TOF-PET IMAGING WITHIN THE FRAMEWORK OF SPARSE RECONSTRUCTION

A Thesis

by

DAPENG LAO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

March 2012

Major Subject: Health Physics

TOF-PET Imaging within the Framework of Sparse Reconstruction

TOF-PET IMAGING WITHIN THE FRAMEWORK OF SPARSE RECONSTRUCTION

A Thesis

by

DAPENG LAO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Gamal Akabani
Committee Members,	Mark W. Lenox
	Lesie A. Braby
Head of Department,	Yassin A. Hassan

March 2012

Major Subject: Health Physics

ABSTRACT

TOF-PET Imaging within the Framework of Sparse Reconstruction. (May 2012)

Dapeng Lao, B.A., Tsinghua University

Chair of Advisory Committee: Dr. Gamal Akabani

Recently, the limited-angle TOF-PET system has become an active topic mainly due to the considerable reduction of hardware cost and potential applicability for performing needle biopsy on patients while in the scanner. However, this kind of measurement configurations oftentimes suffers from the deteriorated reconstructed images, because insufficient data are observed. The established theory of Compressed Sensing (CS) provides a potential framework for attacking this problem. CS claims that the imaged object can be faithfully recovered from highly underdetermined observations, provided that it can be sparse in some transformed domain.

In here a first attempt was made in applying the CS framework to TOF-PET imaging for two undersampling configurations. First, to deal with undersampling TOF-PET imaging, an efficient sparsity-promoted algorithm was developed for combined regularizations of p-TV and l_1-norm, where it was found that (a) it is capable of providing better reconstruction than the traditional EM algorithm, and (b) the 0.5-TV regularization was significantly superior to the regularizations of 0-TV and 1-TV, which are widely investigated in the open literature. Second, a general framework was proposed for sparsity-promoted ART, where accelerated techniques of multi-step and

order-set were simultaneously used. From the results, it was observed that the accelerated sparsity-promoted ART method was capable of providing better reconstruction than traditional ART. Finally, a relationship was established between the number of detectors (or the range of angle) and TOF time resolution, which provided an empirical guidance for designing novel low-cost TOF-PET systems while ensuring good reconstruction quality.

DEDICATION

To my parents, my brothers and sisters, and my girlfriend

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Gamal Akabani, for his support and guidance. I also want to acknowledge the help I received from Dr. Lenox, who gave me many useful suggestions on my thesis.

I want extend my gratitude to all the faculty and staff of the Nuclear Engineering Department. I also want to extend my gratitude to my classmates and friends in this university.

Finally, thanks to my family and girlfriend for their love and encouragement.

NOMENCLATURE

PET	Positron Emission Tomography
TOF	Time-of-Flight
LOR	Line of Response
ART	Algebraic Reconstruction Technique
EM	Expectation Maximization
CS	Compressed Sampling
A	Projection Matrix
f	Images to be Recovered
y	Observed Data
TV	Total Variation
Φ	Sparse Transformed Basis
τ	Time Resolution of TOF-PET System
OS	Order-set

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Introduction to Positron Emission Tomography (PET)

Positron Emission Tomography (PET) is an imaging modality where the distribution of the positron-emitting radionuclide inside the body can be imaged based on the detection of the photons emitted from positron annihilation [1]. In contrast to computed tomography (CT) and magnetic resonance imaging (MRI) that provide detailed anatomical or morphological information, PET shows great superiority in monitoring the functional metabolism in normal and neoplastic tissues [2]. This advantage of indicating metabolic changes in tumors enables their early detection, and has a direct impact on patient treatment and prognosis. Nowadays, PET functional imaging is widely used in clinically oncology, cardiology, and neurology for diagnosing, re-staging, and theragnostics [1, 2].

Radiopharmaceuticals labeled with positron-emitting radionuclides are introduced into the body via intravenous injection. When the radionuclide decays, it emits a positron that will annihilate with an electron after traveling a very short distance. As a consequence, two 511 keV photons are produced traveling in nearly opposite directions. If these two photons are detected by a pair of detectors within a short timing window, a coincidence event is recorded that forms a line of response (LOR) [2, 3]. The choice of positron-emitting radionuclide is a key issue for PET imaging. The most widely used radiopharmaceutical is Fludeoxyglucose (^{18}FDG) [4].

This thesis follows the style of *IEEE Transactions on Medical Imaging*.

A specific feature of current modern current PET systems is time-of-flight (TOF). The applicability of TOF information in PET imaging was initially realized by Brownell et al., for the purpose of whole-body imaging [5]. TOF is the difference between the arrival times at the detectors of two photons produced by a single annihilation event. It is possible to determine the precise location at which the positron annihilation occurred if the time resolution is high enough. The fundamental improvement rendered by TOF is an increase in signal-to-noise ratio (SNR); therefore, TOF can help improve the quality of reconstructed images [5]. The first generation of TOF-PET system was made in the 1980s. It brought resurgence of PET with the development of fast detector system. The current time resolution for TOF-PET is approximately 500 *ps* [6, 7].

In the past several decades, some important developments in PET imaging have been made resulting in high quality images, which promoted its application in clinical diagnosis. These major achievements can be summarized into the following aspects, in particular:

(a) the extension from a 2-dimensional system with septa to a 3-dimensional one with larger axial field of view;

(b) the development of fast scintillators and electronics which made TOF information available yielding a higher spatial resolution and better image reconstruction;

(c) the transition from analytical filtered back-projection (FBP) reconstruction algorithms to fully 3D iterative algorithms with data corrections;

(d) the combination with CT that is capable of providing the fusion of functional and anatomical images for clinical review, and the CT can provide attenuation correction for PET; and

(e) the newly emerging technology of integrating PET and MRI [6,8].

Recently, the development of limited-angle TOF-PET system has been proposed for breast-dedicated imaging [9]. And it has become an active area of research due to its distinct advantages, in particular, the capacity of simultaneously performing a needle biopsy on patients while in the scanner, etc. More importantly, the limited-angle configuration can significantly reduce the cost of data sampling. However, this kind of PET system only can be applicable under the condition that TOF time resolution is high enough. However, artifacts will inevitably appear in the reconstructed images because of the incomplete coverage of the imaged object by a partial PET ring. From a mathematical point of view, this reconstruction scheme is a serious ill-posed inverse problem due to the deficient observations compared with unknowns. Working under the undersampling configurations, the traditional PET reconstruction algorithms usually fail to provide good reconstruction results.

Fortunately, the well-known compressed sampling (CS) developed by Candes et al. in 2005 shows that if a signal is itself sparse or has sparse representation in some transformed domain, it could faithfully be recovered from its undersampling measurements [10]. One fundamental ingredient of CS is the sparsity or inherent low-dimensional structure of the signal. Actually, lots of natural signals are sparse or compressible. The use of the sparse property of the signal can be traced back to the work

of Logan in 1965 [11], Santosa and Symes in 1986 [12], and Donoho and Stark in 1989 [13]. However, it is generally agreed that the foundation of today's CS theory was laid by the three publications [10, 14, 15] in 2005. Since the pioneer work conducted by Candes, Romberg and Tao [10, 14-16], there has been intensive research in the field of compressed sampling.

In the context of medical imaging, it is well known that a large number of medical images are shown to be piecewise constant, and they also have sparse representations in some transform domains such as discrete cosine transformation (DCT) and wavelet [17], or some trained over-complete dictionary [18-21]. Encouraged by these observations, it is expected that CS can be employed to improve the reconstruction of medical imaging, especially for the cases with deficient observation. Actually, since the invention of CS, the sparsity-promoted solution [10, 14] has been widely used in medical imaging, such as MRI [20-22], CT [23, 24], and PET [25-27].

More specifically, regarding the application of sparse reconstruction in PET imaging, research has been focused on the development of more efficient reconstruction algorithms [25-27]. To my best knowledge, the initial work within the context of the sparsity-promoted PET reconstruction can be traced to that made by Harmany et al [25] where photon-limited collections were modeled by a Poisson distribution and the choice of wavelet-based sparse transformation was discussed. Furthermore, to correct the estimation bias caused by the mandatory introduction of the sparsity-promoted term into the data fidelity, Wang and Qi proposed a two-step reconstruction strategy [26], where the sparsity-promoted algorithm was first implemented to determine the basis

components with significant coefficients, and then the final solution was calculated by solving the reduced-dimensional least-square problem. In [27], the authors briefly introduce the application of the first-order primal-dual algorithm to deal with real PET data. In summary, these studies showed that CS was capable of breaking the bottleneck of PET imaging with fewer observations than unknowns.

1.2 Mathematical Model of PET Imaging

This subsection illustrates the model of PET imaging within the two-dimensional context, as shown in Figure 1-1. The relationship between the object of interest $f(x,y)$ and PET observed data $p(\theta,s)$ can be formally represented via the well-known Radon transformation [28], namely,

$$p(\theta,s)=\iint f(x,y)\delta(s-(y\cos\theta-x\sin\theta))dxdy\,. \tag{1.1}$$

Eq. (1.1) shows that the observed data can be understood as the integrals of the object along the lines-of-response (LOR) connecting the pairs of detectors. Conventionally, these collected data (i.e., projections) are stored in the form of a sinogram.

To facilitate the numerical implementation, Eq. (1.1) can be cast in the discretized form as:

$$\mathbf{y}=\mathbf{A}\mathbf{f}+\mathbf{n}\,, \tag{1.2}$$

where the vector $\mathbf{y}\in \mathrm{R}^{M}$ denotes the observed data, $\mathbf{f}\in \mathrm{R}^{N}$ represents the unknown object of interest, and the vector of $\mathbf{n}$ is used to take into consideration the noise, the discrete error, and other possible errors. The projection matrix $\mathbf{A}$ characterizes the

whole imaging system and it represents the probability $p(i,j)$ that an emission from j-th pixel is detected by i-th pair of detectors on PET ring [29].

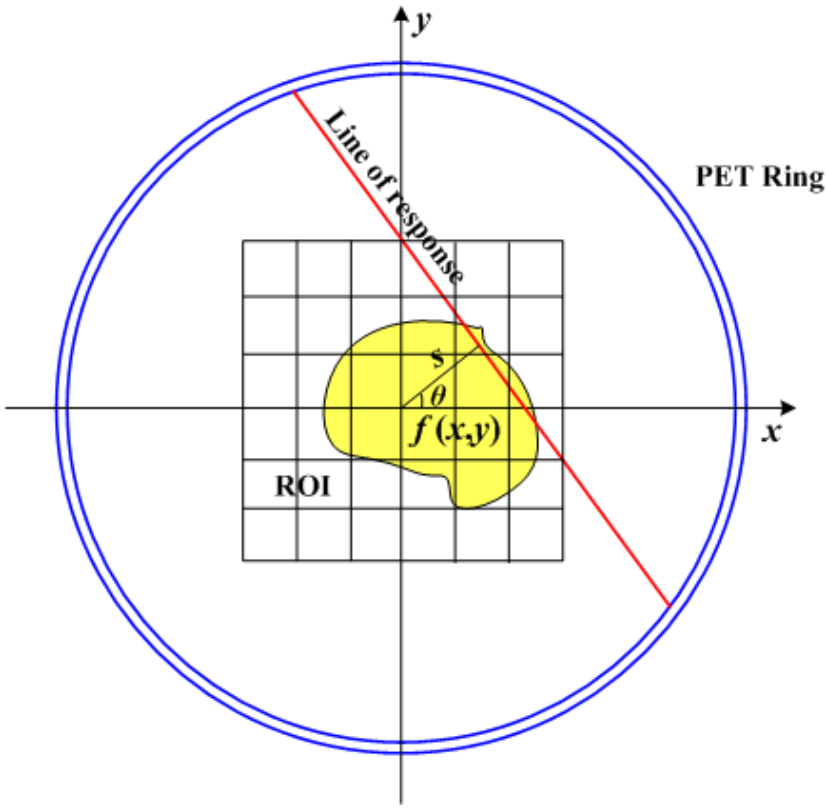


Figure 1-1. Diagram describing the Radon transformation where the red line denotes the LOR (line of response) and the ROI (region of interest) is discretized into $n \times n$ pixels.

With the introduction of time-of-flight, a TOF-PET observation can be approximately represented by a convolution of the object f with a TOF kernel function h along the LOR, in particular [5, 30, 31],

$$y(s,\theta,t) = \int_{-\infty}^{\infty} f(s\cos\theta - l\sin\theta, s\sin\theta + l\cos\theta)h(t-l)dl\,. \tag{1.3}$$

Here, the variables s and θ are the radial and angular coordinates, respectively, and l is the integration variable along the LOR, while t represents the TOF variable. The TOF

kernel h is often considered to be time-shift invariant and modeled as a Gaussian function with the variance σ^2 related with the time resolution τ of the system through the equation $\sigma = \left((c\tau / 2) / 2.35 \right)$ [5, 30 ,31]. From Eq. (1.3) we can observe that if τ is very large, the TOF PET almost returns to a standard PET imaging; in contrast, if system works with ideally small τ, the TOF PET imaging can be obtained directly from the observed data that requires no reconstruction procedure.

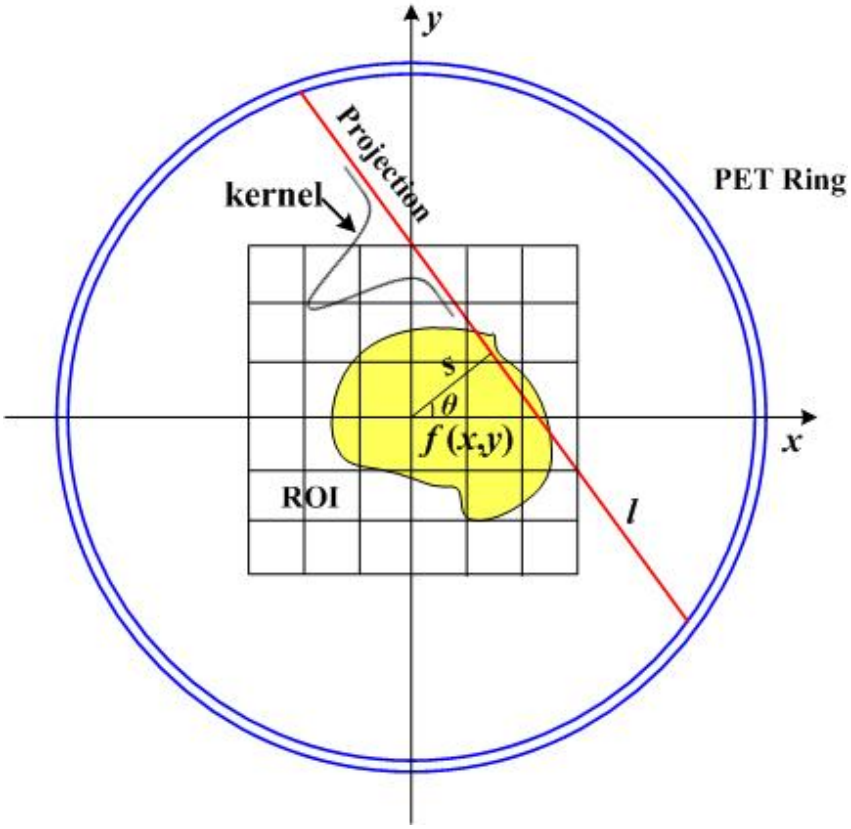


Figure 1-2. Sketch map of TOF-PET imaging system where the red line denotes the LOR (line of response). The curve marked as kernel is the TOF Gaussian function and the ROI (region of interest) is discretized into $n \times n$ pixels.

Accordingly, Eq. (1.3) can also be casted into a series of compact forms,

$$\mathbf{y}_t = \mathbf{A}\mathbf{H}_t\mathbf{f} = \mathbf{A}_t\mathbf{f},$$
$$t = \{-T\Delta t, -(T-1)\Delta t, ..., (T-1)\Delta t, T\Delta t\} \tag{1.4}$$

where Δt is known as TOF bin, and the number of time bins is (2T+1). **A** is the projection matrix of PET system defined previous [29]. In Eq. (1.4), the M-dimensional vector $\mathbf{y}_t \in R^M$ represents the observed data $y(s,\theta,t)$ at time t. The term $\mathbf{H}_t\mathbf{f}$ accounts for the convolution of the Gaussian kernel and the object, where $\mathbf{H}_t$ is the Topletiz matrix generated by h. Note that without otherwise a specific claim, the notation of **f** is a N-dimensional vector, stacked from the two-dimensional image **f** with size of $n \times n$**.**

1.3 Reconstruction Algorithms for PET Imaging

1.3.1 Analytical algorithms

The aim of PET imaging is to recover $f(x,y)$ from the available observation $p(\theta,s)$ through reconstruction algorithms. The estimation of $f(x,y)$ can be directly achieved by implementing the inverse Radon transformation [28], which can be expressed by the following one-dimensional integral of the observed data $p(\theta,s)$

$$\hat{f}(x,y) = \int_0^{\pi} p(\theta, -x\sin\theta + y\cos\theta)d\theta. \tag{1.5}$$

The reconstruction through the above analytical solution is a popular approach due to its relatively low computational cost. However, this straightforward analytical approach seriously suffers from the degraded image quality. Actually, the estimation $\hat{f}_{BP}(x,y)$ can be more exactly expressed by the convolution of the object $f(x,y)$ with a kernel function $R(x,y)$, in particular,

$$\hat{f}(x,y) = f(x,y) \otimes R(x,y), \tag{1.6}$$

where $R(x,y) = 1/\sqrt{x^2+y^2}$ describes the PET system response. One method used to reduce the effect brought by $R(x,y)$ is to incorporate a filter into the reconstruction process; consequently, the reconstruction approach becomes the well-known filtered back-projection (FBP) algorithm [32], which can be formally expressed as

$$f(x,y) = \int_0^{\pi} \mathcal{F}_1^{-1}\left\{|\omega_s| \mathcal{F}\{p(\theta,s)\}\right\} d\theta, \tag{1.7}$$

where $F\{\cdot\}$ and $F^{-1}\{\cdot\}$ denote respectively the forward and inverse Fourier transformations with respect to the variable s and ω_s, where $|\omega_s|$ is normally called a ramp filter. The FBP approach can be roughly considered to be the following four-step strategy: (a) calculating the Fourier transformation of $p(\theta,s)$ with respect to s; (b) filtering $\mathcal{F}\{p(\theta,s)\}$ via the ramp filter $|\omega_s|$; (c) computing the inverse Fourier transformation of $|\omega_s|\mathcal{F}\{p(\theta,s)\}$ with respect to ω_s; and (d) carrying out the integral of $\mathcal{F}_1^{-1}\left\{|\omega_s| \mathcal{F}\{p(\theta,s)\}\right\}$ with respect to the view angle θ.

