, but with the precondition of the accurate imaging system model. The astronomical image model can be described as

$$\mathbf{y} = \mathbf{WCx} + \mathbf{n}, \tag{2.1}$$

where $\mathbf{y}$ represents the column vector of captured image values, $\mathbf{W}$ is the measurement matrix that models degradations caused by motion effects and anti-aliasing filter, $\mathbf{C}$ denotes the convolution matrix which models blur caused by optics, $\mathbf{x}$ is a sparse vector of point sources' positions, and $\mathbf{n}$ is the vector of noise.

2.1.1 Camera calibration

Identification of imaging system parameters is usually part of the complex camera calibration procedure. Camera calibration is a technique of determining the geometric characteristics of the camera that describes the transformation of the 3D scene from the real world into the camera coordinate system and its projection into the 2D image plane [1, 7, 8]. Camera calibration includes determining the intrinsic and extrinsic camera parameters that are part of the linear camera model and the lens distortion parameters that are part of the nonlinear camera model.

The linear camera calibration model is shown in Fig. 2.1. The extrinsic parameters include

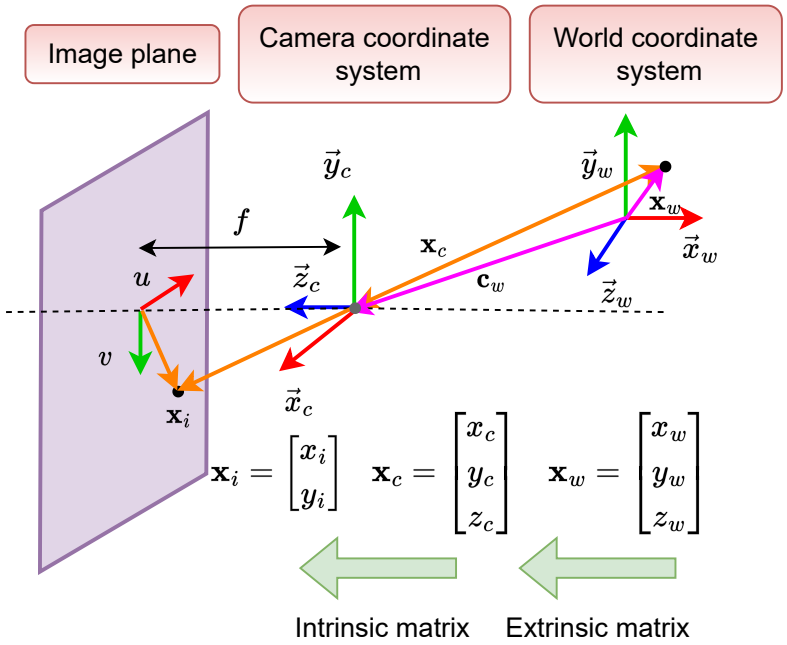


Figure 2.1: Linear camera calibration model

3D rotation and translation of the point from the global (world) coordinate system to the local camera's coordinate system. The point from the word coordinate system $\mathbf{x}_w = (x_w, y_w, z_w)$ is transformed to the camera coordinate system as follows

$$\mathbf{x}_c = \mathbf{R}(\mathbf{x}_w - \mathbf{c}_w) = \mathbf{R}\mathbf{x}_w - \mathbf{R}\mathbf{c}_w = \mathbf{R}\mathbf{x}_w + \mathbf{t}, \tag{2.2}$$

where $\mathbf{R}_{3x3}$ is a rotation matrix that is calculated by the multiplication of fundamental rotation matrices around all three axes (x,y,z) using Euler's angles. This same transformation of the point source from the world to the camera coordinates in homogeneous coordinates is

$$\tilde{\mathbf{x}}_c = \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{3x3} & \mathbf{t}_{3x1} \\ \mathbf{0}_{1x3} & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = \mathbf{M}_{ext}\tilde{\mathbf{x}}_w, \tag{2.3}$$

where $\mathbf{M}_{ext}$ represents the extrinsic matrix. Intrinsic parameters include the camera's focal length, skew coefficient, pixel widths, and principal point that determine the projection of the 3D point from the local camera's coordinate system into the 2D image plane. Perspective projection of the point from the local camera's coordinate system $\mathbf{x}_c = (x_c, y_c)$ to the point in the image

plane $\mathbf{x}_i = (x_i, y_i)$ is obtained from the similarity of triangles as follows

$$\frac{x_i}{f} = \frac{x_c}{z_c} \rightarrow x_i = f\frac{x_c}{z_c}, \tag{2.4}$$

$$\frac{y_i}{f} = \frac{y_c}{z_c} \rightarrow y_i = f\frac{y_c}{z_c}, \tag{2.5}$$

where f is the focal length. The calculated position of the point $\mathbf{x}_i = (x_i, y_i)$ in the image plane is in millimeters. To obtain the corresponding point position in pixels, the pixel densities in both x and y directions, m_x and m_y, have to be taken into account, as well as the position of principal point (o_x, o_y) where the optical axis pierces the sensor. The projected position of the point in pixels is

$$u = m_x x_i = m_x f\frac{x_c}{z_c} + o_x = f_x\frac{x_c}{z_c} + o_x, \tag{2.6}$$

$$v = m_y y_i = m_y f\frac{y_c}{z_c} + o_y = f_y\frac{y_c}{z_c} + o_y, \tag{2.7}$$

where f_x and f_y are focal lengths in pixels in both directions. In homogeneous coordinates, the projected point in the image plane is calculated as

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} = \begin{bmatrix} z_c u \\ z_c v \\ z_c \end{bmatrix} = \begin{bmatrix} f_x & 0 & o_x & 0 \\ 0 & f_y & o_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \mathbf{M}_{int}\tilde{\mathbf{x}}_c. \tag{2.8}$$

The complete linear camera calibration model has the following form

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & o_x & 0 \\ 0 & f_y & o_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R}_{3x3} & \mathbf{t}_{3x1} \\ \mathbf{0}_{1x3} & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = \mathbf{M}_{int}\mathbf{M}_{ext}\tilde{\mathbf{x}}_w. \tag{2.9}$$

Identification of the camera calibration matrix is usually done by using some calibration object. Still, it can also be done from a single night sky image by pairing the positions of stars found in captured images with their positions extracted from the publicly available star catalogs for the same selected field of view (FOV) [1]. Fig. 2.2 shows the linear calibration model used in astronomical images. The position of infinitely far stars in celestial coordinates (α,δ) are

extracted from star catalogs and transformed to the 3D positions on the unit sphere through the following equations:

$$x_w = \cos(\alpha)\cos(\delta), \tag{2.10}$$

$$y_w = \sin(\alpha)\cos(\delta), \tag{2.11}$$

$$z_w = \sin(\delta). \tag{2.12}$$

The 3D positions of stars on the unit sphere are 3D rotated using the Euler rotation matrix to align the world coordinate system with the camera coordinate system. The rotated 3D positions in the camera coordinate system are used to calculate the new right ascension and declination angles (α,δ), which are used to project the point from the camera coordinate system to the image plane. Thereby, the tangential (gnomonic projection) [9] is applied, and the coordinates of the point source in the image plane are

$$u = -f_x \tan(\alpha), \tag{2.13}$$

$$v = f_y \tan(\delta). \tag{2.14}$$

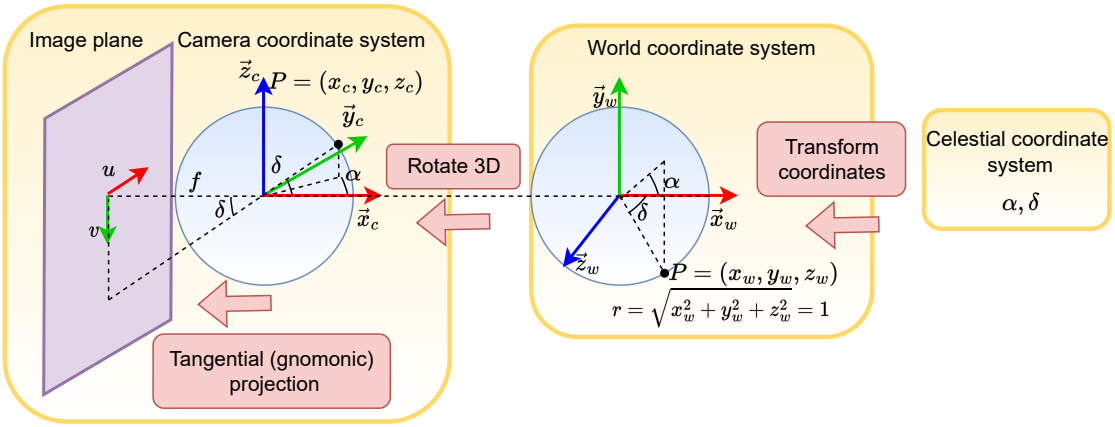


Figure 2.2: Camera calibration model for astronomical imaging

2.1.2 Lens distortions

Lens distortion effects cause the nonlinear displacements of the projected point in the image plane due to the different lens properties such as curvature, decentering, and tilt. The conventionally used nonlinear lens distortion model includes radial, tangential, and thin prism distortions [10, 11]. Distroted point position in the image plane (u_d, v_d) is calculated as

$$u_d = u + \delta_u(u, v), \tag{2.15}$$

$$v_d = v + \delta_v(u, v). \tag{2.16}$$

According to [12], the radial distortion is a consequence of radial lens curvature and depends on the radial distance r of the point (u,v) from the center of radial distortion as

$$\delta_{ur}(u,v) = u(k_1 r^2 + k_2 r^4), \tag{2.17}$$

$$\delta_{vr}(u,v) = v(k_1 r^2 + k_2 r^4). \tag{2.18}$$

Decentering distortion comes from decentering of the lens or other optical components [11] and is described as

$$\delta_{ud}(u,v) = p_1(3u^2 + v^2) + 2p_2 uv, \tag{2.19}$$

$$\delta_{vd}(u,v) = p_2(3v^2 + u^2) + 2p_1 uv, \tag{2.20}$$

where p_1 and p_2 are decentering distortion coefficients. The thin prism distortion is caused by the tilt of a lens concerning the image sensor array [11] and can be written as

$$\delta_{up}(u,v) = s_1(u^2 + v^2), \tag{2.21}$$

$$\delta_{vp}(u,v) = s_2(u^2 + v^2), \tag{2.22}$$

where s_1 and s_2 are thin prism distortion coefficients. The total lens distortion is the sum of individual distortion effects

$$\delta_u(u,v) = \delta_{ur}(u,v) + \delta_{ud}(u,v) + \delta_{up}(u,v), \tag{2.23}$$

$$\delta_v(u,v) = \delta_{vr}(u,v) + \delta_{vd}(u,v) + \delta_{vp}(u,v). \tag{2.24}$$

Even though these equations describe the physical behavior of a single lens, modern optical systems and telescopes contain multiple lenses and mirrors. Hence, the total distortion model requires more complex complete polynomial models in two variables (u and v) to describe such complex optical systems.

2.1.3 Point spread function

The imaging system model should include all physical effects that cause the spread of the point source. Each physical effect that causes the blur has its individual point spread function (PSF). The impulse response of the whole imaging system to a point source corresponds to a convolution of individual PSFs in the spatial domain or multiplication of their spectra in the frequency domain, but only if they are shift-invariant. The Fourier transform of the system PSF is a complex-valued function referred to as the optical transfer function (OTF). Its magnitude is referred to as the modulation transfer function (MTF), while its phase is referred to as the phase

transfer function (PTF). MTF describes the ability of the optical system to transfer, i.e., to capture an object's contrast at some spatial frequency, and thus is a good measure of the optical system's quality. The parameters of the system PSF can be estimated from the stack of captured astronomical images, thus enabling cheap camera calibration.

Diffraction PSF

Diffraction of the light on the lens aperture produces the diffraction PSF, which represents the upper-resolution limit of any imaging system. Diffraction PSF is shift-invariant, while its shape and size depend on the lens aperture's size and shape. In the case of a circular lens aperture, diffraction PSF has a form of an Airy disk (pattern) [13]. The minimum angular displacement of two objects before they blur, which is defined as a displacement when the maximum of the Airy pattern of the first object hits the first minimum of the Airy pattern of the second object, is the Rayleigh resolution criterion and can be described as

$$\sin(\theta) = 1.22\frac{\lambda}{d}, \qquad \frac{x}{f} = 1.22\frac{\lambda}{d}, \qquad x = 1.22\frac{f\lambda}{d}, \tag{2.25}$$

were θ is the angular displacement, x is the distance of two objects' images, f is the focal length, λ is a wavelength of the light, d is a diameter of lens aperture, while the ratio f/d represents the lens f-number. The normalized intensity function of an Airy pattern is a square of radially symmetric Bessel function of the first kind J_1, which is a function of radial distance r as

$$f_{Airy}(r) = \left(\frac{2J_1(r)}{r}\right)^2. \tag{2.26}$$

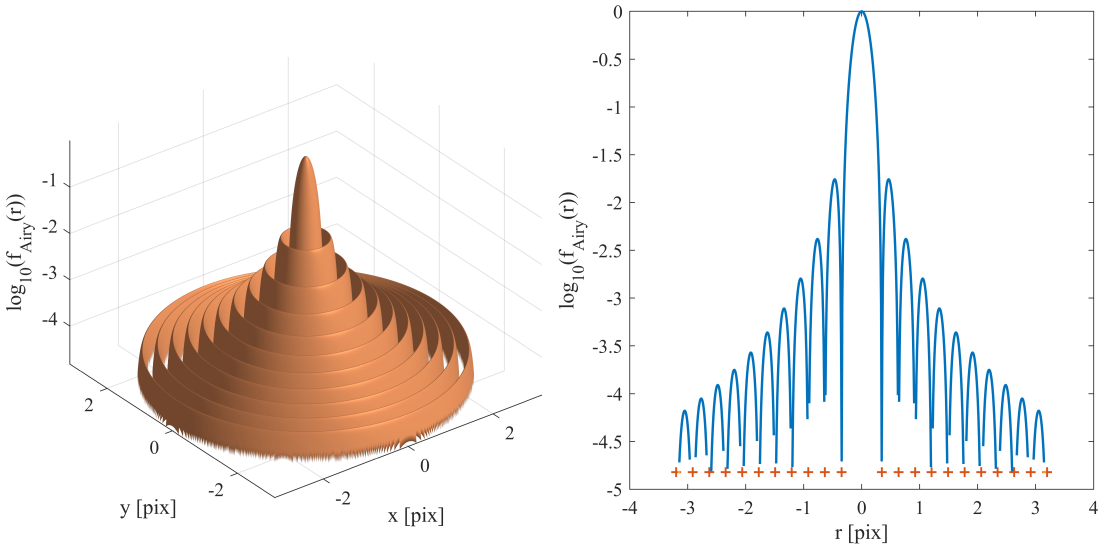
An example of the Airy pattern is shown in Fig. 2.3. MTF of the Airy pattern is the Chinese hat function of the form

$$\text{chat}(f) = \frac{2}{\pi}(\cos^{-1}(f) - f\sqrt{1-f^2}) \tag{2.27}$$

where $0 \leq f \leq 1$ is normalized spatial frequency.

Bivariate Gaussian PSF

2D Gaussian distribution coarsely approximates all other image degradations according to the central limit theorem [14] such as optical monochromatic aberrations (defocus, coma, and astigmatism) and longitudinal chromatic aberrations, motion blur, electrical cross-talk in the sensor, etc. Consequently, the blur of extracted and projected point sources in the image plane can be approximated by the 2D Gaussian profiles. The parameters of such profiles vary across the image frame so they should be separately estimated for each part of the frame and for each of three color channels. In real situations, the Gaussian MTF will determine the system cut-off frequency with its narrower shape in the frequency domain compared to the other MTFs.



(a) Logarithm of 3D Airy pattern

(b) 2D slice of 3D Airy pattern

Figure 2.3: Airy pattern with the parameters: $N_{rings} = 10, f$-number $= 2.8,$ Pixel Pitch $= 6.5\mu m, \lambda = 650nm$

Optical low pass filter

The optical low-pass filter (OLPF) is the anti-aliasing filter that prevents the Moiré effect, which occurs in the spatial domain at high frequencies. The brick-wall filter is an ideal low-pass filter that completely blocks the signal above the Nyquist spatial frequency and passes through the whole signal below it. However, it produces undesirable ringing at the Nyquist frequency. As an alternative, a beam splitter is usually used in digital cameras, which splits the light spot into four spots at the distance of d in both directions, thus blurring of initial spot and reducing the spatial frequency. The cutoff frequency where its MTF hits the first zero depends on the displacement d. Such a filter also attenuates the useful signal below the cutoff frequency, so the displacement d is typically chosen such that the cutoff frequency is above the Nyquist one.

Box function

The box function of the pixel width integrates spatially continuous light field into the pixel intensity. The implementation of this convolution can be numerically efficiently performed by accumulation and differentiation of the over-sampled light field representation (2D CIC filter). The MTF of the box function is sinc^2 function as Fig. 2.5 shows.

System PSF

Fig. 2.4 shows the observation model of astronomical images. The point sources are first extracted from star catalogs and projected into the image plane. Their positions are nonlinearly transformed due to the properties of the lens and the projection of the 3D sphere to the 2D

plane. Point sources in the image plane are convolved with individual PSFs. The convolution of all individual PSFs yields the system PSF of the imaging system. Finally, the additive and multiplicative noise is added, and such a continuous light field is sampled to the selected regular grid. Fig. 2.5 shows individual MTFs and their combination. The multiplication of all MTFs

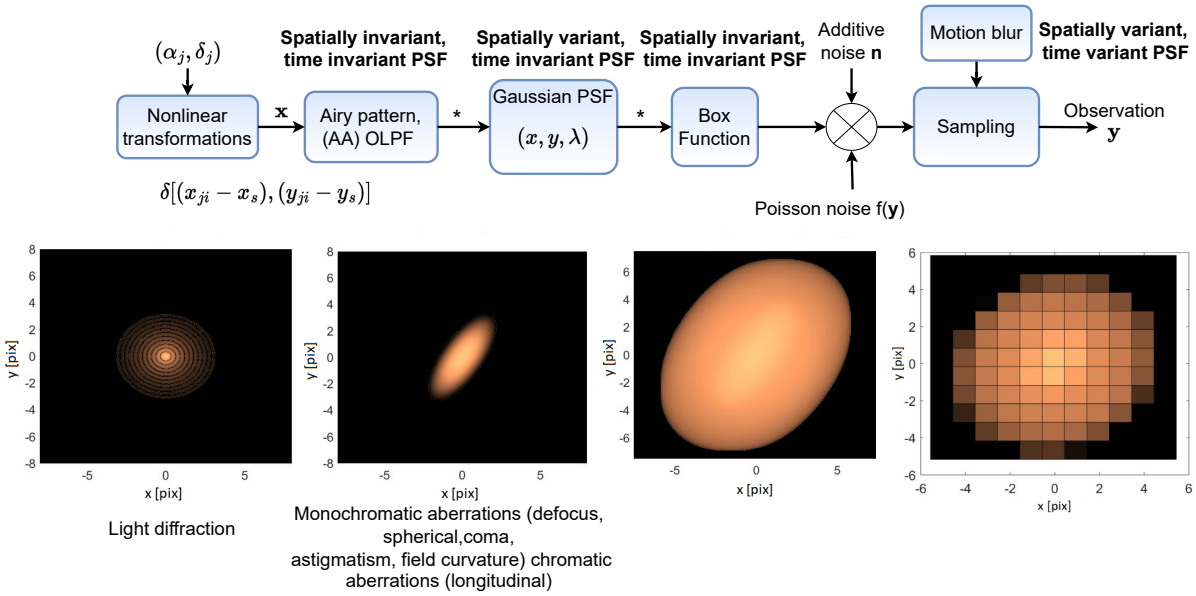


Figure 2.4: The observation model

in the frequency domain represents the system MTF. Even if OLPF primarily determines the cut-off frequency in this example, the Gaussian MTF will usually specify that frequency in real situations.

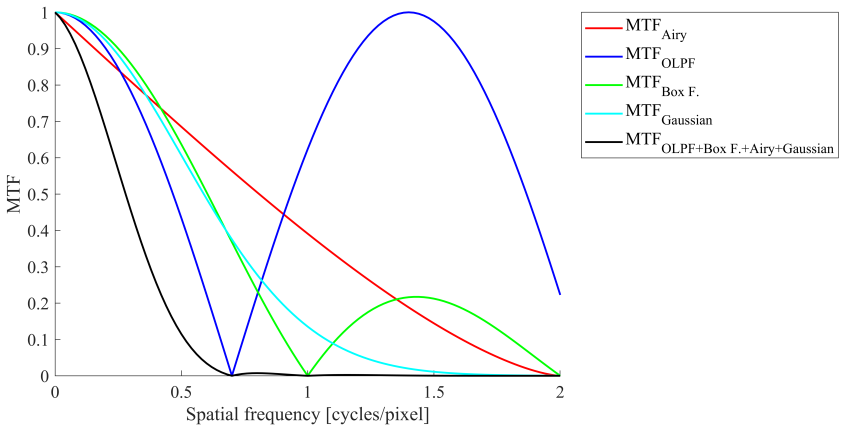


Figure 2.5: MTFs of Airy pattern, box function, OLPF, Gaussian and their combination

The deconvolution of such blurred astronomical images has to be performed to identify the stellar objects and their exact positions. Deconvolution of blurred images is also a part of the usual image restoration procedures. Some deconvolution algorithms used in astrophotography, such as Lucy-Richardson algorithm [15], assume that the composite spread function, i.e., the system PSF, is known in advance. Furthermore, some super-resolution algorithms which recon-

struct high-resolution images from more captured low-resolution images also assume a precise estimate of the system PSF. The system PSF is usually estimated as a part of the imaging system calibration procedure, where each contributing factor of the point spread is modeled separately. However, the exact mathematical model of the system PSF, which includes all possible sources of the blur, is extremely complicated even for an objective with a single lens [13, 16] while in astrophotography, more complex imaging systems are usually used. Also, the parameters of the system PSF are spatially shift-variant, thus meaning that they gradually vary across the image frame due to anisotropic lens behavior. Namely, the blurred stellar objects have a circular shape in the center of the frame, while they are more extended and of elliptic and elongated shapes at the frame edges. Some PSFs are also time-variant, such as blur caused by the motion or atmospheric dispersion. Instead of mathematical modeling and estimating each source of the spread individually, a better approach is estimating a system PSF directly from the captured astronomic images, thus enabling the simple, cheap, and fast calibration of the whole imaging system.

According to the central limit theorem [17], the 2D Gaussian profile can approximate the system PSF, and this assumption is used in this research. Each blurred point source in the image plane can be described by its 2D Gaussian profile. The parameters of the 2D Gaussian profiles vary smoothly and gradually across the image frame, so the spatially localized averaging can be performed to obtain the anisotropic 2D Gaussian model of the system PSF for the whole image frame. The deviation of estimated 2D Gaussian parameters of a particular object from the averaged estimated parameters in the same part of the frame is an indicator that such object is an outlier or some other source of light such as galaxies, nebulae, but also measurement errors (e.g., faulty pixels) or noise (e.g., salt and pepper noise). Hence, such deviations in estimated parameters can be used for outlier detection during stellar object identification. However, the essential prerequisite for precise stellar object detection is an accurate estimate of the composite PSF model for each point source contained in the image.

2.1.4 Image registration

Image registration includes mapping the star catalog to the referent image plane and mapping each captured image to the selected referent one by annotating the coincident stellar objects between each image and the referent one and between the referent image and the projected star catalog, as Fig. 2.6 shows. In each captured image, stellar objects are detected using the method of local extrema, thus yielding the list of N brightest stars. Also, for the selected FOV, the expected stellar objects are extracted from the chosen star catalog. Then, the plate-solving step is applied to find the coincident stellar pairs between the captured image and the catalog by applying the triangle similarity method. The method is based on the fact that relative relations of triangles among stars are preserved independently of linear transformations (scaling,

translating, and rotating). For each of N brightest stars, feature vectors are formed with all combinations of two among seven neighbor stars. The triangle similarity method searches for matching feature vector pairs that infer similar global linear transformation parameters, i.e., linear camera calibration parameters (extrinsic and intrinsic parameters - three Euler's rotation angles and the gnomonic projection parameters (celestial angle pairs (α, δ), and sensor focal lengths f_x, f_y) in the least mean squares sense). The second step is determining the nonlinear 2D polynomial model, which can describe nonlinear distortion at the frame edges, resulting from lens properties and projection of the 3D sphere to the 2D image plane. Lateral chromatic aberrations can also be considered by determining the nonlinear mapping parameters for each color channel separately. Namely, lateral aberrations cause spatially distant projections of the same star position in the image plane and different mapping errors for each of the three color channels.

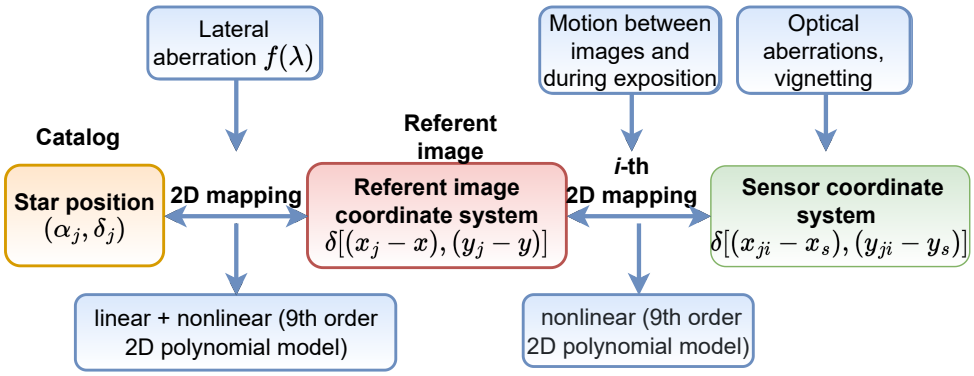


Figure 2.6: Geometric transformation and image registration

2.1.5 Motion estimation

A long exposure, object motions, and unwanted residual motions of the imaging system during exposure are the causes of motion blur. Since stellar objects are static, the motion blur in astronomical images is primarily caused by the residual tracking error of the imaging system (e.g., declination drift, right ascension periodic and non-periodic error). Each pixel's motion trajectory through images is constructed in the image registration step by the cubic spline interpolation (i.e., from the identified 2D mapping model between the referent and i-th image) from which the motion blur is estimated.

2.1.6 Devignetting

Vignetting manifests as a decrease in image intensity from the center to the edges of the image frame. Under homogeneous sky illumination, vignetting is usually modeled with a radially symmetric smooth 2D function, such as the sum of cosines to the fourth power with the initial

center at the center of the image frame. The parameters of the assumed vignetting model are fitted to the median pixel intensity values of the pre-processed image stack for different radial distances from the center of the image frame. The smoothness of the vignetting model can be ensured by keeping only low spatial frequency coefficients. The 2D polynomial model of the 4th order can be estimated and cascaded to the radial model to account for the non-radial vignetting effects. An inverse function is applied to the input images to correct them, smoothing their background.

2.1.7 Denoising

Long exposure time, typical for astronomical images, causes the dark current noise. The dark current frame can be captured once during sensor calibration and subtracted from each captured raw image in post-processing. Alternatively, it can be captured after each captured light frame and immediately removed from it. Since the dark current rises with the exposure time and temperature of the sensor, the latter approach is better for tracking dark current variabilities. Still, it reduces the effective exposure time by half. The quantum nature of the light causes the photon shot noise, which follows Poisson distribution. Poisson noise is multiplicative signal-dependent noise with the variance proportional to the pixel's intensity [18]. One approach for Poisson noise removal is to use Anscombe transformation [19] to transform it into additive Gaussian noise and then remove it. In the second approach, after devignetting, the spatial 2D model of Poisson noise standard deviation is estimated from the image stack. Then the image stack is divided by that model. Such obtained final image stack is slightly attenuated at the edges but has a constant value of noise standard deviation across the whole frame [20].

2.1.8 Astronomical image stack processing pipeline

The main steps of the conventional astronomical image stack processing pipeline for stellar objects identification are

1. Annotation of N brightest stars in each captured image using plate solving and identification of geometric transformation between each captured image and the chosen star catalog
2. Temporal motion vector estimation for each bright star from calculated geometrical transformations (primarily caused by residual imaging system motions - declination drift and right ascension periodic and non-periodic errors)
3. Image registration - 2D mapping of each input image to the chosen referent one by using coincident star pairs found in plate solving step
4. Initial stacking of all captured images that results in the reduction of additive noise

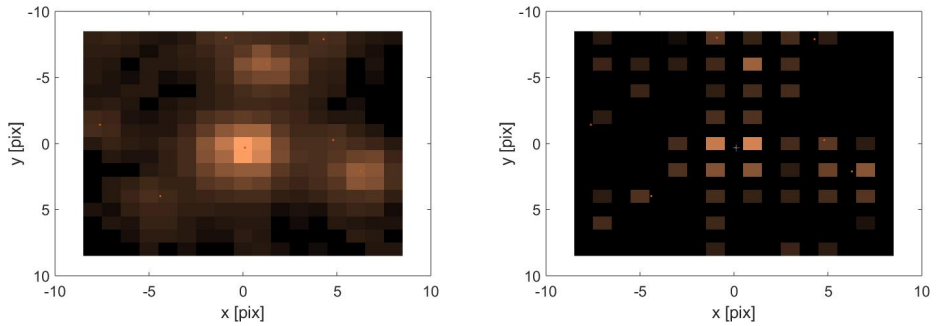
5.Estimation of vignetting and 2D model of Poisson noise standard deviation from an initial image stack
6.Applying the inverse vignetting model to all input frames
7.Application of inverse variance weighting to all input frames to remove Poisson noise
8.New stacking of flattened input frames and flat background removal
9.Separate spatial registration of each of three color channels of the final stack to compensate lateral chromatic aberrations
10.Estimation of spatially smooth temporal noise model and application of its inverse to the image stack
11.Identification of geometrical transformation between the image stack and the chosen catalog (linear and nonlinear 2D mapping)
12.PSF parameter estimation of all expected point sources in the processed image stack, thus yielding spatially variant PSF model for the whole frame
13.Deconvolution of the image stack using estimated spatially variant smooth PSF model

2.1.9 Synthetic astronomical image generation

The results of the proposed approach for synthetic astronomical image sequence generation are given in [20]. Table 2.1 shows the project information about camera parameters and full-stack information. The chosen star catalogs for extraction of expected object positions were the HR, TYCHO-2, and GSC 2.2 catalogs. The selected target object, tracked during the imaging and approximately positioned in the center of the frame, was NGC7635 (also known as Bubble Nebula). Positions of expected stellar objects were extracted from the HR and TYCHO-2 catalogs for the same sky region and were used in initial plate solving. The synthetic image was generated only for a narrow region around the TYC4279-1488-1 target star. Positions of all point sources in this region were then extracted from the GSC 2.2 catalog with six arc minute radius (658 objects), which are in the magnitude range $R_{mag} < 15$ and $B_{jmag} < 16$. The estimated system PSF was placed at the fractional position of the point source on the interpolated sensor grid. Fig. 2.7a shows the blue channel of the processed stack image previously demosaiced in the stack processing pipeline. Fig. 2.7b shows only blue pixels of one input raw image. Fig. 2.7c shows the synthesized continuous light field on the interpolated upsampled grid, and finally, Fig. 2.7d shows the sampled synthesized light field in the sensor image plane. The results demonstrate that the proposed method generates a good representation of the actual image using only the previously estimated PSF parameters from the image stack separately for each of the three color channels. However, more precise synthesis can be achieved if the model parameters are additionally finely tuned to the parameters of the actual image.

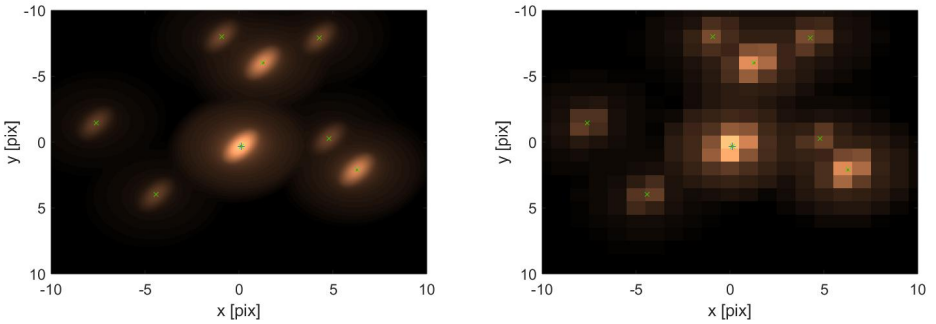
Table 2.1: Project parameters

Target Name	NGC7635
Exposure Time (EXIF)	15.0 sec
f-number (f/Av)	1.4
ISO Speed	800
Camera focal Length	105.0 mm
Sensor Resolution	5505 x 3670 pix
Sensor pixel pitch	6.50 μm
Sensor crop factor	1.01 (35 mm equivalent)
Camera focal Length	105.5 mm
Expected horizontal angle of view	19°20.9’
Expected vertical angle of view	12°59.1’
Expected diagonal angle of view	23°10.1’
Minimum required catalog radius	695.0 min
Image Resolution	5472 x 3648 pix
Expected image scale	12.775 arcsec/pix
Reference star Catalog radius	765.0 min
Number of reference stars in Catalog	62327
Expected number of reference extracted radial segment Catalog stars in frame	30665
N_{star} number of extracted stars from frame for plate solving	1000
Catalog star magnitude	< 15
Number of good light frames in stack	177



(a) Demosaiced processed stack image (blue channel)

(b) One selected input raw image of only blue channel from a series of input images



(c) Synthesised continuous light field on up-sampled grid

(d) Synthesised image sampled on sensor grid

Figure 2.7: Extracted region around TYC4279-1488-1 - Blue channel

Chapter 3

Identification of overlapping stellar components

Precise parameter estimation of 2D Gaussian profiles enables the identification of stellar objects in astronomical images whose PSFs are approximated by such profiles. Conventional parameter estimation procedures seek optimal profile parameters that minimize the error between the empirical input data and the model in either a least-squares (LS) [21] or maximum-likelihood (ML) [22] sense. Overlapping components and noisy measurements are the main issues in the parameter estimation of 2D Gaussian profiles in astronomical images. The overlapping of two or more stellar components makes it difficult to determine the belonging of a pixel to a particular object. For this purpose, spatial image segmentation methods are usually used. However, they typically assume the existence of local minima between overlapping components. An example of a spatial segmentation method is the Voronoi tessellation, but it fails when the overlapping components are closer than the distance defined by the Sparrow resolution limit [23, 24]. Multiple definitions of the resolution limit describe the resolving power of an optical system, such as the Rayleigh, Dawes, and Sparrow criteria. The Rayleigh and Dawes resolution criteria are related to diffraction-limited systems and assume the occurrence of minima of two close Airy disks. For example, the Rayleigh criterion is defined as the distance of centers of two close Airy discs when the first minimum of one Airy disc hits the central maximum of the second Airy disc. Since 2D Gaussian profiles are strictly positive functions, such criteria are not applicable to this type of function. However, the Sparrow resolution criterion is derived from the condition for the occurrence of an inflection point between two Airy discs, thus meaning that it can also be applied to overlapping 2D Gaussian profiles.

3.1 Sparrow resolution criterion for overlapping Gaussian profiles

The Sparrow resolution criterion is derived here for two overlapping 1D Gaussian profiles of the same width σ, but it can also be generalized to the multidimensional case under the assumption that the profiles are rotationally symmetric with the same semiaxes widths in all directions. The mean of the first profile is at the origin ($\mu_1 = 0$), while the mean of the second profile is displaced by $\mu_2 = d = k\sigma$. The component at the origin is the stronger component with a unit amplitude in front of the exponential, while the second component is the weaker of amplitude A. The sum of individual components equals

$$f(x) = \exp\left(\frac{-x^2}{2\sigma^2}\right) + A\exp\left(-\frac{(x-d)^2}{2\sigma^2}\right), A \leq 1. \tag{3.1}$$

By substituting $x = \sqrt{2}\sigma\xi$ and $d = k\sigma$, expression (3.1) becomes

$$f(\xi) = \exp(-\xi^2) + A\exp\left(-\left(\xi - \frac{k}{\sqrt{2}}\right)^2\right). \tag{3.2}$$

The inflection point occurs when the first and the second derivatives of (3.2) are simultaneously zero at the point $x = x_0$:

$$\frac{\delta f(x)}{\delta x}\Big|_{x=x_0} = \frac{\delta f(x)}{\delta \xi}\frac{\delta \xi}{\delta x} = 0, \tag{3.3}$$

$$\frac{\delta^2 f(x)}{\delta x^2}\Big|_{x=x_0} = \frac{\delta^2 f(x)}{\delta \xi^2}\frac{\delta^2 \xi}{\delta x^2} = 0. \tag{3.4}$$

The final expressions for the inflection point and the amplitude of the weaker component are

$$x_0 = \frac{\sigma}{2}\left(k + \sqrt{(k^2-4)}\right), A \leq 1, \tag{3.5}$$

$$A = \frac{\exp\left(\left(-\frac{x_0}{\sqrt{2}\sigma}\right)^2\right)}{\exp\left(-\left(\frac{x_0}{\sqrt{2}\sigma} - \frac{k}{\sqrt{2}}\right)^2\right)}\left(\frac{-\frac{x_0}{\sqrt{2}\sigma}}{\frac{x_0}{\sqrt{2}\sigma} - \frac{k}{\sqrt{2}}}\right). \tag{3.6}$$

The obtained expression implies that for two components displaced by $d = k\sigma$, the inflection point x_0 and the corresponding amplitude A of the weaker component are determined by the chosen value of factor k. The components of the same amplitude ($A = 1$) must be displaced at least by $d = 2\sigma$ with the inflection point at the midpoint ($x_0 = \sigma$). For $A < 1$, the inflection point is always closer to the weaker component, and the separation is always greater than $d = 2\sigma$, i.e., $k > 2$. The inflection point between two overlapping Gaussian profiles determines the resolution upper limit of the imaging system. Conventional methods based on the local minimum/maximum detection can not separate Gaussian components closer than the resolution

limit. Our motivation is to introduce a method that would show improvement even when close components exceed the derived resolution limit. Fig. 3.1 shows the family of curves for six combinations of center-to-center distances d/σ and ratios of their peak amplitudes $1/A$.

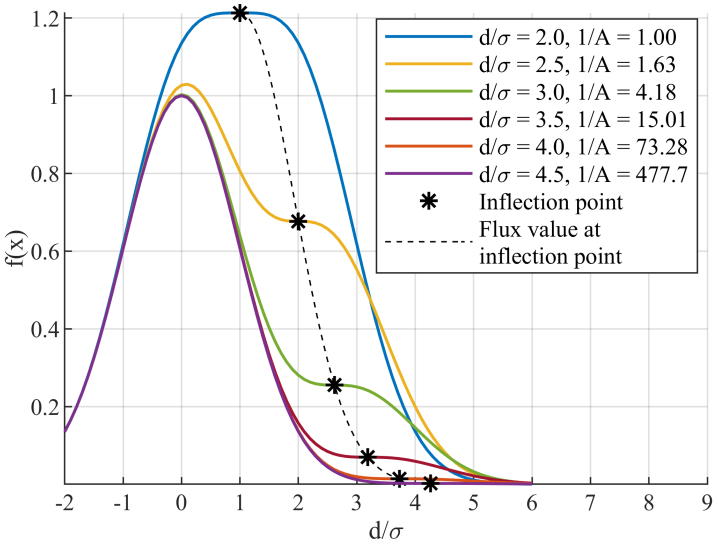


Figure 3.1: Sums of two 1D Gaussian profiles with parameters: $\sigma = 1$, $\mu_1 = 0$ and $\mu_2 = d$ for different center-to-center distances $k = d/\sigma$ and ratios of their peak amplitudes $1/A$.

3.2 Two-dimensional Gaussian profile

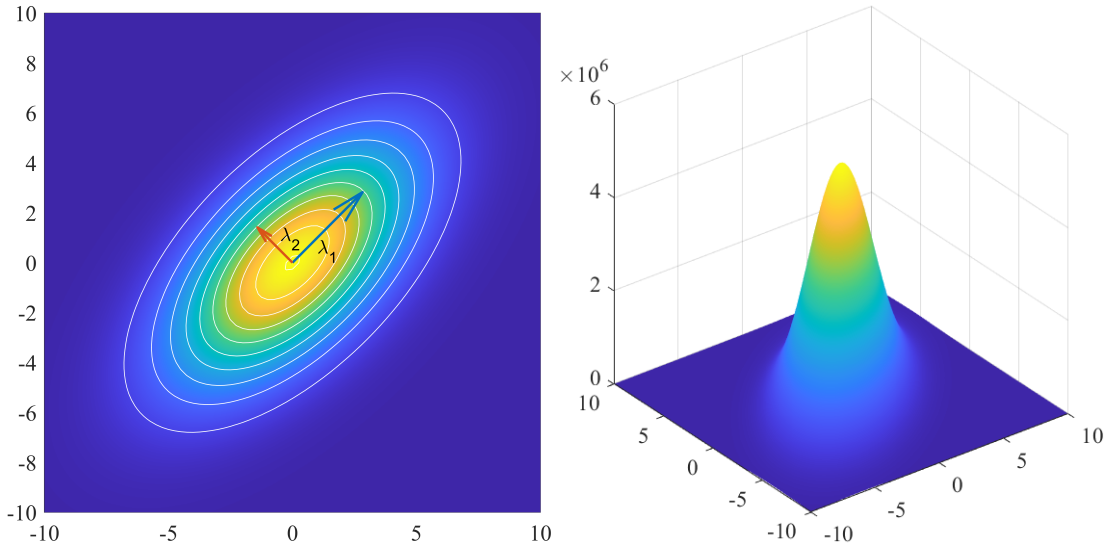
The mathematical form of the 2D Gaussian profile can be written as

$$f(\mathbf{x}_i, \mu, \Sigma, A_l) = \frac{A_l}{2\pi\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}\left((\mathbf{x}_i - \mu)^T \Sigma^{-1} (\mathbf{x}_i - \mu)\right)\right), \tag{3.7}$$

where $\mathbf{x}_i = [x_i, y_i]^T$ denotes the vector of ith of m input sample positions with corresponding readout value z_i. The profile's peak position consists of two terms $\mu = (\mu_x, \mu_y)$, while the covariance matrix has three unique terms $(\Sigma_{11}, \Sigma_{12} = \Sigma_{21}, \Sigma_{22})$. The covariance matrix of the 2D Gaussian profile in uncorrelated form can be calculated as a quadratic form of an orthonormal rotation matrix $\mathbf{R}$ and a diagonal matrix $\mathbf{S}$ of profile's variances, i.e., eigenvalues, as follows

$$\mathbf{R} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}, \mathbf{S} = \begin{bmatrix} \lambda_1^2 & 0 \\ 0 & \lambda_2^2 \end{bmatrix}, \Sigma = \mathbf{RSR}',$$

where λ_1 and λ_2 are the major and minor semiaxes widths, while θ corresponds to the rotation angle. An example of a 2D Gaussian profile is shown in Fig. 3.2.



(a) 2D view (b) 3D view

Figure 3.2: An example of the 2D Gaussian profile with the parameters: $\lambda_1 = 4$, $\lambda_2 = 2$, $\theta = \pi/4$, $\mu_x = \mu_y = 0$, $A = 10^5$

3.3 Astronomical image model and its estimation

The astronomical image, which consists of blurred point sources, can be considered as a sparse signal if it is represented as a linear combination of atomic normalized 2D Gaussian functions of different centroid positions, shapes, and sizes from an infinite parametric dictionary and which are multiplied by their projection coefficients as follows

$$I(x,y) \approx \sum_{i=1}^{N} \alpha_i f(x,y,\mu_i,\Sigma_i,1), \tag{3.8}$$

where α_i corresponds to the projection coefficient which represents the linear scale factor of the ith of N Gaussian atoms, and $f(x,y,\mu_i,\Sigma_i,1)$ denotes the normalized ith 2D Gaussian atom $(A_l = 1)$ which is completely determined by its covariance matrix (shape and size) Σ_i and the centroid position μ_i. An example of a synthetic astronomical image that consists of $N = 1000$ 2D Gaussian atoms of random positions and shape are shown in Fig. 3.3. This approach is similar to the matching pursuit algorithm [25, 26, 27, 28, 29, 30, 31], which, in each iteration, seeks the sparse approximation of the signal by successively selecting the best matching atom function which yields the largest inner product with the signal and thereby minimizes the residual error. Such atomic 2D Gaussian functions can be successively estimated and removed to reduce residual energy. In our approach, the method of local extrema is used to determine the profiles' estimation order. The found N local extrema are sorted in decreasing order by their values and successively estimated and removed from the input signal to minimize the residual energy. The first estimation step is selecting the estimation input domain for the local extremum

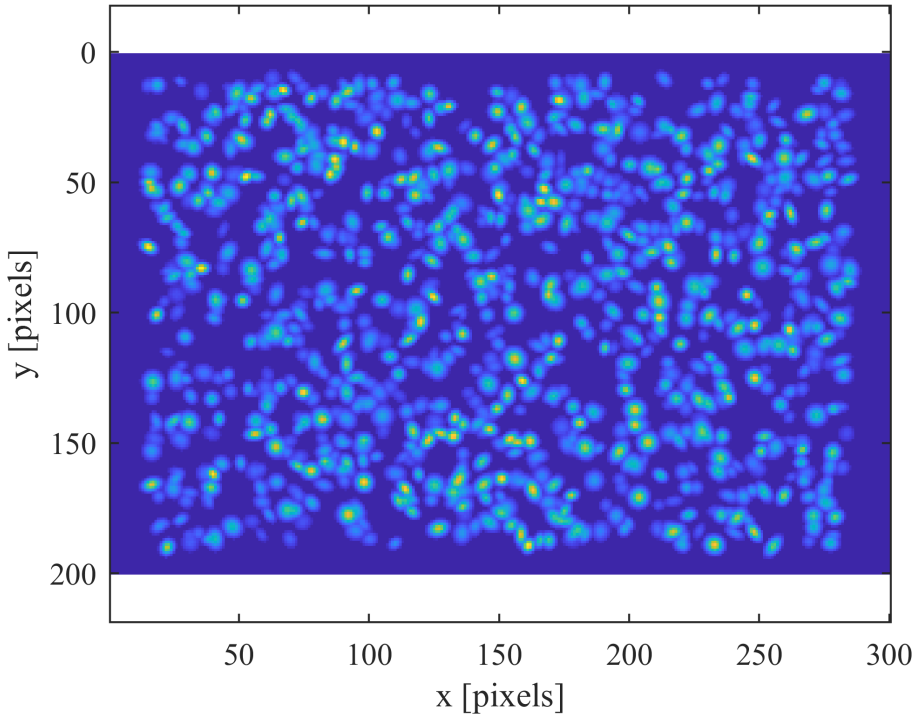
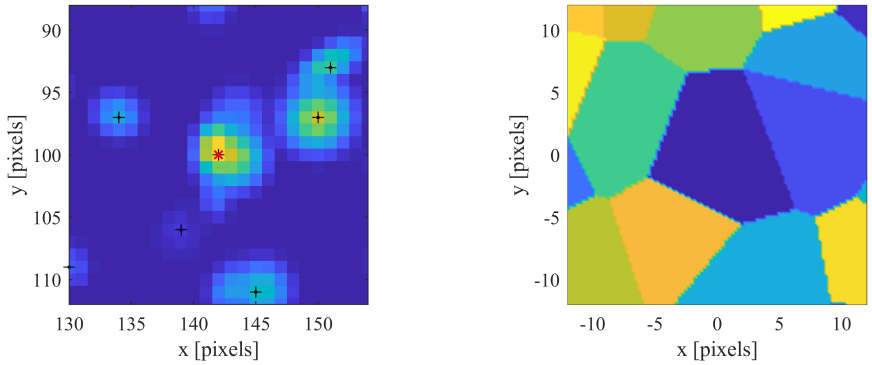


Figure 3.3: Synthetic image with 1000 random 2D Gaussian atoms (due to the high dynamic range, $\mathbf{z}^{1/4}$ is shown)

being estimated. Selection of estimation input domain, so-called region-of-interest (ROI), can be very hard if the stellar objects are densely distributed and overlapped since pixels close to the selected extremum can belong to adjacent objects or the background. The methods for spatial segmentation can help to find boundaries between objects, but only if they are displaced enough that the minimum occurs between them. The example of generalized linear tessellation around the 2D Gaussian profile, which is being estimated, is shown in Fig. 3.4. The boundaries between atoms are determined by their center-to-center distances and differences in their magnitudes.



(a) The region around object selected for estimation

(b) Tessellation around selected object

Figure 3.4: The example of generalized linear tessellation

Our approach is to use a convex input domain placed at the location of the found local

extremum of the radius $k\sigma$ where k represents the factor of Mahalanobis distance. Such selected convex ROI is circular for the rotationally symmetric 2D Gaussian profile with the same major and minor semiaxes widths. For other cases, such an ROI will be elliptical.

The parameters of the 2D Gaussian atoms and the corresponding projection coefficients, which are sequentially estimated and removed, must be accurately estimated to avoid the influence of incorrect atom estimation on subsequent iterations. The parameters of the 2D Gaussian atoms are calculated from the samples belonging to the selected convex ROI. The conventionally used LS method minimizes the sum of squared deviations of the model from the input measurements and obtains a solution for the profile parameters by nonlinear optimization. If the input ROI is wrongly chosen, i.e., if it includes samples belonging to overlapping neighboring profiles, the solution of the LS method gives a significant error because it insists on the simultaneous estimation of overlapping 2D Gaussian profiles with a single profile. After removing such a compromise 2D Gaussian atom, the residual image used in the following estimations is incorrect. It is necessary to introduce a new approach that would recognize samples of neighboring profiles as outliers and deweight them so that they do not affect the current profile estimation but would be used in subsequent iterations.

3.4 Estimation of overlapping 2D Gaussian profiles

Parameters of the 2D Gaussian atoms that need to be estimated from input empirical data and that are contained in the parametric dictionary are the centroid position (μ_x, μ_y), the double rotation angle $\rho = 2\theta, \rho \in [-\pi, \pi]$, the widths of major and minor semiaxes (λ_1, λ_2), and the projection coefficient which is represented by the apparent magnitude A_m. The double rotation angle is introduced as an auxiliary variable to ensure the continuity of parameters in optimization procedures. The vector of parameters β can be written as

$$\beta = (A_m, \lambda_1, \lambda_2, \mu_x, \mu_y, \rho) \tag{3.9}$$

The linear scale A_l of the 2D Gaussian profile in (3.7) represents the total sensed incident radiant flux, i.e., the intensity of the observed point light source invariant of the PSF size and shape. Various PSFs only cause a different spatial spread of the same flux. For astronomical imaging, the dynamic range of such filtered radiant flux is huge, so instead of linear representation, it is usually represented in the reverse logarithmic domain as an apparent magnitude A_m as $A_m = \log_{\sqrt[5]{100}} A_l$. For the example of two objects with apparent magnitude difference $\Delta A = 5^*$, the equivalent ratio of their linear amplitudes is 1/100. Such a logarithmic representation also facilitates the numerical stability of the optimization process and is used in further

*For simplicity, apparent magnitude is denoted by A in the following text

expressions. The readout pixel value z_i is linearly proportional to the part of the radiant flux of the astronomical object that impinges on the ith pixel. The ith pixel error is

$$e_i = z_i - f(\mathbf{x}_i, \beta), \tag{3.10}$$

where $f(\mathbf{x}_i, \beta)$ denotes the ith pixel model value.

3.4.1 State-of-the-art methods for Gaussian profile fitting

Gaussian profiles of different dimensions are widely applied in various engineering fields. In spectrography, 1D Gaussian profiles are used for modeling emission and absorption spectral lines [32, 33, 34, 35]. 2D Gaussian profiles are widely used in engineering fields that involve image processing. A 2D Gaussian profile can approximate a circular Airy disc even in a theoretically ideal diffraction-limited imaging system. It can also be used to model the total blur of a point source. Among others, many papers in astrometry use 2D Gaussian profiles to identify and track stellar or other astronomical objects [34, 36, 37, 38]. 2D Gaussian models are applied in bioimaging and medical imaging. In microscopy, they are used for single particle detection [39, 40] and tracking [41, 42]. Also, 2D and 3D Gaussian profiles are used in computer tomography (CT) [43] and positron emission tomography (PET) [44] to reconstruct the underlying 3D process, i.e., the volume of interest. The generalized multidimensional Gaussian profile has even more extensive applications [45, 46]. The selected approach for parameter estimation of the multidimensional Gaussian profile, especially from the noisy data, depends on the required estimate accuracy, computational complexity and speed, and prior knowledge of the profile's parameters and statistics of noise [42].

Parameters can be estimated in the domain of exponential function values or an exponential function argument. The fastest and simplest method in the domain of exponential function's values is the analytical method of moments where the first moment yields the expected centroid position (mean), while the second moment yields the expected covariance matrix [47, 48]. The method of moments, which is also called the center-of-gravity (CoG) method, can be written as $\hat{\mu} = (\hat{x_0}, \hat{y_0}) = \left(\frac{\sum_{i=1}^{m} x_i I_i^p}{\sum_{i=1}^{m} I_i^p}, \frac{\sum_{i=1}^{m} y_i I_i^p}{\sum_{i=1}^{m} I_i^p} \right)$, where x_i and y_i are input samples' position, and I_i is the intensity of the ith of m input samples within the selected ROI. The variable p denotes the power of pixel intensity, and it is introduced to adapt the method to different noise statistics and underlying signals. The profile's centroid is estimated for $p = 1$, while the centroid of squares is estimated for $p = 2$ [48, 49]. Due to its speed, the method of moments can be used in real-time applications and when the requirements on the accuracy of estimated parameters are lower [41, 50]. However, the method is sensitive to noise and selected ROI. It is also biased with usually the underestimated profile's variances.

Conventionally, optimal parameters of the Gaussian profile are searched for in either the

least squares (LS) or maximum likelihood (ML) sense. Both approaches require solving the overdetermined nonlinear system of equations and consequently require iterative procedures for nonlinear optimization. Examples of such iterative methods are the quasi-Newton method [51], downhill simplex method [41, 52], and Levenberg–Marquardt method [53], [54].

The conventionally used LS method minimizes the sum of squared residuals between the empirical input data and estimated model without any prior knowledge of noise statistics [55]. Due to nonlinear optimization, it requires high computational costs. The main disadvantage of the LS method is that it does not guarantee the global optimum since it can get stuck in some of the local minima depending on the selected initial parameters, step size, etc. The more robust version of the LS method for multidimensional Gaussian fitting was proposed in [56], where the conventional LS method is modified by adding the regularization term that uses the Kullback-Leibler divergence to measure the closeness of the input empirical data and the estimated Gaussian function [57].

The second approach is searching for optimal Gaussian profile parameters in the maximum likelihood sense. The ML method searches for the most likely parameters that yield empirical input data. It yields an unbiased solution of minimum variance, but a precondition is a large number of input samples and a prior knowledge of noise statistics. N. Hagen et al. in [58, 59] estimated the 1D and 2D Gaussian profiles in the ML sense from the observations contaminated with additive Gaussian noise and provided analytical first and second derivatives of the log-likelihood function. Newton's method was used for nonlinear optimization and parameters update. They also provided the estimated parameters' sensitivity analysis. They calculate the Fisher information matrix, which is the negative expectation of the Hessian matrix, and invert it to obtain the Cramer-Rao bound for the covariance matrix of estimated parameters. The ML estimate achieves the Cramer-Rao bound if the number of input samples is large enough.

The other group of methods searches for optimal Gaussian parameters in the exponential function's argument. Both the model and empirical input data are logarithmized, thus reducing the estimation problem for the 1D Gaussian profile to polynomial fitting. R.A. Caruana et al. proposed the fast method for 1D Gaussian fitting in the argument domain in the least squares sense by analytically solving the overdetermined system of linear equations [60]. Such a one-step solution algorithm is high-speed but of reduced accuracy in the presence of additive noise. In the log domain, the small profile values introduce a significant estimation error since the ratio of noise variance and squared profile value occurs in the expectation of the sum squared residuals between the observations and model. It is especially emphasized at wide ROIs where additive noise and small profile values dominate. H.Guo [61] proposed extending R.A. Caruana's method by introducing estimation weights equal to Gaussian profile values to reduce the method's sensitivity to additive noise and the influence of small profile values. The input data values were used to approximate ideal estimation weights since the actual profile's values are

usually unknown. The hard thresholding of small values additionally increases the accuracy of the method. Estimation weights selected as input profile values correspond to the error function's approximation with only the Taylor series's first term. In the case of long-tailed noise contamination, such an approximation is insufficient, especially at wide ROIs. To solve the problem of long-tailed noise contamination, H. Guo introduced the iterative procedure where the weights in each iteration are approximated by the estimated model values from the previous iteration [61]. Al-Nahal et al.proposed a fast, accurate, and separable method for 1D Gaussian fitting in the log domain where the profile's standard deviation (STD) is directly calculated by equating the total area under the Gaussian function obtained numerically and Q-function properties [62]. At the same time, the remaining two parameters were calculated the same as in Guo's method. An iterative procedure is further introduced to solve the problem of long-tailed noise. However, for the 2D and higher profile dimensions, even if the estimation is transformed to the argument domain, the profile's parameters are still nonlinearly coupled, and the estimation procedure is complex. S. M. Anthony et al. [52] estimated the parameters of the 2D Gaussian profile in the log domain, but only the uncorrelated 2D Gaussian profile was considered. Furthermore, the influence of noise was neglected, and it was assumed that the background was removed.

Another approach that avoids iterative procedure and transforms the nonlinear least-squares fitting into a standard linear least-squares fitting was proposed by K. Roonizi [63], which uses differentiation and integration and assumes that the Gaussian function is riding on the polynomial background. The method suffers from accumulated noise error from the numerical integration process, which was solved in [64]. However, both approaches are proposed only for the case of 1D Gaussian.

