THE UNIVERSITY OF CHICAGO

INVESTIGATION OF OPTIMIZATION-BASED ALGORITHMS FOR TOMOGRAPHIC IMAGE RECONSTRUCTION

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

COMMITTEE ON MEDICAL PHYSICS

BY
ZHENG ZHANG

CHICAGO, ILLINOIS
JUNE 2016

To my parents

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Towards the end of accomplishing my thesis, I feel that I would never have been able to finish my dissertation without the guidance of my mentor and committee members, the help from colleagues and friends, and support from my parents.

I would like to express my deepest gratitude to my advisor, Dr. Xiaochuan Pan, for introducing me into the field of medical physics, inspiring me with his rich knowledge, teaching me about critical thinking, providing me with an excellent atmosphere for studying, and offering me with his unwavering support. Looking back of my Ph.D. life, I feel fortunate to be under the guidance of Dr. Pan, because not only the expertise I learned from him, but also the serious and rigorous attitude for doing research. I sincerely appreciate his selfless help.

I would like to thank my committee members, Dr. Chien-Min Kao, Dr. Patrick La Riviere, Dr. Charles A. Pelizzari, and Dr. Emil Y. Sidky, for their inspiring advice for improving my dissertation. In particular, I want to thank Dr. Kao for the insightful discussion on the PET projection; I want to thank Dr. Patrick La Riviere and Dr. Charles A. Pelizzari for showing me the clinical significance of optimization-based reconstruction; and special thanks to Dr. Emil Y. Sidky, for leading me into the forefront of optimization-based reconstruction in medical imaging. I would also like to thank Dr. Ingrid Reiser for helping me on a side project on radiation shielding calculation.

I am thankful to many former and current colleagues in the group: Dr. Xiao Han, Dr. Junguo Bian, Dr. Erik A. Pearson, Dr. Adrian Sanchez, Mr. Buxin Chen, Mr. Andrew Davis, and Mr. Sean Rose. I learned a lot through the discussion with them.

I am also grateful to the support from all previous and current collaborators, including: Dr. Joseph Manak from GE global research for providing GE C-arm data; Dr. Alex Zamyatin from Toshiba medical research institute for providing 320-slice diagnostic-CT data; Dr. Budi Kusnoto from University of Illinois at Chicago for providing dental CBCT data; Dr. Lingxiong Shao, Dr. Jinghan Ye, and Dr. Chi-Hua Tung for offering the sparse-PET idea

and providing numerous PET data sets.

Ultimately, I would like to thank my parents for their understanding and unconditional love throughout all my studies. None of my achievements have been possible without their support.

ABSTRACT

Among all tomographic imaging devices, X-ray cone-beam computed tomography (CBCT) and positron emission tomography are two important tools used widely in numerous medical imaging applications, and both have enjoyed tremendous progress in both hardware and reconstruction algorithm development in the past few decades. Advanced optimization-based algorithms, which have gained fast development recently in CBCT imaging, have demonstrated the capability of accommodating a variety of CBCT imaging configurations, improving image quality, and exploiting image-sparsity properties with the intention of reducing data sampling. In this work, we continue investigating optimization-based reconstructions in CBCT for solving existing, practical issues. Meanwhile, we leverage the experience in CBCT image reconstruction from sparse data, and develop advanced algorithms for enabling the design of innovative PET systems with sparsely populated detectors, while not significantly compromising the PET capability and image quality.

In this dissertation, we investigate numerous optimization programs and solve them by using advanced iterative algorithms. Results of the work suggest that, with appropriately designed optimization programs, optimization-based reconstruction tools can be obtained for *1*) reducing the artifacts existing in the off-middle planes of FDK reconstruction from short-scan CBCT data, *2*) enabling a sparse-PET configuration with reduced crystal while not significantly compromising the image quality, and *3*) allowing iterative reconstruction based on an image with variable resolution. In addition, results indicate that the selection of program/algorithm parameters may have significant impact on the outcome.

CHAPTER 1

INTRODUCTION

1.1 Background: cone-beam computed tomography

1.1.1 Overview of CBCT

X-ray computed tomography (CT), since its introduction in 1972 [1, 2], has turned into an important non-invasive tool in medical imaging to assess the internal structure of the scanned object. It has been widely used in disease screening, diagnosis, and other applications requiring image guidance [3–22]. In the past several decades, CT has enjoyed tremendous progress in both hardware and reconstruction algorithm development, enabling a variety of configurations, improving image quality, and facilitating a wide spectrum of clinical applications.

Recently, cone-beam computed tomography (CBCT) has become an emerging tool in numerous medical imaging applications [23–34]. CBCT consists of a rotating gantry to which an X-ray tube and detector are fixed, and a divergent pyramidal- or cone-shape X-ray beam that transmits through a scanned subject onto a detector. The X-ray tube and detector rotate around a rotation axis following a complete, or partial circular trajectory, and multiple, sequential projection data are acquired during the rotation. CBCT, unlike a conventional, fan-beam CT, acquires a volume of field-of-view (FOV) with only one rotation, and the acquired data are used for reconstructing the three-dimensional volume of the FOV. The advantages of such systems comparing to fan-beam CT scanners include reduction of scan time, reduction of latent artifacts due to patient motion, and boosting X-ray tube efficiency.

In general, there exist two types of CBCT scanners: the first type, such as the C-arm system, the on-board-imager (OBI) system, or the dental CBCT scanner, *etc.*, employ a flat-panel detector [35], which is mounted on a gantry with lower mechanical precision compared to conventional CT scanners; the other type is the state-of-the-art multi-slice CT

(MSCT) scanner, with curved- or spherical-detectors consisting of 256 or 320 slices along axial direction. Although differences exist between the two types of CBCT systems, similar reconstruction algorithms can be developed and adapted to both.

1.1.2 CBCT applications

CBCT scanners possess great flexibility in scanning configuration and yield images containing information of high practical utility. Therefore, they are used in a rapidly increasing number of imaging applications.

Diagnostic imaging Many state-of-the-art diagnostic CT scanners employ multi-slice detectors, such as 256- or 320-slices [36–38], which can be used for the circular cone-beam scan configuration. A great body of work has investigated and demonstrated that MSCT images following circular scanning protocols have virtually equal performance with CT images following helical scanning protocols [39–41]. MSCT is widely used in abdomen/brain disease diagnosis and treatment planning. One of the important applications of MSCT is cardiac imaging [42, 43], which requires high spatial and high temporal resolution. Due to the heartbeat, cardiac imaging often requires several rotations in a short time to reconstruct a series of images in which doctors can observe a dynamic procedure. MSCT with sufficient axial coverage is capable of reconstructing a 3D-volume image including the heart in one rotation (0.35-0.5 seconds), meeting the needs of cardiac imaging.

Image guided surgery During the past decade or so, there has been tremendous interest in using C-arm systems as a CBCT unit to yield tomographic images [26, 44–47], which are reconstructed from projection data acquired over an angular range of as small as π plus fan angle (known as a short-scan). These advantages make C-arm CBCT an enabling technique for a number of image guided surgery (IGS) applications, including tumor resection and endovascular coiling of brain aneurysms during brain surgery. For the whole workflow in

IGS, a C-arm CBCT scanner is used before surgery for procedure planning, during surgery for intraoperative imaging guidance, and after surgery for surgery assessment.

Image-guided radiation therapy Development of CBCT for image-guided radiation therapy (IGRT) is a rapidly growing field [31, 32, 48–50], where CBCT is used for patient positioning and monitoring the variation of a tumor's size or shape [34]. The IGRT is performed in such a way that a CBCT scan of the patient on the treatment couch is obtained before radiation treatment, and the reconstructed CBCT image can then be registered to the CT image taken earlier for treatment planning to facilitate precise repositioning of the patient to the treatment machine [51–57]. Advantages of IGRT technology are high spatial resolution, sufficient soft-tissue contrast, and imaging in treatment position.

General dentistry and orthodontics Dental CBCT, as an emerging imaging modality in dentistry and orthodontics, has enjoyed huge growth and widespread acceptance in the dental field [58–62]. Dental CBCT can provide details of teeth and orofacial bone structures with high spatial resolution, as well as 3D information of their position and orientation, which allow precise understanding of the relationship of the structures in an anatomically complex region. Therefore, the dental CBCT systems have great advantages and have been widely used in many applications in dentistry, for example, examination of teeth and facial structures for orthodontic treatment planning, planning for teeth extraction, assessment of the jaws for placement of dental implants, *etc.*

Additional CBCT applications During the past few years, many CBCT systems have been developed for facilitating other imaging modalities and for specific organ imaging. Examples include CBCT as a component of multimodality systems (e.g., PET-CT and SPECT-CT) [63–67], pre-clinical CBCT systems such as micro-CT [68, 69], breast CT [59, 70–72], extremity CBCT [73–76], and neuro CBCT [77–79]. In particular, dedicated breast CBCT, which provides 3D-volume information of the breast, has demonstrated the potential to

overcome some issues existing in mammography [80] and tomosynthesis [81].

1.1.3 CBCT reconstruction

There exist two general approaches to CBCT image reconstruction: analytic-based and optimization-based reconstruction methods, which are based upon different imaging models, and thus employ different algorithms for image reconstruction.

Analytic-based reconstruction Analytic-based algorithms, such as filtered-backprojection (FBP) or Feldkamp-Davis-Kress (FDK) algorithm [82], are based upon continuous-to-continuous (C-C) linear imaging models, and are used for reconstructing continuous images from continuous data. A general form of an analytic imaging model can be written as:

$$g = \mathcal{F}(f), \tag{1.1}$$

where f is a analytically described image, and the linear function $\mathcal{F}$ represents the X-ray transform and generates the projection data g from image f. Analytic algorithms are then derived in an attempt to invert, or pseudo-invert, function $\mathcal{F}$, which solves Eq. (1.1), yielding:

$$f^\star = \mathcal{U}(g). \tag{1.2}$$

Note that, if $\mathcal{F}^{-1}$, the inversion of function $\mathcal{F}$, exists, then $\mathcal{U} = \mathcal{F}^{-1}$ and $f^\star = f$, which means that the image can be accurately reconstructed; if $\mathcal{F}^{-1}$ does not exist, $f^\star$ can be an approximation of f. Discrete forms of an analytic-based algorithm can be devised so that it can be applied to discrete data in realistic tomographic imaging. Because discretization is applied only to the data space, the algorithms can still yield continuous image reconstructions.

For certain imaging configurations, for example, a fan-beam configuration, or a helical cone-beam configuration, an exact, closed-form solution can be derived to invert the imag-

ing model in Eq. (1.1) [83, 84]. However, for a circular cone-beam configuration, no exact reconstruction algorithm can be derived because of the lack of Tuy's data sufficiency condition [85]. In this scenario, an algorithm named FDK was developed by Feldkamp, Davis, and Kress for approximate image reconstruction, which has been widely used for clinical CBCT image reconstruction.

Further physical, mathematical, and/or statistical considerations often call for pre-conditioning the data by multiplication of a weighting matrix with the data. Among all possible weighting functions, those for normalizing the redundancy in the X-ray transform, a CBCT imaging model, have been used as a leading motivation in image reconstruction [86].

Optimization-based reconstruction Optimization-based (i.e., iterative) algorithms are based on a discrete-to-discrete (D-D) imaging model, which reconstructs discrete images from discrete data. A general form of D-D imaging model can be expressed as:

$$\mathbf{g} = \mathcal{H}\mathbf{f}, \tag{1.3}$$

where $\mathbf{f}$ is a discrete image vector, $\mathbf{g}$ a discrete data vector, and $\mathcal{H}$ a system matrix approximating the X-ray transform.

Based upon the imaging model defined in Eq. (1.3), various optimization programs can be defined, and numerous optimization-based algorithms derived to solve them, including project-onto-convex-set (POCS) [87], expectation-maximization (EM) [88], simultaneous algebraic reconstruction technique (SART) [89], and other algorithms solving programs with image constraints. Optimization-based algorithms generally possess a higher degree of flexibility than analytic-based algorithms in accommodating image reconstruction for a wide variety of imaging conditions of practical significance. With the rapidly increasing computation power and significant advances in optimization theory for linear imaging models, there has been a substantially increased interest in developing and applying optimization-based algorithms to image reconstruction in CBCT.

Optimization-based reconstruction, despite its potential advantages, has not yet been accepted as a practical approach to image reconstruction in various applications. A leading reason for this is that it generally requires substantially longer computational time, and higher computer-memory consumption, than FDK-based reconstruction. This issue becomes even more strenuous when one is interested in images with high spatial resolution represented with an array consisting of a huge number of small-size voxels.

1.2 Background: positron emission tomography

1.2.1 Overview of PET

Positron emission tomography (PET), since its introduction to medical imaging application in the late 1960's, has become a powerful tool in nuclear medicine and functional imaging applications [90–99], because it offers the ability to observe metabolic processes in the body. Before a PET scan, positron-emitting radioactive tracer is injected into the scanned subject. The emitted positrons then annihilate with an electron within the subject to form a pair of photons of 511 keV energy, which travel in opposite direction to form a line-pair. If the line-pair of photons intersect with the PET detector, we refer to this detected line-pair as a coincident event. Note that both photons of the line-pair can be attenuated within the scanned subject.

In general, a PET scanner consists of one or more rings of detector, which is usually circle- or polygon-shape. Between two adjacent rings, if there exist "septa," which are composed of dense materials and stop photons traveling between rings, we referred to this mode as 2D configuration; if no septa exist, we refer to that mode as 3D configuration because coincident event can be detected between rings. The detector materials in PET are scintillation crystals, including bismuth germinate (BGO), lutetium oxyorthosilicate (LSO), and sodium iodide (NaI), which convert the 511 keV photons to low-energy scintillation light. Conventional PET scanners employ photomultiplier tubes (PMTs) to collect scintillation light and detect

the 511-keV photon interaction. The most state-of-the-art PET scanners employ digital detectors, which are capable of determining a photon interaction energy and location with one detector that is linked to a single crystal or a few tiled crystals. Therefore, in digital PET systems, different crystals are independent.

If a pair of detected coincident photons travel directly from an annihilation event, we refer to it as a true coincident event. However, there exist another two forms of additive events which corrupt the true measurements: scatter and randoms events. For a scatter event, one or both of the annihilation photons has undergone Compton scattering, and the detected line-pair is no long the true line-pair. For a random event, the detection is from two independent photons but within the coincidence timing window. Note that the two photons in a scatter event still originate from the same annihilation event, but those in a random event do not.

Back in the 1960's, at the early stage in PET development, people realized that it may be beneficial to utilize the time-of-flight (TOF) information in PET image reconstruction [100]. In the early 1980's, a fully functional TOF-PET was built. The concept of TOF-PET is that we measure the difference in detection time for a pair of coincident photons, and the first detected photon is nearer to the detector than the other. This TOF information can help determine more accurately the location of the annihilated events along the line.

1.2.2 PET applications

Since its inception, PET has been used in a wide spectrum of medical imaging applications such as oncology, neuro-imaging, cardiology imaging, infectious disease diagnosis, pharmacokinetics, and small-animal imaging. We briefly describe below some of the applications.

Oncology In clinical oncology, PET scanning with radio tracer F^{18} fluorodeoxyglucose (FDG), which is referred to as FDG-PET, is widely used. FDG-PET has demonstrated the ability to detect and stage cancers/tumor, and to monitor their treatment. Comparing

to CT scan, FDG-PET shows enhanced sensitivity in detecting/staging of several types of lymphomas and solid cancers, such as lung, breast, colon, and head and neck carcinoma [90, 94, 98, 101].

Neuro-imaging The brain consumes considerable glucose and oxygen, and the uptake of glucose or oxygen is impacted by the brain activity and possible brain disease. In practice, PET scan utilizing O^{15} is often employed in neuro-imaging, which is in an attempt to monitor the brain activity [92,96,102], and FDG-PET has shown the possibility to investigate Alzheimer's disease [103, 104].

Cardiology In clinical cardiology, a PET scan is used to diagnose cardiac disease and detect areas of low blood flow in the heart, as well as identify abnormal heart muscle. Radiotracers with different isotopes are used in cardiac PET imaging, such as Rb^{82}, N^{13}, and O^{15}, which are for different clinical purposes [105–108].

Small animal imaging Dedicated PET scanners designed for small-animal imaging are often referred to as micro-PET [109, 110], which has been extensively used in bio-medical research. Compared to a regular human PET scanner, micro-PET scans subjects of size much smaller than a human, and the fine structures of small animals require high spatial resolution, for example, a sub-millimeter level, to identify the organs of interest. As a result, the size of micro-PET and the the size of its crystals should be adapted to yield such a high-resolution demand.

1.2.3 PET reconstruction

Similarly to CT image reconstruction, PET image reconstruction can also utilize analytic algorithms, such as FBP. However, due to the complex data rebinning and inferior noise property in the reconstruction, FBP has become less used in the clinic nowadays. To date, PET image reconstruction is usually performed by using iterative methods, and the maximum-

likelihood expectation-maximization (MLEM) algorithm [111,112] and its variants have been the dominant methods used in practice. The EM algorithm is a specific procedure for finding an image that maximizes the likelihood of the measurement data that are assumed to be mutually independent Poisson variates. While the algorithm is favored due to its quantitative properties, it has slow convergence and typically the converged solution appears undesirably noisy. Hudson and Larkin introduced in 1994 the ordered-subsets (OS) version of the EM algorithm that can considerably accelerate the reconstruction [113]. This OSEM algorithm, however, no longer guarantees convergence. To address this issue, Brown and De Pierro have developed the row-action ML algorithm (RAMLA) [114] and Hsiao *et al* the convergent OSEM (COSEM) algorithm [115] that guarantee convergence; subsequently, enhancements to these algorithms have been introduced [116–118]. The EM and OSEM based algorithms have also been extended for reconstructing list-mode data [119–122], 4D spatial-temporal data, and parameter images [123–133]. To reduce image noise, the popular noise-regularization strategy is to modify the likelihood function by adding *a priori* condition on the solution to seek the maximum a posteriori (MAP) or penalized ML (PML) solution [111–114,134–141]. Generally, the *a priori* term favors smooth images and initially a quadratic function on the difference between neighboring pixel intensity was used [142, 143]. The use of this *a priori* term often leads to images that are over-smoothed and have spatially variant resolution. Therefore, various priors to preserve edges [141, 144–146] and achieve uniform and isotropic image resolution [126, 147] have been investigated. MAP or PML methods have also been applied for list-mode, spatial-temporal and parametric image reconstruction [129, 148, 149].

1.3 Focus of work

1.3.1 Artifact reduction in short-scan CBCT image reconstruction

In short-scan circular CBCT imaging, the Parker weighting function or its variants can appropriately normalize the redundancy of the X-ray transform for the middle plane, i.e., the plane containing the scanning trajectory, and diagonal weighing matrices designed based upon the Parker weighting functions have been shown to yield adequate reconstructions within the middle plane from discrete data by use of existing analytic algorithms in discrete forms. However, the X-ray transform possesses no redundancy for planes parallel to the middle plane, and diagonal weighting matrices such as those devised for the middle plane may lead to reconstruction artifacts in off-middle planes, depending upon the amount of data, subject structure, and distances of the non-middle planes to the middle plane. Optimization-based algorithms, from previous experience, may have the potential to reduce such artifacts. Therefore, in chapters 2 and 3, we investigate different designs of optimization-based reconstruction along with different weighting functions, and study how they impact artifact reduction, especially in the off-middle planes.

1.3.2 Algorithm enabled image reconstruction from sparse-PET data

In the most state-of-the-art digital PET scanners, the detector materials constitute the most significant cost portion, and therefore it is interesting to consider alternate PET design with a reduced number of detectors. However, this will result in reduced measurements as well as degraded image quality. Recently developed optimization-based algorithms may be used for enabling the design and assessment of innovative PET systems, while not significantly sacrificing the PET capability and image quality.

In particular, among all the optimization-based reconstructions, we consider optimization programs with image total-variation (TV) constraints, because it has demonstrated the potential for exploiting image sparsity. In chapters 4 and 5, we design an innovative

sparse-PET configuration, and investigate optimization-based reconstructions with image-TV constraint. The purpose of the study is to examine whether an image can yield potential utilities with a sparse-PET configuration, and how image quality is affected by different designs of optimization-based reconstruction.

1.3.3 Optimization-based reconstruction on images with variable resolution

In numerous CBCT applications, one is often interested in detailed information only within a particular region-of-interest (ROI) and rough knowledge outside the ROI. Yet, information is sought only about an ROI containing the tumor or disease, while some reference knowledge about the region outside the ROI could suffice. Therefore, one may use an image array in which the ROI and the rest of the support of the imaged subject are represented with fine and coarse voxels, respectively. The use of such an image representation, which is referred to as an image array with variable resolution, can substantially lower computation time and computer-memory consumption in optimization-based reconstruction, thus allowing the exploitation of its potential advantages over existing reconstruction approaches in practical applications.

Current optimization-based reconstructions are based only upon image arrays with uniform resolution, not being applicable to image arrays with variable resolution. Therefore, it is of theoretical and practical significance to develop new optimization-based reconstruction for an image array with variable resolution, thus being capable of efficiently yielding an ROI image with high spatial resolution. In chapters 6 and 7, we develop and investigate optimization-based algorithms for image reconstructions with variable spatial resolution.

CHAPTER 2

OPTIMIZATION-BASED ALGORITHM FOR SHORT-SCAN FLAT-PANEL CBCT

2.1 Introduction

Flat-panel cone-beam computed tomography (CBCT) has widely been used for guidance, monitoring, and assessment tasks in interventional [23–26], surgical [27–30], and therapeutic applications [31–34]. Due to workflow and radiation-dose considerations, it is often desirable to scan a patient with the shortest possible rotation scan that provides sufficient data for reconstruction. It can be shown that data collected over 180°+ the-full-fan-angle, which is typically referred to as the short-scan angular range [86], are sufficient for image reconstruction of the middle plane by use of Feldkamp-Davis-Kress (FDK) algorithm (or its variants) [82, 150].

Optimization-based reconstruction [151–159] remains the focus of current research on CBCT algorithms. In general, optimization programs often are comprised of two important components: data fidelities and image constraints, both of which play important roles in the design of optimization-based reconstruction in CBCT. Among numerous data fidelities, data-ℓ_2, data-ℓ_1, and data-Kullback-Leibler (KL) fidelities have been investigated extensively; and among various image constraints, image-total-variation (TV), image-ℓ_2 norm, and image-ℓ_1 norm are used widely in the design of optimization programs. It is of theoretical as well as practical interest to investigate how different data fidelities and image constraints impact optimization-based reconstructions, especially in short-scan CBCT imaging applications.

Based upon the CBCT-imaging model for analytic algorithms, it can be shown that the data in a short-scan CBCT possesses "redundancy" within the middle plane, i.e., the plane containing the source trajectory. A weighting function such as the Parker weighting function [86] is used widely for normalizing the data-function redundancy so that the FDK algorithm can, when applied to the weighted data-function, yield accurate images within the middle

plane. It is well-known that different weighting functions may lead to FDK reconstructions with different degrees of artifacts especially away from the middle plane [160–162]. Therefore, we also investigate possible impact of different weighting functions, especially those that normalize the data-function redundancy, on optimization-based reconstruction in short-scan CBCT.

In this chapter, we design optimization programs consisting of (1) different data fidelities, (2) different image constraints, and (3) different weighting functions. In an attempt to solve these optimization problems, we employ a primal-dual algorithm developed by Chambolle and Pock (CP) [163], which has been proved mathematically to converge to a solution of a general set of convex optimization problems, to be capable of handling effectively system matrix of large size of practical CT interest [164], and to provide solid convergence conditions to check the solutions. The use of a single algorithm for solving all of the optimization programs may also avoid any reconstruction differences as a result of using different algorithms for solving different optimization programs.

In Sec. 2.2, we introduce the imaging model with weighting matrix, design the optimization programs with different data fidelities and image constraints, and derive the corresponding algorithms. In Sec. 2.3, we carry out numerical studies to verify and demonstrate the capability of the optimization-based algorithm in achieving the designed solution determined by the optimization programs, and the high accuracy of this designed solution compared to the ground truth image. Remarks about the study are given in Sec. 2.4.

2.2 Methods

The imaging model used in the optimization-based reconstruction considered is given by

$$\mathbf{g} = \mathcal{H}\mathbf{f}, \tag{2.1}$$

where vector $\mathbf{g}$ of size M denotes discrete model-data, vector $\mathbf{f}$ of size N the discrete image of N voxels, and $\mathcal{H}$ the system matrix of size $M \times N$. In this work, an entry in $\mathbf{g}$ or $\mathbf{f}$ represents the data or image value within a detector bin or image voxel; and an element in system matrix $\mathcal{H}$ represents an intersection of an X-ray detected with a voxel in the image array. The imaging model in Eq. (2.1), linking a discrete-image vector to a discrete-data vector, is referred to as a discrete-to-discrete (D-D) model. We denote experimentally measured data in practical CBCT imaging as vector $\mathbf{g}_m$ of size M in which an entry represents a particular data sample collected. In contrast, the imaging model upon which analytic algorithms such as the FDK algorithm are developed is referred to a continuous-to-continuous (C-C) model because it maps an object function of continuous variables to a data function of continuous variables.

2.2.1 Design of system matrix

In this part of the study, we will investigate and analyze the system matrix for an image array in a general CBCT imaging model.

For given data and image vectors, the data model is determined completely by system matrix $\mathcal{H}$ [165, 166]. Therefore, different data models can be designed by the ways of calculating $\mathcal{H}$. In the work, we investigate one system matrix whose elements are determined by the intersection segments of a given X-ray with a given voxel. Based upon our previous experience in optimization-based reconstruction, the matrix seems to provide reasonable descriptions of X-ray imaging and can be easily calculated in reconstruction.

Consider a given ray i in Fig. 2.1 that divides the connection between pixels j and $j+1$ into two segments ΔX_{ij} and $W - \Delta X_{ij}$, where W denotes the pixel width. In this case, intersection length L_{ij}, weighted by the ratio of $W - \Delta X_{ij}$ to W, will be used as element H_{ij} of $\mathcal{H}$. Because the size of the system matrix in CBCT imaging is exceedingly large, it is virtually impractical to pre-calculate and store the system matrix. The design of the matrix above allows their efficient calculation, without consumption of excessive memory, in

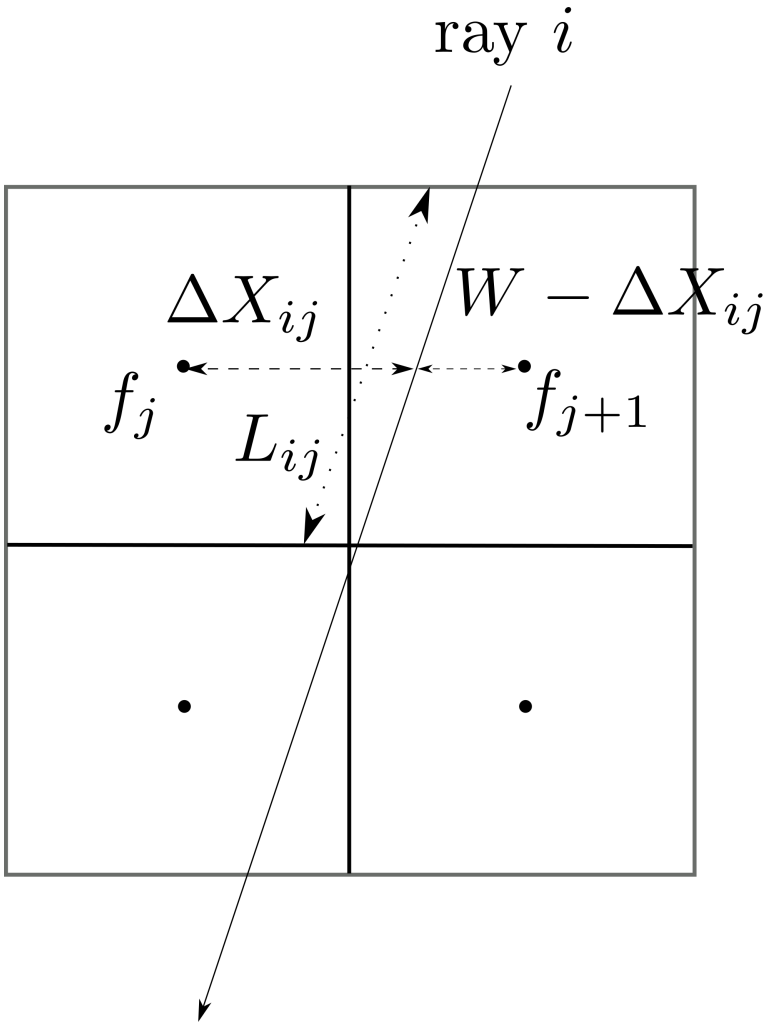


Figure 2.1: Ray i travels through an image with uniform resolution. Element H_{ij} is calculated as the intersection length L_{ij} weighted by the ratio of $W - \Delta X_{ij}$ to W.

the process of image reconstruction.

2.2.2 Design consideration of optimization programs

In optimization-based reconstruction, an image is obtained by solving the optimization program that specifies the image. In this work, we consider optimization programs of form

$$\mathbf{f}^{\star} = \underset{\mathbf{f}}{\mathsf{argmin}} \Phi(\mathcal{W}\mathbf{g}_m, \mathcal{W}\mathcal{H}\mathbf{f}) \quad \text{s.t.} \quad \Psi(\mathbf{f}) \leq \alpha \quad \text{and} \quad f_j \geq 0, \tag{2.2}$$

where $\Phi(\mathcal{W}\mathbf{g}_m, \mathcal{W}\mathcal{H}\mathbf{f})$, referred to data fidelity in the work, denotes a measure of inconsistency between measured data $\mathbf{g}_m$ and model data $\mathbf{g}$, weighted often by a matrix $\mathcal{W}$ of size $M \times M$, $\Psi(\mathbf{f})$ an image constraint function, α a constraint parameter, and f_j the jth entry of the vector $\mathbf{f}$, $j = 1, 2, ..., N$.

Data fidelities

Several data fidelities and image constraints of potential practical interest are considered in the work. Specifically, the data fidelities are given by

(a) data-ℓ_1 fidelity

$$D_{\ell_1}(\mathbf{f}) = ||\mathcal{W}\mathbf{g}_m - \mathcal{W}\mathcal{H}\mathbf{f}||_1, \tag{2.3}$$

where $||\cdot||_1$ denotes ℓ_1-norm;

(b) data-ℓ_2 fidelity

$$D_{\ell_2}(\mathbf{f}) = ||\mathcal{W}\mathbf{g}_m - \mathcal{W}\mathcal{H}\mathbf{f}||_2^2, \tag{2.4}$$

where $||\cdot||_2^2$ denotes the square of ℓ_2-norm; and

(c) data-Kullback-Leibler (KL) fidelity

$$D_{KL}(\mathbf{f}) = \sum_{i=1}^{M}[-\mathcal{W}\mathcal{H}\mathbf{f} + \mathcal{W}\mathbf{g}_m + \mathcal{W}\mathbf{g}_m \ln(\mathcal{W}\mathbf{g}_m) - \mathcal{W}\mathbf{g}_m \ln(\mathcal{W}\mathcal{H}\mathbf{f})]_i, \tag{2.5}$$

where i is the index of the ith entry of data vector $\mathbf{g}_m$, and the entries in vector $(\mathcal{W}\mathcal{H}\mathbf{f})$ that are smaller than 10^{-20} are replaced with 10^{-20}. In contrast to practical FDK reconstruction in which measured data $\mathbf{g}_m$ is pre-conditioned by a diagonal weighting matrix of size $M \times M$, a weighting matrix in an optimization-based reconstruction is multiplied with both measured data $\mathbf{g}_m$ and model data $\mathbf{g}$, thus modulating inconsistency between $\mathbf{g}_m$ and $\mathbf{g}$, instead of $\mathbf{g}_m$ alone.

Ψ \ Φ	$D_{\ell_1}(\mathbf{f})$	$D_{\ell_2}(\mathbf{f})$	$D_{KL}(\mathbf{f})$	α
N/A	D_{ℓ_1}	D_{ℓ_2}	D_{KL}	N/A
$\|\|\mathbf{f}\|\|_{\ell_1}$	D_{ℓ_1}-$\|\|\mathbf{f}\|\|_{\ell_1}$	D_{ℓ_2}-$\|\|\mathbf{f}\|\|_{\ell_1}$	D_{KL}-$\|\|\mathbf{f}\|\|_{\ell_1}$	l_1
$\|\|\mathbf{f}\|\|_{\ell_2}$	D_{ℓ_1}-$\|\|\mathbf{f}\|\|_{\ell_2}$	D_{ℓ_2}-$\|\|\mathbf{f}\|\|_{\ell_2}$	D_{KL}-$\|\|\mathbf{f}\|\|_{\ell_2}$	l_2
$\|\|\mathbf{f}\|\|_{TV}$	D_{ℓ_1}-$\|\|\mathbf{f}\|\|_{TV}$	D_{ℓ_2}-$\|\|\mathbf{f}\|\|_{TV}$	D_{KL}-$\|\|\mathbf{f}\|\|_{TV}$	t_1

Table 2.1: Symbols of optimization programs formed by use of data fidelities in Eqs. (2.3)-(2.5) and image constraints in Eqs. (2.6)-(2.8) to replace Φ and Ψ in Eq. (2.2). For example, using $D_{KL}(\mathbf{f})$ and $||\mathbf{f}||_{TV}$ to replace Φ and Ψ in Eq. (2.2) yields program D_{KL}-$||\mathbf{f}||_{TV}$, as shown in Eq. (2.11) below.

Image constraints

The image constraints considered are given by

(a) image-ℓ_1 constraint

$$||\mathbf{f}||_{\ell_1} = \sum_{j=1}^{N} |f_j|; \tag{2.6}$$

(b) image-ℓ_2 constraint

$$||\mathbf{f}||_{\ell_2} = \sum_{j=1}^{N} f_j^2; \tag{2.7}$$

and

(c) image-TV constraint

$$||\mathbf{f}||_{TV} = ||\,|\nabla \mathbf{f}|_{\text{MAG}}\,||_1, \tag{2.8}$$

where $|\nabla \mathbf{f}|_{\text{MAG}}$ denotes the magnitude image of the spatial gradient of $\mathbf{f}$ defined above Eq. (2.26).

2.2.3 Optimization program with image-TV constraint

Using the data fidelities and image constraints defined above to replace, respectively, Φ and Ψ in Eq. (2.2), we form and study a number of optimization programs in the work, and, for discussion convenience, dubbed with respective symbols summarized in Table 2.1. Image-constraint parameter α corresponding to image constraints $||\mathbf{f}||_{\ell_1}$, $||\mathbf{f}||_{\ell_2}$, or $||\mathbf{f}||_{TV}$ is denoted, respectively, as ℓ_1, ℓ_2, or t_1, as shown in the last column of Table 2.1. For the purpose of clarity and brevity, without loss of generality, we give below explicit forms of optimization programs with the image-TV constraint, shown in row 5 of Table 2.1. In Sec. 3.3.3 we also present results of two additional optimization programs using image-ℓ_1 and image-ℓ_2 constraints, respectively.

Program D_{ℓ_1}-$||\mathbf{f}||_{TV}$: Image-TV-constrained data-ℓ_1 minimization

$$\mathbf{f}^{\star} = \underset{\mathbf{f}}{\mathsf{argmin}} D_{\ell_1}(\mathbf{f}) \quad \text{s.t.} \quad ||\mathbf{f}||_{TV} \leq t_1 \quad \text{and} \quad f_j \geq 0, \tag{2.9}$$

where $t_1 \geq 0$ denotes a pre-selected image-constraint parameter.

Program D_{ℓ_2}-$||\mathbf{f}||_{TV}$: Image-TV-constrained data-ℓ_2 minimization

$$\mathbf{f}^{\star} = \underset{\mathbf{f}}{\mathsf{argmin}} D_{\ell_2}(\mathbf{f}) \quad \text{s.t.} \quad ||\mathbf{f}||_{TV} \leq t_1 \quad \text{and} \quad f_j \geq 0. \tag{2.10}$$

Program D_{KL}-$||\mathbf{f}||_{TV}$: Image-TV-constrained data-KL minimization

$$\mathbf{f}^{\star} = \underset{\mathbf{f}}{\mathsf{argmin}} D_{KL}(\mathbf{f}) \quad \text{s.t.} \quad ||\mathbf{f}||_{TV} \leq t_1 \quad \text{and} \quad f_j \geq 0. \tag{2.11}$$

Program parameters Even if the mathematical form of an optimization program is given, parameters are needed also for completely specifying the program. As mentioned above, system matrix $\mathcal{H}$ needs to be determined, and, in this work, each of its elements is selected as the intersection length of an X-ray with an image voxel. The selection of weighting matrix $\mathcal{W}$, which modulates data inconsistency, is discussed below in Sec.2.2.4. Additionally, the programs in Eqs. (2.9)-(2.11) also depend upon voxel size of the image array and image-TV parameter t_1. When image-ℓ_1 or image-ℓ_2, instead of image-TV, is used as an image constraint, the optimization program depends upon constraint parameter l_1 or l_2 (see row 3 and 4 in Table 2.1). Different selections of parameters specify generally different solutions (i.e., different reconstructions). We discuss in Sec. 3.2.2 below the determination of the program parameter in real-data studies.

2.2.4 Weighting matrix

In short-scan CBCT, let u_m and v_m denote detector sizes in directions tangential to the circular trajectory and perpendicular to the middle plane; and u and v are continuous variables, depicting the Cartesian coordinates of a point on the detector plane, satisfying $|u| \leq u_m$ and $|v| \leq v_m$. At continuous projection angle ϕ, the weighting function normalizing data-function redundancy within the middle plane satisfies

$$w(u, v, \phi) + w(-u, v, \phi + \pi + 2\gamma) = 1, \tag{2.12}$$

where $\gamma = \text{atan}(\frac{u}{L})$, and L the distance from the source to the detector surface. Even though the weighting function is designed for the middle plane, it is used also for normalizing approximate data-function redundancy within the off-middle planes (i.e., $v \neq 0$) in circular CBCT. An additional condition is considered often on the weighting-function smoothness [86] for FDK reconstruction:

$$\begin{cases} \frac{\partial w(u,v,\phi)}{\partial \phi}\Big|_{\phi=2\gamma_m-2\gamma} = \frac{\partial w(u,v,\phi)}{\partial \phi}\Big|_{\phi=\pi-2\gamma} = 0 \\ \\ \frac{\partial w(u,v,\phi)}{\partial u}\Big|_{\phi=2\gamma_m-2\gamma} = \frac{\partial w(u,v,\phi)}{\partial u}\Big|_{\phi=\pi-2\gamma} = 0, \end{cases} \tag{2.13}$$

where $\gamma_m = \text{atan}(\frac{|u_m|}{L})$. It has been shown that the smoothness condition is critical for minimizing sampling artifacts in FDK reconstruction from discrete data. The Parker weighting function [86]

$$\begin{cases} w(u, v, \phi) = \sin^2(\frac{\pi}{4}\frac{\phi}{\gamma_m-\gamma}) & 0 \leq \phi < 2\gamma_m - 2\gamma \\ w(u, v, \phi) = 1 & 2\gamma_m - 2\gamma \leq \phi < \pi - 2\gamma \\ w(u, v, \phi) = \sin^2(\frac{\pi}{4}\frac{\pi+2\gamma_m-\phi}{\gamma_m+\gamma}) & \pi - 2\gamma \leq \phi < \pi + 2\gamma_m \\ w(u, v, \phi) = 0 & \pi + 2\gamma_m \leq \phi < 2\pi. \end{cases} \tag{2.14}$$

satisfies conditions in Eqs. (2.12) and (2.13), and is used widely for image reconstruction in short-scan CBCT.

In analytic reconstruction, a weighting function is used for normalizing the data-function redundancy within the middle plane of short-scan CBCT. In real-data imaging, only discrete data $\mathbf{g}_m$ of size M can be measured, and, a diagonal *weighting matrix* is constructed thus by using M values, in *a concatenated form*, calculated from a weighting function satisfying Eqs. (2.12) and (2.13) at M discrete points in data space (u, v, ϕ) at which $\mathbf{g}_m$ are sampled. The FDK algorithm then reconstructs an image from the measured, discrete data weighted by the diagonal weighting matrix. On the other hand, optimization-based reconstruction can accommodate weighting matrices that are designed based upon physical, statistical, and/or mathematical motivations other than data-redundancy normalization. In an attempt to compare with FDK reconstructions that use the diagonal weighting matrix obtained from the Parker weighting function, the same weighting matrix is used in optimization programs considered in the work, while additional weighting matrices that are designed without consideration of normalizing the data-function redundancy are also studied in Sec. 3.3.3 below.

2.2.5 Algorithm derivation

The optimization programs considered in Table 2.1 are convex, and they can thus be solved by use of the primal-dual algorithms. In the work, we use the primal-dual algorithm developed by Chambolle and Pock (CP) [163, 164] because it appears to be capable of handling effectively system matrix $\mathcal{H}$ of large size of practical CT interest [167], and because it solves not only smooth, but also non-smooth, programs listed in Table 2.1. The use of a single algorithm for solving all of the optimization programs may also avoid any reconstruction differences as a result of using different algorithms for solving different optimization programs. In the following, we will briefly introduce the generic form of the CP algorithm, and adapt it for solving convex optimization programs in Table 2.1.

Generic form of primal and dual objectives The CP algorithm applies to a general form of the primal minimization:

$$\mathbf{x}^\star = \min_{\mathbf{x}} \left\{ F(\mathcal{K}\mathbf{x}) + G(\mathbf{x}) \right\}, \tag{2.15}$$

and a dual maximization:

$$\mathbf{y}^\star = \max_{\mathbf{y}} \left\{ -F^*(\mathbf{y}) - G^*(-\mathcal{K}^\top \mathbf{y}) \right\}, \tag{2.16}$$

where primal variable $\mathbf{x}$ is a vector of size N, and dual variable $\mathbf{y}$ a vector of size M; $\mathcal{K}$ is a linear transform from image space to data space; functions F and G are convex, which map $\mathcal{K}\mathbf{x}$ and $\mathbf{x}$ to non-negative real numbers; the $\top$ superscript denotes matrix transposition; and the superscript '*' in the dual maximization problem refers to convex conjugation, which is given by

$$\begin{aligned} F^*(\mathbf{y}) &= \max_{\mathbf{y}'} \left\{ \mathbf{y}^\top \mathbf{y}' - F(\mathbf{y}') \right\} \\ G^*(\mathbf{y}) &= \max_{\mathbf{y}'} \left\{ \mathbf{y}^\top \mathbf{y}' - G(\mathbf{y}') \right\}. \end{aligned} \tag{2.17}$$

The CP algorithm simultaneously performs the minimization over $\mathbf{x}$ in Eq. (2.15), and the maximization over $\mathbf{y}$ in Eq. (2.16). The difference between the two objectives is referred to as the duality gap, and the convergence is achieved when this gap becomes zero. For the convergence check in practice, we introduce a conditional primal-dual gap (cPD), which is the difference between the primal and the dual objectives ignoring all indicator functions. Theoretically, cPD should approach zero as the algorithm converges.

With the designed primal and dual objectives, we will then derive a general update procedure for the CP algorithm.

CP algorithm for prototyping of convex optimization problems There exist four steps for prototyping a particular convex optimization problem with the CP algorithm: *1*) formulate the optimization problem to the generic minimization problem in Eq. (2.15); *2*) derive the dual maximization problem in the form of Eq. (2.16), by computing the convex conjugate F^* and G^* of function F and G, respectively; *3*) derive the proximal mappings of function F^* and G. The proximal mapping is used to generate a descent direction for a convex function, and its form will be given after Algorithm 1. *4*) obtain the CP algorithm by use of the derived proximal mapping in step *3*).

A general form of CP algorithm is listed in Algorithm 1.

Algorithm 1: Pseudo-code for $\mathcal{N}$ steps of the basic CP algorithm.

$L \leftarrow ||\mathcal{K}||_{SV}$, $\tau \leftarrow 1/L$, $\sigma \leftarrow 1/L$

$\theta \leftarrow 1, n \leftarrow 0$

INITIALIZE: $\mathbf{x}_0$, $\mathbf{y}_0$ to zero

$\bar{\mathbf{x}}_0 \leftarrow \mathbf{x}_0$

repeat

  $\mathbf{y}_{n+1} \leftarrow \mathsf{prox}_{\sigma}[F^*](\mathbf{y}_n + \sigma\mathcal{K}\bar{\mathbf{x}}_n)$

  $\mathbf{x}_{n+1} \leftarrow \mathsf{prox}_{\tau}[G^*](\mathbf{x}_n + \tau\mathcal{K}^T\mathbf{y}_n)$

  $\bar{\mathbf{x}}_{n+1} \leftarrow \mathbf{x}_{n+1} + \theta(\mathbf{x}_{n+1} - \mathbf{x}_n)$

  $n \leftarrow n + 1$

until the practical convergence conditions are satisfied

OUTPUT: image $\mathbf{f}_{\text{conv}} \leftarrow \mathbf{f}_n$

In Algorithm 1, $||\mathcal{K}||_{SV}$ denotes the largest singular value of matrix $\mathcal{K}$, parameters τ, σ, and θ are non-negative CP algorithm parameters, n is the iteration index, and $\mathbf{f}_{\text{conv}}$ the convergent reconstruction when the practical convergence conditions are satisfied. Operators $prox_{\sigma}[F^*](\cdot)$ and $prox_{\tau}[G](\cdot)$ are proximal mappings for functions F^* and G, respectively,

which are defined below:

$$\mathsf{prox}_{\sigma}[F^*](\mathbf{y}) = \underset{\mathbf{y}'}{\mathsf{argmin}}\{F^*(\mathbf{y}') + \frac{||\mathbf{y}-\mathbf{y}'||_2^2}{2\sigma}\}, \tag{2.18}$$

and

$$\mathsf{prox}_{\tau}[G](\mathbf{x}) = \underset{\mathbf{x}'}{\mathsf{argmin}}\{G(\mathbf{x}') + \frac{||\mathbf{x}-\mathbf{x}'||_2^2}{2\tau}\}. \tag{2.19}$$

With the basic CP algorithm listed above, we can formulate all the convex optimization problems in Table 2.1 into the form of Eq. (2.15), derive the corresponding dual objectives and the proximal mapping. After substituting the proximal mapping into Line 5 and 6 in Algorithm 1, a CP algorithm for a specific optimization program can be obtained. Details of the formula derivation can be found in the Appendix of this chapter.

2.2.6 Pseudo-code for CP algorithms

The pseudo-codes for the CP algorithm used are given in Algorithm 2.

In Algorithm 2, vector $\mathbf{g}_m$ of size $M = M_u \times M_v \times M_\phi$ denotes measured data, where M_u and M_v are number of detector bins along the u- and v-axis on the detector plane, and M_ϕ represents the number of views at which data are collected; vector $\mathbf{f}_n$ of size $N = N_x \times N_y \times N_z$ image at the nth iteration, where N_x, N_y, and N_z are the numbers of voxels along the x-, y-, and z-axis, respectively; $\mathbf{f}_{\text{conv}}$ the convergent reconstruction when the practical convergence conditions are satisfied; $\mathcal{W}$ and $\mathcal{H}$ weighting and system matrices; parameter $\nu_{\mathcal{H}} = \frac{||\mathcal{W}\mathcal{H}||_{SV}}{||\mathcal{X}||_{SV}}$, where $||\cdot||_{SV}$ denotes the largest singular value of a matrix [164]; matrix $\mathcal{H}_{\mathcal{X}}$ has a transpose $\mathcal{H}_{\mathcal{X}}^{\top} = (\mathcal{H}^{\top}\mathcal{W}^{\top}, \nu\mathcal{X}^{\top})$; $\mathsf{pos}(\mathbf{f})$ enforces $f_j = 0$, if $f_j < 0$, where f_j denotes the jth entry of vector $\mathbf{f}$ of size N; vectors $\mathbf{p}_n$ and $\mathbf{\Theta}$ are of size M, whereas vectors $\mathbf{q}_n$ and $\mathbf{\Omega}$ have the same size, which depends upon the programs considered. The specifics of vectors $\mathbf{\Theta}$, $\mathbf{\Omega}$, along with matrix $\mathcal{X}$, are discussed below for optimization programs under study. Parameter ζ denotes νl_1, νl_2, and νt_1, respectively, for programs with image-ℓ_1, image-ℓ_2, and image-TV constraints. For each study, the values of l_1, l_2, and t_1 were determined in Sec. 3.2.2.

Algorithm: Pseudo-codes of the CP algorithm for solving programs in Table 2.1

INPUT: $\mathbf{g}_m$; $\mathcal{H}$, $\mathcal{W}$, and $\mathcal{X}$; ζ and λ

1: $\nu \leftarrow \nu_{\mathcal{H}}$, $L \leftarrow ||\mathcal{H}_{\mathcal{X}}||_{SV}$, $\tau \leftarrow 1/L$, $\sigma \leftarrow 1/L$

2: $\theta \leftarrow 1$, $n \leftarrow 0$

3: INITIALIZE: $\mathbf{f}_0$, $\mathbf{p}_0$, and $\mathbf{q}_0$ to zero

4: $\bar{\mathbf{f}}_0 \leftarrow \mathbf{f}_0$

5: **repeat**

6: $\quad \mathbf{p}_{n+1} \leftarrow \mathbf{\Theta}(\lambda, \mathbf{p}_n, \mathcal{W}, \mathcal{H}\bar{\mathbf{f}}_n, \mathbf{g}_m)$

7: $\quad \mathbf{q}_{n+1} \leftarrow \mathbf{\Omega}(\mathbf{q}_n, \nu, \sigma, \mathcal{X}\bar{\mathbf{f}}_n, \zeta)$

8: $\quad \mathbf{f}_{n+1} \leftarrow \mathsf{pos}[\mathbf{f}_n - \tau(\mathcal{H}^T\mathcal{W}^T\mathbf{p}_{n+1} + \nu\mathcal{X}^\top\mathbf{q}_{n+1})]$

9: $\quad \bar{\mathbf{f}}_{n+1} \leftarrow \mathbf{f}_{n+1} + \theta(\mathbf{f}_{n+1} - \mathbf{f}_n)$

10: $\quad n \leftarrow n + 1$

11: **until** the practical convergence conditions are satisfied

12: OUTPUT: image $\mathbf{f}_{\text{conv}} \leftarrow \mathbf{f}_n$

Algorithm parameter λ has no impact mathematically on the feasible solution set specified by an optimization program, but it can affect the convergence rate of the algorithm. We select λ between 0.01 to 1.0 in the study, because previous empirical studies have shown that this selection generally yields a reasonable convergence rate.

Updates in Line 6 Update vector $\mathbf{\Theta}(\lambda, \mathbf{p}_n, \mathcal{W}, \mathcal{H}\bar{\mathbf{f}}_n, \mathbf{g}_m)$ of size M in Line 6 depends upon the specific form of data fidelity considered. For programs in column 2 of Table 2.1,

$$\mathbf{\Theta}(\lambda, \mathbf{p}_n, \mathcal{W}, \mathcal{H}\bar{\mathbf{f}}_n, \mathbf{g}_m) = [\mathbf{p}_n + \sigma\mathcal{W}(\mathcal{H}\bar{\mathbf{f}}_n - \mathbf{g}_m)]/(1 + \sigma/\lambda); \tag{2.20}$$

for optimization programs in column 3 of Table 2.1,

$$\mathbf{\Theta}(\lambda, \mathbf{p}_n, \mathcal{W}, \mathcal{H}\bar{\mathbf{f}}_n, \mathbf{g}_m) = \frac{\mathbf{p}_n + \sigma\mathcal{W}(\mathcal{H}\bar{\mathbf{f}}_n - \mathbf{g}_m)}{\max(\mathbf{1}_D,\, |\mathbf{p}_n + \sigma\mathcal{W}(\mathcal{H}\bar{\mathbf{f}}_n - \mathbf{g}_m)|/\lambda)}, \tag{2.21}$$

where $\mathbf{1}_D$ denotes a vector of size M with all entries set to 1, $|\mathbf{p}|$ takes entry-wise absolute values of vector $\mathbf{p}$ of size M, and $\max(\cdot)$ is performed component-wise; and for programs in column 4 in Table 2.1,

$$\begin{aligned}\boldsymbol{\Theta}(\lambda, \mathbf{p}_n, \mathcal{W}, \mathcal{H}\bar{\mathbf{f}}_n, \mathbf{g}_m) &= \frac{1}{2}[\boldsymbol{\lambda}_D + \mathbf{p}_n + \sigma\mathcal{W}\mathcal{H}\bar{\mathbf{f}}_n] \\ &- \sqrt{[\boldsymbol{\lambda}_D - \mathbf{p}_n - \sigma\mathcal{W}\mathcal{H}\bar{\mathbf{f}}_n]^2 + 4\sigma\lambda\mathcal{W}\mathbf{g}_m}, \end{aligned} \tag{2.22}$$

where $\boldsymbol{\lambda}_D$ is a vector of size M with entries set to λ, and $\sqrt{\mathbf{p}}$ takes entry-wise square roots of vector $\mathbf{p}$ of size M.

Updates in Line 7 For programs in row 2, $\mathcal{X}$ is an identity matrix of size N, and vector $\boldsymbol{\Omega}$ of size N is given by

$$\boldsymbol{\Omega}(\mathbf{q}_n, \nu, \sigma, \mathcal{X}\bar{\mathbf{f}}_n, \zeta) = \mathbf{0}, \tag{2.23}$$

where $\mathbf{0}$ is a vector of size N with entries set to zero.

For program in rows 3 and 4 of Table 2.1, $\mathcal{X}$ is an identity matrix of size N, image-constraint parameter $\zeta = \nu l_1$ for programs in row 3 and $\zeta = \nu l_2$ for programs in row 4, and vector $\boldsymbol{\Omega}$ of size N is given by

$$\boldsymbol{\Omega}(\mathbf{q}_n, \nu, \sigma, \mathcal{X}\bar{\mathbf{f}}_n, \zeta) = \boldsymbol{\varrho}(\mathbf{1}_I - \sigma\mathbf{s}/|\boldsymbol{\varrho}|_{\text{MAG}}), \tag{2.24}$$

where vector $\boldsymbol{\varrho}$ of size N is defined as

$$\boldsymbol{\varrho} = \mathbf{q}_n + \sigma\nu\bar{\mathbf{f}}_n,$$

vector $\mathbf{1}_I$ of size N with entries set to 1, and vector $\mathbf{s}$ of size N is given by

$$\mathbf{s} = \mathbf{ProjOnto}\ell_p\mathbf{Ball}_c(|\boldsymbol{\varrho}|_{\text{MAG}}/\sigma), \tag{2.25}$$

operator $\mathbf{ProjOnto}\ell_p\mathbf{Ball}_c$ projects vector $|\boldsymbol{\varrho}|_{\text{MAG}}/\sigma$ onto the ℓ_p-ball of scale c, $|\boldsymbol{\varrho}|_{\text{MAG}}$ depicts a vector of size N with entry j given by $(|\boldsymbol{\varrho}|_{\text{MAG}})_j = |\varrho_j|$, and ϱ_j indicates the jth entry of $\boldsymbol{\varrho}$. Furthermore, $p = 1$ and $c = \nu l_1$ for programs in row 3 of Table 2.1, and $p = 2$ and $c = \nu l_2$ for programs in row 4 of Table 2.1.

For program in row 5 of Table 2.1, matrix $\mathcal{X} = \nabla$ denotes a spatial gradient matrix of size $3N \times N$, with its transpose $\nabla^\top = (\nabla_x^\top, \nabla_y^\top, \nabla_z^\top)$, where matrices ∇_x, ∇_y, and ∇_z of size $N \times N$ represent the finite difference matrices along x-, y-, and z-axis, respectively, yielding vectors $\nabla_x\mathbf{f}$, $\nabla_y\mathbf{f}$, and $\nabla_z\mathbf{f}$ of size N, which are used to form vector $\nabla\mathbf{f}$ of size $3N$ in a concatenated form in the order of x, y, and z, $|\nabla\mathbf{f}|_{\text{MAG}}$ depicts a vector of size N with entry j given by $(|\nabla\mathbf{f}|_{\text{MAG}})_j = \sqrt{(\nabla\mathbf{f})_j^2 + (\nabla\mathbf{f})_{j+N}^2 + (\nabla\mathbf{f})_{j+2N}^2}$, and $(\nabla\mathbf{f})_j$ indicates the jth entry of vector $\nabla\mathbf{f}$. Moreover, image-constraint parameter $\zeta = \nu t_1$, and vector $\boldsymbol{\Omega}$ of size $3N$ is given by

$$\boldsymbol{\Omega}(\mathbf{q}_n, \nu, \sigma, \mathcal{X}\bar{\mathbf{f}}_n, \zeta) = \boldsymbol{\varrho}(\mathbf{1}_I - \sigma\mathbf{s}/|\boldsymbol{\varrho}|_{\text{MAG}}), \tag{2.26}$$

where vector $\boldsymbol{\varrho}$ of size $3N$ is defined as

$$\boldsymbol{\varrho} = \mathbf{q}_n + \sigma\nu\nabla\bar{\mathbf{f}}_n,$$

vector $\mathbf{1}_I$ is of size N with entries set to 1, and vector $\mathbf{s}$ of size N is given by

$$\mathbf{s} = \mathbf{ProjOnto}\ell_1\mathbf{Ball}_{\nu t_1}(|\boldsymbol{\varrho}|_{\text{MAG}}/\sigma), \tag{2.27}$$

operator $\mathbf{ProjOnto}\ell_1\mathbf{Ball}_{\nu t_1}$ projects vector $|\boldsymbol{\varrho}|_{\text{MAG}}/\sigma$ onto the ℓ_1-ball of scale νt_1 [167], $|\boldsymbol{\varrho}|_{\text{MAG}}$ depicts a vector of size N with entry j given by $(|\boldsymbol{\varrho}|_{\text{MAG}})_j = \sqrt{\varrho_j^2 + \varrho_{j+N}^2 + \varrho_{j+2N}^2}$, and ϱ_j indicates the jth entry of vector $\boldsymbol{\varrho}$.

cPD gap Conditional primal-dual, which is the difference between primal function and dual function, is an important metrics for checking algorithm convergence. Theoretically, cPD should approach 0 as iteration number n increases. We list below the forms of cPD for

all programs listed in Table 2.1. Details of the derivation and symbol explanations can be found in the Appendix of Chapter 2.

Program D_{ℓ_1}

$$\begin{aligned} \text{cPD}(\mathbf{f}_n) &= \lambda||\mathcal{W}\mathcal{H}\mathbf{f}_n - \mathcal{W}\mathbf{g}_m||_1 && (2.28)\\ &- \Big[- \mathbf{g}_m^\top \mathcal{W}^\top \mathbf{p}_n \Big]. && (2.29) \end{aligned}$$

Program D_{ℓ_2}

$$\begin{aligned} \text{cPD}(\mathbf{f}_n) &= \frac{1}{2}\lambda||\mathcal{W}\mathcal{H}\mathbf{f}_n - \mathcal{W}\mathbf{g}_m||_2^2 && (2.30)\\ &- \Big[- \frac{1}{2\lambda}||\mathbf{p}_n||_2^2 - \mathbf{g}_m^\top \mathcal{W}^\top \mathbf{p}_n \Big]. && (2.31) \end{aligned}$$

Program D_{KL}

$$\begin{aligned} \text{cPD}(\mathbf{f}_n) &= \lambda \sum_i [\mathcal{W}\mathcal{H}\mathbf{f}_n - \mathcal{W}\mathbf{g}_m + \mathcal{W}\mathbf{g}_m \ln(\mathcal{W}\mathbf{g}_m) - \mathcal{W}\mathbf{g}_m \ln(\mathcal{W}\mathcal{H}\mathbf{f}_n)]_i && (2.32)\\ &- \Big[- \sum_i [-\lambda \mathbf{g}_m^\top \mathcal{W}^\top \ln(\mathbf{1}_D - \frac{\mathbf{p}_n}{\lambda})]_i \Big]. && (2.33) \end{aligned}$$

Program D_{ℓ_1}-$\mathbf{f}_{\ell_1}$

$$\begin{aligned} \text{cPD}(\mathbf{f}_n) &= \lambda||\mathcal{W}\mathcal{H}\mathbf{f}_n - \mathcal{W}\mathbf{g}_m||_1 && (2.34)\\ &- \Big[- \mathbf{g}_m^\top \mathcal{W}^\top \mathbf{p}_n - \nu l_1 ||\mathbf{q}_n||_\infty \Big]. && (2.35) \end{aligned}$$

Program D_{ℓ_2}-$\mathbf{f}_{\ell_1}$

$$\begin{aligned} \text{cPD}(\mathbf{f}_n) &= \frac{1}{2}\lambda||\mathcal{W}\mathcal{H}\mathbf{f}_n - \mathcal{W}\mathbf{g}_m||_2^2 && (2.36)\\ &- \Big[- \frac{1}{2\lambda}||\mathbf{p}_n||_2^2 - \mathbf{g}_m^\top \mathcal{W}^\top \mathbf{p}_n - \nu l_1 ||\mathbf{q}_n||_\infty \Big]. && (2.37) \end{aligned}$$

Program D_{KL}-$\mathbf{f}_{\ell_1}$

$$\begin{aligned} \text{cPD}(\mathbf{f}_n) &= \lambda \sum_i [\mathcal{W}\mathcal{H}\mathbf{f}_n - \mathcal{W}\mathbf{g}_m + \mathcal{W}\mathbf{g}_m \ln(\mathcal{W}\mathbf{g}_m) - \mathcal{W}\mathbf{g}_m \ln(\mathcal{W}\mathcal{H}\mathbf{f}_n)]_i & (2.38) \\ &- \left[- \sum_i [-\lambda \mathbf{g}_m^\top \mathcal{W}^\top \ln(\mathbf{1}_D - \frac{\mathbf{p}_n}{\lambda})]_i - \nu l_1 ||\mathbf{q}_n||_\infty \right]. & (2.39) \end{aligned}$$

Program D_{ℓ_1}-$\mathbf{f}_{\ell_2}$

$$\begin{aligned} \text{cPD}(\mathbf{f}_n) &= \lambda ||\mathcal{W}\mathcal{H}\mathbf{f}_n - \mathcal{W}\mathbf{g}_m||_1 & (2.40) \\ &- \left[- \mathbf{g}_m^\top \mathcal{W}^\top \mathbf{p}_n - \nu l_2 ||\mathbf{q}_n||_2 \right]. & (2.41) \end{aligned}$$

Program D_{ℓ_2}-$\mathbf{f}_{\ell_2}$

$$\begin{aligned} \text{cPD}(\mathbf{f}_n) &= \frac{1}{2}\lambda ||\mathcal{W}\mathcal{H}\mathbf{f}_n - \mathcal{W}\mathbf{g}_m||_2^2 & (2.42) \\ &- \left[- \frac{1}{2\lambda} ||\mathbf{p}_n||_2^2 - \mathbf{g}_m^\top \mathcal{W}^\top \mathbf{p}_n - \nu l_2 ||\mathbf{q}_n||_2 \right]. & (2.43) \end{aligned}$$

Program D_{KL}-$\mathbf{f}_{\ell_2}$

$$\begin{aligned} \text{cPD}(\mathbf{f}_n) &= \lambda \sum_i [\mathcal{W}\mathcal{H}\mathbf{f}_n - \mathcal{W}\mathbf{g}_m + \mathcal{W}\mathbf{g}_m \ln(\mathcal{W}\mathbf{g}_m) - \mathcal{W}\mathbf{g}_m \ln(\mathcal{W}\mathcal{H}\mathbf{f}_n)]_i & (2.44) \\ &- \left[- \sum_i [-\lambda \mathbf{g}_m^\top \mathcal{W}^\top \ln(\mathbf{1}_D - \frac{\mathbf{p}_n}{\lambda})]_i - \nu l_2 ||\mathbf{q}_n||_2 \right]. & (2.45) \end{aligned}$$

Program D_{ℓ_1}-$\mathbf{f}_{TV}$

$$\begin{aligned} \text{cPD}(\mathbf{f}_n) &= \lambda ||\mathcal{W}\mathcal{H}\mathbf{f}_n - \mathcal{W}\mathbf{g}_m||_1 & (2.46) \\ &- \left[- \mathbf{g}_m^\top \mathcal{W}^\top \mathbf{p}_n - \nu t_1 ||(|\mathbf{q}_n|_{\text{MAG}})||_\infty \right]. & (2.47) \end{aligned}$$

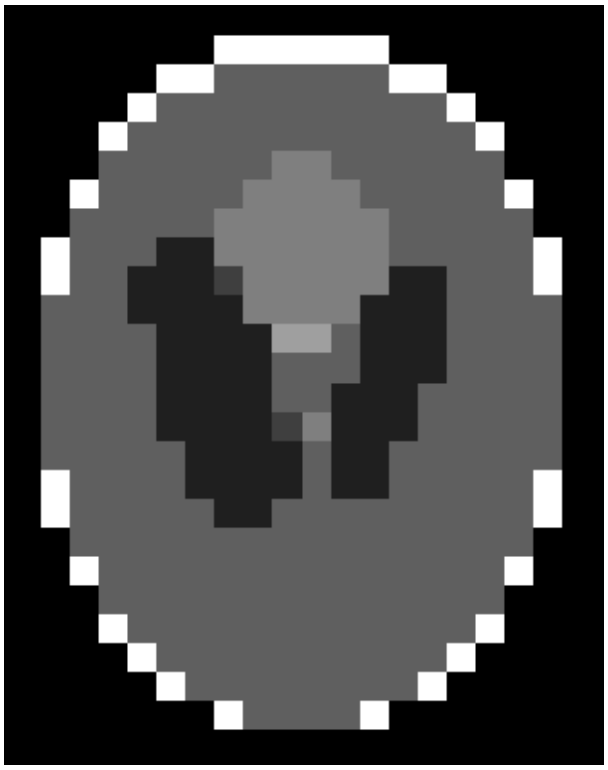


Figure 2.2: Numerical Shepp-Logan phantom used in the inverse crime study. Display window: [0.99, 1.07] cm^{-1}.

