Faculteit Wetenschappen
Departement Fysica

# Local Prior Knowledge in Tomography

---

# Lokale Voorkennis in Tomografie

Proefschrift voorgelegd tot het behalen van de graad van

**Doctor in de Wetenschappen: Fysica**

aan de Universiteit Antwerpen, te verdedigen door

**Tom Roelandts**

**Promotoren**

Prof. Dr. Jan Sijbers
Prof. Dr. Joost Batenburg

Antwerpen, 2013

**Members of the Jury**

Prof. Dr. Sara Bals[1]
Prof. Dr. Joost Batenburg[2]
Prof. Dr. Joris Dirckx[1]
Dr. Xuan Liu[3]
Prof. Dr. Johan Nuyts[4]
Prof. Dr. Jan Sijbers[1]
...

[1]University of Antwerp, Antwerp, Belgium
[2]Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
[3]Bruker microCT, Kontich, Belgium
[4]KU Leuven, Leuven, Belgium

**Contact Information**

Tom Roelandts
✉ Vision Lab, Department of Physics
University of Antwerp (CDE)
Universiteitsplein 1, N-1.14
2610 Wilrijk, Belgium
☎ +32 (0) 3 265 28 40
+32 (0) 3 265 22 45
tom.roelandts@ua.ac.be
http://www.visielab.ua.ac.be/people/tom-roelandts

# Table of Contents

# Samenvatting

## Inleiding

Computertomografie (CT) is een techniek die computerberekeningen gebruikt om een beeld te vormen van de binnenkant van een object of een persoon, door *projecties* van dat object of die persoon te combineren. Het woord *tomografie* is afkomstig van het Griekse *tomos*, wat "snede" of "schijf" betekent. De basis voor computertomografie werd in 1917 gelegd door Johann Radon, een Oostenrijkse wiskundige.

Computertomografie heeft zeer diverse toepassingen, waarvan de meest bekende zeker medische beeldvorming is (de CT-scanner), waar X-stralen gebruikt worden voor het maken van de projecties. De eerste praktische toepassing van CT was echter in de astronomie, door Ronald Bracewell in 1956. Hij gebruikte CT om de resolutie van radio-astronomische waarnemingen te verbeteren. De praktische toepassingen in deze thesis zijn afkomstig uit elektronentomografie, waarbij de opnames gemaakt worden met een elektronenmicroscoop, en in preklinisch onderzoek, waarbij de opnames gemaakt worden met een $\mu$CT-scanner.

Er zijn twee belangrijke technieken voor de *reconstructie* van de beelden, namelijk de analytische en de algebraïsche methodes. In medische scanners wordt zeer veel gebruik gemaakt van *gefilterde terugprojectie*, een analytische methode. Deze thesis bouwt voort op de algebraïsche methodes, waarbij het probleem gesteld wordt als een stelsel van lineaire vergelijkingen $\boldsymbol{W}\boldsymbol{x} = \boldsymbol{p}$, waarbij $\boldsymbol{x}$ het onbekende beeld is, $\boldsymbol{p}$ de projectiedata, en $\boldsymbol{W}$ de *systeemmatrix* die het verband tussen beide beschrijft.

In praktische CT is het niet altijd mogelijk om de projectiebeelden te maken die nodig zijn voor een accurate reconstructie (het aantal, onder welke hoeken, etc.). Hiervoor zijn diverse oorzaken, waarvan een belangrijke het beperken van de stralingsdosis is. Een evident voorbeeld hiervan zijn X-stralen, die schadelijk kunnen zijn voor de mens. Maar ook levenloze objecten kunnen beschadigd worden door straling. Om dat te vermijden, moet dikwijls de dosis verlaagd worden, waardoor de signaal-ruisverhouding afneemt en de reconstructie moeilijker wordt. Sommige beperkingen zijn eigen aan de beeldvormingsmodaliteit zelf. Bij elektronentomografie, bijvoorbeeld, worden de opnames veelal manueel gemaakt, waardoor het opnemen van zelfs een beperkt aantal projecties al erg arbeidsintensief kan zijn. Naast de beperkingen van de microscoop zelf, is er ook het gegeven dat microscopiemonsters meestal vlak zijn, waardoor hun schijnbare dikte toeneemt wanneer ze bekeken worden onder een grote hoek. Hierdoor kunnen er geen opnames gemaakt worden over de volle 180°. Het resultaat van dit soort beperkingen is dat het reconstructiealgoritme moet werken met *beperkte data*, waardoor het on-

mogelijk kan worden om accurate reconstructies te maken.

Een oplossing voor het probleem van de beperkte data is het inbrengen van *voorkennis* in het reconstructiealgoritme. In deze thesis worden twee veronderstellingen gebruikt. De eerste is dat het gescande object opgebouwd is uit homogene gebieden. De tweede is dat de grijswaarden van deze gebieden bekend zijn. Beide veronderstelling samen leiden tot *discrete tomografie* (DT). Enkel de eerste veronderstelling leidt tot algoritmes die werken met *minimalisatie van de totale variatie* (TVMin). De praktische algoritmes die we hiervoor gebruiken in deze thesis zijn DART voor DT, en NESTA en FISTA voor TVMin. Het gebruik van deze algoritmes resulteert in een grote verbetering van de kwaliteit van de reconstructies.

Een probleem met het inbrengen van voorkennis is echter dat deze niet altijd geldt voor het gehele object. De genoemde methodes kunnen dan falen. In deze thesis onderzoeken we de mogelijkheid om de voorkennis enkel op een gedeelte van het object toe te passen. Dit kan op twee manieren, ofwel door het ontwikkelen van een specifiek algoritme voor elke specifieke situatie, ofwel door het ontwikkelen van een algemeen toepasbare techniek om *lokale* versies te maken van bestaande algoritmes. Beide benaderingen worden onderzocht.

## PDART

De *Partieel Discrete Algebraïsche Reconstructietechniek* ("*Partially Discrete Algebraic Reconstruction Technique*", PDART, hoofdstuk 2) werd ontwikkeld voor een specifieke toepassing: de reconstructie en segmentatie van homogene deeltjes met een hoge dichtheid in een (mogelijk) niet-homogene omgeving.

Het PDART-algoritme alterneert SIRT-iteraties met segmentatiestappen. SIRT is een algebraïsche reconstructietechniek die in deze thesis meermaals gebruikt wordt. Tijdens elke segmentatiestap worden alle pixels met een waarde boven een zekere drempel $\tau$ verondersteld deel uit te maken van een homogeen deeltje met hoge dichtheid. De grijswaarde $\rho$ van de deeltjes wordt ook verondersteld gekend te zijn. Na elke segmentatiestap worden de pixels waarvan de waarde de drempel overschrijdt *vastgezet* op de gekende grijswaarde. Dit vastzetten is de cruciale stap waardoor het algoritme resulteert in een accuratere reconstructie. Aangezien het reconstructieprobleem bij de algebraïsche methodes gesteld wordt als een stelsel van lineaire vergelijkingen, komt het vastzetten van pixels erop neer dat onbekenden geschrapt worden uit het stelsel, terwijl het aantal vergelijkingen gelijk blijft, wat betekent dat alle beschikbare data nog steeds gebruikt wordt. Het gevolg is dat de onbepaaldheid van het stelsel afneemt, zodat het dichte én het andere materiaal beter gereconstrueerd wordt.

In de experimenten wordt PDART eerst toegepast op een aantal fantoombeel-

den, om het algoritme te karakteriseren. Daarna wordt het toegepast op twee datasets uit elektronentomografie (in hoofdstuk 2), en een dataset uit preklinisch onderzoek (in hoofdstuk 3). De resultaten geven aan dat het algoritme resulteert in accuratere reconstructies dan SIRT.

## Gereconstrueerde residuele fout

De *gereconstrueerde residuele fout* ("*reconstructed residual error*", hoofdstuk 3) is de reconstructie van het verschil tussen de opgenomen data en de voorwaartse projectie van een gesegmenteerd tomogram. De Moore-Penrose pseudoinverse van de systeemmatrix $\boldsymbol{W}$ wordt gebruikt als wiskundig model voor een algemeen reconstructiealgoritme, om aan te tonen dat de gereconstrueerde residuele fout een goede benadering is van de echte fout in het gesegmenteerde tomogram. De werking van het algoritme is gebaseerd op het feit dat de segmentatie de artefacten van de originele reconstructie onderdrukt. Er worden door de segmentatie weliswaar nieuwe artefacten toegevoegd, maar deze hebben andere eigenschappen. Hierdoor is het mogelijk om opnieuw gebruik te maken van de projectiedata om een beeld te vormen van de fouten in de gesegmenteerde reconstructie.

De techniek heeft verschillende toepassingen. Ze kan de grijswaardeschatting van een gesegmenteerd tomogram verbeteren. Voor reconstructiemethodes die voorkennis van de grijswaarden gebruiken tijdens de reconstructie, kan ook de segmentatie zelf verbeterd worden. Tenslotte kan de gereconstrueerde residuele fout ook gebruikt worden om te bepalen welke de meest accurate van verschillende gesegmenteerde reconstructies is.

In de experimenten wordt het algoritme eerst toegepast op een aantal fantoombeelden, om het te karakteriseren. Daarna wordt het toegepast op een dataset uit preklinisch onderzoek. De resultaten geven aan dat de gereconstrueerde residuele fout in staat is om de fouten in een gesegmenteerde reconstructie te analyseren en te verbeteren.

## Lokale homogeniteit bij reconstructie

Voor de tweede mogelijkheid om voorkennis enkel toe te passen waar ze geldig is, na PDART als voorbeeld van een algoritme dat ontwikkeld werd voor een specifieke situatie, maken we gebruik van de gereconstrueerde residuele fout. Voor objecten die gedeeltelijk homogeen zijn, hebben we een methode ontwikkeld om reconstructiealgoritmes aan te passen zodat hun prior enkel in de homogene gebieden toegepast wordt (hoofdstuk 4).

Uit de gereconstrueerde residuele fout wordt een masker afgeleid door middel van een drempel. Van de pixels met een fout die groter is dan deze drempel wordt

verondersteld dat ze niet tot een homogeen gebied behoren, en dat het beter is om de prior voor deze pixels niet toe te passen. De praktische toepassing hiervan is verschillend voor elk algoritme. In deze thesis worden er lokale versies ontwikkeld van DART en FISTA. Bij *Lokale DART* ("*Local DART*") wordt het algoritme aangepast om er een verzameling niet-discrete pixels aan toe te voegen. Het resulterende algoritme levert in één stap een gelokaliseerd resultaat op. Bij *Lokale FISTA* ("*Local FISTA*") wordt een initiële (klassieke) FISTA-reconstructie gebruikt om, met behulp van de gereconstrueerde residuele fout, de niet-homogene gebieden te bepalen, waarna er een tweede reconstructie met een lokale versie van FISTA volgt.

In de experimenten worden de algoritmes toegepast op een aantal fantoombeelden, om ze te karakteriseren. De resultaten geven aan dat de lokale versies van de algoritmes betere resultaten opleveren dan niet-lokale versies en SIRT.

## Conclusie

Bij objecten die gedeeltelijk uit homogene gebieden bestaan, is het mogelijk om de homogene gebieden in kaart te brengen via de gereconstrueerde residuele fout (hoofdstuk 3). Met zowel specifiek ontwikkelde algoritmes (PDART, hoofdstuk 2) en het lokaal maken van bestaande algoritmes (hoofdstuk 4), is het mogelijk om deze objecten te reconstrueren met behoud van de voorkennis in de gebieden waar ze geldig is. Uit deze thesis blijkt dat ook lokale voorkennis de kwaliteit van reconstructies verbetert.

# Summary

## Introduction

Computed tomography (CT) is a technique that uses computation to form an image of the inside of an object or person, by combining *projections* of that object or person. The word *tomography* is derived from the Greek word *tomos*, meaning "cut" or "slice". The basis for computed tomography was laid in 1917 by Johann Radon, an Austrian mathematician.

Computed tomography has a broad range of applications, the best known being medical imaging (the CT scanner), where X-rays are used for making the projection images. The first practical application of CT was, however, in astronomy, by Ronald Bracewell in 1956. He used CT to improve the resolution of radio-astronomical observations. The practical applications in this thesis are from electron tomography, where the images are made with an electron microscope, and from preclinical research, where the images are made with a $\mu$CT scanner.

There are two important techniques for the *reconstruction* of the images, namely analytical and algebraic methods. Medical scanners mostly employ *filtered backprojection*, an analytical method. This thesis builds on algebraic methods, where the problem is cast as a system of linear equations $\boldsymbol{W}\boldsymbol{x} = \boldsymbol{p}$. Here, $\boldsymbol{x}$ is the unknown image, $\boldsymbol{p}$ represents the projection data, and $\boldsymbol{W}$ the *system matrix*, which describes the relation between both.

In practical CT, it is not always possible to record the projection images that are necessary for creating an accurate reconstruction (their number, from which angles, etc.). There are several reasons for this, an important one being limiting the radiation dose. A straightforward example of this are X-rays, which can be harmful for man, but lifeless objects can also be damaged by radiation. To avoid this, the radiation dose must often be limited, which decreases the signal-to-noise ratio and makes the reconstruction more difficult. The imaging modality itself can also have inherent limitations. In electron tomography, e.g., the projection images are often made manually, so that recording even a small number of projections can be quite labor intensive. Apart from the limits of the microscope itself, there is also the fact that the samples are mostly planar, so that their apparent thickness increases when they are imaged under a large angle. The result of this is that projections cannot be made over the full range of 180°. The consequence of these kinds of limitations is that the reconstruction algorithms only have access to *limited data*, which can make it impossible to create accurate reconstructions.

A solution for the limited data problem is adding *prior knowledge* to the reconstruction algorithm. In this thesis, two assumptions are employed. The first one is that the scanned object consists of homogeneous regions. The second one is that

the gray levels of these regions are known. Together, these assumptions lead to *discrete tomography* (DT). The first assumption by itself leads to algorithms that employ *total variation minimization* (TVMin). The practical algorithms that are used in this thesis are DART for DT, and NESTA and FISTA for TVMin. Using these algorithms results in a large improvement of the quality of the reconstructions.

A problem with applying prior knowledge is that there are cases where the prior knowledge is not valid for the entire object. The methods that were mentioned before can then fail. In this thesis, we investigate the possibility of only applying the prior knowledge to part of the object. There are two ways to do this, either developing a specific algorithm for each specific situation, or developing a general technique for creating *local* versions of existing algorithms. Both approaches are investigated.

## PDART

The *Partially Discrete Algebraic Reconstruction Technique* (PDART, Chapter 2) was developed for a specific application, namely the reconstruction and segmentation of homogeneous particles with a high density in a (possibly) non-homogeneous environment.

The PDART algorithm interleaves SIRT iterations with segmentation steps. SIRT is an algebraic reconstruction technique that is used several times in this thesis. During each segmentation step, all pixels with a value above a certain threshold $\tau$ are assumed to be part of a homogeneous particle with a high density. The gray level $\rho$ of the particles is also assumed to be known. After each segmentation step, the pixels of which the value exceeds the threshold are *fixed* at the known gray level. This fixing is the crucial step through which the algorithm achieves a more accurate reconstruction. Since the reconstruction problem is, for algebraic methods, seen as a system of linear equations, fixing pixels amounts to removing variables from the system, while the number of equations stays the same (which means that all available data is still used). The consequence is that the system becomes less underdetermined, so that the reconstruction quality of both the dense and the other material is improved.

In the experiments, PDART is first applied to a number of phantom images, to characterize the algorithm. It is then applied to two datasets from electron tomography (in Chapter 2) and a dataset from preclinical research (in Chapter 3). The results indicate that the algorithm produces more accurate reconstructions than SIRT.

## Reconstructed Residual Error

The *reconstructed residual error* is the reconstruction of the difference between the recorded data and the forward projection of a segmented tomogram. The Moore-Penrose pseudoinverse of the system matrix $\boldsymbol{W}$ is used as a mathematical model for a general reconstruction algorithm, to show that the reconstructed residual error is a good approximation of the true error of the segmented tomogram. The operation of the algorithm is based on the fact that segmentation suppresses the artifacts of the original reconstruction. The segmentation step does add new artifacts, but these have different properties. This makes it possible to use the projection data again to form an image of the errors in the segmented reconstruction.

The technique has several applications. It can improve the gray level estimates of a segmented tomogram. For reconstruction methods that apply prior knowledge of the gray levels during the reconstruction, the segmentation itself can also be improved. Finally, the reconstructed residual error can be used to select the most accurate of several segmented reconstructions.

In the experiments, the algorithm is first applied to a number of phantom images, to characterize it. It is then applied to a dataset from preclinical research. The results indicate that the reconstructed residual error is able to find and correct the errors in a segmented tomogram.

## Local Homogeneity in Reconstruction

For the second approach for applying prior knowledge only at locations where it is valid, after PDART as an example of an algorithm that was developed for a specific situation, we make use of the reconstructed residual error. For objects that are partially homogeneous, we have developed a method to adapt reconstruction algorithms so that their prior is only applied in the homogeneous regions (Chapter 4).

A mask is determined from the reconstructed residual error by applying a threshold. It is assumed that pixels for which the error is larger than the threshold are not part of a homogeneous region, and that it is better not to apply the prior for those pixels. The practical implementation for this is different for each algorithm. In this thesis, local versions of DART and FISTA are developed. For *Local DART*, the algorithm is adapted by adding a set of non-discrete pixels. The adapted algorithm then provides a localized result in a single step. For *Local FISTA*, an initial (classical) FISTA reconstruction is used to locate the non-homogeneous regions, which is then followed by a second reconstruction that uses a local version of FISTA.

In the experiments, these algorithms are applied to a number of phantom im-

ages, to characterize them. The results indicate that the local versions of the algorithms produce better results than non-local versions and SIRT.

## Conclusions

For objects that consist partially of homogeneous regions, it is possible to locate the homogeneous regions using the *reconstructed residual error* (Chapter 3). Both with specifically developed algorithms (PDART, Chapter 2) and by localizing existing algorithms (Chapter 4), it is possible to reconstruct these objects with preservation of the prior knowledge in the regions where it is valid. This thesis shows that also local prior knowledge increases the quality of reconstructions.

# 1 Introduction

## 1.1 Historical Overview

Computed tomography (CT) is a technique that uses *computation* to reconstruct an image of the inside of a person or an object from (many) *projections*. The word *tomography* is derived from the Greek word *tomos*, meaning "cut" or "slice", so the straightforward meaning of the word is simply that it shows a slice through a person or an object.

CT is practiced today in diverse application fields such as medicine, preclinical research, materials science, astronomy, etc. This thesis focuses on *transmission* CT, where penetrating rays are produced and sent through the person or object that is scanned. This results in *projection images*, which are then used as input to create the CT images from. In medical practice, which is by far the best known field of application, these penetrating rays are most commonly *X-rays*, as discovered by Wilhelm Röntgen (1845–1923) in 1895, and for which he later received the Nobel Prize in Physics in 1901.

The mathematical basis of CT was laid by Johann Radon (1887–1956), an Austrian Mathematician. In 1917, he published a seminal paper [1] in which the tomographic reconstruction problem was, from a mathematical point of view, solved. The importance of this paper as the mathematical basis for tomography was realized only much later. This was probably due to it being presented as a purely mathematical subject, for which no practical applications were identified at the time. The paper being in German possibly also hindered a wide distribution of the results. Since the 1980s, two independent English translations are available, one as a journal article [2], and one in the book by Deans [3].

The first practical application of tomography was, maybe somewhat unexpectedly, in astronomy, by Ronald Bracewell in 1956 [4]. He used CT to improve the resolution of radio-astronomical observations.

We mention the first three-dimensional results from electron microscopy in

this historical overview, since results for several *electron tomography* datasets are described in Chapter 2. Three important papers were published in 1968. In the first one [5], a biological macromolecule is reconstructed from a single projection image containing many copies of the molecule in different orientations, a technique that later became known as *single particle analysis*. In the second and third paper [6, 7], multiple projections are taken, alleviating the need for the object to be symmetrical.

The breakthrough for medical applications arrived through the work of Alan Cormack (1924–1998) and Sir Godfrey Hounsfield (1919–2004), who jointly received the Nobel Prize in Physiology or Medicine for this in 1979. Cormack proposed X-ray scanning for medical applications in 1963 [8]. The first practical CT scanner was built at EMI [9, 10]. It is often claimed that revenues from the sales of The Beatles records in the 1960s helped funding the development of the first CT scanner at EMI [11], although this has recently been disputed [12].

A more extensive overview of the history of tomography than is presented in this Introduction is available in the excellent books by Deans [3] and Buzug [13].

## 1.2 Workflow of Tomography

In this Section, we present a brief overview of the workflow of tomographic imaging. The three major steps are data acquisition, reconstruction, and segmentation.

### 1.2.1 Data Acquisition

The input of the tomographic reconstruction methods (see Section 1.2.2) are *projection images*. In order to create an image of the inside of an object through transmission CT, *penetrating radiation* is needed. Moreover, part of that radiation must be *absorbed* or *scattered* on its path through the object, in order to be able to provide information about it. Different types of radiation have been used for tomography, such as X-rays, electrons, neutrons, etc. Since this thesis features CT reconstructions from both electron and X-ray tomography, the data acquisition process for these imaging modes is briefly discussed below.

In electron tomography, the projection images are created with an electron microscope. More specifically, the examples that appear in this thesis were acquired with a scanning transmission electron microscope (STEM). The penetrating radiation in an electron microscope is a beam of *electrons*, created using an *electron gun*. This beam is then focused by electrostatic and electromagnetic lenses and sent through the sample to image it. Electron microscopes maintain an internal vacuum, which often necessitates extra work to prepare the samples for imaging, since the electron beam would otherwise be scattered by air molecules. After pass-

ing through the sample, the remaining electrons hit a detector, resulting in an image.

In medical and $\mu$CT scanners, the penetrating radiation consists of *X-rays*, which are generated by an *X-ray tube*. The examples that appear in this thesis were acquired with a $\mu$CT scanner. The X-ray tube produces a cone beam of X-rays, which is then sent through the sample to image it. Maintaining a vacuum is not required, since X-ray photons are not affected much by air molecules. After passing through the sample, the remaining X-ray photons hit a detector, resulting in an image.

### 1.2.2 Reconstruction

The mathematical basis for tomography was discovered by the Austrian mathematician Johann Radon. In 1917, he published the seminal paper "Über die Bestimmung von Funktionen durch ihre Intergralwerte längs gewisser Mannigfaltigkeiten" [1] ("On the Determination of Functions from Their Integral Values along Certain Manifolds", English translations are available in [2] and [3]). Given a function on the plane $f : \mathbb{R}^2 \to \mathbb{R}$ with bounded support, Radon showed that it is possible to reconstruct that function from a complete set of line integrals formed by *projecting* that function in all possible directions.

Hence, to enable tomographic reconstruction in practice, a large number of projection images must be acquired from different directions. This is illustrated, for a single slice, in Fig. 1.1. If projections are taken in equiangular steps, the projection of a single point in the object is supported on the graph of a sine wave. For this reason, the graphical representation of a full set of projections (in two dimensions) is called a *sinogram*. From the phantom image from Fig. 1.1a, which contains a number of elliptical particles that are embedded in a disk, a sinogram was created using 360 equiangular projections spaced 0.5° apart (Fig. 1.1b).

**(a)** Phantom **(b)** Sinogram **(c)** SIRT reconstruction

**Figure 1.1:** (a) Phantom image. (b) Sinogram from 360 projections. (c) SIRT reconstruction.

In practice, there are two main approaches for computing a tomographic reconstruction. The first is the one of the analytical methods, which are a direct approximation of Radon's solution. Examples of this are *Filtered Backprojection* (FBP) [13] and the method of Feldkamp, Davis, and Kress (FDK) [14]. The second approach is the *algebraic* (or *iterative*) one. Here, the reconstruction problem is cast as a system of linear equations,

$$\boldsymbol{W}\boldsymbol{x} = \boldsymbol{p}, \tag{1.1}$$

where $\boldsymbol{p} \in \mathbb{R}^m$ contains the projection data and $\boldsymbol{x} \in \mathbb{R}^n$ corresponds to the unknown image (both represented by a one-dimensional vector). The $m \times n$ matrix $\boldsymbol{W}$ is a linear operator that determines the projection geometry. Examples of such iterative algorithms are ART [15], SART [16], and SIRT [17]. SIRT, implemented as defined in [18], is used extensively throughout this thesis. A SIRT reconstruction of the phantom from Fig. 1.1a is shown in Fig. 1.1c.

Analytical methods such as FBP have the advantage that they are fast, but the disadvantage that they need a complete dataset to work correctly. The algebraic methods are much more computationally demanding, but they have the advantage that they are more flexible with regard to the geometry of the scanner and the possibility to exploit extra information that might be available, e.g., about the scanned object. This thesis builds on the algebraic reconstruction methods.

### 1.2.3 Segmentation

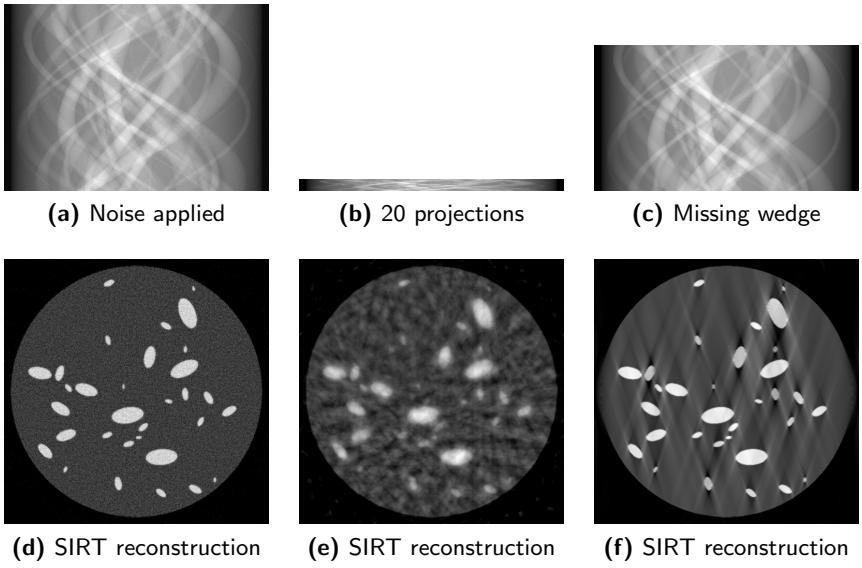
Segmentation is often the final step in the workflow of tomographic imaging. It amounts to the classification of image pixels into distinct classes, based on similarity with respect to some characteristic. Image segmentation is a well established field, and a range of methods has been developed, such as global or local thresholding, region growing, and clustering [19, 20]. Otsu's method [21] is one specific classical technique that is used in several of the experiments in this thesis. It uses the *histogram* of an image to compute optimal thresholds to segment that image into different classes.

For the purpose of this thesis, the segmentation classes are assumed to be directly related to the (modality specific) "density" of the parts of the scanned object.

## 1.3 Limited Data in Tomography

In ideal circumstances, the desirable quality of the reconstruction would determine which projections were made (their number, from which angles, etc.). However, in practical CT, this is not always possible. This gives rise to *limited data problems*.

There are several causes for having only limited data available, three of which are illustrated in Fig. 1.2. An important one is limiting the *radiation dose*. For example, X-rays can be harmful in medical and preclinical applications, and the X-ray dose should be as small as possible. However, the dose can also be a problem for inanimate objects, if those are affected by the radiation during the scan. A particular example of this is beam damage in electron tomography, where, especially for biological samples, the need to avoid radiation damage can strictly limit the acceptable dose. Lowering the radiation dose increases the noise level, since the *signal-to-noise ratio* (SNR) increases with the square root of the number of detected particles (which can be electrons, X-ray photons, etc.). The increased noise degrades the projections (Fig. 1.2a), which subsequently influences the resulting reconstruction (Fig. 1.2d).



**(a)** Noise applied **(b)** 20 projections **(c)** Missing wedge

**(d)** SIRT reconstruction **(e)** SIRT reconstruction **(f)** SIRT reconstruction

**Figure 1.2:** Sinograms (top row) and SIRT reconstructions (bottom row) for several limited data datasets. (a), (d) Noise was applied to the sinogram. (b), (e) The sinogram has only 20 projections. (c), (f) The sinogram has a missing wedge of data.

The imaging modality itself may also have inherent limitations. We take again the example of electron tomography. In that modality, the data is recorded one projection at a time, often with manual adjustment of the settings of the microscope between each projection. This makes creating an electron tomography dataset a very labor-intensive process. A somewhat extreme example of this is shown in Fig. 1.2b. With only 20 projections (spaced 9° apart) instead of 360, it is clear that this sinogram contains much less information. The effect on the SIRT

reconstruction is quite strong (Fig. 1.2e).

A second limitation of electron tomography is that the sample often cannot be tilted over the full 180° range, both because of physical limitations of the microscope and because most samples are flat, which means that their apparent thickness increases greatly at high angles of inclination. Not taking projections over the full 180° leads to a *missing wedge* of data. This is illustrated in Fig. 1.2c, which shows a sinogram with a missing wedge of 40°, while still consisting of equiangular projections spaced 0.5° apart. The only difference with Fig. 1.1b is that 80 projections are missing from Fig. 1.2c (40 lines at the top and 40 lines at the bottom). The effect on the SIRT reconstruction is again quite strong (Fig. 1.2f), even though the amount of data in the sinogram is not much smaller than in Fig. 1.1b.

From these examples, it is clear that having only limited data available can make it very difficult to create accurate reconstructions. This is particularly apparent from Figs. 1.2e and 1.2f. In the next Section, we describe two important approaches for solving the limited data problem, and several practical algorithms that implement them.

## 1.4 Current Approach to the Limited Data Problem

Current approaches to the limited data problem often involve including extra information about the scanned object in the reconstruction algorithm. Exploiting this *prior knowledge* can greatly enhance the reconstruction quality.

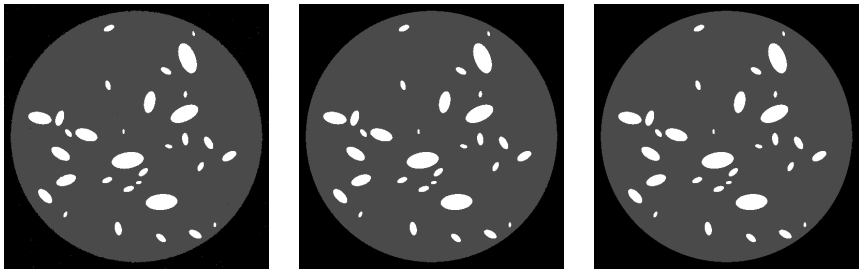
As an illustration of the concept of applying prior knowledge, we first look at the examples from Fig. 1.2 again. For the phantom that is reconstructed in Fig. 1.2, there are two obvious priors that can be applied. The object in the phantom image (Fig. 1.1a) clearly consists of homogeneous materials, both for the particles and for the disk. The black background of the image, which represents air or vacuum, is also homogeneous. Assuming that the gray levels of these three materials are known, leads to *Discrete Tomography* (DT), which is introduced in Section 1.4.1. If only the homogeneity is used as prior knowledge, then *Total Variation Minimization* (TVMin) is an appropriate technique. TVMin does not directly assume that the object consists of homogeneous regions, but it *is* very well suited to reconstruct those kinds of objects. TVMin is introduced in Section 1.4.2.