3.4.2 Iteratively reweighted least squares (IRWLS) method with Huber weights for overlapping 2D Gaussian profiles fitting

We proposed in [65] our adaptation of the conventional iteratively reweighted least squares method (IRWLS) [66] for solving the problem of overlapping 2D Gaussian components. The IRWLS method gradually searches for optimal model parameters $\hat{\beta}$ by minimizing the weighted sum of squared residuals $D(\beta)$ in each iteration for all samples within the selected ROI as follows

$$\hat{\beta} = \arg\min_{\beta} D(\beta), \tag{3.11}$$

where the objective function has the following form

$$D(\beta) = \sum_{i \in ROI} w_i(e_i) e_i^2. \tag{3.12}$$

Unit estimation weights correspond to the usual LS method. In the case of additive noise contamination, the LS method provides a parameter estimate of the minimum variance. However, in the case of overlapping objects, which can be considered long-tailed noise contamination, estimation weights should be introduced to separate valid samples from outliers and reduce outliers' contribution to the final estimate. The LS method with unit weights does not recognize samples belonging to neighboring overlapping objects as outliers and greedily tries to estimate all overlapping components with a single 2D Gaussian profile since it is the currently optimal solution in the LS sense. Consequently, the estimated parameters of such a profile do not correspond to any overlapping actual components. Therefore, the estimation weights must be adjusted to recognize and reduce the impact of outliers. Outliers can be treated in two ways: they can be completely omitted from the estimation procedure, which corresponds to a hard thresholding, or their weights can be reduced, corresponding to a soft thresholding. The hard thresholding also causes the loss of useful information contained in these discarded samples.

For recognition of outliers and reduction of their contribution to the final estimate, Huber weights [67] are used in this research. Such weights are based on robust measures of residual error dispersion from the previous iteration as follows

$$w_i(e_i) = w_H(e_i) = \begin{cases} 1, & \text{if } |e_i| \leq \delta \\ \delta/|e_i|, & \text{if } |e_i| > \delta, \end{cases} \tag{3.13}$$

where e_i denotes the residual error between the observation and model for the ith pixel from the previous iteration, and δ corresponds to the threshold value, which separates the valid samples from outliers. Huber weights represent the combining of L_1 and L_2 norms. The L_1 norm is applied to outliers with absolute error larger than the calculated threshold δ, while the L_2 norm is applied to the rest of the valid samples. The threshold δ is calculated as a product of estimated error dispersion $\hat{\sigma}_e$ and the tuning constant k_{const} as $\delta = k_{const}\hat{\sigma}_e$. The residual error dispersion $\hat{\sigma}_e$ is estimated from the errors from the previous iteration using some robust measure of error dispersion. The most commonly used robust measures of dispersion are the median of the absolute deviations from the data's median (MAD) and inter-quartile range (IQR). The precondition for applying robust statistics is at least 50% valid samples. Robust measures of dispersion are robust and efficient in the case of long-tailed noise contamination at the cost of lower efficiency in the case of normally distributed errors. To achieve the high asymptotic efficiency even in the case of normally distributed errors, the robust error dispersion can be estimated as $\hat{\sigma}_e = MAD/0.6745$ or $\hat{\sigma}_e = IQR/1.349$. We estimated $\hat{\sigma}_e$ as $\hat{\sigma}_e = MAD/0.6745$ and selected the threshold δ as $\delta = 1.345\hat{\sigma}_e$ to achieve the asymptotic efficiency of 95% in the case of additive noise contamination as it is proposed in [68]. The IRWLS method with Huber weights can be generally applied in estimation from data contaminated with long-tailed noise

and outliers but with the prerequisites of 50% of valid samples and the separable bi-modal histogram of residuals. The factor k_{const} determines the share of influence of the L_1 and L_2 norms. Smaller values of the factor k_{const}, such as $k_{const} = 1$, prefer the L_1 norm, which is then applied to many input samples. A higher value of the factor k_{const} increases the value of the threshold δ, and consequently, the L_2 norm is applied to a higher percentage of input samples.

The IRWLS method's pseudocode for estimation of one 2D Gaussian component is given in Algorithm 1, while the method's flowchart is shown in Fig. 3.5. The algorithm consists of inner and outer nested loops. The outer loop calculates estimation weights from residual errors from the previous iteration while the inner loop performs the nonlinear optimization to minimize (3.11) and obtain optimal parameters for fixed estimation weights, which are calculated in the outer loop.

Algorithm 1 IRWLS method for single ROI

Input: positions of all n samples within the ROI and their values $(\mathbf{x}_i, z_i) \in ROI$
Output: parameters $\hat{\beta}$

1: $t \leftarrow 0$
2: $w_i^{(0)} \leftarrow 1, \forall i \in ROI$
3: $\hat{\beta}^{(0)} \leftarrow \{A, \lambda_1, \lambda_2, \mu_x, \mu_y, \rho\}$ ▷ initial LS estimates
4: $t \leftarrow t + 1$
5: **do** ▷ outer loop
6: $\quad e_i^{(t-1)} \leftarrow z_i - f(\mathbf{x}_i, \hat{\beta}^{(t-1)}), \forall i \in ROI$
7: $\quad \delta^{(t)} = 1.345 \cdot MAD(\mathbf{e}^{(t-1)})/0.6745$
8: $\quad w_i^{(t)} \leftarrow w_H(e_i^{(t-1)}, \delta^{(t)}), \forall i \in ROI$ ▷ (3.13)
9: $\quad \hat{\beta}^{(t)} \leftarrow \arg\min_{\beta} D(\beta)$ ▷ inner loop (3.11)
10: $\quad t \leftarrow t + 1$ ▷ outer loop iteration
11: **while** $\frac{|\delta^{(t-1)} - \delta^{(t)}|}{|\delta^{(t-1)}|} > \text{tol} \ \ \& \ \ t < t_{max}$

The analytical gradients of (3.12) concerning each profile's parameter are provided to speed up the nonlinear optimization procedure. The analytical gradients are:

$$\frac{\delta D}{\delta A_m} = \frac{\delta D}{\delta A_l}\frac{\delta A_l}{\delta A_m} = \ln\sqrt[5]{100}\sum_{i=1}^{n}\tau_i,$$

$$\frac{\delta D}{\delta \mu_x} = \sum_{i=1}^{n}\frac{\tau_i}{2\lambda_1^2\lambda_2^2}\Big((x_i-\mu_x)((\lambda_1^2+\lambda_2^2)+\cos\rho(\lambda_2^2-\lambda_1^2))+(y_i-\mu_y)((\lambda_2^2-\lambda_1^2)\sin\rho)\Big),$$

$$\frac{\delta D}{\delta \mu_y} = \sum_{i=1}^{n}\frac{\tau_i}{2\lambda_1^2\lambda_2^2}\Big((y_i-\mu_y)((\lambda_1^2+\lambda_2^2)+\cos\rho(\lambda_1^2-\lambda_2^2))+(x_i-\mu_x)((\lambda_2^2-\lambda_1^2)\sin\rho)\Big),$$

$$\frac{\delta D}{\delta \rho} = \frac{\delta D}{\delta \theta}\frac{\delta \theta}{\delta \rho} = \sum_{i=1}^{n}\frac{\tau_i(\lambda_1^2-\lambda_2^2)}{2\lambda_1^2\lambda_2^2}\Big((y_i-\mu_y)\cos\frac{\rho}{2}-(x_i-\mu_x)\sin\frac{\rho}{2}\Big)$$
$$\Big((y_i-\mu_y)\sin\frac{\rho}{2}+(x_i-\mu_x)\cos\frac{\rho}{2}\Big),$$

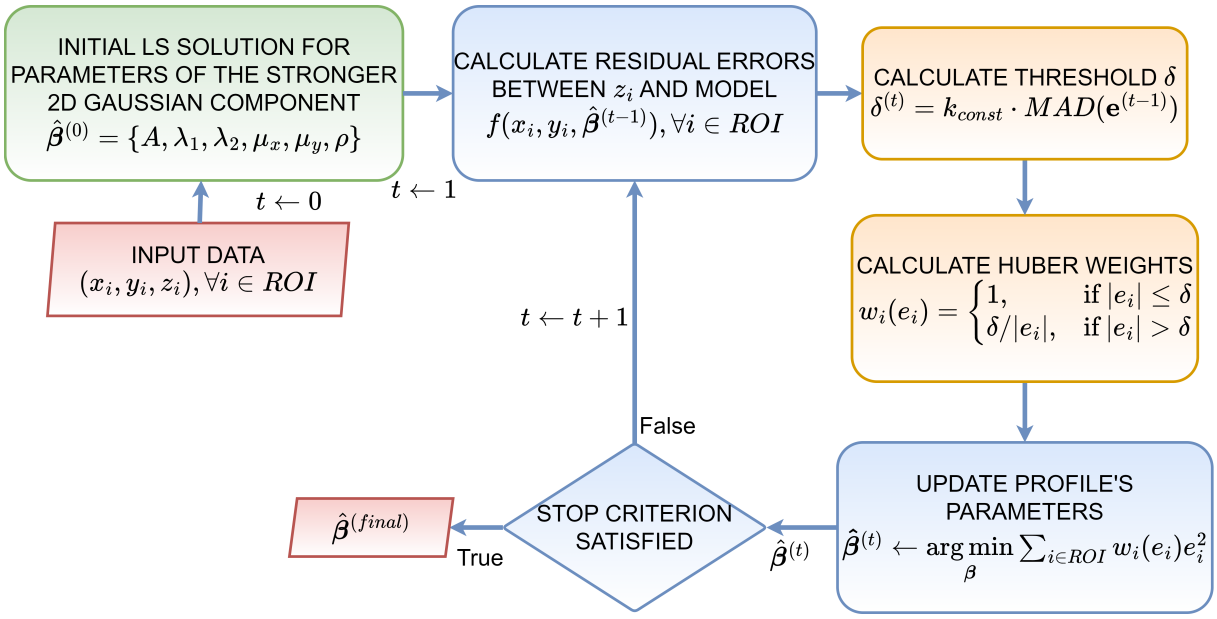


Figure 3.5: IRWLS method in the estimation of the stronger of two overlapped 2D Gaussian components

$$\frac{\delta D}{\delta \lambda_1} = \sum_{i=1}^{n} \frac{\tau_i}{\lambda_1^3} \left((x_i - \mu_x)^2 \cos^2 \frac{\rho}{2} + (y_i - \mu_y)^2 \sin^2 \frac{\rho}{2} + \sin \rho (y_i - \mu_y)(x_i - \mu_x) - \lambda_1^2 \right),$$

$$\frac{\delta D}{\delta \lambda_2} = \sum_{i=1}^{n} \frac{\tau_i}{\lambda_2^3} \left((x_i - \mu_x)^2 \sin^2 \frac{\rho}{2} + (y_i - \mu_y)^2 \cos^2 \frac{\rho}{2} - \sin \rho (y_i - \mu_y)(x_i - \mu_x) - \lambda_2^2 \right),$$

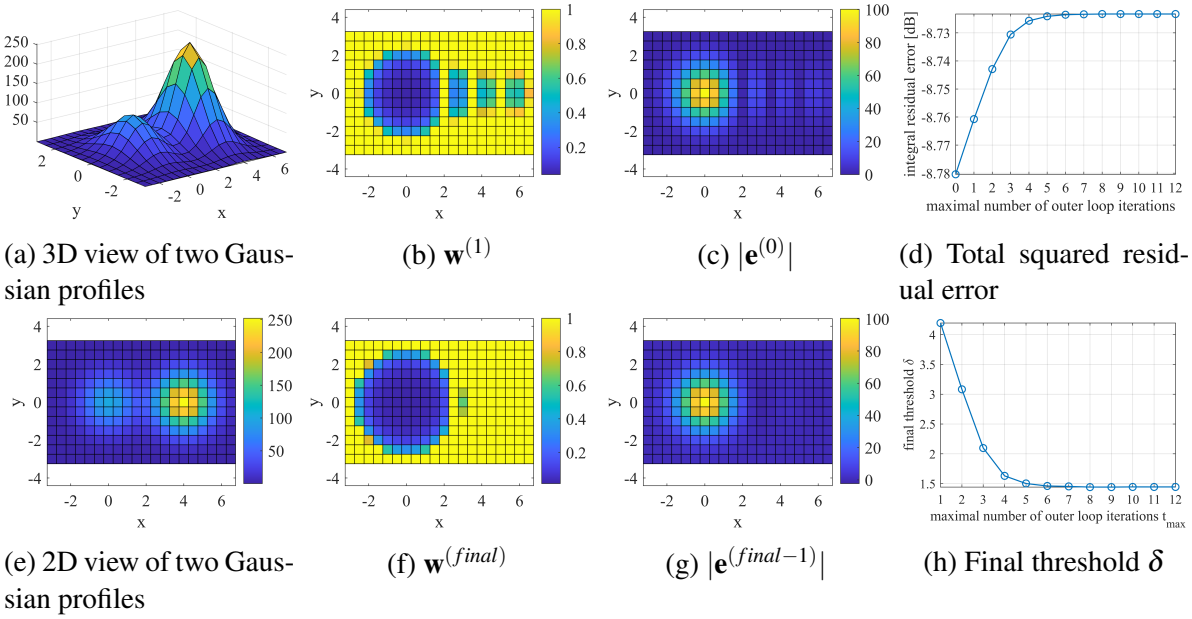
where $\tau_i = 2w_i(f(\mathbf{x}_i, \beta) - z_i) f(\mathbf{x}_i, \beta)$. The nonlinear optimization for the initial LS solution requires a good initial guess for the profile's parameters that can be obtained using the method of moments or the method of local extrema.

3.4.3 Illustration of the IRWLS method performance

An illustration of the IRWLS method in the estimation of the stronger of two overlapping 2D Gaussian profiles is given in Fig. 3.6. For simplicity, two rotationally symmetric 2D Gaussian components with the same semiaxes widths are considered ($\lambda_1 = \lambda_2 = \sigma, \sigma_1 = \sigma_2 = 1$). The given parameters of the stronger components are $\beta_1 = [A_m, \lambda_1, \lambda_2, \mu_x, \mu_y, \rho] = [8, 1, 1, 4, 0, 0]$, while the parameters of the weaker component are $\beta_2 = [7, 1, 1, 0, 0, 0]$. The given parameters show that their peak positions are displaced by $d = 4\sigma$ with the difference in apparent magnitudes $\Delta A = 1$. Components of the given parameters were synthesized over the rectangular grid with the pixel size $p_x = p_y = 0.5$ as Figs. 3.6a and 3.6e show. The given parameters of the stronger component are passed as an initial parameter guess for the LS method to avoid the influence of wrong initialization on the accuracy of the final estimate.

Estimation of the stronger of two overlapping Gaussian components

Figs. 3.6c and 3.6g show the residual errors in the initial and final iteration of the IRWLS method when the stronger of two overlapped 2D Gaussian components is estimated. Figs. 3.6b



(a) 3D view of two Gaussian profiles

(b) $\mathbf{w}^{(1)}$

(c) $|\mathbf{e}^{(0)}|$

(d) Total squared residual error

(e) 2D view of two Gaussian profiles

(f) $\mathbf{w}^{(final)}$

(g) $|\mathbf{e}^{(final-1)}|$

(h) Final threshold δ

Figure 3.6: Estimation of the stronger 2D Gaussian profile in the case of two overlapping components using the IRWLS method with the uniform sampling grid and the pixel size of 0.5 in both directions. The total number of outer loop iterations is 12.

and 3.6f show the corresponding estimation weights calculated from those residuals. Estimation errors in Fig. 3.6c correspond to the error of the LS estimate, which has undesirable ridges at positions where the stronger component is dominant. The same effect can be seen in estimation weights calculated from those errors, indicating that the LS method did not appropriately estimate and remove the stronger component. Regardless, such a wrong LS estimate is used for IRWLS method initialization. In each outer loop iteration, the IRWLS method with Huber weights gradually deweights the weaker component, recognizing its samples as outliers, and assigns unit weights to the samples belonging to the stronger component. By changing the domain of the input samples that participate in the estimation, the IRWLS method gradually converges toward the actual parameters of the stronger component. The residual error of the IRWLS method in the final iteration is close to zero around the peak position of the stronger component, which implies that the stronger component is completely removed while the weaker component is left almost untouched.

Fig. 3.6d shows the increase of the squared residual errors with iterations of the IRWLS method compared to the error in the initial iteration, which corresponds to the LS solution. It is the natural behavior of the IRWLS method. Namely, the method reduces the weighted sum of squared residuals at the expense of increasing the sum of squared residuals. However, the total squared residual error will be smaller than the initial LS estimation error when the weaker component is removed as well.

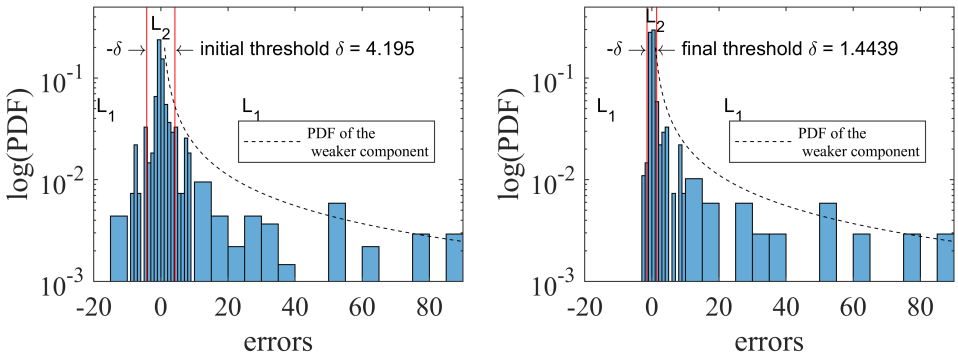
Fig. 3.6h shows the decrease of threshold δ with iterations of the IRWLS method. The threshold δ converges in approximately six iterations, thus indicating the required number of

iterations for the convergence of the IRWLS method.

The estimated parameters of the stronger 2D Gaussian component using LS method were $\hat{\beta}_1 = [8.0569, 1.0912, 1.0000, 3.9557, 0.0000, 0.0000]$, while using the IRWLS method were $\hat{\beta}_1 = [8.0122, 1.0145, 1.0001, 3.9885, 0.0000, 0.0000]$. The IRWLS method yields a solution closer to the actual given profile's parameters than the LS method. The LS method yields an elongated and shifted solution since it tries to fit both Gaussian components simultaneously.

Statistical analysis of residual errors

Fig. 3.7 represents the logarithm of the probability density function (PDF) of residual errors in the initial and final iteration of the IRWLS method when the stronger component is being estimated. The PDFs of errors are approximated from experimentally obtained error histograms.



(a) PDF of errors in the initial iteration (b) PDF of errors in the final iteration

Figure 3.7: PDF of errors when the stronger of two overlapping 2D Gaussian profiles is being estimated using the IRWLS method

Fig. 3.7a corresponds to the residual errors of the LS method that have both positive and negative values. The assumption is that the positive and negative values normally distributed around zero are related to the estimation error of the stronger component, while the significant positive errors correspond to the weaker component's values. The dashed curves denote the PDF of the weaker component's values, which is strictly positive, as will be derived below. Hence, the negative errors indicate that the LS method inappropriately estimated the stronger component due to negative errors which do not belong to the sample values of the weaker component. Red lines denote the positive and negative threshold values δ, which determines the action limits between the L_1 and L_2 norms. The threshold δ reduces from the initial $\delta = 4.195$ to the final value $\delta = 1.444$ since the model is gradually adapted to the stronger component with fewer and fewer undetected outliers. Consequently, the method spatially discriminates against the profiles' support by solely using the information of the residual errors. Fig. 3.7b shows the residual errors in the last iteration of the IRWLS method. The residual errors are mostly positive and match the positive PDF of the weaker component, indicating that the stronger component is properly estimated.

3.4.4 Probability density function (PDF) of Gaussian profile's values

To derive the PDF of the 2D Gaussian profile's values, let us assume the rotationally symmetric 2D Gaussian profile, centered at $\mu = \mathbf{0}$, which can be expressed in polar coordinates as

$$z(r,\theta) = A\exp\left(\frac{-r^2}{2\sigma^2}\right), \tag{3.14}$$

where $\lambda_1 = \lambda_2 = \sigma$ corresponds to the profile's width, A denotes the high of the profile's peak, including the profile's normalization term, while r denotes the radial distance from the profile's peak position. Let us assume that the profile's support is a circular region, centered at the profile's peak position, of radius $r \in [0, k\sigma]$, where k denotes the factor of Mahalanobis distance [†]. The maximal radius of the circular domain determines the co-domain of the 2D Gaussian profile's values as $z \in [A\exp(-k^2/2), A]$. Let us assume that input samples' positions are random variables, X and Y, that are uniformly distributed within the circular ROI, defined as $D = \{(x,y)|x^2+y^2 \leq (k\sigma)^2\}$, and whose joint PDF can be expressed as

$$f_{XY}(x,y) = \begin{cases} \frac{1}{(k\sigma)^2\pi}, & \text{if } (x,y) \in D \\ 0, & \text{otherwise.} \end{cases} \tag{3.15}$$

The method of transformations can be applied to transform this joint PDF $f_{XY}(x,y)$ to polar coordinates $f_{R\theta}(r,\theta)$ as follows

$$x = r\cos(\theta) = h_1(r,\theta), \tag{3.16}$$

$$y = r\sin(\theta) = h_2(r,\theta),\ r \geq 0,\ 0 \leq \theta \leq 2\pi, \tag{3.17}$$

$$f_{R\theta}(r,\theta) = f_{X,Y}(h_1(r,\theta), h_2(r,\theta))|\mathbf{J}|, \tag{3.18}$$

$$|\mathbf{J}| = \det\begin{bmatrix} \frac{\partial h_1}{\partial r} & \frac{\partial h_1}{\partial \theta} \\ \frac{\partial h_2}{\partial r} & \frac{\partial h_2}{\partial \theta} \end{bmatrix} = \det\begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix} = r. \tag{3.19}$$

Finally, the joint PDF $f_{R\theta}(r,\theta)$ is

$$f_{R\theta}(r,\theta) = \begin{cases} \frac{r}{(k\sigma)^2\pi}, & \text{if } r \in [0,k\sigma], \theta \in (0,2\pi) \\ 0, & \text{otherwise.} \end{cases} \tag{3.20}$$

[†]The quadratic form in the argument of the Gaussian function represents the squared Mahalanobis distance of the samples from the peak position. By choosing the ROI within $k\sigma$ around the peak position, we are defining a convex input domain with the radius of the Mahalanobis distance equal to k in all directions. Samples of the profile at the same Mahalanobis distance are the samples of the same values.

The 2D Gaussian value is a random variable Z since it is a function of two random variables, r and θ. Its PDF can be calculated from the cumulative density function (CDF), which equals

$$\begin{aligned} F_Z(z) &= P(Z \leq z) && (3.21)\\ &= \iint_D f_{R\theta}(r,\theta) dr d\theta && (3.22)\\ &= \int_0^{2\pi} \int_r^{k\sigma} \frac{r}{(k\sigma)^2\pi} dr d\theta && (3.23)\\ &= \int_0^{2\pi} \int_{\sqrt{-2\sigma^2 \ln \frac{z}{A}}}^{k\sigma} \frac{r}{(k\sigma)^2\pi} dr d\theta && (3.24)\\ &= 1 + \frac{2}{k^2}(\ln z - \ln A). && (3.25) \end{aligned}$$

The PDF of Z can be calculated by differentiating F_Z

$$f_Z(z) = \frac{dF_z(z)}{dz} = \frac{2}{zk^2}. \tag{3.26}$$

The PDF of value Z corresponds to the log-uniform distribution in the range $[a,b]$ since

$$f_Z(z) = \frac{1}{z \ln \frac{b}{a}}, \tag{3.27}$$

where $a = A\exp(-k^2/2)$ is the profile's value at the edge of circular ROI determined by $r = k\sigma$, while $b = A$ is the maximal profile's peak value. Finally, the PDF of Z can be written as

$$f_Z(z) = \frac{1}{z \ln \frac{A}{A\exp(-k^2/2)}} = \frac{2}{zk^2}. \tag{3.28}$$

with the mean

$$\mu_Z = E[Z] = \int_{A\exp(\frac{-k^2}{2})}^{A} z f_Z(z) dz = \frac{2A}{k^2}\left(1 - \exp\left(\frac{-k^2}{2}\right)\right), \tag{3.29}$$

and the variance

$$\begin{aligned} \mathrm{Var}(Z) &= E[(Z-\mu_z)^2] = \int_{a=A\exp(\frac{-k^2}{2})}^{b=A} (z-\mu_z)^2 f_Z(z) dz \\ &= \frac{a^2-b^2}{2(\log(a)-\log(b))} - \left(\frac{(a-b)^2}{\log(a)-\log(b)}\right)^2 \\ &= \frac{A^2 e^{-k^2}(e^{\frac{k^2}{2}}-1)(e^{\frac{k^2}{2}}(k^2-4)+k^2+4)}{k^4}. \end{aligned} \tag{3.30}$$

The 2D Gaussian profile values over the uniform rectangular region, such as a rectangular region used in the illustrative example in Fig. 3.6, can be considered a realization of a stochastic process drawn from the derived log-uniform distribution. The proof for that assumption is that the empirically obtained PDF from the histogram of positive errors follows the analytically derived PDF of the weaker component, as Fig. 3.7b shows. It confirms that those positive errors belong to the weaker 2D Gaussian component. Regardless of bounded support of log-uniformly distributed positive errors, such contamination can still be considered a heavy-tailed noise compared to the much narrower normally distributed positive and negative modeling errors around the zero, which belong to the stronger component. We can conclude that such a histogram is bi-modal, thus enabling the separation of the errors from the mixture of distributions by a suitably chosen threshold δ. The given example shows the effectiveness of the proposed IRWLS method in handling heavy-tailed noise contamination. After the removal of the stronger component, the estimated parameters of the weaker component by using the IRWLS method are $\hat{\beta}_2 = [6.9913, 1.0002, 0.9914, -0.0089, 0.0000, -3.1416]$, while using the LS method are $\hat{\beta}_2 = [6.9303, 1.0000, 0.9193, -0.0578, 0.0000, -3.1416]$. The results show that the IRWLS method yields a more accurate estimate than the LS method for the weaker component as well.

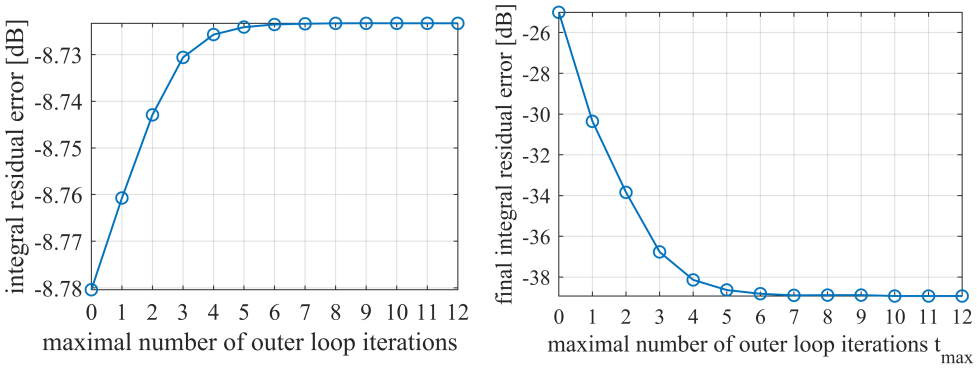
Total modeling error

Instead of performing a sensitivity analysis of each estimated 2D Gaussian profile parameter to changes in input data, the total modeling error is used as a composite fitness measure, indirectly showing the estimation accuracy of all profile parameters. The total modeling error is a relative integral residual error over some arbitrary evaluation domain I, defined as

$$e_{total} = 10\log_{10}\left(\frac{\sum_{\forall i \in I}\left(z_i - \sum_{j=1}^{M} f(\mathbf{x}_i, \hat{\boldsymbol{\beta}}_j)\right)^2}{\sum_{\forall i \in I} z_i^2}\right), \tag{3.31}$$

where z_i corresponds to the ith sample value in the evaluation domain I, $f(\mathbf{x}_i, \hat{\boldsymbol{\beta}}_j)$ is the jth of M estimated models over the selected domain I. This common measure has a sense after removing all expected profiles in the selected evaluation domain I, as Fig. 3.8 shows. Fig. 3.8a shows the total modeling error after removing only the stronger of two overlapped Gaussian profiles from the previously illustrative example in Fig. 3.6. In that case, the LS solution in iteration zero has the smallest total residual error, while this error rises with the iterations of the IRWLS method. It is the natural behavior of the IRWLS method since the integral residual error increases at the expense of reducing the sum of weighted squared residual errors, yielding a better solution for the stronger component. However, the total modeling error after removing both overlapped components reduces with iterations of the IRWLS method, demonstrating the method's efficiency (Fig. 3.8b).

The LS method yields the total modeling error after removing both components -25.0045



(a) Integral residual error after removal of the stronger component

(b) Final integral residual error after removal of both overlapped components

Figure 3.8: Total modeling error in estimating two overlapping 2D Gaussian components over a rectangular uniformly sampled domain. Zero iteration corresponds to the LS solution.

dB, and the IRWLS method -38.9272 dB, as shown in Fig. 3.8b. The total modeling error is achieved after 12 iterations for estimating the stronger of two overlapped components and five iterations for estimating the weaker component. The number of outer loop iterations of the IRWLS method was limited to 12, but 90% total modeling gain was achieved after only four iterations, indicating the IRWLS method's convergence rate.

IRWLS method's efficiency

Since the analytical expressions for parameters' variances are hard to derive due to estimation weights update in each iteration according to residual errors from the previous iteration, the method's efficiency was numerically verified through Monte Carlo simulation. The experimental results were published in [65]. We considered the noiseless case of two overlapping rotationally symmetric 2D Gaussian components of the same semiaxes widths. In the noiseless case, a few contributing factors affect the method's efficiency: center-to-center distance d of profiles' peak positions, their difference in brightness, i.e., the absolute difference in their apparent magnitudes $\Delta A = |A_{m1} - A_{m2}|$, the ratio of the profiles' widths and the pixel size, and the selected tuning factor k_{const}. Namely, center-to-center distance indirectly affects the number of samples that belong to a particular component and the ability of the method to separate the overlapped components. The ratio of the profile's width and the pixel size also affects the total number of input samples and, consequently, variances of the estimated profile's parameters.

In the first experiment, we analyzed the influence of center-to-center distance d and the absolute difference in magnitudes ΔA on the efficiency of the IRWLS method. The considered center-to-center distances d were from 0 to 6σ with the stepsize 0.5σ, while considered differences in absolute magnitudes were from 0 to 5 with the stepsize 0.5. The Monte Carlo simulation was performed for each combination of these factors. The widths of profiles' semiaxes were fixed to $\lambda_1 = \lambda_2 = \sigma = 1$ pix, and the tuning factor k_{const} was set to $k_{const} = 1.345$. The

center-to-center distances are selected from 0 to 6σ since 99.7% In each simulation, one image with 192 sub-images was synthesized. Each sub-image contains two overlapping 2D Gaussian components on distance d and of selected difference in magnitudes ΔA, as shown in Fig. 3.9. The stronger of the two components is positioned at the center of each sub-image with a random fractional shift of ± 0.5 pix. The weaker component is at the distance d from the center of the stronger component but at a random angular position. The introduced random fractional shifts simulate the random profile sampling over a regular grid. Fig. 3.10 shows the uniform distribution of profiles' centers from the nearest integer position for both components. It shows that such synthesized sub-images are good representatives of random sampling in actual images.

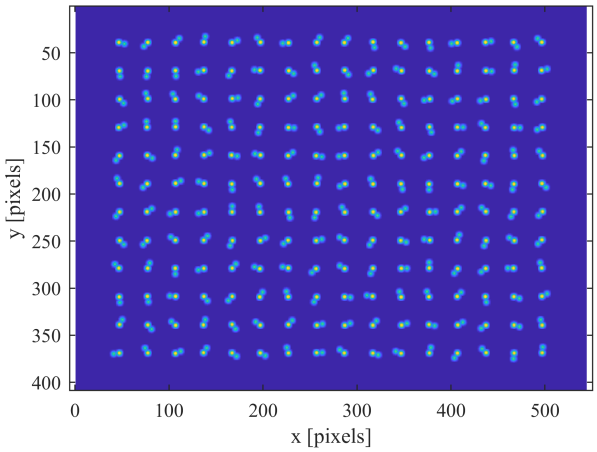
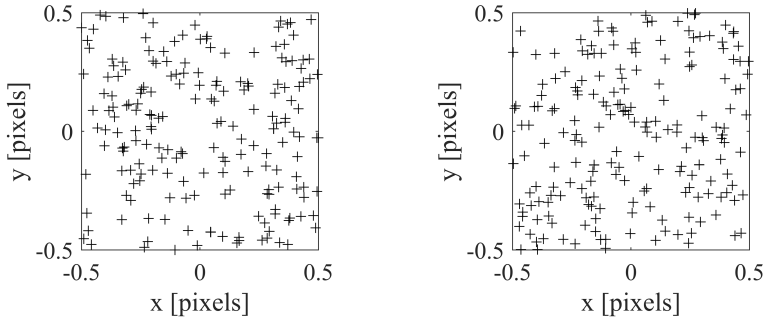


Figure 3.9: Synthetic image of 192 sub-images and 384 objects. Objects' parameters: $\sigma = 1\,pix$, $d = 6\sigma$, $\Delta A = 3$.



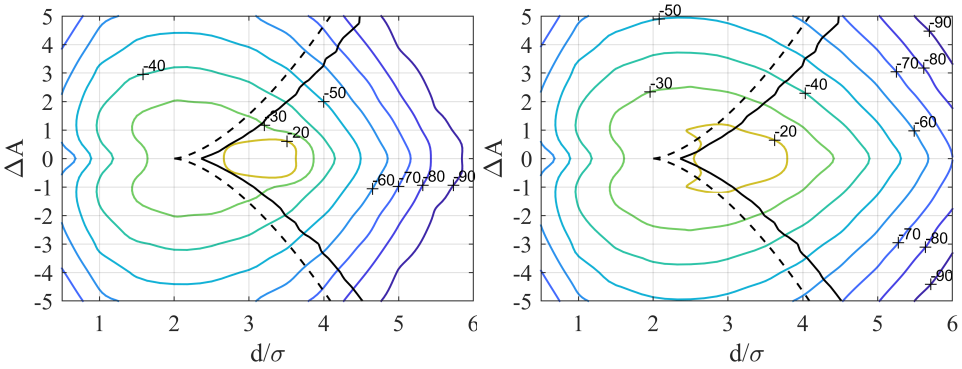
(a) Random fractional shifts of the stronger components from the center of subimage

(b) Distance of the centers of weaker components to the nearest integers

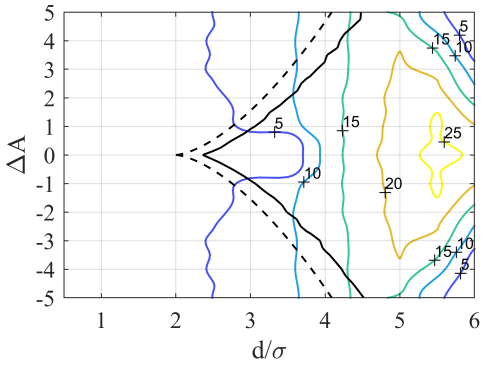
Figure 3.10: Distribution of centers of the profiles concerning the nearest integers

2D Gaussian components were successively extracted, estimated, and removed based on maximum pixel intensity detection. In each iteration, the component located at the currently detected maximum was estimated. Until the total number of components in the synthesized image was not achieved, maxima were iteratively detected, and components were estimated and removed. The estimation ROI in each iteration was selected as a circular region centered at the

detected maximum position with the radius adapted to the expected profile's width. The proposed IRWLS method with Huber weights was compared with the commonly used LS method by using the total modeling error (3.31) as a measure of accuracy. In each simulation, 384 objects from all 192 sub-images were detected, estimated, and removed. The total modeling error was calculated for each sub-image. Medians of these errors are shown in Fig. 3.11 as isolines over uniformly sampled $(d/\sigma, \Delta A)$-plane.



(a) IRWLS method, modeling errors in dB (b) LS method, modeling errors in dB



(c) IRWLS method modeling gain in dB

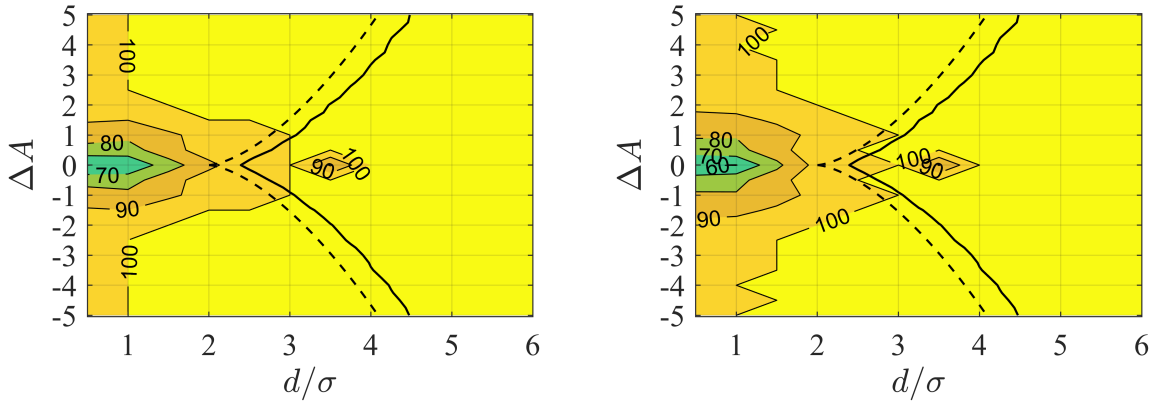
Figure 3.11: The modeling errors of the IRWLS and LS methods and the IRWLS method modeling gain in dB in $d/\sigma\Delta A$-plane. $\sigma = 1$.

The proposed IRWLS method with Huber weights always yields better or the same results as the LS method. For some combination of contributing factors $(d/\sigma, \Delta A)$, the IRWLS method modeling gain is up to 25 dB, such as for $(d/\sigma, \Delta A) = (5.5, 1)$ where the IRWLS method modeling error is -85 dB, while the LS method has the larger modeling error of -60 dB. Consequently, the IRWLS method modeling gain is 25 dB.

When $d/\sigma = 0$, components are completely overlapped and cannot be separated. However, both methods perfectly model such overlapped components with a single profile, and the IRWLS solution corresponds to the LS solution. Contrary, if $d/\sigma = 6$, the components are almost completely separated and modeled as two separated profiles, resulting in high and almost the same accuracy for both methods again.

For the medium range of distances $1 < d/\sigma < 4$, both methods have more significant mod-

eling errors since some estimated components were declared as invalid due to certain rejection criteria (too large estimated semiaxes widths, too small number of input samples within the selected ROI, etc.). Such discarded components were not removed from the synthesized image, thus resulting in a high modeling error. Fig. 3.12. shows percentages of found objects by using the IRWLS and LS methods in the $d/\sigma\Delta A$-plane. The results are similar for both methods. In



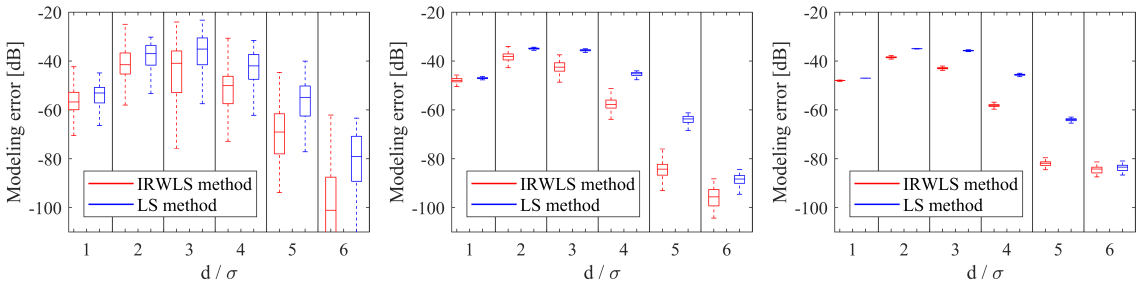
(a) LS method, the percentage of found objects

(b) IRWLS method, the percentage of found objects

Figure 3.12: The percentage of found objects using the IRWLS and LS methods in $d/\sigma\Delta A$-plane defined by (3.5) and (3.6)

an area where distances d/σ and differences of magnitudes ΔA are larger than 2, both methods find 100% of objects. Although it seems that the LS method finds the 100% of objects, the resolution limits, shown in Fig. 3.12, should be considered. The dashed black line represents the theoretical resolution limit in the continuous case according to (3.5) and (3.6), while the solid black line represents the empirical resolution limit in the discrete case with 50% probability that a local minimum exists between profiles. In such a case, after removing the first compromise model, the LS method tries to model the secondary object from the ridges, which correspond to residuals of the first compromise model. In certain cases, such false models still fulfill the chosen validity criterion and are kept as valid secondary objects. On the other hand, with the IRWLS method, both estimated objects are valid since the method is based on the robust analysis of the residual errors and not on the existence of a minimum between overlapped profiles. Furthermore, the IRWLS method gain is visible even in this unresolved area where the IRWLS method achieves 5dB modeling gain, as Fig. 3.11c shows. In the medium range of center-to-center distances $1 < d/\sigma < 4$ and the differences in magnitudes $\Delta A > 1$, the IRWLS modeling gain is at least 5 dB and increases with larger d/σ. The IRWLS modeling gain in this area primarily depends on the center-to-center distance d/σ. If objects have equal magnitudes, they have to be at least $d/\sigma = 2.5$ apart so that the histogram of errors becomes bimodal and the IRWLS method becomes applicable and achieves a modeling gain. The IRWLS method achieves the highest modeling gain for distances $4 < d/\sigma < 6$ due to an accurate estimate of

actual profiles' parameters, as Fig. 3.11c shows.



(a) $\sigma = 0.5, \Delta A = 3$ (b) $\sigma = 1, \Delta A = 3$ (c) $\sigma = 2, \Delta A = 3$

Figure 3.13: Statistical analysis of modeling error for different profile widths σ in pixels

The ratio of the profile's semiaxes width and the pixel size affects the number of input samples and, consequently, variances of estimated parameters. Since parameters are estimated through iterative procedures for nonlinear optimization, the number of input samples has to be much larger than the number of unknowns to achieve higher accuracy and lower variances of estimated parameters. The Monte Carlo simulation with the same setup was performed for profile widths from the set $\sigma \in \{0.5, 1, 2\}$ pix to analyze the influence of different profile widths on estimation accuracy. Fig. 3.13 shows the statistics of modeling errors for different profile widths. The results show that the parameter variances are most prominent for small profile widths such as $\sigma = 0.5$ pix due to a small number of input samples. Still, the median of errors is slightly decreased compared to the cases $\sigma = 1$ pix and $\sigma = 2$ pix. For $\sigma = 1$ pix and $\sigma = 2$ pix, the optimization problem is overdetermined, the number of input samples increases, and the parameter variances decrease, as expected. The statistical analysis of modeling error is shown in Fig. 3.13.

Convergence rate of the IRWLS method

To analyze the convergence rate of the proposed IRWLS algorithm, the elapsed time and the number of outer loop iterations required for the estimation of each of the 384 components in the synthesized image were measured. The termination criteria for the algorithm outer loop was the relative change in threshold δ smaller than 10^{-5}. The threshold δ is used to update estimation weights passed to the algorithm's inner loop, which performs nonlinear optimization. The objective function value tolerance and the current point tolerance were set to 10^{-10}, while the maximal number of iterations was unlimited. The experiment was conducted on a 2.5 GHz Intel(R) Core(TM) i5-7200U CPU and 8 GB of RAM. The nonlinear optimization was performed by using the MATLAB *fminunc* function. Table 3.1 shows the execution time statistics. The mean estimation time is 0.1591 seconds per profile if analytical gradients of the objective function are passed to the optimization procedure. In comparison, it increases to 0.3665 seconds per

profile without supplying analytical gradients. It indicates that the analytic gradients accelerate the estimation process more than two times.

Table 3.1: The execution time required to estimate one 2D Gaussian profile. The average number of input samples per profile was 29, and the average number of outer loop iterations was 20 for the center-to-center distance $d/\sigma = 4$.

	With analytical derivatives	Without analytical derivatives
Mean time [s]	0.1591	0.3665
Median time [s]	0.1313	0.3137
Maximum time [s]	0.5027	1.0865
Minimum time [s]	0.0438	0.1125

Fig. 3.14 shows the median of outer loop iterations in $(d/\sigma, \Delta A)$-plane. The median of outer loop iterations primarily raises with larger center-to-center distances d/σ, with less dependence on the difference in magnitudes ΔA. However, if the outer loop iteration maximum is limited to 5, 80% of the maximum IRWLS method gain is still achieved. It shows that the method can be used in low-complexity applications as well.

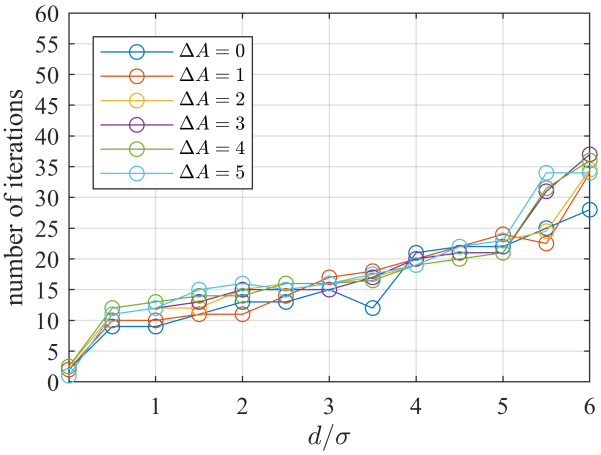


Figure 3.14: Medial number of outer loop iterations for $\sigma = 1$ pix, $tol = 10^{-5}$, different center-to-center distances d/σ, and different ΔA

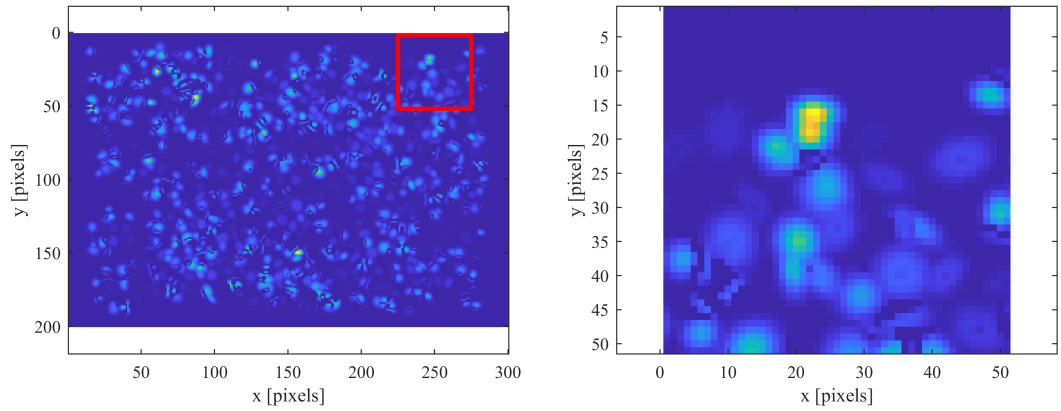
Synthetic astronomical image

In the second experiment, 1000 profiles of random semiaxes widths, rotation angles, centroid positions, and magnitudes were synthesized and placed within the image of 200×300 pixels, including the 10% of padding to simulate the actual astronomical image. An example of such a synthesized image is shown in Fig. 3.3. This experiment aims to precisely estimate the parameters of those densely distributed 2D Gaussian profiles.

The segmentation of the synthesized input image was performed to facilitate the estimation procedure of the LS method and to show that the IRWLS method still achieves a significant

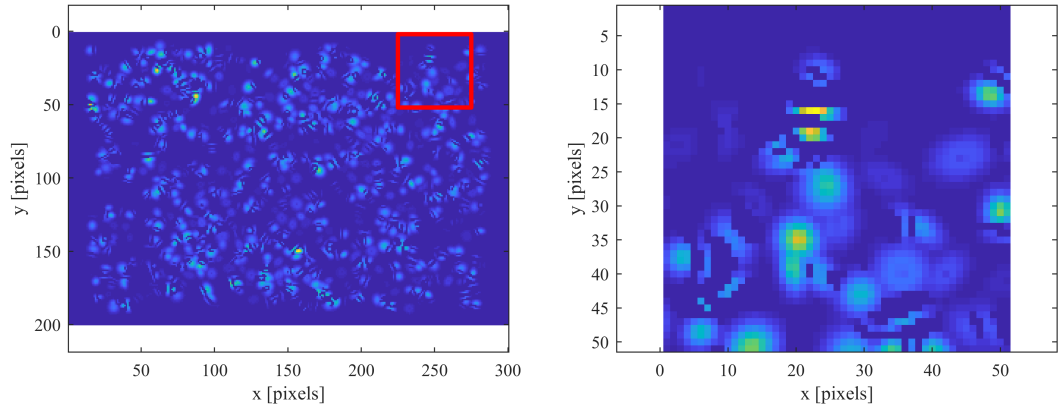
modeling gain. For the segmentation, the generalized Voronoi tessellation is performed, which combines the squared Euclidian distance and the additive term of absolute difference in magnitudes of neighboring profiles as a criterion of pixel closeness to a particular profile. Determination of the pixels' affiliations to the particular profile comes down to searching for bisectors between each two neighboring profiles and finding the convex affiliation regions by using the intersections of these bisectors. In that way, the larger convex region belongs to the stronger component. Fig. 3.4 shows the example of the applied segmentation procedure.

The local extrema method is used to initialize the explained segmentation procedure, where the detected maxima are sequentially declared as potential centroids of 2D Gaussian profiles. For the case of the actual astronomical image, which is noisy, the pre-burring has to be performed to remove isolated wrong pixels or pixels contaminated by salt and pepper noise. The method of moments is used for initial parameter estimation.



(a) Residual error of the IRWLS method with tessellation

(b) Residual error of the IRWLS method (zoomed red square)

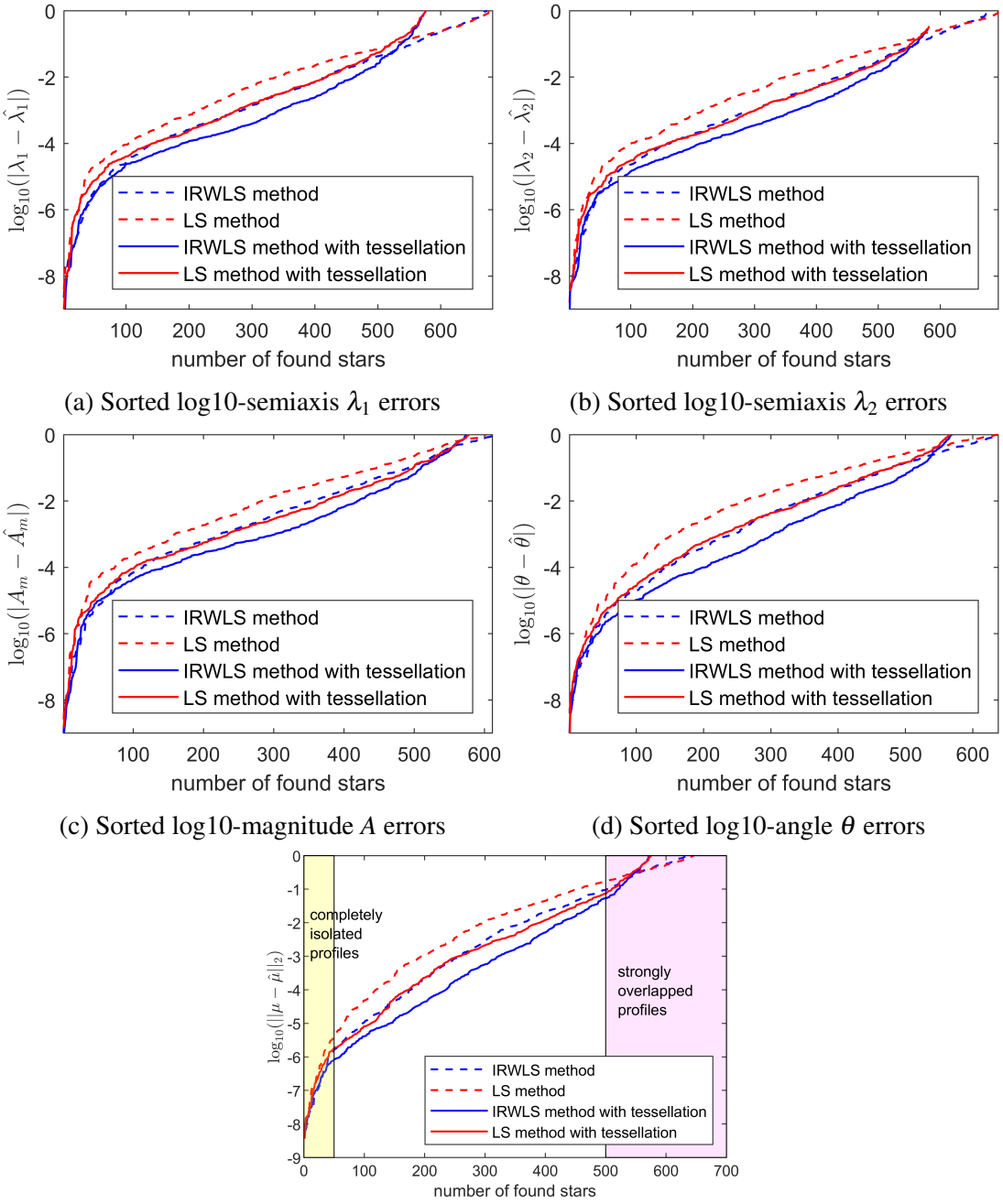


(c) Residual error of the LS method with tessellation

(d) Residual error of the LS method (zoomed red square)

Figure 3.15: Residual errors obtained by the IRWLS and LS methods with tessellation (to emphasize small error values, $|\mathbf{e}|^{1/4}$ is shown).

The ROI was determined either as a convex region around the initial center position of the expected profile's width without tessellation (e.g., $\pm 3\sigma$ around the initial center position) or



(a) Sorted log10-semiaxis λ_1 errors

(b) Sorted log10-semiaxis λ_2 errors

(c) Sorted log10-magnitude A errors

(d) Sorted log10-angle θ errors

(e) Sorted log10-distances of estimated and actual center positions

Figure 3.16: Sorted log10-absolute errors of estimated parameters for two object extraction methods: 1) iterative object removal based on image maximum value (dashed lines) and 2) sorted list of local extrema of the blurred image with generalized linear tessellation (solid lines)

convex regions obtained by using described generalized Voronoi tesselation. Estimated valid models were subtracted from the whole input image, while invalid models were discarded. Profiles were iteratively estimated and removed similarly to the matching pursuit method. Fig. 3.15 shows residual errors of the IRWLS and LS methods when the generalized Voronoi tessellation is used.

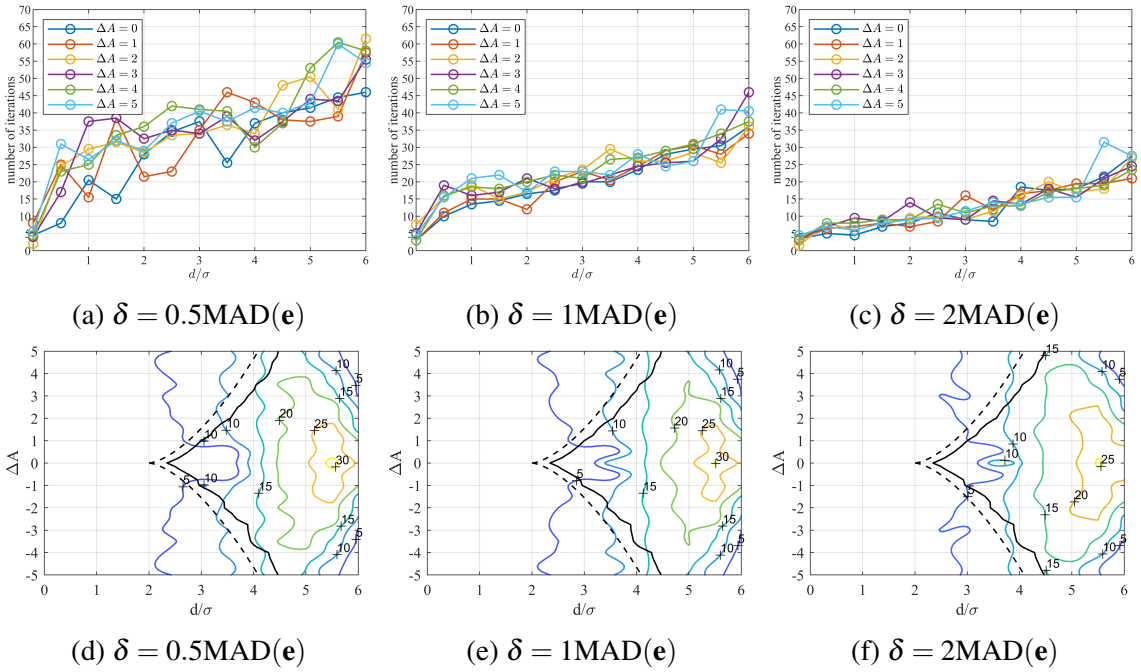
Even if residual errors in Fig. 3.15b and Fig. 3.15d look quite similar, differences can be seen in the zoomed regions. Namely, the LS method produced more false ridges than the IRWLS method and, consequently, the worse parameter estimates. The sorted log10 errors of $\hat{\lambda}_1$, $\hat{\lambda}_2$, $\hat{A}_m$, $\hat{\theta}$ and the L_2 distances of estimated and given centroid positions are shown in Fig. 3.16. Results show that the IRWLS method achieves up to one order of magnitude modeling gain compared to the LS method, even with previously applied segmentation. As Fig. 3.16e shows, three regions of sorted profiles' parameter errors exist. In the left region of Fig. 3.16e are smaller errors of firstly estimated brightest stars which are completely isolated profiles, and the accuracy of the IRWLS method in their estimation is similar to the LS solution. In the right region of Fig. 3.16e are the fainter estimated stars which are almost completely overlapped, and both methods yield similar results again. However, in the central zone of Fig. 3.16e, the method modeling gain of the IRWLS method is significant for a large number of found objects for all estimated parameters.