Program D_{ℓ_2}-$\mathbf{f}_{TV}$

$$\begin{aligned} \mathrm{cPD}(\mathbf{f}_n) &= \frac{1}{2}\lambda||\mathcal{W}\mathcal{H}\mathbf{f}_n - \mathcal{W}\mathbf{g}_m||_2^2 && (2.48) \\ &- \left[-\frac{1}{2\lambda}||\mathbf{p}_n||_2^2 + \mathbf{g}_m^\top \mathcal{W}^\top \mathbf{p}_n - \nu t_1 ||(|\mathbf{q}_n|_{\mathrm{MAG}})||_\infty \right]. && (2.49) \end{aligned}$$

Program D_{KL}-$\mathbf{f}_{TV}$

$$\begin{aligned} \mathrm{cPD}(\mathbf{f}_n) &= \lambda \sum_i [\mathcal{W}\mathcal{H}\mathbf{f}_n - \mathcal{W}\mathbf{g}_m + \mathcal{W}\mathbf{g}_m \ln(\mathcal{W}\mathbf{g}_m) - \mathcal{W}\mathbf{g}_m \ln(\mathcal{W}\mathcal{H}\mathbf{f}_n)]_i && (2.50) \\ &- \left[-\sum_i [-\lambda \mathbf{g}_m^\top \mathcal{W}^\top \ln(\mathbf{1}_D - \frac{\mathbf{p}_n}{\lambda})]_i - \nu t_1 ||(|\mathbf{q}_n|_{\mathrm{MAG}})||_\infty \right]. && (2.51) \end{aligned}$$

2.3 Algorithm validation

In an attempt to verify the implementation of the CP algorithm, and its appropriate applications to solving the optimization programs, we have performed, for each of the optimization programs considered, an inverse-crime study in which model data were generated by use of a system matrix from a known truth image, and the same system matrix is used also in reconstruction. Therefore, in each inverse-crime study, program parameters such as image-voxel size and image-constraint parameters (i.e., t_1, l_1, or l_2) are known [72,168] with which the CP algorithm was validated quantitatively to converge to the truth image numerically through solving the optimization program.

2.3.1 Study design

For the purpose of validation and computation-efficiency consideration, we use the fan-beam CT configuration with circular geometry, the distance from the source to rotation axis is 100 cm, and the distance from the source to detector is 150 cm. The detector consists of 80 detector elements of size 0.45 mm. We generate 168 projections uniformly distributed over 193 degrees (short-scan angle). We discretize the Shepp-Loggan phantom on an image array consisting of 32×32 identical square pixels of size 0.75 mm, shown in Fig. 2.2, which is referred to as the truth image $\mathbf{f}_{\text{true}}$. With the configuration and image array given, system matrix $\mathcal{H}$ can be calculated as described in Sec. 2.2.1 above. The model data $\mathbf{g}_m$ is generated as $\mathbf{g}_m = \mathcal{W}\mathcal{H}\mathbf{f}_{\text{true}}$, where $\mathcal{W}$ denotes the diagonal weighting matrix obtained from the Parker weighting function.

For designing the practical convergence conditions, we introduce normalized data divergence, such as normalized data-ℓ_1 divergence:

$$\overline{D}_{\ell_1}(\mathbf{g}_m, \mathbf{g}) = D_{\ell_1}(\mathbf{g}_m, \mathbf{g}) / D_{\ell_1}(\mathbf{g}_m, \mathbf{0}), \tag{2.52}$$

where $\mathbf{f}_n$ denote reconstructions at iteration n, and vector $\mathbf{0}$ is of size N with all entries set

to 0; normalized data-ℓ_2 divergence:

$$\overline{D}_{\ell_2}(\mathbf{g}_m, \mathbf{g}) = D_{\ell_2}(\mathbf{g}_m, \mathbf{g})/D_{\ell_2}(\mathbf{g}_m, \mathbf{0}); \tag{2.53}$$

and normalized data-KL divergence:

$$\overline{D}_{KL}(\mathbf{g}_m, \mathbf{g}) = D_{KL}(\mathbf{g}_m, \mathbf{g})/D_{KL}(\mathbf{g}_m, \mathbf{g}_\varsigma); \tag{2.54}$$

When computing $D_{KL}(\mathbf{g}_m, \mathbf{g})$ in the work, the entries in $\mathbf{g}$ that are smaller than 10^{-20} are replaced with 10^{-20}, and $\mathbf{g}_\varsigma$ is obtained by replacing all of the entries in $\mathbf{g}$ with 10^{-20}. Here, we define $\overline{D}(\mathbf{f}_n) = \overline{D}_{\ell_2}(\mathbf{f}_n)$, $\overline{D}_{\ell_1}(\mathbf{f}_n)$, and $\overline{D}_{KL}(\mathbf{f}_n)$ for programs containing data-ℓ_1, data-ℓ_2, and data-KL fidelities, respectively.

When image constraint exists, we introduced unitless metrics for normalizing image constraints, respectively, which are defined as:

$$\overline{TV}(\mathbf{f}_n) = |\, \|\mathbf{f}_n\|_{TV} - t_0|/t_0, \tag{2.55}$$

$$\overline{L1}(\mathbf{f}_n) = |\, \|\mathbf{f}_n\|_1 - l_1|/l_1, \tag{2.56}$$

and

$$\overline{L2}(\mathbf{f}_n) = |\, \|\mathbf{f}_n\|_2 - l_2|/l_2. \tag{2.57}$$

Here we define $\overline{C}(\mathbf{f}_n) = \overline{TV}(\mathbf{f}_n)$, $\overline{L1}(\mathbf{f}_n)$, and $\overline{L2}(\mathbf{f}_n)$ for programs containing image-TV, image-ℓ_1, and image-ℓ_2 constraints, respectively.

In addition, we define a normalized cPD gap, which is given by:

$$\overline{\mathrm{cPD}}(\mathbf{f}_n) = |\mathrm{cPD}(\mathbf{f}_n)/\mathrm{cPD}(\mathbf{f}_1)|. \tag{2.58}$$

The mathematical convergence conditions for the CP algorithms include $\overline{C}(\mathbf{f}_n) \to 0$, $\overline{D}(\mathbf{u}_n) \to 0$, and $\overline{\mathrm{cPD}}(\mathbf{f}_n) \to 0$, as $n \to \infty$. They are unachievable, however, due to

limited computer precision and computation time involved in any practical, numerical study. Therefore, for the inverse-crime study considered, we design practical convergence conditions

$$\begin{aligned} \overline{D}(\mathbf{f}_n) &< 10^{-5}, \\ \overline{\mathrm{cPD}}(\mathbf{f}_n) &< 10^{-5}, \\ \overline{\Delta}(\mathbf{f}_n) &< 10^{-5}, \end{aligned} \tag{2.59}$$

where $\overline{\Delta}(\mathbf{f}_n) = \frac{||\mathbf{f}_n - \mathbf{f}_{\text{true}}||_2}{||\mathbf{f}_{\text{true}}||_2}$. When there exists image constraint, we also consider the convergence condition

$$\overline{C}(\mathbf{f}_n) < 10^{-5} \tag{2.60}$$

It requires that the convergence metrics maintain their decaying trends even after that the conditions are satisfied, as n increases.

2.3.2 Study results

We have performed inverse-crime studies on reconstructions based upon the 12 optimization programs in Table 2.1. It can be observed in Fig. (2.3) that all convergent reconstructions are visually identical to the truth image $\mathbf{f}_{\text{true}}$. Also, Figs. 2.4–2.15 show how convergence metrics, for each optimization program, evolve as functions of the iteration number, demonstrating that the practical convergence conditions in Eq. (2.59) are satisfied, which indicates that the reconstruction converges to the truth image.

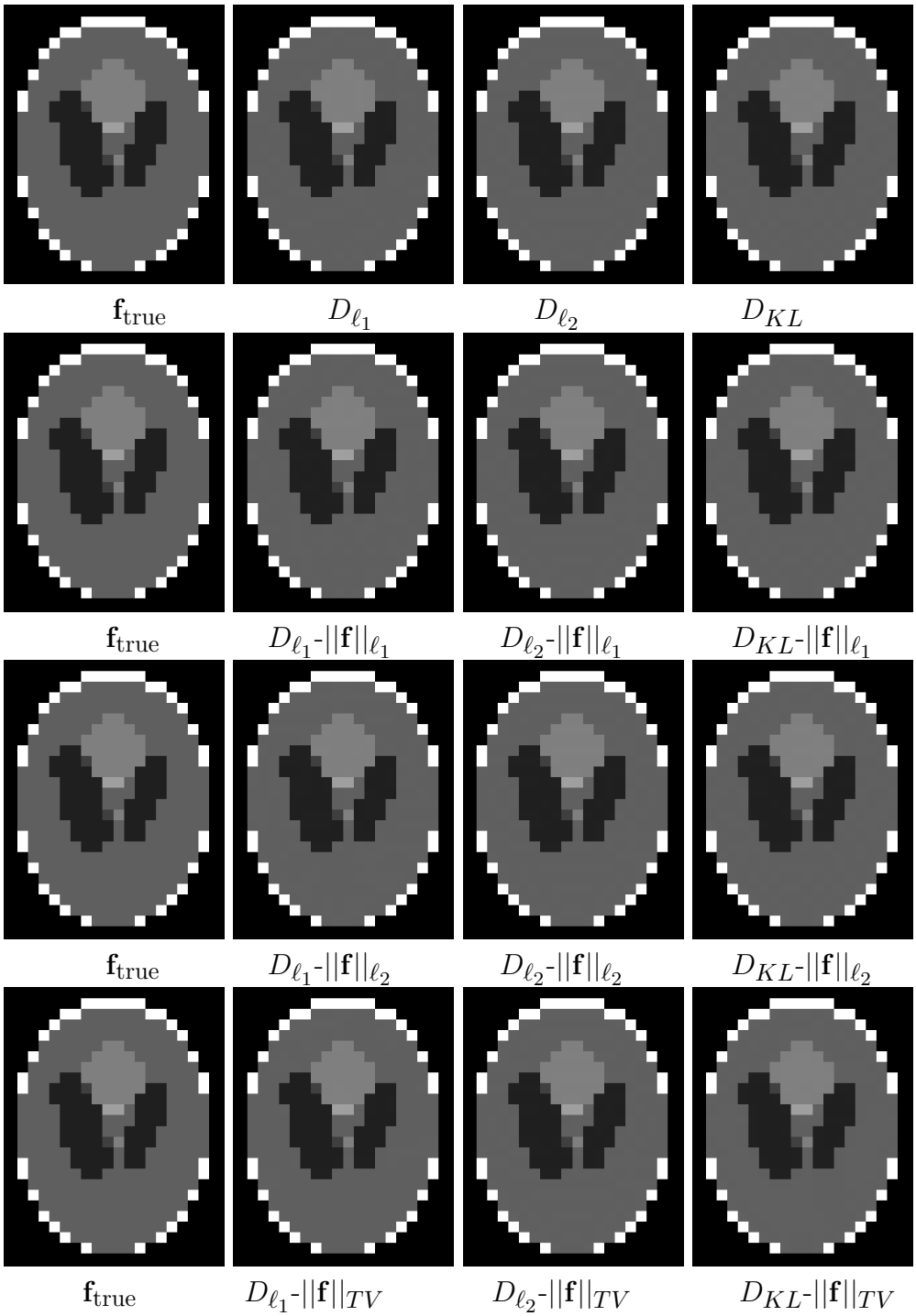


Figure 2.3: Reconstructions of a numerical Shepp-Logan phantom obtained by use of CP algorithm solving, respectively, optimization programs in Table 2.1. Display window: [0.99, 1.07] cm^{-1}.

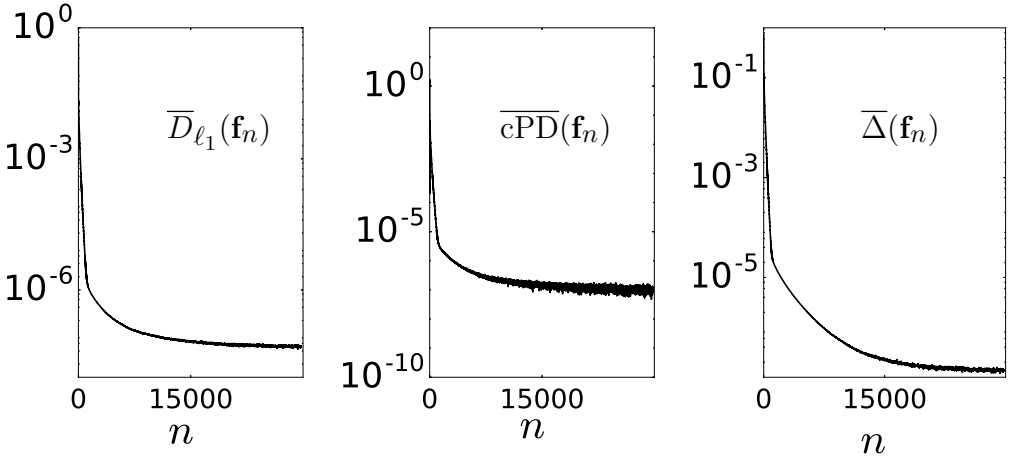


Figure 2.4: Convergence metrics $\overline{D}_{\ell_1}(\mathbf{f}_n)$, $\overline{\mathrm{cPD}}(\mathbf{f}_n)$, and $\overline{\Delta}(\mathbf{f}_n)$ as functions of iteration number n for solving program D_{ℓ_1}.

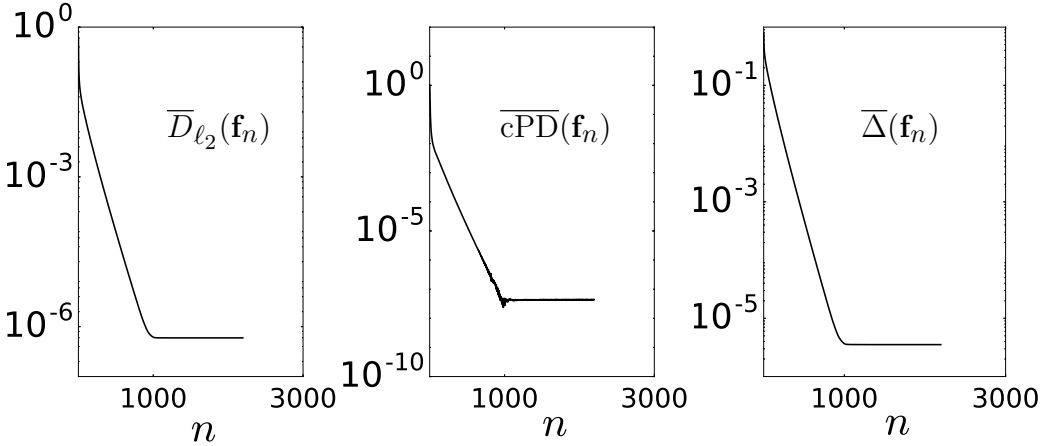


Figure 2.5: Convergence metrics $\overline{D}_{\ell_2}(\mathbf{f}_n)$, $\overline{\mathrm{cPD}}(\mathbf{f}_n)$, and $\overline{\Delta}(\mathbf{f}_n)$ as functions of iteration number n for solving program D_{ℓ_2}.

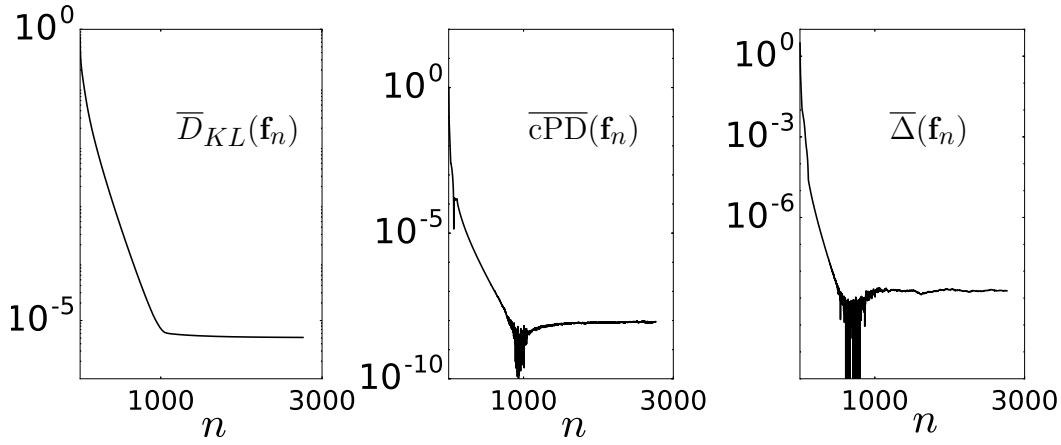


Figure 2.6: Convergence metrics $\overline{D}_{KL}(\mathbf{f}_n)$, $\overline{\mathrm{cPD}}(\mathbf{f}_n)$, and $\overline{\Delta}(\mathbf{f}_n)$ as functions of iteration number n for solving program D_{KL}.

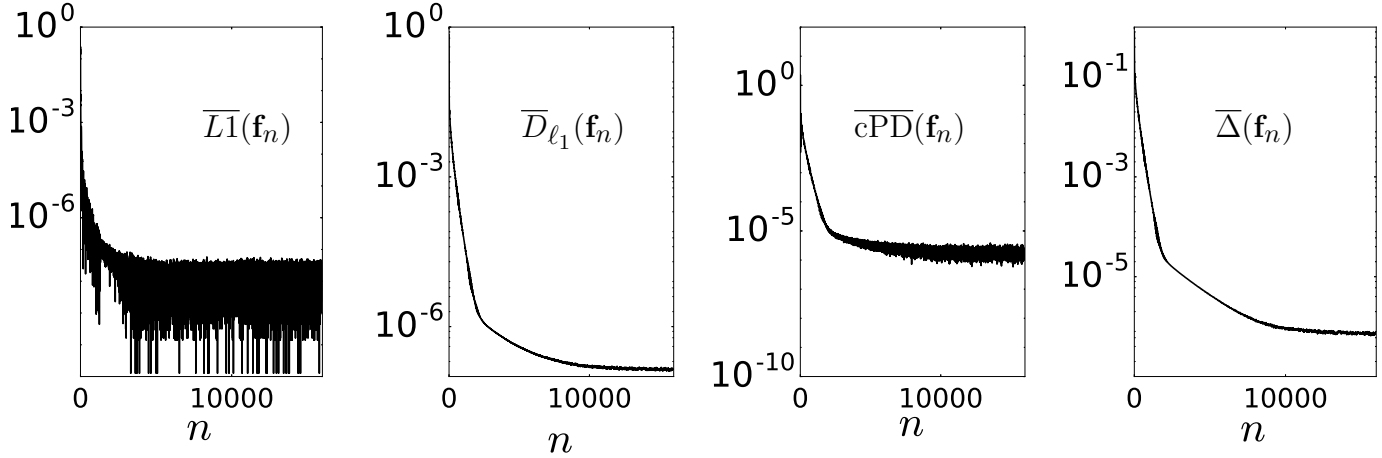


Figure 2.7: Convergence metrics $\overline{L1}(\mathbf{f}_n)$, $\overline{D}_{\ell_1}(\mathbf{f}_n)$, $\overline{\mathrm{cPD}}(\mathbf{f}_n)$, and $\overline{\Delta}(\mathbf{f}_n)$ as functions of iteration number n for solving program D_{ℓ_1}-$||\mathbf{f}||_{\ell_1}$.

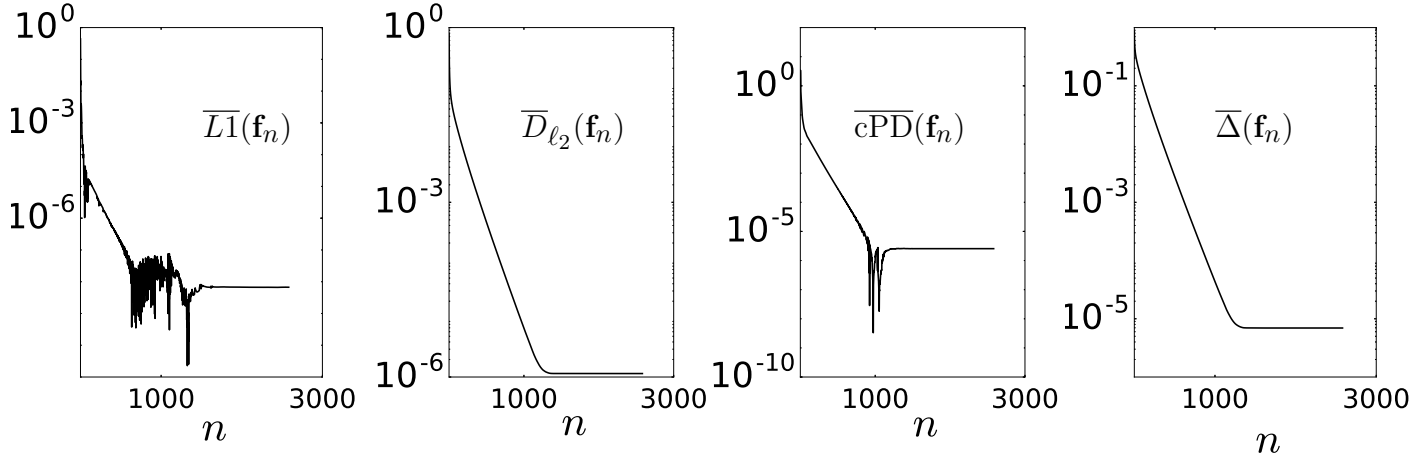


Figure 2.8: Convergence metrics $\overline{L1}(\mathbf{f}_n)$, $\overline{D}_{\ell_2}(\mathbf{f}_n)$, $\overline{\mathrm{cPD}}(\mathbf{f}_n)$, and $\overline{\Delta}(\mathbf{f}_n)$ as functions of iteration number n for solving program D_{ℓ_2}-$||\mathbf{f}||_{\ell_1}$.

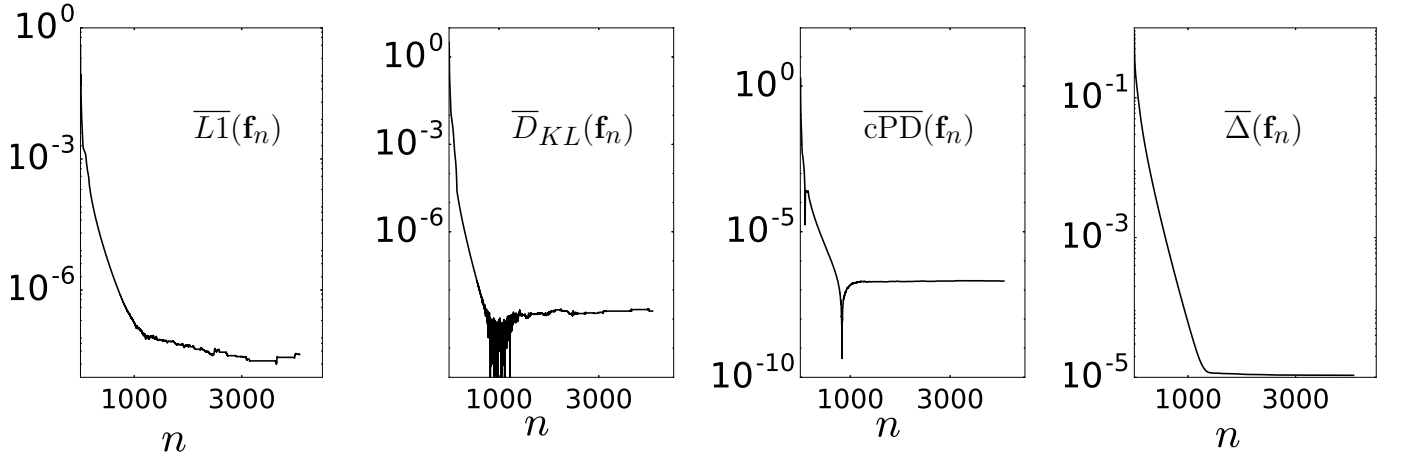


Figure 2.9: Convergence metrics $\overline{L1}(\mathbf{f}_n)$, $\overline{D}_{KL}(\mathbf{f}_n)$, $\overline{\mathrm{cPD}}(\mathbf{f}_n)$, and $\overline{\Delta}(\mathbf{f}_n)$ as functions of iteration number n for solving program D_{KL}-$||\mathbf{f}||_{\ell_1}$.

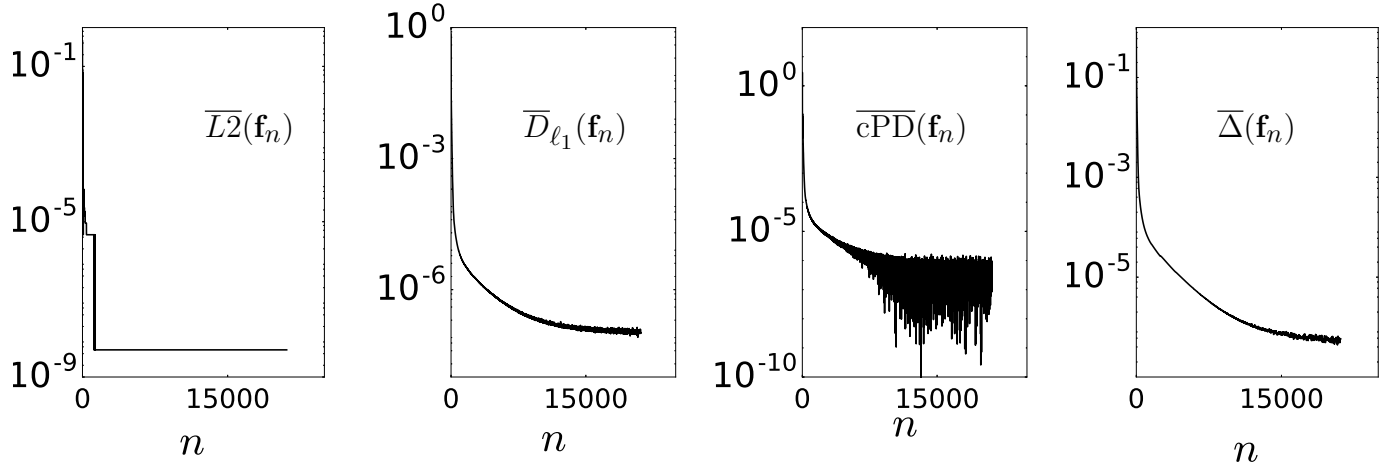


Figure 2.10: Convergence metrics $\overline{L2}(\mathbf{f}_n)$, $\overline{D}_{\ell_1}(\mathbf{f}_n)$, $\overline{\mathrm{cPD}}(\mathbf{f}_n)$, and $\overline{\Delta}(\mathbf{f}_n)$ as functions of iteration number n for solving program D_{ℓ_1}-$||\mathbf{f}||_{\ell_2}$.

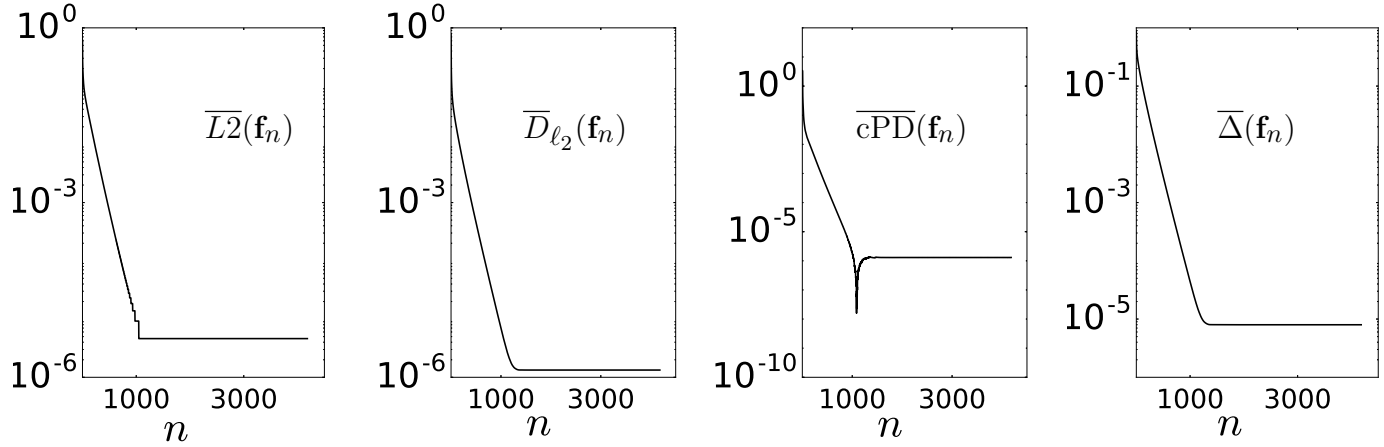


Figure 2.11: Convergence metrics $\overline{L2}(\mathbf{f}_n)$, $\overline{D}_{\ell_2}(\mathbf{f}_n)$, $\overline{\mathrm{cPD}}(\mathbf{f}_n)$, and $\overline{\Delta}(\mathbf{f}_n)$ as functions of iteration number n for solving program D_{ℓ_2}-$||\mathbf{f}||_{\ell_2}$.

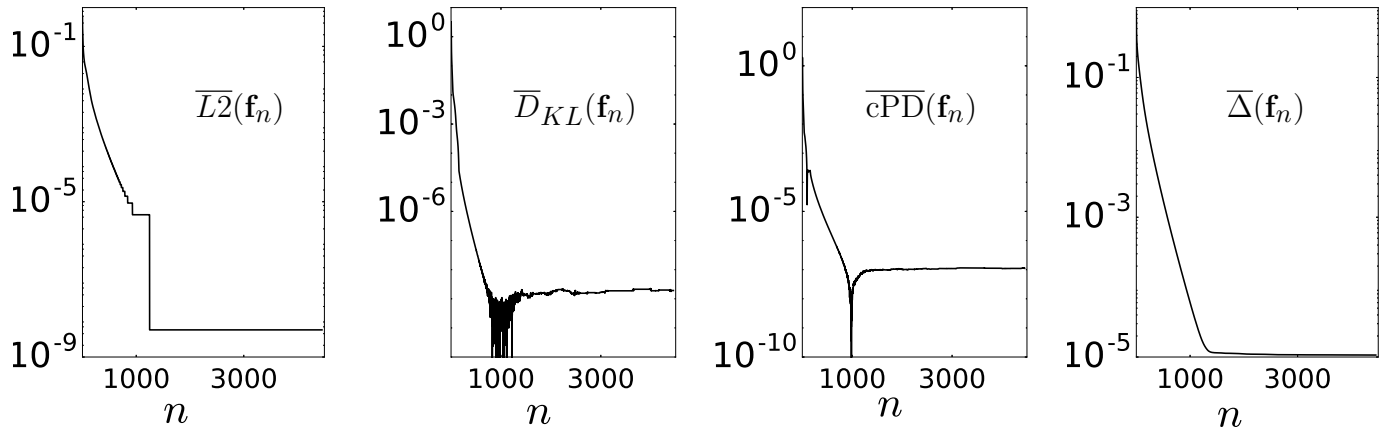


Figure 2.12: Convergence metrics $\overline{L2}(\mathbf{f}_n)$, $\overline{D}_{KL}(\mathbf{f}_n)$, $\overline{\mathrm{cPD}}(\mathbf{f}_n)$, and $\overline{\Delta}(\mathbf{f}_n)$ as functions of iteration number n for solving program D_{KL}-$||\mathbf{f}||_{\ell_2}$.

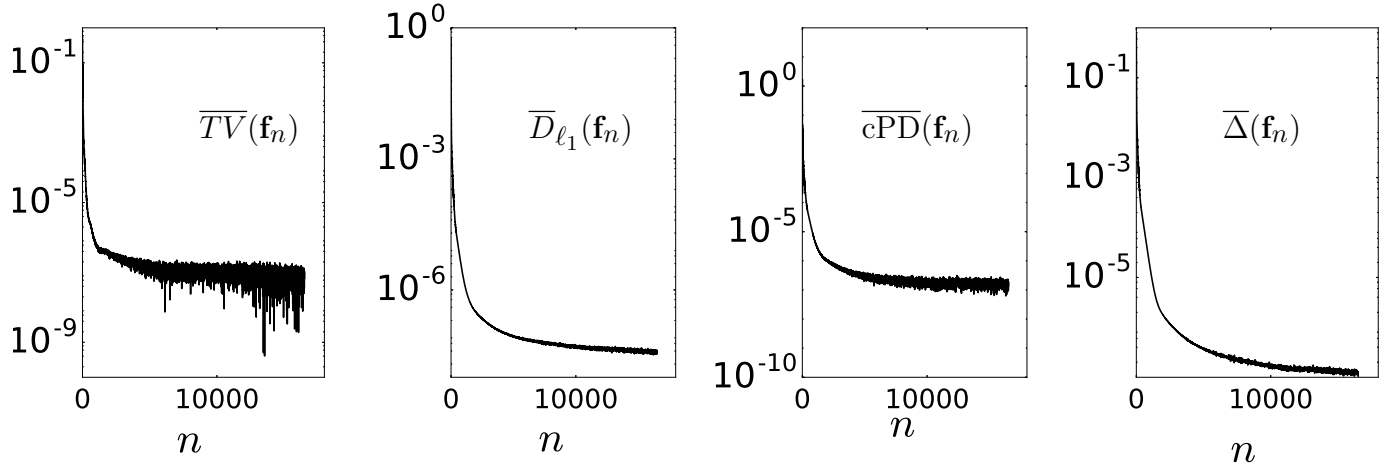


Figure 2.13: Convergence metrics $\overline{TV}(\mathbf{f}_n)$, $\overline{D}_{\ell_1}(\mathbf{f}_n)$, $\overline{\mathrm{cPD}}(\mathbf{f}_n)$, and $\overline{\Delta}(\mathbf{f}_n)$ as functions of iteration number n for solving program D_{ℓ_1}-$||\mathbf{f}||_{TV}$.

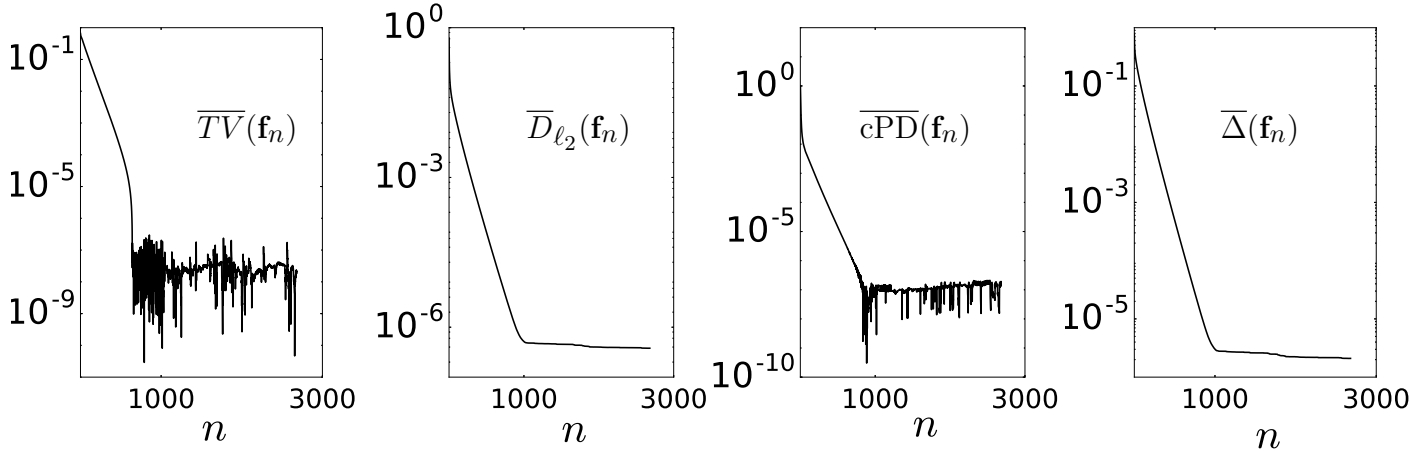


Figure 2.14: Convergence metrics $\overline{TV}(\mathbf{f}_n)$, $\overline{D}_{\ell_2}(\mathbf{f}_n)$, $\overline{\mathrm{cPD}}(\mathbf{f}_n)$, and $\overline{\Delta}(\mathbf{f}_n)$ as functions of iteration number n for solving program D_{ℓ_2}-$||\mathbf{f}||_{TV}$.

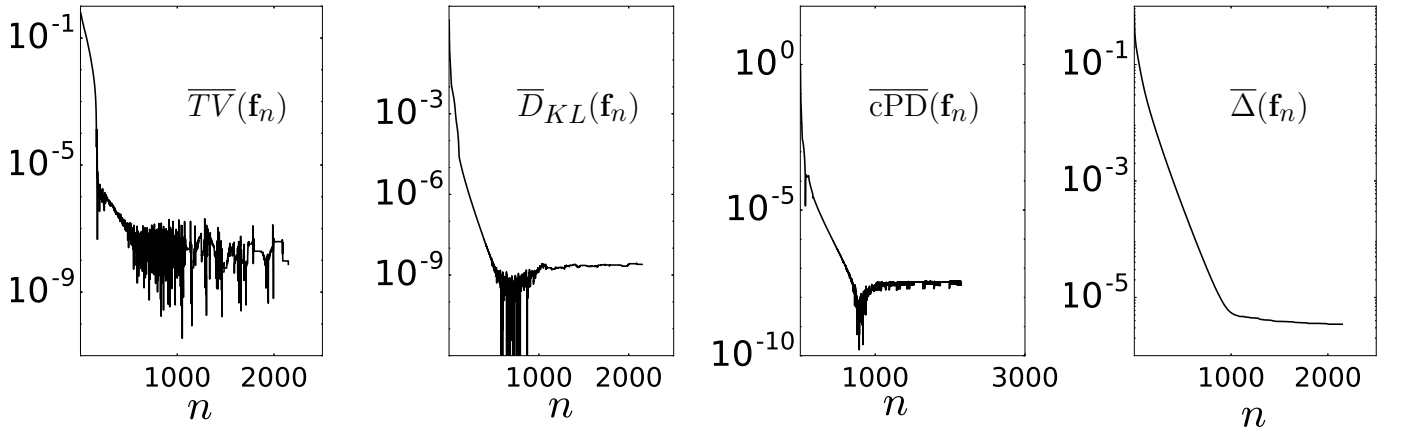


Figure 2.15: Convergence metrics $\overline{TV}(\mathbf{f}_n)$, $\overline{D}_{KL}(\mathbf{f}_n)$, $\overline{\mathrm{cPD}}(\mathbf{f}_n)$, and $\overline{\Delta}(\mathbf{f}_n)$ as functions of iteration number n for solving program D_{KL}-$||\mathbf{f}||_{TV}$.

2.4 Discussion

In this chapter, we have derived CP algorithms tailored to solve numerous convex optimization programs with different data fidelities and image constraints, which are of potential research and application interest. Both data fidelities and image constraints play an important role in the design of optimization-based reconstruction in cone-beam computed tomography (CBCT). We consider data fidelities such as data-ℓ_2, data-ℓ_1, and data-KL fidelities, and image constraints such as image-ℓ_2, image-ℓ_1, and image-TV. It is of theoretical as well as practical interest to investigate how different data fidelities and image constraints impact optimization-based reconstructions.

Because all the optimization programs considered in this work are convex, the investigation in this study is enabled by a primal-dual algorithm that is shown mathematically to solve convex optimization programs, especially for those under consideration in this chapter. By using a single algorithm for solving all of the optimization programs, it may avoid any reconstruction differences as a result of using different algorithms for solving different optimization programs.

In an attempt to validate the derived CP algorithms, we have carried out inverse-crime studies, where data were simulated with a known system matrix, and images were reconstructed by use of the exact system matrix. The results showed that the CP algorithm can accurately solve the designed programs. Therefore, the CP algorithm has been verified for its capability of solving the optimization programs. The inverse-crime study, however, does not imply algorithm properties in reconstruction scenario where data contain inconsistency, which are the cases to be studied in the following chapter.

2.5 Appendix

We take optimization program D_{ℓ_2}-$||\mathbf{f}||_{TV}$ as an example, formulate the primal and dual objectives in accordance to Eqs. (2.15) and (2.16), and derive its proximal-mapping. Derivation

for other optimization programs can be conducted in the same manner.

Derivation for program D_{ℓ_2}-$||\mathbf{f}||_{TV}$ We modify Eq. (2.10) by introducing two positive parameters, λ and ν,

$$\mathbf{f}^\star = \underset{\mathbf{f}}{\mathsf{argmin}} \frac{1}{2}\lambda D_{\ell_2}(\mathbf{f}) \quad \text{s.t.} \quad ||\mathbf{f}||_{TV} \leq \nu t_1 \quad \text{and} \quad f_j \geq 0. \tag{2.61}$$

It is worth noting that, parameter λ and ν have no impact on the designed solutions. We show below the CP algorithm tailored specifically to solve the program in Eq. (2.61), or equivalently, Eq. (2.10), which can be formulated into an unconstrained problem as:

$$\mathbf{f}^\star = \underset{\mathbf{f}}{\mathsf{argmin}} \left\{ \frac{1}{2}\lambda ||\mathcal{W}(\mathcal{H}\mathbf{f} - \mathbf{g})_m||_2^2 + \delta_{\text{Diamond}(\nu t_1)}(|\nu \nabla \mathbf{f}|_{\text{MAG}}) + \delta_P(\mathbf{f}) \right\} \tag{2.62}$$

where ∇ denotes a spatial gradient matrix of size $3N \times N$, with its transpose $\nabla^\top = (\nabla_x^\top, \nabla_y^\top, \nabla_z^\top)$, in which the superscript "$\top$" indicates a transpose operation; matrices ∇_x, ∇_y, and ∇_z of size $N \times N$ represent the finite difference matrices along x-, y-, and z-axis, respectively, yielding vectors $\nabla_x \mathbf{f}$, $\nabla_y \mathbf{f}$, and $\nabla_z \mathbf{f}$ of size N, which are used to form vector $\nabla \mathbf{f}$ of size $3N$ in a concatenated form in the order of x, y, and z; $|\nabla \mathbf{f}|_{\text{MAG}}$ depicts a vector of size N with entry j given by $(|\nabla \mathbf{f}|_{\text{MAG}})_j = \sqrt{(\nabla \mathbf{f})_j^2 + (\nabla \mathbf{f})_{j+N}^2 + (\nabla \mathbf{f})_{j+2N}^2}$, and $(\nabla \mathbf{f})_j$ indicates the jth entry of vector $\nabla \mathbf{f}$; the indicator function $\delta_{\text{Diamond}(\nu t_1)}(\mathbf{x})$ is defined as

$$\delta_{\text{Diamond}(\nu t_1)}(\mathbf{x}) = \begin{cases} 0 & ||\mathbf{x}||_1 \leq \nu t_1 \\ \infty & ||\mathbf{x}||_1 > \nu t_1, \end{cases}. \tag{2.63}$$

and the indicator function $\delta_P(\mathbf{e})$ is written as:

$$\delta_P(\mathbf{x}) = \begin{cases} 0 & \mathbf{x} \geq 0 \\ \infty & \mathbf{x} < 0. \end{cases}. \tag{2.64}$$

For the convenience of discussion, we consider two vectors $\mathbf{a}$ and $\mathbf{b}$ of sizes kN and N, where $k \geq 1$ is a positive integer. Let a_j denote the jth entry of $\mathbf{a}$, $j = 1, 2, ..., kN$, and b_j the jth entry of $\mathbf{b}$, $j = 1, 2, ..., N$, respectively. We use vectors $\mathbf{ab}$ and $\mathbf{a/b}$ of sizes kN to denote the multiplication and division of $\mathbf{a}$ and $\mathbf{b}$, with entry j given by

$$(\mathbf{ab})_j = a_j b_{j_m}$$
$$(\mathbf{a/b})_j = a_j / b_{j_m},$$

where $j = 1, 2, ..., kN$, and $j_m = \mathrm{mod}\,(j, N)$ indicates the remainder of j dividing by N, and $b_{j_m} \neq 0$. In particular, when $k = 1$, $\mathbf{a}$ and $\mathbf{b}$ are of the same size N, and vectors $\mathbf{ab}$ and $\mathbf{a/b}$ thus are of size N and depict element-wise multiplication and division.

In order to match the form of primal minimization in Eq. (2.15), we introduce the following assignments:

$$\begin{cases} G(\mathbf{x}) = \delta_P(\mathbf{x}), \\ F(\mathbf{r}, \mathbf{z}) = F_1(\mathbf{r}) + F_2(\mathbf{z}), \\ F_1(\mathbf{r}) = \frac{1}{2}\lambda ||\mathbf{r} - \mathcal{W}\mathbf{g}_m||_2^2, \\ F_2(\mathbf{z}) = \delta_{\mathrm{Diamond}(\nu t_1)}(|\mathbf{z}|_{\mathrm{MAG}}), \end{cases} \tag{2.65}$$

and the symbols in Eq. (2.15) are defined as:

$$\begin{cases} \mathbf{x} = \mathbf{f}, \\ \mathbf{r} = \mathcal{W}\mathcal{H}\mathbf{f}, \\ \mathbf{z} = \nu\nabla\mathbf{f}, \\ \mathcal{K} = \begin{pmatrix} \mathcal{W}\mathcal{H} \\ \nu\nabla \end{pmatrix}. \end{cases} \tag{2.66}$$

Conjugate functions: For functions F_1, F_2 and G, the convex conjugate functions, F_1^*, F_2^*

and G^*, are given by:

$$\begin{aligned} F_1^*(\mathbf{p}) =& \max_{\mathbf{r}}\{\mathbf{p}^\top \mathbf{r} - F_1(\mathbf{r})\} \\ =& \max_{\mathbf{r}}\{\mathbf{p}^\top \mathbf{r} - \frac{1}{2}\lambda||\mathbf{r} - \mathcal{W}\mathbf{g}_m||_2^2\} \\ =& \frac{1}{2\lambda}||\mathbf{p}||_2^2 + \mathbf{g}_m^\top \mathcal{W}^\top \mathbf{p}, \end{aligned} \tag{2.67}$$

$$\begin{aligned} F_2^*(\mathbf{q}) =& \max_{\mathbf{z}}\{\mathbf{q}^\top \mathbf{z} - F_2(\mathbf{z})\} \\ =& \max_{\mathbf{z}}\{\mathbf{q}^\top \mathbf{z} - \delta_{\text{Diamond}(\nu\gamma)}(|\mathbf{z}|_{\text{MAG}})\}, \\ =& \nu\gamma||(|\mathbf{q}|_{\text{MAG}})||_\infty \end{aligned} \tag{2.68}$$

$$\begin{aligned} G^*(\mathbf{y}) =& \max_{\mathbf{x}}\{\mathbf{y}^\top \mathbf{x} - G(\mathbf{x})\} \\ =& \max_{\mathbf{x}}\{\mathbf{y}^\top \mathbf{x} - \delta_P(\mathbf{x})\} \\ =& \delta_P(-\mathbf{y}) \end{aligned} \tag{2.69}$$

where $||\cdot||_\infty$ norm denotes the largest component of the argument.

Therefore the dual optimization of Eq. (2.62) can be written as

$$\begin{aligned} (\mathbf{p}^*, \mathbf{q}^*) =& \max_{\mathbf{p},\mathbf{q}}\{-F^*(\mathbf{p},\mathbf{q}) - G^*(\mathcal{K}^T \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix})\} \\ =& \max_{\mathbf{p},\mathbf{q}}\{-\frac{1}{2\lambda}||\mathbf{p}||_2^2 - \mathbf{g}_m^\top \mathcal{W}^\top \mathbf{p} - \nu t_1||(|\mathbf{q}|)||_\infty - \delta_P(-\mathcal{H}^\top \mathcal{W}^\top \mathbf{p} - \nu\nabla^\top \mathbf{q})\}. \end{aligned} \tag{2.70}$$

In an attempt to implement Chambolle-Pock algorithm, we need to derive the proximal mapping for functions F_1^*, G, and F_2^*, which are given below:

The proximal mapping for F_1^*:

$$\begin{aligned}\mathsf{prox}_{\sigma}[F_1^*](\mathbf{p}) =& \underset{\mathbf{p}'}{\mathsf{argmin}} \left\{ F_1^*(\mathbf{p}') + \frac{||\mathbf{p}-\mathbf{p}'||_2^2}{2\sigma} \right\} \\ =& \underset{\mathbf{p}'}{\text{argmin}} \left\{ \frac{1}{2\lambda}||\mathbf{p}'||_2^2 + < \mathcal{W}\mathbf{g}_m, \mathbf{p}' > + \frac{||\mathbf{p}-\mathbf{p}'||_2^2}{2\sigma} \right\} \\ =& (\mathbf{p} - \sigma\mathcal{W}\mathbf{g}_m)/(1+\sigma/\lambda)\end{aligned} \tag{2.71}$$

The proximal mapping for F_2^*: By using Moreau identity [169], the proximal mapping $\mathsf{prox}_{\sigma}[F_2^*](\mathbf{q})$ can be written as

$$\begin{aligned}\mathsf{prox}_{\sigma}[F_2^*](\mathbf{q}) =& \underset{\mathbf{q}'}{\mathsf{argmin}} \left\{ F_2^*(\mathbf{q}') + \frac{||\mathbf{q}-\mathbf{q}'||_2^2}{2\sigma} \right\} \\ =& \underset{\mathbf{q}'}{\mathsf{argmin}} \left\{ \nu\gamma||(|\mathbf{q}'|_{\text{MAG}})||_\infty + \frac{||\mathbf{q}-\mathbf{q}'||_2^2}{2\sigma} \right\} \\ =& \mathbf{q} - \sigma\mathsf{prox}_{\frac{1}{\sigma}}[F_2](\frac{\mathbf{q}}{\sigma}) \\ =& \mathbf{q} - \sigma\underset{\mathbf{q}'}{\mathsf{argmin}} \left\{ \delta_{\text{Diamond}(\nu t_1)}(|\mathbf{q}'|) + \frac{\sigma||\mathbf{q}/\sigma - \mathbf{q}'||_2^2}{2} \right\}\end{aligned} \tag{2.72}$$

We rewrite $\mathbf{q} = |\mathbf{q}|_{\text{MAG}}\hat{\mathbf{q}}$ and $\mathbf{q}' = |\mathbf{q}'|_{\text{MAG}}\hat{\mathbf{q}}'$, where $\hat{\mathbf{q}}$ and $\hat{\mathbf{q}}'$ are vectors of $3N$ denoting the direction of $\mathbf{q}$ and $\mathbf{q}'$. Because the indicator function in Eq. (2.72) is independent of $\hat{\mathbf{q}}'$, the second term in Eq. (2.72) is minimized for a given $|\mathbf{q}'|_{\text{MAG}}$ as $\hat{\mathbf{q}}$ and $\hat{\mathbf{q}}'$ are pointing to the same direction. Therefore, Eq. (2.72) can be expressed as:

$$
\begin{aligned}
\mathsf{prox}_{\sigma}[F_2^*](\mathbf{q}) =& \mathbf{q} - \sigma \underset{|\mathbf{q}'|_{\mathrm{MAG}}, \hat{\mathbf{q}}'}{\mathsf{argmin}} \left\{ \delta_{\mathrm{Diamond}(\nu t_1)}(|\mathbf{q}'|_{\mathrm{MAG}}) + \frac{\sigma ||(|\mathbf{q}|_{\mathrm{MAG}}\hat{\mathbf{q}})/\sigma - (|\mathbf{q}'|_{\mathrm{MAG}}\hat{\mathbf{q}}')||_2^2}{2} \right\} \\
=& \mathbf{q} - \sigma \hat{\mathbf{q}} \underset{|\mathbf{q}'|_{\mathrm{MAG}}}{\mathsf{argmin}} \left\{ \delta_{\mathrm{Diamond}(\nu t_1)}(|\mathbf{q}'|_{\mathrm{MAG}}) + \frac{||(|\mathbf{q}|_{\mathrm{MAG}})/\sigma - (|\mathbf{q}'|_{\mathrm{MAG}})||_2^2}{2/\sigma} \right\} \\
=& \mathbf{q} - \sigma \frac{\mathbf{q}}{|\mathbf{q}|_{\mathrm{MAG}}} \mathbf{s} \\
=& \mathbf{q}(\mathbf{1}_I - \sigma \mathbf{s}/|\mathbf{q}|_{\mathrm{MAG}}),
\end{aligned}
\tag{2.73}
$$

where $\mathbf{1}_I$ is a vector of size N with all entries set to be 1, and vector $\mathbf{s}$ of size N is given by:

$$
\begin{aligned}
\mathbf{s} &= \underset{|\mathbf{q}'|_{\mathrm{MAG}}}{\mathsf{argmin}} \left\{ \delta_{\mathrm{Diamond}(\nu t_1)}(|\mathbf{q}'|_{\mathrm{MAG}}) + \frac{||(|\mathbf{q}|_{\mathrm{MAG}})/\sigma - (|\mathbf{q}'|_{\mathrm{MAG}})||_2^2}{2/\sigma} \right\} \\
&= \mathbf{ProjOnto}\ell_1\mathbf{Ball}_{\nu t_1}(|\mathbf{q}|_{\mathrm{MAG}}/\sigma),
\end{aligned}
$$

where operator $\mathbf{ProjOnto}\ell_1\mathbf{Ball}_{\nu t_1}$ projects vector $|\mathbf{q}|_{\mathrm{MAG}}/\sigma$ onto the ℓ_1-ball of scale νt_1.

The proximal mapping for G^*:

$$
\begin{aligned}
\mathsf{prox}_{\tau}[G](\mathbf{x}) =& \underset{\mathbf{x}'}{\mathsf{argmin}} \left\{ G^*(\mathbf{x}') + \frac{||\mathbf{x} - \mathbf{x}'||_2^2}{2\tau} \right\} \\
=& \underset{\mathbf{x}'}{\mathsf{argmin}} \left\{ \delta_P(-\mathbf{x}) + \frac{||\mathbf{x} - \mathbf{x}'||_2^2}{2\tau} \right\} \\
=& [\mathsf{pos}(\mathbf{x})]_i \\
=& \begin{cases} x_i & x_i > 0 \\ 0 & x_i \leq 0 \end{cases}
\end{aligned}
\tag{2.74}
$$

The proximal-mapping for other programs in Table 2.1 can be derived in the manner. It is worth noting that, for programs containing data-ℓ_2 fidelity, the function F_1 in Eq. (2.65)

can be written as:

$$F_1(\mathbf{r}) = \frac{1}{2}\lambda||\mathbf{r} - \mathcal{W}\mathbf{g}_m||_2^2, \tag{2.75}$$

whose conjugate function and proximal-mapping have been derived in Eqs. (2.67) and (2.71), respectively. For programs containing data-ℓ_1 fidelity, the function F_1 in Eq. (2.65) is:

$$F_1(\mathbf{r}) = \lambda||\mathbf{r} - \mathcal{W}\mathbf{g}_m||_1. \tag{2.76}$$

For programs containing data-KL fidelity, the function F_1 in Eq. (2.65) is:

$$F_1(\mathbf{r}) = \lambda \sum_i [\mathbf{r} - \mathcal{W}\mathbf{g}_m + \mathcal{W}\mathbf{g}_m \ln(\mathcal{W}\mathbf{g}_m) - \mathcal{W}\mathbf{g}_m \ln(\mathbf{r})]_i, \tag{2.77}$$

where all entries in vector $\mathbf{r}$ of size M that are smaller than 10^{-20} are set to be 10^{-20}.

We also derived the corresponding conjugate functions and the proximal-mapping for different image constraints in Table 2.1. For programs containing image-TV constraint, the function F_2 in Eq. (2.65) is:

$$\begin{cases} F_2(\mathbf{z}) = \delta_{\text{Diamond}(\nu t_1)}(|\mathbf{z}|), \\ \mathbf{z} = \nu \nabla \mathbf{f}, \end{cases} \tag{2.78}$$

whose conjugate function and proximal-mapping are derived in Eqs. (2.68) and (2.72), respectively. For programs containing image-ℓ_1 constraint, the function F_2 in Eq. (2.65) is:

$$\begin{cases} F_2(\mathbf{z}) = \delta_{\text{Diamond}(\nu l_1)}(|\mathbf{z}|), \\ \mathbf{z} = \nu \mathbf{f}. \end{cases} \tag{2.79}$$

For programs containing image-ℓ_2 constraint, the function F_2 in Eq. (2.65) is:

$$\begin{cases} F_2(\mathbf{z}) = \delta_{\text{Ball}(\nu l_2)}(\mathbf{z}), \\ \mathbf{z} = \nu \mathbf{f} \end{cases} \tag{2.80}$$

Here, $\delta_{\text{Ball}(a)}(\mathbf{x})$ is an indicator function defined as:

$$\delta_{\text{Ball}(a)}(\mathbf{x}) = \begin{cases} 0 & ||\mathbf{x}||_2 \leq a \\ \infty & ||\mathbf{x}||_2 > a \end{cases} \tag{2.81}$$

In the following section, we will derive the conjugate functions and the proximal-mapping of Eqs. (2.76) and (2.77) for function F_1, and Eqs. (2.79) and (2.80) for function F_2.

Data-ℓ_1 fidelity Computing the convex conjugate of Eq. (2.76) yields

$$\begin{aligned} F_1^*(\mathbf{p}) =& \max_{\mathbf{r}} \{\mathbf{p}^\top \mathbf{r} - F_1(\mathbf{r})\} \\ =& \max_{\mathbf{r}} \{\mathbf{p}^\top \mathbf{r} - \lambda ||\mathbf{r} - \mathcal{W}\mathbf{g}_m||_1\} \\ =& \mathbf{g}_m^\top \mathcal{W}^\top \mathbf{p} + \delta_{\text{Box}(\lambda)}(\mathbf{p}). \end{aligned} \tag{2.82}$$

The corresponding proximal-mapping of F_1^* is:

$$\begin{aligned} \text{prox}_\sigma[F_1^*](\mathbf{p}) =& \operatorname*{argmin}_{\mathbf{p}'} \left\{ F_1^*(\mathbf{p}') + \frac{||\mathbf{p} - \mathbf{p}'||_2^2}{2\sigma} \right\} \\ =& \operatorname*{argmin}_{\mathbf{p}'} \left\{ \mathbf{g}_m^\top \mathcal{W}^\top \mathbf{p} + \delta_{\text{Box}(\lambda)}(y) + \frac{||\mathbf{p} - \mathbf{p}'||_2^2}{2\sigma} \right\} \\ =& \frac{(\mathbf{p} - \sigma \mathcal{W}\mathbf{g}_m)}{\max(\mathbf{1}_\text{D},\, |\mathbf{p} - \sigma \mathcal{W}\mathbf{g}_\text{m}|/\lambda)}, \end{aligned} \tag{2.83}$$

where $\mathbf{1}_D$ is a vector of size M with all entries set to be 1.

