Master's Thesis / Pro gradu -tutkielma

X-ray computed tomography of pipe sections by discrete tomography and total variation minimization

Aaro Salosensaari

University of Helsinki
Department of Mathematics and Statistics
Applied Mathematics

Helsinki, Finland
April 18, 2018

Typeset with LuaLaTeX by the author.

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI
MATEMAATTIS-LUONNONTIETEELLINEN TIEDEKUNTA
MATEMATISK-NATURVETENSKAPLIGA FAKULTETEN
FACULTY OF SCIENCE

Tiedekunta – Fakultet – Faculty	Koulutusohjelma – Utbildningsprogram – Degree programme
Faculty of Science	Applied Mathematics

Tekijä – Författare – Author

Aaro Salosensaari

Työn nimi – Arbetets titel – Title

X-ray computed tomography of pipe sections bydiscrete tomography and total variation minimization

Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages
Master's thesis	April 2018	81

Tiivistelmä – Referat – Abstract

The applied mathematical field of inverse problems studies how to recover an unknown function from a set of possibly incomplete and noisy observations. One example of practical, real-life inverse problem is X-ray tomography, where one wishes to recover the internal composition of an object by taking several X-ray projection images from different directions around the object. The tomographic inversion task is a severely ill-posed problem in Hadamard's sense if a limited amount of measurement data is available for the reconstruction. Two examples of limited data tomography are sparse tomography and limited angle tomography. In sparse tomography, tomographic projections are taken with a sparse angular sampling, either because one wishes to reduce the radiation exposure of the object or have a faster imaging process. In limited angle tomography, projections are available only from a limited angular range.

The numerous algorithms that have been proposed for solving the tomographic problem with limited data are based on taking account *a priori* information about the object of interest. In this thesis we study two such algorithms, total variation minimization and discrete tomography. In the total variation minimization the prior assumption is that the total variation of the reconstructed image is small. In the discrete tomography, it is assumed that the target object consists of only discrete amount of different, internally homogeneous materials. Smooth Discrete Algebraic Reconstruction Technique or SDART is a modern method for discrete tomography.

In this thesis we study the problem of recovering the size, shape and the location of small submillimeter-sized voids enclosed in the wall of a metal pipe section. Tomographic reconstructions were found with the total variation minimization and SDART algorithms and also with a combination of both methods, which we call SDART-TV. The study setup is inspired by the problems involved in the radiographic inspection of steel pipe girth welds. During the welding process, various defects may form in the weld seam and affect negatively the structural properties of the weld. In industrial applications it is important to discover if such defects are present in the weld. The X-ray detectors suited for the existing devices for the radiographic pipe weld inspection can be small compared to the pipe diameter, which results in a difficult limited angle imaging geometry. The geometry of pipes also provides another source of prior information for the algorithms, as the space inside the pipe is often known to contain only air or similar matter.

The performance of the total variation and algorithms for discrete tomography in the setup described above was studied with both computational simulations and experiments with an aluminium pipe phantom. The results of the experiments suggest that X-ray tomography algorithms based on *a priori* information could be applicable for non-destructive testing of pipe welds.

Avainsanat – Nyckelord – Keywords

Inverse problems, X-ray tomography, total variation minimzation, discrete tomography, non-destructive testing

Säilytyspaikka – Förvaringställe – Where deposited

Kumpula Campus Library

Muita tietoja – Övriga uppgifter – Additional information

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI
MATEMAATTIS-LUONNONTIETEELLINEN TIEDEKUNTA
MATEMATISK-NATURVETENSKAPLIGA FAKULTETEN
FACULTY OF SCIENCE

Tiedekunta – Fakultet – Faculty	Koulutusohjelma – Utbildningsprogram – Degree programme
Matemaattis-luonnontieteellinen	Soveltava matemaatiikka

Tekijä – Författare – Author

Aaro Salosensaari

Työn nimi – Arbetets titel – Title

X-ray computed tomography of pipe sections bydiscrete tomography and total variation minimization

Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages
Pro gradu	Huhtikuu 2018	81

Tiivistelmä – Referat – Abstract

Inversio-ongelmien tutkimus on soveltavan matematiikan ala jossa tutkitaan tuntemattoman funktion palauttamista joukosta funtiosta tehtyjä mittauksia, jotka saattavat olla epätäydellisiä ja kohinaisia. Esimerkki käytännön sovelluskohteesta inversio-ongelmien alalla on röntgentomografia, missä yritetään nähdä jonkin kohteen sisäinen rakenne ottamalla useita röntgenprojektiokuvia eri puolilta tutkittavaa kohdetta. Tomografinen rekonstruktiotehtävä on erittäin huonosti asetettu ongelma Hadamardin määritelmän mukaan, mikäli käytettävissä on vain rajoitettu määrä projektiokuvia. Esimerkkejä rajoitetun datan tomografiasta ovat harvan kulman tomografia ja rajoitetun kulman tomografia. Harvan kulman tomografiassa projektioiden määrää on rajoitettu tutkittavan kohteen säteilyaltistuksen rajoittamiseksi tai kuvausprosessin nopeuttamiseksi, ja rajoitetun kulman tomografiassa projektioita on saatavilla vain tietystä kulmasta.

Rajoitetun datan tomografiaan on esitetty monia algoritmeja, joista monet usein hyödyntävät ennalta tunnettua *a priori*-informaatiota tutkittavasta kohteesta. Tällaisia algoritmeja ovat totaalivariaatiominimointi ja diskreetin tomografian algoritmit, joita molempia tutkitaan tässä tutkielmassa. Totaalivariaatiominimoinnissa oletetaan että rekonstruoitavan kuvan totaalivariaatio on pieni. Diskreetin tomografian algoritmeissa oletetaan että kohdeobjekti koostuu rakenteeltaan homogeenisista materiaaleista, joita on vain rajoitettu määrä. Eräs modern diskreetin tomografian algoritmi on pehmeä diskreetti algebrallinen rekonstruktiomenelmä SDART.

Tässä tutkielmassa tutkitaan ongelmaa, jossa tavoitteena on palauttaa metalliputken seinämän sisällä sijaitsevien alle millimetrin kokoluokkaa olevien aukkojen koko, muoto ja sijanti röntgentomografian keinoin. Tomografiset rekonstruktiot tehtiin sekä totaalivariaatiominimointimenetelmällä että SDART-algoritmilla, ja lisäksi yhdistelmällä molempia metodeja josta tässä tutkielmassa käytetään nimeä SDART-TV. Tutkimusongelma on peräisin teräsputkien hitsisaumojen laadun tutkimuksesta röntgenkuvantamisella. Hitsausprosessin aikana saumaan saattaa monista syistä syntyä poikkeamia, jotka vaikuttavat haitallisesti sauman kestävyyteen. Monissa teollisuuden sovelluksissa on tärkeää havaita sauman mahdolliset virheet. Putkihitsisaumojen röntgentutkimusta varten kehitettyjen laitteiden kanssa käytettävät röntgenilmaisimet saattavat olla pieniä verrattuna putken halkaisijaan, mikä johtaa vaikeaan rajoitetun kulman kuvausgeometriaan. Putken geometria myös toimii *a priori*-informaation lähteenä jota voidaan hyödyntää algoritmeissa, sillä putkia tutkiessa usein tiedetään että putken sisäosa on tyhjä.

Totaalivariaatio- ja diskreetin tomografian algoritmien toimivuutta kuvatun ongelman ratkaisemisessa tutkittiin sekä laskennallisilla simulaatioilla että kokeilla fyysisellä alumiinisella testikappaleella. Kokeiden tulokset alustavasti vittaavat että modernit röntgentomografiset etukäteistietoa soveltavat algoritmit voivat soveltua putkihitsisaumojen ei-tuhoavaan testaukseen.

Avainsanat – Nyckelord – Keywords

Inversio-ongelmat, röntgentomografia, totaalivariaatiominimointi, diskreetti tomografia, ei-tuhoava testaus

Säilytyspaikka – Förvaringställe – Where deposited

Kumpulan kampuskirjasto

Muita tietoja – Övriga uppgifter – Additional information

Preface and acknowledgments

The text at hand is submitted for the requirements of a degree of Master of Science in the University of Helsinki. It concludes a project to investigate feasibility of discrete tomography algorithms in inspection of quality of steel pipe welds by X-ray imaging, conducted in collaboration between the Inverse Problems research group at the University of Helsinki and Ajat Oy, and participation by Kemppi Oy. The M.Sc. thesis project was supervised by prof. Samuli Siltanen.

I wish to thank Esa Niemi (University of Helsinki) for advice and input during the course of the project. I also thank Henrik Lohman (Ajat Oy), Tuomas Pantsar (Ajat Oy), and Jyri Uusitalo (Kemppi Oy). Additional thanks to Alexander Meaney (University for Helsinki) for help on conducting the practical X-ray measurements, and the people of Department of Physics Workshop for the construction of the physical test phantom.

Notation

symbol	explanation
x	coordinates
$f, f(x)$	ideal signal (image) at infinite precision, function $f : \Omega \to \mathbb{R}^2, \Omega \subset \mathbb{R}^2$
$\Omega \subset \mathbb{R}^2$	image domain, often $[a, b] \times [c, d] \subset \mathbb{R}^2$
$\hat{f}, \mathfrak{F}(f)$	Fourier transform of f
$\mathfrak{R}f(r, \theta)$	Radon transform of f specified by r and θ
$\mathbf{f}$	discretely sampled numerical image, 'dropped' to vector
$\mathbf{A}$	projection matrix
$\boldsymbol{\varepsilon}$	noise term vector
$\mathbf{m}$	measurement vector
$\mathbf{m} = \mathbf{Af} + \boldsymbol{\varepsilon}$	linear forward model
$\mathbf{f}^*$	estimated approximate numerical solution to inverse problem $\mathbf{m} = \mathbf{Af} + \boldsymbol{\varepsilon}$
$\mathbf{f}^k$	kth iterative solution in iterative algorithms
$\epsilon, \epsilon_{\text{limit}}$	(projection) error, error limit
Q	objective function in optimization tasks

List of Theorems and Definitions

List of Figures

Contents

Chapter 1

Introduction

This thesis investigates X-ray computed tomography of hollow, metal pipes or tubes with purpose of recovering small, sub-millimeter size irregularities inside the pipe wall. The task is motivated by in the quality control and inspection of circumferential pipe welds.

1.1 Non-destructive testing of pipe welds

Pipelines constructed of smaller pipe sections joined together by girth welds are common in modern industrial environments. During the welding process, various defects and impurities can be introduced to into the weld seam, affecting integrity and structural soundness of the weld. Many of use cases for such pipes require welds that must be able to withstand high stress and pressure loads. Examples of such safety critical applications range from cooling circuits of nuclear power plants [1] to underwater oil pipelines [2]. Quality control of weld joints in such applications is of utmost importance. Inspection of weld joint quality without breaking or otherwise affecting the object being inspected falls within the scope of the wider field of *non-destructive testing* (NDT, also *non-destructive evaluation* or *inspection*). Various methods for testing quality of welds and pipework are in use today, such as ultrasonic techniques and X-ray radiography. [1, 3]

In this thesis, we study algorithms for X-ray computed tomography of simulated and physical phantoms that are motivated by the girth weld inspection task.

1.2 X-ray computed tomography and inverse problems

Computed tomography (CT) is a branch of applied inverse problems [4]. The field of numerical *inverse* problems concerns itself with finding reconstruction of some unknown function given a set of measurements of the function. If we denote the

unknown function by f, the measurements by m, and let operator A to describe the relation between the unknown and the measurements, the measurement model can be stated as

$$m = A(f).$$

The forward problem is determining m given A and f, and the respective inverse problem is to recover f given m and A. The measurements are often imperfect: the set of measurements may be too small to fully determine f, and in many cases they are to some extent corrupted by noise. Because of these reasons, in many interesting practical applications the inverse problem is ill-posed.

In computed tomography, one wishes to numerically reconstruct a two or three-dimensional view of the internal structure of a given object f from a set of X-ray projection images m (also called *radiographs*). Colloquially described, we wish to *see inside* of the object without breaking it apart. Micrometer-precision image resolution CT (*microtomography* or *micro-CT*) is today an established technique in medical radiology. [See e.g. 5, 6]

Traditional CT algorithms like filtered back-projection are fast, but require a large quantity of full-view projection images sampled at high angular density from all directions around the target. However, often one wishes to work with limited projection data. In medical imaging, taking minimal number of projection images reduces radiation dose received by the patient. In industrial NDT setting the radiation exposure aspect is not as important, but nevertheless smaller number of projection images taken can speed up the imaging process and thus result in more cost-effective procedure. Sometimes projection data can be limited for other reasons, for example projections are available only from certain direction or without full view of the inspected object. The problem of using only a few, sparsely sampled number of projection images is known as *sparse angle tomography*. In some settings projections are available only from restricted angular range (*limited tomography*). Unfortunately CT with classical algorithms becomes ill-posed problem with severely reduced projection data, which motivates the study of algorithms that utilize *a priori* information. [4, 7, 8]

1.3 *A priori* information in CT algorithms

One way to alleviate the problem of ill-posedness in limited data X-ray CT is to augment the reconstruction algorithm with *prior information* about the object of interest. One such source of prior information is knowledge that the object of interest is composed of only a few kinds of different materials and each kind of material itself has uniform composition. The subfield of computed tomography that

studies such problems is known as *discrete tomography* [9] (also sometimes called *discrete-valued* tomography). If we view a 2-dimensional tomographic reconstruction slice as a gray-scale image, in discrete tomography we assume that the pixels in the reconstructed image are allowed to take only certain gray values from a pre-defined (discrete) set. A special case of discrete tomography where there are only two allowed gray values (often normalized to $\{0, 1\}$) is known as *binary tomography*. [9]

The assumptions of discrete tomography hold for a large number of objects in industrial applications, where number and radiological properties of constituent materials of the object can often be known beforehand. This is also true in the case of metal pipe sections welded together: The composition of the pipe base metal is uniform, and each weld can be assumed to have practically uniform structure aside from some discontinuities and defects. The common defects found inside the weld are also often approximately homogeneous, as the they are created by either lack of weld material (i.e. voids and cracks filled with some gas) or presence of extraneous material (slag inclusions).

One popular modern algorithm for discrete tomography is DART, introduced by Batenburg and Sijbers [10], which has spawned a family of related algorithms based on DART [11, 12, 13]. In addition to DART-family of methods, many other algorithms for discrete and binary tomography on have been proposed in the literature in the recent years, including (to name a few) belief-propagation algorithm and Markov random field model from statistical physics [14], hierarchical probabilistic models [15, 16], graph-cut algorithms [17, 18] and various other models and methods [19, 20, 21, 22, 23, 24, 25].

Total variation minimization (TVM, introduced in image reconstruction context by Osher and Rudin [26] and Rudin, Osher, and Fatemi [27]) is another popular way to improve CT reconstruction quality with prior information. It builds upon a similar notion as discrete tomography, but from a slightly different point of view: instead of assuming a discrete set of gray values, TVM is based on the observation that many interesting objects consist only of piecewise constant or smooth areas that have only relatively few sharp discontinuities (that is, edges) between the smooth areas. In TVM, we assume is that the total variation of the gradient magnitude of pixel values in the reconstructed image should be minimal, which corresponds to the characterization of the ideal reconstruction as consisting of relatively homogeneous areas.

The third source of prior information considered in this thesis is the peculiar geometry of the welded pipe itself: As noted above, we can safely assume *a priori* that the metal pipe segments themselves have the uniform attenuation. We can also assume that the hollow pipe segments are empty inside, and contain only uniform intermediate material (which is often air). Moreover, the geometry of the segments

and measurement setup can be known beforehand down to millimeter precision. Only the internal structure of a rather small weld area (wherein the possibly defects lie) is uncertain. In this work we investigate if this information can be used to guide the reconstruction algorithm.

As a summary, in this thesis we study algorithms based on both DART-like discrete tomography heuristic (namely, SDART, which is a smooth variant of DART proposed by Bleichrodt, Tabak, and Batenburg [11]) and the total variation prior (TVM), and methods for combining them with the pipe geometry prior. Recently there has been interest in combining TV and DART, [28, 13, 29], and we also experiment with some conceptually simple ways to combine them. The software implementation of the algorithms was written in MATLAB [30], using ASTRA [31, 32, 33] and Spot toolboxes [34].

1.4 Prior literature on X-ray tomography for NDT of pipes and girth welds

In recent decades there has been some interest in developing methods and tools for X-ray tomography of pipes and welds (both circumferential pipe welds and regular linear welds): Jovanović, Kosec, and Zorc [35] study weld NDT but only with traditional slow CT algorithms. Venkatraman, Raj, and Vaithiyanathan [36] simulated phantoms of various weld defects and used SART and linear system of equations solver for reconstruction. Redmer et al. [37] describe a device for conducting computed laminography and tomography of pipe welds (TomoCAR) and Ewert et al. [38] [1] present further development of similar device (called TomoWELD). For the computed tomography reconstruction part, both TomoCAR and TomoWELD papers [37, 38, 1, 39] discuss traditional algorithms (FBP, ART) only. Vengrinovich et al. [40] describe a toolchain that uses a Bayesian method for weld inspection and especially restoration of cracks. Riis et al. [2] study a limited-view setup similar to the one considered in this work in the context of underwater pipe inspection and use shearlet-based algorithm for reconstruction. Haith, Huthwaite, and Lowe [41] also study underwater pipeline defect detection and characterization by radiography and they apply similar prior constraints from pipe weld geometry, but they use background subtraction and ray-tracing algorithms.

1.5 Outline of the thesis

The rest of the thesis is structured as follows: Chapter 2 provides first an introduction to X-ray computed tomography and classic reconstruction algorithm called filtered backprojection (FBP), followed by an introduction to the linear inverse problems

point-of-view to CT and regularized reconstruction algorithms, and finally also an introduction to discrete tomography. In Chapter 3 we look into the steel pipe welding defects in a more detail and study how X-ray CT could be used in the NDT task of detecting weld defects, and also explain how the weld inspection problem inspired the design of the experiments, such as choice of the imaging geometries and use of *a priori* information. In Chapter 4 the main algorithms and methods considered in this thesis are described. The setup of simulated phantom objects, the construction of the physical aluminum phantom, and other details of the conducted experiments is described in Chapter 5. In Chapter 6 we discuss the results of the experiments, and conclude the thesis in Chapter 7. Details and proofs of some relevant numerical algorithms that fall outside the main scope of the thesis are described in Appendices.

Chapter 2

Theoretical background of inverse problems and X-ray computed tomography

This chapter presents the necessary theoretical background to linear inverse problems in the context of X-ray tomography. We also describe classic reconstruction algorithm (filtered backprojection) and notion of discrete tomography.

2.1 Basics of X-ray computed tomography

In this section we describe how radiographic projection images are formed and the Radon transformation model for computer tomography. Sources were Aarle [42, Ch.1], Müeller and Siltanen [4, Section 2.3], and Kak and Slaney [43, Ch.1].

2.1.1 X-ray imaging

X-rays are a type of electromagnetic radiation capable of passing through materials that appear opaque in the range of the electromagnetic spectrum that is visible to human eye. The latter property makes radiography and other forms of X-ray imaging possible, as discovered by Wilhelm Röntgen in 1895.

X-rays used in radiographic imaging are typically generated by an X-ray tube, which we can understand as a point-like source that shoots a beam of high-energy photons. As the X-rays travel through matter, they interact with it in various ways that depend on the atomic composition of the matter and the energy of X-ray photons. While the actual physical interactions involved are quite complicated, from the mathematical modeling point of view they can be viewed as a phenomenon where the object absorbs or scatters some of the photons that travel through it, but

leaves some photons untouched. As a result, per each unit of distance in matter, a slightly fewer number of photons leave than enter each such unit-length window. The interaction of photons and matter can be approximately characterized by a numerical quantity called *attenuation factor* or coefficient, and we also say that the X-ray is *attenuated* while it traverses in the object being imaged: The matter that absorbs high proportion of X-ray photons has high attenuation coefficient; the matter through which the photons pass through easily is said to have a low attenuation coefficient. In radiographic imaging applications, an object of interest is illuminated with X-ray radiation. The photons that were not absorbed exit the object and arrive on a detector surface; we refer to photon count also as the intensity of the X-ray. Traditional choice of the detector material has been film, but today digital detectors that count the number of photons observed on a single rectangular unit of the detector surface (pixel) are increasingly common. (In this work we are mainly interested in digital tomography.) The projection image formed is known as radiograph.

Following the presentation in [4], the image formation process described above can be modeled mathematically (in the two-dimensional case) in the following way: Let us define the attenuation coefficient of a horizontal cross-section of the object of interest at location $\mathbf{x} = (x_1, x_2) \in B \subset \mathbb{R}^2$ as $f(\mathbf{x})$, where f is a non-negative attenuation coefficient function $f : B \to \mathbb{R}_{\geq 0}$ with a compact support and B is the imaging area.

First we establish the connection between the detected photon counts (i.e. intensity) and the attenuation f. Because X-rays are non-diffracting, we can mathematically model them as straight lines. Assume our X-ray source is located at point $\mathbf{x}_0$ and the ray leaving arrives on a detector pixel located at $\mathbf{x}_1$, and denote the intensity of the X-ray leaving the radiation source and arriving on the detector by $I(\mathbf{x}_0) =: I_0$ and $I(\mathbf{x}_1) =: I_1$, respectively. Let us examine the straight line $\mathbf{s} = \mathbf{s}(t), t \in [0, 1]$ from $\mathbf{x}_0 = \mathbf{s}(0)$ to $\mathbf{x}_1 = \mathbf{s}(1)$. The relative loss in the intensity I of a narrow X-ray that that travels a small distance $\Delta s = |\Delta \mathbf{s}|$ along the line is given by the formula

$$\frac{\Delta I(\mathbf{s})}{I(\mathbf{s})} = -f(\mathbf{s})\Delta s, \tag{2.1}$$

where s is the arc length along the parametrized path $\mathbf{s}$.

By taking the line integral along the path $\mathbf{s}$, we have

$$\int_{\mathbf{s}} f(\mathbf{s})ds = -\int_{\mathbf{s}} \frac{1}{I(\mathbf{s})}\frac{d}{ds}I(\mathbf{s})ds = -\left[\log I(\mathbf{s}(t))\right]_0^1 = \log I_0 - \log I_1. \tag{2.2}$$

Equivalently, this formula can be rearranged as

$$I_1 = I_0 \exp\left(-\int_{\mathbf{s}} f(\mathbf{s})ds\right) \tag{2.3}$$

and is known as a *Beer-Lambert law* for *monochromatic* X-rays. 'Monochromatic' means the model assumes that X-rays are assumed to be of a single energy; true X-rays generated are almost always to some extent polychromatic. Inaccuracy arising from this modeling assumption often causes various *beam hardening* artifacts in the reconstructed images.

The account of the physics of X-ray generation and imaging presented here is simplified, because its main purpose is to motivate how the line integral of attenuation function f can be related to the measured photon count data.

In computed tomography, the inverse problem is to recover the attenuation function f of an object of interest given several image projections taken from multiple angles of view.

2.1.2 Parallel and cone beam geometry

In the preceding subsection, we looked at a single line of one X-ray passing through an object and measured the intensity drop of this single ray. In reality we have a set of measurements from a flat panel detector which can be viewed as a rectangular array of pixels.

The panel captures an image that is formed when X-rays arrive on the detector surface after passing through the object of interest.

An important modeling choice is how we model the path of the X-ray beam. In *parallel beam* model (Figure 4a), we assume that X-rays radiate from a source so far away from the detector that they can be assumed to be (approximately) parallel. In *fan beam* (Figure 4b) model a point-like source is assumed, which results in a more exact model.

In all numerical computations in this thesis, we assume a fan beam model, but in the following sections we will present basic idea of FBP algorithm in parallel beam model, where mathematical treatment is somewhat simpler.

