On the application of the polychromatic statistical reconstruction technique to C-arm CT data

Dissertation

zur Erlangung des akademischen Grades

Doktoringenieur

(Dr.-Ing.)

von Dipl.-Phys. Richard Nikolaus Karl Bismark

geboren am 11. Dezember 1986 in Magdeburg

genehmigt durch die Fakultät für Elektrotechnik und Informationstechnik
der Otto-von-Guericke-Universität Magdeburg.

Gutachter:

Prof. Dr. rer. nat. Georg Rose
Prof. Dr.-Ing. habil. Andreas Maier

Tag der Einreichung am 22. Februar 2021
Promotionskolloquium verteidigt am 02. Juli 2021

Funding

This work is partly funded by the German Ministry of Education and Research (BMBF) within the Forschungscampus *STIMULATE* (13GW0095A).

Zusammenfassung

Computer Tomographie (CT) ist im klinischen Alltag der Krankenhäuser ein essentielles Mittel der Diagnostik und Nachbehandlung. Obwohl es bereits seit Jahrzehnten erfolgreich angewendet und verbessert wird, ist das standard Verfahren, dass aus den Röntgenbildern die 3D Information gewinnt störanfällig und vereinfachend. Auf der Hardware-Ebene wird das Potential der Ingenieurskunst bis ins kleinste Detail ausgereizt, doch auf der Software-Ebene findet die Innovation der Forschung nur langsam Anschluss.

Viele moderne Verfahren benötigen zusätzliche Technik oder alternative Messmethoden, doch auch mit den bereits genutzten Geräten und Messungen könnten bessere klinische CT Bilder erzeugt werden. Durch das stetig wachsende Potential der modernen Computer und deren Rechenleistung ist es heutzutage möglich, enorme Datenmengen zu verarbeiten und Ergebnisse zu ermitteln, für die es vor 20 Jahren noch einige Monate an Rechenzeit gebraucht hätte.

Diese zusätzlichen Rechenkapazitäten können genutzt werden, um die vereinfachten Verfahren durch komplexere Algorithmen zu ersetzen. Ein wesentliches Problem der aktuellen Standard-Methode ist die sogenannte Strahlaufhärtung (engl. beam hardening). Die erzeugte Röntgenstrahlung verfügt über ein Energiespektrum, das verschiedene quantemechanische Wechselwirkungen abdeckt. Die unterschiedlichen Wechselwirkungen tragen dabei alle zum gemessenen Signal bei. Es ist bekannt, dass der Absorptionsprozess der Röntgenstrahlung nicht-linear von dem absorbierenden Objekt abhängt. Eben diese Nicht-Linearität wird bei der standard Rekonstruktionstechnik nicht vollständig berücksichtigt und führt zu systematischen Bildfehlern (sog. Bildartefakten). Historisch war es schlicht nicht möglich, die Nicht-Linearitäten innerhalb eines klinisch vertretbaren Zeitraumes zu berücksichtigen. Nach 50 Jahren angewandter Computer Tomographie ist es mittlerweile allerdings sehr wohl möglich klinische CT Bilder innerhalb vertretbarer Rechenzeiten zu ermitteln. Hierfür bedarf es lediglich einer angemessenen Implementation der modernen Rekonstruktionsalgorithmen.

Diese Arbeit beschäftigt sich mit einem physikalisch motivierten Rekonstruktionsalgorithmus, der bereits 2002 entwickelt wurde aber seit dem nicht auf klinische Daten angewendet wurde. Dieser Algorithmus nutzt sowohl ein physikalisch motiviertes Modell der Röntgenabsorption - die polychromatische Natur der Messung - als auch moderne Techniken der mathematischen Modellierung - die statistische Natur der Messung. Dies wird in der Bezeichnung der Methode deutlich: polychromatisch statistische Rekonstruktionstechnik (PSR).

In der vorliegenden Arbeit wird die Machbarkeit der PSR demonstriert und ihr diagnostischer Mehrwert aufgezeigt. Zum Zeitpunkt der Verfassung dieser Schrift (2020) ist kein Fall bekannt, in dem diese Methode auf klinische Daten angewendet wurde. Mit anderen Worten: zum ersten mal und

weltweit, wurde die PSR auf klinische Daten angewendet und die Bildqualität dadurch enorm verbessert. Es ist insbesondere hervorzuheben, dass der spezielle Fall des C-arm CT eine besondere Herausforderung darstellt, da die Datenqualität wesentlich niedriger ist als im konventionellen spiral CT.

Zusammenfassend kommt die Arbeit zu dem Schluss, dass die PSR sowohl einen erheblichen Mehrwert für die aktuelle klinische Anwendung darstellt, als auch für die weiterführende Forschung und alternativer röntgenbasierten tomographischen Techniken wie Spektral CT oder der zerstörungsfreien Materialprüfung.

Abstract

Computer tomography (CT) is an essential means of diagnosis and aftercare in the clinical routine of hospitals. Although it has been successfully used and improved for decades, the standard reconstruction procedure extracts 3D information from the X-ray images in a way that is both error-prone and simplifying. On the hardware level, the potential of engineering are exploited to the fullest, but on the software level, innovation is only slowly catching up with research.

Many modern procedures require additional technology or alternative measurement methods, but even with the equipment and measurements that are already in use, better clinical CT images could be produced. Due to the constantly growing potential of modern data processing architectures and their computing power, it is now possible to handle enormous amounts of data and to determine results for which it would have taken several months of computing time 20 years ago.

These additional computing capacities can be used to replace the simplified procedures with more complex algorithms. A major problem of the current standard method is the so-called beam hardening. The generated X-rays have an energy spectrum that covers various quantum mechanical interactions. The different interactions contribute to the measured signal. It is known that the absorption process of X-rays depends non-linearly on the absorbing object. It is precisely this non-linearity that is not fully taken into account in standard reconstruction techniques and leads to systematic image errors (so-called artifacts). Historically, it has simply not been possible to acruately consider the non-linearities within a clinically justifiable time frame. However, after 50 years of applied computed tomography, it is now possible to obtain clinical CT images within reasonable computing times. This requires only an appropriate implementation of modern reconstruction algorithms.

This thesis deals with a physically motivated reconstruction algorithm that was developed in 2002 but has not been applied to clinical data since then. This algorithm uses both a physically motivated model of X-ray absorption - the polychromatic nature of the measurement - and modern techniques of mathematical modelling - the statistical nature of the measurement. This is reflected in the name of the method: polychromatic statistical reconstruction technique (PSR).

In the presented work the feasibility of PSR is demonstrated and its enhancement of diagnostic value is shown. At the time of writing this dissertation there is no known case where this method has been applied to clinical data. In other words: for the first time and worldwide, PSR has been applied to clinical data and the image quality has improved enormously. It should be emphasized that the special case of the C-arm CT represents a special challenge, as the data quality is much lower than in conventional CT.

In summary, the work concludes that the PSR represents an added value for current clinical applications as well as for further research and alternative X-ray based tomographic techniques such as spectral CT or non-destructive material testing.

Contents

1 Introduction 15

2 Theoretical background 21
- 2.1 The CT problem . . . 21
  - 2.1.1 Towards CT reconstruction . . . 22
  - 2.1.2 Monochromatic absorption law . . . 26
  - 2.1.3 DIRAC delta distribution and FOURIER-transform . . . 29
  - 2.1.4 FOURIER slice theorem . . . 31
  - 2.1.5 Filtered back projection (FBP) . . . 33
- 2.2 X-ray generation . . . 38
- 2.3 Polychromatic photon absorption law . . . 39
  - 2.3.1 RAYLEIGH scattering . . . 40
  - 2.3.2 Photoelectric absorption . . . 41
  - 2.3.3 COMPTON scattering . . . 43
  - 2.3.4 Pair production . . . 43
  - 2.3.5 Detector technology . . . 44
  - 2.3.6 Beam hardening . . . 46
- 2.4 Statistical reconstruction . . . 49
  - 2.4.1 Iterative monochromatic statistical reconstruction . . . 50
  - 2.4.2 Iterative polychromatic statistical reconstruction . . . 53
- 2.5 Pinhole camera model . . . 59

3 The implementation 63
- 3.1 The forward-projection / the ray caster . . . 63
- 3.2 The backward-projection / the footprint . . . 65
- 3.3 The displacement model - distinguishing f^k by the density . . . 67
- 3.4 The PSR algorithm . . . 69
  - 3.4.1 The initialization of the reconstruction program . . . 70
  - 3.4.2 The forward-projection step . . . 73
  - 3.4.3 The backward-projection step . . . 76
  - 3.4.4 The iteration of the update step . . . 77
  - 3.4.5 The multi-GPU aspects of parallelization and buffer access minimization . . . 79

4 Results 81
- 4.1 Simulations . . . 81
  - 4.1.1 Monochromatic approximation . . . 83

4.1.2 Potential spectral mismatches of $I(E)$ 85
4.1.3 Scattered radiation 88
4.1.4 Detector saturation effects 91
4.1.5 Truncation of the outer volume in the projection space 94
4.2 Real C-arm CT data 97
4.2.1 Pig head experiments 99
4.2.2 5 Euro-cent coin 100
4.2.3 Clinical data 102

5 Discussion **111**

List of Figures

1.0.1 First impression of a metal artifact in a reconstruction of a pig head that includes a steel needle. The left picture shows the result that is provided by the manufacturer of the C-arm system. The image makes it hard to even guess the shape of the object. The right-hand side shows a very first result of the algorithm presented in this thesis made in the year 2016. The window level is $0.0 \ldots 2.0\,\mathrm{g/cm^3}$ which roughly corresponds to 2000 Hounsfield units centered around the grey values of water. 16
1.0.2 The Artis Zeego C-arm angiography system (SIEMENS HEALTHCARE), that was used for the measurements and the reconstruction depicted in Figure 1.0.1 is shown. The drawn axes only illustrate the directions of the volume coordinates but do not correspond to the used world coordinate system nor the coordinate system of the C-arm. Photography by Robert Frysch 17

2.1.1 The goal of the "Candle Steel Shadow Tomographic Setup" is to calculate the position of two balls of full optical opacity (e.g. balls of steel) with projections of two different angles. Since the tomographic problem is ill-posed, it is not possible to find a unique solution. Given the two (projection) measurements one cannot tell if (a) or (b) or another combination represents the reality. 21
2.1.2 If you replace the balls of steel with glasses of coffee, the opacity will drop and the problem becomes more user-friendly. The darker the projection, the more light propagation through coffee has been measured. Thus, the projections of transparent objects contain more information. 22
2.1.3 A simple tomographic candle setup is displayed. The incident light intensity of a candle is 20 candela. The shadow projection measures an intensity of less then 20 candela if the light ray is propagating through an object. The light attenuation depends on the physical size of the object and its light attenuation properties. The darker the coffee, the higher its attenuation. To compute the attenuation value, one needs to calculate the difference of incident intensity and measured intensity. 23

2.1.4 Schematic explanation of the iterative back smearing process that solves the ill-posed example problem. Note that the measured numbers result from the example above. They result from the difference of the incoming 20 candela and the measured light intensity. 24
2.1.5 Schematic description of the absorption process of X-rays propagating through matter is shown. 27
2.1.6 The geometry of parallel beam setup shows an important relation. The 1DFT of the projection with the corresponding angle θ is equal to the line of the same angle in the *2D Fourier space*. 32
2.1.7 The left image shows the sampling of the *Fourier space* needed for a reconstruction without sampling artifacts. The right image shows the radial sampling that is achieved in typical CT setups. 34
2.2.1 Schematic drawing of an X-ray tube is shown. 39
2.2.2 Depiction of two X-ray spectra (calculated by spekCalc and normalized by its maximum value) and the detector response function by Roberts et al. [RHN$^+$08] used in this work can be seen. 39
2.3.1 This is a schematic drawing of the total attenuation of cortical bone as an example material as a sum of the different interactions of the respective energy interval. 41
2.3.2 An overview of three example materials and its energy dependent attenuation properties taken from the NIST data base [HS96]. 42
2.3.3 Schematic depiction of an incident X-ray spectrum (green) $n(E)$ in arbitrary units. After propagation through matter the effective energy of the spectrum (red) is shifted to a higher energy $E_{\text{eff}}^{\text{after}}$. Thus, the propagation process made the beam *harder*. 46
2.3.4 Schematic explanation for the non-linear behavior (beam hardening) of the measured values with respect to the propagation length. 47
2.4.1 The defining properties of a surrogate function in the context of this work is that it always has values above the original function and that its derivative at the value $\vec{\mu}^n$ is the equal to the original. 52

3.1.1 The visualization of the ray casting concept in two arbitrary spatial dimensions illustrates the procedure. 64
3.2.1 The implemented footprint scheme to calculate geometry factors in a back projecting setting 66

3.3.1 As an example parameter set, ($\rho^w = 1.0\,\mathrm{g/cm^3}$, $\rho^b = 1.92\,\mathrm{g/cm^3}$, $\tau' = 0.25$) is used in this work for the clinical head scan data in Section 4.2.3. The plot shows the resulting material fractions of water f^w (red line) and bone f^b (blue line) as a cubic polynomial of the density ρ. 68
3.4.1 PAP of the PSR, pseudo code of the PSR is given in Algorithm 1 at the end of Subsection 3.4.4. 69

4.0.1 Scheme of the 200° Scan 81
4.1.1 GT vs. MSR vs. PSR window $0.95 - 1.05\,\mathrm{g/cm^3}$. The MSR shows the characteristic cupping artifact whereas the PSR is converging accordingly. 83
4.1.2 Sampled X-ray spectra with a corresponding energy range of 28 to 108 keV. The sampling levels are 40,30,20,15,10,9,8,7,6,5,4,3,2 and 1. 85
4.1.3 Head phantom reconstructions with different $I(E)$ that are depicted in Figure 4.1.2. The first row shows the PSR with differently sampled spectra. The second row depicts differences of the MSR that was fed with different monochromatic spectra and compares them with the perfectly sampled spectrum PSR. The third row shows the performance of the PSR when initialized with different detector response functions $D(E)$ and shows the perfectly initialized PSR that was iterated over 5000 times. 86
4.1.4 The cross section plots of the head phantom, which was reconstructed by the PSR, that were initialized with different spectra $I(E)$. 87
4.1.5 Convergency plots of the PSR of the head-disc phantom consisting of the RMSE over the course of iterations and the absolute value of the update step $|\triangle\vec{\rho}|$. Note that the RMSE was only calculated for a single slice while the US considers the whole volume. 87
4.1.6 Disc: GT vs. perfect data PSR and uncorr PSR vs corr PSR 89
4.1.7 Coin: GT vs perfect data PSR and uncorr PSR vs corr PSR . 89
4.1.8 Disc: cross section along the line plotted in Figure 4.1.6 . . . 90
4.1.9 Coin: cross section along the red line indicator in Figure 4.1.7. 90
4.1.10 Coin: cross section along the blue line indicator in Figure 4.1.7. 90
4.1.11 RMSE and US are plotted for the scatter case of the disc phantom. 90
4.1.12 The perfect data PSR vs. uncorr PSR (rotated and non rotated) vs. corr PSR show distinct artifacts. 92
4.1.13 The GT vs. perfect data PSR vs. uncorrected PSR vs. corrected PSR of the coin phantom which were initialized with tanh-saturated projections show only small artifacts. 93

4.1.14 The cross section plot of the disc phantom PSR that was initialized with tanh-saturated projections. 93
4.1.15 The cross section plot along the thickness of the coin phantom phantom PSR that was initialized with tanh-saturated projections. 93
4.1.16 The cross section plot along the diameter of the coin phantom PSR that was initialized with tanh-saturated projections. . . 93
4.1.17 GT vs. PSR with the same size as the GT and the full projection data vs. PSR of lower voxel number, lower resolution and less projection data (window level: $0.00 - 1.92\,\mathrm{g/cm^3}$). . 97
4.1.18 PSR in the standard setting vs. PSR given the perfect OV vs. PSRov (window level: $0.90 - 1.10\,\mathrm{g/cm^3}$). 97
4.2.1 Steel needle pig head reconstructions: FBP vs. MSR vs. PSR, window level: $0.50 - 2.40\,\mathrm{g/cm^3}$ 100
4.2.2 A quasi perpendicular projection views on the 5 Euro-cent coin and the cross section of the extinction values. The measured frontal and side view projections of the coin show an inconsistency in the extinction values from those two directions. The ratio of the extinction value is expected to be proportional to its geometric size (except for beam hardening effects), which is above a factor of the order 10. However, the measured ratio lies around $4.8/4.2 \approx 1.14$. 101
4.2.3 MSR vs. PSR, window level: $0.0 - 7.0\,\mathrm{g/cm^3}$ 102
4.2.4 MSR vs. PSR vs. cross section window 0.0 - 7.0 102
4.2.5 Schematic depiction of the procedure to evaluate cupping reduction of the PSR on the 36 data sets. 103
4.2.6 Data set 12: The PSRov does not improve the image quality but in fact introduced even more cupping. The PSRrsi shows a clear improvement in image quality. 105
4.2.7 Data set 36: The PSRov introduces artifacts over the course of 40 iterations. 106
4.2.8 Data set 7: The PSRov introduces artifacts over the course of 40 iterations. 107
4.2.9 Data set 2: The PSR suffers from strong motion artifacts introduced by the applied motion compensation. Using the uncorrected projection matrices results in less motion artifacts. The ML is less sensitive to such motion. The PSRrsi is only slightly improves the image quality. 107
4.2.10 The native ML vs. bi-material corrected ML vs. native PSR vs. PSRrsi vs. PSRov are depicted. The used window size is relatable to 100HU. 108

4.2.11 The bi-material corrected ML (left-hand side) and PSR (right-hand side), circumscribed mild brain edema is depicted, but barely visible. The blue arrows point out at the absence of sulci. Typically they should be visible as the white arrow suggests. The window level corresponds to 100HU. Here a ring artifact correction has been performed. 108
4.2.12 Data set 28: A typical convergence pattern of the update step when different numbers of iterations are applied. 109
4.2.13 Dataset 28: Difference between 40 iterations of the default setting and the 300 iterations depicted in Figure 4.2.12. . . . 109

1 Introduction

Over the last hundred years of scientific progress, humanity's knowlegde of the underlying fundamental processes of nature has increased drastically. The more accurate physical models became, the more potential for technical workarounds were imaginable. As civilization keeps developing more advanced methods in science, while fine tuning the already achieved ones, there can be also stagnation in various fields. A recent example is the so-called Eroom's law, which is the counterpart to the famous MOORE'S law [Moo65]. Scannell et al. [SBBW12] report their findings in the field of commercial drug research and state,

"*The past 60 years have seen huge advances in many of the scientific, technological and managerial factors that should tend to raise the efficiency of commercial drug research and development (R&D). Yet the number of new drugs approved per billion US dollars spent on R&D has halved roughly every 9 years since 1950, falling around 80-fold in inflation-adjusted terms. There have been many proposed solutions to the problem of declining R&D efficiency. However, their apparent lack of impact so far and the contrast between improving inputs and declining output in terms of the number of new drugs make it sensible to ask whether the underlying problems have been correctly diagnosed.*"

There are a variety of speculations on the causes of Eroom's law in the field of drug research. However, there is no such a thing in the field of programming or medical imaging, even though one could consider that with increasingly faster calculation power and memory capacity, programmers tend to favor a more lazy computing style when it comes to performance aspects. One could interpret such a tendency as a "programmers" PARKINSON'S law [Par55], "*[W]ork expands so as to fill the time available for its completion*".

Additionally, one can see similarities to the field of medical imaging since every aspect of research which is closely related to commercialized applications and industry might be under similar conditions. There have been trends occuring in certain disciplines that range from the inflationary usage of buzzwords such as "compressed sensing" up to conference sessions filling topics about deep learning-based approaches that not only increase the accuracy of diagnosis given a medical image [LJC$^+$17], but even produced better medical images [ZD20]. While new technical developments want to realize the dream of artificial intelligence, those trends may also overshadow potential solutions that were never evaluated nor benchmarked against the proposed method of this thesis due to a limitation in calculation power when those potential solutions came to live.

One of the main techniques to acquire medical images is the computed tomography (CT). It has become a commonly used imaging technique not only in medical imaging, but also in nondestructive testing. It is based on the fact that common substances, such as human tissue, bones, metals and timbers look *transparent* to X-ray photons. In other words, X-rays offer the interested viewer to look *inside* an object instead of just observing its surface.

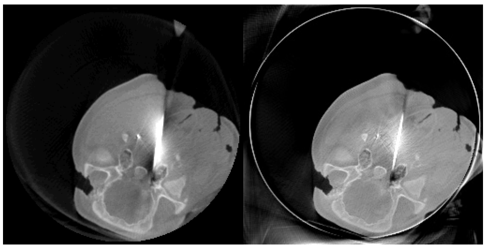


Figure 1.0.1: First impression of a metal artifact in a reconstruction of a pig head that includes a steel needle. The left picture shows the result that is provided by the manufacturer of the C-arm system. The image makes it hard to even guess the shape of the object. The right-hand side shows a very first result of the algorithm presented in this thesis made in the year 2016. The window level is $0.0 \ldots 2.0\,\mathrm{g/cm^3}$ which roughly corresponds to 2000 Hounsfield units centered around the grey values of water.

This thesis focusses on a method of medical image reconstruction originated in 2002 that might have been too calculation-heavy at that time, but certainly is not anymore. A main drawback of X-ray based imaging is the very nature of such X-rays. They belong to the group of ionizing radiation. Such radiation transports enough energy to change the quantum state of electrons in atoms and molecules. Thus, it can change fundamental properties of biological tissue and cause serious damage to living organisms. The contrast information of an X-ray projection *causes damage and delivers a diagnostic value* at the same time. The X-ray dose is proportional to the tissue damage caused by radiation exposure. A main drive to enhance the diagnostic value of X-ray imaging techniques is to minimize the dose while providing the best image quality possible to the radiologist.

State-of-the-art CT systems work with X-ray sources that provide a continuum of photons with different energies. Those photons propagate through the object of interest (e.g. a head of a patient) and are collected by specialized X-ray detectors which consist of materials that absorb those photons and produce visible light photons in the process. This light can be recorded by CCD-chips and transformed to an electrical signal, which is a measure of the X-ray intensity that is not attenuated by the object of interest. Note that this electrical signal is typically a sum of different light photons that are produced by X-ray photons of different energies. This means in particular that the measurement process loses the information about the energy dependency while integrating over this exact attribute.

Additionally, this attenuation process is typically assumed to be linear with the object's size and density. Namely, if the size of the patient doubles, the attenuation of the patient doubles as well. This property only holds true if the incoming X-ray photons have the same energy (which they have not) or the size of the measured object is very small (which is not true either). The attenuation process itself has a non-linear dependency of the photon energy with respect to the object of interest. This means in particular that an X-ray projection image of two identical objects that are scanned consecutively in a row, is not the same as the sum of two separated X-ray projections of the same objects. This non-linearity is called beam hardening and will be explained later in this thesis.

However, it turns out that certain mathematical models are able to correctly incorporate such a non-linear behavior while increasing the overall computational effort to reconstruct the medical image. Maximum likelihood approaches combine statistical knowledge about such processes with the powerful framework of mathematical analysis. The resulting reconstruction methods look similar to established techniques, but usually require prior knowledge and more computational performance. The latter represents a decreasing difficulty due to Moore's law [Moo65].

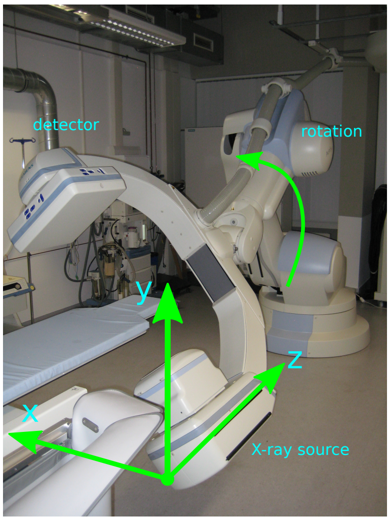


Figure 1.0.2: The Artis Zeego C-arm angiography system (Siemens Healthcare), that was used for the measurements and the reconstruction depicted in Figure 1.0.1 is shown. The drawn axes only illustrate the directions of the volume coordinates but do not correspond to the used world coordinate system nor the coordinate system of the C-arm. Photography by Robert Frysch

A very elegant approach is the so-called polychromatic statistical reconstruction (PSR) by Elbakri and Fessler [EF02]. It combines the statistical character of the maximum likelihood framework with a physically motivated X-ray projection model. Since the model is based on the physics of X-ray attenuation, it can estimate the projection process with regard to the different materials that the object of interest consists of. If the prior knowledge about the materials and X-ray source are given, one can apply the PSR to CT data and increase the image quality dramatically. Thus, instead of applying (dominantly) empirical image processing, it simply models the physics of the CT measurement process in a proper manner and incorporates this into a maximum-likelihood framework. However, Elbakri and Fessler never applied their method to real patient data. An important motivation for this work was that there was no publication found about the PSR and its application on clinical C-arm CT data[1]. To make an algorithm like this work on C-arm CT is particularly challenging, since C-arm data quality is substantially worse in comparison to conventional CT. C-arm systems bring more difficulties into play for a variety of reasons [KRK12]:

- the flat-panel cone-beam geometry cannot grant a sufficiently homogenous illumination of the detector,
- the detector needs to be more flexible and robust due to the nature of the moving C-arm,
- the dynamic bit-range is lower compared to typical detectors used in conventional CT,
- the tube voltage of the X-ray source might vary from 70 kVp to 120 kVp while recording the projection data,
- and the photon flux is changing due to tube current that is not held constant.

In this work, we use clinical head scan data where the tube voltage was kept constant. This is an advantage in comparison with abdomen scans, where the X-ray spectrum is modified in between the different angles.

The “non-linearity problem” of the measurement process becomes more drastic in the case of high-absorbing materials such as bones or metal compounds (e.g. pacemakers or screws) and can potentially render the medical image almost useless, as its diagnostic value decreases drastically due to strong artifacts that are induced due to those high-absorbing materials.

Initially, it was observed that metal artifact reduction (MAR) methods do not work sufficiently well when confronted with metal parts in phantoms consisting mostly of water [GMV$^+$17, LFN09]. Also, there is a lot of image processing needed to make most of the artifact-tackling methods work. For example, MAR methods typically try to estimate projection values caused by metal objects and replace them with the actual measurements in order to reduce metal artifacts in the reconstruction.

Astonishingly, there is only limited real data PSR found in literature. In particular, the method has *never* been evaluated on clinical data. All previous studies have used either simulated data or ‘small’ data sets of test bench flat panel CT.

---

[1] Note that C-arm CT data is recorded with a flat panel detector, thus the C-arm setup is usually categorized as flat-panel cone-beam CT.

Another aspect of such a real data usage is the need for information from ‘hidden measurements’ protected by the manufacturer, such as non-water-corrected projections and projection matrices. Most companies are highly protective of their processing pipeline (e.g. scatter kernels, over exposure correction, normalization, cosine weighting, water correction and post-processing steps such as image modification to reduce ring artifacts). The rare opportunity to collaborate closely with the software development team (specifically the C-arm CT group) of SIEMENS HEALTHCARE enabled this work. Real data is processed in a state-of-the-art manner with respect to certain corrections such as scattered radiation and *beam hardening*. Rewinding the *beam hardening* correction is essential for the PSR and therefore allows for conclusive methodological analysis of the effect under investigation.

These key issues reveal the reasons why Elbakri and Fessler — or someone else — never started to evaluate this very elegant PSR approach on clinical data. Being in close collaboration with radiologists of the university hospital in Magdeburg, it can be assured that the PSR approach does not need any new measurement hardware and should stay compatible with their current C-arm system as well as with real CT scanners and upcoming generations of C-arm systems. A classification within the context of other state-of-the-art methods which tackle beam hardening will be given in the discussion.

Due to MOORE’S law, it has become feasible to apply the computationally expensive PSR to real data sets a decade after its original release. Recent development in modern techniques (e.g. spectral CT or dual energy CT) aimes to overcome the beam hardening problem by new measurement techniques. Since those new methods still apply an X-ray dose, it is of major interest to reduce this dose to a minimum. To not fall into the pothole of Eroom’s law it could be very interesting to apply the accumulated knowledge of the traditional CT community — especially the PSR — to this new approach. This also includes training data sets for potential deep learning approaches and a better understanding of the potential efficiency of information that is measured.

Thus, in order to get the maximum out of anything that will come up in the future of medical imaging, it could be beneficial to fully grasp what was already achieved in the past.

This work describes the implementation of the PSR and its application to a difficult type of data that led to an increase of diagnostic value. The method was implemented in C++ and OpenCL to ensure a feasible time-performance. This includes a ray caster projector that models the multi-energetic absorption process of X-ray photons in matter, a back-projector and buffer management of the multiGPU hardware.

Initially the medical images produced from the head scan data suffered from the patient table that is partially truncated in the X-ray projections. It was found out to play a dramatic role for the image quality of the PSR. Comparable methods show almost no sensivity due to the truncation.

This development should be acknowledged in order to minimize the effects of Eroom’s law with respect to the domain of medical imaging.

2 Theoretical background

There is nothing more practical than a good theory.
- Jürgen Schmidhuber[1]

2.1 The CT problem

The computed tomography (CT) is an imaging technique with high clinical relevance in daily practice. It is based on the X-ray projection modality. The goal is to reconstruct 2D/3D information of a patient with 1D/2D projection measurements. For the sake of explanation, imagine a candle setup as in Figure 2.1.1.

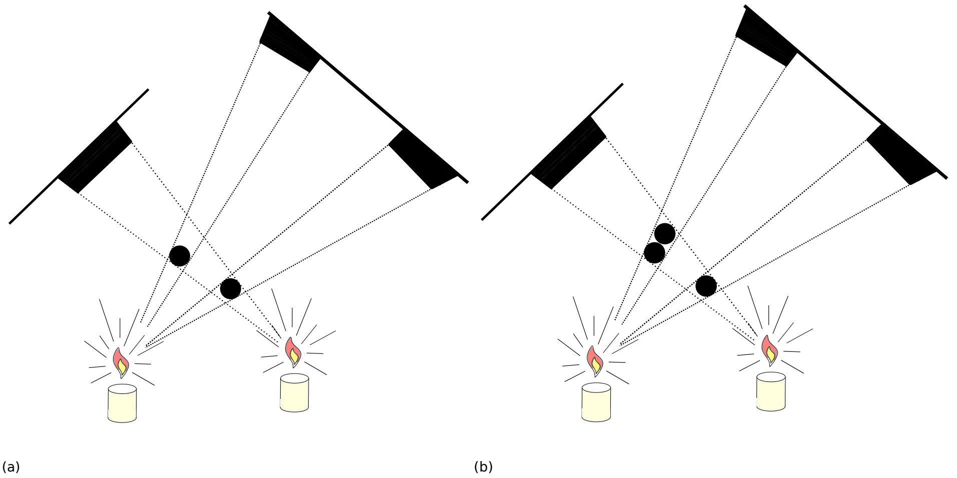
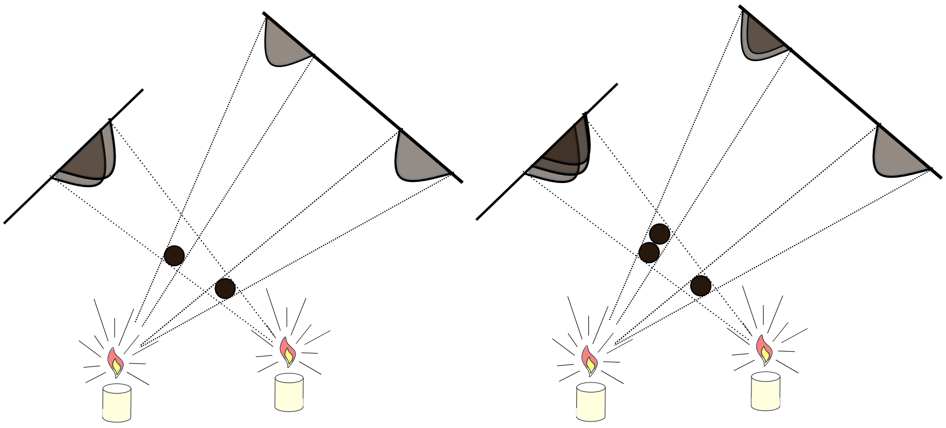


Figure 2.1.1: The goal of the "Candle Steel Shadow Tomographic Setup" is to calculate the position of two balls of full optical opacity (e.g. balls of steel) with projections of two different angles. Since the tomographic problem is ill-posed, it is not possible to find a unique solution. Given the two (projection) measurements one cannot tell if (a) or (b) or another combination represents the reality.

In a 2D setup like this, it becomes very easy to calculate the position of the two objects but if the number of balls is not given and the radius is unknown, one can find more then one solution and the problem becomes *ill-posed*. Just to mention: the solution (a) is called the minimal solution.

---

[1] Lex Fridman (MIT) interviews Jürgen Schmidhuber in 2018: "Juergen Schmidhuber: Godel Machines, Meta-Learning, and LSTMs | Artificial Intelligence (AI) Podcast" https://youtu.be/3FIo6evmweo



(d) (c)

Figure 2.1.2: If you replace the balls of steel with glasses of coffee, the opacity will drop and the problem becomes more user-friendly. The darker the projection, the more light propagation through coffee has been measured. Thus, the projections of transparent objects contain more information.

The projections can be formulated as line integrals along the ray path which sum up how much light got actually absorbed. If there is a given infinite set of such 1D projections, 360° around the unknown 2D object with known light absorption model, one can find an exact mathematical solution for this inverse problem in case of a monochromatic light source (e.g. a laser). It is called the filtered backward-projection (FBP) and is described in Section 2.1.5.

In CT reality, we encounter several issues that corrupt the clean mathematical beauty of that solution. Before we start to elaborate the FBP we motivate the idea behind solution techniques used in such tomography settings and introduce the mathematical models that are mandatory for the FBP.

2.1.1 Towards CT reconstruction

Imagine the candle setup is being modified as depicted in Figure 2.1.3. Further, let's assume that the light intensity measured in candela [cd] is absorbed linearily by each glass of coffee depending on its "blackness". Thus, the stronger the coffee, the darker the measured shadow. The solution with respect to the four glasses of blackness $\vec{x}$ needs to satisfy the equation

$$A\vec{x} = \vec{p}$$

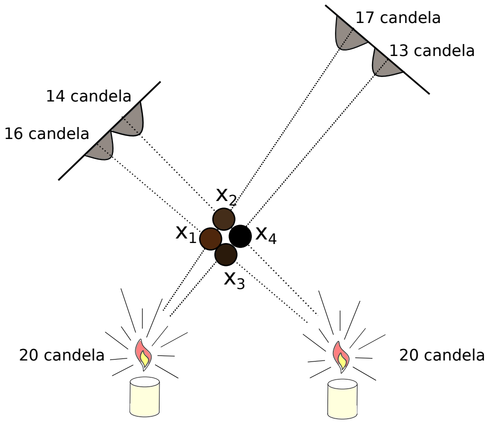


Figure 2.1.3: A simple tomographic candle setup is displayed. The incident light intensity of a candle is 20 candela. The shadow projection measures an intensity of less then 20 candela if the light ray is propagating through an object. The light attenuation depends on the physical size of the object and its light attenuation properties. The darker the coffee, the higher its attenuation. To compute the attenuation value, one needs to calculate the difference of incident intensity and measured intensity.

with $\vec{x} = [x_1, x_2, x_3, x_4]^{\mathrm{T}}$ and the measured absorption[2] $\vec{p} = [p_1, p_2, p_3, p_4]^{\mathrm{T}} = [3, 7, 4, 6]^{\mathrm{T}}$ with respect to the initial intensity of 20 cd. By defining $\vec{p}$ in this manner the matrix A follows as

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

It is obvious that this linear equation system cannot be solved analytically since the system matrix A has a vanishing determinant[3].

To avoid this disturbance one could try to somehow "back smear" the measured numbers. Let's establish a model that assumes the glasses are empty at the first glance. If this were true, the shadows would have to vanish. Since such transmission is not observed (as we measure a decline in light intensity) we try to back propagate the measured blackness distributed to the glasses which contribute to the very same measurement. Since the glasses have the same diameter, one can distribute this blackness equally. One back propagation that regards each measurement once, is called an update step. Since we can repeat the very same procedure and keep the

---

[2] The blackness of one measured shadow results in the difference of the initial light intensity and the residual after propagating through the coffee.

[3] Adding the first and second row delivers the same as adding the third and fourth row. Hence, those rows are not linear independent.

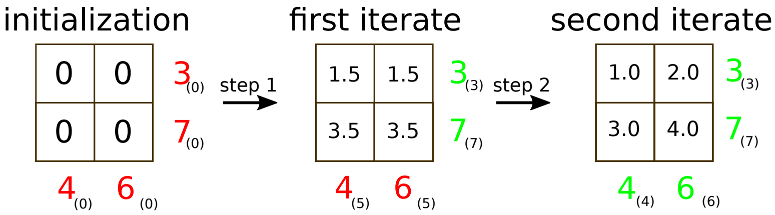


Figure 2.1.4: Schematic explanation of the iterative back smearing process that solves the ill-posed example problem. Note that the measured numbers result from the example above. They result from the difference of the incoming 20 candela and the measured light intensity.

back propagated information, we call this recursive technique an iteration. In our example case, it turns out that we can find a solution after only two iterations as depicted in Figure 2.1.4.

We will stay a little longer in the "candle-light-shadow-analogy-domain" to express the practical problems of the computed tomography.

If we increase the number of glasses and the number of candles and measurement screens, it will be more likely that the dimensions of $\vec{x}$ and $\vec{p}$ differ. In such a case a unique analytical solution is not available by design. Real measurements $\vec{p}$ in different angles might contradict each other due to erronous effects on the measurement setup, such as wind, that temporarily lowers the light intensity of the candles. In a CT scanner this could be a patient who is moving while the scan is performed.

As real measurements are always corrupted by noise, a "clean mathematical" solution might become impossible as well. The noise of the shadow might cause oscillatory behavior within an iteration or in between update steps. For this purpose we will later introduce the concept of regularization.

Imagine the glasses of coffee are measured in arbitrary angles, the system matrix A will contain numbers between 0 and 1 multiplied by its diameter. This is because the cone that is measured may only slightly traverse a glass instead of encompass the full glass. This information is important for the back smearing approach, because different glasses are assigned to different weightings. We call those weightings geometry factors later on.