3.4.5 Influence of the Huber estimator tuning constant on the performance of the IRWLS method

According to [66], the tuning constant k_{const} can be chosen in the range from 1 to 2. Still, the common practice is to select $k_{const} = 1.345$ to achieve high asymptotic efficiency when residual errors follow a normal distribution. The Huber estimator is more resilient to outliers for smaller values of k_{const} at the expense of lower asymptotic efficiency, while it is closer to the LS estimate for larger values of k_{const}. The selection of factor k_{const} represents the trade-off between the L_1 and L_2 norms, i.e., the outlier rejection and asymptotic efficiency. However, the optimal tuning factor k_{const} should be data-driven from the histogram of errors. The assumption is that such a histogram is bi-modal with smaller errors normally distributed and larger errors heavy-tailed distributed and that these two modes can be separated.

The Monte Carlo simulation with the same setup was performed for the tuning constants from the set $k_{const} = \{0.5, 1, 1.345, 2\}$ to analyze the influence of k_{const} on the convergence rate and the efficiency of the IRWLS method. For each combination of d/σ, ΔA, and k_{const}, the image with 20 sub-images was synthesized. Each sub-image contained two rotationally symmetric Gaussian profiles of the widths $\lambda_1 = \lambda_2 = \sigma = 1$. As before, the fractional shifts of the first profile within ± 0.5pm were introduced, as well as the random angular position of the second profile, which is at a distance d from the first one. The experimental results were published in [69]. The results in Fig. 3.17 show that the convergence rate of the IRWLS method for overlapped Gaussian profiles depends on the tuning factor k_{const}. A smaller coefficient k_{const} reduces the convergence rate of the method. For the case of $k_{const} = 0.5$, the convergence rate is reduced approximately two times compared to the case of $k_{const} = 2$. The experimental results

show that the modeling error is not very sensitive to the tuning factor k_{const} if it is in the range from 1 to 2. The smaller k_{const} yields a solution closer to the L_1 solution, while the larger k_{const} yields a solution closer to the L_2 norm.



(a) $\delta = 0.5\text{MAD}(\mathbf{e})$ (b) $\delta = 1\text{MAD}(\mathbf{e})$ (c) $\delta = 2\text{MAD}(\mathbf{e})$

(d) $\delta = 0.5\text{MAD}(\mathbf{e})$ (e) $\delta = 1\text{MAD}(\mathbf{e})$ (f) $\delta = 2\text{MAD}(\mathbf{e})$

Figure 3.17: The influence of the tuning constant k_{const} on the convergence rate and modeling gain of the IRWLS method

3.5 Robustness of the IRWLS method to different noise types

The proposed IRWLS method with Huber weights yields good results in estimating overlapping stellar components in astronomical images without other noise sources. To adapt the proposed IRWLS method to different noise types in the sense of finding optimal estimation weights, it is essential to understand noise sources that occur during digital image acquisition and their distribution models. In this section, three different types of noise distribution models were considered: additive Gaussian noise, multiplicative Poisson noise, and heavy-tailed noise. As two types of estimation weights, Huber and inverse variance weights were considered.

In [58] and [59], 1D and 2D Gaussian profiles were fitted in maximum likelihood sense from data corrupted with additive Gaussian noise. Also, they provided the analytical expressions for parameter variances in the case of Gaussian and Poisson noise using the Cramer-Rao bound. In [61] and [62], the 1D Gaussian profile is fitted in the argument domain of exponential from data contaminated with Gaussian noise through the iterative procedure with the prior knowledge of noise variance. The advantage of the proposed IRWLS method with Huber weights is that such weights do not require prior knowledge of noise statistics since they are based on robust

statistics of residual errors. The assumption is that they successfully solve heavy-tailed noise contamination, such as the contamination by overlapping adjacent objects, as it is described in the previous section. On the other hand, the inverse variance weights require prior knowledge of the variance of each sample, but they can be applied in the case of Poisson noise contamination.

3.5.1 Noise sources

Digital image acquisition corresponds to converting the incoming light into digital pixel values by commonly using the CMOS or CCD imaging sensors [70]. The incoming light is firstly transformed to charge. Afterward, the accumulated charge is transformed into the voltage output. Finally, the A/D conversion is applied to the voltage output to obtain the quantized digital pixel values, as Fig. 3.18 shows. The noise sources present in digital image acquisition can be divided into random and spatial noise. Random noise is time-variant, while spatial noise can be time-invariant and time-variant [71].

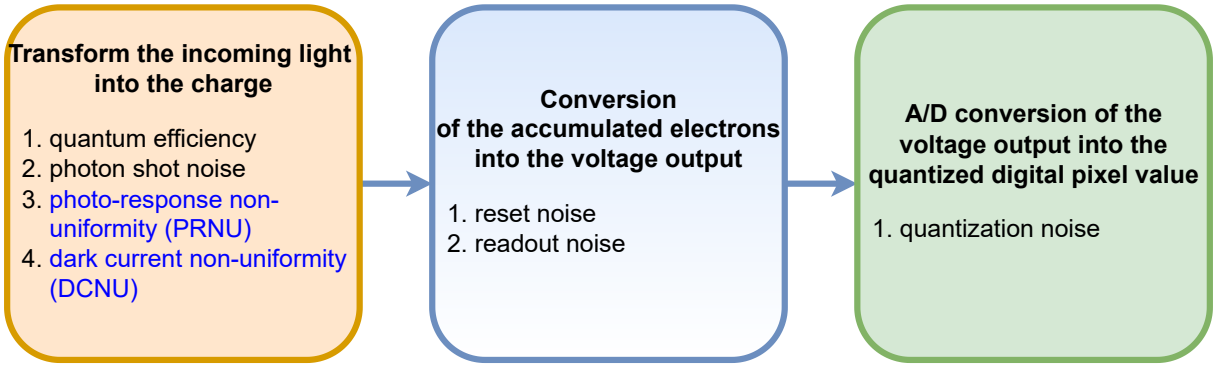


Figure 3.18: Noise sources in digital image acquisition

Random noise

The first acquisition stage is characterized by the ability of the photosensor to transform the incoming light into the charge, known as quantum efficiency [72]. In the first acquisition stage are random photon and dark current shot noise, which follow Poisson distribution [72]. Photon shot noise describes the detection of the photons on the photosensor, while the dark current shot noise is related to undesirable electron generation in the photosensor, even in the absence of light, and raises with the temperature and exposure time.

In the second readout stage, the accumulated electrons are transformed to the voltage output for each pixel as the difference between the initial referent voltage in the absence of the light and the final after-exposure voltage. In this stage are present two noise types that are normally distributed: the reset noise, which is caused by the uncertainty of the referent voltage, and readout noise, which is thermally generated [73].

Spatial noise

The spatially variant noise includes fixed pattern noise (FPN), bad or hot pixels, etc [71]. The FPN differs for CCD and CMOS sensors due to the specifics of transformations of accumulated charge into the voltage and quantization of the output voltage into the digital values. Namely, in CCD sensors, the accumulated charge is transferred across the sensor to the output amplifier, where it is converted to voltage and quantized. Unlike this, in CMOS sensors, each sensor column has its amplifier for charge-to-voltage conversion, increasing FPN noise. FPN includes the photo-response non-uniformity (PRNU) due to variations in generated electrons in pixels under the same illumination and dark current non-uniformity (DCNU) due to variant dark current generation in different sensor pixels [70]. The DCNU follows the heavy-tailed distribution since hor or bad pixels have to be taken into account [72].

In the third A/D phase, the uniform quantization noise is present, which is almost negligible compared to other noise sources.

Inverse variance weights

The inverse variance weights are calculated as

$$w_i(\sigma_i^2) = w_I(\sigma_i^2) = \frac{1}{\sigma_i^2 \sum_{i=1}^{N} \left(\frac{1}{\sigma_i^2}\right)}, \tag{3.32}$$

where σ_i corresponds to the standard deviation of the ith pixel value, while N corresponds to the total number of samples within the selected ROI.

For data contaminated with Poisson noise, the standard deviation of the value of the ith pixel is equal to the square root of the ideal noise-free sample value and can be approximated by the square root of the model value as

$$\sigma_i \approx \sqrt{f(\mathbf{x}_i, \hat{\beta})}. \tag{3.33}$$

The IRWLS method with inverse variance weights approximated by model values will converge to the optimal solution in the case of Poisson noise, but only with the initialization of the model close to the actual solution. If the variance of each sample is known in advance and its inverse is used as the weight, the weighted least squares (WLS) method provides the optimal solution without the need for further iterations.

3.5.2 IRWLS method results in 2D Gaussian profile estimation from noisy data

In the first experiment, the single 2D Gaussian profile parameters were estimated from data contaminated with Poisson and heavy-tailed noise. In both experiments, the 2D Gaussian profile

of the given parameters $\beta = [A_l, \lambda_1, \lambda_2, \mu_x, \mu_y, \theta] = [10^5, 2, 1, 0, 0, \pi/8]$ was firstly synthesized, but with random fractional shifts within ± 0.5 pix to simulate the random profile's sampling over a regular grid. The estimation ROI was chosen as an elliptic region around the profile's peak position within Mahalanobis distance $k = 3$. The vector of noisy measurements $\mathbf{z}$ for all pixels within the ROI was obtained so that the noise $\mathbf{n}$ was scaled with the factor α and added to the noiseless synthesized profile's values $\bar{\mathbf{z}}$ as $\mathbf{z} = \bar{\mathbf{z}} + \alpha\mathbf{n}$ to satisfy the required SNR in dB within the ROI according to

$$SNR = 10\log_{10}\frac{\sum_{i\in ROI}\bar{z}_i^2}{\sum_{i\in ROI}\alpha^2 n_i^2}\ [\text{dB}]. \tag{3.34}$$

where i denotes the ith pixel within the ROI.

The chosen SNR was 30 dB in both experiments. The given profile's parameters were passed to the optimization procedure to avoid the influence of wrong initialization on the final estimate. The maximum number of outer loop iterations of the IRWLS method was set to $t_{max} = 50$, while optimization termination tolerance was set to $tol = 10^{-10}$.

As a measure of accuracy, the total modeling error between the estimated and given profile was used, which can be expressed as

$$e_{total} = 10\log_{10}\frac{\sum_{j\in eval}(f(\mathbf{x}_j, \hat{\beta}) - f(\mathbf{x}_j, \beta))^2}{\sum_{j\in eval} f(\mathbf{x}_j, \beta)^2}\ [\text{dB}] \tag{3.35}$$

where $f(\mathbf{x}_j, \hat{\beta})$ corresponds to the estimated model value while $f(\mathbf{x}_j, \beta)$ corresponds to the given model value of the jth pixel position $\mathbf{x}_j$, which is within the selected evaluation region $eval$. The evaluation region was selected as the estimation ROI but upsampled to better approximate the integral of the continuous error function over the evaluation region. The 2D Gaussian profile parameters were estimated using the LS method, IRWLS method with Huber weights according to (3.13), and IRWLS method with normalized inverse variance weights according to (3.32). The efficiency of these three methods in the profile estimation from noisy data was compared, and the results were published in [74].

Estimation of the 2D Gaussian profile from data contaminated with heavy-tailed noise

In the first experiment, the heavy-tailed noise was added to the 2D Gaussian profile of the given vector of parameters which include the random fractional shift of the profile's peak position: $\beta = [A_l, \lambda_1, \lambda_2, \mu_x, \mu_y, \rho] = [10^5, 2, 1, -0.3082, -0.2817, 0.7854,]$. Fig. 3.19 shows the estimation of the 2D Gaussian profile parameters from data contaminated with heavy-tailed noise by using the IRWLS method with Huber weights, the WLS method with inverse variance weights, and the LS method with unit weights. The long-term statistics of the heavy-tailed noise, i.e., the expected variances of noise samples were generated from generalized Pareto distribution of the shape parameter $\varepsilon = 3$, the scale parameter $\sigma = 1$, and the location parameter $\mu = 0$, using

the MATLAB *random* function. Then, the random noise samples were drawn from the normal distribution of zero mean and unit variance for all pixels within the ROI. Their absolute values were multiplied by the square root of the long-term heavy-tailed noise statistics. Such obtained noise samples were finally scaled with an appropriate factor α to achieve the required SNR in dB.

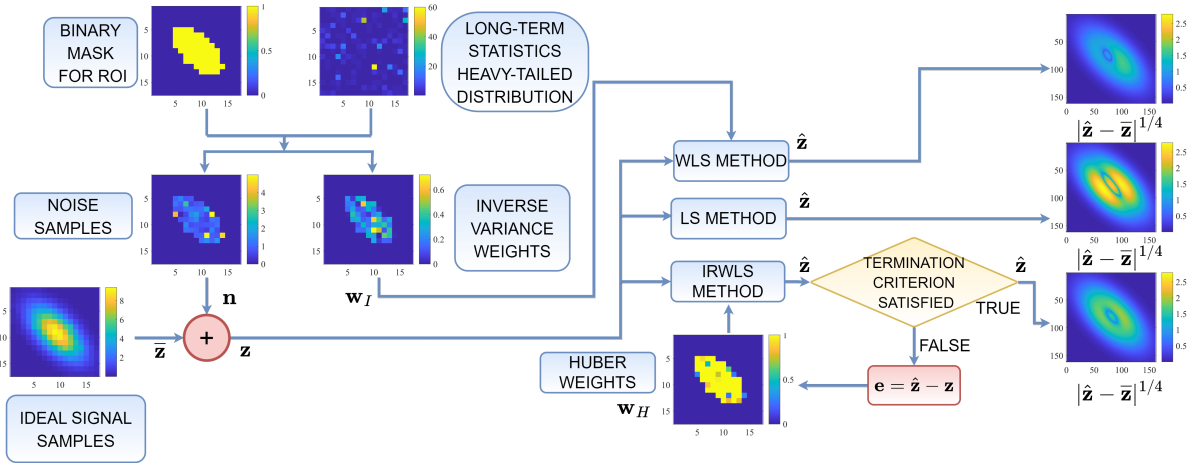
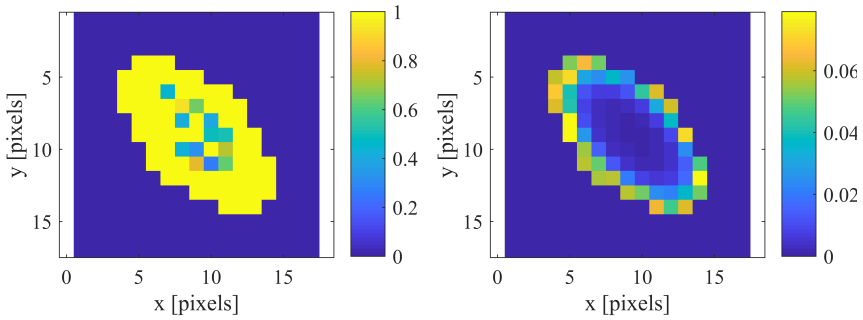


Figure 3.19: The absolute residual errors of the IRWLS method with Huber weights, the WLS method with inverse variance weights, and the LS method in 2D Gaussian profile estimation from data contaminated with heavy-tailed noise

IRWLS method with Huber weights proved to be successful in solving the problem of heavy-tailed noise with high accuracy and without any prior knowledge of noise statistics. The WLS method with inverse variance weights calculated yields a more optimal solution but the noise variance of each sample has to be known in advance, i.e., the spatially variant long-term noise statistics. The total modeling error of the IRWLS method with Huber weights was -64.2273 dB with the estimated vector of the 2D Gaussian profile's parameters $\hat{\beta}_H = [1.0008 \cdot 10^5, 2.0005, 1.0006, 0.7855, -0.3083, -0.2816, 0.7855]$, while the WLS method with the specified inverse variance weights yielded the total modeling error of -82.8051 dB with the estimated vector of parameters $\hat{\beta}_I = [1.0001 \cdot 10^5, 2.0001, 1.0000, -0.3082, -0.2817, 0.7854]$. The LS method results were the worst with the total modeling error of -36.4718 dB with the estimated vector of parameters $\hat{\beta}_{LS} = [1.0136 \cdot 10^5, 1.9950, 1.0175, -0.3199, -0.2864, 0.7823]$.

Poisson noise contamination

In the second experiment, the Poisson noise was applied to the profile with the given parameters: $\beta = [A_I, \lambda_1, \lambda_2, \mu_x, \mu_y, \rho] = [10^5, 2, 1, -0.3639, -0.3530, 0.7854]$. The Poisson noise was generated such that the random samples were drawn from the normal distribution of zero mean and unit variance, and then they were multiplied with the square root of ideal signal values for all pixels within the ROI. Such obtained noise samples were multiplied with the factor α to obtain the required SNR.



(a) Huber weights within the ROI (b) Normalized inverse variance weights within the ROI

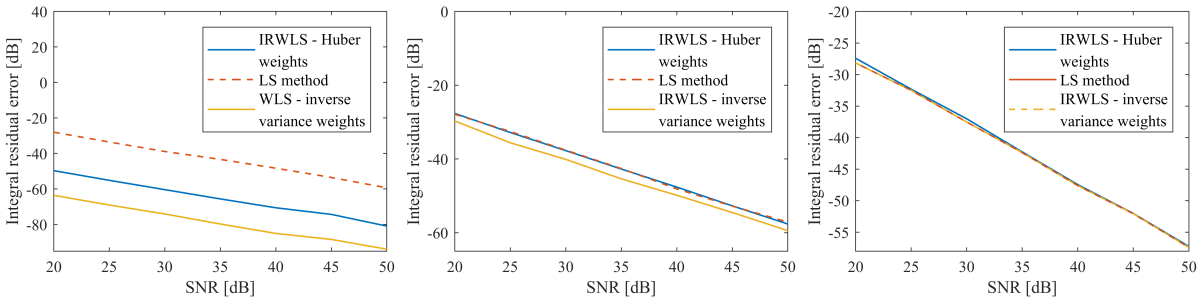
Figure 3.20: Estimation weights in the last iteration of the IRWLS method when the profile was contaminated with **Poisson noise**. The pixel size was 1, and the sampling grid was uniform. The maximum number of outer loop iterations was 50.

The IRWLS method with Huber weights accomplished a total modeling error of -36.9081 dB with estimated parameters $\hat{\beta}_H = [9.9343 \cdot 10^4, 1.9929, 0.9889, -0.3695, -0.3721, 0.7759]$. The better results of -42.7928 dB achieved the IRWLS method with inverse variance weights, calculated in each iteration from the current estimated model values. The estimated vector of the profile's parameters with the IRWLS method with inverse variance weights was: $\hat{\beta}_I = [9.9719 \cdot 10^4, 2.0083, 0.9985, -0.3583, -0.3573, 0.7837]$. The LS method yielded the estimated vector of parameters $\hat{\beta}_{LS} = [9.9374 \cdot 10^4, 1.9961, 0.9931, -0.3678, -0.3734, 0.7739]$ with the largest error of -36.8037 dB. Fig. 3.20 shows Huber weights and inverse variance estimation weights in the last IRWLS method's iteration. The inverse variance weights, calculated in each iteration from the current estimated model, reduce the contribution of larger samples with larger variances which are close to the profile's peak position and emphasize the contribution of samples at the ROI edge where the samples' intensities, as well as variances, are much smaller. The weights at the ROI edge are variable due to coarse discretization and the fractional shifts of the profile. Huber weights are units for most of the regular samples, except for a smaller percentage of central pixels with large errors whose weights are reduced. Even though such Huber weights correspond to the characteristics of Poisson contamination, they are less efficient than the inverse variance weights. However, even if the IRWLS method with inverse variance weights yields the best results, the total modeling errors of all three methods differ by only a few dBs.

Monte Carlo simulation

Monte Carlo simulations with 100 trials were performed for each combination of three noise distributions (heavy-tailed, Poisson, and Gaussian distributions), three types of estimation weights (Huber weights, inverse variance weights, and unit weights), and different SNRs within the ROI in the range from 20 dB to 50 dB with the stepsize of 5 dB to analyze the efficiency of the IR-

WLS method. In each of the 100 trials, the profile with the same parameters as in the previous experiment was synthesized but with the random fractional shift within ± 0.5 pix to simulate random uniform sampling. Fig 3.21 shows medians of estimation modeling errors in 100 trials. The IRWLS method with Huber weights yields a modeling gain of up to 25 dB compared to the LS method, while the WLS method with inverse weights yields even better, close to the optimal solution, but with the priory known long-term noise statistics. The IRWLS method with model-driven inverse variance weights produces a minor modeling error (up to 3 dB better than the LS solution) since such weighting is an MLE solution for Poisson noise. In the case of Gaussian noise, the LS method yields an optimal solution in MSE and MLE sense. The IRWLS method with inverse variance weights yields the same solution as the LS method since all pixels have the same variances.



(a) Total modeling error for heavy-tailed noise contamination

(b) Total modeling error for Poisson noise contamination

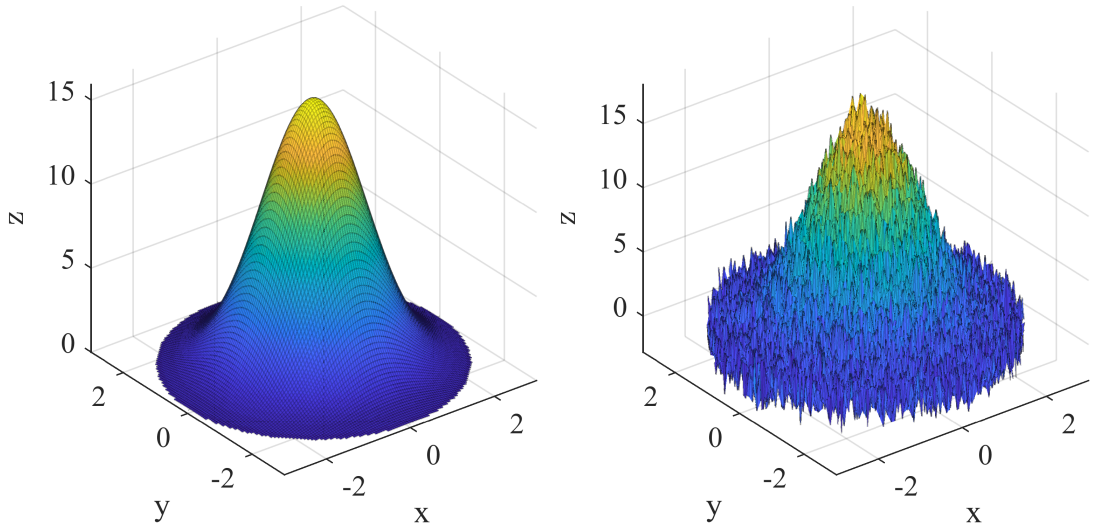
(c) Total modeling error for Gaussian noise contamination

Figure 3.21: Comparison of the IRWLS method with Huber weights, the IRWLS method with inverse variance weights, and the LS method for different SNRs in dB and different noise distributions.

3.6 Optimal ROI width for maximizing 2D Gaussian profile estimation accuracy

The appropriate selection of the estimation input domain's shape and width is crucial for the accurate parameter estimation of the 2D Gaussian profile from the noisy data since it defines the subset of input samples' positions and values which enter the estimation process. Some papers propose selecting the estimation ROI as a square region of fixed size centered at the coarsely estimated profile's peak position [37]. However, in section 3.3, we defined the estimation ROI as a circular or elliptical region around the profile's peak position whose radius corresponds to the product of the profile's width, i.e., its standard deviation σ, and the factor of Mahalanobis distance k. Namely, the Mahalanobis distance k represents the profile's positions of the same values. Such a selected ROI is circular when a profile's semiaxes have the same widths. In contrast, such an ROI is an elliptic convex region for a profile with different semiaxes widths and a full covariance matrix.

Namely, it is noticed that the chosen estimation ROI affects even the estimation of a single 2D Gaussian profile from a fixed number of random input samples contaminated with additive Gaussian noise. In the case of a single 2D Gaussian profile corrupted with additive Gaussian noise, the LS method yields the optimal parameters estimate since it is the ML solution for such contamination type. Still, the accuracy of the solution differs for different ROI widths due to variations in informativity contained in input data. To find the optimal ROI width, the rotationally symmetric 2D Gaussian profile was considered, which is defined in (3.14). The examples of the ideal rotationally symmetric Gaussian profile and the profile contaminated with additive Gaussian noise are shown in Fig. 3.22.



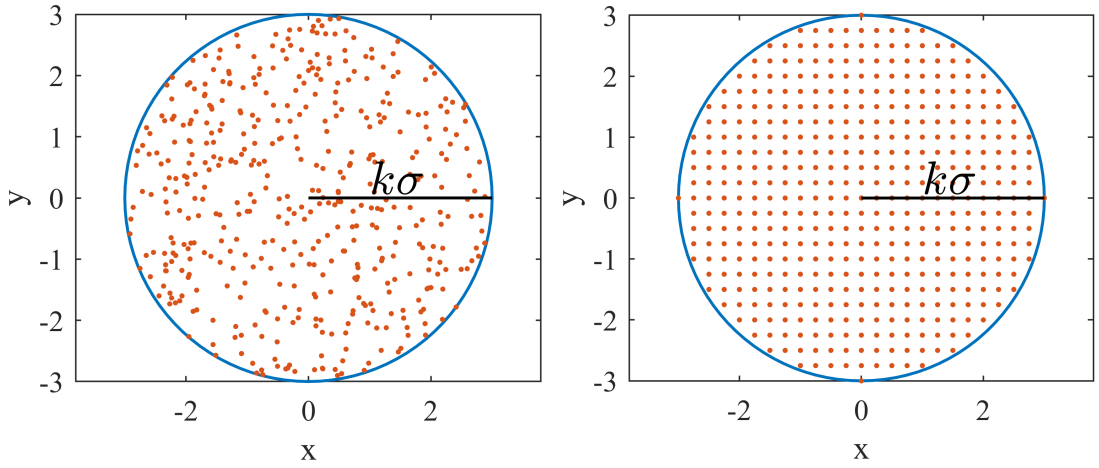
(a) Ideal circular 2D Gaussian profile

(b) Input noisy measurements

Figure 3.22: Circular 2D Gaussian profile in polar coordinates $z(r,\theta) = A\exp\left(\frac{-r^2}{2\sigma^2}\right), z \in [A\exp(-k^2/2), A]$, where $\sigma = \sigma_x = \sigma_y$ is the profile's standard deviation and $r \in [0, k\sigma]$ denotes the radius of circular estimation input domain (ROI)

Both parameters that define the radius of the estimation ROI, the profile's standard deviation σ, and the factor of Mahalanobis distance k have to be precisely determined for accurate parameter estimation of the noisy circular 2D Gaussian profile. If the profile's σ is overestimated, the invalid pixels which belong to the background or adjacent objects enter the estimation process and cause false object identification. On the other hand, the underestimated profile's σ causes the discarding of valid pixels and wrong parameter estimation. Similarly, the smaller values of factor k cause the narrow ROIs that include a small number of pixels around the profile's peak position. In that way, the influence of adjacent objects and the background is avoided. Still, the solution can be highly affected by Poisson noise which is dominant in that region of large profile values. Finally, the trade-off between these two effects can be achieved by finding the optimal value of factor k while assuming that the profile's standard deviation σ is known in advance.

This subsection analyzes the accuracy of the circular 2D Gaussian profile estimation from data contaminated by additive Gaussian noise using the standard LS method. Thereby, two types of estimation ROI sampling were considered: random and uniform. In the case of random sampling, the number of input samples for different ROI widths was fixed, and the required SNR was ensured within the ROI. The input sample positions were considered independent random variables uniformly distributed in both directions whose joint PDF was given in (3.15). In the case of uniform sampling in both directions, the number of input samples varies with different ROI widths. Two sampling types within circular ROI are shown in Fig. 3.23.



(a) Random sampling of circular ROI (b) Uniform sampling of circular ROI

Figure 3.23: Two types of ROI sampling: **random** with a fixed number of uniformly distributed samples and **uniform** with a fixed density of input samples. The estimation ROI is defined as a circular region around the profile's peak position of the radius $k\sigma$.

To explain the differences in accuracy of the LS method for different ROI widths for the constant number of input samples and given SNR, the concept of differential entropy [75] is introduced. Differential entropy is an extended and modified version of Shanon's discrete entropy but for continuous random variables. It is a measure of information, i.e., the randomness of a continuous random variable, and determines the number of bits required for its description. In subsection 3.4.4 is proved that the 2D Gaussian profile values follow the log-uniform distribution. To find the optimal ROI width, differential entropy for the log uniform distribution is derived and maximized concerning the Mahalanobis distance k. The assumption is that the reduction of LS method accuracy for different values of factor k can be predicted from the decrease in differential entropy.

3.6.1 Differential entropy of the log-uniform distribution of Gaussian profile's values

Differential entropy is a relative measure of change in information or randomness of continuous random variables concerning some parameter and can have negative values. High differential entropy implies a large dispersion of sample values within the sample space and vice versa [75].

If positions of input samples are random variables uniformly distributed within the circular ROI with the joint PDF (3.4.4), the circular 2D Gaussian profiles values of those samples follow the log-uniform distribution according to (3.26), where the maximal profile's value is at the ROI center where $r = 0$ and $z = A$, while the minimal profile's value is at ROI edge where $r = k\sigma$ and $z = A\exp(-k^2/2)$. Examples of histograms of Gaussian profile values for different factors of Mahalanobis distance k and corresponding PDFs are shown in Fig. 3.24.

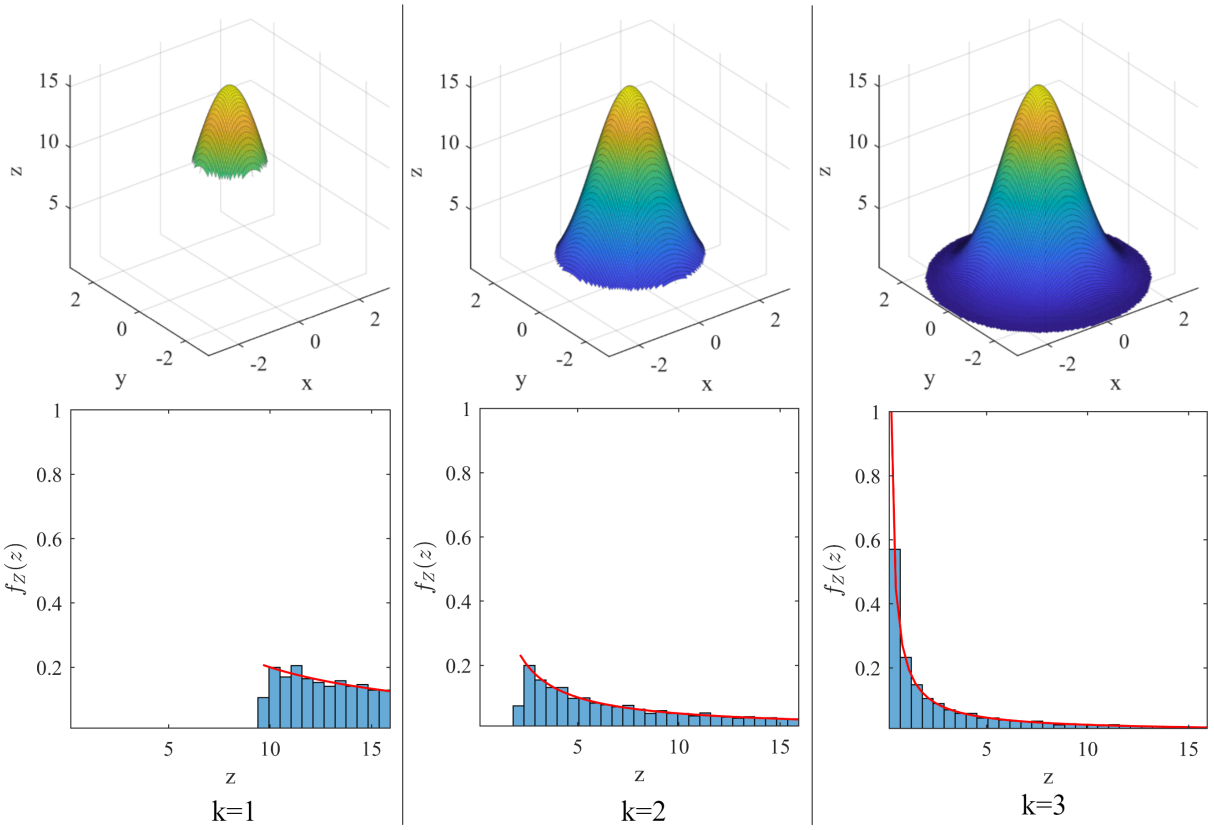


Figure 3.24: Normalized histograms and corresponding log-uniform PDFs of 2D Gaussian profile's values ($f_Z(z) = 2/(zk^2)$) enclosed within ROIs of different widths determined by the factor of Mahalanobis distance k

Generally, the differential entropy of random variable X is defined as the expectation of the negative logarithm of its PDF as follows

$$h(X) = E[-\log(f(X)] = -\int_S f(x)\log_2 f(x)dx \text{ [bits]} = -\int_S f(x)\ln f(x)dx \text{ [nats]} \qquad (3.36)$$

where S denotes the support set of random variable X.

If $a = A\exp(-k^2/2)$ and $b = A$, the differential entropy of 2D Gaussian profile's values $h(Z)$ is

$$h(Z) = -\int_a^b f_Z(z)\ln f_Z(z)dz = \ln\left(\frac{bk^2}{2}\right) - \frac{k^2}{4} \text{ [nats]}. \tag{3.37}$$

It is expected that narrower ROI will contain samples with values close to the profile's peak value but of small variance and, consequently, of slight randomness and informativeness. Contrary, wide ROIs which are equal to or larger than $r = 3\sigma$ contain a large number of samples close to the additive noise values with less informativeness. To find the optimal ROI width, which contains the maximal information about the profile's parameters, differential entropy has to be maximized concerning the factor of Mahalanobis distance k

$$\frac{\partial h_Z}{\partial k} = \frac{-(k^2-4)}{2k} = 0. \tag{3.38}$$

It turns out that the maximal differential entropy is for $k = 2$ and equals

$$h_{Z_{max}} = h_Z(k=2) = \ln(2A) - 1. \tag{3.39}$$

Differential entropy reduction for different values of factor k can be calculated as

$$\Delta h_{Z_{nats}} = h_Z - h_{Z_{max}} = \ln\left(\frac{k^2}{4}\right) - \frac{k^2}{4} + 1, \tag{3.40}$$

$$\Delta h_{Z_{bits}} = \Delta h_{Z_{nats}} / \ln 2, \tag{3.41}$$

$$\Delta h_{Z_{dB}} = \Delta h_{Z_{bits}} \cdot 20\log_{10} 2. \tag{3.42}$$

The difference in differential entropy for different ROI widths depends only on factor k. It indicates how much SNR should be increased compared to the case when $k = 2$ to preserve the same relative ratio between the information contained in samples' values and noise.

In the case of additive noise with spatially uniform distribution within ROI, the nominal SNR can be defined as the ratio of the power of the signal, which corresponds to the squared maximal profile's value A, and the power of noise, which corresponds to the noise variance σ_n^2, as follows

$$SNR_{dB} = 10\log_{10}\frac{b^2}{\sigma_n^2} = 20\log_{10}\frac{b}{\sigma_n}. \tag{3.43}$$

To account for the reduction of differential entropy for non-optimal factors k, the noise standard deviation σ_n has to be reduced according to

$$\sigma_{n_{reduced}} = \sigma_n \cdot 10^{(\Delta h_{z_{dB}}/20)} \tag{3.44}$$

The differential entropy of additive noise normally distributed is only a function of σ_n and is constant regardless of factor k.

The reduction of differential entropy in dBs for different factors k compared to the optimal case when $k = 2$ is shown in Fig. 3.25. For $k = 0.39$ and $k = 4.42$, $\Delta h_Z = -20$dB compared to the optimal case of $k = 2$, thus meaning that the standard deviation of noise has to be 10-fold reduced to preserve the same ratio of informativity in the profile's values and noise.

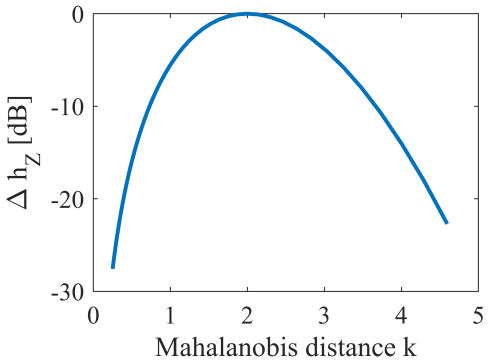


Figure 3.25: Differential entropy reduction compared to the nominal case for $k = 2$

SNR in dBs is a conventional approximate measure of the informativeness of noisy data for a certain signal class. It is defined as the logarithm of the ratio of the signal variance to the noise variance. SNR, as a measure of the information content of noisy data, can be used for signals whose differential entropy is only a function of variance, such as normally distributed signals and noise. However, the values of the Gaussian profile follow a log-uniform distribution whose differential entropy strongly depends on the chosen Mahalanobis distance factor k, so SNR is a too simple measure of the information content of the Gaussian profile values with interference.

By combining the expressions (3.30) and (3.37), we can relate the variance of the log-uniform distribution of Gaussian profile's values and the corresponding differential entropy as

$$\ln\sqrt{\mathrm{Var}(Z)} = h_Z + \frac{\ln\left(k^2\exp\frac{k^2}{2} - 4\exp\frac{k^2}{2} + k^2 + 4\right)}{2} \tag{3.45}$$

$$+ \frac{\ln\left(4\exp\frac{k^2}{2} - 4\right)}{2} - 4\ln k - \frac{k^2}{4}, \tag{3.46}$$

which shows that the logarithm of the standard deviation of the profile's values and differential entropy differ in the offset that depends solely on chosen radius k. This difference is shown in Fig. 3.26. For ROI widths from $k = 0$ to $k = 4.56\sigma$, the difference between the actual differential entropy and its predictions calculated as the logarithm of the standard deviation of the profile samples' values are ± 10.8dB, while equals zero for $k = 3.684$. Consequently, instead of using variance-based SNR to measure information content, the actual derived ROI-dependent differential entropy expression must be considered to circumvent such predicted error

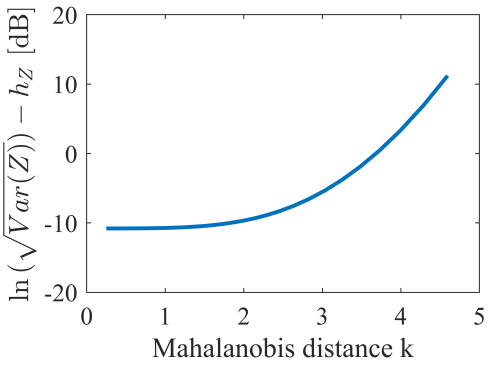


Figure 3.26: Difference between the natural logarithm of the square root of the variance and the differential entropy in nats for different factors of Mahalanobis distance k

of estimation performance.

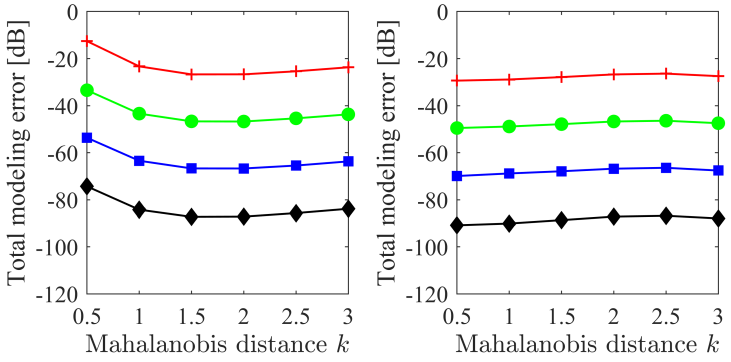
3.6.2 Experiments and results

To analyze the accuracy loss of the LS method in 2D Gaussian profile estimation from noisy data for different non-optimal ROI widths ($k \neq 2$) and to verify that the accuracy loss corresponds to a reduction of differential entropy, two experiments were made, as it is described in [76]. In both experiments, Monte Carlo simulations were performed, where in each trial, the random measurements were generated as a combination of the ideal 2D Gaussian profile values and random additive noise. As a measure of accuracy, the total modeling error from (3.35) was used where the estimated and given models were compared over the chosen evaluation grid. The evaluation grid was selected as uniformly upsampled estimation ROI within 3σ where the samples are displaced for $\sigma/10$ in both directions, where σ corresponds to the given profile's width. The given parameters of the 2D Gaussian profile were $[A_l, \lambda_1, \lambda_2, \mu_x, \mu_y, \theta] = [100, 1, 1, 0, 0, 0]$. The simulations were performed for all combinations of the given SNR from the set $SNR = \{20, 40, 60, 80\}$ and factor of Mahalanobis distance from 0.5 to 3 with the stepsize 0.5, w/wo differential entropy compensation. Additive noise was added to synthesized ideal sample values to satisfy the required SNR within the selected ROI. For the case of differential entropy compensation, the standard deviation of noise σ_n is calculated for the given SNR according to (3.43). It is additionally reduced for the predicted reduction of differential entropy compared to the optimal case for $k = 2$ according to (3.44).

Random sampling of circular ROI

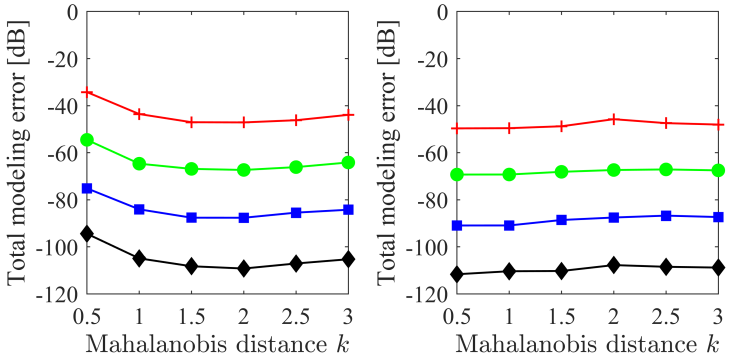
In each trial of the first experiment, 100 and 10000 input samples were randomly picked within the selected ROI, which is determined by the factor k. The random samples' positions were uniformly distributed within the ROI. The number of input samples was fixed, while their density decreased for wider ROIs. For each considered combination of factor k and SNR, 5000 and 50

trials were performed for 100 and 10000 samples, respectively, to ensure the constant product of trials and the number of input samples, and consequently, the constant variance of estimation error. Results are shown in Fig. 3.27. Figs. 3.27a and 3.27c show the mean modeling errors of the LS method without differential entropy compensation. The LS method accuracy varies with k, and it is the smallest for narrow ROIs due to the sample values close to the profile's peak value and of small information content. A similar effect can be seen for wide ROIs where most of the samples have small values close to the additive noise values and, consequently, without much information about the profile's parameters. Estimation errors in Fig. 3.27c are only reduced for 20 dBs ($10\log_{10}100$) due to a 100-fold increase in number of input samples and consequently reduction of noise standard deviation for $\sqrt{100}$. Figs. 3.27b and 3.27d show the results after the compensation of differential entropy reduction. The mean modeling errors became independent of the ROI width, i.e., of the selected factor k, thus proving that the derived expressions for differential entropy reduction of Gaussian profile's values precisely predict the loss of the LS method accuracy.



(a) n_s=100, results **without** differential entropy compensation

(b) n_s=100, results **with** differential entropy compensation



(c) n_s=10000, results **without** differential entropy compensation

(d) n_s=10000, results **with** differential entropy compensation

Figure 3.27: Mean modeling errors in dB for different ROI widths and nominal SNRs denoted with: red - 20 dB, green - 40 dB, blue - 60 dB, black - 80 dB. The number of random input samples is denoted with n_s.

Uniform sampling of circular ROI

In the second experiment, the sampling within the ROI was uniform, with the same pixel size in both directions ($dx = dy$) to ensure the fixed density of samples within the ROI of different widths. The number of input samples consequently rises quadratically with the factor k and the ratio of the profile's width σ and pixel size ($dx = dy$) according to expressions

$$P_{ROI} = (k\sigma)^2, \tag{3.47}$$

$$P_{pix} = dx \cdot dy, \tag{3.48}$$

$$N = \frac{P_{ROI}}{P_{pix}} = \frac{k^2\sigma^2\pi}{dxdy} = k^2 f_\sigma^2 \pi,\ f_\sigma = \frac{\sigma}{dx} = \frac{\sigma}{dy}, \tag{3.49}$$

where P_{ROI} and P_{pix} denote the area of estimation ROI and pixel. Pixel size was selected as $dx = dy = 0.25$, while the given profile's width $\sigma = 1$. Such discretization ensures sufficient input samples for the optimization procedure, even for the narrowest ROI ($k = 0.5$). The number of input samples has to be at least equal to the number of unknowns, but such a solution is, in that case, very sensitive to added noise. The Monte Carlo simulation with 50 trials was performed for each combination of considered (k, SNR) pairs w/wo compensation of differential entropy reduction. Fig. 3.28 shows the quadratic increase in input samples with the factor k. Therefore, the expected total modeling error decreases proportionally to $10\log_{10}(N_1/N_2) = 20\log_{10}(k_1/k_2)$ where $k_1 > k_2$ and $N_1 > N_2$.

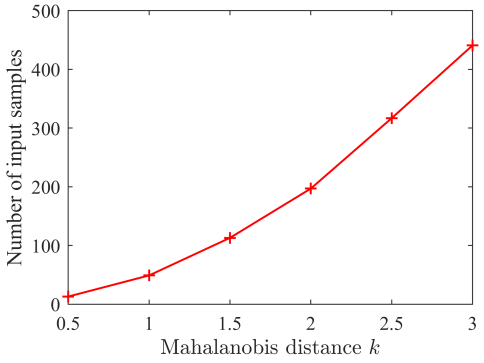
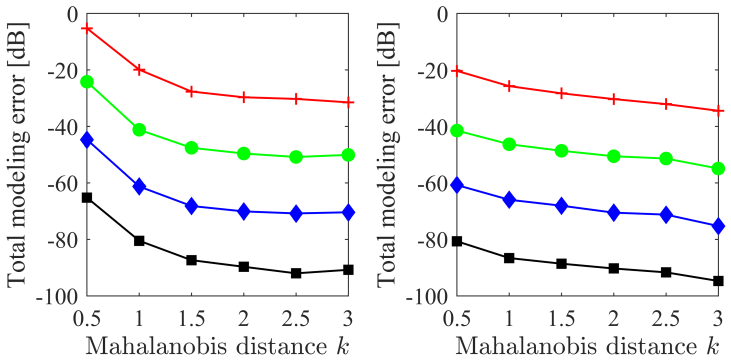


Figure 3.28: Number of input samples for different ROI widths, $\sigma_x = \sigma_y = 1, dx = dy = 0.25$

With the increase of ROI width from the minimal for $k = 0.5$ to the maximum for $k = 3$, the number of input samples raises 36 times since $(k_1^2/k_2^2 = 3^2/0.5^2)$; thus the expected modeling error should reduce by $10\log_{10}(36) = 15.5$dB.

Fig. 3.29a shows the total modeling errors of the LS method without compensation of differential entropy reduction. Total modeling error decreases by almost 30 dB, twice as much as expected. On wider ROIs, the reduction of total modeling error slows down, and the further increase of input samples does not significantly increase the estimation accuracy. However, after the compensation of differential entropy reduction, which is shown in Fig. 3.29b, the total mod-



(a) Without compensation of differential entropy reduction

(b) With compensation of differential entropy reduction

Figure 3.29: Mean of total modeling errors in dB for 50 trials

eling error linearly decreases with the slope $-10\log_{10}(N_1/N_2)$, as expected. The conclusion is that the derived analytical expression for differential entropy reduction completely compensates for the accuracy loss of the LS method caused by the reduction of information content within the uniformly sampled data with constrained spatial density.

Chapter 4

Multidimensional Gaussian profile fitting

This chapter describes the method proposed in [77] for multidimensional Gaussian profile fitting from noisy data in the exponential function's argument domain. As aforementioned, the Gaussian profiles of 2D and higher dimensions have applications in many engineering fields such as astronomical imaging, spectroscopy, and medical imaging [37, 38, 42, 50, 54, 55, 56, 64].

The generalized multidimensional Gaussian profile has the following form

$$f(\mathbf{x},\mu,\Sigma,A) = \frac{A}{\sqrt{(2\pi)^n|\Sigma|}} \exp\Big(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\Big), \tag{4.1}$$

where n denotes the dimension, $\mathbf{x}^{(i)} = [x_1^{(i)},..,x_n^{(i)}]^T$, $\forall i \in (1,m)$ corresponds to the position vector of ith of m input samples within the selected estimation input domain (ROI), A corresponds to the linear scale factor that multiplies the n-dimensional Gaussian probability density function (PDF), $\mu \in \mathbb{R}^n$ is the profile's centroid position, and $\Sigma \in \mathbf{S}_{++}^n$ is the symmetrical, positive definite, invertible and of full rank $n \times n$ covariance matrix. The matrix of vectors of samples' positions within the selected ROI is $\mathbf{X} = (\mathbf{x^1},..,\mathbf{x^m})$, while the vector of corresponding empirical profile values is $\mathbf{z} = [z^{(1)},..,z^{(m)}]^T$. Estimation ROI does not need to be uniformly sampled rectangular grid in n-dimensional space. However, samples within ROI must belong to a convex region symmetrically distributed around the expected profile's peak position.

Several papers have already dealt with the estimation of the 1D Gaussian profile parameters in the log domain from the noisy measurements using the conventional least-squares (LS) method, where the estimation problem comes down to parabola fitting [60, 61, 62]. However, for the multidimensional case, the number of profile parameters quadratically increases, as well as the estimation complexity. The analytical LS solution for all model parameters exists only for the 1D case, while the optimal parameters are non-linearly coupled for higher dimensions. The inverse covariance matrix in the quadratic form and the determinant of the covariance matrix in the normalization term in front of the exponential additionally complicates the estimation of the proper linear profile's scale for higher profile dimensions.

In [77], we propose transforming the problem of multidimensional Gaussian profile fitting

in the exponential argument domain by taking the natural logarithm of both model and input measurements. The solution for optimal profile parameters is searched for by minimizing the weighted sum of the squared residuals in the log domain. The proposed two-step WLS method in the log domain combines the analytical and iterative approaches. In the first step of the proposed method, the parameters of the inverse covariance matrix and one additional unknown, the residual vertical shift of the log target, are estimated by solving a system of linear equations with the one-step analytical solution given an initial centroid. In the second step of the proposed method, the centroid is updated by solving the system of nonlinear equations through the iterative procedure for a given estimated parameters from the first step. Such a two-step procedure is iteratively repeated until the required accuracy or convergence is not achieved. Since most of the profile's parameters are analytically estimated in the first method's step, the proposed method provides a significant reduction of computational costs and the estimation speed up compared to the conventionally used iterative LS method in the domain of values. Additionally, there is also proposed the analytical solution for the centroid update by using the theory of resultants [78, 79], but only for the 2D case. Also, the sensitivity of the proposed iterative two-step method to different noise sources is analyzed, and the method's accuracy and efficiency are compared with the commonly used LS method in the domain of values from [74].

The estimation weights are introduced to ensure that the minimization of the sum of weighted squared errors in the log domain simultaneously minimizes the sum of squared residuals in the domain of values. Two types of estimation weights are considered: model-driven or input data-driven estimation weights. In addition, there is also analyzed the influence of negative input samples and the difference in the informativeness of input data for different ROI widths on the accuracy of the model estimation in the log domain as a function of the input noise level.

4.1 Iterative two-step method for multidimensional Gaussian profile fitting in the exponential argument domain

The illustration of the iterative method for multidimensional Gaussian profile fitting in the domain of exponential argument is given in Fig.4.1. The main steps of the method include

1. Initialization which corresponds to the calculation of ROI width, normalization of input measurements $\mathbf{z}$ with initial scale factor $\hat{A}_{init}$, and calculation of initial centroid μ_{init}
2. Transformation of the estimation problem in the domain of exponential argument by taking the natural logarithm of both the model of multidimensional Gaussian profile $\ln\left(f(\mathbf{x},\mu,\Sigma,1)\right)$ and the normalized input data values $\ln(\mathbf{z}_n)$
3. Determination of normalized estimation weights $\mathbf{w}_n$ as model- or data-driven weights
4. Minimizing the objective function, which is defined as the weighted sum of squared residuals in the log domain, concerning the inverse covariance matrix terms $\hat{\Sigma}^{-1}$ and additional

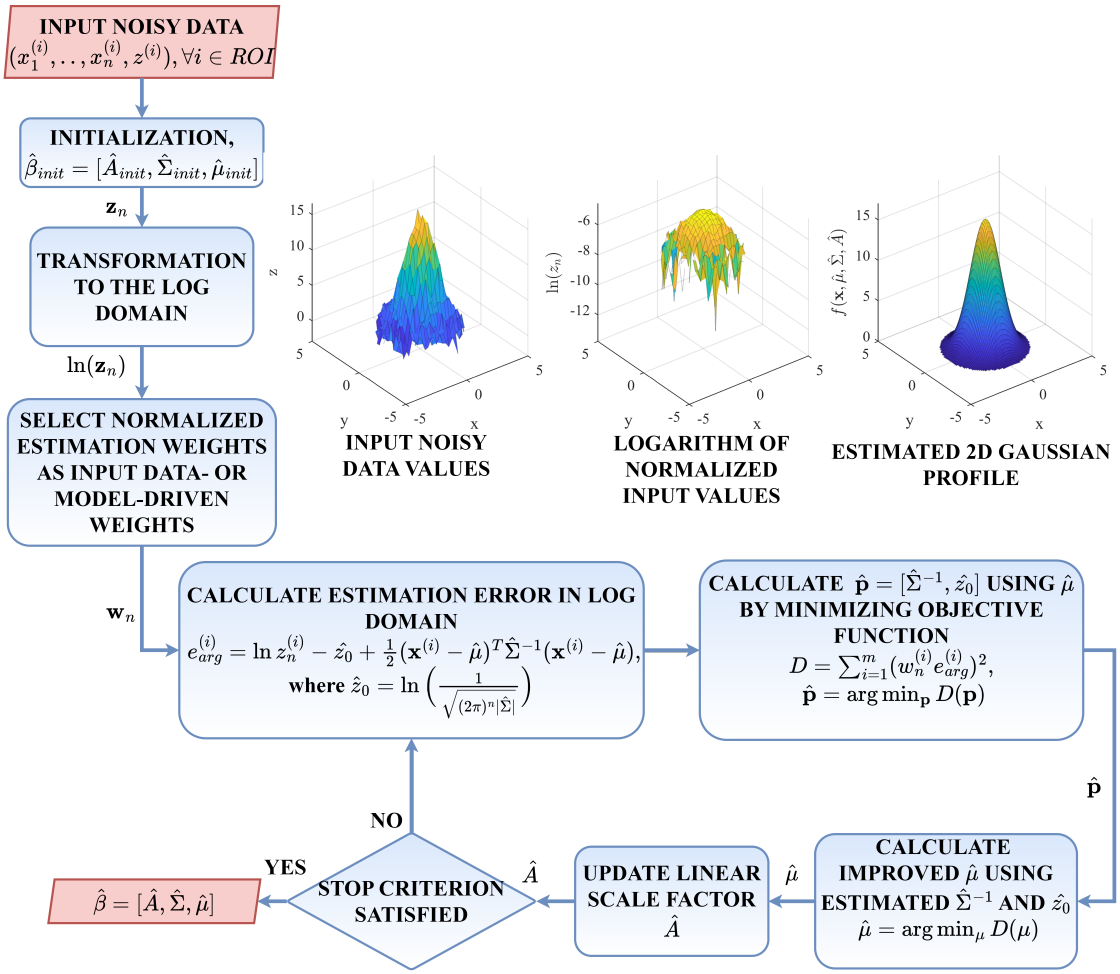


Figure 4.1: Illustration of the proposed method

offset of the log target $\hat{z}_0$ for a given centroid $\hat{\mu}$ by solving the system of linear equations

5. Update centroid position $\hat{\mu}$ by solving the system of nonlinear equations given the estimated parameters $\hat{\Sigma}$ and $\hat{z}_0$ from the previous step

6. Estimate linear scale factor $\hat{A}$ directly from estimated parameters in 4 or by fitting the obtained model to input data as a linear least squares solution as follows

$$\hat{A} = \frac{f(\mathbf{X}, \hat{\mu}, \hat{\Sigma}, 1)^T \mathbf{z}}{f(\mathbf{X}, \hat{\mu}, \hat{\Sigma}, 1)^T f(\mathbf{X}, \hat{\mu}, \hat{\Sigma}, 1)} \tag{4.2}$$

7. Iteratively repeat steps 4-6 until the given termination criterion is not achieved (convergence, maximum number of iterations, etc.)

The proposed method iteratively corrects covariance matrix Σ and centroid position μ to obtain a better Gaussian profile's fit in each iteration to empirical noisy input data. In the absence of noise, independently of selected estimation weights, $m_{min} = \frac{n(n+1)}{2} + 1$ input samples are sufficient to find a unique solution for the terms of the inverse covariance matrix with $n(n+1)/2$ unknowns and residual vertical shift of the log target as one additional unknown. However, in the case of a minimal number of input samples, their positions must not be collinear concerning

the centroid position since the system of linear equations would be of non-full rank and without a unique solution. In estimation from noisy data, increasing the number of input samples is recommended to reduce the noise variance and, consequently, the variance of estimated parameters. Also, due to the nonlinearity of the exponential function, the estimation weights have to be introduced to achieve that minimizing the sum of weighted squared residuals in the log domain minimizes the sum of squared residuals in the domain of values as well.

The main advantage of estimating the Gaussian profile parameters in the log domain is an analytical one-step solution for all terms of the covariance matrix $\hat{\Sigma}$ and residual vertical shift of the log target $\hat{z}_0$ for a given initial centroid. Input data entering the estimation process are normalized using an initial linear scale factor $\hat{A}_{init}$, which is not optimally initialized. Also, there is an unknown determinant of the covariance matrix in the normalization term in front of the exponential. Estimation error in one of these two quantities causes the residual vertical shift of the log target, which must be identified and compensated, thus enabling the proper transformation of the estimation problem in the log domain.