2.1.3 Radon transform

Our measurement model is based on the *Radon transform* of the function f, which is defined as line integral along line path $\mathbf{s}$, when $\mathbf{s}$ is written in terms of an angle θ and a scalar variable $r \in \mathbb{R}$ (with slight abuse of notation)

$$\mathbf{s} = \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{x} \cdot \vec{\theta} = r\}.$$

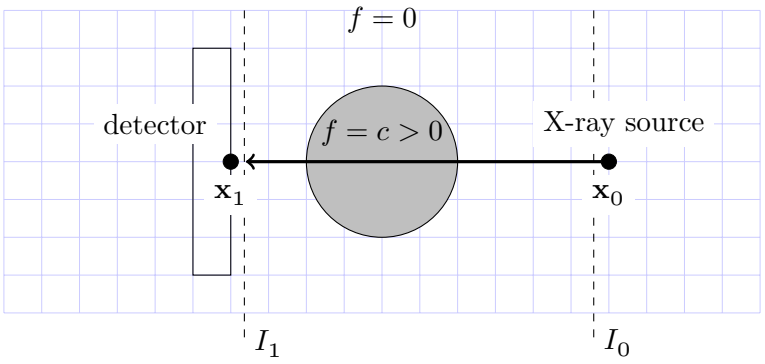


Figure 1: Illustration of the basic attenuation model. A single X-ray beam generated by X-ray tube at point $\mathbf{x}_0$ travels through an object and arrives on the detector surface at point $\mathbf{x}_1$. The object consists of radiologically uniform matter (attenuation coefficient function f is constant $f(\mathbf{x}) = c$) and the attenuation of the air around the object is assumed to be zero.

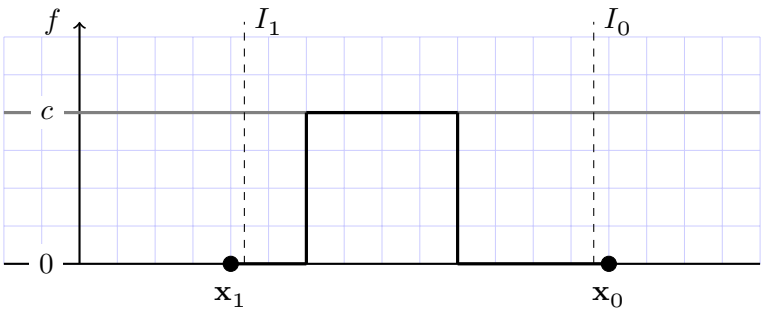


Figure 2: The attenuation profile of f corresponding to the same setup as in Figure 1.

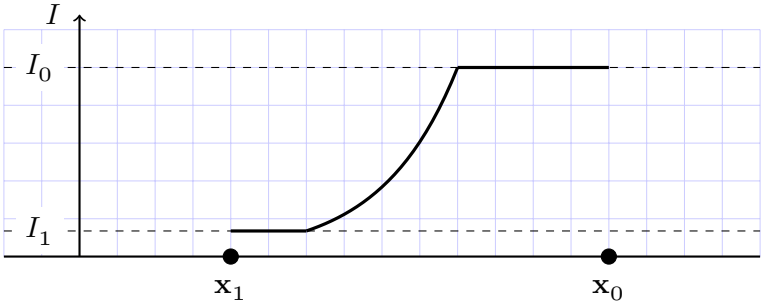


Figure 3: Drop in the X-ray intensity according to the exponential Beer-Lambert model 2.3 in the setup depicted in Figure 1. The X-ray intensity remains the same as long as it travels in air, but as the ray passes through a dense medium, the amount of photons that are not absorbed decays exponentially.

(a) Parallel beam. (b) Fan beam.

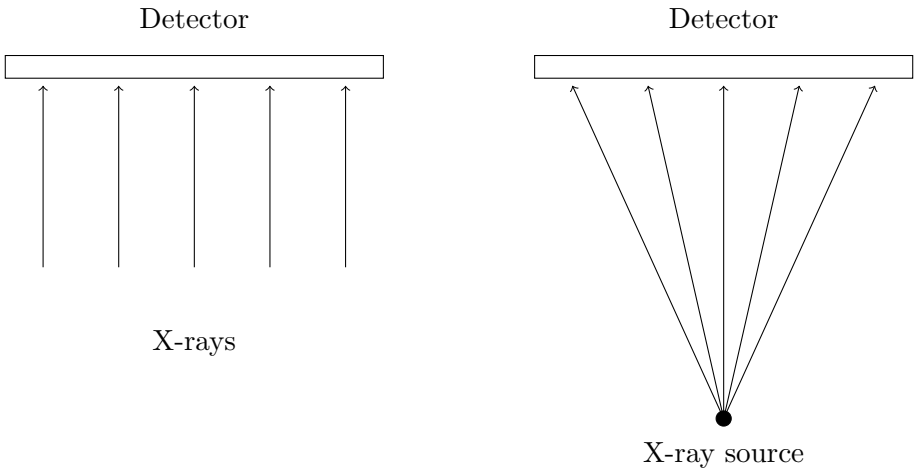


Figure 4: Illustration of parallel (a) and fan beam (b) models. In parallel beam model, X-rays are assumed to arrive at the detector at right angle. In fan beam model, the X-rays are assumed to emanate from one single point-like source.

Here $\vec{\theta}$ is the unit vector tangential to $\mathbf{s}$ and has angle θ with the respect to the x_1 axis

$$\vec{\theta} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \in \mathbb{R}^2.$$

Notice that geometrically r is a length of projection of vector $\mathbf{x}$ to a line determined by the unit vector $\vec{\theta}$. With this notation, we can formally define the *Radon transform*:

Definition 1. The Radon transform $\mathfrak{R}f$ of the function f is the one-dimensional line integral along the line $\mathbf{s}$ specified by parameters (r, θ):

$$\mathfrak{R}f(r,\theta) = \mathfrak{R}(f)(r,\theta) = \int_{\mathbf{s}} f(\mathbf{x})ds = \int_{\mathbf{x}\cdot\vec{\theta}=r} f(\mathbf{x})ds. \tag{2.4}$$

Notice that assuming the parallel beam model, one full projection image consists of several line integrals $\mathfrak{R}f(r,\theta)$ where θ is kept fixed and r tells the distance between the ray and the origin (Alternatively, the length of the vector tangential to $\mathbf{s}$).

The collection of projections sampled from several angles is called a *sinogram*.

It will prove useful to express $f(x_1, x_2) = f'(r, s)$ in a rotated coordinate system (r, s), where r is as above and s is the distance of the displacement along the line $\mathbf{s}$. The coordinate transforms between systems (r, s) and (x_1, x_2) are thus given by

$$\begin{bmatrix} r \\ s \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{2.5}$$

and

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} r \\ s \end{bmatrix}. \tag{2.6}$$

Applying the transform (2.6), the integral in (2.4) becomes

$$\mathfrak{R}f(r,\theta) = \int_{\mathbf{x}\cdot\vec{\theta}=r} f'(r,s)ds = \int_{\mathbf{x}\cdot\vec{\theta}=r} f(x_1(r,s), x_2(r,s))ds \tag{2.7}$$

$$= \int_{-\infty}^{\infty} f(r\cos\theta - s\sin\theta, r\sin\theta + s\cos\theta)ds. \tag{2.8}$$

This latter form (2.8) will become useful in the next section when proving the Fourier slice theorem.

Having now defined the Radon transform, the central problem in computed tomography can be restated as a task of recovering f given the observed Radon transform $\mathfrak{R}(f)(r,\theta)$ of f over multiple θ, r.

2.2 Filtered backprojection

Filtered backprojection (FBP) is a classic tomographic reconstruction algorithm. FBP is fast and very common method in industrial applications. In the context of the weld inspection task, its has a major drawback of not performing well with limited amount of angles nor with extremely noisy data. In the context of weld inspection task, tomographic reconstruction by FBP thus necessitates high amount of projection images (making the imaging process impractically slow).

The FBP algorithm can be described as a filtered, discretized version of the inverse Radon transform. To motivate the rest of the methods in this thesis, we provide here a brief introduction to the theory behind the algorithm. The presentation and the proofs below follow Kak and Slaney [43, Sections 3.2–3.3] and Müeller and Siltanen [4, Section 2.3.3].

2.2.1 Analytical background: Fourier transform and the Radon inversion formula

The theoretical background of filtered backprojection algorithm is provided by a classic result known as the *Radon inversion formula*. The Radon inversion formula can be derived by observing a certain relationship between the Radon transform and the Fourier transform called *Fourier slice theorem*.

Here we derive the formula for the parallel beam case, building upon the exposition in Section 2.1.3.

The Fourier transform is an important integral transform that has many applications in signal and image processing. Intuitively described, the Fourier transformation $\hat{f}$ of a signal f is a representation of the original signal as its constituent sinusoidal frequencies. The connection between the Radon and the Fourier transforms is established in a classic result called *Fourier slice theorem* [4, 43] (also known as *central slice theorem* [4]).

Let us define the *Fourier transform* $\hat{f} = \mathfrak{F}(f)$ of (any) continuous, integrable function $f : \mathbb{R}^n \to \mathbb{R}, f \in L^(\mathbb{R}^2)$ by

$$\hat{f}(\omega) = \mathfrak{F}(f)(\omega) = \int_{\mathbb{R}^n} f(x) e^{-i\omega \cdot x} dx, \tag{2.9}$$

which is a function $\hat{f} : \mathbb{R}^n \to \mathbb{C}$; In a dimension $n = 1$, ω is interpreted as the angular frequency of the signal f. When performing Fourier transform on two-dimensional signals such as images, we have $n = 2$, and ω can be viewed as two dimensional generalization of the frequency. The inverse of the Fourier transform $\hat{f}$ is given by

$$f(x) = \mathfrak{F}^{-1}(\hat{f})(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \hat{f}(\omega) e^{i\omega \cdot x} d\omega. \tag{2.10}$$

Together f and $\hat{f}$ are called a *Fourier transform pair*.

Equipped with these definitions, we are able to state and prove the Fourier slice theorem for the parallel beam case.

Theorem 1 (Fourier slice theorem)**.** *Let the $f = f(x_1, x_2)$ be a compactly supported integrable attenuation function, and $\hat{f}$ its two-dimensional Fourier transform, and let θ be an angular parameter $\theta \in [0, 2\pi]$. Then it holds for all $\omega \in \mathbb{R}$ that*

(2.11)
$$\widehat{\mathfrak{R}f_\theta}(\omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) e^{-i\omega(x_1 \cos\theta + x_2 \sin\theta)} dx_1 dx_2 = \hat{f}(\omega \cos\theta, \omega \sin\theta),$$

where $\hat{f} = \hat{f}(\xi, \eta)$ is the two-dimensional Fourier transform of f at a spatial frequency

$$\xi = \omega \cos\theta$$
$$\eta = \omega \sin\theta.$$

Proof. Let f be the compactly supported attenuation coefficient function with a finite integral, and $\mathfrak{R}f(r, \theta) = \int_{-\infty}^{\infty} f'(r, s) ds$ the Radon transform of f expressed as the (r, s) coordinate system. Let θ be fixed, and write the Radon transform as a function of r, $\mathfrak{R}f(r, \theta) = \mathfrak{R}f_\theta(r)$. The one-dimensional Fourier transform of $\mathfrak{R}f_\theta(r)$

is $\widehat{\mathfrak{R}f_\theta}$, given by

$$\widehat{\mathfrak{R}f_\theta}(\omega) = \int_{-\infty}^{\infty} \mathfrak{R}f_\theta(r) e^{-i\omega \cdot r} dr.$$

By substituting the Radon transform in the rotated coordinate system and applying the change of variables theorem, we get

$$\begin{aligned}
\widehat{\mathfrak{R}f_\theta}(\omega) &= \int_{-\infty}^{\infty} \mathfrak{R}f_\theta(r) e^{-i\omega \cdot r} dr \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f'(r,s) ds \right) e^{-i\omega \cdot r} dr \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(r\cos\theta - s\sin\theta, r\sin\theta + s\cos\theta) ds \right) e^{-i\omega \cdot r} dr \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(r\cos\theta - s\sin\theta, r\sin\theta + s\cos\theta) e^{-i\omega \cdot r} ds dr \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) e^{-i\omega \cdot (x_1 \cos\theta + x_2 \sin\theta)} J dx_1 dx_2,
\end{aligned}$$

where J is the Jacobian,

$$J = \begin{vmatrix} \frac{\partial r}{\partial x_1} & \frac{\partial s}{\partial x_1} \\ \frac{\partial r}{\partial x_2} & \frac{\partial s}{\partial x_2} \end{vmatrix} = (\cos\theta)^2 + (\sin\theta)^2 = 1.$$

□

To summarize, the slice theorem states that the one-dimensional Fourier transform $\widehat{\mathfrak{R}f_\theta} : \mathbb{R} \to \mathbb{C}$ of the Radon transform of f at a fixed angle θ is equal to a restriction of two-dimensional Fourier transform of f, given by $\hat{f}(\omega\cos\theta, \omega\sin\theta)$. Notice in particular that from the equation $\widehat{\mathfrak{R}f_\theta}(\omega) = \hat{f}(\omega\cos\theta, \omega\sin\theta)$ it follows that if we sample the projections of an f at angles $\theta_i, i = 1 \ldots n$, we will know the value of $\hat{f}$ along radial lines at angles θ_i, and with infinite number of samples, $\hat{f}$ would be known fully. (In practice one has finite number of samples and thus one must interpolate.)

With this result, we are now ready to derive the *Radon inversion formula* [4, 43], which describes a simple method for computing inverse Radon transform in the ideal theoretical case.

Theorem 2 (Radon inversion formula)**.** *If $f : \mathbb{R}^2 \to \mathbb{R}$ and its Fourier transform $\hat{f} : \mathbb{R}^2 \to \mathbb{C}$ are absolutely integrable functions, then*

$$f(\mathbf{x}) = f(x_1, x_2) = \frac{1}{(2\pi)^2} \int_0^{\pi} \int_{-\infty}^{\infty} \widehat{\mathfrak{R}f_\theta}(\omega) e^{i\omega \mathbf{x} \cdot \vec{\theta}} |\omega| d\omega d\theta. \tag{2.12}$$

Proof. We first show that Radon transform satisfies symmetry property $\widehat{\mathfrak{R}f}_\theta(r) =$

$\widehat{\mathfrak{R}f}(r,\theta) = \widehat{\mathfrak{R}f}(-r,\theta+\pi)$:

$$\begin{aligned}
\widehat{\mathfrak{R}f}(-r,\theta+\pi) &= \int_{-\infty}^{\infty} \mathfrak{R}f(t,\theta+\pi)e^{-it(-r)}dt \\
&= \int_{-\infty}^{\infty} \mathfrak{R}f(t,\theta+\pi)e^{-i(-t)r}dt \\
&= \int_{-\infty}^{\infty} \mathfrak{R}f(-t,\theta+\pi)e^{-itr}d(-t) \\
&= \int_{-\infty}^{\infty} \mathfrak{R}f(-t,\theta+\pi)e^{-itr}dt \\
&= \int_{-\infty}^{\infty} \mathfrak{R}f(t,\theta)e^{-itr}dt \qquad \text{because } \mathfrak{R}f(-t,\theta+\pi) = \mathfrak{R}f(t,\theta) \\
&= \widehat{\mathfrak{R}}f(r,\theta)
\end{aligned}$$

Now the inversion formula follows from (2.10). Set $\boldsymbol{\xi} = (\omega\cos\theta, \omega\sin\theta)$, and we obtain

$$\begin{aligned}
f(\mathbf{x}) &= \frac{1}{(2\pi)^n}\int_{\mathbb{R}^2} \hat{f}(\boldsymbol{\xi})e^{i\boldsymbol{\xi}\cdot\mathbf{x}}d\boldsymbol{\xi} \quad \text{change of coord.} \\
&= \frac{1}{(2\pi)^n}\int_0^{2\pi}\int_0^{\infty} \hat{f}(\omega\cos\theta,\omega\sin\theta)e^{i\omega(x_1\cos\theta+x_2\sin\theta)}\omega d\omega d\theta \\
&= \frac{1}{(2\pi)^n}\int_0^{2\pi}\int_0^{\infty} \widehat{\mathfrak{R}f_\theta}(\omega)e^{i\omega(x_1\cos\theta+x_2\sin\theta)}\omega d\omega d\theta \\
&= \frac{1}{(2\pi)^n}\int_0^{\pi}\int_{-\infty}^{\infty} \widehat{\mathfrak{R}f_\theta}(\omega)e^{i\omega(x_1\cos\theta+x_2\sin\theta)}|\omega| d\omega d\theta.
\end{aligned}$$

□

2.2.2 Numerical algorithm: Filtered backprojection

The standard tomographic reconstruction algorithm based on the Radon inversion formula is called *filtered backprojection* (FBP).

We sketch the general idea of the FBP algorithm in the parallel projection case[4, 43, p. 3.3] as a numerical version of the theory presented above: First, measure the projection intensities of the object f from all angles θ. By Beer-Lambert law, we can obtain the Radon transform of f from the recorded intensities as $\mathfrak{R}f(r,\theta) = \log I_0 - \log I_1$. Then calculate the (discretized equivalent of) Fourier transform $\widehat{\mathfrak{R}f_\theta}(\omega)$ by FFT (Fast Fourier Transform algorithm). Finally we utilize the formula (2) to obtain the reconstruction in two steps: First compute

$$Q_\theta(t) := \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{\mathfrak{R}f_\theta}(\omega)|\omega|e^{i\omega t}d\omega \tag{2.13}$$

along the line $\mathbf{x} \cdot \vec{\theta} = t$; this is viewed as the *filter* operation that coins the name 'filtered backprojection'. The second step is to 'backproject' the filtered projection $Q_\theta(t)$ for each projection to retrieve reconstruction of f:

$$f(\mathbf{x}) = \frac{1}{2\pi} \int_0^\pi Q_\theta(\mathbf{x} \cdot \vec{\theta}) d\theta. \tag{2.14}$$

Here we skipped many technical details to provide a general idea how FBP works but also an explanation why it can be expected to fail with sparse data (with too sparse angular sample rate θ, Fourier domain $\widehat{f}(\xi, \eta)$ will have information only on a few radial lines and the filter-backprojection step can not be expected to work well).

For discussion of details such as filter design see [44, e.g] or full implementation details (including fan-beam geometries) in [43]. In short, for fan beam geometries two common FBP algorithms exist: First option is to derive a fan-beam analogue of the theory above, which results in a weighted FBP. Another way is to notice that after conducting measurements in the fan beam geometry, by re-ordering all observed line integrals in a particular fashion, the sinogram can be made to look like as if the measurements were made in the parallel beam configuration. This results in a *resorting algorithm*, implemented by e.g. MATLAB `ifanbeam`.

The three-dimensional reconstructions can be achieved by either by extending the theory presented above in two dimensions to the three-dimensional case, or simply stacking two-dimensional cross-section reconstructions. One popular 3D version of FBP for the cone beam geometry is the FDK algorithm (by Feldkamp, Davis, and Kress [45]), which is based on interpreting conebeam as tilted fanbeams [43, p. 3.6.2].

2.3 Linear inverse problems

2.3.1 Motivation: Limited data tomography as an ill-posed inverse problem

The natural filtered backprojection reconstruction method based on the Radon inversion formula 2 (FBP, FDK) has the benefit of being computationally efficient. The main drawback of the method is that it is not robust if the measurement data m is very noisy or only a sparse-angle projection data is available. It is known in the literature [7] that a FBP reconstruction is *numerically stable* given complete data, that is, a large number of projections from all angles around the target object are available. However, as noted, with increasingly limited projection data the FBP reconstruction quality suffers. [7]

There are two common imaging situations where the data may become limited:

either the number of projections may be small (resulting in a *sparse tomography problem*), or the angular range from which the projection data is available is restricted(*limited-angle tomography*). The limited-angle geometries specifically relevant to the weld inspection scenario are described in more detail in Chapter 3.

In many applications the full projection data may not be available because of technical constraints (for example, in some setups it may be physically impossible to take radiographic picture from some particular range of angles), or we wish to reduce the number of projection images taken for a specific reason: for instance, to expose the patient to as minimal dose of ionizing radiation as possible [7][1]. Taking fewer radiographs can also result in an overall faster imaging process.

The problem of creating a tomographic reconstruction from limited projection data is an *ill-posed* problem in the Hadamard's sense[4]. Hadamard's conditions for well-posed problem are:

1. *Existence.* There should be at least one solution.
2. *Uniqueness.* There should be at most one solution.
3. *Stability.* The solution must depend continuously on data.

In an *ill-posed* problem, one or more of the conditions fail.

2.3.2 Definition of linear inverse problem

The tomographic reconstruction problem presented in the Section 2.1 can be regarded as an example of a *linear inverse problem*. The following presentation is based on Müeller and Siltanen [4].

As already noted in the Chapter 1, the field of inverse problems studies the problem of recovering unknown function f from possibly noisy measurements m, the relationship between which is characterized by an operator A, or

$$m = A(f).$$

Recall that in the case of tomography, f is the attenuation function, m are the collection of logarithmic photon count intensities as per the line model (2.2), and A describes the projection process (the Radon transform). It turns out that in computed tomography, the interaction operator A is an approximately linear operator, and so we can write our projection model as

$$m = Af.$$

[1] Alternatively, one could achieve reduced radiation exposure by applying much shorter exposure time windows per projection than usual, but this usually results in poor signal-to-noise ratio per image [6].

Taking into account that measurements m can contain signal noise produced by various sources, we augment the projection model with an additive noise term ε,

$$m = Af + \varepsilon.$$

However, in actual numerical computations we do not observe a continuous function m which depends on a continuous f at infinite precision, but rather we have a discretized vector of measurements $\mathbf{m} = [m_1, \ldots, m_{n_m}]$, one per each constituent pixel on the detector surface per projection view. Likewise the unknown attenuation function is also treated as a n_d dimensional vector $\mathbf{f} = f(\mathbf{x}) = [f_1, \ldots, f_{n_d}]$ sampled at some grid coordinates $\mathbf{x}$. In two-dimensional case (which corresponds to reconstruction of one horizontal slice of the target object), $\mathbf{x}$ is a vector that specifies $n_1 \times n_2 = n_d$ pixel elements, and in the three-dimensional case (the reconstruction task of a whole three-dimensional volume), $\mathbf{x}$ specifies $n_1 \times n_2 \times n_3 = n_d$ voxels.

Taking the above considerations into account, the numerical inverse problem model we are mainly concerned with can be written as a matrix equation

$$\mathbf{m} = \mathbf{A}\mathbf{f} + \boldsymbol{\varepsilon}. \tag{2.15}$$

Here the interaction operator becomes a $n_m \times n_d$ real matrix known as the *projection matrix*, also *sensing* or *measurement matrix*. The vector $\mathbf{f}$ encapsulates the pixel-level information about attenuation coefficient, and the elements a_{ij} of the matrix $\mathbf{A}$ describe how each pixel $f(x, y)$ in the image domain contribute to each measurement m_i. There are various ways to construct projection matrix, depending on the level of modeling precision and computational simplicity desired. (The choice of projection matrix is discussed in more detail in Section 2.7.) The unknown vector $\mathbf{f}$ can be viewed as a two (or three) dimensional pixel (voxel) image subject to linear transformation $\mathbf{A}$. The *inverse problem* is then to recover the vector $\mathbf{f}$ or approximate solution that is "good enough":

Definition 2 (Inverse problem)**.** Given the forward model (2.15) $\mathbf{m} = \mathbf{A}\mathbf{f} + \boldsymbol{\varepsilon}$ and measurement vector $\mathbf{m}$, the *inverse problem* is to recover a solution $\mathbf{f}^* = \mathbf{f}$ or an estimate $\mathbf{f}^*$ that is near the original target, $\mathbf{f} \approx \mathbf{f}^*$.

If the matrix $\mathbf{A}$ were invertible, a naive solution could be computed by

$$\mathbf{f}^*_{\text{naive}} = \mathbf{A}^{-1}\mathbf{m},$$

relying on the intuition that as $\mathbf{m} \approx \mathbf{m} + \boldsymbol{\varepsilon}$,it would appear sensible to assume that also $\mathbf{f}^*_{\text{naive}} = \mathbf{A}^{-1}\mathbf{m} \approx \mathbf{A}^{-1}\mathbf{A}\mathbf{f} = \mathbf{f}$.

However, in vast majority of cases of interest such naive inversion is not feasible

because the problem is *ill-posed*. In the context of model (2.15), its ill-posedness can be characterized in terms of the matrix $\mathbf{A}$: Usually $\mathbf{A}$ is not invertible or its inverse is not continuous, and small changes of $\mathbf{f}$ result in large perturbations of $\mathbf{m}$. For example, even without presence of noise, the linear system defined by matrix equation $\mathbf{Af}+\boldsymbol{\varepsilon}$ can be underdetermined with $n_m < n_d$ (data is undersampled). And as an additional hindrance, often there is at least some of amount of measurement noise $\boldsymbol{\varepsilon}$, often enough to prevent naive reconstruction.