### 1.4.1 Discrete Tomography

There are two kinds of discrete tomography. In the first kind, two constraints of discreteness are combined. It is assumed that the scanned object has only a few different densities, *and* that it is on a grid. Assuming that the object is on a grid

is often not realistic, but it *is* applicable in practice, e.g., in atomic resolution tomography [22]. In the second kind of discrete tomography, the scanned object is still assumed to consist of only a few different densities, but it is no longer assumed to be on a grid. This allows reconstructing more general objects. It is this second kind of discrete tomography that is relevant for this thesis. A thorough review of the field of discrete tomography is available in the books by Herman and Kuba (Eds.) [23, 24].

In practice, the *Discrete Algebraic Reconstruction Technique* (DART) [25] is an efficient algorithm that has been used in fields such as electron tomography [26–29] and X-ray crystallography [30]. It is also being studied for use in the online scanning of food [31]. DART is introduced briefly in Chapter 4, and a detailed description of the algorithm is available in [25]. Fig. 1.3 shows DART reconstructions of the three sinograms from Fig. 1.2. It is clear that all three DART reconstructions are more accurate than the corresponding SIRT reconstructions from Fig. 1.2, showing that DART is able to exploit the available prior knowledge.



**(a)** Noise applied **(b)** 20 projections **(c)** Missing wedge

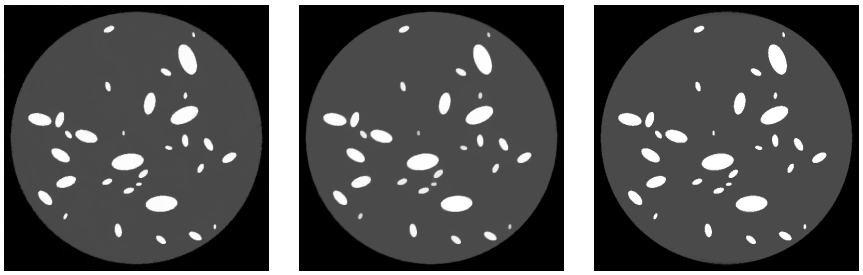
**Figure 1.3:** DART reconstructions of the sinograms from Fig. 1.2. (a) Noise was applied to the sinogram. (b) The sinogram has only 20 projections. (c) The sinogram has a missing wedge of data.

### 1.4.2 Total Variation Minimization

Total Variation Minimization (TVMin) was introduced in [32] as a noise removal algorithm. For TVMin, it is assumed that, out of all possible images that conform to the available data, images with a low total variation (TV) are preferable. The principle of TVMin is readily applicable to tomography.

In practice, both NESTA [33] and the Fast Iterative Shrinkage Thresholding Algorithm (FISTA) [34, 35] are relevant for this thesis. FISTA is introduced in some detail in Chapter 4. We refer to the mentioned articles for a detailed description of NESTA [33] and FISTA [34, 35]. As an aside, we note that NESTA is not an acronym, but short for *Nesterov's algorithm*. The authors of [33] named

it after Nesterov to acknowledge his work [36]. These algorithms have been used for tomography in practice, in fields such as electrical resistance tomography [37] and preclinical imaging [38]. Fig. 1.4 shows FISTA reconstructions of the three sinograms from Fig. 1.2. As for DART, it is clear that all three FISTA reconstructions are more accurate than the corresponding SIRT reconstructions from Fig. 1.2, showing that FISTA is also able to exploit the available prior knowledge.



**(a)** Noise applied **(b)** 20 projections **(c)** Missing wedge

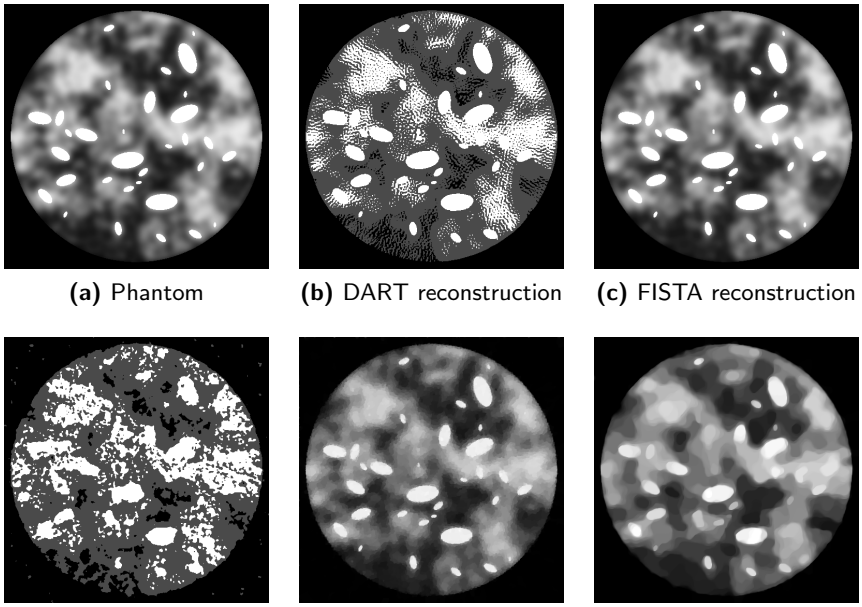
**Figure 1.4:** FISTA reconstructions of the sinograms from Fig. 1.2. (a) Noise was applied to the sinogram. (b) The sinogram has only 20 projections. (c) The sinogram has a missing wedge of data.

### 1.4.3 Limitations of the Current Approach

In Sections 1.4.1 and 1.4.2, we have shown examples of how DT and TVMin can greatly improve the quality of reconstructions from limited data. However, this is with the assumption that the applied prior knowledge is correct, and that it is correct for the *whole object*. In practice, this is not always the case.

This is illustrated in Fig. 1.5. The previous phantom (Fig. 1.1a) was adapted to have non-homogeneous supporting material (Fig. 1.5a). Algorithms from DT, such as DART, are unable to reconstruct non-homogeneous objects, since each pixel of the tomogram will be forcibly set to one of the "known" gray levels. The example for DART in Fig. 1.5b shows this clearly. Algorithms from TVMin are capable of reconstructing non-homogeneous objects if the spatial variation of the object is compatible with the TV prior, and if sufficient data is available. A FISTA reconstruction from 360 projections is quite accurate (Fig. 1.5c).

For the limited data case, for which DT and TVMin are most useful, the results for DART deteriorate further (Fig. 1.5d), while those for FISTA also become inaccurate (Figs. 1.5e and 1.5f). For DART, the white ellipses, which are still present in the reconstruction from 360 projections (Fig. 1.5b), are hardly recognizable in the reconstruction from 20 projections (Fig. 1.5d), due to the greatly decreased amount of data. For FISTA, both the reconstruction from the noisy



**(a)** Phantom **(b)** DART reconstruction **(c)** FISTA reconstruction

**(d)** DART, 20 projections **(e)** FISTA, noise applied **(f)** FISTA, 20 projections

**Figure 1.5:** (a) Phantom image. (b) DART reconstruction from 360 projections. (c) FISTA reconstruction from 360 projections. (d) DART reconstruction from 20 projections. (e) FISTA reconstruction from 360 projections, with noise applied to the sinogram. (f) FISTA reconstruction from 20 projections.

sinogram (Fig. 1.5e) and that from 20 projections (Fig. 1.5f) show different forms of the so-called *staircasing effect*, where the image becomes "cartoon-like" because piecewise constant images have a low TV. FISTA has an important parameter that provides a trade-off between fidelity to the projection data and low TV for the reconstruction. When only limited data is available, or when noise is present, this parameter has to be tuned towards preferring low TV to result in a reasonable reconstruction. The consequence is then that the image may become cartoon-like.

In summary, the major limitation of algorithms that exploit prior knowledge is that they may become ineffective if that knowledge is not valid throughout the scanned object, precluding the use of the prior knowledge. An additional problem is that it will not necessarily be obvious that the resulting reconstructions are incorrect. If DT or TVMin is applied to the reconstruction of objects that are expected to be homogeneous, any resulting smooth reconstruction might be deemed acceptable. In the next Section, we introduce the novel approach that this thesis proposes to solve these problems.

## 1.5 Novel Approach—Local Prior Knowledge

In this thesis, we present *local prior knowledge* as a way of limiting the application of the available prior knowledge to the parts of the scanned object for which it is valid.

In general, there appear to be two major ways to approach this problem. The first one is to develop *a specific new algorithm for each situation*. This allows exploiting exactly the available prior knowledge. With this approach, we have developed the PDART algorithm to reconstruct and segment dense homogeneous particles embedded in (possibly) non-homogeneous material.

The second approach is to *localize existing algorithms*, i.e., to adapt them to only apply their prior knowledge where this is appropriate. This approach needs a way of detecting where the prior knowledge is valid, for which we have developed the *reconstructed residual error*. We then describe a general way in which reconstruction algorithms can be localized, and apply this in practice by creating local variants of DART and FISTA.

The reconstructed residual error itself is also a major result of our work. The method visualizes and improves the quality of a segmented tomogram. We present the algorithm and several applications, in addition to the mentioned localized algorithms.

## 1.6 Outline of this Thesis

**Chapter 1** contains a general introduction.

**Chapter 2** presents the *Partially Discrete Algebraic Reconstruction Technique* (PDART), an algorithm that was developed for the specific task of reconstructing and segmenting dense homogeneous particles that are embedded in (possibly) non-homogeneous material. The algorithm is applied to two experimental datasets from electron tomography and, in Chapter 3, to a $\mu$CT dataset.

**Chapter 3** introduces the *reconstructed residual error*, an algorithm that can visualize and improve the quality of a segmented tomogram. Three applications are presented. First, it is used to improve gray level estimates. Second, for reconstruction algorithms that assume a priori knowledge of the gray levels, it is used to improve the segmentation itself. Finally, the reconstructed residual error is used to select the most accurate of several segmentations, for experimental $\mu$CT data.

**Chapter 4** presents another major application of the reconstructed residual error. It uses the algorithm to develop a general method for creating *localized* variants of existing reconstruction algorithms, which only apply their prior knowledge where this is appropriate. This general principle is then applied to DART and FISTA.

**Chapter 5** draws general conclusions.

# References

[1] J. Radon, "über die bestimmung von funktionen durch ihre intergralwerte längs gewisser mannigfaltigkeiten," *Berichte über die Verhandlungen der Sächsische Akademie der Wissenschaften*, vol. 69, pp. 262–277, 1917.

[2] ——, "On the determination of functions from their integral values along certain manifolds," *IEEE Trans. Med. Imag.*, vol. 5, no. 4, pp. 170–176, 1986.

[3] S. R. Deans, *The Radon Transform and Some of Its Applications*. New York: Wiley, 1983.

[4] R. N. Bracewell, "Strip integration in radio astronomy," *Aust. J. Phys.*, vol. 9, no. 2, pp. 198–217, 1956.

[5] D. J. de Rosier and K. A., "Reconstruction of three dimensional structures from electron micrographs," *Nature*, vol. 217, no. 5124, pp. 130–134, 1968.

[6] W. Hoppe, R. Langer, G. Knesch, and C. Poppe, "Protein-kristallstrukturanalyse mit elektronenstrahlen," *Naturwissenschaften*, vol. 55, no. 7, pp. 333–336, 1968.

[7] R. G. Hart, "Electron microscopy of unstained biological material: The polytropic montage," *Science*, vol. 159, no. 3822, pp. 1464–1467, 1968.

[8] A. M. Cormack, "Representation of a function by its line integrals, with some radiological applications," *J. Appl. Phys.*, vol. 34, no. 9, pp. 2722–2727, 1963.

[9] G. N. Hounsfield, "A method of and apparatus for examination of a body by radiation such as X or gamma radiation," Patent GB 1 283 915, 1972.

[10] ——, "Computerized transverse axial scanning (tomography): Part I. description of system," *Brit. J. Radiol.*, vol. 46, no. 552, pp. 1016–1022, 1973.

[11] P. A. Midgley and R. E. Dunin-Borkowski, "Electron tomography and holography in materials science," *Nature Mater.*, vol. 8, pp. 271–280, 2009.

[12] Z. V. Maizlin and P. M. Vos, "Do we really need to thank the beatles for the financing of the development of the computed tomography scanner?" *J. Comput. Assist. Tomogr.*, vol. 36, no. 2, pp. 161–164, 2012.

[13] T. M. Buzug, *Computed Tomography: From Photon Statistics to Modern Cone-Beam CT*. Berlin/Heidelberg: Springer-Verlag, 2008.

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

[15] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography," *J. Theor. Biol.*, vol. 29, no. 3, pp. 471–481, 1970.

[16] A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm," *Ultrason. Imag.*, vol. 6, pp. 81–94, 1984.

[17] P. Gilbert, "Iterative methods for the three-dimensional reconstruction of an object from projections," *J. Theor. Biol.*, vol. 36, no. 1, pp. 105–117, 1972.

[18] J. Gregor and T. Benson, "Computational analysis and improvement of SIRT," *IEEE Trans. Med. Imag.*, vol. 27, no. 7, pp. 918–924, 2008.

[19] N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern Recogn.*, vol. 26, no. 9, pp. 1277–1294, 1993.

[20] D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," *Annu. Rev. Biomed. Eng.*, vol. 2, pp. 315–337, 2000.

[21] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, 1979.

[22] S. Van Aert, K. J. Batenburg, M. D. Rossell, R. Erni, and G. Van Tendeloo, "Three-dimensional atomic imaging of crystalline nanoparticles," *Nature*, vol. 470, pp. 374–377, 2011.

[23] G. T. Herman and A. Kuba, *Discrete Tomography: Foundations, Algorithms, and Applications*. Boston: Birkhäuser, 1999.

[24] ——, *Advances in Discrete Tomography and Its Applications*. Boston: Birkhäuser, 2007.

[25] K. J. Batenburg and J. Sijbers, "DART: A practical reconstruction algorithm for discrete tomography," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2542–2553, 2011.

[26] S. Bals, K. J. Batenburg, J. Verbeeck, J. Sijbers, and G. Van Tendeloo, "Quantitative three-dimensional reconstruction of catalyst particles for bamboo-like carbon nanotubes," *Nano Lett.*, vol. 7, no. 12, pp. 3669–3674, 2007.

[27] S. Bals, K. J. Batenburg, D. Liang, O. Lebedev, G. Van Tendeloo, A. Aerts, J. A. Martens, and C. E. A. Kirschhock, "Quantitative three-dimensional modeling of zeotile through discrete electron tomography," *J. Am. Chem. Soc.*, vol. 131, no. 13, pp. 4769–4773, 2009.

[28] K. J. Batenburg, S. Bals, J. Sijbers, C. Kübel, P. A. Midgley, J. C. Hernandez, U. Kaiser, E. R. Encina, E. A. Coronado, and G. Van Tendeloo, "3D imaging of nanomaterials by discrete tomography," *Ultramicroscopy*, vol. 109, no. 6, pp. 730–740, 2009.

[29] B. Goris, T. Roelandts, K. J. Batenburg, H. Heidari Mezerji, and S. Bals, "Advanced reconstruction algorithms for electron tomography: From comparison to combination," *Ultramicroscopy*, vol. 127, pp. 40–47, 2013.

[30] K. J. Batenburg, J. Sijbers, H. F. Poulsen, and E. Knudsen, "DART: a robust algorithm for fast reconstruction of three-dimensional grain maps," *J. Appl. Cryst.*, vol. 43, no. 6, pp. 1464–1473, 2010.

[31] L. F. A. Pereira, T. Roelandts, and J. Sijbers, "Inline 3D X-ray inspection of food using discrete tomography," in *Proc. InsideFood Symposium*, Leuven, Belgium, 2013.

[32] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1–4, pp. 259–268, 1992.

[33] S. Becker, J. Bobin, and E. J. Candès, "NESTA: A fast and accurate first-order method for sparse recovery," *SIAM J. Imaging Sci.*, vol. 4, no. 1, pp. 1–39, 2011.

[34] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[35] ——, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, 2009.

[36] Y. Nesterov, "Smooth minimization of non-smooth functions," *Math. Program.*, vol. 103, no. 1, pp. 127–152, 2005.

[37] L. Zhang, H. Wang, and Y. Xu, "A shrinkage-thresholding method for the inverse problem of electrical resistance tomography," in *Proc. Instrumentation and Measurement Technology Conference (I2MTC)*, Graz, Austria, 2012, pp. 2425–2429.

[38] K. Wang, R. Su, A. A. Oraevsky, and M. A. Anastasio, "Investigation of iterative image reconstruction in three-dimensional optoacoustic tomography," *Phys. Med. Biol.*, vol. 57, no. 17, pp. 5399–5423, 2012.

# 2
# PDART

**This chapter has been published as**

T. Roelandts, K. J. Batenburg, E. Biermans, C. Kübel, S. Bals, and J. Sijbers, "Accurate segmentation of dense nanoparticles by partially discrete electron tomography," *Ultramicroscopy*, vol. 114, pp. 96–105, 2012.

**Abstract**—Accurate segmentation of nanoparticles within various matrix materials is a difficult problem in electron tomography. Due to artifacts related to image series acquisition and reconstruction, global thresholding of reconstructions computed by established algorithms, such as weighted backprojection or SIRT, may result in unreliable and subjective segmentations. In this paper, we introduce the *Partially Discrete Algebraic Reconstruction Technique* (PDART) for computing accurate segmentations of dense nanoparticles of constant composition. The particles are segmented directly by the reconstruction algorithm, while the surrounding regions are reconstructed using continuously varying gray levels. As no properties are assumed for the other compositions of the sample, the technique can be applied to any sample where dense nanoparticles must be segmented, regardless of the surrounding compositions. For both experimental and simulated data, it is shown that PDART yields significantly more accurate segmentations than those obtained by optimal global thresholding of the SIRT reconstruction.

## 2.1 Introduction

Electron tomography deals with the reconstruction of a three-dimensional (3D) representation of a microscopy sample from a tilt series of two-dimensional (2D) images. This technique has been applied successfully in materials science since the late 1980s [1]. Several imaging modes have been used for acquiring the projection

images, in particular bright-field TEM [2, 3], annular dark field TEM [4], high-angle annular dark-field scanning TEM (HAADF STEM) [5–9], and energy-filtered TEM (EFTEM) [10–13].

Quantitative interpretation of the reconstructed 3D volume is often hampered by the presence of *artifacts*: structured distortions that do not correspond with the actual sample. In particular, limits on the number of projection images imposed by sample contamination or beam damage give rise to such artifacts. Furthermore, the limited spacing for specimen holders in between the pole pieces of the objective lens often restricts the range of tilt angles to about $\pm 70°$, leading to a missing wedge in the collected data. As a consequence, features perpendicular to the electron beam are better resolved than features parallel to the beam, resulting in anisotropic resolution and distortions of the structure.

For many imaging tasks in materials science, the goal is to obtain an accurate *segmentation* of particular structures (i.e., particles, pores, tubules, etc.). Of particular importance is the problem of segmenting nanoparticles within various matrix materials [5–7, 14]. Due to artifacts related to image acquisition and reconstruction, segmenting these structures from gray level volumes computed by established algorithms, such as weighted backprojection (WBP) or SIRT [15], may result in unreliable and subjective segmentations. In practice, reconstructions are often segmented using a global threshold. Since the threshold is estimated visually, this approach is highly subjective. Moreover, it does not account for the effect that the intensity of the features in the reconstruction strongly depends on their size [16]. Fully manual segmentation may avoid this effect, but remains a time consuming and subjective approach.

Recently, discrete tomography algorithms have demonstrated the ability to overcome some of these limitations by exploiting prior knowledge. Discrete tomography is based on the assumption that the sample consists of only a few different compositions. Two rather different variants of discrete tomography have been applied to electron tomography. The first variant was recently applied to the reconstruction of crystalline nanoparticles at atomic resolution [17, 18]. For this variant, it is assumed that the crystal contains only a few atomic species, and that the atoms lie on a regular grid. Together, these assumptions allow to create a reconstruction from as few as two or three projections. For the second variant, which can be applied at lower resolutions, it is only assumed that the sample consists of a few different compositions, each corresponding to a particular gray level in the reconstructed image. The discrete tomography algorithms that appear in this paper are of the second variant.

Major advantages of discrete tomography algorithms are that they require fewer projection images compared to alternative methods such as SIRT, and that missing wedge artifacts are strongly reduced [19]. Moreover, as the final result of the

reconstruction process is a segmented image, a separate segmentation step is no longer required. The *Discrete Algebraic Reconstruction Technique* (DART) for discrete tomography has been successfully applied to a broad range of materials samples [20–23]. The main restriction for using discrete tomography is evidently that the entire sample must satisfy the discreteness requirement. If the sample contains a mixture of compositions, the results of discrete tomography cannot be relied upon, as the key assumptions are violated.

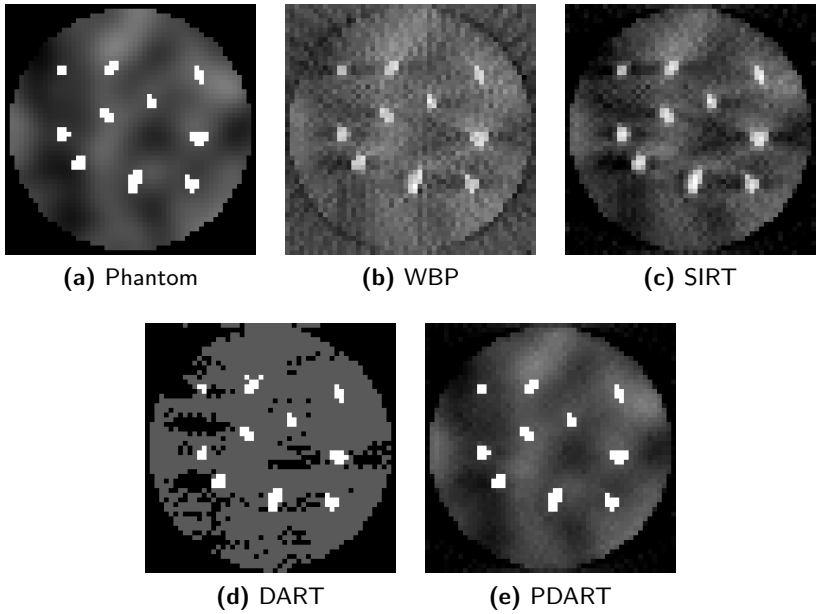
In this paper, we introduce the *Partially Discrete Algebraic Reconstruction Technique* (PDART) for computing accurate segmentations of dense nanoparticles of constant composition, regardless of the compositions in the remaining part of the sample. Embedded nanoparticles such as catalyst particles are often dense structures compared to their surroundings (e.g., porous materials), resulting in a high gray level in the reconstructed image. PDART is based on the assumption that the densest composition occurs in homogeneous regions that have a constant gray level. These dense regions are segmented discretely, while the surrounding regions are reconstructed using continuously varying gray levels. If the assumption of a homogeneous densest composition holds, the imaging mode that is used to record the tilt series—HAADF STEM for both samples in this paper—is not a restriction on the applicability of PDART, as long as the selected imaging mode is compatible with tomography. PDART imposes no restrictions on the nature of the sample (except that the densest composition must be homogeneous), which means that the application of the algorithm is not restricted to any specific type of samples.

This paper is structured as follows. In Section 2.2, the problem of segmenting dense particles is introduced, and the PDART algorithm is defined. Section 2.2 also introduces the figure of merit that is used for quantitative evaluation of the results. It concludes by describing how the parameters of the algorithm can be optimized automatically. In Section 2.3, the capabilities of PDART are assessed using two different experimental datasets and a number of simulation experiments. The results are discussed in Section 2.4 and conclusions are drawn in Section 2.5.

## 2.2 Algorithm

Before describing the PDART algorithm, we start by giving an example of its applicability. Fig. 2.1 illustrates the problem of nanoparticle segmentation. Fig. 2.1a shows a *phantom* (i.e., a simulated image), representing a microscopy sample that contains nanoparticles of only a few pixels each, embedded in a cylinder of varying composition. From this phantom, a synthetic dataset was created by calculating 28 evenly spaced projections in the range of $\pm 70°$. Fig. 2.1 also shows WBP (Fig. 2.1b), SIRT (Fig. 2.1c), DART (Fig. 2.1d), and PDART (Fig. 2.1e) recon-

structions of this dataset.



**(a)** Phantom **(b)** WBP **(c)** SIRT

**(d)** DART **(e)** PDART

**Figure 2.1:** A simulation phantom and several reconstructions. The phantom represents a cylindrical sample that contains nanoparticles of only a few pixels each, embedded in a material of varying composition.

The gray level reconstructions computed by WBP and SIRT have limited visual quality, as a result of the small number of projection angles and their limited angular range. When thresholding these images to determine the size and shape of the particles, it is not clear how the threshold should be chosen in an optimal way. The DART reconstruction shown in Fig. 2.1d is already segmented, yet the segmentation is not accurate at all when compared to the original phantom. The varying composition of the disk surrounding the nanoparticles violates the key discreteness assumption imposed by the DART algorithm. The PDART reconstruction, shown in Fig 2.1e, seems much more accurate than the other reconstructions.

### 2.2.1 Algorithm Description

The PDART algorithm has been designed to allow for accurate particle segmentation in cases where neither continuous methods nor fully discrete tomography leads to good results. The algorithm is based on the assumption that the particles have a constant composition, and that this composition represents the highest gray

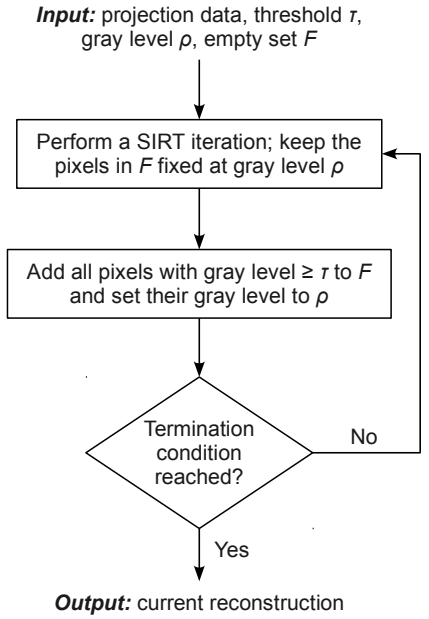



**Figure 2.2:** Flowchart of the PDART algorithm.

level in the reconstructed volume. PDART combines an iterative reconstruction algorithm, such as SIRT, with intermediate segmentation steps. Once pixels have been identified as "particle", they are directly segmented (i.e., their value is set to the constant gray level for the particles) and kept fixed at this value in subsequent SIRT iterations. We note that, throughout this paper, we use the additive variant of SIRT, as described in [24].

Fig. 2.2 shows a flowchart of the PDART algorithm. Besides having the projection data as input, the algorithm has two parameters: a threshold $\tau$ and a gray level $\rho > \tau$, which corresponds to the gray level of the particles. Optimal values for both parameters can be determined automatically, as is outlined in Section 2.2.3.

Initially, the set $F$ of *fixed pixels* is empty. In an iterative loop, the algorithm starts by performing one or more SIRT iterations on the entire image volume. Whenever one or more pixels are assigned a higher gray level than the threshold $\tau$, it is decided that these pixels belong to a particle. Such pixels are added to $F$: their gray level is set to $\rho$ and is kept fixed at this value during all subsequent SIRT iterations. In this way, the set $F$ gradually expands as pixels are added, until some termination condition is satisfied. Typically, one aims for terminating the algorithm when no new pixels have been added to $F$ for a sufficiently large number of iterations.

In its original form, the SIRT algorithm computes a weighted least square solution of the system $\boldsymbol{W}\boldsymbol{x} = \boldsymbol{p}$, where $\boldsymbol{x}$ denotes the unknown image, $\boldsymbol{p}$ denotes the projection data, and $\boldsymbol{W}$ denotes the projection operator. As pixels are added to $F$ within the PDART algorithm, the number of unknowns in this system is reduced while the number of equations remains the same, thereby improving the reconstruction for pixels that do not (yet) belong to $F$.

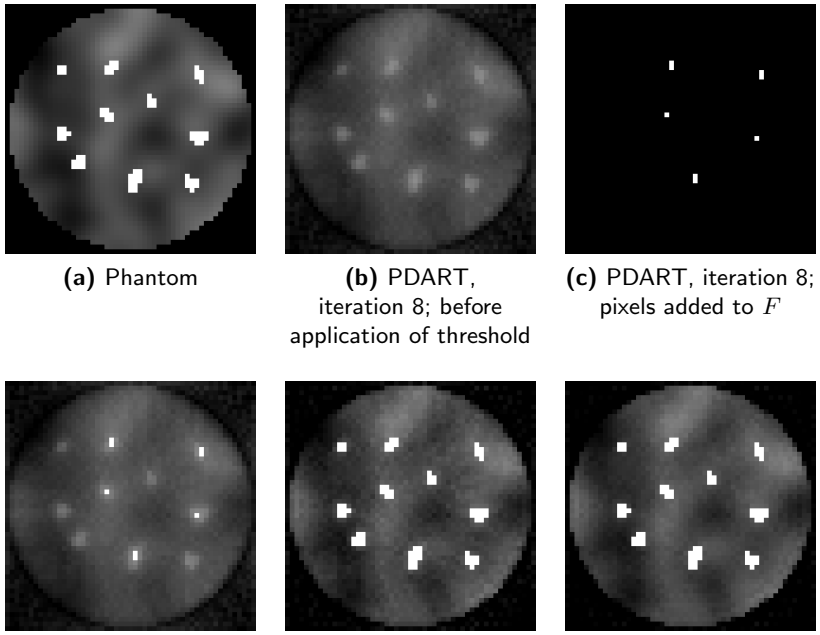
As an illustration, Fig. 2.3 shows several intermediate steps of the PDART algorithm, using the phantom from Fig. 2.1a. In this particular example, the first pixels cross the threshold $\tau$ at the eighth iteration (Fig. 2.3b). At this point, the result is still exactly the same as for SIRT, since no pixels have been fixed yet. The pixels that crossed the threshold are then added to the set $F$ of fixed pixels (Fig. 2.3c) and fixed for the remainder of the reconstruction process (Fig. 2.3d). During the following iterations, the process of discovering and fixing more and more particle pixels continues until all of them are found at, in this example, iteration 57 (Fig. 2.3e). No new pixels are found during the following iterations, although the reconstruction quality of the background keeps improving somewhat. The result after 150 iterations is shown in Fig. 2.3f.

### 2.2.2 Figure of Merit

After computing a reconstruction, the projections of the reconstructed image can be computed, and subsequently compared with the measured projections. The difference between the computed projections and the measured dataset is calculated by taking the sum of squares of the differences for all projection pixels, resulting in a number that indicates how well the reconstruction adheres to the measured projection data. Mathematically, this is known as the *projection distance*, defined as $d_{\mathrm{pr}}(\boldsymbol{x}) = \|\boldsymbol{W}\boldsymbol{x} - \boldsymbol{p}\|_2$. It is this projection distance that we use as a figure of merit. For phantoms, we can also calculate the *phantom distance*, which is defined as $d_{\mathrm{ph}}(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{h}\|_2$, where $\boldsymbol{h}$ denotes the phantom image. The phantom distance directly measures the difference between the reconstruction and the phantom.

### 2.2.3 Parameter Optimization

In the flowchart of Fig. 2.2, the threshold $\tau$ and the gray level $\rho$ are assumed to be known in advance. Their values may be set manually. A good value for $\rho$ can be determined by calculating the average value within one or more particles in a SIRT reconstruction. For $\tau$, a value that is somewhat higher than the highest gray level in the background material generally leads to accurate reconstructions. However, although this manual procedure may lead to satisfactory results, it is also subjective.



