Characterizing and reducing the influence of model errors in inverse problems

Riis, Nicolai Andre Brogaard

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Ph.D. Thesis
Doctor of Philosophy

DTU Compute
Department of Applied Mathematics and Computer Science

Characterizing and reducing the influence of model errors in inverse problems

Nicolai André Brogaard Riis

Kongens Lyngby 2021

DTU Compute
Department of Applied Mathematics and Computer Science
Technical University of Denmark

Matematiktorvet
Building 303B
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk

Summary (English)

In natural sciences, one uses mathematics to model the interaction between things in the real world. However, no model is fully correct, as it is always an approximation of the actual physics. In most cases this *model error* is negligible by design and therefore does not negatively influence the application at hand. However, in other cases it may not be feasible to construct a model with negligible model error. One example of this is inverse problems, where the goal is to *infer* something about a system given an indirect measurement of said system. Inverse problems are well-known to utilize simplified mathematical models to model complex large-scale physical systems for this reason.

In this thesis, we study methods for characterizing and reducing model errors when solving inverse problems. The driving application for this work is the inverse problem of Computed Tomography (CT), where the goal is to reconstruct the interior structure of an object from measurements of X-ray attenuation from different view angles around the object. The model error we consider is caused by uncertainty in the actual view angles for which the data is acquired. We view this application as a general linear inverse problem with the addition of uncertain model parameters and review existing research for this type of problem.

The main contribution of the thesis is the development of a new framework in the form of models and algorithms for handling CT with uncertain view angles. The work can be condensed into two major components: 1) a new method for approximate marginalization of view angle uncertainty and 2) a new method for estimation of the view angles – including uncertainty quantification of the estimate. Ultimately both components are combined to achieve estimation of the view angles as well as marginalization of any remaining uncertainty – which in our simulated experiments completely alleviate the issue with uncertainty in the view angles. Both components are also general enough that they can be applied to other similar inverse problems.

A huge part of making models and algorithms relevant to practical inverse problems is computational efficiency. In this thesis, computational efficiency has been considered throughout and we demonstrate our framework on both 2D and 3D CT problems with more than 10^6 unknowns and 10^6 data points.

Summary (Danish)

I naturvidenskab bruger man matematik til at modellere interaktionen mellem ting i den virkelige verden. Ingen model er dog helt korrekt, da den altid er en tilnærmelse af den faktiske fysik. I de fleste tilfælde en model designed på en sådan måde at *model-fejlen* er ubetydelig og påvirker derfor ikke det anvendte problem. I andre tilfælde er det imidlertid ikke muligt at konstruere en model hvor model-fejlen er ubetydelig. Et eksempel på dette er inverse problemer, hvor målet er at *inferer* viden om et system givet en såkaldt indirekte måling af systemet. Inverse problemer er velkendt for at anvende simple matematiske modeller til store komplekse fysiske systemer af denne grund.

I denne afhandling studerer vi metoder til karakterisering og reduktion af model-fejl i inverse problemer. Den drivende applikation til dette arbejde er computertomografi (CT), hvor målet er at rekonstruere den indvendige struktur af et objekt ud fra målinger af røntgendæmpning fra forskellige synsvinkler omkring objektet. Den model-fejl vi betragter skyldes usikkerhed i de faktiske synsvinkler som objektet er målt ved. Vi undersøger problemet som et generelt lineært inverst problem med usikre model-parametre og gennemgår eksisterende forskning for denne type problemer.

Hovedbidraget i denne afhandling er en ny metode – i form af modeller og algoritmer – til håndtering af CT med usikre synsvinkler. Arbejdet kan kondenseres i to hovedkomponenter: 1) en ny metode til marginalisering af synsvinkelusikkerheden og 2) en ny metode til estimering af synsvinklerne - inklusive usikkerhedskvantificering af estimatet. I sidste ende kombineres begge komponenter for at opnå både estimering af synsvinklerne og marginalisering af eventuel resterende usikkerhed - hvilket i vores simulerede eksperimenter fuldstændigt afhjælper problemet med usikkerhed i synsvinklerne. Begge komponenter er også generelle nok til, at de kan anvendes til andre lignende inverse problemer.

En stor del af arbejdet har haft fokus på beregningseffektivitet og vi demonstrerer vores metode for både 2D- og 3D CT-problemer med mere end 10^6 ukendte og 10^6 datapunkter.

Preface

This PhD thesis was prepared at the department of Applied Mathematics and Computer Science at the Technical University of Denmark in fulfilment of the requirements for acquiring a PhD degree at the PhD School of the university. The research period was between September 1st 2017 and February 5th 2021 at the Section for Scientific Computing under principal supervisor Associate Professor Yiqiu Dong and co-supervisor Professor Per Christian Hansen both from the department. The project was also co-supervised by Professor Jürgen Frikel from OTH Regensburg, Germany. Part of the research was carried out during two short externals stays visiting Professor Jürgen Frikel at OTH Regensburg and Professor Johnathan Bardsley at University of Montana, USA. The majority of the project is funded by an internal scholarship from the university with 3 months co-financed salary from Intereg Europe under the MAX4ESSFUN sub-project.

The thesis covers research performed during the PhD programme related to characterizing and reducing the influence of model errors in inverse problems. The aim is to summarize and present the work of three of the papers written during the PhD programme, as well as provide an overview of the field as a whole. The three papers are all concerned with Computed Tomography with uncertain view angles and the thesis therefore considers the research through this lens. Most of the other papers are referenced and discussed briefly in the thesis, but are not included directly.

Kongens Lyngby, February 5, 2021

Nicolai André Brogaard Riis

Acknowledgements

I would like to thank my supervisors Yiqiu Dong and Per Christian Hansen for their tremendous support – both scientifically and on a personal level – throughout both highs and lows of this project. I thank Jürgen Frikel for giving me a large part of his time during the external stay and help shape the direction of my research early on in the project. I also thank Johnathan Bardsley for an inspiring external stay in Montana, which helped my development on a scientific but also personal level. Finally, I thank the PhD School for their flexibility and openness during the entire programme. This made many of the official tasks of the PhD project much easier to manage.

Finally, I also thank my family for their support in the completion of this project. Without their help I would not have been able to finish the project.

List of papers

The following are lists of papers researched and written during the PhD programme. The thesis summarizes the work of Papers A, B and C with the initial chapters providing background and an overview of the field as a whole. Papers D, E and F are collaborations carried out during the PhD programme, but differ from the main area of research and are therefore not included.

Papers included in thesis

A) **N. A. B. Riis** and Y. Dong "A New Iterative Method for CT Reconstruction with Uncertain View Angles." In: *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer. 2019, pages 156–167.

B) **N. A. B. Riis**, Y. Dong and P. C. Hansen "Computed Tomography Reconstruction with Uncertain View Angles by Iteratively Updated Model Discrepancy." In: *Journal of Mathematical Imaging and Vision* (2020), pages 1–11.

C) **N. A. B. Riis**, Y. Dong and P. C. Hansen "Computed Tomography with View Angle Estimation using Uncertainty Quantification." In: *Inverse Problems* (2020), submitted.

Papers not included in this thesis

D) D. Caviedes-Nozal, F. M. Heuchel, J. Brunskog, **N. A. B. Riis** and E. Fernandez-Grande "A Bayesian spherical harmonics source radiation model for sound field control." In: *The Journal of the Acoustical Society of America* 146.5 (2019), pages 3425–3435.

E) D. Caviedes-Nozal, **N. A. B. Riis**, F. M. Heuchel, J. Brunskog, P. Gerstoft and E. Fernandez-Grande "Gaussian processes for sound field reconstruction." In: *The Journal of the Acoustical Society of America* (2021), to appear.

F) Y. Dong, P. C. Hansen, M. E. Hochstenbach and **N. A. B. Riis** "Fixing nonconvergence of algebraic iterative reconstruction with an unmatched backprojector." In: *SIAM Journal on Scientific Computing* 41.3 (2019), A1822–A1839.

List of symbols

Symbol	Description
$\mathbf{A} \in \mathbb{R}^{m \times n}$	Matrix representing linear forward model.
$\bar{\mathbf{A}} \in \mathbb{R}^{m \times n}$	Matrix representing the exact forward model.
$\tilde{\mathbf{A}} \in \mathbb{R}^{m \times n}$	Matrix representing the perturbed forward model.
$\mathbf{A}^T \in \mathbb{R}^{n \times m}$	Transpose of $\mathbf{A}$.
$a_{i,j}$	Element in ith row and jth column of $\mathbf{A}$.
$\mathbf{A}(\boldsymbol{\theta}) \in \mathbb{R}^{m \times n}$	Discretized CT model parametrized by view angles $\boldsymbol{\theta}$.
$\mathbf{A}(\theta_i) \in \mathbb{R}^{p \times n}$	Discretized CT model parametrized by single view angle θ_i.
$\mathbf{A}_{\boldsymbol{\theta}} \in \mathbb{R}^{m \times n}$	Discretized CT model with fixed view angle $\boldsymbol{\theta}$.
$\mathbf{a}_j(\theta_i) \in \mathbb{R}^{p \times 1}$	jth column of $\mathbf{A}(\theta_i)$.
$\mathbf{A}(\theta_i)_l \in \mathbb{R}^{1 \times n}$	lth row of $\mathbf{A}(\theta_i)$.
$\mathbf{b} \in \mathbb{R}^m$	Right-hand side of linear forward model.
$\bar{\mathbf{b}} \in \mathbb{R}^m$	True right-hand side with no noise and model error.
b_i	ith element of $\mathbf{b}$.
$\mathbf{b}_i \in \mathbb{R}^p$	ith projection of CT model.
$(\mathbf{b}_i)_l$	lth element of $\mathbf{b}_i$.
$\mathbf{C}_{\mathbf{e}} \in \mathbb{R}^{m \times m}$	Covariance of $\mathbf{e}$.
$\mathbf{C}_{\boldsymbol{\nu}\vert\mathbf{x}} \in \mathbb{R}^{m \times m}$	Covariance of $\boldsymbol{\nu}\vert\mathbf{x}$.
$\mathbf{C}_{\boldsymbol{\nu}_i\vert\mathbf{x}} \in \mathbb{R}^{p \times p}$	Covariance of $\boldsymbol{\nu}_i\vert\mathbf{x}$.
$\widetilde{\mathbf{C}}_{\boldsymbol{\nu}\vert\mathbf{x}} \in \mathbb{R}^{m \times m}$	Sample covariance of $\boldsymbol{\nu}\vert\mathbf{x}$.
$\widetilde{\mathbf{C}}_{\boldsymbol{\nu}_i\vert\mathbf{x}} \in \mathbb{R}^{p \times p}$	Sample covariance of $\boldsymbol{\nu}_i\vert\mathbf{x}$.
$\mathbf{c}_{\boldsymbol{\nu}_i\theta_i\vert\mathbf{x}} \in \mathbb{R}^{p \times 1}$	Cross-covariance of θ_i and $\boldsymbol{\nu}_i$.
$\widetilde{\mathbf{c}}_{\boldsymbol{\nu}_i\theta_i\vert\mathbf{x}} \in \mathbb{R}^{p \times 1}$	Sample cross-covariance of θ_i and $\boldsymbol{\nu}_i$.
$\mathcal{D}(\cdot)$	Data-fitting term.
$\delta(\cdot)$	Delta distribution.
$^\circ$	Angular degrees.
$\boldsymbol{\delta} \in \mathbb{R}^q$	Vector with variances for uncertain view angles $\boldsymbol{\theta}$.
$\mathrm{diag}(\boldsymbol{\delta}) \in \mathbb{R}^{q \times q}$	Covariance of view angles $\boldsymbol{\theta}$.
δ_i	ith element of $\boldsymbol{\delta}$.
$\mathbf{e} \in \mathbb{R}^m$	Measurement noise.
$\mathbf{e}_i \in \mathbb{R}^p$	Measurement noise for ith projection.
$\exp(\cdot)$	Exponential function.
e	Exponential.

Symbol	Description
$\boldsymbol{\eta} \in \mathbb{R}^m$	Model discrepancy term.
$\boldsymbol{\eta}_i \in \mathbb{R}^p$	Model discrepancy term for ith projection.
$\mathbf{E} \in \mathbb{R}^{m \times n}$	Linear model error.
f	Continuous representation of unknown object.
$\mathcal{F}_s$	1D Fourier transform.
$g_+(\cdot)$	Non-negativity penalty function.
$\Gamma(\cdot)$	Filter in Fourier domain.
I_ℓ	Final X-ray intensity through line ℓ.
I_0	Initial X-ray intensity.
$\hat{\imath}$	Imaginary number $\sqrt{-1}$.
$\mathbf{I}$	Identity matrix (size determined from context).
$\mathrm{Ker}(\cdot)$	Kernel.
$\kappa(\cdot)$	Condition number.
λ	Regularization parameter.
log	Natural logarithm.
$\mathbf{L}_{\boldsymbol{\nu}\vert\mathbf{x}} \in \mathbb{R}^{m \times m}$	Cholesky factor of inverse covariance $\mathbf{C}^{-1}_{\boldsymbol{\nu}\vert\mathbf{x}}$.
$\mathbf{L}_{\boldsymbol{\nu}_i\vert\mathbf{x}} \in \mathbb{R}^{p \times p}$	Cholesky factor of inverse covariance $\mathbf{C}^{-1}_{\boldsymbol{\nu}_i\vert\mathbf{x}}$.
$\widetilde{\mathbf{L}}_{\boldsymbol{\nu}\vert\mathbf{x}} \in \mathbb{R}^{m \times m}$	Cholesky factor of inverse sample covariance $\widetilde{\mathbf{C}}^{-1}_{\boldsymbol{\nu}\vert\mathbf{x}}$.
$\widetilde{\mathbf{L}}_{\boldsymbol{\nu}_i\vert\mathbf{x}} \in \mathbb{R}^{p \times p}$	Cholesky factor of inverse sample covariance $\widetilde{\mathbf{C}}^{-1}_{\boldsymbol{\nu}_i\vert\mathbf{x}}$.
$\ell(\omega, s)$	Line parametrized by unit vector ω and signed distance s.
$\boldsymbol{\mu}_{\boldsymbol{\theta}} \in \mathbb{R}^p$	Mean of $\boldsymbol{\theta}$.
μ_{θ_i}	Mean of θ_i.
$\boldsymbol{\mu}_{\boldsymbol{\nu}\vert\mathbf{x}} \in \mathbb{R}^m$	Mean of $\boldsymbol{\nu}\vert\mathbf{x}$.
$\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\nu}\vert\mathbf{x}} \in \mathbb{R}^m$	Sample mean of $\boldsymbol{\nu}\vert\mathbf{x}$.
∇	Gradient.
N	Image dimension (2D: $N \times N$, 3D: $N \times N \times N$).
n	Number of elements in discrete CT image.
$\boldsymbol{\nu} \in \mathbb{R}^m$	Combined model discrepancy and noise.
$\boldsymbol{\nu}_i \in \mathbb{R}^p$	Combined model discrepancy and noise for one projection.
ω	Unit vector.
$\omega^\perp$	Perpendicular unit vector.
$\pi(\cdot)$	Probability distribution.
π	Pi.
$p(\ell(\omega, s))$	Projection along line ℓ.
p	Number of detector pixels/elements.
$\mathcal{P}$	Representation of physical system.
prox_f^σ	Proximal operator of f with step size σ.
q	Number of view angles.
$\mathrm{Ran}(\cdot)$	Range.
$R(\cdot)$	Regularization term.
$\mathcal{R}$	Radon transform.
$\sigma_{\max}$	Largest singular value.
$\sigma_{\min}$	Smallest singular value.
σ^2	Variance of measurement noise.

Symbol	Description
s	Signed distance from 0.
S^1	Unit circle.
$\mathcal{S}$	Schwartz space.
S_{CT}	Number of samples of $\boldsymbol{\eta}\|\mathbf{x}$ for CT reconstruction step.
S_{VA}	Number of samples of $\boldsymbol{\eta}\|\mathbf{x}$ for view angle estimation step.
$\text{TV}(\mathbf{x})$	Total variation regularization term.
$\boldsymbol{\theta} \in \mathbb{R}^q$	View angle parameters
θ_i	ith element of $\boldsymbol{\theta}$.
$\hat{\boldsymbol{\theta}} \in \mathbb{R}^q$	Fixed (nominal) view angle parameter.
$\mathcal{U}$	Uncertainty set (Chapter 4), Uniform distr. Appendix D.1.
$\mathcal{V}$	Uncertainty set
$\mathbf{x} \in \mathbb{R}^n$	Representation of unknown object (CT image).
x_j	jth element of $\mathbf{x}$.
$\bar{\mathbf{x}} \in \mathbb{R}^n$	Ground truth (exact object).
$\mathbf{x}^* \in \mathbb{R}^n$	Naive reconstruction.
$\mathbf{x}^*_\lambda \in \mathbb{R}^n$	Regularized solution (Tikhonov).
$\bar{\mathbf{x}}^*_\lambda \in \mathbb{R}^n$	Regularized solution (Tikhonov) from noise-free data.
$\mathbf{x}_{\text{L2-TV}} \in \mathbb{R}^n$	Non-neg TV rec. with 2-norm data-fit.
$\mathbf{x}_{\text{MD-TV}} \in \mathbb{R}^n$	Non-neg TV rec. with marginalized view angle uncertainty.
$\mathcal{X}$	Representation of true object.

Contents

CHAPTER 1

Introduction

The use of mathematical models to describe the behaviour of complex physical processes is the backbone of many fields of research. One goal of scientists in these fields is to select models that capture the essence of the physics and disregards the "unimportant" aspects. A hallmark of a great scientist have long been the ability to devise simple yet effective models of real physical phenomena. In this way, these models are used to solve real world problems despite the inexact nature of the model. A famous quote by the statistician George Box *"All models are wrong, but some are useful"* [9] summarizes this principle well.

However in some practical applications, it is not possible to come up with a *perfect* model that only disregards the negligible aspects of a physical process. Instead, a mathematical model often disregards some aspects of a process that may be important to take into account. Furthermore, even physical processes that are taken into account can only be modelled with limited accuracy in practise. Going back to George Box another quote from some of his earlier works states

> *"Since all models are wrong the scientist must be alert to what is importantly wrong."*
>
> (George E P Box, 1978 [10]).

In this thesis, we are interested in exploring this concept of *model errors* in applied mathematics related to the field *inverse problems*. In particular, we aim to characterize and reduce the influence of model errors, when it is not feasible to simply improve the mathematical model to avoid the error altogether. In an inverse problem the goal is to infer information about an object of interest that can not be directly observed. An inverse problem can be thought of as the inverse of a *forward problem*, for which the goal is to predict the output of a system given some known input object. Hence, the goal of an inverse problem is to *infer* the input given a measured output from a system. This principle is illustrated in Figure 1.1. The inverse problem is solved by *reconstructing* a set of parameters describing the object from the data. It is common to refer to the reconstructed object as the solution to the inverse problem. The reconstruction is achieved by measuring an effect caused by the object and then using a mathematical model to reconstruct the object from said measurement. In practise, the system and data are both known with limited accuracy, which should be taken into account when solving the inverse problem for the best results. In Chapter 2, we give a brief overview of inverse problems.

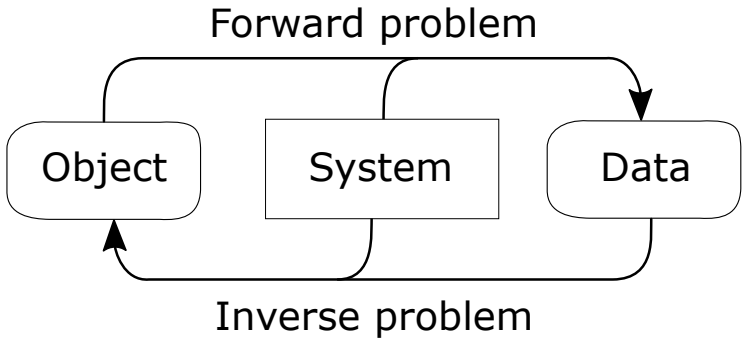


Figure 1.1: A classical illustration of forward and inverse problems. When solving an inverse problem both the data and system is known with limited accuracy in practise. This should be taken into account in the reconstruction of the object for the best results.

1.1 Model errors in inverse problems

Model errors in inverse problems is steadily becoming a popular topic of research in the inverse problems community. It has long been known that errors in the mathematical model can negatively influence the solution of an inverse problem, and there exists bounds on the effect for specific cases as we discover in Chapter 4. However, recent advances in theory, algorithms and computational efficiency have led to more research and theory focused on explicitly taking the model errors into account when solving the inverse problem. In this thesis, we explore some of these methods and discuss their applicability in practise. The main part of the thesis focuses on a specific strategy for characterizing model errors based on an additive *model-discrepancy* term and extend upon existing research in this direction. The approach was originally pioneered in [32] for Bayesian calibration of computer models and later adapted to inverse problems in [31, 46] and is still ongoing research.

1.2 Computed Tomography

There are many important applications which can be viewed as an inverse problem and where reducing the influence of model errors improves reconstruction quality. In this thesis, we select Computed Tomography (CT) as the practical application to validate and test our theory and algorithms on. In CT, the goal is to reconstruct the interior structure of an object by measuring X-ray attenuation from different angles around the object. One calls the measured X-ray attenuation from a specific angle a *projection*. The interior may be a persons chest or the crystal structure of some material for example. The object is reconstructed using a mathematical model that combines the projections for each angle into a single image reflecting the objects X-

ray attenuation at different locations in space. Because different materials present different X-ray attenuation, this provides relevant information about the interior of said object. In Chapter 3 we give an overview of CT.

In CT, model errors may arise for different reasons. In this thesis, we are interested in the important case where the *view angles* of the scan is known with limited accuracy. The parameter representing the view angle in the CT model can therefore be thought of as uncertain. In Chapters 5 and 6 we develop new models and algorithms for handling CT with uncertain view angles.

1.3 Preview of results

The research carried out in this thesis is not necessarily limited to CT alone. Despite this, our research have focused on CT as the practical test case. One nice feature of CT, is that it is easy to compare the performance of new theory and algorithms by visually inspecting the reconstructed images. In this section, we give a preview of the improvement one can obtain in the reconstructed images by using our methods. This may lead one to believe that our research is limited to CT alone, but the hope is that others can utilize approach we have developed on other interesting inverse problems as well.

In Figure 1.2, we show reconstructions for a CT test case where each view angle is assumed uncertain. To be more specific, the *nominal view angles* (equidistant angles around the object) are uncertain and the true view angles are simulated as realizations from a probability distribution with the nominal view angles as the mean. More details on the simulations are given in the later chapters of the thesis and in Appendix D.1.

In the top row of Figure 1.2, we compare CT reconstructions using the nominal view angles with the ground truth image. We show a filtered back projection reconstruction (FBP), which is a commonly used method in practise [14] and a total variation (TV) based reconstruction (L2-TV), which is a state-of-the-art method used for CT [58]. Both are defined in Chapter 3. Indeed, in this example it is clear that the model error introduced by reconstructing with the nominal view angles negatively impacts the CT image. This example provides a relevant test case for the research carried out in this thesis and we specifically try to handle this exact problem in this work.

In the bottom row of Figure 1.2, we show the best CT reconstructions obtained from our research. In papers A and B, the goal is to take the view angle uncertainty into account in the CT reconstruction, while still using the nominal view angles for reconstruction. In paper C, the goal is to also estimate the true view angles to improve reconstruction further. The right-most figure in this row shows the best possible L2-TV reconstruction using the true view angles, which is only possible because the data is simulated. The results show a steady improvement in the reconstruction quality and paper C is very close to the best possible reconstruction using the TV prior.

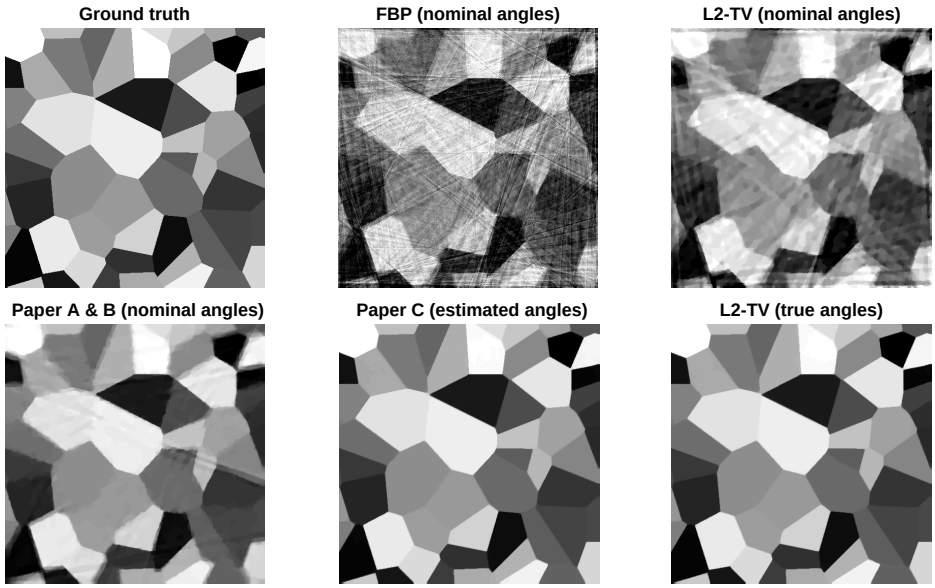


Figure 1.2: Preview of results obtained in thesis for the CT test case. The filtered back projection (FBP) and total variation based reconstruction (L2-TV) using the nominal view angles both show severe artefacts caused by view angle uncertainty. The reconstruction algorithm from paper A & B improves reconstruction quality by taking the view angle uncertainty into account. Finally, by also estimating the view angles (paper C) the reconstruction result is improved to such a degree, that it is comparable to a reconstruction obtained using the true view angles (the best-case scenario).

1.4 Contribution of thesis

In this thesis, we give an overview of existing research concerned with handling model errors in inverse problems in a general setting. The aim is to develop new theory and algorithms in this direction. The main contribution is the development of a new mathematical framework aimed at characterizing and reducing the influence of model errors caused by model parameter uncertainty in inverse problems. The framework is applied on a CT problem with uncertain view angles, but can be extended to other similar inverse problems. The work can be condensed into two major components: 1) a new method for approximate marginalization of view angle uncertainty and 2) a new method for estimation of the view angles – including uncertainty quantification of the estimate. Ultimately both components are combined to achieve estimation of

the view angles as well as marginalization of any remaining uncertainty – which in our simulated experiments completely alleviate the issue with uncertainty in the view angles. The detailed contributions can be listed as follows.

- The thesis provides an overview of the current approaches to handle model errors in inverse problems.
- A new model and algorithm is proposed to take uncertain model parameters into account when solving an inverse problem.
- This approach is able to approximately marginalize model parameter uncertainty by constructing a data-fitting term that takes the model parameter uncertainty into account.
- The marginalization is carried out alternately. That is, each time a new solution to the inverse problem is obtained, the quality of the approximate marginalization can be improved and in turn a new improved solution can be obtained.
- The approach also provides a direct way to estimate the model parameters including uncertainty quantification of the estimates, which can be included with the above-mentioned alternating procedure.
- The new framework is tested and verified on a CT problem, where the view angle parameter is assumed to be uncertain, and shows significant improvement in reconstruction quality in simulated experiments.
- The CT test case also illustrates the computational efficiency of the algorithms, which is important for large-scale inverse problems.

1.5 Structure of thesis

Chapter 2 gives an overview of inverse problems mainly focused on variational regularization, while the final section introduces the Bayesian approach to inverse problems, which is needed to derive the data-fitting term that takes the uncertainty in the model parameters into account.

Chapter 3 gives an overview of Computed Tomography. In particular, we define the CT forward model parametrized by the view angles.

Chapter 4 gives an overview of existing research related to model errors in inverse problems. The chapter also explains why a Bayesian approach was ultimately chosen to handle the CT problem and gives an overview of the approach.

Chapter 5 summarizes the research of papers A and B.

Chapter 6 summarizes the research of paper C.

Chapter 7 is a discussion of further details and comments on the work.

Chapter 8 is the concluding remarks with comments on future work.

The appendix contains Papers A, B and C as well as supplementary material related to implementation details and selected definitions.

CHAPTER 2

Inverse problems

In this chapter, we provide the reader with a brief overview of the field inverse problems. Because inverse problems is a huge area of research, we present only the relevant material as it pertains to our work. By taking a general perspective on inverse problems, we are able to acquire important insight that can be applied to the inverse problem of CT. The chapter begins with a general introduction to inverse problems, ill-posedness and regularization and it ends with an overview of the Bayesian formulation of inverse problems. The Bayesian formulation is a particularly important part of this thesis, as our research heavily utilizes this framework. We refer to the recent review paper on modern regularization theory for inverse problems [7] for a more general treatment of regularization theory and to [31, 61] for a general treatment of inverse problems in the statistical setting and [29] for the discrete (deterministic) setting.

2.1 Mathematical formulation

The term *inverse problem* describes the important category of real-world problems, where the goal is to infer information about an unobservable object from measurements of a quantity affected by said object – a so-called indirect measurement. To provide meaningful insight to an inverse problem, a mathematical model is required. For theoretical analysis it is beneficial to study inverse problems in an infinite dimensional setting. For the purpose of this thesis we only briefly consider the infinite dimensional setting and mostly focus our attention to the finite dimensional case. That is, we consider the inverse problem after discretization. Because we use CT as a test case, we assume a linear relation between the data and the object of interest, i.e. a linear forward model. This inverse problem can be formulated as

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{2.1}$$

where $\mathbf{b} \in \mathbb{R}^m$ is the data, $\mathbf{x} \in \mathbb{R}^n$ is a set of parameters that describe the object of interest and $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the matrix representing the forward model relating the parameters of the object with the data. The matrix $\mathbf{A}$ is often a discretization of a continuous model, as is the case for the CT problem (more details in Chapter 3).

In general, the matrix $\mathbf{A}$ and parametrization of the object $\mathbf{x}$ are mathematical models – i.e. approximations – of the actual physical phenomena that have led to the

observed data $\mathbf{b}$. Therefore, the choice of parametrization and forward model are particularly important to take into account when solving an inverse problem. Furthermore, because the data is measured, it is contaminated by measurement noise. For these reasons we have an approximate relation in practise, i.e.

$$\mathbf{A}\mathbf{x} \approx \mathbf{b}. \tag{2.2}$$

Despite this issue, we consider the inverse problem from the point of view of the relation (2.1) for now and assume that the forward model represented by the matrix $\mathbf{A}$ is exact, but that the data $\mathbf{b}$ can be contaminated by noise. Later in Chapter 4, we return to the issue of model errors and expand the model to include these.

The nomenclature for inverse problems is that *constructing* the data $\mathbf{b}$ from the parametrization $\mathbf{x}$ is called the *forward problem*, whereas *reconstructing* $\mathbf{x}$ from $\mathbf{b}$ is called the *inverse problem*.

A key feature of inverse problems is that while the forward problem has a unique solution, the inverse problem tend not to have one. Furthermore, the inverse problem also tend to be "unstable" with respect to noise on the measured data. To formalize this concept we introduce *ill-posedness*.

2.2 Ill-posedness

The definition of an *ill-posed* inverse problem is commonly attributed to the work of Hadamard on partial differential equations [26]. The definition stems from defining the complementary statement, namely when a problem is *well-posed*. Hadamard stated the following three conditions for a problem to be well-posed.

- **Existence:** A solution to the problem exists.
- **Uniqueness:** The solution is unique.
- **Stability:** The solution depends continuously on the data.

A problem is then said to be ill-posed, if one or more of the above conditions are *not* satisfied.

In the discretized setting, we can assume there exists some ground truth object $\bar{\mathbf{x}} \in \mathbb{R}^n$, which represents the actual (unknown) parameters of the object we aim to reconstruct. Then, the measured data $\mathbf{b}$ is a noisy version of the "clean" data $\bar{\mathbf{b}} = \mathbf{A}\bar{\mathbf{x}}$. Often the measured data is assumed to be affected by additive noise, i.e. $\mathbf{b} = \bar{\mathbf{b}} + \mathbf{e}$, where $\mathbf{e} \in \mathbb{R}^m$ is the noise term. In this case, it is clear that a solution to (2.1) exists if and only if $\mathbf{b} \in \mathrm{Ran}(\mathbf{A})$ and that the solution is unique if and only if $\mathrm{Ker}(\mathbf{A}) = \{\mathbf{0}\}$. That is, measurement noise may move $\mathbf{b}$ out of the range of $\mathbf{A}$, such that no solution exists and e.g. if the system is overdetermined there is a non-trivial null-space and the uniqueness condition fails.

While existence and uniqueness are conditions that are clearly defined, the stability condition can be more difficult to handle and may be interpreted in various ways. A common interpretation is the following.

- **Stability** (interpretation): A small change in the data should only result in a small change in the reconstructed solution.

Given the above interpretation, the stability of a linear inverse problem is related to the condition number of the forward operator $\kappa(\mathbf{A}) = \|\mathbf{A}\|_2\|\mathbf{A}^{-1}\|_2 = \sigma_{\max}/\sigma_{\min}$, where $\|\cdot\|_2$ is the matrix 2-norm defined in Appendix D.3 and $\sigma_{\min}$, $\sigma_{\max}$ are the smallest and largest singular values respectively (see [29] for more details). This is shown in the following theorem.

Theorem 1 ([29, 39])**.** *Consider the linear system* (2.1) *with perturbation* $\mathbf{e}$ *of the data such that* $\mathbf{b} = \bar{\mathbf{b}} + \mathbf{e}$*, where* $\bar{\mathbf{b}} = \mathbf{A}\bar{\mathbf{x}}$ *and let* $\mathbf{A}$ *be square and non-singular. Then for the naive solution* $\mathbf{x}^* = \mathbf{A}^{-1}\mathbf{b}$ *and the exact solution* $\bar{\mathbf{x}} = \mathbf{A}^{-1}\bar{\mathbf{b}}$ *the following bound holds*

$$\frac{\|\mathbf{x}^* - \bar{\mathbf{x}}\|_2}{\|\bar{\mathbf{x}}\|_2} \leq \kappa(\mathbf{A})\frac{\|\mathbf{e}\|_2}{\|\bar{\mathbf{b}}\|_2}. \tag{2.3}$$

Theorem 1 shows that for large condition numbers even small perturbations $\mathbf{e}$ in the data can lead to large changes in the naive solution. Hence, if the forward operator in the inverse problem has a large condition number the stability condition is not satisfied and the problem is considered ill-posed.

Remark. A similar bound holds for perturbations of the forward operator $\mathbf{A}$ (i.e. model errors), and we return to that in Chapter 4.

2.3 Regularization

Most inverse problems do not satisfy any of Hadamard's conditions in practise and are therefore ill-posed. One way to handle ill-posedness is to solve a modified version of the inverse problem that is forced to satisfy the conditions. This technique is known as regularization and the solution to the modified inverse problem is known as a regularized solution. In the following, we give a prelude to regularization and refer to [7, 29] for a more in-depth discussion on the topic.

First, consider the issue of non-existence. To handle this, one can choose to solve a least squares problem

$$\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2, \tag{2.4}$$

which is guaranteed to have at least one solution for all possible $\mathbf{A}$ and $\mathbf{b}$.

While the least squares problem (2.4) provides existence, it does not guarantee uniqueness nor stability. To ensure that the solution is both stable and unique one can add

a *regularization term* to the least squares problem. Perhaps the most simple regularization term is the 2-norm, which defines the regularized least squares problem

$$\arg\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_2^2, \tag{2.5}$$

where $\lambda > 0$ is called the regularization parameter. This formulation is also commonly referred to as Tikhonov regularization due to his early work on this problem [63]. For a fixed regularization parameter $\lambda > 0$, there exists a unique solution to (2.5), which has the closed-form expression

$$\mathbf{x}_\lambda^* = (\mathbf{A}^T\mathbf{A} + \lambda\mathbf{I})^{-1}\mathbf{A}^T\mathbf{b}. \tag{2.6}$$

Perhaps even more importantly, Tikhonov regularization can be shown to alleviate the issue of stability. In a similar way to Theorem 1, this is shown by considering the relative change in the solutions obtained with the exact and perturbed data as in the following theorem.

Theorem 2 (Hansen, 1989 [27])**.** *Consider the linear system* (2.1) *with perturbation* $\mathbf{e}$ *of the data such that* $\mathbf{b} = \bar{\mathbf{b}} + \mathbf{e}$*, where* $\bar{\mathbf{b}} = \mathbf{A}\bar{\mathbf{x}}$*. Then for the Tikhonov solution* $\mathbf{x}_\lambda^*$ *from* (2.6) *and the Tikhonov solution on the noise free data* $\bar{\mathbf{x}}_\lambda^* = (\mathbf{A}^T\mathbf{A} + \lambda\mathbf{I})^{-1}\mathbf{A}^T\bar{\mathbf{b}}$ *we have the following bound*

$$\frac{\|\mathbf{x}_\lambda^* - \bar{\mathbf{x}}_\lambda^*\|_2}{\|\bar{\mathbf{x}}_\lambda^*\|_2} \leq \kappa_\lambda(\mathbf{A})\frac{\|\mathbf{e}\|_2}{\|\mathbf{b}_\lambda\|_2}, \tag{2.7}$$

where $\mathbf{b}_\lambda = \mathbf{A}\mathbf{x}_\lambda$ *and* $\kappa_\lambda(\mathbf{A}) = \|\mathbf{A}\|_2/\lambda$.

The theorem shows that the influence of the error caused by measurement noise in the data can be handled by adding a 2-norm regularization term given an appropriate choice of the regularization parameter λ. The regularization parameter controls the degree in which the problem is regularized as indicated by dividing by λ in the bound. The trade-off is that the noise-free regularized solution $\bar{\mathbf{x}}_\lambda^*$ can stop resembling the ground truth $\bar{\mathbf{x}}$ if λ is too large. One can derive bounds for this so-called *regularization error* in specific cases, but we leave these out of the thesis and refer to [29].

Remark. This presentation was chosen to highlight key elements of inverse problems and regularization and is not meant detail a rigorous theory on the subject. For the interested reader the theory of *regularization methods* present a rigorous treatment of ill-posed inverse problems and provides clear conditions for methods to satisfy well-defined versions of the Hadamard conditions. We leave the mathematical details of regularization methods out of the thesis but refer to [7] for a recent overview. A fourth requirement for regularization methods that is not clear from the current presentation, is that the solution to the modified inverse problem should also converge to the solution of the original inverse problem, when the noise in the data goes to zero.

2.3.1 Variational regularization

Tikhonov regularization (2.5) belongs to a larger group of methods know as *variational regularization* methods. These methods have become popular for solving a wide variety of inverse problems in the last decade, in part due to their simplicity and the development of fast and efficient solvers. The key principle is to define an optimization problem with two terms: 1) a data fidelity/fitting term that measures the fit between the measured data $\mathbf{b}$ and the *forward projected* solution $\mathbf{Ax}$ and 2) a regularization term that includes the prior information about the object of interest, e.g. by favouring or penalizing certain structures in the reconstructed solution. A variational regularization method can thus be seen as solving the optimization problem

$$\min_{\mathbf{x}} \mathcal{D}(\mathbf{b}, \mathbf{Ax}) + \lambda R(\mathbf{x}), \tag{2.8}$$

where again $\lambda > 0$ is the regularization parameter that controls the balance between enforcing prior information in the regularization term R and fitting the data in data-fitting term $\mathcal{D}$. A unique solution to (2.8) is called the regularized solution. One can show that for a number of choices of $\mathcal{D}$ and R that (2.8) constitutes a regularization method see [7] and references therein.

2.3.2 Total variation regularization

For the purpose of this thesis, we are interested in the imaging problem of CT and so we are motivated to select a regularization term (prior) that reflects natural CT images. While Tikhonov regularization have been a popular method for CT [49], enforcing a small 2-norm provides limited prior knowledge of the CT image. Instead, in these types of images one is interested in piece-wise constant solutions that clearly show boundary information. For this reason, we choose a total variation (TV) regularization term. We note that the methods derived in this thesis do not directly depend on the choice of regularization term, which is also the reason we only go into brief detail here.

The method of TV regularization was originally proposed for image denoising [56] and has subsequently been heavily used for CT reconstruction – see e.g. [58]. If we consider $\mathbf{x}$ as the vectorization of a 2D or 3D image, the discrete the TV regularization term can be defined as

$$\mathrm{TV}(\mathbf{x}) = \|\nabla \mathbf{x}\|_{2,1} := \sum_{i=1}^{n} \|[\nabla \mathbf{x}]_i\|_2, \tag{2.9}$$

where $[\nabla \mathbf{x}]_i$ is the discrete gradient of $\mathbf{x}$ at the ith pixel/voxel in the image. One can select different discretization schemes and boundary conditions for the gradient depending on the application. For this thesis, we simply choose to go with a forward difference scheme with reflexive boundary conditions.

2.3.3 Enforcing non-negativity

Another aspect of CT images is that the attenuation coefficients cannot be negative. To enforce this prior, we define an indicator function

$$g_+(\cdot) = \begin{cases} 0 & \text{if } x_i \geq 0, \\ \infty & \text{if } x_i < 0, \end{cases} \tag{2.10}$$

which enforces the non-negativity by penalizing negative values.

2.3.4 Combining the priors in an variational model

By combining these two priors into a variational model, we arrive at the optimization problem

$$\min_{\mathbf{x}} \mathcal{D}(\mathbf{b}, \mathbf{A}\mathbf{x}) + \lambda \mathrm{TV}(\mathbf{x}) + g_+(\mathbf{x}). \tag{2.11}$$

Because of the non-smooth TV term and the non-negativity constraint, an iterative method is required to obtain a solution to (2.11). In this thesis, we have chosen to utilize the popular Primal Dual Hybrid Gradient (PDHG) algorithm [17] and the Stochastic version (SPDHG) [18, 21] for this purpose. We give further details on how we use these algorithms in Chapter 5 and Appendix D.2 once the data-fitting term have been derived, and refer to the referenced papers for a complete overview.

There exists many choices of data fidelity $\mathcal{D}$ depending on the application at hand. A particular set of data fidelity terms, are those that are motivated by statistical arguments. These arise when considering inverse problems in the Bayesian paradigm. In this thesis, we rely on the data-fitting term to take the model and noise error into account and this is again why the Bayesian paradigm is of central importance in our work.

2.4 Bayesian approach to inverse problems

Bayesian inverse problems – or Bayesian estimation – is the methodology of solving an inverse problem by utilizing a probabilistic point-of-view, where the parameters of the object, measurement noise and other variables are considered as random variables. Specific for the Bayesian perspective is that the randomness describes the degree of information for each variable. That is, how certain or uncertain are we of the value of a specific variable.

In its simplest form, we consider the following version of the inverse problem in the Bayesian paradigm

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{e}, \qquad \mathbf{e} \sim \pi_{\mathbf{e}}(\cdot), \quad \mathbf{x} \sim \pi_{\mathbf{x}}(\cdot), \tag{2.12}$$

where $\mathbf{b} \in \mathbb{R}^m$ is the measured data, $\mathbf{A} \in \mathbb{R}^{m \times n}$ are deterministic (fixed) forward model and $\mathbf{e}$, $\mathbf{x}$ are random vectors. Here the data are corrupted by additive noise, which is a decent approximation for CT noise in practise [14].

We are in particular interested in the probability distribution of the object of interest $\mathbf{x}$ given the measured data $\mathbf{b}$, which is the conditional probability distribution $\pi(\mathbf{x}|\mathbf{b})$. This distribution is commonly referred to as the *posterior* distribution. The central equation in Bayesian inverse problems that allow us to infer information about the posterior is Bayes' theorem which state the relation

$$\pi(\mathbf{x}|\mathbf{b}) = \frac{\pi(\mathbf{b}|\mathbf{x})\pi(\mathbf{x})}{\pi(\mathbf{b})} \propto \pi(\mathbf{b}|\mathbf{x})\pi(\mathbf{x}). \tag{2.13}$$

Here $\pi(\mathbf{b}|\mathbf{x})$ is the *likelihood*, $\pi(\mathbf{x})$ the *prior* and $\pi(\mathbf{b})$ the *evidence*. An obvious requirement of Bayes' theorem is that the evidence is greater than zero, i.e. $\pi(\mathbf{b}) > 0$. That is, the probability of observing the data given the assumed model is greater than zero. In general however, the evidence also tend to play the rule of a normalizing constant and can be ignored by considering only the proportionality relation.

Remark. This presentation focuses on the Bayesian approach, but we note that there exists similar theory and derivations by considering "combination of states of information" that is free from the theoretical difficulties associated with conditional probability distributions. For more details on this see e.g. [61].

2.4.1 Deriving the likelihood

One of the tasks in the Bayesian paradigm is to derive the likelihood according to the assumed model. In this case the likelihood is determined according to our model (2.12) by *marginalizing out* the noise term $\mathbf{e}$. That is, the noise term is considered as a *nuisance* parameter, that need not be estimated, but should be taken into account. The marginalization is achieved by integration out the noise term, i.e.

$$\pi(\mathbf{b}|\mathbf{x}) = \int_{\mathbb{R}^m} \pi(\mathbf{b}, \mathbf{e}|\mathbf{x}) \mathrm{d}\mathbf{e}, \tag{2.14}$$

$$= \int_{\mathbb{R}^m} \pi(\mathbf{b}|\mathbf{x}, \mathbf{e})\pi(\mathbf{e}|\mathbf{x}) \mathrm{d}\mathbf{e}. \tag{2.15}$$

Because $\mathbf{x}$ and $\mathbf{e}$ are specified conditioning on these yields $\pi(\mathbf{b}|\mathbf{x}, \mathbf{e}) = \delta(\mathbf{b} - \mathcal{A}(\mathbf{x}) - \mathbf{e})$, where $\delta(\cdot)$ is the delta distribution, see e.g. [31]. This simplifies the likelihood to

$$\pi(\mathbf{b}|\mathbf{x}) = \pi_{\mathbf{e}|\mathbf{x}}(\mathbf{b} - \mathbf{A}\mathbf{x}|\mathbf{x}), \tag{2.16}$$

where $\pi_{\mathbf{e}|\mathbf{x}}$ is the conditional probability distribution of the noise term $\mathbf{e}$ given $\mathbf{x}$. In particular, if one assumes $\mathbf{e}$ to be independent of $\mathbf{x}$ the relation simplifies to

$$\pi(\mathbf{b}|\mathbf{x}) = \pi_{\mathbf{e}}(\mathbf{b} - \mathbf{A}\mathbf{x}), \tag{2.17}$$

which is simply the probability distribution of $\mathbf{e}$ evaluated at $\mathbf{b} - \mathbf{A}\mathbf{x}$.

