

Numerische Verfahren für die PET

Diplomarbeit
zur Erlangung des akademischen Grades einer
Diplom-Mathematikerin
der Naturwissenschaftlichen-Technischen Fakultät I
der Universität des Saarlandes

von
Katja Kruschewski

Saarbrücken
Juli 2004

Betreuer: Univ.-Prof. Dr. A. K. Louis

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, daß ich die vorliegende Arbeit eigenständig und ausschließlich mit den angegebenen Hilfsmitteln angefertigt habe.

Saarbrücken, den 26.07.2004

Katja Kruschewski

Inhaltsverzeichnis

Vorwort	iv
1 Einführung in die PET	1
1.1 Ein medizinisch-geschichtlicher Überblick	1
1.1.1 Die Geschichte der PET von den Anfängen bis zum ersten, industriereifen Tomographen	1
1.1.2 PET in der Onkologie	2
1.2 Physikalische Grundlagen der PET	3
1.3 Gütekriterien	6
1.3.1 Streuung	6
1.3.2 Dämpfung	8
1.3.3 Zufällige Koinzidenzen	10
1.3.4 Totzeit	10
1.3.5 Rauschen	10
2 Mathematische Grundlagen	13
2.1 Grundlagen aus der Funktionalanalysis	13
2.2 Grundlagen aus der Bildrekonstruktion	16
2.2.1 Die Fouriertransformation	16
2.2.2 Die Radontransformation	22
2.2.3 Die Röntgentransformation	24
2.3 Grundlagen aus der Statistik	25
2.4 Grundlagen aus der nichtlinearen Optimierung	29
3 Das D-C-Modell (Grundmodell) und ein darauf beruhender Bildrekonstruktionsansatz	34
3.1 Das D-C-Modell	34
3.2 Lineare inverse Probleme und der Lösungsansatz der verallgemeinerten Inversen	36
3.3 DART	42
3.4 Diskussion des Lösungsansatzes DART und einige Randbemerkungen	45
4 Das C-C-Modell und darauf beruhende Bildrekonstruktionen	47
4.1 Das C-C Modell	47
4.2 Datenformate, Projektionen und Sinogramme	52

Inhaltsverzeichnis

4.3	Datenreduktion	59
4.4	2D PET	61
4.4.1	2D FBP	62
4.4.2	2D CBP	67
4.5	3D PET	68
4.5.1	3D FBP für komplette 2D parallele Projektionen	72
4.5.2	Der 3D Rückprojektions-Algorithmus (3DRP)	84
4.6	Rebinningmethoden	90
4.6.1	FOREX	92
4.6.2	FORE	109
4.6.3	FORE-J	118
4.6.4	Zusammenhang zwischen FORE, FOREX und FORE-J	125
4.7	Vergleich der Algorithmen dieses Kapitels	128
5	Das D-D-Modell und darauf beruhende Bildrekonstruktionen	139
5.1	Das D-D Modell	139
5.2	Allgemeine Komponenten für das D-D Modell	140
5.3	Iterative Methoden	142
5.3.1	EM-Algorithmen	143
5.3.2	ART (Algebraic Reconstruction technique)	164
5.3.3	Zusammenhang zwischen OSEM und ART	173
5.4	Zusammenfassung	176
6	Ausblick	178

Vorwort

Seit Ende der 80'er Jahre wird PET als eine wesentliche 3D Bildgebungstechnik betrachtet. Jedoch erst in den letzten 10-12 Jahren widmete man sich der Entwicklung der Methodik des Datenerwerbs und der Rekonstruktion von PET-Daten im 3D. Der Hauptteil dieser Arbeit ist in verschiedenen naturwissenschaftlichen Journalen erschienen, aber leider wurde bis heute - bis auf das Werk von Bendriem [6], welches aber nur einen sehr dünnen Beitrag zum mathematischen Hintergrund liefert - kein mir bekanntes Werk veröffentlicht, in dem ausschließlich zu der Thematik "PET" die interessanten mathematischen Fragestellungen und Entwicklungen in einem einzigen Exemplar dargelegt wurden. Da die Anwendung der 3D-PET wachsendes Interesse an vielen PET-Zentren erhält, ist diese Arbeit ein Versuch, die oben genannte Lücke zu füllen.

Das Ziel dieser Arbeit ist es zum einen, dem Leser einen Überblick über den historischen und physikalischen Hintergrund der PET zu geben und ihm die Wichtigkeit dieser Technik für die heutige Medizin zu verdeutlichen, zum anderen - und dies ist unser Hauptziel - soll der mathematische Hintergrund eingehend durchleuchtet werden.

In Kapitel 1 werden wir daher einen kurzen geschichtlichen Abriss angeben, die heutige Anwendung der PET in der Onkologie - mit allen ihren Vorteilen gegenüber der CT - etwas genauer beschreiben und zum Schluß die zu Grunde liegende Physik und einige der wichtigsten Gütekriterien für diese Technik darlegen.

Kapitel 2 stellt dann die mathematische Einleitung zu dieser Arbeit dar, wobei wir in diesem Kapitel alles zur Verfügung stellen werden, was in den darauf folgenden Kapiteln Anwendung bzw. Verwendung findet.

Die Idee zur Einteilung der Arbeit in die drei Hauptkapitel 3, 4 und 5 entstammt aus dem Paper von Lewitt [10]. In Kapitel 3 beginnen wir dabei mit der Beschreibung eines mathematischen Modells, das auf der einen Seite so einfach wie möglich gehalten wird, auf der anderen Seite jedoch die Physik des Datenerwerbs möglichst gut wiedergibt. Auf der Grundlage dieses Modells leiten wir dann ein Rekonstruktionsverfahren her, das auf den Ideen der Approximativen Inversen beruht. Dabei wird der mathematische Hintergrund der Approximativen Inversen, die Herleitung des Rekonstruktionsverfahrens und eines dazu modifizierten Verfahrens eingehend besprochen (wobei der 3D-Fall hier völlig analog zum 2D-Fall zu sehen ist). Am Ende dieses Kapitels wird dann der Nutzen dieser Art von Verfahren für die Praxis diskutiert.

Durch gewisse Annahmen und Modifizierungen des Grundmodells aus Kapitel 3 erhalten wir schließlich unser mathematisches Modell für die Verfahren in Kapitel 4. Dieses Modell führt uns zu analytischen Methoden, für die Inversionsformeln hergeleitet werden können. Einteilung und Vorgehensweise in diesem gesamten Kapitel beruhen auf den Ideen von Bendriem [6]. Im Gegensatz zu Kapitel 3 und 5 müssen wir hier sehr darauf

achten, ob wir unsere Daten nur im 2D-Modus oder im 3D-Modus erwerben. Wir werden daher einen Überblick über den Datenerwerb, der sowohl im Sinogramm- als auch im Projektionsformat erfolgen kann, geben und dann auch noch ein paar Worte über die Datenreduktion verlieren (da wir im Vergleich zur 2D-PET bei der 3D-PET wesentlich mehr Daten erwerben, die alle abgespeichert und bearbeitet werden müssen, ist der Vorgang der Datenreduktion durchaus ein paar Zeilen bzw. Gedanken wert). Im Folgenden werden wir in diesem Kapitel die 2D-PET mit ihrer Standardrekonstruktionsmethode, der 2D-FBP, vorstellen und noch kurz anhand eines Beispiels darauf eingehen, wie man andere 2D-Inversionsformeln zur 2D-FBP in Beziehung setzen kann. Danach behandeln wir die sehr aufwendige 3D-PET, stellen das damit in Zusammenhang stehende Problem der abgeschnittenen Daten vor und leiten dann eine 3D-Rekonstruktionsformel her, einmal unter der Annahme, dass unsere Daten komplett sind und ein zweites Mal unter der Annahme, dass die uns fehlenden Daten erst durch Vorwärtsprojektion beschafft werden müssen. Danach kommen wir zu dem letzten großen Abschnitt dieses Kapitels, den Rebinningmethoden. Da die Daten im 3D erworben werden, jedoch so "sortiert" werden, dass nachher nur noch 2D-Rekonstruktionsmethoden benutzt werden, widmen wir diesen Methoden einen eigenen Abschnitt. Wir erläutern, was Rebinning bedeutet und leiten drei der aktuellsten Rebinningmethoden her. Am Ende dieses Kapitels vergleichen wir wiederum alle hergeleiteten Verfahren miteinander.

Das mathematische Modell in Kapitel 5 erhält man ebenfalls aus dem Grundmodell aus Kapitel 3. Das durch Diskretisierung erhaltene Modell berücksichtigt - im Vergleich zu den anderen Modellen - am meisten die statistische Natur des Datenerwerbs und der Großteil der darauf beruhenden Verfahren sind die sogenannten "iterativen Verfahren". Diese wollen wir dann auch näher betrachten, wobei wir dabei nur die beiden iterativen Verfahren herausgreifen, die in der Praxis tatsächlich Anwendung finden. Auch hier können wir den 3D-Fall völlig analog zum 2D-Fall betrachten. Die Beweise bzw. Beweisideen zu diesem Kapitel entstammen dabei hauptsächlich aus Natterer [11]. Eines der beiden vorgestellten Verfahren, der ML-EM Algorithmus hat zu der Einführung vieler verwandter Techniken geführt. Dieser Algorithmus beruht darauf, dass wir eine bedingte Erwartung maximieren wollen. In der Theorie ist die Maximierung dieser bedingten Erwartung äquivalent zur ML-Aufgabe. Da dieser Algorithmus viele Probleme bei der Implementierung mit sich bringt, versuchen wir im Folgenden, Lösungsmöglichkeiten dafür zu finden. Dies führt uns zu weiteren, auf EM beruhenden Algorithmen, die - bis auf OSEM - alle nur kurz angeschnitten werden. Die zweite, hier behandelte iterative Methode beruht auf der Projektion auf affine Unterräume. Diese Methoden (ART) stellen wir ebenfalls ausführlich vor und vergleichen sie schließlich mit OSEM. Zum Ende des Kapitels wird noch einmal in einem kurzen Überblick auf die Vor- und Nachteile der iterativen Methoden eingegangen und dargelegt, welche Lösungsansätze bisher für gewissen Probleme, im Zusammenhang mit diesen Methoden, bereitgestellt wurden.

Kapitel 6 ist nun schließlich ein kurzer Ausblick, der zuerst einmal Anregungen dazu geben soll, wie man die in dieser Arbeit vorgestellten Verfahren noch verbessern kann und der zum anderen die Gegenstände der aktuellen Forschung beschreibt.

Vorwort

An dieser Stelle möchte ich mich ganz herzlich bei Herrn Prof. Dr. A. K. Louis für die interessante Themenstellung und umfassende Betreuung bedanken.

Mein besonderer Dank gilt auch Dr. Peter Jonas, der sich ebenfalls viel Zeit für meine Betreuung genommen hat und mir mit seinen Tipps, Anregungen und Verbesserungsvorschlägen stets eine große Hilfe war.

Ebenso ein Dankeschön an Thomas Weber und Sebastian Ziaja, die mir eine Einführung in LATEX gegeben haben bzw. geduldigst meine Fragen zu auftretenden Problemen in Bezug auf LATEX beantwortet haben.

Ein Wort des Dankes geht auch an Herrn Mark Michel, der mir das Scannen und die Bearbeitung der Abbildungen dieser Arbeit ermöglichte und an Nadine Brozewski, die sich geduldigst meiner Kommasetzung und Rechtschreibung angenommen hat.

Schließlich möchte ich noch meinen Eltern, für die immense Geduld und Unterstützung während meiner Zeit an der Universität danken, meiner Schwester Sandra, die sich meiner Rückenschmerzen während der vielen Zeit des Sitzen angenommen hat und auch Herrn Thorsten Schlegel, der mir immer wieder genügend Motivation, Hilfe und Verständnis während meines gesamten Studiums entgegengebracht hat.

1 Einführung in die PET

1.1 Ein medizinisch-geschichtlicher Überblick

”Zahlreiche medizinische Bildgebungsverfahren beruhen auf der Rekonstruktion eines Bildes aus einem Satz Messungen. Zu solchen Techniken gehören u.a. die Röntgen-Computer- Tomographie (CT), die Ultraschallbildgebung und die nukleare Emissionstomographie.

Emissionstomographie ist ein Terminus, der auf die nuklearmedizinische Technik angewandt wird. Unter Nuklearmedizin verstehen wir die Verwendung von offenen, radioaktiven Stoffen zur Diagnostik und Therapie. Wir unterscheiden 2 Arten/Teilgebiete der Emissionstomographie: SPECT (Single-Photonen-Emissions-Tomographie) und PET (Positronen-Emissions-Tomographie). Während bei der SPECT Radiopharmaka zum Einsatz kommen, die sich analog zu den Stoffwechselsubstraten verhalten, verwenden wir bei der PET Radionuklide, die Teil des Metabolismus (Stoffwechselveränderung des Gewebes bzw. der Zelle) des Körpers werden, so dass Stoffwechselvorgänge vollständig erfasst und quantifiziert werden können.

Die herkömmliche Röntgentechnik ist eine sehr wertvolle Technik, die Bilder von hoher räumlicher Auflösung liefert, wobei aber der Weichgewebekontrast arm und die Strahlenbelastung für den Patienten nicht von der Hand zu weisen ist.

Die Nuklearmedizin liefert sowohl metabolische als auch anatomische Informationen. Ihre Nachteile schließen schlechte räumliche Auflösung, Signalverlust durch distale Bereiche dicker Organe und Abhängigkeit von radioaktiven Stoffen ein.” [4]

1.1.1 Die Geschichte der PET von den Anfängen bis zum ersten, industriereifen Tomographen

Im Folgenden soll ein kurzer zeitlicher Überblick über die Entdeckung der Existenz des Positrons bis zur heutigen Anwendung der PET in der klinischen Routine gegeben werden. Dieser Überblick entstammt aus [9] und [5].

1928: P.A.M. Dirac sagt die Existenz der Positronen voraus.

1932: C.D. Anderson entdeckt Positronen in durch kosmische Strahlung ausgelösten Kernprozessen.

1959: Anger und Rosenthal stellen die erste Kamera vor, die für Positronenstrahler nutzbar ist. Es handelt sich um zwei gegenüberliegende Gammakameras, die in Koinzidenz (genauere Definition des Begriffes siehe Kapitel 1.2) geschaltet sind.

1963: Cormack beschreibt die theoretischen Grundsätze für die Rekonstruktion.

1962: Rankowitz et al. nutzen als erste die ringförmige Detektoranordnung.

1975: Der erste industriereife Positronen-Emissions-Tomograph wird vorgestellt.

Seit **Ende der 80er** Jahre hat die PET in der Hirn- und Herzdiagnostik und später in der Onkologie zunehmende Bedeutung erlangt.

Seit **Mitte der 90er** Jahre gewann die PET auch zunehmende Bedeutung bei der orthopädischen Fragestellungen, wie z. B. der Entzündung des Knochens.

1.1.2 PET in der Onkologie

Der nun folgende kurze, aktuelle, medizinische Abriss ist eine Zusammenfassung aus den Werken [1], [2] und [3]:

Der wichtigste Einsatzbereich der PET liegt heute in der onkologischen Diagnostik, Therapie- und Verlaufskontrolle. Dabei spielt die morphologische Bildgebung (Morphologie = Lehre von der Körperform/Körperstruktur) eine entscheidende Rolle.

Die morphologische Bildgebung, z. B. CT, ist die Basis der meisten etablierten Staging-Methoden. Dabei handelt es sich um Bestandsaufnahmefethoden, die zur Bestimmung der Ausdehnung eines malignen (= bösartigen) Tumors durchgeführt werden. Dieses Staging dient nicht nur als Grundlage für therapeutische Entscheidungen, sondern wird auch zur Risikostratifizierung genutzt. In den 90er Jahren hat PET als metabolische Bildgebungsmethode eine zunehmende Bedeutung in der klinischen Diagnostik von Tumorerkrankungen erlangt.

Onkologische PET basiert auf dem von Warburg entdeckten Prinzip, dass Tumore hypermetabolisch sind und damit durch einen erhöhten Glukosestoffwechsel von nicht-malignem Gewebe unterschieden werden können. Mittels radioaktiv markierter Glucose [F-18-FDG] kann somit der Glucosstoffwechsel im Patienten in vivo verfolgt und quantifiziert werden.

Der klinische Nutzen von PET ist zum größten Teil auf die funktionellen Eigenschaften von FDG zurückzuführen. Dazu zählt unter anderem die Tatsache, dass das Ausmaß der Stoffwechselaktivität - und somit die FDG-Menge, die sich in einem Tumor (= Schwellung, kann gut- oder bösartig sein) anreichert - als Hinweis auf den Malignitätsgrad der Neubildung gilt. Des Weiteren konnte gezeigt werden, dass sich der Tumor-Metabolismus bereits verändern kann, bevor Monate später eine Größen- und Formveränderung des Tumors in der CT dargestellt werden kann. Da eine erhöhte Stoffwechselaktivität also vor anatomischen Veränderungen auftritt und viele Monate früher nachweisbar ist, kann mittels PET eine frühere Diagnose und ein exaktes Staging ermöglicht werden. Die erhaltenen Daten ermöglichen dann eine genaue Strahlentherapie-Planung und helfen, die Auswirkungen verschiedener Behandlungsformen präzise zu beurteilen. Insbesondere lässt sich anhand von PET-Verlaufskontrollen die Wirksamkeit einer Tumorbehandlung bei Einsatz von Chemo- oder Strahlentherapie feststellen.

Mit PET-Untersuchungen kann des Weiteren ein Rezidiv (= wiederkehrender Tumor) nachgewiesen werden. FDG-PET-Aufnahmen zeigen ein solches als heiße Knoten. Zu-

sätzlich eignen sich FDG-PET-Untersuchungen auch dazu, frühzeitig ein Ansprechen auf die Therapie zu erfassen oder die Behandlungsmethode rechtzeitig zu ändern.

Zusammenfassend kann man sagen, dass die PET-Bildgebung umfassende und exakte Staging-Informationen liefern kann, die durch CT oder MRT (Magnetresonanztomographie) nicht bereitgestellt werden können. Damit bietet das Verfahren die Möglichkeit, nicht nur den Behandlungsaufwand für Krebspatienten, sondern auch den Kostenaufwand im Gesundheitswesen zu reduzieren, da PET direkt Aufschluss über die bestmögliche Therapie gibt.

1.2 Physikalische Grundlagen der PET

Bei diesem Abschnitt folgen wir den Quellen [2], [6], [8], [9] und [17].

Der Arzt verabreicht dem Patienten eine positronenemittierende Substanz, die sich in den verschiedenen Regionen des Körpers anreichert. Die jeweilige Menge, die dabei angelagert wird, ist proportional zur Glucoseaufnahme an dieser Stelle. Wenn es uns daher möglich wäre, den Ort jeder Positronenemission zu bestimmen, könnten wir Aussagen über den Glucoseverbrauch der jeweiligen Region machen. Obwohl es uns unmöglich ist, den exakten Ort einer Positronenemission zu identifizieren, ist es uns dennoch möglich - durch die Positionierung von Detektoren um den Patienten herum - ein Volumen zu bestimmen, in welchem die Emission stattgefunden hat. Die Physik dahinter ist die folgende:

Wie wir gesehen haben, werden bei der PET Radionuklide verwendet, die beim Zerfall Positronen aussenden. Positronen werden auch als Antimaterieteilchen bezeichnet, d. h. es handelt sich hierbei um Elementarteilchen, entsprechend dem Elektron, mit gleicher Ruhemasse aber entgegengesetzter positiver Ladung.

	Ruhemasse	Ladung
Elektron	$0.9109 \cdot 10^{-27} g$	$-1.602 \cdot 10^{-19}$
Positron	$0.9109 \cdot 10^{-27} g$	$1.602 \cdot 10^{-19}$

Tabelle 1.1: Ruhemasse und Ladung von Elektron und Positron. Quelle: [9] S. 104.

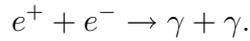
Positronen entstehen nur, wenn instabile, protonenreiche Kerne durch β^+ -Zerfall in einen stabileren Energiezustand übergehen, wobei die Protonen des Kerns sich unter Positronenaussendung in Neutronen umwandeln:



Die Energiemenge eines Positrons beträgt nur ein paar MeV und kann schnell wieder durch Ionisierung und Kollision mit Atomen im umgebenden Gewebe verlorengehen. Eine

1 Einführung in die PET

Begegnung eines Positrons mit einem freien Elektron im Gewebe führt zu einer Teilchen-Antiteilchen-Vernichtung aus der zwei Photonen (Gammaquanten) hervorgehen. Dieser Vorgang wird *Paarvernichtung* oder *Annihilation* genannt. Die bei der Paarvernichtung emittierte Strahlung wird als *Annihilationsstrahlung* oder *Vernichtungsstrahlung* bezeichnet:



Die Wahrscheinlichkeit für eine Reaktion ist allerdings nur dann groß, wenn das Positron in Ruhe oder nahezu in Ruhe ist. In der Reaktion müssen zwei Photonen freigesetzt werden, die in entgegengesetzte Richtung auseinanderfliegen, damit die Impulserhaltung gewährleistet ist. Aus der Energieerhaltung folgt, dass die Energie jedes dieser Photonen 511 keV beträgt. Um die Vernichtungsstrahlung nachzuweisen sind Detektoren notwendig, die sich gegenüberstehen und in der Lage sind, die beiden entstandenen Photonen gleichzeitig nachzuweisen.

Die gleichzeitige Emission von zwei Photonen ist die Basis koinzidenter Detektion und Bildgebung.

Unter *Koinzidenz* versteht man dabei das zeitliche Zusammentreffen zweier Vorgänge, sprich das gleichzeitige Ansprechen zweier sich gegenüberliegender Strahlungsdetektoren auf jeweils einen Gammaquanten. Die Breite des *Koinzidenzfensters* legt das Zeitintervall fest, innerhalb dessen zwei Gammaquanten auftreten müssen, um als gleichzeitig registriert zu gelten. Eine *echte Koinzidenz* liegt dann vor, wenn zwei Photonen aus derselben Positronen-Elektronen-Annihilation ohne Wechselwirkungen zu den Detektoren gelangen und beide innerhalb eines kurzen Zeitfensters (normalerweise 6-12 ns) detektiert werden.

Werden zwei Photonen durch eine Koinzidenzabfrage nachgewiesen, so bezeichnen wir das Volumen, das durch das Detektorpaar, das die Koinzidenz zählt, festgelegt wird, als *Detektortube* oder einfach *Tube*.

Die einzige Information, die wir also erhalten, wenn ein Paar Detektoren eine Koinzidenz zählt, ist die, dass die Annihilation irgendwo innerhalb des Tubes der entsprechenden Detektorelemente stattgefunden hat. Die Menge der Daten, die wir während eines PET-Scans gesammelt haben, sind als Tubezählungen n_1, \dots, n_I , wobei n_i die gesamte Anzahl der Koinzidenzen, die in Detektortube i gezählt wurden, ist und I die Anzahl der Tubes ist.

Die Tubes spielen bei dem Modell in Kapitel 3 eine Rolle. Wir können jedoch auch hingehen - wie bei dem Modell in Kapitel 4 - und anstelle des Volumens die Verbindungslinie zwischen den Wechselwirkungspunkten in den beiden gegenüberliegenden Detektorelementen betrachten. Diese Verbindungslinie bezeichnen wir mit *line of response*, kurz: LOR (siehe Abbildung 1.1).

Wir sollten uns jedoch vor Augen halten, dass die Gesamtanzahl der Koinzidenzen normalerweise viel kleiner als die Gesamtanzahl der tatsächlichen Emissionen ist. Dieses Phänomen ist auf gewisse Effekte wie Streuung, Dämpfung etc. zurückzuführen und wird in Abschnitt 1.3 noch ausführlicher besprochen werden.

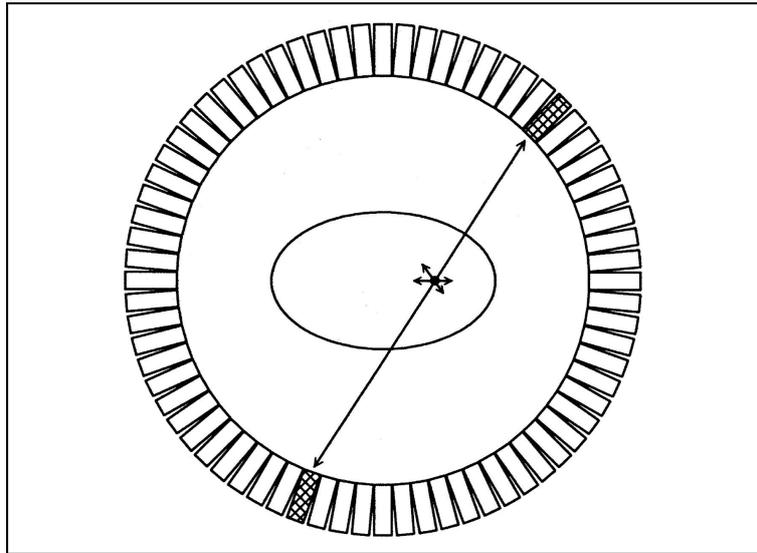


Abbildung 1.1: Koinzidentes Ereignis, das in einem Ring-PET-Scanner detektiert wird mit zugehörigem LOR. Quelle: [8] Figure 2.

Am Ende dieses Abschnitts wollen wir noch festhalten, dass der Aktionsradius des Positrons im Gewebe - gemessen vom Punkt der Emission bis zum Punkt der Vernichtung durch ein Elektron - in Wirklichkeit ein paar Millimeter betragen kann und von der Energie der Emission aus dem Kern abhängig ist. Es kann auch zu Abweichungen vom 180° Winkel kommen, wenn das Positronen-Elektronen-System einen Restimpuls hat, welcher im Vernichtungsprozess enthalten sein muss. Die Photonenpaare werden aber - innerhalb eines festen Winkels - ohne eine bevorzugte Richtung emittiert.

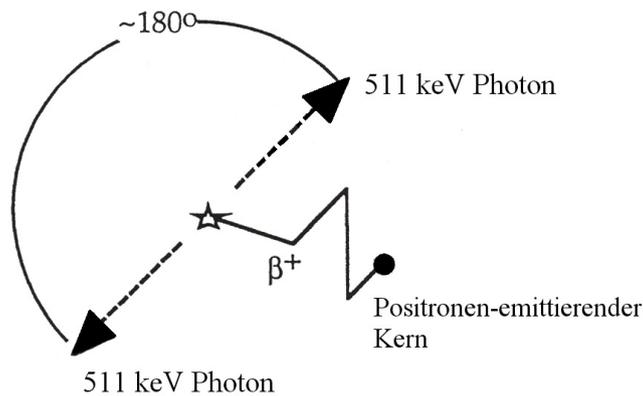


Abbildung 1.2: Schema der Positronen Emission und Annihilation. Quelle: [6] S. 1.

Sobald also der Isotopenindikator injiziert wurde und die zur Konzentration in den interessierenden Organen erlaubte Zeit verstrichen ist, können wir - wie wir im Laufe dieser Arbeit noch sehen werden - die Verteilung des Indikators innerhalb der Organe durch Sammlung koinzidenter Photonenpaare rekonstruieren. Dazu müssen wir aber im Voraus zwei Annahmen treffen, die in der gesamten Arbeit Gültigkeit haben:

1. die Linie, entlang derer die Photonen emittiert werden, enthält den Ort des Kerns, von dem aus das Positron entsteht (d. h. der Aktionsradius des Positrons im Gewebe ist Null).
2. die Photonen, die durch Vernichtung hervorgehen, werden in einem Winkel von genau 180° in entgegengesetzte Richtung ausgesendet.

Die verschiedenen Detektoren sind nun über eine Koinzidenzschaltung miteinander gekoppelt, d. h. erzeugt ein Vernichtungsquant in einem Detektor eine Szintillation, dann erfolgt innerhalb des gewählten Zeitintervalls an allen Detektoren eines Ringes eine Abfrage, ob in einer gegenüberliegenden Winkelposition ebenfalls ein Ereignis registriert wurde. So dient die Koinzidenzmessung zum einen der Lokalisierung der Ereignisse, in diesem Falle spricht man von *elektronischer Kollimierung*. Der Entstehungsort der 511 keV-Strahlung liegt nun - je nachdem welches Modell wir zugrundelegen - entweder auf einer Linie, welche die Wechselwirkungspunkte in den beiden Detektoren verbindet (siehe Abbildung 1.1), oder aber innerhalb des entsprechenden Tubes. Wenn in zwei gegenüberliegenden Detektoren zeitgleich beide Quanten gemessen werden, so ist das ein Nachweis dafür, dass eine Positronenemission stattgefunden hat. Mit Hilfe von Rekonstruktionsalgorithmen, die der Gegenstand dieser Arbeit sein werden, wollen wir nun die Verteilung des Kontrastmittels im Körper des Patienten bestimmen, um damit eine Aussage über den Glucosestoffwechsel in den verschiedenen Regionen des Körpers zu machen.

1.3 Gütekriterien

Die Qualität des von einem PET-System produzierten Bildes wird durch mehrere physikalische Faktoren beeinträchtigt, wobei einige dieser Faktoren korrigiert werden können. Bei der kurzen Darstellung dieser Kriterien richten wir uns nach [8].

1.3.1 Streuung

Wird mindestens eines der beiden Photonen aus einer Annihilation von seinem ursprünglichen Weg abgelenkt und werden trotzdem beide Photonen detektiert, so nennt man diese Koinzidenz eine *Streukoinzidenz*. Die Folge einer Streukoinzidenz ist, dass sich der Ort der Annihilation nun nicht mehr auf der Linie befindet, die die beiden Detektoren,

an denen die Photonen detektiert wurden, verbindet.

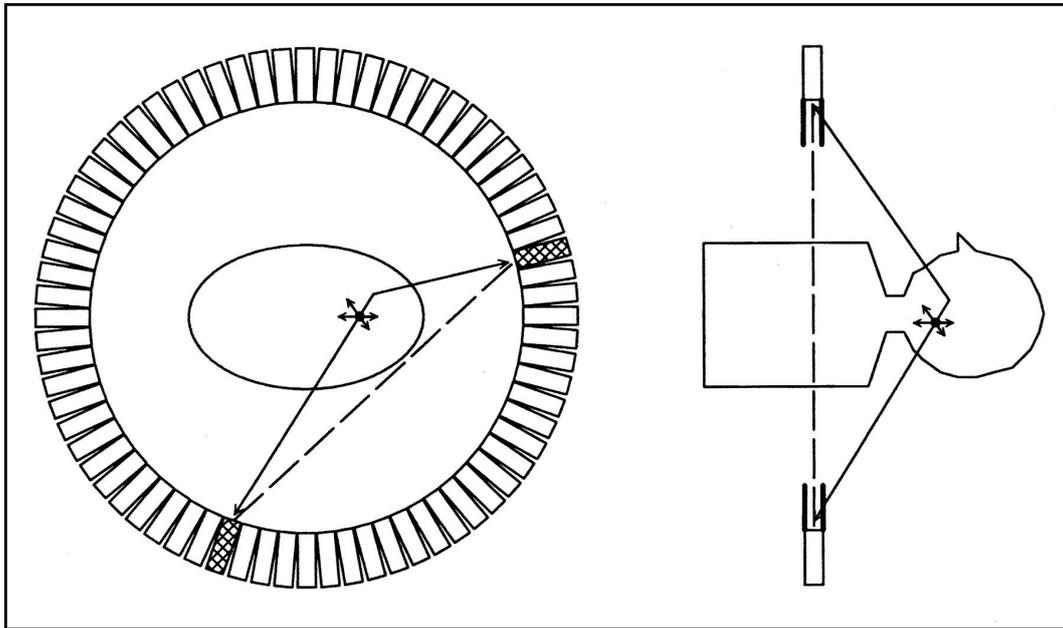


Abbildung 1.3: Streuung. Das linke Bild zeigt Streuung innerhalb der Ebene, das rechte Bild Streuung außerhalb der Ebene. Quelle: [8] Figure 4.

Wie wir in der Abbildung 1.3 erkennen können, gibt es mehrere Arten der Streuung. Auf der linken Seite ist die Streuung innerhalb einer Ebene dargestellt. Wir erkennen, dass ein Photon aus dem Vernichtungsprozess den Körper des Patienten ungestreut verlässt, wobei das zweite Photon streut, bevor es den Körper verlässt. Betrachten wir die zugehörige Koinzidenz, so scheint es, dass die Quelle der radioaktiven Strahlung außerhalb des Körpers liegt.

Eine andere Möglichkeit der Streuung wird auf der rechten Seite der Abbildung gezeigt. In diesem Fall findet die Positronenemission außerhalb der Ebene der Detektorringe statt. Eines der Photonen aus dem Vernichtungsprozess emittiert direkt in Richtung Ring, das andere ist anfänglich von den Detektoren weg gerichtet, streut dann jedoch und bewegt sich ebenfalls in Richtung Detektoren. Hier liegt die Quelle der radioaktiven Strahlung außerhalb der Detektorringebene, scheint aber in der Ebene der Detektoren zu liegen.

Um die Streuung außerhalb der Ebene zu reduzieren, werden sogenannte Septen verwendet (siehe Abbildung 1.4). Diese blockieren nicht nur die radioaktive Strahlung, die außerhalb des FOV (field of view, kurz: FOV; Sichtfeld; Es handelt sich hierbei um die Fläche des Detektors, die zur Bildinformation beiträgt) des Ringes entsteht, sondern minimieren auch andere Effekte der radioaktiven Strahlung, die außerhalb des FOV entstehen, inklusive Totzeit und zufällige Koinzidenzen, die später noch diskutiert werden.

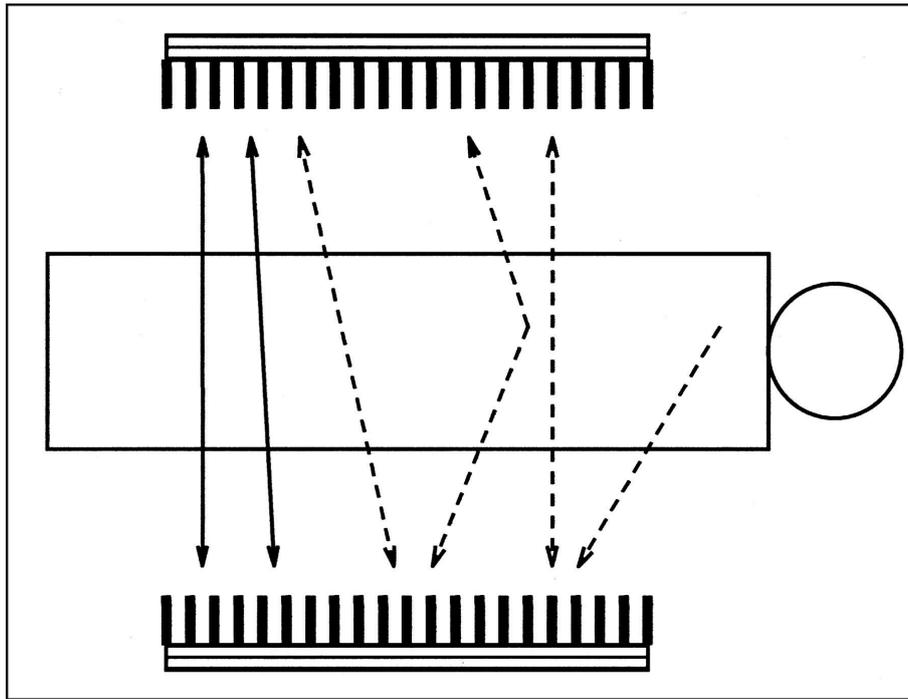


Abbildung 1.4: Der Gebrauch von Septen für die hybrid-PET-Bildgebung. Die durchgezogenen Linien stehen für detektierte Ereignisse, die gestrichelten Linien repräsentieren verschiedene Typen nichtdetektierter Ereignisse. Quelle: [8] Figure 14.

Insgesamt können wir sagen, dass Streuung entweder im Objekt, auf dem Weg Objekt-Detektor oder im Kristall erfolgen kann. Somit werden auch Photonen registriert, die unterhalb des Energiemaximums liegen. Dieser Effekt kann durch eine bessere Geometrie des Detektors und durch eine entsprechende Wahl des Energiefensters reduziert werden. Eine Implementierung eines Streu-Korrektur-Algorithmus, der vor der Rekonstruktion auf die Originaldaten angewendet wird, würde den Rahmen dieser Arbeit sprengen, kann aber in dem Artikel von "M. Bergstrom, L. Eriksson, C. Bohmm, G. Blomqvist, J. Litton, Correction for scattered radiation in a ring detector positron camera by integral transformation of the projections, J. Comput. Assist. Tomogr., 1983, Volume 7, pp. 42-50", nachgelesen werden.

1.3.2 Dämpfung

Dämpfung ist der Verlust wahrer Ereignisse, der auf Streuung und Absorption zurückzuführen ist.

Abbildung 1.5 zeigt ein Ereignis, bei dem die beiden Photonen in Richtung Detektoren emittiert werden, aber eines der beiden Photonen irgendwo auf seinem Weg durch den Körper gestoppt oder abgelenkt wird und somit nicht von dem vorgesehenen Detektor

detektiert wird. Dieses gestreute Photon kann, muss aber nicht in einem anderen Detektor als dem vorgesehenen detektiert werden.

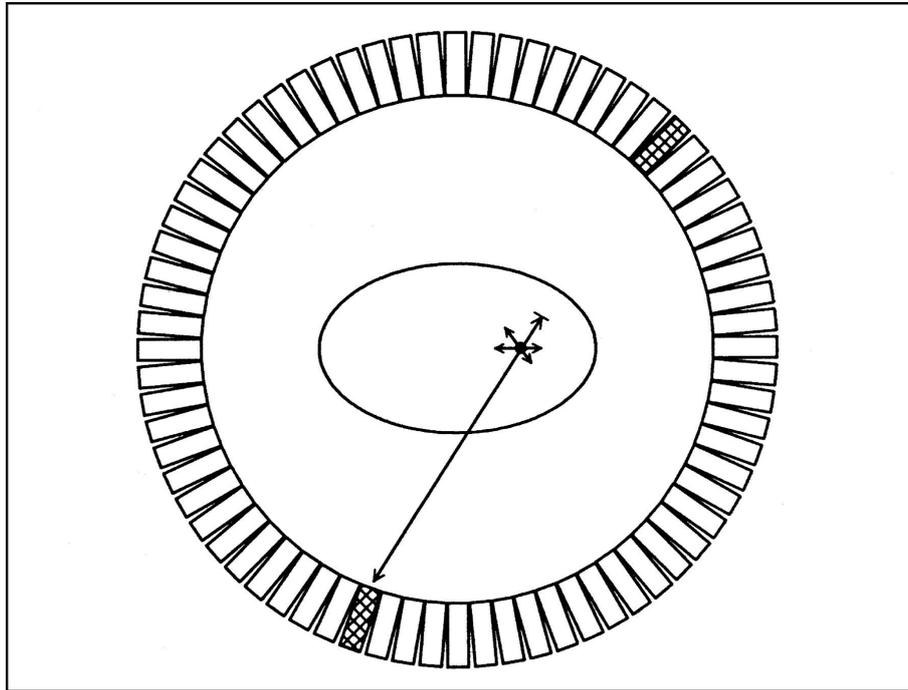


Abbildung 1.5: Dämpfung. Eines der Photonen wird gestoppt oder abgelenkt, bevor es detektiert wird. Quelle: [8] Figure 5.

In der PET müssen **beide** Photonen den Körper ungedämpft verlassen, damit es zu einer echten Koinzidenz kommt. Deshalb ist die Wahrscheinlichkeit, dass ein Ereignis in der PET gedämpft wird wesentlich höher, als in der SPECT.

Charakteristisch für die Dämpfung in der Koinzidenz-Bildgebung ist, dass in den meisten Fällen zumindest eines der emittierten Photonen eine wesentliche Menge an Gewebe durchqueren muss, sogar dann, wenn die Strahlung nahe an den Rändern des Körpers ist.

Der offenkundigste Effekt der Dämpfung ist der Verlust an Zählungen. Daraus resultieren eine Zunahme des Rauschens und eine ungenaue Quantifizierung der radioaktiven Verteilung. Die Rauscheffekte können nicht behoben werden, allerdings können wir durch die Anwendung einer Dämpfungskorrektur die quantitative Genauigkeit zurückbekommen.

Ein anderer Effekt der Dämpfung sind Unregelmäßigkeiten im rekonstruierten Bild. Es ist z. B. wesentlich wahrscheinlicher, dass die Strahlung, die von der Mitte des Körpers aus emittiert wird, gedämpft wird, als die Strahlung, die an den Rändern des Körpers emittiert wird. Das resultierende Bild zeigt daher eine künstlich verringerte Radioaktivität tiefer im Körper. Die Außenkonturen dagegen zeigen eine künstlich hohe Menge an Radioaktivität, weil die radioaktive Strahlung, die tangential zu den Außenkonturen

des Körpers emittiert, nicht gedämpft wird. Da alle restliche radioaktive Strahlung, die innerhalb des Körpers emittiert wird, in allen Richtungen gedämpft wird und keinen einfachen Weg aus dem Körper hinaus hat, wird eine helle Außenkontur beobachtet. Um die Dämpfung zu korrigieren, stehen uns die gemessene und die berechnete Korrektur zur Verfügung. Genaueres dazu werden wir jedoch erst im Ausblick auf Seite 179 ff angeben.

1.3.3 Zufällige Koinzidenzen

Eine *zufällige Koinzidenz* ist das Ergebnis einer Koinzidenzabfrage, bei der beide Photonen von verschiedenen Vernichtungsorten herrühren. Das Verhältnis echte Koinzidenz-zufällige Koinzidenz verbessert sich, je mehr Detektoren in Koinzidenz geschaltet sind. Auf der einen Seite muss das Zeitfenster groß genug sein, damit alle echten Ereignisse festgestellt werden, andererseits muss man aufpassen, denn je größer das Zeitfenster ist, umso mehr zufällige Koinzidenzen kommen hinzu.

Die Rate zufälliger Ereignissen zwischen zwei Detektoren ist:

$$R_R = r * R_1 * R_2,$$

wobei R_1 und R_2 die Raten sind, mit denen Detektor 1 bzw. Detektor 2 Gammastrahlen erhält und r das Zeitfenster ist.

Zufällige Koinzidenz-Zählungen addieren Hintergrundaktivität zum Bild. Zufällige Ereignisse werden dann bedeutsam (verglichen mit wahren Ereignissen), wenn die Detektorraten sehr hoch sind. Sie sind für Detektoren mit geringer Detektionseffizienz und für die dreidimensionale Bildgebung problematischer.

1.3.4 Totzeit

Unter *Totzeit* versteht man bei elektrischen und mechanischen Messsystemen die an der Übergangsfunktion des Systems zu ermittelnde Zeitspanne zwischen dem Wirkungseinsatz einer Sprungfunktion als Eingangssignal und der ersten messbaren Reaktion in der Antwortfunktion des Systems. Wenn die Rate der Photonen, die auf einen Detektor trifft wächst, dann wächst auch, zurückzuführend auf die Totzeit, die Wahrscheinlichkeit, dass Photonen verschwinden bzw. fehlen. Dieses Problem tritt besonders bei der Koinzidenz-Detektion auf, da beide Photonen detektiert werden müssen. Totzeit-Verluste werden von Systemen mit vielen unabhängigen Detektoren minimiert. Die Verluste werden auch bei der Verwendung schnellerer Szintillatoren reduziert.

1.3.5 Rauschen

Ein wichtiger Faktor bei allen nuklearmedizinischen Bildern ist das Rauschen. Bildrauschen (zufällige Variationen in der Pixelintensität) nimmt bei mehr Zählungen ab. Mehr Zählungen können wir durch längeres scannen, höhere Tracerdosen oder durch eine verbesserte Effizienz der Scanner beim Detektieren emittierter radioaktiver Strahlung er-

reichen.

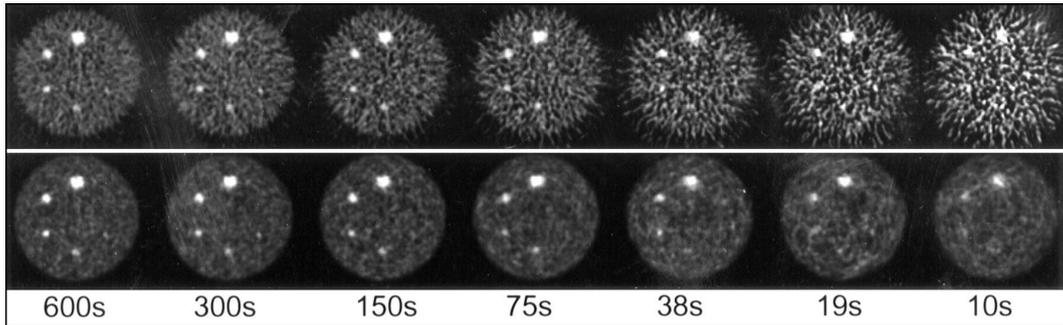


Abbildung 1.6: Bildqualität als eine Funktion von Zählungen. Das gleiche Phantom wird für verschiedene Zeiten, die ungefähr um den Faktor 2 wachsen, dargestellt. Obere Reihe: Bilder, die mit der gefilterten Rückprojektion rekonstruiert wurden. Untere Reihe: Bilder, die mit dem OS-Algorithmus rekonstruiert wurden. Eine genauere Beschreibung der gerade erwähnten Algorithmen erfolgt später. Quelle: [8] Figure 8.

Ein wichtiger Faktor in der Rauschqualität von Daten ist der Grad an Hintergrundaktivität. Die Zählungen, die entlang eines festen LORs während eines PET-Scans gemessen wurden, enthalten echte, zufällige und Streu-Koinzidenzen:

$$P = T + S + R.$$

Die echten Ereignisse erhält man, in dem man Streu- und Zufalls-Korrekturen auf die Gesamtereignisse anwendet:

$$T = P - S - R.$$

Die gesamten Zählungen, die während eines PET-Scans gemessen werden, werden mit P bezeichnet. Die Anzahl der Zählungen, die nach der Streu- und Zufalls-Korrektur verbleibt, wird mit T bezeichnet. Jedoch ist die Anzahl der echten Ereignisse T (nach den Korrekturen) kein adäquater Indikator für die spätere Bildqualität. Z. B. wurden bei einer Studie, in der 1 Millionen Gesamtzählungen ohne irgendeinen Hintergrund (keine Streuung und keine Zufalls-Ereignisse) gesammelt wurden, viel bessere Bilder erzielt als bei einer Studie, bei der 1,5 Millionen Gesamtzählungen gesammelt wurden, von denen 0,5 Millionen Hintergrund sind. Sogar nach der Korrektur des Hintergrunds wiesen die Bilder der ersten Studie eine bessere Qualität auf als die der zweiten. In der Tat sind sie sogar in der Qualität Bildern ähnlich, die mit 0.7 Millionen Gesamtzählungen ohne Hintergrund rekonstruiert wurden.

1 Einführung in die PET

Dämpfung, Streuung und zufällige Koinzidenzen mindern also die Qualität des produzierten Bildes. Trotzdem werden sie in den folgenden Kapiteln sowohl bei der mathematischen Modellierung als auch den daraus resultierenden Algorithmen zur Bestimmung der gesuchten Tracerverteilung oft gar nicht oder nur kurz erwähnt werden. Das liegt daran, dass es nicht das Thema dieser Arbeit sein wird, diese Faktoren ausführlicher zu besprechen, wir jedoch darauf hinweisen wollen, dass sie existieren und berücksichtigt werden müssen. Im Ausblick dieser Arbeit werden wir noch einmal kurz auf ihre Problematik eingehen und mögliche Lösungsvorschläge ansprechen. Eine genauere Ausführung davon könnte zur Themenstellung einer weiteren Diplomarbeit führen.

2 Mathematische Grundlagen

In diesem Kapitel wollen wir Sätze, Lemmata und Definitionen aus den Bereichen Funktionalanalysis, Bildrekonstruktion, Wahrscheinlichkeitstheorie und Optimierung zur Verfügung stellen, die als Hilfsmittel für das Verständnis und die Durchführung von Beweisen in den nachfolgenden Kapiteln dienen werden.

2.1 Grundlagen aus der Funktionalanalysis

Alle in diesem Abschnitt behandelten Sätze, Korollare und Lemmata werden ohne Beweise angegeben, da diese in nahezu jedem Funktionalanalysisbuch nachgeschlagen werden können. Interessierte vermögen dies z. B. ausführlichst in [24] und [39] nachzulesen.

Definition 2.1. (Hilbertraum)

Sei H ein Vektorraum (VR) über $k = \mathbb{R}$ (\mathbb{C}) und $\langle \cdot, \cdot \rangle : H \times H \rightarrow \mathbb{R}$ ein Skalarprodukt auf H .

Ist H bzgl. der induzierten Norm $\| \cdot \|_H = \langle \cdot, \cdot \rangle_H^{1/2}$ vollständig, so heisst $(H, \langle \cdot, \cdot \rangle_H)$ ein **Hilbertraum (HR)**.

Definition 2.2. (orthogonal, orthogonales Komplement)

Sei H ein HR und seien $x, y \in H$, $A \subset H$. Dann definiert man:

$$x, y \text{ heissen } \mathbf{orthogonal} \quad :\Leftrightarrow \langle x, y \rangle = 0 \quad (x \perp y)$$

und es ist

$$A^\perp := \{z \in H; z \perp a \quad \forall a \in A\}$$

das **orthogonale Komplement von A** .

Satz 2.3. (beste Approximation)

Sei H ein HR, $A \subset H$ abgeschlossen und konvex, $x \in H$.

Dann existiert ein eindeutiges $y \in A$ mit:

$$\| x - y \| = \mathit{dist}(x, A) = \inf_{z \in A} \| x - z \| .$$

Korollar 2.4. Sei H ein HR und $A \subset H$ ein linearer TR. Dann gilt:

- a) $\{0\}^\perp = H$,
- b) $H^\perp = \{0\}$,
- c) $\overline{A} = (A^\perp)^\perp$.

Satz 2.5. (Projektionssatz)

Sei H ein HR, $A \subset H$ abgeschlossen. Dann gilt:

jedes $x \in H$ lässt sich eindeutig in der Form $x = y + z$ schreiben, mit $y \in A$ und $z \in A^\perp$ ($H = A \oplus A^\perp$).

Dabei ist y die Bestapproximation von x in A .

Definition 2.6. (orthogonale Projektion)

Sei H ein HR und $A \subset H$ abgeschlossen.

Die Abbildung $P : H \rightarrow A$ mit $P(x) = y$, wobei y die Bestapproximation aus Satz 2.5 ist, heisst **orthogonale Projektion** von x auf A .

Satz 2.7. Sei $A \subset H$ ein abgeschlossener TR und $P : H \rightarrow A$ die orthogonale Projektion auf A . Dann gilt:

- a) $D(P) = H$, $R(P) = A$, $N(P) = A^\perp$,
- b) $\|P\|_{H \rightarrow H} = 1$ falls $A \neq \{0\}$,
- c) $P^2 = P$ und $P^* = P$,
- d) $R(P) = N(P)^\perp$,
- e) $I - P$ ist die orthogonale Projektion auf $R(P)^\perp = N(P) = A^\perp$.

Satz 2.8. Seien E, F normierte Räume, $T : E \rightarrow F$ k -linear. Äquivalent sind:

- i) T ist stetig,
- ii) T ist stetig in 0,
- iii) $\exists C > 0$ mit $\|Tx\|_F \leq C \|x\|_E \quad \forall x \in E$,
- iv) $\sup_{x \in \overline{B}_1(0)} \|Tx\|_F < \infty$.

In diesem Fall ist:

$$C := \sup_{x \in \overline{B_1(0)}} \|Tx\|_F =: \|T\| \text{ die kleinste Zahl } C \geq 0 \text{ mit: } \|Tx\|_F \leq C \|x\|_E \quad \forall x \in E.$$

Definition 2.9. Seien E, F normierte k -Vektorräume. Dann definiert man:

$$\mathcal{L}(E, F) := \{T; T : E \rightarrow F \text{ stetig linear}\}$$

und

$$\mathcal{K}(E, F) := \{T; T : E \rightarrow F \text{ linear und kompakt}\}.$$

Satz 2.10. Kompakte Operatoren sind stetig.

Satz 2.11. Sei $T \in \mathcal{L}(X, Y)$, X, Y HR'e über $k = \mathbb{R} (\mathbb{C})$. Dann gilt:

$$a) N(T^*) = R(T)^\perp \text{ und } N(T^*)^\perp = (R(T)^\perp)^\perp = \overline{R(T)},$$

$$b) N(T) = R(T^*)^\perp \text{ und } N(T)^\perp = (R(T^*)^\perp)^\perp = \overline{R(T^*)}.$$

Satz 2.12. (Prinzip der stetigen Inversen)

Seien X, Y Banachräume, $T \in \mathcal{L}(X, Y)$ bijektiv. Dann gilt:
 $T^{-1} \in \mathcal{L}(Y, X)$.

Definition 2.13. (adjungiert, selbstadjungiert, unitär)

Sei $T \in \mathcal{L}(X, Y)$, X, Y Hilberträume.

Der Operator $T^* : Y \rightarrow X$ mit

$$\langle T^*y, x \rangle_X = \langle y, Tx \rangle_Y \quad \forall x \in X, \forall y \in Y$$

heißt der zu T **adjungierte Operator**.

Gilt $T = T^*$, so heißt T **selbstadjungiert**.

Ist $TT^* = Id_Y$, $T^*T = Id_X$, so heißt T **unitär**.

Satz 2.14. Sei $T \in \mathcal{L}(X, Y)$, X, Y Hilberträume.

Wenn $T : X \rightarrow Y$ surjektiv ist, dann folgt:

Der adjungierte Operator $T^* : Y \rightarrow X$ ist injektiv und $R(T^*) \subset X$ ist abgeschlossen.

2.2 Grundlagen aus der Bildrekonstruktion

Die in diesem Abschnitt vorgestellten Definitionen, Sätze und Lemmata inklusive Beweisen entstammen aus [23] und [40].

2.2.1 Die Fouriertransformation

Definition 2.15. (*Fouriertransformierte, Fouriertransformation*)

Sei $f \in L_1(\mathbb{R}^n)$. Setze

$$(\mathcal{F}f)(\xi) := \hat{f}_n(\xi) := (2\pi)^{-n/2} \int_{\mathbb{R}^n} f(x) e^{-i\langle x, \xi \rangle} dx \quad \forall \xi \in \mathbb{R}^n. \quad (2.1)$$

Die Funktion \hat{f} heißt **Fouriertransformierte** von f , die Abbildung $\mathcal{F}f \mapsto \mathcal{F}f$ **Fouriertransformation**.

Definition 2.16. (*Faltung*)

Sei $f \in L_2(\mathbb{R}^n)$, $g \in L_1(\mathbb{R}^n)$ oder umgekehrt.

$$(f * g) := \int_{\mathbb{R}^n} f(x - y)g(y) dy \quad (2.2)$$

heißt **Faltung** der Funktionen f und g .

Satz 2.17. (*Faltungssatz*)

Für $f \in L_2(\mathbb{R}^n)$, $g \in L_1(\mathbb{R}^n)$ bzw. $f \in L_1(\mathbb{R}^n)$, $g \in L_2(\mathbb{R}^n)$ gilt:

$$(\mathcal{F}(f * g))(\xi) = (2\pi)^{n/2} (\mathcal{F}f)(\xi) (\mathcal{F}g)(\xi). \quad (2.3)$$

Beweis.

$$\begin{aligned} (\mathcal{F}(f * g))(\xi) &\stackrel{(2.1)}{=} (2\pi)^{-n/2} \int_{\mathbb{R}^n} (f * g)(x) e^{-i\langle x, \xi \rangle} dx \\ &\stackrel{(2.2)}{=} (2\pi)^{-n/2} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(x - y)g(y) e^{-i\langle x, \xi \rangle} dy dx \end{aligned}$$

Wir substituieren $x = y + z$ und erhalten somit

$$\begin{aligned} (\mathcal{F}(f * g))(\xi) &= (2\pi)^{-n/2} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(z)g(y)e^{-i\langle y, \xi \rangle} e^{-i\langle z, \xi \rangle} dy dz \\ &= (2\pi)^{-n/2} \int_{\mathbb{R}^n} f(z)e^{-i\langle z, \xi \rangle} \int_{\mathbb{R}^n} g(y)e^{-i\langle y, \xi \rangle} dy dz \\ &= (2\pi)^{n/2} (\mathcal{F}f)(\xi) (\mathcal{F}g)(\xi). \end{aligned}$$

□

Aus funktionalanalytischer Sicht ist es günstig, \mathcal{F} auf einen geeigneten Teilraum von $L_1(\mathbb{R}^n)$, der aus glatten Funktionen besteht, einzuschränken. Optimal ist dabei die Wahl des als nächstes definierten Schwarzraumes $\mathcal{S}(\mathbb{R}^n)$.

Definition 2.18. (Schwarzraum, Schwarzfunktionen)

Eine Funktion $f : \mathbb{R}^n \rightarrow \mathbb{C}$ heißt *schnell fallend*, falls

$$\lim_{|x| \rightarrow \infty} x^\alpha f(x) = 0 \quad \forall \alpha \in \mathbb{N}_0^n.$$

Hier ist x^α durch $x_1^{\alpha_1} \cdot \dots \cdot x_n^{\alpha_n}$ erklärt (Multiindex). Der Raum

$$\mathcal{S}(\mathbb{R}^n) = \{f \in C^\infty(\mathbb{R}^n); d^\beta f \text{ schnell fallend } \forall \beta \in \mathbb{N}_0^n\}$$

heißt *Schwarzraum*, seine Elemente *Schwarzfunktionen*.

Nach der Definition verschwinden Schwarzfunktionen und ihre sämtlichen Ableitungen im Unendlichen schneller als das Reziproke jeden Polynoms. Daher gilt:

$$\mathcal{S}(\mathbb{R}^n) \subset L_p(\mathbb{R}^n) \quad \forall p \geq 1$$

und man kann sogar zeigen, dass $\mathcal{S}(\mathbb{R}^n)$ dicht in $L_p(\mathbb{R}^n)$ für $1 \leq p \leq \infty$ liegt.

Die Bedeutung des Schwarzraumes liegt aber vor allem darin, dass die Fouriertransformation \mathcal{F} eine Bijektion von $\mathcal{S}(\mathbb{R}^n)$ auf $\mathcal{S}(\mathbb{R}^n)$ vermittelt, was für $L_1(\mathbb{R}^n)$ nicht der Fall ist.

Zunächst wollen wir aber noch ein paar wichtige Eigenschaften der Fouriertransformation angeben, bevor wir die obige Behauptung beweisen wollen.

Lemma 2.19. Sei $f \in \mathcal{S}(\mathbb{R}^n)$ und α ein Multiindex. Dann gilt:

a) $\mathcal{F}f \in C^\infty(\mathbb{R}^n)$ und $d^\alpha(\mathcal{F}f)(\xi) = (-i)^{|\alpha|} \mathcal{F}(x^\alpha f)(\xi)$.

b) $\mathcal{F}(d^\alpha f)(\xi) = i^{|\alpha|} \xi^\alpha (\mathcal{F}f)(\xi)$.

c) $\mathcal{F}(T^b f)(\xi) = E^b(\mathcal{F}f)(\xi)$ **Translation**

wobei $T^b f(x) = f(x - b)$ $b \in \mathbb{R}^n$ und $E^b(\mathcal{F}f)(\xi) = e^{-i\langle b, \xi \rangle}(\mathcal{F}f)(\xi)$ ist.

d) $\mathcal{F}(D^a f)(\xi) = D^{1/a}(\mathcal{F}f)(\xi)$ **Dilatation**

wobei $D^a f(x) = |a|^{-n/2} f(a^{-1}x)$ $a \in \mathbb{R} \setminus \{0\}$ ist.

Beweis. a) Sei $f \in \mathcal{S}(\mathbb{R}^n)$.

$$\text{Setze } h(\xi, x) := \underbrace{f(x)e^{-i\langle x, \xi \rangle}}_{\text{integrierbar}} \in C^\infty(\mathbb{R}^n)$$

$$\Rightarrow \frac{\partial^\alpha h}{\partial \xi^\alpha}(\xi, x) = f(x)(-i)^{|\alpha|} x^\alpha e^{-i\langle x, \xi \rangle} \quad \forall \alpha \in \mathbb{N}_0^n.$$

$$\text{Da } \left| \frac{\partial^\alpha h}{\partial \xi^\alpha}(\xi, x) \right| = |f(x)x^\alpha| \underbrace{|(-i)^{|\alpha|}|}_{=1} \underbrace{|e^{-i\langle x, \xi \rangle}|}_{=1} \text{ und } f \in \mathcal{S} \text{ folgt}$$

$$M := \sup |f(x)x^\alpha| < \infty \quad \forall \alpha \in \mathbb{N}_0^n$$

$$\Rightarrow \left| \frac{\partial^\alpha h}{\partial \xi^\alpha}(\xi, x) \right| \leq \underbrace{M}_{\text{Konstanten sind integrabel}} \quad \forall (\xi, x).$$

Mit Hilfe des Konvergenzsatzes von Lebesgue können wir nun also die Differentiation ins Integral ziehen und es gilt:

$$\frac{\partial^\alpha}{\partial \xi^\alpha} (2\pi)^{-n/2} \int_{\mathbb{R}^n} f(x)e^{-i\langle x, \xi \rangle} dx = (2\pi)^{-n/2} \int_{\mathbb{R}^n} f(x) \frac{\partial^\alpha}{\partial \xi^\alpha} e^{-i\langle x, \xi \rangle} dx. \quad (2.4)$$

Somit gilt also

$$\begin{aligned} d^\alpha(\mathcal{F}f)(\xi) &= \frac{\partial^\alpha}{\partial \xi^\alpha} (2\pi)^{-n/2} \int_{\mathbb{R}^n} f(x)e^{-i\langle x, \xi \rangle} dx \\ &\stackrel{(2.4)}{=} (2\pi)^{-n/2} \int_{\mathbb{R}^n} f(x) \frac{\partial^\alpha}{\partial \xi^\alpha} e^{-i\langle x, \xi \rangle} dx \\ &= (2\pi)^{-n/2} (-i)^{|\alpha|} \int_{\mathbb{R}^n} f(x)x^\alpha e^{-i\langle x, \xi \rangle} dx \\ &= (-i)^{|\alpha|} \mathcal{F}(x^\alpha f)(\xi) \quad \forall \alpha \in \mathbb{N}_0^n. \end{aligned}$$

Da dies für alle $\alpha \in \mathbb{N}_0^n$ gilt, ist $\mathcal{F} \in C^\infty(\mathbb{R}^n)$.

- b) Durch partielle Integration erhält man (wegen $x^\beta f \in \mathcal{S}(\mathbb{R}^n)$ verschwinden alle Randterme)

$$\begin{aligned}\mathcal{F}(d^\alpha f)(\xi) &= (2\pi)^{-n/2} \int_{\mathbb{R}^n} (d^\alpha f)(x) e^{-i\langle x, \xi \rangle} dx \\ &= (-1)^{|\alpha|} (2\pi)^{-n/2} \int_{\mathbb{R}^n} f(x) \frac{\partial^\alpha}{\partial x^\alpha} e^{-i\langle x, \xi \rangle} dx \\ &= i^{|\alpha|} \xi^\alpha (\mathcal{F}f)(\xi).\end{aligned}$$

c)

$$\begin{aligned}\mathcal{F}(T^b f)(\xi) &= (2\pi)^{-n/2} \int_{\mathbb{R}^n} T^b f(x) e^{-i\langle x, \xi \rangle} dx \\ &= (2\pi)^{-n/2} \int_{\mathbb{R}^n} f(x-b) e^{-i\langle x, \xi \rangle} dx.\end{aligned}$$

Substituiere $y = x - b$ und es folgt

$$\begin{aligned}\mathcal{F}(T^b f)(\xi) &= (2\pi)^{-n/2} e^{-i\langle b, \xi \rangle} \int_{\mathbb{R}^n} f(y) e^{-i\langle y, \xi \rangle} dy \\ &= E^b(\mathcal{F}f)(\xi).\end{aligned}$$

d)

$$\begin{aligned}\mathcal{F}(D^a f)(\xi) &= (2\pi)^{-n/2} \int_{\mathbb{R}^n} (D^a f)(x) e^{-i\langle x, \xi \rangle} dx \\ &= (2\pi)^{-n/2} \int_{\mathbb{R}^n} |a|^{-n/2} f(x/a) e^{-i\langle x, \xi \rangle} dx.\end{aligned}$$

Substituiere $y = x/a$ und man erhält $dx = |a|^n dy$. Somit folgt

$$\begin{aligned}\mathcal{F}(D^a f)(\xi) &= (2\pi)^{-n/2} |a|^{n/2} \int_{\mathbb{R}^n} f(y) e^{-ia\langle y, \xi \rangle} dy \\ &= D^{1/a}(\mathcal{F}f)(\xi).\end{aligned}$$

□

Lemma 2.20. Wenn $f \in \mathcal{S}(\mathbb{R}^n)$ ist, dann ist auch $\mathcal{F}f \in \mathcal{S}(\mathbb{R}^n)$.

Beweis. Aus Lemma 2.19 a) folgt, dass $\mathcal{F}f \in C^\infty(\mathbb{R}^n)$ ist.

noch zu zeigen: $\lim_{|\xi| \rightarrow \infty} \xi^\alpha d^\beta(\mathcal{F}f)(\xi) = 0$

Es gilt:

$$\begin{aligned} \xi^\alpha d^\beta(\mathcal{F}f)(\xi) &\stackrel{2.19a)}{=} \xi^\alpha (-i)^{|\beta|} \mathcal{F}(x^\beta f)(\xi) \underbrace{(-1)^{|\alpha|} (-i)^{|\alpha|} (-i)^{|\alpha|}}_{=1} \\ &= \xi^\alpha (-i)^{|\alpha|} (-1)^{|\alpha|} \mathcal{F}(x^\beta f)(\xi) (-i)^{|\beta|} (-i)^{|\alpha|} \\ &\stackrel{2.19b)}{=} \mathcal{F}(d^\alpha x^\beta f)(\xi) (-i)^{|\beta|} (-i)^{|\alpha|}. \end{aligned}$$

Aus $f \in \mathcal{S}(\mathbb{R}^n)$ folgt

$$x^\beta f(x) \in \mathcal{S}(\mathbb{R}^n) \text{ und } d^\alpha x^\beta f(x) \in \mathcal{S}(\mathbb{R}^n) \subset L_1(\mathbb{R}^n)$$

und damit sehen wir, dass

$$d^\alpha x^\beta f(x) \in L_1(\mathbb{R}^n)$$

ist. Da allgemein für ein $f \in L_1(\mathbb{R}^n)$ gilt:

$$\mathcal{F}f \in C_0(\mathbb{R}^n) := \{f \in C(\mathbb{R}^n); \lim_{|x| \rightarrow \infty} f(x) = 0\}$$

folgt

$$\mathcal{F}(d^\alpha x^\beta f)(\xi) \in C_0(\mathbb{R}^n),$$

damit erhalten wir

$$\lim_{|\xi| \rightarrow \infty} \mathcal{F}(d^\alpha x^\beta f)(\xi) = 0$$

und schließlich

$$\lim_{|\xi| \rightarrow \infty} \xi^\alpha d^\beta(\mathcal{F}f)(\xi) = 0.$$

□

Der Beweis zu dem folgenden Lemma wird nicht angegeben, da er sich ausführlich in [40] nachlesen lässt.