In the second stage of the method, the centroid position $\hat{\mu}$ is updated by minimizing the same objective function by solving the nonlinear system of equations. The nonlinear system of equations consists of n complete polynomials of third order in variables $[\mu_1..\mu_n]$. Such a nonlinear system requires some iterative procedure for nonlinear optimization. Analytical gradients and the Hessian matrix of the objective function are provided for the optimization procedure to speed up the estimation. The whole estimation process is iteratively repeated (alternately $\hat{\Sigma}$, then centroid $\hat{\mu}$) until convergence.

4.1.1 Selection of estimation weights

In this subsection, the change in the exponential function argument is related to the change in its values to derive the optimal estimation weights that ensure that the minimization of the sum of weighted squared residuals in the log domain simultaneously minimizes the sum of squared residuals in the domain of exponential function values. Let us consider the simplified example of the 1D Gaussian profile, defined as

$$y = K\exp(x), \tag{4.3}$$

where x corresponds to the argument, while K corresponds to the amplitude of the exponential function. Let us add the perturbation dx to the ideal argument x_0 and see how the argument error affects the ideal exponential function value $y_0 = K\exp(x_0)$ as

$$y = y_0 + dy = K\exp(x_0 + dx) = y_0\exp(dx). \tag{4.4}$$

From the above equation follows that the change of the ideal exponential function value dy equals

$$dy = y_0(\exp(dx) - 1). \tag{4.5}$$

Exponential function can be approximated by the first term of the Taylor series for small values of dx as

$$\exp(dx) \simeq 1 + dx. \tag{4.6}$$

Finally, the change of exponential function value with the small perturbations of exponential argument dx equals

$$dy = y_0(1 + dx - 1) = y_0 dx \tag{4.7}$$

It indicates that the ideal estimation weight w equals the noiseless sample value y_0. Since the ideal sample values are usually unknown, the alternatives use empirical input data values or model values as estimation weights. In [77], to regularize the dynamic range of estimated profile parameters, we used normalized input data and, consequently, the normalized estimation weights, which are

$$\mathbf{w}_n = \mathbf{z}_n, \tag{4.8}$$

where $\mathbf{z}_n$ is the normalized input data column vector. In the case of model-driven weights, a normalized initial moment-based model of the form $f(\mathbf{X}, \hat{\mu}_{init}, \hat{\Sigma}_{init}, 1)$ is used. Such chosen weights are calculated only once, and the same weights are used in all iterations of the proposed method. The method can be further improved by updating the estimation weights in each iteration according to the estimated model values from the previous iteration. Such approximation of estimation weights is appropriate only for small argument errors, which also yield small errors in the domain of values.

4.1.2 Objective function

The vector of the multidimensional Gaussian profile' parameters is

$$\beta = (A, \Sigma_{11}, \Sigma_{12}, ..., \Sigma_{nn}, \mu_{x_1}, ..., \mu_{x_n}), \tag{4.9}$$

where n denotes the dimension, A denotes the linear scale factor, $(\Sigma_{11}, \Sigma_{12}, ..., \Sigma_{nn})$ corresponds to the vector of unique terms of the covariance matrix, and $(\mu_{x_1}, ..., \mu_{x_n})$ denotes the vector of centroid position. The unique covariance matrix terms, $(\Sigma_{11}, \Sigma_{12}, ..., \Sigma_{nn})$, are unequivocally defined by their unique inverse terms, $(\Sigma_{11}^{-1}, \Sigma_{12}^{-1}, ..., \Sigma_{nn}^{-1})$ that appear in the quadratic form of the exponential argument. The residual error in the log domain of the ith of m samples within the selected ROI with the corresponding position vector $\mathbf{x}^{(i)}$ and normalized value $z_n^{(i)}$ equals

$$
\begin{aligned}
e_{arg}^{(i)} &= \ln z_n^{(i)} - \ln\Big(f(\mathbf{x}^{(i)},\mu,\Sigma,A)/A\Big) \\
&= \ln z_n^{(i)} - \ln\left(\frac{1}{\sqrt{(2\pi)^n|\Sigma|}}\right) + \frac{1}{2}(\mathbf{x}^{(i)}-\mu)^T\Sigma^{-1}(\mathbf{x}^{(i)}-\mu),\ \forall i \in [1,m].
\end{aligned}
\tag{4.10}
$$

By introducing the auxiliary variable of residual vertical offset of the log target z_0, which is defined as

$$
z_0 = \ln\left(\frac{1}{\sqrt{(2\pi)^n|\Sigma|}}\right), \tag{4.11}
$$

the residual error in the argument domain of the *i*th sample becomes

$$
e_{arg}^{(i)} = \ln z_n^{(i)} - z_0 + \frac{1}{2}(\mathbf{x}^{(i)}-\mu)^T\Sigma^{-1}(\mathbf{x}^{(i)}-\mu). \tag{4.12}
$$

Finally, the objective function, which is defined as a sum of weighted squared residuals in the log domain, has the following form

$$
D = \sum_{i=1}^{m}(w_n^{(i)}e_{arg}^{(i)})^2, \tag{4.13}
$$

where $\mathbf{e}_{arg} = (e_{arg}^{(1)},..,e_{arg}^{(m)})^T$ corresponds to the column vector of errors for all samples within the estimation ROI.

4.1.3 Estimation of the inverse covariance matrix

In the first stage of the proposed method, the inverse covariance matrix Σ^{-1} and the auxiliary variable of the residual vertical shift of the log target z_0 are estimated. Hence, the vector of unknowns $\mathbf{p}$ in the first stage has the following form $\mathbf{p} = (\Sigma_{11}^{-1},\Sigma_{12}^{-1},..,\Sigma_{nn}^{-1},z_0)^T$. The parameters are estimated by minimizing the objective function (4.13) as

$$
\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} D(\mathbf{p}), \tag{4.14}
$$

where

$$
e_{arg}^{(i)} = \ln z_n^{(i)} - z_0 + \frac{1}{2}(\mathbf{x}^{(i)}-\hat{\mu})^T\Sigma^{-1}(\mathbf{x}^{(i)}-\hat{\mu}), \tag{4.15}
$$

where $\hat{\mu}$ corresponds to the initial centroid for the first iteration or estimated centroid position from the previous iteration. The partial derivatives of the objective function (4.13) considering vector $\mathbf{p}$ given an initial centroid $\hat{\mu}$ yield a system of linear equations with an analytical one-step solution. The first derivatives of the objective function considering the vector of unknowns $\mathbf{p}$ in

the matrix form are

$$\nabla D(\mathbf{p}) = \mathbf{A}\mathbf{p} - \mathbf{b}, \tag{4.16}$$

where $\mathbf{A}$ corresponds to the matrix of coefficients and $\mathbf{b}$ denotes the vector of constant terms. The solution vector of this system of linear equations is

$$\hat{\mathbf{p}} = \mathbf{A}^{-1}\mathbf{b}. \tag{4.17}$$

The matrix of coefficients $\mathbf{A}$ is formed as

$$\mathbf{A} = \sum_{i=1}^{m} 2w_n^{(i)} \mathbf{d}^{(i)} \mathbf{d}^{\mathbf{T}(i)}, \tag{4.18}$$

while vector of constant terms $\mathbf{b}$ is

$$\mathbf{b} = \sum_{i=1}^{m} -2w_n^{(i)} \ln(z_n^{(i)}) \mathbf{d}^{(i)}. \tag{4.19}$$

The auxiliary vector $\mathbf{d}^{(i)}$ is the vector of the form

$$\mathbf{d}^{(i)} = \Big(dx_{1,1}^{(i)}/2, dx_{1,2}^{(i)}, dx_{1,3}^{(i)}, \cdots, dx_{1,n}^{(i)}, dx_{2,2}^{(i)}/2, dx_{2,3}^{(i)}, \cdots, dx_{2,n}^{(i)}, \cdots, \tag{4.20}$$

$$dx_{n-1,n-1}^{(i)}/2, dx_{n-1,n}^{(i)}, dx_{n,n}^{(i)}/2, -1 \Big)^T, \tag{4.21}$$

where $dx_j^{(i)} = x_j^{(i)} - \mu_{x_j}$, $dx_{r,s}^{(i)} = dx_r^{(i)} \cdot dx_s^{(i)}$, and the index pairs (r, s), with $1 \leq r \leq n$ and $r \leq s \leq n$, denote the unique row and column indices of the inverse covariance matrix elements, taking into account the property of symmetry.

The Hessian matrix of the objective function (4.13) for the vector of unknowns $\mathbf{p}$ is equal to the matrix of coefficients $\mathbf{A}$. According to (4.18), the matrix of coefficients $\mathbf{A}$ is calculated from the positions of input samples and corresponding estimation weights. The transformation of the estimation process to the log domain requires that noisy input samples have strictly positive values, thereby discarding the rest of the input samples from the estimation ROI. Consequently, the estimation weights, which are selected as normalized input sample values according to (4.8), are also positive. The matrix of coefficients $\mathbf{A}$ is calculated as a weighted sum of sample autocorrelation matrices that are positive semidefinite by definition. Consequently, the matrix $\mathbf{A}$, which is also a Hessian matrix, is positive semidefinite too.

Hence, the solution of the linear system $\hat{\mathbf{p}}$ is the unique and optimal for the global minimum of the objective function (4.13) for the given centroid $\hat{\mu}$ and for the chosen and fixed estimation weights, provided that $\mathbf{A}$ is nonsingular, i.e., that the Hessian is strictly positive definite. The matrix $\mathbf{A}$ can become singular only for an insufficient number of input samples or their collinearity relative to the centroid position, resulting in a non-full rank system. Namely, in

these situations, the determinant of this linear system of equations becomes zero, thus yielding infinitely many solutions. It occurs for the minimal number of input samples (m_{min}) with equal absolute distances from the profile's peak position concerning any one of the n axes ($\text{rank}(\mathbf{A}) < n$ if $|dx_j^{(1)}| = |dx_j^{(2)}| = ... = |dx_j^{(m_{min})}| = \text{const}$, $1 \leq j \leq n$). In that situation, the positions of all of the m_{min} input samples are on a flat hyperplane perpendicular to one of the n axes ($\pm$const from the peak position). Also, in the case of a regular grid, the geometric center of those m_{min} input samples must not coincide with the profile's peak position since, in that case, all input samples are equidistant from the peak position for all axes. In all other cases, the matrix $\mathbf{A}$ and the Hessian matrix are positive definite, ensuring that the obtained solution for the inverse covariance matrix and the residual vertical shift of the log target is unique and represents the global minimum of the objective function for the given centroid position and the chosen fixed estimation weights.

Finally, the optimal solution for the covariance matrix $\hat{\Sigma}$ is found by inverting $\hat{\Sigma}^{-1}$ whose unique elements are contained in $\hat{\mathbf{p}}$ by imposing the symmetry of the inverse. Thus, the solution $\hat{\Sigma}$ will also be a symmetric matrix, but to be a valid covariance matrix, it also has to be positive semidefinite. Theoretically, the estimated covariance matrix $\hat{\Sigma}$ can have some of the axes (eigenvalues) of zero length, thus indicating that the given Gaussian is actually of a lower dimension than the dimension of the input vectors, n. The condition for positive semidefiniteness of $\hat{\Sigma}$ also requires positive definiteness of $\hat{\Sigma}^{-1}$, because if any eigenvalue of $\hat{\Sigma}^{-1}$ is negative, then its reciprocal value, which is the eigenvalue of $\hat{\Sigma}$, will also be negative. Also, none of the eigenvalues of $\hat{\Sigma}^{-1}$ must be zero since, in that case, $\hat{\Sigma}^{-1}$ would be singular, and would not have an inverse. Since the solution for $\hat{\mathbf{p}}$ in (4.17) depends on the input data (sample positions, their values, and weights), the estimated matrices $\hat{\Sigma}^{-1}$ and $\hat{\Sigma}$ might become invalid for estimation from noisy data. Experiments have shown that such exceptions occur only for very low SNRs and a small number of input samples k, comparable with the number of unknowns.

4.1.4 Estimation of improved centroid

The second step is the calculation of the improved centroid $\hat{\mu}$ using the estimated inverse covariance matrix $\hat{\Sigma}^{-1}$ which will further reduce the summary weighted squared error of the argument. The centroid was calculated by minimizing the same objective function (4.13) as

$$\hat{\mu} = \arg\min_{\mu} D(\mu), \tag{4.22}$$

where

$$e_{arg}^{(i)} = \ln z_n^{(i)} - \hat{z_0} + \frac{1}{2}(\mathbf{x}^{(i)} - \mu)^T \hat{\Sigma}^{-1} (\mathbf{x}^{(i)} - \mu). \tag{4.23}$$

If $\mathbf{e}_{arg}$ is the column vector of errors of the exponential function's argument and $\mathbf{w}_n$ is the

column vector of normalized estimation weights with m elements that correspond to the number of input samples, then the expressions for gradients and Hessian matrix of the objective function concerning the centroid position have the following matrix forms,

$$\nabla D(\mu) = -2{\Sigma^{-1}}^T \mathbf{L}^T \mathbf{W}_n \mathbf{e}_{arg}, \tag{4.24}$$

$$\mathbf{H}(D(\mu)) = 2({\Sigma^{-1}}^T \mathbf{L}^T \mathbf{W}_n \mathbf{L} \Sigma^{-1} + \mathbf{w}_n^T \mathbf{e}_{arg} \Sigma^{-1}), \tag{4.25}$$

where $\mathbf{W}_n = diag(\mathbf{w}_n)$, and $\mathbf{dx}_j = (dx_j^{(1)}, .., dx_j^{(m)})^T$ denotes the column vector of distances of all m input samples from the given initial centroid for the jth dimension, and $\mathbf{L} = (\mathbf{dx}_1, .., \mathbf{dx}_n)$ is the matrix of such column vectors for all of n dimensions.

4.1.5 Update linear scale

The previously estimated linear scale $\hat{A}$ is updated either from the solution of the system of linear equations (4.17) by using the estimated $\hat{\Sigma}$ and $\hat{z}_0$ or by fitting the obtained normalized model to input data values according to (4.2) by using the estimated $\hat{\Sigma}$ and $\hat{\mu}$.

In the first case, the previously estimated linear scale is multiplied by the residual scale calculated from (4.11) to obtain the updated scale of the form

$$\hat{A} = \hat{A} \exp(\hat{z_0}) \sqrt{(2\pi)^n |\hat{\Sigma}|}. \tag{4.26}$$

However, it is easier and more accurate to update a linear scale in the value domain according to (4.2), but adding the residual vertical shift of the log target z_0 to the set of unknowns is necessary to decouple the scale and shape of the Gaussian profile in the first stage of the proposed method.

4.1.6 Initialization

Application of the proposed method for multidimensional Gaussian fitting in the log domain requires an appropriate initialization, including ROI selection, estimation of initial scale factor $\hat{A}_{init}$, calculation of normalized input data values $\mathbf{z}_n$ and estimation of the initial centroid position $\hat{\mu}_{init}$.

Estimation ROI selection The estimation ROI used in [77] is defined as a convex region enclosed within an arbitrarily rotated hyper-ellipsoid in n-dimensional space centered at the profile's peak position whose width is described by the chosen Mahalanobis distance k. The Mahalanobis distance factor k is a simple scalar measure of ROI width and denotes the samples'

positions of the same profile's values. The Mahalanobis distance is defined as the square root of the quadratic form in the argument as

$$k = \sqrt{(\mathbf{X} - \mu)^T \Sigma^{-1} (\mathbf{X} - \mu)}, \tag{4.27}$$

where $\mathbf{X}$ is the matrix of all input samples positions in n-dimensional space. The inverse covariance matrix and centroid position must be known in advance to calculate the Mahalanobis distance of each input sample using (4.27).

However, since the values of the samples are known, and these are assumed to belong to the underlying Gaussian profile, simple thresholding of sample values can be used to extract those expected to be within the selected ROI width, determined by the k. For example, if the maximum value of input samples is $\max(\mathbf{z})$, then the chosen marginal threshold $\bar{\mathbf{z}}_{edge}$ will extract only those samples whose Mahalanobis distance to the profile center is k at most, where k equals

$$k = \sqrt{-2 \ln (\bar{\mathbf{z}}_{edge} / \max(\mathbf{z}))}, \tag{4.28}$$

Such selection is invariant concerning the unknown linear scale A since it is canceled in the quotient in (4.28). Moreover, since input sample values may be contaminated with noise, the binary domain matrix obtained through thresholding can be further refined using morphological smoothing to ensure it is convex and homogeneous in all dimensions.

Calculation of initial scale factor A_{init} After extracting only the samples within the chosen bounded hyper-ellipsoidal ROI, these samples approximate the truncated multidimensional Gaussian profile at Mahalanobis distance of k. The total probability of all samples within such a truncated region is known to be $\mathrm{erf}(k/\sqrt{2})$, which can be used to normalize the input data (i.e., to remove the unknown scale A_{init}). In the case of uniformly sampled ROI, the initial scale factor A_{init} can be calculated as a quotient of the total probability under the n-dimensional histogram of input data and the total probability of the truncated Gaussian profile as

$$A_{init} = \frac{\sum \mathbf{z} \cdot \delta x_1 \cdot \ldots \cdot \delta x_n}{\mathrm{erf}(\frac{k}{\sqrt{2}})}, \tag{4.29}$$

where δx_j denotes the distance between neighboring samples for the jth axis in n-dimensional space in the case of regularly and uniformly sampled ROI.

Normalization of input data values The input data values $\mathbf{z}$ are normalized according to

$$\mathbf{z}_n = \frac{\mathbf{z} \cdot \delta x_1 \cdot \ldots \cdot \delta x_n}{A_{init}} = \frac{\mathbf{z} \cdot \mathrm{erf}(\frac{k}{\sqrt{2}})}{\sum \mathbf{z}}, \tag{4.30}$$

where $\mathbf{z}_n$ represents the normalized n-dimensional histogram, approximating the regularly sampled Gaussian profile truncated at maximal Mahalanobis distance k.

Initial centroid estimation The simplest method for initial centroid estimation is the method of moments in the domain of values. The first moment yields the expected centroid, i.e., the profile's peak position, along the x_j-axis as

$$\mu_{x_j} = \frac{\sum_{i=1}^{m} x_j^{(i)} z_n^{(i)}}{\sum_{i=1}^{m} z_n^{(i)}}, \tag{4.31}$$

where the denominator represents the total probability within the enclosed input region of truncated Gaussian, m is the number of input samples within the ROI, and j is the notation of the axis in n-dimensional space ($j \in [1,n]$).

Initial covariance matrix estimation In the case of using the normalized moment-based model of the form $\mathbf{f_1} = f(\mathbf{X}, \hat{\mu}_1, \hat{\Sigma}_1, \hat{A})/\hat{A} = f(\mathbf{X}, \hat{\mu}_1, \hat{\Sigma}_1, 1)$ as estimation weights, the corresponding covariance matrix has to be calculated as well. The second moment yields the variances that form the covariance matrix as follows:

$$\Sigma_{rr} = \sigma_r^2 = \sum_{i=1}^{m} z_n^{(i)} (x_r^{(i)} - \mu_{x_r})^2, \tag{4.32}$$

$$\Sigma_{rs} = \sigma_r \sigma_s \rho_{rs} = \sum_{i=1}^{m} z_n^{(i)} (x_r^{(i)} - \mu_{x_r})(x_s^{(i)} - \mu_{x_s}), \tag{4.33}$$

where σ_r and σ_s denotes the profile's STDs for the r and s axes and $\rho_{rs} \in [-1,1]$ is the Pearson product-moment correlation coefficient of r and s. Since the Gaussian profile is truncated at the maximal Mahalanobis distance k, the calculated profile's STDs have to be additionally scaled using the following expression

$$\sigma_{trunc} = \frac{\sigma}{\sqrt{1 - \frac{k \exp(-k^2/2)}{\sqrt{\frac{\pi}{2}} \mathrm{erf}(\frac{k}{\sqrt{2}})}}} \tag{4.34}$$

Initial scale factor update Additionally, the initial linear scale factor can be updated by fitting the obtained model to empirical data as the ordinary LS estimate in a simple linear regression model as

$$\hat{A}_1 = \frac{\mathbf{f}_1^T \mathbf{z}}{\mathbf{f}_1^T \mathbf{f}_1}, \tag{4.35}$$

where the numerator in (4.35) represents the sample covariance between the initial model and input data values, while the denominator represents the sample variance of the estimated model.

4.1.7 Probability of negative noisy input samples

Transformation of the Gaussian estimation problem to the log domain requires positive input sample values. We considered estimating the Gaussian profile parameters from data contaminated with additive Gaussian noise of zero-mean and σ_n^2 variance.

Samples that belong to a Gaussian profile have strictly positive values. However, after adding Gaussian noise, they can become negative, thus preventing the transformation of the Gaussian profile estimation to the log domain. If we assume that the Gaussian profile value, a function of input sample positions which are random variables uniformly distributed within the chosen ROI, is also a random variable Y that follows log-uniform distribution and that additive noise is a random variable X that follows the normal distribution, let us calculate the probability that the sum of these two random variables ($Z = Y + X$), Gaussian profile value, and additive noise is less than zero, i.e., $P(Z < 0)$. For Z to be less than z, X must be less than $z - Y$. The cumulative distribution function (CDF) of Z can be calculated as

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_{XY}(x,y)dxdy. \tag{4.36}$$

If X and Y are independent, the CDF of Z becomes

$$F_Z(z) = \int_{-\infty}^{\infty} f_Y(y)\Big(\int_{-\infty}^{z-y} f_x(x)dx\Big)dy. \tag{4.37}$$

In our case, $z = 0$ and the PDF of variables X and Y are

$$f_Y(y) = \frac{2}{k^2 y}, A\exp(-k^2/2) \leq y \leq A, \tag{4.38}$$

$$f_X(x) = \frac{1}{\sigma_n\sqrt{2\pi}} \exp\frac{-1}{2}\frac{x^2}{\sigma_n^2}, \mu = 0 \,(\text{zero-mean noise}), \tag{4.39}$$

while the CDF of normal noise is

$$F_X(x) = \frac{1}{2}\left(1 + \text{erf}\Big(\frac{x}{\sigma_n\sqrt{2}}\Big)\right), \tag{4.40}$$

and the CDF of total signal Z is

$$\begin{aligned} F_Z(z) &= \int_{A\exp(-k^2/2)}^{A} f_y(y)\Big(\int_{-\infty}^{-y} f_X(x)dx\Big)dy = \int_{A\exp(-k^2/2)}^{A} f_y(y)F_x(-y)dy \\ &= \int_{A\exp(-k^2/2)}^{A} f_y(y)\frac{1}{2}\Big(1 + \text{erf}\Big(\frac{-y}{\sigma_n\sqrt{2}}\Big)\Big)\Big)dy = \frac{1}{2} - \frac{1}{(2k^2\sigma_n\sqrt{\pi})}\Bigg(2\sqrt{2}A\,{}_2F_2\Big(\Big[\frac{1}{2},\frac{1}{2}\Big],\Big[\frac{3}{2},\frac{3}{2}\Big],\frac{-A^2}{2\sigma_n^2}\Big) \\ &- 2\sqrt{2}A\exp\Big(\frac{-k^2}{2}\Big)\,{}_2F_2\Big(\Big[\frac{1}{2},\frac{1}{2}\Big],\Big[\frac{3}{2},\frac{3}{2}\Big],\frac{-A^2\exp(-k^2)}{2\sigma_n^2}\Big)\Bigg), \end{aligned} \tag{4.41}$$

where ${}_2F_2$ is a hypergeometric function. The derived expression for the CDF of total signal $F_Z(z)$ can be used for predicting the percentage of negative samples depending on the linear scale A, the standard deviation of noise σ_n, and the maximal Mahalanobis distance k.

4.1.8 Analytical 3D Gaussian profile fitting for a given centroid

To analyze the proposed method's efficiency and speed, the parameters of the 3D Gaussian profile were fitted to the synthesized noiseless data and data contaminated with additive Gaussian noise using the proposed method. The results were compared with the results of the commonly used method of moments [48], LS method [54, 80, 81], and ML method [59]. The examples of the 3D Gaussian fitting to the noisy data using the proposed method are shown in Fig. 4.2.

To emphasize the method's speed and to easily compare it with the mentioned methods, we fixed the centroid position and measured the accuracy and the execution time for estimation of all other profile parameters: the covariance matrix and the linear scale. The experiments were executed on a system with Intel(R) Core(TM) i5-7200U CPU @ 2.5 GHz and 8 GB of RAM, with MATLAB implementation of all methods. The solution of the method of moments was provided as an initial guess for the LS and ML methods which minimize the objective function using the iterative optimization procedure. The optimization was performed using the native MATLAB *fminunc* solver without specified analytical gradients. The maximum number of iterations was set to 50, while the optimality tolerance and the current point tolerance were set to 10^{-7}.

The total modeling error e_{total} from (3.35) was used to measure accuracy. The total modeling error represents the sum of the squared residuals between the estimated and given model over the same evaluation grid and indirectly aggregates the accuracy of all model parameters.

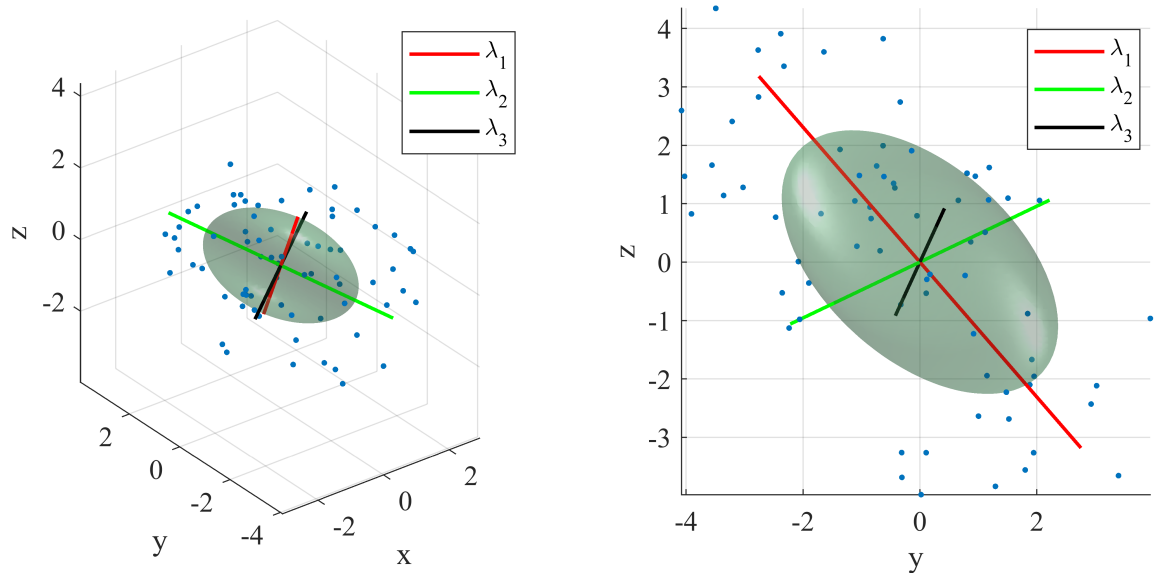
For experimental purposes, to obtain the synthetic noisy measurements, the profile with the given parameters was firstly synthesized, and then the additive noise with the appropriate STD σ_n was added to the profile to ensure the desired SNR within the ROI. In this case, the SNR was defined as PSNR according to the following formula

$$\mathrm{SNR}_{\mathrm{dB}} = 20\log_{10}\frac{A/\sqrt{(2\pi)^n|\Sigma|}}{\sigma_n}. \tag{4.42}$$

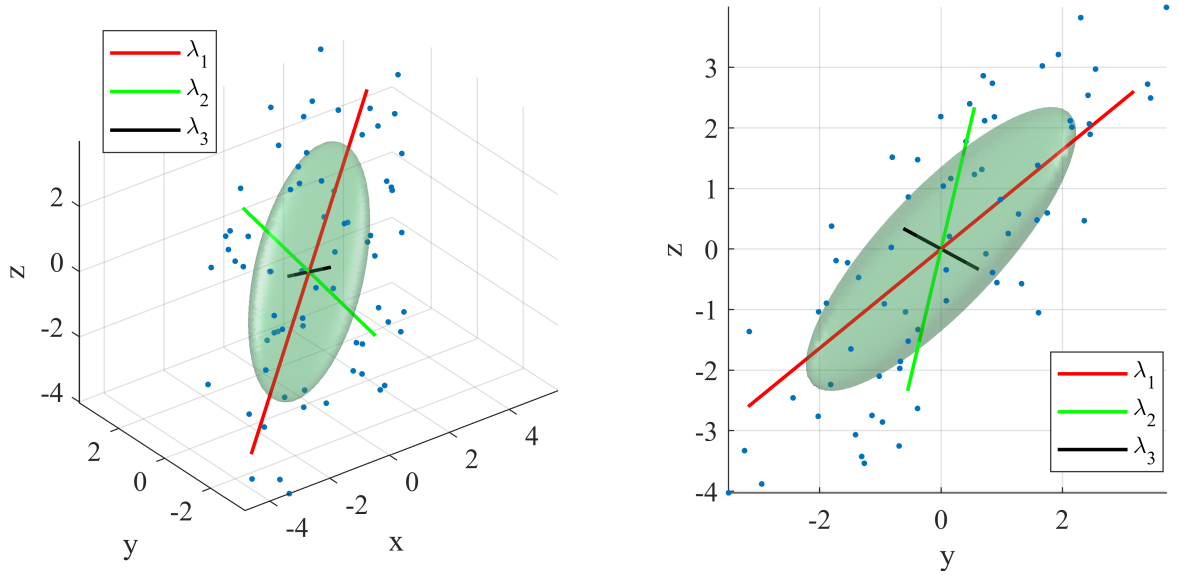
Fig. 4.2a illustrates the given 3D Gaussian profile estimation using the proposed method. Its actual parameters in the uncorrelated form are

$$[A,\lambda_1,\lambda_2,\lambda_3,\theta_1,\theta_2,\theta_3] = [1000,3,2,1,\pi/3,\pi/4,\pi/6],$$

where λ_1, λ_2 and λ_3 denote the semiaxes widths, θ_1, θ_2, θ_3 denote the Euler rotation angles, and A is the linear scale factor. The centroid position was fixed to $\mu = [0,0,0]$ and passed as input



(a) $A = 1000, \lambda_1 = 3, \lambda_2 = 2, \lambda_3 = 1, \theta_1 = \pi/3, \theta_2 = \pi/4, \theta_3 = \pi/6, \hat{\lambda}_1 = 3.01, \hat{\lambda}_2 = 2.03, \hat{\lambda}_3 = 1.02, \hat{\theta}_1 = \pi/3.00, \hat{\theta}_2 = \pi/4.09, \hat{\theta}_3 = \pi/6.21, \hat{A} = 1019.02$



(b) $A = 1000, \lambda_1 = 4, \lambda_2 = 2, \lambda_3 = 0.5, \theta_1 = \pi/5, \theta_2 = -\pi/7, \theta_3 = \pi/3, \hat{\lambda}_1 = 4.13, \hat{\lambda}_2 = 2.00, \hat{\lambda}_3 = 0.50, \hat{\theta}_1 = \pi/5.03, \hat{\theta}_2 = -\pi/7.09, \hat{\theta}_3 = \pi/3.00, \hat{A} = 1020.49$

Figure 4.2: Examples of 3D Gaussian fitting results in 3D and 2D views using the proposed method for a given centroid. Estimation was done from $m = 70$ input samples with the noise level of $\mathrm{SNR}_{\mathrm{dB}} = 30$. Estimated profile parameters are: the linear scale $\hat{A}$, the semiaxes widths $(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3)$, and the rotation angles $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$. The fitted 3-D surfaces represent the points where the estimated model equals the given model value at $k = 1$, which is $f = A\exp(-1/2)$. The blue dots represent only the input samples' positions but not their values (due to the limitation of the 3D example visualization).

to all estimation methods. The actual profile's parameters yield the following rotation matrix **R**

$$\mathbf{R} = \mathbf{R}_z(\theta_1)\mathbf{R}_y(\theta_2)\mathbf{R}_x(\theta_3) = \begin{bmatrix} 0.3536 & -0.5732 & 0.7392 \\ 0.6124 & 0.7392 & 0.2803 \\ -0.7071 & 0.3536 & 0.6124 \end{bmatrix}, \tag{4.43}$$

and the corresponding covariance matrix Σ

$$\Sigma = \mathbf{R} \begin{bmatrix} \lambda_1^2 & 0 & 0 \\ 0 & \lambda_2^2 & 0 \\ 0 & 0 & \lambda_3^2 \end{bmatrix} \mathbf{R}' = \begin{bmatrix} 2.9858 & 0.4609 & -2.6080 \\ 0.4609 & 5.6392 & -2.6801 \\ -2.6080 & -2.6801 & 5.3750 \end{bmatrix}. \tag{4.44}$$

Finally, the vector of the actual profile's parameters can be expressed as

$$\begin{aligned} \beta =& [A, \Sigma_{11}, \Sigma_{12}, \Sigma_{13}, \Sigma_{22}, \Sigma_{23}, \Sigma_{33}] \\ =& [1000, 2.9858, 0.4609, -2.6080, 5.6392, -2.6801, 5.3750]. \end{aligned}$$

The proposed method estimated the unique terms of the covariance matrix by solving the system of linear equations (4.17) and inverting the obtained solution, while the linear scale factor was calculated from the estimated additional unknown z_0 as $\hat{A} = \exp(\hat{z_0})\sqrt{(2\pi)^3|\hat{\Sigma}|}$, thus avoiding the need for initialization. In the noiseless case, the unit estimation weights were used, while for the estimation from the noisy data, the estimation weights were fixed to input sample values.

Table 4.1: The mean total modeling errors in dB for estimating the 3D Gaussian profile given the actual centroid position.

	Mean total modeling error [dB]				
	Ideal noiseless case			$\text{SNR}_{\text{dB}} = 40$	
Number of input samples	7	70	7000	70	7000
Proposed method	**-255.5804**	**-293.3372**	**-293.5277**	**-40.6434**	**-58.8981**
Method of moments	-7.1845	-13.9014	-15.3146	-13.7786	-15.3044
LS method	-67.0825	-134.5511	-135.3360	-41.4402	-61.9902
ML method	-89.9811	-134.5511	-135.3360	-41.4402	-61.9902

The parameters of the given profile were estimated for different numbers of input samples ($m = \{7, 70, 7000\}$) and different SNRs ($\text{SNR}_{\text{dB}} = \{20, 30, 40,$
$50, 60, 70, 80\}$). The Monte Carlo simulation with 1000 trials was performed for each combination of those parameters, including the noiseless case. In each trial, the input samples were

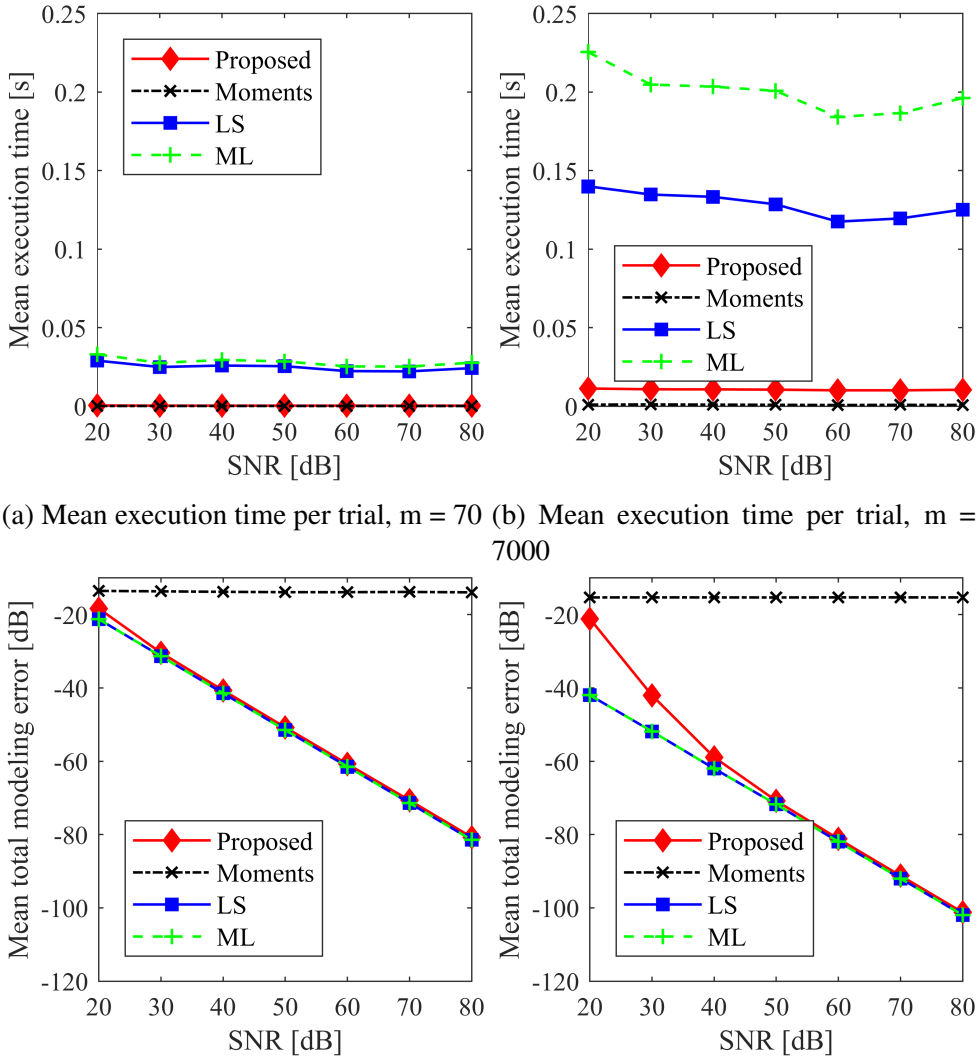
Table 4.2: The mean execution times in seconds for estimating the 3D Gaussian profile given the actual centroid position.

	Mean execution time per trial [s]				
	Ideal noiseless case			$\text{SNR}_{\text{dB}} = 40$	
Number of input samples	7	70	7000	70	7000
Proposed method	**0.0004**	**0.0004**	**0.0101**	**0.0004**	**0.0105**
Method of moments	0.0001	0.0001	0.0007	0.0001	0.0009
LS method	0.0475	0.0349	0.1239	0.0255	0.1286
ML method	0.0638	0.0391	0.1922	0.0285	0.2008

randomly picked within convex ROI determined with maximal Mahalanobis distance $k = 2$ from the actual centroid position. The evaluation region was selected as uniformly sampled ROI enclosed within Mahalanobis distance $k = 3$ with the step size of 0.1 in each direction and is defined by the given model's parameters. For the same 3D example, for $\text{SNR}_{\text{dB}} = 40$, the estimated parameters from 70 random input samples using the proposed method in one trial were $\hat{\beta} = [\hat{A}, \hat{\Sigma}_{11}, \hat{\Sigma}_{12}, \hat{\Sigma}_{13}, \hat{\Sigma}_{22}, \hat{\Sigma}_{23}, \hat{\Sigma}_{33}] = [1004.5, 3.0018, 0.4664, -2.6003, 5.6972, -2.6807, 5.3330]$ with the total modeling error $e_{total} = -42.8741\text{dB}$.

The solution of the method of moments was provided as an initial guess for LS and ML methods. Since the objective of the ML method assumes the presence of stochastic additive Gaussian noise, in the noiseless case, $\sigma_n = 10^{-4}$ was passed as an input to the optimization procedure to mimic the almost ideal measurements ($\text{SNR}_{\text{dB}} = 100$). In the case of noisy measurements, the considered numbers of input samples were only 70 and 7000 since, in that case, the minimum number of input samples which corresponds to the number of unknowns, was insufficient for a valid solution (for the 3D Gaussian with the given centroid position, the number of unknowns is 7). The mean execution times and total modeling errors for different numbers of input samples for the noiseless case and for $\text{SNR}_{\text{dB}} = 40$ are given in Tables 4.1 and 4.2. The trends of mean total modeling errors and execution times for different SNRs and numbers of input samples are shown in Fig. 4.3.

In the noiseless case, the proposed method finds the optimal (ideal) solution of all seven unknowns from only seven randomly positioned input samples with an average modeling error of -255dB for 1000 trials ($e_{total} = -255.58\text{dB}, m = 7$). For such a small number of input samples, the LS and ML estimates are inaccurate due to the wrong initial guess obtained by the method of moments. The LS and ML methods also converge to the optimal solution in the noiseless case for a more significant number of input samples ($m = 70$ and $m = 7000$). However, their speed and accuracy highly depend on the initial guess. The method of moments has poor accuracy since it underestimates all parameters ($e_{total} = -15.31\text{dB}, m = 7000$). In the



(a) Mean execution time per trial, m = 70 (b) Mean execution time per trial, m = 7000

(c) Mean total modeling error, m = 70 (d) Mean total modeling error, m = 7000

Figure 4.3: Comparison of the speed and the accuracy of the following methods: proposed method, method of moments, LS method, and ML method

estimation from noisy measurements, the LS and the ML methods behave similarly considering the accuracy since they yield the same optimal solution for the case of additive Gaussian noise contamination, as can be seen in Fig. 4.3c and Fig. 4.3d. The 100-fold increase in the number of input samples from 70 to 7000 caused the 20dB increase in the accuracy of the LS and ML methods due to averaging, as expected. The proposed method follows the trends of LS and ML methods concerning accuracy, except for very low SNRs (20dB) where the approximation of the exponential function with the first term of the Taylor series for estimation weights determination in the argument domain is insufficient. However, the proposed method requires at least ten times less time to calculate the covariance matrix and linear amplitude scale than LS and ML methods yielding comparable estimates without prior knowledge. Fig. 4.3a and Fig. 4.3b show that the mean execution time of the proposed method is comparable with the mean execution time of the method of moments since both methods do not use the iterative procedure and find

a solution analytically. Both methods are almost 100 times faster than the LS method for the smaller number of input samples (70) and up to 10 times for the larger number of input samples, according to Table 4.2. The LS and ML methods have a similar mean execution time for the small number of input samples, but with the increasing number of samples, the ML method becomes slower than the LS method, as shown in Fig. 4.3b.

To sum up, the proposed method is much faster than the commonly used LS and ML methods and has a mean execution time comparable with the fastest method of moments. Thereby, it yields the estimate of almost the same accuracy as the LS and ML methods except for very low SNRs.

4.1.9 Iterative method's accuracy

In the second experiment, the parameters of the rotationally symmetric 2D Gaussian profile were estimated from noisy data using the proposed method to verify its accuracy. In addition, the results are compared with those previously published in [76], where the same 2D Gaussian profile was estimated in the value domain using the numerical LS method. Although we fitted the 2D Gaussian profile, the proposed method can be applied to estimate the Gaussian profile of arbitrary dimensions and arbitrary shapes, as was demonstrated in the 3D experiment.

In [76] and section 3.6.1, it was shown that the accuracy of the LS method in the estimation of the 2D Gaussian parameters from data corrupted with additive noise varies for different ROI widths as the consequence of the difference in the informativeness of the input data, i.e., the difference in the differential entropy of the input data. The maximal differential entropy is obtained when the Gaussian profile is truncated at Mahalanobis distance $k = 2$. For other non-optimal ROI widths, the reduction of differential entropy compared to the optimal case for $k = 2$ was compensated by the proportional reduction of the given noise level. The difference in differential entropy of the input data precisely predicts a decrease in the LS method accuracy for different ROI widths in the value domain. We performed the same experiments in this section and re-applied the data entropy compensation.

In addition to differential entropy, we also analyzed the influence of negative samples on the method's accuracy since it is necessary to handle negative input sample values before the log transformation. Therefore, the probability of negative samples when a Gaussian profile is contaminated with additive noise was derived and given in subsection 4.1.7. The total number of input samples was increased in proportion to the predicted percentage of negative samples for the given ROI width and SNR to eliminate the loss of method accuracy due to the occurrence of negative samples that cannot participate in the estimation in the log domain. Negative samples were either set to $eps = 2^{-52}$ and logarithmized or simply removed from the input data and the estimation process. However, the compensation of negative samples did not significantly improve the accuracy of the proposed method for the considered ROI widths.

Furthermore, two types of estimation weights were considered and compared. In the first case, the weights in the objective function (4.13) were the values of the input samples, while, in the second case, the weights were selected as the initial moment-based model values for the same sample positions.

Two experiments were made considering circular ROI sampling: random sampling with a fixed number of input samples or uniform sampling with a fixed density of input samples. The total number of performed test cases in each experiment was 16, depending on the input parameters given in Table 4.3. In each test case, the Monte Carlo simulation with 50 trials was performed for all Mahalanobis distances k from the set $k = \{0.5, 1, 1.5, 2, 2.5, 3\}$ and for all given SNRs from the set $\mathrm{SNR}_{\mathrm{dB}} = \{20, 40, 60, 80\}$. As an evaluation grid, we used a uniformly sampled circular region of fixed width $r = 3\sigma$, where $\sigma = \sigma_1 = \sigma_2$ is the profile's STD with equal x and y spacing of $\sigma/10$.

In both experiments, the same six parameters of the specified 2D Gaussian profile were estimated from noisy measurements: profile's STDs σ_1, σ_2, the correlation coefficient ρ_{12}, the profile's peak position $\mu = [\mu_{x_1}, \mu_{x_2}]$, and the linear profile's scale A. The given profile parameters in the correlated form were $[A, \sigma_1, \sigma_2, \rho_{12}, \mu_{x_1}, \mu_{x_2}] = [100, 1, 1, 0, 0, 0]$, i.e., $\beta = [A, \Sigma_{11}, \Sigma_{12}, \Sigma_{22}, \mu_{x_1}, \mu_{x_2}] = [100, 1, 0, 1, 0, 0]$. In the case of compensation of differential entropy reduction compared to the nominal case for $k = 2$, the noise STD calculated from (3.43) was reduced according to the formula in [76], as follows

$$\sigma_{n_{reduced}} = \sigma_n \cdot 10^{(\Delta h_{z_{dB}}/20)}, \tag{4.45}$$

where $\Delta h_{z_{dB}}$ denotes the reduction of differential entropy in dB for the chosen factor k compared to the optimal case for $k = 2$.

Table 4.3: Input parameters combined in experiments

Weights $\mathbf{w_n}$	$\mathbf{z_n}$	$\mathbf{f_1}$
Number of input samples m (only for random sampling)	100	10000
Compensation of predicted number of negative samples	yes	no
Negative input samples	set to eps	removed
Compensation of differential entropy reduction compared to the case for $k = 2$	yes	no

The fixed number of input samples

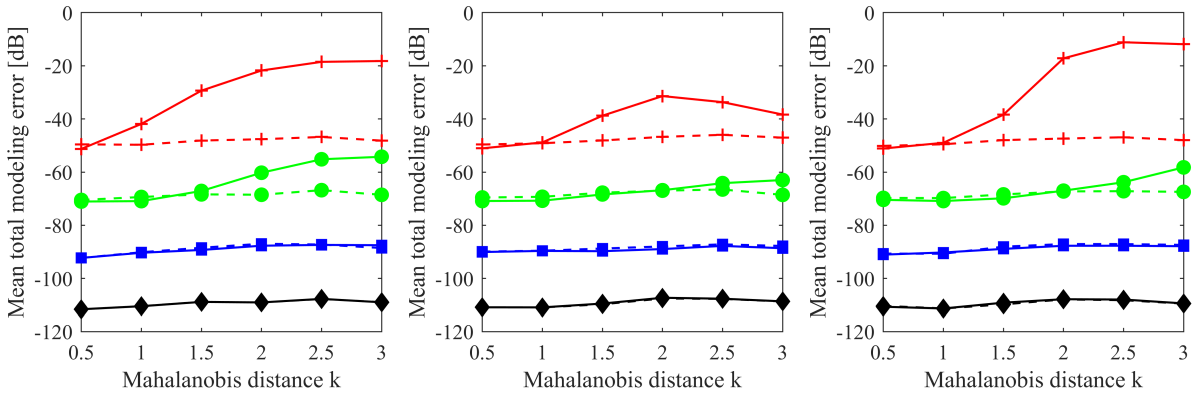
In the first experiment, the estimation input samples were randomly picked within circular ROIs, and their number was fixed for all ROI widths. The number of input samples was chosen as $m_1 =$

10000 or $m_2 = 100$. Thus, the total number of test cases in this experiment was 32, considering all combinations of input parameters in Table 4.3. The comparison of the mean total modeling errors averaged over 50 trials of the LS method in the domain of values, and the proposed method in the argument domain are shown in Fig. 4.4. As shown in [76], by compensating the differential entropy reduction, compared to the nominal optimal case for $k = 2$, the efficiency of the LS method in the domain of values became invariable of ROI widths for each given SNR (flat dashed lines). Additionally, as expected, the 100-fold increase in the number of random input samples increases the accuracy of the LS method by 20 dB $(10\log_{10}(m_1/m_2))$ due to averaging, which effectively reduces the noise level and, therefore, improves model accuracy.

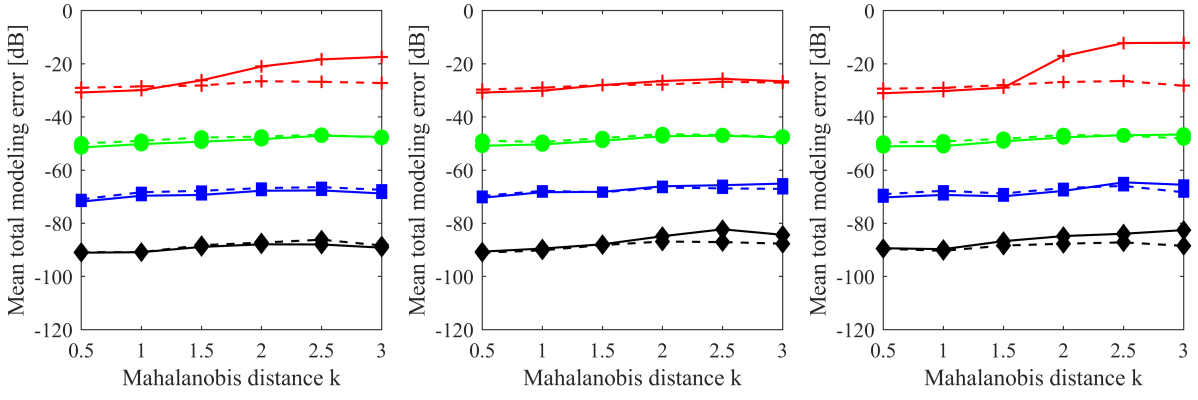
The accuracy of the proposed argument domain method depends on the selected combination of input parameters used in each test case. The results of a few representative test cases are shown in Fig. 4.4 where differential entropy reduction was compensated, and the compensation of the predicted number of negative samples was not applied.

As shown in Fig. 4.4, the accuracy of the proposed method is smaller than the LS method in the value domain only for very low SNR ratios and extensive ROI widths. For sufficiently high SNR ratios and narrower ROIs, the efficiency of the proposed method is practically identical to the efficiency of the LS method in the value domain. In Figs. 4.4a and 4.4d, the input sample values are used as estimation weights in the objective function (4.13), while in Figs. 4.4b- 4.4f the initial model values are used as estimation weights. When input sample values are utilized as estimation weights, both methods' mean total modeling errors are almost the same whether negative samples are removed before estimation or set to eps. However, for the case of using initial model values as weights, the results of the proposed method are closer to the results of the LS method in the value domain, but only if negative samples are removed from the estimation process (Figs. 4.4b and 4.4e). The most significant deviation in the accuracy of the proposed method compared to the LS method in the value domain is observed in the case of replacing negative samples with eps when the initial model values were used as weights, especially for low SNR and wide ROI $(\text{SNR} = 20\,\text{dB}, k = 3)$ as Figs. 4.4c and 4.4f show. Additionally, in Figs. 4.4a, 4.4b and 4.4c is shown that a 100-fold increase in the number of input samples does not increase the estimation accuracy by 20 dB on wide ROIs and low SNRs. The causes of such behavior are inappropriate handling of negative samples and improper weight selection since the assumption that the exponential function can be approximated with the first term of the Taylor series holds only for small errors.

In Fig. 4.4 can be seen that the relative degradation for low SNRs (20 dB) and wide ROIs is not the same for $m_1 = 10000$ as for $m_2 = 100$ but increases more for the larger m and less for the small m. However, it is simply a consequence of direct bias in the input data, which are falsified so that no averaging can help. The averaging can only attenuate a stochastic error and not a consistent one, so the error is almost the same in absolute terms for both cases ($m_1 = 10000$ and



(a) $m_1 = 10000$, $\mathbf{w} = \mathbf{z_n}$ (b) removed negative samples, $m_1 = 10000$, $\mathbf{w} = \mathbf{f_1}$ (c) negative samples set to *eps*, $m_1 = 10000$, $\mathbf{w} = \mathbf{f_1}$



(d) $m_2 = 100$, $\mathbf{w} = \mathbf{z_n}$ (e) removed negative samples, $m_2 = 100$, $\mathbf{w} = \mathbf{f_1}$ (f) negative samples set to *eps*, $m_2 = 100$, $\mathbf{w} = \mathbf{f_1}$

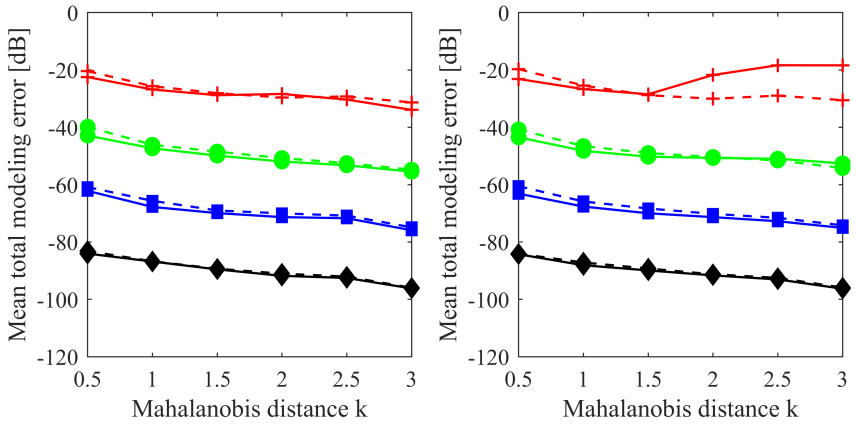
Figure 4.4: Comparison of the mean total modeling errors in dB of the LS method in the value domain (dashed lines) and the proposed method in the argument domain (solid lines) in the case of random sampling with fixed number of input samples for different factors k and different SNRs: 20 dB (red), 40 dB (green), 60 dB (blue), and 80 dB (black). $\mathbf{f_1}$ denotes a vector of initial model values, while $\mathbf{z_n}$ denotes input sample values.

$m_2 = 100$ when SNR $= 20$ dB and ROI $= 3$). It can be concluded that increasing the number of input samples reduces the model error only for very narrow ROIs when the percentage of such negative samples is small, so the contribution of stochastic noise is effectively attenuated. With very low SNRs, the initial moment-based model might also be inaccurate. Using such a model for weights and even throwing out negative samples results in a model denoted with the red line that is not parallel to green, blue, and black but instead increases incrementally (SNR $= 20$ dB).

The fixed density of input samples

The circular ROI was uniformly sampled in the second experiment, so the number of input samples was variable for different ROI widths depending on the ratio between the profile's STD and the pixel size. Since the STD of the given 2D Gaussian profile was $\sigma_1 = \sigma_2 = 1$, the selected pixel width was chosen as $\delta x_1 = \delta x_2 = 0.25$ to ensure the sufficient number of input samples for estimation of all given profile's parameters even for the narrowest ROI ($k = 0.5$), as it was

described in [76]. From the ratio of the ROI size and individual pixel size, it can be concluded that the number of input samples increases quadratically with the factor of Mahalanobis distance k ($m = P_{\mathrm{ROI}}/P_{\mathrm{pix}} = k^2\sigma_1\sigma_2\pi/(\delta x_1 \delta x_2)$ so $m_1/m_2 = k_1^2/k_2^2$) so it is expected that the total modeling error decreases with the increase of ROI width as $10\log_{10}(m_1/m_2) = 10\log_{10}(k_1^2/k_2^2)$. Since the minimal ROI width considered in this experiment is for $k = 0.5$ while the maximal for $k = 3$, the number of input samples increases 36-fold. Consequently, the expected reduction of the total modeling error due to the ROI size increase should be 15.56 dB. The mean total modeling errors are shown in Fig. 4.5. Again, only some of the 16 test cases are shown considering all input parameters in Table 4.3, while the number of input samples is directly determined by k. The results show that the proposed method's accuracy follows the LS method's accuracy trend in the value domain when using initial model values as weights and removing negative samples. The results are worse when input sample values are used as weights for low SNRs (20 dB) and wide ROIs. However, both methods achieve the same accuracy in cases of high SNRs.



(a) negative samples are removed, $\mathbf{w_n} = \mathbf{f_1}$, where $\mathbf{f_1}$ is a vector of initial model values

(b) negative samples are removed, $\mathbf{w_n} = \mathbf{z_n}$, where $\mathbf{z_n}$ is a vector of input sample values

Figure 4.5: Comparison of the mean total modeling errors in dB of the LS method in the value domain (dashed lines) and the proposed method in the argument domain (solid lines) in the case of uniform sampling and a variable number of input samples for different ROI widths determined by the factor k and different SNRs: 20 dB (red), 40 dB (green), 60 dB (blue), and 80 dB (black).

4.1.10 Discussion

This subsection analyses the influence of each input parameter given in Table 4.3 on the estimation accuracy of the proposed method. In the case of using input sample values as estimation weights, the weighted squared residual error is the same whether the negative samples are removed or set to *eps*. Since the squared residual error between the log of *eps* and the log of model value is further weighted with eps^2, it does not contribute to the total error sum, thus achieving the same effect as the negative sample removal. However, when the model values

are used as estimation weights in the log domain, setting negative samples to *eps* or removing them yields different estimation results. Namely, the log transform of such a small value (*eps*) yields a significant negative value. The initial moment-based model is sensitive to noise and imprecise at wide ROIs where negative samples typically occur and yields falsely and potentially more enormous model values at those positions than *eps*. When such a model value is used as a weight at a place of negative input sample set to *eps*, it can significantly degrade the estimation since such weight is false and arbitrary. This can be observed only for the lowest SNR ratio of 20 dB when such negative samples predominantly occur and, of course, only for the widest ROIs. For narrower ROIs, the probability of negative samples is lower, even for such low SNRs, so it is much better to remove negative samples from the estimation process and avoid the problems of weight selection. Therefore, using the initial model values as weights yields better results only if negative samples are removed and SNR is sufficiently high.