Example 1. *To clarify the presentation above, suppose the noise distribution is Gaussian, i.e.* $\mathbf{e} \sim \mathcal{N}(\boldsymbol{\mu}_\mathbf{e}, \mathbf{C}_\mathbf{e})$ *and independent of* $\mathbf{x}$*. Then the likelihood function with respect to the generic additive noise inverse problem* (2.12) *is given by*

$$\pi(\mathbf{b}|\mathbf{x}) = \pi_\mathbf{e}(\mathbf{b} - \mathbf{A}\mathbf{x}) \tag{2.18}$$

$$\propto \exp\left(-\frac{1}{2}(\mathbf{b} - \mathbf{A}\mathbf{x} - \boldsymbol{\mu}_\mathbf{e})^T \mathbf{C}_\mathbf{e}^{-1}(\mathbf{b} - \mathbf{A}\mathbf{x} - \boldsymbol{\mu}_n oise)\right), \tag{2.19}$$

$$\propto \exp\left(-\frac{1}{2}\|\mathbf{b} - \mathbf{A}\mathbf{x} - \boldsymbol{\mu}_\mathbf{e}\|^2_{\mathbf{C}_\mathbf{e}^{-1}}\right), \tag{2.20}$$

$$\propto \exp\left(-\frac{1}{2}\|\mathbf{L}_\mathbf{e}\left(\mathbf{b} - \mathbf{A}\mathbf{x} - \boldsymbol{\mu}_\mathbf{e}\right)\|_2^2\right), \tag{2.21}$$

where $\mathbf{L}_\mathbf{e}$ *is the Cholesky factor of the inverse covariance* $\mathbf{C}_\mathbf{e}^{-1} = \mathbf{L}_\mathbf{e}^T \mathbf{L}_\mathbf{e}$*.*

2.4.2 Connection to variational regularization

Once the likelihood and prior are determined, the goal in Bayesian inverse problems is to obtain samples or point estimates of the posterior distribution. In the following we describe the connection between the maximum a posteriori (MAP) estimator and variational regularization. For a more thorough walk-through on Bayesian inverse problems and sampling-based methods see e.g. [3, 31].

The MAP estimator is defined as maximizing the posterior density (or minimizing the negative logarithm), i.e.

$$\arg\max_\mathbf{x} \left(\pi(\mathbf{b}|\mathbf{x})\pi(\mathbf{x})\right) = \arg\min_\mathbf{x} \left(-\log \pi(\mathbf{b}|\mathbf{x}) - \log \pi(\mathbf{x})\right), \tag{2.22}$$

and is closely linked to variational regularization.

Comparing the variational optimization problem (2.8) and the MAP estimation in (2.22), we can interpret $-\log \pi(\mathbf{b}|\mathbf{x})$ as the data fidelity and $-\log \pi(\mathbf{x})$ as the regularization term. This allows us consider the negative log-likelihood as a data fidelity term based on the statistics of the noise model.

Remark. An important distinction to make here is that the priors defined earlier in Sections 2.3.2 and 2.3.3 are not straightforwardly extended to the Bayesian formulation. We refer to [4, 37] for recent research in this direction for both cases. While this technicality is an important one, we simply choose to consider the problem from a variational point-of-view, where the data-fitting term is defined according to a Bayesian noise model.

One can show that Tikhonov regularization is equivalent to a Bayesian MAP estimate with a Gaussian likelihood and prior. We end this section with that example.

Example 2. *In the case that both the prior and likelihood is Gaussian we have a closed-form expression for the posterior. Assume* $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_\mathbf{x}, \mathbf{C}_\mathbf{x})$ *and* $\mathbf{e} \sim \mathcal{N}(\boldsymbol{\mu}_\mathbf{e}, \mathbf{C}_\mathbf{e})$ *are mutually independent. Then the posterior is given by*

$$\begin{aligned}
\pi(\mathbf{x}|\mathbf{b}) &\propto \pi(\mathbf{b}|\mathbf{x})\pi(\mathbf{x}) && (2.23)\\
&\propto \pi_\mathbf{e}(\mathbf{b} - \mathbf{A}\mathbf{x})\pi_\mathbf{x}(\mathbf{x}) && (2.24)\\
&\propto \exp\left(-\frac{1}{2}\|\mathbf{b} - \mathbf{A}\mathbf{x} - \boldsymbol{\mu}_\mathbf{e}\|^2_{\mathbf{C}_\mathbf{e}^{-1}}\right) \exp\left(-\frac{1}{2}\|\mathbf{x} - \boldsymbol{\mu}_\mathbf{x}\|^2_{\mathbf{C}_\mathbf{x}^{-1}}\right) && (2.25)\\
&\propto \exp\left(-\frac{1}{2}\|\mathbf{L}_\mathbf{e}(\mathbf{b} - \mathbf{A}\mathbf{x} - \boldsymbol{\mu}_\mathbf{e})\|_2^2\right) \exp\left(-\frac{1}{2}\|\mathbf{L}_\mathbf{x}(\mathbf{x} - \boldsymbol{\mu}_\mathbf{x})\|_2^2\right) && (2.26)
\end{aligned}$$

In particular if $\boldsymbol{\mu}_\mathbf{e} = \boldsymbol{\mu}_\mathbf{x} = \mathbf{0}$ *and* $\mathbf{C}_\mathbf{e} = \sigma^2\mathbf{I}$ *and* $\mathbf{C}_\mathbf{x} = \alpha^2\mathbf{I}$ *we arrive at the negative log-posterior*

$$-\log \pi(\mathbf{x}|\mathbf{b}) \propto \frac{1}{2\sigma^2}\|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \frac{1}{2\alpha^2}\|\mathbf{x}\|_2^2, \tag{2.27}$$

which when defining $\lambda = \sigma^2/\alpha^2$ *and taking the minimum yields the Tikhonov regularised least squares problem*

$$\min_\mathbf{x} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_2^2. \tag{2.28}$$

2.5 Summary

In this chapter, we gave a brief overview of inverse problems and regularization. We showed that classical regularization based on 2-norm priors such as Tikhonov regularization can be thought of as as strategy to handle measurement noise in the data. We considered another prior better suited for CT, namely TV. Finally, we showed the Bayesian approach to inverse problems and how the likelihood determined according to the noise model can be used to guide the choice of data-fitting term. Finally, we showed the Tikhonov regularization is equivalent to the MAP estimate of the Bayesian model, where the prior and noise terms are both Gaussian.

CHAPTER 3

Computed Tomography

In this chapter, we give a brief overview of computed tomography (CT). The chapter begins with defining the classical line integral CT model, which is the model we use in our numerical experiments. A special focus is put on defining the model with respect to the view angle parameter, as this is assumed to be known with limited accuracy. Finally, we show that the CT model can be decomposed according to the view angle parameter, and how this can be utilized in reconstruction algorithms.

Computed Tomography is an indispensable imaging method for both clinical and industrial applications. It is the practise of *reconstructing* an interior image of an object by combining X-ray projection images from different *view angles* around the object. The projection images are obtained by illuminating the object with X-rays and measuring the X-ray "shadow" cast by the object. The projection images are then combined to form a single image – in either 2D or 3D depending on the application – that shows the attenuation of different parts of the object.

3.1 X-ray modelling

As mentioned, we are in CT concerned with reconstructing an image of the internal structure of an object from projections at different view angles. We use X-ray attenuation to model the internal structure of the object, because it varies depending on the material it passes through thus allowing the distinction of boundary information on the object. Let $f(x)$ be a function describing the *absorption coefficient* at location x. Lambert-Beer's law states the relation between the initial intensity of the X-ray I_0 and final intensity of the X-ray after having passed through the object $f(x)$ along a line ℓ and is given by

$$I_\ell = I_0 \exp\left(-\int_\ell f(x)\mathrm{d}x\right). \tag{3.1}$$

The derivation of Lambert-Beer's law is shown in various places such as [14].

Equation (3.1) is a severely simplified model for the interaction of X-rays with matter, since it only considers direct absorption and ignores other effects such as scattering or beam-hardening. Furthermore, the model assumes that X-rays attenuate the same in matter for different wavelengths, which is well-known not to be true [14]. Despite these issues, Lambert-Beer's law is still widely used in practise. This is in part due to the fact that in CT, we are interested in reconstructing the attenuation coefficients $f(x)$ by *inverting* the model for X-ray attenuation, which very quickly becomes computationally unfeasible if the CT forward model is a complicated physical model.

3.2 Continuous model

This section is based upon previous work of the author in [54] and the definitions and theorems are directly quoted from there.

From Lambert-Beer's law, we formulate a mathematical model for CT that can be used for reconstruction. To simplify the presentation we focus on the 2D case here. Later in Chapter 7, we discuss the 3D case and give some numerical results.

First, we define the line $\ell(\omega, s)$ parametrized by the unit vector $\omega \in S^1$, where S^1 is the unit circle and by the singed distance to the origin $s \in \mathbb{R}$ as

$$\ell(\omega, s) = \left\{ \omega s + \omega^{\perp} t \,|\, t \in \mathbb{R} \right\}. \tag{3.2}$$

It is sometimes useful to parametrize ω by the polar angle $\theta \in \mathbb{R}$ such that $\omega(\theta) = (\cos(\theta), \sin(\theta))$ is the unit-vector at angle θ and $\omega^{\perp}(\theta) = \omega(\theta + \pi/2)$ is the perpendicular unit vector at angle θ.

The X-ray attenuation along this line is then modelled by Lambert-Beer's law

$$I(\omega, s) = I_0 \exp\left(- \int_{\ell(\omega,s)} f(t) \mathrm{d}t \right). \tag{3.3}$$

By rearranging terms and taking the negative logarithm we arrive at

$$p(\ell(\omega, s)) := -\log\left(\frac{I(\omega, s)}{I_0} \right) = \int_{\ell(\omega,s)} f(t)\mathrm{d}t = \int_{\mathbb{R}} f(\omega s + \omega^{\perp} t)\mathrm{d}t, \tag{3.4}$$

where $p(\ell(\omega, s))$ is the so-called projection along the line $\ell(\theta, s)$. This model for X-ray attenuation is illustrated in Figure 3.1.

For the remainder of the thesis, we use the line integral model in (3.4) unless stated otherwise. Note that this is simply a choice of convenience as the research carried out in the thesis can be directly applied to other commonly used CT models in both 2D and 3D. Again we discuss the 3D case in Chapter 7.

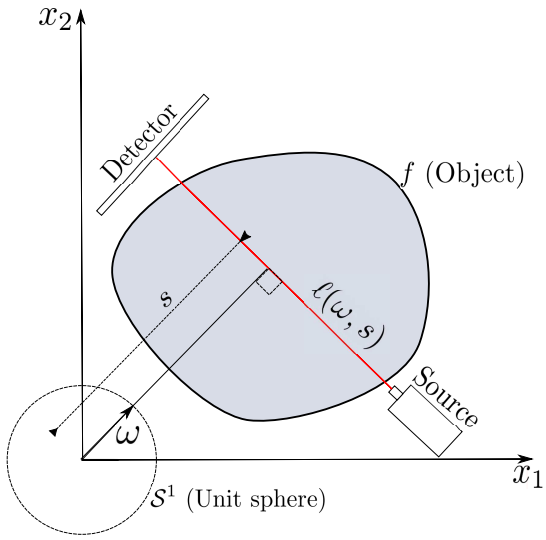


Figure 3.1: Illustration of CT line model in (3.4). This image is a modified version of an illustration by the author in previous work [54].

3.2.1 Direct inversion formula

Given our model (3.4), it may be prudent to ask if it is possible to obtain a reconstruction based on the continuous model. In fact, determining a function f from line integrals (3.4) was considered in 1917 by Johan Radon unrelated to CT. His work was later published in English in 1986 [51]. Following a more recent notation from [36], we define the *Radon transform* as the forward operator that constructs the projection data for all possible directions given the object f as follows.

Definition 1. *The 2D Radon transform $\mathcal{R}f : S^1 \times \mathbb{R} \to \mathbb{R}$ of a function $f \in \mathcal{S}(\mathbb{R}^2)$ is defined by*

$$(\mathcal{R}f)(\omega, s) := \int_{\ell(\omega,s)} f(x)\mathrm{d}x = \int_{\mathbb{R}} f(\omega s + \omega^{\perp} t)\mathrm{d}t. \tag{3.5}$$

Here $\mathcal{S}(\mathbb{R}^2)$ is the Schwartz-space of $\mathbb{R}^2$ described in Appendix D.4, Definition 6.

To arrive at the inversion theorem, we first define the *back projection* operator.

Definition 2. *Given a function $g \in L^1(S^1 \times \mathbb{R})$, we define the back projection, $\mathcal{R}^*$, as*

$$(\mathcal{R}^* g)(x) = \int_{S^1} g(\omega, x \cdot \omega)\mathrm{d}\omega. \tag{3.6}$$

Given the back projector one can show the following inversion theorem.

Theorem 3. *For $f \in \mathcal{S}(\mathbb{R}^2)$ and the 2D Radon transform, $\mathcal{R}f(\omega, s)$, $\omega \in S^1$, $s \in \mathbb{R}$, we have the inversion formula*

$$f(x) = (\mathcal{R}^{-1}\mathcal{R}f)(x) = (\mathcal{R}^*\Lambda_s\mathcal{R}f)(x) \tag{3.7}$$

$$= \frac{1}{2}(2\pi)^{-3/2}\int_{S^1}\int_{-\infty}^{\infty}\mathcal{F}_s\mathcal{R}f(\omega,\sigma)e^{\hat{i}\sigma\langle x,\omega\rangle}|\sigma|\mathrm{d}\sigma\mathrm{d}\omega, \tag{3.8}$$

where $\hat{i} = \sqrt{-1}$ and $\Lambda_s = \mathcal{F}_s^{-1}|\cdot|\mathcal{F}_s$ is a filtering in Fourier domain giving the inversion formula its alias Filtered Back Projection (FBP). The theoretical justification of this inversion formula and the definition of the 1D Fourier transform, $\mathcal{F}_s$, is included in Appendix D.4

3.3 Discrete model

In practise, one measures the intensities I and I_0 using a detector with finitely many camera elements, and so the data will present as finite dimensional. Further, to visualize the CT attenuation image, we are going to need a finite set of pixels. For this reason we are motivated to construct a discrete model for CT.

Suppose we use a discretization of the object $\mathbf{x} = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^n$, where x_i describes the attenuation coefficient for each pixel in an $N \times N$ image, with $n = N^2$. The attenuation along a line is then approximated by summing the attenuation coefficients x_j weighted by the euclidean distance travelled by the X-ray beam through that pixel. Indexing each measured data point by i, we can define

$$b_i = \sum_{j=1}^{n} a_{i,j}x_j, \tag{3.9}$$

where $a_{i,j}$ is the length of the ith beam going through the object and b_i is the measured data from the beam. We can represent the above discrete model in matrix form, which is the linear equation

$$\mathbf{b} = \mathbf{A}\mathbf{x}, \tag{3.10}$$

that we are already familiar with from Chapter 2. Here the ith row and jth column of $\mathbf{A}$ is $a_{i,j}$ and $\mathbf{b} = [b_1, b_2, \ldots, b_m] \in \mathbb{R}^m$.

The CT matrix $\mathbf{A}$ is sparse. This is easily seen from (3.9) as for each ray i most distances $a_{i,j}$ are zero. This is also clear from the illustration in Figure 3.2 showing a single X-ray passing through an object with the discretization on top. In Figure 3.3, we show an example of the non-zero elements of a CT matrix derived from the line model (3.4). In large-scale CT, even storing only the non-zero elements of this CT matrix is prohibitive. For this reason the CT matrix is typically not stored in practise and multiplication with $\mathbf{A}$ or $\mathbf{A}^T$ are carried out matrix-free. We get back to this point later in Chapter 4.

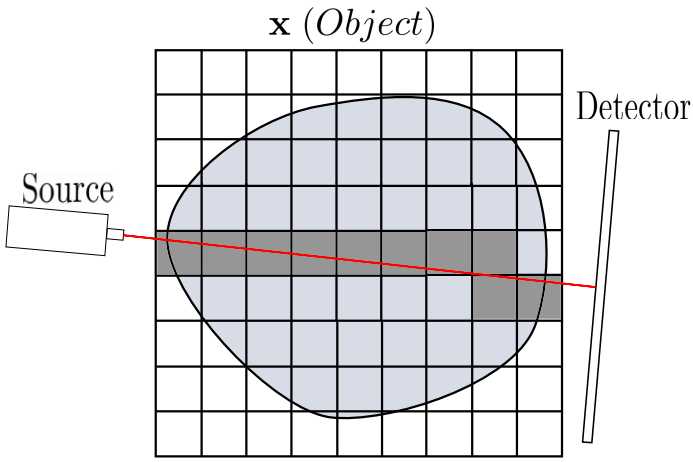


Figure 3.2: Illustration of discrete CT model. This image is an illustration by the author in previous work [54].

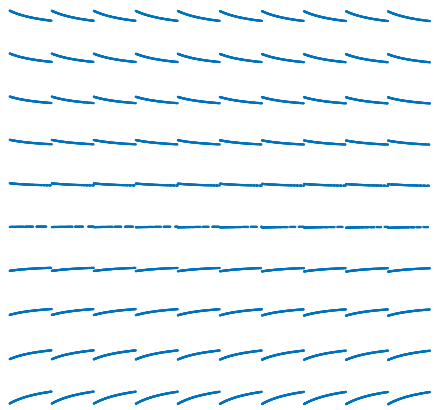


Figure 3.3: Zoom of a small part ($10N \times 10p$) of the non-zero elements of a CT matrix of generated from the line model (3.4). There are about 0.18% non-zero elements in the matrix. Even so storing all the non-zero elements for the full $n \times m$ matrix is not feasible for large-scale CT problems.

3.3.1 View angle parametrization

Because we are interested in the case where the view angle parameter is known with limited accuracy, it becomes useful to explicitly define the CT forward model parametrized by the view angles. To achieve this, we define the vector $\mathbf{b}_i \in \mathbb{R}^p$ representing the ith projection at view angle θ_i for $i = 1, \dots, q$.

That is, the measured data is assumed to come from the CT forward model

$$(\mathbf{b}_i)_l = \mathbf{A}(\theta_i)_l \mathbf{x} \approx \int_{\mathbb{R}} f(\omega(\theta_i) s_l + \omega(\theta_i)^{\perp} t) \mathrm{d}t \tag{3.11}$$

where $\mathbf{A}(\theta_i) \in \mathbb{R}^{p \times n}$ is the CT forward matrix for only the view angle θ_i and $\mathbf{A}(\theta_i)_l \in \mathbb{R}^{1 \times n}$ is the lth row vector of this matrix. Similar to earlier, the elements of the CT matrix is given by the euclidean distance travelled through each pixel. In the integral s_l is the signed distance of the lth detector element.

A single projection at view angle θ_i is thus defined as

$$\mathbf{b}_i = \mathbf{A}(\theta_i)\, \mathbf{x}, \tag{3.12}$$

and the entire CT data set $\mathbf{b}$ is given by

$$\mathbf{b} = \mathbf{A}(\boldsymbol{\theta})\, \mathbf{x}. \tag{3.13}$$

Hence, the full CT forward matrix is decomposed into q projection matrices

$$\mathbf{A}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{A}(\theta_1) \\ \mathbf{A}(\theta_2) \\ \vdots \\ \mathbf{A}(\theta_q) \end{bmatrix}, \tag{3.14}$$

where each projection matrix consists of the row vectors defined in (3.11), i.e.

$$\mathbf{A}(\theta_i) = \begin{bmatrix} \text{——} \mathbf{A}(\theta_i)_1 \text{——} \\ \text{——} \mathbf{A}(\theta_i)_2 \text{——} \\ \vdots \\ \text{——} \mathbf{A}(\theta_i)_p \text{——} \end{bmatrix}. \tag{3.15}$$

3.4 Reconstruction methods

We now aim to obtain a CT image reconstruction given the measured data $\mathbf{b}$ and the CT forward model. First thing to note is that the measured data $\mathbf{b}$ is affected by measurement noise. For low-dose CT the measurement noise is best approximated by a Poisson distribution, due to the counting nature of the detector. However, for larger dose CT the noise is well approximated by additive Gaussian noise and is often used in practise [14].

3.4.1 Filtered back projection

Based on Theorem 9, we can derive the so-called *filtered back projection* (FBP) reconstruction. In the discrete setting, the back projector $\mathcal{R}^*$ is approximated by the transpose of the forward matrix, i.e. $\mathbf{A}^T$ yielding the reconstruction formula

$$\mathbf{x} = \frac{1}{2}(2\pi)^{-3/2}\mathbf{A}^T\Lambda_\Gamma\mathbf{b} \tag{3.16}$$

where Λ_Γ is a filter in Fourier domain, i.e. $\Lambda_\Gamma = \mathcal{F}^{-1}\Gamma(\cdot)\mathcal{F}$, where $\mathcal{F}$ is the Fourier transform defined earlier and $\Gamma(\cdot) = |\cdot|$. This provides an interesting perspective on the effect of the measurement noise. If there is high frequency noise on the measured data, then the filtering will amplify that noise and possibly destroy the reconstruction. This is a well known phenomenon and for this reason a low-pass filter for $\Gamma(\cdot)$ is often used in practise. We again refer to [14] for more details.

The FBP reconstruction works well when the CT data is complete, i.e. when the object is densely illuminated from all scanning angles by X-rays. However, when the data is limited e.g. from sparse angular sampling, the FBP reconstruction does not perform well. This is illustrated in Figure 3.4, where the reconstructions are from obtained from data with angular sampling is every 4° (see Appendix D.1 for details on the simulations). It is clear that certain line artefacts appear in the FBP reconstruction compared to the ground truth, which is also shown in the figure. This phenomenon can be explained by microlocal analysis and we refer to [36, 54] for more details. Note here there is no uncertainty in the view angle parameter because the reconstructions are obtained with the true view angles.

3.4.2 Variational reconstruction based on total variation

Since the CT model fits directly into the linear inverse problem from Chapter 2, we can use the TV-based variational method defined in Section 2.3.2 as well as the non-negativity prior from Section 2.3.3 to define a reconstruction method. The data-fitting term is chosen as a 2-norm due to the Gaussian measurement noise, defined as $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. The variational TV-based CT reconstruction can thus be written as

$$\mathbf{x}_{\text{L2-TV}} = \arg\min_{\mathbf{x}} \frac{1}{2\sigma^2}\|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\text{TV}(\mathbf{x}) + g_+(\mathbf{x}). \tag{3.17}$$

Using the view angle parametrization we can also write the above optimization problem with q data fitting terms (one for each projection) as

$$\mathbf{x}_{\text{L2-TV}} = \arg\min_{\mathbf{x}} \frac{1}{2\sigma^2}\sum_{i=1}^{q}\|\mathbf{b}_i - \mathbf{A}(\theta_i)\mathbf{x}\|_2^2 + \lambda\text{TV}(\mathbf{x}) + g_+(\mathbf{x}). \tag{3.18}$$

This formulation lends itself well to be solved using the recently proposed SPDHG algorithm [18, 21] already eluded to in Chapter 2. Again, we give a few details on

how we utilize this algorithm in Appendix D.2 and otherwise refer to Paper A and B.

The L2-TV reconstruction obtained for a specific regularization parameter is also shown in Figure 3.4. Indeed the TV-based reconstruction performs better than FBP for this test case and the line artefacts are removed. It is a well-known but interesting phenomenon that the TV prior can suppress the line artefacts in the reconstruction.

3.5 Reconstructions with uncertain view angles

In this section, we give a prelude to the issue with only knowing the view angle parameter with limited accuracy. Indeed, assume that we obtain reconstructions using the so-called nominal view angles (equidistant around the object), instead of the true view angles – again see Appendix D.1 for details on the simulation.

In the bottom row of Figure 3.4, we see the reconstructions for the FBP and L2-TV reconstruction methods based on the nominal view angles. The FBP reconstruction looks more "fuzzy" compared to the one from true view angles, indicating the uncertainty in the nominal view angles impacts the quality of the CT reconstruction. Interestingly, the L2-TV reconstruction is no longer able to suppress the line artefacts from the sparse angular sampling. In this case, the regularization parameter was chosen as the one with the lowest relative error $\|\mathbf{x} - \bar{\mathbf{x}}\|_2 / \|\bar{\mathbf{x}}\|_2$, where $\mathbf{x}$ is the reconstruction. In Chapters 5 and 6 we return to this issue.

Remark. We note that it is still possible to suppress the line artefacts in the L2-TV reconstruction by significantly increasing the regularization parameter at the cost of significant loss of image detail.

3.6 Summary

In this chapter, we provided a brief overview of CT focused on the case where the CT forward model was explicitly parametrized by the view angles. We provided two methods of obtaining reconstructions from CT data, one based on the continuous formulation on CT and one based on a variational formulation. We showed that both existing methods for CT reconstruction perform poorly if the view angle parameter is uncertain.

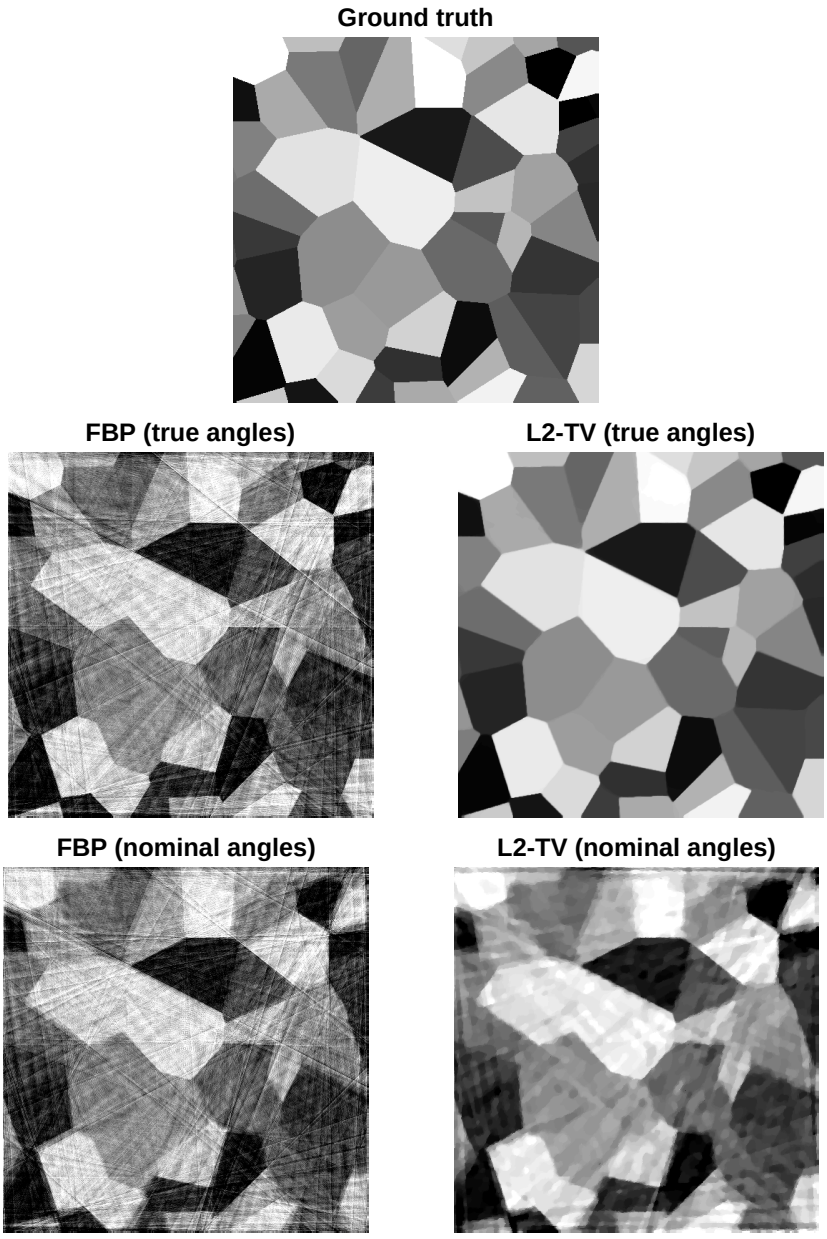


Figure 3.4: CT reconstructions for the simulated test case in Appendix D.1.

CHAPTER 4

Handling model errors in inverse problems

In this chapter, we survey some of the existing methods for handling model errors in inverse problems as they pertain to the work carried out in this thesis. We split the presentation into two parts. The first part is concerned with deterministic methods that characterize the model error by a norm-bounded perturbation of the forward model. The second part is concerned with a Bayesian method, where the effect of the model error is characterized by a model discrepancy term on the data side. The chapter contains an outlook and discussion on which methods are best suited for the CT application.

In Chapters 2 and 3, it was assumed that the forward model in the inverse problem was exact. However, all models are wrong. This is particularly true in inverse problems, where there is a tendency to use precise mathematical models that can be analysed and inverted in the continuous domain – such as the Radon transform (Definition 1, Chapter 3) in the case of CT.

Using a mathematical model to describe a complex physical phenomena presents a discrepancy between the predicted value of the model and the measured data, that can not only be accounted for by measurement noise. When handling such *model errors*, the goal is not to make the existing mathematical model more accurate by including more complex interactions, but rather to acknowledge that the model error exists and find a way to take it into account when solving the inverse problem.

In this thesis, it will prove useful to characterize the model error by its influence on the residual – i.e. the difference between the measurement data and the predicted forward model. Suppose we have a linear model $\mathbf{A}$ as in (2.1), relating the measured data $\mathbf{b}$ with an parametrization $\mathbf{x}$ of the object such that

$$\mathbf{b} \approx \mathbf{A}\mathbf{x}, \tag{4.1}$$

where the symbol $\approx$ is simply used to state that the model is wrong, but approximates the process that generated the data. We can then define the (noisy) residual as

$$\boldsymbol{\nu} := \mathbf{b} - \mathbf{A}\mathbf{x}. \tag{4.2}$$

We assume that the data is corrupted by additive measurement noise, such that a part of the residual comes from said noise. For this reason, we split the residual into two parts

$$\boldsymbol{\nu} = \boldsymbol{\eta} + \mathbf{e}, \tag{4.3}$$

where $\mathbf{e}$ is the measurement noise which is independent of the system and $\boldsymbol{\eta}$ contains the model errors influence on the residual, which we call the *model discrepancy term*. This way of characterizing model errors have been considered before, of which [32] is the earliest work we are aware of. The approach was also later adapted for Bayesian inverse problems in [31] and applied on a number of practical applications [2, 11, 31, 33, 46, 47, 50, 62].

The model discrepancy $\boldsymbol{\eta}$ may arise from a number of different reasons. Using similar terminology as in [32], we distinguish between the following three.

- **Model inadequacy.** Since the mathematical model only models a part of the physical system that generated the data, it is possible – even for data with no measurement noise – that the predicted value of the model will not equal the true value of the physical process for any input. A simple example of this is a finite-dimensional linear model describing a non-linear process.
- **Parameter uncertainty.** The mathematical models do not only depend on the solution to the inverse problem. They also depend on some parameters that may be directly or indirectly related to a physical parameter. In practise, these may only be known with a limited accuracy, but are often assumed to be known exactly, which introduces a model error from said parameter uncertainty.
- **Code error.** Finally, when implementing the forward model on a computer, the implemented code may differ depending on the implementation. We call this *code error*. This could for example be discretization error, but certain choices may also be taken in the implementation that make it such that the code output would not converge to the output of the model even for fine discretization.

Example 3. *Consider the CT case, where the forward model approximates the interaction of X-rays with matter by a line integral model* (3.4) *parametrized by an angle and distance parameter. First, in practise the angle and distance parameters may only be known with a limited accuracy. This introduces a parameter uncertainty in the model. Second, the line integral model ignores physical effects such as scattering, refraction etc. in CT, which introduces a model inadequacy. Finally, the implementation of the line integral model may introduce code error, if certain steps are taken to optimize the code for speed e.g. on the GPU.*

Remark. This is not an exhaustive list for all possible model errors that may occur, but all are relevant to the work carried out in this thesis and other work related to the PhD project. Other reasons worthy of mention are residual and parametric variability, which we do not consider here, but refer to the discussion in [32] for the interested reader.

Another way to think of model errors is as perturbations of the forward model. This characterization is more limited, because it assumes the true process is also linear. However, it provides a way to characterize the effect that existing regularization techniques have on these types of model errors in inverse problems. As regularization techniques are one of the main tools to solve inverse problems, we go into detail on this topic in the following section.

4.1 Norm-bounded linear model error

In this section, we investigate a class of model errors that can be defined by a norm-bounded linear perturbation of the true process. We also still assume that the data is perturbed by a noise term $\mathbf{e}$. That is, suppose that for the linear inverse problem (2.1) both the model and data are affected by linear perturbations such that

$$\mathbf{A} = \bar{\mathbf{A}} + \mathbf{E}, \quad \mathbf{b} = \bar{\mathbf{b}} + \mathbf{e}, \tag{4.4}$$

where we call $\mathbf{E}$ the linear model error and $\mathbf{e}$ the measurement error and assume that both are norm-bounded. Here, we assume that $\bar{\mathbf{A}}$ is the exact model such that the exact data comes from $\bar{\mathbf{b}} = \bar{\mathbf{A}}\bar{\mathbf{x}}$ where $\bar{\mathbf{x}}$ is the so-called ground truth – or exact solution to the inverse problem.

Equation (4.4) can be considered as a special case of the model error characterization introduced in the beginning of this chapter. To see this relation, note that we can express the inverse problem as

$$\begin{aligned} \mathbf{b} &= \bar{\mathbf{A}}\mathbf{x} + \mathbf{e} && (4.5)\\ &= \mathbf{A}\mathbf{x} + (\bar{\mathbf{A}} - \mathbf{A})\mathbf{x} + \mathbf{e} && (4.6)\\ &= \mathbf{A}\mathbf{x} + \boldsymbol{\eta} + \mathbf{e}, && (4.7) \end{aligned}$$

where the exact solution that satisfies the relation is again the ground truth $\bar{\mathbf{x}}$. The linear model error can thus be expressed on the data side by a model discrepancy $\boldsymbol{\eta} = -\mathbf{E}\mathbf{x}$, with the assumption that both the exact and approximate forward model are linear. The negative sign is simply because of the choice of defining the model error $\mathbf{E}$ as a perturbation of the exact forward model, rather than as a perturbation of the model. In fact, to ease notation and follow existing literature, we are going to also define

$$\bar{\mathbf{A}} = \mathbf{A} + \Delta, \quad \bar{\mathbf{b}} = \mathbf{b} + \boldsymbol{\delta}, \tag{4.8}$$

such that $\Delta = -\mathbf{E}$ and $\boldsymbol{\delta} = -\mathbf{e}$. This will become relevant in Section 4.1.2.

This type of linear model error in (4.4) is well-researched in the literature concerned with perturbation theory [27, 29, 60], total least squares [23–25, 43] and robust optimization [6, 8, 66]. The main goal of perpetuation theory is to provide bounds on the solutions to linear systems given perturbations in the matrix or data vector – i.e.

the forward model and measured data. For our application this provides meaningful insight into how model errors of the type in (4.4) can influence reconstructions.

4.1.1 Perturbation bounds for closed-form solutions

Similar to the perturbation bounds in Theorems 1 and 2 in Chapter 2, it is possible to characterize the influence of the linear model error in (4.4) by bounding the relative difference between the solutions obtained with and without the said errors for certain closed-form solutions.

We start by repeating the classical result, which shows the effect that the model error has on the naive solution to the inverse problem.

Theorem 4 ([29])**.** *Consider the linear system* (2.1) *with perturbations defined in* (4.4) *and let* $\bar{\mathbf{A}}$ *be square and non-singular and assume that* $\|\mathbf{E}\|_2 < \sigma_{\min}$*. Then for the naive solution* $\mathbf{x}^* = \mathbf{A}^{-1}\mathbf{b}$ *and the exact solution* $\bar{\mathbf{x}} = \bar{\mathbf{A}}^{-1}\bar{\mathbf{b}}$ *the following bound holds*

$$\frac{\|\mathbf{x}^* - \bar{\mathbf{x}}\|_2}{\|\bar{\mathbf{x}}\|_2} \leq \frac{\kappa(\bar{\mathbf{A}})}{1-\gamma}\left(\frac{\|\mathbf{e}\|_2}{\|\bar{\mathbf{b}}\|_2} + \frac{\|\mathbf{E}\|_2}{\|\bar{\mathbf{A}}\|_2}\right), \tag{4.9}$$

where $\gamma = \|\mathbf{E}\|_2/\sigma_{\min}$ *and* $\sigma_{\min}$ *is the smallest singular value of* $\bar{\mathbf{A}}$*.*

The requirement that $\|\mathbf{E}\|_2 < \sigma_{\min}$ may seem restrictive, but is necessary to ensure that $\mathbf{A}$ is invertible. The theorem shows that small errors in the model can cause large changes in the reconstructed solution, when the condition number of the forward matrix is large. This shows that even if the measurement noise is negligible some type of regularization may be needed to handle the influence of modelling errors in inverse problems with real data.

In Section 2.3, we presented a regularization strategy that was based on solving an optimization problem with a data-fitting and regularization term and saw that in the case both terms are squared 2-norms (Tikhonov regularization) this alleviated the issue of stability with respect to perturbations in the data. We now show the extended result given the perturbations in the model as well.

Theorem 5 (Hansen, 1989 [27])**.** *Consider the linear system* (2.1) *with perturbations defined in* (4.4) *and assume that* $\|\mathbf{E}\|_2 < \lambda$ *for some fixed* $\lambda > 0$*. Then for the Tikhonov solution* $\mathbf{x}_\lambda^*$ *from* (2.6) *and the Tikhonov solution on the noise free data* $\bar{\mathbf{x}}_\lambda^* = (\bar{\mathbf{A}}^T\bar{\mathbf{A}} + \lambda\mathbf{I})^{-1}\bar{\mathbf{A}}^T\bar{\mathbf{b}}$ *with the same* λ *we have the following bound*

$$\frac{\|\mathbf{x}_\lambda^* - \bar{\mathbf{x}}_\lambda^*\|_2}{\|\bar{\mathbf{x}}_\lambda^*\|_2} \leq \frac{\kappa_\lambda(\bar{\mathbf{A}})}{1-\gamma_\lambda}\left(\frac{\|\mathbf{e}\|_2}{\|\bar{\mathbf{b}}_\lambda\|_2} + 2\frac{\|\mathbf{E}\|_2}{\|\bar{\mathbf{A}}\|_2} + \gamma_\lambda\frac{\|\bar{\mathbf{b}} - \bar{\mathbf{b}}_\lambda\|_2}{\|\bar{\mathbf{b}}_\lambda\|_2}\right), \tag{4.10}$$

where $\bar{\mathbf{b}}_\lambda = \bar{\mathbf{A}}\mathbf{x}_\lambda$*,* $\kappa_\lambda(\bar{\mathbf{A}}) = \|\bar{\mathbf{A}}\|_2/\lambda$ *and* $\gamma_\lambda = \|\mathbf{E}\|_2/\lambda$*.*

These bounds illustrate that (Tikhonov) regularization can be thought of as a strategy to handle linear model and data errors in inverse problems. This is because the

influence of the model error in the Tikhonov solution is reduced by increasing the regularization parameter to decrease the common factor in front of the perturbation terms. The cost of reducing the influence of the model error is in the regularization error as already argued in Chapter 2.

However, these perturbation bounds are generally only shown for regularized solutions that have closed-form expressions, due to how the proof is constructed. For this reason there are many regularization techniques that can not be considered using the above mentioned technique. To investigate this further, we turn to a different approach and see that this can be related to a larger class of regularization methods.

4.1.2 A different perspective on regularization

While regularization is commonly introduced to enforce the problem to be well-posed, the regularization term can also be viewed from a different light, namely as a type of robust optimization – see [6] for an overview. To see this consider the linear inverse problem (2.1) with perturbations (4.8) such that

$$\mathbf{b} + \boldsymbol{\delta} = (\mathbf{A} + \Delta)\,\mathbf{x}. \tag{4.11}$$

The goal in robust optimization is to optimize for the worst-case perturbation in the linear forward operator $\mathbf{A} \in \mathbb{R}^{m\times n}$, given that we know some bound for the perturbation Δ. This is achieved by solving the robust-regression problem [8], which is given by

$$\min_{\mathbf{x}\in\mathbb{R}^n} \max_{\Delta\in\mathcal{U}} g(\mathbf{b} - (\mathbf{A} + \Delta)\,\mathbf{x}), \tag{4.12}$$

with *uncertainty set* $\mathcal{U} = \{\Delta \in \mathbb{R}^{m\times n} : \|\Delta\| \leq \lambda\}$ and where g is a semi-norm. Here $\|\Delta\|$ is the induced matrix norm defined in Appendix D.3.

In a sense the robust regression problem can be seen as solving the inverse problem with the worst-case model error bounded by λ. What is particularly interesting is the equivalence between this worst-case regression problem and variational regularization. A recent theorem summarizes this for a large number of data-fitting and regularization terms.

Theorem 6 (Bertsimas & Copenhaver, 2018 [8])**.** *If $g : \mathbb{R}^m \to \mathbb{R}$ is a semi-norm which is not identically zero and $h : \mathbb{R}^n \to \mathbb{R}$ is a norm, then for any $\mathbf{z} \in \mathbb{R}^m$ and $\mathbf{x} \in \mathbb{R}^n$*

$$\max_{\Delta\in\mathcal{U}_{(h,g)}} g(\mathbf{z} + \Delta\mathbf{x}) = g(\mathbf{z}) + \lambda h(\mathbf{x}), \tag{4.13}$$

where $\mathcal{U}_{(h,g)} = \left\{\Delta \in \mathbb{R}^{m\times n} : \|\Delta\|_{(h,g)} \leq \lambda\right\}$.

As a consequence of this theorem, we find a connection between the worst-case robust formulation in (4.12) and variational regularization (2.8) in the case that the data-fit is a semi-norm and the regularization term is a norm.

Corollary 1 (Bertsimas & Copenhaver, 2018 [8]). *For all* $\mathbf{b} - \mathbf{A}\mathbf{x} \in \mathbb{R}^m$*, we have*

$$\max_{\Delta \in \mathcal{U}_{(h,g)}} g(\mathbf{b} - \mathbf{A}\mathbf{x} + \Delta\mathbf{x}) = g(\mathbf{b} - \mathbf{A}\mathbf{x}) + \lambda h(\mathbf{x}), \tag{4.14}$$

and so

$$\min_{\mathbf{x} \in \mathbb{R}^n} \max_{\Delta \in \mathcal{U}_{(h,g)}} g(\mathbf{b} - (\mathbf{A} + \Delta)\mathbf{x}) = \min_{\mathbf{x} \in \mathbb{R}^n} g(\mathbf{b} - \mathbf{A}\mathbf{x}) + \lambda h(\mathbf{x}), \tag{4.15}$$

where $\mathcal{U}_{(h,g)} = \left\{\Delta \in \mathbb{R}^{m \times n} : \|\Delta\|_{(h,g)} \leq \lambda\right\}$.

The corollary shows that the objective functions of robust regression and variational regularization are equivalent when the terms are (semi-) norms.

The above result indicates that if one aims to devise new methods to handle model errors that perform better than standard regularization techniques, one must incorporate something more informative than worst-case perturbations of the forward model.

Remark. This formulation does not capture all variational regularization schemes, as the regularization term is requred to be a norm. Hence, the TV regularization term (2.9) is not included in this formulation. In fact, one can show that for specific seminorms the equality does not hold [8]. We are at this time unaware of any research extending these results to TV. These results also lead to an interesting interpretation of the regularization parameter that could be explored further, but we do not explore these topics further.

A special case of Corollary 1 is if $p, q \in [1, \infty]$ then

$$\min_{\mathbf{x} \in \mathbb{R}^n} \max_{\Delta \in \mathcal{U}_{q,p}} \|\mathbf{b} - (\mathbf{A} + \Delta)\mathbf{x}\|_p = \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_p + \lambda \|\mathbf{x}\|_q, \tag{4.16}$$

where $\mathcal{U}_{q,p} = \{\Delta \in \mathbb{R}^{m \times n} : \|\Delta\|_{q,p} \leq \lambda\}$. In particular for $p = q = 2$, we recover non-squared Tikhonov regularization (2.5). Strictly speaking, we recover similar formulations to Tikhonov regularization, since we are working on the non-squared norms. As argued in [66] there exists a choice of λ such that we reach the same solution. To see this, consider $h = g = \ell_2^2$ such that the corollary yields

$$\min_{\mathbf{x} \in \mathbb{R}^n} \max_{\Delta \in \mathcal{U}_{(2,2)}^2} \|\mathbf{b} - (\mathbf{A} + \Delta)\mathbf{x}\|_2^2 = \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_2^2, \tag{4.17}$$

where $\mathcal{U}_{(2,2)}^2 = \left\{\Delta \in \mathbb{R}^{m \times n} : \|\Delta\|_2^2 \leq \lambda\right\}$ and $\|\Delta\|_2^2$ is the square of the largest singular value of Δ.

4.1.3 Extending to errors on the data

The previous presentation seems to have ignored the perturbation $\boldsymbol{\delta}$ of the data. In this section, we extend the presentation by explicitly considering this perturbation. It may be that these results are obvious to those working in robust optimization, but

we are at this time unaware of any literature going into detail on this, so we present our derivations here.

In [8], it is briefly mentioned that if we have some uncertainty set for the noise $\boldsymbol{\delta} \in \mathcal{V}$, with $\mathcal{V} = \{\boldsymbol{\delta} \in \mathbb{R}^m : \|\boldsymbol{\delta}\| \leq \beta\}$, then we may be tempted to consider

$$\min_{\mathbf{x}\in\mathbb{R}^n} \max_{\substack{\mathbf{e}\in\mathcal{V} \\ \Delta\in\mathcal{U}}} g(\mathbf{b} + \boldsymbol{\delta} - (\mathbf{A} + \Delta)\,\mathbf{x}), \tag{4.18}$$

but may in fact work with a new loss function $\bar{g}$ instead, defined as

$$\bar{g}(z) := \max_{\delta\in\mathcal{V}} g(\mathbf{z} + \boldsymbol{\delta}). \tag{4.19}$$

In this way, the problem can be reformulated in the form

$$\min_{\mathbf{x}} \max_{\Delta\in\mathcal{U}} \bar{g}(\mathbf{b} - (\mathbf{A} + \Delta)\,\mathbf{x}), \tag{4.20}$$

for which the above theory can be applied directly.

In the following, we construct a proof to show this is actually the case. This can be seen from results below.

Theorem 7. *If $g : \mathbb{R}^m \to \mathbb{R}$ is a semi-norm which is not identically zero then for any $\mathbf{z} \in \mathbb{R}^m$,*

$$\max_{\boldsymbol{\delta}\in\mathcal{V}_g} g(\mathbf{z} + \boldsymbol{\delta}) = g(\mathbf{z}) + \beta, \tag{4.21}$$

where $\mathcal{V}_g := \{\boldsymbol{\delta} \in \mathbb{R}^m : g(\boldsymbol{\delta}) \leq \beta\}$.