Several reconstruction methods for recovering $\mathbf{f}$ in (2.15) given $\mathbf{A}$ and $\boldsymbol{\varepsilon}$ have been proposed in the inverse problems literature in general and for the tomographic reconstruction problem in particular. In context of CT, we have already described the analytical filtered backprojection algorithm based on Radon transform. Notice that given the algebraic formulation (2.15), an unfiltered and incorrectly weighted 'backprojection' can be obtained by matrix multiplication of the measurement vector $\mathbf{m}$ by the transpose of the projection matrix,

$$\mathbf{f}_{\text{backp}} = \mathbf{A}^T\mathbf{m}. \tag{2.16}$$

('Incorrectly weighted', because in the backprojection step we would like to set each pixel in $\mathbf{f}$ according to *inverse* of the weight it contributes to respective measurement pixels in $\mathbf{m}$ in the forward projection.)

For the rest of this thesis, we will stick to the vector notation of (2.15) and denote the unknown by $\mathbf{f}$ for consistency's sake. In other literature, $\mathbf{x}$ is also often used signify the values of the unknown attenuation to be recovered (instead of $\mathbf{f}$), but in this text, it refers to the coordinates of the values of the unknown function.

Note that tomographic reconstruction methods that are based on solving the system of equations (2.15) are commonly called *algebraic reconstruction techniques*. We will next review the least squares method for solving (2.15) (LSQ, also known as singular value solution; Section 2.4) which can be improved with regularization (for example TV minimization).

2.4 Least-squares solution

One common matrix inversion method is simply finding the approximate *least squares solution* $\mathbf{f}^*$ (LSQ solution; also *Moore-Penrose pseudo-inverse* solution), defined as (if $\mathbf{A}^T\mathbf{A}$ is invertible[2])

$$\mathbf{f}^* = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{m} := \mathbf{A}^+\mathbf{m}, \tag{2.17}$$

[2]If $\mathbf{A}^T\mathbf{A}$ does not exist, pseudo-inverse $\mathbf{A}^+$ can still be defined (Appendix B) but does not have a similarly simple formula.

which is a minimizing solution to the squared 2-norm error functional

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} \|\mathbf{m} - \mathbf{A}\mathbf{f}\|_2^2 \tag{2.18}$$

$$= \arg\min_{\mathbf{f}} \sum_i \left(m_i - \sum_j a_{ji} f_j \right)^2 \tag{2.19}$$

(thus the name "least squares", from "sum of the least squares".) If there are many solutions $\mathbf{f}$ that minimize (2.18), to uniquely determine we choose the $\mathbf{f}$ with the smallest norm $\|\mathbf{f}\|_2$. Also notice that inclusion of inverse $\mathbf{A}^T\mathbf{A}$ 'fixes' the backprojection equation (2.16).

In large-scale problems constructing and inverting $\mathbf{A}^T\mathbf{A}$ explicitly would require prohibitively large amount of memory and computation time; a more common approach is based on *singular value decomposition* (SVD). In problems of even more larger scale (such as high-resolution tomographic reconstruction considered in this thesis), iterative methods are used instead of SVD. One formulation of such gradient-based minimization algorithm is given in a later chapter. In addition to LSQ, there exists many other iterative algebraic reconstruction methods (ART).

In practice the LSQ solution (whether computed with the SVD algorithm or a gradient-based minimization scheme) can often be of poor quality due to matrix $\mathbf{A}$ being ill-conditioned, and thus we often want to compute regularized solutions (Section 2.5).

2.5 Regularized reconstructions

As discussed in Section 2.4, the simple LQS-SVD method can yield suboptimal results for ill-posed inverse problem: Recall that because the matrix $\mathbf{A}$ is ill-conditioned, the naive 'exact' solution to equation $\mathbf{m} = \mathbf{A}\mathbf{f}$ is sensitive to small perturbations introduced by the noise term $\boldsymbol{\varepsilon}$ in the right-hand side of $\mathbf{m} = \mathbf{A}\mathbf{f} + \boldsymbol{\varepsilon}$. The way to overcome this problem is to apply *regularization* [4, 46]. One typical regularization technique is to add a *regularization term* to the objective function Q. Purpose of the term is to enforce smoothness in the solutions $\mathbf{f}^*$ and thus dampen the irregularities of the naive solution. In classical *Tikhonov or L_2 regularization* [4, 47], one applies L_2 norm, and in L_1 regularization L_1 norm is used, respectively;

To make this talk more concrete, we write down the definitions of Tikhonov and L_1 regularized solutions below. Total variation minimization method covered in Section 4.1 can also viewed as a particular kind of L_1 regularization [27].

Definition 3 (Tikhonov regularization)**.** To compute a *Tikhonov regularized solution*, one minimizes an objective function Q that combines the quadratic sum of squares

with (squared) 2-norm of $\mathbf{f}$,

$$\mathbf{f}^* := \mathbf{f}^*_{\text{Tikhonov}} = \arg\min_{\mathbf{f}} Q_{\text{Tikhonov}} \tag{2.20}$$

$$Q_{\text{Tikhonov}} = \|\mathbf{Af} - \mathbf{m}\|_2^2 + \alpha \cdot \|\mathbf{f}\|_2^2 . \tag{2.21}$$

The sum-of-squares term is called *data fidelity term* and as in unregularized LSQ-SVD, it measures the distance of the current candidate solution $\mathbf{f}$ from the data $\mathbf{m}$. The squared norm $\|\mathbf{f}\|_2^2$ is the regularization term which encourages the solutions to be small, alleviating the instability. The constant α is a regularization parameter.

Definition 4 (L_1 regularization)**.** L_1 regularized solution is computed quite the same way as Tikhonov, but replace 2-norm with 1-norm,

$$\mathbf{f}^* := \mathbf{f}^*_{L_1} = \arg\min_{\mathbf{f}} Q_{L_1} \tag{2.22}$$

$$Q_{L_1} = \|\mathbf{Af} - \mathbf{m}\|_2^2 + \alpha \cdot \|\mathbf{f}\|_1 . \tag{2.23}$$

2.6 Discrete tomography

The field of *discrete tomography* is concerned with a subset of tomographic reconstruction problems where one can utilize *a priori* knowledge that elements of $\mathbf{f}$ are only allowed to have values from a discrete domain of K real values, $(\rho_1, \dots, \rho_K) \in \mathbb{R}^K$ for some fixed $K \in \mathbb{N}$. Formally, the discrete tomography problem can be stated as version of inverse problem (Definition 2) where values of vector $\mathbf{f}^*$ are restricted to some set of scalars:

Definition 5 (Discrete-valued inverse problem, discrete tomography)**.** Given the forward model (2.15) $\mathbf{m} = \mathbf{Af} + \boldsymbol{\varepsilon}$ and measurement vector $\mathbf{m}$, recover a solution $\mathbf{f}_{\text{discrete}} = \mathbf{f}$ or an estimate $\mathbf{f}^*_{\text{discrete}}$ that is near the original target $\mathbf{f} \approx \mathbf{f}^*_{\text{discrete}}$ such that $\mathbf{f}^*_{\text{discrete}} \in \{\rho_1, \dots, \rho_K\}^N, \rho_k \geq 0$.

Alternatively, in discrete tomography one seeks to find $\mathbf{f}^*$ that minimizes an objective function

$$\mathbf{f}^* = \arg\min_{f_n \in \{\rho_1, \dots, \rho_K\}} \|\mathbf{m} - \mathbf{Af}\| \tag{2.24}$$

with respect to some norm $\|\cdot\|$ or norm-like measure. The discrete-value assumption can also be seen as a particular kind of sparsity-enforcing restriction. In this viewpoint, the sparse basis is the set of K scalar values $\{\rho_1, \dots, \rho_K\}$, and are image the component vectors; the major difficulty is then in finding the possibly very complicated component vectors.

2.7 Projection matrix and kernels

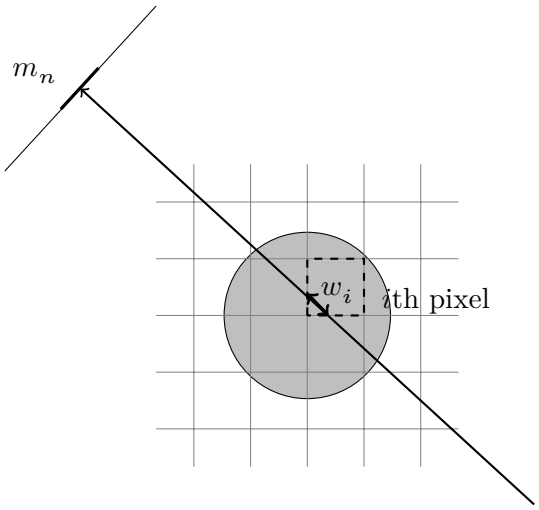


Figure 5: Line kernel projection model. We assume a single zero width X-ray line per pixel m_n on the flat panel detector surface. Reconstruction domain is gridded, and each voxel/pixel i (dashed rectangle) in the grid is assumed to contribute to intensity observed at detector pixel m_n according to the length w_i of the line section that travels through it.

Let us look in more detail how the matrix $\mathbf{A}$ is formed in various measurement setups. On a discretized grid, integral in the line or pencil beam model of (2.2) can be replaced with a weighted sum of discrete elements, illustrated in Figure 5. The path of the X-ray $\mathbf{s}_n$ that arrives on the detector pixel m_n travels through particular pixels f_i in the grid, and the negative logarithm of the observed intensity at the detector:

$$m_n = \sum_{i \in \text{line } \mathbf{s}_n} w_i f_i = \log I_0 - \log I_1. \tag{2.25}$$

Depending on desired level of model precision, different *projection kernels* can be chosen to determine the weight w_i attributed to each pixel through which the X-ray line travels. The projection geometry and the kernel uniquely determine the matrix $\mathbf{A}$.

The simplest possible model ignores the effect of projection angles in a grid and sets $w_i = 1$ for all i, resulting in a very simple approximate formula

$$m_n = \sum_{i \in \text{line } \mathbf{s}_n} f_i,$$

which results in the very simple projection matrix $\mathbf{A}$ (a matrix consisting of row blocks of 1s), but for most purposes this is deemed to be a far too coarse model.

More refined kernels are usually used [see 48, also 42, p. 13]. Some common kernels (that are supported by ASTRA Toolbox) are line kernel (which is illustrated in Figure 5), strip kernel (instead of assuming a zero-width line, also the area of the pixels is considered), and linear interpolation (Joseph kernel). [42, 31, 32]

In all numerical computations carried out for this work, ASTRA toolbox implementation of the line kernel was used.

2.8 Noise model

While working with real-world data generated by a physical system, the measurements $\mathbf{m}$ are often affected by some amount of *noise* from various sources. In the model equation

$$\mathbf{m} = \mathbf{A}\mathbf{f} + \boldsymbol{\varepsilon}, \tag{2.26}$$

the (additive) noise term $\boldsymbol{\varepsilon}$ characterizes our noise model assumptions. In our simulated experiments we assume Gaussian white noise, which is a reasonable approximation for typical noise profile. [7, pp. 1458–1459]

Chapter 3

The motivating problem of X-ray tomography of pipe weld seams

In this chapter we describe the engineering background that inspired the simulated and physical phantom studies in the later chapters. In the first section we review the different kind of defects that may arise in welding process, and in the following sections discuss some practical considerations such as relevant imaging geometries for tomographic imaging of pipe-like target objects.

3.1 Welding process and weld defects

While a full technical treatment on welding is not within the scope of this thesis, it stands to reason to provide some examples of defects we want to investigate and a brief description of the process in which they may arise.

There are many different welding processes: Phillips [3, Ch. 1] mentions over 75, including various types of arc welding, solid state welding and resistance welding. Here we are mainly interested in *arc welding*. While also a great variety of arc welding processes exist, they all share the same common working principle: molten weld filler material is injected between the heated pieces of metal ('base metal') that are to be joined together. Application of heat causes the base metal and the filler metal to fuse together. After the filler material has cooled, a lasting metallic bond has been created. In some processes also high pressure or other energy sources are applied; in arc welding the heat is produced by electric arc that completes an electric circuit between the welding power source, electrode and the work piece. Filler metal is usually supplied by a wire feed mechanism and the weld formation process is often protected from oxidization with a shield gas (in gas shielded processes) or flux

material (flux shielded processes). [See 3, Ch. 1, Sec. 2.1]

Many kinds of discontinuities and flaws may be created in the weld depending on the exact details of the welding process. When the severity or number of imperfections passes a threshold defined in the standardized code applicable for each particular intended use case, it is classified as a defect [3, Sec. 13.1]. Here we list some common types of imperfections and defects that of interest to us:

Porosity consists of rounded voids that contain gaseous materials; one common cause for them is moisture. *Cracks* are thin discontinuities that may be caused by multitude of reasons. [3, p. 234]

In some scenarios several passes of weld are required: for example, after a first weld pass, another one is required on top of the first one until a joint of desired thickness is achieved. If the weld is shielded with flux (provided in e.g. filler material wire). When the filler melts, the flux reacts with impurities in the weld and forms a protective coating layer of *slag* on the top of the weld, which should be removed before the next weld pass. If the slag is imperfectly cleaned and some slag is left in place before the next welding pass (and thus some slag remains inside the weld), a defect called *slag inclusion* is created. [3, p. 13]

For simulation modeling point of view, we treat all defects mentioned above (pores, cracks, slag inclusions) as sub-mm size or smaller differently-shaped regions in the weld area that have different atomic composition than the surrounding weld material.

3.2 Practical considerations regarding X-ray imaging of metal welds

3.2.1 Issues due to composition of the target object and X-ray physics

The common industrial non-destructive testing tasks involve radiographic imaging of metals such as steel and similar materials of high atomic density: inspection of such objects by radiographic imaging will require high energies and long exposure times so that any attenuation drop due to small defects can be registered reliably. This difficulty inspires the choice of looking into sparse-angle tomography: if the procedure of taking a single radiographic image is slow and expensive, could the whole imaging process made faster by reducing the number of projections needed?

An additional difficulty is posed by reconstruction artifacts that are caused by inaccurate model assumptions. As already mentioned in previously in Section 2.1, usually in CT it is assumed that X-rays are monochromatic (i.e. X-rays have a single energy), when in reality they are polychromatic (mixture of X-rays of various

energies). As polychromatic X-ray beam enters the material, the high-energy photons in the beam are absorbed at a different (lower) rate than low-energy photons and thus the beam is 'hardened'. The mathematical model, however, assumes a simple linear attenuation. This modeling inaccuracy tends to cause *beam hardening* artifacts in CT solutions, most prominent of which are 'cupping' or 'whitening' artifacts and 'streaking artifacts'. [see 43, pp. 118 – 121] As an example, some amount of cupping effect can be seen in the ground truth FDK reconstruction of the physical test phantom in Figure 11. Presence of beam hardening effect and other similar model inaccuracies may complicate use of algorithms that make strong *a priori* assumptions and do not account for the possibility of such effects.

3.2.2 Imaging devices and detectors

Practical imaging geometries for the CT tasks may sometimes necessitate dealing with limited field-of-view data. In common weld CT setups it is often desirable to use a detector that is small compared to the size of the pipe being imaged. For example, the TomoWELD[1] hardware supports detector with sensitive area of 100 mm x 50 mm but the system is intended for inspection of pipes that have diameter as large as 220mm.

In the following two subsections we present two different, limited-angle imaging geometries relevant in setups like this: a tomosynthesis-like setup where the detector or source moves parallel along the length of the pipe, and a regular tomography-like setup where the detector-source system rotates around the pipe. In both cases, we discuss only the two-dimensional case, that is, reconstruction of a cross-section 'slice' of the weld.

3.2.3 Limited angle tomography in a tomosynthesis-like setup

The first setup is familiar from the traditional translational laminography and tomosynthesis [4, Sec. 9.3.1] and illustrated in Figure 6. In this setup, we have a point-like X-ray source that moves relative to pipe in direction parallel to its longitudinal axis (i.e. horizontally in the Figure 6). (Alternatively, functionally equivalent geometry can be achieved by keeping the location of X-ray source table and moving the detector and / or pipe. Important feature of the geometry is that acquisition of projection images of the weld seam from as wide angular range as possible.) This imaging geometry that corresponds to traditional tomosynthesis setup and can be viewed as a special case of traditional limited-angle tomography or tomosynthesis imaging, thus we will refer to as 'tomosynthesis geometry' throughout the rest of thesis. The setup can be readily extended to a three-dimensional setting by noting that the X-ray source and the detector will rotate around the imaged

pipe-like object, so one can construct a 3D reconstruction by repeating the 2d limited angle reconstruction slice by slice.

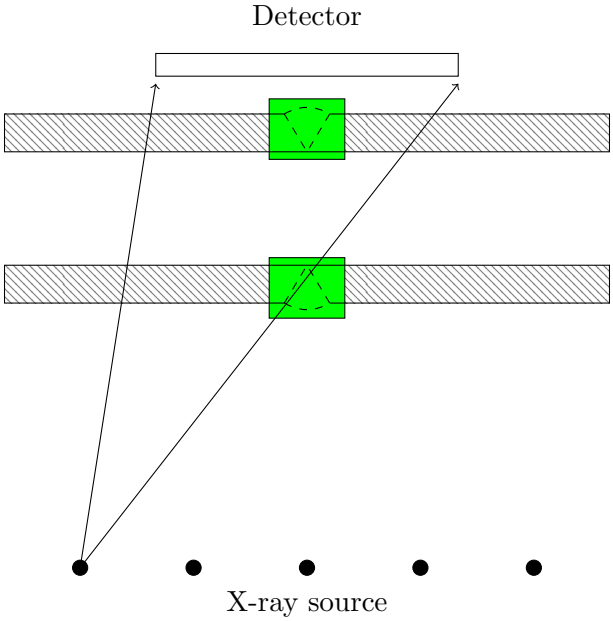


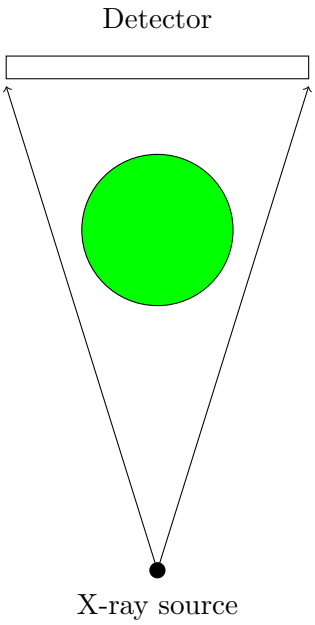
Figure 6: Schematic illustration of the limited angle tomography in tomosynthesis-like geometry: X-ray source moves relative to the detector and the area of interest, and multiple projection images are taken. Region of interest denoted in constant light green and it contains the weld seam (denoted by dashed lines) and the nearby area. The rest of the pipe is in light gray pattern.

3.2.4 Tomography with a limited width detector

In case of the second tomographic reconstruction geometry, we are interested in conducting tomographic reconstructions of the pipe weld with a measurement device similar to TomoCAR [37] or TomoWELD [38] The basic working principle of such devices is described as follows: X-ray radiation source and detector instrument are mounted to the device 180 degrees from each other so that the pipe lies in between of them. The manipulator is slowly rotated around the pipe (more precisely, the weld) and required amount of radiographic images of weld seam are taken. As already discussed, the measurement instrument may be substantially smaller than the diameter of the pipe being imaged[1]. Aforementioned devices inspire our second imaging geometry that can be viewed as a limited-view version of regular (circular full-angle) tomography, which is illustrated in Figure 8b.

[1]One could, of course, obtain a full view projection image even with a small detector by shifting it in a direction tangential to the pipe surface and taking several limited-width projection images which would be combined into a single projection image, but such procedure would be slow and require a more complicated mechanical device.

(a) Regular tomography.



(b) Local tomography.

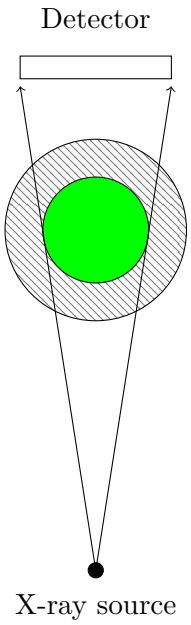


Figure 7: Illustration of (a) regular *full* or *global* tomography and (b) *local* or *region-of-interest* tomography. The region of interest to be recovered is denoted with a constant color (light green in color version). In (b), the uninteresting area where the radiation exposure should be reduced is marked by gray pattern.

(a) Exterior tomography. (b) Our narrow field-of-view tomography for pipes.

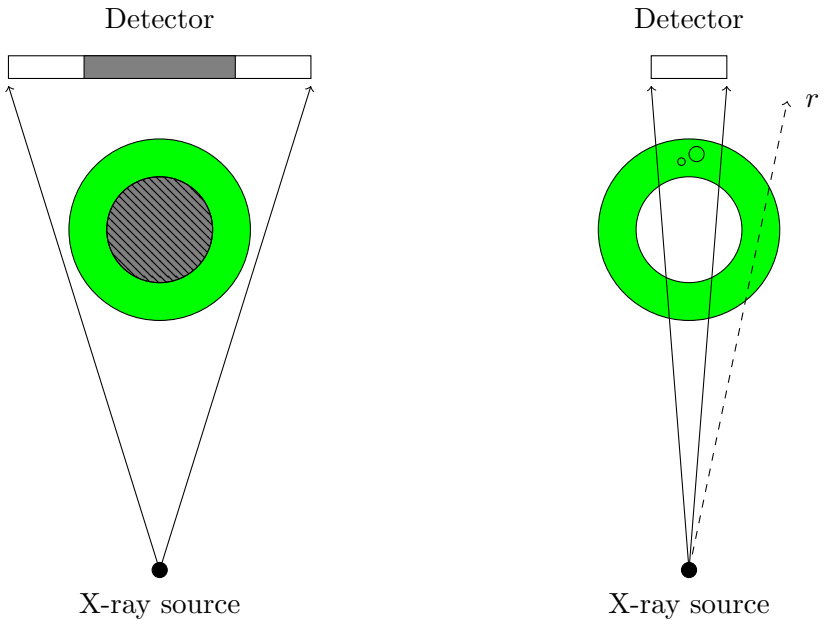


Figure 8: Illustration of (a) *exterior* tomography and (b) our narrow field-of-view setup for tomography of pipes with a small detector. Region of interest in green. In (a), the obscuring region of radiologically oblique material is denoted by dark pattern. Unused portion of detector in constant gray. In (b), the space inside the object known to be empty is in white and the undesirable path of X-ray that would travel a prohibitively long distance through heavily attenuating matter is marked with the letter r (dashed line). Hypothetical defects in the ROI denoted as small circles: projection data about defects will be available only from limited angle.

To summarize, in the limited-view tomographic geometry we are interested in reconstructing a ring-like area (the unknown contents of the weld) in the transversal cross-section of the pipe with a detector that is not wide enough to capture image of the whole object in a single view. This is in contrast to the traditional tomography (Figure 7a), where the common algorithms (such as the ones based on Radon transform) often require *full field-of-view* tomography to reconstruct the object reliably. (Also the full-width data would not necessary be of much use if it was available: as illustrated in Figure 8b, the X-rays that travel along the path r would become so heavily attenuated that impractical exposure times would be needed to observe reliable photon count data.) The algorithms presented in Chapter 4 seek to address this difficulty by utilizing a variety of sources of prior information about the object.

Above we described only a two-dimensional reconstruction setup. A simple way to adapt it to the three-dimensional case can be obtained by moving the device by a small distance parallel to the longitudinal pipe axis and make several measurements and reconstructions, then stack the two-dimensional reconstruction slices on the top of each other. A more sophisticated three-dimensional CT could be achieved by noting that the detector is two-dimensional flat surface, and by constructing a model for a three-dimensional geometry (resulting in the so called *true* three-dimensional reconstruction in a cone-beam model). The circular tomographic setup could also be combined with the 'tomosynthesis geometry' described in the previous section by letting the X-ray device rotate around the pipe once per each tomosynthesis projection. However, such complicated geometries are not investigated in this work.

3.2.5 Related limited data geometries: Region of interest and exterior tomography

We note that the two setups described above both belong in the wider category of limited data X-ray reconstruction problems, but are different than some other related classes of limited data problems that are also common in literature.

In *interior* or *local* or *region-of-interest* tomography (ROI, Figure 7b) one is interested in the reconstruction of only a small area inside a larger object. In medical contexts one often wants to expose the uninteresting areas to as little radiation as possible [49, 7, p. 1439].