Compared to the direct back-projection algorithm (i.e., inverse Radon transformation), it does not result in considerable extra computational cost, as the operations of forward and inverse Fourier transformation introduced in FBP can be efficiently implemented by the use of the well-known fast Fourier transformation (FFT) technique. In addition, it has been theoretically proved that the object $f(x,y)$ can be reconstructed perfectly when the sufficient noise-free data is available.

Nevertheless, we would like to specifically emphasize that the FBP is noise sensitive, as the ramp filter of FBP will amplify the noise or error; consequently, it will substantially distort the reconstructed images. To attack this problem, the strategy used commonly is to incorporate a suitable window function (such as, cosine function, Hanning function, etc. [33, 34]) to suppress the noise.

Encouraged by the improvements on sensitivity and signal-to-noise ratio, PET data acquisition and reconstruction methods have been successfully transited from 2-dimensional to 3-dimensional systems. However, regarding this transition, two major problems had to be overcome: (1) the spatially varying nature; and (2) the huge computation cost.

The spatially varying nature of the 3D data sets makes the reconstruction more challenging [35]. One popular approach to avoid the data variance is the 3-dimensional re-projection (3DRP) algorithm developed by Kinahan and Rogers [36]. The working principle of 3DRP is to conjecture the unmeasured data through forward projections, which are based on the initial estimates of every transverse slice obtained through the 2-dimensional FBP reconstruction.

However, the 3DRP algorithm is much more time consuming than the 2-dimensional slice-by-slice FBP algorithm due to the considerable amount of increasing LORs. To accelerate the 3DRP algorithm, a re-binning technique has been developed. The re-binning operation is to sort (rebin) the 3-dimensional observed data into many sets of 2-dimensional data prior to the implementation of the FBP algorithm. It can be considered as a pre-procedure to the reconstruction [37]. Three commonly used re-

binning algorithms have been introduced, (1) the single-slice re-binning (SSRB) algorithm, (2) the multi-slice re-binning (MSRB) algorithm, and (3) the Fourier re-binning (FORE) algorithm. Among these three, the SSRB algorithm is considered to be the simplest one; the MSRB algorithm is more accurate than the SSRB algorithm, but less stable in the presence of noise, and the FORE algorithm is considered to be the most accurate [38-40]. In summary, the family of FBP algorithms is computationally attractive in terms of practical consideration, and has played an important role in the clinic use.

1.3.2 Iterative algorithms

Recalling Eq. (1.2), the basic purpose of reconstruction is to recover **f** from **y** given the knowledge of **A**. The iterative approach is well designed to address this kind of finite-dimensional inverse problem. Compared to analytical approaches, iterative approaches have two distinct advantages: (a) it is relatively easy to incorporate prior knowledge into the reconstruction; and (b) they are suitable for the undersampling imaging problem with some specific modifications.

To date, a large number of iterative algorithms have been developed, and they can be divided into two major groups: (a) the algebraic reconstruction algorithms, and (b) the statistical reconstruction algorithms. In here, a brief description of algebraic reconstruction algorithms is given followed by an extensive discussion of the statistical approaches using Expectation Maximization (EM).

1.3.2.1 Algebraic algorithms

In this subsection, the basic algebraic reconstruction technique (ART), which was widely used in the early stages of PET imaging, will be briefly discussed. Simply, the iterative formula of basic ART can be written as following [41]:

$$\mathbf{f}^{(new)} = \mathbf{f}^{(old)} - \frac{\mathbf{y}_i - \left\langle \mathbf{a}_i, \mathbf{f}^{(old)} \right\rangle}{\left\| \mathbf{a}_i \right\|_2^2} \mathbf{a}_i, \tag{1.8}$$

where $i = k \bmod M + 1$, $\mathbf{a}_i$ is the i-th row of $\mathbf{A}$, $\mathbf{y}_i$ is the i-th element of $\mathbf{y}$. There are several variants of ART, such as (a) simultaneous iterative reconstruction technique (SART), and (b) multiplicative simultaneous algebraic reconstruction technique (MART) [42]. As for these variants of ART, a more detailed description of these variants is given in [41, 42].

In the terminology of mathematics, ART also can be regarded as incremental gradient [43]. In the context of convex optimization, the solution to Eq. (1.2) is usually cast as

$$\hat{\mathbf{f}} = \min_{\mathbf{x} \geq \mathbf{0}} \left\{ \mathcal{J}(\mathbf{f}) = \frac{1}{2} \left\| \mathbf{A}\mathbf{f} - \mathbf{b} \right\|_2^2 \right\}. \tag{1.9}$$

The iterative formula of the gradient-based approach to solving Eq. (1.9) at the k-th iteration is

$$\mathbf{f}^{(k)} \Leftarrow \mathbf{f}^{(k-1)} - \lambda_k \cdot \mathbf{d}^{(k-1)}, \tag{1.10}$$

where $\mathbf{d}^{(k-1)} = \mathbf{A}^T(\mathbf{A}\mathbf{f}^{(k-1)} - \mathbf{b})$ is the gradient of $\mathcal{J}(\mathbf{f}) = \frac{1}{2} \left\| \mathbf{A}\mathbf{f} - \mathbf{b} \right\|_2^2$ with respect to $\mathbf{f}^{(k-1)}$. The step size λ_k is determined through the standard linear search technique given as

$$\lambda_k = \frac{\left\|\mathbf{d}^{(k-1)}\right\|_2^2}{\left\|\mathbf{A}\mathbf{d}^{(k-1)}\right\|_2^2}. \tag{1.11}$$

For the consideration of computation simplicity, λ_k is usually approximated as $\lambda_k = 1/\|\mathbf{A}\|_2^2$.

Comparing Eq. (1.8) and Eq. (1.10), one can immediately see the difference between ART and the traditional steep descent method, in particular only one measurement is used each iteration in ART while all measurements are used in each iteration in the steep descent approach. Similar to the steep descent approach, ART often converges slowly for ill-conditional problems, especially for large-scale cases. To overcome this problem, some pre-conditioners, i.e. diagonal scaling matrices, Toeplitz matrices, etc. should be incorporated [45].

1.3.2.2 Statistical algorithms

This subsection introduces the statistical reconstruction algorithms, focusing on the best-known Expectation Maximization (EM) algorithm. The EM algorithm was developed by Dempster, Laird, and Rubin in 1977 [46], and gradually attracted great attention due to two distinct advantages over the analytical methods and algebraic iterative methods: (a) it is essentially based on the maximum likelihood or maximum a posterior (MAP), which is capable of offering unbiased, minimum variance estimates when sufficient measurements are available; (b) it remains a basis for some popular statistical reconstruction methods. The EM iteration is performed through alternations between an expectation (*E*) step and a maximization (*M*) step. More explicitly, the

expectation (E) step computes the expectation of the log-likelihood using the current estimate of the parameters, while the maximization (M) step computes the parameters maximizing the expected log-likelihood. These parameter-estimates are then used to determine the probability distribution of the latent variables for the next E-step.

We can consider the maximum likelihood estimator:

$$\hat{\mathbf{f}} = \arg\max_{\mathbf{f}} L(\mathbf{f}), \tag{1.12}$$

where

$$L(\mathbf{f}) = \log p(\mathbf{y} \mid \mathbf{f}), \tag{1.13}$$

and recall the assumption made in PET imaging that the observed data $\mathbf{y}$ conditional on $\mathbf{f}$ obeys the Poisson distribution:

$$p(\mathbf{y} \mid \mathbf{f}) = \prod_{i=1}^{m} \frac{\exp(-\sum_{j=1}^{n} p(i,j) f_j)(\sum_{j=1}^{n} p(i,j) f_j)^{y_i}}{(y_i)!}, \tag{1.14}$$

where f_j is the j-th element of vector $\mathbf{f}$, y_i is the i-th element of observations. Normally, the E-step and M-step of EM algorithm can be expressed as

$$E-\text{step}: z_{ij} = \frac{y_i p(i,j) f_j}{\sum_{j=1}^{n} p(i,j) f_j} \tag{1.15}$$

and

$$M-\text{step}: \tilde{f}_j = \frac{\sum_{i=1}^{m} z(i,j)}{\sum_{i=1}^{m} p(i,j)}. \tag{1.16}$$

If the above two-step iterative operations are integrated, the iterative formula frequently used in PET imaging can be obtained as [47, 48]

$$\tilde{f}_j^{new} = \frac{\tilde{f}_j^{old} \sum_{i=1}^{m} y_i p(i,j) \Big/ (\sum_{j=1}^{n} p(i,j) f_j)}{\sum_{i=1}^{m} p(i,j)}. \tag{1.17}$$

As pointed out previously, EM approach can provide unbiased and minimum variance estimation. However, it is extremely sensitive to noise. Therefore, it is necessary to take some steps to address this problem. A mainstream strategy is the Bayesian/penalized method, which will be discussed as follows [49-52].

Different from Eq. (1.12), which is based on maximum likelihood, the Bayesian strategy can be used to find the solution of maximum a posterior (MAP) problem. We consider the MAP estimator

$$\hat{\mathbf{f}} = \arg\max_{\mathbf{f}} L(\mathbf{f}), \tag{1.18}$$

where

$$L(\mathbf{f}) = \log p(\mathbf{f} \mid \mathbf{y}) \propto \log p(\mathbf{y} \mid \mathbf{f}) + \log p(\mathbf{f}). \tag{1.19}$$

In Eq. (1.18), the prior term of $p(\mathbf{f}) \propto \exp(-\beta V(\mathbf{f}))$ has been taken into account, which renders the benefit that the reconstructed image quality may be greatly improved if proper prior is used. Several successful priors have been proposed by different researchers [52, 53].

The update iterative formula turns to be

$$\tilde{f}_j^{new} = \frac{\tilde{f}_j^{old} \sum_{i=1}^{m} y_i p(i,j) \Big/ (\sum_{j=1}^{n} p(i,j) f_j)}{\sum_{i=1}^{m} p(i,j) + \beta \frac{\partial}{\partial f_j} V(\mathbf{f}) | \tilde{f}_j^{old}}. \tag{1.20}$$

Comparing Eq. (1.17) and Eq. (1.20), we can readily observe that there is an extra term $\beta \frac{\partial}{\partial f_j} V(\mathbf{f})$ resulting from the introduction of the prior $p(\mathbf{f})$. Obviously, this term plays the role of a filter that suppresses noise, especially for the case where $\sum_j p(i,j) \to 0$.

Regarding the convergence of the EM algorithm, a strict theoretical analysis on stable convergence has been made. Normally, EM-based algorithms require approximately 20 to 50 iterations to reach an acceptable solution. Though capable of providing much more accurate reconstruction than the FBP algorithm, EM algorithms require the implementations of one forward projection and one back projection in each iteration; consequently, it is sometimes computational intractable for clinical use. To make the EM algorithms more practical, Hudson and Larkin [54] proposed in 1994 the order-set Expectation Maximization (OSEM) algorithm, which can efficiently reduce the computational cost.

Assuming that the observed data is grouped into *n* subsets prior to the reconstruction operation and $S_i (i = 1,2,...,n)$ denote the ordered subsets of the observed data, the iterative formula of OS-EM is

$$\tilde{f}_j^{new} = \frac{\tilde{f}_j^{old} \sum_{t \in S_i} y_t p(t,j) \Big/ (\sum_{j=1}^{n} p(t,j) f_j)}{\sum_{t \in S_i} p(t,j)} \tag{1.21}$$

for the observed data $t \in S_i$. Different from the standard EM algorithm, only one subset is used in the update formula for OSEM each iteration, using part of rows of the projection matrix. It is appropriate to point out that when the number of subsets equals one, it turns back to the standard EM algorithm. Many numerical simulations have shown that with suitable choice of *n*, the OSEM can provide a better convergence at the early stage of iteration than the standard EM algorithm. However, at the late stage of iteration, the use of order-set technique will result in oscillation. It should be very careful of selecting step size to ensure stable convergence [40, 54, 55].

2. BASICS OF SPARSE RECONSTRUCTION

2.1 Ill-posed Linear Problem and Its Regularization

Recalling that the PET imaging problem can be mathematically cast into the following linear equation, i.e.,

$$\mathbf{y} = \mathbf{A}\mathbf{f} \tag{2.1}$$

where $\mathbf{A} \in R^{M \times N}$ is the projection matrix, which relates the object denoted by $\mathbf{f} \in R^{N}$ and the observed data denoted by $\mathbf{y} \in R^{M}$. Provided that $\mathbf{A}$ is well conditioned, which requires that the independent observations should be sufficient comparing to unknowns; thus, the solution to Eq. (2.1) can be straightforward achieved as

$$\mathbf{f} = \mathbf{A}^{-1}\mathbf{y}. \tag{2.2}$$

However, in the context of undersampling PET imaging, the problem (2.1) turns to be ill-posed. It is usually challenging due to the infinite number of non-meaningful solutions matching the data, especially for the case of noisy measurement. One popular method to obtain proper $\mathbf{x}$ is the regularization, whose basic idea is to incorporate prior information to narrow the feasible solution space. The best-known regularization strategy is the Tikhonov regularization, by which the solution with minimized squared Euclidean norm (l_2-norm) is desirable [18]. Formally, the regularized version of Eq. (2.1) after using the Tikhonov regularization can be written as

$$\min_{\mathbf{f}} \|\mathbf{f}\|_2^2 \quad s.t. \quad \mathbf{y} = \mathbf{A}\mathbf{f}. \tag{2.3}$$

One important property of (2.3) is the convexity that ensures the uniqueness of the solution. The analytical solution to Eq. (2.3) can be obtained as

$$\mathbf{f} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{y}. \tag{2.4}$$

Due to the simplicity and non-iterative property of the solution (2.4), the l_2-norm regularization has been widely used in various engineering fields. Nonetheless, the researchers have realized that the use of l_2-norm regularization is problematic in many situations, for example, the object to be recovered is non-smooth or mathematically out of H^2-space. Actually, this assumption of minimum energy for the solution is broken in many cases. It makes the use of l_2-norm regularization often leading to results that cannot be accepted by the engineers or customers [18].

Great efforts have been made to find more suitable regularizations [44], among which the sparsity-promoted regularization has been extensively studied. The sparse reconstruction has become a well-established field with clear theoretical foundations and extensive applications. Within the framework of sparse reconstruction, a representative sparsity regularized form is

$$\min \left\|\boldsymbol{\theta}\right\|_1 \quad s.t. \quad \left\|\mathbf{A}\boldsymbol{\Phi}\boldsymbol{\theta} - \mathbf{y}\right\|_2 \le \varepsilon, \tag{2.5}$$

where ε relates to the noise level of the data, $\boldsymbol{\Phi}$ is the transformed domain in which the signal of interest $\theta = \Phi^{-1}\mathbf{f}$ is sparse or compressible [10]. Note that equation (2.5) is convex and it can be addressed through well-established convex optimization algorithms. In the remainder of this chapter, some basic concepts of sparse signal will be briefly introduced at first. Then the restricted isometry property (RIP) and its resulting

theorem are introduced, which provide the guarantee for the accurate recovery of sparse signal through solving l_1-norm regularized convex problem.

2.2 Sparsity and *lp*-norm Measure

For convenience, some necessary concepts and notations are introduced at first. In discrete context, our signals or images can be viewed as vectors endowed in an *N*-dimensional Euclidean space, denoted by $\mathbb{R}^N$. The l_p-norms of vector **x** for $p \in [1,\infty)$ are defined as [18, 56]

$$\|x\|_p = \begin{cases} \left(\sum_{i=1}^{N} |x_i|^p \right)^{1/p} & p \in [1,\infty). \\ \max_i |x_i| & p = \infty. \end{cases} \tag{2.6}$$

Furthermore, the definition of l_p-norm has also been heuristically extended to the case of $p \in [0,1)$. Specially point out that the l_p-norms for $p \in [0,1)$ are just quasi-norms as they fail to satisfy the triangle inequality. The l_p-norms ($p \in [0,\infty)$) have notably different properties for different values of p. To illustrate this, the unit spheres induced by different norms in $\mathbb{R}^2$ are provided in Figure 2-1.