Proof. First note by the triangle inequality that for $\boldsymbol{\delta} \in \mathcal{V}_g$, we have

$$g(\mathbf{z} + \boldsymbol{\delta}) \leq g(\mathbf{z}) + g(\boldsymbol{\delta}) \leq g(\mathbf{z}) + \beta. \tag{4.22}$$

We then show that there exists $\hat{\boldsymbol{\delta}} \in \mathcal{V}_g$ such that $g(\mathbf{z} + \hat{\boldsymbol{\delta}}) = g(\mathbf{z}) + \beta$, since then $\max_{\boldsymbol{\delta}\in\mathcal{V}_g} g(\mathbf{z} + \boldsymbol{\delta}) = g(\mathbf{z}) + \beta$.

i) Assume first that $g(\mathbf{z}) \neq 0$. Define $\hat{\boldsymbol{\delta}} := \beta/g(\mathbf{z})\mathbf{z}$, and note that $g(\hat{\boldsymbol{\delta}}) = g(\beta/g(\mathbf{z})\mathbf{z}) = \beta$, so $\hat{\boldsymbol{\delta}} \in \mathcal{V}_g$. Then

$$g(\mathbf{z} + \hat{\boldsymbol{\delta}}) = g(\mathbf{z} + \beta/g(\mathbf{z})\mathbf{z}) = (1 + \beta/g(\mathbf{z}))g(\mathbf{z}) = g(\mathbf{z}) + \beta, \tag{4.23}$$

ii) Assume instead that $g(\mathbf{z}) = 0$. Let $\mathbf{y} \in \mathbb{R}^m$ such that $g(\mathbf{y}) = 1$. Define $\hat{\boldsymbol{\delta}} := \beta\mathbf{y}$. Then by the triangle inequality

$$g(\mathbf{z} + \hat{\boldsymbol{\delta}}) \leq g(\mathbf{z}) + g(\hat{\boldsymbol{\delta}}) = g(\mathbf{z}) + \beta g(\mathbf{y}) = g(\mathbf{z}) + \beta, \tag{4.24}$$

and by the reverse triangle inequality

$$g(\mathbf{z} + \hat{\boldsymbol{\delta}}) \geq g(\hat{\boldsymbol{\delta}}) - g(\mathbf{z}) = g(\hat{\boldsymbol{\delta}}) = \beta = g(\mathbf{z}) + \beta. \tag{4.25}$$

Hence, $g(\mathbf{z} + \hat{\boldsymbol{\delta}}) = g(\mathbf{z}) + \beta$ and since $g(\hat{\boldsymbol{\delta}}) = \beta$, $\hat{\boldsymbol{\delta}} \in \mathcal{V}_g$ and so we have shown the theorem. □

The theorem leads to the following result.

Corollary 2. *For all* $\mathbf{b} - (\mathbf{A} + \Delta)\,\mathbf{x} \in \mathbb{R}^m$ *we have*

$$\max_{\boldsymbol{\delta} \in \mathcal{V}_g} g(\mathbf{b} + \boldsymbol{\delta} - (\mathbf{A} + \Delta)\,\mathbf{x}) = g(\mathbf{b} - (\mathbf{A} + \Delta)\,\mathbf{x}) + \beta, \tag{4.26}$$

hence

$$\min_{\mathbf{x} \in \mathbb{R}^n} \max_{\boldsymbol{\delta} \in \mathcal{V}_g} g(\mathbf{b} + \boldsymbol{\delta} - (\mathbf{A} + \Delta)\,\mathbf{x}) = \min_{\mathbf{x} \in \mathbb{R}^n} g(\mathbf{b} - (\mathbf{A} + \Delta)\,\mathbf{x}) + \beta. \tag{4.27}$$

This provides a way to directly characterize Tikhonov regularization as a robust optimization problem by substituting the following.

$$g(\cdot) := \|\cdot\|_2^2, \quad h(\cdot) := \|\cdot\|_2^2, \tag{4.28}$$

$$\bar{g}(\cdot) := \max_{\boldsymbol{\delta} \in \mathcal{V}_2} \|\cdot + \boldsymbol{\delta}\|_2^2, \quad \mathcal{V}_2 := \{\boldsymbol{\delta} \in \mathbb{R}^m : \|\boldsymbol{\delta}\|_2^2 \leq \beta\}, \tag{4.29}$$

$$\mathcal{U}_{(2,2)}^2 := \{\Delta \in \mathbb{R}^{m \times n} : \|\Delta\|_2^2 \leq \lambda\}. \tag{4.30}$$

Then by Corollary 1 and 2 we have that

$$\min_{\mathbf{x} \in \mathbb{R}^n} \max_{\substack{\boldsymbol{\delta} \in \mathcal{V}_p \\ \Delta \in \mathcal{U}_{(2,2)}^2}} \|\mathbf{b} + \boldsymbol{\delta} - (\mathbf{A} + \Delta)\,\mathbf{x}\|_2^2 = \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|x\|_2^2 + \beta, \tag{4.31}$$

and since the Tikhonov problem is strictly convex we have

$$\arg\min_{\mathbf{x} \in \mathbb{R}^n} \max_{\substack{\boldsymbol{\delta} \in \mathcal{V}_2 \\ \Delta \in \mathcal{U}_{(2,2)}^2}} \|\mathbf{b} + \boldsymbol{\delta} - (\mathbf{A} + \Delta)\,\mathbf{x}\|_2^2 = \arg\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|x\|_2^2. \tag{4.32}$$

That is, Tikhonov regularization can be seen as a type of worse-case regression, where the perturbations of $\mathbf{A}$ and $\mathbf{b}$ are norm-bounded and the regularization parameter can be related to this bound.

4.2 The difficulty with characterizing model errors deterministically

In this section, we provide a perspective on some of the difficulties with deterministic model errors, we have found during this PhD project and give the reason why the majority of our work ended up using a statistical (Bayesian) approach instead.

The deterministic approach to handle model errors that is considered in this chapter, provides an easy way to characterize and handle the model error by a norm-bound. However, as we have also seen, this approach is equivalent to standard regularization in many cases, which is not satisfactory if one aims to improve upon reconstruction quality of existing methods.

In some of the early work of this PhD project, we considered the model error introduced by an unmatched transpose in CT [20] (Paper F), which can be viewed as a type of code error. This was an interesting direction from a theoretical perspective, but from a practical perspective it essentially also boiled down to adding a small amount of regularization to "fix" the issue that arose from the unmatched transpose. The approach did not significantly improve reconstruction quality in practise, because existing approaches already use regularization in some form. Because of this, we are motivated to pursue methods that add more structural prior information of the model error to determine if it is possible to improve the reconstruction result compared to existing regularization techniques.

To add structural priors on the model error in a deterministic setting raises many challenges. Despite this there are some recent research that considers this case such as structural TLS [30] as well as some of the methods in robust optimization [6]. Of particular interest is also the recent series of papers in [12, 13, 34, 35]. Here the authors characterize model errors by a partial order relation, which is essentially equivalent to providing upper and lower bounds on each element of the model matrix in our case. The work in [12] even discusses the case arising from uncertain model parameters. This approach shows great promise and has well-developed theory, but it comes with a huge practical challenge in the case of CT for two reasons. First, it is in most cases not feasible to store upper and lower bounds for each element of a forward matrix in large-scale CT, even if the CT matrix is sparse as argued in Chapter 3. Second, in the case that the view angles are uncertain, these upper and lower bounds may be so large that they are meaningless to consider. For example, we found that one could only bound the length of an X-ray through a pixel by zero and the maximum length through that pixel for most pixels in our CT simulations, due to the relatively large uncertainty in the view angle parameter.

For these reasons, we are motivated to take a different perspective on model errors that instead describe the statistical characteristics of the model error rather than a deterministic approach.

4.3 A Bayesian approach to model errors

In this section, we summarize a Bayesian approach for characterizing model errors using the model discrepancy formulation eluded to at the beginning of this chapter. We are going to distinguish between the model error introduced by approximating a true physical process $\mathcal{P}$ and two mathematical models, which we denote the parametrized model $\mathbf{A}(\boldsymbol{\theta})$ and the fixed model $\mathbf{A}_{\hat{\boldsymbol{\theta}}}$. The physical process $\mathcal{P}$ takes the true representation of the object, denoted $\mathcal{X}$ as input, while the mathematical models take a vector $\mathbf{x}$ as input. The goal is now to characterize the model error introduced by using these models. As earlier in the chapter, we split this into three scenarios as follows.

4.3.1 Model inadequacy

First, we consider the model error caused by the model inadequacy compared to the true physical process. To study this, two modelling choices need to be taken. First, a representation for the physical unknown object $\mathcal{X}$ is needed. In general, it is common to start with a continuous model, where the object is represented by an element in some abstract space and later discretize the representation arriving at the vector representation $\mathbf{x} \in \mathbb{R}^n$. These choices are heavily dependent on the application, so we simply assume that a parametrization $\mathbf{x}$ of $\mathcal{X}$ have been selected e.g. as in Chapter 3 for CT. Second, the physical system $\mathcal{P}$ needs to be represented by a mathematical model. This will generally be a model that depend on some parameters, which we represent by the vector $\boldsymbol{\theta} \in \mathbb{R}^q$. We consider the matrix $\mathbf{A}(\boldsymbol{\theta}) \in \mathbb{R}^{m \times m}$, which represent the discretized linear forward model given a set of model parameters $\boldsymbol{\theta}$.

With these two modelling choices the mathematical model of the inverse problem approximates the physical relation, i.e.

$$\mathcal{P}(\mathcal{X}) \approx \mathbf{A}(\boldsymbol{\theta})\,\mathbf{x}. \tag{4.33}$$

Further, the measured data is assumed to come from the physical model with some added independent measurement noise, i.e.,

$$\mathbf{b} = \mathcal{P}(\mathcal{X}) + \mathbf{e}. \tag{4.34}$$

Remark. To even consider (4.34) may seem uninformative, because for practical purposes we can not know $\mathcal{P}$ nor $\mathcal{X}$. However, we can select a model and parametrization that approximate (4.34) and approximate the modelling error introduced by these choices as indicated below.

In the spirit of [32], we characterize this model error introduced by *model inadequacy* using the reformulation

$$\begin{aligned} \mathbf{b} &= \mathcal{P}(\mathcal{X}) + \mathbf{e}, && (4.35)\\ &= \mathbf{A}(\boldsymbol{\theta})\,\mathbf{x} + (\mathcal{P}(\mathcal{X}) - \mathbf{A}(\boldsymbol{\theta})\,\mathbf{x}) + \mathbf{e}, && (4.36)\\ &= \mathbf{A}(\boldsymbol{\theta})\,\mathbf{x} + \boldsymbol{\eta} + \mathbf{e}, && (4.37) \end{aligned}$$

and we call $\boldsymbol{\eta}$ the model discrepancy term – introduced by model inadequacy in this case.

To our knowledge this is the first type of model discrepancy considered, in particular by the authors in [32]. Here $\boldsymbol{\eta}$ was modelled as a Gaussian Process with a known mean and covariance function based on an assumed prior. In some of our earlier work in the PhD project, we collaborated on a number of projects that utilize this approach in sound field reconstruction [15, 16] (Papers D and E). We do not go into further detail in this thesis.

For large-scale inverse problems Gaussian Processes may be prohibitive due to the high dimensionality. For the purpose of this thesis and the CT test case, we ignore

the influence of model errors introduced by model inadequacy and refer to the above-mentioned papers for the interested reader.

4.3.2 Parameter uncertainty

In inverse problems, it is often the case that the forward model is "fixed" in the sense that physical parameters are assumed to be known exactly, as we have already seen with the CT test case in Chapter 3. This essentially means that we select some parameters $\hat{\boldsymbol{\theta}}$ from a larger set of possible parameters, e.g. a point estimate of a probability distribution $\boldsymbol{\theta} \sim \pi_{\boldsymbol{\theta}}(\cdot)$ such that

$$\mathbf{A}(\boldsymbol{\theta})\,\mathbf{x} \approx \mathbf{A}_{\hat{\boldsymbol{\theta}}}\,\mathbf{x}. \tag{4.38}$$

This can be reflected in the model discrepancy formulation by writing

$$\mathbf{b} = \mathbf{A}_{\hat{\boldsymbol{\theta}}}\,\mathbf{x} + \boldsymbol{\eta} + \mathbf{e}, \tag{4.39}$$

where $\boldsymbol{\eta} = \mathbf{A}(\boldsymbol{\theta})\,\mathbf{x} - \mathbf{A}_{\hat{\boldsymbol{\theta}}}\,\mathbf{x}$. We go into further detail on this formulation in Chapters 5 and 6 as this is the key formulation of the models in our research.

Remark. In the above presentation it is assumed that the data comes from the forward model $\mathbf{A}(\boldsymbol{\theta})\,\mathbf{x}$ with some added measurement noise, which means the model inadequacy is effectively ignored as stated earlier. We leave it to future work to also consider the model inadequacy combined with this model.

4.3.3 Code error

The final type of model error we describe in this way is code error. In the following, we assume that this comes from a type of discretization error, i.e.,

$$\mathbf{A}\mathbf{x} \approx \mathbf{A}^h\mathbf{x}^h, \tag{4.40}$$

where $\mathbf{A}^h \in \mathbb{R}^{m\times h}$ and $\mathbf{x} \in \mathbb{R}^h$ is the forward model with a courser discretization compared to $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{x} \in \mathbb{R}^n$ with $h < n$. Similar to above, we can write the model discrepancy introduced by this discretization error as

$$\mathbf{b} = \mathbf{A}^h\mathbf{x}^h + \boldsymbol{\eta} + \mathbf{e}, \tag{4.41}$$

where $\boldsymbol{\eta} = \mathbf{A}\mathbf{x} - \mathbf{A}^h\mathbf{x}^h$ and again any other model errors are ignored in the formulation.

Remark. The majority of the work concerned with the approximation error approach [31] considers similar formulations to this type of discretization-based code error. In this thesis, we do not explore this topic further.

4.4 Taking model error into account using approximate marginalization

In this section, we describe how to take the model error into account using the model discrepancy formulation above and approximate marginalization. We illustrate this with the parameter uncertainty case, as this is the one relevant to our CT test case, but the approach works for all three cases.

4.4.1 Model and likelihood

We consider the model where the parameter uncertainty is moved to a model discrepancy term (4.39) and aim to derive a likelihood according to this model. We derive the likelihood by defining the combined model discrepancy and measurement noise term $\boldsymbol{\nu} = \boldsymbol{\eta} + \mathbf{e}$ such that (4.39) becomes

$$\mathbf{b} = \mathbf{A}_{\hat{\boldsymbol{\theta}}}\,\mathbf{x} + \boldsymbol{\eta} + \mathbf{e} = \mathbf{A}_{\hat{\boldsymbol{\theta}}}\,\mathbf{x} + \boldsymbol{\nu}, \tag{4.42}$$

where we assume probability distributions for the variables $\boldsymbol{\theta} \sim \pi_{\boldsymbol{\theta}}(\cdot)$, $\mathbf{e} \sim \pi_{\mathbf{e}}(\cdot)$ and $\mathbf{x} \sim \pi_{\mathbf{x}}(\cdot)$. Similar to Section 2.4 the likelihood is then derived by marginalizing out $\boldsymbol{\nu}$ to arrive at

$$\pi(\mathbf{b}|\mathbf{x}) \propto \pi_{\boldsymbol{\nu}|\mathbf{x}}(\mathbf{b} - \mathbf{A}_{\hat{\boldsymbol{\theta}}}\,\mathbf{x}|\mathbf{x}), \tag{4.43}$$

where importantly $\boldsymbol{\nu}$ is not independent of $\mathbf{x}$.

This likelihood takes the model error and noise error into account because of the marginalized variable $\boldsymbol{\nu}$. The issue is that this term can be complicated to determine in practise.

4.4.2 Gaussian approximation

As in [31, 46] a Gaussian approximation is used for the combined term, i.e.

$$\boldsymbol{\nu}|\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x}}, \mathbf{C}_{\boldsymbol{\nu}|\mathbf{x}}), \tag{4.44}$$

leading to a closed-form expression of the likelihood

$$\pi(\mathbf{b}|\mathbf{x}) \propto \exp\left(-\frac{1}{2}\|\mathbf{b} - \mathbf{A}_{\hat{\boldsymbol{\theta}}}\mathbf{x} - \boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x}}\|^2_{\mathbf{C}^{-1}_{\boldsymbol{\nu}|\mathbf{x}}}\right), \tag{4.45}$$

$$\propto \exp\left(-\frac{1}{2}\|\mathbf{L}_{\boldsymbol{\nu}|\mathbf{x}}(\mathbf{b} - \mathbf{A}_{\hat{\boldsymbol{\theta}}}\mathbf{x} - \boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x}})\|^2_2\right), \tag{4.46}$$

where $\mathbf{L}_{\boldsymbol{\nu}|\mathbf{x}}$ is the Cholesky factor of the inverse covariance $\mathbf{C}^{-1}_{\boldsymbol{\nu}|\mathbf{x}} = \mathbf{L}^T_{\boldsymbol{\nu}|\mathbf{x}}\mathbf{L}_{\boldsymbol{\nu}|\mathbf{x}}$.

This approximate likelihood with the marginalized model discrepancy and noise term $\boldsymbol{\nu}$, can then be used when solving the inverse problem either by sampling-based methods or as a data-fitting term in a variational formulation. The main challenge to achieving this is to determine the mean $\boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x}}$ and covariance $\mathbf{C}_{\boldsymbol{\nu}|\mathbf{x}}$ of the combined term $\boldsymbol{\nu}$. This can be achieved by sampling from the prior distributions.

4.4.3 Sampling the model discrepancy

We can generate S samples of model discrepancy by

$$\boldsymbol{\eta}^s = \mathbf{A}(\boldsymbol{\theta}^s)\,\mathbf{x}^s - \mathbf{A}_{\hat{\boldsymbol{\theta}}}\,\mathbf{x}^s, \quad \mathbf{x}^s \sim \pi_{\mathbf{x}}(\cdot),\, \boldsymbol{\theta}^s \sim \pi_{\boldsymbol{\theta}}(\cdot), \tag{4.47}$$

and compute the sample mean

$$\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}|\mathbf{x}} = \frac{1}{S}\sum_{s=1}^{S} \boldsymbol{\eta}^s \tag{4.48}$$

and sample covariance

$$\widetilde{\mathbf{C}}_{\boldsymbol{\eta}|\mathbf{x}} = \frac{1}{S-1}\sum_{s=1}^{S} (\boldsymbol{\eta}^s - \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}|\mathbf{x}})(\boldsymbol{\eta}^s - \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}|\mathbf{x}})^T. \tag{4.49}$$

Assuming that the measurement noise is Gaussian, i.e. $\mathbf{e} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{e}}, \mathbf{C}_{\mathbf{e}})$, we can define the sample mean $\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\nu}|\mathbf{x}} = \boldsymbol{\mu}_{\mathbf{e}} + \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}|\mathbf{x}}$ and sample covariance $\widetilde{\mathbf{C}}_{\boldsymbol{\nu}|\mathbf{x}} = \mathbf{C}_{\mathbf{e}} + \widetilde{\mathbf{C}}_{\boldsymbol{\eta}|\mathbf{x}}$. This leads to the approximate likelihood with the sample mean and covariance, i.e.

$$\pi(\mathbf{b}|\mathbf{x}) \propto \exp\left(-\frac{1}{2}\|\widetilde{\mathbf{L}}_{\boldsymbol{\nu}|\mathbf{x}}(\mathbf{b} - \mathbf{A}_{\hat{\boldsymbol{\theta}}}\mathbf{x} - \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\nu}|\mathbf{x}})\|_2^2\right), \tag{4.50}$$

where $\widetilde{\mathbf{L}}_{\boldsymbol{\nu}|\mathbf{x}}$ is the Cholesky factor of the inverse sample covariance $\widetilde{\mathbf{C}}^{-1}_{\boldsymbol{\nu}|\mathbf{x}}$. For more details see e.g. [31, 46].

Remark. The above approach defines the model discrepancy according to the assumed prior distributions – effectively defining a prior model discrepancy. However, if one has prior distributions that are not particularly informative, this may not be an effective strategy and it may be better to determine the model discrepancy based on the measured data. In Chapter 5, we adapt the above approach to the CT problem with uncertain view angles, in particular by defining the model discrepancy based on a reconstructed solution obtained from the measured data.

4.5 Summary

In this chapter, we gave an overview of various techniques to handle model errors in inverse problems. The first section focused on the case where the model error was seen as a linear perturbation and was norm-bounded. Here, we described the equivalence relation between this formulation and a large group of regularization techniques. Finally, we described a Bayesian approach to handle model errors based on a model discrepancy formulation and related this approach to three types of model errors.

CHAPTER 5

Computed Tomography with view angle marginalization

Relevant papers:

A) **N. A. B. Riis** and Y. Dong "A New Iterative Method for CT Reconstruction with Uncertain View Angles." In: *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer. 2019, pages 156–167. [52]

B) **N. A. B. Riis**, Y. Dong and P. C. Hansen "Computed Tomography Reconstruction with Uncertain View Angles by Iteratively Updated Model Discrepancy." In: *Journal of Mathematical Imaging and Vision* (2020), pages 1–11. [53]

5.1 Motivation and goal

In CT, there is often a hidden assumption that the geometry of the physical set-up is known exactly, which is never the case in practise. One way to characterize the uncertainty of the geometry is to consider the parameters of the CT forward model such as the view angle and distance parameters in (3.4) as being uncertain. As already mentioned, we are going to focus on the case where the view angle parameter is uncertain and assume that other geometric parameters have negligible uncertainty. That is, we formulate the *uncertain view angle CT problem* as

$$\mathbf{b} = \mathbf{A}(\boldsymbol{\theta})\,\mathbf{x} + \mathbf{e}, \quad \boldsymbol{\theta} \sim \pi_{\text{angles}}(\cdot), \quad \mathbf{e} \sim \pi_{\text{noise}}(\cdot), \tag{5.1}$$

where $\mathbf{b} \in \mathbb{R}^m$ is the measured data corrupted by additive measurement noise $\mathbf{e} \in \mathbb{R}^m$ and $\mathbf{A}(\boldsymbol{\theta}) \in \mathbb{R}^{m\times n}$ is the CT forward model explicitly parametrized by the view angles $\boldsymbol{\theta} \in \mathbb{R}^q$ as defined in Chapter 3. We describe our lack of knowledge of the parameters by assigning probability distributions π_{angles} and π_{noise} for the view angle and measurement noise respectively. The CT set-up is illustrated in Figure 5.1.

Uncertainty in the view angles of a CT scan can arise from several sources such as inexact estimates from a calibration procedure or uncertainty in the mechanical positioning of the source and detector. For medical imaging these uncertainties may be small, but from experience this uncertainty can be much larger in custom industrial inspection such as scanning of oil pipes on the sea-bed [55] for example. Another interesting source of view angle uncertainty is related to rigid body motion of the object during the scanning procedure. Here the motion of the object can be viewed as uncertainty in the scanning geometry, which can lead to larger uncertainty in the view angles [45]. In some applications such as tomography of viruses using electron microscopy, the view angles may even be known with such limited accuracy that they are considered unknown in practise [5]. In this case, procedures for estimating the angles may not produce accurate results and estimates may only be accurate to within 1° [41]. In this case it may be necessary to take the remaining uncertainty of the view angles into account during the reconstruction to obtain the best results.

In this chapter, the goal is to obtain a candidate CT reconstruction to (5.1), which takes the uncertainty in the view angles and measurement noise into account during reconstruction. We therefore do not aim to estimate neither the view angles nor the measurement noise directly. The distributions of the view angles and measurement noise are assumed to be known or estimated beforehand.

Because of the above-mentioned goal, our work differs from the majority of existing research in the literature concerned with uncertain view angles in CT. The reason for this is that the common approach when the view angles are uncertain or unknown, is to estimate the view angles along with the CT image either jointly or by first estimating the view angles and later the CT image. To our knowledge there is no work concerned with taking the view angle uncertainty into account in the CT reconstruction without direct estimation. Later in Chapter 6, we consider the issue of view angle estimation and provide related references.

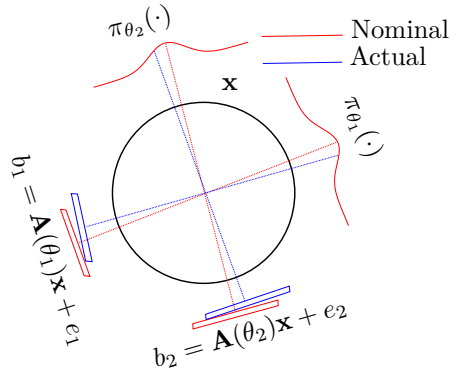


Figure 5.1: Illustration of uncertainty in the view angles for the CT set-up.

5.1.1 CT reconstructions using the nominal view angles

It is commonly the case that the physical uncertainty of the view angles is simply ignored in the CT reconstruction by selecting a forward model based on the nominal view angles as already discussed in 3. This introduces a model error when the true view angles differ from the nominal ones as illustrated in Figure 5.1. However, this approach is not necessarily unjustified as already argued in Chapter 4, since regularization is shown to limit the influence of model errors in the reconstruction. But in some cases, the CT images may end up over-regularized, leading to a loss of detail and it is therefore necessary to take the view angle uncertainty into account in a more direct approach.

To see the effect of ignoring view angle uncertainty, we repeat the results from Chapter 3 for the uncertain view angle CT problem (5.1) by showing a reconstruction using the nominal view angles and compare it with a reconstruction using the actual *true* view angles that the data was acquired with. More details of the simulated CT set-up are given in Appendix D.1. First, in Figure 5.2 we see 3 possible simulated data-sets generated from realizations of uncertain view angles mentioned. To the naked eye there is very little difference between the 3 data sets, but each data set is generated from view angles that vary up to 2°.

Looking at a CT reconstructions in Figure 5.3 obtained using the non-negative total variation variational formulation shown in (3.17) for the left-most CT data set, we see that using the nominal view angles severely deteriorates the CT reconstruction compared to using the true view angles. Here the regularization parameter is chosen as the one giving the lowest relative error. Since the actual true view angles are unknown in practise it may therefore be necessary to take the view angle uncertainty into account to improve the CT reconstruction.

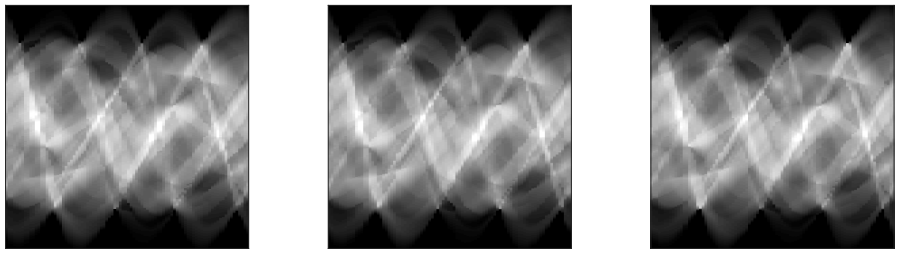


Figure 5.2: Measured data from three different simulations of CT with uncertain view angles.

5.2 Contribution

The main contribution in this chapter and of papers A and B may be summarized as follows. We propose a new model and algorithm for CT reconstruction when the view angles are uncertain by taking the uncertainty of the view angles into account in the CT reconstruction. This is achieved by reformulating the uncertain angle CT problem into one where the forward model is defined by the nominal view angles and the uncertainty in the view angles is moved to a model discrepancy term, similar to the Bayesian approximation error approach described in Chapter 4. Given the new reformulated problem, we derive a new iterative CT reconstruction algorithm, where the model discrepancy is approximately marginalized in the data-fit. To achieve this result a number of issues had to be addressed. We list these as follows.

- The model discrepancy term may have a complicated distribution, but is approximated with a Gaussian, which allows for a closed-form expression of the data-fit.
- A variational optimization problem is defined using the above-mentioned data-fit. The the model discrepancy in the data-fit depend on the unknown CT image. We propose an alternating update scheme to handle this issue.
- In the data-fit, we need to sample and calculate the mean and covariance of the model discrepancy as well as the Cholesky factorization of the inverse covariance. This is carried out block-wise for each view angle, which greatly reduces the computational load of the method.
- The CT image reconstruction is computed using a stochastic primal dual hybrid gradient algorithm to make use of the fact that our variational optimization problem can be separated into data-fitting terms for each view angle. This also greatly reduces the computational work required of our method.

5.3 Method

This section briefly summarizes the CT model and reconstruction algorithm in papers A and B. The main principle of the methods used in our work is a reformulation of the uncertain angle CT problem (5.1), where the model error introduced by using a forward model based on the nominal view angles is characterized by a model discrepancy term. Similar to Chapter 4, this reformulating is given by

$$\mathbf{b} = \mathbf{A}_{\hat{\boldsymbol{\theta}}}\mathbf{x} + \boldsymbol{\eta} + \mathbf{e}, \quad \boldsymbol{\eta} \sim \pi_{\text{discrep}}(\cdot), \quad \mathbf{e} \sim \pi_{\text{noise}}(\cdot), \tag{5.2}$$

where $\boldsymbol{\eta} = \mathbf{A}(\boldsymbol{\theta})\,\mathbf{x} - \mathbf{A}_{\hat{\boldsymbol{\theta}}}\mathbf{x}$ is the model discrepancy term and $\hat{\boldsymbol{\theta}}$ are chosen as the nominal view angles of the scan.

We assume that the view angles are mutually independent, leading to distributions $\theta_i \sim \pi_{\text{angle(i)}}$ for $i = 1, \ldots, q$ and that these can be sampled.

5.3.1 Problem decomposition

A key aspect – which was partly realized in paper A and fully utilized in paper B – is that the assumed independence of the view angles allows the problem to be decomposed into q parts – one for each projection – such that the problem can be written

$$\mathbf{b}_i = \mathbf{A}_{\hat{\theta}_i}\mathbf{x} + \boldsymbol{\eta}_i + \mathbf{e}_i, \quad \text{for } i = 1, \ldots, q, \tag{5.3}$$

where the model discrepancy $\boldsymbol{\eta}_i \sim \pi_{\boldsymbol{\eta}_i}(\cdot)$ is mutually independent. In addition to the view angles being independent, we also assume that the measurement noise is mutually independent, $\mathbf{e}_i \sim \pi_{\mathbf{e}_i}(\cdot)$.

Remark. In the papers the independence of the model discrepancy was seen from a different point of view. This presentation generalizes the presentation a little.

5.3.2 Likelihood

Given the decomposition in (5.3), we can define the Bayesian likelihood for the model as the product of the likelihoods associated with each view angle, i.e.

$$\pi(\mathbf{b}|\mathbf{x}) = \prod_{i=1}^{q} \pi(\mathbf{b}_i|\mathbf{x}). \tag{5.4}$$

As already shown in Section 4.4 the likelihood can then be determined by defining the combined model discrepancy and noise term $\boldsymbol{\nu}_i = \boldsymbol{\eta}_i + \mathbf{e}_i$ and *marginalizing out* this term to arrive at

$$\pi(\mathbf{b}_i|\mathbf{x}) = \pi_{\boldsymbol{\nu}_i|\mathbf{x}}(\mathbf{b}_i - \mathbf{A}_{\hat{\theta}_i}\mathbf{x}|\mathbf{x}). \tag{5.5}$$

5.3.3 Gaussian approximations

Because each of the terms in (5.5) can have a complicated distribution, we approximate them by Gaussian distributions to arrive at a closed-form expression for the likelihood.

That is, we assume that $\boldsymbol{\nu}_i|\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\nu}_i|\mathbf{x}}, \mathbf{C}_{\boldsymbol{\nu}_i|\mathbf{x}})$, which leads to the likelihoods

$$\pi(\mathbf{b}_i|\mathbf{x}) \propto \exp\left(-\frac{1}{2}\|\mathbf{L}_{\boldsymbol{\nu}_i|\mathbf{x}}(\mathbf{b}_i - \mathbf{A}_{\hat{\theta}_i}\mathbf{x} - \boldsymbol{\mu}_{\boldsymbol{\nu}_i|\mathbf{x}})\|_2^2\right), \tag{5.6}$$

where $\mathbf{L}_{\boldsymbol{\nu}_i|\mathbf{x}}$ is the Cholesky factor of the inverse covariance $\mathbf{C}^{-1}_{\boldsymbol{\nu}_i|\mathbf{x}}$.

For a fixed $\mathbf{x}$, the mean $\boldsymbol{\mu}_{\boldsymbol{\nu}_i|\mathbf{x}}$ and covariance $\mathbf{C}_{\boldsymbol{\nu}_i|\mathbf{x}}$ can be estimated by sampling

$$\boldsymbol{\eta}_i^s = \mathbf{A}(\theta_i^s)\mathbf{x} - \mathbf{A}_{\hat{\theta}_i}\mathbf{x}, \quad \theta_i^s \sim \pi_{\text{angle(i)}}, \quad s = 1, \ldots, S, \tag{5.7}$$

and computing the sample mean

$$\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}_i|\mathbf{x}} = \frac{1}{S}\sum_{s=1}^{S} \boldsymbol{\eta}_i^s \tag{5.8}$$

and sample covariance

$$\widetilde{\mathbf{C}}_{\boldsymbol{\eta}_i|\mathbf{x}} = \frac{1}{S-1}\sum_{s=1}^{S} (\boldsymbol{\eta}_i^s - \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}_i|\mathbf{x}})(\boldsymbol{\eta}_i^s - \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}_i|\mathbf{x}})^T, \tag{5.9}$$

followed by adding the measurement noise and covariance, i.e. $\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\nu}_i|\mathbf{x}} = \boldsymbol{\mu}_{\mathbf{e}_i} + \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}_i|\mathbf{x}}$ and covariance $\widetilde{\mathbf{C}}_{\boldsymbol{\nu}_i|\mathbf{x}} = \mathbf{C}_{\mathbf{e}_i} + \widetilde{\mathbf{C}}_{\boldsymbol{\eta}_i|\mathbf{x}}$ similar to Section 4.4.

Remark. If the mean and covariance of the measurement noise is not known or is not Gaussian, one may also approximate $\mathbf{e}$ by the sample mean and covariance from samples of $\mathbf{e}_i \sim \pi_{\mathbf{e}_i}$.

Basing the model discrepancy on a fixed $\mathbf{x}$ is the key difference between this approach and previous work. In principle, this $\mathbf{x}$ can then be determined from a CT reconstruction method before sampling the model discrepancy.

5.3.4 Reconstruction problem

As we already saw in Chapter 2, we can use the likelihood to define a data-fitting term by taking the negative logarithm. We choose to combine this with the total variation regularization term (Section 2.3.2) and non-negativity to arrive at the CT reconstruction problem with approximately marginalized view angle uncertainty

$$\mathbf{x}_{\text{MD-TV}} = \arg\min \frac{1}{2}\sum_{i=1}^{q} \|\widetilde{\mathbf{L}}_{\boldsymbol{\nu}_i|\mathbf{x}}(\mathbf{b}_i - \mathbf{A}_{\hat{\theta}_i}\mathbf{x} - \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\nu}_i|\mathbf{x}})\|_2^2 + \lambda \text{TV}(\mathbf{x}) + g_+(\mathbf{x}), \tag{5.10}$$

where $\widetilde{\mathbf{L}}_{\boldsymbol{\nu}_i|\mathbf{x}}$ is the Cholesky factor of the inverse sample covariance $\widetilde{\mathbf{C}}^{-1}_{\boldsymbol{\nu}_i|\mathbf{x}}$.

The optimization problem in (5.10) is efficiently solved using the stochastic primal dual hybrid gradient (SPDHG) algorithm [18, 21] as described in Appendix D.2.

5.3.5 Reconstruction algorithm

The optimization problem (5.10) raises a number of challenges. The main one is that the sample mean and covariance depend on the CT image $\mathbf{x}$ and vice versa. It is not feasible to recompute the sample mean and covariance each time the data-fitting term needs to be evaluated. For this reason, we choose to alternately update $\mathbf{x}$ and $(\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\nu}_i|\mathbf{x}}, \widetilde{\mathbf{L}}_{\boldsymbol{\nu}_i|\mathbf{x}})$. This approach is shown in both papers A and B and we briefly summarize the idea of the algorithm here.

First, set initial guess $\mathbf{x}^0$, and calculate $\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\nu}_i|\mathbf{x}^0}$ and $\widetilde{\mathbf{L}}_{\boldsymbol{\nu}_i|\mathbf{x}^0}$ from samples (5.7). The alternating algorithm then repeats the following three steps:

1. Update CT reconstruction $\mathbf{x}^{k+1}$ by solving (5.10) with $\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\nu}_i|\mathbf{x}^k}$ and $\widetilde{\mathbf{L}}_{\boldsymbol{\nu}_i|\mathbf{x}^k}$.
2. Sample S samples of model discrepancy $\boldsymbol{\eta}_i|\mathbf{x}^{k+1}$ from (5.7).
3. Compute mean $\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\nu}_i|\mathbf{x}^{k+1}}$ and Cholesky factor of inverse covariance $\widetilde{\mathbf{L}}_{\boldsymbol{\nu}_i|\mathbf{x}^{k+1}}$.

5.4 Results

In this section, we briefly summarize the performance of the proposed model and algorithm and refer to the papers for more details. In Figure 5.3 we show the CT reconstruction after 10 iterations (repeating steps 1-3 10 times) of the algorithm (MD-TV) using $S = 100$ samples of the model discrepancy and regularization parameter chosen to give the lowest relative error. We see that the reconstruction quality is improved compared to ignoring the view angle uncertainty and most of the line artefacts disappear from the reconstruction.

In papers A and B a thorough numerical study was carried out for the method on a similar test problem to the one described in this thesis. Here, we repeat the main results from the papers. First, by comparing relative error and regularization parameter in the top left of Figure 5.4, we see that our method outperforms the L2-TV method for almost all regularization parameters. This indicates that the method is robust towards the choice of regularization parameter. In the top right of Figure 5.4, we also compare the relative error vs epochs – where epochs are defined as one multiplication with the forward matrix or its transpose. Here the black line indicates when the model discrepancy is sampled requiring epochs but not improving reconstruction quality immediately. From our experiments it seems that algorithm is convergences numerically after 3 iterations. Finally, in the bottom row of Figure 5.4, we compare the computational load in terms of epochs for the algorithm using either the PDHG or SPDHG algorithms. We see that by using the SPDHG algorithm, we can carry out 10 iterations of the algorithm before the PDHG finished the first one, which is a significant improvement. We refer to Papers A and B for a more thorough discussion.

5.5 Summary

In this chapter, we summarized papers A and B. We proposed a new model and algorithm for CT reconstruction where the view angles are uncertain. The algorithm takes view angle uncertainty into account by approximate marginalization and improves existing reconstruction techniques based on the nominal view angles. We note there is still some room for improvement in reconstruction quality compared to using the true view angles.

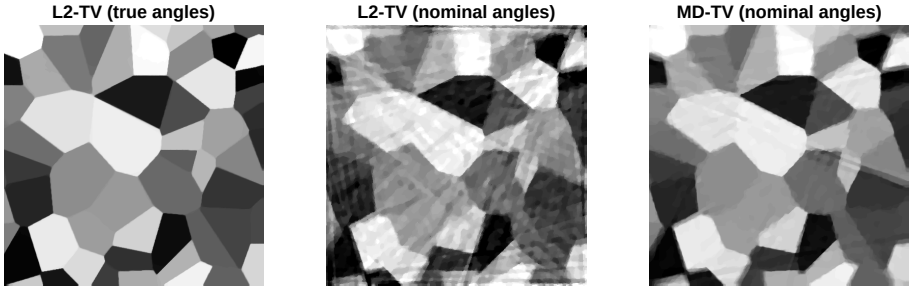


Figure 5.3: We compare reconstructions from the L2-TV method (3.17) for both the nominal and true view angles with the proposed MD-TV method (5.10) using the nominal view angles.

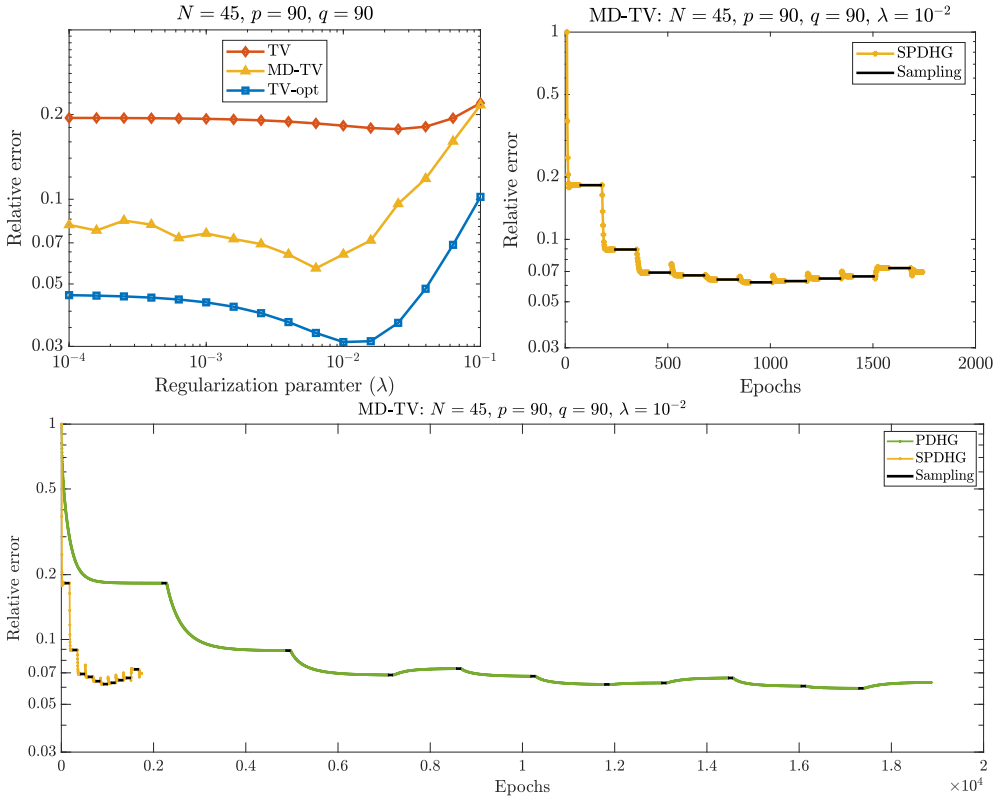


Figure 5.4: Figures from Paper B [53] reprinted with permission. ©2019 Springer. See text for details.

CHAPTER 6

Computed Tomography with view angle estimation

Relevant paper:

C) **N. A. B. Riis**, Y. Dong and P. C. Hansen "Computed Tomography with View Angle Estimation using Uncertainty Quantification." In: *Inverse Problems* (2020), submitted.

6.1 Motivation and goal

In Chapter 5, we proposed a new model and algorithm for CT reconstruction given the view angles of the CT scan were uncertain. In this chapter, we extend on upon this work by proposing an updated approach, that additionally estimates the view angles and also quantifies uncertainty of the estimate. This addition is significant, because it allows for both marginalization and estimation of the view angles, which in turn further improves image reconstruction quality as we will discover.

Compared to Chapter 5, we explicitly require both the view angles and measurement noise to follow Gaussian distributions now. Thus, we consider the uncertain view angle CT problem

$$\mathbf{b} = \mathbf{A}(\boldsymbol{\theta})\,\mathbf{x} + \mathbf{e}, \quad \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \mathrm{diag}(\boldsymbol{\delta})), \quad \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}). \tag{6.1}$$

The Gaussian assumption on the view angles is key to how the view angle estimation is carried out in this work. We assign the mean $\boldsymbol{\mu}_{\boldsymbol{\theta}} \in \mathbb{R}^q$ as the nominal view angles of the scan, which in our experiments are assumed equidistant in the scanning domain. The uncertainty of the nominal view angles is then reflected by the diagonal covariance matrix $\mathrm{diag}(\boldsymbol{\delta}) \in \mathbb{R}^{q\times q}$. The measurement noise is assumed to be i.i.d. Gaussian with zero mean and σ^2 variance and independent of $\boldsymbol{\theta}$. Hence, each view angle and the measurement noise have the assumption of mutual independence.

6.1.1 Previous work

The CT problem in (6.1) can be handled in many ways. For example, for small variances $\boldsymbol{\delta}$ and σ^2 in the view angles and measurement noise respectively, one can utilize direct inversion methods [44] based on e.g. the inversion theorem in Chapter 3. These methods perform well with enough measurement data, but will perform poorly with increasing uncertainty in these parameters even with enough measurement data.

As an alternative to direct methods, variational methods can be robust towards measurement noise and uncertainty in the model by incorporating prior information via a regularization term, as we showed in Chapter 4. However, Figure 5.3 in Chapter 5 illustrated that this broad approach to handle the model error introduced by the view angle uncertainty does come at the cost of reduced image detail.

Instead of using classical regularization, we proposed in Chapter 5 to approximately marginalize the uncertainty of the view angles to improve reconstruction quality further. This improves the CT reconstruction, but some artefacts still remain. To further improve upon this, we therefore propose to additionally estimate the view angles in this chapter.

Scan parameters such as the view angles we are interested in here, can be estimated directly by attaching markers to the scanned object. The issue is that these markers can create artefacts in the reconstruction [59]. An alternative approach is to estimate the view angles after the scan. This can be carried out before the CT reconstruction in a pre-processing step such as analytic methods based on specific geometries [5, 19, 22] or certain Bayesian methods [42]. The issue with these methods is that they suffer from error propagation in the sense that any errors in the view angle estimate propagates to the CT reconstruction.

Recently many approaches to joint estimation of view angles and CT reconstruction have been proposed. The most popular approach for joint estimation are the so-called projection matching (PM) algorithms. These methods try to match the acquired projection data with projections from the forward model using various schemes [38, 45, 48, 64]. In Paper C, we compare our method with a gradient-based projection matching approach recently proposed in [38]. The main difficulty with these methods is that the matching is determined based on a best fit in some metric (typically a 2-norm), which depending on the regularization is a non-convex optimization problem leading to errors in the view angle estimates as seen in [38] (in that paper the view angles are called the tomographic angle).

6.1.2 Goal

The goal of this work is to devise a model and reconstruction algorithm which overcomes the above-mentioned issues. That is, we aim to estimate the view angles with a higher accuracy than existing methods, while also quantifying any remaining uncertainty of the view angle estimate. In the CT reconstruction where the forward model

is based on the estimated angles, we can then take this remaining uncertainty into account using the approach proposed in Chapter 5. This hopefully improves the image quality compared to existing methods with acceptable computational overhead.

6.2 Contribution

The main contribution of this work is a new model and algorithm for CT reconstruction with an added view angle estimation component. The method used for view angle estimation can also straightforwardly be extended to other geometric parameters, but we leave that for future work. We summarize our contributions as follows.

- We propose a new CT reconstruction model and algorithm with an added view angle estimation component.
- Our method is able to perform uncertainty quantification of the view angle estimate and take this uncertainty into account in the CT image reconstruction.
- Our method is fast and computationally efficient both for the angle estimation and CT image reconstruction.
- Numerical results show improved angle estimates and image reconstruction compared to existing work.
- Finally, we show that the angle estimation is robust towards the choice of regularization parameter on the TV prior.