In addition, the experimental results show that the contribution of differential entropy compensation is not ideal. For example, in the case of a fixed number of input samples, an alignment of the error curves similar to the estimation in the value domain is achieved, but these curves are still not wholly flat across all ROI widths.

Noise level compensation performed by predicting the expected number of negative samples for a given width and SNR also does not significantly improve the estimation accuracy in the log domain since these negative samples occur mainly for very low SNRs and wide ROIs. In these cases, the problem is not a lack of valid positive samples but rather a poor approximation of quadratic error in the value domain by the weighted quadratic error in the argument domain using the weights that describe only the first term of Taylor's series of that approximation. Such simplified weights are insufficient for considerable errors when measurements significantly deviate from the given model.

4.1.11 Method's complexity

The complexity of the proposed two-step method in the argument domain is comparable to the complexity of only a single iteration of the LS method in the domain of values that simultaneously estimates all profile parameters using Newton's optimization technique that requires the calculation of analytical derivatives and Hessian matrix concerning all profile parameters.

The complexity of both methods is parameterized with the dimension of the Gaussian profile n and the number of estimation input samples m. The number of input samples m must be at least equal to the number of unknowns, which is of order n^2. When the number of input samples is of order n^2, the complexity of both methods in a big O notation is $O(n^6)$. For the number of input samples of order n^3 or higher, the complexity of both methods in a big O notation equals $O(mn^4)$.

Although the complexity of a single iteration is comparable, we have experimentally de-

termined that the proposed method converges in at most four such iterations since most of the optimal model parameters are found analytically, which explains the significant advantage of this method.

4.1.12 Iterative method's convergence

This subsection explains the proposed iterative method's convergence, where the centroid and the covariance matrix are updated alternately. If the given initial centroid is close enough to the actual solution and the estimation weights are fixed, the proposed method's first step finds a unique solution that minimizes the objective function by solving a system of linear equations. The proof of this statement with the description of exceptions is given in subsection 4.1.3.

In the second step of the method, we search for a better centroid that gives a smaller weighted squared error than the centroid from the previous iteration for the calculated covariance matrix by solving the system of coupled nonlinear equations. Due to the minimization criterion itself and the fixed estimation weights, an iterative optimization procedure such as Newton's method returns, in the worst case, the current centroid, and the iterative procedure terminates. In all other situations, the new centroid will yield a better fit (smaller value of the objective function), and a new estimation of the covariance matrix and scale can be performed for the new centroid position, further improving the model fit. The method can get stuck in a local minimum if the initial centroid is far from the real solution or if SNR is low. However, the method's first step with the analytical solution and the fixed estimation weights ensure convergence, at least according to the local minimum.

4.2 Analytical method for centroid estimation

There are many papers which deal with centroid estimation of Gaussian profile from noisy measurements [41, 48, 48, 49, 49, 52, 52, 55, 56, 59, 62, 82, 83, 84, 85, 86]. This section represents an extension of the proposed two-step method for multidimensional Gaussian profile fitting in the argument domain where the estimation of the centroid in the second stage of the method is also analytical, but only for the 2D case. The same objective function (4.13) is minimized to find the optimal solution for centroid where the *i*th sample's error is

$$e_{arg}^{(i)} = \ln z_n^{(i)} - \hat{z_0} + \frac{1}{2}(\mathbf{x}^{(i)} - \mu)^T \hat{\Sigma}^{-1}(\mathbf{x}^{(i)} - \mu), \tag{4.46}$$

$w_n^{(i)} = z_n^{(i)}, \forall i \in ROI$, and the total number of input samples is m.

The centroid estimation is reduced to fitting an elliptical paraboloid of the known shape determined by the covariance matrix $\hat{\Sigma}$ and the residual vertical shift of the log target $\hat{z}_0$ to the logarithm of normalized input noisy data by finding the solution for an optimal centroid

position $\hat{\mu} = [x_0, y_0]$ in the weighted least squares sense [87]. The logarithm of the normalized noisy input data is shown in Fig 4.6.

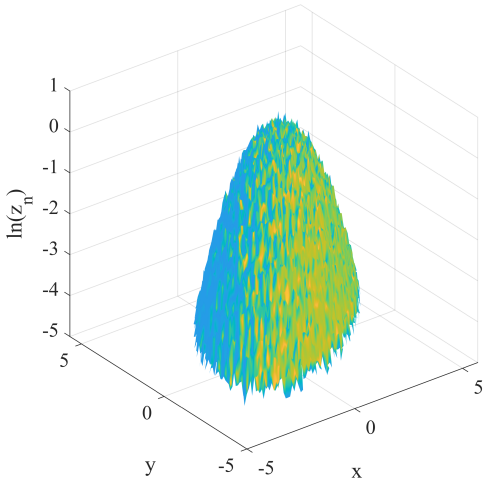


Figure 4.6: Logarithm of the noisy 2D Gaussian profile values

To minimize such an objective function, which is the 4th order polynomial in two variables of the form

$$D(x_0, y_0) = d_{40}x_0^4 y_0^0 + d_{31}x_0^3 y_0 + \ldots + d_{01}x_0^0 y_0^1 + d_{00}, \tag{4.47}$$

it is necessary to find the centroid position where the gradients of the objective function concerning both x_0 and y_0 variables are simultaneously zeros ($\nabla D(\mu) = 0$) and where the Hessian matrix is positive definite, and in the case of multiple solutions, the one with the minimum objective function value has to be selected as the global minimum.

The gradients of the objective function (4.13) with respect to the centroid position (x_0, y_0) in the matrix form are given in (4.24) which are two bivariate cubic polynomials in $\mu = [x_0, y_0]$ as follows

$$\begin{aligned} g_1(x_0, y_0) = \frac{\partial D}{\partial x_0} &= a_1 x_0^3 + a_2 x_0^2 + a_3 x_0 + a_4 y_0^3 + a_5 y_0^2 \\ &+ a_6 y_0 + a_7 x_0^2 y_0 + a_8 x_0 y_0^2 + a_9 x_0 y_0 + a_{10}, \end{aligned} \tag{4.48}$$

$$\begin{aligned} g_2(x_0, y_0) = \frac{\partial D}{\partial y_0} &= b_1 x_0^3 + b_2 x_0^2 + b_3 x_0 + b_4 y_0^3 + b_5 y_0^2 \\ &+ b_6 y_0 + b_7 x_0^2 y_0 + b_8 x_0 y_0^2 + b_9 x_0 y_0 + b_{10}, \end{aligned} \tag{4.49}$$

with ten nonzero coefficients in each gradient polynomial: $a_1, .., a_{10}$, for $g_1(x_0, y_0)$, and $b_1, .., b_{10}$ for $g_2(x_0, y_0)$. The unique terms of the inverse covariance matrix Σ^{-1} are for simplicity denoted

with C_{11}, C_{12}, and C_{22}. The coefficients of the gradient polynomials are

$$
\begin{aligned}
a_1 &= C_{11}^2 s_1, & b_4 &= C_{22}^2 s_1,\\
a_2 &= -3C_{11} s_2, & b_5 &= -3C_{22} s_3,\\
a_3 &= 2s_4 + 2C_{11} s_7, & b_3 &= 2s_6 + 2C_{12} s_7,\\
a_4 &= C_{12} C_{22} s_1, & b_1 &= C_{11} C_{12} s_1,\\
a_5 &= -C_{22} s_2 - 2C_{12} s_3, & b_2 &= -C_{11} s_3 - 2C_{12} s_2,\\
a_6 &= 2s_6 + 2C_{12} s_7, & b_6 &= 2s_5 + 2C_{22} s_7,\\
a_7 &= 3C_{11} C_{12} s_1, & b_8 &= 3C_{22} C_{12} s_1,\\
a_8 &= (C_{11} C_{22} + 2C_{12}^2) s_1, & b_7 &= (C_{11} C_{22} + 2C_{12}^2) s_1,\\
a_9 &= -4C_{12} s_2 - 2C_{11} s_3, & b_9 &= -4C_{12} s_3 - 2C_{22} s_2,\\
a_{10} &= -2s_8, & b_{10} &= -2s_9,
\end{aligned}
$$

which contain, except the terms of the inverse covariance matrix, the nine composite sums $(s_1, .., s_9)$ as functions of the terms of inverse covariance matrix again and input samples' positions, values, and weights, which form the 13 auxiliary sums $s_{jkl} = \sum_{i=1}^{m} w_i x_i^j y_i^k z_i^l$ as follows:

$$
\begin{aligned}
s_1 &= s_{000},\\
s_2 &= C_{11} s_{100} + C_{12} s_{010},\\
s_3 &= C_{12} s_{100} + C_{22} s_{010},\\
s_4 &= C_{11}^2 s_{200} + 2C_{11} C_{12} s_{110} + C_{12}^2 s_{020},\\
s_5 &= C_{12}^2 s_{200} + 2C_{12} C_{22} s_{110} + C_{22}^2 s_{020},\\
s_6 &= C_{11} C_{12} s_{200} + (C_{11} C_{22} + C_{12}^2) s_{110} + C_{12} C_{22} s_{020},\\
s_7 &= 1/2(C_{11} s_{200} + 2C_{12} s_{110} + C_{22} s_{020}) - s_{001},\\
s_8 &= 1/2(C_{11}^2 s_{300} + 3C_{11} C_{12} s_{210} + (C_{11} C_{22} + 2C_{12}^2) s_{120}\\
&\quad + C_{12} C_{22} s_{030}) - C_{11} s_{101} - C_{12} s_{011},\\
s_9 &= 1/2(C_{22}^2 s_{030} + 3C_{12} C_{22} s_{120} + (C_{11} C_{22} + 2C_{12}^2) s_{210}\\
&\quad + C_{11} C_{12} s_{300}) - C_{12} s_{101} - C_{22} s_{011}.
\end{aligned}
$$

Finding a minimum of an objective function $D(\mu)$ concerning the centroid μ is a non-trivial problem since it requires, in this case, solving the nonlinear system of two coupled bivariate cubic polynomials. The solution of such a nonlinear system of equations is usually searched for by utilizing iterative optimization techniques such as Newton's method or trust region algorithm [88]. However, the iterative techniques require high computational costs and highly depend on an initial guess for the centroid.

4.2.1 Method of resultants

To analytically estimate the optimal solution for the centroid by minimizing (4.47), the necessary condition on $\hat{\mu}$ is $\nabla D(\hat{\mu}) = \mathbf{0}$, i.e., gradients in both directions must be zero. To solve this system of nonlinear equations, the theory of resultants [78] is applied. The resultant is a determinant of the Sylvester matrix [89], formed from the coefficients of two gradient polynomials. The auxiliary resultant polynomial, obtained by calculating this determinant, is equal to zero only when both gradient polynomials have a common root [90]. To solve the bivariate nonlinear system, two Sylvester matrices and two resultants were calculated where each resultant is a function of one variable (x_0 or y_0), while the other variable is treated as a constant,

$$\mathrm{Res}(g_1, g_2, y_0) = \begin{vmatrix} a_{y3} & 0 & 0 & b_{y3} & 0 & 0 \\ a_{y2} & a_{y3} & 0 & b_{y2} & b_{y3} & 0 \\ a_{y1} & a_{y2} & a_{y3} & b_{y1} & b_{y2} & b_{y3} \\ a_{y0} & a_{y1} & a_{y2} & b_{y0} & b_{y1} & b_{y2} \\ 0 & a_{y0} & a_{y1} & 0 & b_{y0} & b_{y1} \\ 0 & 0 & a_{y0} & 0 & 0 & b_{y0} \end{vmatrix} = r_{y_5} x_0^5 + r_{y_4} x_0^4 + r_{y_3} x_0^3 + \ldots + r_{y_0}, \quad (4.50)$$

where $a_{y3} = a_4, a_{y2} = a_5 + a_8 x_0, a_{y1} = a_6 + a_9 x_0 + a_7 x_0^2, a_{y0} = a_{10} + a_3 x_0 + a_2 x_0^2 + a_1 x_0^3, b_{y3} = b_4, b_{y2} = b_5 + b_8 x_0, b_{y1} = b_6 + b_9 x_0 + b_7 x_0^2, b_{y0} = b_{10} + b_3 x_0 + b_2 x_0^2 + b_1 x_0^3$, and

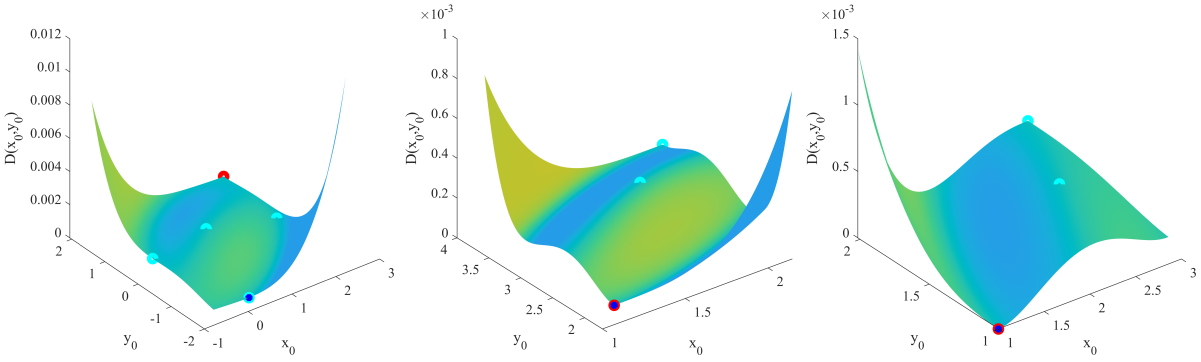
$$\mathrm{Res}(g_1, g_2, x_0) = \begin{vmatrix} a_{x3} & 0 & 0 & b_{x3} & 0 & 0 \\ a_{x2} & a_{x3} & 0 & b_{x2} & b_{x3} & 0 \\ a_{x1} & a_{x2} & a_{x3} & b_{x1} & b_{x2} & b_{x3} \\ a_{x0} & a_{x1} & a_{x2} & b_{x0} & b_{x1} & b_{x2} \\ 0 & a_{x0} & a_{x1} & 0 & b_{x0} & b_{x1} \\ 0 & 0 & a_{x0} & 0 & 0 & b_{x0} \end{vmatrix} = r_{x_5} y_0^5 + r_{x_4} y_0^4 + r_{x_3} y_0^3 + \ldots + r_{x_0}, \quad (4.51)$$

where $a_{x3} = a_1, a_{x2} = a_2 + a_7 y_0, a_{x1} = a_3 + a_9 y_0 + a_8 y_0^2, a_{x0} = a_{10} + a_6 y_0 + a_5 y_0^2 + a_4 y_0^3, b_{x3} = b_1, b_{x2} = b_2 + b_7 y_0, b_{x1} = b_3 + b_9 y_0 + b_8 y_0^2, b_{x0} = b_{10} + b_6 y_0 + b_5 y_0^2 + b_4 y_0^3$.

The determinants of these Sylvester matrices are found symbolically, yielding two resultant polynomials with 6 coefficients in each polynomial ($r_{y_0}, .., r_{y_5}, r_{x_0}, .., r_{x_5}$). These coefficients are functions of 9 composite sums ($s_1, .., s_9$) and the three unique terms of the inverse covariance

matrix (C_{11}, C_{12}, C_{22}). Generally, two bivariate cubic polynomials yield two resultant polynomials of the 9th order. However, we obtained two resultants of the reduced 5th due to common terms in coefficients of gradient polynomials. Since the resultant is a polynomial of 5th order in one variable, each resultant yields 5 roots. Only the real roots should be considered, which can be 1,3, or 5. Even though the maximal number of possible combinations of solution is 25, we have experimentally verified that it is always at most 5. The second derivative test can be applied to multiple stationary points to inspect which of the stationary points are minima indeed. Finally, the objective function has to be evaluated for all minima to find the global one. The explained procedure has to be performed only if the number of input samples is very small and comparable with the number of unknowns since, only in those situations, up to 5 stationary points may occur. When the number of input samples is larger, there is only one combination of real roots, which is also the global minimum indeed.

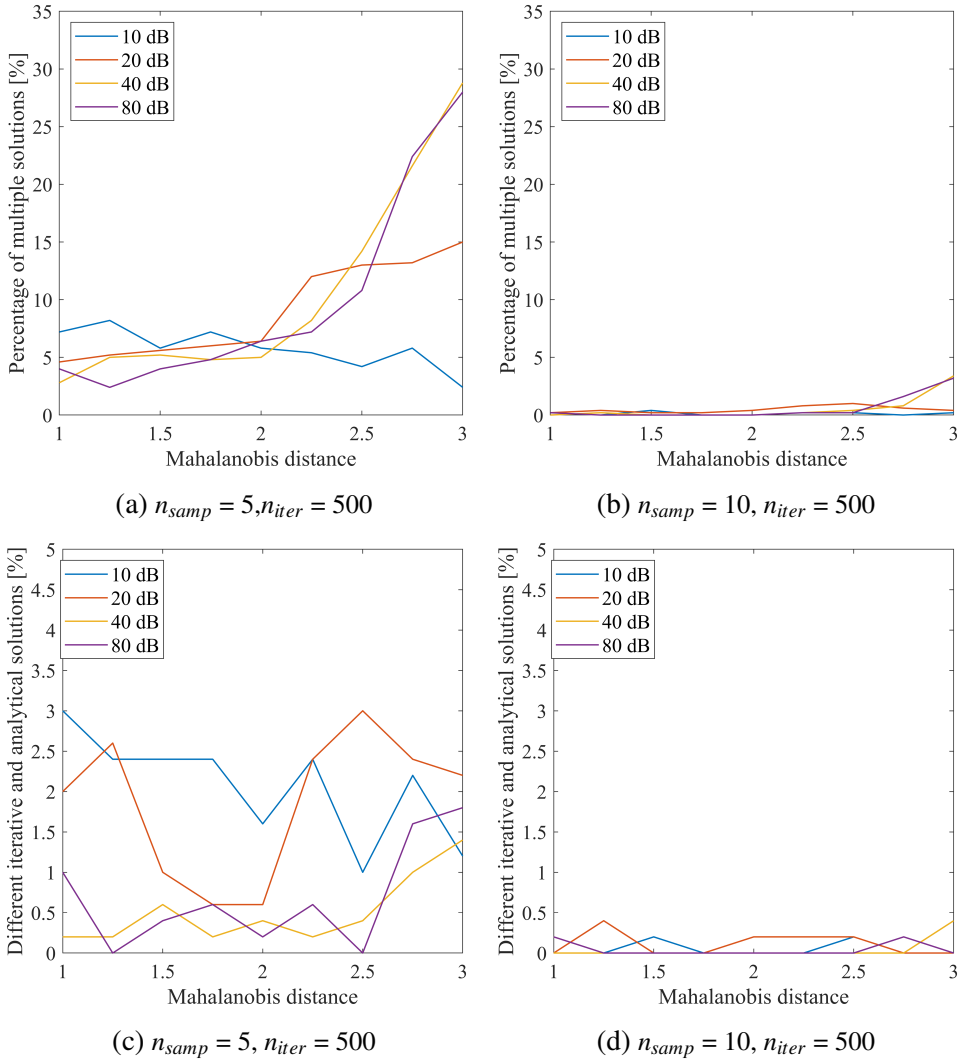
The analytical results of centroid estimation were compared with those obtained by *fminunc* function in MATLAB, which uses the iterative procedure for nonlinear optimization, and the results are shown in Figs. 4.7 and 4.8.



(a) SNR = 10dB, n_{samp} = 2, k = 1 (b) SNR = 20dB, n_{samp} = 2, k = 2 (c) SNR = 80dB, n_{samp} = 5, k = 3

Figure 4.7: Objective function and stationary points (cyan), iteratively obtained minimum (red), and analytically obtained minimum (blue)

The results show that the number of stationary points, and consequently possible difference of iterative and analytical solution for centroid, depends on the given signal-to-noise ratio (SNR), the number of input samples, and the closeness of the input samples' positions to the actual centroid position. The number of extrema increases with the expansion of the input domain from the actual centroid position. The number of extrema and the equality of analytical and iterative solutions also depend on the given SNR. The experimental results show that for the case of low SNRs, the iterative and analytical procedures yield different solutions more often since the iterative methods can get stuck in the local minimum while the proposed analytical approach always yields the global minimum. The number of extrema reduces when the number of samples is much larger than the number of unknowns. In most cases, when the number of



(a) n_{samp} = 5,n_{iter} = 500

(b) n_{samp} = 10, n_{iter} = 500

(c) n_{samp} = 5, n_{iter} = 500

(d) n_{samp} = 10, n_{iter} = 500

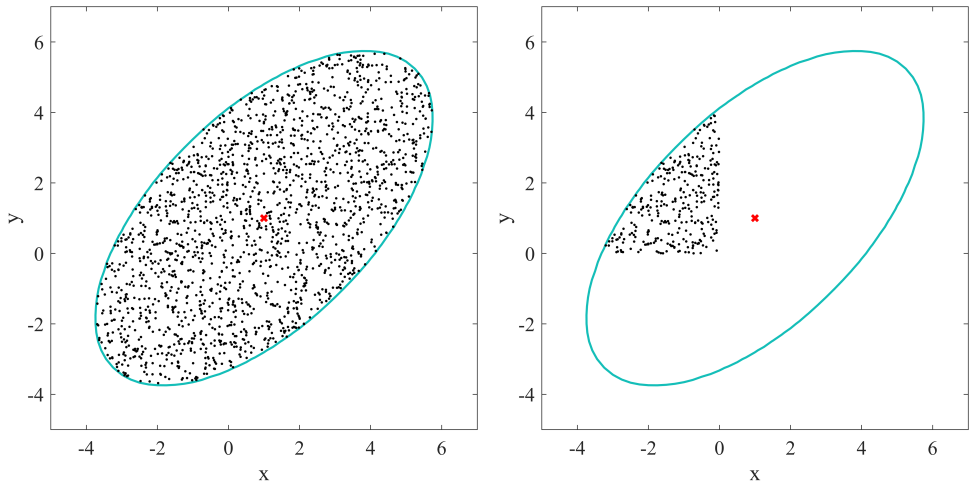
Figure 4.8: Percentage of multiple real roots and different iterative and analytical solutions for different factors of Mahalanobis distance and SNRs

input samples is large enough, there is only one extremum, which is the minimum, and both analytical and iterative procedures yield the same solution.

4.2.2 Accuracy of the method of resultants in centroid estimation

To analyze the accuracy and the speed of the proposed method of resultants in centroid estimation of the 2D Gaussian profile from noisy data, the proposed method results were compared with the results of the center-of-gravity (CoG) method in the domain of values and the iterative LS method w/wo provided analytical gradients of the objective function. The actual 2D Gaussian profile parameters were: the centroid position $\mu = [x_0, y_0] = [1, 1]$, the linear $A = 1000$, and the unique inverse covariance matrix terms $(C_{11}, C_{12}, C_{22}) = (5/8, -3/8, 5/8)$. The profile with the specified parameters was synthesized and first evaluated over uniformly and randomly sampled ROI. The additive Gaussian noise ε_i was added to the ideal sample value to satisfy the

required signal-to-noise ratio (SNR) defined as $\mathrm{SNR}_{\mathrm{dB}} = 20\log_{10}\frac{A/(2\pi\sqrt{|\mathbf{C}|})}{\sigma_\varepsilon}$, where σ_ε denotes the standard deviation of the added noise. Two types of estimation domains were considered: symmetrical input domain within maximal Mahalanobis distance $k = 3$ from the given centroid position, and asymmetrical and truncated input domain as shown in Figs. 4.9a and 4.9b.



(a) Symmetric estimation domain ($k \leq 3, m = 1897$).

(b) Truncated estimation domain ($k \leq 3, x < 0, y > 0, m = 306$).

Figure 4.9: Estimation domain with uniformly distributed positions of input samples

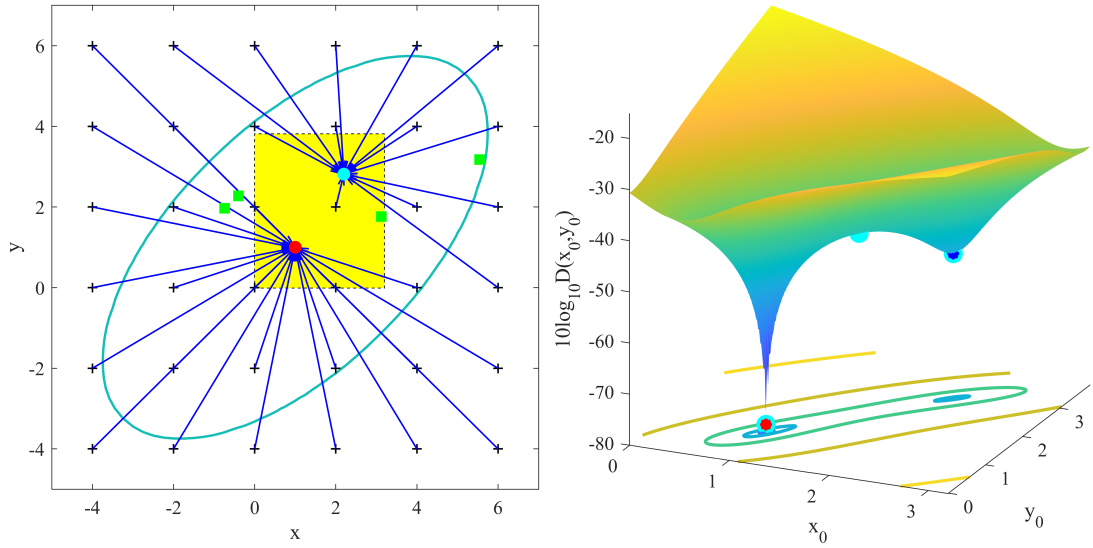
The Monte Carlo simulation was performed for different numbers of input samples from the set $m = \{4, 40, 400, 4000\}$ and different SNRs from the set $\mathrm{SNR}_{\mathrm{dB}} = \{10, 20, 30, 40, 50, 60\}$. The number of trials in Monte Carlo simulations depended on the number of input samples and was $n_{iter} = \{50000, 5000, 500, 50\}$, respectively, so that the product of the number of trials and the number of input samples was constant, making the centroid estimator equally reliable for all cases.

The experiments were conducted on a 2.5 GHz Intel Core i5 processor with 8 GB of RAM under the Matlab environment. The solution of the CoG method was provided as an initial guess for the iterative LS method, which uses the native MATLAB *fminunc*. For the case of specified analytical gradients, the gradient descent method was used. The maximal number of function evaluations was limited to 500, while the optimality tolerance and the current point tolerance were set to 10^{-14}. As a measure of accuracy, the mean squared error (MSE) of the form

$$MSE(\hat{\mu}) = E[||\hat{\mu} - \mu||^2] \approx \frac{1}{n_{iter} - 1}\sum_{i=1}^{n_{iter}} ||\hat{\mu}_i - \mu||^2 \tag{4.52}$$

was used to simultaneously take into account the influences of both the estimator's variance and its bias.

Fig. 4.10 clearly illustrates the advantage of the proposed method of resultants in centroid estimation. Namely, when the number of input samples m is small and comparable with the number of unknowns and the SNR is high, the multiple stationary points occur, and the iterative



(a) Initial and final points of the LS method (blue arrows). Plus signs denote the initial centroids. Yellow shows the surface domain in (b).

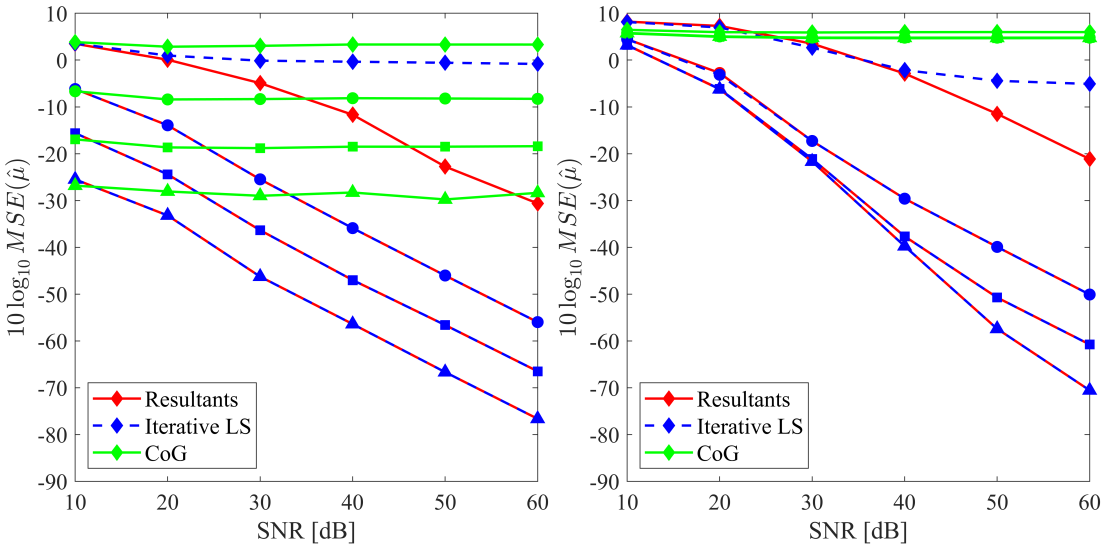
(b) The objective with two minima and one inflection (cyan) (LS minimum (blue), minimum of resultants (red)).

Figure 4.10: Comparison of the LS method and the method of resultants for different initial centroids using the same 4 input samples (green). $\mathrm{SNR}_{\mathrm{dB}} = 60$.

LS method can get stuck in a local minimum if the initial guess for the centroid position is far from the actual (Fig. 4.10a). However, the proposed method of resultants calculates the first and the second derivative tests for all stationary points to find the global minimum without any initialization.

MSEs in centroid estimation of the 2D Gaussian profile for the case of symmetric and asymmetric estimation domains are shown in Fig. 4.11. For the case of the symmetric input domain, the MSEs are primarily caused by the estimators' variances. The variances and, consequently, MSEs of the CoG method are invariant of SNR and depend only on m. The variances and the MSEs of the iterative LS method and the method of resultants reduce with the increase of m, but also with the increase of SNR. The 10-fold increase in m reduces the estimation variance 10 times (10 dB MSE reduction, as expected). The iterative LS and the resultant methods achieve the same MSEs for larger m ($m \geq 40$). However, for a small m, as for $m = 4$, the method of resultants has a much smaller MSE due to optimality, especially for higher SNRs. Namely, in those situations, the objective has multiple stationary points, and the iterative LS method can end up in a local minimum, depending on the initial guess for the centroid. The method of resultants always yields the global minimum, even for the small m (red line).

For the case of the asymmetric estimation domain, the CoG method has large and constant MSEs for all SNRs, primarily due to the inherent bias. The iterative method and the method of resultants exhibit nonlinear behavior with lower asymptotic efficiency, especially for small m and low SNRs, due to higher bias and higher variance. However, for small m and high SNRs, the method of resultants yields much smaller MSEs again.



(a) **Symmetric input domain** ($k \leq 3$)

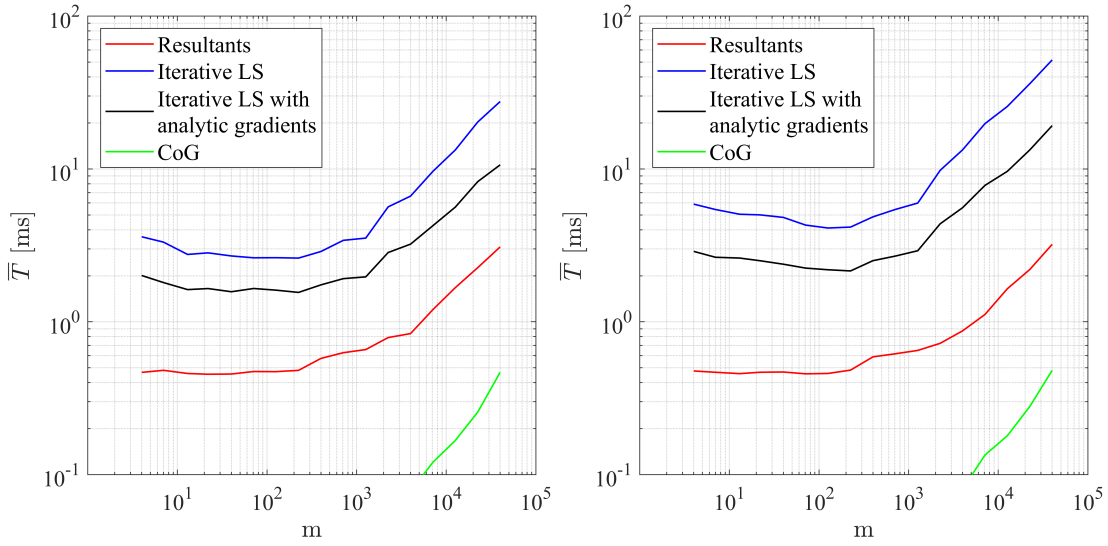
(b) **Asymmetric input domain** ($k \leq 3, x < 0, y > 0$).

Figure 4.11: MSEs in dB of the method of resultants (red), the LS method (blue), and the CoG method (green) for different numbers of input samples m: 4 - diamonds, 40 - circles, 400 - squares, 4000 - triangles ($m \cdot n_{iter} = 200000$).

4.2.3 Complexity of the method of resultants in centroid estimation

The computational complexity of the resultant method consists of variable and fixed parts. The variable part is related to the calculation of 9 composite sums $(s_1, .., s_9)$. It linearly depends on the number of input samples m and equals $12m + 44$ multiplications and $13m + 11$ additions, which are MAC operations. The fixed part is related to the formation of coefficients in resultant polynomials in (4.50) and (4.51). The calculation of coefficients requires 9296 multiplications and 1594 additions. Using the composite sums $(s_1, .., s_9)$ instead of simple sums s_{jkl} in analytical expressions reduces the total number of numerical operations by 43%. Finally, the fixed part also includes the polynomial rooting complexity.

The complexity of the method of resultants is compared with the iterative LS method, with and without analytical gradients, and with the CoG method by measuring the average elapsed time for different numbers of input samples. The log-log plots of average elapsed times for symmetric and asymmetric input domains are shown in Fig. 4.12. The CoG method is the fastest but with the least accuracy. The proposed method of resultants is much faster than the iterative LS method, especially when the gradients are calculated numerically, 5 to 9 times in the case of symmetric and 8 to 19 times in the case of asymmetric input domain. When the analytical gradients are provided, this relative difference is reduced, but the method of resultant is still 3 to 4 times faster in the case of symmetric and 4 to 7 times faster in the case of asymmetric input domain.



(a) Symmetric input domain (4.9a)

(b) Asymmetric input domain (4.9b)

Figure 4.12: Log-log plots of average elapsed times of the resultant method, the LS method with and without analytical gradients, and the CoG method for different numbers of input samples m ($\text{SNR}_{\text{dB}} = 30$ and $m \cdot n_{iter} = 400000$).

Chapter 5

Low-dose 2D PET imaging

Positron emission tomography (PET) is a medical imaging technique widely used for diagnosing and monitoring various diseases [91, 92]. The accurate and fast reconstruction of underlying unknown volume from the acquired PET data is the subject of many research works. In this section, only the 2D PET imaging is considered, i.e., only the transversal slices of the unknown 3D function are considered. In the proposed algorithms for the 2D PET reconstruction, the acquired data are coincident events proportional to the activity of the radiotracer injected into the patient's body. Namely, the radiotracer is metabolized in the places of the unknown process samples, thus generating the annihilation photons that travel in opposite directions along the straight lines, also called the lines of response (LORs). Two photons detected in the sufficiently short time window make a coincident event detected by the PET scanners typically placed at the ring surrounding the patient's body and the edge of the field of view (FOV). In the ideal noiseless case, the number of coincident events detected by the pair of detectors is proportional to the activity of the radiotracer in the space between those detectors. The detected coincident events represent the acquired PET data, which are further used in the reconstruction process.

The aim of this research is to reduce the dose of injected radiotracer by proposing a method for 2D PET image reconstruction from a small number of process samples using the intersections of LORs. Each point source of the underlying unknown process generates a series of LORs at random angles. The number of generated LORs in a given time depends on the activity of the particular point source, i.e., the amount of radiotracer comprised in that spatial point. Information about the point source's activity is obtained through a long-term acquisition process due to the quantum nature of coincident events, i.e., an integer number of generated LORs. Such a long-term acquisition aims to obtain activity information about all point sources within the volume of interest throughout the entire acquisition period. A shorter acquisition time limits the number of generated LORs, directly impacting the accuracy of measurements and reconstruction. The parameters of LORs, the signed distance s between the LOR and the origin of the coordinate system, and angle θ, which specifies the LOR's orientation, are deter-

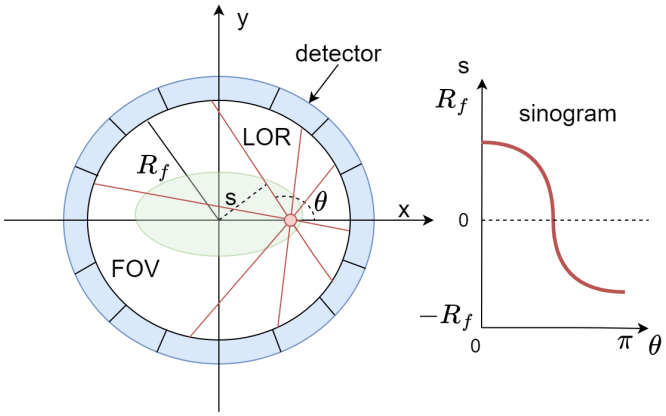
mined in the detection process. At the same time, the information on the actual point source position along the LOR remains unknown. The motivation for using the intersections of LORs as starting points to generate virtual LORs is their quadratically larger number and the fact that the distribution of process samples is related to the distribution of intersections of LORs, as described later in this chapter. Namely, the number of intersections of LORs N_c is associated with the number of coincident events (LORs) N as $N_c = N(N-1)/2$. It is especially interesting to use the intersections of LORs for generating virtual LORs in the cases of extra-low-dose PET imaging where the number of coincident events is extremely small. It is proposed that the intersections of LORs can be used as a new technique of PET data acquisition since the mean and the covariance matrix of the process samples and those of intersections of LORs are related. Moreover, the reconstruction from such a small number of process samples can significantly improve the temporal resolution of the reconstruction. It can be used in dynamic PET imaging, where fast reconstruction is extremely important.

This chapter consists of several sections. Section 5.1 describes 2D PET imaging and provides an overview of existing methods for reconstructing the unknown underlying process in 2D PET. Section 5.2 describes a proposed method for data acquisition in low-dose 2D PET images, which uses intersections of LORs to generate more virtual LORs and thereby increase the available dataset. In the following, the analytical relationship between the spatial statistical properties of the process and the intersections of LORs is described in detail. In addition, few experiments were made to demonstrate the effectiveness of the proposed method. At the end of the chapter, there is a discussion about the advantages and drawbacks of the proposed method and suggestions for further research directions.

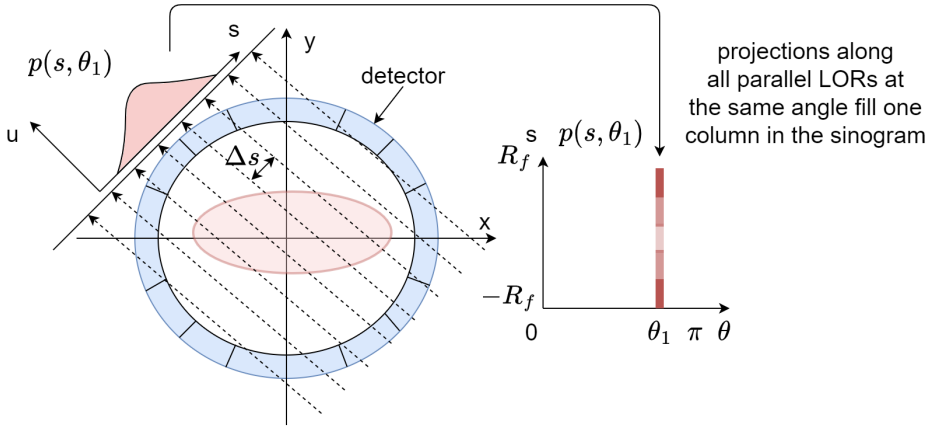
5.1 2D PET imaging

In 2D PET imaging, the detected LORs are used for the reconstruction of 2D images, i.e., slices that can be stacked to form the unknown 3D volume activity. Each LOR is parametrized with the angle θ and the signed distance s of the LOR from the origin of the coordinate system located at the center of FOV. The LORs are generated at places of the underlying process samples and are recorded in the 2D function called a sinogram. The sinogram is a 2D histogram of coincidental events for all (θ, s) pairs across all angles $\theta \in [0, \pi]$ and distances $|s| < R_f$ where R_f is the radius of the FOV. Geometrically, the sinogram represents the set of projections, a line integrals of the unknown 2D object activity over all angles θ and distances $|s| < R_f$. The name sinogram comes from the fact that a single point source of the unknown 2D object is mapped into a single sinusoid in the sinogram, as shown in Fig. 5.1a. One column of the sinogram represents a projection at a single angle along all parallel LORs at distances $|s| < R_f$ (Fig. 5.1b). Unlike other medical imaging techniques such as CT or MRI, the choice of desired projection angle

θ is not feasible during PET imaging, but a 2D histogram is formed by simply counting the discrete coincident events that belong to particular discretized angle θ and signed distance s. The suitable discretization step of θ and s sinogram parameters is determined by the number of detectors N_d at the edge of FOV and their mechanical properties. The number of detected coincident events from which the underlying process is reconstructed is usually huge and can exceed 10^9 with modern PET scanners.



(a) Projection of the one process sample across all angles



(b) Projection of the unknown process at a single angle along all parallel LORs

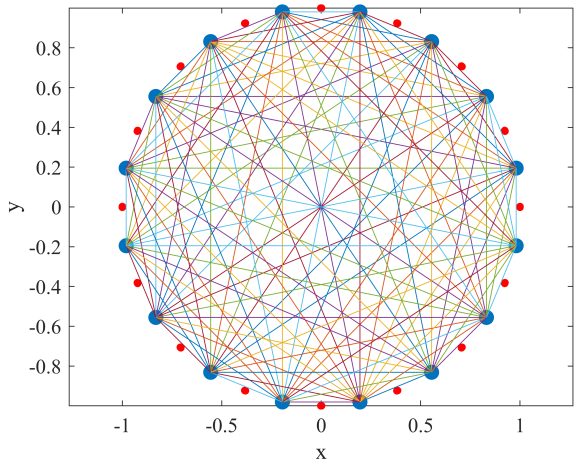
Figure 5.1: 2D PET imaging

5.1.1 Data acquisition in 2D PET

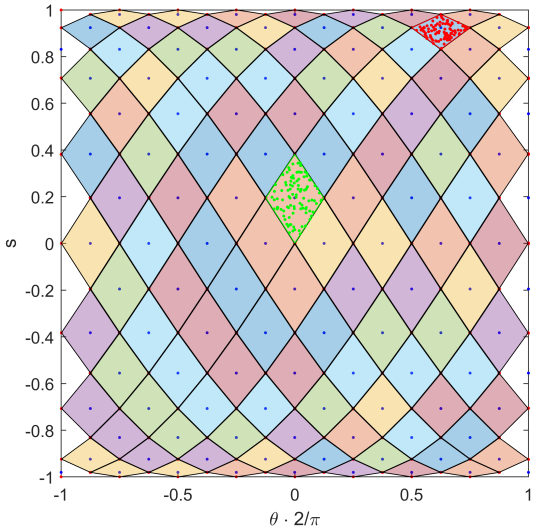
The 2D PET data are projections **p**, i.e., the line integrals of the unknown 2D function $f(x,y)$ along the LORs. The projection at the specified distance s and angle θ can be calculated as $p(s,\theta) = Xf(x,y)$, where the operator X is called the X-ray transform [93], i.e., Radon transform of the function $f(x,y)$ along the LOR as follows

$$p(s,\theta) = Xf = \int_{u=-\infty}^{\infty} f(x = s\cos\theta - u\sin\theta, y = s\sin\theta + t\cos\theta)du, \; |s| < R_f, \theta \in [0,\pi] \quad (5.1)$$

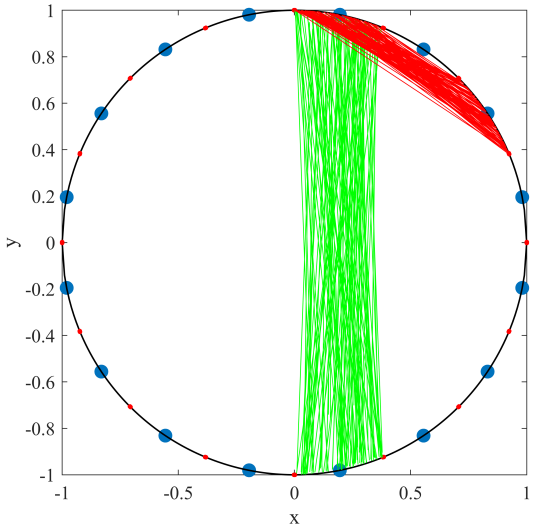
where u is the integration variable along the chosen integration line (LOR) with the angle θ. Due to the uniform distribution of detectors over the edge of the circular FOV and their finite number N_d, the set of possible pairs of variables s and θ in the sinogram is finite and predetermined. However, the uniform distribution of detectors over the circular FOV causes a non-uniform distribution of areas covered by the individual detector pairs in the spatial domain and, consequently, the non-uniform distribution of corresponding projection areas in the sinogram. Examples of the uniform distribution of 16 detectors over the ring of the FOV and the non-uniform distribution of projection areas in the sinogram, which are associated with individual detector pairs, are shown in Fig. 5.2.



(a) PET scanner with $N_d = 16$ detectors uniformly distributed over the ring of FOV and LORs that connect all possible detector pairs. Detector centers are denoted with blue dots, while edges are denoted with red dots



(b) Non-uniform distribution of acquired 2D PET data in sinogram and the projection areas related to particular detector pairs



(c) Green and red LORs which correspond to uniformly sampled green and red polygons in sinogram

Figure 5.2: 2D PET data acquisition

In order to use the acquired sinogram data in analytical reconstruction algorithms, it is often necessary to resample the data on a uniform rectangular grid in the variables s and θ. In [94], it is recommended to use the following discretization of the sinogram

$$s_k = k\Delta s,\ -N_s < k < N_s, N_s = N_d/\pi, \Delta s = R_f\pi/N_d, \tag{5.2}$$

$$\theta_j = j\Delta\theta,\ j = 0,..,N_\theta - 1, \Delta\theta = 2\pi/N_d, N_\theta = N_d/2. \tag{5.3}$$

The coincident event determined with its actual measured s and θ value pair joins the nearest bin in a uniformly sampled sinogram based on the smallest distance to the center of that bin. Therefore, the accuracy of the spatial localization of the LOR depends on the sensor's geometrical dimensions and the number of sensors in the ring.

5.1.2 2D PET image reconstruction techniques

The reconstruction problem involves determining the unknown distribution $f(x,y)$ from the set of projections $\mathbf{p}$. The proposed techniques for 2D PET data reconstruction from projections can be divided into analytical and iterative reconstruction techniques. The review of the proposed methods can be found in [94, 95, 96, 97]. Analytical reconstruction techniques assume that the data are deterministic, the acquisition process is noiseless, i.e., data are pre-corrected by removing the scattering, attenuation, and random events, and the reconstruction can be described as a linear inverse problem [94]. These methods are based on the central slice theorem [98], which states that the 1D Fourier transform of projection at angle θ across all parallel LORs is identical to the section through the center of the 2D Fourier transform of the 2D object activity at the same angle θ. The most widely used technique from this group is the Filtered back-projection (FBP) [93, 98] due to the fast reconstruction and the ability of noise control.

The other group of techniques includes iterative reconstruction techniques, which iteratively correct the unknown 2D image to achieve that the projections of a corrected image are the most similar to the given projections. The Maximum Likelihood Expectation Maximization (MLEM) and Ordered Subset Expectation Maximization (OSEM) [94] are the most important methods from this group.

Many researchers attempt to reconstruct the standard-dose PET image from the low-dose PET inputs to reduce the risks of radiation exposure in PET imaging. Since low-dose PET imaging implies the reduction of the injected radiotracer, the main challenge in low-dose PET imaging is the low signal-to-noise ratio (SNR) and, consequently, the low quality of the reconstructed image. The hardware performance or software techniques should be improved to achieve the standard-dose PET image quality from the low-dose inputs. The hardware performance can be improved by enhancing the sensitivity of PET sensors [99], thus increasing the percentage of detected events and reducing the attenuation and scattering. On the other hand,

there are many software techniques based on the image pre- or post-reconstruction denoising [100] or techniques that combine the low-dose PET inputs with inputs from other imaging modalities such as CT or MRI [101]. In [102], an iterative method was proposed that considers the low-dose reconstruction problem as a convex optimization problem combining a sinogram with the regularization term, which suppresses noise. In recent times, increasingly popular become machine learning [103, 104] and deep learning methods [105, 106, 107] which use both low-dose and standard-dose images to train models that can predict standard-dose images from low-dose inputs. However, the computational complexity of the mentioned reconstruction techniques is higher than the complexity of the simple analytic filtered back-projection algorithm.

5.1.3 Discrete Filtered Backprojection

Discrete Filtered Backprojection (FBP) method implies reconstruction of 2D function $f(x,y)$ by the back projection X^* of filtered projections p^F as

$$f(x,y) = X^* p^F(x,y) \approx \Delta\theta \sum_{j=0}^{N_\theta - 1} p^F(s,\theta_j) \tag{5.4}$$

where the variable $s = x\cos(\theta_j) + y\sin(\theta_j)$ and the filtered projections are calculated from the convolution of projections with the apodized ramp filter h as

$$p^F(k\Delta s,\theta_j) \approx \Delta s \sum_{k'=-N_s}^{N_s} p(k'\Delta s,\theta_j) h((k-k')\Delta s), \tag{5.5}$$

$$k = -N_s,..,N_s \tag{5.6}$$

$$h(s) = \int_{-\infty}^{\infty} \exp(2\pi j s\omega_s) w(\omega_s) |\omega_s| d\omega_s, \tag{5.7}$$

where ω_s is the frequency related to the variable s, and $w(\omega_s)$ is the low-pass filter that suppresses the noise dominant at high frequencies [95, 108]. A frequently used low-pass filter is the Hamming window

$$w_{ham}(\omega_s) = \begin{cases} (1+\cos(\pi\omega_s/\omega_c))/2, & |\omega_s| < \omega_c \\ 0, & |\omega_s| > \omega_c, \end{cases} \tag{5.8}$$

where ω_c denotes the cut-off frequency. The filtered projection at a specified point (s,θ_j) can be obtained by linear interpolation of the closest available projections as

$$p^F(s,\theta_j) \approx (k+1-\frac{s}{\Delta s}) p^F(k\Delta s,\theta_j) + (\frac{s}{\Delta s}-k) p^F((k+1)\Delta s,\theta_j), \tag{5.9}$$

where $k\Delta s \leq s < (k+1)\Delta s$. The back-projection operator X^* is the dual of the X-ray transform. Filtering sinogram projections is necessary since the back projection of sinogram projections

without filtering would yield a blurred reconstructed image. The blurring of the reconstructed image can be described as a convolution of an ideally reconstructed image with the cone filter [98] as a consequence of line integration and loss of information about the sample position along the LOR. Namely, the constant value is back-projected along the whole LOR in the back-projection step. In the ideal continuous case, the filtering of projections with a 1D ramp filter enhances the high frequencies and enables the reconstruction of the ideal 2D image but also enhances the noise, which is dominant at high frequencies. Therefore, the low-pass filter $w(\omega_s)$ is introduced to suppress the noise influence at high frequencies. Determining the cutoff frequency of such a filter represents the trade-off between noise and spatial resolution.

5.1.4 Maximum likelihood expectation maximization algorithm for 2D PET image reconstruction

The maximum likelihood expectation maximization (MLEM) is an iterative algorithm similar to the Lucy-Richardson deconvolution algorithm for astronomical images. The MLEM algorithm consists of a series of successive projections and back projections, which are used for iterative image correction, and can be described in a simplified manner as

$$\text{Image}^{(next)} = \text{Image}^{(current)} \cdot \frac{\text{Backproject}\left\{\frac{\text{Measured projections}}{\text{Projections of Image}^{(current)}}\right\}}{\text{Backproject}\{\mathbf{1}\}}, \tag{5.10}$$

where the measured projections represent the sinogram of detected coincident events (LORs), $\text{Image}^{(current)}$ is the current estimate of the reconstructed image, and $\mathbf{1}$ is the matrix of ones whose size is equal to the sinogram size. The algorithm assumes that the number of detected LORs, i.e., the number of photons emitted from each image pixel is a Poisson random variable. The objective function of the MLEM algorithm is the Poisson likelihood function, which is the joint PDF of Poisson random variables. The solution, i.e., the reconstructed image, is iteratively searched for in two steps. The expectation of this likelihood function for the current image estimate and measured projections is calculated in the first "E" step. The expected likelihood function is maximized to update the current image estimate in the second "M" step. These two steps are iteratively exchanged until convergence. This algorithm is described in detail in [109, 110].

5.1.5 Low-dose 2D PET imaging

The application of analytical reconstruction algorithms assumes the uniformly resampled sinogram obtained from a large number of coincident events. The reconstruction spatial resolution and accuracy depend on the sinogram spatial resolution but also on the SNR of the sinogram val-

ues (i.e., counts). In the low-dose 2D PET imaging, the dose of radiotracer and, consequently, the number of coincident events is extremely small. The small number of coincident events produces a sparse sinogram with many empty bins as consequence of the detection process that follows the Poisson distribution. In that case a further increase of the sinogram spatial resolution can not improve the reconstruction accuracy. Contrarily, the pre-undersampling of such a sinogram by grouping and summing the consecutive rows of the sinogram can enhance the quality of the reconstruction but at the expense of lower spatial resolution. The number of coincident events is directly proportional to the integration time and volume activity that depends on the amount of injected radiotracer. The main idea of this research is to reduce the amount of injected radiotracer while still enabling the detection and reconstruction of the underlying 2D process representing a chosen slice of the inspected volume. Furthermore, monitoring the dynamic underlying process requires high temporal resolution. The spatial resolution, temporal resolution, and the amount of injected dose are in triangular relation, as shown in Fig. 5.3. The reduction

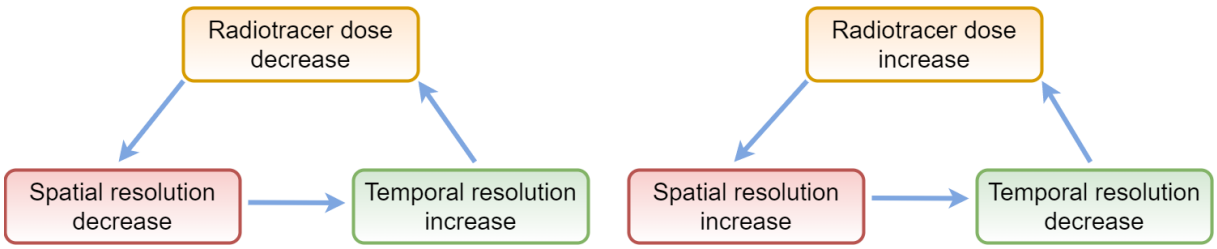


Figure 5.3: The relation between the radiotracer dose, spatial resolution, and temporal resolution

of radiotracer dose or reduction of the integration time decreases the number of acquired coincident events, thus causing a necessary reduction of the sinogram spatial resolution to achieve the required SNR (events per bin), but at the same time increasing the temporal resolution due to shorter integration. To achieve higher spatial resolution, either dose must be increased or integration lengthened, thus causing lower temporal resolution for dynamic events. Finally, the reconstruction accuracy is also limited by the minimum number of required samples (and corresponding LORs) that should be large enough to represent the unknown underlying process statistically.

5.2 Proposed method for low-dose 2D PET imaging

The peculiarity of this research represents the utilization of intersections of LORs as starting points for the artificial generation of a more significant number of coincident events whose statistical properties are inherited from the properties of the underlying unknown process. To the best of our knowledge, this is the first mention of using the intersections of LORs for boosting the number of coincident events and PET image reconstruction. In this research, the virtual LORs are generated at the places of intersections of original LORs to boost the available PET

dataset. The relation between the spatial distribution of intersections of LORs and the distribution of process samples from which those LORs originate is derived. It sets the ground for reconstructing an unknown 2D function from a small number of process samples and corresponding LORs, i.e., by using a very low dose of the injected radiotracer. Otherwise, the reconstruction from such a small number of LORs using conventional methods would be almost impossible, as it is demonstrated in the following sections. However, such an approach has a drawback since the image reconstructed from virtual LORs is biased due to the difference between spatial distributions of intersections and the original process samples, and it also has an estimation variance due to the finite (small) number of underlying process samples (LORs) that are used for reconstruction. This variance can be reduced by increasing the integration time (i.e., using a larger sample size) or by averaging several neighboring reconstructions from virtual LORs drawn from different process samples, but the estimation bias remains. Since the number of virtual LORs is quadratically larger than the number of process samples, a relatively high spatial and temporal resolution can be achieved simultaneously by using the virtual LORs, while the amount of injected radiotracer can be significantly reduced.

5.2.1 The first and the second moments of intersections of LORs and their relationship to corresponding moments of the original process

This section describes how the parameters of the real-valued 2D unknown distribution, the first and the second moment are related to the first and the second moments of intersections of LORs, where the LORs pass through the random process samples. Let $z_1 = (x_1, y_1)^T$ and $z_2 = (x_2, y_2)^T$ be two points drawn from an arbitrary distribution with the finite first and the second moments, i.e., the covariance matrix $\mathbf{C} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$ and the mean value $\mu = [\mu_x, \mu_y]$. The process mean μ and the covariance matrix $\mathbf{C}$ can be computed as $\mu = E[\mathbf{z}]$ and $\mathbf{C} = E[(\mathbf{z}-\mu)(\mathbf{z}-\mu)^T]$, where E denotes the expectation.