**(a)** Phantom

**(b)** PDART, iteration 8; before application of threshold

**(c)** PDART, iteration 8; pixels added to $F$

**(d)** PDART, iteration 8; after application of threshold

**(e)** PDART, iteration 57

**(f)** PDART, iteration 150

**Figure 2.3:** Illustration of several stages of the PDART algorithm.

A more objective way to determine optimal values for $\tau$ and $\rho$ is to search for those values that result in a reconstruction that corresponds maximally to the measured tilt series. After computing the PDART reconstruction for particular values $(\tau, \rho)$, the projection distance can be computed, resulting in a number that indicates the quality of a particular pair $(\tau, \rho)$. By optimizing the reconstruction quality over the space of possible values for $\tau$ and $\rho$, their optimal values can be determined. In the case studies that follow, this optimization was performed by applying an unconstrained nonlinear optimization, using the derivation-free Nelder-Mead simplex algorithm [25]. This procedure needs in the order of 100 reconstructions to reach a precision of three significant digits for the threshold and gray level.

## 2.3 Experiments and Results

In this section, we report on a series of experiments that were performed to assess the capabilities of the PDART algorithm. The experiments were conducted based on phantom objects, as well as experimental electron tomography datasets of two different samples.

The phantom study in Section 2.3.1 illustrates the basic differences between PDART and DART, and establishes that PDART can be an alternative for DART in cases where that algorithm is not applicable.

In Sections 2.3.2 and 2.3.3, PDART is applied to two different experimental datasets, to further investigate the properties of the algorithm. The reconstructions from the experimental datasets also show that PDART is applicable in practice.

The first experimental sample is a heterogeneous catalyst, consisting of metal nanoparticles on a mesoporous silica support, acquired using an angular range of $-70°$ to $+72°$. The second sample consists of Pb nano inclusions in a crystalline Si matrix, acquired using an On-Axis Rotation Tomography Holder, allowing for image acquisition over the full angular range.

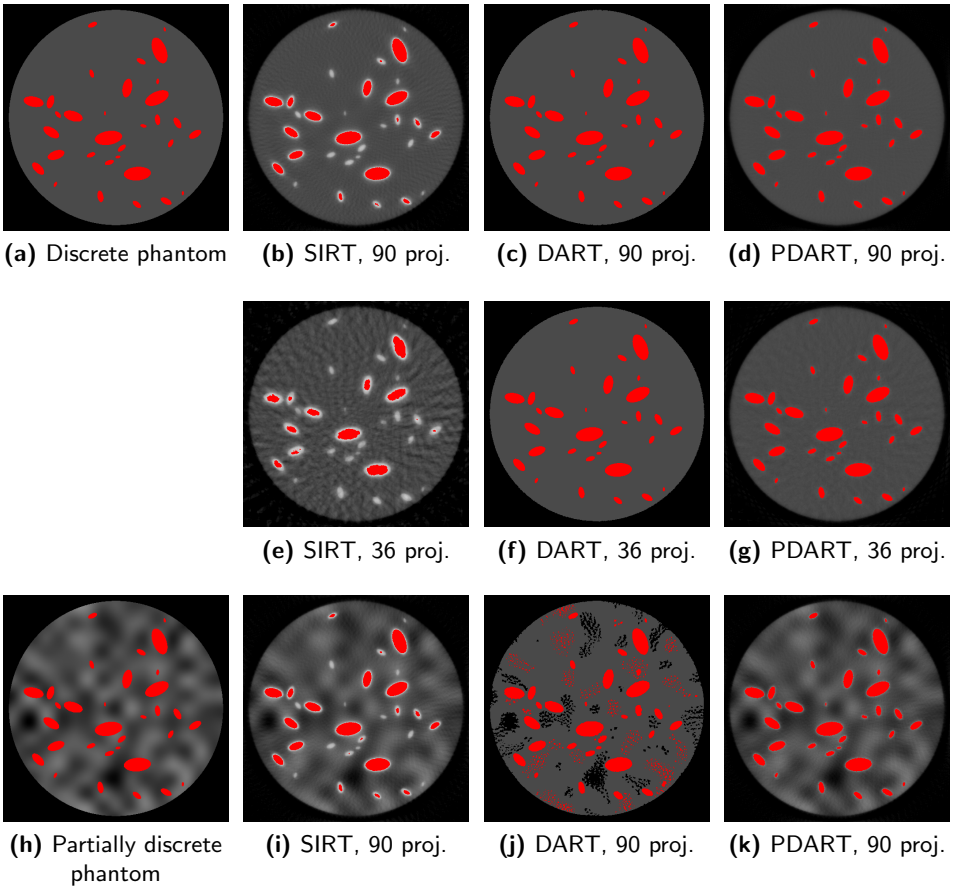
To validate the reconstruction results for both datasets, simulation phantoms that resemble the experimental sample were designed. For these phantoms, reconstructions with known ground truth were computed, thereby allowing a direct comparison with the original object.

Throughout this section, the results for PDART are compared with SIRT and, for the phantom study, with DART. Since both PDART and DART result in reconstructions that are already segmented, the results from those algorithms are compared with segmented SIRT reconstructions. The segmentation was performed by taking a pair $(\tau, \rho)$, and setting all pixels with a value that exceeds $\tau$ to $\rho$. The optimal pair $(\tau, \rho)$ was determined by minimizing the projection distance, using the optimization procedure from Section 2.2.3.

### 2.3.1 Phantom Study

For the phantom study, two phantoms containing dense elliptical particles were designed. For the first phantom (Fig. 2.4a), the particles are embedded in a homogeneous material; for the second one (Fig. 2.4h), in a continuously varying material. The gray value of the homogeneous background from Fig. 2.4a is the mean of the gray values of the background from Fig. 2.4h. The dimensions of both phantoms are $512 \times 512$ pixels.

From the discrete phantom (Fig. 2.4a), a synthetic dataset using 90 projections at evenly spaced $2°$ intervals was created. The resulting reconstructions are shown in the top row of Fig. 2.4. Visually, the results for DART (Fig. 2.4c) and



**(a)** Discrete phantom **(b)** SIRT, 90 proj. **(c)** DART, 90 proj. **(d)** PDART, 90 proj.

**(e)** SIRT, 36 proj. **(f)** DART, 36 proj. **(g)** PDART, 36 proj.

**(h)** Partially discrete phantom **(i)** SIRT, 90 proj. **(j)** DART, 90 proj. **(k)** PDART, 90 proj.

**Figure 2.4:** A discrete and a partially discrete phantom and several reconstructions. The first two rows show the discrete phantom and reconstructions, using 90 and 36 projections for the first and second row, respectively. The third row shows the partially discrete phantom and reconstructions, using 90 projections.

PDART (Fig. 2.4d) seem to be of comparable quality. In the SIRT reconstruction (Fig. 2.4b) the size of the particles seems to be underestimated, and reconstruction artifacts are visible in the background material. Table 2.1 (rows 1–4) shows the numerical results for this experiment. In Table 2.1, the values for $\tau$ and $\rho$ that yield the minimum projection distance $d_{\mathrm{pr}}$ are shown, together with the phantom distance $d_{\mathrm{ph}}$ for the same pair $(\tau, \rho)$. The numerical results from Table 2.1 (rows 1–4) confirm the visual assessment, but indicate that DART adheres better to the projection data than PDART for this fully discrete dataset (the value for $d_{\mathrm{pr}}$ is lower), even though this is not clear from the visual appearance of the reconstruction. This is confirmed by the phantom distance $d_{\mathrm{ph}}$.

| Phantom | Algorithm | $\tau$ | $\rho$ | $d_{\mathrm{pr}}$ | $d_{\mathrm{ph}}$ |
|---|---|---|---|---|---|
| Discrete, | | N/A | 1 | 0 | 0 |
| 90 proj. | SIRT | 0.772 | 0.928 | 641 | 32.0 |
| | DART | 0.710 | 1.00 | 130 | 8.54 |
| | PDART | 0.368 | 0.99 | 161 | 9.79 |
| Discrete, | SIRT | 0.843 | 0.977 | 334 | 38.6 |
| 36 proj. | DART | 0.608 | 1.00 | 42.4 | 5.55 |
| | PDART | 0.368 | 1.00 | 103 | 15.8 |
| Partially | | N/A | 1 | 0 | 0 |
| discrete, | SIRT | 0.766 | 0.926 | 654 | 31.9 |
| 90 proj. | DART | 0.516 | 0.964 | 702 | 54.9 |
| | PDART | 0.494 | 1.00 | 266 | 23.0 |

**Table 2.1:** Optimal values for thresholds and gray levels, and the corresponding projection and phantom distances.

From the same discrete phantom of Fig. 2.4a, a second synthetic dataset using 36 projections at evenly spaced 5° intervals was created. The resulting reconstructions are shown in the middle row of Fig. 2.4. The DART reconstruction (Fig. 2.4f) is virtually identical to that from Fig. 2.4c, even though the number of projections was reduced from 90 to 36. The quality of the PDART reconstruction (Fig. 2.4g) has decreased somewhat, mainly in the background material. This effect is more obvious for SIRT (Fig. 2.4e), as is the underestimation of the particle sizes for that algorithm. The numerical results in Table 2.1 (rows 5–7) confirm this, and show that the difference between DART and PDART has increased, both for $d_{\mathrm{pr}}$ and for $d_{\mathrm{ph}}$. The results for both synthetic datasets based on the discrete phantom suggest that, for a fully discrete sample, DART should be the algorithm of choice.

From the partially discrete phantom (Fig. 2.4h), a third synthetic dataset using

90 projections at evenly spaced 2° intervals was created. The resulting reconstructions are shown in the last row of Fig. 2.4. For this third dataset, the results are markedly different. It is clear that the quality of the DART reconstruction (Fig. 2.4j) is quite bad. The reason that DART fails is that the model that is imposed by DART, namely that the sample is fully discrete, does not apply. The model that PDART (Fig. 2.4k) assumes is correct: the dense particles are discrete, and they are embedded in a continuously varying material. The visual results are confirmed by the numerical results in Table 2.1 (rows 8–11). Both $d_{\mathrm{pr}}$ and $d_{\mathrm{ph}}$ are now higher for DART than for SIRT (Fig. 2.4i), while the values for PDART are still lower, as for the preceding experiments. This implies that PDART should be the algorithm of choice for a partially discrete sample that contains homogeneous dense particles.

### 2.3.2 Sample I: Heterogeneous Catalyst

The sample is a heterogeneous catalyst, consisting of metal nanoparticles on a mesoporous silica support (called Catalyst hereafter) [16].

An HAADF STEM tilt series was acquired using an FEI Tecnai F20 ST microscope operated at an acceleration voltage of 200 kV. The sample was mounted on a Fischione Model 2020 Advanced Tomography Holder. The series was recorded using 2° tilt angle increments over a range of −70° to +72°. Fig. 2.5 shows a single projection image from the tilt series. The size of the projection images is $1280 \times 1280$ pixels. To increase the SNR, the projection images were downsampled by a factor of 2 in both dimensions.

In principle, the Catalyst dataset appears to be suitable for a fully discrete DART reconstruction, as it has just two compositions. However, as Fig. 2.6 shows, it was not possible to reconstruct the supporting particle as a uniform structure using DART. The fine porous structure of the supporting particle results in heavy partial volume effects, such that the particle cannot be properly represented by a constant gray level. As a result of the mismatch within the support, the segmentation of the catalyst particles also degrades. If one would only assume discreteness of the catalyst particles, the support would be reconstructed using continuous gray levels, thereby mitigating this problem. We therefore expect that PDART will be more suitable.

The dataset was reconstructed in 3D using both SIRT and PDART. Optimal parameters $\tau$ and $\rho$ were determined from the projection data, using the procedure outlined in Section 2.2.3. Both algorithms were run for 100 iterations. For experimental samples, the effect of noise has to be taken into account when determining the optimal number of iterations. This is due to the well known effect that for iterative algorithms, such as SIRT, the influence of noise in the projection data starts dominating the reconstruction after a certain number of iterations (known as

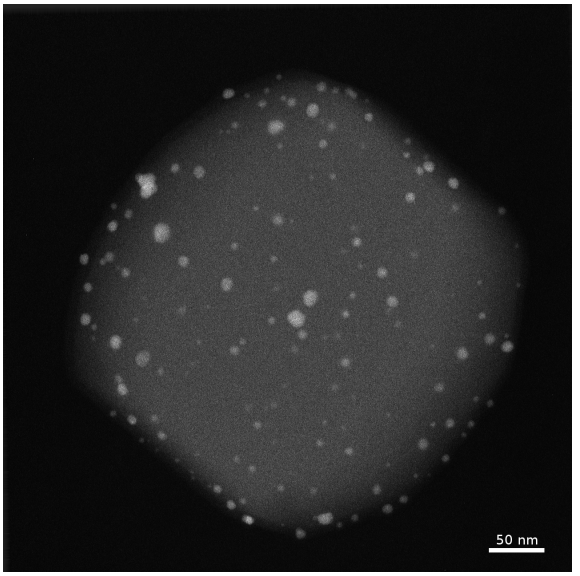



**Figure 2.5:** HAADF STEM projection image from the tilt series of the Catalyst sample.

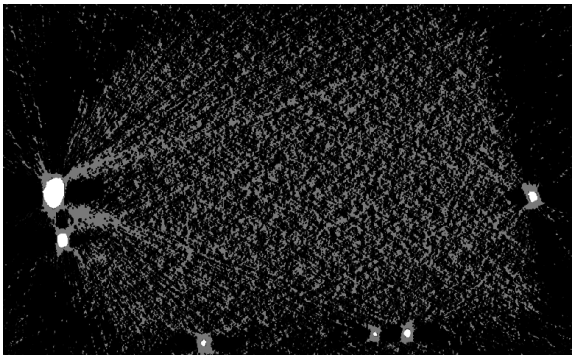
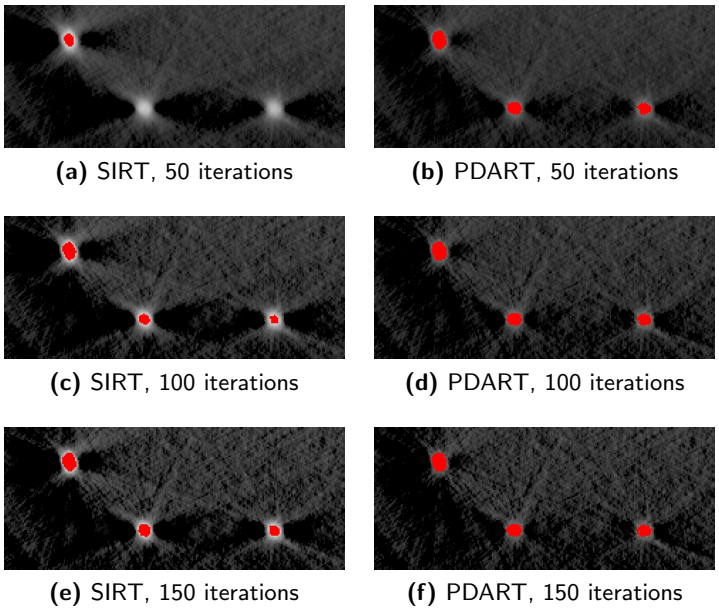


**Figure 2.6:** DART reconstruction of a slice of the Catalyst dataset in the xz-plane. The brightness of the support material was increased for clarity.

semi-convergence). This effect carries over to PDART, since it uses SIRT. Fig. 2.7 shows that the number of iterations is less important for PDART than it is for SIRT. For PDART at 50 iterations (Fig. 2.7b), there is still a hint of the typical blur that surrounds the particles in the PDART reconstruction during the early stages of reconstruction (also see Fig. 2.3d for an extreme example of this effect). At 100 iterations (Fig. 2.7d), this effect has greatly diminished. The difference

with the reconstruction at 150 iterations is small, although the effect of the noise starts to become apparent.



**(a)** SIRT, 50 iterations **(b)** PDART, 50 iterations

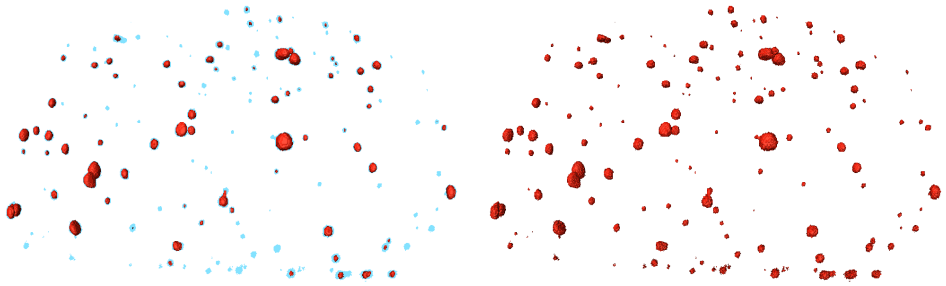
**(c)** SIRT, 100 iterations **(d)** PDART, 100 iterations

**(e)** SIRT, 150 iterations **(f)** PDART, 150 iterations

**Figure 2.7:** Detail of the reconstruction of the Catalyst dataset at different numbers of iterations. The brightness of the support material was increased for clarity.

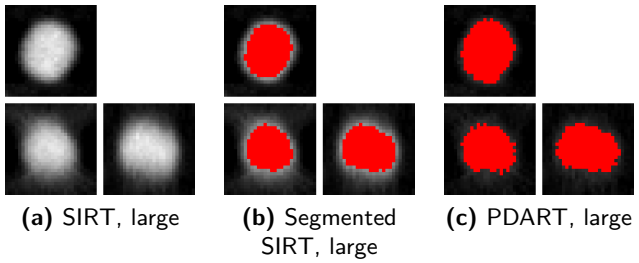
Fig. 2.8 shows isosurface renderings of the catalyst particles in the resulting reconstructions. From Fig. 2.8, it can be observed that the thresholded SIRT reconstruction is lacking many of the small particles. Fig. 2.9 shows a set of 2D slices through the reconstructed volume, both for a large catalyst particle (top row), and for a small one (bottom row). The same thresholds were used as for the isosurface renderings from Fig. 2.8. From the middle column of Fig. 2.9, it can be clearly observed that the threshold for SIRT, which appears suitable for the large particle, is not at all suitable for segmenting the small particle. As a consequence, no single threshold can be found for which the entire volume is segmented with reasonable accuracy. On the other hand, the PDART reconstructions for both particles, which were also computed using a single pair $(\tau, \rho)$, do not suffer from this problem.

To validate the results for the experimental dataset, a detailed simulation of a mesoporous support particle with catalyst particles on its surface was performed in cooperation with the Fraunhofer ITWM, Germany. The dimensions of the phantom are $548 \times 548 \times 325$ pixels. From the phantom, a synthetic dataset was created,

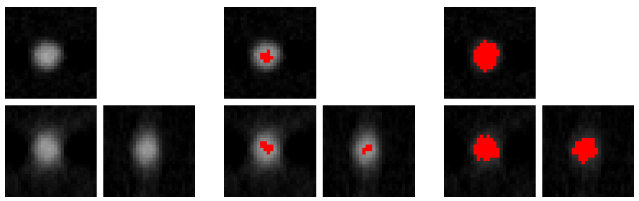


**(a)** SIRT **(b)** PDART

**Figure 2.8:** Isosurface rendering of the SIRT and PDART reconstructions of the Catalyst sample. The PDART reconstruction is shown behind the SIRT reconstruction for comparison.



**(a)** SIRT, large **(b)** Segmented SIRT, large **(c)** PDART, large



**(d)** SIRT, small **(e)** Segmented SIRT, small **(f)** PDART, small

**Figure 2.9:** Three orthogonal slices through a large catalyst particle (top row) and a small particle (bottom row) in the reconstruction of the Catalyst sample. The same threshold was used for the large and small particle.

consisting of 72 projections at evenly spaced 2° intervals, producing the same 38° missing wedge as in the original tilt series of the Catalyst dataset. Poisson noise was applied to the simulated projection images. This setup resembles the conditions under which the experimental dataset was recorded. The SIRT and PDART algorithms were both run for 100 iterations. The segmented volumes for the phantom image and for both algorithms are shown in Fig. 2.10. In accordance with the

reconstructions from experimental data, the segmentation computed by PDART contains particles of all sizes, whereas from the segmented SIRT reconstruction a number of small particles is missing.

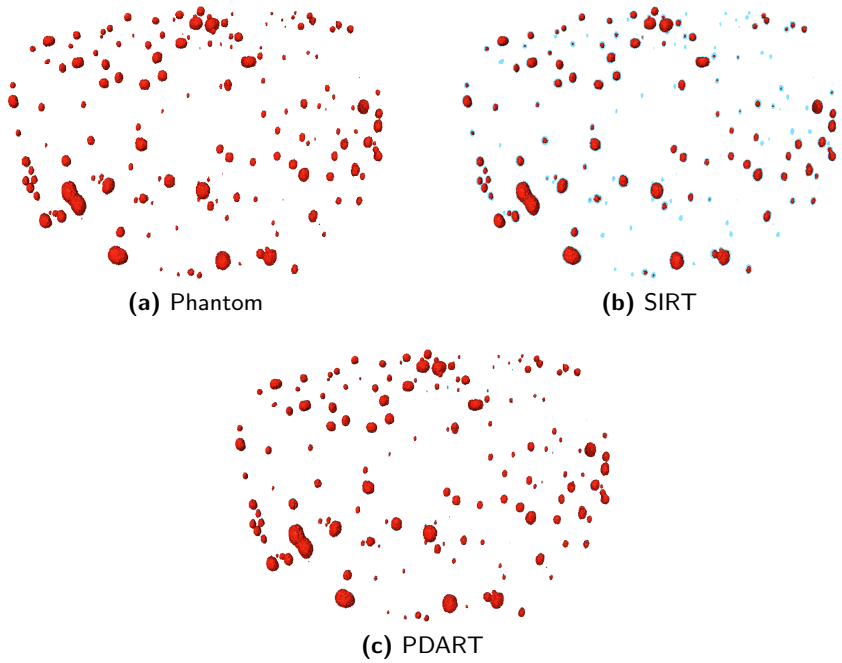



**Figure 2.10:** Isosurface rendering of the Catalyst phantom and reconstructions for SIRT and PDART. The phantom is shown behind the SIRT and PDART reconstructions for comparison.

The numerical results for the reconstruction quality are summarized in Table 2.2. The projection distance $d_{\mathrm{pr}}$ for PDART is lower than for the other reconstructions, which means that it corresponds more accurately to the projection data. This suggests that the value for $\rho$ that was found by PDART is also closer to the true value than the value that was found by SIRT. For the phantom, the numbers for $d_{\mathrm{ph}}$ confirm the results for $d_{\mathrm{pr}}$. Moreover, the value for $\rho$ is indeed closer to the correct value of 1.

### 2.3.3 Sample II: Pb in Si Inclusions

The second sample consists of Pb nano inclusions in a crystalline Si matrix (called Pb–Si hereafter). A micro-pillar sample was prepared by FIB milling, using an FEI Nova Nanolab 200 DualBeam system. The procedure to prepare these dedicated micro-pillar samples is explained in more detail in [26].

| Sample | Algorithm | $\tau$ | $\rho$ | $d_{\mathrm{pr}}$ | $d_{\mathrm{ph}}$ |
|---|---|---|---|---|---|
| Catalyst | SIRT | 1.16 | 1.81 | 10259 | N/A |
| | PDART | 0.403 | 2.03 | 10018 | N/A |
| Phantom | | N/A | 1 | 0 | 0 |
| | SIRT | 0.531 | 0.828 | 3311 | 210 |
| | PDART | 0.164 | 0.977 | 2880 | 149 |

**Table 2.2:** Optimal values for thresholds and gray levels, and the corresponding projection and phantom distances.

The sample was mounted on a Fischione Model 2050 On-Axis Rotation Tomography Holder, which allowed to acquire a series with a tilt range of $\pm 90°$. An HAADF STEM tilt series was recorded using 2° tilt angle increments, using a JEOL JEM-3000F microscope operated at an acceleration voltage of 300 kV. Fig. 2.11 shows a single projection image from the tilt series. The size of the projection images is $512 \times 512$ pixels.

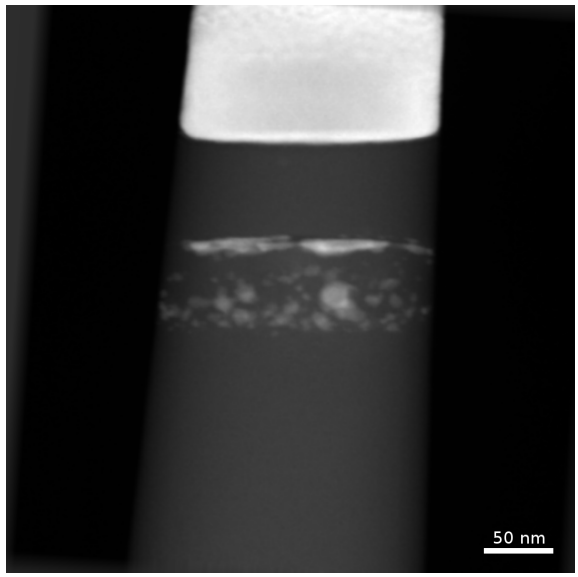



**Figure 2.11:** HAADF STEM projection image from the tilt series of the Pb–Si sample, showing the needle shaped structure of the sample.

As for the Catalyst dataset, the Pb–Si dataset appears to be suitable for a fully discrete DART reconstruction, as it has just two compositions. However, Fig. 2.12

demonstrates that also in this case discrete tomography failed to reconstruct the matrix material as a uniform structure. Artifacts related to the inherent difficulty of aligning the projections of micro-pillar samples have as a result that the Si matrix cannot be properly represented by a constant gray level. If we assume that only the Pb particles are discrete, we can avoid degrading the accuracy of the reconstruction of those particles by the artifacts in the Si matrix. Hence, PDART is again expected to be more suitable.

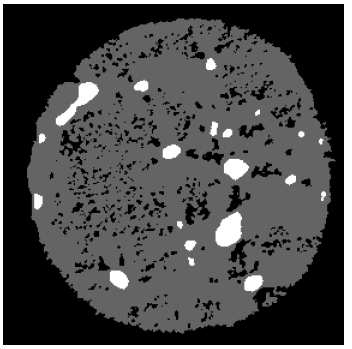
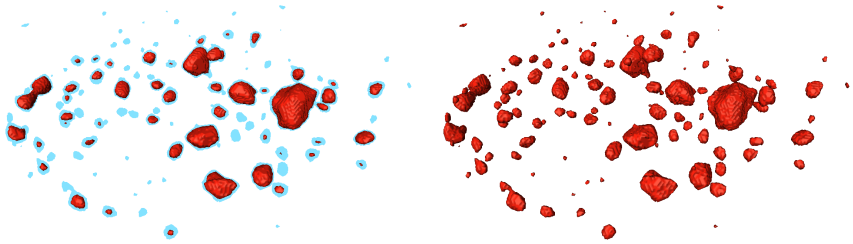


**Figure 2.12:** DART reconstruction of a slice of the Pb–Si dataset in the xz-plane. The brightness of the Si matrix material was increased for clarity.

Using 70 iterations for both SIRT and PDART, the experimental dataset was reconstructed in 3D. Stopping the reconstruction at 70 iterations avoided emphasizing the noise, while producing clearly delineated particles. As before, optimal parameters $\tau$ and $\rho$ were determined from the projection data, using the procedure outlined in Section 2.2.3. Fig. 2.13 shows isosurface renderings of the Pb particles in the resulting reconstructions. It is apparent that the segmented SIRT reconstruction is lacking many of the small particles. This is confirmed by Fig. 2.14, which shows three orthogonal slices through the reconstruction.

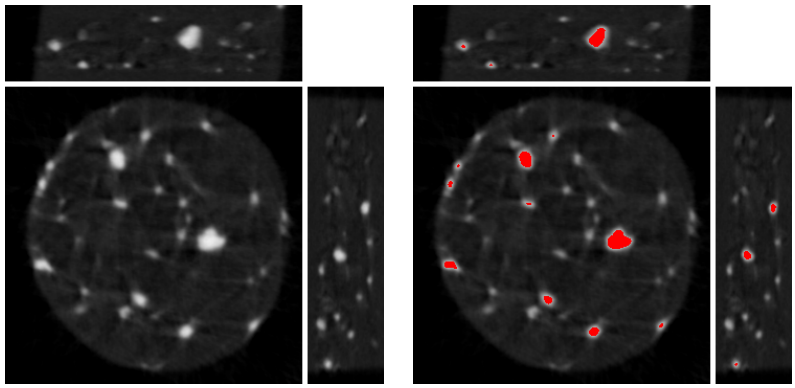
The results for this experimental sample were validated using a 3D phantom modeled after the original Pb–Si dataset. The dimensions of the phantom are $332 \times 332 \times 73$ pixels. From the phantom, a synthetic dataset was created, consisting of 90 projections at evenly spaced $2^\circ$ intervals. Poisson noise was applied to the simulated projection images. This setup resembles the conditions under which the experimental dataset was recorded. The SIRT and PDART algorithm were both run for 70 iterations. The segmented volumes for the phantom image and for both algorithms are shown in Fig. 2.15. As for the reconstructions from experimental data, the smaller particles seem to be missing from the segmented SIRT reconstruction.

The numerical results for the reconstruction quality are summarized in Ta-

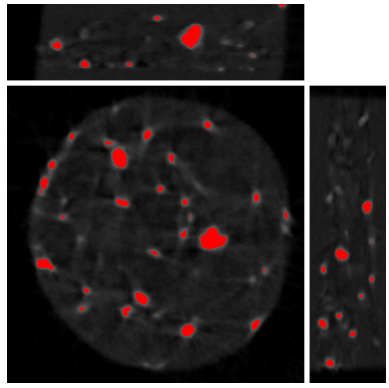


**(a)** SIRT **(b)** PDART

**Figure 2.13:** Isosurface rendering of the SIRT and PDART reconstructions of the Pb–Si sample. The PDART reconstruction is shown behind the SIRT reconstruction for comparison.

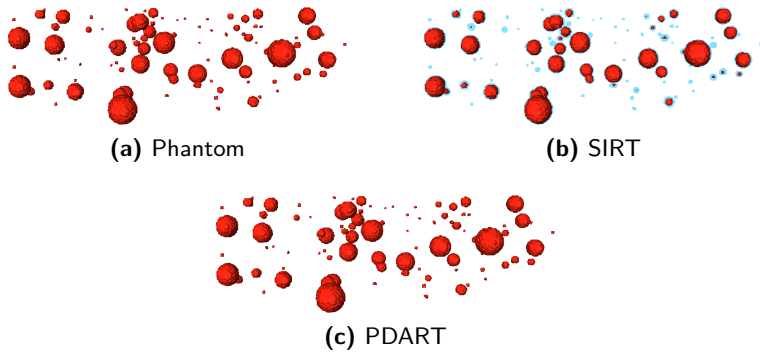


**(a)** SIRT **(b)** Segmented SIRT



**(c)** PDART

**Figure 2.14:** Three orthogonal slices through the reconstructions of the Pb–Si sample for SIRT, segmented SIRT and PDART; xz-plane (squares), xy-plane (horizontal rectangles), yz-plane (vertical rectangles).



