Multi-channel computed tomographic image reconstruction by exploiting structural similarities

Suxer Lazara Alfonso Garcia

▶ **To cite this version:**

Suxer Lazara Alfonso Garcia. Multi-channel computed tomographic image reconstruction by exploiting structural similarities. Medical Imaging. Université de Bretagne occidentale - Brest, 2022. English. NNT : 2022BRES0020 . tel-03845150

HAL Id: tel-03845150
https://theses.hal.science/tel-03845150

Submitted on 9 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DOCTORAT BIOLOGIE BRETAGNE SANTE LOIRE

Université de Bretagne Occidentale

THESE DE DOCTORAT DE

L'UNIVERSITE
DE BRETAGNE OCCIDENTALE

ECOLE DOCTORALE N° 605
Biologie Santé
Spécialité : *Analyse et Traitement de l'Information et des Image Médicale*

Par
Suxer Lazara ALFONSO GARCIA

Multi-channel Computed Tomographic Image Reconstruction by Exploiting Structural Similarities

Thèse présentée et soutenue à Brest, le 31 Mars 2022
Unité de recherche : LATIM, INSERM, U1101

Rapporteurs avant soutenance :

Michel DEFRISE	Professeur	University Hospital AZ-VUB, Belgium.
Voichita MAXIM	Enseignant-Chercheur	CREATIS INSERM U1294, Lyon, France.

Composition du Jury :

Président :	Dimitris VISVIKIS	Directeur de recherche, INSERM, Brest, France.
Examinateurs :	Michel DEFRISE	Professeur, University Hospital AZ-VUB, Belgium.
	Voichita MAXIM	Enseignant-Chercheur, CREATIS INSERM U1294, Lyon, France.
Dir. de thèse :	Alexandre BOUSSE	Chargé de recherche, LATIM INSERM U1101, Brest, France.

Invité(s)
Alessandro PERELLI — Assistant professor, University of Dundee, United Kingdom.

DECLARATION OF AUTHORSHIP

I, Suxer Lazara ALFONSO GARCIA here by declare that the thesis entitled "Multi-channel Computed Tomographic Image Reconstruction by Exploiting Structural Similarities" submitted by me, for the award of the degree of *Doctor of Philosophy* to LATIM INSERM U1101, Université de Bretagne Occidentale, Brest, France is a record of the work carried out by me under the supervision of Alexande BOUSSE, PhD, HDR.

I declare that the work reported in this thesis has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

I declare that where I have consulted the published work of others has always been clearly attributed.

I declare that where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Place: Brest

Date:14/02/2022 **Suxer Lazara ALFONSO GARCIA**

ABSTRACT

The multi-channel joint reconstruction technique is a highly suited method for multi-modal medical imaging reconstruction. In the technique, the unknown images are reconstructed simultaneously by solving a single combined inverse problem and exploiting structural similarities between the images. The hypothesis behind this approach is that the image modalities inform each other during the reconstruction allowing artifact reduction and image quality enhancement. The present thesis develops three image reconstruction models for multi-channel image reconstruction. The first methodology consists of a Coupled Image-Motion Dictionary Learning algorithm for Motion Estimation-Compensation in Cone-Beam Computed Tomography (CBCT). Standard CBCT motion estimation techniques from the literature enforce uniform motion smoothing, which can be sub-optimal (e.g., sliding motion between organs). This approach proposes a motion estimation-compensation algorithm by penalized-likelihood function with a coupled dictionary learning as a regularization. The advantage of the methodology is that the image and the motion can inform each other, thus allowing for noise reduction and learning features such as sliding motion at organ boundaries. The dictionaries are learned from a set of images and their corresponding Deformation Vector Fields (DVF) at each respiratory gate. Results show the ability of the proposed coupled dictionary learning algorithm to learn from both dictionaries simultaneously and exploit data dependencies.

The second approach proposes a Multi-channel Convolutional Analysis Operator Learning (MCAOL) for Dual-Energy CT (DECT) Reconstruction. The method exploits standard spatial features within attenuation images at different energies and proposes an optimization method that jointly reconstructs the attenuation images at low and high energies with a mixed norm regularization on the sparse features. In particular, the regularization term promotes the joint sparsity between features obtained by pre-trained convolutional filters through the Convolutional Analysis Operator Learning (CAOL). Extensive experiments with simulated and real CT data were performed to validate the effectiveness of the proposed meth-

ods. Qualitative and quantitative results on sparse-views and low-dose DECT demonstrate that the proposed MCAOL method outperforms both CAOL applied on each energy independently and several existing state-of-the-art model-based iterative methods.

In the third technique, we focus on the sparse view single-source fast KVp switching acquisition set-up in Dual Energy CBCT to reduce the total dose delivered during a CT acquisition. We propose to exploit the Joint Total Variation regularization in the reconstruction problem, between low and high energy images, to reduce the artifacts due to the under-sampling of the angular views. Through numerical experiments and patient data, we show the benefit of the proposed method for material decomposition and estimation both qualitatively and quantitatively compared to regularization on the images separately.

Keywords: *X-ray Computed Tomography*, *Dictionary Learning*, *Image Reconstruction*, *Iterative Methods*, *Optimization*.

ACKNOWLEDGEMENT

With immense pleasure and deep sense of gratitude, I wish to express my sincere thanks to my supervisor **Dr. Alexandre Bousse**, without his motivation and continuous encouragement, this research would not have been successfully completed. His immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life.

I am grateful to the Operational Director of LaTIM, **Dr. Dimitris Visvikis**, for motivating me to carry out research and also for providing me with infrastructural facilities and many other resources needed for my research.

I express my sincere thanks to **Dr. Alessandro Perelli**, Assistant professor, University of Dundee, UK, for his advice on many technical aspects of the thesis. I like to acknowledge the support rendered by **my colleagues** at LaTIM in several ways throughout my research work.

I wish to extend my profound sense of gratitude to **my parents** for all the sacrifices they made during my research and also providing me with moral support and encouragement whenever required.

Place: Brest

Date: 14/02/2022 **Suxer Lazara ALFONSO GARCIA**

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF TERMS AND ABBREVIATIONS

1D One-dimentional

2D two-dimentional

3D three-dimentional

4D Four-dimentional

4D-CBCT Four-dimensional Cone Beam Computed Tomography

4D-CT Four-dimentional Computed Tomography

ADMM Alternating Direction Method of Multipliers

BPEG-M Block Proximal Extrapolated Gradient method using a Majorizer

CAOL Convolutional Analysis Operator Learning

CB Cone-Beam

CBCT Cone-Beam Computed Tomography

CDL Convolutional Dictionary Learning

CNN Convolutional Neural Network

CONVOLT CONVolutional Operator Learning Toolbox

CS Compressed Sensing

CT Computed Tomography

DCT Discrete Cosine Transform

DE-CBCT Dual-Energy Cone Beam Computed Tomography

DECT Dual-Energy Computed Tomography

DL Dictionary Learning

DVF Deformation Vector Fields

EBCT Electron Beam Computerized Tomography

EM expectation-maximization

EP edge-preserving

LIST OF TERMS AND ABBREVIATIONS

ESF Edge Spread Function

ET emission tomography

FBP Filtered Back Projection

FDK Feldkamp– Davis– Kress

FFT Fast Fourier Transform

FOV Field of View

FWHM Full Width at Half Maximum

GT Ground Truth

IGRT Image-Guided Radiation Therapy

IHT Iterative Hard Thresholding

IST Iterative Soft Thresholding

JRM Joint Reconstruction and Motion estimation

JTV Joint Total Variation

K-SVD K-means Singular Value Decomposition

KVp Peak Kilo-Voltage

L-BFGS Limited-memory Broyden-Fletcher-Goldfarb-Shanno

LAC Linear Attenuation Coefficient

LSF Line Spread Function

MAP Maximum a Posteriori

MBIR Model-based iterative reconstructions

MCAOL Multi-channel Convolutional Analysis Operator Learning

MEC-MDL Motion Estimation-Compensation with Multi-channel Dictionary Learning

MEC-SDL Motion Estimation-Compensation with Single Dictionary Learning

ML Maximum Likelihood

MLTR Maximum-likelihood reconstruction for transmission tomography

MOD Method of Optimal Directions

LIST OF TERMS AND ABBREVIATIONS

MSCT Multi-slice Computed Tomography

MTF Modulation Transfer Function

NLL Negative Log-Likelihood

OMP Orthogonal Machine Pursuit

PMOC Proximal Mapping with Orthogonality Constraint

PRISM prior rank, intensity and sparsity model

PSNR Peak Signal-to-Noise Ratio

PWLS penalized weighted least-squares

RIC Restricted Isometry Constant

RIP Restricted Isometry Property

RMSE Root Mean Square Error

ROI Regions of Interest

SMEIR Simultaneous Motion Estimation and Image Reconstruction

SNR Signal-to-Noise Ratio

SOUP-DIL Sum of Outer Products Dictionary Learning

SPS separable paraboloidal surrogate

SSIM Structural Similarity Index Measure

STD Standard Deviation

TV Total Variation

WT Wavelet Transform

XCAT Extended Cardiac-Torso

CHAPTER 1

General Introduction

1.1 Introduction

Computed Tomography CT has become an invaluable imaging tool in clinical practice. It was the first non-invasive means of obtaining images of the human body's interior that were not distorted by the superposition of different anatomical features, as in planar X-ray fluoroscopy. As a result, CT produces images with better contrast compared to traditional radiography. This was a giant stride forward in advancing diagnostic capabilities in medicine throughout the 1970s (Buzug 2008).

CT has proven an effective imaging technique for detecting potential cancers or lesions in the abdomen. A CT scan of the heart may be requested when various cardiac illnesses or anomalies are detected. CT scans of the head can detect injuries, tumors, blood clots that cause strokes, bleeding, and other diseases. It can examine the lungs to see malignancies, pulmonary embolisms (blood clots), excess fluid, and other illnesses, including emphysema or pneumonia (NIBIB 2021).

CT scanners have gone through seven generations of development and research. From the first generation to the seventh generation, CT has continually improved in speed, spatial resolution, and density resolution. Currently, these three aspects of CT are still goals of manufacturers, but the fourth aspect, low-dose scanning, is what manufacturers are focused on and is their main direction for CT development. In general, X-ray CT has been trending towards low-dose CT, ultra-low-dose CT, and spectral CT, which have an accurate positioning and qualitative diagnosis using the least amount of radiation possible (Liu 2018).

Common strategies to lower X-ray radiation dose are: lowering the X-ray exposure in each view by adjusting the tube current; decreasing the number of projection angles (sparse-view)-CT. However, reducing the number of projection angles leads to inaccuracy in the resultant image. More sophisticated methods are needed to process the raw data from CT systems to reduce radiation while still producing good quality images. These methods are known as image reconstruction and are one of the main topics of research in the CT field. Researchers constantly develop

new, faster, and more accurate image reconstruction algorithms.

1.2 Aim of the Thesis

The present thesis aims to develop sophisticated X-ray CT image reconstruction algorithms to improve image quality while keeping the radiation dose as low as possible. The objective is to deploy new model-based iterative reconstruction algorithms based on the Compressed Sensing (CS) theory and Machine Learning. Proof-of-concept methods are developed with an emphasis on the joint reconstruction of multi-channel modalities to exploit structural similarities in the images.

1.3 Structure of the Thesis

The present manuscript is composed of five main chapters in addition to a general introduction presented in Chapter 1, the Conclusions presented in Chapter 7 and the appendices.
Chapter 2 explains the physics and mathematical principles of the X-ray CT. It provides detailed information about the X-ray production (Bremsstrahlung radiation) and the main interaction process of the X-ray with the matter at the diagnostic energies (Photoelectric effect, Compton scattering, and Coherent scattering). Furthermore, it goes through the advancement of the CT generations by describing the seven generations and their main characteristics. The second part of the chapter explains the mathematics behind image reconstruction in CT starting with the analytical methods and continuing with the Model-based iterative reconstructions (MBIR). It explains in detail the main optimization algorithms for MBIR used in the thesis (e.g., Newton approaches with Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm).
Chapter 3 describes the CS theory and provides a literature review of the main sparse recovery algorithms used in the thesis (Orthogonal Machine Pursuit (OMP), Iterative Soft Thresholding (IST), Iterative Hard Thresholding (IHT)). It describes the Total Variation (TV) semi norm and its implication in the CS theory. The Dictionary Learning (DL) problem and the main optimization algorithms for patch-based dictionary learning are explained in detail in this chapter. The Convolutional Dictionary Learning (CDL) and, more specifically, the Convolutional Analysis Operator Learning (CAOL) are explained. The Block Proximal Extrapolated Gradient method using a Majorizer (BPEG-M) algorithm for the optimization of the CAOL algorithm is detailed as well as its application to the CAOL problem.
Chapter 4 depicts the first contribution of this thesis. We deploy a new approach for motion-estimation compensation in Cone-Beam Computed Tomography (CBCT)

by learning joint image-motion dictionaries in order to correct sliding motion at organs boundaries. First, the state-of-the-art of the motion estimation-compensation with sliding correction in CBCT is presented. Then, the model is explained in detail and the methods used for comparison. Details on the experiments performed are explained, and the more relevant results are discussed. An extensive discussion section details the follow-up projects of the proposed approach.
Chapter 5, which is the major contribution of the thesis, proposes the Multi-channel Convolutional Analysis Operator Learning (MCAOL) method for Dual-Energy Computed Tomography (DECT). It proposes an optimization algorithm that jointly reconstructs the attenuation images at low and high energies with a mixed semi-norm regularization on the sparse features. First, it details the state-of-the-art of joint reconstruction within the CS theory. Then, the methodology is explained in detail and the algorithm used for comparison. The experiments performed with low-dose CT and sparse-view CT for the Extended Cardiac-Torso (XCAT) phantom and the clinical data are detailed in the chapter guaranteeing their reproducibility.
Chapter 6 proposes a methodology for sparse-view image reconstruction in single-source rapid Peak Kilo-Voltage (KVp) switching in Dual-Energy Cone Beam Computed Tomography (DE-CBCT). The Joint Total Variation (JTV) regularization is implemented and used within a MBIR to encode the low and high energy images. The performance of the reconstructed images for material decomposition is evaluated and compared with the single reconstruction utilizing TV and the Huber prior.

CHAPTER 2

Computed Tomography and Image Reconstruction

2.1 X-ray Tomographic Imaging

The earliest diagnostic imaging technology with X-rays was created immediately after Roentgen discovered X-rays in 1895. X-rays are electromagnetic radiation that propagates through matter and interacts with it through various physical processes. Planar radiography and CT utilize differential absorption of X-rays while traveling through human tissue. For example, bones absorb X-rays more efficiently than soft tissue. Therefore, the interaction of X-rays with matter can be used as a non-invasive alternative to imaging an object. In radiography, an X-ray beam irradiates an object providing a two-dimentional (2D) image, which is the "shadow" of the 3D object. The projection becomes a superposition of internal structures, making it difficult for the radiologist to identify them. Moreover, it is quite challenging to differentiate low-contrast structures in tissue.
CT was developed to overcome these limitations and to be able to acquire a fully three-dimensional image. The CT machine consists of an X-ray source and a radiation detector with multiple rows placed in the opposite direction to the source. The source and the detector rows are rotated in synchronization around the patient. A set of 2D projections are acquired and further reconstructed to form the 3D images (Smith and Webb 2010). Figure 2.1 shows the basic principle of CT scanner and a picture of a modern multi-detector helical scanner.

2.1.1 X-ray Generation

The X-ray source consist of an X-ray tube. The X-rays photons are produced when accelerated electrons hit a target with a high number of protons.
The tube is composed of an electron source, the cathode, commonly a heated filament, and an anode, usually made of tungsten and contained in an evacuated glass envelope. First, a high voltage is applied between the cathode and the anode. This voltage accelerates the electrons in a range from 30 to 140 kilo-volts. This accelerating voltage is also known as the Peak-Kilo-voltage (kVp). When the

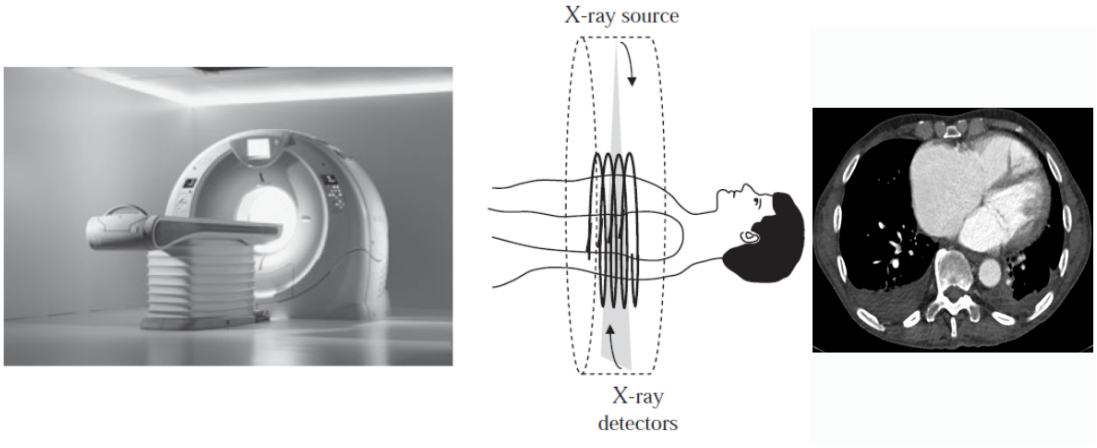


Fig. 2.1 A modern helical CT scanner (left). The basis principle of CT. The X-ray source and detector set-up(center). A 3D reconstructed volume of the heart utilizing CT scanning(right). Reprint from Smith and Webb (2010), O'Donnell (2022)

high-energy electrons collide with the target (tungsten anode), they pass close to the nucleus in the atoms. They are influenced by its electric field (Allisy-Roberts and Williams 2007). They are decelerated, deflecting their trajectories, decreasing the electron's kinetic energy. The energy "lost" by the electron in this process is emitted as X-rays photons or *bremsstrahlung* radiation (bremsstrahlung is German for "braking radiation"). The incident electron also loses energy throughout the tungsten target by ionization, interacting with other electrons in the matter. Thus, the mean energy lost by the electron in a material of thickness dx can quantitatively be described by

$$\left(\frac{\mathrm{d}E}{\mathrm{d}x}\right) = \left(\frac{\mathrm{d}E}{\mathrm{d}x}\right)_{\text{ionization}} + \left(\frac{\mathrm{d}E}{\mathrm{d}x}\right)_{\text{bremsstrahlung}} \tag{2.1}$$

where $\left(\frac{\mathrm{d}E}{\mathrm{d}x}\right)_{\text{ionization}}$ is given by the Bethe-Bloch equation

$$\left(\frac{\mathrm{d}E}{\mathrm{d}x}\right)_{\text{ionization}} = -4\pi N_A \rho \frac{Z}{A_r}\left(\frac{e^4}{m_\mathrm{e}c^2}\right)\frac{z^2}{\beta^2}\cdot \left[\frac{1}{2}\ln\left(\frac{2m_\mathrm{e}c^2\beta^2\gamma^2 T_{\max}}{I_m}\right) - \beta^2 - \frac{\delta}{2}\right] \tag{2.2}$$

with N_A denoting the Avogadro constant, ρ is the material density, Z is the atomic number, A_r is the atomic weight of the material; e and m_e the electron charge and rest mass, respectively. The electron velocity is expressed in units of light speed, i.e. $\beta = v/c$; γ represents the Lorentz factor $\gamma = (1-\beta^2)^{-1/2}$ and $T_{\max}$ is equal to the tube voltage times the electron charge and represents the maximum kinetic energy that may be transmitted in a single collision; δ is a density correction of

the ionization energy, and I_m is the mean ionization energy of the material (Buzug 2008).
The second term in 2.1 is the Bremsstrahlung photons energy and is given by quantum electrodynamics (QED) (Buzug 2008)

$$\left(\frac{\mathrm{d}E}{\mathrm{d}x}\right)_{\text{bremsstrahlung}} = -4\alpha N_{\mathrm{A}}\rho\frac{Z^2}{A}\left(\frac{e^2}{m_{\mathrm{e}}c^2}\right)^2 E\ln\left(\frac{183}{Z^{1/3}}\right), \tag{2.3}$$

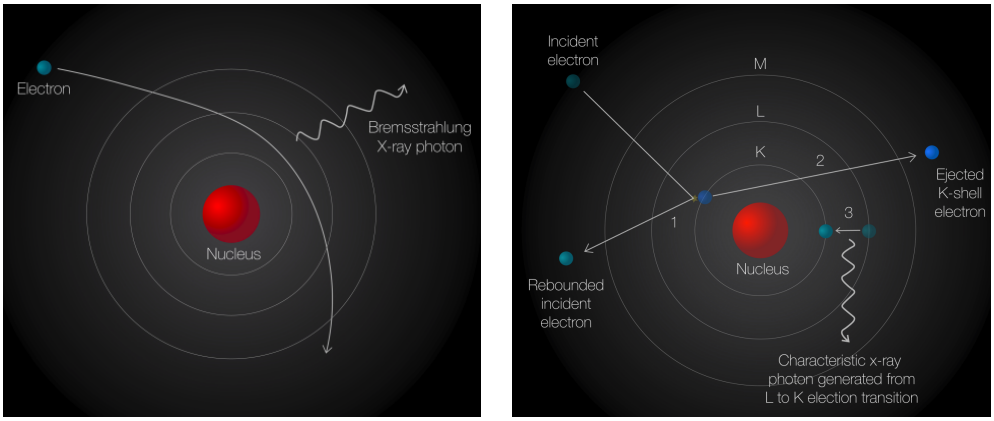
where α denotes the fine-structure constant.
The bremsstrahlung radiation has a continuous spectrum with an energy range from zero to the maximum kinetic energy of the bombarding electron depending on how much the nucleus electric field impacts the electrons. Figure 2.2 illustrates the bremsstrahlung emission mechanisms in the atom and the X-ray energy spectrum from tungsten anode operating at 120 KVp accelerating voltage. The spectrum contains vertical lines corresponding to *characteristics* X-ray. These *characteristics* X-rays are created when a bombarding electron collides with a K-shell electron in the tungsten anode. If the incident electron's energy is bigger than the binding energy of the K-shell electron, the electron in K-shell is ejected, leaving a hole in the shell. An electron coming from more external shells (L-shell, M-shell) fills the hole. During the desexcitation process, a characteristic photon is emitted with an energy level equal to the binding energy difference between the outer and inner shell electron involved in the transition. Figure 2.2a show the characteristic X-rays production mechanism (Hapugoda 2020*a*).
The efficiency of converting kinetic electron energy to bremsstrahlung energy is given by (Buzug 2008)

$$\eta = KZU_a, \tag{2.4}$$

where is $K = 9.2 \cdot 10^{-7} kV^{-1}$ the Kramers constant (Kramers 1923), U_a is the accelerating voltage in the X-ray tube, and Z is the atomic number of the anode material. Following equation 2.4, the quantum efficiency of the conversion from kinetic energy into X-ray radiation, within a tungsten anode ($Z = 74$), and operating with an acceleration voltage of $Ua = 140kV$ is $\eta = 0.01$. This efficiency implies that only 1% of the kinetic energy is converted to bremsstrahlung radiation. The other 99% is transmitted locally to the lattice, causing the anode to heat up. As a consequence, CT X-rays tubes may suffer from overheating (Buzug 2008).

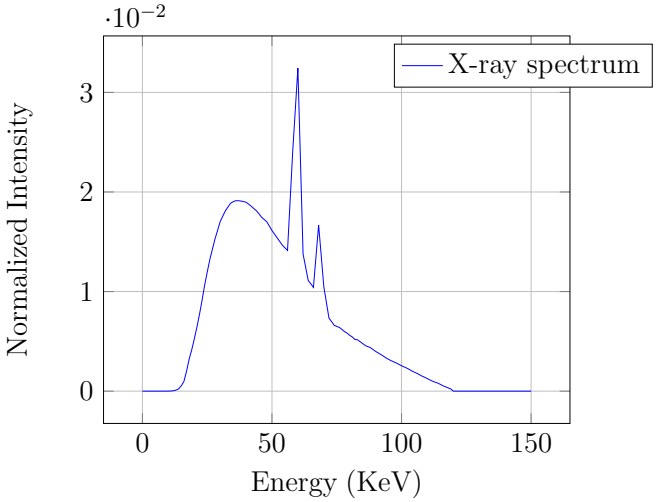
2.1.2 Interaction of X-ray with the matter

The X-rays produced in the tube irradiate the patient or the studied anatomical region. They interact with the tissue via three main processes, which depend on the photons energy, atomic number of the material, and the density of the material



Bremsstrahlung X-rays Characteristic X-rays

(a) X-rays production mechanisms



(b) Energy spectrum of Bremsstrahlung X-rays

Fig. 2.2 a) Bremsstrahlung (Reprint from Hapugoda (2020*a*)) and characteristics (Reprint from (Hapugoda 2020*b*)) X-rays production mechanisms in the atom. b) X-ray energy spectrum from tungsten anode operating at 120 KVp. (Punnoose et al. 2016)

in the body. At the diagnostic energies, the primary interaction processes are photoelectric effect, incoherent (Compton) scattering, and coherent (Rayleigh) scattering. The interaction of the photons with matter is a stochastic process. The probability of the interaction depends on the atomic cross-section. We denote σ_{FE} the atomic cross-section for the photoelectric effect, σ_{R} for coherent scattering and, σ_{KN} for incoherent scattering.

2.1.2.1 Photoelectric effect

The photo-effect or photoelectric effect is the process where an incident photon interacts with a binding electron in the atom. Albert Einstein introduced the photoelectric effect theory in 1905, based on Max Planck's idea that light consists

of small packets of energy known as photons or light quanta with energy $h\nu$ proportional to the frequency ν of the corresponding electromagnetic wave and the Planck's constant h. In CT the incident photons come from the X-ray beam generated in the X-ray tube. The incident photon is absorbed, leaving the atom in an excited state. One of the electrons attached to the nucleus is ejected, releasing the extra energy in the collision. The ejected electron is called a *photo-electron* and leaves the atom with a kinetic energy

$$T = h\nu - E_{\mathrm{s}} \tag{2.5}$$

where E_{s} is the binding energy of the electron shell where the electron was located, and ν is the incident photon frequency (Dance et al. 2014). Thus, the photoelectric effect occurs only when the incident photon energy is greater than the binding energy. The electron shell that satisfies these criteria and is closest to the nucleus (with the highest binding energy) is the most likely to lose an electron. The photoelectric effect cross-section is obtained through Quantum Mechanics, and it is proportional to fourth power atomic number (Z) and inversely proportional to photon energy ($h\nu$). In the diagnostic photon energy range, a typical dependency of σ_{FE} is

$$\sigma_{\mathrm{FE}} \sim \frac{Z^4}{(h\nu)^3} \tag{2.6}$$

The photoelectric effect is the most likely process for low energy photons and high Z materials. It plays an essential role in CT and is the reason why bone tissue is easily visible in CT images (Dance et al. 2014).

2.1.2.2 Coherent (Rayleigh) scattering

The Rayleigh scattering mechanisms consists of scattering of photons by non-free electrons. In the Rayleigh scattering the photon is scattered slightly resulting in a small change in energy. The differential cross-section can be written as

$$\frac{\mathrm{d}\sigma_R}{\mathrm{d}\Omega} = \frac{r_0^2}{2}\left(1 + \cos^2\theta\right)\left[F(q, Z)\right]^2 \tag{2.7}$$

where θ is the photon scattering angle, r_0 is the classical electron radius, and $F(q, Z)$ is a coherent factor calculated utilizing Quantum Mechanical Models with $q = \frac{\sin(\theta/2)}{\lambda}$. Denoting λ the wavelength of the incident photon and $\mathrm{d}\Omega$ is the solid angle (Figure 2.3). The total atomic cross-section in the Rayleigh scattering is second power inversely proportional to the energy of the photon and directly

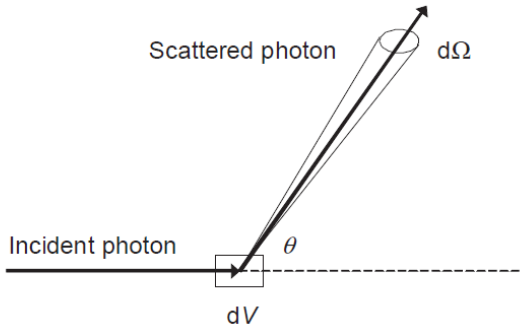


Fig. 2.3 The solid and scatter angles. A photon incident on a tiny volume element dV is scattered into the solid angle element $d\Omega$ through angle σ. (Reprint from (Dance et al. 2014))

proportional to the atomic number

$$\sigma_R \propto \frac{Z^2}{(h\nu)^2} \tag{2.8}$$

Since the incident photon loses no energy during Rayleigh scattering, the process does not deliver a radiation dose to matter. Rayleigh scattering is more likely to occur in photon beams with lower energy (Dance et al. 2014).

2.1.2.3 Incoherent (Compton) scattering

The Compton scattering, as Rayleigh scattering, is the interaction between the incident photons and the electrons in the matter, where the electron receives an energy transfer during the process. Figure 2.4 depicts the interaction geometry. An incident photon with energy $h\nu$ collides (Billiard-ball-like collision) with the electron and is scattered through an angle θ. The photon energy after the collision becomes $h\nu'$. The electron recoils with kinetic energy T_e at angle ϕ

$$T_\mathrm{e} = h\nu - h\nu' \tag{2.9}$$

The differential cross-section can be calculated by assuming the electron is "free" (unbound). Klein and Nishina first derived it in 1928 utilizing the Dirac theory of the electron (Klein and Nishina 1929). The expression estimates the differential cross-section for the scattering of photons by a free electron.

$$\frac{\mathrm{d}\sigma_\mathrm{KN}}{\mathrm{d}\Omega} = \frac{r_0^2}{2}\left(1+\cos^2\theta\right) f_\mathrm{KN} \tag{2.10}$$

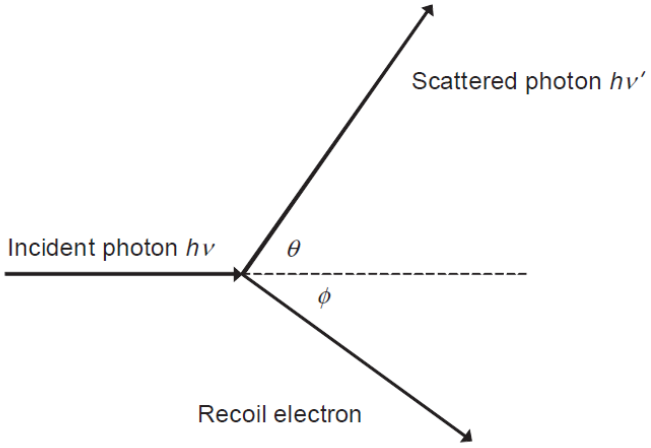


Fig. 2.4 Compton scattering geometry. Reprint from (Dance et al. 2014)

where

$$f_{\mathrm{KN}} = \left\{ \frac{1}{1+\alpha(1-\cos\theta)} \right\}^2 \left\{ 1 + \frac{\alpha^2(1-\cos\theta)^2}{[1+\alpha(1-\cos\theta)]\,[1+\cos^2\theta]} \right\} \tag{2.11}$$

where $\alpha = h\nu/m_0c^2$, with c the speed of the light in vacuum and m_0 denoting the electron rest mass.
Integrating over all the scattered angles, the total cross-section becomes

$$\begin{aligned} \sigma_{\mathrm{KN}}(h\nu) = 2\pi r_0^2 \left\{ \left(\tfrac{1+\alpha}{\alpha^2}\right) \left(\tfrac{2(1+\alpha)}{1+2\alpha} - \tfrac{\ln(1+2\alpha)}{\alpha} \right) \right\} + \\ 2\pi r_0^2 \left\{ \tfrac{\ln(1+2\alpha)}{2\alpha} - \tfrac{1+3\alpha}{(1+2\alpha)^2} \right\} \end{aligned} \tag{2.12}$$

In 2.1.2.3 is assumed that the electron is free. We can see from this equation that the attenuation coefficient per electron is independent of the atomic number and is solely reliant on the photon energy.

2.1.2.4 Linear attenuation coefficient

The total cross-sections mentioned above concern the interaction of photons with an individual atom. It is necessary to consider the macroscopic properties of a photon beam when traversing the matter. Consider a photon beam incident generally on a thin uniform slab of material with thickness dl. The probability that a photon interacts in this thin slab is given by

$$N_{\mathrm{a}}\sigma dl \tag{2.13}$$

where N_{a} is the total number of atoms in a substance per unit volume, and σ is the total atomic cross-section, which can be calculated utilizing the "or rule" for probabilities

$$\sigma = \sigma_{\mathrm{FE}} + \sigma_{\mathrm{R}} + \sigma_{\mathrm{KN}} \tag{2.14}$$

The quantity $N_{\mathrm{a}}\sigma$ is the linear attenuation coefficient, and it is denoted by μ in the manuscript. The estimation of the number of atoms N_{a} can be performed utilizing the Avogadro constant N_{A}, the material density ρ, and the atomic weight A_r

$$\mu = N_{\mathrm{a}}\sigma = \frac{N_{\mathrm{A}}\rho}{A_{\mathrm{r}}}\sigma \tag{2.15}$$

The dimensions of μ in the International System of Units is m^{-1}, although it is common to use cm^{-1} (Dance et al. 2014).

Exponential attenuation

Let us consider a thick slab of material of thickness l and $I_f(l)$ the fluence of photons that have passed the slab and have not interacted. The variation in the fluence, $\mathrm{d}I_f$, after passing the thickness $\mathrm{d}l$ is given by

$$\begin{aligned} \mathrm{d}I_f &= -I_f \mu \mathrm{d}l \\ \mathrm{d}I_f/I_f &= -\mu \mathrm{d}l \end{aligned} \tag{2.16}$$

The negative sign implies that the fluence I_f decreases with l and μ. Integrating each side of the equation 2.16

$$\begin{aligned} &\textstyle\int_{I_{f0}}^{I_f} \mathrm{d}I_f/I_f = \int_0^L \mu \mathrm{d}l \\ &I_f = I_{f0} e^{-\mu L} \end{aligned} \tag{2.17}$$

where L denotes the slab's thickness and I_{f0} is the initial fluence. The resulting relation in 2.17 is known as Beer's law and describes the exponential attenuation of a photon beam. More specifically, I_f represents the number of photons that pass through the slab without interaction. In the CT energy ranges, other photons may be present in the detector after passing the slab (Dance et al. 2014). To account for these photons, we add a background term s to equation 2.17

$$I_f = I_{f0} e^{-\mu L} + s \tag{2.18}$$

The expression 2.18 holds for mono-energetic X-rays and assumes the slab of thickness L is composed of a unique material. Let us denote $\mu(l)$ the variation of the linear attenuation coefficient through a medium with different materials. Thus, after crossing a multi-material slab of length L, the fluence is given by (Buzug

2008)

$$I_f = I_{f0}\mathrm{e}^{-\int_0^L \mu(l)\mathrm{d}l} + s \tag{2.19}$$

Taking into account the energy dependency of the attenuation values, equation 2.19 must be extended to

$$I_f = \int_0^{E_{\max}} I_{f0}(E)\varepsilon(\mathrm{E})\mathrm{e}^{-\int_0^L \mu(E,l)\mathrm{d}l}\mathrm{d}E \tag{2.20}$$

where ε(E) denotes the detector efficiency. The relation 2.19 is the most common used for image reconstruction. Therefore, in this thesis, it will be used to model the projection dataset.

2.1.3 Radiation Detection

In the previous sections, we have described how the incident photons coming from the X-ray tube interact with the body. The photons that cross the body are collected in a device known as detectors. Specific materials in the detector are used to convert the X-ray energy of the photons into lower-energy forms. For instance, optical photons in the case of scintillator detectors or electron-hole pairs in the case of semiconductor detectors. In the detection process, thousands of secondary quanta per primary incident photon are generated, which have energies of a few electron volts. The low energy quanta generated produce an electrical current which is further conditioned utilizing an electronic amplifier. Then, the signal passes through an analog-to-digital converter which converts it into a digital number. These digital numbers are the raw projection data which is further reconstructed utilizing an appropriated reconstruction algorithm (Drzezo 2016).

2.1.4 CT Configuration and Generations

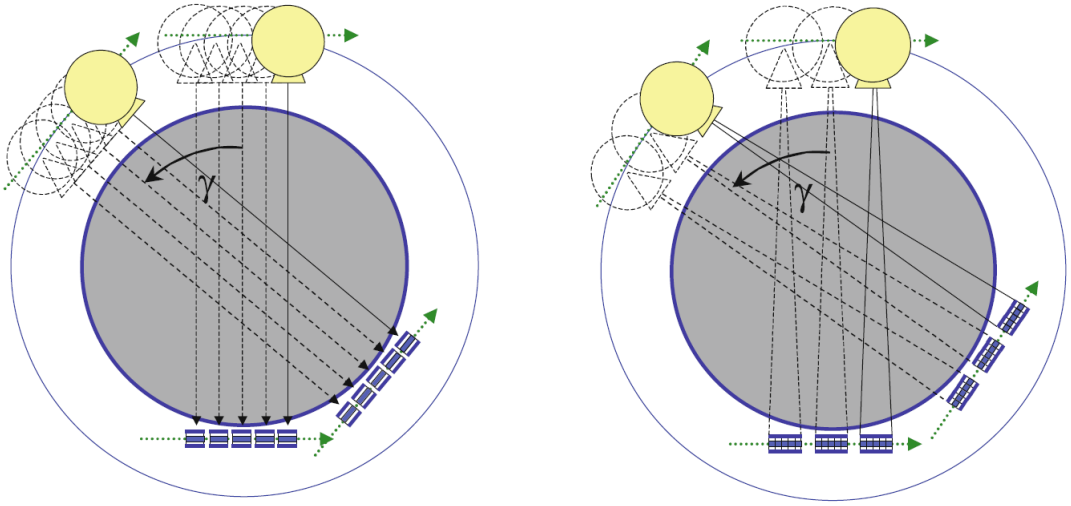
Several CT configurations have been implemented based on the physics principles above explained. The CT configurations have gone through multiple enhancements focusing on an increase in the number of detectors and a reduction in scan time. The first generation design consists of a single X-ray source emitting a single needle-like X-ray beam and rigidly coupled single detector cell. The pencil beam is translated across the patient to obtain a set of parallel projections at one angle. Then the system rotates γ degrees, and another set of parallel projections is collected by translating the system across the patient. The process is repeated until they acquired 180 projections with a Field of View (FOV) of 24 *cm* approximately. This type of scanner is known as parallel beam translate-rotate scanners (Buzug 2008). Figure 2.5a (left) illustrates the configuration of the first tomograph generation.

The second tomographs generation features an X-ray source with a narrow fan beam and a short detector array of about 30 elements. Because the fan beam aperture is small, the X-ray tube and the detector array needs to be linearly translated and rotated as in the first generation. The fan angle on the earliest second-generation CT scanners was 10 degrees. This type of scanner is known as Narrow Fan Beam Rotation–Translation scanners. Figure 2.5a (right) illustrates the configuration of the second tomograph generation. The first and second-generation are quite slow in acquisition time per slice. Thus these scanners were mainly restricted to use in imaging the cranium.
The third generation focuses on decreasing the acquisition time to less than 20 seconds, which allows to acquire an image of the abdomen while the patient holds their breath. The main improvement in the third generation is the extension of fan beam angle to a range between 40 to 60 degrees and the detector array to an arc of 400 to 1000 elements (Figure 2.5b left). For each projection angle γ, the system can simultaneously irradiate the full measuring field, which is wide enough to encompass the torso. Thus, the third-generation scanner eliminates the linear translation of the X-ray source and the detector (Buzug 2008).
The rotation-fix with closed detector ring CT is the fourth generation of scanners. The X-ray source remains the same as in the third generation, a fan-beam source rotating continuously around the measuring field. However, the detector is fixed, making a ring with around 5000 elements. The X-ray tube rotates inside the detector ring. Figure 2.5b (right) illustrates the configuration of the fourth tomograph generation.

Other modifications of tomographs have been developed to improve the older generations. For example, Rotation in Spiral Path Scanner, the Electron Beam Computerized Tomography, and Rotation in Cone-Beam Geometry. Many authors identify them as the fifth, sixth, and seventh generations. However, there is no precise classification (Buzug 2008)).

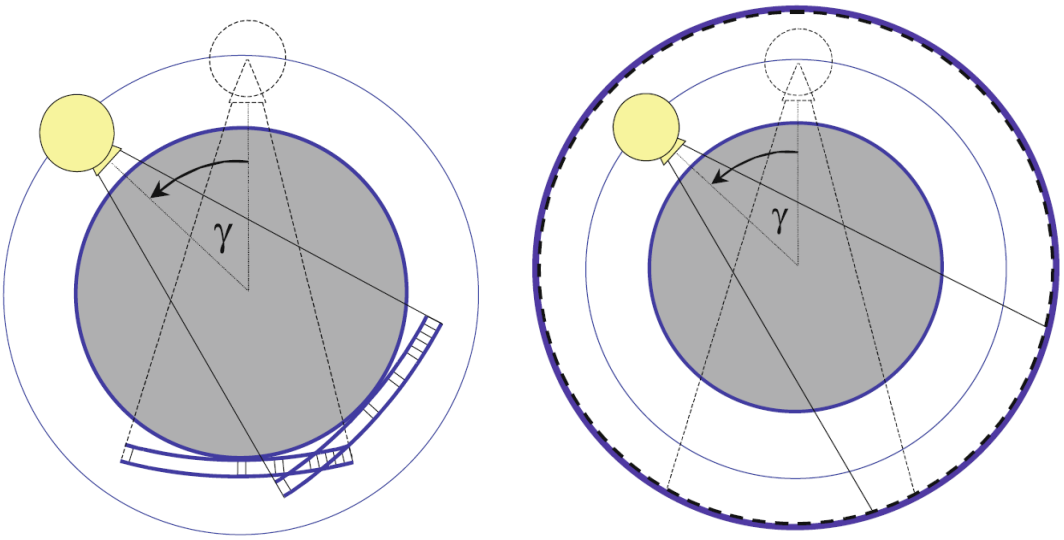
Electron Beam Computerized Tomography

One approach to decrease the acquisition time is to use the Electron Beam Computerized Tomography (EBCT) system. It was introduced for cardiac imaging and was capable of acquiring an image slice in $50ms$. In EBCT an electron beam is focused onto tungsten target rings which are positioned in a half-circle around the patient and generate a fan beam. A stationary detector ring is used to measure the X-ray irradiation (Buzug 2008). The main application of this type of tomographs is in cardiology to search calcium build-up in the heart arteries. The EBCT is also referred as the "cine CT" system, and some authors have categorized it as the fifth generation.



First CT Generation Second CT Generation

(a) Rotate-Translate CT scanners



Third CT Generation Fourth CT Generation

(b) Rotation of a fan-beam CT scanners

Fig. 2.5 a) The first and second generation of X-ray CT scanners utilize the rotate-translate principle. The source and the detector are moved linearly and rotated at an angle γ. b) Third and fourth generation of CT scanners which irradiate with a wide fan beam, and the X-ray source rotates continuously without any linear displacement. In the third generation, the detector has an arc shape with around 1000 elements, while in the fourth generation, the detector has a ring shape and is fixed. Reprint from Buzug (2008)

Rotation in Spiral Path

In the previous CT generations, after each 360° rotation, the gantry has to stop and reverse direction. Mainly because of the cables connecting the rotating components to the rest of the gantry. They are spooled onto a drum, then released and re-

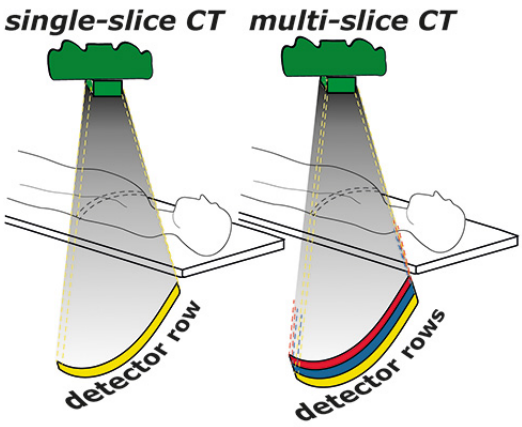


Fig. 2.6 Single-slice CT (left) versus multi-slice CT. Reprint from Annelies van der Plas (2016)

spooled during rotation and reversal. The scanning, braking, and reverse process needs at least 8-10 seconds, with just 1-2 seconds spent on data acquisition. As a result, the scan required considerable acquisition time, and the temporal resolution is poor (Goel 2015). The invention of slip-ring technology eliminates this problem and led to what Bushberg et al. (2003) identified as the sixth generation scanners. In this technology, the electrical power is provided via sliding contacts outside the gantry, allowing the X-ray tube and the detector (in the third generation) to rotate continuously. Since the gantry can now rotate non-stop, it become possible to acquire data in the shape of a spiral by translating the patient table through the gantry. This powerful idea, also known as helical CT or spiral CT, enables quick scans of entire z-axis regions of interest, in some circumstances within a single breath hold (Buzug 2008). However, as seen in section 2.1.1 the X-ray tubes suffer from overheating.