Very strong coffee of high opacity might produce an almost black shadow and becomes indistinguishable from another coffee glass of similar strength or simply conceal brighter glasses (e.g. Espresso vs. Americano). In other words, the shadow resolution (or "bandwidth") needs to be sufficient with an increasing number of glasses. This can be identified as the dynamic range of the detector. To work around this problem, one could consider increasing the light intensity (e.g. a flood light from a lighthouse). In industrial CT, where no patient is scanned, this might be a viable option, but living organisms are severely damaged by ionizing radiation such as X-rays.

The candle produces a polychromatic spectrum of visible light that even contains infrared radiation. The light reflexion and absorption (dispersion) is different for every photon energy and is particular non-linear. This makes the interpretation of the "shadow signal" ambiguous and renders the model of linear absorption invalid. This non-linearity is later identified as beam hardening and is the main motivation of this work.

Furthermore the measurements are usually not infinite in terms of angular resolution. We will see that finite angular resolution corrupts the filtered backward-projection in the "first order" due to missing angular information for the back smearing process and in the "second order", as it cannot be re-sorted to transforming a data set from fan beam geometry to a parallel beam geometry.

Despite those practical flaws, the idea of back smearing is the core of every classical reconstruction technique[4]. Iterative methods revolve around a forward model that predicts the measurements given a current iterate. Depending on the mathematical formulation and the solution of such framework, each resulting reconstruction technique attains weaknesses and advantages compared to others.

The practical CT problem belongs to the class of *ill-posed inverse problems* which implies a broad range of concepts that solve such problems for a given configuration of measurements. Additional *a priori* knowledge might be incorporated as well.

Tomography is realized as an ensemble of different point of views of an X-ray source that has a detector system attached in a way that the object gets traversed by the radiation. Instead of visible light, higher frequencies of the so-called X-radiation is of interest.

Early X-ray based applications where focussed on single projections, as they already deliver information about the *inner structure* of objects. Nowadays, X-ray projections are still a viable tool, but if depth information is mandatory, reconstruction techniques become of importance. CT systems consist of at least one X-ray source and an X-ray detector. The projection images are recorded while those components are moving around the object of interest. The common source-detector trajectories describe a circle whose center point lies within the object. Since the beginning in 1973 [Hou73], there are two different medical applications: interventional and diagnostic CT.

The latter are limited to traditional trajectories since they provide a heavy gantry and a curved detector. Such gantries are able to rotate with 4 GHz [MSCH18] and can support high quality detectors with fast read out times (high spatial and temporal resolution). They are used to generate medical images of high quality for the radiologist. At the same time, this makes it infeasible for a physician to perform an intervention.

---

[4] This might not apply in this very same sense to deep learning approaches. Even though the learning process is always the difference of ground truth back propagated and weighted by the gradient of the network architecture.

Interventional CT supports the physician during a surgery. It needs to acquire reconstructions in a very short time. C-arm devices are commonly used for such intraoperative settings. They offer a more flexible source-to-detector geometry, but rely on less heavy components due to stability issues. C-arm devices record projections with a flat panel detector.

Depending on the actual reconstruction technique, one might adjust the amount and the position of such "emitter-receiver" configurations properly. Before we get to an analytical approach that solves the "standard 180° scan" problem and furthermore the CT problem, we need to specify the X-ray measurements and the implications of the underlying physics and algebra.

2.1.2 Monochromatic absorption law

The easiest model for X-rays propagating in materials along straight lines can be found with the Lambert-Beer absorption Law. It is a heuristically motivated model that describes the decreasing of the number of photons n_0 in homogenous matter. In this work we will always talk of an incident spectrum $n(E)$ which gives the actual number of photons with the energy E or of an incident photon intensity

$$I(E) := \frac{n(E)}{\int_{\mathbb{R}^+} n(E) \mathsf{d}E},$$

which is the normalized version of $n(E)$. Imagine a monochromatic beam of photons $n_0(E) = N_0 \delta(E - E_0)$ (i.e. all N_0 photons have the exact same energy E_0, details follow in Subsection 2.1.3) that is partially scattered and absorbed in a thin slice of matter characterized by the linear attenuation coefficient μ. This coefficient is an indicator of how much interaction of matter and photons is happening for a certain photon energy. The attenuation coefficients μ are catalogued for different energies and materials [HS96].

According to Lambert-Beer, the difference $\Delta n := n - n_0$ of the photon count is stated to be proportional to the incident number n_0 and the small thickness of material Δx. Transition onto the infinitesimal thickness $\mathsf{d}x$ yields

$$\mathsf{d}n \propto -n \cdot \mathsf{d}x$$

$$\Rightarrow \mathsf{d}n = -\mu(x, E_0) \cdot n \cdot \mathsf{d}x$$

$$\Rightarrow \int_{n_0}^{n} \frac{\mathsf{d}n'}{n'} = -\int_{x_0}^{x} \mu(x', E_0) \, \mathsf{d}x'.$$

Note that we state that $\mu(x, E)$ is a function of the spatial coordinate x, but not of the photon count n. Furthermore, μ is a function of the photon

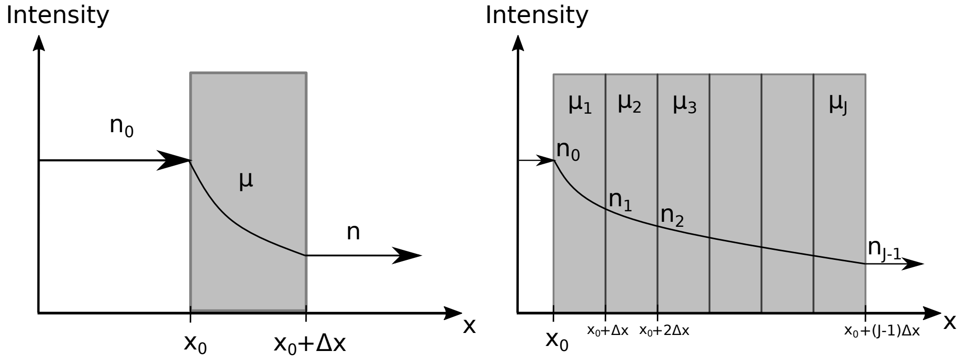


Figure 2.1.5: Schematic description of the absorption process of X-rays propagating through matter is shown.

energy E_0, but we can neglect that dependence, since all photons have the same energy (monochromatic) in this example. From the above we can now formulate the transmitted photon intensity/count and define the extinction ϵ as

$$\int_{x_0}^{x} \mu\left(x', E_0\right) \mathsf{d}x' = \ln n_0\left(E_0\right) - \ln n\left(x, E_0\right) = \ln \frac{n_0\left(E_0\right)}{n\left(x, E_0\right)} =: \epsilon\left(x, E_0\right). \tag{2.1.1}$$

An alternative derivation of the absorption law, compared to the one we have above, can be found in the book of Buzug [Buz08] (chapter: “Problems with Lambert-Beer’s Law”). The exponential decay of the intensity that corresponds to the line integral over the volume $\vec{\mu}$ can also be seen as a cascade of multiplications of infinitesimal absorption steps through the medium. Combining binomial and Poisson distribution leads to the same linearity of the measurements with respect to the volume.

In computed tomography the extinction values ϵ_i are the measurements. The index i continues over all detector pixels and all measurement angles. Thus we can formulate a 2D line $\mathcal{L}_i = \left\{\vec{x_i}\left(t\right) = \vec{\xi_i} + t\vec{r_i}\,|t \in \mathbb{R}\right\}$ respectively that goes through the X-ray source $\vec{\xi_i}$ and the i-th detector pixel. The integral in the above Equation 2.1.1 reformulates as

$$\bar{\epsilon}_i = \int_{\mathcal{L}_i} \mu \mathsf{d}s_i = \int_0^\infty \mu\left(\vec{x}_i\left(t\right)\right) \mathsf{d}t. \tag{2.1.2}$$

We identify the measurements ϵ_i as line integrals along ray paths $\mathcal{L}_i$.

The 2D CT problem is to find the scalar function $\mu\left(x, y\right) : \mathbb{R}^2 \to \mathbb{R}$ that minimizes the error of the measurements ϵ_i to its forward-projections $\bar{\epsilon}_i = \int_{\mathcal{L}_i} \mu \mathsf{d}s_i$. We use the bar $\bar{\epsilon}$ to emphasize that Equation 2.1.2 models what kind of measurement is expected, given an object that is characterized by the scalar function $\mu(\vec{x})$.

Imagine that we discretize $\mu\left(x,y\right) \longrightarrow \mu_j,\, j = 1, ..., N_j$ into N_j pixels, which are partially constant over a short range[5]. Note that the index j at N_j denotes that it refers to the reconstructed object and not that there is a sequence of $N_{j=1,2,3,...}$ in any viable interpretation. Typically N_j ranges from $256 \times 256 = 65,536$ to $2048 \times 2048 = 4,194,304$ [HYH$^+$18]. The information of voxels that get in touch, and how much, with the pencil beam, that we approximate as the line $\mathcal{L}_i$, can be stored in the so-called system matrix $A = \{a_{ij}\}$. Details about the geometry factors a_{ij} follow in chapter 3.1. The algebraic CT problem formulation can be stated as

$$A\vec{\mu} = \vec{\epsilon} \tag{2.1.3}$$

or, more specifically, the N_i equations

$$\sum_{j=1}^{N_j} a_{ij}\mu_j = \epsilon_i,\; i = 1, ..., N_i. \tag{2.1.4}$$

As we will see later, the size of the matrix A is in fact *too large* to store for this forseeable future, even taking MOORE'S law into account. And even if it were possible to store the amount of data, the access to the data would cause serious problems as well. Hence, to solve such a practical CT system of linear equations by pseudoinverse multiplication is not an option any time soon, one needs to find iterative methods that somehow follow the idea presented with the candle setting in Figure 2.1.3. The goal is to find the minimizer $\vec{\mu}^*$ with non-negative coefficients

$$\vec{\mu}^* = \underset{\mu_j \geq 0,\, \forall j}{\arg\min} \sum_{i=1}^{N_i} \left[\sum_{j=1}^{N_j} a_{ij}\mu_j - \epsilon_i\right]^2. \tag{2.1.5}$$

Since the algebraic CT problem is overdetermined, the classical iterative KACZMARZ [Kac37] algorithm gives a practical solution. If there exist a unique solution (if the system is not overdetermined), it is proven that this iterative algorithm converges to this solution [Tan71].

By denoting $\vec{a}_i$ as the i-th row of A

$$\epsilon_i = \vec{a}_i \cdot \vec{\mu}$$

it is proven that

$$\vec{\mu}^* = \lim_{k\to\infty} \vec{\mu}_k$$

with

$$\vec{\mu}_{k+1} = \vec{\mu}_k + \vec{a}_{i_k}^{\mathrm{T}} \cdot \frac{\epsilon_i - \vec{a}_{i_k} \cdot \vec{\mu}_k}{\vec{a}_{i_k} \cdot \vec{a}_{i_k}^{\mathrm{T}}} \tag{2.1.6}$$

---

[5] Technically speaking, we project the continuous function $\mu(\vec{x})$ on to the basis function of a pixel or a voxel.

and the index i_k that is found [SV08] to be sampled at random for each iteration k. This algorithm computes the forward-projection $\vec{a}_{i_k} \cdot \vec{\mu}_k$ of the current volume $\vec{\mu}_k$, compares this with the measured pixel ϵ_i and back projects this difference into the current volume. The back projection is realized by the column vector $\vec{a}^{\mathrm{T}}_{i_k} / \left(\vec{a}_{i_k} \cdot \vec{a}^{\mathrm{T}}_{i_k} \right)$ that points only to volume elements that interfers with line $\mathcal{L}_i$. The structure of this approach is similar to the structure of the reconstruction technique used in this thesis and the reason for introducing the back smearing metaphor first.

In fact, new research [HWF17] suggests that reconstructions can be improved if the forward-projection is modelled more accurately than the linear model in Equation 2.1.2, while leaving the rest of the algorithm almost unchanged.

An additional adaptation of the Kaczmarz method is called SART (simultaneous algebraic reconstruction technique). Here, several update steps (of different subsets) are performed simultaneously and weighted accordingly within one iteration. This leads to a substantial speed up with respect to computing time.

Iterative methods are almost[6] independent of the scanning geometry. The following section derives an analytical solution for a certain scanning geometry that is the most commonly used reconstruction method in hospitals to this day. Their main advantage compared to the problem statement 2.1.5 lies in their low computational effort. It is helpful to read through that topic to see the “bigger picture” of the CT reconstruction problem.

2.1.3 Dirac delta distribution and Fourier-transform

This subsection supports the reader with a recap of the mathematical background information for the sake of completeness[7]. To have an understanding of the underlying algebra that describes the phenomena which lead to the measurements, we need to specify a few concepts of *complex analysis*. A mathematical in-depth guide for this topic is the book of Helgason [Hel99].

With the definition of the Heaviside-function $H : \mathbb{R} \to \{0, 1\}$

$$H(x) := \begin{cases} 1 & , x \geq 0 \\ 0 & , x < 0 \end{cases}$$

and a test function $\phi \in C_0^\infty(\mathbb{R})$ (that has a *compact support* and *infinite continuous* derivatives) we can define the *Dirac delta distribution* as a *weak*

---

[6] In this work, the system matrix A is of full rank and the system may not be underdetermined, but there are iterative approaches that tackle underdetermined problem settings as well.

[7] However, it is not mandatory to read through, but it can come in handy when reading further into derivation of the reconstruction technique used in this thesis (see Section 2.4.2).

derivative of H. Weak differentiation means, it follows the same rules as a classical derivative except for a finite number of points which do not impact the integral over the function. Thus it satisfies in particular the partial integration (Equation 2.1.7). With the notation[8] $\langle\cdot,\cdot\rangle$ we define the scalar product of two functions $f, g : \mathbb{R} \to \mathbb{R}$ or distributions as $\langle f, g\rangle := \int_{\mathbb{R}} f(x) \cdot g(x)\,\mathrm{d}x$ and similarily the edges can be expressed as $[F]_{-\infty}^{\infty} := F(\infty) - F(-\infty)$. Due to the product rule $\frac{\mathrm{d}}{\mathrm{d}x}(f \cdot g) = g \cdot \frac{\mathrm{d}}{\mathrm{d}x} f + f \cdot \frac{\mathrm{d}}{\mathrm{d}x} g$ follows the partial integration[9]

$$[f \cdot g]_{-\infty}^{\infty} = \frac{\mathrm{d}}{\mathrm{d}x}\langle f, g\rangle = \left\langle \frac{\mathrm{d}}{\mathrm{d}x} f, g\right\rangle + \left\langle f, \frac{\mathrm{d}}{\mathrm{d}x} g\right\rangle . \tag{2.1.7}$$

We define the δ-distribution as follows

$$\left\langle \frac{\mathrm{d}}{\mathrm{d}x} H, \phi\right\rangle \overset{2.1.7}{=} [H \cdot \phi]_{-\infty}^{\infty} - \left\langle H, \frac{\mathrm{d}}{\mathrm{d}x}\phi\right\rangle \overset{\phi \in C_0^{\infty}(\mathbb{R})}{=} - \left\langle H, \frac{\mathrm{d}}{\mathrm{d}x}\phi\right\rangle$$

$$\Rightarrow \left\langle \frac{\mathrm{d}}{\mathrm{d}x} H, \phi\right\rangle = -\int_0^{\infty} \mathrm{d}x \left(\frac{\mathrm{d}}{\mathrm{d}x}\phi\right) = -[\phi]_0^{\infty} = \phi(0) =: \langle \delta_0, \phi\rangle . \tag{2.1.8}$$

With this, one can write weak derivatives of this δ-distribution as

$$\left\langle \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^n \delta_0, \phi\right\rangle = (-1)^n \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^n \phi(0) .$$

Since ϕ is a test function,

$$\phi \in S(\mathbb{R}^n) := \left\{ \phi \in C^{\infty}(\mathbb{R}^n) \middle| \forall a, b \in \mathbb{N}_0^n : \sup_{x \in \mathbb{R}^n} \left| x^a \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^b \phi(x) \right| < \infty \right\}$$

is a SCHWARTZ-function[10] as well. ϕ has infinite derivatives and is rapidly decreasing at infinity (the derivatives are rapidly decreasing at infinity as well, $a, b \in \mathbb{N}_0^n$ are multi-indices). Since it is commonly known that a derivative of a *Fourier-transform* is equal to multiplication with its argument and a constant in the *dual* space[11], we can identify the SCHWARTZ-functions with such an attribute. If $\phi \in S\left(\mathbb{R}^n\right)$, we call the following the *Fourier-transform* $\mathcal{F} : S\left(\mathbb{R}^n\right) \rightarrow S\left(\mathbb{R}^n\right)$ with

---

[8] This notation is commonly used in quantum mechanics and statistical physics. It emphasizes the scalar product character as a full analogy from the vector space (linear algebra) that can be formulated for the space of functions with certain properties such as the famous HILBERT space.

[9] In some literature, it is referred as divergence theorem or GAUSS's theorem or OSTROGRADSKY's theorem. This is due to the fact that n-dimensional volume integrals of a divergence always can be written as surface integrals which have the dimension of $n-1$: $\int\!\cdots\!\int_{\Omega^n} \mathrm{d}V_{\Omega} \left(\nabla^{\mathrm{T}} \cdot \vec{f}\right) = \oint \cdots \oint_{\partial\Omega^{n-1}} \mathrm{d}S_{\Omega} \left(\vec{f}^{\mathrm{T}} \cdot \vec{n}_{\partial\Omega}\right)$. This reduces to GREEN's theorem ($n = 2$) and to the fundamental theorem of calculus ($n = 1$), hence the partial integration.

[10] Such a function does not need to have a compact support, e.g. it could be a *Fourier* basis function or a GAUSSIAN function.

[11] There exists an isomorphism between this specific space and its dual space. Here we will talk about the frequency domain $\vec{u}$ and its spatial dual space and its elements $\vec{x}$.

$$\mathcal{F}\left[\phi\right]\left(\vec{u}\right) = \tilde{\phi}\left(\vec{u}\right) := \int_{\mathbb{R}^n} \mathrm{d}V_x \left(\phi\left(\vec{x}\right) \exp\left\{-2\pi i \vec{x}^{\mathrm{T}} \vec{u}\right\}\right)$$

and its inverse

$$\mathcal{F}^{-1}\left[\tilde{\phi}\right]\left(\vec{x}\right) = \tilde{\phi}\left(\vec{x}\right) := \int_{\mathbb{R}^n} \mathrm{d}V_u \left(\tilde{\phi}\left(\vec{u}\right) \exp\left\{2\pi i \vec{u}^{\mathrm{T}} \vec{x}\right\}\right).$$

The *Fourier-transform* is an automorphism in the space of tempered distributions $S'\left(\mathbb{R}^n\right)$. Thus, a *Fourier-transform* of an element of $S'\left(\mathbb{R}^n\right)$ will still be an element of $S'\left(\mathbb{R}^n\right)$. With this formulation of $\mathcal{F}$ and $S'\left(\mathbb{R}^n\right)$, we can write down that the *Fourier-transform* of a distribution $\partial \in S'\left(\mathbb{R}^n\right)$ is

$$\left\langle \tilde{\partial}, \phi \right\rangle := \left\langle \partial, \tilde{\phi} \right\rangle$$

which is an isomorphism on $S'\left(\mathbb{R}^n\right)$. This is used to calculate

$$\left\langle \tilde{\delta}_0, \phi \right\rangle = \left\langle \delta_0, \tilde{\phi} \right\rangle = \tilde{\phi}\left(0\right) = \int_{\mathbb{R}} \phi\left(x\right) \mathrm{d}x = \left\langle 1, \phi \right\rangle$$

which implies that the *Fourier-transform* of the *Dirac delta distribution* is a constant

$$\Rightarrow \tilde{\delta}_0 \equiv 1.$$

Keep in mind that this identity is a *weak* identity, since we only showed that it holds true in the brackets $\left\langle \tilde{\delta}_0, \phi \right\rangle$. This is sufficient, since the δ-distribution is an integral operation by design (see Equation 2.1.8).

2.1.4 Fourier slice theorem

Before we arrive at an analytical solution to the problem statement Equation 2.1.3, we need to draw a connection between the line integrals of an object and its *2D-Fourier-transform.* We consider our measurements ϵ_i ($i = 1, ..., N_i$) of an unknown object $\mu\left(x, y\right) : \mathbb{R}^2 \rightarrow \mathbb{R}$ to be line integrals of the form of Equation 2.1.2. In contrast to Figure 2.1.2, we assume our source to be very far away. Hence it produces beams in a parallel manner. In doing so, we can use the parameters D, θ as in Figure 2.1.6 to characterize the 2D line

$$\mathcal{L}_{\theta,D} = \left\{x \cos\theta + y \sin\theta = D \,|\forall x, y \in \mathbb{R}\right\} \tag{2.1.9}$$

and rewrite the measurements from 2.1.2 as

$$\epsilon_{\theta,D} = \int_{\mathcal{L}_{\theta,D}} \mu\left(s_{\theta,D}\right) \mathsf{d}s_{\theta,D} = \int_{\mathbb{R}^2} \mu\left(x, y\right) \delta\left(x \cos\theta + y \sin\theta - D\right) \mathsf{d}x\mathsf{d}y =: P_\theta\left(D\right) \tag{2.1.10}$$

using the *Dirac delta distribution*[12] δ that is characterized by the property

$$\int \mathsf{d}X\, \delta(X - X_0) f\left(X\right) = f\left(X_0\right). \tag{2.1.11}$$

We call $P_\theta\left(D\right)$ the *Radon transform* of $\mu\left(x, y\right)$. For constant θ, the measurements represent the parallel projection of the object μ for that specific angle θ. An interesting property of $P_\theta\left(D\right)$ is that its 1D *Fourier-transform* (1DFT) is equal to the slice of the 2DFT of μ at the angle θ.

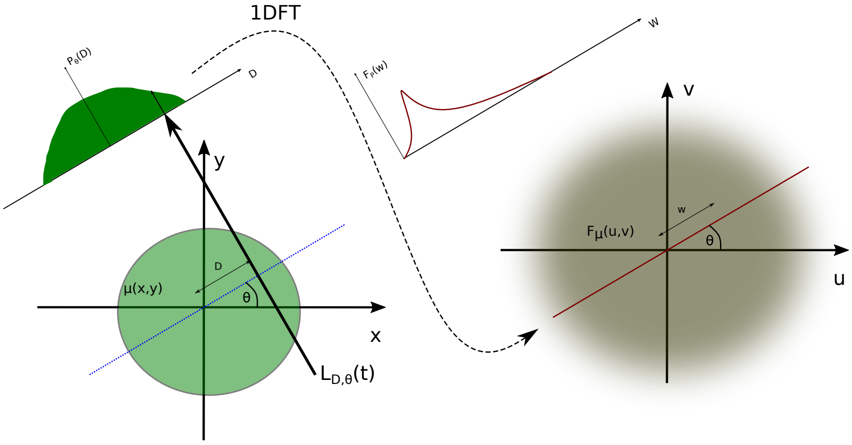


Figure 2.1.6: The geometry of parallel beam setup shows an important relation. The 1DFT of the projection with the corresponding angle θ is equal to the line of the same angle in the *2D Fourier space*.

To show this property, we write down the definition of the 2DFT of μ first.

$$F_\mu\left(u, v\right) := \int_{\mathbb{R}^2} \mathsf{d}x\mathsf{d}y \left\{\mu\left(x, y\right) \exp\left[-i2\pi\left(ux + vy\right)\right]\right\}$$

The 1DFT of $P_\theta\left(D\right)$ is defined as

$$F_P\left(w, \theta\right) := \int_{\mathbb{R}} \mathsf{d}D \left\{P_\theta\left(D\right) \exp\left[-i2\pi w D\right]\right\}. \tag{2.1.12}$$

To see that the union set of $F_P\left(w\right)$ of each pair $\left(D, \theta\right)$ represents the full set of $F_\mu\left(u, v\right)$ for each $\left(u, v\right)$, we need to identify w as the radius in the *Fourier space*

$$\begin{aligned} u &= w \cos\theta \\ v &= w \sin\theta \end{aligned}\;. \tag{2.1.13}$$

---

[12] see Subsection 2.1.3

Plugging that into the definition of $F_P(w)$ and using *Fubini's theorem* results in:

$$\begin{aligned}
F_P(w,\theta) &= \int_{\mathbb{R}} \mathsf{d}D \left\{ \left[\int_{\mathbb{R}^2} \mu(x,y)\,\delta(x\cos\theta + y\sin\theta - D)\,\mathsf{d}x\mathsf{d}y \right] \exp\left[-i2\pi w D\right] \right\} \\
&= \int_{\mathbb{R}^2} \mathsf{d}x\mathsf{d}y \int_{\mathbb{R}} \mathsf{d}D \left[\mu(x,y)\,\delta(x\cos\theta + y\sin\theta - D)\right] \exp\left[-i2\pi w D\right] \\
&\overset{\text{eq. }\underline{2.1.11}}{=} \int_{\mathbb{R}^2} \mathsf{d}x\mathsf{d}y\mu(x,y) \exp\left[-i2\pi w\left(x\cos\theta + y\sin\theta\right)\right] \\
&= \int_{\mathbb{R}^2} \mathsf{d}x\mathsf{d}y\mu(x,y) \exp\left[-i2\pi\left(xw\cos\theta + yw\sin\theta\right)\right] \\
&\overset{\text{eq. }\underline{2.1.13}}{=} \int_{\mathbb{R}^2} \mathsf{d}x\mathsf{d}y\mu(x,y) \exp\left[-i2\pi\left\{xu + yu\right\}\right] = F_\mu(u,v)\,. \qquad (2.1.14)
\end{aligned}$$

Equation 2.1.14 is called the *Fourier slice theorem.* An important result of 2.1.14 is that one could solve the reconstruction problem 2.1.2 by the inverse 2DFT of the 1DFT of the *Radon transform* 2.1.12, which is given by the measurements ϵ_i. Thus, given infinite angles would make that analytically possible, but still not feasible (due to infinite amount of data).

To make use of the practical implications of the well known *fast Fourier transforms*, typically one needs equidistant points in the 2D *Fourier plane.*With our measurement setup, we obtain finite and angular equidistant measurements as drawn in Figure 2.1.7. Those practical implications result in totally different solution approaches. Note the following property of $F_\mu(u,v) = F_\mu(w\cos\theta, w\sin\theta)$ in polar coordinates holds true due to the definitions of $\sin x$ and $\cos x$:

$$F_\mu\left(w\cos(\theta+\pi), w\sin(\theta+\pi)\right) = F_\mu\left(-w\cos\theta, -w\sin\theta\right). \qquad (2.1.15)$$

2.1.5 Filtered back projection (FBP)

The most common reconstruction method for CT is the filtered back projection algorithm (FBP). It is derived analytically and is inherently efficient, thus can be executed in a very short timeon simple CPU hardware. The derivation of the FBP from Equation 2.1.14 is a good example. It shows how to attain a drastically different algorithm by reformulating the fundamental expressions for the underlying theory.

If we take a look at Figure 2.1.7 again, we will recognize that the different lines we measure in the space domain represent nearly independent regions in the *Fourier plane*. The higher the frequency $\sqrt{u^2+v^2}$, the more independent the sample points of the measurements become. Keep in mind that the one dimensional *Fourier-transform* of the projection data represents one line

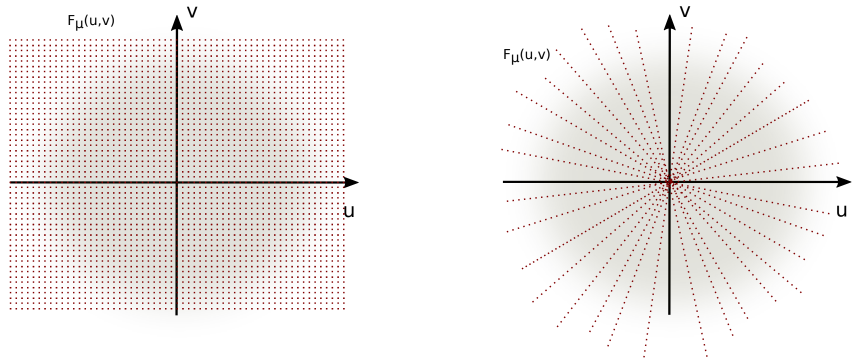


Figure 2.1.7: The left image shows the sampling of the *Fourier space* needed for a reconstruction without sampling artifacts. The right image shows the radial sampling that is achieved in typical CT setups.

in the *Fourier space*. If we “fill” this space with our measurements of all available angles (not an infinite amount any more) and assume that the “untouched” area has the value zero, we will be able to reconstruct the image by simply performing the inverse 2D *Fourier tansformation* of that data.

This dropping to zeroes will distort the reconstruction even more than the lack of equidistant data points, but the forward-projection of that kind of reconstruction will still be in good agreement with the measurements [FBKR19]. Thus, the simple “back transformation” has practical limitations.

To arrive at the FBP algorithm, we rewrite the problem to calculate μ as the inverse *Fourier-transform* and use polar coordinates as well as the Equation 2.1.15

$$\mu\left(x,y\right)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}F_{\mu}\left(u,v\right)\exp\left[i2\pi\left(ux+vy\right)\right]\mathsf{d}u\mathsf{d}v$$

$$=\int_{0}^{2\pi}\int_{0}^{\infty}F_{\mu}\left(w\cos\theta,w\sin\theta\right)\exp\left[i2\pi w\left(x\cos\theta+y\sin\theta\right)\right]w\mathsf{d}w\mathsf{d}\theta$$

$$=\int_{0}^{\pi}\int_{0}^{\infty}F_{\mu}\left(w\cos\theta,w\sin\theta\right)\exp\left[i2\pi w\left(x\cos\theta+y\sin\theta\right)\right]w\mathsf{d}w\mathsf{d}\theta$$

$$+\int_{\pi}^{2\pi}\int_{0}^{\infty}F_{\mu}\left(w\cos\theta,w\sin\theta\right)\exp\left[i2\pi w\left(x\cos\theta+y\sin\theta\right)\right]w\mathsf{d}w\mathsf{d}\theta$$

$$=\int_{0}^{\pi}\int_{0}^{\infty}F_{\mu}\left(w\cos\theta,w\sin\theta\right)\exp\left[i2\pi w\left(x\cos\theta+y\sin\theta\right)\right]w\mathsf{d}w\mathsf{d}\theta$$

$$+\int_0^\pi \int_0^\infty F_\mu\left(w\cos\left(\theta+\pi\right), w\sin\left(\theta+\pi\right)\right)\exp\left[i2\pi w\left(x\cos\left(\theta+\pi\right)+y\sin\left(\theta+\pi\right)\right)\right] w\mathsf{d}w\mathsf{d}\theta$$

$$\overset{\text{eq. 2.1.15}}{=} \int_0^\pi \int_0^\infty F_\mu\left(w\cos\theta, w\sin\theta\right)\exp\left[i2\pi w\left(x\cos\theta+y\sin\theta\right)\right] w\mathsf{d}w\mathsf{d}\theta$$

$$+\int_0^\pi \int_0^\infty F_\mu\left(-w\cos\theta, -w\sin\theta\right)\exp\left[i2\pi\left(-w\right)\left(x\cos\theta+y\sin\theta\right)\right] w\mathsf{d}w\mathsf{d}\theta$$

$$= \int_0^\pi \int_0^\infty F_\mu\left(w\cos\theta, w\sin\theta\right)\exp\left[i2\pi w\left(x\cos\theta+y\sin\theta\right)\right] w\mathsf{d}w\mathsf{d}\theta$$

$$+\int_0^\pi \int_{-\infty}^0 F_\mu\left(w\cos\theta, w\sin\theta\right)\exp\left[i2\pi w\left(x\cos\theta+y\sin\theta\right)\right]\left(-w\right)\mathsf{d}w\mathsf{d}\theta$$

$$= \int_0^\pi \int_{-\infty}^\infty F_\mu\left(w\cos\theta, w\sin\theta\right)\exp\left[i2\pi wD\left(\theta;x,y\right)\right]\left|w\right|\mathsf{d}w\mathsf{d}\theta$$

$$\overset{\text{eq. 2.1.14}}{=} \int_0^\pi \left\{\int_{-\infty}^\infty F_P\left(w,\theta\right)\left|w\right|\exp\left[i2\pi wD\left(\theta;x,y\right)\right]\mathsf{d}w\right\}\mathsf{d}\theta. \qquad (2.1.16)$$

Here we expressed D of Equation 2.1.9 as a function of θ and dependent on the parameters x, y. The above integral 2.1.16 can be expressed as

$$\mu\left(x,y\right) = \int_0^\pi \tilde{F}_\theta\left(x\cos\theta+y\sin\theta\right)\mathsf{d}\theta \qquad (2.1.17)$$

where

$$\tilde{F}_\theta\left(D\right) := \int_{-\infty}^\infty F_P\left(w,\theta\right)\left|w\right|\exp\left[i2\pi wD\right]\mathsf{d}w. \qquad (2.1.18)$$

The approximation of $\mu\left(x,y\right)$ becomes feasible when the transformed projection data $F_P\left(w,\theta\right)$ is given according to Equation 2.1.18. This integral can be interpreted as the *Fourier-transform* of $F_P\left(w,\theta\right)$ multiplied by the ramp filter $|w|$, which is basically a filtering operation that is back projected into the space domain. This coins the name “filtered back projection”. The operation described by 2.1.17 can be verbally expressed as:

The filtered projection $\tilde{F}_\theta\left(D\right)$ gives a constant contribution to the reconstructed image μ along the line $D = x\cos\theta + y\sin\theta$. Thus, to approximate the value of μ at the point $\left(x,y\right)$, one needs to back project the filtered measurements $\tilde{F}_\theta$ over the range of 180°.

Note that we derived a solution where the quality of the reconstruction still depends on the angular resolution, but does not need the dropping of unkown points in the *Fourier space* to zero or an interpolation step of those unknown points of $F_\mu\left(u,v\right)$.

For the sake of completeness, there are some crucial practical implications worth mentioning, considering the feasibility of the FBP algorithm stated in Equation 2.1.17.

The *Fourier transform* of a product of two functions is equivalent to the convolution of the two *Fourier transformed* functions in the dual space. This is named the "*convolution theorem*". It becomes even more relevant, because one of the two functions is already a *Fourier transform*, namely $F_P(w, \theta)$. Hence, if we had the *Fourier transform* of $|w|$ we could work in the projection domain. In fact, the 1DFT of $|w|$ diverges at $w = 0$.

Nevertheless, we could find a "dirty version" of that transformation if we clamp the integration boundaries from infinity to a real number. In practice, the filter $|w|$ along the boundaries of the integral 2.1.18 is limited due to the sampling theorem anyway[13]. This gives us the opportunity to simplify the problem 2.1.17 further, at least from an algorithmic point of view. Just to mention, the limitation of the spatial frequency w is given by the maximum frequency of the measured projections, and hence depends on the sampling of D. Finally, this is limited by the size of the detector pixels. Wu et al. convincingly suppressed sampling artifacts caused in an FBP setting [WMYF15, MMMF15].

Additionally, we assumed to have a measurement setup that gives the line integrals of parallel lines $\mathcal{L}_i$ in such a way that we can go through the whole projection by varying the parameter D at a fixed angle θ. If we consider our point sized X-ray source that creates an isotropic distribution of radiation at a finite distance from the object, we have to deal with the so-called fan beam geometry[14]. Theoretically, we could re-sort our data set to produce parallel data. In practice, that is not viable due to the limited sampling of D.

The 3D cone-beam generalization of the FBP algorithm is called the FDK algorithm (named after Feldkamp LA, Davis LC, and Kress JW) [FDK84a]. The beauty of the analytical solution of Equation 2.1.5 is that one gets a better understanding of the reconstruction artifacts that are caused by the violation of the ideal circumstances. Such analytical methods are limited by the constraints of the measurement setup (e.g. the physics of the absorption process or NYQUIST limit of the data sampling) and specific requirements such as the source trajectory. The following sections adress the details which lead to the more sophisticated PSR algorithm described in Section 2.4.2.

Optional side note on the connection to Kaczmarz

The following derivation is not mathematically accurate for the sake of brevity. The actual derivation needs to discuss more details about the attributes of the system matrix and convergence properties [Nat86].

---

[13] The sampling theorem gives an upper boundary for the frequency component of a signal. If the detector has the pixel size Υ, the maximum frequency of that signal that can be reconstructed is $f_\Upsilon = (2\Upsilon)^{-1}$.

[14] Sir Godfrey Newbold Hounsfields first scanner was working with parallel beams back in 1975.

This side note supports the reader with more background and supports the intuition about fundamental properties of the mathematics behind the iterative reconstruction approaches.

In CT, the inversion of the problem $A\vec{\mu} = \vec{\epsilon}$ with the pseudo inverse $A^{\dagger}$

$$\vec{\mu} = A^{\dagger}\vec{\epsilon} = A^{\mathrm{T}}\left(AA^{\mathrm{T}}\right)^{-1}\vec{\epsilon}$$

cannot be performed due to the size[15] of A. Nevertheless, in the case of 2D parallel beam geometry, $\left(AA^{\mathrm{T}}\right)^{-1}$ can be identified as the convolution with the ramp filter and A^{T} is the back projection. Thus, a discretized FBP is the analytical solution to the problem statement above.

Rewriting Equation 2.1.5

$$\vec{\mu}^{*} = \underset{\mu_j \geq 0, \forall j}{\arg\min} \sum_{i=1}^{N_i} \left[\sum_{j=1}^{N_j} a_{ij}\mu_j - \epsilon_i\right]^2$$

which is equivalent of finding the

$$\min_{\mu_j \geq 0, \forall j} \quad \frac{1}{2}\left|\vec{\epsilon} - A\vec{\mu}\right|^2 .$$

Differentiating gives

$$0 = A^{\mathrm{T}}\left(\vec{\epsilon} - A\vec{\mu}\right).$$

Interestingly we can interpret this expression as a fix-point $\frac{\mathrm{d}}{\mathrm{d}t}\vec{\mu} = 0$ of the system of ordinary differential equations[16] with the inhomogeneity $A^{\mathrm{T}}\vec{\epsilon}$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\vec{\mu} = A^{\mathrm{T}}\left(\vec{\epsilon} - A\vec{\mu}\right).$$

On the other hand, Kaczmarz yields (see Equation 2.1.6):

$$\vec{\mu}_{k+1} = \vec{\mu}_k + \vec{a}_{i_k}^{\mathrm{T}} \cdot \frac{\epsilon_i - \vec{a}_{i_k} \cdot \vec{\mu}_k}{\vec{a}_{i_k} \cdot \vec{a}_{i_k}^{\mathrm{T}}}$$

$$\Longrightarrow \frac{\vec{\mu}_{k+1} - \vec{\mu}_k}{1} = \left(\kappa_{i_k}\right)^{-1} \cdot \vec{a}_{i_k}^{\mathrm{T}} \cdot \left\{\epsilon_i - \left(\vec{a}_{i_k} \cdot \vec{\mu}_k\right)\right\}.$$

The left-hand side of this equation can be interpreted as a discretized differential quotient with time step length of 1. The right-hand side is similar to the right-hand side $\frac{\mathrm{d}}{\mathrm{d}t}\vec{\mu} = A^{\mathrm{T}}\left(\vec{\epsilon} - A\vec{\mu}\right)$ divided by a weighting that is dependent on the geometry factors of the i-th ray $\kappa_{i_k} = \vec{a}_{i_k} \cdot \vec{a}_{i_k}^{\mathrm{T}}$. Note that $\vec{a}_{i_k}$ is the i-th row of A, thus κ_{i_k} is a scalar. This constant could be factored in at $0 = A^{\mathrm{T}} (\vec{\epsilon} - A\vec{\mu})$ and can be interpreted as a so-called relaxation parameter.