6.3 Method

We aim to estimate the view angles according to the model (6.1) and then reconstruct the CT image using our existing approach (5.10) proposed in Chapter 5 given the estimated view angles.

To achieve this, we use the same reformulation as earlier by moving uncertainty in the view angles to the model discrepancy term, i.e.

$$\mathbf{b}_i = \mathbf{A}_{\mu_{\theta_i}}\mathbf{x} + \boldsymbol{\eta}_i + \mathbf{e}_i, \quad \text{for } i = 1, \dots, q, \tag{6.2}$$

where the model discrepancy is given by $\boldsymbol{\eta}_i = \mathbf{A}(\theta_i)\mathbf{x} - \mathbf{A}_{\mu_{\theta_i}}\mathbf{x}$ for $\theta_i \sim \mathcal{N}(\mu_{\theta_i}, \delta_i)$ and q is the number of projections.

Similar to Chapter 5 the model discrepancy conditioned on $\mathbf{x}$ is assumed to follow a Gaussian distribution $\boldsymbol{\eta}_i \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\eta}_i|\mathbf{x}}, \mathbf{C}_{\boldsymbol{\eta}_i|\mathbf{x}})$.

6.3.1 Joint Gaussian distribution

Given the model-discrepancy based formulation in (6.2) with the combined model discrepancy and measurement noise $\boldsymbol{\nu}_i = \boldsymbol{\eta}_i + \mathbf{e}_i$ conditioned on $\mathbf{x}$ approximated by a Gaussian, i.e. $\boldsymbol{\nu}_i|\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\nu}_i|\mathbf{x}}, \mathbf{C}_{\boldsymbol{\nu}_i|\mathbf{x}})$, we find natural connection between $\mathbf{x}$ and θ_i given the joint Gaussian distribution

$$\begin{bmatrix} \theta_i \\ \boldsymbol{\nu}_i|\mathbf{x} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_{\theta_i} \\ \boldsymbol{\mu}_{\boldsymbol{\nu}_i|\mathbf{x}} \end{bmatrix}, \begin{bmatrix} \delta_i & \mathbf{c}^T_{\boldsymbol{\nu}_i\theta_i|\mathbf{x}} \\ \mathbf{c}_{\boldsymbol{\nu}_i\theta_i|\mathbf{x}} & \mathbf{C}_{\boldsymbol{\nu}_i|\mathbf{x}} \end{bmatrix} \right), \tag{6.3}$$

where $\boldsymbol{\mu}_{\boldsymbol{\nu}_i|\mathbf{x}} = \boldsymbol{\mu}_{\boldsymbol{\eta}_i|\mathbf{x}} \in \mathbb{R}^p$ because the mean of $\mathbf{e}$ is zero, the cross-covariance with respect to a single angle θ_i is $\mathbf{c}_{\boldsymbol{\nu}_i\theta_i|\mathbf{x}} = \mathbf{c}_{\boldsymbol{\eta}_i\theta_i|\mathbf{x}} \in \mathbb{R}^{p\times 1}$ because $\mathbf{e}$ is independent of $\boldsymbol{\theta}$ and $\mathbf{C}_{\boldsymbol{\nu}_i|\mathbf{x}} = \mathbf{C}_{\boldsymbol{\eta}_i|\mathbf{x}} + \sigma^2\mathbf{I} \in \mathbb{R}^{p\times p}$.

6.3.2 Angle estimation

Given the joint distribution (6.3), we estimate the view angles and associated uncertainty given $\mathbf{x}$ from the conditional mean and covariance of the joint Gaussian distribution. Given a reconstruction $\mathbf{x}$ the combined term can be estimated by $\boldsymbol{\nu}_i = \mathbf{b}_i - \mathbf{A}_{\mu_{\theta_i}}\mathbf{x}$ for each $i = 1, \ldots, q$ and the conditional mean is given by

$$\mu_{\theta_i|\boldsymbol{\nu}_i=\hat{\boldsymbol{\nu}}_i} = \mu_{\theta_i} + \mathbf{c}^T_{\boldsymbol{\eta}_i\theta_i|\mathbf{x}}\mathbf{C}^{-1}_{\boldsymbol{\nu}_i|\mathbf{x}}\left(\mathbf{b}_i - \mathbf{A}_{\mu_{\theta_i}}\mathbf{x} - \boldsymbol{\mu}_{\boldsymbol{\nu}_i|\mathbf{x}}\right), \tag{6.4}$$

and the conditional covariance by

$$\delta_{\theta_i|\boldsymbol{\nu}_i=\hat{\boldsymbol{\nu}}_i} = \delta_i - \mathbf{c}^T_{\boldsymbol{\eta}_i\theta_i|\mathbf{x}}\mathbf{C}^{-1}_{\boldsymbol{\nu}_i|\mathbf{x}}\mathbf{c}_{\boldsymbol{\eta}_i\theta_i|\mathbf{x}}. \tag{6.5}$$

That is, given a candidate CT image $\mathbf{x}$ we can estimate the view angles and associated uncertainty using the above conditional parameters given the joint distribution.

6.3.3 Sampling the cross-covariance

Similar to Chapter 5, we need to sample the model discrepancy to compute the sample means and covariances used in the estimates above. That is the sampling of the model discrepancy is carried out exactly as in (5.7) and computation of the sample mean and covariance of the model discrepancy as in (5.8) and (5.9) respectively.

The parameter estimate also requires the cross-covariance of θ_i and $\boldsymbol{\eta}_i$. Given $\mathbf{x}$ these are also estimated by sampling as follows.

$$\widetilde{\mathbf{c}}_{\boldsymbol{\eta}_i\theta_i|\mathbf{x}} = \frac{1}{S-1}\sum_{s=1}^{S}(\boldsymbol{\eta}_i^s - \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}_i|\mathbf{x}})(\theta_i^s - \widetilde{\mu}_{\theta_i}), \tag{6.6}$$

with sample mean

$$\widetilde{\mu}_{\theta_i} = \frac{1}{S}\sum_{s=1}^{S}\theta_i^s, \tag{6.7}$$

where θ_i^s follows the prior distribution and S is the number of samples.

6.3.4 Relaxing the covariance estimate

A difficulty is that the sampled matrices for the covariance estimate depend on $\mathbf{x}$. For this reason, if the reconstruction $\mathbf{x}$ is poor and far from the ground truth $\bar{\mathbf{x}}$, we cannot expect the estimates to be perfect. This issue was illustrated and discussed in paper C, where numerical results showed that the variance estimate was overly optimistic. We reprint the illustration in Figure 6.1.

To handle this issue, we add a relaxation parameter $\alpha \in [0, 1]$ to the covariance update such that the estimate of the variance becomes

$$\delta_{\theta_i|\boldsymbol{\nu}_i=\hat{\boldsymbol{\nu}}_i} = \delta_i - \alpha \widetilde{\mathbf{c}}^T_{\boldsymbol{\eta}_i\theta_i|\mathbf{x}} \widetilde{\mathbf{C}}^{-1}_{\boldsymbol{\nu}_i|\mathbf{x}} \widetilde{\mathbf{c}}_{\boldsymbol{\eta}_i\theta_i|\mathbf{x}}, \tag{6.8}$$

such that if $\alpha = 0$ the variance is not updated and if $\alpha = 1$ the covariance is fully updated. In our implementation we use $\alpha = 0.5$ to conservatively update the covariance estimate.

6.3.5 Algorithm

Because the angle estimation using the conditional mean (6.4) and covariance (6.5) depends on the CT image and vice versa, we propose to alternately update the estimates. The proposed algorithm, denoted CT-VAE, is detailed in Paper C and we briefly summarize it here. The algorithm can be summarized by repeating the following two steps (note the updates are done for all $i = 1, \ldots, q$).

1. **View angle estimation step**
   Sample S_{VA} samples $\boldsymbol{\eta}_i^s = \mathbf{A}(\theta_i^s)\mathbf{x}^k - \mathbf{A}_{\theta_i^k}\mathbf{x}^k, \quad \theta_i^s \sim \mathcal{N}(\theta_i^k, \delta_i^k)$.
   Compute sample mean $\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\nu}_i|\mathbf{x}^k}$, covariance $\widetilde{\mathbf{C}}_{\boldsymbol{\nu}_i|\mathbf{x}^k}$ and cross-covariance $\widetilde{\mathbf{c}}_{\boldsymbol{\eta}_i\theta_i|\mathbf{x}^k}$ from (5.8), (5.9) and (6.6) respectively.
   Update angle estimate θ_i^{k+1} and variance δ_i^{k+1} according to (6.4) and (6.8).

2. **Image reconstruction step**
   Sample S_{CT} samples $\boldsymbol{\eta}_i^s = \mathbf{A}(\theta_i^s)\mathbf{x}^k - \mathbf{A}_{\theta_i^{k+1}}\mathbf{x}^k, \quad \theta_i^s \sim \mathcal{N}(\theta_i^{k+1}, \delta_i^{k+1})$.
   Compute mean $\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\nu}_i|\mathbf{x}^{k+1}}$ and Cholesky factor of inverse covariance $\widetilde{\mathbf{L}}_{\boldsymbol{\nu}_i|\mathbf{x}^{k+1}}$ from (5.8) and (5.9) respectively.
   Update reconstruction $\mathbf{x}^{k+1}$ by solving (5.10) with $\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\nu}_i|\mathbf{x}^k}$, $\widetilde{\mathbf{L}}_{\boldsymbol{\nu}_i|\mathbf{x}^k}$, $\hat{\theta}_i := \theta_i^{k+1}$.

The algorithm alternately updates the estimate of the view angles (including uncertainty estimates) and the CT reconstruction. For each step the model discrepancy is sampled to either 1) update the covariance matrices used in the view angle estimate or 2) update the data-fit in the CT reconstruction. Again, for more details see Paper C.

6.4 Results

In this section, we show simulated numerical experiments to demonstrate the performance of our proposed method. The details of the simulations are in Appendix D.1.

In Figure 6.2 we compare the CT reconstruction obtained using our proposed algorithm (CT-VAE) after 10 iterations compared with the method in Chapter 5 (MD-TV) after 10 iterations both using 100 samples for the model discrepancy. We also compare with the method in (3.17) (L2-TV) given the nominal and true view angles. The reconstruction quality of CT-VAE is similar to the best-case scenario which is the L2-TV method using the true view angles. This means that the view angle uncertainty is completely handled by the combined view angle estimation and marginalization – at least in this simulated case.

In Paper C, we conducted a thorough numerical study of the CT-VAE method including the influence of various parameter choices. In Figures 6.3 and 6.4 we repeat the main results of the study. Note that L2-TV-opt is the L2-TV reconstruction given the true view angles.

First, by comparing relative error with regularization parameter we see in Figure 6.3 that for most choices of regularization parameter the proposed method provides reconstructions with similar relative error compared to using the true view angles. This indicates that the method is robust toward the choice of regularization parameter and completely handles the view angle uncertainty for all these choices.

We show the angle estimates of the algorithm in Figure 6.4. Here, we see that the estimated view angles are far closer to the true view angles compared to initial nominal view angles.

In the paper we also show the comparison with the projecting matching algorithm in [38] and show that our method outperforms this algorithm in both CT reconstruction quality and angle estimation. We refer to paper C for the results and discussion on this comparison.

6.5 Summary

In this chapter, we summarized the work of Paper C proposing a new model and algorithm for CT reconstruction with uncertain view angles. Compared to our previous work this research proposed to also estimate the view angles. We show that our algorithm is able to give reconstructions with similar quality to those obtained with the same TV prior using the true view angles. Hence, the model and algorithm completely takes care of the view angle uncertainty in the simulated tests. Finally, we compare to existing work and show that our algorithm is able to give both better reconstructions (in terms of relative error) and angle estimates than the existing methods for our test case.

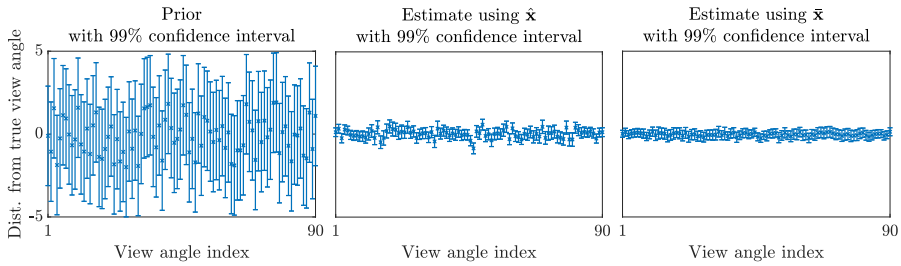


Figure 6.1: Comparing the angle estimation using (6.4), (6.5) based on a CT reconstruction $\hat{\mathbf{x}}$ and the ground truth $\bar{\mathbf{x}}$. Figures from Paper C reprinted with permission.

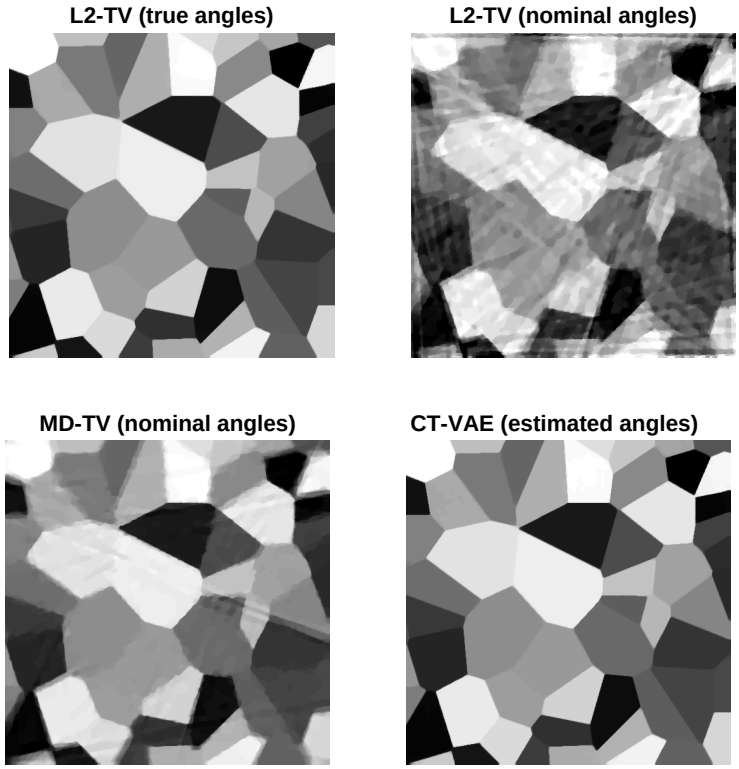


Figure 6.2: CT reconstruction for the uncertain view angle CT test problem.

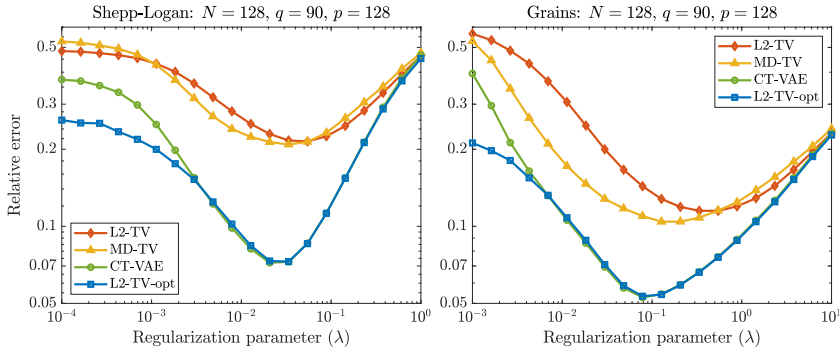


Figure 6.3: Comparing relative error vs regularization parameter for the CT reconstruction methods. Figures from Paper C reprinted with permission.

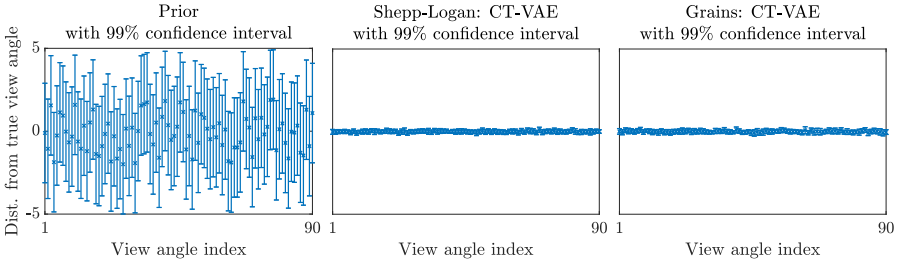


Figure 6.4: Angle estimates (including uncertainty) after 10 iterations of the CT-VAE algorithm. Figures from Paper C reprinted with permission.

CHAPTER 7

Discussion

In this chapter, we give a discussion of selected topics and new research directions with proof-of-concepts that are currently not part of any existing work.

In general, the aim of the PhD project was to study methods for characterizing and reducing model errors in inverse problems broadly construed. A large part of this work was tested on a CT problem with uncertain view angles, which steered the direction of the project. A great benefit was the authors previous experience with CT problems. The CT problem also provided practically relevant challenges in terms of handling large-scale problems. In the following, we discuss the validity of the Gaussian approximation of the model discrepancy term and give a proof-of-concept of how to extend our work to 3D CT.

7.1 Is the Gaussian approximation of the model discrepancy term valid?

Before going further it is worthwhile to discuss if the Gaussian approximation of the model discrepancy $\boldsymbol{\eta}_i$ in (5.3) and (6.2) is even valid for CT. Recall, that the model discrepancy is defined given an image reconstruction $\mathbf{x}$ and some fixed view angle $\hat{\theta}_i$ as

$$\boldsymbol{\eta}_i = \mathbf{A}(\theta_i)\mathbf{x} - \mathbf{A}_{\hat{\theta}_i}\mathbf{x}, \quad \theta_i \sim \pi_{\text{angles}}(\cdot). \tag{7.1}$$

Assuming $\mathbf{x}$ is fixed, the stochastic part of $\boldsymbol{\eta}_i$ comes from the view angle distribution π_{angles}, which is transformed by applying the CT forward model $\mathbf{A}(\cdot)$ and multiplying with the CT image $\mathbf{x}$. Essentially, the uncertainty in $\boldsymbol{\theta}$ is translated to uncertainties in elements of the CT matrix – which consists of the lengths that each X-ray passes through all pixel – and then multiplied by $\mathbf{x}$. One way to see this is to write each projection as a sum similar to Chapter 3, i.e.

$$\mathbf{A}(\theta_i)\mathbf{x} = \sum_{j=1}^{n} \mathbf{a}_j(\theta_i)x_j, \tag{7.2}$$

where $\mathbf{a}_j(\theta_i) \in \mathbb{R}^p$ is a vector of lengths with elements corresponding to the jth column of $\mathbf{A}(\theta_i)$. Even though most of these lengths are zero, a large number non-zero stochastic scalars for each detector element in the projection is summarized. In theory,

if all these have the same variance, the sum is well-approximated by a Gaussian due to the central limit theorem. In practise, this may also be a reasonable approximation even if the variances are not exactly the same. However, after multiplication with $\mathbf{x}$ this will depend on the structure of the CT image, which makes the analysis more difficult.

Instead of a theoretical analysis, we simply investigate histograms of the model discrepancy in Figure 7.1. Here we use 1000 samples for 9 randomly selected pixels given the *initial* CT reconstruction from the nominal view angles (left) and samples given a CT reconstruction after 10 iterations of the algorithm in Chapter 6 (right) where the view angles are also updated. Given the initial CT reconstruction, some of the histograms are clearly not Gaussian from a visual inspection. On the other hand, after running 10 iterations of the CT-VAE algorithm the histograms of the model discrepancy do resemble Gaussian distributions. It is unclear exactly why this trend happens, but we conjecture that it is due to the reduced view angle uncertainty combined with an artefact free CT reconstruction. It would be interesting to explore this topic further.

Remark. Even if the Gaussian assumption is violated, it is still possible to approximately marginalize the model discrepancy using the Gaussian distribution, and the results in Chapters 5 and 6 show that the reconstructions obtained do take the view angle uncertainty into account, which indicates that even if the Gaussian approximation is not perfect the method could still be used.

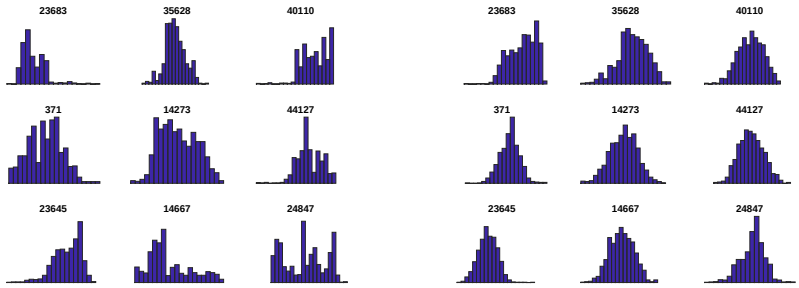


Figure 7.1: Histograms of the sampled model discrepancy for 9 randomly selected detector pixels. Left: given initial CT reconstruction, right: given CT reconstruction after 10 iterations of algorithm in Chapter 6.

We are at this time unaware of other papers investigating the Gaussian assumption on the model discrepancy in detail other than the recent work in [40]. Here the authors do not check if the discrepancy is Gaussian, but rather visually compare the *true model discrepancy* in their simulations with the sampled model discrepancy. In Figure 7.2 we do the same comparison in our case. Indeed the sampled model discrepancy

$\boldsymbol{\eta} = \mathbf{A}(\boldsymbol{\theta}^s)\mathbf{x} - \mathbf{A}_{\hat{\boldsymbol{\theta}}}\mathbf{x}$, where $\boldsymbol{\theta}^s \sim \pi_{\text{angles}}$ and $\mathbf{x}$ is the L2-TV CT reconstruction from the nominal view angles resembles the true model discrepancy defined as $\mathbf{A}_{\bar{\boldsymbol{\theta}}}\bar{\mathbf{x}} - \mathbf{A}_{\hat{\boldsymbol{\theta}}}\bar{\mathbf{x}}$.

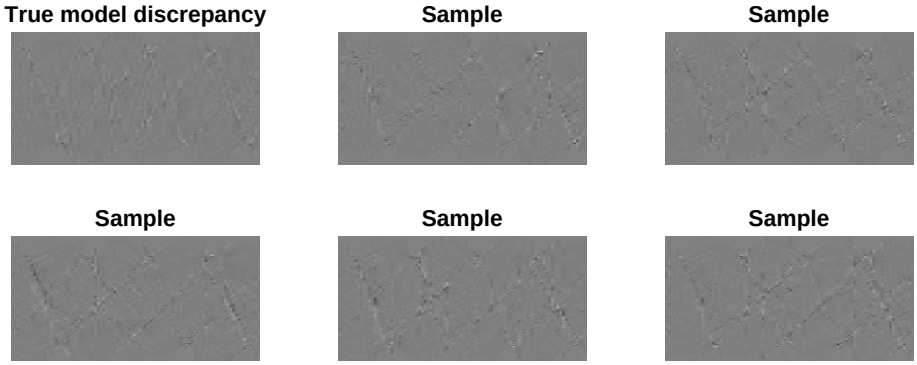


Figure 7.2: Comparing the true model discrepancy with samples of the model discrepancy given an L2-TV CT reconstruction.

7.2 Extending to 3D

We now discuss how our algorithm is straightforwardly extended to a 3D CT case. First note that for cone-beam tomography with a fixed axis rotation, we can formulate the 3D CT problem with uncertain view angles exactly as previously, i.e.,

$$\mathbf{b} = \mathbf{A}(\boldsymbol{\theta})\,\mathbf{x} + \mathbf{e}, \quad \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \operatorname{diag}(\boldsymbol{\delta})), \quad \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}), \tag{7.3}$$

where $\mathbf{A}(\boldsymbol{\theta}) \in \mathbb{R}^{m\times n}$ now represents a discretized cone-beam CT forward model parametrized by the view angles $\boldsymbol{\theta} \in \mathbb{R}^q$, which rotate around a fixed axis. The cone-beam model is illustrated in various places, but we like the illustrations in [1]. Note here that the voxels of the 3D object is represented as a vector $\mathbf{x} \in \mathbb{R}^n$, which now has $n = N^3$ elements, where N is the number of voxels in the image width, height and depth. Similarly, the detector is a 2D panel with p^2 elements such that $m = p^2 q$ is the total number of measured data.

7.2.1 Numerical results

Reusing most of our existing code, but switching the ASTRA forward projector to a 3D cone-beam projection and modifying a few plotting functions, we can handle the 3D case. Here we use $N = 100$, $q = 90$, and $p = 120$ such that both n and m have more than 10^6 elements. We decide to run the CT-VAE algorithm from Chapter 6 with $S_{\text{CT}} = 0$ and $S_{\text{VA}} = 500$, i.e. we do not marginalize the view angle uncertainty

in the CT reconstruction but update the view angles using 500 samples of the model discrepancy. We run 10 iterations of the algorithm. The 3D TV term is computed similar to (2.9), where the gradient is computed using a forward difference for all 6 neighbouring voxels.

In Figure 7.4, we see the CT reconstructions using the nominal, estimated and true view angles similar to the figures we compare in the earlier chapters. The phantom is a 3D version of the Shepp-Logan phantom from [57]. Similar to the previous results, we also see a visual improvement in the reconstructions when estimating the view angles. In Figure 7.3 we see the convergence of the algorithm in terms of relative error. Note that the algorithm reaches the lowest relative error after just 3 iterations.

7.2.2 Even more efficient view angle estimation

For large-scale CT, computing S_{VA} full samples to estimate the view angles, followed by solving a $p^2 \times p^2$ linear system in 3D – i.e. multiplying with the inverse covariance $\mathbf{C}^{-1}_{\boldsymbol{\nu}_i|\mathbf{x}}$) in (6.4)) – may be too computationally expensive even for a single projection. In this case, we can reduce the computational load further by *reducing* the forward model for the view angle estimation. This can simply be carried out by defining a selection function $F : \mathbb{R}^{p^2} \mapsto \mathbb{R}^{p_r^2}$ where $p_r < p$ and then using the reduced model $\mathcal{R}_r(\theta_i)\mathbf{x} = F(\mathcal{R}(\theta_i)\mathbf{x})$. In practise, only the selected p_r^2 X-ray transforms need to be computed. This is feasible because the statistical model does not require the entire CT data, since the angle estimation in (6.4) uses a sampled covariance, which in turn would be reduced to a matrix of size $p_r^2 \times p_r^2$. We show the effect of this reduction in Figure 7.3 where $p_r^2 = p^2/50$ and the selection was carried out by only sampling every 50th element of the model discrepancy. The results show that the CT reconstruction quality in terms of relative error is similar, even when using 1/50th of the data for the angle estimation. It is important to note that the CT reconstruction step still uses the full data set.

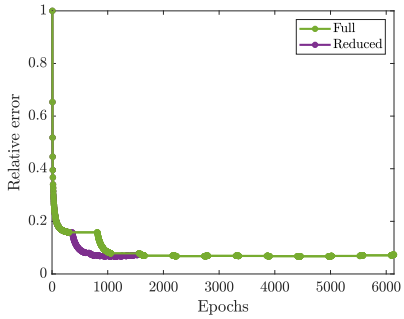


Figure 7.3: Showing the computational gain of sampling 1/50th of the parameters for the model discrepancy used in the view angle estimation.

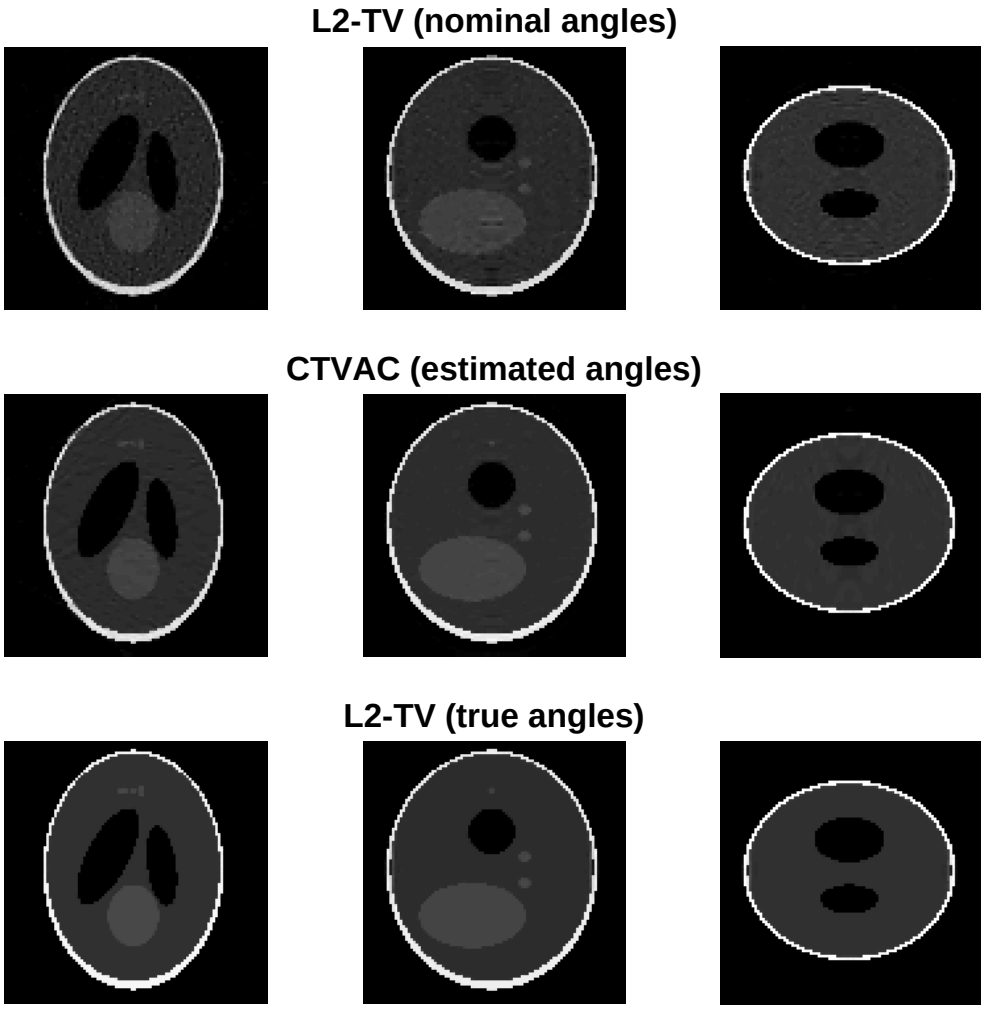


Figure 7.4: Comparing CT reconstructions for the 3D test case.

CHAPTER 8

Concluding remarks

In this chapter we give a perspective on the PhD project as a whole and provide possible directions for future work. Specific contributions and conclusions can be found in the individual chapters as well as the attached papers.

During the PhD project, we studied methods for characterizing and reducing the influence of model errors in inverse problems. A large part of the PhD project focused on a Computed Tomography (CT) application, where the model error was caused by uncertainty in the view angles of the scan. For this reason, the thesis is seen through the lens of CT.

As we saw in Chapter 4, methods for characterizing model errors by perturbations of the forward model have long been studied and have a strong theoretical foundation. In fact, in the chapter we saw that many existing regularization methods can be shown to reduce the influence of such model errors. In Section 4.1.2 we also showed how researchers recently have been able to find a connection between worst-case linear perturbations of the forward model and a large class of existing regularization techniques, further solidifying the concept that existing regularization techniques can work as methods for handling model errors in inverse problems.

The material in Chapter 4 provided an important perspective on the direction of the PhD project, as the goal of developing new methods for reducing the influence of model errors would need to include more structural information of the model errors to be able to improve upon existing techniques. We discussed the difficulties with incorporating this structural information in a deterministic setting for the CT problem, which ultimately led the PhD project in the direction of statistical (Bayesian) methods instead.

In Section 4.4, we discussed a Bayesian method that provided an approach for characterizing any model error by its influence on the data residual. More specifically, the model error was characterized by a model discrepancy term and assumed to be a random variable. This spawned a number of projects which ultimately led to the work of Papers A, B and C (part of thesis) as well as papers F and G (not a part of thesis).

In Papers A, B and C, we applied the model discrepancy characterization to a CT problem with model error caused by uncertain view angles. In Paper A and B (Chapter 5), we used the characterization to derive a reconstruction algorithm which took

the uncertainty in the view angle into account by approximate marginalization. In Paper C (Chapter 6), we used the same characterization to also estimate the view angles and associated uncertainty. Both approaches assumed very little about the CT model – except for independence of the view angles – and we argue that they can therefore be relatively straightforwardly extended to other similar inverse problems.

Combined this work provides a new framework for handling model errors caused by uncertain model parameters in inverse problems. The approach allows for both estimation and marginalization of the uncertain parameters. We showed in simulated numerical experiments, that the framework completely reduces the influence from uncertainty in the view angles for the CT problem with acceptable computational overhead.

8.1 Possible future work

During the PhD project a number of interesting questions came up that we unfortunately did not have the time to consider in detail. Some of these were already briefly discussed in Chapter 7. In the following we describe other possible directions for future work.

An Achilles' heel of this PhD project is the lack of theoretical analysis of the two proposed algorithms in Chapter 5 and Chapter 6. A major reason for this, is that we combine statistically- and optimization-based approaches, which makes analysis difficult. We propose as a future study, to investigate the convergence properties of these methods. A possible direction to investigate is whether it is possible to connect our model to Kalman filters [65].

Another direction of possible future research is to validate our framework on other inverse problems. A straightforward extension is to include more geometric parameters in the CT model such as axis shift/tilt similar to problems considered in [38, 45].

Finally, another extension of our work is also to validate the method on real data. A possible avenue of research is the scanning of oil pipes on the sea-bed [55], which we know from experience has uncertainty in the view angles.

APPENDIX A

A New Iterative Method for CT Reconstruction with Uncertain View Angles

In: International Conference on Scale Space and Variational Methods in Computer Vision - SSVM 2019. Editors: Lellman J., Burger M., Modersitzki J., pp. 156-167.
First published online 05 June 2019.
Part of the Lecture Notes in Computer Science book series (LNCS, volume 11603)
DOI: https://doi.org/10.1007/978-3-030-22368-7_13

Riis, N. A. B., & Dong, Y.

A New Iterative Method for CT Reconstruction with Uncertain View Angles

Nicolai André Brogaard Riis(✉) and Yiqiu Dong

Department of Applied Mathematics and Computer Science, Technical University of Denmark, Richard Petersens Plads, 2800 Kongens Lyngby, Denmark
nabr@dtu.dk

Abstract. In this paper, we propose a new iterative algorithm for Computed Tomography (CT) reconstruction when the problem has uncertainty in the view angles. The algorithm models this uncertainty by an additive model-discrepancy term leading to an estimate of the uncertainty in the likelihood function. This means we can combine state-of-the-art regularization priors such as total variation with this likelihood. To achieve a good reconstruction the algorithm alternates between updating the CT image and the uncertainty estimate in the likelihood. In simulated 2D numerical experiments, we show that our method is able to improve the relative reconstruction error and visual quality of the CT image for the uncertain-angle CT problem.

Keywords: Computed Tomography · Uncertain view angles · Model error · Variational methods · Total variation · Model discrepancy

1 Introduction

In this paper, we consider Computed Tomography (CT) where the geometry of the physical set-up is only known approximately. The goal is to achieve reconstructions that are stable in the presence of uncertainties in the geometric parameters. We restrict our attention to uncertainty in the view angles. We assume that the actual view angles are realizations of some known probability distribution $\pi_{\text{angles}}(\cdot)$ and that the measured sinogram is corrupted by additive Gaussian noise with known mean and covariance.

With the above assumptions, we formulate the CT reconstruction problem under uncertain view angles as estimating the unknown attenuation coefficient image $\mathbf{x} \in \mathbb{R}^n$ from a measured (noisy) sinogram $\mathbf{b} \in \mathbb{R}^m$ following the model

$$\mathbf{b} = \mathbf{A}(\boldsymbol{\theta})\,\mathbf{x} + \mathbf{e}, \quad \boldsymbol{\theta} \sim \pi_{\text{angles}}(\boldsymbol{\theta}),\ \mathbf{e} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{e}}, \mathbf{C}_{\mathbf{e}}), \tag{1}$$

The work was supported by the National Natural Science Foundation of China via Grant 11701388.

J. Lellmann et al. (Eds.): SSVM 2019, LNCS 11603, pp. 156–167, 2019.
https://doi.org/10.1007/978-3-030-22368-7_13

where $\mathbf{e} \in \mathbb{R}^m$ is the additive Gaussian noise with mean $\boldsymbol{\mu}_\mathbf{e}$ and symmetric positive definite covariance $\mathbf{C}_\mathbf{e}$. The parameterized matrix $\mathbf{A}(\boldsymbol{\theta}) \in \mathbb{R}^{m\times n}$ is the discrete approximation of the Radon transform with view angles $\boldsymbol{\theta} \in \mathbb{R}^q$. The measurement at detector element l from the view angle θ_i with $i = 1, \ldots, q$, i.e., $(\mathbf{A}(\theta_i)\mathbf{x})_l$, is a discretization of $(\mathcal{R}f)(\theta_i, \mathrm{s}_l) = \int_{\mathbb{R}} f(\mathrm{s}_l V(\theta_i) + tV^{\perp}(\theta_i))\,\mathrm{d}t$, where s_l gives the position of the lth pixel on the detector with $l = 1, \cdots, p$, and $m = qp$. Moreover, the function f is the continuous representation of $\mathbf{x}$, $\mathrm{V}(\theta) = (\cos\theta, \sin\theta)$ and $\mathrm{V}^{\perp}(\theta) = V(\theta + \pi/2)$. For more details on the mathematical model of CT see e.g. [1]. We emphasize that the goal in this work is to estimate the CT image $\mathbf{x}$ from a measured sinogram $\mathbf{b}$ according to the model (1) with uncertain view angles $\boldsymbol{\theta}$ and noise $\mathbf{e}$. Here, $\boldsymbol{\theta}$ and $\mathbf{e}$ are considered as nuisance or uninteresting parameters, and they are only taken into account when reconstructing $\mathbf{x}$ without being explicitly estimated.

1.1 Previous Work

Many variational methods have been proposed for CT reconstruction, see e.g., [2–4]. In general, variational models in these methods consist of a data-fitting term and a regularization term, and these two terms are balanced by a regularization parameter. In order to deal with the ill-posedness in CT reconstruction problems, the choice of regularization is very important. Different regularization techniques have been applied, for example, total variation (TV) regularization [5] and framelet representations [2]. But these methods do not take parametric uncertainty such as uncertainty in view angles into account. Therefore, good performance of the methods is not guaranteed if the view angles are uncertain.

The CT reconstruction problem with uncertain view angles in (1) is generally solved by estimating the view angles from some measurements. Geometrical calibration of models in CT has been studied, see e.g. [6] for a review. Typically such methods are based on reference objects or reference instruments for the calibration. Recently, in the case of uncertain or unknown view angles a few reconstruction methods that only use the measured sinogram without reference objects or instruments have been proposed, see [7–9]. These methods aim to estimate the view angles $\boldsymbol{\theta}$ in addition to the CT image $\mathbf{x}$, and can be categorized into two groups: (1) Estimating view angles directly from projection data and then estimating the CT image and (2) simultaneously estimating view angles and CT image.

In [9] it has been shown that if the scanned objects are asymmetrical then view angles can be uniquely determined by sinogram measurements. According to this result, we can estimate angles directly from complete measurements. However, if the object is partly symmetrical or the measurements are not sufficient, we cannot obtain an accurate angle estimation, see [8]. Then, due to error propagation, an inaccurate angle estimation would lead to an unsatisfactory reconstruction.

The simultaneous methods such as Bayesian sampling-based methods [7] can effectively avoid error propagation, but they are limited by computational complexity and generally require many evaluations of the forward model (1), which makes them unfeasible for large-scale problems.

There are also a few methods for characterizing and reducing model errors in general inverse problems, see [10–15]. Most of these methods are based on the statistical description of the model error in a Bayesian setting. This leads to a natural way of incorporating uncertainties and modelling errors in the model. However, full Bayesian methods also suffer from computational complexity issues except in cases when the object is assumed to follow a Gaussian distribution, in which case closed-form solutions exist.

1.2 Our Contribution

In this paper, we propose a new iterative algorithm for CT reconstruction with uncertain view angles. The main step in the algorithm is based on a variational method, which combines the state-of-the-art regularization such as TV with a modified data-fitting term, that includes the uncertainty in the view angles via a model-discrepancy term. Since the model-discrepancy term depends on the estimated reconstruction, we update it and the reconstruction alternately. The simulated numerical results show that the new algorithm is able to reduce the relative error and improve the visual quality of the reconstructions.

2 Our Method

The CT reconstruction with uncertain view angles is formulated in (1) with an assumption of the probability distribution on the view angles. By including the known expected view angles, $\hat{\boldsymbol{\theta}}$, we can reformulate the problem as follows.

$$\mathbf{b} = \mathbf{A}(\hat{\boldsymbol{\theta}})\,\mathbf{x} + \boldsymbol{\eta} + \mathbf{e}, \quad \mathbf{e} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{e}}, \mathbf{C}_{\mathbf{e}}), \tag{2}$$

where the new random variable $\boldsymbol{\eta} = \boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{x}) = \mathbf{A}(\boldsymbol{\theta})\,\mathbf{x} - \mathbf{A}(\hat{\boldsymbol{\theta}})\,\mathbf{x}$ with $\boldsymbol{\theta} \sim \pi_{\text{angles}}(\boldsymbol{\theta})$ models the uncertainties associated with the view angles. Note that (2) is consistent with (1). By this reformulation, we basically shift the uncertainties in the view angles to the model-discrepancy term $\boldsymbol{\eta}$, which will be used to derive our variational model.

Defining modelling errors as an additive model-discrepancy term was first applied in [16] in the field of model calibration of physical and computer models. The distribution of $\boldsymbol{\eta}$ was assumed as a Gaussian Process and determined as a model correction term in addition to $\mathbf{x}$. In [11,17], this idea was applied in Bayesian inverse problems and named as the Approximation Error Approach (AEA). The main differences are that in the AEA $\boldsymbol{\eta}$ is used to represent the difference between two grid systems instead of a model discrepancy and it is marginalized out in the likelihood function. The outputs of the AEA are the distributions of $\mathbf{x}$ and $\boldsymbol{\eta}$. To further improve the results, in [14] an iterative scheme was introduced where the distributions of $\mathbf{x}$ and $\boldsymbol{\eta}$ are updated alternately.

Inspired by the ideas of the AEA, we derive the likelihood according to the model (2) by marginalizing out both $\boldsymbol{\eta}$ and $\mathbf{e}$. Define $\boldsymbol{\nu} = \boldsymbol{\eta}+\mathbf{e}$, and the likelihood is given by

$$\pi(\mathbf{b}\,|\,\mathbf{x}) = \int_{\mathbb{R}^m} \pi(\mathbf{b},\boldsymbol{\nu}\,|\,\mathbf{x})\mathrm{d}\boldsymbol{\nu} = \int_{\mathbb{R}^m} \pi(\mathbf{b}\,|\,\mathbf{x},\boldsymbol{\nu})\pi(\boldsymbol{\nu}\,|\,\mathbf{x})\mathrm{d}\boldsymbol{\nu} = \pi_{\boldsymbol{\nu}\,|\,\mathbf{x}}(\mathbf{b}-\mathbf{A}(\hat{\boldsymbol{\theta}})\,\mathbf{x}\,|\,\mathbf{x}). \tag{3}$$

The conditional distribution of $\boldsymbol{\nu}\,|\,\mathbf{x}$ may be rather complicated, but we can approximate it by a simpler distribution such as a Gaussian. Gaussian approximations has been shown experimentally to be reasonable in many applications [11,13,14,17,18]. Here, we assume that $\boldsymbol{\eta}\,|\,\mathbf{x}$ follows a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\eta}\,|\,\mathbf{x}}, \mathbf{C}_{\boldsymbol{\eta}\,|\,\mathbf{x}})$ with mean $\boldsymbol{\mu}_{\boldsymbol{\eta}\,|\,\mathbf{x}}$ and covariance $\mathbf{C}_{\boldsymbol{\eta}\,|\,\mathbf{x}}$ and $\mathbf{e}$ is independent of $\mathbf{x}$. Then we obtain the negative log-likelihood function

$$-\log\pi(\mathbf{b}\,|\,\mathbf{x}) \propto \frac{1}{2}\|\mathbf{b}-\mathbf{A}(\hat{\boldsymbol{\theta}})\,\mathbf{x}-\boldsymbol{\mu}_{\boldsymbol{\nu}\,|\,\mathbf{x}}\|^2_{\mathbf{C}^{-1}_{\boldsymbol{\nu}\,|\,\mathbf{x}}} = \frac{1}{2}\|\mathbf{L}_{\boldsymbol{\nu}\,|\,\mathbf{x}}(\mathbf{b}-\mathbf{A}(\hat{\boldsymbol{\theta}})\,\mathbf{x}-\boldsymbol{\mu}_{\boldsymbol{\nu}\,|\,\mathbf{x}})\|^2_2, \tag{4}$$

where $\boldsymbol{\mu}_{\boldsymbol{\nu}\,|\,\mathbf{x}} = \boldsymbol{\mu}_{\mathbf{e}} + \boldsymbol{\mu}_{\boldsymbol{\eta}\,|\,\mathbf{x}}$, $\mathbf{C}_{\boldsymbol{\nu}\,|\,\mathbf{x}} = \mathbf{C}_{\mathbf{e}} + \mathbf{C}_{\boldsymbol{\eta}\,|\,\mathbf{x}}$ is the combined covariance of the measurement noise and model discrepancy, and $\mathbf{L}^T_{\boldsymbol{\nu}\,|\,\mathbf{x}}\mathbf{L}_{\boldsymbol{\nu}\,|\,\mathbf{x}} = \mathbf{C}^{-1}_{\boldsymbol{\nu}\,|\,\mathbf{x}}$ is the Cholesky factorization of the inverse covariance. Applying regularization techniques, we can formulate a variational model for (2) that gives a stable solution with respect to the uncertain view angles and measurement noise using the likelihood (4). TV regularization has shown good performance in CT [2], and thus we use it as regularization term and obtain the following variational model

$$\min_{\mathbf{x}\geq 0} \ \frac{1}{2}\|\mathbf{L}_{\boldsymbol{\nu}\,|\,\mathbf{x}}(\mathbf{b}-\mathbf{A}(\hat{\boldsymbol{\theta}})\,\mathbf{x}-\boldsymbol{\mu}_{\boldsymbol{\nu}\,|\,\mathbf{x}})\|^2_2 + \lambda\mathrm{TV}(\mathbf{x}), \tag{5}$$

with minimizer $\mathbf{x}_{\mathrm{STV}}$ and regularization parameter $\lambda > 0$. We use $\mathrm{TV}(\mathbf{x}) = \sum_i \|[\nabla\mathbf{x}]_i\|_2$, where $\|[\nabla\mathbf{x}]_i\|_2 = \sqrt{(\nabla_h\mathbf{x})_i^2 + (\nabla_v\mathbf{x})_i^2}$, with $(\nabla_h\mathbf{x})_i$ and $(\nabla_v\mathbf{x})_i$ denoting the derivatives of $\mathbf{x}_i$ along horizontal and vertical directions with symmetric boundary condition, respectively. A non-negativity constraint is added because the attenuation coefficients $\mathbf{x}$ cannot be negative.