**(a)** Phantom **(b)** SIRT **(c)** PDART

**Figure 2.15:** Isosurface rendering of the Pb–Si phantom and reconstructions for SIRT and PDART. The phantom is shown behind the SIRT and PDART reconstructions for comparison.

ble 2.3. As for the Catalyst dataset, the projection distance for PDART is lower than for the other reconstructions, which means that it has closer correspondence to the projection data. This again suggests that the value for $\rho$ that was found by PDART is closer to the true value than the value that was found by SIRT.

| Sample | Algorithm | $\tau$ | $\rho$ | $d_{\mathrm{pr}}$ | $d_{\mathrm{ph}}$ |
|---|---|---|---|---|---|
| Pb–Si | SIRT | 0.631 | 0.829 | 1832 | N/A |
| | PDART | 0.347 | 0.855 | 1704 | N/A |
| Phantom | | N/A | 1 | 0 | 0 |
| | SIRT | 0.653 | 0.925 | 1157 | 74.8 |
| | PDART | 0.221 | 0.989 | 751 | 30.5 |

**Table 2.3:** Optimal values for thresholds and gray levels, and the corresponding projection and phantom distances.

## 2.4 Discussion

The case studies in this paper show two substantially different experimental datasets that could not be reconstructed successfully using discrete tomography. Since they both contain dense particles as their main composition of interest, they are suitable for application of the PDART algorithm. The algorithm is straightforward to implement, which makes it applicable in practice. The results from PDART seem to improve on the segmentation of SIRT reconstructions that were created using a threshold that was optimized to match the projection data.

In both studies, the particles are accurately segmented, regardless of their size. The reconstructions of the different samples and phantoms also show that the parameters of the method can be optimized in an objective manner. This makes the algorithm an alternative to manual segmentation, since, even if a comparable segmentation is created manually, there will always be subjective judgment involved.

The results for the experimental datasets, together with the phantom study, suggest that PDART is a useful alternative to DART, since the algorithm allows samples for which a fully discrete reconstruction is not possible to still benefit from the techniques of discrete tomography. The four algorithms that were applied in this paper are shown in an algorithm selection tree in Fig. 2.16. If a full set of projections is available, or if computing resources are limited, WBP can still be the algorithm of choice. However, if sufficient computing resources are available, iterative techniques have compelling advantages, like a lower sensitivity to noise. Another important advantage is that they allow to exploit prior knowledge, which leads to (partially) discrete algorithms. If a sample is fully discrete, DART is the best option. If only the densest material is homogeneous, PDART retains a number of the advantages of discrete tomography, while providing a SIRT reconstruction of the background. If no assumptions can be made, SIRT is the most generally applicable algorithm, which is also widely available.

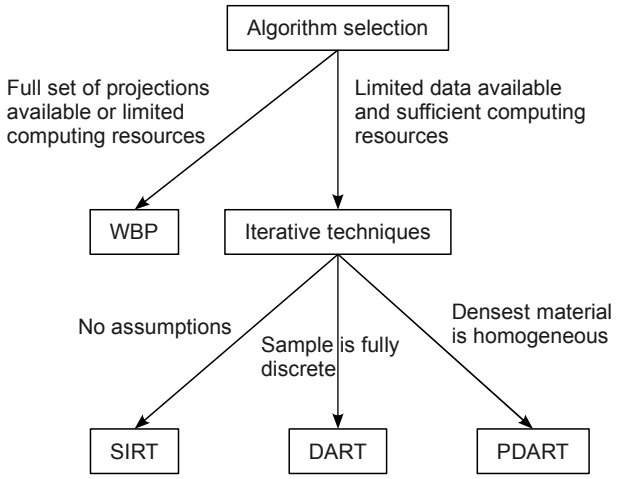



**Figure 2.16:** Algorithm selection tree.

## 2.5 Conclusions

We have presented a novel reconstruction algorithm for partially discrete tomography. The new method has two main benefits.

First, a partially discrete technique is useful to expand the set of samples for which concepts of discrete tomography can be applied. The method presented in this paper is a practical one in this regard, since it is applicable to the common problem of dense particle segmentation. The method retains two advantages of discrete tomography: the densest composition is automatically segmented, and it can produce reconstructions that are more accurate than SIRT.

Second, the method automatically determines the gray level that should be used for the densest composition in an objective manner, by locating the value that makes the resulting reconstruction closest to the original projection images. This property of the algorithm should make it easier to draw quantitative conclusions from the reconstructions.

## Acknowledgments

The authors would like to thank M. Van Bael and H. Wang from the K.U.Leuven (Belgium) for providing the Pb–Si sample. They would also like to thank R. Cieslinski and S. Rozeveld from the The Dow Chemical Company (USA) for providing the Catalyst sample. The authors also gratefully acknowledge the assistance of K. Schladitz and M. Godehardt from the Fraunhofer ITWM (Germany) with the analysis of the Catalyst sample. This work was financially supported by the Research Foundation (FWO), Flanders, Belgium (Contract no. G.0247.08).

## References

[1] P. A. Midgley and R. E. Dunin-Borkowski, "Electron tomography and holography in materials science," *Nature Mater.*, vol. 8, pp. 271–280, 2009.

[2] R. J. Spontak, M. C. Williams, and D. A. Agard, "Three-dimensional study of cylindrical morphology in a styrene-butadiene-styrene block copolymer," *Polymer*, vol. 29, no. 3, pp. 387–395, 1988.

[3] A. J. Koster, U. Ziese, A. J. Verkleij, A. H. Janssen, and K. P. de Jong, "Three-dimensional transmission electron microscopy: A novel imaging and characterization technique with nanometer scale resolution for materials science," *J. Phys. Chem. B*, vol. 104, no. 40, pp. 9368–9370, 2000.

[4] S. Bals, G. Van Tendeloo, and C. Kisielowski, "A new approach for electron tomography: Annular dark-field transmission electron microscopy," *Adv. Mater.*, vol. 18, no. 7, pp. 892–895, 2006.

[5] P. A. Midgley, M. Weyland, J. M. Thomas, and B. F. G. Johnson, "Z-contrast tomography: a technique in three-dimensional nanostructural analysis based on Rutherford scattering," *Chem. Commun.*, vol. 10, pp. 907–908, 2001.

[6] E. P. W. Ward, T. J. V. Yates, J.-J. Fernández, D. E. W. Vaughan, and P. A. Midgley, "Three-dimensional nanoparticle distribution and local curvature of heterogeneous catalysts revealed by electron tomography," *J. Phys. Chem. C*, vol. 111, no. 31, pp. 11 501–11 505, 2007.

[7] M. Weyland, T. J. V. Yates, R. E. Dunin-Borkowski, L. Laffont, and P. A. Midgley, "Nanoscale analysis of three-dimensional structures by electron tomography," *Scripta Mater.*, vol. 55, no. 1, pp. 29–33, 2006.

[8] T. J. V. Yates, J. M. Thomas, J.-J. Fernández, O. Terasaki, R. Ryoo, and P. A. Midgley, "Three-dimensional real-space crystallography of MCM-48 mesoporous silica revealed by scanning transmission electron tomography," *Chem. Phys. Lett.*, vol. 418, no. 4–6, pp. 540–543, 2006.

[9] K. Kaneko, K. Inoke, K. Sato, K. Kitawaki, H. Higashida, I. Arslan, and P. A. Midgley, "TEM characterization of Ge precipitates in an Al–1.6 at% Ge alloy," *Ultramicroscopy*, vol. 108, no. 3, pp. 210–220, 2008.

[10] G. Möbus and B. J. Inkson, "Three-dimensional reconstruction of buried nanoparticles by element-sensitive tomography based on inelastically scattered electrons," *Appl. Phys. Lett.*, vol. 79, no. 9, pp. 1369–1371, 2001.

[11] M. Weyland and P. A. Midgley, "Extending energy-filtered transmission electron microscopy (EFTEM) into three dimensions using electron tomography," *Microsc. Microanal.*, vol. 9, no. 6, pp. 542–555, 2003.

[12] M. H. Gass, K. K. K. Koziol, A. H. Windle, and P. A. Midgley, "Four-dimensional spectral tomography of carbonaceous nanocomposites," *Nano Lett.*, vol. 6, no. 3, pp. 376–379, 2006.

[13] B. Goris, S. Bals, W. Van den Broek, J. Verbeeck, and G. Van Tendeloo, "Exploring different inelastic projection mechanisms for electron tomography," *Ultramicroscopy*, vol. 111, no. 8, pp. 1262–1267, 2011.

[14] L. C. Gontard, R. E. Dunin-Borkowski, and D. Ozkaya, "Three-dimensional shapes and spatial distributions of Pt and PtCr catalyst nanoparticles on carbon black," *J. Microsc.*, vol. 232, no. 2, pp. 248–259, 2008.

[15] P. Gilbert, "Iterative methods for the three-dimensional reconstruction of an object from projections," *J. Theor. Biol.*, vol. 36, no. 1, pp. 105–117, 1972.

[16] C. Kübel, D. Niemeyer, R. Cieslinski, and S. Rozeveld, "Electron tomography of nanostructured materials – towards a quantitative 3D analysis with nanometer resolution," *Mater. Sci. Forum*, vol. 638–642, pp. 2517–2522, 2010.

[17] S. Van Aert, K. J. Batenburg, M. D. Rossell, R. Erni, and G. Van Tendeloo, "Three-dimensional atomic imaging of crystalline nanoparticles," *Nature*, vol. 470, pp. 374–377, 2011.

[18] S. Bals, M. Casavola, M. A. van Huis, S. Van Aert, K. J. Batenburg, G. Van Tendeloo, and D. Vanmaekelbergh, "Three-dimensional atomic imaging of colloidal core-shell nanocrystals," *Nano Lett.*, vol. 11, no. 8, pp. 3420–3424, 2011.

[19] K. J. Batenburg and J. Sijbers, "DART: A practical reconstruction algorithm for discrete tomography," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2542–2553, 2011.

[20] K. J. Batenburg, S. Bals, J. Sijbers, C. Kübel, P. A. Midgley, J. C. Hernandez, U. Kaiser, E. R. Encina, E. A. Coronado, and G. Van Tendeloo, "3D imaging of nanomaterials by discrete tomography," *Ultramicroscopy*, vol. 109, no. 6, pp. 730–740, 2009.

[21] S. Bals, K. J. Batenburg, J. Verbeeck, J. Sijbers, and G. Van Tendeloo, "Quantitative three-dimensional reconstruction of catalyst particles for bamboo-like carbon nanotubes," *Nano Lett.*, vol. 7, no. 12, pp. 3669–3674, 2007.

[22] S. Bals, K. J. Batenburg, D. Liang, O. Lebedev, G. Van Tendeloo, A. Aerts, J. A. Martens, and C. E. A. Kirschhock, "Quantitative three-dimensional modeling of zeotile through discrete electron tomography," *J. Am. Chem. Soc.*, vol. 131, no. 13, pp. 4769–4773, 2009.

[23] K. J. Batenburg, J. Sijbers, H. F. Poulsen, and E. Knudsen, "DART: a robust algorithm for fast reconstruction of three-dimensional grain maps," *J. Appl. Cryst.*, vol. 43, no. 6, pp. 1464–1473, 2010.

[24] J. Gregor and T. Benson, "Computational analysis and improvement of SIRT," *IEEE Trans. Med. Imag.*, vol. 27, no. 7, pp. 918–924, 2008.

[25] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder-Mead simplex method in low dimensions," *SIAM J. Optim.*, vol. 9, no. 1, pp. 112–147, 1998.

[26] X. Ke, S. Bals, A. R. Negreira, T. Hantschel, H. Bender, and G. Van Tendeloo, "TEM sample preparation by FIB for carbon nanotube interconnects," *Ultramicroscopy*, vol. 109, no. 11, pp. 1353–1359, 2009.

# 3

# Reconstructed Residual Error

**This chapter has been submitted for publication as**

T. Roelandts, K. J. Batenburg, and J. Sijbers, "The Reconstructed Residual Error Visualizes and Improves the Quality of a Segmented Tomogram," *IEEE Transactions on Image Processing*, 2013.

**Abstract**—In this paper, we present the reconstructed residual error, which visualizes and improves the quality of a segmented tomogram. The method, which is independent of the reconstruction and segmentation algorithms, reconstructs the difference between the recorded data and the forward projection of the segmented tomogram. The Moore-Penrose pseudoinverse of the projection matrix is used as a mathematical model for the reconstruction process, showing that segmentation removes reconstruction artifacts, and that the new artifacts that are introduced by the segmentation itself can be studied and improved by extracting new information from the recorded data. The properties and applications of the algorithm are verified experimentally through simulations and experimental $\mu$CT data. The experiments show that the reconstructed residual error is close to the true error. Three applications of the technique are introduced. First, it is used to improve gray level estimates. Second, for reconstruction algorithms that assume a priori knowledge of the gray levels, it is used to improve the segmentation itself. Finally, the reconstructed residual error is used to select the most accurate of several segmentations, for experimental $\mu$CT data.

## 3.1 Introduction

In many applications of tomography, the tomographic reconstruction (the *tomogram*) must be segmented before the results can be analyzed. Segmentation

amounts to the classification of image pixels into distinct classes, based on similarity with respect to some characteristic. Image segmentation is a well established field, and a range of methods has been developed, such as global or local thresholding, region growing, and clustering [1, 2].

The current paper focuses on reconstruction problems for which the segmentation is based on the density of the scanned object, where we use the term *density* to refer to the particular physical property of the object of which linear projections are acquired during the scanning process (e.g., mass density, X-ray attenuation, electron beam scattering, etc.). In this common case, each material in the tomogram is expected to have a (relatively) constant gray level, since each gray level represents, e.g., a type of tissue (in medical applications) or a composition (in materials science). This allows exploiting the projection images during the segmentation step, since those represent projections of that density. Most current segmentation methods do not exploit this information, since they are not specific towards the modality that was used to acquire the image.

Recently, global and local thresholding methods were proposed that do employ the projection data to improve the selection of threshold parameters [3, 4]. In addition, reconstruction methods were developed for which the segmentation was directly incorporated into the reconstruction algorithm. These *discrete* and *partially discrete* algorithms assume that the gray levels to be used for the reconstruction are known *a priori*. The recently proposed DART [5, 6] and PDART [7, 8] are examples of such algorithms. In [9], the projection data is used to estimate the parameters of DART.

This paper introduces the *reconstructed residual error* of a segmented tomogram, which reconstructs the difference between the recorded data and the forward projection of that segmented tomogram. The result is a reconstruction of the segmentation errors, which provides a visual map of those errors. The technique is independent of the reconstruction and segmentation algorithms, since it takes the final segmented tomogram as an input. Hence, an error map can be computed regardless of the algorithms that were used to create that segmented tomogram. The reconstructed residual error allows to detect and correct errors in the gray levels of the segmented image. For reconstruction algorithms that assume a priori knowledge of the gray levels, it allows correcting the segmentation itself.

We use the Moore-Penrose pseudoinverse (see, e.g., [10, Section 7.3]) of the projection matrix as a mathematical model, both for the original reconstruction and for the reconstructed residual error, allowing for mathematical rigor. As many iterative reconstruction methods converge to this pseudoinverse, it can serve as a model for the behavior of actual reconstruction methods. An overview of recent developments in iterative reconstruction algorithms is given in [11].

Various types of iterative methods exist. The examples from *discrete tomogra-*

*phy* that were mentioned before are iterative algorithms that exploit prior knowledge of the gray levels. *Model-based* algorithms try to model the acquisition process as closely as possible, including, e.g., the geometry of the scanner and the X-ray beam being polychromatic [12–14]. *Statistical* methods incorporate counting statistics of the detected photons, leading to algorithms such as expectation-maximization [15]. The most basic type of iterative reconstruction is done by algorithms such as ART, SART, and SIRT [16–18]. The Moore-Penrose pseudoinverse of the projection matrix is the limit case of this last type of solutions. It provides exactly the shortest least squares (LS) solution of the reconstruction problem. Conjugate gradient methods can be used to compute this shortest LS solution in practice [19]. However, an algorithm such as SIRT is often preferable, since it has a regularizing effect on the reconstruction by providing a *weighted* LS solution. SIRT can also be interpreted as the shortest LS solution of a *rescaled* system of equations [20]. Algorithms such as SART and SIRT are still the subject of active research, to improve their performance, to add new functionality, and to implement them on graphics processing units (GPUs) [21–25].

The pseudoinverse is used to show that the artifacts of a reconstruction cannot be removed using the recorded data. However, it is then shown that a segmentation step mostly replaces these artifacts with ones that *can* be removed. The reconstructed residual error exploits this by creating a reconstruction of the segmentation artifacts, thereby extracting new information from the recorded data, which is the reason that the proposed technique can provide superior results over techniques that operate solely in the reconstruction domain. The approach is then generalized, to show that any suitable reconstruction algorithm can be used to compute the reconstructed residual error.

The structure of this paper is as follows. In Section 3.2, the reconstructed residual error is defined, its properties are described in detail, and several applications are introduced. Section 3.3 reports on the results of experiments, using both simulations and experimental $\mu$CT data. These results are discussed in Section 3.4, and a conclusion is reached in Section 3.5.

## 3.2 Reconstructed Residual Error

Here, we describe the reconstructed residual error. We first present an intuitive overview of the algorithm, before giving a complete description of the algorithm and its properties.

### 3.2.1 Overview

Fig. 3.1 presents an overview of the computation of the reconstructed residual error. Its inputs are the original projection data (Fig. 3.1b) and the segmented reconstruction (Fig. 3.1e). The original (non-segmented) reconstruction (Fig. 3.1c) is not used. The original projection data (Fig. 3.1b) is a sinogram, acquired by rotating around the object (Fig. 3.1a). The segmented reconstruction is computed by thresholding Fig. 3.1c, and subsequently choosing gray levels. To compute the reconstructed residual error, the segmented reconstruction (Fig. 3.1e) is first projected forward. The result of this operation (Fig. 3.1d) is then subtracted from the original projections, resulting in the *residual projection error* (Fig. 3.1g).

The residual projection error is then reconstructed to provide the reconstructed residual error (Fig. 3.1h), which provides a visual map of the segmentation error. From Fig. 3.1h, it is clear that both gray levels of the segmented reconstruction are incorrect. The erroneous lines and dots in the segmented reconstruction (Fig. 3.1e), which are caused by artifacts in the reconstruction (Fig. 3.1c), are also clearly visible in Fig. 3.1h. Note that the true error (Fig. 3.1i), which is the difference between the original object (Fig. 3.1a) and the segmented reconstruction, is quite close to the reconstructed residual error.

An alternative that might be considered, is to simply compute the difference between the original and the segmented reconstruction (Fig. 3.1f). However, this difference can be expected to show much more reconstruction artifacts, as is explained in Section 3.2.4. An intuitive way to see this is that, for phantom experiments, the segmented reconstruction can potentially be identical to the phantom, in which case the residual projection error (Fig. 3.1g) will be zero. Hence, a perfect segmentation results in the reconstructed residual error (Fig. 3.1h) being zero everywhere, while the difference between the original and the segmented reconstruction (Fig. 3.1f) will always exhibit the reconstruction artifacts of the original reconstruction (Fig. 3.1c).

### 3.2.2 Notation and Concepts

The projection process in tomography can be modeled as a linear operator that is determined by the projection geometry. This leads to a system of linear equations,

$$\boldsymbol{W}\boldsymbol{x} = \boldsymbol{p}, \tag{3.1}$$

where $\boldsymbol{p} \in \mathbb{R}^m$ contains the projection data and $\boldsymbol{x} \in \mathbb{R}^n$ corresponds to the image. The linear operator is represented by the $m \times n$ matrix $\boldsymbol{W}$, the *projection matrix*. An approximate solution $\hat{\boldsymbol{x}} \in \mathbb{R}^n$ of (3.1) can then be computed, for example by minimizing some norm $\|\boldsymbol{W}\boldsymbol{x} - \boldsymbol{p}\|$.

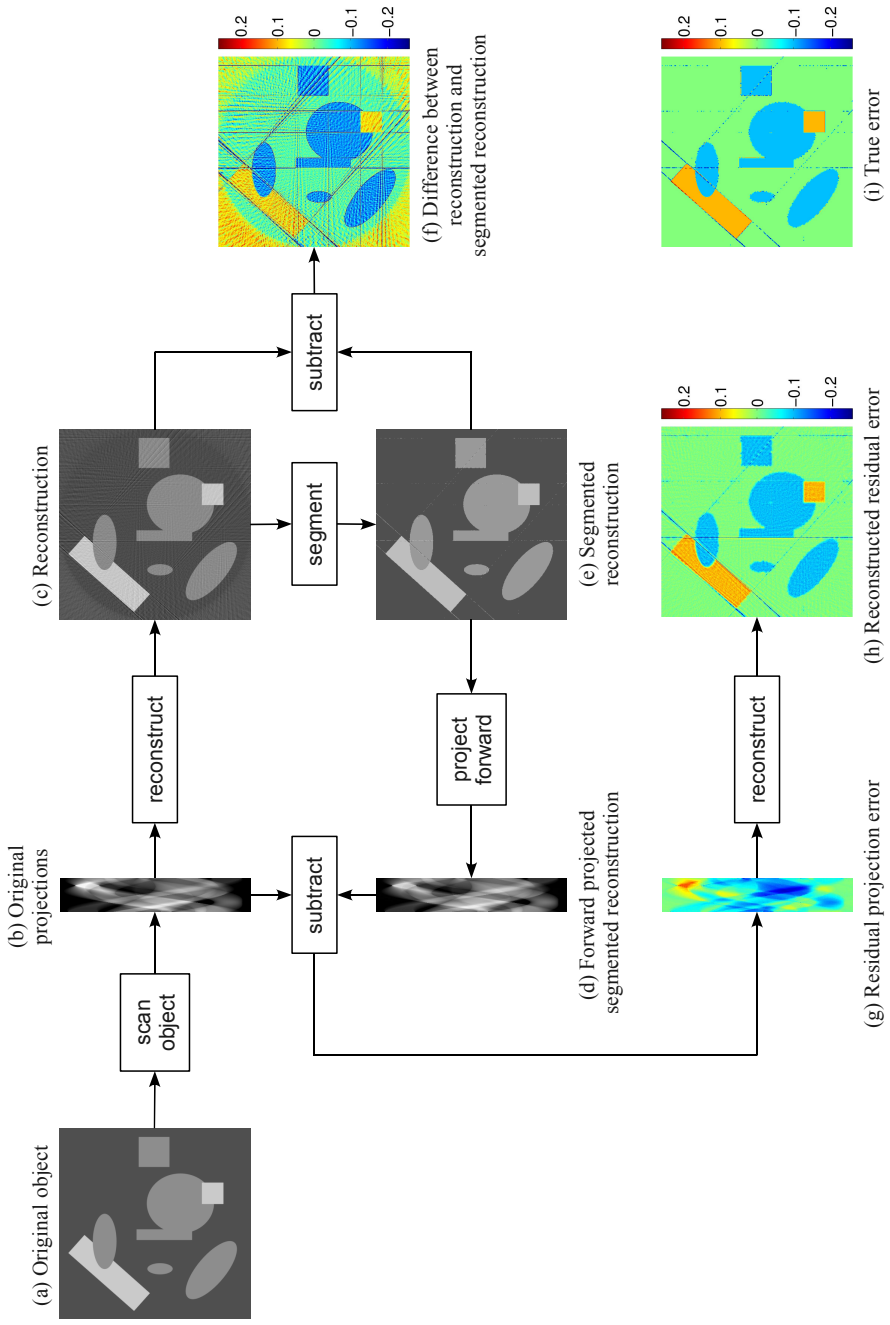



**Figure 3.1:** Overview of the computation of the reconstructed residual error.

We explicitly refer to the ground truth object as $\boldsymbol{g} \in \mathbb{R}^n$, which leads to $\boldsymbol{p} = \boldsymbol{W}\boldsymbol{g}$. We denote a general reconstruction algorithm as an operator $\boldsymbol{\mathcal{R}} : \mathbb{R}^m \to \mathbb{R}^n$, which leads to $\hat{\boldsymbol{x}} = \boldsymbol{\mathcal{R}}\boldsymbol{W}\boldsymbol{g}$. The (shortest) least squares solution is denoted as $\boldsymbol{x}^+$, where $\boldsymbol{x}^+ = \boldsymbol{W}^+\boldsymbol{p}$, with $\boldsymbol{W}^+$ the Moore-Penrose pseudoinverse of $\boldsymbol{W}$. We denote the row space of $\boldsymbol{W}$ by $R(\boldsymbol{W})$, and the null space by $N(\boldsymbol{W})$. See, e.g., [10], for details on these concepts from linear algebra.

The final step in the tomographic workflow is often *segmentation*. A segmentation method essentially partitions the pixels of an image into sets $Y_1, \ldots, Y_d$, where $d$ is the number of classes (segments) in the segmented image. If it is assumed that the scanned object consists of homogeneous regions, then the segmented image should have the same gray level for all pixels in a set $Y_k$. Hence, from the reconstruction $\hat{\boldsymbol{x}}$, a segmented image $\boldsymbol{s} \in \mathbb{R}^n$ can be created by assigning a gray level $\rho_k \in \mathbb{R}$ to all pixels in a set $Y_k$, for each $k \in \{1, \ldots, d\}$. If the values of $\rho_1, \ldots, \rho_d$ are not known, the mean of all pixel values in a class can be used as an estimate,

$$\hat{\rho}_k = \frac{1}{|Y_k|} \sum_{y \in Y_k} y, \text{ for each } k \in \{1, \ldots, d\}. \tag{3.2}$$

### 3.2.3 Definition

Here, we define the reconstructed residual error. Inputs are the segmented image $\boldsymbol{s}$ and the projection data $\boldsymbol{p}$. First, $\boldsymbol{s}$ is forward projected to give $\boldsymbol{p}_{\boldsymbol{s}} \in \mathbb{R}^m$, so $\boldsymbol{p}_{\boldsymbol{s}} = \boldsymbol{W}\boldsymbol{s}$. The *residual projection error* $\boldsymbol{p}_{\boldsymbol{e}} \in \mathbb{R}^m$ is then defined as $\boldsymbol{p}_{\boldsymbol{e}} = \boldsymbol{p} - \boldsymbol{p}_{\boldsymbol{s}}$. Second, the residual error is reconstructed by computing an approximate solution of the system

$$\boldsymbol{W}\boldsymbol{e} = \boldsymbol{p} - \boldsymbol{p}_{\boldsymbol{s}}, \tag{3.3}$$

as for the system of (3.1). Here, $\boldsymbol{e} \in \mathbb{R}^n$ corresponds to the (unknown) error image. The approximate solution $\hat{\boldsymbol{e}} \in \mathbb{R}^n$ is then defined to be the reconstructed residual error. The least squares solution is again denoted by $\boldsymbol{e}^+$, as for $\boldsymbol{x}^+$.

### 3.2.4 Least Squares Approach

In this Section, we study the reconstructed residual error under the assumption that the Moore-Penrose pseudoinverse $\boldsymbol{W}^+$ is used for the reconstruction, both for the original tomogram and for the reconstructed residual error. Doing so makes the reconstruction algorithm a well-defined operator, and allows us to be mathematically rigorous.

First, we formally define what we mean by the term *artifact*. Typically, a tomogram contains different types of artifacts, with causes such as scanner imperfections, noise, the reconstruction algorithm, the segmentation algorithm, etc. In

the current paper, we consider all differences between the ground truth object and the (possibly segmented) final image to be artifacts.

*Definition 1: The artifacts of an image* $\boldsymbol{y} \in \mathbb{R}^n$ *are* $\boldsymbol{g} - \boldsymbol{y}$*, where* $\boldsymbol{g} \in \mathbb{R}^n$ *is the ground truth object.*

Each image $\boldsymbol{y} \in \mathbb{R}^n$ can be split into two orthogonal components, with the first in the row space and the second in the null space of $\boldsymbol{W}$. We write this as $\boldsymbol{y} = \boldsymbol{y}_R + \boldsymbol{y}_N$, with $\boldsymbol{y}_R, \boldsymbol{y}_N \in \mathbb{R}^n$, and $\boldsymbol{y}_R \perp \boldsymbol{y}_N$. Below, we reiterate a few basic properties of $\boldsymbol{W}$ and $\boldsymbol{W}^+$, for convenience.

*Property 1:* $R(\boldsymbol{W}) \perp N(\boldsymbol{W})$.

*Property 2:* $\boldsymbol{W}\boldsymbol{W}^+\boldsymbol{W} = \boldsymbol{W}$.

*Property 3:* $\boldsymbol{W}^+\boldsymbol{W}\boldsymbol{y} = \boldsymbol{y}_R$.

Property 3 implies that the shortest least squares reconstruction $\boldsymbol{x}^+$ exactly recovers the row space component of $\boldsymbol{g}$, since $\boldsymbol{x}^+ = \boldsymbol{W}^+\boldsymbol{p} = \boldsymbol{W}^+\boldsymbol{W}\boldsymbol{g} = \boldsymbol{g}_R$.

*Proposition 1:* $\boldsymbol{g} - \boldsymbol{x}^+ \in N(\boldsymbol{W})$*, i.e., the artifacts of the shortest least squares solution are in the null space of* $\boldsymbol{W}$.

We have $\boldsymbol{x}^+ = \boldsymbol{W}^+\boldsymbol{p} = \boldsymbol{W}^+\boldsymbol{W}\boldsymbol{g}$. It is instructive to split $\boldsymbol{x}^+$ into the ground truth and a component that represents the artifacts, as $\boldsymbol{x}^+ = \boldsymbol{W}^+\boldsymbol{W}\boldsymbol{g} = \boldsymbol{g} + (\boldsymbol{W}^+\boldsymbol{W} - \boldsymbol{I})\boldsymbol{g}$. On forward projection, the second component disappears, since $\boldsymbol{W}\boldsymbol{W}^+\boldsymbol{W} - \boldsymbol{W} = \boldsymbol{0}$. Hence, $\boldsymbol{W}\boldsymbol{x}^+ = \boldsymbol{W}\boldsymbol{g}$, and the artifacts are indeed in $N(\boldsymbol{W})$.□

The implication of Proposition 1 is that the reconstruction artifacts of $\boldsymbol{x}^+$ cannot be studied or further reduced by comparing them with the projection data $\boldsymbol{p}$ after forward projection.

This changes after the reconstruction is segmented. Reconstruction artifacts are typically non-homogeneous structures such as streaks and blurring. Segmentation into regions that each have a given gray level $\rho_k$, however, results in an image that is composed of homogeneous regions, since the segmentation removes the small variations of the gray levels within each region. Hence, it removes a large part of the reconstruction artifacts, while, of course, also introducing new artifacts. These new artifacts are typically very different from reconstruction artifacts, and mainly consist of erroneous gray levels and wrongly classified pixels at the edges of different regions in the image. In contrast with reconstruction artifacts, which are in the null space of $\boldsymbol{W}$, segmentation artifacts are not expected to have zero ray sums. In a homogeneous region with a gray level offset, individual pixels all share the same error, resulting in a nonzero total contribution. This means that the artifacts of $\boldsymbol{s}$ are expected to have a (large) component in $R(\boldsymbol{W})$, and *can* be studied and further reduced by comparing them with the projection data after forward projection. The reconstructed residual error is an application of this.

Denote the segmentation artifacts $\boldsymbol{a} \in \mathbb{R}^n$ of $\boldsymbol{s}$ as $\boldsymbol{a} \equiv \boldsymbol{g} - \boldsymbol{s}$. We then have the following proposition.

*Proposition 2:* $\boldsymbol{e}^+ = \boldsymbol{a}_R$*, i.e., the reconstructed residual error exactly recovers*

*the row space component of* $\boldsymbol{a}$.