Lemma 2.21. Für $f \in \mathcal{S}(\mathbb{R}^n)$ gilt:

$$(\mathcal{F}\mathcal{F}f)(x) = f(-x) \quad \forall x \in \mathbb{R}^n. \quad (2.5)$$

Satz 2.22. Die Fouriertransformation ist eine Bijektion von $\mathcal{S}(\mathbb{R}^n)$ auf $\mathcal{S}(\mathbb{R}^n)$, der inverse Operator ist durch

$$(\mathcal{F}^{-1}f)(x) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} f(\xi) e^{i\langle x, \xi \rangle} d\xi \quad \forall x \in \mathbb{R}^n \quad (2.6)$$

gegeben. Ferner gilt:

$$\langle \mathcal{F}f, \mathcal{F}g \rangle_{L_2(\mathbb{R}^n)} = \langle f, g \rangle_{L_2(\mathbb{R}^n)} \quad \forall f, g \in \mathcal{S}(\mathbb{R}^n). \quad (2.7)$$

Insbesondere gilt:

$$\| \mathcal{F}f \|_{L_2(\mathbb{R}^n)} = \| f \|_{L_2(\mathbb{R}^n)} \quad \forall f \in \mathcal{S}(\mathbb{R}^n). \quad (2.8)$$

Beweis. Aus Lemma 2.21 folgt, dass $\mathcal{F}^4 = Id_{\mathcal{S}(\mathbb{R}^n)}$ ist. Daraus wiederum folgt, dass \mathcal{F} bijektiv ist und $\mathcal{F}^{-1} = \mathcal{F}^3$ ist. Das liefert uns:

$$(\mathcal{F}^{-1}f)(x) = (\mathcal{F}^2(\mathcal{F}f))(x) = (\mathcal{F}f)(-x).$$

Als unmittelbare Konsequenz des Satzes von Fubini gilt:

$$\int_{\mathbb{R}^n} (\mathcal{F}f)(\xi) \overline{(\mathcal{F}g)(\xi)} d\xi = \int_{\mathbb{R}^n} f(x) \overline{(\mathcal{F}(\mathcal{F}g))(x)} dx \quad \forall f, g \in \mathcal{S}(\mathbb{R}^n).$$

Für $h := \mathcal{F}g$ erhält man dann:

$$\begin{aligned} (\mathcal{F}\bar{h})(x) &= (2\pi)^{-n/2} \int_{\mathbb{R}^n} \overline{h(\xi)} e^{-i\langle x, \xi \rangle} d\xi \\ &= (2\pi)^{-n/2} \overline{\int_{\mathbb{R}^n} h(\xi) e^{i\langle x, \xi \rangle} d\xi} \\ &= \overline{(\mathcal{F}^{-1}h)(x)} \\ &= \overline{g(x)}. \end{aligned}$$

Damit erhält man also

$$\begin{aligned} \langle \mathcal{F}f, \mathcal{F}g \rangle_{L_2} &= \int_{\mathbb{R}^n} f(x) \overline{g(x)} dx \\ &= \langle f, g \rangle_{L_2} \end{aligned}$$

und insbesondere gilt:

$$\| \mathcal{F}f \|_{L_2} = \| f \|_{L_2} \quad \forall f \in \mathcal{S}(\mathbb{R}^n).$$

□

Der Operator \mathcal{F} ist also auf dem Teilraum $\mathcal{S}(\mathbb{R}^n)$ von $L_2(\mathbb{R}^n)$ wohldefiniert, bijektiv und bzgl. $\|\cdot\|_{L_2}$ isometrisch. Da $\mathcal{S}(\mathbb{R}^n)$ dicht in $L_2(\mathbb{R}^n)$ liegt, kann man \mathcal{F} zu einem isometrischen Operator auf $L_2(\mathbb{R}^n)$ fortsetzen. Man nennt diese Fortsetzung **Fourier-Plancherel-Transformation** und kennzeichnet sie vorübergehend mit \mathcal{F}_2 . Wegen Satz 2.22 ist $\mathcal{F}_2 : L_2(\mathbb{R}^n) \rightarrow L_2(\mathbb{R}^n)$ ein isometrischer Isomorphismus und es gilt die **Plancherel-Gleichung**

$$\langle \mathcal{F}_2 f, \mathcal{F}_2 g \rangle_{L_2} = \langle f, g \rangle_{L_2} \quad \forall f, g \in L_2(\mathbb{R}^n).$$

Wir sollten bemerken, dass $\mathcal{F}_2 f$ für $f \in L_2(\mathbb{R}^n)$ nicht durch (2.1) gegeben ist, denn für $f \in L_2(\mathbb{R}^n)$ braucht das Integral in (2.1) nicht zu existieren. (Ebenso ist \mathcal{F}_2 nach Konstruktion eine Äquivalenzklasse von Funktionen, während (2.1) wirklich eine Funktion definiert).

Im Folgenden wollen wir nur noch kurz den Zusammenhang von \mathcal{F} und \mathcal{F}_2 angeben, ohne auf den Beweis einzugehen, den man in [40] nachlesen kann.

Für $B_R := \{x \in \mathbb{R}^n : |x| \leq R\}$, $f \in L_2(\mathbb{R}^n)$ und

$$g_R(\xi) = (2\pi)^{-n/2} \int_{B_R} f(x) e^{-i\langle x, \xi \rangle} dx$$

gilt:

Satz 2.23. a) Für $f \in L_1(\mathbb{R}^n) \cap L_2(\mathbb{R}^n)$ gilt:

$$(\mathcal{F}_2 f)(\xi) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} f(x) e^{-i\langle x, \xi \rangle} dx \quad \text{fast überall.}$$

b) Für $f \in L_2(\mathbb{R}^n)$ gilt:

$$\mathcal{F}_2 f = \lim_{R \rightarrow \infty} g_R,$$

wobei die Konvergenz im Sinne von $\|\cdot\|_{L_2}$ vorliegt.

2.2.2 Die Radontransformation

Die Definitionen und Sätze dieses Abschnitts entstammen aus [11].

Definition 2.24. (Radontransformation)

Sei $C^n := \{(\omega, s); \omega \in S^{n-1}, s \in \mathbb{R}\}$, $n \geq 2$.

Dann heißt $R : \mathcal{S}(\mathbb{R}^n) \rightarrow \mathcal{S}(C^n)$, $f \mapsto Rf$ mit

$$(Rf)(\omega, s) := \int_{\mathbb{R}^n} f(x) \delta(s - x^\top \omega) dt \tag{2.9}$$

die **Radontransformation**,

wobei $\mathcal{S}(C^n) := \{g \in C^\infty(C^n); s^l \frac{\partial^k}{\partial s^k} g(\omega, s) \text{ beschränkt}; l, k = 0, 1, \dots\}$ ist.

Eine alternative Notation ist:

$$(Rf)(\omega, s) = \int_{\omega^\perp} f(s\omega + y) dy \quad (2.10)$$

$$\text{mit } \omega^\perp = \{x \in \mathbb{R}^n; \langle x, \omega \rangle = 0\}$$

und speziell für den Fall $n = 2$ schreiben wir:

$$(Rf)(\omega, s) = \int_{\mathbb{R}} f(s\omega + t\omega^\perp) dt. \quad (2.11)$$

Definition 2.25. (Faltung und Fouriertransformation)

Auf C^n sind Faltung und Fouriertransformation wie folgt definiert:

$$(\mathcal{F}g)(\omega, \sigma) := (2\pi)^{-1/2} \int_{\mathbb{R}} g(\omega, s) e^{-is\sigma} ds \text{ und} \quad (2.12)$$

$$(g * h)(\omega, s) := \int_{\mathbb{R}} g(\omega, s - t) h(\omega, t) dt \quad (2.13)$$

für $g, h \in \mathcal{S}(C^n)$.

Satz 2.26. (Projektionssatz)

Sei $f \in \mathcal{S}(\mathbb{R}^n)$, dann gilt:

$$(R_\omega f)^\wedge(\sigma) = (2\pi)^{(n-1)/2} \hat{f}(\sigma\omega), \quad \sigma \in \mathbb{R}. \quad (2.14)$$

Dabei entspricht die linke Seite der Gleichung einer 1D FT auf C^n und die rechte Seite einer FT im \mathbb{R}^n .

Beweis. Siehe [12] S. 11. □

Satz 2.27. (Rückprojektion)

Sei $R : \mathcal{S}(\mathbb{R}^n) \rightarrow \mathcal{S}(C^n)$, $g \in \mathcal{S}(C^n)$ mit $Rf = g$, dann ist der adjungierte Operator R^* gegeben durch:

$$(R^*g)(x) = \int_{S^{n-1}} g(\omega, x^\top \omega) d\omega \quad \text{Rückprojektion.} \quad (2.15)$$

Beweis. Siehe [12] S. 13. □

Satz 2.28. (Konsistenzbedingungen von Helgason, Ludwig, Gelfaud-Vilenkis)

Wenn $g \in \mathcal{S}$ im Bild von R ist, dann sind folgende Aussagen äquivalent:

i) $g(-\omega, -s) = g(\omega, s),$

ii) $\int s^m g(\omega, s) ds$ ist ein homogenes Polynom vom Grad $\leq m$ in ω .

Wenn zusätzlich $g(\omega, s) = 0 \quad \forall |s| \geq 1$ ist, dann gilt:

f hat Träger in $V(0, 1)$.

Beweis. Da der Beweis hierfür viel zu umfangreich und für die Arbeit irrelevant ist, verzichten wir auf ihn. Interessierte können ihn jedoch in der Originalarbeit von Helgason, Ludwig und Gelfaud nachlesen!

□

2.2.3 Die Röntgentransformation

Auch dieser Abschnitt entstammt aus [11].

Definition 2.29. (Röntgentransformation)

Sei $T^n := \{(\omega, a); \omega \in S^{n-1}, a \in \omega^\perp\}, n \geq 2$.

Dann heißt $P : \mathcal{S}(\mathbb{R}^n) \rightarrow \mathcal{S}(T^n), f \mapsto Pf$ mit

$$(Pf)(\omega, a) = \int_{\mathbb{R}} f(a + t\omega) dt \tag{2.16}$$

die **Röntgentransformation**,

wobei $\mathcal{S}(T^n) = \{g \in C^\infty(T^n); a^\alpha d^\beta g(\omega, a)$ beschränkt, $\alpha, \beta \geq 0\}$ ist.

Definition 2.30. (Faltung und Fouriertransformation)

Auf T^n sind Faltung und Fouriertransformation wie folgt definiert:

$$(\mathcal{F}g)(\omega, \xi) := (2\pi)^{(1-n)/2} \int_{\omega^\perp} g(\omega, a) e^{-i\langle a, \xi \rangle} da \quad \xi \in \omega^\perp \tag{2.17}$$

$$(g * h)(\omega, a) := \int_{\omega^\perp} g(a - y)h(y) dy \quad a \in \omega^\perp \tag{2.18}$$

für $g, h \in \mathcal{S}(T^n)$.

Satz 2.31. (Projektionssatz)

Sei $f \in \mathcal{S}(\mathbb{R}^n)$, dann gilt:

$$(P_\omega f)^\wedge(\eta) = (2\pi)^{1/2} \hat{f}(\eta) \quad \eta \in \omega^\perp. \tag{2.19}$$

Beweis. Siehe [12] S. 11. □

Satz 2.32. (Rückprojektion)

Sei $P : \mathcal{S}(\mathbb{R}^n) \rightarrow \mathcal{S}(T^n)$, $g \in \mathcal{S}(T^n)$, mit $Pf = g$, dann ist der adjungierte Operator P^* gegeben durch:

$$(P^*g)(a) = \int_{S^{n-1}} g(\omega, E_\omega a) d\omega \quad \text{Rückprojektion,} \quad (2.20)$$

wobei E_ω die orthogonale Projektion auf ω^\perp ist (z. B. $E_\omega a = a - (a^\top \omega)\omega$).

Beweis. Siehe [12] S. 13 und 14. □

2.3 Grundlagen aus der Statistik

Da die in Kapitel 5 vorgestellten Algorithmen auf statistischen Modellen und Methoden beruhen, auf die wir auch näher eingehen werden, wollen wir hier einen kurzen Überblick über die wichtigsten Grundbegriffe der Statistik geben, der sich an [25], [26] und [27] orientiert.

Definition 2.33. (Wahrscheinlichkeitsraum)

Seien Ω , \mathcal{F} Mengen, $P : \mathcal{F} \rightarrow R$ eine Abbildung.

Das Tripel (Ω, \mathcal{F}, P) heißt **Wahrscheinlichkeitsraum (WR)**, wenn gilt:

- $\Omega \neq \{\}$
- $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ mit
 - $\Omega \in \mathcal{F}$
 - $A \in \mathcal{F} \Rightarrow A^C \in \mathcal{F}$
 - $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^\infty A_i \in \mathcal{F}$

(d. h. \mathcal{F} ist σ -Algebra in Ω)

- $P : \mathcal{F} \rightarrow R$ mit

- $A \in \mathcal{F} \Rightarrow PA \geq 0$ (Nichtnegativität)
- $A_1, A_2, \dots \in \mathcal{F}$ mit $A_i \cap A_j = \{\}$ für $i \neq j$
 $\Rightarrow P \bigcup_{i=1}^{\infty} A_i = \sum_{i=1}^{\infty} PA_i$ (σ -Additivität)
- $P\Omega = 1$ (Normierung)

Dabei heißen Ω **Ergebnisraum**, **Stichprobenraum** oder **Grundgesamtheit**, \mathcal{F} **Ergebnisraum** ($\mathcal{F} \subseteq \mathcal{P}(\Omega)$), wobei die Potenzmenge oft zu dick ist und man deshalb $\mathcal{F} \subset \mathcal{P}(\Omega)$ nimmt) und P **Wahrscheinlichkeitsmaß**.

Definition 2.34. (*n*-dimensionale Zufallsvariable)

Sei (Ω, \mathcal{F}, P) ein WR.

Eine Abbildung $X : \Omega \rightarrow \mathbb{R}^n$ heißt ***n*-dimensionale Zufallsvariable** über (Ω, \mathcal{F}, P) wenn

$$(\forall B \in \mathcal{B}_n)(X^{-1}B \in \mathcal{F}) \quad (\text{Meßbarkeit})$$

ist. Dabei ist \mathcal{B}_n die Borelsche σ -Algebra auf \mathbb{R}^n (kleinste σ -Algebra auf \mathbb{R}^n , die alle offenen Mengen $U \subseteq \mathbb{R}^n$ enthält).

Anmerkung 2.35. Sei $X : \Omega \rightarrow \mathbb{R}^n$ eine *n*-dimensionale Zufallsvariable über einem WR (Ω, \mathcal{F}, P) . Es gilt:

$$X(\omega) = (X_1(\omega), \dots, X_n(\omega)) \quad \forall \omega \in \Omega,$$

wobei

$$X = (X_1, \dots, X_n)$$

ist.

Definition 2.36. (*gemeinsame Wahrscheinlichkeitsverteilung*)

Zu (Ω, \mathcal{F}, P) ist durch die \mathcal{F} - \mathcal{B}_n -meßbare Abbildung $X : \Omega \rightarrow \mathbb{R}^n$ ein neuer Wahrscheinlichkeitsraum $(\mathbb{R}^n, \mathcal{B}_n, Q_X)$ gegeben.

Es gilt $\forall B \in \mathcal{B}_n$:

$$Q_X B := PX^{-1}B = P\{\omega \in \Omega; X(\omega) \in B\} =: P\{X \in B\}.$$

Man bezeichnet Q_X als **gemeinsame Wahrscheinlichkeitsverteilung** der Zufallsvariablen X_1, \dots, X_n .

Definition 2.37. (*i*-te Projektion, *i*-te Komponente und Randverteilung)

Sei $X : \Omega \rightarrow \mathbb{R}^n$ eine n -dimensionale Zufallsvariable über einem WR (Ω, \mathcal{F}, P) . Die Abbildung

$$pr_i : \mathbb{R}^n \rightarrow \mathbb{R}^1$$

mit

$$pr_i(x_1, \dots, x_n) = x_i$$

heißt die ***i*-te Projektion** des \mathbb{R}^n auf den \mathbb{R}^1 .

Man bezeichnet

$$X_i := pr_i \circ X : \Omega \rightarrow \mathbb{R} \quad i = 1, \dots, n$$

als ***i*-te Komponente** von X und die Wahrscheinlichkeitsverteilung Q_{X_i} der i -ten Komponente X_i als **Randverteilung** ($i = 1, \dots, n$).

Man unterscheidet folgende Zufallsvariablen:

1. diskrete,
2. stetige,
3. singuläre,
4. Mischtypen der obigen drei.

Da wir uns in Kapitel 5 nur mit diskreten Zufallsvariablen beschäftigen, werden wir im Folgenden auch nur diese näher betrachten.

Definition 2.38. (Diskrete Wahrscheinlichkeit und diskreter Wahrscheinlichkeitsraum)

Sei $\Omega \neq \{\}$ und $(\omega_i)_{i \in I \subseteq \mathbb{N}}$ eine abzählbare Familie von Elementen aus Ω , $\omega_i \neq \omega_j$ für $i \neq j$, $p_i \in]0, 1]$ $\forall i \in I$ mit $\sum_{i \in I} p_i = 1$ und \mathcal{F} σ -Algebra in Ω .

Dann ist durch:

$$(\forall A \in \mathcal{F})(PA = \sum_{\omega_i \in A} p_i)$$

eine Wahrscheinlichkeit $P : \mathcal{F} \rightarrow \mathbb{R}$ definiert und (Ω, \mathcal{F}, P) ein Wahrscheinlichkeitsraum.

Insbesondere gilt: $P\{\omega_i\} = p_i \forall i \in I$.

Dieser Wahrscheinlichkeitsraum und auch die Wahrscheinlichkeit P heißen **diskret**, die ω_i heißen **Trägerpunkte** von P und die p_i die zugehörigen **Punktwahrscheinlichkeiten** oder **Punktmassen**.

Definition 2.39. (Diskrete Zufallsvariable)

Sei X eine n -dimensionale Zufallsvariable mit der Wahrscheinlichkeitsverteilung Q_X . Die Zufallsvariable und auch ihre Wahrscheinlichkeitsverteilung heißen **diskret**, wenn $(R^n, \mathcal{B}_n, Q_X)$ ein diskreter Wahrscheinlichkeitsraum ist.

Definition 2.40. (Unabhängigkeit)

Seien $X_1, X_2, \dots, X_r, n_1, n_2, \dots, n_r$ -dimensionale Zufallsvariablen. Sie heißen (**stochastisch**) **unabhängig**, wenn gilt:

$$Q_{(X_1, \dots, X_r)}(B_1 \times \dots \times B_r) = Q_{X_1} B_1 \cdot \dots \cdot Q_{X_r} B_r \quad \forall B_1 \in \mathcal{B}_{n_1}, \dots, B_r \in \mathcal{B}_{n_r}.$$

Diskrete Verteilungen:

1. Binomialverteilung ($B(n, p)$)

Trägerpunkte: $i \in \{0, 1, \dots, n\}$

Punktwahrscheinlichkeiten: $p_i = Q_X\{i\} = \binom{n}{i} p^i (1-p)^{n-i}$ mit $0 < p < 1$

$EX = np, VarX = np(1-p)$

2. Poisson-Verteilung

Trägerpunkte: $i \in N_0$

Punktwahrscheinlichkeiten: $p_i = Q_X\{i\} = \frac{\lambda^i}{i!} e^{-\lambda}$ mit $\lambda > 0$

$EX = \lambda, VarX = \lambda$

Definition 2.41. (Erwartungswert, Varianz)

Sei X eine diskrete eindimensionale Zufallsvariable mit Trägerpunkten x_i und Punkt-
wahrscheinlichkeiten $Q_X\{x_i\} =: p_i, i \in J \subseteq N$.

a) gilt:

$$\sum_{i \in J} |x_i| p_i < \infty$$

so heißt

$$EX := \sum_{i \in J} x_i p_i$$

der **Erwartungswert (Mittelwert)** der Zufallsvariablen X .

b)

$$E(X - EX)^2 =: VarX =: \sigma^2$$

heißt **Streuung** oder **Varianz** und

$$\sigma := \sqrt{E(X - EX)^2}$$

heißt **Standardabweichung**.

Definition 2.42. (Erwartungswertvektor, Varianz-Kovarianz-Matrix)

Sei $X = (X_1, \dots, X_n)$ eine n -dimensionale Zufallsvariable.

Man bezeichnet

$$\mu := (\mu_1, \dots, \mu_n) := (EX_1, \dots, EX_n)$$

als **Erwartungswertvektor** von X und

$$\Sigma := (\sigma_{i,j}) := (E[(X_i - EX_i)(X_j - EX_j)])_{i,j=1,\dots,n}$$

als **Varianz-Kovarianz-Matrix**.

Anmerkung 2.43. Die Poisson-Verteilung (in Kapitel 5 angenommene Verteilung) kommt nicht unmittelbar, sondern nur mittelbar "in der Natur" vor. Sie ist vor allem dadurch interessant, dass sie eine Approximationsmöglichkeit für die $B(n, p)$ -Verteilung bietet und zwar im Falle kleiner p -Werte und großer n -Werte.

Infolgedessen findet man die Poisson-Verteilung dann empirisch besonders gut bestätigt, wenn man registriert, wie oft ein bei einmaliger Durchführung sehr unwahrscheinliches Ereignis bei vielen Wiederholungen eintritt. Die Poisson-Verteilung wird aus diesem Grund auch als Verteilung der seltenen Ereignisse bezeichnet.

Beispielsweise kann man die Anzahl

- der pro Zeiteinheit an einer Autobahntankstelle tankenden PKW's,
- der "Sechser" pro Ausspielung im Zahlenlotto,
- der pro Zeiteinheit von einer Versicherung zu regulierenden Schadensfälle,
- der pro Zeiteinheit zerfallenden Atome eines spaltbaren Materials, usw.

als Poisson-verteilt annehmen. Man ersetzt in derartigen Fällen die eigentlich vorliegende $B(n, p)$ -Verteilung durch die Poisson-Verteilung $P(np)$, wobei der passende Parameterwert λ der Poisson-Verteilung also das Produkt von n und p ist. Diese Approximation ist brauchbar für $n \geq 50$, $p \leq 0,1$ und $np \leq 10$.

2.4 Grundlagen aus der nichtlinearen Optimierung

In diesem Abschnitt werden wir einige Definitionen und Lemmata betrachten, die später in Kapitel 5 noch benötigt werden. Es handelt sich hier um einen kurzen Abriss, der aus [22] entstanden ist.

I. Allg. müssen wir bei dem NLP-Problem (nonlinear programming problem) etwas maximieren oder minimieren. Jedoch schränken gewisse Nebenbedingungen unsere Aktionen, um das Maximum (oder Minimum) zu erhalten, ein. Mathematisch bedeutet dies, dass wir eine Funktion

$$f : E^n \rightarrow E^1, x \mapsto f(x)$$

betrachten, wobei E^n ein n -dimensionaler euklidischer Raum ist.

Ziel: Finde x , welches $f(x)$ maximiert (oder minimiert).

Wir können x jedoch nicht willkürlich wählen, da wir noch m Bedingungsfunktionen $g_i(x)$, $i = 1, \dots, m$ gegeben haben. In unserem speziellen, für Kapitel 5 notwendigen Fall, muss x den Bedingungen

$$g_i(x) \geq 0, \quad i = 1, \dots, m$$

genügen.

Wir versuchen also $f(x)$ über der Menge, die durch die obigen Bedingungen abgegrenzt ist, zu maximieren (minimieren).

Ein Punkt, der f unter den Nebenbedingungen maximiert, wird **optimal** genannt. Standardformulierung des Problems:

$$\max f(x)$$

$$\text{unter den Nebenbedingungen } g_i(x) \geq 0, \quad i = 1, \dots, m$$

mit $f, g_i : E^n \rightarrow E^1$, $i = 1, \dots, m$.

f wird **Kostenfunktion**, g_i die **Bedingungen** genannt.

Ein x mit $g_i(x) \geq 0 \quad \forall i = 1, \dots, m$ heißt **realisierbar**. Die Menge aller realisierbaren Punkte heißt **realisierbare Menge F**.

Definition 2.44. (Konvexe Menge)

Eine Menge $C \subset E^n$ ist **konvex**, wenn gilt:

$$x_1, x_2 \in C \Rightarrow w := \theta x_1 + (1 - \theta)x_2 \in C \quad \forall \theta \text{ mit } 0 \leq \theta \leq 1.$$

Beispiel 2.45. \mathbb{R}^n ist konvex.

Definition 2.46. (konkave und konvexe Funktionen)

Sei C eine konvexe Menge.

Eine Funktion h auf der Menge C ist **konkav**, wenn gilt:

$$x_1, x_2 \in C$$

$$\Rightarrow h(\theta x_1 + (1 - \theta)x_2) \geq \theta h(x_1) + (1 - \theta)h(x_2) \quad \forall \theta \text{ mit } 0 \leq \theta \leq 1.$$

Eine Funktion h auf C ist **konvex**, wenn $-h$ konkav ist.

Definition 2.47. (*negativ (semi)definit, positiv (semi)definit*)

Sei $A \in M(n \times n, \mathbb{R})$ eine symmetrische $n \times n$ -Matrix.

a) Die Matrix A heißt **positiv definit**, falls

$$\langle x, Ax \rangle > 0 \quad \forall x \in \mathbb{R}^n \setminus \{0\}$$

ist.

b) Die Matrix A heißt **negativ definit**, falls

$$\langle x, Ax \rangle < 0 \quad \forall x \in \mathbb{R}^n \setminus \{0\}$$

ist.

c) Die Matrix A heißt **positiv semidefinit**, falls

$$\langle x, Ax \rangle \geq 0 \quad \forall x \in \mathbb{R}^n$$

ist.

d) Die Matrix A heißt **negativ semidefinit**, falls

$$\langle x, Ax \rangle \leq 0 \quad \forall x \in \mathbb{R}^n$$

ist.

Lemma 2.48. Sei h zweimal stetig partiell differenzierbar.

Dann gilt:

h konkav \Leftrightarrow Hessematrix H ist negativ semidefinit.

Beweis. Mit Hilfe der Taylorentwicklung erhalten wir:

$\exists \theta \in [0, 1]$ mit:

$$h(y) = h(x) + \nabla h(x)^\top (y - x) + \frac{1}{2} (y - x)^\top H(x + \theta(y - x))(y - x), \quad (2.21)$$

und daraus folgt offensichtlich:

$$(y - x)^\top H(x + \theta(y - x))(y - x) \leq 0 \Leftrightarrow h(y) \leq h(x) + \nabla h(x)^\top (y - x). \quad (2.22)$$

” \Leftarrow ” Sei H negativ semidefinit

$$\begin{aligned} &\stackrel{\text{Def. 2.47 d)}}{\Rightarrow} (y-x)^\top H(x+\theta(y-x))(y-x) \leq 0 \\ &\stackrel{(2.22)}{\Rightarrow} h(y) \leq h(x) + \nabla h(x)^\top (y-x) \end{aligned} \quad (2.23)$$

setze $x := \theta x_1 + (1-\theta)x_2$ und $y := x_1$ bzw. $y := x_2$, dann folgt:

$$\begin{aligned} h(x_1) &\leq h(\theta x_1 + (1-\theta)x_2) + \nabla h(\theta x_1 + (1-\theta)x_2)^\top \underbrace{[x_1 - (\theta x_1 + (1-\theta)x_2)]}_{=(1-\theta)(x_1-x_2)} \quad \text{und} \\ h(x_2) &\leq h(\theta x_1 + (1-\theta)x_2) + \nabla h(\theta x_1 + (1-\theta)x_2)^\top \underbrace{[x_2 - (\theta x_1 + (1-\theta)x_2)]}_{=-\theta(x_1-x_2)}. \end{aligned}$$

Damit erhalten wir:

$$\begin{aligned} h(x_1) &\leq h(x) + \nabla h(x)^\top (1-\theta)(x_1-x_2) \quad \| \cdot \theta \quad (\geq 0) \quad \text{und} \\ h(x_2) &\leq h(x) - \nabla h(x)^\top \theta(x_1-x_2) \quad \| \cdot (1-\theta) \quad (\geq 0) \end{aligned}$$

und durch Aufsummieren der beiden Ungleichungen folgt die Behauptung.

” \Rightarrow ” Sei h konkav

$$\begin{aligned} &\Rightarrow h(\theta y + (1-\theta)x) \geq \theta h(y) + (1-\theta)h(x) \quad x, y \in \mathbb{R}^n, \text{ und } \theta \in [0, 1] \\ &\stackrel{\theta \geq 0}{\Rightarrow} \frac{h(\theta y + (1-\theta)x) - h(x)}{\theta} \geq h(y) - h(x) \\ &\Rightarrow \lim_{\theta \rightarrow 0^+} \frac{h(\theta y + (1-\theta)x) - h(x)}{\theta} \geq h(y) - h(x) \\ &\stackrel{h \text{ diff'bar}}{\Rightarrow} \nabla h(x)^\top (y-x) \geq h(y) - h(x) \\ &\Rightarrow h(y) \leq \nabla h(x)^\top (y-x) + h(x) \end{aligned} \quad (2.24)$$

Annahme: $\exists x$ mit $(y-x)^\top H(x)(y-x) > 0$

Wegen der Stetigkeit der Hessematrix können wir y so nahe an x wählen, dass für alle θ mit $0 \leq \theta \leq 1$ gilt:

$$(y-x)^\top H(x+\theta(y-x))(y-x) > 0$$

$$\stackrel{(2.22)}{\Rightarrow} h(y) > h(x) + \nabla h(x)^\top (y-x) \quad \blacktriangleright \text{ zu (2.24).}$$

□

Hat man ein Optimierungsproblem ohne Nebenbedingungen, so gilt:

Sei f diff'bar.

Wenn x^* f über E^n maximiert, dann gilt:

$$\nabla f(x^*) = 0.$$

Die sogenannten **Kuhn-Tucker-Bedingungen** (K-T-Bedingungen) sind nun die verallgemeinerte Form des obigen Satzes für Optimierungsprobleme mit Nebenbedingungen. Intuitiv besagen sie, dass wenn wir uns in eine beliebige Richtung von x^* aus weg bewegen, die Kostenfunktion nicht wachsen kann, solange man in der realisierbaren Region bleibt.

Da wir jetzt nicht noch tiefer in die Materie der Optimierung unter Nebenbedingungen eindringen wollen, geben wir die K-T-Bedingungen und eine hinreichende Bedingung, so dass x^* optimal ist, als Sätze ohne Beweise an. Interessierte können die entsprechenden Beweise in [22], Seite 40/41 bzw. 49/50 nachlesen.

Satz 2.49. (die Kuhn-Tucker-Bedingungen)

Gegeben sei das folgende NLP-Problem:

$$\max f(x)$$

$$\text{unter den Nebenbedingungen } g_i(x) \geq 0 \quad i = 1, \dots, m$$

wobei alle Funktionen diff'bar sind.

Sei x^* eine optimale Lösung und nimmt man an, dass $\mathcal{D}(x^*) = \overline{\mathcal{D}}(x^*)$ ist, wobei

$$\mathcal{D}(x) = \{d; \nabla g_i(x)^\top d \geq 0 \quad \forall i \in \mathcal{A}(x)\} \text{ mit } \mathcal{A}(x) = \{i; g_i(x) = 0\} \text{ und}$$

$$D(x) = \{d; \exists \sigma > 0 \text{ so dass für alle } 0 \leq \tau \leq \sigma \text{ gilt: } x + \tau d \in F\} \text{ und}$$

$$\overline{\mathcal{D}}(x) \text{ ist der Abschluß von } D(x).$$

Dann gilt:

1. x^* ist realisierbar

$$\text{und } \exists \lambda_i \geq 0 \quad i = 1, \dots, m \text{ so, dass}$$

2. $\lambda_i g_i(x^*) = 0 \quad i = 1, \dots, m$

und

$$3. \nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla g_i(x^*) = 0.$$

Satz 2.50. (hinreichende Bedingung)

Es seien $f, g_i, i = 1, \dots, m$ konkav, $x^* \in E^n$ und x^* genügt den K-T-Bedingungen.

Dann gilt:

x^* ist optimal für das NLP-Problem.

3 Das D-C-Modell (Grundmodell) und ein darauf beruhender Bildrekonstruktionsansatz

In diesem Kapitel wollen wir zuerst das physikalische Grundmodell der PET, auch D-C-Modell (discrete-continuous model) oder LSV-Modell (linear, spatially-variant) genannt, etwas eingehender betrachten. Danach geben wir eine mathematische Einführung in die Problemstellung linearer, inverser Probleme und stellen den Lösungsansatz der verallgemeinerten Inversen vor. Mit Hilfe dieser Grundlagen können wir dann das aus dem D-C-Modell abgeleitete inverse Problem eindeutig und stabil lösen und die Vor- und Nachteile dieses Lösungsansatzes in der Praxis diskutieren.

3.1 Das D-C-Modell

In diesem Abschnitt folgen wir dem Paper von R. Lewitt, [10] S. 3 ff.

Können die interessierenden Eigenschaften eines Objektes nicht direkt gemessen werden, sondern muss man von indirekten Beobachtungen auf diese Größe zurückschließen, so sprechen wir von einem inversen Problem.

Definition 3.1. (*inverses Problem*)

Sei $A \in \mathcal{L}(X, Y)$, X, Y Banachräume.

Ein Problem der Form:

gegeben sei $g \in Y$ (Daten),

suche ein $f \in X$ (Lösung, Rekonstruktion) mit

$$Af = g$$

nennen wir ein *inverses Problem*.

Anwendungen, bei denen diese sogenannte indirekte Bildgebung verwendet wird, inklusive der ET (Emissions Tomography), beziehen also diskrete Messungen, die einer Integraltransformation über einer in ihren räumlichen Variablen stetigen Funktion entsprechen, in die Bildgebung mit ein. Diese Funktion und das Bild, das aus den Daten rekonstruiert werden soll, repräsentieren beide die räumliche Verteilung der interessierenden physikalischen Eigenschaften innerhalb des zu betrachtenden Objekts. Bei diesen Anwendungen

wird der Prozess der Datensammlung durch ein diskretes-stetiges (discrete-continuous, kurz: D-C) Modell repräsentiert, welches die diskreten Daten zu der in ihren räumlichen Variablen stetigen Funktion $f(x, y, z)$ in Beziehung setzt.

Im Folgenden wird angenommen, dass es eine lineare, räumlich-variante (linear, spatially-variant, kurz: LSV) Beziehung zwischen den gemessenen Daten und der Funktion $f(x, y, z)$, die die räumliche Verteilung repräsentiert, gibt.

Das LSV-Modell oder D-C-Modell

$$\bar{g}_i^{LSV} = \int \int \int_{\Omega} f(x, y, z) h_i(x, y, z) dx dy dz \quad i = 1, \dots, I \quad (3.1)$$

wobei	Ω	FOV
	$f \in C^\infty(\Omega)$	Verteilungsfunktion des Kontrastmittels
	\bar{g}_i^{LSV}	Komponente der i-ten Messung
	$h_i(x, y, z) \in L_2(\Omega)$	Integrationskern

ist nun eine gute Beschreibung der zu Grunde liegenden Physik der Datensammlung in der ET, wobei die Detektorantwort die Annahme der "linearen Response" erfüllt, jedoch Schwanken in Bezug auf die Position hat.

Wir wollen hier nun gleich den 3D Fall betrachten. (Alle nachfolgenden Überlegungen dieses Kapitels können vollkommen analog auf den 2D Fall übertragen werden).

Ziel: Finde eine Schätzung für die exakte Verteilungsfunktion $f(x, y, z)$ bei gegebener Datemenge $\{\bar{g}_1^{LSV}, \dots, \bar{g}_I^{LSV}\}$.

Die Daten werden dabei auf einem diskreten, endlichen Gitter gesammelt. Der Integrationskern $h_i(x, y, z)$ ist außerhalb des i-ten Tubes Null und nimmt Werte ungleich Null innerhalb des i-ten Tubes an. Er kann als Beitrag eines Punktes (x, y, z) zur i-ten Messung angesehen werden.

In Worten bedeutet das obige Modell also nichts anderes, als dass die i-te Messung als Integral über $f(x, y, z)$ mit einem Gewicht $h_i(x, y, z)$ über dem i-ten Tube modelliert ist.

Von diesem Punkt an können wir von dem D-C-Modell ausgehen und in drei verschiedene Richtungen fortfahren:

1. Wir können uns weiter mit dem D-C-Modell beschäftigen. Dies führt zu einem Ansatz, bei dem die Menge der Funktionen $\{h_i(x, y, z)\}_{i=1}^I$ zur Menge der Basisfunktionen im Bildraum wird.
In den nächsten beiden Abschnitten wird darauf genauer eingegangen werden.
2. In dem wir die gemessenen diskreten Daten $\{\bar{g}_i^{LSV}\}_{i=1}^I$ als Realisationen einer im Messraum stetigen Funktion interpretieren, erhalten wir aus dem D-C-Modell

das C-C-Modell. Für die einfachste Form des Integrationskerns $h_i(x, y, z)$ und der Messgeometrie ist es möglich, eine Inversionsformel für $f(x, y, z)$ herzuleiten.

Die numerische Entwicklung der Inversionsformel liefert Realisationen $f(x_k, y_k, z_k)$ auf einem Gitter von Punkten im Bildraum, unter Verwendung der diskreten Daten $\{\bar{g}_i^{LSV}\}_{i=1}^I$.

Wir diskutieren dies in Kapitel 4 der Arbeit.

3. Stellt man die unbekannt Funktion $f(x, y, z)$ als Linearkombination einer endlichen Anzahl von Basisfunktionen dar, so gelangt man vom D-C-Modell zum D-D-Modell. Jede der Basisfunktionen wird mit einem Koeffizienten multipliziert, der anfänglich unbekannt ist und aus den Daten mit Hilfe eines speziellen Algorithmus bestimmt wird.

Dieser Ansatz ist Gegenstand des 5. Kapitels der Arbeit.

Damit man obiges inverses Problem besser in einen mathematischen Rahmen einordnen und schließlich auch lösen kann, geben wir im nächsten Kapitel eine allgemeine mathematische Einführung in die Problemstellung inverser Probleme und stellen den Lösungsansatz der verallgemeinerten Inversen vor.

3.2 Lineare inverse Probleme und der Lösungsansatz der verallgemeinerten Inversen

Im Folgenden beziehen wir uns ausschließlich (teilweise auch wortwörtlich) auf das Buch und die Vorlesung von Prof. A. K. Louis, [13] und [23] und auf die Vorlesung von Dr. Th. Schuster, [24].

Allgemeine Einführung in die Problemstellung und ein Lösungsansatz:

Gegeben ist als mathematisches Modell eine Abbildung A von der Menge X der Parameter in die Menge Y der Resultate, also

$$A : X \rightarrow Y.$$

Die Lösung des inversen Problems besteht in der Interpretation der Daten $g \in Y$, also in der Konstruktion des Urbildes.

Diese Lösung ist sehr einfach, wenn A eine Bijektion ist und A^{-1} stetig bzgl. geeigneter Topologien in X und Y ist.

Die 1. Bedingung garantiert, dass das Problem

$$Af = g$$

für alle $g \in Y$ eindeutig lösbar ist.

Die 2. Bedingung liefert die Stabilität des Problems in dem Sinne, dass kleine Änderungen in den Daten g nur zu kleinen Änderungen in der Lösung f führen.

Diese Eigenschaften haben zu folgender Begriffsbildung geführt:

Definition 3.2. (Gut gestellt und schlecht gestellt)

Sei $A : X \rightarrow Y$ mit topologischen Räumen X und Y .

Das Problem (A, X, Y) heißt **gut gestellt**, wenn:

i) $Af = g$ für alle $g \in Y$ eine Lösung hat,

ii) diese Lösung eindeutig ist,

iii) die Lösung stetig von den Daten abhängt.

Ist eine der Bedingungen nicht erfüllt, so nennen wir das Problem **schlecht gestellt**.

Kennzeichnend für inverse Probleme ist, dass sie in obigem Sinne schlecht gestellt sind. (In diesem, sowie in allen nachfolgenden Kapiteln dieser Arbeit behandeln wir also, nach obiger Definition, schlecht gestellte, inverse Probleme).

Da nicht für jedes $g \in Y$ eine Lösung existiert und der Wertebereich

$$R(A) = \{g \in Y; \exists f \in X \text{ mit } Af = g\}$$

sehr klein sein kann, betrachten wir statt der Gleichung

$$Af = g$$

das Problem, den Abstand

$$\| Af - g \|^2$$

zu minimieren.

Dadurch definieren wir für eine größere Menge von Elementen in Y einen Lösungsbegriff. Im Folgenden nehmen wir an, dass X und Y Hilberträume sind und dass A linear und stetig ist.

Wenn $g \in R(A)$, dann ist der Abstand Null, ist $g \in R(A)^\perp$, so ist ebenfalls eine Minimierung möglich. Ist A nicht injektiv, so gibt es unendlich viele Elemente, die den Abstand minimieren. Unter all diesen Elementen wählen wir dasjenige, das selbst kleinste Norm hat. Dies führt uns zu folgender Definition:

Definition 3.3. (Verallgemeinerte Inverse und Verallgemeinerte Lösung)

$A^+ : \mathcal{D}(A^+) := R(A) \oplus R(A)^\perp \subset Y \rightarrow X$ mit A^+g minimiert den Defekt $\| Af - g \|$ bzgl. f und hat unter allen minimierenden Elementen kleinste Norm.

Dann heißt A^+ **verallgemeinerte Inverse** oder **Pseudoinverse** und $f^+ = A^+g$ heißt **verallgemeinerte Lösung** oder **Moore-Penrose Lösung** oder **Minimum-Norm-Lösung**.

Satz 3.4. Die Minimum-Norm-Lösung $f_{MP}^+ = A^+g$, $g \in \mathcal{D}(A^+)$ ist die eindeutige Lösung der Normalgleichung

$$A^*Af = A^*g \quad (3.2)$$

in $N(A)^\perp = \overline{R(A^*)}$.

Beweis. 1.) Sei $g \in \mathcal{D}(A^+)$.

Somit können wir g eindeutig darstellen als:

$$g = g_1 + g_2 \text{ mit } g_1 \in R(A), g_2 \in R(A)^\perp.$$

Daraus folgt dann:

$$J^2(u) := \| Au - g \|^2 = \| Au - g_1 - g_2 \|^2 \stackrel{\text{Pythag.}}{=} \| Au - g_1 \|^2 + \| g_2 \|^2.$$

Für das minimierende f_{MP}^+ ist also:

$$J(f_{MP}^+) = \| g_2 \| \quad \text{und} \quad Af_{MP}^+ = g_1.$$

Damit erhalten wir:

$$\langle Af_{MP}^+ - g, Au \rangle = \langle \underbrace{Af_{MP}^+ - g_1}_{=0} - g_2, Au \rangle = - \langle \underbrace{g_2}_{\in R(A)^\perp}, \underbrace{Au}_{\in R(A)} \rangle = 0 \quad \forall u \in X.$$

Es folgt weiter:

$$\langle A^*(Af_{MP}^+ - g), u \rangle = \langle A^*Af_{MP}^+ - A^*g, u \rangle = 0 \quad \forall u \in X,$$

und da $X^\perp = \{0\}$ ist gilt:

$$A^*Af_{MP}^+ - A^*g = 0,$$

woraus sich die Behauptung ergibt.

2.) Eindeutigkeit der Lösung in $N(A)^\perp$:

z.z.: A^*A ist injektiv in $N(A)^\perp$.

dazu: Sei $h \in N(A)^\perp$ und $A^*Ah = 0$

$$\begin{aligned} &\Rightarrow Ah \in N(A^*) = R(A)^\perp \\ &\Rightarrow Ah \in R(A)^\perp \text{ und } Ah \in R(A) \\ &\Rightarrow Ah \in R(A)^\perp \cap R(A) = \{0\} \\ &\Rightarrow Ah = 0 \\ &\Rightarrow h \in N(A) \text{ und nach Voraussetzung } h \in N(A)^\perp \\ &\Rightarrow h \in N(A) \cap N(A)^\perp = \{0\} \\ &\Rightarrow h = 0. \end{aligned}$$

3.) Die Minimum-Norm-Lösung liegt in $N(A)^\perp$:

Unter allen Lösungen der NG hat $f \in N(A)^\perp = \overline{R(A^*)}$ kleinste Norm, denn sei

$$\tilde{f} = f + f_0 \quad \text{mit } f_0 \in N(A),$$

dann ist $\langle f, f_0 \rangle = 0$ und $A\tilde{f} = Af$, aber

$$\|\tilde{f}\|^2 = \|f + f_0\|^2 \stackrel{Pyth.}{=} \|f\|^2 + \|f_0\|^2 > \|f\|^2 \quad \text{für } f_0 \neq 0.$$

□

Einige Eigenschaften der verallgemeinerten Inversen sollen im Folgenden noch zusammengestellt werden:

Satz 3.5. Seien X, Y HR'e, dann gilt:

a) $\mathcal{D}(A^+) = Y \Leftrightarrow R(A) = \overline{R(A)}$

b) $R(A^+) = N(A)^\perp$

c) A^+ ist linear

d) A^+ ist stetig $\Leftrightarrow R(A) = \overline{R(A)}$

Beweis. a) " \Leftarrow " Da $Y = \overline{R(A)} \oplus R(A)^\perp$ folgt

$$\mathcal{D}(A^+) := \underbrace{R(A)}_{\stackrel{Vor.}{=} \overline{R(A)}} + R(A)^\perp \stackrel{\text{Satz 2.5}}{=} Y.$$

" \Rightarrow " Sei $Y = \mathcal{D}(A^+) \stackrel{Def.}{=} R(A) \oplus R(A)^\perp$.

Da Y ein Hilbertraum und $R(A)^\perp \subset Y$ ein abgeschlossener Teilraum ist folgt

$$Y = R(A)^\perp \oplus \underbrace{(R(A)^\perp)^\perp}_{= \overline{R(A)}},$$

3 Das D-C-Modell (Grundmodell) und ein darauf beruhender Bildrekonstruktionsansatz

und damit gilt:

$$\overline{R(A)} = R(A).$$

b) " \subseteq " $R(A^+) \subset N(A)^\perp$ nach Satz 3.4.

" \supseteq " Sei $\varphi \in N(A)^\perp$, $g := A\varphi$

$\Rightarrow P_A g = g = A\varphi$ mit $P_A : Y \rightarrow \overline{R(A)} \subset Y$ orthogonale Projektion

$\Leftrightarrow 0 = A\varphi - P_A g = P_A A\varphi - P_A g$

$\Leftrightarrow P_A(A\varphi - g) = 0$

$\Leftrightarrow A\varphi - g \in N(P_A) \stackrel{\text{Satz 2.7 a)}}{=} R(A)^\perp \stackrel{\text{Satz 2.11 a)}}{=} N(A^*)$

$\Leftrightarrow A^*(A\varphi - g) = 0$

$\Leftrightarrow A^*A\varphi = A^*g.$

Sei nun ψ eine weitere beliebige Lösung der Normalengleichung, dann gilt:

$$\begin{aligned} & A(\varphi - \psi) = 0, \text{ da } P_A g = A\varphi \text{ und } P_A g = A\psi \\ \Rightarrow & \varphi - \psi \in N(A) \\ \Rightarrow & \|\psi\|^2 \stackrel{2^{Phyt.}}{=} \|\varphi\|^2 + \|\psi - \varphi\|^2 \geq \|\varphi\|^2 \\ \Rightarrow & \varphi = A^+g \\ \Rightarrow & N(A)^\perp \subset R(A^+). \end{aligned}$$

c) • Es seien $\varphi, \psi \in \mathcal{D}(A^+)$. Daraus folgt:

$$P_A \varphi = AA^+ \varphi \text{ und } P_A \psi = AA^+ \psi$$

und somit:

$$A \underbrace{A^+(\varphi + \psi)}_{\in R(A^+) \stackrel{b)}{=} N(A)^\perp} = P_A(\varphi + \psi) \stackrel{P_A, A \text{ linear}}{=} A \underbrace{(A^+ \varphi + A^+ \psi)}_{\in R(A^+) \stackrel{b)}{=} N(A)^\perp}. \quad (3.3)$$

Man kann A in (3.3) weglassen, da A injektiv auf $N(A)^\perp$ ist. Somit folgt also:

$$A^+ \varphi + A^+ \psi = A^+(\varphi + \psi).$$

• Analog für $A^+(\alpha\varphi) = \alpha A^+ \varphi$.

d) " \Rightarrow " Sei A^+ stetig

Da $\mathcal{D}(A^+)$ dicht in Y liegt, kann man A^+ stetig auf Y fortsetzen. Es gibt also ein

$$\tilde{A} \in \mathcal{L}(Y, X) \text{ mit } A\tilde{A}y = P_A y \quad \forall y \in Y.$$

Daraus folgt:

$$\underbrace{R(P_A)}_{\stackrel{\text{Satz 2.7 a)}}{=} \overline{R(A)} \subseteq R(A)$$

und somit erhalten wir die Gleichheit:

$$\overline{R(A)} = R(A).$$

” \Leftarrow ” Es sei $R(A) = \overline{R(A)}$.

Wir betrachten nun die Einschränkung von A auf $N(A)^\perp$, d.h. $\hat{A} := A|_{N(A)^\perp}$:

- \hat{A} ist stetig und linear als Einschränkung eines stetig linearen Operators
- \hat{A} ist injektiv auf $N(A)^\perp$; dort gilt:

$$\hat{A}f = Af = 0 \quad \Rightarrow \quad f \in N(A) \text{ und nach Voraussetzung } f \in N(A)^\perp.$$

Daraus folgt

$$f \in N(A) \cap N(A)^\perp = \{0\},$$

also ist $f = 0$ und damit ist \hat{A} injektiv.

- $\hat{A} : N(A)^\perp \rightarrow R(A) \stackrel{\text{Vgl.}}{=} \overline{R(A)}$ ist surjektiv

Wir haben also $\hat{A} \in \mathcal{L}(N(A)^\perp, \overline{R(A)})$, wobei $\overline{R(A)}$ und $N(A)^\perp$ vollständig normierte Räume, also Banachräume sind und \hat{A} ist bijektiv.

Somit folgt aus Satz 2.12, dass

$$\hat{A}^{-1} \in \mathcal{L}(\overline{R(A)}, N(A)^\perp)$$

ist und somit erhalten wir:

$$\|A^+y\| = \|\hat{A}^{-1}(\hat{A}A^+y)\| \stackrel{\text{Satz 2.8}}{\leq} \|\hat{A}^{-1}\| \|\hat{A}A^+y\| \stackrel{A^+ \in N(A)^\perp}{=} \|\hat{A}^{-1}\| \|AA^+y\|. \quad (3.4)$$

Da $y = y_1 + y_2$ eine eindeutige Zerlegung mit $y_1 \in R(A)$ und $y_2 \in R(A)^\perp$ ist, $P_A y = y_1$ folgt mit Pythagoras:

$$\|y\| \geq \|y_1\| = \|P_A y\| \quad \text{und} \quad \|P_A y\| = \|AA^+y\| \stackrel{(3.4)}{\geq} \|\hat{A}^{-1}\|^{-1} \|A^+y\|$$

und durch Multiplikation mit $\|\hat{A}^{-1}\|$ auf beiden Seiten ergibt sich:

$$\|A^+y\| \leq \|\hat{A}^{-1}\| \underbrace{\|P_A y\|}_{\leq \|y\|} \leq \|\hat{A}^{-1}\| \|y\| \quad \forall y \in Y.$$

Mit Hilfe von Satz 2.8 erhalten wir dann schließlich:

$$A^+ \in \mathcal{L}(Y, X) \text{ mit } \|A^+\| \leq \|\hat{A}^{-1}\|.$$

□

Das Problem der verallgemeinerten Inversen ist i. Allg., dass $R(A)$ nicht abgeschlossen und daher nach A^+ nicht stetig ist. Da A^+ auf die gemessenen Daten angewendet wird und diese Messfehler enthalten, kann die Unstetigkeit von A^+ riesige Artefakte im rekonstruierten Bild verursachen.

Die aus diesem Problem entstandene Idee, A^+ durch eine Folge stetiger Operatoren zu ersetzen, soll hier nicht weiter ausgeführt werden, da in unserem speziellen Fall (D-C-Modell) $R(A) = \overline{R(A)}$ sein wird und wir deshalb den Ansatz der verallgemeinerten Inversen verwenden können.

Nachdem wir uns nun den mathematischen Hintergrund der inversen Probleme und einen möglichen Lösungsansatz vor Augen geführt haben, können wir diesen nun auf unser spezielles Problem, abgeleitet aus dem D-C-Modell, anwenden. Dies führt uns zum nächsten Abschnitt.

3.3 DART

Dieser Abschnitt ist eine, auf den Ideen von [10] S. 5 ff. und [11] S. 125 ff. beruhende Ausarbeitung.

Der in Abschnitt 3.1 vorgestellte Ansatz (3.1) führt uns zu einer direkten Lösung des D-C-Problems, ohne das weitere Diskretisierungen oder Approximationen eingeführt werden müssen. Obwohl dieser Ansatz diese attraktive theoretische Eigenschaft hat, werden die darauf beruhenden Methoden kaum bzw. selten in der PET benutzt. Wir werden daher nur eine Rekonstruktionsmethode (die auf der Moore-Penrose-Lösung beruhende Methode) für das D-C-Modell herleiten und uns i. Allg., verglichen mit den folgenden Modellen und Methoden, eher kürzer halten.

Zu lösen ist bei obigem Ansatz das inverse Problem:

$$Pf = \bar{g}^{LSV} \text{ mit gegebenem Datenvektor } \bar{g}^{LSV} \text{ und}$$

$$P : C^\infty(\Omega) \rightarrow \mathbb{R}^I \text{ linear und kompakt (also stetig nach Satz 2.10).}$$

Wegen der Dichtheit von $C^\infty(\Omega)$ in $L_2(\Omega)$ folgt:

Es existiert eine stetig lineare Fortsetzung:

$$P : L_2(\Omega) \rightarrow \mathbb{R}^I \text{ mit } Pf = \left(\langle f, h_1 \rangle_{L_2(\Omega)}, \dots, \langle f, h_I \rangle_{L_2(\Omega)} \right)^\top = (\bar{g}_1^{LSV}, \dots, \bar{g}_I^{LSV})^\top.$$

Wir interpretieren \bar{g}_i^{LSV} in (3.1) also als das Skalarprodukt der Funktionen f und h_i in $L_2(\Omega)$.

(In Worten: P sei also der lineare, stetige Operator, der auf $f(x, y, z)$ gemäß (3.1) operiert und den Datenvektor $\bar{g}^{LSV} = (\bar{g}_1^{LSV}, \dots, \bar{g}_I^{LSV})^\top$ erzeugt.)

Es ist klar, dass eine endliche Anzahl von Messungen alleine nicht ausreicht, um eine unbekannte Funktion $f \in L_2(\Omega)$ zu bestimmen. Das inverse Problem ist also schlecht

3 Das D-C-Modell (Grundmodell) und ein darauf beruhender Bildrekonstruktionsansatz

gestellt und man benötigt weitere Bedingungen an die Funktion, um eine eindeutige Lösung zu erhalten.

Eine mögliche eindeutige Lösung des unterbestimmten Systems

$$\bar{g}_i^{LSV} = P_i f \quad i = 1, \dots, I$$

ist die Moore-Penrose erzeugte Lösung. Sie ist die eindeutige Funktion minimaler Norm unter denen, die die Norm

$$\| Pf - \bar{g}^{LSV} \|$$

minimieren.

Lemma 3.6. *Der zu $P : L_2(\Omega) \rightarrow \mathbb{R}^I$ adjungierte Operator lautet:*

$$P^* : \mathbb{R}^I \rightarrow L_2(\Omega) \text{ mit } (P^*r)(x, y, z) = \sum_{i=1}^I h_i(x, y, z)r_i. \quad (3.5)$$

Beweis.

$$\begin{aligned} \langle Pf, \tilde{r} \rangle_{\mathbb{R}^I} &= \sum_{i=1}^I \int \int \int_{\Omega} f(x, y, z) h_i(x, y, z) \tilde{r}_i \, dx \, dy \, dz \\ &= \int \int \int_{\Omega} f(x, y, z) \underbrace{\sum_{i=1}^I \tilde{r}_i h_i(x, y, z)}_{=P^*\tilde{r}} \, dx \, dy \, dz \\ &= \langle f, P^*\tilde{r} \rangle_{L_2(\Omega)}. \end{aligned}$$

□

Satz 3.7. *Die Moore-Penrose-Lösung des obigen inversen Problems ist*

$$f_{MP}^+ = P^*r, \quad (3.6)$$

wobei r die Lösung des linearen Gleichungssystems $PP^*r = \bar{g}^{LSV}$ ist.

$PP^*r = \bar{g}^{LSV}$ braucht nicht notwendigerweise eine eindeutige Lösung zu haben. Im Falle der Nichteindeutigkeit ist es egal, welche Lösung man nimmt.

Beweis. z.z.:

- 1.) P^*r löst NG
- 2.) $P^*r \in N(P)^\perp$

dazu:

$$1.) P^* P f_{MP}^+ = P^* \underbrace{PP^* r}_{=\bar{g}^{LSV}} = P^* \bar{g}^{LSV}$$

$$2.) \text{ Ann.: } P^* r \in N(P) \quad L_2 = N(P) \oplus N(P)^\perp$$

$$\Rightarrow \underbrace{PP^* r}_{=\bar{g}^{LSV}} = 0$$

$$\Rightarrow \bar{g}^{LSV} = 0 \quad \text{⚡}$$

$$\Rightarrow P^* r \in N(P)^\perp.$$

Daraus folgt, dass $P^* r$ die eindeutige Lösung der NG in $N(P)^\perp$ ist und damit ist nach Satz 3.4 $P^* r$ die Moore-Penrose-Lösung. □

$S := PP^* \in M(I \times I, \mathbb{R})$ ist die **Gram-Matrix** der h_i mit Elementen

$$S_{i,j} = \int \int \int_{\Omega} h_i(x, y, z) h_j(x, y, z) dx dy dz.$$

Da $R(P) < \infty$ ($R(P) \subset \mathbb{R}^I$), ist $R(P) = \overline{R(P)}$ und deshalb ist

$$P^+ : R(P) \oplus N(P^*) = \mathbb{R}^I \rightarrow L_2(\Omega)$$

mit

$$P^+ \bar{g}^{LSV} := P^* r = f_{MP}^+ \text{ (wobei } r \text{ die Lösung von } PP^* r = \bar{g}^{LSV} \text{ ist)}$$

nach Satz 3.5 d) stetig.

Wir haben damit also eine eindeutige Lösung gefunden, die stetig von den Daten abhängt.

Anmerkung 3.8. Anstelle der Moore-Penrose-Lösung hätten wir z. B. auch die regularisierte Tikhonov-Phillips-Lösung nehmen können.

Man müsste in diesem Fall

$$\| Pf - \bar{g}^{LSV} \|^2 + \gamma \| f \|^2$$

mit dem Regularisierungsparameter $\gamma > 0$ minimieren.

Die Lösung - analog zum den vorherigen Betrachtungen - wäre in diesem Fall

$$f_{TP}^+ = P^* r \text{ mit } (S + \gamma I)r = \bar{g}^{LSV}.$$

Diese Art von Algorithmen, bei denen Messungen \bar{g}_i^{LSV} als beschränkte (stetige) lineare Funktionale $\bar{g}_i^{LSV} = P_i f$ auf $L_2(\Omega)$ angesehen werden, heißen ” **direct algebraic algorithms**”, kurz: **DART**.

Wie wir gesehen haben, müssen wir bei diesem Ansatz eine ” Lösung ” des unterbestimmten Systems

$$\bar{g}_i^{LSV} = P_i f \quad i = 1, \dots, I$$

bestimmen. Dies kann - abhängig von den Anforderungen, die wir an unser f stellen - z. B. die Moore-Penrose-Lösung oder aber auch die regularisierte Tikhonov-Phillips-Lösung sein. (Weitere Lösungen sind möglich, werden jedoch hier nicht weiter erwähnt werden).

Zu den Vor- und Nachteilen von DART nehmen wir im nächsten Abschnitt Stellung.

3.4 Diskussion des Lösungsansatzes DART und einige Randbemerkungen

Die Anregungen zur folgenden Diskussion entstammen ebenfalls aus [10] S. 5 und [11] S. 125.

Zu den Vorteilen von DART zählen wir, dass die Strahlengeometrie leicht modelliert werden kann und dass es keine Diskretisierungsfehler gibt.

Die Nachteile sind, dass die Implementierung nicht ganz einfach ist und sehr große Matrizen, abhängig von der Scan-Geometrie, vorberechnet und gespeichert werden müssen. Das Hauptproblem liegt also in der Notwendigkeit, das große, nicht-wenig besetzte, lineare System $Sr = \bar{g}^{LSV}$ zu konstruieren und zu lösen, um r zu erhalten. Die Berechnung der Matrixelemente $S_{i,j}$ ist ebenfalls keine triviale Aufgabe. Für 3D Geometrien benötigt dieser Ansatz, verglichen mit anderen Bildrekonstruktionsmethoden aus 3D-Daten, eine übertrieben Menge an Berechnungen. Die Matrix S ist sehr groß ($I \times I$, wobei I die Anzahl der verschiedenen Messungen ist) und dicht (relativ wenig Elemente sind Null), weil die Funktionen $h_i(x, y, z)$ nicht lokalisiert sind. Diese Funktionen sind Basisfunktionen, die überlagert werden, um das Bild zu gestalten ($f_{MP}^+(x, y, z) = \sum_{i=1}^I h_i(x, y, z)r_i$).

Dieselben Vor- und Nachteile hat man auch bei der regularisierten Tikhonov-Phillips-Lösung. Zuerst muss man auch hier eine Variante des linearen Systems $Sr = \bar{g}^{LSV}$ lösen. Die Dimension dieses Systems entspricht ebenfalls der Anzahl der Messungen und ist normalerweise sehr groß. Damit ist eine direkte Lösung dieses Systems ebenfalls nicht realisierbar.

Zum Ende dieses Kapitels nun noch einige Randbemerkungen zu den Basisfunktionen und zu einer nichtlinearen Rekonstruktionsmethode:

Anmerkung 3.9. • Die h_i 's sind charakteristische Strahlenfunktionen, d. h. für die Strahlen typische, charakteristische Funktionen endlicher Breite/Ausdehnung. Die h_i 's können auch allgemeinere, von den Strahlen unterstützte Funktionen sein, die die uneinheitliche Detektorsensitivität und ähnliches modellieren.

3 Das D-C-Modell (Grundmodell) und ein darauf beruhender Bildrekonstruktionsansatz

- Basisfunktionen, die den Integrationskernen des Messprozesses in der Tomography entsprechen heissen "natürliche Pixel" und ihr Gebrauch bei linearen Rekonstruktionsmethoden hat eine solide Grundlage in der Theorie der linearen, inversen Probleme mit diskreten Daten.
Dieser lineare Inversionsansatz ist ziemlich allgemein, jedoch sind die meisten Entwicklungen in Richtung Rekonstruktion in der SPECT gerichtet.
- Eine nichtlineare Rekonstruktionsmethode wurde auch formuliert und wird weitgehend in gewissen nicht-medizinischen Anwendungen der tomographischen Bildgebung benutzt, wo nur eine kleine Anzahl an Messungen notwendig ist.

4 Das C-C-Modell und darauf beruhende Bildrekonstruktionen

Der Ansatz des C-C-Modells (continuous-countinuous model) führt zu schnelleren Methoden, die weitgehend in der Praxis verwendet werden. Allerdings ist ihre Durchführung sicherlich weniger optimal, da dies lineare Methoden sind, die weder detaillierte Modelle der Physik des Messprozesses noch die signalabhängige Statistik der Messungen berücksichtigen. Im Folgenden konzentrieren wir uns auf die Geometrie der Datenerfassung, die zu parallelen Projektionen führt, da dieser Fall von großer praktischer Bedeutung in der PET ist. (Andere Datenerfassungsgeometrien, die zu Projektionen führen, die aus nichtparallelen Strahlen bestehen, findet man in der SPECT).

In den weiteren Abschnitten wollen wir nun das C-C-Modell vorstellen und dann auf dieser Grundlage sowohl im 2D als auch im 3D Fall eine altbekannte Inversionsformel (FBP) mit Hilfe des Projektionssatzes herleiten. Wichtig ist hier, dass wir, im Gegensatz zum vorangegangenen und zum folgenden Kapitel, genaustens zwischen 2D und 3D PET unterscheiden müssen und die beiden Fälle nicht einfache durch "analoge" Betrachtung ineinander übergehen. Schließlich wollen wir uns dann mit Verfahren beschäftigen, die man unter dem Begriff Rebinningverfahren zusammenfassen kann. Ein Vergleich der vorgestellten Verfahren miteinander beendet auch dieses Kapitel.

4.1 Das C-C Modell

Die Ideen zu diesem Abschnitt stammen aus [6] und [10].

Die Herleitung des C-C-Modells aus dem D-C Ausdruck (3.1) beginnt mit der Interpretation der Komponenten \bar{g}_i^{LSV} des gemessenen Vektors als eine Realisation einer im Messraum stetigen Funktion.

Die nächsten Schritte bestehen nun darin, Vereinfachungen an dem D-C Modell vorzunehmen, bis man ein C-C Modell erhält, für welches eine analytische Inversionsformel hergeleitet werden kann.

Die erste Vereinfachung beinhaltet das "Schrumpfen" jeder Tube-geformten Region der Integration auf eine unendlich lange dünne Linie. Dies hat zur Folge, dass das 3D Integral über dem Volumen des Tubes durch ein 1D Integral entlang einer Linie ersetzt wird. Als nächstes wird der Integrationskern $h_i(x, y, z)$ auf denselben konstanten Wert für alle Punkte entlang der Linie und für alle Linien $i = 1, \dots, I$ gesetzt.

Diese Vereinfachung approximiert das LSV-Modell nicht nur durch ein lineares, räumliches invariantes (LSI-linear, spatially invariant) Modell, sondern auch durch einen spe-

ziellen Fall des LSI-Modells, nämlich den, der die Integration von $f(x, y, z)$ entlang von Linien mit sich bringt.

Dieser sogenannte C-C Ansatz führt uns in Richtung paralleler Projektionen.

Um den Ort der Linie im 3D Raum zu bestimmen, definieren wir uns Rotationskoordinaten (x_r, y_r, z_r) , die wie folgt entstehen:

- Rotation der Originalkoordinaten (x, y, z) um den Winkel $90^\circ - \Phi$ um die z -Achse in negativer Richtung, gefolgt von
- einer Rotation um den Winkel $90^\circ - \theta$ um die neue entstandene x -Achse in negativer Richtung und schließlich
- eine Rotation um 180° um die neue z -Achse (um die positive Richtung der neuen x - und y -Achse zu ändern).

In der Matrixschreibweise bedeutet das also nichts anderes, also dass

$$\begin{pmatrix} x_r \\ y_r \\ z_r \end{pmatrix} = \begin{pmatrix} -\sin \phi & \cos \phi & 0 \\ -\cos \phi \sin \theta & -\sin \phi \sin \theta & \cos \theta \\ \cos \phi \cos \theta & \sin \phi \cos \theta & \sin \theta \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (4.1)$$

ist.