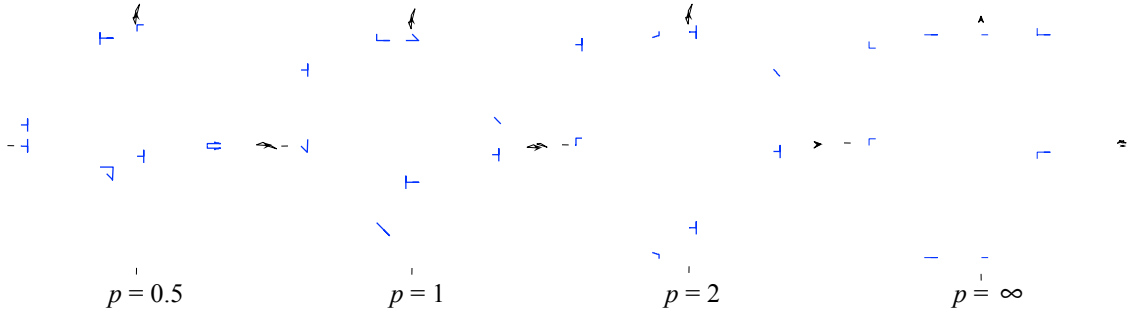
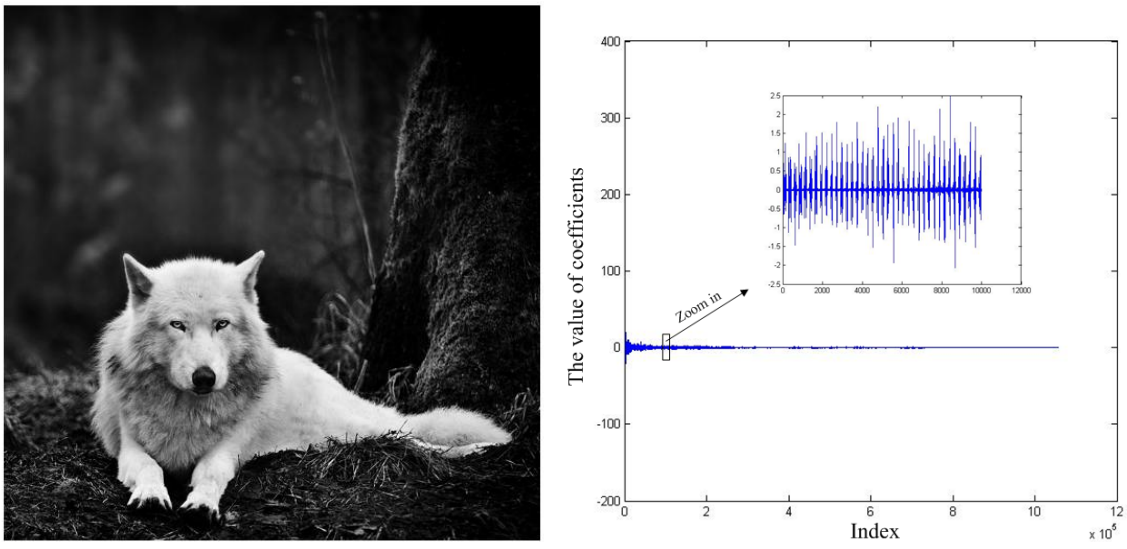


Figure 2-1. Unit spheres induced by different norms in $\mathbb{R}^2$.

Mathematically, a signal **f** is considered as *k*-sparse when at most its *k* entries are nonzero or significant ($\|\mathbf{f}\|_0 \le k$). For the signals that are not themselves sparse but admit sparse representations in some basis or dictionary **Φ**, they are still viewed as *k*-sparse in case that the inequality $\|\mathbf{c}\|_0 \le k$ holds after the transformation of $\mathbf{c} = \mathbf{\Phi f}$ [18, 56]. The sparse signals can often be well approximated as a linear combination of just a few elements in **Φ**.

Lots of natural images and signals have sparse representations in some transformed domain. Taking the natural image shown by Figure 2-2(a) as an example, we can observe that only small fraction of its wavelet transformed coefficients are significant, while the others are very small that can be ignored, as illustrated in Figure 2-2(b).



(a) (b)

Figure 2-2. (a) A natural image and (b) its wavelet transformation coefficients

Hence, good approximations of this image can be obtained by keeping its few largest coefficients, as shown in Figure 2-3(a) and (b), where 1% and 5% its largest coefficients are used, respectively.

(a) (b)

Figure 2-3. The approximations of Figure 2.2(a) by keeping (a) 1% and (b) 5% of its largest wavelet transformation coefficients

2.3 Restricted Isometry Property and Its Implication in Sparse Reconstruction

Candes and Tao [10] introduced the restricted isometry property (RIP) conditional on matrices A and discussed the important role it plays in the recovery of sparse signal.

The definition of the RIP is given as follows **[10]**. A matrix **A** satisfies the restricted isometry property (RIP) of order *k* if there exists a constant $\delta_k \in (0,1)$ such that

$$(1-\delta_k)\|\mathbf{f}\|_2^2 \le \|\mathbf{Af}\|_2^2 \le (1 \quad \delta_k)\|\mathbf{f}\|_2^2 \tag{2.7}$$

holds for all k-sparse $\mathbf{f}$.

If a matrix $\mathbf{A}$ satisfies the RIP of order $2k$, we can interpret Eq. (2.7) as saying that $\mathbf{A}$ approximately preserves the distance between any pair of k-sparse vectors. This will clearly have fundamental implications concerning robustness to noise. It is important to note that in the above definition of the RIP the bounds are assumed to be symmetric about 1 for notational convenience. More generally, arbitrary bounds can be considered [10]

$$0 < \alpha \|\mathbf{f}\|_2^2 \le \|\mathbf{A}\mathbf{f}\|_2^2 \le \beta \|\mathbf{f}\|_2^2 < \quad , \tag{2.8}$$

where $0 < \alpha \le \beta < \quad$.

With RIP, the guarantee of perfectly reconstructing a sparse signal from undersampling noisy measurements is stated by the following theorem [56].

Suppose that A satisfies the RIP of order $2k$ with $\delta_{2k} < \sqrt{2} - 1$ and let $\mathbf{y} = \mathbf{A}\mathbf{f} + \mathbf{e}$ where $\|\mathbf{e}\|_2 \le \varepsilon$. When $B(\mathbf{y}) = \{\mathbf{z} : \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \le \varepsilon\}$, the solution $\hat{\mathbf{f}}$ to the l_1-minization optimization problem of Eq. (2.5) obeys

$$\left\|\mathbf{f} - \hat{\mathbf{f}}\right\|_2 \le C_0 \frac{\sigma_k(\mathbf{f})}{\sqrt{k}} + C_1 \varepsilon \tag{2.9}$$

where

$$C_0 = 2\frac{1 - \left(1 - \sqrt{2}\right)\delta_{2k}}{1 - \left(1 + \sqrt{2}\right)\delta_{2k}},\ C_0 = \frac{4\sqrt{1 + \delta_{2k}}}{1 - \left(1 + \sqrt{2}\right)\delta_{2k}}. \tag{2.10}$$

Finally, it is appropriate to say that due to well-behavioral convex property of problem (2.5), it can be easily solved through a large number of strategies well

developed in the field of convex optimization, such as, iterative shrinkage-thresholding approach [57-59], sparse Bayesian learning approach [60], iteratively reweighed approach [61，62], linear programming [63] and so on.

2.4 Representative Sparse Reconstruction Algorithms

To date, many algorithms have been developed to achieve fast, accurate and stable reconstruction of sparse signals with underdetermining observations. In this subsection three representative sparsity-promoted algorithms are introduced, in particular, the iterative reweighting algorithm (IRA) [62], the iterative shrinkage-thresholding algorithm (ISTA) [57, 58] and its accelerated version (Fast ISTA) [59, 64], and the primal-dual algorithm (PD) [44].

2.4.1 Iterative reweighted algorithm

With sparsity-promoted regularization in terms of l_p-norm ($0 \le p \le 1$) used in the reconstruction, the optimization problem can be formulated as

$$\arg\min_{\mathbf{f}} \left(\left\| \mathbf{Af} - \mathbf{y} \right\|_2^2 + \gamma \left\| \mathbf{f} \right\|_p^p \right), \tag{2.11}$$

where the first term represents the data fidelity and the second term is the sparse penalty. The contributions of these two terms are traded through the regularization factor γ, which should be carefully determined.

An important problem involved in this optimization problem is the non-smoothness of $\left\| \mathbf{f} \right\|_p^p$. To address this problem, we can equivalently express $\left\| \mathbf{f} \right\|_p^p$ as [61]

$$\|\mathbf{f}\|_p^p = \min_{\mathbf{w}\geq 0} \frac{p}{2}\left(\sum_{i=1}^{N} \mathbf{f}_i^2 \mathbf{w}_i + \sum_{i=1}^{N} \frac{2-p}{p} \mathbf{w}_i^{-\frac{p}{2-p}} \right). \tag{2.12}$$

Then Eq. (2.11) can be reformulated into the following two-fold optimization problem,

$$\left(\hat{\mathbf{f}}, \hat{\mathbf{w}}\right) := \arg\min_{\mathbf{x},\mathbf{w}\geq 0} \|\mathbf{y} - \mathbf{A}\mathbf{f}\|_2^2 + \lambda \left[\frac{p}{2} \sum_{i=1}^{N} \mathbf{f}_i^2 \mathbf{w}_i + \frac{2-p}{2} \sum_{i=1}^{N} \mathbf{w}_i^{-\frac{p}{2-p}} \right]. \tag{2.13}$$

The procedure of minimizing Eq. (2.13) can be described as follows. With initial $\mathbf{w}^0$ and $\mathbf{f}^0$, one can go through the following recursion to find $\mathbf{f}$ and $\mathbf{w}$:

Step 1: Update $\mathbf{f}^{(k)}$ *with fixed* $\mathbf{w}^{(k)}$

With $\mathbf{w}^{(k)}$ fixed, $\mathbf{f}^{(k)}$ can be achieved through the following optimization problem:

$$\mathbf{f}^{(k)} = \arg\min_{\mathbf{f}} \|\mathbf{y} - \mathbf{A}\mathbf{f}\|_2^2 \quad \gamma \frac{p}{2} \sum_{i=1}^{N} \mathbf{f}_i^2 \mathbf{w}_i^{(k)}. \tag{2.14}$$

Readily the solution to (2.14) can be directly obtained to be

$$\begin{aligned} &\arg\min_{\mathbf{f}} \left(\|\mathbf{A}\mathbf{f} - \mathbf{y}\|_2^2 + \gamma \cdot \frac{p}{2} \cdot \|\mathbf{f}\|_w^2 \right) \\ \Rightarrow\ &\mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{f}) + \gamma p \mathbf{w}\mathbf{f} = 0 \\ \Rightarrow\ &\mathbf{f}^{(k)} = \left(\mathbf{A}^T\mathbf{A} + \gamma p \mathbf{w}^{(k)} \right)^{-1} \mathbf{A}^T\mathbf{y}. \end{aligned} \tag{2.15}$$

Step 2: Update $\mathbf{w}^{(k+1)}$ *with fixed* $\mathbf{f}^{(k)}$

With $\mathbf{f}^{(k)}$ fixed, $\mathbf{w}^{(k)}$ can be achieved through the following optimization problem:

$$\mathbf{w}^{(k+1)} := \arg\min_{\mathbf{w}\geq 0} \frac{p}{2} \sum_{i=1}^{N} \left(\mathbf{f}_i^{(k)} \right)^2 \mathbf{w}_i + \frac{2-p}{2} \sum_{i=1}^{N} \mathbf{w}_i^{-\frac{p}{2-p}}, \tag{2.16}$$

where the closed form solution for (2.16) is $\mathbf{w}_i^{(k+1)} = \left|\mathbf{f}_i^{(k)}\right|^{p-2}$. Moreover, for the purpose of avoiding singularity of the term of $\left|\mathbf{f}_i^{(k)}\right|^{p-2}$, a slight modification on the solution to $\mathbf{w}_i^{(k+1)}$ should be made as illustrated as

$$\mathbf{w}_i^{(k+1)} = \left(\left(\mathbf{f}_i^{(k)}\right)^2 + \varepsilon\right)^{-\frac{2-p}{2}}, \tag{2.17}$$

where $\varepsilon > 0$ is a small positive value.

By adaptively controlling the dynamic range of **f** through **w** during the iterations, the above algorithm can provide a superior solution. The complete procedure for the iterative reweighted algorithm is summarized in Table 2-1.

Table 2-1. Procedure for the iterative reweighted algorithm

Given: γ; $\mathbf{y}$
Initialization: $\mathbf{w} = \mathbf{I}$; $\mathbf{f} = \mathbf{0}$
for $k = 1,2,3,\cdots,n$
$\mathbf{f}^{(k)} = \left(\mathbf{A}^T\mathbf{A} + \gamma p \mathbf{w}^{(k)}\right)^{-1}\mathbf{A}^T\mathbf{y}$
$\mathbf{w}_i^{(k+1)} = \left(\left(\mathbf{f}_i^{(k)}\right)^2 + \varepsilon\right)^{-\frac{2-p}{2}}$
End

The iterative reweighted algorithm belongs to the class of second-order optimization methods; consequently, it can provide more accurate reconstruction than the first-order algorithms, but at the cost of huge computation.

2.4.2 Iterative shrinkage-thresholding algorithm

Among the family of sparsity-promoted algorithms, another popular approach is the iterative shrinkage-thresholding algorithm (ISTA) [57, 58], which has been widely investigated. Many improved versions have also been proposed, such as the FISTA proposed by A. Beck [59] and NESTA proposed by Stephen Becker [64]. For convenience, the objective function with l_1-norm regularization considered by the iterative shrinkage-thresholding algorithm can be expressed as:

$$\min_{\mathbf{f}} \left\{ \mathcal{J}(\mathbf{f}) + \gamma \left\| \mathbf{f} \right\|_1 \right\}. \tag{2.18}$$

Here, $\mathcal{J}(\mathbf{f})$ is a smooth convex function that represents the data fidelity. Regarding the fact that at any fixed point $\mathbf{f}^{(k)}$ there is an upper bound of $\mathcal{J}(\mathbf{f})$, $\mathcal{J}(\mathbf{f})$ can be approximated as

$$\mathcal{J}(\mathbf{f}) \le \mathcal{J}(\mathbf{f}^{(k)}) + \left\langle \nabla \mathcal{J}\left(\mathbf{f}^{(k)}\right), \mathbf{f} - \mathbf{f}^{(k)} \right\rangle + \frac{1}{2L} \left\| \mathbf{f} - \mathbf{f}^{(k)} \right\|_2^2, \tag{2.19}$$

where L is the Lipschtiz constant. Eq. (2.19) can be equivalently represented as

$$\mathbf{f}^{(k+1)} = \arg\min_{\mathbf{x}} \left\{ \mathcal{J}(\mathbf{f}^{(k)}) + \left\langle \nabla \mathcal{J}\left(\mathbf{f}^{(k)}\right), \mathbf{f} - \mathbf{f}^{(k)} \right\rangle + \frac{1}{2L} \left\| \mathbf{f} - \mathbf{f}^{(k)} \right\|_2^2 + \gamma \left\| \mathbf{f} \right\|_1 \right\}. \tag{2.20}$$

Recalling that $\mathcal{J}\left(\mathbf{f}\right) = \frac{1}{2} \left\| \mathbf{A}\mathbf{f} - \mathbf{y} \right\|_2^2$ and ignoring the constant terms with respect to $\mathbf{f}$, we have that

$$\mathbf{f}^{(k+1)} = \mathbf{f}^{(k)} - L^{-1}\mathbf{A}^T\left(\mathbf{A}\mathbf{f}^{(k)} - \mathbf{y}\right) - \gamma L^{-1} sign(\mathbf{f}^{(k\ 1)}). \tag{2.21}$$

Readily, the update formula for **f** can be obtained through the shrinkage-thresholding operation as

$$\mathbf{f}^{(k+1)} = SoftThreshold\left(\mathbf{f}^{(k)} - L^{-1}\mathbf{A}^T\left(\mathbf{A}\mathbf{f}^{(k)} - \mathbf{y}\right), \gamma L^{-1}\right), \tag{2.22}$$

where

$$SoftThreshold(x,c) = \begin{cases} x - c & x > c. \\ 0 & |x| \le c. \\ x + c & x < \ c. \end{cases} \tag{2.23}$$

In here, Table 2-2 provides the complete procedure for ISTA.

Table 2-2. Procedure for the iterative shrinkage-thresholding algorithm

Given: γ; $\mathbf{y}$; $\mathbf{t}^{(k)}$
Initialization: $\mathbf{f} = \mathbf{0}$
for $k = 1,2,3,\cdots,n$
$\mathbf{f}^{(k+1)} = SoftThreshold\left(\mathbf{f}^{(k)} - L^{-1}\mathbf{A}^T\left(\mathbf{A}\mathbf{f}^{(k)} - \mathbf{y}\right), \gamma L^{-1}\right)$
End

Regarding the convergence of ISTA, it is well known that ISTA has the computation complexity on the order of $1/k$. With the practical use taken into consideration, great efforts have been made in improving ISTA to get lower computation complexity. Here, an accelerated version named as FISTA (fast ISTA) is introduced. It is claimed to be capable of reducing the computation complexity to the order of $1/k^2$ [59].

The FISTA begins with the update of **f** using the operation of shrinkage-thresholding introduced previously:

$$\mathbf{f}^{(k+1)} = SoftThreshold\left(\mathbf{z}^{(k)} - L^{-1}\mathbf{A}^T\left(\mathbf{A}\mathbf{z}^{(k)} - \mathbf{y}\right), \gamma L^{-1}\right). \tag{2.24}$$

The major difference of the FISTA with the standard ISTA is illustrated by the following steps:

$$t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}, \tag{2.25}$$

$$\mathbf{z}^{(k+1)} = \mathbf{f}^{(k+1)} + \left(\frac{t_k - 1}{t_{k+1}}\right)\left(\mathbf{f}^{(k+1)} \quad \mathbf{f}^{(k)}\right). \tag{2.26}$$

It can be observed that a sequence of accelerate factor t_k are involved in the iteration to make use of the value at the previous iteration. The complete procedure for FISTA is exhibited in Table 2-3.