Data-KL fidelity Computing the convex conjugate of Eq. (2.77) yields

$$
\begin{aligned}
F_1^*(\mathbf{p}) =& \max_{\mathbf{r}}\{\mathbf{p}^\top \mathbf{r} - \lambda \sum_i [\mathbf{r} - \mathcal{W}\mathbf{g}_m + \mathcal{W}\mathbf{g}_m \ln(\mathcal{W}\mathbf{g}_m) - \mathcal{W}\mathbf{g}_m \ln(\mathbf{r})]_i\} \\
=& \sum_i [-\lambda \mathbf{g}_m^\top \mathcal{W}^\top \ln(\mathbf{1}_D - \frac{\mathbf{p}}{\lambda})]_i + \delta_P(\boldsymbol{\lambda}_D - \mathbf{p}),
\end{aligned}
\tag{2.84}
$$

where $\boldsymbol{\lambda}_D$ is a vector of size M with all entries set to be 1, and all entries in vector $(\mathbf{1}_D - \frac{\mathbf{p}}{\lambda})$ that are smaller than 10^{-20} are set to be 10^{-20}.

The corresponding proximal-mapping of F_1^* is:

$$
\begin{aligned}
\mathrm{prox}_\sigma[F_1^*](\mathbf{p}) =& \operatorname*{argmin}_{\mathbf{p}'} \left\{ F_1^*(\mathbf{p}') + \frac{||\mathbf{p} - \mathbf{p}'||_2^2}{2\sigma} \right\} \\
=& \operatorname*{argmin}_{\mathbf{p}'} \left\{ \sum_i [-\lambda \mathbf{g}_m^\top \mathcal{W}^\top \ln(\mathbf{1}_D - \frac{\mathbf{p}'}{\lambda})]_i + \delta_P(\boldsymbol{\lambda}_D - \mathbf{p}') + \frac{||\mathbf{p} - \mathbf{p}'||_2^2}{2\sigma} \right\} \\
=& \frac{1}{2}(\boldsymbol{\lambda}_D + \mathbf{p} \pm \sqrt{(\lambda - \mathbf{p})^2 + 4\sigma\lambda\mathcal{W}\mathbf{g}_m} \\
\overset{\lambda - \mathbf{p}' > 0}{=}& \frac{1}{2}(\boldsymbol{\lambda}_D + \mathbf{p} - \sqrt{(\lambda - \mathbf{p})^2 + 4\sigma\lambda\mathcal{W}\mathbf{g}_m}
\end{aligned}
\tag{2.85}
$$

Image-ℓ_1 constraint Computing the convex conjugate of Eq. (2.79) yields

$$
\begin{aligned}
F_2^*(\mathbf{q}) =& \max_{\mathbf{z}}\{\mathbf{q}^\top \mathbf{z} - F_2(\mathbf{z})\} \\
=& \max_{\mathbf{z}}\{\mathbf{q}^\top \mathbf{z} - \delta_{\mathrm{Diamond}(\nu l_1)}(\mathbf{z})\}, \\
=& \nu l_1 ||\mathbf{q}||_\infty,
\end{aligned}
\tag{2.86}
$$

where $\mathbf{q}$ is a vector of size N.

The corresponding proximal-mapping of F_2^* is:

$$
\begin{aligned}
\mathrm{prox}_\sigma[F_2^*](\mathbf{q}) &= \mathbf{q} - \sigma \mathrm{prox}_{\frac{1}{\sigma}}[F_2](\frac{\mathbf{q}}{\sigma}) \\
&= \mathbf{q} - \sigma \operatorname*{argmin}_{\mathbf{q}'} \left\{ \delta_{\mathrm{Diamond}(\nu l_1)}(|\mathbf{q}'|) + \frac{\sigma||\mathbf{q}/\sigma - \mathbf{q}'||_2^2}{2} \right\}
\end{aligned}
\tag{2.87}
$$

We rewrite $\mathbf{q} = |\mathbf{q}|\hat{\mathbf{q}}$ and $\mathbf{q}' = |\mathbf{q}'|\hat{\mathbf{q}}'$, $|\mathbf{q}|$ depicts a vector of size N with entry $(|\mathbf{q}|)_j = |q_j|$, and q_j is the jth entry of vector $\mathbf{q}$. Then Eq. (2.87) is expressed as:

$$\begin{aligned}
\mathsf{prox}_{\sigma}[F_2^*](\mathbf{q}) =& \mathbf{q} - \sigma \underset{|\mathbf{q}'|,\hat{\mathbf{q}}'}{\mathsf{argmin}} \left\{ \delta_{\mathrm{Diamond}(\nu l_1)}(|\mathbf{q}'|) + \frac{\sigma||(|\mathbf{q}|\hat{\mathbf{q}})/\sigma - (|\mathbf{q}'|\hat{\mathbf{q}}')||_2^2}{2} \right\} \\
=& \mathbf{q} - \sigma\hat{\mathbf{q}} \underset{|\mathbf{q}'|}{\mathsf{argmin}} \left\{ \delta_{\mathrm{Diamond}(\nu l_1)}(|\mathbf{q}'|) + \frac{\sigma||(|\mathbf{q}|)/\sigma - (|\mathbf{q}'|)||_2^2}{2} \right\} \\
=& \mathbf{q}(\mathbf{1}_I - \sigma \mathbf{s}/|\mathbf{q}|),
\end{aligned} \tag{2.88}$$

where vector $\mathbf{s}$ of size N is given by:

$$\begin{aligned}
\mathbf{s} &= \underset{|\mathbf{q}'|}{\mathsf{argmin}} \left\{ \delta_{\mathrm{Diamond}(\nu l_1)}(|\mathbf{q}'|) + \frac{||(|\mathbf{q}|)/\sigma - (|\mathbf{q}'|)||_2^2}{2/\sigma} \right\} \\
&= \mathbf{ProjOnto}\ell_1\mathbf{Ball}_{\nu l_1}(|\mathbf{q}|/\sigma),
\end{aligned}$$

where operator $\mathbf{ProjOnto}\ell_1\mathbf{Ball}_{\nu l_1}$ projects vector $|\mathbf{q}|/\sigma$ onto the ℓ_1-ball of scale νl_1.

Image-ℓ_2 constraint Computing the convex conjugate of Eq. (2.80) yields

$$\begin{aligned}
F_2^*(\mathbf{q}) =& \max_{\mathbf{q}'}\{\mathbf{q}^\top\mathbf{q}' - F_2(\mathbf{q}')\} \\
=& \max_{\mathbf{q}'}\{\mathbf{q}^\top\mathbf{q}' - \delta_{\mathrm{Ball}(\nu l_2)}(\mathbf{q}')\}, \\
=& \nu l_2||\mathbf{q}||_2,
\end{aligned} \tag{2.89}$$

where $\mathbf{q}$ is a vector of size N.

The corresponding proximal-mapping of F_2^* is:

$$\begin{aligned}
\mathsf{prox}_{\sigma}[F_2^*](\mathbf{q}) =& \mathbf{q} - \sigma \mathsf{prox}_{\frac{1}{\sigma}}[F_2](\frac{\mathbf{q}}{\sigma}) \\
=& \mathbf{q} - \sigma \underset{\mathbf{q}'}{\mathsf{argmin}} \left\{ \delta_{\mathrm{Ball}(\nu l_2)}(\mathbf{q}') + \frac{\sigma ||\mathbf{q}/\sigma - \mathbf{q}'||_2^2}{2} \right\} \\
=& \mathbf{q} - \sigma \underset{|\mathbf{q}'|, \hat{\mathbf{q}}'}{\mathsf{argmin}} \left\{ \delta_{\mathrm{Ball}(\nu l_2)}(|\mathbf{q}'|) + \frac{\sigma ||(|\mathbf{q}|\hat{\mathbf{q}})/\sigma - (|\mathbf{q}'|\hat{\mathbf{q}}')||_2^2}{2} \right\} \\
=& \mathbf{q} - \sigma \hat{\mathbf{q}} \underset{|\mathbf{q}'|}{\mathsf{argmin}} \left\{ \delta_{\mathrm{Ball}(\nu l_2)}(|\mathbf{q}'|) + \frac{\sigma ||(|\mathbf{q}|)/\sigma - (|\mathbf{q}'|)||_2^2}{2} \right\} \\
=& \mathbf{q}(\mathbf{1}_I - \sigma \mathbf{s}/|\mathbf{q}|),
\end{aligned} \tag{2.90}$$

where $\mathbf{q} = |\mathbf{q}|\hat{\mathbf{q}}$ and $\mathbf{q}' = |\mathbf{q}'|\hat{\mathbf{q}}'$; $|\mathbf{q}|$ depicts a vector of size N with entry $(|\mathbf{q}|)_j = |q_j|$, and q_j is the jth entry of vector $\mathbf{q}$; and vector $\mathbf{s}$ of size N is given by:

$$\begin{aligned}
\mathbf{s} &= \underset{|\mathbf{q}'|}{\mathsf{argmin}} \left\{ \delta_{\mathrm{Ball}(\nu l_2)}(|\mathbf{q}'|) + \frac{||(|\mathbf{q}|)/\sigma - (|\mathbf{q}'|)||_2^2}{2/\sigma} \right\} \\
&= \mathbf{ProjOnto}\ell_2\mathbf{Ball}_{\nu l_2}(|\mathbf{q}|/\sigma),
\end{aligned}$$

where operator $\mathbf{ProjOnto}\ell_2\mathbf{Ball}_{\nu l_2}$ projects vector $|\mathbf{q}|/\sigma$ onto the ℓ_2-ball of scale νl_2.

CHAPTER 3

ARTIFACT REDUCTION IN SHORT-SCAN CBCT BY USE OF OPTIMIZATION-BASED RECONSTRUCTION

3.1 Introduction

The previous chapter focused on optimization problem design and derivation of the CP algorithm instances, which were verified numerically in inverse-crime studies. In this chapter, we carry out qualitative characterization and evaluation studies on the optimization algorithms by employing data acquired from Catphan and Rando phantoms.

This study serves two purposes. The first is to characterize and evaluate the CP algorithm for solving different optimization programs in practical CBCT imaging applications, where data unavoidably contain inconsistencies. In the Catphan-phantom study, the ground truth is available from the phantom manufacturer, thus allowing us to verify the reconstruction against the known truth. In the Rando-phantom study, the scanned object contains more realistic and complex structures, which yielding CBCT images of practical interest.

The second purpose of the study is to investigate how different designs of optimization-based reconstruction impact artifact reduction, especially in the off-middle plane. In clinical CBCT applications, although the FDK algorithm can provide images of high practical utility for a wide class of applications, it has also been observed to yield reconstructions with noticeable artifacts especially in the off-middle planes for short-scan CBCT with data insufficiency and inconsistency [170, 171]. Evidence from previous experience seems to suggest that appropriately designed optimization-based algorithms may be less susceptible to data inconsistency and insufficiency than analytic algorithms such as the FDK algorithm, thus possibly yielding reconstructions with reduced artifacts [86,170]. In this work, we carry out a study on optimization-based reconstructions from short-scan CBCT data with an emphasis on demonstrating whether different designs of optimization programs of potential practical interest impact the reduction of artifacts observed in the FDK reconstruction.

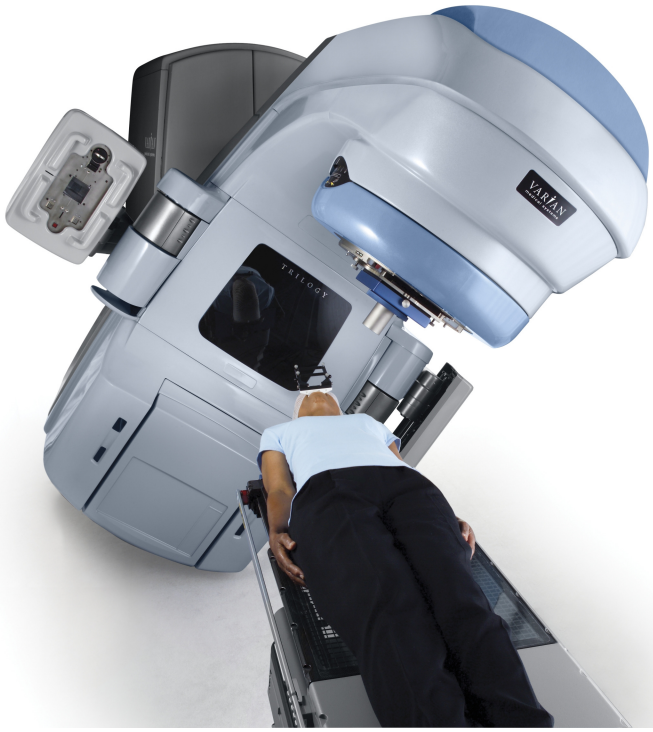


Figure 3.1: On-board Imaging (OBI) system on a Trilogy LINAC.

This chapter starts with an introduction in Sec. 3.1, followed by the study design in Sec. 3.2. Results of real-data studies and final remarks are presented in Secs. 3.3 and 3.4, respectively.

3.2 Study Design and Materials

3.2.1 Data collection and FDK reference

Short-scan CBCT data We collected data from two physical phantoms, the RANDO man and the Catphan 504 (The Phantom Laboratory, Salem NY), using a Varian on-board-imager (OBI) on a Trilogy LINAC (Varian Medical System, Palo Alto CA), shown in Fig. 3.1. The two phantoms are selected because they contain both anthropomorphic structures in RANDO and geometric, quantitative structures in the Catphan that can be used for evaluating the impact of reconstruction algorithms. The OBI CBCT system consists of

a keV X-ray source and a flat-panel detector of sizes $2u_m = 39.73$ cm and $2v_m = 29.80$ cm, composed of 1024×768 identical square-shaped-detector bins of size 0.0388 cm. The distances from the source to the rotation axis, and to the flat-detector surface, are $S = 100$ cm and $L = 150$ cm, respectively. Data were acquired at 353 and 347 views, respectively, for the Rando and Catphan phantoms, uniformly distributed over a short-scan angular range of 196 degrees. Therefore, the sizes of data vectors considered are $M = 1024 \times 768 \times 353$ and $1024 \times 768 \times 347$ for the cases involving, respectively, the Rando and Catphan phantoms.

FDK reference For comparison, we also carried out reconstructions by using the FDK algorithm, with a Hanning filter and a cutoff at 0.5 of the Nyquist frequency, from weighted real-data $\mathcal{W}\mathbf{g}_m$. For revealing details of artifact reduction, soft-tissue display windows are used for reconstructions; and, as shown in Fig. 3.2, two regions of interest (ROIs) enclosed by rectangles within each of the FDK reconstructions of the Rando and Catphan phantoms are selected for showing ROI images in zoomed-in views.

3.2.2 Program-parameter selection

Weighting matrix and image-voxel size Results below were obtained with a diagonal weighting matrix that is formed from discrete values of the Parker weighting function, except for the two cases in Sec. 3.3.3, in which different weighting matrices were used. In a real-data study, image-voxel size and image-constraint parameter need to be estimated, because only measured data $\mathbf{g}_m$ are available experimentally [72,168]. In the Rando and Catphan studies, we use image arrays of sizes $N = 512 \times 512 \times 451$ and $512 \times 512 \times 239$, respectively, consisting of cubic voxels of size 0.488 mm, identical to that used in clinical IGRT CBCT imaging, which is 1.258 times of the detector-bin size of OBI CBCT scanner used for collecting data.

Image-constraint parameter We performed multiple reconstructions from each data set collected, with different t_1 values. Based upon a visual inspection of these reconstructions, we then select t_1 that yields a reconstruction with reduced artifacts comparing to the FDK

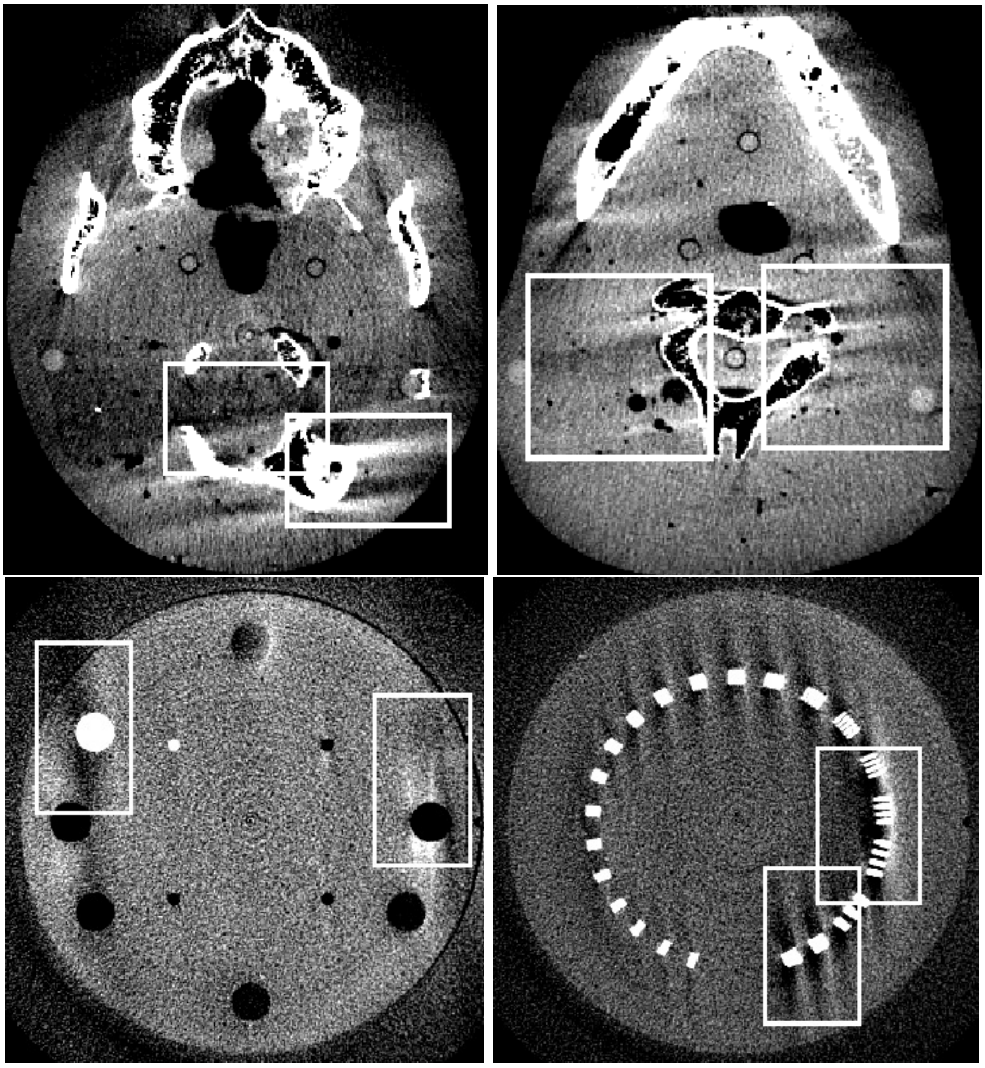


Figure 3.2: ROIs indicated by rectangles within transverse slices at 3.37 cm (top left) and 7.32 cm (top right) from the middle plane of the Rando phantom, and at 4.15 cm (bottom left) and 6.0 cm (bottom right) from the middle plane of the Catphan phantom. ROI images are displayed in zoomed-in views below.

image, yet without worsened, or with improved, contrast and spatial resolution. Such a way of selecting t_1 is admittedly subjective, but it seems sufficient for the purpose of the work which is qualitatively demonstrating the potential of image-TV-constrained reconstruction for reduction of artifacts observed in FDK reconstruction from short-scan CBCT data. Similarly, we determined image-constraint parameters l_1 or l_2 in optimization programs using image-ℓ_1 or image-ℓ_2 constraint by performing reconstructions with varying values of l_1 or of l_2 instead, and then selecting the l_1 or l_2 that yields the reconstruction with visually the least significant artifacts. The specific selection of an image-constraint parameter is generally data- and object-dependent.

Practical convergence conditions The CP algorithm has been shown to converge mathematically to solutions specified by a convex optimization program, including the optimization programs considered, but with an infinite number of iterations, which cannot be achieved in practice. Instead, using the mathematical convergence conditions as a guide, we developed practical convergence conditions such that they can be achieved within a finite number of iterations [72, 168]. Of course, practical convergence conditions should be interpreted as additional program parameters because they play now, along with other program parameters, a role in specifying feasible solutions. We consider in real-data studies three practical convergent conditions to be satisfied collectively as functions of iteration numbers: *1*) the image-ℓ_1, image-ℓ_2, or image-TV constraint of the reconstructed image arrives at its corresponding, designed image constraint value; *2*) the data fidelity data-ℓ_2, data-ℓ_1, or data-KL sufficiently plateau; and *3*) the conditional primal-dual gap (cPD) [164] continuously decreases when conditions *1*) and *2*) are reached. (The mathematical cPD condition on the CP-algorithm convergence is that the cPD approaches zero. Depending upon the property of the system matrix, data condition, optimization program, and computer precision, it is often unrealistic to achieve numerically the cPD condition in a practical reconstruction. As the results below show, however, the three practical convergence conditions defined above are able to yield

reconstructions of practical interest.)

3.3 Study Results

From the perspective of a C-C data model for circular CBCT, short-scan data sufficient for accurate reconstruction within the middle plane are insufficient for non-middle planes. In a real-data study, data inconsistency, including noise and scatter, can lead to artifacts within the middle plane. However, for non-middle planes, artifacts result from not only data inconsistency but also data insufficiency. The results below highlight three major studies: (1) We first study optimization programs in Table 2.1 for reconstructions within the middle plane, which contains no data-insufficiency artifacts. (2) Based upon the study results, we select the optimization programs that are shown to perform well based upon visual inspection for further investigation of their potential in reduction of artifacts observed within non-middle planes of FDK reconstructions caused by data insufficiency. (3) Additional studies are performed on how the artifact reduction is affected by the selection of additional optimization programs and weighting matrices.

3.3.1 Impact of optimization programs on reconstructions

Reconstructions based upon optimization programs in row 2 of Table 2.1 We reconstructed images from Rando data by solving the optimization programs in row 2 of Table 2.1, and show the convergent reconstructions in row 1 of Fig. 3.3. As expected, because the optimization programs contain only data fidelities (except for an image positivity constraint,) the convergent reconstructions are all highly noisy, with image textures considerably different than that in the FDK reconstruction. Visual differences exist among the reconstructions because different designs of optimization programs necessarily respond differently to data inconsistencies such as data noise. It can be observed that the reconstruction of program D_{ℓ_1} with data-ℓ_1 fidelity appears with sprinkled-salt texture, and is noisier than those obtained

with programs D_{ℓ_2} and D_{KL} containing data-ℓ_2 and data-KL fidelities. Results similar to Rando reconstructions have been obtained also for Catphan reconstructions, which are not shown. The study suggests the need of adequate image constraints for image reconstruction from real data.

Reconstructions based upon optimization programs in rows 3-5 of Table 2.1 We then performed reconstructions from Rando data by solving the optimization programs in rows 3-5 of Table 2.1, each of which containing a data fidelity and an image constraint (in addition to the image positivity constraint.) As discussed in Sec. 3.2.2, for each program studied, we performed a number of reconstructions with different values of its image-constraint parameter; and, based upon visual inspection of the reconstructions, we selected a final value and used it to yield convergent reconstructions, as shown in rows 2-4 of Fig. 3.3, according to the practical convergent conditions discussed in 3.2.2. Result in row 2 reveals that reconstructions with image-ℓ_1 constraint retain a noisy, sprinkled-salt texture. This is not surprising because image-ℓ_1 constraint encourages sparse solutions, while the Rando-phantom image is substantially non-sparse. Conversely, as the image-ℓ_2 constraint promotes non-sparse solutions, optimization programs with image-ℓ_2 constraint thus lead to reconstructions of the Rando phantom with visually reasonable appearance, as shown in row 3 of Fig. 3.3. On the other hand, reconstructions with an image-TV constraint in row 4 of Fig. 3.3 appear comparable to those obtained with image-ℓ_2, but with bony structures of enhanced spatial resolution. For the cases considered, image constraints appear to impact reconstruction textures more significantly than do data fidelities, as reconstructions vary visually more significantly within each column than within each row. Observations similar to those for reconstructions of the Rando phantom can also be made for reconstructions of the Catphan phantom.

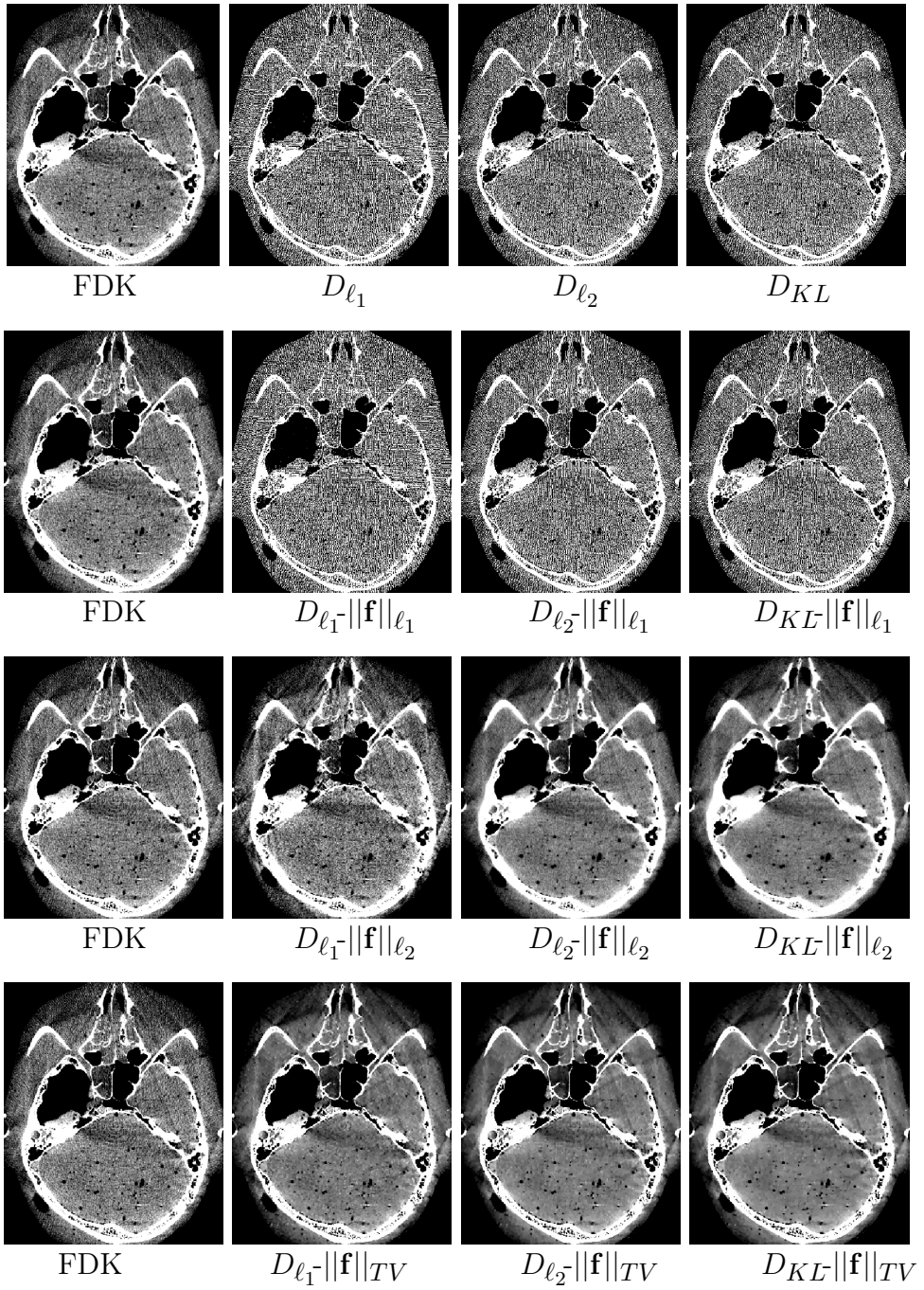


Figure 3.3: Reconstructions within the middle plane of the Rando phantom obtained by use of the FDK algorithm, and by use of the CP algorithm solving, respectively, optimization programs shown in Table 2.1. Display window: [0.22, 0.30] cm^{-1}.

3.3.2 Optimization-based reconstructions and artifact reduction

We investigate below optimization-based reconstructions for possible reduction of artifacts within non-middle planes in FDK reconstructions.

Reconstructions As Fig. 3.3 shows, optimization programs with image-TV constraint appear to yield reconstructions with visually reasonable spatial and contrast resolution as well as image texture. Therefore, for brevity, without loss of generality, we present below results of reconstructions specified by optimization programs with image-TV constraint in row 5 of Table 2.1 (i.e., Eqs. (2.9)-(2.11)). In the study, using the strategy described in Sec. 3.2.2, we determined image parameter t_1 for the Rando data, and display in Fig. 3.4 convergent reconstructions, accompanied by the corresponding FDK reconstructions. For revealing reconstruction details, zoomed-view images within the ROIs defined in Fig. 3.2 are depicted immediately below each corresponding reconstruction. Also, in Fig. 3.5, we display convergent reconstructions of the Catphan phantom obtained by solving the programs with image-TV constraint in row 5 of Table 2.1. Similar results have also been obtained for reconstructions within other transverse, coronal, and sagittal slices of the two phantoms, which are not shown for brevity.

Artifact reduction Results in Figs. 3.4 and 3.5 indicate that reconstructions specified by the programs with image-TV constraint in row 5 of Table 2.1 appear to suppress effectively artifacts observed in FDK reconstructions. In particular, as a result of artifact reduction, low-contrast structures, highlighted by arrows, within the Rando phantom can now be observed in Fig. 3.4. Some residual artifacts remain visible in the reconstructions due to data insufficiency and inconsistency, such as beam-hardening and scatter, which are not included in the imaging model. It is also interesting to notice that image texture obtained with the programs appear differently, because they respond differently to data inconsistency and insufficiency. Reconstructions with image-TV constraint seem to be robust relative to different

data fidelities in terms of compensating for the artifacts. Again, remarks similar to those for the Rando reconstructions above can be made for the Catphan reconstructions. Furthermore, inspection of Figs. 3.4 and 3.5 indicates that the performance of the optimization programs considered appears to be comparably effective for objects such as the Rando and Catphan phantoms with distinct anatomies.

Reconstruction evolution as a function of iterations The discussion above focuses on convergent reconstructions satisfying the practical convergent conditions, obtained at about 1090 iterations for the Rando phantom, and about 950 iterations for the Catphan phantom, respectively. For demonstrating the reconstruction convergence, we show in Fig. 3.6 convergence metrics, including data-ℓ_2, image-TV, and cPD, as functions of iterations for the Rando reconstruction in solving program D_{KL}-$||\mathbf{f}||_{TV}$ in Eq. (2.10). It can be observed that the convergence metrics plateau considerably before 1090 iterations. It is possible, however, that artifacts are compensated effectively for in reconstructions at earlier iterations. In Fig. 3.7, we display reconstructions of the Rando phantom at iterations 30, 50, and 150, and of the Catphan phantom at iterations 10, 30, and 100, as well as the corresponding convergent reconstructions, i.e., the final reconstructions, shown in column 4. It can be observed that Rando and Catphan reconstructions at, e.g., about iterations 150 already resemble visually their respective convergent reconstructions in terms of effective reduction of artifacts in FDK reconstructions.

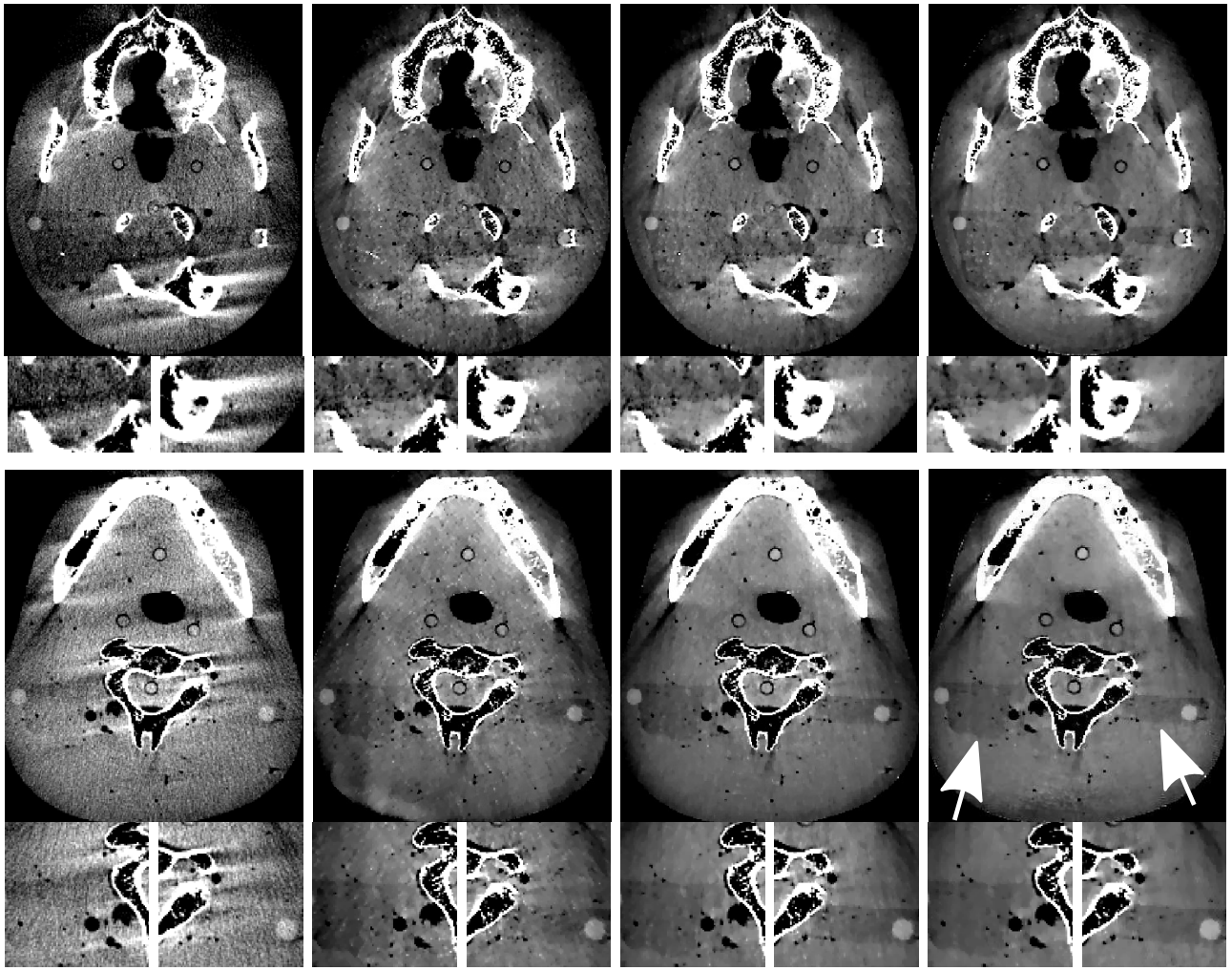


Figure 3.4: Reconstructions, together with zoomed-in images within ROIs defined in Fig. 3.2, of the Rando phantom within transverse slices at 3.37 cm (row 1) and 7.32 cm (row 2) from the middle plane, reconstructed by use of the FDK algorithm and the CP algorithm solving, respectively, optimization programs in row 5 of Table 2.1. Arrows in the reconstruction in panel 4 of row 2 highlight the low-contrast structures in the Rando phantom. Display window: [0.22, 0.30] cm^{-1}.

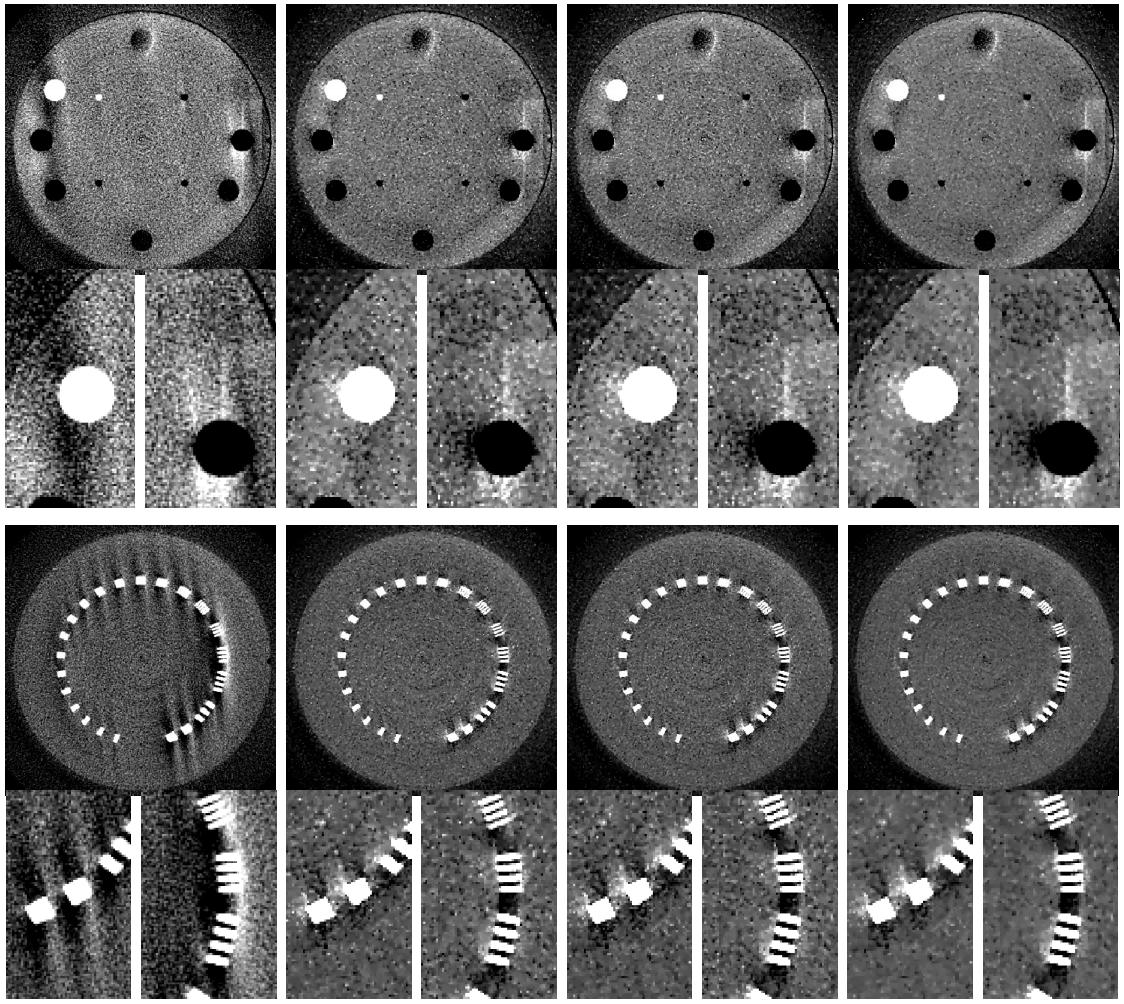


Figure 3.5: Reconstructions, together with zoomed-in images within ROIs defined in Fig. 3.2, of the Catphan phantom within transverse slices at 4.15 cm (row 1) and 6.00 cm (row 2) from the middle plane, reconstructed by use of the FDK algorithm and the CP algorithm solving, respectively, optimization programs in row 5 of Table 2.1. Display window: [0.22, 0.32] cm^{-1} for row 1 and [0.22, 0.35] cm^{-1} for row 2.

3.3.3 Additional study results

Reconstructions with image-ℓ_1 and image-ℓ_2 constraints We have investigated artifact reduction with the optimization programs in Table 2.1, and show results above obtained from the optimization programs with image-TV constraint. In the discussion below, we present reconstructions obtained from optimization programs with image-ℓ_1 and image-ℓ_2 constraints instead, i.e.,

$$\mathbf{f}^{\star} = \underset{\mathbf{f}}{\mathsf{argmin}} D_{\ell_2}(\mathbf{f}) \quad \text{s.t.} \quad ||\mathbf{f}||_{\ell_1} \leq l_1 \quad \text{and} \quad f_j \geq 0, \tag{3.1}$$

and

$$\mathbf{f}^{\star} = \underset{\mathbf{f}}{\mathsf{argmin}} D_{\ell_2}(\mathbf{f}) \quad \text{s.t.} \quad ||\mathbf{f}||_{\ell_2} \leq l_2 \quad \text{and} \quad f_j \geq 0, \tag{3.2}$$

which are denoted as D_{ℓ_2}-$||\mathbf{f}||_{\ell_1}$ and D_{ℓ_2}-$||\mathbf{f}||_{\ell_2}$, respectively, in Table 2.1. For each study, image-constraint parameters ℓ_1 and ℓ_2 were determined as described in Sec. 3.2.2. Using the CP algorithm to solve the programs, we reconstruct images from the Rando and Catphan data, and display them in Fig. 3.8. Reconstructions in Fig. 3.8 demonstrate that optimization programs with image-ℓ_1 and image-ℓ_2 constraints appear to be considerably less effective than the optimization programs with image-TV constraint in reducing the FDK-reconstruction artifacts. Moreover, as expected, reconstructions with an image-ℓ_1 constraint appear with noisy, sprinkled-salt textures because both Rando and Catphan are highly non-sparse. It is also interesting to notice that reconstructions based upon program D_{ℓ_2}-$||\mathbf{f}||_{\ell_2}$ in Eq. (3.1) have an appearance resembling that of the corresponding FDK reconstructions. This is because program D_{ℓ_2}-$||\mathbf{f}||_{\ell_2}$ without image-positivity constraint is equivalent to the Tikhonov-Phillips method, which has been shown to be an analog of the filtered-back projection (i.e., the FDK) algorithm [172, 173].

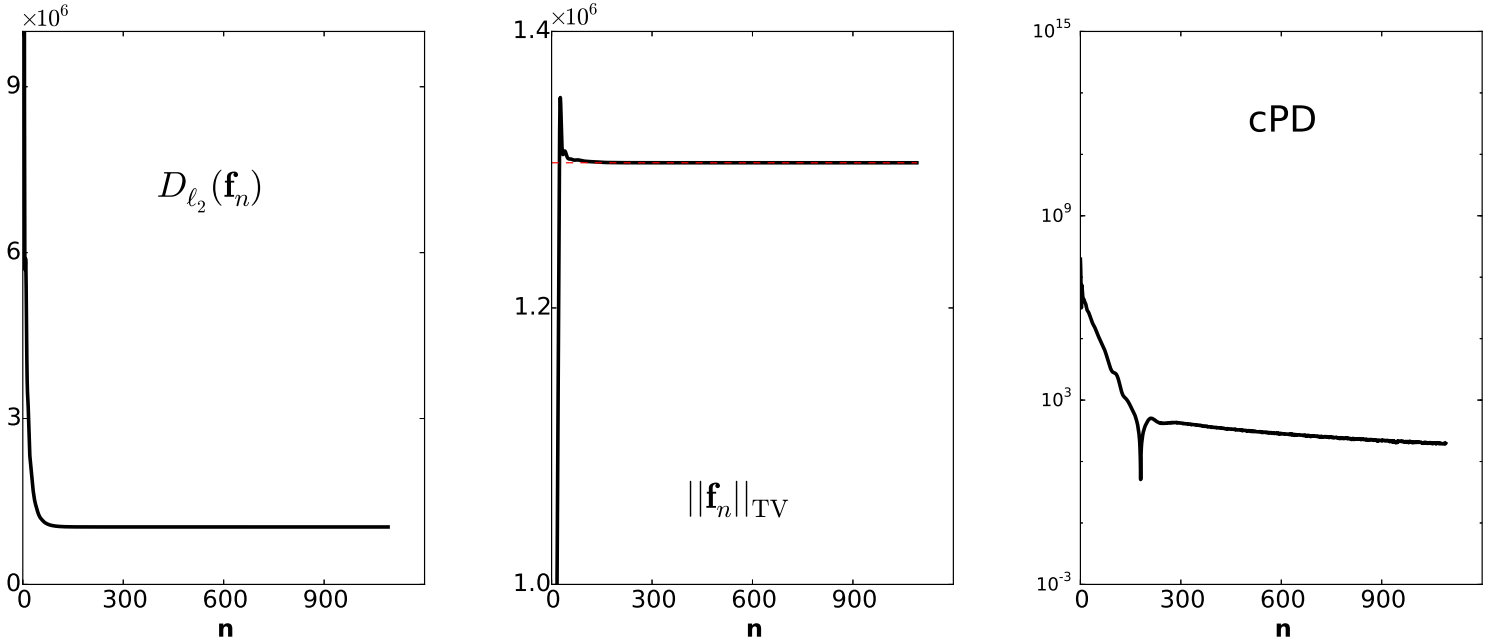


Figure 3.6: Data-ℓ_2 (left), image-TV (middle), and conditional primal-dual gap (cPD) (right), as functions of iteration number n for Rando-phantom reconstruction by use of the CP algorithm solving program D_{ℓ_2}-$|\mathbf{f}|_{TV}$ in Eq. (2.10). Dashed line in image-TV plot (middle) is specified by image-constraint parameter t_1 determined. The plot scales are in arbitrary units.

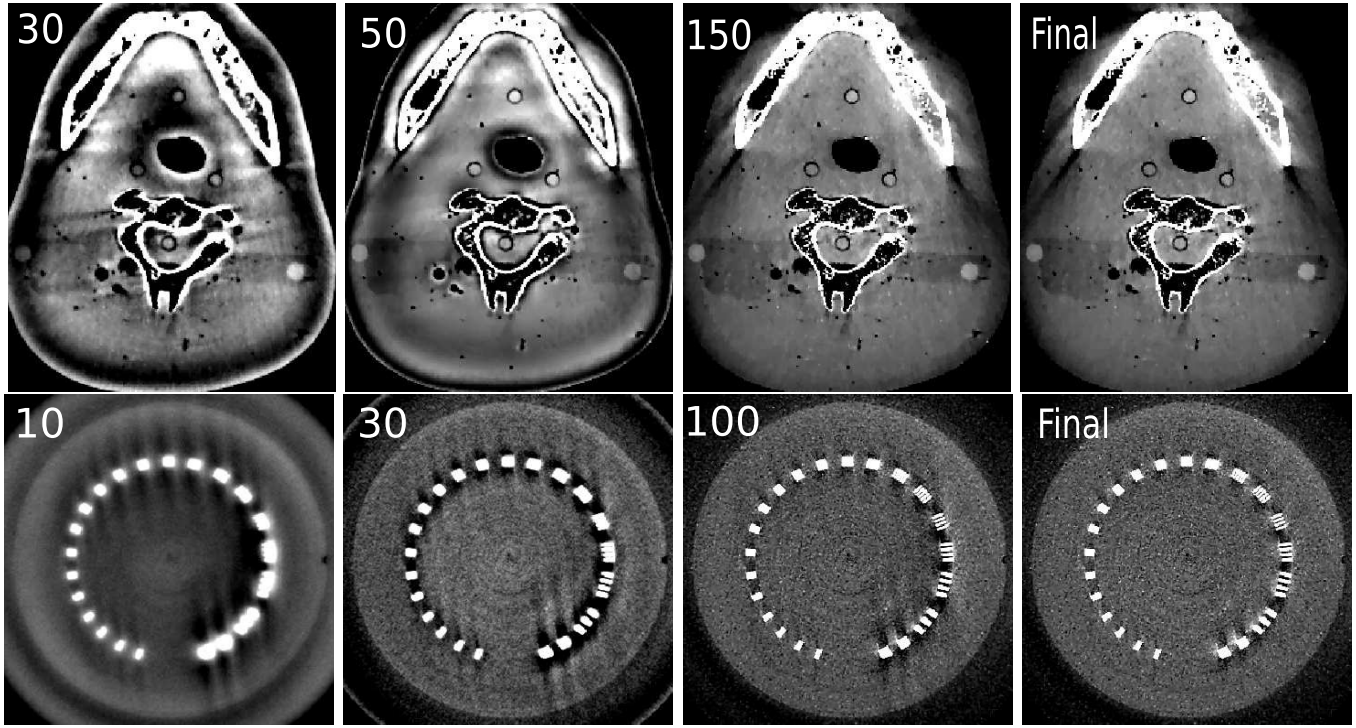


Figure 3.7: Reconstructions of the Rando phantom (row 1) within transverse slice at 7.32 cm (row 1), and of Catphan (row 2) phantom within transverse slice at 6.0 cm, from their respective middle planes at iteration numbers indicated, along with their respective, convergent reconstructions (column 4), obtained with the CP algorithm solving program D_{ℓ_2}-$|\mathbf{f}|_{TV}$ in Eq. (2.10). Display windows [0.22, 0.30] cm^{-1} and [0.22, 0.35] cm^{-1} are for Rando and Catphan images, respectively.

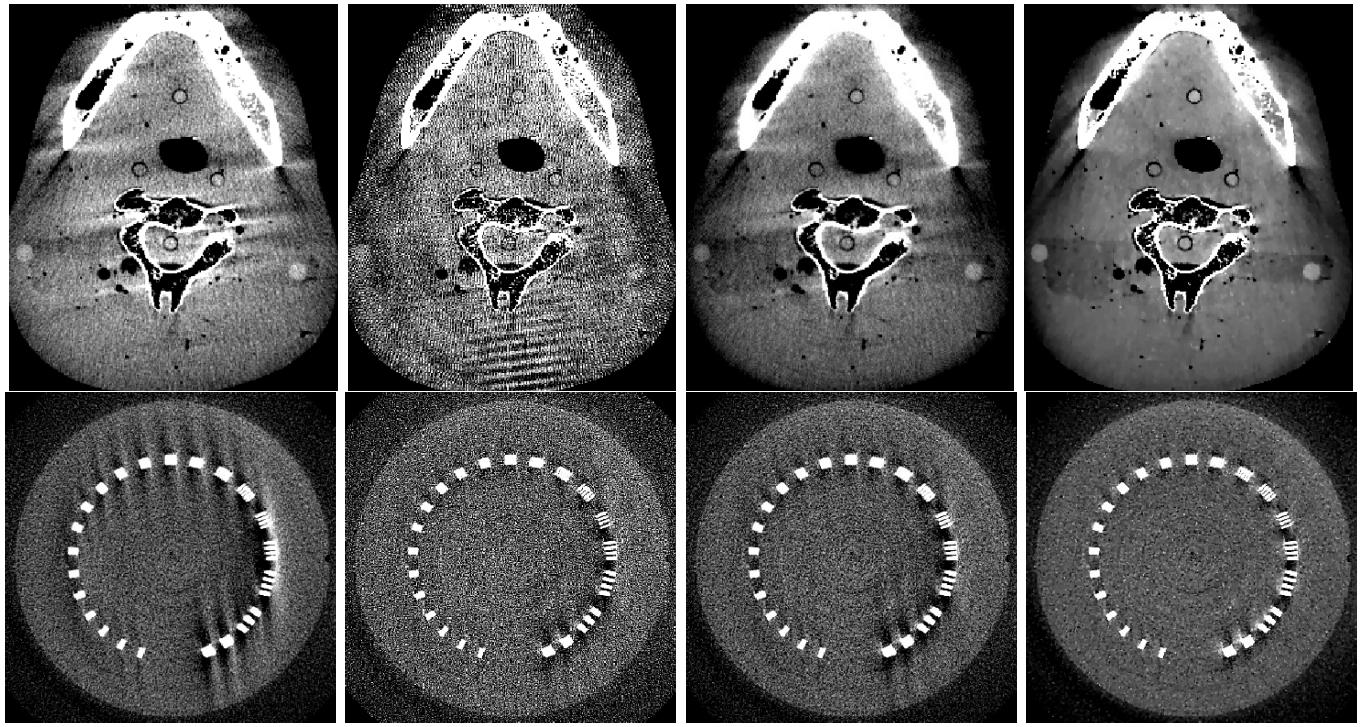


Figure 3.8: Reconstructions of the Rando phantom (row 1) within transverse slice at 7.32 cm, and of the Catphan phantom (row 2) within transverse slice at 6.0 cm, from their respective middle planes, reconstructed by use of the FDK algorithm, and by use of the CP algorithm solving optimization programs in column 3 of Table 2.1, respectively. Display windows [0.22, 0.30] cm^{-1} and [0.22, 0.35] cm^{-1} are for the Rando and Catphan images, respectively.

Effect of different weighting matrices on image reconstruction In analytic, such as FDK, reconstruction, a weighting function used for designing a weighting matrix is required to satisfy the normalization and continuity conditions in Eqs. (2.12) and (2.13) for avoiding significant numerical artifacts. We have investigated weighting functions with different forms in optimization reconstructions. Using a weighting matrix formed from a weighting function

$$\begin{cases} w(u,v,\phi) = \frac{1}{2} & 0 \leq \phi < 2\gamma_m - 2\gamma \\ w(u,v,\phi) = 1 & 2\gamma_m - 2\gamma \leq \phi < \pi - 2\gamma \\ w(u,v,\phi) = \frac{1}{2} & \pi - 2\gamma \leq \phi < \pi + 2\gamma_m \\ w(u,v,\phi) = 0 & \pi + 2\gamma_m \leq \phi < 2\pi, \end{cases} \tag{3.3}$$

which satisfies the normalization condition in Eq. (2.12), but not the continuity condition in Eq. (2.13), we carried out reconstruction based upon the optimization programs considered, and display example reconstructions in Fig. 3.9 obtained by solving program $D_{\ell_2^-}|\mathbf{f}|_{TV}$ in Eq. (2.10) from the Rando and Catphan data. As shown in row 1 of Fig. 3.9, optimization-based reconstructions are virtually free of the prominent numerical artifacts seen in FDK reconstructions, resulted from the weighting-function discontinuity in ϕ. Furthermore, we studied reconstructions using an identity matrix, corresponding to a uniform weighting function without satisfying Eq. (2.12), and display them in row 2 of Fig. 3.9. Again, optimization-based reconstructions appear free of the numerical artifacts seen in the FDK reconstruction, as a consequence of the uniform weighting function not satisfying the normalization condition. The results of the investigation indicate weighting matrices designed by use of weighting functions satisfying, or not satisfying, Eqs. (2.12) and/or (2.13) appear to have little impact on the performance of optimization-based reconstructions.

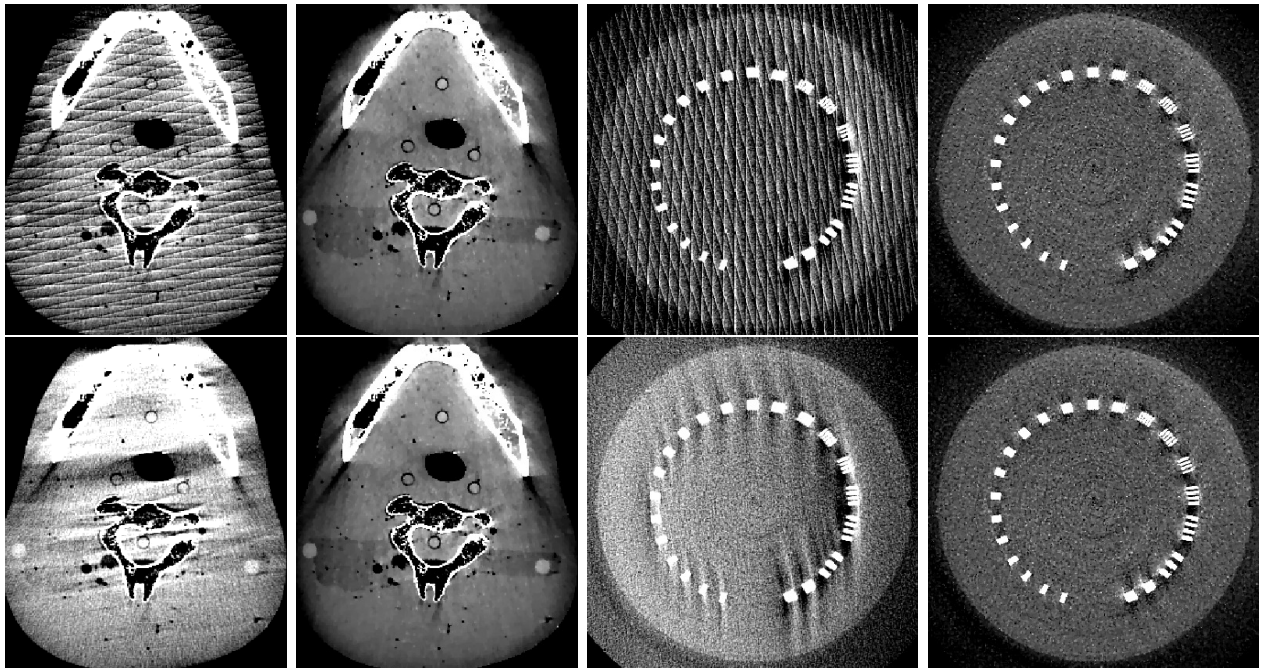


Figure 3.9: Images of the Rando phantom within transverse slice at 7.32 cm, and of the Catphan phantom within transverse slice at 6.0 cm, from their respective middle planes, reconstructed by use of the FDK algorithm (column 1 & 3), and by use of the CP algorithm (column 2 & 4) solving optimization program D_{ℓ_2}-$||\mathbf{f}||_{TV}$ in Eq. (2.10) with weighting matrices calculated from a non-smooth weighting function in Eq. (3.3) (row 1) and a smooth, but uniform, weighting function (row 2), respectively. Display windows [0.22, 0.35] cm^{-1} and [0.22, 0.30] cm^{-1} are for the Rando and Catphan images, respectively.

3.4 Discussion

In this work, we have investigated optimization-based reconstructions by using the CP algorithm for possible reduction of image artifacts often observed in FDK reconstruction from short-scan CBCT data. In an attempt to appreciate how different designs of optimization-based reconstruction may impact the artifact reduction, numerous convex optimization programs of potential research and application interest were considered. The focus of the study is not on quantitative assessment of an optimization-based reconstruction in a specific, clinically or practically, relevant task; it is tailored, instead, to demonstrating whether optimization-based reconstructions can be designed for reducing artifacts observed in FDK reconstruction. The study confirms that optimization-based reconstruction with appropriate image constraints such as image-TV constraint can reduce FDK-reconstruction artifacts in CBCT with a short-scan configuration. As the results demonstrated, a direct benefit of such an artifact reduction can be an improved contrast level of low-contrast anatomic structures that are otherwise obscured by the artifacts.

The performance of optimization-based reconstructions in a specific imaging task is dependent upon the forms and parameters specifying the optimization programs, as the study results demonstrated. In terms of the reduction of artifacts observed in FDK reconstruction from short-scan CBCT data, visual inspection of the study results reveals that reconstructions based upon image-TV-constrained data-discrepancy minimization appear to be capable of reducing the artifacts. Furthermore, although the selection of different weighting matrices considered in this work has a significant impact on FDK reconstruction, it appears to have insignificant effect on optimization-based reconstructions. Using two phantoms of IGRT CBCT relevance, we show that the image-TV-constrained reconstruction remains effective in reduction of off-middle-plane artifacts in FDK reconstructions from short-scan CBCT data for objects with substantially distinct anatomies.

We considered in the study reconstruction only from short-scan CBCT data. However, the work has implications for reconstruction from data acquired with other scanning con-

figurations in CBCT. It is not uncommon that a scan with an angular range between those of short-scan and full-scan configurations is used in practical applications. The investigation and analysis approach in the work can readily be tailored to investigating appropriate optimization-based reconstruction for possible reduction of artifacts that may appear in the corresponding FDK reconstruction. The study method and results can also be adopted to investigate optimization-based reconstruction from data collected with an offset detector in CBCT.

The evaluation of the study results was based largely on visual inspection, which serves adequately the focus of the study, which is on whether optimization-based reconstruction can reduce some of the prominent artifacts observed in FDK reconstructions from short-scan CBCT data. However, as the results reveal, different optimization-based reconstructions may behave distinctly differently, depending upon numerous factors, including their design parameters, data conditions, and object properties. More importantly, their performance is likely to be task-specific and thus should be measured with metrics of task-performance utility. Quantitative study of task-specific performance of optimization-based reconstructions can be a topic of theoretical and practical interest for future research.

CHAPTER 4

OPTIMIZATION-BASED ALGORITHMS WITH AN IMAGE-TOTAL-VARIATION CONSTRAINT FOR DIGITAL PET

4.1 Introduction

In Chapters 2 and 3, we have investigated numerous designs of optimization problems and their applications to flat-panel CBCT imaging applications. The studies have demonstrated that appropriately designed optimization-based algorithms may have the potential to improve image quality in CBCT image reconstruction, and possess the flexibility to accommodate many non-conventional CBCT imaging configurations, including some employing sparse data sampling. In particular, extensive efforts have been devoted to exploiting image-sparsity properties by employing an image-TV regularizer, with the intention of reducing data sampling. Recently, there exists much interest in designing an innovative PET scanner with sparsely populated crystals. This is because detector materials constitute a significant cost portion of a clinical PET scanner, and reducing the number of detectors could significantly reduce cost. However, using reduced number of detectors results in reduced measurements, which may be challenging in image reconstruction for current clinically used algorithms. It is therefore an interesting idea to leverage the experience in CBCT image reconstruction from sparse data, and develop advanced algorithms for enabling the design of innovative PET systems with sparsely populated detectors, while not significantly compromising the PET capability and image quality.

Image reconstruction is a key component in the development and application of advanced PET imaging. A great body of work exists on PET-image reconstruction algorithms of theoretical significance and practical value, including those developed based upon expectation-maximization (EM), row-action maximum-likelihood (RAMLA), maximum a

posteriori (MAP), and penalized maximum-likelihood (PML) algorithms [111–114, 134–141]. The algorithms have also been extended to list-mode [119–122], time-of-flight (TOF) [174–177], and 4D spatial-temporal/parametric image reconstructions [123–133]. Strong interests remain, however, in research on reconstruction algorithm development for further improvement on image quality in PET imaging, and for enabling, together with empowering detection and electronic technologies, innovative PET systems tailored possibly to enhance existing, and facilitate new, applications [178–185]. In this work, we investigate tailoring optimization-based reconstructions with an image-total-variation (TV) constraint, which has been demonstrated to have potential utility for CT imaging in recent years [153, 154, 156, 165, 167, 168], for PET scans with standard and non-standard scan configurations.