In *exterior tomography* (see Figure8a), one is (like in our situation) only interested in an exterior area or 'shell', but contrary to our scenario, the center area is both uninteresting and radiologically oblique compared to the interesting area (or possibly the object is of same density throughout but simply so large in diameter that X-rays are not able to penetrate trough the central region). Because of this the central

area casts a shadow on the detector and renders the mid-section of the sinogram unusable. For example, a classical example of an exterior tomography problem is inspection of an outer part of heavy machinery such as rocket motors. [50]

3.2.6 Limited angle geometry: typical artifacts and detectability of cracks

Both geometries discussed above can be viewed as variations of the limited angle CT, where one conducts classical rotational tomography but projections are available only from limited angular range. In the tomosynthesis geometry the similarity is self-evident: only difference is in of the X-ray source locations (source moves linearly instead of rotating).

In the limited-view tomography geometry, the similarity can be readily noticed if we restrict our attention to any particular fixed small region of interest in the weld cross-section. (See Figure 8b.) As the detector-source pair rotates around the center of the object, any region (for example, a hypothetical small defect) will eventually pass away from the detector view, so projections images where such defect is illuminated are available only from few angles. Thus the reconstruction task with a small detector can be viewed as a series of many limited angle reconstructions solved at once.

The connection between our geometries and the generic limited-angle problem is worth noting for two reasons. Firstly, in limited angle problems the reconstructions by the traditional algorithms based on backprojection are plagued by characteristic 'stretching' aberrations: the details in the reconstructions are elongated in the directions from which data is not available. The severity of the deviations in the resulting reconstruction image increases when the angular range measured decreases (and thus the data becomes more limited). Thus one challenge for the reconstruction algorithms we investigate here is to recover details in the weld with minimal possible amount of such stretching artifacts.

Secondly, it also should be noted that detectability of thin cracks in the weld depends greatly on the positions of X-ray source and detector. Detecting certain kinds of cracks can become an extremely difficult problem in limited angle problems: thin cracks positioned at high angles versus the ray beam are practically invisible, because the minimal drop in the intensity in the X-ray that pass through them. [7, p. 1440] Consequently in geometries similar to what were presented in preceding section, it is reasonable to expect to recover only cracks that are perpendicular to weld surface.

Chapter 4

Tomographic reconstruction methods with discrete and total variation priors

The classical reconstruction methods presented in Chapter 2 often do not achieve the desired reconstruction quality if the available measurement data is noisy or limited. In this chapter, we describe two families of algorithms that try to solve the ill-posed underdetermined tomography problem by the use of prior knowledge: The first algorithm is total variation minimization, which is viewed as a particular kind of regularization method. The second algorithm is DART family of algorithms for discrete tomography.

4.1 Total variation regularization

As already noted in the Chapter 1, total variation minimization (TVM) is a popular regularization method in CT and other image processing applications. The classical Tikhonov or L_2 regularization described in Section 2.5 often manages to successfully suppress noise, but it tends to result in smooth reconstructions. Especially in case of many man-made objects, it is sensible to assume a *piece-wise constant* or *piece-wise smooth* image model. Sharp transition edges present in such images are not always captured very well by the Tikhonov method. The main idea of TVM is to tune the regularization term so that it guides the numerical solution $\mathbf{f}^*$ towards images that are more consistent with our prior knowledge. This is done by replacing the L_2 regularization term with a term that measures the *total variation* of the image, which has a characterization as L_1 norm of magnitude of the image gradient.

4.1.1 Basic mathematical notions of total variation

The concept of total variation is related to the notion of functions of bounded variation. We want to define a model for images (mathematically, functions) $f : \Omega \to \mathbb{R}, \Omega \subset \mathbb{R}^n, n = 2$ that include natural-looking 'piece-wise smooth' images, but not others. We do this by first defining total variation in a mathematically rigorous manner, and then defining our set of desired images as such f that have finite total variation, that is, functions of bounded variation. We present first in one dimension, where the definition is remarkably simple, and then state the equivalent definition in several dimensions.

First we define TV in one dimension $n = 1$. Following [e.g. 4, 51, 52, 53], we define total variation as follows:

Definition 6 (Total variation in one dimension)**.** Let f be a real-valued function $f : [a, b] \to \mathbb{R}$. Total variation of f is defined to be

$$TV(f) = \sup_{P_k : k \in \mathbb{N}} \sum_{i=1}^{k} |f(x_i) - f(x_{i-1})|, \tag{4.1}$$

where the least upper bound (if it exists) is taken over all finite partitions $P_k = (x_0, \dots, x_k)$ of the interval $[a, b]$, defined by

$$x_{i-1} < x_i \forall i \in \{1, \dots, k\}, x_0 = a, x_k = b.$$

If $TV(f) < \infty$ exists, f is said to be of *bounded variation*. The space of functions of bounded variation is denoted by $f \in BV([a, b])$.

If f is absolutely continuous on $[a, b]$, it can be shown [27, 52] that its total variation has a particularly suggestive form

$$TV(f) = \int_{[a,b]} |f'(x)| dx. \tag{4.2}$$

We generalize the concept of total variation in one dimension to $\mathbb{R}^n$ by replacing the derivative in (4.2) with gradient of f, where f is a function $f : \Omega \to \mathbb{R}, \Omega \subset \mathbb{R}^n$,yielding

$$TV(f) = \int_{\Omega} |\nabla f(x_1, \dots, x_n)| dx_1, \dots, dx_n. \tag{4.3}$$

Obviously we are interested in the case $n = 2$. Equation (4.3) is well-defined only for functions f for which gradient ∇f exists. However, we want to define $BV(\Omega)$ in a way that the space $BV(\Omega)$ would include images f with discontinuities such as edges, and consequently a more rigorous definition is warranted. The classical

definition [54, 52, 55] is given as follows:

Definition 7 (Total variation in several dimensions)**.** Let f be a real-valued function $f : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^n$, $f \in L^1$. The total variation of f is defined to be

$$TV(f) := \sup \left\{ \int_\Omega u(x) \operatorname{div} \phi(x) : \phi \in C_0^1(\Omega, \mathbb{R}^2), |\phi| \leq 1 \forall x \in \Omega \right\},$$

where C_0^1 is the space of functions in C^1 with compact support. The function f is said to be of *bounded variation* if its total variation is finite, $TV(f) < \infty$. The space of functions of bounded variation on Ω is denoted by $BV(\Omega)$.

When the gradient notation ∇f can be justified, for example if $f \in C^1(\Omega)$ or $f \in W^{1,1}(\Omega)$ the definition coincides with equation (4.3). We state the relationship as a theorem, but omit the proof.

Theorem 3 (Equivalence of characterizations of total variation)**.** *If $f \in C^1(\Omega)$, then*

$$\int_\Omega |\nabla f(x_1, \ldots, x_n)| dx_1, \ldots, dx_n = TV(f).$$

Proof. Omitted [55]. □

Proof in case of $f \in W^{1,1}(\Omega)$ would require a more detailed discussion on theory of Sobolev spaces and weak (or generalized) derivatives, which falls outside the scope of this thesis.

4.1.2 Total variation for images

In order to carry out any numerical computations a definition of total variation that applies to discrete approximations $\mathbf{f}$ of the ideal images $f \in BV(\mathbb{R}^2)$ is needed. This can be done by discretizing (4.3) with finite differences, which corresponds to taking a L_1-norm of a so-called magnitude of (discrete) image gradient.

Definition 8 (Image gradient)**.** We define the gradient $\nabla \mathbf{f}$ of two-dimensional real-valued image $\mathbf{f}$ by discrete approximation by differences. For each (x, y), let

$$\nabla \mathbf{f}(x, y) = \begin{bmatrix} D_x \mathbf{f}(x, y) \\ D_y \mathbf{f}(x, y) \end{bmatrix} = \begin{bmatrix} f(x+1, y) - f(x, y) \\ f(x, y+1) - f(x, y) \end{bmatrix}. \tag{4.4}$$

Image gradient also provides helpful intuition why minimal total variation variation captures desirable properties of 'cartoon-like' piece-wise constant images: As we see soon below, TV for digital images is defined as sum over the magnitude of the image gradient values. From the definition of the image gradient we can see that in piece-wise constant images, there are only few transition surfaces and the

sum will be small; in unnatural noisy images, the gradient magnitude will be large because of many transitions between homogeneous areas.[1]

Two common discretization schemes, *isotropic* and *anisotropic total variation* are commonly used in literature (and often confusingly both are called 'TV norm'). The difference between them lies in how one chooses to characterize the notion of 'magnitude' so that total variation can be computed for discretely sampled (grid-like) images. Here we define isotropic TV discretization as follows:

Definition 9 (Total variation of image)**.** Given a two-dimensional image $\mathbf{f}$ indexed by (x, y), the isotropic (or 2-1) numerical total variation $TV(\mathbf{f})_{2,1}$ of image $\mathbf{f}$ is given by

$$\begin{aligned} TV(\mathbf{f})_{2,1} &= \|\nabla \mathbf{f}\|_1 = \sum_{x,y} |\nabla \mathbf{f}(x,y)|, \quad \text{where} \\ |\nabla \mathbf{f}(x,y)| &:= \|\nabla \mathbf{f}(x,y)\|_2 \\ &= \sqrt{(f(x+1,y) - f(x,y))^2 + (f(x,y+1) - f(x,y))^2}. \end{aligned}$$

In addition *anisotropic* discretization [see e.g. 56, 57, 58] is defined as follows:

Definition 10 (Anisotropic total variation)**.** Given two-dimensional image $\mathbf{f}$ indexed by (x, y), the anisotropic numerical total variation $TV(\mathbf{f})_{1\text{-ani}}$ of image $\mathbf{f}$ is given by

$$\begin{aligned} TV(\mathbf{f})_{1\text{-ani}} &= \|\nabla \mathbf{f}\|_{1\text{-ani}} = \sum_{x,y} |\nabla \mathbf{f}(x,y)|_{\text{ani}}, \quad \text{where} \\ |\nabla \mathbf{f}(x,y)|_{\text{ani}} &:= \|\nabla \mathbf{f}(x,y)\|_1 \\ &= |f(x+1,y) - f(x,y)| + |f(x,y+1) - f(x,y)|. \end{aligned}$$

While both definitions involve L_1-norm sum $\sum_{x,y} |\nabla(\mathbf{f}(x,y))|$ over the discrete gradient of the image $\mathbf{f}$, in the isotropic version we compute the regular image gradient magnitude with the Euclidian L_2-norm, and in the anisotropic case we instead have the L_1-norm of the gradient. Consequently the methods yield different results for a given image $\mathbf{f}$. While the isotropic discretization measures the perimeter of the objects and the shapes in the image, the anisotropic discretization sums the lengths of the edges along the coordinate axis x, y, which encourages 'blocky' reconstructions that might be desirable in certain applications [57].[2] However, in this text, "total variation" refers to the isotropic approximation unless otherwise indicated.

[1] See Rudin [51] for extended discussion why this is an applicable model for natural images.

[2] Also other characterizations exist. For example, motivated by compressed sensing literature Lou et al. [58] propose combining the two formulations by subtracting the isotropic term from the anisotropic one, yielding regularization term that that better approximates L_0.

4.1.3 Practical algorithm for total variation minimization

Computing a TV-regularized solution can be stated as a task where one minimizes an objective function Q, in which the data fidelity term is regularized by a L_1 total variation term:

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} Q_{TV}(\mathbf{f}) \quad \text{where} \tag{4.5}$$

$$Q_{TV}(\mathbf{f}) = \|\mathbf{A}\mathbf{f} - \mathbf{m}\|_2^2 + \alpha \cdot TV(\mathbf{f})_{2,1}. \tag{4.6}$$

Many algorithms have been proposed for either minimization of Q_{TV} or solving some other alternative characterization that results in a TV-regularized solution (to name a few, projection-onto-convex-set algorithms POCS and ASD-POCS [see eg 59, 6], primal-dual algorithm [60] and many others). Gradient descent methods are efficient choice for the large-scale computations involved here, and simple to implement. Here we use a gradient-descent algorithm similar to Müeller and Siltanen [4, pp. 90–92] and Hämäläinen et al. [61, Sec. 2.4], in particular, a steepest descent method with the step-length choice method of Barzilai and Borwein [62] and non-negativity projection.

Applying gradient-based algorithm to minimization of Q_{TV} as-is, however, is not straightforward. The issue is that the TV term in Q_{TV} is not differentiable at 0 because of the Euclidian norm inside the TV norm,

$$TV(\mathbf{f})_{2,1} = \sum_{x,y} |\nabla \mathbf{f}(x,y)| = \sqrt{(f(x+1,y) - f(x,y))^2 + (f(x,y+1) - f(x,y))^2},$$

which are not differentiable at 0. One efficient practical numerical algorithm can be found by replacing the $|\cdot|$ term in TV-norm with a smooth approximation

$$|x| \approx |x|_\beta = \sqrt{x^2 + \beta}, \quad \beta > 0 \text{ small.}$$

Applying this approximation yields an approximate the energy functional Q_β,

$$Q_\beta = \|\mathbf{A}\mathbf{f} - \mathbf{m}\|_2^2 + \alpha \cdot \sum_{x,y} |\nabla \mathbf{f}(x,y)|_\beta, \tag{4.7}$$

which is everywhere differentiable, and thus fast gradient-based iterative numerical algorithms can be utilized.

In detail, the gradient ∇Q_β with respect to $\mathbf{f} = [f_1, \ldots, f_{n_d}]$ is defined as

$$\nabla Q_\beta(\mathbf{f}) = \nabla_{\mathbf{f}} Q_\beta(\mathbf{f}) = \left[\frac{\partial}{\partial \mathbf{f}(x,y)} Q_\beta(\mathbf{f}(x,y)) \right]_{x,y} \tag{4.8}$$

and we wish to evaluate

$$\nabla_{\mathbf{f}} Q_\beta(\mathbf{f}) = \nabla_{\mathbf{f}} \left(\|\mathbf{A}\mathbf{f} - \mathbf{m}\|_2^2 \right) + \nabla_{\mathbf{f}} \alpha \cdot \sum_{x,y} |\nabla \mathbf{f}(x,y)|_\beta .$$

As the 2-norm is a quadratic form, $\|\mathbf{A}\mathbf{f} - \mathbf{m}\|_2^2 = (\mathbf{A}\mathbf{f} - \mathbf{m})^T (\mathbf{A}\mathbf{f} - \mathbf{m})$, the first term of the gradient can be evaluated as

$$\nabla_{\mathbf{f}} \left(\|\mathbf{A}\mathbf{f} - \mathbf{m}\|_2^2 \right) = 2\mathbf{A}^T (\mathbf{A}\mathbf{f} - \mathbf{m}) \tag{4.9}$$

$$= 2\mathbf{A}^T \mathbf{A}\mathbf{f} - 2\mathbf{A}^T \mathbf{A}\mathbf{m}. \tag{4.10}$$

The gradient of second term is found by applying chain derivative rule, resulting in expression

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{f}(x,y)} \sum_{x,y} |\nabla \mathbf{f}(x,y)|_\beta &= \frac{\partial}{\partial \mathbf{f}(x,y)} \sum_{x,y} \sqrt{(f(x+1,y) - f(x,y))^2 + (f(x,y+1) - f(x,y))^2 + \beta} \\
&= \frac{f(x+1,y) - f(x,y)}{\sqrt{(f(x+1,y) - f(x,y))^2 + (f(x,y+1) - f(x,y))^2 + \beta}} \\
&\quad + \frac{f(x,y+1) - f(x,y)}{\sqrt{(f(x+1,y) - f(x,y))^2 + (f(x,y+1) - f(x,y))^2 + \beta}} \\
&\quad + \frac{2f(x,y) - f(x+1,y) - f(x,y+1)}{\sqrt{(f(x+1,y) - f(x,y))^2 + (f(x,y+1) - f(x,y))^2 + \beta}}.
\end{aligned}$$

The gradient descent method of Barzilai and Borwein [62] is a variation of the classic steepest descent algorithm which itself dates back to Cauchy. In the steepest descent method, the iterative solutions $\mathbf{f}^k$ are found by taking a small step in direction in which the objective function Q seems to decrease fastest, as indicated by the gradient $\nabla Q(\mathbf{f}^{k-1})$ evaluated at the previous location $\mathbf{f}^{k-1}$:

$$\mathbf{f}^k = \mathbf{f}^{k-1} - s_{k-1} \nabla_{\mathbf{f}} Q_\beta(\mathbf{f}^{k-1}). \tag{4.11}$$

The convergence to global optimum is clearly guaranteed if the objective function Q is globally convex. In traditional steepest descent, the optimal step-length s_k at each step is chosen by *line search*, or equivalently, solving minimization problem

$$s_k = \arg\min_s Q(\mathbf{f}^k - s \nabla_{\mathbf{f}} Q(\mathbf{f}^k)). \tag{4.12}$$

In Newton's method for quadratic optimization task, the optimal step-length is given by inverse of a Hessian matrix[3] of Q and in so called quasi-Newton methods,

[3]Defined for a function $Q(x,y)$ as matrix of second order partial derivatives, $\left[\frac{\partial^2}{\partial x \partial y} Q\right]_{x,y}$

by an approximate Hessian.

In the method of Barzilai and Borwein [62], the line search is replaced with a heuristic that results in an algorithm that is faster to compute and is known to be more robust for quadratic objectives. Let $\mathbf{f}^k$ be the reconstruction after kth iteration, and $\mathbf{y}_k = \mathbf{f}^k - \mathbf{f}^{k-1}$. $\mathbf{g}_k = \nabla_{\mathbf{f}} Q_\beta(\mathbf{f}^k) - \nabla_{\mathbf{f}} Q_\beta(\mathbf{f}^{k-1})$. In the BB method one selects the step-length for the $k+1$th iteration s_k to be

$$s_k = \frac{\mathbf{y}_k^T \mathbf{y}_k}{\mathbf{y}_k^T \mathbf{g}_k},$$

and the parameter s_k is the step-length on the next iteration. The BB method can be justified as way of choosing step-length that best corresponds to the Hessian approximation of a second-order quasi-Newton methods, but without the computational overhead involved in finding the (approximate) Hessian. For details, see Appendix A.

In addition to the standard BB steepest descent described above, we apply a projection step to enforce an element-wise non-negativity constraint $\mathbf{f} > 0$. Let $\mathbf{f}'$ be the reconstruction obtained after the gradient descent step (4.11) of the kth iteration,

$$\mathbf{f}' = \mathbf{f}^{k-1} - s_{k-1} \nabla_{\mathbf{f}} Q_\beta(\mathbf{f}^{k-1}).$$

Then the kth reconstruction itself is obtained after a projection operation

$$\mathbf{f}^k(x, y) = \max\{0, \mathbf{f}'(x, y)\} \quad \text{for all } (x, y).$$

We run BB either until the algorithm hits some predefined convergence limit ($\|\mathbf{A}\mathbf{f}^k - \mathbf{m}\| > \epsilon_{\text{residual}}$ or $\|\nabla_{\mathbf{f}} Q_\beta(\mathbf{f}^{k-1})\| > \epsilon_{\text{g}} \cdot \|\nabla_{\mathbf{f}} Q_\beta(\mathbf{f}^0)\|$) or until a maximum number of iterations is met.

The final resulting algorithm is given in Algorithm 1.

Algorithm 1 Gradient descent TVM with Barzilai-Borwein step-length selection

Input:

Projection matrix $\mathbf{A}$.
Measurement vector $\mathbf{m}$.
Initial reconstruction $\mathbf{f}^0$.
Maximum number of iterations $k_{\max}$.
Convergence limits $\epsilon_{\text{residual}} > 0, \epsilon_{\text{g}}$.
Regularization parameter $\alpha > 0$.
TV-norm approximation parameter $\beta > 0$.
Initial (small) step-length $s_0 > 0$.

Output:

TV-regularized reconstruction $\mathbf{f}^*$.

procedure TVM($\mathbf{A}, \mathbf{m}, \mathbf{f}^0, k_{\max}, \epsilon_{\text{residual}}, \epsilon_{\text{g}}, \alpha, \beta, s_0$)
  $k \leftarrow 1$.
  while not converged or $k < k_{\max}$ **do**
    $\mathbf{f}' \leftarrow \mathbf{f}^{k-1} - s_{k-1} \nabla_{\mathbf{f}} Q_\beta(\mathbf{f}^{k-1})$.
    $\mathbf{f}^k(x, y) \leftarrow \max\{0, \mathbf{f}'(x, y)\}$ for all pixels (x, y).
    $\mathbf{y}_k \leftarrow \mathbf{f}^k - \mathbf{f}^{k-1}$.
    $\mathbf{g}_k \leftarrow \nabla_{\mathbf{f}} Q_\beta(\mathbf{f}^k) - \nabla_{\mathbf{f}} Q_\beta(\mathbf{f}^{k-1})$.
    $s_k = \mathbf{y}_k^T \mathbf{y}_k / \mathbf{y}_k^T \mathbf{g}_k$.
    $k \leftarrow k + 1$.
  end while
  $\mathbf{f}^* \leftarrow \mathbf{f}^k$.
end procedure

4.2 Algorithms for discrete tomography

4.2.1 Discrete Algebraic Reconstruction Technique

The Discrete Algebraic Reconstruction Technique (DART) was introduced by Batenburg and Sijbers [10]. DART and its derivatives are all heuristic algorithms. In this thesis we did not study the original DART algorithm in detail, but opted for an improved soft constrained version called SDART [11] and our own variant of it that we call SDART-TV. However, as they are both based on DART and utilize similar heuristics, their presentation benefits from a brief description of the original algorithm.

DART is combines any continuous reconstruction algorithm M which iteratively solves (or finds an approximate solution to) the inverse problem $\mathbf{Af} + \boldsymbol{\varepsilon} = \mathbf{m}$ with a heuristic based on the discreteness assumption (that is, the knowledge of the target object being composed of limited number of homogeneous materials). The reconstruction method M used as subroutine could be for example algebraic solver (such as SART or SIRT), or TVM; in principle, any suitable iterative (algebraic) method will suffice.

The fundamental assumption of the DART is that the first intermediate reconstructions already obtained by the iterative method M are often somewhat *near* the best possible reconstruction that the algorithm M would eventually converge to. If the intermediate reconstruction images are segmented by thresholding to discrete-valued images (with pixel values from a discrete set $\boldsymbol{\rho}$ that is known *a priori*), we hope that large swathes of area in the thresholded image will be approximately correct, and any errors in the reconstruction will be located near the boundaries of different constant-valued regions. Thus less computational resources needed are needed by constraining the algorithm M to work only with pixels inside such boundary areas.

4.2.2 Description of DART algorithm

In more detail, DART proceeds as follows: First, the iterative method M is run for several steps to create the first intermediate reconstruction $\mathbf{f}^1$ which is thresholded to pre-set levels $\rho_1, \ldots, \rho_K$ by thresholds $\tau_1, \ldots, \tau_{K-1}$. Then the thresholded image $\mathbf{f}_s^1$ in divided into two sets, *free* and *fixed* set, depending on whether the pixels in $\mathbf{f}_s^1$ are near the boundaries of discrete-valued regions or not. Any pixel that has neighboring pixel with different gray value than the pixel itself (within some pre-defined radius $r > 0$) falls in the boundary region B and is set as *free* (hereafter denoted by U); the rest of the pixels (that are located inside constant-valued areas in $\mathbf{f}_s^1$) are fixed to their current value. In addition, some pixels in the non-boundary set

are randomly added to the set of free pixels U with probability $1-p$, where $0 < p \leq 1$ is called *fix probability*. Then the iterative method M is run again for some number of steps, except it is allowed to change values only of the pixels in the free set U: the pixels not in U retain their thresholded values. This produces a new reconstruction $\mathbf{f}^2$, which is then thresholded again; the process is repeated until convergence or maximum number of iterations. The combination of running of algorithm M on the free pixels for $k'_{\max}$ iterations and the subsequent thresholding and boundary detection steps constitutes one DART iteration. Depending on exact details of the algorithm chosen as subroutine, the procedure M may produce noisy reconstructions, so Batenburg and Sijbers [10] also recommend applying a smoothing operation (i.e. a blurring effect) once per each DART iteration before the thresholding step. The resulting algorithm is summarized as a pseudocode as Algorithm 2.

4.2.3 SDART algorithm

'Soft DART' or SDART, as defined by Bleichrodt, Tabak, and Batenburg [11], is a relaxed variant of DART with soft constraints instead of hard constraint of keeping non-boundary pixels fixed. This gives us a more robust algorithm that handles noise in measurements better.