From (3.3) and $\boldsymbol{p_s} = \boldsymbol{W}\boldsymbol{s}$, we have $\boldsymbol{e}^+ = \boldsymbol{W}^+(\boldsymbol{p}-\boldsymbol{p_s}) = \boldsymbol{W}^+\boldsymbol{W}\boldsymbol{g}-\boldsymbol{W}^+\boldsymbol{W}\boldsymbol{s} = \boldsymbol{W}^+\boldsymbol{W}\boldsymbol{a} = \boldsymbol{a}_R$. □

This means that the reconstructed residual error is a reconstruction of the segmentation artifacts in the same way that the original tomogram is a reconstruction of the scanned object, i.e., we have $\boldsymbol{e}^+ = \boldsymbol{a}_R$ and $\boldsymbol{x}^+ = \boldsymbol{g}_R$. The main implication of Proposition 2 is that the reconstructed residual error can be used to study the segmentation artifacts directly, starting from the segmented tomogram.

Finally, we briefly return to the alternative of using the difference between the reconstruction and the segmented reconstruction, $\boldsymbol{x}^+ - \boldsymbol{s}$, as an estimate of the segmentation artifacts $\boldsymbol{a}$, as mentioned in Section 3.2.1. We compare the artifacts of $\boldsymbol{x}^+ - \boldsymbol{s}$ with those of $\boldsymbol{e}^+$. The artifacts of $\boldsymbol{e}^+$ are $\boldsymbol{a}-\boldsymbol{e}^+$. The artifacts of $\boldsymbol{x}^+ - \boldsymbol{s}$, seen as an estimate of $\boldsymbol{a}$, are exactly the artifacts of $\boldsymbol{x}^+$, since $\boldsymbol{a} - (\boldsymbol{x}^+ - \boldsymbol{s}) = \boldsymbol{g} - \boldsymbol{s} - \boldsymbol{x}^+ + \boldsymbol{s} = \boldsymbol{g} - \boldsymbol{x}^+$. And, since $\|\boldsymbol{a}\| \ll \|\boldsymbol{g}\|$, we typically expect that $\|\boldsymbol{a} - \boldsymbol{e}^+\| \ll \|\boldsymbol{g} - \boldsymbol{x}^+\|$, which would imply that the artifacts of $\boldsymbol{x}^+ - \boldsymbol{s}$ make it a poor estimate of $\boldsymbol{a}$.

### 3.2.5 General Reconstruction Algorithms

Leaving behind the "perfect" reconstruction algorithm that is the Moore-Penrose pseudoinverse, we now investigate how the results from Section 3.2.4 change when $\boldsymbol{W}^+$ is replaced by a practical reconstruction algorithm. In Section 3.3, we show experimentally that doing so still results in useful maps of the segmentation artifacts. There, we employ a diverse set of reconstruction algorithms, not restricted to linear ones for the reconstruction of the original tomogram. For the reconstructed residual error itself, however, we restrict ourselves to SIRT [18], to keep the results easily comparable.

If a general, possibly non-linear, reconstruction algorithm is used for the reconstruction of the original tomogram, then Proposition 1 is no longer applicable. However, the reconstructed residual error uses the *segmented* reconstruction $\boldsymbol{s}$, which is still expected to have a (large) component in $R(\boldsymbol{W})$. Hence, the effectiveness of the reconstructed residual error is not affected. Moreover, depending on the component of the artifacts that is in $R(\boldsymbol{W})$, it might be possible to study the reconstruction artifacts of the original tomogram directly, without a separate segmentation step. An obvious example of this are discrete algorithms such as DART, which directly produce a fully segmented result [5, 6].

A general reconstruction algorithm can also replace $\boldsymbol{W}^+$ for the computation of the reconstructed residual error itself. The result is then no longer the row space component of $\boldsymbol{a}$. We have the following proposition.

*Proposition 3: For a general reconstruction algorithm* $\boldsymbol{\mathcal{R}}$, $\hat{\boldsymbol{e}} = \boldsymbol{\mathcal{R}}\boldsymbol{W}\boldsymbol{a}$, *i.e., the reconstructed residual error is the reconstruction of the segmentation artifacts* $\boldsymbol{a}$.

From (3.3) and $\boldsymbol{p_s} = \boldsymbol{W}\boldsymbol{s}$, we have $\hat{\boldsymbol{e}} = \boldsymbol{\mathcal{R}}(\boldsymbol{p}-\boldsymbol{p_s}) = \boldsymbol{\mathcal{R}}(\boldsymbol{W}\boldsymbol{g}-\boldsymbol{W}\boldsymbol{s}) = \boldsymbol{\mathcal{R}}\boldsymbol{W}\boldsymbol{a}$.□

Proposition 3 states that $\hat{\boldsymbol{e}}$ is a reconstruction of $\boldsymbol{a}$ in the same sense that $\hat{\boldsymbol{x}}$ is a reconstruction of the ground truth, since $\hat{\boldsymbol{x}} = \boldsymbol{\mathcal{R}}\boldsymbol{W}\boldsymbol{g}$. Together with the shortest least squares solution often not being a desirable solution anyway, due to noise in the projection data, this implies that $\boldsymbol{W}^+$ can be replaced by a suitable more general reconstruction algorithm for computing the reconstructed residual error.

### 3.2.6 Applications

This Section presents three applications of the reconstructed residual error. These applications also appear in Section 3.3, where they are supported with experiments.

*1) Improving Gray Level Estimates:* As mentioned in the Introduction, we focus on problems for which the segmentation is based on the density of the scanned object. This gives the gray levels a physical meaning, in that they represent the density of each region. It is therefore important to determine them accurately. Let $\rho_k$ denote the true gray levels. In practice, the $\rho_k$ are often unknown, which means that they have to be estimated from the tomogram $\hat{\boldsymbol{x}}$. However, any estimates $\hat{\rho}_k$ that are computed in this way will be influenced by the reconstruction artifacts of $\hat{\boldsymbol{x}}$.

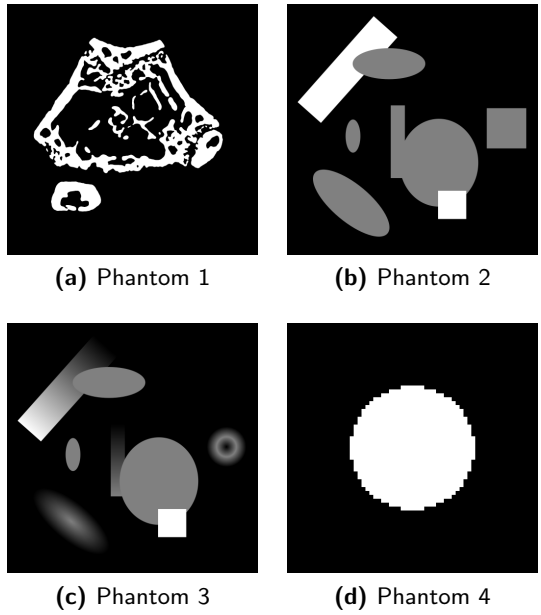
The reconstructed residual error avoids the reconstruction artifacts of $\hat{\boldsymbol{x}}$, since it directly estimates the segmentation artifacts. Therefore, it allows improving the estimates of the gray levels $\hat{\rho}_k$ of the segmented image. If (3.2) is applied to $\hat{\boldsymbol{e}}$, with the segmentation classes of $\boldsymbol{s}$ as the sets $Y_k$, the result is an estimate of the gray level errors $\hat{\varepsilon}_k$. Therefore, the gray level estimates can be improved by setting $\hat{\rho}'_k = \hat{\rho}_k + \hat{\varepsilon}_k$, for each $k \in \{1, \ldots, d\}$. The updated gray levels can then be used to improve the segmented image.

*2) Improving Segmentation:* Discrete reconstruction algorithms exploit a priori knowledge of the gray levels to directly create a segmented tomogram $\boldsymbol{s}$, combining reconstruction and segmentation. These a priori gray levels are often estimated from an initial (non-discrete) reconstruction, which again means that reconstruction artifacts limit their accuracy. Of course, since the gray levels of $\boldsymbol{s}$ are now *inputs* of the reconstruction, they influence the classification of the pixels in $\boldsymbol{s}$ (the partition $Y_1, \ldots, Y_d$). Hence, after improving the gray level estimates as before, rerunning the reconstruction algorithm might result in a more accurate segmented image $\boldsymbol{s}'$ (i.e., with a different partition $Y'_1, \ldots, Y'_d$). Moreover, repeating these steps may produce further improvements, since the accuracy of the segmentation, in turn, influences the accuracy of the gray levels that are computed using (3.2).

*3) Selecting the Most Accurate Segmentation:* This is a straightforward application, since a visual map of the segmentation error can clearly be used to discriminate between several segmented tomograms. In Section 3.3, this is demonstrated using experimental $\mu$CT Data.

## 3.3 Experiments and Results

In this Section, we describe the experiments, for both simulated and experimental data, which were carried out to investigate the properties of the reconstructed residual error. Four *phantom images* were created, the first three of size $2048 \times 2048$ pixels (Figs. 3.2a, 3.2b, and 3.2c). These phantoms have a higher resolution than the reconstruction grid to reduce the effect of the pixelation on the reconstructions. The fourth phantom (Fig. 3.2d) is of size $64 \times 64$ pixels, to allow exact least squares computations using $\boldsymbol{W}^{+}$. The simulation experiments using Phantoms 1–3 were performed on a square reconstruction grid of size $512 \times 512$ pixels. The experiment using Phantom 4 was performed on a grid of size $64 \times 64$ pixels.



**(a)** Phantom 1 **(b)** Phantom 2

**(c)** Phantom 3 **(d)** Phantom 4

**Figure 3.2:** Phantom images. (a) Two gray levels, based on the FBP reconstruction of a mouse femur. (b) Three gray levels. (c) Continuous grayscale. (d) Two gray levels, $64 \times 64$ pixels.

To demonstrate that the effectiveness of the reconstructed residual error does not depend on the reconstruction method, the following diverse set of algorithms was used in the experiments.

1) FBP is an analytical technique that is directly related to the inverse Radon transform. A standard implementation of FBP was used, with a Ram-Lak filter and linear interpolation in the projection domain.

2) CGLS is an algebraic technique that computes the shortest least squares solution of (3.1). We used 300 iterations of CGLS to approximate the effect of $\boldsymbol{W}^{+}$, allowing for the use of a larger reconstruction area than is possible when using the pseudoinverse directly.

3) SIRT is an algebraic technique that computes a *weighted* least squares solution of (3.1). SIRT was implemented as defined in [21], performing 300 iterations.

4) DART [5, 6] is a discrete algebraic technique that exploits prior knowledge of the gray levels. SIRT was used as the (embedded) algebraic reconstruction method. A total of 300 SIRT iterations was performed, divided between 100 initial and 200 update iterations. The fix probability was kept constant at 0.9.

5) PDART [7, 8] is a partially discrete technique that assumes that only the densest material is homogeneous. The gray level of the densest material is exploited as prior knowledge. A total of 300 iterations was performed.

6) TVMin solves (3.1) while minimizing the total variation (TV) seminorm of the solution. There several algorithms that implement TVMin, such as FISTA [26, 27] and NESTA [28]. The NESTA algorithm was used for the experiments, with the regularization parameter $\lambda$ set to 0.05, and using 100 iterations. We used the MATLAB implementation of the algorithm that can be downloaded from the NESTA website[1].

The results from DART and PDART are already segmented (partially for PDART). Otsu's method [29] was used for the algorithms that do not produce a segmented image, i.e., FBP, SIRT, and TVMin. CGLS was only used to compute row space components of images, so no segmentation was needed. Gray levels are prior knowledge for DART and, for the densest material, for PDART. For the other algorithms, they were estimated using (3.2) on the original reconstruction, using the classes from the segmented reconstruction as the sets $Y_k$.

The correctness of the DART reconstructions that are computed in Section 3.3.3 are assessed using the *relative number of misclassified pixels* (rNMP), defined as the ratio between the number of misclassified pixels and the total number of pixels in an image.

For all experiments, the reconstructed residual error was computed using 300 iterations of SIRT, to make the results comparable.

We now provide an overview of the experiments that follow. First, the basic properties of the algorithm are studied (Section 3.3.1). Then, three applications are presented. It is demonstrated that the technique can be used to detect and correct errors in the estimate of the gray levels (Section 3.3.2). The segmentation itself (the classification of the pixels) can be improved for reconstruction algorithms that use a priori knowledge of the gray levels (Section 3.3.3). Finally, the technique is applied to experimental $\mu$CT Data (Section 3.3.4), to select the most accurate

[1] http://www-stat.stanford.edu/~candes/nesta/

of several segmentations.

### 3.3.1 Properties of the Algorithm

This first experiment demonstrates the least squares approach from Section 3.2.4. Both the original and the segmented reconstruction were split into their row and null space components. To be able to compute $\boldsymbol{W}^+$ in practice, a small reconstruction area of size $64 \times 64$ pixels was combined with taking only 32 projections. This results in a system matrix $\boldsymbol{W}$ of size $2048 \times 4096$, and $\boldsymbol{W}^+$ of size $4096 \times 2048$. A synthetic dataset was created from Phantom 4 (Fig. 3.2d), also shown in Fig. 3.3d, using 32 equiangular parallel beam projections. Different from the other simulation experiments, noise was not added to this dataset. The reconstruction (Fig. 3.3e) was then computed using the Moore-Penrose pseudoinverse $\boldsymbol{W}^+$, making it the exact shortest least squares solution. The reconstruction was segmented using Otsu's method (Fig. 3.3g). The gray levels of the segmented reconstruction were determined using (3.2) on the pixels of the original reconstruction, with the Otsu segmentation classes as $Y_k$ $(k = 1, 2)$.

The artifacts of the reconstruction are shown separately (Fig. 3.3f), and split into row and null space components. As shown theoretically by Proposition 1, the artifacts are completely in the null space component (Fig. 3.3j). The row space component (Fig. 3.3a) is zero everywhere. The artifacts of the segmented reconstruction are also shown separately (Fig. 3.3h). The main assumption that was made in Section 3.2.4, was that the segmentation artifacts have a (large) component in the row space of $\boldsymbol{W}$. This assumption is confirmed by again splitting these artifacts into row and null space components. The null space component (Fig. 3.3k) is not zero, but most of the artifacts are clearly in the row space component (Fig. 3.3b). Using the least squares approach of Section 3.2.4, this row space component is exactly $\boldsymbol{e}^+$. However, when the reconstructed residual error is computed using SIRT (Fig. 3.3c), the result is very close to $\boldsymbol{e}^+$, indicating that SIRT is an acceptable practical alternative for $\boldsymbol{W}^+$. The true error (Fig. 3.3m), which is the difference between the original object (Fig. 3.3d) and the segmented reconstruction (Fig. 3.3g), is close to both $\boldsymbol{e}^+$ and the reconstructed residual error as computed with SIRT. The difference between the original and the segmented reconstruction (Fig. 3.3i) can clearly not be used as an alternative.

For the second experiment, a synthetic dataset was created from Phantom 1 (Fig. 3.2a), using 90 parallel beam projections, evenly spaced at $2^\circ$ intervals. A detector with 512 pixels was used, to simulate the practical situation where the detector pixel size equals the width of the reconstruction grid. Poisson noise, corresponding to $10^5$ initial photons per detector pixel, was applied to the synthetic dataset. The dataset was then reconstructed using 300 iterations of SIRT (Fig. 3.4a), and segmented using Otsu's method (Fig. 3.4d). The gray levels of

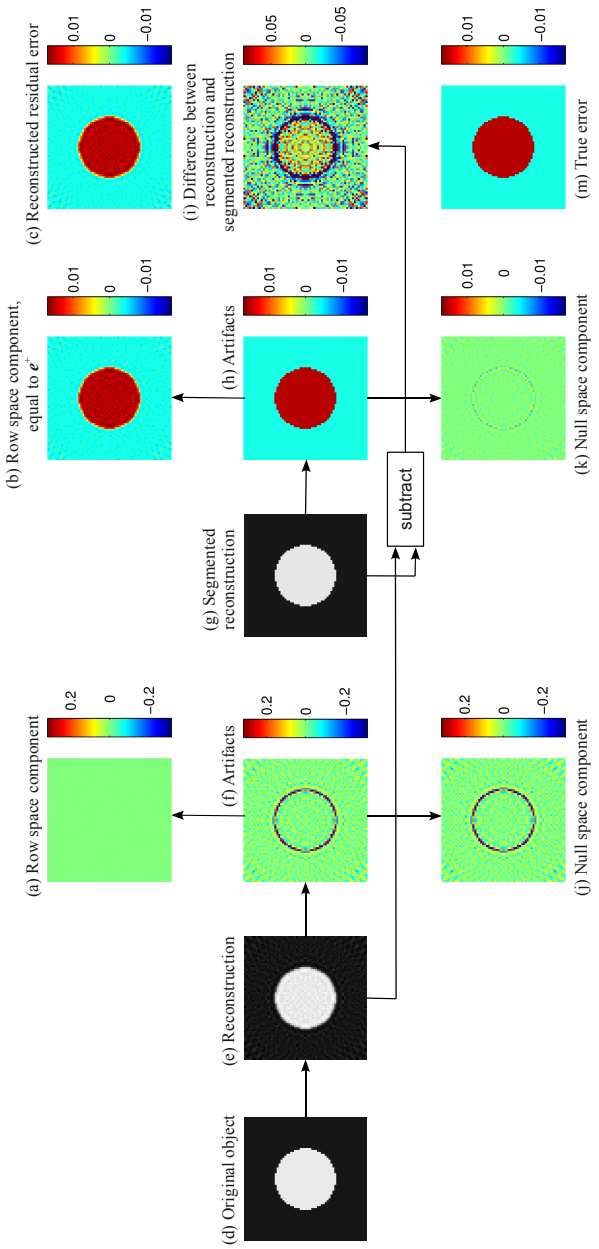
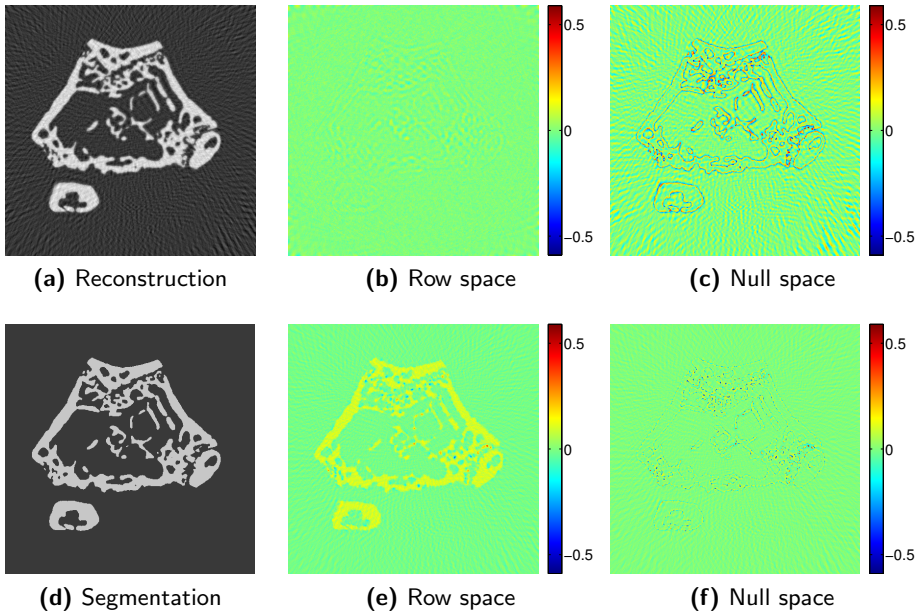



**Figure 3.3:** Complete results for a simple phantom, using a least squares approach. Artifacts of both the original and the segmented reconstruction are split into row and null space components.

Fig. 3.4d were determined using (3.2) on the pixels of Fig. 3.4a, with the Otsu segmentation classes as $Y_k$ $(k = 1, 2)$.




**(a)** Reconstruction **(b)** Row space **(c)** Null space

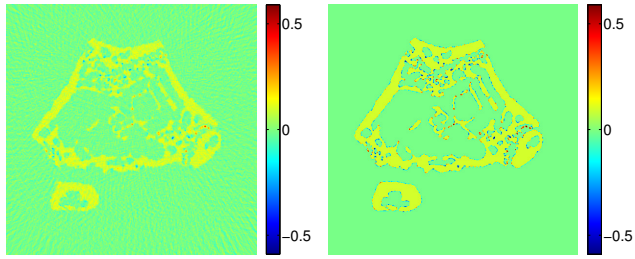
**(d)** Segmentation **(e)** Row space **(f)** Null space

**Figure 3.4:** (a) SIRT reconstruction of Phantom 1. (b) Row space component of the reconstruction artifacts. (c) Null space component of the reconstruction artifacts. (d) Segmented SIRT reconstruction of Phantom 1. (e) Row space component of the segmentation artifacts. (f) Null space component of the segmentation artifacts.

Using 300 iterations of CGLS as an approximation of $\boldsymbol{W}^+$, both the original and the segmented reconstruction were then split into their row and null space components. The artifacts of the reconstruction are again mainly in the null space component (Fig 3.4c), even though SIRT has replaced $\boldsymbol{W}^+$ as the reconstruction algorithm. The row space component (Fig. 3.4b) is small. For the artifacts of the segmented reconstruction, this is clearly different. The gray level error of the "bone" material, which is quite subtle since the gray levels were computed from the original reconstruction, clearly shows as an artifact in the row space component (Fig. 3.4e). The null space component (Fig. 3.4f) is small.

If the reconstructed residual error is computed using SIRT (Fig. 3.5a), then the result is close to the row space component of the segmentation artifacts (Fig 3.4e). The reconstructed residual error is also close to the true error (Fig. 3.5b).

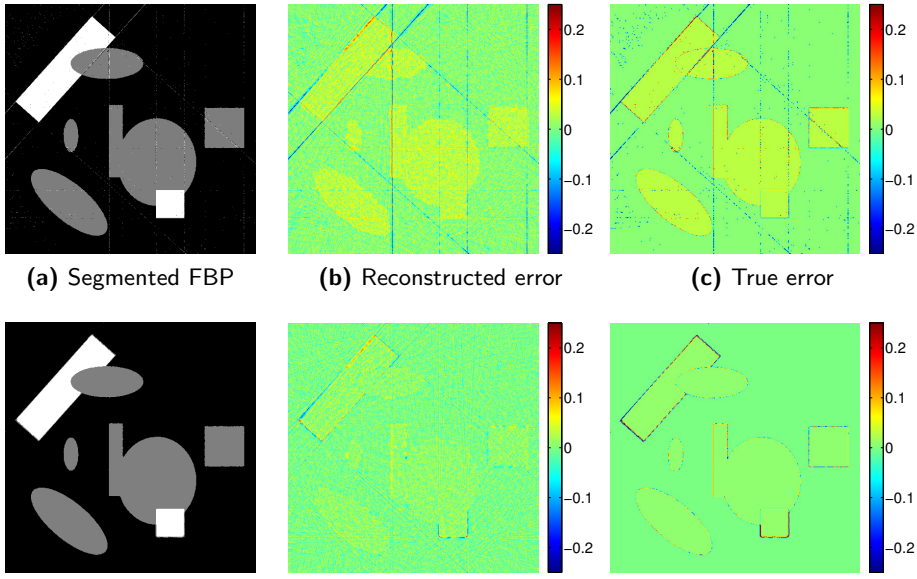
The third experiment is based on Phantom 2. A synthetic dataset was created, again using 90 parallel beam projections. Poisson noise, again corresponding to




**(a)** Reconstructed error **(b)** True error

**Figure 3.5:** (a) Reconstructed residual error for the segmented reconstruction of Phantom 1. (b) True error.

$10^5$ initial photons per detector pixel, was applied. The dataset was then reconstructed twice, once using FBP (Fig. 3.6a), and once using 300 iterations of SIRT (Fig. 3.6d). The gray levels were again computed from the original reconstruction, using the Otsu segmentation classes.




**(a)** Segmented FBP **(b)** Reconstructed error **(c)** True error

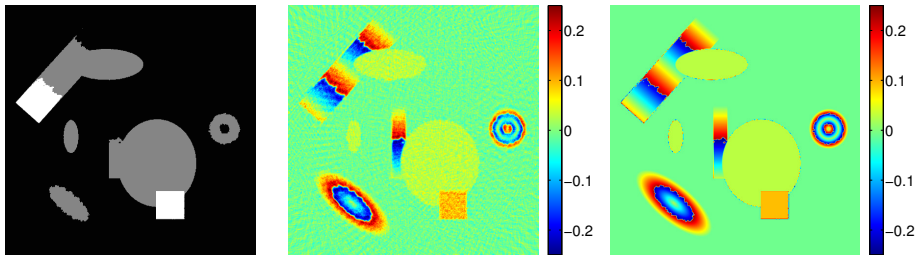
**(d)** Segmented SIRT **(e)** Reconstructed error **(f)** True error

**Figure 3.6:** (a) Segmented FBP reconstruction of Phantom 2. (b) Reconstructed residual error of (a). (c) True error of (a). (d) Segmented SIRT reconstruction of Phantom 2. (e) Reconstructed residual error of (d). (f) True error of (d).

As before, the reconstructed residual error (Figs. 3.6b and 3.6e) is close to the true error (Figs. 3.6c and 3.6f, respectively). The gray level errors are apparently quite small, in particular for the segmented SIRT reconstruction (Fig. 3.6e). However, we have not increased the gray level errors artificially to make the result more clear, since we also use this result in Section 3.3.2 to show numerically that it can still be used to correct the gray levels.

For the final experiment in this Section, a synthetic dataset was created from Phantom 3 (Fig. 3.2c), again using 90 parallel beam projections, evenly spaced at 2° intervals. Poisson noise, corresponding to $10^5$ initial photons per detector pixel, was applied. This synthetic dataset was then reconstructed using TVMin, and segmented using Otsu's method (Fig. 3.7a). The gray levels in Fig. 3.7a were, as before, determined using (3.2) on the pixels of the original reconstruction, with the Otsu segmentation classes as $Y_k$.

From Fig. 3.7a, it is not clear which objects are homogeneous and which are not. However, from the reconstructed residual error (Fig. 3.7b) this is obvious, since the non-homogeneous objects have large structured errors. Objects that seem merged in the segmented reconstruction of Fig. 3.7a can also be recognized as separate in Fig. 3.7b. The reconstructed residual error is again quite close to the true error (Fig. 3.7c).




**(a)** Segm. TVMin **(b)** Reconstructed error **(c)** True error

**Figure 3.7:** (a) Segmented TVMin reconstruction of Phantom 3. (b) Reconstructed residual error of (a). (c) True error.

### 3.3.2 Application—Improving Gray Level Estimates

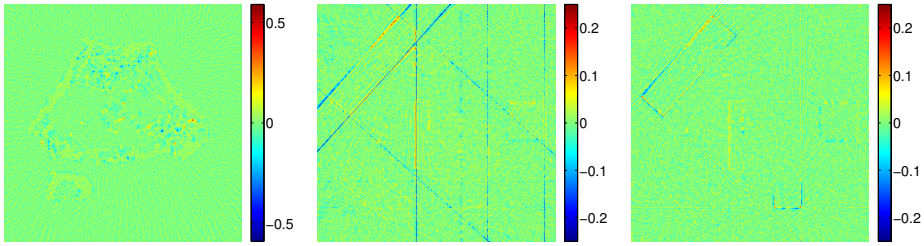
In this Section, we demonstrate that the gray level estimates of a segmented reconstruction can be improved using the reconstructed residual error. Visually, Figs. 3.3c, 3.5a, and 3.6b already suggest that the gray levels of the different objects are not correct, since there is a clear and relatively uniform error inside the objects. In Fig. 3.6f, this effect is less obvious, but Table 3.1 shows that this result can still be used to correct the gray levels.

Table 3.1 provides a numerical overview of the corrections that were achieved for the experiments of Section 3.3.1. The true gray levels $\rho_k$ were taken from the phantoms (Fig. 3.2), with $k = 1, 2$ for Phantoms 1 and 4, and $k = 1, 2, 3$ for Phantom 2. The estimated gray levels $\hat{\rho}_k$ are the gray levels of the segmented reconstructions, computed using (3.2) from the original reconstructions. The estimated gray level errors $\hat{\varepsilon}_k$ were computed from the reconstructed residual error using (3.2), the same procedure as for the $\hat{\rho}_k$. The sum of $\hat{\rho}_k$ and $\hat{\varepsilon}_k$ is shown as the corrected gray level $\hat{\rho}'_k$.

| Reconstruction | Param. | $k = 1$ | $k = 2$ | $k = 3$ |
|---|---|---|---|---|
| Phantom 4 | $\rho_k$ | 0.000 | 1.000 | N/A |
| CGLS | $\hat{\rho}_k$ | 0.004 | 0.985 | N/A |
| | $\hat{\varepsilon}_k$ | −0.004 | 0.015 | N/A |
| | $\hat{\rho}'_k$ | 0.000 | 1.000 | N/A |
| Phantom 1 | $\rho_k$ | 0.000 | 1.000 | N/A |
| SIRT | $\hat{\rho}_k$ | 0.012 | 0.904 | N/A |
| | $\hat{\varepsilon}_k$ | −0.011 | 0.071 | N/A |
| | $\hat{\rho}'_k$ | 0.000 | 0.975 | N/A |
| Phantom 2 | $\rho_k$ | 0.000 | 0.502 | 1.000 |
| FBP | $\hat{\rho}_k$ | −0.005 | 0.474 | 0.973 |
| | $\hat{\varepsilon}_k$ | 0.004 | 0.024 | 0.024 |
| | $\hat{\rho}'_k$ | −0.001 | 0.498 | 0.998 |
| Phantom 2 | $\rho_k$ | 0.000 | 0.502 | 1.000 |
| SIRT | $\hat{\rho}_k$ | 0.003 | 0.495 | 0.992 |
| | $\hat{\varepsilon}_k$ | −0.003 | 0.006 | 0.010 |
| | $\hat{\rho}'_k$ | −0.000 | 0.502 | 1.002 |

**Table 3.1:** Improving gray level estimates, showing true gray levels ($\rho_k$), estimated gray levels ($\hat{\rho}_k$), estimated gray level errors ($\hat{\varepsilon}_k$), and corrected gray levels ($\hat{\rho}'_k$).

From comparison of the corrected gray levels $\hat{\rho}'_k$ with the true gray levels $\rho_k$ from the phantom, it is clear that $\hat{\varepsilon}_k$ is a good estimate of the difference between $\hat{\rho}_k$ and $\rho_k$, and that it can be used to correct $\hat{\rho}_k$. The corrections that are applied to the gray levels are relatively small. However, they still have a visible effect on the segmented reconstructions. This is clear from the reconstructed residual error (Fig. 3.8) for reconstructions in which the corrected gray levels were used.