It is worth noting that, there exist differences in CT and PET imaging: *1*) the transition between different uptake regions in the PET activity map is expected to be smoother than that in the CT attenuation map, and *2*) PET data often contain more significant noise than conventional CT data, which may become an issue when implementing TV minimization for image reconstruction. Therefore, additional regularization other than TV are needed for image reconstruction from PET data. Recently, *Wolf et al.*, proposed an algorithm intended for few-view single photon emission computed tomography (SPECT) reconstruction, where the SPECT object is modeled as piecewise constant subject to a Gaussian blurring operation. Results of that work suggest that by appropriately modeling the blurring parameter, further sampling reduction may be enabled, and artifacts such as streak and noise may be reduced. In this work, we will design and investigate image-TV constrained optimization programs containing such a Gaussian blurring operation. In addition, we will formulate different designs of the optimization programs with different data divergences and derive the corresponding CP algorithms for the PET reconstruction.

In Sec. 4.2, we introduce the PET imaging model, design the optimization programs with different data fidelities, and derive the corresponding CP algorithms. In Sec. 4.3, we carry out numerical studies to verify and demonstrate the capability of the optimization-based

algorithm in achieving the designed solution determined by the optimization programs, and the high accuracy of this designed solution compared to the ground truth image. Remarks of the study are given in Sec. 4.4.

4.2 Methods

We describe briefly in this section the PET configurations in a digital prototype PET/CT system and optimization-based reconstructions considered in the work.

4.2.1 PET configurations

As shown in Fig. 4.1, a full-scan configuration is formed by tightly assembling P identical detector modules, each of which itself is a flat panel containing $L \times K$ identical tiles of square shape, on a cylindrical surface of radius R. In particular, the centers of the modules are placed on the same circular ring on the cylindrical surface, while the horizontal central lines of the modules are parallel to the central axis of a cylindrical surface. Therefore, the full-scan configuration consists effectively of L rings of $P \times K$ tiles. Each of the tightly congregated tiles itself within a module is composed of $J \times J$ identical crystal bins of square shape and length d. A straight line connecting the centers of two crystal detector-bins forms a line-of-response (LOR) along which a data sample is collected. A PET configuration with M distinct LORs thus yields a data set of at most M samples. In the work, this configuration is referred to as the full-scan configuration, and data collected with the configuration the *full data*.

In the Monte-Carlo-simulation and real-data studies performed below, PET data were collected by use of a PET configuration, in a digital prototype PET/CT system, identical to that described above, with $P = 18$, $L = 5$, $K = 4$, $J = 8$, and $d = 4$ mm, yielding a total of $M = 153,446,400$ distinct LORs. We also investigated image reconstruction from data collected from a sparse-scan configuration that is formed by removing odd- or

even-numbered tiles in odd- or even-numbered rings in the full-scan configuration. The sparse-scan configuration thus consists of L rings each of which is composed only of $P/2 \times K$ uniformly, but sparsely, distributed tiles, as depicted in Fig. 4.1, and we refer to data collected with the sparse-scan configuration as *sparse data*. It can be shown that the sparse-scan configuration has a total of $M = 38,361,600$ distinct LORs. The investigation of the sparse-scan configuration is motivated by an interest in PET imaging with a considerably reduced number of crystals and associated electronics (relative to the full-scan configuration.)

4.2.2 PET-data model

In this work, we consider a discrete image defined on a three-dimensional (3D) array containing $N = N_x \times N_y \times N_z$ identical voxels of cubic shape, where N_x, N_y, and N_z denote the numbers of voxels along the Cartesian x-, y-, and z-axis, respectively. The z-axis coincides with the central axis of the PET configuration shown in Fig. 4.1. We use vector $\mathbf{u}$ of size N to denote the image in a concatenated form. The PET-data model used in the study is given by

$$\mathbf{g} = \mathcal{H}\mathbf{u} + \mathbf{g}_s + \mathbf{g}_r, \tag{4.1}$$

where vector $\mathbf{g}$ of size M denotes the model data, $\mathcal{H}$ is an $M \times N$ system matrix in which element h_{ij} is the intersection length of LOR i with voxel j, vectors $\mathbf{g}_s$ and $\mathbf{g}_r$ of size M denote scatter and random events, which are assumed to be known in the work. We use vector $\mathbf{g}_m$ of size M to denote the measured data. In this work, $\mathbf{g}_m$, $\mathbf{g}_s$, and $\mathbf{g}_r$ are corrected for the effect of photon attenuation. The goal of PET-image reconstruction is to determine (i.e., reconstruct) $\mathbf{u}$ from knowledge of $\mathbf{g}_m$, $\mathcal{H}$, $\mathbf{g}_s$, and $\mathbf{g}_r$.

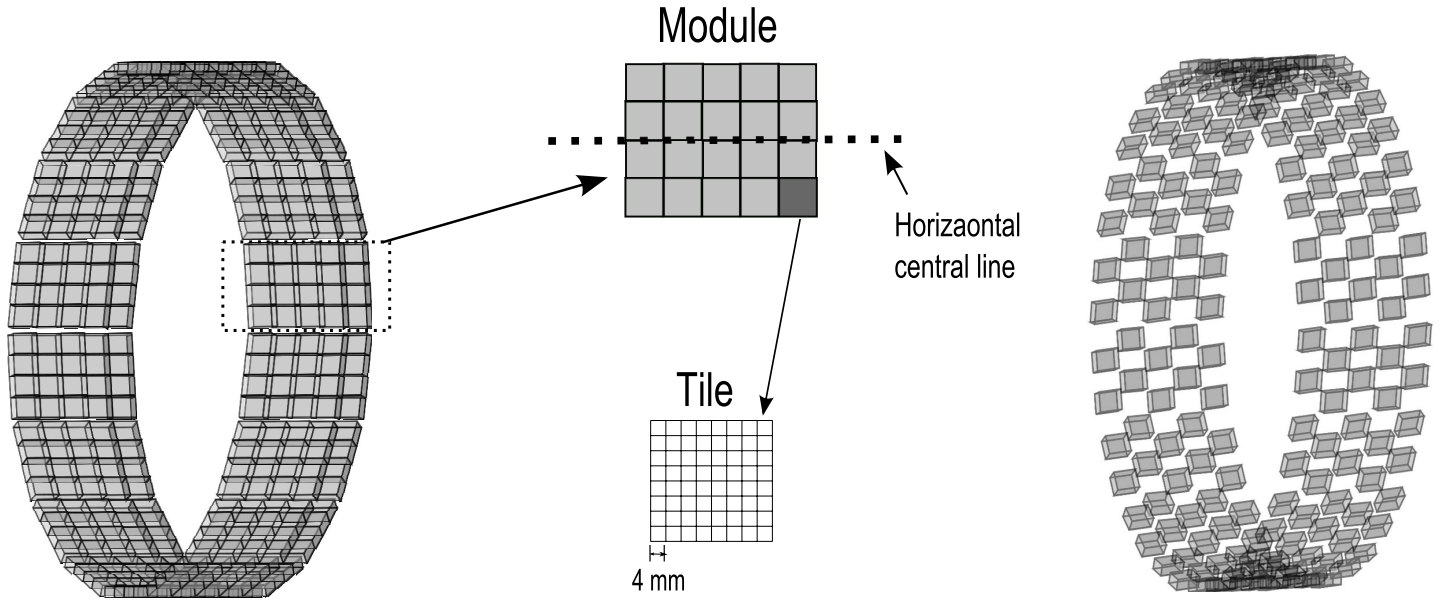


Figure 4.1: Full- and sparse-scan PET configurations considered. In the full-scan configuration, 18 modules, each of which consists of 5×4 identical tiles of square shape, are assembled tightly on a cylindrical surface, forming effectively 5 full rings of tiles. Each tile itself is composed of 8×8 identical, square-shaped crystal bins of size 4 mm. The sparse-scan configuration is obtained by the removal of odd- or even-numbered tiles in an odd- or even-numbered ring in the full-scan configuration.

4.2.3 Optimization programs

Using Eq. (4.1), we form an optimization program in the form of

$$\mathbf{u}^{\star} = \underset{\mathbf{u}}{\mathsf{argmin}} D(\mathbf{g}_m, \mathbf{g}) \quad \text{s.t.} \quad ||\mathbf{f}||_{TV} \leq t_1 \quad \text{and} \quad f_j \geq 0, \tag{4.2}$$

where $D(\mathbf{g}_m, \mathbf{g})$ denotes a measure of data fidelity between measured data $\mathbf{g}_m$ and model data $\mathbf{g}$, image $\mathbf{u}$ to be reconstructed is related to latent image vector $\mathbf{f}$ of size N through [186]

$$\mathbf{u} = \mathcal{G}\mathbf{f}, \tag{4.3}$$

$\mathcal{G}$ a matrix of size $N \times N$, $||\mathbf{f}||_{TV}$ image TV (see Eq. (2) in Ref. [165]), t_1 TV constraint parameter, f_j the jth element of vector $\mathbf{f}$, and $j = 1, 2, ..., N$.

The program in the form of Eq. (4.2) has been investigated for CT imaging in which $\mathcal{G}$ is generally an identity matrix. As the selection of data divergence $D(\mathbf{g}_m, \mathbf{g})$ can significantly impact PET reconstructions, we consider below three different data divergences of potential research and practical interest. In PET imaging, data generally have a signal-to-noise ratio (SNR) substantially lower than that of typical CT data, and the gradient-magnitude image of a typical CT image of practical interest is generally sparser than that of a typical PET image (i.e., $\mathbf{u}$). Therefore, we formulate PET image $\mathbf{u}$ desired as latent image $\mathbf{f}$ blurred by matrix $\mathcal{G}$, and apply the image-TV constraint to $\mathbf{f}$ instead of $\mathbf{u}$. This formulation can potentially avoid yielding an image with significant patchy textures when data have a low level of SNR [167].

In the work, the blurring matrix is obtained as follows: for a 3D, isotropic Gaussian function centered at a given voxel in the image array, we calculate its values at the center locations of N voxels within the image array, and use the calculated values in a concatenated form identical to that of vector $\mathbf{u}$ to create a vector of size N. Repeating the calculation for each of the N voxels in the image array, N such vectors can be formed; and matrix $\mathcal{G}$ of size

$N \times N$ can subsequently be built in which a row is the transpose of one of the N vectors, and the N rows are in an order consistent with the concatenated order of entries of $\mathbf{u}$. The unit of standard deviation of the Gaussian function is defined in terms of voxel size. With standard deviation taking zero value, $\mathcal{G}$ becomes identity matrix $\mathcal{I}$.

(1) Optimization program DKL-fTV Using the data-Kullback-Leibler (KL) divergence to replace $D(\mathbf{g}_m, \mathbf{g})$ in Eq. (4.2), we obtain the program

$$\mathbf{u}^\star = \underset{\mathbf{u}}{\mathsf{argmin}} D_{KL}(\mathbf{g}_m, \mathbf{g}) \quad \text{s.t.} \quad ||\mathbf{f}||_{TV} \leq t_1 \quad \text{and} \quad f_j \geq 0, \tag{4.4}$$

where $D_{KL}(\mathbf{g}_m, \mathbf{g})$ denotes the data-KL divergence given by

$$D_{KL}(\mathbf{g}_m, \mathbf{g}) = \sum_{i}^{M} [\mathbf{g} - \mathbf{g}_m + \mathbf{g}_m \ln \mathbf{g}_m - \mathbf{g}_m \ln \mathbf{g}]_i, \tag{4a}$$

$[\cdot]_i$ denotes the ith element of vector $[\cdot]$, and $\ln\mathbf{g}$ a vector of size M in which entry i is $\ln g_i$, where g_i indicates the ith entry of vector $\mathbf{g}$. When computing $D_{KL}(\mathbf{g}_m, \mathbf{g})$ in this work, the entries in $\mathbf{g}$ that are smaller than 10^{-20} are replaced with 10^{-20}. We refer to the optimization program in Eq. (4.4) as *program* DKL-fTV.

(2) Optimization program DL2-fTV We also consider the program

$$\mathbf{u}^\star = \underset{\mathbf{u}}{\mathsf{argmin}} D_{\ell_2}(\mathbf{g}_m, \mathbf{g}) \quad \text{s.t.} \quad ||\mathbf{f}||_{TV} \leq t_1 \quad \text{and} \quad f_j \geq 0, \tag{4.5}$$

which is obtained by the replacement of $D(\mathbf{g}_m, \mathbf{g})$ in Eq. (4.2) with

$$D_{\ell_2}(\mathbf{g}_m, \mathbf{g}) = ||\mathbf{g}_m - \mathbf{g}||_2. \tag{5a}$$

This fidelity metric takes the ℓ_2-norm of the difference between the measured data and model data. We refer to the optimization program in Eq. (4.5) as *program* DL2-fTV.

(3) Optimization program DL1-fTV In addition to $D_{KL}(\mathbf{g}_m, \mathbf{g})$ and $D_{\ell_2}(\mathbf{g}_m, \mathbf{g})$ that have been used widely for formulating optimization programs in image reconstruction, we consider an optimization program below

$$\mathbf{u}^\star = \underset{\mathbf{u}}{\mathsf{argmin}} D_{\ell_1}(\mathbf{g}_m, \mathbf{g}) \quad \text{s.t.} \quad ||\mathbf{f}||_{TV} \leq t_1 \quad \text{and} \quad f_j \geq 0, \tag{4.6}$$

which is obtained by the replacement of $D(\mathbf{g}_m, \mathbf{g})$ in Eq. (4.2) with

$$D_{\ell_1}(\mathbf{g}_m, \mathbf{g}) = ||\mathbf{g}_m - \mathbf{g}||_1, \tag{6a}$$

denoting the ℓ_1-norm of difference between measured data and model data. We refer to the optimization program in Eq. (4.5) as *program* DL1-fTV.

4.2.4 Row-action maximum-likelihood algorithm (RAMLA)

Reconstruction techniques based upon the RAMLA algorithm [114] are used frequently in PET research and application for yielding reconstructions of practical utility. It can be viewed as a relaxed ordered-subset algorithm in which the step size is subset-independent and gradually decreases to zero. Under certain conditions, the RAMLA algorithm is mathematically equivalent to the EM algorithm [114], but they involve different implementation procedures and can lead to different solutions when a finite number of iterations are used as in all practical reconstructions.

In the work we applied the RAMLA algorithm to reconstructing from *full data* PET images $\mathbf{u}_{\text{ref}}$, and use them as *reference reconstructions* to gauge the relevance of the study considered. Specifically, the RAMLA implementation in the study consists of subsets with the number of LORs varying from 291 to 291×40, and yields the reconstruction after two full iterations, as is done typically in practical research and clinical applications. RAMLA reconstructions from sparse data were also carried out in each of the studies described below, and were observed of quality substantially lower than that of the reference reconstructions.

Therefore, they are not included in the work.

4.2.5 Derivation of CP algorithms

Optimization programs DKL-fTV, DL2-fTV, and DL1-fTV in Eqs. (4.4)-(4.6) are convex, and can be solved in the CP framework. The explicit forms of the CP algorithms for the programs considered can readily be derived by taking the approach in Ref. [164] and the Appendix in Chapter 2. In particular, we will take optimization program DKL-fTV as an example, formulate the primal and dual objectives in accordance to Eqs. (2.15) and (2.16), and derive its proximal-mapping. Derivation for other optimization programs can be conducted in the same manner.

Derivation for program DKL-fTV We modify Eq. (4.4) by introducing two constants, λ and ν,

$$\mathbf{u}^{\star} = \underset{\mathbf{u}}{\mathsf{argmin}} \lambda D_{KL}(\mathbf{g}_m, \mathbf{g}) \quad \text{s.t.} \quad ||\mathbf{f}||_{TV} \leq \nu t_1 \quad \text{and} \quad f_j \geq 0. \tag{4.7}$$

By introducing an indicator function and in accordance to Eq. (2.15), the above constrained optimization problem can be formed as:

$$\mathbf{u}^{\star} = \underset{\mathbf{f}}{\mathsf{argmin}} \left\{ \sum_{i}^{M} [\mathbf{g} - \mathbf{g}_m + \mathbf{g}_m \ln \mathbf{g}_m - \mathbf{g}_m \ln \mathbf{g}]_i + \delta_{\text{Diamond}(\nu t_1)}(|\nu \nabla \mathbf{f}|_{\text{MAG}}) + \delta_P(\mathbf{f}) \right\}. \tag{4.8}$$

where ∇ denotes a spatial gradient matrix of size $3N \times N$, and $|\nabla \mathbf{f}|_{\text{MAG}}$ depicts a vector of size N. The definition of ∇, $|\nabla \mathbf{f}|_{\text{MAG}}$, and indicator function δ_{Diamond} can be found in the Appendix of Chapter 2.

In order to match the form of primal minimization in Eq. (2.15), we introduce the following assignments:

$$\begin{cases} G(\mathbf{x}) = \delta_P(\mathbf{x}), \\ F(\mathbf{s}, \mathbf{z}) = F_1(\mathbf{s}) + F_2(\mathbf{z}), \\ F_1(\mathbf{s}) = \sum_i^M [\mathbf{s} + \mathbf{g}_r + \mathbf{g}_s - \mathbf{g}_m + \mathbf{g}_m \ln \mathbf{g}_m - \mathbf{g}_m \ln(\mathbf{s} + \mathbf{g}_r + \mathbf{g}_s)]_i, \\ F_2(\mathbf{z}) = \delta_{\text{Diamond}(\nu t_t)}(|\vec{\mathbf{z}}|_{\text{MAG}}), \end{cases} \tag{4.9}$$

and the symbols in Eq. (2.15) are defined as:

$$\begin{cases} \mathbf{x} = \mathbf{f}, \\ \mathbf{s} = \mathcal{H}\mathcal{G}\mathbf{f}, \\ \mathbf{z} = \nu \nabla \mathbf{f}, \\ \mathcal{K} = \begin{pmatrix} \mathcal{H}\mathcal{G} \\ \nu \nabla \end{pmatrix}, \end{cases} \tag{4.10}$$

where $\mathbf{z}$ is a vector of size $3N$. The convex conjugate function for functions F_2 and G can be found in the Appendix of Chapter 2. For function F_1, its convex conjugate functions can be written as:

$$\begin{aligned} F_1^*(\mathbf{p}) &= \max_{\mathbf{s}} \{\mathbf{p}^\top \mathbf{s} - F_1(\mathbf{s})\} \\ &= \max_{\mathbf{s}} \{\mathbf{p}^\top \mathbf{s} - \sum_i^M [\mathbf{s} + \mathbf{g}_r + \mathbf{g}_s - \mathbf{g}_m + \mathbf{g}_m \ln \mathbf{g}_m - \mathbf{g}_m \ln(\mathbf{s} + \mathbf{g}_r + \mathbf{g}_s)]_i\} \\ &= \sum_i [-(\mathbf{g}_r^\top + \mathbf{g}_s^\top)\mathbf{p} - \lambda \mathbf{g}_m \ln(\mathbf{1}_D - \frac{\mathbf{p}}{\lambda})]_i + \delta_P(\boldsymbol{\lambda}_D - \mathbf{p}), \end{aligned}$$

where $\mathbf{1}_D$ is a vector of size M with all entries set to be 1, $\boldsymbol{\lambda}_D$ is a vector of size M with all entries set to be λ, and all entries in vector $(\mathbf{1}_D - \frac{\mathbf{p}}{\lambda})$ of size M that are smaller than 10^{-20} are set to 10^{-20}.

Therefore the dual optimization of Eq. (4.8) can be written as

$$\begin{aligned}(\mathbf{p}^*, \mathbf{z}^*) &= \max_{\mathbf{p},\mathbf{z}}\{-F^*(\mathbf{p},\mathbf{z}) - G^*(\mathcal{K}^\top \begin{pmatrix} \mathbf{p} \\ \mathbf{z} \end{pmatrix})\} \\ &= \max_{\mathbf{p},\mathbf{z}}\{\sum_i [(\mathbf{g}_r^\top + \mathbf{g}_s^\top)\mathbf{p} + \lambda \mathbf{g}_m^\top \ln(\mathbf{1}_D - \frac{\mathbf{p}}{\lambda})]_i - \delta_P(\boldsymbol{\lambda}_D - \mathbf{p}) \\ &\quad -\nu t_1 ||(|\mathbf{z}|_{\text{MAG}})||_\infty - \delta_P(-\mathcal{G}^\top \mathcal{H}^\top \mathbf{p} - \nu \nabla^\top \mathbf{z})\}.\end{aligned}$$

In an attempt to implement Chambolle-Pock algorithm, we need to derive the proximal mapping for functions F_1^*, G, and F_2^*. The proximal-mappings for F_2^* and G can be found in Appendix in Chapter 2, and that for function F_1^* is:

$$\begin{aligned}\mathsf{prox}_\sigma[F_1^*](\mathbf{p}) &= \operatorname*{\mathsf{argmin}}_{\mathbf{p}'} \left\{ F_1^*(\mathbf{p}') + \frac{||\mathbf{p} - \mathbf{p}'||_2^2}{2\sigma} \right\} \\ &= \operatorname*{\mathsf{argmin}}_{\mathbf{p}'} \left\{ \sum_i [-\mathbf{p}^\top (\mathbf{g}_r + \mathbf{g}_s) - \lambda \mathbf{g}_m \ln(\mathbf{1} - \frac{\mathbf{p}}{\lambda})]_i + \delta_P(\boldsymbol{\lambda}_D - \mathbf{p}) + \frac{||\mathbf{p} - \mathbf{p}'||_2^2}{2\sigma} \right\} \\ &= \frac{1}{2}(\boldsymbol{\lambda}_D + \mathbf{p} + \sigma(\mathbf{g}_r + \mathbf{g}_s) - \sqrt{[\boldsymbol{\lambda}_D - \mathbf{p} - \sigma(\mathbf{g}_r + \mathbf{g}_s)]^2 + 4\sigma\lambda\mathbf{g}_m})\end{aligned}$$

Derivation for program DL2-fTV The proximal-mapping for other programs DL2-fTV and DL1-fTV can be derived in the manner. It is worth noting that, the only difference between those two program and program DKL-fTV is that they possess different F_1 function. For programs containing data-ℓ_2 fidelity, the function F_1 in Eq. (2.65) is:

$$F_1(\mathbf{s}) = \frac{1}{2}\lambda ||\mathbf{s} + \mathbf{g}_r + \mathbf{g}_s - \mathbf{g}_m||_2^2. \tag{4.11}$$

Computing the convex conjugate of Eq. (4.11) yields

$$\begin{aligned} F_1^*(\mathbf{p}) &= \max_{\mathbf{s}}\{\mathbf{p}^\top \mathbf{s} - F_1(\mathbf{s})\} \\ &= \max_{s}\{\mathbf{p}^\top \mathbf{s} - \frac{1}{2}\lambda||\mathbf{s}+\mathbf{g}_r+\mathbf{g}_s-\mathbf{g}_m||_2^2\} \\ &= \frac{1}{2\lambda}||\mathbf{p}||_2^2 + (\mathbf{g}_m^\top - \mathbf{g}_r^\top - \mathbf{g}_s^\top)\mathbf{p} \end{aligned}$$

The corresponding proximal-mapping of F_1^* is:

$$\begin{aligned} \mathsf{prox}_\sigma[F_1^*](\mathbf{p}) &= \operatorname*{argmin}_{\mathbf{p}'}\left\{F_1^*(\mathbf{p}') + \frac{||\mathbf{p}-\mathbf{p}'||_2^2}{2\sigma}\right\} \\ &= \operatorname*{argmin}_{\mathbf{p}'}\left\{\frac{1}{2\lambda}||\mathbf{p}'||_2^2 + \mathbf{p}'^\top(\mathbf{g}_m-\mathbf{g}_r-\mathbf{g}_s) + \frac{||\mathbf{p}-\mathbf{p}'||_2^2}{2\sigma}\right\} \\ &= (\mathbf{p}-\sigma(\mathbf{g}_m-\mathbf{g}_r-\mathbf{g}_s))/(1+\sigma/\lambda) \end{aligned}$$

Derivation for program DL1-fTV For programs containing data-ℓ_1 fidelity, the function F_1 in Eq. (2.65) is:

$$F_1(\mathbf{s}) = \lambda||\mathbf{s}+\mathbf{g}_r+\mathbf{g}_s-\mathbf{g}_m||_1. \tag{4.12}$$

Computing the convex conjugate of Eq. (4.12) yields

$$\begin{aligned} F_1^*(\mathbf{p}) &= \max_{\mathbf{s}}\{\mathbf{p}^\top \mathbf{s} - F_1(\mathbf{s})\} \\ &= \max_{\mathbf{s}}\{\mathbf{p}^\top \mathbf{s} - \lambda||\mathbf{s}+\mathbf{g}_r+\mathbf{g}_s-\mathbf{g}_m||_1\} \\ &= (\mathbf{g}_m^\top - \mathbf{g}_r^\top - \mathbf{g}_s^\top)\mathbf{p} + \delta_{\mathrm{Box}(\lambda)}(\mathbf{p}). \end{aligned}$$

The corresponding proximal-mapping of F_1^* is:

$$\begin{aligned}\mathbf{prox}_{\sigma}[F_1^*](\mathbf{p}) &= \underset{\mathbf{p}'}{\mathsf{argmin}} \left\{ F_1^*(\mathbf{p}') + \frac{||\mathbf{p}-\mathbf{p}'||_2^2}{2\sigma} \right\} \\ &= \underset{\mathbf{p}'}{\mathsf{argmin}} \left\{ \mathbf{p}^\top(\mathbf{g}_m - \mathbf{g}_r - \mathbf{g}_s) + \delta_{\mathrm{Box}(\lambda)}(\mathbf{y}) + \frac{||\mathbf{p}-\mathbf{p}'||_2^2}{2\sigma} \right\} \\ &= \frac{\mathbf{p} - \sigma(\mathbf{g}_m - \mathbf{g}_r - \mathbf{g}_s)}{\max(\mathbf{1}_D,\, |\mathbf{p} - \sigma\mathbf{g}_m|/\lambda)}.\end{aligned}$$

4.2.6 Pseudo codes of CP Reconstruction algorithms

The explicit forms of the CP algorithms for the programs considered can readily be written as:

Algorithm: Pseudo codes of the CP algorithm for solving Eqs. (4.4)-(4.6)

INPUT: $\mathbf{g}_m$, $\mathbf{g}_s$, and $\mathbf{g}_r$; $\mathcal{H}$ and $\mathcal{G}$; t_1; and λ

$\nu \leftarrow \nu_{\mathcal{H}}$, $L \leftarrow ||\mathcal{H}_\nabla||_{SV}$, $\tau \leftarrow 1/L$, $\sigma \leftarrow 1/L$

$\theta \leftarrow 1$, $n \leftarrow 0$

INITIALIZE: $\mathbf{f}_0$, $\mathbf{u}_0$, $\mathbf{p}_0$, and $\mathbf{q}_0$ to zero

$\bar{\mathbf{f}}_0 \leftarrow \mathbf{f}_0$

repeat

  $\mathbf{p}_{n+1} \leftarrow \mathbf{\Theta}(\mathbf{p}_n, \bar{\mathbf{g}}_n, \mathbf{g}_m)$

  $\boldsymbol{\varrho} \leftarrow \mathbf{q}_n + \sigma\nu\nabla\bar{\mathbf{f}}_n$

  $\mathbf{q}_{n+1} \leftarrow \boldsymbol{\varrho}(\mathbf{1}_I - \sigma\mathbf{s}/|\boldsymbol{\varrho}|_{\mathrm{MAG}})$

  $\mathbf{f}_{n+1} \leftarrow \mathsf{pos}(\mathbf{f}_n - \tau(\mathcal{G}^\top\mathcal{H}^\top\mathbf{p}_{n+1} + \nu\nabla^\top\mathbf{q}_{n+1}))$

  $\mathbf{u}_{n+1} \leftarrow \mathcal{G}\mathbf{f}_{n+1}$

  $\bar{\mathbf{f}}_{n+1} \leftarrow \mathbf{f}_{n+1} + \theta(\mathbf{f}_{n+1} - \mathbf{f}_n)$

  $n \leftarrow n + 1$

until the practical convergence conditions are satisfied

OUTPUT: image $\mathbf{u}_{\mathrm{conv}} \leftarrow \mathbf{u}_n$

For the convenience of discussing the operations in the pseudo codes, we consider two

vectors $\mathbf{a}$ and $\mathbf{b}$ of sizes kN and N, where $k \geq 1$ is a positive integer. Let a_j denote the jth entry of $\mathbf{a}$, $j = 1, 2, ..., kN$, and b_j the jth entry of $\mathbf{b}$, $j = 1, 2, ..., N$, respectively. We use vectors $\mathbf{ab}$ and $\mathbf{a/b}$ of sizes kN to denote the multiplication and division of $\mathbf{a}$ and $\mathbf{b}$, with entry j given by

$$(\mathbf{ab})_j = a_j b_{j_m}$$
$$(\mathbf{a/b})_j = a_j / b_{j_m},$$

where $j = 1, 2, ..., kN$, and $j_m = \text{mod}\,(j, N)$ indicates the remainder of j dividing by N, and $b_{j_m} \neq 0$. In particular, when $k = 1$, $\mathbf{a}$ and $\mathbf{b}$ are of the same size N, and vectors $\mathbf{ab}$ and $\mathbf{a/b}$ thus are of size N and depict element-wise multiplication and division.

In the pseudo codes, algorithm parameter λ is used for potentially improving the convergence rate of the CP algorithms. Parameter $\nu_{\mathcal{H}} = \frac{||\mathcal{H}\mathcal{G}||_{SV}}{||\nabla||_{SV}}$, where $||\cdot||_{SV}$ represents the largest singular value of a matrix [164]; ∇ denotes a spatial gradient matrix of size $3N \times N$, with its transpose given by $\nabla^\top = (\nabla_x^\top, \nabla_y^\top, \nabla_z^\top)$, in which the superscript "$\top$" indicates a transpose operation; matrices ∇_x, ∇_y, and ∇_z of size $N \times N$ represent the finite difference matrices along x-, y-, and z-axis, respectively, yielding vectors $\nabla_x \mathbf{f}$, $\nabla_y \mathbf{f}$, and $\nabla_z \mathbf{f}$ of size N, which are used to form vector $\nabla \mathbf{f}$ of size $3N$ in a concatenated form in the order of x, y, and z; matrix $\mathcal{H}_\nabla$ has a transpose $\mathcal{H}_\nabla^\top = (\mathcal{G}^\top \mathcal{H}^\top, \nabla^\top)$; vector $\bar{\mathbf{g}}_n$ of size M is given by $\bar{\mathbf{g}}_n = \mathcal{H}\mathcal{G}\bar{\mathbf{f}}_n + \mathbf{g}_s + \mathbf{g}_r$; vector $\mathbf{1}_I$ denotes a vector of size N with all entries set to 1; vectors $\mathbf{p}_n$ and $\boldsymbol{\Theta}$ are of size M, whereas vectors $\boldsymbol{\varrho}$ and $\mathbf{q}_n$ are of size $3N$; vector $\mathbf{s}$ of size N is given by

$$\mathbf{s} = \mathbf{ProjOnto}\ell_1\mathbf{Ball}_{\nu t_1}(|\boldsymbol{\varrho}|_{\text{MAG}}/\sigma), \tag{4.13}$$

where operator $\mathbf{ProjOnto}\ell_1\mathbf{Ball}_{\nu t_1}$ projects vector $|\boldsymbol{\varrho}|_{\text{MAG}}/\sigma$ onto the ℓ_1-ball of scale νt_1 [167], $|\boldsymbol{\varrho}|_{\text{MAG}}$ depicts a vector of size N with entry j given by $(|\boldsymbol{\varrho}|_{\text{MAG}})_j = \sqrt{\varrho_j^2 + \varrho_{j+N}^2 + \varrho_{j+2N}^2}$, and ϱ_j indicates the jth entry of vector $\boldsymbol{\varrho}$; operator $\mathsf{pos}(\mathbf{f})$ enforces $f_j = 0$, if $f_j < 0$, and keeps f_j unchanged if $f_j \geq 0$, where f_j represent the jth entry of vector $\mathbf{f}$ of size N.

We note that all of the lines in the pseudo codes are identical for the three optimization programs in Eqs. (4.4)-(4.6), except for vector $\boldsymbol{\Theta}$ of size M in line 6, which may vary depending upon the specific data divergence considered [164]. For the program in Eq. (4.4),

$$\begin{aligned}\boldsymbol{\Theta}(\mathbf{p}_n, \bar{\mathbf{g}}_n, \mathbf{g}_m) = \frac{1}{2}[\boldsymbol{\lambda}_D + \mathbf{p}_n + \sigma\bar{\mathbf{g}}_n] \\ - \sqrt{(\boldsymbol{\lambda}_D - \mathbf{p}_n - \sigma\bar{\mathbf{g}}_n)^2 + 4\sigma\lambda\mathbf{g}_m},\end{aligned} \tag{4.14}$$

where $\boldsymbol{\lambda}_D$ is a vector of size M with all entries set to λ, and $\sqrt{\mathbf{p}}$ takes entry-wise square roots of vector $\mathbf{p}$ of size M. For the optimization program in Eq. (4.5),

$$\boldsymbol{\Theta}(\mathbf{p}_n, \bar{\mathbf{g}}_n, \mathbf{g}_m) = [\mathbf{p}_n + \sigma(\bar{\mathbf{g}}_n - \mathbf{g}_m)]/(1 + \sigma/\lambda); \tag{4.15}$$

and for optimization program in Eq. (4.6),

$$\boldsymbol{\Theta}(\mathbf{p}_n, \bar{\mathbf{g}}_n, \mathbf{g}_m) = \frac{\mathbf{p}_n + \sigma(\bar{\mathbf{g}}_n - \mathbf{g}_m)}{\max(\mathbf{1}_D,\, |\mathbf{p}_n + \sigma(\bar{\mathbf{g}}_n - \mathbf{g}_m)|/\lambda)}, \tag{4.16}$$

where $\max(\cdot)$ is performed element-wise, $\mathbf{1}_D$ is a vector of size M with all entries set to 1, and $|\mathbf{p}|$ takes entry-wise absolute values of vector $\mathbf{p}$ of size M.

The iterative reconstruction stops when the appropriate practical convergence conditions defined in Secs. 4.3 or 5.2 below are satisfied. The reconstruction and its latent image at the iteration are denoted as $\mathbf{u}_{\text{conv}}$ and $\mathbf{f}_{\text{conv}}$.

cPD gap In the below, we list the forms of cPD for all programs considered.

Program D_{KL}-$\mathbf{f}_{TV}$

$$\text{cPD}(\mathbf{f}_n) = \lambda \sum_i [\mathbf{g}_n - \mathbf{g}_m + \mathcal{W}\mathbf{g}_m \ln(\mathcal{W}\mathbf{g}_m) - \mathcal{W}\mathbf{g}_m \ln(\mathcal{W}\mathcal{H}\mathbf{f}_n)]_i \tag{4.17}$$

$$- \left[- \sum_i [-(\mathbf{g}_r^\top + \mathbf{g}_s^\top)\mathbf{p} - \lambda \mathbf{g}_m \ln(\mathbf{1}_D - \frac{\mathbf{p}}{\lambda})]_i - \nu t_1 ||(|\mathbf{q}_n|_{\text{MAG}})||_\infty \right]. \tag{4.18}$$

Program D_{ℓ_2}-$\mathbf{f}_{TV}$

$$\begin{aligned} \text{cPD}(\mathbf{f}_n) &= \frac{1}{2}\lambda||\mathbf{g}_n - \mathbf{g}_m||_2^2 && (4.19) \\ &- \left[-\frac{1}{2\lambda}||\mathbf{p}||_2^2 - (\mathbf{g}_m^\top - \mathbf{g}_r^\top - \mathbf{g}_s^\top)\mathbf{p} - \nu t_1||(|\mathbf{q}_n|_{\text{MAG}})||_\infty \right]. && (4.20) \end{aligned}$$

Program D_{ℓ_1}-$\mathbf{f}_{TV}$

$$\begin{aligned} \text{cPD}(\mathbf{f}_n) &= \lambda||\bar{\mathbf{g}}_n - \mathbf{g}_m||_1 && (4.21) \\ &- \left[-(\mathbf{g}_m^\top - \mathbf{g}_r^\top - \mathbf{g}_s^\top)\mathbf{p} - \nu t_1||(|\mathbf{q}_n|_{\text{MAG}})||_\infty \right]. && (4.22) \end{aligned}$$

4.2.7 Reconstruction parameters

Each of the optimization programs in Eqs. (4.4)-(4.6) includes numerous parameters such as system matrix $\mathcal{H}$, image array (i.e., voxel shape and size), and TV-constraint parameter t_1, scatter and random components $\mathbf{g}_s$ and $\mathbf{g}_r$, and blurring matrix $\mathcal{G}$. Furthermore, a reconstruction depends upon the specific form of an optimization program. Therefore, the optimization program itself constitutes a program parameter. This is the reason that we study programs with different data divergences in Eqs. (4.4)-(4.6). We refer to these parameters as *program parameters* in the work, because their different choices specify different feasible solutions (i.e., reconstructions). In contrast, the CP algorithm derived also involves parameter λ, which is referred to the *algorithm parameter*. Unlike program parameters, algorithm parameters have no impact on the specification of feasible solutions. Instead, they can affect the algorithm's convergence path and rate. For the studies carried out in the work, $\lambda = 0.01$ was used because it has been shown in the study to yield a reasonable convergence rate. We discuss below the determination of the program parameters for the studies carried out in the work.

As mentioned above, system matrix $\mathcal{H}$ contains M row vectors of size N in which each entry depicts the intersection of an LOR with a voxel in the image array. In the verification

and Monte-Carlo-simulation studies in Secs. 4.3 and 5.2, scatter and random events $\mathbf{g}_s$ and $\mathbf{g}_r$ are not considered; whereas in the IEC and human studies performed in Secs. 5.3 and 5.4, the single-scatter simulation method [187] and the delayed coincidence method [188] were employed for estimating $\mathbf{g}_s$ and $\mathbf{g}_r$, respectively. A voxel size of 4 mm was selected for the studies because it is used often in clinical studies. Results of previously relevant [186] and current studies indicate that matrix $\mathcal{G}$ generated by use of a 3D isotropic Gaussian function with a standard deviation of 2.4 mm, which is 0.6 times the image-voxel size, appears to yield appropriate reconstructions for data conditions considered.

The CP algorithm solves mathematically the convex optimization programs in Eqs. (4.4)-(4.6), but like any other iterative algorithms, they only converge to a solution in the limit of infinite iterations. In a practical study, due to the limitation of computer precision and computation time, one can obtain reconstructions only at finite iterations. Therefore, it is necessary to specify practical convergence conditions under which reconstructions can be achieved within a finite number of iterations; and the practical convergence conditions thus play a role in defining an actual solution set achievable within a finite number of iterations. For designing the practical convergence conditions, we introduce two unitless metrics as

$$\overline{D}(\mathbf{u}_n) = D(\mathbf{g}_m, \mathbf{g}_n)/D(\mathbf{g}_m, \mathbf{g}_1) \tag{4.23}$$

$$\overline{\mathrm{TV}}(\mathbf{f}_n) = |\, \|\mathbf{f}_n\|_{TV} - t_1|/t_1, \tag{4.24}$$

where $\mathbf{u}_n$ and $\mathbf{f}_n$ denote reconstructions at iteration n, $\mathbf{g}_n = \mathcal{H}\mathbf{u}_n + \mathbf{g}_s + \mathbf{g}_r$ the model data estimated at the nth iteration, obtained by replacing $\mathbf{u}$ with $\mathbf{u}_n$ in Eq. (4.1); and $D(\mathbf{g}_m, \mathbf{g}_n)$ in $\overline{D}(\mathbf{u}_n)$ takes the form of $D_{KL}(\mathbf{g}_m, \mathbf{g}_n)$, $D_{\ell_2}(\mathbf{g}_m, \mathbf{g}_n)$, and $D_{\ell_1}(\mathbf{g}_m, \mathbf{g}_n)$, respectively, for programs DKL-fTV, DL2-fTV, and DL1-fTV. Practical convergence conditions are devised with $\overline{\mathrm{TV}}(\mathbf{f}_n)$ and $\overline{D}(\mathbf{u}_n)$ for the studies performed in Secs. 4.3-5.4 below.

With the program parameters and practical convergence conditions discussed, only image-TV-constraint parameter t_1 remains to be discussed. We illustrate the determination of t_1

in each specific study carried out below, because different data conditions in the studies can impact differently on the appropriate selection of t_1.

4.3 Algorithm Validation

We first performed a study to validate that under ideal, sufficient data conditions, the CP algorithms and their computer implementation numerically recover the truth image by solving the optimization programs. In the study, which is referred to as an inverse-crime study, a known truth image represented on a discrete array of identical voxels is used for generating model data by use of a known, system matrix, which is used also in reconstruction from the model data on the same image array. It should be pointed out that an inverse-crime study reveals no information about the practical utility of an optimization reconstruction in real-data studies. Such information can be obtained only through carefully designed evaluation studies of task-specific significance.

4.3.1 Study design

For the purpose of verification and computation-efficiency consideration, we use the full-scan configuration described in Sec. 4.2.1 but with only one ring of tiles, and a 3D image array consisting of $N = 50 \times 50 \times 8$ identical cubic voxels of size 4 mm. With the configuration and image array given, system matrix $\mathcal{H}$, and Gaussian matrix $\mathcal{G}$ can be formed, as described in Sec. 4.2.6 above. According to Eq. (4.3), we create truth image $\mathbf{u}_{\text{true}}$ shown in Fig. 4.2a by application of matrix $\mathcal{G}$ to a known, numerical IEC phantom $\mathbf{f}_{\text{true}}$, i.e., the truth latent image. Subsequently, model data $\mathbf{g}$ is generated from $\mathbf{u}_{\text{true}}$ by use of system matrix $\mathcal{H}$. Without loss of generality, random and scatter are not considered (i.e., $\mathbf{g}_s = 0$ and $\mathbf{g}_r = 0$) in the study.

In the inverse-crime study, the mathematical convergence conditions for the CP algorithms include $\overline{D}(\mathbf{u}_n) \to 0$, $\overline{\text{TV}}(\mathbf{f}_n) \to 0$, and $\overline{\text{cPD}}(\mathbf{u}_n) \to 0$, as $n \to \infty$, where $\overline{\text{cPD}}(\mathbf{u}_n) =$

$|\text{cPD}(\mathbf{u}_n)/\text{cPD}(\mathbf{u}_1)|$, and $\text{cPD}(\mathbf{u}_n)$ denotes the conditional primal-dual (cPD) gap [163,189]. The explicit form of a cPD depends upon the specific form of the corresponding optimization program. Using the approach in Ref. [163], we have derived the cPD as a function of n for each of the programs in Eqs. (4.4)-(4.6). The mathematical convergence conditions are unachievable in practical, numerical study, however, due to limited computer precision and computation time. Instead, for each of the programs Eqs. (4.4)-(4.6), we design the practical convergence conditions

$$\begin{aligned} \overline{D}(\mathbf{u}_n) &< 10^{-5}, \\ \overline{\text{TV}}(\mathbf{f}_n) &< 10^{-5}, \\ \overline{\text{cPD}}(\mathbf{u}_n) &< 10^{-5}, \end{aligned} \tag{4.25}$$

in the inverse-crime study. Obviously, practical convergence conditions tighter or looser than those in Eq. (4.25) can readily be designed, depending upon the amount of computation resources to be invested for the study.

4.3.2 Study results

We have performed inverse-crime studies on reconstructions based upon the three optimization programs in Eqs. (4.4)-(4.6) in which $t_1 = \| \mathbf{f}_{\text{true}} \|_{\text{TV}}$ is computed from truth-latent-image $\mathbf{f}_{\text{true}}$. For brevity, we show results obtained only for program DKL-fTV in Eq. (4.4), as similar results are obtained also for programs in Eqs. (4.5) and (4.6). It can be observed in Fig. (4.2) that convergent reconstruction $\mathbf{u}_{\text{conv}}$ is visually virtually identical to truth image $\mathbf{u}_{\text{true}}$. Also, Fig. 4.3 shows how convergence metrics evolve as functions of the iteration number, demonstrating that the practical convergence conditions in Eq. (4.25) are satisfied. Because $\mathbf{u}_{\text{true}}$ is known, we also calculated $\overline{\Delta}(\mathbf{u}_n) = \frac{||\mathbf{u}_n - \mathbf{u}_{\text{true}}||_2}{||\mathbf{u}_{\text{true}}||_2}$, and display it in Fig. 4.3, which indicates that the reconstruction converges to the truth image. Furthermore, it can be observed that the convergent metrics maintain their decaying trends beyond the convergence

conditions in Eq. (4.25). Therefore, the results of the inverse-crime study numerically assure the correctness of computer implementation of the CP algorithms.

4.4 Discussions

In this chapter, we have investigated image reconstruction by using the CP algorithm, a primal-dual optimization algorithm, for solving image-TV-constrained optimization problems associated with different data fidelities. In particular, we studied configurations containing less detectors than a conventional PET configuration. Again, we have carried out inverse-crime studies to verify whether the CP algorithm can accurately solve the designed optimization programs. In the next chapter, we will evaluate the performance of the CP algorithm in reconstruction scenarios where the data contain inconsistency. It is worth noting that, the behavior of the optimization-based reconstruction may depend on numerous factors, which include the selection of program parameters (e.g., image constraint γ and voxel size), different designs of optimization problem (e.g., different data fidelities and image constraints), and data conditions, *e*tc.

Optimization programs with different data divergences were investigated in the work, because evidence exists indicating their potential theoretical and practical utility in CT imaging. In each of the programs considered, an image-TV constraint was used, because accumulating results in CT imaging appear to suggest that the TV-constrained programs may yield reconstructions superior, or at least comparable, to those obtained with other optimization-based reconstructions. It is meritorious to investigate additional optimization programs with different data divergences and image constraints.

A Gaussian blurring matrix is included in the designed optimization programs. This is because data collected in PET imaging generally have SNR considerably lower than that of data in typical CT imaging, and transitions among different uptake regions in a PET-activity map are observed to be generally not as sharp as that among some anatomic regions in a CT-attenuation map. Therefore, a desired PET image is formulated as a product of a

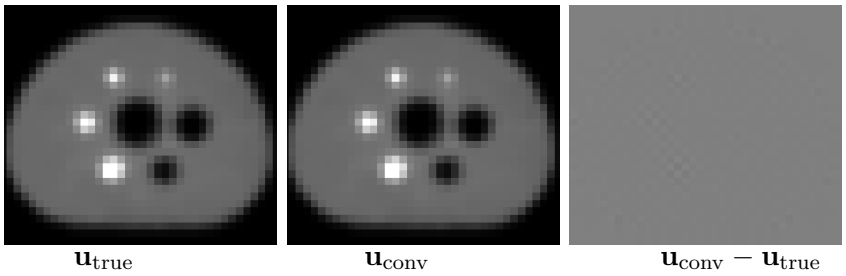


Figure 4.2: Truth image $\mathbf{u}_{\text{true}}$, convergent reconstruction $\mathbf{u}_{\text{conv}}$, and difference $\mathbf{u}_{\text{conv}} - \mathbf{u}_{\text{true}}$. Display window [0.0, 0.1] for the truth image and reconstruction, and display window [-0.0001, 0.0001] for their difference. Image values are in arbitrary units.

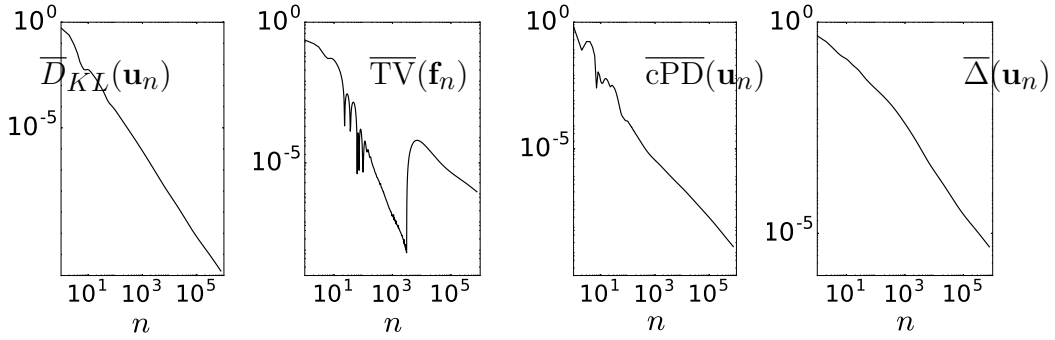


Figure 4.3: Convergence metrics $\overline{D}_{KL}(\mathbf{u}_n)$, $\overline{\text{TV}}(\mathbf{f}_n)$, $\overline{\text{cPD}}(\mathbf{u}_n)$, and $\overline{\Delta}(\mathbf{u}_n)$ as functions of iteration number n.

latent image and Gaussian blurring matrix, and the TV constraint is applied to the latent image instead of the desired PET image. This formulation thus allows for a latent image with sparser gradient magnitude image than the desired image, and avoids yielding an image with significant patchy textures when the data have low SNR [186].

CHAPTER 5

INVESTIGATION OF OPTIMIZATION-BASED RECONSTRUCTION IN PET

5.1 Introduction

The previous chapter focused on designing optimization programs containing *1)* image-TV constraint, *2)* Gaussian blurring operation, and *3)* different data fidelities, such as data-KL, data-ℓ_2, and data-ℓ_1. The investigation of the proposed optimization problems is facilitated by CP algorithm instances. In an attempt to verify the correctness of the derived algorithms as well as their implementation, we carried out inverse-crime studies where the system matrix used in image reconstruction is exactly the same as that used in data generation. In this chapter, we will further carry out quantitative characterization and qualitative evaluation studies on the optimization algorithms.

The focus of this chapter is to investigate technical characteristics of TV-constrained, optimization reconstructions of PET images from Monte-Carlo-simulation/real data, instead of showing whether it is superior to existing reconstruction techniques. Specifically, we study how different designs of the optimization programs with different data divergences affect PET-image properties, and we identify clearly parameters involved in, and demonstrate their impact on, the optimization-based reconstructions considered. With the understanding of the characteristics of TV-constrained optimization programs and their associated parameters, insights may be gained for development of practically useful algorithms for PET-image reconstruction in preclinical and clinical applications.

The robustness of the TV-constrained, optimization-based reconstruction was investigated by applying it to objects with considerably different uptake maps, and to data with different levels of quality/quantity. In this work, we conducted a Monte-Carlo-simulation study from a digital Jaszczak phantom, a real-data study using a physical IEC phantom, and a patient study. Moreover, a study was conducted for image reconstruction from data

collected by use of a PET configuration with sparsely populated detectors, amounting to only half the number of detectors in a digital prototype PET/CT system. The study demonstrates the potential for algorithm-enabled design of a PET configuration with a reduced number of detectors. The saved detectors can be exploited, e.g., for extending the longitudinal coverage of a PET scanner. It is worth noting that, in the Jaszcak-phantom study, the knowledge of the truth image is known, and in the IEC-phantom study, the ground truth is available from the phantom manufacturer. Therefore, quantitative characterization of the algorithm can be performed via comparing the reconstruction to the truth image, or true information about the phantom. In the patient study, visual assessment is conducted to evaluate the potential practical utilities of the reconstructions.

In this chapter, we carried out a Monte-Carlo-simulation study in Sec. 5.2, a physical IEC-phantom study in Sec. 5.3, and a human-data study in Sec. 5.4. Sections 5.2, 5.3, and 5.4 follow a similar structure in which a brief description of the data is provided followed by investigations of TV-constraint, different optimization programs, convergent reconstructions from full and sparse data, and reconstructions as a function of iteration. Remarks about the work are given in Sec. 5.5.

5.2 Monte-Carlo-Simulation Study

For studies in Secs. 5.2-5.4 below, we design practical convergence conditions

$$\begin{aligned} \left| \frac{\partial \overline{D}(\mathbf{u}_n)}{\partial n} \right| &< 10^{-3} \\ \overline{\mathrm{TV}}(\mathbf{f}_n) &< 10^{-3}, \end{aligned} \tag{5.1}$$

which are less restrictive than those in Eq. (4.25) for the inverse-crime study. The conditions are designed based upon the practical considerations: (a) in a real-data study, because some inconsistency always exists between measured data $\mathbf{g}_m$ and model data $\mathbf{g}$, $\overline{D}(\mathbf{u}) > 0$ is generally non-zero, (b) due to the considerably large size of data and image arrays in real-

data studies, it is often challenging computationally to achieve convergence conditions as tight as those given in Eq. (4.25) used for the inverse-crime study, and (c) our study also suggests that the convergence conditions in Eq. (5.1) yield reconstructions similar to those obtained with a tighter convergence conditions. Also, because the inconsistency between the measured and model data is generally unknown, $\overline{D}(\mathbf{u}_n)$ converges to an unknown non-zero number. It is thus unclear how an upper bound on $\overline{D}(\mathbf{u}_n)$ can be determined. Conversely, we expect that, as the algorithm converges, the absolute difference in $\overline{D}(\mathbf{u}_n)$ between two successive iterations diminishes. Therefore, an upper bound (i.e., 10^{-3}) on $\left|\frac{\partial \overline{D}(\mathbf{u}_n)}{\partial n}\right|$ is set as a practical convergence condition. Moreover, unlike metrics $\overline{\mathrm{TV}}(\mathbf{f})$ and $\overline{D}(\mathbf{u})$ that provide direct measures of physical properties of reconstruction and data estimation in a practical study, metric $\overline{\mathrm{cPD}}(\mathbf{u}_n)$ yields a check on the mathematical correctness of the algorithm implementation. Consequently, once the implementation correctness is verified in the inverse-crime study, metric $\overline{\mathrm{cPD}}(\mathbf{u}_n)$ is not used in real-data studies, and practical convergence conditions in Eq. (5.1) thus were used instead in Monte-Carlo-simulation and real-data studies.

Prior to physical-phantom and human studies, we conducted a Monte-Carlo-simulation study in which full data of ∼200 million total counts were generated from the digital Jaszczak phantom by using the GATE simulation package [190] for the full-scan configuration discussed in Sec. 4.2.1. From the full data, we also extracted sparse data and carried out reconstructions for the sparse-scan configuration. Without loss of generality, the simulation study includes only truth events. Images are reconstructed on a 3D array of $N = 70 \times 70 \times 41$ identical cubic voxels of size 4 mm. Digital phantom $\mathbf{u}_{\text{true}}$ consists of cold- and hot-rod sections each of which contains six types of cylindrical-shaped rods of diameters 4.8, 6.4, 7.9, 9.5, 11.1, and 12.7 mm. Although GATE data contain only true coincidence events, they are inconsistent with the model data in Eq. (4.1) due to noise and other physical factors included by GATE, but not in the data model. Using GATE data and knowledge of truth image $\mathbf{u}_{\text{true}}$, we characterize how optimization-based reconstruction responds to data inconsistency prior

to its application to real data in which knowledge of the truth image is unavailable.

5.2.1 Determination of TV-constraint parameter t_1

In the study, given the convergence conditions in Eq. (5.1), all of the program parameters are determined, as described in Sec. 4.2.7, except for the TV-constraint parameter t_1, which is determined by use of the root-mean-square error (RMSE),

$$\text{RMSE} = \frac{1}{\sqrt{N}}||\mathbf{u}_{\text{true}} - \mathbf{u}_{\text{conv}}||_2, \tag{5.2}$$

between truth image $\mathbf{u}_{\text{true}}$ and convergent reconstruction $\mathbf{u}_{\text{conv}}$. For each of a set of t_1 values, we solve program DKL-fTV in Eq. (4.4) to obtain convergent reconstruction $\mathbf{u}_{\text{conv}}$ from full data and calculate the RMSE. Repeating the reconstruction and calculation for all values of t_1, we obtain an RMSE of t_1, as depicted in Fig. 5.1. Based upon the RMSE result, we select $t_1 = 13488$, which yields a minimum RMSE, for obtaining reconstruction results in subsequent sections. The RMSE metric was used for selecting t_1 in the study simply because knowledge of truth image $\mathbf{u}_{\text{true}}$ is available, and the recovery of the truth image is the "utility" goal in the study.

5.2.2 Reconstructions based upon different optimization programs

Using the selected t_1, we obtain reconstructions from full data by solving programs DKL-fTV, DL2-fTV, and DL1-fTV in Eqs. (4.4)-(4.6), and display them in Fig. 5.2, along with the reference image reconstructed by use of RAMLA introduced in Sec. 4.2.4. It can be observed that all of the convergent reconstructions $\mathbf{u}_{\text{conv}}$ for the cold-rod section appear visually comparable, only with slightly different noise textures, while program DKL-fTV yields a reconstruction of the hot-rod section with spatial resolution slightly superior to those obtained with programs DL2-fTV and DL1-fTV. However, the DL1-fTV reconstruction of the hot-rod section appears to contain prominent zero-valued artifacts. In fact, the DL1-fTV

reconstructions from data of the IEC phantom and human subject also comprise significant artifacts, as shown in Figs. 5.8 and 5.15 below.

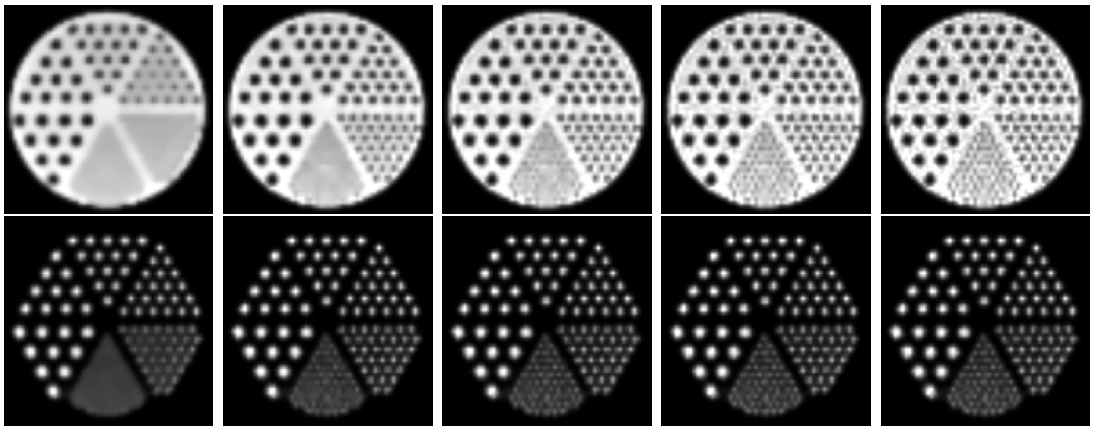
In an attempt to illustrate the artifact source, we define

$$\mathbf{g}'_m = \mathbf{g}_m - \mathbf{g}_s - \mathbf{g}_r \qquad \text{and} \qquad \mathbf{g}' = \mathcal{H}\mathbf{u}_{\text{conv}}, \tag{5.3}$$

where $\mathbf{g}'_m$ denotes the measured data with scatter/random corrected for, and $\mathbf{g}'$ the model data estimated from the convergent reconstruction $\mathbf{u}_{\text{conv}}$, also with scatter/random corrected for. In particular, we use $\mathbf{g}'_{KL}$, $\mathbf{g}'_{\ell_2}$, and $\mathbf{g}'_{\ell_1}$ to specify explicitly the model data estimated, respectively, by use of programs DKL-fTV, DL2-fTV, and DL1-fTV, and display in Fig. 5.3 their example profiles for the Jaszczak and IEC phantoms and the human subject. The profile results suggest that the minimization of data ℓ_1-norm in program DL1-fTV yields the estimated model data "biased" toward zero due to the prevalence of zero or small-valued measurements, thus producing artifacts observed in the DL1-fTV-based reconstruction of the hot-rod section in Fig. 5.2. It is possible that artifacts in DL1-fTV reconstructions may be different when program parameters are different than those used in the work.