---

[15] Due to $\dim\left(\vec{\epsilon}\right) > \dim\left(\vec{\mu}\right)$, $\left(AA^{\mathrm{T}}\right)^{-1}$ does not exist in a typical CT setting either, but it still helps to see the connection established in this subsection.

[16] Technically speaking, the problem statement includes the non-negativity constraint on $\vec{\mu}$ which renders the problem non-linear. A general discussion about the non-negativity constraint can be found in Haase et al. [HHS+19].

Alternatively, the above can be derived as a special case of the LANDWEBER iteration [Lan51] with ordered subsets, specifically when each ray is a single update and the relaxation parameter is equal to κ_{i_k}.

2.2 X-ray generation

This section will give a brief introduction about the commonly used X-ray sources. X-rays are a side product of the so-called *bremsstrahlung*. Figure 2.2.1 is a schematic drawing of an X-ray tube. The procedure to generate X-rays is the following:

- A cathode and an anode are placed within a vacuum tube.
- Tube voltage up to $U = 125\text{kV}$ is applied between them[17].
- A cathode creates free electrons (electric charge of $q = -e$) by either ionizing the air around (cold cathode tube, pressure: $10^{-1} \sim 10^{-3}\text{Pa}$) or ionizing itself (hot cathode tube, pressure: 10^{-4}Pa).
- These electrons are accelerated towards the anode reaching it with a kinetic energy of eU.
- Colliding electrons interact with the anode material (typically made of tungsten), i.e. with other electrons, nuclei and ions.
- 1% of their initial kinetic energy is lost due to the deceleration of the Coulomb field of the particles of the anode.
- The deceleration emits the X-ray photons to a maximum energy of eU.
- The rest is heating the anode[18].
- The tube is covered in a box made of lead. A window that is typically made of glass or beryllium foil helps to aim the radiation.
- Depending on the purpose, certain filters and collimators are used to design the X-rays in a desired manner.

Low energy photons are very likely to be absorbed by tissue. To avoid unnecessary dose, filters made of Al and Cu are commonly applied to reduce low energy photons that are not delivering contrast information in the

---

[17] This is the maximum tube voltage for long term use of the SIEMENS Artis Zeego system depicted in Figure 1.0.2.

[18] This requires a serious amount of cooling and clever technical applications.

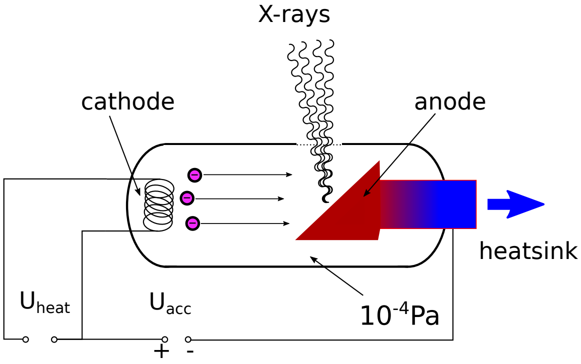


Figure 2.2.1: Schematic drawing of an X-ray tube is shown.

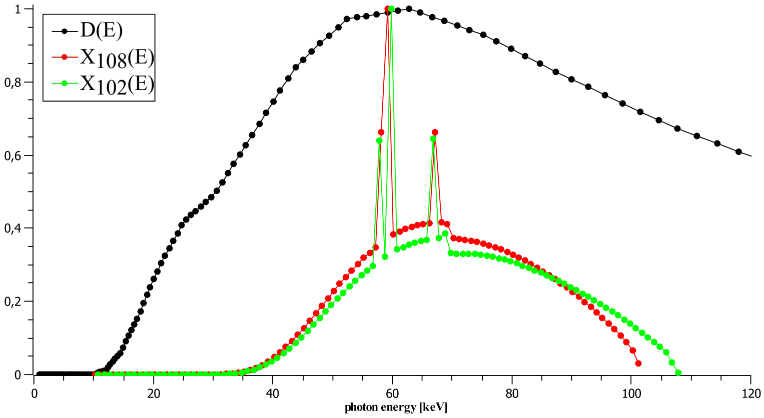


Figure 2.2.2: Depiction of two X-ray spectra (calculated by spekCalc and normalized by its maximum value) and the detector response function by Roberts et al. [RHN^{+}08] used in this work can be seen.

later projection. Figure 2.2.2 shows the spectra used within our algorithmic framework. Characteristic peaks in X-ray spectra result of the so-called K_α- and K_β-radiation. These are caused by electrons that are dropping from orbits of higher energy to lower energy levels. As these energy differences are material-specific the emitted photons are called characteristic radiation.

2.3 Polychromatic photon absorption law

This subsection is based on the book of Buzug [Buz08] and the book of Barrett and Swindell [BS]. It provides all interactions of photon-matter-interaction, which lead to the X-ray measurements.

The linear absorption coefficient μ in Equation 2.1.1 consists of different photon-matter interaction terms. In general, those interactions change the energy of the photon and its direction. In this chapter, we list a brief description of the interactions that are dominant at the energy domain relevant for X-ray imaging and describe the main characteristics of how and why those alter the X-ray photon in direction and energy. We recall that the energy $E_\gamma = hf$ of the photon is a function of its frequency, hence if the photon undergoes a change in f, its corresponding energy is changing. We can write the linear attenuation coefficient as

$$\mu = \frac{\rho N_A}{M} \cdot \sigma_{\text{total}} = n \cdot \sigma_{\text{total}}.$$

Here, ρ is the density and M is the atomic weight of the corresponding material. N_A is the Avogadro constant. It states that the attenuation of the incident photons depends on the number of target atoms per volume $n := \frac{\rho N_A}{M}$ and the total photon atomic cross-section σ_{total}. The different interactions can be separated into absorption and/or scattering. The latter typically changes the direction of single photons, resulting in blurred X-ray projections. Note that this chapter contains equations which are mentioned only for the sake of qualitative understanding of the different effects that take place and defy the simple idea of mono-energetic photon absorption to be the correct model for the description of X-ray imaging. The relevant photon energy lies in the interval of $10^{-2} \sim 10^{-1}$ MeV(see Figure 2.2.2). The impact of the different attenuation mechanisms is depicted in Figure 2.3.1 for the case of bone and in Figure 2.3.2 for different reference materials.

2.3.1 Rayleigh scattering

If the order of the wavelength of the incident photon is above the diameter of the nucleus, the most dominant scatter effect is the elastic (no energy loss) Rayleigh scattering. Here, the photon gets absorbed by an outer electron which then jumps back into its bounded state and emits the exactly the same amount of energy as electromagnetic radiation (a photon) into a new direction. This effect can be described by the classical field theory of electromagnetism [Fli08] or [LDL75] and does not take into account any quantum properties of the participating particles. For completeness we note that the cross-section of this scatter event is

$$\sigma_{\text{Rayleigh}} = \frac{8}{3}\pi r_e^2 \frac{f^4}{\left(f^2 - f_0^2\right)^2 + \Gamma^2 f^2}$$

with the classical electron radius $r_e = \frac{e^2}{4\pi\varepsilon_0 m_e c^2}$, ε_0 the permittivity of the vacuum, m_e the electron rest mass and Γ a dissipation coefficient of the oscillation of the electron and $\Gamma/f_0 < 0.3$ for common cases. Here, f_0 is the natural frequency of the participating electron. Rayleigh scattering has one special case which is called the Thompson scattering. If the electron is only weakly bound to the atom or completely unbounded, the respective

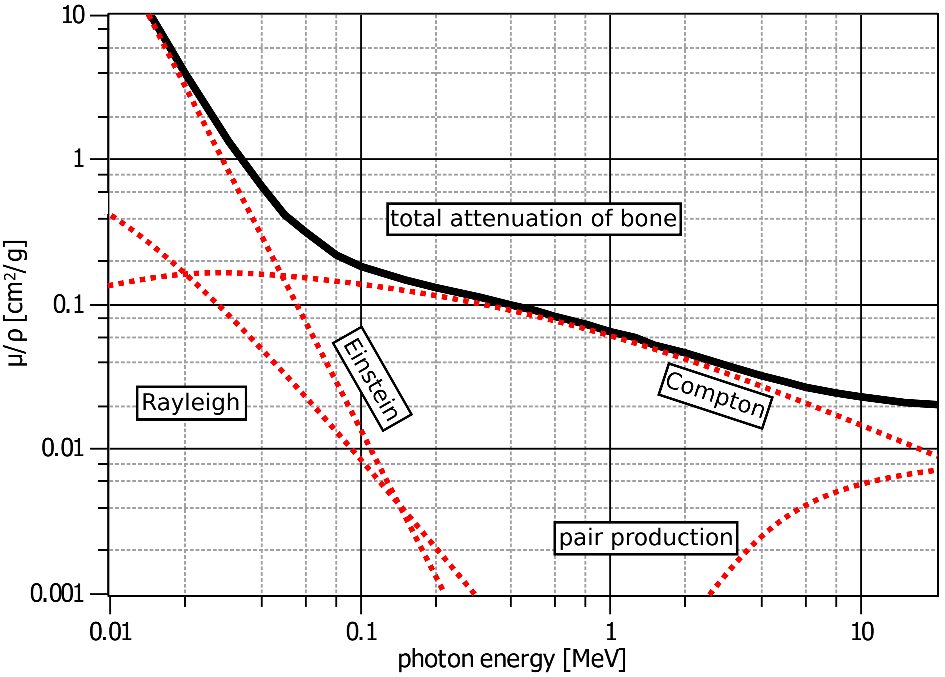


Figure 2.3.1: This is a schematic drawing of the total attenuation of cortical bone as an example material as a sum of the different interactions of the respective energy interval.

f_0 approaches zero and σ_{Rayleigh} becomes a constant. For higher photon energies ($f \gg f_0$), σ_{Rayleigh} becomes a constant as well. This scattering effect is most dominant in the domain of the visible photon spectrum.

$$\frac{\sigma^{\text{blue}}_{\text{Rayleigh}}}{\sigma^{\text{red}}_{\text{Rayleigh}}} = \left(\frac{f^{\text{blue}}_{\text{Rayleigh}}}{f^{\text{red}}_{\text{Rayleigh}}} \right)^4 \approx 10$$

Rayleigh scattering is the reason why we can perceive clouds, but not the water vapor in the air and why the sky is blue at day time and red in the evening[19]. The X-ray interactions of the following subsections take the quantum properties of photons and matter into account and consider higher photon energies.

2.3.2 Photoelectric absorption

If the photon energy is above the binding energies of atomic electrons, the photon can be absorbed entirely by such an atom. This process ionizes the

---

[19] The evening also shows a blue sky, but the respective effect that is dominant at this domain is called Chappuis absorption.

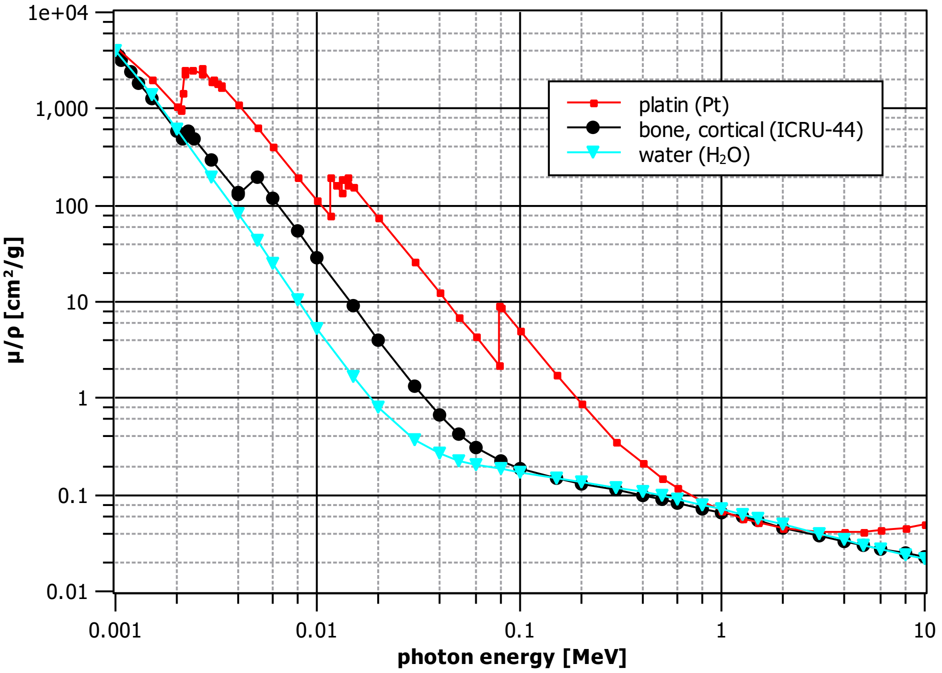


Figure 2.3.2: An overview of three example materials and its energy dependent attenuation properties taken from the NIST data base [HS96].

atom, leaves a free electron with kinetic energy equal to the difference of binding energy and photon energy. This lost energy can be observed as heat on the local lattice of the material. Electrons of outer shells or the free electron energy band are then fill the vacant electron spot. The recombination emit is characteristic radiation of specific energy levels depending on the absorbing material. If the emitted photon has enough energy to break an outer electron from outlying orbitals, this creates a free and mono-energetic Auger electron. The absorption coefficient μ_{Einstein} of the photoelectric absorption has the important dependence

$$\mu_{\text{Einstein}} \propto Z^n f^{-3},$$

where n lies between 4 and 5 [Dav65]. The dependence of Z^n has practical consequences for applications that utilize contrast agent, e.g. iodine and barium ($Z_{\text{I}} = 53$,$Z_{\text{Ba}} = 56$) or X-ray shielding with lead ($Z_{\text{Pb}} = 82$). The dependence of f^{-3} shows the rapid decay for higher energies. At high X-ray energies the following processes become more relevant.

2.3.3 Compton scattering

The most dominant scatter event in the current X-ray imaging domain is the COMPTON scattering process. Weaker bound electrons (i.e. the orbit that correspond to the smallest binding energy) can interact with X-ray photons by colliding and transfering only a proportion of the incident energy to their kinetic energy. Thus, the process is almost independent on the actual material, but dependend on the density ρ of the material the photon is propagating through. The lost photon energy is in direct correlation to the scattering angle. The important relation comes down to

$$\Delta\lambda = \frac{h}{m_e c} \left(1 - \cos\Delta\theta\right),$$

where $\Delta\theta$ is the angle of the scattered photon that increases $\Delta\lambda$ of its wavelength during that process, hence decreases the photon energy $E_\gamma = hf \propto \lambda^{-1}$. In literature one can find that the constant $\frac{h}{m_e c}$ is referred to as the COMPTON wavelength.

The ratio of scattered X-ray photons can be derived by the KLEIN-NISHINA formula[20], so we can state that the cross-section of this process is

$$\sigma_{\text{Compton}} = 2\pi r_e^2 \left\{ \frac{1+\xi f}{(\xi f)^2} \left(2\frac{1+\xi f}{1+2(\xi f)^2} - \frac{\ln(1+2\xi f)}{\xi f} \right) + \frac{\ln(1+2\xi f)}{2\xi f} - \frac{1-3\xi f}{(1+2\xi f)^2} \right\}$$

where $\xi = \frac{h}{m_e c^2}$ that could be called the inverse COMPTON frequency. With this, we can write down the attenuation coefficient as

$$\mu_{\text{Compton}} = n \cdot \sigma_{\text{Compton}}.$$

2.3.4 Pair production

If the photon energy E_γ is above $E_{\text{Pair}} = 1.022\,\text{MeV} = 2m_e c^2$ the X-ray photon can interact with the electromagnetic field of an electron or a nucleus to create an electron-positron pair with the combined kinetic energy of $E_{\text{kin}} = E_\gamma - E_{\text{Pair}}$. The probability increases proportionally to E_γ and field-strength. This effect is called **pair production**. Since the diagnostic energies of X-ray imaging lie in the order of $10^{-2} \sim 10^{-1}\,\text{MeV}$(see 2.2.2), this process is not of importance for the typical μ of our problem statements but it is worth mentioning that the side effect, called pair annihilation, becomes important in positron emission tomography (PET) imaging. Since the positron will meet another electron after a very short time, they will disintegrate and produce a pair of photons, traveling 180° away from each other.

---

[20] Equation 2.188, page 155 in [LR09]

2.3.5 Detector technology

As Cunningham and Judy [CJ00] state, the X-ray detector used in a CT setup needs

- to have a high overall efficiency to reduce the dose (offer a large dynamic energy range),
- to be stable over time,
- and to have a constant sensivity with respect to temperature variations (e.g. in the gantry).

The so-called geometric efficiency, quantum efficiency, and conversion efficiency are the important factors that ensure the high overall efficiency of the X-ray detector. The higher the ratio of sensitive to unsensitive area of the detector element becomes, the more its geometric efficiency increases. Having already introduced all photon-matter interactions in the subsections above, we can easily recognize that the fraction of incident X-rays on the detector element contributing to the measured signal is a statistical property of the material and the photon energy.

The quantum efficiency is the number of such photons that contribute to the measured signal.

Finally, the conversion efficiency quantifies the response that accurately converts the absorbed X-ray signal into an electrical signal. The overall efficiency is basically the product of these three numbers and it usually lies between 0.45 and 0.85. Of course, a number that is smaller than 1.0 refers to an imperfect measurement system, which increases the dose needed to maintain a certain CT image quality. For this reason, this number is called dose efficiency as well.

Gas ionization detectors are stacked chambers that contain compressed gas. Xenon is commonly used with a pressure up to 30 bar. The X-ray photons ionize the gas, which is then collected by a high voltage tungsten septum. These kind of detectors have a lower quantum efficiency than solid-state detectors.

Solid-state detectors consist of an array of **scintillator crystals**. They produce visible light photons when absorbing an X-ray photon, which is then collected by a CCD chip. The X-ray photon is producing an electron with high kinetic energy, which excites other electrons and recombines with another ion. Eventually the recombination emits light photons that are measured to create the signal. Commonly used materials are either cadmium tungstate or Gd_2O_2S ceramic made of rare earth oxides. The reason for the much higher quantum efficiency is reflected by the pure density that the solid materials have, which are three magnitudes higher than that of the compressed gas.

The average obese person that is scanned can absorb X-ray intensities I/I_0 to ranges below 10^{-4}. Thus, an accurate data-acquisition needs to encode floating point precision of at least 10^{-5}. Due to the necessary log conversion, this number is dramatically insufficient in the presence of highly absorbent materials as metal, as there will be almost no X-ray quanta measured. Manufacturers use precision preamplifiers, current-to-voltage converters, analog integrators, multiplexers and analog-to-digital converters to transmit the values to the reconstruction system. Depending on the manufacturer the logarithmic operation (Equation 2.1.1) is performed by either an analog log amplifier or a carefully sampled lookup table.

It is worth mentioning that all detectors suffer from two effects, namely vignetting and distortion.

Vignetting (or veiling glare) is a systematic decrease in brightness towards the outer boundaries of the detector. It is caused by scattering of the light photons emitted by the scintillator crystals. Scattered light photons are distributed isotropically, but as they are more likely to get measured within the center of the detector, the probability of these corrupted measurements decreases with the distance to that central point. As this problem occurs in optical images as well and it is dependent on the measurement itself, it is an ongoing topic of interest in contemporary research [KPB16, YLS06, KW00]. Within the domain of CT, this problem is tackled with calibration [NBC⁺15]. Distortion due to external electromagnetic field strengths[21] needs to be measured and corrected by image processing as well.

---

[21] In addition, geometric distortion [WLDK12] and pincushion distortion need to be tackled as well [LMX10].

2.3.6 Beam hardening

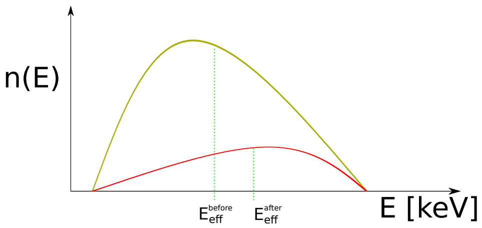


Figure 2.3.3: Schematic depiction of an incident X-ray spectrum (green) $n(E)$ in arbitrary units. After propagation through matter the effective energy of the spectrum (red) is shifted to a higher energy $E_{\text{eff}}^{\text{after}}$. Thus, the propagation process made the beam *harder*.

As photons of lower energy are more likely to be attenuated (see Figure 2.3.2), propagation of a beam of photons that contains an X-ray spectrum of different energies through matter will always lead to a shift of the effective energy. The mean energy of the spectrum increases. Hence, the X-ray beam loses energy but gets *harder* during the process.

Equation 2.1.2 models the measurements ϵ along a line $\mathcal{L}$ to be dependent of the photon energy E_0. Thus, the ratio of the measured X-ray intensity $I\left(E_0\right)$ to the incident X-ray intensity $I_0\left(E_0\right)$ of photon energy E_0 follows the Lambert-Beer absorption law

$$\frac{I\left(E_0\right)}{I_0\left(E_0\right)} = \exp\left\{-\int_{\mathcal{L}} \mu\left(\vec{x}, E_0\right) \mathsf{d}l\right\}. \tag{2.3.1}$$

If the X-ray source only emits photons of one energy E_0, the forward model $\bar{\epsilon} = -\ln {}^{I(E_0)}/_{I_0(E_0)}$ would be a linear operator over the unknown object properties $\mu\left(\vec{x}, E_0\right)$. Since X-ray sources produce a broad spectrum of photons with different energies, this model becomes non-linear. This is because the measurements are in fact a sum over all energies. More precisely, the detector measures different energies by different gain values which result in an even more complicated signal. The monoenergetic absorption law holds true for a single energy. So the actual measured intensity $\bar{Y}$ sums over different energies E and can be written as

$$\bar{Y} = \int_{\mathbb{R}^+} D\left(E\right) \cdot X\left(E\right) \exp\left\{-\int_{\mathcal{L}} \mu\left(\vec{x}, E\right) \mathsf{d}l\right\} \mathsf{d}E \tag{2.3.2}$$

where $D\left(E\right)$ is the detector response[22] depending on the photon energy and $X\left(E\right)$ is the X-ray spectrum produced by the source. Thus, both parts contribute to what we can interpret as the incident intensity $I_0 = D \cdot X$. The extinction value $\bar{\epsilon}_{\text{poly}}$ would be calculated as

$$\bar{\epsilon}_{\text{poly}} = -\ln\frac{\bar{Y}}{\bar{Y}_0} = -\ln\left[\frac{\int_{\mathbb{R}^+} D\left(E\right) \cdot X\left(E\right) \exp\left\{-\int_{\mathcal{L}} \mu\left(\vec{x}, E\right) \mathsf{d}l\right\} \mathsf{d}E}{\int_{\mathbb{R}^+} D\left(E\right) \cdot X\left(E\right) \mathsf{d}E}\right]. \tag{2.3.3}$$

---

[22] The C-arm systems that provided measurements in this thesis feature a CsI scintillator-based detectors. Responses of NaI(Tl), CsI(Tl) and CsI(Na) detectors above 20keV behave qualitatively similarly [ABYZ67].

Even though the actual incident intensity is X alone, one can read the above equation like this: *The incident X-ray intensity X gets attenuated while propagating through the object μ and measured by an integrating detector D.*

To emphasize the consequence of such a non-linearity (in μ) of the projection operation, we write down the following "informal" equations and neglect the red parts at first:

$$\tilde{P}(\mu) := I = \int I_0(E) e^{-\int_0^\alpha \mu(x,E)\mathrm{d}x} \mathrm{d}E$$

$$\Rightarrow P(\mu; E) := \ln \frac{I_0(E)}{I(E)} = \int_0^\alpha \mu(x, E)\, \mathrm{d}x$$

$$\Rightarrow \tilde{P}(\mu_1 + \mu_2) \cancel{=} \tilde{P}(\mu_1) + \tilde{P}(\mu_2)\,. \tag{2.3.4}$$

Here, α is the thickness of the volume that is projected (in Figure 2.1.5 this means $\alpha = \triangle x$). Interpreting the projection operation P as a function of μ shows that the introduction of the red colored parts (Equation 2.3.2) destroy the linearity with respect to the projected volume. It is important to recognize that if $D(E) = D_0 \delta(E - E_0)$, the integrating detector would be the perfect bandpass filter for the photon energy E_0. Thus, one would reduce the projection problem to the linear structure of Equation 2.3.1. The very same would happen if we consider a monochromatic X-ray source $X(E) = X_0 \delta(E - E_0)$ and keep $D(E)$ arbitrary.

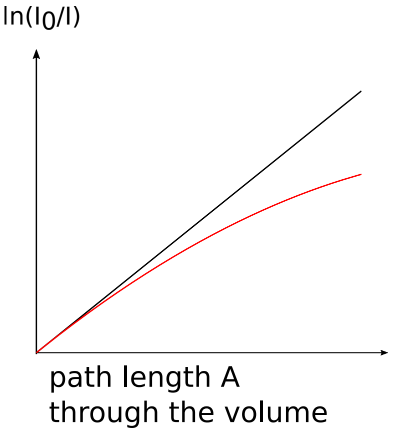


Figure 2.3.4: Schematic explanation for the non-linear behavior (beam hardening) of the measured values with respect to the propagation length.

To tackle the problem that is introduced by the non-linearity (beam hardening) of the projection that is depicted in Figure 2.3.4, one could try to find a method that maps the measured red signal to the theoretical signal one would expect from a monochromatic measurement. Such a mapping is called **water-correction.** Because the human body consists of water for the most part, it is the most commonly used idea to compensate for beam hardening. This method only needs a calibration with a appropriate water phantom.

In terms of non destructive testing of objects which are made out of a single material, this method leads to sufficient reduction of beam hardening. However, the human body consists at least of water-like tissue and bones. This distinction destroys the easy mapping approach, as a very thin slice of metal can have the same extinction value as a bottle of water, but would still need to be mapped to a different linear function in Figure 2.3.4 extinction values. A modern approach to separate these different types of extinction values is explained in Abdurahman et al. [AFB$^+$18] or in Cao et al. [CHF$^+$19].

2.4 Statistical reconstruction

This section introduces the theory about the reconstruction technique used in this work and gives an advanced version of the method described with Equation 2.1.5. It is based on the book of Buzug [Buz08]. The fundamental assumption of this kind of methods is that the measured number of X-ray photons n_i is a Poisson distributed random variable. This assumption is usually justified by all the covered physical effects in Subsection 2.3. Equation 2.3.1 motivates the expected value of the measurements to be as the following:

$$\bar{n}_i = n_0 e^{-\sum_j a_{ij}\bar{\mu}_j}. \tag{2.4.1}$$

The Poisson distribution P is defined as

$$P(n) = \frac{\bar{n}^n}{n!} e^{-\bar{n}}.$$

To calculate the probability to measure $\bar{n}_i$, under the circumstance that the volume has the values $\bar{\mu}_j$, one needs the product probability over all projections

$$P(\vec{n}|\vec{\bar{\mu}}) = \prod_i P(n_i|\bar{\mu}_j) = \prod_i \frac{\left(n_0 e^{-\sum_j a_{ij}\bar{\mu}_j}\right)^{n_i}}{n_i!} \exp\left\{-n_0 e^{-\sum_j a_{ij}\bar{\mu}_j}\right\} =: L\left(\vec{\bar{\mu}}\right).$$

The function L is called the likelihood function of $\vec{\bar{\mu}}$. If we find a way of maximizing L over $\vec{\bar{\mu}}$, we find the most consistent $\vec{\bar{\mu}}$ in agreement with the measurements. To find the maximum value of L, we first take its logarithm. Due to the monotonic behavior of the logarithm, a maximum of $\ln L$ will still be a maximum of L. We write down

$$\ln L\left(\vec{\bar{\mu}}\right) = \sum_i \left\{ n_i \ln\left(n_0 e^{-\sum_j a_{ij}\bar{\mu}_j}\right) - \ln(n_i!) - n_0 e^{-\sum_j a_{ij}\bar{\mu}_j} \right\}$$

and neglect the constant number $\ln(n_i!)$, because it has no impact on the demanded maximum, which together with Equation 2.4.1 leads to

$$\ln L\left(\vec{\bar{\mu}}\right) = \sum_i \left\{ n_i \ln\left(\bar{n}_i\right) - \bar{n}_i \right\}. \tag{2.4.2}$$

Note that this is only a function of the measurements n_i and the unknown volume $\bar{\mu}_j$. With $[A\vec{\mu}]_i$ denoting the i-th component of the vector $[A\vec{\mu}]$, we now assume that the measured monochromatic X-ray intensity Y_i^M is Poisson distributed as well. We can use Equation 2.4.2 and Equation 2.3.1 to find a desired L as follows:

$$Y_i^M \sim \text{Poisson}\left\{I_i e^{-[A\vec{\mu}]_i} + r_i\right\}$$

$$\Rightarrow L(\vec{\mu}) = \sum_i \left\{ Y_i \ln \left(I_i e^{-[A\vec{\mu}]_i} + r_i \right) - \left(I_i e^{-[A\vec{\mu}]_i} + r_i \right) \right\}. \tag{2.4.3}$$

Here, $A \in \mathbb{R}^{\text{number of voxels}\times\text{number of (pixels } \times \text{ views)}} =: \mathbb{R}^{N_j \times N_i}$ is the system matrix which represents the geometry factors $a_{ij} \geq 0$ in matrix form. In reality, we have to deal with noisy data (indicated by the noise term r_i), which usually leads to noisy reconstruction images. To mitigate those problems, we add a regularization term $R(\mu)$ characterized by — a yet to be defined — potential function $\psi(\cdot)$ to the likelihood function. We define

$$R(\vec{\mu}) := \sum_{j=1}^{N_j} \sum_{k \in \mathcal{N}_j} \psi(\mu_j - \mu_k) \tag{2.4.4}$$

which takes into account only a certain (usually the 26 surrounding voxels) neighborhood $\mathcal{N}_j$ around the voxel μ_j and arrive at the penalized likelihood function

$$\Phi(\vec{\mu}) := L(\vec{\mu}) - \beta R(\vec{\mu}). \tag{2.4.5}$$

The goal of the reconstruction technique becomes to maximize the function Φ under the condition that the reconstruction volume has only positive numbers of the linear attenuation coefficient μ

$$\hat{\vec{\mu}} = \underset{\mu_j \geq 0\, \forall j}{\arg\max}\, \Phi(\vec{\mu}). \tag{2.4.6}$$

2.4.1 Iterative monochromatic statistical reconstruction

This and the next subsection are derived from Elbarki and Fessler and show an in-depth view on their theoretical work published in [EF02]. In order to solve the problem Equation 2.4.6, we wish to have an easy re-formulation which is quadratic (easy to maximize) in the argument and separable (easy to parallelize) in the volume domain j. In order to make the reformulation separable, we will introduce the so-called surrogate functions. First, we will achieve the quadratic approximation as we start to rewrite Equation 2.4.3 and use the substitution $\hat{l}_i = \ln(\frac{I_i}{Y_i - r_i})$, $w_i = \frac{(Y_i - r_i)^2}{Y_i}$ to

$$-L(\mu) = \sum_{i=1}^{N_i} g_i\left([A\vec{\mu}]_i\right) \tag{2.4.7}$$

$$g_i(l) = -Y_i \ln \left(I_i e^{-l} + r_i \right) + \left(I_i e^{-l} + r_i \right)$$

$$g_i(l) \approx g_i\left(\hat{l}_i\right) + \dot{g}_i(\hat{l}_i)(l - \hat{l}_i) + \frac{\ddot{g}_i(\hat{l}_i)}{2}(l - \hat{l}_i)^2$$

$$g_i(l) \approx \underbrace{(Y_i - Y_i \ln Y_i)}_{=\text{const}} + \frac{\ddot{g}_i(\hat{l}_i)}{2}(l - \hat{l}_i)^2$$

$$\Rightarrow \Phi(\vec{\mu}) \approx \sum_{i=1}^{N_i} \frac{w_i}{2}([A\vec{\mu}]_i - \hat{l}_i)^2 + \beta R(\vec{\mu}). \qquad (2.4.8)$$

Note that we just seek a minimum now, since we added an artificial minus sign in Equation 2.4.7, which is called the negative log likelihood. Depending on R, this cost function is already quadratic, but still not separable in j. For this we need to use the convexity trick of De Pierro [Pie93, Pie95] that is also referred to in the literature as the optimization transfer principle [JF07, LHY00]. We are going to rewrite the approximated line integral 2.1.4, which is contained in Equation 2.4.7,

$$[A\vec{\mu}]_i = \sum_{j=1}^{N_j} a_{ij}\mu_j = \sum_{\substack{j=1 \\ a_{ij}\neq 0}}^{N_j} \alpha_{ij} \left\{ \frac{a_{ij}}{\alpha_{ij}}(\mu_j - \mu_j^n) + [A\vec{\mu}^n]_i \right\}$$

with

$$\sum_{j=1}^{N_j} \alpha_{ij} = 1\, \forall i, \alpha_{ij} \geq 0. \qquad (2.4.9)$$

Now we use the convexity of

$$\left\{ [A\vec{\mu}]_i - \hat{l}_i \right\}^2 \leq \sum_{\substack{j=1 \\ a_{ij}\neq 0}}^{N_j} \alpha_{ij} \frac{w_i}{2} \left\{ \frac{a_{ij}}{\alpha_{ij}}(\mu_j - \mu_j^n) + [A\vec{\mu}^n]_i - \hat{l}_i \right\}^2 \qquad (2.4.10)$$

which leads to a separable quadratic surrogate function

$$Q(\vec{\mu}; \vec{\mu}^n) := \sum_{i=1}^{N_i} \sum_{\substack{j=1 \\ a_{ij}\neq 0}}^{N_j} \alpha_{ij} \left\{ \frac{a_{ij}}{\alpha_{ij}}(\mu_j - \mu_j^n) + [A\vec{\mu}^n]_i - \hat{l}_i \right\}^2.$$

Here μ_j^n refers to the voxel j of the iteration n. We can derive a surrogate function for the penality $R \leq S(\vec{\mu}; \vec{\mu}^n)$ with ease. In this work we use the HUBER loss function $\psi_a(\cdot)$. This loss function is defined to behave quadratically in a small regime a around zero and linear outside

$$\psi_a(x) = \begin{cases} x^2/2 & , |x| < a \\ a\,(|x| - a/2) & , \text{else} \end{cases} \qquad (2.4.11)$$

$$S(\vec{\mu}; \vec{\mu}^n) = \sum_{j=1}^{N_j} \sum_{\xi_j \in \mathcal{N}_j} w_\xi \psi_a\left(\mu_j - \mu_{\xi_j}^n\right). \qquad (2.4.12)$$

Again, $\xi_j \in \mathcal{N}_j$ is the multi index that encodes the coordinates that belong to the specific neighborhood[23] of the voxel μ_j^n, w_ξ is the weighting of the

---

[23] In this work we have chosen the 26 surrounding voxels and weighted them accordingly to their distance. These voxels are called the nearest neighbors (N.N.).

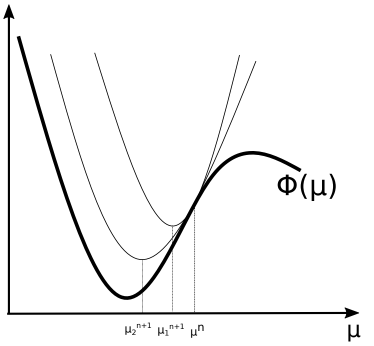


Figure 2.4.1: The defining properties of a surrogate function in the context of this work is that it always has values above the original function and that its derivative at the value $\vec{\mu}^n$ is the equal to the original.

different neighbors distances with respect to the central voxel. The overall surrogate function that we want to minimize is

$$\phi\left(\vec{\mu};\vec{\mu}^n\right) := Q\left(\vec{\mu};\vec{\mu}^n\right) + \beta S\left(\vec{\mu};\vec{\mu}^n\right) \tag{2.4.13}$$

and has the attributes depicted in Figure 2.4.1:

$$\phi\left(\vec{\mu}^n;\vec{\mu}^n\right) = \Phi(\vec{\mu}^n)$$

$$\left.\frac{\partial \phi}{\partial \mu_j}\left(\vec{\mu};\vec{\mu}^n\right)\right|_{\mu=\mu^n} = \left.\frac{\partial \Phi}{\partial \mu_j}\left(\vec{\mu}\right)\right|_{\vec{\mu}=\vec{\mu}^n}$$

$$\phi\left(\vec{\mu};\vec{\mu}^n\right) \geq \Phi\left(\vec{\mu}\right), \text{ for } \mu_j \geq 0\,\forall j.$$

According to Newtons method, we can find the minimum of Equation 2.4.13 with the iterative update step of the form

$$\mu_j^{n+1} = \left[\mu_j^n - \frac{\left.\frac{\partial \phi}{\partial \mu_j}\left(\vec{\mu};\vec{\mu}^n\right)\right|_{\vec{\mu}=\vec{\mu}^n}}{\left.\frac{\partial^2 \phi}{\partial \mu_j^2}\left(\vec{\mu};\vec{\mu}^n\right)\right|_{\vec{\mu}=\vec{\mu}^n}}\right]_+ , \; j = 1, \ldots, N \tag{2.4.14}$$

$$\left.\frac{\partial \phi}{\partial \mu_j}\left(\vec{\mu};\vec{\mu}^n\right)\right|_{\vec{\mu}=\vec{\mu}^n} = \sum_{i=1}^{N_i} a_{ij} w_i \left(\left[A\vec{\mu}^n\right]_i - \hat{l}_i\right) + \beta \left.\frac{\partial S}{\partial \mu_j}\right|_{\vec{\mu}=\vec{\mu}^n} \tag{2.4.15}$$

$$\left.\frac{\partial^2 \phi}{\partial \mu_j^2}\left(\vec{\mu};\vec{\mu}^n\right)\right|_{\vec{\mu}=\vec{\mu}^n} = \sum_{i=1}^{N_i} \frac{a_{ij}^2 w_i}{\alpha_{ij}} + \beta \left.\frac{\partial^2 S}{\partial \mu_j^2}\right|_{\vec{\mu}=\vec{\mu}^n}. \tag{2.4.16}$$

As we can see, the larger the nominator in Equation 2.4.15 and the smaller the denominator in 2.4.16, the larger the step size of 2.4.14. A suitable interpretation of Figure 2.4.1 shows the same behavior for smaller second derivatives 2.4.16. Since we still have the choice of the explicit form of the α_{ij}, we can adapt that idea into the denominator. Since $a_{ij}^2 w_i$ is independent of the iteration, we also would like to have iteration independent α_{ij}. A typical choice subject to Equation 2.4.9 is

$$\alpha_{ij} := \frac{a_{ij}}{\sum_j a_{ij}}. \tag{2.4.17}$$

To accelerate the algorithm 2.4.14, we use the ordered subsets concept. With the number M of used subsets per update step [KB98], we define the penalized weighted least squares ordered subsets (PWLS-OS) algorithm for monochromatic X-ray tomography

$$\mu_j = \left[\mu_j - \frac{M \sum_{i\in\eta} a_{ij} w_i \left([A\vec{\mu}]_i - \hat{l}_i \right) + \beta \left. \frac{\partial S}{\partial \mu_j} \right|_{\vec{\mu}=\vec{\mu}_{\text{old}}}}{\sum_{i=1} \frac{a_{ij}^2 w_i}{\alpha_{ij}} + \beta \left. \frac{\partial^2 S}{\partial \mu_j^2} \right|_{\vec{\mu}=\vec{\mu}_{\text{old}}}} \right]_+ . \tag{2.4.18}$$

Here, the $\sum_i \frac{a_{ij}^2 w_i}{\alpha_{ij}}$ is iteration independent and can be pre-computed once and stored for the whole reconstruction. The term $\sum_{i\in\eta} a_{ij} w_i \left([A\vec{\mu}]_i - \hat{l}_i \right)$ basically compares the measurements $\hat{l}_i$ with the forward-projection of the current iterate $[A\vec{\mu}]_i$ and back projects a multiple of that information. The β-terms represent the penalization. For completeness, this method could be refered as penalized weighted least squares method with ordered subsets (PWLS-OS). The next section delivers a polychromatic generalization of Equation 2.4.18 which is the main subject of this thesis.