2.1 An Iterative Algorithm

The variational model defined in (5) still leaves the question of how the mean and covariance of $\boldsymbol{\eta}\,|\,\mathbf{x}$ are determined. Given a reconstruction $\mathbf{x}$, one can generate N_{samp} samples of $\boldsymbol{\eta}\,|\,\mathbf{x}$ by drawing samples $\boldsymbol{\theta}^s$ following the distribution $\pi_{\mathrm{angles}}(\boldsymbol{\theta})$, and then evaluate the model discrepancy by

$$\boldsymbol{\eta}^s = \mathbf{A}(\boldsymbol{\theta}^s)\,\mathbf{x} - \mathbf{A}(\hat{\boldsymbol{\theta}})\,\mathbf{x}, \quad \boldsymbol{\theta}^s \sim \pi_{\mathrm{angles}}(\boldsymbol{\theta}),\ s = 1,\ldots,N_{\mathrm{samp}}. \tag{6}$$

The sample mean and covariance given $\mathbf{x}$ can then be calculated by

$$\boldsymbol{\mu}^{\mathrm{samp}}_{\boldsymbol{\eta}\,|\,\mathbf{x}} = \frac{1}{N_{\mathrm{samp}}}\sum_{s=1}^{N_{\mathrm{samp}}} \boldsymbol{\eta}^s, \tag{7}$$

and

$$\mathbf{C}^{\text{samp}}_{\boldsymbol{\eta}\,|\,\mathbf{x}} = \frac{1}{N_{\text{samp}}-1}\sum_{s=1}^{N_{\text{samp}}}(\boldsymbol{\eta}^s - \boldsymbol{\mu}_{\boldsymbol{\eta}\,|\,\mathbf{x}})(\boldsymbol{\eta}^s - \boldsymbol{\mu}_{\boldsymbol{\eta}\,|\,\mathbf{x}})^T. \tag{8}$$

If we have a good estimate of $\mathbf{x}$, we can obtain good samples of the model discrepancy, and then we can use the sample mean and covariance in the model (5) to further improve the reconstruction result. This leads to an iterative scheme for alternately updating the estimate of $\mathbf{x}$ and the estimates of mean and covariance of $\boldsymbol{\eta}\,|\,\mathbf{x}$. The iterative scheme is shown in Algorithm 1.

Algorithm 1. Iterative update of reconstruction and likelihood

Inputs: $\mathbf{b}$, λ, $\hat{\boldsymbol{\theta}}$, and $\pi_{\text{angles}}(\boldsymbol{\theta})$. Initial choice of $\mathbf{L}^{[0]}_{\boldsymbol{\nu}\,|\,\mathbf{x}}$ and $\boldsymbol{\mu}^{[0]}_{\boldsymbol{\nu}\,|\,\mathbf{x}}$.

Output: $\mathbf{x}^{[K]}_{\text{STV}}$

for $k = 1, 2, \ldots, K$

  $\mathbf{x}^{[k+1]}_{\text{STV}} = \arg\min_{\mathbf{x}\geq\mathbf{0}} \frac{1}{2}\|\mathbf{L}^{[k]}_{\boldsymbol{\nu}\,|\,\mathbf{x}}(\mathbf{b} - \mathbf{A}(\hat{\boldsymbol{\theta}})\,\mathbf{x} - \boldsymbol{\mu}^{[k]}_{\boldsymbol{\nu}\,|\,\mathbf{x}})\|_2^2 + \lambda\text{TV}(\mathbf{x})$

  for $s = 1, 2, \ldots, N_{\text{samp}}$

    $\boldsymbol{\eta}^s = \mathbf{A}(\boldsymbol{\theta}^s)\,\mathbf{x}^{[k+1]}_{\text{STV}} - \mathbf{A}(\hat{\boldsymbol{\theta}})\,\mathbf{x}^{[k+1]}_{\text{STV}}$ for $\boldsymbol{\theta}^s \sim \pi_{\text{angles}}(\boldsymbol{\theta})$

  end

  Estimate $\boldsymbol{\mu}_{\boldsymbol{\eta}\,|\,\mathbf{x}}$ and $\mathbf{C}_{\boldsymbol{\eta}\,|\,\mathbf{x}}$ according to (7) and (8)

  $\boldsymbol{\mu}^{[k+1]}_{\boldsymbol{\nu}\,|\,\mathbf{x}} = \boldsymbol{\mu}_{\mathbf{e}} + \boldsymbol{\mu}_{\boldsymbol{\eta}\,|\,\mathbf{x}}$

  $\mathbf{L}^{[k+1]}_{\boldsymbol{\nu}\,|\,\mathbf{x}} = \text{chol}\left((\mathbf{C}_{\mathbf{e}} + \mathbf{C}_{\boldsymbol{\eta}\,|\,\mathbf{x}})^{-1}\right)$

end

Here, $\text{chol}(\mathbf{C}^{-1})$ gives the Cholesky factorization of the inverse covariance $\mathbf{C}^{-1}$, i.e., $\mathbf{L}^T\mathbf{L} = \mathbf{C}^{-1}$. In the initialization, we use the measurement noise mean and covariance to initialize $\mathbf{L}^{[0]}_{\boldsymbol{\nu}\,|\,\mathbf{x}} = \mathbf{L}_{\mathbf{e}}$ and $\boldsymbol{\mu}^{[0]}_{\boldsymbol{\nu}\,|\,\mathbf{x}} = \boldsymbol{\mu}_{\mathbf{e}}$, where $\mathbf{L}^T_{\mathbf{e}}\mathbf{L}_{\mathbf{e}} = \mathbf{C}^{-1}_{\mathbf{e}}$.

Compared with the AEA proposed in [14], our method has two main differences. First, our method deals with the uncertainties in the model parameters and the accurate model is unknown, while in the AEA the accurate model is known and the discrepancy is between two different grid systems. Secondly, in our method, we apply a variational method incorporated with regularization techniques to obtain a reconstruction result, which can be solved by many advanced optimization methods, while in the AEA the distribution of the reconstruction is obtained by applying Bayesian inversion methods, which leads to much higher computational complexity.

2.2 Approximation of $\mathbf{L}_{\boldsymbol{\nu}\,|\,\mathbf{x}}$

Because of the computational complexity in calculating the inverse and Cholesky factorization of the covariance matrix $\mathbf{C}_{\mathbf{e}} + \mathbf{C}_{\boldsymbol{\eta}\,|\,\mathbf{x}}$, our method can be of limited use for solving large-scale CT problems. To overcome this limitation, we

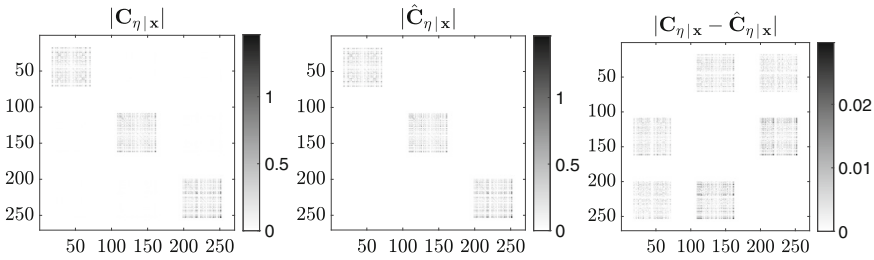


Fig. 1. An example of the absolute value of the full covariance $\mathbf{C}_{\boldsymbol{\eta}\,|\,\mathbf{x}}$, its approximation $\hat{\mathbf{C}}_{\boldsymbol{\eta}\,|\,\mathbf{x}}$ according to (9), and their absolute difference for the uncertain angle CT problem (2). We conclude that the approximation is reasonable for this problem.

approximate the covariance matrix $\mathbf{C}_{\boldsymbol{\eta}\,|\,\mathbf{x}} \in \mathbb{R}^{m\times m}$ by a block diagonal matrix $\hat{\mathbf{C}}_{\boldsymbol{\eta}\,|\,\mathbf{x}}$ given by

$$\hat{\mathbf{C}}_{\boldsymbol{\eta}\,|\,\mathbf{x}} = \begin{bmatrix} \mathbf{C}_{\boldsymbol{\eta}\,|\,\mathbf{x},11} & & & \\ & \mathbf{C}_{\boldsymbol{\eta}\,|\,\mathbf{x},22} & & \\ & & \ddots & \\ & & & \mathbf{C}_{\boldsymbol{\eta}\,|\,\mathbf{x},qq} \end{bmatrix}, \tag{9}$$

where $\mathbf{C}_{\boldsymbol{\eta}\,|\,\mathbf{x},ii} \in \mathbb{R}^{p\times p}$ are the block diagonal parts of $\mathbf{C}_{\boldsymbol{\eta}\,|\,\mathbf{x}}$, $q = m/p$ is the number of view angles and p is the number of detector pixels. Then, if the Gaussian measurement noise $\mathbf{e}$ is i.i.d., i.e., $\mathbf{C}_{\mathbf{e}} = \sigma^2\mathbf{I}_m$, we can compute the Cholesky factorization of the approximate inverse covariance $(\mathbf{C}_{\mathbf{e}} + \hat{\mathbf{C}}_{\boldsymbol{\eta}\,|\,\mathbf{x}})^{-1}$ block-wise as follows

$$\hat{\mathbf{L}}_{\boldsymbol{\nu}\,|\,\mathbf{x}} = \begin{bmatrix} \mathrm{chol}((\mathbf{C}_{\boldsymbol{\eta}\,|\,\mathbf{x},11} + \sigma^2\mathbf{I}_p)^{-1}) & & \\ & \ddots & \\ & & \mathrm{chol}((\mathbf{C}_{\boldsymbol{\eta}\,|\,\mathbf{x},qq} + \sigma^2\mathbf{I}_p)^{-1}) \end{bmatrix}. \tag{10}$$

If the full covariance was used multiplication of an vector with $\mathbf{L}_{\boldsymbol{\nu}\,|\,\mathbf{x}}$ would require $\mathcal{O}(m^2) = \mathcal{O}(p^2q^2)$ operations, whereas multiplication with $\hat{\mathbf{L}}_{\boldsymbol{\nu}\,|\,\mathbf{x}}$ would only be $\mathcal{O}(p^2q)$ operations. Additionally, the matrix inversion and Cholesky factorization is reduced from $\mathcal{O}(m^3) = \mathcal{O}(p^3q^3)$ to $\mathcal{O}(p^3q)$ operations.

In Fig. 1, we show the absolute values of the full covariance $\mathbf{C}_{\boldsymbol{\eta}\,|\,\mathbf{x}}$, its approximation $\hat{\mathbf{C}}_{\boldsymbol{\eta}\,|\,\mathbf{x}}$ according to (9), and their absolute difference. The values in the off-diagonal parts are much smaller than those in the block diagonal parts of $\mathbf{C}_{\boldsymbol{\eta}\,|\,\mathbf{x}}$. Hence, the approximation is reasonable for this problem. In the following experiments line 8 in Algorithm 1 is therefore approximated by (10).

3 Numerical Experiments

In this section, we present simulated 2D numerical results to show the performance of our method. The experiments are carried out in MATLAB and we use

Table 1. The physical parameters in the simulated CT experiments.

Parameter	Value
Beam type	Fan-beam
Reconstruction domain size	50 cm × 50 cm
Source to center distance	50 cm
Source to detector distance	100 cm
Detector length	130 cm
Small example	
Image pixels	$n = 45 \times 45$
Detector pixels	$p = 90$
Number of view angles	$q = 90$
View angle standard deviation	$\delta = 1.2^\circ$
Larger example	
Image pixels	$n = 135 \times 135$
Detector pixels	$p = 270$
Number of projection angles	$q = 270$
View angle standard deviation	$\delta = 0.4^\circ$

the ASTRA toolbox [19] and "Spot Operators" [20] for matrix-free forward- and back-projections, i.e., for multiplication with $\mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{A}(\boldsymbol{\theta})^T$. In the simulated CT problems arising from (1), the physical parameters are shown in Table 1. In both examples, the distribution of the view angles is assumed to be i.i.d. Gaussian with equidistant view angles from 0° to 360°, denoted by $\boldsymbol{\theta}^{\text{equid}}$, as mean and δ^2 as variance. These examples illustrate the physical case where the measurements are acquired at equidistant view angles, but each measurement is associated with some independent uncertainty. Note that the "small example" has 90 view angles and standard deviation $\delta = 1.2^\circ$, whereas the "larger example" has 270 view angles with $\delta = 0.4^\circ$. This is to ensure that the view angles are unlikely to switch positions relative to each other from the added uncertainty. In our numerical tests, we generate the measurements according to

$$\mathbf{b} = \mathbf{A}(\boldsymbol{\theta}^{\text{machine}})\bar{\mathbf{x}} + \mathbf{e}, \tag{11}$$

where $\bar{\mathbf{x}}$ is either the Shepp-Logan or Grains phantom generated from AIR Tools II [21], and $\boldsymbol{\theta}^{\text{machine}}$ denotes the actual view angles, which is a realization of $\mathcal{N}(\boldsymbol{\theta}^{\text{equid}}, \delta^2\mathbf{I})$. Here $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, where $\sigma = 0.005\|\mathbf{A}(\boldsymbol{\theta}^{\text{machine}})\bar{\mathbf{x}}\|_2/\sqrt{m}$. We solve the TV minimization problem using the Chambolle-Pock algorithm in [22] and stopping when the relative change in the objective function is below 10^{-6}. In our method, we set the maximum iteration number $K = 10$ and the number of samples $N_{\text{samp}} = 5000$.

We compare our results with the non-negative TV reconstruction that does not take the uncertainty in the view angles into account, i.e.,

$$\mathbf{x}_{\mathrm{TV}} = \arg\min_{\mathbf{x}\geq 0} \frac{1}{2\sigma^2}\|(\mathbf{b} - \mathbf{A}(\hat{\boldsymbol{\theta}})\,\mathbf{x})\|_2^2 + \lambda \mathrm{TV}(\mathbf{x}), \tag{12}$$

where $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{\mathrm{equid}}$. In addition, we also show the results from the non-negative TV reconstruction with the actual view angles $\boldsymbol{\theta}^{\mathrm{machine}}$, which would be the best-case scenario:

$$\mathbf{x}_{\text{TV-opt}} = \arg\min_{\mathbf{x}\geq 0} \frac{1}{2\sigma^2}\|(\mathbf{b} - \mathbf{A}(\boldsymbol{\theta}^{\mathrm{machine}})\,\mathbf{x})\|_2^2 + \lambda \mathrm{TV}(\mathbf{x}). \tag{13}$$

3.1 The Small Example

In Fig. 2 we show the expected view angles $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{\mathrm{equid}}$ and 3 realizations of $\pi_{\mathrm{angles}}(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}^{\mathrm{equid}}, \delta^2\mathbf{I})$, i.e. 3 examples of $\boldsymbol{\theta}^{\mathrm{machine}}$. The realizations are used to generate noisy sinograms according to (11). We compare the reconstructions $\mathbf{x}_{\mathrm{STV}}$ from our method (Algorithm 1) with the ones obtained by solving (12) and (13). In the left column of Fig. 3 we plot the relative error $\frac{\|\mathbf{x}-\bar{\mathbf{x}}\|_2}{\|\bar{\mathbf{x}}\|_2}$ of the three methods with the regularization parameter λ varying from 10^{-6} to 10^{-2}. We can see that except for large λ, where the influence of the data-fitting term becomes weak, the reconstructions from our method has lower relative errors compared to $\mathbf{x}_{\mathrm{TV}}$ from (12). With the optimal λ choice, which gives the smallest relative error, the improvement by our method is significant. It shows the importance of taking the uncertainty in the view angles into account. In the right column of Fig. 3 we numerically show the convergence of the relative errors in our method.

In order to visually compare the reconstructions from these three methods, in Fig. 4 we show the reconstruction results for the same λ values, which corresponds to the optimal choice in (13). It is clear that our method can effectively reduce the artifacts due to the uncertain view angles.

For this small example, in our method we can also compute the full covariance $\mathbf{C}_{\boldsymbol{\eta}\,|\,\mathbf{x}} \in \mathbb{R}^{m\times m}$ instead of using the approximation $\hat{\mathbf{C}}_{\boldsymbol{\eta}\,|\,\mathbf{x}}$ introduced in Sect. 2.2. Since the relative errors by using the full covariance are almost identical to using the approximation, we do not show them here.

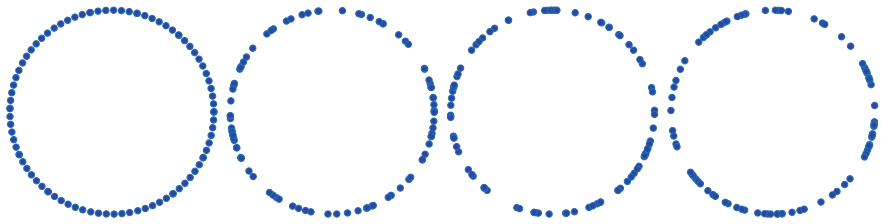


Fig. 2. For the small experiment in Table 1. From left to right: The expected view angles $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{\mathrm{equid}}$ and three realizations of $\pi_{\mathrm{angles}}(\boldsymbol{\theta})$.

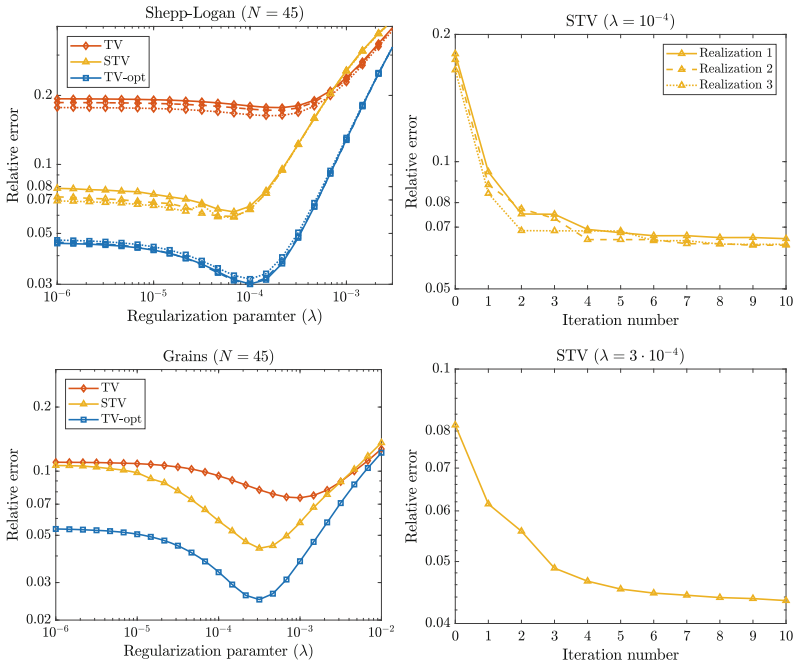


Fig. 3. For $N = 45$ in Table 1. Top: Shepp-Logan. Bottom: Grains. Left: Relative error vs. regularization parameter. Right: Relative error vs. iteration number in Algorithm 1.

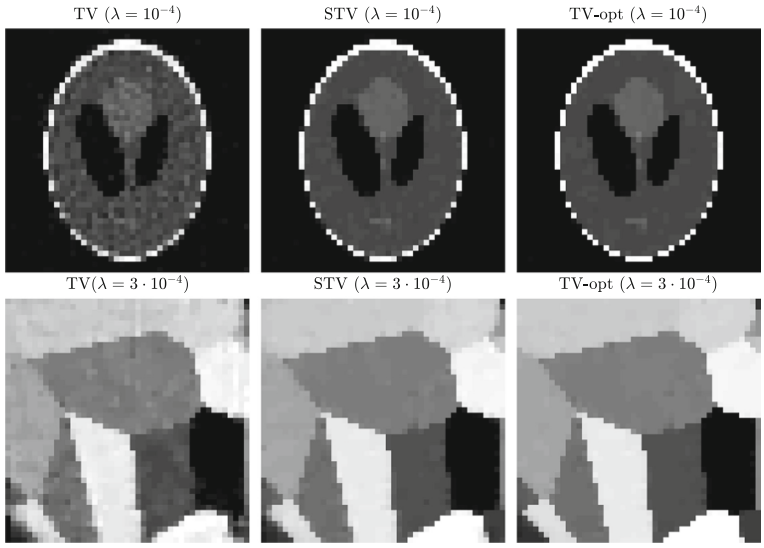


Fig. 4. Reconstructions for $N = 45$ in Table 1. Top: Shepp-Logan. Bottom: Grains.

3.2 The Larger Example

We also compute a larger example according to the parameters in Table 1. In Fig. 5 we show a zoomed part of the expected view angles $\hat{\boldsymbol{\theta}}$ as well as a realization of $\pi_{\text{angles}}(\boldsymbol{\theta})$ that is used to generate the data. In Fig. 6 we show the plots of the relative errors with different choices of λ and along the iterations in our method. In this case the difference between $\mathbf{x}_{\text{TV}}$ from solving (12) and $\mathbf{x}_{\text{TV}-\text{opt}}$ from solving (13) is not as big as in the small example, and the main reason is that the variance δ^2 in the view angles is much smaller. However, we can still clearly see that our method improves the reconstruction quality in terms of relative error. To compare the reconstruction visually, in Fig. 7 we give the reconstruction results from three methods.

Fig. 5. For the larger experiment in Table 1. Left: the expected view angles $\hat{\boldsymbol{\theta}}$. Right: a realization of $\pi_{\text{angles}}(\boldsymbol{\theta})$. Here we zoomed in on a part of the view angles.

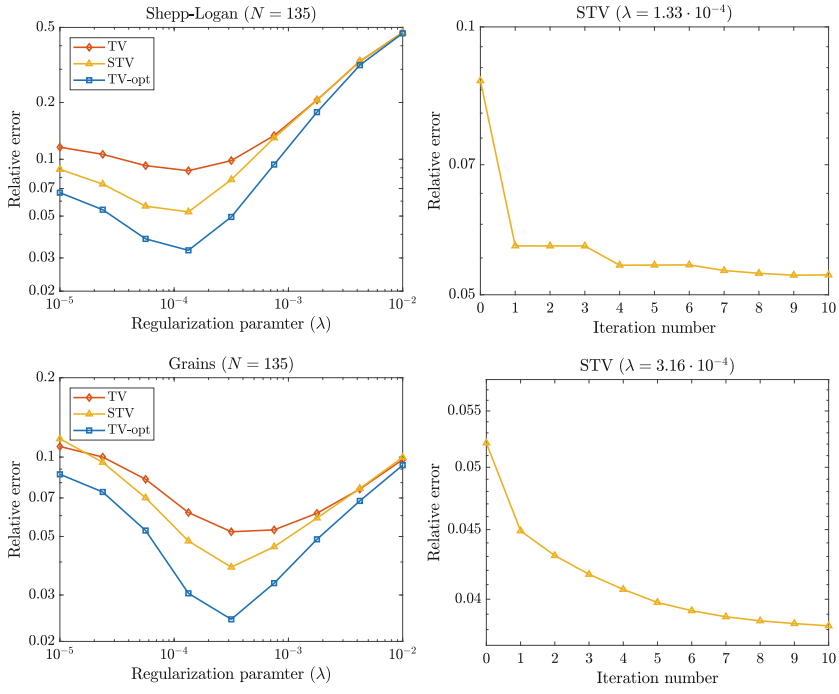


Fig. 6. For the experiment with $N = 135$ in Table 1. Top: Shepp-Logan. Bottom: Grains. Left: Relative error vs. regularization parameter. Right: Relative error vs. iteration number in Algorithm 1.

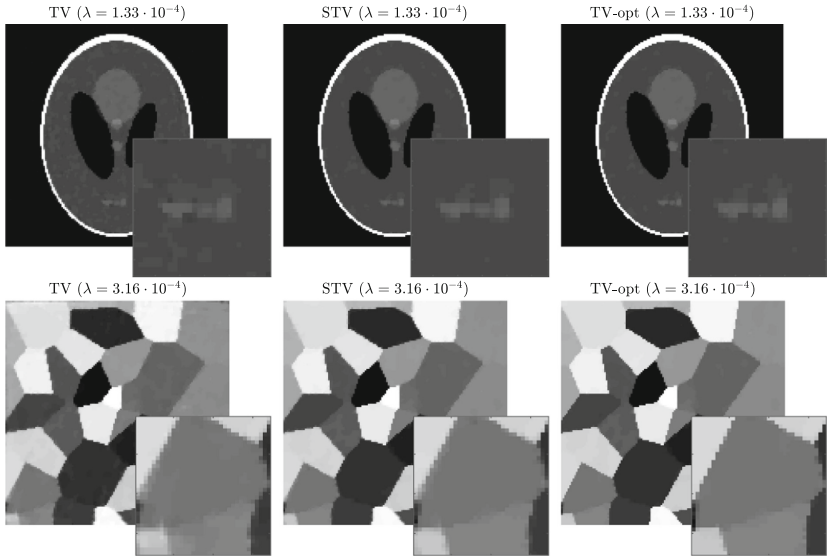


Fig. 7. Reconstructions for $N = 135$ in1 Table 1. Top: Shepp-Logan. Bottom: Grains.

Based on our tests, if we increase the image size n and keep the same number of measurements, the quality of the reconstruction by our method gets closer to the one from (12). The reason is that the reconstruction problem becomes more ill-posed and therefore more difficult to deal with. In this case, we would need a better initial guess on $\hat{\mathbf{L}}^{[0]}_{\boldsymbol{\nu}\,|\,\mathbf{x}}$ and $\boldsymbol{\mu}^{[0]}_{\boldsymbol{\nu}\,|\,\mathbf{x}}$ in order to obtain a good estimate of $\bar{\mathbf{x}}$. Another idea would be to update the estimate of $\hat{\boldsymbol{\theta}}$ and $\pi_{\text{angles}}(\cdot)$ in each iteration. We leave this to the future study.

4 Conclusion

We proposed a new iterative algorithm for the uncertain angle CT problem. The method models the uncertainty of the view angles in the likelihood function. We showed numerically that combining this likelihood with a strong prior such as total variation can significantly improve the relative error and visual quality of reconstructions. Furthermore, we showed a method for approximating the likelihood by a block-diagonal approximation of the covariance, which leads to an algorithm that can run on large-scale CT problems.

References

1. Natterer, F.: The Mathematics of Computerized Tomography. Wiley, Chicago (1986)
2. Benning, M., Burger, M.: Modern regularization methods for inverse problems. Acta Numerica **27**, 1–111 (2018)

3. Riis, N.A.B., Frøsig, J., Dong, Y., Hansen, P.C.: Limited-data X-ray CT for underwater pipeline inspection. Inverse Prob. **34**(3), 034002 (2018)
4. Vandeghinste, B., et al.: Iterative CT reconstruction using shearlet-based regularization. IEEE Tran. Nucl. Sci. **60**(5), 3305–3317 (2013)
5. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**(1–4), 259–268 (1992)
6. Ferrucci, M., Leach, R.K., Giusca, C., Carmignato, S., Dewulf, W.: Towards geometrical calibration of X-ray computed tomography systems-a review. Measur. Sci. Technol. **26**(9), 092003 (2015)
7. Mallick, S.P., Agarwal, S., Kriegman, D.J., Belongie, S.J., Carragher, B., Potter, C.S.: Structure and view estimation for tomographic reconstruction: a bayesian approach. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2(1), pp. 2253–2260 (2006). 1641029
8. Fang, Y., Murugappan, S., Ramani, K.: Estimating view parameters from random projections for tomography using spherical MDS. BMC Med. Imaging **10**(1), 12 (2010)
9. Basu, S., Bresler, Y.: Uniqueness of tomography with unknown view angles. IEEE Trans. Image Process. **9**(6), 1094–1106 (2000)
10. Korolev, Y., Lellmann, J.: Image reconstruction with imperfect forward models and applications in deblurring. SIAM J. Imaging Sci. **11**(1), 197–218 (2018)
11. Kaipio, J., Somersalo, E.: Statistical and Computational Inverse Problems. Springer, New York (2005). https://doi.org/10.1007/b138659
12. Madsen, R.B., Hansen, T.M.: Estimation and accounting for the modeling error in probabilistic linearized amplitude variation with offset inversion. Geophysics **83**(2), N15–N30 (2018)
13. Hansen, T.M., Cordua, K.S., Holm Jacobsen, B., Mosegaard, K.: Accounting for imperfect forward modeling in geophysical inverse problems exemplified for crosshole tomography. Geophysics **79**(3), H1–H21 (2014)
14. Calvetti, D., Dunlop, M., Somersalo, E., Stuart, A.: Iterative updating of model error for bayesian inversion. Inverse Prob. **34**(2), 025008 (2018)
15. Kolehmainen, V., Tarvainen, T., Arridge, S.R., Kaipio, J.P.: Marginalization of uninteresting distributed parameters in inverse problems - application to diffuse optical tomography. Int. J. Uncertainty Quantification **1**(1), 1–17 (2011)
16. Kennedy, M.C., O'Hagan, A.: Bayesian calibration of computer models. J. Roy. Stat. Soc. Ser. B: Stat. Methodol. **63**(3), 425–450 (2001)
17. Kaipio, J., Somersalo, E.: Statistical inverse problems: discretization, model reduction and inverse crimes. J. Comput. Appl. Math. **198**(2), 493–504 (2007)
18. Nissinen, A., Heikkinen, L.M., Kaipio, J.P.: The bayesian approximation error approach for electrical impedance tomography - experimental results. Measur. Sci. Technol. **19**(1), 015501 (2008)
19. van Aarle, W., et al.: Fast and flexible X-ray tomography using the ASTRA toolbox. Opt. Express **24**(22), 25129–25147 (2016). www.astra-toolbox.com
20. van den Berg, E., Friedlander, M.P.: Spot – A Linear-Operator Toolbox: MATLAB software. www.cs.ubc.ca/labs/scl/spot/
21. Hansen, P.C., Jørgensen, J.S.: AIR tools II: algebraic iterative reconstruction methods, improved implementation. Numer. Algorithms **79**(1), 107–137 (2017)
22. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. **40**(1), 120–145 (2011)

APPENDIX B

Computed Tomography Reconstruction with Uncertain View Angles by Iteratively Updated Model Discrepancy

Journal of Mathematical Imaging and Vision, Special issue on Scale Space and Variational Methods in Computer Vision. pp 1-11.
First published online 04 July 2020.
DOI: https://doi.org/10.1007/s10851-020-00972-7

Riis, N. A. B., Dong, Y., & Hansen, P. C.

Journal of Mathematical Imaging and Vision
https://doi.org/10.1007/s10851-020-00972-7

Computed Tomography Reconstruction with Uncertain View Angles by Iteratively Updated Model Discrepancy

Nicolai André Brogaard Riis[1] · **Yiqiu Dong**[1] · **Per Christian Hansen**[1]

Received: 15 October 2019 / Accepted: 8 June 2020

Abstract
We propose a new model and a corresponding iterative algorithm for Computed Tomography (CT) when the view angles are uncertain. The uncertainty is described by an additive model discrepancy term which is included in the data fidelity term of a total variation regularized variational model. We approximate the model discrepancy with a Gaussian distribution. Our iterative algorithm alternates between updating the CT reconstruction and parameters of the model discrepancy. By assuming that the uncertainties in the view angles are independent we achieve a covariance matrix structure that we can take advantage of in a stochastic primal dual method to greatly reduce the computational work compared to classical primal dual methods. Using simulations with 2D problems we demonstrate that our method is able to reduce the reconstruction error and improve the visual quality, compared to methods that ignore the uncertainties in the angles.

Keywords Computed Tomography · Uncertain view angles · Model error · Model discrepancy

Mathematics Subject Classification 65F22 · 65K10

1 Introduction

In this paper we consider Computed Tomography (CT) reconstruction where the view angles are only known approximately. This uncertainty may arise from several sources, e.g., inaccuracies in the physical set-up and inexact estimates from a calibration procedure. The goal is to reconstruct a CT image in such a way that we take into account both measurement noise and the uncertainty associated with the view angles.

We formulate the *CT reconstruction problem with uncertain view angles* as estimating the unknown attenuation coefficient image, represented by the vector $\mathbf{x} \in \mathbb{R}^n$, from a measured and noisy sinogram represented by the vector $\mathbf{b} \in \mathbb{R}^m$. We use the model

$$\mathbf{b} = \mathbf{R}(\boldsymbol{\theta})\,\mathbf{x} + \mathbf{e}, \quad \boldsymbol{\theta} \sim \pi_{\text{angles}}(\cdot), \quad \mathbf{e} \sim \pi_{\text{noise}}(\cdot), \tag{1}$$

where the uncertainty in the view angles and the measured sinogram are characterized by the probability distributions π_{angles} and π_{noise}, respectively.

The CT forward model (the forward projection) is represented by the matrix $\mathbf{R}(\boldsymbol{\theta}) \in \mathbb{R}^{m\times n}$, which is the discrete approximation of the Radon transform $\mathcal{R}$ explicitly parameterized by the view angles $\boldsymbol{\theta} \in \mathbb{R}^q$. Specifically, if f is the continuous representation of the attenuation image, the discretization satisfies

$$(\mathbf{R}(\theta_i)\mathbf{x})_l \approx (\mathcal{R}f)(\theta_i, \mathrm{s}_l) = \int_{\mathbb{R}} f(\mathrm{s}_l \overrightarrow{\mathbf{v}}(\theta_i) + t\,\overrightarrow{\mathbf{v}}^{\perp}(\theta_i))\,\mathrm{d}t, \tag{2}$$

where s_l with $l = 1, \ldots, p$ is the position of the lth pixel on the detector, θ_i with $i = 1, \ldots, q$ is the view angle, and hence $m = qp$. Moreover, $\overrightarrow{\mathbf{v}}(\theta) = (\cos\theta, \sin\theta)$ and $\overrightarrow{\mathbf{v}}^{\perp}(\theta) = \overrightarrow{\mathbf{v}}(\theta + \pi/2)$. For more details on the physical and mathematical models of CT see, e.g., [5,17].

This work was supported by a Villum Investigator Grant (No. 25893) from The Villum Foundation and by the National Natural Science Foundation of China via Grant 11701388.

✉ Nicolai André Brogaard Riis
nabr@dtu.dk

Yiqiu Dong
yido@dtu.dk

Per Christian Hansen
pcha@dtu.dk

[1] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark

Published online: 04 July 2020

We emphasize that the goal in this work is to reconstruct the CT image $\mathbf{x}$ from a measured sinogram $\mathbf{b}$ according to the model (1) with uncertain view angles $\boldsymbol{\theta}$ and noise $\mathbf{e}$. We consider $\boldsymbol{\theta}$ and $\mathbf{e}$ as nuisance or uninteresting parameters, and they are only taken into account when reconstructing $\mathbf{x}$ without being explicitly estimated.

While our work focuses on 2D CT reconstruction, our methodology can also be applied to 3D CT and other applications, because there are no restrictions on the matrix in the forward model. We emphasize the computational advantage of having a block representation of the covariance matrix for the model discrepancy introduced in the next section (cf. Sect. 3).

1.1 Related Work

Variational methods have been proposed for CT reconstruction problems that, in addition to a data fidelity term which incorporates the noise model, explicitly incorporate prior information via a regularization term, see the survey [3]. Different regularization techniques can be applied, for example, Tikhonov regularization [23,29] or total variation (TV) regularization [26,27]. But these methods do not take parametric uncertainty in the model into account. Therefore, good performance of these methods is not guaranteed if the view angles are uncertain.

Reconstruction methods that aim to estimate the view angles $\boldsymbol{\theta}$ in addition to the CT image $\mathbf{x}$ using only the measured sinogram without reference objects or instruments are proposed in [2,10,18,22]. These methods can be categorized into two groups: estimating view angles directly from projection data and then estimating the CT image, and simultaneously estimating view angles and CT image. In the former case an inaccurate angle estimation could lead to an unsatisfactory reconstruction due to error propagation. Algorithms for the latter case (such as the Bayesian sampling-based method [22]) can effectively avoid error propagation, but they are limited by large computing times which makes them unfeasible for large-scale problems.

There are also a few methods for characterizing and reducing model errors in general inverse problems, see [6,11,12,15,16,21]. Most of these methods are based on Bayesian sampling-based methods which may also suffer from large computing times, except in the special case of a Gaussian prior for the solution where a closed-form expression exist.

1.2 Our Contribution

We propose a new model and algorithm for CT reconstruction when the view angles are uncertain. Our model incorporates the uncertainty of the view angles in the data fidelity term of a variational method. This provides a CT reconstruction where the uncertainty is accounted for and we avoid error propagation without the extra cost of simultaneously estimating the view angles and CT image.

Compared to our preliminary work in [25] the main contribution in this work is the formulation and utilization of block-structure of the computational problem. This allows us to utilize a stochastic primal dual hybrid gradient method [9] to significantly reduce the computational work and allows the method to solve large-scale problems. Moreover, we use a numerically stable and efficient approach to factorize and invert the covariance matrix required in our data fidelity term.

2 Iteratively Updated Model Discrepancy

In this section we summarize our previous work in [25] where we reformulate (1) by fixing the forward model with an estimate $\hat{\boldsymbol{\theta}}$ of the view angles. This is done to avoid dealing with a distribution of models $\mathbf{R}(\boldsymbol{\theta})$, $\boldsymbol{\theta} \sim \pi_{\text{angles}}(\cdot)$ in the CT reconstruction. The angles $\hat{\boldsymbol{\theta}}$ are set to the nominal angles of the scanner, and we obtain

$$\mathbf{b} = \mathbf{R}(\hat{\boldsymbol{\theta}})\,\mathbf{x} + \boldsymbol{\eta} + \mathbf{e},\ \boldsymbol{\eta} \sim \pi_{\text{discrep}}(\cdot),\ \mathbf{e} \sim \pi_{\text{noise}}(\cdot), \tag{3}$$

where $\boldsymbol{\eta}$ acquires its push-forward distribution from

$$\boldsymbol{\eta} = \eta(\boldsymbol{\theta}, \mathbf{x}) = \mathbf{R}(\boldsymbol{\theta})\,\mathbf{x} - \mathbf{R}(\hat{\boldsymbol{\theta}})\,\mathbf{x}, \tag{4}$$

and we call $\boldsymbol{\eta}$ the *model discrepancy term*. The advantage of doing so is that the uncertainty in the forward model is moved into $\boldsymbol{\eta}$ and the CT model $\mathbf{R}(\hat{\boldsymbol{\theta}})$ is fixed. If we ignore the fact that $\boldsymbol{\eta}$ depends on $\mathbf{x}$ and consider $\boldsymbol{\eta}$ as independent additive noise, then (3) becomes a standard CT reconstruction model with the fixed view angles $\hat{\boldsymbol{\theta}}$.

Representing the model uncertainties as an additive model discrepancy term was first proposed in [14], where $\boldsymbol{\eta}$ was assumed a Gaussian process and used for calibration from a physical model to a computational model. In [12,13] a similar idea was applied to Bayesian inverse problems and referred to as the approximation error approach (AEA), in which $\boldsymbol{\eta}$ denotes the model discrepancy. This approach is used successfully in a number of different imaging applications such as fine-to-coarse mesh approximation in diffuse optical tomography [1], unknown domain boundaries in electric impedance tomography [20], unknown scattering in both diffuse optical tomography and quantitative photoacoustic tomography [15,24], truncation errors in magnetic particle imaging [4] and uncertain sound speed in photoacoustic tomography [28].

Inspired by AEA, we derive the likelihood for (3) by marginalizing both $\boldsymbol{\eta}$ and $\mathbf{e}$. Let $\boldsymbol{\nu} = \boldsymbol{\eta} + \mathbf{e}$, then

$$\begin{aligned}\pi(\mathbf{b}|\mathbf{x}) &= \int_{\mathbb{R}^m} \pi(\mathbf{b}, \boldsymbol{\nu}|\mathbf{x})\mathrm{d}\boldsymbol{\nu} \\ &= \int_{\mathbb{R}^m} \pi(\mathbf{b}|\mathbf{x}, \boldsymbol{\nu})\pi(\boldsymbol{\nu}|\mathbf{x})\mathrm{d}\boldsymbol{\nu} \\ &= \pi_{\boldsymbol{\nu}|\mathbf{x}}(\mathbf{b} - \mathbf{R}(\hat{\boldsymbol{\theta}})\,\mathbf{x}|\mathbf{x}). \end{aligned} \tag{5}$$

This formulation raises two main issues:

(1) The distribution $\pi_{\boldsymbol{\nu}|\mathbf{x}}$ can be rather complicated and may not have a closed-form expression, so the evaluation of the likelihood becomes difficult.
(2) The distribution $\pi_{\boldsymbol{\nu}|\mathbf{x}}$ depends on the unknown $\mathbf{x}$.

To deal with the former issue, we introduce a Gaussian approximation to $\pi_{\boldsymbol{\nu}|\mathbf{x}}$, i.e., we simplify the model. The latter issue, on the other hand, is dealt with from an algorithmic point of view by using a computational method that alternates between updating $\pi_{\boldsymbol{\nu}|\mathbf{x}}$ and $\mathbf{x}$.

2.1 Gaussian Approximation

The distribution $\pi_{\boldsymbol{\nu}|\mathbf{x}}$ may not have a closed-form expression, so we approximate it by a simple distribution. Experiments suggest that Gaussian approximations are useful in many applications related to model discrepancies [1,4,6,11–13,15,19,20,24,28]. In this work, we therefore assume $\boldsymbol{\eta}|\mathbf{x}$ follows a Gaussian distribution, i.e.,

$$\boldsymbol{\eta}|\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\eta}|\mathbf{x}}, \mathbf{C}_{\boldsymbol{\eta}|\mathbf{x}}) \tag{6}$$

with mean $\boldsymbol{\mu}_{\boldsymbol{\eta}|\mathbf{x}}$ and covariance $\mathbf{C}_{\boldsymbol{\eta}|\mathbf{x}}$, both depending on $\mathbf{x}$. Moreover, we assume

$$\mathbf{e} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{e}}, \mathbf{C}_{\mathbf{e}}), \tag{7}$$

with mean $\boldsymbol{\mu}_{\mathbf{e}}$ and covariance $\mathbf{C}_{\mathbf{e}}$, which are independent on $\mathbf{x}$. Following these two assumptions, we obtain the distribution of $\boldsymbol{\nu}|\mathbf{x}$

$$\pi_{\boldsymbol{\nu}|\mathbf{x}} = \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\eta}|\mathbf{x}} + \boldsymbol{\mu}_{\mathbf{e}}, \mathbf{C}_{\boldsymbol{\eta}|\mathbf{x}} + \mathbf{C}_{\mathbf{e}}). \tag{8}$$

Then, the likelihood (5) admits a closed-form expression. By taking the negative logarithm we get

$$\begin{aligned}-\log \pi(\mathbf{b}|\mathbf{x}) &\propto \tfrac{1}{2}\|\mathbf{b} - \mathbf{R}(\hat{\boldsymbol{\theta}})\,\mathbf{x} - \boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x}}\|^2_{\mathbf{C}^{-1}_{\boldsymbol{\nu}|\mathbf{x}}} \\ &= \tfrac{1}{2}\|\mathbf{L}_{\boldsymbol{\nu}|\mathbf{x}}(\mathbf{b} - \mathbf{R}(\hat{\boldsymbol{\theta}})\,\mathbf{x} - \boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x}})\|_2^2, \end{aligned} \tag{9}$$

where $\boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x}} = \boldsymbol{\mu}_{\boldsymbol{\eta}|\mathbf{x}} + \boldsymbol{\mu}_{\mathbf{e}}$, $\mathbf{C}_{\boldsymbol{\nu}|\mathbf{x}} = \mathbf{C}_{\boldsymbol{\eta}|\mathbf{x}} + \mathbf{C}_{\mathbf{e}}$, and $\mathbf{L}^T_{\boldsymbol{\nu}|\mathbf{x}}\mathbf{L}_{\boldsymbol{\nu}|\mathbf{x}} = \mathbf{C}^{-1}_{\boldsymbol{\nu}|\mathbf{x}}$ is the Cholesky factorization of the inverse of the combined covariance matrix.

2.2 Alternate Updates

Since the distribution $\pi_{\boldsymbol{\nu}|\mathbf{x}}$ depends on the unknown $\mathbf{x}$, the question is how the mean $\boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x}}$ and the covariance $\mathbf{C}_{\boldsymbol{\nu}|\mathbf{x}}$ of the combined uncertainties can be determined.

Considering a given reconstruction $\hat{\mathbf{x}}$, one can generate samples of $\boldsymbol{\eta}|\mathbf{x}$ with $\mathbf{x} = \hat{\mathbf{x}}$ by drawing samples $\boldsymbol{\theta}^s$ with $s = 1, \ldots, S$ following the distribution π_{angles}, and evaluate the model discrepancy term by

$$\boldsymbol{\eta}^s_{\hat{\mathbf{x}}} = \mathbf{R}(\boldsymbol{\theta}^s)\,\hat{\mathbf{x}} - \mathbf{R}(\hat{\boldsymbol{\theta}})\,\hat{\mathbf{x}}, \quad s = 1, \ldots, S. \tag{10}$$

Then, the sample mean and sample covariance can be calculated by

$$\boldsymbol{\mu}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}} = \frac{1}{S}\sum_{s=1}^{S} \boldsymbol{\eta}^s_{\hat{\mathbf{x}}} \tag{11}$$

and

$$\mathbf{C}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}} = \frac{1}{S-1}\sum_{s=1}^{S} (\boldsymbol{\eta}^s_{\hat{\mathbf{x}}} - \boldsymbol{\mu}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}})(\boldsymbol{\eta}^s_{\hat{\mathbf{x}}} - \boldsymbol{\mu}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}})^T. \tag{12}$$

In (9), we can then use the sample mean and the sample covariance to approximate $\boldsymbol{\mu}_{\boldsymbol{\eta}|\mathbf{x}}$ and $\mathbf{C}_{\boldsymbol{\eta}|\mathbf{x}}$, respectively.