5.2.3 Reconstructions based upon optimization program DKL-fTV

The results above indicate that overall, program DKL-fTV appears to produce reconstructions of reasonable visual textures for both cold- and hot-rod sections. Program DKL-fTV was chosen for further investigation based on visualization of converged patient data reconstructions — the data set used in section 5.4 — at different values of t_1 for each optimization program. As can be seen from the example in Figure 5.15, the DKL-fTV reconstructions tended to have better delineated boundaries than the DL2-fTV reconstructions, while DL1-fTV yielded zero-valued reconstructions. Convergent reconstructions $\mathbf{u}_{\text{conv}}$ obtained from both full and sparse data by use of the CP algorithm to solve DKL-fTV with $t_1 = 13488$ are displayed Fig. 5.4, along with the reference reconstruction. The results indicate that,



$\mathbf{u}_{\text{conv}}$, t_1=6225 $\mathbf{u}_{\text{conv}}$, t_1=10375 $\mathbf{u}_{\text{conv}}$, t_1=13488 $\mathbf{u}_{\text{conv}}$, t_1=17638 $\mathbf{u}_{\text{conv}}$, t_1=25938

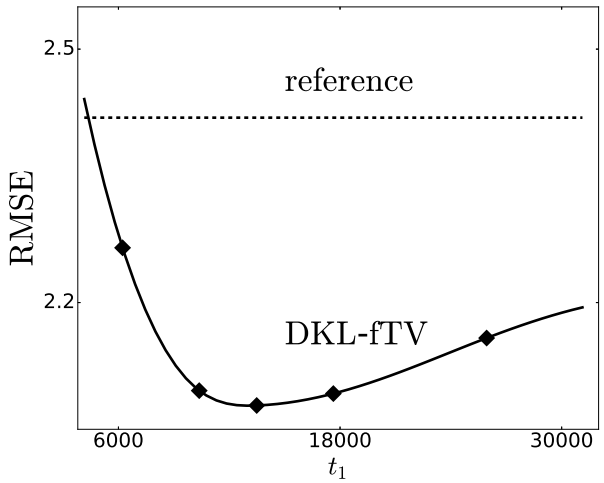
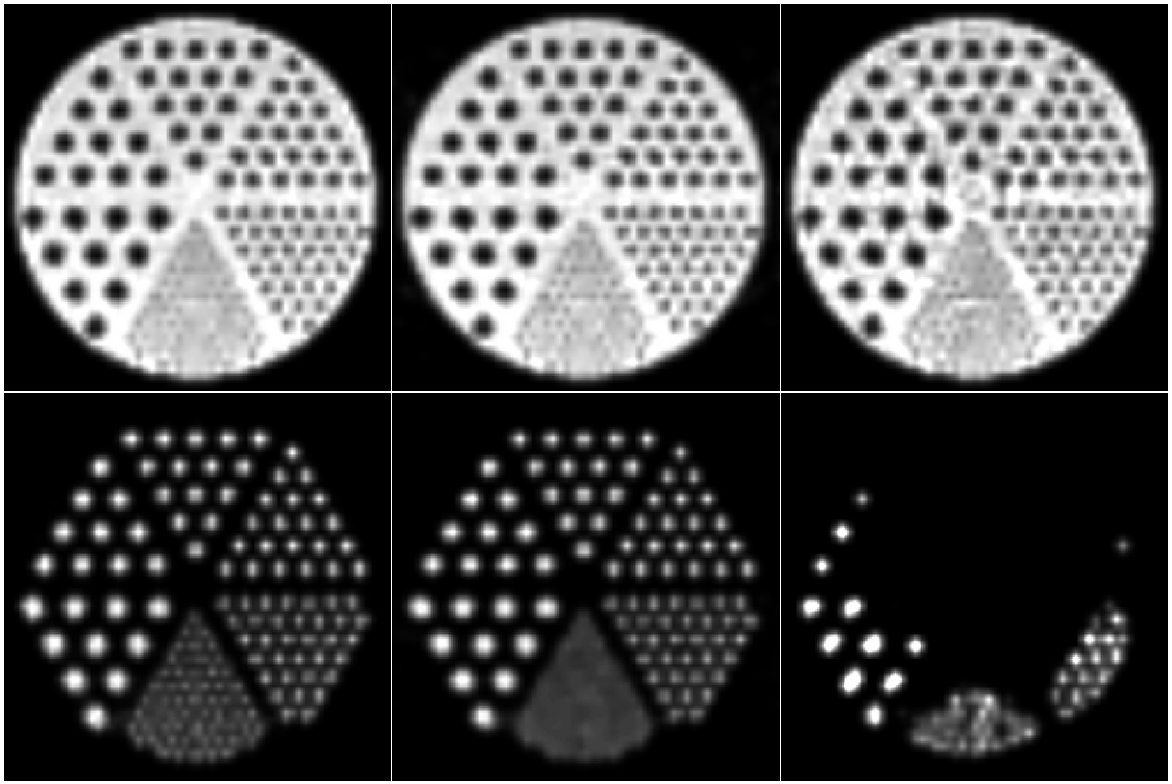


Figure 5.1: Convergent reconstructions $\mathbf{u}_{\text{conv}}$ within transverse slices containing cold (row 1) and hot (row 2) rods in the Jaszczak phantom obtained from full data with program DKL-fTV for different t_1 values. The plot shows metrics RMSE calculated from $\mathbf{u}_{\text{conv}}$ (solid) and the reference reconstruction (dashed), as functions of t_1.

in general, reconstructions from full data appear visually to possess better spatial resolution and lower noise level than do those from sparse data, and that reconstructions obtained with program DKL-fTV seem to reveal enhanced spatial and contrast resolution in which hot/cold rods of size 6.4 mm appear to remain resolved. As shown in the pseudo code for the CP algorithm, latent image $\mathbf{f}$ defined in Eq. (4.3) can also be reconstructed. It is interesting to display convergent reconstructions $\mathbf{u}_{\text{conv}}$ and $\mathbf{f}_{\text{conv}}$ of desired and latent images $\mathbf{u}$ and $\mathbf{f}$. As shown in Fig. 5.5, it can be observed that $\mathbf{f}_{\text{conv}}$ tends to have noisier textures than $\mathbf{u}_{\text{conv}}$.



$\mathbf{u}_{\text{conv}}$, DKL-fTV $\mathbf{u}_{\text{conv}}$, DL2-fTV $\mathbf{u}_{\text{conv}}$, DL1-fTV

Figure 5.2: Convergent reconstructions $\mathbf{u}_{\text{conv}}$ within transverse slices containing cold (row 1) and hot (row 2) rods in the Jaszczak phantom obtained from full data with programs DKL-fTV, DL2-fTV, and DL1-fTV in Eqs. (4.4)-(4.6), respectively. Display windows: [0, 40000] (row 1) and [0, 15000] (row 2). Note that the DL1-fTV reconstruction of the hot-rod section contains a large, zero-valued region.

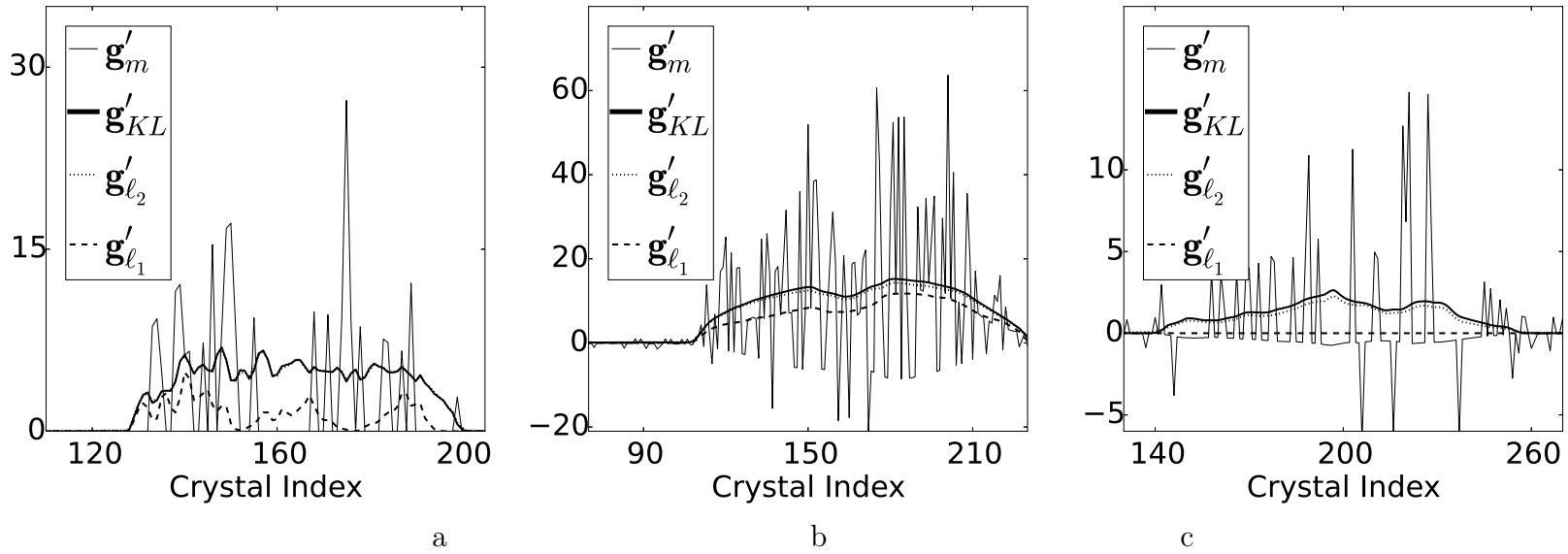
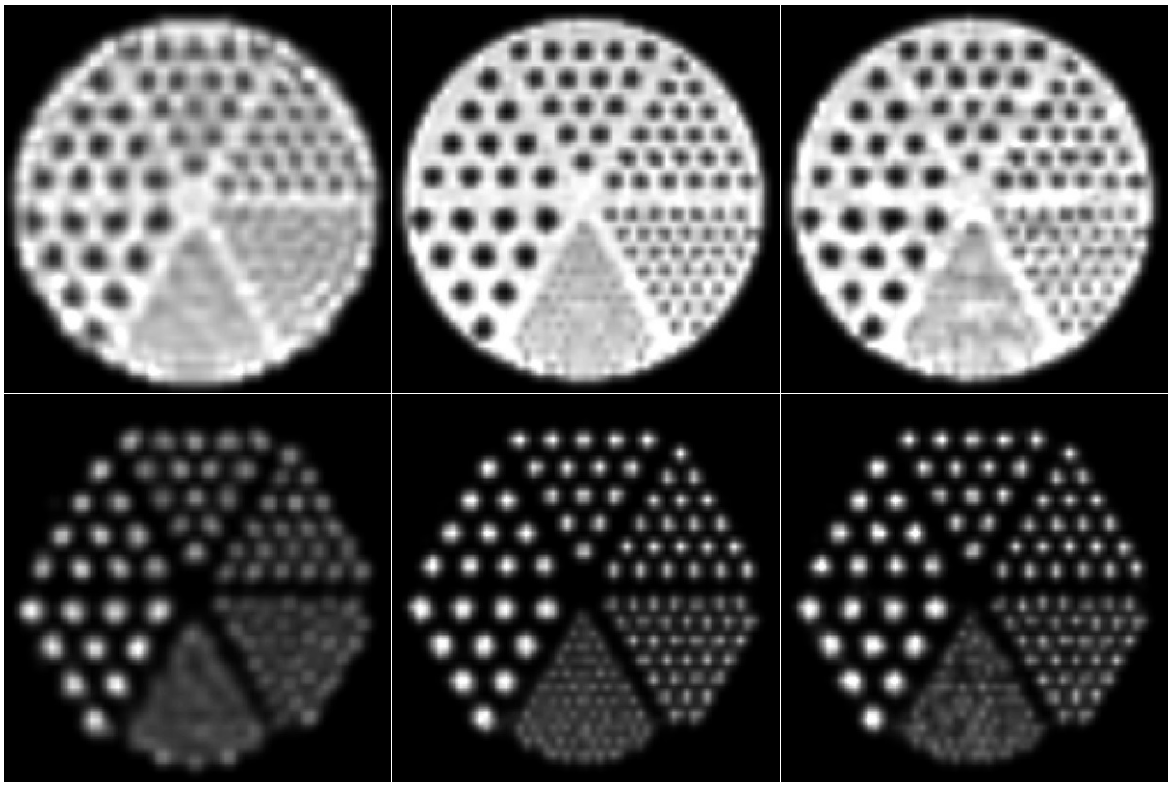
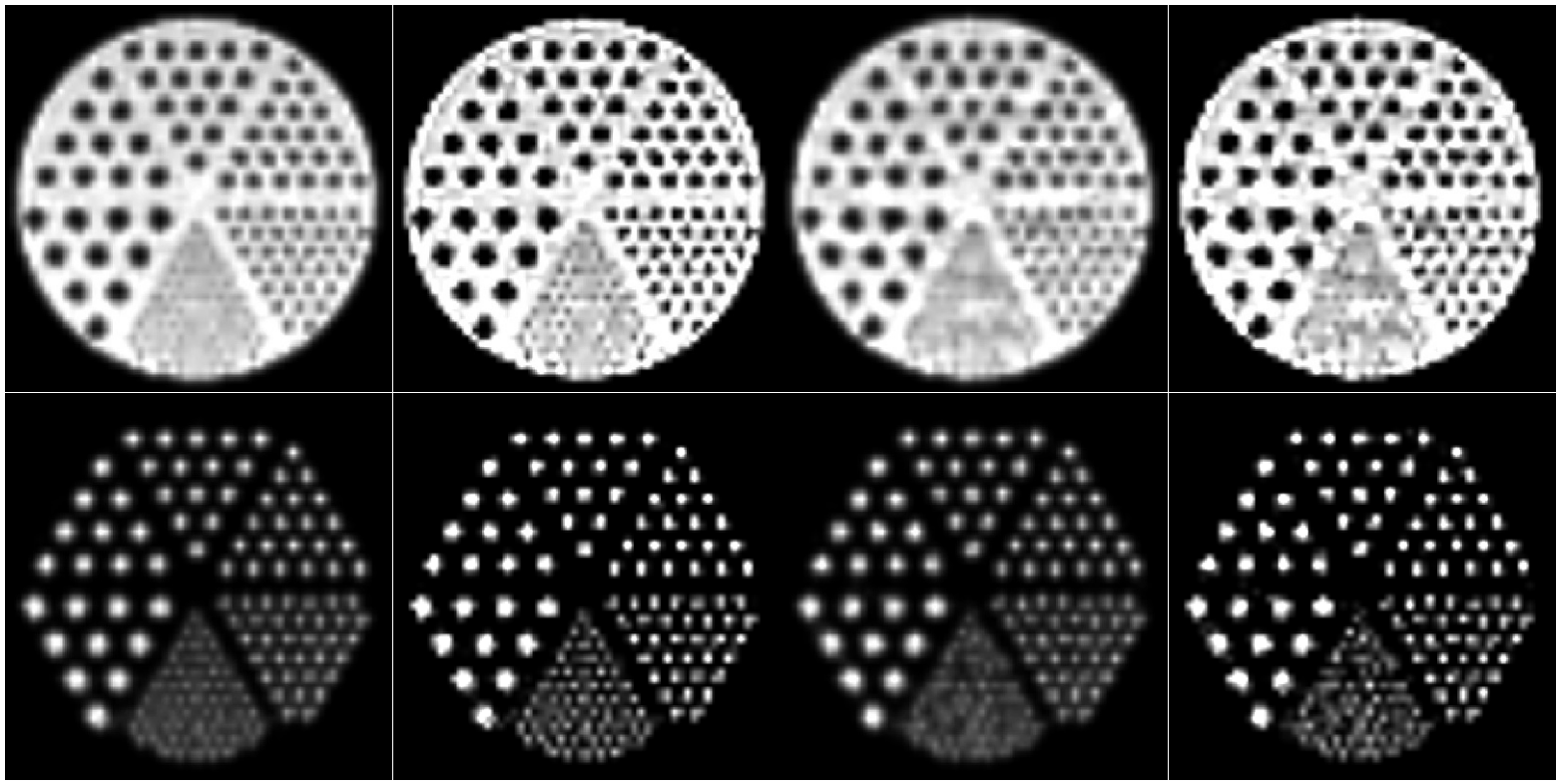


Figure 5.3: Profiles of measured data $\mathbf{g}'_m$ (thin solid), and model data $\mathbf{g}'_{KL}$ (thick solid), $\mathbf{g}'_{\ell_2}$ (dotted), and $\mathbf{g}'_{\ell_1}$ (dashed) obtained with programs DKL-fTV, DL2-fTV, and DL1-fTV for (a) the Jaszczak phantom, (b) the IEC phantom, and (c) the human subject.



$\mathbf{u}_{\text{ref}}$ $\mathbf{u}_{\text{conv}}$, full data $\mathbf{u}_{\text{conv}}$, sparse data

Figure 5.4: Reference reconstructions (column 1), and convergent reconstructions $\mathbf{u}_{\text{conv}}$ within transverse slices containing cold (row 1) and hot (row 2) rods in the Jaszczak phantom obtained from full (column 2) and sparse (column 3) data with program DKL-fTV. Display windows: [0, 40000] (row 1) and [0, 15000] (row 2).



$\mathbf{u}_{\text{conv}}$, full data $\mathbf{f}_{\text{conv}}$, full data $\mathbf{u}_{\text{conv}}$, sparse data $\mathbf{f}_{\text{conv}}$, sparse data

Figure 5.5: Convergent reconstructions $\mathbf{u}_{\text{conv}}$ and $\mathbf{f}_{\text{conv}}$ within transverse slices containing cold (row 1) and hot (row 2) rods in the Jaszczak phantom obtained from full (columns 1 & 2) and sparse (columns 3 & 4) data with program DKL-fTV. Display windows: [0, 40000] (row 1) and [0, 15000] (row 2).

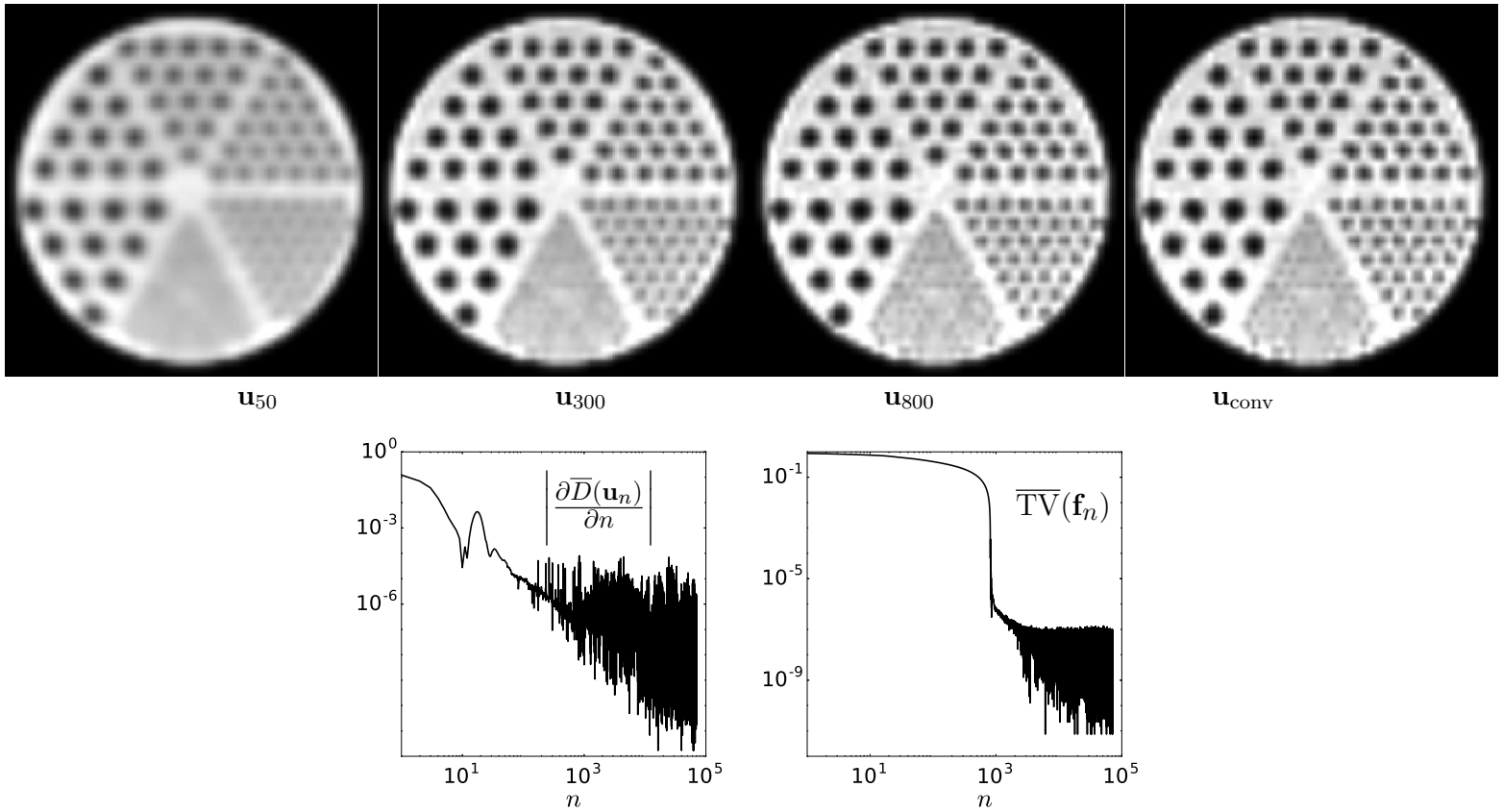


Figure 5.6: Reconstructions $\mathbf{u}_n$ at iteration n =50, 300, and 800 within a transverse containing cold rods in the Jaszczak phantom from full data with program DKL-fTV, along with convergent reconstruction $\mathbf{u}_{\text{conv}}$ (iteration 71160). Display windows: [0, 40000]. The two plots display convergence metrics $\left|\frac{\partial \overline{D}(\mathbf{u}_n)}{\partial n}\right|$ and $\overline{\text{TV}}(\mathbf{f}_n)$ as functions of iterations n.

5.2.4 Reconstruction as a function of iterations

The reconstructions above were obtained when the convergence conditions in Eq. (5.1) were satisfied. In an attempt to inspect how they evolve as functions of iterations, we display in Fig. 5.6 those obtained from full data with program DKL-fTV in Eq. (4.4) at iterations n =50, 300, and 800, as well as convergent reconstruction $\mathbf{u}_{\text{conv}}$. Furthermore, convergence metrics $\overline{D}_{KL}(\mathbf{u}_n)$ and $\overline{\text{TV}}(\mathbf{f}_n)$, computed from the reconstructions are shown in Fig. 5.6 as functions of iterations. Similar reconstructions and plots were obtained also for the sparse-data study, but are not shown here. These results reveal that the reconstruction at, e.g., about iterations 300 is visually similar to the convergent reconstruction.

5.3 Physical-Phantom-Data Study

The IEC phantom is adopted widely for quantitative characterization of PET-image reconstructions. We collected full data of ∼100 million total counts from the phantom by using a full-scan configuration, as depicted in Sec. 4.2.1, in a digital prototype PET/CT system. From the full data, we extracted sparse data to mimic data collected with the sparse-scan configuration. As shown in the Appendix, the IEC phantom is composed of 6 fillable spheres of diameters 10, 13, 17, 22, 28, and 37 mm, respectively, in which the two largest spheres have zero activity, while the other four are filled with positron-emitter activity at a concentration level 4 times the background-activity level. Scatter and random events were measured, and used as known components in the study. Images are reconstructed on a 3D array of $100 \times 100 \times 41$ identical cubic voxels of size 4 mm.

5.3.1 Determination of image-constraint parameter t_1

In the study, given the practical convergence conditions in Eq. (5.1), all program parameters but image-constraint parameter t_1 were determined as described in Sec. 4.2.7. Percent contrasts of hot and cold spheres and percent background variability, described in the Appendix,

are standard metrics designed for evaluation of reconstruction quality of the IEC-phantom. In this study, full knowledge of the truth image is unknown. Therefore, combining the metrics, we form in the Appendix a single quality metric, referred to as the QNR, for determination of t_1. For a set of t_1 values, convergent reconstructions from full data of the IEC phantom were obtained by use of the CP algorithm to solve program DKL-fTV in Eq. (4.4); and the QNRs calculated from the reconstructions are plotted in Fig. 5.7, along with the reconstructions with QNR s highlighted by black diamonds. Based upon the QNR result, we select $t_1 = 977$, which appears to yield a maximum QNR, for reconstructions from full and sparse data of the IEC phantom below.

5.3.2 Reconstructions based upon different optimization programs

Using the selected t_1, we reconstruct IEC images from full data by solving programs DKL-fTV, DL2-fTV, and DL1-fTV in Eqs. (4.4)-(4.6), and display them in Fig. 5.8. Again, the observation can be made that convergent reconstructions obtained with programs DKL-fTV and DL2-fTV appear visually comparable, while the latter has slightly noisier textures than the former, and that program DL1-fTV yields reconstructions with prominent cupping artifacts. Similar to that in the Jaszczak-phantom study, the data-ℓ_1-norm minimization yields the estimated model data of the IEC phantom "biased" toward zero, as shown in Fig. 5.3b, and such a data-estimate "bias" thus gives rise to the artifacts observed in the DL1-fTV-based reconstruction of the IEC phantom in Fig. 5.8. Again, artifacts in the DL1-fTV reconstruction may vary with program parameters different than those used in the work.

5.3.3 Reconstructions based upon optimization program DKL-fTV

For reasons described in section 5.2.3, we select program DKL-fTV for further study on image reconstruction from full and sparse data of the IEC phantom. In Fig. 5.9, we show convergent reconstructions $\mathbf{u}_{\text{conv}}$ obtained with program DKL-fTV from full and sparse data of the IEC phantom. As expected, the reconstruction from full data appears to have a level

of spatial and contrast resolution slightly higher than the reconstruction from sparse data, while both convergent reconstructions have a relatively low level of background noise in which the hot spot of the smallest size (i.e., diameter of 10 mm) remains visible. We subsequently computed from the reconstructions quantitative metrics described in the Appendix, and display them in Fig. 5.10. These results suggest that convergent reconstructions from sparse data are slightly inferior to those from full data, as expected, and that they appear to be largely comparable to, or better than, the reference reconstructions from full data. Again, we display in Fig. 5.11 convergent reconstructions $\mathbf{u}_{\text{conv}}$ and $\mathbf{f}_{\text{conv}}$ of desired and latent images $\mathbf{u}$ and $\mathbf{f}$, and observe that the latter is slightly noisier than the former.

$\mathbf{u}_{\text{conv}}$, t_1=760 $\mathbf{u}_{\text{conv}}$, t_1=814 $\mathbf{u}_{\text{conv}}$, t_1=977 $\mathbf{u}_{\text{conv}}$, t_1=1628

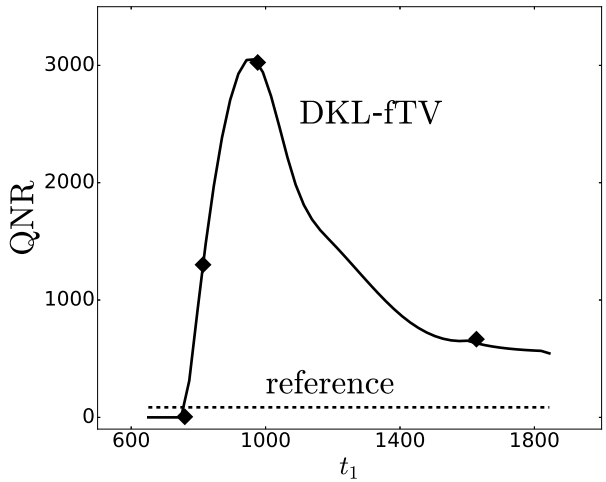
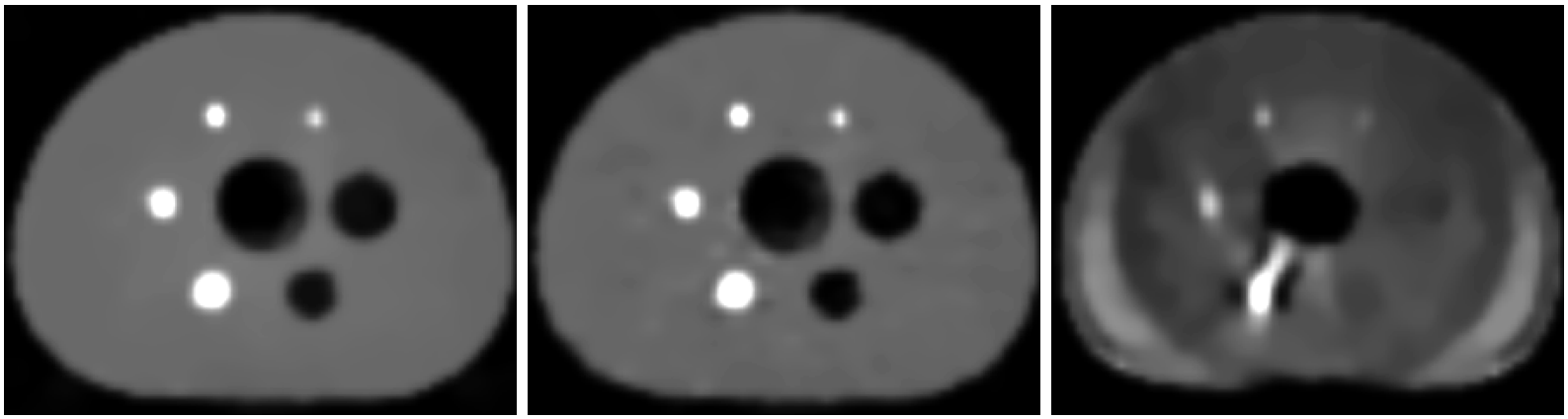
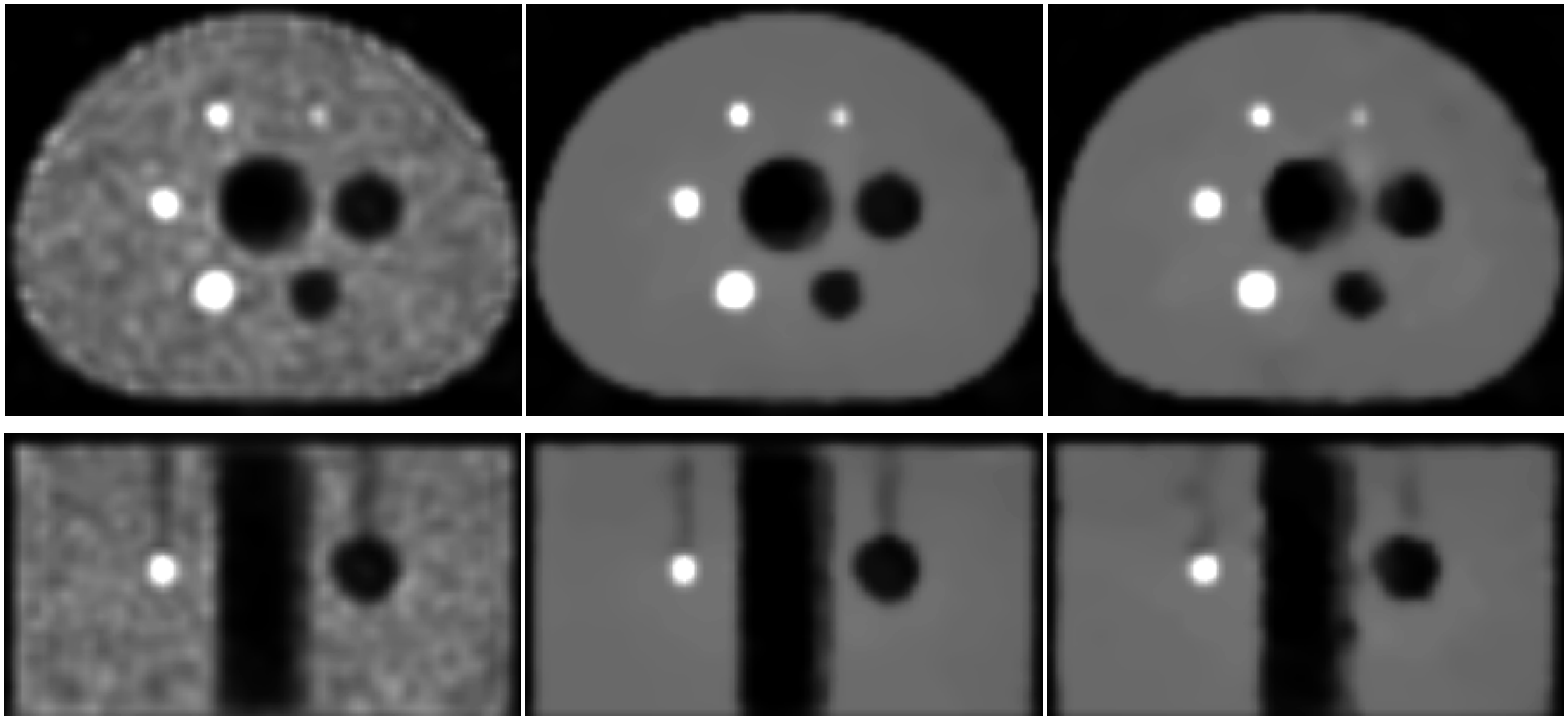


Figure 5.7: Convergent reconstructions $\mathbf{u}_{\text{conv}}$ within the central transverse slice of the IEC phantom obtained from full data with program DKL-fTV for different t_1 values. The plot shows metrics QNR calculated from $\mathbf{u}_{\text{conv}}$ (solid) and the reference reconstruction (dashed), as functions of t_1.



$\mathbf{u}_{\text{conv}}$, DKL-fTV $\mathbf{u}_{\text{conv}}$, DL2-fTV $\mathbf{u}_{\text{conv}}$, DL1-fTV

Figure 5.8: Convergent reconstructions $\mathbf{u}_{\text{conv}}$ within the central transverse slice of the IEC phantom obtained from full data with programs DKL-fTV, DL2-fTV, and DL1-fTV, respectively. Display window: [0, 16000]. Note that prominent artifacts can be observed in the DL1-fTV reconstruction.



$\mathbf{u}_{\text{ref}}$ $\mathbf{u}_{\text{conv}}$, full data $\mathbf{u}_{\text{conv}}$, sparse data

Figure 5.9: Reference reconstructions (column 1), and convergent reconstructions $\mathbf{u}_{\text{conv}}$, within the central transverse (row 1) and central coronal (row 2) slices of the IEC phantom, obtained from full (column 2) and sparse (column 3) data with program DKL-fTV. Display window: [0, 16000].

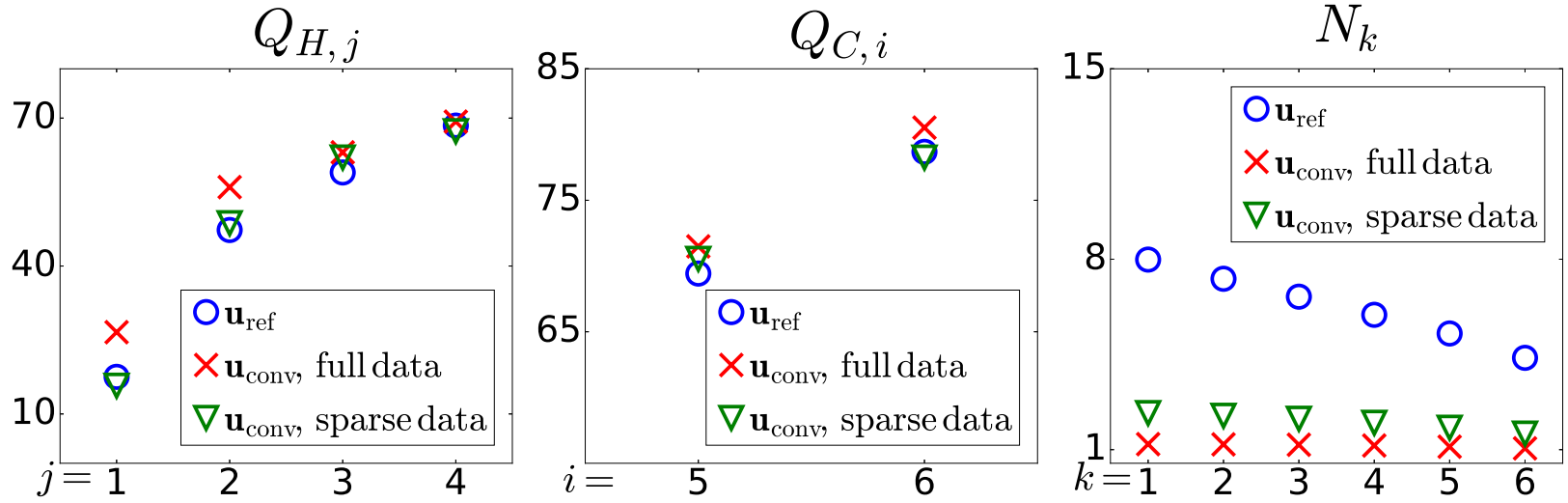
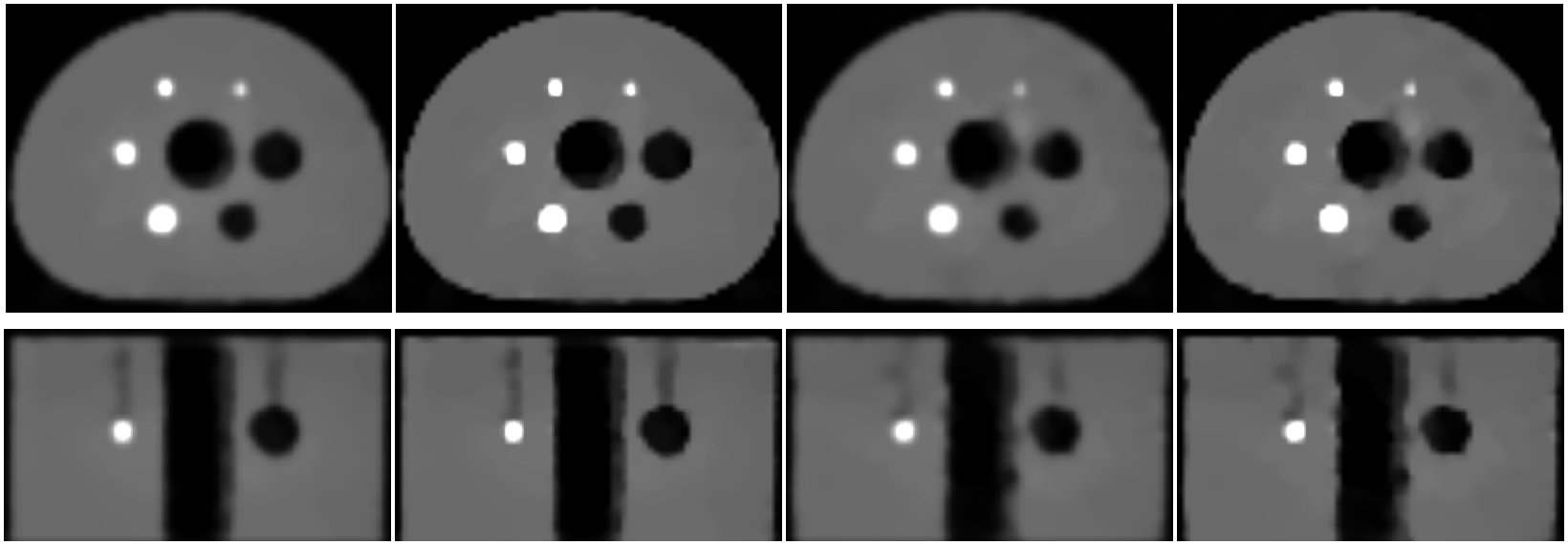
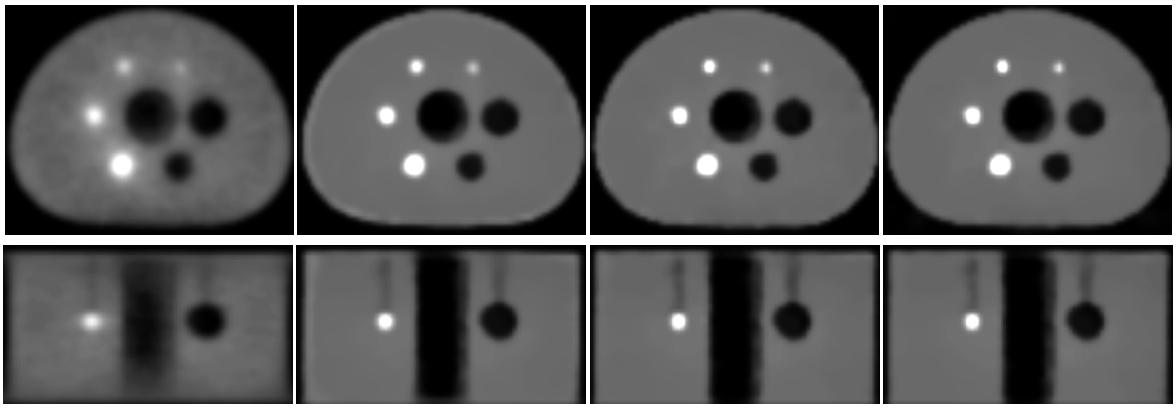


Figure 5.10: Percent contrast $Q_{H,j}$ of hot spheres, where j =1, 2, 3, and 4, percent contrast $Q_{C,i}$ of cold spheres, where i =5 and 6, and percent background variability N_k, where k =1, 2, 3, 4, 5, and 6, respectively, calculated based upon reference reconstruction (◯) and convergent reconstructions $\mathbf{u}_{\text{conv}}$ from full (X) and sparse (∇) data of the IEC phantom shown in Fig. 5.9.



$\mathbf{u}_{\text{conv}}$, full data $\mathbf{f}_{\text{conv}}$, full data $\mathbf{u}_{\text{conv}}$, sparse data $\mathbf{f}_{\text{conv}}$, sparse data

Figure 5.11: Convergent reconstructions $\mathbf{u}_{\text{conv}}$ and $\mathbf{f}_{\text{conv}}$ within the central transverse (row 1) and central coronal (row 2) slices of the IEC phantom obtained from full (columns 1 & 2) and sparse (columns 3 & 4) data with program DKL-fTV. Display window: [0, 16000].



$\mathbf{u}_{10}$ $\mathbf{u}_{50}$ $\mathbf{u}_{200}$ $\mathbf{u}_{\text{conv}}$

Figure 5.12: Reconstructions $\mathbf{u}_n$ at iteration n =10, 50, and 200, along with the convergent reconstruction (iteration 2701), within the central transverse slice (row 1) and central coronal slice (row 2) of the IEC phantom obtained from full data with program DKL-fTV. Display window: [0, 16000].

5.3.4 Reconstruction as a function of iterations

Convergent reconstructions $\mathbf{u}_{\text{conv}}$ of the IEC phantom above were obtained again when the convergence conditions in Eq. (5.1) are satisfied. In an attempt to reveal the reconstruction evolution, we display in row 1 of Fig. 5.12 images of the IEC phantom reconstructed from full data at iterations n =10, 50, and 200 as well as convergent reconstruction $\mathbf{u}_{\text{conv}}$. It can be observed that the reconstruction at, e.g., about iteration 200 visually resembles convergent reconstruction $\mathbf{u}_{\text{conv}}$. Similar observations can be made for reconstructions obtained from sparse data, which are not shown here. We also calculated quality metrics described in the Appendix from full- and sparse-data reconstructions at different iterations, and show in Fig. 5.13 $Q_{H,2}$ of the hot sphere 2 with a diameter of 13 mm, $Q_{C,5}$ of the cold sphere 5 with a diameter of 28 mm, and N_3 within the ROIs of size 17 mm of sphere 3. Results in Fig. 5.13 confirm that reconstructions at about iteration 200 resemble their corresponding convergent reconstructions.

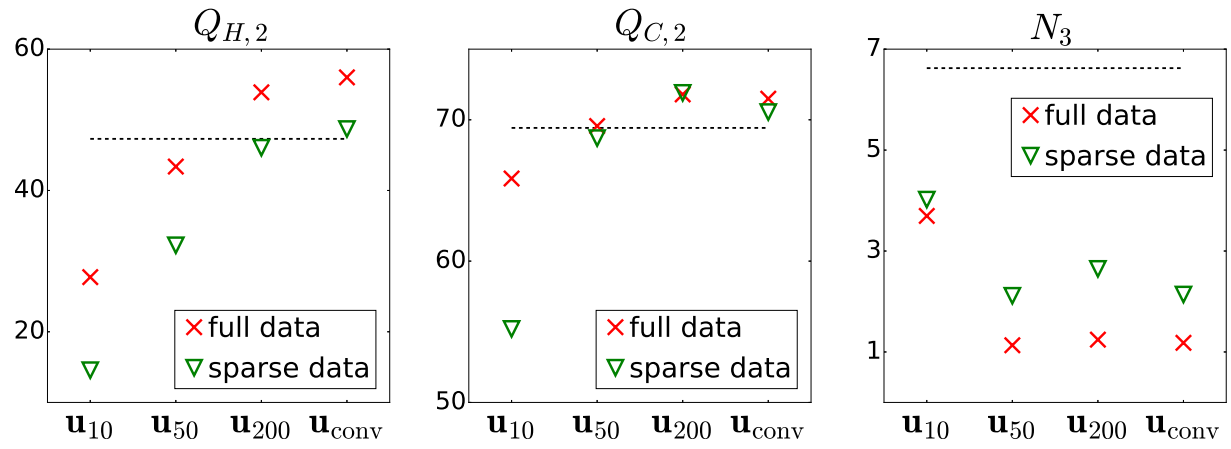


Figure 5.13: Percent contrast $Q_{H,2}$ of hot sphere 2, percent contrast $Q_{C,5}$ of cold sphere 5, and percent background variability N_3 of an ROI of the size of sphere 3, respectively, as functions of iterations n, computed based upon $\mathbf{u}_n$, and upon converged reconstruction $\hat{\mathbf{u}}^*$, obtained from full (X) and sparse (∇) data with program DKL-fTV, along with results obtained from the reference reconstruction (dotted).

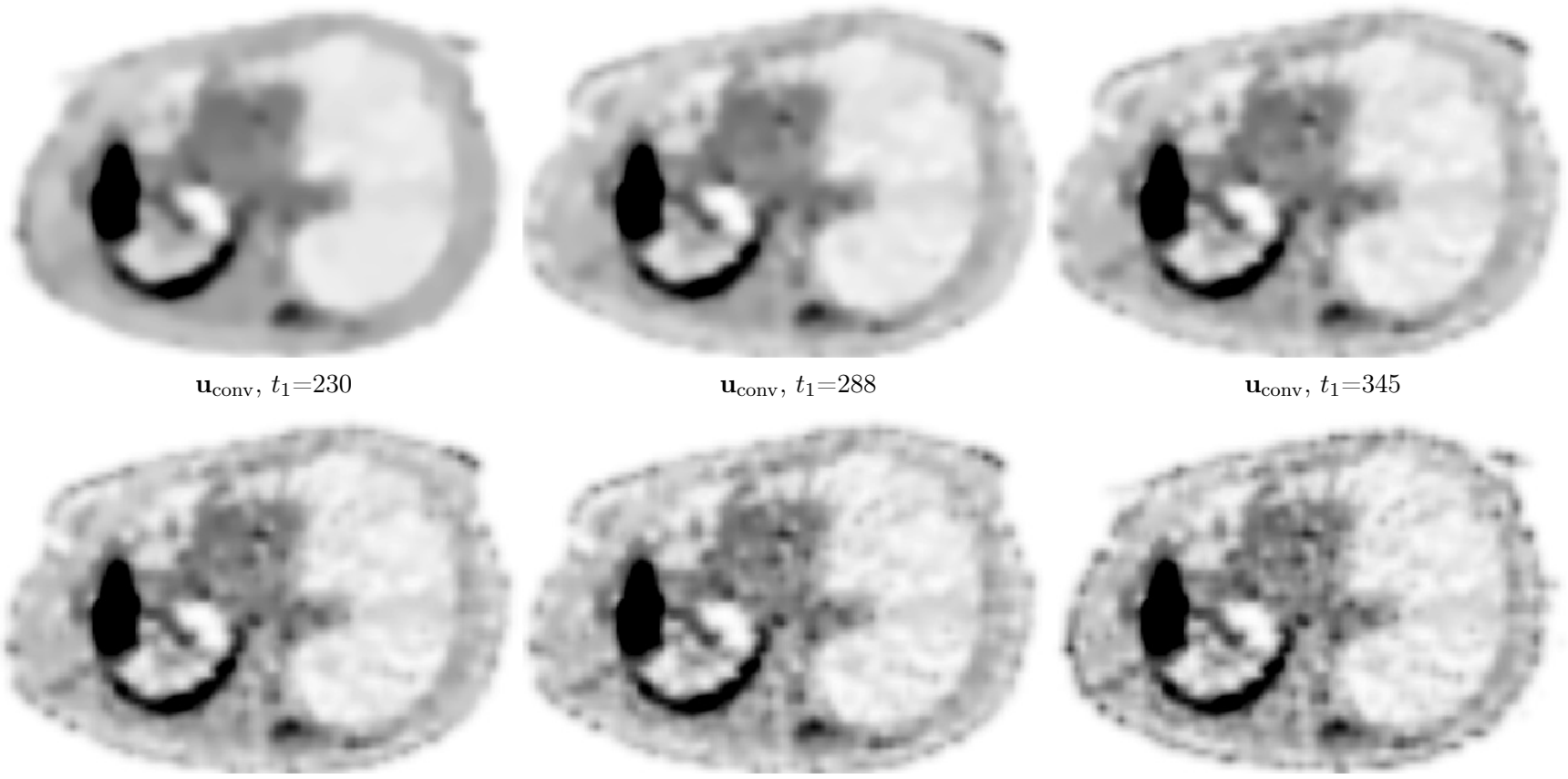
5.4 Human-Data Study

In the human-data study, full data were acquired from the subject by use of a full-scan configuration, as depicted in Sec. 4.2.1, in a digital prototype PET/CT system at a single patient-bed position. In an effort to increase the longitudinal coverage of the human subject, 5 full scans were performed at 5 bed positions along the central axis of the system with a separation of 70 mm between the centers of two consecutive bed positions. The 5 sets of full data acquired contain approximately 23.9, 25.0, 26.7, 24.3, and 22.1 million total counts. Again, we also obtained five sparse data sets each of which was extracted from one of the five sets of full data by use of the method described in Sec. 4.2.1. Scatter and random measurements were included in the study. We reconstruct the human image on a 3D array consisting of $N = 144 \times 144 \times 41$ identical cubic voxels of size 4 mm.

5.4.1 Determination of image-constraint parameter t_1

In the study, again, given the practical convergence conditions in Eq. (5.1), all of the program parameters, but image-constraint parameter t_1, are determined as described in Sec. 4.2.7.

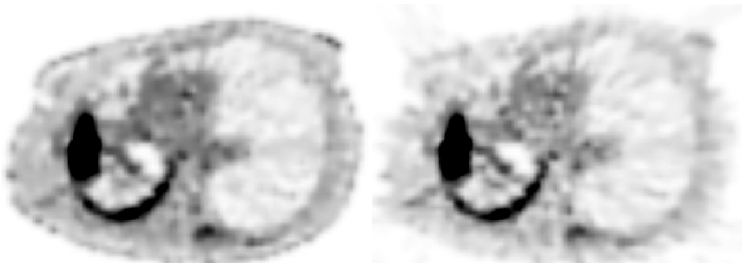
Unlike the studies on the digital Jaszczak phantom in Sec. 5.2 and the physical IEC phantom in Sec. 5.3 in which quantitative metrics were used for selecting t_1, we selected t_1 for the human study based simply upon visual inspection. (It is not entirely clear how a technical efficacy metric analogous to those in the phantom studies can be devised meaningfully for the human study. Although clinical efficacy metrics for a specific clinical task may be devised and used meaningfully for the determination of t_1 optimal for the clinical task, it is beyond the scope of the current work.) For full data collected at bed position 3, we reconstructed human images by solving program DKL-fTV in Eq. (4.4) for a set of t_1 values, and display in Fig. 5.14 convergent reconstructions, obtained with t_1 =230, 287, 345, 460, 575, and 690, respectively. As expected, high t_1 values yield reconstructions with noisy texture while revealing some structural details; conversely, low t_1 values lead to reconstructions with smoothed texture but missing structural details. Based upon the visual inspection of the reconstructions, we select $t_1 = 460$ because it appears to yield reconstructions with a reasonable balance between structural details and image-noise texture. Using the same method, t_1's were determined for reconstructions from other full- and sparse-data sets.



$\mathbf{u}_{\text{conv}}$, t_1=230 $\mathbf{u}_{\text{conv}}$, t_1=288 $\mathbf{u}_{\text{conv}}$, t_1=345

$\mathbf{u}_{\text{conv}}$, t_1=460 $\mathbf{u}_{\text{conv}}$, t_1=575 $\mathbf{u}_{\text{conv}}$, t_1=690

Figure 5.14: Negative convergent reconstructions $\mathbf{u}_{\text{conv}}$ within a transverse slice of the human subject obtained from full data at bed position 3 with program DKL-fTV for different t_1 values. Display window: [-1600, 0].



$\mathbf{u}_{\text{conv}}$, DKL-fTV $\mathbf{u}_{\text{conv}}$, DL2-fTV $\mathbf{u}_{\text{conv}}$, DL1-fTV

Figure 5.15: Negative convergent reconstructions $\mathbf{u}_{\text{conv}}$ within a transverse slice of the human subject obtained from full data at bed position 3 with programs DKL-fTV, DL2-fTV, and DL1-fTV. Display window: [-1600, 0]. Note that program DL1-fTV yields a zero-valued reconstruction.

5.4.2 Reconstructions based upon different optimization programs

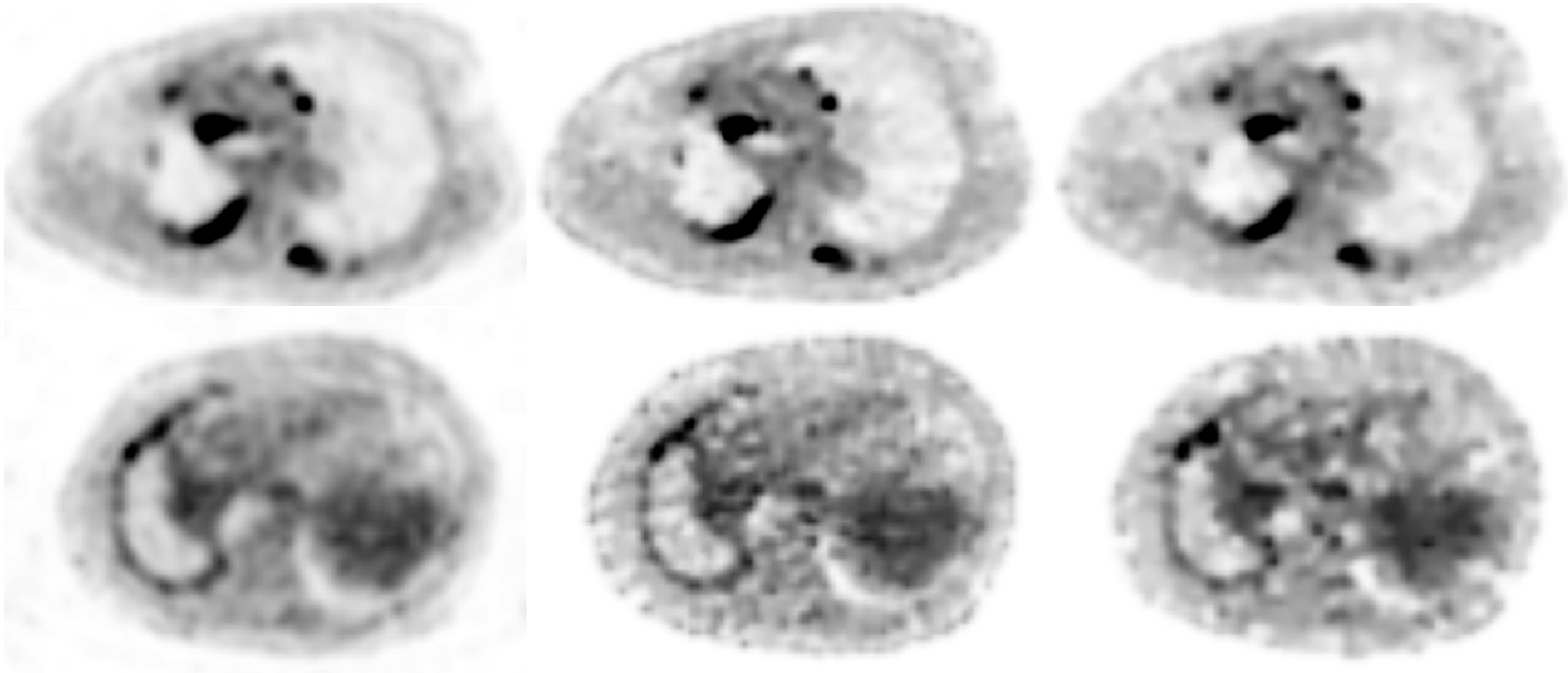
In an attempt to investigate the impact of optimization programs, we carried out a study on reconstructions by using the CP algorithms to solve programs DKL-fTV, DL2-fTV, and DL1-fTV in Eqs. (4.4)-(4.6). For full data collected at a single bed position, we selected t_1 as described above and then reconstruct human images by solving the three programs. We show in Fig. 5.15 the reconstructions from full data collected at bed position 3. The reconstruction obtained with program DKL-fTV appears to possess better delineated boundaries, more structural details, and lower texture noise than does that obtained with program DL2-fTV. However, program DL1-fTV yields a convergent reconstruction with strong artifacts, i.e., with numerically zero values, as depicted in panel 3 of Fig. 5.15. Again, the estimated model data become negatively "biased" as a result of the data-ℓ_1-norm minimization and scatter correction. Consequently, the zero-valued estimated model data revealed in Fig. 5.3c and the zero-valued $\mathbf{u}_{\text{conv}}$ (or, equivalently, zero-valued $\mathbf{f}_{\text{conv}}$) shown in Fig. 5.15, were obtained with the positivity constraint. Clearly, when different program parameters such as t_1 are used, program DL1-fTV may yield reconstructions different than a zero-valued image. Similar results were obtained for reconstructions from full and sparse data collected at this and other bed positions.

5.4.3 Reconstructions based upon optimization program DKL-fTV

Based upon the observation above, we select program DKL-fTV for further study on image reconstruction from human data. For a single bed position, using t_1 selected in Sec. 5.4.1 we reconstructed images from full data collected by using the CP algorithm to solve program DKL-fTV. Repeating this for all of the five bed positions, we obtained five convergent reconstructions, and then concatenated them to form a *final-convergent* reconstruction with an extended longitudinal coverage. Following the same procedure, we also obtained the final-convergent reconstruction of the human subject from the five sets of sparse data. In Figs. 5.16-5.18, we display the final-convergent reconstructions within transverse, coronal, and sagittal slices of the human subject. It can be observed that program DKL-fTV appears to yield reconstructions from full data with well-delineated boundaries and suppressed noise in the background region. Similar observation can be drawn for reconstructions from sparse data as well, despite the fact that reconstructions from sparse data are visually somewhat noisier than those from full data, as expected. We also display in Fig. 5.19 final-convergent reconstructions $\mathbf{u}_{\text{conv}}$ and $\mathbf{f}_{\text{conv}}$ of desired and latent images $\mathbf{u}$ and $\mathbf{f}$.

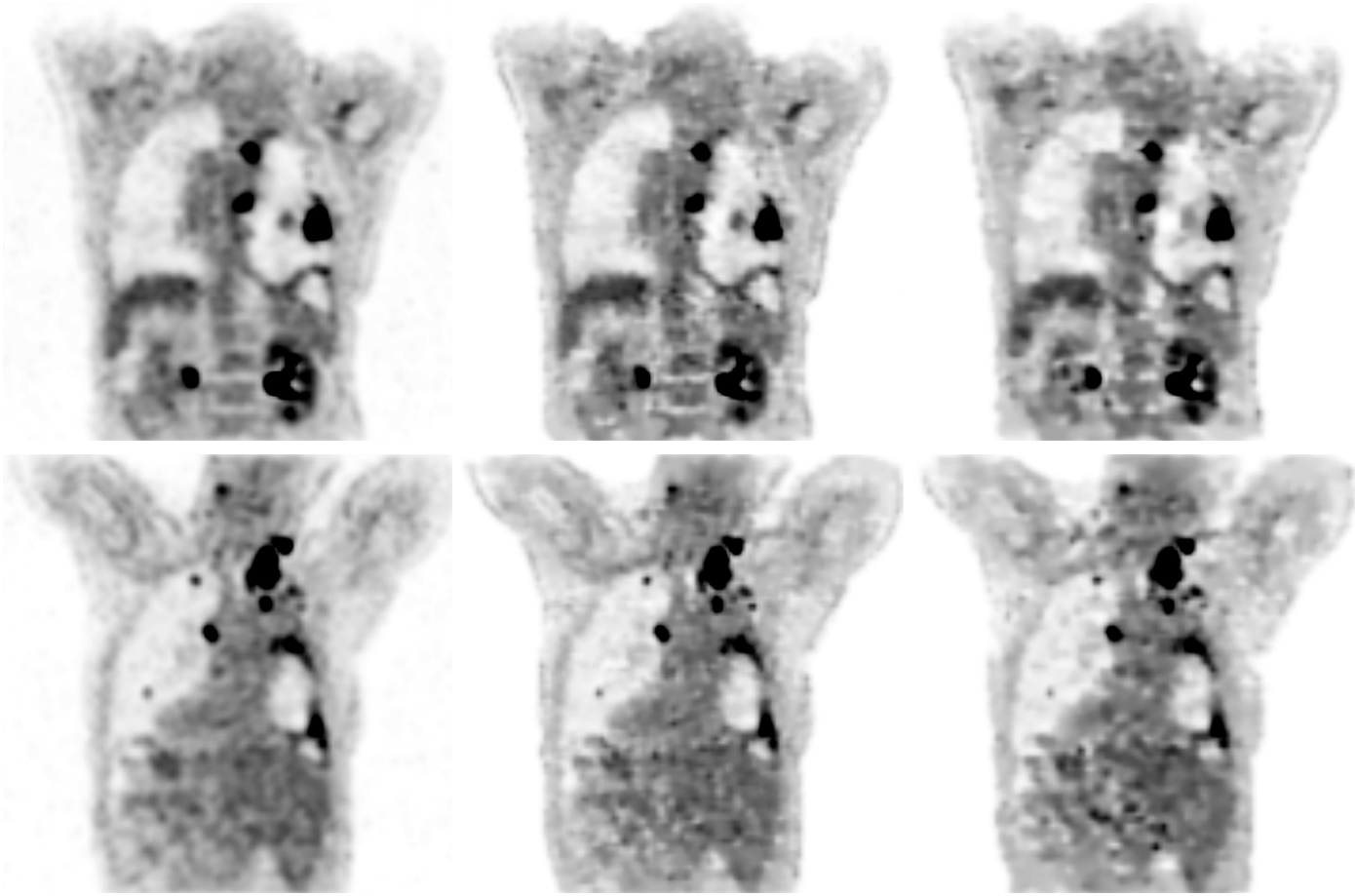
5.4.4 Reconstruction as a function of iterations

Reconstructions of the human images above were obtained again when the practical convergence conditions in Eq. (5.1) are satisfied. We also investigated how the concatenated reconstruction of the human subject evolves as a function of the iteration number. We show in Fig. 5.20 concatenated reconstructions within a coronal slice of the human subject at iteration numbers n =10, 100, 300, 500, 600, 800, and 1000, together with final-convergent reconstruction $\mathbf{u}_{\text{conv}}$. It can be observed that the reconstruction at, e.g., about iteration 600, visually resembles the convergent reconstruction. Similar observations can also be made for reconstructions obtained from sparse data of the human subject.