**(a)** Phantom 1, SIRT **(b)** Phantom 2, FBP **(c)** Phantom 2, SIRT

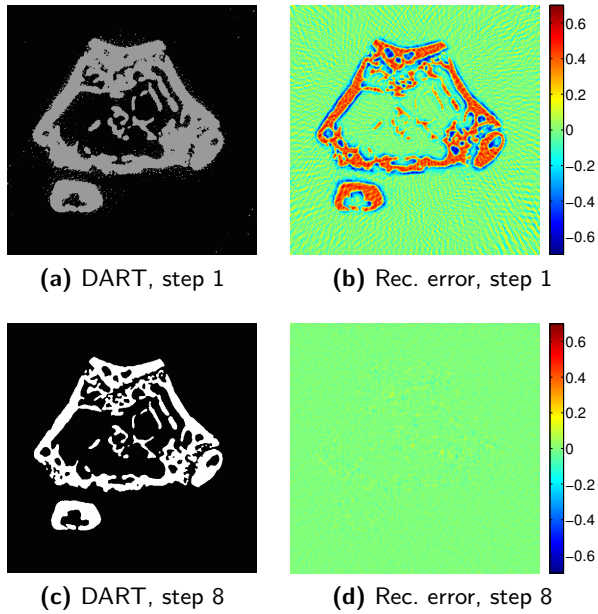
**Figure 3.8:** Reconstructed residual error for reconstructions that use corrected gray levels. (a) Phantom 1, using SIRT. (b) Phantom 2, using FBP. (c) Phantom 2, using SIRT.

### 3.3.3 Application—Improving Segmentation

In the application of the previous Section, the reconstructed residual error was used to improve the gray level estimates. In this second application, we show how these improved gray levels can improve the segmentation itself, for reconstruction algorithms that exploit a priori knowledge of those gray levels. The simulations in this Section are based on Phantoms 1 and 2. From both phantoms, synthetic datasets were created as before, using 90 projections and a detector with 512 pixels. Poisson noise, corresponding to $10^5$ initial photons per detector pixel, was applied to the datasets. Both datasets were then reconstructed with DART, using gray levels that were different from their value in the phantom, to simulate the practical situation where the density of a scanned object is not known precisely.

For Phantom 1, the gray level of the "bone" material was set to 0.6 instead of 1.0 (the value from the phantom). This results in a reconstruction (Fig. 3.9a) in which the bone structures are clearly too wide. This is confirmed by the reconstructed residual error (Fig. 3.9b). The positive error in the interior of the material indicates that its gray level is too low. The blue regions of negative errors at the edges of the material indicate that it is too wide (i.e., the negative error suggests that bone is not the appropriate material for those regions). As before, the gray levels can be corrected by using (3.2). Moreover, since DART uses the gray levels as prior knowledge, a new reconstruction can be computed, using the corrected values. If this procedure is repeated, it can lead to a very accurate reconstruction (Fig. 3.9c shows the result after eight steps), for which the reconstructed residual error does not show a systematic error anymore (Fig. 3.9d). The evolution of the gray levels (Fig. 3.10, plots for Phantom 1) shows that they quickly approach the true values. The rNMP (Fig. 3.12, plot for Phantom 1) also shows quick improvement during the first few steps of this procedure.

The results for Phantom 2 are similar. For the reconstruction of this phantom,




**(a)** DART, step 1 **(b)** Rec. error, step 1

**(c)** DART, step 8 **(d)** Rec. error, step 8

**Figure 3.9:** (a) Initial DART reconstruction of Phantom 1, using an incorrect gray level. (b) Reconstructed residual error for (a). (c) DART reconstruction after eight correction steps. (d) Reconstructed residual error for (c).

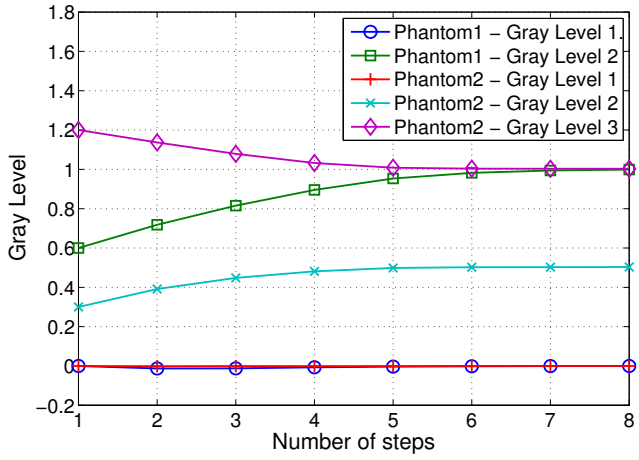
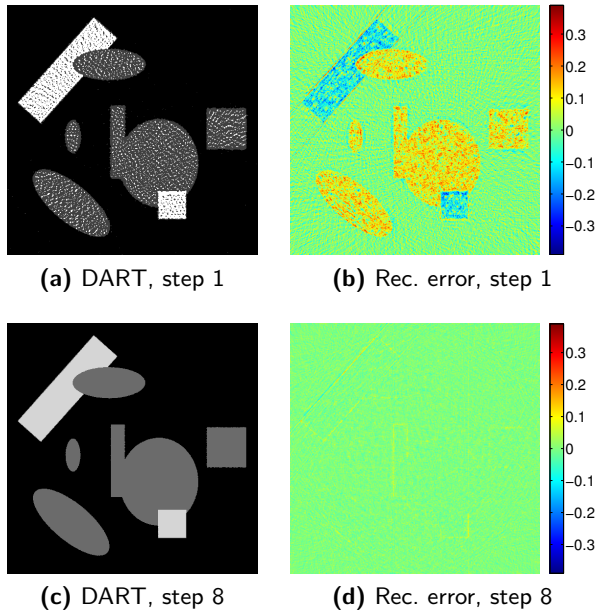



**Figure 3.10:** Gray levels of the DART reconstructions of Phantoms 1 and 2.

the gray levels of both materials were set to an incorrect value (0.3 instead of 0.5 and 1.2 instead of 1.0). This leads to a reconstruction (Fig. 3.11a) in which the free pixels in the interior of the objects deviate from the surrounding fixed pixels, which already suggests that the gray levels are not correct. This is confirmed by the reconstructed residual error (Fig. 3.11b), which clearly shows that the middle gray level is too low (the error is positive) and that the brightest gray level is too high (the error is negative). The gray levels were again corrected using (3.2), and DART was rerun using the new values. As for Phantom 1, repeating this procedure quickly leads to a more accurate estimate of the gray levels (Fig. 3.10, plots for Phantom 2), and reduces the rNMP (Fig. 3.12, plot for Phantom 2).




**(a)** DART, step 1 **(b)** Rec. error, step 1

**(c)** DART, step 8 **(d)** Rec. error, step 8

**Figure 3.11:** (a) Initial DART reconstruction of Phantom 2, using incorrect gray levels. (b) Reconstructed residual error for (a). (c) DART reconstruction after eight correction steps. (d) Reconstructed residual error for (c).

### 3.3.4 Application—Selecting the Most Accurate Segmentation

In this experiment, the reconstructed residual error was computed for experimental cone beam data. A dataset of an ex vivo mouse femur was acquired with a SkyScan 1172 $\mu$CT scanner using 376 projections at 0.5° intervals, at a detector resolution of 5 $\mu$m. The SkyScan NRecon software package was used to correct for ring and

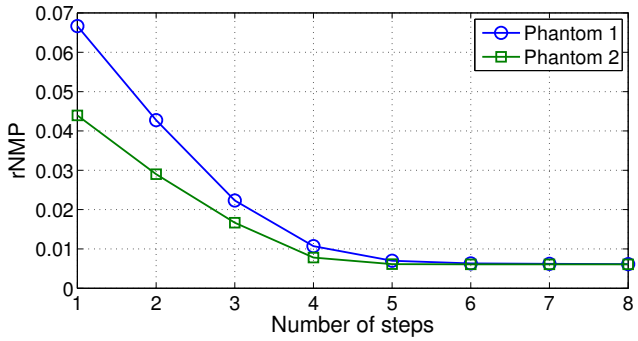
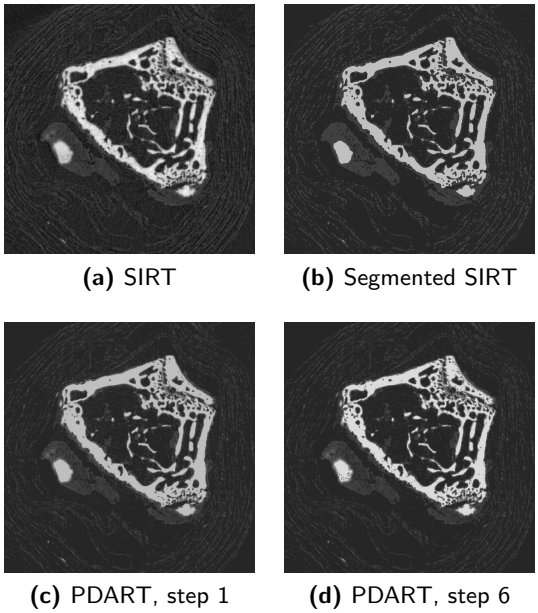



**Figure 3.12:** Relative number of misclassified pixels for the DART reconstructions of Phantoms 1 and 2.

beam-hardening artifacts.

The dataset was reconstructed using two different algorithms. First, it was reconstructed using 300 iterations of SIRT (Fig. 3.13a) and segmented using Otsu's method (Fig. 3.13b). The gray levels of the segmented reconstruction were estimated using (3.2) and the SIRT reconstruction. Second, it was reconstructed using 300 iterations of PDART (Fig. 3.14c), using the gray level from the bone material of the segmented SIRT reconstruction as prior knowledge. The partially discrete PDART algorithm was used instead of the fully discrete DART, because of the non-discrete background of the sample.

For the segmented SIRT reconstruction, the reconstructed residual error (Fig. 3.14a) shows that the gray level for the bone is underestimated. Corrections for the gray levels were computed by applying (3.2) to the reconstructed residual error of Fig. 3.14a. With corrected gray levels for the segmented tomogram, the reconstructed residual error (Fig. 3.14b) does not show a systematic error anymore, in the sense that the bone now contains both positive and negative errors. For the initial PDART reconstruction (Fig. 3.13c), the reconstructed residual error (Fig. 3.14c) also shows that the gray level for the bone is underestimated. Additionally, Fig. 3.14c suggests that the bone structures are too wide, since they have a region of negative errors at their edges (compare with Fig. 3.9b, where this effect is also visible). From Fig. 3.14c, the gray levels can again be corrected using (3.2). Fig. 3.13d shows the resulting PDART reconstruction after six such correction steps. It is clear from Fig. 3.14d that this final PDART reconstruction is the most accurate of all four.



**(a)** SIRT **(b)** Segmented SIRT

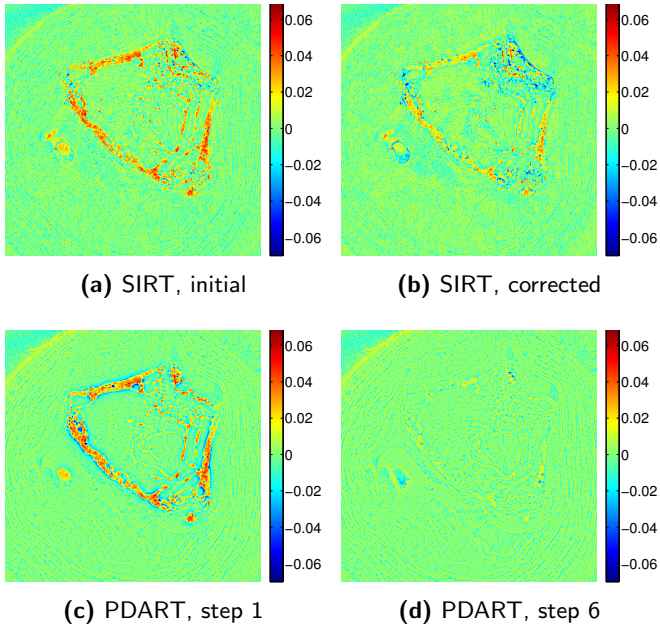
**(c)** PDART, step 1 **(d)** PDART, step 6

**Figure 3.13:** (a) SIRT reconstruction of a single slice of the femur dataset. (b) The reconstruction from (a), segmented using Otsu's method. (c) PDART reconstruction, using the initial gray level estimate. (d) PDART reconstruction after six correction steps.

## 3.4 Discussion

Most existing segmentation algorithms that are used in tomography do no exploit the projection data during segmentation. The few algorithms that do exploit this information [3, 4], use it to compute a quality measure that is a single number (the *projection distance* in [3] and the *segmentation inconsistency* in [4]). The segmentation is then optimized by minimizing this number. In contrast, the reconstructed residual error creates a spatial map of the segmentation quality. This allows studying local variations of the error, which is not possible using a single number.

The reconstructed residual error can exploit the projection data because the artifacts of the reconstruction, which are typically largely in the null space of the system matrix (Figs. 3.3j and 3.4c), are largely in the row space after the segmentation step (Figs. 3.3b and 3.4e). Computing the exact row space component of the segmentation artifacts is infeasible in practice, due to computational constraints and noise. However, a regular reconstruction algorithm such as SIRT can be substituted without much degradation of the results (compare Figs. 3.3c and 3.5a with




**(a)** SIRT, initial **(b)** SIRT, corrected

**(c)** PDART, step 1 **(d)** PDART, step 6

**Figure 3.14:** Reconstructed residual error for the femur dataset for (a) SIRT with the initial gray level estimate, (b) SIRT with the corrected gray level estimate, (c) PDART with the initial gray level estimate, and (d) PDART after six correction steps.

Figs. 3.3b and 3.4e, respectively). This allows adding the reconstructed residual error as a new step in the tomographic workflow, including for large practical datasets.

The reconstructed residual error seems able to detect even small errors in the gray levels (e.g., Figs. 3.6b and 3.6e), as is evident from Table 3.1. Large errors in the gray levels are visible too (e.g., Fig. 3.9b), but more important in this case is that structural problems in the original reconstruction (Fig. 3.9a) are also revealed. Non-homogeneous regions in the scanned object can be detected (Fig. 3.7b). In the particular case of an algorithm that uses a priori knowledge of the gray levels, the segmentation itself can be improved (Figs. 3.9 and 3.11). Moreover, the correction step can be repeated for added improvements, as is illustrated by the evolution of the gray levels (Fig. 3.10) and the rNMP (Fig. 3.12).

The results for experimental $\mu$CT data illustrate another type of analysis that is enabled by the proposed method. After correction of the initial gray levels of the segmented SIRT reconstruction (Fig. 3.13b), the reconstructed residual error (Fig. 3.14b) suggests an uneven density of the bone. The conclusion might be that this effect is real, or that it is due to the scanning process. However, when

the PDART algorithm is used (Fig. 3.13d), the uneven density almost completely disappears (Fig. 3.14d). Hence, this segmentation might be preferable, since it *does* correspond with the bone material having a homogeneous density. Given the relatively small differences between the segmented images (Figs. 3.13b and 3.13d), it is not obvious how to choose a preferred one between them without access to the reconstructed residual error.

## 3.5 Conclusions

We have introduced the reconstructed residual error, a simple way to analyze and improve the segmentation quality of a tomogram. We have used the Moore-Penrose pseudoinverse as a mathematical model for investigating the properties of the technique, and then generalized this approach to practical reconstruction algorithms.

The reconstructed residual error provides an accurate map of the errors in a segmented tomogram. Visually, this map can be used to study the distribution of the errors. Furthermore, it can be used to improve the gray level estimates, and, for certain specific algorithms, the segmentation itself. The method is applicable to experimental datasets.

From a practical point of view, the technique is trivial to implement, since the only necessary tools are a forward projector and a reconstruction algorithm. The computational cost of the basic algorithm is modest, since only a single forward projection and a single reconstruction are needed. These practical aspects should lower the threshold for adopting the method.

## Acknowledgments

The mouse femur dataset is courtesy of Phil Salmon, Bruker microCT. This work was financially supported by the IWT TomFood project (IWT is the agency for Innovation by Science and Technology–Flanders, Belgium) and by the NWO (the Netherlands Organisation for Scientific Research–The Netherlands, research programme 639.072.005).

## References

[1] N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern Recogn.*, vol. 26, no. 9, pp. 1277–1294, 1993.

[2] D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," *Annu. Rev. Biomed. Eng.*, vol. 2, pp. 315–337, 2000.

[3] K. J. Batenburg and J. Sijbers, "Optimal threshold selection for tomogram segmentation by projection distance minimization," *IEEE Trans. Med. Imag.*, vol. 28, no. 5, pp. 676–686, 2009.

[4] W. van Aarle, K. J. Batenburg, and J. Sijbers, "Optimal threshold selection for segmentation of dense homogeneous objects in tomographic reconstructions," *IEEE Trans. Med. Imag.*, vol. 30, no. 4, pp. 980–989, 2011.

[5] K. J. Batenburg, S. Bals, J. Sijbers, C. Kübel, P. A. Midgley, J. C. Hernandez, U. Kaiser, E. R. Encina, E. A. Coronado, and G. Van Tendeloo, "3D imaging of nanomaterials by discrete tomography," *Ultramicroscopy*, vol. 109, no. 6, pp. 730–740, 2009.

[6] K. J. Batenburg and J. Sijbers, "DART: A practical reconstruction algorithm for discrete tomography," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2542–2553, 2011.

[7] T. Roelandts, K. J. Batenburg, and J. Sijbers, "PDART: A partially discrete algorithm for the reconstruction of dense particles," in *Proc. 11th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine (Fully 3D)*, Potsdam, Germany, 2011, pp. 451–454.

[8] T. Roelandts, K. J. Batenburg, E. Biermans, C. Kübel, S. Bals, and J. Sijbers, "Accurate segmentation of dense nanoparticles by partially discrete electron tomography," *Ultramicroscopy*, vol. 114, pp. 96–105, 2012.

[9] W. van Aarle, K. J. Batenburg, and J. Sijbers, "Automatic parameter estimation for the discrete algebraic reconstruction technique (DART)," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4608–4621, 2012.

[10] G. Strang, *Introduction to Linear Algebra.* Wellesley, MA, USA: Wellesley-Cambridge Press, 2009.

[11] M. Beister, D. Kolditz, and W. A. Kalender, "Iterative reconstruction methods in X-ray CT," *Phys. Medica*, vol. 28, pp. 94–108, 2012.

[12] S. Moehrs, M. Defrise, N. Belcari, A. Del Guerra, A. Bartoli, S. Fabbri, and G. Zanetti, "Multi-ray-based system matrix generation for 3D PET reconstruction," *Phys. Med. Biol.*, vol. 53, no. 23, pp. 6925–6945, 2008.

[13] Y. Long, J. A. Fessler, and J. M. Balter, "3D forward and back-projection for X-ray CT using separable footprints," *IEEE Trans. Med. Imag.*, vol. 29, no. 11, pp. 1839–1850, 2010.

[14] I. A. Elbakri and J. A. Fessler, "Segmentation-free statistical image reconstruction for polyenergetic x-ray computed tomography with experimental validation," *Phys. Med. Biol.*, vol. 48, no. 15, pp. 2453–2477, 2003.

[15] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assist. Tomogr.*, vol. 8, no. 2, pp. 306–316, 1984.

[16] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography," *J. Theor. Biol.*, vol. 29, no. 3, pp. 471–481, 1970.

[17] A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm," *Ultrason. Imag.*, vol. 6, pp. 81–94, 1984.

[18] P. Gilbert, "Iterative methods for the three-dimensional reconstruction of an object from projections," *J. Theor. Biol.*, vol. 36, no. 1, pp. 105–117, 1972.

[19] Y. Saad, *Iterative Methods for Sparse Linear Systems*. Philadelphia, PA, USA: SIAM, 2003.

[20] A. van der Sluis and H. A. van der Vorst, "SIRT- and CG-type methods for the iterative solution of sparse linear least-squares problems," *Linear Algebra Appl.*, vol. 130, pp. 257–303, 1990.

[21] J. Gregor and T. Benson, "Computational analysis and improvement of SIRT," *IEEE Trans. Med. Imag.*, vol. 27, no. 7, pp. 918–924, 2008.

[22] S. Kiencke, Y. M. Levakhina, and T. M. Buzug, "Greedy projection access order for sart simultaneous algebraic reconstruction technique," in *Proc. Bildverarbeitung für die Medizin*, ser. Informatik aktuell, Heidelberg, Germany, 2013, pp. 93–98.

[23] G. Van Eyndhoven, J. Sijbers, and J. Batenburg, "Combined motion estimation and reconstruction in tomography," in *Proc. 12th European Conference on Computer Vision (ECCV)*, ser. Lect. Notes Comput. Sc., Florence, Italy, 2012, vol. 7583, pp. 12–21.

[24] W. Xu, F. Xu, M. Jones, B. Keszthelyi, J. Sedat, D. Agard, and K. Mueller, "High-performance iterative electron tomography reconstruction with long-object compensation using graphics processing units (GPUs)," *J. Struct. Biol.*, vol. 171, no. 2, pp. 142–153, 2010.

[25] W. J. Palenstijn, K. J. Batenburg, and J. Sijbers, "Performance improvements for iterative electron tomography reconstruction using graphics processing units (GPUs)," *J. Struct. Biol.*, vol. 176, no. 2, pp. 250–253, 2011.

[26] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[27] ——, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, 2009.

[28] S. Becker, J. Bobin, and E. J. Candès, "NESTA: A fast and accurate first-order method for sparse recovery," *SIAM J. Imaging Sci.*, vol. 4, no. 1, pp. 1–39, 2011.

[29] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, 1979.

# 4
# Localized Priors

**This chapter is in preparation for publication as**

T. Roelandts, K. J. Batenburg, and J. Sijbers, "Localized Priors in Tomography using the Reconstructed Residual Error," *IEEE Transactions on Image Processing*, 2013.

**Abstract**—Limited data problems form an important challenge in tomography. For objects that consist of homogeneous regions, incorporating certain priors in the reconstruction can reduce the necessary number of projections. However, existing reconstruction algorithms that exploit such prior knowledge do not work well when the scanned object also contains non-homogeneous regions. In this paper, we present a way to automatically apply a *localized* prior, by means of the recently proposed *reconstructed residual error*. The approach creates a map that indicates the errors in the reconstruction, which is then used to determine the mask of pixels for which the prior is likely to be unsuitable. The technique is applied to two practical reconstruction algorithms, namely DART and FISTA. Based on simulation experiments, it is shown that a localized prior results in more accurate reconstructions for these algorithms.

## 4.1 Introduction

Limited data problems form an important challenge in tomography. For reasons such as beam damage, scanning time, etc., it is sometimes difficult or impossible to gather enough data to create an acceptable reconstruction. A solution for this problem is to exploit prior knowledge about the scanned object during the reconstruction. This can reduce the amount of data that is needed for an accurate reconstruction to a feasible number of projections.

Objects that consist of only a few different materials that appear in homogeneous regions are particularly suitable for this approach. Several algorithms are available for reconstructing these kinds of objects, e.g., DART [1] and FISTA [2, 3]. DART is an algorithm for *discrete tomography* (DT), which assumes prior knowledge of the gray levels that represent each of the materials in the scanned object. An overview of the field of DT is available in [4, 5]. FISTA employs *total variation minimization* (TVMin), which assumes that the total variation of the scanned object is low. TVMin was introduced in [6] as a noise removal algorithm, and was later applied in tomography [7–9].

When applied to objects that contain not only homogeneous, but also non-homogeneous regions, algorithms such as DART and FISTA may result in incorrect reconstructions. For DART, reconstructing non-homogeneous regions will always produce incorrect results, since all pixels of the reconstruction will be forcibly set to one of the "known" gray levels. For FISTA, the result might be incorrect if the non-homogeneous regions are such that the total variation (TV) prior is not suitable, e.g., when very small structures are present. An additional problem is that it will not necessarily be obvious that the resulting reconstructions are incorrect. If DT or TVMin is applied to the reconstruction of objects that are expected to be homogeneous, any resulting smooth reconstruction might be deemed acceptable.

Recently, the *reconstructed residual error* was introduced as a way of detecting these kinds of problems [10]. The reconstructed residual error reconstructs the difference between the recorded data and the forward projection of a segmented tomogram. The result is a map that indicates the errors in that segmented tomogram. However, instead of simply not using algorithms such as DART or FISTA for objects that contain both homogeneous and non-homogeneous regions, it would be better if these algorithms could still be used for the part of the reconstruction for which their priors are valid, thereby possibly retaining some of their strengths.

In this paper, we present an approach that *localizes* the prior of algorithms from DT and TVMin, based on the reconstructed residual error. We augment the reconstruction area with a mask that defines where the prior is not valid. This mask is determined by assuming that regions for which the reconstructed residual error is higher than some given threshold do not sufficiently conform to the prior, and should be reconstructed without applying it. This approach allows reconstructing complex objects that contain both homogeneous and non-homogeneous regions from relatively limited data. In the current paper, we apply this general principle in practice to DART and FISTA by developing localized versions of those algorithms, but it could also be applied to other algorithms from DT or TVMin (e.g., NESTA [11], which is based on [12]).

The remainder of this paper is structured as follows. In Section 4.2, the technique is introduced, and applied to DART and FISTA. Section 4.3 reports on

the results of several simulation experiments. These results are discussed in Section 4.4, and a conclusion is reached in Section 4.5.

## 4.2 Methods

In this Section, we first present an overview of how the prior of a reconstruction algorithm can be localized using the reconstructed residual error. We then describe in general how we determine the regions for which a prior should not be applied, and show how a reconstruction algorithm can be adapted to use a local prior. This general principle is then applied to DART and FISTA.

### 4.2.1 Overview

Fig. 4.1 presents an overview of how a prior can be applied locally by using the reconstructed residual error to locate non-homogeneous regions in the reconstruction. The original projection data (Fig. 4.1b) is a sinogram, acquired by rotating around the object (Fig. 4.1a). In this example, the sinogram is reconstructed using FISTA (Fig. 4.1c) and segmented using a global threshold (Fig. 4.1f). The reconstructed residual error (Fig. 4.1i) is computed by reconstructing the difference (Fig. 4.1h) between the original projections (Fig. 4.1b) and the forward projection of the segmented reconstruction (Fig. 4.1e). The set of non-homogeneous pixels is then determined by thresholding the reconstructed residual error, resulting in a mask of pixels for which the prior should not be applied (Fig. 4.1g). The original projections (Fig. 4.1b) and this mask are then used as input for a modified FISTA algorithm that implements a local prior (Fig. 4.1d).

### 4.2.2 Notation and Concepts

The projection process in tomography can be modeled as a linear operator that is determined by the projection geometry. This leads to a system of linear equations,

$$\boldsymbol{W}\boldsymbol{x} = \boldsymbol{p}, \tag{4.1}$$

where $\boldsymbol{p} \in \mathbb{R}^m$ contains the projection data and $\boldsymbol{x} \in \mathbb{R}^n$ corresponds to the image. The linear operator is represented by the $m \times n$ matrix $\boldsymbol{W}$, the *projection matrix*. An approximate solution $\tilde{\boldsymbol{x}} \in \mathbb{R}^n$ of (4.1) can then be computed iteratively. A classical approach is the least squares (LS) estimate,

$$\hat{\boldsymbol{x}}_{\mathrm{LS}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{W}\boldsymbol{x} - \boldsymbol{p}\|^2, \tag{4.2}$$

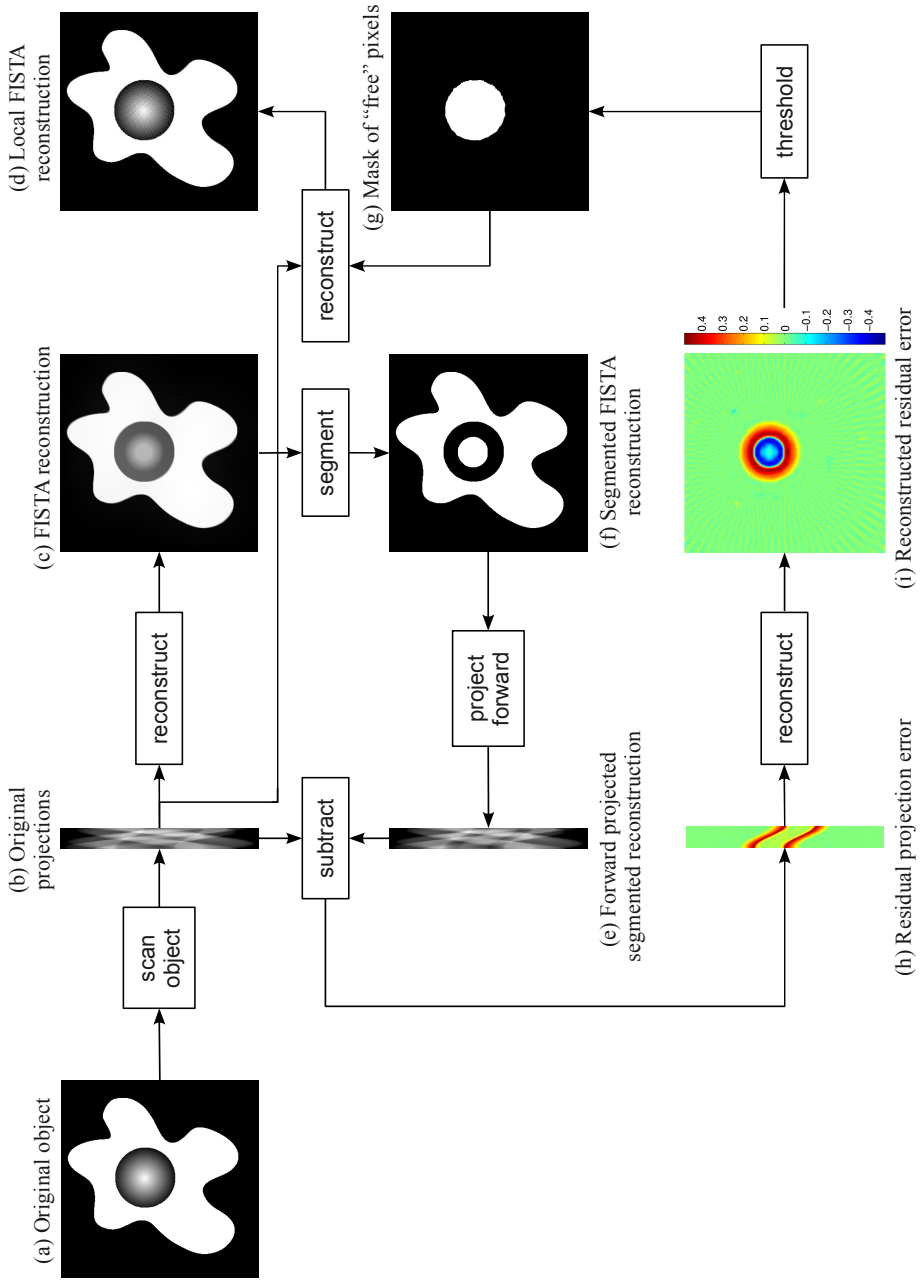



**Figure 4.1:** Overview of how a prior can be localized by using the reconstructed residual error to locate non-homogeneous regions in a reconstruction.

which minimizes the projection distance. In tomography, the well-known SIRT algorithm [13], which results in a *weighted* least squares solution [14], is often used. In this paper, SIRT is used both to reconstruct the residual error and as a subroutine of DART.