The solution is to employ the X-ray beam more efficiently. For instance, the X-ray beam has a cone shape by nature. The pencil and fan-beam are created utilizing appropriate pin-hole or slit collimators. Thus, a distinctive approach would be to widen the beam in the z-direction (slice thickness) and adapt multiple detectors rows to collect the data for more than one slice at a time. This idea is the principle of Multi-slice Computed Tomography (MSCT), which was an extension of the third generation of tomographs (tube and detector bank linked and rotating together). The detectors in MSCT are further separated along the z-axis, allowing for the acquisition of many sections per rotation at the same time. As a result, with smaller section widths, MSCT delivers more and quicker z-axis coverage each rotation (Goldman 2008). Figure 2.6 illustrates the difference between single-slice CT (left) and multi-slice CT which utilizes multiple detector rows. After the introduction of MSCT in the 1990s, many detector array configurations were

exploited depending on the number of sections acquired at each rotation. For example, for 4 data channels, the system acquires 4 slice at a time. From this point forward, manufacturers started developing 16-channel (16-slices), 64-channel (64-slices) scanners with different detector configurations. The total number of detector rows and z-axis coverage varies amongst CT manufacturers (Goldman 2008).

2.1.5 Cone-Beam Computed Tomography

Following the previous idea of exploiting more efficiently the X-ray beam the next step in the development of CT scanners was the use of a cone-shaped X-ray beam, which is already created in the X-ray tube. A flat-panel detector, which did not exist at the time, had to be created to replace the line or multi-line detector arrays to employ the cone beam. This type of scanners are referred as the seventh generation of CT scanners and are denominated as CBCT (Bushberg et al. 2003). The X-ray source and the bank flat-panel detector synchronously rotate around the patients to acquire between 150 and 600 sequential planar projections in a single sweep in 180°–360° of gantry rotation. The main application of CBCT is in dentistry and maxillofacial scan. It produces images of contrasted structures, which makes it well-suited to imaging skeletal structures in the craniofacial region. Another major application of CBCT is for Image-Guided Radiation Therapy (IGRT). In external beam radiotherapy treatments, the machines come with a CBCT device attached to the gantry. The CBCT machine is used to ensure optimum patient setup and as image guidance tools in IGRT, by providing a volumetric image of a patient in the treatment position. With proper calibration, the CBCT image can be used for dose calculation during the radiotherapy and re-planning the treatment in case of anatomical changes in the patient. Nevertheless, the image quality is inferior compared to diagnostic CT. The cone beam irradiates more volume in the patient. Consequently, a large amount of scattering signal reaches the detector. The large scatter-to-primary ratio substantially degrades the reconstructed image. Moreover, depending on the frequency of the acquired CBCT (given that radiotherapy treatments typically involve 30-50 fractions), the dose to the patient may become significant. Decreasing the dose, therefore, increases the noise due to low photon counts, which creates artifacts in the image resulting in random thin bright and dark streaks that appear preferentially along the direction of most significant attenuation (Boas et al. 2012).
There is an increasing interest in working with low-dose CBCT acquisitions without compromising the overall resulting image quality. Additionally, the gantry rotation in a CBCT acquisition for radiotherapy takes around 1 minute for a 360 degrees

scan, and the respiratory cycle is up to 6 seconds. The patient breaths ten times during the acquisition, introducing respiratory motion artifacts in the image (Yoon et al. 2019). In Chapter 4 we propose a novel algorithm for motion-estimation and motion-compensation in CBCT to improve the image quality of a CBCT mounted on the gantry of a linear accelerator used in radiation therapy.

2.1.6 Dual Energy Computed Tomography

Advances in CT continued moving in the direction of improving the visualization of the images and obtaining better contrast and image quality. One approach toward enhancing tissue visualization in CT-CBCT is the dual-energy acquisition. The fundamental concept behind imaging with two energy spectra is that understanding how a material behaves at two different energies can reveal information about tissue composition. As seen in Section 2.1.2, the photoelectric effect depends on the incident photons energy, and its probability or cross-section increases as the incident photon energy approximate the K-shell binding energy of an electron in the matter. The K-shell binding energy is different for each element, increasing with atomic number (Z). The term "K-edge" refers to the increase in attenuation at energy levels just above the K-shell binding due to increased photoelectric absorption. This variability of the K-edges for each material and the energy dependence of the photoelectric effect are the basis of dual-energy imaging techniques.
Let us consider a simple example to illustrate the ideas underpinning dual-energy approaches. Assume hypothetical elements A and B, with K edges of 90 keV and 190 keV, respectively. Now assume four unknown substances, each containing unknown quantities of A and B. We irradiate the unknown substances at two different voltages, 100 kVp and 200 kVp, to determine the amount of element A or B in each unknown substance. The results are shown in Figure 2.7. Substance 1 does not attenuate at either 100KVp or 200KVp. Therefore it contains neither A nor B. Substance 2 attenuates more at 200KVp than at 100KVp; consequently, mainly contains B because 200 kVp is just above 190 keV, the K edge of element B. Substance 3 attenuates more at 100KVp than at 200KVp; therefore it mainly contains A, because 100 kVp is close to 90 keV, the K-edge of element A. Substance 4 attenuates similarly to 100KVp and 200KVp; thus, it contains a similar amount of A and B. (Coursey et al. 2010).

In DECT, it is desirable to have the least possible overlap between spectra, therefore the lowest and highest potentials offered by the scanner should be used. A voltage below 60kV would not be useful because most of the radiation would be absorbed by the human body. Due to heating limitations, X-ray tubes are not capable of using voltages above 150 KV. Furthermore, the material to be

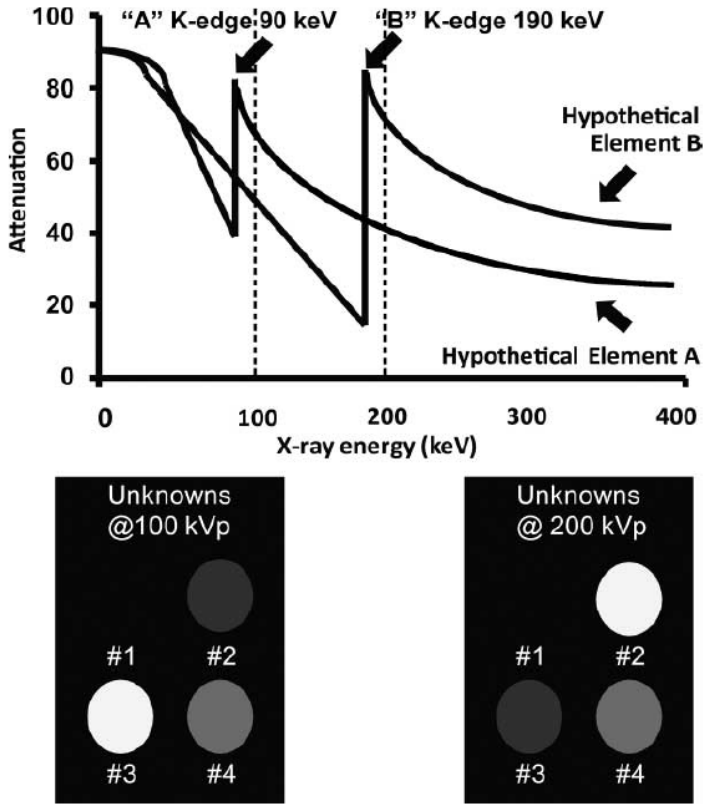


Fig. 2.7 Attenuation of elements A and B as a function of energy level (top). Behavior of substances 1, 2, 3 and 4 at 100kV and 200kV (botton). Reprint from Coursey et al. (2010).

studied must have a sufficient difference in spectral properties. Only elements with considerably different atomic numbers can be distinguished by their spectral properties (Johnson et al. 2011).

2.1.6.1 DECT acquisition methods

There are multiple CT scanner configuration to acquire dual energy projection data: Sequential Acquisition, Rapid Voltage Switching, Dual-Source CT, Dual Layer Detector and Multi-spectral CT with energy discriminating detectors.

The sequential acquisition can be achieved as two subsequent helical or CBCT scans one scan at high kilo-voltage and a second scan at low kilo-voltage. Alternatively, it can be acquired by subsequent rotations at alternating tube voltages and step-wise table feed. This strategy may make sense in systems with wide detectors, but the relatively significant latency between both acquisitions is a drawback. The delay is too lengthy to avoid artifacts caused by cardiac or respiratory movements and variations in contrast material specifications. However, for clinical DECT applications that do not need contrast material, such as metal artifact removal or kidney stone distinction, the sequential acquisition should be a feasible choice

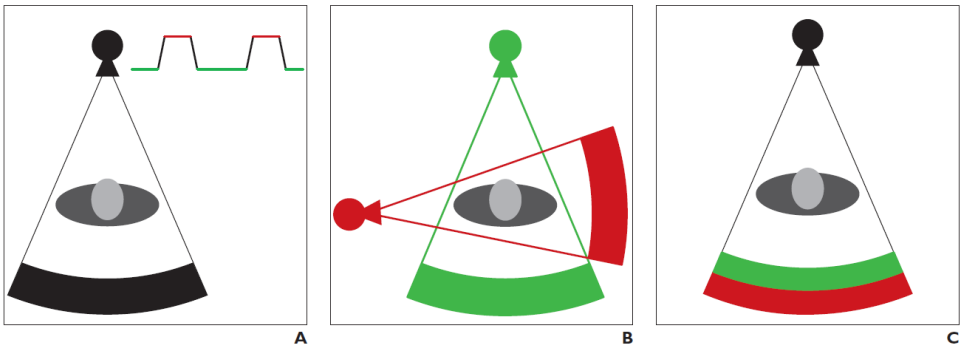


Fig. 2.8 Dual Energy CT acquisition configurations. A. Only one tube and one detector are used in the rapid kilo-voltage switching device. The voltage is rapidly cycled between two levels. B. Dual-source CT system with two tubes operating at different tube voltages and two detectors mounted orthogonally. C. Dual-layer detection setup consisting of two layers detectors with different sensitivity profiles and one X-ray tube. Reprint from (Johnson 2012)

(Johnson 2012).

Dual-Source CT utilizes two tubes operating at different voltages, and corresponding detectors mounted orthogonally in the gantry (Figure 2.8B). This solution needs double the hardware cost, yet it provides significant DECT benefits: voltage, current, and filter settings can be selected independently for each tube to ensure optimal spectral contrast, sufficient transmission, and the least amount of overlap; despite the angular offset between both spiral paths, the data acquisition does not require a time offset because equivalent z-axis positions are scanned at the same time in both orthogonal systems. The main issue with orthogonal setups is cross-scatter radiation, which partially hits non-corresponding detectors and needs to be corrected. However, dual-source CT systems use specific detector elements for measuring and correcting cross-scatter radiation (Johnson 2012).

The dual-layer detection approach uses a two layers energy-resolving detector and the polychromatic spectrum of one X-ray tube (Figure 2.8C). The scintillator material in a layer detector determines the sensitivity of the two layers. For example, ZnSe or CsI should be used in the top layer, while Gd_2O_2S should be used in the bottom layer. The scintillator materials determine the spectral resolution, but sensitivity profiles have a rather broad overlap since the available materials have overlapping sensitivity profiles (Johnson 2012).

In rapid voltage switching, one X-ray source is used, with the tube voltage alternating between high and low voltages. The transmission data are collected twice for every projection or, in practice, for consecutive projections. The additional projections and rise and fall times of the voltage modulation require a slower rotation speed. Another downside is the low photon output at low voltages, which causes excessive noise and necessitates the use of a relatively large current and,

therefore, dose to the patient (Johnson 2012).

A multi-spectral CT with photon-counting detectors that discriminate energy may be a robust solution for dual-energy, or multi-energy, data acquisition. The spectral CT technique uses photon-counting detectors, which can acquire spectral information for several bins of energy simultaneously.

2.1.6.2 Application of Spectral CT

As discussed in Section 2.1.6, dual-energy and spectral CT imaging allow discriminating the transmitted photons between different energies. The technique will enable us to bypass many of the limitations of conventional CT approaches and opens up many new application possibilities. From Dual Energy CT it is possible to obtain material-nonspecific and material-specific energy-dependent information, and both evaluations can be qualitative or quantitative. The material-nonspecific energy-dependent information includes virtual mono-energetic imaging for beam hardening suppression, effective atomic map, and electron density map. The material-specific energy-dependent information includes material decomposition, material labeling, and material highlighting (Goo and Goo 2017). Detailed material decomposition methods will be introduced in Chapter 6.

2.2 Image Reconstruction Techniques

The previous sections described the different CT configurations in which the human body can be scanned and how incident photons are transmitted and collected. The next challenge lies in reconstructing images from the collected data. This is the fundamental problem of computed tomography: from an object tomographic measurement, or more precisely, its projection, reconstruct the object. This problem is a mathematical problem that has been addressed utilizing analytical methods, iterative statistical methods, and, more recently, machine learning approaches.

2.2.1 Analytical methods

Analytical methods are the pioneers in medical image reconstruction. They offer fast and accurate reconstruction. However, they are based on simplified models that are somehow unrealistic. For example, the measurement noise is ignored and treated utilizing filtering operations. Analytical methods generally provide integral-form solutions by assuming the measurements follow a continuous behavior. Moreover, they required specific standard geometries (e.g., parallel beam and complete sampling in radial and angular coordinates) (Fessler 2009).

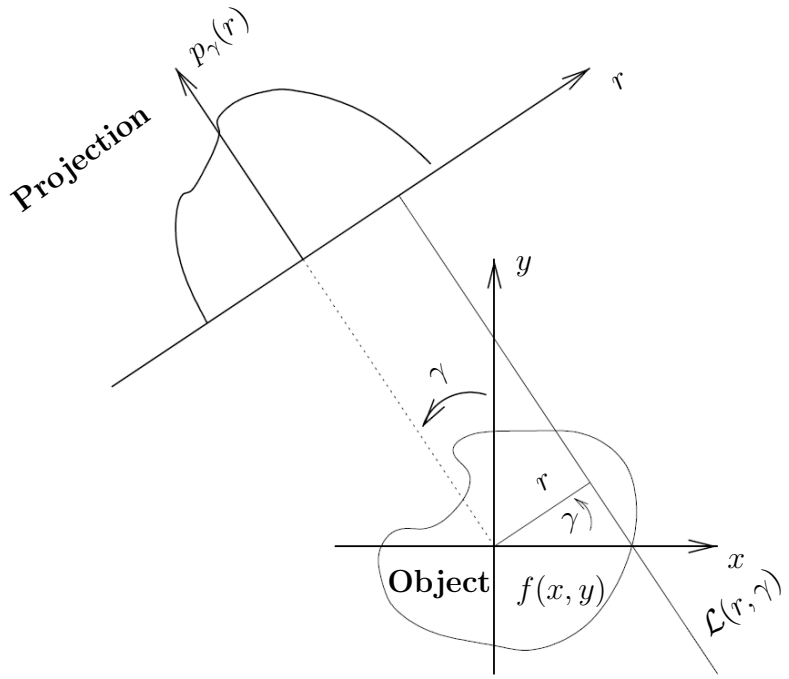


Fig. 2.9 Schematic representation of the line integrals associated with the Radon transform. Reprint from Fessler (2009)

2.2.1.1 Radon Transform

Reconstruction methods based on analytical approach are based on the *Radon transform*, which relates 2D functions $f(x, y)$ to a collection of *line integrals* of those functions. It was first introduced by the Austrian Mathematician Johann Radon in April 1917 at the annual meeting of the Royal Saxonian Society of Physical and Mathematical Sciences (Radon 1986).
Assuming an idealized scanner system, the scanner detector measurements can be represented according to Beers Law:

$$I_f(r, \gamma) = I_{f0}\, e^{-\int_{\mathcal{L}(r,\gamma)} f(x,y)\mathrm{d}x\mathrm{d}y} \tag{2.21}$$

where $\mathcal{L}(r, \gamma)$ denotes the line in the Euclidean plane forming an angle γ with the y-axis and at distance r from the origin (Fessler 2009):

$$\begin{aligned} \mathcal{L}(r, \gamma) &= \left\{(x, y) \in \mathbb{R}^2 : x \cos\gamma + y \sin\gamma = r\right\} \\ &= \{(r\cos\gamma - \ell \sin\gamma, r \sin\gamma + \ell\cos\gamma) : \ell \in \mathbb{R}\} \end{aligned} \tag{2.22}$$

The line integral through the object $f(x, y)$ along the line $\mathcal{L}(r, \gamma)$ takes the form

$$p_\gamma(r) = \int_{\mathcal{L}(r,\gamma)} f(x, y)\mathrm{d}\ell \tag{2.23}$$

$$= \int_{-\infty}^{\infty} f(r\cos\gamma - \ell\sin\gamma, r\sin\gamma + \ell\cos\gamma)\mathrm{d}\ell \tag{2.24}$$

$$\tag{2.25}$$

Thus the Radon transform of function $f(x, y)$ is defined through the operator $\mathcal{R} \to \mathcal{R}f$ with $\mathcal{R}f(x, y) = p_\gamma(r)$. The projection of $f(x, y)$ at the gantry rotation angle γ is the function $p_\gamma(\cdot)$. The 2D image reconstruction problems consist of recovering $f(x, y)$ from its projection $p_{\gamma(\cdot)}$. The Radon transform models the system imaging. In transmission tomography the scanner detector measurement is defined as

$$I_f(r, \gamma) = I_{f0}\, e^{-p_\gamma(r)} \tag{2.26}$$

Radon transform properties

The following is a list of the most notable properties of the Radon transform. We use the notation from Fessler (2009); i.e $f(x, y) \overset{\mathcal{R}}{\leftrightarrow} p_\gamma(r)$ is $\mathcal{R}f(x, y) = p_\gamma(r)$

- **Linearity**
  If $g(x, y) \overset{\mathcal{R}}{\leftrightarrow} q_\gamma(r)$, then $\alpha f(x, y) + \beta g(x, y) \overset{\mathcal{R}}{\leftrightarrow} \alpha p\gamma(r) + \beta q\gamma(r)$
- **Shift / translation**
  $f(x - x_0, y - y_0) \overset{\mathcal{R}}{\leftrightarrow} p_\gamma(r - x_0\cos\gamma - y_0\sin\gamma)$
- **Rotation**
  $f(x\cos\gamma' + y\sin\gamma', -x\sin\gamma' + y\cos\gamma') \overset{\mathcal{R}}{\leftrightarrow} p_{\gamma-\gamma'}(r)$
- **Magnification/minification**
  $f(\alpha x, \alpha y) \overset{\mathcal{R}}{\leftrightarrow} \frac{1}{|\alpha|} p_\gamma(\alpha r), \quad \alpha \neq 0$
- **Flip**
  $f(x, -y) \overset{\mathcal{R}}{\leftrightarrow} p_{\pi-\gamma}(-r)$
  $f(-x, y) \overset{\mathrm{R}}{\leftrightarrow} p_{\pi-\gamma}(r)$
  $p_\gamma(-r) = p_\gamma + \pi(r)$
- **Laplacian**
  $\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) f(x, y) \overset{\mathcal{R}}{\leftrightarrow} \frac{\partial^2}{\partial r^2} p_\gamma(r)$

If we display the projections $p_\gamma(r)$ of a 2D Dirac impulse, where usually r and γ are the horizontal and vertical axes respectively, then the projection image is

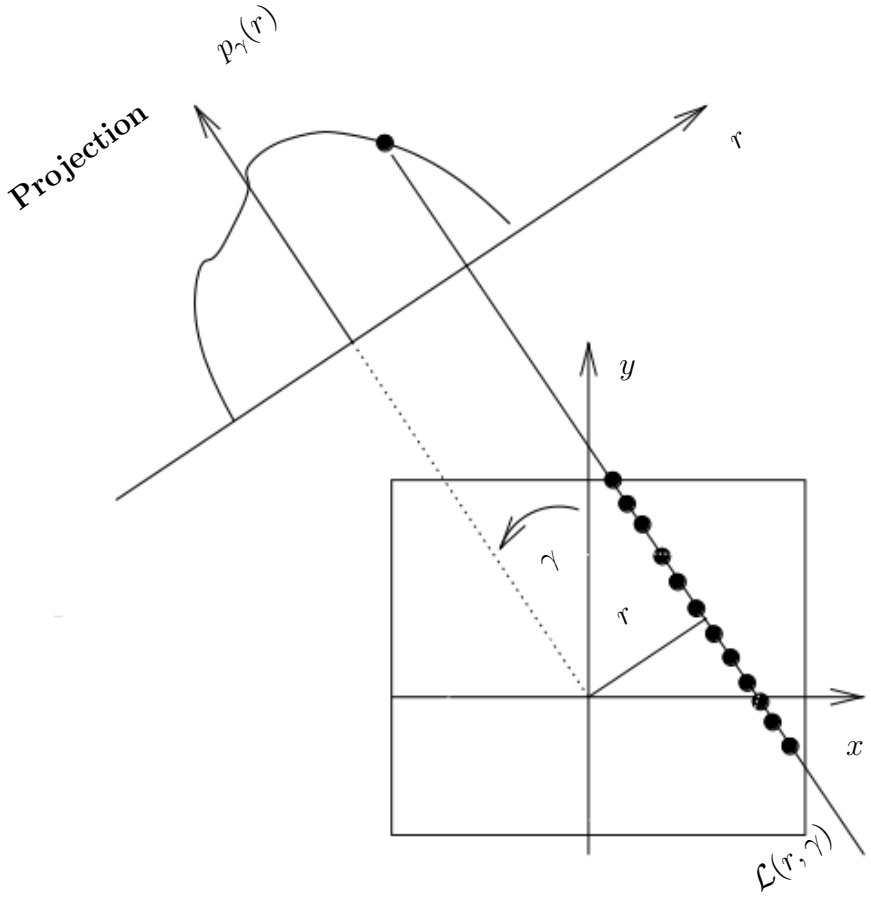


Fig. 2.10 Schematic representation of the back projection operation for a single projection view. Reprint from Fessler (2009)

a sinusoid corresponding to the function $r = x_0 \cos\gamma + y_0 \sin\gamma$. This function is termed sinograms, and represents the raw data used to reconstruct an image.

Back projection

The straightforward approach to recover the object represented by the function $f(x, y)$ from the projections $p_\gamma(r)$ is to take each sinogram value and spread it back into the object space along the line integral (Figure 2.10). In image reconstruction, this operation is named back projection. However, this operation does not retrieve the object $f(x, y)$. It produces a blurred version of the object $f_\mathrm{b}(x; y)$ which is called laminogram.

The back projection operation can be written as

$$f_\mathrm{b}(x, y) = \int_0^\pi p_\gamma \left(x\cos\gamma + y\sin\gamma\right)\mathrm{d}\gamma, \tag{2.27}$$

which corresponds to the transpose of the Radon transform. The practical backprojection are performed utilizing four distinct approaches: rotation-based backprojection, ray-driven backprojection, pixel-driven backprojection and distance-driven backprojection (De Man and Basu 2002).

2.2.1.2 Inverse Radon Transform

In order to recover the object $f(x, y)$, one must compute the Inverse Radon transform. There exist several alternatives, e.g. direct Fourier reconstruction based on the Fourier-slice theorem, the back project-filter method based on the laminogram and Filtered Back Projection (FBP) method. FBP is one of the most popular and used method in image reconstruction. The following section describes the FBP algorithm.

Filtered back projection

The filtered back projection approach is based on the Fourier-slice theorem, also known as the central-slice theorem or projection-slice theorem. It states the following: "If $p_\gamma(r)$ is the Radon transform of the function $f(x; y)$, then the One-dimentional (1D) Fourier transform of $p_\gamma(r)$ equals the slice at angle γ through the 2D Fourier transform of $f(x; y)$". Mathematically, if we denote $P_\gamma(\nu)$ as the 1D Fourier transform of $p_\varphi(r)$:

$$P_\gamma(\nu) = \int_{-\infty}^{\infty} p_\gamma(r) \mathrm{e}^{-\imath 2\pi\nu r} \, \mathrm{d}r \tag{2.28}$$

and $F(u, v)$ the 2D Fourier transform of $f(x, y)$

$$F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \mathrm{e}^{-\imath 2\pi(ux+vy)} \mathrm{d}x \, \mathrm{d}y \tag{2.29}$$

then the Fourier-slice theorem can be written as follow

$$P_\varphi(\nu) = F(\nu \cos\varphi, \nu \sin\varphi) \quad \forall\nu \in \mathbb{R}, \quad \forall\varphi \in \mathbb{R}, \tag{2.30}$$

The FBP uses the Fourier Slice theorem as follow

$$\begin{aligned} f(x, y) &= \iint F(u, v) \mathrm{e}^{\imath 2\pi(xu+yv)} \mathrm{d}u \, \mathrm{d}v && (2.31)\\ &= \int_0^{\pi} \int_{-\infty}^{\infty} F(\nu \cos\gamma, \nu \sin\gamma) \mathrm{e}^{\imath 2\pi\nu(x\cos\gamma + y\sin\gamma)} |\nu| \mathrm{d}\nu \mathrm{d}\gamma && (2.32)\\ &= \int_0^{\pi} \int_{-\infty}^{\infty} P_\gamma(\nu) \mathrm{e}^{\imath 2\pi\nu(x\cos\gamma + y\sin\gamma)} |\nu| \mathrm{d}\nu \mathrm{d}\gamma && (2.33)\\ &= \int_0^{\pi} \check{p}_\gamma(x\cos\gamma + y\sin\gamma) \mathrm{d}\gamma && (2.34) \end{aligned}$$

where the filtered projection $\check{p}_\gamma$ is defined as

$$\check{p}_\gamma(r) = \int_{-\infty}^{\infty} P_\gamma(\nu) |\nu| \mathrm{e}^{\imath 2\pi\nu r} \, \mathrm{d}\nu \tag{2.35}$$

where $|\nu|$ represents the Ramp filter (due to its shape) applied to the frequency domain. The FBP method summarizes as follow (Fessler 2009)

$f \rightarrow$ Projection $\rightarrow p_\gamma \rightarrow$ Ramp filters $\rightarrow \check{p}_\gamma \rightarrow$ Backprojection $\rightarrow \hat{f}$

- Compute the 1D Fourier transform of the projection $p_\gamma(\cdot)$ at each projection angle γ to obtain $P_\gamma(\nu)$
- Compute $\check{P}_\gamma$ by multiplying P_γ and the Ramp filter $|\nu|$, i.e., $\check{P}_\gamma(\nu) = |\nu| P_\gamma(\nu)$
- For each angle γ compute the inverse 1D Fourier transform $\check{P}_\gamma(\nu)$ to obtain the filtered projection $\check{p}_\gamma(r)$ (Equation 2.35)
- Backproject the filtered sinogram using 2.27 to obtain $\hat{f}(x, y)$, i.e.

$$\hat{f}(x, y) = \int_0^\pi \check{p}_\gamma(x \cos\gamma + y \sin\gamma) \mathrm{d}\gamma. \tag{2.36}$$

2.2.1.3 Model Based Iterative Reconstruction

Analytical methods, which are based on model simplicity, are limited by many drawbacks as outlined in Section 2.2.1. Statistical image reconstruction techniques can help overcome these limitations. These iterative statistical reconstructions provide accurate physics models that include the X-ray spectrum and scatter, which can improve beam hardening artifacts. It is possible to incorporate detector characteristics such as the focal spot size and spatial detector response into the model, which improves spatial resolution. By incorporating the spectral detector response (e.g., photon-counting detectors), one can improve the contrast between different materials. Statistical methods can model non-standard geometries, including irregular angular sampling in "next-generation" geometries, limited angular range, and "missing" data such as sparse views. Object constraint can be incorporated which, helps to reduce image artifacts (e.g., non-negativity constraints, object support, piece-wise smoothness, object sparsity, motion models, dynamic models). Several statistical models have been proposed in the literature. Table 2.1 illustrates a list of the most popular statistical reconstruction methods for X-ray CT.

2.2.1.4 Discrete model

As discussed in section 2.1.2 the incident photons that interact with the human body follows Lamber Beer's law. We derivated Beer's law in a continuous formulation. This section will derive Beer's law in its discrete form.

Model Based Iterative Reconstruction for X-ray CT
Algebraic reconstruction technique (ART) (Gordon et al. 1970)
Simultaneous Algebraic reconstruction technique (SART) (Andersen and Kak 1984)
Simultaneous iterative reconstruction technique (SIRT) (Gilbert 1972)
Multiplicative algebraic reconstruction technique (Lent and Censor 1991, Badea and Gordon 2004)
Iterative coordinate descent (Thibault et al. 2007, Sauer and Bouman 1993, Bouman and Sauer 1996)
Roughness regularized Least Square for tomography (Kashyap and Mittal 1975)
Ordered-subsets algorithms (Erdogan and Fessler 1999, Beekman and Kamphuis 2001, Lee 2000)

Table 2.1 Statistical reconstruction methods for X-ray CT.

Let i denote the index of the pixel detector locations, where $i = 1, \ldots, n$. Generally in transmission scans and in modern X-ray CT systems $n \approx 10^5 - 10^6$. Let b_i denote the number of photons collected in the detector when there is no patient (blank scan). This value b_i depends on the X-ray source intensity, the scan duration, and the detector efficiency at the source photon energy [1].

Denote y_i a random variable representing the number of photons counted in the detector for the *ith* ray. A statistical model for the transmission measurement assumes that they are independent Poisson random variables with means given by (Fessler 2000).

$$E[Y_i] = b_i \exp\left(-\int_{L_i} \mu_0(\vec{x})dl\right) + s_i \tag{2.37}$$

where s_i represents the background events (such as random coincidences, scatter, and cross-talk). The reconstruction problem consists of estimating μ from the measurement realizations $\{y_i = Y_i\}_{i=1}^{N_Y}$ (the discrete sinograms). Image reconstruction naturally becomes a statistical problem due to the primary concern of noise. Moreover, since the numbers of measurements is finite $\boldsymbol{\mu}$ can be represented with a finite parametrization. An approach to parameterize the linear attenuation coefficient map is through a finite basis expansion as follow

$$\mu_0(\vec{x}) = \sum_{j=1}^{N_P} \mu_j \chi_j(\vec{x}) \tag{2.38}$$

[1] The detector efficiency is the ratio of the number of photons measured by the detector to the number of incident photons.

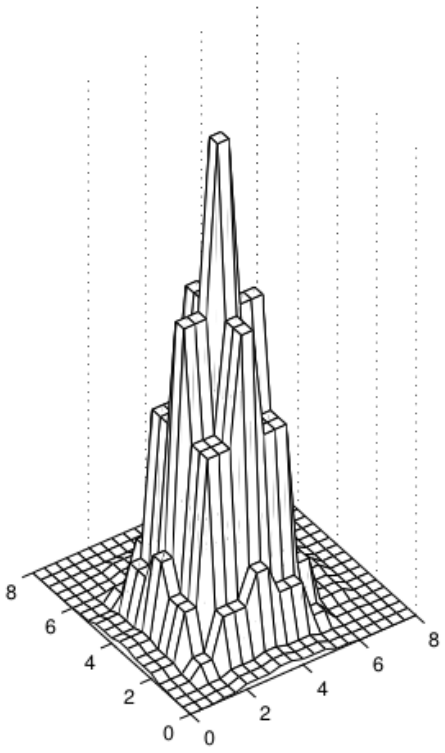


Fig. 2.11 Illustration of the function $\mu(x, y)$ parametrized utilizing pixel basis functions. Reprint from (Fessler 2000)

where N_{P} denotes the number of coefficients μ_j , and $\chi_j(\vec{x})$ the basis functions. Since $\mu \geqslant 0$, one would like to represent the basis functions as non-negative functions. Conventionally, these basic functions are the "pixels" or "voxels". The pixel basis function $\chi_j(\vec{x})$ is 1 inside the jth pixel and 0 everywhere else (Fessler 2000).

$$\chi_j(x, y) = \mathrm{rect}\left(\frac{x - x_j}{\Delta}\right) \mathrm{rect}\left(\frac{y - y_j}{\Delta}\right) \tag{2.39}$$

where Δ is the pixel width and (x_j, y_j) is the center of the jth pixel. Figure 2.11 illustrates the parametrization. It provides piece-wise-constant approximation to μ.

At this stage, the problem of estimating the linear attenuation coefficients map reduces to estimating the vector $\boldsymbol{\mu} = \left[\mu_1, \ldots, \mu_{N_p}\right]$ from the set of measurements $\boldsymbol{y} = [Y_1, \ldots, Y_{N_Y}]$ and the line integral becomes

$$\int_{L_i} \mu_0(\vec{x}) dl = \int_{L_i} \sum_{j=1}^{N_p} \mu_j \chi_j(\vec{x}) dl = \sum_{j=1}^{N_p} \mu_j \int_{L_i} \chi_j(\vec{x}) dl = \sum_{j=1}^{N_p} a_{ij} \mu_j, \tag{2.40}$$

where a_{ij} denotes the normalized strip integrals (Lo 1988) along the ith ray passing

through the jth pixel

$$a_{ij} \triangleq \int_{L_i} \chi_j(\vec{x})dl \tag{2.41}$$

The discrete measurement model simplifies to

$$y_i \sim \text{ Poisson } \{\bar{y}_i(\boldsymbol{\mu}_{\text{true}})\}, i = 1, \ldots, N_Y \tag{2.42}$$

where

$$\bar{y}_i(\boldsymbol{\mu}) \triangleq b_i \mathrm{e}^{-[\boldsymbol{A}\boldsymbol{\mu}]_i} + s_i \tag{2.43}$$

with

$$[\boldsymbol{A}\boldsymbol{\mu}]_i \triangleq \sum_{j=1}^{N_p} a_{ij}\mu_j \tag{2.44}$$

where $\boldsymbol{A} = \{a_{ij}\}$ is the system matrix.

2.2.1.5 Maximum Likelihood estimation

Maximum Likelihood (ML) estimation is a probabilistic approach for estimating $\boldsymbol{\mu}$ from the observable $\boldsymbol{y}$. A maximum likelihood estimate of $\boldsymbol{\mu}$ is the value $\hat{\boldsymbol{\mu}}$ that maximizes the likelihood function (Fessler 2000).

$$\hat{\boldsymbol{\mu}} = \arg\max_{\boldsymbol{\mu}\geq 0} L(\mu), \quad L(\mu) \triangleq \log P[Y = y; \mu]. \tag{2.45}$$

Utilizing the Poisson Model 2.42 the measurement joint probability mass function is

$$P[\boldsymbol{Y} = \boldsymbol{y}; \boldsymbol{\mu}] = \prod_{i=1}^{N_Y} P\left[Y_i = y_i; \boldsymbol{\mu}\right] = \prod_{i=1}^{N_Y} \frac{e^{-\bar{y}_i(\boldsymbol{\mu})}\left[\bar{y}_i(\boldsymbol{\mu})\right]^{y_i}}{y_i!} \tag{2.46}$$

Applying the log to the condition probability 2.46 the log-likelihood function takes the form

$$L(\boldsymbol{\mu}) = \sum_{i=1}^{N_Y} \left(y_i \ln \bar{y}_i(\boldsymbol{\mu}) - \bar{y}_i(\boldsymbol{\mu}) - \ln y_i!\right) \tag{2.47}$$

The term $\ln y_i!$ is constant and may be neglected for optimization. Thus, the log-likelihood takes the form

$$L(\boldsymbol{\mu}) = \sum_{i=1}^{N_Y} y_i \ln \bar{y}_i(\boldsymbol{\mu}) - \bar{y}_i(\boldsymbol{\mu}) \tag{2.48}$$

Having the likelihood function, the challenge will be finding an appropriate optimization algorithm to maximize 2.48.

2.2.1.6 Penalized Maximum Likelihood estimation

If we maximize the log-likelihood function alone, the result will lead to a noisy image because the transmission tomography is an ill-conditioned problem.

An alternative could be to include a penalty function that favors reconstructed images that are piece-wise smooth. This procedure is known as regularization. The expected value of the attenuation coefficients map is obtained by maximizing the penalized-likelihood objective function

$$\hat{\boldsymbol{\mu}} \triangleq \arg\max_{\boldsymbol{\mu}\geq 0} \Phi(\boldsymbol{\mu}), \quad \Phi(\boldsymbol{\mu}) \triangleq L(\boldsymbol{\mu}) - \beta R(\boldsymbol{\mu}), \tag{2.49}$$

where $R(\boldsymbol{\mu})$ denotes the penalty term and β is a parameter which controls the relative contributions of the data fidelity term (the log-likelihood function) and of the penalty term.

Bayesian approach

The Bayes rule applied to the likelihood probability also leads to objective functions of the form 2.49. The Bayes rules is mathematically formulated as follow (Bayes 1763)

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \tag{2.50}$$

where

- $P(A \mid B)$: Conditional probability defined as the likelihood of an event A occurring if B is true. The posterior probability of A given B is another name for it.
- $P(B \mid A)$: Conditional probability defined as the probability of event B occurring given that A is true. It can also be interpreted as the probability of A given a fixed B because $P(B \mid A) = L(A \mid B)$.
- $P(A)$ and $P(B)$ are the likelihood of observing A and B respectively without a given conditions; they are known as the prior probability.

Let assume $\boldsymbol{\mu}$ is a random vector corresponding to a prior distribution $f(\boldsymbol{\mu})$ that is proportional to $e^{-\beta R(\boldsymbol{\mu})}$. (Markov Random Field models for images entail such priors by nature (Besag 1986)). The Maximum a Posteriori (MAP) estimate of $\boldsymbol{\mu}$ is the value that maximizes the posterior distribution $f(\boldsymbol{\mu}|\boldsymbol{y})$. By Bayes rule:

$$f(\boldsymbol{\mu} \mid \boldsymbol{y}) = \frac{f(\boldsymbol{y} \mid \boldsymbol{\mu})f(\boldsymbol{\mu})}{f(\boldsymbol{y})} \tag{2.51}$$

and applying the logarithm, the log posterior takes the form

$$\log f(\boldsymbol{\mu} \mid \boldsymbol{y}) \equiv \log f(\boldsymbol{y} \mid \boldsymbol{\mu}) + \log f(\boldsymbol{\mu}) \equiv L(\boldsymbol{\mu}) - \beta R(\boldsymbol{\mu}) \tag{2.52}$$

It's worth noting that $f(\boldsymbol{y} \mid \boldsymbol{\mu})$ is proportional to the likelihood function, with the exception of a constant that makes it a proper density. Furthermore, the marginal probability $f(\boldsymbol{y})$ serves as a normalizing constant, ensuring that the posterior density is appropriate. Therefore, the MAP estimation is computationally equivalent to the penalized maximum likelihood estimation.

Penalty function: For many authors the attenuation coefficients maps are considered piece-wise smooth functions. If attenuation maps are piece-wise smooth, it makes sense for the penalty function $R(\boldsymbol{\mu})$ to discourage images that are too rough. The most basic penalty function for roughness discouragement examines the differences between nearby pixel values:

$$R(\boldsymbol{\mu}) = \sum_{j=1}^{N_p} \frac{1}{2} \sum_{k=1}^{N_p} w_{jk} \psi\left(\mu_j - \mu_k\right) \tag{2.53}$$

where $w_{jk} = w_{kj}$.
For the four horizontal and vertical neighboring pixels $w_{jk} = 1$ and for diagonal neighboring pixels $w_{jk} = 1/\sqrt{2}$. Typical choices of the potential function ψ are the Quadratic prior (O'Meara 2013), the Huber prior (Huber 1964) and the Geman prior (Geman 1987). The Huber potential is detailed in Chapter 6.

2.2.2 Optimization Algorithms

After defining the objective function, an optimization algorithm needs to be developed to maximize the objective function. If one ignores the non-negativity constraint, one could try to find $\hat{\boldsymbol{\mu}}$ analytically by zeroing the gradient of the objective function. Unfortunately, there is not closed solution to this problem, even without taking into account the non-negativity constraint and the prior. Here is where iterative methods play a roll, in order to find the maximizer of the objective function. An iterative method is a mathematical procedure which begins with an initial estimation of $\boldsymbol{\mu}^{(0)}$ of the linear attenuation coefficient and generates a sequences of improved $\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}, ...,$. The iterates $\boldsymbol{\mu}^{(n)}$ should converge as fast as possible to the solution $\hat{\boldsymbol{\mu}}$. For the purpose of designing an algorithm to optimize a penalized maximum-likelihood objective function, some characteristics must be taken into account (Fessler 2000):

- Non-negativity constraint: $(\mu \geq 0)$

- Convergence rate: (The fewer iterations the better)
- Computation time per iteration: (Minimize the number of floating point operations)
- Storage requirements: (Minimize memory usage as much as possible)

2.2.3 Quasi-Newton algorithm

One of the algorithms for optimization are the Newton methods. In order to understand the Quasi-Newton algorithm it is necessary to introduce the Newton-Raphson method. Let us consider the case of a 1D variable objective function $\Phi(\mu)$ which is twice differentiable $\Phi : \mathbb{R} \to \mathbb{R}$. One attempt to solve the optimization problem:

$$\min_{\mu \in \mathbb{R}} \Phi(\mu) \tag{2.54}$$

Newton's methods solve optimization problem 2.54 by building a sequences of $\{\mu_t\}$ from an initial guess [2] (μ_0). It uses a succession of second-order Taylor approximations of $\Phi(\mu)$ around the iterates μ_t to converge towards a minimizer. The second-order Taylor expansion of $\Phi(\mu)$ takes the form

$$\begin{aligned} \Phi(\mu) \approx & \frac{\Phi(\mu_t)}{0!}(\mu - \mu_t)^0 + \frac{\Phi'(\mu_t)}{1!}(\mu - \mu_t)^1 \\ & + \frac{\Phi''(\mu_t)}{2!}(\mu - \mu_t)^2 \end{aligned} \tag{2.55}$$

thus

$$\Phi(\mu) \approx \Phi(\mu_t) + \Phi'(\mu_t)(\mu - \mu_t) + \frac{\Phi''(\phi_t)}{2}(\mu - \mu_t)^2 \tag{2.56}$$

where $\Phi'(\cdot)$ and $\Phi''(\cdot)$ denotes the first and second derivative of $\Phi(\mu)$. The minimum can be achieved by setting the derivative to zero $(0 = \frac{\partial \Phi}{\partial \mu})$.

$$\begin{aligned} &\tfrac{\partial \Phi}{\partial \mu} = \Phi'(\mu_t) + \Phi''_{\mu_t}(\mu - \mu_t) \\ &0 = \Phi'(\mu_t) + \Phi''(\mu_t)(\mu - \mu_t] \\ &-\tfrac{\Phi'(\mu_t)}{\Phi''(\mu_t)} = \mu - \mu_t \\ &\mu_{t+1} = \mu_t - \tfrac{\Phi'(\mu_t)}{\Phi''(\mu_t)} \end{aligned} \tag{2.57}$$

The Newton methods then estimates the minimum at each iteration by computing μ_{t+1} as in 2.57. In the multi-dimensional case, equation 2.57 can be expressed as

$$\boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}_t - \frac{\nabla \Phi(\boldsymbol{\mu}_t)}{\nabla^2 \Phi(\boldsymbol{\mu}_t)} \tag{2.58}$$

[2] In transmission tomography, the initial guess may be an initial reconstruction of the sinograms utilizing an analytical method, for example FBP.

where ∇ denotes the gradient and ∇^2 denotes the exact Hessian.