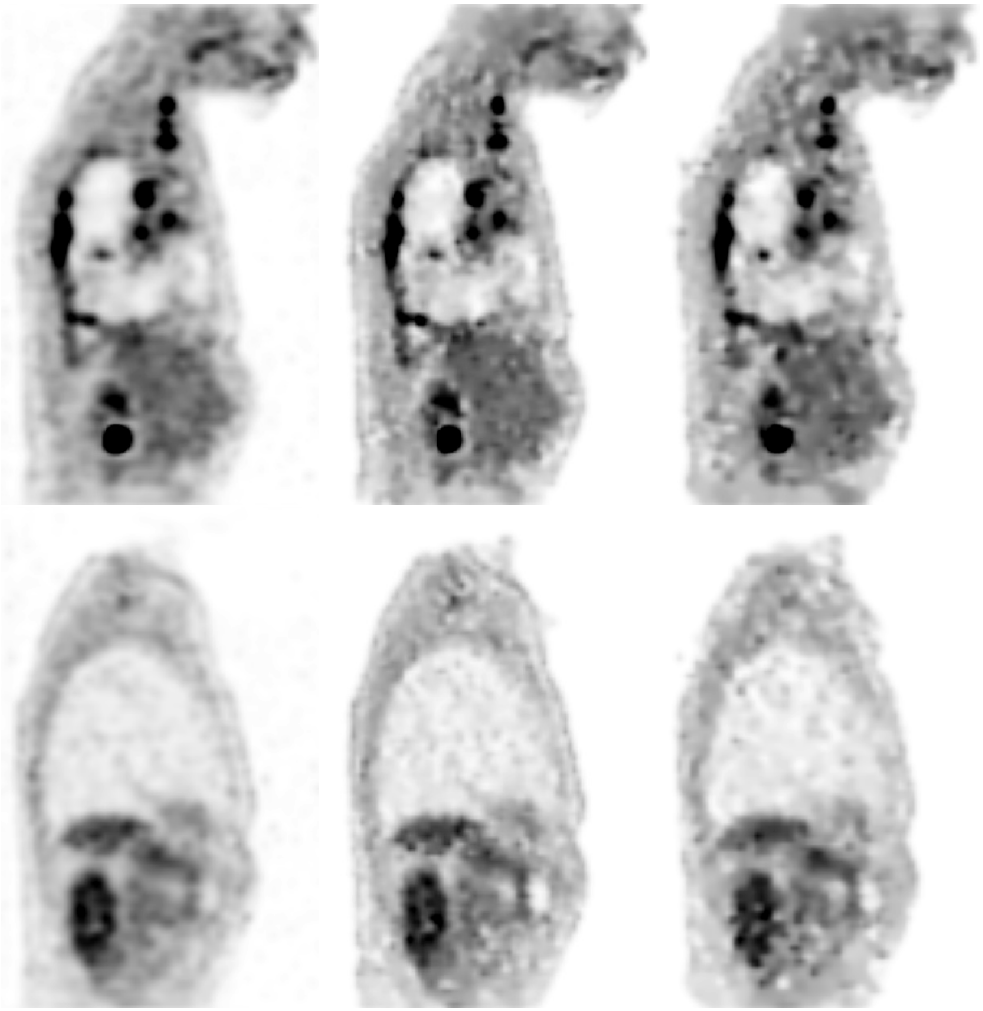
$\mathbf{u}_{\text{ref}}$ $\mathbf{u}_{\text{conv}}$, full data $\mathbf{u}_{\text{conv}}$, sparse data

Figure 5.16: Negative reference reconstructions (column 1), and negative final-convergent reconstructions $\mathbf{u}_{\text{conv}}$, within two transverse slices (rows 1 and 2) of the human subject obtained from full (column 2) and sparse (column 3) data with program DKL-fTV. Display window: [-1600, 0].



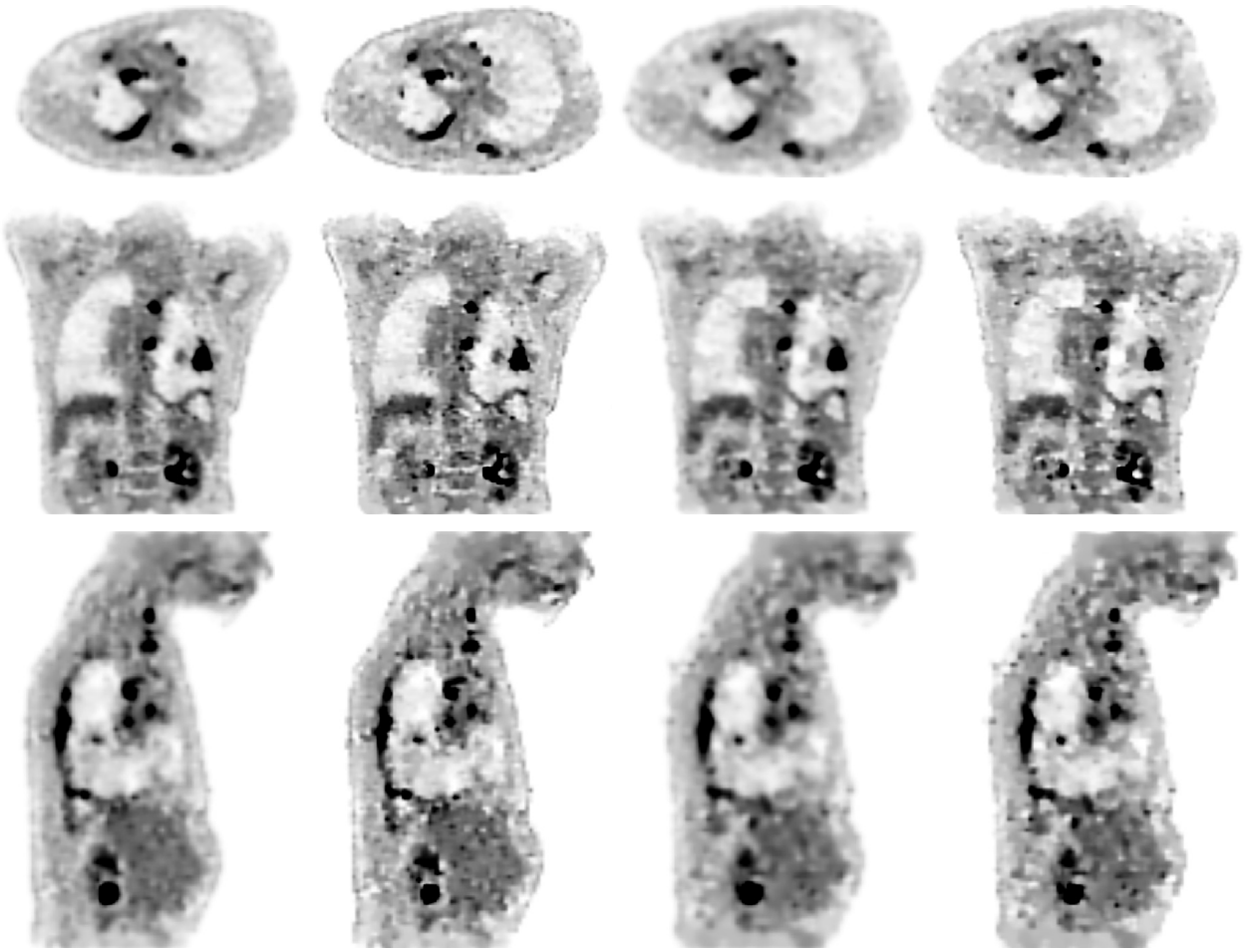
$\mathbf{u}_{\text{ref}}$ $\mathbf{u}_{\text{conv}}$, full data $\mathbf{u}_{\text{conv}}$, sparse data

Figure 5.17: Negative reference reconstructions (column 1), and negative final-convergent reconstructions $\mathbf{u}_{\text{conv}}$, within two coronal slices (rows 1 and 2) of the human subject obtained from full (column 2) and sparse (column 3) data with program DKL-fTV. Display window: [-1600, 0].



$\mathbf{u}_{\text{ref}}$ $\mathbf{u}_{\text{conv}}$, full data $\mathbf{u}_{\text{conv}}$, sparse data

Figure 5.18: Negative reference reconstructions (column 1), and negative final-convergent reconstructions $\mathbf{u}_{\text{conv}}$, within two sagittal slices (rows 1 and 2) of the human subject obtained from full (column 2) and sparse (column 3) data with program DKL-fTV. Display window: [-1600, 0].



$\mathbf{u}_{\text{conv}}$, full data $\mathbf{f}_{\text{conv}}$, full data $\mathbf{u}_{\text{conv}}$, sparse data $\mathbf{f}_{\text{conv}}$, sparse data

Figure 5.19: Negative final-convergent reconstructions $\mathbf{u}_{\text{conv}}$ and $\mathbf{f}_{\text{conv}}$ within a transverse (row 1), coronal (row 2), and sagittal (row 3) slices of the human subject obtained from full (columns 1 & 2) and sparse (columns 3 & 4) data with program DKL-fTV. Display window: [-1600, 0].

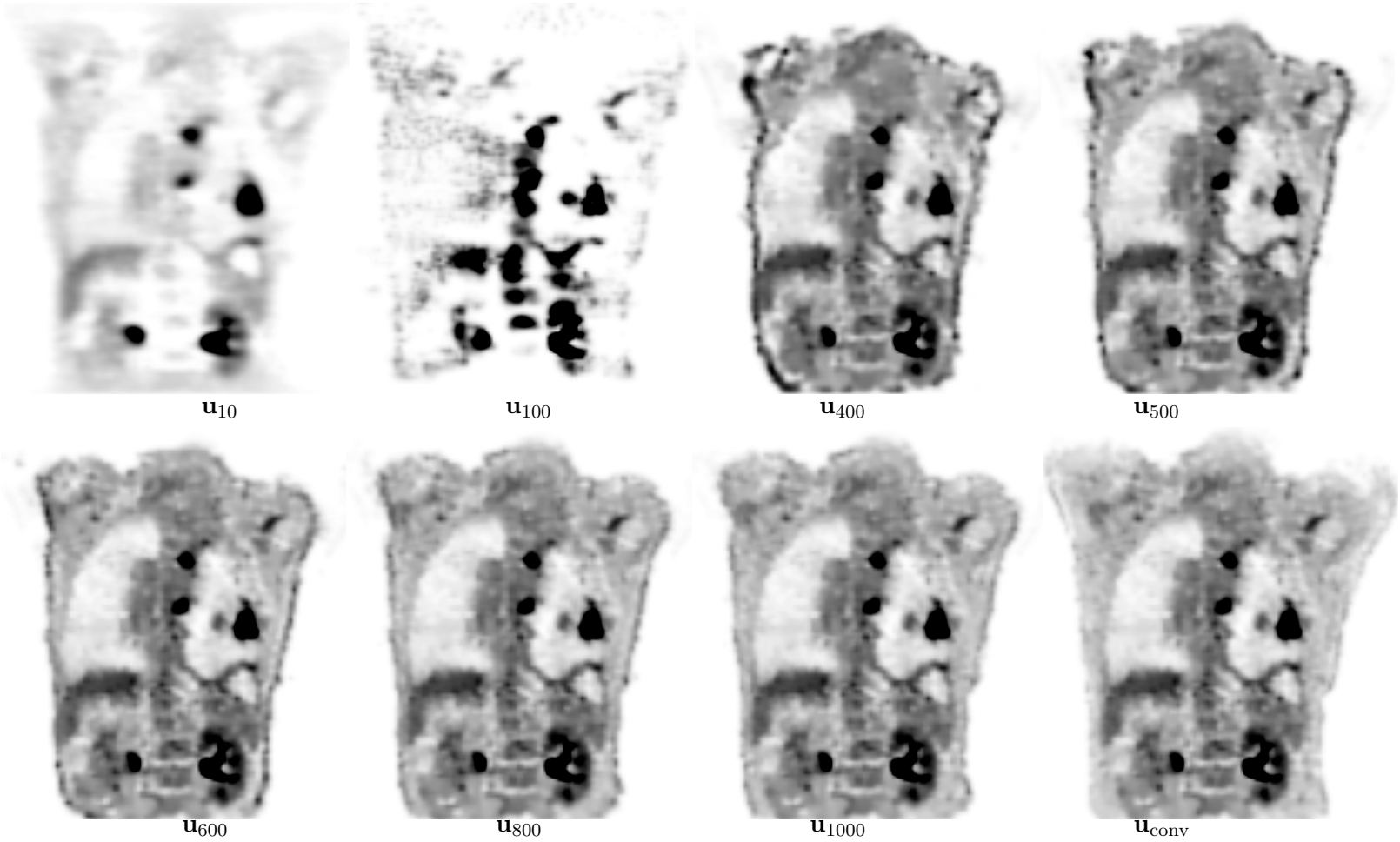


Figure 5.20: Negative reconstructions $\mathbf{u}_n$ at iteration n =10, 100, 400, 500, 600, 800, and 1000, along with negative final-convergent reconstruction $\mathbf{u}_{\text{conv}}$, within a coronal slice of the human subject obtained from full data of all bed positions with program DKL-fTV. Display window: [-1600, 0].

5.5 Discussions

In this chapter, we have performed an investigation in which TV-constrained optimization techniques were applied to the reconstruction of PET images. In particular, we have studied whether they can be exploited for enabling a PET configuration with sparsely populated detectors by extracting a sparse data from full data. We point out that, in a realistic sparse-scan configuration, the sensitivity of detector blocks and scatter/random behavior may be considerably different than that of the simulated sparse-scan configuration considered. The work is not intended to demonstrate whether the optimization-based reconstruction investigated is superior to existing reconstruction techniques, because such a demonstration, which is beyond the scope of the work, can meaningfully be carried out only in carefully designed evaluation studies in which imaging tasks and the associated utility metrics are clearly defined. Instead, the goal of the work is to investigate technical characteristics in the application of TV-constrained optimization reconstructions to PET imaging.

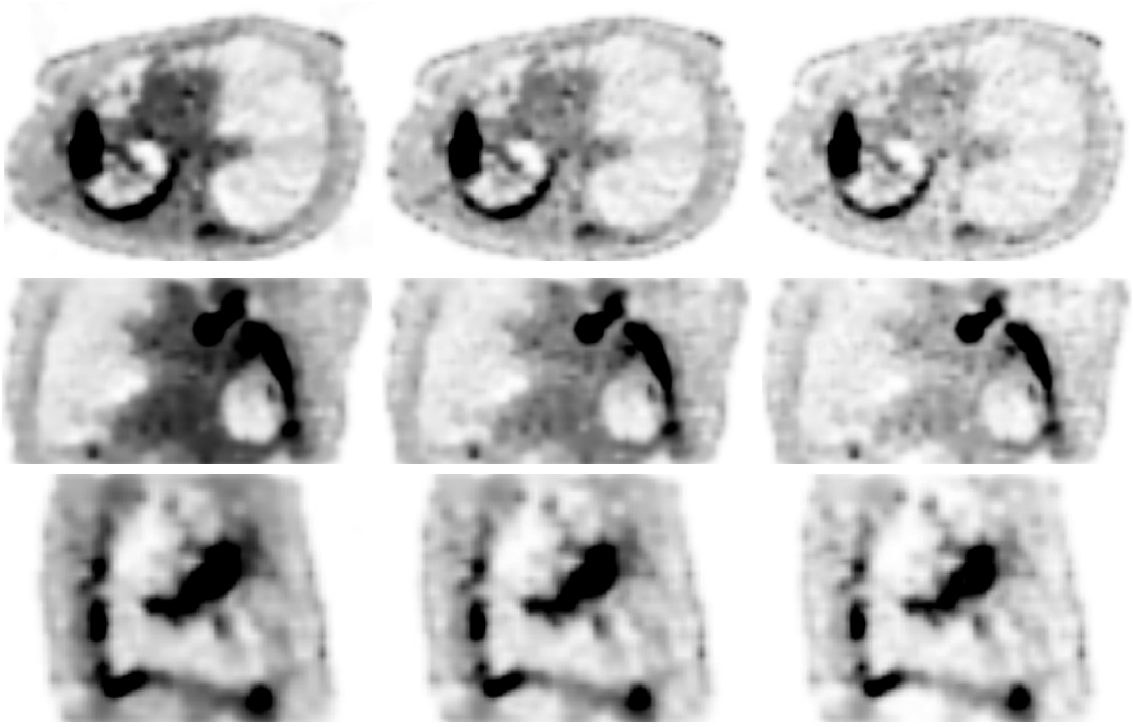
A central focus of this work is to clarify parameters involved in, and to demonstrate their impact on, optimization-based reconstructions. While it is natural to appreciate the impact of parameters in a given optimization program, the form of the optimization program itself is also a parameter that can significantly affect PET-image reconstruction. This is the motivation of the study carried out on optimization programs with different data divergences.

In addition to the program form, numerous parameters used for specification of a program can have a significant impact on the final reconstruction. Numerous parameters were identified explicitly and investigated, as discussed in Sec. 4.2.7. Among the parameters, image-TV-constraint parameter t_1 remains a leading factor affecting reconstruction properties. Therefore, we have specifically investigated the effect of parameter t_1 in real-data studies, and on devising the selection of t_1 specifically tailored to a set of given study conditions, involving different scanning configurations, imaged objects, and data properties. Although some evidence was given in the work that justifies the selection of other parameters, including voxel size, matrix $\mathcal{G}$, scatter/random events, and convergence conditions, it

remains interesting and necessary to investigate more thoroughly their optimal selections for specific applications.

For example, scatter component $\mathbf{g}_s$ with attenuation-effect corrected is considered a program parameter that is estimated from experimental measurements in the work. Clearly, the degree of estimation variability of $\mathbf{g}_s$ can impact the reconstruction. In an attempt to inspect the impact, we repeated the DKL-fTV reconstruction from full data of the human subject at bed position 3. Using $\mathbf{g}_s$ obtained experimentally at bed position 3, we created hypothetically "under"- and "over"-estimated scatter events by scaling $\mathbf{g}_s$ with a factor ranging from 0 to 2. In Fig. 5.21, we show convergent reconstructions from the data corrected with scatter events 0, $\mathbf{g}_s$, and $1.8\mathbf{g}_s$, respectively. The visual difference among these images indicates the effect of different estimates of the scatter events on the reconstruction. The study can conversely be exploited for fine-tuning the estimate of the scatter events for yielding desired reconstructions.

We have carried out image reconstructions in different studies involving objects with considerably different activity-uptake distributions of practical relevance and data with different quality/quantity conditions of interest. The results show that the reconstruction based upon program DKL-fTV appears to be robust for the different activity uptakes and data sets under consideration. Moreover, a study was conducted for image reconstruction from data collected with a PET configuration containing only half of the detectors in a digital prototype PET/CT scanner. The study reveals, on one hand, the robustness of DKL-fTV reconstruction with respect to significantly different data conditions, and, on the other hand, its potential for enabling PET-scanner configuration with a sparse detector distribution. It is of merit to investigate possible enabling of additional designs of PET configurations, which are of application-, workflow-, and/or cost-reduction-interest, by use of optimization-based reconstructions. Such an investigation can also provide a further assessment of the robustness of optimization-based reconstructions under additional, different imaging and data conditions.



$\hat{\mathbf{u}}^*$, $0.0\mathbf{g}_s$ $\hat{\mathbf{u}}^*$, $\mathbf{g}_s$ $\hat{\mathbf{u}}^*$, $1.8\mathbf{g}_s$

Figure 5.21: Negative convergent reconstructions $\hat{\mathbf{u}}^*$ within a transverse (row 1), coronal (row 2), and sagittal (row 3) slices of the human subject obtained with program DKL-fTV from full data at bed position 3 corrected for scatter events by use of $0.0\mathbf{g}_s$ (column 1), $\mathbf{g}_s$, and $1.8\mathbf{g}_s$, respectively. Display window: [-1600, 0].

In this study, the investigation was based upon a data model in which list-mode and/or TOF information is not considered because an element of the system matrix is designed as the intersection of an LOR with a voxel in the image array. Therefore, the data model and optimization programs considered cannot be applied to list-mode and/or TOF-PET data. We are currently developing optimization-based reconstruction to incorporate list-mode and TOF information, and expect to report the results elsewhere in the future. It would be of high interest to exploit list-mode and TOF information for possibly enabling the design, investigation, and implementation of innovative PET configurations of workflow-, cost-reduction-, and/or application-significance.

5.6 Appendix

We display in Fig. 5.22 a transverse slice of the IEC phantom used widely for evaluating PET-image quality. Six spheres labeled as s_1, s_2, s_3, s_4, s_5, and s_6 are embedded within the background and are with diameters of 10, 13, 17, 22, 28, and 37 mm, respectively. Spheres s_1-s_4, referred to hot spheres, have an identical concentration level of positron emitters, which is 4 times of that in the background, whereas spheres s_5 and s_6 referred to as cold spheres contain no positron emitters. The dark circle at the center displays a transverse cross section of the cylinder containing zero activity in the phantom. Additionally, 12 identical circular background ROIs of diameter 37 mm are drawn in the slice, as shown in Fig. 5.22, and also in each of its four nearest neighboring slices, thus amounting to a total of 60 background ROIs. Within each ROI, 6 sub-ROIs of sizes corresponding to those of the hot and cold spheres are also drawn, as indicated in the top ROI, and thus a total of $T = 60$ sub-ROIs for each of the sphere sizes is obtained, which are used below for calculating the average background activity for the corresponding sphere.

Using $C_{B,t,k}$ to denote the average background activity within sub-ROI t of the size of sphere k, we define an average background activity corresponding to sphere k as $C_{B,k} = \sum_{t=1}^{T} C_{B,t,k}$. With this, percent contrasts $Q_{H,j}$ and $Q_{C,i}$ for hot sphere j, where $j = 1, 2, 3,$

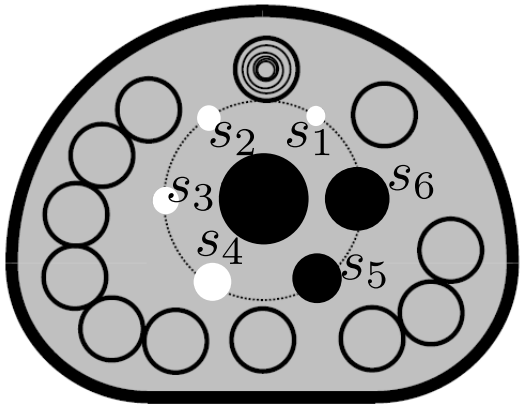


Figure 5.22: A transverse slice of the IEC phantom containing six spheres, labeled as s_1, s_2, s_3, s_4, s_5, and s_6. Hot spheres s_1-s_4 hold an identical level of activity concentration, 4 times of that in the background, whereas cold spheres s_5 and s_6 contain no activity. The dark circle at the center of the phantom displays a cross section of the central cylinder containing zero activity. Also, 12 identical circular background ROIs of diameter 37 mm are drawn. Within each ROI, 6 concentric sub-ROIs of sizes of the 6 spheres are also drawn, as indicated in the top ROI.

and 4, and cold sphere i, where $i = 5$ and 6, and background variability N_k for sphere k, where $k = 1, 2, 3, 4, 5$ and 6, defined in NEMA NU 2-2012 [191] are calculated as

$$\begin{aligned} Q_{H,j} &= \frac{C_{H,j} - C_{B,j}}{a_H - a_B} \times 100\%, \\ Q_{C,i} &= (1 - \frac{C_{C,i}}{C_{B,i}}) \times 100\%, \end{aligned} \tag{5.4}$$

$$N_k = \frac{\sqrt{\sum_{t=1}^{T}(C_{B,t,k} - C_{B,k})^2/(T-1)}}{C_{B,k}} \times 100\%, \tag{5.5}$$

where a_H and a_B denote truth activity concentrations in a hot sphere and background, $a_H = 4 \times a_B$, and $C_{H,j}$ and $C_{C,i}$ are the average activities within hot sphere j and cold sphere i.

We define a metric, which takes into account the trade-off between contrast and background noise, as

$$\text{QNR} = \frac{Q_{H,2}}{N_2}\frac{Q_{C,5}}{N_5}, \tag{5.6}$$

where $Q_{H,2}$ and $Q_{C,5}$ denote percent contrasts for hot sphere s_2 of diameter 13 mm and for cold sphere s_5 of diameter 28 mm, and N_2 and N_5 percent background variabilities corresponding to the two spheres.

CHAPTER 6

OPTIMIZATION-BASED ALGORITHM WITH VARIABLE SPATIAL RESOLUTION

6.1 Introduction

In medical imaging, including CBCT imaging applications, one is interested generally in detailed information only within a particular ROI and rough knowledge outside the ROI. For example, in image-guided-surgery (IGS) and image-guided-radiation-therapy (IGRT), information about an ROI containing the tumor or lesion is of clinical significance, while some reference knowledge about the region outside the ROI could suffice. Therefore, one may use an image array with variable resolution in which the ROI and the rest of the support of the imaged subject are represented with fine and coarse voxels, respectively. In this chapter and the next, we investigate optimization-based image reconstruction with variable resolution in CBCT applications.

Attempts to develop iterative reconstruction with variable resolution exist [192, 193]. However, the attempts generally involve multiple analytic and iterative reconstructions as well as a considerable number of parameters; their optimization and generalization for different practical problems remain unclear. Conversely, our proposed approach yields a reconstruction directly from data via solution of a single optimization problem, thus reducing the steps and parameters involved in the existing iterative approaches to the problem, and allowing its adoption to various practical applications. Therefore, in addition to its potential clinical significance, the approach proposed is also of technical significance. Among optimization-based algorithms developed recently, algorithms exploiting image-sparsity properties have attracted considerable attention as they have shown some potential to yield reconstructions of practical utility. The adaptive-steepest-descent (ASD)-projection-onto-convex-set (POCS) algorithm is one such algorithm, and it has been applied to numerous CBCT imaging applications. In this work, we focus on tailoring the ASD-POCS algorithm

to reconstructing images with variable resolution.

The chapter is organized as below. Following the introduction in Sec. 6.1, we describe in Sec. 6.2 the imaging model, optimization program, and ASD-POCS algorithm for image reconstruction with variable resolution, in both 2-dimension and 3-dimension scenarios. In Sec. 6.3 we carry out inverse crime studies for validating the proposed algorithms. Remarks of this chapter is given in Sec. 6.4.

6.2 Development of optimization-based image reconstruction with variable resolution

In an optimization-based reconstruction, three key components need to be developed: (1) a data model that relates the data and the underlying image of interest, (2) an optimization program that employs a data model and information about data and image to design reconstructions, and (3) algorithms that can achieve the designed reconstructions through solving the optimization program. Furthermore, an optimization program and its associated algorithms generally involve a number of parameters such as the number of voxels which can considerably affect the reconstruction quality. In the following, we will describe and discuss each of the components and the impact of program/algorithm parameters.

6.2.1 Data model for optimization-based reconstruction with variable resolution

In this part, based upon a data model for an image array with uniform resolution, we will investigate and develop a data model for an image array with variable resolution.

Data model and system matrix The extension from data model to system matrix is relatively straightforward, involving the design of schemes for calculating forward and backward projection of an X-ray through an image array with spatially variable resolution. Mathe-

matically, the data model [165] can be written as

$$\mathbf{g}_0 = \mathcal{H}\mathbf{f}, \tag{6.1}$$

where $\mathcal{H}$ denotes a system matrix of size $M \times N$ that relates the model data $\mathbf{g}_0$ and the image vector $\mathbf{f}$. The model data $\mathbf{g}_0$ contain M measurements, which is determined by the number of detector bins N_{bin} and the number of views N_{view}. It is important to notice that the properties of the system matrix depend not only on the total number of measurements M, but also on the number of voxels N of the image array.

6.2.2 Uniform- and variable-resolution image arrays

In a practical optimization-based reconstruction, an image array consisting of voxels of identical sizes is generally used. We refer to such an array as *a uniform-resolution array*. In the work, instead, we develop optimization-based reconstructions that use an image array formed by voxels of different sizes, and thus refer to it as *a variable-resolution array*. In Fig. 6.1, we display a 2D version of a variable-resolution array, which includes a rectangle-shaped ROI consisting of pixels of size a, and the region surrounding the ROI composing of pixels of size b. Obviously, when $a = b$, a variable-resolution array becomes a uniform-resolution array. However, if $a \neq b$, different levels of resolution are obtained for the ROI and its surrounding region. In particular, when $a < b$, the array thus includes an ROI with resolution higher than that in the region surrounding the ROI. When a variable-resolution array is used, the calculation method of the system matrix remains largely unchanged except that care should be taken for calculations involving voxels along boundaries between an ROI and its surrounding region. The same concept can be applied to a 3D image array and an illustration of 3D image array of variable resolution is shown in Fig. 6.2. In the below, we define the ratio of b to a as w, that is, $w = b/a$.

Calculation of system matrix For given data and image vectors, the data model is

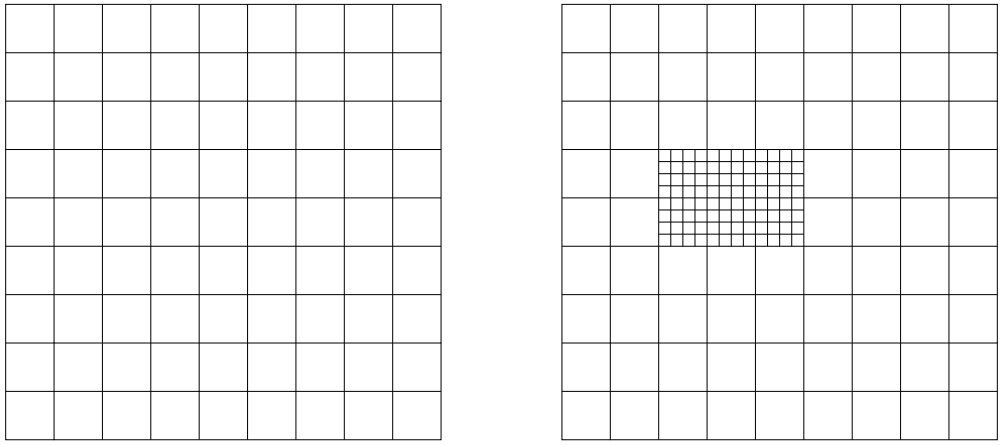


Figure 6.1: 2D Image arrays with uniform resolution (left) and variable resolution (right). The variable-resolution array includes a rectangle-shaped ROI consisting of pixels of size a, and a region surrounding the ROI composing of pixels of size b. When $a < b$, different levels of resolution are obtained for the ROI and its surrounding region.

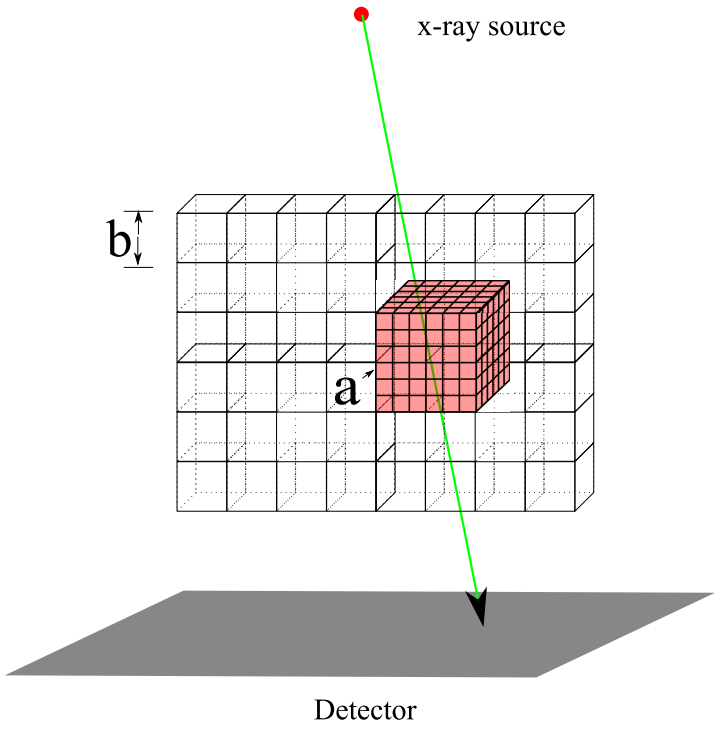


Figure 6.2: Illustration of a 3D image with variable resolution. Voxels within ROI have size of a and those outside the ROI have size of b.

determined completely by system matrix $\mathcal{H}$ [165, 166, 194], and different data models can be designed by the different ways of calculating $\mathcal{H}$. In the work, we will investigate a system matrix whose elements are determined by the intersection segments of a given X-ray.

Without loss of generality, the calculation of the elements of system matrix $\mathcal{H}$ based on an image of variable resolution can follow similar procedures as that based on an image of uniform resolution in Sec. 2.2.1. It is worth noting that the calculation involves the intersection length of ray j within a pixel or a row, which is proportional to the pixel size. Therefore, the elements of the system matrix based upon an image of variable resolution should be weighted differently due to the different pixel sizes inside and outside the ROI. Moreover, the calculation of the elements of the system matrix for the boundary pixels becomes non-trivial, and should be treated differently. Details of the system matrix calculation will be discussed in Sec. 6.2.5.

6.2.3 Optimization programs for reconstructions with variable resolution

For a selected data model, we investigate and design optimization programs for image reconstruction with variable resolution in CBCT.

Optimization programs Using data and a data model, one can form data metrics to be used in an optimization program. In addition, image metrics that exploit information about the underlying image can also be devised for use in the optimization programs. Optimization programs can be divided into two broad classes: constrained and unconstrained optimization programs [153, 154]. In this chapter, we will focus primarily on constrained optimization programs because they allow natural incorporation of physical constraints into the programs. A constrained optimization program can be written in a general form as

$$\mathbf{f}^{\star} = \text{argmin}\Phi(\mathbf{g}, \mathbf{f}) \qquad \text{s.t.} \qquad \Psi(\mathbf{g}, \mathbf{f}), \tag{6.2}$$

where $\Phi(\mathbf{g}, \mathbf{f})$ denotes the objective function to be minimized under the constraint $\Psi(\mathbf{g}, \mathbf{f})$.

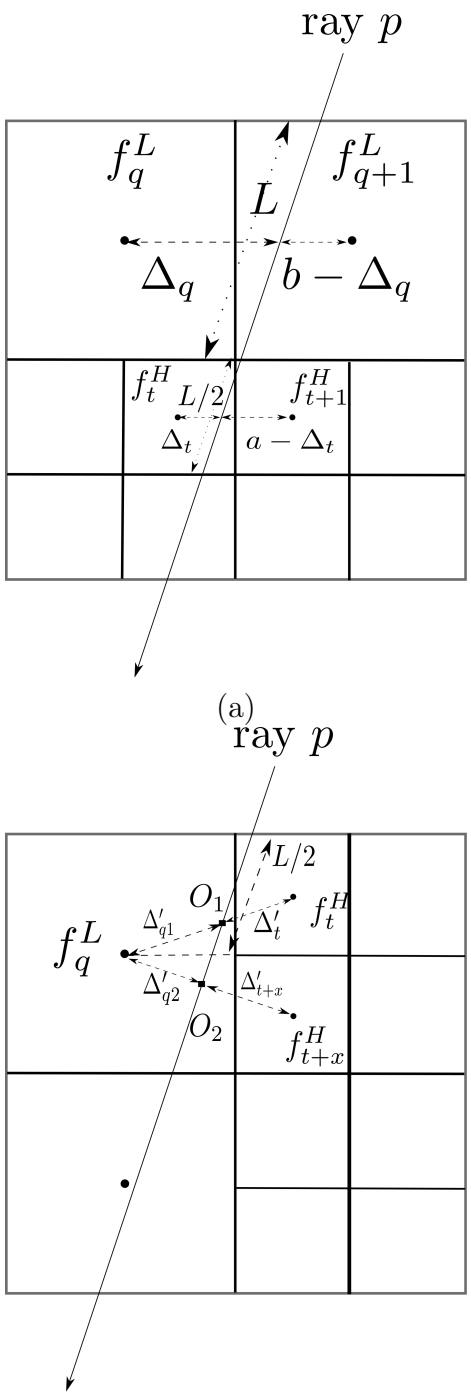


Figure 6.3: Ray p through an image with variable resolution: (a) linear interpolation is applied between high-resolution (or low-resolution) pixels only; (b) linear interpolation is applied between high-resolution and low-resolution pixels.

Both objective and constraint can be dependent on data, data model, image metrics, and additional parameters, which will be discussed in the two example constrained optimization programs below.

Constrained TV-minimization program The constrained TV-minimization program has been studied extensively in CBCT and MRI in the last several years, and yielded promising results for CBCT reconstructions with uniform resolution. This program can be expressed mathematically as [153]

$$\mathbf{f}^{\star} = \mathsf{argmin} \parallel \mathbf{f} \parallel_{TV} \qquad s.t. \qquad D_{\ell_2}(\mathbf{g}_m, \mathbf{g}) \leq \epsilon \quad and \quad \mathbf{f} \geq 0, \tag{6.3}$$

where $\mathbf{g}_m = \mathcal{H}\mathbf{f}_{\text{true}}$, $\mathbf{g} = \mathcal{H}\mathbf{f}$, $D_{\ell_2}(\mathbf{g}_m, \mathbf{g}) = \parallel \mathbf{g}_m - \mathbf{g} \parallel_2$ denotes the ℓ_2-norm of the difference between data and data model, and ϵ a pre-selected, positive parameter accounting for the minimum inconsistency between data $\mathbf{g}$ and data model $\mathcal{H}\mathbf{f}$. In the work, we will extend the program to reconstructions with variable resolution. It is known that, for given data, the complete specification of the program for reconstruction with uniform resolution involves several key parameters: the voxel size, the total number of voxels, the number of measurements, the system matrix $\mathcal{H}$, and the parameter ϵ. In reconstructions with variable resolution, however, additional parameters will be involved in specifying the program: the sizes and numbers of the two types of voxels. Therefore, the extension of the program to reconstruction with variable resolution will be be accomplished by incorporating appropriately these new parameters into the calculation of system matrix $\mathcal{H}$ and image TV.

6.2.4 Algorithms for optimization-based reconstruction with variable resolution

When an optimization program is determined, it in effect specifies a set of feasible solutions. Algorithms can be developed for achieving the solution set through solving the optimization program. In this part of the research, we will develop algorithms that reconstruct images

with variable resolution through solving the corresponding optimization programs. In practice, only iterative algorithms can be developed for optimization programs considered in the project.

Calculation of data and image metrics Several basic components need to be developed for use in algorithm development, and they have been investigated and implemented for reconstructions with uniform resolution. The focus of this part of the research is to extend these components to accommodate reconstructions with variable resolution. The first component is the calculation of forward and backward projections based upon a given data model. A generalization of the calculation for an image array with uniform resolution will be carried out for an image array with variable resolution. The generalization is straightforward except for computing the part when a ray passes through the boundary between voxels with different sizes. In the design of optimization programs, in addition to regular data divergence, some kind of difference between the derivatives of the data and the data model will be used. In this case, we will carefully design the calculation of the derivative of the data model so that it can be carried out for an image array with variable resolution. We will also consider a number of image metrics for use in the design of the optimization program. A leading image metric to be considered is the image TV. The generalization of the image TV to an image array with variable resolution is, again, straightforward when the voxel sizes are taken appropriately into the calculation. When derivatives of an image TV are calculated during its minimization, however, extra care needs to be taken for dealing with the boundaries between regions with voxels of different sizes so that they result in no artifacts in reconstructed images. We have performed a preliminary study involving 2D reconstruction with variable resolution, and learned that an adequate design of numerical calculation of image TV derivatives is critical in eliminating reconstruction artifacts.

ASD-POCS algorithms It has been demonstrated in extensive studies [153, 156, 165, 195, 196] that the ASD-POCS algorithm can be effective in solving constrained optimization programs such as those in Eqs. (6.3) for reconstructions with uniform resolution. A primary

focus of the work will be on the extension of the ASD-POCS algorithm to reconstruct images with variable resolution. In this algorithm, POCS and gradient descent (GD) algorithms will be used adaptively for reducing data divergence and image TV. Other algorithms such as conjugate gradient will also be used for robustly lowering data divergence in the vicinity of a feasible solution set designed by the optimization program. As described above, the basic components developed specifically for reducing data divergence and image TV based upon an image array with variable resolution will be incorporated into the ASD-POCS algorithm. An emphasis will be placed on the investigation of an adequate design of a balanced reduction of data divergence and image TV, thus enhancing the algorithm robustness and efficiency. In particular, a scheme will be investigated that specifically incorporates the boundary information between different types of voxel sizes in the determination of the step sizes and relaxation parameters at each iteration of the ASD-POCS algorithm. Convergence conditions of the original ASD-POCS will also be generalized to accommodate reconstructions with variable resolution. A relationship between data divergence and image TV, as a function of iteration numbers, will be used for monitoring and assessing algorithm convergence. An additional convergence condition

$$c_{\alpha} = \frac{\mathbf{d}_{\mathrm{TV}} \cdot \mathbf{d}_{\mathrm{D}}}{|\mathbf{d}_{\mathrm{TV}}||\mathbf{d}_{\mathrm{D}}|}, \tag{6.4}$$

which accounts for the direction between the data divergence gradient descent $\mathbf{d}_{\mathrm{TV}}$ and image TV gradient descent $\mathbf{d}_{\mathrm{D}}$, has been shown to be a very useful convergence condition. In this part of the work, it will be generalized to monitor and evaluate the convergence of reconstructions with variable resolution.

6.2.5 Forward Projection on an image with variable resolution

Without loss of generality, we consider a 2D image array with variable resolution, and explain how to modify the forward projection procedure for yielding such a variable-resolution image.

Forward projection for 3D images with variable resolution can be adapted in the same manner.

In this work, we employ a 2D variable-resolution image consisting of two sets of pixels: one set, labeled as $\mathbf{f}^H$, is within the ROI, with pixels of size a; and the other, labeled as $\mathbf{f}^L$, is outside the ROI, with pixels of size b.

Forward projection in high- and low-resolution regions In Fig. 6.3(a), ray p travels through an image with variable resolution. To make the calculation below simple, we define $b = 2a$, and $w = b/a = 2$. In general, the element of system matrix $\mathcal{H}$ can be considered as the contribution of pixels to a rays, where we use ray-driven model and linear interpolation method. In order to differentiate the index for the pixels in the high- and low-resolution regions, respectively, we index the former with symbol q and the later with t. Thus, an pixel in the high resolution region can be labeled as f_t^H, and one in the low-resolution region f_q^L.

The first row in Fig. 6.3(a) that ray p travels through contains low-resolution pixels only, and the second row contains high-resolution pixels only. In this scenario, the elements of $\mathcal{H}$ corresponding to ray p can be calculated in the high- and low-resolution regions independently. For example, in Fig.6.3(a), the line integral of ray p in the first row (low-resolution region) is calculated as the length of ray p within the row, L, multiplied by a value interpolated from pixel values f_q^L and f_{q+1}^L:

$$f_q^L \frac{b - \Delta_p}{b} L + f_{q+1}^L \frac{\Delta_q}{b} L, \tag{6.5}$$

where b is the size of the low-resolution pixels, and Δ_p is the distance from pixel f_q^L to the intersection point of ray j and the line connecting f_q^L and f_{q+1}^L. Thus, we can write the two elements of system matrix $\mathcal{H}$, corresponding to two low-resolution pixels, as:

$$\begin{aligned} H_{p,q} &= \frac{b - \Delta_q}{b} L \\ H_{p,q+1} &= \frac{\Delta_q}{b} L \end{aligned} \tag{6.6}$$

Similarly, the line integral of ray p in the second row (low-resolution region) is calculated as the length of ray p within the row, L/w, multiplied by a value interpolated from pixel values f_t^H and f_{t+1}^H:

$$f_t^H \frac{a-\Delta_t}{a}\frac{L}{2} + f_{t+1}^H \frac{\Delta_t}{a}\frac{L}{2}, \tag{6.7}$$

where a is the size of the high-resolution pixels, and Δ_t is the distance from pixel f_t^H to the intersection point of ray j and the line connecting f_t^H and f_{t+1}^H. We then can write the two elements of system matrix $\mathcal{H}$, corresponding to two high-resolution pixels, as:

$$\begin{aligned} H_{p,t} &= \frac{a-\Delta_t}{a}\frac{L}{2} \\ H_{p,t+1} &= \frac{\Delta_t}{a}\frac{L}{2} \end{aligned} \tag{6.8}$$

It is worth noting that the value of system matrix elements are proportional to the travel length of ray p within the row of pixels, which is proportional to the pixel size.

Forward projection at the boundaries of high- and low-resolution regions If an image row consists of both high- and low-resolution pixels, and when ray p travels though this row, part of the segment is in the low-resolution region, and the other in the high-resolution region, as illustrated in Fig. 6.3(b). Again, we assume $b = 2a$, and $w = b/a = 2$. In Fig. 6.3(b), high-resolution pixels, f_t^H and f_{t+x}^H, sit on the right hand side of a low-resolution pixel, f_q^L, and ray p travels between pixels f_q^L and f_t^H, f_{t+x}^H. For calculating the system matrix $\mathcal{H}$, a ray-driven model and linear interpolation method are used.

We draw a line connecting pixels f_q^L and f_t^H in Fig. 6.3(b) which intersects with ray p at point O_1, then we connect pixels f_q^L and f_{t+x}^H and the line intersects with ray p at point O_2. We label the distance between pixel f_q^L to O_1 as Δ'_{q1}, that between pixel f_q^L to O_2 as Δ'_{q2}, that between pixel f_t^H to O_1 as Δ'_t, and that between f_{t+x}^H to O_2 as Δ'_{t+x}. In this scenario, the line integral of ray p in the first row of low-resolution region is calculated as the summation of (1) the length of ray p within the first row of high-resolution region, $L/2$, multiplied by a value interpolated from pixel values f_q^L and f_t^H, and (2) the length of ray

p within the second row of high-resolution region, $L/2$, multiplied by a value interpolated from pixel values f_q^L and f_{t+x}^H, which can be written as:

$$\begin{aligned} & f_q^L \frac{\Delta'_t}{\Delta'_{q1}+\Delta'_t}\frac{L}{2} + f_t^H \frac{\Delta'_{q1}}{\Delta'_{q1}+\Delta'_t}\frac{L}{2} + f_q^L \frac{\Delta'_{t+x}}{\Delta'_{q2}+\Delta'_{t+x}}\frac{L}{2} + f_{t+x}^H \frac{\Delta'_{q2}}{\Delta'_{q2}+\Delta'_{t+x}}\frac{L}{2} \\ =& f_q^L (\frac{\Delta'_t}{\Delta'_{q1}+\Delta'_t} + \frac{\Delta'_{t+x}}{\Delta'_{q2}+\Delta'_{t+x}})\frac{L}{2} + f_t^H \frac{\Delta'_{q1}}{\Delta'_{q1}+\Delta'_t}\frac{L}{2} + f_{t+x}^H \frac{\Delta'_{q2}}{\Delta'_{q2}+\Delta'_{t+x}}\frac{L}{2}. \end{aligned} \tag{6.9}$$

Thus, we can formulate three elements of the system matrix $\mathcal{H}$ as:

$$\begin{aligned} H_{p,q} &= (\frac{\Delta'_t}{\Delta'_{q1}+\Delta'_t} + \frac{\Delta'_{t+x}}{\Delta'_{q2}+\Delta'_{t+x}})\frac{L}{2} \\ H_{p,t} &= \frac{\Delta'_{q1}}{\Delta'_{q1}+\Delta'_t}\frac{L}{2} \\ H_{p,t+x} &= \frac{\Delta'_{q2}}{\Delta'_{q2}+\Delta'_{t+x}}\frac{L}{2}. \end{aligned} \tag{6.10}$$

In the same manner, we can derive all the element values in the system matrix $\mathcal{H}$.

6.2.6 Total Variation of an image with variable resolution

Again, for the modification of image total variation for a variable-resolution image, we consider a 2D case as an example. Firstly, we consider a continuous 2D image f and its TV can be formulated as:

$$f_{TV} = \iint \sqrt{(\frac{\partial f(x,y)}{\partial x})^2 + (\frac{\partial f(x,y)}{\partial y})^2}\,\mathrm{d}x\,\mathrm{d}y. \tag{6.11}$$

A discrete image can be considered as the expansion of the continuous image in a finite set of basis, which are generally square pixels.

Then we consider an image on uniform grids with pixel sizes of Δx and Δy, which are in the x and y dimensions respectively. TV of such a discrete image can be formulated as:

$$\| \mathbf{f} \|_{TV} = \sum_{i,j} \sqrt{(\frac{f_{i,j}-f_{i-1,j}}{\Delta x})^2 + \frac{f_{i,j}-f_{i,j-1}}{\Delta y})^2}. \tag{6.12}$$

Image of uniform resolution If the pixels are isotropic and we assume $\Delta x = \Delta y = 1$, then Eq.(6.12) can be further simplified to be:

$$\| \mathbf{f} \|_{TV} = \sum_{i,j} \sqrt{(f_{i,j} - f_{i-1,j})^2 + (f_{i,j} - f_{i,j-1})^2} = \sum_{i,j} ||(\nabla f)_{i,j}||_2. \tag{6.13}$$

Let $\mathbf{f}$ represent an $N_x \times N_y$ image and $f_{i,j}$ refer to the (i,j)th voxel of $\mathbf{f}$, we introduce the differencing images $\Delta_x f$ and $\Delta_y f$.

$$(\nabla_x f)_{i,j} = \begin{cases} f_{i+1,j} - f_{i,j} & i < N_x \\ -f_{i,j} & i = N_x - 1 \end{cases} \tag{6.14}$$

$$(\nabla_y f)_{i,j} = \begin{cases} f_{i,j+1} - f_{i,j} & i < N_y \\ -f_{i,j} & i = N_y - 1 \end{cases} \tag{6.15}$$

Using the definitions above, ∇ can be written as

$$(\nabla f)_{i,j} = \begin{pmatrix} (\nabla_x f)_{i,j} \\ (\nabla_y f)_{i,j} \end{pmatrix} \tag{6.16}$$

Images of variable resolution Now we consider image-TV for an 2D image with variable resolutions. Remember that in a variable-resolution image there are at least two sets of pixels: in the high-resolution region the pixel size is $\Delta x^H \times \Delta y^H$ and in the low-resolution region the pixel size is $\Delta x^L \times \Delta y^L$. Then the TV of such a discrete image can be expressed as the expansion of Eq. (6.11) with two basis sets:

$$\begin{aligned} \| \mathbf{f} \|_{TV} = & \sum_{r,s} \sqrt{(\frac{f^H_{r,s} - f^H_{r-1,s}}{\Delta x^H})^2 + (\frac{f^H_{r,s} - f^H_{r,s-1}}{\Delta y^H})^2} \Delta x^H \Delta y^H \\ & + \sum_{k,l} \sqrt{(\frac{f^L_{k,l} - f^L_{k-1,l}}{\Delta x^L})^2 + (\frac{f^L_{k,l} - f^L_{k,l-1}}{\Delta y^L})^2} \Delta x^L \Delta y^L, \end{aligned} \tag{6.17}$$

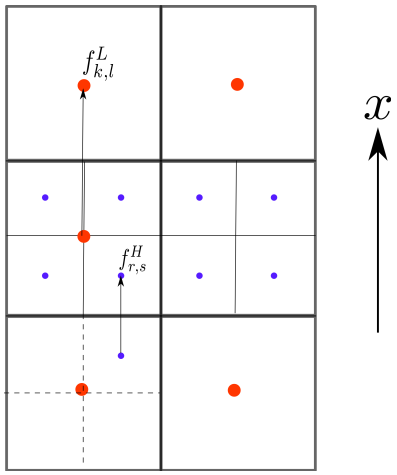


Figure 6.4: Finite differencing along x direction of a 2D image with variable resolution.

where $f^H_{r,s}$ are pixel indexes in high-resolution region and $f^L_{k,l}$ are pixel indexes in low-resolution region.

If the pixels in each basis set are isotropic and we assume $\Delta x^H = \Delta y^H = 1$ and $\Delta x^L = \Delta y^L = w$, then Eq. (6.17) can be simplified as:

$$\begin{aligned} \| \mathbf{f} \|_{TV} = & \sum_{r,s} \sqrt{(f^H_{r,s} - f^H_{r-1,s})^2 + (f^H_{r,s} - f^H_{r,s-1})^2} \\ & + w^2 \times \sum_{k,l} \sqrt{(f^L_{k,l} - f^L_{k-1,l})^2 + (f^L_{k,l} - f^L_{k,l-1})^2} \\ & = \sum_{r,s} ||(\nabla f^H)_{r,s}||_2 + w^2 \sum_{k,l} ||(\nabla f^L)_{k,l}||_2, \end{aligned} \tag{6.18}$$

where $(\nabla f^H)_{r,s}$ and $(\nabla f^L)_{k,l}$ can be calculated in the way shown in Eq. (6.16).

Boundary conditions in image-TV calculation Without loss of generality, we can take the finite differencing along x-direction, $\nabla_x f$, for example, and the calculation of the finite differencing along y-direction, $\nabla_y f$ can be derived in the same manner.

In Fig. 6.4, we show an image with variable resolution: the top and bottom two rows of pixels in the low-resolution region, and the middle of pixels are in the high-resolution region. The direction for taking the derivative along x is indicated as the direction of the arrow.

We first consider a pixel $f^H_{r,s}$ in the high-resolution region, and its finite differencing along x-direction is:

$$(\nabla_x f^H)_{r,s} = \begin{cases} f^H_{r+1,s} - f^H_{r,s} & \text{for} \quad r < N^h_x - 1 \\ f^L_{[\frac{r}{w}]+1,[\frac{s}{w}]} - f^H_{r,s} & \text{for} \quad r = N^h_x - 1, \end{cases} \tag{6.19}$$

where operator $[\cdot]$ calculates the largest integer not greater than the argument.

We then consider a pixel $f^L_{k,l}$ in the low-resolution region, and its finite differencing along x-direction is:

$$(\nabla_x f^L)_{k,l} = \begin{cases} f^L_{k+1,l} - f^L_{k,l} & \text{for} \quad r < N^L_x - 1 \\ \frac{1}{w^2} \sum\limits_{wk,wl}^{w(k+1),w(l+1))} f^H_{r,s} - f^L_{k,l} & \text{for} \quad r = N^L_x - 1. \end{cases} \tag{6.20}$$

In this section, we have derived image-TV calculation and considered scenarios for the low-resolution region, high-resolution region, and their boundaries. The calculation of image-TV is a preparation for calculating the gradient of image-TV, which will be introduced in the following part.

6.2.7 TV gradient descent of an image with variable resolution

TV gradient descent of uniform-resolution image We first consider an image with uniform resolution, and the gradient of image TV can be written as:

$$\begin{aligned} \nabla \sum_{i,j} ||(\nabla f)_{i,j}||_2 =& \nabla^T diag\{||(\nabla f)_{i,j}||_2\}^{-1} \nabla f \\ =& \nabla_x^T diag\{||(\nabla f)_{i,j}||_2\}^{-1} \nabla_x f \\ &+ \nabla_y^T diag\{||(\nabla f)_{i,j}||_2\}^{-1} \nabla_y f \end{aligned} \tag{6.21}$$

and ∇_x^T and ∇_y^T are:

$$(\nabla_x^T a)_{i,j} = \begin{cases} -a_{i,j} & \text{for} \quad i = 0 \\ a_{i-1,j} - a_{i,j} & \text{for} \quad i > 0 \end{cases} \tag{6.22}$$

$$(\nabla_y^T a)_{i,j} = \begin{cases} -a_{i,j} & \text{for} \quad j = 0 \\ a_{i,j-1} - a_{i,j} & \text{for} \quad j > 0 \end{cases} \tag{6.23}$$

TV gradient descent of variable-resolution image Similarly, the gradient of image TV with variable resolution can be written as:

$$\begin{aligned} &\nabla(\sum_{r,s} ||(\nabla f^H)_{r,s}||_2 + w^2 \sum_{k,l} ||(\nabla f^L)_{k,l}||_2) \\ =&\nabla^T diag\{||(\nabla f^H)_{r,s}||_2\}^{-1} \nabla f^H + w^2 \nabla^T diag\{||(\nabla f^L)_{k,l}||_2\}^{-1} \nabla f^L \end{aligned} \tag{6.24}$$

where $(\nabla_x^T a^H)_{r,s}$ and $(\nabla_x^T a^L)_{k,l}$ are:

$$(\nabla_x^T a^H)_{r,s} = \begin{cases} \frac{1}{w^2} a^L_{[\frac{r}{w}]+1,[\frac{s}{w}]} - a^H_{r,s} & \text{for} \quad r = 0 \\ \frac{1}{w^2} a^L_{[\frac{r}{w}]+1,[\frac{s}{w}]} + a^H_{r-1,s} - a^H_{r,s} & \text{for} \quad 0 < r < w \\ a^H_{r-1,s} - a^H_{r,s} & \text{for} \quad w \leq r < N_x^h \end{cases} \tag{6.25}$$

$$(\nabla_x^T a^L)_{k,l} = \begin{cases} -a^L_{k,l} & \text{for} \quad k = 0 \\ \sum_{e=0}^{e=w-1} a^H_{wk-1,w(l-1)+e} - a^L_{k,l} & \text{for} \quad wk = N_x^H \\ a^L_{k-1,l} - a^L_{k,l} & \text{for} \quad otherwise \end{cases} \tag{6.26}$$

6.3 Verification of the developed optimization-based reconstruction with variable resolution

It is important to verify that the algorithms and their implementation perform as they are designed to. Because optimization-based algorithms can, in general, solve their corresponding optimization programs iteratively, they and their implementation can be verified only numerically. In this section, we will design and carry out numerical verification studies of the optimization-based algorithms and their computer implementation.

Inverse-crime study In a verification study, data will be generated from a discrete image represented on an image array, and image reconstruction will be carried out by use of the same system matrix in the data generation and on the same image array. We refer to such a study as an *inverse-crime* study. In essence, in an inverse-crime study, the data and the data model are completely consistent, and the difference between the reconstruction and the truth image used for generating the data should be numerically diminished when sufficient data are used. Although inverse-crime studies do not incorporate any physical factors, they can verify the proposed algorithms, examine their robustness, and provide guidance for image reconstructions from real data. It is reasonable to assume that, if the reconstruction program and algorithm fail to perform in an inverse-crime scenario, it is unlikely that they would perform well in real-data studies where inconsistencies exist. Inverse-crime studies based on images of uniform resolution have already been carried out to verify the proposed ASD-POCS algorithm. However, images of variable resolution have not been used for image reconstruction before, and the properties of the system matrix as well as the algorithms based on images of variable resolution have barely been analyzed. In the work, we focus on inverse-crime studies based upon images of variable resolution.

Inverse-crime study design and results In the inverse-crime study, we use a numerical, cadaver CT image as a truth phantom, which is shown in the left panel of Fig. 6.5. The numerical cadaver image consists of square pixels of size 0.4 mm in the left region and

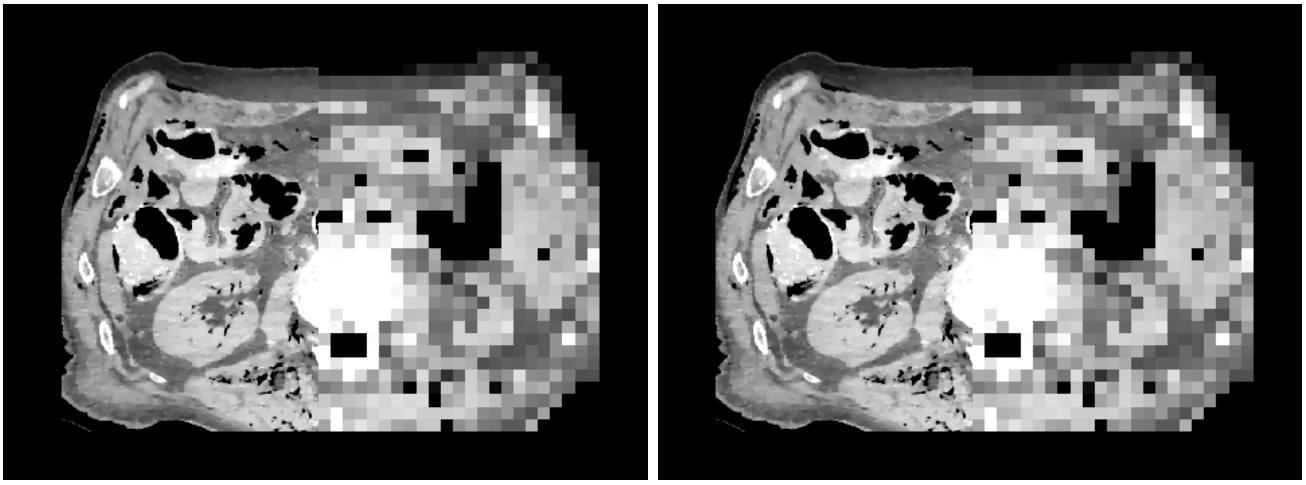


Figure 6.5: Truth images (left) and ASD-POCS reconstructions of a numerical cadaver phantom. Display window: [40, 55].

square pixels of size 4 mm in the right region. We mimick a fan-beam CT configuration with the source-to-detector distance of 150 cm, and the source-to-rotation-axis distance of 100 cm. The 2-dimensional detector consists of 800 detector element of size 0.4 mm. We generate data from the numerical cadaver image at 721 views uniformly distributed over 2π.

For given data, we verify numerically the ASD-POCS algorithm to solve the constrained TV-minimization program in Eq. (6.3). Practical convergence conditions in this study include

$$\begin{aligned} \overline{D}_{\ell_2}(\mathbf{g}_m, \mathbf{g}_n) &< 10^{-6} \\ \overline{\Delta}(\mathbf{f}_n, \mathbf{f}_{\text{true}}) &< 10^{-6} \end{aligned} \tag{6.27}$$

where $\mathbf{g}_n = \mathcal{H}\mathbf{f}_n$, $\overline{D}_{\ell_2}(\mathbf{g}_m, \mathbf{g}_n) = \frac{||\mathbf{g}_n - \mathbf{g}_m||_2}{||\mathbf{g}_m||_2}$ and $\overline{\Delta}(\mathbf{f}_n, \mathbf{f}_{\text{true}}) = \frac{||\mathbf{f}_n - \mathbf{f}_{\text{true}}||_2}{||\mathbf{f}_{\text{true}}||_2}$. We show in the right panel of Fig. 6.5 the reconstructed images, and in Fig. 6.6 the convergence conditions defined in Eq. (6.27), for the inverse-crime studies. It can be observed that the reconstructions are virtually identical to their corresponding truth images, demonstrating that the ASD-POCS (and its computer implementation) can solve the constrained TV-minimization program in Eq. (6.3).