Let each of the two points has an associated LOR

$$y - y_1 = \frac{s_1}{c_1}(x - x_1), y - y_2 = \frac{s_2}{c_2}(x - x_2), \tag{5.11}$$

where $s_1 = \sin(\theta_1)$, $s_2 = \sin(\theta_2)$, $c_1 = \cos(\theta_1)$, $c_2 = \cos(\theta_2)$, while LOR angles θ_1 and θ_2 are assumed to be random variables with uniform distribution in the range from $-\pi/2$ to $\pi/2$. The intersection point $z_s = (x_s, y_s)$ can be found as the intersection of these two lines:

$$y_s - y_1 = \frac{s_1}{c_1}(x_s - x_1), y_s - y_2 = \frac{s_2}{c_2}(x_s - x_2), \tag{5.12}$$

or in the matrix form

$$\mathbf{z_s} = \begin{bmatrix} x_s \\ y_s \end{bmatrix} = \frac{1}{c_1 s_2 - c_2 s_1} \cdot \begin{bmatrix} -c_2 s_1 & c_1 c_2 & c_1 s_2 & -c_1 c_2 \\ -s_1 s_2 & c_1 s_2 & s_1 s_2 & -c_2 s_1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \end{bmatrix}. \tag{5.13}$$

The expectation of intersections is

$$\begin{aligned} E[x_s] &= E\left[\frac{c_1 c_2 y_1 - c_1 c_2 y_2 - c_2 s_1 x_1 + c_1 s_2 x_2}{c_1 s_2 - c_2 s_1}\right] \\ &= E\left[\frac{c_1 c_2 - c_1 c_2}{c_1 s_2 - c_2 s_1}\right] \mu_y + E\left[\frac{-c_2 s_1 + c_1 s_2}{c_1 s_2 - c_2 s_1}\right] \mu_x = \mu_x, \end{aligned} \tag{5.14}$$

$$\begin{aligned} E[y_s] &= E\left[\frac{c_1 s_2 y_1 - c_2 s_1 y_2 - s_1 s_2 x_1 + s_1 s_2 x_2}{c_1 s_2 - c_2 s_1}\right] \\ &= E\left[\frac{c_1 s_2 - c_2 s_1}{c_1 s_2 - c_2 s_1}\right] \mu_y + E\left[\frac{s_1 s_2 - s_1 s_2}{c_1 s_2 - c_2 s_1}\right] \mu_x = \mu_y. \end{aligned} \tag{5.15}$$

The equations above show that the expectation of intersections is equal to the process mean: $\mu_s = E[\mathbf{z_s}] = \mu$. Once we have determined the first moment, we can also derive the second moment of intersections which is calculated as

$$\mathbf{C_s} = E[(\mathbf{z_s} - \mu)(\mathbf{z_s} - \mu)^T],$$

where $E(x_1^2) = E(x_2^2) = C_{11}, E(x_1 y_1) = E(x_2 y_2) = E(y_1 x_1) = E(y_2 x_2) = C_{12}, E(y_1^2) = E(y_2^2) = C_{22}$, while mixed expectations equal zero. Finally, terms of the covariance matrix of the intersections $\mathbf{C_s} = \begin{bmatrix} C_{s11} & C_{s12} \\ C_{s21} & C_{s22} \end{bmatrix}$ are connected with the terms of the covariance matrix of the process $\mathbf{C}$ via the expectation of the transformation matrix $\mathbf{M}$ as follows

$$\begin{bmatrix} C_{s11} \\ C_{s12} \\ C_{s22} \end{bmatrix} = E[\mathbf{M}] \begin{bmatrix} C_{11} \\ C_{12} \\ C_{22} \end{bmatrix}, \tag{5.16}$$

where transformation matrix $\mathbf{M}$ is

$$\mathbf{M} = k^2 \begin{bmatrix} c_1^2 s_2^2 + c_2^2 s_1^2 & -2 s_2 c_1^2 c_2 - 2 s_1 c_1 c_2^2 & 2 c_1^2 c_2^2 \\ c_2 s_1^2 s_2 + c_1 s_1 s_2^2 & -4 c_1 c_2 s_1 s_2 & s_2 c_1^2 c_2 + s_1 c_1 c_2^2 \\ 2 s_1^2 s_2^2 & -2 c_2 s_1^2 s_2 - 2 c_1 s_1 s_2^2 & c_1^2 s_2^2 + c_2^2 s_1^2 \end{bmatrix}, \tag{5.17}$$

where (s_1, s_2, c_1, c_2) are sine and cosine functions of random angles θ_1 and θ_2, while the extracted term k^2 in front of the transformation matrix is

$$k^2 = \frac{1}{(c_1 s_2 - c_2 s_1)^2} = \frac{1}{\sin^2(\theta_1 - \theta_2)}. \tag{5.18}$$

According to (5.16), the terms of the covariance matrix of an underlying 2D process $\mathbf{C}$ can be calculated from the inverse of the analytically derived optimal transformation matrix $E[\mathbf{M}]^{-1}$ and the estimated terms of the covariance matrix of intersections $\mathbf{C_s}$ as

$$\begin{bmatrix} C_{11} \\ C_{12} \\ C_{22} \end{bmatrix} = E[\mathbf{M}]^{-1} \begin{bmatrix} C_{s11} \\ C_{s12} \\ C_{s22} \end{bmatrix}. \tag{5.19}$$

Expectation of the transformation matrix $\mathbf{M}$ depends on statistics of LOR angles. Throughout this derivation, we assumed that the LOR angle is statistically independent of the point source position from which the LOR originates. In subsection 5.2.3 the expectation of the matrix $\mathbf{M}$ will be derived for chosen characteristic joint distribution of θ_1 and θ_2.

5.2.2 Underlying PDF reconstruction from intersections of LORs

To describe and compare the reconstruction of the unknown underlying 2D function from the original and virtual LORs, two experiments were made. In both experiments, the discrete FBP algorithm is used. In the first baseline experiment, the underlying process is reconstructed from two sinograms: the sinogram obtained from counting the LORs at random angles that pass through a large number of process samples, and the ideal sinogram obtained from geometrical projections of the ideal continuous underlying process. The second experiment is an example of extra-low-dose 2D PET imaging. It compares the reconstruction from sinograms obtained from a small number of process samples by utilizing virtual LORs.

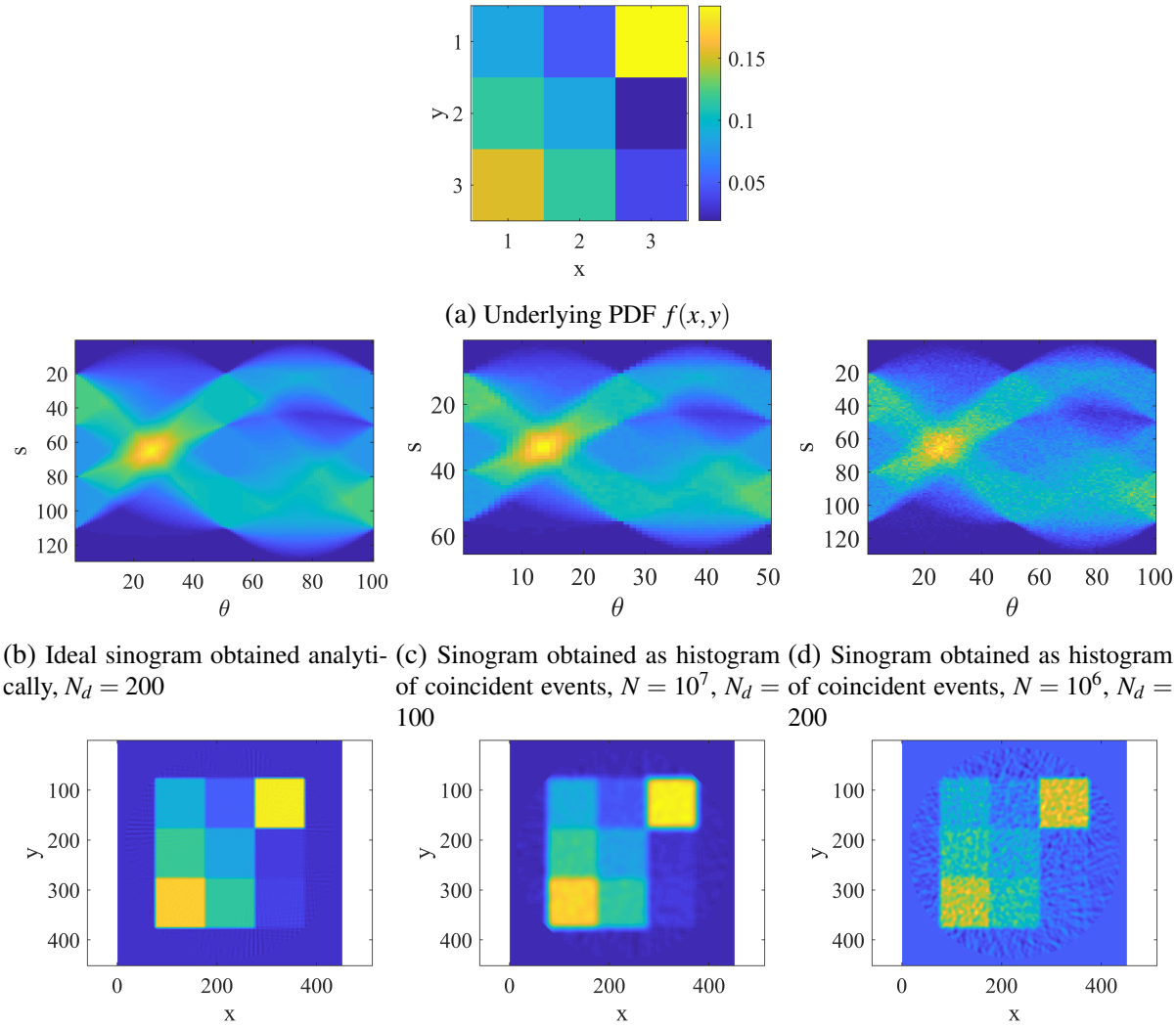
Baseline FBP reconstruction from a large number of LORs

The first experiment was performed to establish the baseline using the classical FBP reconstruction algorithm on a synthetic 2D process with uniform distribution over 3×3 patches. Fig. 5.4a shows the used 2D PDF $f(x,y)$ which is a set of 9 uniform distributions with probabilities within rectangular regions

$$\mathbf{p} = [0.0864, 0.1152, 0.1537, 0.0480, 0.0864, 0.1152, 0.1921, 0.0192, 0.0384].$$

The goal of the experiment was to identify the required number of coincident events necessary to reconstruct the given process from the sinogram. The efficiency of the discrete FBP method in the reconstruction of an underlying process consisting of a large number of samples was analyzed.

The process was synthesized by drawing 10^7 and 10^6 random samples from this distribution. Each random sample generated one LOR at a random angle from the range $\theta \in [-\pi/2, \pi/2]$. It is assumed that the PET detectors are on a circle of radius $R_f = 3\sqrt{2}/2$ to ensure the whole process fits within the FOV. Depending on the experiment, the number of detectors in the ring was selected as $N_d = 100$ or $N_d = 200$. The discretization per variables s and θ was determined using (5.2)-(5.3). The LOR with the positive segment on the y axis of the Cartesian coordinate system has a positive sign of distance s and vice versa. The sinogram was constructed in two ways: analytically and statistically. The analytical sinogram was made by calculating the projection matrix from the cross-sectional areas of the original process grid (Fig. 5.4a) and the projection grid. The projection grid rotates with an angular increment of $\Delta\theta$, and its pixel dimensions are equal to $\Delta s \times 2R_f$. The results of intersections of these two grids is a projection matrix that is ideal for the given parameters and determines the contributions of individual process regions in each projection determined with the θ_j and s_k. The dimension of the projection matrix is $(2N_s + 1) \times N_p \times N_\theta$ where N_p is the number of specified rectangular regions with uniform probabilities (9 in our case). The ideal sinogram for the given parameters is shown in Fig. 5.4b and was obtained by matrix multiplication of the projection matrix with the specified process probability column over all angles and center displacements. The statistically obtained sinograms for different numbers of process samples shown in Figs. 5.4c and 5.4d are obtained by counting the LORs belonging to each angle index where the indices go from 0 to N_θ and to each distance index where the indices go from $-N_s$ to N_s. Reconstructions were made from the obtained sinograms which were filtered with an apodized ramp filter. The larger number of samples yields the reconstructed image closer to the one reconstructed from the ideal sinogram. It is further shown that increasing the number of sensors N_d, and thus the number of segments on the s axis of the sinogram, only improves the reconstruction if there are enough events. Such a sinogram and the reconstructed image are more grainy, i.e., high-frequency noise components are present. In such cases, the reconstruction can be improved by introducing a low-pass filter,



(a) Underlying PDF $f(x,y)$

(b) Ideal sinogram obtained analytically, $N_d = 200$

(c) Sinogram obtained as histogram of coincident events, $N = 10^7$, $N_d = 100$

(d) Sinogram obtained as histogram of coincident events, $N = 10^6$, $N_d = 200$

(e) Reconstructed image from the analytically obtained sinogram

(f) Reconstructed image from sinogram (c) filtered with the apodized ramp filter

(g) Reconstructed image from sinogram (d) filtered with the apodized ramp filter

Figure 5.4: Comparison of the reconstructed underlying PDF using the analytically and statistically obtained sinogram from different numbers of LORs N and detectors N_d. $w_{cutoff} = \pi$.

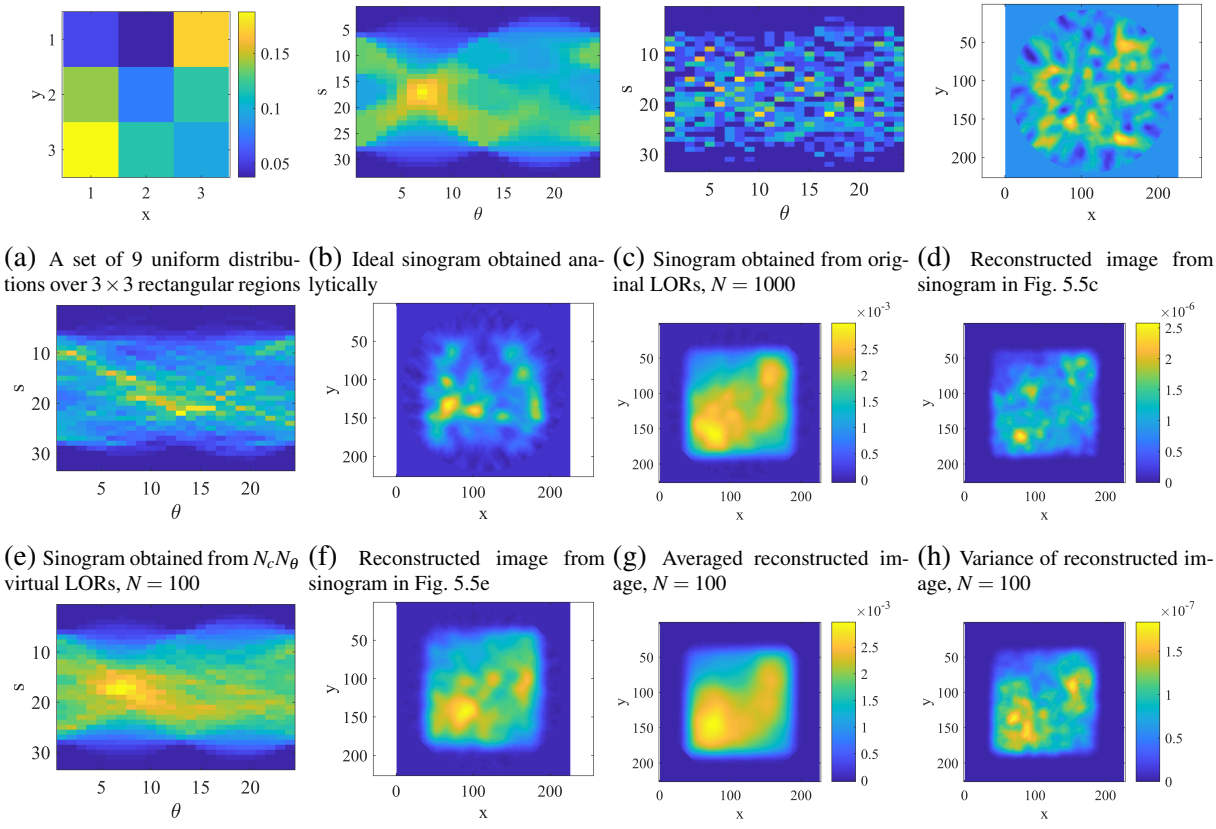
i.e., by applying an apodized ramp filter, which reconstructs a more blurred image. This experiment shows that high-resolution image reconstruction can only be achieved from a very large number of events, thus requiring the injection of a high dose of radiotracer into the patient's body or a very long integration.

FBP reconstruction from intersections of LORs

The same procedure as in the baseline experiment was performed but on a much smaller number of LORs (1000). Vector of probabilities within rectangular regions was in this experiment

$$\mathbf{p} = [0.0539, 0.1409, 0.1876, 0.0374, 0.0767, 0.1170, 0.1716, 0.1189, 0.0960].$$

The given 2D distribution shown in Fig. 5.5a was reconstructed from a very small number of samples using the virtual LORs at places of intersections of the original LORs. The number of process samples was 1000 and 100. The corresponding sparse sinogram and the FBP reconstruction from original LORs that pass through the process samples are shown in Figs. 5.5c and 5.5d. The figures show that the sinogram is very sparse, and the conventional reconstruc-



(a) A set of 9 uniform distributions over 3×3 rectangular regions (b) Ideal sinogram obtained analytically (c) Sinogram obtained from original LORs, $N = 1000$ (d) Reconstructed image from sinogram in Fig. 5.5c

(e) Sinogram obtained from $N_c N_\theta$ virtual LORs, $N = 100$ (f) Reconstructed image from sinogram in Fig. 5.5e (g) Averaged reconstructed image, $N = 100$ (h) Variance of reconstructed image, $N = 100$

(i) Sinogram obtained from $N_c N_\theta$ virtual LORs, $N = 1000$ (j) Reconstructed image from sinogram in Fig. 5.5i (k) Averaged reconstructed image, $N = 1000$ (l) Variance of reconstructed image, $N = 1000$

Figure 5.5: 2D underlying PDF reconstruction using the virtual LORs for a different number of process samples. Each intersection of original LORs was used as a point source to generate $N_c N_\theta$ virtual LORs over the predetermined angles. The Monte Carlo simulation with 50 trials was performed and the averaged reconstructed images are shown. $N_d = 48$.

tion does not resemble the original process for such a small and insufficient number of samples. However, the result is different if the virtual LORs are used. Based on the theoretically derived statistical relationship in subsection 5.2.1 between intersections and process samples, the intersections were used to generate virtual LORs. Only intersections within the square area of the process were used in the reconstruction. Instead of passing a single virtual LOR through each intersection with a random angle, we have generated N_θ virtual LORs at angles $\theta = k\Delta\theta$, where $k = 0, .., N_\theta$ (i.e., by adding the whole sinusoid in the sinogram for each valid intersection). The sinograms of those virtual LORs are shown in Figs. 5.5e and 5.5i for $N = 100$ and $N = 1000$ original process samples, respectively. It can be seen that the sinogram in Fig. 5.5i obtained from original samples using the virtual LORs is more similar to the ideal analytically obtained

sinogram in Fig. 5.5b than the sinogram in Fig. 5.5c obtained from the same number of samples but using the conventional methods and 1000 original LORs. Reconstructed images from such sinograms are shown in Figs. 5.5f and 5.5j. It can be seen that the reconstruction is better if more process samples are used, and consequently, more intersections and more virtual LORs, although such a reconstruction is far from ideal. The Monte Carlo simulation with 50 trials was performed. The mean and the variance of the reconstructed images for 50 trials were obtained. It is shown that by averaging several reconstructed images, an averaged image increasingly resembles the initial process, even though, in each iteration, only 100 or 1000 random samples were drawn from the given distribution. Also, by increasing the number of process samples from 100 to 1000, the variance of reconstruction decreases by 10 times. However, the averaged image is biased concerning the initial process because the covariance matrices of the process and intersections are connected via expectation of the transformation matrix, i.e., although they are related, they are not identical. This experiment shows that with a much lower dose of radiotracers, a lot of information about the initial process can be obtained, and the temporal resolution of the reconstruction can be increased.

5.2.3 Expectation of the transformation matrix

In this subsection, we will derive the expectation of the transformation matrix $\mathbf{M}$ which relates terms of the covariance matrix of the underlying process $\mathbf{C}$ and terms of the covariance matrix of intersections of LORs $\mathbf{C}_s$. This expectation of tranformation matrix $\mathbf{M}$ is found by computing the expectation of each of its terms $M_{i,j}(\theta_1,\theta_2)\forall i,j \in [1,3]$, as follows

$$E(M_{i,j}(\theta_1,\theta_2)) = \int_{\theta_1}\int_{\theta_2} M_{i,j}(\theta_1,\theta_2)\cdot f_{\theta_1,\theta_2}(\theta_1,\theta_2)d\theta_1 d\theta_2, \tag{5.20}$$

where $f_{\theta_1,\theta_2}(\theta_1,\theta_2)$ corresponds to the joint PDF of random angles θ_1 and θ_2.

Expressions (5.17) and (5.18) show that the multiplicative factor k in the transformation matrix $\mathbf{M}$ diverges to infinity when the difference of angles between two intersected LORs are $\theta_1-\theta_2=0$ or $\theta_1-\theta_2=\pm\pi$. Consequently, the expectation of the transformation matrix $\mathbf{M}$ and the covariance matrix of intersections of LORs $\mathbf{C}_s$ also diverge to infinity. On the other hand, $k=1$ for $\theta_1-\theta_2=\pm\pi/2$. To circumvent the divergence to infinity of the expectation of transformation matrix $\mathbf{M}$ and, consequently, of the covariance matrix of intersections $\mathbf{C}_s$, which occurs if the whole range of input angles is allowed $\theta_{1,2}\in[-\pi/2,\pi/2]$, the regularization of input angles has to be introduced.

Intersecting LORs angle pairs, which are in the range $\pm a$ around the perpendicular LORs, are declared as valid LOR angle pairs as Fig 5.6 shows. Red bands represent the areas of valid angle combinations. In that way, the multiplicative factor k is limited, as well as the values of the covariance matrix of intersections $\mathbf{C}_s$. This criterion can be used to limit the intersections only

to those with sufficient perpendicularity between LOR pairs, which limits the spatial expansion of intersections.

If θ_1 and θ_2 are independent random variables uniformly distributed within the range $[-\pi/2, \pi/2]$, and if the regularization band of width a around the line where $\theta_1 \perp \theta_2$ is introduced, the θ_2 can be expressed as a function of random variables θ_1 and a_x, where a_x is a random variable uniformly distributed within $\pm a$,

$$\theta_2 = \theta_1 + \pi/2 + a_x. \tag{5.21}$$

The multiplicative factor k^2 then becomes a function of a_x according to

$$k^2 = \frac{1}{\sin^2(\theta_1 - \theta_2)} = \frac{1}{\sin^2(\theta_1 - \theta_1 - \pi/2 - a_x)} = \frac{1}{\sin^2(-\pi/2 - a_x)^2} = \frac{1}{\cos^2(a_x)} \tag{5.22}$$

The joint PDF of valid angles can be derived from Fig. 5.6 from the total area enclosed within red bands of valid angle pairs $P_{valid} = 2\pi a$ as $f_{\theta_1,a_x}(\theta_1, a_x) = \frac{1}{2\pi a}$. By substituting $\theta_2 = \theta_1 +$

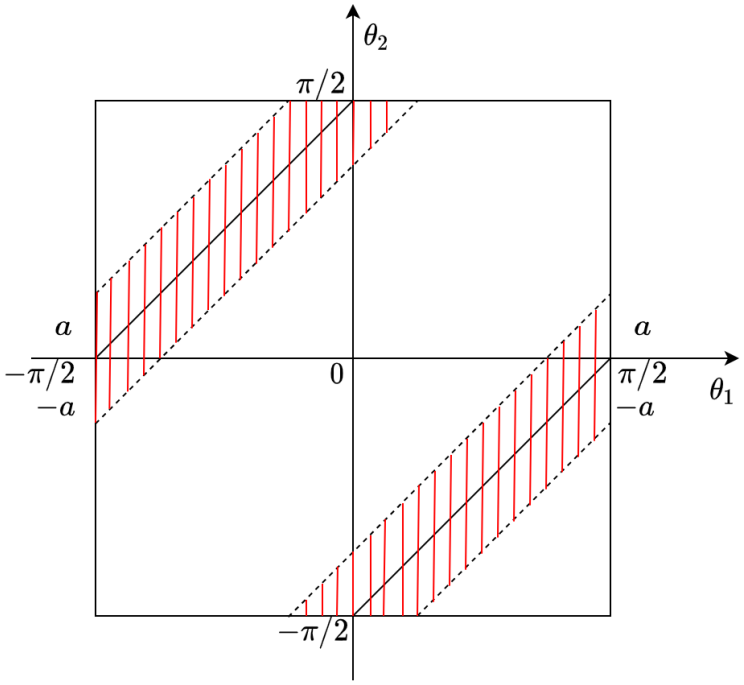


Figure 5.6: The band regularization (the red active area)

$\pi/2 + a_x$, the expectations of the terms in the transformation matrix $\mathbf{M}$ are

$$\begin{aligned} E(M_{1,1}) &= \frac{1}{2\pi a}\int_{-a}^{a}\int_{-\pi/2}^{\pi/2} \frac{c_1^2 s_2^2 + c_2^2 s_1^2}{\cos^2 a_x} d\theta_1 da_x \\ &= \frac{1}{2\pi a}\int_{-a}^{a}\left(\frac{\pi}{2} + \frac{\pi}{4\cos^2 a_x}\right) da_x \\ &= \frac{1}{2} + \frac{\tan a}{4a}, \end{aligned} \tag{5.23}$$

$$E(M_{1,2}) = \frac{1}{2\pi a}\int_{-a}^{a}\int_{-\pi/2}^{\pi/2}\frac{-2(s_2c_1^2c_2+s_1c_1c_2^2)}{\cos^2 a_x}d\theta_1 da_x = 0, \tag{5.24}$$

$$\begin{aligned} E(M_{1,3}) &= \frac{1}{2\pi a}\int_{-a}^{a}\int_{-\pi/2}^{\pi/2}\frac{2c_1^2c_2^2}{\cos^2 a_x}d\theta_1 da_x \\ &= \frac{1}{2\pi a}\int_{-a}^{a}\Big(\frac{3\pi}{8\cos^2 a_x}-\frac{\pi}{4}\Big)da_x \\ &= \frac{3\tan a}{8a}-\frac{1}{4}, \end{aligned} \tag{5.25}$$

$$E(M_{2,1}) = \frac{1}{2\pi a}\int_{-a}^{a}\int_{-\pi/2}^{\pi/2}\frac{(c_2s_1^2s_2+c_1s_1s_2^2)}{\cos^2 a_x}d\theta_1 da_x = 0, \tag{5.26}$$

$$\begin{aligned} E(M_{2,2}) &= \frac{1}{2\pi a}\int_{-a}^{a}\int_{-\pi/2}^{\pi/2}\frac{c_1c_2s_1s_2}{\cos^2 a_x}d\theta_1 da_x \\ &= \frac{1}{2\pi a}\int_{-a}^{a}\Big(\frac{\pi}{8\cos^2(a_x)-\frac{\pi}{4}}\Big)da_x \\ &= \frac{\tan a}{8a}-\frac{1}{4}, \end{aligned} \tag{5.27}$$

$$E(M_{3,1}) = E(f_{1,3}) = \frac{3\tan a}{8a}-\frac{1}{4}. \tag{5.28}$$

Finally, the expectation of the transformation matrix $\mathbf{M}$ for the chosen band regularization a becomes

$$E(\mathbf{M}) = \begin{bmatrix} E(M_{1,1}) & -2E(M_{1,2}) & 2E(M_{1,3}) \\ E(M_{2,1}) & -4E(M_{2,2}) & E(M_{1,2}) \\ 2E(M_{3,1}) & -2E(M_{2,1}) & E(M_{1,1}) \end{bmatrix}. \tag{5.29}$$

By introducing the auxiliary variable $t = \frac{\tan(a)}{a}$, the expectation of the transformation matrix gets the following form

$$E(\mathbf{M}) = \begin{bmatrix} \frac{1}{2}+\frac{t}{4} & 0 & \frac{3t}{4}-\frac{1}{2} \\ 0 & 1-\frac{t}{2} & 0 \\ \frac{3t}{4}-\frac{1}{2} & 0 & \frac{1}{2}+\frac{t}{4} \end{bmatrix}. \tag{5.30}$$

5.2.4 Estimation of process parameters using transformation matrix and intersection statistics

In this subsection, the parameters of the 2D Gaussian process realization, the sample mean and sample covariance, are estimated from the intersections of LORs and the expectation of the transformation matrix $\mathbf{M}$. The given parameters of the 2D Gaussian distribution were

$$[\lambda_1, \lambda_2, \mu_x, \mu_y, \phi] = [3, 1, 4, 7, -\pi/4],$$

where (μ_x, μ_y) is the distribution mean, λ_1 and λ_2 are the major and minor semiaxes widths, while ϕ denotes the rotation angle of the major semiaxis relative to the x-axis of coordinate system. Fig. 5.7 shows the realization of the given 2D Gaussian process with 500 random samples drawn from the distribution with the given parameters. The calculated sample mean

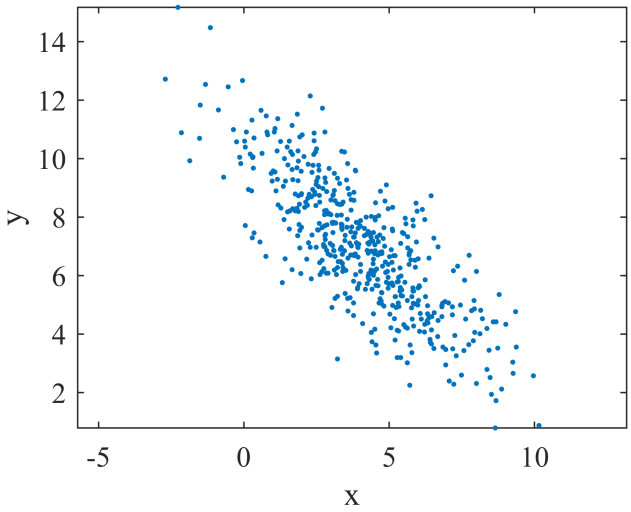


Figure 5.7: Realization of the 2D Gaussian process with the given parameters $[\lambda_1, \lambda_2, \mu_x, \mu_y, \phi] = [3, 1, 4, 7, -\pi/4]$. Number of samples was $N = 500$.

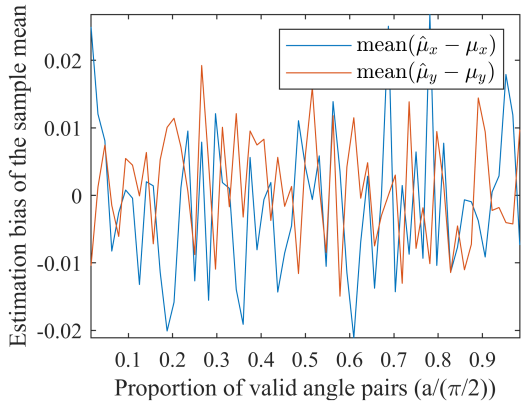
and the sample covariance matrix for this process realization were

$$C = \begin{bmatrix} 4.9255 & -3.9920 \\ -3.9920 & 5.0545 \end{bmatrix}, \mu = [4.000, 7.000], \tag{5.31}$$

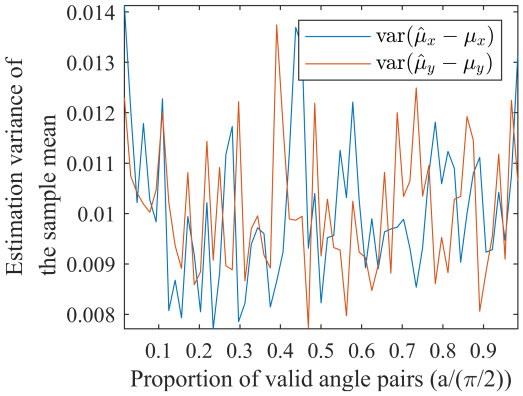
and they represented the ground truth of the process parameters. Monte Carlo simulation with 100 trials was performed for different values of regularization factor a. The theoretical range of regularization band a is from $a = 0$, where only the intersections obtained by perpendicular LORs are valid, to the $a = \pi/2$, where all intersections are declared valid. The values of regularization band a varied in this experiment from a minimum of $a = \pm\pi/128$, which corresponds to $k_{max} = 1.0003$ with 1.56% of valid pairs of angles, to a maximum of $a = \pm 63/128\pi$ which corresponds to $k_{max} = 40.7$ with 98.4% of valid angle pairs. Small values of the regularization band a caused only a small percentage of angle pairs to be valid but also low dispersion of intersections from the underlying process. Consequently, the covariance matrix was prevented from spreading to infinity. Conversely, larger absolute values of the regularization band $\pm a$ yielded more valid angle pairs and intersections. Also, the consequence was a larger dispersion of intersections from the underlying process since some distant intersections were declared regular.

In each trial of Monte Carlo simulation, one LOR at a random angle was generated through each process sample. The LORs' angles were uniformly distributed in the range $[-\pi/2, \pi/2]$.

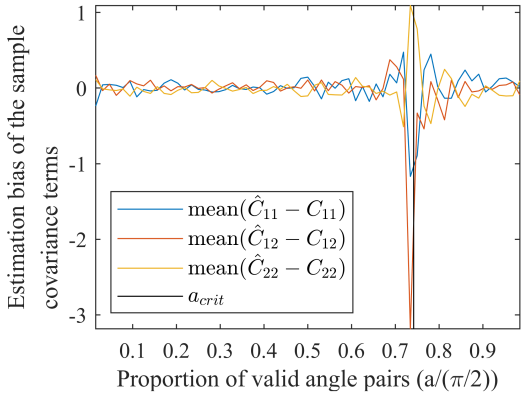
All possible intersections of LORs were found, but only those that satisfied the regularization condition determined by factor a were retained. The sample mean and covariance matrix were calculated for the process of intersections. Also, for each factor a, the expectation of the transformation matrix was calculated by using (5.30), and the inverse of this expectation was applied to the calculated parameters of the process of intersections to estimate the original process parameters (sample mean and the sample covariance matrix). Such an estimator was verified by calculating the bias and variance of the estimated sample mean and covariance matrix and the results are shown in Fig. 5.8.



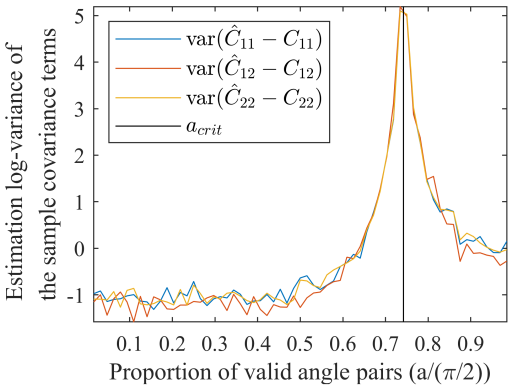
(a) Bias of the estimated intersection mean concerning the sample mean



(b) Variance of the intersection mean estimation concerning the sample mean



(c) Bias of the covariance estimation concerning the sample covariance



(d) Log-variance of the covariance estimation concerning the sample covariance

Figure 5.8: Bias and variance of the mean and covariance estimation concerning the sample mean and sample covariance

Figs. 5.8a shows that the estimation bias of the sample mean has small positive and negative values close to zero, thus indicating that the estimation of the sample mean is insensitive to the selected factor a. Figs. 5.8b shows that the estimation variance of the sample covariance terms is also small and almost constant for all regularization factors a. Figs. 5.8c and 5.8d show that for small values of the regularization band a ($a < \pi/4$), the estimated sample covariance is unbiased with small variances of its terms. For the medium-range regularization bands $a = [\pi/4, 3\pi/4]$,

the estimation variance of the sample covariance terms increases with the maximal value at so-called critical point where $\tan(a) = 2a$. This critical point of the largest error is for $\pm a_{crit} = 66^\circ$ which corresponds to $k_{max} = 2.537$ and is denoted with a vertical line in Figs. 5.8c and 5.8d. For a_{crit}, the covariance matrix of intersections is diagonal with the same widths of semiaxes. In such a diagonal matrix, the information about the rotation of the original process is lost, thus preventing the reconstruction by inversion (the determinant of the transformation matrix $\mathbf{M}$ becomes zero). The estimation variance of the covariance terms decreases for values larger than a_{crit}, but they are still higher than for $a < \pi/4$. Consequently, choosing the regularization band in the range $0 < a < \pi/4$ is recommended. The estimation bias of the covariance terms is close to zero for most values of regularization factor a except for those in the vicinity of a_{crit}.

This experiment also shows that an increase in the number of valid angle pairs and intersections does not reduce the estimation variance, which indicates that intersections do not introduce new information and that the total information about the process is contained in the process samples.

5.2.5 Probability density function of intersections for a pair of point sources

In this subsection, we will derive the probability density function (PDF) of intersections for a pair of samples drawn from some distribution with a finite covariance matrix, i.e., a finite second moment. The total PDF of intersections for the process with N point sources can be found as the sum of PDFs of all $N_c = N(N-1)/2$ possible combinations of unique point source pairs. Let us assume two process samples drawn from some underlying distribution at positions A and B as Fig. 5.9 shows. The origin of the coordinate system O is placed exactly at half of the junction of these two points. Both samples are at the distance c from the origin O. The $\overrightarrow{AB}$ junction is at angle ρ concerning the x-axis of the coordinate system. Two LORs at angles θ_1 and θ_2 pass through the samples A and B and intersect at the point $T = re^{j\theta}$, which is described by polar coordinates, radius r and the rotation angle θ. The LOR angles θ_1 and θ_2 can be expressed in polar coordinates (r, θ) from the geometry of Fig. 5.9 as follows

$$\overrightarrow{OB} + \overrightarrow{BT} = \overrightarrow{OT}, \tag{5.32}$$

$$\overrightarrow{OA} + \overrightarrow{AT} = \overrightarrow{OT}, \tag{5.33}$$

$$\theta_1 = \measuredangle\overrightarrow{BT} = \measuredangle(\overrightarrow{OT} - \overrightarrow{OB}) = \measuredangle(r\exp(j\theta) - c\exp(j\rho)) = \arctan\frac{r\sin\theta - c\sin\rho}{r\cos\theta - c\cos\rho}, \tag{5.34}$$

$$\theta_2 = \measuredangle\overrightarrow{AT} = \measuredangle(\overrightarrow{OT} - \overrightarrow{OA}) = \measuredangle(r\exp(j\theta) + c\exp(j\rho)) = \arctan\frac{r\sin\theta + c\sin\rho}{r\cos\theta + c\cos\rho}. \tag{5.35}$$

The mapping of each intersection point $T = re^{j\theta}$ to a pair of angles (θ_1, θ_2) is bijective if it is parametrized with the rotation angle of the junction of two point sources ρ, and the correspond-

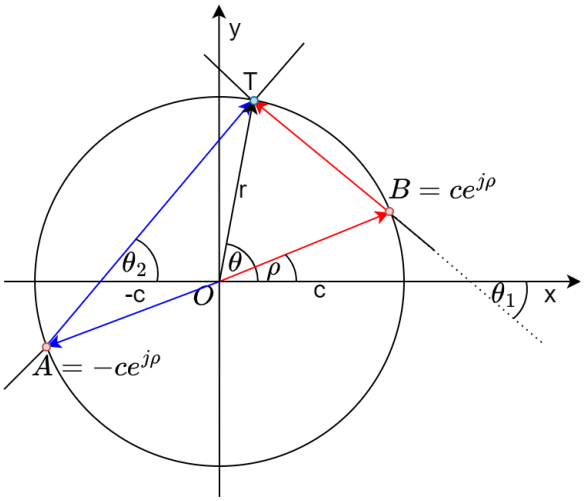


Figure 5.9: Intersection of two LORs

ing halfwidth c as follows

$$\theta_{1,2}(r,\theta,[c,\rho]) = \arctan\frac{r\sin\theta \pm c\sin\rho}{r\cos\theta \pm c\cos\rho}. \tag{5.36}$$

The PDF of intersections for a pair of process samples in polar coordinates (r,θ) can be calculated from the joint PDF of angles of LORs f_{θ_1,θ_2} and the ratio of the surface differentials in both mapping domains, $\theta_1\theta_2$- and $r\theta$-domains. Since the surface differential in domain $\theta_1\theta_2$ is

$$P = \left(\frac{\partial\theta_1}{\partial\theta}\frac{\partial\theta_2}{\partial r} - \frac{\partial\theta_2}{\partial\theta}\frac{\partial\theta_1}{\partial r}\right)\Big|_{r_0,\rho_0}\Delta r\Delta\theta, \tag{5.37}$$

while the surface differential in the domain $r\theta$ is $P_a = r\Delta\theta\Delta r$, the PDF of the intersections for a pair of samples in domain $r\theta$ equals

$$\begin{aligned} f_{r,\theta}(r,\theta,[c,\rho]) &= f_{\theta_1,\theta_2}\frac{P}{P_a} \\ &= f_{\theta_1,\theta_2}\frac{\left(\frac{\partial\theta_1}{\partial\theta}\frac{\partial\theta_2}{\partial r} - \frac{\partial\theta_2}{\partial\theta}\frac{\partial\theta_1}{\partial r}\right)\Big|_{r_0,\rho_0}\Delta r\Delta\theta}{r\Delta r\Delta\theta} \\ &= f_{\theta_1,\theta_2}\frac{2cr|\sin(\rho-\theta)|}{(c^4 - 2\cos(2\rho-2\theta)c^2r^2 + r^4)}. \end{aligned} \tag{5.38}$$

The alternative expression for the intersection PDF for a pair of point sources is

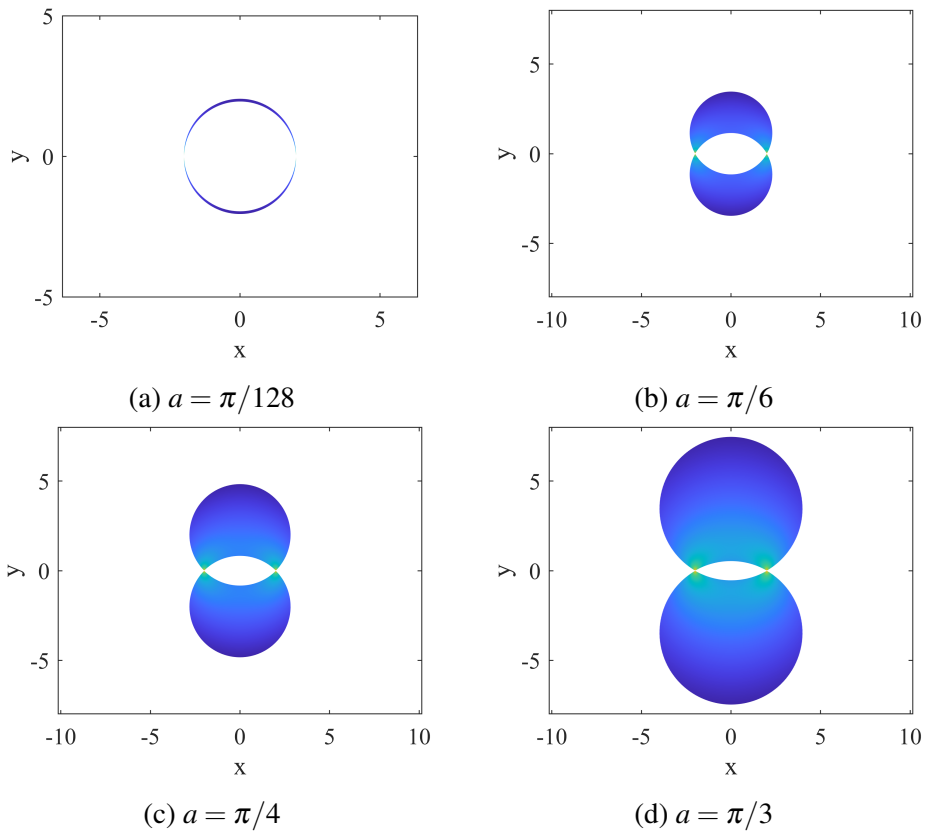
$$f_{r,\theta}(r,\theta,[c,\rho]) = f_{\theta_1,\theta_2}\frac{2cr|\sin(\rho-\theta)|}{(c^2-r^2)^2 + 4r^2c^2\sin^2(\rho-\theta)}, \tag{5.39}$$

which shows that the denominator is always positive or at least equals zero when $r = c$, and $\rho - \theta = 0$ or $\rho - \theta = \pi$. There is also assumed that $r > 0$ and $c > 0$.

In the limit when the radius r tends to c, the PDF of intersections gets the following form

$$\lim_{r\to c} f_{r,\theta}(r,\theta,[c,\rho]) = f_{\theta_1,\theta_2}\frac{|\sin(\rho-\theta)|}{2c^2\sin^2(\rho-\theta)}. \tag{5.40}$$

The joint PDF f_{θ_1,θ_2} of uniformly distributed random angles θ_1 and θ_2 in the range $[-\pi/2, \pi/2]$ is $f_{\theta_1,\theta_2} = \frac{1}{\pi^2}$. In the case of the band regularization, the angles θ_1 and θ_2 satisfy the following expression $\theta_2 = \theta_1 + \pi/2 + a_x$. The area of valid pairs of angles, in that case, is equal to $\frac{a}{\pi/2}\pi^2 = 2\pi a$ with the joint PDF $f_{\theta_1,a_x} = \frac{1}{2\pi a}$. The area of valid angles achieves the maximum value π^2 when all pairs of angles are allowed and $a = \pi/2$, while for $a = 0$, only perpendicular LORs are declared valid. The intersections are in band regularization spatially limited to two "regularization balls" whose common secant is the junction of these two point sources. The radius of these balls is ck_{max}, where $k_{max} = 1/\cos(a)$. In the limit when a tends to zero, $k_{max} = 1$, θ_1 and θ_2 are perpendicular, and these two balls become completely overlapped and turned into a circle of radius c. Contrary, when a tends to $\pi/2$, then k_{max} tends to ∞, so the "regularization balls" have an infinite radius and cover the entire $\mathbb{R}^2$ domain of intersections. Valid intersections with the described band regularization of $\pm a$ around the perpendicular angles are never within the intersection area of the two "regularization balls." The examples of "regularization balls" for different values of regularization factor a are shown in Fig. 5.10.



(a) $a = \pi/128$ (b) $a = \pi/6$

(c) $a = \pi/4$ (d) $a = \pi/3$

Figure 5.10: PDF of intersections for a pair of point sources in the spatial domain for different regularization factors a

5.2.6 Analysis of the behavior of the PDF of intersections for a pair of point sources

The partial fraction decomposition of (5.39) yields the following expression

$$f_{r,\theta}(r,\theta,[c,\rho]) = f_{\theta_1,\theta_2}\frac{|\sin(\rho-\theta)|}{2\cos(\rho-\theta)}\left(\frac{1}{c^2-2cr\cos(\rho-\theta)+r^2}-\frac{1}{c^2+2cr\cos(\rho-\theta)+r^2}\right), \tag{5.41}$$

or alternatively

$$\begin{aligned} f_{r,\theta}(r,\theta,[c,\rho]) &= f_{\theta_1,\theta_2}\frac{|\sin(\rho-\theta)|}{2\cos(\rho-\theta)}\left(\frac{1}{(c-r)^2+4rc\sin^2\frac{(\rho-\theta)}{2}}-\frac{1}{(c+r)^2-4rc\sin^2\frac{(\rho-\theta)}{2}}\right) \\ &= \frac{|\sin(\rho-\theta)|}{2\cos(\rho-\theta)}\left(\frac{1}{d_1}-\frac{1}{d_2}\right). \end{aligned} \tag{5.42}$$

The expression obtained by partial fraction decomposition describes the individual contributions of point sources to the final PDF of intersections. The denominator d_1 in (5.42) is always positive or at least equal to zero for each combination of ρ and θ angles since r and c are positive real numbers. When the intersection hits the position of the first point source, $r = c$ and $\rho = \theta$, and d_1 becomes zero. In that case, $1/d_1$ tends to infinity, as well as the whole PDF in the vicinity of the first point source. The second denominator d_2 is also non-negative and achieves the smallest value when $\sin^2((\rho-\theta)/2) = 1$. In that case, $d_2 = (c^2 - 2rc + r^2) = (c-r)^2$, which can not be less than zero. Thus, the second denominator achieves zero value only when $c = r$ and $\rho - \theta = \pm\pi$.

Since d_1 and d_2 are non-negative, the difference of two fractions $(1/d_1 - 1/d_2)$ is positive when $1/d_1 > 1/d_2$, i.e., $d_1 < d_2$. According to (5.41), this inequality is satisfied if $2\cos(\rho - \theta) > 0$. Expression $2\cos(\rho-\theta)$ is the same as the denominator of the factor $\frac{|\sin(\rho-\theta)|}{2\cos(\rho-\theta)}$ which multiplies the difference $(1/d_1 - 1/d_2)$. As the numerator of this multiplicative factor is also always non-negative $|\sin(\rho-\theta)|$, it follows that the overall PDF will also be non-negative for each combination of parameters $(r > 0, c > 0, \rho, \theta)$. Namely, the denominator of the factor $\frac{|\sin(\rho-\theta)|}{(2\cos(\rho-\theta))}$ as well as the difference $(1/d_1 - 1/d_2)$ are either at the same time positive or at the same time negative. Therefore, the overall expression for the intersection PDF will always be positive, as expected, since the intersection PDF is derived from the ratio of two surface differentials, which must be necessarily positive or at least zero.

In addition, the horizontal and vertical cuts of the intersection PDF from (5.38) are derived to determine the function decay rate in the directions of the principal axes of the local PDF coordinate system. To calculate these cuts, the auxiliary variable $\delta = \rho - \theta$ is introduced. A

vertical cut is the limit of (5.38) when δ tends to $\pm\pi/2$ as follows

$$\lim_{(\rho-\theta)\to\pm\pi/2} f_{r,\theta}(r,\theta,[c,\rho]) = \frac{2rc}{(r^2+c^2)^2}. \tag{5.43}$$

For $r >> c$, (5.43) tends to $f_{r,\theta}(r,\pi/2) = \frac{2c}{r^3}$, which describes the cubic decay of the PDF of intersections with the radial distance r from the center of the junction.

The horizontal cut is in the direction of the junction of two point sources, which corresponds to $\theta = \rho$. The horizontal cut has the right side $\rho - \theta = 0$ and left side for $\rho - \theta = \pm\pi$. To calculate the horizontal cut, the $|\sin(\rho-\theta)|$ is removed from the numerator in (5.38) since it surely goes to 0, but the rest of the PDF potentially goes to ∞ if the left or right point source is hit. The limit of the rest of the PDF expression with the removed sine term when one or another point sources ($r \to c$) is approached equals

$$\lim_{(\rho-\theta)\to 0} \frac{f_{r,\theta}(r,\theta,[c,\rho])}{\sin(\delta)} = \frac{1}{2(c^2-2cr+r^2)} - \frac{1}{2(c^2+2cr+r^2)}. \tag{5.44}$$

To determine the dominant term in the denominator of the (5.44), the inverse of the (5.44) is expanded in the Taylor series, yielding the following expression

$$\frac{(c-r)^4}{2c^2} + \frac{(c-r)^5}{2c^3} + 2(c-r)^2. \tag{5.45}$$

For $(c-r) << c$, the last term in the Taylor expansion is dominant, so the rest of the PDF tends with the power of two to infinity when we approach the point $r = c$ according to the expression $1/(2(c-r)^2)$. For all other values of r, the PDF at the horizontal junction is equal to zero, because $\sin(\rho-\theta) = 0$, and the rest of the PDF expression is finite.

Marginal PDF of intersections for a pair of point sources

The marginal PDF of intersections for a pair of point sources as a function of radius r can be obtained by integrating the joint PDF of intersections $f(r,\theta,[c,\rho])$ from (5.41) over the variable $\delta = \rho - \theta$ in the range from 0 to $\pi/2$, where all terms $(\sin(\delta), \cos(\delta), 1/d_1 - 1/d_2)$ are positive and multiply the obtained integral by 4 to cover the whole range of angles. This integral for non-regularized case when $a = \pi/2$ is

$$\begin{aligned} f_r(r) &= f_{\theta_1,\theta_2} 4r \int_{\delta=0}^{\pi/2} \frac{|\sin(\delta)|}{(2\cos(\delta))} \Big(\frac{1}{c^2 - 2cr\cos(\delta) + r^2} - \frac{1}{c^2 + 2cr\cos(\delta) + r^2} \Big) d\delta \\ &= f_{\theta_1,\theta_2} 4r \frac{\log \frac{(c+r)^2}{(c-r)^2}}{2c^2 + 2r^2} \end{aligned} \tag{5.46}$$

For non-regularized case, the joint PDF $f_{\theta_1,\theta_2} = \frac{1}{\pi^2}$, while for the case of band regularization $f_{\theta_1,\theta_2} = \frac{1}{2\pi a}$. In the case of band regularization, the marginal PDF is obtained by integration from the initial angle $\delta = \psi$ to the final angle $\delta = \pi/2$ as follows

$$f_r(r) = f_{\theta_1,\theta_2} 4r \int_{\delta=\psi}^{\pi/2} \frac{|\sin(\delta)|}{(2\cos(\delta))} \left(\frac{1}{c^2 - 2cr\cos(\delta) + r^2} - \frac{1}{c^2 + 2cr\cos(\delta) + r^2} \right) d\delta = \frac{r \log\left(\frac{A+1}{A-1}\right)}{\pi a (c^2 + r^2)}, \tag{5.47}$$

where

$$A = \tan(a)(c^2 + r^2) \sqrt{-\frac{1}{-4\tan^2(a)c^2r^2 + c^4 - 2c^2r^2 + r^4}}.$$

The initial angle ψ is analytically derived in the next subsection.

5.2.7 Regularization of angles of LORs

The band regularization in the domain of angles $\theta_1\theta_2$ is mapped in the spatial xy domain into the disjunctive union of two intersecting circles which corresponds to their union without their intersection, as shown in Fig. 5.11 as two blue regions. In the boundary case with the regularization factor $a = 0$, only the perpendicular LORs are declared valid, and their intersections are mapped to a circle of the radius c in the spatial domain. For the maximal value of the regularization band $a = \pi/2$ all angles of LORs are valid and the intersection domain is infinite.

It is necessary to find the intersection of the small circle of an arbitrary radius r and large circle from Fig. 5.11 to find the initial integration angle ψ for the calculation of marginal PDF of intersections $f(r)$ from (5.47). Firstly, the lengths b, e, and d are calculated from the geometry

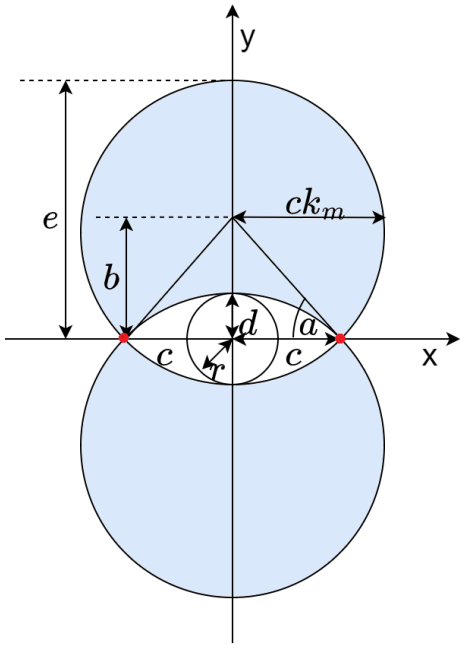


Figure 5.11: The intersection PDF shape with the regularization bandwidth a

of Fig. 5.11 as follows

$$c^2 + b^2 = c^2 k_{max}^2 = \frac{c^2}{\cos^2(a)}, \tag{5.48}$$

$$b^2 = c^2\Big(\frac{1}{\cos^2(a)} - 1\Big) = c^2\Big(\frac{1-\cos^2(a)}{\cos^2(a)}\Big) = c^2\tan^2(a), \tag{5.49}$$

$$b = c|\tan(a)|, a \to 0, b \to 0, a \to \pm\pi/2, b \to \infty, \tag{5.50}$$

$$d = ck_{max} - b = \frac{c}{\cos(a)} - c\frac{\sin(a)}{\cos(a)} = c\Big(\frac{1-\sin(a)}{\cos(a)}\Big), a \to 0, d \to c, a \to \pm\pi/2, d \to 0, \tag{5.51}$$

$$e = ck_{max} + b = \frac{c}{\cos(a)} + \frac{c\sin(a)}{\cos(a)} = c\Big(\frac{1+\sin(a)}{\cos(a)}\Big). \tag{5.52}$$

If $r < d$ or $r > e$, there is no intersection between a small circle centered at the origin of the local coordinate system and a large circle centered at $(0,b)$. The equations of the large and small circles are, respectively

$$x^2 + (y-b)^2 = \frac{c^2}{\cos^2(a)}, \tag{5.53}$$

$$x^2 + y^2 = r^2, \tag{5.54}$$

with the intersection points that satisfy the following equations

$$y = \frac{r^2 - c^2}{2\tan(a)c}, \tag{5.55}$$

$$x^2 = r^2 - \frac{(c^2 - r^2)^2}{4\tan(a)^2 c^2}. \tag{5.56}$$

The condition $x^2 > 0$, i.e., $r^4 - (2c^2 + 4b^2)r^2 + c^4 < 0$, should be full-filled to find a real intersection between two circles. If $a_t = \tan(a)$, a solution of this inequality exists for the following range of radius r

$$d^2 < r^2 < e^2, \tag{5.57}$$

$$c^2(a_t - \sqrt{a_t^2 + 1})^2 < r^2 < c^2(a_t + \sqrt{a_t^2 + 1})^2. \tag{5.58}$$

Finally, the corresponding initial angle where the large and small circles intersect for a given regularization factor $a_t = \tan(a)$ and for the distance from the center of the junction r is

$$\psi = \arctan\Big(\frac{y}{x}\Big) = \arctan\left(\frac{|c^2 - r^2|}{2a_t c\sqrt{r^2 - \frac{(c^2-r^2)^2}{4a_t c^2}}}\right). \tag{5.59}$$

The term under the root must be positive, which is ensured only if $d < r < e$.