In SDART, the alternating steps of first running the subroutine M for $k'_{\max}$ steps followed by a "threshold and detect boundaries" operation is replaced with an optimization task, where DART-like behavior is achieved by imposing soft constraints on the objective function. In more detail, the algorithm proceeds as follows:

Let $\mathbf{f}^{k-1}$ be the result of previous iteration (or the initial reconstruction $\mathbf{f}^0$). Like in DART, a segmented image $\mathbf{f}_s^{k-1}$ is obtained by thresholding $\mathbf{f}^{k-1}$, but instead of setting pixels either free or fixed, one builds a diagonal *weight* or *penalty matrix* $\mathbf{D}$ where each pixel in $\mathbf{f}$ is given a non-negative scalar value that determines how easily the pixel is allowed to change from its current thresholded value during the optimization phase. If d_{ii} is near 0, the corresponding pixel is effectively "free"; the larger penalty makes the pixels more "fixed" Like in DART, pixels that have different-valued neighbors are to be set "free" and pixels in constant regions are to be "fixed", but in SDART the diagonal entry d_{ii} is function of the number of different neighbors. In detail, if we denote the neighboring pixels of ith pixel x_i by $N_r(x_i)$, the values of penalty matrix are given by formula

$$d_{ii} = \frac{100}{c^{b_i}}, \qquad \text{where } b_i = \sum_{x_j \in N_r(x_i)} \mathbb{1}(x_j \neq x_i) \tag{4.13}$$

for each diagonal element in $\mathbf{D}$ that corresponds to each pixel $i = 1, \ldots, n_d$ in the reconstruction image domain; non-diagonal elements are set to be zero. The 'penalty

Algorithm 2 Discrete Algebraic Reconstruction Technique DART.

Input:

Projection matrix $\mathbf{A}$.
Measurement vector $\mathbf{m}$.
Initial (empty) reconstruction $\mathbf{f}^0$.
Maximum number of DART iterations $k_{\max}^{\mathrm{DART}}$.
Subroutine reconstruction method M.
Number of iterations k'_{init} for subroutine M for initial step.
Number of iterations $k'_{\max}$ for subroutine M per each DART iteration.
Thresholding operation T.
Segmentation values $\boldsymbol{\rho} = [\rho_1, \dots, \rho_K]$
Respective segmentation thresholds $\boldsymbol{\tau} = [\tau_1, \dots, \tau_K]$.
Convergence limit $\epsilon_{\mathrm{limit}} > 0$.
Boundary radius $r > 0$.
Fix probability $0 < p \leq 1$.

Output:

DART reconstruction $\mathbf{f}^*$.

procedure DART($\mathbf{A}, \mathbf{m}, \mathbf{f}^0, k_{\max}^{\mathrm{DART}}, M, k'_{\mathrm{init}}, k'_{\max}, T, \boldsymbol{\rho}, \boldsymbol{\tau}, \epsilon_{\mathrm{limit}}, r$)
  $k \leftarrow 1$.
  Compute $\mathbf{f}^1 \leftarrow M(\mathbf{f}^0, k'_{\mathrm{init}})$.
  while not converged ($\|\mathbf{A}\mathbf{f}^k - \mathbf{m}\| > \epsilon_{\mathrm{limit}}$) or $k < k_{\max}^{\mathrm{DART}}$ **do**
    Compute the thresholded image $\mathbf{f}_s^k \leftarrow T(\mathbf{f}^k, \boldsymbol{\rho}, \boldsymbol{\tau})$.
    Find the pixels in the boundary region B.
    Set $U \leftarrow B$.
    Add pixels $\notin B$ to set U with probability $1 - p$.
    Compute $\mathbf{f}^{k+1} \leftarrow M(\mathbf{f}^k, k'_{\max})$ while keeping the pixels $\notin U$ fixed.
    Apply a smoothing operation to pixels in U.
    $k \leftarrow k + 1$.
  end while
  $\mathbf{f}^*_{\mathrm{DART}} \leftarrow \mathbf{f}^k$.
end procedure

constant' c is set to some positive value: Bleichrodt, Tabak, and Batenburg [11] propose $c = 3$, we found that $c = 2$ yielded better results in our problems.

After computing $\mathbf{D}$, one then finds the next reconstruction image $\mathbf{f}^k$ by solving a minimization task

$$\mathbf{f}^k = \arg\min_{\mathbf{f}} Q_{\mathrm{SDART}}(\mathbf{f}) \quad \text{where} \tag{4.14}$$

$$Q_{\mathrm{SDART}}(\mathbf{f}) = \|\mathbf{A}\mathbf{f} - \mathbf{m}\|_2^2 + \alpha \cdot \left\|\mathbf{D}(\mathbf{f} - \mathbf{f}_s^{k-1})\right\|_2^2. \tag{4.15}$$

To summarize, one SDART step consists of application of the thresholding operation, computation of penalty matrix, and finally solving minimization problem. In the original paper [11], the authors propose using CGLS or SIRT as solver. We found that running the Barzilai-Borwein gradient descent for constant number of steps was robust and worked well. Pseudocode of our implementation is given as Algorithm 3.

To derive the BB algorithm for minimization of Q_{SDART}, the gradient of the objective with the respect to $\mathbf{f}$ is needed. The gradient of the squared norm of the residual was already computed for TVM; on the other hand, treating $\mathbf{D}$ and $\mathbf{f}_s^{k-1}$ as fixed, the gradient of the smooth penalty term is

$$\begin{aligned}
\nabla_{\mathbf{f}} \left\|\mathbf{D}(\mathbf{f} - \mathbf{f}_s^{k-1})\right\|_2^2 &= \nabla_{\mathbf{f}} (\mathbf{D}^T(\mathbf{f} - \mathbf{f}_s^{k-1}))^T \mathbf{D}^T(\mathbf{f} - \mathbf{f}_s^{k-1}) \\
&= \nabla_{\mathbf{f}} (\mathbf{f} - \mathbf{f}_s^{k-1})^T \mathbf{D}\mathbf{D}^T(\mathbf{f} - \mathbf{f}_s^{k-1}) \\
&= 2\mathbf{D}^T\mathbf{D}(\mathbf{f} - \mathbf{f}_s^{k-1}).
\end{aligned}$$

In numerical computations instead of matrix-vector product $\mathbf{D}(\mathbf{f} - \mathbf{f}_s^{k-1})$, one can work with vector $\mathrm{diag}(\mathbf{D})$ and element-wise vector products. Likewise, because $\mathbf{D}$ is diagonal, $\mathbf{D}\mathbf{D}^T = \mathbf{D}^T\mathbf{D} =$ elements of $\mathrm{diag}(\mathbf{D})$ squared element-wise.

The reason why SDART handles noisy data better than regular DART is because of the use of the penalty matrix $\mathbf{D}$, which allows the algorithm to distribute noise and other reconstruction uncertainty "smoothly" over the whole image: the DART-like behavior is implemented via the term $\left\|\mathbf{D}(\mathbf{f} - \mathbf{f}_s^{k-1})\right\|_2^2$, which penalizes non-discrete solutions and solutions that are far from intermediate discrete solutions, but its effect is balanced against a data fidelity term. Unlike in DART, even the values of relatively fixed pixels could change.

The main downside of SDART is that the numerical speed-ups achieved by not needing to compute full matrix equations is lost, but in the setups considered in this thesis, this did not prove a serious hindrance.

Algorithm 3 Soft DART.

Input:

Projection matrix $\mathbf{A}$.
Measurement vector $\mathbf{m}$.
Initial (empty) reconstruction $\mathbf{f}^0$.
Maximum number of (S)DART iterations $k_{\max}^{\mathrm{SDART}}$.
Number of BB iterations $k'_{\max}$ per each SDART iteration.
Thresholding operation T.
Segmentation values $\boldsymbol{\rho} = [\rho_1, \dots, \rho_K]$
Respective segmentation thresholds $\boldsymbol{\tau} = [\tau_1, \dots, \tau_K]$.
Convergence limit $\epsilon_{\mathrm{limit}} > 0$.
Boundary radius $r > 0$.

Output:

SDART reconstruction $\mathbf{f}^*_{\mathrm{SDART}}$.

procedure SDART($\mathbf{A}, \mathbf{m}, \mathbf{f}^0, k_{\max}^{\mathrm{SDART}}, k'_{\max}, T, \boldsymbol{\rho}, \boldsymbol{\tau}, \epsilon_{\mathrm{limit}}, r$)
  $k \leftarrow 1$.
  while not converged ($\|\mathbf{A}\mathbf{f}^k - \mathbf{m}\| > \epsilon_{\mathrm{limit}}$) or $k < k_{\max}^{\mathrm{SDART}}$ **do**
    Compute the thresholded image $\mathbf{f}_s^k \leftarrow T(\mathbf{f}^k, \boldsymbol{\rho}, \boldsymbol{\tau})$.
    Count the number of different-valued neighbors b_i for each pixel in $\mathbf{f}_s^k$.
    Compute $\mathbf{D}$ given $\mathbf{b}$.
    $\mathbf{f}^{k+1} \leftarrow \arg\min_{\mathbf{f}} Q_{\mathrm{SDART}}$ with BB for $k'_{\max}$ iterations starting from $\mathbf{f}^k$.
    $k \leftarrow k + 1$.
  end while
  $\mathbf{f}^*_{\mathrm{SDART}} \leftarrow T(\mathbf{f}^k, \boldsymbol{\rho}, \boldsymbol{\tau})$.
end procedure

4.2.4 Concerning some issues with DART-like heuristics

As a heuristic algorithm, DART (and DART-like algorithms such as SDART) may function remarkably well in some scenarios and yet give less than satisfactory results in some others: the suitability of the algorithm to each individual reconstruction task must be confirmed by experiments. One typical problem with the DART heuristic manifests itself when attempting reconstruction from extremely limited and noisy data: if the DART algorithm misclassifies a large area of pixels during the initial phase of reconstruction (or the initial reconstruction $\mathbf{f}^0$ fed to DART is far from the true solution) and the edges of piecewise constant areas in $\mathbf{f}^1$ are far from their correct shape, DART may correct the error very slowly. This is because only the boundary pixels are set as free, algorithm tends to manipulate mostly only the pixels in the boundary area per one DART iteration. (This is despite the random additions to set B with probability p.) In other words, during one DART iteration the erroneous boundary moves usually towards its correct location only by the width of the boundary defined by the neighborhood criterion r. Issues like this may even lead the algorithm becoming stuck at a wrong solution. Similar issues have been also reported at least by Zhuge, Palenstijn, and Batenburg [13, p. 456]. We noticed similar issues also to some extent in the case of SDART.

The impact of such undesirable behavior could be reduced if the DART algorithm could be improved to be less prone to in introduce prohibitively major reconstruction errors in the initial and intermediate reconstructions. In case of the SDART, we can try to manipulate objective function landscape to the effect of steering the optimization process away from suboptimal reconstruction paths right from the beginning. Below we discuss how to improve SDART process with total variation prior.

4.3 Augmenting SDART with total variation minimization

In recent years there has been some interest in improving performance of DART-like and other discrete tomography algorithms with TV minimization or related methods. On the other hand, TVM methods do not utilize *a priori* gray level information and often have tendency to yield solutions that are not exactly discrete-valued but have some amount of smooth variation, which is obviously non-ideal. Zhuge, Palenstijn, and Batenburg [13] propose SDART-like variation of DART, where DART-like behavior is induced with a choice of suitable objective function which is optimized with a gradient method. Demircan-Tureyen and Kamasak [28] propose finding an initial reconstruction by TVM (however, instead of the method presented in

Section 4.1.3, they use the TVAL3 method of Li [63]) and then continuing with regular DART iterations with refined on pixel fixing scheme and particular automatic threshold and gray-value estimation. TVM initialization has also been suggested by Goris et al. [64] and Zhuge, Palenstijn, and Batenburg [13]. In a more recent paper [29], Demircan-Tureyen and Kamasak incorporate DART-like pixel fixing with 1-D TV-minimization.

In this thesis we study two different ways to combine TVM and SDART: firstly, initializing the SDART-algorithms with TV-reconstructed solution (not unlike [28]); and secondly, augmenting the regular SDART objective function with a TV term. We call the latter algorithm SDART-TV and describe it in more detail in the next section.

4.3.1 SDART-TV algorithm

The approach proposed here is similar to and inspired by the TVR-DART method proposed by Zhuge, Palenstijn, and Batenburg [13], but the author believes it's conceptually simpler and possibly may prove more robust. Simply described, we accompany the SDART optimization objective Q_{SDART} (which we view as an heuristic implementation of a "discreteness prior" in Bayesian probabilistic terms) with a total variation prior (or the smooth β-approximation), resulting in updated objective function

(4.16)
$$Q_{\text{SDART-TV}} = \|\mathbf{Af} - \mathbf{m}\|_2^2 + \alpha_{\text{SDART}} \cdot \left\|\mathbf{D}(\mathbf{f} - \mathbf{f}_s^{k-1})\right\|_2^2 + \alpha_{\text{TV}} \cdot TV(\mathbf{f})_{2,1}$$

(4.17)
$$\approx \|\mathbf{Af} - \mathbf{m}\|_2^2 + \alpha_{\text{SDART}} \cdot \left\|\mathbf{D}(\mathbf{f} - \mathbf{f}_s^{k-1})\right\|_2^2 + \alpha_{\text{TV}} \cdot \sum_{x,y} |\nabla \mathbf{f}(x,y)|_\beta.$$

Algorithm otherwise proceeds in the same way as regular SDART, alternating between computation of $\mathbf{D}$ and minimization of the objective with a gradient method. If SDART is viewed as a smooth relaxation of DART where M is replaced by minimization of a LSQ residual, the SDART-TV is nothing more than relaxation of DART where M is replaced by TVM. This is in contrast to TVR-DART, where the TV assumption and DART-like discreteness prior are incorporated into a single objective function,

(4.18)
$$Q_{\text{TVR-DART}} = \left\|\mathbf{A}S_{\boldsymbol{\tau},\boldsymbol{\rho}}(\mathbf{f}) - \mathbf{m}\right\|_2^2 + \alpha \cdot TV(S_{\boldsymbol{\tau},\boldsymbol{\rho}}(\mathbf{f})),$$

where $S_{\boldsymbol{\tau},\boldsymbol{\rho}}$ is a smooth segmentation ('staircase') function. Again, the objective $Q_{\text{TVR-DART}}$ is minimized by applying a gradient method. The idea is to have a differentiable optimization function that penalizes segmented solutions $S(\mathbf{f})$ that both result in large residual error and have large TV norm, and there is no separate

step for boundary detection / computation of penalty matrix $\mathbf{D}$. Full treatment on TVR-DART falls outside the scope of this thesis, but some remarks about differences between TVR-DART and SDART-TV are in order:

Because of the chain derivatives involved in the gradient $\nabla_{\mathbf{f}} Q_{\text{TVR-DART}}$, the numerical gradient of $Q_{\text{TVR-DART}}$ may become very small (and thus the current value very slow to update!) for pixel (x, y) if $f(x, y) \approx S(f(x, y))$ while $f(x, y)$ is still far away from the ideal solution. Authors of [13] anticipate this behavior and recommend starting from initial reconstruction $\mathbf{f}^0$ that is initialized from a TVM solution. In the SDART-TV algorithm, both data fidelity term and the TV regularization term depend directly on $\mathbf{f}$, so the gradient will always carry information which is the locally best update direction for $\mathbf{f}$ in TVM sense; the discrete solutions are encouraged separately with the SDART penalty term. The assumptions are balanced against each other by the choice of regularization parameters α.

4.3.2 TV initialization of discrete algorithms

Another way to combine TV and DART(-like) algorithms is to build an intermediate reconstruction by TVM and then feed it as the initial reconstruction $\mathbf{f}^0$ to the discrete-valued algorithm. Here the use of DART-like algorithm can be viewed as an intelligent way to do the segmentation of a continuous solution to a discrete-valued image with the help of our imaging model $\mathbf{A}$ and measurements $\mathbf{m}$. This and other similar approaches are quite common: the first step of DART (Algorithm 2) is to build an initial reconstruction with a subroutine M (which could be TVM). See also e.g. [13, 28].

4.4 Use of global *a priori* knowledge about pipe geometry

In addition to local neighborhood prior such as TVM and DART-like priors, *a priori* information of the geometrical structure of the target object can be leveraged to improve reconstruction quality and speed. As already discussed before, in case of pipe welds we know that the region inside the pipe is empty and the non-weld parts of the pipe are supposed to consist of uniform matter (such as steel).

One way to take this kind of information into account is by application of a simple 'mask-like' constraints on the solution space: the areas in the reconstruction image domain $\mathbf{f}$ that are known to have uniform composition are fixed to a constant value during all iterations. Also, the pixels that are fixed (or 'masked away') can be dropped out from the gradient computations altogether to speed up the computations: because the pixels in masked region are forced to have constant value,

we know that the gradient inside those regions naturally would be zero.

Unfortunately, while this method to reduce the number of unknowns in the matrix equations appears very attractive, it has some practical problems. In theory, we would like to set the masked pixels inside the pipe to the attenuation coefficient of air (which is in practice radiologically opaque matter and thus 0). However, application of a *a priori* mask naively can cause numerical problems: Firstly, we need to account for noise and cupping-like artifacts that arise from the imperfect radiological model. Setting the areas known to be empty in $\mathbf{f}$ to exact zero can cause additional artifacts when the algorithm tries to use the pixels in non-masked area to explain observations in $\mathbf{m}$ that are best explained in our numerical model by producing a cupping artifact inside the masked area in the reconstructed image. To accommodate for this, we usually set the fixed ('masked') pixels in $\mathbf{f}$ to small constant $b > 0$, in practice $b = 0.0001..0.0005$ instead of $b = 0$, which appears to help somewhat in reducing artifacts. Additionally the numerical gradient can become ill-behaved at the transition boundaries between masked/not-masked areas: a fixed pixel is usually going to have a slightly different value than its neighboring free pixels. We address the numerical problems in gradient calculations by enforcing a boundary criterion where the pixels near the mask transition boundary in $\mathbf{f}$ are also forced to have zero gradient (i.e. do not change their value). In our test cases, this approach seem to result in reasonable reconstructions that are improved over the non-masked ones. For detailed discussion on the numerical results, see Chapter 6.

4.5 Estimation the gray-value thresholds

The discrete reconstruction methods studied in this thesis require segmentation by thresholding and thus method for obtaining sensible values for thresholding parameters $\boldsymbol{\tau}, \boldsymbol{\rho}$ is needed. In simulation studies we naturally can use the exact values $\boldsymbol{\rho}$ that were used to generate the projection data in the first place. For physical objects with known material composition (such as our test object, see Section 5.4), the estimates could be derived from the chemical properties of the constituent materials and other parameters of the imaging setup, but we found that physics-agnostic methods that utilize the measured sinogram data work reasonably well. Initial experiments with automatic parameter estimation methods (such as PDM-DART [65] or TVR-DART [13]) were found to be finicky, and are thus not discussed in this thesis.

4.5.1 Manual estimation of thresholding parameters from sinogram

We can easily calculate an estimate for ρ from the projection data with the Beer-Lambert equation 2.3 if we can assume that (i) X-rays are monochromatic, (ii) that

they pass through an area that is known to be constant or piece-wise constant, and (iii) we have exact knowledge of the geometry of the area the rays pass through.

In case of metal pipes, the measurement of the intensity I_0 at the unobstructed area of detector surface and the intensity I_1 after a portion of pipe known to be solid both are almost always available. Thus finding the estimate is a matter of simple calculation

$$\rho_{\text{est}} = \frac{\log I_0 - \log I_1}{s}$$

where s is distance traveled by the X-ray inside the solid metal (s can be easily calculated from the knowledge of the imaging geometry).

After obtaining estimates for ρ_i and ρ_{i+1}, a reasonable choice of the thresholding parameter τ_i can be chosen as the midpoint between them, $\tau_i = (\rho_i + \rho_{i+1})/2$. However, in studies with the physical test piece where data did not exactly conform to the discrete-value assumption, it turned out to be necessary to often tweak to the initial choices of $\boldsymbol{\rho}, \boldsymbol{\tau}$ found be the procedure described above by trial and error.

4.6 Choice of regularization parameter

All methods defined in preceding sections incorporate a regularization parameter α that controls the 'strength' of the prior relative to the residual norm (also known as data fidelity term). An optimal choice of the parameter α results in a solution where neither the prior assumption is too weak (where noise and artifacts are not suppressed) nor too strong (where the prior dominates at the expense of loss of reconstruction detail). While many methods that seek to find either an optimal value α or at least a principled guess have been proposed (for example, L-curve method [4, Section 5.4.2]), here we chose suitable value of parameter α by manual experimentation and visual inspection of the intermediate reconstructions after a couple of iterations.

Chapter 5

Simulation and experimental studies

In this chapter we describe the experiments that were conducted and the evaluation metrics used to assess the results. The feasibility of the algorithms was first assessed in tests with simulated phantom, which is described in Section 5.3. The simulation study was followed by an experimental study with a purposed-made physical phantom. The construction of the test object and the other details of experimental setup of X-ray study are described in Section 5.4. The results of the experiments are presented and analyzed in Chapter 6.

Technical limitations necessitate use of a rather small physical object compared to real-world industrial scale pipe welds. To maintain consistency between the results of the simulations and the physical experiments with the test object, the simulated phantoms were scaled to match the test object instead of the hypothetical industrial cases.

5.1 Summary of the algorithms studied

The theoretical groundwork of the algorithms studied in this thesis has been laid out the preceding chapters. The algorithms that were included in the experiments are summarized in the list below:

1. FDK. In the experimental study of circular X-ray tomography of the physical object, the standard FDK algorithm (see Section 2.2) and full, dense circular projection data was used to build the high-resolution, high-quality ground truth reconstructions.

2. TV. The TV algorithm described in the subsection 4.1.3. We set the absolute value smooth approximation parameter $\beta = 0.001$. The effect of regularization

parameter α on reconstruction quality was evaluated in preliminary tests where the value of α ranged through $\alpha \in (0.01, 100)$. We performed n iterations, which was usually more than enough to reach acceptable convergence.

3. TV-S. The basic TV algorithm is continuous. To compare it with the discrete-valued algorithms, the resulting TV-reconstruction was segmented by a threshold operation.

4. TV(-S)-A. This is used to denote the TV(-S) reconstructions that utilize the *a priori* mask.

5. SDART. The SDART algorithm [11] as presented in Subsection 4.2.3. As in the case of TVM, the reconstructions were again computed for different values of the regularization parameter α. The results marked 'SDART' denote the final reconstruction $\mathbf{f}^k$ obtained by minimizing Q_{SDART} after the final round before the last thresholding step.

6. SDART-S. The thresholded SDART reconstruction $\mathbf{f}^*_{\mathrm{SDART}} = T(\mathbf{f}^k, \boldsymbol{\tau}, \boldsymbol{\rho})$.

7. SDART-TV. The variant of SDART with the inclusion of TV regularization term in the objective function, as described in Subsection 4.3.1. Like SDART, 'SDART-TV' is the 'continuous' result before the final thresholding operation.

8. SDART-TV-S. The thresholded SDART-TV reconstruction.

As discussed in Section 4.3.2, we also investigated initializing discrete (SDART) algorithms with a preliminary TV-reconstruction or from default matrix filled with constant 0, $\mathbf{f}^0 = 0$. The details of the initialization methods used for each particular algorithm is mentioned in the Results chapter.

5.2 Evaluation metrics

To assess the quality of the reconstruction images, we quantitatively inspect the images and also employ several quantitative metrics. In numerical studies, the reconstructions could be compared against the original numerical phantom that was used to generate the simulated projection data in the first place. In the experimental study with the physical object, similar perfect reference image is not available. When applicable, we used a very high-quality FDK reconstruction as our ground truth reference. Also the geometry of the test object was known down to mm precision.

5.2.1 Visual comparison

An intuitive method to evaluate the quality of the results is simply the visual inspection of the reconstruction images, and attempting to evaluate which defects

known to be inside the target (whether in the simulated phantom or the physical object) are present in the reconstructed image, and how exactly their location and size are recovered.