Table 2-3. Procedure for FISTA

Given: γ; L
Initialization: $\mathbf{z} = \mathbf{f} = \mathbf{0}$; $t_1 = 1$
for $k = 1,2,3,\cdots,n$
$\mathbf{f}^{(k+1)} = SoftThreshold\left(\mathbf{z}^{(k)} - L^{-1}\mathbf{A}^T\left(\mathbf{A}\mathbf{z}^{(k)} - \mathbf{y}\right), \gamma L^{-1}\right)$
$t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$
$\mathbf{z}^{(k+1)} = \mathbf{f}^{(k+1)} + \left(\frac{t_k - 1}{t_{k+1}}\right)\left(\mathbf{f}^{(k+1)} \quad \mathbf{f}^{(k)}\right)$
End

2.4.3 Primal-dual algorithm

Another popular first-order algorithm to solve problem (2.5) is the primal-dual approach. Repeatedly, the convex optimization problem to be addressed is

$$\min \|\mathbf{f}\|_1 \quad s.t. \quad \mathbf{Af} = \mathbf{y}. \tag{2.27}$$

In the context of primal-dual, the above equation can be reformulated into

$$\max_{\lambda} \min_{\mathbf{f}} \left(\langle \boldsymbol{\lambda}, \mathbf{Af} - \mathbf{y} \rangle + \|\mathbf{f}\|_1 \right). \tag{2.28}$$

Note that due to the non-smoothness of $\|\mathbf{f}\|_1$, available expression for updating $\mathbf{f}$ is intractable. Therefore, the l_1-norm term in Eq. (2.28) is reformulated into

$$\|\mathbf{f}\|_1 = \frac{1}{2}\left(\|\mathbf{f}\|_1 + \|\mathbf{f}\|_1\right) = \frac{1}{2}\min_{\mathbf{w}} \left(\|\mathbf{f}\|_1 + \frac{1}{2}\left(\|\mathbf{f}\|_{\mathbf{w}}^2 + \sum_{i=1}^{N} \mathbf{w}_i^{-1} \right) \right). \tag{2.29}$$

Note that $\|\mathbf{f}\|_1 = \frac{1}{2}\left(\|\mathbf{f}\|_1 + \|\mathbf{f}\|_1\right)$ also can be expressed in other form, for example, $\|\mathbf{f}\|_1 = \alpha\|\mathbf{f}\|_1 + (1 \quad \alpha)\|\mathbf{f}\|_1$ where $\alpha \in [0,1]$. Now the resulting primal-dual objective function becomes

$$\{\mathbf{f}, \mathbf{w}, \lambda\} = \arg\max_{\lambda} \arg\min_{\mathbf{w}} \arg\min_{\mathbf{f}} \left(\langle \boldsymbol{\lambda}, \mathbf{Af} - \mathbf{y} \rangle + \frac{1}{2}\left(\|\mathbf{f}\|_1 + \frac{1}{2}\left(\|\mathbf{f}\|_{\mathbf{w}}^2 + \sum_{i=1}^{N} \mathbf{w}_i^{-1} \right) \right) \right). \tag{2.30}$$

The update formulas for $\mathbf{f}$, $\boldsymbol{\lambda}$ and $\mathbf{w}$ are respectively shown as follows.

Step 1: Update of $\boldsymbol{f}$

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} \left(\langle \boldsymbol{\lambda}, \mathbf{Af} - \mathbf{y} \rangle + \frac{1}{2}\left(\|\mathbf{f}\|_1 + \frac{1}{2}\|\mathbf{f}\|_{\mathbf{w}}^2 \right) \right) \tag{2.31}$$

The closed-form solution to Eq. (2.31) can be straightforward derived as

$$\mathbf{f}^{(k+1)} = 2\mathbf{w}^{-1} SoftThreshold\left(\mathbf{A}^T \boldsymbol{\lambda}, 0.5 \right). \tag{2.32}$$

Step 2: Update of $\mathbf{w}$

By ignoring the constant terms with respect to $\mathbf{w}$, we can renew $\mathbf{w}$ through the following equation:

$$\{\mathbf{f}, \mathbf{w}, \lambda\} = \arg\min_{\mathbf{w}} \left(\|\mathbf{f}\|_{\mathbf{w}}^2 + \sum_{i=1}^{N} \mathbf{w}_i^{-1} \right) \Rightarrow \mathbf{w}_{ii} \quad 1/|\mathbf{f}_i|. \tag{2.33}$$

Specifically, the l_1-norm term in Eq. (2.27) is replaced by $\|\mathbf{f}\|_{\mathbf{w}}^2 = \sum_i \frac{|\mathbf{f}_i|^2}{w_i + \varepsilon}$, where w_i is determined by the value of f_i obtained through previous iteration, and ε is a small positive small real to void the singularity.

Step 3: Update of $\boldsymbol{\lambda}$

The $\boldsymbol{\lambda}$ is updated by $\boldsymbol{\lambda} = \boldsymbol{\lambda} - \mu(\mathbf{A}\mathbf{f} - \mathbf{y})$ where μ is a suitable step size. Finally, the complete procedure for the triple iteration approach is summarized in Table 2-4.

Table 2-4. Procedure for the triple iteration algorithm

Given: μ ; $\mathbf{y}$
Initialization: $w = \mathbf{I}$; $\mathbf{f} = \mathbf{0}$; $\lambda = 1$
for $k = 0,1,2,\cdots,n$
$\mathbf{f}^{(k+1)} = 2\mathbf{w}^{-1} SoftThreshold\left(\mathbf{A}^T\boldsymbol{\lambda}, 0.5\right)$
$\mathbf{w}_{ii}^{(k+1)} = 1/\left\|\mathbf{f}_i^{(k+1)}\right\|$
$\boldsymbol{\lambda}^{(k+1)} = \boldsymbol{\lambda}^{(k)} - \mu\left(\mathbf{A}\mathbf{f}^{(k+1)} - \mathbf{y}^{(k+1)}\right)$
End

3. PROPOSED METHODOLOGY

3.1 Improved ART Algorithm

3.1.1 Order-set ART

It is well known that the order-set technique is an accelerated strategy widely used, for example, the order-set EM (OS-EM) algorithm in the realm of medical imaging. The order-set ART (OS-ART) algorithm developed in [65] is briefly discussed below. For convenience, the update formula for the standard ART is repeated here,

$$\mathbf{f}^{(new)} = \mathbf{f}^{(old)} - \frac{\mathbf{y}_i - \left\langle \mathbf{a}_i, \mathbf{f}^{(old)} \right\rangle}{\left\| \mathbf{a}_i \right\|_2^2} \mathbf{a}_i . \tag{3.1}$$

Seen from the above equation, only one element of the observed data is used each iteration. With the order-set technique employed, the observed data are grouped into n subsets; consequently, the iteration formula of Eq. (3.1) turns to be

$$\mathbf{f}^{(new)} = \mathbf{f}^{(old)} - \begin{pmatrix} \mathbf{a}_{i\cdot m+1} \\ \mathbf{a}_{i\cdot m+2} \\ \vdots \\ \mathbf{a}_{(i+1)\cdot m} \end{pmatrix}^T \frac{\begin{pmatrix} \mathbf{y}_{i\cdot m+1} \\ \mathbf{y}_{i\cdot m+2} \\ \vdots \\ \mathbf{y}_{(i+1)\cdot m} \end{pmatrix} - \begin{pmatrix} \mathbf{a}_{i\cdot m+1} \\ \mathbf{a}_{i\cdot m+2} \\ \vdots \\ \mathbf{a}_{(i+1)\cdot m} \end{pmatrix} \mathbf{f}^{(old)}}{\left\| \begin{pmatrix} \mathbf{a}_{i\cdot m+1} \\ \mathbf{a}_{i\cdot m+2} \\ \vdots \\ \mathbf{a}_{(i+1)\cdot m} \end{pmatrix} \right\|_2^2}, \tag{3.2}$$

where $i = k \bmod n + 1$ at the step k, and $m = \mathrm{int}(M/n)$. To point out specifically, it returns to be the standard ART algorithm when $n = M$; on the contrary, it becomes the traditional Newton-based algorithm if n is set to be 1.

Note that the convergence rate of iterative algorithms using order-set technique is highly sensitive to the choice of n. Therefore, it's appealing to choose a suitable n to get optimal convergence in practice.

3.1.2 Sparsity-promoted order-set ART algorithm

Due to the lack of prior information used, ART usually performs poorly with deficient observation and suffers from the serious salt and pepper noise. To mitigate this problem, the ART+TV reconstruction method has been investigated by several authors [66-68]. This approach can be viewed as a sparsity-promoted ART algorithm.

In here, the more general framework of sparsity-promoted ART is built and its objective function can be formulated as

$$\min \left\| \mathbf{\Phi f} \right\|_1 \quad s.t. \quad \mathbf{a}_i \mathbf{f} = \mathbf{y}_i \quad (i = 1, 2, \mathrm{L}\ n). \tag{3.3}$$

The FISTA algorithm discussed in previous chapter is adopted to solve Eq. (3.3). The complete procedure of the proposed accelerated order-set ART algorithm is summarized in Table 3-1. Two distinct benefits are rendered by this proposed method, in particular: (a) the sparsity prior of the imaged object is readily incorporated; (b) the accelerated advantages of the order-set technique and multiple-step strategy used in FISTA are simultaneously taken. Consequently, the convergence of the proposed method can be greatly improved, which will be verified through the numerical simulations provided next chapter.

Table 3-1. Procedure for the proposed accelerated sparsity-promoted order-set ART algorithm

---

Given: $\gamma; L$

Initialization: $\mathbf{z} = \mathbf{f} = \mathbf{0}$; $t_1 = 1$

for $k = 1,2,3,\cdots,n$

$$\mathbf{z}^{(k)} = \mathbf{z}^{(k)} - \begin{pmatrix} \mathbf{a}_{i\cdot m+1} \\ \mathbf{a}_{i\cdot m+2} \\ \vdots \\ \mathbf{a}_{(i+1)\cdot m} \end{pmatrix}^T \frac{\begin{pmatrix} \mathbf{y}_{i\cdot m+1} \\ \mathbf{y}_{i\cdot m+2} \\ \vdots \\ \mathbf{y}_{(i+1)\cdot m} \end{pmatrix} - \begin{pmatrix} \mathbf{a}_{i\cdot m+1} \\ \mathbf{a}_{i\cdot m+2} \\ \vdots \\ \mathbf{a}_{(i+1)\cdot m} \end{pmatrix} \mathbf{z}^{(k)}}{\left\| \begin{pmatrix} \mathbf{a}_{i\cdot m+1} \\ \mathbf{a}_{i\cdot m+2} \\ \vdots \\ \mathbf{a}_{(i+1)\cdot m} \end{pmatrix} \right\|_2^2}$$

$$\mathbf{f}^{(k+1)} = \mathbf{\Phi}^T SoftThreshold\left(\mathbf{\Phi}\left(\mathbf{z}^{(k)} - L^{-1} \sum_t \mathbf{A}_t^T \left(\mathbf{A}_t \mathbf{z}^{(k)} - \mathbf{y}_t \right) \right), \gamma L^{-1} \right)$$

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$$

$$\mathbf{z}^{(k+1)} = \mathbf{f}^{(k+1)} + \left(\frac{t_k - 1}{t_{k+1}}\right)\left(\mathbf{f}^{(k+1)} \quad \mathbf{f}^{(k)}\right)$$

End

3.2 The Sparsity-promoted Reconstruction with Combined Regularizations of l_1-norm and p-TV

The TOF-PET imaging problem of Eq. (1.4) can be addressed through the following least square problem

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} \sum_{t=-T}^{T} \left\| \mathbf{y}_t - \mathbf{A}_t \mathbf{f} \right\|_2^2 \tag{3.4}$$

As pointed out previously, without the use of suitable prior information, almost all of methods fail to produce meaningful solution to Eq. (3.4) for undersampling TOF-PET imaging, especially for the case of noisy observations. To circumvent this problem, two regularizations of p-TV and l_1-norm are incorporated into the above objective function; consequently,

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} \left\{ \sum_{t} \left\| \mathbf{A}_t \mathbf{f} - \mathbf{y}_t \right\|_2^2 + \gamma_1 \left\| \mathbf{f} \right\|_{p-TV} + \gamma_2 \left\| \mathbf{\Phi f} \right\|_1 \right\}. \tag{3.5}$$

Here, the p-TV and l_1-norm terms are used to take advantages of the correlated-structure sparse and piecewise-constant properties of the object **f**. Furthermore, for TOF-PET imaging, the use of p-TV regularization is more efficient to detect a "hot" region, implying a tumor, which manifests a sudden change boundary with respect to its surrounding normal tissue. In Eq. (3.5), $\left\| \mathbf{f} \right\|_{p-TV}$ is defined as

$$\left\| \mathbf{f} \right\|_{p-TV} = \sum_{i,j} \left(\left(\Delta^x \mathbf{f}_{i,j} \right)^2 + \left({}^y \mathbf{f}_{i,j} \Delta \right)^2 \right)^{p/2} \tag{3.6}$$

and

$$\Delta^x \mathbf{f}_{i,j} = \mathbf{f}_{i+1,j} - \mathbf{f}_{i,j}, \Delta^y \mathbf{f}_{i,j} = \mathbf{f}_{i,j+1} - \mathbf{f}_{i,j} \tag{3.7}$$

where **f** in Eqs. (3.6) and (3.7) is the $n \times n$ matrix representing the discrectized object. The factors of γ_1 and γ_2 allow the trade-off between the observed data and the priors of **f**. The transformation domain $\mathbf{\Phi}$ is specified as a discrete cosine transformation (DCT) here.

It is noted that the objective function of Eq. (3.5) is analogous to the one proposed by Lustig et al. [22]. However, compared to [22] and other relevant works, this work presents two specific differences: (a) instead of implementing a conjugate gradient algorithm to the whole objective function, an alternative approach was provided to solve Eq. (3.5), i.e., separately dealing with *p*-TV regularization term with an iteratively reweighted method, while applying operation of shrinkage-thresholding on l_1-norm constraint; (b) the generalized TV, i.e., *p*-TV, instead of standard TV regularization is adopted. As a matter of fact, $\|\mathbf{f}\|_{p\text{-}TV}$ is a straightforward extension of the standard TV regularization, which was initially mentioned by Rodriguez and Wohlberg [69]. In [69], only the 1-TV and 2-TV are numerically studied for imaging painting and de-noising. In the subsequent section we discuss the performances of different *p*-TV adapted for our TOF-PET imaging problem, which show that the 0.5-TV is optimal.

To efficiently solve the Eq. (3.5), variable splitting technique is applied on the term of $\gamma_2 \|\mathbf{\Phi f}\|_1$; consequently, the Eq. (3.5) can be reformulated into a two-fold optimization problem, namely,

$$\min_{\mathbf{f},\mathbf{d}} \left\{ \sum_t \|\mathbf{A}_t\mathbf{f} - \mathbf{y}_t\|_2^2 + \gamma_1 \|\mathbf{f}\|_{p-TV} + \gamma_2 \|\mathbf{d}\|_1 + \gamma_3 \|\mathbf{d} - \mathbf{\Phi f}\|_2^2 \right\} \tag{3.8}$$

Intuitively, the classical alternatively iterative strategy can be exploited to treat Eq. (3.8). The working procedure consists of alternation between performing the update of **f** using the current estimate for **d**, and computing **d** based on the update of **f**.

In summary, the updates for **f** and **d** at the *n*th iteration are made by solving the following two sub-optimization problems

$$\mathbf{f}^{(n)} = \arg\min_{\mathbf{f}} \left\{ \sum_t \left\| \mathbf{A}_t \mathbf{f} - \mathbf{y}_t \right\|_2^2 + \gamma_1 \left\| \mathbf{f} \right\|_{p-TV} + \gamma_3 \left\| \mathbf{d}^{(n-1)} \quad \mathbf{\Phi f} \right\|_2^2 \right\}. \tag{3.9}$$

and

$$\mathbf{d}^{(n)} = \arg\min_{\mathbf{d}} \left\{ \gamma_2 \left\| \mathbf{d} \right\|_1 + \gamma_3 \left\| \mathbf{d} \quad \mathbf{\Phi f}^{(n)} \right\|_2^2 \right\} \tag{3.10}$$

respectively.