5.2.8 Vector formulation of the PDF of intersections for a pair of samples in global coordinate system

In this section, the vector formulation of the intersection PDF for a pair of point sources is derived. In Fig. 5.12, r and θ are the polar coordinates of intersection in the local $x'y'$ coordinate system with the origin at the center of the junction of these two points, c represents the junction halfwidth, and ρ is the rotation angle of the junction concerning the x' axis.

The intersection PDF for a pair of point sources in the local $x'y'$ coordinate system is described in (5.38) and in (5.39). To transform this PDF to global coordinate system xy, the point sources can be considered as complex numbers ($\mathbf{r}_1 = r_1 \exp(j\theta_1)$ and $\mathbf{r}_2 = r_2 \exp(j\theta_2)$). Also, a complex vector of the intersection point in the global coordinate system is $\mathbf{z}$. According to Fig. 5.12, the vector from the origin of the local coordinate system to the second point source can be found as $\mathbf{c} = \frac{\mathbf{r}_2 - \mathbf{r}_1}{2}$, while the vector from the origin of the global to the origin of the local coordinate system is $\mathbf{r}_0 = \frac{\mathbf{r}_1 + \mathbf{r}_2}{2}$. The vector of intersection in the local coordinate system $\vec{r}$ is in the global coordinate system $\mathbf{z} = \mathbf{r} + \mathbf{r}_0$. The intersection PDF after the partial decomposition

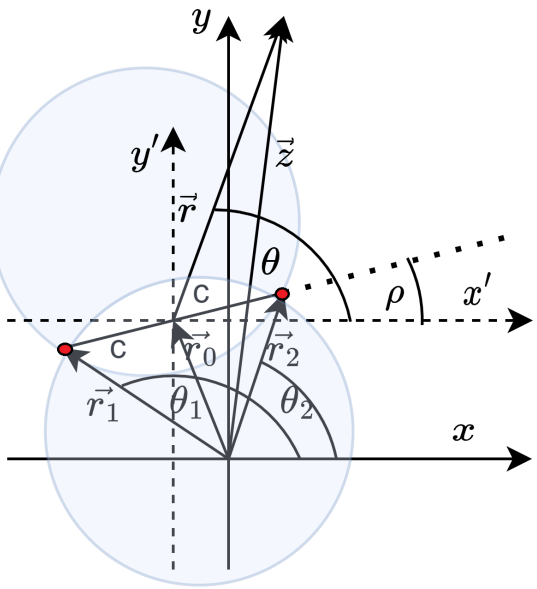


Figure 5.12: Vector formulation of the intersection PDF

in vector formulation in the local coordinate system can be expressed as

$$f(\mathbf{r},[\mathbf{c}]) = f_{\theta_1,\theta_2} \frac{|\mathbf{c}\mathbf{r}^* - \mathbf{c}^*\mathbf{r}|}{2(\mathbf{c}\mathbf{r}^* + \mathbf{c}^*\mathbf{r})} \left[\frac{1}{(\mathbf{r}-\mathbf{c})(\mathbf{r}-\mathbf{c})^*} - \frac{1}{(\mathbf{r}+\mathbf{c})(\mathbf{r}+\mathbf{c})^*} \right], \tag{5.60}$$

since

$$\begin{aligned}\mathbf{c}\mathbf{r}^* - \mathbf{c}^*\mathbf{r} &= [c\exp(j\rho)r\exp(-j\theta) - c\exp(-j\rho)r\exp(j\theta)] \\ &= rc(\exp(j(\rho-\theta)) - \exp(-j(\rho-\theta))) \\ &= 2rcj\frac{\exp(j(\rho-\theta)) - \exp(-j(\rho-\theta))}{2j} \\ &= 2rcj\sin(\rho-\theta),\end{aligned} \tag{5.61}$$

$$\begin{aligned}\mathbf{c}\mathbf{r}^* + \mathbf{c}^*\mathbf{r} &= rc(\exp(j(\rho-\theta)) + \exp(-j(\rho-\theta))) \\ &= 2rc\cos(\rho-\theta),\end{aligned} \tag{5.62}$$

$$\begin{aligned}(\mathbf{r}-\mathbf{c})(\mathbf{r}-\mathbf{c})^* &= [r\exp(j\theta) - c\exp(j\rho)][r\exp(-j\theta) - c\exp(-j\rho)] \\ &= r^2 - rc\exp(j(\rho-\theta)) - rc\exp(-j(\rho-\theta)) + c^2 \\ &= r^2 - 2rc\frac{\exp(j(\rho-\theta)) + \exp(-j(\rho-\theta))}{2} + c^2 \\ &= r^2 - 2rc\cos(\rho-\theta) + c^2,\end{aligned} \tag{5.63}$$

$$\begin{aligned}(\mathbf{r}+\mathbf{c})(\mathbf{r}+\mathbf{c})^* &= [r\exp(j\theta) + c\exp(j\rho)][r\exp(-j\theta) + c\exp(-j\rho)] \\ &= r^2 + c^2 + 2rc\frac{\exp(j(\rho-\theta)) + \exp(-j(\rho-\theta))}{2} \\ &= r^2 + c^2 + 2rc\cos(\rho-\theta).\end{aligned} \tag{5.64}$$

Finally, the PDF of intersection, parametrized with the pair of point sources ($\mathbf{r}_1$,$\mathbf{r}_2$), in the global coordinate system for the nonregularized case has the following form

$$f(\mathbf{z},[\mathbf{r}_1,\mathbf{r}_2]) = \frac{|\mathbf{c}\mathbf{r}^* - \mathbf{c}^*\mathbf{r}|}{2(\mathbf{c}\mathbf{r}^* + \mathbf{c}^*\mathbf{r})}\left(\frac{1}{(\mathbf{r}-\mathbf{c})(\mathbf{r}-\mathbf{c})^*} - \frac{1}{(\mathbf{r}+\mathbf{c})(\mathbf{r}+\mathbf{c})^*}\right) \tag{5.65}$$

$$= -\frac{|t_2 - t_1|}{2(t_2+t_1)}\left(\frac{1}{\left(\frac{|\mathbf{r}_1|^2}{\mathbf{r}_1} - \frac{|\mathbf{z}|^2}{\mathbf{z}}\right)(\mathbf{r}_1-\mathbf{z})} - \frac{1}{\left(\frac{|\mathbf{r}_2|^2}{\mathbf{r}_2} - \frac{|\mathbf{z}|^2}{\mathbf{z}}\right)(\mathbf{r}_2-\mathbf{z})}\right), \tag{5.66}$$

where

$$t_1 = \left(\frac{\mathbf{r}_1}{2} - \frac{\mathbf{r}_2}{2}\right)\left(\frac{|\mathbf{r}_1|^2}{2\mathbf{r}_1} + \frac{|\mathbf{r}_2|^2}{2\mathbf{r}_2} - \frac{|\mathbf{z}|^2}{\mathbf{z}}\right), \tag{5.67}$$

$$t_2 = \left(\frac{|\mathbf{r}_1|^2}{2\mathbf{r}_1} - \frac{|\mathbf{r}_2|^2}{2\mathbf{r}_2}\right)\left(\frac{\mathbf{r}_1}{2} + \frac{\mathbf{r}_2}{2} - \mathbf{z}\right). \tag{5.68}$$

The regularization within the $\pm a$ around the perpendicular LORs in the vector formulation is achieved by the following condition

$$\frac{|(\mathbf{z}-\mathbf{r}_1)(\mathbf{z}-\mathbf{r}_2)^* + (\mathbf{z}-\mathbf{r}_1)^*(\mathbf{z}-\mathbf{r}_2)|}{|(\mathbf{z}-\mathbf{r}_1)(\mathbf{z}-\mathbf{r}_2)^* - (\mathbf{z}-\mathbf{r}_1)^*(\mathbf{z}-\mathbf{r}_2)|} < |\tan(a)|. \tag{5.69}$$

5.2.9 Alternative expression for the PDF of intersections for a pair of point sources in the global coordinate system

To find the general expression for the intersection PDF of two point sources in the global coordinate system, we can start from Fig. 5.12, equations (5.34), (5.35), and the following vectors

$$\mathbf{r}_1 = r_{1r} + r_{1im}i, \tag{5.70}$$

$$\mathbf{r}_2 = r_{2r} + r_{2im}i, \tag{5.71}$$

$$\mathbf{z} = x + yi, \tag{5.72}$$

where i denotes the imaginary unit, (r_{1r}, r_{1im}) and (r_{2r}, r_{2im}) are real and imaginary parts of the first and the second point source respectively, while x and y represent the intersection coordinates in the global coordinate system. Then equations (5.34) and (5.35) become functions of random variables x and y as follows

$$\theta_1 = \arctan\frac{r_{1im} - y}{r_{1r} - x}, \tag{5.73}$$

$$\theta_2 = \arctan\frac{r_{2im} - y}{r_{2r} - x}. \tag{5.74}$$

The joint PDF of intersections f_{xy} can be found by using the method of transformations

$$f_{xy}(x, y, [r_{1r}, r_{1im}, r_{2r}, r_{2im}]) = f_{\theta_1\theta_2}|J| = \frac{1}{\pi^2}|J|, \tag{5.75}$$

where

$$J = \det\begin{bmatrix} \frac{\partial\theta_1}{\partial x} & \frac{\partial\theta_1}{\partial y} \\ \frac{\partial\theta_2}{\partial x} & \frac{\partial\theta_2}{\partial y} \end{bmatrix}. \tag{5.76}$$

The final expression for the PDF of intersections for a pair of point sources $f_{xy}(x, y)$, parametrized with real and imaginary parts of point sources in the global coordinate system, has the following form

$$f_{xy}(x, y, [r_{1r}, r_{1im}, r_{2r}, r_{2im}]) = f_{\theta_1\theta_2}\frac{|r_{1im}r_{2r} - r_{2im}r_{1r} - r_{1im}x + r_{2im}x + r_{1r}y - r_{2r}y|}{((r_{1im} - y)^2 + (r_{1r} - x)^2)((r_{2im} - y)^2 + (r_{2r} - x)^2)}. \tag{5.77}$$

5.2.10 Total PDF of intersections

The total PDF of intersections for all pairs of point sources drawn from some underlying 2D distribution can be numerically found as a sum of PDFs of intersections for all pairs of process samples. For the process with N samples, there are $N(N-1)/2$ combinations of unique sample pairs. Fig. 5.13 shows the example of the total PDF of intersections for three process samples.

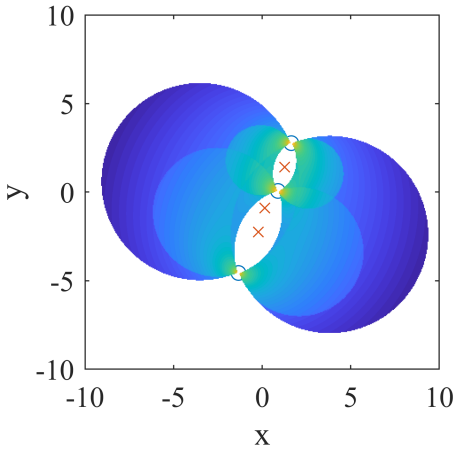
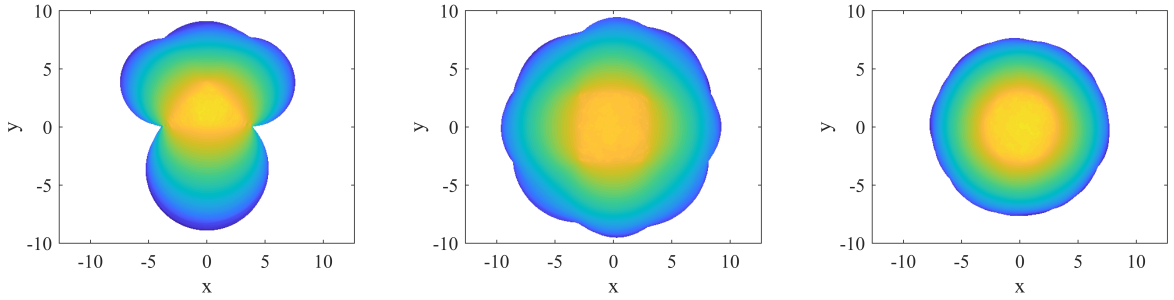


Figure 5.13: The total PDF of intersections with regularization for the process with three samples

An analytical expression for a total PDF of intersections is the expectation of the PDF of intersections as follows

$$E(f_{xy}(x,y,[r_{1r},r_{1im},r_{2r},r_{2im}])) = \int_{r_{2im}}\int_{r_{2r}}\int_{r_{1im}}\int_{r_{1r}} f_{xy}(x,y,[r_{1r},r_{1im},r_{2r},r_{2im}])f_{r_{1r}}f_{r_{1im}}f_{r_{2r}}f_{r_{2im}}dr_{1r}dr_{1im}dr_{2r}dr_{2im}, \tag{5.78}$$

where (x,y) represents the intersection position and $(r_{1r},r_{1im},r_{2r},r_{2im})$ are real and imaginary parts of process samples, which are random variables with their PDFs $f_{r_{1r}}, f_{r_{1im}}, f_{r_{2r}}$, and $f_{r_{2im}}$. The analytical expression for this expectation is very hard to obtain since it is necessary to calculate the quadruple integral over the real and imaginary positions of the point sources which are random variables from some distribution. Even for assumed uniform distributions of point source positions within a circle or square, the quadruple integral was not calculated and the closed form of this expectation was not obtained. Fig. 5.14 shows the numerically obtained total PDFs of intersections for uniformly distributed processes within triangle, square, and circle with 500 samples drawn from such distributions. The total PDFs were obtained as sums of individual



(a) Uniformly distributed process over a triangle (b) Uniformly distributed process over a square (c) Uniformly distributed process over a circle

Figure 5.14: Distributions of intersections for uniformly distributed processes with 500 samples. Regularization was $a = \pi/4$.

contributions of PDFs for all unique sample pairs. Also, the regularization term was used, which was $a = \pi/4$.

5.2.11 PDF of intersections for a pair of point sources in the local sinogram domain

The intersection PDF for a pair of point sources in the local and global sinogram domains is derived in this subsection. As shown in Fig. 5.15, the PDF of intersections in the local sinogram domain can be obtained by rotating the junction of two point sources in the local spatial domain by changing its rotation angle ρ in the range $[-\pi/2, \pi/2]$ and by integrating over y. At the same time, the local coordinate system in the spatial domain xy remains static, and its variable x is simultaneously the variable of signed distance in the local sinogram domain, while the junction rotation angle ρ denotes the angle in the local sinogram.

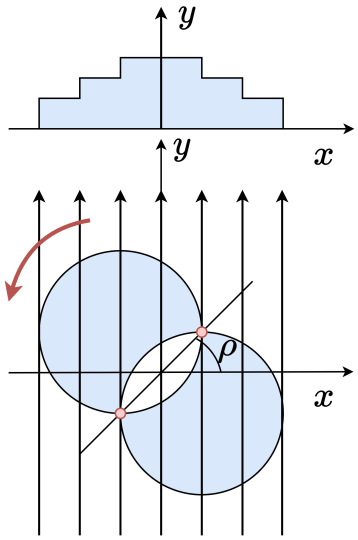


Figure 5.15: Calculation of the PDF of intersections in the local sinogram $x\rho$ domain by integrating (5.82) over y for all angles $\rho = [-\pi/2, \pi/2]$

Let us start from the expression for the PDF of intersections for a pair of point sources (5.38) in the local spatial xy domain, but in the vector form and without regularization

$$f(\mathbf{r}, [\mathbf{c}]) = \frac{1}{\pi^2} \frac{|\mathbf{r}^*\mathbf{c} - \mathbf{r}\mathbf{c}^*|}{|\mathbf{r} - \mathbf{c}|^2 |\mathbf{r} + \mathbf{c}|^2}, \tag{5.79}$$

where $\mathbf{r}$ is a vector of an arbitrary point in that local coordinate system xy and $\mathbf{c}$ is the vector of one of two point sources as follows

$$\mathbf{c} = c_r + c_{im} i, \tag{5.80}$$

$$\mathbf{r} = x + yi, \tag{5.81}$$

where $c_r = |\mathbf{c}| \cos\rho, c_{im} = |\mathbf{c}| \sin\rho$. After including (5.80) and (5.81) in (5.79), the expression for the PDF of intersections for a pair of point sources in the local spatial xy domain becomes a function of spatial x and y variables parametrized with the local coordinates of the point source

c_r and c_{im} as follows

$$f(x,y,[c_r,c_{im}]) = \frac{1}{\pi^2}\frac{2|c_{im}x - c_r y|}{((c_r-x)^2+(c_{im}-y)^2)((c_r+x)^2+(c_{im}+y)^2)}. \tag{5.82}$$

To obtain the PDF of intersections for a pair of point sources in the local $x\rho$ sinogram domain, (5.82) is integrated over y

$$\begin{aligned} I_y &= \int f(x,y,[c_r,c_{im}])dy \\ &= -\frac{c_{im}x - c_r y}{\pi^2|c_{im}x - c_r y|(4c_{im}^2+4x^2)}(2c_{im}\,\text{atan2}(c_r-x,c_{im}-y) \\ &+ x\log\frac{(c_r-x)^2+(c_{im}-y)^2}{(c_r+x)^2+(c_{im}-y)^2} + 2c_{im}\,\text{atan2}(c_r+x,c_{im}+y)). \end{aligned} \tag{5.83}$$

In order to calculate the definite form of this integral, limits were used to omit the exact positions of point sources where this PDF tends to infinity ($y_0 = c_{im}x/c_r$). Finally, the PDF of intersections for a pair of point sources in the local sinogram domain has the following form

$$f(x,\rho,[c_r,c_{im}]) = \lim_{y\to y_0^-} I_y - \lim_{y\to-\infty} I_y + \lim_{y\to\infty} I_y - \lim_{y\to y_0^+} I_y = \frac{x\log\frac{(c_r+x)^2}{(c_r-x)^2} + 4c_{im}\rho}{2\pi^2(c_{im}^2+x^2)}. \tag{5.84}$$

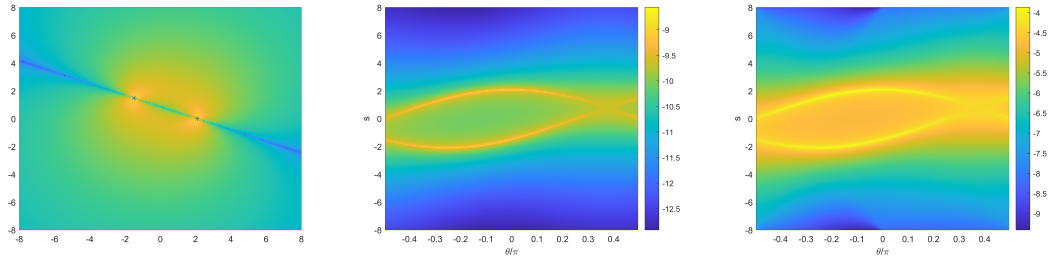
This PDF is transformed to the global $s\theta$ sinogram domain by relating the local and global sinogram coordinates as follows

$$x = s - |\mathbf{r_0}|\cos(\theta - \angle\mathbf{r_0}), \tag{5.85}$$

$$\rho = \theta - \angle\mathbf{c},\ \rho \in [-\pi/2,\pi/2], \tag{5.86}$$

where $\mathbf{c} = (\mathbf{r}_2 - \mathbf{r}_1)/2$, $\mathbf{r}_0 = (\mathbf{r}_1 + \mathbf{r}_2)/2$, $\theta \in [-\pi/2,\pi/2]$, and $\mathbf{r}_1$ and $\mathbf{r}_2$ are the vectors of the first and the second point sources in the global coordinate system in the spatial domain. Fig. 5.16 shows the sinogram obtained analytically by using (5.84) and the sinogram obtained by transforming the PDF of intersections for a pair of point sources from the spatial to the sinogram domain by using *radon* function in Matlab. The results show that the analytically obtained expression for the PDF of intersections for a pair of point sources in the global sinogram domain is correct.

Fig. 5.17 shows the realization of the input process in the spatial domain that consists of two circular 2D Gaussians (5.17a), the total PDF of intersections for that process in the global sinogram domain (5.17b), and the integrals of individual sinograms for each unique pair of point sources (5.17c). The input process consists of the first component with 10 samples drawn from the bivariate normal distribution with the centroid $\mu_1 = [1,3]$ and the semiaxes widths $\lambda_1 = \lambda_2 = 1/3$ and of the second component with 20 samples drawn from the bivariate



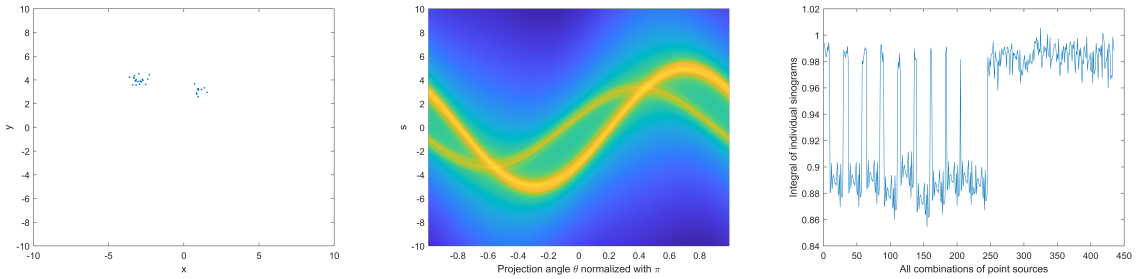
(a) Log of input PDF of intersections for a pair of point sources in the spatial domain

(b) Log of normalized analytical sinogram obtained by using (5.84) of input PDF of intersections

(c) Log of PDF of intersections in sinogram obtained by using *radon* function in Matlab

Figure 5.16: Reconstruction of the input intersection PDF for a pair of sources using transformed coordinates

normal distribution with the centroid $\mu_2 = [-3,4]$ and the same semiaxes widths. The total PDF of intersections in the global sinogram domain was obtained as the sum of the analytically calculated individual sinograms of PDFs of intersections for all unique pairs of point sources. Fig. 5.17c shows numerical integrals of individual sinograms. The numerical integrals close to one indicate the PDFs of intersections for pairs of point sources that originate from the same Gaussian component, while the PDFs of intersections obtained from a combination of point sources from two Gaussian components are dispersed and of smaller integrals.



(a) Input process in the spatial domain

(b) Sinogram of the intersection density for all pairs of point sources

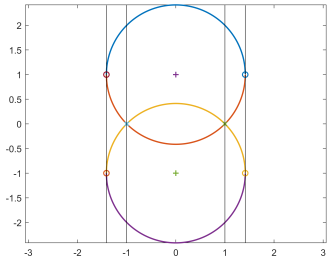
(c) Numerical integral of individual sinograms for each pair of point sources

Figure 5.17: Sinogram of the PDF of intersections for all pairs of point sources obtained by transformations from the global to local coordinates

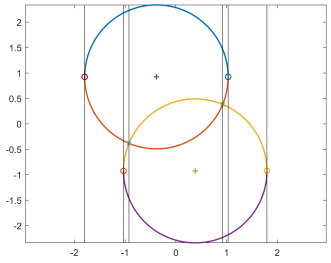
PDF of intersections for a pair of point sources in sinogram for a regularized case

To understand how the regularization of angles of LORs is reflected in the sinogram domain, the calculation of the definite integral of (5.83) for the regularized case is considered for the selected regularization band $a = \pi/4$ and the junction semi-width $c = 1$. Integration by parts of (5.83) includes several cases depending on the relationship between the regularization angle a and the junction angle ρ. Fig. 5.18 shows typical cases that should be considered in integration

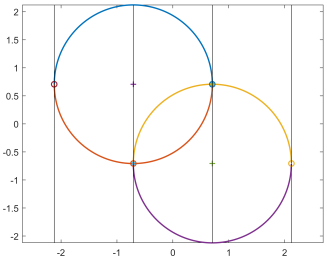
by parts: $\rho = 0$, $\rho < a$, $\rho = a$, $\rho > a$, $\rho = \pi/2$. Integration areas are within the union of two circles without their intersection. The radius of each circle is equal to $r = c/\cos(a)$, while



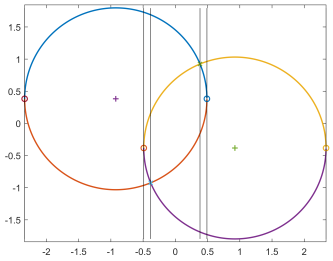
(a) $\rho = 0$, $a = \pi/4$



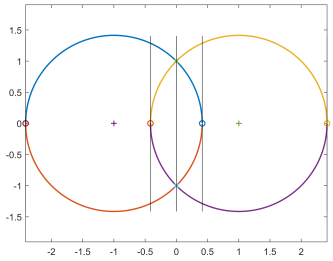
(b) $\rho = \pi/8$, $a = \pi/4$



(c) $\rho = \pi/4$, $a = \pi/4$



(d) $\rho = 3\pi/8$, $a = \pi/4$



(e) $\rho = \pi/2$, $a = \pi/4$

Figure 5.18: Different domains for the regularized case of intersection PDF of two point sources

the distance of the junction of point sources to the center of each circle equals $b = c|\tan a|$. The centers of upper and lower circles are $[x_{0_1}, y_{0_1}] = [-b\sin\rho, b\cos\rho], [x_{0_2}, y_{0_2}] = [-x_{0_1}, -y_{0_1}]$ $\forall\rho \in [-\pi/2, \pi/2]$, while the equations of upper and lower semicircles which represent the integration upper and lower boundaries are:

$$y1_{upper} = y_{0_1} + \sqrt{r^2 - (s - x_{0_1})^2}, \qquad y1_{lower} = y_{0_1} - \sqrt{r^2 - (s - x_{0_1})^2}, \tag{5.87}$$

$$y2_{upper} = y_{0_2} + \sqrt{r^2 - (s - x_{0_2})^2}, \qquad y2_{lower} = y_{0_2} - \sqrt{r^2 - (s - x_{0_2})^2}. \tag{5.88}$$

The sinogram of the PDF of intersections for a pair of point sources is calculated for a regularized case by including the upper and lower boundaries for y in (5.84). Fig. 5.19 represents such an analytically obtained local sinogram of the PDF of intersections for a pair of point sources for $a = \pi/3$ and $c = 1$.

5.2.12 Parameter estimation of rotationally symmetrical 2D Gaussian process from the total PDF of intersections

In this subsection, the rotationally symmetric 2D Gaussian process parameters were estimated from the empirically obtained marginal total PDF of intersections. The underlying PDF was the rotationally symmetric 2D Gaussian distribution with the same semiaxes widths $\sigma_1 = \sigma_2 = \sigma$

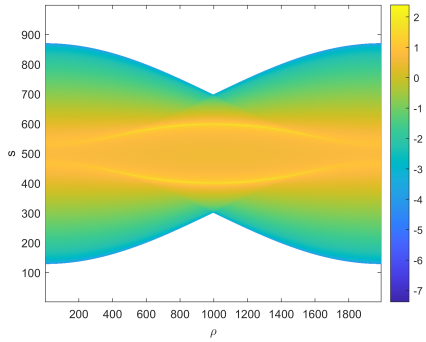


Figure 5.19: Analytically obtained local sinogram of the PDF of intersections for a pair of point sources for regularized case ($a = \pi/3, c = 1$)

and the center at the origin $\mu = [0,0]$. Monte Carlo simulation with 10 trials was performed for each semiaxis width from the set $\sigma = \{0.5,1,2,4\}$. In each trial, $N = 200$ samples were drawn from the given 2D Gaussian distribution. The regularization factor was chosen as $a = \pi/4$. For each unique sample pair, the PDF of intersections $f(x,y,[r_{1r},r_{1im},r_{2r},r_{2im}])$ was calculated over a uniformly sampled grid in polar coordinates (r,θ). The sum of $N(N-1)/2$ synthesized PDFs was calculated and integrated over the angle θ to obtain the mean cut of PDFs of intersections, i.e., the marginal total PDF of intersections as a function of the radius r from the origin of coordinate system.

It was assumed that the mean cut of the total PDF of intersections can be described as the mixture of underlying PDF and the bivariate generalized normal distribution, integrated over $\theta \in [0,\pi]$ as

$$f(r) = \pi(m_1 f_1(r,\sigma_1,\beta) + m_2 f_2(r,\sigma)), \tag{5.89}$$

where m_1 and m_2 are contributions of components to the mixture, $f_1(r,\sigma_1,\beta)$ corresponds to the bivariate generalized normal distribution according to [111] with the β as a shape parameter and σ_1 as a scale parameter, while $f_2(r,\sigma_2)$ corresponds to the rotationally symmetric bivariate normal distribution of the underlying process. Finally, this mixture can be expressed as

$$f(r) = \pi\left(\frac{m_1\beta}{2^{(1/\beta)}\pi\sigma_1^2\Gamma(1/\beta)}\exp\left(\frac{-1}{2}\left(\frac{r}{\sigma_1}\right)^{2\beta}\right) + \frac{m_2}{2\pi\sigma^2}\exp\left(\frac{-1}{2}\left(\frac{r}{\sigma}\right)^2\right)\right), \tag{5.90}$$

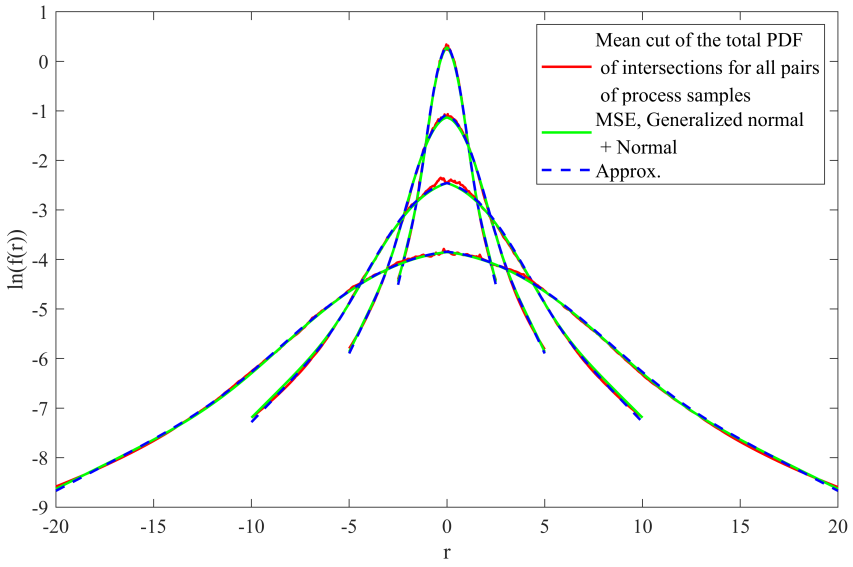
where $\Gamma(\cdot)$ denotes the gamma function.

The given model was fitted to the empirically obtained mean cut of the total PDF of intersections in the least squares sense in the logarithmic domain as

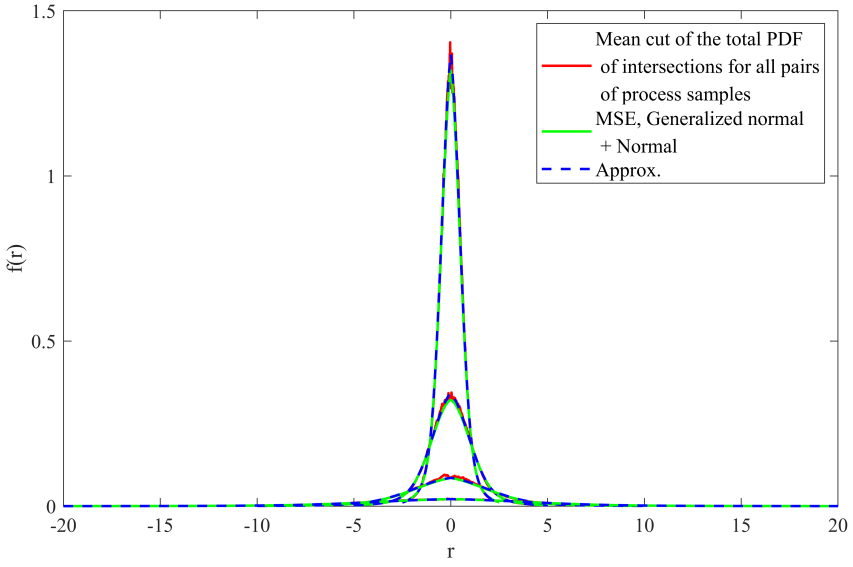
$$D = \sum_r (\ln f(r) - \ln z(r))^2, \tag{5.91}$$

where $f(r)$ corresponds to the model, while $z(r)$ corresponds to the empirical input data.

Experiments have shown that the estimated mixture model parameters depend on the chosen width of the underlying 2D normal distribution. To obtain the general form of this mixture of the underlying and the contamination distributions that fits the empirical data, the estimated model parameters were approximated as $m_1 = m_2 = 0.45$, $\sigma_1 = 0.8\sigma$, and $\beta = 0.55$. The factors $m_1 = m_2 = 0.45$ were selected to obtain the same integral as by integrating the empirical data (≈ 0.9). The estimation results for different widths σ of the underlying normal distribution are shown in Fig. 5.20.



(a) Log of estimated and approximated models that were fitted to empirical data for different widths σ



(b) Estimated (green dashed curve) and approximated models (blue curve) that were fitted to empirical data (red curve) for different widths σ

Figure 5.20: Fitting of a mixture of the generalized bivariate normal distribution (close to Laplace distribution) and the normal bivariate distribution of the underlying PDF to the empirical marginal PDF of intersections $f(r)$, obtained by summing the individual PDFs of intersections for each pair of samples and integrating over all angles θ.

The results show that the estimated model fits the empirical data very well. This means that the blurred PDF of intersections can be described as the sum of the underlying process distribution and a heavy-tailed distribution whose shape parameter is close to that of the Laplace distribution. The mean of estimated parameters for different factors σ of the underlying 2D process are shown in Table. 5.1.

Table 5.1: The mean estimated parameters of a mixture of generalized normal distribution and normal distributions

	$\hat{m}_1$	$\hat{m}_2$	$\hat{\sigma}_1$	$\hat{\beta}$
σ=0.5	0.4408	0.4520	0.4271	0.5562
σ=1	0.4177	0.4746	1.0377	0.6085
σ=2	0.4735	0.4249	1.4424	0.5214
σ=4	0.4343	0.4607	3.1392	0.5365

5.2.13 Expectation maximization algorithm for parameter estimation of the 2D Gaussian process from intersections

This section describes the application of the expectation-maximization (EM) algorithm for parameter estimation of the 2D Gaussian process from the intersections of LORs generated by such a process. The EM algorithm is an iterative two-step method to find the statistical model's parameters that maximize the log-likelihood of observations. In the first step, the EM calculates the expectation of the log-likelihood for the current estimate of the model's parameters. In the second step, the log-likelihood is maximized, i.e., the negative log-likelihood is minimized, to update the model's parameters. These two steps alternate until convergence, at least to the local minimum.

The given parameters of the 2D Gaussian process in the uncorrelated form were: $\theta = [\mu_x, \mu_y, \lambda_1, \lambda_2, \phi] = [1, 1, 2, 1, \pi/4]$, where (μ_x, μ_y) is the mean, (λ_1, λ_2) are the major and minor semiaxes widths, while ϕ denotes the rotation angle with the x-axis of the coordinate system. The number of process samples was $N = 100$. The Monte Carlo simulation with 10 trials was performed. In each trial, $N = 100$ random samples were drawn from the 2D Gaussian process of the given parameters. Each process sample generated one LOR at a random angle in the range $[-\pi/2, \pi/2]$. All possible $N_c = N(N-1)/2$ intersections of LORs were found and only intersections within the rectangular region of $\pm 10\lambda_1$ were declared valid and passed to the EM algorithm.

The initial parameters of the 2D Gaussian process and the valid intersections were passed to the EM algorithm. The initial process parameters were selected as 90% of the actual parameter values. In the objective function of the EM algorithm, 100 random samples were generated

from the 2D Gaussian process with the initial parameters. The total PDF of intersections was calculated as a sum of individual PDFs of intersections for all unique sample pairs. Such a total PDF was evaluated for all observed intersections. The sum of individual PDFs of intersections for all unique sample pairs represented the approximation of expectation of PDF of intersections, whose analytical form remained unknown. If $D = \{x_i, y_i\} \forall i \in [1, N_s]$ was the set of all N_s valid intersections that entered optimization, the probability of intersection point (x_i, y_i) was approximated with the total PDF of intersections as follows

$$p(x_i, y_i) \approx f_{tot} \approx \sum_{j,k} f_{x_i y_i}(x_i, y_i, [r_{jr}, r_{jim}, r_{kr}, r_{kim}]) \tag{5.92}$$

$$= \sum_{j,k} \frac{1}{\pi^2} \frac{|r_{jim} r_{kr} - r_{kim} r_{jr} - r_{jim} x_i + r_{kim} x_i + r_{jr} y_i - r_{kr} y_i|}{((r_{jim} - y_i)^2 + (r_{jr} - x_i)^2)((r_{kim} - y)^2 + (r_{kr} - x_i)^2)}, \tag{5.93}$$

where $r_{jr}, r_{jim}, r_{kr}, r_{kim}$ are real and imaginary parts of N_c unique (j,k) sample pairs of generated process in the objective function. This approximation is closer to the actual expectation of the actual intersection PDF for a larger number of process samples.

The second step was maximization of the log-likelihood function, i.e., the minimization of the negative log-likelihood function, which was calculated as

$$-\mathscr{L}(\theta|D) = -\frac{1}{N_s} \sum_{i=1}^{N_s} \ln p(x_i, y_i), \tag{5.94}$$

where N_s denotes the number of valid intersections. This objective function was minimized to update the 2D Gaussian process parameters, and these two steps were iteratively repeated until the convergence of the EM algorithm. The objective function achieved different scores for the same model parameters due to the random process sample generation in each evaluation of the objective function. In order to reduce the stochastic component in the objective function, the median of 10 negative log-likelihood values for different realizations of the 2D Gaussian process was the final objective function score in each objective function evaluation.

The minimum of such a stohastic objective function can be found through the algorithms for stochastic optimization, which do not use derivatives of the objective function to find an optimal point. In this example, Matlab *patternsearch* function with *classic* algorithm was used for stochastic optimization. The stochastic optimization is performed for the stochastic non-smooth objective functions where the conventional gradient-based optimization method can not be applied.

Since the PDF of intersections for a pair of point sources goes to infinity at the places of process samples, only 90% of the lowest values of the total PDF of intersections were used to calculate the log-likelihood.

The results of parameter estimation of the 2D Gaussian process in 10 trials of Monte Carlo

simulation are shown in Table. 5.2. The number of iterations of the *patternsearch* method, objective function values, and distribution similarity measures (Euclidian, cosine, and Kullback–Leibler divergence) when in the objective function 100 samples were each time drawn from a distribution with current parameters are given in Tables 5.2 and 5.3. The results of the same experiments but with the generation of 1000 random samples in each evaluation of the objective function are given in Tables 5.4 and 5.5. The estimated parameters with the best distribution similarity measures are denoted with green, while the worst estimates are denoted with red.

Table 5.2: Estimated parameters and objective function values (N=100)

	μ_x	μ_y	λ_1	λ_2	ϕ	iter	$-\mathscr{L}(\hat{\theta}\|D)$
θ	1	1	2	1	0.7854		
$\hat{\theta}_1$	1.1539	1.4000	2.0500	0.9000	0.7069	30	4.8447
$\hat{\theta}_2$	1.4000	0.9000	2.5500	0.9000	0.7069	28	5.1452
$\hat{\theta}_3$	0.8999	0.4000	1.5500	1.4000	0.9549	44	4.9707
$\hat{\theta}_4$	1.4000	0.9000	1.9250	1.4000	1.2069	38	5.1121
$\hat{\theta}_5$	0.9000	1.2750	2.3156	1.1500	0.8944	36	5.2699
$\hat{\theta}_6$	0.9000	0.9000	1.8000	1.4000	0.4583	30	5.1203
$\hat{\theta}_7$	0.6500	0.9000	2.1750	0.9156	1.2069	40	5.0461
$\hat{\theta}_8$	1.1969	1.0875	2.5656	1.1500	0.7694	50	5.3317
$\hat{\theta}_9$	0.9000	1.1500	2.1750	1.1500	0.9569	34	5.2567
$\hat{\theta}_{10}$	0.9000	1.4000	2.8000	0.9000	0.9569	28	5.3648

Table 5.3: Similarity measures between given and estimated distributions (N=100)

	KL divergence	Cosine similarity	Euclidian distance
$\hat{\theta}_1$	0.0766	0.9763	0.0045
$\hat{\theta}_2$	0.2329	0.9467	0.0064
$\hat{\theta}_3$	0.4279	0.9191	0.0079
$\hat{\theta}_4$	0.4655	0.9308	0.0074
$\hat{\theta}_5$	0.1445	0.9718	0.0052
$\hat{\theta}_6$	0.2971	0.9597	0.0058
$\hat{\theta}_7$	0.4145	0.9234	0.0078
$\hat{\theta}_8$	0.1497	0.9768	0.0053
$\hat{\theta}_9$	0.1182	0.9798	0.0043
$\hat{\theta}_{10}$	0.4293	0.9271	0.0075

Table 5.4: Estimated parameters and objective function values (N=1000)

	μ_x	μ_y	λ_1	λ_2	ϕ	iter	$-\mathscr{L}(\hat{\theta}\|D)$
θ	1	1	2	1	0.7854		
$\hat{\theta}_1$	1.1500	0.9000	2.0539	1.4000	0.7069	40	5.2829
$\hat{\theta}_2$	1.0406	0.7750	2.0500	0.9000	0.6600	54	4.9757
$\hat{\theta}_3$	1.0240	0.9000	1.3000	2.0250	2.1444	56	5.1499
$\hat{\theta}_4$	1.2749	1.1510	2.1086	1.2750	0.7069	54	5.1464
$\hat{\theta}_5$	1.0250	0.9000	2.5656	0.9625	1.1444	42	5.2802
$\hat{\theta}_6$	0.9000	0.9000	2.3000	1.3844	0.7069	30	5.2982
$\hat{\theta}_7$	0.7750	0.6696	2.5500	1.2750	0.7069	48	5.3522
$\hat{\theta}_8$	1.1656	1.1500	2.3000	1.1500	0.7069	38	5.1403
$\hat{\theta}_9$	0.7750	0.9313	1.3000	1.9000	2.2069	38	5.1495
$\hat{\theta}_{10}$	1.1500	0.9039	2.2961	1.2750	0.9569	48	5.2552

Table 5.5: Similarity measures between given and estimated distributions (N=1000)

	KL divergence	Cosine similarity	Euclidian distance
$\hat{\theta}_1$	0.1652	0.9665	0.0058
$\hat{\theta}_2$	0.0506	0.9806	0.0041
$\hat{\theta}_3$	0.1270	0.9743	0.0049
$\hat{\theta}_4$	0.0944	0.9772	0.0048
$\hat{\theta}_5$	0.3405	0.9360	0.0071
$\hat{\theta}_6$	0.1648	0.9675	0.0061
$\hat{\theta}_7$	0.1730	0.9611	0.0065
$\hat{\theta}_8$	0.0584	0.9860	0.0041
$\hat{\theta}_9$	0.1126	0.9748	0.0047
$\hat{\theta}_{10}$	0.1459	0.9688	0.0057

5.2.14 Estimation of the 2D Gaussian mixture model parameters from virtual LORs at intersections

The PDF of the bivariate Gaussian mixture model (GMM) of two components has the following form

$$f(\mathbf{x}^{(i)}|\mu_k,\Sigma_k) = \sum_{k=1}^{2} w_k \frac{1}{2\pi\sqrt{|\Sigma_k|}} \exp\left(-\frac{1}{2}\left((\mathbf{x}^{(i)}-\mu_k)^T\Sigma_k^{-1}(\mathbf{x}^{(i)}-\mu_k)\right)\right), \tag{5.95}$$

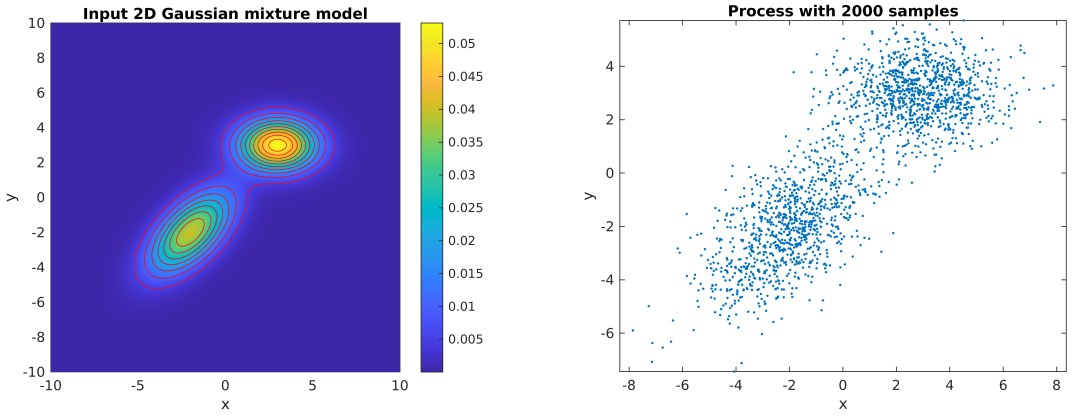
where $\mathbf{x}^{(i)}$ denotes the ith sample position, Σ_k denotes the covariance matrix, μ_k denotes the mean, while w_k is the weight, i.e., the proportion of the kth component in the mixture.

The given parameters of the 2D Gaussian components were:

$$\theta_1 = [\mu_x, \mu_y, \lambda_1, \lambda_2, \phi] = [-2, -2, 2, 1, \pi/4], w_1 = 0.5, \Sigma_1 = \begin{bmatrix} 2.5 & 1.5 \\ 1.5 & 2.5 \end{bmatrix}, \tag{5.96}$$

$$\theta_2 = [\mu_x, \mu_y, \lambda_1, \lambda_2, \phi] = [3, 3, 1.5, 1, 0], w_2 = 0.5, \Sigma_2 = \begin{bmatrix} 2.25 & 0 \\ 0 & 1 \end{bmatrix}. \tag{5.97}$$

2000 random samples were drawn from the GMM distribution with the given parameters. The given GMM probability density function and the generated process samples are shown in Fig. 5.21.



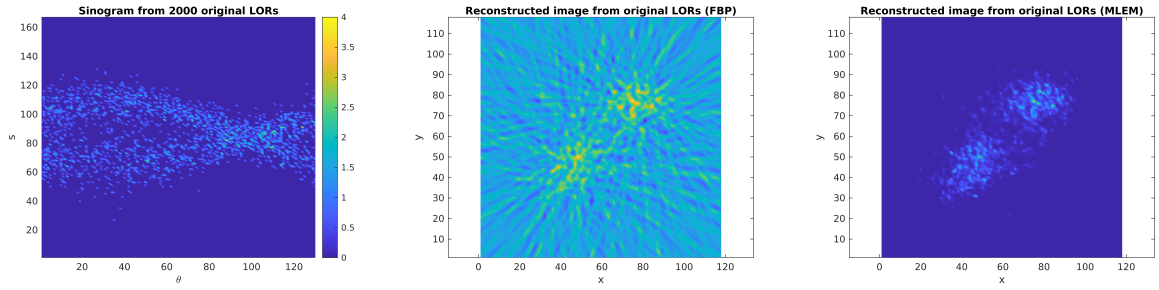
(a) The given GMM model

(b) Generated process with $N = 2000$ samples

Figure 5.21: The input GMM model and generated process samples

Each process sample at position $(x^{(i)}, y^{(i)})$ generated one LOR at random angle from the range $\theta = [-\pi/2, \pi/2]$. The signed distance s of each LOR was calculated as $s = \frac{l}{\sqrt{k^2+1}}$, where $k = \tan(\theta)$ and $l = y^{(i)} - kx^{(i)}$. The sinogram was obtained as a 2D histogram of s and θ parameters of all N original LORs. The discretization of the sinogram angle θ and signed distance s parameters was determined by the number of detectors at the edge of the circular FOV. The number of detectors was $N_d = 260$, the number of discretized angles was $N_\theta = N_d/2$, while the number of discretized signed distance s bins was $2N_s + 1$ where $N_s = N_d/\pi$. The sinogram bin widths were $ds = R_f \pi / N_d$ and $d\theta = 2\pi/N_d$, where $R_f = (\max(x) - \min(x))/2\sqrt{2}$. The filtered back-projection (FBP) and maximum likelihood expectation maximization (MLEM) algorithms [112] were used for the reconstruction. The underlying process was reconstructed by using the FBP algorithm with linear interpolation and Hamming filter and by using the MLEM algorithm with 10 iterations starting from the white image. Fig. 5.22 shows the sinogram of the original LORs and the images reconstructed from that sinogram using the FBP and MLEM

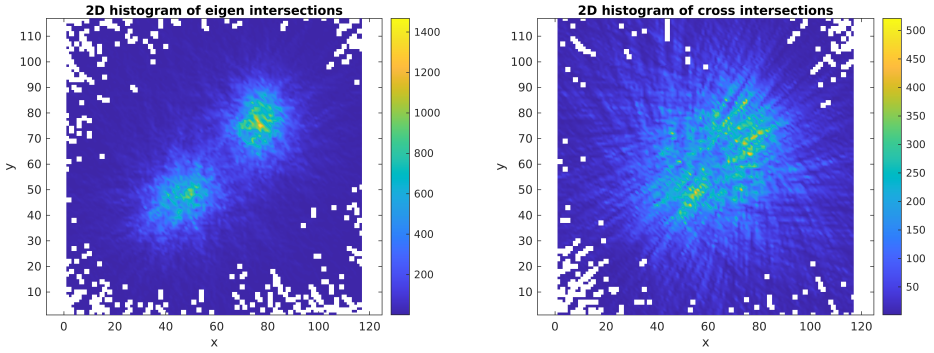
algorithms. All possible $Nc = N(N-1)/2$ intersections of original LORs were found. 2D



(a) Sinogram from original LORs (b) Reconstructed image (FBP) (c) Reconstructed image (MLEM)

Figure 5.22: Sinogram from 2000 original LORs and corresponding reconstructions by using FBP and MLEM algorithms

histograms of eigen intersections which were generated by the LORs of the same 2D Gaussian component, and the cross intersections which were generated by the LORs from different 2D Gaussian components are shown in Fig. 5.23.

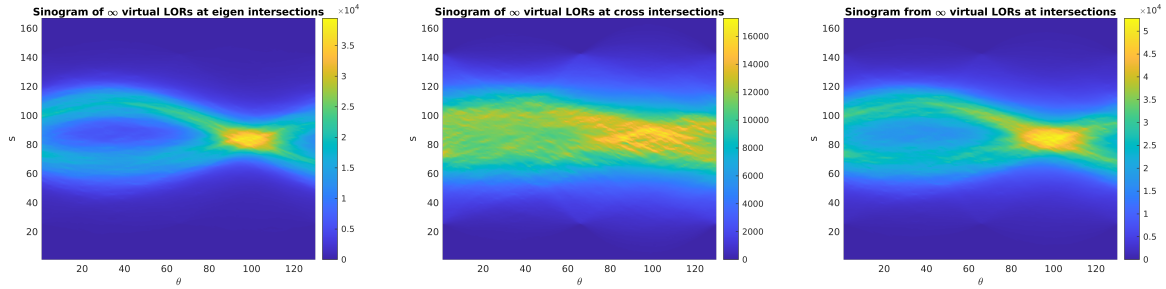


(a) 2D histogram of eigen intersections (b) 2D histogram of cross intersections

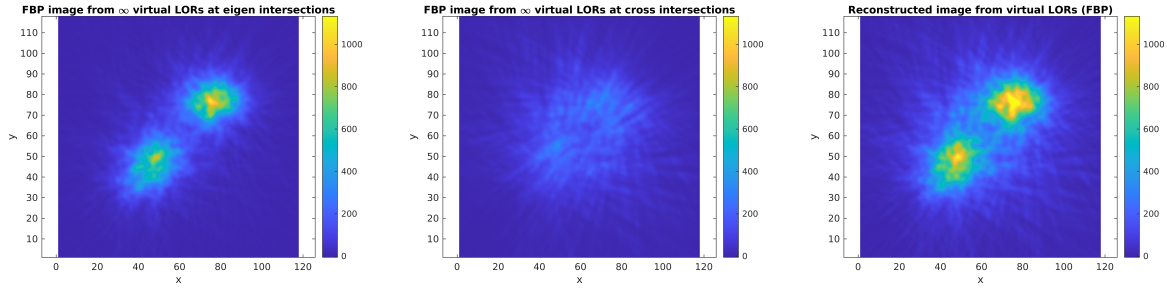
Figure 5.23: 2D histograms of intersections of LORs of the same or cross 2D Gaussian components

Each intersection generated one sinusoid which was recorded in the sinogram. Recording sinusoids for intersections is equivalent to an infinite number of virtual LORs at each intersection point. In this way, the number of events in the sinogram was increased, which enabled the reconstruction of an unknown process from a reduced number of input data. Although the additional input data are artificially invented in this way, it is acceptable because the spatial mean of the intersection is proven to be equal to the spatial mean of the underlying 2D process. Each intersection (xs_i, ys_i) uniquely determined the sinusoid $s = -r\sin(\theta - \phi)$, where $r = \sqrt{xs_i^2 + ys_i^2}$, and $\phi = \text{atan2}(ys_i, xs_i)$. The sinogram of these intersections was obtained by counting the number of sinusoids that intersect a certain sinogram bin. Fig. 5.24 shows sinograms of sinusoids of eigen, cross, and total intersections and corresponding reconstructions using FBP and MLEM algorithms.

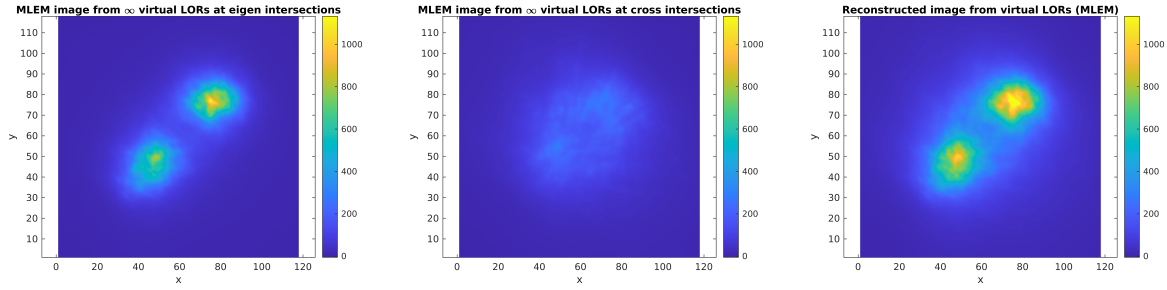
The reconstructed images using the MLEM algorithm were used to fit the GMM model. 10000 random samples were generated from the normalized reconstructed 2D histograms from



(a) Sinogram of sinusoids for eigen intersections (b) Sinogram of sinusoids for cross intersections (c) Sinogram of sinusoids for all intersections



(d) Reconstructed image from (a) using FBP (e) Reconstructed image from (b) using FBP (f) Reconstructed image from (c) using FBP



(g) Reconstructed image from (a) using MLEM (h) Reconstructed image from (b) using MLEM (i) Reconstructed image from (c) using MLEM

Figure 5.24: Sinograms from infinitely many virtual LORs and corresponding reconstructed images by using FBP and MLEM algorithms

sinograms of sinusoids belonging to eigen and total intersections. The cumulative distributions of the reconstructed 2D histograms were calculated. 10000 random samples were generated from the uniform distribution, and then a class was assigned to each sample by summing the number of pixels where the random sample value was larger than the cumulative distribution value. After assigning a class (bin) to each sample, the sample position was randomly selected within the bin.