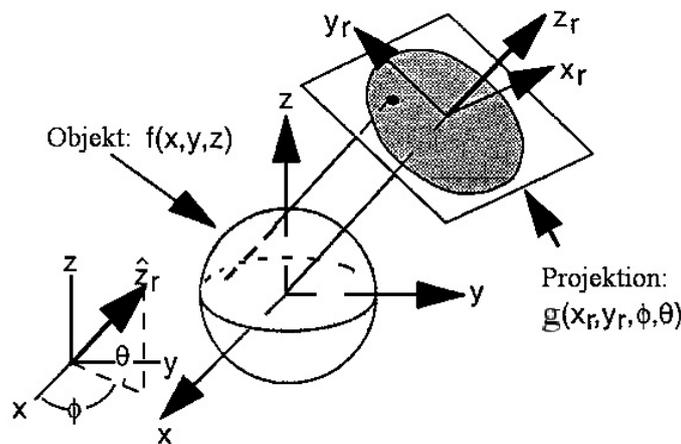


Abbildung 4.1: 2D Projektion. Wir definieren eine 2D Projektion als die Sammlung aller parallelen Linienintegrale für feste Winkel ϕ und θ . Quelle: [6] S. 14.

Wir definieren die Linie im Raum also als Parallele zur z_r -Achse, die durch den Punkt $(x_r, y_r, 0)$ geht. Somit wird die Linie durch die vier Parameter (x_r, y_r, ϕ, θ) bestimmt.

Wir erhalten nun das C-C Modell für 3D Daten in der "sehr simpel ausgedrückten" Form

$$g^{LSI}(x_r, y_r, \phi, \theta) = \int_{\mathbb{R}} f(x, y, z) dz_r, \quad (4.2)$$

wobei (x, y, z) die Koordinaten eines Punktes in dem ursprünglichen System sind, der Punkt auf der Integrationslinie L liegt und durch (x_r, y_r, z_r) in den rotierten Koordinaten spezifiziert wird.

Die Funktion $g^{LSV}(x_r, y_r, \phi, \theta)$ ist eine in den Variablen des Messraumes stetige Funktion und die physikalische Komponente der i -ten Messung ist der Wert dieser Funktion im Punkt $(x_{r_i}, y_{r_i}, \phi_i, \theta_i)$ des Messraumes.

Ersetzen wir nun $(x, y, z)^\top$ durch:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -\sin \phi & -\cos \phi \sin \theta & \cos \phi \cos \theta \\ \cos \phi & -\sin \phi \sin \theta & \sin \phi \cos \theta \\ 0 & \cos \theta & \sin \theta \end{pmatrix} \begin{pmatrix} x_r \\ y_r \\ z_r \end{pmatrix} \quad (4.3)$$

dann gilt:

$$\begin{aligned} g^{LSI}(x_r, y_r, \phi, \theta) &= \int_{\mathbb{R}} f(x, y, z) dz_r \\ &= \int_{\mathbb{R}} f(-x_r \sin \phi - y_r \cos \phi \sin \theta + z_r \cos \phi \cos \theta, x_r \cos \phi - y_r \sin \phi \sin \theta \\ &\quad + z_r \sin \phi \cos \theta, y_r \cos \theta + z_r \sin \theta) dz_r \\ &= \int_{\mathbb{R}} f(a + t\omega) dt \\ &\stackrel{(2.16)}{=} (Pf)(\omega, a), \end{aligned} \quad (4.4)$$

mit

$$a = x_r \begin{pmatrix} -\sin \phi \\ \cos \phi \\ 0 \end{pmatrix} + y_r \begin{pmatrix} -\cos \phi \sin \theta \\ -\sin \phi \sin \theta \\ \cos \theta \end{pmatrix} \text{ und } \omega = \begin{pmatrix} \cos \phi \cos \theta \\ \sin \phi \cos \theta \\ \sin \theta \end{pmatrix}. \quad (4.5)$$

Wir können nachrechnen, dass

$$a \in \omega^\perp \text{ liegt}$$

und erkennen, dass unsere vorliegenden Daten nichts anderes als die in Definition (2.29) eingeführte 3D Röntgentransformation beschreiben.

Da wir in diesem Kapitel eine einheitliche Notation benutzen wollen und gerade für die Beweise bei dem Rebinningverfahren FORE-J (4.4) die sinnvollere Parametrisierung ist, wollen wir im Folgenden (4.4) beibehalten, aber trotzdem immer im Hinterkopf behalten, dass man das Ganze auch anders notieren kann (was z. B. bei der Durchführung

einiger Beweise, die nicht zu FORE-J gehören, durchaus von Vorteil ist und in diesen speziellen Fällen dann auch angewendet wird).

Da wir unsere Betrachtungen zuerst im 2D durchführen wollen, bevor wir den 3D Fall angehen, stellen wir nun an dieser Stelle noch kurz das C-C Modell für den 2D Fall vor. Analog zum 3D Fall gilt:

Um den Ort der Linie in der 2D Ebene zu bestimmen, definieren wir Rotationskoordinaten (x_r, y_r) , wobei diese Koordinaten durch Drehung der Originalkoordinaten (x, y) um den Winkel ϕ gegen den Uhrzeigersinn entstehen. Das bedeutet also, dass

$$\begin{pmatrix} x_r \\ y_r \end{pmatrix} = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (4.6)$$

ist.

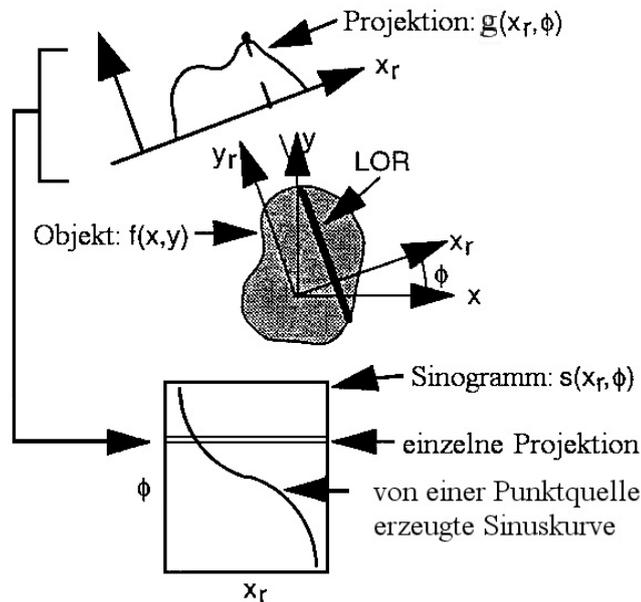


Abbildung 4.2: 1D Projektion. Idealisierte Beschreibung einer 1D Projektion eines 2D Objektes als die Sammlung aller paralleler Linienintegrale für einen Winkel ϕ . Die Beziehung einer Projektion zum Sinogramm wird im unteren Teil des Bildes verdeutlicht. Quelle: [6] S. 13.

Wir definieren die Linie in der Ebene als Parallel zur y_r -Achse, die durch den Punkt $(x_r, 0)$ geht. Somit wird die Linie durch die zwei Parameter (x_r, ϕ) bestimmt.

Das C-C Modell für die 2D Daten lautet daher:

$$g^{LSI}(x_r, \phi) = \int_{\mathbb{R}} f(x, y) dy_r, \quad (4.7)$$

wobei (x, y) die Koordinaten eines Punktes im ursprünglichen System sind, der Punkt auf der Integrationslinie L liegt und durch (x_r, y_r) in den rotierten Koordinaten spezifiziert wird.

Die Funktion $g^{LSI}(x_r, \phi)$ ist eine in den Variablen des Messraumes stetige Funktion und die physikalische Komponente der i -ten Messung ist der Wert dieser Funktion in dem Punkt (x_{r_i}, ϕ_i) . Auch hier haben wir einen "alten Bekannten" vorliegen. Ersetzen wir nämlich $(x, y)^\top$ durch:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} x_r \\ y_r \end{pmatrix} \quad (4.8)$$

dann gilt:

$$\begin{aligned} g^{LSI}(x_r, \phi) &= \int_{\mathbb{R}} f(x, y) dy_r \\ &= \int_{\mathbb{R}} f(x_r \cos \phi - y_r \sin \phi, x_r \sin \phi + y_r \cos \phi) dy_r \\ &= \int_{\mathbb{R}} f(x_r \omega + y_r \omega^\perp) dy_r \\ &\stackrel{(2.11)}{=} (Rf)(\omega, x_r) \end{aligned} \quad (4.9)$$

mit

$$\omega = \begin{pmatrix} \cos \phi \\ \sin \phi \end{pmatrix} \text{ und } \omega^\perp = \begin{pmatrix} -\sin \phi \\ \cos \phi \end{pmatrix}. \quad (4.10)$$

Ist im 2D Fall der Winkel $\phi \in [0, \pi]$ und sind im 3D Fall die Winkel $\phi, \theta \in [0, \pi]$ fest vorgegeben, so bezeichnet man die Funktion der verbleibenden Variablen x_r bzw. x_r und y_r als die **Parallelprojektion** von $f(x, y)$ bzw. $f(x, y, z)$ für diese gegebene Richtung der Integrationslinien.

Wir schreiben:

$$g_\phi(x_r) := g(x_r, \phi) := g^{LSI}(x_r, \phi), \quad (4.11)$$

$$g_{\phi, \theta}(x_r, y_r) := g(x_r, y_r, \phi, \theta) := g^{LSI}(x_r, y_r, \phi, \theta). \quad (4.12)$$

Für Parallelprojektionen gibt es eine einfache Beziehung zwischen der 1D bzw. 2D Fouriertransformation einer Projektion bei gegebenem Winkel ϕ bzw. gegebenen Winkeln ϕ und θ und einer entsprechenden Linie bzw. Ebene innerhalb der 2D bzw. 3D Fouriertransformation der Funktion $f(x, y)$ bzw. $f(x, y, z)$.

Auf diese Beziehungen werden wir in den Abschnitten 4.4 und 4.5.1 der Arbeit noch genauer eingehen.

Bevor wir jedoch damit beginnen, wollen wir zuerst einmal klären, in welchem Format die Daten gesammelt werden und die damit in Beziehung stehenden Begriffe erläutern. Dies führt uns zum nächsten Abschnitt.

4.2 Datenformate, Projektionen und Sinogramme

Dieser Abschnitt entstand mit Hilfe von [6], [7] und [10].

Es gibt zwei verschiedene Methoden, während eines PET-Scans Informationen über die Ereignisse zu sammeln. Bei der für die PET gebräuchlicheren Methode wird im Speicher des Computers ein 3D Feld für jedes LOR, das während des Scans gemessen werden kann, bereitgestellt. Anfänglich sind alle diese Elemente auf Null gesetzt. Für jede Koinzidenz, die während des Scans detektiert wird, wird das dem LOR entsprechende Feldelement um Eins erhöht. Dieses Feld, das die gesamten Zählungen als eine Funktion von Positionsvariablen speichert, bezeichnet man als **Histogramm**. Die individuellen Speicherelemente nennt man **Bins**. Wird der Datenraum also in I Bins aufgeteilt, so ist I die Anzahl von Paaren, bestehend aus entgegengesetzten Detektorelementen.

In der Praxis ist die Anzahl I der Histogrammbins jedoch gewöhnlich viel kleiner als die Anzahl der Paare entgegengesetzter Detektorelemente, die normalerweise 75-200 Millionen bei klinischen Scannern beträgt. Es wäre nicht effizient, diese riesige Anzahl an Histogrammbins zu speichern und zu bearbeiten, da die Anzahl der gemessenen Koinzidenzen gewöhnlich viel kleiner ist (normalerweise liegt sie nur bei 20-50 Millionen).

Der Gebrauch einer kleineren Anzahl an Histogrammbins resultiert jedoch in einem gewissen Verlust an räumlicher Auflösung, da die Zählungen, die mit einer Anzahl benachbarter Linien im Raum assoziiert werden, (alle) zusammen in demselben Histogrammbin gesammelt werden.

Bei der zweiten Methode, bekannt als **List-Mode-Methode**, werden die Detektorkoordinaten für jede einzelne Zählung in einer langen Liste abgespeichert und zwar in der Reihenfolge, in der Photonenpaare detektiert werden.

List-Mode Daten können einige Zeit nach der Datensammlung ins Histogrammformat überführt werden und dann durch herkömmliche Algorithmen rekonstruiert werden. Die Rekonstruktion kann aber auch direkt von den List-Mode Daten gemacht werden.

Es gibt jedoch nur eine beschränkte Anzahl von Algorithmen für die Bildrekonstruktion von List-Mode Daten in der PET, im Gegensatz zu einer breiten Auswahl an Algorithmen für Daten, die im herkömmlichen Histogrammformat vorliegen. Das Problem der Rekonstruktion von List-Mode Daten führt zu einer ML Formulierung (siehe Genaueres dazu in Abschnitt 5.3.1) und schließlich zu Algorithmen, die die ML-Lösung bei gegebenen List-Mode Daten schätzen.

Als einen notwendigen ersten Schritt für die Bildrekonstruktionsalgorithmen in diesem Kapitel nehmen wir an, dass die Daten im Histogrammformat vorliegen. (In Kapitel 5 werden wir dahingegen annehmen, dass die Daten im List-Mode Format vorliegen).

Diese Histogrammbinwerte entsprechen dann den zugehörigen Linienintegralen durch das Objekt.

Die 2D Daten können in ein 2D Feld (Matrix) gespeichert und angezeigt werden. Dieses Feld (Matrix) wird durch x_r und ϕ indiziert und wird **Sinogramm** genannt, da eine Punktquelle im FOV des Scanners zu einem auf die Seite gedrehten Sinuskurvenmus-

ter führt. Genauer gesagt bedeutet dies, dass wenn man bei einer großen Anzahl von LOR's, die alle von demselben Punkt bzw. Pixel im FOV des Scanners ausgehen, die Histogrammbinwerte der entsprechenden LOR's an die dafür vorgesehene Stelle in dem 2D Feld einträgt, eine halbe, auf die Seite gedrehte Sinuskurve erhält. (siehe Abbildung 4.2).

Da man i. Allg. komplizierte Objekte betrachtet, die über mehrere Pixel hinweggehen, besteht das zugehörige Sinogramm aus einer großen Anzahl sich überlappender Sinuskurven.

Jeder Pixelwert entlang einer Reihe im Sinogramm entspricht also der Summe aller Ereignisse, die entlang des zugehörigen LOR's stattgefunden haben. Solch eine Sammlung von LOR-Summen in einer Reihe entspricht gemäß unseres C-C Modells der Sammlung aller, zu einer vorgegebenen Richtung paralleler, Linienintegrale und führt uns wieder zu dem Begriff der Parallelprojektion.

Die in (4.9) vorgestellten 2D Projektionsdaten werden also in einem Sinogramm gespeichert und angezeigt. Jede Reihe im Sinogramm entspricht der Parallelprojektion von $f(x, y)$ mit dem der Reihe entsprechenden Projektionswinkel ϕ .

Zusammengefasst kann man also sagen, dass ein Sinogramm die 2D Darstellung **aller** 1D Projektionen einer Objektschicht ist. Dabei ist die eine Koordinate dieser Darstellung die Ortskoordinate der Projektion, die andere der Projektionswinkel. Die Daten werden nur für die Winkel $0 \leq \phi \leq \pi$ gespeichert, da für $\pi \leq \phi \leq 2\pi$ die Projektionen folgende Symmetrien aufweisen:

$$g(x_r, \phi + \pi) = g(-x_r, \phi).$$

Im Gegensatz zur 2D PET-Bildgebung gibt es zwei Formate, die für die Speicherung der 4D Projektionsdaten in der 3D PET verwendet werden. Diese werden in der folgenden Abbildung illustriert. Wir bezeichnen sie als **Projektions-Modus** und **schräger Sinogramm Modus**.

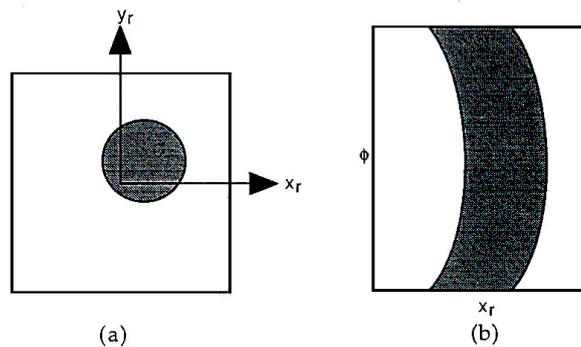


Abbildung 4.3: Die beiden Formate für die Speicherung der 4D Projektionsdaten, die in der 3D PET Bildgebung gesammelt wurden sind: a) das Projektionsformat (siehe auch Abbildung 4.1) für einen festen Wert (ϕ, θ) und b) das schräge Sinogrammformat für einen festen Wert von (y_r, θ) . Quelle: [6] S. 15.

Beide Formate repräsentieren dieselben Daten, jedoch mit einer unterschiedlichen Anordnung der Parameter. Welches der beiden Formate nun benutzt wird hängt von der gewählten Datenverarbeitung ab.

In der 3D PET sind Projektionen 2D Mengen paralleler Linienintegrale mit gegebenem azimuthalen Winkel ϕ und copolaren Winkel θ .

Auf die exakte Beziehung zwischen den beiden Formaten wird nun im Folgenden genauer eingegangen.

Um das Sampling-Muster in einem Sinogramm besser zu verstehen, betrachten wir die folgende Abbildung:

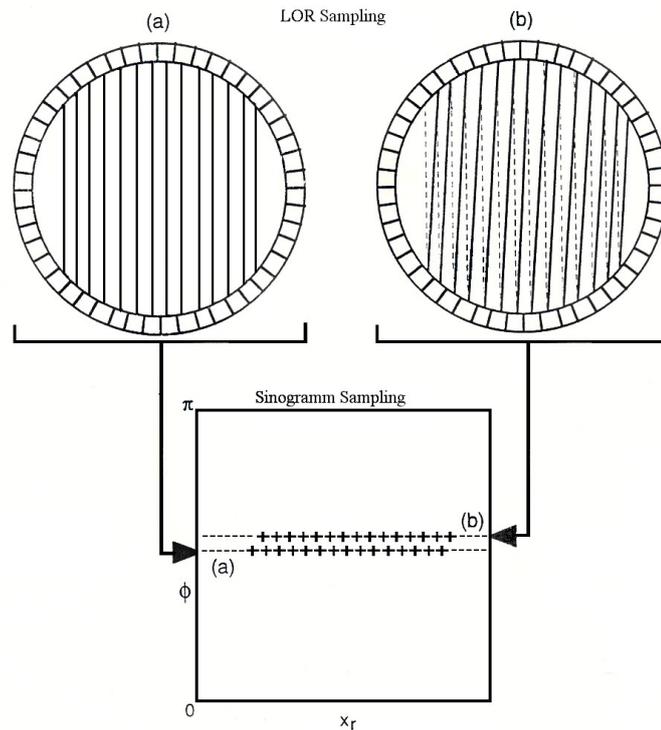


Abbildung 4.4: Sinogramm-Sampling-Muster für einen Scanner mit einer geraden Anzahl horizontaler Detektorelemente. In b) werden die LOR's durch eine Verschiebung von nur einem der beiden Detektorendpunkte verändert. Die ursprünglichen LOR's sind durch gestrichelte Linien dargestellt. a) Die zwei zentralen LOR's überbrücken das genaue Zentrum des FOV. b) Ein LOR geht durch das genaue Zentrum des FOV. Dies führt zu einem Versetzen der x_r -Koordinate bei benachbarten Reihen um $\Delta x_r/2$ im horizontalen Sampling-Muster. Quelle: [6] S. 18.

In der axialen Richtung wird das Sampling im Wesentlichen durch den Zentrum-zu-

Zentrum Abstand Δdz der Detektorelemente in axialer Richtung bestimmt, wobei

$$\Delta dz = \frac{L_D}{N_R}$$

wobei N_R die Anzahl der Detektorringe und L_D die Höhe des Detektors ist. Damit erhält man N_R^2 mögliche Ringkombinationen für schräge LOR's.

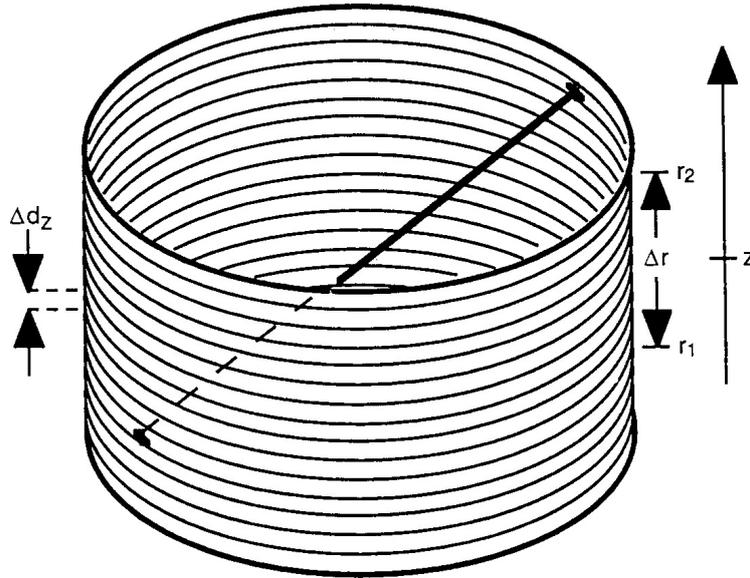


Abbildung 4.5: Die $(z, \Delta r)$ Koordinaten eines schrägen Sinogramms. Quelle: [6] S. 19.

Diese Kombinationen können durch (r_1, r_2) Koordinaten oder durch $(z, \Delta r)$ Koordinaten mit einer durchschnittlichen axialen Koordinate $z = (r_1 + r_2)/2$ und der **Ringdifferenz** $\Delta r = (r_1 - r_2)$ dargestellt werden. Die Beziehung zwischen Projektionen und schrägen Sinogrammen ist gegeben durch:

$$g(x_r, y_r, \phi, \theta) = s(x_r, \phi, z, \Delta r),$$

$$\text{mit } y_r = z \cos \theta \text{ und } \tan \theta = \frac{\Delta r}{2\sqrt{R_D^2 - x_r^2}}, \quad (4.13)$$

wobei R_D der Radius des Ringdetektors ist.

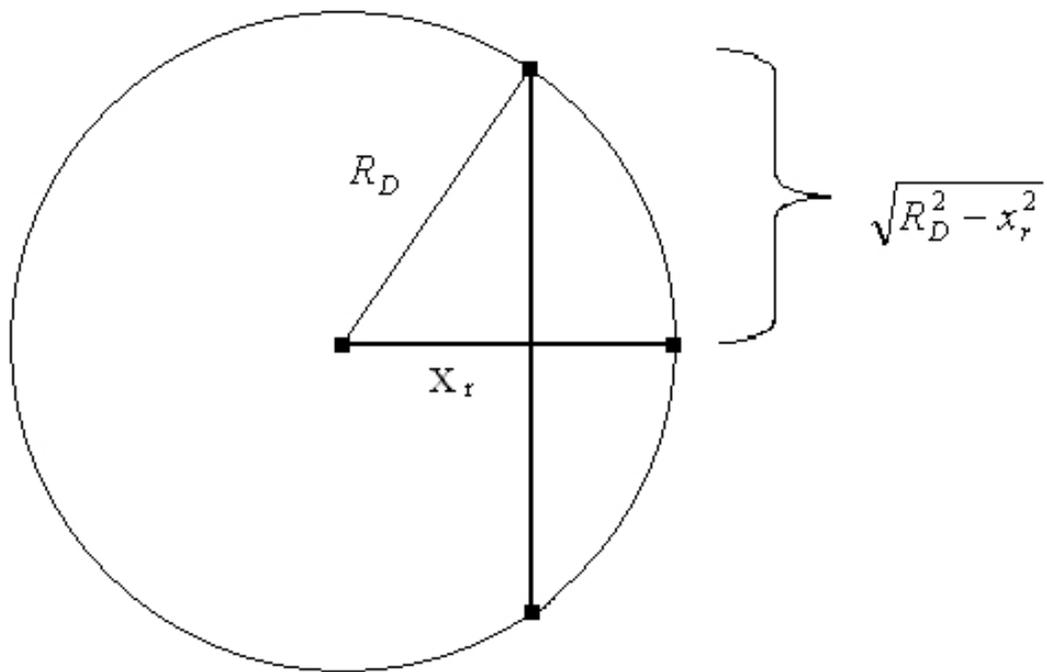


Abbildung 4.6: Erklärung des Nenners von $\tan \theta = \frac{\Delta r}{2\sqrt{R_D^2 - x_r^2}}$.

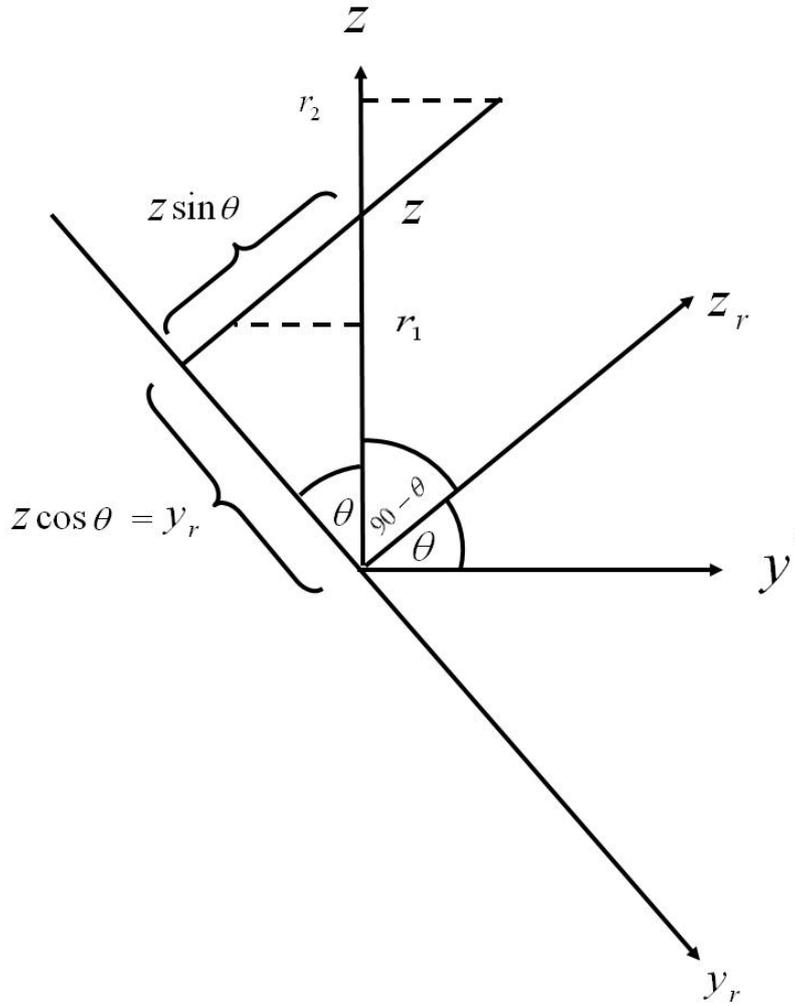


Abbildung 4.7: Erklärung der Variablen y_r .

Wenn der Radius des FOV, R , und die maximale axiale Ausdehnung des Tomographen - verglichen mit R_D - klein sind, dann kann man $\tan \theta$ durch $\tan \theta \approx \frac{\Delta r}{2R_D}$ approximieren. Dadurch ist θ nicht länger von x_r abhängig und jede Reihe im Sinogramm ist eine 1D parallele Projektion.

So kann ein schräges Sinogramm für gegebene z , Δr -Werte dadurch entstehen, dass man für festes $y_r = z \cos \theta$ (mit $\theta = \arctan \frac{\Delta r}{2R_D}$) eine Reihe aus einer Originalprojektionsebene mit festem Winkel ϕ herausnimmt und diese Reihe bei der entsprechenden ϕ -Koordinate im Sinogramm einsetzt (siehe nachfolgende Abbildung).

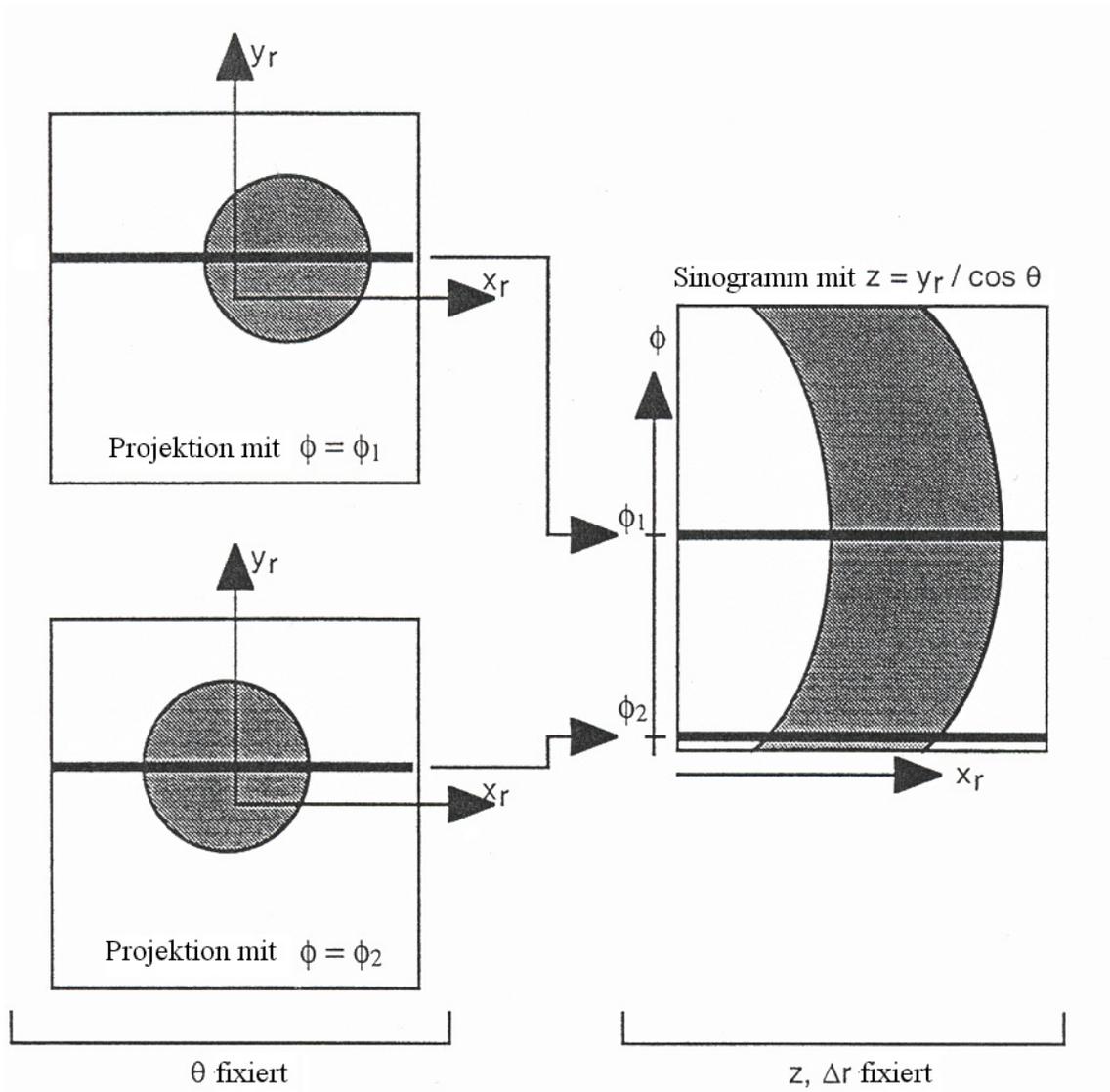


Abbildung 4.8: Übereinstimmung zwischen Projektionen und schrägen Sinogrammen in dem Falle, wo die Approximationen $|x_r|, |y_r| \ll R_D$ gültig sind. Quelle: [6] S. 20.

Wenn die oben genannten Approximationen nicht gelten, müssen die schrägen PET-Sinogramm-Daten interpoliert werden, um echte parallele Projektionen zu erzielen. Von nun an nehmen wir an, dass die 2D und 3D Projektionen exakt die wahren Orte der LOR's reflektieren.

4.3 Datenreduktion

Wir folgen [6] und erkennen Folgendes:

Die gesamte Größe der Projektionsdatenmenge ist ungefähr $N_T \approx N_R^2 \times (N_d/2) \times N_b$, wobei N_b die Anzahl der Bytes pro abgespeichertem Bin und N_d die gerade Anzahl der horizontalen Detektorelemente ist.

Wegen des großen Umfangs einer gesamten 3D PET Datenmenge ist es wichtig, über das Reduzieren der Größe der Datenmenge und die damit verbundenen Auswirkungen Bescheid zu wissen.

In der Praxis wird die maximale räumliche Frequenz, die wiedererlangt werden kann, normalerweise durch eine Abschneidefrequenz $\nu_c < \nu_N$ (ν_N ist die **Nyquistfrequenz**) eingeschränkt, um den Effekt des statistischen Rauschens zu reduzieren. In diesem Fall kann die Anzahl der abgetasteten Winkel N_ϕ reduziert werden, ohne die Bildqualität ungünstig zu beeinflussen.

Die reduzierte Anzahl N'_ϕ ist typischerweise durch $N'_\phi = N_\phi \cdot 2^{-m}$ mit $\Delta\phi' = \Delta\phi 2^m$ gegeben, wobei $m = 0, 1, 2 \dots$ ist.

Um die Anzahl der abgetasteten Winkel um den Faktor 2 zu reduzieren ($m = 1$), wird jede zweite Zeile im Sinogramm zu der vorhergehenden Zeile dazuaddiert und so die Anzahl der Reihen im Sinogramm halbiert. (Die Anzahl der Zählungen pro Reihe steigt hingegen). Diese Methode wird als **Sinogramm-Mashing** bezeichnet.

Das bedeutet, dass die Originaldaten, die im Sinogramm vorliegen, nur noch halbsoviel Speicherplatz benötigen. Sinogramm-Mashing wird in Faktoren von 1/2 durchgeführt. Der **Mashing-Faktor** gibt an, wie viele dieser Faktoren angewendet wurden. Ein Mashing-Faktor von 1 (2 oder 3) reduziert die Daten also um den Faktor 2 (4 oder 8). Es wäre ebenso möglich die Anzahl der Abtastungen in x_r -Richtung durch wachsendes Δx_r zu reduzieren, jedoch ist die Bildqualität empfindlicher gegenüber diesem Typ der Datenreduktion und deshalb wird er selten angewendet.

Für die 3D PET Bildgebung ist die Analyse zur Bestimmung eines aus- bzw. hinreichenden Samplings in allen vier Koordinaten sehr komplex, und eine vollständige Analyse ist bis jetzt noch nicht veröffentlicht. Die Anzahl der Abtastungen in den x_r -, y_r -Koordinaten wird i. Allg. nicht (von der eigentlichen Datensamplinggröße) reduziert. Für die ϕ -Koordinate kann man Mashing, wie oben beschrieben, durchführen. In der θ -Richtung ist die Situation allerdings ein wenig komplexer. Ein verwendbares Hilfsmittel um Mashing in der θ -Koordinate graphisch zu illustrieren, ist das **Michelogramm**. Dies ist die graphische Darstellung der Menge aller erlaubten axialen Datenkombinationen, wobei die Detektorringe der einen Seite des Scanners auf der x -Achse und die Detektorringe der gegenüberliegenden Seite auf der y -Achse aufgetragen werden. Erlaubte Koinzidenzen werden durch einen Punkt in dem Kästchen markiert, das den Schnittpunkt von Zeile und Spalte, die mit der speziellen Detektorringkombination assoziiert werden, repräsentiert.

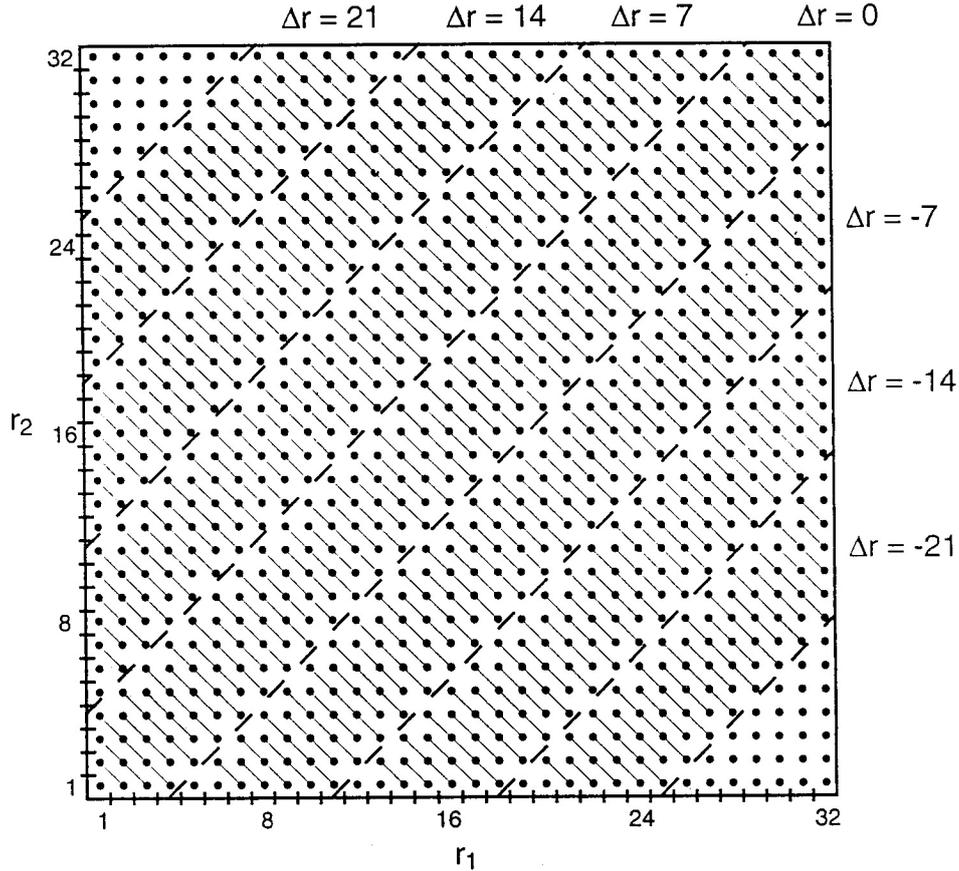


Abbildung 4.9: Illustration eines Michelogramms für einen hypothetischen 32-Ring Scanner. Jeder Punkt repräsentiert das Sinogramm für das Ringpaar (r_1, r_2) . Sinogramme mit denselben z -Werten und angrenzenden θ -Werten werden zusammengefasst/gruppirt, um ein in θ -Richtung größeres Bin zu formen. Dieses Zusammenfassen wird durch die diagonalen Linien, die die Punkte miteinander verbinden, angezeigt. In diesem Beispiel wurde ein Span von 7 gewählt, was zu 7 Segmenten ($\Delta r = \{0, \pm 7, \pm 14, \pm 21\}$) und einer maximalen Ringdifferenz Δr_{max} von 24 führt. Quelle: [6] S. 21.

Um die Anzahl der θ -Abtastungen zu reduzieren, werden in θ -Richtung benachbarte LOR's zusammengefasst. Dabei wird die Größe $\Delta\theta$ des Projektionsbins in der θ -Richtung **Span** genannt. Er ist also ein Ausdruck, um das Ausmaß der axialen Datenkombinationen zu beschreiben. Die maximale axiale Ausdehnung von Koinzidenzen, die in der 3D PET erlaubt ist, heißt **maximale Ringdifferenz**. Dieser maximale Wert von Δr kann auch auf einen Wert geringer als $N_R - 1$ reduziert werden, obwohl es einen Zielkonflikt zwischen allgemeiner Sensitivität und Rekonstruktionszeit gibt. Nachdem man einen maximalen Wert für Δr gewählt hat sind nur noch gewisse Spans möglich, da das resultierende N_θ ein Integer-Wert sein muss.

Als einen abschließenden Punkt hierzu sollten wir noch bemerken, dass wachsendes $\Delta\theta$ nicht notwendigerweise zu Aliasing-Artefakten (durch untersampeln) führen muss, da das Bild nur von den geraden Projektionen ($\theta = 0$) rekonstruiert werden kann, indem man 2D Rekonstruktionsmethoden benutzt (siehe nächster Abschnitt). Wenn $\Delta\theta$ wächst, wird das Bild unter geometrischen Verzerrungen und Unregelmäßigkeiten leiden, die ähnlich zu denen sind, die durch SSRB-Methoden (Single-Slice-Rebinning-Methoden, siehe Seite 116) verursacht werden.

4.4 2D PET

Auch in diesem Abschnitt - inklusive Unterabschnitte - richten wir uns ausschließlich nach [6].

Werden PET-Daten auf eine Art und Weise erworben, so dass sie schon von sich aus in einer Menge paralleler, 2D, horizontaler Schichten organisiert sind und können diese Schichten unabhängig voneinander mit 2D Algorithmen rekonstruiert werden, so wird dies als **2D PET** bezeichnet.

Wir können die 2D Bildrekonstruktionsmethoden in zwei Klassen einteilen:

- iterative Methoden und
- transformierte oder analytische Methoden.

Transformierte Methoden basieren auf diskreten Implementierungen analytischer Lösungen und liefern eine "Einschrittlösung". Das bedeutet, dass die Bildschätzung direkt von den Daten berechnet wird.

Iterative Methoden dagegen beziehen die diskrete Natur des Datensamplings und des Rekonstruktionsproblems sowie statistische Modelle des Datenerwerbsprozesses mit ein. Das so erhaltene Gleichungssystem ist jedoch so groß, dass man bei diesen Methoden versucht eine akzeptable Lösung durch eine Reihe nacheinander gebildeter Schätzungen anzunähern.

Das bei den transformierten Methoden zu Grunde liegende Modell ist das in Abschnitt 4.1 vorgestellte C-C Modell. (Auf die iterativen Methoden werden wir erst in Kapitel 5 genauer eingehen).

Eine der bekanntesten und gebräuchlichsten transformierten Bildrekonstruktionsmethoden im 2D ist die 2D gefilterte Rückprojektion (filtered backprojection, kurz: 2D FBP). Alle anderen 2D transformierten Methoden können zu der 2D FBP-Methode in Beziehung gesetzt werden. Aus diesem Grund wollen wir im nächsten Abschnitt das Verfahren der 2D FBP herleiten und im übernächsten Abschnitt zeigen, wie die 2D FBP zur CBP-Methode (convolution with backprojection) in Beziehung gesetzt werden kann.

4.4.1 2D FBP

Um dieses Verfahren herzuleiten benötigen wir als erstes das Resultat des **Projektionssatzes** 2.26 für den Fall $n = 2$, welcher besagt, dass die 1D FT einer Projektion mit festem Winkel ϕ den Werten auf der Geraden durch den Ursprung (mit demselben Winkel ϕ) der 2D FT von $f(x, y)$ entspricht.

Im Falle $n = 2$ mit $\omega = (\cos \phi, \sin \phi)^\top$ und $v_{x_r} := \sigma$ lautet der Projektionssatz 2.26 also:

$$\hat{g}_1(v_{x_r}, \phi) = \underbrace{(2\pi)^{1/2} \hat{f}_2(v_{x_r} \cos \phi, v_{x_r} \sin \phi)}_{=(2\pi)^{1/2} \hat{f}_2(v_x, v_y)|_{v_{y_r}=0}}, \quad (4.14)$$

wobei wir auf der linken Seite die 1D FT bzgl. x_r und auf der rechten Seiten die 2D FT bzgl. x und y haben.

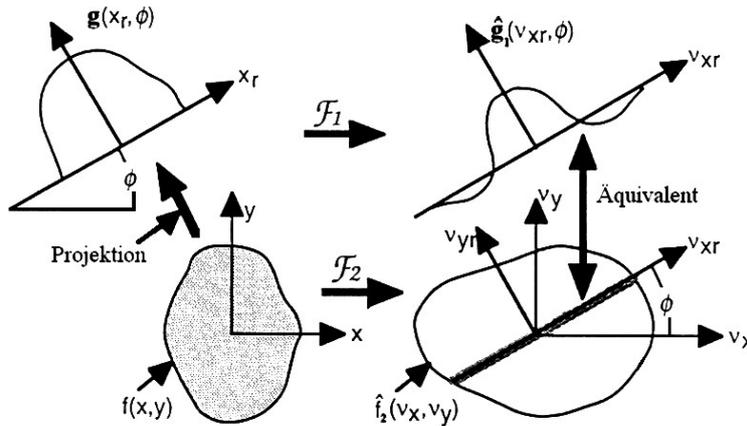


Abbildung 4.10: Illustration des 2D Projektionssatzes. Quelle: [6] S. 23.

Um nun $f(x, y)$ bestimmen zu können benötigen wir $g(x_r, \phi)$ für alle $|x_r| \leq R$ und für alle ϕ mit $0 \leq \phi < \pi$.

Dies folgt, da die FT auf $\mathcal{S}(\mathbb{R}^2)$ eine bijektive Abbildung ist, denn hat man $\hat{g}_1(v_{x_r}, \phi)$ für alle ϕ , so kann man mit Hilfe des Projektionssatzes (4.14) $\hat{f}_2(v_x, v_y)$ für alle (v_x, v_y) bestimmen und somit kann man, da \mathcal{F}^{-1} bijektiv ist, auf jedes $f(x, y)$ eindeutig zurückschließen.

Anmerkung 4.1. Im Folgenden nehmen wir nun an, dass $f \in \mathcal{S}(\Omega)$, Ω FOV, $f = 0$ außerhalb des FOV und somit $g = 0$ für $x_r > R$ ist. Ausserdem soll $g(x_r, \phi)$ für alle $|x_r| \leq R$ und $0 \leq \phi < \pi$ gemessen worden sein, also bekannt sein.

Dies bedeutet, dass die Daten für den 2D Fall **komplett** sind. (Eine genauere Erklärung des Begriffes "komplett" kann in Abschnitt 4.5 nachgelesen werden.)

Satz und Definition 4.2. (2D FBP und Ramp-Filter)

Es gelten die Annahmen aus Anmerkung 4.1. Dann gilt:

$$f(x, y) = (2\pi)^{-3/2} \int_0^\pi \int_{\mathbb{R}} \hat{g}_1(v_{x_r}, \phi) |v_{x_r}| e^{ix_r v_{x_r}} dv_{x_r} d\phi. \quad (4.15)$$

In diesem Fall heisst $|v_{x_r}|$ der **Ramp-Filter**.

Man nennt diese Rekonstruktionsmethode die **2D gefilterte Rückprojektion**, kurz: **2D FBP**.

Beweis. Wir wenden die inverse FT (2.6) an und erhalten:

$$f(\underbrace{x, y}_{=: \tilde{x}}) = (2\pi)^{-1} \int_{\mathbb{R}^2} \hat{f}(\xi) e^{i\langle \tilde{x}, \xi \rangle} d\xi.$$

Wenn wir nun das Integral über \mathbb{R}^2 durch ein Sphärenintegral ausdrücken, indem wir Polarkoordinaten einführen, dann erhalten wir für $\xi = \sigma\omega$ mit $\omega \in S^1$

$$\begin{aligned} & \stackrel{[12] \text{ S. 188}}{=} (2\pi)^{-1} \int_0^\infty \sigma \int_{S^1} \hat{f}(\sigma\omega) e^{i\langle \tilde{x}, \sigma\omega \rangle} d\omega d\sigma \\ & \stackrel{(2.14)}{=} (2\pi)^{-1} \int_0^\infty \sigma \int_{S^1} (2\pi)^{-1/2} (R_\omega f)_1^\wedge(\sigma) e^{i\langle \tilde{x}, \sigma\omega \rangle} d\omega d\sigma \end{aligned}$$

Für $v_{x_r} := \sigma$ und ω wie gehabt erhalten wir (da $\tilde{x}^\top \omega = x_r$ und $(R_\omega f)_1^\wedge(\sigma) = \hat{g}_1(\sigma, \phi)$)

$$= (2\pi)^{-3/2} \int_0^{2\pi} \int_0^\infty \hat{g}_1(v_{x_r}, \phi) e^{ix_r v_{x_r}} v_{x_r} dv_{x_r} d\phi.$$

Wir substituieren im Integral von π bis 2π $v_{x_r} = -v_{x_r}$, $\phi = \phi + \pi$ und benutzen die Symmetrie $\hat{g}_1(-v_{x_r}, \phi + \pi) = \hat{g}_1(v_{x_r}, \phi)$ und erhalten schließlich:

$$= (2\pi)^{-3/2} \int_0^\pi \int_{\mathbb{R}} \hat{g}_1(v_{x_r}, \phi) |v_{x_r}| e^{ix_r v_{x_r}} dv_{x_r} d\phi.$$

□

Der Wert des Bildes im Punkt (x, y) kann also durch Filterung der Projektion im Frequenzraum mit dem Ramp-Filter, 1D inverser FT der gefilterten Projektion und anschließender Rückprojektion erzielt werden.

Man beachte, dass die Rückprojektion für $n = 2$ mit $\omega = (\cos \phi, \sin \phi)^\top$ folgendermaßen lautet:

$$\begin{aligned}
 (R^*g) \underbrace{(x, y)}_{:=\tilde{x}} &\stackrel{(2.15)}{=} \int_{S^1} g(\omega, \underbrace{\tilde{x}^\top \omega}_{=:x_r}) d\omega \\
 &= \int_{S^1} g(\omega, x_r) d\omega \\
 &= \int_0^\pi g(x_r, \phi) d\phi.
 \end{aligned} \tag{4.16}$$

Algorithmus

Der Algorithmus besteht also also folgenden Schritten:

1. Fouriertransformation der Projektion für ein gegebenes ϕ :

$$\hat{g}_1(v_{x_r}, \phi) = (2\pi)^{-1/2} \int_{\mathbb{R}} g(x_r, \phi) e^{-ix_r v_{x_r}} dx_r$$

2. Filterung der Projektion im Frequenzraum:

$$\hat{g}_1^F(v_{x_r}, \phi) = |v_{x_r}| \hat{g}_1(v_{x_r}, \phi)$$

3. Inverse Fouriertransformation der gefilterten Projektion:

$$g^F(x_r, \phi) = (2\pi)^{-1/2} \int_{\mathbb{R}} \hat{g}_1^F(v_{x_r}, \phi) e^{ix_r v_{x_r}} dv_{x_r}$$

4. Rückprojektion der gefilterten Projektion:

$$f(x, y) = f(x, y) + (\Delta\phi g^F(x_r, \phi)) \tag{4.17}$$

5. Wiederhole die Schritte 1) - 4) für jedes ϕ mit $0 \leq \phi < \pi$.

6. Multiplikation von $f(x, y)$ mit $(2\pi)^{-1}$

In Gleichung (4.17) erhalten wir x_r durch das Lösen des Gleichungssystems (4.6). Da ϕ fest ist ist die 2×2 -Matrix bekannt und wenn wir den Punkt $(x, y)^\top$ haben wollen dann müssen wir das entsprechende x_r dazu ausrechnen.

Allerdings gibt es mehrere Probleme mit diesem Algorithmus:

1. Obwohl die Dichte als nicht negativ bekannt ist, liefert der Algorithmus negative Werte, insbesondere, wenn die Daten verrauscht sind.
2. Modelle für die Detektorantwort müssen räumlich invariant sein und können nur als eine Faltung mit begleitender Rauschverstärkung in den Algorithmus integriert werden.
3. Das Powerspektrum der exakten, rauschfreien Projektionen fällt typischerweise schneller als das entsprechende Powerspektrum des statistischen Rauschens. Deshalb wird die FT von $\hat{g}_1(v_{x_r}, \phi)$ oberhalb gewisser Frequenzen, welche vom Verhältnis Signal-zu-Rauschen (Signal-to-noise ratio, kurz: SNR) abhängen, vom Rauschen dominiert und die Anwendung des Ramp-Filter auf höhere Frequenzen resultiert in einer Verstärkung der Rauschstärke, was wiederum die SNR des rekonstruierten Bildes vermindert.

Um diese Probleme zu vermeiden wird ein Tiefpass-Filter $W(v_{x_r})$ benutzt. Somit werden die Beiträge von Frequenzen über einer vorbestimmten Abschneidefrequenz v_c eliminiert. Die geeignete Wahl von v_c wird durch die Samplingdistanz Δx_r und die Photonenstatistik bestimmt. Die Wahl der Funktion $W(v_{x_r})$ und der Abschneidefrequenz v_c wird die räumliche Auflösung des rekonstruierten Bildes sowie seine Rauschcharakteristik bestimmen.

Eine typische "Window-Funktion", die für PET-Rekonstruktionen benutzt wird, ist das allgemeine **Hamming-Window**:

$$W_H(v_{x_r}) = \begin{cases} \alpha + (1 - \alpha) \cos\left(\frac{\pi v_{x_r}}{v_c}\right) & ; |v_{x_r}| < v_c \\ 0 & ; |v_{x_r}| \geq v_c, \end{cases}$$

wobei α ($1/2 \leq \alpha \leq 1$) die Form und $v_c \leq v_N$ die maximale Frequenz kontrolliert.

Korollar 4.3. *Es gelten die Annahmen aus Anmerkung 4.1. Damit folgt:*

$$f(x, y) \approx (2\pi)^{-3/2} \int_0^\pi \int_{\mathbb{R}} \hat{g}_1(v_{x_r}, \phi) |v_{x_r}| W(v_{x_r}) e^{ix_r v_{x_r}} dv_{x_r} d\phi, \quad (4.18)$$

wobei $W(v_{x_r})$ ein Tiefpassfilter ist (z. B. das Hamming-Window).

In obigem Korollar wurde die Gleichheit durch eine Schätzung ersetzt, was auf den Verlust der hochfrequenten Terme, verursacht durch den Gebrauch von $W(v_{x_r})$, zurückzuführen ist. (4.18) repräsentiert nun den 2D FBP Algorithmus, der die gebräuchlichste Methode der Bildrekonstruktion in der 2D PET Bildgebung ist.

Der abschließende Schritt besteht nun darin, aus den 2D Bildern ein "Volumenbild" zu gestalten.

Dabei werden mit Hilfe von Septen benachbarte Ebenen dargestellt, wobei jede dieser Ebenen separat rekonstruiert wurde. Diese rekonstruierten Ebenen werden dann "gestapelt", um ein Volumenbild zu erhalten.

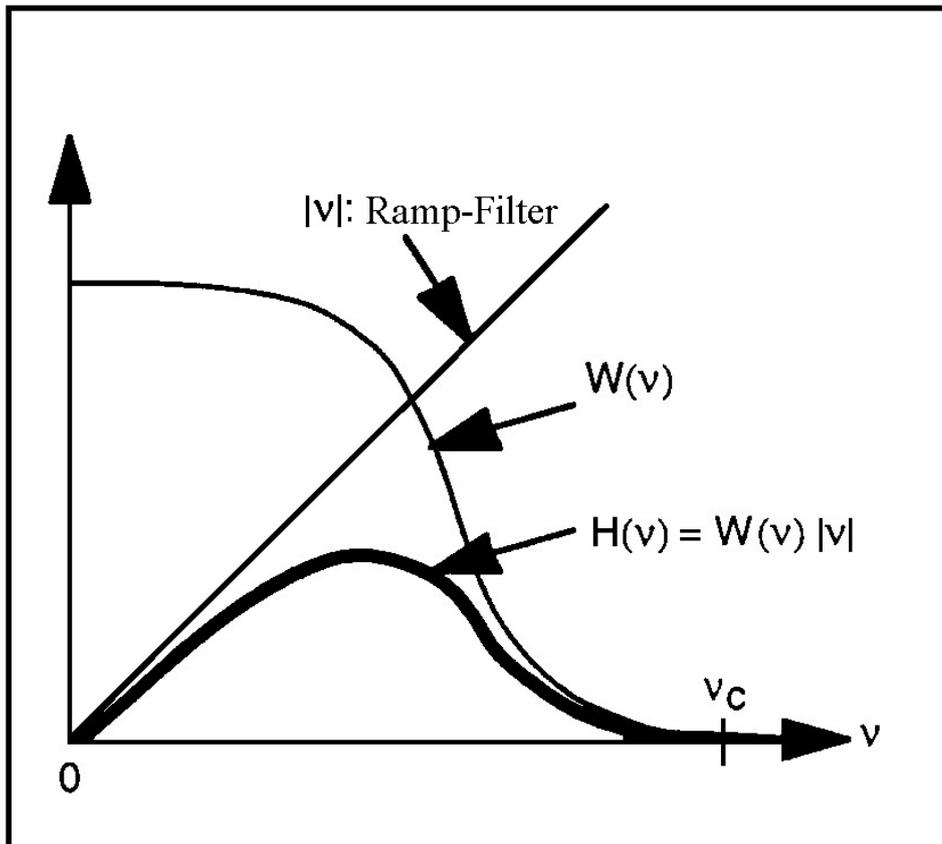


Abbildung 4.11: Modifikation des Ramp-Filters durch Multiplikation mit einem Hamming-Window. Quelle: [6] S. 28.

4.4.2 2D CBP

Um dieses Verfahren herzuleiten und in Beziehung zur 2D FBP zu setzen, müssen wir uns als erstes das Resultat des Faltungssatzes (2.3) in Erinnerung rufen. Dann können wir folgenden Satz formulieren:

Satz und Definition 4.4. (2D CBP)

Es seien dieselben Voraussetzungen wie in den Sätzen zuvor gegeben. Dann gilt:

$$f(x, y) \approx (2\pi)^{-3/2} \int_0^\pi g(x_r, \phi) * h(x_r) d\phi, \quad (4.19)$$

wobei $h(x_r) := (\mathcal{F}_1^{-1}W(v_{x_r}) | v_{x_r} |)(x_r)$ ist.

Diese Rekonstruktionsmethode wird **Faltung mit Rückprojektion** (convolution with backprojection, kurz: **CBP**) genannt.

Beweis. Aus (4.18) folgt:

$$f(x, y) \approx (2\pi)^{-3/2} \int_0^\pi \underbrace{\int_{\mathbb{R}} \hat{g}_1(v_{x_r}, \phi) | v_{x_r} | W(v_{x_r}) e^{ix_r v_{x_r}} dv_{x_r}}_{=(2\pi)^{1/2}(\mathcal{F}_1^{-1}\{W(v_{x_r})|v_{x_r}|\hat{g}_1(v_{x_r}, \phi)\})(x_r, \phi)} d\phi. \quad (4.20)$$

z.z.: $(2\pi)^{-1/2}g(x_r, \phi) * \mathcal{F}_1^{-1}\{W(v_{x_r}) | v_{x_r} |\} = \mathcal{F}_1^{-1}\{W(v_{x_r}) | v_{x_r} | \mathcal{F}_1\{g(x_r, \phi)\}\}$

dazu: Aus dem Faltungssatz folgt:

$$\mathcal{F}_1\{g(x_r, \phi) * \mathcal{F}_1^{-1}\{W(v_{x_r}) | v_{x_r} |\}\} = (2\pi)^{1/2}\mathcal{F}_1\{g(x_r, \phi)\} \cdot \underbrace{\mathcal{F}_1\{\mathcal{F}_1^{-1}\{W(v_{x_r}) | v_{x_r} |\}\}}_{=W(v_{x_r})|v_{x_r}|}.$$

Damit folgt:

$$\mathcal{F}_1^{-1}\{\mathcal{F}_1\{g(x_r, \phi) * \mathcal{F}_1^{-1}\{W(v_{x_r}) | v_{x_r} |\}\}\} = \mathcal{F}_1^{-1}\{(2\pi)^{1/2}\mathcal{F}_1\{g(x_r, \phi)\}W(v_{x_r}) | v_{x_r} |\}$$

und wir erhalten:

$$g(x_r, \phi) * \mathcal{F}_1^{-1}\{W(v_{x_r}) | v_{x_r} |\} = (2\pi)^{1/2}\mathcal{F}_1^{-1}\{W(v_{x_r}) | v_{x_r} | \mathcal{F}_1\{g(x_r, \phi)\}\}.$$

Setzen wir diese Überlegungen in die Gleichung (4.20) ein so ergibt sich:

$$\begin{aligned} f(x, y) &= (2\pi)^{-3/2} \int_0^\pi (2\pi)^{1/2}\mathcal{F}_1^{-1}\{W(v_{x_r}) | v_{x_r} | \mathcal{F}_1\{g(x_r, \phi)\}\}(x_r, \phi) d\phi \\ &= (2\pi)^{-3/2} \int_0^\pi g(x_r, \phi) * \mathcal{F}_1^{-1}\{W(v_{x_r}) | v_{x_r} |\}(x_r) d\phi \\ &= (2\pi)^{-3/2} \int_0^\pi g(x_r, \phi) * h(x_r) d\phi. \end{aligned}$$

□

4.5 3D PET

Ebenso wie in den Abschnitten zuvor folgen wir hier [6] und [10].

Wie im 2D Fall ist auch hier bei den transformierten Methoden das zu Grunde liegende Modell das in Abschnitt 4.1 vorgestellte C-C Modell. Bei dem auf diesem Modell beruhenden Bildrekonstruktionsansatz hat die Geometrie der Datensammlung einen Haupteinfluss auf alle Aspekte des Ansatzes, nämlich auf das Modell selbst, auf die auf dem Modell basierende Inversionsformel und die numerische Beurteilung der Inversionsformel.

Im Folgenden behandeln wir nun einige der spezifischen Probleme, die im 3D Fall auftreten, wenn der Ansatz auf dem C-C Modell basiert.

Die zwei Hauptunterschiede zwischen der 2D und der 3D PET Bildgebung sind die **Datenredundanz** und die **räumliche Varianz**. Diese beiden Punkte verkomplizieren die 3D Bildrekonstruktion im Vergleich zum 2D Fall. (Außer den beiden oben genannten Unterschieden gibt es natürlich auch noch andere Abweichungen, z. B. die wachsende Anzahl an Streukoinzidenzen im 3D, etc. Wir nehmen jedoch an, dass diese Effekte bereits korrigiert wurden und dass die Projektionsdaten weiterhin den Linienintegralen über die 3D Tracerverteilung entsprechen, siehe (4.4)).

Um die Datenredundanz der 3D PET Bildgebung zu verdeutlichen, betrachten wir die normale 2D Bildgebung eines Volumenbildes. Dort (siehe z. B. 2D FBP (4.15)) hatten wir ausreichende Informationen, um die Tracerverteilung zu rekonstruieren. (Zur Erinnerung: Zur Erhaltung eines Volumenbildes rekonstruierten wir jede der 2D parallelen Ebenen mit Hilfe einer 2D Rekonstruktionsmethode separat für sich und "stapelten" dann diese 2D parallelen Ebenen zu einem Volumenbild). Bei der Durchführung der 3D Bildgebung desselben Volumenbildes sammeln wir also eine Obermenge der 2D Daten. Deshalb müssen die zusätzlich zu den 2D Daten gesammelten Daten aus redundanten Informationen bestehen. Der springende Punkt hierbei ist, dass die Sammlung dieser redundanten Informationen entlang der zusätzlichen LOR's die Bild-SNR durch eine Reduzierung des statistischen Rauschens verbessern können.

Das zweite geometrische Hauptproblem, das bei der PET paralleler Projektionen mit schrägen Winkeln ($\theta \neq 0$) auftritt, ist die endliche axiale Ausdehnung des Scanners.

Wie wir in der nachfolgenden Abbildung erkennen können, beinhalten Projektionen in schräge Richtungen Regionen, markiert durch x im Schaubild, die schrägen Linien entsprechen, bei denen eines der Photonen den Detektor verfehlt. Solche Projektionen bezeichnet man als "**abgeschnitten**".

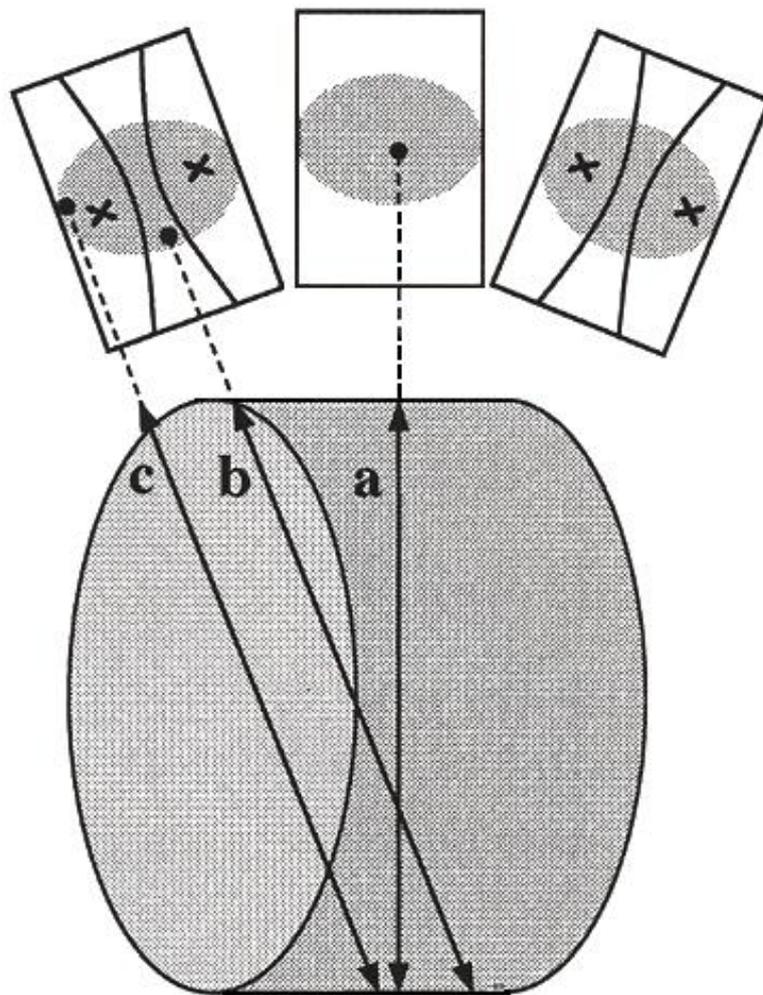


Abbildung 4.12: Illustration der geometrischen Anordnung von Projektionsdaten für Photonenpaare, die nicht streuen. a) repräsentiert ein Photonenpaar, das in nicht-schräger Richtung ($\theta = 0$) ausgestrahlt wird (die Ausstrahlung ist senkrecht zur z -Achse). b) repräsentiert ein Photonenpaar, das in schräger Richtung ($\theta \neq 0$) ausgestrahlt wird. Dabei wird jedes der beiden Photonen auch detektiert. c) repräsentiert ein Photonenpaar, das in schräger Richtung ausgestrahlt wird und wobei eines der beiden Photonen nicht detektiert wird. Quelle: [10] S. 13.

Um die Theorie und die Algorithmen, die zur Rekonstruktion der Tracerverteilung benötigt werden zu entwickeln, müssen wir annehmen, dass $g(x_r, y_r, \phi, \theta)$ für alle $(x_r, y_r) \in \mathbb{R}^2$ bekannt ist. In anderen Worten müssen wir alle LOR's messen, die parallel zu

$$\hat{z}_r := \omega = (\cos \phi \cos \theta, \sin \phi \cos \theta, \sin \theta)^\top$$

sind und den Bereich $D_f = \{(x, y, z); f(x, y, z) \neq 0\}$ schneiden.