5.2.2 Mean squared error of reconstruction

The first quantitative error measure we use is the classic MSE, defined as follows:

Definition 11 (MSE)**.** Denoting the ideal image as $\mathbf{f}$ and the computed solution as $\mathbf{f}^*$, the *mean squared error* or *deviation* is defined as

$$MSE(\mathbf{f}^*) = \frac{\|\mathbf{f} - \mathbf{f}^*\|_2^2}{n_d} = \frac{\sum_{x,y}(\mathbf{f}(x,y) - \mathbf{f}^*(x,y))^2}{n_d}.$$

5.2.3 Total reconstruction pixel error after segmentation

The second numerical measure we employ is the total reconstruction pixel error after segmentation, or simply the pixel error of a reconstruction [10].

Definition 12 (Pixel error)**.** Let us denote the ideal (discrete-valued) image by $\mathbf{f}$, where $\mathbf{f}(x,y) \in \{\rho_1, \dots, \rho_K\}$ and the computed (also discrete-valued) solution as $\mathbf{f}^*$, the pixel error is defined as

$$Error(\mathbf{f}^*) = \sum_{x,y} \mathbb{1}(\mathbf{f}(x,y) \neq \mathbf{f}^*(x,y)),$$

where $\mathbb{1}(x \neq y) = 1$ when $x \neq y$ and 0 elsewhere.

While both MSE and pixel error measure the fundamentally the same property (how different $\mathbf{f}^*$ is from the ideal true reconstruction $\mathbf{f}$ i.e. the phantom), notice that unlike in MSE, the pixel error is well-defined for discrete-valued images. Thus the image reconstructions by the algorithms that produce smooth reconstructions (such as TV, SDART, and SDART-TV) need to be thresholded first for the error measure to be meaningful. Furthermore, the error score is influenced only by the *amount* of misclassified pixels; the extent of the error (how far the pixel is from the correct value) does not have an effect.

To enable easy comparison of results between reconstructions of different size, we also define *relative pixel error*, which is the pixel error normalized by the number of non-zero pixels in the phantom:

Definition 13 (Relative pixel error)**.** Relative pixel error[1] is defined as

$$RelError(\mathbf{f}^*) = \frac{Error(\mathbf{f}^*)}{\sum_{x,y} \mathbb{1}(\mathbf{f}(x,y) \neq 0)}.$$

[1] Also called *relative Number of Misclassified Pixels (rNMP)* [65].

5.3 Materials for simulation studies

Here we describe the simulated test phantom created for the computational experiments.

The first series of experiments consisted of a study with simulated numerical phantom to validate the soundness of the methods. The phantom A.1 in Figure 9 corresponds to the regular circular tomography of a pipe weld with some small, localized voids. Simulations were carried out in Matlab and with ASTRA toolbox.

As discussed in Section 2.8, the real-world projection data $\mathbf{m}$ always has some amount of noise. To study the robustness of the algorithms in the presence of noise, we added varying amounts of i.i.d. Gaussian noise to the simulated projection data $\mathbf{m}$.

5.3.1 Simulated phantom for tomography geometry

For simulations the circular tomography setup, we created one phantom (A.1) that simulates a single two-dimensional cross-section along the girth weld seam of the pipe. As already discussed, we want the simulation to match the physical experiment, and thus the object is assumed to be rather small with diameter of 50mm and thickness 2mm. The pipe weld is assumed to contain only a single group of two small void-like defects: the voids have diameters 0.3 mm and 0.5 mm , their centers have distance of 0.6 mm and 1.0 mm from the outer surface and they have horizontal displacement of 0.0 mm and 1.0 mm (as measured from the vertical center line), respectively.

5.3.2 Geometry in simulation study

In all simulations with the phantom A.1, the focal point of the X-ray source was assumed to be located 200 mm from the center of the object and the detector (width of 100 mm, 10 pixels per 1 mm) mirrored exactly on the other side of the object, also 200 mm from the center of the object. We assumed a fan beam model and line kernel.

5.4 Experimental study of physical object

A physcial test object was created at the Workshop of the Department of Physics of University of Helsinki for radiographic experimental study. Two sets of measurements were done: one series to obtain the regular circular tomography data, and another to study the tomosynthesis-like limited angle problem. All measurements were conducted in University of Helsinki X-ray Micro-Imaging Laboratory with a GE Phoenix Nanotom device.

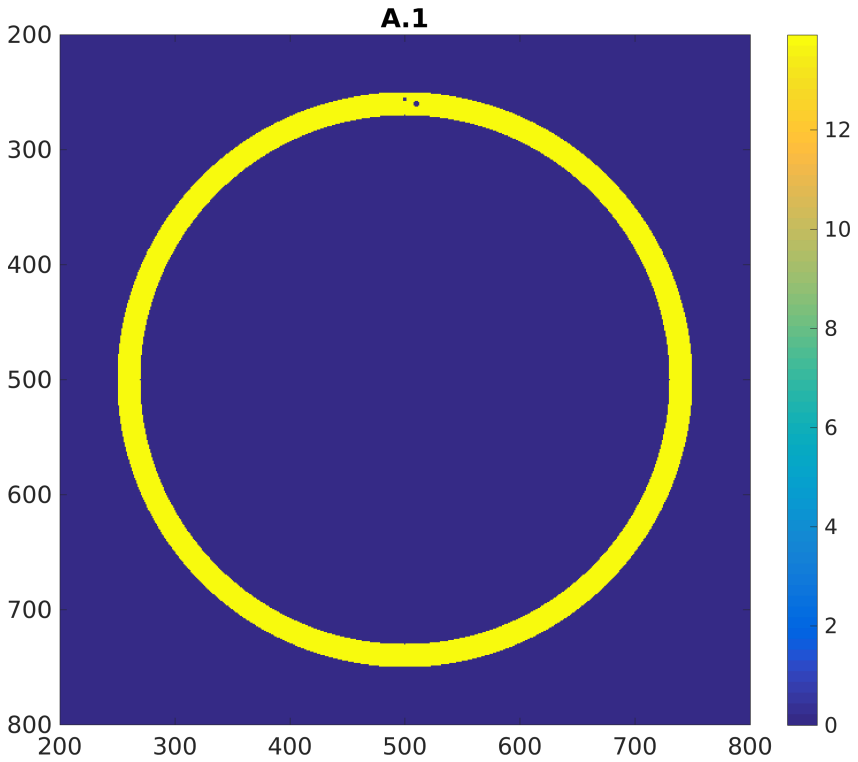


Figure 9: Phantom A.1 (single area of localized defects).

5.4.1 Description of the test object

The physicial phantom object consists of two uniform aluminum pipe sections (length 100mm each, outer diameter 50mm, wall thickness 2mm), which when joined together form a single pipe that is 200mm long. The transversal surfaces of both segments were milled with CNC machine to be as smooth and level as possible (so that the pipe sections could be joined practically seamlessly to create a single pipe), and three holes were drilled into one surface to simulate voids: the holes become voids enclosed inside a 'weld' when the sections are joined. (The voids had outer diameters 0.3 mm, 0.5mm and 0.8mm, each 1mm deep, located at uniform 120 degree intervals from each other.) The segments were joined and held together by wrapping aluminum coated adhesive tape around the seam (which also simulates rough outer weld surface). The construction of the test object is illustrated in the Figure 10.

5.4.2 Circular tomography geometry and FDK ground truth

In the first series of measurements a regular CT scan was performed: the object was attached in up-right position to a sample manipulator located between a X-ray tube

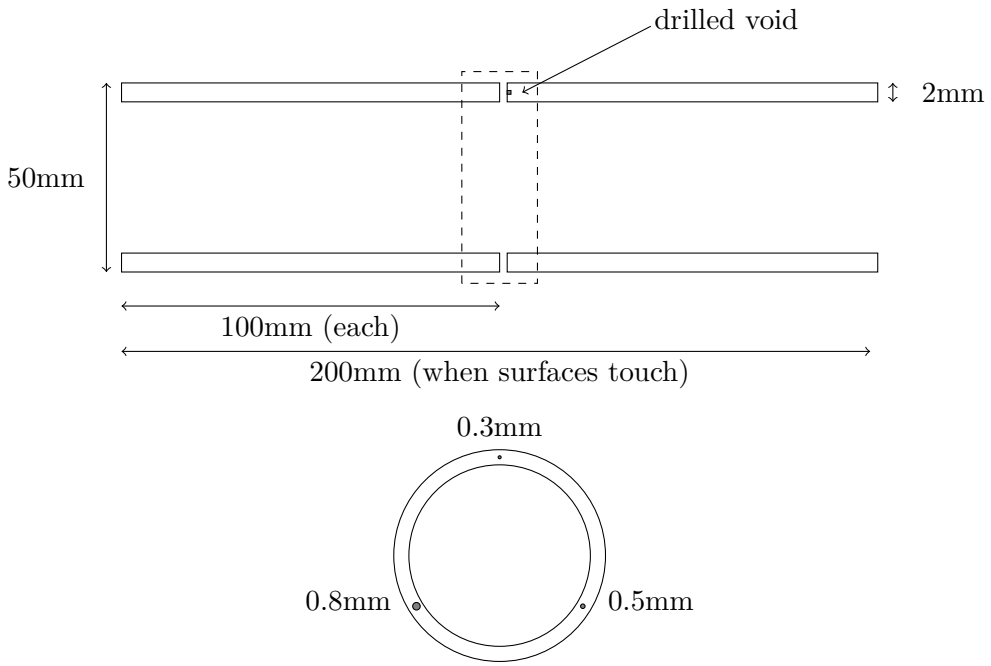


Figure 10: Illustration of the test object construction (not to scale). Two 100mm long pipe segments are pressed together to form a single 200mm pipe with three 'voids' enclosed in a seam; approx 25mm wide and <1mm thick layer of aluminum adhesive tape (dashed rectangle) is wrapped around the seam to keep them joined and simulate the rough weld surface.

and a detector panel.[2] The object was rotated full 360 degrees, and total of 1080 projections were taken (a 1/3 degree step between each subsequent projection).[3]

The full tomography data was used to build a three-dimensional FDK reconstruction of the object. The relevant horizontal slice of the reconstruction was extracted and it serves as our ground truth reference image in our reconstruction algorithm studies. The reference FDK reconstruction slice is shown in the Figure 11.

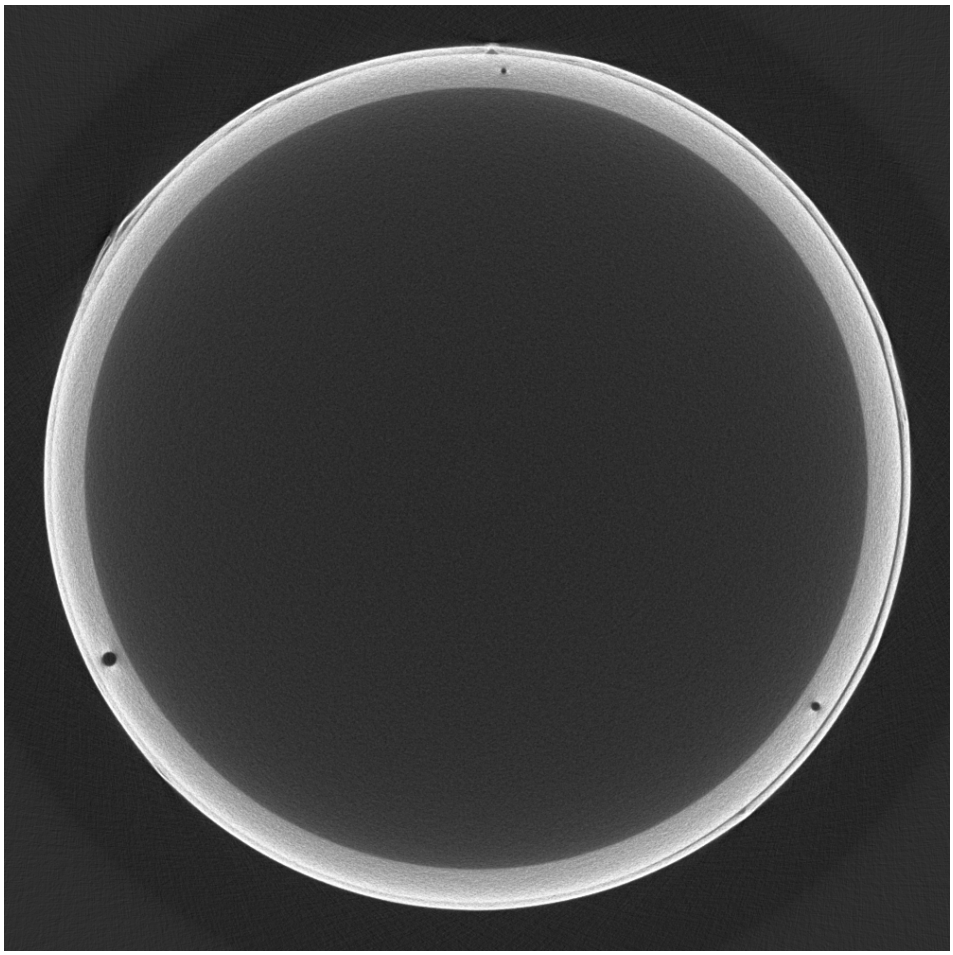


Figure 11: Transversal slice of FDK reconstruction at the void depth midpoint. Notice the tape layer and the light cupping artifact. (Color curves adjusted for printing.)

[2]The exact geometry was as follows: Distance between X-ray source and the center of object (focus-origin distance, FOD) was 220 mm, distance between X-ray source and detector (focus-detector distance, FDD) 400 mm, detector width 12 cm.

[3]The X-ray tube's operation parameters were as follows: acceleration voltage $90kV$, X-ray tube current at 200μA; no filtering was used. Parameters for the imaging device: single projection imaged constructed by 4 images taken at 250 ms exposure each and computing an average; voxel size 0.055 mm

5.4.3 Tomosynthesis geometry

The second series of measurements were conducted in the tomosynthesis geometry: The phantom object was placed in a horizontal position between the X-ray tube and a (simulated[4]) 36cm wide detector. While in the description of the tomosynthesis geometry outlined in Section 3.2.3 we discuss a setup where X-ray source moves in relation to a pipe and a detector that remain in a fixed position (see Figure 6), this particular arrangement was not possible with the machinery available in the laboratory, so the equivalent geometry was achieved by moving the pipe and cutting the sinogram accordingly. A total of 31 projection images were taken with 10 mm step (and thus the total pipe movement was 300mm). We note that the computer controlled manipulator was not available and the phantom was shifted manually.

Based on the previous experience from the first measurement series, the X-ray tube and imaging parameters were also adjusted slightly.[5]

[4] To enable as large 'angle of view' as possible, each single projection image was in fact a composition of three images. The detector that is 12cm wide was moved horizontally by its width between each measurement.

[5] Acceleration voltage $70kV$, tube current 150μA. Exposure time for a single projection was 1000 ms.

Chapter 6

Results and discussion

In the previous chapter we have described the simulation and experimental studies that were conducted. Here we present and discuss the results of the aforementioned experiments.

6.1 Reconstructions in regular circular tomography

In regular tomographic geometry, even the smallest voids both in simulations and physical experiment could be recovered in regular tomography setup with relatively sparse angular sampling and full-width detector. With a limited-width detector, some precision is lost but the 0.5mm pore in the physical phantom could be reconstructed quite reliably. We first present the results of simulated phantom with full detector width and two angular sampling schemes to validate and compare the different methods, followed by validation with physical phantom (Subsection 6.1.1), and then the results with physical phantom with reduced FOV and sparse sampling (Subsection 6.1.2). For the simulation studies, here we assumed that segmentation parameters $\boldsymbol{\rho}, \boldsymbol{\tau}$ are approximately known beforehand.

6.1.1 Sparse-angle tomography and the effect of the global prior

Simulated phantom

The results of the algorithms with varying number of projections are presented in Figures 12,13,14 and 15 for the phantom A.1; in this series of reconstructions, we assumed a full-width detector and a prior mask. To demonstrate the effectiveness of the prior mask, also the ground truth and the results of TV reconstruction without prior mask are included.

The details of the numerical results for the simulations studies (both the notes and quantitative error measures could be computed against the simulated phantom)

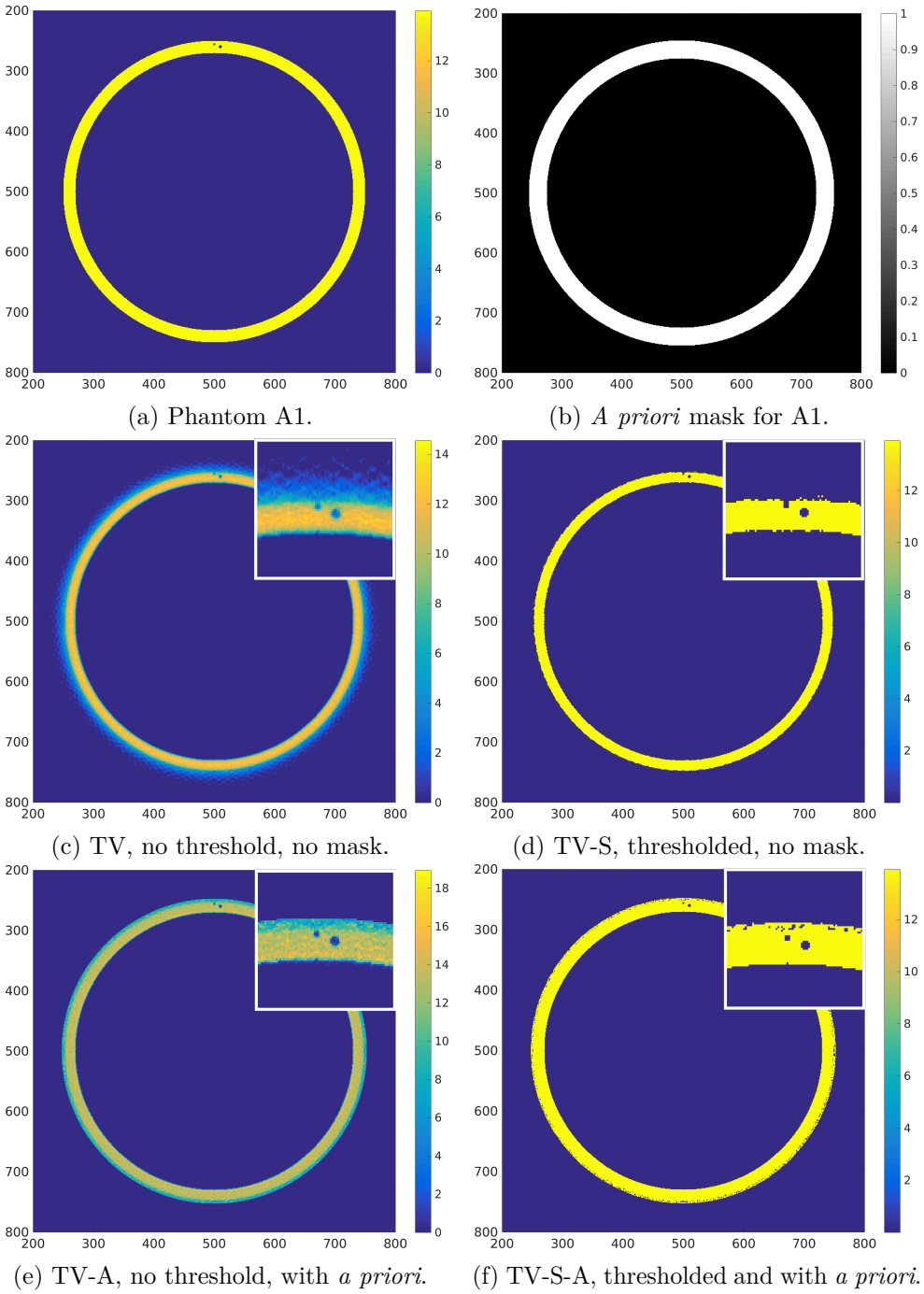
are included in Table 6.1. Notice wide discrepancy between the numerical errors between TV and SDART(-TV): while both methods recover the voids with similar precision, the ring-like band in the TV solution is slightly too wide.

Looking at the simulation study results (Figures 12 and 13), we can see that guiding TV with *a priori* mask is useful, resulting in less pronounced artifacts after $k = 20$ BB iterations (and details such as small void near the weld surface is correctly rendered as an enclosed inside the object and not as a hole on its surface as in 12d). Similar improvement can also be seen in the numerical measures recorded in Table 6.1.

Table 6.1: Numerical details of reconstructions of phantom A.1. (Figures 12, 13,14, 15.)

Algorithm	Notes	No. pr.	MSE	Error[1]
TV	$\alpha = 80$, no *a priori* mask	108	739.5213	-
	$\mathbf{f}^0 = 0$, iterations $k_{\max} = 20$	54	1183.3203	-
TV-S	same as TV above +	108	926.566	32.3993
	manual thresholding params $\boldsymbol{\tau}, \boldsymbol{\rho}$	54	1220.8376	32.4499
TV-A	same as TV but with *a priori* mask,	108	453.7563	-
	no thresholding	54	516.3895	-
TV-S-A	same as TV but with *a priori* mask	108	693.4730	32.2450
	+ ideal threshold params $\boldsymbol{\tau}, \boldsymbol{\rho}$	54	687.4714	32.2492
SDART	$\alpha = 0.01$, $r = 2$, $k_{\max}^{\text{SDART}} = 5$, $k'_{\max} = 15$,	108	309.3478	-
	$k'_{\text{init}} = 10$, $\mathbf{f}^0 = 0$, *a priori* mask, $c = 2$,	54	388.4435	-
	manual threshold params $\boldsymbol{\tau}, \boldsymbol{\rho}$			
SDART-S	same as SDART above +	108	318.5870	0.5183
	final threshold step	54	462.3124	0.5393
SDART-TV	$\alpha_{\text{TV}} = 5, \alpha_{\text{SDART}} = 0.01$, $\mathbf{f}^0 = 0$	108	244.2580	-
	(all others same as in SDART)	54	417.9092	-
SDART-TV-S	same as above + final threshold	108	176.7547	0.5520
		54	478.9612	0.5481

[1] relative pixel error after threshold



(a) Phantom A1.

(b) *A priori* mask for A1.

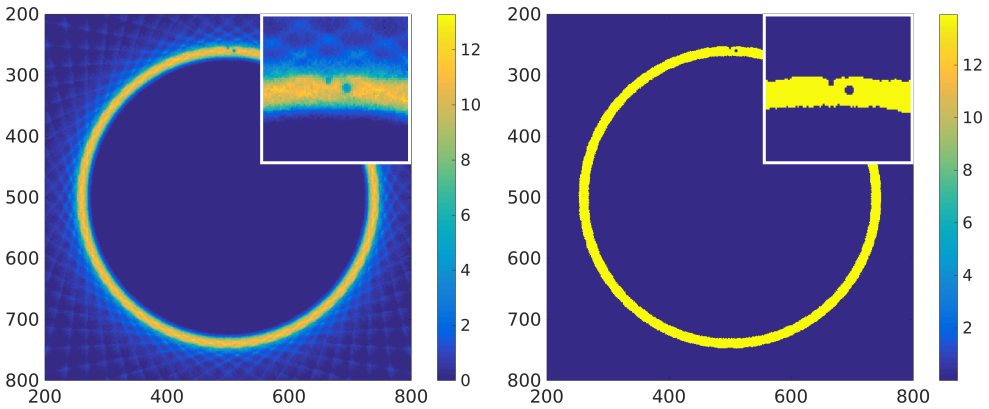
(c) TV, no threshold, no mask.

(d) TV-S, thresholded, no mask.

(e) TV-A, no threshold, with *a priori*.

(f) TV-S-A, thresholded and with *a priori*.

Figure 12: Comparison of TV reconstruction results (108 projections, full detector width) on simulated phantom A1.



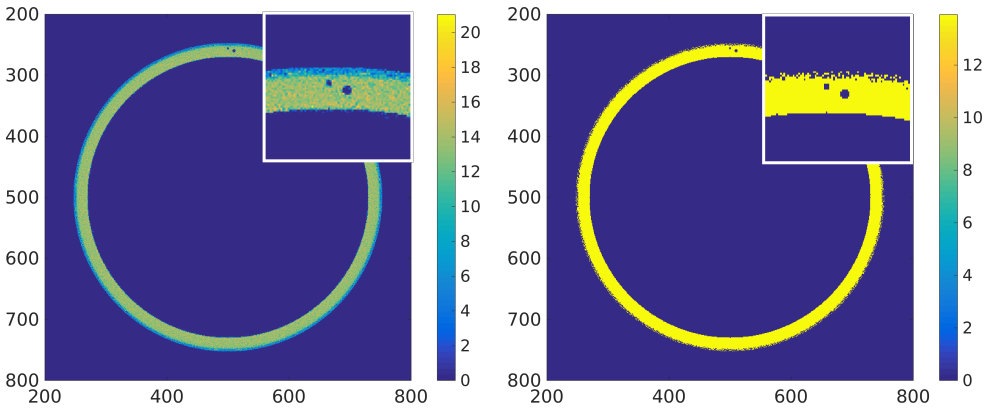
(a) TV, no threshold, no mask.