Good estimation of the model discrepancy term strongly relies on a reconstruction that resembles the ground truth. Variational methods with TV regularization have shown good performance in large-scale CT reconstruction [27]. Adding a TV regularization term to the negative log-likelihood in (9) we obtain the variational model

$$\min_{\mathbf{x}\geq 0} \tfrac{1}{2}\|\mathbf{L}_{\boldsymbol{\nu}|\mathbf{x}}(\mathbf{b} - \mathbf{R}(\hat{\boldsymbol{\theta}})\,\mathbf{x} - \boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x}})\|_2^2 + \lambda \mathrm{TV}(\mathbf{x}), \tag{13}$$

where $\lambda > 0$ denotes the regularization parameter, and the TV term is defined as

$$\mathrm{TV}(\mathbf{x}) = \|\nabla \mathbf{x}\|_{2,1} \equiv \sum_{i=1}^{n} \|[\nabla \mathbf{x}]_i\|_2\,, \tag{14}$$

in which $[\nabla \mathbf{x}]_i$ denotes the discrete gradient of $\mathbf{x}$ at the ith pixel, computed via a forward difference scheme with reflexive boundary conditions. The non-negativity constraint represents the fact that the attenuation coefficients $\mathbf{x}$ cannot be negative. Since $\mathbf{L}_{\boldsymbol{\nu}|\mathbf{x}}$ depends on the unknown $\mathbf{x}$, the objective function in (13) is non-convex, which means that a solution relies on the initialization as well as the numerical algorithm.

According to (13), the reconstruction $\mathbf{x}$ also depends on $\boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x}}$ and $\mathbf{C}_{\boldsymbol{\nu}|\mathbf{x}}$. To further strengthen the relation between $\mathbf{x}$ and $(\boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x}}, \mathbf{C}_{\boldsymbol{\nu}|\mathbf{x}})$ we introduce an alternately updating scheme **MD-TV** shown in Algorithm 1.

Algorithm 1: Iteratively Updated Model Discrepancy with TV (MD-TV)

Input: $\mathbf{b}$, λ, $\hat{\boldsymbol{\theta}}$, π_{angles}, $\boldsymbol{\mu}_{\mathbf{e}}$, $\mathbf{C}_{\mathbf{e}}$. Initialize $\mathbf{L}^0_{\boldsymbol{\nu}|\mathbf{x}}$, $\boldsymbol{\mu}^0_{\boldsymbol{\nu}|\mathbf{x}}$.
Output: $\mathbf{x}$

1: **for** $k = 1, 2, \ldots, K$
2: $\quad \mathbf{x}^k = \arg\min_{\mathbf{x}\geq 0} \frac{1}{2}\|\mathbf{L}^{k-1}_{\boldsymbol{\nu}|\mathbf{x}}(\mathbf{b} - \mathbf{R}(\hat{\boldsymbol{\theta}})\,\mathbf{x} - \boldsymbol{\mu}^{k-1}_{\boldsymbol{\nu}|\mathbf{x}})\|_2^2 + \lambda \mathrm{TV}(\mathbf{x})$
3: $\quad$ **for** $s = 1, 2, \ldots, S$
4: $\quad\quad \boldsymbol{\eta}^s_{\mathbf{x}^k} = \mathbf{R}(\boldsymbol{\theta}^s)\,\mathbf{x}^k - \mathbf{R}(\hat{\boldsymbol{\theta}})\,\mathbf{x}^k$ with $\boldsymbol{\theta}^s \sim \pi_{\text{angles}}(\cdot)$
5: $\quad$ **end**
6: $\quad \boldsymbol{\mu}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\mathbf{x}^k} = \frac{1}{S}\sum_{s=1}^S \boldsymbol{\eta}^s_{\mathbf{x}^k}$
7: $\quad \mathbf{C}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\mathbf{x}^k} = \frac{1}{S-1}\sum_{s=1}^S (\boldsymbol{\eta}^s_{\mathbf{x}^k} - \boldsymbol{\mu}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\mathbf{x}^k})(\boldsymbol{\eta}^s_{\mathbf{x}^k} - \boldsymbol{\mu}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\mathbf{x}^k})^T$
8: $\quad \boldsymbol{\mu}^k_{\boldsymbol{\nu}|\mathbf{x}} = \boldsymbol{\mu}_{\mathbf{e}} + \boldsymbol{\mu}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\mathbf{x}^k}$
9: $\quad \mathbf{L}^k_{\boldsymbol{\nu}|\mathbf{x}} = \mathrm{chol}\big((\mathbf{C}_{\mathbf{e}} + \mathbf{C}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\mathbf{x}^k})^{-1}\big)$
10: **end**

Here, $\mathbf{L} = \mathrm{chol}(\mathbf{C}^{-1})$ gives the Cholesky factor of the inverse covariance $\mathbf{C}^{-1}$, i.e., $\mathbf{L}^T\mathbf{L} = \mathbf{C}^{-1}$. Note that although the minimization problem for obtaining $\mathbf{x}$ is convex now, it does not guarantee that the overall algorithm converges.

3 Block-Wise Representation

In each iteration of Algorithm 1 we require S forward projections to generate samples (10) of $\boldsymbol{\eta}|\mathbf{x}$ in step 4, which can dominate the cost of each iteration for large S. Moreover, even for small S the algorithm can be of limited use in solving large-scale CT problems in its "naive" form due to the amount of work in solving the TV optimization problem in step 2 and inverting and factorizing the covariance matrix $\mathbf{C}_{\mathbf{e}} + \mathbf{C}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}} \in \mathbb{R}^{m\times m}$ in step 9.

In our previous work [25], the latter issue was partly addressed by a block-diagonal representation of the covariance matrix. This greatly reduces the computational burden of the inversion and factorization and it reduces the cost of multiplication with the factorized matrix when solving the TV optimization problem. In this work, we further extend these ideas by reformulating the entire variational model using the block diagonal representation. Our new formulation can thus benefit from algorithms that utilize this structure as well as parallelization in the block-wise sampling, inversion and factorization of the covariance matrix.

3.1 Block Covariance Matrix Representation

The covariance matrix $\mathbf{C}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}}$ is block-diagonal when the view angles are independent. The sampled matrix in (12) may not be block diagonal due to the finite sampling.

Our experience is that the norm of the off-diagonal blocks decreases as the number of samples S increases. Therefore, we replace the sample covariance matrix $\mathbf{C}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}} \in \mathbb{R}^{m\times m}$ by a block diagonal matrix $\widetilde{\mathbf{C}}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}}$ given by

$$\widetilde{\mathbf{C}}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}} = \begin{bmatrix} \mathbf{C}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}},1} & & \\ & \ddots & \\ & & \mathbf{C}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}},q} \end{bmatrix}, \tag{15}$$

where $\mathbf{C}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}},i} \in \mathbb{R}^{p\times p}$ are the block diagonal parts of $\mathbf{C}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}}$. Then, assuming the Gaussian measurement noise is independent identically distributed, i.e., $\mathbf{C}_{\mathbf{e}} = \sigma^2\mathbf{I}$, we can compute the Cholesky factor of the inverse covariance matrix $(\mathbf{C}_{\mathbf{e}} + \widetilde{\mathbf{C}}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}})^{-1}$ block-wise as follows

$$\widetilde{\mathbf{L}}^{\text{sample}}_{\boldsymbol{\nu}|\mathbf{x}=\hat{\mathbf{x}}} = \begin{bmatrix} \mathrm{chol}\big((\sigma^2\mathbf{I}_p + \widetilde{\mathbf{C}}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}},1})^{-1}\big) & & \\ & \ddots & \\ & & \mathrm{chol}\big((\sigma^2\mathbf{I}_p + \widetilde{\mathbf{C}}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}},q})^{-1}\big) \end{bmatrix}. \tag{16}$$

These aspects are illustrated in Fig. 1, which shows that the block-diagonal representation captures the structure of the actual covariance matrix with fewer samples compared to the full version.

Furthermore, to achieve a computationally efficient and numerically stable inversion and factorization, we note that we can interchange the inversion and factorization and utilize that the covariance matrix can be written as a low-rank update of a scaled identity matrix,

$$\widetilde{\mathbf{C}}^{\text{sample}}_{\boldsymbol{\nu}|\mathbf{x}=\hat{\mathbf{x}}} = \sigma^2\mathbf{I}_m + \sum_{s=1}^S \mathbf{w}_s\mathbf{w}_s^T, \tag{17}$$

where $\mathbf{w}_s = \frac{\boldsymbol{\eta}^s_{\hat{\mathbf{x}}} - \boldsymbol{\mu}^{\text{sample}}_{\boldsymbol{\eta}|\mathbf{x}=\hat{\mathbf{x}}}}{\sqrt{S-1}}$. Therefore, the computation of the inverse Cholesky factors that constitute the blocks in $\widetilde{\mathbf{L}}^{\text{sample}}_{\boldsymbol{\nu}|\mathbf{x}=\hat{\mathbf{x}}}$ merely consists of low-rank updates of $\sigma^{-2}\mathbf{I}_m$. Updating an inverse Cholesky factor is discussed in, e.g., [8]; this approach is numerically more stable than using the Sherman-Morrison formula for updating the inverse covariance matrix.

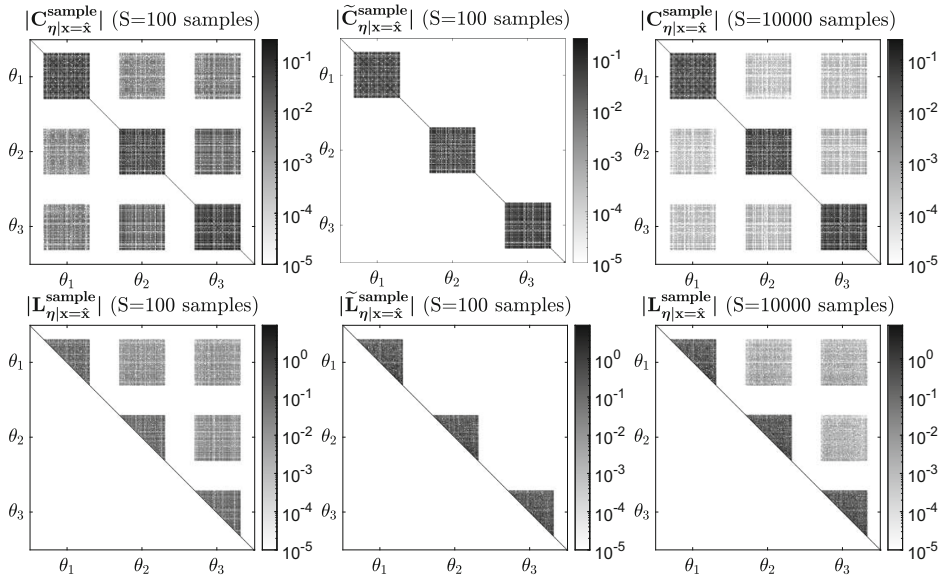


Fig. 1 These plots show the small top-left submatrix of the entire matrices. Each plot shows the absolute value of the matrix elements, and to visualize the zero-nonzero structure we use a logarithmic colormap. Left: the full sampled covariance matrix $\mathbf{C}^{\text{sample}}_{\eta|\mathbf{x}=\hat{\mathbf{x}}}$ and the Cholesky factor $\mathbf{L}^{\text{sample}}_{\eta|\mathbf{x}=\hat{\mathbf{x}}}$ of its inverse with $S = 100$ samples. Middle: the block-diagonal representation $\widetilde{\mathbf{C}}^{\text{sample}}_{\eta|\mathbf{x}=\hat{\mathbf{x}}}$ according to (15) and the Cholesky factor $\widetilde{\mathbf{L}}^{\text{sample}}_{\eta|\mathbf{x}=\hat{\mathbf{x}}}$ of its inverse with $S = 100$ samples. Right: full sampled covariance matrix and Cholesky factor of its inverse with $S = 10000$ samples. We see that the block-diagonal representation captures the structure of the actual covariance matrix with fewer samples compared to the full version

3.2 Block-Wise Variational Model

Given the block-wise approximation of the covariance matrix, the variational model in (13) can be formulated as

$$\min_{\mathbf{x}\geq\mathbf{0}} \frac{1}{2} \left\| \begin{bmatrix} \widetilde{\mathbf{L}}_{\boldsymbol{\nu}|\mathbf{x},1} & & \\ & \ddots & \\ & & \widetilde{\mathbf{L}}_{\boldsymbol{\nu}|\mathbf{x},q} \end{bmatrix} \left(\begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_q \end{bmatrix} - \begin{bmatrix} \mathbf{R}(\hat{\theta}_1) \\ \vdots \\ \mathbf{R}(\hat{\theta}_q) \end{bmatrix} \mathbf{x} - \begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x},1} \\ \vdots \\ \boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x},2} \end{bmatrix} \right) \right\|_2^2 + \lambda \mathrm{TV}(\mathbf{x}), \tag{18}$$

where we use the short-hand notation $\boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x},i}$ and $\widetilde{\mathbf{L}}_{\boldsymbol{\nu}|\mathbf{x},i}$ for the blocks of the sample mean (11) and Cholesky factor of the inverse sample covariance matrix (16).

We can conveniently utilize the block structure to arrive at a variational model using q data-fitting terms, i.e.,

$$\min_{\mathbf{x}\geq 0} \frac{1}{2} \sum_{i=1}^{q} \| \widetilde{\mathbf{L}}_{\boldsymbol{\nu}|\mathbf{x},i} (\mathbf{b}_i - \mathbf{R}(\hat{\theta}_i)\, \mathbf{x} - \boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x},i}) \|_2^2 + \lambda \mathrm{TV}(\mathbf{x}). \tag{19}$$

This formulation allows us to take the block structure into account via a stochastic optimization algorithm, namely, the stochastic primal dual hybrid gradient (SPDHG) method [9]. This greatly reduces the number of outer iterations and thus the overall computational work, compared to classical primal dual methods. The SPDHG algorithm solves the generic optimization problem

$$\min_{\mathbf{x}} \sum_{i=1}^{q+1} f_i(\mathbf{A}_i \mathbf{x}) + g(\mathbf{x}), \tag{20}$$

where for (19) we specifically have

$$f_{1,\dots,q}(\cdot) = \tfrac{1}{2}\|\cdot + \widetilde{\mathbf{L}}_{\boldsymbol{\nu}|\mathbf{x},i}(\mathbf{b}_i - \boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x},i})\|_2^2, \tag{21}$$

$$f_{q+1}(\cdot) = \lambda\,\|\cdot\|_{2,1}, \tag{22}$$

$$\mathbf{A}_{1,\dots,q} = -\widetilde{\mathbf{L}}_{\boldsymbol{\nu}|\mathbf{x},i}\mathbf{R}(\hat{\theta}_i), \tag{23}$$

$$\mathbf{A}_{q+1} = \nabla, \tag{24}$$

$$g(\cdot) = \begin{cases} 0 & \text{if } x_i \geq 0, \\ \infty & \text{if } x_i < 0. \end{cases} \tag{25}$$

The procedure for solving the generic problem (20) is shown in Algorithm 2, where we specially use the proximal operators:

$$\operatorname{prox}_g^{\tau}(\mathbf{x}) = \max(\mathbf{x}, \mathbf{0}), \tag{26}$$

$$\operatorname{prox}_{f_i^*}^{\omega}(\mathbf{y}) = \frac{1}{1+\omega}\left(\mathbf{y} + \omega\widetilde{\mathbf{L}}_{\boldsymbol{\nu}|\mathbf{x},i}(\mathbf{b}_i - \boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x},i})\right), \quad i = 1,\dots,q, \tag{27}$$

$$\operatorname{prox}_{f_{q+1}^*}^{\omega}(\mathbf{y}) = \frac{1}{\lambda}\frac{\mathbf{y}}{\max(1, \|\mathbf{y}\|_2)}. \tag{28}$$

Algorithm 2: SPDHG

Input: Initial $\mathbf{x}$, step parameters ω_i, τ.
Output: $\mathbf{x}$

1: $\mathbf{y} = \mathbf{0}, \mathbf{z} = \bar{\mathbf{z}} = \mathbf{A}^T\mathbf{y} = \mathbf{0}$
2: **for** $k = 1, \dots$
3: $\quad \mathbf{x} = \operatorname{prox}_g^{\tau}(\mathbf{x} - \tau\bar{\mathbf{z}})$
4: $\quad$ Select $i \in 1, \dots, q$ with probability P_i
5: $\quad \mathbf{y}_i^+ = \operatorname{prox}_{f_i^*}^{\omega_i}(\mathbf{y}_i + \omega_i\mathbf{A}_i\mathbf{x})$
6: $\quad \boldsymbol{\Delta}\mathbf{z} = \mathbf{A}_i^T(\mathbf{y}_i^+ - \mathbf{y}_i)$
7: $\quad \mathbf{z} = \mathbf{z} + \boldsymbol{\Delta}\mathbf{z}, \quad \mathbf{y} = \mathbf{y}^+$
8: $\quad \bar{\mathbf{z}} = \mathbf{z} + \frac{1}{P_i}\boldsymbol{\Delta}\mathbf{z}$
9: **end**

A key choice here is the selection of the probabilities P_i in step 4. First note that if $P_i = 1$ then we obtain the standard the primal dual hybrid gradient (PDHG) aka. the Chambolle-Pock algorithm [7]. However, if $P_i < 1$ only a few blocks are updated in the dual variable before $\mathbf{x}$ is updated, which reduced the computational work in each iteration. To ensure that the algorithm puts equal weight on both regularization and the data-fitting blocks, we choose

$$P_i = \begin{cases} \frac{1}{2q} & \text{if } \ i = 1,\dots,q, \\ \frac{1}{2} & \text{if } \ i = q+1. \end{cases} \tag{29}$$

The step sizes are selected according to

$$\omega_i = \gamma\frac{\rho}{\|\mathbf{A}_i\|_2}\mathbf{I}, \quad \tau_i = \gamma^{-1}\frac{\rho P_i}{\|\mathbf{A}_i\|_2}\mathbf{I}, \quad \tau = \min\tau_i, \tag{30}$$

where $\rho < 1$ and $\gamma > 0$. In our implementation we choose $\gamma = 1$ and $\rho = 0.999$, i.e., we put equal balance between primal and dual variable and as large a step size as possible.

4 Numerical Experiments

We present simulated numerical experiments in 2D to show the performance of our method and compare it to our previous non-block version in [25]. The simulations are carried out in MATLAB and we use the ASTRA Toolbox [30] for the matrix-free forward and back projections, i.e., for multiplication with $\mathbf{R}(\boldsymbol{\theta})$ and $\mathbf{R}(\boldsymbol{\theta})^T$. The block formulation involves multiplication with the block matrices $\mathbf{R}(\theta_i) \in \mathbb{R}^{p\times n}$ and their transpose for $i = 1, \dots q$.

We assume that the view angles are independent and distributed according to $\pi_{\text{angles}} = \mathcal{N}(\boldsymbol{\theta}^{\text{equid}}, \delta^2\mathbf{I})$, where $\boldsymbol{\theta}^{\text{equid}}$ are equidistant view angles in [0° to 360°). We generate the noisy data according to

$$\mathbf{b} = \mathbf{R}(\bar{\boldsymbol{\theta}})\bar{\mathbf{x}} + \mathbf{e}, \tag{31}$$

where $\bar{\mathbf{x}}$ represents MATLAB's Shepp-Logan phantom and $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ with $\sigma = 0.005\,\|\mathbf{R}(\bar{\boldsymbol{\theta}})\bar{\mathbf{x}}\|_2/\sqrt{m}$. Moreover, $\bar{\boldsymbol{\theta}}$ is a realization of $\mathcal{N}(\boldsymbol{\theta}^{\text{equid}}, \delta^2\mathbf{I})$.

As in [25] we compare our results with a TV reconstruction that does not take the uncertainty into account, and a TV reconstruction that uses the true view angles $\bar{\boldsymbol{\theta}}$. That is, we compare

Table 1 The physical and discretization parameters in the simulated CT experiments

Parameter	Value
Scan geometry	Fan-beam
Reconstruction domain size	50 cm × 50 cm
Source to center distance	50 cm
Source to detector distance	100 cm
Detector length	130 cm
Small example	
Image pixels	$n = 45^2$
Detector pixels	$p = 90$
Number of view angles	$q = 90$
View angle standard deviation	$\delta = 1.2°$
Medium example	
Image pixels	$n = 135^2$
Detector pixels	$p = 270$
Number of projection angles	$q = 270$
View angle standard deviation	$\delta = 0.4°$

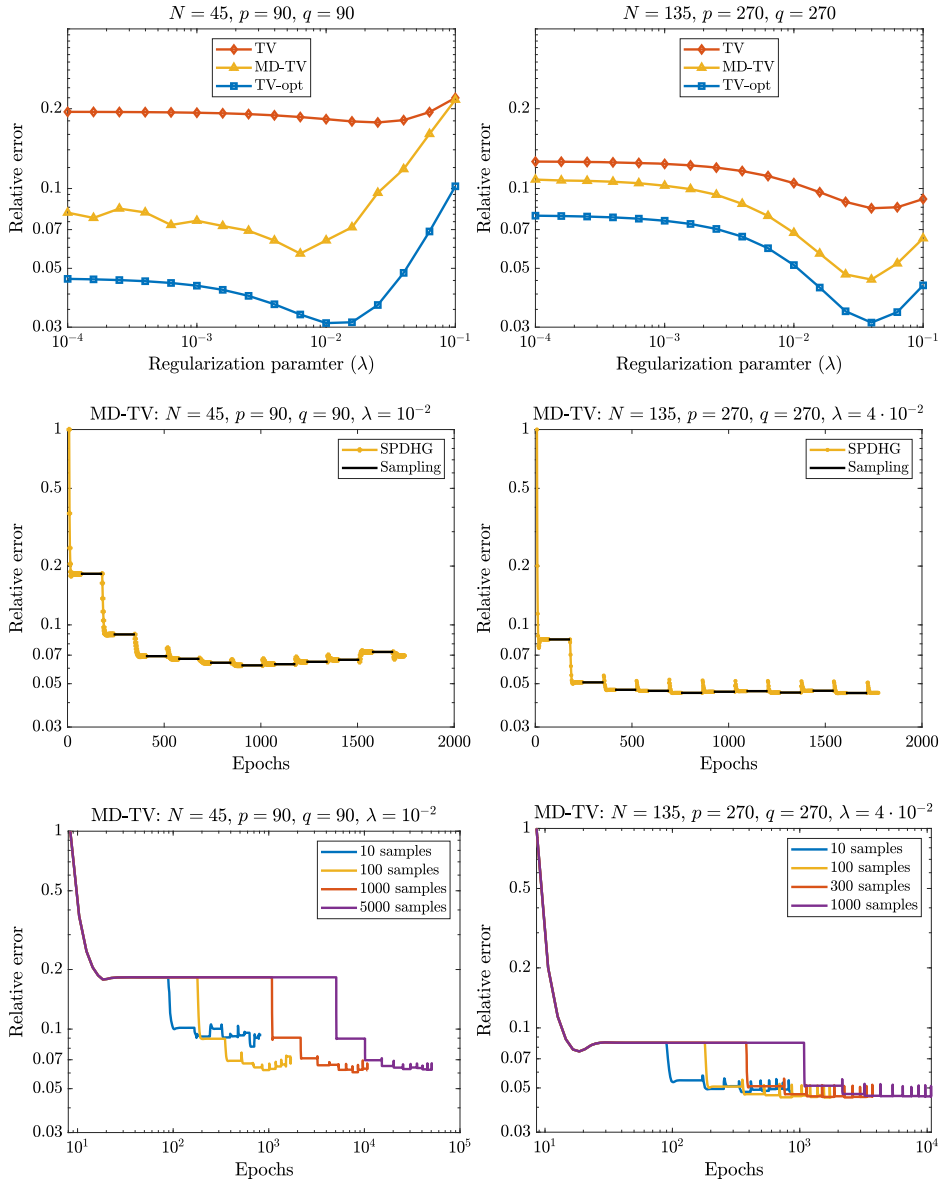


Fig. 2 We compare relative error (y-axis) with regularization parameter choice (top) for all 3 methods, show convergence history of our algorithm (middle) and compare number of samples of the model discrepancy for our method (bottom)

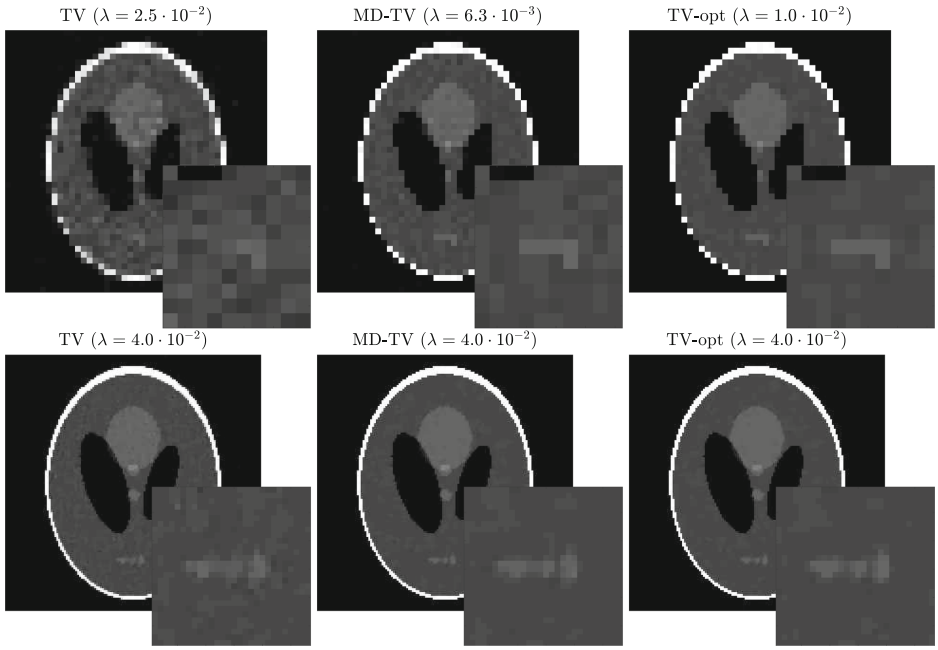


Fig. 3 We compare reconstruction quality of the solutions for all 3 methods with the optimal regularization parameter choice. All images are shown with the same greyscale. Top row $N = 45$, bottom row $N = 135$. We see that our method (MD-TV) improves reconstruction quality visually compared to a standard TV reconstruction, and is visually similar to the TV reconstruction using the true view angles

$$\mathbf{x}_{\text{TV}} = \arg\min_{\mathbf{x}\geq 0} \frac{1}{2}\sum_{i=1}^{q} \|\mathbf{L}_{\mathbf{e},i}(\mathbf{b}_i - \mathbf{R}(\hat{\theta}_i)\,\mathbf{x} - \boldsymbol{\mu}_{\mathbf{e},i})\|_2^2 + \lambda\,\text{TV}(\mathbf{x}), \tag{32}$$

$$\mathbf{x}_{\text{MD-TV}} = \arg\min_{\mathbf{x}\geq 0} \frac{1}{2}\sum_{i=1}^{q} \|\widetilde{\mathbf{L}}_{\boldsymbol{\nu}|\mathbf{x},i}(\mathbf{b}_i - \mathbf{R}(\hat{\theta}_i)\,\mathbf{x} - \boldsymbol{\mu}_{\boldsymbol{\nu}|\mathbf{x},i})\|_2^2 + \lambda\,\text{TV}(\mathbf{x}), \tag{33}$$

$$\mathbf{x}_{\text{TV-opt}} = \arg\min_{\mathbf{x}\geq 0} \frac{1}{2}\sum_{i=1}^{q} \|\mathbf{L}_{\mathbf{e},i}(\mathbf{b}_i - \mathbf{R}(\bar{\theta}_i)\,\mathbf{x} - \boldsymbol{\mu}_{\mathbf{e},i})\|_2^2 + \lambda\,\text{TV}(\mathbf{x}), \tag{34}$$

where $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{\text{equid}}$ and MD-TV is our method.

In all cases, when solving for $\mathbf{x}$ we have a choice between using SPDHG (Algorithm 2) and PDHG. We use $K = 10$ outer iterations and $S = 100$ samples unless stated otherwise. According to our numerical tests, after $K = 10$ outer iterations the reconstruction error levels off, see Fig. 2 (middle plots). The stopping criterion in Algorithm 2 when using SPDHG is

$$\frac{\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2}{\|\mathbf{x}^k\|_2} \leq \frac{10^{-6}}{q}, \tag{35}$$

and when using PDHG we use the stopping criterion

$$\frac{\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2}{\|\mathbf{x}^k\|_2} \leq 10^{-6}. \tag{36}$$

We found empirically that dividing by q in (35) for SPDHG provides similar reconstructions as using (36) for PDHG.

In Table 1, we summarize the parameters for our simulated experiments. We consider the same two examples from [25]. We compare the quality of a solution $\mathbf{x}$ using the relative error $\|\mathbf{x} - \bar{\mathbf{x}}\|_2 / \|\bar{\mathbf{x}}\|_2$ and visual quality.

4.1 Reconstructions

In the top row of Fig. 2 we show the relative error versus the regularization parameter λ for the two examples and

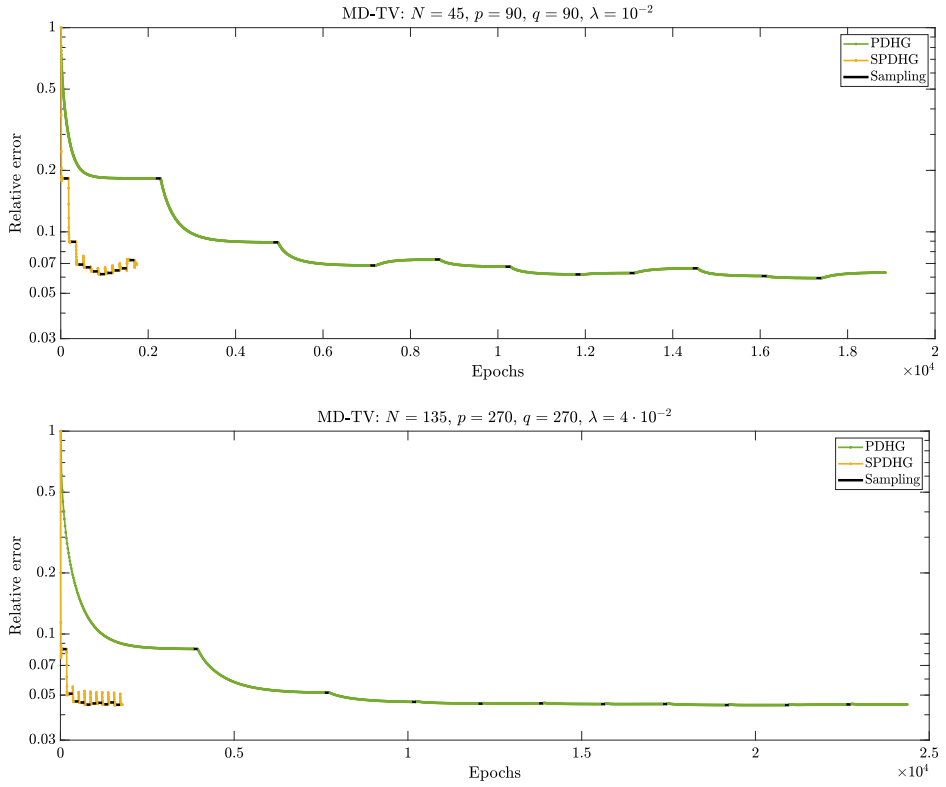


Fig. 4 We compare the convergence history of Algorithm 1 using either PDHG or SPDHG on our block-structure problem (19). Every "corner" in these plots indicates the beginning of a new outer iteration. We see that SPDHG greatly decreases the number of epochs required for this problem

all three methods. First note that for the ranges shown our method (MD-TV) always has a lower relative error compared to the TV reconstruction (TV) from (32). Furthermore, we see that the optimal choice of regularization parameter for MD-TV is similar to the one from the TV reconstruction using the true view angles (TV-opt) from (34). This indicates that the uncertainty due to the view angles is better handled by incorporating it in the data fidelity, rather than increasing the amount of TV regularization.

In the middle row of Fig. 2 we show the convergence history of MD-TV for the $K = 10$ outer iterations. The SPDHG algorithm uses one multiplication with $\mathbf{R}(\hat{\boldsymbol{\theta}})$ and its transpose every $2q$ iterations in expectation. Hence, to show convergence and compare the amount of work we use the unit of an *epoch* defined as the work involved in one multiplication with $\mathbf{R}(\hat{\boldsymbol{\theta}})$ or its transpose, since the computation is dominated by the forward and back projections.

The SPDHG algorithm computes q times as many gradients (multiplication with ∇) as the PDHG algorithm. This leads to a small increase in computational cost that is not accounted for when only comparing epochs. However, evaluation of each finite difference approximation $[\nabla \mathbf{x}]_i$ in (14) requires summing only the neighboring pixels in the CT reconstruction, compared to summing all pixels along each line for the forward and back projections. Hence—except for very underdetermined problems—computation of the gradient is significantly less expensive than computing the forward projection. In our numerical tests using MATLAB, we found that a direct computation of the gradient is 1000 times faster than computing a forward projection with the optimized CPU-ASTRA library. The convergence plot includes the

work associated with sampling the model discrepancy which requires $S + 1$ multiplications of $\mathbf{R}(\boldsymbol{\theta})$.

In the bottom row of Fig. 2 we show the convergence history of MD-TV for different number of samples S. We note that relative error of the final solution does not improve when using more than about $S = 100$ samples, so we use that for our other experiments.

Finally, in Fig. 3 we show the reconstructions obtained from the regularization parameter that yielded the lowest relative error using all 3 methods.

4.2 Convergence

In Fig. 4, we compare the convergence history of our algorithm using SPDHG which takes the block structure into account with our previous algorithm using PDHG, which does not. Every "corner" in these plots indicates the beginning of a new outer iteration, and we see a significant decrease in the number of epochs required in each outer iteration. We see for example that in the small example the SPDHG algorithm reaches convergence in the first outer iteration using 69 epochs compared to PDHG with 2174 epochs.

5 Conclusion

We proposed a new model and iterative algorithm for CT reconstruction when the view angles are uncertain. The uncertainty is described by a model discrepancy term and included in the data fidelity term of a TV regularized variational model. To establish the new model, we have overcome two key difficulties: (i) the probability distribution for the combined measurement noise and model discrepancy may not have a closed-form expression leading to difficulties for deriving a data fidelity term and (ii) the combined noise and discrepancy depend on the unknown CT reconstruction. We handle these issues by approximating the model discrepancy by a Gaussian, leading to a closed-form expression for the data-fidelity, and we alternately update the reconstruction and the parameters of the model discrepancy. 2D numerical experiments show that this approach improves the reconstruction quality in terms of relative error and visual quality compared to a standard TV reconstruction. Furthermore, our model admits a block structure, which we take advantage of to greatly reduce the overall computational work.

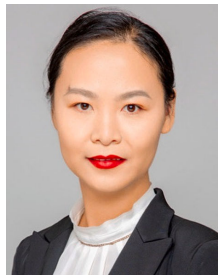
Acknowledgements We thank Aku Seppänen for a helpful discussion regarding the block formulation of our variational method. We also thank Matthias Ehrhardt for a discussion on the implementation details of SPDHG for this problem. Finally, we thank the referees for comments and suggestions that helped to improve the presentation.

References

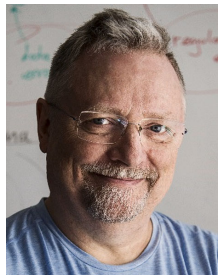
1. Arridge, S.R., Kaipio, J.P., Kolehmainen, V., Schweiger, M., Somersalo, E., Tarvainen, T., Vauhkonen, M.: Approximation errors and model reduction with an application in optical diffusion tomography. Inverse Probl. **22**(1), 175–195 (2006)
2. Basu, S., Bresler, Y.: Uniqueness of tomography with unknown view angles. IEEE Trans. Image Process. **9**(6), 1094–1106 (2000)
3. Benning, M., Burger, M.: Modern regularization methods for inverse problems. Acta Numer. **27**, 1–111 (2018)
4. Brandt, C., Seppänen, A.: Recovery from errors due to domain truncation in magnetic particle imaging: approximation error modeling approach. J. Math. Imaging Vis. **60**(8), 1196–1208 (2018)
5. Buzug, T.M.: Computed Tomography: From Photon Statistics to Modern Cone-Beam CT. Springer, Berlin (2008)
6. Calvetti, D., Dunlop, M., Somersalo, E., Stuart, A.: Iterative updating of model error for Bayesian inversion. Inverse Probl. **34**(2), 025008 (2018)
7. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. **40**(1), 120–145 (2011)
8. Dongarra, J., Moler, C., Bunch, J., Stewart, G.: LINPACK Users' Guide. SIAM, Philadephia (1979)
9. Ehrhardt, M.J., Markiewicz, P.J., Schönlieb, C.B.: Faster PET reconstruction with non-smooth priors by randomization and preconditioning. Phys. Med. Biol. **64**(22), 225019 (2019)
10. Fang, Y., Murugappan, S., Ramani, K.: Estimating view parameters from random projections for tomography using spherical MDS. BMC Med. Imaging **10**(1), 12 (2010)
11. Hansen, T.M., Cordua, K.S., Holm Jacobsen, B., Mosegaard, K.: Accounting for imperfect forward modeling in geophysical inverse problems—exemplified for crosshole tomography. Geophysics **79**(3), H1–H21 (2014)
12. Kaipio, J., Somersalo, E.: Statistical and Computational Inverse Problems. Springer, New York (2005)
13. Kaipio, J., Somersalo, E.: Statistical inverse problems: discretization, model reduction and inverse crimes. J. Comput. Appl. Math. **198**(2), 493–504 (2007)
14. Kennedy, M.C., O'Hagan, A.: Bayesian calibration of computer models. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **63**(3), 425–464 (2001)
15. Kolehmainen, V., Tarvainen, T., Arridge, S.R., Kaipio, J.P.: Marginalization of uninteresting distributed parameters in inverse problems—application to diffuse optical tomography. Int. J. Uncertain. Quantif. **1**(1), 1–17 (2011)
16. Korolev, Y., Lellmann, J.: Image reconstruction with imperfect forward models and applications in deblurring. SIAM J. Imaging Sci. **11**(1), 197–218 (2018)
17. Natterer, F.: The Mathematics of Computerized Tomography. Wiley, Chicago (1986)
18. Niebler, S., Schömer, E., Tjaden, H., Schwanecke, U., Schulze, R.: Projection-based improvement of 3D reconstructions from motion-impaired dental cone beam CT data. Med. Phys. **46**(10), 4470–4480 (2019)
19. Nissinen, A., Heikkinen, L.M., Kaipio, J.P.: The Bayesian approximation error approach for electrical impedance tomography—experimental results. Meas. Sci. Technol. **19**(1), 015501 (2007)
20. Nissinen, A., Kolehmainen, V., Kaipio, J.P.: Reconstruction of domain boundary and conductivity in electrical impedance tomography using the approximation error approach. Int. J. Uncertain. Quantif. **1**(3), 203–222 (2011)
21. Madsen, R.B., Hansen, T.M.: Estimation and accounting for the modeling error in probabilistic linearized amplitude variation with offset inversion. Geophysics **83**(2), N15–N30 (2018)

22. Mallick, S.P., Agarwal, S., Kriegman, D.J., Belongie, S.J., Carragher, B., Potter, C.S.: Structure and view estimation for tomographic reconstruction: a Bayesian approach. In: Fitzgibbon, A., Taylor, C.J., Lecun, Y. (eds.) 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), **2**(1):2253–2260 (2006)
23. Peng, C.B., Rodi, W.L., Toksöz, M.N.: A Tikhonov regularization method for image reconstruction. In: Wei, Y., Gu, B. (eds.) Acoustical Imaging, vol. 20, pp. 153–164. Springer, Berlin (1993)
24. Pulkkinen, A., Kolehmainen, V., Kaipio, J.P., Cox, B.T., Arridge, S.R., Tarvainen, T.: Approximate marginalization of unknown scattering in quantitative photoacoustic tomography. Inverse Probl. Imaging **8**(3), 811–829 (2014)
25. Riis, N.A.B., Dong, Y.: A new iterative method for CT reconstruction with uncertain view angles. In: Lellmann, J., Burger, M., Modersitzki, J. (eds.) Scale Space and Variational Methods in Computer Vision. SSVM 2019. Lecture Notes in Computer Science, vol. 11603, pp. 156–167. Springer, Berlin (2019)
26. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**(1–4), 259–268 (1992)
27. Sidky, E.Y., Pan, X.: Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. Phys. Med. Biol. **53**(17), 4777–4807 (2008)
28. Tick, J., Pulkkinen, A., Tarvainen, T.: Modelling of errors due to speed of sound variations in photoacoustic tomography using a Bayesian framework. Biomed. Phys. Eng. Express **6**(1), 015003 (2019)
29. Tikhonov, A.N.: Solution of incorrectly formulated problems and the regularization method. Soviet Math. Dokl. **4**:1035–1038 (1963); English translation of Dokl. Akad. Nauk. SSSR **151**:501–504 (1963)
30. van Aarle, W., Palenstijn, W.J., Cant, J., Janssens, E., Bleichrodt, F., Dabravolski, A., Sijbers, J.: Fast and flexible X-ray tomography using the ASTRA toolbox. Opt. Express **24**(22), 25129–25147 (2016)

Nicolai André Brogaard Riis is a Ph.D. researcher at the Technical University of Denmark. His research focuses on the mathematical modelling of inverse problems with uncertainty in the physical models.



Yiqiu Dong is associate professor at the Technical University of Denmark. She received the B.Sc. degree in mathematics from Yantai University, Yantai, China, in 2002 and the Ph.D. degree in mathematics from Peking University, Beijing, China, in 2007. Her research area includes mathematical modeling and algorithms in imaging sciences, inverse problems and variational methods, matrix application and computation, and optimization methods.



Per Christian Hansen is professor of scientific computing at the Technical University of Denmark. His research focuses on numerical analysis, iterative reconstruction methods, and computational methods for inverse problems. He is the author of four books and several MATLAB packages, and he is a SIAM Fellow.

APPENDIX C

Computed Tomography with View Angle Estimation using Uncertainty Quantification

Submitted to: Inverse Problems, 2020.

Riis, N. A. B., Dong, Y., & Hansen, P. C.

Computed Tomography with View Angle Estimation using Uncertainty Quantification

Nicolai André Brogaard Riis, Yiqiu Dong and Per Christian Hansen

Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark

E-mail: nabr@dtu.dk, yido@dtu.dk, pcha@dtu.dk

October 2020

Abstract. We consider computed tomography (CT) with uncertain measurement geometry, with a focus on the case where the view angles are uncertain and where estimation of these angles improves the reconstruction. We propose a new reconstruction model and a corresponding algorithm that has an additional view-angle estimation component, allowing us to determine the angles solely from the measured CT data. A key component of our approach is that we quantify the uncertainty of the view angles via a model-discrepancy formulation, allowing us to take the uncertainty into account in the image reconstruction. This approach generalizes in a straightforward way to other cases of uncertain geometry. Our method is computationally efficient since we can utilize a block-structure in our model for estimation of both the CT image and the view angles. The joint image/angle reconstruction problem is non-convex which leads to difficulties in recently proposed algorithms, and we demonstrate numerically that our method seems to avoid these difficulties. Simulations show that our method, with a total variation (TV) prior that reflects our phantoms, is able to achieve reconstructions whose quality is similar to ones obtained with the correct view angles (the ideal scenario).

Keywords: Computed Tomography, View Angle Estimation, Uncertainty Quantification, Block Representation, Alternating Updates.

Submitted to: *Inverse Problems*

1. Introduction

In this paper we consider Computed Tomography (CT) where – in addition to reconstructing an X-ray attenuation image of an object – estimation of certain parameters of the measurement geometry leads to an improved reconstruction. We focus on the important case of uncertain view angles, but our approach will also work for other types of uncertainties in the measurement geometry. In the standard models for CT the view angles of the set-up are assumed to be known exactly, but in practice they may only be known with a limited accuracy. For specific applications, such as nano and micro X-ray tomography [21, 32] or motion impaired CT [17] this uncertainty in the view angles is important to take into account to improve the quality of the reconstruction. The goal in this paper is to study a new method for simultaneous image reconstruction and view angle estimation solely from measured CT data with no machine or object calibration.

The generic version of the discretized CT reconstruction model takes the form

$$\mathbf{b} = \mathbf{R}(\boldsymbol{\theta})\,\mathbf{x} + \mathbf{e}, \tag{1}$$

where $\mathbf{b} \in \mathbb{R}^m$ is the measured (noisy) CT data, $\mathbf{x} \in \mathbb{R}^n$ is the unknown X-ray attenuation coefficients (CT image) that we want to reconstruct, and $\mathbf{e} \in \mathbb{R}^m$ is the measurement noise. The CT forward model – or forward projection – is represented by the matrix $\mathbf{R}(\boldsymbol{\theta}) \in \mathbb{R}^{m\times n}$, which is parameterized by the vector $\boldsymbol{\theta} \in \mathbb{R}^q$ that represents the geometry – here the view angles. Specifically, we consider CT geometries where each position of the source is defined by a single angle given by the view angles in $\boldsymbol{\theta}$. This includes 2D parallel- and fan-beam as well as 3D cone-beam with a fixed axis rotation. Any other parameter of the measurement geometry is assumed to be known with high accuracy and are therefore not parameterized explicitly in the forward model. In the case that other parameters are uncertain these can be added to the parameterization $\boldsymbol{\theta}$. For more details on the physical and mathematical models of CT see e.g., [5, 16].

1.1. Previous work

The CT problem in (1) can be handled in many different ways. Direct inversion methods [16] are fast, but may perform poorly with increasing measurement noise or uncertainty in the parameters of the scan, e.g., in the view angles. Variational methods can incorporate prior information via regularization terms, see, e.g., the survey [3]. These methods are robust toward measurement noise because of the regularization techniques. Techniques such as Tikhonov regularization [22, 29] or total variation (TV) [25, 26] are used for the image reconstruction problem. However, these methods do not take the uncertainty of the view angles into account, which means that good performance is not guaranteed if the view angles are only known with limited accuracy.

View angles and other scan parameters can be estimated by attaching markers to the scanned object that can then be tracked during the measurement process. In addition to being time-consuming to attach, these markers are made of a material with

a high X-ray absorption, which can obscure areas of interest in the scanned object and create artefacts in the reconstruction [27]. Alternatively, marker-free view angle estimation approaches that are based only on the measured CT data have been studied analytically for specific geometries in e.g., [2, 7, 9]. This has – along with a number of practical methods – led to a large body of work on estimating view angles and other scan parameters in CT such as [2, 7, 9, 17, 20, 31, 33]. These methods can be categorized into two groups: i) estimating the view angles directly from the CT data followed by a reconstruction, and ii) jointly estimating the view angles and image reconstruction.

Methods such as [9] that first estimate the view angles based purely on the CT data and then reconstruct the CT image are computationally efficient, but they suffer from error propagation since any errors introduced in the view angle estimation propagate to the reconstruction. Jointly estimating the view angles and CT image, e.g., via Bayesian sampling methods [20], avoid this error propagation, but these methods can be computationally expensive due to the sampling-based nature of the approach.