2.4.2 Iterative polychromatic statistical reconstruction

The update step 2.4.18 is based on the forward-projection model Equation 2.3.1. Thus, it does not incorporate the physics of the X-ray absorption process mentioned in Section 2.3.

Given the more accurate model Equation 2.3.2, the minimizing of a similar cost function of 2.4.5 becomes a more delicate task.

The forward-projection model states that the measurement along the line i can be written as

$$\bar{Y}_i(\vec{s}) = \int_{\mathbb{R}^+} I_i(E) \exp\left\{ -\vec{m}(E) \cdot \vec{s}_i \right\} \mathsf{d}E. \tag{2.4.19}$$

Note that $\vec{m}(E) \cdot \vec{s}$ is the scalar product $\vec{m}^{\mathsf{T}} \vec{s} = \sum_k m_k s^k$. Further note that Y_i represents the intensity measurements and $\bar{Y}_i$ can be interpreted as the forward-projection of the volume $\vec{\rho}$. The k-th component of $\vec{s}_i$ represents the line integral along the line $\mathcal{L}_i$ of the k-th material. In other words: $\vec{s}_i$ is the classic line integral of the k-th material with K the total number of different materials. Furthermore, we need to define the following variables:

$$\mu_j(E) = \sum_{k=1}^{K} m_k(E)\,\rho_j f_j^k \tag{2.4.20}$$

$$a_{ij}^k := a_{ij} f_j^k \tag{2.4.21}$$

$$s_i^k(\vec{\rho}) := \int_{\mathcal{L}_i} \rho(x,y,z)\, f^k(x,y,z)\,\mathsf{d}l = \sum_{j=1}^{N_j} \rho_j a_{ij}^k \tag{2.4.22}$$

$$\vec{s}_i(\vec{\rho}) := \left[s_i^1(\vec{\rho}), s_i^2(\vec{\rho}), \dots, s_i^K(\vec{\rho})\right]^{\mathsf{T}}$$

$$E[Y_i] = \int_{\mathbb{R}^+} I_i(E)\exp\left\{-\vec{m}(E)\cdot\vec{s}_i(\vec{\rho})\right\}\mathsf{d}E + r_i = \bar{Y}_i + r_i. \tag{2.4.23}$$

One main idea is the product approach 2.4.20, which splits the energy dependence $m_k(E)$ from the spatial information, stating which voxel consists of how much material f_j^k. This is the bone attenuation property that is plotted in Figure 2.3.1. The segmentation f_j^k has to be known for each voxel j or it has to be subject to certain physical model assumptions. Note that $\sum_k f_j^k = 1\,\forall j$.

The polychromatic equivalent of 2.4.6 is

$$\hat{\vec{\rho}} = \underset{\rho_j \geq 0\,\forall j}{\arg\min}\,\Phi(\vec{\rho}) \tag{2.4.24}$$

$$\Phi(\vec{\rho}) = -L(\vec{\rho}) + \beta R(\vec{\rho}).$$

The negative log likelihood can be expressed as

$$-L(\vec{\rho}) = \sum_{i=1}^{N_i} h_i\left(\bar{Y}_i(\vec{s}_i(\vec{\rho})) + r_i\right)$$

$$h_i(t) = -Y_i \ln(t) + t$$

$$\Rightarrow \frac{\partial}{\partial t} h_i(t) = 1 - \frac{Y_i}{t} \tag{2.4.25}$$

$$t_i := \exp\left\{-\vec{m}(E)\cdot\vec{s}_i\right\} + \bar{r}_i \tag{2.4.26}$$

$$\bar{r}_i := \frac{r_i}{I_0}$$

$$I_0 := \int_{\mathbb{R}^+} I(E)\,\mathsf{d}E.$$

Note that the argument of h is non-linear, hence it is not as easy to move the argument out of the integral as in Equation 2.4.10. With the definition

$$b_i^n(E) := \frac{\bar{Y}_i(\vec{s}_i(\vec{\rho}^n)) + r_i}{t_i(E, \vec{s}_i(\vec{\rho}^n))} \tag{2.4.27}$$

$$\begin{aligned}
\Rightarrow \int_{\mathbb{R}^+} \frac{I_i(E)}{b_i^n(E)} \mathsf{d}E &= \int_{\mathbb{R}^+} \frac{I_i(E)\, t_i(E, \vec{s}_i(\vec{\rho}^n))}{\bar{Y}_i(\vec{s}_i(\vec{\rho}^n)) + r_i} \mathsf{d}E \\
= \int_{\mathbb{R}^+} & \frac{I_i(E)\, t_i(E, \vec{s}_i(\vec{\rho}^n))}{\int_{\mathbb{R}^+} I_i\left(\widetilde{E}\right) \exp\left\{-\vec{m}\left(\widetilde{E}\right) \cdot \vec{s}_i(\vec{\rho}^n)\right\} \mathsf{d}\widetilde{E} + r_i} \mathsf{d}E \\
&= \frac{\int_{\mathbb{R}^+} I_i(E) \left[\exp\left\{-\vec{m}(E) \cdot \vec{s}_i(\vec{\rho}^n)\right\} + \bar{r}_i\right] \mathsf{d}E}{\int_{\mathbb{R}^+} I_i\left(\widetilde{E}\right) \exp\left\{-\vec{m}\left(\widetilde{E}\right) \cdot \vec{s}_i(\vec{\rho}^n)\right\} \mathsf{d}\widetilde{E} + r_i} \\
= \frac{\int_{\mathbb{R}^+} I_i(E) \exp\left\{-\vec{m}(E) \cdot \vec{s}_i(\vec{\rho}^n)\right\} \mathsf{d}E + \int_{\mathbb{R}^+} I_i(E) \frac{r_i}{\int_{\mathbb{R}^+} I\left(\check{E}\right) \mathsf{d}\check{E}} \mathsf{d}E}{\int_{\mathbb{R}^+} I_i\left(\widetilde{E}\right) \exp\left\{-\vec{m}\left(\widetilde{E}\right) \cdot \vec{s}_i(\vec{\rho}^n)\right\} \mathsf{d}\widetilde{E} + r_i} & \\
= \frac{\int_{\mathbb{R}^+} I_i(E) \exp\left\{-\vec{m}(E) \cdot \vec{s}_i(\vec{\rho}^n)\right\} \mathsf{d}E + r_i \cancelto{1}{\frac{\int_{\mathbb{R}^+} I_i(E) \mathsf{d}E}{\int_{\mathbb{R}^+} I\left(\check{E}\right) \mathsf{d}\check{E}}}}{\int_{\mathbb{R}^+} I_i\left(\widetilde{E}\right) \exp\left\{-\vec{m}\left(\widetilde{E}\right) \cdot \vec{s}_i(\vec{\rho}^n)\right\} \mathsf{d}\widetilde{E} + r_i} &= 1
\end{aligned}$$

and as we are in the intensity domain, the scatter signal cannot change the positivity of $I_i(E)$ so that we can conclude

$$\frac{I_i(E)}{b_i^n(E)} \geq 0.$$

We can now write

$$\bar{Y}_i(\vec{s}_i(\vec{\rho}^n)) + r_i = \int_{\mathbb{R}^+} I_i(E)\, t_i(E, \vec{s}_i(\vec{\rho}^n))\, \mathsf{d}E = \int_{\mathbb{R}^+} \frac{I_i(E)}{b_i^n(E)} t_i(E, \vec{s}_i(\vec{\rho}^n))\, b_i^n(E)\, \mathsf{d}E$$

$$\begin{aligned}
\Rightarrow h_i\left(\bar{Y}_i(\vec{s}_i(\vec{\rho}^n)) + r_i\right) &= h_i\Big(\int_{\mathbb{R}^+} \frac{I_i(E)}{b_i^n(E)} t_i(E, \vec{s}_i(\vec{\rho}^n))\, b_i^n(E)\, \mathsf{d}E\Big) \\
&\leq \int_{\mathbb{R}^+} \frac{I_i(E)}{b_i^n(E)} h_i\left(t_i(E, \vec{s}_i(\vec{\rho}^n))\, b_i^n(E)\right) \mathsf{d}E =: Q_1(\vec{\rho}, \vec{\rho}^n).
\end{aligned}$$

We want to arrive at a surrogate like Equation 2.4.13 and need a quadratic version of Q_1 as a function of $l_i^n(\vec{\rho}) := \vec{m}(E) \cdot \vec{s}_i(\vec{\rho}^n)$, which is easier to minimize. Hence, we go for the standard TAYLOR expansion as follows

$$\begin{aligned}
h_i\left(t_i(E, \vec{s}_i(\vec{\rho}^n))\, b_i^n(E)\right) &= h_i\left(b_i^n(E) \exp\left\{-\vec{m}(E) \cdot \vec{s}_i(\vec{\rho}^n)\right\} + b_i^n(E)\, \bar{r}_i\right) \\
&=: g_i^n\left(\vec{m}(E) \cdot \vec{s}_i(\vec{\rho}^n), E\right) \leq q_i^n\left(\vec{m}(E) \cdot \vec{s}_i(\vec{\rho}^n), E\right) \\
q_i^n(l, E) &= g_i^n(l_i^n, E) + \dot{g}_i^n(l_i^n, E)\left[l - l_i^n\right] + \frac{1}{2} C_i^n(E)\left[l - l_i^n\right]^2.
\end{aligned} \tag{2.4.28}$$

Here, we call C_i^n the curvature of g_i^n, as it is basically its second derivative $\ddot{g}_i^n = \partial^2/\partial l^2 g_i^n$ with respect to the line integral $l_i^n(\vec{\rho})$. It is meant to be a helpful function of convergence as depicted in Figure 2.4.1.

We define Q_2, $\lambda_{ij}(E)$ and its matrix $\Lambda(E) = \{\lambda_{ij}(E)\} \in \mathbb{R}^{N_j \times N_i}$ and use the convexity trick in its first argument of Q_2 as follows:

$$Q_2\left(\vec{\rho};\vec{\rho}^n\right) := \sum_{i=1}^{N_i} \int_{\mathbb{R}^+} \frac{I_i(E)}{b_i^n(E)} q_i^n\left(\vec{m}(E)\cdot\vec{s}_i\left(\vec{\rho}^n\right), E\right) \mathsf{d}E$$

$$\vec{m}(E)\cdot\vec{s}_i(\vec{\rho}) = \sum_{j=1}^{N_j}\sum_{k=1}^{K} m_k(E)\, a_{ij}^k \rho_j = \sum_j \lambda_{ij}(E)\,\rho_j$$

$$\lambda_{ij}(E) := \sum_{k=1}^{K} m_k(E)\, a_{ij}^k$$

$$\Rightarrow \vec{m}(E)\cdot\vec{s}_i(\vec{\rho}) \overset{2.4.17}{=} \sum_{\substack{j=1\\ \lambda_{ij}\neq 0}}^{N_j} \alpha_{ij}\left(\frac{\lambda_{ij}(E)}{\alpha_{ij}}\left(\rho_j-\rho_j^n\right) + \left[\Lambda(E)\,\vec{\rho}^n\right]_i\right)$$

$$\Rightarrow q_i^n\left(\vec{m}(E)\cdot\vec{s}_i, E\right) = q_i^n\left(\left(\frac{\lambda_{ij}(E)}{\alpha_{ij}}\left(\rho_j-\rho_j^n\right) + \left[\Lambda(E)\,\vec{\rho}^n\right]_i\right), E\right)$$

$$\leq \sum_{\substack{j=1\\ \lambda_{ij}\neq 0}}^{N_j} \alpha_{ij} q_i^n\left(\frac{\lambda_{ij}(E)}{\alpha_{ij}}\left(\rho_j-\rho_j^n\right) + \left[\Lambda(E)\,\vec{\rho}^n\right]_i, E\right).$$

Now we are closer to the formulation of an update step similar to Equation 2.4.14, since we can define a surrogate Q and its derivatives ($\frac{\partial}{\partial s_i^k}$ denotes the k-th component of the gradient)

$$Q\left(\vec{\rho};\vec{\rho}^n\right) := \sum_{\substack{i=1\\ \lambda_{ij}\neq 0}}^{N_i} \sum_{\substack{j=1\\ \lambda_{ij}\neq 0}}^{N_j} \int_{\mathbb{R}^+} \frac{I_i(E)}{b_i^n(E)} \alpha_{ij} q_i^n\left(\frac{\lambda_{ij}(E)}{\alpha_{ij}}\left(\rho_j-\rho_j^n\right) + \left[\Lambda(E)\,\vec{\rho}^n\right]_i, E\right) \mathsf{d}E$$

$$\left.\frac{\partial Q}{\partial \rho_j}\left(\vec{\rho};\vec{\rho}^n\right)\right|_{\vec{\rho}=\vec{\rho}^n} = \sum_{\substack{i=1\\ \lambda_{ij}\neq 0}}^{N_i} \frac{\partial}{\partial \rho_j}\left\{\left.\int_{\mathbb{R}^+} \frac{I_i(E)}{b_i^n(E)} \alpha_{ij} q_i^n\left(\frac{\lambda_{ij}(E)}{\alpha_{ij}}\left(\rho_j-\rho_j^n\right) + \left[\Lambda(E)\,\vec{\rho}^n\right]_i, E\right) \mathsf{d}E\right|_{\vec{\rho}=\vec{\rho}^n}\right\}.$$

At this point, it helps to remember that $\frac{\lambda_{ij}(E)}{\alpha_{ij}}\left(\rho_j-\rho_j^n\right)$ is the only relevant term (in the first argument of q) for that derivative. This basically results in a factor of $\dot{g}_i^n$ and its inner derivative, because of the taylor expansion 2.4.28. The definition 2.4.26 as well as Equation 2.4.25 helps to write down[24]

$$\left.\frac{\partial Q}{\partial \rho_j}\left(\vec{\rho};\vec{\rho}^n\right)\right|_{\vec{\rho}=\vec{\rho}^n} = \sum_{i=1}^{N_i}\int_{\mathbb{R}^+}\frac{I_i\left(E\right)}{b_i^n\left(E\right)}\lambda_{ij}\dot{g}_i^n\left(\vec{m}\left(E\right)\cdot\vec{s}_i\left(\vec{\rho}^n\right),E\right)\mathsf{d}E$$

$$=\sum_{i=1}^{N_i}\int_{\mathbb{R}^+}\frac{I_i\left(E\right)}{b_i^n\left(E\right)}\lambda_{ij}\left\{b_i^n\left(E\right)\left(1-\frac{Y_i}{\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right)+r_i}\right)\right\}\exp\left(-\vec{m}\left(E\right)\cdot\vec{s}_i\left(\vec{\rho}^n\right)\right)\mathsf{d}E$$

$$=\sum_{i=1}^{N_i}\left(1-\frac{Y_i}{\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right)+r_i}\right)\int_{\mathbb{R}^+}\lambda_{ij}\left(E\right)I_i\left(E\right)\exp\left(-\vec{m}\left(E\right)\cdot\vec{s}_i\left(\vec{\rho}^n\right)\right)\mathsf{d}E$$

$$\Rightarrow\left.\frac{\partial Q}{\partial \rho_j}\left(\vec{\rho};\vec{\rho}^n\right)\right|_{\vec{\rho}=\vec{\rho}^n}=\sum_{i=1}^{N_i}\sum_{k=1}^{K}a_{ij}f_j^k\left(1-\frac{Y_i}{\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right)+r_i}\right)\frac{\partial}{\partial s_i^k}\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right). \tag{2.4.29}$$

The second derivative delivers another factor $\frac{\lambda_{ij}(E)}{\alpha_{ij}}$ and the curvature $C_i^n\left(E\right)$, finally resulting in

$$\Rightarrow\left.\frac{\partial^2 Q}{\partial {\rho_j}^2}\left(\vec{\rho};\vec{\rho}^n\right)\right|_{\vec{\rho}=\vec{\rho}^n}=\sum_{\substack{i=1\\ \lambda_{ij}\neq 0}}^{N_i}\int_{\mathbb{R}^+}\frac{I_i\left(E\right)}{b_i^n\left(E\right)}\frac{\lambda_{ij}^2\left(E\right)}{\alpha_{ij}}C_i^n\left(E\right)\mathsf{d}E. \tag{2.4.30}$$

A solid choice is to approximate the most relevant parts of the integral over the energy live around the effective energy E_{eff},, such that

$$\left.\frac{\partial^2 Q}{\partial {\rho_j}^2}\left(\vec{\rho};\vec{\rho}^n\right)\right|_{\vec{\rho}=\vec{\rho}^n}=\sum_{\substack{i=1\\ \lambda_{ij}\neq 0}}^{N_i}\int_{\mathbb{R}^+}\frac{I_i\left(E\right)\left(\exp\left\{-\vec{m}\left(E\right)\cdot\vec{s}_i\left(\vec{\rho}^n\right)\right\}+\bar{r}_i\right)}{\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right)+r_i}\frac{\lambda_{ij}^2\left(E\right)}{\alpha_{ij}}C_i^n\left(E\right)\mathsf{d}E$$

$$=\sum_{\substack{i=1\\ \lambda_{ij}\neq 0}}^{N_i}\int_{\mathbb{R}^+}\frac{\lambda_{ij}^2\left(E\right)}{\alpha_{ij}}C_i^n\left(E\right)\left\{\cancelto{\text{eq.2.4.19}}{\frac{I_i\left(E\right)\exp\left\{-\vec{m}\left(E\right)\cdot\vec{s}_i\left(\vec{\rho}^n\right)\right\}+r_i}{\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right)+r_i}}\right\}\mathsf{d}E$$

$$\approx\sum_{\substack{i=1\\ \lambda_{ij}\neq 0}}^{N_i}\frac{\lambda_{ij}^2\left(E_{\text{eff}}\right)}{\alpha_{ij}}C_i^n\left(E_{\text{eff}}\right)$$

---

[24] The original paper [EF02] includes a minus sign error, which propagates through the final formula 2.4.34 of $\hat{N}_j$. Also, the original paper has "$+r_i$" missing in b_i^n of Equation 2.4.27.

$$\left.\frac{\partial^2 Q}{\partial {\rho_j}^2}\left(\vec{\rho};\vec{\rho}^n\right)\right|_{\vec{\rho}=\vec{\rho}^n} \approx \sum_{\substack{i=1\\ \lambda_{ij}\neq 0}}^{N_i} \frac{\lambda_{ij}^2\left(E_{\text{eff}}\right)}{\alpha_{ij}} Y_i = \sum_{\substack{i=1\\ a_{ij}\neq 0}}^{N_i} \frac{\left[\sum_{k=1}^{K} m_k\left(E_{\text{eff}}\right) a_{ij}^k\right]^2}{\alpha_{ij}} Y_i$$

$$\Rightarrow \left.\frac{\partial^2 Q}{\partial {\rho_j}^2}\left(\vec{\rho};\vec{\rho}^n\right)\right|_{\vec{\rho}=\vec{\rho}^n} \approx \sum_{k=1}^{K} m_k^2\left(E_{\text{eff}}\right) \sum_{i=1}^{N_i} a_{ij} Y_i \left[\sum_{j=1}^{N_j} a_{ij}\right] =: d_j. \qquad (2.4.31)$$

Here, we assumed that the main contribution of that integral lies around the centroid

$$E_{\text{eff}} := \frac{\int_{\mathbb{R}^+} E \cdot I_i\left(E\right) \mathsf{d}E}{\int_{\mathbb{R}^+} I_i\left(E\right) \mathsf{d}E}. \qquad (2.4.32)$$

A more mathematically sophisticated explanation of Equation 2.4.31 can be found in [EF02] and [KPTF13]. Note that $\sum_{j=1}^{N_j} a_{ij} =: \gamma_i$ can be interpreted as the path length of the i-th ray through the volume. The above definition of d_j has the form of a rayweighted back projection of the measurements Y_i with a constant factor defined as

$$M^K := \sum_{k=1}^{K} m_k^2\left(E_{\text{eff}}\right)$$

that is only dependent on the considered materials in the reconstruction volume. Since that number increases even if there is only a single voxel in the volume of a certain material, it can be useful for the convergence to ignore that material. Imagine a given set of materials. We could sort those materials increasingly by their value $m_k\left(E_{\text{eff}}\right)$ over the index k in most of the cases [25]

$$m_1\left(E_{\text{eff}}\right) < m_2\left(E_{\text{eff}}\right) < \ldots < m_K\left(E_{\text{eff}}\right)$$

such that we use only a decreased amount of materials for a d_j^K, defined in the same manner as

$$d_j^K := M^K \sum_{i=1}^{N_i} \gamma_i a_{ij} Y_i. \qquad (2.4.33)$$

In fact, for the reconstructions done in this work, d_j^1 was preferred over $d_j^{K>1}$ due to convergence speed. This holds true, because smaller d_j result in larger step-length (see Equation 2.4.35), and thus in faster convergence. The relevant objects we want to reconstruct consist mostly of water. Hence, the first material is meant to be water.

---

[25] The case of different materials with equal densities is excluded, but becomes an increasing problem considering contrast agent in perfusion imaging. Such material can overlap in the range from water to bone and contradicts the easy distinguishing by one value.

With Equation 2.4.29 the analog update step to 2.4.18 can be written with the definition of

$$\hat{N}_j := \sum_{i\in\eta}\sum_{k=1}^{K} a_{ij} f_j^k \left(1 - \frac{Y_i}{\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho^n}\right)\right) + \bar{r}_i}\right)\frac{\partial}{\partial s_i^k}\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho^n}\right)\right) \tag{2.4.34}$$

to the final update step of this PWLS-OS algorithm as

$$\rho_j^{n+1} = \left[\rho_j^n - \frac{M\hat{N}_j + \beta\frac{\partial S}{\partial \rho_j}}{d_j^1 + \beta\frac{\partial^2 S}{\partial {\rho_j}^2}}\right]_+ . \tag{2.4.35}$$

The subset $\eta \ni i$ needs to be defined by the user and typically lies in the interval of one angle per subset to all angles in one big subset (see Section 3.4). In this work, it is not considered to consist of only parts of a projection of a single angle.

As a firm reminder to the beginning of this subsection, technically speaking, every single event that leads to the measurement of a single signal is a statistical event of higher order. It is worth noting that this likelihood approach is only a first order of statistical accuracy. Thus, a more precise statistical model to account for every statistical aspect of this problem would be more complicated (e.g. a bigger product in the likelihood function). A more complicated likelihood function, and thus a very difficult update step can be found in [EF03a]: "*Unlike the Poisson distribution, compound Poisson statistics have a complicated likelihood that impedes direct use of statistical reconstruction*".

2.5 Pinhole camera model

This section provides the reader with the concept of the encoding and decoding of the acquisition setup. As the X-ray source is approximated as a point source that projects a three dimensional volume onto a two dimensional plane (flat panel detector), the projective geometry of the pinhole camera model can accurately describe the setup. This section gives an idea of how the position of the X-ray source, detector and reconstrcution volume are positioned in relation to each other.

C-arm CT scanners use the so-called projection matrices to encode the geometry information of the scanning protocol into the 12 entries of a matrix P. The pinhole camera model is used to project world coordinates $\vec{x} = [x, y, z, 1]^{\mathrm{T}}$ onto detector pixel "u, v"-coordinates[26].

$$P\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \tag{2.5.1}$$

---

[26] Often called "s,t"-coordinates in common literature.

The projection matrix $P \in \mathbb{R}^{3\times 4}$ incorporates the scanning geometry of the C-arm device. It consists of a scaling factor $s \in \mathbb{R}^{+}, [s] = \frac{1}{\mathrm{mm}}$, a calibration matrix $\Omega \in \mathbb{R}^{3\times 3}$, a rotation $R \in \mathbb{R}^{3\times 3}$ and a translation $\vec{t} = [t_x, t_y, t_z]^{\mathrm{T}} \in \mathbb{R}^3$ of world coordinates

$$P = s \cdot \Omega \cdot \left[R | \, \vec{t} \right] . \tag{2.5.2}$$

The calibration Ω contains the information of the source-to-detector distances in units of the detector pixels.

$$\Omega = \begin{pmatrix} f_u & a & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{pmatrix}$$

$f_{u,v}$ gives the distance of the source to the center point of the detector. $c_{u,v}$ are the coordinates of the center point of the detector. Note that the pixel lengths in u and v direction depend on the actual detector and may specify a non-quadratic pixel size, which would result in $f_u \neq f_v$. If the detector is rectangular (which is the case in this work), the scalar a is equal to zero.

The pinhole camera model has very practical use, i.e. for the voxel based footprint projector (Section 3.2). Furthermore, to calculate the lines $\mathcal{L}_i$ in world coordinates of the X-ray source $\vec{\xi} = [\xi_x, \xi_y, \xi_z, 1]^{\mathrm{T}}$ to the detector pixels, we need to solve the equation

$$P\vec{\xi} = \vec{0} \tag{2.5.3}$$

and plug that into the equation of a straight line in 3D respectively (sec 3.1 Equation 3.1.2).

To extract the direction $\vec{r}_i$ of specific ray i, we need to find one more point $\vec{x}_i$ of the world coordinate system that lies on this straight line. From linear algebra follows, we can write the projection matrix as a product of Q and R, where Q is an orthogonal matrix that can be diagonal and R is an upper triangle matrix. From Equation 2.5.1 follows then

$$QR \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}$$

$$\Rightarrow R \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = Q^{-1} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \overset{Q\text{orth.}}{=} Q^{\mathrm{T}} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} . \tag{2.5.4}$$

Since R is an upper triangular matrix, one can calculate $\vec{r}_i = [x_i, y_i, z_i, 1]^{\mathrm{T}}$ for a given u, v-pair und thus build the needed equation of the line that represents the X-ray propagation from the source to the detector pixel. In this work we derive R and Q via HOUSEHOLDER method that is proven to be numerically stable [Hou58, Wat82].

Depending on the projection angle, the world coordinates of the X-ray source R and Q are stored and utilized by the ray caster (Subsection 3.1).

3 The implementation

This chapter gives the reader an in-depth view of the implemented version of the PSR technique. It starts by introducting the main ideas that contribute to the program building blocks that are needed by the algorithm Equation 2.4.35. Section 3.4 draws the connection of those main ideas into the software driven aspects that end up representing the whole algorithm that represents the core of this thesis and that produced the reconstruction results presented in chapter 4.

For the sake of clarity, whenever the opportunity for a reference of the source code provided with this thesis appears, the reader is referred via footnotes and the brackets
["name of the function/class" @ "name of the cpp/cl source file"].

3.1 The forward-projection / the ray caster

The main goal of this ray caster is to project a voxelbased volume

$$\vec{\rho} = \{\rho_j\} = \left[\rho_1, \rho_2, \dots, \rho_{N_j}\right]^{\mathsf{T}}$$

onto a flat panel detector from a point sized source (flat panel cone beam geometry). The forward-projection s_i can be written as

$$s_i := c_i \sum_j \rho_j a_{ij} \tag{3.1.1}$$

with a scaling factor c_i which is a constant for every direction $i = 1, \dots, (N_u \cdot N_v \cdot N_p)$ with $N_{u,v}$ the number of detector pixels in u, v direction and N_p the number of sampling angles (views). In order to calculate the geometry factors a_{ij}, one needs to check the intersection points of the ray with the voxel edges of the volume in arbitrary directions. For every line $\vec{x}_i : \mathbb{R} \to \mathbb{R}^3$, we calculate the direction on the surface of the 3D ball $\vec{r}_i \in \varsigma^2$ and the source coordinate $\vec{\xi}_i \in \mathbb{R}^3$ from the projection matrices (pinhole camera model Section 2.5 Equation 2.5.3). The parameter representation of the line $\vec{x}_i(t)$ has the form

$$\vec{x}_i(t) = \vec{\xi}_i + \vec{r}_i t \quad \text{with } t \in \mathbb{R}^+. \tag{3.1.2}$$

In general, the three voxel dimensions $\vec{v} = \{v_1, v_2, v_3\} \in \mathbb{R}^{3+}$ only scale up the line integral s_i with the factor $c_i = |\vec{r}_i \cdot \vec{v}|$. Since $\vec{r}_i$ is given for each i-th line and $v_m = \text{const}$, we can transform the grid to the unit cube.

This enables us to break down the intersection problem to simple integer calculations and case analysis depending on the signs of the three components of $\vec{r}_i = \{r_{i1}, r_{i2}, r_{i3}\}$.

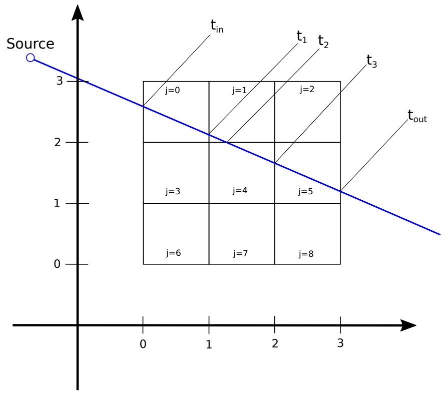


Figure 3.1.1: The visualization of the ray casting concept in two arbitrary spatial dimensions illustrates the procedure.

The a_{ij} can be calculated as differences of the parameter t at the voxel edges. First, we start to calculate the two parameters $t_{\text{in}} < t_{\text{out}}$ (where the ray hits the outer voxels of the volume) and use them as starting and ending criteria for the ray casting loop. This can be done by plugging zeroes $\vec{0} = \{0, 0, 0\}$ and the number of volume voxels $\vec{N} = \{N_1, N_2, N_3\} \in \mathbb{N}^3$ into the three Equations 3.1.2 and analyzing the resulting parameters t by cross-checking the other two equations about the resulting $\vec{x}_i(t)$ and weither if they lie in the interval between $\vec{0}$ and $\vec{V}$. Case analysis delivers the smaller of the two resulting parameters which represent t_{in} and t_{out} respectively.

Now we need to check the dimension in which we are intersecting with the next voxel edge. This means which component of $\vec{x}_i(t_{n+1})$ is an integer that has exactly ± 1 value difference. Thus, we look in the direction $\vec{r}_i$ in which dimension the next integer value of $\vec{x}_i(t_{n+1})$ lies. The geometry factors can be calculated as the differences $\triangle t = t_{n+1} - t_n$ and assigned to the voxel in between $\vec{x}_i(t_{n+1})$ and $\vec{x}_i(t_n)$. Practically we calculate the mean value of those two points and use it to refer to the specific location in the array, which represents the volume point μ_j. That needs the coordinates of the volume to be positive numbers. This is always possible - as we can shift the volume coordinate system to the first octant of the world coordinate system.

For example, the a_{ij} in Figure 3.1.1 are associated to $a_{i0} = t_1 - t_{\text{in}}, a_{i1} = t_2 - t_1, a_{i2} = a_{i3} = 0, a_{i4} = t_3 - t_2, a_{i5} = t_{\text{out}} - t_3$. During this loop, one can easily build the sum of Equation 3.1.1 and multiply it with c_i until the loop ends.

Certain special cases have to be considered and checked everytime a new ray is calculated. For example, the null set, if the ray lies *exactly* (with respect to the used floating point precision) in a 2D or 3D corner. A small number $\varepsilon > 0$ was defined, depending on $\vec{v}, \vec{N}$ and the number of voxels $\vec{N} = \{N_1, N_2, N_3\}$. Note that $v_m N_m = V_m$. To check if the coordinates $\vec{x}_i(t)$ are nearly integer values, we use the very same tolerance level ε. That small number was determined by trial and error and is correlated with the number of voxels. If ε becomes too small, the kernels will not terminate due to an infinite loop in the respective corner, resulting in a crash of the program. Such an infinite loop can be avoided, if the next t_{n+1} is increased by a number having the same magnitude.

If ε is too large, the accuracy of a_{ij} suffers and the quality of the forward-projection becomes poor. The accuracy of the ray caster has low impact in comparison with real cone beam CT data, because of the inaccuracy of this kind of real world measurements (e.g. the voxelization assumption is a much harder restriction then the accuracy of the intersection calculus). This is the reason why typical ray casters use the GPU-specific voxel interpolation and constant step length for very fast ray casting on GPU devices for gaming or rendering software [KW03, Sch05].

A high accuracy (small ε) was used, because we need to match the high accuracy of the footprint back projector, which will be explained in detail in the next section.

3.2 The backward-projection / the footprint

All reconstruction methods have in common that they need to reverse the projection step of the volume. This reversal step is usually called the backward-projection (BP). Here, we call the terms $\sum_{i=1}^{N_i} \gamma_i a_{ij} Y_i$ in Equation 2.4.33 and $\sum_{i \in \eta} a_{ij} f_j^k$ in Equation 2.4.34 a backward-projection. If we compare this to $\sum_j \mu_j a_{ij}$ in Equation 3.1.1, we can easily recognize that the important difference is the sum over i instead of j. Hence, we need to calculate the sum over all detector pixels that receive contributions of voxel j with respect to all views. This is a very delicate task, since we would need to check every single ray i with a method like a ray caster (Section 3.1), for the case of propagation through the voxel j.

Because it is a very high computational effort to backward-project with a ray caster, we use a different approach. A ray caster sits on a line and follows its way through the volume. It is an efficient method, if we want to know which voxels contribute to a ray, because we do not need to check *every* voxel, but instead only those which are contributing. With an increasing number of voxels N_j, the ratio of such actively contributing voxels compared to N_j decreases dramatically. The geometric interpretation of this relation can be formulated as the decreasing probability for a randomly chosen voxel to touch a randomly chosen line through the volume. This probability converges to zero with increasing N. For the reverse problem we need a method

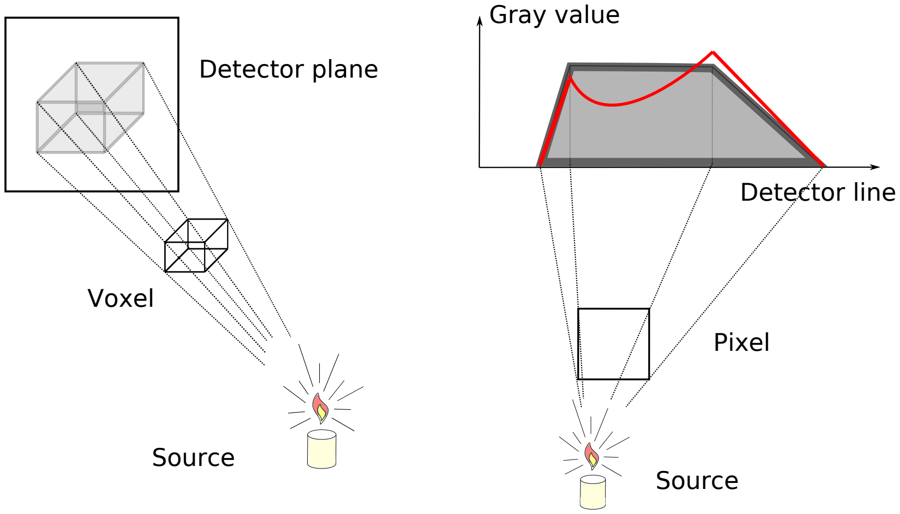


Figure 3.2.1: The implemented footprint scheme to calculate geometry factors in a back projecting setting

to evaluate all pixels that are influenced by a voxel when forward-projected. If we know the pixels and their corresponding geometry factors $a_{ij} > 0$, we can calculate the backward-projecting step.

The idea is to calculate the “shadow-approximation” of a voxel on the detector plane, which tells us which pixels are influenced by a certain voxel. This section gives a brief description of that footprint idea.

We approximate the footprint with a trapezoid. This method was developed by Long and Fessler [LFB10] and is used by several research groups [FPB$^+$14]. The error that is indicated by the red lines in Figure 3.2.1 is corrected in a later calculation step that is called “the angle correction”. If we combine the 1D approximation and state that the product of two trapezoid functions builds the 2D approximation of a “voxel shadow”, we achieve a backward-projecting step with a high computing performance. The accuracy of this method is discussed in the origin [LFB10] and the performance in [FPR13]. Thus, the current implementation of the PSR is using non-adjugated operators that represent the forward-projection (FP, see Section 3.1) and the backward-projection (BP, see Section 3.2). Using the corresponding projection operators A_{BP} and A^{FP}, this property can be written as

$$\vec{\mu}^{\mathrm{T}}\left(A_{\mathrm{BP}}\vec{p}\right) \neq \vec{p}^{\mathrm{T}}\left(A^{\mathrm{FP}}\vec{\mu}\right)$$

or

$$A_{\mathrm{BP}}A^{\mathrm{FP}} \neq \operatorname{diag}\left\{1\right\}.$$

It is also worth mentioning that the early stages of the PSR framework a ray caster based approach was used for the backward-projection. Even though this back projector is actually the perfect adjugate to the FP, the BP operator leads to undersampling artifacts[1] that become visible if one backward-projects a constant.