Wenn die Projektionen nicht abgeschnitten sind, werden sie **komplett** genannt.

Wären alle Projektionen komplett, dann wäre die Sensitivität des Scanners unabhängig von der Position im FOV, das Rekonstruktionsproblem wäre also verschiebungs-invariant und könnte mit Standard-Fourier-Entwicklungstechniken gelöst werden. Im Gegensatz dazu ist das 2D Rekonstruktionsproblem bei einem Multi-Ring-Scanner, der mit ausgefahrenen Septen operiert, verschiebungs-invariant.

Wie wir in den Abbildungen 4.13 und 4.14 erkennen können, ist eine 2D Projektion, die von einem zylindrischen Scanner gemessen wird, nur dann komplett, wenn $|\theta| < \theta_c$ ist, wobei θ_c von der Größe der Region D_f relativ zu der Scannerdimension abhängt.

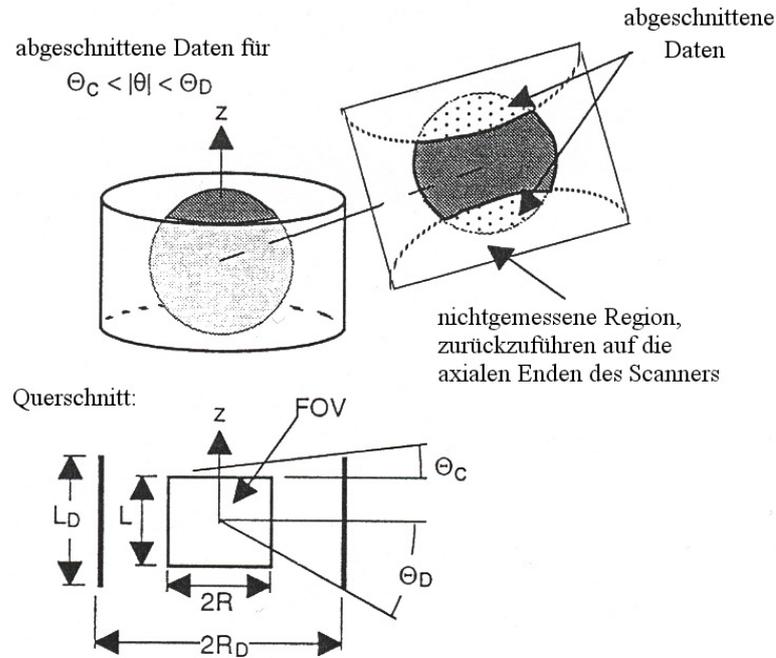


Abbildung 4.13: Illustration der Region des kompletten Datenerwerbs für einen 3D PET Scanner. Quelle: [6] S. 32.

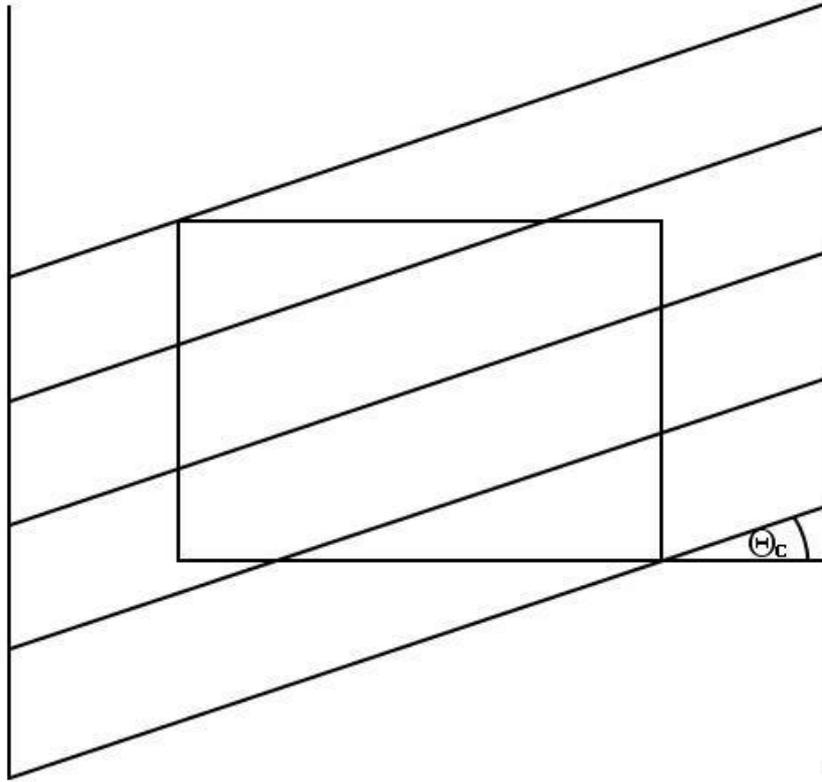


Abbildung 4.14: Nochmalige Darstellung des Winkels Θ_c .

Gewöhnlich erstreckt sich D_f axial über das FOV des Scanners hinaus und θ_c ist dann gleich Null. In solch einem Fall sind die einzigen kompletten Projektionen diejenigen, die einer Ringdifferenz von $\Delta r = 0$ oder $\Delta r = \pm 1$ entsprechen.

Durch die Abgeschnittenheit wächst die Komplexität der Rekonstruktion, weil die gefilterte Rückprojektion nicht auf unkomplette Daten angewendet werden kann. Die Standardlösung zu dieser Schwierigkeit besteht nun darin, in einem ersten Schritt die abgeschnittenen Daten zu schätzen und somit eine Menge kompletter 2D Projektionen entlang aller Richtungen

$$\hat{z}_r(\phi, \theta) = (\cos \phi \cos \theta, \sin \phi \cos \theta, \sin \theta)^\top \quad 0 \leq \phi < \pi, |\theta| \leq \Theta$$

zu erhalten.

Der Schätzungsschritt und die Wahl von Θ ($\Theta_c \leq \Theta \leq \Theta_D$) werden nach dem nächsten Abschnitt diskutiert werden.

Als nächstes beschreiben wir den zweiten Schritt, die Rekonstruktion von $f(x, y, z)$ von kompletten Projektionen.

4.5.1 3D FBP für komplette 2D parallele Projektionen

Dieser Abschnitt beschreibt die Theorie und die Algorithmen, die zur Rekonstruktion der Tracerverteilung $f(x, y, z)$ aus 3D PET-Daten, die aus einer Menge 2D paralleler Projektionen bestehen, benötigt wird.

Wir nehmen in diesem Abschnitt an, dass $g(x_r, y_r, \phi, \theta)$ für $0 \leq \phi < \pi$, $|\theta| \leq \Theta$ und für alle $(x_r, y_r) \in \mathbb{R}^2$ bekannt ist, die Projektionen also komplett sind.

Der Grundstein der 3D Bildrekonstruktion ist der 3D Projektionssatz der Röntgentransformation 2.31, welcher den 2D Projektionssatz aus Abschnitt 4.4.1 verallgemeinert. Dieser besagt, dass die 2D FT von einer Projektion mit festen Winkeln ϕ und θ den Werten auf der Ebene durch den Ursprung senkrecht zu \hat{z}_r der 3D FT von $f(x, y, z)$ entspricht.

Wir haben also für $n = 3$ mit

$$\omega = \begin{pmatrix} \cos \phi \cos \theta \\ \sin \phi \cos \theta \\ \sin \theta \end{pmatrix} \text{ und } v := \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} = \begin{pmatrix} -v_{x_r} \sin \phi - v_{y_r} \cos \phi \sin \theta \\ v_{x_r} \cos \phi - v_{y_r} \sin \phi \sin \theta \\ v_{y_r} \cos \theta \end{pmatrix}, \quad v \in \omega^\perp$$

den Projektionssatz in folgender Form:

$$\underbrace{(2\pi)^{1/2} \hat{f}_3(v)}_{=(2\pi)^{1/2} \hat{f}_3(v_x, v_y, v_z)|_{v_{z_r}=0}} = \hat{g}_2(v_{x_r}, v_{y_r}, \phi, \theta). \quad (4.21)$$

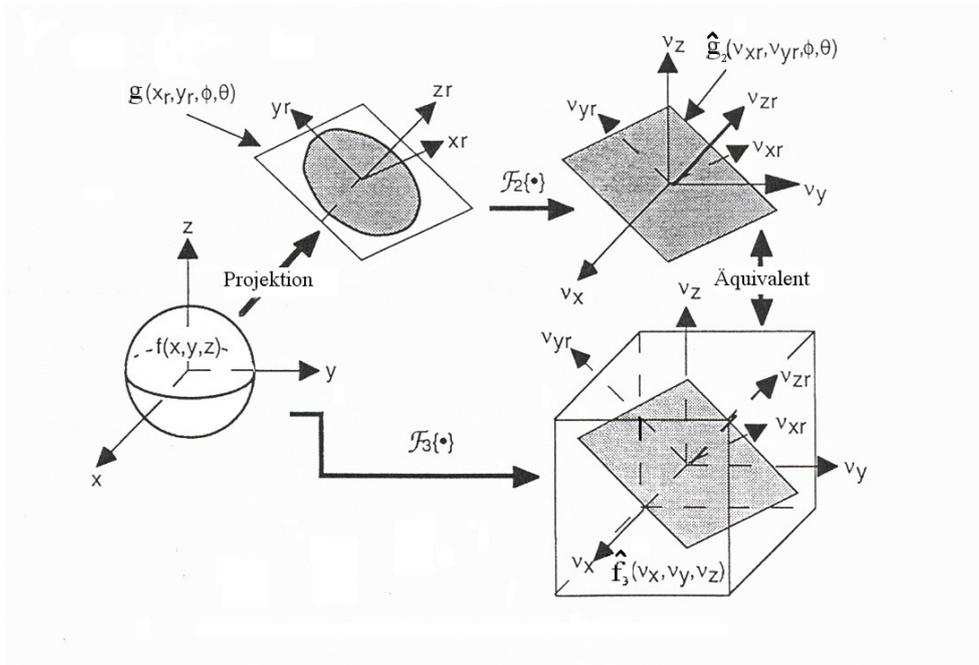


Abbildung 4.15: Illustration des 3D Projektionssatzes. Quelle: [6] S. 33.

Für eine gegebene 2D Projektion senkrecht zu $\hat{z}_r(\phi, \theta)$ kann die Gleichung

$$\hat{g}_2(v_{x_r}, v_{y_r}, \phi, \theta) = (2\pi)^{1/2} \hat{f}_3(v_x, v_y, v_z) |_{v_{z_r}=0}$$

also dazu dienen, die 3D FT des Bildes für eine beliebige Frequenz (v_x, v_y, v_z) , die in der Ebene v_{z_r} liegt, zu berechnen. Wie man leicht sieht, ist jede solche Frequenz orthogonal zur Projektionsrichtung $\hat{z}_r(\phi, \theta)$. Geometrisch bedeutet dies, dass eine 2D parallele Projektion ausreicht, um $\hat{f}_3(v)$ auf einer Ebene, die den Ursprung enthält und orthogonal zu $\hat{z}_r(\phi, \theta)$ (im 3D Bildfrequenzraum) ist, zu erhalten.

Um noch einmal genauer auf die Redundanz der 3D Daten einzugehen, betrachten wir die Abbildung 4.16.

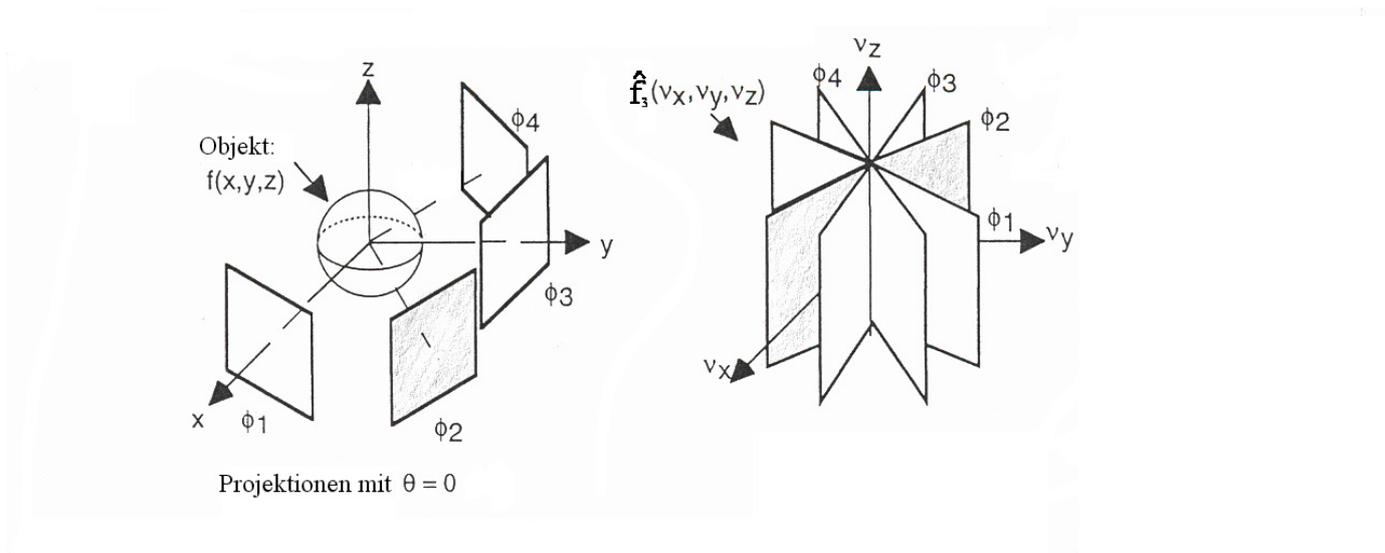


Abbildung 4.16: Illustration, wie der 3D Projektionssatz benutzt werden kann, um $\hat{f}_3(v_x, v_y, v_z)$ aus $g(x_r, y_r, \phi, \theta)$ zu rekonstruieren, indem man nur Projektionen mit $\theta = 0$ benutzt. Quelle: [6] S. 34.

Beim 2D Erwerb, bei dem $g(x_r, y_r, \phi, \theta)$ nur für $\theta = 0$ gemessen wird, liefert der 3D Projektionssatz (4.21) für jedes $\phi \in [0, \pi]$ $\hat{f}_3(v)$ auf einer Ebene, die die v_z -Achse enthält. Für abweichende $\phi \in [0, \pi]$ rotiert diese Ebene um die v_z -Achse und tastet so den gesamten 3D Frequenzraum ab. Das erlaubt einem, die 3D FT von $f(x, y, z)$ zu berechnen und daher das Bild selbst zu rekonstruieren.

Im 2D Fall kann eine gegebene Frequenz $v = (v_x, v_y, v_z)^T$ von einer einzigen Projektion, definiert durch $\phi = \arctan\left(-\frac{v_x}{v_y}\right)$ wiedergefunden werden. Deshalb sind die 2D Daten nicht redundant und dies resultiert in einem, bis auf die Wahl des Regularisierungsfensters, eindeutigen Ramp-Filter.

Um das Sampling der 3D FT in dem allgemeineren Fall der 3D PET Bildgebung zu verstehen, ist es hilfreich die von **Orlov** 1976 eingeführte Notation zu benutzen. Wie in Abbildung 4.17 Teil a) gezeigt wird, kann die Orientierung einer einzelnen, 2D Projektionsebene durch die Lokalisation des Endpunktes des Einheitsvektors $\hat{z}_r(\phi, \theta)$ auf der Einheitskugel beschrieben werden. Die Menge solcher Endpunkte wird mit Ω bezeichnet. In Teil b) der Abbildung wird Ω für die einzelne Projektion aus Teil a) angegeben. Ω_0 aus Teil c) entspricht den Projektionen aus der vorhergehenden Abbildung, während Teil d) Ω_Θ für die generelle 3D Bildgebung zeigt.

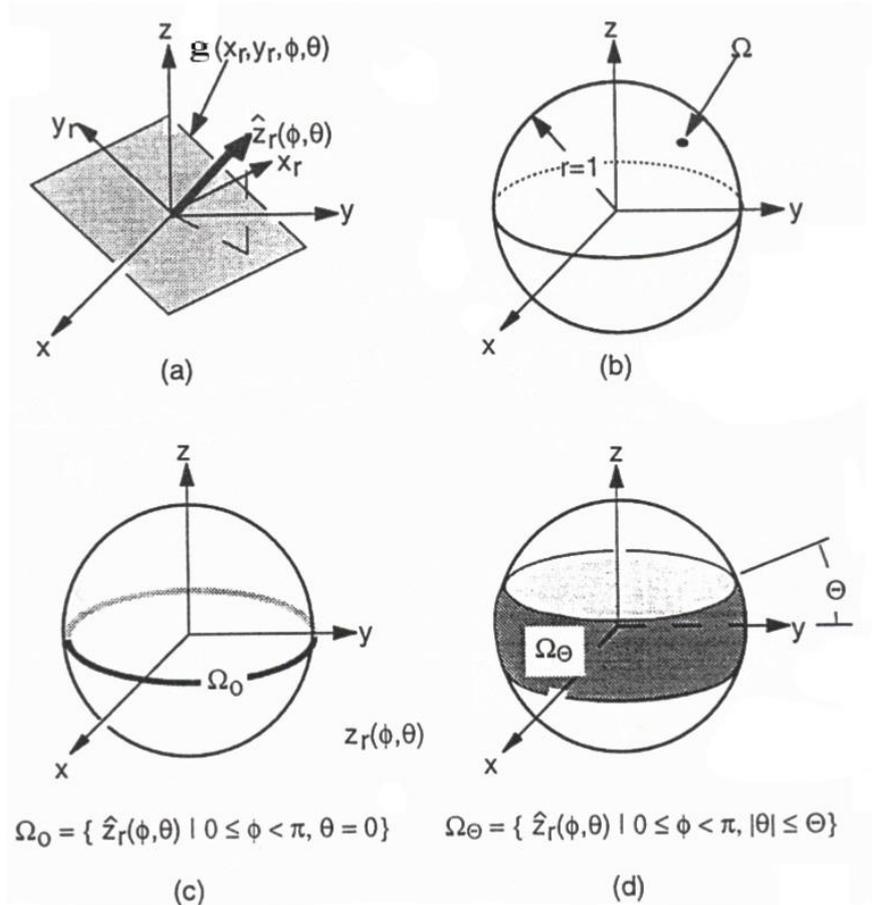


Abbildung 4.17: Orlov-Sphären für verschieden Scan-Geometrien. Quelle: [6] S. 35.

Betrachtet man nun die 3D Datenmenge

$$g(x_r, y_r, \phi, \theta) = \int_{\mathbb{R}} f(x, y, z) dz_r$$

dann folgt:

für gegebenes $\bar{v} := (v_x, v_y, v_z)^\top$ kann man $\hat{f}_3(\bar{v})$ durch die Anwendung des 3D Projek-

tionsatzes auf einer 2D gemessenen Projektion errechnet werden, wobei $\hat{z}_r(\phi, \theta)$ orthogonal zu \bar{v} sein muss.

Wir haben also:

$$\bar{v}^\top \hat{z}_r(\phi, \theta) = v_x \cos \phi \cos \theta + v_y \sin \phi \cos \theta + v_z \sin \theta = 0 \quad (4.22)$$

Im Allgemeinen hat diese Gleichung eine Reihe von Lösungen innerhalb des gemessenen Ranges $0 \leq \phi < \pi$, $|\theta| \leq \Theta$, $\Theta \in [0, \pi/2[$. Deshalb sind die 3D Daten redundant.

Satz 4.5. *Wird eine gegebene Frequenz $\bar{v} = (v_x, v_y, v_z)^\top$ mit Hilfe von Polarkoordinaten (v, α, ψ) ausgedrückt, dann wird aus der Gleichung (4.22) mit $|\theta| \leq \Theta$ und $\Theta \in [0, \pi/2[$ die Gleichung:*

$$\cos(\phi - \alpha) = -\tan \psi \tan \theta, \quad v \neq 0, \psi \in]-\pi/2, \pi/2[. \quad (4.23)$$

Diese Gleichung hat genau dann eine Lösung für ϕ , wenn $|\sin \theta| \leq |\cos \psi|$ ist. Für $\phi \in [0, \pi]$ ist diese Lösung dann auch noch eindeutig.

Beweis. Wenn wir \bar{v} in Polarkoordinaten ausdrücken, so erhalten wir:

$$\bar{v} = (v_x, v_y, v_z)^\top = (v \cos \alpha \cos \psi, v \sin \alpha \cos \psi, v \sin \psi)^\top.$$

Aus der Gleichung (4.22) wird also

$$\begin{aligned} & v_x \cos \phi \cos \theta + v_y \sin \phi \cos \theta + v_z \sin \theta = 0 \\ \Leftrightarrow & v \cos \alpha \cos \psi \cos \phi \cos \theta + v \sin \alpha \cos \psi \sin \phi \cos \theta + v \sin \psi \sin \theta = 0 \\ \stackrel{v \neq 0}{\Leftrightarrow} & \cos \alpha \cos \psi \cos \phi \cos \theta + \sin \alpha \cos \psi \sin \phi \cos \theta + \sin \psi \sin \theta = 0 \\ \Leftrightarrow & (\cos \psi \cos \theta)(\cos \alpha \cos \phi + \sin \alpha \sin \phi) = -\sin \psi \sin \theta \\ \stackrel{\psi, \theta \in]-\frac{\pi}{2}, \frac{\pi}{2}[}{\Leftrightarrow} & \cos \alpha \cos \phi + \sin \alpha \sin \phi = -\frac{\sin \psi \sin \theta}{\cos \psi \cos \theta} \\ \stackrel{\text{Add.theoreme}}{\Leftrightarrow} & \cos(\phi - \alpha) = \tan \psi \tan \theta \end{aligned}$$

" \Leftarrow " Sei $|\sin \theta| \leq |\cos \psi|$, dann folgt:

$$\sin^2 \theta \leq \cos^2 \psi. \quad (4.24)$$

Da

$$\cos^2 \theta = 1 - \sin^2 \theta \stackrel{(4.24)}{\geq} 1 - \cos^2 \psi = \sin^2 \psi$$

folgt:

$$|\sin \psi| \leq |\cos \theta|$$

und damit erhalten wir:

$$\left| \frac{\sin \theta}{\cos \psi} \right| \leq 1 \text{ und } \left| \frac{\sin \psi}{\cos \theta} \right| \leq 1.$$

Es gilt also:

$$\left| \frac{\sin \theta}{\cos \psi} \right| \left| \frac{\sin \psi}{\cos \theta} \right| = \left| \frac{\sin \psi \sin \theta}{\cos \psi \cos \theta} \right| = |\tan \psi \tan \theta| \leq 1,$$

und daher hat die Gleichung

$$\cos(\phi - \alpha) = -\tan \psi \tan \theta$$

eine Lösung für ϕ .

" \Rightarrow " Sei $|\sin \theta| > |\cos \psi|$, dann folgt analog zur Rückrichtung, dass

$$|\sin \psi| > |\cos \theta|$$

ist, und ebenfalls analog zur Rückrichtung erhalten wir dann:

$$|\tan \psi \tan \theta| > 1,$$

und daher hat die Gleichung:

$$\cos(\phi - \alpha) = -\tan \psi \tan \theta$$

keine Lösung für ϕ .

□

Anmerkung 4.6. Im Folgenden nehmen wir an, dass komplette Projektionen vorliegen, d. h. dass für jedes $0 \leq \phi < \pi$ und $|\theta| \leq \Theta$, $g(x_r, y_r, \phi, \theta)$ für alle $(x_r, y_r) \in D_g$ gemessen wurde, also bekannt ist. Dabei ist D_g der Träger der Projektion des Objektes. Der Träger der orthogonalen Projektion eines Zylinders entlang einer Richtung $\hat{z}_r(\phi, \theta)$ wird in Abschnitt 4.5.2 genauer spezifiziert werden.

Des Weiteren nehmen wir an, dass $f \in \mathcal{S}(\Omega)$, Ω das FOV und $f = 0$ außerhalb des FOV ist.

Lemma 4.7. Die Projektionen seien komplett, $\psi \in]-\frac{\pi}{2}, \frac{\pi}{2}[$ und $\Theta \in [0, \frac{\pi}{2}[$.

Dann kann

$$\hat{f}_3(\bar{v}) = (2\pi)^{-3/2} \int_{\mathbb{R}^2} g(x_r, y_r, \phi, \theta) e^{i(x_r v_{x_r} + y_r v_{y_r})} dx_r dy_r \quad (4.25)$$

mit einer beliebigen Projektion (ϕ, θ) berechnet werden, wobei $\theta \in [-\Theta_{lim}, \Theta_{lim}]$ ist, ϕ die eindeutige Lösung von Gleichung (4.23) und

$$\Theta_{lim} := \begin{cases} \Theta & ; & |\psi| < \frac{\pi}{2} - \Theta \\ \frac{\pi}{2} - \psi & ; & |\psi| \geq \frac{\pi}{2} - \Theta \end{cases} \quad (4.26)$$

ist.

Beweis. 1. Sei $\underbrace{|\psi|}_{[0, \frac{\pi}{2}[} < \underbrace{\frac{\pi}{2} - \Theta}_{\in]0, \frac{\pi}{2}]}$, dann ist nach Definition

$\Theta_{lim} = \Theta$ und da der Kosinus streng monoton fallend in $[0, \frac{\pi}{2}]$ ist folgt:

$$\cos |\psi| = \cos \psi > \cos\left(\frac{\pi}{2} - \Theta\right) = \sin \Theta. \quad (4.27)$$

Wir haben also $\theta \in [-\Theta, \Theta]$ beliebig, aber fest.

- Wenn $\theta \in [0, \Theta]$ ist dann folgt, da der Sinus streng monoton wachsend in $[0, \frac{\pi}{2}]$ ist, dass:

$$\sin \theta < \sin \Theta \stackrel{(4.27)}{<} \cos \psi.$$

- Wenn $\theta \in [-\Theta, 0]$ ist dann folgt, da der Sinus streng monoton wachsend in $[-\frac{\pi}{2}, 0]$ ist, dass:

$$\begin{aligned} \sin -\Theta &< \sin \theta \\ \Rightarrow \sin \Theta &> -\sin \theta \\ \Rightarrow -\sin \theta &< \sin \Theta \stackrel{(4.27)}{<} \cos \psi. \end{aligned}$$

Insgesamt erhalten wir also:

$$|\sin \theta| < \cos \psi = |\cos \psi|$$

und damit hat die Gleichung (4.23) eine eindeutige Lösung für $\phi \in [0, \pi]$. Für festes θ mit $|\theta| \leq \Theta$ und zugehöriges $\phi_\theta \in [0, \pi]$, welches Gleichung (4.23) eindeutig löst, sind die Daten nach Voraussetzung für alle $(x_r, y_r) \in D_g$ bekannt (außerhalb gilt: $g = 0$), und deshalb ist die Integration über x_r und y_r möglich und wir erhalten mit Hilfe des Projektionssatzes der Form (4.21)

$$\hat{f}_3(\bar{v}) = (2\pi)^{-3/2} \int_{\mathbb{R}^2} g(x_r, y_r, \phi, \theta) e^{i(x_r v_{x_r} + y_r v_{y_r})} dx_r dy_r. \quad (4.28)$$

Dabei berechnet sich v_{x_r} und v_{y_r} durch:

$$\begin{pmatrix} v_{x_r} \\ v_{y_r} \\ v_{z_r} = 0 \end{pmatrix} = \begin{pmatrix} -\sin \phi & \cos \phi & 0 \\ -\cos \phi \sin \theta & -\sin \phi \sin \theta & \cos \theta \\ \cos \phi \cos \theta & \sin \phi \cos \theta & \sin \theta \end{pmatrix} \begin{pmatrix} x_r \\ y_r \\ z_r \end{pmatrix}.$$

2. Sei nun $|\psi| \geq \frac{\pi}{2} - \Theta$, dann führen wir dieselben Überlegungen vollständig analog zum Fall 1) durch und erhalten ebenso die Behauptung. □

Die folgende Abbildung verdeutlicht die Wahl von Θ_{lim} in Lemma 4.7:

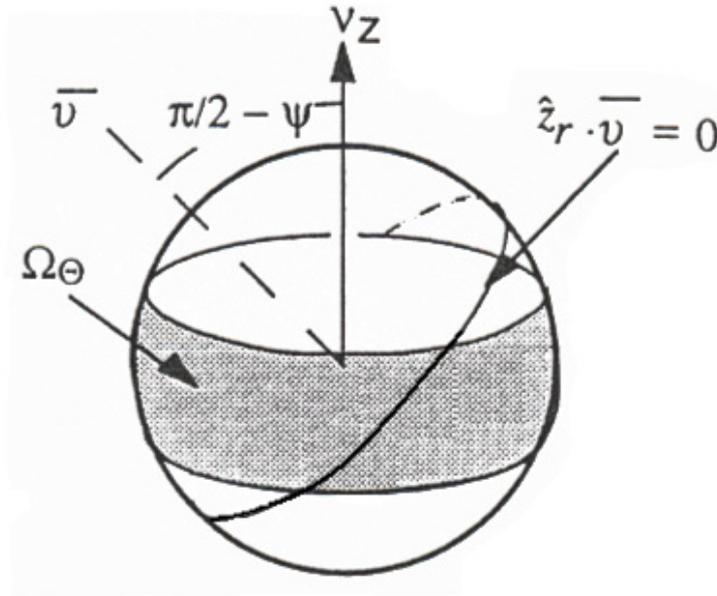


Abbildung 4.18: Koordinaten für die Filtergleichung. Quelle: [6] S. 36.

Die Schnittmenge von Ω_Θ mit den Lösungen der Gleichung

$$\hat{z}_r \bar{v} = 0$$

für gegebenes \bar{v} ist die Menge von Projektionen, von denen $\hat{f}_3(\bar{v})$ berechnet werden kann.

Wenn die Daten konsistent wären, dann würden alle diese unabhängigen Schätzungen von $\hat{f}_3(\bar{v})$ gleich sein.

(Für jedes θ und dazu entsprechende ϕ_θ bekämen wir unabhängig Schätzung von $\hat{f}_3(\bar{v})$.) Da wir jedoch Rauschen in den Daten haben, sind die Daten nicht konsistent und eine optimale SNR wird durch Mittelung über alle verfügbaren Schätzungen von $\hat{f}_3(\bar{v})$ erreicht.

Der Standardansatz führt uns zu einem gewichteten Mittel mit einem Gewichtungsfaktor von

$$\frac{\cos \theta}{\sqrt{\cos^2 \psi - \sin^2 \theta}}$$

und einem Normierungsfaktor von $H(\Theta, \psi)$, welcher sicherstellt, dass $\hat{f}_3 = 1$ ist, wenn $\hat{g}_2 = 1$ ist.

Wir erhalten also:

$$\hat{f}_3(\bar{v}) = H(\Theta, \psi) \int_{-\Theta_{lim}}^{\Theta_{lim}} \hat{g}_2(v_{x_r}, v_{y_r}, \phi, \theta) \frac{2 \cos \theta}{\sqrt{\cos^2 \psi - \sin^2 \theta}} d\theta. \quad (4.29)$$

Lemma 4.8. *Der Normierungsfaktor ist:*

$$\frac{1}{H(\Theta, \psi)} = \begin{cases} 4 \arcsin\left(\frac{\sin \Theta}{\cos \psi}\right) & ; \quad |\psi| < \frac{\pi}{2} - \Theta \\ 2\pi & ; \quad |\psi| \geq \frac{\pi}{2} - \Theta. \end{cases} \quad (4.30)$$

Beweis. Sei $\hat{g}_2(v_{x_r}, v_{y_r}, \phi, \theta) = 1$, dann soll gelten:

$$H(\Theta, \psi) \int_{-\Theta_{lim}}^{\Theta_{lim}} \frac{2 \cos \theta}{\sqrt{\cos^2 \psi - \sin^2 \theta}} d\theta = 1.$$

Daraus folgt:

$$\begin{aligned} \frac{1}{H(\Theta, \psi)} &= 2 \int_{-\Theta_{lim}}^{\Theta_{lim}} \frac{\cos \theta}{\sqrt{\cos^2 \psi - \sin^2 \theta}} d\theta \\ &= 2 \left[\arcsin\left(\frac{\sin \theta}{\cos \psi}\right) \right]_{-\Theta_{lim}}^{\Theta_{lim}} \\ &= 4 \arcsin\left(\frac{\sin \Theta_{lim}}{\cos \psi}\right) \\ &\stackrel{\text{Lemma 4.7}}{=} \begin{cases} 4 \arcsin\left(\frac{\sin \Theta}{\cos \psi}\right) & ; \quad |\psi| < \frac{\pi}{2} - \Theta \\ 4 \arcsin\left(\frac{\overbrace{\sin\left(\frac{\pi}{2} - |\psi|\right)}^{=\cos|\psi|}}{\cos|\psi|}\right) & ; \quad |\psi| \geq \frac{\pi}{2} - \Theta \end{cases} \\ &= \begin{cases} 4 \arcsin\left(\frac{\sin \Theta}{\cos \psi}\right) & ; \quad |\psi| < \frac{\pi}{2} - \Theta \\ 4 \frac{\pi}{2} = 2\pi & ; \quad |\psi| \geq \frac{\pi}{2} - \Theta. \end{cases} \end{aligned}$$

□

Die nachfolgende Abbildung zeigt die zwei verschiedenen Regionen im Frequenzraum und die Bogenlängen, die dem Normierungsfaktor für das Sampling der 3D FT entsprechen.

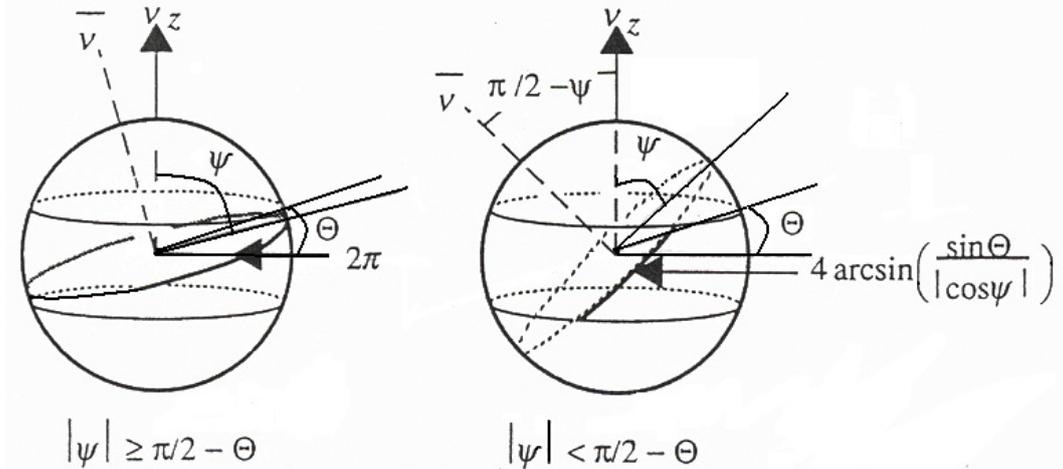


Abbildung 4.19: Quelle: [6] S. 36.

Satz und Definition 4.9. (3D FBP)

Es gelten die Annahmen aus der Anmerkung 4.6. Somit erhalten wir:

$$f(x, y, z) = (2\pi)^{-3} \int_{-\Theta}^{\Theta} \int_0^{\pi} \int_{\mathbb{R}^2} \hat{g}_2(v_{x_r}, v_{y_r}, \phi, \theta) \sqrt{v_{x_r}^2 + v_{y_r}^2} H(\Theta, \psi) e^{i(x_r v_{x_r} + y_r v_{y_r})} \cos \theta dv_{x_r} dv_{y_r} d\phi d\theta. \quad (4.31)$$

In diesem Fall heißt:

$$H_C(v_{x_r}, v_{y_r}, \theta) := \sqrt{v_{x_r}^2 + v_{y_r}^2} H(\Theta, \psi) \quad (4.32)$$

der **Colsher-Filter**.

Diese Rekonstruktionsmethode wird die **3D gefilterte Rückprojektion**, kurz: **3D FBP** genannt.

Beweis. Wir wenden die inverse FT (2.6) an und erhalten:

$$\begin{aligned}
 \underbrace{f(x, y, z)}_{:=\tilde{x}} &= (2\pi)^{-3/2} \int_{\mathbb{R}^3} \hat{f}_3(\xi) e^{i\langle \tilde{x}, \xi \rangle} d\xi \\
 &\stackrel{\text{Natterer [12] S. 190}}{=} (2\pi)^{-3/2} \frac{1}{|S^1|} \int_{S^2} \int_{\omega^\perp} |\eta| \hat{f}_3(\eta) e^{i\langle \tilde{x}, \eta \rangle} d\eta d\omega \\
 &\stackrel{\text{Proj.satz 2.31}}{=} (2\pi)^{-3/2} \frac{1}{2\pi} \int_{S^2} \int_{\omega^\perp} |\eta| (2\pi)^{-1/2} (\mathcal{F}_2 P f)(\omega, \eta) e^{i\langle \tilde{x}, \eta \rangle} d\eta d\omega \\
 &= (2\pi)^{-3} \int_{S^2} \int_{\omega^\perp} |\eta| (\mathcal{F}_2 P f)(\omega, \eta) e^{i\langle \tilde{x}, \eta \rangle} d\eta d\omega \\
 &\approx (2\pi)^{-3} \int_{S^2} \int_{\omega^\perp} |\eta| \underbrace{F(\eta) (\mathcal{F}_2 P f)(\omega, \eta)}_{=:G(\omega, \tilde{x})} e^{i\langle \tilde{x}, \eta \rangle} d\eta d\omega \\
 &= (2\pi)^{-3} \int_{S^2} G(\omega, \tilde{x}) d\omega \\
 &= (2\pi)^{-3} \int_0^{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} G(\omega, \tilde{x}) \cos \theta d\theta d\phi.
 \end{aligned}$$

Da

$$(\mathcal{F}_2 P f)(\omega, \eta) = \hat{g}_2(v_{x_r}, v_{y_r}, \phi, \theta)$$

und

$$\eta = \begin{pmatrix} -v_{x_r} \sin \phi - v_{y_r} \cos \phi \sin \theta \\ v_{x_r} \cos \phi - v_{y_r} \sin \phi \sin \theta \\ v_{y_r} \cos \theta \end{pmatrix}, \text{ also } |\eta| = \sqrt{v_{x_r}^2 + v_{y_r}^2}$$

ist, der Filter wie folgt aussieht:

$$F(\eta) = H(\Theta, \psi) W(v_{x_r}, v_{y_r}),$$

und das Skalarprodukt

$$\langle \eta, \tilde{x} \rangle = x_r v_{x_r} + y_r v_{y_r}$$

ergibt, erhalten wir:

$$\begin{aligned}
 G(\omega, \tilde{x}) &= \int_{\omega^\perp} |\eta| F(\eta) (\mathcal{F}_2 P f)(\omega, \eta) e^{i\langle \tilde{x}, \eta \rangle} d\eta \\
 &= \int_{\mathbb{R}^2} \hat{g}_2(v_{x_r}, v_{y_r}, \phi, \theta) \sqrt{v_{x_r}^2 + v_{y_r}^2} H(\Theta, \psi) W(v_{x_r}, v_{y_r}) e^{i(x_r v_{x_r} + y_r v_{y_r})} dv_{x_r} dv_{y_r}.
 \end{aligned}$$

4 Das C-C-Modell und darauf beruhende Bildrekonstruktionen

Damit folgt also mit $\phi \in [0, \pi)$ und $|\theta| \leq \Theta$, dass

$$f(x, y, z) = (2\pi)^{-3} \int_{-\Theta}^{\Theta} \int_0^{\pi} \int_{\mathbb{R}^2} \hat{g}_2(v_{x_r}, v_{y_r}, \phi, \theta) \sqrt{v_{x_r}^2 + v_{y_r}^2} H(\Theta, \psi) W(v_{x_r}, v_{y_r}) e^{i(x_r v_{x_r} + y_r v_{y_r})} \cos \theta dv_{x_r} dv_{y_r} d\phi d\theta$$

ist. □

Der Wert des Bildes im Punkt $f(x, y, z)$ kann also durch Filterung der Projektion im Frequenzraum und Multiplikation mit dem Normierungsfaktor, 2D inverser FT der gefilterten Projektion und anschließender Rückprojektion erzielt werden.

Man beachte, dass die Rückprojektion für $n = 3$ mit $\omega = (\cos \phi \cos \theta, \sin \phi \cos \theta, \sin \theta)^\top$ folgendermaßen lautet:

$$\begin{aligned} (P^*g)(\underbrace{x, y, z}_{:=\tilde{x}}) &\stackrel{(2.20)}{=} \int_{S^2} g(\omega, E_\omega \tilde{x}) \\ &= \int_0^{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} g(\omega, E_\omega \tilde{x}) \cos \theta d\theta d\phi \end{aligned} \quad (4.33)$$

Da $\Theta \in [0, \frac{\pi}{2})$ und $\phi \in [0, \pi)$ folgt:

$$(P^*g)(\tilde{x}) = \int_0^{\pi} \int_{-\Theta}^{\Theta} g(\omega, E_\omega \tilde{x}) \cos \theta d\theta d\phi.$$

Da

$$E_\omega \tilde{x} = \tilde{x} - (\tilde{x}^\top \omega) \omega = \begin{pmatrix} -x_r \sin \phi - y_r \cos \phi \sin \theta \\ x_r \cos \phi - y_r \sin \phi \sin \theta \\ y_r \cos \theta \end{pmatrix}$$

folgt:

$$\begin{aligned} (P^*g)(\tilde{x}) &= \int_0^{\pi} \int_{-\Theta}^{\Theta} \underbrace{g(\omega, E_\omega \tilde{x})}_{=\int_{\mathbb{R}} f(E_\omega \tilde{x} + z_r \omega) dz_r \stackrel{(4.4)}{=} g(x_r, y_r, \phi, \theta)} \cos \theta d\theta d\phi \\ &= \int_{-\Theta}^{\Theta} \int_0^{\pi} g(x_r, y_r, \phi, \theta) \cos \theta d\phi d\theta. \end{aligned}$$

Abgesehen von der höheren Dimension ist der 3D FBP Algorithmus ähnlich zu dem 2D FBP Algorithmus:

Algorithmus

1. Filtern der Projektionen für festes (ϕ, θ) :

a) 2D FT der Projektion:

$$\hat{g}_2(v_{x_r}, v_{y_r}, \phi, \theta) = (2\pi)^{-1} \int_{\mathbb{R}^2} g(x_r, y_r, \phi, \theta) e^{i(x_r v_{x_r} + y_r v_{y_r})} dx_r dy_r$$

b) Multiplikation mit einem 2D Filter H_C und einem 2D Window W , die beide weiter unten noch genauer beschrieben werden:

$$\hat{g}_2^F(v_{x_r}, v_{y_r}, \phi, \theta) = \hat{g}_2(v_{x_r}, v_{y_r}, \phi, \theta) H_C(v_{x_r}, v_{y_r}, \theta) W(v_{x_r}, v_{y_r})$$

c) 2D inverse FT um eine gefilterte Projektion zu erhalten:

$$g^F(x_r, y_r, \phi, \theta) = (2\pi)^{-1} \int_{\mathbb{R}^2} \hat{g}_2^F(v_{x_r}, v_{y_r}, \phi, \theta) e^{i(x_r v_{x_r} + y_r v_{y_r})} dv_{x_r} dv_{y_r}$$

2. 3D Rückprojektion der gefilterten Projektionen:

$$f(x, y, z) = f(x, y, z) + (\cos \theta \Delta \theta \Delta \phi) g^F(x_r, y_r, \phi, \theta) \quad (4.34)$$

3. Wiederholen der Schritte 1) - 2) für jedes ϕ mit $0 \leq \phi < \pi$

4. Wiederholen der Schritte 1) - 3) für jedes θ mit $-\Theta \leq \theta \leq \Theta$

5. Multiplikation von $f(x, y, z)$ mit $(2\pi)^{-2}$.

In Gleichung (4.34) erhalten wir x_r und y_r durch das Lösen des Gleichungssystems (4.1). (Da ϕ und θ fest sind ist die 3×3 -Matrix bekannt und wenn wir f in dem Punkt (x, y, z) haben wollen, dann müssen wir die entsprechenden x_r und y_r dazu ausrechnen.)

Der 2D Filter $H_C(v_{x_r}, v_{y_r}, \theta)$ beachtet die Redundanz der Daten und steht in Beziehung zum Normierungsfaktor des gewichteten Mittels aus Gleichung (4.29), $H(\Theta, \psi)$ haben wir in Gleichung (4.30) dargestellt und ψ kann mit Hilfe der Projektionsfrequenzen v_{x_r} und v_{y_r} dargestellt werden:

Da

$$\begin{aligned} v_x & \stackrel{v_{z_r}=0}{=} -v_{x_r} \sin \phi - v_{y_r} \cos \phi \sin \theta \\ v_y & = v_{x_r} \cos \phi - v_{y_r} \sin \phi \sin \theta \\ v_z & = v_{y_r} \cos \theta \end{aligned} \quad (4.35)$$

folgt:

$$\tan \psi = \frac{v_z}{\left| \begin{pmatrix} v_x \\ v_y \end{pmatrix} \right|}$$

und damit erhalten wir:

$$\begin{aligned} \tan^2 \psi &= \frac{v_z^2}{v_x^2 + v_y^2} \\ &\stackrel{(4.35)}{=} \frac{v_{y_r}^2 \cos^2 \theta}{v_{x_r}^2 + v_{y_r}^2 \sin^2 \theta}. \end{aligned}$$

Den Winkel ψ bekommen wir also durch:

$$\psi = \arctan \left(\sqrt{\frac{v_{y_r}^2 \cos^2 \theta}{v_{x_r}^2 + v_{y_r}^2 \sin^2 \theta}} \right). \quad (4.36)$$

Die Winkel- und Radialabhängigkeiten des Colsher-Filters werden deutlich in Gleichung (4.32) getrennt. Der radiale Faktor ist der Betrag der Frequenz

$$|v| = \sqrt{v_{x_r}^2 + v_{y_r}^2 + \underbrace{v_{z_r}^2}_{=0}} = \sqrt{v_{x_r}^2 + v_{y_r}^2}$$

und ähnelt dem Ramp-Filter $|v_{x_r}|$ in der 2D FBP.

Um die Rauschverstärkung des Colsher-Filters zu drosseln, wird $W(v_{x_r}, v_{y_r})$ konstruiert. Wie in der 2D Bildrekonstruktion hängt die Wahl von W vom Grad des Rauschens (Rauschlevel) ab und wird gewöhnlich empirisch bestimmt. Üblicherweise benutzt man das "Hamming-Window" mit einer geeigneten Abschneidefrequenz, entweder in der Form eines Kartesischen Produktes $W_H(v_{x_r})W_H(v_{y_r})$ oder als ein radial symmetrisches "Window" $W_H(\sqrt{v_{x_r}^2 + v_{y_r}^2})$. Die erstgenannte Wahl erlaubt die Auswahl derselben "Apodisation" wie in der 2D Rekonstruktion, während die letztere Wahl mit der 3D Natur des Problems konsistent ist.

4.5.2 Der 3D Rückprojektions-Algorithmus (3DRP)

Als erstes wollen wir uns kurz mit der Frage beschäftigen: "Was ist Abgeschnittenheit und wie beeinflusst sie die Rekonstruktion?". Danach stellen wir eine 3D gefilterte Rückprojektionsmethode vor, den 3DRP Algorithmus, der durch Diskretisierung einer analytischen Rekonstruktionsformel erzielt wird und bei dem die vom Scanner nicht gemessenen Daten geschätzt werden.

Im vorangegangenen Abschnitt sahen wir, wie das Bild $f(x, y, z)$ aus den 2D parallelen Projektionen $g(x_r, y_r, \phi, \theta)$ rekonstruiert wurde, die für eine Menge von Richtungen $\Omega_\Theta = \{(\theta, \phi); 0 \leq \phi < \pi, |\theta| \leq \Theta\}$ gemessen worden waren. Der 3D FBP Algorithmus wurde unter der Annahme hergeleitet, dass komplette Projektionen vorlagen, was bedeutet, dass für jede Richtung $\hat{z}_r(\phi, \theta) \in \Omega$, $g(x_r, y_r, \phi, \theta)$ für alle $(x_r, y_r) \in D_g$ gemessen

worden war, wobei D_g der Träger der Projektion des Objektes ist. Z. B. ist der Träger der orthogonalen Projektion eines Zylinders entlang einer Richtung $\hat{z}_r(\phi, \theta)$:

$$D_g = \{(x_r, y_r); |x_r| < R, |y_r| < \frac{L_D \cos \theta}{2} + |\sin \theta| \sqrt{R_D^2 - x_r^2}\} \quad (4.37)$$

wenn $f(x, y, z) \neq 0$ in einem Zylinder mit derselben Länge L_D wie der Scanner und einem Radius $R < R_D$ ist.

Man beachte, dass D_g unabhängig von ϕ ist, was auf die Rotationssymmetrie des Scanners zurückzuführen ist.

Ein zylindrischer Scanner misst die Projektion $g(x_r, y_r, \phi, \theta)$ jedoch nur in der Region M_g :

$$M_g = \{(x_r, y_r); |x_r| < R, |y_r| < \frac{L_D \cos \theta}{2} - |\sin \theta| \sqrt{R_D^2 - x_r^2}\}. \quad (4.38)$$

Ersichtlich wird die Abschätzung für M_g aus Folgendem:

$$|y_r| \stackrel{(4.13)}{=} |z| |\cos \theta| \stackrel{\theta \in [0, \frac{\pi}{2}[}{=} |z| \cos \theta.$$

Da z für $|z| \stackrel{(4.42)}{\leq} \frac{L_D - |\Delta r|}{2}$ gemessen wird folgt:

$$\begin{aligned} |y_r| &= |z| \cos \theta \leq \frac{L_D}{2} \cos \theta - \frac{|\Delta r|}{2} \cos \theta \\ &\stackrel{(4.13)}{=} \frac{L_D}{2} \cos \theta - \frac{|\tan \theta| 2\sqrt{R_D^2 - x_r^2}}{2} \cos \theta \\ &= \frac{L_D}{2} \cos \theta - |\sin \theta| \sqrt{R_D^2 - x_r^2}. \end{aligned}$$

Für D_g können wir analog rechnen, nur sind die Daten komplett, d. h. $|z| \leq \frac{L_D + |\Delta r|}{2}$.

Man beachte bei M_g , dass im Zentrum des FOV ($x_r = 0$) y_r mit wachsendem θ sinkt und verschwindet, wenn $|\theta| = \Theta_D$ ist.

$$\left. \begin{array}{l} \theta \uparrow \Rightarrow \cos \theta \downarrow \Rightarrow \frac{L_D \cos \theta}{2} \downarrow \\ \theta \uparrow \Rightarrow |\sin \theta| \uparrow \Rightarrow |\sin \theta| \sqrt{R_D^2} \uparrow \end{array} \right\} \Rightarrow \frac{L_D \cos \theta}{2} - |\sin \theta| R_D \downarrow$$

Wie man in Abbildung 4.13 und für $\theta \neq 0$ sieht, gilt:

$$M_g \subset D_g.$$

Die gemessene Region M_g enthält nicht die Projektion des Objektes D_g und daher ist die Projektion abgeschnitten.

Wendet man den 3D FBP Algorithmus auf die abgeschnittenen Projektionen an, so weist das resultierende Bild $f(x, y, z)$ heftige Artefakte auf, welche hauptsächlich die Abnahme der Scannersensitivität reflektieren, wenn man sich vom Zentrum ($z = 0$) in Richtung Ecken des FOV ($z = \pm \frac{L_D}{2}$) bewegt. Leider kann man diese Artefakte nicht dadurch korrigieren, dass man einfach in jedem Punkt die FBP Rekonstruktion durch die Sensitivität in diesem Punkt dividiert.

Der 3D-Rückprojektions-Algorithmus (3D reprojection algorithm, 3DRP) ist nun die Standardmethode, die zur Rekonstruktion abgeschnittener 3D PET-Daten benutzt wird. Bei dieser Methode erzielt man komplette 2D Projektionen ($g_{\phi, \theta}(x_r, y_r)$ ist 2D, da nur x_r und y_r Variablen sind), indem man $g(x_r, y_r, \phi, \theta)$ in der nicht-gemessenen Region $\{(x_r, y_r); (x_r, y_r) \in D_g, (x_r, y_r) \notin M_g\}$ schätzt. Die gefilterte Rückprojektion kann dann auf die kompletten Projektionen angewendet werden, die zum Teil gemessen und zum anderen Teil künstlich synthetisiert (hergestellt) sind.

Die abgeschnittenen Daten werden durch die Berechnung des Linienintegrals

$$g(x_r, y_r, \phi, \theta) = \int_{\mathbb{R}} f_{2D}(x, y, z) dz_r \quad (4.39)$$

über $f_{2D}(x, y, z)$ geschätzt, wobei $f_{2D}(x, y, z)$ aus den 2D Projektionen mit $\theta = 0$ rekonstruiert wird (z. B. aus der Datenteilmenge, die dem 2D Erwerb entspricht). Diese Projektionen sind nicht abgeschnitten, denn für $\theta = 0$ gilt:

$$D_g = M_g$$

und wir können dann $f_{2D}(x, y, z)$ z. B. unter Verwendung der 2D FBP rekonstruieren.

In der Praxis werden die 2D Projektionen für diskrete Werte des Winkels θ gemessen, die den Integerwerten der Ringdifferenz entsprechen (siehe Abbildung 4.9). Bei der Implementierung des 3DRP Algorithmus benötigen wir die Angabe der Datenteilmenge, die zur Rekonstruktion von $f_{2D}(x, y, z)$ benutzt wird. Die minimale Teilmenge besteht aus den Koinzidenzen mit den Ringdifferenzen $\Delta r = 0$ und $\Delta r = \pm 1$, welche ausreichend sind um das gesamte FOV zu rekonstruieren. Um die statistische Genauigkeit des geschätzten Bildes zu verbessern, nimmt man gewöhnlich auch Koinzidenzen mit größeren Werten von Δr , z. B. $|\Delta r| \leq 5$ hinzu (dabei folgen wir den standardmäßigen 2D Rekonstruktionsverfahren, die angewendet werden, wenn der Scanner mit ausgefahrenen Septen operiert). Die Wahl des maximalen Wertes von Δr führt zu einem Kompromiss zwischen systematischen und statistischen Fehlern für $f_{2D}(x, y, z)$.

Empirische Erfahrungswerte haben gezeigt, dass die Qualität des endgültigen Bildes $f(x, y, z)$ nicht so stark von dem maximalen Wert von Δr abhängt, der zur Schätzung von $f_{2D}(x, y, z)$ benutzt wird.

Algorithmus

Der 3DRP-Algorithmus besteht aus folgenden Schritten:

1. Auswahl/Entnahme der 2D Daten aus der gesamten Datenmenge (um die 2D Rekonstruktion durchführen zu können)

$$\theta = 0 \stackrel{(4.13)}{\Rightarrow} z = y_r \Rightarrow g(x_r, y_r = z, \phi, \theta = 0) = g_{\theta=0}^{y_r=z}(x_r, \phi)$$

2. Rekonstruktion einer vorläufigen Bildschätzung $f_{2D}(x, y, z)$ durch Anwendung der 2D FBP für jede Ebene z :

$$\begin{aligned} f_{2D}(x, y, z) &= f_z(x, y) \\ &\stackrel{(4.15)}{=} (2\pi)^{-3/2} \int_0^\pi \int_{\mathbb{R}} \hat{g}_1(v_{x_r}, \phi) |v_{x_r}| e^{ix_r v_{x_r}} dv_{x_r} d\phi \\ &= (2\pi)^{-2} \int_0^\pi \int_{\mathbb{R}^2} g_{\theta=0}^{y_r=z}(x_r, \phi) e^{-ix_r v_{x_r}} dx_r |v_{x_r}| e^{ix_r v_{x_r}} dv_{x_r} d\phi \end{aligned}$$

3. Durchführung der nachfolgenden Schritte für jede Projektion (ϕ, θ) mit $0 \leq \phi < \pi$ und $|\theta| \leq \Theta$:

- a) Schätzung der abgeschnittenen Daten mit Hilfe der Gleichung (4.2):

$$g(x_r, y_r, \phi, \theta) = \int_{\mathbb{R}} f_{2D}(x, y, z) dz_r \text{ für } \{(x_r, y_r); (x_r, y_r) \in D_g, (x_r, y_r) \notin M_g\}$$

- b) Zusammenfassen der geschätzten und gemessenen Daten (\Rightarrow komplette Daten)

Die nachfolgenden Schritte entsprechen der 3D FBP Methode:

- c) Berechnung der 2D FFT um $\hat{g}_2(v_{x_r}, v_{y_r}, \phi, \theta)$ zu erhalten
- d) Multiplikation von $\hat{g}_2(v_{x_r}, v_{y_r}, \phi, \theta)$ mit dem apodisierten Colsher-Filter:

$$H_c(v_{x_r}, v_{y_r}, \theta)W(v_{x_r}, v_{y_r})$$

- e) Berechnung der 2D inversen FFT um $g^F(x_r, y_r, \phi, \theta)$ zu erhalten
- f) Rückprojektion durch das 3D Volumenbild entsprechend Gleichung (4.34)
- g) Multiplikation des erhaltenen f 's mit $(2\pi)^{-2}$.

Den Schritt 3. a) bezeichnet man auch als **Vorwärtsprojektion**.

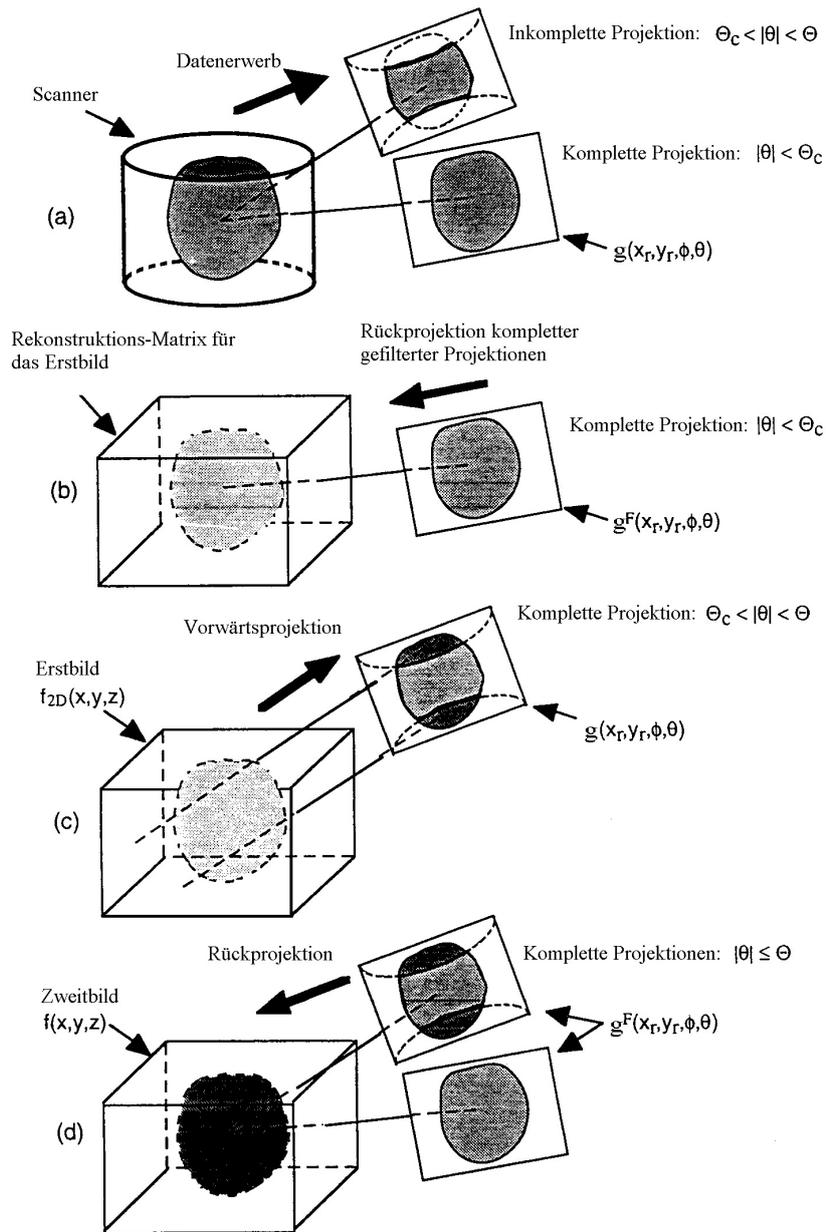


Abbildung 4.20: Der 3DRP-Algorithmus. a) Sammlung der abgeschnittenen Projektionen durch den Scan. b) Rekonstruktion des vorläufigen Bildes $f_{2D}(x, y, z)$ unter Verwendung der kompletten Projektionsdaten $|\theta| \leq \Theta_c$. c) Schätzung der abgeschnittenen Projektionsdaten durch Vorwärtsprojektion durch das vorläufige Bild. d) Rekonstruktion des endgültigen Bildes $f(x, y, z)$ unter Verwendung der kompletten Projektionsmenge. Beachte, dass beim Schritt b) in der Praxis $\Theta_c = 0$ ist! Quelle: [6] S. 41.

Der maximale Wert von θ , der in der Rekonstruktion berücksichtigt wird bzw. auftritt, Θ , ist gewöhnlich kleiner als die Scanneröffnung $\Theta_D = \arctan\left(\frac{L_D}{2R_D}\right)$.

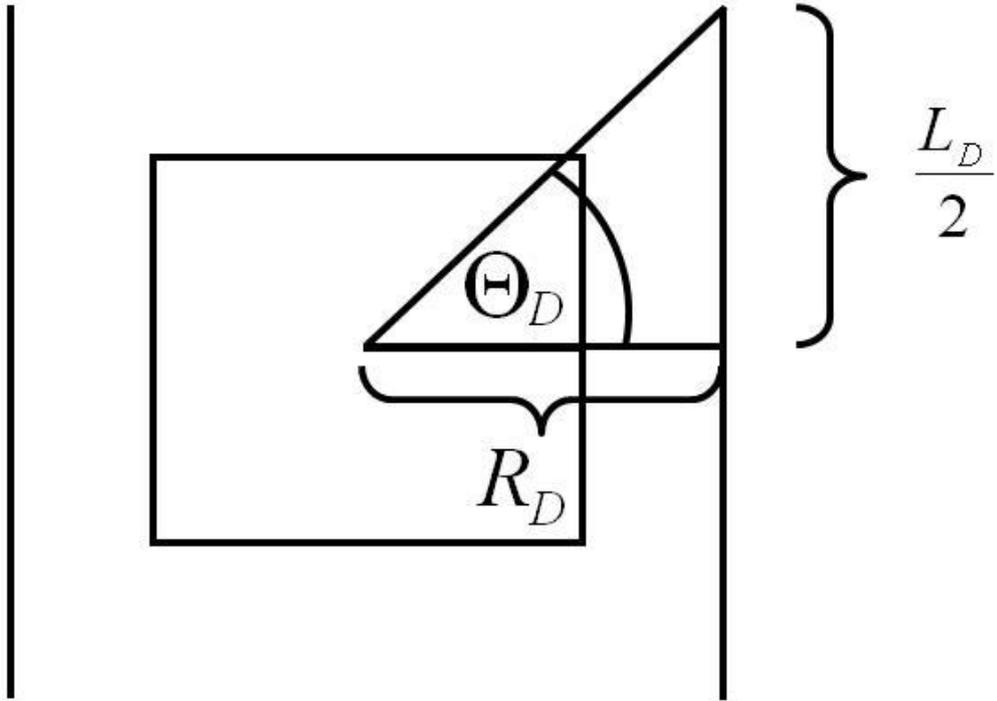


Abbildung 4.21: $\Theta_D = \arctan\left(\frac{L_D}{2R_D}\right)$.

Der Grund dafür ist folgender:

Wenn $|\theta| \approx \Theta_D$, dann ist die gemessene Region M_g wesentlich kleiner als die benötigte Region D_g (siehe Gleichung (4.37) und (4.38) und Abbildung 4.13). Für große $|\theta|$ bestehen also die kompletten Projektionsdaten hauptsächlich aus geschätzten Daten und nicht aus gemessenen Daten. Reduzieren wir den Wert von Θ , sparen wir erhebliche Rechenzeit ohne auf viel Sensitivität verzichten zu müssen. Eine maximale Ringdifferenz von 17 ist typisch für einen 24-Ring-Scanner, was einem Winkel Θ von 12° entspricht. In diesem Fall sind volle 93% der Daten in die Rekonstruktion miteinbezogen.

4.6 Rebinningmethoden

Im Folgenden richten wir uns - bis nichts anderes vermerkt wird - nach dem Paper von Michel Defrise, [14].

Rebinningmethoden sind, wie wir gleich sehen werden, Methoden, die wir zum Teil zur 3D PET und zum anderen Teil auch zur 2D PET zählen können. Aus diesem Grund widmen wir ihnen einen eigenen Abschnitt.

Die beachtlichen Erfahrungswerte, die wir seit 1990 mit dem 3DRP-Algorithmus gemacht haben, da er in den meisten PET-Zentren, die 3D Daten erwerben, zum Einsatz kam, haben die Genauigkeit und Robustheit dieses Algorithmus nur bestätigt. Es wurden jedoch trotzdem weitere, alternative Algorithmen entwickelt, entweder um die Rekonstruktionszeit zu verringern oder um die Bildqualität zu erhöhen. Die Zeit, die zur Rekonstruktion 3D Daten mit dem 3DRP-Algorithmus benötigt wird ist um ein Vielfaches länger als die Zeit, die man zur Rekonstruktion 2D Daten, die mit Septen erworben wurden, benötigt. Dies ist auf den beachtlichen Zuwachs der Anzahl der LOR's im 3D Datenerwerb, die rückprojiziert werden müssen, zurückzuführen. Des Weiteren sollte man die Menge an Speicherplatz, die beim 3D Datenerwerb benötigt wird, nicht außer Acht lassen. Ein weiterer Grund für die wachsende Komplexität ist der, dass die Rekonstruktion des 3D Bildes beim 3DRP nicht in die Rekonstruktion einer Menge unabhängiger Ebenen zerlegt werden kann. Signifikante Verbesserungen bei der Rekonstruktionszeit wurden nun dadurch erreicht, dass verschiedene Kombinationen der drei folgenden Ansätze benutzt wurden:

1. Die Einführung schnellerer, aber auch teurerer Hardware und die Optimierung der Implementierung des 3DRP-Algorithmus.
2. Eine Reduzierung des Sampling der 3D Daten um die Anzahl der LOR's, die rückprojiziert werden müssen, zu verringern.
Reduziertes Sampling wird durch das Zusammenfassen von Gruppen angrenzender LOR's erzielt und zwar so, dass der resultierende Verlust räumlicher Auflösung für einen gegebenen Studientyp "akzeptabel" bleibt. (siehe Sinogramm-Mashing in Abschnitt 4.3).
3. Die Einführung approximativer bzw. exakter Algorithmen, die auf dem Rebinning der 3D Daten in eine 2D Datenmenge basieren.

Die Rebinningoperation wird aus dem C-C Modell hergeleitet.

Unter einem **Rebinning-Algorithmus** versteht man einen Algorithmus, der 3D Daten in einen Stapel 2D Datenmengen sortiert (engl.: rebins), wobei für jede transaxiale Schicht die 2D Daten als Sinogrammdaten angeordnet sind.