GMM model was then fitted to the generated random samples, whose distribution followed the reconstructed 2D histogram, by using the MLEM algorithm. The Matlab function $fitgmdist$ was used to estimate the parameters of the GMM model. Due to cross intersections, the reconstructed 2D histogram using total intersections was necessarily blurred, and the estimated GMM parameters did not correspond to the actual parameters. In the first case, the spatial median was

subtracted from the reconstructed 2D histogram and then a 2D GMM model with two components was fitted to the generated random samples whose distribution followed the modified 2D histogram. The estimated parameters were

$$\hat{w}_1 = 0.5134, \hat{\Sigma}_1 = \begin{bmatrix} 4.2994 & -0.1739 \\ -0.1739 & 4.2920 \end{bmatrix}, \hat{\mu}_1 = [-1.2280, -1.1977], \tag{5.98}$$

$$\hat{w}_2 = 0.4866, \hat{\Sigma}_2 = \begin{bmatrix} 3.5816 & -0.6965 \\ -0.6965 & 3.1033 \end{bmatrix}, \hat{\mu}_2 = [3.0365, 3.1924] \tag{5.99}$$

In the second case, the GMM model with three components was fitted to the generated samples which followed the distribution of the originally reconstructed 2D histogram, where the third wide component described the spatial blurring due to cross intersections. The estimated GMM parameters were:

$$\hat{w}_1 = 0.5135, \hat{\Sigma}_1 = \begin{bmatrix} 17.1461 & -0.5129 \\ -0.5129 & 17.6351 \end{bmatrix}, \hat{\mu}_1 = [0.5992, 0.6021], \tag{5.100}$$

$$\hat{w}_2 = 0.2310, \hat{\Sigma}_2 = \begin{bmatrix} 2.6477 & 0.8224 \\ 0.8224 & 3.1927 \end{bmatrix}, \hat{\mu}_2 = [-1.6856, -1.5928], \tag{5.101}$$

$$\hat{w}_3 = 0.2555, \hat{\Sigma}_3 = \begin{bmatrix} 2.5139 & -0.0222 \\ -0.0222 & 1.9073 \end{bmatrix}, \hat{\mu}_3 = [3.3219, 3.3221]. \tag{5.102}$$

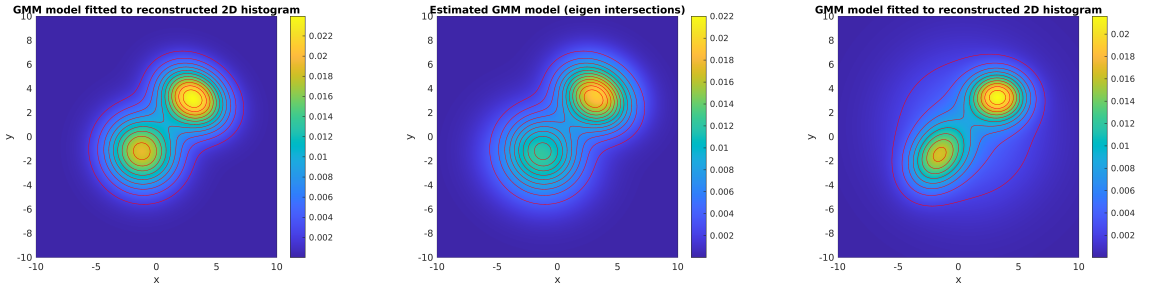
In the third case, the GMM model with 2 components was fitted to the reconstructed image from the sinogram that was obtained from counting the sinusoids that belong to eigen intersections. The estimated parameters were:

$$\hat{w}_1 = 0.5123, \hat{\Sigma}_1 = \begin{bmatrix} 7.1165 & -0.4595 \\ -0.4595 & 6.2084 \end{bmatrix}, \hat{\mu}_1 = [-1.4663, -1.6689], \tag{5.103}$$

$$\hat{w}_2 = 0.4877, \hat{\Sigma}_2 = \begin{bmatrix} 4.4542 & -0.6739 \\ -0.6739 & 3.7758 \end{bmatrix}, \hat{\mu}_2 = [3.3145, 3.5496]. \tag{5.104}$$

The estimated GMM model with three components and GMM model with two components

fitted to the reconstructed 2D histogram from eigen intersections yield the parameters closest to the given model parameters. The estimated GMM models with two and three components are shown in Fig. 5.25.



(a) GMM model with 2 components and removed spatial median

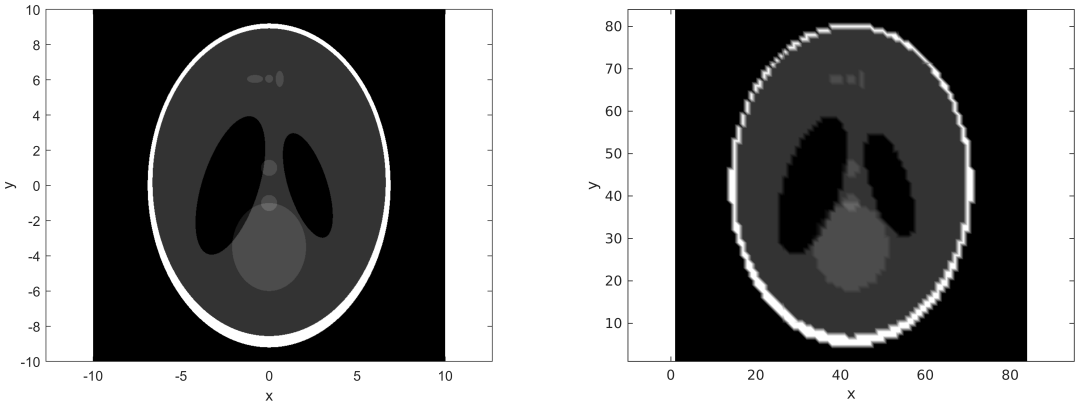
(b) GMM model with 2 components fitted to the reconstructed image of eigen intersections

(c) GMM model with 3 components

Figure 5.25: GMM models fitted to images reconstructed using MLEM algorithm

5.2.15 Head phantom image reconstruction

The modified Shepp-Logan phantom was reconstructed from a small number of original LORs ($N = 1000$) and from infinitely many virtual LORs at the positions of intersections of original LORs by using the previously described FBP and MLEM algorithms. The high and low-resolution versions of the modified Shepp-Logan phantom are shown in Fig. 5.26. The dimensions of the reconstructed image correspond to the low-resolution version of the modified Shepp-Logan phantom which, consequently, represents the ground truth image. The dimensions of the reconstructed image were 84×84, which was determined by the number of detectors at the edge of the FOW $N_d = 188$.

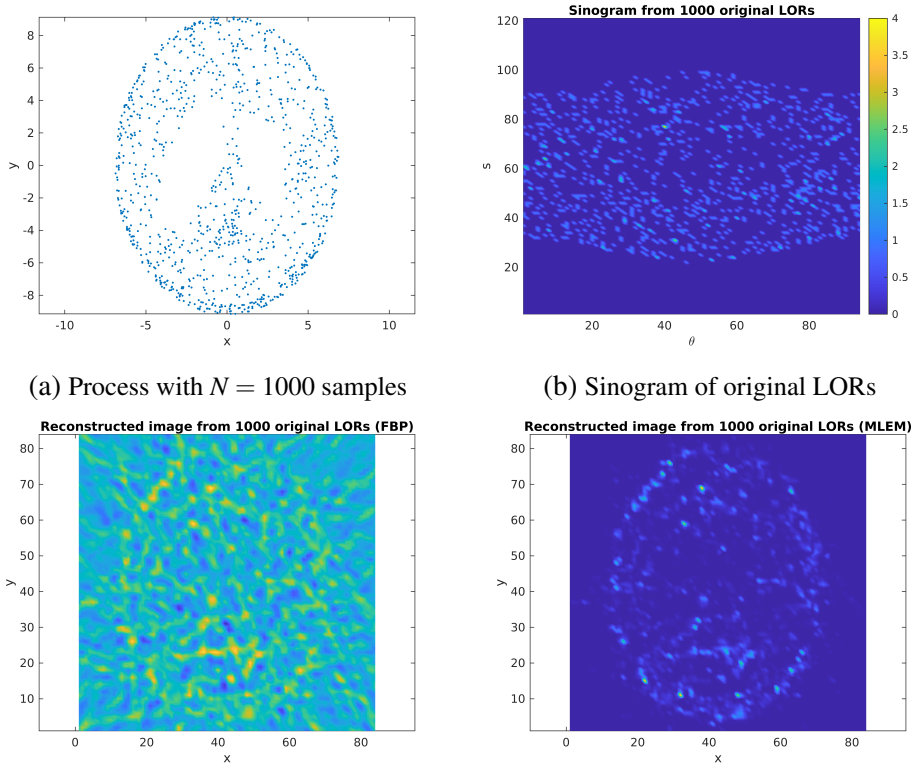


(a) High-resolution version

(b) Low-resolution image

Figure 5.26: Modified Shepp-Logan head phantom

Two versions of the Monte Carlo simulation with 1000 trials were performed. In the first version, in each trial of the Monte Carlo simulation, new 1000 random samples were drawn from the distribution which corresponds to the high-resolution version of the modified Shepp-Logan head phantom. From each process sample, one original LOR at a random angle in the range $[-\pi/2, \pi/2]$ was generated. The process with 1000 random samples and the corresponding sinogram of original LORs are shown in Fig. 5.27. In the second version, the 1000 random samples were drawn from the high-resolution phantom distribution only once, while in each trial of the Monte Carlo simulation, only the angles of LORs were randomly picked.



(a) Process with $N = 1000$ samples

(b) Sinogram of original LORs

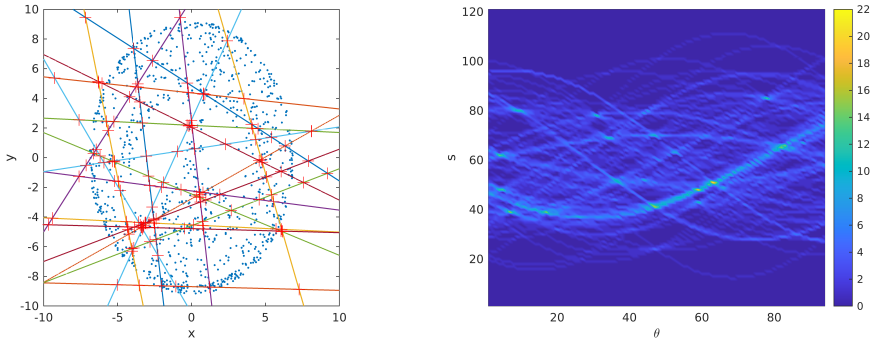
(c) Reconstructed image from 1000 original LORs (FBP)

(d) Reconstructed image from 1000 original LORs (MLEM)

Figure 5.27: Process with $N = 1000$ samples, sinogram of original LORs, and reconstructed images using the FBP and MLEM algorithms

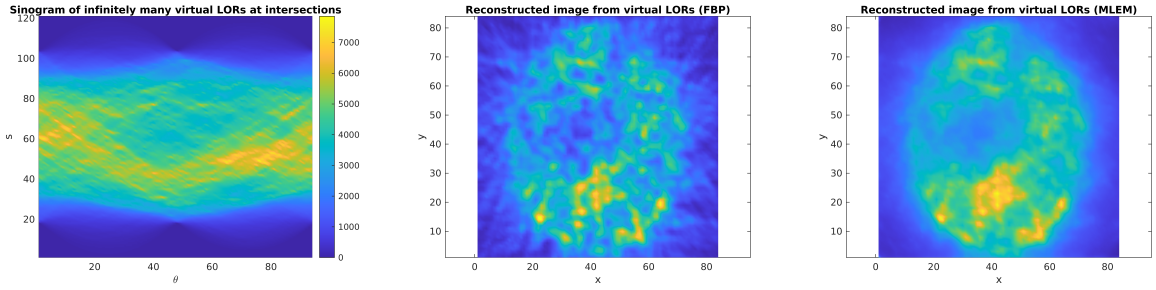
The intersections of original LORs were found. For each intersection, one sinusoid was generated. The sinusoids were recorded in the sinogram by counting the number of sinusoids that intersect a particular sinogram bin. This corresponds to the virtual generation of infinitely many virtual LORs over the whole range of angles at the positions of intersections. The generation of 20 original LORs, their intersections, and sinograms of sinusoids at intersection points and reconstructed images are shown in Fig. 5.28.

The underlying 2D distribution was reconstructed from sinograms of original and virtual LORs by using the FBP and MLEM algorithms. The mean reconstructed images from 10 such trials are shown in Fig. 5.29.



(a) 20 original LORs from process samples, corresponding LORs, and intersections

(b) Sinogram of sinusoids for 128 valid of 190 intersections of 20 LORs



(c) Sinogram of virtual LORs at intersections for all 1000 process samples

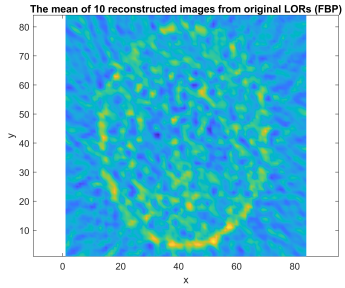
(d) Reconstructed image from virtual LORs (FBP)

(e) Reconstructed image from virtual LORs (MLEM)s

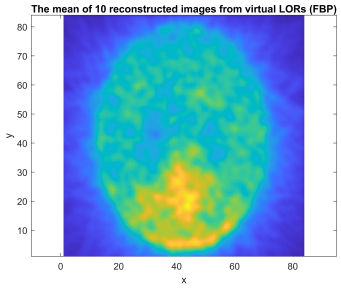
Figure 5.28: Illustration of the proposed method for infinitely many virtual LORs generation at the positions of intersections, sinogram, and reconstructed images from virtual LORs

The averaged reconstructed image from virtual LORs resembles the original phantom distribution more than the grainy averaged reconstructed image from original LORs. The reconstructed image from virtual LORs is blurred due to the difference in covariances of original process samples and intersections, which is especially noticed in the averaged reconstructed images from all 1000 trials which are shown in Fig. 5.30. However, the reconstruction from virtual LORs can be used in real-time applications for fast scanning of the underlying process from a small number of detected coincident events. To reduce the blurring of the averaged reconstructed image, the image of the phantom's skull was reconstructed from the virtual LORs using the FBP and MLEM algorithms and was subtracted from the averaged reconstructed image of the phantom. Results are shown in Fig. 5.30.

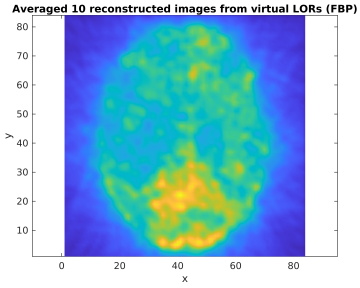
In order to estimate the vignetting function, the uniformly distributed 2D circular process within the FOV with 10,000 samples was generated. From each process sample, one LOR at a random angle was generated. Intersections of generated LORs were calculated and recorded in the 2D histogram. The averaged 2D histogram of intersections of LORs for 10 trials is shown in Fig 5.31d where in each trial new 10000 random LORs were generated from the same process samples. Fig. 5.31b shows the 2D Gaussian vignetting model, which was fitted to the averaged 2D histogram of intersections. The scaled inverse vignetting model was applied to the averaged



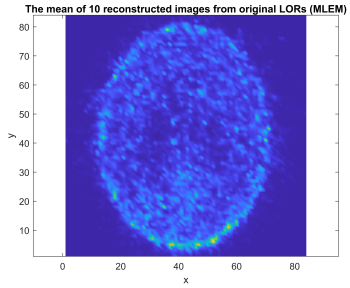
(a) Mean reconstructed image from original LORs (FBP)



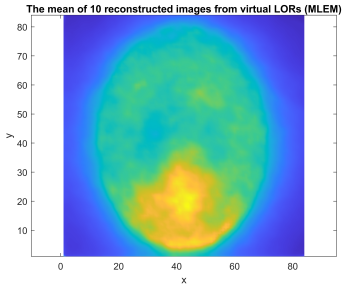
(b) Mean reconstructed image from virtual LORs, new 1000 random samples in each trial (FBP)



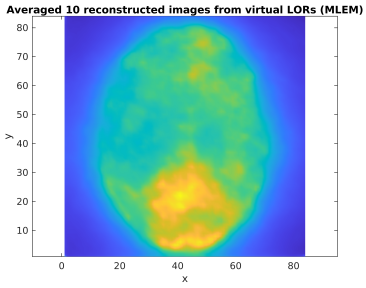
(c) Mean reconstructed image from virtual LORs, same 1000 random samples in each trial (FBP)



(d) Mean reconstructed image from original LORs (MLEM)

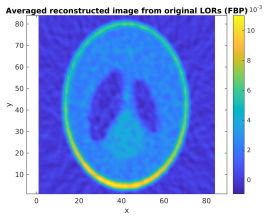


(e) Mean reconstructed image from virtual LORs, new 1000 random samples in each trial (MLEM)

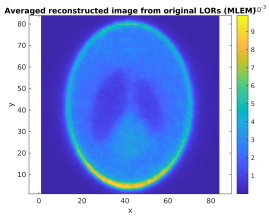


(f) Mean reconstructed image from virtual LORs, same 1000 random samples in each trial (MLEM)

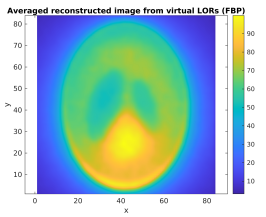
Figure 5.29: Averaged 10 reconstructed images from original and virtual LORs



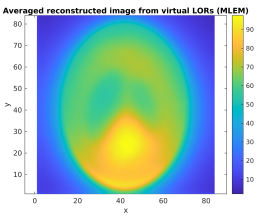
(a) The mean reconstructed image from original LORs (FBP)



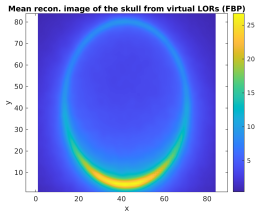
(b) The mean reconstructed image from original LORs (MLEM)



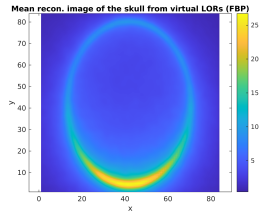
(c) The mean reconstructed image from virtual LORs (FBP)



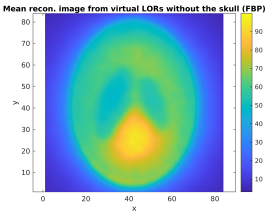
(d) The mean reconstructed image from virtual LORs (MLEM)



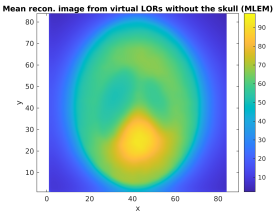
(e) The mean reconstructed image of the skull from virtual LORs (FBP)



(f) The mean reconstructed image of the skull from virtual LORs (MLEM)



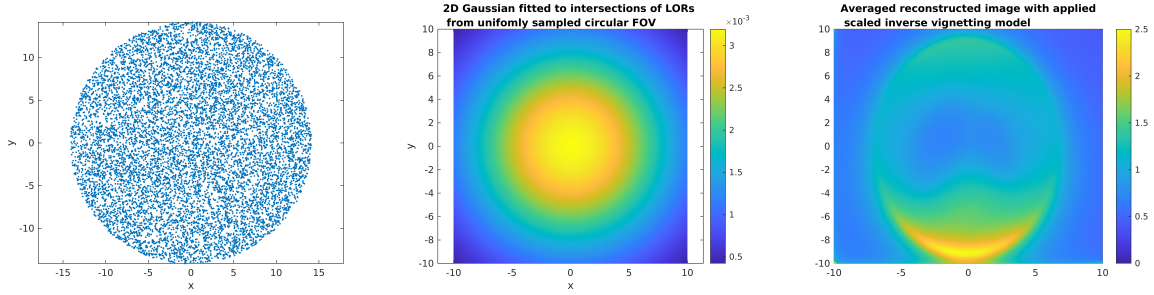
(g) The mean reconstructed image from virtual LORs without the skull (FBP)



(h) The mean reconstructed image from virtual LORs without the skull (MLEM)

Figure 5.30: Averaged reconstructed images from original and virtual LORs for 1000 trials where in each trial new 1000 random samples were generated

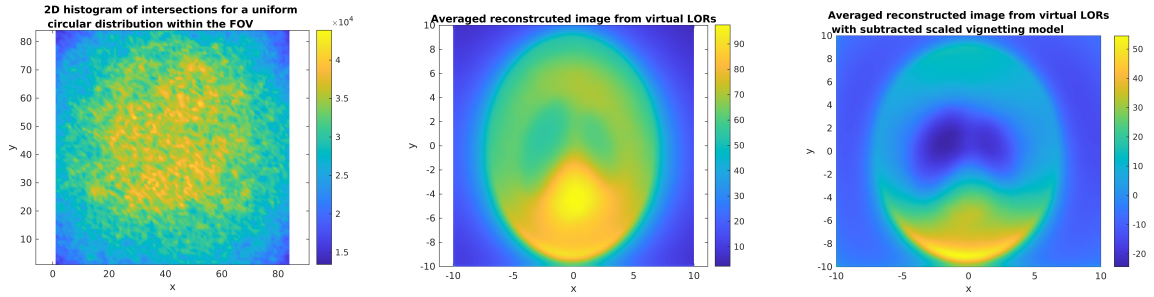
reconstructed phantom image in Fig. 5.31e and the results are shown in Fig. 5.31c. In the other version, the scaled vignetting model was subtracted from Fig. 5.31e and the results are shown in Fig. 5.31f.



(a) 2D circular process with 10000 samples

(b) 2D Gaussian fitted to the 2D histogram of intersections

(c) Application of inverse scaled vignetting model to (d)



(d) Averaged 2D histogram of intersections of LORs for 10 trials

(e) Averaged reconstructed images from virtual LORs (MLEM)

(f) Subtracted scaled vignetting model from (d)

Figure 5.31: Devignetting of averaged reconstructed image from virtual LORs with a 2D Gaussian model of intersections of LORs for uniformly sampled circular FOV. Estimated parameters of the 2D Gaussian: $[\hat{\mu_x}, \hat{\mu_y}] = [0.0706 - 0.0574]$, $\hat{\Sigma} = \begin{pmatrix} 43.4248 & 0.1253 \\ 0.1253 & 43.6020 \end{pmatrix}$, $\hat{A} = 2.486 \cdot 10^4$

Results show that the averaged reconstructed image from original LORs is better than the reconstructed image from virtual LORs if the effective number of process samples is large or if number of trials is large (1000 in this case). The image reconstructed from virtual LORs is blurred, as expected, but such a blur can be reduced by devignetting. However, if only 10 reconstructed image is averaged, the image obtained by using virtual LORs resembles the given phantom distribution more than the image obtained from a still insufficient number of original LORs.

5.2.16 Discussion

The first and second moments, i.e., the mean and the covariance matrix of the underlying 2D process, can be calculated from the first and second moments of intersections of LORs that pass through those process samples, but only for a single underlying 2D process. However, the analytical PDF of the underlying 2D process still remains unknown. By calculating the higher moments of the intersections of LORs, more statistical information about the underlying 2D process could be obtained. On the other hand, if the expectation of the PDF of intersections of

LORs were known, the deconvolution algorithms could be applied to reconstruct the underlying 2D process. For example, the expectation-maximization algorithm could be applied to find the process that is most likely to have a given total PDF of intersections. Since the distribution of intersections of LORs is related to each pair of process samples, simple deconvolution algorithms that assume that each point of the process is blurred with a specific PSF are inapplicable. The PDF of intersections for a pair of process samples was derived in the spatial and sinogram domains. The expectation of such a PDF was searched for by assuming that process samples come from some simple uniform 2D distribution. Even with the assumption that the process samples are uniformly distributed within a circle or rectangle, the analytical integral of the PDF of the intersections, i.e., the expectation of the PDF of intersections, was not found. Experimentally, this integral can be calculated by summing the individual PDFs of intersections for each unique pair of process samples. The shape of the underlying process can be guessed from such experimentally obtained total PDF of intersections of LORs, but it is blurred. In the case of two or more underlying 2D processes, the total PDF of the intersections of LORs is additionally more complex due to cross-intersections of LORs that originate from different components. Experiments showed that the blurring of a rotationally symmetric process that follows a 2D normal distribution can be approximated by a mixture of the distribution of that process and a heavy-tailed generalized normal distribution with a shape factor close to the shape factor of the Laplace distribution and with the scale factor related to the width of the underlying 2D normal distribution.

The experiments with the reconstructions of the Gaussian mixture model and the Shepp-Logan head phantom from intersections of LORs show that such reconstructions, although blurred and potentially biased, can be used for fast scanning in real-time applications since they require an extremely small number of process samples. Such reconstructions resemble the original underlying 2D processes and recognize the regions of the underlying processes' activities since the first moments of intersections of LORs and of original process samples are equal.

Chapter 6

Conclusion

The synthetic astronomical image that includes all expected image degradations of a typical imagining system was generated. The current implementation includes optical distortion, aberrations and vignetting, camera motion effects, electrical sensor properties, sensor noise, and spectral response of the optics and sensor. Parameters of individual degradation models are derived from a typical astronomical stack processing pipeline and are fed to the synthesis algorithm. Parameters need to be finely tuned to the actual imaging parameters for accurate modeling. Experimental results show that the synthesized image is close to the real one, but further parameter manipulation with objective similarity measures is necessary to obtain even more precise results. Such a synthetic image represents the ground-truth or high-resolution image which can be used for further research.

The main advantage of the proposed robust iteratively reweighted least squares (IRWLS) method for the estimation of overlapped 2D Gaussian components in astronomical images compared to the conventional least squares (LS) method is the more precise estimation of Gaussian profile parameters when neither the centers of the profiles nor their total number is known. Since the method is based only on the histogram of residual errors, it recognizes the outliers and applies soft thresholding, which with an appropriate threshold δ provides unbiased and efficient estimates as it is demonstrated through Monte Carlo analysis. It was shown that the method achieves greater accuracy up to 25 dB compared to the conventional least squares method for a certain set of differences in profiles' magnitudes and center-to-center distances. Moreover, the 80% of total improvement is achieved in only five iterations. As expected, the modeling gain is primarily determined by the distances of the profiles' centers. For the case of rotationally symmetric profiles, the minimum ratio of the effective profile width and the pixel size should be around one to have a sufficient number of samples for the optimization process and smaller variances of the estimated parameters. Through the synthetic example, the applicability of the proposed method for the analysis of astronomical images is demonstrated. The tuning constant of the Huber estimation in the relatively wide range from 1 to 2 results in approximately the

same IRWLS method efficiency. However, it primarily affects the convergence rate where the smaller values of the tuning constant yield a solution closer to the L_1 norm with a slower convergence rate, while the greater values of the tuning constant yield a solution closer to the L_2 norm with the faster convergence rate. The main benefit of the IRWLS method with Huber weights in the estimation from the noisy data is that a priori knowledge about the noise distribution is not required. The method can achieve a large modeling gain of up to 25 dB in the case of heavy-tailed noise contamination. Furthermore, the appropriate choice of the threshold δ yields the high asymptotic efficiency of the IRWLS method in the case of additive Gaussian noise contamination. For the case of Poisson distributed noise, the IRWLS method with inverse variance weights has the largest modeling gain of up to 3 dB compared to the IRWLS method with Huber weights and the LS method since the inverse variance weighting is optimal in the MLE sense for this type of contamination. For the case of additive Gaussian noise, the LS solution is optimal as expected. Differences in differential entropy for different ROI widths predict the reduction of estimation accuracy very well. The maximizing of differential entropy concerning the factor of Mahalanobis distance k yields the optimal ROI width for $k = 2$. Random samples uniformly sampled within such an ROI contain the maximal informativity and randomness in the case of additive Gaussian noise contamination. Smaller and larger values of factor k than $k = 2$, reduce the estimation accuracy. In the case of uniformly spaced samples with fixed ROI density, the total modeling error in dB reduces with wider ROI with the slope of 10 dB for each 10-fold increase in the number of samples after the compensation of variations in data informativity. Future work should be focused on the verification of the proposed IRWLS method to actual astronomical images.

In the case of multidimensional Gaussian profile fitting in the log domain, the experimental results show that in the case of high SNRs, both the WLS method in the log domain and the LS method in the domain of Gaussian profile values achieve the same accuracy. Still, the proposed method in the log domain converges faster, especially for the Gaussian profiles of higher dimensions. The main advantage of the proposed method is a one-step solution for the covariance matrix and linear scale for a given centroid position, thus avoiding the nonlinear optimization for the estimation of those parameters. For many practical applications where the initial centroid is already sufficiently accurately determined, only one single iteration of the proposed algorithm is sufficient to determine all remaining model parameters using the analytical procedure without the need for further iteration. Experiments showed that in the case of the 3D Gaussian profile estimation from data contaminated with additive Gaussian noise, the covariance matrix and the linear scale were estimated one or even two orders of magnitude faster using the analytical solution of the proposed method than using the iterative LS method in the domain of values. At the same time, for high SNRs ($\geq$ 40dB) the proposed method achieved almost the same total modeling error (with suboptimality of less than 3dB) without prior knowledge

except for the assumed centroid position in comparison to the iterative LS method in the domain of values, which yields an optimal solution for the case of Gaussian noise contamination. Moreover, experiments showed that such a model could be further improved by the described correction of the initial centroid through a maximum of four iterations of this two-step method. For comparison, numerical optimization procedures in the value domain require a significantly larger number of iterations than the proposed method to find all model parameters, which explains the significant advantage of our method. Due to its rapid convergence, the method can be used in real-time applications. The LS method in the domain of values is more accurate than the proposed WLS method in the case of low SNRs ($\leq 20\,\mathrm{dB}$) and wide ROIs. The loss of precision of the proposed method is dominantly caused by small profile values on such wide ROIs that can introduce significant errors in the log domain when the approximation of the exponential function with the first term of the Taylor series does not hold. The proposed method achieves higher accuracy in the case of model-driven estimation weights, which are selected as values of the initial, moment-based model. Such selected estimation weights are fixed for all iterations of the proposed method, thus guaranteeing the method's convergence. However, in the case of low SNRs, the initial model can significantly deviate from the specified model, and using the initial model weights yields even worse parameter estimates than using the input data-driven weights. This issue can be approached by iterating the estimation procedure where the weights in each iteration are selected as estimated model values from the previous step. Under the assumption of getting a better model in each iteration and consequently more appropriate weights, the updated estimates in each iteration will be closer to the given model. Such a procedure with adaptive weights should converge for high SNRs. As noise gradually grows and SNR reduces, the estimation should still converge and get a usable solution due to robustness until the noise does not become so high that the estimation diverges and becomes unfeasible. However, the guarantee of convergence is much more difficult to prove in the case of an iterative procedure with adaptive weights since in each iteration, in addition to the model, the optimization criterion also changes. However, this approach is undoubtedly worth future work. The proposed analytical method of resultants for the centroid estimation of the 2D Gaussian profile from noisy data yields up to 30 dB smaller MSE than the iterative LS method for a small number of input samples symmetrically distributed around the centroid. The robustness of the proposed method is validated for asymmetrically distributed samples, where the analytical solution for centroid reduces the MSE by up to 15 dB. The proposed method is up to 19 times faster than the iterative LS method without analytical gradients when the number of input samples is large ($m = 40000$), independently of the given SNR. Thereby, the method of resultants guarantees the finding of the global minimum even from sparsely sampled truncated estimation domain with low complexity.

The reconstruction of the underlying process in low-dose 2D PET imaging is possible by using the intersections of LORs which are quadratically larger than the number of process sam-

ples from which those LORs originate. Thereby, it is possible to simultaneously obtain a reconstruction of high spatial and high temporal resolution with extra small doses of the received radiotracer. The baseline experiment shows that the FBP algorithm gives an accurate reconstruction provided that the number of LORs is of the order of 10^7. Such a large number of LORs, i.e., process samples, very well represent the unknown distribution, but a reduction in the number of samples by an order of magnitude gives a grainy reconstructed image. The experimental results show that even in the case of a very small number of process samples whose number is hardly sufficient to represent an unknown distribution, the intersections of LORs can be successfully used as starting points for artificial generation of virtual LORs. Although at first glance the reconstructed image from virtual LORs at positions of intersections of original LORs is far from the original process, by averaging only 50 such reconstructions, a stable image of small variance of the order of 10^{-7} is obtained, thereby the variance decreases with the number of samples. It was proved that the mean of the process samples is equal to the mean of intersections and that the covariance matrix of the process is in analytical relation via the expectation of transformation matrix with the covariance of the intersections. The averaged reconstructed image primarily has an unwanted bias concerning the original process due to the difference in the covariance matrix of the process and intersections. This bias concerning the given density function could be compensated by nonlinear deconvolution by knowing the analytical relation of process and intersections spatial distributions. The PDF of intersections is derived for a pair of process samples in the spatial and the sinogram domain. The total PDF of intersections can be found as the sum of individual PDFs for all pairs of point sources. The regularization of initial angles is introduced to limit the dispersion of the intersections and enable the reconstruction of the underlying process. The regularization ensures the finite first and second moments of the distribution of intersections, enabling the calculation of the mean and the covariance matrix of the underlying process as well. The attempt to analytically derive the expectation of PDFs of intersections was unsuccessful even for a simple circular of rectangular 2D uniform distributions. If this expectation could be calculated, a reconstruction of any 2D input process could be obtained by applying expectation maximization or deconvolution algorithms. In the case of numerically calculated total PDF of intersections of underlying 2D Gaussian process, the distribution of its intersections is approximated by the mixture of the generalized normal distribution close to Laplace distribution and the normal distribution of the underlying process. Even though the reconstruction from intersections of LORs is blurred, it can be used as a fast scanning technique since it can recognize the regions of the underlying process activity from a very small number of coincident events.

Bibliography

[1]Klaus, A., Bauer, J., Karner, K., Elbischger, P., Perko, R., Bischof, H., "Camera calibration from a single night sky image", in Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., Vol. 1. IEEE, 2004, str. I–I.

[2]Johnson, H. L., Mitchell, R., Iriarte, B., Wisniewski, W., "UBVRIJKL Photometry of the Bright Stars", Communications of the lunar and planetary laboratory, Vol. 4, 1966, str. 99–110.

[3]Hoffleit, D., Warren Jr, W., "The Bright Star Catalogue", Astronomical Data Center Bulletin, Vol. 1, No. 4, 1987, str. 285–294.

[4]Hog, E., Fabricius, C., Makarov, V. V., Urban, S., Corbin, T., Wycoff, G., Bastian, U., Schwekendiek, P., Wicenec, A., "The Tycho-2 catalogue of the 2.5 million brightest stars", Naval Observatory Washington DC, Tech. Rep., 2000.

[5]Lasker, B. M., Sturch, C. R., McLean, B. J., Russell, J. L., Jenkner, H., Shara, M. M., "The Guide Star Catalog. i-Astronomical foundations and image processing", The Astronomical Journal, Vol. 99, 1990, str. 2019–2058.

[6]Jenkner, H., Lasker, B. M., Sturch, C. R., McLean, B. J., Shara, M. M., Russel, J., "The Guide Star Catalog. iii-Production, database organization, and population statistics", The Astronomical Journal, Vol. 99, 1990, str. 2082–2154.

[7]Remondino, F., Fraser, C., "Digital camera calibration methods: considerations and comparisons", International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 36, No. 5, 2006, str. 266–272.

[8]Zhang, Z., "A flexible new technique for camera calibration", IEEE Transactions on pattern analysis and machine intelligence, Vol. 22, No. 11, 2000, str. 1330–1334.

[9]Calabretta, M. R., Greisen, E. W., "Representations of celestial coordinates in FITS", Astronomy & Astrophysics, Vol. 395, No. 3, 2002, str. 1077–1122.

[10]Ricolfe-Viala, C., Sánchez-Salmerón, A.-J., "Robust metric calibration of non-linear camera lens distortion", Pattern Recognition, Vol. 43, No. 4, 2010, str. 1688–1699.

[11]Wang, J., Shi, F., Zhang, J., Liu, Y., "A new calibration model of camera lens distortion", Pattern recognition, Vol. 41, No. 2, 2008, str. 607–615.

[12]Weng, J., Cohen, P., Herniou, M. *et al.*, "Camera calibration with distortion models and accuracy evaluation", IEEE Transactions on pattern analysis and machine intelligence, Vol. 14, No. 10, 1992, str. 965–980.

[13]Goodman, J. W., Introduction to Fourier optics. Roberts and Company Publishers, 2005.

[14]Kwak, S. G., Kim, J. H., "Central limit theorem: the cornerstone of modern statistics", Korean journal of anesthesiology, Vol. 70, No. 2, 2017, str. 144–156.

[15]Green, P. J., "Bayesian reconstructions from emission tomography data using a modified EM algorithm", IEEE transactions on medical imaging, Vol. 9, No. 1, 1990, str. 84–93.

[16]Wyant, J. C., Creath, K., "Basic wavefront aberration theory for optical metrology", Applied optics and optical engineering, Vol. 11, No. part 2, 1992, str. 28–39.

[17]Montgomery, D. C., Runger, G. C., Applied statistics and probability for engineers. John Wiley & Sons, 2010.

[18]Hasinoff, S. W., "Photon, Poisson Noise.", 2014.

[19]Makitalo, M., Foi, A., "Optimal inversion of the Anscombe transformation in low-count Poisson image denoising", IEEE transactions on Image Processing, Vol. 20, No. 1, 2010, str. 99–109.

[20]Gribl, A., Petrinovi ć, D., "Synthetic astronomical image sequence generation", in 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO). IEEE, 2020, str. 1103–1108.

[21]Levenberg, K., "A method for the solution of certain non-linear problems in least squares", Quarterly of applied mathematics, Vol. 2, No. 2, 1944, str. 164–168.

[22]Richards, F. S., "A method of maximum-likelihood estimation", Journal of the Royal Statistical Society Series B: Statistical Methodology, Vol. 23, No. 2, 1961, str. 469–475.

[23]Sparrow, C. M., "On spectroscopic resolving power", The Astrophysical Journal, Vol. 44, 1916, str. 76.

[24]Asakura, T., “Resolution of two unequally bright points with partially coherent light”, Nouvelle Revue d’Optique, Vol. 5, No. 3, 1974, str. 169.

[25]Mallat, S. G., Zhang, Z., “Matching pursuits with time-frequency dictionaries”, IEEE Transactions on signal processing, Vol. 41, No. 12, 1993, str. 3397–3415.

[26]Bergeaud, F., Mallat, S., “Matching pursuit of images”, in Proceedings., International Conference on Image Processing, Vol. 1. IEEE, 1995, str. 53–56.

[27]Rebollo-Neira, L., Bowley, J., “Sparse representation of astronomical images”, JOSA A, Vol. 30, No. 4, 2013, str. 758–768.

[28]Mendels, F., Vandergheynst, P., Thiran, J.-P., “Rotation and scale invariant shape representation and recognition using matching pursuit”, in Object recognition supported by user interaction for service robots, Vol. 4. IEEE, 2002, str. 326–329.

[29]Phillips, P. J., “Matching pursuit filters applied to face identification”, IEEE Transactions on Image Processing, Vol. 7, No. 8, 1998, str. 1150–1164.

[30]Hsu, S.-H., Huang, C.-L., “Road sign detection and recognition using matching pursuit method”, Image and Vision Computing, Vol. 19, No. 3, 2001, str. 119–129.

[31]Díaz-Hernández, R., Peregrina-Barreto, H., Altamirano-Robles, L., González-Bernal, J., Ortiz-Esquivel, A., “Automatic stellar spectral classification via sparse representations and dictionary learning”, Experimental Astronomy, Vol. 38, No. 1-2, 2014, str. 193–211.

[32]Kazovsky, L., “Beam position estimation by means of detector arrays”, Optical and Quantum Electronics, Vol. 13, No. 3, 1981, str. 201–208.

[33]Lenz, D. D., Ayres, T. R., “Errors associated with fitting Gaussian profiles to noisy emission-line spectra”, Publications of the Astronomical Society of the Pacific, Vol. 104, No. 681, 1992, str. 1104.

[34]Condon, J., “Errors in elliptical Gaussian fits”, Publications of the Astronomical Society of the Pacific, Vol. 109, No. 732, 1997, str. 166.

[35]Ireland, J., “Precision limits to emission-line profile measuring experiments”, The Astrophysical Journal, Vol. 620, No. 2, 2005, str. 1132.

[36]Auer, L., Van Altena, W., “Digital image centering. ii”, The Astronomical Journal, Vol. 83, 1978, str. 531–537.

[37]Delabie, T., Schutter, J. D., Vandenbussche, B., "An accurate and efficient Gaussian fit centroiding algorithm for star trackers", The Journal of the Astronautical Sciences, Vol. 61, 2014, str. 60–84.

[38]Lei, Z., Chen, X., Lei, T., "Sub-pixel location of motion blurred weak celestial objects in optical sensor image based on elliptical 2D Gaussian surface fitting", in 2016 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII). IEEE, 2016, str. 100–104.

[39]Thompson, R. E., Larson, D. R., Webb, W. W., "Precise nanometer localization analysis for individual fluorescent probes", Biophysical journal, Vol. 82, No. 5, 2002, str. 2775–2783.

[40]Babcock, H., Sigal, Y. M., Zhuang, X., "A high-density 3D localization algorithm for stochastic optical reconstruction microscopy", Optical nanoscopy, Vol. 1, No. 1, 2012, str. 1–10.

[41]Cheezum, M. K., Walker, W. F., Guilford, W. H., "Quantitative comparison of algorithms for tracking single fluorescent particles", Biophysical journal, Vol. 81, No. 4, 2001, str. 2378–2388.

[42]Shen, H., Tauzin, L. J., Baiyasi, R., Wang, W., Moringo, N., Shuang, B., Landes, C. F., "Single particle tracking: from theory to biophysical applications", Chemical reviews, Vol. 117, No. 11, 2017, str. 7331–7376.

[43]Chen, Y.-C., Furenlid, L. R., Wilson, D. W., Barrett, H. H., "Calibration of scintillation cameras and pinhole SPECT imaging systems", in Small-Animal SPECT Imaging. Springer, 2005, str. 195–201.

[44]Okada, K., Comaniciu, D., Krishnan, A., "Robust anisotropic Gaussian fitting for volumetric characterization of pulmonary nodules in multislice CT", IEEE Transactions on Medical Imaging, Vol. 24, No. 3, 2005, str. 409–423.

[45]Chen, Y., Ma, J., Zhang, P., Liu, F., Mei, S., "Robust state estimator based on maximum exponential absolute value", IEEE Transactions on Smart Grid, Vol. 8, No. 4, 2015, str. 1537–1544.

[46]Chen, Y., Liu, F., Mei, S., Ma, J., "Toward adaptive robust state estimation based on mcc by using the generalized gaussian density as kernel functions", International Journal of Electrical Power & Energy Systems, Vol. 71, 2015, str. 297–304.

[47]Feng, Y., Goree, J., Liu, B., "Accurate particle position measurement from images", Review of scientific instruments, Vol. 78, No. 5, 2007, str. 053704.

[48]Campbell, S., “Comparison of the generalized centroid with Gaussian and quadratic peak localization methods”, arXiv preprint arXiv:1807.08355, 2018.

[49]Vyas, A., Roopashree, M., Prasad, B. R., “Centroid detection by Gaussian pattern matching in adaptive optics”, arXiv preprint arXiv:0910.3386, 2009.

[50]Spille, J.-H., Kaminski, T., Königshoven, H.-P., Kubitscheck, U., “Dynamic three-dimensional tracking of single fluorescent nanoparticles deep inside living tissue”, Optics express, Vol. 20, No. 18, 2012, str. 19 697–19 707.

[51]Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P., Numerical recipes 3rd edition: The art of scientific computing. Cambridge university press, 2007.

[52]Anthony, S. M., Granick, S., “Image analysis with rapid and accurate two-dimensional Gaussian fitting”, Langmuir, Vol. 25, No. 14, 2009, str. 8152–8160.

[53]Smith, C. S., Joseph, N., Rieger, B., Lidke, K. A., “Fast, single-molecule localization that achieves theoretically minimum uncertainty”, Nature methods, Vol. 7, No. 5, 2010, str. 373–375.

[54]Zhu, X., Zhang, D., “Efficient parallel Levenberg-Marquardt model fitting towards real-time automated parametric imaging microscopy”, PloS one, Vol. 8, No. 10, 2013, str. e76665.

[55]Leutenegger, M., Weber, M., “Least-squares fitting of Gaussian spots on graphics processing units”, arXiv preprint arXiv:2106.02045, 2021.

[56]Chouzenoux, E., Lau, T. T.-K., Lefort, C., Pesquet, J.-C., “Optimal multivariate Gaussian fitting with applications to PSF modeling in two-photon microscopy imaging”, Journal of Mathematical Imaging and Vision, Vol. 61, No. 7, 2019, str. 1037–1050.

[57]Basseville, M., Cardoso, J.-F., “On entropies, divergences, and mean values”, in Proceedings of 1995 IEEE International Symposium on Information Theory, 1995, str. 330-.

[58]Hagen, N., Kupinski, M., Dereniak, E. L., “Gaussian profile estimation in one dimension”, Applied optics, Vol. 46, No. 22, 2007, str. 5374–5383.

[59]Hagen, N., Dereniak, E. L., “Gaussian profile estimation in two dimensions”, Applied optics, Vol. 47, No. 36, 2008, str. 6842–6851.

[60]Caruana, R. A., Searle, R. B., Heller, T., Shupack, S. I., “Fast algorithm for the resolution of spectra”, Analytical chemistry, Vol. 58, No. 6, 1986, str. 1162–1167.

[61]Guo, H., “A simple algorithm for fitting a Gaussian function [DSP tips and tricks]”, IEEE Signal Processing Magazine, Vol. 28, No. 5, 2011, str. 134–137.

[62]Al-Nahhal, I., Dobre, O. A., Basar, E., Moloney, C., Ikki, S., “A fast, accurate, and separable method for fitting a Gaussian function [tips & tricks]”, IEEE Signal Processing Magazine, Vol. 36, No. 6, 2019, str. 157–163.

[63]Roonizi, E. K., “A new algorithm for fitting a Gaussian function riding on the polynomial background”, IEEE Signal Processing Letters, Vol. 20, No. 11, 2013, str. 1062–1065.

[64]Kwak, M., Lkhagvasuren, B., Sun, X., “An efficient method for fitting Gaussian functions”, Iranian Journal of Science and Technology, Transactions A: Science, Vol. 45, No. 3, 2021, str. 1043–1056.

[65]Gribl, A., Petrinovi ć, D., “A robust method for gaussian profile estimation in the case of overlapping objects”, IEEE Access, 2021, str. 1-1.

[66]Burrus, C. S., “Iterative reweighted least squares”, OpenStax CNX. Available online: http://cnx. org/contents/92b90377-2b34-49e4-b26f-7fe572db78a1, Vol. 12, 2012.

[67]Huber, P. J., “Robust estimation of a location parameter: Annals mathematics statistics, 35”, 1964.

[68]Fox, J., Weisberg, S., “Robust regression”, An R and S-Plus companion to applied regression, Vol. 91, 2002.

[69]A. Gribl, D. P., “The influence of the Huber estimator tuning constant on the performance of the iteratively reweighted least squares method”, 2020, str. 12-14.

[70]Aguerrebere, C., Delon, J., Gousseau, Y., Musé, P., “Study of the digital camera acquisition process and statistical modeling of the sensor raw data”, 2013.

[71]Theuwissen, A. J., “Better pictures through physics”, IEEE Solid-State Circuits Magazine, Vol. 2, No. 2, 2010, str. 22–28.

[72]Konnik, M., Welsh, J., “High-level numerical simulations of noise in CCD and CMOS photosensors: review and tutorial”, arXiv preprint arXiv:1412.4031, 2014.

[73]Costantini, R., Susstrunk, S., “Virtual sensor design”, in Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications V, Vol. 5301. International Society for Optics and Photonics, 2004, str. 408–419.

[74]Gribl, A., Petrinovi ć, D., “The influence of noise on 2D Gaussian profile parameters estimation”, in 2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO). IEEE, 2021, str. 1032–1037.

[75]Michalowicz, J. V., Nichols, J. M., Bucholtz, F., Handbook of differential entropy. Crc Press, 2013.

[76]Koš čević, A. G., Petrinović, D., “Maximizing accuracy of 2D Gaussian profile estimation using differential entropy”, in 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO). IEEE, 2022, str. 888–893.

[77]Koš čević, A. G., Petrinović, D., “A fast method for fitting a multidimensional Gaussian function”, IEEE Access, Vol. 10, 2022, str. 106 921–106 935.

[78]Macaulay, F. S., “Some formulae in elimination”, Proceedings of the London Mathematical Society, Vol. 1, No. 1, 1902, str. 3–27.

[79]Gribl Koš čević, A., Petrinović, D., “Fitting an elliptical paraboloid with the known shape to the empirical data”, 2022, str. 13-15.

[80]Merriman, M., A List of Writings Relating to the Method of Least Squares: With Historical and Critical Notes. Academy, 1877, Vol. 4.

[81]Johnson, M. L., Faunt, L. M., “[1] Parameter estimation by least-squares methods”, in Methods in enzymology. Elsevier, 1992, Vol. 210, str. 1–37.

[82]Fisher, R., Naidu, D., “A comparison of algorithms for subpixel peak detection”, in Image technology. Springer, 1996, str. 385–404.

[83]Thomas, S., Fusco, T., Tokovinin, A., Nicolle, M., Michau, V., Rousset, G., “Comparison of centroid computation algorithms in a Shack-Hartmann sensor”, Monthly Notices of the Royal Astronomical Society, Vol. 371, No. 1, 2006, str. 323–336.

[84]Baker, K., Moallem, M., “Iteratively weighted centroiding for Shack-Hartmann wavefront sensors”, Optics express, Vol. 15, No. 8, 2007, str. 5147–5159.

[85]Parthasarathy, R., “Rapid, accurate particle tracking by calculation of radial symmetry centers”, Nature methods, Vol. 9, No. 7, 2012, str. 724–726.

[86]Mukherjee, D., Miao, L., Stone, G., Alem, N., “mpfit: a robust method for fitting atomic resolution images with multiple Gaussian peaks”, Advanced Structural and Chemical Imaging, Vol. 6, No. 1, 2020, str. 1–12.

[87]Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., Meester, L. E., A Modern Introduction to Probability and Statistics: Understanding why and how. Springer, 2005, Vol. 488.

[88]Avriel, M., Nonlinear programming: analysis and methods. Courier Corporation, 2003.

[89]Akritas, A., Malaschonok, G., Vigklas, P., "Sturm sequences and modified subresultant polynomial remainder sequences", Serdica Journal of Computing, Vol. 8, No. 1, 2014, str. 29–46.

[90]Gelfand, I. M., Kapranov, M. M., Zelevinsky, A. V., Gelfand, I. M., Kapranov, M. M., Zelevinsky, A. V., A-discriminants. Springer, 1994.

[91]Bailey, D. L., Maisey, M. N., Townsend, D. W., Valk, P. E., Positron emission tomography. Springer, 2005, Vol. 2.

[92]Weber, W. A., "Use of PET for monitoring cancer therapy and for predicting outcome", Journal of Nuclear Medicine, Vol. 46, No. 6, 2005, str. 983–995.

[93]Natterer, F., Wübbeling, F., Mathematical methods in image reconstruction. SIAM, 2001.

[94]Tong, S., Alessio, A. M., Kinahan, P. E., "Image reconstruction for PET/CT scanners: past achievements and future challenges", Imaging in medicine, Vol. 2, No. 5, 2010, str. 529.

[95]Defrise, M., Kinahan, P. E., Michel, C. J., "Image reconstruction algorithms in PET", in Positron Emission Tomography. Springer, 2005, str. 63–91.

[96]Alessio, A., Kinahan, P. *et al.*, "PET image reconstruction", Nuclear medicine, Vol. 1, No. 1, 2006, str. 1–22.

[97]Reader, A. J., Zaidi, H., "Advances in PET image reconstruction", PET clinics, Vol. 2, No. 2, 2007, str. 173–190.

[98]Kak, A. C., Slaney, M., Wang, G., "Principles of computerized tomographic imaging", 2002.

[99]Catana, C., "The dawn of a new era in low-dose PET imaging", str. 657–658, 2019.

[100]Yu, S., Muhammed, H. H., "Noise type evaluation in positron emission tomography images", in 2016 1st International Conference on Biomedical Engineering (IBIOMED). IEEE, 2016, str. 1–6.

[101]Chen, K. T., Gong, E., de Carvalho Macruz, F. B., Xu, J., Boumis, A., Khalighi, M., Poston, K. L., Sha, S. J., Greicius, M. D., Mormino, E. *et al.*, "Ultra–low-dose 18F-florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs", Radiology, Vol. 290, No. 3, 2019, str. 649–656.

[102]Wang, C., Hu, Z., Shi, P., Liu, H., "Low dose PET reconstruction with total variation regularization", in 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2014, str. 1917–1920.

[103]Gong, K., Berg, E., Cherry, S. R., Qi, J., "Machine learning in PET: from photon detection to quantitative image reconstruction", Proceedings of the IEEE, Vol. 108, No. 1, 2019, str. 51–68.

[104]Wang, Y., Zhang, P., An, L., Ma, G., Kang, J., Shi, F., Wu, X., Zhou, J., Lalush, D. S., Lin, W. *et al.*, "Predicting standard-dose PET image from low-dose PET and multimodal mr images using mapping-based sparse representation", Physics in Medicine & Biology, Vol. 61, No. 2, 2016, str. 791.

[105]Reader, A. J., Corda, G., Mehranian, A., da Costa-Luis, C., Ellis, S., Schnabel, J. A., "Deep learning for PET image reconstruction", IEEE Transactions on Radiation and Plasma Medical Sciences, Vol. 5, No. 1, 2020, str. 1–25.

[106]Feng, Q., Liu, H., "Rethinking PET image reconstruction: Ultra-low-dose, sinogram and deep learning", in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2020, str. 783–792.

[107]Xu, J., Gong, E., Pauly, J., Zaharchuk, G., "200x low-dose PET reconstruction using deep learning", arXiv preprint arXiv:1712.04119, 2017.

[108]Wernick, M. N., Aarsvold, J. N., Emission tomography: the fundamentals of PET and SPECT. Elsevier, 2004.

[109]Shepp, L. A., Vardi, Y., "Maximum likelihood reconstruction for emission tomography", IEEE transactions on medical imaging, Vol. 1, No. 2, 1982, str. 113–122.

[110]Meng, X.-L., Van Dyk, D., "The EM Algorithm—an Old Folk-song Sung to a Fast New Tune", Journal of the Royal Statistical Society: Series B (Methodological), Vol. 59, No. 3, 01 2002, str. 511-567, dostupno na: https://doi.org/10.1111/1467-9868.00082

[111]Zhou, J., Said, S., Berthoumieu, Y., "Online estimation of MGGD: the Riemannian averaged natural gradient method", in 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). IEEE, 2019, str. 515–519.

[112]Zeng, G. L., Medical image reconstruction. Springer, 2010, Vol. 530.

Biography

Anita Gribl Koščević was born in 1995 in Slavonski Brod, Croatia. She received BSc and MSc degrees in Electrical Engineering and Information Technology from the University of Zagreb Faculty of Electrical Engineering and Computing, in 2016 and 2018, respectively.

From November 2018 to September 2019, she was employed as a research associate at the Department of Electronics, Microelectronics, Computer and Intelligent Systems, University of Zagreb Faculty of Electrical Engineering and Computing, where she participated in the research-development project supported by European Regional Development Fund "Security of Cloud Computing and Cloud-Based Mobile Applications". She enrolled in doctoral study in March 2019.

Since October 2019, she has been employed as a research assistant at the Department of Electronic Systems and Information Processing at the same faculty, where she participated in the scientific projects Beyond Nyquist Limit, IP-2014-09-2625, Renaissance of the Sampling Theory, IP-2019-04-6703, and DOK-01-2018 funded by Croatian Science Foundation.

Her research interests include image processing, estimation theory, and applied statistics. She has been involved in teaching activities in the Embedded Systems and Embedded System Design courses. She is also a student member of the IEEE Association and author of two journal papers, four conference papers, and two abstracts in conference proceedings.

List of publications

Journal papers

1. A. Gribl Koš čević and D. Petrinović, "A Fast Method for Fitting a Multidimensional Gaussian Function," in IEEE Access, vol. 10, pp. 106921-106935, 2022, doi: 10.1109/ACCESS.2022.3212388.
2. A. Gribl and D. Petrinovi ć, "A Robust Method for Gaussian Profile Estimation in the Case of Overlapping Objects," in IEEE Access, vol. 9, pp. 21071-21084, 2021, doi: 10.1109/ACCESS.2021.3055282.

Conference papers

1. A. Gribl Koš čević and D. Petrinović, "Extra-low-dose 2D PET imaging," 2021 12th International Symposium on Image and Signal Processing and Analysis (ISPA), Zagreb, Croatia, 2021, pp. 84-90, doi: 10.1109/ISPA52656.2021.9552059.
2. A. Gribl Koš čević and D. Petrinović, "Maximizing accuracy of 2D Gaussian profile estimation using differential entropy," 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 2022, pp. 888-893, doi: 10.23919/MIPRO55190.2022.9803382.
3. A. Gribl and D. Petrinovi ć, "The Influence of Noise on 2D Gaussian Profile Parameters Estimation," 2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 2021, pp. 1032-1037, doi: 10.23919/MIPRO52101.2021.9596661.
4. A. Gribl and D. Petrinovi ć, "Synthetic astronomical image sequence generation," 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 2020, pp. 1103-1108, doi: 10.23919/MIPRO48935.2020.9245294.

Abstracts in conference proceedings

1. A. Gribl Koš čević and D. Petrinović, "Fitting an elliptical paraboloid with the known shape to the empirical data," Abstract Book, 7th International Workshop on Data Science (IWDS 2022), Zagreb, Croatia, October 26, 2022, pp. 13-15
2. A. Gribl and D. Petrinovi ć, "The influence of the Huber estimator tuning constant on the performance of the iteratively reweighted least squares method," Abstract Book, Fifth International Workshop on Data Science (IWDS 2020), Zagreb, Croatia, November 24, 2020, pp. 12-14

Životopis

Anita Gribl Koščević rođena je 1995. godine u Slavonskom Brodu, Hrvatska. Na Fakultetu elektrotehnike i računarstva Sveučilišta u Zagrebu stekla je diplome sveučilišnog prvostupnika i magistra inženjera elektrotehnike i informacijske tehnologije 2016., odnosno 2018. godine.

Od studenog 2018. do rujna 2019. zaposlena je kao znanstvena suradnica na Zavodu za elektroniku, mikroelektroniku, računalne i inteligentne sustave Sveučilišta u Zagrebu, Fakulteta elektrotehnike i računarstva, gdje je sudjelovala na istraživačko-razvojnom projektu podržanom od strane Europskog fonda za regionalni razvoj „Sigurnost računarstva u oblaku prilikom korištenja mobilnih aplikacija“. Doktorski studij upisala je u ožujku 2019. godine.

Od listopada 2019. zaposlena je kao asistent na Zavodu za elektroničke sustave i obradbu informacija na istom fakultetu, gdje je sudjelovala na znanstvenim projektima Iznad Nyquistove granice, IP-2014-09-2625, Renesansa teorije uzorkovanja, IP-2019-04-6703, i DOK-01-2018 koje financira Hrvatska zaklada za znanost. Njezini znanstveni interesi uključuju obradu slike, teoriju estimacije i primijenjenu statistiku.

Sudjelovala je u nastavnim aktivnostima na kolegijima Ugradbeni računalni sustavi i Projektiranje ugradbenih računalnih sustava. Također je studentski IEEE član i autorica dva rada u časopisima, četiri rada na konferencijama i dva sažetka u konferencijskim zbornicima radova.