3.3 The displacement model - distinguishing f^k by the density

The algorithm Equation 2.4.35 requires calculation of the term

$$\frac{M\hat{N}_j + \beta \frac{\partial S}{\partial \rho_j}}{d_j^1 + \beta \frac{\partial^2 S}{\partial {\rho_j}^2}},$$

which includes the approximation of the X-ray image that would be the *most likely* outcome, if the current volume ρ_j^n were scanned within the measurement setup. The important numbers $\hat{N}_j$

$$\hat{N}_j = \sum_{i \in \eta} \sum_{k=1}^{K} a_{ij} f_j^k \left(1 - \frac{Y_i}{\bar{Y}_i \left(\vec{s}_i \left(\vec{\rho^n} \right) \right) + \bar{r}_i} \right) \frac{\partial}{\partial s_i^k} \bar{Y}_i \left(\vec{s}_i \left(\vec{\rho^n} \right) \right)$$

tell the algorithm if voxel values are too high or too low with respect to the measurements Y_i. We can identify the structure of $\hat{N}_j = \sum_{i \in \eta} a_{ij} \cdot \ldots$ as the backward-projection operation (see Section 3.2) with $i \in \eta$ that encodes the subset of used projections.

The information f_j^k which voxel j consists of what kind of material k, is needed in the backward-projecting operation as well as in the forward-projection. Elbakri and Fessler expanded the PSR with a “displacement model” [EF03c], which allows material mixtures in voxels via sigmoids (e.g. cubic splines or trigonometric functions).

Similar to the idea proposed to $m_k(E_{\text{Eff}})$ in Equation 2.4.33, we assume that the given set of K materials can be identified by their density ρ^k as $\rho^1 < \rho^2 < \ldots < \rho^K$. To specify the overlap of at most two of the K materials in a single voxel, we define $K-1$ cubic splines. With $0 < \tau\prime \leq 0.5$ we define a smoothness parameter

$$\tau^k := \tau' \left(\rho^{k+1} - \rho^k \right) \tag{3.3.1}$$

and additionally the central point

$$R^k := \frac{1}{2} \left(\rho^k + \rho^{k+1} \right).$$

---

[1] Those artifacts are similar to Moiré patterns.

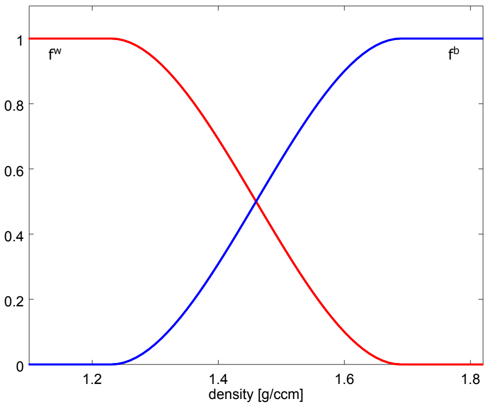


Figure 3.3.1: As an example parameter set, ($\rho^w = 1.0\,\mathrm{g/cm}^3$, $\rho^b = 1.92\,\mathrm{g/cm}^3$, $\tau' = 0.25$) is used in this work for the clinical head scan data in Section 4.2.3. The plot shows the resulting material fractions of water f^w (red line) and bone f^b (blue line) as a cubic polynomial of the density ρ.

Then we can define the threshold values $\rho^k_{min} := R^k - \tau^k$ and $\rho^k_{max} := R^k + \tau^k$ which finally enables us to write down the overlap of two materials as

$$f^k(\rho) = \begin{cases} 1 & \rho^k_{min} < \rho < \rho^k_{max} \\ f\left(\rho; R^k, \tau^k\right) & \rho^k_{max} < \rho < \rho^{k+1}_{min} \\ 0 & \rho > \rho^{k+1}_{min} \end{cases}. \tag{3.3.2}$$

Of course, the fraction of the complementary material follows as $f^{k+1}(\rho) = 1 - f^k(\rho)$ within the respective interval.

The function $f\left(\rho; R^k, \tau^k\right)$ is used to smoothly connect the edges of the fraction of the interval between ρ^k_{max} and ρ^{k+1}_{min} as shown in Figure 3.3.1

$$f\left(\rho; R^k, \tau^k\right) := \frac{\left(\rho - R^k\right)^3}{4\left(\tau^k\right)^3} - \frac{3\left(\rho - R^k\right)}{4\tau^k} + \frac{1}{2}. \tag{3.3.3}$$

In this work we limited our implementation to a maximum of $K = 3$ (water, bone and a metal compound). This is due to specific aspects of the RGBA standard representation of the OpenCL buffers. More details are described in the Subsection 3.4.2.

3.4 The PSR algorithm

This subsection gives a brief overview of the building blocks comprising the PSR. As every iterative reconstruction technique, it basically compares the measurements with the forward-projection of the volume. In the following, we call the volume of the current iteration just *the iterate*[2] and we use the abbrevation US to refer to the update step and FP/BP to the forward-/backward projection.

The comparison of the measurements with the FP of the iterate determines the contribution that is backward-projected in order to introduce a change (represented by the US) of the current volume that is more consistent with the measurements. The program begins by allocating all the needed memory. Of course, this includes memory for the measurements Y_i (introduced in Section 2.4.2), the current iterate ρ_j and the US of Equation 2.4.35, the FP of ρ_j namely $\bar{Y}_i$ (Equation 2.4.23), and its K derivatives[3] $\frac{\partial}{\partial s_i^k}\bar{Y}_i\left(\vec{s}_i^{\,n}\right)$. Thus, the amount of memory can be expressed by the numbers N_j and N_i introduced in Section 2.1.2. To give the reader a better understanding for such numbers, a typical head scan measured at a hospital (Section 4.2.3) results in $N_i = 1280 \times 960 \times 496 \approx 6 \cdot 10^8$ and $N_j = 512 \times 512 \times 400 \approx 10^8$, a short scan protocol of the SIEMENS Artis Zeego[4] used in Section 4.2.1 yields $N_i = 640 \times 480 \times 133 \approx 4 \cdot 10^7$ and $N_j = 256 \times 256 \times 199 \approx 10^6$. The very first CPU-implementation[5] of the PSR took around one week to execute 100 iterations for a short scan data set. Note that the system

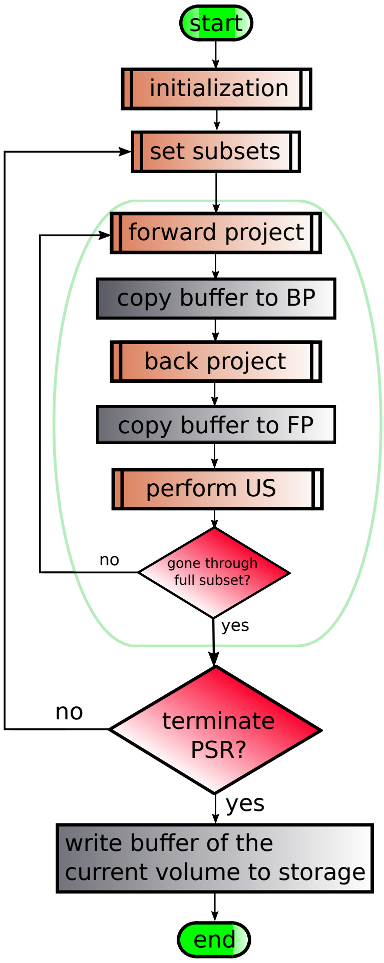

start
initialization
set subsets
forward project
copy buffer to BP
back project
copy buffer to FP
perform US
no
gone through full subset?
yes
no
terminate PSR?
yes
write buffer of the current volume to storage
end

Figure 3.4.1: PAP of the PSR, pseudo code of the PSR is given in Algorithm 1 at the end of Subsection 3.4.4.

---

[2] The current iteration $(n+1)$ refers to the current volume ρ_j^n of Equation 2.4.35.

[3] In this work K is either 1,2 or 3.

[4] Artis Zeego C-arm system SIEMENS Healthcare, Germany

[5] This version suffered from a very inefficient backward-projector.

matrix A of the above setting corresponds to $6 \cdot 10^8 \cdot 10^8 = 6 \cdot 10^{16}$ 32-bit floating point numbers. A has a lot of vanishing elements, but storing of 210PB is practically impossible. Thus, it is a necessity to calculate its elements a_{ij} on-the-fly.

The green line in Figure 3.4.1 indicates that the important memory is represented by OpenCL buffers due to the multi-GPU implementation of the algorithm 2.4.35. Those OpenCL buffers need to be written and read accordingly. 20 iterations of PSR need around one hour on GPU hardware that was bought in 2013. In the following subsections of this chapter the most important aspects of the reconstruction routine are presented.

To decide what is mandatory after the iterations and when to stop the iterative reconstruction (“terminate PSR?”) is described in the Subsection 3.4.4.

3.4.1 The initialization of the reconstruction program

The implementation used in this work is based on a simple “parameter text file”. The user configurates the program by adjusting an ASCII based file. One can categorize the user defined specifications (see Section 2.5) as:

- file paths to the C-arm CT data,
- geometry information,
- prior knowledge and
- hardware and software specific parameters.

File paths to the C-arm CT data which is stored as extinctions ϵ (32 bit precision) and projection matrices Ω (64 bit precision). Typically such raw data is just a set of floating point numbers. In order to extract the single detector pixel values, one needs to know how many pixels the detector consists of and copy them into the memory as a series accordingly.
Geometry information such as the geometric size of the volume $\vec{v}, \vec{V}, \vec{N}$ and detector settings are needed by the FP and the BP. The number of views N_i, voxels N_j, and number of detector pixels in u, v-direction are needed as well and will be read and copied to their memory counterparts.
Prior knowledge is mandatory in order to make the PSR work. File paths to the known energy dependent material absorption $m(E)$, detector response function $D(E)$, and X-ray spectrum $X(E)$ tables have to be set by the user.

Iterative reconstruction methods always include the feature of a proper initialization. Usually this initialization can be the empty volume that is basically $\vec{\rho} = 0$ or an FBP reconstruction which is almost free in terms of computational overload. To transform the volume from HOUNSFIELD units[6] to density, we take the easy way out and assume that the object consists of the first and most important material (e.g. water). Reminding the reader of d_j^1 in Equation 2.4.33, we assume that the initialization volume $\vec{\mu}_{\text{FBP}} \approx m^1 \left(E_{\text{Eff}}\right) \vec{\rho}$ consists of this very material for the most part and that the reconstruction leads to linear attenuation values that are contributed from the effective energy given by $X(E)$, $D(E)$ and Equation 2.4.32.

For simulation studies (supported with a given ground truth) it might come in handy to calculate the root mean square error (RMSE) of the iterate during the reconstruction. The user can enable a feature that will export the RMSE to the ground truth and the current iteration number to a text file.

If the user has reconstructed this data set before, the object d_j^1 was stored for later use. The user can refer to the file path as well and just load these numbers instead of calculating them by the backward-projecting operator one more time.

Hardware specific parameters evolve as a consequence of the GPU devices that are used in order to increase the computational performance. If the system consists of multiple GPU devices and the user wishes to use only a certain number of those devices he or she can specify the hardware specific name of the chip. The device offset can be utilized in case the user wishes to keep the first devices open for the OS or other applications, e.g. another instance of the PSR that is reconstructing a different data set. The default list of OpenCL devices is used in case the user did not type a correct device name[7].

Software specific parameters contribute the largest amount of features that the implementation offers to the user.

- If the user wants to ignore outer pixel areas of the detector - e.g. because of collimator blur - he or she can define 4 numbers, which determine the new rectangle that is cut out of the measured projections[8]. When enabled, the feature also will change the projection matrices Ω accordingly.

- The algorithm relies on spectral information encoded by the minimum energy that is accounted for, and the sampling interval width of the data tables. Those two numbers have to be defined[9], i.e. if the tube voltage was set to 108keV, the used setting for such a scan was $E_{\text{min}} =$ 28keV. Since the default value of sampling points is 40 energy bins (should not be changed, but can be changed in the parameter file), thus $\triangle E = 2$keV.

---

[6] Hounsfield units define the CT number by the attenuation of water μ_w. If μ is equal to μ_w the corresponding CT number is equal to zero. This number is a measure of how much promille the attenuation differs from water. Air has the CT number of -1000. This defines the linear transformation of the original attenuation $\mu \longmapsto \mu^{[\text{HU}]} = \frac{\mu - \mu_w}{\mu_w - \mu_{\text{air}}} \cdot 1000$.

[7] [createDevList(); setContributions(); @polyreco.cpp]

[8] [truncateProjectionsUniform(QRect sector, std::vector<float> *); @systemgeometry.cpp]

[9] [loadSpec(QString); @systemspectrum.cpp]

- The displacement model has the degree of freedom represented by the smoothness parameter $\tau\prime$ and the density threshold values ρ_{tissue} defined in 3.3.2 and 3.3.3[10].

- The parameter β of the regularization and the feature of averaging at the end of one iteration step is enabled with a Boolean parameter.

- The number of iterations that the single angle subset is to be applied can be chosen as well. Once these iterations are finished, the subset size will successively be doubled and repeat the iterations by a different amount given by another parameter. The subset size increases in this manner until it contains a user defined number. Its default value is 1, but can be changed due to storage issues that are discussed in Subsection 3.4.4.

- To terminate the iterative process the user can choose the maximum amount of iterations or give a threshold $\varepsilon_{\text{terminate}}$ value of the total induced change

$$\varepsilon_{\text{terminate}} > \left|\triangle\rho^{n+1}\right| = \sum_{j=1}^{N_j} \sqrt{\left(\rho_j^{n+1} - \rho_j^{n}\right)^2} \tag{3.4.1}$$

  to the volume defined by the update step 2.4.35.

- Since the C-arm system varies the electrical current that feeds the X-ray source depending on the projection angle, the implementation can use this information by basically weighting the induced change to the volume depending on the specific number of the “current-time-product”[11].

- To account for scatter or “detector handicaps“ such as over-saturation within the forward-projection model, certain features can be enabled by a Boolean parameter that represents a switch to apply different OpenCL kernels[12], which are explained in Subsection 3.4.2.

- It can be recommended to use the so-called **“outer volume feature”** in certain scenarios. It feeds the forward-projector a second volume memory that represents a surrounding voxel volume that is not visible in every single projection angle. The utilization of this feature is justified and described in Section 4.2.3.

---

[10] [loadM_E(QString,QString,QString); defineFracMat(); @systemspectrum.cpp]

[11] [loadTubeVoltageModulation(QString, QVector<float> *); @polyreco.cpp]

[12] [initKernelsAndPrograms(); @polyreco.cpp]

- To export intermediate reconstruction data in-between iterations, the user can define the number of iterations after which such an export shall happen. To avoid overflow of memory, the user can also choose to just append one or more z-slices of the iterate to a file. This feature is useful to get an understanding how exactly the volume $\rho_j|_{z=\text{const}}$ is changing over the course of the iterations.

3.4.2 The forward-projection step

For every single detector pixel u, v-pair the ray caster needs to calculate[13] the geometry information discussed in Equation 3.1.2 and derived from the HOUSEHOLDER method as paraphrased in Subsection 2.5. The routine of the ray caster is applied[14] to calculate the forward-projection (FP) as Equation 2.4.19 models the expected intensity depending on the volume (and prior knowledge) to be

$$\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right) = \int_{\mathbb{R}+} I_i\left(E\right) \exp\left\{-\vec{m}\left(E\right) \cdot \vec{s}_i\left(\vec{\rho}^n\right)\right\} \mathsf{d}E.$$

The integral over the photon energy E becomes a sum over discrete and equidistant energy levels e_M. In this work we used 40 equidistant sampling points in the interval of interest. For a typical head scan this means we sample the functions $X(E)$ and $D(E)$ plotted in Figure 2.2.2 by simply averaging the corresponding tabulated numbers of the specific interval. The same applies to $m_k(E)$ for each material. The data points are given by the NIST data base for the energy-dependent attenuation coefficients [HS96]. At first, the ray caster has to calculate

$$\vec{m}\left(E\right) \cdot \vec{s}_i\left(\vec{\rho}^n\right) = \left[m_1(E), m_2(E), m_3(E)\right]^{\mathrm{T}} \cdot \left[s_i^1\left(\vec{\rho}^n\right), s_i^2\left(\vec{\rho}^n\right), s_i^3\left(\vec{\rho}^n\right)\right]$$

for each energy interval e_ω. This is represented by 40 different 32 bit floating point values which in the end get summed to calculate $\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right)$ and its $k = \{1, 2, 3\}$ derivatives $\frac{\partial}{\partial s_i^k}\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right)$. In order to make the update step less computationally intensive, one could consider lowering the amount of energy intervals e_ω to 10 or even 3 as Wu et al. [WYMF14] indicated in their work. Here the 40 intervals were chosen - basically to maximize the accuracy - depending on the actual capacities of the GPU hardware that was utilized[15]. As a short reminder, the three components m_k represent the different material-depending attenuation tables. From this vantage point, the three numbers stored in $\left[s_i^1\left(\vec{\rho}^n\right), s_i^2\left(\vec{\rho}^n\right), s_i^3\left(\vec{\rho}^n\right)\right]$ can be interpreted as the forward-projections of those three specific materials that are mapped only by the density values with respect to the model $f^k\left(\rho\right)$ (see Equation 3.3.2), which is realized by one function within the OpenCL kernel and two *if* cases[16].

---

[13] [void writeATandQRinto(double* , double*); @spaghettipseudoinverse.cpp]

[14] [PSR_set_Y(...); @PSR_set_Y.cl]

[15] 8 GPU devices were used and made of 1 to 4 AMD FirePro S10000 graphics cards each with two 'Tahiti' chips.

After the ray caster has estimated the $\omega = 1, 2, \dots 40$ numbers $\vec{m}\left(e_\omega\right) \cdot \vec{s}_i\left(\vec{\rho}^n\right)$, we can approximate the four numbers that represent the integrals of $\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right)$ and $\frac{\partial}{\partial s_i^k}\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right)$ as follows

$$\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right) = \int_{\mathbb{R}+} I_i\left(E\right)\exp\left\{-\vec{m}\left(E\right)\cdot\vec{s}_i\left(\vec{\rho}^n\right)\right\}\mathsf{d}E$$

$$\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right) \approx \triangle E \cdot \sum_{\omega=1}^{40}\exp\left\{-\vec{m}\left(e_\omega\right)\cdot\vec{s}_i\left(\vec{\rho}^n\right)\right\}D(e_\omega)X(e_\omega) \tag{3.4.2}$$

$$\frac{\partial}{\partial s_i^k}\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right) \approx \triangle E \cdot \sum_{\omega=1}^{40}\left\{-m_k\left(e_\omega\right)\right\}\exp\left\{-\vec{m}\left(e_\omega\right)\cdot\vec{s}_i\left(\vec{\rho}^n\right)\right\}D(e_\omega)X(e_\omega).$$

Of course, the interval width is determined by the overall energy range

$$\triangle E = \left(E_{\max} - E_{\min}\right)/40.$$

Because the forward-projection consists of four numbers per detector pixel, one can use the RGBA standard that GPU hardware is optimized for. Thus, one forward-projection memory contains not only the intensity values of the forward-projection $\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right)$ for each detector pixel, but also the three material specific derivatives $\frac{\partial}{\partial s_i^k}\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right)$ which represent a specific weighting in the backward-projection 2.4.34

$$\hat{N}_j\left(\vec{\rho}^n\right) = \sum_{i\in\eta}\sum_{k=1}^{K} a_{ij} f_j^k\left(1 - \frac{Y_i}{\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right) + \bar{r}_i}\right)\frac{\partial}{\partial s_i^k}\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right).$$

This “factor 4 memory” can limit the used subset size in the current implementation. With a higher count of angles used per subiteration, one would need to increase the memory linearly to that number. The current implementation of the forward step is working “angle wise”, but uses one big read-write memory. This problem can be addressed by possible implementation changes, but was not considered. No memory overflow was observed when working on all the data sets that got processed in order to get results with the PSR.

---

[16] [float frac_W(const float ,const float , const float); @PSR_set_Y.cl]

Due to the need to calculate $\frac{\partial}{\partial s_i^k}\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right)$ within the OpenCL kernel and the factor 4 of memory read/write access it could be a *very interesting* research topic to find a viable approximation $y_Y\left(\cdot\right)$ of $\frac{\partial}{\partial s_i^k}\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right) \approx y_Y\left(\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right)\right)$. A motivation for such an idea could be a fast converging series expansion of the integral around the centroid E_{Eff} as in d_j Equation 2.4.31.

Additional features We note the possibility of adding a **scatter estimation** r_i in Equation 2.4.23 and eventually in Equation 2.4.34, a feature of the PSR that works in simulations to reduce scatter artifacts. If one had an approximation provided by scatter kernel convolution of the projection space or another viable method, one could possibly support the reconstruction with such information to further increase the image quality.

In a similar fashion, the PSR routine can approximate **detector saturation**. In the presence of objects consisting of high absorbing materials (a 5 Euro-cent coin or a metallic screw), it was observed that extinction values are not consistent with the geometric size of such high absorbing objects (see Subsection 4.2.2). The saturation model approximates such a behavior by a hyperbolic tangent function $\bar{\epsilon} \longrightarrow a' \tanh\left(\bar{\epsilon}/a'\right)$ that becomes over-saturated when the extinction value $\bar{\epsilon}$ reaches the value a'. The user can define a' or choose the "auto correcture" mode. Here the largest measured extinction value is identified to be a'. The transformation of the extinction values is performed within the OpenCL kernel[17]. Such an adaptation of the forward-projection can be interpreted as a first order approximation of the detector saturation, as this aspect is not covered within the mathematical derivation of the update step.

The used OpenCL kernels that execute the ray caster are able to increase their **accuracy** by casting more rays per detector pixel. The default value of $n = 1$ one ray that targets at the center of the pixel can be increased by the user. It is possible to cast any number n^2 of rays per pixel element. This is implemented by an outer loop in the OpenCL kernels that can adjust the beam direction onto 4, 9, 16,... equidistant points on the detector pixel. When terminating the loop, one needs just to take the sum of all calculated forward-projections and multiply it by a factor of n^{-2}. With large numbers ($n > 10$) the used AMD FirePro S10000 that were used started to crash due to the long execution times by single kernels, which is a well known behavior in GPU parallelization[18].

The algorithm also supports the option to forward-project a second volume that surrounds the main volume. The **outer volume feature** reduces the parts of the ray casting that account for this region to constant step length sampling. The information gain is used when calculating the final intensity. This enables the proper consideration of objects that are visible in only a limited number of X-ray projections (e.g. a patient table). The necessity of such a feature is introduced in the Section 4.1.5.

---

[17] [PSR_set_Y_tanh(...); @PSR_set_Y_tanh.cl]

[18] NVIDIA's cuda has the same problems with shutting down when a kernel exceeds the threshold of the valid execution time.

Spectral forward-projection The framework of the forward-projection algorithm was utilized by another program designed to calculate forward-projections of arbitrary objects and geometries. An interesting purpose is a proper simulation of a bowtie filter or the recently developed granting based spectral CT of Stayman et al. [STG^{+}20]. To add the effects of such a filter to the forward-projections, one needs to save the spectral absorption caused by the filter in front of the X-ray source. In this case, the 40 energy bins worth of information need to be stored. To project an object within this setting, this data needs to be loaded as an initial state to each casted ray (with respect to the geometry) to be considered by the final step of calculating the exctinction values or the intensity.

Also this feature would add the possibility to forward-project more then 3 materials, as different materials could be projected in a series and added linearly in those 40 energy bins.

This forward-projector could be of interest for acquiring simulated data for studies or even to enlarge a data base for certain machine learning tasks.

3.4.3 The backward-projection step

The backward-projecting[19] (BP) operation $\hat{N}_j = \sum_{i \in \eta} a_{ij} \ldots$ is performed when the FP operation has terminated. *In general*, it is not known which voxels j are connected to which detector pixel i in the current subset of angles. This is the reason why the “shadow projection” concept of the voxel-driven BP in Section 3.2 is used.

The RGBA memory that represents $\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho^n}\right)\right)$ and $\frac{\partial}{\partial s_i^k}\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho^n}\right)\right)$ is handed over as a constant reference. This is possible because the kernels share the same GPU memory. Since η is a set of indices which encode *full projection angles*, one needs to hand over the full FP. The computational effort could be reduced if prior knowledge about the specific geometry is used. The circle trajectory is the most common trajectory in C-arm CT protocols. The rotation around the z-axis of the source-detector setup could be used to find a fast solution that computes which parts of the memory that represent the FP are never used. Since the geometry is known beforehand (see Section 2.5), the corresponding memory could be identified and ignored. Just to ignore these parts of the memory could increase the performance by reducing the read/write operations.

---

[19] [PSR_set_N(...); @PSR_set_N.cl]

3.4.4 The iteration of the update step

In order to calculate the next iterate $\vec{\rho}^{n+1}$, we still need to apply the regularization that results in the term βR defined in 2.4.4, hence the need of calculating the two terms $\beta \frac{\partial S}{\partial \rho_j}$ and $\beta \frac{\partial^2 S}{\partial \rho_j^2}$ in Equation 2.4.35. The HUBER loss function $\psi_{\triangle\rho_{\text{tissue}}}(\triangle\rho_j)$ of the difference of density $\triangle\rho_j$ with respect to the neighbor voxels around j was chosen, which is quadratic for small and linear for large differences. In terms of image processing this property combines high quality of the low contrast of the tissue range and the sharpness of edges.

$$\psi_{\triangle\rho_{\text{tissue}}}(\triangle\rho_j) = \begin{cases} \triangle\rho_j^2/2 & , |\triangle\rho_j| < \triangle\rho_{\text{tissue}} \\ \triangle\rho_{\text{tissue}}\left(|\triangle\rho_j| - \triangle\rho_{\text{tissue}}/2\right) & , \text{else} \end{cases}$$

Note that we adapted the parameter ($a \longrightarrow \triangle\rho_{\text{tissue}}$) compared to Equation 2.4.11 and 2.4.12 as well. In the early version of this regularization approach we approximated this loss function by a simple and analytical surrogate $\cosh(x) := \{e^x + e^{-x}\}/2$ approach, but we found that the direct *if-else case* and the arithmetic operations of $\psi_{\triangle\rho_{\text{tissue}}}(\triangle\rho_j)$ were speeding up the execution of the respective OpenCL kernel[20]. Since we only need derivatives of

$$S(\vec{\rho}; \vec{\rho}^n) = \sum_{j=1}^{N_j} \sum_{\xi_j \in \text{N.N. of } j} w_\xi \psi_{\triangle\rho_{\text{tissue}}}\left(\rho_j - \rho_{\xi_j}^n\right),$$

the first derivative

$$\left.\frac{\partial S}{\partial \rho_j}\right|_{\vec{\rho}=\vec{\rho}^n} = \sum_{j=1}^{N_j} \sum_{\xi_j \in \text{N.N. of } j} w_\xi \left.\left(\frac{\partial}{\partial \rho_j} \psi_{\triangle\rho_{\text{tissue}}}\left(\rho_j - \rho_{\xi_j}^n\right)\right)\right|_{\vec{\rho}=\vec{\rho}^n}$$

can be implemented as a clamp operation[21] on the interval of $[-\triangle\rho_{\text{tissue}}/2, \triangle\rho_{\text{tissue}}/2]$ and

$$\left.\frac{\partial^2 S}{\partial \rho_j^2}\right|_{\vec{\rho}=\vec{\rho}^n} = \sum_{j=1}^{N_j} \sum_{\xi_j \in \text{N.N. of } j} w_\xi \left.\left(\frac{\partial^2}{\partial \rho_j^2} \psi_{\triangle\rho_{\text{tissue}}}\left(\rho_j - \rho_{\xi_j}^n\right)\right)\right|_{\vec{\rho}=\vec{\rho}^n}$$

becomes only a sign operation and a multiplication with the clamp operation that was already done in the previous step. For the sake of clarity we can state that the *variable* ρ_j in the argument becomes ρ_j^n *after* the differentiation which implies that the differences of the nearest neighbors are calculated on the very same volume memory of the current iterate ρ_j^n.

---

[20] [PSR_set_reg(...); @PSR_set_reg.cl]

[21] The clamp operation usually assigns a constant c_b to each input x below a threshold value T_b and another constant c^a to each x above a threshold value T^a. Between those thresholds, x is not changed. Hence, such an operation can be interpreted as a sigmoid function that consists of three linear piecewise functions of which two are the trivial constants and the "connection" is the identity operator.

Once the mandatory memory has been updated according to the previous subsections, the term $\rho_j^{n+1}\left(\vec{\rho}^n\right)$ can be approximated with the representations of $\hat{N}_j\left(\vec{\rho}^n\right)$, $\left.\frac{\partial S}{\partial \rho_j}\right|_{\vec{\rho}=\vec{\rho}^n}$, and $\left.\frac{\partial^2 S}{\partial \rho_j^2}\right|_{\vec{\rho}=\vec{\rho}^n}$ to

$$\rho_j^{n+1}\left(\vec{\rho}^n\right) = \left[\rho_j^n - \frac{M\hat{N}_j + \beta \left.\frac{\partial S}{\partial \rho_j}\right|_{\vec{\rho}=\vec{\rho}^n}}{d_j^1 + \beta \left.\frac{\partial^2 S}{\partial \rho_j^2}\right|_{\vec{\rho}=\vec{\rho}^n}}\right]_+ .$$

Again, d_j^1 and β are iteration-independent constants and M is the number of subsets η.

Algorithm 1 PSR

```
Compute or load existing $d_j$ utilizing Equation 2.4.33
Initialize volume $\rho_j^0$ with zero or load a reconstruction
Initialize subsets $\eta$
repeat                ▷ if all subsets were used once one iteration is finished
    for all subsets $\eta = 1, \ldots, M$ do
        Compute $\bar{Y}_i$ and $\frac{\partial}{\partial s_i^k}\bar{Y}_i$ using Equation 2.4.22 and 3.4.2
      $s_i^k\left(\vec{\rho}^n\right) = \sum_{j=1}^{N_j} \rho_j^n a_{ij}^k$
      $\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right) \approx \triangle E \cdot \sum_{\omega=1}^{40} \exp\left\{-\vec{m}\left(e_\omega\right) \cdot \vec{s}_i\left(\vec{\rho}^n\right)\right\} D(e_\omega)X(e_\omega)$
      $\nabla_k\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right) \approx \triangle E \cdot \sum_{\omega=1}^{40} \left\{-m_k\left(e_\omega\right)\right\} \exp\left\{-\vec{m}\left(e_\omega\right) \cdot \vec{s}_i\left(\vec{\rho}^n\right)\right\} D(e_\omega)X(e_\omega)$
        Compute $\hat{N}_j$ using Equation 2.4.34
      $\hat{N}_j = \sum_{i\in\eta} \sum_{k=1}^{K} a_{ij} f_j^k \left(1 - \frac{Y_i}{\bar{Y}_i(\vec{s}_i(\vec{\rho}^n)) + \bar{r}_i}\right) \frac{\partial}{\partial s_i^k}\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right)$
        Compute $\triangle\rho_j^{n+1}$ for all $j$ using Equation 2.4.35
      $\rho_j^{n+1} = \left[\rho_j^n - \frac{M\hat{N}_j + \beta\frac{\partial S}{\partial \rho_j}}{d_j^1 + \beta\frac{\partial^2 S}{\partial \rho_j^2}}\right]_+$
    end for
until $n + 1 >$ Maximum Iterations Or $\varepsilon_{\text{terminate}} > \left|\triangle\rho^{n+1}\right|$
```

3.4.5 The multi-GPU aspects of parallelization and buffer access minimization

This subsection provides inside knowledge about the OpenCL parallelization and the buffer management approach respectively. It will guide the interested reader with explanations of the source code that executes the subiterations framed in green in Figure 3.4.1 and explain some specific aspects one needs to consider when using multi-GPU approaches.

Three important OpenCL buffers represent the iterate ρ_j^n:

- the BP $\hat{N}_j$ which allocates the same amount of memory as the iterate,
- the FP that includes $\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right)$ and its $k = \{1, 2, 3\}$ derivatives $\frac{\partial}{\partial s_i^k}\bar{Y}_i\left(\vec{s}_i\left(\vec{\rho}^n\right)\right)$ which require the amount of detector pixels multiplied by the factor of 4 and of the maximum amount of angles that the biggest subset can contain.

If the subset η contains more then one projection angle, the FP kernels are looped over this exact amount of times and write at the position of the buffer with respect to the blockwise memory assignment. The maximum buffer size needed is determined by the user's choice of the biggest subset size were the default corresponds to one single subset.

The FP is divided in v-direction of the detector and split between all GPU devices that are actively picked by the user. For a circular scan trajectory this kind of separation gives best performance results, because the v-direction is usually approximately perpendicular to the axis of rotation of this specific trajectory. The FP workgroup consists of single pixel coordinates u, v. Thus, for each detector pixel one instance of the FP kernel is running. After the FP calculation is finished, scatter estimations can be added if available and chosen by the user.

The "volume" buffers that represent ρ_j^n and $\hat{N}_j$ are divided into y-blocks and split between all GPU devices as well. Then, the BP workgroups consist of single x, y-pairs and the kernels "crawl up" in z-direction. Proceeding in this way, the execution times of those kernels will have low variance, so this scheme is chosen for performance reasons as well. Since it is not known which v-blocks of the FP contribute to the BP, one needs to copy *each* FP buffer to *each* BP y-block. So there is an inherent trade off between number of GPU devices and the detector size. This is a bottleneck to the performance of the PSR implementation on multiple GPU's with "small" detector sizes and number of projections, where one device will be enough. In fact, on two state-of-the-art NVIDIA GeForce RTX 2080 Ti graphics cards there is not even an observable benefit by parallelizing over these devices.

Technically it is possible to perform the regularization $\beta{\partial S}/{\partial \rho_j}$ and $\beta{\partial^2 S}/{\partial \rho_j^2}$ step after finishing the calculation of $\hat{N}_j$. This would require copying the iterate ρ_j^n to the same kernels that calculate the BP. A GPU-coder "rule of thumb" is that it is almost never a problem to start more kernels with smaller tasks then to perform all tasks in one kernel, when there is no overhead produced by copying of buffers. This holds true, because the regularization kernels and the BP kernels share the *same* volume buffer. Thus, it will lower the variance of kernel execution times as well.

Motivated by this fact, the update step is done in one kernel (see Subsection 3.4.4) that is given the buffer of the iterate and the BP. The workgroup is selected in the same way as the BP is parallelized. At the very moment those calculations are finished, the BP buffer contains the next iterate which needs to be copied to the FP buffer and the "old volume" buffer ρ_j^n of the regularization kernels.

If the user has enabled the averaging feature, this new iterate gets added in a separately allocated memory in the RAM and will be used after the full subset has undergone this procedure. The averaging feature will trigger at the end of the subset cycle (which is one iteration) and basically averages over every single volume that was produced in each subiteration of this subset.

It is worth mentioning that before a new iteration starts, the current set of projection angles is permuted randomly[22] to increase induced change of the update step. The theory of compressed sensing suggests that shuffling the projection angles creates the most beneficial update step on average. For the sake of clarity this means that the actual sequence that decides which angle is inducing a change in $\hat{N}_j$ is chosen randomly from the subset that is determined by subset size of the current position in the ordered subset scheme. Candes, Romberg and Tao provided groundbreaking results in 2006, showing that the reconstruction of an undersampled object has the best outcome if the system matrix A is a Gaussian random matrix [CRT06]. Hence, the more random your sampling points are, the better the reconstruction can become.

At the end of the iteration, certain procedures are executed, e.g. saving the iterate on the hard disk. Features like averaging, calculating check sums (like an RMSE to a ground truth) or writing the overall change of the volume into a text file can be applied as well.

---

[22] [permuteSubsets(); @polyreco.cpp]

4 Results

Confronting the PSR with projection data recorded by CT devices can be a delicate task. Note that the current implementation is based around the geometry described in Section 2.5. Hence, to work with CT data, one would need to adapt the geometry of the curved detector first.

This chapter describes the different experiments that were carried out as well as the application of the PSR on real data received from the university hospital Magdeburg. Real C-arm CT data was used to show the viability of the PSR as a potential enhancement of diagnostics in hospitals. Specific simulations were performed to study detailed aspects of the implementation and its bottlenecks.The used projection matrices are given by the C-arm systems. They cover a 200° angle of a circular scan trajectory sampled with 496 or 248 views. The detector resolution is 1240×960 with $0.308 \times 0.308\,\mathrm{mm}^2$ or 616×480 with $0.616 \times 0.616\,\mathrm{mm}^2$. In this section we call the first setting “geometry A” and the latter setting “geometry B”.

The first subsets consisted of a single angle. Thus, the FP of a single angle is performed, the BP within that view is performed. Repeating this for each angle once terminates one iteration. Such an iteration is depicted with the green closed line in Figure 3.4.1. At the beginning of each iteration the subsets are randomized with respect to the specific subset level. After a certain amount of iterations, the subset level was increased to contain more angles. The procedures were repeated until the amount of iterations reaches a user-defined maximum number.

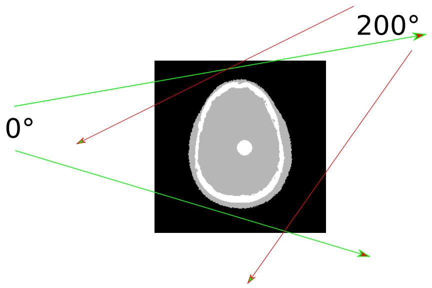


Figure 4.0.1: Scheme of the 200° Scan

4.1 Simulations

As already described in Subsection 3.2, the current implementation of the PSR uses non-adjugated operators that represent the ray based forward-projection (FP, see Section 3.1) and the voxel based backward-projection (BP, see se. 3.2). Hence, the RMSE of reconstructions of perfect projection data cannot converge to zero. Thus, “perfect projection data” refers to the fact, that such data was produced by reading out the forward-projection storage filled by the FP of the ground truth volume. Software phantoms were used which consisted of water and bone [AMN+13]. In some cases a metal object was manually added to the phantom. The forward-projector described in Section 3.4.2 and $\bar{\epsilon}_{\text{poly}}$ in Equation 2.3.3 was used to create artificial extinction projection data. It is assumed to deliver appropriate projection values that are similar to the real C-arm data that is received from the devices, as the very same projection geometry was used. Note that the projection matrices used for those kind of simulations originated from the C-arm systems.