Diese rebinnten Daten sind zu den Daten, die im konventionellen 2D Modus erworben wurden geometrisch äquivalent und können deshalb durch die separate Anwendung

eines beliebigen 2D Rekonstruktionsalgorithmus auf jede Ebene rekonstruiert werden. Rebinning zerlegt also das 3D Rekonstruktionsproblem in eine Menge unabhängiger 2D Radontransformationen.

Vorausgesetzt das Rebinningverfahren selbst ist effizient, so wird die Rekonstruktion "fast so schnell" wie im 2D Modus, wohingegen jedoch die durch den 3D Erwerb gewachsene Sensitivität beibehalten wird. Dies liegt daran, dass die komplette Menge der 3D LOR's für die Rekonstruktion benutzt wird.

Im Folgenden wollen wir nun drei der aktuellsten Rebinningmethoden, die aus den vergangenen 6-7 Jahren stammen, vorstellen:

- **FOREX:**

- zerlegt das 3D Rekonstruktionsproblem in eine Menge 2D Probleme
- exakter Algorithmus (d. h. beruht auf der Diskretisierung einer exakten Inversionsformel)
- die nicht gemessenen Teile der Projektionen (abgeschnittene Daten) müssen geschätzt werden (es wird also derselbe zeitaufwendige Vorwärtsprojektionsschritt wie bei der 3DRP benötigt)

- **FORE:**

- ähnliche Zerlegung wie bei FOREX
- approximativer Algorithmus
- operiert direkt (ohne einen Schätzungsschritt) auf den axial abgeschnittenen Daten
- nur geeignet, wenn der maximale Winkel zwischen LOR's und transaxialen Ebenen (**Winkelöffnung**) 25° nicht übersteigt (diese Schwelle hängt natürlich vom Rauschlevel und der Natur des Bildes ab!)

- **FORE-J:**

- ähnliche Zerlegung wie bei FOREX und FORE
- exakter Algorithmus
- operiert direkt auf den axial abgeschnittenen Daten (es werden also die Vorteile von FOREX und FORE kombiniert)

Im Anschluss an die exakte Herleitung der drei Verfahren werden wir ihren Zusammenhang untereinander klären und in einem abschließenden Schritt alle Verfahren dieses Kapitels miteinander vergleichen.

4.6.1 FOREX

Da wir die ganze Zeit über Daten im Projektionsformat betrachtet haben, bei Rebinningverfahren jedoch die Daten in Sinogrammen vorliegen, werden wir zu Beginn dieses Abschnitts noch einmal kurz auf die Geometrie der Sinogrammdaten eingehen (zur Erinnerung: die Beziehung zwischen Projektionen und schrägen Sinogrammen wurde in Abschnitt 4.2 beschrieben), bevor wir die Frage behandeln: "Was ist ein Rebinning-Algorithmus?", eine exakte Rebinningformel herleiten und die entsprechende Implementierung beschreiben.

Sinogrammdaten für einen zylindrischen PET-Scanner

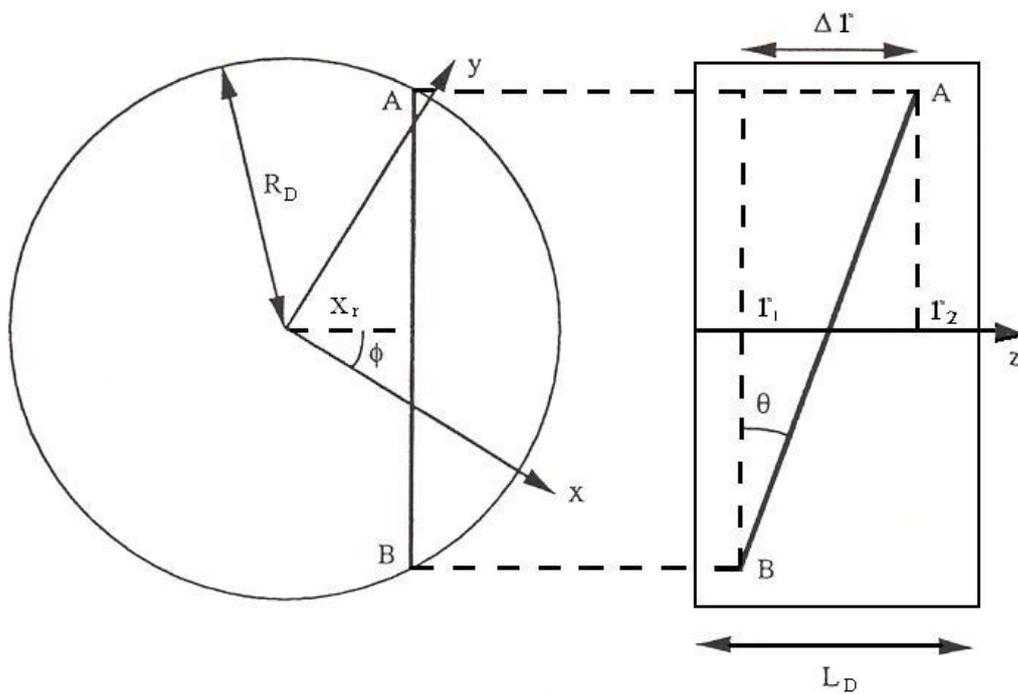


Abbildung 4.22: Geometrien eines zylindrischen PET-Scanners. Transaxiale Sicht (links) mit den Standard-Sinogramm-Variablen x_r und ϕ , die die gerade Linie \overline{AB} parametrisieren. Längsschnitt (rechts) mit den axialen Variablen z und Δr . (Man beachte, dass die x - und y -Achse für illustrative Zwecke gedreht/rotiert wurden). Quelle: [14] S. 146.

- Wir betrachten einen zylindrischen Scanner mit Radius R_D und Länge L_D .
- Die z -Achse definiert die axiale Richtung.

- Jede zur z -Achse orthogonale Ebene wird **transaxiale Ebene** genannt.
- Wir betrachten ein zylindrisches FOV Ω mit Radius $R < R_D$ und derselben Länge L_D wie der Scanner.
- **gesucht:** die Verteilungsfunktion des Kontrastmittels $f(x, y, z) \in \mathcal{S}(\mathbb{R}^3)$ mit $\text{supp } f \subset \Omega$.
- Das Linienintegral (3D Röntgentransformation) zwischen den beiden Detektoren A und B wird folgendermaßen parametrisiert (die zylindrische Geometrie repräsentiert):

$$s(x_r, \phi, z, \Delta r) = \int_{\mathbb{R}} f(x_r \cos \phi + tu_x, x_r \sin \phi + tu_y, z + tu_z) dt, \quad (4.40)$$

wobei

$$u = (u_x, u_y, u_z)^\top = \frac{1}{\sqrt{1 + \frac{(\Delta r)^2}{4(R_D^2 - x_r^2)}}} \left(-\sin \phi, \cos \phi, \frac{\Delta r}{2\sqrt{R_D^2 - x_r^2}} \right)^\top$$

ein Einheitsvektor entlang der Integrationslinie ist, $z = \frac{r_1 + r_2}{2}$, $\Delta r = (r_1 - r_2)$, x_r die Distanz zwischen z -Achse und der Projektion der Linie in eine transaxiale Ebene ist und ϕ der Winkel zwischen dieser Projektion und der y -Achse ist.

Wir erkennen also, dass z die axiale Koordinate des Mittelpunktes zwischen den beiden Detektoren und Δr der axiale Abstand zwischen den zwei Detektoren ist.

- Wir definieren eine Datenmenge, die einem festen Paar $(z, \Delta r)$ entspricht, als **schräges Sinogramm**.
- Wenn $\Delta r = 0$ ist wird das Sinogramm **gerades Sinogramm** genannt.
- Bei einem PET-Scanner mit N Ringen entspricht jedes Ringpaar einem festen Paar $(z, \Delta r)$ und daher setzen sich die Daten, die im 3D Modus erworben wurden aus insgesamt N^2 Sinogrammen zusammen. (N gerade Sinogramme und $N(N - 1)$ schräge Sinogramme).
- Für die vier Variablen gilt:

$$\begin{aligned} -R &< x_r < R \\ 0 &\leq \phi < \pi \\ |\Delta r| &\leq L_D \\ |z| &\leq \frac{L_D - |\Delta r|}{2} \end{aligned}$$

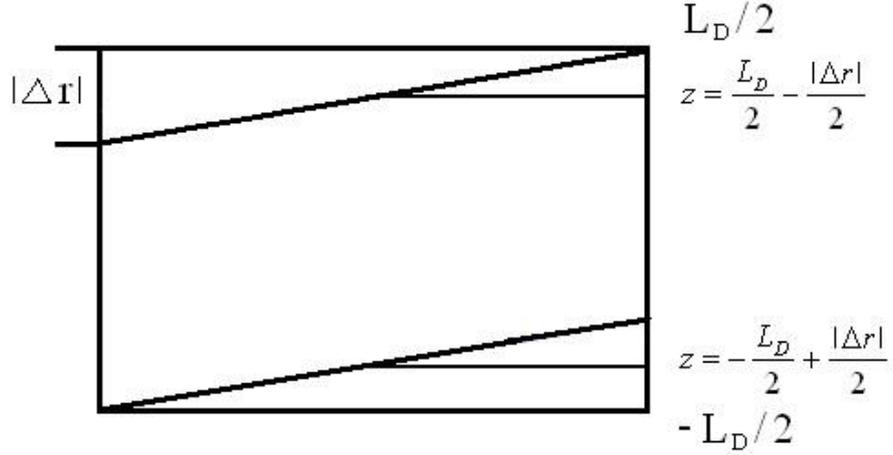


Abbildung 4.23: Verdeutlichung der Variablen $|z|$ und $|\Delta r|$.

- Wegen der endlichen Länge des zylindrischen Scanners hängt z von Δr ab und bedeckt nicht das gesamte FOV (ausgenommen für $\Delta r = 0$). Abgeschnittenheit der 3D Daten: Verkleinerung von z mit wachsendem Δr .

Die Rebinningformel, die nun hergeleitet werden soll, basiert auf folgender Parametrisierung:

$$\begin{aligned}
 & s(x_r, \phi, z, \Delta r = 2\delta\sqrt{R_D^2 - x_r^2}) \\
 = & \int_{\mathbb{R}} f\left(x_r \cos \phi - t \frac{1}{\sqrt{1+\delta^2}} \sin \phi, x_r \sin \phi + t \frac{1}{\sqrt{1+\delta^2}} \cos \phi, z + t \frac{1}{\sqrt{1+\delta^2}} \delta\right) dt.
 \end{aligned}$$

Wir substituieren $\tilde{t} = \frac{t}{\sqrt{1+\delta^2}}$, dividieren beide Seiten durch $\sqrt{1+\delta^2}$ und erhalten:

$$\begin{aligned}
 p(x_r, \phi, z, \delta) & := \frac{s(x_r, \phi, z, \Delta r = 2\delta\sqrt{R_D^2 - x_r^2})}{\sqrt{1+\delta^2}} \\
 & = \int_{\mathbb{R}} f(x_r \cos \phi - \tilde{t} \sin \phi, x_r \sin \phi + \tilde{t} \cos \phi, z + \tilde{t} \delta) d\tilde{t}, \quad (4.41)
 \end{aligned}$$

wobei $\delta := \tan \theta = \frac{\Delta r}{2\sqrt{R_D^2 - x_r^2}}$ und θ der Winkel zwischen dem LOR und einer transaxialen Ebene ist.

Da in einem Multiringscanner die Variable δ proportional zur Differenz zwischen den Indexen der zwei Koinzidenzringe ist, nennt man auch δ - mit einem kleinen Missbrauch

der Terminologie - die **Ringdifferenz**.

Für z und δ gilt nun:

$$|z| \leq \frac{L_D}{2} - \frac{|\Delta r|}{2} \stackrel{\Delta r=2\delta\sqrt{R_D^2-x_r^2}}{=} \frac{L_D}{2} - |\delta| \sqrt{R_D^2 - x_r^2}, \quad (4.42)$$

$$|\delta| \leq \frac{L_D}{2\sqrt{R_D^2 - x_r^2}}. \quad (4.43)$$

Aus Vereinfachungsgründen benutzen wir den maximalen Wert, der unabhängig von x_r ist ($x_r = 0$) und erhalten:

$$|z| \leq \frac{L_D}{2} - |\delta| R_D, \quad (4.44)$$

$$|\delta| \leq \frac{L_D}{2R_D}. \quad (4.45)$$

Was ist ein Rebinning-Algorithmus?

Wenn ein Scanner im 2D Modus operiert, liegen die gemessenen LOR's in einer transaxialen Ebene, und zwar so, dass $r_1 = r_2$ ist. Deshalb wird eine 2D Datenmenge auch durch (4.41) beschrieben, aber mit $\delta = 0$:

$$p_{reb}(x_r, \phi, z) := p(x_r, \phi, z, 0). \quad (4.46)$$

Diese 2D Daten können wir in geraden Sinogrammen anordnen (für jede Schicht z ein Sinogramm), welche durch 2D FBP und unabhängig voneinander rekonstruiert werden können. Es ist nicht weiter verwunderlich, dass die Ebene-um-Ebene Rekonstruktion einer 3-Parameter-Datenmenge schneller als die Rekonstruktion einer 4-parametrischen 3D Datenmenge mit dem 3DRP-Algorithmus ist.

Diese Betrachtung führt zu einem alternativen Ansatz der 3D Rekonstruktion, bei dem die 3D Daten nicht direkt konstruiert werden, sondern dazu dienen, die 2D Daten zu schätzen, aus denen dann das Bild durch Benutzung irgendeines 2D Rekonstruktionsalgorithmus (z. B. 2D FBP) konstruiert werden kann. Deshalb definieren wir einen Rebinningalgorithmus auch als eine Methode, um $p_{reb}(x_r, \phi, z)$ aus $p(x_r, \phi, z, \delta)$ zu schätzen. Ein nützlicher Rebinning-Algorithmus sollte wie folgt sein:

1. schnell (ansonsten würde man keinen Vorteil in Hinsicht auf andere Techniken haben),
2. genau und im Idealfall auf einer exakten analytischen Inversionsformel beruhend,
3. stabil in Hinsicht auf das Rauschen (wie bei der 3DRP Methode).

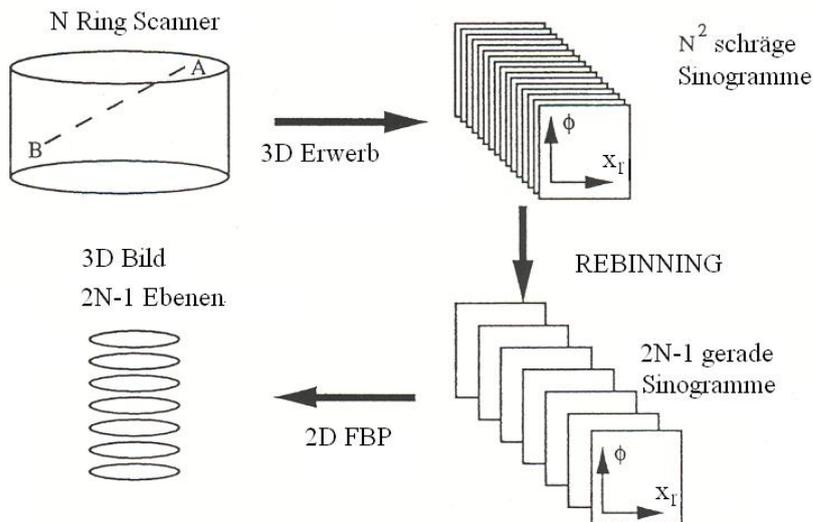


Abbildung 4.24: Das Prinzip eines Rebinning-Algorithmus, dargestellt anhand des typischen Ebenen-Samplings eines Multi-Ring-Scanners: N gerade Ebenen, die in den Ebenen der N Detektorringe liegen und $N - 1$ Ebenen, die zwischen benachbarten Detektorringen liegen. Quelle: [14] S. 147.

Man beachte die Wichtigkeit der letzten Forderung:

Rauschfreie 3D PET Daten sind in dem Sinne redundant, dass die geraden Sinogramme ($\delta = 0$) für eine exakte Rekonstruktion ausreichend sind. Folglich könnte, ohne Forderung 3., Rebinning trivialerweise durch das Herausnehmen der direkten Sinogramme aus den gesamten 3D Daten erreicht werden (gemäß (4.46)). Da wir jedoch nicht von rauschfreien Daten ausgehen können, würde dieser Ansatz (Herausnehmen der direkten Sinogramme) den Vorteil der wachsenden Sensitivität, der durch den 3D Erwerb erreicht wurde, wieder zunichte machen. Eine optimale SNR im rekonstruierten Bild kann nur erreicht werden, wenn die Rebinningmethode die gesamten 3D Daten miteinbezieht, so wie es der 3DRP-Algorithmus macht.

Zwei Herleitungen einer exakten Rebinningformel

Im Folgenden zeigen wir nun zwei Möglichkeiten auf, die Gleichung, die die Basis des FOREX-Algorithmus ist, herzuleiten. Bei der 1. Herleitung nehmen wir die **Besselfunktion** und ihre Eigenschaften zur Hilfe, bei der 2. Herleitung den 3D Projektionssatz. Diese Basisgleichung des FOREX-Algorithmus ist dann nicht nur Bestandteil der Inversionsformel, sondern kann auch zum Schätzen der abgeschnittenen Daten benutzt werden.

Herleitung 1:

Das inverse Problem, welches durch Integral (4.41) definiert ist, ist rotationsinvariant um die z -Achse und translationsinvariant entlang der radialen Variablen x_r . Diese Invarianzen können benutzt werden, um für jedes feste Paar (z, δ) die kontinuierliche FT der schrägen Sinogramme für die Variablen x_r und ϕ zu berechnen.

Lemma 4.10. *Mit den Bezeichnungen wie zuvor gilt:*

$$p(x_r, \phi + \pi, z, \delta) = p(-x_r, \phi, z, -\delta). \quad (4.47)$$

Beweis. $p(x_r, \phi + \pi, z, \delta)$

$$\begin{aligned} &= \int_{\mathbb{R}} f(x_r \cos(\phi + \pi) - t \sin(\phi + \pi), x_r \sin(\phi + \pi) + t \cos(\phi + \pi), z + t\delta) dt \\ &= \int_{\mathbb{R}} f(-x_r \cos \phi + t \sin \phi, -x_r \sin \phi - t \cos \phi, z + t\delta) dt. \end{aligned}$$

Substituieren wir nun $\tilde{t} = -t$, so erhalten wir:

$$\begin{aligned} &= \int_{\mathbb{R}} f(-x_r \cos \phi - \tilde{t} \sin \phi, -x_r \sin \phi + \tilde{t} \cos \phi, z + \tilde{t}(-\delta)) d\tilde{t} \\ &= p(-x_r, \phi, z, -\delta). \end{aligned}$$

□

Anmerkung 4.11. Sei $T^4 := \{(x_r, \phi, z, \delta) ; |x_r| < R, 0 \leq \phi < \pi, |z| \leq \frac{L_D}{2} - |\delta| R_D, |\delta| \leq \frac{L_D}{2R_D}\}$. In Zukunft - wenn es nicht anders erwähnt wird - sei p für alle $(x_r, \phi, z, \delta) \in T^4$ bekannt. Wegen Lemma 4.10 ist p somit auch bekannt für alle $|x_r| < R, 0 \leq \phi < 2\pi, |z| \leq \frac{L_D}{2} - |\delta| R_D$ und $|\delta| \leq \frac{L_D}{2R_D}$. Das bedeutet, dass wenn wir p in Zukunft bzgl. ϕ integrieren, dass wir dies von 0 bis 2π tun können.

Lemma 4.12. *Sei p bekannt für alle $(x_r, \phi, z, \delta) \in T^4$.*

Dann lautet die 2D kontinuierliche FT in Bezug auf die Variablen x_r und ϕ :

$$\begin{aligned} \text{a) } \hat{p}_2(\omega, k, z, \delta) &= (2\pi)^{-1} \int_0^\pi e^{-ik\phi} \int_{-R}^R e^{-i\omega x_r} p(x_r, \phi, z, \delta) \\ &\quad + (-1)^k e^{i\omega x_r} p(x_r, \phi, z, -\delta) dx_r d\phi, \end{aligned}$$

$$\text{b) } \hat{p}_2(\omega, k, z, \delta) = (2\pi)^{-1} \int_0^R \int_0^{2\pi} \int_0^{2\pi} \sigma e^{-ik\beta} e^{-i(k\phi + \omega\sigma \cos \phi)} .$$

$$f(\sigma \cos \beta, \sigma \sin \beta, z - \sigma\delta \sin \phi) d\phi d\beta d\sigma.$$

Beweis. Aus der Voraussetzung folgt:

$$\hat{p}_2(\omega, k, z, \delta) = (2\pi)^{-1} \int_0^{2\pi} \int_{-R}^R e^{-i(k\phi + \omega x_r)} p(x_r, \phi, z, \delta) dx_r d\phi,$$

wobei $\omega \in \mathbb{R}$ die radiale Frequenz und $k \in \mathbb{Z}$ der Fourierindex ist.

a) Da die schrägen Sinogramme nur für $\phi \in [0, \pi[$ gesampelt sind, wird die obige Gleichung in der Praxis unter Verwendung der Symmetrie aus dem vorangegangenen Lemma berechnet:

$$\begin{aligned} \hat{p}_2(\omega, k, z, \delta) &= (2\pi)^{-1} \int_0^{\pi} \int_{-R}^R e^{-i(k\phi + \omega x_r)} p(x_r, \phi, z, \delta) dx_r d\phi \\ &\quad + (2\pi)^{-1} \int_{\pi}^{2\pi} \int_{-R}^R e^{-i(k\phi + \omega x_r)} p(x_r, \phi, z, \delta) dx_r d\phi. \end{aligned}$$

Wir substituieren im 2. Summanden $\phi = \tilde{\phi} + \pi$ und $x_r = -\tilde{x}_r$ und erhalten:

$$\begin{aligned} \hat{p}_2(\omega, k, z, \delta) &= (2\pi)^{-1} \int_0^{\pi} \int_{-R}^R e^{-i(k\phi + \omega x_r)} p(x_r, \phi, z, \delta) dx_r d\phi \\ &\quad + (2\pi)^{-1} \int_0^{\pi} \int_R^{-R} e^{-ik\phi} \underbrace{e^{ik\pi}}_{=(-1)^k} e^{i\omega x_r} \underbrace{p(-x_r, \phi + \pi, z, \delta)}_{\substack{\text{Lemma 4.10} \\ p(x_r, \phi, z, -\delta)}} (-1) dx_r d\phi \\ &= (2\pi)^{-1} \int_0^{\pi} e^{-ik\phi} \int_{-R}^R e^{-i\omega x_r} p(x_r, \phi, z, \delta) + (-1)^k e^{i\omega x_r} p(x_r, \phi, z, -\delta) dx_r d\phi. \end{aligned}$$

b)

$$\begin{aligned} \hat{p}_2(\omega, k, z, \delta) &\stackrel{\text{Anmerkung 4.11}}{=} (2\pi)^{-1} \int_0^{2\pi} \int_{-R}^R e^{-i(k\phi + x_r \omega)} p(x_r, \phi, z, \delta) dx_r d\phi \\ &= (2\pi)^{-1} \int_0^{2\pi} \int_{-R}^R \int_{\mathbb{R}} e^{-i(k\phi + x_r \omega)} f(x_r \cos \phi - t \sin \phi, x_r \sin \phi + t \cos \phi, \\ &\quad z + t\delta) dt dx_r d\phi. \end{aligned}$$

Wir setzen $x_r = x \cos \phi + y \sin \phi$ und $t = -x \sin \phi + y \cos \phi$ und erhalten für die Funktionalmatrix $D\phi(x, y)$:

$$D\phi(x, y) = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix}.$$

Die Determinante dieser Matrix ist 1 und somit erhalten wir:

$$= (2\pi)^{-1} \int_0^{2\pi} \int_{\Omega} \int_{\Omega} e^{-ik\phi - i\omega(x \cos \phi + y \sin \phi)} f(x, y, z + \delta(-x \sin \phi + y \cos \phi)) dx dy d\phi.$$

Wir benutzen nun ebene Polarkoordinaten und substituieren $x = \sigma \cos \beta$, $y = \sigma \sin \beta$ mit $\sigma \in]0, R]$ und $\beta \in]0, 2\pi]$ und erhalten:

$$= (2\pi)^{-1} \int_0^{2\pi} \int_0^R \int_0^{2\pi} \sigma e^{-ik\phi - i\omega \cos(\beta - \phi)} f(\sigma \cos \beta, \sigma \sin \beta, z - \delta \sigma \sin(\phi - \beta)) d\phi d\sigma d\beta.$$

Wir substituieren nun $\phi = \tilde{\phi} + \beta$ und somit wird das obige Integral zu:

$$= (2\pi)^{-1} \int_0^R \int_0^{2\pi} \int_{-\beta}^{2\pi - \beta} \sigma e^{-ik\phi} e^{-ik\beta} e^{-i\omega \sigma \cos \phi} f(\sigma \cos \beta, \sigma \sin \beta, z - \delta \sigma \sin \phi) d\phi d\beta d\sigma.$$

Das Integral von $-\beta$ bis $2\pi - \beta$ zerlegen wir in die Summe der Integrale von $-\beta$ bis 0 und von 0 bis $2\pi - \beta$, dann substituieren wir den ersten Summanden mit $\phi = \phi - 2\pi$ und erhalten dasselbe Integral wie zuvor, nur in den Grenzen von $2\pi - \beta$ bis 2π . Die Summe dieses neuen Integrals mit dem alten zweiten Summanden kann man auch als das Integral von 0 bis 2π zusammenfassen und somit haben wir:

$$= (2\pi)^{-1} \int_0^R \int_0^{2\pi} \int_0^{2\pi} \sigma e^{-ik\phi} e^{-ik\beta} e^{-i\omega \sigma \cos \phi} f(\sigma \cos \beta, \sigma \sin \beta, z - \delta \sigma \sin \phi) d\phi d\beta d\sigma.$$

□

Wegen der abgeschnittenen Daten ist (4.41) nicht invariant für axiale Translation, da die gemessene z -Variable auf $|z| \leq \frac{L_D}{2} - |\delta| R_D$ eingeschränkt ist.

Zur Erinnerung: der 3DRP-Algorithmus überwindet dieses Problem, indem anfangs die noch fehlenden Daten geschätzt werden und somit anschließend die Daten für alle z mit $|z| \leq \frac{L_D}{2} + |\delta| R_D$ vorliegen. Das Schätzen der abgeschnittenen Daten wurde durch Vorwärtsprojektion verwirklicht.

Ähnlich dazu nehmen wir jetzt hier an, dass die Daten für alle z extrapoliert wurden. Die Invarianz für axiale Translation ist somit wiederhergestellt und wir können das folgende Lemma formulieren:

Lemma 4.13. Sei $f \in \mathcal{S}(\mathbb{R}^3)$, $\text{supp } f \subset \text{Zylinder mit Radius } R$, $p(x_r, \phi, z, \delta)$ bekannt für alle $-R < x_r < R$, $0 \leq \phi < \pi$, $|z| \leq \frac{L_D}{2} + |\delta| R_D$ und $|\delta| \leq \frac{L_D}{2R_D}$. Dann lautet die 3D FT in Bezug auf die Variablen x_r , ϕ und z :

$$\hat{p}_3(\omega, k, \omega_z, \delta) = (2\pi)^{-3/2} \int_0^R \int_0^{2\pi} \sigma e^{-ik\beta} F(\sigma, \beta, \omega_z) e^{-ik \arctan(\frac{\delta\omega_z}{\omega})} 2\pi (-i)^k \cdot J_k(\sigma |\omega| \sqrt{1 + \frac{\delta^2 \omega_z^2}{\omega^2}}) d\beta d\sigma, \quad (4.48)$$

wobei

$$F(\sigma, \beta, \omega_z) := \int_{-L_D/2}^{L_D/2} e^{-i\omega_z z} f(\sigma \cos \beta, \sigma \sin \beta, z) dz \quad (4.49)$$

und

$$J_k = \frac{1}{2\pi i} \int_C e^{x/2(z-1/z)} z^{-k-1} dz$$

die **Besselfunktion** ist. Dabei ist $k \in \mathbb{Z}$, C ein Kreis mit Radius r um den Ursprung und wenn wir $z = re^{i\phi}$ setzen, $0 \leq \phi \leq 2\pi$, dann können wir J_k auch in folgender Form ausdrücken:

$$J_k = \frac{1}{2\pi} r^{-k} \int_0^{2\pi} e^{x/2(re^{i\phi} - r^{-1}e^{-i\phi}) - ik\phi} d\phi.$$

Beweis.

$$\begin{aligned} \hat{p}_3(\omega, k, \omega_z, \delta) &= (2\pi)^{-1/2} \int_{\mathbb{R}} e^{-iz\omega_z} \hat{p}_2(\omega, k, z, \delta) dz \\ &\stackrel{\text{Lemma 4.12 b)}}{=} (2\pi)^{-3/2} \int_0^R \int_0^{2\pi} \int_0^{2\pi} \int_{\mathbb{R}} \sigma e^{-iz\omega_z} e^{-ik\beta} e^{-ik\phi} e^{-i\omega\sigma \cos \phi} \cdot \\ &\quad f(\sigma \cos \beta, \sigma \sin \beta, z - \sigma\delta \sin \phi) dz d\phi d\beta d\sigma. \end{aligned}$$

Wir substituieren jetzt $z = \tilde{z} + \sigma\delta \sin \phi$ und berücksichtigen, dass $f(x, y, z) = 0$ ist, wenn

$(x, y, z) \notin \Omega$ ist und erhalten dadurch:

$$\begin{aligned}
 &= (2\pi)^{-3/2} \int_0^R \int_0^{2\pi} \int_0^{2\pi} \sigma e^{-i(\omega_z \sigma \delta \sin \phi + k\beta + k\phi + \omega \sigma \cos \phi)} \underbrace{\left[\int_{-L_D/2}^{L_D/2} e^{-iz\omega_z} f(\sigma \cos \beta, \sigma \sin \beta, z) dz \right]}_{=: F(\sigma, \beta, \omega_z)} \\
 &\quad d\phi d\beta d\sigma \\
 &= (2\pi)^{-3/2} \int_0^R \sigma \int_0^{2\pi} e^{-ik\beta} F(\sigma, \beta, \omega_z) \int_0^{2\pi} e^{-i(k\phi + \omega \sigma \cos \phi + \omega_z \sigma \delta \sin \phi)} d\phi d\beta d\sigma.
 \end{aligned}$$

Da

$$\int_0^{2\pi} e^{-ik\phi - i\omega \sigma \cos \phi - i\omega_z \sigma \delta \sin \phi} d\phi = \int_0^{2\pi} \underbrace{e^{-ik\phi}}_{=: Y_k(\phi)} e^{-i\|v\| \langle v_e, \omega(\phi) \rangle} d\phi$$

mit

$$\omega(\phi) = \begin{pmatrix} \cos \phi \\ \sin \phi \end{pmatrix}, \quad v = \begin{pmatrix} \omega \sigma \\ \omega_z \sigma \delta \end{pmatrix},$$

$\|v\| = |\omega| \sigma \sqrt{1 + \frac{\omega_z^2 \delta^2}{\omega^2}}$ und $v_e = \frac{1}{\|v\|} v$, können wir Formel (3.19) aus [12] S. 198 anwenden und erhalten mit $n = 2$:

$$\begin{aligned}
 \int_0^{2\pi} Y_k(\phi) e^{-i\|v\| \langle v_e, \omega(\phi) \rangle} d\phi &= 2\pi (-i)^k J_k(\|v\|) \underbrace{Y_k(v_e)}_{=e^{ik \arctan\left(\frac{v_{e,2}}{v_{e,1}}\right)}} \\
 &= 2\pi (-i)^k J_k(\|v\|) e^{ik \arctan\left(\frac{\omega_z \sigma \delta |\omega| \sigma \sqrt{1 + \frac{\delta^2 \omega_z^2}{\omega^2}}}{\omega \sigma |\omega| \sigma \sqrt{1 + \frac{\delta^2 \omega_z^2}{\omega^2}}}\right)} \\
 &= 2\pi (-i)^k J_k(\|v\|) e^{ik \arctan\left(\frac{\delta \omega_z}{\omega}\right)},
 \end{aligned}$$

mit $\arctan : \mathbb{R} \rightarrow]-\frac{\pi}{2}, \frac{\pi}{2}[$. Damit folgt die Behauptung. □

Die exakte Rebinningformel ist eine direkte Konsequenz aus (4.48). Wir beachten, dass der zweite Exponentialfaktor und die Quadratwurzel unabhängig von den Integrationsvariablen σ und β sind und schreiben (4.48) für den speziellen Fall des geraden Sinogramms mit $\delta = 0$:

Satz 4.14. (Basisgleichung des FOREX-Algorithmus)

Es gelten dieselben Voraussetzungen wie in Lemma 4.13.

Dann gilt für die Beziehung zwischen der 3D FT der schrägen und der geraden Sino-gramme:

$$\hat{p}_3(\omega, k, \omega_z, \delta) = e^{-ik \arctan\left(\frac{\delta\omega_z}{\omega}\right)} \underbrace{\hat{p}_3\left(\omega\sqrt{1 + \frac{\delta^2\omega_z^2}{\omega^2}}, k, \omega_z, \delta = 0\right)}_{=:\hat{p}_{reb}(\omega\sqrt{1 + \frac{\delta^2\omega_z^2}{\omega^2}}, k, \omega_z)} . \quad (4.50)$$

Für $\omega' := \omega\sqrt{1 + \frac{\delta^2\omega_z^2}{\omega^2}}$, $|\omega'| \geq |\delta\omega_z|$ und $\arcsin :]-1, 1] \rightarrow [-\frac{\pi}{2}, \frac{\pi}{2}]$ ist die obige Formel äquivalent zu:

$$\hat{p}_{reb}(\omega', k, \omega_z) = e^{ik \arcsin\left(\frac{\delta\omega_z}{\omega'}\right)} \hat{p}_3\left(\omega'\sqrt{1 - \frac{\delta^2\omega_z^2}{\omega'^2}}, k, \omega_z, \delta\right). \quad (4.51)$$

Beweis. Aus Lemma 4.13 folgt für $\delta = 0$:

$$\hat{p}_{reb}(\omega', k, \omega_z) = (2\pi)^{-3/2} \int_0^R \sigma \int_0^{2\pi} e^{-ik\beta} F(\sigma, \beta, \omega_z) 2\pi(-i)^k J_k(\sigma | \omega' |) d\beta d\sigma. \quad (4.52)$$

Wir setzen in (4.52) $\omega' := \omega\sqrt{1 + \frac{\delta^2\omega_z^2}{\omega^2}}$ (d. h. ω und ω' haben dasselbe Vorzeichen), multiplizieren beide Seiten der Gleichung mit $e^{-ik \arctan\left(\frac{\delta\omega_z}{\omega}\right)}$ und erhalten:

$$\begin{aligned} & e^{-ik \arctan\left(\frac{\delta\omega_z}{\omega}\right)} \hat{p}_{reb}\left(\omega\sqrt{1 + \frac{\delta^2\omega_z^2}{\omega^2}}, k, \omega_z\right) \\ = & (2\pi)^{-3/2} \int_0^R \sigma \int_0^{2\pi} e^{-ik\beta} F(\sigma, \beta, \omega_z) e^{-ik \arctan\left(\frac{\delta\omega_z}{\omega}\right)} 2\pi(-i)^k \cdot \\ & J_k(\sigma | \omega | \sqrt{1 + \frac{\delta^2\omega_z^2}{\omega^2}}) d\beta d\sigma \end{aligned}$$

$$\stackrel{\text{Lemma 4.13}}{=} \hat{p}_3(\omega, k, \omega_z, \delta).$$

Für $\omega' := \omega\sqrt{1 + \frac{\delta^2\omega_z^2}{\omega^2}}$ mit $|\omega| > |\delta\omega_z|$ gilt:

$$\omega'^2 = \omega^2 + \delta^2\omega_z^2$$

und somit gilt für ω (da ω und ω' dasselbe Vorzeichen haben):

$$\omega = \omega' \sqrt{1 - \frac{\delta^2\omega_z^2}{\omega'^2}}. \quad (4.53)$$

Setzen wir (4.53) in

$$\hat{p}_3(\omega, k, \omega_z, \delta) = e^{-ik \arctan\left(\frac{\delta\omega_z}{\omega}\right)} \hat{p}_{reb}\left(\omega \sqrt{1 + \frac{\delta^2\omega_z^2}{\omega^2}}, k, \omega_z\right)$$

ein, so folgt:

$$\begin{aligned} & \hat{p}_3\left(\omega' \sqrt{1 - \frac{\delta^2\omega_z^2}{\omega'^2}}, k, \omega_z, \delta\right) \\ = & e^{-ik \arctan\left(\frac{\delta\omega_z}{\omega' \sqrt{1 - \frac{\delta^2\omega_z^2}{\omega'^2}}}\right)} \underbrace{\hat{p}_{reb}\left(\omega' \sqrt{1 - \frac{\delta^2\omega_z^2}{\omega'^2}} \sqrt{1 + \frac{\delta^2\omega_z^2}{\omega'^2(1 - \frac{\delta^2\omega_z^2}{\omega'^2})}}, k, \omega_z\right)}_{=\omega'} \end{aligned}$$

Da

$$\frac{\delta\omega_z}{\omega' \sqrt{1 - \frac{\delta^2\omega_z^2}{\omega'^2}}} = \frac{\frac{\delta\omega_z}{\omega'}}{\sqrt{1 - \sin^2(\arcsin \frac{\delta\omega_z}{\omega'})}} = \frac{\sin(\arcsin \frac{\delta\omega_z}{\omega'})}{\cos(\arcsin \frac{\delta\omega_z}{\omega'})} = \tan(\arcsin \frac{\delta\omega_z}{\omega'})$$

erhalten wir:

$$\begin{aligned} \hat{p}_3(\omega, k, \omega_z, \delta) &= e^{-ik \arctan\left(\frac{\delta\omega_z}{\omega}\right)} \hat{p}_{reb}\left(\omega \sqrt{1 + \frac{\delta^2\omega_z^2}{\omega^2}}, k, \omega_z\right) \\ \Leftrightarrow \hat{p}_3\left(\omega' \sqrt{1 - \frac{\delta^2\omega_z^2}{\omega'^2}}, k, \omega_z, \delta\right) &= e^{-ik \arctan(\tan \arcsin \frac{\delta\omega_z}{\omega'})} \hat{p}_{reb}(\omega', k, \omega_z) \\ \Leftrightarrow e^{ik \arcsin\left(\frac{\delta\omega_z}{\omega'}\right)} \hat{p}_3\left(\omega' \sqrt{1 - \frac{\delta^2\omega_z^2}{\omega'^2}}, k, \omega_z, \delta\right) &= \hat{p}_{reb}(\omega', k, \omega_z). \end{aligned}$$

□

Herleitung 2:

Bei dieser Herleitung halten wir uns an die Vorgaben von [15].

Auch hier nehmen wir zuerst einmal an, dass die Daten für alle x_r und z bekannt sind, denn dann können wir die 2D FT der 2D parallelen Projektion, die einem fixierten Paar (ϕ, δ) entspricht, berechnen.

Um die exakte Rebinningformel herleiten zu können, starten wir mit dem 3D Projektionssatz.

Satz 4.15. (Projektionssatz für den 3D Fall und Daten im Sinogrammformat)

Sei $f \in \mathcal{S}(\mathbb{R}^3)$, $\text{supp } f \subset \text{Zylinder mit Radius } R$, $p(x_r, \phi, z, \delta)$ bekannt für alle $-R < x_r < R$, $0 \leq \phi < \pi$, $|z| \leq \frac{L_D}{2} + |\delta| R_D$ und $|\delta| \leq \frac{L_D}{2R_D}$, dann gilt:

$$\hat{p}_2(\omega, \phi, \omega_z, \delta) = (2\pi)^{1/2} \hat{f}_3(\omega \cos \phi + \omega_z \delta \sin \phi, \omega \sin \phi - \omega_z \delta \cos \phi, \omega_z), \quad (4.54)$$

wobei auf der linken Seite die 2D FT bzgl. x_r und z und auf der rechten Seite die 3D FT von $f(x, y, z)$ steht.

Beweis. Aus dem Projektionssatz 2.31 mit

$$\omega = \begin{pmatrix} -\sin \phi \\ \cos \phi \\ \delta \end{pmatrix}, \quad a = \begin{pmatrix} x_r \cos \phi \\ x_r \sin \phi \\ z \end{pmatrix}, \quad \eta = \begin{pmatrix} \omega \cos \phi + \omega_z \delta \sin \phi \\ \omega \sin \phi - \omega_z \delta \cos \phi \\ \omega_z \end{pmatrix}$$

folgt:

$$\begin{aligned} \hat{p}_2(\omega, \phi, \omega_z, \delta) &= (2\pi)^{-1} \int_{\mathbb{R}^2} p(x_r, \phi, z, \delta) e^{-i(x_r \omega + z \omega_z)} dx_r dz \\ &= (2\pi)^{-1} \int_{\omega^\perp} (Pf)(\omega, a) e^{-i(a, \eta)} da \\ &\stackrel{\text{Proj.satz 2.31}}{=} (2\pi)^{1/2} \hat{f}_3(\eta). \end{aligned}$$

□

Korollar 4.16. Für Gleichung (4.54) gilt:

$$\hat{p}_2(\omega, \phi, \omega_z, \delta) = (2\pi)^{1/2} \hat{f}_3(\omega' \cos(\phi - \sigma), \omega' \sin(\phi - \sigma), \omega_z), \quad (4.55)$$

wobei $\sigma := \arctan\left(\frac{\delta \omega_z}{\omega}\right)$ und $\omega' := \omega \sqrt{1 + \frac{\delta^2 \omega_z^2}{\omega^2}}$ ist.

Beweis. 1. Sei $\frac{\delta \omega_z}{\omega} \geq 0$, also $\sigma \in [0, \frac{\pi}{2}]$.

Da

$$\begin{aligned} \omega' \cos(\phi - \sigma) &= \omega \sqrt{1 + \tan^2 \sigma} \cos \phi \cos \sigma + \omega \sqrt{1 + \tan^2 \sigma} \sin \phi \sin \sigma \\ &= \omega \cos \phi \underbrace{\sqrt{\cos^2 \sigma + \tan^2 \sigma \cos^2 \sigma}}_{=1} + \omega \sin \phi \underbrace{\sqrt{\sin^2 \sigma + \tan^2 \sigma \sin^2 \sigma}}_{=\sqrt{\tan^2 \sigma} = \frac{\sqrt{\delta^2 \omega_z^2}}{\sqrt{\omega^2}}} \\ &= \omega \cos \phi + \delta \omega_z \sin \phi \end{aligned}$$

und

$$\begin{aligned}
 \omega' \sin(\phi - \sigma) &= \omega \sqrt{1 + \tan^2 \sigma} \sin \phi \cos \sigma - \omega \sqrt{1 + \tan^2 \sigma} \cos \phi \sin \sigma \\
 &= \omega \sin \phi \underbrace{\sqrt{\cos^2 \sigma + \sin^2 \sigma}}_{=1} - \omega \cos \phi \sqrt{\tan^2 \sigma} \\
 &= \omega \sin \phi - \delta \omega_z \cos \phi
 \end{aligned}$$

folgt die Behauptung.

2. Für $\frac{\delta \omega_z}{\omega} < 0$, also $\sigma \in] -\frac{\pi}{2}, 0[$ erhalten wir analog zum obigen Fall dieselben Ergebnisse und so folgt auch hier die Behauptung. □

Die Rebinninggleichung wird erzielt, indem man die 1D FT von (4.55) bzgl. der Variablen ϕ berechnet.

Lemma 4.17. *Es seien dieselben Voraussetzungen wie in Satz 4.15 gegeben. Dann lautet die 3D FT in Bezug auf die Variablen x_r , ϕ und z :*

$$\hat{p}_3(\omega, k, \omega_z, \delta) = e^{-ik\sigma} \int_0^{2\pi} \hat{f}_3(\omega' \cos \phi, \omega' \sin \phi, \omega_z) e^{-ik\phi} d\phi, \quad (4.56)$$

wobei σ und ω' wie im vorangegangenen Korollar definiert sind.

Beweis. Aus Korollar 4.16 und anschließender Substitution von $\phi = \sigma + \alpha$ erhalten wir:

$$\begin{aligned}
 \hat{p}_3(\omega, k, \omega_z, \delta) &= \int_0^{2\pi} \hat{f}_3(\omega' \cos(\phi - \sigma), \omega' \sin(\phi - \sigma), \omega_z) e^{-ik\phi} d\phi \\
 &= \int_{-\sigma}^0 \hat{f}_3(\omega' \cos \alpha, \omega' \sin \alpha, \omega_z) e^{-ik\sigma} e^{-ik\alpha} d\alpha \\
 &\quad + \int_0^{2\pi-\sigma} \hat{f}_3(\omega' \cos \alpha, \omega' \sin \alpha, \omega_z) e^{-ik\sigma} e^{-ik\alpha} d\alpha.
 \end{aligned}$$

Substituieren wir nun im 1. Integral $\alpha = \phi - 2\pi$ so folgt:

$$\begin{aligned}
 \hat{p}_3(\omega, k, \omega_z, \delta) &= \int_{2\pi-\sigma}^{2\pi} \hat{f}_3(\underbrace{\omega' \cos(\phi - 2\pi)}_{=\cos \phi}, \underbrace{\omega' \sin(\phi - 2\pi)}_{=\sin \phi}, \omega_z) e^{-ik\sigma} e^{-ik\phi} \underbrace{e^{ik2\pi}}_{=1} d\phi \\
 &\quad + \int_0^{2\pi-\sigma} \hat{f}_3(\omega' \cos \alpha, \omega' \sin \alpha, \omega_z) e^{-ik\sigma} e^{-ik\alpha} d\alpha \\
 &= e^{-ik\sigma} \int_0^{2\pi} \hat{f}_3(\omega' \cos \phi, \omega' \sin \phi, \omega_z) e^{-ik\phi} d\phi.
 \end{aligned}$$

□

Setzt man in der Gleichung (4.56) $\delta = 0$ und vergleicht dieses Resultat mit der Gleichung für beliebiges δ , so führt uns dieser Vergleich ebenfalls zu der Basisgleichung der FOREX-Algorithmus und wir erhalten wieder Satz 4.14.

Beweis. Hier wollen wir nun Satz 4.14 mit Hilfe des Lemmas 4.17 beweisen:

Aus Lemma 4.17 folgt für $\delta = 0$ und $\omega' = \omega \sqrt{1 + \frac{\delta^2 \omega_z^2}{\omega^2}}$:

$$\begin{aligned}
 \hat{p}_3(\omega', k, \omega_z) &\stackrel{(4.56)}{=} e^0 \int_0^{2\pi} \hat{f}_3(\underbrace{\omega' \sqrt{1 + \frac{0}{\omega^2}} \cos \phi}_{=\omega' \cos \phi}, \underbrace{\omega' \sqrt{1 + \frac{0}{\omega^2}} \sin \phi}_{\omega' \sin \phi}, \omega_z) e^{-ik\phi} d\phi \\
 &\stackrel{(4.56)}{=} e^{ik\sigma} \hat{p}_3(\omega, k, \omega_z, \delta) \\
 &= e^{ik\sigma} \hat{p}_3(\omega' \sqrt{1 - \frac{\delta^2 \omega_z^2}{\omega^2}}, k, \omega_z, \delta).
 \end{aligned}$$

Der Beweis der Äquivalenz ist vollkommen identisch zur ersten Version des Beweises von Satz 4.14.

□

Der FOREX-Algorithmus

Ab hier orientieren wir uns wieder an [14].

Das Ziel eines Rebinning-Algorithmus ist es, aus den 3D Daten (4.41) einen Stapel gerader Sinogramme $p_{reb}(x_r, \phi, z)$ zu schätzen.

Wie wir gerade gesehen haben liefert die Basisgleichung des FOREX-Algorithmus für jeden Wert der Ringdifferenz δ , der der Bedingung $|\omega'| \geq |\delta \omega_z|$ genügt, eine unabhängige Schätzung von $\hat{p}_3(\omega', k, \omega_z, 0)$. In Abwesenheit von Rauschen sind alle diese Schätzungen gleich.

Da die Daten jedoch verrauscht sind, sollte man die rebinnten Daten unter Verwendung der gesamten 3D Datenmenge schätzen, um eine größere SNR zu erhalten. Der FOREX-Algorithmus erreicht dieses Ziel durch Mittelung über alle Schätzungen:

$$\hat{p}_{reb}(\omega', k, \omega_z) = \frac{1}{\delta_1(\omega', \omega_z)} \int_0^{\delta_1} \hat{p}_3(\omega' \sqrt{1 - \frac{\delta^2 \omega_z^2}{\omega'^2}}, k, \omega_z, \delta) e^{ik \arcsin(\frac{\delta \omega_z}{\omega'})} d\delta, \quad (4.57)$$

wobei $\delta \leq |\frac{\omega'}{\omega_z}|$ und $|\delta| \leq \frac{L_D}{2R_D}$ ist, also

$$\delta_1(\omega', \omega_z) := \min \left(\left| \frac{\omega'}{\omega_z} \right|, \frac{L_D}{2R_D} \right)$$

ist.

Wir beachten, dass in (4.57) nur positive Werte von δ erscheinen. Dies liegt daran, dass die Sinogramme mit negativen δ bei der Berechnung der 2D FT von Lemma 4.12 a) darin verschmolzen sind.

Leider kann Gleichung (4.57) wegen der abgeschnittenen Daten nicht direkt benutzt werden (da die Daten $p(x_r, \phi, z, \delta \neq 0)$ nur für $|z| \leq \frac{L_D}{2} - |\delta| R_D$ bekannt sind, kann die axiale FT nicht durchgeführt werden). Deshalb müssen, wie auch beim 3DRP, zuerst die abgeschnittenen Daten geschätzt werden, um komplette 2D Projektionen zu erhalten. Für diese Schätzung der abgeschnittenen Daten kann die 2D Datenmenge $p_{reb}(x_r, \phi, z)$ und somit die Basisgleichung des FOREX-Algorithmus verwendet werden. Dies führt uns zu folgendem Algorithmus:

Algorithmus

1. Initialisiere eine Menge rebinnter, fouriertransformierter Sinogramme:

$$\hat{p}_{reb}(\omega', k, \omega_z) = 0.$$

2. Berechne die 3D FT (bzw. 3D FFT) der gemessenen 2D Sinogrammteilmengen $p_{reb}(x_r, \phi, z)$:

$$\hat{p}_{reb}(\omega', k, \omega_z) = (2\pi)^{-3/2} \int_{-R}^R \int_0^{2\pi} \int_{-L_D/2}^{L_D/2} p_{reb}(x_r, \phi, z) e^{-i(\omega' x_r + k\phi + \omega_z z)} dz d\phi dx_r,$$

(da für $\delta = 0$ die Daten komplett sind, also $z \in [-\frac{L_D}{2}, \frac{L_D}{2}]$).

3. Betrachte der Reihe nach jeden Wert von δ und berechne:

- a) Berechne die 2D FT (bzw. 2D FFT) der Daten bzgl. x_r und ϕ um $\hat{p}_2(\omega, k, z, \delta)$ für alle gemessenen z : $|z| \leq \frac{L_D}{2} - |\delta| R_D$ zu erhalten:

$$\hat{p}_2(\omega, k, z, \delta) = (2\pi)^{-1} \int_{-R}^R \int_0^{2\pi} p(x_r, \phi, z, \delta) e^{-i(\omega x_r + k\phi)} d\phi dx_r.$$

4 Das C-C-Modell und darauf beruhende Bildrekonstruktionen

- b) Schätze $\hat{p}_3^{geschätzt}(\omega, k, \omega_z, \delta)$ aus $\hat{p}_{reb}(\omega', k, \omega_z)$ unter Verwendung der Basisgleichung:

$$\hat{p}_3^{geschätzt}(\omega, k, \omega_z, \delta) = e^{-ik \arctan\left(\frac{\delta\omega_z}{\omega}\right)} \hat{p}_{reb}\left(\underbrace{\omega \sqrt{1 + \frac{\delta^2\omega_z^2}{\omega^2}}}_{=\omega'}, k, \omega_z\right).$$

- c) Berechne die inverse 1D FT (bzw. 1D FFT) von $\hat{p}_3^{geschätzt}(\omega, k, \omega_z, \delta)$ bzgl. ω_z , um eine Schätzung von $\hat{p}_2(\omega, k, z, \delta)$ über dem fehlenden Bereich von z : $\frac{L_D}{2} - |\delta| R_D < |z| \leq \frac{L_D}{2} + |\delta| R_D$ zu erhalten:

$$\hat{p}_2^{geschätzt}(\omega, k, z, \delta) = (2\pi)^{-1/2} \int_{\mathbb{R}} \hat{p}_3^{geschätzt}(\omega, k, \omega_z, \delta) e^{i\omega_z z} d\omega_z.$$

- d) Nimm die Resultate aus 3 a) und 3 c) zusammen, um $\hat{p}_2(\omega, k, z, \delta)$ über dem kompletten Rang von z : $|z| \leq \frac{L_D}{2} + |\delta| R_D$ zu erhalten.
e) Berechne die 1D FT (bzw. 1D FFT) bzgl. z um $\hat{p}_3(\omega, k, \omega_z, \delta)$ zu erhalten:

$$\hat{p}_3(\omega, k, \omega_z, \delta) = (2\pi)^{-1/2} \int_{\mathbb{R}} \hat{p}_2(\omega, k, z, \delta) e^{-i\omega_z z} dz.$$

- f) **Rebinning:** Für jedes (ω', k, ω_z) benutze die Basisgleichung, um die rebinnten Daten $\hat{p}_{reb}(\omega' = \omega \sqrt{1 + \frac{\delta^2\omega_z^2}{\omega^2}}, k, \omega_z)$ zu erzeugen:

$$\hat{p}_{reb}(\omega', k, \omega_z) = \hat{p}_3(\omega, k, \omega_z, \delta) e^{ik \arctan\left(\frac{\delta\omega_z}{\omega}\right)}.$$

(Bei diesem Schritt verwendet man eine 1D Interpolation in ω .)

4. Wegen der verrauschten Daten berechne $\hat{p}_3(\omega', k, \omega_z)$ als Mittelung über alle in f) errechneten $\hat{p}_{reb}(\omega', k, \omega_z)$:

$$\hat{p}_3(\omega', k, \omega_z) := \frac{1}{\delta_1(\omega', \omega_z)} \int_0^{\delta_1} \hat{p}_3(\omega' \sqrt{1 - \frac{\delta^2\omega_z^2}{\omega'^2}}, k, \omega_z, \delta) e^{ik \arcsin\left(\frac{\delta\omega_z}{\omega'}\right)} d\delta.$$

5. Berechne die inverse 3D FT (bzw. 3D FFT) von $\hat{p}_3(\omega', k, \omega_z)$:

$$p_{reb}(x_r, \phi, z) = (2\pi)^{-3/2} \int_{\mathbb{R}^3} \hat{p}_3(\omega', k, \omega_z) e^{i(\omega' x_r + k\phi + \omega_z z)} d\omega_z dk d\omega'.$$

Wir erhalten also einen Stapel rebinnter Sinogramme:

$$p_{reb}(x_r, \phi, z) = p(x_r, \phi, z, \delta = 0).$$

6. Rekonstruiere jede Ebene separat und unabhängig voneinander unter Verwendung irgendeines beliebigen 2D Rekonstruktionsalgorithmus.
7. Stapele die rekonstruierten Ebenen, um ein Volumenbild zu erhalten.

Die Schritte 3b) - 3d) entsprechen dem Vorwärtsprojektionsschritt im 3DRP-Algorithmus.

4.6.2 FORE

Die Datendarstellung bei FORE und FOREX ist identisch. Die einzige Veränderung besteht darin, dass aus der exakten Rebinninggleichung FOREX eine Approximation, FORE, hergeleitet wird, die zu einer schnelleren Implementierung führen soll.

Wir setzen:

$$\alpha := \frac{\delta\omega_z}{\omega} \quad (4.58)$$

$$\Delta\Phi(\alpha) := k \arctan \alpha \quad \textit{Phasenshift} \quad (4.59)$$

$$\chi(\alpha) := \sqrt{1 + \alpha^2} \quad (4.60)$$

und aus unserer Basisgleichung für den FOREX-Algorithmus (4.50) wird:

$$\hat{p}_3(\omega, k, \omega_z, \delta) = e^{-i\Delta\Phi(\alpha)} \hat{p}_{reb}(\omega\chi(\alpha), k, \omega_z). \quad (4.61)$$

Eine Klasse von Rebinning-Algorithmen mit wachsender Genauigkeit erreichen wir, indem wir die abgeschnittene Taylorentwicklung um Null in α des Phasenshifts und von $\chi(\alpha)$ betrachten.

Fourier Rebinning Approximation

FORE ist ein Algorithmus, der darauf beruht, dass wir nur bis zum linearen Term α^1 entwickeln.

Satz 4.18. (Taylorentwicklung)

Für $\chi(\alpha)$ und $\Delta\Phi(\alpha)$ wie oben definiert gilt:

$$\chi(\alpha) = 1 + O(\alpha^2), \quad \Delta\Phi(\alpha) = k\alpha + O(\alpha^3).$$

Beweis. Taylorentwicklung ist möglich, da $\chi(\alpha), \Delta\Phi(\alpha) \in C^\infty(\mathbb{R})$.

Wir erhalten also:

$$\chi(0 + \alpha) = \sum_{i=0}^n \frac{\chi^{(i)}(0)}{i!} \alpha^i + \underbrace{\frac{\chi^{(n+1)}(\nu)}{(n+1)!} \alpha^{n+1}}_{\xrightarrow{n \rightarrow \infty} 0} \xrightarrow{n \rightarrow \infty} \sum_{i=0}^{\infty} \frac{\chi^{(i)}(0)}{i!} \alpha^i$$

und somit gilt:

$$\begin{aligned} \chi(0 + \alpha) &= \sum_{i=0}^{\infty} \frac{\chi^{(i)}(0)}{i!} \alpha^i \\ &= 1 + \alpha \cdot 0 + \frac{\alpha^2}{2} \cdot 1 + \sum_{i=3}^{\infty} \frac{\chi^{(i)}(0)}{i!} \alpha^i \\ &= 1 + O(\alpha^2). \end{aligned}$$

Mit denselben Überlegungen erhalten wir:

$$\begin{aligned}
 \Delta\Phi(\alpha) &= \sum_{i=0}^{\infty} \frac{\Delta\Phi^{(i)}(0)}{i!} \alpha^i \\
 &= 0 + \alpha \cdot k + \frac{\alpha^2}{2} \cdot 0 + \frac{\alpha^3}{6} (-2k) + \sum_{i=4}^{\infty} \frac{\Delta\Phi^{(i)}(0)}{i!} \alpha^i \\
 &= \alpha k + O(\alpha^3).
 \end{aligned}$$

□

Satz und Definition 4.19. (Basisgleichung des FORE-Algorithmus)

Ausgehend von Gleichung (4.61) können wir zeigen, dass gilt:

$$\hat{p}_2(\omega, k, z, \delta) \approx \underbrace{\hat{p}_2\left(\omega, k, z - \frac{k\delta}{\omega}, 0\right)}_{=:\hat{p}_{reb}(\omega, k, z - \frac{k\delta}{\omega})}. \quad (4.62)$$

Diese Gleichung heißt **Fourier-Rebinning-Approximation**.

Beweis. Ausgangsgleichung ist Gleichung (4.61). Setzen wir nun dort, wie in Satz 4.18 gezeigt, $\chi(\alpha) = 1 + O(\alpha^2)$ und $\Delta\Phi(\alpha) = k\alpha + O(\alpha^3)$ so folgt:

$$\begin{aligned}
 \hat{p}_3(\omega, k, \omega_z, \delta) &= e^{-ik\alpha} \underbrace{e^{-iO(\alpha^3)}}_{=1+O(\alpha^3)} (2\pi)^{-1/2} \int_{-R}^R \hat{p}_{reb}(x_r, k, \omega_z) \underbrace{e^{-ix_r(\omega(1+O(\alpha^2)))}}_{=e^{-ix_r\omega(1+O(\alpha^2))}} dx_r \\
 &= e^{-ik\alpha} (1 + O(\alpha^3)) [\hat{p}_{reb}(\omega, k, \omega_z) (1 + O(\alpha^2))] \\
 &= e^{-ik\alpha} \hat{p}_{reb}(\omega, k, \omega_z) + O(\alpha^3) + O(\alpha^2) + O(\alpha^5) \\
 &= e^{-ik\alpha} \hat{p}_{reb}(\omega, k, \omega_z) + O(\alpha^2).
 \end{aligned}$$

Damit erhalten wir:

$$\begin{aligned}
 \hat{p}_2(\omega, k, z, \delta) &= (2\pi)^{-1/2} \int_{\mathbb{R}} \left[e^{-ik\frac{\delta\omega_z}{\omega}} \hat{p}_{reb}(\omega, k, \omega_z) + O(\alpha^2) \right] e^{i\omega_z z} d\omega_z \\
 &= (2\pi)^{-1/2} \int_{\mathbb{R}} \hat{p}_{reb}(\omega, k, \omega_z) e^{i\omega_z(z - \frac{k\delta}{\omega})} d\omega_z + O(\alpha^2) (2\pi)^{-1/2} \int_{\mathbb{R}} e^{i\omega_z z} d\omega_z \\
 &\approx \hat{p}_{reb}\left(\omega, k, z - \frac{k\delta}{\omega}\right).
 \end{aligned}$$

□

Diese Basisgleichung bringt die 2D FT eines schrägen Sinogramms (z, δ) mit der 2D FT des geraden Sinogramms einer Ebene - axial geshiftet durch einen frequenzabhängigen Ausgleich $\Delta z = -\frac{k\delta}{\omega}$ - in Verbindung.

Im Gegensatz zum exakten Rebinning, basierend auf der Basisgleichung des FOREX-Algorithmus (4.50), benötigt Fourier-Rebinning keine FT bzgl. z . Dies führt zu einer beachtlichen Vereinfachung der Implementierung und zwar nicht nur wegen der kleineren Anzahl an FFT's, sondern eher wegen der Daten, die nicht länger für alle Werte von z bekannt sein müssen. Somit benötigen wir also keine Schätzung der abgeschnittenen Daten, was wiederum den Aufwand der Implementierung stark reduziert.

Schätzung der Genauigkeit von Fourier-Rebinning

Entwickeln wir bis zum Term α^2 so kommen wir auf folgendes Ergebnis:

Satz 4.20. *(Taylorentwicklung bis zum Term α^2)*

Für $\chi(\alpha)$ und $\Delta\Phi(\alpha)$ wie in Satz 4.18 gilt:

$$\chi(\alpha) = 1 + \frac{\alpha^2}{2} + O(\alpha^4), \quad \Delta\Phi(\alpha) = k\alpha + O(\alpha^3).$$

Beweis.

$$\begin{aligned} \chi(\alpha) &= \sum_{i=0}^{\infty} \frac{\chi^{(i)}(0)}{i!} \alpha^i \\ &= 1 + \alpha \cdot 0 + \frac{\alpha^2}{2} \cdot 1 + \frac{\alpha^3}{6} \cdot 0 - 3 \frac{\alpha^4}{24} + \sum_{i=5}^{\infty} \frac{\chi^{(i)}(0)}{i!} \alpha^i \\ &= 1 + \frac{\alpha^2}{2} + O(\alpha^4). \end{aligned}$$

Für den zweiten Term können wir dieselbe Rechnung wie im Beweis von Satz 4.18 durchführen und erhalten die Behauptung. □

Satz 4.21. *Ausgehend von Gleichung (4.61) können wir des Weiteren zeigen, dass gilt:*

$$\hat{p}_2(\omega, k, z, \delta) \approx \hat{p}_{reb}(\omega, k, z - \frac{k\delta}{\omega}) - \frac{\delta^2}{2\omega} \frac{\partial}{\partial \omega} \frac{\partial^2}{\partial z^2} \hat{p}_{reb}(\omega, k, z - \frac{k\delta}{\omega}). \quad (4.63)$$

Beweis. Ausgangsgleichung ist auch hier Gleichung (4.61). Setzen wir nun dort, wie in Satz 4.20 gezeigt, $\chi(\alpha) = 1 + \frac{\alpha^2}{2} + O(\alpha^4)$ und $\Delta\Phi(\alpha) = k\alpha + O(\alpha^3)$ so folgt:

$$\begin{aligned}
 \hat{p}_3(\omega, k, z, \delta) &= \underbrace{e^{-i(k\alpha + O(\alpha^3))}}_{=e^{-ik\alpha + O(\alpha^3)}} \underbrace{\hat{p}_{reb}(\omega(1 + \frac{\alpha^2}{2} + O(\alpha^4)), k, \omega_z)}_{=\hat{p}_{reb}(\omega(1 + \frac{\alpha^2}{2}), k, \omega_z) + O(\alpha^4)} \\
 &= e^{-ik\alpha} \hat{p}_{reb}(\omega(1 + \frac{\alpha^2}{2}), k, \omega_z) + O(\alpha^3) + O(\alpha^4) + O(\alpha^7) \\
 &= e^{-ik\alpha} \hat{p}_{reb}(\omega(1 + \frac{\alpha^2}{2}), k, \omega_z) + O(\alpha^3) \\
 &\stackrel{\text{Taylor entw. um } \omega}{=} e^{-i\frac{k\delta\omega_z}{\omega}} \left(\hat{p}_{reb}(\omega, k, \omega_z) + \frac{\delta^2\omega_z^2}{2\omega} \frac{\partial}{\partial\omega} \hat{p}_{reb}(\omega, k, \omega_z) \right) + O(\alpha^3).
 \end{aligned}$$

Damit erhalten wir:

$$\begin{aligned}
 \hat{p}_2(\omega, k, z, \delta) &= (2\pi)^{-1/2} \int_{\mathbb{R}} \hat{p}_{reb}(\omega, k, \omega_z) e^{i\omega_z(z - \frac{k\delta}{\omega})} d\omega_z \\
 &\quad + (2\pi)^{-1/2} \int_{\mathbb{R}} e^{-i\frac{k\delta\omega_z}{\omega}} \frac{\delta^2\omega_z^2}{2\omega} \frac{\partial}{\partial\omega} \hat{p}_{reb}(\omega, k, \omega_z) e^{i\omega_z z} d\omega_z + O(\alpha^3) \\
 &= \hat{p}_{reb}(\omega, k, z - \frac{k\delta}{\omega}) - (2\pi)^{-1/2} \frac{\delta^2}{2\omega} \frac{\partial}{\partial\omega} \int_{\mathbb{R}} \underbrace{\hat{p}_{reb}(\omega, k, \omega_z) e^{i\omega_z(z - \frac{k\delta}{\omega})} (-\omega_z^2)}_{=\frac{\partial^2}{\partial z^2} \hat{p}_{reb}(\omega, k, \omega_z) e^{i\omega_z(z - \frac{k\delta}{\omega})}} d\omega_z \\
 &\quad + O(\alpha^3) \\
 &\approx \hat{p}_{reb}(\omega, k, z - \frac{k\delta}{\omega}) - \frac{\delta^2}{2\omega} \frac{\partial}{\partial\omega} \frac{\partial^2}{\partial z^2} \underbrace{(2\pi)^{-1/2} \int_{\mathbb{R}} \hat{p}_{reb}(\omega, k, \omega_z) e^{i\omega_z(z - \frac{k\delta}{\omega})} d\omega_z}_{=\hat{p}_{reb}(\omega, k, z - \frac{k\delta}{\omega})}.
 \end{aligned}$$

□

Der zweite Term auf der rechten Seite von Gleichung (4.63) ist eine Korrektur der Basisgleichung FORE (4.62) und seine Größe ist eine Schätzung der Genauigkeit der Fourier-Rebinning-Approximation.