The final step in the tomographic reconstruction process is often *segmentation*. A segmentation method essentially partitions the pixels of an image into sets $Y_1, \ldots, Y_d$, where $d$ is the number of classes (segments) in the segmented image. A straightforward way to perform segmentation is *global thresholding*, which is also the technique that is employed by DART. Let $\{\rho_1, \ldots, \rho_d\}$ be a set of gray levels, with $\rho_1 < \rho_2 < \ldots < \rho_d$. Also define a set of thresholds $\{\tau_1, \ldots, \tau_{d-1}\}$. A *thresholding function* $r : \mathbb{R} \times \mathbb{R}^{d-1} \times \mathbb{R}^d \to \mathbb{R}$ can then be defined as

$$r(v, \tau, \rho) = \begin{cases} \rho_1 & \text{if } v < \tau_1 \\ \rho_2 & \text{if } \tau_1 \leq v < \tau_2 \\ \vdots & \\ \rho_d & \text{if } \tau_{d-1} \leq v. \end{cases} \tag{4.3}$$

The thresholding function for an *image* $\boldsymbol{x} \in \mathbb{R}^n$ is defined as

$$\boldsymbol{r}(\boldsymbol{x}, \tau, \rho) = (r(x_1, \tau, \rho) \ \cdots \ r(x_n, \tau, \rho))^T. \tag{4.4}$$

From the segmentation classes $Y_k$, a segmented image $\boldsymbol{s} \in \mathbb{R}^n$ can be created by assigning a gray level $\rho_k$ to all pixels in the set $Y_k$, for each $k \in \{1, \ldots, d\}$. For global thresholding as defined in (4.4), which we use in this paper, such a segmented image is the output of the thresholding step. However, other segmentation algorithms can also be used, if all pixels in each class $Y_k$ are assigned the same gray level $\rho_k$ afterwards. If the values of $\rho_1, \ldots, \rho_d$ are not known, the mean of all pixel values in a class can be used as an estimate,

$$\hat{\rho}_k = \frac{1}{|Y_k|} \sum_{y \in Y_k} y, \text{ for each } k \in \{1, \ldots, d\}. \tag{4.5}$$

The *reconstructed residual error*, introduced in [10], is a technique that creates a visual map of the errors in a segmented tomogram. Inputs are a segmented tomogram $\boldsymbol{s}$ and the projection data $\boldsymbol{p}$. First, $\boldsymbol{s}$ is forward projected to give $\boldsymbol{p_s} \in \mathbb{R}^m$. The system of linear equations $\boldsymbol{W e} = \boldsymbol{p} - \boldsymbol{p_s}$, where $\boldsymbol{e} \in \mathbb{R}^n$ corresponds to the (unknown) error image, can then be solved to produce an approximate solution $\tilde{\boldsymbol{e}} \in \mathbb{R}^n$, the reconstructed residual error.

We define a *local* (or *localized*) *prior* as a prior that is only applied to part of the reconstruction area. When the word *local* is used with respect to a reconstruction

algorithm, e.g., *Local DART*, it means that the prior of the original algorithm is only applied locally.

### 4.2.3 Locating Non-Homogeneous Regions

Here, we describe how the reconstructed residual error is used to determine the mask of non-homogeneous pixels. For this general presentation of the approach, we assume that a segmented tomogram exists. Sections 4.2.5 and 4.2.6 below describe how that is determined for the particular cases of DART and FISTA, respectively.

The algorithm that determines the mask of non-homogeneous pixels is shown in Fig. 4.2. Let $\boldsymbol{s} \in \mathbb{R}^n$ be a segmented tomogram. The reconstructed residual error $\boldsymbol{e}$ can then be computed from $\boldsymbol{s}$ and $\boldsymbol{p}$. Define a mask $\boldsymbol{m} \in \mathbb{R}^n$, with $m_j \in \{0, 1\}$ for each $j \in \{1, \ldots, n\}$, that indicates the location of the homogeneous ($m_j = 0$) and non-homogeneous pixels ($m_j = 1$). For convenience, we also define the *set* $M \subseteq \{1, \ldots, n\}$ of non-homogeneous pixels (i.e., for which $m_j = 1$). The mask $\boldsymbol{m}$ (and the corresponding set $M$) can be determined from $\boldsymbol{e}$ in a straightforward manner, by setting $\boldsymbol{m} = \boldsymbol{r}(|\boldsymbol{e}|, c, \{0, 1\})$, for a given threshold $c \in \mathbb{R}$, and where $|\cdot|$ denotes element-wise absolute value. The optimal value of $c$ is different for each reconstruction, since $|\boldsymbol{e}|$ depends on $\boldsymbol{s}$ and $\boldsymbol{p}$. Morphological operations on $M$ are optional, but can be helpful to correct for imperfections in $\boldsymbol{e}$. In all experiments in this paper, the morphological operations that are shown in Fig. 4.2 were applied.

### 4.2.4 Reconstruction Algorithms with a Local Prior

After locating the non-homogeneous regions of an object, i.e., the set $M$, using the reconstructed residual error, the next step is modifying the reconstruction algorithm to take this information into account. The implementation of this is different for each algorithm. The general approach is to modify the reconstruction algorithm to no longer apply its prior to the whole reconstruction $\boldsymbol{x}$, but only to the pixels $x_j$ for which $j \notin M$. In Sections 4.2.5 and 4.2.6, this principle is applied to *DART* and *FISTA*, respectively.

The set $M$ of non-homogeneous pixels must be computed from a segmented reconstruction $\boldsymbol{s}$ (Fig. 4.2). There are two approaches for this. The first is computing a suitable $\boldsymbol{s}$ from an intermediate reconstruction $\boldsymbol{x}$, *during* the reconstruction, and possibly repeating that several times. The adapted version of DART that is presented in Section 4.2.5 uses this approach. A second approach is to create an initial reconstruction $\boldsymbol{x}^0$ first, using an unmodified reconstruction algorithm. After segmenting $\boldsymbol{x}^0$ as $\boldsymbol{s}^0$, it can then be used to determine $M$. The adapted version of FISTA that is presented in Section 4.2.6 uses this approach.

**Input:**

projection data $\boldsymbol{p}$

segmented reconstruction $\boldsymbol{s}$

error threshold $c$

**Algorithm:**

Reconstruct the residual error $\boldsymbol{e}$ of $\boldsymbol{s}$

Compute the mask $\boldsymbol{m} := \boldsymbol{r}(|\boldsymbol{e}|, c, \{0, 1\})$

Determine the set $M$ from $\boldsymbol{m}$ by including all pixels for which $m_j = 1$

Perform optional morphological operations to clean up $M$; in the current paper, the following was done

1) Remove small clusters
2) Dilate with a $4 \times 4$ square as structuring element
3) Remove small holes
4) Erode with a $4 \times 4$ square as structuring element

**Output:** $M$

**Figure 4.2:** Algorithm that uses the reconstructed residual error to determine the set $M$ of non-homogeneous pixels, starting from a segmented reconstruction.

### 4.2.5 Application—Local DART

We give an overview of the DART algorithm here, and refer to [1] for a detailed description. The prior that DART uses is that the scanned object is assumed to consist of only a few different materials that appear in homogeneous regions, and that the gray level for each of these materials is known. The idea behind DART is illustrated in Fig. 4.3. A synthetic dataset was created from a phantom (Fig. 4.3a), using only five equiangular projections. If this dataset is reconstructed with SIRT (Fig. 4.3b), the quality is very poor. However, if the SIRT reconstruction is thresholded (Fig. 4.3c) using the a priori known gray levels, it appears that the pixels that are not too close to the boundary of the object are correctly classified as either object or background. DART exploits this observation by first determining that boundary (Fig. 4.3d), and then only updating the boundary pixels in subsequent iterations. Repeating this a number of times results in a final reconstruction (Fig. 4.3e) that is very accurate.

DART implements its prior by interleaving iterations of an algebraic reconstruction method (SIRT is used in this paper) with segmentation steps. In each of those steps, a segmented image $\boldsymbol{s}$ is determined from the intermediate reconstruc-

tion $\boldsymbol{x}$ by setting $\boldsymbol{s} = \boldsymbol{r}(\boldsymbol{x}, \tau, \rho)$, where $\rho = \{\rho_1, \ldots, \rho_d\}$ is the set of a priori known gray levels, and $\tau_k = (\rho_k + \rho_{k+1})/2$, for each $k \in \{1, \ldots, d-1\}$. The pixels of $\boldsymbol{s}$ are then divided into two groups. The first group, the set of *free pixels* $F \subseteq \{1, \ldots, n\}$, contains the boundary pixels of the segmented reconstruction. The second group contains the remaining pixels, which are kept fixed during subsequent SIRT iterations. Fixing pixels, which decreases the number of variables in the reconstruction problem while still using all recorded data, is done as follows. Consider the system of linear equations

$$\begin{pmatrix} | & & | \\ \boldsymbol{w}_1 & \cdots & \boldsymbol{w}_n \\ | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \boldsymbol{p}. \tag{4.6}$$

A pixel $x_j$ can be fixed at the value $v_j \in \mathbb{R}$ by transforming (4.6) into

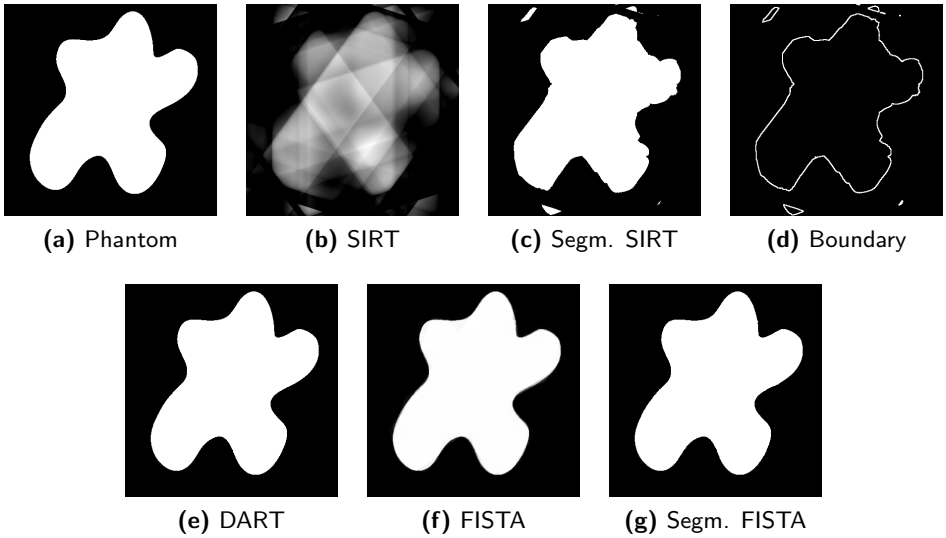
$$\begin{pmatrix} | & & | & | & & | \\ \boldsymbol{w}_1 & \cdots & \boldsymbol{w}_{j-1} & \boldsymbol{w}_{j+1} & \cdots & \boldsymbol{w}_n \\ | & & | & | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_{j-1} \\ x_{j+1} \\ \vdots \\ x_n \end{pmatrix} = \boldsymbol{p} - v_j \boldsymbol{w}_j. \tag{4.7}$$

With all the pixels in $F$ fixed in this way, the next SIRT iterations are then run on only the boundary pixels. After these SIRT iterations, the boundary pixels are smoothed (see again [1]) before the reconstruction is segmented again and the next DART iteration starts. Over several DART iterations, the position of the border is optimized, resulting in a discrete reconstruction that incorporates the prior knowledge about the gray levels.

To *localize* DART, the algorithm must be changed in two ways. First, the set $M$ of non-homogeneous pixels must be determined at some point. As mentioned in Section 4.2.4, this is done *during* the reconstruction for DART. Second, the algorithm must be adapted to stop applying its prior to the pixels of $M$. An overview of the *Local DART* algorithm is shown in Fig. 4.4.

For computing $M$, a segmented reconstruction is needed (Fig. 4.2). However, DART already computes such a reconstruction during its regular iterations, in the statement $\boldsymbol{s}^t := \boldsymbol{r}(\boldsymbol{x}^t, \tau, \rho)$ (Fig. 4.4). Hence, the computation of $M$ can simply be added at that location in the algorithm.

Not applying the prior for the pixels of $M$ is also straightforward, since DART already maintains a set $F$ of pixels that are not kept fixed during DART iterations. Hence, $M$ can simply be added to $F$ after each segmentation step (the statement $F^t := F^t \cup M^t$ in Fig. 4.4). If the set $M$ is known a priori, it can be specified



**(a)** Phantom **(b)** SIRT **(c)** Segm. SIRT **(d)** Boundary

**(e)** DART **(f)** FISTA **(g)** Segm. FISTA

**Figure 4.3:** Demonstration of the DART and FISTA algorithms. (a) Phantom image. (b) SIRT reconstruction. (c) Segmented SIRT reconstruction. (d) Boundary pixels of (c). (e) DART reconstruction. (f) FISTA reconstruction. (g) Segmented FISTA reconstruction. (b–e) were created from a synthetic dataset with five equiangular projections, (f–g) from one with ten.

as an input parameter. Setting $M$ to the empty set recovers the original DART algorithm. Depending on how $M$ is obtained, two variants are discerned, namely *Automatic Local DART* (if $M$ is computed) and *Manual Local DART* (if $M$ is known a priori).

### 4.2.6 Application—Local FISTA

We give an overview of the FISTA algorithm here, and refer to [3] for a detailed description. The prior that FISTA uses is that the scanned object is assumed to have a low total variation. This is done by adding a TV regularization term to the LS approach of (4.2). This results in

$$\hat{\boldsymbol{x}}_{\mathrm{FISTA}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{W}\boldsymbol{x} - \boldsymbol{p}\|^2 + 2\lambda\|\boldsymbol{x}\|_{\mathrm{TV}}. \tag{4.8}$$

Here, $\|\cdot\|_{\mathrm{TV}}$ is the TV seminorm. The regularization parameter $\lambda$ provides a trade-off between the LS and TV parts of the equation. The general nonsmooth convex optimization problem that corresponds to (4.8) is the minimization of the

**Input:**

projection data $\boldsymbol{p}$

thresholds $\tau$ and gray levels $\rho$

error threshold $c$ ***or*** set of free pixels $M$

**Algorithm:**

**if** $M$ is not provided

  Set $M^0$ to the empty set

**else**

  $M^0 := M$

**endif**

$t := 0$

Compute an initial SIRT reconstruction $\boldsymbol{x}^0$ and an initial segmented image $\boldsymbol{s}^0 := \boldsymbol{r}(\boldsymbol{x}^0, \tau, \rho)$

**while** stop condition is not met

  Initialize the set $F^t$ with the boundary pixels of $\boldsymbol{s}^t$

  $F^t := F^t \cup M^t$

  Compute the image $\boldsymbol{y}^t$ from $\boldsymbol{x}^t$ and $\boldsymbol{s}^t$, setting $y_j^t := x_j^t$ if $j \in F^t$ and $y_j^t := s_j^t$ otherwise

  Compute the SIRT reconstruction $\boldsymbol{x}^{t+1}$, with $\boldsymbol{y}^t$ as the start solution, while keeping the pixels not in $F^t$ fixed

  Smooth the pixels of $\boldsymbol{x}^{t+1}$ that are in $F^t$

  $t := t + 1$

  Compute the segmented image $\boldsymbol{s}^t := \boldsymbol{r}(\boldsymbol{x}^t, \tau, \rho)$

  **if** $M$ is not provided

    Compute $M^t$ from $\boldsymbol{s}^t$, $\boldsymbol{p}$, and $c$ (Fig. 4.2)

  **else**

    $M^t := M^{t-1}$

  **endif**

**endwhile**

Compute the image $\boldsymbol{z}^t$ from $\boldsymbol{x}^t$ and $\boldsymbol{s}^t$, setting $z_j^t := x_j^t$ if $j \in M^t$ and $z_j^t := s_j^t$ otherwise

**Output:** $\boldsymbol{z}^t$

**Figure 4.4:** The Local DART algorithm. All statements that use or compute $M$ are new, the rest of the algorithm is identical to the original DART, which is recovered by setting $M$ to the empty set.

sum of two convex functions

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) + g(\boldsymbol{x}), \tag{4.9}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is smooth and $g : \mathbb{R}^n \to (-\infty, +\infty]$ is nonsmooth. In (4.8), we have $f(\boldsymbol{x}) \equiv \|\boldsymbol{W}\boldsymbol{x} - \boldsymbol{p}\|^2$ and $g(\boldsymbol{x}) \equiv 2\lambda\|\boldsymbol{x}\|_{\mathrm{TV}}$. The key assumption that is used in the current paper to adapt FISTA is that $g$ is a closed proper convex function. See [3] for other assumptions and the general presentation of (4.9) as the underlying problem that is solved by FISTA.

Each iteration of FISTA consists of two steps. The first is a gradient step in the direction of the minimum of $f$. The second is the solution of a *denoising* subproblem, in which the current guess $\boldsymbol{x}_{\mathrm{c}} \in \mathbb{R}^n$ for the solution is optimized using TVMin. This subproblem,

$$\min_{\boldsymbol{x}} \|\boldsymbol{x} - \boldsymbol{x}_{\mathrm{c}}\|^2 + g(\boldsymbol{x}), \tag{4.10}$$

is a complete minimization problem in itself, which must be solved using an iterative method. For FISTA, the denoising subproblem is solved using a dual approach, as was already proposed in [15].

A FISTA reconstruction is not discrete. However, when applied to an object that consists of homogeneous regions, such as the phantom from Fig. 4.3a, it results in a reconstruction (Fig. 4.3f, created from ten equiangular projections) that is much more homogeneous than SIRT (Fig. 4.3b), and that can be segmented easily (Fig. 4.3g).

To *localize* FISTA, two changes to the algorithm are needed, as for DART. The algorithm must be adapted to only apply its prior to the pixels that are not in $M$, and the set $M$ itself must be determined at some point.

Not applying the prior for the pixels of $M$ is done by adapting the definition of $g$. Define the diagonal matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{n\times n}$, with $\lambda_{jj} \in \{0, 1\}$ for each $j \in \{1, \ldots, n\}$. We know that, for a closed proper convex function $g$ on $\mathbb{R}^n$ and a linear transformation $\boldsymbol{L} : \mathbb{R}^n \to \mathbb{R}^n$, the function $g\boldsymbol{L}(\boldsymbol{x}) \equiv g(\boldsymbol{L}\boldsymbol{x})$ is also a closed proper convex function on $\mathbb{R}^n$ [16, Theorems 5.7 and 9.5]. This means that we can safely adapt the definition of $g$ to

$$g_{\boldsymbol{\Lambda}}(\boldsymbol{x}) \equiv 2\lambda\|\boldsymbol{\Lambda}\boldsymbol{x}\|_{\mathrm{TV}}. \tag{4.11}$$

The matrix $\boldsymbol{\Lambda}$ is computed from $M$ by setting $\lambda_{jj} = 0$ if $j \in M$, and $\lambda_{jj} = 1$ otherwise. The solution of (4.10), with $g_{\boldsymbol{\Lambda}}$ substituted for $g$, will then no longer depend on the pixels of $\boldsymbol{x}$ that are in $M$, because of the multiplication $\boldsymbol{\Lambda}\boldsymbol{x}$. Hence, the TV will no longer be minimized for those pixels. We denote the variant of FISTA that takes $\boldsymbol{\Lambda}$ into account as FISTA$_{\boldsymbol{\Lambda}}$. In practice, FISTA$_{\boldsymbol{\Lambda}}$ differs little

from the original algorithm. The matrix $\mathbf{\Lambda}$, which is fixed during the entire run of FISTA$_{\mathbf{\Lambda}}$, is simply used to exclude the pixels of $M$ from the TV minimization steps.

An overview of the *Local FISTA* algorithm is shown in Fig. 4.5. The set $M$ is computed from an initial FISTA reconstruction. Hence, Local FISTA consists of two separate invocations of the algorithm. As for DART, the reconstructed residual error must be computed from a segmented reconstruction. The initial FISTA reconstruction is relatively easy to segment by thresholding, since FISTA minimizes TV. Additionally, for the initial FISTA reconstruction, a "high" value for $\lambda$ is used. This makes the reconstruction even easier to threshold, since the TV of the solution becomes smaller. For the final FISTA$_{\mathbf{\Lambda}}$ reconstruction, the specified $\lambda$ is used unchanged.

**Input:**

projection data $\boldsymbol{p}$

regularization parameter $\lambda$

thresholds $\tau$ and gray levels $\rho$

error threshold $c$ ***or*** set of free pixels $M$

**Algorithm:**

**if** $M$ is not provided

  Set $\lambda_{\mathrm{h}}$ to a high value; in the current paper $\lambda_{\mathrm{h}} := 10\lambda$

  Compute an initial FISTA reconstruction $\boldsymbol{x}^0$, using $\lambda_{\mathrm{h}}$

  Compute the segmented image $\boldsymbol{s}^0 := \boldsymbol{r}(\boldsymbol{x}^0, \tau, \rho)$

  Compute $M$ from $\boldsymbol{s}^0$ and $c$ (see Fig. 4.2)

**endif**

Compute the diagonal matrix $\mathbf{\Lambda}$ from $M$, setting
  $\lambda_{jj} := 0$ if $j \in M$ and $\lambda_{jj} := 1$ otherwise

Compute the final FISTA$_{\mathbf{\Lambda}}$ reconstruction $\boldsymbol{x}$, using $\lambda$ and $\mathbf{\Lambda}$

**Output:** $\boldsymbol{x}$

**Figure 4.5:** The Local FISTA algorithm. FISTA$_{\mathbf{\Lambda}}$ is a version of FISTA that takes $\mathbf{\Lambda}$ into account, i.e., it uses $g_{\mathbf{\Lambda}}$ from (4.11).

Depending on how $M$ is obtained, two variants are again discerned, namely *Automatic Local FISTA* (if $M$ is computed) and *Manual Local FISTA* (if $M$ is known a priori).

## 4.3 Experiments and Results

In this Section, the properties of Local DART and Local FISTA are studied using simulation experiments. All reconstructions were performed on a square grid of size $512 \times 512$ pixels. The quality of a reconstruction $\boldsymbol{x}$ is measured using the *relative Euclidean phantom distance* $d_{\text{ph}} : \mathbb{R}^n \to \mathbb{R}$, defined as

$$d_{\text{ph}}(\boldsymbol{x}) = \frac{\|\boldsymbol{x} - \boldsymbol{h}\|_2}{\|\boldsymbol{h}\|_2}, \tag{4.12}$$

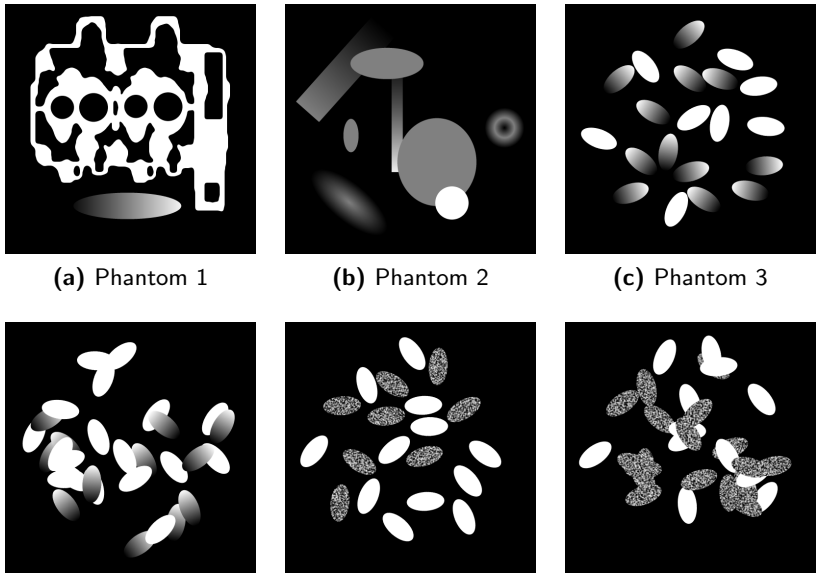
where $\boldsymbol{h} \in \mathbb{R}^n$ is the original phantom image. For the discrete part of a localized DART reconstruction, the quality is also measured using the *relative number of misclassified pixels* (rNMP), defined as the ratio between the number of misclassified pixels and the total number of pixels in the image. We also use the phantom distance and the rNMP (for DART) restricted to part of an image.

### 4.3.1 Phantom Images

Phantom images of size $2048 \times 2048$ pixels were created (Fig. 4.6). The resolution of the phantoms is higher than that of the reconstruction grid to reduce the effect of the pixelation on the reconstructions.

Phantoms 1 (Fig. 4.6a) and 2 (Fig. 4.6b) are basic phantoms. Phantoms 3–6 (Figs. 4.6c–4.6f) each represent a *set* of phantoms. These sets are composed of phantoms with a varying ratio of homogeneous to non-homogeneous elliptical particles, allowing to investigate the properties of the localized algorithms with different amounts of non-homogeneous material. The examples shown in Fig. 4.6 have 40% (Figs. 4.6c and 4.6f) or 60% (Figs. 4.6d and 4.6e) of homogeneous particles. In the phantom sets represented by Phantoms 3 and 5, the ellipses are non-overlapping. In the sets represented by Phantoms 4 and 6, they are allowed to overlap. In all four sets, the arrangement of the ellipses is random and different for each phantom. This may cause some extra variation in the resulting graphs (Figs. 4.9, 4.10, 4.12, and 4.13), but it also ensures that the results are not skewed by any particular arrangement.

Phantoms 1 (Fig. 4.6a) and 2 (Fig. 4.6b) are used in Section 4.3.2 to demonstrate the general principle of reconstruction with a local prior. The sets represented by Phantoms 3 (Fig. 4.6c) and 4 (Fig. 4.6d) are used in Section 4.3.3 to study Local DART. Phantom sets 3 and 4 could also have been used for FISTA. However, algorithms that apply TVMin can still result in accurate reconstructions for samples that contain continuous grayscale objects, if the spatial variation is low. Therefore, Phantom sets 5 (Fig. 4.6e) and 6 (Fig. 4.6f) were designed to have high spatial variation, and are used in Section 4.3.4 to study Local FISTA.



**(a)** Phantom 1 **(b)** Phantom 2 **(c)** Phantom 3

**(d)** Phantom 4 **(e)** Phantom 5 **(f)** Phantom 6

**Figure 4.6:** Phantom images. (a) and (b) are basic phantoms. Phantoms 3–6 (c–f) each represent a *set* of phantoms, with each element containing elliptical particles with an increasing ratio of homogeneous to non-homogeneous ones. (c) and (f) have $40\%$ of homogeneous particles, (d) and (e) have $60\%$ of homogeneous particles.

### 4.3.2 Reconstruction Algorithms with a Local Prior

Using DART as an example, this Section illustrates the mechanism of creating reconstructions with a local prior, both based on the reconstructed residual error and on the true mask of non-discrete pixels. The experiments in this Section are based on Phantoms 1 and 2 (Figs. 4.6a and 4.6b). From these phantoms, synthetic datasets were created using parallel beam equiangular projections. A total of 40 projections was used for Phantom 1, and 55 for Phantom 2. Poisson noise corresponding to $10^5$ photons per detector pixel was applied to the data.

Automatic Local DART was then used to reconstruct these datasets (Figs. 4.7f and 4.7l). The sets of free pixels $M$, from the last DART iteration, are shown in Figs. 4.7c and 4.7i. These were computed from the reconstructed residual error (Figs. 4.7b and 4.7h). The reconstructed residual error itself was computed from the segmented images that are computed internally by the DART algorithm (Figs. 4.7a and 4.7g). The gray levels $\rho$ that are input parameters of DART were taken from the phantom, and the thresholds $\tau$ were computed as $\tau_k = (\rho_k + \rho_{k+1})/2$. For experimental datasets, the gray levels can be optimized

manually [1], or automatically using the projection data [17]. The error threshold $c$ was set to 0.06 for Phantom 1 and to 0.04 for Phantom 2. SIRT (Figs. 4.7d and 4.7j) and (regular) DART (Figs. 4.7e and 4.7k) reconstructions are shown for comparison.

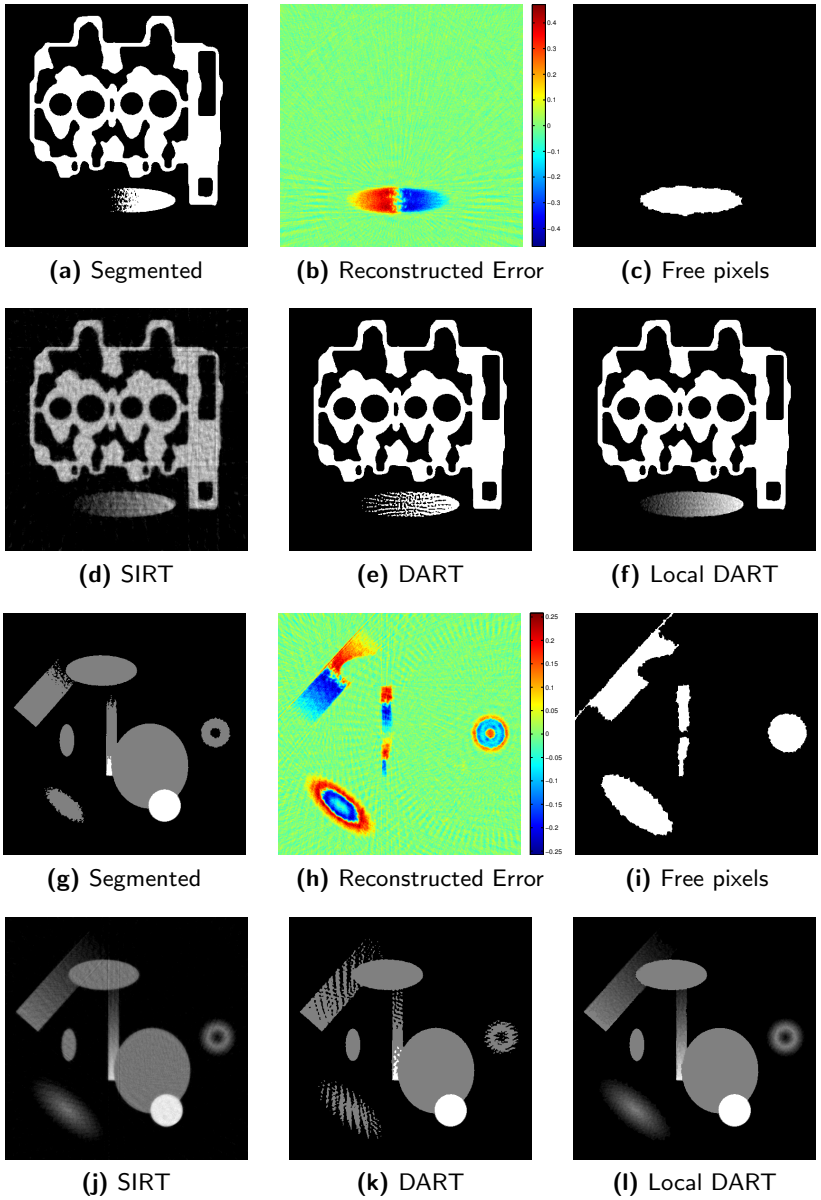
Using the true masks (Figs. 4.8a and 4.8d) increases the accuracy, or allows using more limited data while keeping the same accuracy, of the results over approximate masks that are computed using the reconstructed residual error (Figs. 4.7c and 4.7i). This is illustrated in Fig. 4.8, where Manual Local DART reconstructions are shown (Figs. 4.8c and 4.8f). For these reconstructions, the number of projections was decreased to 20 for Phantom 1 and 35 for Phantom 2. Although, in practice, the true mask is unknown, using the true mask with a simulated dataset illustrates one of the theoretical limits of the approach. SIRT reconstructions are shown for comparison (Figs. 4.8b and 4.8e).