A more generic version of 2.57 is what is known as the line search technique, of which Newton's method is an example. Line search iterations compute the search direction p_t and decides how far to move along it (Nocedal and Wright 2006*a*).

$$\boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}_t + \eta_t p_t, \tag{2.59}$$

where η_t is called the step length. The search direction p_t has the general form

$$p_t = -\boldsymbol{B}_t^{-1} \nabla \Phi(\boldsymbol{\mu}_t) \tag{2.60}$$

where $\boldsymbol{B}_t$ is a symmetric and non-singular matrix. With $\boldsymbol{B}_t$ being the identity matrix, one have the steepest gradient descent algorithm (Nocedal and Wright 2006*a*). For the gradient descent the iterates take the form

$$\boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}_t - \eta_t \nabla \Phi(\boldsymbol{\mu}_t), \tag{2.61}$$

In Newton's method as showed in 2.58 , $\boldsymbol{B}_t$ is the exact Hessian.

Gradient descent uses the first-order Taylor expansion at the current optimized location to estimate the shape of the optimization space, while Newton's approach uses the second-order Taylor expansion. From the graphic point of view, Newton's uses a quadratic "bowl" with local curvature to approximate the shape of the presently optimized point. The second derivative informs the curvature at the current point and takes steps that are inversely proportional to the degree of "steepness" (very steep →tiny steps, extreme flat →huge steps). Therefore, Newton's method advances to the minimum more rapidly than gradient descent, which requires more iterations. Figure 2.12 shows a comparison between gradient descent and Newton's Method.

The Hessian must be determined on the first iteration and then fully recalculated on subsequent iterations, making Newton's technique computationally expensive. The Newton technique requires iteratively solving a linear system of equations, which is memory demanding and time consuming. Quasi-Newton techniques are an alternative to Newtonian procedures. In quasi-Newton techniques, $\boldsymbol{B}$ is an estimate of the Hessian that is updated using a low-rank formula at each iteration (Nocedal and Wright 2006*a*). An example of a quasi-Newton algorithm is the Broyden–Fletcher–Goldfarb–Shanno algorithm proposed by Charles George Broyden, Roger Fletcher, Donald Goldfarb and David Shanno (Broyden 1970, Goldfarb 1970, Shanno 1970).

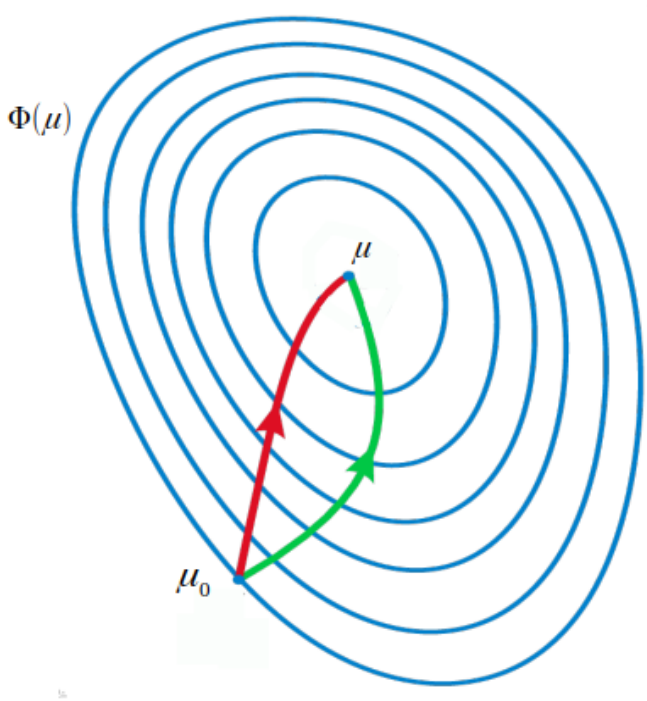


Fig. 2.12 A gradient descent (green) versus Newton's method (red) comparison for minimizing a function. Newton's method relies on curvature information (i.e. the second derivative) to reach more direct the minimum. Reprint from (Alexandrov 2021)

2.2.4 L-BFGS

The L-BFGS iterative solver estimates $\boldsymbol{\mu}^t$ starting the previous iterate $\boldsymbol{\mu}^{t-1}$. We define the first estimate as $\boldsymbol{\mu}^{(0)}$. Given a current estimate $\boldsymbol{\mu}^{(t)}$, the new estimate $\boldsymbol{\mu}^{(t+1)}$ is obtained as

$$\begin{aligned} \boldsymbol{\mu}^{(t+1)} &= \boldsymbol{\mu}^{(t)} - s^{\star}(\boldsymbol{B}^{-1})^{(t)}\nabla\Phi(\mu^{(t)}) \\ \text{with} \quad s^{\star} &= \underset{s\in[0,1]}{\arg\max}\,\chi(s) \\ \text{and} \quad \chi(s) &= \Phi\left(\mu^{(t)} - s(\boldsymbol{B}^{-1})^{(t)}\nabla\Phi(\mu^{(t)})\right) \end{aligned} \tag{2.62}$$

where $(\boldsymbol{B}^{-1})^{(t)}$ is an approximate inverse Hessian of Φ evaluated at $\boldsymbol{\mu}^{(t)}$. The matrix/vector product $(\boldsymbol{B}^{-1})^{(t)}\nabla\Phi(\boldsymbol{\mu}^{(t)})$ in (2.62) is directly computed (without storing $(\boldsymbol{B}^{-1})^{(t)}$) from the m previous iterates $\boldsymbol{\mu}^{(t-p)}$, $p = 0, \ldots, m-1$. An approximate solution of the line-search sub-problem is obtained by backtracking to match the *Wolfe Conditions*. The iterative scheme (i.e., w.r.t. t) is repeated until a convergence criterion is met. A more detailed explanation can be found in Bousse et al. (2019).

In L-BFGS the next step direction is calculated as the approximate inverse Hessian times the gradient, but it only needs to store the last several gradient

updates, not the approximate inverse Hessian.

In the experiments presented in this thesis, we employed the L-BFGS method to optimize the objective function with regard the linear attenuation coefficient.

CHAPTER 3

Sparse Regularization for Inverse Problem

The present chapter describes the CS theory and provides a literature review of the main sparse recovery algorithms used in the thesis (OMP, IST, IHT). It describes the Total Variation semi norm, the Dictionary Learning approaches and the main optimization algorithms for patch-based dictionary learning. The CAOL and the Block Proximal Extrapolated Gradient with a Majorizer algorithm for the optimization of the CAOL algorithm are detailed.

3.1 Compressed Sensing Theory

CS is a technique for recovering a signal from fewer samples than the Nyquist-Shannon sampling theorem requires (Sher 2019). The essential assumption of the CS theory is that most signals in real applications have a sparse representation in a certain transform domain with just a few of them being significant and the rest being zero or negligible. Another essential condition is that measurements in the signal acquisition domain are incoherent. That is, the distances between sparse signals are roughly preserved as the distances between the observations made by the sampling process (Orović et al. 2016, Marques et al. 2018). CS assists in reducing the energy required for transmission and storage by projecting the information into a smaller dimensional space. It reduces power consumption by lowering the sampling rate to the signal's information content rather than its bandwidth (Marques et al. 2018, Donoho 2006). The CS process is divided into three fundamental steps as shown in 3.1.

The *sparse representation* step is the process of expressing a signal using a small number of projections on an appropriate basis. A vector signal $\boldsymbol{x} \in \mathbb{R}^N$ is *s-sparse* if s elements of its entries are non-zero, where s is denoted as the sparsity level. Mathematical, this can be written as (Draganic et al. 2017)

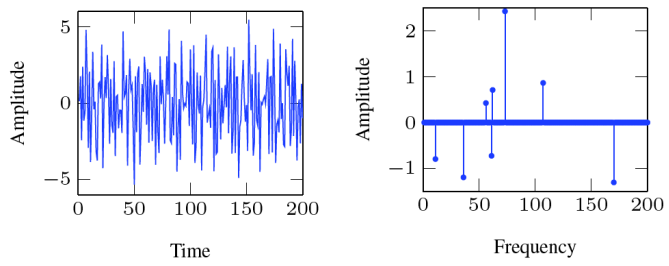
$$\|\boldsymbol{x}\|_0 = \lim_{p \to 0} \sum_{i=1}^{N} |x_i|^p \leqslant s \tag{3.1}$$

Fig. 3.1 The primary technique of compressive sensing.

If a signal is not sparse, it can be sparsified by simply representing it as a suitable basis. For instances, a linear combination of $s \ll N$ basis vectors. Fundamentally, the signal $\boldsymbol{x}$ can be represented with N basis vectors $\{\Upsilon_i\}_{i=1}^{N}$. Let $\boldsymbol{\Upsilon}$ be an $N \times N$ basis matrix, the sparse representation of the signal $\boldsymbol{x}$ becomes the vector $\boldsymbol{z}$ as (Marques et al. 2018)

$$\boldsymbol{x} = \boldsymbol{\Upsilon}\boldsymbol{z} \tag{3.2}$$

A visual example of how the sparse representation works is illustrated in Figure 3.2. It depicts a 200-sample time-domain signal with 8 different sinusoids. It is a frequency domain representation of an 8-sparse signal. That means, there are only 8 non-zero values among the 200 frequency (Marques et al. 2018). Other examples of sparse representation are Wavelet Transform (WT), Fast Fourier Transform (FFT), and Discrete Cosine Transform (DCT).



(a) Time domain (b) Frequency domain

Fig. 3.2 8 sinusoidal samples in (a) time and (b) frequency domains. Reprint from Marques et al. (2018).

The *CS Acquisition* process entails obtaining a few measurements $\boldsymbol{y} \in \mathbb{R}^M$ from the sparse signal, with $M \ll N$. Obtaining the measurements consist of sampling the signal $\boldsymbol{x}$ according to a matrix $\boldsymbol{\Phi} \in \mathbb{R}^{M \times N}$, where ϕ_i denotes the i^{th} column of $\boldsymbol{\Phi}$:

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{n} \tag{3.3}$$

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{\Upsilon}\boldsymbol{z} + \boldsymbol{n} \tag{3.4}$$

$$\boldsymbol{y} = \boldsymbol{D}\boldsymbol{z} + \boldsymbol{n} \tag{3.5}$$

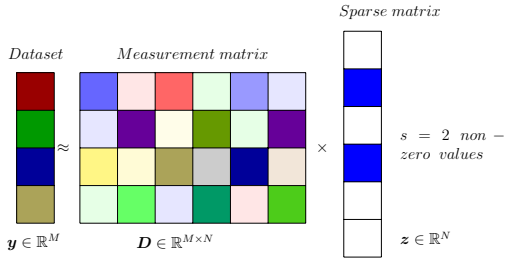


Fig. 3.3 Matrix representation of the Compressive Sensing metrics.

where $\boldsymbol{n}$ is the noise and $\boldsymbol{D} \in \mathbb{R}^{M \times N}$ ($\boldsymbol{D} = \boldsymbol{\Phi}\Upsilon$) is the measurement matrix. To reconstruct the original signal from the few selected measurements, the reduction from N to M signals must preserve the information stored in the s-space.

The last step in the CS process is the *sparse recovery*. It consists of recovering the sparse signal from the few measurements $\boldsymbol{y}$ through a sparse recovery algorithm (Arjoune et al. 2017)

CS theory covers two major issues: the design of the matrix $\boldsymbol{D}$; and implementation of an efficient sparse recovery method for the estimation of the sparse vector $\boldsymbol{z}$, given $\boldsymbol{y}$, $\boldsymbol{D}$ and s.

The measurement matrix design must be in such a way that the relevant information of any s-sparse signal is contained in this matrix. The ultimate objective is to create a suitable measurement matrix with $M \approx s$.
Moreover, the measurement matrix should satisfy the Restricted Isometry Property (RIP) (Donoho 2006):

$$(1 - \delta_s) \|\boldsymbol{z}\|_2^2 \leq \|\boldsymbol{D}\boldsymbol{z}\|_2^2 \leq (1 + \delta_s) \|\boldsymbol{z}\|_2^2 \tag{3.6}$$

where $\delta_s \in (0, 1)$ is the Restricted Isometry Constant (RIC) value and denotes the lowest number that satisfies 3.6. If the measurement matrix $\boldsymbol{D}$ fulfills the RIP, an accurate estimation of the sparse signal $\boldsymbol{z}$ can be obtained using a recovery technique, such as solving an l_p-norm problem (Wen et al. 2015). Once the measurement matrix is appropriately defined, the sparse recovery consists of finding the sparse vector $\boldsymbol{z}$ by solving:

$$\min_{\boldsymbol{z}} \|\boldsymbol{z}\|_p \quad s.t. \quad \boldsymbol{D}\boldsymbol{z} = \boldsymbol{y} \tag{3.7}$$

where $\|\boldsymbol{z}\|_p$ is the l_p-norm of $\boldsymbol{z}$ with $0 < p < 2$. The system 3.7 contains an unlimited number of solutions when $M < N$, with some exceptions. The problem 3.7 is NP-hard (non-deterministic polynomial-time hardness): there are no algorithms that can ensure it will always be solved (Dumitrescu and Irofti 2018). Figure 3.3 shows the connection between the variables in the noiseless scenario. Each column of the matrix $\boldsymbol{D}$ is called atom.

Several methods for sparse recovery have been proposed in the literature.

<table>
<tr><th colspan="3">Sparse Recovery Algorithms</th></tr>
<tr><th>Convex
Relaxation</th><th>Non-Convex
Optimization</th><th>Greedy
Algorithms</th></tr>
<tr><td>Approximate
Message
Passing
(Donoho et al. 2010)</td><td rowspan="2">Bayesian
Compressive
Sensing
(Ji et al. 2008)</td><td>Matching
Pursuit
(Mallat and Zhang 1993)</td></tr>
<tr><td>Basis
Pursuit
(Chen and Donoho 1994)</td><td>Orthogonal
Matching
Pursuit
(Pati et al. 1993)</td></tr>
<tr><td>Least
Angle
Regression
(Efron et al. 2004)</td><td rowspan="2">Focal
Under determined
System Solution
(Gorodnitsky and Rao 1997)</td><td>Subspace
Pursuit
(Dai and Milenkovic 2009)</td></tr>
<tr><td>Gradient
Projection
for Sparse
Reconstruction
(Figueiredo et al. 2007)</td><td>Stage-wise
Orthogonal
Matching Pursuit
(Donoho et al. 2012)</td></tr>
<tr><td>Least Absolute
Shrinkage
and Selection
Operator
(Tibshirani 1996a)</td><td rowspan="2">Iterative
Reweighted
Least Squares
(Burrus et al. 1994)</td><td>Regularized
Orthogonal
Matching Pursuit
(Needell D. 2009)</td></tr>
<tr><td>Iterative Soft
Thresholding
(Daubechies et al. 2004)</td><td>Iterative Hard
Thresholding
(Blumensath and Davies 2008)</td></tr>
</table>

Table 3.1 List of sparse recovery algorithms according to their classification in Convex Relaxation, Non-Convex Optimization and Greedy Algorithms.

They are mainly classified into three categories: convex relaxations, non-convex optimization techniques, and greedy algorithms (Marques et al. 2018). The convex relaxations algorithms replace the l_p-norm by a smooth approximation. For instance, replacing it by l_1-norm or by a smooth function (Elad 2010). The non-convex optimization techniques solve the challenge of sparse recovery by using prior knowledge of the sparse signal distribution. The greedy techniques recover the sparse signal iteratively (Donoho 2006). They are usually extremely speedy. Table 3.1 displays a list of sparse recovery algorithms based on the previous categorization. We selected the most relevant algorithms among the extensive approaches existing in the literature. The following sections provide a detailed explanation of the most relevant sparse recovery algorithms covered in the thesis.

3.1.1 Orthogonal Machine Pursuit: OMP

The optimization algorithm to be solved takes the form

$$\begin{aligned} \min_{\boldsymbol{z}} \quad & \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{z}\|^2 \\ \text{s.t.} \quad & \|\boldsymbol{z}\|_0 \leq s \end{aligned} \tag{3.8}$$

If we had an appropriate guess of the sparsity level s the solution to 3.8 would be straightforward. However, in the majority of applications there is not an exact sparsity level estimate. Therefore the choice of s is based on try-and-error approach. For instance, imposing an error bound

$$\begin{aligned} \min_{\boldsymbol{z}} \quad & \|\boldsymbol{z}\|_0 \\ \text{s.t.} \quad & \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{z}\| \leq \varepsilon \end{aligned} \tag{3.9}$$

The problem 3.9 may not have a sparse solution if ε is too small. If ε is large, the solution is over sparse. Dumitrescu and Irofti (2018) suggest ε being larger than the square root of the noise variance, but of the same order of magnitude.

OMP (Pati et al. 1993) builds the sparse representation support by finding the column $d_j \in \mathbb{R}^M$ (called atom) which is best aligned with the residual vector $\boldsymbol{r}$. It chooses the atoms one by one in order to minimize the approximation error as much as possible (greedily). At each iteration, OMP adds the atom with the largest projection value to the augmented support matrix $\boldsymbol{D}_{\mathcal{S}}$, with $\mathcal{S}$ containing the indices of the selected atoms. Assuming that we know the atom coefficients at the current iteration or representation, the residual takes the form (Dumitrescu and Irofti 2018)

$$\boldsymbol{r} = \boldsymbol{y} - \sum_{j \in \mathcal{S}} z_j d_j \tag{3.10}$$

The augmented matrix is void in the first iteration. As a result, the residual is the signal $\boldsymbol{r} = \boldsymbol{y}$. The first step in the OMP algorithm consists of finding the next atom to be added. This step is performed by projecting the matrix $\boldsymbol{D}$ onto the residual or the signal in the first iteration. This is accomplished by determining which atom has the highest inner product with the residual and storing the atom index in $\mathcal{S}$. The new atom is designated as $\boldsymbol{d}_k$

$$\left|\boldsymbol{r}^T \boldsymbol{d}_k\right| = \max_{j \notin \mathcal{S}} \left|\boldsymbol{r}^T \boldsymbol{d}_j\right| \tag{3.11}$$

Then, we include the selected atom in the augmented matrix $\boldsymbol{D}_{\mathcal{S}}$, which is used to minimize the next residual. Thus, the next $\mathcal{S}$ would be $\mathcal{S} \leftarrow \mathcal{S} \cup \{k\}$.
The second step in the OMP algorithm consists of computing the new sparse

Algorithm 1: Orthogonal Matching Pursuit algorithm

Data: Measurement matrix $\boldsymbol{D} \in \mathbb{R}^{M \times N}$;
Signal $\boldsymbol{y} \in \mathbb{R}^{M}$;
Sparsity level s;
Stopping error ε;
Result: Support $\mathcal{S}$; Sparse solution $\boldsymbol{z}$
Initialization $\mathcal{S} = \emptyset, \boldsymbol{r} = \boldsymbol{y}$;
while $|\mathcal{S}| < s$ *and* $\|\boldsymbol{r}\| > \varepsilon$ **do**
  Find the index: $k = \arg\max_{j \notin \mathcal{S}} \left|\boldsymbol{r}^T \boldsymbol{d}_j\right|$;
  Build the support:$\mathcal{S} \leftarrow \mathcal{S} \cup \{k\}$;
  Find the sparse solution: $\boldsymbol{z}_{\mathcal{S}} = \min_{\boldsymbol{z}} \|\boldsymbol{y} - \boldsymbol{D}_{\mathcal{S}}\boldsymbol{z}\|$;
  Find the residual: $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{D}_{\mathcal{S}}\boldsymbol{z}_{\mathcal{S}}$
end

representation coefficients utilizing the matrix $\boldsymbol{D}_{\mathcal{S}}$ at the current iteration. These coefficients are the solution of the least squares optimization problem

$$\min_{\boldsymbol{z}} \|\boldsymbol{y} - \boldsymbol{D}_{\mathcal{S}}\boldsymbol{z}\| \tag{3.12}$$

where $\boldsymbol{D}_{\mathcal{S}}$ are the atoms of $\boldsymbol{D}$ which had the largest projection (the atoms with indices in $\mathcal{S}$). The analytical solution to 3.12 can be written as

$$\boldsymbol{z}_{\mathcal{S}} = \left(\boldsymbol{D}_{\mathcal{S}}^T \boldsymbol{D}_{\mathcal{S}}\right)^{-1} \boldsymbol{D}_{\mathcal{S}}^T \boldsymbol{y} \tag{3.13}$$

where $\boldsymbol{z}_{\mathcal{S}}$ is a vector of $|\mathcal{S}|$ dimension and contains the current non-zero values of the sparse representation $\boldsymbol{z}$. It is worth noting that at each step of OMP, all of the non-zero coefficients of the sparse representation are recalculated.
The third and last step in the OMP algorithm computes the new residual which will be used in the next iteration

$$\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{D}_{\mathcal{S}}\boldsymbol{z}_{\mathcal{S}} \tag{3.14}$$

The algorithm then continues to iterate until the stopping criteria is reached. There are two stopping criteria enclosing the optimization problem 3.8 and 3.9. The first criteria imposes a value of s as the maximum sparse level to be reached. Once the algorithm reaches the sparsity level, it stops disregarding the error bound ε. The second stopping criteria sets the error bound ε and increases the sparsity level at each iteration until the error becomes the error bound. The choice of the error bound is critical for this criteria since the sparsity level can increase to the point where the solution is no longer sparse. In Chapter 4 we implemented a GPU-accelerated version of the OMP algorithm described in algorithm 1.

3.1.2 Iterative Soft Thresholding: ISTA

The IST algorithm (Daubechies et al. 2004) is a convex relaxation method which relaxes the l_0-norm in 3.8 by an l_1-norm regularization. It solves the optimization problem

$$\min_{\boldsymbol{z}} \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{z}\|^2 + \lambda\|\boldsymbol{z}\|_1 \tag{3.15}$$

The above optimization (3.15) can be seen as a more general problem

$$\min_{\boldsymbol{z}} f(\boldsymbol{z}) + g(\boldsymbol{z}) \tag{3.16}$$

where $f, g : \mathbb{R}^m \to \mathbb{R}$ are convex functions, but only f is differentiable. Therefore, $f(\boldsymbol{z}) = \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{z}\|^2$ and $g(\boldsymbol{z}) = \lambda\|\boldsymbol{z}\|_1$. It can be solved utilizing a proximal gradient descent method (Rockafellar 1970).
The general approach of the proximal gradient descent method for minimizing a convex function $h(\boldsymbol{x})$ can be defined as

$$\boldsymbol{x}_{t+1} = \text{prox}_{\eta h}\left(\boldsymbol{x}_t - \eta\nabla h\left(\boldsymbol{x}_t\right)\right) \tag{3.17}$$

where t is the current iteration, η is the step size and prox is the proximal operator. The proximal operator applied to a function h can be defined as

$$\text{prox}_h(\boldsymbol{z}) = \arg\min_{\boldsymbol{x}} h(\boldsymbol{x}) + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{z}\|_2^2 \tag{3.18}$$

Applying the proximal gradient to equation 3.16 the IST algorithm takes the form

$$\begin{aligned} \boldsymbol{z}_{t+1} &= \arg\min_z \left\{ f\left(\boldsymbol{z}_t\right) + \left(\boldsymbol{z} - \boldsymbol{z}_t\right)^T \nabla f\left(\boldsymbol{z}_t\right) + \tfrac{1}{2\eta}\left\|\boldsymbol{z} - \boldsymbol{z}_t\right\|^2 + g(\boldsymbol{z}) \right\} \\ &= \arg\min_z \left\{ \tfrac{1}{2\eta}\left\|\boldsymbol{z} - \left(\boldsymbol{z}_t - \eta\nabla f\left(\boldsymbol{z}_t\right)\right)\right\|^2 + g(\boldsymbol{z}) \right\} \end{aligned} \tag{3.19}$$

Focusing in our specific case $g(\boldsymbol{z}) = \lambda\|\boldsymbol{z}\|_1$ and denoting $\tilde{\boldsymbol{z}}_t = \boldsymbol{z}_t - \eta\nabla f\left(\boldsymbol{z}_t\right)$ the problem 3.19 becomes

$$\boldsymbol{z}_{t+1} = \arg\min_{\boldsymbol{z}} \left\{ \frac{1}{2\eta}\left\|\boldsymbol{z} - \tilde{\boldsymbol{z}}_t\right\|^2 + \lambda\|\boldsymbol{z}\|_1 \right\} \tag{3.20}$$

which can be solved utilizing a *soft thresholding operator* applied to each element on vectors.

$$\boldsymbol{z}_{t+1} = \text{S}_{\text{th}}\left(\tilde{\boldsymbol{z}}_t, \eta\lambda\right) \tag{3.21}$$

where

$$\mathrm{S_{th}}(\xi, \alpha) = \begin{cases} \xi + \alpha, \text{ if } \xi < -\alpha \\ 0, \quad \text{if } -\alpha \leq \xi \leq \alpha \\ \xi - \alpha, \text{ if } \xi > \alpha \end{cases} \tag{3.22}$$

At each iteration the *soft thresholding operator* pulls ξ towards the origin by α. The IST algorithm is guaranteed to converge, however it convergence rate is slow. Several variations, such as the "fast ISTA" (FISTA), which uses a Nesterov's Accelerated Gradient Descent algorithm, have been developed to speed it up (Beck and Teboulle 2009).

3.1.3 Iterative Hard Thresholding: IHT

The IHT algorithm (Blumensath and Davies 2008) solve the optimization problem

$$\min_{\boldsymbol{z}} \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{z}\|^2 + \lambda \|\boldsymbol{z}\|_0 \tag{3.23}$$

At each iteration the solution is computed as

$$\boldsymbol{z}_{t+1} = H_{\lambda^{0.5}} \left(\boldsymbol{z}_t + \boldsymbol{D}^T \left(\boldsymbol{y} - \boldsymbol{D}\boldsymbol{z}_t \right) \right) \tag{3.24}$$

where $H_{\lambda^{0.5}}$ is the element-by-element hard thresholding operation

$$H_{\lambda^{0.5}}\left(z_i\right) = \begin{cases} 0 & \text{if } |z_i| \leq \lambda^{0.5} \\ z_i & \text{if } |z_i| > \lambda^{0.5} \end{cases} \tag{3.25}$$

The IHT algorithm can either finish after a set number of iterations or when the sparse vector does not change significantly between iterations. Blumensath and Davies (2009) proves that under the assumption $\|\boldsymbol{D}\|_2 < 1$ the algorithm converges to a local minimum of 3.23.

3.2 Total Variation

The TV problem recovers a signal which is sparse in its gradient transform domain. As raw images generally assume that their gradients are sparse, TV-based approaches have been widely used in practical image reconstruction. The TV problem can be posed as

$$\min_{\boldsymbol{z}} \|\boldsymbol{z}\|_{TV} \quad \text{s.t.} \quad \boldsymbol{y} = \boldsymbol{A}\boldsymbol{z} \tag{3.26}$$

or in its regularized version

$$\min_{\boldsymbol{z}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{z}\|_2^2 + \lambda\|\boldsymbol{z}\|_{TV} \tag{3.27}$$

where the measurement matrix $\boldsymbol{A}$ describes the system model (e.g. Fourier transform or Radon transform), $\boldsymbol{z}$ is the signal or image to be recovered and $\boldsymbol{y}$ is the incomplete measurement data (Poon 2015). The TV semi-norm introduced by Rudin et al. (1992) in the context of image denoising can be defined for smooth functions as

$$\|\boldsymbol{z}\|_{TV} := \int_{\Omega} \sqrt{|\nabla \boldsymbol{z}|^2}\ \mathrm{d}x\mathrm{d}y = \int_{\Omega} \|\nabla \boldsymbol{z}\|_2\ \mathrm{d}x\mathrm{d}y \tag{3.28}$$

where ∇ indicates the gradient of $\boldsymbol{z}$ and $\Omega \subset \mathbb{R}^m$ is the m-dimensional signal domain.

In Chapters 4 and 5 we define the l_2-TV semi norm defined as

$$\|\boldsymbol{z}\|_{TV} := \sum_{j=1}^{m} \sum_{k \in \mathcal{N}_k} \omega_{j,k} \sqrt{(z_j - z_k)^2} \tag{3.29}$$

where $\mathcal{N}_j$ denotes the 8 nearest neighboring pixels of pixel j and $\omega_{j,k}$ are weights ($\omega_{j,k} = 1$ for axial neighbors and $\omega_{j,k} = 1/\sqrt{2}$ for diagonal neighbors). In this case we represent the image $\boldsymbol{z}(j)$ as a 2D matrix with pixel index (j). In Chapter 6 we defined an l_2-TV semi norm where the gradient is computed utilizing the finite difference approximation and taking the image $\boldsymbol{z}(i,j)$ as:

$$\|\boldsymbol{z}\|_{TV} := \sum_{i,j}^{m} \sqrt{|z_{i+1,j} - z_{i,j}|^2 + |z_{i,j+1} - z_{i,j}|^2} \tag{3.30}$$

The $\|\boldsymbol{z}\|_{TV}$ semi norm can be also written as an l_1-TV semi norm

$$\|\boldsymbol{z}\|_{TV} := \|\boldsymbol{z}\|_1 = \sum_{j=1}^{m} \sum_{k \in \mathcal{N}_k} \omega_{j,k} |z_j - z_k| \tag{3.31}$$

For this case one exploits the sparsity of the gradient when solving the optimization problem 3.27.

Therefore, there are two possible interpretations of the TV regularization. First, it can be seeing as a sparsity promoting norm in the gradient domain due to the l_1-norm. Secondly, it can be seeing as a regularizer that penalizes the oscillations in the output signal. By taking the gradient, one work in a domain where the values themselves are less important than their relationships with their neighbors. TV regularization measures how much the neighboring point or pixels differ from each other and forces the neighboring pixels to have similar values. TV-based

models have the advantage that the image edges are preserved, which is important for many imaging problems (Chambolle and Pock 2011). The problem 3.27 can be solved utilizing the first-order primal-dual algorithm for convex problems proposed by Chambolle and Pock (2011). The FISTA algorithm explained in Section 3.1.2 can be used specially when the l_1-TV semi norm is used as regularization. The Augmented Lagrange approaches can solve problem 3.27, for example the Alternating Direction Method of Multipliers (ADMM) (Hestenes 1969).

3.3 Dictionary Learning

The sparse recovery problem 3.7 assumes that the measurement matrix $\boldsymbol{D}$ is a fixed transform such as WT, DCT or Fourier Transform. In the Dictionary Learning (DL) approach the measurement matrix $\boldsymbol{D}$ is learned from a training dataset and adapted to the class of signal at hand. The adaptation process is called *Dictionary Learning* and can be posed as an optimization problem. Let us consider a set of P training signals $\boldsymbol{y}$ and build the matrix $\boldsymbol{Y} \in \mathbb{R}^{m \times P}$ whose columns are the training signals. DL solves the optimization problem

$$\begin{aligned} \min_{D,Z} \quad & \|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{Z}\|_2^2 \\ \text{s.t.} \quad & \|\boldsymbol{z}_\ell\|_0 \leq s, \quad \ell = 1:P \\ & \|\boldsymbol{d}_j\| = 1, \quad j = 1:n \end{aligned} \tag{3.32}$$

where the sparsity level s is given a priori, $\boldsymbol{D} \in \mathbb{R}^{m \times n}$ is the dictionary and $\boldsymbol{Z} \in \mathbb{R}^{n \times P}$ is a matrix containing the sparse vectors. The first constraint in 3.32 enforces each column of $\boldsymbol{Z}$ to contain at most s non-zero values. It sets the sparsity level of the representation to be the same for each signal. The second constraint normalizes the dictionary atoms to have unit norm. This constraint is inherited from orthogonal transforms. The normalization constraint aims to remove indetermination caused by a possible multiplicative factor that can multiply $\boldsymbol{D}$ and divide $\boldsymbol{Z}$ without changing the objective function.
In general, the dictionary learning method involves finding the dictionary $\boldsymbol{D}$ and the sparse representation $\boldsymbol{Z}$ such that $\boldsymbol{Y} \approx \boldsymbol{D}\boldsymbol{Z}$ is as good as possible (Dumitrescu and Irofti 2018). This problem is extremely difficult since it is non-convex and has a sparsity constraint which makes it NP-hard.

The normalization constraint opens the possibility for a sign flip in the dictionary atoms and the sparse representation. Moreover, the problem can be indeterminate due to the fact that a permutation of atoms can be combined with a permutation of representations to produce an identical objective function. Therefore, if $(\boldsymbol{D}, \boldsymbol{Z})$ is a solution of problem, 3.32 then $(\boldsymbol{D}\boldsymbol{P}, \boldsymbol{P}^{-1}\boldsymbol{Z}$ is also a solution, where $\boldsymbol{P}$ is a permutation matrix with nonzero elements equal ± 1. As a consequence, there

will be multiple local minima with the same value. Nevertheless, sign flipping and atom permutations do not hinder the optimization since identifying one of these minima is sufficient. The uniqueness of the solution is an issue in DL problem. One may wonder, under which conditions $\boldsymbol{D}$ and $\boldsymbol{Z}$ are the unique matrices whose product is $\boldsymbol{Y}$. Dumitrescu and Irofti (2018) summarized as follow: (i) If $|| y || 0 < spark^1(D)/2$ then the solution is the sparsest possible. As a result, once $\boldsymbol{D}$ is known, the matrix $\boldsymbol{Z}$ must be unique, due to the sparse nature of the support. (ii) There must be enough training signals in order to have information to retrieve a unique solution. Technically, there must be $s+1$ signals that are linear combinations of these atoms for each set of s atoms. That would be $P \geq (s+1)\begin{pmatrix} n \\ s \end{pmatrix}$ signals, which is practically impossible. However, reducing the number of signals to $2n(s+1)$ ensures that $\boldsymbol{D}$ is unique.

Another shortcoming of the DL problem is the multiple local minima. There is a distinct solution for each sparsity pattern. By considering only dictionaries with exactly s nonzero elements, it is clear that the DL problem, for each sparsity pattern, has at least one distinct local minimum (Dumitrescu and Irofti 2018).

3.3.1 Optimization algorithm in Dictionary Learning

Several approaches have been developed in the literature for the DL optimization problem. The straightforward and more successful approach is the alternate optimization. The optimization is performed by iteratively alternating between solving the sparse code keeping the dictionary fixed and updating the dictionary fixing the sparse representation variables. The strategy is also called block coordinate descent. Algorithm 2 shows how DL problem can be split into two optimization sub-problems (sparse coding and dictionary update) utilizing alternate optimization approach (Dumitrescu and Irofti 2018).

The most popular algorithms developed for dictionary learning are Method of Optimal Directions (MOD) and K-means Singular Value Decomposition (K-SVD). The MOD was introduced by Engan et al. (1999) in 1999. The MOD iteratively alternates between the sparse-code step and the dictionary updates. The sparse recovery step is performed for each signal using any standard sparse recovery technique presented in 3.1. The dictionary update step is analytically solved by computing $\boldsymbol{D} = \boldsymbol{Y}\boldsymbol{Z}^{-1}$ with $\boldsymbol{Z^{-1}}$ denoting the inverse. The MOD is an extremely effective algorithm that only requires a few iterations to converge. However, due to the complexities of matrix inversion, the procedure is quite difficult.

The K-SVD algorithm was developed by Aharon et al. (2006) in 2005. Similar to

[1]The spark of a dictionary $\boldsymbol{D}$ is the smallest number of columns that are linearly dependent.

Algorithm 2: DL by Alternate Optimization

Data: Training signals set $\boldsymbol{Y} \in \mathbb{R}^{m \times P}$;
Sparsity level s ;
Number of iterations T
Result: Trained dictionary $\boldsymbol{D}$; Sparse representation $\boldsymbol{Z}$
Initialization: Initial dictionary $\boldsymbol{D}_0$;
Initial sparse representation $\boldsymbol{Z}_0$;
for $t = 1, \ldots, T$ **do**
  Sparse Coding Update;
  Keeping $\boldsymbol{D}$ fixed, solve 3.32 to compute the sparse representations $\boldsymbol{Z}$
  Dictionary Update;
  Keeping the nonzero pattern fixed $\boldsymbol{Z}$, solve 3.32 to compute new dictionary
  Perform atoms normalization : $\boldsymbol{d}_j \leftarrow \frac{\boldsymbol{d}_j}{\|\boldsymbol{d}_j\|}, j = 1 : n$
end

MOD, the K-SVD algorithm performs alternate optimization, updating the sparse representation individually. The main contribution of K-SVD concerns the dictionary update step. Instead of utilizing the matrix inversion the dictionary update is performed atom by atom. The current atom and its related sparse coefficients are both updated at the same time, which provides even more acceleration. As a result, the method is both fast and efficient, and it is significantly less demanding than the MOD. For each atom the quadratic term in 3.32 is reformulated as

$$\left\| \boldsymbol{Y} - \sum_{j \neq k} \boldsymbol{d}_j \boldsymbol{z}_j^T - \boldsymbol{d}_k \boldsymbol{z}_k^T \right\|_F^2 = \left\| \boldsymbol{E}_k - \boldsymbol{d}_k \boldsymbol{z}_k^T \right\|_F^2 \tag{3.33}$$

where $\boldsymbol{z}_j^T$ are the rows of the sparse representation matrix, and $\boldsymbol{E}_k$ is the residual matrix. The atoms are updated by minimizing 3.33 with respect to $\boldsymbol{z}_j^T$ and $\boldsymbol{d}_k$ via a simple rank-1 approximation of $\boldsymbol{E}_k$. Updates are performed only for examples whose current representations use the atom $\boldsymbol{d}_k$. These methods are appropriate for image patches since they produce a non-structured dictionary (Rubinstein et al. 2010).

Ravishankar et al. (2017) also proposed a powerful approach for DL called Sum of Outer Products Dictionary Learning (SOUP-DIL). In order to approximate the training signals $\boldsymbol{Y}$, they use sparse rank-one matrices or outer products. In Chapter 4 we explain SOUP-DIL algorithm in detail. We implemented a multi-channel GPU version of the SOUP-DIL for multi-channel dictionary training.

3.4 Convolutional Dictionary Learning

The dictionary learning technique uses overlapping patches across the training signals. This approach leads to spatially redundant atoms that are essentially shifts of a basic atom type to "enforce" the expected spatial-invariance of the representation. Patch-domain methods suffer from memory limitation, especially when large dataset is used.

The convolutional dictionary learning approach, instead, replaces the non-structured dictionary $\boldsymbol{D}$ with a set of convolutional filters. In this method, one can utilize the entire image instead of small patches and learn filters and obtaining (sparse) representations directly from the original signals without storing many overlapping patches (Garcia-Cardona and Wohlberg 2018*a*).

In the CDL approach convolutional kernels are used to sparsely represent the signal Chun and Fessler (2017*a*, 2019*a*). The signal $\boldsymbol{y}$ can be synthesized by performing convolution of the filters and the sparse component:

$$\boldsymbol{y} = \sum_{k=1}^{K} \boldsymbol{d}_k \circledast \boldsymbol{z}_k \,. \tag{3.34}$$

where the signal-dimension vector $\boldsymbol{z}_k \in \mathbb{R}^m$ contains sparse signal features; and $\boldsymbol{d}_k \in \mathbb{R}^R$ (R represents filters dimensions) are filters regrouped in a dictionary $\boldsymbol{D} = \{\boldsymbol{d}_k\}$.

The mapping $\boldsymbol{S_D} : \{\boldsymbol{z}_k\} \mapsto \sum_{k=1}^{K} \boldsymbol{d}_k \circledast \boldsymbol{z}_k$ is called synthesis operator, since it synthesizes the signal $\boldsymbol{y}$ from the sparse vector $\boldsymbol{z}_k$.

Alternatively, the sparse vector $\boldsymbol{z}_k$ can be represented as a convolution of the signal $\boldsymbol{y}$ and the filters $\boldsymbol{d}_k$ Chun and Fessler (2019*a*), i.e.,

$$\mathbf{d}_k \circledast \boldsymbol{y} = \boldsymbol{z}_k, \quad \forall k = 1, ..., K \tag{3.35}$$

Hence, the mapping $\boldsymbol{A_D} : \boldsymbol{y} \mapsto \boldsymbol{d}_k \circledast \boldsymbol{y}$ is the analysis operator which coincides with the synthesis operator transpose, $\boldsymbol{A_D} = \boldsymbol{S}_{\boldsymbol{D}}^{\top}$.

A dataset of signals $\{\boldsymbol{y}_l \in \mathbb{R}^m : l = 1 \dots P\}$ is used to train the filters $\mathbf{d}_k$. The training cost synthesis function Γ_s and its analysis counterpart Γ_a are defined as follow (as defined in Chun and Fessler (2017*a*, 2019*a*)):

$$\Gamma_s(\boldsymbol{D}, \{\boldsymbol{z}_{k,l}\}) = \sum_{l=1}^{P} \frac{1}{2} \|\boldsymbol{y}_l - \sum_{k=1}^{K} \boldsymbol{d}_k \circledast \boldsymbol{z}_{k,l}\|_2^2 + \alpha \sum_{k=1}^{K} \|\boldsymbol{z}_{k,l}\|_r. \tag{3.36}$$

and,

$$\Gamma_a(\boldsymbol{D}, \{z_{k,l}\}) = \sum_{l=1}^{P} \sum_{k=1}^{K} \frac{1}{2} \|\boldsymbol{d}_k \circledast \boldsymbol{y}_l - \boldsymbol{z}_{k,l}\|_2^2 + \alpha \|z_{k,l}\|_r. \tag{3.37}$$

where $\boldsymbol{z}_{k,l}$ is the feature sparse vector associated to the training examples ($\boldsymbol{y}_l$) and the filter $\mathbf{d}_k$, $\|\cdot\|_r$ is a norm promoting sparsity (i.e., $r = 0, 1$) and $\alpha >> 0$ is a penalty weight controlling the sparsity of the features vector $\boldsymbol{z}_{k,l}$. Thus the training consists of finding a set of filters $\hat{\boldsymbol{D}} = \{\hat{\boldsymbol{d}}_k\}$ that "best" sparsify the set of training images such that (Chun and Fessler 2019*a*):

$$\hat{\boldsymbol{D}} = \arg\min_{\boldsymbol{D}} \min_{\{\boldsymbol{z}_{k,l}\}} \Gamma(\boldsymbol{D}, \{\boldsymbol{z}_{k,l}\}), \tag{3.38}$$

In Chun and Fessler (2019*a*) the constraint enforces the filter to satisfy the *tight-frame* condition to promote filter diversity: $C = \{\{\boldsymbol{d}_k\} : [\boldsymbol{d}_1, ...\boldsymbol{d}_K][\boldsymbol{d}_1, ...\boldsymbol{d}_K]^\top = \frac{1}{R}\mathbf{I}_K\}$ where $\mathbf{I}_K$ is the $K \times K$ identity matrix (see (Hines 2010)). The *tight-frame* condition forces the filters to be orthogonal, ensuring diversity.

3.4.1 Optimization Algorithms in CDL

As in DL optimization problem, the approach used in CDL alternates between the sparse code and the dictionary update. The most common method used are the Augmented Lagrangian approaches (Chun and Fessler 2017*a*, 2019*a*). The first application of Augmented Lagrangian methods in CDL was proposed by Bristow et al. (2013).

In Heide et al. (2015), Wohlberg (2015, 2016) a spatial domain ADMM framework was utilized to solve the CDL problem. These methods use alternate optimization between the sparse code and the dictionary (i.e., a two-block update), using augmented Lagrange (or ADMM) methods for each inner update (Chun and Fessler 2017*a*). The sparse coding step can be performed utilizing a suitable sparse recovery algorithm (e.g IHT) as presented 3.1, while the dictionary update can be addressed utilizing proximal gradient methods. For example, Chun and Fessler (2017*a*, 2019*a*) introduced a new optimization approach (BPEG-M) for solving block multi-nonconvex problems as the convolutional analysis and synthesis operator learning. The following section describe the (BPEG-M) algorithm applied to CAOL.