Das Verhalten des Korrekturterms bei niedrigen Frequenzen ω kann durch die Betrachtung der **Konsistenzbedingung** für die 2D Radontransformation verstanden werden.

Satz 4.22. (Konsistenzbedingung)

Für die 2D FT eines Sinogramms ($z, \delta = 0$) gilt:

$$\hat{p}_{reb}(\omega, k, z) \approx 0 \quad \text{für } |\omega| < \frac{|k|}{R}, \quad (4.64)$$

wobei R der Radius des FOV ist.

Beweis. Der Beweis dieses Satzes führt zu weit, nachzulesen ist er in: "P. R. Edholm, R. M. Lewitt and B. Lindholm, Novel properties of Fourier decomposition of the sinogram, Int. Workshop on Physics and Engineering of Computerised Multidimensional Imaging and Processing (SPIE 671), 1986, pp. 8-18."

□

Die Gleichung (4.64) zeigt, dass für große Werte von k die Konsistenzbedingung die Daten zwingt, genau in der niedrig- ω -Region, wo die Fourier-Rebinning-Approximation wegen der $\frac{1}{\omega}$ Abhängigkeit im Korrekturterm von (4.63) ungültig wäre, Null zu sein.

Drei andere Eigenschaften lassen sich ebenfalls aus Gleichung (4.63) ableiten:

1. Fourier Rebinning ist exakt, wenn $f(x, y, z)$ linear von z abhängt. In diesem Fall ist $\hat{p}_2(\omega, k, z, \delta)$ auch linear in z und daher verschwindet $\frac{\partial^2}{\partial z^2} \hat{p}_{reb}(\omega, k, z - \frac{k\delta}{\omega})$.
2. Der zweite Term auf der rechten Seite von Gleichung (4.63) (=Korrekturterm) ist proportional zum Quadrat der Ringdifferenz δ ($\delta = \tan \theta$). Die Fourier-Rebinning-Approximation "bricht daher zusammen", wenn die axiale Öffnung des Scanners zu groß wird. In diesem Fall sollte der exakte Rebinning-Algorithmus anstelle der Fourier-Rebinning-Approximation benutzt werden. (Für große δ wird der Korrekturterm zu groß und damit ist das Weglassen des Korrekturterms, um die Basisgleichung des FORE zu erhalten, nicht mehr gerechtfertigt).
3. Die tatsächliche Berechnung des Korrekturterms (um die Genauigkeit der Rekonstruktion zu verbessern) ist unpraktikabel, da die zweite Ableitung nach z das Rauschen sehr verstärken würde.

Das Prinzip des Algorithmus

Wie wir in dem Abschnitt "Fourier-Rebinning-Approximation" gesehen haben, basiert Fourier-Rebinning auf einer Hochfrequenz-Approximation. Deshalb ist es nötig, die hohen und niedrigen Frequenzen getrennt zu behandeln. Dazu unterteilen wir die (ω, k) -Ebene in drei Regionen, die durch die beiden Parameter k_{lim} und ω_{lim} definiert werden. In jeder Region wird eine andere Methode zur Schätzung von $\hat{p}_{reb}(\omega, k, z)$ angewendet.

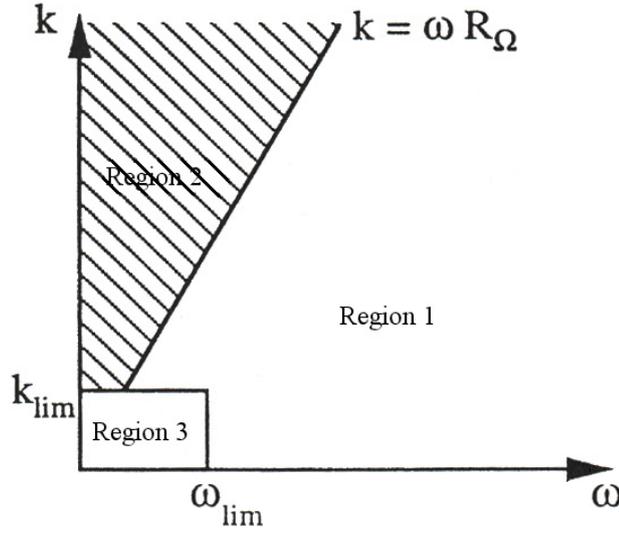


Abbildung 4.25: Ein Quadrant der (ω, k) -Ebene und seine Unterteilung in drei Unterregionen. In Region 1 wird Fourier-Rebinning, in Region 2 die Konsistenzbedingung und in Region 3 SSRB angewendet. Quelle: [14] S. 152.

Region 1: In dieser Hochfrequenzregion werden die rebinnten Daten mit Fourier-Rebinning geschätzt.
Für jedes δ liefert

$$\hat{p}_2(\omega, k, z, \delta) \approx \hat{p}_{reb}(\omega, k, z - \frac{k\delta}{\omega} := z_0)$$

bzw.

$$\hat{p}_{reb}(\omega, k, z_0) \approx \hat{p}_2(\omega, k, z_0 + \frac{k\delta}{\omega}, \delta) \quad (4.65)$$

eine unabhängige Schätzung von $\hat{p}_{reb}(\omega, k, z_0 = z - \frac{k\delta}{\omega})$.

Läge kein Rauschen vor, dann wären alle diese Schätzungen gleich (bis auf die Genauigkeit der Fourier-Rebinning-Approximation). Da die Daten jedoch verrauscht sind, mitteln wir wieder über alle Schätzungen, um die SNR zu verbessern. Die rebinnten Daten werden also wie folgt berechnet:

$$\hat{p}_{reb}(\omega, k, z_0) = \frac{1}{\delta_1(\omega, k, z_0)} \int_0^{\delta_1} \hat{p}_2(\omega, k, z_0 + \frac{k\delta}{\omega}, \delta) d\delta, \quad (4.66)$$

für $|\frac{k}{\omega}| < R$ und ($|k| > k_{lim}$ oder $|\omega| > \omega_{lim}$). Es gilt weiter (siehe Gleichung (4.44)):

$$-\frac{L_D}{2} + |\delta| R_D \stackrel{1)}{\leq} z = z_0 + \frac{k\delta}{\omega} \stackrel{2)}{\leq} \frac{L_D}{2} - |\delta| R_D$$

und somit folgt:

$$\begin{aligned}
 1) \quad & -\frac{L_D}{2}\omega - z_0\omega \leq k\delta - |\delta| R_D\omega = \delta \underbrace{(k - R_D\omega)}_{<0} \\
 & \Rightarrow \delta \leq \frac{\left(\frac{L_D}{2} + z_0\right)\omega}{R_D\omega - k}, \\
 2) \quad & z_0\omega + k\delta \leq \frac{L_D}{2}\omega - |\delta| R_D\omega \\
 & \Rightarrow \delta \underbrace{(k + R_D\omega)}_{>0} \leq \left(\frac{L_D}{2} - z_0\right)\omega \\
 & \Rightarrow \delta \leq \frac{\left(\frac{L_D}{2} - z_0\right)\omega}{k + R_D\omega}.
 \end{aligned}$$

Damit erhalten wir für δ_1 :

$$\delta_1(\omega, k, z_0) = \min \left\{ \frac{\frac{L_D}{2} - z_0}{R_D + \frac{k}{\omega}}, \frac{\frac{L_D}{2} + z_0}{R_D - \frac{k}{\omega}}, \delta_{max} \right\}, \quad (4.67)$$

wobei R_D der Radius des Detektors und δ_{max} die größte Ringdifferenz ($\delta_{max} \leq \frac{L_D}{2R_D}$) ist.

Es erscheinen auch nur positive Werte von δ in Gleichung (4.66), da das Sinogramm mit negativem δ bei der Berechnung der 2D FT, (siehe Lemma 4.12 a), miteinbezogen wurde.

Region 2: In dieser Hochfrequenzregion ist die Konsistenzbedingung nicht erfüllt und daher gilt:

$$\hat{p}_{reb}(\omega, k, z) = 0 \quad \text{für } \left| \frac{k}{\omega} \right| \geq R \text{ und } (|k| > k_{lim} \text{ oder } |\omega| > \omega_{lim}).$$

Region 3: In der Niedrigfrequenzregion ist Fourier-Rebinning nicht anwendbar. Deshalb werden wir in dieser Region die Rebinningdaten nur unter Zuhilfenahme der schrägen Sinogramme mit einem kleinen δ -Wert, $\delta \leq \delta_{lim}$, geschätzt.

Wegen der kleinen Werte von δ kann der axiale Shift vernachlässigt werden und daher gilt:

$$\hat{p}_{reb}(\omega, k, z) = \frac{1}{\delta_2(z)} \int_0^{\delta_2} \hat{p}_2(\omega, k, z, \delta) d\delta \quad \text{für } |k| \leq k_{lim} \text{ und } |\omega| \leq \omega_{lim}, \quad (4.68)$$

wobei gilt:

$$|z| \leq \frac{L_D}{2} - |\delta| R_D \quad \Rightarrow \quad |\delta| \leq \frac{\left(\frac{L_D}{2} - |z|\right)}{R_D}$$

$$\Rightarrow \delta_2(z) = \min \left\{ \frac{\frac{L_D}{2} - |z|}{R_D}, \delta_{lim} \right\}. \quad (4.69)$$

Gleichung (4.68) beruht auf der **SSRB-Approximation** (SSRB=Single-Slice-Rebinning). Setzen wir in Gleichung (4.61) $\alpha = 0$, so erhalten wir:

$$\hat{p}_3(\omega, k, \omega_z, \delta) = \hat{p}_{reb}(\omega, k, \omega_z) + O(\alpha)$$

und wenn wir die 3D inverse FT in Hinsicht auf ω , k und ω_z berechnen folgt:

$$p(x_r, \phi, z, \delta) \approx p_{reb}(x_r, \phi, z). \quad (4.70)$$

und diese Gleichung (4.70) heisst nun **SSRB-Approximation**.

Wir müssen δ_{lim} klein genug wählen, um die systematischen Fehler, die auf den Gebrauch der SSRB-Approximation zurückzuführen sind, zu minimieren. Üblicherweise wählt man dasselbe δ_{lim} wie dieses, welches die Datenteilmenge definiert, die gewöhnlich zur Schätzung der abgeschnittenen Daten im 3DRP-Algorithmus benutzt wird.

Der Algorithmus hängt also von den drei Parametern ω_{lim} , k_{lim} und δ_{lim} ab. Diese Parameter müssen so gewählt werden, dass sie einen guten Kompromiss zwischen dem systematischem Fehler (Genauigkeit) und dem statistischer Fehler (Rauschen) liefern. Die Möglichkeit eine Menge von Parametern zu finden, die eine Bildqualität liefern, die vergleichbar mit der ist, die beim 3DRP-Algorithmus erzielt wurde, hängt von den zwei folgenden Beobachtungen ab, die auf einer großen Anzahl von Simulationen basieren:

1. Obwohl die Fourier-Rebinning-Approximation im Prinzip eine asymptotische Beziehung ist, die nur bei hohen Frequenzen gültig ist, stellte es sich heraus, dass diese Approximation sogar überraschenderweise gute Ergebnisse bei sehr niedrigen Frequenzen für gewisse PET-Scanner-Geometrien hervorbringt.
2. Die Tatsache, dass die Niedrigfrequenzkomponenten in Region 3 unter Verwendung einer kleinen Teilmenge aller verfügbarer Daten ($\delta \leq \delta_{lim}$) berechnet werden hat nur einen kleinen Einfluß auf die SNR. Dies liegt daran, dass Rauschen in einer tomographischen Rekonstruktion hauptsächlich aus Hochfrequenzverstärkung durch den Ramp-Filter entsteht.

Deshalb sollten kleine Werte von ω_{lim} und k_{lim} gewählt werden. Mit Scannern wie dem ECAT EXACT HR wurden gute Rekonstruktionen mit $k_{lim} = 2$ und $\omega_{lim} = 2 \Delta \omega$ ($\Delta \omega =$ Frequenzsampling) erzielt. Jedoch gibt es keine systematische Optimierung der drei Parameter ω_{lim} , k_{lim} und δ_{lim} und allgemeiner gesprochen, der Aufteilung der (ω, k) -Ebene.

Algorithmus

Eine Diskretisierung des δ -Integrals (4.66) mit einer Standardquadratur würde dazu führen, dass nicht jedes schräge Sinogramm unabhängig bearbeitet werden kann. Deshalb folgen wir bei der Implementierung von FORE dem unten angegebenen Ansatz:

1. Initialisiere eine Menge rebinnter, 2D, fouriertransformierter Sinogramme:

$$\hat{p}_{reb}(\omega, k, z_0) = 0.$$

2. Betrachte der Reihe nach jedes Paar schräger Sinogramme (z, δ) und $(z, -\delta)$:

- a) Wenn nötig, fülle in x_r auf und interpoliere in ϕ , um geeignete Dimensionen N'_{x_r} und N'_ϕ zu erhalten (Erklärung dazu siehe unten).
 b) Berechne die 2D FT (bzw. 2D FFT) bzgl. x_r und ϕ :

$$\begin{aligned} \hat{p}_2(\omega, k, z, \delta) &\stackrel{\text{Lemma 4.12 a)}}{=} (2\pi)^{-1} \int_0^\pi e^{-ik\phi} \int_{-R}^R e^{-i\omega x_r} p(x_r, \phi, z, \delta) \\ &\quad + (-1)^k e^{i\omega x_r} p(x_r, \phi, z, -\delta) dx_r d\phi. \end{aligned}$$

- c) Für jedes Paar (ω, k) in Region 1 berechne $z = z_0 + \frac{k\delta}{\omega}$ und man benutze lineare Interpolation zwischen den 2 am nächsten gelegenen Ebenen, um den Wert von $\hat{p}_{reb}(\omega, k, z_0)$ zu erhöhen:
 Für $z_1 < z \leq z_2$ folgt somit:

$$\hat{p}_{reb}(\omega, k, z_0) = \hat{p}_{reb}(\omega, k, z_0) + \frac{\hat{p}_2(\omega, k, z_2, \delta)(z - z_1) + \hat{p}_2(\omega, k, z_1, \delta)(z_2 - z)}{z_2 - z_1}.$$

- d) Wenn $\delta \leq \delta_{lim}$, dann betrachtet man jedes Paar (ω, k) in Region 3 und berechnet:

$$\hat{p}_{reb}(\omega, k, z_0) = \hat{p}_{reb}(\omega, k, z_0) + \hat{p}_2(\omega, k, z_0, \delta).$$

3. Normiere die rebinnten Daten \hat{p}_{reb} . Der Normierungsfaktor wird bestimmt, indem wir 2c) - 2d) auf die Einheitsdaten $\hat{p}_2(\omega, k, z, \delta) = 1$ anwenden.
 4. Berechne die inverse 2D FT (bzw. FFT) von $\hat{p}_{reb}(\omega, k, z_0)$ um eine Menge rebinnter Sinogramme $p_{reb}(x_r, \phi, z_0)$ zu erhalten:

$$p_{reb}(x_r, \phi, z_0) = (2\pi)^{-1} \int_{\mathbb{R}^2} \hat{p}_{reb}(\omega, k, z_0) e^{i\omega x_r} e^{ik\phi} dk d\omega.$$

5. Rekonstruiere jede Ebene separat und unabhängig voneinander unter Verwendung irgendeines beliebigen 2D Rekonstruktionsalgorithmus.
 6. Stapele die rekonstruierten Ebenen, um ein Volumenbild zu erhalten.

Erklärung zu 2 a):

Die diskrete 2D FT $\hat{p}_2(\omega, k, z, \delta)$ wird durch Gruppierung von Paaren schräger Sinogramme mit entgegengesetzten δ -Werten, Lemma 4.12 a), berechnet. Jede schräge Sinogramm-Matrix hat N_{x_r} radiale Werte und N_ϕ Winkelwerte mit

$$N_\phi = \frac{N_R}{2^{m+1}},$$

wobei N_R die Anzahl der Detektorringe und 2^m der Mashing-Faktor ist (Integerwert, durch den die Samplingrate in ϕ -Richtung reduziert wird).

Der FFT-Algorithmus ist nur dann effizient, wenn die Dimensionen N_{x_r} und N_ϕ Zweierpotenzen entsprechen oder zumindest einem Produkt einer Zweierpotenz mit einem kleinen Faktor (≤ 10) entsprechen.

Wenn N_{x_r} dieser Einschränkung nicht genügt, dann kann die Sinogramm-Matrix mit Nullen aufgefüllt werden, um ein größeres Feld mit einer Dimension N'_{x_r} zu erhalten, die einer Zweierpotenz am nächsten an N_{x_r} entspricht.

Leider ist das Auffüllen von Nullen ("Zero Padding") nicht in der ϕ -Richtung möglich, da das Sinogramm periodisch in ϕ ist und die Transformation eine Fourier-Reihe ist. Um dieses Problem zu lösen, interpolieren wir linear in ϕ , um ein Sinogramm mit einer Dimension N'_ϕ zu erzielen, die gleich der am nächstgelegenen Zweierpotenz größer als N_ϕ ist.

4.6.3 FORE-J

Im Folgenden halten wir uns an [16].

In diesem Abschnitt leiten wir den exakten Rebinning-Algorithmus FORE-J unter Zuhilfenahme der Gleichung von John her.

Wir betrachten dieselbe zylindrische Geometrie und dieselbe Darstellung des Linienintegrals wie in den beiden Abschnitten zuvor. Ebenso ist $f \in \mathcal{S}(\mathbb{R}^3)$, $\text{supp } f \subset \text{Zylinder}$ mit Radius R und $g \in \mathcal{S}(T^4)$. Dabei gilt auch hier:

$$T^4 = \{(x_r, \phi, z, \delta); |x_r| < R, 0 \leq \phi < \pi, |z| \leq \frac{L_D}{2} - |\delta| R_D, |\delta| \leq \frac{L_D}{2R_D}\}$$

und des Weiteren wurde g für alle $(x_r, y_r, \phi, \delta) \in T^4$ gemessen und kann somit als bekannt vorausgesetzt werden.

Wir haben also wiederum die Darstellung der 3D Röntgentransformation in der Form:

$$p(x_r, \phi, z, \delta) = \int_{\mathbb{R}} f(x_r \cos \phi - t \sin \phi, x_r \sin \phi + t \cos \phi, z + t\delta) dt. \quad (4.71)$$

Die 2D Radontransformation in Schichten senkrecht zur z -Achse impliziert auch hier $\delta = 0$ und es gilt:

$$p_{reb}(x_r, \phi, z) := p(x_r, \phi, z, 0), \quad (4.72)$$

wobei hier $|x_r| < R, 0 \leq \phi < \pi$ und $-\frac{L_D}{2} \leq z \leq \frac{L_D}{2}$ ist.

Satz 4.23. (Konsistenzbedingungen)

Für die oben definierten Daten gilt:

$$\begin{aligned}
 a) \quad & p_{reb}(x_r, \phi, z) = p_{reb}(-x_r, \phi + \pi, z) \quad \forall 0 \leq \phi < \pi \\
 & \Leftrightarrow e^{ik\pi} \hat{p}_{reb}(\omega, k, z) = \hat{p}_{reb}(-\omega, k, z) \quad \forall k \in \mathbb{Z}, \omega \in \mathbb{R}.
 \end{aligned}$$

b) Da das x_r -Integral von $p_{reb}(x_r, \phi, z)$ unabhängig von ϕ ist gilt:

$$\hat{p}_{reb}(\omega = 0, k, z) = (2\pi)^{-1} \int_0^{2\pi} \int_{\mathbb{R}} p_{reb}(x_r, \phi, z) e^{i \cdot 0 \cdot x_r} dx_r e^{-ik\phi} d\phi = 0 \text{ für } k \neq 0.$$

Beweis. a) " \Rightarrow " Sei $p_{reb}(x_r, \phi, z) = p_{reb}(-x_r, \phi + \pi, z)$, dann folgt:

$$e^{ik\pi} \hat{p}_{reb}(\omega, k, z) = (2\pi)^{-1} \int_{\mathbb{R}} \int_0^{2\pi} p_{reb}(x_r, \phi, z) e^{-ik(\phi-\pi)} d\phi e^{-i\omega x_r} dx_r.$$

Substituieren wir $x_r = -\tilde{x}_r$ und $\phi = \tilde{\phi} + \pi$ so erhalten wir mit Hilfe der Voraussetzung

$$e^{ik\pi} \hat{p}_{reb}(\omega, k, z) = (2\pi)^{-1} \int_{\mathbb{R}} \int_{-\pi}^{\pi} p_{reb}(x_r, \phi, z) e^{-ik\phi} e^{i\omega x_r} d\phi dx_r,$$

und wenn wir nun noch einmal $x_r = -\tilde{x}_r$ setzen so folgt:

$$\begin{aligned}
 e^{ik\pi} \hat{p}_{reb}(\omega, k, z) &= (2\pi)^{-1} \int_{\mathbb{R}} \int_{-\pi}^{\pi} p_{reb}(-x_r, \phi, z) e^{-ik\phi} e^{-i(-\omega)(-x_r)} d\phi dx_r \\
 &= \hat{p}_{reb}(-\omega, k, z).
 \end{aligned}$$

" \Leftarrow " Es sei $e^{ik\pi} \hat{p}_{reb}(\omega, k, z) = \hat{p}_{reb}(-\omega, k, z)$. Das ist äquivalent mit:

$$\begin{aligned}
 & (2\pi)^{-1} \int_{\mathbb{R}} \int_0^{2\pi} e^{ik\pi} \hat{p}_{reb}(\omega, k, z) e^{ik\phi} dk e^{i\omega x_r} d\omega \\
 &= (2\pi)^{-1} \int_{\mathbb{R}} \int_0^{2\pi} \hat{p}_{reb}(-\omega, k, z) e^{ik\phi} dk e^{i(-\omega)(-x_r)} d\omega \\
 \Leftrightarrow & \quad p_{reb}(x_r, \phi + \pi, z) = p_{reb}(-x_r, \phi, z).
 \end{aligned}$$

b)

$$\hat{p}_{reb}(\omega = 0, k, z) = (2\pi)^{-1} \int_0^{2\pi} \underbrace{\int_{\mathbb{R}} p_{reb}(x_r, \phi, z) e^{-i \cdot 0 \cdot x_r} dx_r}_{= \text{konstant}} e^{-ik\phi} d\phi = 0 \quad k \neq 0$$

Der konstante Term folgt aus den Konsistenzbedingungen von Helgason-Ludwig (Satz 2.28) und da die e -Funktion 2π -periodisch ist, wird das Integral insgesamt Null. \square

Satz und Definition 4.24. (John's Gleichung)

Sei p die 3D Röntgentransformation wie in (4.71), dann gilt:
 p genügt der partiellen Differentialgleichung 2. Ordnung:

$$\frac{\partial^2 p(x_r, \phi, z, \delta)}{\partial z \partial \phi} + \frac{\partial^2 p(x_r, \phi, z, \delta)}{\partial \delta \partial x_r} - x_r \delta \frac{\partial^2 p(x_r, \phi, z, \delta)}{\partial z^2} = 0.$$

Diese Gleichung heißt **John's Gleichung**.

Beweis. Da wir Differentiation und Integration vertauschen können folgt:

$$\begin{aligned} & \left(\frac{\partial^2}{\partial z \partial \phi} + \frac{\partial^2}{\partial x_r \partial \delta} + x_r \delta \frac{\partial^2}{\partial z^2} \right) \int_{\mathbb{R}} f(x_r \cos \phi - t \sin \phi, x_r \sin \phi + t \cos \phi, z + t\delta) dt \\ = & \int_{\mathbb{R}} \left[(-x_r \sin \phi - t \cos \phi) \frac{\partial^2 f}{\partial z \partial x} + (x_r \cos \phi - t \sin \phi) \frac{\partial^2 f}{\partial z \partial y} \right. \\ & \left. + t \cos \phi \frac{\partial^2 f}{\partial x \partial z} + t \sin \phi \frac{\partial^2 f}{\partial y \partial z} + x_r \delta \frac{\partial^2 f}{\partial z^2} \right] dt \\ = & \int_{\mathbb{R}} \left[-x_r \sin \phi \frac{\partial^2 f}{\partial z \partial x} + x_r \cos \phi \frac{\partial^2 f}{\partial z \partial y} + x_r \delta \frac{\partial^2 f}{\partial z^2} \right] dt \\ = & x_r \int_{\mathbb{R}} \frac{d}{dt} \frac{\partial f}{\partial z} dt \\ \stackrel{f \in \mathcal{S}}{=} & 0. \end{aligned}$$

\square

Korollar 4.25. (Basisgleichung für FORE-J, FOREX und FORE)

Es seien dieselben Voraussetzungen wie in Satz 4.24 gegeben. Dann gilt:

$$k \frac{\partial \hat{p}_2(\omega, k, z, \delta)}{\partial z} + \omega \frac{\partial \hat{p}_2(\omega, k, z, \delta)}{\partial \delta} = -\delta \frac{\partial^3 \hat{p}_2(\omega, k, z, \delta)}{\partial \omega \partial z^2}. \quad (4.73)$$

Insbesondere ist für eine Schicht z_0 und für ein $\omega \neq 0$:

$$\frac{\partial \hat{p}_2(\omega, k, z_0 + \frac{k\delta}{\omega}, \delta)}{\partial \delta} = -\frac{\delta}{\omega} \frac{\partial^3 \hat{p}_2(\omega, k, z_0 + \frac{k\delta}{\omega}, \delta)}{\partial \omega \partial z^2}. \quad (4.74)$$

Beweis. Es gelte John's Gleichung. Bilden wir nun die FT bzgl. der ersten beiden Argumente, vernachlässigen die multiplikative Konstante $(2\pi)^{-1}$, Nutzen die Aditivität des Integrals aus und ziehen die unabhängige Ableitung heraus, so ist John's Gleichung äquivalent zu:

$$\begin{aligned} & \frac{\partial}{\partial z} \int_{\mathbb{R}} \int_0^{2\pi} \underbrace{\frac{\partial p(x_r, \phi, z, \delta)}{\partial \phi} e^{-ik\phi} d\phi}_{\text{Lemma 2.19 b)}_{ik\hat{p}_1(x_r, k, z, \delta)}} e^{-i\omega x_r} dx_r + \frac{\partial}{\partial \delta} \int_0^{2\pi} \int_{\mathbb{R}} \underbrace{\frac{\partial p(x_r, \phi, z, \delta)}{\partial x_r} e^{-i\omega x_r} dx_r}_{\text{Lemma 2.19 b)}_{i\omega\hat{p}_1(\omega, \phi, z, \delta)}} e^{-ik\phi} d\phi \\ &= -\delta \frac{\partial}{\partial z^2} \int_0^{2\pi} \int_{\mathbb{R}} \underbrace{x_r p(x_r, \phi, z, \delta) e^{-i\omega x_r} dx_r}_{\text{Lemma 2.19 a)}_{i\frac{\partial}{\partial \omega}\hat{p}_1(\omega, \phi, z, \delta)}} e^{-ik\phi} d\phi. \end{aligned}$$

Damit erhalten wir nun folgende äquivalente Gleichung:

$$ik \frac{\partial}{\partial z} \hat{p}_2(\omega, k, z, \delta) + i\omega \frac{\partial}{\partial \delta} \hat{p}_2(\omega, k, z, \delta) = -i\delta \frac{\partial^3 \hat{p}_2(\omega, k, z, \delta)}{\partial z^2 \partial \omega}$$

und teilen durch i liefert uns die Gleichung (4.73).

Gleichung (4.74) folgt direkt aus Gleichung (4.73). □

Aus den beiden Gleichungen aus Korollar 4.25 lassen sich nun die drei Rebinningmethoden FORE, FOREX und FORE-J ableiten.

Integration der Gleichung (4.74) von 0 bis δ_1 liefert uns exaktes Rebinning mit in z -Richtung abgeschnittenen Daten:

Satz 4.26. (Rebinninggleichung)

Es seien die Voraussetzungen wie in Satz 4.24. Dann gilt:

$$\hat{p}_{reb}(\omega, k, z_0) = \hat{p}_2(\omega, k, z_0 + \frac{k}{\omega} \delta_1, \delta_1) + \int_0^{\delta_1} \frac{\delta}{\omega} \frac{\partial^3 \hat{p}_2(\omega, k, z_0 + \frac{k}{\omega} \delta, \delta)}{\partial \omega \partial z^2} d\delta \quad (4.75)$$

für alle $\delta_1 \in [0, \delta_{max}(\omega, k, z_0)]$ mit:

$$\delta_{max}(\omega, k, z_0) = \begin{cases} \frac{\frac{L_D}{2} - z_0}{k} & ; \quad \frac{k}{\omega} \geq \frac{-2R_D z_0}{L_D} \\ R_D + \frac{1}{\omega} & \\ \frac{\frac{L_D}{2} + z_0}{k} & ; \quad \frac{k}{\omega} < \frac{-2R_D z_0}{L_D} \\ R_D - \frac{1}{\omega} & \end{cases} .$$

Spezialfall: $\hat{p}_{reb}(0, k, z_0) = \hat{p}_2(0, k, z_0, 0)$.

Beweis. • Wenn wir Gleichung (4.74) auf beiden Seiten von 0 bis δ_1 integrieren, so erhalten wir folgende Äquivalenz:

$$\int_0^{\delta_1} \frac{\partial \hat{p}_2(\omega, k, z_0 + \frac{k\delta}{\omega}, \delta)}{\partial \delta} d\delta = \int_0^{\delta_1} -\frac{\delta}{\omega} \frac{\partial^3 \hat{p}_2(\omega, k, z_0 + \frac{k\delta}{\omega}, \delta)}{\partial \omega \partial z^2} d\delta$$

$$\Leftrightarrow \hat{p}_2(\omega, k, z_0 + \frac{k\delta_1}{\omega}, \delta_1) - \underbrace{\hat{p}_2(\omega, k, z_0, 0)}_{=\hat{p}_{reb}(\omega, k, z_0)} = - \int_0^{\delta_1} \frac{\delta}{\omega} \frac{\partial^3 \hat{p}_2(\omega, k, z_0 + \frac{k\delta}{\omega}, \delta)}{\partial \omega \partial z^2} d\delta$$

und daraus folgt die Behauptung.

- δ_{max} wird dadurch bestimmt, dass alle Frequenzkomponenten im Definitionsbereich sind, der durch

$$0 \leq \delta \leq \frac{L_D}{2R_D} \text{ und } |z| \leq \frac{L_D}{2} - |\delta| R_D$$

bestimmt wird.

(Wir haben dieselbe Herleitung wie auf Seite 115 für:

$$z = z_0 + \frac{k\delta}{\omega} \geq 0$$

$$\Leftrightarrow 0 \leq z_0 + \frac{k\delta}{\omega} \leq z_0 + \frac{kL_D}{2R_D\omega}$$

$$\Leftrightarrow \frac{-z_0 2R_D}{L_D} \leq \frac{k}{\omega},$$

und für:

$$z = z_0 + \frac{k\delta}{\omega} < 0$$

$$\Leftrightarrow \frac{k}{\omega} < \frac{-z_0 2R_D}{L_D}).$$

- Die Nullfrequenzkomponenten der Rebinningdaten kann nur von der $\delta = 0$ Teilmenge bestimmt werden, denn für

$$\omega \rightarrow 0 \Rightarrow \delta_{max} \rightarrow 0 \Rightarrow \int_0^{\delta_1} \frac{\delta}{\omega} \frac{\partial^3 \hat{p}_2(\omega, k, z_0 + \frac{k\delta}{\omega}, \delta)}{\partial \omega \partial z^2} d\delta \rightarrow 0.$$

□

Wie immer erhält man eine Verbesserung der SNR, wenn man mittelt:

Satz 4.27. (Verbessertes Rebinning)

Es seien die Voraussetzung wie in Satz 4.24, dann gilt:

$$\hat{p}_{reb}(\omega, k, z_0) = \begin{cases} \frac{1}{\delta_{max}} \left[\int_0^{\delta_{max}} \hat{p}_2(\omega, k, z_0 + \frac{k\delta}{\omega}, \delta) d\delta \right. \\ \left. + \int_0^{\delta_{max}} (\delta_{max} - \delta) \frac{\delta}{\omega} \frac{\partial^3 \hat{p}_2(\omega, k, z_0 + \frac{k\delta}{\omega}, \delta)}{\partial \omega \partial z^2} d\delta \right] & ; \quad \omega \neq 0 \\ \hat{p}_2(0, 0, z_0, 0) & ; \quad \omega = k = 0 \\ 0 \text{ (Gleichung (4.64))} & ; \quad \omega = 0, k \neq 0. \end{cases}$$

Beweis. Durch Mittelung und anschließende partielle Integration erhält man:

$$\begin{aligned} \hat{p}_{reb}(\omega, k, z_0) &\stackrel{\text{Mittelung}}{=} \underbrace{\frac{1}{\delta_{max}} \int_0^{\delta_{max}} \hat{p}_2(\omega, k, z_0 + \frac{k\delta_1}{\omega}, \delta_1) d\delta_1}_{=:A} \\ &= \frac{1}{\delta_{max}} \int_0^{\delta_{max}} \underbrace{1}_{=:u'} \cdot \underbrace{\int_0^{\delta_1} \frac{\delta}{\omega} \frac{\partial^3 \hat{p}_2(\omega, k, z_0 + \frac{k\delta}{\omega}, \delta)}{\partial \omega \partial z^2} d\delta}_{=:v(\delta_1)} d\delta_1 \\ &\stackrel{\text{part. Integr.}}{=} A + \frac{1}{\delta_{max}} \left([\delta_1 v(\delta_1)]_0^{\delta_{max}} - \int_0^{\delta_{max}} \delta_1 \frac{\delta_1}{\omega} \frac{\partial^3 \hat{p}_2(\omega, k, z_0 + \frac{k\delta_1}{\omega}, \delta_1)}{\partial \omega \partial z^2} d\delta_1 \right) \\ &= A + \frac{1}{\delta_{max}} \left(\delta_{max} v(\delta_{max}) - 0 - \int_0^{\delta_{max}} \frac{\delta^2}{\omega} \frac{\partial^3 \hat{p}_2(\omega, k, z_0 + \frac{k\delta}{\omega}, \delta)}{\partial \omega \partial z^2} d\delta \right) \\ &= A + \frac{1}{\delta_{max}} \int_0^{\delta_{max}} \frac{(\delta_{max} - \delta) \delta}{\omega} \frac{\partial^3 \hat{p}_2(\omega, k, z_0 + \frac{k\delta}{\omega}, \delta)}{\partial \omega \partial z^2} d\delta. \end{aligned}$$

□

Der FORE-J Algorithmus kann direkt auf die abgeschnittenen Daten angewendet werden (ohne sie vorher zu schätzen), da keine FT in z -Richtung gemacht wird und der Wert von δ_{max} sicherstellt, dass nur wirklich gemessene Daten benutzt werden.

Algorithmus

Die Implementierung des FORE-J Algorithmus folgt der Implementierung des FORE Algorithmus. Der einzige Unterschied besteht in der Addition des zusätzlichen Termes Q . Ansonsten haben wir dieselbe Einteilung der Frequenzkomponenten ω und k und betrachten dieselben Regionen 1, 2 und 3 wie beim FORE-Algorithmus. Normalerweise setzen wir auch hier $k_{lim} = 2$, $\omega_{lim} = 2 \Delta \omega$ und $\delta_{lim} = 2 \Delta z / R_D$, wobei $\Delta \omega$ das radiale Frequenzsampling ist.

1. Initialisiere eine Menge rebinnter, 2D, fouriertransformierter Sinogramme:

$$\hat{p}_{reb}(\omega, k, z_0) = 0.$$

2. Betrachte der Reihe nach jedes Paar schräger Sinogramme (z, δ) und $(z, -\delta)$ und vereinige sie, um ein 2π -Sinogramm $p(x_r, \phi, z, \delta)$ zu bilden.:

- a) Wenn nötig, fülle in x_r auf und interpoliere in ϕ , um Matrixdimensionen zu erhalten, die einer Zweierpotenz entsprechen.
 b) Berechne die 2D FT (bzw. 2D FFT) bzgl. x_r und ϕ :

$$\begin{aligned} \hat{p}_2(\omega, k, z, \delta) &\stackrel{\text{Lemma 4.12 a)}}{=} (2\pi)^{-1} \int_0^\pi e^{-ik\phi} \int_{-R}^R e^{-i\omega x_r} p(x_r, \phi, z, \delta) \\ &\quad + (-1)^k e^{i\omega x_r} p(x_r, \phi, z, -\delta) dx_r d\phi. \end{aligned}$$

- c) Berechne den Term $Q(\omega, k, z, \delta)$ wie folgt:

$$\begin{aligned} Q(\omega, k, z, \delta) &= -i \frac{\delta}{\omega} \left(\delta_{max}(\omega, k, z - \underbrace{\frac{k\delta}{\omega}}_{=:z_0}) - \delta \right) \\ &\quad (2\pi)^{-1} \int_{\mathbb{R}} \int_0^{2\pi} e^{-i\omega x_r} e^{-ik\phi} x_r \frac{\partial^2 p(x_r, \phi, z, \delta)}{\partial z^2} d\phi dx_r \end{aligned}$$

mit δ_{max} wie in Satz 4.26 und approximiere $\frac{\partial^2 p(x_r, \phi, z, \delta)}{\partial z^2}$ durch:

$$\frac{\partial^2 p(x_r, \phi, z, \delta)}{\partial z^2} \approx \frac{p(x_r, \phi, z + \Delta z, \delta) + p(x_r, \phi, z - \Delta z, \delta) - 2p(x_r, \phi, z, \delta)}{\Delta z^2},$$

wobei Δz dem axialen Sinogramm-Sampling entspricht.

- d) Für jedes Paar (ω, k) in Region 1 berechne $z = z_0 + \frac{k\delta}{\omega}$ und man benutze lineare Interpolation zwischen den 2 am nächsten gelegenen Ebenen, um den Wert von $\hat{p}_{reb}(\omega, k, z_0)$ zu erhöhen:

Für $z_1 < z \leq z_2$ folgt somit:

$$\begin{aligned} \hat{p}_{reb}(\omega, k, z_0) &= \hat{p}_{reb}(\omega, k, z_0) + \{[\hat{p}_2(\omega, k, z_2, \delta) + Q(\omega, k, z_2, \delta)](z - z_1) \\ &\quad + [\hat{p}_2(\omega, k, z_1, \delta) + Q(\omega, k, z_1, \delta)](z_2 - z)\} / (z_2 - z_1). \end{aligned}$$

- e) Wenn $\delta \leq \delta_{lim}$, dann betrachtet man jedes Paar (ω, k) in Region 3 und berechnet:

$$\hat{p}_{reb}(\omega, k, z_0) = \hat{p}_{reb}(\omega, k, z_0) + \hat{p}_2(\omega, k, z_0, \delta).$$

3. Normiere die rebinnten Daten. Der Normierungsfaktor ist der Faktor $\frac{1}{\delta_{max}(\omega, k, z_0)}$ aus Satz 4.26.

Da eine diskrete Implementierung jedoch auch die Effekte der Interpolation in Schritt 2 d) berücksichtigen muss, benutzt man lieber einen genaueren Normierungsfaktor, der sich durch die Anwendung der Rebinningprozedur (Schritt 2 d)-2 e)) auf die Einheitsdaten $\hat{p}_2(\omega, k, z, \delta)$ und $Q(\omega, k, z, \delta) = 0$ ergibt.

4. Berechne die inverse 2D FT (bzw. 2D FFT) von $\hat{p}_{reb}(\omega, k, z_0)$ um für jede transaxiale Ebene z_0 ein rebinntes Sinogramm $p_{reb}(x_r, \phi, z_0)$ zu erhalten:

$$p_{reb}(x_r, \phi, z_0) = (2\pi)^{-1} \int_{\mathbb{R}^2} \hat{p}_{reb}(\omega, k, z_0) e^{i\omega x_r} e^{ik\phi} dk d\omega.$$

5. Rekonstruiere jede Ebene separat und unabhängig voneinander unter Verwendung irgendeines beliebigen 2D Rekonstruktionsalgorithmus.
6. Stapele die rekonstruierten Ebenen, um ein Volumenbild zu erhalten.

4.6.4 Zusammenhang zwischen FORE, FOREX und FORE-J

Für alle drei Algorithmen kann Gleichung (4.73) bzw. (4.74) als Ausgangsgleichung angesehen werden:

- FORE-J:
- Integration der Gleichung (4.74) von 0 bis δ_1 über δ
 - Mittelung
- ⇒ FORE-J

- FORE:
- Wähle δ klein genug
- ⇒ Rechte Seite von Gleichung (4.74) wird durch Null approximiert
- Integration dieser neuen Gleichung (rechte Seite ist Null) von 0 bis δ_1 über δ und daraus folgt:

$$\hat{p}_{reb}(\omega, k, z_0) \approx \hat{p}_2(\omega, k, z_0 + \frac{k\delta_1}{\omega}, \delta_1) \quad \omega \neq 0$$

- Mittelung:

$$\hat{p}_{reb}(\omega, k, z_0) \approx \frac{1}{\delta_{max}} \int_0^{\delta_{max}} \hat{p}_2(\omega, k, z_0 + \frac{k\delta}{\omega}, \delta) d\delta$$

- FOREX:
- Annahme, dass die Daten komplett sind oder durch Vorwärtsprojektion bekannt sind

- FT von Gleichung (4.73) bzgl. z (beachte Lemma 2.19 b):

$$\Rightarrow -k\omega_z \hat{p}_3(\omega, k, \omega_z, \delta) + \underbrace{i\omega \frac{\partial \hat{p}_3(\omega, k, \omega_z, \delta)}{\partial \delta}}_{\text{Hauptteil der Dgl.}} = i\omega_z^2 \delta \frac{\partial \hat{p}_3(\omega, k, \omega_z, \delta)}{\partial \omega} \quad (4.76)$$

- Lösen der Differentialgleichung:

$$\begin{aligned} \hat{p}_3(\omega, k, \omega_z, \delta) &= e^{-ik \arctan\left(\frac{\delta\omega_z}{\omega}\right)} U \left(\overbrace{\sqrt{\omega^2 + \delta^2\omega_z^2}}{=\omega_0}, k, \omega_z \right) \quad \omega > 0 \\ &= e^{-ik \arctan\left(\frac{\delta\omega_z}{\omega}\right)} \hat{p}_{reb}\left(\omega \sqrt{1 + \frac{\delta^2\omega_z^2}{\omega^2}}, k, \omega_z\right), \end{aligned}$$

und

$$\begin{aligned} \hat{p}_3(\omega, k, \omega_z, \delta) &= e^{ik\pi - ik \arctan\left(\frac{\delta\omega_z}{\omega}\right)} U \left(\overbrace{\sqrt{\omega^2 + \delta^2\omega_z^2}}{=\omega_0}, k, \omega_z \right) \quad \omega < 0 \\ &= e^{-ik \arctan\left(\frac{\delta\omega_z}{\omega}\right)} \hat{p}_{reb}\left(\omega \sqrt{1 + \frac{\delta^2\omega_z^2}{\omega^2}}, k, \omega_z\right), \end{aligned}$$

wobei das Lösen der Differentialgleichung und U im folgenden Lemma 4.28, Satz 4.29 und Korollar 4.30 erklärt werden.

- Mittelung
 \Rightarrow FOREX

Um also die obige Differentialgleichung zu lösen, benötigen wir noch ein paar Zusatzinformationen:

Für jedes feste Paar $(k, \omega_z \neq 0)$ sind die charakteristischen Kurven dieser partiellen Differentialgleichung 1. Ordnung (4.76) konzentrische Ellipsen $\omega^2 + \delta^2\omega_z^2 = \omega_0^2$ in der (ω, k) Ebene und zwar für alle $\omega_0 \in (0, \infty)$. Wenn wir einen festen Wert ω_0 betrachten und die Position entlang der entsprechenden Ellipse parametrisieren, so können wir die partielle Differentialgleichung in eine gewöhnliche Differentialgleichung überführen, von der man die exakte Lösung kennt. Rücksubstitution liefert dann die Lösung der partiellen Differentialgleichung.

Lemma 4.28. *Es seien $k, \omega_z \neq 0$ fixiert, $\omega_0 \in (0, \infty)$ fest, g die 3D Röntgentransformation. Weiterhin seien:*

$$\omega := \omega(t) := \omega_0 \cos t, \quad \delta := \delta(t) := \frac{\omega_0}{\omega_z} \sin t \quad t \in [0, 2\pi),$$

dann gilt: Gleichung (4.76) kann geschrieben werden als:

$$\frac{d}{dt} \hat{p}_3(\omega(t), k, \omega_z, \delta(t)) = -ik \hat{p}_3(\omega(t), k, \omega_z, \delta(t)). \quad (4.77)$$

Beweis. Gleichung (4.77) ist äquivalent zu:

$$\begin{aligned} & \underbrace{\left(\frac{-\omega_0 \sin t}{= -\delta(t)\omega_z}\right)}_{= -\delta(t)\omega_z} \frac{\partial \hat{p}_3(\omega(t), k, \omega_z, \delta(t))}{\partial \omega} + \underbrace{\left(\frac{\omega_0 \cos t}{\omega_z}\right)}_{= \frac{\omega(t)}{\omega_z}} \frac{\partial \hat{p}_3(\omega(t), k, \omega_z, \delta(t))}{\partial \delta} \\ &= -ik\hat{p}_3(\omega(t), k, \omega_z, \delta(t)), \end{aligned}$$

und wenn wir nun diese Gleichung auf beiden Seiten mit $i\omega_z$ durchmultiplizieren, so erhalten wir Gleichung (4.76) als äquivalente Gleichung. □

Satz 4.29. Die Lösung von Gleichung (4.77) ist:

$$\hat{p}_3(\omega(t), k, \omega_z, \delta(t)) = e^{-ikt}U(\omega_0, k, \omega_z),$$

wobei $U(\omega_0, k, \omega_z) = \hat{p}_{reb}(\omega_0, k, \omega_z) = e^{-ik\pi}\hat{p}_{reb}(-\omega_0, k, \omega_z)$ ist.

Beweis.

$$\frac{d}{dt} \left(e^{-ikt} \underbrace{U(\omega_0, k, \omega_z)}_{\text{unabh. von } z} \right) = -ike^{-ikt}U(\omega_0, k, \omega_z).$$

An sich kann U eine beliebige glatte Funktion für $\omega_0 > 0$ sein. Spezifiziert wird U jedoch durch das AWP:

$$\begin{aligned} \hat{p}_3(\omega(0), k, \omega_z, \delta(0)) &= \underbrace{\hat{p}_3(\omega_0, k, \omega_z, 0)}_{= \hat{p}_{reb}(\omega_0, k, \omega_z)} \\ &= e^0U(\omega_0, k, \omega_z). \end{aligned}$$

Es folgt daher:

$$\begin{aligned} U(\omega_0, k, \omega_z) &= \hat{p}_{reb}(\omega_0, k, \omega_z) \\ &\stackrel{\text{Satz 4.23 a)}}{=} e^{-ik\pi}\hat{p}_{reb}(-\omega_0, k, \omega_z) \quad \text{für } \omega_0 > 0. \end{aligned}$$

□

Korollar 4.30. Die allgemeine Lösung von Gleichung (4.76) lautet:

$$\hat{p}_3(\omega, k, \omega_z, \delta) = \begin{cases} e^{-ik \arctan\left(\frac{\delta\omega_z}{\omega}\right)} U(\sqrt{\omega^2 + \delta^2\omega_z^2}, k, \omega_z) & ; \quad \omega > 0 \\ e^{ik\pi - ik \arctan\left(\frac{\delta\omega_z}{\omega}\right)} \underbrace{U(\sqrt{\omega^2 + \delta^2\omega_z^2}, k, \omega_z)}_{= \omega_0} & ; \quad \omega < 0 \end{cases}$$

wobei $U(\omega_0, k, \omega_z) = \hat{p}_{reb}(\omega_0, k, \omega_z)$ eine glatte Funktion ist.

Beweis. Rücksubstitution ergibt:

$$\frac{\delta}{\omega} = \frac{1}{\omega_z} \frac{\sin t}{\cos t} \quad \Rightarrow \quad \arctan\left(\frac{\delta\omega_z}{\omega}\right) = t$$

und

$$\sqrt{\omega^2 + \delta^2\omega_z^2} = (\omega_0^2 \cos^2 t + \frac{\omega_0^2}{\omega_z^2} \sin^2 t \omega_z^2)^{1/2} = \sqrt{\omega_0^2} \stackrel{\omega_0 \geq 0}{=} \omega_0.$$

Aus Satz 4.29, t und ω_0 wie gerade gezeigt, folgt dann die Behauptung. \square

4.7 Vergleich der Algorithmen dieses Kapitels

Die nun folgende Diskussion orientiert sich an den Ideen und Ergebnissen aus den Papern [14], [15], [16] und [41].

Der Übergang vom 2D zum 3D Erwerb führt uns zu einer wesentlichen Verbesserung der Scanner-Sensitivität. Die ist auf die wachsende Anzahl der gemessenen LOR's zurückzuführen.

Gewöhnlich werden die 3D-Daten unter Verwendung des 3DRP, eine 3D gefilterte Rückprojektionsmethode, die man durch Diskretisierung einer analytischen Rekonstruktionsformel erzielt, rekonstruiert. Wegen der beachtlichen Anzahl an LOR's im 3D-Modus ist es auch nicht weiter verwunderlich, dass der 3DRP viel mehr Zeit benötigt als die 2D FBP. Ein weiterer Grund für die wachsende Komplexität ist der, dass die Rekonstruktion des 3D Bildes nicht in die Rekonstruktion einer Menge unabhängiger Ebenen zerlegt werden kann.

Da die Rekonstruktion der 3D PET-Daten eine, was die Rechenzeit angeht, intensive Aufgabe ist, kann gerade dies die klinische Anwendung in speziellen Bereichen, wie z. B. der dynamischen Ganzkörperstudie, beschränken. Dabei werden nämlich mehrere Scans desselben Patienten benötigt und abhängig von der verfügbaren Hardware, dem Studientyp und dem Patientendurchsatz kann die Rechenzeit für eine akzeptable klinische Anwendung einfach zu hoch sein. Dies motiviert zu weiterführender Forschung, um schnellere Rekonstruktionsalgorithmen zu erzielen.

Die bisher schnellsten Methoden basieren auf Rebinning. Dies bedeutet, dass die redundanten 3D Daten in einen Stapel unabhängiger 2D Datenmengen (Sinogramme) sortiert werden, welche unter Verwendung der 2D FBP oder 2D iterativen Algorithmen rekonstruiert werden können.

Eine praktische Rebinningmethode ist FORE. Die Genauigkeit von FORE sinkt mit zunehmendem Wert des Winkels θ . (Erklärung siehe Seite 113). Trotzdem wird dieser Algorithmus routinemäßig von mehreren Gruppen angewendet und es konnte gezeigt werden, dass er hinreichend genau für die meisten klinischen Studien mit Multi-Ring-Scannern ist, die einen Öffnungswinkel von ungefähr $\theta_{max} \approx 15^\circ$ nicht überschreiten.

Zusätzlich zu seinem approximativen Charakter gibt eine zweite Einschränkung von FORE, nämlich diese, dass er statistische Eigenschaften der Daten modifiziert: Im Gegensatz zu den gemessenen 3D-Daten sind die rebinnten Sinogramme nicht als unabhängige

Poisson-Variablen verteilt und deshalb sollten sie im Prinzip nicht unter Verwendung eines standardmäßigen statistischen Algorithmus, so wie z.B. OSEM aus Kapitel 5, rekonstruiert werden. (Erstaunlicherweise ist OSEM jedoch ein sehr robuster Algorithmus, der mit guten Resultaten auf die 2D-Daten, rebinnt mit FORE, angewendet werden kann).

Wegen dieser zwei Einschränkungen - approximativer Charakter und komplexer Wirkung auf die Datenstatistik - ist es wahrscheinlich, dass 3D iterative Algorithmen (Näheres dazu finden wir im nächsten Kapitel) FORE bzw. die Rebinningalgorithmen überhaupt ersetzen werden (Rebinning, ob approximativ oder genau hat immer eine komplexe Wirkung auf die Datenstatistik), zumindest wenn die verfügbare Rechenleistung schneller wächst als die Anzahl der LOR's, die von den Scannern erworben werden. Jedoch solange sich OSEM noch in der Erprobung der klinischen Routine befindet (siehe Ausblick S. 181) sind die Rebinningmethoden, gerade wegen ihrer Schnelligkeit, für Scanner wie dem ECAT HRRT von Vorteil. Dieser erwirbt Datenmengen, die noch zu groß für eine routinemäßige Anwendung der 3D iterativen Algorithmen sind. Jedoch übersteigt der Öffnungswinkel 20° , was die axiale Auflösung bei der Verwendung von FORE erniedrigen kann. (Dies wurde mit Phantom-Studien auf dem ECAT HRRT demonstriert).

Diese Beobachtungen regten dazu an, den Vorteil der Anwendung eines exakten Rebinningalgorithmus, FORE-J, auf dem HRRT zu untersuchen. (Eine Darstellung der Ergebnisse dieser Untersuchung finden wir unter dem Punkt "Genauigkeit mit gemessenen Daten"). Dieser Algorithmus ist exakt wie FOREX, kann jedoch im Gegensatz zu FOREX direkt auf die axial abgeschnittenen Daten angewendet werden. Er vermeidet also den Vorwärtsprojektionsschritt, der bei der 3DRP und FOREX benötigt wird.

FORE-J basiert auf der Eigenschaft, dass die 3D Röntgentransformation einer Funktion die Lösung einer Differentialgleichung 2. Ordnung sein muss, die zuerst von John studiert wurde. Vom praktischen Gesichtspunkt ist FORE-J einfach zu implementieren, da er dieselbe Struktur wie FORE hat. Der einzige Unterschied ist eine kleine Korrektur, die vor dem Rebinning zu jedem schrägen Sinogramm dazuaddiert wird. Die Berechnung dieser Korrektur ist numerisch gesehen schnell, beinhaltet jedoch eine 2. Ableitung der Daten bzgl. der axialen Variablen z , eine Größe, die empfindlich für Rauschen ist.

Im Folgenden wollen wir nun die 3DRP, FORE, FOREX und FORE-J bzgl. Rekonstruktionszeit, Rauschen und Genauigkeit miteinander vergleichen:

Rekonstruktionszeiten

Wir betrachten einen zylindrischen Multi-Ring-Scanner mit den Parametern, die in Tabelle 4.1 dargestellt werden.

Die Daten werden simuliert, in dem wir analytisch die Linienintegrale über die mathematischen Phantome berechnen. Wir nehmen weiterhin an, dass die Daten bzgl. Detektorsensitivität, Streuung und Dämpfung vorkorrigiert wurden und dass die Reihen mit konstantem Winkel ϕ in jedem Sinogramm Mengen paralleler LOR's entsprechen.

Ringe	36	Detektorabstand	$\Delta z = 8 \text{ mm}$
Ebenen	71	Ebenenabstand	$\Delta z/2 = 4 \text{ mm}$
Ringradius	$R_D = 256 \text{ mm}$	Öffnungswinkel	$\theta_{max} = 29^\circ$
Winkelmessungen	128	Radius FOV	$R = 124 \text{ mm}$
Radialmessungen	128	radiales Sampling	$\Delta x_r = 2 \text{ mm}$
Ebenengröße	128×128	Voxelgröße	$\Delta x = \Delta y = 2 \text{ mm}$

Tabelle 4.1: Quelle: [16], Tabelle 1, S. 1059.

Die Algorithmen werden für eine "UltraSparc Workstation (170 MHz)" mit 128 Mbyte Speicherplatz implementiert und den Ergebnissen aus [16] folgend erhalten wir:

Die Zeit, um die Daten zu rebinnen beträgt 148 s für FORE-J, 59 s, für FORE und 567 s für FOREX. Nach 27 s, die benötigt werden, um die Menge von 71 Ebenen mit FBP zu rekonstruieren, erhalten wir das Verhältnis der Rekonstruktionszeiten von:

$$\frac{FOREX}{FORE - J} = \frac{567 + 27}{148 + 27} = \frac{594}{175} \approx 3,4$$

$$\frac{FORE - J}{FORE} = \frac{148 + 27}{59 + 27} = \frac{175}{86} \approx 2,0$$

Zum Vergleich: Die gesamte Zeit für den 3DRP ist 5600 s.

Wir sollten jedoch beachten, dass eine schnellere Implementierung dieses Algorithmus vorgeschlagen wurde und in dem Paper von "M. L. Egger, C. Joseph, C. Morel, Incremental beamwise backprojection using geometrical symmetries for 3D PET reconstruction in a cylindrical scanner geometry, Phys. Med. Biol. Volume 43, 1998, pp. 3009-24" nachzulesen ist.

Rauschen

Das Hauptziel der 3D PET besteht darin, die SNR zu verbessern, indem eine redundante Menge der 3D Röntgentransformation gemessen wird. Deshalb sollte ein praktikabler Rebinningalgorithmus nicht nur genau sein, sondern er sollte auch die redundanten Daten verwerten, um so gut wie es geht die Varianz im rekonstruierten Bild zu reduzieren.

Leider handelt es sich bei den folgenden Abbildungen um unterschiedliche verrauschte Datenmengen, so dass ein direkter Vergleich nur bedingt möglich ist.

Aus Abbildung 4.26 können wir die Standardabweichung in der Rekonstruktion eines gleichmäßigen Zylinders, simuliert mit einem 32-Ring-Scanner mit großer Öffnung, ablesen. (Je kleiner die Standardabweichung, umso größer ist die SNR). Während die Standardabweichung, die mit den beiden Algorithmen FORE und 3DRP rekonstruiert wird, in den zentralen Ebenen des Scanners fast identisch ist, wächst die Standardabweichung der FORE Rekonstruktion in den äußeren Ebenen (in Übereinstimmung mit dem ungleichmäßigen axialen Sensitivitätsprofil des Volumenscanners). Diese Eigenschaft von FORE wird auch bei dem SSRB-Algorithmus beobachtet. Im Gegensatz dazu resultiert der 3DRP-Algorithmus in einer fast gleichmäßigen Standardabweichung über das gesamte axiale FOV.

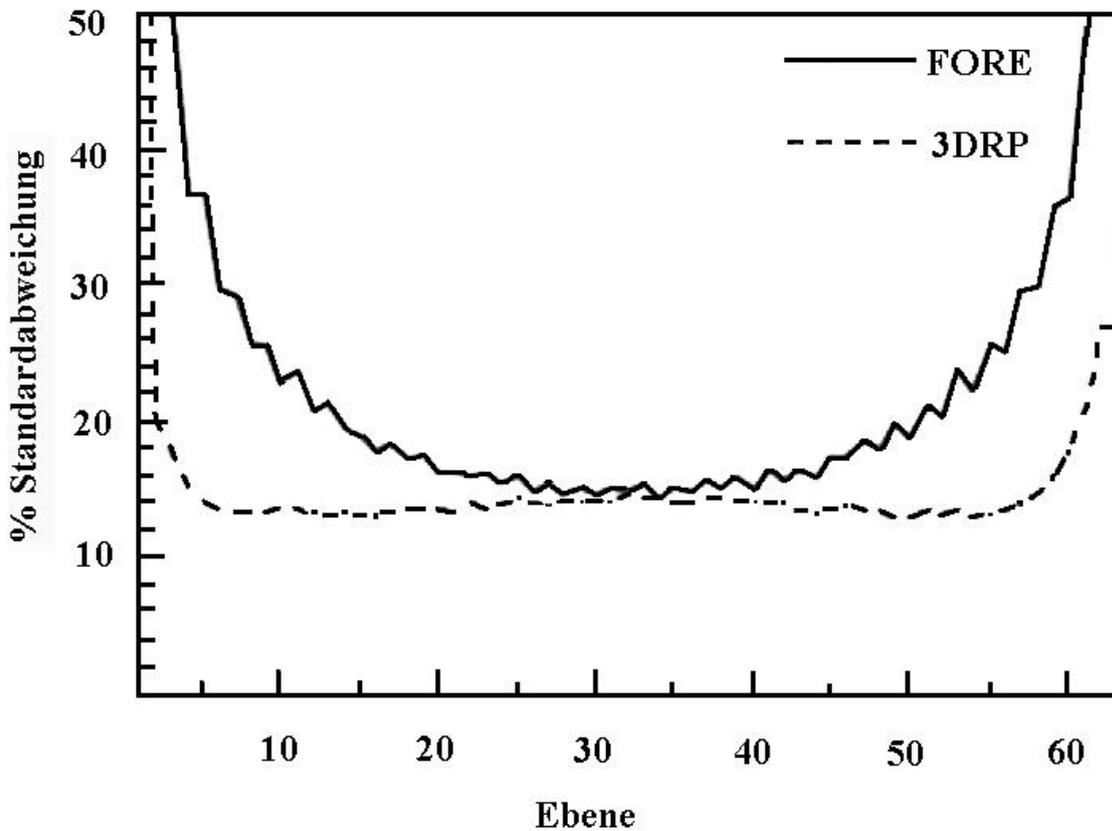


Abbildung 4.26: Rauscheigenschaften des 3DRP und des FORE Algorithmus für einen gleichmäßigen 20 cm Zylinder mit großer Scanneröffnung. Die relative Standardabweichung in einer Region (kreisförmig), die von Interesse ist, wird dabei auf der y-Achse abgetragen, der Ebenenindex auf der x-Achse. Quelle: [14], Figur 10, S. 157.

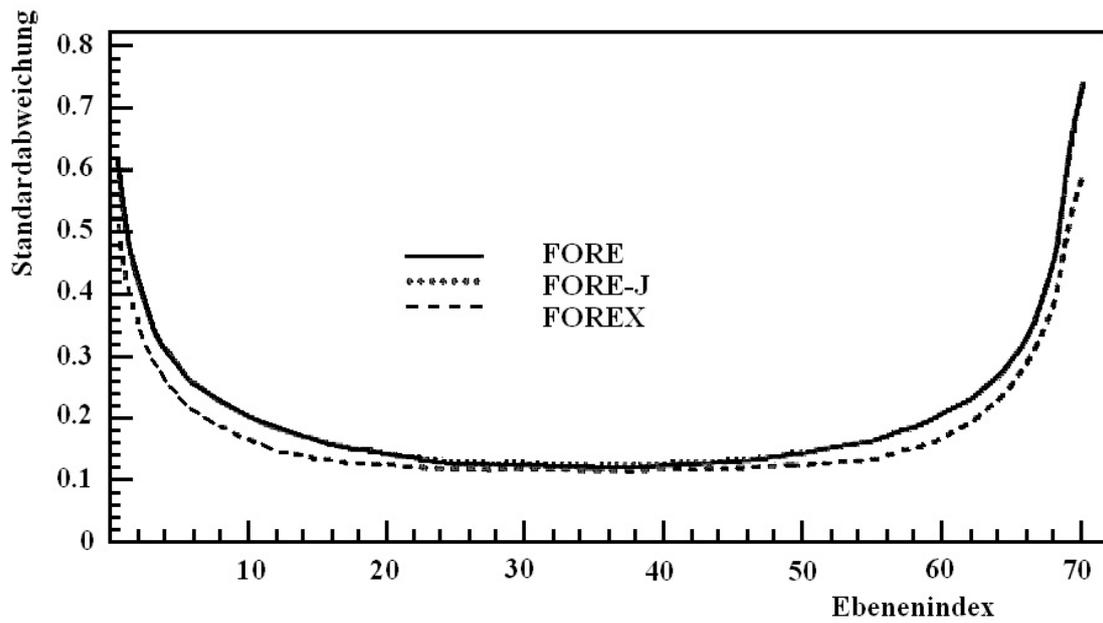


Abbildung 4.27: Relative Standardabweichung bei der Rekonstruktion simulierter verauschter Daten eines gleichmäßigen Zylinders als Funktion der axialen Koordinate z (Ebenenindex). Die Kurven von FORE und FORE-J sind fast identisch. Quelle: [16], Figure 8, S. 1061.

Bei den verrauschten Daten aus Abbildung 4.27 handelt es sich um die simulierten Daten eines gleichmäßigen Zylinders ($r = 120\text{ mm}$, Scanner aus Tabelle 4.1), auf die wir Zufalls-Poisson-Rauschen addiert haben, welches 100 Millionen Koinzidenzen entspricht. Wir schätzen dann die Varianz der Rekonstruktionen, indem wir in jeder transaxialen Ebene die Varianz der Voxel, die in einem Gebiet von $r = 100\text{ mm}$ enthalten sind, berechnen. Die Resultate bestätigen, dass die SNR von FORE und FORE-J fast identisch sind und FOREX eine etwas höhere SNR in den äußeren Ebenen liefert.

Genauigkeit mit simulierten, rauschfreien Daten

Wir betrachten ein mathematisches Phantom, das aus einem gleichmäßigen Zylinder besteht, der ein paar dünne Scheiben mit Kontrasten, die zwischen 0,1 und 0,5 variieren, enthält. Die Daten für diesen Zylinder werden simuliert und mit verschiedenen Algorithmen verarbeitet.

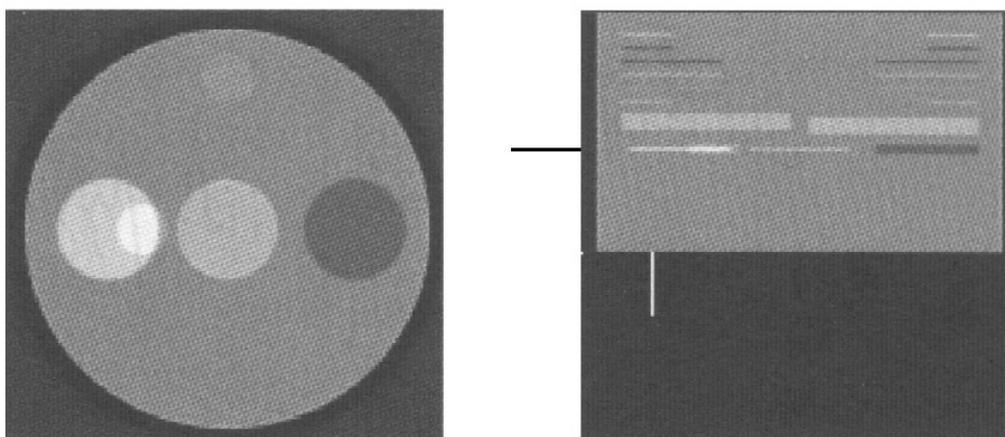


Abbildung 4.28: Transaxiale und axiale Ansicht des mathematischen Phantoms eines gleichmäßigen Zylinders mit mehreren Niedrigkontrastscheiben. Die Profile aus den Abbildungen 4.29 und 4.30 wurden entlang der Linien, die in der rechten Abbildung dargestellt sind, gemacht. Quelle: [16], Figure 6, S. 1059.