Simulation experiments were carried out to obtain an intuition about the underlying model of physics. If the PSR prior knowledge is given *false* information about the materials and X-ray spectrum $X(E)$ it is expected to observe certain artifacts in the reconstruction. The commonly used subset scheme runs 200 iterations.

We focus only on one slide of the reconstruction to emphasize the characteristics of the PSR and discuss certain aspects such as RMSE or cross section plots that show cupping or blurring.

The number of iterations is set to be very high in simulations. This is to ensure convergence but also to develop a better intuition for the metrics and the compared appearence of the images. When reconstructing real data, the number of iterations is lower compared to the simulations. As real data is corrupted by noise, detector inhomogeneities, image corrections (e.g. scatter correction and over exposure correction) and not “perfect” compared to the forward-projection model, it is impossible to reach the accuracy of simulation results.

Software phantoms

The software phantoms used represent the central slice of a reconstruction analog to a real setting. A slice of a head phantom [AMN+13] was used and copied into 5 z-slices of $256 \times 256 \times 5$ voxels with a spatial resolution of $1.0 \times 1.0 \times 1.0\,\text{mm}^3$. The tissue region inside the skull can contain low contrast information as in Figure 4.1.1. Additionally the tissue can be represented by a constant but contain a high absorbing metal object as in Figure 4.1.8.

To investigate the special case of the measurements done with real data, an additional software phantom was used that mimics the setting with the parameters $256 \times 256 \times 200$ voxels with a spatial resolution of $1.0 \times 1.0 \times 1.0\,\text{mm}^3$ and a disc that contains iron (Fe).

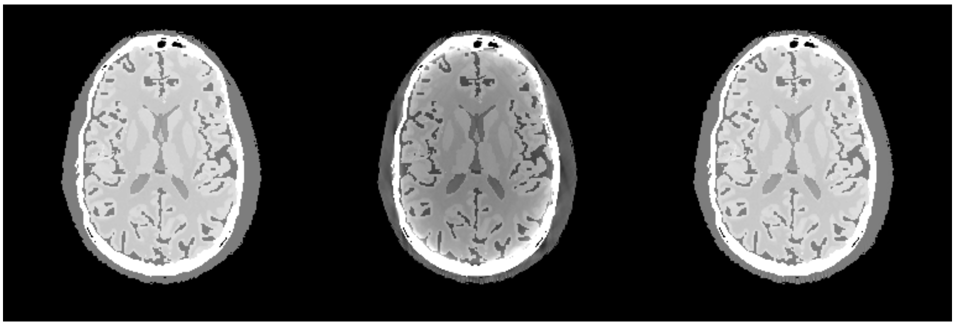


Figure 4.1.1: GT vs. MSR vs. PSR window $0.95 - 1.05\,\mathrm{g/cm^3}$. The MSR shows the characteristic cupping artifact whereas the PSR is converging accordingly.

4.1.1 Monochromatic approximation

An extreme case of such false prior knowledge is the usual assumption of monochromatic X-ray sources. The so-called MSR is defined by giving the PSR only a monochromatic spectrum and a single material as prior knowledge. Thus, the PSR reduces to a monochromatic but statistical maximum likelihood reconstruction.

Imported data The algorithm recieves polychromatic projection data that was produced with the full information about the software phantom including the materials and the X-ray spectrum. Geometry A was used.

PSR configuration The material- and energy dependent attenuation tables are initialized as $m_1(E) = m_2(E) = m_3(E) = m_{\text{water}}(E)$ and the X-ray spectrum is initialized as a single peak $X(E) = X_0 \delta\left(E - E_{\text{eff}}\right)$. Note that the effective energy lies around 66keV. The first subsets were iterated 40 times, after that each higher subset level was iterated 20 times. The overall number of iterations was capped at 200.

Results The software phantom (the ground truth) consists of an outer tissue area, a skull area (cortical bone) and inner tissue area with 4 different constant value types. The MSR shows clear cupping and spill-over of the bones. The small dots of higher density values are less visible due to the lower contrast of the MSR.

Remarks The MSR shows the characteristic beam hardening induced artifacts such as cupping and spill over. The magnitude of the cupping is comparable to cupping that is visible in real head scan data (e.g. in the first row of Figure 4.2.10). As the PSR covers the physics of the absorption process it is not surprising that beam hardening artifacts occur when falsely initialized with a monochromatic spectrum.

The next interesting question is concerned with the accuracy of the prior knowledge about the X-ray spectrum.

4.1.2 Potential spectral mismatches of $I(E)$

Prior knowledge about the X-ray spectrum is a necessity to the PSR. Since it is impossible to perfectly measure the spetrum that was used in a CT scan , it is of interest how the convergence of the PSR is changing with increasing uncertainty of the spectrum. Instead of the exact spectrum that was used to create the projections of the above experiment, the PSR got initialized with lower sampled piecewise constant spectra. Here, we approximate the spectrum with only one constant, to 10 constants and then 15,20,30. As the PSR uses 40 equidistant sampling points of the spectrum, it is expected that the resulting polychromatic reconstructions will suffer from those mismatches. Additionally to the experiment above, it was also initialized with a monochromatic spectrum that has a peak at a different position and is zero elsewhere. Here it was 30 keV and 90 keV and at $E_{\text{eff}} = 66\,\text{keV}$.

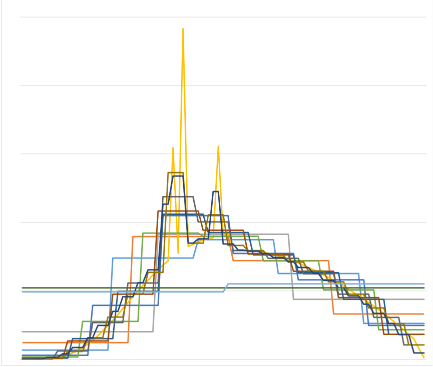


Figure 4.1.2: Sampled X-ray spectra with a corresponding energy range of 28 to 108 keV. The sampling levels are 40,30,20,15,10,9,8,7,6,5,4,3,2 and 1.

As the detector response impacts $I(E) = D(E) \cdot X(E)$ in the same way as the spectrum, different approximations were tested as well. A constant, a linear and a "ramp" function were used. The ramp is defined as two piecewise linear functions that are zero on the first and the last energy bin of the ramp, growing until $E_{\text{eff}} = 66\,\text{keV}$ and decaying linearly again.

Imported data The algorithm recieves polychromatic projection data that was produced with the full information about the software phantom including the materials and the X-ray spectrum. Geometry A was used.

PSR configuration 200 iterations except for one dataset were used. The density is set to $\rho_1 = 1.0\,\text{g/cm}^3$ and $\rho_2 = 1.9\,\text{g/cm}^3$ accordingly. The spectra initialization or the detector response differ for each experiment. As Figure 4.1.2 suggests, several different spectra were used. Additionally three monochromatic spectra and three different detector response functions were used. The "perfect" spectrum was used in these three cases.

Finally the perfect case was iterated over 5,000 iterations to compare convergence properties. A similar reconstruction has been done with 10,000

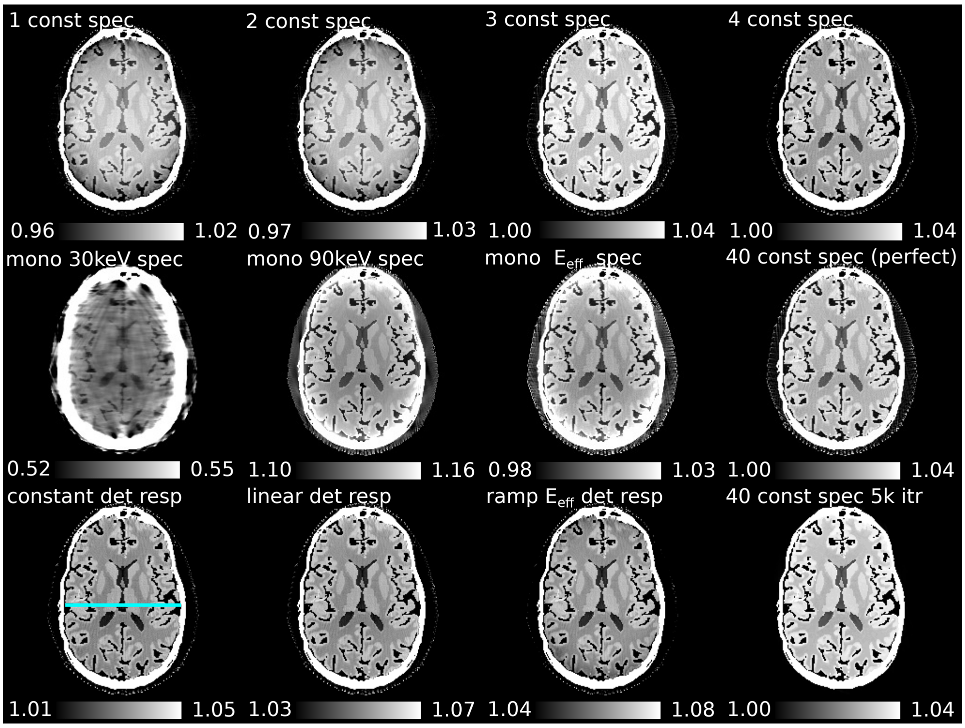

1 const spec
2 const spec
3 const spec
4 const spec
0.96
1.02
0.97
1.03
1.00
1.04
1.00
1.04
mono 30keV spec
mono 90keV spec
mono E_{eff} spec
40 const spec (perfect)
0.52
0.55
1.10
1.16
0.98
1.03
1.00
1.04
constant det resp
linear det resp
ramp E_{eff} det resp
40 const spec 5k itr
1.01
1.05
1.03
1.07
1.04
1.08
1.00
1.04

Figure 4.1.3: Head phantom reconstructions with different $I(E)$ that are depicted in Figure 4.1.2. The first row shows the PSR with differently sampled spectra. The second row depicts differences of the MSR that was fed with different monochromatic spectra and compares them with the perfectly sampled spectrum PSR. The third row shows the performance of the PSR when initialized with different detector response functions $D(E)$ and shows the perfectly initialized PSR that was iterated over 5000 times.

iterations on the skull phantom that contains a disc of nitinol. This reconstruction also was given the perfect projection data.

Results Using only one or two constants as an approximation of the X-ray spectrum results in an inverse cupping effect. With three or more intervals, a proper reconstruction with the PSR is feasible. These results are observed in Figure 4.1.3.

Using the monochromatic spectrum with $E_0 = 30\,\text{keV}$, we can observe a huge loss in low-contrast image quality and an overall decrease of voxel values. Since every material is increasing in terms of attenuation when lowering the photon energy[1], the density of the resulting reconstruction needs to drop compared to the values of the reconstruction that is initialized with $E_0 = 90\,\text{keV}$.

---

[1] see Figure 2.3.1 and Figure 2.3.2

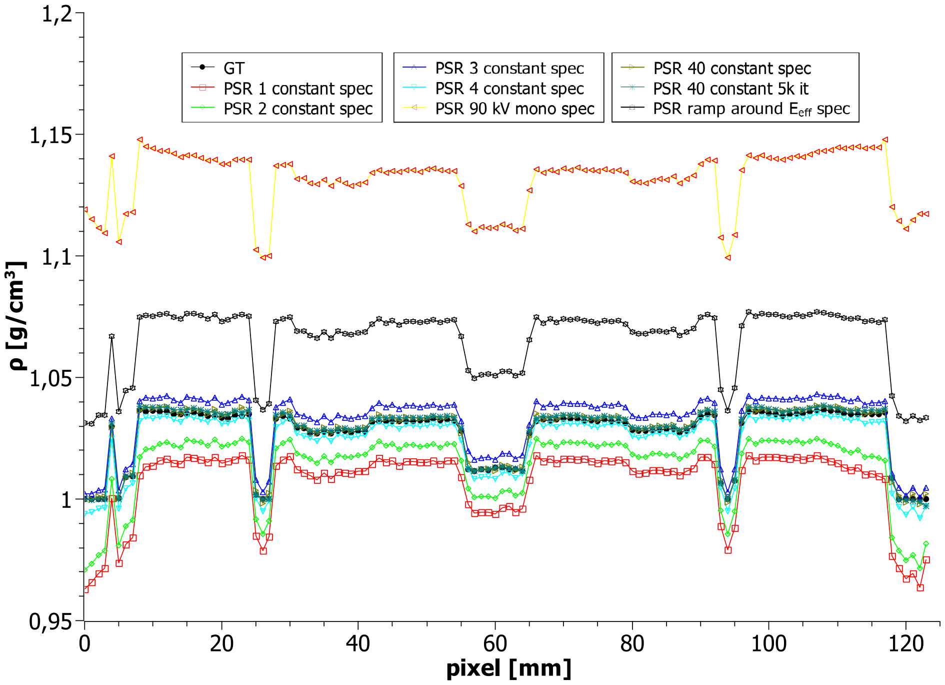


Figure 4.1.4: The cross section plots of the head phantom, which was reconstructed by the PSR, that were initialized with different spectra $I(E)$.

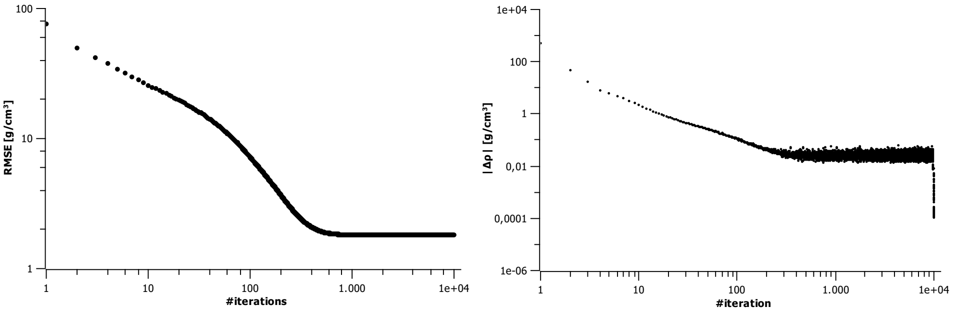


Figure 4.1.5: Convergency plots of the PSR of the head-disc phantom consisting of the RMSE over the course of iterations and the absolute value of the update step $|\triangle\vec{\rho}|$. Note that the RMSE was only calculated for a single slice while the US considers the whole volume.

Peaking the spectrum at E_{eff} leads to average tissue values comparable to the true values of the ground truth (averaging over the cupping). Those observations are the result of the separation approach established with Equation 2.4.20: $\mu = m \cdot \rho$ and support the ansatz that sampling around the effective energy is a good first order approximation.

Remarks Sampling the X-ray spectrum with at least 3 points drastically improves the reconstruction quality. This observation can be explained by the fact that a typical X-ray spectrum "looks like a quadratic function", which has 3 degrees of freedom. This idea is already in use as Wu et al. have shown massive performance boosts in their beam hardening correction approach [WYMF14].

The skull phantom that is used in the following subsections was reconstructed with the perfect setting. The convergence plots in Figure 4.1.5 show a stagnation after a few hundred iterations. The RMSE remains constant after the reconstruction "has finished".

4.1.3 Scattered radiation

Since typical X-ray spectra deliver photon energies between 10 and 100 keV, COMPTON scattering is the most dominant process that attenuates photons while propagating along the original direction. These photons lose energy and change direction (see Section 2.3.3). Detector pixels that are hit by such photons falsely interpret the corresponding signal as non attenuated photons of the original pencil beam that is targeting this detector element. Small scattering angles are more likely to be observed because the energy loss of the scattered photon increases with the angle. Thus, scattered radiation blurs the measurement input.

There were two software phantoms scanned with an artificial scatter signal to see how the PSR performs under such circumstances.

Imported data A skull phantom that is filled with perfect homogenous voxels that consist of water, was modified by an additional metal (NiTi) object represented by density values of $6.45\,\mathrm{g/cm^3}$. The used shape mimics a disc.

A second phantom mimics the 5 Euro-cent coin setup in Subsection 4.2.2. It is a cuboid of water and a coin of iron (Fe) placed within. The density is set to $1.0\,\mathrm{g/cm^3}$ or $7.874\,\mathrm{g/cm^3}$ accordingly.

Polychromatic projections of these phantoms were produced. The used approximation to add "scatter signal" was a convolution with an appropriate kernel in the projection domain (represented by extinction values). This is a viable approach that is used in state-of-the-art methods to estimate for scatter radiation in X-ray simulation techniques [BTK99,SSL10]. We applied

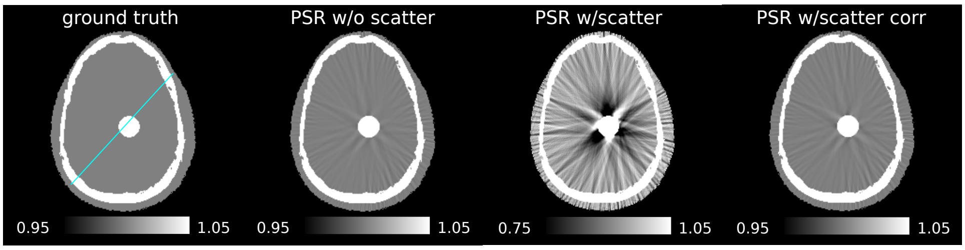


Figure 4.1.6: Disc: GT vs. perfect data PSR and uncorr PSR vs corr PSR

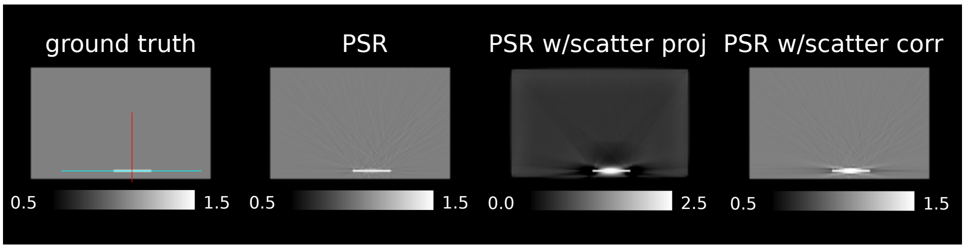


Figure 4.1.7: Coin: GT vs perfect data PSR and uncorr PSR vs corr PSR

a convolution with a Gaussian function of $\sigma = 50\,\text{pixel}$ to the forward-projection of the software phantom. The magnitude of scattered extinction values 50% of the original signal. Glover reports magnitudes of around 10% [Glo82]. These numbers were stored separately to feed them into the PSR to enable the scatter correction feature and were subtracted of the original projections to build the projections containing the scatter signal.

PSR configuration It is observable that on simulated data, the single angle subset iterations lead to the biggest change in the volume domain, hence $|\triangle\vec{\rho}| = \left|\vec{\rho^{n+1}} - \vec{\rho^{n}}\right|$ is largest when using only one angle. This is the reason why the PSR was iterated over the single angle subset for 900 times before reducing the subset level according to the scheme presented in Section 3.4. Setting B and threshold values according to TiNi or Fe were chosen. The regularization parameter was set to $\beta = 0.1$.

Results The p-values for testing the null hypothesis that there is no relationship between the different reconstructions and the ground truth is identified as a measure of correlation between those images. Scatter typically reduced the overall voxel values. This is due to the scatter approximation that is basically lowering the extinction values.

The error introduced due to scatter can be reduced but not eliminated. The coin seems to be a more difficult task for the PSR as the perfect data reconstruction still has an RMSE that is more then 3 times higher than the

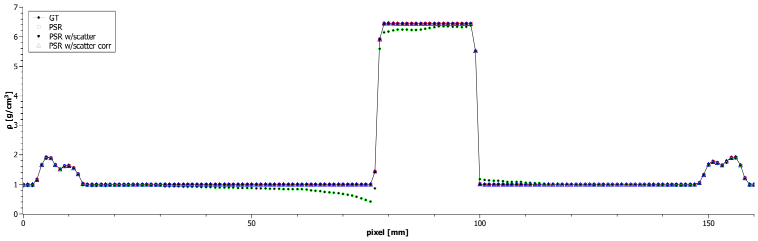


Figure 4.1.8: Disc: cross section along the line plotted in Figure 4.1.6

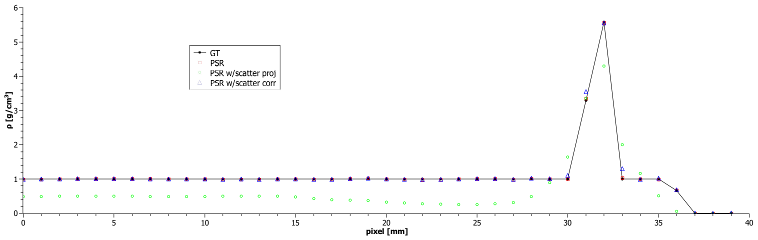


Figure 4.1.9: Coin: cross section along the red line indicator in Figure 4.1.7.

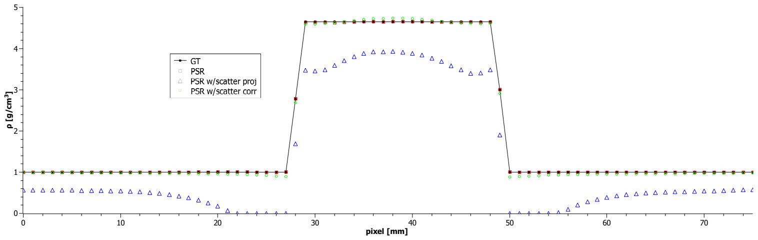


Figure 4.1.10: Coin: cross section along the blue line indicator in Figure 4.1.7.

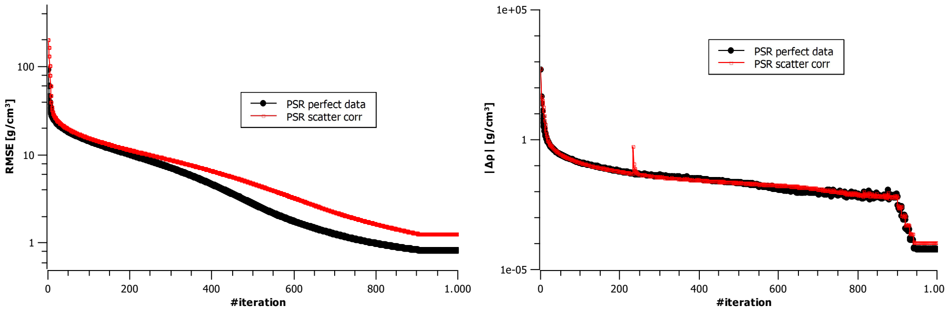


Figure 4.1.11: RMSE and US are plotted for the scatter case of the disc phantom.

disc phantom	RMSE of slice $[g/cm^3]$	min value$[g/cm^3]$	correlation
PSR perfect data	0.83	0.99	0.9999982
uncorrected PSR	-	0.22	0.9976777
corrected PSR	1.25	0.987	0.9999971
coin phantom	RMSE of slice $[g/cm^3]$	min value$[g/cm^3]$	correlation
PSR perfect data	4.13	0.93	0.999956
uncorrected PSR	-	0.0	0.9473
corrected PSR	4.34	0.78	0.9996

Table 4.1: Computed metrics of reconstructed phantoms with differently initialized projections with regard to the scatter model. The RMSE of the uncorrected reconstructions is left out because of the overall offset observed in grey values depicted in Figure 4.1.6 and Figure 4.1.7. The min value refers to the smallest voxel value around the metal object. The correlation was calculated by the MATLAB command corrcoef(A,B).

RMSE of the disc reconstruction. This is the RMSE of the depicted slices with respect to the known ground truth. The smallest voxel value around the metal object is an indicator of the quality of the reconstruction as this value should lie around the value of the water phantom of $1.0 g/cm^3$.

Remarks The feature of scatter artifact reduction is viable when supported with a proper scatter estimation. Here, the perfect scatter estimation was given, resulting in an almost identical reconstruction compared to the scatter free PSR. The uncorrected PSR suffers from heavy shadow artifacts surrounding the metal object. The PSR can compensate for scatter induced artifacts when supported with a proper estimation. This is indicated by the similarity between the scatter free reconstruction and the scatter compensated version in Figure 4.1.11.

The sudden stagnation of the absolute value of the US is caused by the changing subset level. One can explain this by the fact that the number of update steps performed in one iteration of a certain subset level is decreasing the more angles a subset consists of. This means that the factor M is not sufficiently chosen in terms of convergence. A design change in $M(|\xi|)$ or a manual relaxation (see Equation 2.4.35) with respect to the subset level could enhance the convergence speed while conserving the image quality. Alternatively one could change the design of d_j. A discussion by Fessler can be found in [KPTF13].

4.1.4 Detector saturation effects

As metallic objects cause even stronger beam hardening artifacts then bone and tissue, it is of importance to observe the behavior of the PSR in presence

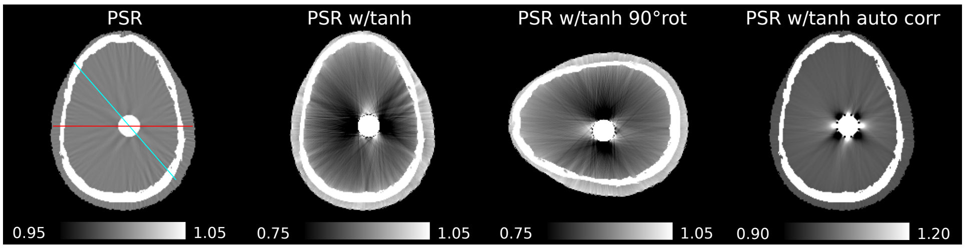


Figure 4.1.12: The perfect data PSR vs. uncorr PSR (rotated and non rotated) vs. corr PSR show distinct artifacts.

of very high absorption rates and even in cases of photon starvation. In presence of such objects (a coin or a metallic screw) it is observed that extinction values are not consistent with the geometric size of such high absorbing objects (see Subsection 4.2.2). A simple saturation model approximates such a behavior by a hyperbolic tangent function $\bar{\epsilon} \longmapsto a' \tanh\left(\bar{\epsilon}/a'\right)$ that becomes over-saturated when the extinction value $\bar{\epsilon}$ reaches the value a'.

Around this extinction value a' the forward-projection becomes non-linear in its argument due to the nature of the used $a' \tanh\left(\bar{\epsilon}/a'\right)$ that reaches the value a' when $\bar{\epsilon} \longrightarrow \infty$.

Imported data This model was applied as post processing on the perfect data produced with the forward-projection of the skull phantom that includes a disc and the water cuboid that mimics the coin setup. a' was set to 5.0. Geometry B was used with respect to the real data case in Section 4.2.2.

The skull phantom is the same that is used in Section 4.1.3. Since the artifact that is caused by the non-linearity is found to be non-isotropic, the PSR was also initialized with projections that have the same geometry but the phantom was transposed slice per slice. Note that this is not a real rotation, but more a reflection of the image on its diagonal (e.g. with a HOUSEHOLDER transformation).

PSR configuration Iterations, thresholds and regularization were chosen to be identical to the scatter experiments in Section 4.1.3.

Results The modelled saturation leads to characteristic artifacts within the PSR. It is observed that there are areas of higher and lower voxel values. These areas depend on the shape of the phantom.

Rotating the disc phantom results in a rotation of the artifact as well (see Figure 4.1.12). The orientation of the brighter and darker areas changes when rotating the coin out of the plane of rotation of the circular scanning trajectory.

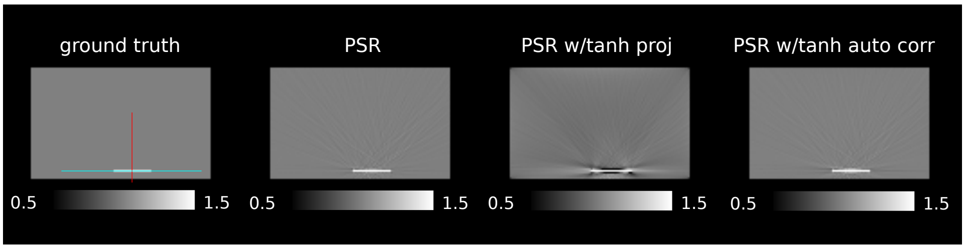

ground truth
PSR
PSR w/tanh proj
PSR w/tanh auto corr
0.5
1.5
0.5
1.5
0.5
1.5
0.5
1.5

Figure 4.1.13: The GT vs. perfect data PSR vs. uncorrected PSR vs. corrected PSR of the coin phantom which were initialized with tanh-saturated projections show only small artifacts.

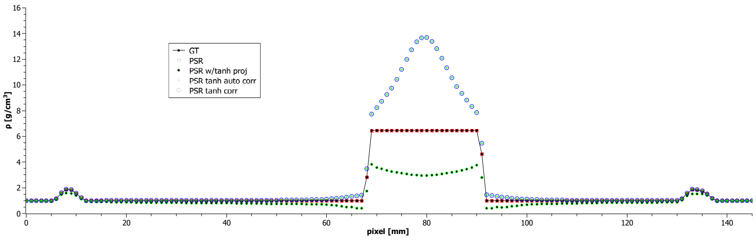


Figure 4.1.14: The cross section plot of the disc phantom PSR that was initialized with tanh-saturated projections.

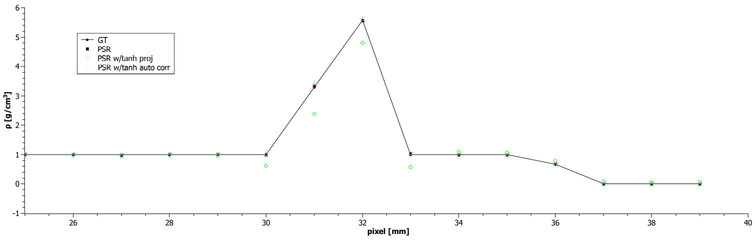


Figure 4.1.15: The cross section plot along the thickness of the coin phantom phantom PSR that was initialized with tanh-saturated projections.

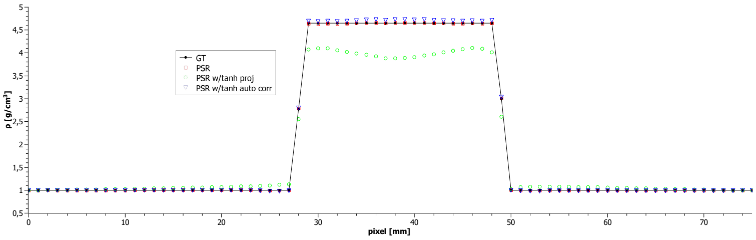


Figure 4.1.16: The cross section plot along the diameter of the coin phantom PSR that was initialized with tanh-saturated projections.

disc phantom	RMSE of slice $[\mathrm{g/cm^3}]$	min value$[\mathrm{g/cm^3}]$	correlation
PSR perfect data	0.83	0.99	0.9999982
uncorrected PSR	156.36	0.0	0.9618974
corrected PSR	126.59	0.0	0.9717095

coin phantom	RMSE of slice $[\mathrm{g/cm^3}]$	min value$[\mathrm{g/cm^3}]$	correlation
PSR perfect data	4.13	0.93	0.999956
uncorrected PSR	6.86	$4 \cdot 10^{-5}$	0.9980199
corrected PSR	4.12	0.88	0.999925

Table 4.2: Computed metrics of reconstructed phantoms with differently initialized projections with regard to the detector saturation model. The min value refers to the smallest voxel value around the metal object. The correlation was calculated by the MATLAB command corrcoef(A,B).

Remarks Since the saturation effect is modelled as a non-linearity, the induced artifacts are similar to beam hardening artifacts. Shadow artifacts and brighter areas perpendicular to these shadows are observable in Figure 4.1.12. As such artifacts are not isotropic, a 90° rotation of this phantom was scanned and the artifact does not change its relative direction. Hence, the artifacts are independent on the scanning geometry depicted in Figure 4.0.1. The artifact changes when the coin is placed perpendicular to the plane of rotation of the scanning protocol.

Comparing the tendency of increasing and decreasing voxel values around the metal object with the scatter experiments, one can see that these artifacts appear to cancel each other out (compare with Figure 4.1.7).

During the iterations it was observed that early iterates showed a dark shadow artifact similar to artifacts that are caused by scatter right where the higher iterations cause brighter regions. This leads to the conclusion that the statistical weightings of the PSR have a larger impact on the early iterations.

4.1.5 Truncation of the outer volume in the projection space

As clinical data always contains the patient table in some projections but never in each view, we can speak of truncation of such a table in the projection space. Typically the tables are designed to add very little attenuation to the measurements and common reconstruction techniques do not suffer alot from this truncation [BBR19]. However, the PSR is sensitive to such a setting. As these objects are typically outside of the reconstruction volume, the PSR tends to “smear” increased voxel values to the borders of the volume. Since the object contributes only to a limited number of views, but the forward-projection step accounts for those voxels in each view, such a truncation setting leads to inconsistencies of the forward projection model.

Ultimately, this results in reconstruction artifacts. It is not needed to accurately reconstruct the truncated objects but only to somehow approximate its impact on the measurements compared to the forward-projection.

This is mainly caused by the structure of the denominator d_j^1 within the US (Equation 2.4.34). Comparing this US to the standard ART approach of the Kaczmarz method (Equation 2.1.6) one can state that the weighting with the reciprocal of $\vec{a}_{i_k}^{\mathrm{T}} \cdot \vec{a}_{i_k}$ can be roughly interpreted as the square of the path length of the i-th ray:

$$\vec{a}_i^{\mathrm{T}} \cdot \vec{a}_i = \sum_j (a_{ij})^2 \leq \left(\sum_j |a_{ij}| \right)^2 = \gamma_i^2 .$$

The backward-projection

$$d_j^1 \propto \sum_{i=1}^{N_i} \gamma_i a_{ij} Y_i$$

on the other hand is only weighted with the path length which makes a difference in the outer volume (OV) parts that are not fully covered by each projection [FBMR17].

To remove the inconsistency of the forward projection, the algorithm needs to support the accurate position of the attenuating voxels that are “out of scope” in the default settings. The first idea is to simply enlarge the reconstruction volume. However, the latter would reduce the resolution and the first would increase the needed memory and computational effort. Thus, Moore’s law would solve the problem at some point but this thesis cannot wait any longer. The idea proposed in this work amalgamates both of the two ideas.

A second volume is defined to represent the OV. The forward-projector is handed the OV that is designed to fully cover the reconstruction volume (RV). The ray caster is modified to incorporate the OV. To reduce the induced computational overhead, the OV is designed to be a constant. Thus, not to be object of the update step at the end of each iteration. To decrease memory, the OV can have a very coarse voxel resolution. Remember, the accurate reconstruction of the OV is not of interest. Additionally, the ray caster does not need to be as accurate as presented in Section 3.4.2. The current implementation uses the built-in feature of interpolating between nearest neighbor voxels as such an operation comes basically for free on OpenCL devices.

Directly measuring the OV (e.g. by a CAD model of the table, photographing or other possible complicated techniques) and trying to calibrate this kind of prior knowledge would probably lead to a bottleneck of alignment problems and might be over the top in terms of needed quality.

In order to obtain a sufficient approximation of the OV that is consistent within the geometry of the PSR, one can simply enlarge the voxel size until the object fits in the RV and perform a few iterations of ordinary PSR to get a good approximation of the shape and position. This reconstruction is then consistent with the forward model. To shorten the reconstruction time of the larger volume, it is considered to use only a fraction of views. Additionally, a high regularization can be applied to smoothen potential artifacts caused by this angular undersampling. Hence, enlarging the physical size of the reconstruction volume gives a hint about the truncated structures which might already represent a sufficient approximation and eventually remove the truncation artifacts. This approach is an extension of Ziegler et al. [ZNG08].

Imported data The software phantom is modified by adding a box of low attenuating voxels at the bottom of the doubled volume. These voxels mimic the truncated table in the OV. Polychromatic forward-projections were calculated. Note that this means in particular, the geometry matrices were unchanged, hence the table was truncated similar to a realistic setup.

PSR configuration We performed 200 iterations of PSR with different settings such as standard size and one with enabled outer volume feature (with the perfect outer volume and an estimated outer volume). The forward-projection step of the PSR is handed an additional volume that is “out of scope” (see Section 3.4.2).

Only every 4th projection angle is considered for the rough estimation of the OV reconstruction. It was calculated with only 20 iterations of PSR of doubled physical voxel size and an increased regularization parameter $100 \cdot \beta$. In this way, the table is represented by a smooth object within the OV.

Results The lower sampled PSR of the doubled volume size gives a rough estimate of the table phantom depicted in Figure 4.1.17. The usage of this estimation delivers appropriate PSR results in the setting of table truncation compared with the PSR that uses the information provided by the ground truth (see Figure 4.1.18).

Remarks Truncation of an object in a realistic setting leads to a characteristic artifact within the PSR. Such a characteristic behavior is observed in the real data case (see Section 4.2.3). The proposed idea of sampling the outer volume with a less computationally heavy PSR as a prior step delivers sufficient image quality.

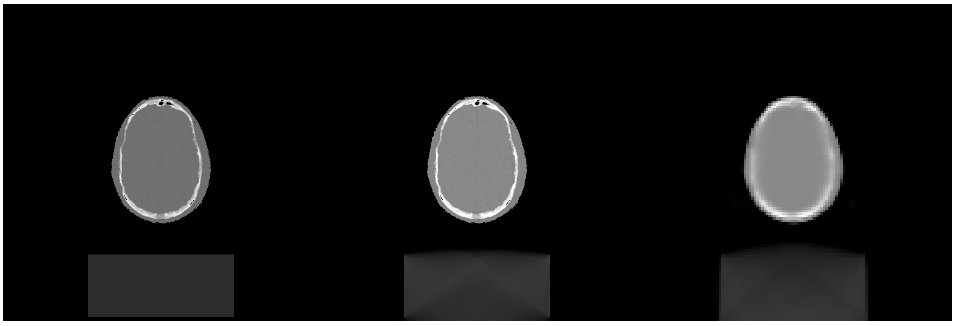


Figure 4.1.17: GT vs. PSR with the same size as the GT and the full projection data vs. PSR of lower voxel number, lower resolution and less projection data (window level: $0.00 - 1.92\,\mathrm{g/cm^3}$).

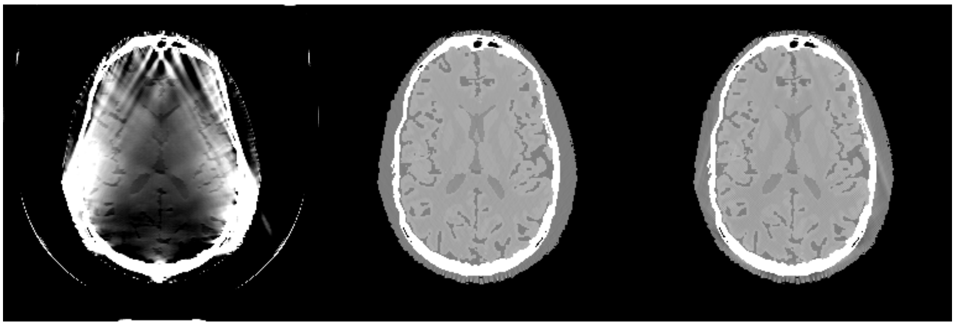


Figure 4.1.18: PSR in the standard setting vs. PSR given the perfect OV vs. PSRov (window level: $0.90 - 1.10\,\mathrm{g/cm^3}$).