3.4.1.1 Block Proximal Extrapolated Gradient method using a Majorizer

The BPEG-M solve *the block multi-nonconvex* optimization problem:

$$\min F\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_B\right) := f\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_B\right) + \sum_{b=1}^{B} g_b\left(\boldsymbol{x}_b\right) \tag{3.39}$$

where $\boldsymbol{x}$ is decomposed into B blocks $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_B$ ($\{\boldsymbol{x}_b \in \mathbb{R}^{n_b} : b = 1, \ldots, B\}$). The function f is differentiable while the set of functions $\{g_b : b = 1, \ldots, B\}$ are not necessarily differentiable. For example, g_b could be a non-convex l_p quasi-norm, $0 \leqslant p < 1$). The constraint $\boldsymbol{x}_b \in \mathcal{X}_b$ can be incorporated to the function g_b by allowing them to be extended-valued. (Extended value means $g_b(\boldsymbol{x}_b) = +\infty$ if $\boldsymbol{x}_b \in dom(g_b)$, for $b = 1, \ldots, B$. In particular, g_b can be indicator functions of convex sets) (Chun and Fessler 2019*a*).
Chun and Fessler (2019*a*) prove in Section V that the CAOL model 3.38 satisfies the BPEG-M conditions. Thus it can be solved for two blocks, the $\boldsymbol{z}_{k,l}$-block and the $\boldsymbol{D}$-block. From equation 3.37 and 3.39 can be inferred that

$$f\left(\boldsymbol{z}, \boldsymbol{D}\right) = \sum_{l=1}^{P} \sum_{k=1}^{K} \frac{1}{2} \left\| \boldsymbol{d}_k \circledast \boldsymbol{y}_l - \boldsymbol{z}_{l,k} \right\|_2^2 \tag{3.40}$$

with

$$g_1\left(\boldsymbol{z}\right) = \alpha \sum_{l=1}^{P} \sum_{k=1}^{K} ||\boldsymbol{z}_{l,k}||_0 \tag{3.41}$$

and

$$g_2(\boldsymbol{D}) = \begin{cases} 0 & \boldsymbol{D}\boldsymbol{D}^T = \frac{1}{R} I_K \\ +\infty & \text{otherwise.} \end{cases} \tag{3.42}$$

The BPEG-M method employs an optimization transfer approach with a quadratic majorization matrix of the Hessian. A general definition of a Quadratic Majorization is explained in Lemma 4.2 in (Chun and Fessler 2019*a*):
Let $f : \mathbb{R}^n \to \mathbb{R}$. If ∇f is $\boldsymbol{M}$-Lipschitz (The definition of $\boldsymbol{M}$-Lipschitz continuity can be found in Chun and Fessler (2019*a*) Definition 4.1) continuous, then

$$f(x) \leq f(y) + \langle \nabla f(y), x - y \rangle + \frac{1}{2} \|x - y\|_{\boldsymbol{M}}^2, \quad \forall x, y \in \mathbb{R}^n. \tag{3.43}$$

If f satisfies 3.43, it is also satisfied with $\widetilde{\boldsymbol{M}}$ that $\widetilde{\boldsymbol{M}} > \boldsymbol{M}$ with $\widetilde{\boldsymbol{M}}$ being a diagonal matrix. A diagonal matrix provides an easy-to-minimize majorizer function. In the CAOL case f is quadratic, thus, $\widetilde{\boldsymbol{M}}$ is a majorizer of the Hessian matrix. Thus, BPEG-M solves 3.39 by minimizing a majorizer of F cyclically with respect to each block $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_B$ while fixing the remaining blocks at their previously updated

variables. For each block the iterations take the form:

$$\boldsymbol{x}_b^{(i+1)} = \arg\min_{\boldsymbol{x}_b} \frac{1}{2} \left\| \boldsymbol{x}_b - \left(\boldsymbol{x}_b^{(i)} - \left[\widetilde{\boldsymbol{M}}_b^{(i)} \right]^{-1} \nabla_{\boldsymbol{x}_b} f^{(i)} \boldsymbol{x}_b^{(i)} \right) \right\|_{\widetilde{\boldsymbol{M}}_b^{(i)}}^2 + g_b(\boldsymbol{x}_b) \tag{3.44}$$

which is a proximal gradient update similar to the IST algorithm. The BPEG-M block updates applied to CAOL performs the sparse code update, then the dictionary update.

Sparse code: $\boldsymbol{z}_{k,l}$-block

Given the current estimate of the dictionary $\boldsymbol{D}$, the optimization problem for the sparse code is written as follow:

$$\boldsymbol{z}_{l,k} = \arg\min_{\{\boldsymbol{z}\}} \sum_{l=1}^{P} \sum_{k=1}^{K} \frac{1}{2} \left\| \boldsymbol{d}_k \circledast \boldsymbol{y}_l - \boldsymbol{z} \right\|_2^2 + \alpha ||\boldsymbol{z}||_0 \tag{3.45}$$

This problem can be optimally solved utilizing the hard thresholding sparse recovery algorithm presented in 3.1.3.

$$\boldsymbol{z}_{l,k}^{(i+1)} = H_{\sqrt{2\alpha}} \left(\boldsymbol{d}_k \circledast \boldsymbol{y}_l \right) \tag{3.46}$$

Chun and Fessler (2019*a*) Section V.B shows how the Hard-Thresholding optimization is equivalent to apply the BPEG-M to problem 3.45.

Dictionary Update: $\boldsymbol{D}$-block

The filters update step consist of solving

$$\arg\min_{\{\boldsymbol{d}_k\}} \sum_{l=1}^{P} \sum_{k=1}^{K} \frac{1}{2} \left\| \boldsymbol{d}_k \circledast \boldsymbol{y}_l - \boldsymbol{z}_{l,k} \right\|_2^2 + \beta g(\boldsymbol{D}) \tag{3.47}$$

given the current estimate of the sparse component $\boldsymbol{z}_{l,k}$. Defining $\Psi_l \boldsymbol{d} = \boldsymbol{y}_l \circledast \boldsymbol{d} \quad \forall\, \boldsymbol{d}$ the filters update problem can be written as

$$\underset{\{\boldsymbol{d}_k\}}{\operatorname{argmin}} \sum_{k=1}^{K} \sum_{l=1}^{P} \frac{1}{2} \left\| \Psi_l \boldsymbol{d}_k - \boldsymbol{z}_{l,k} \right\|_2^2 + \beta g(\boldsymbol{D}) \tag{3.48}$$

Next step consists of designing the majorizer. One option is utilizing a diagonal majorization matrix $\widetilde{\boldsymbol{M}}_{\boldsymbol{D}} \in \mathbb{R}^{R \times R}$ that satisfies $\widetilde{\boldsymbol{M}}_{\boldsymbol{D}} \succeq \sum_{l=1}^{P} \Psi_l^T \Psi_l$

$$\widetilde{\boldsymbol{M}}_{\boldsymbol{D}} = \operatorname{diag} \left(\sum_{l=1}^{P} \left| \Psi_l^T \right| |\Psi_l| \, 1_R \right) \tag{3.49}$$

The majorization matrix in CAOL is pre-computed before optimization. However in the general case of BPEG-M the majorizer is updated at each iteration since it

depends on the sparse code.
After computing the majorization matrices one applies them in the proximal mapping problem to update the filters. The Proximal Mapping with Orthogonality Constraint (PMOC) is obtained by applying 3.44 to the optimization problem 3.48. Thus, the proximal mapping problem is written as

$$\boldsymbol{d}_k^{(i+1)} = \operatorname{argmin} \sum_{k=1}^{K} \frac{1}{2} \left\| \boldsymbol{d}_k - \left(\boldsymbol{d}_k^{(i)} - \left[\widetilde{\boldsymbol{M}}_{\boldsymbol{D}} \right]^{-1} \sum_{l=1}^{P} \Psi_l^T \left(\Psi_l \boldsymbol{d}_k^{(i)} - \boldsymbol{z}_{l,k} \right) \right) \right\|_{\widetilde{\boldsymbol{M}}_{\boldsymbol{D}}}^2 \tag{3.50}$$

Representing $\nu = \left(\boldsymbol{d}_k^i - \left[\widetilde{\boldsymbol{M}}_{\boldsymbol{D}} \right]^{-1} \sum_{l=1}^{P} \Psi_l^T \left(\Psi_l \boldsymbol{d}_k^{(i)} - \boldsymbol{z}_{l,k} \right) \right)$ the problem 3.50 can be re-written as

$$\begin{gathered} \left\{ \boldsymbol{d}_k^{(i+1)} \right\} = \underset{\{\boldsymbol{d}_k\}}{\operatorname{argmin}} \sum_{k=1}^{K} \frac{1}{2} \left\| \boldsymbol{d}_k - \nu_k^{(i)} \right\|_{\widetilde{\boldsymbol{M}}_{\boldsymbol{D}}}^2, \\ \text{subject to } \boldsymbol{D}\boldsymbol{D}^H = \frac{1}{R} \cdot \boldsymbol{I}, \end{gathered} \tag{3.51}$$

Proposition 5.4 in Chun and Fessler (2019*a*) considers the optimization problem

$$\min_{\boldsymbol{D}} \left\| \widetilde{\boldsymbol{M}}_{\boldsymbol{D}}^{1/2} \boldsymbol{D} - \widetilde{\boldsymbol{M}}_{\boldsymbol{D}}^{1/2} \mathcal{V} \right\|_{\mathrm{F}}^2, \quad \text{subj. to } \boldsymbol{D}\boldsymbol{D}^H = \frac{1}{R} \cdot \boldsymbol{I} \tag{3.52}$$

where $\mathcal{V} = \left[\nu_1^{(i+1)} \cdots \nu_K^{(i+1)} \right] \in \mathbb{R}^{R \times K}$, which can be solved using value decomposition of $\widetilde{\boldsymbol{M}}_{\boldsymbol{D}} \mathcal{V}$.

CHAPTER 4

Coupled Dictionary Learning Algorithm for Motion Estimation-Compensation in Cone-Beam CT

The present chapter proposes a new approach for motion-estimation compensation in CBCT. The main idea is to learn joint image-motion dictionaries in order to capture sliding motion at organs boundaries. An image dictionary and a set of DVF at different respiratory gates are learned jointly, thus, allowing the motion and the image to share structural similarities. The learned dictionaries are used within a MBIR algorithm to perform direct motion-estimation motion compensation. The preliminary results show the ability of the coupling dictionaries to capture structural similarities. The method performs well in terms of noise controlling. However, we have found many drawbacks to this methodology. The idea is at early research stage. This chapter related to the work presented in IEEE Nuclear Science Symposium and Medical Imaging Conference. Boston, USA. *Oral Presentation.*

4.1 Introduction

Due to the limited gantry rotation speed in acquiring the CBCT projection data, respiratory motion causes severe blurring artifacts, affecting the image quality of the reconstructed volume and the accuracy of dose planning and delivery (Zhi et al. 2019). Four-dimensional Cone Beam Computed Tomography (4D-CBCT) has been developed to address this issue, in which the acquired full-sampled projections are sorted into different respiratory phases. Thereafter the phase-resolved projections are reconstructed independently (Liu et al. 2015). The above technique is the so-called phase-correlated reconstruction technique which reconstructs the 3D image at each phase from gated data and concatenates the reconstructed images to obtain a Four-dimentional (4D) reconstruction. These reconstruction techniques include the respiration-correlated variants of the Feldkamp– Davis– Kress (FDK) (Feldkamp et al. 1984, Sonke et al. 2005) and simultaneous algebraic reconstruction (Andersen and Kak 1984) approaches (Mory et al. 2016). Nevertheless, the insufficient cone-beam projections per respiratory phase cause streak artifacts in the reconstructed

images due to the Nyquist-Shannon theorem (Rit, Wolthaus, van Herk and Sonke 2009). Several iterative reconstruction algorithms utilizing regularization can mitigate this shortcoming. The TV regularization is one of the most widespread regularization used to reconstruct sparse sample projection data. However, when the number of measurements is insufficient, the reconstruction frequently result in over-smoothing, especially for low-contrast regions (Wang and Gu 2013).
Alternative solutions are the motion-compensated reconstruction techniques where the motion is estimated by computing the DVF either from scout images or from the projection data. Earlier research works from (Li et al. 2007), (Rit, Wolthaus, van Herk and Sonke 2009), (Rit, Sarrut and Desbat 2009) and (Rit et al. 2011) use an a priori motion estimation from the Four-dimentional Computed Tomography (4D-CT) to back-project along curved trajectories. These approaches are highly dependent on the a priori estimation, meaning the motion-compensation is as good as the motion-estimation used as input. The motion estimation from the CBCT projection data, also known as joint motion-estimation and motion-compensated reconstruction methods, estimate the DVF from the gated Cone-Beam (CB) projections and perform a motion-compensated reconstruction. Examples of joint motion-estimation and motion-compensation methods are: the Simultaneous Motion Estimation and Image Reconstruction (SMEIR) algorithm introduced in (Wang and Gu 2013) and the Motion-Compensated 4D-CBCT approach proposed in (Brehm et al. 2011). Another approach, known as regularized 4D reconstruction techniques, reconstructs the entire cycle at once, using all of the projection data, and impose some similarities between subsequent frames by regularizing along time (Mory et al. 2016). These techniques include (Jia et al. 2010) and (Ritschl et al. 2012).
The majority of the motion-compensated 4D-CBCT reconstruction typically impose isotopic smoothing of the DVF, which can be inaccurate, at regions where different organs are in contact such as the lung-to-thoracic interface or lung-to-heart interface, where we observe a sliding motion between the organs. In the literature researchers have addressed this issue by zeroing motion regularization or adding a different motion constraint at boundaries between organs, but a segmentation of the organs is required prior to motion estimation (Dang et al. 2016). For example, (Werner et al. 2009) and (Wu et al. 2008) based motion estimation on the segmentation of areas that slide along each other. The work from (Schmidt-Richberg et al. 2012) introduces directional dependent regularization for the DVF estimation. They differentiate between the normal and tangential motion direction according to the boundary of the sliding regions. The normal-directed motion regularizer prevents overlaps, whereas the tangential regularization allows sliding motion. However,

segmentation of sliding organs is still needed (Delmon et al. 2013).
The present work proposes a novel algorithm for motion-estimation and motion-compensation in CBCT, based on image-motion dictionary coupling. Each atom of the dictionary represents a portion of the CBCT image and the associated motion. The hypothesis behind this approach is that the image and the motion can inform each other, thus not only allowing for noise reduction but also to learn features such as sliding motion at organ boundaries. We treat the motion vectors field as an image and learn dictionaries such that they can inform which motion takes place at which region of the body. For example, we would like the method to "learn" the sliding motion at lung boundaries.
The implementation of the proposed methodology concerns two stages: coupled dictionary learning and motion estimation-compensation. The first step consists of learning an image-motion coupled dictionary from a training dataset of images with a pre-estimated motion DVF dataset at different respiratory phases, using a modified SOUP-DIL algorithm (Ravishankar et al. 2017). In the second step, we utilize the learned dictionaries as an image-motion prior within a motion-compensated iterative reconstruction algorithm. This proposed methodology was validated using a training dataset generated from XCAT phantom (Kainz et al. 2019).

4.2 Direct Motion Compensation by Penalized-Likelihood

The model from (Zeng et al. 2005) was used to describe the image acquisition with gated motion. We assume that the measurement data are regrouped into L respiratory gates $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_L$, where for all $\ell = 1, \ldots, L$, $\boldsymbol{y}_\ell = [y_{1,\ell}, \ldots, y_{1,\ell}]^\top \in \mathbb{R}^n$ is the projection data (sinogram) corresponding to the ℓ-th respiratory gate. The discrete attenuation image to reconstruct takes the form of a vector $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_m]^\top \in \mathbb{R}^m$, where m is the number of voxels in the image. For all $j = 1, \ldots, m$, the coordinate of the j-th voxel is denoted $\boldsymbol{r}_j = [x_j, y_j, z_j]^\top \in \mathbb{R}^3$, and $\boldsymbol{G} = \{\boldsymbol{r}_j\}_{j=1}^m$ denotes the voxel grid. At each respiratory gate ℓ, $\boldsymbol{G}$ is deformed by a mapping $\varphi_\ell \colon \mathbb{R}^3 \to \mathbb{R}^3$. The discrete DVF is denoted $\boldsymbol{M}_\ell = \{\vec{\boldsymbol{m}}_{j,\ell}\}_{j=1}^m \in \mathbb{R}^{3\times m}$, with $\vec{\boldsymbol{m}}_{j,\ell} = \varphi_\ell(\boldsymbol{r}_j) - \boldsymbol{r}_j \in \mathbb{R}^3$ and $\varphi(\boldsymbol{r}_j)$ defined as:

$$\varphi(\boldsymbol{r}_j) = \boldsymbol{r}_j + \begin{bmatrix} \sum_{n=1}^{n_c} \alpha_n^X B\left(\frac{\boldsymbol{r}-\tilde{\boldsymbol{r}}_{\boldsymbol{n}}}{\sigma}\right) \\ \sum_{n=1}^{n_c} \alpha_n^Y B\left(\frac{\boldsymbol{r}-\tilde{\boldsymbol{r}}_{\boldsymbol{n}}}{\sigma}\right) \\ \sum_{n=1}^{n_c} \alpha_n^Z B\left(\frac{\boldsymbol{r}-\tilde{\boldsymbol{r}}_{\boldsymbol{n}}}{\sigma}\right) \end{bmatrix} \tag{4.1}$$

where n_c are the numbers of control points, B is the cubic B-spline function, σ is the distance between control points and $\boldsymbol{\alpha}^X = \left(\alpha_n^X\right)_{n=1}^{n_c}, \boldsymbol{\alpha}^Y = \left(\alpha_n^Y\right)_{n=1}^{n_c}$, $\boldsymbol{\alpha}^{\boldsymbol{Z}} = \left(\alpha_n^Z\right)_{n=1}^{n_c}$ are the motion B-spline coefficients along each axis X, Y and Z.

Using the cubic B-spline interpolation, the image deformation operator is a $m \times m$ square matrix entirely determined by $\boldsymbol{M}_\ell$, and is denoted $\boldsymbol{W}_\ell$ (warping operator). The respiratory-deformed attenuation image at gate ℓ is the matrix/vector product $\boldsymbol{W}_\ell \boldsymbol{\mu} \in \mathbb{R}^m$.
In a simplified setting (no electronic noise), each sinogram $\boldsymbol{y}_\ell$ is a random vector following a Poisson distribution with independent entries,

$$y_{i,\ell} \sim \text{Poisson}(\bar{y}_{i,\ell}(\boldsymbol{W}_\ell \boldsymbol{\mu}))\,, \tag{4.2}$$

with

$$\bar{y}_{i,\ell}(\boldsymbol{\mu}) = I_0 \exp(-[\boldsymbol{A}\boldsymbol{\mu}]_i) + \boldsymbol{s}_{i,\ell} \tag{4.3}$$

where $\boldsymbol{A}$ is a matrix modeling the CBCT system, $\boldsymbol{s}_{i,\ell}$ is a background term and I_0 is the blank scan.
Direct motion compensation is achieved by penalized maximum-likelihood joint estimation of the image $\boldsymbol{\mu}$ and the motion fields $\boldsymbol{\mathcal{M}}$ from the gated sinograms $\boldsymbol{y}_\ell$:

$$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\mathcal{M}}}) = \underset{\boldsymbol{\mu} \geq \boldsymbol{0}, \boldsymbol{\mathcal{M}}}{\arg\max}\, L(\boldsymbol{\mu}, \boldsymbol{\mathcal{M}}) - \beta R(\boldsymbol{\mu}, \boldsymbol{\mathcal{M}}) \tag{4.4}$$

where $\boldsymbol{\mathcal{M}} = \{\boldsymbol{M}_\ell\}_{\ell=1}^L$, and $R(\boldsymbol{\mu}, \boldsymbol{\mathcal{M}})$ is a noise-controlling penalty on the image and the motion, and the log-likelihood L is defined as:

$$L(\boldsymbol{\mu}, \boldsymbol{\mathcal{M}}) = \sum_{\ell=1}^{L} \sum_{i=1}^{n} y_{i,l} \log \bar{y}_{i,\ell}([\boldsymbol{W}_\ell \boldsymbol{\mu}]) - \bar{y}_{i,\ell}([\boldsymbol{W}_\ell \boldsymbol{\mu}])\,. \tag{4.5}$$

The maximization problem 4.4 can be solved with the Joint Reconstruction and Motion estimation (JRM) technique proposed in Bousse et al. (2016) for example.

4.2.1 Coupled-Dictionary Penalty

In this work, we used a penalty term $R(\boldsymbol{\mu}, \boldsymbol{\mathcal{M}})$ derived from coupled image-motion dictionary learning inspired from (Song et al. 2019):

$$\begin{aligned} R(\boldsymbol{\mu}, \boldsymbol{\mathcal{M}}) = \underset{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_P \in \mathbb{R}^d}{\arg\min} \sum_{p=1}^{P} \Bigg\{ & \|\boldsymbol{P}_p^{\text{im}}(\boldsymbol{\mu}) - \boldsymbol{D}^{\text{im}} \boldsymbol{z}_p\|_2^2 \\ & + \sum_{\ell=1}^{L} \|\boldsymbol{P}_p^{\text{mtn}}(\boldsymbol{M}_\ell) - \boldsymbol{D}_\ell^{\text{mtn}} \boldsymbol{z}_p\|_2^2 + \gamma \|\boldsymbol{z}_p\|_0 \Bigg\} \end{aligned} \tag{4.6}$$

where $\boldsymbol{D}^{\text{im}}$ (resp. $\boldsymbol{D}^{\text{mtn}}$) is a $\tilde{m} \times d$ (resp. $3\tilde{m} \times d$) image (resp. motion) dictionary matrix composed of d $\tilde{m}$-dimensional atoms and $\boldsymbol{P}_p^{\text{im}} \colon \mathbb{R}^m \to \mathbb{R}^{\tilde{m}}$ (resp $\boldsymbol{P}_p^{\text{mtn}} \colon \mathbb{R}^{3m} \to \mathbb{R}^{3\tilde{m}}$) is the p-th image (resp. motion) patch extractor. The penalty

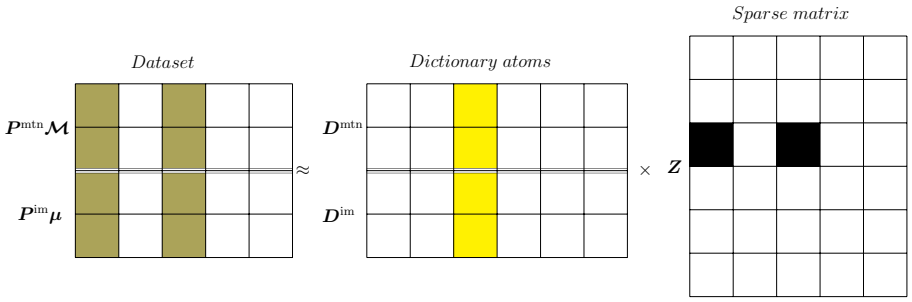


Fig. 4.1 Matrix representation of the coupled dictionary learning approach. $(\boldsymbol{P}^{\text{im}}\boldsymbol{\mu})$ and $(\boldsymbol{P}^{\text{mtn}}\boldsymbol{\mathcal{M}})$ represent the training examples, $\boldsymbol{D}^{\text{im}}$ and $\boldsymbol{D}^{\text{mtn}}$ the dictionaries and $\boldsymbol{Z}$ the common sparse matrix shared by the dictionaries. The sparse vector selects the same signal from $(\boldsymbol{P}^{\text{im}}\boldsymbol{\mu})$ and $(\boldsymbol{P}^{\text{mtn}}\boldsymbol{\mathcal{M}})$ to update the same atom in $\boldsymbol{D}^{\text{im}}$ and $\boldsymbol{D}^{\text{mtn}}$.

weight γ controls the overall sparsity. We treat each $\boldsymbol{M}_\ell$ as a 3-channel image containing the DVF in x, y, z coordinates. Therefore, if we train the dictionaries for 2 respiratory gates the total number of dictionaries would be 7.

At each respiratory gate, the image and motion dictionaries are trained to fit the training image-motion dataset $\{\boldsymbol{\mu}^k, \boldsymbol{\mathcal{M}}^k\}_{k=1}^K$, using a common sparse encoding $\boldsymbol{Z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_P] \in \mathbb{R}^{d \times P}$ which is used to encode the image and the motion simultaneously. This can be formulated as a constrained optimization problem:

$$\begin{aligned} \min_{\boldsymbol{D}^{\text{im}}, \boldsymbol{D}^{\text{mtn}}, \{\boldsymbol{Z}^k\}} \sum_{k=1}^{K} \Big\{ & \|\boldsymbol{P}^{\text{im}}(\boldsymbol{\mu}^k) - \boldsymbol{D}^{\text{im}}\boldsymbol{Z}^k\|_2^2 \\ & + \|\boldsymbol{P}^{\text{mtn}}(\boldsymbol{\mathcal{M}}^k) - \boldsymbol{D}^{\text{mtn}}\boldsymbol{Z}^k\|_2^2 \Big\} \\ \text{s.t} \quad \|\boldsymbol{Z}^k\|_0 \leq s \quad \forall k \quad & \text{and} \quad \|\boldsymbol{d}_q^{\text{im}}\|_2 = \|\boldsymbol{d}_q^{\text{mtn}}\|_2 = 1 \quad \forall q \end{aligned} \tag{4.7}$$

where $\boldsymbol{P}^{\text{im}}\colon\ \colon \mathbb{R}^m \to \mathbb{R}^{\tilde{m} \times P}$ (resp. $\boldsymbol{P}^{\text{mtn}}\colon\ \colon \mathbb{R}^{3m} \to \mathbb{R}^{3\tilde{m} \times P}$) is the image (resp. motion) patch extractor, $\boldsymbol{d}_q^{\text{im}}$ (resp. $\boldsymbol{d}_q^{\text{mtn}}$) is the q-th column of $\boldsymbol{D}^{\text{im}}$ (resp. $\boldsymbol{D}^{\text{mtn}}$) and s denotes the maximum sparsity level (number of non-zeros in $\boldsymbol{Z}^k$). We solved this problem with a modified version of the SOUP-DIL algorithm (Ravishankar et al. 2017).

The training task consists of finding a common sparse component $\boldsymbol{Z}$ that is shared by the image and motion dictionaries $\boldsymbol{D}^{\text{im}}, \boldsymbol{D}^{\text{mtn}}$. We enforce the dictionaries to "fit the image and motion datasets simultaneously". Hence, if one atom of $\boldsymbol{D}^{\text{im}}$ is a linear combination of the first and third signal of $\boldsymbol{P}^{\text{im}}\boldsymbol{\mu}^k$, then the same atom of $\boldsymbol{D}^{\text{mtn}}$ will be a linear combination of the first and third signal of $\boldsymbol{P}^{\text{mtn}}\boldsymbol{\mathcal{M}}^k$ as well. Figure 4.1 shows a representation of the aforementioned hypothesis.

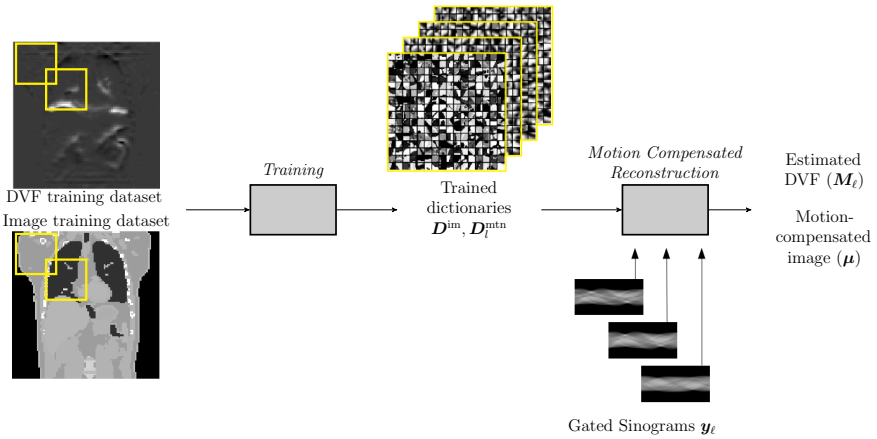


Fig. 4.2 Diagram of coupled dictionary learning algorithm for motion estimation-compensation consisting of the dictionary learning training and the motion-estimation and motion-compensation module.

4.2.1.1 Methodology

The approach proposed in this work is divided into two stages: *a*) the *training* consisting of estimating the image and motion dictionaries and *b*) the *motion compensated reconstruction* where we utilize the trained dictionaries to estimate the DVF and perform motion compensated reconstruction. Figure 4.2 shows the diagram describing coupled dictionary learning algorithm for motion estimation/compensation.

For the *training* step, we extended the SOUP-DIL algorithm to multi-channel dictionary learning. We stake-up the dictionaries and the training dataset as if we had only one dictionary and only one dataset:

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{D}^{\text{im}} \\ \boldsymbol{D}_{\ell}^{\text{mtn}} \end{bmatrix} \tag{4.8}$$

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{M}_{\ell} \end{bmatrix} \tag{4.9}$$

The SOUP-DIL replaces the sparsity constraint in 4.7 with an l_0 penalty term and introduces $\boldsymbol{C} = \boldsymbol{Z}^H \in \mathbb{R}^{P \times d}$ where $(\cdot)^H$ is the Hermitian (conjugate transpose). Then the product $\boldsymbol{DZ}$ can be written as a Sum of OUter Products ($\boldsymbol{DZ} = \boldsymbol{DC}^H = \sum_{q=1}^{d} \boldsymbol{d}_q \boldsymbol{c}_q^H$) where $\boldsymbol{c}_q$ is the q column of $\boldsymbol{C}$. Thus, the optimization problem 4.7 is written as:

$$\min_{d_q, c_q} \|\boldsymbol{P}(\boldsymbol{Y}) - \sum_{q=1}^{d} d_q c_q^H\|_2^2 + \lambda^2 \sum_{q=1}^{d} \|c_q\|_0; \tag{4.10}$$
$$\text{s.t} \quad \|c_q\|_\infty \leqslant \Omega; \quad \|\boldsymbol{d}_q^{\text{im}}\|_2 = \|\boldsymbol{d}_q^{\text{mtn}}\|_2 = 1 \quad \forall q$$

where $\lambda > 0$ is a weight to control the sparsity level and $\boldsymbol{P}$ is the patch-extraction operator. The constraint $\|c_q\|_\infty \leqslant \Omega$ with $\Omega > 0$ (e.g. $\Omega = \|\boldsymbol{P}\boldsymbol{Y}\|_2$) makes the objective function invariant to (arbitrarily) large scaling of c_q (i.e., non-coercive objective). See section II in Ravishankar et al. (2017). We constrain the columns of $\boldsymbol{D}^{\text{im}}$ and $\boldsymbol{D}^{\text{mtn}}$ to have unit norm individually since they have different physical units.

For each dictionary atom q, we solve equation 4.10 using the block coordinate descent algorithm. We first update the sparse vector c_q keeping the dictionaries fixed. We refer to this step as the *sparse coding update*. Then, we update the dictionaries keeping the sparse matrix constant. We refer to this step as the *dictionaries update*. Given $\boldsymbol{E}_q \triangleq \boldsymbol{P}\boldsymbol{Y} - \sum_{t \neq q} \mathrm{d}_t \boldsymbol{c}_t^H$ the *sparse coding update* is achieved with truncated hard-thresholding operation (Ravishankar et al. 2017):

$$\hat{\boldsymbol{c}}_q = \min\left(\left|H_\lambda\left(\boldsymbol{E}_q^H \boldsymbol{d}_q\right)\right|, \Omega \mathbf{1}_N\right) \odot e^{\angle \boldsymbol{E}_q^H \boldsymbol{d}_q} \tag{4.11}$$

with

$$(H_\lambda(x)) = \begin{cases} 0 & |x| < \lambda \\ x & |x| \geqslant \lambda \end{cases} \tag{4.12}$$

and $\mathbf{1}_P$ is the vector of ones of length P. "$\odot$" denotes the element-wise multiplication. The supplementary material in Ravishankar et al. (2017) provides a detailed explanation on how to solve 4.11.

The *dictionaries update* consist of finding $\boldsymbol{d}_q$ from 4.10. SOUP-DIL applies the global minimizer:

$$\hat{\boldsymbol{d}}_q = \begin{cases} \frac{\boldsymbol{E}_q \boldsymbol{c}_q}{\|\boldsymbol{E}_q \boldsymbol{c}_q\|_2}, & \text{if } \boldsymbol{c}_q \neq 0 \\ \boldsymbol{v}, & \text{if } \boldsymbol{c}_q = 0 \end{cases} \tag{4.13}$$

where $\boldsymbol{v}$ can contain random or unit values. (See Section III in Ravishankar et al. (2017))

Following the estimation of the dictionaries, we perform motion-compensated reconstruction by iteratively alternating between (i) updating the common sparse vector $\boldsymbol{Z}$ using the OMP algorithm (Rubinstein et al. 2008), and (ii) updating the image $\boldsymbol{\mu}$ and the motion fields $\boldsymbol{\mathcal{M}}$ with a L-BFGS algorithm. The 4D-CBCT reconstruction is achieved by warping the reconstructed images at the reference gate utilizing the estimated DVF. The pseudo-code for motion compensation

Algorithm 3: Motion Compensated Reconstruction algorithm

Input: Pre-trained dictionaries ($\boldsymbol{D}^{\text{im}}, \boldsymbol{D}_\ell^{\text{mtn}}$), initial DVF ($\boldsymbol{\mathcal{M}}^0$), initial sparse matrix ($\boldsymbol{Z}^0$), initial image ($\boldsymbol{\mu}^0$), penalty weight (β), gated sinograms ($\boldsymbol{y}_\ell$), forward operator ($\boldsymbol{A}$), patch-extraction operators ($\boldsymbol{P}^{\text{mtn}}$,$\boldsymbol{P}^{\text{im}}$), error threshold for OMP (χ) ;

$\boldsymbol{D} = [\boldsymbol{D}^{\text{im}}; \boldsymbol{D}_\ell^{\text{mtn}}]$;

#outer iterations N_{outer}.

Output: DVF estimation, ($\hat{\boldsymbol{\mathcal{M}}}$), Motion-compensated reconstructed image ($\hat{\boldsymbol{\mu}}$), Sparse matrix ,($\hat{\boldsymbol{Z}}$)

for $t = 1, \ldots, N_{\text{outer}} - 1$ **do**

Sparse code update

$\boldsymbol{Z}^t \leftarrow \text{OMP}(\boldsymbol{Z}^{t-1}, \boldsymbol{\mathcal{M}}^{t-1}, \boldsymbol{\mu}^{t-1}, \boldsymbol{D}, \boldsymbol{P}^{\text{mtn}}, \boldsymbol{P}^{\text{im}}, \chi)$

Motion field update (For each respiratory phase)

$\boldsymbol{M}_\ell^t \leftarrow \text{L-BFGS}(\boldsymbol{M}_\ell^{t-1}, \boldsymbol{Z}^t, \boldsymbol{\mu}^{t-1}, \boldsymbol{D}_\ell^{\text{mtn}}, \boldsymbol{y}_\ell, \boldsymbol{P}^{\text{mtn}}, \boldsymbol{A}, \beta)$

Image update

$\boldsymbol{\mu}^t \leftarrow \text{L-BFGS}(\boldsymbol{\mu}^{t-1}, \boldsymbol{\mathcal{M}}^t, \boldsymbol{Z}^t, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_L, \boldsymbol{D}^{\text{im}}, \boldsymbol{P}^{\text{im}}, \boldsymbol{A}, \beta)$

end

$\hat{\boldsymbol{\mathcal{M}}} \leftarrow \{\boldsymbol{M}_\ell^{t=N_{\text{outer}}}\}_{\ell=1}^L$;

$\hat{\boldsymbol{\mu}} \leftarrow \boldsymbol{\mu}^{t=N_{\text{outer}}}$;

reconstruction is summarized in Algorithm 3.

4.2.2 Non-Coupled Dictionary Penalty

We also investigate the use of a non-coupled dictionary penalty term in which the sparse vectors are not updated jointly:

$$R(\boldsymbol{\mu}, \boldsymbol{\mathcal{M}}) = \underset{\substack{\boldsymbol{z}_1^{im},\ldots,\boldsymbol{z}_P^{im} \in \mathbb{R}^d \\ \boldsymbol{z}_1^{mtn},\ldots,\boldsymbol{z}_P^{mtn} \in \mathbb{R}^d}}{\arg\min} \sum_{p=1}^{P} \Bigg\{ \|\boldsymbol{P}_p^{\text{im}}(\boldsymbol{\mu}) - \boldsymbol{D}^{\text{im}} \boldsymbol{z}_p^{im}\|_2^2 + \sum_{\ell=1}^{L} \|\boldsymbol{P}_p^{\text{mtn}}(\boldsymbol{M}_\ell) - \boldsymbol{D}_\ell^{\text{mtn}} \boldsymbol{z}_{p,\ell}^{mtn}\|_2^2 + \kappa \|\boldsymbol{z}_p^{im}\|_0 + \varepsilon \|\boldsymbol{z}_p^{mtn}\|_0 \Bigg\} \tag{4.14}$$

where κ and ε are penalty weight controlling the overall sparsity.

The dictionaries are trained separately, the motion and image dictionaries are not sharing information. Thus, there is a different sparse matrix for the motion and the image dictionaries. The motion dictionary is obtained by solving:

$$\begin{aligned} &\min_{\boldsymbol{D}^{\text{im}}, \{(\boldsymbol{Z}^{im})^k\}} \sum_{k=1}^{K} \left\{ \|\boldsymbol{P}^{\text{im}}(\boldsymbol{\mu}^k) - \boldsymbol{D}^{\text{im}} (\boldsymbol{Z}^{im})^k\|_2^2 \right\} \\ &\text{s.t} \quad \|(\boldsymbol{Z}^{im})^k\|_0 \leq s \quad \forall k \quad \text{and} \quad \|\boldsymbol{d}_q^{\text{im}}\|_2 = 1 \quad \forall q \end{aligned} \tag{4.15}$$

while the image dictionary is obtained by solving:

$$\min_{\boldsymbol{D}_\ell^{\text{mtn}},\{(\boldsymbol{Z}_\ell^{mtn})^k\}} \sum_{k=1}^{K} \left\{ \|\boldsymbol{P}^{\text{mtn}}(\boldsymbol{\mu}^k) - \boldsymbol{D}_\ell^{\text{mtn}}(\boldsymbol{Z}_\ell^{mtn})^k\|_2^2 \right\}$$
$$\text{s.t} \quad \|(\boldsymbol{Z}_\ell^{mtn})^k\|_0 \leq s \quad \forall k \quad \text{and} \quad \|\boldsymbol{d}_{q,\ell}^{\text{mtn}}\|_2 = 1 \quad \forall q \tag{4.16}$$

As for the coupled dictionary learning algorithm, we perform motion-compensated reconstruction by iteratively alternating between (i) updating the sparse vectors $\boldsymbol{Z}^{im}$ and $\boldsymbol{Z}^{mtn}$ separately, using the OMP algorithm (Rubinstein et al. 2008), and (ii) updating the image $\boldsymbol{\mu}$ and the motion fields $\boldsymbol{\mathcal{M}}$ with a L-BFGS algorithm. During the reconstruction, the image and the motion penalty terms do not share information.

4.2.3 Algorithms used for Comparison

We compare the methodology proposed in this work against a motion-estimation motion-compensation approach utilizing the EP regularizer:

$$R(\boldsymbol{\mu}) = \sum_{j=1}^{m} \sum_{t \in \mathcal{N}_j} \omega_{j,t} \sqrt{(\boldsymbol{\mu}_j - \boldsymbol{\mu}_t)^2 + \varpi} \tag{4.17}$$

where $\mathcal{N}_j$ denotes the 8 nearest neighboring pixels of pixel j and $\omega_{j,t}$ are weights ($\omega_{j,t} = 1$ for axial neighbors and $\omega_{j,t} = 1/\sqrt{2}$ for diagonal neighbors), and $\varpi > 0$ is a small real value to ensure differentiability. The objective function takes the form:

$$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\mathcal{M}}}) = \arg\max_{\boldsymbol{\mu} \geq \boldsymbol{0}, \boldsymbol{\mathcal{M}}} L(\boldsymbol{\mu}, \boldsymbol{\mathcal{M}}) - \beta R(\boldsymbol{\mu}) \tag{4.18}$$

We used the L-BFGS solver to estimate $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\mathcal{M}}}$.

4.3 Experiments

The training data consists of a collection of 3-mm pixel-width $90 \times 90 \times 90$ torso axial images generated from the XCAT phantom. Patient size, organs size, and the maximum extension of the diaphragm were modified to assure diversity in the dataset. The image dataset corresponds to the phantom at the reference gate. For each phantom at the reference gate, we obtained their corresponding DVF at each respiratory gate with a standard deformable registration (Bousse et al. 2016). The DVF correspond to the displacement of each pixel from the reference gate to the pointed gate in x, y, z coordinates. For example, if the reference gate is 5%-inhalation and the pointed gate is 50%-inhalation, the DVF is the pixel displacements between the two respiratory phases. We utilized 10 3D images as

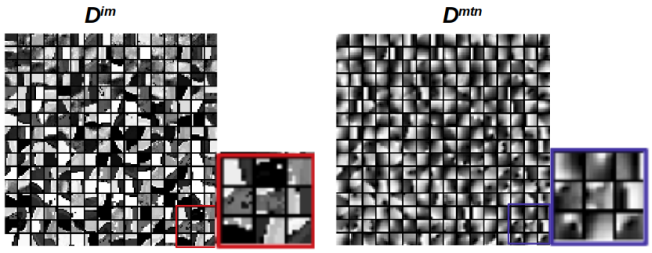


Fig. 4.3 Trained coupled dictionaries from the image dataset ($\boldsymbol{D}^{\text{im}}$) and the DVFs dataset ($\boldsymbol{D}^{\text{mtn}}$) along x-axis.

training examples. The dictionaries were trained on $8 \times 8 \times 8$-pixel patches and contained 7680 atoms. We trained a total of 7 dictionaries corresponding to the image dictionary $\boldsymbol{D}^{\text{im}}$; the motion dictionaries at gate $\ell = 1$ in the x, y, z axes $(\boldsymbol{D}^{\text{mtn}})^{X}_{\ell=1}; (\boldsymbol{D}^{\text{mtn}})^{Y}_{\ell=1}; (\boldsymbol{D}^{\text{mtn}})^{Z}_{\ell=1}$ and the motion dictionaries at gate $\ell = 2$ in the x, y, z axes $(\boldsymbol{D}^{\text{mtn}})^{X}_{\ell=2}; (\boldsymbol{D}^{\text{mtn}})^{Y}_{\ell=2}; (\boldsymbol{D}^{\text{mtn}})^{Z}_{\ell=2}$. The gated CB projection data was generated by forward projection of 3-mm pixel-width $90 \times 90 \times 90$ torso axial images generated from the XCAT phantom at each respiratory phase. We modeled the projector $\boldsymbol{A}$ with a cone-beam CT system utilizing the Astra Toolbox (van Aarle et al. 2016, van Aarle et al. 2015, Palenstijn et al. 2011). For each sinogram, we use a monochromatic source with 10^3 incident photons and 100 background events. We initialized the image using the Maximum-likelihood reconstruction for transmission tomography (MLTR) algorithm (Nuyts et al. 1998), which maximizes the likelihood function without regularization. The motion vectors were initialized with zeros. The implementation of the OMP and the modified SOUP-DIL algorithms were GPU-accelerated and directly callable from Matlab (Release 2018).

4.4 Results on XCAT phantom

In this section we report results with 3D images. We compare the performance of MEC-MDL (Section 4.2.1) with MEC-SDL (Section 4.2.2). We also compare with EP regularizer (Section 4.2.1) applied in the image update.