Auch hier stammen die Vergleiche aus zwei unterschiedlichen Papern. Es handelt sich zwar augenscheinlich um dasselbe Phantom, allerdings müssen wir mit zwei Abbildungen Vorlieb nehmen, da nicht alle Algorithmen in einer einzigen Abbildung miteinander verglichen werden und unterschiedliche Scanner zum Simulieren der Daten benutzt werden.

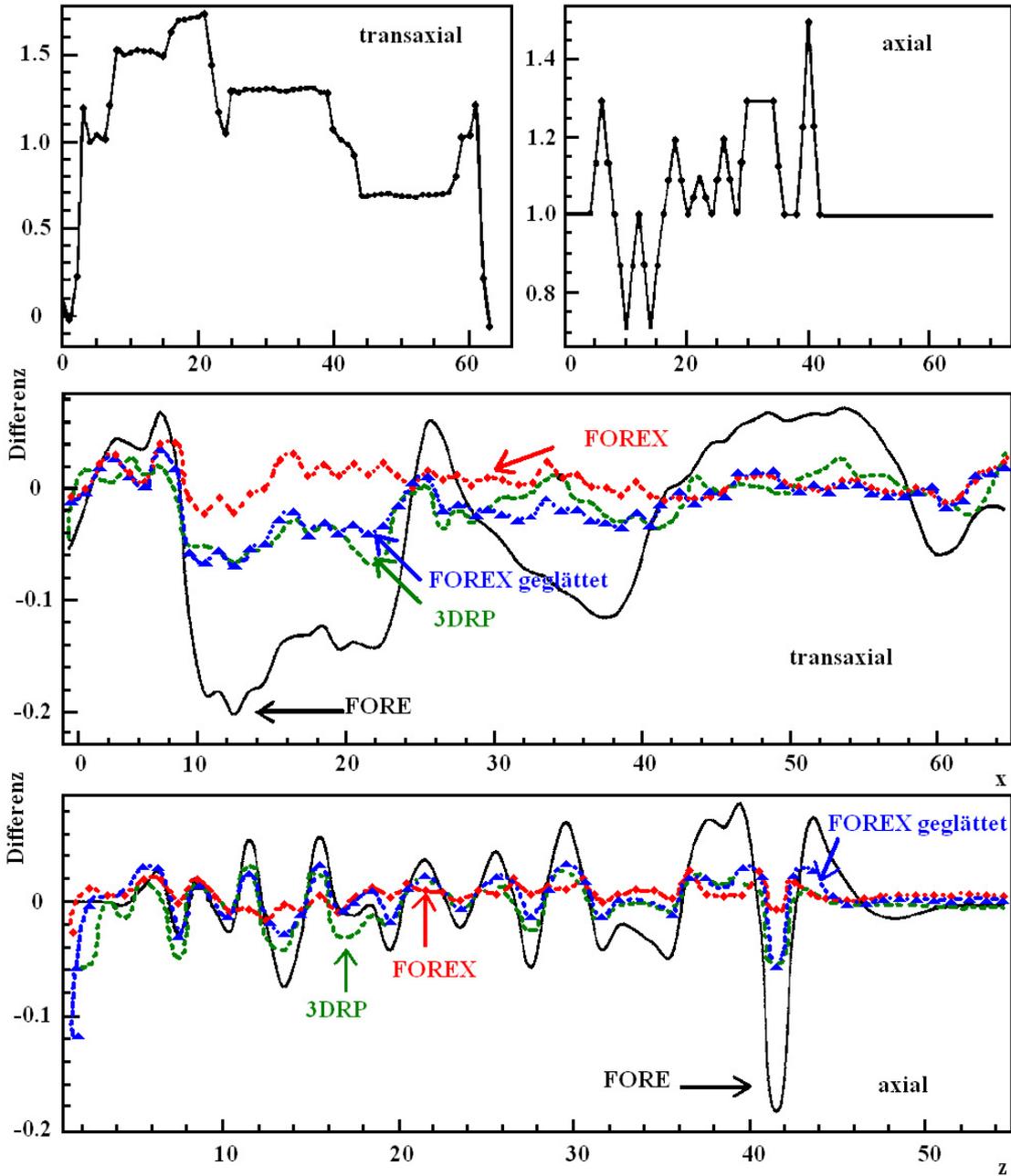


Abbildung 4.29: Transaxiale (links) und axiale (rechts) Profile des simulierten Phantoms aus Abbildung 4.28.

Obere Reihe: Die Profile stammen von der 2D Simulation, die als Referenz dient.

Mittlere und untere Reihe: Die Differenz (Profil, das von dem entsprechenden Algorithmus erzielt wurde minus Referenzprofil) der Profile in der transaxialen Ebene wird in der mittleren Reihe graphisch dargestellt, die untere Reihe steht dementsprechend für die Differenz in axialer Richtung. Quelle: [15], Figure 10, S. 663.

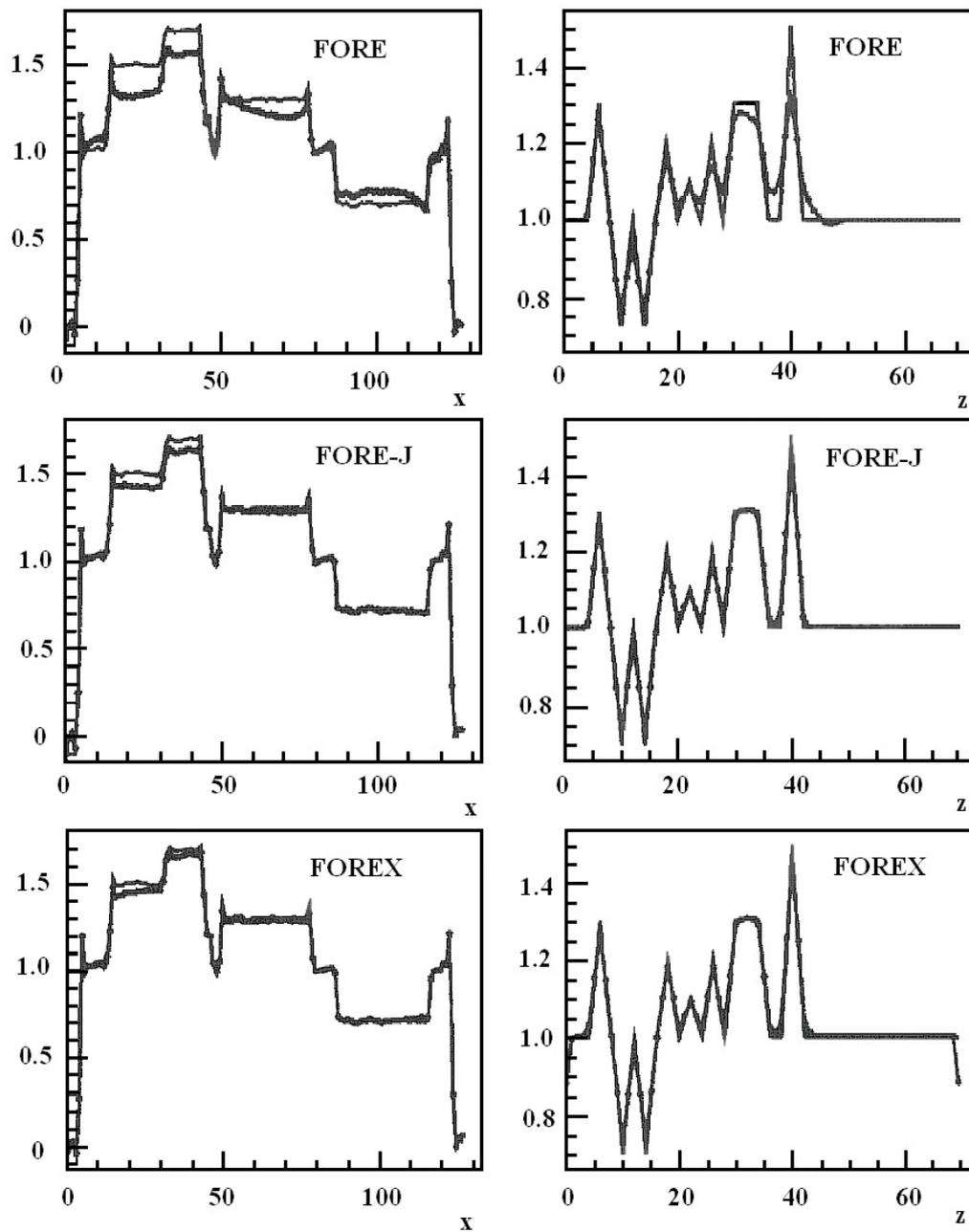


Abbildung 4.30: Transaxiale und axiale Profile des simulierten Phantoms aus Abbildung 4.28. In jeder Reihe sind die Profile mit den vollen Kurven von der 2D Simulation, die als Referenz diente. Die Profile mit Punkten wurden von den Bildern genommen, die mit FORE (1. Reihe), FORE-J (2. Reihe) und FOREX (3. Reihe) rekonstruiert wurden. Quelle: [16], Figure 7, S. 1060.

Die simulierten, rauschfreien Daten ($\delta_{max} = 25^\circ$) für das Phantom aus Abbildung 4.28 werden mit FORE, FOREX und 3DRP verarbeitet. (Dabei werden die Algorithmen auf einer SUN Ultra Sparc 1 Station (170 MHz) mit 128-Mb RAM implementiert und getestet). Wie wir in Abbildung 4.29 erkennen können, besitzt FORE (durchgezogene Linie, schwarz) sowohl im axialen wie auch dem transaxialen Bereich die größten Fehler. FOREX hat eine verbesserte Genauigkeit gegenüber FORE und ist vergleichbar mit dem 3DRP (gestrichelte Linie, grün).

Bei dem axial geglätteten FOREX werden mit FOREX rebinnte Sinogramme axial mit einem Drei-Punkt Filter geglättet. Dieser Filter wird empirisch erstellt, um zumindest in den zentralen Ebenen des Scanners die Wirkung der axialen, linearen Interpolation in FORE und 3DRP nachzuahmen. Der axial geglättete FOREX (gestrichelte Linie mit Dreiecken, blau) liegt sehr nahe bei dem 3DRP, wohingegen der nicht geglättete FOREX (gestrichelte Linie mit Punkten, rot) die beste Auflösung hat.

Bei Abbildung 4.30 wurden die simulierten, rauschfreien Daten für das Phantom aus Abbildung 4.28 (Scanner aus Tabelle 4.1) mit den drei Algorithmen FORE, FORE-J und FOREX rebinnt. Jedes rebinnte Sinogramm wurde dann mit 2D FBP und einem Ramp-Filter, abgeschnitten bei der Nyquistfrequenz $\frac{1}{2\Delta x_r}$ rekonstruiert. Transaxiale und axiale Profile durch die drei rekonstruierten Bilder demonstrieren die verbesserte Genauigkeit von FORE-J im Vergleich zu FORE. Einige Diskrepanzen zwischen FORE-J und FOREX werden ebenfalls im transaxialen Profil beobachtet. Obwohl wir weitere Studien zur Bestätigung dieses Sachverhaltes benötigten, könnten wir diese kleine Diskrepanz mit der begrenzten Genauigkeit der Berechnung der 2. Ableitung in "Q" erklären.

Genauigkeit mit gemessenen Daten

Nach der Durchführung eines FDG-Gehirn-Scans mit dem ECAT EXACT HR rekonstruieren wir die Daten einmal mit dem axialen geglätteten FOREX und einmal mit 3DRP. (Die Algorithmen werden dabei wieder auf SUN Ultra Sparc 1 (170 MHz) mit 128-Mb RAM implementiert). Die Differenz zwischen den zwei Rekonstruktionen ist so klein, dass sie vom Rauschen nicht zu unterscheiden ist. Beachten sollten wir jedoch die Tatsache, dass die Unterschiede in den äußeren Ebenen größer sind.

Die gleichen Beobachtungen können wir beim Vergleich von FOREX und FORE machen.

Im Folgenden wollen wir, wie schon zuvor angekündigt, den praktischen Nutzen des Algorithmus FORE-J bewerten. Dazu verwenden wir die Daten, die von dem ECAT HRRT erzeugt wurden. Dieser Scanner hat folgende Eigenschaften:

- ein axiales FOV von 25,2 cm,
- 104 Detektorringe,
- jeder Ring hat eine oktagonale Geometrie,
- die List-Mode Daten des HRRT werden in Sinogrammen mit 256 radialen Elementen und 288 Winkelmessungen reorganisiert,

- das radiale Sinogramm-Samplingintervall beträgt 1,22 mm.

Wegen den Verbindungen zwischen den acht Seiten eines oktagonalen Detektors können die Sinogramme nicht in acht diagonalen Streifen, "gaps" genannt, gemessen werden. Um die Menge der Daten, die gespeichert und verarbeitet wird zu reduzieren, wird normalerweise axiales Winkeluntersampeln (Kompression) benutzt. Hier benutzen wir die aktuelle Sinogramm-Menge, die aus 2209 Sinogrammen (Span = 9, Ringdifferenz = 67, 325,7 Mbyte) besteht und 88% der LOR-Informationen abdeckt.

Vor der Anwendung von fourierbasiertem Rebinning werden die "gaps" in jedem Sinogramm, mittels linearer Interpolation in Winkelrichtung, aufgefüllt.

Auch hier führen wir einen FDG-Gehirn-Scan durch und rekonstruieren die Daten sowohl mit FORE als auch mit FORE-J (Voxelgröße: 1,22 mm). Die gesamte Rekonstruktionszeit für die zwei Algorithmen beträgt 500 s für FORE bzw. 717 s für FORE-J. Die Rekonstruktionen sind sich sehr ähnlich, jedoch weist FORE-J geringe Artefakte auf, die als horizontale Linien in der koronalen und sagittalen Ebene und als Ringe in der transaxialen Ebene erscheinen. Diese Artefakte sind mit FORE nicht sichtbar und wir schreiben sie versuchsweise der Rauschvermehrung bei der Berechnung der 2. Ableitung im "Q-Term" zu.

Bei den geglätteten Bildern, die in der Praxis benutzt werden, sind diese Artefakte kaum sichtbar. Eine genaue Betrachtung dieser Bilder und ein Vergleich mit einer 3D OSEM-Rekonstruktion mit 8 Teilmengen, 2 Iterationen und ungefähr 4 Stunden pro Iteration (OSEM siehe Kapitel 5) deckt mehrere Details in der Kortikalstruktur auf, die besser von FORE-J als von FORE reproduziert werden.

Zusammenfassung

Wie wir gesehen haben ist der approximative FORE also der bei weitem schnellste Algorithmus, gefolgt von dem genauen FORE-J, dem genauen FOREX und dem weit abgeschlagenen 3DRP. Die Unterschiede in der Rekonstruktionszeit der Rebinningalgorithmen lassen sich wohl damit erklären, dass FOREX im Vergleich zu FORE und FORE-J die abgeschnittenen Daten schätzt. FORE-J ist langsamer als FORE, da der zusätzliche Term "Q" berechnet werden muss.

Was das Rauschen angeht konnten wir erkennen, dass die SNR von FORE und FORE-J fast identisch waren, FOREX in den zentralen Ebenen ebenfalls dicht an FORE bzw. FORE-J ist, jedoch in den äußeren Ebenen eine etwas höhere SNR liefert. Ebenso haben wir gesehen, dass der 3DRP in den zentralen Ebenen fast mit FORE identisch ist, in den äußeren Ebenen jedoch eine höhere SNR als FORE liefert und über das gesamte FOV eine fast gleichmäßige SNR hat. Die verbesserte SNR von FOREX und 3DRP in den äußeren Ebenen kann damit erklärt werden, dass dort eine noch größere 3D Datenmenge - durch die Vorwärtsprojektion - für die beiden Algorithmen bereitsteht.

Bei der Genauigkeit mit rauschfreien, simulierten Daten haben wir festgestellt, dass der approximative FORE sowohl im axialen wie auch transaxialen Bereich der ungenaueste Algorithmus ist. FOREX hat die beste Auflösung und dazwischen liegen FORE-J, 3DRP

und der axial geglättet FOREX.

Bei einem Vergleich der Genauigkeit mit gemessenen Daten war die Differenz von FOREX und 3DRP in den zentralen Ebenen so gering, dass sie kaum vom Rauschen zu unterscheiden war. Nur in den äußeren Ebenen sind die Unterschiede größer. Dasselbe gilt für einen Vergleich von FOREX mit FORE. Ein Vergleich von FORE-J mit FORE (andere Daten und anderer Scanner als zuvor) zeigte, dass, obwohl die räumliche Auflösung durch die Anwesenheit von "gaps" beim HRRT und durch axiales "untersampeln" beschränkt ist, FORE-J die axiale Auflösung für den ECAT HRRT bedeutend verbessert. Die Implementierung von FORE-J benötigt nur eine geringe Modifizierung von FORE, die nur etwas die gesamte Rekonstruktionszeit erhöht. Es scheint, dass FORE-J einige feine Details beim Gehirnsan besser aufdeckt. Trotzdem führt die vorläufige Implementierung von FORE-J zu geringen Artefakten. Alternative Methoden, um die 2. Ableitung im "Q-Term" zu diskretisieren, werden untersucht und es werden zusätzliche Tests mit Phantom- und Klinikdaten unternommen und in naher Zukunft präsentiert werden.

Wir haben also gesehen, dass FOREX eine ähnliche Genauigkeit wie 3DRP hat und um einiges schneller ist. Die Tragweite der Verbesserung, verglichen mit FORE, ist auf die Scannergeneration aus den letzten Jahren beschränkt, aber FOREX ist eine nützliche Alternative zum 3DRP für Scanner mit großem Öffnungswinkel. Der Geschwindigkeitsvorteil jedoch könnte, im Vergleich mit einer schnelleren Implementierung des 3DRP oder mit einer direkten 3D Fouriermethode, reduziert werden.

FOREX Hauptnachteil ist die Tatsache, dass die Matrixdimension aus Zweierpotenzen oder Produkten von kleinen Primzahlen mit Zweierpotenzen bestehen muss, um die FFT zu optimieren. Dies kann durch "Zero-Padding", "Trimming" und/oder "Resampling" erreicht werden, obwohl Resampling in dem Verlust von Auflösung resultieren könnte. FORE-J ist dahingegen der erste Rebinningalgorithmus, der die folgenden Eigenschaften kombiniert:

1. er ist exakt, bis auf Diskretisierungsfehler und
2. er kann direkt auf axial abgeschnittene Daten angewendet werden, ohne zuerst die Daten für einen virtuell längeren Scanner zu schätzen.

Resultate demonstrierten, dass FORE-J fast so genau wie der exakte FOREX ist und dreimal schneller als FOREX ist. Ebenso ist FORE-J genauer als FORE, jedoch doppelt so langsam.

Daraus erkennen wir, dass FORE-J ein wertvolles Werkzeug für zukünftige PET-Scanner mit großem Öffnungswinkel ist.

5 Das D-D-Modell und darauf beruhende Bildrekonstruktionen

ET ist im wesentlichen stochastischer Natur. Wegen der "geringen" Anzahl an Zählungen, ist der stochastische Aspekt viel ausgeprägter als in der Transmissionstomographie. Deshalb wurden neben den Modellen, die Integraltransformationen benutzen (siehe auch Kapitel 4), auch stochastische Modelle für die ET entwickelt. Diese Modelle sind komplett diskret.

Im Folgenden werden wir nun das D-D-Modell (discrete-discrete model) vorstellen, fünf allgemeine Komponenten für Bildrekonstruktionsansätze zu diesem Modell nennen und dann den Ansatz von Shepp und Vardi (1982), auf dem viele weitere Ansätze beruhen, die Maximum Likelihood Rekonstruktion, vorstellen. Danach kommen wir auf einige Probleme, die dieser Ansatz mit sich bringt, zu sprechen und werden Lösungsvorschläge (u. a. OSEM), sowie einen Alternativansatz, der auf der Projektion auf affine Unterräume beruht (ART), vorstellen.

Der letzte Teil dieses Kapitels, also vor allem die Vorstellung von OSEM und ART, entspricht auch der momentanen Forschungsarbeit auf diesem Gebiet. Sie besteht zum Großteil darin, mathematische Ansätze zu finden, um mit Hilfe der immer besser und schneller werdenden Rechenanlagen die iterativen Methoden, deren Hauptproblem in ihrer Langsamkeit besteht, zu beschleunigen.

5.1 Das D-D Modell

Auch in diesem Abschnitt folgen wir den Ideen von Lewitt, [10].

Bei diesem Ansatz approximieren wir die gesuchte Funktion $f(x, y, z)$ durch eine endliche Linearkombination von Basisfunktionen $b_j(x, y, z)$, $j \in J$ (J endlich).

Seien \tilde{f}_j die Koeffizienten, die mit der entsprechenden Basisfunktion b_j multipliziert werden, so erhält man:

$$f(x, y, z) \approx \sum_{j=1}^J \tilde{f}_j b_j(x, y, z). \quad (5.1)$$

Ersetzt man $f(x, y, z)$ in unserem Grundmodell (3.1) durch die obige Summe, so folgt:

$$\begin{aligned}
 \tilde{g}_i^{LSV} &= \int \int \int_{\Omega} \sum_{j=1}^J \tilde{f}_j b_j(x, y, z) h_i(x, y, z) dx dy dz \quad i = 1, \dots, I \\
 &= \sum_{j=1}^J \tilde{f}_j \underbrace{\int \int \int_{\Omega} b_j(x, y, z) h_i(x, y, z) dx dy dz}_{=: a_{i,j}} \\
 &= \sum_{j=1}^J \tilde{f}_j a_{i,j} \quad i = 1, \dots, I.
 \end{aligned} \tag{5.2}$$

Somit folgt also:

$$\tilde{g}^{LSV} = A \tilde{f} \tag{5.3}$$

mit

$$A = (a_{i,j})_{\substack{i=1,\dots,I \\ j=1,\dots,J}}, \tilde{f} = (\tilde{f}_1, \dots, \tilde{f}_J)^\top, \tilde{g}^{LSV} = (\tilde{g}_1^{LSV}, \dots, \tilde{g}_I^{LSV})^\top.$$

Ziel: Finde einen Vektor \tilde{f} , so dass $A\tilde{f}$ möglichst nahe (entsprechend eines gewählten Maßes) bei dem Datenvektor \tilde{g}^{LSV} liegt.

Anmerkung 5.1. Für das im Folgenden vorgestellte Poisson-Modell der Datenstatistik ist die Poisson Likelihood Funktion ein solches natürliches Maß. Sie wächst, sobald die Vektoren ähnlicher werden.

5.2 Allgemeine Komponenten für das D-D Modell

Bildrekonstruktionsansätze, die auf dem D-D Modell basieren, bestehen aus fünf allgemeinen Komponenten (siehe [10] S. 6 ff.). Innerhalb jeder dieser allgemeinen Komponenten muss aus einem breiten Feld an Möglichkeiten eine spezielle Wahl getroffen werden. Diese fünf allgemeinen Komponenten lauten:

1. *Eine Menge von Basisfunktionen*

Diese sollen ermöglichen, dass das zu repräsentierende Bild durch eine endliche Anzahl von Parametern dargestellt wird. Es gibt zwei verschiedene Arten von Basisfunktionen: lokale und globale.

Das wesentliche Beispiel einer lokalen Basisfunktion ist das Voxel, für den die Basisfunktion einen konstanten Wert innerhalb eines kleinen Würfels hat und außerhalb dieses Würfels null ist.

Globale Funktionen dagegen, wie z. B. die Basisfunktionen von einer Fourierreihenentwicklung wirken (d. h. liefern Beiträge) über die gesamte Bildregion.

2. *Ein Modell der Physik des Messprozesses*

Das LSV-Modell (Kapitel 3) ist solch ein Modell, das die Spezifizierung des Kerns $h_i(x, y, z)$ zur Folge hat. Der Integrationskern $h_i(x, y, z)$ enthält die Effekte der Photonendämpfung sowie den radiometrischen Sensitivitätsfaktor für die Detektion von Photonen, die aus dem Punkt (x, y, z) hervorgehen. (Diese wurden bei den Algorithmen in Kapitel 4 nicht berücksichtigt.)

3. *Ein Modell für die gemessene Unsicherheit*

Dabei handelt es sich z. B. um ein Modell der Wahrscheinlichkeitsverteilung von jeder Messung um ihren Erwartungswert. Die Poisson-Verteilung ist solch ein Modell.

4. *Eine Kostenfunktion*

Diese Funktion besteht aus einem Bildeigenschaftskriterium und einem Dateneinbaukriterium. Sie gibt ein Maß dafür an, wie gut ein Bild zu den gegebenen Daten passt und wie gut dieses Bild den gewünschten Bildeigenschaften entspricht.

Wir nennen diese Funktion vorläufig $\Phi(\tilde{f}, g) = L(\tilde{f}, g) - \beta R(\tilde{f})$, wobei $L(\tilde{f}, g)$ die Logarithmus-Likelihood Funktion ist, auf die wir im Folgenden noch näher eingehen werden. Der Parameter β kontrolliert dabei die Balance zwischen dem Dateneinbaukriterium und dem Bildeigenschaftskriterium, sprich den Austausch zwischen räumlicher Auflösung und Rauschen im rekonstruierten Bild. $R(\tilde{f})$ ist ein Regularisierungsterm, der signifikante Abweichungen von einem gegebenen Bildmodell oder "Bildunebenheiten" bestraft.

5. *Einen numerischen Algorithmus*

Dieser soll die Werte der Koeffizienten der Basisfunktionen, die man in 1. gewählt hat, liefern unter Berücksichtigung der in 2. bis 4. getroffenen Kriterien.

Der numerische Algorithmus ist so konstruiert, dass er genau den Bildvektor liefert, der die Kostenfunktion maximiert (oder minimiert).

Wenn die Datenmenge relativ gering ist und eine einfache Kostenfunktion in 4. gewählt wurde, dann wäre es machbar einen nicht-iterativen Algorithmus zu benutzen. In den meisten Fällen jedoch existiert kein nicht-iterativer Algorithmus oder falls er existieren würde, dann würde er eine nicht realisierbare Menge an Berechnungen benötigen.

Daher gilt: In den meisten Fällen, bei denen die Bildrekonstruktion auf dem D-D Modell basiert, benötigen wir einen iterativen Algorithmus, der sukzessive Schätzungen der Koeffizienten der Basisfunktionen liefert.

Der Algorithmus wird so konstruiert, dass er eine Folge von Schätzungen liefert, die gegen eine Lösung (Maximum oder Minimum) der Kostenfunktion konvergieren.

Im Prinzip sollte die Wahl des numerischen Algorithmus verhältnismäßige Routine sein, wenn einmal die anderen Wahlen getroffen wurden. Das Abschlussbild sollte dann nur von diesen Wahlen und nicht von der Wahl des Algorithmus abhängig sein. In der Praxis ist es jedoch so, dass man den iterativen Algorithmus im Allgemeinen stoppt, bevor die

Folge von Schätzungen konvergiert. Somit führen verschiedene Algorithmen zu verschiedenen Abschlussbildern.

Historisch gesehen erhielt die Wahl des numerischen Algorithmus während der Entwicklung von Bildrekonstruktionsansätzen, die auf dem D-D Modell basieren, sehr viel Beachtung. Den Wahlen jenseits der einfachsten Möglichkeiten in den Kategorien 1. bis 4. wurde dahingegen kaum Beachtung geschenkt. Es ist jedoch mittlerweile allgemein bekannt, dass die in 1. bis 4. getroffenen Wahlen einen großen Einfluß auf die Charakteristiken des Abschlussbildes haben und dass der Gebrauch besserer Modelle zu besseren Bildern führt.

5.3 Iterative Methoden

Wie wir im vorangegangenen Abschnitt gesehen haben, benötigen wir in den meisten Fällen der auf dem D-D-Modell basierenden Bildrekonstruktionen einen iterativen Algorithmus. Aus diesem Grund werden wir hier auch nur die iterativen Methoden vorstellen. Im Gegensatz zu den transformierten Rekonstruktionsmethoden (siehe Kapitel 4), versuchen wir bei den iterativen Methoden stufenweise Schätzungen des Bildes f zu erhalten. Diese Methoden erlauben also eine genauere Modellierung des Datenerwerbs im Vergleich zu den Linienintegral-Modellen. Darüber hinaus kann die Aufnahme eines genaueren Modells des statistischen Rauschens bedeutende Vorteile haben. Ein weiterer Vorteil der iterativen Methoden ist die Möglichkeit, "a priori Informationen" über das Bild in den Rekonstruktionsprozess miteinzubeziehen. Solche "a priori Informationen" können darin besonders nützlich sein, dass sie den Rekonstruktionsalgorithmus gegen eine akzeptable Lösung führen, auch wenn die gewöhnlichen Bedingungen, die durch die Daten geliefert werden, schwach sind, z. B. in dem Falle geringer SNR. Eine solche Bedingung ist die Nichtnegativität, welche gewährleistet, dass sukzessive Schätzungen von f überall nicht negativ sind. (Das ist vernünftig, da f die Tracerverteilung repräsentiert). Eine andere Möglichkeit ist die Einbeziehung anatomischer Informationen, die man aus MRI oder CT-Studien über denselben Patienten erreicht.

Im Gegensatz dazu können transformierte Algorithmen, so wie die FBP, nur die Glattheitsbedingungen ausnutzen, die durch das "Apodization Window" erzwungen werden. Von der Vielzahl an iterativen Methoden, die für die numerische Lösung von diskreten Modellen in der Tomography vorgeschlagen wurden, wenden wir uns, [11] S. 110 ff., nur den beiden iterativen Methoden zu, die einen weitverbreiteten Gebrauch gefunden haben:

- EM (expectation-maximization),
- ART (algebraic reconstruction technique).

Im Anschluss daran werden wir noch eine kurze Zusammenfassung über die aktuellen Forschungen bzw. Forschungsergebnisse auf diesem Gebiet geben.

5.3.1 EM-Algorithmen

Das im Anschluß an diese Vorbemerkung vorgestellte Verfahren, der ML-EM Algorithmus, hat zu der Einführung vieler verwandter Techniken geführt, von denen wir ebenfalls einige genauer betrachten werden. Der ML-EM beruht darauf, dass wir eine bedingte Erwartung maximieren wollen. In der Theorie ist die Maximierung dieser bedingten Erwartung äquivalent zur ML-Aufgabe (Aufgabe, die Likelihood-Funktion zu maximieren). Deshalb liefert der EM Algorithmus ein praxisnahes Verfahren zur Berechnung iterativer ML-Schätzer. ML (Maximum Likelihood) ist dabei eine Technik, um Schätzer zu finden und ML-Schätzer sind die Werte der unbeobachteter Parameter eines theoretischen Modells, basierend auf physikalischen und statistischen Prozessen, die in die Studie einbezogen sind, und die die gemessenen Daten höchstwahrscheinlich machen.

Modellierung

Vorlagen zur Modellierung lieferten uns die Quellen [17], [18] und [11], S. 45 und 46.

Bei der nun folgenden Herleitung betrachten wir gleich den 3D Fall. (Der 2D Fall würde völlig analog dazu aussehen).

gesucht: $f = f(x, y, z) \geq 0$ unbekannte Emissionsdichte.

Sei g_i die Anzahl der in Detektorpaar i gemessenen Koinzidenzen (gemessene Daten), $i = 1, \dots, n$.

Problem: Schätze f aus $g = (g_1, \dots, g_n)^\top$.

Wir suchen nun einen Schätzer \hat{f} von f , der die Wahrscheinlichkeit $p(g | f)$ maximiert, wobei $p(g | f)$ die Wahrscheinlichkeit der beobachteten, aktuellen Detektorzählraten g über alle möglichen Dichten f ist. Dabei gehen wir - uns an Abschnitt 5.2 orientierend - folgendermaßen vor:

Unser erster Schritt besteht nun in der Diskretisierung der zu rekonstruierenden Region in Voxel (im 2D Fall in Pixel), um eine Menge lokaler Basisfunktionen zu erhalten. Es gilt also:

$$f = (f_1, \dots, f_m)^\top,$$

und zu lösen ist somit das lineare, inverse Problem:

$$Af = g$$

mit:

$f = (f_1, \dots, f_m)^\top$, den gemessenen Daten $g = (g_1, \dots, g_n)^\top$ und der Matrix A , die unten noch genauer spezifiziert wird. Unser zweiter Schritt besteht nun in der Wahl des

Modells der Physik des Messprozesses. Ein "nahezu perfektes" mathematisches Modell der obigen Physik des Messprozesses ist das **Poisson-Modell** :

Es sei $n_j, j = 1, \dots, m$ die wahre, unbekannte Anzahl an Emissionen in Voxel (Pixel) j und wir nehmen an, dass die n_j 's unabhängige, Poisson-verteilte ZV'en mit (unbekanntem) Erwartungswert

$$En_j = f_j \quad j = 1, \dots, m$$

sind. Es gilt also:

$$n_j \sim \text{Poisson} \{En_j = f_j, \text{Var } n_j = f_j\} \quad \text{für } j = 1, \dots, m,$$

mit:

$$p(n_j = k) = e^{-f_j} \frac{f_j^k}{k!} \quad k = 0, 1, 2, \dots$$

(Siehe dazu auch noch einmal Abschnitt 2.3, Anmerkung 2.43 und die Darstellung der Poisson-Verteilung als Beispiel einer diskreten Verteilung auf Seite 28.)

Der Erwartungswert $En_j = f_j$ ist hier also ein Maß für die Aktivität in Voxel (Pixel) j und unser Problem nach der Diskretisierung und Modellbildung lautet nun:

Problem: Schätze f_j oder n_j aus den beobachteten Daten $g_i, i = 1, \dots, n$, für jedes $j = 1, \dots, m$.

Nehmen wir nun weiter an, dass sobald eine Emission in Voxel (Pixel) j stattgefunden hat, die bedingte Wahrscheinlichkeit, dass sie in Detektorpaar i detektiert wird, unabhängig von anderen Emissionen ist und durch:

$$a_{ij} := P(\text{detektiert in } i \mid \text{emittiert in } j) \quad i = 1, \dots, n, j = 1, \dots, m$$

gegeben ist, wobei die a_{ij} 's bekannte, nichtnegative Konstanten sind mit:

$$0 < a_{.j} = \sum_{i=1}^n a_{ij} \leq 1 \quad j = 1, \dots, m.$$

Die Matrix A wird entweder theoretisch oder durch Messungen bestimmt [11]. Beeinflusst werden die a_{ij} 's durch eine Anzahl von Faktoren, von denen wir hier nur einige aufzählen werden:

- geometrische Beziehung zwischen Voxel (Pixel) und Detektorpaar
- Laufzeit des Datenerwerbs
- Rate des Radioaktivitätszerfalls
- Detektorantwort
- Streuung

- Dämpfung

⋮

Dann erhalten wir

$$n(i, j) = \text{Anzahl der Emissionen, die in } j \text{ stattgefunden haben} \\ \text{und in } i \text{ detektiert wurden} \quad i = 1, \dots, n \quad j = 1, \dots, m$$

durch "Ausdünnung" jedes n_j gemäß der Wahrscheinlichkeiten a_{ij} , $i = 1, \dots, n$, sprich:

$$n(i, j) = n_j a_{ij} \quad i = 1, \dots, n \quad j = 1, \dots, m$$

und deshalb sind die $n(i, j)$'s Poisson-verteilte ZV'en, die unabhängig voneinander sind und den Erwartungswert:

$$\begin{aligned} En(i, j) &= E(n_j \underbrace{a_{ij}}_{\text{Konstante}}) \\ &= a_{ij} \underbrace{En_j}_{=f_j} \\ &= a_{ij} f_j \quad i = 1, \dots, n \quad j = 1, \dots, m \end{aligned} \quad (5.4)$$

haben.

Damit können wir die gemessenen Daten X_i , $i = 1, \dots, n$, als

$$\begin{aligned} X_i &= n(i, \cdot) = \sum_{j=1}^m n(i, j) \\ &= \text{Gesamtzahl der Emissionen, die in } i \text{ detektiert wurden} \quad i = 1, \dots, n \end{aligned}$$

ausdrücken.

Als Summe unabhängiger Poisson-verteilter ZV'en sind nun auch alle X_i , $i = 1, \dots, n$, unabhängig und Poisson-verteilt und wir betrachten den Datenvektor $g = (g_1, \dots, g_n)^\top$ als die Realisation der ZV'en

$$X = (X_1, \dots, X_n).$$

Mit Hilfe der Rechenregeln für den Erwartungswert erhalten wir dann:

$$\begin{aligned} EX_i &= E\left(\sum_{j=1}^m n(i, j)\right) \\ &= \sum_{j=1}^m En(i, j) \\ &\stackrel{(5.4)}{=} \sum_{j=1}^m f_j \cdot a_{ij} \\ &= (Af)_i. \end{aligned}$$

Wenn wir weiter

$$\begin{aligned} n(\cdot, j) &= \sum_{i=1}^n n(i, j) \\ &= \text{Gesamtanzahl der Emissionen, die in Voxel (Pixel) } j \text{ stattgefunden haben} \\ &\quad \text{und detektiert wurden} \quad j = 1, \dots, m \end{aligned}$$

setzen, dann gilt:

$$\begin{aligned} En(\cdot, j) &= E \left(\sum_{i=1}^n n(i, j) \right) \\ &= \sum_{i=1}^n En(i, j) \\ &\stackrel{(5.4)}{=} f_j \underbrace{\sum_{i=1}^n a_{ij}}_{=a_{.j}}. \end{aligned}$$

Da $a_{.j}$ bekannt ist, sehen wir, dass das PET-Rekonstruktionsproblem äquivalent zu dem Problem ist, den Erwartungswert von $n(\cdot, j)$, $j = 1 \dots, m$, auf der Basis von $n(i, \cdot) = X_i$, $i = 1, \dots, n$ zu schätzen.

Wir haben also gesehen, dass für die unabhängigen ZV'en X_i , $i = 1, \dots, n$ gilt:

$$X_i \sim \text{Poisson}\{EX_i = (Af)_i, \text{Var } X_i = (Af)_i\}$$

und somit folgt:

$$X \sim \text{Poisson}\{EX = Af, \text{Var } X = Af\}.$$

Das Maximum-Likelihood-Prinzip zur Schätzung des unbekanntem Parameters f der gemeinsamen Verteilungen der Stichprobenvariablen X_1, \dots, X_n besagt nun Folgendes: Man wähle zum Stichprobenergebnis $g = (g_1, \dots, g_n)^\top$ denjenigen Wert \hat{f} als Schätzwert für f , unter dem die Wahrscheinlichkeit für das Eintreten dieses Ereignisses am größten ist [27]. Das führt uns zu den folgender zwei Definitionen Sätzen, die aus [25] entnommen wurden:

Definition 5.2. (Likelihood-Funktion)

Es sei X ein n -dimensionale Zufallsvariable und die X_i , $i = 1, \dots, n$ unabhängig. Die Wahrscheinlichkeit $p(g | f)$ der Beobachtungen (Realisationen) $g = (g_1, \dots, g_n)^\top$ ist eine Abbildung $L(\cdot | g) : \Theta \rightarrow [0, 1]$ mit

$$L(f | g_1, \dots, g_n) = \prod_{i=1}^n p(g_i | f). \tag{5.5}$$

Diese Abbildung heisst **Likelihood-Funktion** (LF), wobei Θ die Menge aller als möglich erachteten f -Werte ist.

Ziel: Finde einen Schätzer \hat{f} von f mit $L(\hat{f} | g_1, \dots, g_n) = \max_f L(f | g_1, \dots, g_n)$.

Definition 5.3. (Maximum-Likelihood-Schätzwert)

Der Schätzer \hat{f} heisst **Maximum-Likelihood-Schätzwert** von f bei gegebenen Daten g falls gilt:

$$L(\hat{f} | g_1, \dots, g_n) = \max_{f \in \Theta} L(f | g_1, \dots, g_n). \quad (5.6)$$

Satz 5.4. Für $L(f) := L(f | g)$ gilt:

$$L(f) = \prod_{i=1}^n \frac{(Af)_i^{g_i}}{g_i!} e^{-(Af)_i}. \quad (5.7)$$

Beweis.

$$\begin{aligned} L(f) &= L(f | g_1, \dots, g_n) \\ &= \prod_{i=1}^n \underbrace{p(g_i | f)}_{= p(X_i=g_i|f)} \\ &\stackrel{X_i \text{ Poisson-verteilt}}{=} \prod_{i=1}^n e^{-(Af)_i} \frac{(Af)_i^{g_i}}{g_i!}. \end{aligned}$$

□

Bis hierhin haben wir jetzt also die Kategorien 1) - 3) aus Abschnitt 5.2 gewählt und die Vorbereitungen für die Wahl der Kostenfunktion getroffen. Nun werden wir noch einen Satz angeben, der uns im Folgenden die Arbeit erleichtern wird und dann werden wir einen zu den gewählten Komponenten passenden numerischen Algorithmus, den ML-EM herleiten.

Satz 5.5. (Logarithmierte Likelihood-Funktion (LLF))

Folgende Äquivalenz gilt:

$$\begin{aligned} L(\hat{f} | g_1, \dots, g_n) &= \max_{f \in \Theta} L(f | g_1, \dots, g_n) \\ \Leftrightarrow \ln L(\hat{f} | g_1, \dots, g_n) &= \max_{f \in \Theta} \ln L(f | g_1, \dots, g_n). \end{aligned}$$

Man nennt $\ln L(\cdot | g_1, \dots, g_n)$ die **logarithmierte Likelihood-Funktion** (LLF).

Beweis. Wir nutzen hier aus, dass der Logarithmus eine strikt monoton wachsende Funktion ist und daher für $L(f | g_1, \dots, g_n) > 0$ gilt:

$$L(\hat{f} | g_1, \dots, g_n) \geq L(f | g_1, \dots, g_n) \Leftrightarrow \ln L(\hat{f} | g_1, \dots, g_n) \geq \ln L(f | g_1, \dots, g_n).$$

□

Kostenfunktion und numerischer Algorithmus ML-EM

Die Herleitung des ML-EM, insbesondere die die dazu nötigen Lemmata und Sätze inklusive Beweise sind Ausführungen, die auf den Ideen von Natterer, [11], beruhen.

Wie wir in Satz 5.5 gesehen haben, entspricht das Maximum der LF dem Maximum der LLF, welches i. Allg. einfacher zu bestimmen ist. Deshalb werden wir ab jetzt nur noch die LLF

$$\begin{aligned} \ln L(f) &= \ln \prod_{i=1}^n e^{-(Af)_i} \frac{(Af)_i^{g_i}}{g_i!} \\ &= \sum_{i=1}^n \ln e^{-(Af)_i} + \ln (Af)_i^{g_i} - \ln g_i! \\ &= \sum_{i=1}^n (g_i \ln (Af)_i - (Af)_i) - \sum_{i=1}^n \ln g_i! \end{aligned}$$

betrachten.

Da man die Konstante beim Maximierungsproblem auch weglassen kann, genügt es auch die Funktion

$$l(f) := \sum_{i=1}^n (g_i \ln (Af)_i - (Af)_i) \quad (5.8)$$

zu maximieren. Mit (5.8) haben wir nun also die Wahl unserer **Kostenfunktion** getroffen!

Es ist weiterhin sinnvoll anzunehmen, dass $f_j \geq 0 \forall j = 1, \dots, m$ ist ("Nichtnegativität der Dichte"). Wir haben also folgendes Optimierungsproblem:

$$\max l(f) \text{ unter den Nebenbedingungen } h_j(f) := f_j \geq 0 \quad j = 1, \dots, m,$$

wobei $l, h_j : \mathbb{R}^m \rightarrow \mathbb{R}$ mit dem Euklidischen Raum \mathbb{R}^m sind, l die Kostenfunktion ist und h_j die Bedingungen sind (siehe Kapitel 2.4).

Weitere Annahmen: (um technische Schwierigkeiten zu vermeiden)

- $a_{ij} > 0 \quad \forall i, j,$
- $g_i > 0 \quad \forall i.$

Lemma 5.6. *Seien $a_{i,j}, g_i > 0$. Dann gilt:
 $l(f)$ ist konkav.*

5 Das D-D-Modell und darauf beruhende Bildrekonstruktionen

Beweis. z.z.: Die Hessematrix ist negativ semidefinit ($\stackrel{\text{Lemma 2.48}}{\Rightarrow} l$ ist konkav).

dazu: Die partielle Ableitung bzgl. f_j liefert uns:

$$\begin{aligned} \frac{\partial l(f)}{\partial f_j} &= \sum_{i=1}^n a_{i,j} \left(\frac{g_i}{(Af)_i} - 1 \right) \\ &= - \sum_{i=1}^n a_{i,j} + \sum_{i=1}^n \frac{a_{i,j} g_i}{\sum_{k=1}^m a_{ik} f_k}, \end{aligned}$$

und wenn wir noch einmal bzgl. f_l partiell ableiten, dann erhalten wir:

$$\begin{aligned} \underbrace{\frac{\partial^2 l(f)}{\partial f_j \partial f_l}}_{=: h_{jl}} &= 0 - \sum_{i=1}^n \frac{a_{ij} g_i}{(\sum_{k=1}^m a_{ik} f_k)^2} a_{il} \\ &= - \sum_{i=1}^n \frac{a_{ij} a_{il} g_i}{(\sum_{k=1}^m a_{ik} f_k)^2}. \end{aligned}$$

$H := (h_{jl})_{\substack{j=1, \dots, m \\ l=1, \dots, m}}$ ist die Hessematrix von $l(f)$ und für jeden Vektor $x \in \mathbb{R}^m \setminus \{0\}$ gilt:

$$\begin{aligned} \langle x, Hx \rangle &= \sum_{j=1}^m x_j \sum_{l=1}^m h_{jl} x_l \\ &= \sum_{j=1}^m \sum_{l=1}^m x_j x_l h_{jl} \\ &= \sum_{j=1}^m \sum_{l=1}^m x_j x_l \left(- \sum_{i=1}^n \frac{a_{ij} a_{il} g_i}{(\sum_{k=1}^m a_{ik} f_k)^2} \right) \\ &= - \sum_{i=1}^n g_i \frac{\sum_{j=1}^m \sum_{l=1}^m x_j x_l a_{ij} a_{il}}{(\sum_{k=1}^m a_{ik} f_k)^2} \\ &= - \sum_{i=1}^n g_i \frac{\sum_{j=1}^m x_j a_{ij} \sum_{l=1}^m x_l a_{il}}{(\sum_{k=1}^m a_{ik} f_k)^2} \\ &= - \sum_{i=1}^n \underbrace{g_i}_{>0} \underbrace{\frac{\left(\sum_{j=1}^m x_j a_{ij} \right)^2}{\left(\sum_{k=1}^m a_{ik} f_k \right)^2}}_{\geq 0} \\ &\leq 0. \end{aligned}$$

H ist also negativ semidefinit und damit ist l wegen Lemma 2.48 konkav. □

Da

$$h_j(\theta f + (1 - \theta)w) = \theta f_j + (1 - \theta)w_j = \theta h_j(f) + (1 - \theta)h_j(w)$$

folgt, dass h_j konkav und konvex ist.

Wenn also $f^* \in \mathbb{R}^m$ ist und den Kuhn-Tucker-Bedingungen genügt, dann folgt aus Satz 2.50, dass f^* optimal für unser Maximierungsproblem mit Nebenbedingungen ist.

Nach Satz 2.49 bedeutet f^* genügt den K-T-Bedingungen, dass:

$$1) \quad h_j(f^*) = f_j^* \geq 0 \quad j = 1, \dots, m \quad \text{und}$$

es $\exists \lambda_1, \dots, \lambda_m \geq 0$ so, dass

$$2) \quad \lambda_j h_j(f^*) = \lambda_j f_j^* = 0 \quad j = 1, \dots, m$$

und

$$3) \quad \nabla l(f^*) + \sum_{j=1}^m \lambda_j \underbrace{\nabla h_j(f^*)}_{\substack{=(0,0,\dots,0, \\ \underbrace{1}_{j\text{-te Stelle}}, 0,\dots,0)}} = 0.$$

Um zu zeigen, dass f^* optimal ist, müssen wir also zeigen, dass Folgendes gilt:

1. $f_j^* \geq 0 \quad \forall j = 1, \dots, m$ (dies folgt aus dem Algorithmus).
2. $\frac{\partial l}{\partial f_j}(f^*) = 0$, wenn $f_j^* > 0$ ist.
3. $\frac{\partial l}{\partial f_j}(f^*) \leq 0$, wenn $f_j^* = 0$ ist.

Da $\frac{\partial l(f)}{\partial f_i} = \sum_{l=1}^n a_{il} \left(\frac{g_i}{(Af)_i} - 1 \right)$ folgt:

$$\nabla l(f) = \begin{pmatrix} \sum_{i=1}^n a_{i1} \left(\frac{g_i}{(Af)_i} - 1 \right) \\ \sum_{i=1}^n a_{i2} \left(\frac{g_i}{(Af)_i} - 1 \right) \\ \vdots \\ \sum_{i=1}^n a_{im} \left(\frac{g_i}{(Af)_i} - 1 \right) \end{pmatrix} .$$

Im Folgenden sind alle arithmetischen Operationen zwischen Vektoren komponentenweise zu verstehen!

Dann folgt aus den vorangegangenen Überlegungen mit $1 = (1, \dots, 1)^\top \in \mathbb{R}^n$, $\frac{g}{(Af)} = \left(\frac{g_1}{(Af)_1}, \dots, \frac{g_n}{(Af)_n} \right)^\top$, dass jedes globale Maximum f von l mit $f \geq 0$ den K-T-Bedingungen genügt und somit erhalten wir als notwendige Bedingung:

$$f \nabla l(f) = f A^\top \left(\frac{g}{(Af)} - 1 \right) = 0. \tag{5.9}$$

Normieren wir A so, dass jede Spaltensumme 1 ist ($A^\top \mathbf{1} = \mathbf{1}$), so erhält man:

$$fA^\top \frac{g}{Af} = fA^\top \mathbf{1} = f. \quad (5.10)$$

Der **ML-EM Algorithmus** ist nun die einfachste iterative Methode um (5.10) zu lösen:

$$f^{k+1} = f^k A^\top \frac{g}{Af^k} \quad k = 0, 1, 2, \dots \quad (5.11)$$

Um die Konvergenz dieses Algorithmus zu beweisen, benötigen wir noch ein paar kurze Vorbemerkungen.

Satz 5.7. (Jensen'sche Ungleichung)

Sei I ein Intervall oder $I = \mathbb{R}$, $f : I \rightarrow \mathbb{R}$ konvex (konkav), $n \in \mathbb{N}$ und $x_1, \dots, x_n \in I$. Für $\lambda_1, \dots, \lambda_n$ gelte $\lambda_i > 0$ und $\sum_{i=1}^n \lambda_i = 1$. Dann gilt:

$$f\left(\sum_{i=1}^n \lambda_i x_i\right) \leq \sum_{i=1}^n \lambda_i f(x_i) \quad \text{bzw.} \quad f\left(\sum_{i=1}^n \lambda_i x_i\right) \geq \sum_{i=1}^n \lambda_i f(x_i). \quad (5.12)$$

Beweis. Der Beweis zu diesen Ungleichungen lässt sich ganz einfach mit Hilfe der vollständigen Induktion durchführen. □

Satz 5.8. (Kullback-Leibler-Distanz und ihre Eigenschaften)

Seien $x, y \in \mathbb{R}^n$, $x, y \geq 0$ und $t \log t = 0$ für $t = 0$. Dann heisst:

$$L(x, y) := \sum_{i=1}^n \left(x_i \log \left(\frac{x_i}{y_i} \right) + y_i - x_i \right) \quad (5.13)$$

die **Kullback-Leibler-Distanz** und es gilt:

- 1.) $L(x, y) \geq 0$,
- 2.) $L(x, y) = 0 \Leftrightarrow x = y$,
- 3.) $L(x, y) \rightarrow L(x, x) \Rightarrow y \rightarrow x$,

wobei "log" hier für jeden beliebigen Logarithmus steht.

Beweis. 1.) $L(x, y) = -(-L(x, y))$

$$\begin{aligned}
 &= - \left[\sum_{i=1}^n x_i \underbrace{\log \left(\frac{y_i}{x_i} \right)}_{\leq \frac{y_i}{x_i} - 1 \text{ da } \log x \leq x-1} - y_i + x_i \right] \\
 &\geq - \left[\sum_{i=1}^n x_i \left(\frac{y_i}{x_i} - 1 \right) - y_i + x_i \right] \\
 &= - \left[\sum_{i=1}^n y_i - x_i - y_i + x_i \right] \\
 &= 0.
 \end{aligned}$$

2.) " \Leftarrow " Sei $x = y \Rightarrow L(x, y) = \sum_{i=1}^n x_i \log \left(\frac{x_i}{x_i} \right) + x_i - x_i = 0$.

$$\begin{aligned}
 \text{"} \Rightarrow \text{"} \quad \text{Da} \quad &-x_i \log \left(\frac{y_i}{x_i} \right) + y_i - x_i \\
 &= -x_i \left(\log \left(\frac{y_i}{x_i} \right) + 1 \right) + y_i \\
 &= x_i \underbrace{\left(-1 - \log \left(\frac{y_i}{x_i} \right) \right)}_{\geq -\frac{y_i}{x_i}} + y_i \\
 &\geq x_i \left(-\frac{y_i}{x_i} \right) + y_i = 0 \quad \forall i = 1, \dots, n
 \end{aligned}$$

folgt: wenn $L(x, y) = 0$ dann gilt:

$$-x_i \log \left(\frac{y_i}{x_i} \right) + y_i - x_i = 0 \quad \forall i.$$

1. Fall: $x_i = 0$

$\Rightarrow y_i = 0$.

2. Fall: $x_i \neq 0$

$$\Rightarrow -x_i \log \left(\frac{y_i}{x_i} \right) + y_i - x_i = 0 \quad \parallel: (-x_i)$$

$$\Rightarrow \log \left(\frac{y_i}{x_i} \right) - \frac{y_i}{x_i} + 1 = 0.$$

Da die Funktion $f(x) = \log x - x + 1$ in dem Punkt $(1/f(1) = 0)$ ihren Hochpunkt hat und die Funktion auf \mathbb{R}^+ rechtsgekrümmt ist liegt die einzige Nullstelle der Funktion bei $x = 1$. Daraus folgt:

$$\log \left(\frac{y_i}{x_i} \right) - \frac{y_i}{x_i} + 1 = 0 \Leftrightarrow \frac{y_i}{x_i} = 1,$$

und daraus folgt, dass das eindeutige Minimum der Funktion $L(x, y)$ liegt bei $x = y$.

3.) z.z.: $y \not\rightarrow x \Rightarrow L(x, y) \not\rightarrow L(x, x) = 0$.

dazu: $y \not\rightarrow x$ bedeutet:

$$\begin{pmatrix} y_1^k \\ \vdots \\ y_n^k \end{pmatrix} \xrightarrow{k \rightarrow \infty} \begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_n \end{pmatrix} \quad \text{wobei mindestens ein } \bar{y}_i \neq x_i \text{ ist.}$$

Daraus folgt:

$$L(x, y^k) = \sum_{i=1}^n x_i \log \left(\frac{x_i}{y_i^k} \right) + y_i^k - x_i \xrightarrow{k \rightarrow \infty, L \text{ stetig}} L(x, \bar{y}) > 0 = L(x, x),$$

da mindestens ein $\bar{y}_i \neq x_i$ ist.

□

Lemma 5.9. Sei $(f^k)_k$ die Folge, die aus der Anwendung der ML-EM Methode (5.11) resultiert. Dann gilt:

$$l(f^{k+1}) \geq l(f^k) \quad k = 0, 1, 2, \dots \quad (5.14)$$

Beweis. Für $h, f \in \mathbb{R}^m$, $h, f > 0$ gilt ([11] folgend):

a) Wir haben:

$$\begin{aligned} \sum_{i=1}^n g_i \ln(Af)_i &= \sum_{i=1}^n g_i \ln(Af)_i \frac{1}{(Ah)_i} \underbrace{\left(\sum_{l=1}^m a_{il} h_l \right)}_{=(Ah)_i} \\ &= \sum_{i=1}^n \sum_{l=1}^m g_i \ln(Af)_i \frac{a_{il} h_l}{(Ah)_i} \\ &= \sum_{i=1}^n g_i \sum_{l=1}^m \frac{a_{il} h_l}{(Ah)_i} \ln(a_{il} f_l) - \frac{a_{il} h_l}{(Ah)_i} \ln(a_{il} f_l) + \frac{a_{il} h_l}{(Ah)_i} \ln(Af)_i \\ &= \sum_{i=1}^n g_i \sum_{l=1}^m \frac{a_{il} h_l}{(Ah)_i} \left[\ln(a_{il} f_l) - \ln \left(\frac{a_{il} f_l}{(Af)_i} \right) \right]. \end{aligned}$$

b) für $k > 0$ gilt:

$$\begin{aligned}
 \sum_{i=1}^n (Af^k)_i &= \sum_{j=1}^m a_{1j} f_j^k + \cdots + \sum_{j=1}^m a_{nj} f_j^k \\
 &= f_1^k \underbrace{\left(\sum_{i=1}^n a_{i1} \right)}_{=1, \text{ da } A^T \mathbf{1} = \mathbf{1}} + \cdots + f_m^k \underbrace{\left(\sum_{i=1}^n a_{im} \right)}_{=1} \\
 &= \sum_{l=1}^m f_l^k
 \end{aligned}$$

und

$$\begin{aligned}
 \sum_{l=1}^m f_l^k &\stackrel{(5.11)}{=} f_1^{k-1} \sum_{i=1}^n a_{i1} \frac{g_i}{(Af^{k-1})_i} + \cdots + f_m^{k-1} \sum_{i=1}^n a_{im} \frac{g_i}{(Af^{k-1})_i} \\
 &= \frac{g_1}{(Af^{k-1})_1} (Af^{k-1})_1 + \cdots + \frac{g_n}{(Af^{k-1})_n} (Af^{k-1})_n \\
 &= \sum_{i=1}^n g_i .
 \end{aligned}$$

Daraus folgt also:

$$\sum_{i=1}^n (Af^k)_i = \sum_{l=1}^m f_l^k = \sum_{i=1}^n g_i . \quad (5.15)$$

c) daher gilt für $k > 0$:

$$\begin{aligned}
 l(f^{k+1}) - l(f^k) &= \sum_{i=1}^n g_i \ln(Af^{k+1})_i - \sum_{i=1}^n (Af^{k+1})_i - \sum_{i=1}^n g_i \ln(Af^k)_i + \sum_{i=1}^n (Af^k)_i \\
 &\stackrel{b)}{=} \sum_{i=1}^n g_i \ln(Af^{k+1})_i - \sum_{i=1}^n g_i \ln(Af^k)_i .
 \end{aligned}$$

Wir benutzen a), wobei wir in der 1. Summe $h := f^k$, $f := f^{k+1}$ und in der

2. Summe $h := f^k$, $f := f^k$ gesetzt haben, und erhalten:

$$\begin{aligned}
 l(f^{k+1}) - l(f^k) &= \sum_{i=1}^n g_i \sum_{l=1}^m \frac{a_{il} f_l^k}{(A f^k)_i} \left[\ln(a_{il} f_l^{k+1}) - \ln \left(\frac{a_{il} f_l^{k+1}}{(A f^{k+1})_i} \right) \right] \\
 &\quad - \sum_{i=1}^n g_i \sum_{l=1}^m \frac{a_{il} f_l^k}{(A f^k)_i} \left[\ln(a_{il} f_l^k) - \ln \left(\frac{a_{il} f_l^k}{(A f^k)_i} \right) \right] \\
 &= \sum_{i=1}^n g_i \sum_{l=1}^m \frac{a_{il} f_l^k}{(A f^k)_i} [\ln(A f^{k+1})_i - \ln(A f^k)_i] \\
 &= \sum_{i=1}^n g_i \sum_{l=1}^m \frac{a_{il} f_l^k}{(A f^k)_i} \left[\ln \left(\frac{f_l^{k+1}}{f_l^k} \right) - \ln \left(\frac{f_l^{k+1} (A f^k)_i}{f_l^k (A f^{k+1})_i} \right) \right] \\
 &= \sum_{i=1}^n g_i \sum_{l=1}^m \frac{a_{il} f_l^k}{(A f^k)_i} \ln \left(\frac{f_l^{k+1}}{f_l^k} \right) \\
 &\quad - \sum_{i=1}^n g_i \sum_{l=1}^m \frac{a_{il} f_l^k}{(A f^k)_i} \ln \left(\frac{f_l^{k+1} (A f^k)_i}{f_l^k (A f^{k+1})_i} \right) \\
 &= \sum_{l=1}^m \ln \left(\frac{f_l^{k+1}}{f_l^k} \right) \underbrace{\sum_{i=1}^n \frac{g_i}{(A f^k)_i} a_{il} f_l^k}_{\stackrel{(5.11)}{=} f_l^{k+1}}} \\
 &\quad - \sum_{i=1}^n \underbrace{g_i}_{>0} \underbrace{\sum_{l=1}^m \frac{a_{il} f_l^k}{(A f^k)_i} \ln \left(\frac{f_l^{k+1} (A f^k)_i}{f_l^k (A f^{k+1})_i} \right)}_{\stackrel{\text{Satz 5.7}}{\leq} \ln \left(\sum_{l=1}^m \frac{a_{il} f_l^k f_l^{k+1} (A f^k)_i}{(A f^k)_i f_l^k (A f^{k+1})_i} \right)} \\
 &\geq \sum_{l=1}^m f_l^{k+1} \ln \left(\frac{f_l^{k+1}}{f_l^k} \right) - \sum_{i=1}^n g_i \underbrace{\ln \left(\frac{1}{(A f^{k+1})_i} \sum_{l=1}^m a_{il} f_l^{k+1} \right)}_{=1} \\
 &= \sum_{l=1}^m f_l^{k+1} \ln \left(\frac{f_l^{k+1}}{f_l^k} \right) \\
 &= \sum_{l=1}^m f_l^{k+1} \ln \left(\frac{f_l^{k+1}}{f_l^k} \right) + \underbrace{\sum_{l=1}^m f_l^k - \sum_{l=1}^m f_l^{k+1}}_{\stackrel{b)}{=} 0}
 \end{aligned}$$

$$= L(f^{k+1}, f^k)$$

$$\stackrel{\text{Satz 5.8}}{\geq} 0.$$

□

Lemma 5.10. Für jeden Häufungspunkt f^* der Folge $(f^k)_k$ gilt:

$$L(f^*, f^{k+1}) \leq L(f^*, f^k) \quad k = 0, 1, 2 \dots \quad (5.16)$$

Beweis. a) definiere x^l, y^l wie folgt:

$$x_i^l := \frac{a_{il}g_i/(Af^*)_i}{(A^\top(g/Af^*))_l},$$

$$y_i^l := \frac{a_{il}g_i/(Af^k)_i}{(A^\top(g/Af^k))_l} \quad \text{für } i = 1, \dots, n.$$

Daraus folgt:

$$\begin{aligned} \sum_{i=1}^n x_i^l &= \sum_{i=1}^n \frac{a_{il}g_i}{(Af^*)_i} \frac{1}{\left(A^\top \frac{g}{(Af^*)}\right)_l} \\ &= \frac{1}{\left(A^\top \frac{g}{(Af^*)}\right)_l} \underbrace{\sum_{i=1}^n \frac{a_{il}g_i}{(Af^*)_i}}_{=(A^\top \frac{g}{Af^*})_l} \\ &= 1. \end{aligned}$$

ebenso: $\sum_{i=1}^n y_i^l = 1.$

$$\text{b) } 0 \leq \sum_{l=1}^m \underbrace{f_l^*}_{>0} L(x^l, y^l)$$

$$\stackrel{\text{Def. von } L}{=} \sum_{l=1}^m f_l^* \sum_{i=1}^n x_i^l \ln \left(\frac{x_i^l}{y_i^l} \right) + y_i^l - x_i^l$$

$$\stackrel{\text{a)}}{=} \sum_{l=1}^m f_l^* \sum_{i=1}^n x_i^l \ln \left(\frac{x_i^l}{y_i^l} \right)$$

5 Das D-D-Modell und darauf beruhende Bildrekonstruktionen

$$\begin{aligned}
&= \sum_{l=1}^m f_l^* \sum_{i=1}^n \frac{a_{il}g_i/(Af^*)_i}{(A^\top(g/Af^*))_l} \ln \left(\frac{a_{il}g_i/(Af^*)_i}{(A^\top(g/Af^*))_l} \frac{(A^\top(g/Af^k))_l}{a_{il}g_i/(Af^k)_i} \right) \\
&= \sum_{l=1}^m f_l^* \sum_{i=1}^n \frac{a_{il}g_i/(Af^*)_i}{(A^\top(g/Af^*))_l} \ln \left(\frac{(Af^k)_i \left(A^\top \frac{g}{Af^k} \right)_l}{(Af^*)_i \left(A^\top \frac{g}{Af^*} \right)_l} \right) \\
&= \sum_{l=1}^m f_l^* \sum_{i=1}^n \frac{a_{il}g_i/(Af^*)_i}{(A^\top(g/Af^*))_l} \ln \left(\frac{\overbrace{f_l^k \left(A^\top \frac{g}{Af^k} \right)_l}^{=f_l^{k+1}} (Af^k)_i}{f_l^k \left(A^\top \frac{g}{Af^*} \right)_l (Af^*)_i} \right) \\
&= \sum_{l=1}^m f_l^* \sum_{i=1}^n \frac{a_{il}g_i/(Af^*)_i}{(A^\top(g/Af^*))_l} \ln \left(\frac{f_l^{k+1} (Af^k)_i f_l^*}{f_l^k \left(A^\top \frac{g}{Af^*} \right)_l (Af^*)_i f_l^*} \right) \\
&= \sum_{l=1}^m f_l^* \sum_{i=1}^n \frac{a_{il}g_i/(Af^*)_i}{(A^\top(g/Af^*))_l} \ln \left(\frac{f_l^* f_l^{k+1} (Af^k)_i}{f_l^k f_l^* (Af^*)_i} \right).
\end{aligned}$$

Da

$$f^* = f^* A^\top \left(\frac{g}{Af^*} \right)$$

folgt:

$$\left(A^\top \frac{g}{Af^*} \right)_l = 1 \quad \text{wenn } f_l^* > 0.$$

Somit gilt:

$$\begin{aligned}
0 &\leq \sum_{l=1}^m f_l^* \sum_{i=1}^n \frac{a_{il}g_i/(Af^*)_i}{(A^\top(g/Af^*))_l} \ln \left(\frac{f_l^{k+1} (Af^k)_i}{f_l^k (Af^*)_i} \right) \\
&= \sum_{l=1}^m f_l^* \sum_{i=1}^n a_{il} \frac{g_i}{(Af^*)_i} \left[\ln \left(\frac{(Af^k)_i}{(Af^*)_i} \right) + \ln \left(\frac{f_l^{k+1}}{f_l^k} \right) \right]
\end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=1}^n \frac{g_i}{(Af^*)_i} \ln \left(\frac{(Af^k)_i}{(Af^*)_i} \right) \underbrace{\sum_{l=1}^m a_{il} f_l^*}_{=(Af^*)_i} + \sum_{l=1}^m f_l^* \ln \left(\frac{f_l^{k+1}}{f_l^k} \right) \underbrace{\sum_{i=1}^n \frac{a_{il} g_i}{(Af^*)_i}}_{=(A^\top \frac{g}{Af^*})_l=1} \\
 &= \sum_{i=1}^n g_i \ln \left(\frac{(Af^k)_i}{(Af^*)_i} \right) + \sum_{l=1}^m f_l^* \ln \left(\frac{f_l^{k+1}}{f_l^k} \right) \\
 &= \sum_{i=1}^n g_i \ln(Af^k)_i - \sum_{i=1}^n g_i \ln(Af^*)_i + \sum_{l=1}^m f_l^* (\ln f_l^{k+1} - \ln f_l^k) \\
 &\quad - \underbrace{\sum_{l=1}^m f_l^{k+1} + \sum_{i=1}^n (Af^*)_i}_{\stackrel{\text{Lemma 5.9 b)}}{=} 0 \\
 &= l(f^k) - l(f^*) + L(f^*, f^k) - L(f^*, f^{k+1}).
 \end{aligned}$$

Damit erhält man:

$$L(f^*, f^{k+1}) \leq L(f^*, f^k) + \underbrace{l(f^k) - l(f^*)}_{\stackrel{(5.14)}{\leq 0}} \leq L(f^*, f^k).$$

□

Satz 5.11. (Konvergenz des EM-Algorithmus)

Sei $f^0 > 0$. Dann gilt:

Die Folge $(f^k)_k$ konvergiert gegen f^* , welches $l(f)$ maximiert, d. h. die Folge $(f^k)_k$ konvergiert gegen den Wert, der l maximiert.