A popular approach for joint estimation is the so-called projection matching (PM) methods. Here, given a reconstruction, the view angles are estimated by matching the measured projections with ones obtained from the forward model. Different approaches for finding the matching projections have been proposed and investigated such as cross-correlation [33] or joint optimization using, e.g., exhaustive search [21], gradient-based methods [31] or derivative-free methods [17]. In general, the goal of projection matching using joint optimization is to obtain the best fitting view angles by comparing the distance of the forward model projection with the CT data in some metric (typically a 2-norm data-fit). This leads to a non-linear problem and depending on the regularization it can be a non-convex problem with many local minima, often leading to inaccuracies in the view angle estimation [31].

In our previous work [24], we proposed a CT reconstruction method that takes potential uncertainty in the view angles into account using a model-discrepancy term based on the approximation error approach (AEA) [13]. This approach worked well, but was lacking in two major areas which we address in this paper: i) we assumed that the model discrepancy followed a Gaussian distribution and hence the quality of the reconstruction was limited by this assumption and ii) the uncertainty in the view angles is marginalized in the likelihood and hence even if the correct distribution of the model discrepancy term is known, the quality of the reconstruction is impacted by the marginalization of the uncertainty.

1.2. Our contribution

To address the above-mentioned issues, we propose a new CT reconstruction method that estimates the view angles in addition to reconstructing the CT image. A key contribution of our method is that we are able to quantify the uncertainty of the view angles and incorporate this into the image reconstruction with acceptable computational overhead. The uncertainty is quantified by utilizing a model-discrepancy term, which is

used to estimate the view angles as well as to incorporate any remaining uncertainty in the likelihood for the CT data. We summarize our contributions as follows:

- We propose a new CT reconstruction method with a view-angle estimation component that uses uncertainty quantification.
- We quantify the uncertainty of the view angles and take this into account in the image reconstruction.
- We achieve a fast and computationally efficient angle estimation procedure by utilizing a block-structure of the CT reconstruction model.
- We show numerically that our method gives reconstructions of similar quality to ones obtained with the exact view angles.
- We show numerically that our method is robust towards the choice of the regularization parameter on the TV prior.
- We show numerically that our algorithm seems to avoid the difficulties associated with the non-convexity of the joint image/angle reconstruction problem, found in other recent methods.

1.3. Structure of paper

Our paper is structured as follows. In Section 2 we summarize our previous work in [24], where we derive a CT reconstruction method that takes uncertainty in the view angles into account by marginalization. In Section 3 we propose our new method which additionally estimates the view angles and the associated uncertainty by jointly estimating the view angles and the CT image in an iterative procedure. In Section 4 we present simulated numerical results that show the performance of our method and finally Section 5 is the conclusion.

2. Review of CT algorithm with marginalized view angle uncertainty

In this section we summarize our previous work in [24] to arrive at a statistical model for the CT problem (1), where the view angles are known with a limited accuracy and are marginalized in the likelihood. We assume Gaussian distributions for the view angles $\boldsymbol{\theta}$ and measurement noise $\mathbf{e}$. The CT model with uncertain view angles is then given by

$$\mathbf{b} = \mathbf{R}(\boldsymbol{\theta})\,\mathbf{x} + \mathbf{e}, \quad \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \mathrm{diag}(\boldsymbol{\delta})), \quad \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}), \tag{2}$$

where $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ are chosen as the nominal view angles of the scanner with variances $\mathrm{diag}(\boldsymbol{\delta})$. The measurement noise $\mathbf{e}$ is assumed to be independent identically distributed zero-mean Gaussian with standard deviation σ.

To avoid working with a distribution of forward models $\mathbf{R}(\boldsymbol{\theta})$, $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \mathrm{diag}(\boldsymbol{\delta}))$, we reformulate (2) using a fixed forward model $\mathbf{R}(\boldsymbol{\mu}_{\boldsymbol{\theta}})$ and obtain the modified CT model

$$\mathbf{b} = \mathbf{R}(\boldsymbol{\mu}_{\boldsymbol{\theta}})\,\mathbf{x} + \boldsymbol{\eta} + \mathbf{e}, \quad \boldsymbol{\eta} \sim \pi_{\mathrm{discrep}}(\cdot), \quad \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}), \tag{3}$$

where the so-called *model-discrepancy term* $\boldsymbol{\eta}$ is given by

$$\boldsymbol{\eta}(\boldsymbol{\theta}, \mathbf{x}) = \mathbf{R}(\boldsymbol{\theta})\,\mathbf{x} - \mathbf{R}(\boldsymbol{\mu}_{\boldsymbol{\theta}})\,\mathbf{x}. \tag{4}$$

It can be advantageous to use the model (3) instead of (2) because the uncertainty in the view angles is moved to $\boldsymbol{\eta}$ while the forward model $\mathbf{R}(\boldsymbol{\mu}_{\boldsymbol{\theta}})$ is fixed. If we ignore that $\boldsymbol{\eta}$ depends on $\mathbf{x}$ and consider $\boldsymbol{\eta}$ as independent additive noise then (3) becomes a standard CT image reconstruction model with two additive noise terms.

Representing model uncertainty as a model-discrepancy term has been successfully applied to many imaging applications, such as fine-to-coarse mesh approximation in diffuse optical tomography [1], unknown domain boundaries in electric impedance tomography [19], unknown scattering in both diffuse optical tomography and quantitative photoacoustic tomography [15, 23], truncation errors in magnetic particle imaging [4], and uncertain sound speed in photoacoustic tomography [28]. The idea of a model-discrepancy term was originally proposed in [14] where $\boldsymbol{\eta}$ was modelled as a Gaussian process and was later adapted to Bayesian inverse problems in [12, 13] and named the Approximation Error Approach (AEA).

2.1. Gaussian approximation of model discrepancy

The distribution π_{discrep} may not have a closed-form expression, so we approximate it by a simple distribution. Gaussian approximation of the model discrepancy have been shown experimentally to be useful for this application [24] and many other applications [1, 4, 6, 10, 12, 13, 15, 18, 19, 23, 28]. Hence, we simplify the model (3) by approximating the distribution π_{discrep} conditioned on $\mathbf{x}$ by a Gaussian distribution, i.e., $\boldsymbol{\eta}|\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\eta}|\mathbf{x}}, \mathbf{C}_{\boldsymbol{\eta}|\mathbf{x}})$ with mean and covariance depending on $\mathbf{x}$.

Given a reconstruction $\hat{\mathbf{x}}$ one can generate samples of the model discrepancy $\boldsymbol{\eta}|\mathbf{x} = \hat{\mathbf{x}}$ by drawing samples $\boldsymbol{\theta}^s$ following the prior distribution $\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \operatorname{diag}(\boldsymbol{\delta}))$, and evaluating the model discrepancy

$$\boldsymbol{\eta}^s_{\hat{\mathbf{x}}} = \mathbf{R}(\boldsymbol{\theta}^s)\,\hat{\mathbf{x}} - \mathbf{R}(\boldsymbol{\mu}_{\boldsymbol{\theta}})\,\hat{\mathbf{x}}, \quad s = 1, \ldots, S. \tag{5}$$

The sample mean and covariance of the model discrepancy are then given by

$$\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}|\hat{\mathbf{x}}} = \frac{1}{S}\sum_{s=1}^{S} \boldsymbol{\eta}^s_{\hat{\mathbf{x}}}, \tag{6}$$

$$\widetilde{\mathbf{C}}_{\boldsymbol{\eta}|\hat{\mathbf{x}}} = \frac{1}{S-1}\sum_{s=1}^{S} (\boldsymbol{\eta}^s_{\hat{\mathbf{x}}} - \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}|\hat{\mathbf{x}}})(\boldsymbol{\eta}^s_{\hat{\mathbf{x}}} - \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}|\hat{\mathbf{x}}})^T. \tag{7}$$

2.2. CT image reconstruction

To arrive at a good reconstruction we choose a method based on a variational formulation. Variational methods require a data-fitting term that incorporate the forward model and regularization terms that incorporate prior information about the CT image, see e.g., the survey [3]. For our purpose the data-fitting term is derived through

the likelihood function associated with the model (3) using the Gaussian approximation of the model discrepancy in Section 2.1. Letting $\boldsymbol{\nu} = \boldsymbol{\eta} + \mathbf{e}$ we marginalize with respect to $\boldsymbol{\nu}$ to arrive at the likelihood

$$\pi(\mathbf{b}|\mathbf{x}) = \int_{\mathbb{R}^m} \pi(\mathbf{b}, \boldsymbol{\nu}|\mathbf{x}) \mathrm{d}\boldsymbol{\nu} = \pi_{\boldsymbol{\nu}|\mathbf{x}}(\mathbf{b} - \mathbf{R}(\boldsymbol{\mu}_{\boldsymbol{\theta}})\mathbf{x}|\mathbf{x}). \tag{8}$$

By taking the negative logarithm and using the Gaussian distributions of $\boldsymbol{\eta}|\mathbf{x}$ and $\mathbf{e}$, we obtain the closed-form expression of the negative log-likelihood

$$-\log \pi(\mathbf{b}|\mathbf{x}) \propto \frac{1}{2} \|\mathbf{L}_{\boldsymbol{\nu}|\mathbf{x}}(\mathbf{b} - \mathbf{R}(\boldsymbol{\mu}_{\boldsymbol{\theta}})\mathbf{x} - \boldsymbol{\mu}_{\boldsymbol{\eta}|\mathbf{x}})\|_2^2, \tag{9}$$

where $\mathbf{L}_{\boldsymbol{\nu}|\mathbf{x}}$ is the Cholesky factor of the combined inverse covariance $(\mathbf{C}_{\boldsymbol{\eta}|\mathbf{x}} + \sigma^2\mathbf{I})^{-1}$. Combining this with total variation (TV) regularization, which can work well for large-scale CT problems [25, 26], we arrive at the CT image reconstruction problem

$$\mathbf{x}_{\text{MD-TV}} = \arg\min_{\mathbf{x} \geq 0} \frac{1}{2} \|\mathbf{L}_{\boldsymbol{\nu}|\mathbf{x}}(\mathbf{b} - \mathbf{R}(\boldsymbol{\mu}_{\boldsymbol{\theta}})\mathbf{x} - \boldsymbol{\mu}_{\boldsymbol{\eta}|\mathbf{x}})\|_2^2 + \lambda \mathrm{TV}(\mathbf{x}), \tag{10}$$

where $\lambda > 0$ denotes the regularization parameter, and the TV term is defined as

$$\mathrm{TV}(\mathbf{x}) = \|\nabla \mathbf{x}\|_{2,1} \equiv \sum_{i=1}^{n} \|[\nabla \mathbf{x}]_i\|_2 , \tag{11}$$

in which $[\nabla \mathbf{x}]_i$ denotes the discrete gradient of $\mathbf{x}$ at the ith pixel using reflexive boundary conditions. The non-negativity constraint represents the fact that the attenuation coefficients $\mathbf{x}$ cannot be negative.

The optimization problem in (10) is solved efficiently by the stochastic primal-dual hybrid gradient (SPDHG) algorithm [8] due to the separable nature of the data-fitting term. More details on the block-representation of the covariance matrix and efficient computation of the Cholesky factor of the inverse covariance, as well as the solver for (10), can be found in our previous work [24].

3. New CT algorithm with View Angle Estimation

The quality of the reconstruction from (10) is limited by the Gaussian assumption on the model-discrepancy term $\boldsymbol{\eta}$. If this assumption is not valid or the uncertainty in the view angles is too large then – even when we take the uncertainty into account – we cannot obtain satisfactory results. To overcome this limitation and obtain a better reconstruction, we have to estimate the view angles in addition to the image reconstruction. In this section, we propose a new method with added view angle estimation based only on the measured CT data.

3.1. View angle estimation

First of all, we consider how to estimate the view angles $\boldsymbol{\theta}$, more specifically the mean $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and the associated uncertainty according to the CT model (2), given the

measured CT data $\mathbf{b}$ and an obtained reconstruction $\hat{\mathbf{x}}$. Recall that the distribution of $\boldsymbol{\nu} = \boldsymbol{\eta} + \mathbf{e}$ conditioned on $\mathbf{x}$ is approximated by a Gaussian distribution, i.e., $\boldsymbol{\nu}|\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\eta}|\mathbf{x}}, \mathbf{C}_{\boldsymbol{\eta}|\mathbf{x}} + \sigma^2\mathbf{I})$. This provides a natural connection between $\mathbf{x}$ and $\boldsymbol{\theta}$. With the Gaussian approximation and assuming that $\mathbf{e}$ is independent of $\boldsymbol{\eta}$ as well as $\boldsymbol{\theta}$, we have a joint Gaussian distribution

$$\begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\nu}|\hat{\mathbf{x}} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{\theta}} \\ \boldsymbol{\mu}_{\boldsymbol{\nu}|\hat{\mathbf{x}}} \end{bmatrix}, \begin{bmatrix} \operatorname{diag}(\boldsymbol{\delta}) & \mathbf{C}^T_{\boldsymbol{\nu\theta}|\hat{\mathbf{x}}} \\ \mathbf{C}_{\boldsymbol{\nu\theta}|\hat{\mathbf{x}}} & \mathbf{C}_{\boldsymbol{\nu}|\hat{\mathbf{x}}} \end{bmatrix} \right), \tag{12}$$

where $\boldsymbol{\mu}_{\boldsymbol{\nu}|\hat{\mathbf{x}}} = \boldsymbol{\mu}_{\boldsymbol{\eta}|\hat{\mathbf{x}}}$ because the mean measurement noise is zero, $\mathbf{C}_{\boldsymbol{\nu\theta}|\hat{\mathbf{x}}} = \mathbf{C}_{\boldsymbol{\eta\theta}|\hat{\mathbf{x}}}$ because the measurement noise is independent of $\boldsymbol{\theta}$, and $\mathbf{C}_{\boldsymbol{\nu}|\hat{\mathbf{x}}} = \mathbf{C}_{\boldsymbol{\eta}|\hat{\mathbf{x}}} + \sigma^2\mathbf{I}$.

The joint distribution provides a closed-form expression (shown below) for the conditional parameters – i.e., mean and variance – of the view angles given the reconstruction $\hat{\mathbf{x}}$, through the combined model-discrepancy and noise term $\boldsymbol{\nu}|\hat{\mathbf{x}}$. This is in contrast to determining the conditional parameters, e.g., from $\pi(\boldsymbol{\theta}|\mathbf{b}, \mathbf{x})$, which would require a distributional prior on $\mathbf{x}$. Thus, in turn, requires sampling-based methods that are unfeasible for large-scale CT. Moreover, deriving the conditional parameters provide natural uncertainty estimates of the view angles that can be used in the image reconstruction from Section 2 – a feature which is not present in, e.g., projection matching methods which only provide a single point estimate through optimization.

To derive the conditional parameters according to the joint distribution we need an estimate of the combined model discrepancy and measurement noise term $\boldsymbol{\nu}$. Given a reconstruction $\hat{\mathbf{x}}$ the combined term can be estimated by $\hat{\boldsymbol{\nu}} = \mathbf{b} - \mathbf{R}(\boldsymbol{\mu}_{\boldsymbol{\theta}})\hat{\mathbf{x}}$. To obtain an estimate of the view angles the conditional mean of $\boldsymbol{\theta}$ given $\hat{\boldsymbol{\nu}}$ is

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{\theta}|\boldsymbol{\nu}=\hat{\boldsymbol{\nu}}} &= \boldsymbol{\mu}_{\boldsymbol{\theta}} + \mathbf{C}^T_{\boldsymbol{\nu\theta}|\hat{\mathbf{x}}}(\mathbf{C}_{\boldsymbol{\eta}|\hat{\mathbf{x}}} + \sigma^2\mathbf{I})^{-1}(\hat{\boldsymbol{\nu}} - \boldsymbol{\mu}_{\boldsymbol{\nu}|\hat{\mathbf{x}}}), \\ &= \boldsymbol{\mu}_{\boldsymbol{\theta}} + \mathbf{C}^T_{\boldsymbol{\eta\theta}|\hat{\mathbf{x}}}(\mathbf{C}_{\boldsymbol{\eta}|\hat{\mathbf{x}}} + \sigma^2\mathbf{I})^{-1}(\mathbf{b} - \mathbf{R}(\boldsymbol{\mu}_{\boldsymbol{\theta}})\,\hat{\mathbf{x}} - \boldsymbol{\mu}_{\boldsymbol{\eta}|\hat{\mathbf{x}}}). \end{aligned} \tag{13}$$

Furthermore, the conditional covariance is given by

$$\begin{aligned} \mathbf{C}_{\boldsymbol{\theta}|\boldsymbol{\nu}=\hat{\boldsymbol{\nu}}} &= \operatorname{diag}(\boldsymbol{\delta}) - \mathbf{C}^T_{\boldsymbol{\nu\theta}|\hat{\mathbf{x}}}(\mathbf{C}_{\boldsymbol{\eta}|\hat{\mathbf{x}}} + \sigma^2\mathbf{I})^{-1}\mathbf{C}_{\boldsymbol{\nu\theta}|\hat{\mathbf{x}}}, \\ &= \operatorname{diag}(\boldsymbol{\delta}) - \mathbf{C}^T_{\boldsymbol{\eta\theta}|\hat{\mathbf{x}}}(\mathbf{C}_{\boldsymbol{\eta}|\hat{\mathbf{x}}} + \sigma^2\mathbf{I})^{-1}\mathbf{C}_{\boldsymbol{\eta\theta}|\hat{\mathbf{x}}}. \end{aligned} \tag{14}$$

The mean $\boldsymbol{\mu}_{\boldsymbol{\eta}|\hat{\mathbf{x}}} \in \mathbb{R}^m$ and the covariance matrix $\mathbf{C}_{\boldsymbol{\eta}|\hat{\mathbf{x}}}\mathbb{R}^{m\times m}$ can be approximated by the sample mean and sample covariance defined in (6) and (7), respectively. Furthermore, the cross-covariance matrix $\mathbf{C}_{\boldsymbol{\eta\theta}|\hat{\mathbf{x}}} \in \mathbb{R}^{m\times q}$ of the view angles $\boldsymbol{\theta} \in \mathbb{R}^q$ and the model-discrepancy term $\boldsymbol{\eta}|\hat{\mathbf{x}} \in \mathbb{R}^m$ can be approximated by the sample cross-covariance

$$\widetilde{\mathbf{C}}_{\boldsymbol{\eta\theta}|\hat{\mathbf{x}}} = \frac{1}{S-1}\sum_{s=1}^{S}(\boldsymbol{\eta}^s_{\hat{\mathbf{x}}} - \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}|\hat{\mathbf{x}}})(\boldsymbol{\theta}^s - \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}})^T \tag{15}$$

with sample mean

$$\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}} = \frac{1}{S}\sum_{s=1}^{S}\boldsymbol{\theta}^s, \tag{16}$$

where $\boldsymbol{\theta}^s$ follows the assumed prior distribution $\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \operatorname{diag}(\boldsymbol{\delta}))$.

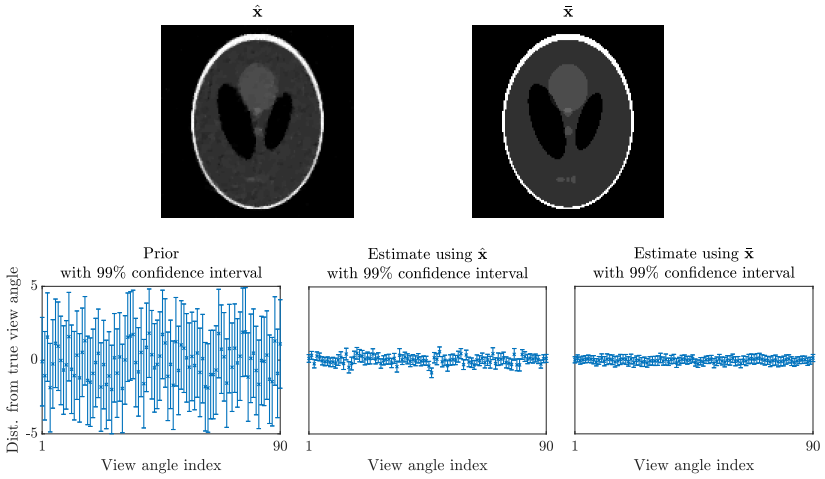


Figure 1. The dependence of the conditional mean and covariance estimates on the CT image used. Top: example of reconstruction using (10) (left) and exact CT image (right) used to calculate the sample mean and covariances shown below. Bottom: the mean and confidence intervals according to the prior (left) and the conditional mean and covariance calculated from (13) and (14), respectively, using $\hat{\mathbf{x}}$ (middle) and $\bar{\mathbf{x}}$ (right).

The quality of the conditional mean and variance estimates depend heavily on the reconstruction $\hat{\mathbf{x}}$ both for estimating $\hat{\boldsymbol{\nu}}$ and for the sampled mean and covariances in (6), (7) and (15). For this reason, we cannot expect the conditional mean (13) to be close to the true underlying view angles if the estimate of $\hat{\mathbf{x}}$ is poor. A similar argument holds for the conditional covariance in (14)

To illustrate this point, Figure 1 shows the conditional mean and 99% confidence intervals calculated from (13) and (14), respectively, using a reconstruction $\hat{\mathbf{x}}$ from solving (10) and from the ground truth CT image $\bar{\mathbf{x}}$. The object is the Shepp-Logan phantom and the number of samples is $S = 100$. For comparison, Figure 1 also shows the mean and 99% confidence intervals from the prior distribution $\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \mathrm{diag}(\boldsymbol{\delta}))$. Note that (13) and (14) yield estimates of the view angles that are closer to the ground truth with smaller uncertainty (variance) compared to the prior. Furthermore, we see that even when using a reconstruction $\hat{\mathbf{x}}$ rather than the ground truth $\bar{\mathbf{x}}$ it is possible to estimate the view angles.

However, the variance does not capture the true uncertainty when using the reconstruction. This suggests that the conditional mean can be used to update the current estimate of the view angles, but the conditional covariance is too aggressively reduced when using $\hat{\mathbf{x}}$ rather than $\bar{\mathbf{x}}$ and therefore can not be fully trusted. To avoid this issue we add a relaxation parameter $\alpha \in [0, 1]$ to the term that reduces the variance, to take into account that we are using an estimated CT image and not the ground truth,

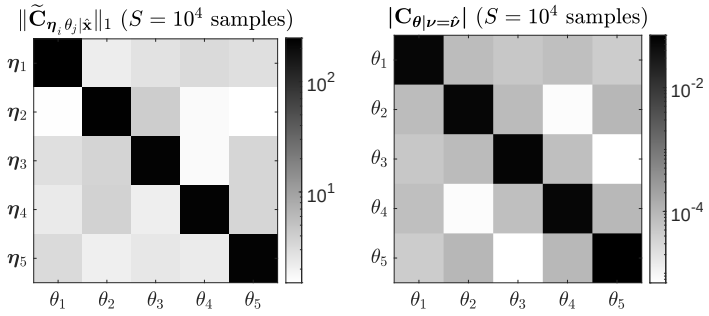


Figure 2. Left: The matrix $\widetilde{\mathbf{C}}_{\boldsymbol{\eta}\boldsymbol{\theta}|\hat{\mathbf{x}}} \in \mathbb{R}^{m\times q}$ is difficult to show because it is very rectangular ("tall and skinny"), and therefore we show the sum of the absolute values of the sample covariance between $\boldsymbol{\eta}_i$ and θ_j, which is identical to $\|\widetilde{\mathbf{C}}_{\boldsymbol{\eta}_i\boldsymbol{\theta}_j|\hat{\mathbf{x}}}\|_1$, shown here for $i, j = 1, 2, 3, 4, 5$. Right: Zoom of the top-left 5×5 submatrix of the computed conditional covariance from (14). The that the colormap in both images is in log-scale.

as assumed by the model. The relaxed conditional covariance is therefore

$$\mathbf{C}^{\alpha}_{\boldsymbol{\theta}|\boldsymbol{\nu}=\hat{\boldsymbol{\nu}}} = \operatorname{diag}(\boldsymbol{\delta}) - \alpha\mathbf{C}^{T}_{\boldsymbol{\eta}\boldsymbol{\theta}|\hat{\mathbf{x}}}(\mathbf{C}_{\boldsymbol{\eta}|\hat{\mathbf{x}}} + \sigma^2\mathbf{I})^{-1}\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\theta}|\hat{\mathbf{x}}}. \tag{17}$$

When $\alpha = 1$ we fully trust the conditional covariance, i.e., $\mathbf{C}^{\alpha}_{\boldsymbol{\theta}|\boldsymbol{\nu}=\hat{\boldsymbol{\nu}}} = \mathbf{C}_{\boldsymbol{\theta}|\boldsymbol{\nu}=\hat{\boldsymbol{\nu}}}$ and when $\alpha = 0$ we do not update the covariance, i.e., $\mathbf{C}^{\alpha}_{\boldsymbol{\theta}|\boldsymbol{\nu}=\hat{\boldsymbol{\nu}}} = \operatorname{diag}(\boldsymbol{\delta})$.

3.2. Block Representation

In this paper we consider CT geometries where in each projection, the position of the source is defined by a single independent angle that needs to be estimated. Therefore the elements of $\boldsymbol{\theta}$ are assumed mutually independent and the mean and covariance defined in (13) and (14) can be computed element-wise.

Our work can immediately be extended to estimation of other parameters in the measurement geometry that have independent realizations for each projection. The importance of this independence can be seen in the block-structure of the estimated covariance matrices as illustrated in Figure 2 where we use the CT set-up in Table 1 with the Shepp-Logan phantom. Note that the colormap here is in log-scale. The matrix $\widetilde{\mathbf{C}}_{\boldsymbol{\eta}\boldsymbol{\theta}|\hat{\mathbf{x}}} \in \mathbb{R}^{m\times q}$ is very rectangular ("tall and skinny") and therefore it is difficult to display it properly; hence the left image in Figure 2 shows the sum of the absolute values of the covariance between $\boldsymbol{\eta}_i$ and θ_j. This amounts to plotting $\|\widetilde{\mathbf{C}}_{\boldsymbol{\eta}_i\theta_j|\hat{\mathbf{x}}}\|_1$ for each pair of $\boldsymbol{\eta}_i$ and θ_j which is done here for $i, j = 1, 2, 3, 4, 5$. This confirms that there is little covariance between $\boldsymbol{\eta}_i$ and θ_j when $j \neq i$ due to the independence of the view angles.

The conditional covariance $\mathbf{C}_{\boldsymbol{\theta}|\boldsymbol{\nu}=\hat{\boldsymbol{\nu}}}$ from (14) has diagonal elements that are much larger than the off-diagonal elements. Therefore the computation of the conditional mean and covariance according to (13) and (14) is carried out element-wise as follows

$$\left(\boldsymbol{\mu}_{\boldsymbol{\theta}|\boldsymbol{\nu}=\hat{\boldsymbol{\nu}}}\right)_i = \mu_{\theta_i} + \mathbf{c}^{T}_{\boldsymbol{\eta}_i\theta_i|\hat{\mathbf{x}}}(\mathbf{C}_{\boldsymbol{\eta}_i|\hat{\mathbf{x}}} + \sigma^2\mathbf{I})^{-1}$$

$$\left(\mathbf{b}_i - \mathbf{R}(\mu_{\theta_i})\,\hat{\mathbf{x}} - \boldsymbol{\mu}_{\boldsymbol{\eta}_i|\hat{\mathbf{x}}}\right), \tag{18}$$

$$\left(\mathbf{C}_{\boldsymbol{\theta}|\boldsymbol{\nu}=\hat{\boldsymbol{\nu}}}\right)_{ii} = \delta_i - \boldsymbol{c}^T_{\boldsymbol{\eta}_i\theta_i|\hat{\mathbf{x}}}(\mathbf{C}_{\boldsymbol{\eta}_i|\hat{\mathbf{x}}} + \sigma^2\mathbf{I})^{-1}\boldsymbol{c}_{\boldsymbol{\eta}_i\theta_i|\hat{\mathbf{x}}}, \tag{19}$$

for $i = 1, \ldots, q$. Here, $\mu_{\theta_i} = (\boldsymbol{\mu}_{\boldsymbol{\theta}})_i$ and $\boldsymbol{c}_{\boldsymbol{\eta}_i\theta_i|\hat{\mathbf{x}}} \in \mathbb{R}^p$ is the ith column vector of the cross-covariance

$$\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\theta}|\hat{\mathbf{x}}} = \begin{bmatrix} \boldsymbol{c}_{\boldsymbol{\eta}_1\theta_1|\hat{\mathbf{x}}} & & \\ & \ddots & \\ & & \boldsymbol{c}_{\boldsymbol{\eta}_q\theta_q|\hat{\mathbf{x}}} \end{bmatrix} \in \mathbb{R}^{p\times q}, \tag{20}$$

and $(\mathbf{C}_{\boldsymbol{\eta}_i|\hat{\mathbf{x}}} + \sigma^2\mathbf{I})^{-1}$ is the sub-matrix of the ith block in

$$(\mathbf{C}_{\boldsymbol{\eta}|\hat{\mathbf{x}}} + \sigma^2\mathbf{I})^{-1} = \begin{bmatrix} (\mathbf{C}_{\boldsymbol{\eta}_1|\hat{\mathbf{x}}} + \sigma^2\mathbf{I})^{-1} & & \\ & \ddots & \\ & & (\mathbf{C}_{\boldsymbol{\eta}_q|\hat{\mathbf{x}}} + \sigma^2\mathbf{I})^{-1} \end{bmatrix}. \tag{21}$$

The smaller covariance sub-matrices can be approximated by the sampling the model discrepancy for each view angle, i.e., for each $i = 1, \ldots, q$ we sample

$$\boldsymbol{\eta}_i^s = \mathbf{R}(\theta_i^s)\,\hat{\mathbf{x}} - \mathbf{R}(\mu_{\theta_i})\,\hat{\mathbf{x}}, \quad s = 1, \ldots, S, \tag{22}$$

where θ_i^s follows the assumed prior distribution $\mathcal{N}(\mu_{\theta_i}, \delta_i)$. We then calculate the blocks of the sample mean and covariance

$$\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}_i|\hat{\mathbf{x}}} = \frac{1}{S}\sum_{s=1}^{S}\boldsymbol{\eta}_i^s, \tag{23}$$

$$\widetilde{\mathbf{C}}_{\boldsymbol{\eta}_i|\hat{\mathbf{x}}} = \frac{1}{S-1}\sum_{s=1}^{S}(\boldsymbol{\eta}_i^s - \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}_i|\hat{\mathbf{x}}})(\boldsymbol{\eta}_i^s - \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}_i|\hat{\mathbf{x}}})^T, \tag{24}$$

as well as the cross-covariance vector

$$\widetilde{\mathbf{c}}_{\boldsymbol{\eta}_i\theta_i|\hat{\mathbf{x}}} = \frac{1}{S-1}\sum_{s=1}^{S}(\boldsymbol{\eta}_i^s - \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}_i|\hat{\mathbf{x}}})(\theta_i^s - \widetilde{\mu}_{\theta_i})^T. \tag{25}$$

with sample mean

$$\widetilde{\mu}_{\theta_i} = \frac{1}{S}\sum_{s=1}^{S}\theta_i^s. \tag{26}$$

Similarly, as we show in [24], the data-fitting term in (10) can be split into q data-fitting terms arriving as the image reconstruction model

$$\mathbf{x}_{\text{MD-TV}} = \arg\min_{\mathbf{x}\geq 0}\frac{1}{2}\sum_{i=1}^{q}\|\mathbf{L}_{\boldsymbol{\nu}_i|\mathbf{x}}(\mathbf{b}_i - \mathbf{R}\left(\mu_{\theta_i}\right)\mathbf{x} - \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}_i|\mathbf{x}})\|_2^2 + \lambda\text{TV}(\mathbf{x}), \tag{27}$$

where the Cholesky factor $\mathbf{L}_{\boldsymbol{\nu}_i|\mathbf{x}}$ is obtained by Cholesky factorization of the inverse covariance $(\widetilde{\mathbf{C}}_{\boldsymbol{\eta}_i|\hat{\mathbf{x}}} + \sigma^2\mathbf{I})^{-1}$. This provides a block representation on both the image reconstruction and view angle estimation problem.

3.3. Alternating scheme for image reconstruction and view angle estimation

The quality of the CT image reconstruction introduced in Section 2 depends on how close the nominal view angles $\boldsymbol{\mu_\theta}$ are to the true underlying view angles that generated the CT data, as well as the corresponding variances $\boldsymbol{\delta}$. Similarly, the quality of the conditional view angles and corresponding variances introduced in this section depends on how close the reconstruction is to the ground truth that generated the CT data. Hence, to take advantage of this relationship we propose an alternating scheme (CT-VAE) shown in Algorithm 1, where we reconstruct the CT image using (10) and estimate the view angles using the conditional mean and variance in (13) and (17) alternately.

The optimization problem with respect to the reconstruction $\mathbf{x}$ in line 12 is solved by using a stochastic primal-dual hybrid gradient algorithm [8]. For more details on solving the CT problem see [24]. The sampling and computation of the conditional mean and covariances in lines 2–7 as well as 8–11 can be carried out in parallel for each $i = 1, \ldots, q$. Note that different number of samples S_{VA} and S_{CT} may be used when sampling the model discrepancy term for the view angle estimation step and image reconstruction step respectively.

4. Numerical Experiments

In this section, we show simulated numerical experiments to illustrate the performance of our method on a CT problem where the view angles are only known with a limited accuracy. All simulation tests were run in MATLAB using the ASTRA toolbox [30] for computing the matrix-free forward and back projections, i.e., multiplication with $\mathbf{R}(\boldsymbol{\theta}) \in \mathbb{R}^{m\times n}$ and its transpose. The CT matrix is obtained by the line model and it represents discretized line integrals parameterized by the view angles $\boldsymbol{\theta} \in \mathbb{R}^q$. That is, letting f be the continuous version of the CT image $\mathbf{x}$, the discretization satisfies

$$(\mathbf{R}(\theta_i)\mathbf{x})_l \approx (\mathcal{R}f)(\theta_i, s_l) = \int_{\mathbb{R}} f(\mathrm{s}_l \overrightarrow{\mathbf{v}}(\theta_i) + t\overrightarrow{\mathbf{v}}^{\perp}(\theta_i))\, \mathrm{d}t, \tag{28}$$

where s_l with $l = 1, \cdots, p$ is the position of the lth pixel on the detector and θ_i with $i = 1, \cdots, q$ are the view angles, and thus $m = qp$. Moreover, $\overrightarrow{\mathbf{v}}(\theta) = (\cos\theta, \sin\theta)$ is the direction of the view angle θ and $\overrightarrow{\mathbf{v}}^{\perp}(\theta) = \overrightarrow{\mathbf{v}}(\theta + \pi/2)$ is the perpendicular direction in which the X-ray is attenuated. The sub-matrix $\mathbf{R}(\theta_i) \in \mathbb{R}^{p\times n}$ represents a single projection at view angle θ_i. In our simulations we use a fan-beam geometry and the nominal view angles $\boldsymbol{\mu_\theta}$ are chosen as equidistant in $[0°, 360°)$. We generate the noisy CT data according to

$$\mathbf{b} = \mathbf{R}(\bar{\boldsymbol{\theta}})\bar{\mathbf{x}} + \bar{\mathbf{e}}, \tag{29}$$

where $\bar{\mathbf{x}}$ is either MATLAB's Shepp-Logan phantom or the "grains" phantom from AIR Tools II [11], which consists of piece-wise constant Voronoi cells. The true underlying view angle $\bar{\theta}_i$ is a realization from the uniform distribution $\mathcal{U}(\mu_{\theta_i} - 2\sqrt{\delta_i}, \mu_{\theta_i} + 2\sqrt{\delta_i})$ for each $i = 1, \ldots, q$ to avoid inverse crime. The prior for each view angle θ_i is $\mathcal{N}(\mu_{\theta_i}, \delta_i)$ and therefore the true view angles fit within a 95% confidence interval of the prior, but

Algorithm 1: Computed Tomography with View Angle Estimation (CT-VAE)

Input : Measured CT data $\mathbf{b}$, CT matrix $\mathbf{R}(\cdot)$, noise variance σ^2, number of outer iterations K, regularization parameter $\lambda > 0$, relaxation parameter $\alpha \in [0, 1]$, initial reconstruction $\mathbf{x}^0$, mean of view angles $\boldsymbol{\theta}^0$ and variance of view angles $\boldsymbol{\delta}^0$.

Output: Final reconstruction $\mathbf{x}^K$, mean of view angles $\boldsymbol{\theta}^K$ and variance of view angles $\boldsymbol{\delta}^K$.

1 **for** $k = 1, 2, \ldots, K$ **do**

// View angle estimation step

2 **for** $i = 1, 2, \ldots, q$ **do**

3 Sample $\boldsymbol{\eta}_i^s = \mathbf{R}(\theta_i^s)\mathbf{x}^k - \mathbf{R}(\theta_i^k)\mathbf{x}^k$, $\theta_i^s \sim \mathcal{N}(\theta_i^k, \delta_i^k)$, $s = 1, \ldots, S_{\text{VA}}$.

4 Calculate $\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}_i|\mathbf{x}^k}$, $\widetilde{\mathbf{C}}_{\boldsymbol{\eta}_i|\mathbf{x}^k}$ and $\widetilde{\boldsymbol{c}}_{\boldsymbol{\eta}_i\theta_i|\mathbf{x}^k}$ from (23), (24) and (25).

5 $\theta_i^{k+1} = \theta_i^k + \widetilde{\boldsymbol{c}}^T_{\boldsymbol{\eta}_i\theta_i|\mathbf{x}^k}(\widetilde{\mathbf{C}}_{\boldsymbol{\eta}_i|\mathbf{x}^k} + \sigma^2\mathbf{I})^{-1}(\mathbf{b}_i - \mathbf{R}(\theta_i^k)\,\mathbf{x}^k - \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}_i|\mathbf{x}^k})$.

6 $\delta_i^{k+1} = \delta_i^k - \alpha\widetilde{\boldsymbol{c}}^T_{\boldsymbol{\eta}_i\theta_i|\hat{\mathbf{x}}}(\widetilde{\mathbf{C}}_{\boldsymbol{\eta}_i|\hat{\mathbf{x}}} + \sigma^2\mathbf{I})^{-1}\widetilde{\boldsymbol{c}}_{\boldsymbol{\eta}_i\theta_i|\hat{\mathbf{x}}}$

7 **end**

// Image reconstruction step

8 **for** $i = 1, 2, \ldots, q$ **do**

9 Sample $\boldsymbol{\eta}_i^s = \mathbf{R}(\theta_i^s)\mathbf{x}^k - \mathbf{R}(\theta_i^{k+1})\mathbf{x}^k$, $\theta_i^s \sim \mathcal{N}(\theta_i^{k+1}, \delta_i^{k+1})$, $s = 1, \ldots, S_{\text{CT}}$.

10 Calculate $\widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}_i|\mathbf{x}^k}$ from (23).

11 Calculate the Cholesky factor $\mathbf{L}_{\boldsymbol{\nu}_i|\mathbf{x}^k}$ of $(\widetilde{\mathbf{C}}_{\boldsymbol{\eta}_i|\hat{\mathbf{x}}} + \sigma^2\mathbf{I})^{-1}$ from (24).

12 **end**

13 $\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}\geq 0} \frac{1}{2}\sum_{i=1}^{q} \|\mathbf{L}_{\boldsymbol{\nu}_i|\mathbf{x}^k}(\mathbf{b}_i - \mathbf{R}(\theta_i^{k+1})\,\mathbf{x} - \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\eta}_i|\mathbf{x}^k})\|_2^2 + \lambda\text{TV}(\mathbf{x})$

14 **end**

are drawn from a different distribution. Finally the measurement noise $\bar{\mathbf{e}}$ is a realization from $\mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ with $\sigma = 0.005\|\mathbf{R}(\bar{\boldsymbol{\theta}})\bar{\mathbf{x}}\|_2/\sqrt{m}$. The physical parameters of the CT model are summarized in Table 1. We note in particular that the image reconstruction problems here are under-determined since $n = 128^2$ and $m = pq$ with $p = 128$ and $q = 90$, which could not be dealt with in [24].

We use $S_{\text{VA}} = S_{\text{CT}} = 100$ samples, but note that different number of samples could be used when sampling the model discrepancy in the two stages, i.e., for the view angle estimation (line 2) and image reconstruction (line 9). We set $K = 10$ outer iterations and relaxation parameter $\alpha = 0.5$. We apply the SPDHG algorithm [8] to solve the minimization problem for reconstructing $\mathbf{x}$ (line 12), and stop the inner iterations when the relative change in $\mathbf{x}$ is small, i.e., when $\|\mathbf{x}^l - \mathbf{x}^{l-1}\|_2/\|\mathbf{x}^l\|_2 < 10^{-5}$ where l is the iteration index for the inner loop. We found no significant difference in reconstruction quality when using 10^{-5} and a smaller relative error 10^{-6} (which was used in [24]). We compare a reconstruction $\mathbf{x}$ to the ground truth $\bar{\mathbf{x}}$ using relative error (RE) defined as

Table 1. The physical and discretization parameters in the simulated CT experiments.

Parameter	Value
Scan geometry	Fan-beam
Reconstruction domain size	50 cm × 50 cm
Source to center distance	50 cm
Source to detector distance	100 cm
Detector length	130 cm
Image pixels	$n = 128^2$
Detector pixels	$p = 128$
Number of view angles	$q = 90$
View angle standard deviation	$\sqrt{\delta} = 1^\circ$

$\|\mathbf{x} - \bar{\mathbf{x}}\|_2 / \|\bar{\mathbf{x}}\|_2$ and compare the view angles $\boldsymbol{\theta}$ to the true underlying view angles $\bar{\boldsymbol{\theta}}$ by the mean absolute difference $\sum_{i=1}^{q} |\theta_i - \bar{\theta}_i|/q$ or direct numerical difference $\theta_i - \bar{\theta}_i$.

To benchmark the performance of our method, we compare the reconstruction obtained from our new method CT-VAE with the ones obtained from our previous algorithm MD-TV [24] from (10). We also compare with a classical non-negative TV regularized reconstruction using the nominal view angles $\boldsymbol{\mu_\theta}$ (L2-TV) as well as the true underlying view angles $\bar{\boldsymbol{\theta}}$ (L2-TV-opt), obtained from solving the optimization problems

$$\mathbf{x}_{\text{L2-TV}} = \arg\min_{\mathbf{x}\geq 0} \frac{1}{2\sigma^2} \sum_{i=1}^{q} \|\mathbf{b}_i - \mathbf{R}(\mu_{\theta_i})\,\mathbf{x}\|_2^2 + \lambda\,\text{TV}(\mathbf{x}), \tag{30}$$

$$\mathbf{x}_{\text{L2-TV-opt}} = \arg\min_{\mathbf{x}\geq 0} \frac{1}{2\sigma^2} \sum_{i=1}^{q} \|\mathbf{b}_i - \mathbf{R}(\bar{\theta}_i)\,\mathbf{x}\|_2^2 + \lambda\,\text{TV}(\mathbf{x}). \tag{31}$$

Here, $\mathbf{x}_{\text{L2-TV-opt}}$ represents the optimal reconstruction obtainable with the non-negative TV prior, since in (31) the true underlying view angles $\bar{\boldsymbol{\theta}}$ are used.

Finally, since our method also estimates the view angles in the CT problems, at the end of this section we also compare our view angle estimation results with the projection matching method proposed in [31] that solves for the best-matching projections using a gradient descent method.

4.1. Comparison of reconstructions

Figure 3 shows the best reconstructions obtained by selecting the optimal regularisation parameter that gave rise to the lowest relative error RE from our new method CT-VAE, our previous method MD-TV, as well as the L2-TV and L2-TV-opt methods. The images all have the same colormap. Comparing CT-VAE and L2-TV-opt with L2-TV and MD-TV, we see that the last two look visually inferior and have a larger relative error than the first two. Note especially the sharper white boundary in the Shepp-Logan phantom and the sharper edges in the grains phantom. In particular, our

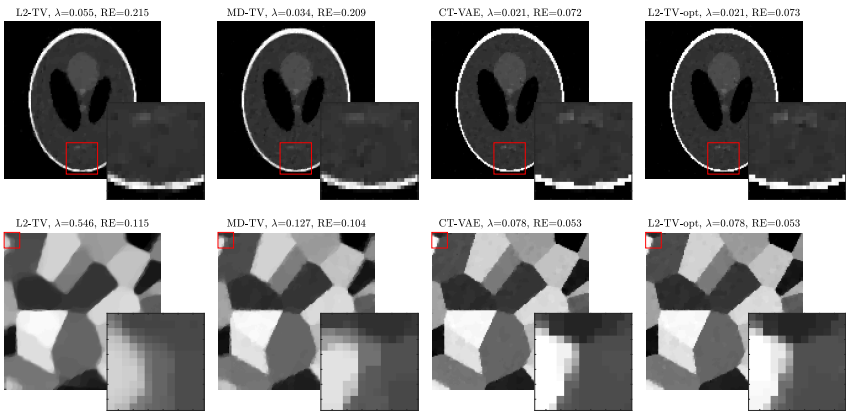


Figure 3. Best reconstructions obtained by selecting the optimal regularization parameter in terms of relative error for our method (CT-VAE) in Algorithm 1 compared to our previous method MD-TV (10) in [24], L2-TV (30) and L2-TV-opt (31) for the Shepp-Logan phantom (top) and the grains phantom (bottom).

method provides reconstructions that are very similar to the benchmark reconstruction obtained with the true view angles (L2-TV-opt) both in terms of visual quality and relative error. Especially, with respect to the relative error our method performs the same or even slightly better than L2-TV-opt.

This comparison shows that for the uncertain view angle CT problem (2) our method outperforms the methods that do not use view angle estimation, and it can provide reconstructions with the same quality as the ones obtained using the true view angles. Furthermore, there is also a difference in the value of the regularization parameter λ, where L2-TV and MD-TV both require more regularization – i.e., larger λ – to avoid artefacts from reconstructing with the nominal view angles.

In Figure 4 we compare the reconstruction quality in terms of relative error with respect to varying regularization parameters. We can see that in terms of relative error our method provides comparable results to the benchmark (L2-TV-opt) reconstructions except for really small regularization parameters. In particular, our method and L2-TV-opt have the same optimal regularization parameter. The reason is that any remaining uncertainty in the view angle estimate is dealt with by the image reconstruction step from Section 2 and therefore additional regularization from the prior is not needed. Furthermore, our method always provides reconstructions with lower relative errors compared to L2-TV and MD-TV. In the case with small regularization parameters the reconstructions are under-regularized, i.e., covered by artefacts from measurement noise and view angle errors, and those artefacts would have a strong effect on the results of the conditional mean and covariance and leads to poor view angle estimates. On the other hand, when we use very large regularization parameters, the reconstructions are

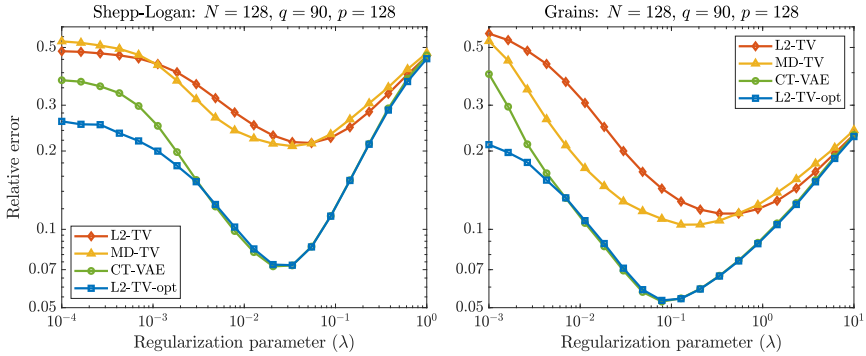


Figure 4. Comparison of relative error vs regularization parameter for the Shepp-Logan phantom (left) and the grains phantom (right) for our method (CT-VAE) in Algorithm 1 compared to our previous method MD-TV (10) in [24], L2-TV (30) and L2-TV-opt (31).

so over-regularized that the uncertainty in the view angles is negligible, therefore all four methods end up with the same relative errors. From these results it is also evident that joint estimation of the view angles and CT image provides better reconstructions with lower relative error compared to reconstruction obtained using the nominal view angles.