4.4.1 Training

Figure 4.3 shows the image and motion dictionaries at the end-of-inhalation respiratory phase. Each atom is represented in the figure as an 8×8 square patch. Most of them exhibit structural similarities but different values, as $\boldsymbol{D}^{\text{im}}$ represents linear attenuation coefficient values and $\boldsymbol{D}^{\text{mtn}}$ represents motion amplitude. This confirms that the coupled dictionaries are able to capture similarities between image and DVFs. Figure 4.4 shows the dictionaries trained using an independent sparse vector for each of them. The atoms are not showing structural similarities

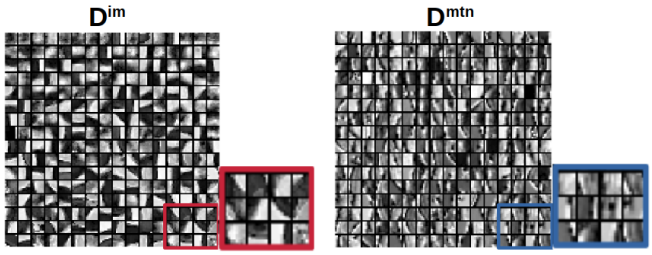


Fig. 4.4 Trained dictionaries from the image dataset ($\boldsymbol{D}^{\text{im}}$) and the DVFs dataset ($\boldsymbol{D}^{\text{mtn}}$) along x-axis using a different sparse vectors for each dictionary.

since they are not constrained to have the same number of supports in the sparse vector.

4.4.2 Motion Estimation-Compensation

Figure 4.5 shows a coronal view of the motion-compensated images utilizing MEC-MDL, MEC-SDL and the EP regularizer. It also includes the no-motion compensated image and the ground truth. The no-motion-compensated image presents noise artifacts and motion artifacts around moving areas. A visual analysis confirms that the MEC-SDL and MEC-MDL algorithms suppress noise and remove motion artifacts correctly. However, the images are still blurry, making it difficult to distinguish between organ features. We selected the ribs-to-lung contact area as Regions of Interest (ROI) to evaluate the impact of the sliding motion correction. A profile along the $x - axis$ on the dashed line (Figure 4.6) shows that Linear Attenuation Coefficient (LAC) values estimated using MEC-MDL are notably biased compare with the ground truth. The MEC-MDL method scores lower performance compare with MEC-SDL and the reconstruction with EP regularizer. The MEC-MDL method and the reconstruction utilizing EP regularizer achieves similar performance, and their LAC values are close to the ground truth.
Figure 4.7 shows a sagittal view of the reconstructed images at the end-of-inhalation respiratory phase. The ROI around the spherical lesion shows a small modification in the tumor shape for the MEC-MDL method. The tumor shape and position in the MEC-SDL image results closer to the ground truth in comparison with EP regularizer.
We quantitatively evaluated the performance of the reconstruction methods by computing the RMSE and the PSNR in the selected ROI. Table 4.1 shows the values of the PSNR and RMSE computed using $m = 60$ pixels in the selected ROI.

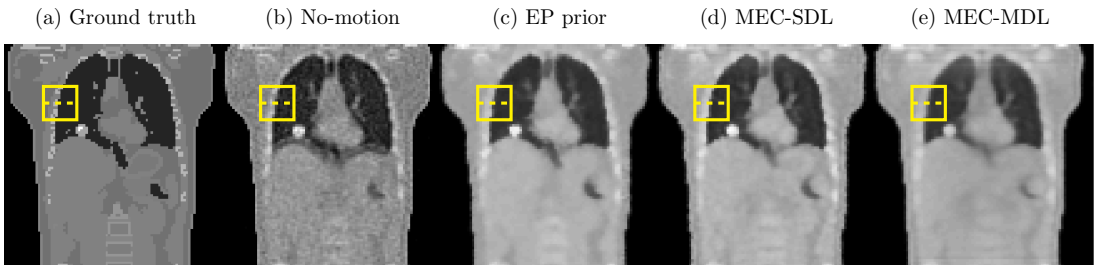


Fig. 4.5 Coronal view of the: a) Ground truth image; b) No motion compensated image (no prior); c) Reconstructed image utilizing the EP prior ; d) Reconstructed image utilizing the MEC-SDL method; e) Reconstructed image utilizing the MEC-MDL method.

The PSNR was computed as:

$$\text{PSNR(dB)} = 10 \cdot \log_{10} \left(\frac{\left\{ \max(\hat{\mu}^{GT}) \right\}^2}{\sum_{j=1}^{m} \frac{1}{m} \left(\hat{\mu}_j - \hat{\mu}_j^{GT} \right)^2} \right) \tag{4.19}$$

and the RMSE as:

$$RMSE = \sqrt{\frac{1}{m} \sum_{j=1}^{m} \left(\hat{\mu}_j^{GT} - \hat{\mu}_j \right)^2} \tag{4.20}$$

The MEC-MDL images scores lower PSNR and higher RMSE than the MEC-SDL images and the image reconstructed utilizing the EP prior. The MEC-SDL shows better performance compare with the image reconstructed utilizing the EP prior. It scores higher PSNR and lower RMSE. The gain in PSNR was 1.01%. (Gain(%) $= 100 \cdot$ (MEC-SDL $-$ EP)/EP).
Overall, the reconstructed images correct motion artifacts and noise acceptably. However, the images are blurred which made difficult to evaluate the sliding artifacts correction. The MEC-MDL methods show lower performance compare with MEC-SDL method and the EP regularizer.

4.5 Discussion and Conclusion

We proposed a method for direct motion-compensated CBCT reconstruction by penalized maximum likelihood using a coupled (image-motion) dictionary learning regularization term. The image to reconstruct and the motion fields image utilize the same encoding in order to capture structural similarities between the image and the DVF. The coupled and single dictionary learning algorithms perform well in terms of noise controlling in the reconstructed image and both estimate the motion correctly. For the coupled dictionary learning algorithm, the dictionaries exhibit structural similarities which confirms that they are able to capture similarities

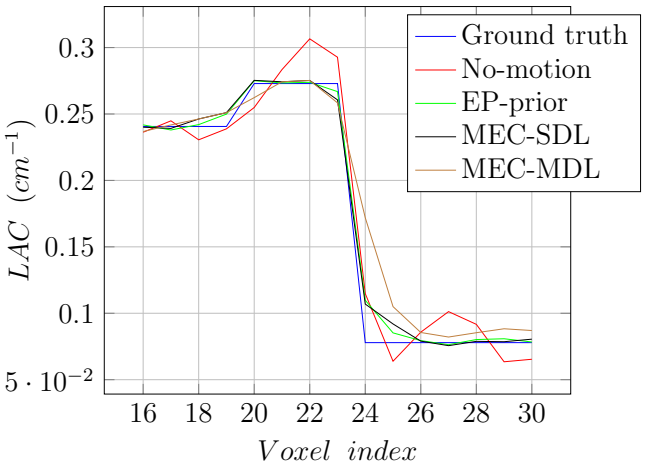


Fig. 4.6 Reconstructed image profile along the $x-axis$ on the dashed line showed in figure 4.5.

(a) Ground truth (b) EP prior (c) MEC-SDL (d) MEC-MDL

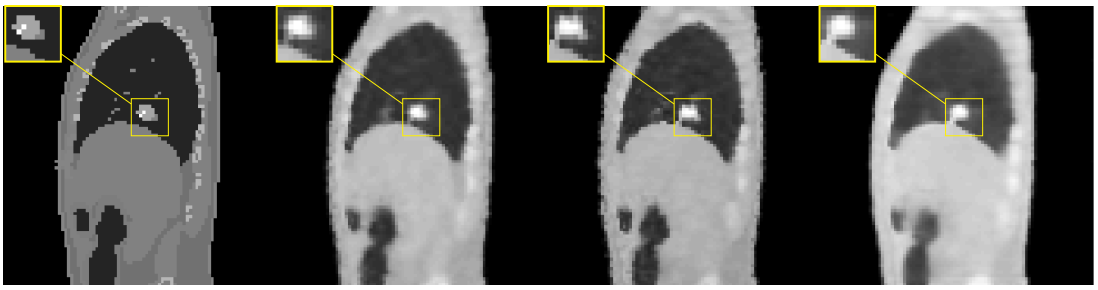


Fig. 4.7 Sagittal view of the: a) Ground truth image; b) Reconstructed image utilizing the EP prior ; c) Reconstructed image utilizing the MEC-SDL method; d) Reconstructed image utilizing the MEC-MDL method.

Method	PSNR(dB)	RMSE
EP-prior	39.3	0.0123
MEC-SDL	39.7	0.0102
MEC-MDL	38.1	0.0127

Table 4.1 Peak Signal-to-Noise Ratio (PSNR) in dB and the Root Mean Square Error (RMSE) for the reconstructed image utilizing EP regularizer, MEC-SDL method and MEC-MDL method.

between image and DVFs. However, the reconstructed image is still blurred. The single dictionary learning algorithm performs better in terms of noise controlling with an improvement in comparison with the EP regularizer.

The authors consider that the MEC-MDL algorithm can potentially work and perform better than existing state-of-the-art methods for sliding artifact correction. Further improvements need to be performed to achieve such accomplishment.

It is critical to ensure that the ground truth DVF estimation accounts for the sliding between organs boundaries. In the present work, we use the B-spline interpolation to estimate the DVF, which can be sub-optimal to account for sliding

motion along organs boundary. Furthermore, it is widely known that in motion-compensation techniques, B-spline interpolation over-smooths the reconstructed images. The authors suggest using the demons registration (Thirion 1998) to better account for the sliding artifacts and avoid over-smoothness.

We generate the CB projection data utilizing few projection angles and extremely low counts, making the image reconstruction task even more ill-posed. The authors suggest increasing the number of projection angles and perform less challenging experiments. Thus, It will be possible to evaluate the effectiveness of the MEC-MDL for the sliding artifacts correction task.

Moreover, the MEC-MDL reconstruction could be enhanced by fine-tuning the regularization parameters during the training and the reconstruction. Other factors that must be considered are the following: the number of atoms in the dictionary, the sparsity level, and the number of patches or training examples. However, the time required to train the dictionaries and reconstruct the images makes it quite challenging to tuning the parameters. The authors suggest the optimization of the algorithm to enhanced execution speed.

For the coupled image-motion dictionary learning, since the DVF and the attenuation images have different values (motion amplitude and attenuation coefficients respectively), constraining both datasets to have the same sparse coefficients could be a strong constraint, a less constraining model is to constrains only the supports (locations of zeros and non-zeros) of each sparse vector to be identical but their values could be different.

CHAPTER 5

Multi-channel Convolutional Analysis Operator Learning for Dual-Energy CT Reconstruction

Summary

The present Chapter proposes the multi-channel convolutional analysis operator learning MCAOL method for DECT to exploit common spatial features within attenuation images at different energies. It proposes an optimization algorithm which jointly reconstructs the attenuation images at low and high energies with a mixed norm regularization on the sparse features. The convolutional filters are pre-trained through the MCAOL algorithm and used within an MBIR, where the unknown images are reconstructed simultaneously by solving one combined optimization problem. As of the authors knowledge, this is the first time MCAOL is applied to DECT image reconstruction and we reported increased reconstruction accuracy compared to CAOL and iterative methods with single and joint total-variation JTV regularization. This work has been published in the peer-reviewed journal Physics in Medicine and Biology.

5.1 Introduction

The dual-source acquisition technique in DECT requires two helical scans at two different tube voltages; therefore, two sets of projection data at different energy levels are collected and further reconstructed. However, as the number of incident photons increases when irradiating with two sources the same anatomical region, the radiation dose increases proportionally (Sajja et al. 2020). A reduction in radiation exposure can be achieved by decreasing the number of projection angles. However, aliasing artifacts can appear in the reconstructed images if the number of projection angles does not follow the Nyquist sampling theorem. Moreover, it is more challenging to achieve high-resolution, high-contrast image reconstruction due to the low Signal-to-Noise Ratio (SNR) (Zhang et al. 2020).

In the literature, most of the development on low-dose CT reconstruction has focused on single image. Among the main techniques, MBIR methods are the

most popular. These techniques exploit models of the imaging system's physics (forward models) along with statistical models of the measurements and noise and often simple object priors. They iteratively optimize model-based cost functions to estimate the underlying unknown image (Elbakri and Fessler 2002). Typically, such cost functions consist of a data-fidelity term, e.g., least squares or NLL, capturing the imaging forward model and the measurement/noise statistical model and a regularizer term promoting smoothness, low-rank or sparsity (Kim et al. 2014). The Total Variation (Sidky et al. 2006, Sidky and Pan 2008) has been proposed to solve incomplete projection data reconstruction problems and achieved good performance. However, TV reconstruction results in undesired patchy effects. Data-driven and learning-based approaches have gained much interest in recent years for biomedical image reconstruction. These methods learn representations of images and are used in combination with MBIR techniques to perform complex mappings between limited or corrupted measurements and high-quality images. Among those algorithms, data-driven sparse transforms such as DL (Xu et al. 2012) use a training dataset of high-resolution and denoised images to learn features, in an unsupervised manner, that can be used to reconstruct new images. These features take the form of "atoms", which are regrouped into dictionaries and are used to sparsely represent the image (Aharon et al. 2006). DL-based image reconstruction integrates the learned atoms with the raw scanner data within a regularized MBIR context (Ravishankar et al. 2017, Zheng et al. 2018). Other closely related methods include sparsifying transform learning (Ravishankar and Bresler 2012) and the connection between data-adaptive models and convolutional deep learning algorithms (Ravishankar et al. 2019) with an increase interest in methods that leverage both learning-based and MBIR tools.

However, most DL methods are patch-based, and the learned features often contain shifted versions of the same features. The resulting learned dictionary may be over-redundant and therefore are memory demanding, which makes it difficult to utilize in 3D multi-modal imaging. To address these problems, CDL techniques utilize shift-invariant filters, providing a convenient and memory-efficient alternative to conventional DL techniques (Chun and Fessler 2017*b*). CDL approaches can be combined with MBIR by providing unsupervised prior knowledge of the target image. The CDL approach can also be formulated from an analysis point of view (Chun and Fessler 2019*b*) (sparse convolution) and is known as CAOL. Despite the rapidly expanding research, the application of CDL to multi-channel images has received little attention (Degraux et al. 2017, Garcia-Cardona and Wohlberg 2018*b*). Image reconstruction from DECT sparse-views or low-dose requires algorithms more advanced than the standard approach where attenuation at each measured

energy is reconstructed independently. Notable models in the literature designed to promote structural similarity of images are JTV (Ehrhardt et al. 2014*a*), spectral patch-based penalty for the maximum-likelihood method (Kim et al. 2015), tensor-based and coupled dictionary learning (Wu et al. 2018, Song et al. 2019), parallel level sets (Kazantsev et al. 2018) and the prior rank, intensity and sparsity model (PRISM) (Yang et al. 2017).

We extend the CAOL approach to multi-channel settings and we develop a MCAOL framework that can exploit direct joint reconstruction, given the low-dose DECT measurements, where all the unknown images are reconstructed simultaneously by solving one combined optimization problem. As of the author knowledge, this is the first time that MCAOL is applied to DECT image reconstruction and we demonstrate its superiority with respect to CAOL. Furthermore, MCAOL requires considerably less memory compared to alternative DL approaches. The joint reconstruction approach is developed for a low-dose data acquisition protocol which consists of collecting data using a sparse angular sampling, using a different X-ray energy in consecutive steps and low X-ray photon counts.

In DECT, a reasonable prior assumption is that attenuation images at different energies can be expected to be *structurally* similar in the sense that an edge (e.g., an organ boundary) that is present at one energy, is likely to be at same location and alignment with the other energies as well, even though the contrast between materials will be different at each energy. MCAOL technique reconstructs attenuation images from the projection data combined with multi-channel filters trained on a dataset of reconstructed images. The central idea of MCAOL is to learn unsupervised DECT multi-channel convolutional dictionaries that can provide a joint sparse representation of the underlined images by jointly learning filters for the different energies: each atom not only carries individual information for each energy individually but also inter-energy information. By reconstructing DECT images using MBIR techniques in conjunction with MCAOL, the multi-energy information can be optimally used by allowing the images to "talk to each other" during the reconstruction process through the learned joint dictionaries, reducing noise while preserving image resolution. In order to deal with the extreme low-dose scenario, we model the Poisson and we solve the image optimization problem by using approximated quasi-Newton method with constrained memory to achieve accurate joint reconstruction with limited computational complexity.

5.2 Learning Convolutional Regularizers for Image Reconstruction: CAOL

In this Section we review the foundation of CAOL for MBIR.
MBIR is achieved by solving an optimization problem of the form

$$\min_{\boldsymbol{\mu}\in\mathbb{R}^m} L(\boldsymbol{\mu}, \boldsymbol{y}) + \beta R(\boldsymbol{\mu}) \tag{5.1}$$

where $\boldsymbol{\mu} \in \mathbb{R}^m$ is the 2D or 3D image to reconstruct, $\boldsymbol{y} \in \mathbb{R}^n$ is the observed measurement, L is a data-fidelity term that incorporates the measurement model –generally taking the form of a NLL function– and R is a regularizer weighted by $\beta > 0$; n and m are respectfully the dimension of the measurement (number of detectors) and dimension of the image (number of pixels). The minimization is carried out with the help of iterative algorithms such as modified expectation-maximization (EM) for emission tomography (ET) (De Pierro 1995) or PWLS combined with separable paraboloidal surrogate (SPS) for CT (Elbakri and Fessler 2002).

The regularizer R is designed such that the reconstructed image $\hat{\boldsymbol{\mu}}(\boldsymbol{y})$ has desired properties, such as smoothness and sparsity of the gradient. It can be also trained so that $\hat{\boldsymbol{\mu}}(\boldsymbol{y})$ can be sparsely represented as a linear combination of basic elements, or atoms, regrouped in a dictionary.
We consider the CAOL approach (Chun and Fessler 2019*b*) where the image is sparsely represented with convolutional kernels (filters). In the analysis model, the image is represented with "sparsifying" filters $\boldsymbol{d}_k \in \mathbb{R}^R$ by the analysis operator $\mathcal{A}_{\boldsymbol{D}} : \boldsymbol{\mu} \mapsto \{\boldsymbol{d}_k \circledast \boldsymbol{\mu}\}$, such that

$$\boldsymbol{d}_k \circledast \boldsymbol{\mu} = \boldsymbol{z}_k, \quad \forall k = 1, \ldots, K. \tag{5.2}$$

where $\boldsymbol{z}_k \in \mathbb{R}^m$ is a sparse feature image vector of the same dimension as the image $\boldsymbol{\mu}$, and "$\circledast$" denotes the 2D convolution operator. The filters $\boldsymbol{d}_k \in \mathbb{R}^R$ are vectorized images of dimension $R \ll m$ that are regrouped in a dictionary $\boldsymbol{D} = \{\boldsymbol{d}_k\} \in \mathbb{R}^{R\times K}$.
Learning the dictionary $\boldsymbol{D}$ from a dataset of training images $\{\boldsymbol{\mu}_l \in \mathbb{R}^m : l = 1, \ldots, P\}$ corresponds to finding a collection of filters $\boldsymbol{D}^\star = \{\boldsymbol{d}_k^\star\}$ obtained by the following non-convex optimization problem

$$\boldsymbol{D}^\star = \arg\min_{\boldsymbol{D}\in C} \min_{\{\boldsymbol{z}_{l,k}\}} F_{\mathrm{a}}(\boldsymbol{D}, \{\boldsymbol{z}_{l,k}\}) \tag{5.3}$$

with the training analysis objective function F_{a} defined as

$$F_{\mathrm{a}}(\boldsymbol{D}, \{\boldsymbol{z}_{l,k}\}) = \sum_{l=1}^{P} \sum_{k=1}^{K} \frac{1}{2} \left\| \boldsymbol{d}_k \circledast \boldsymbol{\mu}_l - \boldsymbol{z}_{l,k} \right\|_2^2 + \alpha \left\| \boldsymbol{z}_{l,k} \right\|_0 \tag{5.4}$$

where $\boldsymbol{z}_{l,k} \in \mathbb{R}^m$ is the feature image associated to the training image $\boldsymbol{\mu}_l$ and the filter $\boldsymbol{d}_k$, $\|\cdot\|_0$ is the sparsity-promoting l_0 semi-norm defined for all $\boldsymbol{z} = [z_1, \ldots, z_m]^\top \in \mathbb{R}^m$ as

$$\|\boldsymbol{z}\|_0 = \sum_{j=1}^{m} \mathbf{1}_{[0,+\infty]}(|z_j|) \tag{5.5}$$

where $1_A : \mathbb{R} \to \{0, 1\}$ denotes the indicator function of a set $A \subset \mathbb{R}$, which is defined as $1_A(\xi) = 1$ if $\xi \in A$ and $1_A(\xi) = 0$ if $\xi \notin A$, and $\alpha > 0$ is a weight balancing between accuracy and sparsity and C is the constrain on $\boldsymbol{D} = \{\boldsymbol{d}_k\}$. In Chun and Fessler (2019*b*) the filters are enforced to satisfy the *tight-frame* conditions, i.e.,

$$C = \left\{ \{\boldsymbol{d}_k\} : [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_K][\boldsymbol{d}_1, \ldots, \boldsymbol{d}_K]^\top = \frac{1}{R} \boldsymbol{I}_K \right\} \tag{5.6}$$

where $\boldsymbol{I}_K$ is the $K \times K$ identity matrix, to promote filters diversity. The entire optimization problem 5.3 is solved by the BPEG-M utilizing two blocks as described in Section 3.4.1.1. The minimization in $\boldsymbol{D}$ is achieved with a PMOC algorithm which can be implemented using the CONVolutional Operator Learning Toolbox (CONVOLT) (Chun and Fessler 2019*b*, Chun 2019). The minimization in $\boldsymbol{z}$ is achieved with a hard-thresholding operator $\mathcal{T} \colon \mathbb{R}^m \times \mathbb{R}_+^* \to \mathbb{R}^m$ defined at each row j as:

$$[\mathcal{T}(\boldsymbol{a}, \beta)]_j = \begin{cases} a_j & \text{if } \frac{1}{2} a_j^2 \geq \beta \\ 0 & \text{otherwise} \end{cases} \tag{5.7}$$

for all $\boldsymbol{a} = [a_1, \ldots, a_m]^\top \in \mathbb{R}^m$ and for all $\beta > 0$, which provides a global minimizer for $\boldsymbol{z} \mapsto \frac{1}{2}\|\boldsymbol{a} - \boldsymbol{z}\|_2^2 + \beta \|\boldsymbol{z}\|_0$, in such a way that

$$\mathcal{T}\left(d_k \circledast x_l, \alpha\right) = \arg\min_{z_{l,k}} \frac{1}{2} \left\| d_k \circledast x_l - z_{l,k} \right\|_2^2 + \alpha \left\| z_{l,k} \right\|_0 \tag{5.8}$$

Finally the regularizer R in the minimization problem 5.1 is derived from the learned filters $\boldsymbol{D}^\star$ as

$$R(\boldsymbol{\mu}) = \min_{\{\boldsymbol{z}_k\}} \sum_{k=1}^{K} \frac{1}{2} \left\| \boldsymbol{d}_k^\star \circledast \boldsymbol{\mu} - \boldsymbol{z}_k \right\|_2^2 + \alpha \left\| \boldsymbol{z}_k \right\|_0 . \tag{5.9}$$

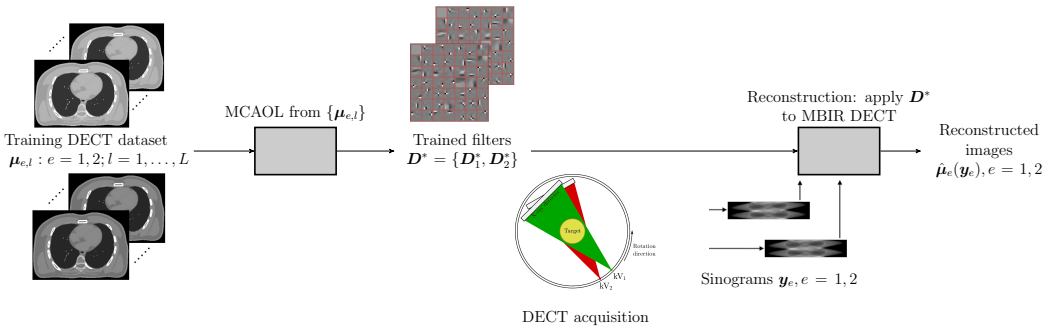


Fig. 5.1 Diagram of MCAOL consisting of the unsupervised filter learning phase and the model-based iterative DECT reconstruction module.

5.3 Multi-channel Convolutional Analysis Operator Learning

MBIR can be generalized to multi-channel imaging. Assuming we wish to reconstruct two images $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{R}^m$ of the same "object" from two independent measurements $\boldsymbol{y}_1 \in \mathbb{R}^{n_1}$ and $\boldsymbol{y}_2 \in \mathbb{R}^{n_2}$ corresponding to two modalities, multi-channel MBIR can be achieved by using an iterative algorithm to solve

$$\min_{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{R}^m} \rho_1 L_1(\boldsymbol{\mu}_1, \boldsymbol{y}_1) + \rho_2 L_2(\boldsymbol{\mu}_2, \boldsymbol{y}_2) + R_{\text{mc}}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \tag{5.10}$$

where L_1 and L_2 are the data-fidelity terms for $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, R_{mc} is a multi-channel regularizer and $\rho_1, \rho_2 > 0$ are weights. R_{mc} is designed to exploit the inference between the 2 channels $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, for example to promote structural similarities as proposed in (Ehrhardt et al. 2014*a*).

MCAOL is a generalization of CAOL where the training is performed jointly on a set of images obtained from imaging modalities as depicted in figure 5.1 for DECT. Let $\{(\boldsymbol{\mu}_{1,l}, \boldsymbol{\mu}_{2,l}) \in \mathbb{R}^m \times \mathbb{R}^m : l = 1, \ldots, P\}$ be a training dataset consisting of P pairs of images.

MCAOL learns the sparsifying filter pairs

$$(\boldsymbol{d}_{1,k}, \boldsymbol{d}_{2,k}) \in \mathbb{R}^R \times \mathbb{R}^R : k = 1, \ldots, K \tag{5.11}$$

together with the extracted feature pairs

$$(\boldsymbol{z}_{1,l,k}, \boldsymbol{z}_{2,l,k}) \in \mathbb{R}^m \times \mathbb{R}^m : k = 1, \ldots, K,\ l = 1, \ldots, P\,. \tag{5.12}$$

MCAOL is achieved by solving the following optimization problem, given the training image set $(\boldsymbol{\mu}_{1,l}, \boldsymbol{\mu}_{2,l})$

$$(\boldsymbol{D}_1^\star, \boldsymbol{D}_2^\star) = \underset{\boldsymbol{D}_1, \boldsymbol{D}_2, \in C}{\arg\min} \underset{\substack{\{\boldsymbol{z}_{1,l,k}\} \\ \{\boldsymbol{z}_{2,l,k}\}}}{\min} F_{\text{mc}}\left(\boldsymbol{D}_1, \boldsymbol{D}_2, \{\boldsymbol{z}_{1,l,k}\}, \{\boldsymbol{z}_{2,l,k}\}\right) \tag{5.13}$$

$$F_{\mathrm{mc}}(\boldsymbol{D}_1, \boldsymbol{D}_2, \{\boldsymbol{z}_{1,l,k}\}, \{\boldsymbol{z}_{2,l,k}\}) = \sum_{l=1}^{P} \sum_{k=1}^{K} \frac{\gamma_1}{2} \left\| \boldsymbol{d}_{1,k} \circledast \boldsymbol{\mu}_{1,l} - \boldsymbol{z}_{1,l,k} \right\|_2^2 + \frac{\gamma_2}{2} \left\| \boldsymbol{d}_{2,k} \circledast \boldsymbol{\mu}_{2,l} - \boldsymbol{z}_{2,l,k} \right\|^2 + \left\| (\boldsymbol{z}_{1,l,k}, \boldsymbol{z}_{2,l,k}) \right\|_{1,0} \tag{5.14}$$

where $\gamma_1, \gamma_2 > 0$ are weights and the semi-norm $\| \cdot \|_{1,0}$ on $\mathbb{R}^m \times \mathbb{R}^m$ is defined for all $\boldsymbol{z}_1 = [z_{1,1}, \ldots, z_{1,m}]^\top \in \mathbb{R}^m$ and for all $\boldsymbol{z}_2 = [z_{2,1}, \ldots, z_{2,m}]^\top \in \mathbb{R}^m$ as

$$\| (\boldsymbol{z}_1, \boldsymbol{z}_2) \|_{1,0} = \sum_{j=1}^{m} \mathbf{1}_{[0,+\infty]} \left(|z_{1,j}| + |z_{2,j}| \right) \tag{5.15}$$

$\| \cdot \|_{1,0}$ denotes the $l_{1,0}$ norm. It promotes joint sparsity, i.e., with zero and non-zero values at the same locations, of image features in all the modalities, that are encoded by the multi-channel dictionary $\boldsymbol{D}_1, \boldsymbol{D}_2$.

To solve (5.13) we utilize the BPEG-M algorithm (Chun and Fessler 2019*b*, Chun 2019) with 3 blocks: 1) the block which updates the sparse codes jointly $(\boldsymbol{z}_{1,l,k}, \boldsymbol{z}_{2,l,k})$; 2) the block for the first dictionary $(\boldsymbol{D}_1,)$; and 3) the block for the second dictionary $(\boldsymbol{D}_2)$. The 2 dictionary blocks are updated utilizing PMOC algorithm Chun and Fessler (2019*b*), Chun (2019) while for the update of the sparse codes we deploy a multi-channel hard-thresholding operator $\mathcal{T}_{\mathrm{mc}} \colon \mathbb{R}^m \times \mathbb{R}^m \times (\mathbb{R}_+^*)^2 \to \mathbb{R}^m \times \mathbb{R}^m$ defined at each row j as

$$[\mathcal{T}_{\mathrm{mc}}(\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{\gamma})]_j = \begin{cases} (a_{1,j}, a_{2,j}) & \text{if } \frac{1}{2}\gamma_1 a_{1,j}^2 + \frac{1}{2}\gamma_2 a_{2,j}^2 \geq 1 \\ (0,0) & \text{otherwise} \end{cases} \tag{5.16}$$

for all $\boldsymbol{a}_1 = [a_{1,1}, \ldots, a_{1,m}]^\top \in \mathbb{R}^m$, $\boldsymbol{a}_2 = [a_{2,1}, \ldots, a_{2,J}]^\top \in \mathbb{R}^m$ and for all $\boldsymbol{\gamma} = (\gamma_1, \gamma_2) \in (\mathbb{R}_+^*)^2$, which provides a global minimizer for $(\boldsymbol{z}_1, \boldsymbol{z}_2) \mapsto \frac{\gamma_1}{2}\|\boldsymbol{a}_1 - \boldsymbol{z}_1\|_2^2 + \frac{\gamma_2}{2}\|\boldsymbol{a}_2 - \boldsymbol{z}_2\|_2^2 + \|\boldsymbol{z}_1, \boldsymbol{z}_2\|_{1,0}$ (Xu et al. 2011, Section 3), in such a way that

$$\mathcal{T}_{\mathrm{mc}} \left(\boldsymbol{d}_{1,k} \circledast \boldsymbol{x}_{1,l}, \boldsymbol{d}_{2,k} \circledast \boldsymbol{x}_{2,l}, \boldsymbol{\gamma} \right) = \arg\min_{\boldsymbol{z}_{1,l,k}, \boldsymbol{z}_{2,l,k}} \left\{ \frac{\gamma_1}{2} \left\| \boldsymbol{d}_{1,k} \circledast \boldsymbol{x}_{1,l} - \boldsymbol{z}_{1,l,k} \right\|_2^2 + \frac{\gamma_2}{2} \left\| \boldsymbol{d}_{2,k} \circledast \boldsymbol{x}_{2,l} - \boldsymbol{z}_{2,l,k} \right\|^2 + \left\| (\boldsymbol{z}_{1,l,k}, \boldsymbol{z}_{2,l,k}) \right\|_{1,0} \right\} \tag{5.17}$$

Finally the regularizer R_{mc} in the minimization problem 5.10 is derived from

Algorithm 4: MCAOL Training Algorithm

Input: DE Training Dataset $\boldsymbol{\mu}_{e,l}$, $l = 1, \ldots, P$, $e = 1, 2$, joint sparsity weights $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$, #outer iterations N_{outer}

Output: Learned filters $(\boldsymbol{D}_1^\star, \boldsymbol{D}_2^\star)$

$(\boldsymbol{D}_1^0, \boldsymbol{D}_2^0) \leftarrow$ Normalized random initialization ;

for $t = 0, \ldots, N_{\text{outer}} - 1$ **do**

  Update sparse codes (in parallel) ;

  for *k,l=1,1,. . . ,K,P* **do**

    $(\boldsymbol{z}_{1,l,k}^{t+1}, \boldsymbol{z}_{2,l,k}^{t+1}) \leftarrow \mathcal{T}_{\text{mc}}(\boldsymbol{d}_{1,k}^{t+1} \circledast \boldsymbol{\mu}_{1,l}, \boldsymbol{d}_{2,k}^{t+1} \circledast \boldsymbol{\mu}_{2,l}, \boldsymbol{\gamma})$;

  end

  Update Filters ;

  $\boldsymbol{D}_1^{t+1} \leftarrow \text{PMOC}(\boldsymbol{\mu}_{1,l}, \boldsymbol{z}_{1,k}^{t+1})$;

  $\boldsymbol{D}_2^{t+1} \leftarrow \text{PMOC}(\boldsymbol{\mu}_{2,l}, \boldsymbol{z}_{2,k}^{t+1})$;

end

$\boldsymbol{D}_1^\star \leftarrow \boldsymbol{D}_1^{N_{\text{outer}}}$;

$\boldsymbol{D}_2^\star \leftarrow \boldsymbol{D}_2^{N_{\text{outer}}}$;

the learned filters $(\boldsymbol{D}_1^\star, \boldsymbol{D}_2^\star)$ as

$$R_{\text{mc}}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = \min_{\substack{\{\boldsymbol{z}_{1,k}\} \\ \{\boldsymbol{z}_{2,k}\}}} \sum_{k=1}^{K} \frac{\gamma_1}{2} \left\| \boldsymbol{d}_{1,k}^\star \circledast \boldsymbol{\mu}_1 - \boldsymbol{z}_{1,k} \right\|_2^2 + \frac{\gamma_2}{2} \left\| \boldsymbol{d}_{2,k}^\star \circledast \boldsymbol{\mu}_2 - \boldsymbol{z}_{2,k} \right\|_2^2 + \left\| (\boldsymbol{z}_{1,k}, \boldsymbol{z}_{2,k}) \right\|_{1,0} \tag{5.18}$$

The pseudo-code for MCAOL training procedure is summarized in Algorithm 4.

5.4 Dual-Energy CT Reconstruction with Multi-Channel CAOL

5.4.1 X-ray CT Discrete Model

In this section, we describe the CT discrete physical measurement process with the spectrum of the X-ray source beams composed of two different energies. We consider the case of 2D slice-by-slice imaging systems. For image reconstruction we assume that the continuous attenuation image $\mu_e(\boldsymbol{r})$ which denotes the linear attenuation coefficient at position $\boldsymbol{r} \in \mathbb{R}^2$ and the energy level $e = 1, 2$, can be represented by a linear combination of basis functions $\{b_j\}$ associated to a discrete sampling on a $\sqrt{m} \times \sqrt{m}$ Cartesian grid,

$$\mu_e(\boldsymbol{r}) = \sum_{j=1}^{m} \mu_{e,j} b_j(\boldsymbol{r}) \,, \tag{5.19}$$

where $\mu_{e,j} > 0$ for all $j = 1, \ldots, m$ and all $e = 1, 2$. The line integral becomes a summation:

$$\int_{\mathbb{R}} \mu_e(\boldsymbol{\nu}_{i(l)})\, dl = \sum_{j=1}^{m} \mu_{e,j} \int_{\mathbb{R}} b_j(\boldsymbol{\nu}_i(l))\, dl = \sum_{j=1}^{m} a_{i,j}\mu_{e,j} \tag{5.20}$$

where $\boldsymbol{\nu}_i(l) = \boldsymbol{s}_i + l\vec{\boldsymbol{\epsilon}}_i \in \mathbb{R}^2$ is a parametrization of the i-th ray emitted from the source $\boldsymbol{s}_i$ with direction $\vec{\boldsymbol{\epsilon}}_i$, $a_{i,j} \triangleq \int_{\mathbb{R}} b_j(\boldsymbol{\nu}_i(l))\, dl$ is the contribution of the j-th pixel to the i-th ray. The system matrix $\boldsymbol{A}$ is constructed as an under-determined matrix of dimensions $n \times m$ where $n = N_{\mathrm{d}} \times N_\theta$ with N_{d} and N_θ being respectively the number of detectors and N_θ and the number of angles (projections), and is defined as $[\boldsymbol{A}]_{i,j} = a_{i,j}$, $\forall\, i = 1, \ldots, n,\ \forall\, j = 1, \ldots, m$. The spectral X-ray mathematical discrete model is based on the Beer's law which provides the X-ray intensity after transmission. The expected number of detected photons $\bar{y}_{i,e}$ is then redefined as a function of the discrete image $\boldsymbol{\mu}_e$ as

$$\bar{y}_{i,e}(\boldsymbol{\mu}_e) = S_e \mathrm{e}^{-[\boldsymbol{A}\boldsymbol{\mu}_e]_i} + \eta_{e,i} \tag{5.21}$$

where $\boldsymbol{\mu}_e = [\mu_{e,1}, \ldots, \mu_{e,m}]^\top \in \mathbb{R}^m$ is the vector of attenuation coefficients at source energy e, S_e is the mean photons flux at the e-th energy bin, as we assume a mono-energetic intensity, and $\eta_{e,i} \in \mathbb{R}^+$ is a known additive term representing the expected number of background events (primarily from scatter). In the case of normal exposure, the number of detected photons follows a Poisson distribution, i.e.,

$$y_{i,e} \sim \mathrm{Poisson}(\bar{y}_{i,e}(\boldsymbol{\mu}_e)) \tag{5.22}$$

and the measurements at each energy bin $e = 1, 2$ are stored in a vector $\boldsymbol{y}_e = [y_{e,1}, \ldots, y_{e,N_{\mathrm{d}} \cdot N_\theta}]^\top$.

Although monochromatic X-ray source does not usually hold for scanners in clinical practice, a common effective strategy consists of applying a polychromatic-to-monochromatic source correction pre-processing step (Whiting et al. 2006), and in the rest of the paper we will therefore assume that we have a monoenergetic source or that it has already been appropriately corrected.

5.4.2 Low-Dose CT Reconstruction

In case of low X-ray dose, since the photons counts can be very limited, the Gaussian approximation is no longer applicable as the logarithm of the data cannot be computed. We therefore chose to perform sparse view CT reconstruction from the raw measurements $(\boldsymbol{y}_1, \boldsymbol{y}_2)$ by solving the minimization problem 5.10, with positivity constraints on $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$, using the Poisson NLL functions L_1 and L_2 defined as

$$- L_e(\boldsymbol{\mu}_e, \boldsymbol{y}_e) = \sum_{i=1}^{n} y_{e,i} \log \bar{y}_{i,e}(\boldsymbol{\mu}_e) - \bar{y}_{i,e}(\boldsymbol{\mu}_e), \quad e = 1, 2 \tag{5.23}$$

and the trained regularizer R_{mc} derived from the learned filters $(\boldsymbol{D}_1^\star, \boldsymbol{D}_2^\star)$ as in 5.18.

Therefore, substituting 5.23 and 5.18 into the minimization 5.10, we obtain the following explicit expression for the MCAOL DECT reconstruction problem:

$$\begin{aligned}(\boldsymbol{\mu}_1^\star, \boldsymbol{\mu}_2^\star) = \underset{\boldsymbol{\mu}_e \geq 0}{\arg\min} \sum_{e=1}^{2} \rho_e \underbrace{\sum_{i=1}^{n} y_{e,i} \log \bar{y}_{i,e}(\boldsymbol{\mu}_e) - \bar{y}_{i,e}(\boldsymbol{\mu}_e)}_{L_e(\boldsymbol{\mu}_e, \boldsymbol{y}_e)} \\ + \underbrace{\min_{\substack{\{\boldsymbol{z}_{1,k}\} \\ \{\boldsymbol{z}_{2,k}\}}} \sum_{k=1}^{K} \left\{ \sum_{e=1}^{2} \frac{\gamma_e}{2} \left\| \boldsymbol{d}_{e,k}^\star \circledast \boldsymbol{\mu}_e - \boldsymbol{z}_{e,k} \right\|_2^2 \right\} + \left\| (\boldsymbol{z}_{1,k}, \boldsymbol{z}_{2,k}) \right\|_{1,0}}_{R_{\text{mc}}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)}\end{aligned} \tag{5.24}$$

We solve the minimization problem (5.24) by the alternating estimation of the sparse feature images and the linear attenuation images $\{\boldsymbol{\mu}_e : e = 1, 2\}$. Given the current estimates of the sparse coefficients $\{\boldsymbol{z}_k^t : k = 1, \ldots, K\}$, the image update $\boldsymbol{\mu}_e^t$ at iteration t is obtained through the following minimization problem

$$\boldsymbol{\mu}_e^t = \underset{\boldsymbol{\mu}_e \in (\mathbb{R}^+)^m}{\arg\min}\ \Phi_e^t(\boldsymbol{\mu}_e) \tag{5.25}$$

$$\text{with } \Phi_e(\boldsymbol{\mu}_e) = \rho_e L_e(\boldsymbol{\mu}_e, \boldsymbol{y}_e) + \frac{\gamma_e}{2} \sum_{k=1}^{K} \left\| \boldsymbol{d}_{e,k}^\star \circledast \boldsymbol{\mu}_e - \boldsymbol{z}_{e,k}^t \right\|_2^2 .$$

In this work, we utilized a L-BFGS algorithm (Nocedal and Wright 2006*b*, Chapter 7) to solve 5.4.2. We utilized the implementation proposed in Zhu et al. (1997). We also used the L-BFGS algorithm to minimize $L_1(\cdot, \boldsymbol{y}_1)$ and $L_2(\cdot, \boldsymbol{y}_2)$ (without penalty) in order to obtain initial images $\boldsymbol{\mu}_1^0$ and $\boldsymbol{\mu}_2^0$.

The other part of the alternating scheme is to update the sparse features $\boldsymbol{z}_{e,k}^t$ given the current estimate of $\boldsymbol{\mu}_e^t$. This step is achieved using the multi-channel thresholding operator defined in 5.16.

The pseudo-code for MCAOL reconstruction algorithm is detailed in Algorithm 5.

Algorithm 5: MCAOL Reconstruction Algorithm

Input: Initial images $(\boldsymbol{\mu}_1^0, \boldsymbol{\mu}_2^0)$, DECT learned filters $\boldsymbol{D}^\star = (\boldsymbol{D}_1^\star, \boldsymbol{D}_2^\star)$, joint sparsity weight $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$, penalty weights $\boldsymbol{\rho} = (\rho_1, \rho_2)$, DE sinogram $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2)$, system matrix $\boldsymbol{A}$, intensities (S_1, S_2) , #outer iterations N_{outer}.

Output: Reconstructed images $(\boldsymbol{\mu}_1^\star, \boldsymbol{\mu}_2^\star)$

for $t = 0, \ldots, N_{\text{outer}} - 1$ **do**

  Update sparse codes (in parallel) ;

  for $k, \ldots, K$ **do**

    $(\boldsymbol{z}_{1,k}^{t+1}, \boldsymbol{z}_{2,k}^{t+1}) \leftarrow \mathcal{T}_{\text{mc}}(\boldsymbol{d}_{1,k}^\star \circledast \boldsymbol{\mu}_1^t, \boldsymbol{d}_{2,k}^\star \circledast \boldsymbol{\mu}_2^t, \boldsymbol{\gamma})$;

  end

  Update linear attenuation images ;

  $\boldsymbol{\mu}_1^{t+1} \leftarrow \text{L-BFGS}(\Phi_1^t, \text{init} = \boldsymbol{\mu}_1^t \mid \boldsymbol{y}_1, \boldsymbol{z}_1^{t+1}, \boldsymbol{D}_1^\star, \boldsymbol{A}, S_1, \rho_1, \gamma_1)$;

  $\boldsymbol{\mu}_2^{t+1} \leftarrow \text{L-BFGS}(\Phi_2^t, \text{init} = \boldsymbol{\mu}_2^t \mid \boldsymbol{y}_2, \boldsymbol{z}_2^{t+1}, \boldsymbol{D}_1^\star, \boldsymbol{A}, S_2, \rho_2, , \gamma_2)$;

end

$\boldsymbol{\mu}_1^\star \leftarrow \boldsymbol{\mu}_1^{N_{\text{outer}}}$;

$\boldsymbol{\mu}_2^\star \leftarrow \boldsymbol{\mu}_2^{N_{\text{outer}}}$;

5.5 Validation

We validated the proposed methods on two different DECT low-dose acquisition setup. In particular, we analyzed the case of sparse-view DECT reconstruction with normal photon dose and the case of extreme low-photon counts with increased number of views. By approximating the dose as the product of the number of views and photon counts, the latter case represents a more challenging scenario since the overall dose considered is lower than the sparse-view case. Our implementation was based on CONVOLT (Chun 2019).