4.2 Real C-arm CT data

Working with real data can be a challenging task in many fields of applied sciences. Disturbances within real data such as noise, systematic errors, misalignment and insufficient calibrations might occur. Finding the root of certain obstacles might take a long time or render the problem unsolvable. Additionally, medical devices manufactured for clinical routines are typically very protective of their processing pipeline and internal procedures. Exporting CT projection data requires additional information about the exact measurement geometry. Such information is represented as projection matrices in the case of cone beam flat panel CT (see Section 2.5).

With regards to beam hardening, the bottleneck of the exported data is the built-in water correction (water correction mapping discussed in the end of Subsection 2.3.6). Since the raw measurement data is processed by a multi-step pipeline of modifications (e.g. over exposure correction, scatter correction, cosine weighting and more), only the manufacturer can provide projection data that is not corrupted by the water correction, which is mandatory for the PSR.

The experiments presented in this work were made possible by the supportive group of Siemens Healthcare AX (Erlangen, Germany), in particular by Dr. Manhart.

A common CT head scan suffers from the well known cupping artifact and at the edge of the dense skull of the bone which “spills over” into brain tissue. This artifact in particular makes clinical diagnosis in this area almost impossible in some cases, as critical information is not visible to the radiologist. A similar problem occurs at the edges of metal implants or screws. If the physicians need to make sure that tissue around such implants is not infected, they need to take other methods into consideration which eventually leads to the necessity of an invasive surgery. This is not just a matter of involuntary costs, but an ancillary increase of the risk of additional infections.

Of course, infections are only visible in a high tissue contrast CT image, which is typically disturbed by beam hardening induced artifacts. In other words, an important goal of beam hardening corrections is the sufficient reconstruction of voxels that are on the edges of materials with different absorption properties such as bone and tissue. An easier goal for beam hardening correction could be the approximation of the shapes of high attenuating objects, as the sufficient approximation of the shape can open up other beam hardening reduction methods to become viable [KKF08, CHF+19]. In some cases it might be important to make a proper segmentation of different materials possible. Suppose the target is to specify the shape of an object because the position is important for a 2D-3D registration. In this case, the possible overestimation of the physical size of the object due to the spill over artifact might cause serious trouble.

Since high absorbing materials can lead to photon starvation, the corresponding measurement is zero. Reconstruction algorithms tend to overestimate the resulting attenuation to be infinite. Such a behavior corrupts common reconstruction techniques and is a separate problem that needs to be adressed whenever photon starvation happens in CT imaging [Hsi98, XM12, WKC+14].

Time critical tasks such as stroke diagnosis are typically performed with CT systems. Here it is desired to have the information of the position of the stroke and if it is caused by a cerebral bleeding or by an obstruction that might prevent the blood supply to the brain. As both types cause the same symptoms but need to be treated in mutually exclusive ways, it is crucial to chose the correct treatment. A cerebral bleeding becomes fatal if the patient is diagnosed with an obstruction, hence will probably get injected anticoagulant which would increase the bleeding even more. In some cases, it is very subtle changes in the low contrast image that change the diagnostic value of an image. Thus the image quality needs to be as good as possible.

The fundamental problem in the image reconstruction domain is, the missing ground truth information. New methods are evaluated on hardware phantoms containing reliable information. As this has been done with the PSR already, the next step is to work on real data and evaluate the resulting diagnostic value. The gold standard in medical imaging are model observer studies [BYRM93, HP13, WNGH14]. Radiologists compare different methods on a set of different images and assign a rating or an assessment. This is typically a measure of the diagnostic value of the image.

Unfortunately it was not possible to perform such an experimental validation within this thesis. The proposal in this work is either checking for a hardware phantom attribute, such as the physical size or a quantitative analysis of the cupping artifact by polynomial fitting.

4.2.1 Pig head experiments

The very first proposal for an experimental testing of the PSR was designed to be a metal containing object was an object consisting of metal, tissue and bone. The natural choice was a head of a pig cadaver. A needle of steel was placed within the head and scanned with a Siemens Zeego C-arm device. Compared to bone and tissue, the steel needle has a very high X-ray attenuation. Hence, it was expected to cause clearly visible artifacts caused by beam hardening.

Imported data A short scan protocol with 4-binning of detector pixels was used. This corresponds to the measurement setup of geometry B.

PSR configuration The volume was initialized with zero, iterated 20 times with only one angle per subset. The subset level changed every 10 iterations after the 20 iterations procedure. The reconstruction was terminated after 100 over-all iterations. The density thresholds had to be set at $\rho_w = 1.0\,\mathrm{g/cm^3}$, $\rho_b = 1.45\,\mathrm{g/cm^3}$and $\rho_m = 1.9\,\mathrm{g/cm^3}$. The smoothness parameter was set to $\tau\prime = 0.25$.

Results The FDK reconstruction that is a built-in feature of the C-arm system suffers from strong shadow artifacts and a heavy spill over of the needle. The shape of the needle representation is hard to distinguish from the induced beam hardening artifacts. The MSR compensates for the shadow artifacts but suffers from heavy spill over of the needle voxels into the tissue. The PSR of the needle is more accurate compared to the FBP and the MSR. The estimated diameter of the needle is in very good agreement with the true value. The PSR tissue-to-bone-contrast is lower compared to the other two reconstructions.

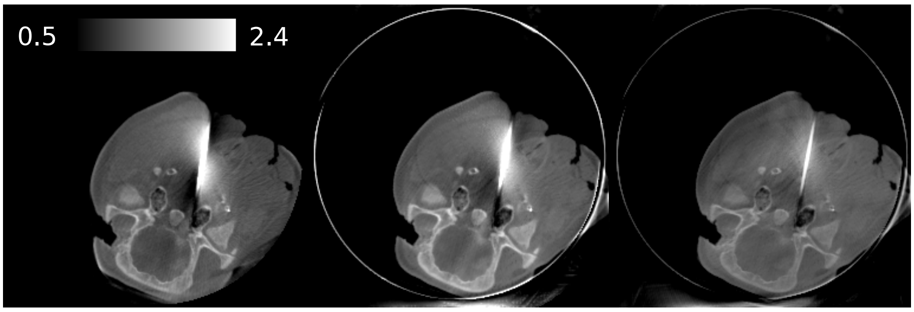


Figure 4.2.1: Steel needle pig head reconstructions: FBP vs. MSR vs. PSR, window level: $0.50 - 2.40\,\mathrm{g/cm^3}$

Remarks The missing shadow artifacts in the MSR are probably caused by the statistical weightings. So it seems that such artifacts occur due to the high absorption and the wrong weighting of such extinction values within the FDK framework. The enabling of the beam hardening compensation leads to a sharper representation of the metal object. The tendency to reconstruct the shape of high absorbing objects in a proper manner could lead to a potential application of the PSR as a prior step for other methods that rely on high quality segmentable volumes (e.g. a scatter estimation).

The MSR and PSR show a bright ring structure around the volume. This is caused by the collimator. Its edges become visible in the outer detector pixels. In other words, this collimator ring artifact indicates the region of interest of this scanning geometry. Because the outer volume is not sufficiently sampled, the FDK algorithm does not provide reliable voxel values. They are probably set to zero by the built-in reconstruction of the C-arm device.

This promising first result of the PSR to basically "get more out of the poor data" can only be an advantage. Even if the PSR is only utilized for other reconstruction methods as a segmentation or serves for other estimation purposes.

4.2.2 5 Euro-cent coin

To get a feeling about quality of measured projections, a 5 Euro-cent coin was placed in a water basin and scanned with a Siemens Artis Zeego C-arm system. This experiment is designed to evaluate the convergence behavior when the detector is under-saturated. The coin is placed in a way that the projections are done from both extreme view points.

Imported data The coin was scanned with the geometry B. The projection values show an inconsistency produced. As the extinction is defined to be linearly proportional to the physical size of the homogenous object (except

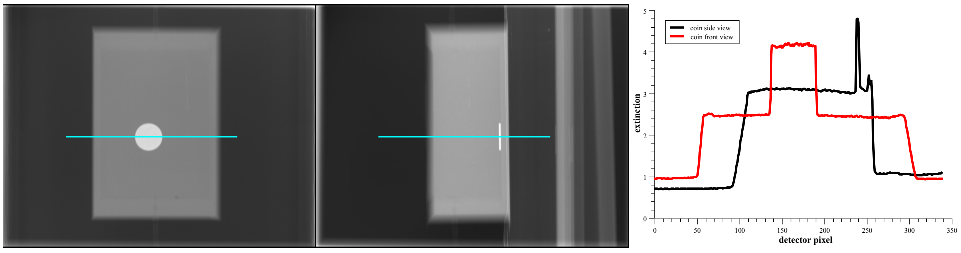


Figure 4.2.2: A quasi perpendicular projection views on the 5 Euro-cent coin and the cross section of the extinction values. The measured frontal and side view projections of the coin show an inconsistency in the extinction values from those two directions. The ratio of the extinction value is expected to be proportional to its geometric size (except for beam hardening effects), which is above a factor of the order 10. However, the measured ratio lies around $4.8/4.2 \approx 1.14$.

for beam hardening effects), the ratio of the measured grey values should correspond to the ratio of the diameter and its thickness. Since this behavior is not observed, one can argue that in this setting a detector saturation is established or scattered radiation lowers the extinction values.

PSR configuration The volume was initialized with zero, iterated 20 times with only one angle per subset. The subset level changed every 10 iterations after the 20 iterations procedure. The reconstruction was terminated after 100 over-all iterations. The density thresholds were set to $\rho_w = 1.0\,\mathrm{g/cm}^3$, $\rho_m = 3.0\,\mathrm{g/cm}^3$ and the smoothness parameter was set to $\tau\prime = 0.25$.

Results The MSR (not to confuse with the FDK) shows beam hardening induced artifacts such as a blurred shape of the coin. The dark streak artifacts that are perpendicular to these artifacts are caused by scattered radiation, detector saturation and beam hardening. The PSR can reduce some of these artifacts as it is clearly visible in Figure 4.2.3. The PSR estimation of coin diameter and coin thickness are in good agreement to the real values. Shadow artifacts remain but are heavily reduced compared to the MSR.

Remarks The MSR shows typical beam hardening induced artifacts. The cross section (Figure 4.2.4) visualizes the proper representation of the coin as a result of the PSR. The remaining artifacts of the PSR might be caused by scattering or detector saturation. As Figure 4.2.2 shows, the measured extinctions are corrupted by inconsistencies. The observed artifact is similar to the simulation results of the scatter experiment, as the utilized saturation model leads to brighter artifacts.

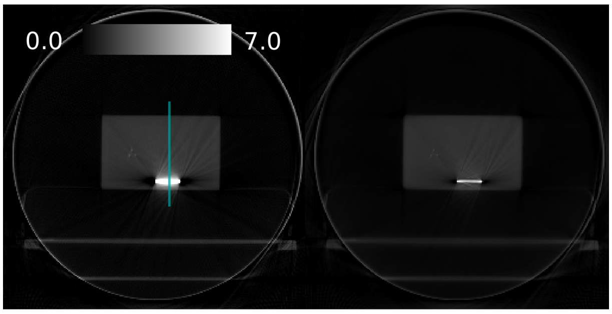


Figure 4.2.3: MSR vs. PSR, window level: 0.0 – 7.0 g/cm^3

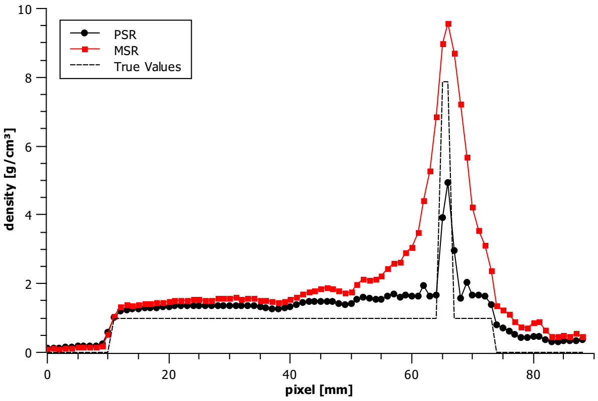


Figure 4.2.4: MSR vs. PSR vs. cross section window 0.0 - 7.0

The projection data from the C-arm system is scatter corrected by the firmware. Such scatter correction is calibrated for an object consisting of water. Hence, it will probably not work well on metal objects and their high extinction values. Scatter can explain the low extinction values of the coin in the side view, but it should reduce the values around the edges as well.

Because the geometry of the coin is known it is concluded that the measurements taken in the corresponding angles should be somehow neglected as they contain no useful information.

Nevertheless, the shape of the coin can be reconstructed in good agreement to the real values.

4.2.3 Clinical data

On a first trial in 2016, a single data set was converted as the PSR requires the absence of water correction and cosine weighting. The PSRrsi was showing promising improvements to the radiologist. Later, 35 more data sets where provided and converted accordingly.

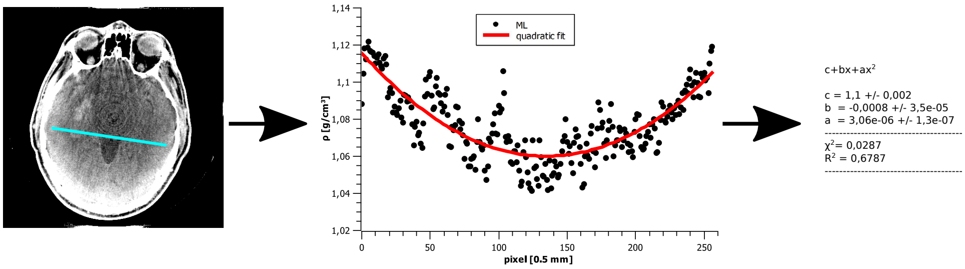


Figure 4.2.5: Schematic depiction of the procedure to evaluate cupping reduction of the PSR on the 36 data sets.

Methods To characterize the ability of reducing beam hardening induced cupping within the brain tissue, a quadratic function was fitted into a cross section of the reconstruction. This procedure is outlined in Figure 4.2.5. Here, the coefficient a measures the cupping. The higher the number, the bigger the cupping artifact within the brain area. Since the estimation depends on the line that was chosen manually for each reconstruction, it is assured that the same voxels were acquired to compare the curvature of the grey values. The PSR is expected to deliver smaller coefficients a then the ML reconstruction of the same data set, as it is not corrected for beam hardening. Thus, the quotient of those two coefficients should yield a number ${}^{a_{\mathrm{PSRov}}}/_{a_{\mathrm{ML}}} < 1$.

For comparison, clinical data was given to a maximum likelihood (ML) reconstruction method that does not account for beam hardening [FPB$^+$14] and the PSR

$$\rho_j^{n+1}\left(\vec{\rho}^{\,n}\right) = \left[\rho_j^n - \Lambda \left\{ \frac{M\hat{N}_j + \beta \left.\frac{\partial S}{\partial \rho_j}\right|_{\vec{\rho}=\vec{\rho}^{\,n}}}{d_j^1 + \beta \left.\frac{\partial^2 S}{\partial \rho_j^2}\right|_{\vec{\rho}=\vec{\rho}^{\,n}}} \right\}\right]_+ . \tag{4.2.1}$$

Here we add a relaxation parameter Λ to enable a workaround that delivers sufficient image quality. The first experiments only delivered appropriate image quality when regularized by early stopping after the very first iteration of the update step. At this time the PSR was initialized with the corresponding ML reconstruction. Earlier studies [BBR19] suggested that the patient table that is present in the head scan projections corrupts the PSR results while comparable iterative ML [FPB$^+$14] reconstructions of the same data sets are not affected in such a manner. The reason for such a weakness of the PSR lies in the specific choice of the weighting d_j^1. The iterative ML reconstruction does not reducing beam hardening artifacts by design, but were given state-of-the-art bi-material beam hardening corrected projections [AFB$^+$18]. To enable a proper comparison of the PSR with a statistical but monochromatic iterative approach, the projection data sets the ML were given are the uncorrected[2] projections and the bi-material corrected projections.

The behavior of the PSR was studied in simulations and two solutions to reduce the table truncation artifact are proposed.

At first a specific early stopping routine was found to be working sufficiently. Here the relaxation parameter was set to $\Lambda = 0.05$. Since this PSR was initialized with the ML, it reduced the beam hardening artifacts when considering every projection angle once and performing the update step right after each angle. Hence, this method is relies on a proper initialization.

The second approach uses the feature of the outer volume (OV) discussed in Subsection 4.1.5. Thus, this approach is a two step algorithm. It reconstructs a native PSR with very few iterations on a low resolution and a high regularization parameter to smooth streak artifacts that are caused by the low angular sampling. This native PSR also used only the half of the measured projections as well.

Such a native PSR is used to initialize the OV buffer of the respective feature (PSRov) of the forward-projection step described in Section 3.4.2. Simulations of the measurement setup were done to support the outer volume strategy (see Subsection 4.1.5).

Imported data The collimator of such real data was causing bright detector pixels at the edges. This forced the cutting out of the outer 30 pixels. Going down from 1240 × 960 detector pixels to 1180 × 900 detector pixels and changing the projection matrices accordingly to this new geometry.

The projection matrices of all data sets were corrected by a state-of-the-art motion compensation technique [FR15]. The data was acquired between 2013 and 2015.

PSR configuration The PSR was initialized with $\rho_j^0 = 0 \ \forall j$, the US of the above equation was performed 40 times with $\Lambda = 1$. This is called the "native PSR". In Figure 4.2.10 the truncation artifact can be observed as the native PSR is showing the characteristic shadows explained in Subsection 4.1.5.

To eliminate these artifacts within the framework of the PSR, it was accounted for the truncation as mentioned. The OV, that was twice of the size of the original reconstruction volume, included the representation of the table. Figure 4.1.17 visualizes the representation of the table. Such approximation was utilized to reduce the inconsistencies between the forward-projection of the current iterate and the measured projections.

Firstly, the proposed method undersamples the OV (that is initialized with zeroes) by performing the US 8 times that is using larger but less voxels.

---

[2] This means that these projections are not water corrected.

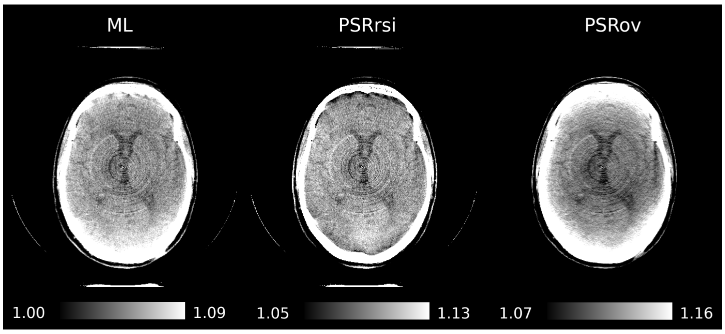


Figure 4.2.6: Data set 12: The PSRov does not improve the image quality but in fact introduced even more cupping. The PSRrsi shows a clear improvement in image quality.

The parameters used were $128 \times 128 \times 176$ voxels with a spatial resolution of $4.0 \times 4.0 \times 1.0\,\mathrm{mm}^3$ instead of $464 \times 464 \times 176$ voxels with a spatial resolution of $0.5 \times 0.5 \times 1.0\,\mathrm{mm}^3$ and 248 views of 1240×960 detector pixels instead of 496 views while keeping the detector size the same. We increase the regularization parameter β by a factor of 100 to smoothen the image.

Secondly, the forward-projector was given this undersampled (but physically larger OV) in addition to the current iterate ρ_j^n. Note that the OV is not affected by the US. This methodology and its resulting reconstruction is called PSRov.

The downsampleing of the OV leads to a smoother approximation but also scales down the reconstruction time of the OV as well as the PSRov. Due to the interpolation feature of GPU devices and the constant step length within the downsampled OV, the forward-projection of the additional OV within the PSRov-US increases the computational load of the forward-projection just moderately. Here the voxel edge length of the OV is enlarged by a factor of 8 compared to the standard setting.

Additionally, the workaround of regularization by early stopping was performed. Initializing the PSR with the native ML and performing only one iteration of equation 4.2.1 the US with $\Lambda = 0.05$ including all measured projection once. This approach is called “relaxed single iteration” (PSRrsi).

Results On one data set (number 12 in table 4.3), the PSRov did not reduce the cupping artifact, yielding $a_{\mathrm{PSRov}}/a_{\mathrm{ML}} \approx 2$. This data set is depicted in Figure 4.2.6. The PSRrsi on the other hand yields a clear reduction of beam hardening in the very same slice.

Figure 4.2.7 shows certain artifacts in the PSRov as well. Such an observation is made in other data sets as well. Figure 4.2.8 and 4.2.9 yield distinct artifacts in the PSRov and the native PSR.

#	a_{ML}	a_{PSRov}	$a_{\mathrm{PSRov}}/a_{\mathrm{ML}}$
1	3,59E-06	2,11E-06	0,59
2	4,60E-06	2,67E-06	0,58
3	2,66E-06	1,03E-06	0,39
4	3,22E-06	1,78E-06	0,55
5	2,97E-06	1,46E-06	0,49
6	3,29E-06	1,52E-06	0,46
7	2,92E-06	5,87E-07	0,20
8	1,70E-06	7,33E-07	0,43
9	2,54E-06	1,69E-06	0,66
10	4,36E-06	2,47E-06	0,57
11	3,85E-06	2,40E-06	0,62
12	1,52E-06	3,00E-06	1,97
13	3,90E-06	2,46E-06	0,63
14	2,40E-06	1,22E-06	0,51
15	2,10E-06	1,01E-06	0,48
16	3,84E-06	2,71E-06	0,71
17	3,86E-06	2,52E-06	0,65
18	2,19E-06	9,84E-07	0,45
19	1,99E-06	8,63E-07	0,43
20	2,80E-06	1,43E-06	0,51
21	4,37E-06	3,31E-06	0,76
22	3,13E-06	1,55E-06	0,50
23	1,64E-06	3,94E-07	0,24
24	4,32E-06	2,76E-06	0,64
25	2,18E-06	7,10E-07	0,33
26	3,77E-06	1,93E-06	0,51
27	3,67E-06	2,27E-06	0,62
28	4,03E-06	2,49E-06	0,62
29	2,90E-06	2,04E-06	0,70
30	4,29E-06	2,52E-06	0,59
31	4,36E-06	3,32E-06	0,76
32	2,64E-06	1,56E-06	0,59
33	2,84E-06	2,10E-06	0,74
34	3,06E-06	1,77E-06	0,58
35	3,12E-06	1,98E-06	0,63
36	3,10E-06	1,57E-06	0,51

Table 4.3: The resulting fit parameters of the quadratic fits of the 36 data sets. The average quotient is estimated as $\overline{a_{\mathrm{PSRov}}/a_{\mathrm{ML}}} = 0.589$ with a standard deviation of 0.27.

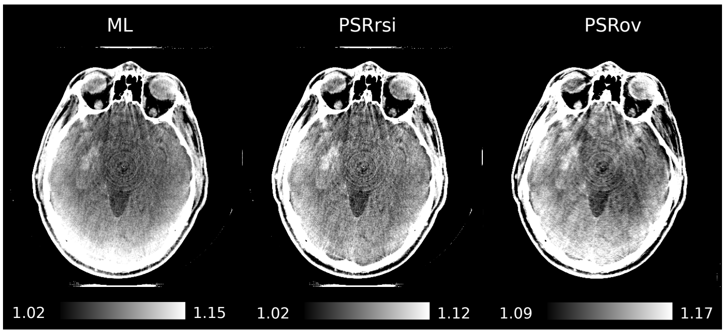


Figure 4.2.7: Data set 36: The PSRov introduces artifacts over the course of 40 iterations.

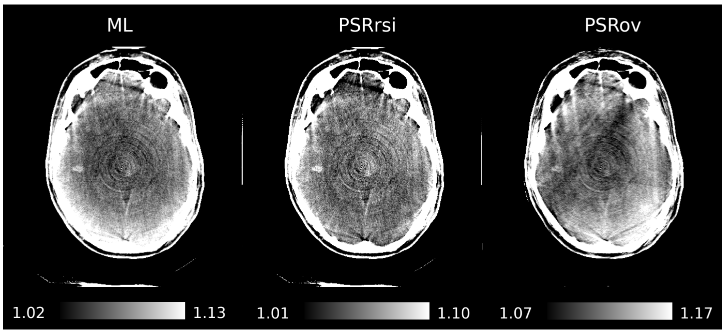


Figure 4.2.8: Data set 7: The PSRov introduces artifacts over the course of 40 iterations.

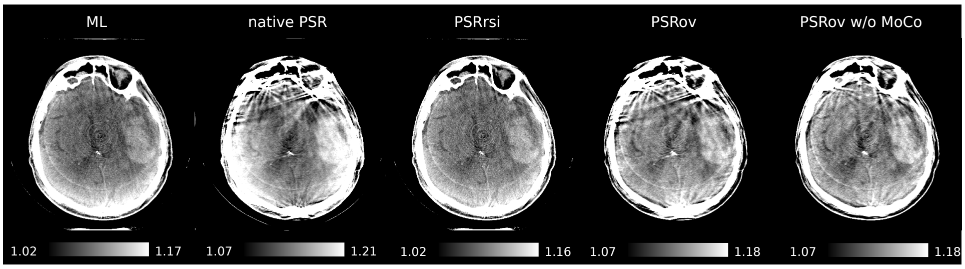


Figure 4.2.9: Data set 2: The PSR suffers from strong motion artifacts introduced by the applied motion compensation. Using the uncorrected projection matrices results in less motion artifacts. The ML is less sensitive to such motion. The PSRrsi is only slightly improves the image quality.

The native ML is distorted by beam hardening induced cupping as well as the spill over effect which blurs the dense bone of the skull into tissue. The bi-material corrected ML cannot fully reduce cupping but clarifies the shape of the bone and reduces its spilling over. The table truncation can be observed in the native PSR. The two proposed workarounds to that problem (PSRrsi and PSRov) overcome these erroneous characteristics, yielding a similar beam hardening reduction as the bi-material corrected ML.

Remarks Ring artifacts caused by defect detector pixels or insufficient homogeneity of detector dynamics are typical for C-arm imaging. These artifacts are reduced in a post processing step [ZSRB05]. One approach is to transform the x-y-slice into cylindrical coordinates and median filter over the radius component. Back transforming results in an almost ring artifact free image. However, the point $r = 0$ is a singularity of cylindrical coordinates. Thus, a small area with $r < \varepsilon$ is not sufficiently sampled by this approach.

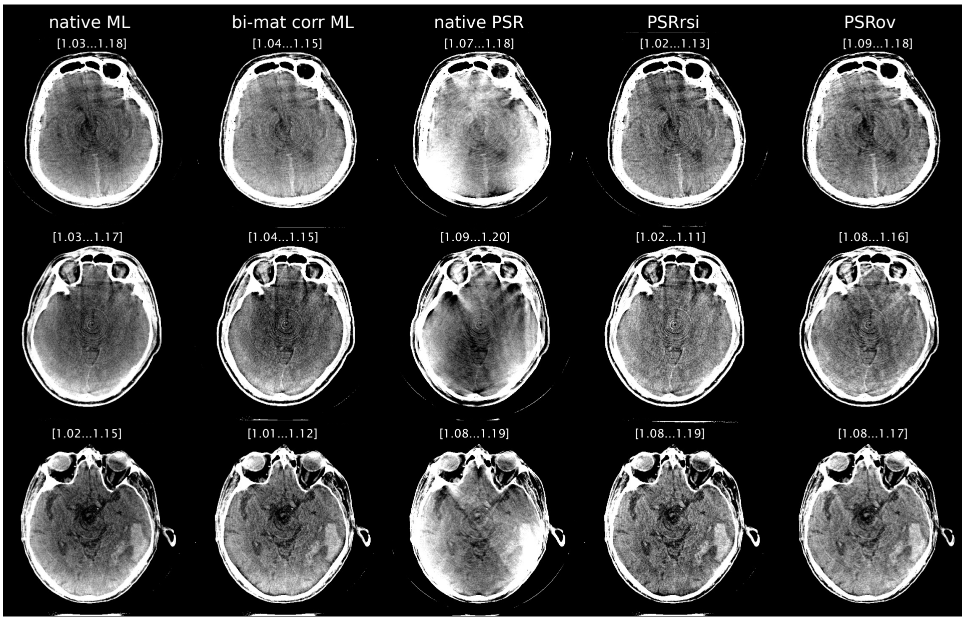

native ML
bi-mat corr ML
native PSR
PSRrsi
PSRov
[1.03...1.18]
[1.04...1.15]
[1.07...1.18]
[1.02...1.13]
[1.09...1.18]
[1.03...1.17]
[1.04...1.15]
[1.09...1.20]
[1.02...1.11]
[1.08...1.16]
[1.02...1.15]
[1.01...1.12]
[1.08...1.19]
[1.08...1.19]
[1.08...1.17]

Figure 4.2.10: The native ML vs. bi-material corrected ML vs. native PSR vs. PSRrsi vs. PSRov are depicted. The used window size is relatable to 100HU.

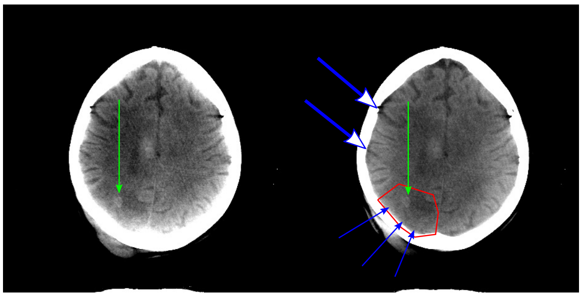


Figure 4.2.11: The bi-material corrected ML (left-hand side) and PSR (right-hand side), circumscribed mild brain edema is depicted, but barely visible. The blue arrows point out at the absence of sulci. Typically they should be visible as the white arrow suggests. The window level corresponds to 100HU. Here a ring artifact correction has been performed.

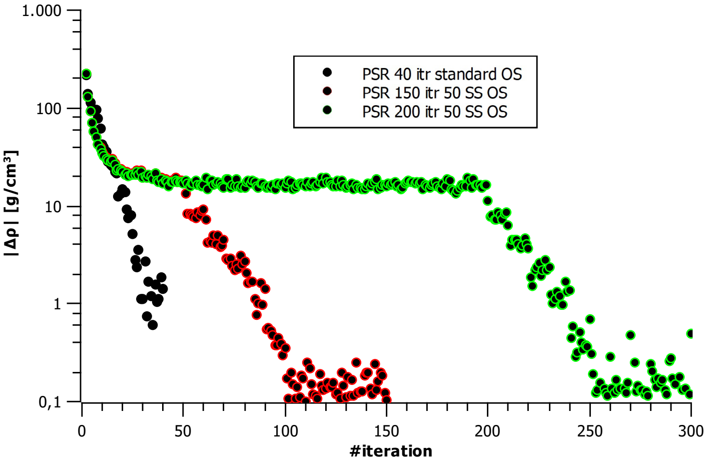


Figure 4.2.12: Data set 28: A typical convergence pattern of the update step when different numbers of iterations are applied.

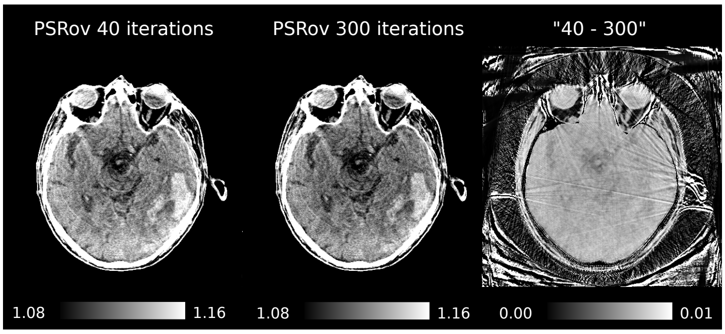


Figure 4.2.13: Dataset 28: Difference between 40 iterations of the default setting and the 300 iterations depicted in Figure 4.2.12.

This correction was only applied to the reconstruction depicted in Figure 4.2.11.

Clinical CT scans include additional objects as a patient table. Such objects are designed to cause the lowest possible extinction values to projections and typical reconstruction techniques tend to increase voxel values around the reconstruction object in order to minimize the inconsistency between the forward-projection while iterating but leaving the important region of interest almost unaffected. This behavior can be observed in the first column of Figure 4.2.10. The native PSR creates increased pixel values at the bottom of the images, a thin "ring" structure that is indicating the region of interest. Those thin "ring" artifacts around the head and the increased pixel values at the bottom vanish in the PSRov as intended.

Despite that for every data set (except number 12) the cupping ratio $a_{\text{PSRov}}/a_{\text{ML}}$was below 1.0, the image quality did not necessaarily improve with respect to the overall appearance. Three examples are shown in Figure 4.2.7, 4.2.8 and 4.2.9. In the latter, strong artifacts are introduced by the PSR. Those artifacts "look like" motion artifacts. The ML is working with the very same data and motion-compensated projection matrices. Using the original projection matrices clarifies that the applied motion compensation can sabotage the PSR whereas the ML is not affected in the same way. Figure 4.2.12 and 4.2.13 depict the development of the US in three different ordered subset settings of another data set. In the default setting, 8 iterations were performed with the single angle subset followed by 4 iterations that double the number of angles with each recursion. The difference image shows that subtle changes are introduced with higher numbers of iterations. Those changes are manifestations of the small inconsistencies within the projections, as the absolute value of the US is not changing over the course of the additional iterations.

The reason for such inconsistent convergence is not fully understood. C-arm systems typically vary the time-current product, hence the photon flux changes depending on the projection angle. Even the tube voltage is not guaranteed to stay constant during the scan. Thus, the X-ray spectrum $X(E) \longrightarrow X_i(E)$ might be object to change as well. Enabling the feature to differently weight projection angles with respect of the applied time-current product did not affect the reconstruction significantly.

Furthermore, calibration related issues of the C-arm system cannot be excluded as the time periods between new calibrations were irregular.

Nevertheless, there are cases where the PSRov and PSRrsi improve image quality and converge accordingly to the results of the simulations in Subsection 4.1.18. An example case of enhanced diagnostic value can be found in Figure 4.2.11. This example emphasizes the problems caused by the spill over as well as the cupping artifact. As the missing sulci cannot be identified without the beam hardening correction.

The reconstruction time does not increase dramatically by the additional forward-projecting in each iteration step. The results are calculated on 1 to 4 AMD FirePro S10000 graphics cards with two 'Tahiti' chips each (bought 2013). A single nVidia GeForce RTX 2080 Ti outperforms these by a factor of 4.

5 Discussion

The underlying theory to reconstruct an image slice with a given set of projections was established by Johann Radon in 1917 [Rad17]. Almost 50 years ago, CT devices started to enhance clinical diagnostics [Kal06]. Since then, the research community has been working to increase the diagnostic value of the CT images by making them more accurate while reducing patient dose. As long as the fundamental measuring process is not altered and a new generation of CT devices, that do not integrate over the different energy bins are established, the non-linearity of the projection process will corrupt the CT images.

To reduce dose while keeping image quality, one needs to apply statistical models.

To reduce beam hardening or photon scattering related artifacts, one needs to apply physical models.

This is the core of the PSR, which was introduced in 2002 by Elbakri and Fessler [EF02, EF03c]. However, this method has not been developed further, nor has it been tested on clinical data. For the first time the PSR was used on real cone-beam CT data yielding promising results.

Back in the early 2000s statistical and iterative reconstruction methods where not feasible due to the enormous amount of memory and calculation requirements. Since 1965, calculation poer grew exponentially according to MOORE'S law [Moo65]. The viability of the PSR has been shown in this thesis.

- Without new measurement techniques or measurement equipment the PSR can enhance image quality on C-arm devices. Figure 4.2.10 compares the PSR with an alternative statistical reconstruction technique.
- The general tendency of the PSR to reduce the beam hardening induced cupping artifact on 35 data sets is shown in Table 4.3.
- Figure 4.2.4 emphasizes the opportunity to enable better segmentation of high absorbing objects. As the beam hardening induced spill over artifact is eliminated. A technical example about the increased diagnostic value is represented in Figure 4.2.11.

The results presented in Section 4.2 are good examples for the superiority of iterative reconstruction techniques that become technically feasible due to

Moore's law. The increased requirements of the PSR can be satisfied with better hardware that is available even for private use such as modern graphics cards, random access memory and CPU power. The prior knowledge is typically available in a practical setting as in a hospital environment.

One bottleneck that appears in daily practice is the need for projection data that is not manipulated by a water correction. This correction is mandatory for the standard reconstruction methods that do not accurately factor in the polychromatic nature of the X-rays.

The outer volume (OV) feature shows consistent results with the simulations. It is essential to account for the outer volume on the clinical head scan data as Figure 4.2.10 suggests. This observation raises the question about volume-of-interest (VOI) imaging to reduce dose by reducing the field of view only to the region that is of interest for the physician. The sensitivity of the PSR with regard to even the smallest truncation points out that VOI imaging is probably not applicable to the PSR in its current implementation.

The varying image quality of the 36 datasets suggests that patient motion is a bottleneck of the PSR, as it seems to be more sensitive then the ML reconstruction. Figure 4.2.9 (data set 12) emphasizes this sensitivity.

Most FP-CBCT systems adapt the X-ray intensity during the scan to reduce radiation exposure and utilize the limited dynamic range of the detector. Thus, the approximation about $I(E)$ to be independent of the projection angle is violated. This might corrupt the statistical assumptions as well as the polyenergetic specifications. The observed stability of the PSR under uncertainties about the X-ray spectrum holds only when the induced change is constant for each view (see Figure 4.1.3). These observations support similar findings of Wu et al. [WYMF14].

The displacement model (see Section 3.3) requires a density value in order to calculate the thresholds that determine the material of a voxel. These thresholds need to be sufficiently distinguishable from each other. The PSR would not perform well when materials overlap in density, e.g. contrast agent (CA). Here, a proper segmentation is needed. In perfusion imaging a mask scan is performed to subtract the voxels that do not change in time. After such a subtraction a distinction could be feasible.