Step 1: Update of **f**

In this research the strategy of iteratively reweighted approach is used to solve Eq. (3.9). $\left\| \mathbf{f} \right\|_{p-TV}$ can be alternatively expressed as

$$\begin{aligned} \left\| \mathbf{f} \right\|_{p-TV} &= \sum_{i,j} \left(\left(\Delta^x \mathbf{f}_{i,j} \right)^2 + \left(\quad {}^y \mathbf{f}_{i,j} \right)^2 \right)^{p/2} \\ &= \sum_{i,j} \frac{\left(\Delta^x \mathbf{f}_{i,j} \right)^2}{\left(\left(\Delta^x \mathbf{f}_{i,j} \right)^2 + \left(\Delta^y \mathbf{f}_{i,j} \right)^2 \right)^{1-p/2}} + \sum_{i,j} \frac{\left(\Delta^x \mathbf{f}_{i,j} \right)^2}{\left(\left(\Delta^x \mathbf{f}_{i,j} \right)^2 + \left(\quad {}^y \mathbf{f}_{i,j} \right)^2 \right)^{1-p/2}} \\ &= \left\| \mathbf{D}_x \mathbf{f} \right\|_{\mathbf{w}}^2 + \left\| \mathbf{D}_y \mathbf{f} \right\|_{\mathbf{w}}^2 \\ &\approx \left\| \mathbf{D}_x \mathbf{f} \right\|_{\mathbf{w}^{(n-1)}}^2 + \left\| \mathbf{D}_y \mathbf{f} \right\|_{\mathbf{w}^{(n-1)}}^2 \end{aligned} \tag{3.11}$$

where

$$\left(\mathbf{D}_x \mathbf{f} \right)_{i,j} = \Delta^x \mathbf{f}_{i,j}, \left(\mathbf{D}_y \mathbf{f} \right)_{i,j} = \Delta^y \mathbf{f}_{i,j} \tag{3.12}$$

and

$$\mathbf{w}^{(n-1)} = diag\left(\left(\left(\Delta^x \mathbf{f}_{i,j}^{(n-1)} \right)^2 + \left(\quad {}^y \mathbf{f}_{i,j}^{(n-1)} \right)^2 \right)^{p/2-1} \right) \tag{3.13}$$

In order to avoid singularity of $\left(\left(\Delta^x \mathbf{f}_{i,j}^{(n-1)} \right)^2 + \left(\quad {}^y \mathbf{f}_{i,j}^{(n-1)} \right)^2 \right)$, a small positive real ε is introduced to modify **w**, i.e.,

$$\mathbf{w}^{(n-1)} = diag\left(\left(\left(\Delta^x \mathbf{f}_{i,j}^{(n-1)} \right)^2 + \left(\Delta^y \mathbf{f}_{i,j}^{(n-1)} \right)^2 + \varepsilon \right)^{p/2-1} \right) \tag{3.14}$$

Now, we can readily derive the iterative solution to Eq. (3.9) as

$$\mathbf{f}^{(n)} = \arg\min_{\mathbf{f}} \left\{ \begin{array}{l} \sum_t \left\| \mathbf{A}_t \mathbf{f} - \mathbf{y}_t \right\|_2^2 + \gamma_3 \left\| \mathbf{d}^{(n-1)} \quad \mathbf{\Phi f} \right\|_2^2 \\ + \gamma_1 \left(\left\| \mathbf{D}_x \mathbf{f} \right\|_{\mathbf{w}^{(n-1)}}^2 + \left\| \mathbf{D}_y \mathbf{f} \right\|_{\mathbf{w}^{(n-1)}}^2 \right) \end{array} \right\} \tag{3.15}$$

Explicitly, the closed-form estimation of $\mathbf{f}^{(n)}$ can be achieved as

$$\begin{aligned} \mathbf{f}^{(n)} = & \left(\sum_t \mathbf{A}_t' \mathbf{A}_t + \gamma_1 \left(\mathbf{D}_x' \mathbf{w}^{(n-1)} \mathbf{D}_x + \mathbf{D}_y' \mathbf{w}^{(n-1)} \mathbf{D}_y \right) + \gamma_3 \mathbf{I} \right)^{-1} \\ & \times \left(\gamma_3 \mathbf{\Phi}' \mathbf{d}^{(n-1)} + \sum_t \mathbf{A}_t' \mathbf{y}_t \right) \end{aligned} \tag{3.16}$$

To avoid the overwhelming computation of calculating matrix inverse involved in Eq. (3.16), the classical preconditioned conjugate gradient (PCG) method was explored to solve Eq. (3.15) in this research.

Step 2: Update of $\mathbf{d}$

Through the standard implementation widely used in the literature of compressed sampling, the following shrinkage solution $\mathbf{d}^{(n)}$ to problem (3.10) can be readily achieved as [59]

$$\begin{aligned} \mathbf{d}^{(n)} &= \arg\min_{\mathbf{d}} \left\{ \gamma_2 \left\| \mathbf{d} \right\|_1 + \gamma_3 \left\| \mathbf{d} \quad \mathbf{\Phi f}^{(n)} \right\|_2^2 \right\} \\ &= Soft\_Thr\left(\mathbf{\Phi f}^{(n)}, \frac{\gamma_2}{\gamma_3} \right) \end{aligned} \tag{3.17}$$

The whole procedure for proposed algorithm for solving sparsity-promoted TOF-PET reconstruction has been summarized in Table 3-2.

Table 3-2. Procedure for the sparsity-promoted reconstruction with combined regularizations of l_1-norm and p-TV

Initial setup: $\mathbf{d}=\mathbf{0}$; $\mathbf{w}=\mathbf{I}$; $\mathbf{f}=\mathbf{0}$; $n=1$

While $Err \geq 10^{-5}$ & $n \leq 50$

$$\mathbf{f}^{(n)} = \left(\sum_t \mathbf{A}_t' \mathbf{A}_t + \gamma_1 \left(\mathbf{D}_x' \mathbf{w}^{(n-1)} \mathbf{D}_x + \mathbf{D}_y' \mathbf{w}^{(n-1)} \mathbf{D}_y \right) + \gamma_3 \mathbf{I} \right)^{-1} \times \left(\gamma_3 \mathbf{\Phi}' \mathbf{d}^{(n-1)} + \sum_t \mathbf{A}_t' \mathbf{y}_t \right)$$

$$\mathbf{d}^{(n)} = \arg\min_{\mathbf{d}} \left\{ \gamma_2 \|\mathbf{d}\|_1 + \gamma_3 \|\mathbf{d} \quad \mathbf{\Phi} \mathbf{f}^{(n)}\|_2^2 \right\} = Soft\_Thr\left(\mathbf{\Phi} \mathbf{f}^{(n)}, \frac{\gamma_2}{\gamma_3} \right)$$

$$\mathbf{w} = diag\left(\left(\left(\Delta^x \mathbf{f}_{i,j} \right)^2 + \left(\Delta^y \mathbf{f}_{i,j} \right)^2 \quad \varepsilon \right)^{p/2-1} \right)$$

$$Err = \frac{\sum_t \|\mathbf{A}_t \mathbf{f} - \mathbf{y}_t\|_2^2}{\sum_t \|\mathbf{y}_t\|_2^2}$$

$n = n+1$

End While

Furthermore, it is well known that to ensure meaningful solution we have to carefully choose regularization factor γ_1. Specifically, if γ_1 is big, the resulting solution will seriously stray from the data fidelity; on the other hand, small γ_1 will cause Eq. (9) highly ill-conditioned, and extremely slow convergence for PCG. The popular

continuation technique has been adopted to address this issue by initially setting a much larger γ_1 and gradually decreasing its value when the iteration is preceded [70]. Therefore, the optimized solution with previous value of γ_1 provides the warm start for the next iteration. In the proposed method, the continuation technique was adopted and empirically chosen as $\gamma_1 \Leftarrow 0.8 \cdot \gamma_1$.

4. NUMERICAL SIMULATIONS AND DISCUSSIONS

4.1 Setup of Numerical Simulation

In this section, numerical experiments were carried out to investigate the benefits brought by the sparsity-promoted method to the undersampling TOF-PET imaging. Two different undersampling measurement configurations are proposed. The first one is parallel to the current traditional PET system, except fewer detectors are sparsely and uniformly distributed over a complete PET ring, as sketched in Figure 4-1(b). The second one is the limited-angle PET system represented in [9], where the detectors are furnished with the density of $384/360^{o}$ over two opposite partial PET rings, as shown in Figure 4-1(c). For convenience, the first undersampling configuration is referred as configuration *A* while the second one is called configuration *B*. In this work, the radius of PET ring is set to be 35cm, while the scale of ROI into which the phantom is embedded is 30 cm by 30 cm. In addition, the physical effects like scattering are not taken into consideration.

As shown in Figure 4-2(a), the phantom used in this research is the well-known Shepp-Logan phantom simulated using MATLAB tool (i.e., phantom). In the context of numerical implementation, this phantom is discretized into 128×128 pixels. The horizontal and vertical profiles of this phantom are defined and provided in Figure 4-2(b) and Figure 4-2(c), respectively.

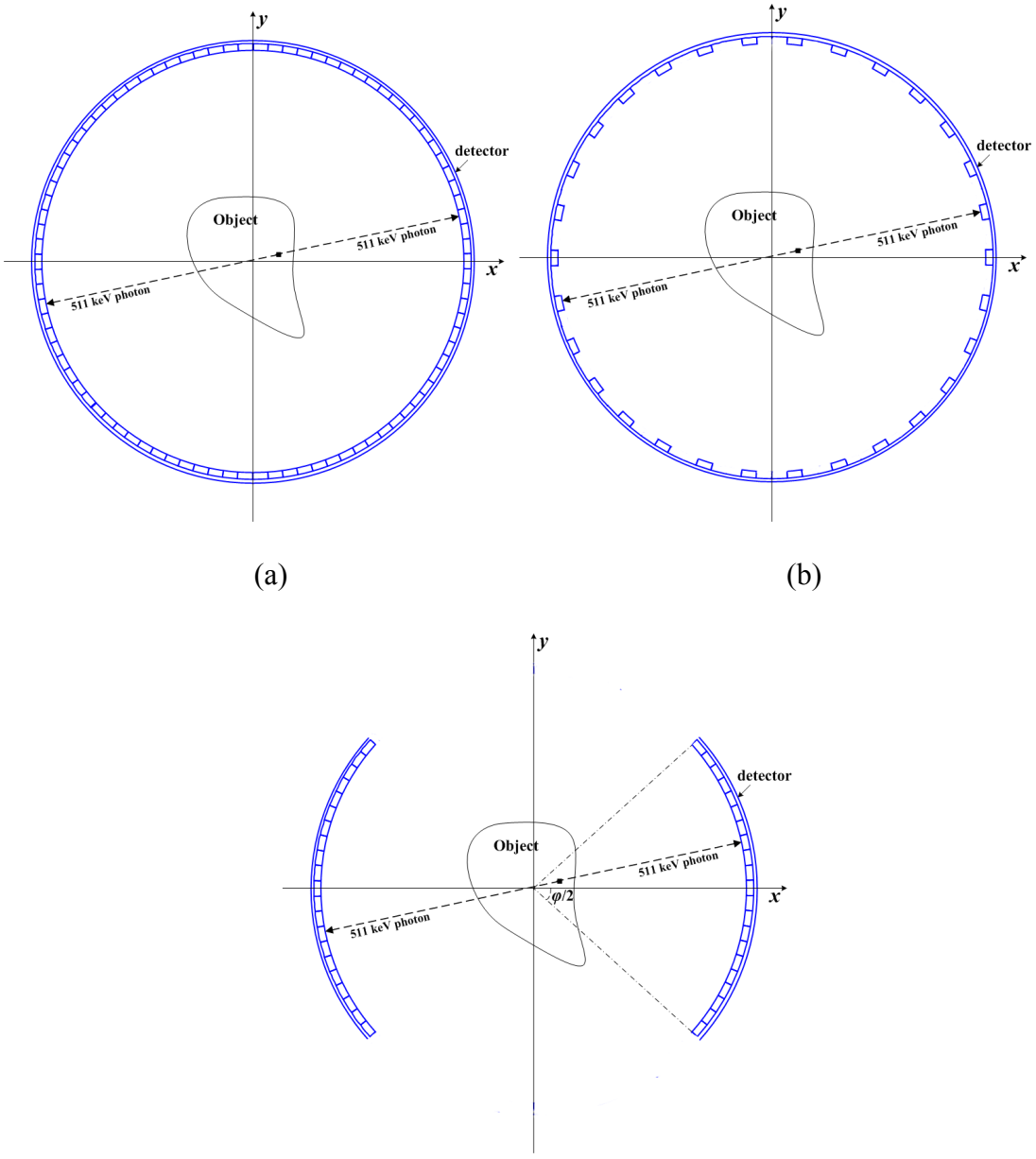
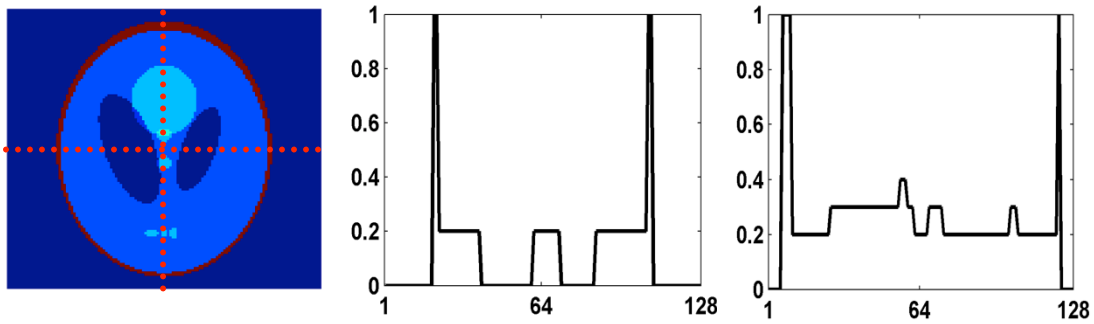


Figure 4-1. Schematic diagrams of (a) traditional PET configuration where detectors are densely distributed in a complete ring, (b) where detectors are sparsely and uniformly distributed over a complete PET ring (Configuration A), and (c) where detectors are compactly distributed over two opposite partial rings (Configuration B). Compared with the traditional PET configuration, both configuration *A* and *B* are undersampling system and will bring incomplete observations.



(a) (b) (c)

Figure 4-2. (a) Ground truth of 128×128 Shepp–Logan phantom used in numerical simulations; (b) horizontal profile that corresponds to the slice indicated by horizontal dotted line in panel (a); (c) vertical profile which corresponds to the slice indicated by vertical dotted line in panel (a).

In here, RMSE (Root of Mean Square Error) and SSIM (Structural Similarity) [71] are used as criteria to assess the reconstruction quality, and their definitions are

$$RMSE = \frac{\left\| \mathbf{f}_{rec} - \mathbf{f}_{true} \right\|}{\left\| \mathbf{f}_{true} \right\|} \tag{4.1}$$

and

$$SSIM = \frac{\left(2\mu_{rec}\mu_{true} + c_1\right)\left(2\sigma_{cross} + c_2\right)}{\left(\mu_{rec}^2 + \mu_{true}^2 + c_1\right)\left(\sigma_{rec}^2 + \sigma_{true}^2 + c_2\right)}, \tag{4.2}$$

respectively. Here, μ_{rec} and μ_{true} are the average of reconstructed image $\mathbf{f}_{rec}$ and true image $\mathbf{f}_{true}$, respectively; σ_{rec}^2 and σ_{true}^2 are the associated variances of them, σ_{cross} is the covariance of f_{rec} and f_{true}, c_1 and c_2 are two variables to stabilize the division with weak dominator.

Additionally, two criterions of the maximum iteration number and the minimum tolerance error are employed to stop the iterative procedure. The tolerance error is then defined as

$$Err = \frac{\sum_t \left\| \mathbf{y}_t - \mathbf{A}_t \mathbf{f}_{rec} \right\|_2}{\sum_t \left\| \mathbf{y}_t \right\|_2}. \tag{4.3}$$

The remainder of this chapter is arranged as follows. First, the performance of the accelerated sparsity-promoted ART method proposed in section 3.1 was investigated; through the comparison with the traditional ART algorithm and the study on convergence. Second, a set of numerical experiments was carried out to find the optimal value of p for p-TV in the sparsity-promoted algorithm with combined regularizations of l_1-norm and p-TV, which was proposed in section 3.2. Third, using the optimal p value, the superiority of this sparsity-promoted method in dealing with undersampling observation was demonstrated, through the comparison with traditional EM algorithm. Finally, based on the sparsity-promoted reconstruction with combined regularizations of l_1-norm and p-TV, the reconstruction performances dependent on the time resolution τ, and the number of detectors N (or the range of angle φ) were investigated in section 4.5 and 4.6, respectively.

4.2 Numerical Test 1: Investigation on the Accelerated Sparsity-promoted ART Algorithm

In this subsection, a set of numerical experiments under configuration A are carried out to examine the performance of the proposed sparse ART algorithm developed in section 3.1, in particular: (a) to investigate the benefits brought by the proposed accelerated sparsity-promoted algorithm through the comparisons with the traditional ART; (b) to study the improved convergence of the proposed method brought

by the use of the order-set technique and multi-step technique mentioned in FISTA. In this set of simulation, 190 detectors are used; the time resolution and TOF bin are set to be 500 *ps* and 67 *ps*, respectively.

The technique of order-set has been adopted and the sizes of sets are chosen to be 1, 10, 100, 1000 and M, where M is the number of total measurements. In Figure 4-3, the TOF-PET images with different sizes of sets reconstructed by ART (Figure 4-3(a)) and the proposed accelerated sparsity-promoted ART (Figure 4-3(b)) are compared. Their associated profiles are also reported in red solids, and the truth profiles are represented in black solids for comparison. The RMSEs and (1-SSIM)s corresponding to Figure 4-3 are reported in Table 4-1.

From the above results, we can observe that the reconstruction quality is sensitive to the size of sets. Furthermore, with suitable sizes of sets, the proposed accelerated sparsity-promoted algorithm can provide better reconstruction results. For example, when the size is chosen as '100', the RMSE and 1-SSIM are enhanced from 0.0735 and 0.046 to 0.0316 and 0.0053 by the use of accelerated sparsity-promoted operation. Furthermore, the profiles of the images reconstructed through the proposed accelerated sparsity-promoted ART algorithm are much smoother than those obtained through the traditional ART algorithm.