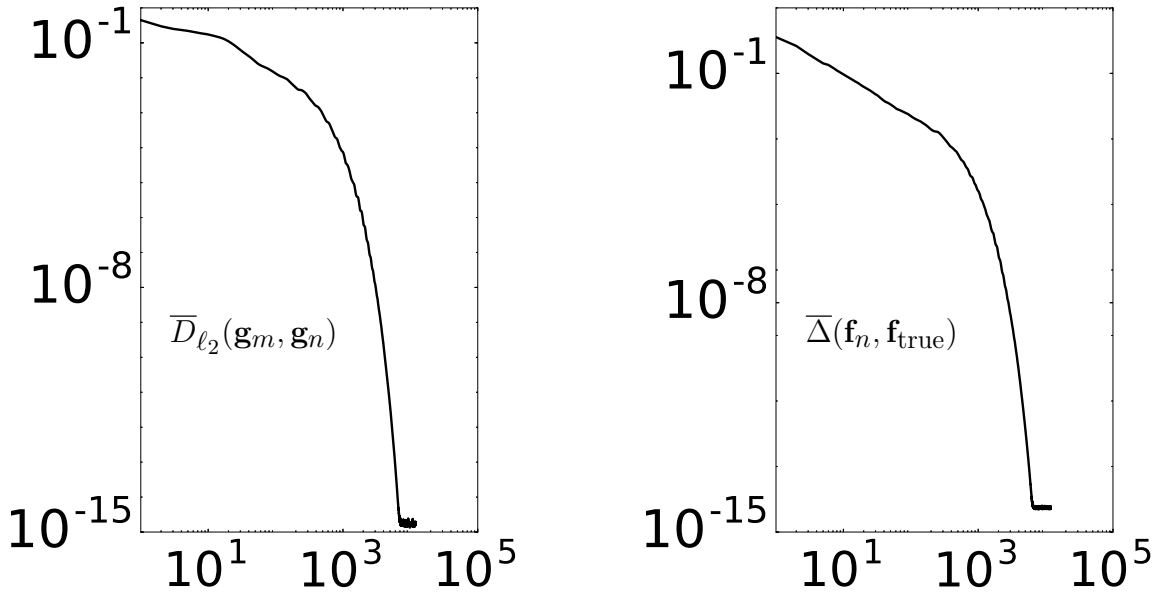


Figure 6.6: Convergence metrics $\overline{D}_{\ell_2}(\mathbf{g}_m, \mathbf{g}_n)$ (left) and $\overline{\Delta}(\mathbf{f}_n, \mathbf{f}_{\text{true}})$ (right) as functions of iteration number n.

6.4 Discussion

In this chapter, we have proposed the concept and design of optimization-based reconstruction with variable resolution. In particular, we designed the data-ℓ_2-constrained image-TV-minimization program, and investigated the ASD-POCS algorithm for reconstruction with variable resolution. The calculation of key components of the algorithm, such as the forward and backward projections, the image TV, and the gradient of image TV, have been tailored to images of variable resolution. An inverse-crime study was carried out to validate the implementation of the modified ASD-POCS algorithm. Results have demonstrated the feasibility of the ASD-POCS algorithm employing variable resolution for image reconstruction. It is worth noting that, the total number of voxels in images of variable resolution is less than that in images of uniform, high resolution (Assume that the size of voxels in the ROI of the variable-resolution images is the same as that in the uniform, high-resolution images.)

Although the focus of the work is to develop the optimization-based approach to recon-

struction with variable resolution in CBCT, the concept and techniques developed can readily be generalized to other linear-model-based imaging technologies such as magnetic resonance imaging (MRI) and nuclear medicine imaging for yielding reconstruction with variable resolution and for enabling the design and optimization of innovative scans tailored to specific applications. With the perspectives and techniques developed, existing iterative algorithms such as expectation maximization (EM) [111, 197], simultaneous iterative reconstruction technique (SIRT) [198], simultaneous algebraic reconstruction technique (SART) [89], and primal-dual algorithms can be generalized to reconstruct images with variable resolution.

CHAPTER 7

OPTIMIZATION-BASED IMAGE RECONSTRUCTION WITH VARIABLE SPATIAL RESOLUTION

7.1 Introduction

In Chapter 6, we have designed optimization-based program with images of variable resolution, and used the ASD-POCS algorithm to solve it. Such optimization-based reconstruction with variable resolution may be useful in many CBCT imaging applications because *a)* one is often interested in detailed information only within a particular ROI and rough knowledge outside the ROI; *b)* it allows the exploitation of potential advantages of optimization-based approaches over existing, analytic-based reconstruction methods; and *c)* the use of an image array with variable resolution can substantially lower computation time and computer-memory consumption of optimization-based reconstruction.

Specifically, in IGRT and IGS, repeated imaging of the subject is performed for the purpose of acquiring detailed information about the changes within the ROI containing the tumor/lesion [49, 50, 199]. Such detailed information is necessary for planning, monitoring, and assessing the treatment of the tumor/lesion within the ROI, while rough knowledge outside the ROI could suffice in providing some global positioning reference. Therefore, the proposed optimization-based reconstruction with variable resolution can be of particularly high practical merit, in terms of reducing computational burden and improving workflow in repeated imaging involved in IGRT and IGS.

In certain CBCT imaging applications, there exist special requirements which may result in non-conventional CBCT imaging configurations. For example, in repeated CBCT imaging, the amount of radiation dose delivered to the patients remains a significant concern. It is thus desirable to develop imaging approaches using reduced radiation. Imaging radiation can be lowered through the reduction of the number of projections, amounting to sparse-view CBCT imaging [156, 196, 200–202]. In addition, for practically optimal workflow in

applications of IGS [23], it is often of interest to relax the requirement on the angular range of data sampling [203]. Image reconstruction from such limited-angle-range data, however, becomes mathematically an ill-posed problem. Existing work [204, 205] suggests that when applied to sparse-view or limited-angular-range data, the image quality of reconstructions are degraded due to artifacts. The proposed optimization-based reconstruction based on arrays of variable resolution may yield high-resolution ROI with substantially reduced artifacts from sparse-view or limited-angular-range data.

In this chapter, by using the developed tools, we will carry out 2D simulation studies by using a FORBILD head phantom and modified Shepp-Logan phantom, a physical-phantom study with 2D bench-top CBCT data, animal/patient study by using 2D diagnostic-CT data, and a patient study with 3D dental CBCT data. Remarks about this study will be given in Sec. 7.6.

7.2 Simulation studies with variable resolution

In this section, we conduct simulation studies, by using analytic described phantoms, for characterizing the algorithm performance under realistic CBCT imaging conditions. We focus in the characterization study on image reconstruction from data collected by mimicking CBCT imaging configurations with a modified Shepp-Logan phantom and a FORBILD head phantom. The Shepp-Logan phantom is chosen because it is used widely for testing algorithm development in the CT field. In an attempt to increase its complexity degree, we add a set of fine structures of various shapes in the ROI, as shown in Fig. 7.1. The FORBILD head phantom is the second phantom to be used in the study, which contains numerous structures with challenging spatial and contrast resolution. In particular, a region simulating the inner ear volume is included in the phantom, as shown in Fig. 7.3, and the fine geometric shapes of varying sizes in the region poses a significant challenge to the algorithms.

A particular emphasis is placed on the adaptation and tailoring of the proposed algorithms to simulation-data reconstruction through careful determination of program and

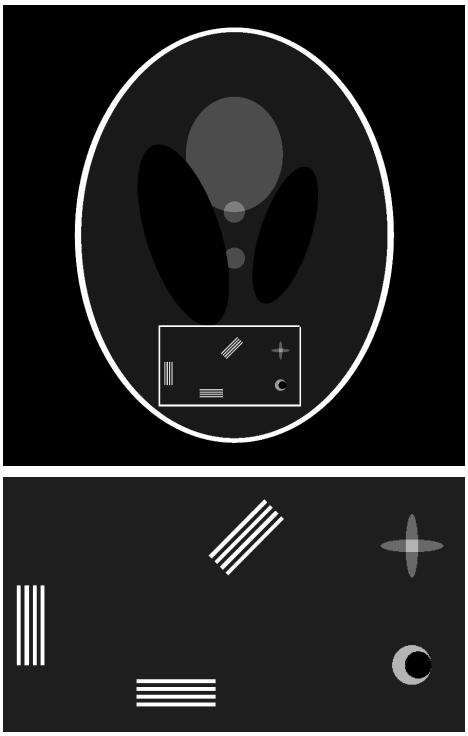


Figure 7.1: A modified Shepp-Logan phantom (top) discretized on a uniform-resolution array with pixel size of 0.0195 cm, with an ROI enclosed by a white box; A zoomed-in view of the ROI image (bottom). Display window: [0.6, 1.4] cm^{-1}.

algorithm parameters. We then characterize the properties of the reconstructed images by using visualization assessment.

7.2.1 Modified Shepp-Logan phantom study

Study design We mimic a C-arm CBCT configuration with a source-to-detector distance of 119.7 cm, and a source-to-rotation-axis distance of 72.4 cm. The 2-dimensional detector consists of 512 detector element of size 0.02 cm. We generate data from an analytic Shepp-Logan phantom, shown in Fig. 7.1, at 720 views uniformly distributed over 2π. The modified shepp-logan phantom contains three bar objects with different orientations, a cross object and a moon-shape object. The goal is to reconstruct an image with the bars resolved and the shape of other objects accurately recovered.

We consider variable-resolution arrays determined by the selection of different combina-

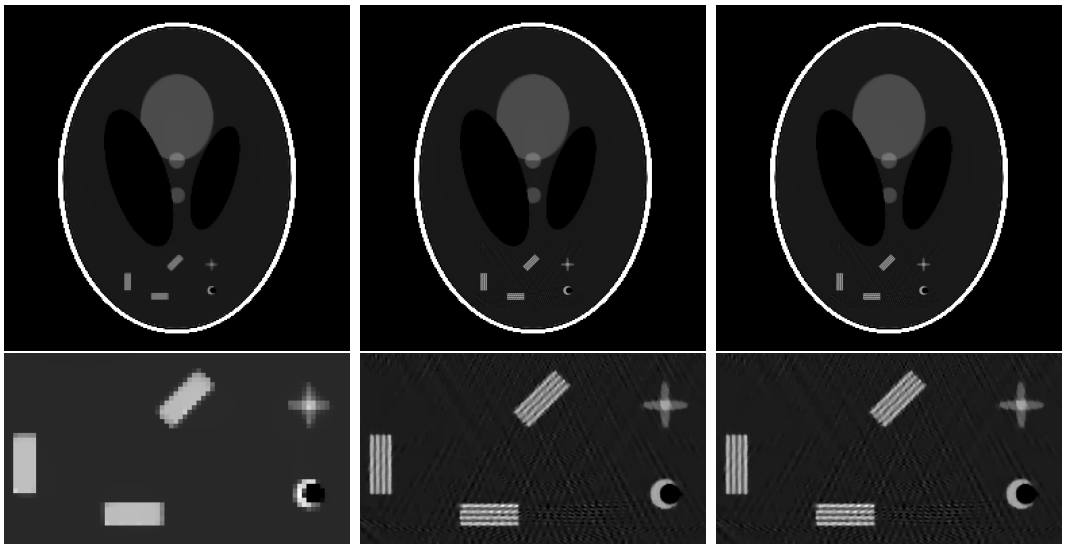


Figure 7.2: Image of modified Shepp-Logan phantom (top row) reconstructed by use of the ASD-POCS algorithm on a uniform-resolution array with pixel size 0.078 cm (left), a uniform resolution array with pixel size of 0.0195 cm (middle), and a variable-resolution array specified by the combination of $a = 0.0195$ cm and $b = 0.078$ cm (right). Display window: [0.6, 1.4] cm^{-1}.

tions of pixel sizes a and b. Specifically, in this study, we select $a = 0.0195$ cm to combine with $b = 0.0195$ cm and 0.078 cm, respectively. Clearly, the first combination yields a uniform-resolution array, whereas the second combination results in a variable-resolution array with a decreasing ratio between a and b.

Study results In order to better resolve the bar phantom and reveal the shape of other fine structures, we firstly performed image reconstruction on a uniform-resolution array consisting of pixels of size $a = 0.0195$ cm, and showed the high-resolution ASD-POCS result in the middle panel of Fig. 7.2. The ROI image in a zoomed-in view was also illustrated in the second row of Fig. 7.2. We also performed ASD-POCS reconstruction on a uniform low-resolution array with pixel size of 0.078 cm, which was shown in the left panel of Fig. 7.2. Comparing the reconstructed images with the truth, it is obvious that the high-resolution image can resolve all the four bars for each bar object, and make the shape of other fine structures comparable to the truth; but in the low-resolution image, the bar objects become rectangular blocks and the shape of other fine structures are severely distorted.

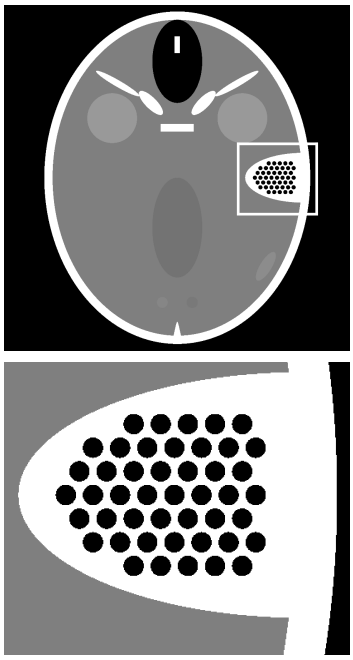


Figure 7.3: A FORBILD head phantom (top) discretized on a uniform-resolution array with pixel size of 0.01225 cm, with an ROI enclosed by a white box; A zoomed-in view of the ROI image (bottom). Display window: [1.0, 1.1] cm^{-1}.

We also performed image reconstruction on a variable-resolution array, with the pixel size outside the ROI 4 times as large as the pixel size within the ROI, which is shown in the right panel of Fig. 7.2. Comparing the reconstruction on a variable-resolution array with the reconstructions on uniform-resolution arrays, we observed that the ROI in the variable-resolution image is comparable to the ROI in the uniform high-resolution image, while the region outside the ROI in the variable-resolution image is comparable to the corresponding region in the uniform low-resolution image. In such a variable-resolution image, all the bar phantoms can be resolved and the shape of fine structures are well recovered and comparable to the truth. Although the ROI images in the variable-resolution image and in the uniform high-resolution image are identical to each other, the total number of elements in the former array is much less than in the latter one.

7.2.2 FORBILD head phantom study

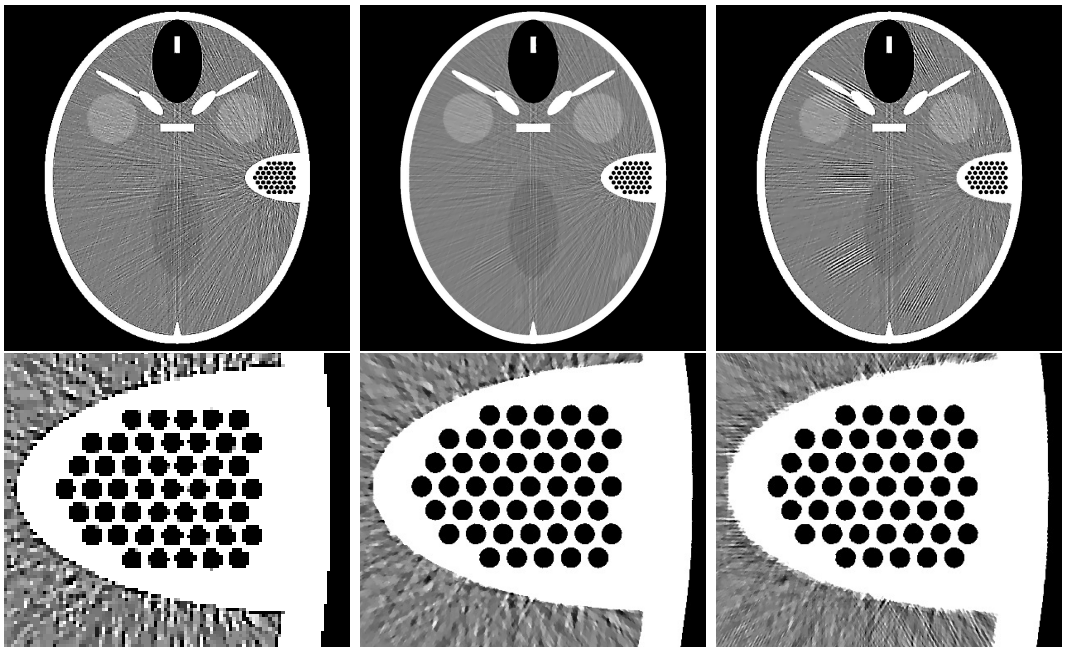
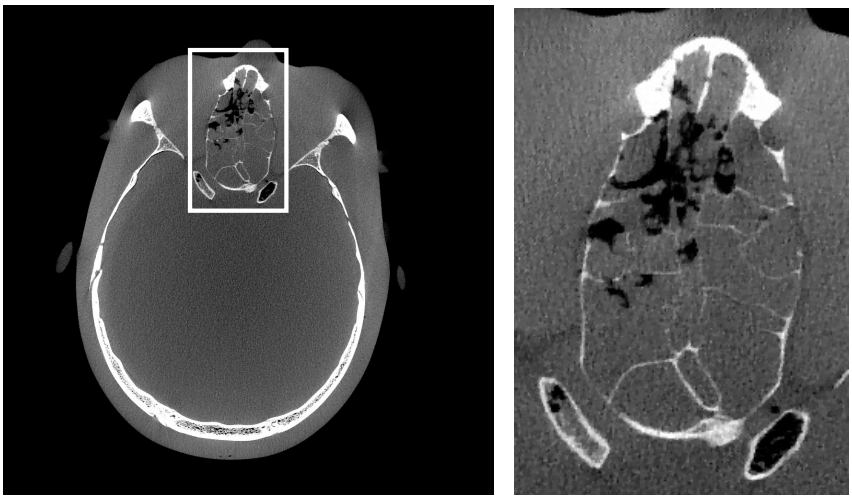


Figure 7.4: Image of FORBILD head phantom (top row) reconstructed by use of the ASD-POCS algorithm on a uniform-resolution array with pixel size 0.049 cm (left), a uniform resolution array with pixel size of 0.01225 cm (middle), and a variable-resolution array specified by the combination of $a = 0.01225$ cm and $b = 0.049$ cm (right). Display window: [1.0, 1.1] cm^{-1}.

Study design We mimic a C-arm CBCT configuration with a source-to-detector distance of 200 cm, and a source-to-rotation-axis distance of 100 cm. The 2-dimensional detector consists of 512 detector element of size 0.02 cm. We generate data from an analytic FORBILD phantom, shown in Fig. 7.1, at 360 views uniformly distributed over 2π.

In this study, we select image arrays with $a = 0.01225$ cm to combine with $b = 0.01225$ cm and 0.049 cm, respectively. Clearly, the first combination yields a uniform-resolution array, whereas the second combination results in a variable-resolution array with a decreasing ratio between a and b.

Study results Again, we firstly reconstruct images by using of ASD-POCS algorithm on two uniform-resolution arrays, the pixel size of one array is four times as large as the pixel size of the other one. Similar to the previous case, the ROI in the uniform high-resolution image is comparable to the ROI in the truth: the boundary of each ear-hole is round, but



(a) (b)

Figure 7.5: An anthropomorphic head phantom image (a) reconstructed by use of ASD-POCS on a uniform-resolution array with pixel size of 0.024 cm, with an ROI enclosed by a white box; A zoomed-in view of the ROI image (b). Display window: [0.1, 0.3] cm^{-1}.

the shape ear-holes in the uniform low-resolution image is changed: the boundaries are not round any more and become stair-shaped or flat.

We also performed optimization-based image reconstruction on a variable-resolution image array, illustrated in the right panel of Fig. 7.4. In this result, although the structures in the region outside the ROI become distorted, the ear-holes within the ROI are still comparable to those in the high-resolution reconstruction result and the truth image. Therefore, to obtain a high-resolution ROI image, it is not necessary to perform image reconstruction with uniform fine pixels and a coarse grid outside the ROI is enough. But the total number of elements in the array can be significantly reduced.

7.3 Physical-phantom studies with variable resolution

7.3.1 Data acquisition

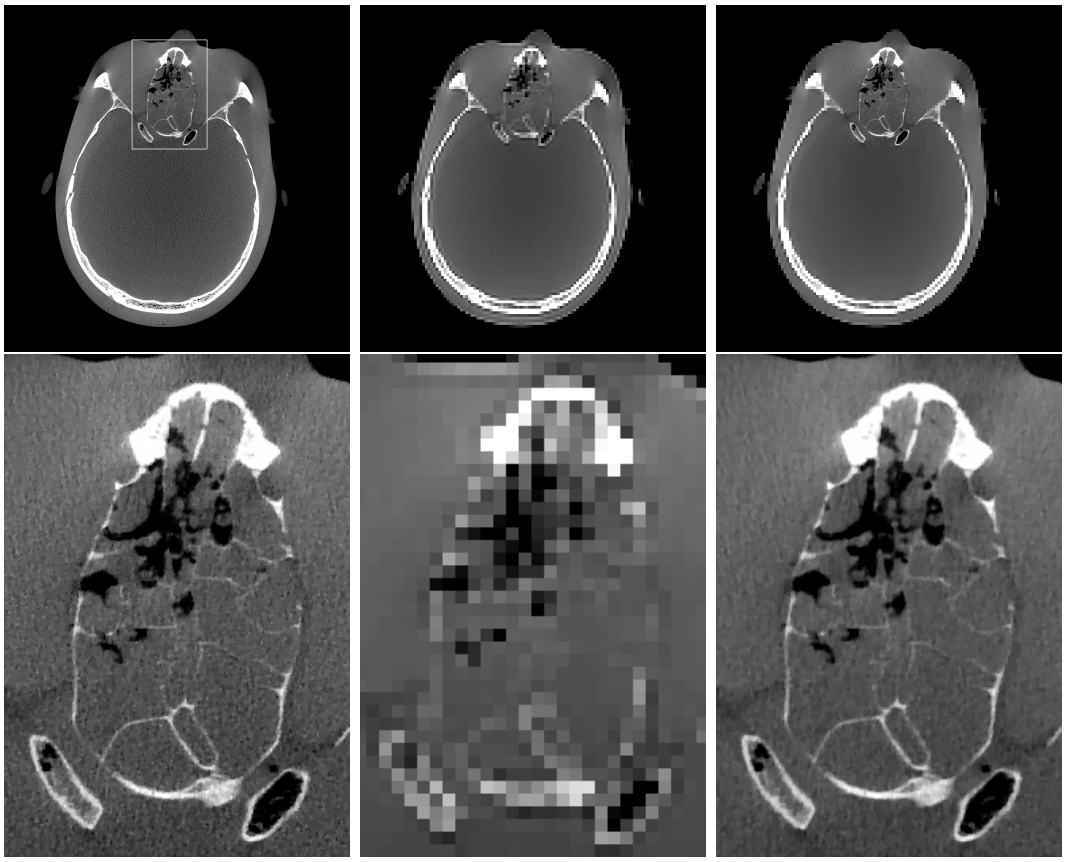
As a preparation towards investigating animal/patient image reconstruction from data acquired with clinical CBCT systems, we first conducted a study by using physical-phantom data collected with a bench-top CBCT system. Full knowledge about the phantom and the

system is available, which offers a unique advantage in algorithm development/tailoring and reconstruction characterization.

In this study, data were collected from an anthropomorphic head phantom, which is constructed based upon real human skeleton, with low-contrast spheres mimicking soft-tissue organs embedded in the brain region. The scanner is a turn-table-based CBCT system consisting of an X-ray source operated at 120 kVp and a 1024×1024-bin flat-panel detector. A circular scanning geometry was employed, in which X-ray exposures were taken at 960 views uniformly distributed over a 2π angular range.

7.3.2 Reconstruction results

We reconstructed from the phantom data three images by using the ASD-POCS algorithm. First, two images were reconstructed with uniform-resolution arrays, a high-resolution array of 0.024 cm pixels and a low-resolution array of 0.192 cm pixels. The image reconstructed on the high-resolution array is displayed in Fig. 7.5(a). An ROI is defined in Figs. 7.5(a) by a white box, which is centered at the sinus, a region of rich details of bone structures often critical for performing IGS tasks. We then performed the third image reconstruction from the same data on a variable-resolution array, by using the approach developed. The variable-resolution array is composed by pixels of size $a = 0.024$ cm within the ROI, and pixels of size $b = 0.192$ cm outside of the ROI. The images within the ROI of those reconstructions are displayed in Fig. 7.6. Observation can be made that the image quality of the variable-resolution reconstruction is visually comparable to that of the uniform high-resolution reconstruction within the ROI. This result suggests that the considerable reduction on the number of image pixels outside of the ROI appears to have little impact on the reconstruction quality within the ROI, while a large amount of computation burden and memory consumption can be spared.



(a) (b) (c)

Figure 7.6: ROI images reconstructed by use of ASD-POCS algorithm from data of the anthropomorphic head phantom on arrays of uniform-resolution array with pixel size of 0.024 cm(a), a uniform-resolution array with pixel size of 0.196 cm (b), and a variable-resolution array specified by the combination of $a = 0.024$ cm and $b = 0.196$ cm (c). Display window: [0.1, 0.3] cm^{-1}.

Figure 7.7: Toshiba Aquilion ONE^{TM} CT scanner.

7.4 Animal/patient studies from diagnostic-CBCT data

7.4.1 Data acquisition

In the work, we focus on image reconstruction from patient and swine data collected with an Aquilion ONE^{TM} CT scanner (Toshiba American Medical Systems, Tustin, CA, USA), shown in Fig. 7.7. In the CT scanner, the distance between the X-ray source and the center of rotation is 60 cm; the detector consists of 320 rows of bins, each of the 320 rows composes 896 bins, thus forming a fan angle of 49.2°.

The patient and swine data were collected at 1200 views over 2π. Specifically, in the patient study, the subject was scanned with X-ray energy of 135 kVp and current at 100 mAs, whereas for the swine study, the animal was scanned with X-ray energy of 100 kVp and current at 275 mAs. The projection data were corrected for scatter and beam-hardening by use of Toshiba's standard method installed on the scanner. In this study, data of the middle row was extracted for image reconstruction.

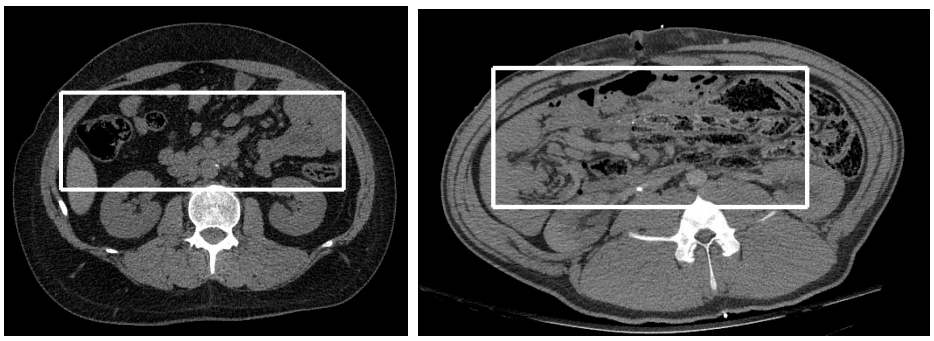


Figure 7.8: Patient (left) and swine (right) images reconstructed by use of the ASD-POCS algorithm in which ROIs are indicated by the white solid lines. The pixel sizes of 0.064 cm and 0.0625 cm are used in the uniform-resolution arrays, respectively, for the patient and swine reconstructions.

7.4.2 Selection of image arrays

In our studies below, without loss of generality, we consider four variable-resolution arrays determined by the selection of four different combinations of pixel sizes a and b, as shown in Fig. 1. Specifically, for the patient study, we select $a = 0.064$ cm to combine with $b = 0.064$ cm, 0.128 cm, 0.256 cm, and 0.512 cm, respectively. Clearly, the first combination yields a uniform-resolution array, whereas the other three combinations result in three variable-resolution arrays with a decreasing ratio between a and b. Similarly, for the swine study, four image arrays were determined by the four combinations of $a = 0.0625$ cm with $b = 0.0625$ cm, 0.125 cm, 0.250 cm, and 0.500 cm, respectively. Again, the first combination produces a uniform-resolution array, whereas the other three combinations result in three variable-resolution arrays, also with a decreasing ratio between a and b.

7.4.3 Results

Reconstruction from patient data We first performed optimization-based reconstructions on a uniform-resolution array consisting of pixels of size $a = 0.064$ cm, and show the ASD-POCS reconstruction in the left panel of Fig. 7.8. The ROI image enclosed by the white lines is also shown in a zoomed-in view in the first rows of Figs. 7.9-7.11, and it is used as a gold standard against which ROI reconstructions on variable-resolution ar-

rays are compared. We then carried out ASD-POCS reconstructions on three additional uniform-resolution arrays with pixel sizes of 0.128 cm, 0.256 cm, and 0.512 cm, respectively, and display the corresponding ROI images in the second rows of Figs. 7.9-7.11. Clearly, reconstruction resolution decreases as the pixel size increases from 0.064 cm to 0.512 cm.

We also performed optimization-based reconstructions on three different, variable-resolution arrays, in the patient study, as described in Sec. 7.4.2. In row 3 of Figs. 7.9-7.11, we display the respective ROI images reconstructed by using the ASD-POCS algorithm.

Comparing ROI reconstructions of Figs. 7.9-7.11 with the gold standard in the first row of Figs. 7.9-7.11, we observe that images within regions outside the ROI, with increased pixel sizes, rapidly lose their resolution, as expected. However, image resolution within the ROI remains virtually unchanged, and is comparable to the gold standard resolution, even though the ratio between the pixel sizes within and outside the ROI has decreased considerably.

Reconstructions from swine data Again, we carried out optimization-based reconstructions on a uniform-resolution array consisting of pixels of size $a = 0.0625$ cm, and show the ASD-POCS reconstruction in the right panel of Fig. 7.8. The ROI image enclosed by the white lines is also shown in a zoomed-in view in the first rows of Figs. 7.12-7.14, and it is also used as a gold standard against which ROI reconstructions on variable-resolution arrays are compared. We also did ASD-POCS reconstructions on three uniform-resolution arrays with pixel sizes of 0.125 cm, 0.25 cm, and 0.5 cm, respectively, and display the ROI images in the second rows of Figs. 7.12-7.14. It can be observed that reconstruction resolution decreases as the pixel size increases from 0.0625 cm to 0.5 cm.

We then performed optimization-based reconstructions on three different, variable-resolution arrays, in the swine study, as described in Sec. 7.4.2. In row 3 of Figs. 7.12-7.14, we display the respective ROI images reconstructed by using the ASD-POCS reconstruction algorithm.

Comparison of ROI reconstructions in row 3 of Figs. 7.12-7.14 with the gold standard reconstruction displayed in row 1 of Figs. 7.12-7.14 reveals that observations similar to those made for the patient study above can also be made for the swine study.

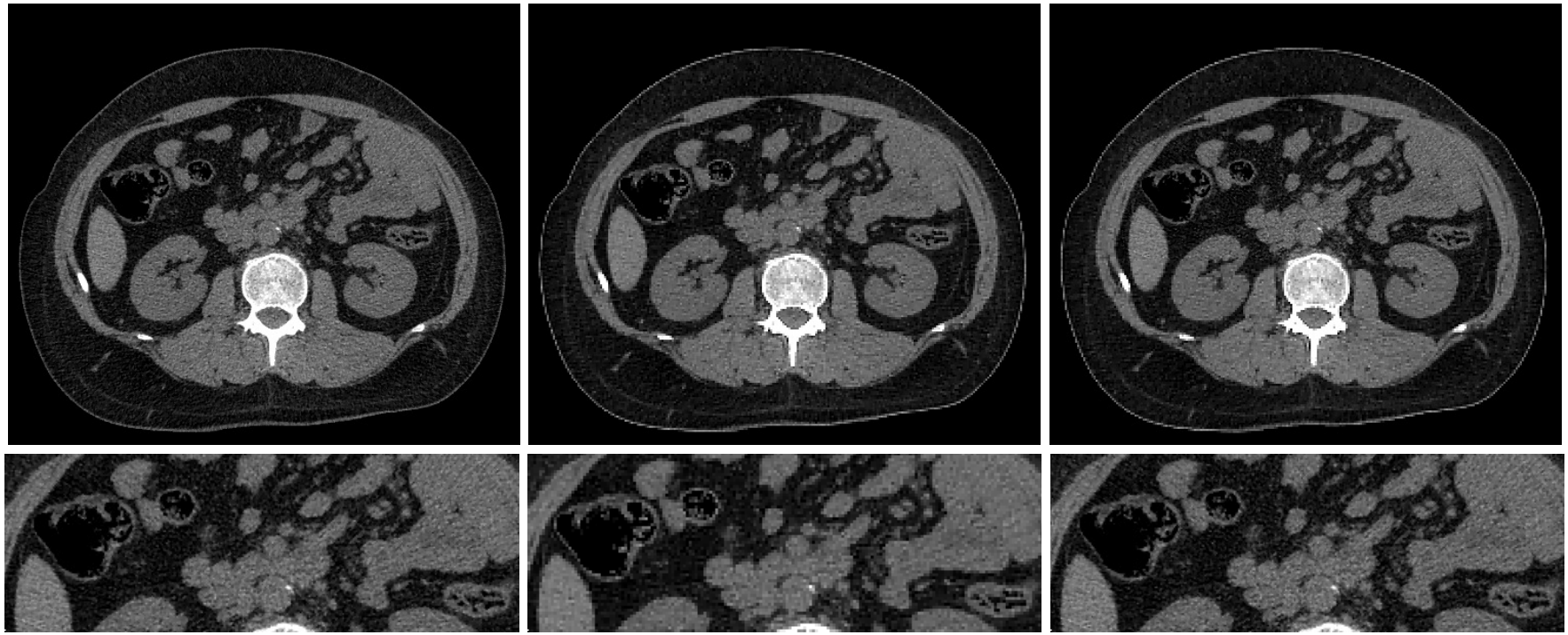


Figure 7.9: Images reconstructed by use of the ASD-POCS algorithm from patient data on a uniform-resolution array with pixel size 0.064 cm (column 1), a uniform-resolution array with pixel size 0.128 cm (column 2), and a variable-resolution array specified by the combination of $a = 0.064$ cm and $b = 0.128$ cm (column 3). A zoom-in view of ROI images are displayed in the bottom row. A display window [-6, 13] is used.

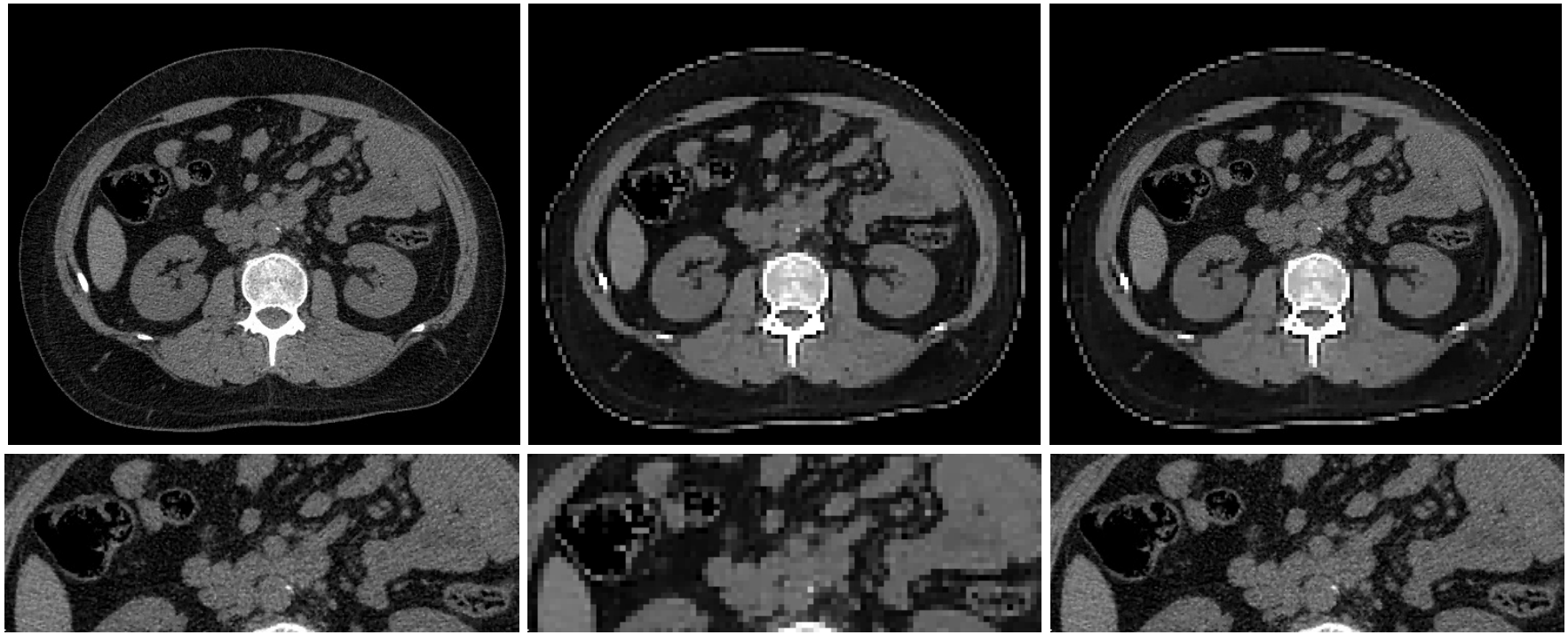


Figure 7.10: Images reconstructed by use of the ASD-POCS algorithm from patient data on a uniform-resolution array with pixel size 0.064 cm (column 1)), a uniform-resolution array with pixel size 0.256 cm (column 2), and a variable-resolution array specified by the combination of $a = 0.064$ cm and $b = 0.256$ cm (column 3). A zoom-in view of ROI images are displayed in the bottom row. A display window [-6, 13] is used.

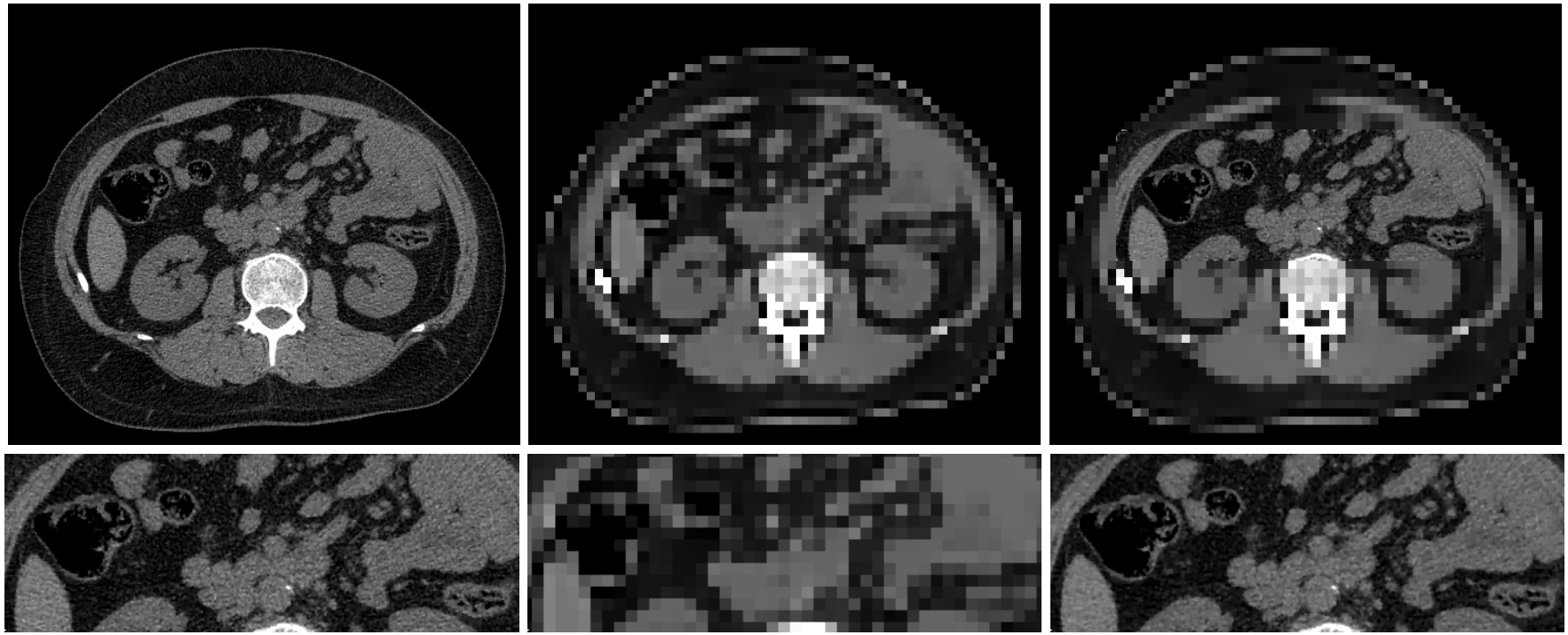


Figure 7.11: Images reconstructed by use of the ASD-POCS algorithm from patient data on a uniform-resolution array with pixel size 0.064 cm (column 1), a uniform-resolution array with pixel size 0.512 cm (column 2), and a variable-resolution array specified by the combination of $a = 0.064$ cm and $b = 0.512$ cm (column 3). A zoom-in view of ROI images are displayed in the bottom row. A display window [-6, 13] is used.

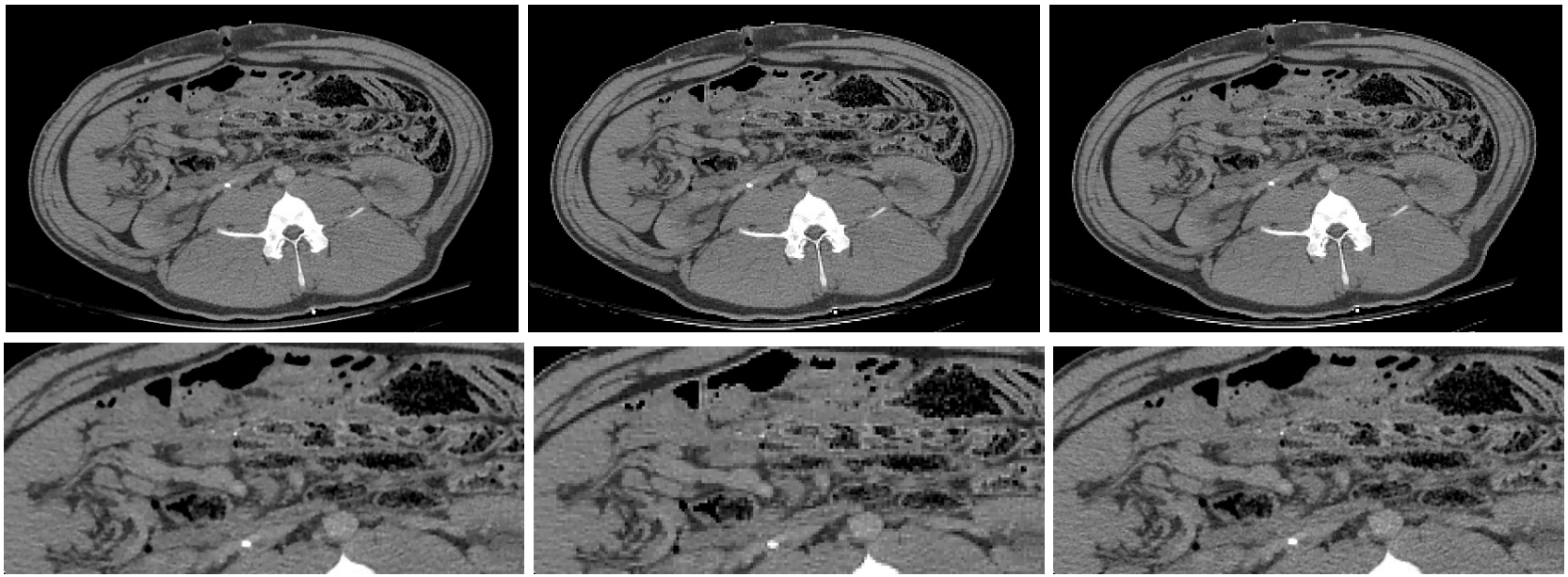


Figure 7.12: Images reconstructed by use of the ASD-POCS algorithm from swine data on a uniform-resolution array with pixel size 0.0625 cm (column 1), a uniform-resolution array with pixel size 0.125 cm (column 2), and a variable-resolution array specified by the combination of $a = 0.0625$ cm and $b = 0.125$ cm (column 3). A zoom-in view of ROI images are displayed in the bottom row. A display window [-8.5, 16] is used.

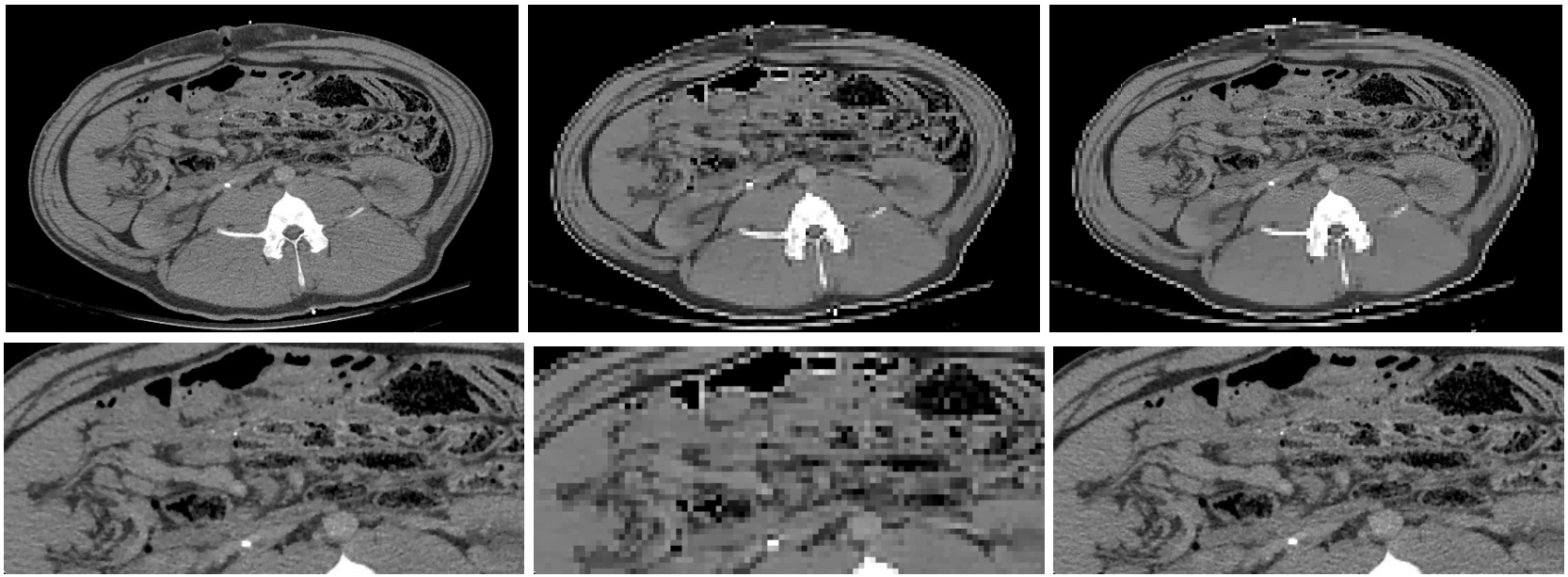


Figure 7.13: Images reconstructed by use of the ASD-POCS algorithm from swine data on a uniform-resolution array with pixel size 0.0625 cm (column 1), a uniform-resolution array with pixel size 0.25 cm (column 2), and a variable-resolution array specified by the combination of $a = 0.0625$ cm and $b = 0.25$ cm (column 3). A zoom-in view of ROI images are displayed in the bottom row. A display window [-8.5, 16] is used.

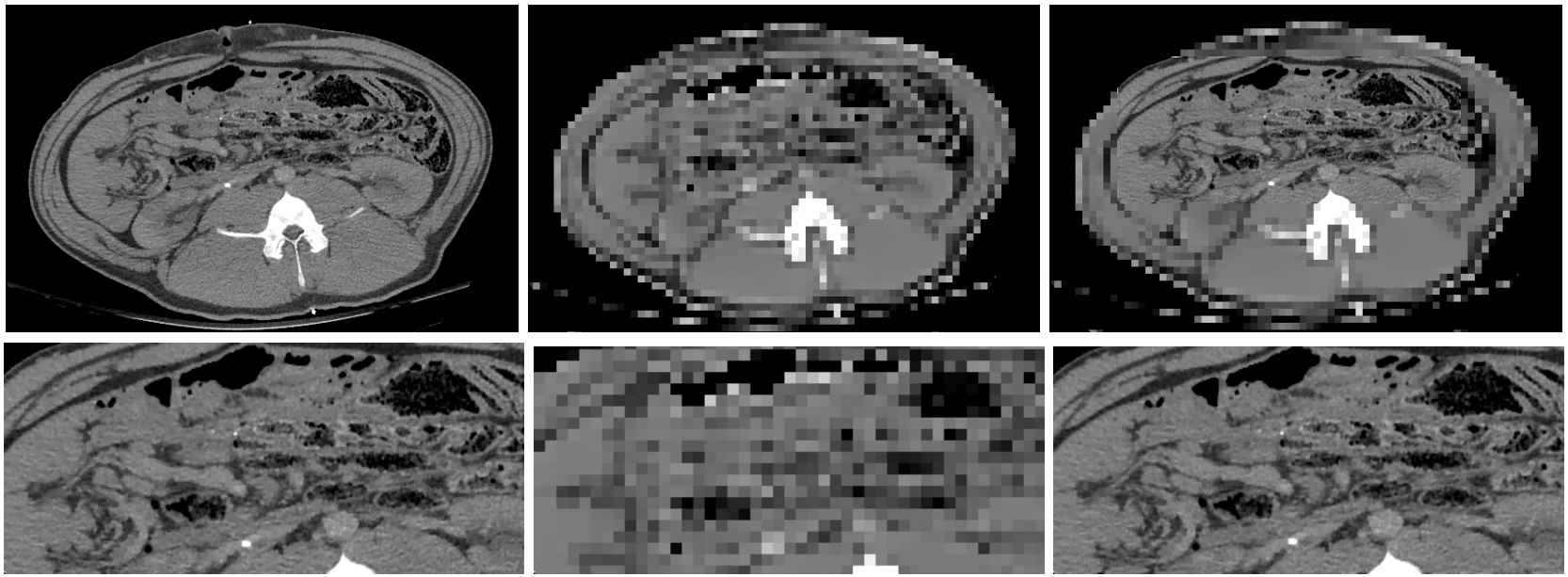


Figure 7.14: Images reconstructed by use of the ASD-POCS algorithm from swine data on a uniform-resolution array with pixel size 0.0625 cm (column 1), a uniform-resolution array with pixel size 0.5 cm (column 2), and a variable-resolution array specified by the combination of $a = 0.0625$ cm and $b = 0.5$ cm (column 3). A zoom-in view of ROI images are displayed in the bottom row. A display window [-8.5, 16] is used.

7.5 Patient Studies from dental-CBCT data

7.5.1 CBCT imaging system and data acquisition

We collected patient data with an iCAT dental CBCT system, shown in Fig. 7.15, operated with X-ray energy of 120 kVp and tube current of 20 mAs. The distances from the source to the rotation axis and to the detector are 49.77 cm and 71.69 cm, respectively. The detector consists of 480×384 elements of size 0.508×0.508 mm^2. The diameter of the field-of-view (FOV) is about 13 cm within a transverse plane. In this study, the patient was scanned over 2π angular range at 300 projection views. The data contain truncation due to the limited size of FOV.

7.5.2 Image arrays of variable resolution

In this study, we employ 3D arrays of variable resolution, which includes a rectangle-shape ROI consisting of voxels of size a, and volume surrounding the ROI composing of voxels of size b. When $a = b$, a variable-resolution array becomes a uniform-resolution array. If $a \neq b$, as the ratio a/b varies, different levels of resolution are obtained for the ROI and its surrounding volume. In particular, when $a < b$, the array thus includes an ROI with resolution higher than that in the region surrounding the ROI.

In our studies below, we consider four arrays determined by the selection of four different combinations of voxel sizes a and b. Specifically, we select $a = 0.3$ mm to combine with $b = 0.3$ mm, 1.2 mm, 2.4 mm, and 4.8 mm, respectively. The first combination yields a uniform-resolution array, whereas the other three combinations result in three variable-resolution arrays with an increasing ratio between b and a.

7.5.3 Reconstruction results

We first performed optimization-based reconstructions on a uniform-resolution array consisting of voxels of size $a = 0.3$ mm, and showed the ASD-POCS reconstruction in the top

Figure 7.15: i-CAT dental CT scanner.

left panel of Figs. (7.16)-(7.18). The ROI image is enclosed by the white box, and it is used as a reference to those ROI reconstructions on variable-resolution arrays.

We then carried out reconstructions by using ASD-POCS on three different, variable-resolution arrays, as shown in Figs. (7.16)-(7.18). Comparing reconstructions on variable- and uniform-resolution image arrays, we observe that images within regions outside the ROI lose rapidly the resolution with increased voxel sizes, as expected, while image resolution within the ROI remains virtually unchanged. For instance, in Figs. (7.16) and (7.17), the ROI images on variable-resolution arrays is comparable to that on uniform-resolution array, and detailed information of root canals of teeth and sinus are clearly revealed. It is worth noting that, although image quality within the ROI are comparable in the four reconstructions, the ratio between the voxel sizes outside and inside the ROI has increased considerably.

7.6 Discussion

In this chapter, we have investigated 2D and 3D optimization-based reconstructions by using the modified ASD-POCS algorithm on images with variable resolution, from both simulation and real-data. The results demonstrate that properly designed iterative algorithms can reconstruct an image on an array with variable resolution, yielding an ROI image with quality comparable to that obtained with an image array with uniform, high resolution. Results of this work may have implications for practical applications of iterative reconstruction algorithms as they can be exploited for reducing computation memory and time.

It is worth noting that, in certain CBCT imaging applications, there may exist special requirements which may result in non-conventional CBCT imaging configurations, such as sparse-view CBCT imaging, imaging with limited angular range, or truncated data. Existing work suggests that when applied to those data, the image quality of reconstructions are degraded due to artifacts. Due to the reduced total number of pixels, the proposed optimization-based reconstruction based on arrays of variable resolution may yield high-

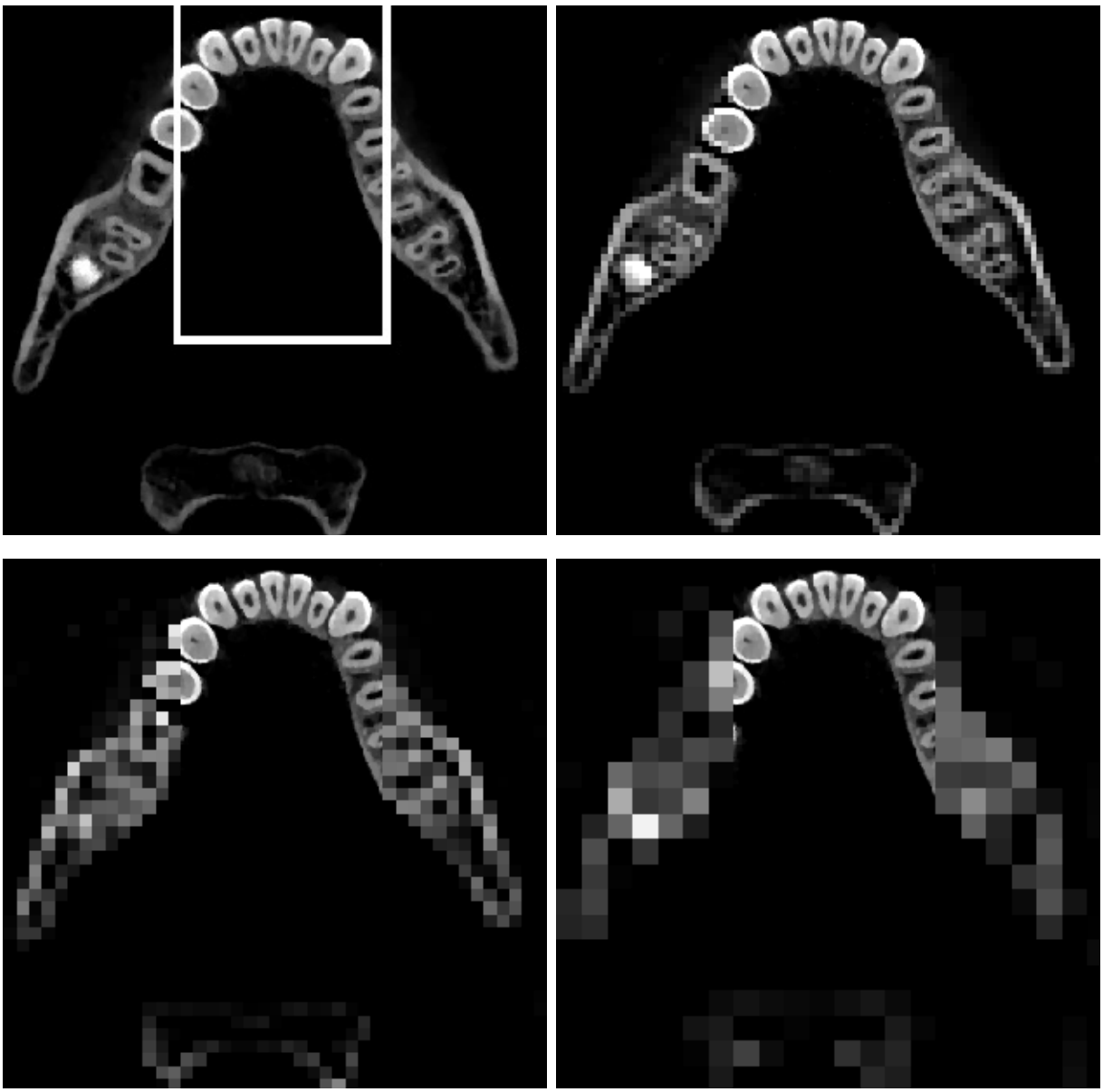


Figure 7.16: A transverse slice of images reconstructed by use of the ASD-POCS algorithm from patient data on a uniform-resolution array with voxels of size 0.3 mm (top left), a variable-resolution array with voxels of size $a = 0.3$ mm and $b = 1.2$ mm (top right), a variable-resolution array with voxels of size $a = 0.3$ mm and $b = 2.4$ mm (bottom left), and a variable-resolution array with voxels of size $a = 0.3$ mm and $b = 4.8$ mm (bottom right). ROI is enclosed within a white box. A display window [0.2, 0.5] cm^{-1} is used.

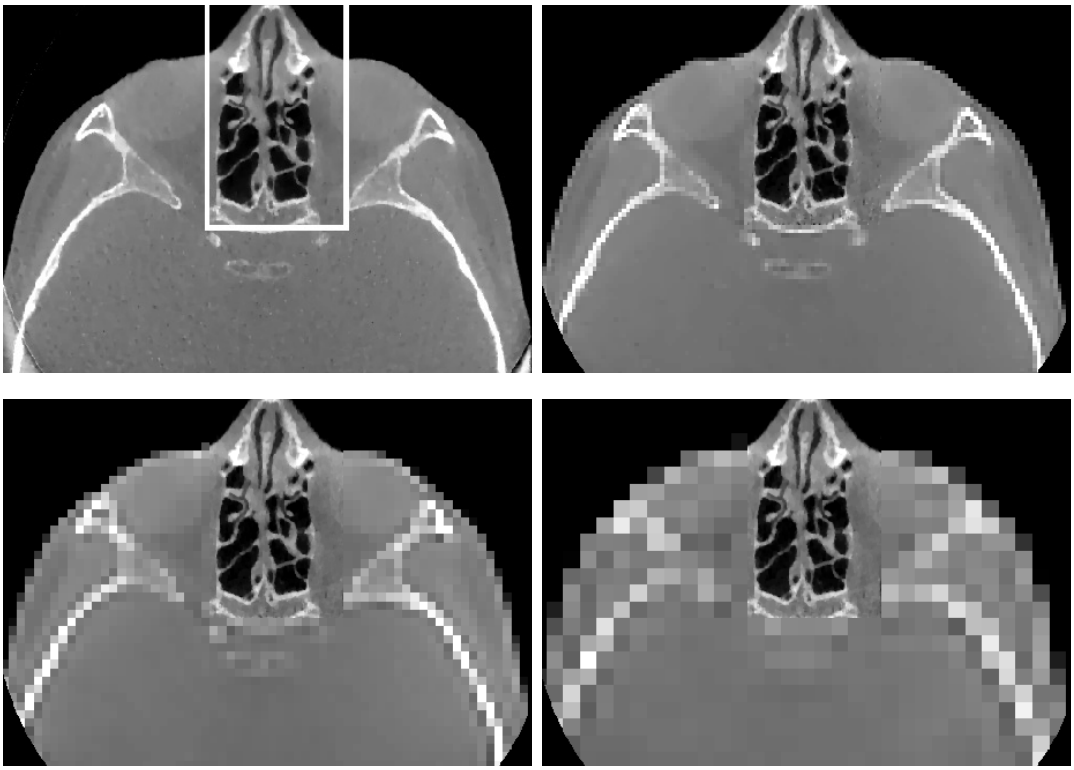


Figure 7.17: A transverse slice of images reconstructed by use of the ASD-POCS algorithm from patient data on a uniform-resolution array with voxel size of 0.3 mm (top left), a variable-resolution array with voxels of size $a = 0.3$ mm and $b = 1.2$ mm (top right), a variable-resolution array with voxels of size $a = 0.3$ mm and $b = 2.4$ mm (bottom left), and a variable-resolution array with voxels of size $a = 0.3$ mm and $b = 4.8$ mm (bottom right). ROI is enclosed within a white box. A display window [0.05, 0.3] cm^{-1} is used.