Beweis. [11] und [17] folgend gilt:

z.z.:

1. $(f^k)_k$ konvergiert gegen einen eindeutigen Grenzwert $(f^k \xrightarrow{k \rightarrow \infty} f^*)$.
2. $f^k \xrightarrow{k \rightarrow \infty} f^* \Rightarrow f^*$ maximiert $l(f)$ ($\max_f l(f) = l(f^*)$).

dazu:

1. Wegen der Kompaktheit existiert eine konvergente Teilfolge $(f^{k_s})_s$ von $(f^k)_k$ mit:

$$f^{k_s} \xrightarrow{s \rightarrow \infty} f^*.$$

Da

$$L(x, y) = 0 \Leftrightarrow x = y$$

folgt mit Satz 5.8 2.):

$$L(f^*, f^{k_s}) \xrightarrow{s \rightarrow \infty, L \text{ stetig}} L(f^*, f^*) = 0.$$

Damit ergibt sich aus Lemma 5.10 (L nicht wachsend), dass:

$$L(f^*, f^k) \xrightarrow{k \rightarrow \infty} 0$$

und aus Satz 5.8 3.), dass:

$$f^k \xrightarrow{k \rightarrow \infty} f^*.$$

2. Wegen Satz 2.50 genügt es z.z.: f^* erfüllt die K-T-Bedingungen.

dazu:

a) Da $f^* = f^* A^\top \frac{g}{Af^*}$ folgt:

$$f_l^* \stackrel{(5.11)}{=} f_l^* \sum_{i=1}^n a_{il} \frac{g_i}{(Af^*)_i}.$$

Wir erhalten mit $A^\top 1 = 1$:

$$f_l^* \underbrace{\sum_{i=1}^n a_{il} \left(\frac{g_i}{(Af^*)_i} - 1 \right)}_{= \frac{\partial l}{\partial f_l}(f^*)} = 0$$

und daraus ergibt sich:

$$\frac{\partial l}{\partial f_l}(f^*) = 0 \quad \text{für } f_l^* > 0.$$

b) Sei $f_l^* = 0$ und definiere $x_l^k := \left(A^\top \frac{g}{Af^k} \right)_l$.

Es gilt:

$$\underbrace{f_l^0}_{>0} \underbrace{\left(A^\top \frac{g}{Af^0} \right)_l}_{>0} \cdot \underbrace{\left(A^\top \frac{g}{Af^1} \right)_l}_{>0} \cdot \dots \cdot \underbrace{\left(A^\top \frac{g}{Af^k} \right)_l}_{=x_l^k} = f_l^{k+1} \xrightarrow{k \rightarrow \infty} f_l^* = 0.$$

Daraus folgt, dass

$$x_l^k \xrightarrow{k \rightarrow \infty} \left(A^\top \frac{g}{Af^*} \right)_l = 0,$$

und damit erhalten wir:

$$\begin{aligned} \frac{\partial l}{\partial f_i}(f^*) &= \sum_{i=1}^n a_{il} \left(\frac{g_i}{(Af^*)_i} - 1 \right) \\ &= \underbrace{\sum_{i=1}^n a_{il} \frac{g_i}{(Af^*)_i}}_{=0} - \sum_{i=1}^n \underbrace{a_{il}}_{>0} \\ &\leq 0 \quad \text{für } f_i^* = 0. \end{aligned}$$

□

Nachteile des ML-EM und ein Ausweg (OSEM)

Dieser Abschnitt ist aus [11], [21] und [28] entstanden.

Seit über einem Jahrzehnt ist bekannt, dass der ML-EM Algorithmus Bilder guter Qualität im 2D produzieren kann, aber wegen verschiedener Probleme nicht benutzt wird. Trotzdem wird er von vielen als der "goldene Standard" Rekonstruktionsalgorithmus angesehen.

Die drei Hauptprobleme, die den Gebrauch von ML-EM in kommerziell produzierten Scannern verhindern, sind:

- der große Speicherbedarf, welcher auf die Größe der Maximum-Likelihood-Matrix zurückzuführen ist,
- kein gutes Stoppkriterium (bedenklichstes Problem) und
- die Berechnungskomplexität einer einzelnen Iteration.

Wenn der relative Fehler klein genug ist, ist das natürliche Abbruchkriterium eines iterativen Algorithmus dafür bekannt, dass es verrauschte (schneeige) Bilder produziert. Zusammengefasst kann man also sagen, dass ML-EM einen hohen Speicherbedarf hat, berühmt-berüchtigt wegen seiner langsamen Konvergenz ist und es ihm an Glattheit mangelt. Viel Rechenzeit wird benötigt um eine akzeptable Lösung zu erzielen, da man für jeden neuen Schätzer einen kompletten Zyklus durch die gesamten Daten benötigt.

Lösungen:

- Die Konvergenz kann beschleunigt werden, in dem man A und g in Teilmatrizen A_j und Teilvektoren g_j für $j = 1, \dots, p$ aufteilt, so dass:

$$A = \begin{pmatrix} A_1 \\ \vdots \\ A_p \end{pmatrix} \quad \text{und} \quad g = \begin{pmatrix} g_1 \\ \vdots \\ g_p \end{pmatrix} \quad (5.17)$$

mit $A_j : \mathbb{R}^m \rightarrow \mathbb{R}^{n_j}$, $g_j \in \mathbb{R}^{n_j}$ ist, und (5.11) für jede Teilmatrix individuell durchgeführt wird.

Annahme: die A_j sind ausgeglichen (balanced) ($a = A_j^\top 1$ unabhängig von j .)

Damit erhalten wir:

$$\begin{aligned} f^{k,0} &= f^k \\ f^{k,j} &= f^{k,j-1} \frac{1}{a} A_j^\top \frac{g_j}{A_j f^{k,j-1}} \quad j = 1, \dots, p \\ f^{k+1} &= f^{k,p} . \end{aligned} \quad (5.18)$$

Dieser Algorithmus heißt **OSEM** (ordered subset EM or block iterative) und wurde von Hudson und Larkin 1994 hergeleitet.

Die Konvergenzeigenschaften von OSEM sind schwer festzustellen und würden den Rahmen dieser Arbeit sprengen.

Informationshalber geben wir nur ein paar kurze Bemerkungen dazu:

Wenn für alle Folgen

$$(f^{k,j}) \xrightarrow{k \rightarrow \infty} f^*$$

gelten würde, dann wäre f^* ein gemeinsamer Maximierer der LLF für jedes der Systeme

$$A_j f = g_j \quad j = 1, \dots, p.$$

Da diese Systeme bei konkreten Anwendungen jedoch wahrscheinlich nicht konsistent sind, kann man vernünftigerweise nicht erwarten, dass solch ein gemeinsamer Maximierer existiert. Deshalb ist es nicht wahrscheinlich, dass OSEM konvergiert, ausgenommen im konsistenten Fall.

Die Konvergenz im konsistenten Fall kann durch eine einfache Ausweitung des Beweises von Satz 5.11, unter der Voraussetzung, dass die A_j ausgeglichen (balanced) sind, bewiesen werden. [siehe auch [21], S. 608/609].

OSEM verarbeitet die Daten also in Teilmengen (oder Blöcken). Das Verfahren beschleunigt die Konvergenz um einen Faktor proportional zur Anzahl der Teilmengen. Tests ergaben, dass OSEM in einem Bruchteil der Zeit Bilder liefert, die in der Qualität ähnlich zu denen sind, die durch dem EM-Algorithmus produziert wurden.

Begründung, warum OSEM schneller ist:

Die Teilmengen bestehen normalerweise aus Projektionsansichten, die durch einen festen Winkel voneinander getrennt sind. Da in der PET alle Tube-Zählungen gleichzeitig gesammelt werden, kann OSEM am besten nach der vollständigen Sammlung der Zählungen angewendet werden. Die Koinzidenzen können dann in parallele Familien rebinnt werden, welche Projektionsansichten definieren und unser Algorithmus kann dann darauf angewendet werden.

Die Ordnung, in der die Projektionen verarbeitet werden ist willkürlich, obwohl es für die Qualität der gelieferten Rekonstruktion von Vorteil sein könnte, eine spezielle Ordnung zu wählen. Beispielsweise könnten wir bedeutende neue Informationen, die so schnell wie möglich eingebracht werden sollen, dadurch begünstigen, dass wir zuerst die Projektion, die der Richtung der größten Veränderlichkeit im Bild entspricht, auswählen und danach eine zweite Projektion, die senkrecht zur ersten ist und eine dritte und vierte Projektion, die mittig zwischen diesen ist, aussuchen, etc.

Normalerweise bestehen die Teilmengen aus zwei Mengen paralleler Projektionsansichten, die durch einen festen Winkel getrennt sind. Z. B.:

$$Y_{\Theta_i} := \{g_{i1}, \dots, g_{im_i}\}$$

wobei Y_{Θ_i} die Menge ist, die aus der Anzahl der parallelen Emissionen besteht, die in Detektorpaar $1, \dots, m_i$ verzeichnet werden. Die Teilmengen A_1, \dots, A_p könnten dann folgendermaßen aussehen (fester Winkel ist $\Pi/2$):

$$A_1 := \{Y_0, Y_{\Pi/2}\}, A_2 := \{Y_{\Pi/4}, Y_{3\Pi/4}\}, \text{ usw.}$$

Allgemeiner:

$$A_1 := \{Y_{\Theta_1}, Y_{\Theta_2}\}, A_2 := \{Y_{\Theta_3}, Y_{\Theta_4}\}, \dots$$

wobei $c := \Theta_1 - \Theta_2 = \Theta_3 - \Theta_4 = \dots$ konstant ist. In dieser Struktur ist ML-EM (kurz: EM) ein Spezialfall von OSEM ($p = 1$).

Für den Fall rauschfreier Projektionen wurde gezeigt, dass jede OSEM-Schätzung, die auf einer Teilmenge von Projektionen basiert, soweit gegen eine Maximum-Likelihood-Lösung konvergiert, wie eine volle Iteration von EM, die alle Projektionen benutzt.

In anderen Worten: Wenn die Projektionsdaten in p -Teilmengen unterteilt sind, dann wurden einmal alle Projektionen in einer einzigen Iteration von OSEM benutzt und ein Schätzwert wurde erhalten, der ähnlich zu dem ist, der durch p Iterationen von EM produziert wurde.

Diese Eigenschaft verleiht OSEM eine beachtliche Beschleunigung der Konvergenz, verglichen mit EM.

I. Allg. gilt: je höher der Level der Unterteilungen der Projektionsdaten ist, umso größer ist die erreichte Beschleunigung.

Es wurde jedoch auch gezeigt, dass sich die Resultate nach einer gewissen Anzahl von Teilmengen verschlechtern. Es wird jedoch angenommen, dass 16 Teilmengen ausreichend sind, um OSEM eine sechzehnfache Beschleunigung, verglichen mit dem EM-Algorithmus, zu geben.

Nachteile von OSEM:

Wenden wir OSEM auf reale, verrauschte Daten an, so hat er die Tendenz, zwischen mehreren Lösungen zu kreisen und nicht gegen das Maximum der LF zu

konvergieren. Obwohl es scheint, dass dieser Mangel in der Praxis nur einen geringen Einfluss hat, gibt es einen Weg, um ihn zu überwinden. Browne et al (1996) [29] schlugen eine Struktur vor, wodurch die OSEM-Schrittgröße kontrolliert wird und mit jeder "äußeren" Iteration neu berechnet wird. Die Autoren behaupten, dass dieser Algorithmus, der als Kombination von ART (siehe Abschnitt 5.3.2) und OSEM angesehen werden kann, von der Ausführung her besser als Standard EM oder OSEM ist.

Ihr Algorithmus, "**row action maximum likelihood algorithm**", **Ramla** genannt, kann wie folgt interpretiert werden:

$$f^{neu-Ramla} = (1 - \Theta)f^{alt} + \Theta f^{neu-OSEM} \quad 0 \leq \Theta \leq 1,$$

wobei Θ die Schrittgröße ($\Theta = 1$: Standard OSEM) und $f^{neu-OSEM}$ das Resultat der Anwendung einer Iteration des Algorithmus (indem man einen signifikant höheren Likelihood erreicht, als mehrere Iterationen von ML-EM) ist.

- Nun wollen wir uns dem Problem des Mangels an Glattheit zuwenden. Wir betrachten hier nur den EM-Algorithmus. Die Ausweitung auf OSEM ist dann offensichtlich.

Die einfachste Möglichkeit besteht nun darin, die Iteration früher zu stoppen. Eine andere Möglichkeit ist die, nach jeder Iteration einen Glättungsschritt durchzuführen (**EMS-Algorithmus** von Silvermann, Jones und Nachka 1990, siehe [30]).

Letztendlich kann man aber auch einen Strafterm " $-B(f)$ " zu $l(f)$ dazuzugaddieren, und maximiert dann " $l(f) - B(f)$ " anstelle von $l(f)$. Wir interpretieren $B(f)$ als **Bayesian** Struktur.

Normalerweise ist:

$$B(f) := (f - \bar{f})^\top B(f - \bar{f}),$$

wobei \bar{f} das Referenzbild und B eine positiv definite Matrix ist.

$$\max l(f) - B(f) \text{ unter der Nebenbedingung } f \geq 0$$

führt dann zu:

$$f = fA^\top \frac{g}{Af} - f\nabla B(f) \tag{5.19}$$

(mit der Kostenfunktion " $l(f) - B(f)$ " und denselben Überlegungen wie zuvor beim EM-Algorithmus).

Es werden dabei in der Literatur verschiedene Methoden zur Lösung von (5.19) vorgeschlagen. Im Folgenden wollen wir nun einige Beispiele kurz darstellen:

- **OSL-Algorithmus** von Green 1990, [31]:

$$f^{k+1} = f^k A^\top \frac{g}{A f^k} - f^{k+1} \nabla B(f^k)$$

oder

$$f^{k+1} = \frac{f^k}{1 + \nabla B(f^k)} A^\top \frac{g}{Af^k}.$$

Dieser Algorithmus bewahrt die Positivität, ebenso wie der Original EM-Algorithmus.

- **SOR Newton-Methode** von Setzepfandt 1992, [32]:

$$f^{k+1} = f^k A^\top \frac{g}{Af^k} - f^{k+1} \nabla B(f^{k+1}).$$

Anmerkung 5.12. Wenn B eine Diagonalmatrix ist, dann ist diese Gleichung von 2. Ordnung in f^{k+1} und kann eindeutig gelöst werden. (siehe dazu Levitan und Herman 1987, [33]).

5.3.2 ART (Algebraic Reconstruction technique)

Auch hier folgen wir dem Buch von Natterer und Wübeling, [11] S. 110 ff.

ART ist einfach die wohlbekannte Kaczmarz-Methode (Projektion auf affine Unterräume) zur Lösung über- oder unterbestimmter linearer Systeme.

Wie wir später noch sehen werden, ist der OSEM-Algorithmus nur eine multiplikative Version der Kaczmarz-Methode für die Lösung $Af = g$. Wir werden also im Folgenden ART vorstellen und die Konvergenz für eine allgemeinere Form von ART beweisen. Nachdem wir diese Überlegungen getroffen haben, werden wir im nächsten Abschnitt die Konvergenzgeschwindigkeiten von ART und OSEM miteinander vergleichen.

Der numerische Algorithmus ART für vollständige Daten

Sei $R_j f = g_j$ ein System mit:

$$R_j : H \rightarrow H_j \quad j = 1, \dots, p$$

beschränkte, lineare Operatoren, also $R_j \in \mathcal{L}(H, H_j)$ mit $H, H_j, j = 1, \dots, p$ Hilberträume.

Da wir von vollständigen Daten ausgehen, ist R_j eine Surjektion für $j = 1, \dots, p$.

Sei

$$R = \begin{pmatrix} R_1 \\ \vdots \\ R_p \end{pmatrix}, \quad g = \begin{pmatrix} g_1 \\ \vdots \\ g_p \end{pmatrix} \quad \text{und} \quad Rf = g$$

wobei g die gegebenen Daten und $f \in H$ gesucht ist.

Sei $P_j : H \rightarrow M$ die Projektion in den affinen Unterraum $R_j f = g_j$.

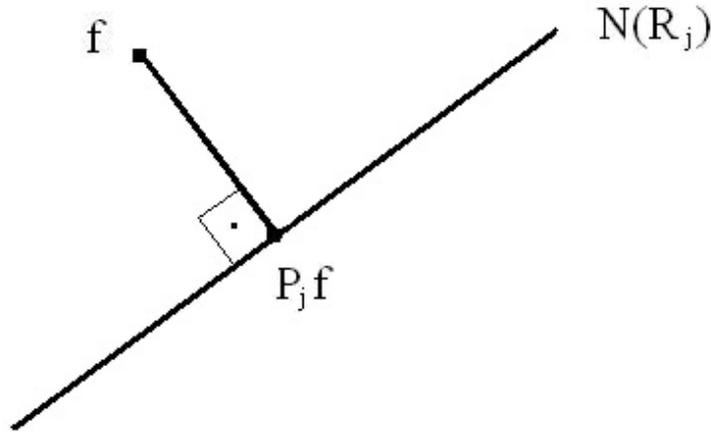


Abbildung 5.1: Projektion in den affinen Unterraum $R_j f = g_j$.

Für $f \in H = \underbrace{N(R_j)}_{=:M} \oplus \underbrace{\overline{R(R_j^*)}}_{=:M^\perp} \stackrel{\text{Satz 2.14}}{=} N(R_j) \oplus R(R_j^*)$ gilt somit:

$$H \ni f = \underbrace{f_1}_{\in M} + \underbrace{f_2}_{\in M^\perp} \quad \text{eindeutige Zerlegung}$$

$$\Rightarrow \exists u_j \in H_j \text{ mit } f = \underbrace{P_j f}_{\in M} + \underbrace{R_j^* u_j}_{\in M^\perp}$$

$$\Rightarrow R_j f = \underbrace{R_j P_j f}_{= g_j} + R_j R_j^* u_j$$

da $P_j f$ im affinen UR

$$\Rightarrow R_j f - g_j = R_j R_j^* u_j .$$

Da R_j surjektiv, also $R(R_j) = H_j$ ist, folgt mit Satz 2.11 und Korollar 2.4:

$$N(R_j^*) = R(R_j)^\perp = H_j^\perp = \{0\}.$$

Der adjungierte Operator R_j^* ist also injektiv, und da

$$R(R_j^*) \subset \overline{R(R_j^*)} \subset N(R_j)^\perp$$

ist, R_j jedoch auf $N(R_j)^\perp$ injektiv ist, erhalten wir, dass $R_j R_j^*$ als Hintereinanderausführung injektiver Operatoren injektiv ist.

Damit gilt:

$$u_j = (R_j R_j^*)^{-1} (R_j f - g_j)$$

und daraus folgt wiederum:

$$\begin{aligned} P_j f &= f - R_j^* u_j \\ &= f - R_j^* (R_j R_j^*)^{-1} (R_j f - g_j) \\ &= f + R_j^* (R_j R_j^*)^{-1} (g_j - R_j f). \end{aligned}$$

Dieses Ergebnis fassen wir in dem folgenden Satz zusammen:

Satz 5.13. *Die orthogonale Projektion P_j von H in den affinen Unterraum $R_j f = g_j$ lautet:*

$$P_j f = f + R_j^* (R_j R_j^*)^{-1} (g_j - R_j f). \quad (5.20)$$

Andere Formen des Verfahrens von Kaczmarz:

Setzen wir für $\omega > 0$

$$P_j^\omega = (1 - \omega)I + \omega P_j, \quad P^\omega = P_p^\omega \cdot \dots \cdot P_1^\omega,$$

dann lautet das Kaczmarz-Verfahren mit Relaxationsfaktor ω zur Lösung von $Rf = g$

$$f^{k+1} = P^\omega f^k,$$

genauer:

$$\begin{aligned} f^{k,0} &= f^k \\ f^{k,j} &= P_j^\omega f^{k,j-1} \\ &= (1 - \omega)I f^{k,j-1} + \omega P_j f^{k,j-1} \end{aligned} \quad (5.21)$$

$$= f^{k,j-1} + \omega R_j^* (R_j R_j^*)^{-1} (g_j - R_j f^{k,j-1}) \quad j = 1, \dots, p$$

$$f^{k+1} = f^{k,p}. \quad (5.22)$$

Für $\omega = 1$ und ein lineares System $Rf = g$, das aus p skalaren Gleichungen in n Unbekannten besteht, wurde diese Methode von Kaczmarz 1937 vorgeschlagen.

Setzt man:

$$\begin{aligned} f^{k,j} &= P_j f^{k,j-1} = f^{k,j-1} + R_j^* u_j \quad j = 1, \dots, p \\ &= \dots = f^{k,0} + \sum_{l=1}^j R_l^* u_l \end{aligned} \quad (5.23)$$

in die Iterationsvorschrift (5.21) ein, so erhält man:

$$\underbrace{f^{k,0} + \sum_{l=1}^j R_l^* u_l}_{f^{k,j}} = \underbrace{f^{k,0} + \sum_{l=1}^{j-1} R_l^* u_l}_{f^{k,j-1}} + \omega R_j^* (R_j R_j^*)^{-1} \left(g_j - R_j f^{k,0} - \sum_{l=1}^{j-1} R_j R_l^* u_l \right)$$

und daraus ergibt sich:

$$R_j^* u_j = \omega R_j^* (R_j R_j^*)^{-1} \left(g_j - R_j f^{k,0} - \sum_{l=1}^{j-1} R_j R_l^* u_l \right).$$

Da R_j^* injektiv ist, kann man es auch weglassen und es folgt:

$$u_j = \omega (R_j R_j^*)^{-1} \left(g_j - R_j f^{k,0} - \sum_{l=1}^{j-1} R_j R_l^* u_l \right).$$

Damit erhalten wir:

$$(R_j R_j^*) u_j = \omega \left(g_j - R_j f^{k,0} - \sum_{l=1}^{j-1} R_j R_l^* u_l \right).$$

In der Matrixschreibweise bedeutet dies:

$$u = \begin{pmatrix} u_1 \\ \vdots \\ u_p \end{pmatrix}, \quad RR^* = \begin{pmatrix} R_1 R_1^* & \cdots & R_1 R_p^* \\ \vdots & & \vdots \\ R_p R_1^* & \cdots & R_p R_p^* \end{pmatrix} = D + L + L^*$$

mit

$$D = \begin{pmatrix} R_1 R_1^* & & 0 \\ & \ddots & \\ 0 & & R_p R_p^* \end{pmatrix}, \quad L = \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ R_2 R_1^* & \ddots & & \\ \vdots & & \ddots & \\ R_p R_1^* & \cdots & R_p R_{p-1}^* & 0 \end{pmatrix}.$$

Es gilt also:

$$Du = \omega (g - Rf^{k,0} - Lu),$$

und wenn wir diese Gleichung nach u auflösen, dann ergibt sich:

$$u = (D + \omega L)^{-1} (g - Rf^{k,0}).$$

Somit gilt für unsere Iteration:

$$\begin{aligned} f^{k+1} &= f^{k,p} \stackrel{(5.23)}{=} f^{k,0} + \sum_{l=1}^p R_l^* u_l \\ &= f^{k,0} + R^* u \\ &\stackrel{u \text{ eingesetzt}}{=} f^{k,0} + \omega R^* (D + \omega L)^{-1} (g - Rf^{k,0}) \\ &= \underbrace{(I - \omega R^* (D + \omega L)^{-1} R)}_{=: B_\omega} f^{k,0} + \underbrace{\omega R^* (D + \omega L)^{-1} g}_{=: b_\omega} \end{aligned}$$

und daraus folgt:

$$f^{k+1} = B_\omega \underbrace{f^{k,0}}_{=: f^k} + b_\omega = B_\omega f^k + b_\omega.$$

Bei der Anwendung von ART hätten wir einen guten Startwert für $f^0 \in R(R^*)$, z. B. $f^0 = 0$ oder $f^0 = R^*g$.

Den Konvergenzbeweis führen wir nun für eine etwas allgemeinere Methode als (5.21) durch. Dabei ersetzen wir den Operator $R_j R_j^*$ in der Formel (5.21) durch einen willkürlichen positiv definiten Operator C_j in H_j .

Somit erhalten wir:

$$\begin{aligned} f^{k,0} &= f^k \\ f^{k,j} &= f^{k,j-1} + \omega R_j^* C_j^{-1} (g_j - R_j f^{k,j-1}) \quad j = 1, \dots, p \\ f^{k+1} &= f^{k,p}. \end{aligned} \tag{5.24}$$

Spezialfall: $p = 1$, $C_1 = I$, **Landweber Iteration** (Hanke, Neubauer und Scherzer 1995, [34])

Wir betrachten nun den konsistenten Fall, bei dem $Rf = g$ eine Lösung hat. Um den Satz über die Konvergenz von f^k beweisen zu können, benötigen wir zuerst noch folgenden Hilfssatz.

Hilfssatz 5.14. *Sei Q ein selbstadjungierter Operator in einem Hilbertraum mit $\|Q\| \leq 1$ und $\|Qf\| < \|f\|$ für $f \neq 0$. Dann gilt:*

$$Q^k \xrightarrow{k \rightarrow \infty} 0.$$

Beweis. Sei $\epsilon_k := \| Q^k f \|$. Da jede beschränkte und monotone Folge reeller Zahlen konvergiert und

- $0 \leq \epsilon_k \leq \epsilon_0 \quad \Rightarrow (\epsilon_k)_k$ beschränkt
- $\epsilon_0 \geq \epsilon_1 \geq \epsilon_2 \geq \dots \quad \Rightarrow (\epsilon_k)_k$ nicht wachsend

konvergiert $(\epsilon_k)_k$ mit $\lim_{k \rightarrow \infty} \epsilon_k =: \epsilon$.

Da auch alle Teilfolgen gegen ϵ konvergieren folgt:

$$\lim_{k \rightarrow \infty} \| Q^{2k} f \| = \epsilon$$

und

$$\lim_{k \rightarrow \infty} \| Q^{2k+1} f \| = \epsilon.$$

Da

$$\begin{aligned} \| Q^{2k} f - Q^{2k+2m} f \|^2 &= \| Q^{2k} f \|^2 + \| Q^{2k+2m} f \|^2 - 2 \langle Q^{2k+m} f, Q^{2k+m} f \rangle \\ &= \underbrace{\epsilon_{2k}^2}_{\rightarrow \epsilon} + \underbrace{\epsilon_{2k+2m}^2}_{\rightarrow \epsilon} - 2 \underbrace{\epsilon_{2k+m}^2}_{\rightarrow \epsilon} \\ &\rightarrow 0 \quad \forall m \end{aligned}$$

folgt: $(Q^{2k} f)_k$ ist eine Cauchy-Folge.

Da jeder Hilbertraum vollständig ist, konvergiert $(Q^{2k} f)_k$ und es gilt:

$$\lim_{k \rightarrow \infty} Q^{2k} f =: g.$$

Wegen der Stetigkeit der Norm und des Operators Q erhalten wir

$$\lim_{k \rightarrow \infty} \| Q^{2k} f \| = \| g \|$$

und

$$\lim_{k \rightarrow \infty} \| Q^{2k+1} f \| = \lim_{k \rightarrow \infty} \| Q Q^{2k} f \| = \| Qg \|.$$

Das bedeutet:

$$\begin{aligned} \epsilon &= \lim_{k \rightarrow \infty} \| Q^{2k} f \| = \| g \| \quad \text{und} \\ \epsilon &= \lim_{k \rightarrow \infty} \| Q^{2k+1} f \| = \| Qg \|. \end{aligned}$$

Also:

$$\| Qg \| = \| g \|.$$

Da laut Voraussetzung $\| Qg \| < \| g \|$ für $g \neq 0$ ist, folgt:

$$g = 0,$$

und daraus ergibt sich mit Hilfe der Normeigenschaften, dass

$$\epsilon = 0$$

ist. Weiterhin gilt dann:

$$\|Q^k\| := \sup_{f \neq 0} \frac{\overbrace{\|Q^k f\|}^{\rightarrow \epsilon = 0}}{\underbrace{\|f\|}_{\neq 0}} \xrightarrow{k \rightarrow \infty} 0,$$

und daraus folgt dann die Behauptung:

$$Q^k \xrightarrow{k \rightarrow \infty} 0.$$

□

Satz 5.15. *Sei C_j beschränkt und positiv definit und $0 < R_j R_j^* \leq C_j$, $j = 1, \dots, p$. Sei $Rf = g$ konsistent.*

Dann konvergiert die Iteration (5.24) und

$$f^k \rightarrow P_R f^0 + R^+ g, \quad (5.25)$$

wobei P_R die orthogonale Projektion auf $N(R)$ und R^+ die verallgemeinerte Inverse (siehe auch Definition 3.3) ist.

Beweis. Sei f^+ die Minimum-Norm-Lösung von $Rf = g$, also $Rf^+ = g$ (da konsistent) und $f^+ \in N(R)^\perp$ nach Satz 3.4. Sei $e^k := f^k - f^+$. Daraus folgt:

$$e^{k+1} = Qe^k = \dots = Q^{k+1}e^0 \quad \text{mit} \quad Q = Q_p \cdot \dots \cdot Q_1 \quad \text{und} \quad (5.26)$$

$$Q_j = I - \omega R_j^* C_j^{-1} R_j \quad j = 1, \dots, p. \quad (5.27)$$

Daraus erkennen wir:

$$Q = I \text{ auf } N(R).$$

z.z.:

1. $Q^k \xrightarrow{k \rightarrow \infty} 0$ auf $N(R)^\perp$.
2. $f^k \xrightarrow{k \rightarrow \infty} P_R f^0 + R^+ g$.

dazu:

- 1.

$$\begin{aligned} \|Q_j f\|^2 &= \|f\|^2 - 2\omega \langle f, R_j^* C_j^{-1} R_j f \rangle + \omega^2 \langle R_j^* C_j^{-1} R_j f, R_j^* C_j^{-1} R_j f \rangle \\ &= \|f\|^2 - 2\omega \langle R_j f, C_j^{-1} R_j f \rangle + \omega^2 \langle R_j R_j^* C_j^{-1} R_j f, C_j^{-1} R_j f \rangle. \end{aligned}$$

Da $R_j R_j^* \leq C_j$ ist folgt:

$$\langle R_j R_j^* C_j^{-1} R_j f, C_j^{-1} R_j f \rangle \leq \langle C_j C_j^{-1} R_j f, C_j^{-1} R_j f \rangle$$

und somit gilt:

$$\begin{aligned} \| Q_j f \|^2 &\leq \| f \|^2 - 2\omega \langle R_j f, C_j^{-1} R_j f \rangle + \omega^2 \langle R_j f, C_j^{-1} R_j f \rangle \\ &= \| f \|^2 - \omega(2 - \omega) \langle R_j f, C_j^{-1} R_j f \rangle. \end{aligned} \quad (5.28)$$

Da C_j^{-1} positiv definit ist gilt:

$$\langle R_j f, C_j^{-1} R_j f \rangle = 0 \Leftrightarrow R_j f = 0 \Leftrightarrow f \in N(R_j).$$

Für $0 < \omega < 2$ gilt damit:

$$\| Q_j f \| \leq \| f \|$$

und

$$f \in N(R_j) \Leftrightarrow \| Q_j f \| = \| f \|.$$

Damit erhalten wir:

$$\| Q f \| \leq \| f \| \quad \text{und} \quad \| Q f \| = \| f \| \Leftrightarrow f \in N(R).$$

Für $0 \neq f \in N(R)^\perp$ gilt somit: $\| Q f \| < \| f \|$.

1. Fall: $N(R)^\perp$ endlich dimensional

$$\Rightarrow Q \text{ ist eine Kontraktion auf } N(R)^\perp$$

$$\Rightarrow Q^k \xrightarrow{k \rightarrow \infty} 0 \text{ auf } N(R)^\perp.$$

2. Fall: $N(R)^\perp$ unendlich dimensional.

Da

$$\begin{aligned} (I - \omega R_j^* C_j^{-1} R_j)^* &= I^* - \omega R_j^* \underbrace{(C_j^*)^{-1}}_{=C_j^{-1}} R_j \\ &= I - \omega R_j^* C_j^{-1} R_j \\ &= Q_j \end{aligned}$$

folgt, dass Q selbstadjungiert ist und da

$$\| Q \| := \sup_{f \neq 0} \underbrace{\frac{\| Q f \|}{\| f \|}}_{\leq 1} \leq 1$$

und

$$\| Q f \| < \| f \| \quad \text{für } 0 \neq f \in N(R)^\perp \text{ (Hilbertraum),}$$

sind also alle Voraussetzungen des Hilfssatz 5.14 erfüllt und somit erhalten wir:

$$Q^k \xrightarrow{k \rightarrow \infty} 0 \text{ auf } N(R)^\perp.$$

2. Um den Beweis nun zu beenden, schreiben wir:

$$\begin{aligned} e^0 = f^0 - f^+ &= P_R f^0 + I f^0 - P_R f^0 - f^+ \\ &= P_R f^0 + (I - P_R) f^0 - f^+. \end{aligned}$$

Da $Q = I$ auf $N(R)$ und

$$\begin{aligned} f^k - f^+ &= e^k \stackrel{(5.26)}{=} Q^k e^0 \\ &= Q^k (P_R f^0 + (I - P_R) f^0 - f^+) \\ &= \underbrace{Q^k P_R f^0}_{= P_R f^0} + Q^k ((I - P_R) f^0 - f^+) \\ &\quad \text{da } P_R f^0 \text{ orthogonale Proj.} \\ &\quad \text{auf } N(R) \text{ und } Q = I \text{ auf } N(R) \\ &= P_R f^0 + Q^k \underbrace{\left(\underbrace{(I - P_R) f^0}_{\in N(R)^\perp} - \underbrace{f^+}_{\in N(R)^\perp} \right)}_{Q^{k \rightarrow \infty} \text{ auf } N(R)^\perp} \\ &\xrightarrow{k \rightarrow \infty} P_R f^0 \end{aligned}$$

folgt:

$$f^k \xrightarrow{k \rightarrow \infty} P_R f^0 + f^+.$$

□

Ein Nebenprodukt des obigen Beweises ist das folgende Korollar:

Korollar 5.16. *Es seien dieselben Voraussetzungen gegeben wie in Satz 5.15. Dann gilt:*

$$\| f^{k,j} - f \|^2 \leq \| f^{k,j-1} - f \|^2 - \omega(2 - \omega) \langle R_j f^{k,j-1} g_j, C_j^{-1} (R_j f^{k,j-1} - g_j) \rangle$$

mit $f = P_R f^0 + R^+ g$.

Beweis.

$$\begin{aligned} \| f^{k,j} - f \|^2 &= \| f^{k,j-1} + \omega R_j^* C_j^{-1} (g_j - R_j f^{k,j-1}) - f \|^2 \\ &= \| Q_j f^{k,j-1} + \omega R_j^* C_j^{-1} \underbrace{g_j}_{= R_j f} - f \|^2 \\ &= \| Q_j f^{k,j-1} - Q_j f \|^2 = \| Q_j (f^{k,j-1} - f) \|^2 \\ &\stackrel{(5.28)}{\leq} \| f^{k,j-1} - f \|^2 - \omega(2 - \omega) \langle R_j (f^{k,j-1} - f), C_j^{-1} R_j (f^{k,j-1} - f) \rangle \\ &= \| f^{k,j-1} - f \|^2 - \omega(2 - \omega) \langle R_j f^{k,j-1} - g_j, C_j^{-1} (R_j f^{k,j-1} - g_j) \rangle. \end{aligned}$$

□

Der Fehler wird also kleiner, solange $R_j f^{k,j-1} - g_j$ nicht verschwindet.

5.3.3 Zusammenhang zwischen OSEM und ART

Die Ideen für diesen Abschnitt entstammen ebenfalls wieder aus [11] S. 123 ff.

Wenn wir nun - wie gefordert - das lineare inverse Problem $Af = g$ lösen wollen, wobei A als bekannt vorausgesetzt wird (siehe Seite 144), und dieses Problem wie bei OSEM in (5.17) in das System:

$$A_j f = g_j \quad j = 1, \dots, p$$

mit den linearen Abbildungen:

$$A_j : \mathbb{R}^m \rightarrow \mathbb{R}^{n_j}, \quad g_j \in \mathbb{R}^{n_j}$$

zerlegt wird, so können wir zur Lösung des Systems nicht nur (5.18), sondern auch ART verwenden. Dazu versehen wir das ganze mit den Skalarprodukten:

$$(x, y)_f := \left\langle \frac{a}{f} x, y \right\rangle, \quad (x, y)_{g_j} := \left\langle \frac{1}{g_j} x, y \right\rangle,$$

wobei $\langle \cdot, \cdot \rangle$ das natürliche Skalarprodukt auf dem \mathbb{R}^m bzw. \mathbb{R}^{n_j} ist. Dann gilt:

$$\begin{aligned} (A_j \tilde{f}, h)_{g_j} &= \left\langle \frac{1}{g_j} A_j \tilde{f}, h \right\rangle_{\mathbb{R}^{n_j}} = \left\langle A_j \tilde{f}, \frac{1}{g_j} h \right\rangle_{\mathbb{R}^{n_j}} \\ &= \left\langle \tilde{f}, A_j^\top \frac{1}{g_j} h \right\rangle_{\mathbb{R}^m} \\ &= \left\langle \frac{a}{f} \tilde{f}, A_j^\top \frac{1}{g_j} h \right\rangle_{\mathbb{R}^m} \\ &= \left(\frac{f}{a} \tilde{f}, A_j^\top \frac{1}{g_j} h \right)_f = \left(\tilde{f}, \frac{f}{a} A_j^\top \frac{1}{g_j} h \right)_f \end{aligned}$$

und somit haben wir:

$$A_j^* = \frac{f}{a} A_j^\top \frac{1}{g_j}. \quad (5.29)$$

Wir erkennen also, dass OSEM nur eine multiplikative Version der Kaczmarz-Methode für die Lösung des inversen Problems $Af = g$ ist.

Hier noch einmal beide Algorithmen für obiges System auf einen Blick:

OSEM:

$$\begin{aligned} f^{k,0} &= f^k \\ f^{k,j} &= f^{k,j-1} \frac{1}{a} A_j^\top \frac{g_j}{A_j f^{k,j-1}} \quad j = 1, \dots, p \\ f^{k+1} &= f^{k,p} \end{aligned}$$

ART:

$$\begin{aligned} f^{k,0} &= f^k \\ f^{k,j} &= f^{k,j-1} + \omega A_j^* (A_j A_j^*)^{-1} (g_j - A_j f^{k,j-1}) \quad j = 1, \dots, p \\ f^{k+1} &= f^{k,p} \end{aligned}$$

mit A_j^* wie in (5.29).

Vergleich der Konvergenzgeschwindigkeiten von OSEM und ART

- **OSEM**: Führen wir auch hier, wie von Tanaka 1987 in [35] vorgeschlagen wurde, einen Relaxationsparameter ω ein und setzen:

$$G_j(f) := f \left(\frac{1}{a} A_j^\top \frac{g}{A_j f} \right)^\omega \quad j = 1, \dots, p,$$

dann gilt für den OSEM-Algorithmus mit Relaxationsparameter ω :

$$f^{k+1} = G^\omega(f^k), \quad G^\omega = G_p^\omega \circ \dots \circ G_1^\omega, \quad (5.30)$$

wobei "o" Komposition bedeutet.

Für $\omega = 1$ erhalten wir den Standard OSEM-Algorithmus.

Für $p = 1$ erhalten wir den Standard EM mit Relaxation:

$$G^\omega = G_1^\omega \Rightarrow f^{k+1} = G^\omega(f^k) = G_1^\omega(f^k) = \underbrace{f^k \left(\frac{1}{a} A_1^\top \frac{g}{A_1 f^k} \right)^\omega}_{\text{relaxierter EM}}.$$

Wenn wir Konvergenz annehmen, dann wird die Konvergenzgeschwindigkeit durch die Jakobi-Matrix $(G^\omega)'(f)$ für festes f von (5.30) bestimmt.

Wir treffen also folgende **Annahmen**:

- Konvergenz,
- $Af = g$.

Damit erhalten wir:

$$G'(f) = \prod_{j=1}^p \left(I - \omega \frac{f}{a} A_j^\top \frac{1}{g_j} A_j \right). \quad (5.31)$$

- ART: Wir setzen:

$$Q_j(f) := f + \omega A_j^* C_j^{-1} (g_j - A_j f) \quad j = 1, \dots, p,$$

dann gilt für ART:

$$f^{k+1} = Q(f^k), \quad Q = Q_p \circ \dots \circ Q_1, \quad (5.32)$$

wobei "o" Komposition bedeutet. Damit erhalten wir:

$$\begin{aligned} Q'(f) &= \prod_{j=1}^p (I - \omega A_j^* C_j A_j) \\ &\stackrel{(5.29)}{=} \prod_{j=1}^p \left(I - \omega \frac{f}{a} A_j^T \frac{1}{g_j} C_j A_j \right). \end{aligned} \quad (5.33)$$

Für $C_j = I$ gilt:

$$G'(f) = Q'(f)$$

und wir können nun schließen, dass die Konvergenzgeschwindigkeiten von OSEM und ART dieselben sind, und dass im Großen und Ganzen die Konvergenzanalyse von ART aus [11] S. 110 ff., angewendet auf OSEM, auch dieselbe ist.

Da eine ausführliche Konvergenzanalyse den Rahmen dieser Arbeit sprengen würde und unter der angegebenen Quelle [11] auch gut nachzulesen ist, wollen wir im Folgenden nur die Ergebnisse in einer Bemerkung kurz darstellen.

Bemerkung 5.17. • *OSEM ist schneller als EM in demselben Sinne wie ART schneller als Landweber ist.*

- *Für relaxierten OSEM sind sehr kleine Relaxationsfaktoren vorteilhaft.*
- *Auch die Konvergenzgeschwindigkeit von OSEM hängt von der Anordnung der Gleichungen ab (genau auf dieselbe Art und Weise wie bei ART).*
- *Für geeignete Anordnungen der Gleichungen und geeignete Relaxationsparameter konvergiert ART sehr schnell [28].*
- *Herman et al (1995) [36] führten mehrere Experimente mit verschiedenen Relaxationsparametern durch und folgerten schließlich, dass die beste Methode, eine Menge von Parametern zu finden, experimentell ist. Um ART zu optimieren sollte man zu jeder Iteration den besten ω -Parameter zum Fortsetzen herausfinden. Wenn eine geeignete Reihe von Regularisierungsparametern ω^k gewählt wurden, wird eine Rekonstruktion von guter Qualität schon nach 2 bis 3 Iterationen des ART-Algorithmus erreicht [28].*

5.4 Zusammenfassung

Gemäß [20] können wir Folgendes festhalten:

Den Nutzen, den wir von iterativen Algorithmen erwarten, besteht in einer Verbesserung der Auflösung, die durch eine genauere Modellierung der physikalischen Messungen erreicht wird. Abgesehen von dem System-Modell, welches zu der Geometrie des Datenerwerbs in Beziehung steht, sind iterative Algorithmen unabhängig von der 2D oder 3D Natur des Problems. Die Wahl und die Durchführung eines Algorithmus werden dennoch durch die Eigenschaften des 3D Problems beeinflusst. Die Haupteigenschaft ist die riesige Datenmenge: gewöhnlich hat man über 10 Millionen LOR's. Man könnte meinen, dass die große Dimension des Problems zu inakzeptablen langen Rekonstruktionszeiten führt und in der Tat benötigen viele Methoden, so wie der EM-Algorithmus, der bei jeder Iteration die gesamte Datenmenge miteinbezieht, für jede Iteration annähernd dieselbe Zeit wie die gesamte 3D FBP. Die Forschung zu modellbasierten Bildrekonstruktionsmethoden steht also vor zwei Problemen:

1. Wie wählen wir eine Kostenfunktion, die Bilder mit den gewünschten Eigenschaften produzieren soll, und
2. wie finden wir diese Bilder möglichst schnell.

In der 1. Kategorie wird an der statistischen und physikalischen Modellierung der Daten, der Wahl von Bildglättungstermen, die die Lösung regularisieren sollen und der Wahl von Kostenfunktionen, die über dem Bildraum optimiert werden müssen, gearbeitet.

In der 2. Kategorie geht es darum, möglichst schnell eine Lösung zu finden, sobald eine Kostenfunktion ausgewählt wurde. Prinzipiell sind die Lösungen des konkaven Maximierungsproblems, die uns üblicherweise in der Bildrekonstruktion begegnen, unabhängig von dem numerischen Algorithmus, der ausgewählt wurde, um die Lösung zu finden. In der Praxis jedoch werden die schnellen Algorithmen oft vor der Konvergenz abgebrochen, so dass die Lösung eine Funktion des Algorithmus wird. Trotzdem ist es nützlich, den Unterschied zwischen den Klassen von Algorithmen zu behalten, die angeblich dieselbe Lösung berechnen und denen, die verschiedene "Kostenkriterien" optimieren und deshalb in verschiedenen Lösungen resultieren.

Wir werden uns im Folgenden nun hauptsächlich mit der Wahl des iterativen Algorithmus befassen und die Probleme, die die Kostenfunktionswahl betreffen außer Acht lassen.

Wie wir gesehen haben war Shepp and Vardi's ML-Algorithmus, der auf EM (expectation maximization) basiert, unter den ersten, die explizit die Poisson-Verteilung in photonbeschränkten Bildgebungssystemen bei der Modellierung miteinbezogen haben. Die EM-Formulierung dieses Problems führte zu einer Gleichung, die an den frühen multiplikativen ART-Algorithmus erinnert.

Die Verbesserungen in der Bildqualität, die der ML-EM Algorithmus lieferte, hatten eine enorme nachfolgenden Forschungstätigkeit zur Folge. Viele dieser Arbeiten wendeten sich den Problemen der langsamen Konvergenz und den Instabilitäten, die sich durch

die hohe Varianz der geschätzten Voxelintensität erklären lassen, zu.

Das erste Problem ließ sich durch die Einführung blockiterativer Versionen (OSEM) lösen. OSEM kann weder auf einzelne Blöcke angewendet werden, noch konvergiert er für beliebig gewählte Blöcke. Er ist jedoch in vielen Fällen von praktischem Interesse. Er lässt sich so gut wie ML-EM durchführen, jedoch mit einer geringeren Anzahl an Iterationen. Manglos et al bemerkten in [37] jedoch, dass verrauschte Rekonstruktionen entstehen können, wenn zuviele Teilmengen bei OSEM benutzt werden. Sowohl bei OSEM als auch bei ART kann die sorgsame Auswahl der Teilmengenanordnungen die Konvergenz außerordentlich beschleunigen. Der Mangel an Konvergenz von OSEM ist in gewissem Maße nur von theoretischem Interesse und Wichtigkeit, da der Algorithmus in der Praxis eh nach nur ein paar Iterationen abgebrochen wird.

Um die "limit cycles", die bei verrauschten Daten entstehen zu entfernen, wurde von Browne and De Pierro der Algorithmus RAMLA eingeführt (Modifikation von OSEM mit starker unter-Relaxation).

Das zweite Problem kann in der Praxis durch frühzeitige Beendigung des Algorithmus oder durch Glättung der Bilder nach der Rekonstruktion kontrolliert werden. Ein weiterer Ansatz zur Kontrollierung der Instabilitäten besteht in der Regularisierung der Lösung (Maximum a posteriori (MAP) Bayesian oder penalized ML). Dabei wird ein Glättungsmaß zu der LLF hinzugefügt, um übertriebenes Rauschen in den Bildern "zu bestrafen". Die MAP-Lösung kann durch Verallgemeinerungen des ML-EM Algorithmus berechnet werden.

Eine Kurzzusammenfassung über weitere Verbesserungsmethoden findet der interessierte Leser in [20].

6 Ausblick

Zuerst einmal soll in diesem Kapitel auf die Dinge hingewiesen werden, die nur kurz, bzw. gar nicht in den vorangegangenen Kapiteln besprochen wurden, und die durchaus noch einer genaueren Ausführung bedürften, jedoch den Rahmen dieser Arbeit überschritten hätten.

Zunächst einmal zählt darunter die Verfeinerung der Modelle aus Kapitel 3 und 5. Die Modelle werden genauer und besser, wenn wir die gestreuten und zufälligen Koinzidenzen noch mitbetrachten und unsere Überlegungen nocheinmal mit diesen neuen, verbesserten Modellen durchführen würden. (Siehe dazu auch Lewitt, [10] S. 4 und 6.)

Ich möchte auch darauf hinweisen, dass der Inversionsansatz in Kapitel 4 auf dem einfachsten C-C-Modell der ET basiert. (Zur Erinnerung: Die Daten entsprechen Linienintegralen über der gesuchten Funktion). Inversion ist auch für ein realistischeres C-C-Modell der ET möglich, welches die Dämpfung der Photonen zwischen Emissionspunkt und Detektoren miteinbezieht. Dieses Modell soll im Folgenden kurz skizziert werden. Eine genaue Herleitung dieses Modells kann in der Staatsexamensarbeit von Frau Elke Weidig [38] nachgelesen werden.

$$\begin{aligned}g(x_r, y_r, \phi, \theta) &= \int_L f(x, y, z) e^{-\int_{L_+(x,y,z)} \mu(l) dl} e^{-\int_{L_-(x,y,z)} \mu(l) dl} dz_r \\ &= e^{-\int_L \mu(l) dl} g^{LSI}(x_r, y_r, \phi, \theta).\end{aligned}$$

L_+ und L_- bezeichnen die zwei Abschnitte von L zwischen dem Punkt (x, y, z) und den zwei entgegengesetzten Detektoren.

$\mu(l)$ ist der lineare Dämpfungskoeffizient im Punkt l entlang der Linie L .

$e^{-\int_L \mu(l) dl}$ gibt die Wahrscheinlichkeit an, dass keine Wechselwirkungen mit dem dämpfenden Gewebe für Photonen, die entlang des LOR's L emittiert werden, bestehen.

Das obige C-C-Modell der PET enthält nur Dämpfungseffekte, jedoch nichts zu Streuungs- und Zufallskoinzidenzen und könnte auch noch bzgl. dieser Faktoren ausgeweitet und verbessert werden.

Des Weiteren ergeben sich auch noch ein paar interessante Fragestellungen zur mathematischen Behandlung des Effektes der Dämpfung. Anhand des obigen Modells erkennen wir, dass die Detektionswahrscheinlichkeit für einen Zerfall irgendwo entlang der Linie nicht vom Ort des Zerfalls entlang der Linie, sondern nur von dem Integral über die Dämpfung über die gesamte Linie abhängt. Inversionsmethoden, die für $g(x_r, y_r, \phi, \theta)$ entwickelt wurden, können weiterhin benutzt werden, nachdem man eine Vorverarbeitung der Daten, um den Effekt der Dämpfung zu korrigieren, vorgenommen hat. Man benötigt also ein Bild von der räumlichen Verteilung der Dämpfungskoeffizienten, bevor man die Emissionsverteilung rekonstruiert. Wie schon in Abschnitt 1.3.2 erwähnt wurde, stehen uns zwei Methoden zur Dämpfungskorrektur zur Verfügung: die gemessene und die berechnete Korrektur.

Bei der letzteren Dämpfungskorrektur nehmen wir an, dass die Außenkontur bekannt ist und, dass innerhalb dieser Kontur die Dämpfungseigenschaften konstant sind (z. B. keine Lunge oder keine wesentlichen Knochen). Die Aussenkontur kann automatisch aus den Daten bestimmt werden oder definiert werden, indem ein Bild ohne Dämpfungskorrektur benutzt wird.

Bei einer gemessenen Dämpfungskorrektur wird ein zusätzlicher Scan durchgeführt. Man benutzt eine externe Quelle radioaktiver Strahlung, die die Photonen durch den Körper des Patienten zu den Detektoren sendet.

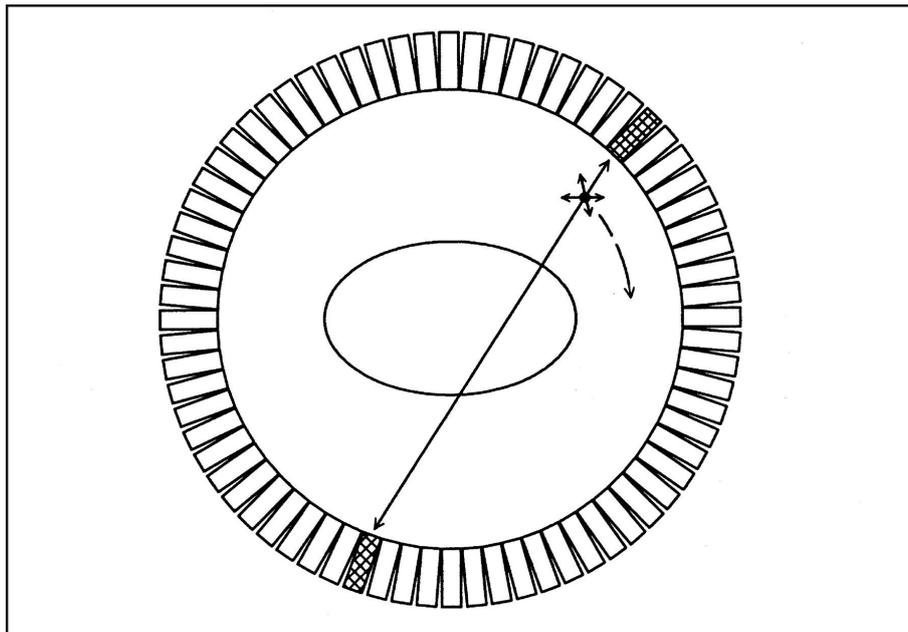


Abbildung 6.1: Rotierende Quelle für den Transmissions-Scan. Quelle: [8] Figure 7.

Diese sog. Transmissions-Daten werden zu den Daten ins Verhältnis gesetzt, die entstehen, wenn kein Patient im Scanner ist. Dieses Verhältnis wird für jede mögliche Koinzidenz einzeln bestimmt und liefert eine Schätzung für die Wahrscheinlichkeit, dass keine Wechselwirkungen mit dem dämpfenden Gewebe für Photonen, die entlang des entspre-

chenden LOR's emittiert werden, bestehen (z. B. entspricht dies $e^{-\int_L \mu(l) dl}$). Im Prinzip könnte diese Schätzung direkt in der PET benutzt werden, da der Kehrwert des obigen Verhältnisses eine Schätzung des Dämpfungskorrekturfaktors für die Daten ergibt. Das Problem hierbei liegt jedoch darin, dass der größte Korrekturfaktor aus den am wenigsten verfügbaren Zählungen resultiert und dies zu sehr verrauschten Schätzungen führt. Ein möglicher Ansatz zur Behebung dieses Problems besteht darin, den Logarithmus von jedem dieser Verhältnisse zu nehmen, um Schätzungen des Linienintegrals der Dämpfungskoeffizientenverteilung (wie in der CT) zu erhalten. Dann rekonstruiert man die Verteilung aus diesen Daten unter Verwendung eines CT-Algorithmus für die Transmissionsgeometrie.

Einige kürzlich entwickelte Scanner kombinieren einen CT-Scanner mit einem ET-Scanner, um kombinierte CT-PET-Systeme zu bilden. Das Resultat der Rekonstruktion aus den Transmissionsdaten kann geglättet werden oder in diskrete Regionen zerlegt werden, um eine sogenannte "attenuation map" zu entwickeln. Unter Verwendung dieser "attenuation map" können die Integrale für komplette Linien berechnet werden, entweder durch Integration in das D-D-Modell oder durch Vorverarbeitung des Emissionsdaten.

Dieser Ansatz ist gut, wenn die Transmissionsdaten eine große Anzahl an Zählungen haben. Das Problem hierbei ist jedoch, dass die meisten PET-Scanner vergleichsweise schwache radioaktive Quellen benutzen, um die Transmissionsdaten zu erhalten. Daraus ergibt sich dann auch nur eine geringe Anzahl an Zählungen. In diesem Fall ist es alles andere als optimal, den Logarithmus der geringzähligen Daten zu nehmen, um Schätzungen des Linienintegrals zu erhalten, da solche Poissondaten nicht symmetrisch um den Erwartungswert sind und Messungen beinhalten, die Null Zählungen haben. Außerdem beinhalten die Daten in der Praxis zusätzliche Beiträge von Streuung und Emissionsaktivität, welche oft im Patienten zur selben Zeit, in der auch der Transmissions-Scan durchgeführt wird, präsent sind. Schätzungen dieser Beiträge werden oft von den Daten subtrahiert, was dann zu negativen Zählungen und somit zu einem Problem bei der Anwendung des Logarithmus führen kann.

Für geringzählige Daten ist es deshalb ratsam, ein statistisches Modell des Messprozesses zu formulieren und einen Algorithmus herzuleiten, der direkt auf den Original-Daten operiert, um die "attenuation map" zu erhalten.

Eine genaue Ausführung eines solchen Modells ist z. B. in dem Artikel von "C. Y. Bai, P. E. Kinahan, D. Brasse, C. Comtat, and D. W. Townsend, Postinjection single photon transmission tomography with ordered-subset algorithms for whole-body PET imaging, IEEE Transactions on Nuclear Science, Volume 49, Number 1, pp. 74-81, 2002", nachzulesen.

Insgesamt können wir festhalten, dass alle Methoden, die Transmissionsdaten verarbeiten, zu einer Vermehrung von Rauschen und Messfehlern führen.

Das Rauschen und die Messfehler aus den Transmissionsdaten gelangen in die rekonstruierte "attenuation map" und von dort in die korrigierten Emissionsdaten oder die Systemmatrix der Emissionsrekonstruktion und schließlich in das rekonstruierte Bild der Emissionsaktivität.

Die Analyse der Vermehrung von Rauschen und Fehlern aus den Transmissionsdaten ist ein für die Zukunft herausforderndes Problem, ebenso die Weiterentwicklung der CT-PET-Systeme.

Auf dem Gebiet der iterativen Methoden wird zum gegenwärtigen Zeitpunkt sowie in naher Zukunft Forschungstätigkeit geleistet bzw. geleistet werden. Wie wir bereits gesehen haben, stellen iterative Methoden große Anforderungen an die Rechenzeit, besonders wenn Datenmengen benutzt werden, die von der aktuellen Scanner-Generation, die eine sehr hohe Auflösung haben, produziert werden. Die laufende Arbeit besteht nun darin, die bereits bekannten Algorithmen auf Klustern der genutzten PC-Prozessoren zu implementieren.

Wir wollen diese Arbeit nun mit der Bemerkung beenden, dass der OSEM-Ansatz kürzlich von Firmen, die nukleare medizinische Bildgebungssysteme herstellen, angenommen wurde. Nun, da die Türen für den Gebrauch iterativer Ansätze in klinischen Einrichtungen geöffnet wurden, wird es interessant sein zu sehen, wie die jeweiligen Leistungen von OSEM und anderen blockiterativen Ansätzen bzgl. klinisch relevanter Messungen in der Bildqualität und den Kosten sein werden.

Literaturverzeichnis

- [1] Rollo F.D. und Hines H., Positronen-Emissions-Tomographie (PET), Medica Mundi, Ausgabe 3, Februar 2002, Seite 36-39
- [2] Waxman A.D., PET: Funktionelle Bildgebung in der Onkologie, Medica Mundi, Ausgabe 4, Oktober 2002, Seite 12-18
- [3] Freudenberg L.S., Antoch G., Beyer T., Bockisch A., Debatin J., Erste klinische Erfahrungen mit einem kombinierten PET/CT-Tomographen an der Universitaet Essen, Electromedica 70, 2002, Heft 1, Seite 68-73
- [4] Broschüre der Firma Siemens zum Thema "Die Anwendungen der MR in der Medizin", Abschnitt: Vergleich der MR mit anderen bildgebenden Verfahren, S. 60
- [5] Schirrmeister H., Die PET in der Skelettdiagnostik, MTA Dialog 5 (2002) 3, Seite 374-377
- [6] Bendriem B., Townsend D.W., The Theory and Practice of 3D PET, Kluwer Academic Publishers, 1998
- [7] Fahey Frederic H., Data Acquisition in PET Imaging, Journal of Nuclear Medicine Technology, Volume 30, Number 2, 2002, pp. 39-49
- [8] Turkington Timothy G., Introduction to PET Instrumentation, Journal of Nuclear Medicine Technology, Volume 29, Number 1, 2001, pp.4-11
- [9] Threm K., Donsch P., SPECT und PET MTAR-Lehrbuch, Verlag Wissenschaftliche Scripten, 1998

- [10] Lewitt Robert M., Matej Samuel, Overview of Methods for Image Reconstruction from Projections in Emission Computed Tomography, Proceedings of the IEEE, Vol. 91, No. 9, September 2003, (Issue Number Subject to Change!!)
- [11] Natterer Frank, Wübbeling Frank, Mathematical Methods in Image Reconstruction, SIAM, Philadelphia, 2001
- [12] Natterer Frank, The Mathematics of Computerized Tomography, SIAM, Philadelphia, 2001
- [13] Louis Alfred K., Inverse und schlecht gestellte Probleme, Teubner, Stuttgart, 1989
- [14] Defrise Michel, Kinahan P. E., Townsend D. W., Exact and Approximate Rebinning Algorithms for 3-D PET Data, IEEE Transactions on Medical Imaging, Vol. 16, No. 2, April 1997, pp. 145-158
- [15] Liu Xuan, Defrise Michel, Michel Christian, Sibomana Merence, Comtat Claude, Kinahan Paul, Townsend David, Exact Rebinning Methods for Three-Dimensional PET, IEEE Transactions on Medical Imaging, Vol. 18, No. 8, August 1999, pp. 657-664
- [16] Defrise Michel, Liu Xuan, A fast rebinning algorithm for 3D positron emission tomography using John's equation, Inverse Problems, Vol. 15, No. , February 1999, pp. 1047-1065
- [17] Vardi Y., Shepp L. A., Kaufman L., A Statistical Model for Positron Emission Tomography, Journal of the American Statistical Association, Vol. 80, No. 389, March 1985, pp. 8-37
- [18] Vardi Y., Shepp L. A., Maximum Likelihood Reconstruction for Emission Tomography, IEEE Transactions on Medical Imaging, Vol. MI-1, No. 2, October 1982, pp. 113-122
- [19] Ollinger John M., Fessler Jeffrey A., Positron-Emission Tomography, IEEE Signal Processing Magazine, Vol. 14, No. 1, January 1997, pp. 43-55
- [20] Leahy R., Byrne C., Editorial Recent Developments in Iterative Image Reonstruction for PET and SPECT, IEEE Transactions on Medical Imaging, Vol. 19, No. 4,

April 2000, pp. 257-259

- [21] Hudson H. Malcolm, Larkin Richard S., Accelerated Image Reconstruction Using Ordered Subsets of Projection Data, IEEE Transactions on Medical Imaging, Vol. 13, No. 4, December 1994, pp. 601-609

- [22] Zangwill Willard, Nonlinear Programming, Prentice-Hall, 1969

- [23] Skript zur Vorlesung: Mathematische Methoden in der Medizintechnik, gehalten von: Prof. A. K. Louis, Universität des Saarlandes, WS 1999/2000

- [24] Skript zur Vorlesung: Inverse Probleme in der CT, gehalten von: Dr. Thomas Schuster, Universität des Saarlandes, SS 2002

- [25] Skript zur Vorlesung: Grundzüge der Statistik Teil A und Teil B, gehalten von: Prof. Volker Steinmetz, Universität des Saarlandes, WS 1997/1998 und SS 1998

- [26] Skript zur Vorlesung: Theoretische Statistik, gehalten von: Prof. Volker Steinmetz, Universität des Saarlandes, WS 1999/2000

- [27] Bamberg Günter, Baur Franz, Statistik, R. Oldenbourg, München, 1996

- [28] Barrett T. et al., A Review of Methods and Techniques used in 3D Positron Emission Tomography (PET) Image Reconstruction, submitted for publication to Medical Image Analysis, 1998

- [29] Browne J., Depierro A., Row-action alternative to the EM algorithm for maximizing likelihood in emission tomography, IEEE Transactions on Medical Imaging, Volume 15, Number 5, 1996, pp. 687-699

- [30] Silverman B.W., Jones M.C., Nychka D.W., A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography, J. Roy. Statist. Soc., B 52, 1990, pp. 271-324

- [31] Green P. J., On the use of the EM algorithm for penalized likelihood estimation, J. Roy. Statist. Soc., B 52, 1990, pp. 443-352

- [32] Setzepfandt B., ESNM: Ein rauschunterdrückendes EM-Verfahren für die Emissionstomographie, Dissertation, Fachbereich Mathematik und Informatik, Universität Münster, Münster Germany, 1992
- [33] Levitan E., Herman G. T., A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography, IEEE Transactions on Medical Imaging, Volume 6, 1987, pp. 185-192
- [34] Hanke M., Neubauer A., Scherzer O., A convergence analysis of Landweber iteration for nonlinear ill-posed problems, Numerische Mathematik, Volume 72, Number 1, November 1995, pp. 21-37
- [35] Tanaka E., A fast reconstruction algorithm for stationary positron emission tomography based on a relaxed EM algorithm, IEEE Transactions on Medical Imaging, Volume 6, 1987, pp. 98-105
- [36] Herman G., Image reconstruction from projections, Real Time Imaging, Volume 1, Issue 1, 1995, pp. 3-18
- [37] Manglos S. H., Gagne G. M., Krol A., Thomas F. D., Narayanaswami R., Transmission maximum-likelihood reconstruction with ordered subsets for cone beam CT, Physics of Medical Biology, Volume 40, 1995, pp. 1225-1241
- [38] Weidig Elke, Die gedämpfte Radontransformation mit Anwendung in der ET, Examensarbeit an der Universität des Saarlandes, Betreuer: Prof. A.K. Louis, Januar 2002
- [39] Skript zur Vorlesung: Funktionalanalysis I, gehalten von: Prof. Jörg Eschmeier, Universität des Saarlandes, WS 2000/2001
- [40] Werner Dirk, Funktionalanalysis, Springer-Lehrbuch, Springer, 1995
- [41] Liu Xuan, Michel Christian, Vollmar Stefan, Wienhard Klaus, Casey Mike, Defrise Michel, Evaluation of an exact Fourier rebinning algorithm for a large aperture PET scanner, <http://cfi.lbl.gov/3D-2001/abstracts/09-2.pdf>