### 4.3.3 Local DART

The experiments in this Section are based on phantom sets 3 (Fig. 4.6c) and 4 (Fig. 4.6d). As for the experiments in Section 4.3.2, synthetic datasets were created using parallel beam equiangular projections. A total of 55 projections was used for all datasets, with Poisson noise corresponding to $10^5$ photons per detector pixel applied to each dataset. As before, the gray levels $\rho$ were taken from the phantoms, and the thresholds $\tau$ were computed as $\tau_k = (\rho_k + \rho_{k+1})/2$. The error thresholds $c$ were optimized manually, and set to 0.7 for the phantoms with 0% and 20% of homogeneous particles, and 0.6 for the other ones. The two variants, Automatic and Manual Local DART, were compared with regular DART and SIRT for all the phantoms in the set.

The quality of the overall reconstructions is given by the phantom distance (Figs. 4.9a and 4.10a). Regular DART does not perform well here, since it cannot accurately reconstruct the non-homogeneous part of the phantoms. This disadvantage decreases when the percentage of homogeneous particles is increased. When all particles are homogeneous (data point 1 in Figs. 4.9a and 4.10a), the phantom distances of regular DART and the two variants of Local DART are the same. This implies that Automatic Local DART correctly determined that there were no non-homogeneous particles left. If the true mask is used (in Manual Local DART), the phantom distance is the lowest. For SIRT with this set of phantoms, the phantom distance seems almost completely independent of the percentage of homogeneous particles, and it is higher than for both variants of Local DART.

The quality of both parts of the reconstructions (i.e., the homogeneous and the non-homogeneous part) can also be measured separately. The phantom distance of the non-homogeneous part (Figs. 4.9b and 4.10b) suggests that Automatic Local DART results in a reconstruction quality that is close to that of SIRT. For



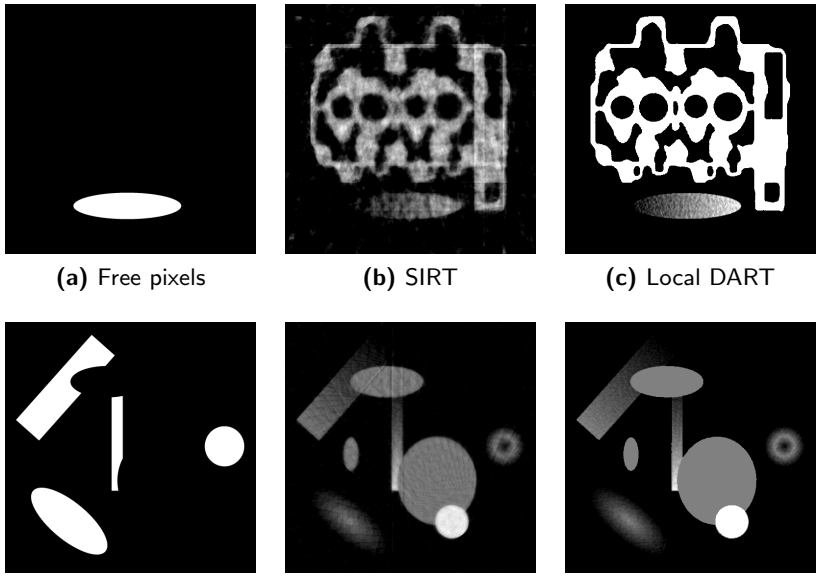
**(a)** Segmented **(b)** Reconstructed Error **(c)** Free pixels

**(d)** SIRT **(e)** DART **(f)** Local DART

**(g)** Segmented **(h)** Reconstructed Error **(i)** Free pixels

**(j)** SIRT **(k)** DART **(l)** Local DART

**Figure 4.7:** Automatic Local DART reconstructions of Phantoms 1 (top row) and 2 (bottom row), where the mask of free pixels was computed *during* the reconstruction. (a), (g) Segmented reconstruction (internal to the DART algorithm). (b), (h) Reconstructed residual error. (c), (i) Mask of free pixels. (d), (j) SIRT reconstruction. (e), (k) DART reconstruction. (f) ,(l) Local DART reconstruction.



**(a)** Free pixels **(b)** SIRT **(c)** Local DART

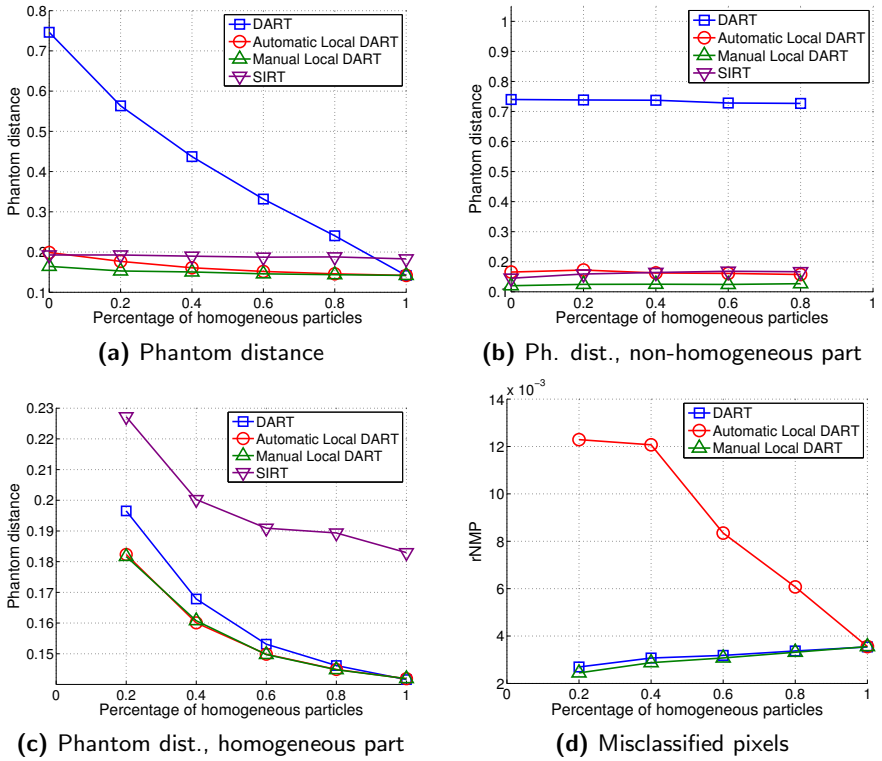
**(d)** Free pixels **(e)** SIRT **(f)** Local DART

**Figure 4.8:** Manual Local DART reconstructions of Phantoms 1 (top row) and 2 (bottom row), where the mask of free pixels was provided manually.

phantoms for which the percentage of homogeneous particles is close to zero (data points 0 and 0.2 in Figs. 4.9b and 4.10b), SIRT seems to be slightly better. If the percentage of homogeneous particles is close to one, Automatic Local DART seems to have a slight advantage. Manual Local DART provides the highest accuracy for all phantoms of the set. The data point for 1 is missing, since there is no non-homogeneous part for that point.

From the phantom distance of the homogeneous part of the reconstructions (Figs. 4.9c and 4.10c), it appears that Automatic and Manual Local DART are very close together. Both show a lower phantom distance than SIRT and DART, and the three variants of DART again coincide at the same value at data point 1. The data point for 0 is missing, since the only homogeneous part there is the (zero valued) background, making (4.12) no longer meaningful.

The rNMP for regular DART is *lower* than for Automatic Local DART (Figs. 4.9d and 4.10d). This is despite the fact that the Local DART reconstruction (Fig. 4.11b, shown for the phantom with 40% of homogeneous particles) is clearly visually closer to the phantom (Fig. 4.6c) than the regular DART reconstruction (Fig. 4.11a). The reason for this is that the mask that is computed automatically during the reconstruction is not exact (Fig. 4.11c), since it is estimated from the




**(a)** Phantom distance **(b)** Ph. dist., non-homogeneous part

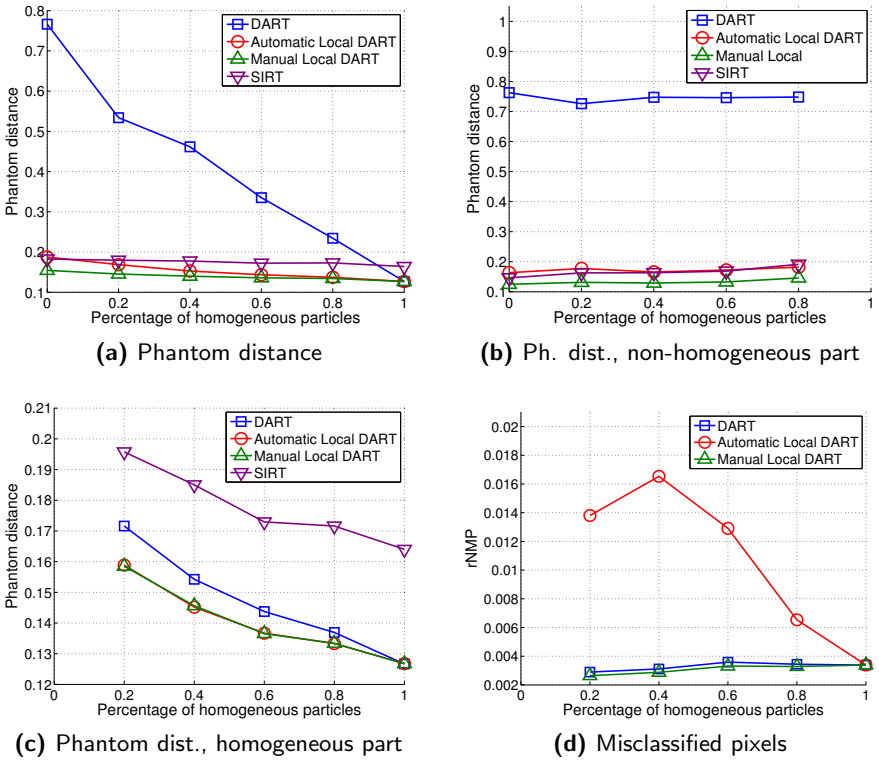
**(c)** Phantom dist., homogeneous part **(d)** Misclassified pixels

**Figure 4.9:** Results for DART, using Phantom set 3. (a) Phantom distance of the complete reconstruction. (b) Phantom distance of the non-homogeneous part. (c) Phantom distance of the homogeneous part. (d) Relative number of misclassified pixels of the homogeneous part.

reconstructed residual error. The rNMP for the homogeneous part, however, is computed using the true mask (Fig. 4.11f), causing a lot of pixels at the edges of the non-homogeneous particles to be marked as misclassified (Fig. 4.11e), even though their value is close to the true value. For regular DART, the pixels in the non-homogeneous particles have completely wrong values (Fig. 4.11a), but these happen to be concentrated inside of the true mask, which results in a low number of misclassified pixels (Fig. 4.11d) in the homogeneous part.

### 4.3.4 Local FISTA

The experiments in this Section are based on phantom sets 5 (Fig. 4.6e) and 6 (Fig. 4.6f). As before, synthetic datasets were created using parallel beam equian-




**(a)** Phantom distance

**(b)** Ph. dist., non-homogeneous part

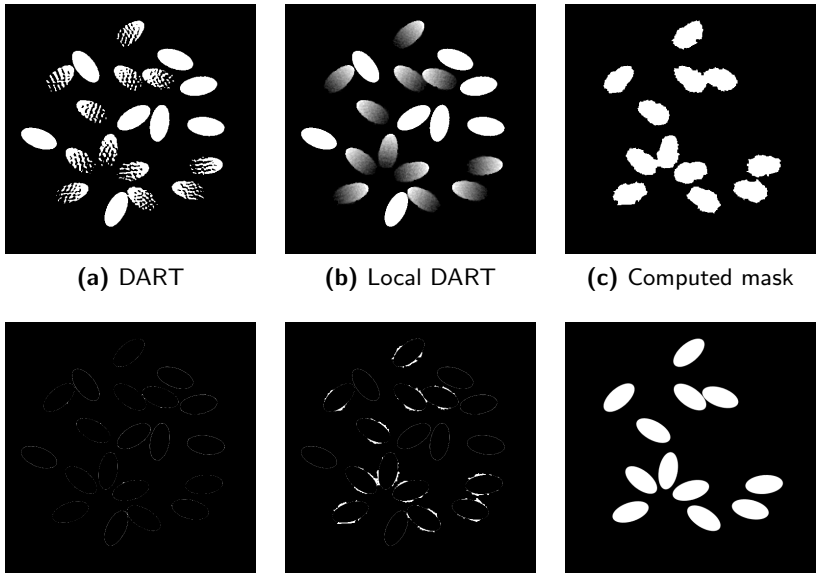
**(c)** Phantom dist., homogeneous part

**(d)** Misclassified pixels

**Figure 4.10:** Results for DART, using Phantom set 4. (a) Phantom distance of the complete reconstruction. (b) Phantom distance of the non-homogeneous part. (c) Phantom distance of the homogeneous part. (d) Relative number of misclassified pixels of the homogeneous part.

gular projections. A total of 90 projections was used for all datasets, with Poisson noise corresponding to $10^5$ photons per detector pixel applied to each dataset. During the initial run of FISTA, the value of $\lambda$ was 0.25. The value of $\lambda$ for the final run was 0.025. As before, the gray levels $\rho$ were taken from the phantoms, and the thresholds $\tau$ were computed as $\tau_k = (\rho_k + \rho_{k+1})/2$. The error thresholds $c$ were optimized manually, and set to decrease linearly from 0.135 for the phantoms with 0% of homogeneous particles to 0.095 for those with 100%. The two variants, *Automatic* and *Manual Local FISTA*, are compared with regular FISTA and SIRT for all the phantoms in the set.

The quality of the overall reconstructions (Figs. 4.12a and 4.13a) is better for both variants of Local FISTA than for regular FISTA and SIRT. As for DART, the advantage of the local algorithms decreases for phantoms with an increasing



**(a)** DART **(b)** Local DART **(c)** Computed mask

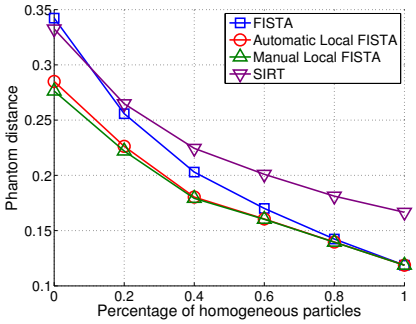
**(d)** Misclassified pixels **(e)** Misclassified pixels **(f)** True mask

**Figure 4.11:** (a) DART reconstruction of Pantom 3. (b) Automatic Local DART reconstruction of Phantom 3. (c) Computed mask of non-discrete pixels for (b). (d) Misclassified pixels for (a). (e) Misclassified pixels for (b). (f) True mask.

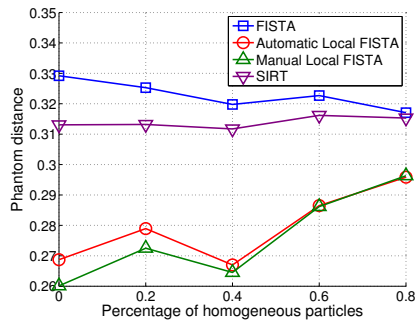
percentage of homogeneous particles, again resulting in an equal phantom distance for the fully discrete phantoms (data point 1 in Figs. 4.12a and 4.13a). The advantage of the Local FISTA reconstructions over SIRT remains about the same for the entire set of phantoms, while the accuracy improves for both when the percentage of homogeneous particles increases.

For the non-homogeneous part of the reconstructions (Figs. 4.12b and 4.13b), both local variants of FISTA improve upon the result of regular FISTA. The results are also better than for SIRT, which performs better than regular FISTA for the non-homogeneous part for this set of phantoms.

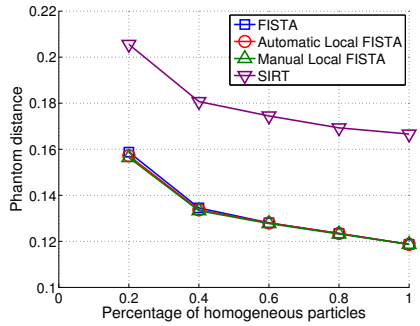
From the phantom distance of the homogeneous part of the reconstructions (Figs. 4.12c and 4.13c), it is apparent that the advantage of Local FISTA is small there. This is as expected, since algorithms that minimize TV are very suitable to reconstruct homogeneous objects. We do note that the advantage from the non-homogeneous parts does not impede the quality of the homogeneous part.




**(a)** Phantom distance




**(b)** Ph. dist., non-homogeneous part



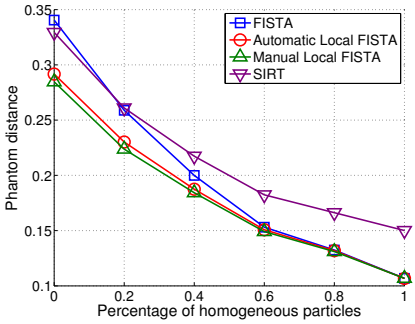

**(c)** Phantom dist., homogeneous part

**Figure 4.12:** Results for FISTA, using Phantom set 5. (a) Phantom distance of the complete reconstruction. (b) Phantom distance of the non-homogeneous part. (c) Phantom distance of the homogeneous part.

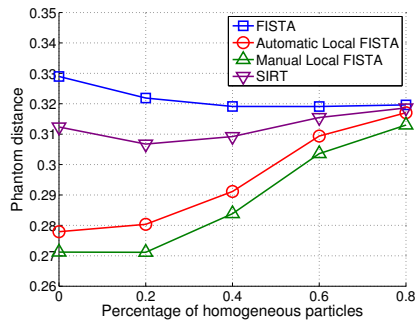
## 4.4 Discussion

Reconstruction algorithms that exploit prior knowledge can greatly improve the accuracy of a reconstruction (Fig. 4.3). However, if the prior is not valid for the complete object, as for the phantoms from Fig. 4.6, then those algorithms can lead to incorrect results (Figs. 4.7e and 4.7k). The reconstructed residual error (Figs. 4.7b and 4.7h) can detect non-homogeneous regions, starting from a segmented reconstruction (Figs. 4.7a and 4.7g). From this, a mask of non-homogeneous pixels can be computed (Figs. 4.7c and 4.7i). If the prior of the reconstruction algorithm, where that prior does not necessarily have to be homogeneity, is then no longer applied for the pixels of that mask, the accuracy of the resulting reconstruction is improved (Figs. 4.7f and 4.7l).

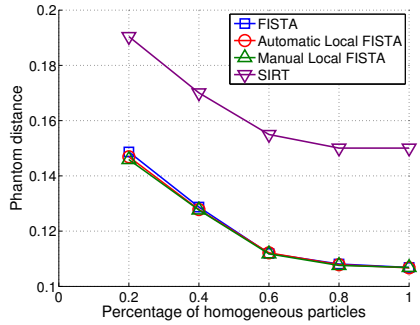
For each reconstruction algorithm, this general principle must be implemented




**(a)** Phantom distance




**(b)** Ph. dist., non-homogeneous part




**(c)** Phantom dist., homogeneous part

**Figure 4.13:** Results for FISTA, using Phantom set 6. (a) Phantom distance of the complete reconstruction. (b) Phantom distance of the non-homogeneous part. (c) Phantom distance of the homogeneous part.

separately. Although it can probably not be applied to all possible reconstruction algorithms, the two algorithms that are localized in the current paper do exploit priors that are quite different. DART uses a hard constraint of discreteness, which results in reconstructions that are exactly homogeneous. DART is completely unable to reconstruct non-homogeneous objects, resulting in large differences in performance between regular DART on the one hand, and Local DART and a classical method such as SIRT on the other hand (Figs. 4.9 and 4.10). FISTA only assumes that the TV should be low, which does not result in discrete reconstructions. FISTA is capable of reconstructing non-homogeneous objects, as long as the spatial variation of those objects is small. Therefore, the overall results for FISTA are closer to Local FISTA and SIRT, although there is still a clear advantage from using Local FISTA (Figs. 4.12 and 4.13).

DART has been used extensively in practice, in fields such as electron tomog-

raphy and X-ray crystallography [18–21]. It is likely that Local DART could make (partially) discrete tomography available for additional datasets, since the results that are described in the current paper (Figs. 4.9 and 4.10) indicate that Local DART improves upon the results of SIRT. The recently proposed PDART algorithm also improves upon SIRT, but its use is restricted to the specific case of the densest material being homogeneous [22]. Local DART is more generally applicable, since it does not impose restrictions on the discrete region, which may contain a mix of (homogeneous) materials. FISTA has also been used in diverse applications such as electrical resistance tomography [23] and mouse whole-body imaging using optoacoustic tomography [24]. As for DART, it seems likely that Local FISTA can make additional datasets available for use with FISTA.

## 4.5 Conclusions

We have presented a way of determining how prior knowledge that is exploited by a reconstruction algorithm can be localized, based on the reconstructed residual error. The application to the DART and FISTA algorithms shows that this approach is feasible in practice, and that it results in more accurate reconstructions.

Implementing the procedure is not complicated. The first step, determining the mask of pixels for which the prior knowledge should not be applied, is performed by computing the reconstructed residual error of a segmented reconstruction and then thresholding it. The second step is different for each algorithm. For DART, the mask is determined during the execution of the algorithm. For FISTA, the mask is determined from an initial reconstruction. Both these changes are simple, compared to the DART and FISTA algorithms themselves.

As discrete tomography and total variation minimization are becoming more established as reconstruction methods, we expect that the technique from the current paper could be useful for a subset of the reconstruction problems that cannot be solved accurately by using algorithms like DART or FISTA unmodified.

## Acknowledgments

This work was financially supported by the IWT TomFood project (IWT is the agency for Innovation by Science and Technology–Flanders, Belgium) and by the NWO (the Netherlands Organisation for Scientific Research–The Netherlands, research programme 639.072.005).

# References

[1] K. J. Batenburg and J. Sijbers, "DART: A practical reconstruction algorithm for discrete tomography," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2542–2553, 2011.

[2] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[3] ——, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, 2009.

[4] G. T. Herman and A. Kuba, *Discrete Tomography: Foundations, Algorithms, and Applications.* Boston: Birkhäuser, 1999.

[5] ——, *Advances in Discrete Tomography and Its Applications.* Boston: Birkhäuser, 2007.

[6] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1–4, pp. 259–268, 1992.

[7] V. Y. Panin, G. L. Zeng, and G. T. Gullberg, "Total variation regulated em algorithm," *IEEE Trans. Nucl. Sci.*, vol. 46, no. 6, pp. 2202–2210, 1999.

[8] M. Persson, D. Bone, and H. Elmqvist, "Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography," *Phys. Med. Biol.*, vol. 46, no. 3, pp. 853–866, 2001.

[9] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, no. 17, pp. 4777–4807, 2008.

[10] T. Roelandts, K. J. Batenburg, and J. Sijbers, "Visualizing the segmentation error of a tomogram using the residual projection error," in *Proc. The Second International Conference on Image Formation in X-Ray Computed Tomography*, Salt Lake City, UT, USA, 2012, pp. 293–296.

[11] S. Becker, J. Bobin, and E. J. Candès, "NESTA: A fast and accurate first-order method for sparse recovery," *SIAM J. Imaging Sci.*, vol. 4, no. 1, pp. 1–39, 2011.

[12] Y. Nesterov, "Smooth minimization of non-smooth functions," *Math. Program.*, vol. 103, no. 1, pp. 127–152, 2005.

[13] P. Gilbert, "Iterative methods for the three-dimensional reconstruction of an object from projections," *J. Theor. Biol.*, vol. 36, no. 1, pp. 105–117, 1972.

[14] J. Gregor and T. Benson, "Computational analysis and improvement of SIRT," *IEEE Trans. Med. Imag.*, vol. 27, no. 7, pp. 918–924, 2008.

[15] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imaging Vis.*, vol. 20, no. 1, pp. 89–97, 2004.

[16] R. T. Rockafellar, *Convex Analysis*. Princeton: Princeton University Press, 1970.

[17] W. van Aarle, K. J. Batenburg, and J. Sijbers, "Automatic parameter estimation for the discrete algebraic reconstruction technique (DART)," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4608–4621, 2012.

[18] K. J. Batenburg, S. Bals, J. Sijbers, C. Kübel, P. A. Midgley, J. C. Hernandez, U. Kaiser, E. R. Encina, E. A. Coronado, and G. Van Tendeloo, "3D imaging of nanomaterials by discrete tomography," *Ultramicroscopy*, vol. 109, no. 6, pp. 730–740, 2009.

[19] S. Bals, K. J. Batenburg, J. Verbeeck, J. Sijbers, and G. Van Tendeloo, "Quantitative three-dimensional reconstruction of catalyst particles for bamboo-like carbon nanotubes," *Nano Lett.*, vol. 7, no. 12, pp. 3669–3674, 2007.

[20] S. Bals, K. J. Batenburg, D. Liang, O. Lebedev, G. Van Tendeloo, A. Aerts, J. A. Martens, and C. E. A. Kirschhock, "Quantitative three-dimensional modeling of zeotile through discrete electron tomography," *J. Am. Chem. Soc.*, vol. 131, no. 13, pp. 4769–4773, 2009.

[21] K. J. Batenburg, J. Sijbers, H. F. Poulsen, and E. Knudsen, "DART: a robust algorithm for fast reconstruction of three-dimensional grain maps," *J. Appl. Cryst.*, vol. 43, no. 6, pp. 1464–1473, 2010.

[22] T. Roelandts, K. J. Batenburg, E. Biermans, C. Kübel, S. Bals, and J. Sijbers, "Accurate segmentation of dense nanoparticles by partially discrete electron tomography," *Ultramicroscopy*, vol. 114, pp. 96–105, 2012.

[23] L. Zhang, H. Wang, and Y. Xu, "A shrinkage-thresholding method for the inverse problem of electrical resistance tomography," in *Proc. Instrumentation and Measurement Technology Conference (I2MTC)*, Graz, Austria, 2012, pp. 2425–2429.

[24] K. Wang, R. Su, A. A. Oraevsky, and M. A. Anastasio, "Investigation of iterative image reconstruction in three-dimensional optoacoustic tomography," *Phys. Med. Biol.*, vol. 57, no. 17, pp. 5399–5423, 2012.

# 5 Conclusions

Computed tomography is an established imaging modality, with applications in diverse fields such as medical imaging, preclinical research, materials science, non-destructive testing, astronomy, etc. Having only a limited amount of recorded data, however, can lead to inaccurate reconstructions. Prior knowledge about the scanned object can help to improve the reconstruction quality, and several algorithms for discrete tomography and total variation minimization are currently used in practice. However, those algorithms are generally limited to objects for which the prior is valid for the whole object.

In this thesis, we have explored two approaches that use *local prior knowledge* to solve the *limited data problem*. The first one is to create a specific new algorithm that directly applies the local prior knowledge that is available for a given problem. Developing an algorithm for a particular problem allows carefully tuning it to maximize the reconstruction quality, but it is not at all straightforward and its application will be limited to the specific setting that it was developed for. The second approach is to devise a general scheme that allows adapting existing algorithms, to enable their use for additional applications, thereby leveraging the existing body of reconstruction algorithms.

In the context of the first approach, we have developed the novel *PDART* algorithm (Chapter 2) for the specific setting of dense homogeneous particles within (possibly) non-homogeneous surroundings. In electron tomography and preclinical imaging, we encountered several datasets for which DART did not provide accurate reconstructions, even though some of the objects that were imaged were expected to consist of homogeneous materials. When PDART was applied to these datasets (for electron tomography in Chapter 2 and for preclinical imaging in Chapter 3), the PDART reconstructions were more accurate than SIRT. This implies that PDART expands the set of samples for which concepts of discrete tomography can be applied, and that it succeeds in retaining some of the strengths of discrete tomography.

In the context of the second approach for solving the limited data problem, we have developed the *reconstructed residual error* (Chapter 3), which creates a map of the errors in a segmented tomogram. It can be used to create *local* versions of existing reconstruction algorithms. However, the reconstructed residual error also has several applications by itself, as shown in Chapter 3. It can be used to improve the gray level estimates of a segmented tomogram, and, for algorithms that assume prior knowledge of those gray levels, to improve the segmentation itself. It can also be used in a straightforward manner to select the most appropriate of several available reconstructions.

The reconstructed residual error can also be used to *localize the available prior knowledge* (Chapter 4), through providing a map of the errors in a segmented reconstruction. If the segmented reconstruction was created using prior knowledge, then the reconstructed residual error can be thresholded to create a mask of pixels that indicates where that prior knowledge was apparently not valid. This leads to a general method for localizing reconstruction algorithms. As applications, we have developed local versions of two practical algorithms from discrete tomography and total variation minimization, namely *Local DART* and *Local FISTA*, respectively. These local variants expand the set of samples for which the strengths of discrete tomography and total variation minimization can be used.

Building on the techniques that were developed in this thesis, we see several possibilities for future research. It would be interesting to compare PDART with Local DART, since there clearly is some overlap in their applicability. Comparing reconstructions from several phantoms and experimental datasets might result in a general way to decide on the appropriate algorithm to use for each object. Another interesting research direction is exploring the reconstructed residual error with other segmentation techniques that are used in tomography. The technique might be useful for selecting the most accurate of several (classical) segmentation methods, such as region growing and clustering, in each particular application.

# Appendix

# A

# Scientific Contributions

## A.1 Journal Articles

T. Roelandts, K. J. Batenburg, E. Biermans, C. Kübel, S. Bals, and J. Sijbers, "Accurate segmentation of dense nanoparticles by partially discrete electron tomography," *Ultramicroscopy*, vol. 114, pp. 96–105, 2012.

B. Goris, T. Roelandts, K. J. Batenburg, H. Heidari Mezerji, and S. Bals, "Advanced reconstruction algorithms for electron tomography: From comparison to combination," *Ultramicroscopy*, vol. 127, pp. 40–47, 2013.

T. Roelandts, K. J. Batenburg, and J. Sijbers, "Visualizing and improving the quality of a segmented tomogram using the reconstructed residual error," *IEEE Trans. Image Process.*, submitted, 2013.

T. Roelandts, K. J. Batenburg, and J. Sijbers, "Local homogeneity in tomography using the reconstructed residual error," *IEEE Trans. Image Process.*, in preparation, 2013.

## A.2 Conference Proceedings (full paper)

T. Roelandts, K. J. Batenburg, and J. Sijbers, "PDART: A partially discrete algorithm for the reconstruction of dense particles," in *Proc. 11th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine (Fully 3D)*, Potsdam, Germany, 2011, pp. 451–454.

T. Roelandts, K. J. Batenburg, and J. Sijbers, "Visualizing the segmentation error of a tomogram using the residual projection error," in *Proc. The Second Interna-*

*tional Conference on Image Formation in X-Ray Computed Tomography*, Salt Lake City, UT, USA, 2012, pp. 293–296.

T. Roelandts, K. J. Batenburg, and J. Sijbers, "Localizing DART using the reconstructed residual error," in Proc. *1st International Conference on Tomography of Materials and Structures*, Ghent, Belgium, 2013.

L. F. Alves Pereira, T. Roelandts, and J. Sijbers, "Inline 3D X-ray inspection of food using discrete tomography," in *Proc. InsideFood Symposium*, Leuven, Belgium, 2013.

## A.3 Conference Proceedings (abstract)

T. Roelandts, K. J. Batenburg, E. Biermans, S. Bals, and J. Sijbers, "Partially discrete tomography for the reconstruction of dense particles," in *17th International Microscopy Congress (IMC17)*, Rio de Janeiro, Brazil, 2010, pp. I7.19.

K. J. Batenburg, S. Bals, S. Van Aert, T. Roelandts, and J. Sijbers, "Ultra-high resolution electron tomography for materials science: a roadmap," in *Microscopy and Microanalysis 2011*, vol. 17, no. S2, Nashville, TN, USA, 2011, pp. 934–935.