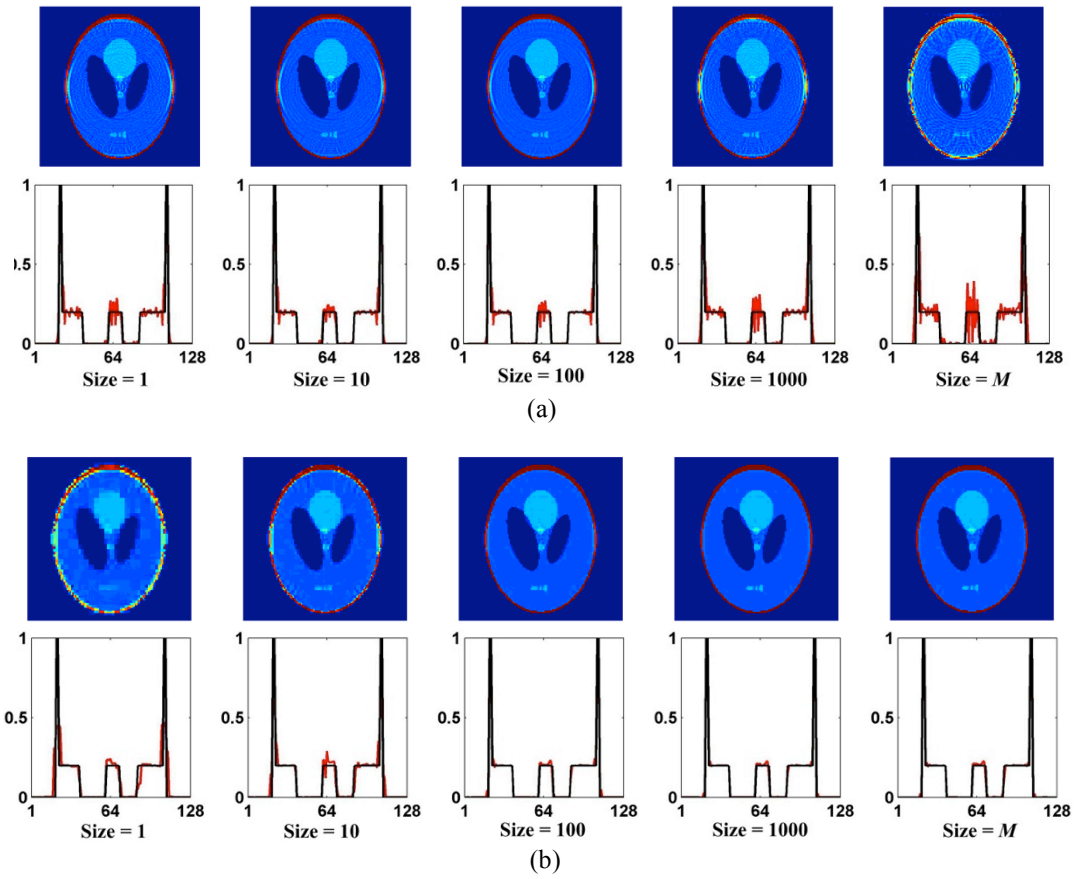


Figure 4-3. TOF-PET images and associated profiles obtained through (a) order-set ART and (b) accelerated order-set sparsity-promoted ART. The sizes of set are used as 1, 10, 100, 1000, *M*.

Table 4-1. RMSEs and (1-SSIM)s that correspond to figure 4-3

Size of sets		1	10	100	1000	*M*
TOF	RMSE	0.0982	0.0864	0.0735	0.1615	0.2739
ART	1-SSIM	0.0681	0.0469	0.0316	0.0652	0.1307
TOF	RMSE	0.2510	0.1361	0.046	0.0284	0.0231
Acc-SART	1-SSIM	0.0964	0.0283	0.0053	0.0029	0.0044

Figure 4-4(a) reports the convergences of order-set ART with different sizes of sets. We can observe that suitable choices of the size of sets do greatly accelerate the convergence at the early stage of iteration. And '100' shows to be the best one among 1, 10, 100, 1000 and *M*. As for the sparsity aided ART algorithms, the convergences of the

proposed order-set sparsity-promoted ART with and without the multiple-step accelerated strategy are compared in Figure 4-4(b).

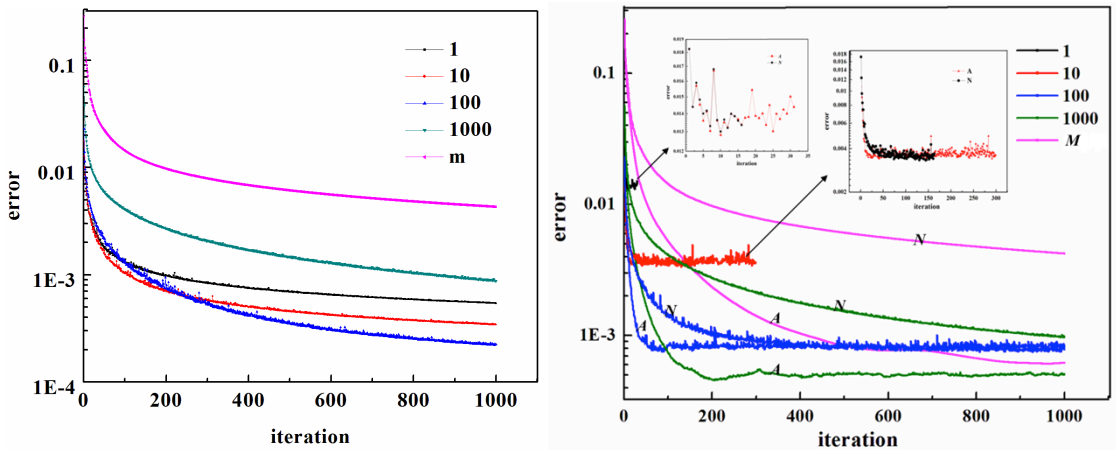


Figure 4-4. Convergence curves of standard ART, OS-ART, and accelerated OS-ART algorithms, where the sizes of sets are chosen as 1, 10, 100, 1000 and *M*.

In this figure, the lines marked with the symbol '*N*' correspond to the cases without multi-step accelerated strategy, while the lines with symbol '*A*' correspond to the cases with multi-step strategy. Comparing Figure 4-4(b) with 4-4(a), we can observe that the converged values of objective function with sparsity-promoted constraint will be slightly larger than those without the use of sparsity penalty, which is resulted from the trade-off between the data fidelity and penalty, as shown in Eq. (3.3).

Furthermore, some important conclusions can be deducted through the above figures: (a) the order-set technique with suitable size of sets can provide improved convergence performance; (b) as pointed out in the literatures, the order-set technique can improve the convergence rate at the early stage of iteration, however, it usually cause the oscillation at the later stage especially for the cases of small size of sets; (c) the

use of multi-step accelerated technique can further improve the convergence of order-set sparsity-promoted ART within whole iteration process; and (d) the combination of the multiple-step technique and '1000' as the size of sets provides the best convergence performance.

4.3 Numerical Test 2: Investigation on *p*-TV Regularization

Regarding the approach of combined regularizations of *p*-TV and l_1-norm proposed in section 3.2, it is well known that different choices of variable *p* involved in *p*-TV will produce different effects on the TOF-PET reconstruction. Roughly, the sparse solution can be promoted when $0 \le p \le 1$, in contrast, no sparsity is enforced otherwise. Furthermore, it is well known that $p = 0$, $p = 1$ and $p = 2$ are three candidates most widely used in dealing with inverse problem. Specifically, $p = 0$ corresponds to the so-called support detector [72]. It usually presents a computational challenge due to the intrinsic non-convexity; thus, when $p = 1$ it is known as the standard TV regularizer [73], and when $p = 2$ it is the known Tikhonov regularization. So far, it hasn't been clear yet what the optimal value of *p* is for efficiently dealing with our problem. In this subsection, the aim is to find the optimal value of *p* for our undersampling TOF-PET reconstruction problem.

These numerical experiments are implemented under configuration *A* with 70 detectors used, and the time resolution and TOF bin are set to be 500 *ps* and 67 *ps*, respectively. Figure 4-5 shows a series of reconstructed TOF-PET images corresponding

to p = 0, 0.5, 1, 1.5 and 2 and Figure 4-6 shows their associated convergences. Their corresponding RMSEs and (1-SSIM)s are also reported in Table 4-1.

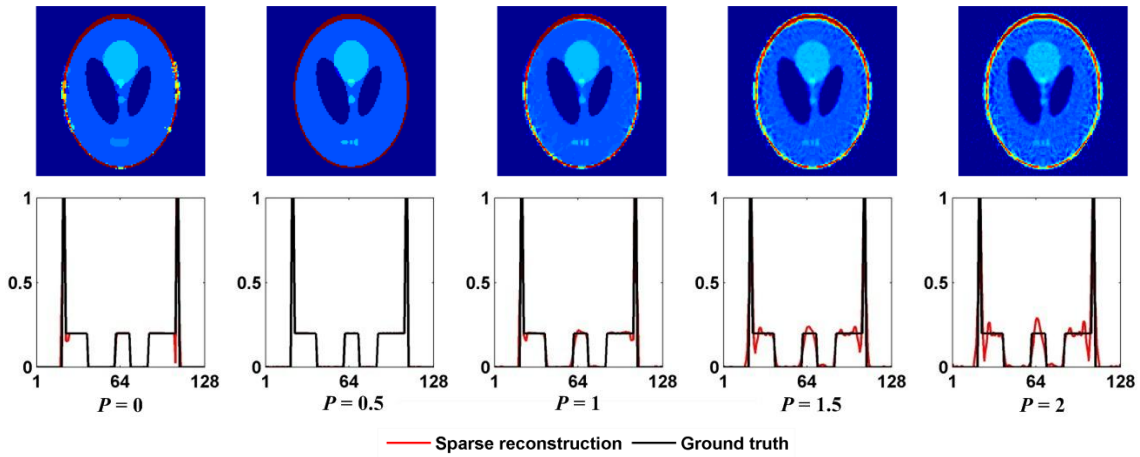


Figure 4-5. Reconstructed TOF-PET images and their associated horizontal profiles corresponding to different values of p = 0, 0.5, 1, 1.5 and 2 for p-TV.

Table 4-2. RMSEs and (1-SSIM)s that correspond to Figure 4-5

P	0.0	0.5	1.0	1.5	2.0
RMSE	0.1658	0.0001	0.1515	0.2094	0.2303
1-SSIM	0.0209	5.72×10^{-6}	0.0434	0.1490	0.2329

From Figure 4-6, we can observe that p = 0.5 is capable of achieving a solution that bests matches the observed data; consequently, it has the strongest capability of avoiding trapping in the local minimum. At the same time, Figure 4-7(a) and (b) give the dependences of RMSE and 1-SSIM on the values of p varying from 0 through 2 with interval of 0.1, which support the standpoint that p = 0.5 is the best candidate to achieve good reconstruction result for our problem. In summary, we can conclude that p = 0.5 provides best TOF-PET reconstruction from Figure 4-5 through Figure 4-7.

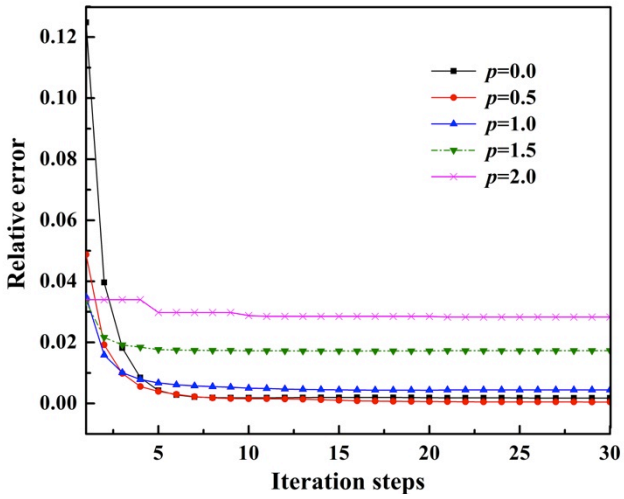


Figure 4-6. Convergence curves of the sparsity-promoted algorithm for different values of p = 0, 0.5, 1, 1.5 and 2 for p-TV.

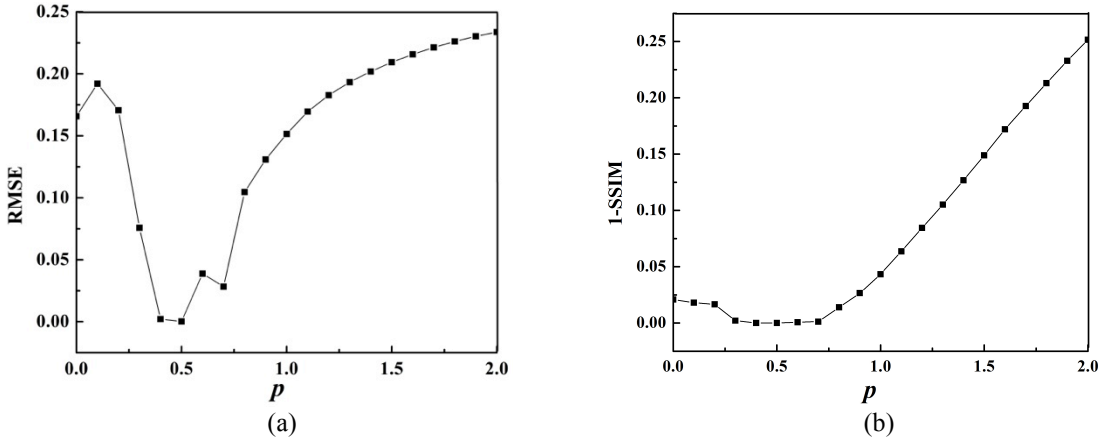


Figure 4-7. Reconstruction performances with the different choices of p value for p-TV under configuration A with 70 detectors. (a) RMSE versus p, and (b) 1-SSIM versus p.

4.4 Numerical Test 3: Sparsity-promoted Algorithm with Combined Regularizations Versus EM Algorithm

In here a set of numerical simulations under configuration A are carried out to demonstrate the superiority of the proposed sparsity-promoted solver with combined regularizations of l_1-norm and p-TV over the traditional EM algorithm, where the cases

with and without TOF information are considered.

The modality without TOF information is considered at first. The results are shown as follows. Figure 4-8(a) shows the images and their associated horizontal profiles obtained through the proposed sparsity-promoted algorithm under configuration *A* using 50, 70, 110, 190 and 270 detectors. For comparison, the corresponding results obtained through the traditional EM algorithm are reported in Figure 4-8(b).

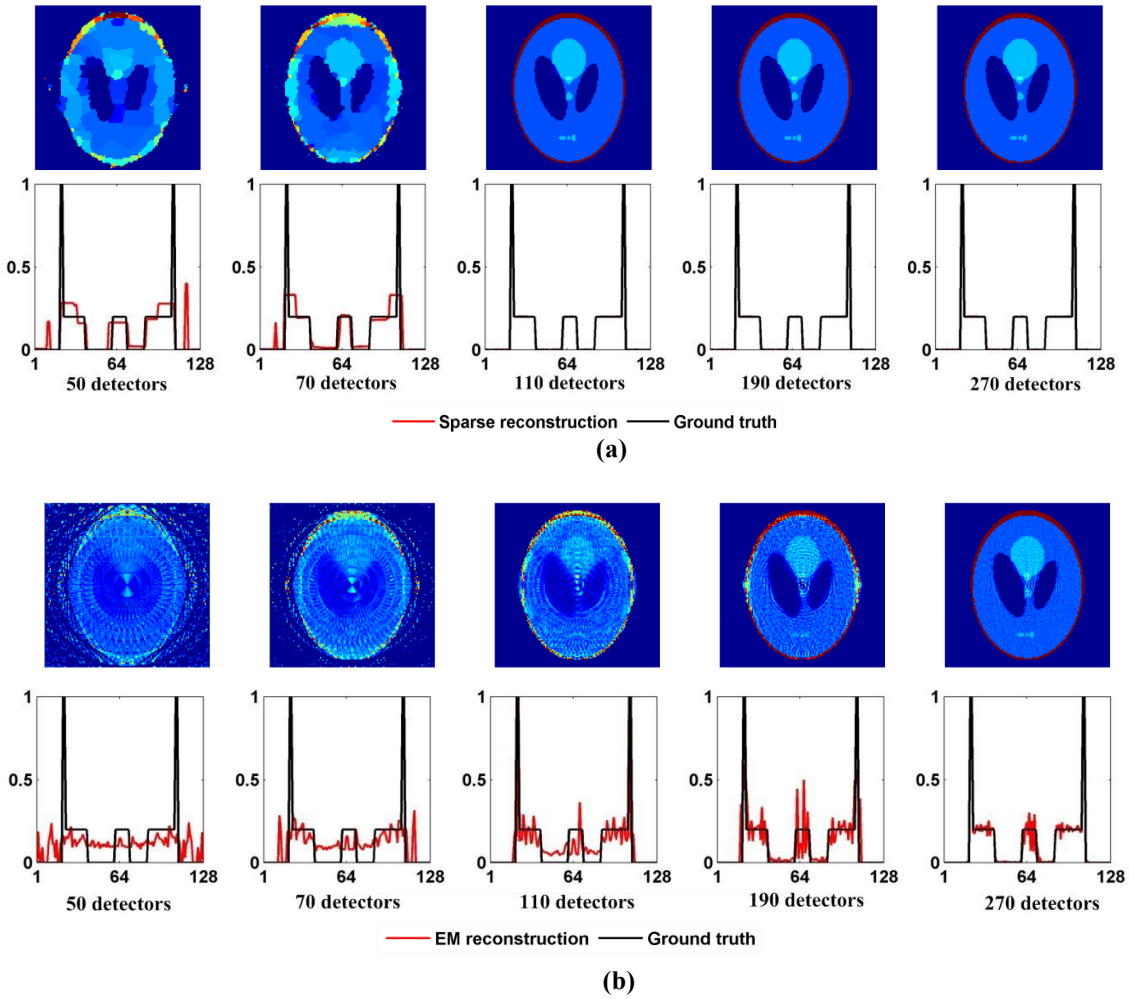


Figure 4-8. PET images reconstructed using (a) the proposed sparsity-promoted and (b) the traditional EM algorithms for different number of detectors used. The corresponding horizontal profiles are also illustrated via red solid line.

A similar analysis on the modality with TOF information is also carried out, where the time resolution and TOF bin are set as 500 *ps* and 67 *ps*, respectively. The reconstructed images and their associated horizontal profiles through proposed sparsity-promoted and EM algorithms are shown in Figure 4-9(a) and 4-9(b), respectively. The RMSEs and (1-SSIM)s corresponding to the reconstructed PET images in Figure 4-8 and Figure 4-9 are reported in Table 4-3.