5.5.1 Methods Used for Comparison

The objective of the simulations with sparse views and normal X-ray source intensity is to demonstrate that MCAOL achieves improved accuracy compared to reconstructing each energy separately by solving 5.1 with the CAOL regularizer defined in 5.9 and with the TV regularizers, as well as simultaneously by solving 5.10 with the JTV regularizer, respectfully defined as

$$R_{\text{tv}}(\boldsymbol{\mu}) = \sum_{j=1}^{m} \sum_{k \in \mathcal{N}_j} \omega_{j,k} \sqrt{(\mu_j - \mu_k)^2 + \varepsilon} \tag{5.26}$$

and

$$R_{\text{jtv}}(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = \sum_{j=1}^{m} \sum_{k \in \mathcal{N}_j} \omega_{j,k} \sqrt{(\mu_{1,j} - \mu_{1,k})^2 + (\mu_{2,j} - \mu_{2,k})^2 + \varepsilon} \tag{5.27}$$

where $\mathcal{N}_j$ denotes the 8 nearest neighboring pixels of pixel j and $\omega_{j,k}$ are weights ($\omega_{j,k} = 1$ for axial neighbors and $\omega_{j,k} = 1/\sqrt{2}$ for diagonal neighbors), and $\varepsilon > 0$ is a small real value to ensure differentiability. For each method, we used the L-BFGS solver to estimate $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$.

The experiment with extreme low-counts aims at demonstrating that considering a weighted least-squares approximation of the log-likelihood function no longer guarantees effective reconstruction results, instead the exact Poisson statistics should be accounted. This results in a degradation of the performance of CAOL when optimized through the PWLS solver while using the quasi-Newton solver L-BFGS leads to improved qualitative and quantitative results.

5.5.2 Methodology

All experiments were validated by generating the DECT measurements as in equation 5.21 and then running $M = 20$ Poisson noise instances as in equation 5.22 from a Ground Truth (GT) image $\boldsymbol{\mu}_e^{GT} = [\mu_{e,1}^{GT}, \ldots, \mu_{e,m}^{GT}]^\top \in \mathbb{R}^m$, $e = 1, 2$. As performance metrics, we considered the mean absolute bias (AbsBias) error function defined as follows

$$\text{AbsBias} = \frac{1}{N_{\mathcal{R}}} \frac{1}{N_{\text{noise}}} \sum_{j \in \mathcal{R}} \sum_{M=1}^{N_{\text{noise}}} \left| \mu_{e,j}^{[M]} - \mu_{e,j}^{GT} \right| \tag{5.28}$$

where $\mu_{e,j}^{[M]}$ indicates the reconstructed linear attenuation coefficient at image pixel j from the M-th Poisson noise replicate, $\mathcal{R}$ is the spatial region of interest and $N_{\mathcal{R}}$ is the number of pixels in the region $\mathcal{R}$. Furthermore, we compute the Standard Deviation (STD) defined as

$$\text{STD} = \frac{1}{N_{\mathcal{R}}} \sum_{j \in \mathcal{R}} \sqrt{\frac{1}{N_{\text{noise}}} \sum_{M=1}^{N_{\text{noise}}} \left(\mu_{e,j}^{[M]} - \bar{\mu}_{e,j} \right)^2} \tag{5.29}$$

where $\bar{\mu}_{e,j} = \frac{1}{N_{\text{noise}}} \sum_{M=1}^{N_{\text{noise}}} \mu_{e,j}^{[M]}$. In this work $\mathcal{R}$ corresponds to the non-negative pixels region of $\boldsymbol{\mu}_e^{GT}$ and is the same for both energy levels.

The simulations were repeated for all the methods, for different values of the regularization parameters in the objective functions 5.1 and 5.10 in order to plot AbsBias/STD curves. The quality of the reconstruction is assessed by the proximity of the curve to the origin. Training and reconstruction were performed according to the below-described settings.

Training The optimization problem (5.13) is minimized using the BPEG-M algorithm Chun and Fessler (2019*b*) with normalized input dataset. To investigate

(E_1), $\gamma_1 = 800$ (E_2), $\gamma_2 = 800$ (E_1), $\alpha = 0.01$ (E_2), $\alpha = 0.01$

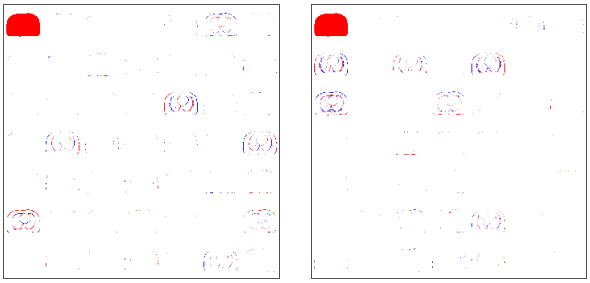
(a) Learned $l_{1,0}$ filters via MCAOL **(b)** Learned l_0 filters via CAOL

Fig. 5.2 Learned filters $\{(\boldsymbol{d}_{1,k}, \boldsymbol{d}_{2,k})\}$ with $R = K = 49$ using the XCAT training dataset, for a MCAOL and b CAOL.

the trade-off between accuracy and features sparsity, we tested (5.13) with different values of $\gamma_1 = \gamma_2$ with filter $(\boldsymbol{d}_{1,k}, \boldsymbol{d}_{2,k})$ of dimension $R = 49$ and number of filters $K = 49$. For each simulation, we tuned $\gamma_1 = \gamma_2$ by testing different values to investigate the effect. For all the datasets, we have used a training of $P = 25$ images for each energy e. Regarding the BPEG-M algorithm, we set the tolerance value equal to 10^{-4} and the maximum number of iterations to 3×10^3. For the CAOL training algorithm, we have used the same settings as detailed for MCAOL except that we tuned a single regularization weight α in the optimization problem 5.4 for each separate energy channel.

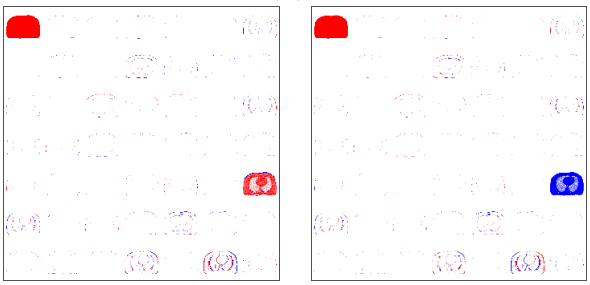
Reconstruction MCAOL and JTV reconstructions were achieved by solving 5.10 with R_{mc} defined as 5.18 and 5.27 respectively, while CAOL and TV reconstructions by solving 5.1 for each energy bin $e = 1, 2$ separately with R defined as 5.9 and 5.26 respectively. MCAOL and CAOL were achieved using $N_{\text{outer}} = 300$ outer iterations while the inner image update is obtained using the L-BFGS algorithm with 300 iterations. The (γ_1, γ_2)-values and β-values were the same as for training. TV and JTV reconstructions were achieved with the L-BFGS algorithm with 300 iterations. The measurements were obtained from the GT images $\boldsymbol{\mu}_e^{GT}$ outside the training set and the reconstructions were repeated for each noise instance M, for a range of (ρ_1, ρ_2)-values with $\rho_1 = \rho_2$ and for a range of β-values, in order to obtain AbsBias-versus-STD curves.

We performed sparse-views and low-dose experiments on a simulated XCAT phantom and clinical data to assess the potential of the method for medical practice as detailed below. The experiments were conducted with fixed X-ray dose amount, i.e., by selecting the number of angles and the X-ray source intensity, and we evaluated the quality of the linear attenuation images reconstructed with different methods, both qualitatively and quantitatively.



$\boldsymbol{z}_{1,k}$ CAOL $\boldsymbol{z}_{2,k}$ CAOL

(a)



$\boldsymbol{z}_{1,k}$ MCAOL $\boldsymbol{z}_{2,k}$ MCAOL

(b)

Fig. 5.3 XCAT Phantom: estimated sparse feature maps $\boldsymbol{z}_{2,k}$ for $e = 1, 2$ and $k = 1, ..., 49$ using CAOL (a) and MCAOL (b); color scale: red for positive values, blue for negative values.

5.5.3 Results on XCAT Phantom

For the unsupervised MCAOL and CAOL training, the numerical data consists of 1-mm pixel-width 512×512 torso axial slice images generated from the XCAT phantom for 60 keV and 120 keV energies.

We utilized 20 slice pairs from the XCAT phantom, each pair consisting of a slice at $E1 = 60$ KeV and a slice at $E2 = 120$ KeV, to train the filters. An additional slice pair—not part of the training dataset—was used to generate the projection data. We used the MCAOL weights parameters $\gamma_1 = \gamma_2 = 800$ and the CAOL parameter $\alpha = 0.01$.

Figure 5.2 shows the pairs $(\boldsymbol{d}_{1,k}, \boldsymbol{d}_{2,k})$ of learned convolutional filters obtained by MCAOL (Fig. 5.2a) and separate learning with CAOL (Figure 5.2b). From a qualitative point of view, it is possible to highlight how the MCAOL filter pairs $\boldsymbol{d}_1, \boldsymbol{d}_2$ look to share a strong coupling as the edges are identical in the 2 energy images compared to the CAOL filters.

In order to generate the sparse-view DECT projection measurements 5.21, we modeled the projector $\boldsymbol{A}$ with a 2-mm Full Width at Half Maximum (FWHM) resolution parallel beam system and we used a 1-mm pixel-width 406×406 GT torso axial-slice images with attenuation coefficients $\boldsymbol{\mu}_1^\star, \boldsymbol{\mu}_2^\star$ at energies 120 keV (high) and

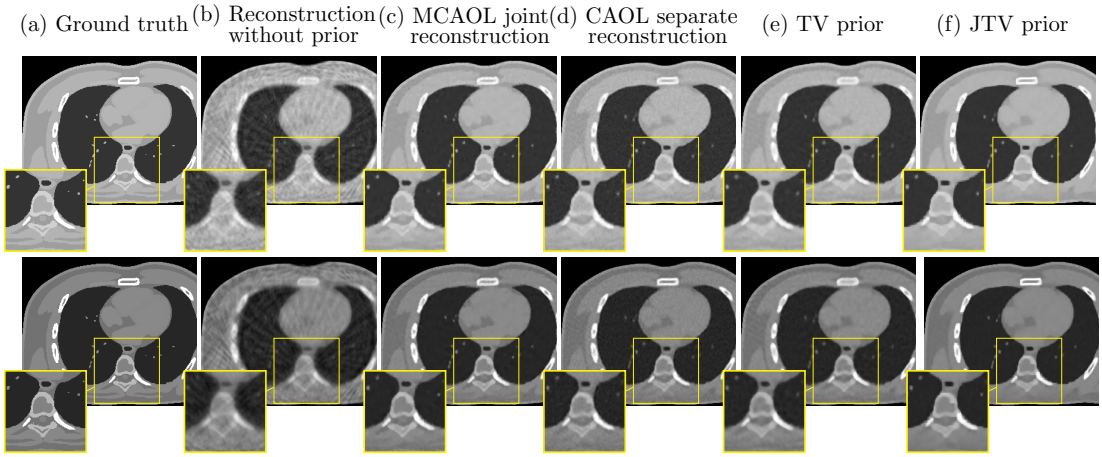
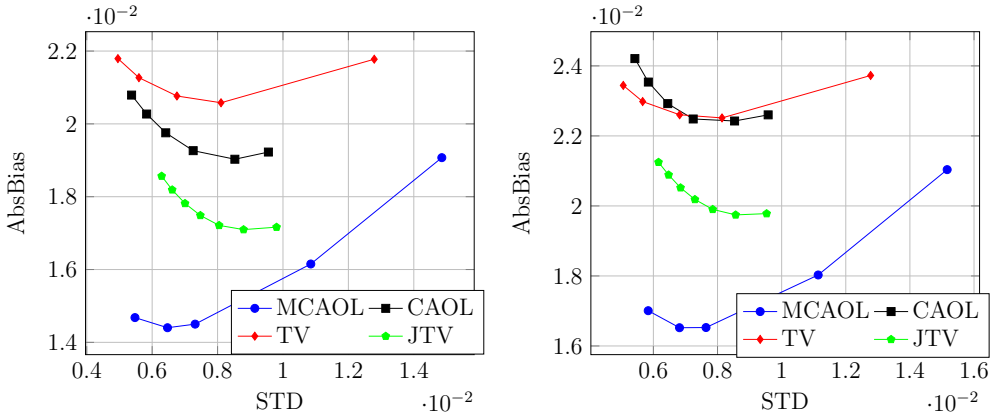


Fig. 5.4 Comparison of reconstructed XCAT phantom from different reconstruction methods for sparse-view CT with top row corresponding to high energy $E_1 = 120$ keV and bottom row to low energy $E_2 = 60$ keV: (a) Ground truth XCAT test image, (b) minimization of the NLL function without prior, (c) MCAOL reconstruction, (d) CAOL reconstruction, (e) separate reconstruction using TV prior and (f) joint reconstruction using JTV prior.



(a) 60 keV **(b)** 120 keV

Fig. 5.5 Plot of the mean absolute bias (AbsBias) versus the standard deviation (STD) for the XCAT phantom at a low X-ray source energy (60 keV) and b high X-ray source energy (120 keV).

60 keV (low) which differs from the training examples. The simulation consisted on generating sparse-view sinograms with 406 detector pixels and 60 regularly spaced projection angles, where 360° is the full view rotation. A monochromatic source with $\bar{S}_e = 10^5$ incident photons and 100 background events was used to generate each sinogram.

To support the statement that joint sparsity allows both images to inform each other, which makes the estimation of $\boldsymbol{z}_1, \boldsymbol{z}_2$ more robust, we show the estimated feature maps in figure 5.3 obtained using the XCAT data. By comparing the estimated sparse feature maps $\boldsymbol{z}_{e,k}$ for $e = 1, 2$ and $k = 1, ..., 49$ for separate reconstructions (5.3 (a)) and joint reconstruction (5.3 (b)), there are no similarities between

the feature maps obtained from separate reconstructions while the feature maps obtained from joint reconstruction have similar structures.

Figure 6.1 shows the XCAT GT and the reconstruction images for both 60 keV and 120 keV energies obtained by MCAOL and the other algorithms used for comparison. The images are obtained using the parameters which corresponds to the minimum AbsBias shown in Figures 5.5a and 5.5b. It is worth noting that MCAOL manages to substantially reduce the noise as compared with CAOL.

Figure 5.5a and Fig. 5.5b show the AbsBias against the STD results respectively for low and high X-ray source energy. Among the methods used for comparison, TV promotes sparsity of the gradient, while JTV promotes joint sparsity of the 2 gradients and therefore are particularly well-suited for XCAT. Despite this observation, it is possible to show that the minimum AbsBias obtained by MCAOL outperforms all other algorithms, or in other words by fixing the STD, the AbsBias achieved by MCAOL is always lower while it is possible to claim that by fixing the AbsBias, the STD of MCAOL is reduced.

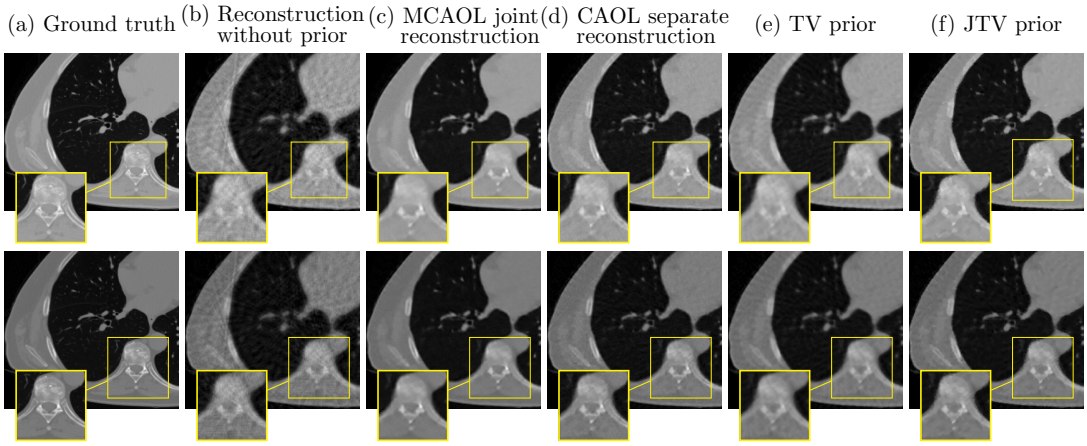


Fig. 5.6 Comparison of reconstructed clinical data from different reconstruction methods for sparse-view CT with top row corresponding to high energy $E_1 = 140$ keV and bottom row to low energy $E_2 = 70$ keV: (a) Ground truth clinical test image, (b) minimization of the NLL function without prior, (c) MCAOL reconstruction, (d) CAOL reconstruction, (e) separate reconstruction using TV prior and (f) joint reconstruction using JTV prior.

5.5.4 Results on Simulation from Clinical Data

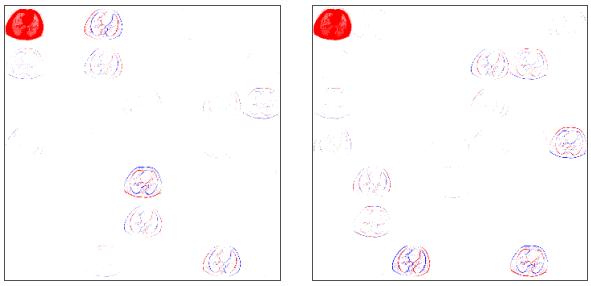
We utilized images reconstructed from data acquired on Philips IQon Spectral CT and reconstructed with a MBIR technique (*Philips IQon Elite Spectral CT product specifications* 2018). All patients provided signed permission for the use of their clinical data for scientific purposes and anonymous publication of data. The experiment was conducted in a similar fashion as for the XCAT simulation. We selected 22 slice pairs from a full body patient scan with 0.902-mm pixel-width

(E_1), $\gamma_1 = 10^4$ (E_2), $\gamma_2 = 10^4$ (E_1), $\alpha = 10^{-4}$ (E_2), $\alpha = 10^{-4}$

(a) Learned $l_{1,0}$ filters via MCAOL **(b)** Learned l_0 filters via CAOL

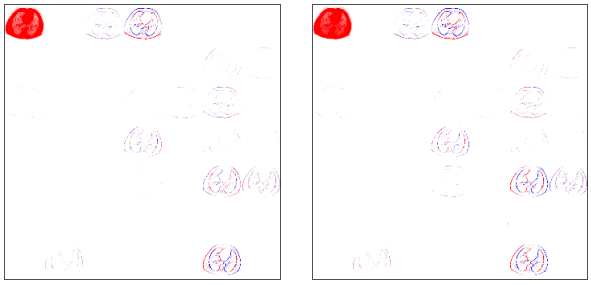
Fig. 5.7 Learned filters $\{(\boldsymbol{d}_{1,k}, \boldsymbol{d}_{2,k})\}$ with $R = K = 49$ using the clinical training dataset, for a MCAOL and b CAOL.

and 512×512 image size for the training dataset corresponding to thorax. The energies used in this study are 70 keV and 140 keV. An additional slice pair was used to generate the projection data for reconstruction, as detailed below. The pair of trained filters $(\boldsymbol{d}_{1,k}, \boldsymbol{d}_{2,k})$ obtained by both MCAOL and CAOL unsupervised learning is shown in Figure 5.7; we used the parameters $\gamma_1 = \gamma_2 = 10^4$ and the CAOL parameter $\alpha = 10^{-4}$.



$\boldsymbol{z}_{1,k}$ CAOL $\boldsymbol{z}_{2,k}$ CAOL

(a)

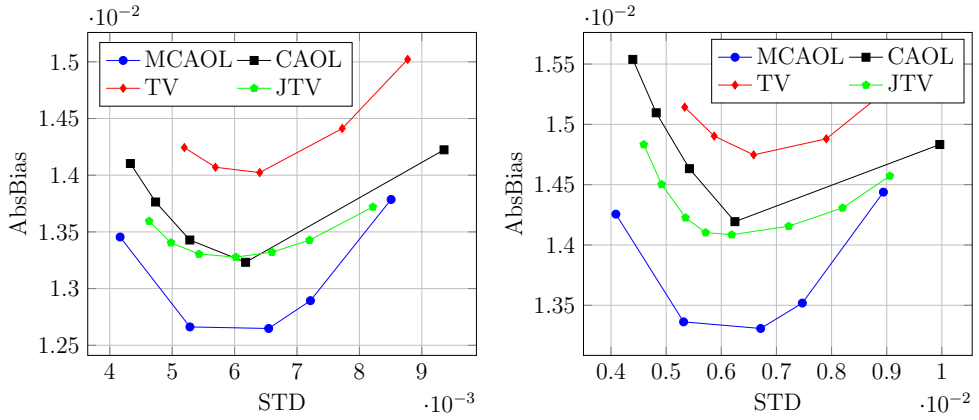


$\boldsymbol{z}_{1,k}$ MCAOL $\boldsymbol{z}_{2,k}$ MCAOL

(b)

Fig. 5.8 Clinical data: estimated sparse feature maps $\boldsymbol{z}_{e,k}$ for $e = 1, 2$ and $k = 1, ..., 49$ using CAOL (a) and MCAOL (b); color scale: red for positive values, blue for negative values.

To generate the sparse-view DECT measurements 5.21, we used the same geometrical and noise settings as for the XCAT simulation except that we used 451 detector pixels and 451×451GT thorax images with attenuation coefficients $\boldsymbol{\mu}_1^\star, \boldsymbol{\mu}_2^\star$ at energies 140 keV (high) and 70 keV (low) which differs from the training



(a) 70 keV. **(b)** 140 keV.

Fig. 5.9 Plot of the mean absolute bias (AbsBias) versus the standard deviation (STD) for the clinical data at a low X-ray source energy (70 keV) and b high X-ray source energy (140 keV).

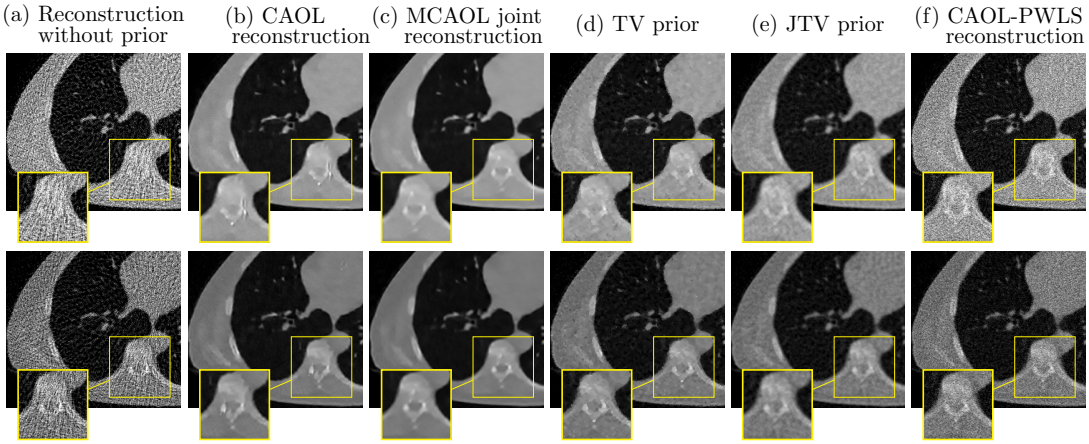


Fig. 5.10 Comparison of reconstructed clinical data from different reconstruction methods for low-dose CT with top row corresponding to high energy $E_1 = 140$ keV and bottom row to low energy $E_2 = 70$ keV: (a) Ground truth clinical test image, (b) minimization of the NLL cost function without prior, (c) MCAOL joint reconstruction, (d) energy separate reconstruction using TV prior, (e) JTV prior and (f) CAOL-PWLS reconstruction.

examples. In 5.8 we show the estimated feature maps obtained using the clinical data. As already noted previously with the XCAT data, by comparing the estimated sparse feature maps $\boldsymbol{z}_{e,k}$ for $e = 1, 2$ and $k = 1 \ldots 49$ for separate reconstructions (5.8 (a)) and joint reconstruction (5.8 (b)), there are no similarities between the feature maps obtained from separate reconstructions while the feature maps obtained from joint reconstruction have similar structures.

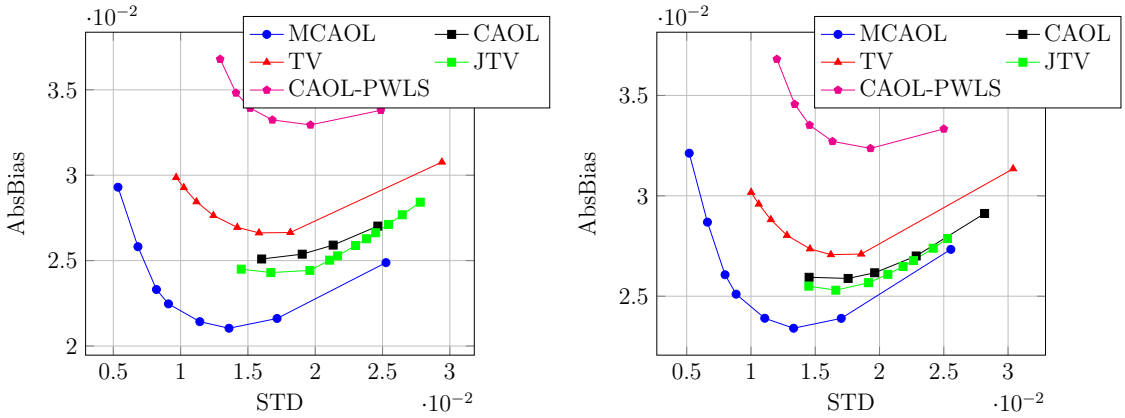
In Figure 5.6 the GT image and the reconstruction images for both energies and the different methods are shown; it is worth noting that the MCAOL reconstruction is less noisy than the CAOL reconstruction.

Figures 5.9a and 5.9b report the AbsBias versus the STD plots and we obtain

a similar behavior compared to the XCAT simulations; although the relative distance of the AbsBias among the simulated algorithms is reduced, MCAOL is still outperforming all other methods constantly either by fixing AbsBias or STD, while the performance of CAOL is improving and it is close to the JTV solution accuracy.

5.5.5 Results for Low-Dose DECT

We conducted DECT reconstruction on the same set of data as in Section 5.5.4 but with a different CT acquisition setup; we substantially decreased the initial photon counts to $I_0 = 10^3$ (reduction of 2 orders of magnitude compared to the previous experiments) and we doubled the number of views to 120. By approximating the total delivered X-ray dose as the product of the photons intensity times the number of views, it turns out that this scenario is considerably more challenging in terms of ill-posed problem with a total dose reduction of 50 times.



(a) 70 keV. **(b)** 140 keV.

Fig. 5.11 Plot of the mean absolute bias (AbsBias) versus the standard deviation (STD) for the low-dose ($I_0 = 10^3$) reconstruction with clinical data at a low X-ray source energy (70 keV) and b high X-ray source energy (140 keV).

We use this simulation to prove that MCAOL returns a more accurate solution compared to other priors. Furthermore, we prove that despite the higher computational complexity to minimize the exact Poisson NLL in 5.24 compared to solving the problem with a weighted least-squares approximated NLL, i.e., PWLS data-fit cost function, MCAOL achieves substantial improved bias accuracy compared to the PWLS solution. To perform these experiments, we used the same optimal learned convolutional filters as obtained by the MCAOL training procedure detailed in Section 5.5.4 and the GT images in Figure 5.6(a).

Figure 5.10 show the reconstruction images for both energies and different methods; MCAOL accurately reconstruct the image features compared to all other methods and it is confirmed that the PWLS model performs poorly.

Figures. 5.11a and 5.11b show either that MCAOL is consistently outperforming the other methods in terms of accuracy and variance and that the Poisson NLL formulation leads to a noticeable improvement compared to the PWLS formulation as it is indicated by comparing CAOL and CAOL-PWLS.

5.6 Discussion and Conclusions

In this work, we have extended the convolutional analysis operator framework to multi-channel imaging and we have applied and extensively analyzed the proposed method to the DECT application. The presented results show that by using the information coming from both energies and allowing the channels to "talk to each other" a more accurate solution of the reconstruction problem can be achieved together with a reduction of the noise in the estimate. The coupling between energies is encapsulated by using an $l_{1,0}$ sparse mixed norm in the MCAOL optimization problems both for training and reconstruction. We obtain consistently better performances across different DECT acquisition scenarios from sparse-views to low-dose photon counts.

The bias-variance trade-off analysis of the estimation results over the regularization parameters confirms that MCAOL allows to achieve the minimum absolute bias compared to CAOL and other MBIR state-of-the-art methods and also reduce standard deviation. Furthermore, MCAOL has the benefit of requiring less memory respect to DL methods because of the convolutional structure of the trained filters.

The MCAOL framework allows to utilize any mixed norms for the jointly sparse regularization and other norms, such as the $l_{2,1}$ -norm which as proposed by Degraux et al. (2017) for convolutional synthesis operator learning, may also be considered.

In our experiments we have considered the product between the X-ray source intensity and the number of projection angles as an empirical measure for the total transmitted X-ray dose. While this metric gives a good approximation of the dose, we consider the analysis of the standardized measure of radiation dose, i.e., CT dose index (CTDI), as well as the absorbed dose as a follow-up study.

We account the open problems of how to optimally select both the regularization norm and regularization parameter according to the dataset for future algorithm development.

Although this work focuses on the multi-channel imaging reconstruction problem, we believe that our proposed method can be utilized in conjunction to DECT to task-oriented material decomposition problems. In particular, while an approach would be to design a material decomposition module in the image space which takes as input the MCAOL reconstructed images, a more compelling strategy would be

designing a direct approach from sinograms to material images through MCAOL.

Furthermore, MCAOL method can be exploited for other multi-modal imaging application such as PET/CT and PET/MRI. In the multi-modal case, given the different intensity range on each channel, a further analysis on how to choose the NLL weights $\gamma_1 \neq \gamma_2$ in (5.10) should be conducted to properly balancing the information coming from the different modalities.

Finally, from a learning point of view, MCAOL training can be seen as a multi-channel single layer unsupervised convolutional autoencoder (Chun and Fessler 2019*b*, Appendix A) which paves the way to extend this approach to deeper autoencoder architectures to capture more complex features such as textures.

The analysis and comparison of the proposed MCAOL approach with other supervised deep learning approaches is planned as a follow-up study. It is important to stress that MCAOL inherits a precise mathematical derivation and therefore it should not be susceptible of instabilities in the reconstruction which have been proven to occur with deep learning methods (Antun et al. 2020).

We consider these problems as future development of the proposed algorithm.

CHAPTER 6

Sparse-View Joint Reconstruction and Material Decomposition for Dual-Energy Cone-Beam CT

The present work proposes a methodology for sparse-view image reconstruction in single-source rapid KVp switching in DE-CBCT. The idea is to reconstruct the low and high energy images jointly in order to exploit structural similarities, thus they inform each other during the reconstruction. The JTV regularization was used within a MBIR to encode the low and high energy images. We demonstrate the superiority of JTV regularization in comparison with TV and the Huber edge preserving prior. We evaluate the performance of the reconstructed images for material decomposition. This work was performed in parallel with the MCAOL algorithm presented in the previous chapter and it was published in the 16th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine 2021, also known as Fully3D.

6.1 Introduction

In DE-CBCT, the rapid potential switching allows consecutive projection measurements with alternating tube potentials where both the low and high energy projection data are acquired throughout a whole gantry rotation (Garnett 2020, Forghani and Mukherji 2018). The tube voltage varies between high and low, and transmission data is acquired twice for adjacent projection angles.
The major disadvantage of this method is the need of reducing the rotation speed of the system to acquire the extra projections and to account for the rise and fall times required for voltage modulation (Lam et al. 2015). Due to fast switching it is not possible to modulate the tube current between high and low energy simultaneously. It remains constant during the acquisition. Thus, the tube current needs to be increased to reduce the noise on images obtained with lower peak voltage, which results in an increase of the radiation dose (Johnson 2012, Goo and Goo 2017).
Sparse-view projection angles can reduce the radiation dose, since the total number of photons (emitted during the whole acquisition) decreases. Image reconstruction

from under-sampled projection data is now possible thanks to the advancement of CS theory. Several MBIR have been proposed based on the CS theorem. The TV penalty, which promotes sparsity in the image gradient transform domain, has been widely used as a regularization in MBIR. It successfully suppresses the streak artifacts arising from sparse-view CT data, nevertheless, it attempts to penalize the image gradient equally, regardless the underlying image structures. Thus, low contrast regions are often over smoothed (Yu et al. 2017, Zhu et al. 2013).
Aside from l_1 sparsity (Tibshirani 1996*b*), other prominent sparsity representation options include a mixture of l_1 and TV ($l_1 + TV$)(Tibshirani et al. 2005, Gao and Zhao 2010), wavelet (Mallat 1998), and tight frame (Daubechies et al. 2003). The majority of these algorithms reconstruct a single image by maximizing an objective function composed of the data fidelity term and the sparse regularization term. A multi-channel joint reconstruction technique is a highly suited method for Dual Energy Sparse CBCT.
The TV regularization can be generalized for multi-channel image reconstruction. The most simplified technique to generalize the TV in multi-channel reconstruction is to sum the total variation of the individual channels, as proposed by (Xu et al. 2014) and (Sawatzky et al. 2014) for spectral CT reconstruction. One important theoretical shortcoming of this strategy is that it independently penalizes each channel, despite the fact that strong inter-channel correlations often exist (Rigie and La Rivière 2015). A few generalizations of TV, which impose coupling in the images have been investigated, e.g the Total Nuclear Variation for spectral CT (Rigie and La Rivière 2015) or in earlier research works for color image restoration (Lefkimmiatis et al. 2013), (Holt 2014), (Keren and Gotlib 1998).
The present work proposes a methodology for sparse view image reconstruction in single-source rapid KVp switching DE-CBCT by exploiting structural similarities using the isotropic scalar JTV regularization proposed by (Sapiro and Ringach 1996) in the context of color images processing. The hypothesis behind this approach is that the low- and high-energy images can inform each other giving room for dose reduction and enhancing the spatial resolution deficit due to the down-sampled projection data.
High-quality reconstructed images allow accurate estimation of the basis materials when performing material decomposition, which constitutes the main clinical application of DE-CBCT. Thus, the present work, in addition, evaluates the performance of the reconstructed images for material decomposition. The two most common methods for reconstructing material-specific volumes from dual-energy CBCT are projection-based and image-based. The projection-based material decomposition can be conducted either by decomposing the acquired data into

material-specific projections and further reconstruct them independently (known as two-step projection-based method) or by reconstructing the decomposed images from the dual-energy sinograms in one-step inversion (known as one step material-decomposition) (Mory et al. 2018). In two-step image-based algorithms, each energy sinogram is log-transformed and reconstructed producing one volume per energy bin, which is then decomposed into material-specific volumes.
This work aims to achieve high quality reconstructed images in fast KVp switching DE-CBCT to lead to accurate material decomposition images utilizing the two-step image-based methodology.

6.2 Dual Energy Image Reconstruction

Assuming a simplified single-source rapid KVp switching DE-CBCT setting, each sinogram $\boldsymbol{y}_\ell \in \mathbb{R}^n$, obtained from the energies $\ell \in \{1, 2\}$ (low and high), is modeled by a random vector $\boldsymbol{y}_\ell = [y_{1,\ell}, \ldots, y_{n,\ell}]^\top$ with independent entries, where n is the number of detector pixels. At each detector pixel $i \in \{1, \ldots, n\}$, the number of detected photons $y_{i,\ell}$ follows a Poisson distribution:

$$y_{i,\ell} \sim \text{Poisson}(\bar{y}_{i,\ell}(\boldsymbol{\mu}_\ell)) \,, \tag{6.1}$$

with

$$\bar{y}_{i,\ell}(\boldsymbol{\mu}_\ell) = b_i \exp(-[\boldsymbol{A}\boldsymbol{\mu}_\ell]_i) + s_{i,\ell} \tag{6.2}$$

where $\boldsymbol{\mu}_\ell \in \mathbb{R}^m$ is the attenuation image at energy ℓ, $\boldsymbol{A}$ is a $n \times m$ matrix modeling the system, $s_{i,\ell}$ is a background term and m is the number of voxels in the image.

We propose to reconstruct the low- and high-energy attenuation images $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ by penalized maximum-likelihood joint estimation from the sinograms $(\boldsymbol{y}_1, \boldsymbol{y}_2)$:

$$(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2) = \underset{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \geq \boldsymbol{0}}{\arg\max}\, L_1(\boldsymbol{\mu}_1, \boldsymbol{y}_1) + L_2(\boldsymbol{\mu}_2, \boldsymbol{y}_2) - \beta R(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \tag{6.3}$$

where $R(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ is a joint regularization term, β is the regularization parameter and $L(\boldsymbol{\mu}_\ell, \boldsymbol{y}_\ell)$ is the log-likelihood defined as:

$$L(\boldsymbol{\mu}_\ell, \boldsymbol{y}_\ell) = \sum_{i=1}^{n} y_{i,\ell} \log \bar{y}_{i,\ell}(\boldsymbol{\mu}_{i,\ell}) - \bar{y}_{i,\ell}(\boldsymbol{\mu}_{i,\ell}) \,. \tag{6.4}$$

The Quasi-Newton maximization problem (6.3) is solved using a L-BFGS algorithm (Zhu et al. 1997).

6.2.1 Joint Total Variation Regularization

In the present work, we used the JTV penalty term $R(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ inspired from (Ehrhardt et al. 2014*b*) and (Sapiro and Ringach 1996). The JTV regularization term can be written as:

$$R(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = \sum_{j=1}^{m} \left(\|[\nabla \boldsymbol{\mu}_1]_j\|^2 + \|[\nabla \boldsymbol{\mu}_2]_j\|^2 + \gamma^2 \right)^{1/2} \tag{6.5}$$

where $\nabla \boldsymbol{\mu}_\ell \in \mathbb{R}^{m \times d}$ $(d = 2, 3)$ is the gradient image of $\boldsymbol{\mu}_\ell$ and $[\nabla \boldsymbol{\mu}_\ell]_j \in \mathbb{R}^d$ is the gradient at voxel j, and $\gamma > 0$ tunes the smoothness of the prior (for differentiability). The image $\boldsymbol{\mu}_l$ is reshaped in a matrix, then we compute ∇ as the finite differences along x and y axis as shown in Section 3.2, equation 3.30.

The role of this prior is to promote structural similarities by enforcing joint sparsity of the 2 gradient images. We compared the proposed approach of jointly reconstruct the images with JTV against reconstructing separately with TV as follows:

$$\hat{\boldsymbol{\mu}}_\ell = \underset{\boldsymbol{\mu}_\ell \geq \mathbf{0}}{\arg\max}\, L(\boldsymbol{\mu}_\ell, \boldsymbol{y}_\ell) - \delta S(\boldsymbol{\mu}_\ell) \tag{6.6}$$

with

$$S(\boldsymbol{\mu}_\ell) = \sum_{j=1}^{m} \left(\|[\nabla \boldsymbol{\mu}_\ell]_j\|^2 + \eta^2 \right)^{1/2} \tag{6.7}$$

where δ and η play the same roles as β and γ respectively.

Moreover, we compared against existing edge preserving prior (e.g. Huber Prior):

$$\hat{\boldsymbol{\mu}}_\ell = \underset{\boldsymbol{\mu}_\ell \geq \mathbf{0}}{\arg\max}\, F(\boldsymbol{\mu}_\ell, \boldsymbol{y}_\ell) - \rho U(\boldsymbol{\mu}_\ell) \tag{6.8}$$

with

$$U(\boldsymbol{\mu}_\ell) = \sum_{j=1}^{m} \sum_{k \in N_j} \omega_{j,k} \Phi(\mu_\ell^j - \mu_\ell^k) \tag{6.9}$$

where N_j are the neighborhood of j and ρ controls the weight of the regularization term; $\omega_{j,k}$ are weights ($\omega_{j,k} = 1$ for axial pixels and $\omega_{j,k} = 1/\sqrt{2}$ for diagonal pixels. For the Huber prior the typical choice of $\Phi(x)$ are (Nuyts et al. 2002)

$$|x| \leqslant \sigma : \quad \Phi(x) = \frac{x^2}{2\sigma^2} \quad |x| > \sigma : \quad \Phi(x) = \frac{|x| - \sigma/2}{\sigma} \tag{6.10}$$

The Huber prior compares the difference between neighboring pixels with the value of the parameter σ (Nuyts et al. 2002).

With these two approaches using TV and Huber priors, each energy image is reconstructed independently without sharing structural information.

Algorithm 6: JTV Reconstruction Algorithm

Input: Initial images $(\boldsymbol{\mu}_1^0, \boldsymbol{\mu}_2^0)$, penalty weight β, prior smoothness $\gamma > 0$, dual-energy sinogram $(\boldsymbol{y}_1, \boldsymbol{y}_2)$, forward operator $\boldsymbol{A}$, intensity $(\boldsymbol{b}_1, \boldsymbol{b}_2)$

#outer iterations N_{outer}.

Output: Reconstructed images $(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2)$

for $t = 1, \ldots, N_{\text{outer}} - 1$ **do**

Update low energy CBCT image

$\boldsymbol{\mu}_1^t \leftarrow \text{L-BFGS}(\boldsymbol{\mu}_1^{t-1}, \boldsymbol{\mu}_2^{t-1}, \boldsymbol{y}_1, \boldsymbol{A}, \boldsymbol{b}_1, \beta, \gamma)$

Update high energy CBCT image

$\boldsymbol{\mu}_2^t \leftarrow \text{L-BFGS}(\boldsymbol{\mu}_2^{t-1}, \boldsymbol{\mu}_1^t, \boldsymbol{y}_2, \boldsymbol{A}, \boldsymbol{b}_2, \beta, \gamma)$

end

$\hat{\boldsymbol{\mu}}_1 \leftarrow \boldsymbol{\mu}_1^{N_{\text{outer}}}$;

$\hat{\boldsymbol{\mu}}_2 \leftarrow \boldsymbol{\mu}_2^{N_{\text{outer}}}$;

6.3 Experiments

We performed the dual-energy image reconstruction by iteratively alternating between (i) updating the low-energy image $\boldsymbol{\mu}_1$ and (ii) updating the high-energy image $\boldsymbol{\mu}_2$ using the L-BFGS algorithm. We initialized the images using a MLTR algorithm (Nuyts et al. 1998) without explicit prior. The pseudo-code for JTV reconstruction algorithm is detailed in Algorithm 6.

6.3.1 Results on XCAT phantom

The numerical down-sampled projection data was modeled by forward projection of a 0.85-mm pixel width 512×512 torso axial slice images generated from the XCAT phantom at two energy levels (Segars et al. 2010). We modeled the projector $\boldsymbol{A}$ with a 1-mm FWHM resolution fan beam system. We simulated sparse-view 60-angle sinograms, where 360 is the number of angles in full view. We distributed the projection angles such that, in a single gantry rotation, one projection angle corresponds to the low energy, and the consecutive corresponds to the high energy projection. For each sinogram, we use a monochromatic source with 10^5 incident photons and 100 background events. In this work, the values of the linear attenuation coefficients at each phantom were generated assuming X-ray energies of 70-KeV (low) and 140-KeV (high).