(b) TV-S, thresholded, no mask.

(c) TV-A, no threshold, with *a priori*.

(d) TV-S-A, thresholded and with *a priori*.

Figure 13: Continuation of Figure 12; Comparison of TV reconstruction results (54 projections, full detector width) on simulated phantom A1.



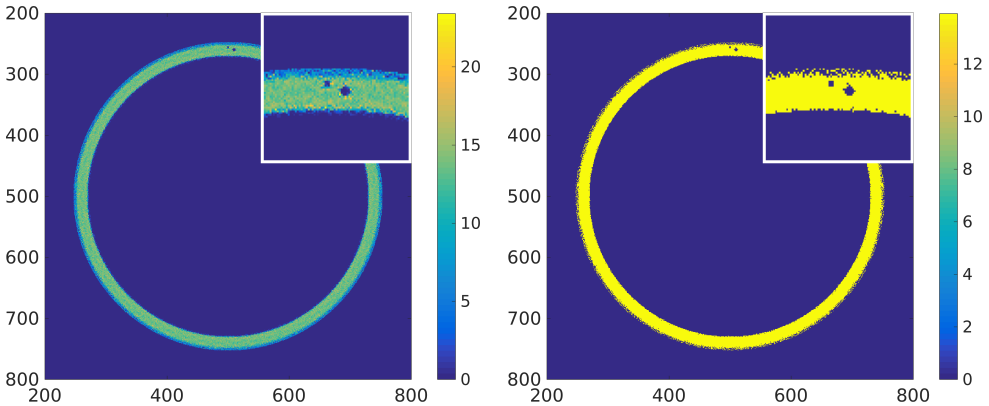
(a) SDART before final threshold.

(b) SDART-S (after final threshold).

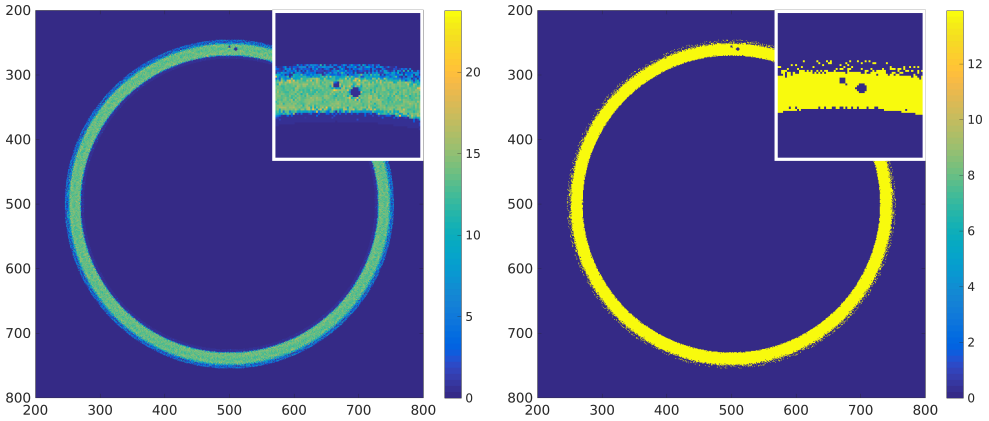
(c) SDART-TV before final threshold.

(d) SDART-TV-S (after final threshold).

Figure 14: SDART and SDART-TV reconstruction results (108 projections, full detector width) on simulated phantom A1.

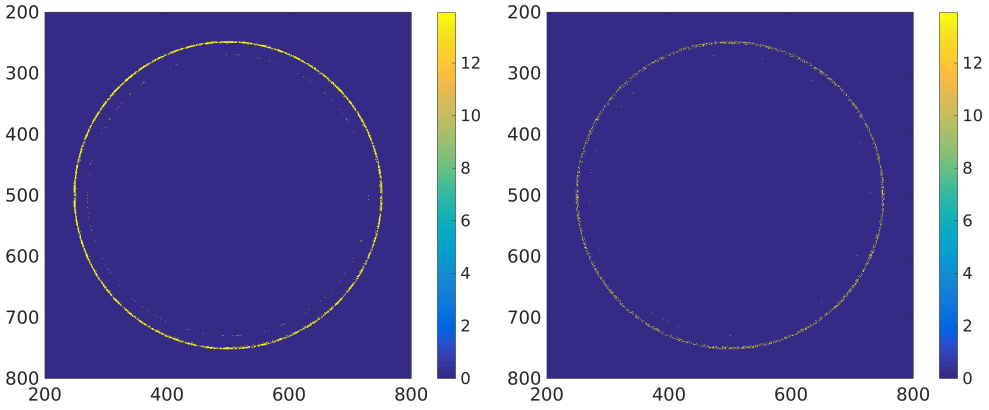


(a) SDART before final threshold. (b) SDART-S (after final threshold).



(c) SDART-TV before final threshold. (d) SDART-TV-S (after final threshold).

Figure 15: SDART and SDART-TV reconstruction results (54 projections, full detector width) on simulated phantom A1.



(a) abs(TV-S-A - A.1). (b) abs(SDART-S - A.1) .

Figure 16: Highlight of reconstruction differences: absolute difference between TV-S-A/SDART-S and simulated phantom A1 (108 projections, full detector width).

Physical test piece phantom

The results for physical phantom in full data (108projections , full detector width) are presented below. The numerical details are given in Table 6.2. Results are in line with the simulation experiments, but the reconstruction task is more difficult.

In the full-angle reconstructions with TV Figure 17a, even the smallest sub-micrometer anomalies (such as air bubble under the tape that holds the pipe segments together) can be seen.

When the number of projections available for reconstruction is decreased, predictably some of the precision is lost. Nevertheless, even the smallest 0.3mm void can be recovered.

6.1.2 Effect of width of the detector and TV initialization

We studied how our algorithms fared in the limited-width detector task by symmetrically cutting off portion of the observed full-view sinogram to 301 px 'limited-view' sinogram (which corresponds to using a detector that is approximately 30mm wide; given the pipe diameter and the distances between the phantom and the detector and the X-ray tube, in thus restricted view about 34% of the pipe can be seen). Here we present only results for the physical phantom.

The results of TV, SDART and SDART-TV algorithm with 54 projection images and with simulated 301 px detector width are presented in Figures 18, 21 and 23. The numeric parameters used to create reconstructions are recorded in Table 6.3. The prior mask method was applied as indicated.

As expected by the geometric analysis presented in Section 3.2.4 of Chapter 3, limiting the width of the detector increases the difficulty of the recovery task for all algorithms. To further illustrate the effect of the mask, in Figure 18 we have included a comparison between a reconstruction by TV-algorithm for 301px case without the mask and with a constant zero mask.

As discussed in Section 4.3.2, given already a good intermediate reconstruction obtained by TVM, a couple of SDART or SDART-TV iterations seem to function as an improved segmentation method compared to a naive thresholding of TV reconstruction for same $\boldsymbol{\rho}, \boldsymbol{\tau}$. This is demonstrated in Figures 21b and 21c.

Table 6.2: Numerical details of reconstructions of test object (tomography setup, full detector width) in Figures 17, 19 and 22. The algorithms ran full $k_{\max}$ iterations unless otherwise indicated.

Algorithm	Notes	No. pr.	Det. width
TV-A	$\alpha = 10$, $k_{\max} = 20$, with *a priori* mask, $\mathbf{f}^0 = 0.002$, no threshold, $\epsilon_{\text{residual}} = \epsilon_{\text{g}} = 0.01$	108	full 1128 px
TV	same as above but no mask, converged at $k = 12$	108	full
SDART	$\alpha = 0.1$, $\boldsymbol{\rho} = [0.0001, 0.01]$, $\boldsymbol{\tau} = [0.006]$ and $[0.008]$ $r = 2$, $k_{\max}^{\text{SDART}} = 5$, $k'_{\max} = 15$, $\mathbf{f}^0 = 0.0001$	108	full
SDART-S	same as SDART but with final threshold step	108	full
SDART-TV	$\alpha_{\text{TV}} = 5, \alpha_{\text{SDART}} = 0.1$, $\boldsymbol{\rho} = [0.0001, 0.01]$, $\boldsymbol{\tau} = [0.008]$ $r = 2$, $k_{\max}^{\text{SDART}} = 5$, $k'_{\max} = 10$, $\mathbf{f}^0 = 0$	108	full
SDART-TV-S	same as SDART-TV but with final threshold step	108	full

Table 6.3: Numerical details of reconstructions of test object (tomography setup, limited detector width).

Algorithm	Notes	No. pr.	Det. width
TV	$\alpha = 10$, $k_{\max} = 20$, $\mathbf{f}^0 = 0.002$, converged at $k = 13$ no threshold, $\epsilon_{\text{residual}} = \epsilon_{\text{g}} = 0.01$	5	301 px
TV-A	same as above but with mask, converg. $k = 13$	54	301 px
SDART	same as full data SDART but $\boldsymbol{\tau} = [0.008]$, $\mathbf{f}^0$ =TV-A	54	301 px
SDART-S	same as SDART but with final threshold step	54	301 px
SDART-TV	same as full data SDART-TV but $\mathbf{f}^0$ =TV-A	54	301 px
SDART-TV-S	same as SDART-TV but with final threshold step	54	301 px

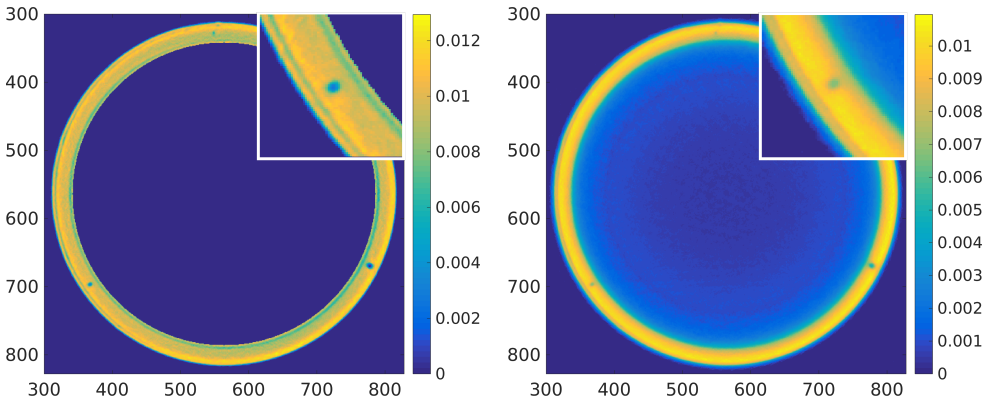
6.2 Reconstructions in limited-angle tomosynthesis geometry

We also present some reconstructions of the test object in the tomosynthesis geometry. Because of the limitations of the test setup (as discussed in Section 5.4.3, projection geometry had to measured manually which incurred some loss of precision), the results are not as good as in the tomographic geometry (where we were able to move the test piece be moved with computer guided machinery). However, the single 0.5 mm void can be seen in the TV solution, and its outline can be improved in the segmented solution by SDART-TV with appropriate choice of masks and threshold parameters. The reconstructions are shown in Figures 24.

[2]hit convergence limit at $k = 3$

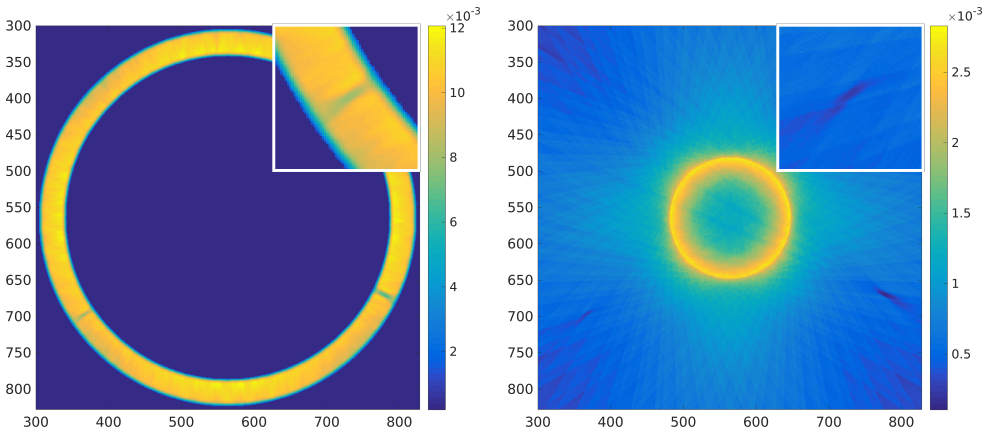
[3]manually estimated from TV solution

[4]See Fig. 26b: changes are allowed only in the region of interest.



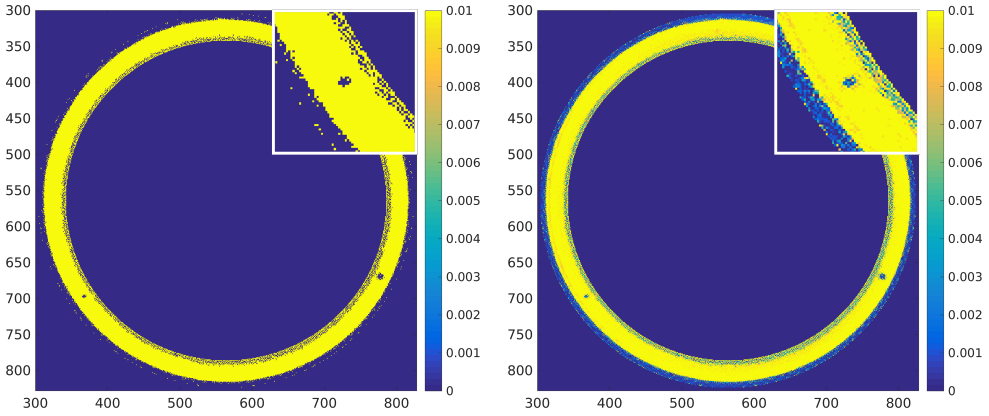
(a) TV-A, 108 proj., full width. (b) TV, 108 proj., full width.

Figure 17: TV reconstructions, 108 projections and full detector. Notice that while TV-A reconstructs the voids more clearly than non-mask TV, this comes at the price of ring-like artifact near the inner mask boundary.



(a) TV-A, 54 proj., 301 px width. (b) TV, 54 proj., 301 px width.

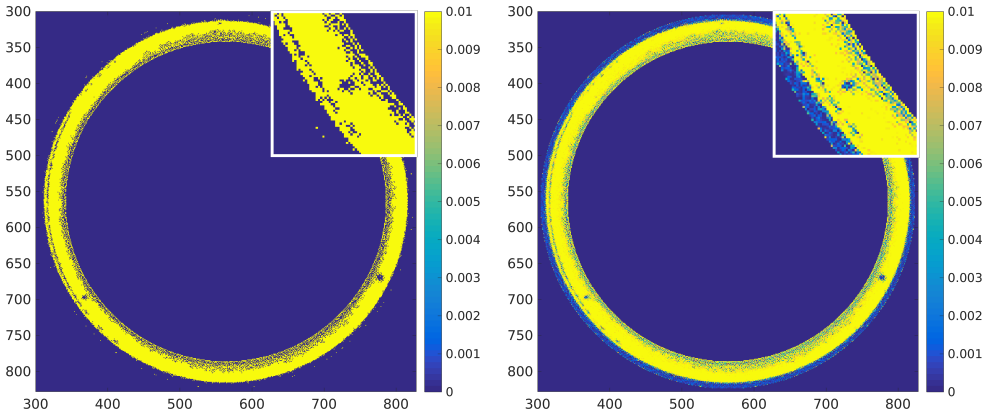
Figure 18: TV reconstructions, 54 projections and limited detector. Comparing the with-mask reconstruction (TV-A) to regular TV reconstruction that does not utilize *a priori* information, it is clear that the prior constraints become crucial for obtaining a solution of any quality. The voids in TV-A reconstruction are elongated along the angular directions which correspond to the limited-width projection images $\mathbf{m}$ where the respective void is within the 301 px field of view.



(a) SDART-S, 108, full width, with mask, $\tau = 0.006$.

(b) SDART, 108, full width, with mask (before final threshold).

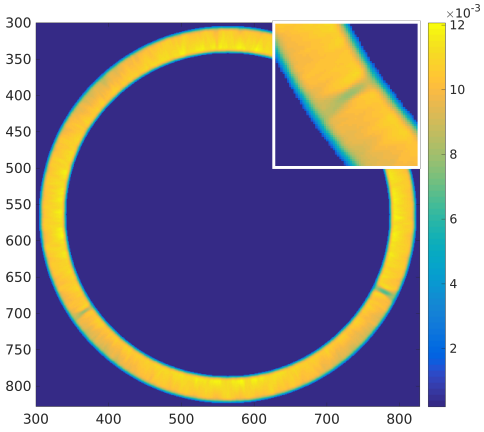
Figure 19: SDART reconstructions, 108 projections and full detector width. Compared to TV, the SDART reconstructions are noisier but conform better to the discrete-valued assumption. The smallest void can be observed in the continuous reconstruction, but thresholding.



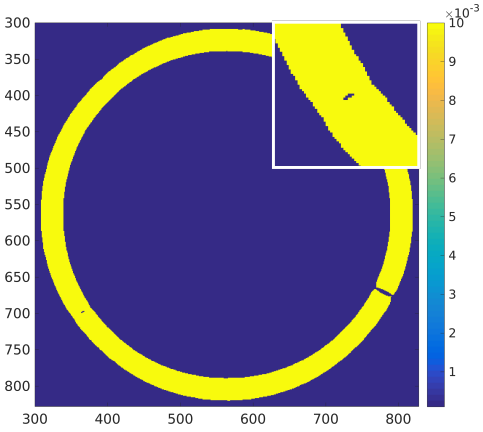
(a) SDART-S, 108, full width, $\tau = 0.008$.

(b) SDART, 108, full width, $\tau = 0.008$.

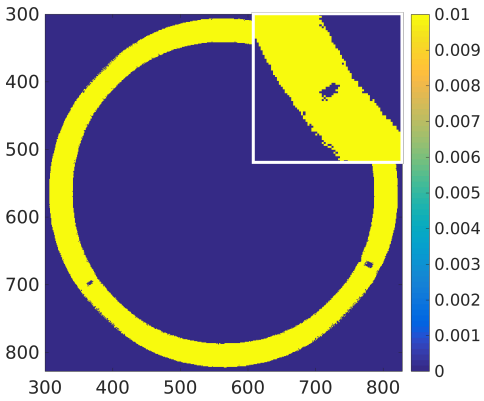
Figure 20: SDART reconstructions, 108 projections. Otherwise same as in Fig. 19, but with different threshold parameter for comparison against Fig. 22.



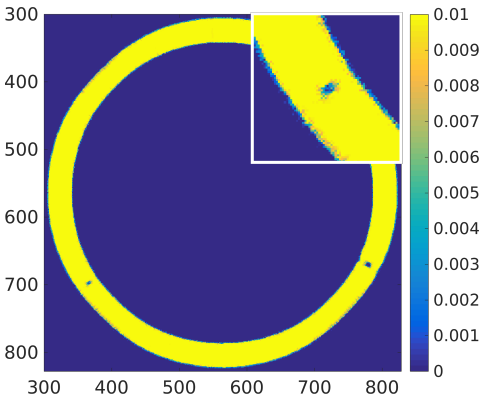
(a) Unsegmented TV-A reconstruction used as $\mathbf{f}^0$.



(b) Thresholded TV-A reconstruction with $\tau = 0.008$.

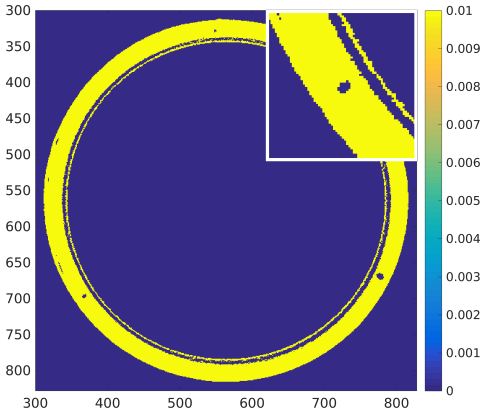


(c) SDART-S, 54, 301 px.

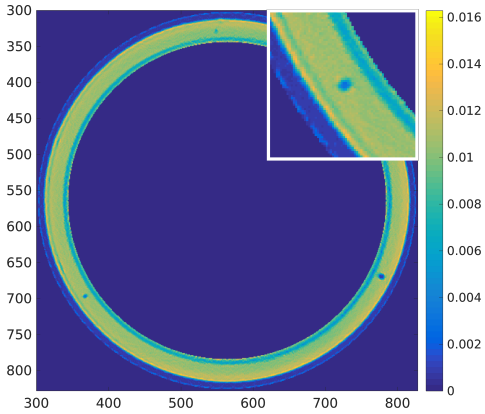


(d) SDART, 54, 301 px, before final threshold.

Figure 21: SDART reconstructions, 54 projections and limited data. In limited data case SDART needs an already good initial solution $\mathbf{f}^0$, but the algorithm manages to improve the segmentation result (Fig. 21c) compared to naive thresholding (Fig. 21b).

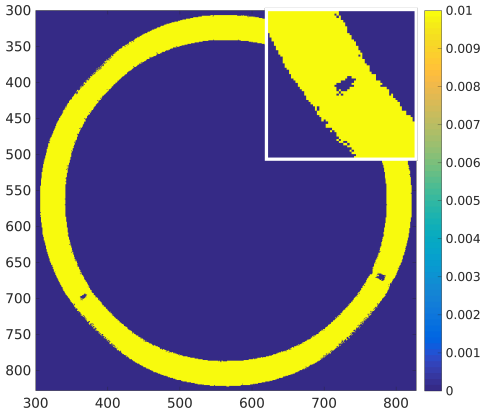


(a) SDART-TV-S, 108, full width, with mask.

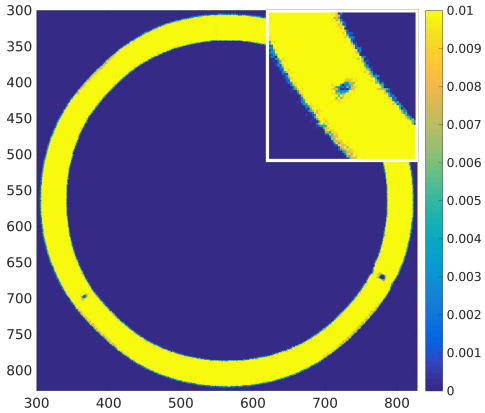


(b) SDART-TV, 108, full width, with mask (before final threshold).

Figure 22: SDART-TV reconstructions, 108 projections and full data. Compared to SDART solutions with same projection data (even to SDART with same threshold parameters as in Fig 20), the smallest void can be seen more clearly. As a drawback the ring-like artifact near the inner mask boundary is stronger.

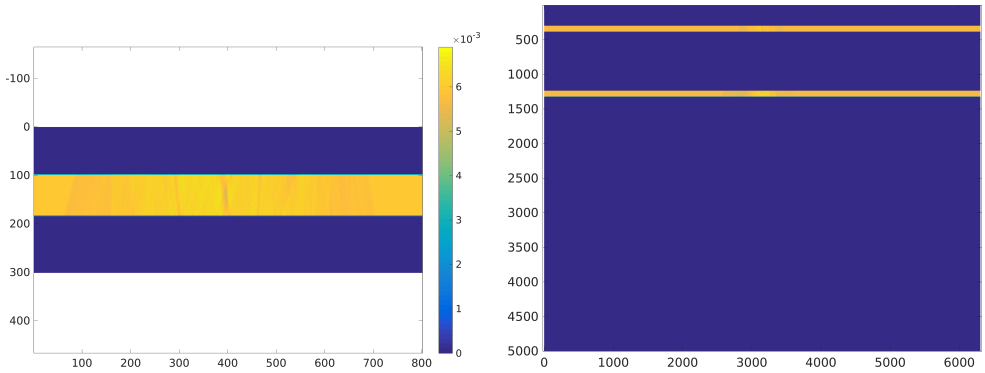


(a) SDART-TV-S, 54, limited width, with mask.