4.2. View angle estimation

In addition to reconstructing the CT image, our new method CT-VAE also estimates the view angles using the conditional mean (13) and covariance (17). In Figure 5 we show these estimates of the view angles obtained from our method and compare with the prior mean and covariance. It is obvious that for both the Shepp-Logan and the grains phantoms the estimated view angles are significantly closer to the true underlying view angles compared with the prior. In the prior, the errors in the mean of the view angles are in the range $[0, 2.5°]$, but using our method the range of the errors is reduced to $[0, 0.15°]$. Furthermore, the variance estimates accurately reflect the distance from the true underlying view angles, i.e., the true view angles lie within a 99% confidence interval in the figure.

In Figure 6 we can also compare the average angle error $\sum_{i=1}^{q} |\theta_i^K - \bar{\theta}_i|/q$ from the conditional mean (13) in terms of varying quality of reconstruction. This is achieved by varying the regularization parameter for the TV prior. Here we see that the smallets angle error is obtained at around the same regularization parameters that provide the best CT reconstruction (smallest relative error RE) and that the angle error is the same for a large range of regularization parameters. The average angle error of the nominal view angles is $\sum_{i=1}^{q} |\mu_{\theta_i} - \bar{\theta}_i|/q = 0.90°$, hence even for the largest and smallest regularization parameters the estimated view angles are closer to the true underlying

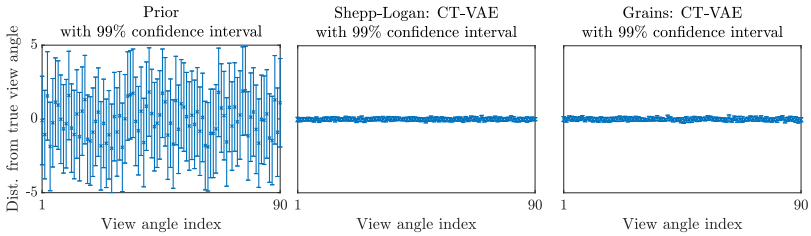


Figure 5. Distance to the true view angles and 99% confidence interval for the prior (left), and CT-VAE estimates for the Shepp-Logan (middle) and grains (right) phantoms. The scale on the y-axis is the same in all three figures.

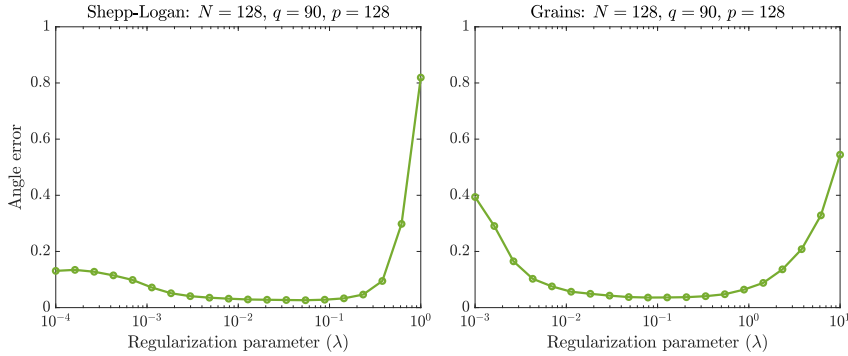


Figure 6. The angle estimation error $\sum_{i=1}^{q} |\theta_i^K - \bar{\theta}_i|/q$ from CT-VAE vs the regularization parameter λ in reconstructing $\mathbf{x}$.

view angles than the nominal view angles. This shows that the angle estimation of our method is robust towards the quality of the CT reconstruction. Essentially, if the reconstruction is not heavily over- or under-regularized, we are able to obtain a good angle estimate.

4.3. Convergence history

Here we consider the convergence history of our method and compare the computational work of different parts of the algorithm. For this purpose we use the unit of an *epoch* defined as the work involved in one multiplication with $\mathbf{R}(\cdot)$ or its transpose, since the computation is dominated by the forward and back projections as discussed in [24].

Figure 7 shows the convergence history of our method in terms of relative errors for the CT reconstruction and angle estimation errors. We note that the horizontal lines of no change in relative error happen when sampling the model-discrepancy in Algorithm 1 (lines 3 and 9), since this does not change $\mathbf{x}$ but requires forward evaluations of $\mathbf{R}(\cdot)$. The beginning of the horizontal lines is where a new outer iteration starts. We see

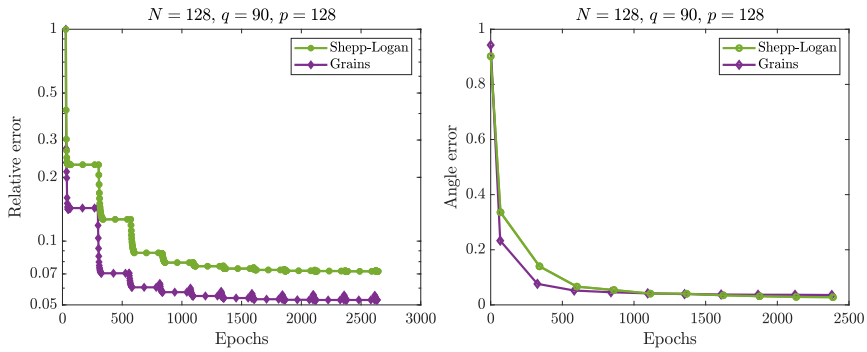


Figure 7. The convergence history of the CT-VAE method.

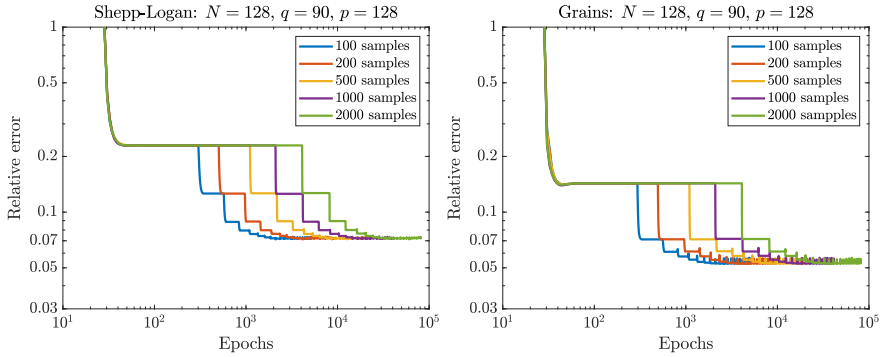


Figure 8. The convergence history of $\mathbf{x}$ for varying number of samples $S_{\text{VA}} = S_{\text{CT}}$. Note that we use log-scale for epochs (x-axis) in these plots.

that for both phantoms the relative error and the angle error levels off after at most 10 iterations.

4.4. Number of model-discrepancy samples

Figure 8 shows the convergence history of $\mathbf{x}$ for varying number of samples $S_{\text{VA}} = S_{\text{CT}}$. We see that the relative error of the final reconstruction does not improve when using more than 100 samples, so we use that in our experiments. If we use fewer than 100 samples the covariances are no longer accurately estimated by the sample covariances (24) and (25). This leads to the term $\widetilde{\boldsymbol{c}}_{\boldsymbol{\eta}_i\theta_i|\hat{\mathbf{x}}}^T(\widetilde{\mathbf{C}}_{\boldsymbol{\eta}_i|\hat{\mathbf{x}}} + \sigma^2\mathbf{I})^{-1}\widetilde{\boldsymbol{c}}_{\boldsymbol{\eta}_i\theta_i|\hat{\mathbf{x}}}$ in the conditional covariance estimate (17) to sometimes become larger than δ_i and thus giving a covariance estimate that is negative. Since $S = 100$ was found to be a good sample size in [24] for this particular problem, we do not explore the sample size selection further.

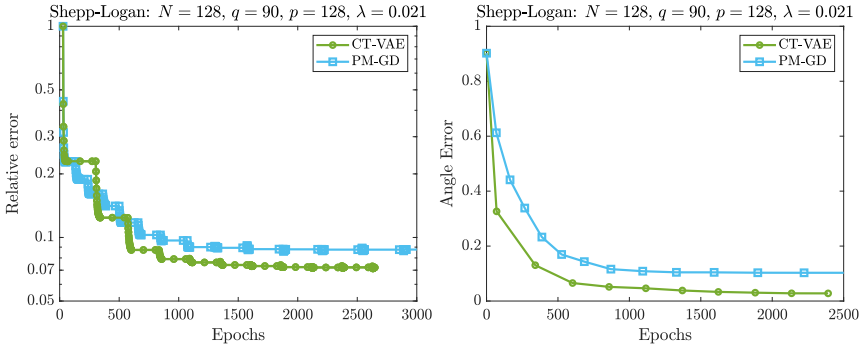


Figure 9. The convergence history of our method CT-VAE and the projection matching method with gradient descent methods PM-GD.

4.5. Comparison with projection matching using gradient descent

We conclude by comparing our method with the projection matching scheme proposed in [31]. The proposed scheme solves the non-linear least-squares optimization problem

$$\min_{\boldsymbol{\theta}, \mathbf{x}>\mathbf{0}} \frac{1}{2\sigma^2} \|\mathbf{b} - \mathbf{R}(\boldsymbol{\theta})\,\mathbf{x}\|_2^2 + \lambda \mathrm{TV}(\mathbf{x}). \tag{32}$$

The algorithm to solve (32) is based on variable projection and essentially alternates between an image reconstruction step and a gradient step for the view angles. Hence, this alternating scheme is similar to our method, and the main difference lies in how the reconstruction and the view angles are estimated. For our implementation of the projection matching scheme, we use SPDHG for the image reconstruction step and a simple gradient step with line search for the view angle estimation step.

In Figure 9 we compare the convergence history of our method with the projection matching scheme using gradient descent (PM-GD). We see that both in terms of relative error vs epochs and angle error vs epochs our method provides better results except in the first few iterations where our method spends computational work sampling the model discrepancy. Furthermore, we see that our method reaches a lower relative error and angle error compared with PM-GD. To further investigate why our method yields a lower final relative error, we plot the objective function (32) with respect to θ_i given the CT reconstruction $\mathbf{x}^0$ as well as the angle estimates from our method and the projection matching scheme. Note that the objective function is highly non-convex with many local minima. It is therefore clear that unless the initial view angles are close to the true view angles, the projection matching scheme can get stuck in local minima.

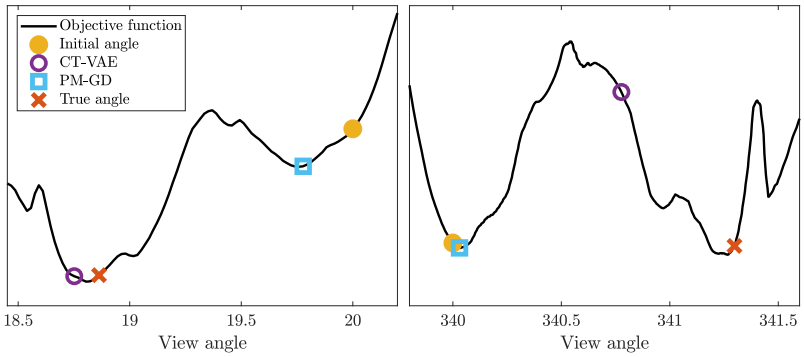


Figure 10. Comparison of the values obtained in the first angle estimation step θ^1 given the initial CT reconstruction $\mathbf{x}^0$ for our method (CT-VAE) and the projection matching scheme from [31] with gradient descent (PM-GD) for the Shepp-Logan example with $\lambda = 0.021$. The objective function used for the projection matching scheme (32) is shown in grey to illustrate the non-convexity of the objective.

5. Conclusion

We propose a new model and a corresponding algorithm for CT reconstruction with an added view angle estimation component, in order to handle uncertain angles (other geometric uncertainties can be handled in a similar way and are in some cases a straightforward extension of this work). We show numerically that our method is able to outperform gradient descent-based projection-matching methods for the joint image/angle reconstruction problem. In particular, our method, with a TV prior that reflects our phantoms, is able to provide reconstructions that are of similar quality as ones obtained with the exact view angles, which is the best-case scenario. Furthermore, our method is able to run on large-scale CT problems because of an inherent block-structure in our CT model. At this time, we can demonstrate numerically that our method converges; it is a topic of future research to prove this rigorously.

References

[1] Arridge S R, Kaipio J P, Kolehmainen V, Schweiger M, Somersalo E, Tarvainen T and Vauhkonen M 2006 Approximation errors and model reduction with an application in optical diffusion tomography *Inverse Problems* **22** 175–95

[2] Basu S and Bresler Y 2000 Uniqueness of tomography with unknown view angles *IEEE Trans. Image Process.* **9** 1094–106

[3] Benning M and Burger M 2018 Modern regularization methods for inverse problems *Acta Numer.* **27** 1–111

[4] Brandt C and Seppänen A 2018 Recovery from errors due to domain truncation in magnetic particle imaging: approximation error modeling approach *J. Math. Imaging Vis.* **60** 1196–208

[5] Buzug T M 2008 *Computed Tomography: From Photon Statistics to Modern Cone-Beam CT* (Berlin: Springer-Verlag Berlin Heidelberg)

[6] Calvetti D, Dunlop M, Somersalo E and Stuart A 2018 Iterative updating of model error for Bayesian inversion *Inverse Problems* **34** 025008

[7] Clackdoyle R and Desbat L 2015 Data consistency conditions for truncated fanbeam and parallel projections *Med. Phys.* **42** 831–45

[8] Ehrhardt M J, Markiewicz P J and Schönlieb C B 2019 Faster PET reconstruction with non-smooth priors by randomization and preconditioning *Phys. Med. Biol.* **64** 225019

[9] Fang Y, Murugappan S and Ramani K 2010 Estimating view parameters from random projections for tomography using spherical MDS *BMC Med. Imaging* **10** 12

[10] Hansen T M, Cordua K S, Jacobsen B H and Mosegaard K 2014 Accounting for imperfect forward modeling in geophysical inverse problems – exemplified for crosshole tomography *Geophysics* **79** H1–21

[11] Hansen P C and Jørgensen J S 2017 AIR Tools II: algebraic iterative reconstruction methods, improved implementation *Numer. Algorithms* **79** 107–37

[12] Kaipio J and Somersalo E 2005 *Statistical and Computational Inverse Problems* (New York: Springer)

[13] Kaipio J and Somersalo E 2007 Statistical inverse problems: discretization, model reduction and inverse crimes *J. Comput. Appl. Math.* **198** 493–504

[14] Kennedy M C and O'Hagan A 2002 Bayesian calibration of computer models *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **63** 425–64

[15] Kolehmainen V, Tarvainen T, Arridge S R and Kaipio J P 2011 Marginalization of uninteresting distributed parameters in inverse problems – application to diffuse optical tomography *Int. J. Uncertain. Quantif.* **1** 1–17

[16] Natterer F 1986 *The Mathematics of Computerized Tomography* (Chicago: Wiley)

[17] Niebler S, Schömer E, Tjaden H, Schwanecke U and Schulze R 2019 Projection-based improvement of 3D reconstructions from motion-impaired dental cone beam CT data *Med. Phys.* **46** 4470–80

[18] Nissinen A, Heikkinen L M and Kaipio J P 2007 The Bayesian approximation error approach for electrical impedance tomography – experimental results *Meas. Sci. Technol.* **19** 015501

[19] Nissinen A, Kolehmainen V and Kaipio J P 2011 Reconstruction of domain boundary and conductivity in electrical impedance tomography using the approximation error approach *Int. J. Uncertain. Quantif.* **1** 203–22

[20] Mallick S P, Agarwal S, Kriegman D J, Belongie S J, Carragher B and Potter C S 2006 Structure and view estimation for tomographic reconstruction: a Bayesian approach *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* eds Fitzgibbon A, Taylor C J and Lecun Y (New York: IEEE) pp 2253–60

[21] Parkinson Y D, Knoechel C, Yang C, Larabell C A and Le Gros M A 2012 Automatic alignment and reconstruction of images for soft X-ray tomography *J. Struct. Biol.* **177** 259–66

[22] Peng C B, Rodi W L and Toksöz M N 1993 A Tikhonov regularization method for image reconstruction *Acoustical Imaging* eds Wei Y and Gu B (Berlin: Springer) pp 153–63

[23] Pulkkinen A, Kolehmainen V, Kaipio J P, Cox B T, Arridge S R and Tarvainen T 2014 Approximate marginalization of unknown scattering in quantitative photoacoustic tomography *Inverse Probl. Imaging* **8** 811–29

[24] Riis N A B, Dong Y and Hansen P C 2020 Computed tomography reconstruction with uncertain view angles by iteratively updated model discrepancy *J. Math. Imaging Vis.* 1–11

[25] Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica D* **60** 259–68

[26] Sidky E Y and Pan X 2008 Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization *Phys. Med. Biol.* **53** 4777–807

[27] Song K, Comolli L R and Horowitz M 2012 Removing high contrast artifacts via digital inpainting in cryo-electron tomography: an application of compressed sensing *J. Struct. Biol.* **178** 108–20

[28] Tick J, Pulkkinen A and Tarvainen T 2019 Modelling of errors due to speed of sound variations in photoacoustic tomography using a Bayesian framework *Biomed. Phys. Eng. Express* **6** 015003

[29] Tikhonov A N 1963 Solution of incorrectly formulated problems and the regularization method *Soviet Math. Dokl.* **4** 1035–38; 1963 English translation of Dokl. Akad. Nauk. SSSR **151** 501–4

[30] Van Aarle W, Palenstijn W J, Cant J, Janssens E, Bleichrodt F, Dabravolski A and Sijbers J 2016 Fast and flexible X-ray tomography using the ASTRA toolbox *Opt. Express* **24** 25129–47

[31] Van Leeuwen T, Maretzke S and Batenburg K J 2018 Automatic alignment for three-dimensional tomographic reconstruction *Inverse Problems* **34** 024004

[32] Wang J, Karen Chen Y C, Yuan Q, Tkachuk A, Erdonmez C, Hornberger B and Feser M 2012 Automated markerless full field hard X-ray microscopic tomography at sub-50 nm 3-dimension spatial resolution *Appl. Phys. Lett.* **100** 143107

[33] Wang C C, Chiang C C, Liang B, Yin G C, Weng Y T and Wang L C 2017 Fast Projection Matching for X-ray Tomography *Sci. Rep.* **7** 3691

APPENDIX D

Supplementary material

D.1 Overview of simulated CT experiment

The simulations are carried out in MATLAB using the ASTRA Toolbox [1] for computation of the CT forward and back projections. As in the papers the data is generated from

$$\mathbf{b} = \mathbf{A}_{\bar{\boldsymbol{\theta}}}\bar{\mathbf{x}} + \bar{\mathbf{e}}. \tag{D.1}$$

In this case for $\bar{\mathbf{x}}$ we use the grains phantom from AIR Tools II [28] similar to the one used in the paper. The true view angles $\bar{\theta}_i$ are sampled from

$$\mathcal{U}(\mu_{\theta_i} - 2, \mu_{\theta_i} + 2), \tag{D.2}$$

to avoid inverse crime, since the prior of the view angles is Gaussian, $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \operatorname{diag}(\boldsymbol{\delta}))$. Here $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ is the nominal view angles and $\delta_i = 1°$. The measurement noise is sampled from $\mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ with $\sigma = 0.005\|\mathbf{A}_{\bar{\boldsymbol{\theta}}}\bar{\mathbf{x}}\|_2/\sqrt{m}$. The CT problem is of size $n = 512 \times 512$ with $m = p \times q$ with $p = 512$ and $q = 90$.

D.2 (Stochastic) Primal Dual Hybrid Gradient algorithm for CT

In this thesis we consider a variational regularization problem on the form

$$\mathbf{x}_{\text{MD-TV}} = \arg\min \frac{1}{2}\sum_{i=1}^{q} \|\widetilde{\mathbf{L}}_{\boldsymbol{\nu}_i|\mathbf{x}}(\mathbf{b}_i - \mathbf{A}_{\hat{\theta}_i}\mathbf{x} - \widetilde{\boldsymbol{\mu}}_{\boldsymbol{\nu}_i|\mathbf{x}})\|_2^2 + \lambda \text{TV}(\mathbf{x}) + g_+(\mathbf{x}), \tag{D.3}$$

This is solved using the Primal Dual Hybrid Gradient [17] and Stochastic Primal Dual Hybrid Gradient [18, 21] algorithms. These algorithms solve the generic (non-smooth) optimization problem

$$\mathbf{x}^* \in \arg\min_{\mathbf{x}\mathbb{R}^n} \left\{ \sum_{i=1}^{q+1} f_i(\mathbf{B}_i\mathbf{x}) + g(\mathbf{x}) \right\}, \tag{D.4}$$

where in our case

$$f_{1,\dots,q}(\cdot) = \tfrac{1}{2}\|\cdot + \widetilde{\mathbf{L}}_{\boldsymbol{\nu}_i|\mathbf{x}}(\mathbf{b}_i - \boldsymbol{\mu}_{\boldsymbol{\nu}_i|\mathbf{x}})\|_2^2, \tag{D.5}$$

$$f_{q+1}(\cdot) = \lambda \|\cdot\|_{2,1}, \tag{D.6}$$

$$\mathbf{B}_i = -\widetilde{\mathbf{L}}_{\boldsymbol{\nu}_i|\mathbf{x}}\mathbf{A}_{\hat{\theta}_i} \tag{D.7}$$

$$\mathbf{B}_q + 1 = \nabla \tag{D.8}$$

$$g(\cdot) = g_+(\cdot). \tag{D.9}$$

The PDHG and SPDHG algoritms are summarized below.

PDHG

Inputs: Initial $\mathbf{x}$, step parameters σ_i, τ.
Output: $\mathbf{x}$

$\mathbf{y} = \mathbf{0}$, $\mathbf{z} = \bar{\mathbf{z}} = \mathbf{B}^T\mathbf{y} = \mathbf{0}$
for $k = 1, \dots$
  $\mathbf{x} = \text{prox}_g^\tau(\mathbf{x} - \tau\bar{\mathbf{z}})$
  for $i = 1, \dots, q+1$
    $\mathbf{y}_i^+ = \text{prox}_{f_i^*}^{\sigma_i}(\mathbf{y}_i - \sigma_i\mathbf{B}_i\mathbf{x})$
  $\boldsymbol{\Delta}\mathbf{z} = \sum_{i=1}^{q+1}\mathbf{B}_i^T(\mathbf{y}_i^+ - \mathbf{y}_i)$
  $\mathbf{z} = \mathbf{z} + \boldsymbol{\Delta}\mathbf{z}, \quad \mathbf{y} = \mathbf{y}^+$
  $\bar{\mathbf{z}} = \mathbf{z} + \boldsymbol{\Delta}\mathbf{z}$
end

SPDHG

Inputs: Initial $\mathbf{x}$, step parameters σ_i, τ.
Output: $\mathbf{x}$

$\mathbf{y} = \mathbf{0}$, $\mathbf{z} = \bar{\mathbf{z}} = \mathbf{B}^T\mathbf{y} = \mathbf{0}$
for $k = 1, \dots$
  $\mathbf{x} = \text{prox}_g^\tau(\mathbf{x} - \tau\bar{\mathbf{z}})$
  Select i with probability P_i
  $\mathbf{y}_i^+ = \text{prox}_{f_i^*}^{\sigma_i}(\mathbf{y}_i - \sigma_i\mathbf{B}_i\mathbf{x})$
  $\boldsymbol{\Delta}\mathbf{z} = \mathbf{B}_i^T(\mathbf{y}_i^+ - \mathbf{y}_i)$
  $\mathbf{z} = \mathbf{z} + \boldsymbol{\Delta}\mathbf{z}, \quad \mathbf{y} = \mathbf{y}^+$
  $\bar{\mathbf{z}} = \mathbf{z} + \frac{1}{P_i}\boldsymbol{\Delta}\mathbf{z}$
end

In the above $\text{prox}_f^S(x)$ defines the proximal operator of f with step size S given by

Definition 3. *Let S be a symmetric and positive definite matrix. Then we define the proximal operator of f with step size S as*

$$\text{prox}_f^S(x) := \arg\min_z \left\{\|z - x\|_S^2 + f(z)\right\}. \tag{D.10}$$

For our case this is given by

$$\text{prox}_{f_i^*}^\sigma(\mathbf{y}) = \frac{1}{1+\sigma}\left(\mathbf{y} + \sigma\widetilde{\mathbf{L}}_{\boldsymbol{\nu}_i|\mathbf{x}}(\mathbf{b}_i - \boldsymbol{\mu}_{\boldsymbol{\nu}_i|\mathbf{x}})\right), i = 1, \dots, q, \tag{D.11}$$

$$\text{prox}_{f_{q+1}^*}^\sigma(\mathbf{y}) = \frac{1}{\lambda}\frac{\mathbf{y}}{\max(1, \|\mathbf{y}\|_2)} \tag{D.12}$$

$$\text{prox}_g^\tau(\mathbf{x}) = \max(\mathbf{x}, \mathbf{0}). \tag{D.13}$$

We refer to the discussion in Paper B for step-sizes τ, σ_i, and selection probability P_i.

D.3 Definitions used in Section 4.1.2

We start with some necessary definitions.

Definition 4. *For a norm on $\mathbb{R}^n$ we define its dual, denoted $\|\cdot\|_*$, to be*

$$\|x\|_* := \max_{y\in\mathbb{R}^n} \frac{y^T x}{\|y\|}. \tag{D.14}$$

For a norm on $\mathbb{R}^{m\times n}$ we define its dual, denoted $\|\cdot\|_*$, to be

$$\|\Delta\|_* := \max_{A\in\mathbb{R}^{m\times n}} \frac{\langle A,\Delta\rangle}{\|A\|}, \quad \langle A,\Delta\rangle = Tr(A^T\Delta). \tag{D.15}$$

Example 4. *The ℓ_p norms $\|x\|_p^p := \sum_i |x_i|^p$ for $p\in[1,\infty)$ and $\|x\|_\infty := \max_i |x_i|$ satisfy the well-known duality relation*

$$\ell_{p^*} \text{ is dual to } \ell_p, \text{ where } p^*\in[1,\infty] \text{ with } \tfrac{1}{p}+\tfrac{1}{p^*}=1. \tag{D.16}$$

We call p^* the conjugate of p. Finally, we need some definitions of matrix norms:

Definition 5. *We define the following matrix norms.*

- *The p-Frobenius norm:*
$$\|\Delta\|_{F_p}^p := \sum_{ij} |\Delta_{ij}|^p. \tag{D.17}$$
- *The p-spectral norm:*
$$\|\Delta\|_{\sigma_p} := \|\boldsymbol{\mu}(\Delta)\|_p, \tag{D.18}$$
where $\boldsymbol{\mu}(\Delta)$ denotes the vector containing the singular values of Δ.
- *The h,g-induced norm (where h and g are norms):*
$$\|\Delta\|_{(h,g)} := \max_{x\in\mathbb{R}^n} \frac{g(\Delta x)}{h(x)}. \tag{D.19}$$

D.4 Definitions used in Section 3.2

This section is qouted from earlier work of the author [54]. Thus, it does not stand as any contribution to the thesis, but is included for completeness.

Definition 6. *The Schwartz space $\mathcal{S}(\mathbb{R}^n)$ is defined as the set of all functions $f\in C^\infty(\mathbb{R}^n)$ for which*

$$\|f\|_{\alpha,\beta} := \sup_{x\in\mathbb{R}^n} \left|x^\beta D^\alpha f(x)\right| < \infty \tag{D.20}$$

for all multi-indices $\alpha,\beta\in\mathbb{N}_0^n$, where $D^\alpha = \partial^\alpha/\partial x^\alpha$.

Definition 7 (Fourier Transform)**.** *The Fourier transform of a function $f \in L^1(\mathbb{R}^n)$ is defined by the operator $\mathcal{F} : L^1(\mathbb{R}^n) \to C_0(\mathbb{R}^n)$ of a function*

$$\mathcal{F}f(\xi) := \hat{f}(\xi) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} f(x) e^{-i\langle x,\xi\rangle} \mathrm{d}x. \tag{D.21}$$

Definition 8 (Inverse Fourier Transform)**.** *The inverse Fourier transform of a function $g \in L^1(\mathbb{R}^n)$ is defined by*

$$\mathcal{F}^{-1}g(x) := (2\pi)^{-n/2} \int_{\mathbb{R}^n} g(\xi) e^{i\langle \xi,x\rangle} \mathrm{d}\xi. \tag{D.22}$$

Hence, if $\hat{f} \in L^1(\mathbb{R}^n)$ then the equality $\mathcal{F}^{-1}\hat{f}(x) = f(x)$ holds. In particular it can be shown that if $f \in \mathcal{S}(\mathbb{R}^n)$ then $\hat{f} \in \mathcal{S}(\mathbb{R}^n) \subset L^1(\mathbb{R}^n)$.

One of the most central results in tomography is the Fourier Slice Theorem. It relates the Radon transform of a function to the function itself by Fourier transforms. In the following we use the 1D Fourier transform of $\mathcal{R}f(\phi, s)$ along s and denote it by $\widehat{\mathcal{R}f}(\phi, \sigma) = \mathcal{F}_s\mathcal{R}f(\phi, \sigma)$. Whenever an angle, ϕ, is fixed we denote the Radon transform for a fixed angle by $\mathcal{R}_\phi f(s)$.

Theorem 8 (Fourier Slice Theorem)**.** *For $f \in \mathcal{S}(\mathbb{R}^n)$ and $\phi \in S^{n-1}$ we have*

$$\widehat{\mathcal{R}_\phi f}(\sigma) = (2\pi)^{(n-1)/2} \hat{f}(\sigma\phi), \quad \sigma \in \mathbb{R}. \tag{D.23}$$

Proof. Applying the definition of the one-dimensional Fourier transform of $\mathcal{R}_\phi f$ over s and then the Radon transform we get

$$\begin{aligned}
\widehat{\mathcal{R}_\phi f}(\sigma) &= (2\pi)^{-1/2} \int_{\mathbb{R}} \mathcal{R}_\phi f(s) e^{-is\sigma} \mathrm{d}s \\
&= (2\pi)^{-1/2} \int_{\mathbb{R}} \int_{\phi^\perp} f(\phi s + y) \mathrm{d}y \quad e^{-is\sigma} \mathrm{d}s \\
&= (2\pi)^{-1/2} \int_{\mathbb{R}} \int_{\phi^\perp} e^{-is\sigma} f(\phi s + y) \mathrm{d}y \mathrm{d}s.
\end{aligned}$$

Now, let $\varphi : \phi^\perp \times \mathbb{R} \to \mathbb{R}^n$, be defined by $\varphi(y, s) = s\phi + y$, and let $u = (y, s)$. Note that $s = x \cdot \phi = \phi \cdot x$ over $\phi^\perp$ and that $\phi^\perp \times \mathbb{R}$ is an open set. Then by change of variables we see that

$$\begin{aligned}
\widehat{\mathcal{R}_\phi f}(\sigma) &= (2\pi)^{-1/2} \int_{\mathbb{R}} \int_{\phi^\perp} e^{-is\sigma} f(\varphi(u)) \mathrm{d}u \\
&= (2\pi)^{-1/2} \frac{1}{|\det D\varphi(u)|} \int_{\mathbb{R}^n} e^{-i\sigma\phi\cdot x} f(x) \mathrm{d}x.
\end{aligned}$$

Noting that $|\det D\varphi(u)| = 1$, we get the desired result

$$\begin{aligned}\widehat{\mathcal{R}_\phi f}(\sigma) &= (2\pi)^{-1/2}(2\pi)^{n/2}\left((2\pi)^{-n/2}\int_{\mathbb{R}^n} e^{-i\sigma\phi\cdot x} f(x)\mathrm{d}x\right)\\ &= (2\pi)^{(n-1)/2}\hat{f}(\sigma\phi).\end{aligned}$$

□

Using the Fourier Slice Theorem, one can derive an inversion formula for the Radon transform.

Theorem 9 (Inverse Radon Transform)**.** *For $f \in \mathcal{S}(\mathbb{R}^n)$ and the Radon transform $\mathcal{R}f(\phi, s)$, $\phi \in S^{n-1}$, $s \in \mathbb{R}$ we have that f is given by*

$$f(x) = \frac{1}{2}(2\pi)^{-n+1/2}\int_{S^{n-1}}\int_{-\infty}^{\infty} \mathcal{F}_s\mathcal{R}f(\phi,\sigma)e^{i\sigma\langle x,\phi\rangle}|\sigma|^{n-1}\mathrm{d}\sigma\mathrm{d}\phi. \tag{D.24}$$

Proof. Assume $f \in \mathcal{S}(\mathbb{R}^n)$. Then by applying the inverse Fourier transform on $\hat{f} \in \mathcal{S}(\mathbb{R}^n)$, we can write

$$f(x) = (\mathcal{F}^{-1}\hat{f})(x) = (2\pi)^{-n/2}\int_{\mathbb{R}^n}\hat{f}(\xi)e^{i\langle x,\xi\rangle}\mathrm{d}\xi. \tag{D.25}$$

Let $U = S^{n-1}\times(0,\infty)$ and $V = \mathbb{R}^n\backslash\{0\}$. Note, U and V are open sets and since the singleton $\{0\}$ has measure zero in $\mathbb{R}^n$ the integral over U and $\mathbb{R}^3$ are equal. Furthermore, let

$$\varphi : U \to V, \quad (\sigma,\phi) \mapsto \sigma\phi = \xi \tag{D.26}$$

for $\phi \in S^{n-1}$ and $\sigma \in (0,\infty)$. Then we can write any $\xi \in V$ as $\xi = \varphi(\sigma,\phi) = \sigma\phi$. Noting that $|\det[(D\varphi)(\sigma,\phi)]| = |\sigma|^{n-1}$ we have by change of variables on (D.25) the following.

$$f(x) = (2\pi)^{-n/2}\int_{S^{n-1}}\int_0^{\infty}\hat{f}(\sigma\phi)e^{i\langle x,\sigma\phi\rangle}|\sigma|^{n-1}\mathrm{d}\sigma\mathrm{d}\phi. \tag{D.27}$$

The Fourier Slice Theorem 8 and $\mathcal{R}f$ being even then yields

$$f(x) = 2\pi^{-n+1/2}\int_{S^{n-1}}\int_0^{\infty}\mathcal{F}_s\mathcal{R}f(\phi,\sigma)e^{i\langle x,\sigma\phi\rangle}|\sigma|^{n-1}\mathrm{d}\sigma\mathrm{d}\phi \tag{D.28}$$

$$= \frac{1}{2}(2\pi)^{-n+1/2}\int_{S^{n-1}}\int_{-\infty}^{\infty}\mathcal{F}_s\mathcal{R}f(\phi,\sigma)e^{i\langle x,\sigma\phi\rangle}|\sigma|^{n-1}\mathrm{d}\sigma\mathrm{d}\phi \tag{D.29}$$

□

Bibliography

[1] W. van Aarle et al. “Fast and flexible X-ray tomography using the ASTRA toolbox.” In: *Optics express* 24.22 (2016), pages 25129–25147.

[2] S. R. Arridge et al. “Approximation errors and model reduction with an application in optical diffusion tomography.” In: *Inverse problems* 22.1 (2006), page 175.

[3] J. M. Bardsley. *Computational uncertainty quantification for inverse problems*. Volume 19. Computational Science and Engineering. SIAM, 2018.

[4] J. M. Bardsley and P. C. Hansen. “MCMC algorithms for computational UQ of nonnegativity constrained linear inverse problems.” In: *SIAM Journal on Scientific Computing* 42.2 (2020), A1269–A1288.

[5] S. Basu and Y. Bresler. “Uniqueness of tomography with unknown view angles.” In: *IEEE Transactions on Image Processing* 9.6 (2000), pages 1094–1106.

[6] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust optimization*. Volume 28. Princeton Series in Applied Mathematics. Princeton University Press, 2009.

[7] M. Benning and M. Burger. “Modern regularization methods for inverse problems.” In: *Acta Numerica* 27 (2018), pages 1–111.

[8] D. Bertsimas and M. S. Copenhaver. “Characterization of the equivalence of robustification and regularization in linear and matrix regression.” In: *European Journal of Operational Research* 270.3 (2018), pages 931–942.

[9] G. E. P. Box. “Robustness in the strategy of scientific model building.” In: *Robustness in statistics*. Edited by R. L. Launer and G. N. Wilkinson. Academic Press, 1979, pages 201–236.

[10] G. E. P. Box. “Science and statistics.” In: *Journal of the American Statistical Association* 71.356 (1976), pages 791–799.

[11] C. Brandt and A. Seppänen. “Recovery from errors due to domain truncation in magnetic particle imaging: approximation error modeling approach.” In: *Journal of Mathematical Imaging and Vision* 60 (2018), pages 1196–1208.

[12] L. Bungert et al. “Variational regularisation for inverse problems with imperfect forward operators and general noise models.” In: *Inverse Problems* 36.12 (2020), page 125014.

[13] M. Burger, Y. Korolev, and J. Rasch. “Convergence rates and structure of solutions of inverse problems with imperfect forward models.” In: *Inverse Problems* 35.2 (2019), page 024006.

[14] T. M. Buzug. *Computed tomography: from photon statistics to modern cone-beam CT*. Springer, Berlin, Heidelberg, 2008.

[15] D. Caviedes-Nozal et al. “A Bayesian spherical harmonics source radiation model for sound field control.” In: *The Journal of the Acoustical Society of America* 146.5 (2019), pages 3425–3435.

[16] D. Caviedes-Nozal et al. “Gaussian processes for sound field reconstruction.” In: *The Journal of the Acoustical Society of America* (2021), *to appear*.

[17] A. Chambolle and T. Pock. “A first-order primal-dual algorithm for convex problems with applications to imaging.” In: *Journal of mathematical imaging and vision* 40 (2011), pages 120–145.

[18] A. Chambolle et al. “Stochastic primal-dual hybrid gradient algorithm with arbitrary sampling and imaging applications.” In: *SIAM Journal on Optimization* 28.4 (2018), pages 2783–2808.

[19] R. Clackdoyle and L. Desbat. “Data consistency conditions for truncated fan-beam and parallel projections.” In: *Medical physics* 42.2 (2015), pages 831–845.

[20] Y. Dong et al. “Fixing nonconvergence of algebraic iterative reconstruction with an unmatched backprojector.” In: *SIAM Journal on Scientific Computing* 41.3 (2019), A1822–A1839.

[21] M. J. Ehrhardt, P. Markiewicz, and C.-B. Schönlieb. “Faster PET reconstruction with non-smooth priors by randomization and preconditioning.” In: *Physics in Medicine & Biology* 64.22 (2019), page 225019.

[22] Y. Fang, S. Murugappan, and K. Ramani. “Estimating view parameters from random projections for tomography using spherical MDS.” In: *BMC Medical Imaging* 10 (2010), pages 1–9.

[23] R. D. Fierro et al. “Regularization by truncated total least squares.” In: *SIAM Journal on Scientific Computing* 18.4 (1997), pages 1223–1241.

[24] G. H. Golub, P. C. Hansen, and D. P. O’Leary. “Tikhonov regularization and total least squares.” In: *SIAM journal on matrix analysis and applications* 21.1 (1999), pages 185–194.

[25] G. H. Golub and C. F. van Loan. “An analysis of the total least squares problem.” In: *SIAM journal on numerical analysis* 17.6 (1980), pages 883–893.

[26] J. Hadamard. *Lectures on Cauchy’s Problem in Linear Partial Differential Equations*. 1923.

[27] P. C. Hansen. “Perturbation bounds for discrete Tikhonov regularisation.” In: *Inverse problems* 5.4 (1989), page L41.

[28] P. C. Hansen and J. S. Jørgensen. "AIR Tools II: algebraic iterative reconstruction methods, improved implementation." In: *Numerical Algorithms* 79 (2018), pages 107–137.

[29] Per Christian Hansen. *Discrete inverse problems: insight and algorithms*. Volume 7. Fundamentals of Algorithms. SIAM, 2010.

[30] S. van Huffel and P. Lemmerling. *Total least squares and errors-in-variables modeling: analysis, algorithms and applications*. Springer Science & Business Media, 2013.

[31] J. Kaipio and E. Somersalo. *Statistical and computational inverse problems*. Volume 160. Applied Mathematical Sciences. Springer Science & Business Media, 2006.

[32] M. C Kennedy and A. O'Hagan. "Bayesian calibration of computer models." In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.3 (2001), pages 425–464.

[33] V. Kolehmainen et al. "Marginalization of uninteresting distributed parameters in inverse problems-application to diffuse optical tomography." In: *International Journal for Uncertainty Quantification* 1.1 (2011).

[34] Y. Korolev. "Making use of a partial order in solving inverse problems: II." In: *Inverse Problems* 30.8 (2014), page 085003.

[35] Y. Korolev and A. Yagola. "Making use of a partial order in solving inverse problems." In: *Inverse Problems* 29.9 (2013), page 095012.

[36] V. P. Krishnan and E. T. Quinto. "Microlocal analysis in tomography." In: *Handbook of mathematical methods in imaging*. Edited by O. Scherzer. Springer, New York, NY, 2015, pages 847–902.

[37] M. Lassas and S. Siltanen. "Can one use total variation prior for edge-preserving Bayesian inversion?" In: *Inverse Problems* 20.5 (2004), page 1537.

[38] T. van Leeuwen, S. Maretzke, and K. J. Batenburg. "Automatic alignment for three-dimensional tomographic reconstruction." In: *Inverse Problems* 34.2 (2018), page 024004.

[39] C. F. van Loan and G. H. Golub. *Matrix computations*. Johns Hopkins University Press Baltimore, 1983.

[40] R. B. Madsen and T. M. Hansen. "Estimation and accounting for the modeling error in probabilistic linearized amplitude variation with offset inversion." In: *Geophysics* 83.2 (2018), N15–N30.

[41] E. Malhotra and A. Rajwade. "Tomographic reconstruction from projections with unknown view angles exploiting moment-based relationships." In: *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2016, pages 1759–1763.

[42] S. P. Mallick et al. “Structure and view estimation for tomographic reconstruction: A Bayesian approach.” In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. IEEE. 2006, pages 2253–2260.

[43] I. Markovsky and S. van Huffel. “Overview of total least-squares methods.” In: *Signal processing* 87.10 (2007), pages 2283–2302.

[44] F. Natterer. *The mathematics of computerized tomography*. SIAM, 2001.

[45] Elmar Niebler S.and Schömer et al. “Projection-based improvement of 3D reconstructions from motion-impaired dental cone beam CT data.” In: *Medical physics* 46.10 (2019), pages 4470–4480.

[46] A. Nissinen, L. M. Heikkinen, and J. P. Kaipio. “The Bayesian approximation error approach for electrical impedance tomography – experimental results.” In: *Measurement Science and Technology* 19.1 (2007), page 015501.

[47] A. Nissinen, V. Kolehmainen, and J. P. Kaipio. “Reconstruction of domain boundary and conductivity in electrical impedance tomography using the approximation error approach.” In: *International Journal for Uncertainty Quantification* 1.3 (2011).

[48] D. Y. Parkinson et al. “Automatic alignment and reconstruction of images for soft X-ray tomography.” In: *Journal of structural biology* 177.2 (2012), pages 259–266.

[49] C. Peng, W. L. Rodi, and M. N. Toksöz. “A Tikhonov regularization method for image reconstruction.” In: *Acoustical Imaging*. Springer, 1993, pages 153–164.

[50] A. Pulkkinen et al. “Approximate marginalization of unknown scattering in quantitative photoacoustic tomography.” In: *Inverse Problems and Imaging* 8.3 (2014), page 811.

[51] J. Radon. “On the determination of functions from their integral values along certain manifolds.” In: *IEEE transactions on medical imaging* 5.4 (1986), pages 170–176.

[52] N. A. B. Riis and Y. Dong. “A New Iterative Method for CT Reconstruction with Uncertain View Angles.” In: *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer. 2019, pages 156–167.

[53] N. A. B. Riis, Y. Dong, and P. C. Hansen. “Computed Tomography Reconstruction with Uncertain View Angles by Iteratively Updated Model Discrepancy.” In: *Journal of Mathematical Imaging and Vision* (2020), pages 1–11.

[54] N. A. B. Riis and J. Frøsig. *Computed Tomography for Region-of-Interest Problems with Limited Data*. M.Sc. thesis at Technical University of Denmark. 2017.

[55] N. A. B. Riis et al. “Limited-data X-ray CT for underwater pipeline inspection.” In: *Inverse Problems* 34.3 (2018), page 034002.

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

[57] M. Schabel. *3D Shepp-Logan phantom*. Online; MATLAB Central File Exchange. Retrieved February 3, 2021.

[58] E. Y. Sidky and X. Pan. “Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization.” In: *Physics in Medicine & Biology* 53.17 (2008), page 4777.

[59] K. Song, L. R. Comolli, and M. Horowitz. “Removing high contrast artifacts via digital inpainting in cryo-electron tomography: an application of compressed sensing.” In: *Journal of structural biology* 178.2 (2012), pages 108–120.

[60] G. W. Stewart. *Matrix perturbation theory*. Computer Science and Scientific Computing. Academic Press, 1990.

[61] A. Tarantola. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005.

[62] J. Tick, A. Pulkkinen, and T. Tarvainen. “Modelling of errors due to speed of sound variations in photoacoustic tomography using a Bayesian framework.” In: *Biomedical Physics & Engineering Express* 6.1 (2019), page 015003.

[63] A. N. Tihonov. “Solution of incorrectly formulated problems and the regularization method.” In: *Soviet Math.* 4 (1963), pages 1035–1038.

[64] C.-C. Wang et al. “Fast projection matching for X-ray tomography.” In: *Scientific reports* 7.1 (2017), pages 1–10.

[65] G. Welch and G. Bishop. “An introduction to the Kalman filter.” 1997.

[66] H. Xu, C. Caramanis, and S. Mannor. “Robust Optimization in Machiine Learning.” eng. In: *Optimization for Machine Learning*. MIT Press, 2011, pages 369–399.

{2.7182818284}