Figure 6.1 shows the reconstructed images using JTV regularization, TV, the Huber prior and without prior. In absence of prior, the images suffer from under-sampling artifacts. The selected ROI in the images show the improved performance of JTV as compared with TV and the Huber edge preserving prior. Low-contrast features can be better identified with JTV.

70 KeV	PSNR	SSIM	140 KeV	PSNR	SSIM
JTV	64.85	0.9996	**JTV**	66.66	0.9998
TV	62.01	0.9993	**TV**	63.01	0.9992
Gain(%)	4.58	0.030	**Gain(%)**	5.79	0.06
Huber	60.32	0.9987	**Huber**	62.98	0.9990
Gain(%)	7.51	0.083	**Gain(%)**	5.84	0.080

Table 6.1 Peak Signal-to-Noise Ratio (PSNR) in dB and the Structural Similarity Index (SSIM) for the JTV, TV and Huber reconstruction algorithms at low-energy (70 KeV) and high-energy (140 KeV). The Gain is calculated as $\text{Gain}(\%) = 100 \cdot (\text{JTV} - \text{TV})/\text{TV}$ in the case of TV regularization and $\text{Gain}(\%) = 100 \cdot (\text{JTV} - \text{Huber})/\text{Huber}$ in the case of Huber prior

Furthermore, we quantitatively evaluated the performance of JTV using the PSNR defined as:

$$\text{PSNR(dB)} = 10 \cdot \log_{10} \left(\frac{\left\{ \max(\hat{\mu}^{GT}) \right\}^2}{\sum_{j=1}^{m} \frac{1}{m} \left(\hat{\mu}_j - \hat{\mu}_j^{GT} \right)^2} \right) \tag{6.11}$$

where $\hat{\mu}_j$ and $\hat{\mu}^{GT}$ represent the intensity value at the pixel j in the reconstructed image and the ground truth respectively.

We utilized the SSIM to measure the visual impact of three characteristics in the reconstructed image: luminance, contrast and structure. The SSIM between two images (x, y) can be defined as: (Kawahara et al. 2020, Wang et al. 2004)

$$\text{SSIM}(x, y) = \frac{\left(2\eta_x \eta_y + C_1 \right) (2\sigma_{xy} + C_2)}{\left(\eta_x^2 + \eta_y^2 + C_1 \right) \left(\sigma_x^2 + \sigma_y^2 + C_2 \right)} \tag{6.12}$$

where $\eta_x, \eta_y, \sigma_x, \sigma_y$, and σ_{xy} are the local means, standard deviations, and cross-covariance for images x, y. The constants C_1 , C_2 are used to prevent a zero denominator and to avoid instability for image regions where the local mean or standard deviation is close to zero.

Table 6.1 shows the values of the metrics mentioned above for the reconstructed images utilizing the XCAT phantom. At both energy levels, the JTV approach results in higher PSNR and SSIM.

For the low-energy image the JTV gain with respect to TV was 4.58% in PSNR and 0.03% in SSIM while for the high energy image the gain was 5.79% in PSNR and 0.06% in SSIM. Regarding the Huber prior, the gain was 7.51% and 5.84% in PSNR for the low and high energy image respectively, while the gain in SSIM was 0.08% for both, the low and high energy images.

We also analyzed the bias/variance trade-off of JTV, TV and Huber prior on

(a) Ground truth (b) No prior (c) Huber prior (d) TV prior (e) JTV prior

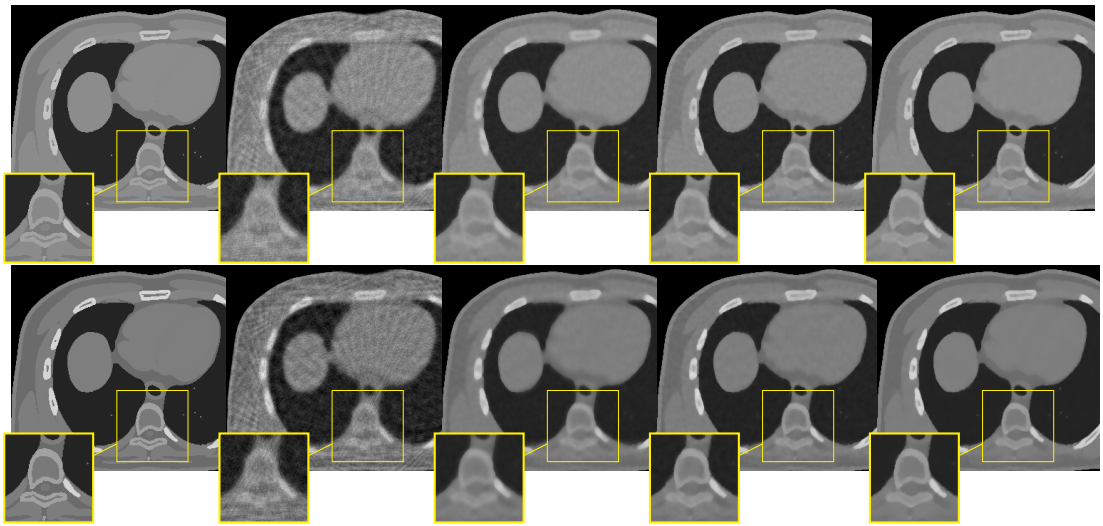


Fig. 6.1 Comparison of reconstructed XCAT phantom from different reconstruction methods for sparse-view DE-CBCT with top row corresponding to high energy ($E = 140$ KeV) and bottom row to low energy ($E = 70$ KeV): (a) Ground truth, (b) reconstruction without prior, (c) reconstruction utilizing Huber prior, (d)TV reconstruction, (e) joint reconstruction using JTV prior

the low and high energy images by plotting the absolute bias (AbsBias) against the variance of the total image, based on $K = 30$ realizations of $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, for each value of the regularization parameter β, δ and ρ.

$$\begin{aligned} \text{AbsBias} &= \frac{1}{K}\frac{1}{m}\sum_{k=1}^{K}\sum_{j=1}^{m}\left|\hat{\mu}_j^k - \hat{\mu}_j^{GT}\right| \\ \text{Var} &= \frac{1}{K}\frac{1}{m}\sum_{k=1}^{K}\sum_{j=1}^{m}\left(\hat{\mu}_j^k - \hat{\mu}_j^{\eta}\right)^2 \\ &\text{with} \quad \hat{\mu}_j^{\eta} = \frac{1}{K}\sum_{k=1}^{K}\hat{\mu}_j^k \end{aligned} \tag{6.13}$$

where $\hat{\mu}_j^k$ the reconstructed image at pixel j for the noise realization k and $\hat{\mu}_j^{GT}$ is the ground truth.

Figure 6.2 and 6.3 show that JTV achieves lower absolute bias for any variance level in the two energy images.

6.4 Results on simulation from Clinical Data

The clinical dataset is acquired from the Philips IQon Spectral CT scanner from the Poitiers University Hospital. All patients used in the study provided signed permission to use their clinical data for scientific purposes and anonymous publication

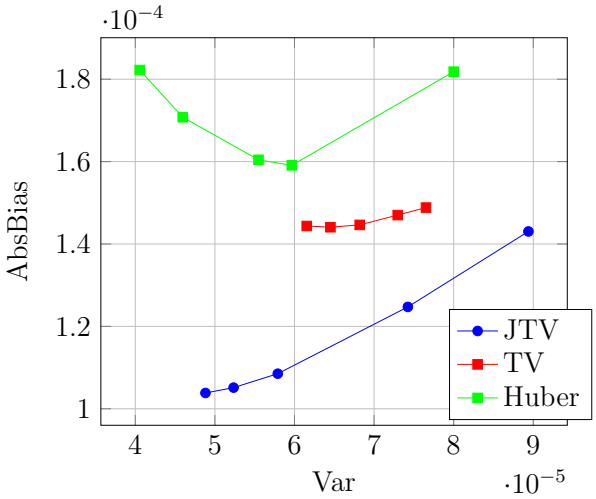


Fig. 6.2 Plot of the Absolute Bias (AbsBias) versus the Variance (VAR) for the sparse-view reconstruction with XCAT data and high X-ray source energy, 140 keV. Each point on the curve corresponds to a value of the regularization parameter β, δ and ρ.

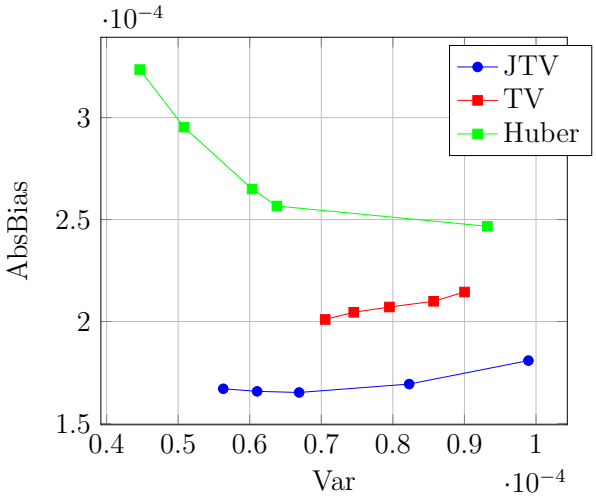


Fig. 6.3 Plot of the Absolute Bias (AbsBias) versus the Variance (Var) for the sparse-view reconstruction with XCAT data and low X-ray source energy, 70 keV. Each point on the curve corresponds to a value of the regularization parameter β, δ and ρ.

of data.

We selected 2D slices from a full body patient scan with 0.902-mm pixel-width and 512×512 image size corresponding to thorax. The monochromatic energies used in this study are 70 keV and 140 keV. To generate the sparse-view DE-CBCT measurements we used the same geometrical and noise settings as for the XCAT simulation.

In Figure 6.4 we observe that JTV outperforms TV for clinical data; TV-reconstructed images shows aliasing artifacts.

Moreover, we computed the PSNR and SSIM for clinical data. Table 6.2 shows

(a) Ground truth (b) No prior (c) Huber prior (d) TV prior (e) JTV prior

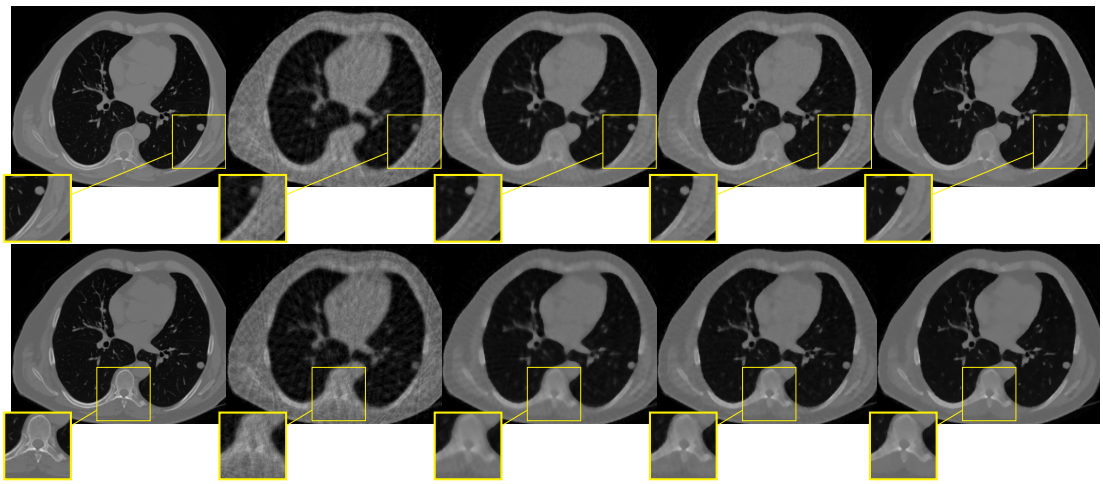


Fig. 6.4 Comparison of reconstructed Clinical data from different reconstruction methods for sparse-view with top row corresponding to high energy ($E = 140$ KeV) and bottom row to low energy ($E = 70$ KeV): (a) Ground truth, (b) reconstruction without prior, (c)reconstruction utilizing Huber prior (d)TV reconstruction, (e) joint reconstruction using JTV prior.

70 KeV	**PSNR**	**SSIM**	**140 KeV**	**PSNR**	**SSIM**
JTV	63.81	0.9993	**JTV**	66.24	0.9994
TV	62.27	0.9989	**TV**	65.23	0.9992
Gain(%)	2.45	0.040	**Gain(%)**	1.54	0.02
Huber	59.74	0.9978	**Huber**	63.68	0.9990
Gain(%)	6.80	0.150	**Gain(%)**	4.02	0.04

Table 6.2 Peak Signal-to-Noise Ratio (PSNR) in dB and the Structural Similarity Index (SSIM) for the JTV, TV and Huber reconstruction algorithms at low energy (70 KeV) and high energy (140 KeV). The Gain is calculated as $\text{Gain}(\%) = 100 \cdot (\text{JTV} - \text{TV})/\text{TV}$ for TV and $\text{Gain}(\%) = 100 \cdot (\text{JTV} - \text{Huber})/\text{Huber}$ for the Huber prior.

the SSIM and PSNR values are higher for JTV for the low and high energy images.

Figures 6.5 and 6.6 report the AbsBias versus the Var plots. We obtain a similar behavior compared to the XCAT simulations, JTV outperforms TV and the Huber edge preserving prior.

6.4.0.1 Modulation Transfer Function

The spatial resolution of the DE-CBCT images reconstructed utilizing the different algorithms was measured by computing the MTF derived from an edge measurement. The MTF is a metric that indicates how efficiently a system transmits contrast across spatial-frequencies. In this work we utilized the slanted edge technique to measure the MTF (Richard et al. 2012). Initially, an Edge Spread Function (ESF) was obtained at the slanted edge between the trachea and the lung. The ESF was

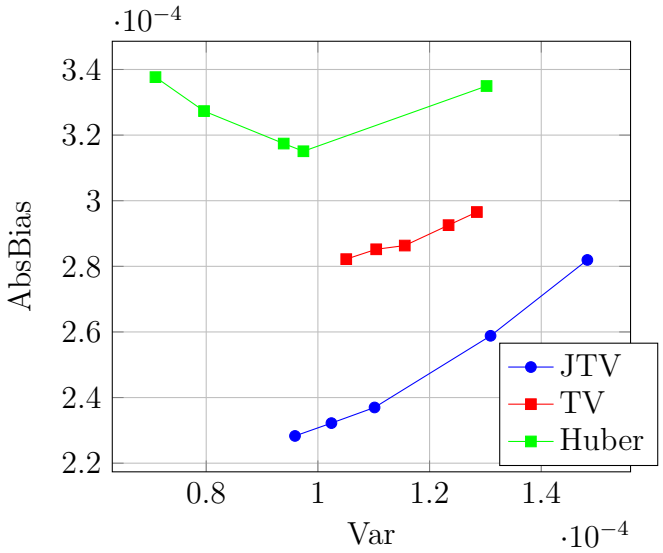


Fig. 6.5 Plot of the Absolute Bias (AbsBias) versus the Variance (Var) for the sparse-view reconstruction with Clinical Data and high X-ray source energy, 140 keV.

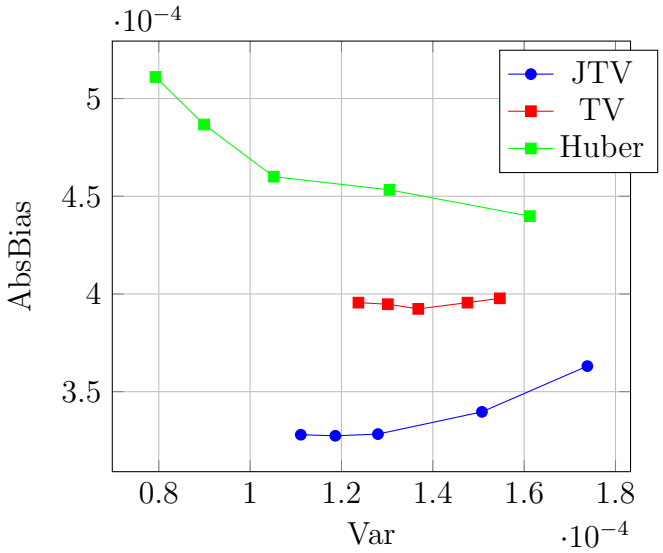


Fig. 6.6 Plot of the Absolute Bias (AbsBias) versus the Variance (Var) for the sparse-view reconstruction with Clinical Data and low X-ray source energy, 70 keV.

re-sampled using linear interpolation and averaged across multiple ESF realizations to reduce noise in ESF. Then, a Line Spread Function (LSF) was estimated by taking the derivative of the ESF as:

$$LSF(x) = \frac{\partial(ESF)}{\partial x} \tag{6.14}$$

Finally, the MTF was obtained by applying the Fourier Transform to the LSF Zhang et al. (2017), Richard et al. (2012).

$$MTF(t) = \mathscr{F}\{LSF(x)\} \tag{6.15}$$

Figures 6.7 and 6.8 show the MTF of the images reconstructed utilizing TV, JTV and the Huber regularization for high- and low-energy images respectively. We observe that JTV produces higher spatial resolution than TV and the Huber prior. The spatial resolution analysis reveals that JTV increases detectability and edge-preservation in comparison to TV.

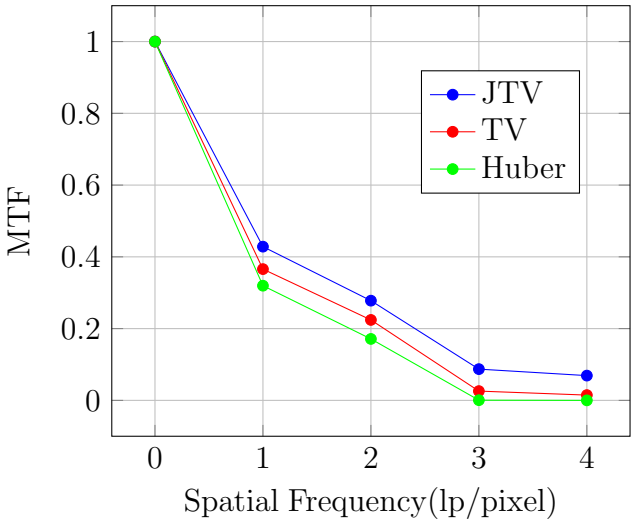


Fig. 6.7 MTF obtained from the reconstructed images utilizing JTV, TV and the Huber priors for high-energy clinical data , 140 keV.

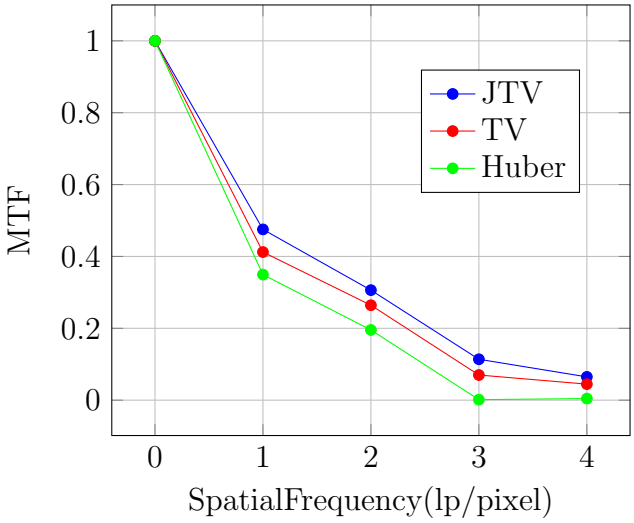


Fig. 6.8 MTF obtained from the reconstructed images utilizing JTV, TV and the Huber priors for low-energy clinical data , 70 keV.

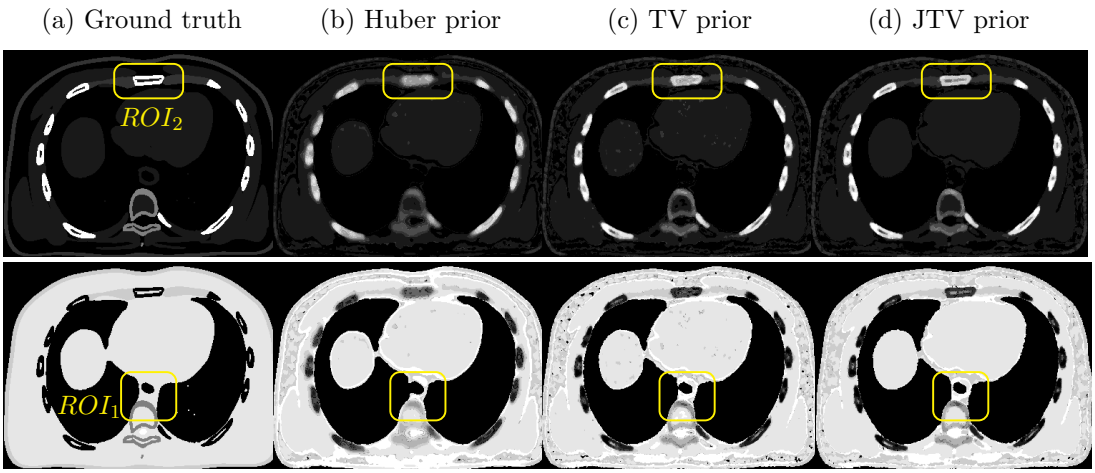


Fig. 6.9 Decomposed images into Bone (top row) and Soft Tissue (bottom row) basis materials utilizing the XCAT images obtained form the (a) ground truth, (b) Huber prior reconstruction (c) reconstruction with TV and (d) reconstruction using JTV prior

6.5 Results for Material Decomposition

An important application of DE-CBCT is material decomposition. It relies on the approximation of the linear attenuation coefficient at each pixel in the CT image by a linear combination of the attenuation values of basis materials. Thus, the material decomposition can be written as:

$$\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \mu_{11} & \mu_{21} \\ \mu_{12} & \mu_{22} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \tag{6.16}$$

where the subscripts $p \in (1, 2)$ indicate two basis materials, μ_{pl} is the linear attenuation coefficient of material p at the energy $l \in (1, 2)$, z_1 and z_2 are the volume fractions of the basis materials at the same position of two basis material images and μ_1 and μ_2 are the low and high energy reconstructed images. The aim of material decomposition algorithms is to estimate the volume fractions knowing the linear attenuation coefficient of the basis materials. In the present study we utilize the methodology proposed in Friedman et al. (2012).

We decompose into Soft Tissue (z_1): Breast Tissue 308 ICRU-44, $1.00 g/cm^3$; and Bone: B-100 Bone-Equivalent Plastic, $1.50 g/cm^3$). (of Standards and Technology 2001)

Figure 6.9 shows the image decomposition into bone and soft tissue basis material from the reconstructed images using JTV, TV, the Huber Prior and the ground truth. We observe that small bone structures can be better identified in the bone-decomposed image obtained from the JTV reconstruction.

We quantitatively compared the performance of JTV for material decomposition

Bone	RMSE	Soft Tissue	RMSE
JTV	0.1722	**JTV**	0.1710
TV	0.2142	**TV**	0.1757
Huber	0.2672	**Huber**	0.2395

Table 6.3 Root Mean Square Error (RMSE of the soft tissue (ROI_1) and bone ROI_2 images decomposed utilizing JTV,TV and The Huber regularization.

(a) Ground truth (b) Huber prior (c) TV prior (d) JTV prior

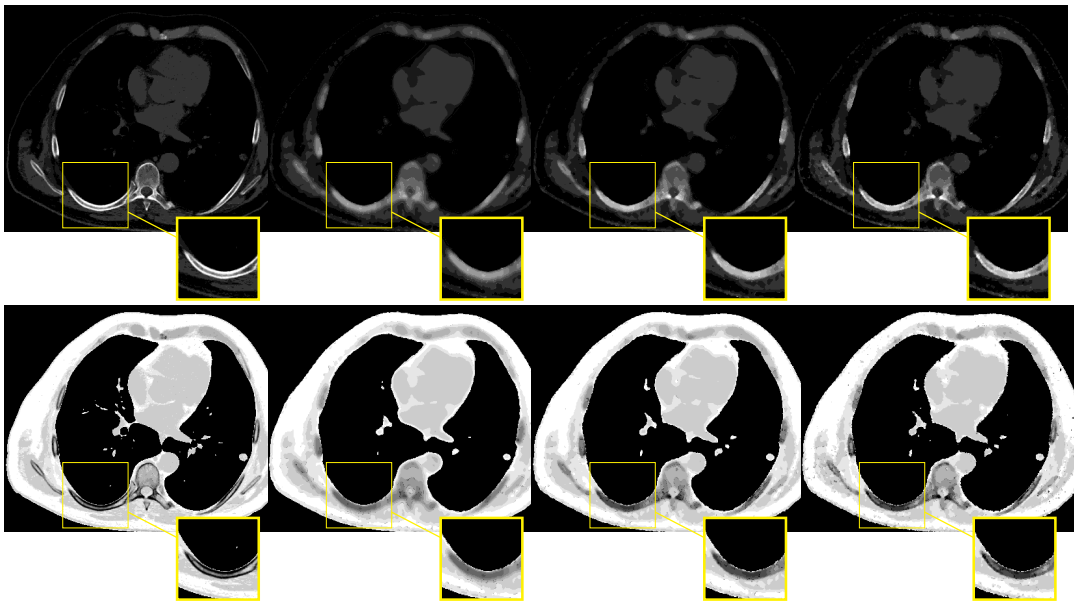


Fig. 6.10 Decomposed images into Bone (top row) and Soft Tissue (bottom row) basis materials utilizing the clinical images obtained form the (a) ground truth, (b) Huber prior reconstruction (c) reconstruction with TV and (d) reconstruction using JTV prior

by computing the RMSE in the selected ROI as:

$$RMSE = \sqrt{\frac{1}{m}\sum_{j=1}^{m}(\hat{\mu}_j - \hat{\mu}_j^{GT})^2} \tag{6.17}$$

Table 6.3 show the values of the RMSE calculated in ROI_1 and ROI_2 as shown in 6.9. For both basis materials, the decomposition utilizing JTV reconstructed images scores lower RMSE compared with TV and Huber reconstructed images.

We performed material decomposition from the reconstructed images utilizing clinical data. Figure 6.10 shows the images decomposed into soft tissue and bone. We observe similar behavior to the results obtained with the XCAT data. Small bone structures are better identified from the image obtained with JTV reconstruction.

6.6 Discussion and Conclusions

The present work proposes an image reconstruction methodology for sparse-view DE-CBCT using a JTV regularization. The coupled regularizer exploits structural similarities between the two images acquired at low- and high-energy. We compared the performance of the proposed approach against the reconstruction of each image separately using TV regularization and the Huber edge preserving prior. Reconstruction with JTV resulted in improved contrast and spatial resolution as well as improved material decomposition.

By using JTV and coupling the low- and high-energy images, is possible to incorporate joint structural information between the 2 energies. This allows to reconstruct images from the same object where some features are missing due to the down-sampling projection data, for instance. The results presented in this work show the ability of the JTV regularization to improve sparse-view reconstruction, even when the number projection angles are 6 times less than that of a full-view setting, which allows a significant decrease the radiation dose to the patient. In comparison with TV regularization and the Huber prior, JTV leads to improved accuracy both in reconstruction and material decomposition. The reconstruction with JTV results in better contrast and spatial resolution. The results obtained with patient data or more textured phantoms corroborate the high performance of JTV compared to TV. Further analysis will involve using the proposed reconstruction framework in new CT scanner technologies, like photon-counting spectral CT, where the algorithm can leverage the joint structural similarities from an increased number of images at different energies, leading to an overall improved quantitative estimation even with a further reduction of the acquired projection angles.

CHAPTER 7

Conclusion and Perspectives

7.1 Conclusions

The present thesis proposes image reconstruction techniques for different X-ray Computed Tomography modalities. The main objective is the reduction of artifacts and the dose delivered to the patient while maintaining the image quality. We have designed new MBIR methods using data-driven approaches and machine learning. We have exploit the multi-channel joint reconstruction approaches by reconstructing the unknown images simultaneously. We solved a single combined inverse problem and exploit structural similarities between the images. We designed three multi-channel image reconstruction for: (i) Sliding motion artefact correction in CBCT utilizing sparse dictionary learning methods (Chapter 4); (ii) Dual energy CT reconstruction utilizing convolutional dictionary learning approaches (Chapter 5); (iii) Dual energy CBCT reconstruction utilizing the joint total variation technique (Chapter 6). The three approaches exploit the hypothesis that the input channels share structural similarities thus thy can "inform" each other during the reconstruction. The proposed methodologies were compared with the state-of-the-art in low dose and sparse-view CT image reconstruction methods.

- In Chapter 4 we proposed a coupled image-motion dictionary learning technique for sliding motion estimation-compensation in CBCT. The image and the DVF are simultaneously encoded in order to capture structural similarities between the image and the motion, especially the sliding motion at organs boundaries. The first step consisted of learning a set of coupled image-motion dictionaries from a training data set of DVF and image at different respiratory gates. The second step uses the trained dictionaries as sparse regularization within a MBIR which performs direct motion estimation-compensation from the projection data. We also proposed a single dictionary learning approach where the image and the motion dictionaries are trained separately, thus, they do not share the same sparse component. Both methodologies perform

well in terms of noise controlling and both estimate the motion field correctly. The resulting dictionaries learned with the coupled image-motion technique exhibit structural similarities. However, still further improvement are needed in order to capture sliding motion at organ boundaries and image denoising. Because the single dictionary learning performed better than coupled dictionary learning, we may conclude that restricting the algorithm so that both dictionaries have the same sparse vector is a very strong constraint. Another option is to restrict only the support, which means having two sparse vectors but with zeros and non-zero coefficients at the same position. Moreover, for each respiratory phase, an image dictionary and three DVF dictionaries (corresponding to the DVF along each axis x, y, z) are learned. The approach uses a significant amount of memory. Convolutional dictionary learning is one technique which could mitigate this issue. Another option is to use CNN to fine-tune the DVF. We discuss these concepts in depth in the next section.

- In Chapter 5 we proposed a multi-channel convolutional analysis operator learning framework as an extension of the CAOL method. We applied the MCAOL method to DECT. MCAOL learns convolutional dictionaries of the underlined images by jointly learning filters for the different modalities. In the DECT application, each atom not only carries individual information for each energy individually but also inter-energy information. We utilize two sparse vector coupled through using an $l_{1,0}$ sparse mixed semi norm in the MCAOL optimization problems both for training and reconstruction. We performed extensive experiments for sparse-view and low dose CT. We evaluated through many experiments the superior performance of MCAOL compared to independent optimization of each input energy. MCAOL resulted in higher quality images than state-of-the-art methods. The bias versus variance trade off showed how MCAOL archives the minimum bias and reduces the variance. The proposed methodologies can be seen as a general multi-channel framework. It can be applied to other modalities such as PET/CT, PET/MRI and SPECT/CT. Moreover, it can be extended to multi-energies or spectral CT by training energy dictionaries and combine the sparse vectors $z_{e,k}$ ($e : 1, \ldots, E$, with E is the number of energies) in a mixed norm. The reconstructed images from MCAOL can be used as follow up for image-based material decomposition in Spectral CT.

- In Chapter 6 we implemented the JTV and applied to the fast KVp switching set up in DE-CBCT. We simulated sparse-view CB projection data such that, in a single gantry rotation, one projection angle corresponds to the low energy, and the consecutive corresponds to the high energy. The main purpose was

to asses the spacial resolution improvement with joint reconstructions in alternating projection angles. We compared the reconstruction obtained with JTV against the single reconstruction utilizing TV and the Huber prior. The bias versus variance trade-off showed the out-performance of JTV, which scores lower bias and variance for different values of the regularization parameter. We also evaluated the performance of the JTV reconstructed images in material decomposition. The findings revealed that JTV regularization may enhance sparse-view reconstruction even when the number of projection angles is 6 times lower than in a full-view case, resulting in a considerable reduction in the patient's radiation exposure. We used the reconstructed images to perform image base material decomposition. The qualitative and quantitative results showed the effectiveness of JTV for this task as well as its superior performance compared to TV and Huber prior.

7.2 Perspectives

The methodologies implemented in this thesis can be considered proof-of-concept. The most remarkable continuation of the three methodologies would be the use of raw projection data. Thus, other issues such as beam hardening and scatter will be considered, especially in CBCT, where the scatter may be a significant problem. For the motion estimation compensation presented in Chapter 4, the methodology can be improved utilizing CDL (e.g. MCAOL extended to more than 2 channels). Another approach could be utilizing CNN which have shown promising results in image processing task.

7.2.1 Sliding motion correction utilizing Neural Networks

CNN have proven to be quite efficient in image processing tasks, such as segmentation, pattern recognition, classification etc. The work from Zhang et al. (2019) uses CNN to improve the accuracy of intra-lung DVF.
The sliding motion estimation presented in Chapter 4 can be driven using CNN. The general framework is presented in Figure 7.1. The motion DVF can be estimated from the CB projection data. This estimation can be encoded as the network input (Figure 7.2). The output could be the DVF with sliding motion. The CNN needs to be trained beforehand. The DVF output of the network is used to perform the motion compensation. When convergence is reached, the motion compensated CBCT image and the DVF with sliding motion at organs boundary will be obtained. We account the open problem of choosing the neural network, although we believe that U-net (Ronneberger et al. 2015) could perform the task.

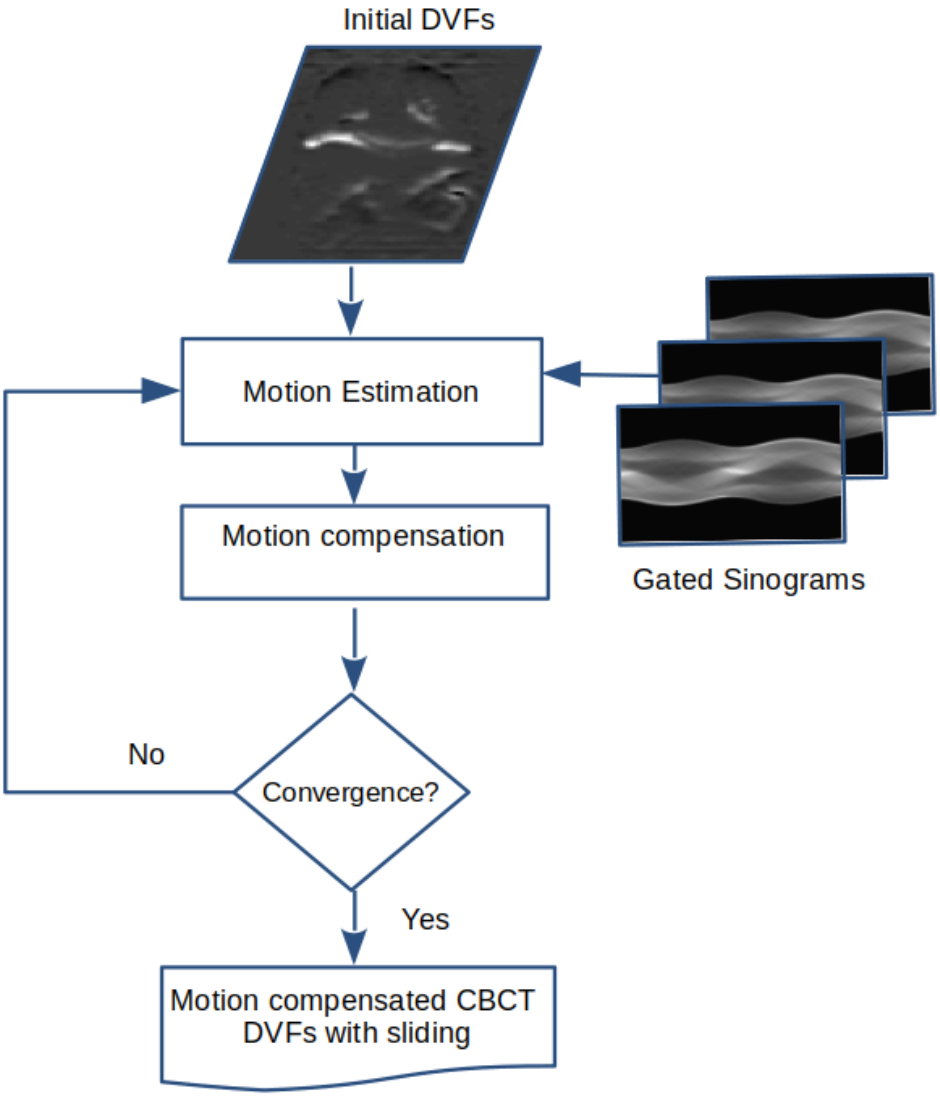


Fig. 7.1 General framework of the sliding motion estimation compensation in CBCT.

7.2.2 MCAOL extension to Spectral CT

The MCAOL algorithm can be extended to multi-energies since the $l_{1,0}$ semi-norm can be defined for a set of d-vectorized feature maps $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_d$. Each $\boldsymbol{z}_i, i = 1, \ldots, d$ is a column vector of dimension $J \times 1$. Then the joint $l_{1,0}$ semi-norm is defined as

$$\|(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_d)\|_{1,0} = \sum_{j=1}^{J} \mathbf{1}_{]0,+\infty[}\left(|z_{1,j}| + \ldots + |z_{d,j}|\right) \tag{7.1}$$

A more compact form it is obtainable using the matrix form, i.e., by collecting all column vectors $\boldsymbol{z}_i$ in matrix form as $\boldsymbol{X} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_d]$. Then the semi-norm can be written as

$$\|\boldsymbol{X}\|_{1,0} = \sum_{j=1}^{J} \|\boldsymbol{z}_{j,:}\|_0 \tag{7.2}$$

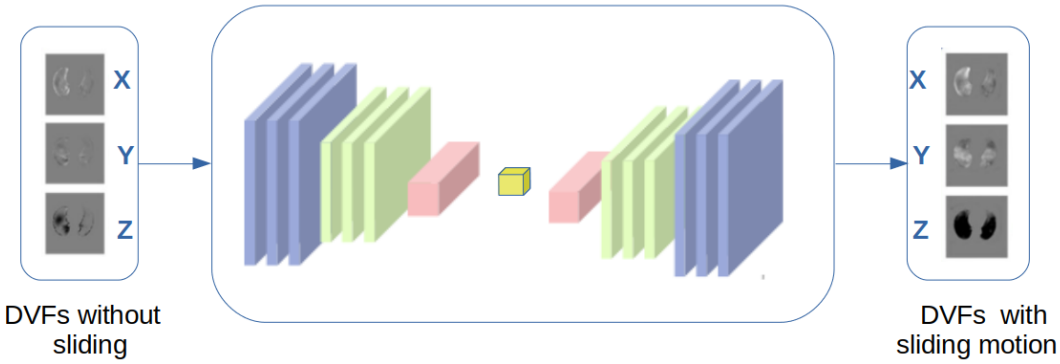


Fig. 7.2 The set up of the sliding motion correction with CNN.

where $\boldsymbol{x}_{j,:}$ is the j-th row of $\boldsymbol{X}$. While the statistical noise tends to be higher in the multi-energy case, on each sub-band the contribute of the noise is reduced since the noise is split on more energy bands. Therefore, evaluating the joint norm, i.e., non-zeros elements in the feature vectors in overlapping positions for all energies at the same time will reduce the degradation due to the increased overall noise.

REFERENCES

Aharon, M., Elad, M. and Bruckstein, A. (2006), 'K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation', *IEEE Transactions on signal processing* **54**(11), 4311–4322.

Alexandrov, O. (2021), 'Newton's method in optimization'.

Allisy-Roberts, P. J. and Williams, J. (2007), *Farr's physics for medical imaging*, Elsevier Health Sciences.

Andersen, A. H. and Kak, A. C. (1984), 'Simultaneous algebraic reconstruction technique (sart): a superior implementation of the art algorithm', *Ultrasonic imaging* **6**(1), 81–94.

Annelies van der Plas, M. r. M. U. (2016), 'X-ray/ct technique', `https://www.startradiology.com/the-basics/x-rayct-technique/index.html`.

Antun, V., Renna, F., Poon, C., Adcock, B. and Hansen, A. C. (2020), 'On instabilities of deep learning in image reconstruction and the potential costs of ai', *Proceedings of the National Academy of Sciences* **117**(48), 30088–30095.

Arjoune, Y., Kaabouch, N., El Ghazi, H. and Tamtaoui, A. (2017), Compressive sensing: Performance comparison of sparse recovery algorithms, *in* '2017 IEEE 7th annual computing and communication workshop and conference (CCWC)', IEEE, pp. 1–7.

Badea, C. and Gordon, R. (2004), 'Experiments with the nonlinear and chaotic behaviour of the multiplicative algebraic reconstruction technique (mart) algorithm for computed tomography', *Physics in Medicine and Biology* **49**(8), 1455.

Bayes, T. (1763), 'Lii. an essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfr s', *Philosophical transactions of the Royal Society of London* (53), 370–418.

Beck, A. and Teboulle, M. (2009), A fast iterative shrinkage-thresholding algorithm with application to wavelet-based image deblurring, *in* '2009 IEEE International Conference on Acoustics, Speech and Signal Processing', IEEE, pp. 693–696.

Beekman, F. J. and Kamphuis, C. (2001), 'Ordered subset reconstruction for x-ray ct', *Physics in medicine and biology* **46**(7), 1835.

Besag, J. (1986), 'On the statistical analysis of dirty pictures', *Journal of the Royal Statistical Society: Series B (Methodological)* **48**(3), 259–279.

Blumensath, T. and Davies, M. E. (2008), 'Iterative thresholding for sparse approximations', *Journal of Fourier analysis and Applications* **14**(5-6), 629–654.

Blumensath, T. and Davies, M. E. (2009), 'Iterative hard thresholding for compressed sensing', *Applied and computational harmonic analysis* **27**(3), 265–274.

Boas, F. E., Fleischmann, D. et al. (2012), 'Ct artifacts: causes and reduction techniques', *Imaging Med* **4**(2), 229–240.

Bouman, C. A. and Sauer, K. (1996), 'A unified approach to statistical tomography using coordinate descent optimization', *IEEE Transactions on image processing* **5**(3), 480–492.

Bousse, A., Bertolli, O., Atkinson, D., Arridge, S., Ourselin, S., Hutton, B. F. and Thielemans, K. (2016), 'Maximum-likelihood joint image reconstruction/motion estimation in attenuation-corrected respiratory gated PET/CT using a single attenuation map', *IEEE transactions on medical imaging* **35**(1), 217–228.

Bousse, A., Courdurier, M., Émond, É., Thielemans, K., Hutton, B. F., Irarrazaval, P. and Visvikis, D. (2019), 'Pet reconstruction with non-negativity constraint in projection space: Optimization through hypo-convergence', *IEEE transactions on medical imaging* **39**(1), 75–86.

Brehm, M., Berkus, T., Oehlhafen, M., Kunz, P. and Kachelrieß, M. (2011), Motion-compensated 4d cone-beam computed tomography, *in* '2011 IEEE Nuclear Science Symposium Conference Record', IEEE, pp. 3986–3993.

Bristow, H., Eriksson, A. and Lucey, S. (2013), Fast convolutional sparse coding, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition', pp. 391–398.

Broyden, C. G. (1970), 'The convergence of a class of double-rank minimization algorithms 1. general considerations', *IMA Journal of Applied Mathematics* **6**(1), 76–90.

Brunton, S. L. and Kutz, J. N. (2019), *Data-driven science and engineering: Machine learning, dynamical systems, and control*, Cambridge University Press.

Burrus, C., Barreto, J. and Selesnick, I. (1994), 'Iterative reweighted least-squares design of fir filters', *IEEE Transactions on Signal Processing* **42**(11), 2926–2936.

Bushberg, J. T., Seibert, J. A., Leidholdt Jr, E. M., Boone, J. M. and Goldschmidt Jr, E. J. (2003), 'The essential physics of medical imaging'.

Buzug, T. (2008), 'Computed tomography: From photon statistics to modern cone-beam ct. springer-verlag berlin heidelberg'.

Chambolle, A. and Pock, T. (2011), 'A first-order primal-dual algorithm for convex problems with applications to imaging', *Journal of mathematical imaging and vision* **40**(1), 120–145.

Chen, S. and Donoho, D. (1994), Basis pursuit, *in* 'Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers', Vol. 1, pp. 41–44 vol.1.