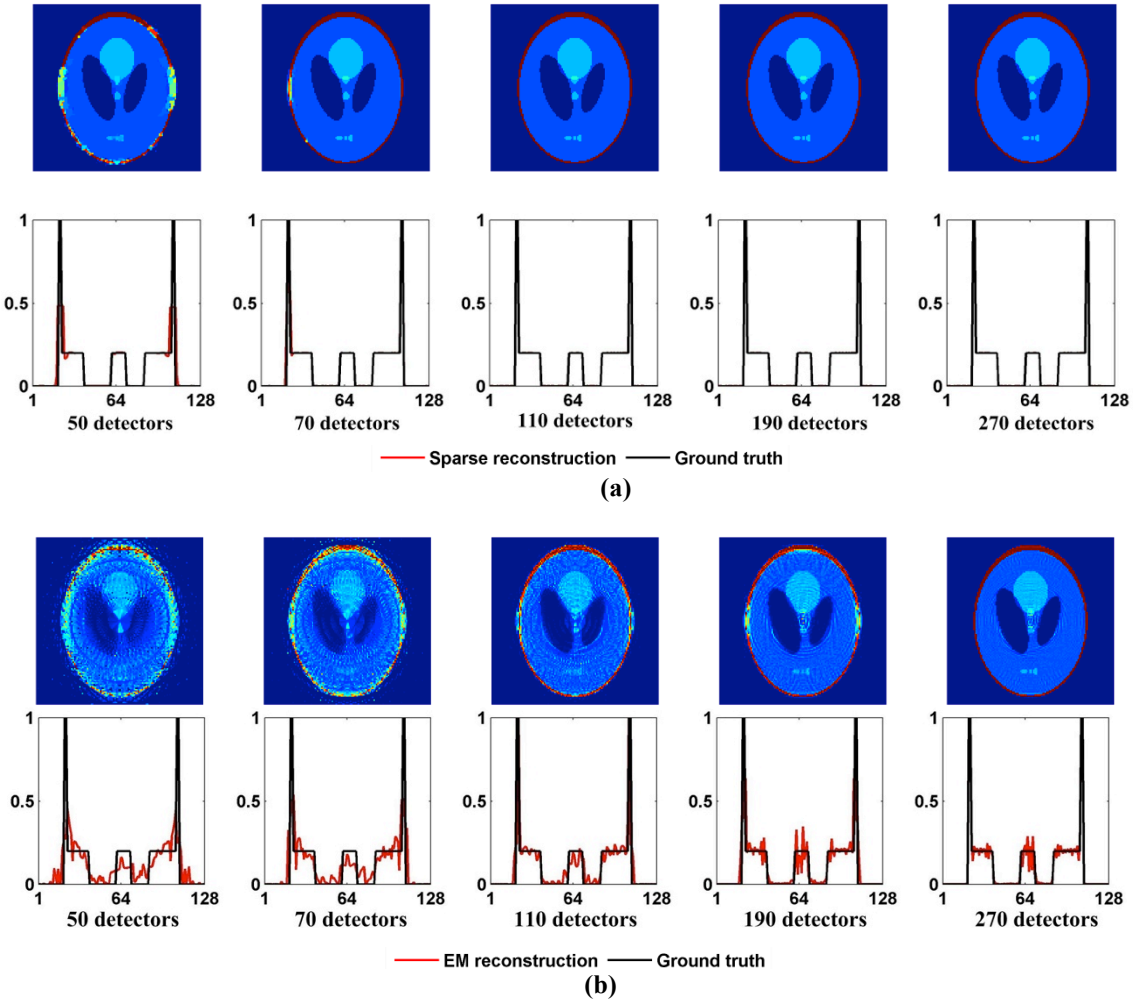


Figure 4-9. TOF-PET images reconstructed using (a) the proposed sparsity-promoted and (b) the traditional EM algorithms for different number of detectors used. The corresponding horizontal profiles are also illustrated via red solid line.

Table 4-3. RMSEs and (1-SSIM)s that correspond to Figure 4-8 and Figure 4-9

Number of detector		50	70	110	190	270
Non-TOF EM	RMSE	0.4563	0.3994	0.2364	0.1797	0.0338
	1-SSIM	0.7547	0.3144	0.2142	0.1681	0.0470
Non-TOF Sparse	RMSE	0.3976	0.3126	4.44×10^{-4}	1.92×10^{-4}	1.8×10^{-4}
	1-SSIM	0.1119	0.1251	6.08×10^{-6}	5.40×10^{-6}	5.33×10^{-6}
TOF EM	RMSE	0.3095	0.2515	0.2156	0.2177	0.0320
	1-SSIM	0.2361	0.1706	0.1445	0.0914	0.0351
TOF Sparse	RMSE	0.2599	0.1046	3.24×10^{-4}	1.79×10^{-4}	1.66×10^{-4}
	1-SSIM	0.0407	0.0024	5.54×10^{-6}	5.33×10^{-6}	5.31×10^{-6}

From the above results, it can be easily verified that TOF information is useful by enhancing image quality. More importantly, the sparsity-promoted approach to PET imaging performs much better than the EM algorithm, in the sense of whether visual observation or RMSE and 1-SSIM. Specifically, for TOF-PET imaging the sparse solver can provide a visually acceptable solution when 70 detectors are used, in contrast, 190 detectors are required by traditional EM algorithm. Even in the absence of TOF information, the sparsity-promoted approach can provide almost exact reconstruction with only 110 detectors; however, the reconstruction images gotten from EM algorithm suffer from serious salt and pepper noise, even with 270 detectors.

From Table 4-3, we can quantitatively observe the benefits that rendered by the sparsity prior and TOF information. Overall, the RMSE and 1-SSIM levels of the reconstructed images using the sparsity-promoted algorithm are much lower than those obtained using the EM algorithm. For instance, with 110 detectors used, RMSEs of images reconstructed by EM algorithm are 0.2156 and 0.2364 for the modalities with and without TOF information. However, they can be enhanced to 3.24×10^{-4} and

4.44×10^{-4}, respectively, with the sparsity-promoted algorithm implemented. For the cases of much fewer detectors, e.g., 50 detectors and 70 detectors, the sparsity-promoted algorithm still significantly outperforms EM algorithm.

Regarding their associated horizontal profiles shown in Figures 4-8(a), 4-9(a) and 4-8(b), 4-9(b) corresponding to the reconstructed images from sparsity-promoted and EM algorithms, we can observe that the profiles obtained from the sparsity-promoted algorithm are more exact and smooth than those obtained from EM algorithm, and almost indistinguishable from the ground truth with more than 110 detectors used. Furthermore, it demonstrates the capability of the sparsity-promoted algorithm in reducing the artifacts produced in reconstruction process.

From the above results we can conclude that the combination of TOF information and sparsity-promoted algorithm is expected to bring a substantial improvement on the reconstruction quality.

4.5 Numerical Test 4: Investigation on Undersampling Configuration *A*

Through a set of numerical experiments carried out under configuration *A*, the relationship between the RMSE and 1-SSIM of images reconstructed through sparsity-promoted algorithm, the time resolution τ of TOF-PET system, and the numbers of detectors *N* will be studied. The basic purpose of this investigation is to find the sufficient condition of getting acceptable TOF-PET reconstructions with limited detectors used. It is appropriate to mention that the parameters of γ_1, γ_2 and γ_3 in our

sparsity-promoted algorithm are empirically tuned and set to be 1×10^{-2}, 1×10^{-4}, and 1×10^{-8}, respectively.

Figure 4-10 shows the reconstructed results with different time resolutions of 100 *ps*, 700 *ps*, 1300 *ps*, 1900 *ps* and 2500 *ps* where 70 detectors were used under configuration *A*. Their associated horizontal profiles are also provided in Figure 4-10. Their corresponding RMSEs and (1-SSIM)s are reported in Table 4-4. From Figure 4-10 and Table 4-4 one can observe, as expected, that the reconstruction quality gets worse as the time resolution increases.

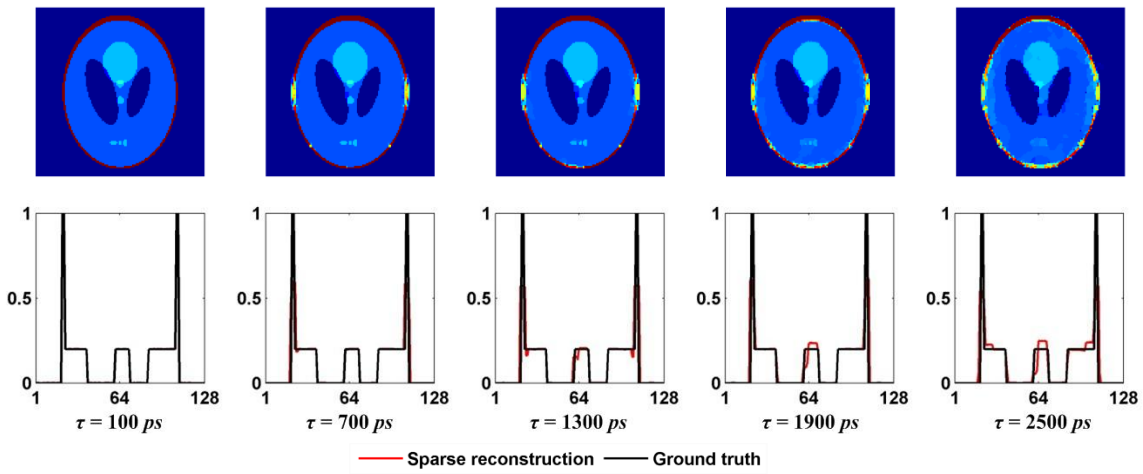


Figure 4-10. TOF-PET images reconstructed through sparsity-promoted algorithm and their corresponding horizontal profiles for different TOF time resolutions under configuration *A* using 70 detectors.

Table 4-4. RMSEs and (1-SSIM)s that correspond to Figure 4-10

Time resolution (*ps*)	100	700	1300	1900	2500
RMSE	1.99×10^{-4}	0.1887	0.2263	0.2274	0.2422
1-SSIM	5.04×10^{-6}	0.0067	0.0148	0.0284	0.0463

The generic dependences of RMSEs and (1-SSIM)s on the number of detectors N and the TOF time resolution τ are shown in Figure 4-11(a) and (b), respectively, where x-axis denotes the time resolution τ in ps while y-axis for number of detectors N. The TOF time resolution τ ranges from 100 ps to 3000 ps, with interval of 100 ps, while N varies from 50 through 110 with interval of 10.

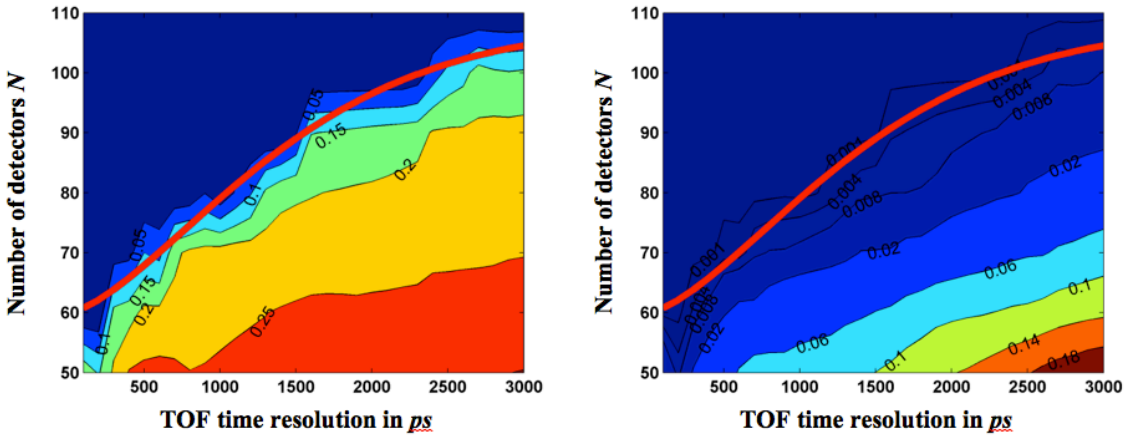


Figure 4-11. Contours of (a) RMSE and (b) 1-SSIM of TOF-PET reconstructed images as a function of numbers of detectors N and TOF resolution τ, for configuration A. The red solid lines represent the phase transition curves with a RMSE threshold of 0.1.

From Figure 4-11 we can easily observe that: (a) the more the detectors, the better the reconstruction, and (b) the higher the TOF time resolution, the better the reconstruction. More importantly, we can choose suitable critical values for RMSE and 1-SSIM as a threshold according to the requirements of reconstruction quality; therefore, for each τ there is a critical value N_c above which the reconstructed TOF-PET images can be considered acceptable. Appealingly, it is desirable to get the explicit dependence of N_c on τ for given threshold values of RMSE or 1-SSIM. For instance, if the threshold

of RMSE is chosen as 0.1, after carrying out standard least square method $N_c(\tau)$ can be derived as

$$N_c(\tau) = 108 - 48\exp(-0.016\tau^{1.5}) \tag{4.4}$$

For convenience, this curve represented by Eq. (20) is referred as transition phase curve, which has been superposed in Figure 4-11(a) by red line. Through this curve, we can make the trade-off between the number of detectors used and time resolution of the system for the manufacture of PET systems. In addition, this curve is very close to that obtained by the similar analysis on 1-SSIM with the threshold of 0.004, as shown by red line in Figure 4-11(b).

4.6 Numerical Test 5: Investigation on Undersampling Configuration *B*

With almost the same computational setup as those used in numerical test 4, the parallel numerical investigation under configuration *B* was conducted to explore the relationship between the reconstruction performance represented by RMSE and 1-SSIM, TOF time resolution *t*, and the range of angle φ. Different from numerical test 4, γ_1, γ_2 and γ_3 are set to be 1×10^{-3}, 1×10^{-4}, and 1×10^{-8}, respectively.

Figure 4-12 shows the TOF-PET reconstructed images through sparsity-promoted algorithm with the range of angle fixed at 60^o and corresponding horizontal profiles for TOF time resolutions of 100 *ps*, 700 *ps*, 1300 *ps*, 1900 *ps* and 2500 *ps*. Their associated RMSEs and (1-SSIM)s are also reported in Table 4-5. From Figure 4-12 and Table 4-5, we can get the following two conclusions: (a) for a fixed range of angle, the reconstruction quality gets worse as the time resolution increases, similar as the one

drawn in numerical test 4; (b) the reconstructions close to the upper and bottom edges of phantom are distorted, mainly resulting from deficient rays traveling through these two regions.

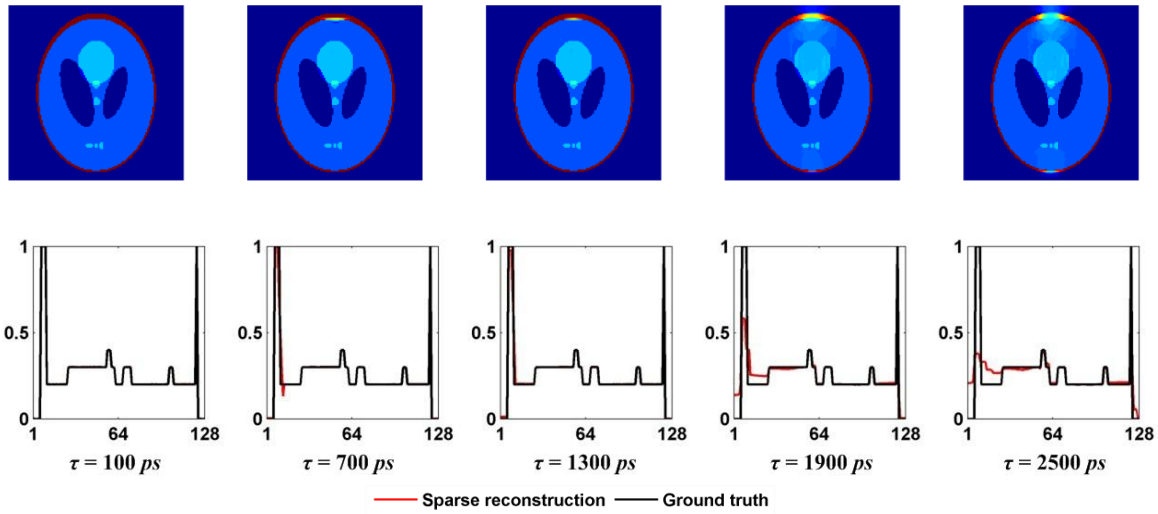


Figure 4-12. TOF-PET images and their corresponding vertical profiles for different time resolutions, under configuration *B* with range of angle being 60^{o}.

Table 4-5. RMSEs and (1-SSIM)s that correspond to Figure 4-12

Time resolution (*ps*)	100	700	1300	1900	2500
RMSE	1.79×10^{-4}	0.0763	0.0527	0.1832	0.2584
1-SSIM	4.93×10^{-6}	6.84×10^{-4}	3.58×10^{-4}	0.0082	0.0164

The relations between RMSEs and (1-SSIM)s of reconstructed images, the range of angle φ and the TOF resolution τ have been illustrated in Figure 4-13(a) and (b), respectively. Here, the range of angle varies between 40^{o} and 110^{o} with interval of 10^{o}. In Figure 4-13, the *x*-axis denotes the time resolution τ in *ps* while the range of angle φ is for *y*-axis. From Figure 4-13, we notice that for each τ there is a critical value ϕ_c

above which the reconstructed TOF-PET images can be acceptable. Similarly, if the threshold of RMSE is specified as 0.05, the phase transition curve of $\phi_c(\tau)$ can be fitted into

$$\phi_c(\tau) = -510 - 3284a\tan(9\tau + 300) \tag{4.5}$$

Again this curve is close to that fitted for 1-SSIM with threshold of 0.0001. These curves have been superposed in Figure 4-13(a) and (b) by solid red line, respectively.