(b) SDART-TV, 54, limited width, with mask (before final threshold).

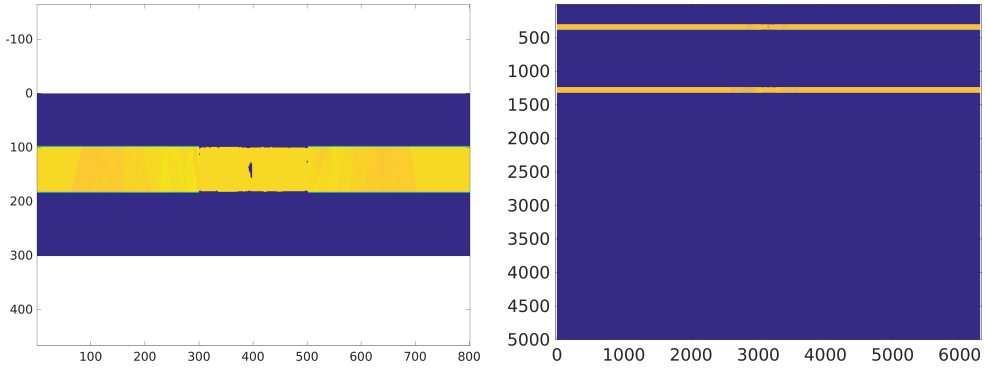
Figure 23: SDART-TV reconstructions, 54 projections and limited data. Also initialized with TV-A solution. In limited data case we don't have much improvement over regular SDART.



(a) TV-A, zoomed in to region of interest.

(b) The full TV-A solution.

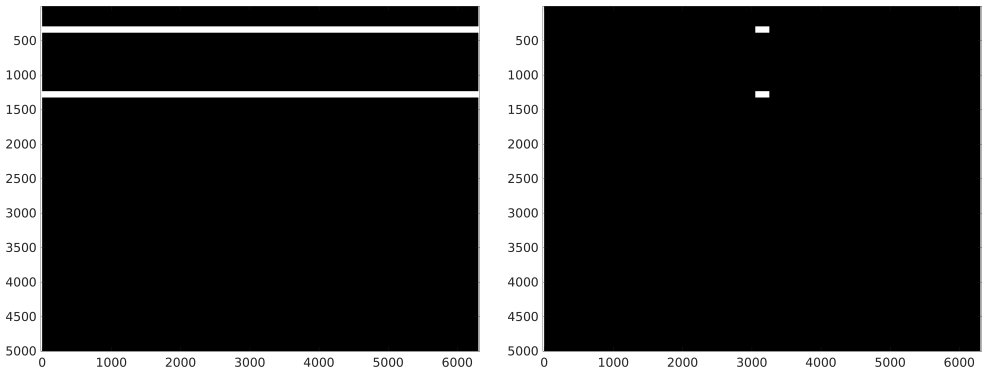
Figure 24: TV reconstruction in limited-angle tomosynthesis geometry.



(a) SDART-TV, zoomed in to region of interest.

(b) The full SDART-TV solution.

Figure 25: SDART-TV reconstruction in limited-angle tomosynthesis geometry. Choice of masks results in some artifacts near the mask boundaries (and masked parts of initial solution are not updated), but the void can be seen clearly.



(a) The mask used in TV-A reconstruction.

(b) The mask used in SDART-TV reconstruction.

Figure 26: Masks used in limited-angle tomosynthesis geometries.

Table 6.4: Numerical details of reconstructions of test object (tomosynthesis geometry) in Figures 24 and 25.

Algorithm	Notes	Projections
TV	$\alpha = 5$, with *a priori* mask, $\epsilon_{\text{resid.}} = \epsilon_{\text{g}} = 0.5^2$ $\mathbf{f}^0 = 0.006\times$ mask	5 projections, 3cm step
SDART-TV-S	$\alpha_{\text{TV}} = 5$, $\alpha_{\text{SDART}} = 0.01$, $\mathbf{f}^0 =$ TV solution $\boldsymbol{\rho} = [0.0001, 0.006], \boldsymbol{\tau} = [0.0055]^3$ and a more restrictive mask[4]	(same as above)

6.3 Discussion on gray level and threshold estimation in discrete algorithms

To produce the results presented above, the segmentation parameters were assumed to be known *a priori*, either from the construction of phantoms (in simulation studies) or estimated from the sinogram (Section 4.5.1) and initial TV reconstructions with the knowledge of structure of the target object (experimental study). The best results for DART-like algorithms could be obtained only by the manually selecting the optimal thresholds with the help of initial continuous reconstructions and knowledge of object structure, which can be impractical and unreliable in more realistic imaging situations.

Chapter 7

Conclusions

In this work we have studied how certain tomographic reconstruction algorithms (TVM, SDART, SDART-TV) can be used to solve the problem of determining sizes and locations of small, gaseous voids in metal tubes with computerized X-ray tomography. The problem is inspired by real-life task of non-destructive testing and quality control of steel pipe welds. The algorithms studied in this thesis were based on total variation regularization and DART-like methods of discrete tomography. We have also given a thorough introduction to theoretical background of X-ray computed tomography and basics of linear inverse problems, which included description of the classical FBP algorithm based on the Radon inversion formula.

The performance of the algorithms was investigated in numerical experiments with simulated phantom and X-ray imaging experiments with a physical aluminum tube phantom. For example, the algorithm SDART-TV was capable of distinguishing a 0.5mm empty void inside a 2mm thick aluminum pipe of diameter 50mm using 54 projections around the object, compared to classical FBP reconstruction which require 1080 projections.

However, all methods studied come with limitation that they require some amount of manual calibration. In TVM, SDART, and SDART-TV, a regularization parameter must be chosen. We here used *ad hoc* method where we chose the best parameter value after constructing several initial tries; in future studies, investigation of a more principled method for choosing the parameter would be in order.

The algorithms based on discrete-value assumption (that is, DART-like algorithms SDART and SDART-TV) also require a method for choosing the segmentation parameters (gray value levels and thresholds). In our experiments with imaging of a physical test piece, it was found that obtaining optimal choices of ρ, τ required some amount of hand-tuning; such difficulties may be attributed to the presence of distortions in measurement data and model that mildly violate the discrete-value assumption in the reconstruction image domain. However, the amount of

hand-tuning required is significant for any practical applications. This suggests that in order to successfully implement DART-like algorithms for NDT tasks in industrial environments, a (possibly lengthy) calibration step may be required for each individual configuration of measurement equipment and object of interest.

Also, while the prior information about pipe geometry was utilized and deemed useful, use of it was implemented by simply applying fixed masks (fixed in both shape and value), which can become problematic in real-life in-service situations because full exact knowledge might not be available (as opposed to a study of purpose-made phantom). A proper probabilistic prior model that quantifies both our assumptions and their uncertainty could prove fruitful.

The experiments with computer simulations and physical phantom suggest that computed tomography provides a plausible method for detecting porous defects, and modern algorithms such as TVM and DART-like methods enable use of relatively limited data. However, both simulation and physical models considered here were relatively simplistic, and a full proof-of-concept work of non-destructive testing of steel pipe welds would require more extensive models and larger scale tests with real-sized steel objects.

Appendix A

Gradient descent with Barzilai-Borwein step-length selection

Here we explain the reasoning for BB step-length rule of Barzilai and Borwein [62] in more detail.

A.1 Quasi-Newton algorithms

As already discussed in Section 4.1.3, solving for line search equation (4.12) results in Newton iteration step

$$\mathbf{f}^k_{\text{Newton}} = \mathbf{f}^{k-1} - \mathbf{B}^{-1}\nabla Q(\mathbf{f}^{k-1}). \tag{A.1}$$

In regular Newton's method $\mathbf{B} = \mathbf{B}_{k-1}$ is Hessian of Q at point $\mathbf{f}^{k-1}$. In quasi-Newton methods, an approximation of $\mathbf{B}_k$ is chosen to satisfy *the secant equation*,

$$\mathbf{B}_k \cdot (\mathbf{f}^k - \mathbf{f}^{k-1}) = \nabla Q(\mathbf{f}^k) - \nabla Q(\mathbf{f}^{k-1}) \tag{A.2}$$

$$\mathbf{B}_k \mathbf{y}_k = \mathbf{g}_k, \tag{A.3}$$

where $\mathbf{y}_k = \mathbf{f}^k - \mathbf{f}^{k-1}$ and $\mathbf{g}_k = \nabla_{\mathbf{f}} Q_\beta(\mathbf{f}^k) - \nabla_{\mathbf{f}} Q_\beta(\mathbf{f}^{k-1})$.

A.2 Derivation of BB algorithm

In the Barzilai-Borwein method, we want to choose a step-length s_{k-1} such that the BB iteration

$$\mathbf{f}^k_{\text{BB}} = \mathbf{f}^{k-1} - s_{k-1}\nabla Q(\mathbf{f}^{k-1}) \tag{A.4}$$

is near the quasi-Newton iteration, $\mathbf{f}^k_{\text{BB}} \approx \mathbf{f}^k_{\text{Newton}}$ or

$$\mathbf{B}^{-1}\nabla Q(\mathbf{f}^{k-1}) \approx s_{k-1}\nabla Q(\mathbf{f}^{k-1}).$$

From the secant equation (A.3) we can derive

$$\begin{aligned}\mathbf{B}_k\mathbf{y}_k &= \mathbf{g}_k\\ (s_k I)^{-1}\mathbf{y}_k &\approx \mathbf{g}_k\end{aligned}$$

and thus s_k that approximates quasi-Newton update best can be found by solving

$$\begin{aligned}s_k^{-1} &= \arg\min_{s^{-1}} \left\|s^{-1}\mathbf{y}_k - \mathbf{g}_k\right\|_2^2\\ &= \frac{\mathbf{y}_k^T\mathbf{g}_k}{\mathbf{y}_k^T\mathbf{y}_k}\\ s_k &= \frac{\mathbf{y}_k^T\mathbf{y}_k}{\mathbf{y}_k^T\mathbf{g}_k}.\end{aligned}$$

Symmetrically one can also derive an alternative BB update rule s'_k by noting that equivalently one can write

$$\begin{aligned}\mathbf{B}_k\mathbf{y}_k &= \mathbf{g}_k\\ \mathbf{y}_k &= \mathbf{B}_k^{-1}\mathbf{g}_k \approx s'_k\mathbf{g}_k.\end{aligned}$$

The s'_k then can be found by solving

$$s'_k = \arg\min_{s} \left\|\mathbf{y}_k - s\mathbf{g}_k\right\|_2^2, \tag{A.5}$$

which results in

$$s'_k = \frac{\mathbf{y}_k^T\mathbf{g}_k}{\mathbf{g}_k^T\mathbf{g}_k}.$$

One can choose to use either the rule s_k or the rule s'_k, or alternate between them.

Appendix B

Moore-Penrose pseudo-inverse

Here we define the Moore-Penrose pseudo-inverse [66, p. 4] $\mathbf{A}^+$ as a matrix that satisfies

$$\begin{aligned}\mathbf{A}\mathbf{A}^+\mathbf{A} &= \mathbf{A}\\ \mathbf{A}^+\mathbf{A}\mathbf{A}^+ &= \mathbf{A}^+\\ (\mathbf{A}\mathbf{A}^+)^H &= \mathbf{A}\mathbf{A}^+\\ (\mathbf{A}^+\mathbf{A})^H &= \mathbf{A}^+\mathbf{A}\end{aligned}$$

where $\mathbf{B}^H$ denotes the complex conjugate of matrix $\mathbf{B}$. The equation (2.17) for invertible $\mathbf{A}^T\mathbf{A}$ then follows

$$\begin{aligned}\mathbf{A}\mathbf{f} &= \mathbf{m}\\ \mathbf{A}^T\mathbf{A}\mathbf{f} &= \mathbf{A}^T\mathbf{m}\\ \mathbf{f}^* &= (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{m}.\end{aligned}$$

References

[1] U. Ewert et al. “X-ray tomographic in-service-Testing of Circumferential Pipe Welds -The European project TomoWELD”. In: *19th World Conference on Non-Destructive Testing*. NDT.net, 2016. url: http://ndt.net/?id=19544.

[2] N. A. B. Riis et al. “Limited-data x-ray CT for underwater pipeline inspection”. *Inverse Problems* 34.3 (2018). doi: https://doi.org/10.1088/1361-6420/aaa49c.

[3] D. H. Phillips. *Welding Engineering: An Introduction*. New York: Wiley, 2016.

[4] J. Müeller and S. Siltanen. *Linear and Nonlinear Inverse Problems with Practical Applications*. Computational Science & Engineering Series. Philadelphia: SIAM, 2012.

[5] B. P. Flannery et al. “Three-dimensional X-ray microtomography”. *Science* 237.4821 (1987), pp. 1439–1444.

[6] X. Han et al. “Algorithm-enabled low-dose micro-CT imaging”. *IEEE transactions on medical imaging* 30.3 (2011), pp. 606–620. url: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3645946/.

[7] S. Siltanen et al. “Statistical inversion for medical x-ray tomography with few radiographs: I. General theory”. *Physics in Medicine and Biology* 48.10 (2003), pp. 1437–1463.

[8] J. S. Jørgensen et al. “Empirical average-case relation between undersampling and sparsity in x-ray CT”. *Inverse problems and imaging* 9.2 (2015), p. 431.

[9] G. T. Herman and A. Kuba. “Discrete tomography: A historical overview”. In: *Discrete tomography: Foundations, algorithms, and applications*. 1999.

[10] K. J. Batenburg and J. Sijbers. “DART: A Practical Reconstruction Algorithm for Discrete Tomography”. *IEEE Transactions on Image Processing* 20.9 (2011), pp. 2542–2553.

[11] F. Bleichrodt, F. Tabak, and K. J. Batenburg. “SDART: An algorithm for discrete tomography from noisy projections”. *Computer Vision and Image Understanding* 129 (2014), pp. 63–74.

[12] L. Brabant et al. “EDART, a discrete algebraic reconstructing technique for experimental data obtained with high resolution computed tomography”. *Journal of X-ray Science and Technology* 22.1 (2014), pp. 47–61.

[13] X. Zhuge, W. J. Palenstijn, and K. J. Batenburg. “TVR-DART: A More Robust Algorithm for Discrete Tomography From Limited Projection Data With Automated Gray Value Estimation”. *IEEE Transactions on Image Processing* 25.1 (2016), pp. 455–468.

[14] E. Gouillart et al. “Belief-propagation reconstruction for discrete tomography”. *Inverse Problems* 29.3 (2013).

[15] S.-i. Maeda et al. “Maximum a posteriori X-ray computed tomography using graph cuts”. In: *Journal of Physics: Conference Series*. Vol. 233. 1. IOP Publishing. 2010, p. 012023.

[16] L. Wang et al. *Computed tomography reconstruction based on a hierarchical model and variational Bayesian method*. IEEE, 2016.

[17] A. Tuysuzoglu et al. “Graph-cut based discrete-valued image reconstruction”. *IEEE Transactions on Image Processing* 24.5 (2015), pp. 1614–1627.

[18] J. H. Kappes et al. “TomoGC: binary tomography by constrained graphcuts”. In: *German Conference on Pattern Recognition*. Springer. 2015, pp. 262–273.

[19] K. J. Batenburg. “A network flow algorithm for reconstructing binary images from continuous X-rays”. *Journal of Mathematical Imaging and Vision* 30.3 (2008), pp. 231–248.

[20] S. Weber et al. “A benchmark evaluation of large-scale optimization approaches to binary tomography”. In: *Discrete Geometry for Computer Imagery*. Springer. 2006, pp. 146–156.

[21] L. Varga, P. Balázs, and A. Nagy. “An energy minimization reconstruction algorithm for multivalued discrete tomography”. In: *Computational Modelling of Objects Represented in Images III. Fundamentals, Methods, and Applications*. Ed. by P. Di Giamberardino and D. Iacoviello. CRC Press, 2012.

[22] M. Storath et al. “Joint image reconstruction and segmentation using the Potts model”. *Inverse Problems* 31.2 (2015), p. 025003.

[23] M. Zisler et al. “Discrete Tomography by Continuous Multilabeling Subject to Projection Constraints.” In: *German Conference on Pattern Recognition*. 2016, pp. 261–272.

[24] A. Stefanoiu et al. “Joint Segmentation and Shape Regularization With a Generalized Forward–Backward Algorithm”. *IEEE Transactions on Image Processing* 25.7 (2016), pp. 3384–3394.

[25] A. Tuysuzoglu, Y. Khoo, and W. C. Karl. “Variable splitting techniques for discrete tomography”. In: *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE. 2016, pp. 1764–1768.

[26] S. Osher and L. I. Rudin. “Feature-oriented image enhancement using shock filters”. *SIAM Journal on Numerical Analysis* 27.4 (1990), pp. 919–940.

[27] L. I. Rudin, S. Osher, and E. Fatemi. “Nonlinear total variation based noise removal algorithms”. *Physica D: Nonlinear Phenomena* 60.1-4 (1992), pp. 259–268.

[28] E. Demircan-Tureyen and M. E. Kamasak. “A compressed sensing based approach on Discrete Algebraic Reconstruction Technique”. In: *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*. IEEE. 2015, pp. 7494–7497.

[29] E. Demircan-Tureyen and M. E. Kamasak. “A discretized tomographic image reconstruction based upon total variation regularization”. *Biomedical Signal Processing and Control* 38 (2017), pp. 44–54.

[30] *MATLAB*. Version 9.1.0 (R2016b). Natick, Massachusetts: The MathWorks Inc., 2016.

[31] W. van Aarle et al. “Fast and Flexible X-ray Tomography Using the ASTRA Toolbox”. *The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography* 157 (2015), pp. 35–47. doi: `http://dx.doi.org/10.1016/j.ultramic.2015.05.002`.

[32] W. van Aarle et al. “Fast and Flexible X-ray Tomography Using the ASTRA Toolbox”. *Optics Express* 24.22 (2016), pp. 25129–25147. doi: `http://dx.doi.org/10.1364/OE.24.025129`.

[33] W. J. Palenstijn, K. Batenburg, and J. Sijbers. “Performance improvements for iterative electron tomography reconstruction using graphics processing units (GPUs)”. *Journal of structural biology* 176.2 (2011), pp. 250–253.

[34] E. van den Berg and M. P. Friedlander. *Spot Toolbox*. Version 1.2. 2013. url: `http://www.cs.ubc.ca/labs/scl/spot/`.

[35] M. Jovanović, L. Kosec, and B. Zorc. “Examination of weld defects by Computed Tomography”. *Metalurgija* 51.2 (2012).

[36] B. Venkatraman, M. A. Raj, and V. Vaithiyanathan. “Weld Defect Detection using Iterative Image Reconstruction Methods”. *Indian Journal of Science and Technology* 6.4 (2013), pp. 4378–4383.

[37] B. Redmer et al. “Tomographic 3D-Radiometry for the visualisation and measurement of the defects of Girth Seams”. In: *9th European Conference on NDT, Proceedings of*. Vol. 25. 2006, p. 29.

[38] U. Ewert et al. “X-ray tomographic in-service inspection of girth welds-The European project TomoWELD”. In: *AIP Conference Proceedings*. Vol. 1650. 1. AIP. 2015, pp. 525–533.

[39] *The TomoWELD project*. 2018. url: `http://www.tomoweld.eu/home/`.

[40] V. L. Vengrinovich et al. “Bayesian restoration of crack images in welds from incomplete and noisy X-ray data”. In: *AIP Conference Proceedings*. Vol. 509. 1. AIP. 2000, pp. 635–642.

[41] M. I. Haith, P. Huthwaite, and M. J. S. Lowe. “Defect characterisation from limited view pipeline radiography”. *NDT&E International* 86 (2017), pp. 186–198. doi: `http://dx.doi.org/10.1016/j.ndteint.2016.12.007`.

[42] W. V. Aarle. “Tomographic segmentation and discrete tomography for quantitative analysis of transmission tomography data”. PhD thesis. University of Antwerpen, 2012.

[43] A. C. Kak and M. Slaney. *Principles of computerized tomographic imaging*. Philadelphia: SIAM, 2001. A republication of the work first published by IEEE Press, New York, 1988.

[44] A. Meaney. “Design and Construction of an X-ray Computed Tomography Imaging System”. Master’s thesis. University of Helsinki, 2015.

[45] L. Feldkamp, L. Davis, and J. Kress. “Practical cone-beam algorithm”. *JOSA A* 1.6 (1984), pp. 612–619.

[46] P. C. Hansen. *Discrete Inverse Problems. Insight and Algorithms*. Fundamentals of Algorithms Series. Philadelphia: SIAM, 2010.

[47] G. H. Golub, P. C. Hansen, and D. P. O’Leary. “Tikhonov regularization and total least squares”. *SIAM Journal on Matrix Analysis and Applications* 21.1 (1999), pp. 185–194.

[48] F. Xu and K. Mueller. “A comparative study of popular interpolation and integration methods for use in computed tomography”. In: *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*. IEEE. 2006, pp. 1252–1255.

[49] L. C. B. dos Santos, E. Gouillart, and H. Talbot. “Combining interior tomography reconstruction and spatial regularization”. In: *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE. 2014, pp. 1768–1772.

[50] E. T. Quinto. “Exterior and limited-angle tomography in non-destructive evaluation”. *Inverse Problems* 14.2 (1998), pp. 339–353.

[51] L. Rudin. “Images, numerical analysis of singularities and shock filters”. PhD thesis. California Institute of Technology, 1987. CalTech C.S. Dept. Report TR:5250:87.

[52] A. Chambolle et al. “An introduction to total variation for image analysis”. *Theoretical foundations and numerical methods for sparse recovery* 9.263-340 (2010), p. 227.

[53] A. N. Kolmogorov and S. V. Fomin. *Introductory Real Analysis*. Trans. by R. A. Silverman. New York: Dover, 1975. Unabridged and corrected republication of the work published by Prentice-Hall, Eaglewood Cliffs N.J., 1970.

[54] A. Chambolle. “An Algorithm for Total Variation Minimization and Applications”. *Journal of Mathematical Imaging and Vision* 20 (2004), pp. 89–97.

[55] E. Giusti. *Minimal Surfaces and Functions of Bounded Variation*. Monographs in Mathematics. Boston, Basel, Stuttgart: Birkhäuser, 1984.

[56] A. Beck and M. Teboulle. “Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems”. *IEEE Transactions on Image Processing* 18.11 (2009), pp. 2419–2434.

[57] R. Choksi, Y. van Gennip, and A. Oberman. “Anisotropic Total Variation Regularized L_1-Approximation and Denoising/Deblurring of 2D Bar Codes”. *preprint* (2010). arXiv:1007.1035. Previous version published in Inverse Problems and Imaging 5.3 (2010), pp. 591-617.

[58] Y. Lou et al. “A weighted difference of anisotropic and isotropic total variation model for image processing”. *SIAM Journal on Imaging Sciences* 8.3 (2015), pp. 1798–1823.

[59] E. Y. Sidky and X. Pan. “Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization”. *Physics in medicine and biology* 53.17 (2008), p. 4777.

[60] A. Chambolle and T. Pock. “A first-order primal-dual algorithm for convex problems with applications to imaging”. *Journal of mathematical imaging and vision* 40.1 (2011), pp. 120–145.

[61] K. Hämäläinen et al. “Total variation regularization for large-scale X-ray tomography”. *International Journal of Tomography & Simulation* 25.1 (2014), pp. 1–25.

[62] J. Barzilai and J. M. Borwein. “Two-point step size gradient methods”. *IMA Journal of Numerical Analysis* 8.1 (1988), pp. 141–148.

[63] C. Li. “An efficient algorithm for total variation regularization with applications to the single pixel camera and compressive sensing”. Master’s thesis. Rice University, 2009.

[64] B. Goris et al. “Advanced reconstruction algorithms for electron tomography: from comparison to combination”. *Ultramicroscopy* 127 (2013), pp. 40–47.

[65] W. van Aarle, K. J. Batenburg, and J. Sijbers. “Automatic Parameter Estimation of the Discrete Algebraic Reconstruction Technique (DART)”. *IEEE Transactions on Image Processing* 21.11 (2012), pp. 4608–4621.

[66] J. C. A. Barata and M. S. Hussein. “The Moore-Penrose Pseudoinverse. A Tutorial Review of the Theory” (2011). arxiv:1110.6882. url: `https://arxiv.org/abs/1110.6882`. ArXiv preprint.