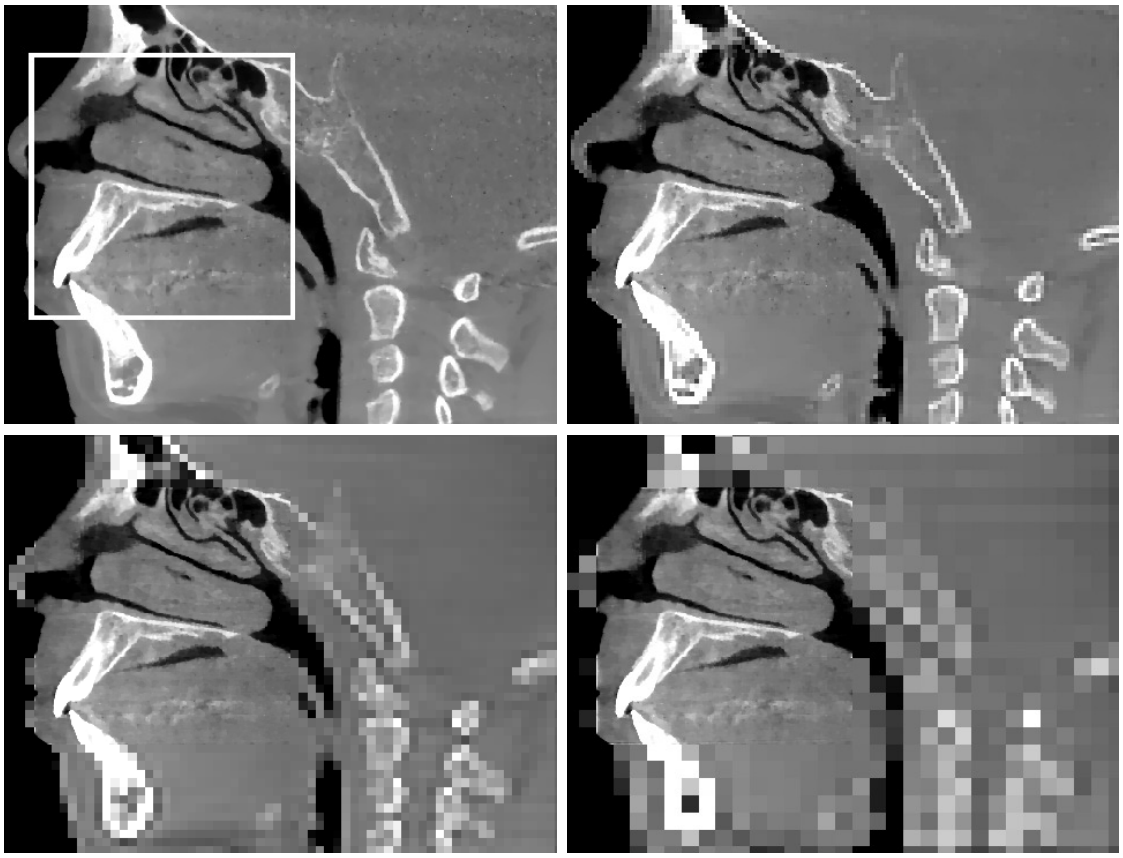


Figure 7.18: A sagittal slice of images reconstructed by use of the ASD-POCS algorithm from patient data on a uniform-resolution array with voxel size of 0.3 mm (top left), a variable-resolution array with voxels of size $a = 0.3$ mm and $b = 1.2$ mm (top right), a variable-resolution array with voxels of size $a = 0.3$ mm and $b = 2.4$ mm (bottom left), and a variable-resolution array with voxels of size $a = 0.3$ mm and $b = 4.8$ mm (bottom right). ROI is enclosed within a white box. A display window [0.05, 0.3] cm^{-1} is used.

resolution ROI with substantially reduced artifacts in such scan conditions.

CHAPTER 8

DISCUSSION

In this thesis, we have developed and utilized optimization-based reconstruction tools for enabling non-conventional tomographic imaging configurations, and for remedying existing challenges of interest in practical applications. We have demonstrated that, with appropriately designed optimization programs, optimization-based reconstruction tools can be obtained for *1*) reducing the artifacts existing in the off-middle plane of FDK reconstruction from short-scan CBCT data, *2*) enabling a sparse-PET configuration with reduced crystal while not significantly compromising the image quality, and *3*) allowing iterative reconstruction based on an image with variable resolution.

In Chapters 2 and 3, we have investigated optimization-based reconstruction from short-scan CBCT data and focused on artifact reduction at the off-middle plane. We first designed numerous optimization programs, with data fidelities such as data-ℓ_2, data-ℓ_1, and data-KL fidelities, and image constraints such as image-ℓ_2, image-ℓ_1, and image-TV. It is worth studying both data fidelities and image constraints, because both play an important role in the design of optimization programs in CBCT. We then employ a primal-dual algorithm developed by Chambolle and Pock (CP) for solving those convex optimization programs. By using a single algorithm for solving all of the optimization programs, it may avoid any reconstruction differences arising as a result of using different algorithms for solving different optimization programs. We further carried out inverse-crime studies to validate the derived CP algorithms. After the algorithm derivation and validation, we have utilized the developed tools to investigate optimization-based reconstructions by using CP algorithm for possible reduction of image artifacts often observed in FDK reconstruction from real short-scan CBCT data. The study confirms that optimization-based reconstruction with appropriate image constraints such as the image-TV constraint can reduce FDK-reconstruction artifacts in CBCT with a short-scan configuration. As the results demonstrated, a direct benefit of such an artifact reduction can be an improved contrast level of low-contrast anatomic structures

that are otherwise obscured by the artifacts. Furthermore, we notice that, although the selection of different weighting matrices in Chapters 2 and 3 has a significant impact on FDK reconstruction, it appears to have insignificant effect on optimization-based reconstructions.

In Chapters 4 and 5, we have designed a sparse-PET system containing fewer detectors than that in a conventional PET configuration, and we investigated image reconstruction by using the CP algorithm for solving image-TV-constrained optimization problems associated with different data fidelities, such as data-ℓ_2, data-ℓ_1, and data-KL fidelities. A Gaussian blurring matrix is included in the designed optimization programs. This is because data collected in PET imaging generally have SNR considerably lower than that of data in typical CT imaging, and transitions among different uptake regions in a PET-activity map is observed to be generally not as sharp as that among some anatomic regions in a CT-attenuation map. Therefore, a desired PET image is formulated as a product of a latent image and Gaussian blurring matrix, and the TV constraint is applied to the latent image instead of the desired PET image. This formulation thus allows for a latent image with sparser gradient magnitude image than the desired image, and avoids yielding an image with significant patchy textures when the data have low SNR. Again, we have carried out inverse-crime studies to verify whether the CP algorithm can accurately solve the designed optimization programs. Employing the developed CP algorithms, we have performed an investigation on applying TV-constrained optimization techniques to reconstructing PET images from real data, with both full- and sparse-PET configurations. Results show that the proposed CP algorithm can be exploited for enabling a PET configuration with sparsely populated detectors by extracting the sparse data from the full data. Further, we show that the behavior of the optimization-based reconstruction depend on numerous factors, which include the selection of program parameters (e.g., image constraint and voxel size), different designs of optimization problem (e.g., different data fidelities and image constraints), and data conditions, *e*tc.

In Chapters 6 and 7, we have studied optimization-based reconstruction on images

with variable resolution. In particular, we designed the data-ℓ_2-constrained image-TV-minimization program, and investigated the ASD-POCS algorithm for reconstruction with image of variable resolution. The calculation of key components of the algorithm, such as the forward and backward projections, the image TV, and the gradient of image TV, have been tailored to images of variable resolution. An inverse-crime study was carried out to validate the implementation of the modified ASD-POCS algorithm. Further, we have investigated 2D and 3D optimization-based reconstructions from both simulation and real-data, by using the modified ASD-POCS algorithm on images with variable resolution. Results show that images of variable resolution can be used in optimization-based reconstruction, and the total number of image voxels can be reduced, thus relieving the computational burden. It is worth noting that, although the focus of the work is to develop the ASD-POCS with variable resolution in CBCT, the concept and techniques developed can readily be generalized to other existing iterative algorithms such as EM, SART, and CP algorithms.

In summary, optimization-based reconstruction in tomographic imaging offers great flexibility in program design, shows capacity for improving image quality, and demonstrates the potential for enabling innovative imaging configurations. To date, many optimization algorithms, such as CP and ASD-POCS algorithm, have demonstrated to be able to solve those convex optimization problems, mathematically or practically. However, it is worth noting that, optimization algorithms often involve numerous parameters for image reconstruction, the selection of which may have huge impact on the outcome. Therefore, in studies of optimization-based reconstruction, it is of interest and significance to carefully investigate program/algorithm/convergence parameters, and how those parameters affect image reconstruction.

REFERENCES

[1] A. M. Cormack, "Recollections of my work with computer assisted tomography," *Molec. Cell. Bio.*, vol. 32, no. 2, pp. 57–61, 1980.

[2] G. Hounsfield, "The emi scanner," *Proc. Royl. Soc. of London B: Bio. Sci.*, vol. 195, no. 1119, pp. 281–289, 1977.

[3] K. A. Miles, "Perfusion imaging with computed tomography: brain and beyond," *Euro. Radio. Suppl.*, vol. 16, no. 7, pp. M37–M43, 2006.

[4] M. J. Haydel, C. A. Preston, T. J. Mills, S. Luber, E. Blaudeau, and P. M. DeBlieux, "Indications for computed tomography in patients with minor head injury," *N. Eng. Jourl. Med.*, vol. 343, no. 2, pp. 100–105, 2000.

[5] B. Jacobs, T. Beems, T. M. van der Vliet, R. R. Diaz-Arrastia, G. F. Borm, and P. E. Vos, "Computed tomography and outcome in moderate and severe traumatic brain injury: hematoma volume and midline shift revisited," *J. Neurotra.*, vol. 28, no. 2, pp. 203–215, 2011.

[6] A. S. Agatston, W. R. Janowitz, F. J. Hildner, N. R. Zusmer, M. Viamonte, and R. Detrano, "Quantification of coronary artery calcium using ultrafast computed tomography," *J. the Amer. Col. of Cardio.*, vol. 15, no. 4, pp. 827–832, 1990.

[7] G. L. Raff, M. J. Gallagher, W. W. O′Neill, and J. A. Goldstein, "Diagnostic accuracy of noninvasive coronary angiography using 64-slice spiral computed tomography," *J. the Amer. Col. of Cardio.*, vol. 46, no. 3, pp. 552–557, 2005.

[8] N. R. Mollet, F. Cademartiri, C. A. van Mieghem, G. Runza, E. P. McFadden, T. Baks, P. W. Serruys, G. P. Krestin, and P. J. de Feyter, "High-resolution spiral computed tomography coronary angiography in patients referred for diagnostic conventional coronary angiography," *Circulation*, vol. 112, no. 15, pp. 2318–2323, 2005.

[9] S. Achenbach, D. Ropers, A. Kuettner, T. Flohr, B. Ohnesorge, H. Bruder, H. Theessen, M. Karakaya, W. G. Daniel, and W. Bautz, "Contrast-enhanced coronary artery visualization by dual-source computed tomography-initial experience," *Euro. J. Radio.*, vol. 57, no. 3, pp. 331–335, 2006.

[10] A. J. Einstein, M. J. Henzlova, and S. Rajagopalan, "Estimating risk of cancer associated with radiation exposure from 64-slice computed tomography coronary angiography," *Jama*, vol. 298, no. 3, pp. 317–323, 2007.

[11] J. H. Warrick, M. Bhalla, S. I. Schabel, and R. M. Silver, "High resolution computed tomography in early scleroderma lung disease." *J. Rheumat.*, vol. 18, no. 10, pp. 1520–1528, 1991.

[12] S. Sone, S. Takashima, F. Li, Z. Yang, T. Honda, Y. Maruyama, M. Hasegawa, T. Yamanda, K. Kubo, and K. Hanamura, "Mass screening for lung cancer with mobile spiral computed tomography scanner," *The Lancet*, vol. 351, no. 9111, pp. 1242–1245, 1998.

[13] M. N. Gurcan, B. Sahiner, N. Petrick, H.-P. Chan, E. A. Kazerooni, P. N. Cascade, and L. Hadjiiski, "Lung nodule detection on thoracic computed tomography images: preliminary evaluation of a computer-aided diagnosis system," *Med. Phy.*, vol. 29, no. 11, pp. 2552–2558, 2002.

[14] T. Sobue, N. Moriyama, M. Kaneko, M. Kusumoto, T. Kobayashi, R. Tsuchiya, R. Kakinuma, H. Ohmatsu, K. Nagai, and H. Nishiyama, "Screening for lung cancer with low-dose helical computed tomography: anti-lung cancer association project," *J. Clinic. Oncol.*, vol. 20, no. 4, pp. 911–920, 2002.

[15] P. B. Bach, J. R. Jett, U. Pastorino, M. S. Tockman, S. J. Swensen, and C. B. Begg, "Computed tomography screening and lung cancer outcomes," *Jama*, vol. 297, no. 9, pp. 953–961, 2007.

[16] G. Brancatelli, S. Katyal, M. P. Federle, and P. Fontes, "Three-dimensional multislice helical computed tomography with the volume rendering technique in the detection of vascular complications after liver transplantation," *Transplantation*, vol. 73, no. 2, pp. 237–242, 2002.

[17] C. Dromain, T. de Baere, J. Lumbroso, H. Caillet, A. Laplanche, V. Boige, M. Ducreux, P. Duvillard, D. Elias, and M. Schlumberger, "Detection of liver metastases from endocrine tumors: a prospective comparison of somatostatin receptor scintigraphy, computed tomography, and magnetic resonance imaging," *J. Clini. Oncol.*, vol. 23, no. 1, pp. 70–78, 2005.

[18] K. L. Killeen, K. Shanmuganathan, P. A. Poletti, C. Cooper, and S. E. Mirvis, "Helical computed tomography of bowel and mesenteric injuries," *J. Trauma Acute Care Surg.*, vol. 51, no. 1, pp. 26–36, 2001.

[19] D. Beall, B. Fortman, B. Lawler, and F. Regan, "Imaging bowel obstruction: a comparison between fast magnetic resonance imaging and helical computed tomography," *Clinic. Radiol.*, vol. 57, no. 8, pp. 719–724, 2002.

[20] T. Uematsu, M. Sano, K. Homma, M. Shiina, and S. Kobayashi, "Three-dimensional helical ct of the breast: accuracy for measuring extent of breast cancer candidates for breast conserving surgery," *Breast Canc. Resear. Treat.*, vol. 65, no. 3, pp. 249–257, 2001.

[21] S. Akashi-Tanaka, T. Fukutomi, N. Sato, and K. Miyakawa, "The role of computed tomography in the selection of breast cancer treatment," *Breast Can*, vol. 10, no. 3, pp. 198–203, 2003.

[22] J. L. Antevil, M. J. Sise, D. I. Sack, B. Kidder, A. Hopper, and C. V. Brown, "Spiral computed tomography for the initial evaluation of spine trauma: a new standard of care?" *Journal of Trauma and Acute Care Surgery*, vol. 61, no. 2, pp. 382–387, 2006.

[23] K. Wiesent, K. Barth, N. Navab, P. Durlak, T. Brunner, T. Schuetz, and W. Seissler, "Enhanced 3-D-reconstruction algorithm for C-arm systems suitable for interventional procedures," *IEEE Trans. Med. Imag.*, vol. 19, pp. 391 – 403, 2000.

[24] G. Lauritsch, J. Boese, L. Wigström, H. Kemeth, and R. Fahrig, "Towards cardiac C-arm computed tomography," *IEEE Trans. Med. Imag.*, vol. 25, no. 7, pp. 922–934, 2006.

[25] M. J. Wallace, M. D. Kuo, C. Glaiberman, C. A. Binkert, R. C. Orth, and G. Soulez, "Three-dimensional C-arm cone-beam CT: applications in the interventional suite," *J. Vasc. Interv. Radiol.*, vol. 19, no. 6, pp. 799–813, 2008.

[26] R. C. Orth, M. J. Wallace, and M. D. Kuo, "C-arm cone-beam CT: general principles and technical considerations for use in interventional radiology," *J. Vasc. Interv. Radiol.*, vol. 19, no. 6, pp. 814–820, 2008.

[27] M. Grass, R. Koppe, E. Klotz, R. Proksa, M. Kuhn, H. Aerts, J. O. De Beek, and R. Kemkers, "Three-dimensional reconstruction of high contrast objects using C-arm image intensifier projection data," *Comput. Med. Imag. Grap.*, vol. 23, no. 6, pp. 311–321, 1999.

[28] J. H. Siewerdsen, D. J. Moseley, S. Burch, S. K. Bisland, A. Bogaards, B. C. Wilson, and D. A. Jaffray, "Volume CT with a flat-panel detector on a mobile, isocentric C-arm: Pre-clinical investigation in guidance of minimally invasive surgery," *Med. Phys.*, vol. 32, pp. 241–254, 2005.

[29] J. S. Hott, V. R. Deshmukh, J. D. Klopfenstein, V. K. Sonntag, C. A. Dickman, R. F. Spetzler, and S. M. Papadopoulos, "Intraoperative Iso-C C-arm navigation in craniospinal surgery: the first 60 cases," *Neurosur.*, vol. 54, no. 5, pp. 1131–1137, 2004.

[30] W. De Vos, J. Casselman, and G. Swennen, "Cone-beam computerized tomography (CBCT) imaging of the oral and maxillofacial region: a systematic review of the literature," *Int. J. Oral. Max. Surg.*, vol. 38, no. 6, pp. 609–625, 2009.

[31] D. A. Jaffray, J. H. Siewerdsen, J. W. Wong, and A. A. Martinez, "Flat-panel cone-beam computed tomography for image-guided radiation therapy," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 53, pp. 1337–1349, 2002.

[32] M. Oldham, D. Létourneau, L. Watt, G. Hugo, D. Yan, D. Lockman, L. H. Kim, P. Y. Chen, A. Martinez, and J. W. Wong, "Cone-beam-CT guided radiation therapy: A model for on-line application," *Radiother. Oncol.*, vol. 75, no. 3, pp. 271–E1, 2005.

[33] M. H. Smitsmans, J. De Bois, J.-J. Sonke, A. Betgen, L. J. Zijp, D. A. Jaffray, J. V. Lebesque, and M. Van Herk, "Automatic prostate localization on cone-beam CT scans for high precision image-guided radiotherapy," *Int. J. Radiat. Oncol.*, vol. 63, no. 4, pp. 975–984, 2005.

[34] I. S. Grills, G. Hugo, L. L. Kestin, A. P. Galerani, K. K. Chao, J. Wloch, and D. Yan, "Image-guided radiotherapy via daily online cone-beam CT substantially reduces margin requirements for stereotactic lung radiotherapy," *Int. J. Radiat. Oncol.*, vol. 70, no. 4, pp. 1045–1056, 2008.

[35] D. Jaffray and J. Siewerdsen, "Cone-beam computed tomography with a flat-panel imager: initial performance characterization," *Med. Phy.*, vol. 27, no. 6, pp. 1311–1323, 2000.

[36] S. Mori, R. Hara, T. Yanagi, G. C. Sharp, M. Kumagai, H. Asakura, R. Kishimoto, S. Yamada, S. Kandatsu, and T. Kamada, "Four-dimensional measurement of intrafractional respiratory motion of pancreatic tumors using a 256 multi-slice ct scanner," *Radioth. & Oncol.*, vol. 92, no. 2, pp. 231–237, 2009.

[37] F. J. Rybicki, H. J. Otero, M. L. Steigner, G. Vorobiof, L. Nallamshetty, D. Mitsouras, H. Ersoy, R. T. Mather, P. F. Judy, T. Cai *et al.*, "Initial evaluation of coronary images from 320-detector row computed tomography," *Internat. J Cardiov. Imag.*, vol. 24, no. 5, pp. 535–546, 2008.

[38] M. Dewey, E. Zimmermann, F. Deissenrieder, M. Laule, H.-P. Dübel, P. Schlattmann, F. Knebel, W. Rutsch, and B. Hamm, "Noninvasive coronary angiography by 320-row computed tomography with lower radiation exposure and maintained diagnostic accuracy comparison of results with cardiac catheterization in a head-to-head pilot investigation," *Circulation*, vol. 120, no. 10, pp. 867–875, 2009.

[39] T. Yamashiro, T. Miyara, M. Takahashi, A. Kikuyama, H. Kamiya, H. Koyama, Y. Ohno, H. Moriya, M. Matsuki, and Y. Tanaka, "Lung image quality with 320-row wide-volume ct scans: the effect of prospective ecg-gating and comparisons with 64-row helical ct scans," *Academic radiology*, vol. 19, no. 4, pp. 380–388, 2012.

[40] J. H. Johnston, D. J. Podberesky, T. T. Yoshizumi, E. Angel, G. Toncheva, D. B. Larson, J. C. Egelhoff, C. Anderson-Evans, G. B. Nguyen, and A. Barelli, "Comparison of radiation dose estimates, image noise, and scan duration in pediatric body imaging for volumetric and helical modes on 320-detector ct and helical mode on 64-detector ct," *Pediatric radiology*, vol. 43, no. 9, pp. 1117–1127, 2013.

[41] S. P. Jadhav, F. Golriz, L. A. Atweh, W. Zhang, and R. Krishnamurthy, "Ct angiography of neonates and infants: comparison of radiation dose and image quality of target mode prospectively ecg-gated 320-mdct and ungated helical 64-mdct," *American Journal of Roentgenology*, vol. 204, no. 2, pp. W184–W191, 2015.

[42] T. Kido, A. Kurata, H. Higashino, Y. Sugawara, H. Okayama, J. Higaki, H. Anno, K. Katada, S. Mori, S. Tanada *et al.*, "Cardiac imaging using 256-detector row four-dimensional ct: preliminary clinical report," *Radiation medicine*, vol. 25, no. 1, pp. 38–44, 2007.

[43] S. Mori, K. Nishizawa, C. Kondo, M. Ohno, K. Akahane, and M. Endo, "Effective doses in subjects undergoing computed tomography cardiac imaging with the 256-multislice ct scanner," *Europ. J Radiol.*, vol. 65, no. 3, pp. 442–448, 2008.

[44] R. Fahrig, A. Fox, S. Lownie, and D. Holdsworth, "Use of a c-arm system to generate true three-dimensional computed rotational angiograms: preliminary in vitro and in vivo results." *Americ. J. Neuro.*, vol. 18, no. 8, pp. 1507–1514, 1997.

[45] U. Linsenmaier, C. Rock, E. Euler, S. Wirth, R. Brandl, D. Kotsianos, W. Mutschler, and K. J. Pfeifer, "Three-dimensional ct with a modified c-arm image intensifier: feasibility 1," *Radiology*, vol. 224, no. 1, pp. 286–292, 2002.

[46] J. Siewerdsen, D. Moseley, S. Burch, S. Bisland, A. Bogaards, B. Wilson, and D. Jaffray, "Volume CT with a flat-panel detector on a mobile, isocentric C-arm: Pre-clinical investigation in guidance of minimally invasive surgery," *Med. Phys.*, vol. 32, pp. 241 – 254, 2005.

[47] M. A. Rafferty, J. H. Siewerdsen, Y. Chan, D. J. Moseley, M. J. Daly, D. A. Jaffray, and J. C. Irish, "Investigation of C-arm cone-beam CT-guided surgery of the frontal recess," *Laryngoscope*, vol. 115, pp. 2138–2143, 2005.

[48] D. Létourneau, J. W. Wong, M. Oldham, M. Gulam, L. Watt, D. A. Jaffray, J. H. Siewerdsen, and A. A. Martinez, "Cone-beam-ct guided radiation therapy: technical implementation," *Radiotherapy and oncology*, vol. 75, no. 3, pp. 279–286, 2005.

[49] L. Xing, B. Thorndyke, E. Schreibmann, Y. Yang, T. F. Li, G. Kim, G. Luxton, and A. Koong, "Overview of image-guided radiation therapy," *Med. Dosi.*, vol. 31.2, pp. 91–112, 2006.

[50] L. A. Dawson and D. A. Jaffray, "Advances in Image-Guided Radiation Therapy," *Ame. Soc. Clin. Onc.*, vol. 25.8, pp. 938–946, 2007.

[51] T. S. Hong, W. A. Tomé, R. J. Chappell, P. Chinnaiyan, M. P. Mehta, and P. M. Harari, "The impact of daily setup variations on head-and-neck intensity-modulated radiation therapy," *Internat. J. Radiat. Oncol. Bio. Phy.*, vol. 61, no. 3, pp. 779–788, 2005.

[52] M. Guckenberger and M. Flentje, "Intensity-modulated radiotherapy (imrt) of localized prostate cancer," *Strahlentherapie und Onkologie*, vol. 183, no. 2, pp. 57–62, 2007.

[53] J. Wang, S. Bai, N. Chen, F. Xu, X. Jiang, Y. Li, Q. Xu, Y. Shen, H. Zhang, Y. Gong *et al.*, "The clinical feasibility and effect of online cone beam computer tomography-guided intensity-modulated radiotherapy for nasopharyngeal cancer," *Radioth. & Oncol.*, vol. 90, no. 2, pp. 221–227, 2009.

[54] R. B. Den, A. Doemer, G. Kubicek, G. Bednarz, J. M. Galvin, W. M. Keane, Y. Xiao, and M. Machtay, "Daily image guidance with cone-beam computed tomography for head-and-neck cancer intensity-modulated radiotherapy: a prospective study," *Internat. J. Radiat. Oncol. Bio. Phy.*, vol. 76, no. 5, pp. 1353–1359, 2010.

[55] G. X. Ding, D. M. Duggan, C. W. Coffey, M. Deeley, D. E. Hallahan, A. Cmelak, and A. Malcolm, "A study on adaptive IMRT treatment planning using kV cone-beam CT," *Radioth. & Oncol.*, vol. 85, no. 1, pp. 116–125, 2007.

[56] J. Nijkamp, F. J. Pos, T. T. Nuver, R. De Jong, P. Remeijer, J.-J. Sonke, and J. V. Lebesque, "Adaptive radiotherapy for prostate cancer using kilovoltage cone-beam computed tomography: first clinical results," *Internat. J. Radiat. Oncol. Bio. Phy.*, vol. 70, no. 1, pp. 75–82, 2008.

[57] D. Paquin, D. Levy, and L. Xing, "Multiscale registration of planning CT and daily cone beam CT images for adaptive radiation therapy," *Medical physics*, vol. 36, no. 1, pp. 4–11, 2009.

[58] J. Ludlow, L. Ludlow, S. Brooks, and W. Howerton, "Dosimetry of 3 CBCT devices for oral and maxillofacial radiology: CB Mercuray, NewTom 3G and i-CAT," *Dentomaxillofacial Radio.*, vol. 35, pp. 219 – 226, 2005.

[59] J. Ludlow and M. Ivanovic, "Comparative dosimetry of dental CBCT devices and 64-slice CT for oral and maxillofacial radiology," *Oral surg., Oral med., Oral Path., Oral Radio., & Endoont.*, vol. 106, pp. 106 – 114, 2008.

[60] J. Ludlow, L. Ludlow, and S. Brooks, "Dosimetry of two extraoral direct digital imaging devices: NewTom cone beam CT and Orthophos Plus DS panoramic unit," *Dentomaxillofacial Radio.*, vol. 32, pp. 229 – 234, 2003.

[61] J. Roberts, N. Drage, J. Davies, and D. Thomas, "Effective dose from cone beam CT examinations in dentistry," *British journ. Radio.*, vol. 82, pp. 35 – 40, 2008.

[62] P. Kaur, "Evaluation radiographic orthodontic records image quality derived from cbct," 2014.

[63] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "Pet-ct image registration in the chest using free-form deformations," *IEEE Trans. Med. Imag.*, vol. 22, no. 1, pp. 120–128, 2003.

[64] G. Antoch, J. Stattaus, A. T. Nemat, S. Marnitz, T. Beyer, H. Kuehl, A. Bockisch, J. F. Debatin, and L. S. Freudenberg, "Non–small cell lung cancer: Dual-modality pet/ct in preoperative staging 1," *Radiology*, vol. 229, no. 2, pp. 526–533, 2003.

[65] R. Bar-Shalom, N. Yefremov, L. Guralnik, D. Gaitini, A. Frenkel, A. Kuten, H. Altman, Z. Keidar, and O. Israel, "Clinical performance of pet/ct in evaluation of cancer: additional value for diagnostic imaging and patient management," *J Nucl. Med.*, vol. 44, no. 8, pp. 1200–1209, 2003.

[66] K. Tharp, O. Israel, J. Hausmann, L. Bettman, W. Martin, M. Daitzchman, M. Sandler, and D. Delbeke, "Impact of 131i-spect/ct images obtained with an integrated system in the follow-up of patients with thyroid carcinoma," *Europ. J Nucl. Med. Molecul. Imag.*, vol. 31, no. 10, pp. 1435–1442, 2004.

[67] D. Utsunomiya, S. Shiraishi, M. Imuta, S. Tomiguchi, K. Kawanaka, S. Morishita, K. Awai, and Y. Yamashita, "Added value of spect/ct fusion in assessing suspected bone metastasis: Comparison with scintigraphy alone and nonfused scintigraphy and ct 1," *Radiology*, vol. 238, no. 1, pp. 264–271, 2006.

[68] S. M. Jorgensen, O. Demirkaya, and E. L. Ritman, "Three-dimensional imaging of vasculature and parenchyma in intact rodent organs with x-ray micro-ct," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 275, no. 3, pp. H1103–H1114, 1998.

[69] E. L. Ritman, "Micro-computed tomographyâĂŤcurrent status and developments," *Annu. Rev. Biomed. Eng.*, vol. 6, pp. 185–208, 2004.

[70] J. M. Boone, T. R. Nelson, K. K. Lindfors, and J. A. Seibert, "Dedicated breast CT: Radiation dose and image quality evaluation," *Radiol.*, vol. 221, pp. 657–667, 2001.

[71] B. Chen and R. Ning, "Cone-beam volume CT breast imaging: Feasibility study," *Med. Phys.*, vol. 29, pp. 755–770, 2002.

[72] J. Bian, K. Yang, J. M. Boone, X. Han, E. Y. Sidky, and X. Pan, "Investigation of iterative image reconstruction in low-dose breast CT," *Phys. Med. Biol.*, vol. 59, no. 11, p. 2659, 2014.

[73] W. Zbijewski, P. De Jean, P. Prakash, Y. Ding, J. Stayman, N. Packard, R. Senn, D. Yang, J. Yorkston, A. Machado *et al.*, "A dedicated cone-beam ct system for musculoskeletal extremities imaging: design, optimization, and initial performance characterization," *Med. Phy.*, vol. 38, no. 8, pp. 4700–4713, 2011.

[74] P. Prakash, W. Zbijewski, G. Gang, Y. Ding, J. Stayman, J. Yorkston, J. Carrino, and J. Siewerdsen, "Task-based modeling and optimization of a cone-beam ct scanner for musculoskeletal imaging," *Med. Phy.*, vol. 38, no. 10, pp. 5612–5629, 2011.

[75] S. K. Koskinen, V. V. Haapamäki, J. Salo, N. C. Lindfors, M. Kortesniemi, L. Seppälä, and K. T. Mattila, "CT arthrography of the wrist using a novel, mobile, dedicated extremity cone-beam CT (CBCT)," *Skelet. Radiol.*, vol. 42, no. 5, pp. 649–657, 2013.

[76] E. K. Tuominen, J. Kankare, S. K. Koskinen, and K. T. Mattila, "Weight-bearing ct imaging of the lower extremity," *Americ. J. Roentgen.*, vol. 200, no. 1, pp. 146–148, 2013.

[77] Z. Rumboldt, W. Huda, and J. All, "Review of portable ct with assessment of a dedicated head ct scanner," *Americ. J. Neuro.*, vol. 30, no. 9, pp. 1630–1636, 2009.

[78] L. Yu, T. J. Vrieze, M. R. Bruesewitz, J. M. Kofler, D. R. DeLone, J. F. Pallanch, E. P. Lindell, and C. H. McCollough, "Dose and image quality evaluation of a dedicated cone-beam ct system for high-contrast neurologic applications," *Americ. J. Roentgen.*, vol. 194, no. 2, pp. W193–W201, 2010.

[79] M. Kamran, S. Nagaraja, and J. V. Byrne, "C-arm flat detector computed tomography: the technique and its applications in interventional neuro-radiology," *Neuroradiology*, vol. 52, no. 4, pp. 319–327, 2010.

[80] E. D. Pisano, C. Gatsonis, E. Hendrick, M. Yaffe, J. K. Baum, S. Acharyya, E. F. Conant, L. L. Fajardo, L. Bassett, C. D'Orsi *et al.*, "Diagnostic performance of digital versus film mammography for breast-cancer screening," *New Eng. J Med.*, vol. 353, no. 17, pp. 1773–1783, 2005.

[81] L. T. Niklason, B. T. Christian, L. E. Niklason, D. B. Kopans, D. E. Castleberry, B. Opsahl-Ong, C. E. Landberg, P. J. Slanetz, A. A. Giardino, R. Moore *et al.*, "Digital tomosynthesis in breast imaging." *Radiology*, vol. 205, no. 2, pp. 399–406, 1997.

[82] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am. A*, vol. 1, pp. 612–619, 1984.

[83] A. Katsevich, "Theoretically exact filtered backprojection-type inversion algorithm for spiral CT," *SIAM J. Appl. Math.*, vol. 62, pp. 2012–2026, 2002.

[84] Y. Zou and X. Pan, "Exact image reconstruction on PI-lines from minimum data in helical cone-beam CT," *Phys. Med. Biol.*, vol. 49, pp. 941–959, 2004.

[85] H. K. Tuy, "An inversion formula for cone-beam reconstruction," *SIAM J. Appl. Math.*, vol. 43, pp. 546–552, 1983.

[86] D. L. Parker, "Optimal short scan convolution reconstruction for fan beam CT," *Med. Phys.*, vol. 9, no. 2, pp. 254–257, 1982.

[87] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography," *J. Theor. Biol.*, vol. 29, pp. 471–481, 1970.

[88] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *J Royal Stat. Socit. Series B*, pp. 1–38, 1977.

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

[90] S. L. Rauch, B. A. van der Kolk, R. E. Fisler, N. M. Alpert, S. P. Orr, C. R. Savage, A. J. Fischman, M. A. Jenike, and R. K. Pitman, "A symptom provocation study of posttraumatic stress disorder using positron emission tomography and script-driven imagery," *Archives of General Psychiatry*, vol. 53, no. 5, pp. 380–387, 1996.

[91] R. S. Frackowiak, G.-L. Lenzi, T. Jones, and J. D. Heather, "Quantitative measurement of regional cerebral blood flow and oxygen metabolism in man using 15o and positron emission tomography: theory, procedure, and normal values." *Journal of computer assisted tomography*, vol. 4, no. 6, pp. 727–736, 1980.

[92] H. T. Chugani, M. E. Phelps, and J. C. Mazziotta, "Positron emission tomography study of human brain functional development," *Annals of neurology*, vol. 22, no. 4, pp. 487–497, 1987.

[93] E. Tulving, S. Kapur, F. Craik, M. Moscovitch, and S. Houle, "Hemispheric encoding/retrieval asymmetry in episodic memory: positron emission tomography findings," *Proceedings of the National Academy of Sciences*, vol. 91, no. 6, pp. 2016–2020, 1994.

[94] S. S. Gambhir, "Molecular imaging of cancer with positron emission tomography," *Nature Reviews Cancer*, vol. 2, no. 9, pp. 683–693, 2002.

[95] S. T. Grafton, M. A. Arbib, L. Fadiga, and G. Rizzolatti, "Localization of grasp representations in humans by positron emission tomography," *Experimental Brain Research*, vol. 112, no. 1, pp. 103–111, 1996.

[96] M. Corbetta, F. M. Miezin, S. Dobmeyer, G. L. Shulman, and S. E. Petersen, "Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography," *The Journal of Neuroscience*, vol. 11, no. 8, pp. 2383–2402, 1991.

[97] M. K. Gould, C. C. Maclean, W. G. Kuschner, C. E. Rydzak, and D. K. Owens, "Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions: a meta-analysis," *Jama*, vol. 285, no. 7, pp. 914–924, 2001.

[98] M. E. Juweid, S. Stroobants, O. S. Hoekstra, F. M. Mottaghy, M. Dietlein, A. Guermazi, G. A. Wiseman, L. Kostakoglu, K. Scheidhauer, A. Buck *et al.*, "Use of positron emission tomography for response assessment of lymphoma: consensus of the imaging subcommittee of international harmonization project in lymphoma," *Journal of Clinical Oncology*, vol. 25, no. 5, pp. 571–578, 2007.

[99] D. Lardinois, W. Weder, T. F. Hany, E. M. Kamel, S. Korom, B. Seifert, G. K. von Schulthess, and H. C. Steinert, "Staging of non–small-cell lung cancer with integrated positron-emission tomography and computed tomography," *New England Journal of Medicine*, vol. 348, no. 25, pp. 2500–2507, 2003.

[100] D. T. Delpy, M. Cope, P. van der Zee, S. Arridge, S. Wray, and J. Wyatt, "Estimation of optical pathlength through tissue from direct time of flight measurement," *Physics in medicine and biology*, vol. 33, no. 12, p. 1433, 1988.

[101] D. Rischin, R. J. Hicks, R. Fisher, D. Binns, J. Corry, S. Porceddu, and L. J. Peters, "Prognostic significance of [18f]-misonidazole positron emission tomography–detected tumor hypoxia in patients with advanced head and neck cancer randomly assigned to chemoradiation with or without tirapazamine: a substudy of trans-tasman radiation oncology group study 98.02," *Journal of Clinical Oncology*, vol. 24, no. 13, pp. 2098–2104, 2006.

[102] R. Cabeza, C. L. Grady, L. Nyberg, A. R. McIntosh, E. Tulving, S. Kapur, J. M. Jennings, S. Houle, and F. I. Craik, "Age-related differences in neural activity during memory encoding and retrieval: a positron emission tomography study," *The Journal of Neuroscience*, vol. 17, no. 1, pp. 391–400, 1997.

[103] N. L. Foster, T. N. Chase, P. Fedio, N. J. Patronas, R. A. Brooks, and G. Di Chiro, "Alzheimer's disease focal cortical changes shown by positron emission tomography," *Neurology*, vol. 33, no. 8, pp. 961–961, 1983.

[104] R. Duara, C. Grady, J. Haxby, M. Sundaram, N. Cutler, L. Heston, A. Moore, N. Schlageter, S. Larson, and S. Rapoport, "Positron emission tomography in alzheimer's disease," *Neurology*, vol. 36, no. 7, pp. 879–879, 1986.

[105] N. Tamaki, Y. Yonekura, K. Yamashita, H. Saji, Y. Magata, M. Senda, Y. Konishi, K. Hirata, T. Ban, and J. Konishi, "Positron emission tomography using fluorine-18 deoxyglucose in evaluation of coronary artery bypass grafting," *The American journal of cardiology*, vol. 64, no. 14, pp. 860–865, 1989.

[106] L. A. Piérard, C. M. De Landsheere, C. Berthe, P. Rigo, and H. E. Kulbertus, "Identification of viable myocardium by echocardiography during dobutamine infusion in patients with myocardial infarction after thrombolytic therapy: comparison with positron emission tomography," *Journal of the American College of Cardiology*, vol. 15, no. 5, pp. 1021–1031, 1990.

[107] D. Eitzman, Z. Al-Aouar, H. L. Kanter, J. vom Dahl, M. Kirsh, G. M. Deeb, and M. Schwaiger, "Clinical outcome of patients with advanced coronary artery disease after viability studies with positron emission tomography," *Journal of the American College of Cardiology*, vol. 20, no. 3, pp. 559–565, 1992.

[108] J. C. Wu, I. Y. Chen, G. Sundaresan, J.-J. Min, A. De, J.-H. Qiao, M. C. Fishbein, and S. S. Gambhir, "Molecular imaging of cardiac cell transplantation in living animals using optical bioluminescence and positron emission tomography," *Circulation*, vol. 108, no. 11, pp. 1302–1305, 2003.

[109] I. Y. Chen, J. C. Wu, J.-J. Min, G. Sundaresan, X. Lewis, Q. Liang, H. R. Herschman, and S. S. Gambhir, "Micro–positron emission tomography imaging of cardiac gene expression in rats using bicistronic adenoviral vector-mediated gene delivery," *Circulation*, vol. 109, no. 11, pp. 1415–1420, 2004.

[110] A. A. Memon, S. Jakobsen, F. Dagnaes-Hansen, B. S. Sorensen, S. Keiding, and E. Nexo, "Positron emission tomography (pet) imaging with [11c]-labeled erlotinib: a micro-pet study on mice with lung tumor xenografts," *Cancer research*, vol. 69, no. 3, pp. 873–878, 2009.

[111] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. 1, pp. 113–122, 1982.

[112] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography." *J. Comput. Assist. Tomogr.*, vol. 8, pp. 306–316, 1984.

[113] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601–609, 1994.

[114] J. Browne and A. R. De Pierro, "A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography," *IEEE Trans. Med. Imag.*, vol. 15.5, pp. 687 – 699, 1996.

[115] T. Hsiao, A. Rangarajan, and G. R. Gindi, "Provably convergent osem-like reconstruction algorithm for emission tomography," in *Medical Imaging 2002*. International Society for Optics and Photonics, 2002, pp. 10–19.

[116] T. Hsiao, A. Rangarajan, P. Khurd, and G. Gindi, "An accelerated convergent ordered subsets algorithm for emission tomography," *Physics in medicine and biology*, vol. 49, no. 11, p. 2145, 2004.

[117] T. Hsiao, P. Khurd, A. Rangarajan, and G. Gindi, "An overview of fast convergent ordered-subsets reconstruction methods for emission tomography based on the incremental em algorithm," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 569, no. 2, pp. 429–433, 2006.

[118] E. Tanaka and H. Kudo, "Subset-dependent relaxation in block-iterative algorithms for image reconstruction in emission tomography," *Physics in medicine and biology*, vol. 48, no. 10, p. 1405, 2003.

[119] H. H. Barrett, T. White, and L. C. Parra, "List-mode likelihood," *J. Opt. Soc. Amer.*, vol. 14, no. 11, pp. 2914–2923, 1997.

[120] R. H. Huesman, G. J. Klein, W. W. Moses, J. Qi, B. W. Reutter, and P. R. Virador, "List-mode maximum-likelihood reconstruction applied to positron emission mammography (PEM) with irregular sampling," *IEEE Trans. Med. Imag.*, vol. 19, no. 5, pp. 532–537, 2000.

[121] R. Levkovilz, D. Falikman, M. Zibulevsky, A. Ben-Tal, and A. Nemirovski, "The design and implementation of COSEN, an iterative algorithm for fully 3-D listmode data," *IEEE Trans. Med. Imag.*, vol. 20, no. 7, pp. 633–642, 2001.

[122] A. Rahmim, M. Lenox, A. Reader, C. Michel, Z. Burbar, T. Ruth, and V. Sossi, "Statistical list-mode image reconstruction for the high resolution research tomograph," *Phys. Med. Biol.*, vol. 49, no. 18, p. 4239, 2004.

[123] C.-M. Kao, J. T. Yap, J. Mukherjee, and M. N. Wernick, "Image reconstruction for dynamic PET based on low-order approximation and restoration of the sinogram," *IEEE Trans. Med. Imag.*, vol. 16, pp. 738–749, 1997.

[124] J. Matthews, D. Bailey, P. Price, and V. Cunningham, "The direct calculation of parametric images from dynamic PET data using maximum-likelihood iterative reconstruction," *Phys. Med. Biol.*, vol. 42, no. 6, p. 1155, 1997.

[125] S. R. Meikle, J. C. Matthews, V. J. Cunningham, D. L. Bailey, L. Livieratos, T. Jones, and P. Price, "Parametric image reconstruction using spectral analysis of PET projection data," *Phys. Med. Biol.*, vol. 43, no. 3, p. 651, 1998.

[126] T. E. Nichols, J. Qi, E. Asma, and R. M. Leahy, "Spatiotemporal reconstruction of list-mode PET data," *IEEE Trans. Med. Imag.*, vol. 21, no. 4, pp. 396–404, 2002.

[127] A. J. Reader, F. C. Sureau, C. Comtat, R. Trébossen, and I. Buvat, "Joint estimation of dynamic PET images and temporal basis functions using fully 4D ML-EM," *Phys. Med. Biol.*, vol. 51, no. 21, p. 5455, 2006.

[128] C. Tsoumpas, F. E. Turkheimer, and K. Thielemans, "A survey of approaches for direct parametric image reconstruction in emission tomography," *Med. Phys.*, vol. 35, no. 9, pp. 3963–3971, 2008.

[129] G. Wang and J. Qi, "Generalized algorithms for direct reconstruction of parametric images from dynamic PET data," *IEEE Trans. Med. Imag.*, vol. 28, no. 11, pp. 1717–1726, 2009.

[130] Y. Rakvongthai, J. Ouyang, B. Guerin, Q. Li, N. M. Alpert, and G. El Fakhri, "Direct reconstruction of cardiac PET kinetic parametric images using a preconditioned conjugate gradient approach," *Med. Phys.*, vol. 40, no. 10, p. 102501, 2013.

[131] J. Ouyang, Q. Li, and G. El Fakhri, "Magnetic resonance-based motion correction for positron emission tomography imaging," *Semin. in Nucl. Med.*, vol. 43, no. 1, pp. 60–67, 2013.

[132] W. Zhu, Q. Li, B. Bai, P. S. Conti, and R. M. Leahy, "Patlak image estimation from dual time-point list-mode PET data," *IEEE Trans. Med. Imag.*, vol. 33, pp. 913–924, 2014.

[133] R. Manber, K. Thielemans, B. Hutton, A. Barnes, S. Ourselin, S. Arridge, C. O'Meara, S. Wan, and D. Atkinson, "Practical PET Respiratory Motion Correction in Clinical PET/MR," *J. Nucl. Med.*, p. 114, 2015.

[134] P. E. Kinahan and J. Rogers, "Analytic 3D image reconstruction using all detected events," *IEEE Trans. Nucl. Sci.*, vol. 36, no. 1, pp. 964–968, 1989.

[135] S. Matej, G. Herman, T. Narayan, S. Furuie, R. Lewitt, and P. Kinahan, "Evaluation of task-oriented performance of several fully 3D PET reconstruction algorithms," *Phys. Med. Biol.*, vol. 39, no. 3, p. 355, 1994.

[136] E. Ü. Mumcuoğlu, R. Leahy, S. R. Cherry, and Z. Zhou, "Fast gradient-based methods for bayesian reconstruction of transmission and emission PET images," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 687–701, 1994.

[137] J. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 13, no. 2, pp. 290–300, 1994.

[138] W. Wang and G. Gindi, "Noise analysis of MAP-EM algorithms for emission tomography," *Phys. Med. Biol.*, vol. 42, pp. 2215–2232, 1997.

[139] X. Liu, C. Comtat, C. Michel, P. Kinahan, M. Defrise, and D. Townsend, "Comparison of 3-D reconstruction with 3D-OSEM and with FORE+ OSEM for PET," *IEEE Trans. Med. Imag.*, vol. 20, no. 8, pp. 804–814, 2001.

[140] D. J. Kadrmas, "LOR-OSEM: statistical PET reconstruction from raw line-of-response histograms," *Phys. Med. Biol.*, vol. 49, no. 20, p. 4731, 2004.

[141] J. Qi and R. M. Leahy, "Iterative reconstruction techniques in emission computed tomography," *Phys. Med. Biol.*, vol. 51, no. 15, p. R541, 2006.

[142] J. A. Fessler and A. O. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms," *IEEE Trans. Imag. Proc.*, vol. 4, pp. 1417–1429, 1995.

[143] D. F. Yu and J. A. Fessler, "Edge-preserving tomographic reconstruction with nonlocal regularization," *Medical Imaging, IEEE Transactions on*, vol. 21, no. 2, pp. 159–173, 2002.

[144] X. Ouyang, W. H. Wong, V. E. Johnson, X. Hu, and C.-T. Chen, "Incorporation of correlated structural images in pet image reconstruction," *Medical Imaging, IEEE Transactions on*, vol. 13, no. 4, pp. 627–640, 1994.

[145] J. Cheng-Liao and J. Qi, "Pet image reconstruction with anatomical edge guided level set prior," *Physics in medicine and biology*, vol. 56, no. 21, p. 6899, 2011.

[146] A. Mehranian, A. Rahmim, M. R. Ay, F. Kotasidis, and H. Zaidi, "An ordered-subsets proximal preconditioned gradient algorithm for edge-preserving PET image reconstruction," *Med. Phys.*, vol. 40, no. 5, p. 052503, 2013.

[147] J. Qi, "Comparison of statistical reconstructions with isotropic and anisotropic resolution," in *Nuclear Science Symposium Conference Record, 2004 IEEE*, vol. 6. IEEE, 2004, pp. 3624–3628.

[148] E. Asma, R. Manjeshwar, and K. Thielemans, "Theoretical comparison of motion correction techniques for PET image reconstruction," in *IEEE NSS/MIC conf. Rec.*, vol. 3. IEEE, 2006, pp. 1762–1767.

[149] G. Wang, L. Fu, and J. Qi, "Maximum a posteriori reconstruction of the Patlak parametric image from sinograms in dynamic PET," *Phys. Med. Biol.*, vol. 53, no. 3, p. 593, 2008.

[150] R. H. Johnson, H. Hu, S. T. Haworth, P. S. Cho, C. A. Dawson, and J. H. Linehan, "Feldkamp and circle-and-line cone-beam reconstruction for 3D micro-CT of vascular networks," *Phys. Med. Biol.*, vol. 43, no. 4, p. 929, 1998.

[151] A. Delaney, Y. Bresler, and C. Sunnyvale, "Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography," *IEEE Trans. Imag. Proc.*, vol. 7, pp. 204–221, 1998.

[152] I. Elbakri and J. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE Trans. Med. Imag.*, vol. 21, no. 2, pp. 89–99, 2002.

[153] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, pp. 4777–4807, 2008.

[154] X. Pan, E. Y. Sidky, and M. Vannier, "Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?" *Inverse Probl.*, vol. 25, p. 123009, 2009.

[155] J. Tang, B. E. Nett, and G.-H. Chen, "Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms," *Phys. Med. Biol.*, vol. 54, pp. 5781–5804, 2009.

[156] J. Bian, J. H. Siewerdsen, X. Han, E. Y. Sidky, J. L. Prince, C. A. Pelizzari, and X. Pan, "Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT," *Phys. Med. Biol.*, vol. 55, pp. 6575–6599, 2010.

[157] D. Stsepankou, A. Arns, S. Ng, P. Zygmanski, and J. Hesser, "Evaluation of robustness of maximum likelihood cone-beam CT reconstruction with total variation regularization," *Phys. Med. Biol.*, vol. 57, no. 19, p. 5955, 2012.

[158] A. S. Wang, J. W. Stayman, Y. Otake, G. Kleinszig, S. Vogt, G. L. Gallia, A. J. Khanna, and J. H. Siewerdsen, "Soft-tissue imaging with C-arm cone-beam CT using statistical reconstruction," *Phys. Med. Biol.*, vol. 59, no. 4, p. 1005, 2014.

[159] H. Nien and J. A. Fessler, "Fast splitting-based ordered-subsets X-ray CT image reconstruction," *Proc. 3rd Intl. Mtg. Imag. Form. X. CT*, pp. 291–4, 2014.

[160] M. Grass, T. Köhler, and R. Proksa, "Angular weighted hybrid cone-beam CT reconstruction for circular trajectories," *Phys. Med. Biol.*, vol. 46, no. 6, p. 1595, 2001.

[161] X. Tang, J. Hsieh, A. Hagiwara, R. A. Nilsen, J.-B. Thibault, and E. Drapkin, "A three-dimensional weighted cone beam filtered backprojection (CB-FBP) algorithm for image reconstruction in volumetric CT under a circular source trajectory," *Phys. Med. Biol.*, vol. 50, no. 16, p. 3889, 2005.

[162] S. Mori, M. Endo, S. Komatsu, S. Kandatsu, T. Yashiro, and M. Baba, "A combination-weighted Feldkamp-based reconstruction algorithm for cone-beam CT," *Phys. Med. Biol.*, vol. 51, no. 16, p. 3953, 2006.

[163] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, pp. 1 – 26, 2011.

[164] E. Y. Sidky, J. H. Jørgensen, and X. Pan, "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm," *Phys. Med. Biol.*, vol. 57, no. 10, p. 3065, 2012.

[165] E. Y. Sidky, C.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *J. X-Ray Sci. and Technol.*, vol. 14, pp. 119–139, 2006.

[166] J. S. Jørgensen, E. Y. Sidky, and X. Pan, "Quantifying admissible undersampling for sparsity-exploiting iterative image reconstruction in x-ray ct," *IEEE Trans. Med. Imag.*, vol. 32.2, pp. 460 – 473, 2013.

[167] E. Y. Sidky, D. N. Kraemer, E. G. Roth, C. Ullberg, I. S. Reiser, and X. Pan, "Analysis of iterative region-of-interest image reconstruction for X-ray computed tomography," *J. Med. Imag.*, vol. 1, no. 3, pp. 031 007–031 007, 2014.

[168] X. Han, E. Pearson, C. Pelizzari, H. Al-Hallaq, E. Y. Sidky, J. Bian, and X. Pan, "Algorithm-enabled exploration of image-quality potential of cone-beam CT in image-guided radiation therapy," *Phys. Med. Biol.*, vol. 60, no. 12, p. 4601, 2015.

[169] J.-J. Moreau, "Fonctions convexes duales et points proximaux dans un espace hilbertien," *CR Acad. Sci. Paris Sér. A Math*, vol. 255, pp. 2897–2899, 1962.

[170] F. Dennerlein, F. Noo, H. Schöndube, G. Lauritsch, and J. Hornegger, "A factorization approach for cone-beam reconstruction on a circular short-scan," *IEEE Trans. Med. Imag.*, vol. 27, no. 7, pp. 887–896, 2008.

[171] C. Maaß, F. Dennerlein, F. Noo, and M. Kachelrieß, "Comparing short scan CT reconstruction algorithms regarding cone-beam artifact performance," *IEEE Nucl. Sci. Sym. & Med. Imag. Conf. Record*, pp. 2188–2193, 2010.

[172] F. Natterer and K. Hadeler, "Efficient implementation of 'optimal' algorithms in computerized tomography," *Math. Method Appl. Sci.*, vol. 2, no. 4, pp. 545–555, 1980.

[173] F. Natterer and F. Wübbeling, *Mathematical methods in image reconstruction*. Siam, 2001.

[174] T. K. Lewellen, "Time-of-flight PET," *Semin. Nucl. Med.*, vol. 28, no. 3, pp. 268–275, 1998.

[175] W. W. Moses, "Recent advances and future advances in time-of-flight PET," *Nucl. Instr. Meth. Phy. Resch.*, vol. 580, no. 2, pp. 919–924, 2007.

[176] J. S. Karp, S. Surti, M. E. Daube-Witherspoon, and G. Muehllehner, "Benefit of time-of-flight in PET: experimental and clinical results," *J. Nucl. Med.*, vol. 49, no. 3, pp. 462–470, 2008.

[177] B. Jakoby, Y. Bercier, M. Conti, M. Casey, B. Bendriem, and D. Townsend, "Physical and clinical performance of the mCT time-of-flight PET/CT scanner," *Phys. Med. Biol.*, vol. 56, no. 8, p. 2375, 2011.

[178] R. Freifelder, J. S. Karp, M. Geagan, and G. Muehllehner, "Design and performance of the HEAD PENN-PET scanner," *IEEE Trans. Nucl. Sci.*, vol. 41, no. 4, pp. 1436–1440, 1994.

[179] Y. Shao, S. R. Cherry, K. Farahani, K. Meadors, S. Siegel, R. W. Silverman, and P. K. Marsden, "Simultaneous pet and mr imaging," *Phys. Med. Biol.*, vol. 42, pp. 1965–1970, 1997.

[180] W.-H. Wong, J. Uribe, H. Li, H. Baghaei, Y. Wang, M. Aykac, Y. Liu, T. Xing, D. Bilgen, and R. Farrell, "The design of a high-resolution transformable wholebody PET camera," *IEEE Trans. Nucl. Sci.*, vol. 49, no. 5, pp. 2079–2084, 2002.

[181] P. Crespo, T. Barthel, H. Frais-Kölbl, E. Griesmayer, K. Heidel, K. Parodi, J. Pawelke, and W. Enghardt, "Suppression of random coincidences during in-beam PET measurements at ion beam radiotherapy facilities," *IEEE Trans. Nucl. Sci.*, vol. 52, no. 4, pp. 980–987, 2005.

[182] Y.-C. Tai, A. Ruangma, D. Rowland, S. Siegel, D. F. Newport, P. L. Chow, and R. Laforest, "Performance evaluation of the microPET focus: a third-generation microPET scanner dedicated to animal imaging," *J. Nucl. Med.*, vol. 46, no. 3, pp. 455–463, 2005.

[183] C. S. Levin and H. Zaidi, "Current trends in preclinical PET system design," *PET Clinics*, vol. 2, pp. 125–160, 2007.

[184] T. Yamaya, T. Inaniwa, S. Minohara, E. Yoshida, N. Inadama, F. Nishikido, K. Shibuya, C. F. Lam, and H. Murayama, "A proposal of an open PET geometry," *Phys. Med. Biol.*, vol. 53, no. 3, p. 757, 2008.

[185] A. Salomon, D. Truhn, R. Botnar, F. Kiessling, and Y. Schulz, "Sparse crystal setting and large axial FOV for integrated whole-body PET/MR," *IEEE NSS/MIC conf. Rec.*, pp. 2521–2523, 2011.

[186] P. A. Wolf, J. S. Jørgensen, T. G. Schmidt, and E. Y. Sidky, "Few-view single photon emission computed tomography (SPECT) reconstruction based on a blurred piecewise constant object model," *Phys. Med. Biol.*, vol. 58, no. 16, p. 5629, 2013.

[187] C. C. Watson, D. Newport, and M. E. Casey, "A single scatter simulation technique for scatter correction in 3d pet," in *3D Img. Recon. Radiol. Nucl. Med.* Springer, 1996, pp. 255–268.

[188] R. Badawi, M. Miller, D. Bailey, and P. Marsden, "Randoms variance reduction in 3D PET," *Phys. Med. Biol.*, vol. 44, no. 4, p. 941, 1999.

[189] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, Priceton, New Jersey, 1999.

[190] S. Jan, G. Santin, D. Strul, S. Staelens, K. Assie, D. Autret, S. Avner, R. Barbier, M. Bardies, and P. Bloomfield, "GATE: a simulation toolkit for PET and SPECT," *Phys. Med. Biol.*, vol. 49, no. 19, p. 4543, 2004.

[191] National Electrical Manufacturers Association, *NEMA NU 2-2012: Performance Measurements of Positron Emission Tomographs*. National Electrical Manufactures Association, Washington, DC, 2012.

[192] C. Maab, M. Knaup, and M. Kachelrieb, "New approaches to region of interest computed tomography," *Med. Phys.*, vol. 38, pp. 2868–2878, 2011.

[193] A. Zamyatin, M. Dinu, and D. Shi, "Multi-scale iterative reconstruction," *Proc. MIC*, pp. MIC21.S–111, 2011.

[194] J. S. Jørgensen, E. Y. Sidky, P. C. Hansen, and X. Pan, "Quantitative study of undersampled reconverability for sparse images in computed tomography," 2012, arxiv preprint arxiv 1211.5658: http://arxiv.org/abs/1211.5658.

[195] X. Han, J. Bian, D. R. Eaker, T. L. Kline, E. Y. Sidky, E. L. Ritman, and X. Pan, "Algorithm-enabled low-dose micro-CT imaging," *IEEE Trans. Med. Imag.*, vol. 30, pp. 606–620, 2011.

[196] J. Bian, J. Wang, X. Han, E. Y. Sidky, J. Ye, L. Shao, and X. Pan, "Reconstruction from sparse data in offset-detector CBCT," in *Proceedings of the First International Meeting on Image Formation in X-Ray Computed Tomography*, 2010, pp. 96–100.

[197] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assisted Tomogr.*, vol. 8, pp. 306–316, 1984.

[198] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imaging*, vol. 13, pp. 601–609, 1994.

[199] M. J. Daly, J. H. Siewerdsen, D. J. Moseley, D. A. Jaffray, and J. C. Irish, "Intraoperative cone-beam CT for guidance of head and neck surgery: assessment of dose and image quality using a C-arm prototype," *Med. Phys.*, vol. 33, pp. 3767–3780, 2006.

[200] J. Bian, X. Han, E. Sidky, G. Cao, J. Lu, O. Zhou, and X. Pan, "Investigation of Sparse Data Mouse Imaging Using Micro-CT with a Carbon-Nanotube-Based X-ray Source," *Tsinghua Sci. Technol.*, vol. 15, pp. 74 – 78, 2010.

[201] X. Han, J. Bian, E. L. Ritman, E. Y. Sidky, and X. Pan, "Optimization-based reconstruction of sparse images from few-view projections," *Phys. Med. Biol.*, vol. 57, pp. 5245–5274, 2012.

[202] M. Li, H. Yang, and H. Kudo, "An accurate iterative reconstruction algorithm for sparse objects: application to 3D blood vessel reconstruction from a limited number of projections," *Phys. Med. Biol.*, vol. 47, pp. 2599–2609, 2002.

[203] Z. Zhang, X. Han, J. Bian, J. J. Manak, E. Y. Sidky, and X. Pan, "Initial experience in image reconstruction from limited-angle c-arm cbct data," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. IEEE, 2011, pp. 3977–3979.

[204] K. Hanson and G. Wecksung, "Bayesian approach to limited-angle reconstruction in computed tomography," *J. Opt. Soc. Am*, vol. 73, pp. 1501–1509, 1983.

[205] A. H. Andersen, "Algebraic reconstruction in CT from limited views," *IEEE Trans. Med. Imag.*, vol. 8.1, pp. 50–55, 1989.