The PSR in relation to other reconstruction techniques

Common beam hardening reduction approaches are based either on preprocessing of the X-ray projections, such as a water correction [NL79, JS78a], or they try to manipulate the reconstruction as a post-processing step [BC76, HMDJ00, MMS90]. The former are working best when the scanned object consists only of one material. The latter are heuristic image processing applications by design and might be too unreliable for clinical applications.

Huang et al. recently provided an insightful article about the limitations of deep learning based approaches [HWB$^+$18].

Another approach is based on the idea of splitting the main contributing physical effects by acquiring two separate measured sets of X-ray projections. The most prominent case is the dual energy CT [AM76, FESC02]. Proper filtering and doubling the recording could have the same effect. Two energy-dependent measurements of the same object with respect to the identical scanning geometry. The main drawback of those kind of approaches is that they are reliant on new hardware and will also most likely increase the patient's X-ray dose.

Recent developments in the field of energy resolved photon-counting detectors want to overcome the problem of beam hardening [GL11]. As we can see in Figure 4.1.3, the PSR can deliver viable reconstructions with three energy bins. However, to fully neglect the beam hardening by this detector technology one needs a higher energy resolution. This would certainly be the case when reaching a high resolution because the non-linearity is only introduced to the CT problem by the integrating detector in Equation 2.3.2. Additionally, the counting rate seems to be a bottleneck to real applications as well.

With respect to the beam hardening induced cupping effect, Kachelrieß et al. propose an empirical post-processing correction. Their approach does not need any prior knowledge such as X-ray spectrum, size or position of the object [KSK06].

Compared to the PSR, different iterative techniques have been proposed to tackle beam hardening. They can reduce beam hardening but rely on a proper segmentation [YWB$^+$00, LFN09]. Xu et al. [XUK$^+$17] propose to overcome the segmenting problem by a 2D-3D registration of the metallic objects when scanned in air or given a CAD model. They also do not need the prior knowledge about the absorption properties but include unknown polynomial absorption into their framework. These assumptions might lead to limitations regarding the workflow or feasibility in a clinical setting. Similar to the OV feature of the PSR, their approach is supposed to work on Volume-of-Interest imaging.

Humphries et al. [HWF17] introduced the same polychromatic forward-projection model[1] to a classic weighted least squares framework which also features TV-superiorization. It is used on 2D simulations within a limited angle setting. Due to the strong similarity of needed implementation work it would be very interesting to test the available clinical data sets on this algorithm.

Lately, several beam hardening correction methods evolved around consistency conditions in the projection domain [AFB$^+$18, THC$^+$18].

---

[1] The main difference is that they use three piecewise linear functions as sigmoid functions instead of the polynoms of the displacement model.

The connecting factor of the PSR

Scattered radiation

As argued in Subsection 4.1.3, scattered radiation causes artifacts which can be reduced drastically when given a proper scatter estimation as prior knowledge. If the region of interest lies directly at the edge of a high absorbing object the PSR cannot sufficiently enhance the information gain of the image. This can be observed in Figure 4.2.3. It shows the remaining shadow artifacts caused by the 5 Euro-cent coin. The bottleneck of state-of-the-art scatter estimation methods is a proper segmentation. We can conclude by observation, that such a segmentation is made feasible by the PSR itself.

This opens up the opportunity for a two step approach to the solution of the scatter problem. A first PSR of the scan is expected to deliver an accurate estimation of the physical size of the object[2]. A third party program could give a proper scatter estimation [MPND18] and feed this information into a second PSR which is initialized with this estimation and the first PSR. To speed up the second PSR, one can use the outer volume feature to ensure that only the area around the object is changed by the update step.

Alternatively it would be of interest to test the assumptions about the impact of scattered radiation. Bier et al. [BBM$^+$17] developed a hardware-driven approach to eliminate scatter by modulating primary radiation. They evaluated their method on the same kind of C-arm system. Performing a similar experiment as the 5 cent coin measurements should be easy. Their results could be incorporated very easily due to the same geometry.

Dual energy CT

Dual energy CT captures two separate X-ray projections of different angles at the same time[3]. They are intended to enable the segmentation of two different materials within the reconstruction. This is possible because different materials will be reconstructed with different attenuation values depending on the effective energy of the respective spectrum. These measurements can split the attenuation into its two main contributors $\mu = \mu_{\text{Compton}} + \mu_{\text{Einstein}}$, which enables their modelling feasible.

For example, dual energy CT can be used to reliably measure either bone density or plaque. Within this domain, the PSR could be used to effectively half the dose while keeping the image quality constant. The idea would be to leave both source-detector systems only a 90° trajectory and combine the measurements. As the PSR will reconstruct the energy independent density of the object, it is sufficient to combine both scans to a 180° half circle.

---

[2] The accuracy is limited by the voxel size.

[3] This also enables a better time resolution when both sources work with the same tube voltage. A good application is perfusion imaging with CT. There, the dynamic flow of contrast agent is reconstructed and the CT problem becomes a 4D reconstruction problem.

This idea is similar to the recent paper of Mechlem, Ehn et al. where they use the information of all measured energy levels of a spectral CT in one update step of the reconstruction [MES$^+$18].

Non-destructive testing

X-ray based methods are of interest for manufacturers of a variety of objects such as chassis, tools, screws, airplanes and so on. These measurements need to be as accurate as possible. Usually the manufacturers need to know the exact number of blisters to ensure a certain stability, reliable quantitive estimations of density and the cleanliness of weldseams.

As these measurements are corrupted by beam hardening and scattering as well, the PSR could lead to an increase of the testing accuracy. Note that in this type of applications, calculation time is not the most important factor, which the PSR certainly needs more then alternative methods.

This topic is of special interest because of the slightly shifted settings with regards to materials and required accuracy. The projection based beam hardening approaches might be a perfect fit as the number of different high absorbing materials is known and prior knowledge such as CAD models of the objects are a given. Those conditions will also make scatter estimations more accurate. Since the “diagnostic value” of such images is not potentially lethal to patients, it might also be a good training ground for new machine learning based approaches [CHF$^+$19].

The PSR can help to reduce dose. Even though the X-ray dose is not harmful to the objects that are scanned, it is still of interest to reduce the dose. To reduce dose certainly helps to reduce costs for the manufacturer, as the X-ray sources have a larger lifespan.

Future work

A performance evaluation study of the PSR in the clinical setting is of interest. Model observers [BYRM93] would need to evaluate the image quality compared to other modern methods that deal with beam hardening.

As the detector performance of CT devices is higher compared to C-arm devices, it is of interest to evaluate the full potential of the PSR within the CT environment. Here, a major change of the implementation will be necessary due to the changing detector geometry from the flat panel to the curved panel.

As high absorbing materials can lead to photon starvation (described in Section 4.1.4), statistical assumptions become violated. This problem is of major concern for all reconstruction techniques. As the PSR is a flexible framework, it is of interest to adapt certain ideas such as MAR techniques [GMV+17] to reduce resulting artifacts.

To tackle detector saturation effects in the presence of metal objects as observed in Figure 4.2.2, it could be required to model this aspect inside the forward-projection model. Repeating the mathematical derivation of the update step as a result of the maximization problem statement of Equation 2.4.24 is a non-trivial task. Certain approximations will be needed and an accurate saturation model needs to be evaluated first. Alternatively, a similar approach to the classic of Joseph and Spital look-up table [JS78b] or the modern adaptation of Cao et al. [CHF+19] could solve the problem. As the path lengths through the objects are known, it could be a viable approach to replace the measured projections with corrected expectation values, as the exact pixels that are affected by the metal object are known after an initial PSR.

As stated multiple times in this work, segmentation is a problem of interest in many fields within the CT community. The mentioned topics such as scatter estimation/reduction or other beam hardening corrections are desperately in need of a proper segmentation. The PSR might work as an additional tool of the toolkits that are available to improve solutions that are already evaluated.

Conclusion

Polychromatic statistical reconstruction can enhance clinical diagnosis by providing higher quality images while using the very same CT devices. In other words, CT scanners could enhance their image quality by investing in more GPU devices which are cheap compared to new detector technology.

To tackle an emerging result of of Eroom's law [SBBW12] in CT imaging, it should be of serious consideration to leave the FDK based imaging and head towards modern approaches that became viable due to MOORE'S law [Moo65]. This work gives a detailed insight into the method and depicts its value with respect to CT image enhancement.

For the first time and more then a decade after its original publication, the PSR was tested on real cone-beam CT data . The PSR needs neither a prior segmentation nor new CT devices. It can work in a real hospital setting and reduce beam hardening related artifacts. The PSR has substantial potential for adaptation and offers applications in diverse domains of CT reconstuction.

Declaration of Honor

„I hereby declare that I produced this thesis without prohibited external assistance and that none other than the listed references and tools have been used. I did not make use of any commercial consultant concerning graduation. A third party did not receive any nonmonetary perquisites neither directly nor indirectly for activities which are connected with the contents of the presented thesis. All sources of information are clearly marked, including my own publications. In particular I have not consciously:

- Fabricated data or rejected undesired results,
- Misused statistical methods with the aim of drawing other conclusions than those warranted by the available data,
- Plagiarized data or publications,
- Presented the results of other researchers in a distorted way.

I do know that violations of copyright may lead to injunction and damage claims of the author and also to prosecution by the law enforcement authorities. I hereby agree that the thesis may need to be reviewed with an electronic data processing for plagiarism. This work has not yet been submitted as a doctoral thesis in the same or a similar form in Germany or in any other country. It has not yet been published as a whole.”

Magdeburg, July 2nd 2021
Richard Bismark

List of Publications

- Sebastian Bannasch, Robert Frysch, Richard Bismark, Gerald Warnecke, Georg Rose: An Optimal Relaxation of the Algebraic Reconstruction Technique for CT Imaging, Fully3D 2015, S. 622-625

- Richard Bismark, Robert Frysch, Georg Rose: Reduktion von Strahlaufhärtungsartefakten beim C-Arm CT mittels statistisch-iterativer polychromatischer Rekonstruktion, 50. Jahrestagung der DGNR, 2015

- Richard Bismark, Robert Frysch, Georg Rose: Reduction of beam hardening artifacts on real C-Arm CT data using statistical polyenergetic image reconstruction, In: 4th International Conference on Image Formation in X-Ray Computed Tomography, Bamberg Germany, 2016

- Shiras Abdurahman, Robert Frysch, Richard Bismark, Michael Friebe, Georg Rose: Calibration free Beam Hardening Correction using Grangeat-based Consistency Measure, IEEE MIC 2016

- Shiras Abdurahman, Robert Frysch, Richard Bismark, Oliver Beuing, Michael Friebe, Georg Rose: Strahlaufhärtungskorrektur mithilfe von Grangeat-Konsistenzbedingungen für Kegelstrahl-CT, NeuroRad 2017

- Robert Frysch, Richard Bismark, Andreas Maier, Georg Rose: Ray-Density Weighted Algebraic Reconstruction for Volume-of-Interest CT, Fully3D 2017

- Shiras Abdurahman, Robert Frysch, Richard Bismark, Oliver Beuing, Georg Rose: Beam harening correction using Grangeat-based consistency measure, In: IGIC 2017: 3. Image-Guided Interventions Conference, 6. - 7. November 2017, Abstract Book, Magdeburg, S. 26

- Richard Bismark, Oliver Beuing, Georg Rose: Untersuchung von Artefaktverursachenden Einflüssen auf die polychromatische statistische Rekonstruktion, In: IGIC 2017: 3. Image-Guided Interventions Conference, 6. - 7. November 2017, Abstract Book, Magdeburg, S. 10-11

- Shiras Abdurahman, Robert Frysch, Richard Bismark, Steffen Melnik, Oliver Beuing and Georg Rose: Beam hardening correction using cone beam consistency conditions, IEEE Trans Med Imaging. 2018 May 24, doi: 10.1109/TMI.2018.2840343

- Shiras Abdurahman, Robert Frysch, Richard Bismark, Oliver Beuing, Georg Rose: A complete scheme of empirical beam hardening correction using Grangeat consistency condition, IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC), Sydney, Australia, 2018, pp. 1-5

- Tim Pfeiffer, Robert Frysch, Richard N. K. Bismark, Georg Rose: CTL: Modular Open-source C++-Library for CT-Simulations, Fully3D 2019
- Richard N. K. Bismark, Oliver Beuing, Georg Rose: Truncation artifacts caused by the patient table in polyenergetic statistical reconstruction on real C-arm CT data, Fully3D 2019
- Shiras Abdurahman, Robert Frysch, Tim Pfeiffer, Richard Bismark, Oliver Beuing, Georg Rose: CTRS: A 3D reconstruction software for cone beam and multi-slice CT, IGIC 2019, Abstract Book, Mannheim, S. 34
- Richard N. K. Bismark, Georg Rose: Reduktion des Patiententischartefakts in der polychromatisch statistischen Rekonstruktion durch Außenraumunterabtastung, IGIC 2019, Abstract Book, Mannheim, S. 28
- Richard N.K. Bismark, Robert Frysch, Shiras Abdurahman, Oliver Beuing, Manuel Blessing, Georg Rose: Reduction of beam hardening artifacts on real C-arm CT data using polychromatic statistical image reconstruction, Zeitschrift für Medizinische Physik, 2019
- Richard N.K. Bismark, Oliver Beuing, Georg Rose: Overcoming truncation artifacts caused by the patient table in polyenergetic statistical reconstruction on clinical C-arm CT data, In: 6th International Conference on Image Formation in X-Ray Computed Tomography, Regensburg, Germany, 2020

Bibliography

[ABW$^+$15] André Aichert, Martin Berger, Jian Wang, Nicole Maass, Arnd Doerfler, Joachim Hornegger, and Andreas K. Maier, *Epipolar consistency in transmission imaging*, IEEE Trans. Med. Imaging **34** (2015), no. 11, 2205–2219.

[ABYZ67] D.W. Aitken, B.L. Beron, G. Yenicay, and H.R. Zulliger, *The fluorescent response of NaI(Tl), CsI(Tl), CsI(Na) and CaF2(Eu) to X-Rays and low energy gamma rays*, Nuclear Science, IEEE Transactions on **14** (1967), no. 1, 468–477.

[AFB$^+$16] Shiras Abdurahman, Robert Frysch, Richard Bismark, Michael Friebe, and Georg Rose, *Calibration free beam hardening correction using grangeat-based consistency measure*, IEEE NSS/MIC 2016, 2016.

[AFB$^+$18] Shiras Abdurahman, Robert Frysch, Richard Bismark, Steffen Melnik, Oliver Beuing, and Georg Rose, *Beam hardening correction using cone beam consistency conditions*, IEEE Transactions on Medical Imaging (2018), 1–1.

[AM76] R E Alvarez and A Macovski, *Energy-selective reconstructions in X-Ray computerised tomography*, Physics in Medicine and Biology **21** (1976), no. 5, 733.

[AMN$^+$13] A. Aichert, M. T. Manhart, B. K. Navalpakkam, R. Grimm, J. Hutter, A. Maier, J. Hornegger, and A. Doerfler, *A realistic digital phantom for perfusion c-arm ct based on mri data*, 2013 IEEE Nuclear Science Symposium and Medical Imaging Conference (2013 NSS/MIC), Oct 2013, pp. 1–2.

[BBM$^+$17] Bastian Bier, Martin Berger, Andreas Maier, Marc Kachelrieß, Ludwig Ritschl, Kerstin Müller, Jang-Hwan Choi, and Rebecca Fahrig, *Scatter correction using a primary modulator on a clinical angiography C-arm CT system*, Medical Physics **44** (2017), no. 9.

[BBR19] Richard Bismark, Oliver Beuing, and Georg Rose, *Truncation artifacts caused by the patient table in polyenergetic statistical* *reconstruction on real C-arm CT data*, The Fifteenth International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine (Pennsylvania, PA, USA), 2019, p. 28.

[BC76] R A Brooks and G Di Chiro, *Beam hardening in X-Ray reconstructive tomography*, Physics in Medicine and Biology **21** (1976), no. 3, 390.

[BFA+20] Richard N.K. Bismark, Robert Frysch, Shiras Abdurahman, Oliver Beuing, Manuel Blessing, and Georg Rose, *Reduction of beam hardening artifacts on real c-arm ct data using polychromatic statistical image reconstruction*, Zeitschrift für Medizinische Physik **30** (2020), no. 1, 40 – 50.

[BFR16] Richard Bismark, Robert Frysch, and Georg Rose, *Reduction of beam hardening artifacts on real C-Arm CT data using statistical polyenergetic image reconstruction*, CT-Meeting 2016: the 4th International Meeting on Image Formation in X-Ray Computed Tomography (Bamberg Germany), 2016.

[BS] H. H. Barrett and W. Swindell, *Radiological imaging: The theory of image formation, detection, and processing*, Cooper Medical Center, Camden, New Jersey.

[BTK99] Ohnesorge B., Flohr T, and Klingenbeck-Regn K., *Efficient object scatter correction algorithm for third and fourth generation ct scanners*, Eur Radiol **9** (1999), 563–569.

[Buz08] Thorsten M. Buzug, *Computed tomography - from photon statistics to modern conea-beam ct*, Springer, 2008.

[BYRM93] H H Barrett, J Yao, J P Rolland, and K J Myers, *Model observers for assessment of image quality*, Proceedings of the National Academy of Sciences **90** (1993), no. 21, 9758–9765.

[CHF+19] Wenchao Cao, Sam Hawker, Gemma Fardell, Ben Price, and Wim Dewulf, *An improved segmentation method for multi-material beam hardening correction in industrial x-ray computed tomography*, Measurement Science and Technology **30** (2019), no. 12, 125403.

[Cie82] Robert Cierniak, *Computed tomography: Some history and recent developments*, Proc Symp Appl Math, 1982.

[CJ00] Ian A. Cunningham and Philip F. Judy, *Computed tomography - the biomedical engineering handbook: Second edition*, Ed. Joseph D. Bronzino, Boca Raton: CRC Press LLC, 2000.

[CRT06] Emmanuel J. CandÃšs, Justin K. Romberg, and Terence Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Communications on Pure and Applied Mathematics **59** (2006), no. 8, 1207–1223.

[Dav65] C. M. Davisson, *Alpha-, beta- and gamma-ray spectroscopy*, 1 ed., pp. 37–78, North-Holland Publishing Company, 1965.

[EF99] H. Erdogan and J. A. Fessler, *Monotonic algorithms for transmission tomography*, IEEE Transactions on Medical Imaging **18** (1999), no. 9, 801–814.

[EF02] I.A. Elbakri and J.A. Fessler, *Statistical image reconstruction for polyenergetic X-Ray computed tomography*, Medical Imaging, IEEE Transactions on **21** (2002), no. 2, 89–99.

[EF03a] Idris A. Elbakri and Jeffrey A. Fessler, *Efficient and accurate likelihood for iterative image reconstruction in x-ray computed tomography*, Medical Imaging 2003: Image Processing (Milan Sonka and J. Michael Fitzpatrick, eds.), vol. 5032, International Society for Optics and Photonics, SPIE, 2003, pp. 1839 – 1850.

[EF03b] Idris A. Elbakri and Jeffrey A. Fessler, *Efficient and accurate likelihood for iterative image reconstruction in x-ray computed tomography*, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 5032, pp. 1839–1850, 2003.

[EF03c] Idris A Elbakri and Jeffrey A Fessler, *Segmentation-free statistical image reconstruction for polyenergetic X-Ray computed tomography with experimental validation*, Physics in Medicine and Biology **48** (2003), no. 15, 2453.

[FBKR19] Robert Frysch, Sebastian Bannasch, Vojtech Kulvait, and Georg Rose, *Efficient nullspace-constrained modifications of incompletely sampled CT images*, 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine (Samuel Matej and Scott D. Metzler, eds.), vol. 11072, International Society for Optics and Photonics, SPIE, 2019, pp. 493 – 497.

[FBMR17] Robert Frysch, Richard Bismark, Andreas Maier, and Georg Rose, *Ray-Density Weighted Algebraic Reconstruction for Volume-of-Interest CT*, Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine (Xi'an, China), vol. 14, June 2017.

[FDK84a] L. A. Feldkamp, L. C. Davis, and J. W. Kress, *Practical cone-beam algorithm*, J. Opt. Soc. Am. A **1** (1984), no. 6, 612–619.

[FDK84b] ———, *Practical cone-beam algorithm*, J. Opt. Soc. Am. A **1** (1984), no. 6, 612–619.

[FESC02] Jeffrey A. Fessler, Idris A. Elbakri, Predrag Sukovic, and Neal H. Clinthorne, *Maximum-likelihood dual-energy tomographic image reconstruction*, vol. 4684, 2002, pp. 38–49.

[Fli08] Torsten Fließbach, *Elektrodynamik*, 5 ed., Spektrum Akademischer Verlag Heidelberg, 2008.

[FPB+14] Robert Frysch, Tim Pfeiffer, Sebastian Bannasch, Steffen Serowy, Sebastian Gugel, Martin Skalej, and Georg Rose, *C-Arm perfusion imaging with a fast penalized maximum-likelihood approach*, vol. 9033, 2014, pp. 90332M–90332M–8.

[FPR13] Robert Frysch, Tim Pfeiffer, and Georg Rose, *Scalable OpenCL Accelerated Multi-GPU Projection for Cone Beam Computed Tomography with a Highly Accurate Separable Footprint Method*, WFITN (Buenos Aires), 2013.

[FR15] Robert Frysch and Georg Rose, *Rigid motion compensation in interventional c-arm ct using consistency measure on projection data*, Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 - Volume 9349 (New York, NY, USA), Springer-Verlag New York, Inc., 2015, pp. 298–306.

[GL11] B. Gonzales and D. Lalush, *Full-spectrum ct reconstruction using a weighted least squares algorithm with an energy-axis penalty*, IEEE Transactions on Medical Imaging **30** (2011), no. 2, 173–183.

[Glo82] G. H. Glover, *Compton scatter effects in ct reconstructions*, Medical Physics **9** (1982), no. 6, 860–867.

[GMV+17] Drosoula Giantsoudi, Bruno De Man, Joost Verburg, Alexei Trofimov, Yannan Jin, Ge Wang, Lars Gjesteby, and Harald Paganetti, *Metal artifacts in computed tomography for radiation therapy planning: dosimetric effects and impact of metal artifact reduction*, Physics in Medicine and Biology **62** (2017), no. 8, R49–R80.

[Hel99] Sigurdur Helgason, *The radon transform*, Springer, Boston, MA, 1999.

[HHS$^+$19] Viktor Haase, Katharina Hahn, Harald Schöndube, Karl Stierstorfer, Andreas Maier, and FrÃ©dÃ©ric Noo, *Impact of the non-negativity constraint in model-based iterative reconstruction from ct data*, Medical Physics **46** (2019), no. 12, e835–e854.

[HMDJ00] Jiang Hsieh, Robert C. Molthen, Christopher A. Dawson, and Roger H. Johnson, *An iterative approach to the beam hardening correction in cone beam CT*, Medical Physics **27** (2000), no. 1, 23–29.

[Hou58] Alston S. Householder, *Unitary Triangularization of a Nonsymmetric Matrix*, Journal of the ACM **5** (1958), no. 4, 33942.

[Hou73] G. N. Hounsfield, *Computerized transverse axial scanning (tomography): Part 1. description of system*, The British Journal of Radiology **46** (1973), no. 552, 1016–1022, PMID: 4757352.

[HP13] X He and S Park, *Model observers in medical imaging research*, Theranostics **3** (2013), no. 10, 774–786.

[HS96] J. H. Hubbell and S. M. Seltzer, *Tables of X-Ray mass attenuation coefficients and mass energy-absorption coefficients from 1 kev to 20 mev for elements z = 1 to 92 and 48 additional substances of dosimetric interest*, 1989, 1990, 1996.

[Hsi98] J Hsieh, *Adaptive streak artifact reduction in computed tomography resulting from excessive x-ray photon noise*, Med Phys **25** (1998), no. 11, 2139–2147.

[HWB$^+$18] Yixing Huang, Tobias Würfl, Katharina Breininger, Ling Liu, Günter Lauritsch, and Andreas Maier, *Some investigations on robustness of deep learning in limited angle tomography*, Medical Image Computing and Computer Assisted Intervention – MICCAI 2018 (Cham) (Alejandro F. Frangi, Julia A. Schnabel, Christos Davatzikos, Carlos Alberola-López, and Gabor Fichtinger, eds.), Springer International Publishing, 2018, pp. 145–153.

[HWF17] T Humphries, J Winn, and A Faridani, *Superiorized algorithm for reconstruction of CT images from sparse-view and limited-angle polyenergetic data*, Physics in Medicine & Biology **62** (2017), no. 16, 6762–6783.

[HXH$^+$17] Xiaolin Huang, Yan Xia, Yixing Huang, Joachim Hornegger, and Andreas K. Maier, *Overexposure correction by mixed one-bit compressive sensing for c-arm ct*, Bildverarbeitung für die Medizin, 2017.

[HYH$^+$18] Akinori Hata, Masahiro Yanagawa, Osamu Honda, Noriko Kikuchi, Tomo Miyata, Shinsuke Tsukagoshi, Ayumi Uranishi, and Noriyuki Tomiyama, *Effect of matrix size on the image quality of ultra-high-resolution ct of the lung: Comparison of 512x512, 1024x1024, and 2048x2048*, Academic Radiology **25** (2018), no. 7, 869 – 876.

[JF07] M. W. Jacobson and J. A. Fessler, *An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms*, IEEE Transactions on Image Processing **16** (2007), no. 10, 2411–2422.

[JS78a] Peter M. Joseph and Robin D. Spital, *A Method for Correcting Bone Induced Artifacts in Computed Tomography Scanners*, Journal of Computer Assisted Tomography **2** (1978), 100–108.

[JS78b] PM Joseph and RD Spital, *A method for correcting bone induced artifacts in computed tomography scanners*, Journal of computer assisted tomography **2** (1978), no. 1, 100–108.

[Kac37] Stefan Kaczmarz, *Angenährte aufloesung von systemen linearer gleichungen*, Bulletin International de l'Academie Polonaise des Sciences et des Lettres. Classe des Sciences Mathematiques et Naturelles. Serie A, Sciences Mathematiques **35** (1937), 355.

[Kal06] W Kalender, *Computertomographie. Grundlagen, Gerätetechnologie, Bildqualität, Anwendungen. 2., überarb. und erw. Auflage.*, Publicis Corporate Publishing, Erlangen, Germany, 2006.

[KB98] C. Kamphuis and F. J. Beekman, *Accelerated iterative transmission ct reconstruction using an ordered subsets convex algorithm*, IEEE Transactions on Medical Imaging **17** (1998), no. 6, 1101–1105.

[KKF08] M. Krumm, S. Kasperl, and M. Franz, *Reducing non-linear artifacts of multi-material objects in industrial 3d computed tomography*, NDT and E International **41** (2008), no. 4, 242–251.

[KPB16] Andrzej Kordecki, Henryk Palus, and Artur Bal, *Practical vignetting correction method for digital camera with measurement of surface luminance distribution*, Signal, Image and Video Processing **10** (2016), 1417–1424.

[KPTF13] D. Kim, D. Pal, J. B. Thibault, and J. A. Fessler, *Accelerating ordered subsets image reconstruction for x-ray ct using spatially nonuniform optimization transfer*, IEEE Transactions on Medical Imaging **32** (2013), no. 11, 1965–1978.

[KRK12] Michael Knaup, Ludwig Ritschl, and Marc Kachelrieß, *Digitization and visibility issues in flat detector ct: A simulation study*, 10 2012, pp. 2661–2666.

[KSK06] Marc Kachelrieß, Katia Sourbelle, and Willi A. Kalender, *Empirical cupping correction: A first-order raw data precorrection for cone-beam computed tomography*, Medical Physics **33** (2006), no. 5, 1269–1274.

[KW00] Sing Bing Kang and Richard Weiss, *Can we calibrate a camera using an image of a flat,textureless lambertian surface?*, Computer Vision — ECCV 2000 (Berlin, Heidelberg) (David Vernon, ed.), Springer Berlin Heidelberg, 2000, pp. 640–653.

[KW03] J. Kruger and R. Westermann, *Acceleration techniques for gpu-based volume rendering*, Proceedings of the 14th IEEE Visualization 2003 (VIS03) (USA), VIS03, IEEE Computer Society, 2003, p. 38.

[Lan51] L. Landweber, *An iteration formula for fredholm integral equations of the first kind*, American Journal of Mathematics **73** (1951), no. 3, 615–624.

[LDL75] E. M. Lifshitz L. D. Landau, *The classical theory of fields*, 4 ed., vol. 2, Butterworth-Heinemann, 1975.

[LFB10] Yong Long, Jeffrey A. Fessler, and James M. Balter, *3D forward and back-projection for X-Ray CT using separable footprints*, IEEE Trans. Med. Imaging **29** (2010), no. 11, 1839–1850.

[LFN09] C. Lemmens, D. Faul, and J. Nuyts, *Suppression of metal artifacts in CT using a reconstruction procedure that combines MAP and projection completion*, IEEE Transactions on Medical Imaging **28** (2009), no. 2, 250–260.

[LHY00] Kenneth Lange, David R. Hunter, and Ilsoon Yang, *Optimization transfer using surrogate objective functions*, Journal of Computational and Graphical Statistics **9** (2000), no. 1, 1–20.

[LJC$^+$17] June-Goo Lee, Sanghoon Jun, Young-Won Cho, Hyunna Lee, Guk Bae Kim, Joon Beom Seo, and Namkug Kim, *Deep learning in medical imaging: General overview*, Korean journal of radiology **18** (2017), no. 28670152, 570–584 (eng).

[LMX10] Tong Liu, A. A. Malcolm, and Jian Xu, *Pincushion distortion correction in x-ray imaging with an image intensifier*, Fourth International Conference on Experimental Mechanics (Chenggen

Quan, Kemao Qian, Anand Krishna Asundi, and Fook Siong Chau, eds.), vol. 7522, International Society for Optics and Photonics, SPIE, 2010, pp. 1085 – 1093.

[LR09] Claude Leroy and Pier-Giorgio Rancoita, *Principles of radiation interaction in matter and detection*, 2nd edition ed., p. page 155, World Scientific Publishing Co. Pte. Ltd., 2009, equation 2.188.

[MES$^+$18] K. Mechlem, S. Ehn, T. Sellerer, E. Braig, D. MÃŒnzel, F. Pfeiffer, and P. B. Noël, *Joint statistical iterative material image reconstruction for spectral computed tomography using a semi-empirical forward model*, IEEE Transactions on Medical Imaging **37** (2018), no. 1, 68–80.

[MMMF15] Wu Meng, Kerstin Müller, Michael P. Marks, and Rebecca Fahrig, *Evaluation of two-pass view aliasing artifact suppression algorithm using clinical data*, World Congress on Medical Physics and Biomedical Engineering, June 7-12, 2015, Toronto, Canada (David A. Jaffray, ed.), 2015.

[MMP16] Thomas G. Mayerhoefer, Harald Mutschke, and Jürgen Popp, *Employing theories far beyond their limits - the case of the (boguer-) beer-lambert law*, ChemPhysChem **17** (2016), no. 13, 1948–1955.

[MMS90] J. M. Meagher, C. D. Mote, and H. B. Skinner, *CT image correction for beam hardening using simulated projection data*, IEEE Transactions on Nuclear Science **37** (1990), no. 4, 1520–1524.

[Moo65] G Moore, *Cramming more components onto integrated circuits*, Electronics **38** (1965), no. 8, 114–117.

[MPND18] Jonathan H Mason, Alessandro Perelli, William H Nailon, and Mike E Davies, *Quantitative cone-beam CT reconstruction with polyenergetic scatter model fusion*, Physics in Medicine & Biology **63** (2018), no. 22, 225001.

[MSCH18] A. Maier, S. Steidl, V. Christlein, and J. Hornegger, *Medical imaging systems - an introductory guide*, 2018.

[Nat86] Frank Natterer, *The mathematics of computerized tomography*, Vieweg+Teubner Verlag, 1986.

[NBC+15] Vincent Van Nieuwenhove, Jan De Beenhouwer, Francesco De Carlo, Lucia Mancini, Federica Marone, and Jan Sijbers, *Dynamic intensity normalization using eigen flat fields in x-ray imaging*, Opt. Express **23** (2015), no. 21, 27975–27989.

[NL79] O Nalcioglu and R Y Lou, *Post-reconstruction method for beam hardening in computerised tomography*, Physics in Medicine and Biology **24** (1979), no. 2, 330.

[Par55] C. Northcote Parkinson, *Parkinson's law*, The Economist **177** (1955), no. 5856, 635–637.

[PFGR13] Tim Pfeiffer, Robert Frysch, Sebastian Gugel, and Georg Rose, *Ml reconstruction of cone-beam projections acquired by a flat-panel rotational x-ray device*, 2013.

[Pie93] A. R. De Pierro, *On the relation between the ISRA and the EM algorithm for positron emission tomography*, IEEE Transactions on Medical Imaging **12** (1993), no. 2, 328–333.

[Pie95] ______, *A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography*, IEEE Transactions on Medical Imaging **14** (1995), no. 1, 132–137.

[PLD+09] G Poludniowski, G Landry, F DeBlois, P M Evans, and F Verhaegen, *Spekcalc : a program to calculate photon spectra from tungsten anode X-Ray tubes*, Physics in Medicine and Biology **54** (2009), no. 19, N433.

[Rad17] Johann Radon, *Uber die bestimmung von funktionen durch ihre integralwerte langs gewisser mannigfaltigkeiten*, Ber. Saechsische Akad. Wiss **29** (1917), 262.

[RHN+08] DA Roberts, VN Hansen, AC Niven, MG Thompson, J Seco, and PM Evans, *A low Z linac and flat panel imager: comparison with the conventional imaging approach*, PHYSICS IN MEDICINE AND BIOLOGY **53** (2008), no. 22, 6305 – 6319.

[SBBW12] Jack W. Scannell, Alex Blanckley, Helen Boldon, and Brian Warrington, *Diagnosing the decline in pharmaceutical r&d efficiency*, Nature Reviews Drug Discovery **11** (2012), no. 3, 191–200.

[Sch05] Henning Scharsach, *Advanced gpu raycasting*, In Proceedings of CESCG 2005, 2005, pp. 69–76.

[SFG86] J G Stears, J P Felmlee, and J E Gray, *Half-value-layer increase owing to tungsten buildup in the x-ray tube: fact or fiction.*, Radiology **160** (1986), no. 3, 837–838, PMID: 3737925.

[Sla88] Avinash C. Kak Malcolm Slaney, *Principles of computerized tomographic imaging*, New York: IEEE Press, 1988.

[SSL10] M Sun and J M Star-Lack, *Improved scatter correction using adaptive scatter kernel superposition*, Physics in Medicine and Biology **55** (2010), no. 22, 6695–6720.

[STG$^+$20] J. Webster Stayman, Matthew Tivnan, Grace J. Gang, Wenying Wang, Nadav Shapira, and Peter B. Noel, *Grating-Based Spectral CT using Small Angle X-Ray Beam Deflections*, CT-Meeting 2016: the 6th International Meeting on Image Formation in X-Ray Computed Tomography (Regensburg Germany), 2020.

[SV08] Thomas Strohmer and Roman Vershynin, *A randomized kaczmarz algorithm with exponential convergence*, Journal of Fourier Analysis and Applications **15** (2008), no. 2, 262.

[Tan71] Kunio Tanabe, *Projection method for solving a singular system of linear equations and its applications*, Numerische Mathematik **17** (1971), 203–214.

[THC$^+$18] Shaojie Tang, Kuidong Huang, Yunyong Cheng, Xuanqin Mou, and Xiangyang Tang, *Optimization based beam-hardening correction in CT under data integral invariant constraint*, Physics in Medicine & Biology **63** (2018), no. 13, 135015.

[Wat82] David S. Watkins, *Understanding the qr algorithm*, SIAM Review **24** (1982), no. 4, 427–440.

[WKC$^+$14] M Wu, A Keil, D Constantin, J Star-Lack, L Zhu, and R Fahrig, *Metal artifact correction for x-ray computed tomography using kv and selective mv imaging*, Med Phys **41** (2014), no. 12, 121910.

[WLDK12] Daniel Weiss, R. Lonardoni, A. Deffner, and Christoph Kuhn, *Geometric image distortion in flat-panel x-ray detectors and its influence on the accuracy of ct-based dimensional measurements*, 09 2012.

[WMM$^+$11] X Wang, D Meier, S Mikkelsen, G E Maehlum, D J Wagenaar, B M W Tsui, B E Patt, and E C Frey, *MicroCT with energy-resolved photon-counting detectors*, Physics in Medicine and Biology **56** (2011), no. 9, 2791–2816.

[WMYF15] Meng Wu, Andreas Maier, Qiao Yang, and Rebecca Fahrig, *A novel filtered backprojection-based algorithm for sparse view ct image reconstruction*, Fully3D, 06 2015.

[WNGH14] Adam Wunderlich, Frederic Noo, Brandon Gallas, and Marta Heilbrun, *Exact confidence intervals for channelized hotelling observer performance in image quality studies*, IEEE transactions on medical imaging **34** (2014).

[WYMF14] Meng Wu, Qiao Yang, Andreas Maier, and Rebecca Fahrig, *A practical statistical polychromatic image reconstruction for computed tomography using spectrum binning*, Proc. SPIE Medical Imaging 2014 (SPIE, ed.), 2014, pp. 9033–26.

[XM12] W Xu and K Mueller, *Efficient low-dose ct artifact mitigation using an artifact-matched prior scan*, Med Phys **39** (2012), no. 8, 4748–4760.

[XUK$^+$17] S Xu, A Uneri, A Jay Khanna, J H Siewerdsen, and J W Stayman, *Polyenergetic known-component CT reconstruction with unknown material compositions and unknown x-ray spectra*, Physics in Medicine and Biology **62** (2017), no. 8, 3352–3374.

[YLS06] Yuanjie Zheng, S. Lin, and Sing Bing Kang, *Single-image vignetting correction*, 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), vol. 1, 2006, pp. 461–468.

[YWB$^+$00] Chye Hwang Yan, R. T. Whalen, G. S. Beaupre, S. Y. Yen, and S. Napel, *Reconstruction algorithm for polychromatic CT imaging: application to beam hardening correction*, IEEE Transactions on Medical Imaging **19** (2000), no. 1, 1–11.

[ZD20] Hai-Miao Zhang and Bin Dong, *A review on deep learning in medical image reconstruction*, Journal of the Operations Research Society of China (2020).

[ZNG08] Andy Ziegler, Tim Nielsen, and Michael Grass, *Iterative reconstruction of a region of interest for transmission tomography*, Medical Physics **35** (2008), no. 4, 1317–1327.

[ZSRB05] M. Zellerhoff, B. Scholz, E.-P. Ruehrnschopf, and T. Brunner, *Low contrast 3D reconstruction from C-arm data*, 2005, pp. 646–655.