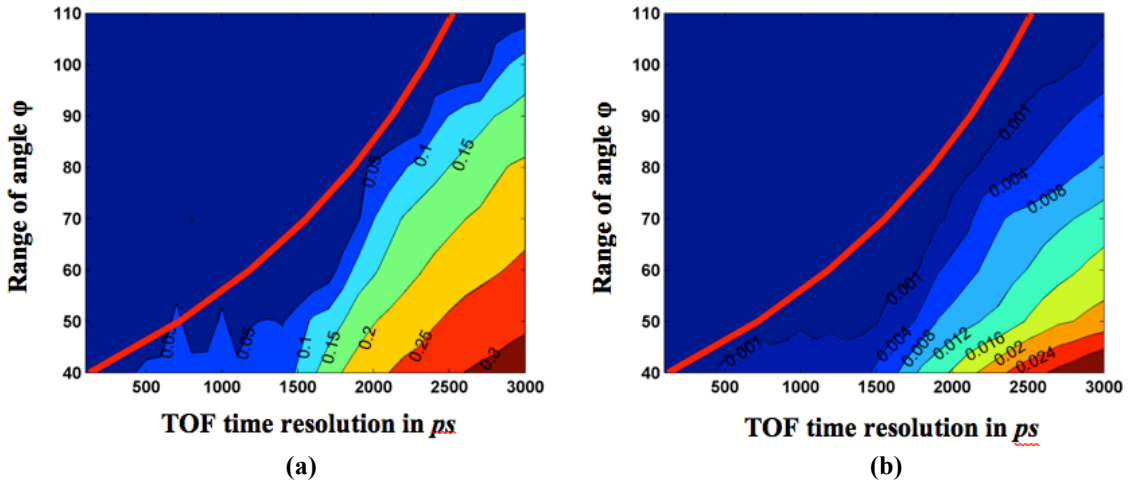


Figure 4-13. Contours of (a) RMSE and (b) 1-SSIM of TOF-PET reconstructed images as a function of the range of angle φ, and the TOF resolution τ, for configuration *B*. The red solid lines represent the phase transition curves with a RMSE threshold of 0.05.

Finally it should be pointed out that besides the dependence on specific choice of thresholds involved in Eqs. (4.4) and (4.5) to meet the specific requirement on imaging quality, these phase transition curves are also highly dependent on the structural complexity (or information content) of the phantom used. In the context of CS, the "sparsity degree" can be used to measure it. Therefore, it is appealing and instructive to carry out the further analyses of $N_c(\tau)$ and $\phi_c(\tau)$ on more phantoms.

5. CONCLUSIONS

Owing to the considerations of reducing the cost of data sampling and improving the reconstruction quality, the benefits brought by sparsity-promoted reconstruction to two undersampling TOF-PET configurations were studied. The findings obtained in this research are summarized as follows;

(1) An efficient algorithm for the combined sparse regularizations to deal with TOF-PET imaging problem was developed, where the combined regularizations of *p*-TV and l_1-norm were used to take advantage of the piecewise constant and structure-correlated sparse properties of the images. Additionally, the l_1-norm regularization was found to suppress the artifacts introduced by *p*-TV, while *p*-TV helped reduce the edge-blurring affect brought by DCT used in l_1-norm. The results showed that the proposed approach is superior to the traditional EM algorithm.

(2) It is well known that 0-TV and 1-TV have been widely used to promote the piece-wise constant of medical images. In here the *p*-TV was investigated with values of *p* varying from 0 to 2 with an interval of 0.1. The results showed that a *p* value of 0.5 was optimal for our undersampling TOF-PET imaging problem.

(3) A general framework of sparsity-promoted ART was proposed and further improved by incorporating the accelerated techniques of multi-step and order-set. Simulation results showed that the proposed sparsity-promoted ART was capable of providing better reconstruction than the traditional ART algorithm. In addition, a super-convergence performance was achieved.

(4) The relationship between the reconstruction quality, number of detectors (or the range of angle) and TOF time resolution was built, which provides an empirical guide for designing novel low-cost TOF-PET systems while ensuring good reconstruction quality.

To summarize, the representative numerical experiments made in this research have shown that sparsity-promoted approaches can significantly improve the performance of the PET imaging, which demonstrated the promising applicability of undersampling TOF-PET imaging.

REFERENCES

[1] D. G. Jamieson, J. H. Greenberg, "Positron emission tomography of the brain," *Comput. Med. Imaging Graphics*, vol. 13, no. 1, pp. 61-79, 1989.

[2] J. M. Ollinger and J. A. Fessler, "Positron-emission tomography," *IEEE Signal Process. Mag.*, vol. 14, no. 1, pp. 43-55, January 1997.

[3] H. Jadvar and J. A. Parker, *Clinical PET and PET/CT*, New York: Springer Press, 2005.

[4] T. Torizuka, N. Tamaki, T. Inokuma, Y. Magata, S. Sasayama, etc., "In vivo assessment of glucose metabolism in hepatocellular carcinoma with FDG-PET," *J. Nucl. Med.*, vol. 36, no. 10, pp. 1811-1817, 1995.

[5] M. Conti, B. Bendriem, M. Casey, M. Chen, F. Kehren, etc., "First experimental results of time-of-flight reconstruction on an LSO PET scanner," *Phys. Med. Biol.*, vol. 50, pp. 4507-4526, 2005.

[6] J. S. Karp, S. Surti, M. E. Daube-Witherspoon, and G. Muehllehner, "Benefit of Time-of-Flight in PET: Experimental and Clinical Results," *J. Nucl. Med.*, vol.49, no. 3, pp. 462-470, March, 2008.

[7] G. Muehllehner and J. S. Karp, "Positron Emission Tomograpy," *Phys. Med. Biol.*, vol. 51, pp. R117-R137, 2006.

[8] M. S. Judenhofer, H. F. Wehrl, D. F. Newport, C. Catana, S. B. Siegel, etc., "Simultaneous PET-MRI: a new approach for functional and morphological imaging," *Nat. Med.*, vol. 14, no. 4, pp. 459-465, April 2008.

[9] S. Surti and J. S. Karp, "Design considerations for a limited angle, dedicated breast, TOF PET scanner," *Phys. Med. and Biol.*, vol. 53, pp. 2911-2921, 2008.

[10] E. Candes, J. Romberg and T.Tao, "Stable signal recovery from incomplete and inaccurate measurements", *Commun. Pure Appl. Math.*, vol.59, pp.1027-1223, 2006.

[11] B. F. Logan, "properties of high-pass signals," Ph. D. Dissertation, Columbia University, New York, 1965.

[12] F. Santosa and W. W. Symes, "Linear Inversion of Band-limited Reflection Seismograms," *SIAM J. Sci. Stat. Comput.*, vol. 7, no. 4, pp. 1307-1330, October 1986.

[13] D. L. Donoho and P. B. Stark, "Uncertainty Principles and Signal Recovery," *SIAM J. Appl. Math*, vol 49, no. 3, pp. 906-931, June 1989.

[14] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal recognition from highly incomplete frequency information," *IEEE Trans. Info. Theory*, vol. 52, no. 2, pp. 489-509, February 2006.

[15] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies," *IEEE Trans. Info. Theory*, vol. 52, no. 12, pp. 5406-5425, Dec. 2006.

[16] E. J. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969-985, 2007.

[17] S. Mallat, *A Wavelet Tour of Signal Processing*, London: Academic-Press, 1998.

[18] M. Elad, *Sparse and Redundant Representations*, New York: Springer-Press, 2010.

[19] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation", *IEEE Trans. Signal Prcoess.*, Vol.54, pp.4311, Nov. 2006.

[20] K. Lee, S. Tak and J. Ye, "A data-driven sparse GLM for fMRI analysis using sparse dictionary learning with MDL criterion", *IEEE Trans. Med. Imag.*, vol.30, pp.1076-1089, May, 2011.

[21] S. Ravishankar and Y. Bresler, "MR image reconsruction from highly undersampled k-space data by dictionary learning", *IEEE Trans. Med. Imag.*, vo.30, pp.1028-1041, 2011.

[22] M. Lustig, D. Donoho, and J. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Reson. Med.*, vol. 58, pp. 1182–1195, April 2007.

[23] J. Bian, J. Siewedsen, X. Han , E. Y. Sidky, J. L. Prince, et al., "Evaluatio of sparse-view reconstruction from flat-panel-detector cone-beam CT", *Phys. Med. Biol.*, vol.55, pp. 6575, 2010.

[24] X. Han, J. Bian, D. R. Eaker, T. L. Kline, E. Y. Sidky, et al., "Algorithm-enabled low-dose micro-CT imaging", *IEEE Trans. Med. Imag.*, vol.30, pp.606-620, March, 2011.

[25] Z. T.Harmany, R. F. Marcia and R. M. Willett, "Sparsity-regularized photon-limited imaging," *IEEE International Symposium on Biomedical Imaging From Nano to Macro*, pp. 772–775, 2010.

[26] G. Wang, and J. Qi, "Direct reconstruction of dynamic PET parameteric images using sparse spectral representation," *Proc IEEE Int. Symp. Biomed. Imaging*, pp. 867-870, 2009.

[27] S. Ahthoine, J. F. Aujol, Y. Broursier and C. Melot, "On the efficiency of proximal methods for CBCT and PET reconstruction with sparity constraint," *25, 4TH Workshop on signal processing with adaptive sparse structured representations*, 2011.

[28] A. C. Kak and M. Slaney, *Principle of Computed Tomography*, IEEE Press, 1988.

[29] S.Vandenberghe, Y. D'Asseler, R. Van de Walle, T. Kuppinen, M. Koole, etc., "Iterative reconstruction algorithms in nuclear medicine," *Comput. Med. Imaging Graphics*, vol. 25, pp. 105-111, 2001.

[30] S. Cho, S. Ahn, Q. Li and R. M. Leahy, etc., "Analytical properties of time-of-flight PET data," *Phys. Med. Biol.*, vol. 53, pp. 2809-2821, 2008.

[31] A. Mallon, and P. Grangeat, "Three-dimensional PET reconstruction with time-of-flight measurement," *Phys. Med. Biol*, vol. 37, no. 3, pp. 717-729, 1992.

[32] N. Baba and K. Murata, "Filtering for image reconstruction from projections," *J. Opt. Sot. Amer.*, vol. 67, pp. 662-668, 1977.

[33] U. Tun, S. Peltonen and U. Ruotsalainen, "Gap-Filling for the High-Resolution PET Sinograms With a Dedicated DCT-Domain Filter," *IEEE Trans. Med. Imag.*, vol. 29, no. 3, pp. 830-839, March 2010.

[34] D. A. Chesler and S. J. Riederer, "Ripple Suppression during Reconstruction in Transverse Tomography," *Phys. Med. Biol.*, vol. 20, no. 4, pp. 632-636, 1975.

[35] J. G. Rogers, R. Harrop and P. E. Kinahan, "the Theory of Three-Dimensional Image Reconstruction for PET," *IEEE Trans. on Med. Imag.*, vol. MI-6, no. 3, September 1987.

[36] P. E. Kinahan and J. G. Rogers, "Analytic 3D image reconstruction using all detected events," *IEEE Trans. Nucl. Sci.*, vol. 36, pp. 964-968, 1989.

[37] M. Defrise, P. E. Kinahan, D. W. Townsend, C. Michel, M. Sibomana and D. F. Newport, "Exact and Approximate Rebinning Algorithms for 3-D PET Data," *IEEE Trans. on Med. Imag.*, vol. 16, no. 2, April 1997.

[38] M. Defrise, M. Sibomana, C. Michel, and D. Newport, "3-D PET reconstruction with the ECAT EXACT HR using Fourier rebinning," in *Conf. Rec. IEEE 1995 Nucl. Sci. Symp.*, San Francisco, CA, 1996, pp. 1316–1320.

[39] M. E. Daube-Witherspoon and G. Muehllehner, "Treatment of axial data in 3D PET," *J. Nucl. Med.*, vol. 28, pp. 1717-1724, 1987.

[40] X. Liu, C. Comtat, C. Michel, P. Kinahan, M. Defrise, etc., "Comparison of 3-D Reconstruction with 3D-OSEM and with FORE+OSEM for PET," *IEEE Trans. on Med. Imag.*, vol. 20, no. 8, pp. 804-814, August 2001.

[41] G. T. Herman and L. B. Meyer, "Algebraic Reconstruction Techniques Can Be Made Computationally Efficient," *IEEE Trans. Med. Imag.*, vol. 12, no. 3, September 1993.

[42] X. Xu, J. Liow and S. C. Strother, "Iterative algebraic reconstruction algorithms for emission computed tomography: A unified framework and its application to positron emission tomography," *Med. Phys.*, vol. 20, no. 6, pp. 1675-1684, December, 1993.

[43] D. Blatt, A. Hero and H. Gauchman, "A Convergent Incremental Gradient Method with a Constant Step Size," *SIAM J. Optimization*, vol. 18, pp. 29-51, 2008.

[44] S. Boyd and L. Vandenberghe, *Convex Optimization*, New York: Cambrige University Press, 2004.

[45] J. A. Fessler and S. D. Booth, "Conjugate-Gradient Preconditioning Methods for Shift-Variant PET Image Reconstruction," *IEEE Trans. Image Process.*, vol. 8, no. 5, pp. 688-699, May 1999.

[46] G. Gindi, M. Lee, A. Rangarajan, G. I. Zubal, "Bayesian Reconstruction of Functional Images Using Anatomical Information as Priors," *IEEE Trans. Med. Imag.*, vol. 12, no. 4, pp. 670-680, December 1993.

[47] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for Emission Tomography," *IEEE Trans. Med. Imag.*, vol. MI-1, no. 2, pp. 113-122, October 1982.

[48] L. Kaufman, "Maximum Likelihood, Least Squares, and Penalized Least Squares for PET," *IEEE Trans. Med. Imag.*, vol. 12, no. 2, pp. 200-214, June 1993.

[49] M. Bierlaire, Ph. Toint, and D. Tuyttens, "On iterative algorithms for linear least squares problems with bound constraints," *Linear Algebra Appl.*, pp. 111-143, 1991.

[50] T. Hebert and R. Leahy, "A Generalized EM Algorithm for 3-D Bayesian Reconstruction from Poisson Data Using Gibbs Priors," *IEEE Trans. Med. Imag.*, vol. 8, no. 2, pp. 194-202, 1989.

[51] R. Leahy and X. Yan, *Incorporation of Anatomical MR Data for Improved Functional Imaging with PET*, in *Information Processing in Medical Imaging*, A. C. F. Colchester and D. J. Hawks, eds., pp. 105-120. New York: Springer-Verlag, 1991.

[52] P. J. Green, "Bayesian Reconstruction from Emission Tomography Data Using a Modified EM Algorithm," *IEEE Trans. Med. Imag.*, vol. 9, no. 1, pp. 84-93, March 1990.

[53] S. Geman and D. E. McClure, "Bayesian image analysis: An application to single photon emission tomography," *Proc. Amer. Statist. Assoc., Stat. Comp., sect.*, 12-18.

[54] H. M. Hudson and R. S. Larkin, "Accelerated Image Reconstruction Using Ordered Subsets of Projection Data," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601-609, December 1994.

[55] R. Yao, J. Seidel, C. A. Johnson, M. E. Daube-Witherspoon, M. V. Green, etc., "Performance Characteristics of the 3-D OSEM Algorithm in the Reconstruction of Small Animal PET Images," *IEEE Trans. Med. Imag.*, vol. 19, no. 8, pp. 798-804, August 2000.

[56] R. Baraniuk, M. A. Davenport, M. F. Duarte and C. Hegde, "An introduction to compressive sensing," Connexions, Rice University, Houston, Texas. On line: http://cnx.org/content/col11133/1.5/

[57] Y. Nesterov, "Gradient methods for minimizing composite objective function," CORE discussion paper, no. 76, University of Catholic Louvain, Belgium, Sept. 2007.

[58] Y. Nesterov, *Introductory Lectures on Convex Optimization – A Basic Course*, Boston: Kluwer Academic Publishers, 2004.

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

[60] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol.1, pp. 211-244, 2001.

[61] I. Daubechies, R. DeVore, M. Fornasie and C. S. Gunturk, "Iteratively re-weighted least squares minimization for sparse reconvery," *Commun. Pure Appl. Math.*, vol. 63, no. 1, pp. 1-38, 2010.

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

[63] E. Candès and J. Romberg, "l_1-MAGIC: Recovery of sparse signals via convex programming," *Caltech*, October 2005.

[64] S. Becker, J. Bobin and E. Candes, "NESTA: a fast and accurate first-order method for sparse recovery," *SIAM J. Imag. Sci.*, vol. 4, no. 1, pp. 1-39, 2011.

[65] G. Wang and M. Jiang, "Ordered-subset simultaneous algebraic reconstruction techniques (OS-SART)," *J. X-ray sci. technol.*, vol. 12, pp. 169-177, 2004.

[66] G. H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med. Phys.*, vol. 35, no. 2, pp. 660–663, 2008.

[67] E. Y. Sidky, and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, pp. 4777–4807, 2008.

[68] Xin jin, L. Li, Z. Chen, L. Zhang and Y. Xing, "Anisotropic total variation for limited-angle CT reconstruction," *Nuclear Science Symposium Conference Record (NSS/MIC), 2010 IEEE*, pp. 2232-2238, 2010.

[69] P. Rodriguez and B. Wohlberg, "Efficient minimization method for a generalized total variation functional", *IEEE Trans. Image Process.*, vol.18, pp.322-332, Feb. 2009.

[70] S. Richter and R. DeCarlo, "Continuation methods: Theory and applications", *IEEE Trans. Autom. Control*, vol. 28, no. 6, pp. 660–665, 1983.

[71] Z. Wang and A. C. Bovik, "Image quality assessment: from error visibility to structural similarity", *IEEE Trans. Image Process.*, vol. 13, no. 4, pp.1-14, 2004.

[72] M. Zhdanov and E. Tolstaya, "Minimum support nonlinear parameterization in the solution of a 3D magnetotelluric inverse problem", *Inverse problems*, vol. 20, pp. 937-952, 2004.

[73] L. Rudin, S. Osher and E. Fatemi, "Nonlinear total variation based noise removal algorithms", *Physics D*, vol. 60, pp. 259-268, July, 1992.

VITA

Name: Dapeng Lao

Address: 337 Zachry Engineering Center,
3133 TAMU,
College Station, TX 77843

Email Address: lakobers@neo.tamu.edu

Education: B.A., Tsinghua University, 2009