Chun, I. Y. (2019), Convolt: CONVolutional Operator Learning Toolbox.
URL: *https://github.com/mechatoz/convolt*

Chun, I. Y. and Fessler, J. A. (2017*a*), 'Convolutional dictionary learning: Acceleration and convergence', *IEEE Transactions on Image Processing* **27**(4), 1697–1712.

Chun, I. Y. and Fessler, J. A. (2017*b*), 'Convolutional dictionary learning: Acceleration and convergence', *IEEE Transactions on Image Processing* **27**(4), 1697–1712.

Chun, I. Y. and Fessler, J. A. (2019*a*), 'Convolutional analysis operator learning: Acceleration and convergence', *IEEE Transactions on Image Processing* **29**, 2108–2122.

Chun, I. Y. and Fessler, J. A. (2019*b*), 'Convolutional analysis operator learning: Acceleration and convergence', *IEEE Transactions on Image Processing* **29**, 2108–2122.

Coursey, C. A., Nelson, R. C., Boll, D. T., Paulson, E. K., Ho, L. M., Neville, A. M., Marin, D., Gupta, R. T. and Schindera, S. T. (2010), 'Dual-energy multidetector ct: how does it work, what can it tell us, and when can we use it in abdominopelvic imaging?', *Radiographics* **30**(4), 1037–1055.

Dai, W. and Milenkovic, O. (2009), 'Subspace pursuit for compressive sensing signal reconstruction', *IEEE Transactions on Information Theory* **55**(5), 2230–2249.

Dance, D., Christofides, S., Maidment, A., McLean, I. and Ng, K. (2014), 'Diagnostic radiology physics: A handbook for teachers and students. endorsed by: American association of physicists in medicine, asia-oceania federation of organizations for medical physics, european federation of organisations for medical physics'.

Dang, J., Yin, F.-F., You, T., Dai, C., Chen, D. and Wang, J. (2016), 'Simultaneous 4d-cbct reconstruction with sliding motion constraint', *Medical physics* **43**(10), 5453–5463.

Daubechies, I., Defrise, M. and De Mol, C. (2004), 'An iterative thresholding algorithm for linear inverse problems with a sparsity constraint', *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* **57**(11), 1413–1457.

Daubechies, I., Han, B., Ron, A. and Shen, Z. (2003), 'Framelets: Mra-based constructions of wavelet frames', *Applied and computational harmonic analysis* **14**(1), 1–46.

De Man, B. and Basu, S. (2002), Distance-driven projection and backprojection,

in '2002 IEEE Nuclear Science Symposium Conference Record', Vol. 3, IEEE, pp. 1477–1480.

De Pierro, A. R. (1995), 'A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography', *IEEE Transactions on Medical Imaging* **14**(1), 132–137.

Degraux, K., Kamilov, U. S., Boufounos, P. T. and Liu, D. (2017), Online convolutional dictionary learning for multimodal imaging, *in* '2017 IEEE International Conference on Image Processing (ICIP)', pp. 1617–1621.

Delmon, V., Rit, S., Pinho, R. and Sarrut, D. (2013), 'Registration of sliding objects using direction dependent b-splines decomposition', *Physics in Medicine and Biology* **58**(5), 1303.

Donoho, D. L. (2006), 'Compressed sensing', *IEEE Transactions on information theory* **52**(4), 1289–1306.

Donoho, D. L., Maleki, A. and Montanari, A. (2010), Message passing algorithms for compressed sensing: I. motivation and construction, *in* '2010 IEEE information theory workshop on information theory (ITW 2010, Cairo)', IEEE, pp. 1–5.

Donoho, D. L., Tsaig, Y., Drori, I. and Starck, J.-L. (2012), 'Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit', *IEEE Transactions on Information Theory* **58**(2), 1094–1121.

Draganic, A., Orovic, I. and Stankovic, S. (2017), 'On some common compressive sensing recovery algorithms and applications-review paper', *arXiv preprint arXiv:1705.05216* .

Drzezo (2016), 'Basic principles of computed tomography physics and technical considerations', `https://radiologykey.com/basic-principles-of-computed-tomography-physics-and-technical-considerations/#R13-1`.

Dumitrescu, B. and Irofti, P. (2018), *Dictionary learning algorithms and applications*, springer.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004), 'Least angle regression', *The Annals of Statistics* **32**(2), 407 – 499.
URL: *https://doi.org/10.1214/009053604000000067*

Ehrhardt, M. J., Thielemans, K., Pizarro, L., Atkinson, D., Ourselin, S., Hutton, B. F. and Arridge, S. R. (2014*a*), 'Joint reconstruction of pet-mri by exploiting structural similarity', *Inverse Problems* **31**(1), 015001.

Ehrhardt, M. J., Thielemans, K., Pizarro, L., Atkinson, D., Ourselin, S., Hutton, B. F. and Arridge, S. R. (2014*b*), 'Joint reconstruction of pet-mri by exploiting structural similarity', *Inverse Problems* **31**(1), 015001.

Elad, M. (2010), 'Sparse and redundant representations: from theory to applications in signal and image processing'.

Elbakri, I. A. and Fessler, J. A. (2002), 'Statistical image reconstruction for polyenergetic X-ray computed tomography', *IEEE Transactions on Medical Imaging* **21**(2), 89–99.

Engan, K., Aase, S. O. and Husoy, J. H. (1999), Method of optimal directions for frame design, *in* '1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258)', Vol. 5, IEEE, pp. 2443–2446.

Erdogan, H. and Fessler, J. A. (1999), 'Ordered subsets algorithms for transmission tomography', *Physics in Medicine and Biology* **44**(11), 2835–2851.
URL: *https://doi.org/10.1088/0031-9155/44/11/311*

Feldkamp, L. A., Davis, L. C. and Kress, J. W. (1984), 'Practical cone-beam algorithm', *Josa a* **1**(6), 612–619.

Fessler, J. (2000), 'Statistical image reconstruction methods for transmission tomography', *Handb Med Imaging* **2**.

Fessler, J. (2009), 'Analytical tomographic image reconstruction methods', *Image Reconstruction: Algorithms and Analysis* **66**, 67.

Figueiredo, M. A. T., Nowak, R. D. and Wright, S. J. (2007), 'Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems', *IEEE Journal of Selected Topics in Signal Processing* **1**(4), 586–597.

Forghani, R. and Mukherji, S. (2018), 'Advanced dual-energy ct applications for the evaluation of the soft tissues of the neck', *Clinical radiology* **73**(1), 70–80.

Friedman, S. N., Nguyen, N., Nelson, A. J., Granton, P. V., MacDonald, D. B., Hibbert, R., Holdsworth, D. W. and Cunningham, I. A. (2012), 'Computed tomography (ct) bone segmentation of an ancient egyptian mummy a comparison of automated and semiautomated threshold and dual-energy techniques', *Journal of computer assisted tomography* **36**(5), 616–622.

Gao, H. and Zhao, H. (2010), 'Multilevel bioluminescence tomography based on radiative transfer equation part 2: total variation and l1 data fidelity', *Optics Express* **18**(3), 2894–2912.

Garcia-Cardona, C. and Wohlberg, B. (2018*a*), 'Convolutional dictionary learning: A comparative review and new algorithms', *IEEE Transactions on Computational Imaging* **4**(3), 366–381.

Garcia-Cardona, C. and Wohlberg, B. (2018*b*), Convolutional dictionary learning for multi-channel signals, *in* '2018 52nd Asilomar Conference on Signals, Systems, and Computers', IEEE, pp. 335–342.

Garnett, R. (2020), 'A comprehensive review of dual-energy and multi-spectral computed tomography', *Clinical Imaging* .

Geman, S. (1987), 'Statistical methods for tomographic image reconstruction', *Bull. Int. Stat. Inst* **4**, 5–21.

Gilbert, P. (1972), 'Iterative methods for the three-dimensional reconstruction of an object from projections', *Journal of theoretical biology* **36**(1), 105–117.

Goel, M. (2015), Generation of ct. Slideshow presented at JR3, 9 January, 2015.

Goldfarb, D. (1970), 'A family of variable-metric methods derived by variational means', *Mathematics of computation* **24**(109), 23–26.

Goldman, L. W. (2008), 'Principles of ct: multislice ct', *Journal of nuclear medicine technology* **36**(2), 57–68.

Goo, H. W. and Goo, J. M. (2017), 'Dual-energy ct: new horizon in medical imaging', *Korean journal of radiology* **18**(4), 555–569.

Gordon, R., Bender, R. and Herman, G. T. (1970), 'Algebraic reconstruction techniques (art) for three-dimensional electron microscopy and x-ray photography', *Journal of theoretical Biology* **29**(3), 471–481.

Gorodnitsky, I. and Rao, B. (1997), 'Sparse signal reconstruction from limited data using focuss: a re-weighted minimum norm algorithm', *IEEE Transactions on Signal Processing* **45**(3), 600–616.

Hapugoda, S. (2020*a*), 'Bremsstrahlung radiation: Case courtesy of dr sachintha hapugoda, radiopaedia.org, rid: 51794', `https://radiopaedia.org/articles/bremsstrahlung-radiation`.

Hapugoda, S. (2020*b*), 'Characteristic radiation: Case courtesy of dr sachintha hapugoda, radiopaedia.org, rid: 51796', `https://radiopaedia.org/articles/characteristic-radiation?lang=gb`.

Heide, F., Heidrich, W. and Wetzstein, G. (2015), Fast and flexible convolutional sparse coding, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition', pp. 5135–5143.

Hestenes, M. R. (1969), 'Multiplier and gradient methods', *Journal of optimization theory and applications* **4**(5), 303–320.

Hines, T. (2010), 'An introduction to frame theory'.

Holt, K. M. (2014), 'Total nuclear variation and jacobian extensions of total variation for vector fields', *IEEE Transactions on Image Processing* **23**(9), 3975–3989.

Huber, P. J. (1964), 'Robust estimation of a location parameter: Annals mathematics statistics, 35'.

Ji, S., Xue, Y. and Carin, L. (2008), 'Bayesian compressive sensing', *IEEE Transactions on signal processing* **56**(6), 2346–2356.

Jia, X., Lou, Y., Dong, B., Tian, Z. and Jiang, S. (2010), 4d computed tomography reconstruction from few-projection data via temporal non-local regularization, *in* 'International Conference on Medical Image Computing and Computer-Assisted Intervention', Springer, pp. 143–150.

Johnson, T., Fink, C., Schönberg, S. O. and Reiser, M. F. (2011), *Dual energy CT in clinical practice*, Springer Science and Business Media.

Johnson, T. R. (2012), 'Dual-energy ct: general principles', *American Journal of Roentgenology* **199**(5_supplement), S3–S8.

Kainz, W., Neufeld, E., Bolch, W. E., Graff, C. G., Kim, C. H., Kuster, N., Lloyd, B., Morrison, T., Segars, P., Yeom, Y. S., Zankl, M., Xu, X. G. and Tsui, B. M. W. (2019), 'Advances in computational human phantoms and their applications in biomedical engineering—a topical review', *IEEE Transactions on Radiation and Plasma Medical Sciences* **3**(1), 1–23.

Kashyap, R. L. and Mittal, M. C. (1975), 'Picture reconstruction from projections', *IEEE Transactions on Computers* **100**(9), 915–923.

Kawahara, D., Saito, A., Ozawa, S. and Nagata, Y. (2020), 'Image synthesis with deep convolutional generative adversarial networks for material decomposition in dual-energy ct from a kilovoltage ct', *Computers in Biology and Medicine* **128**, 104111.

Kazantsev, D., Jørgensen, J. S., Andersen, M. S., Lionheart, W. R. B., Lee, P. D. and Withers, P. J. (2018), 'Joint image reconstruction method with correlative multi-channel prior for x-ray spectral computed tomography', *Inverse Problems* **34**(6), 064001.

Keren, D. and Gotlib, A. (1998), 'Denoising color images using regularization and "correlation terms"', *Journal of Visual Communication and Image Representation* **9**(4), 352–365.

Kim, K., Ye, J. C., Worstell, W., Ouyang, J., Rakvongthai, Y., El Fakhri, G. and Li, Q. (2014), 'Sparse-view spectral CT reconstruction using spectral patch-based low-rank penalty', *IEEE Transactions on Medical Imaging* **34**(3), 748–760.

Kim, K., Ye, J. C., Worstell, W., Ouyang, J., Rakvongthai, Y., Fakhri, G. E. and Li., Q. (2015), 'Sparse-view spectral CT reconstruction using spectral patch-based low-rank penalty', *IEEE Transactions on Medical Imaging* **34**(3), 748–760.

Klein, O. and Nishina, Y. (1929), 'Über die streuung von strahlung durch freie elektronen nach der neuen relativistischen quantendynamik von dirac', *Zeitschrift für Physik* **52**(11), 853–868.

Kramers, H. A. (1923), 'Xciii. on the theory of x-ray absorption and of the continuous x-ray spectrum', *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **46**(275), 836–871.

Lam, S., Gupta, R., Kelly, H., Curtin, H. D. and Forghani, R. (2015), 'Multiparametric evaluation of head and neck squamous cell carcinoma using a single-source dual-energy ct with fast kvp switching: state of the art', *Cancers* **7**(4), 2201–2216.

Lee, S.-J. (2000), Accelerated coordinate descent methods for bayesian reconstruction using ordered subsets of projection data, *in* 'Mathematical Modeling, Estimation, and Imaging', Vol. 4121, International Society for Optics and Photonics, pp. 170–181.

Lefkimmiatis, S., Roussos, A., Unser, M. and Maragos, P. (2013), Convex generalizations of total variation based on the structure tensor with applications to inverse problems, *in* 'International Conference on Scale Space and Variational Methods in Computer Vision', Springer, pp. 48–60.

Lent, A. and Censor, Y. (1991), 'The primal-dual algorithm as a constraint-set-manipulation device', *Mathematical Programming* **50**(1), 343–357.

Li, T., Koong, A. and Xing, L. (2007), 'Enhanced 4d cone-beam ct with inter-phase motion model', *Medical physics* **34**(9), 3688–3695.

Liu, J., Zhang, X., Zhang, X., Zhao, H., Gao, Y., Thomas, D., Low, D. A. and Gao, H. (2015), '5d respiratory motion model based image reconstruction algorithm for 4d cone-beam computed tomography', *Inverse Problems* **31**(11), 115007.

Liu, Y. (2018), Research status and prospect for ct imaging, *in* M. S. Ghamsari, ed., 'State of the Art in Nano-bioimaging', IntechOpen, Rijeka, chapter 5.
URL: *https://doi.org/10.5772/intechopen.73032*

Lo, S.-C. (1988), 'Strip and line path integrals with a square pixel matrix: A unified theory for computational ct projections', *IEEE transactions on medical imaging* **7**(4), 355–363.

Mallat, S. (1998), 'A wavelet tour of signal p rocessing (academic p ress, new york 1 999); i', *Daubechies Ten L ectures on Wavelets (SIAM, P hiladelphia, 1 992).[1 0]. J ouault, F. S é bille and V. de la Mota, Nucl. P hys. A* **628**(1), 998.

Mallat, S. and Zhang, Z. (1993), 'Matching pursuits with time-frequency dictionaries', *IEEE Transactions on Signal Processing* **41**(12), 3397–3415.

Marques, E. C., Maciel, N., Naviner, L., Cai, H. and Yang, J. (2018), 'A review of sparse recovery algorithms', *IEEE access* **7**, 1300–1322.

Mory, C., Janssens, G. and Rit, S. (2016), 'Motion-aware temporal regularization for improved 4d cone-beam computed tomography', *Physics in Medicine and Biology* **61**(18), 6856.

Mory, C., Sixou, B., Si-Mohamed, S., Boussel, L. and Rit, S. (2018), 'Comparison of five one-step reconstruction algorithms for spectral ct', *Physics in Medicine and Biology* **63**(23), 235001.

Needell D., V. R. (2009), 'Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit', *Foundations of Computational Mathematics* **9**, 317–334.

NIBIB (2021), 'Computed tomography (ct)', `https://www.nibib.nih.gov/science-education/science-topics/computed-tomography-ct`.

Nocedal, J. and Wright, S. (2006*a*), *Numerical optimization*, Springer Science and Business Media.

Nocedal, J. and Wright, S. J. (2006*b*), *Numerical Optimization*, 2nd edn, Springer Science and Business Media.

Nuyts, J., Beque, D., Dupont, P. and Mortelmans, L. (2002), 'A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography', *IEEE Transactions on nuclear science* **49**(1), 56–60.

Nuyts, J., De Man, B., Dupont, P., Defrise, M., Suetens, P. and Mortelmans, L. (1998), 'Iterative reconstruction for helical ct: a simulation study', *Physics in Medicine and Biology* **43**(4), 729.

O'Donnell, C. (2022), 'Radiopaedia: Hepatic congestion secondary to right heart failure', `https://radiopaedia.org/cases/hepatic-congestion-secondary-to-right-heart-failure`.

of Standards, N. I. and Technology (2001), Security requirements for cryptographic modules, Technical Report Federal Information Processing Standards Publications (FIPS PUBS) 140-2, Change Notice 2 December 03, 2002, U.S. Department of Commerce, Washington, D.C.

Orović, I., Papić, V., Ioana, C., Li, X. and Stanković, S. (2016), 'Compressive sensing in signal processing: algorithms and transform domain formulations', *Mathematical Problems in Engineering* **2016**.

O'Meara, O. T. (2013), *Introduction to quadratic forms*, Vol. 117, Springer.

Palenstijn, W., Batenburg, K. and Sijbers, J. (2011), 'Performance improvements for iterative electron tomography reconstruction using graphics processing units (gpus)', *Journal of Structural Biology* **176**(2), 250–253.
URL: *https://www.sciencedirect.com/science/article/pii/S1047847711002267*

Pati, Y., Rezaiifar, R. and Krishnaprasad, P. (1993), Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, *in* 'Proceedings of 27th Asilomar Conference on Signals, Systems and Computers', pp. 40–44 vol.1.

Philips IQon Elite Spectral CT product specifications (2018).

Poon, C. (2015), 'On the role of total variation in compressed sensing', *SIAM Journal on Imaging Sciences* **8**(1), 682–720.

Punnoose, J., Xu, J., Sisniega, A., Zbijewski, W. and Siewerdsen, J. (2016), 'spektr 3.0—a computational tool for x-ray spectrum modeling and analysis', *Medical physics* **43**(8Part1), 4711–4717.

Radon, J. (1986), 'On the determination of functions from their integral values along certain manifolds', *IEEE transactions on medical imaging* **5**(4), 170–176.

Ravishankar, S. and Bresler, Y. (2012), 'Learning sparsifying transforms', *IEEE Transactions on Signal Processing* **61**(5), 1072–1086.

Ravishankar, S., Nadakuditi, R. R. and Fessler, J. A. (2017), 'Efficient sum of outer products dictionary learning (soup-dil) and its application to inverse problems', *IEEE transactions on computational imaging* **3**(4), 694–709.

Ravishankar, S., Ye, J. C. and Fessler, J. A. (2019), 'Image reconstruction: From sparsity to data-adaptive methods and machine learning', *Proceedings of the IEEE* **108**(1), 86–109.

Release, M. A. (2018), 'of the matlab and simulink product families,(nd)'.

Richard, S., Husarik, D. B., Yadava, G., Murphy, S. N. and Samei, E. (2012), 'Towards task-based assessment of ct performance: system and object mtf across different reconstruction algorithms', *Medical physics* **39**(7Part1), 4115–4122.

Rigie, D. S. and La Rivière, P. J. (2015), 'Joint reconstruction of multi-channel, spectral ct data via constrained total nuclear variation minimization', *Physics in Medicine and Biology* **60**(5), 1741.

Rit, S., Nijkamp, J., van Herk, M. and Sonke, J.-J. (2011), 'Comparative study of respiratory motion correction techniques in cone-beam computed tomography', *Radiotherapy and Oncology* **100**(3), 356–359.

Rit, S., Sarrut, D. and Desbat, L. (2009), 'Comparison of analytic and algebraic methods for motion-compensated cone-beam ct reconstruction of the thorax', *IEEE transactions on medical imaging* **28**(10), 1513–1525.

Rit, S., Wolthaus, J. W., van Herk, M. and Sonke, J.-J. (2009), 'On-the-fly motion-compensated cone-beam ct using an a priori model of the respiratory motion', *Medical physics* **36**(6Part1), 2283–2296.

Ritschl, L., Sawall, S., Knaup, M., Hess, A. and Kachelrieß, M. (2012), 'Iterative 4d cardiac micro-ct image reconstruction using an adaptive spatio-temporal sparsity prior', *Physics in Medicine and Biology* **57**(6), 1517.

Rockafellar, R. T. (1970), 'Convex analysis princeton university press', *Princeton, NJ* .

Ronneberger, O., Fischer, P. and Brox, T. (2015), U-net: Convolutional networks for biomedical image segmentation, *in* 'International Conference on Medical image computing and computer-assisted intervention', Springer, pp. 234–241.

Rubinstein, R., Bruckstein, A. M. and Elad, M. (2010), 'Dictionaries for sparse representation modeling', *Proceedings of the IEEE* **98**(6), 1045–1057.

Rubinstein, R., Zibulevsky, M. and Elad, M. (2008), Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit, Technical report, Computer Science Department, Technion.

Rudin, L. I., Osher, S. and Fatemi, E. (1992), 'Nonlinear total variation based noise removal algorithms', *Physica D: nonlinear phenomena* **60**(1-4), 259–268.

Sajja, S., Lee, Y., Eriksson, M., Nordström, H., Sahgal, A., Hashemi, M., Mainprize, J. G. and Ruschin, M. (2020), 'Technical principles of dual-energy cone beam computed tomography and clinical applications for radiation therapy', *Advances in Radiation Oncology* **5**(1), 1–16.

Sapiro, G. and Ringach, D. L. (1996), 'Anisotropic diffusion of multivalued images with applications to color filtering', *IEEE transactions on image processing* **5**(11), 1582–1586.

Sauer, K. and Bouman, C. (1993), 'A local update strategy for iterative reconstruction from projections', *IEEE Transactions on Signal Processing* **41**(2), 534–548.

Sawatzky, A., Xu, Q., Schirra, C. O. and Anastasio, M. A. (2014), 'Proximal admm for multi-channel image reconstruction in spectral x-ray ct', *IEEE transactions on medical imaging* **33**(8), 1657–1668.

Schmidt-Richberg, A., Werner, R., Handels, H. and Ehrhardt, J. (2012), 'Estimation of slipping organ motion by registration with direction-dependent regularization', *Medical image analysis* **16**(1), 150–159.

Segars, W. P., Sturgeon, G., Mendonca, S., Grimes, J. and Tsui, B. M. (2010), '4d xcat phantom for multimodality imaging research', *Medical physics* **37**(9), 4902–4915.

Shanno, D. F. (1970), 'Conditioning of quasi-newton methods for function minimization', *Mathematics of computation* **24**(111), 647–656.

Sher, Y. (2019), 'Review of algorithms for compressive sensing of images', *arXiv preprint arXiv:1908.01642* .

Sidky, E. Y., Kao, C. and Pan, X. (2006), 'Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT', *J. X-ray Sci. Technol* **14**(2), 119–139.

Sidky, E. Y. and Pan, X. (2008), 'Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization', *Physics in Medicine and Biology* **53**(17), 4777.

Smith, N. B. and Webb, A. (2010), *Introduction to medical imaging: physics, engineering and clinical applications*, Cambridge university press.

Song, P., Weizman, L., Mota, J. F. C., Eldar, Y. C. and Rodrigues, M. R. D. (2019), 'Coupled dictionary learning for multi-contrast MRI reconstruction', *IEEE Transactions on Medical Imaging* **39**(3), 621–633.

Sonke, J.-J., Zijp, L., Remeijer, P. and Van Herk, M. (2005), 'Respiratory correlated cone beam ct', *Medical physics* **32**(4), 1176–1186.

Thibault, J.-B., Sauer, K. D., Bouman, C. A. and Hsieh, J. (2007), 'A three-dimensional statistical approach to improved image quality for multislice helical ct', *Medical physics* **34**(11), 4526–4544.

Thirion, J.-P. (1998), 'Image matching as a diffusion process: an analogy with maxwell's demons', *Medical image analysis* **2**(3), 243–260.

Tibshirani, R. (1996*a*), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society. Series B (Methodological)* **58**(1), 267–288.
URL: *http://www.jstor.org/stable/2346178*

Tibshirani, R. (1996*b*), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society: Series B (Methodological)* **58**(1), 267–288.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005), 'Sparsity and smoothness via the fused lasso', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(1), 91–108.

van Aarle, W., Palenstijn, W. J., Cant, J., Janssens, E., Bleichrodt, F., Dabravolski, A., Beenhouwer, J. D., Batenburg, K. J. and Sijbers, J. (2016), 'Fast and flexible x-ray tomography using the astra toolbox', *Opt. Express* **24**(22), 25129–25147.
URL: *http://www.osapublishing.org/oe/abstract.cfm?URI=oe-24-22-25129*

van Aarle, W., Palenstijn, W. J., De Beenhouwer, J., Altantzis, T., Bals, S., Batenburg, K. J. and Sijbers, J. (2015), 'The astra toolbox: A platform for advanced algorithm development in electron tomography', *Ultramicroscopy* **157**, 35–47.
URL: *https://www.sciencedirect.com/science/article/pii/S0304399115001060*

Wang, J. and Gu, X. (2013), 'Simultaneous motion estimation and image reconstruction (smeir) for 4d cone-beam ct', *Medical physics* **40**(10), 101912.

Wang, Z., Bovik, A. C., Sheikh, H. R. and Simoncelli, E. P. (2004), 'Image quality assessment: from error visibility to structural similarity', *IEEE transactions on image processing* **13**(4), 600–612.

Wen, J., Li, D. and Zhu, F. (2015), 'Stable recovery of sparse signals via lp-minimization', *Applied and Computational Harmonic Analysis* **38**(1), 161–176.

Werner, R., Ehrhardt, J., Schmidt-Richberg, A. and Handels, H. (2009), Validation and comparison of a biophysical modeling approach and non-linear registration for estimation of lung motion fields in thoracic 4d ct data, *in* 'Medical Imaging 2009: Image Processing', Vol. 7259, International Society for Optics and Photonics, p. 72590U.

Whiting, B., Massoumzadeh, P., Earl, O., O'Sullivan, J., Snyder, D. and Williamson, J. (2006), 'Properties of preprocessed sinogram data in X-ray CT', *Medical physics* **33**(3), 3290–3303.

Wohlberg, B. (2015), 'Efficient algorithms for convolutional sparse representations', *IEEE Transactions on Image Processing* **25**(1), 301–315.

Wohlberg, B. (2016), Boundary handling for convolutional sparse representations, *in* '2016 IEEE International Conference on Image Processing (ICIP)', IEEE, pp. 1833–1837.

Wu, W., Zhang, Y., Wang, Q., Liu, F., Chen, P. and Yu., H. (2018), 'Low-dose spectral CT reconstruction using image gradient $\ell 0$–norm and tensor dictionary', *Applied Mathematical Modelling* **63**, 538–557.

Wu, Z., Rietzel, E., Boldea, V., Sarrut, D. and Sharp, G. C. (2008), 'Evaluation of deformable registration of patient lung 4dct with subanatomical region segmentations', *Medical physics* **35**(2), 775–781.

Xu, L., Lu, C., Xu, Y. and Jia, J. (2011), Image smoothing via l_0 gradient minimization, *in* '2011 SIGGRAPH Asia Conference', pp. 1–12.

Xu, Q., Sawatzky, A., Anastasio, M. A. and Schirra, C. O. (2014), 'Sparsity-regularized image reconstruction of decomposed k-edge data in spectral ct', *Physics in Medicine and Biology* **59**(10), N65.

Xu, Q., Yu, H., Mou, X., Zhang, L., Hsieh, J. and Wang, G. (2012), 'Low-dose X-ray CT reconstruction via dictionary learning', *IEEE Transactions on Medical Imaging* **31**(9), 1682–1697.

Yang, Q., Cong, W. and Wang, G. (2017), 'Superiorization-based multi-energy CT image reconstruction', *Inverse Problems* **33**(4).

Yoon, S., Katsevich, A., Frenkel, M., Munro, P., Paysan, P., Seghers, D. and Strzelecki, A. (2019), A motion estimation and compensation algorithm for 4d cbct of the abdomen, *in* '15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine', Vol. 11072, International Society for Optics and Photonics, p. 110720E.

Yu, W., Wang, C., Nie, X., Huang, M. and Wu, L. (2017), 'Image reconstruction for few-view computed tomography based on l0 sparse regularization', *Procedia Computer Science* **107**, 808–813.

Zeng, R., Fessler, J. A. and Balter, J. M. (2005), 'Respiratory motion estimation from slowly rotating x-ray projections: Theory and simulation', *Medical physics* **32**(4), 984–991.

Zhang, H., Zeng, D., Lin, J., Zhang, H., Bian, Z., Huang, J., Gao, Y., Zhang, S., Zhang, H., Feng, Q. et al. (2017), 'Iterative reconstruction for dual energy ct with an average image-induced nonlocal means regularization', *Physics in Medicine and Biology* **62**(13), 5556.

Zhang, W., Liang, N., Wang, Z., Cai, A., Wang, L., Tang, C., Zheng, Z., Li, L., Yan, B. and Hu, G. (2020), 'Multi-energy ct reconstruction using tensor nonlocal similarity and spatial sparsity regularization', *Quantitative Imaging in Medicine and Surgery* **10**(10), 1940.

Zhang, Y., Huang, X. and Wang, J. (2019), 'Advanced 4-dimensional cone-beam computed tomography reconstruction by combining motion estimation, motion-compensated reconstruction, biomechanical modeling and deep learning', *Visual Computing for Industry, Biomedicine, and Art* **2**(1), 1–15.

Zheng, X., Ravishankar, S., Long, Y. and Fessler, J. A. (2018), 'PWLS-ULTRA: An efficient clustering and learning-based approach for low-dose 3D CT image reconstruction', *IEEE Transactions on Medical Imaging* **37**(6), 1498–1510.

Zhi, S., Jiang, B., Kachelrieß, M. and Mou, X. (2019), Artifacts reduction method in 4dcbct based on a weighted demons registration framework, *in* '15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and

Nuclear Medicine', Vol. 11072, International Society for Optics and Photonics, p. 110722R.

Zhu, C., Byrd, R. H., Lu, P. and Nocedal, J. (1997), 'Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization', *ACM Transactions on Mathematical Software (TOMS)* **23**(4), 550–560.

Zhu, Z., Wahid, K., Babyn, P., Cooper, D., Pratt, I. and Carter, Y. (2013), 'Improved compressed sensing-based algorithm for sparse-view ct image reconstruction', *Computational and mathematical methods in medicine* **2013**.

LIST OF PUBLICATIONS

Journal papers

1. Alessandro Perelli, Suxer Lazara Alfonso Garcia , Alexandre Bousse, Jean-Pierre Tasu, Nikolaos Efthimiadis, Dimitris Visvikis. **Multi-channel convolutional analysis operator learning for dual-energy CT reconstruction**. Phys Med Biol. 2022 Jan 17.
   DOI:10.1088/1361-6560/ac4c32

Conference papers, peer-reviewed

1. Suxer Alfonso Garcia, Alessandro Perelli , Alexandre Bousse, and Dimitris Visvikis. **Sparse-View Joint Reconstruction and Material Decomposition for Dual-Energy Cone-Beam Computed Tomography**. 16th International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine,Leuven, Belgium, 2021.
   https://arxiv.org/abs/2110.04143)

International Conference Abstracts

1. Suxer Alfonso Garcia, Alexandre Bousse, and Dimitris Visviki. **A coupled image-motion dictionary learning algorithm for motion estimation-compensation in cone-beam computed tomography**. Oct 31, 2020. IEEE Nuclear Science Symposium and Medical Imaging Conference. Boston, USA. *Oral Presentation*
   https://www.eventclass.org/contxt_ieee2020/online-program/session?s=M-02

2. Suxer Alfonso Garcia, Alessandro Perelli , Alexandre Bousse, and Dimitris Visvikis **Dual-Energy CT Reconstruction with Convolutional Analysis Operator Learning**. Oct 16, 2021. IEEE Nuclear Science Symposium

and Medical Imaging Conference. *Poster Presentation*
`https://www.eventclass.org/contxt_ieee2021/scientific/online-program/poster-session?s=M-16`

Appendices

Appendix A

Sparsity Promoting Norms

This section discusses why the l_0 pseudo-norm and l_1 norm promote sparsity. The l_p-norm of a vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ measures its size and can be computed as

$$\|\boldsymbol{x}\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \tag{A.1}$$

The norms can be geometrically represented as shown in figure A.1. The points (vector) on the red "star" have l_1 norm equal 1. For the Euclidean distance measure with the l_2 norm every point on the circumference is a vector with l_2 norm equal 1.

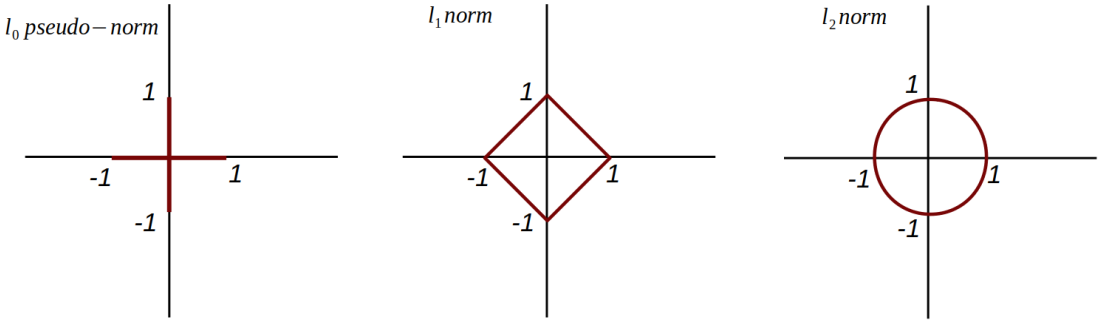


Fig. A.1 Geometric properties of l_0 pseudo-norm, l_1 and l_2 norm. Every vector on the red shape has respectively l_0 pseudo-norm, l_1 and l_2 norm equal 1. Reprint from Brunton and Kutz (2019)

Let us consider the following system of equation:

$$\boldsymbol{y} = \boldsymbol{D}\boldsymbol{z} \tag{A.2}$$

where $\boldsymbol{y}$ and $\boldsymbol{D}$ are known. This is an undetermined system of equation with multiple solutions $\boldsymbol{z}_k$. Figure A.2 depicts the solutions $\boldsymbol{z}_k$ as a blue line.

If an l_p-norm constraint in $\boldsymbol{z}$ is added, the optimization problem takes the form

$$\min_{\boldsymbol{z}} \|\boldsymbol{z}\|_p \quad \text{s.t.} \quad \boldsymbol{D}\boldsymbol{z} = \boldsymbol{y} \tag{A.3}$$

The solution is constrained to the vector with the smallest l_p norm among the possible solutions $\boldsymbol{z}_k$.

When $p = 2$ (Euclidean norm), the selected vector is the intersection point between the red circle and the blue line, as shown in A.2. This point has the two coordinates with non-zero values, thus it is not the sparsest solution.

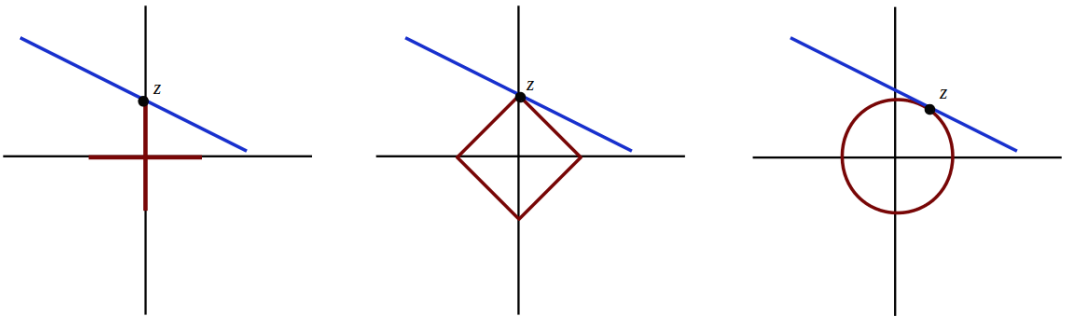


Fig. A.2 The minimum norm point on a line in different l_p norms. The red curves show the minimum-norm level sets that cross the blue line for different norms, while the blue line represents the solution set of an under-determined system of equations. According to the l_0 and l_1 norms, the minimal norm solution also corresponds to the sparsest solution, i.e., with just one active coordinate. There is no sparsity in the l_2 minimum-norm solution, as all coordinates are active. Reprint from Brunton and Kutz (2019)

When $p = 0$ the point with the smallest l_0 pseudo-norm is on the axis, which has one of its coordinates equal zero. Thus, the l_0 pseudo-norm, due to the geometrical shape, selects the sparsest solution among all the possibles values $\boldsymbol{z}_k$. That is the reason why l_0 pseudo-norm promote sparse solutions. The l_0 pseudo-norm would be the ideal case-scenario to induce sparsity. However with this norm the optimization problem becomes highly combinatorial, NP-hard and extremely difficult to solve.

One approach to relax the optimization problem is to replace the l_0 pseudo-norm by the l_1. As shown in A.2, the l_1 norm constraint also selects the sparsest solution among all the possibles values $\boldsymbol{z}_k$.

Appendix B

CAOL PWLS Objective Function

With X-ray CT high/normal exposure, a common practice is to use a quadratic approximation of equation (5.23) which leads to a Weighted Least Squares (WLS) approximation (Elbakri and Fessler 2002) based on taking the logarithm of the data

$$y_{e,i} = \log\left(\frac{S_e}{p_{e,i} - \eta_{e,i}}\right) \tag{B.1}$$

This is equivalent to observing $\boldsymbol{u}_e$ corrupted with a data-dependent Gaussian noise, $\boldsymbol{n}_e$,

$$\boldsymbol{y}_e = \boldsymbol{u}_e + \boldsymbol{n}_e = \boldsymbol{A}\boldsymbol{\mu}_e + \boldsymbol{n}_e \tag{B.2}$$

where $\boldsymbol{y}_e = [y_{e,1}, \ldots, y_{e,I}]$ and $\boldsymbol{n}_e \sim \mathcal{N}(0, \boldsymbol{W}_e^{-1})$, with inverse covariance $\boldsymbol{W}_e \in \mathbb{R}^{I \times I}$ defined as follows

$$\boldsymbol{W}_e = \operatorname{diag}\left[\frac{(p_{e,i} - \eta_{e,i})^2}{p_{e,i}}\right] \tag{B.3}$$

The NLL can then be approximated as:

$$-L(\boldsymbol{\mu}_e) \approx \text{const.} + \frac{1}{2}\Big(\boldsymbol{A}\boldsymbol{\mu}_e - \boldsymbol{y}_e\Big)^T \boldsymbol{W}_e \Big(\boldsymbol{A}\boldsymbol{\mu}_e - \boldsymbol{y}_e\Big) \tag{B.4}$$

Using the learned dictionaries $(\boldsymbol{D}_1^\star, \boldsymbol{D}_2^\star) = (\{\boldsymbol{d}_{1,k}^\star\}, \{\boldsymbol{d}_{2,k}^\star\})$ obtained as the solution the CAOL optimization (5.13) with a set of high-quality CT images, i. e. normal-dose and full views, we aim at reconstruct dual energy images independently $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) \in \mathbb{R}^J \times \mathbb{R}^J$ from the post-log measurements $(\boldsymbol{y}_1, \boldsymbol{y}_2) \in \mathbb{R}^I \times \mathbb{R}^I$. We use a model-based objective function with a penalty term for ($\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$) that can be solved through the following multi-nonconvex optimization problem

$$(\boldsymbol{\mu}_e) = \underset{\boldsymbol{\mu}_e > 0}{\arg\min} - \sum_{e=1}^{2} \gamma_e L(\boldsymbol{\mu}_e, \boldsymbol{y}_e) + R_e(\boldsymbol{\mu}_e) \tag{B.5}$$

with

$$R(\boldsymbol{\mu}_e) = \sum_{k=1}^{K} \frac{\beta_1}{2} \left\| \boldsymbol{d}_{e,k}^\star \circledast \boldsymbol{\mu}_e - \boldsymbol{z}_{e,k} \right\|_2^2 - \left\| (\boldsymbol{z}_{e,k}) \right\|_0 \tag{B.6}$$

Substituting the NLL expression in (B.4) we obtain for each energy $e = 1, 2$ the following optimization problems

$$\begin{aligned} \boldsymbol{\mu}_e = \underset{\boldsymbol{\mu}_e \geq 0}{\arg\min} \frac{\gamma_e}{2} \left\| \boldsymbol{y}_e - \boldsymbol{A}\boldsymbol{\mu}_e \right\|_{\boldsymbol{W}_e}^2 \\ + \min_{\boldsymbol{z}_{e,k}} \sum_{k=1}^{K} \frac{\beta_e}{2} \left\| \boldsymbol{d}_{e,k}^{\star} \circledast \boldsymbol{\mu}_e - \boldsymbol{z}_{e,k} \right\|_2^2 + \left\| (\boldsymbol{z}_{e,k}) \right\|_0 \end{aligned} \tag{B.7}$$

The minimization problem (B.7) is solved through a gradient-based two-block solver which alternating estimates the sparse feature images and the linear attenuation images $\{\boldsymbol{\mu}_e : e = 1, 2\}$ as in Chun and Fessler (2019*b*).

Titre: Reconstruction d'images Tomographiques Multicanaux en Exploitant les Structures Similaire des Images.

Mots clés : Tomographie à rayons X, Apprentissage par dictionnaires, Reconstruction d'images, Méthodes itératives, Optimisation.

Résumé : La technique de reconstruction multicanal est une méthode adaptée à la reconstruction multimodale en imagerie médicale. Dans la technique, les images inconnues sont reconstruites simultanément en résolvant un seul problème inverse et en exploitant les similitudes structurelles entre les images. L'hypothèse sous-jacente à cette approche est que les modalités de l'image s'informent mutuellement lors de la reconstruction permettant la réduction des artefacts et l'amélioration de la qualité de l'image. La présente thèse développe trois modèles de reconstruction d'images multicanaux. La première méthodologie consiste un algorithme d'apprentissage du dictionnaire couplé image et mouvement pour l'estimation et la compensation du mouvement en Cone Beam Computed Tomography (CBCT). La deuxième approche propose un apprentissage d'opérateur d'analyse convolutive multicanal (MCAOL) pour la reconstruction CT bi-énergie (DECT). Dans la troisième technique, nous nous concentrons sur la configuration d'acquisition de commutation KVp rapide à source unique à vue sparse dans le CBCT à double énergie pour réduire la dose totale délivrée lors d'une acquisition CT. Les méthodologies proposées ont été comparées aux algorithmes de reconstruction de pointe actuels pour la tomographie à faible dose et à vue sparse. Les trois méthodologies surpassent les méthodes utilisées par comparaison. Ils ont été publiés dans des revues à comité de lecture et des conférences internationales.

Title: Multi-channel Computed Tomographic Image Reconstruction by Exploiting Structural Similarities

Keywords: X-ray Computed Tomography, Dictionary Learning, Image Reconstruction, Iterative Methods, Optimization.

Abstract: The multi-channel joint reconstruction technique is a highly suited method for multi-modal medical imaging reconstruction. In the technique, the unknown images are reconstructed simultaneously by solving a single combined inverse problem and exploiting structural similarities between the images. The hypothesis behind this approach is that the image modalities inform each other during the reconstruction allowing artifact reduction and image quality enhancement. The present thesis develops three image reconstruction models for multi-channel image reconstruction. The first methodology consists of a Coupled Image-Motion Dictionary Learning algorithm for Motion Estimation Compensation in Cone-Beam Computed Tomography (CBCT). The second approach proposes a Multi-channel Convolutional Analysis Operator Learning (MCAOL) for Dual-Energy CT (DECT) Reconstruction. In the third technique, we focus on the sparse view single source fast KVp switching acquisition set-up in Dual Energy CBCT to reduce the total dose delivered during a CT acquisition. The proposed methodologies were compared with the current state-of-the-art reconstruction algorithms for sparse-view and low-dose CT. The three methodologies outperform the methods used for comparison. They were published in peer-reviewed journals and international conferences.