A Mathematical Model for Extremely Low Dose Adaptive Computed Tomography Acquisition

Oren Barkan[1], Amir Averbuch[1], Shai Dekel[2,3], and Yaniv Tenzer[3]

[1] School of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel
[2] School of mathematical sciences, Tel-Aviv University, Tel-Aviv 69978, Israel
[3] GE Healthcare, Israel

Abstract. One of the main challenges in Computed Tomography is to balance the amount of radiation exposure to the patient at the time of the scan with high image quality. We propose a mathematical model for adaptive Computed Tomography acquisition whose goal is to reduce dosage levels while maintaining high image quality at the same time. The adaptive algorithm iterates between selective limited acquisition and improved reconstruction, with the goal of applying only the dose level needed for sufficient image quality. The theoretical foundation of the algorithm is nonlinear Ridgelet approximation and a discrete form of Ridgelet analysis is used to compute the selective acquisition steps that best capture the image edges. We show experimental results where the adaptive model produces significantly higher image quality, when compared with known non-adaptive acquisition algorithms, for the same number of projection lines.

Keywords: Adaptive compressed sensing, Ridgelets.

1 Introduction

In the last decade, several studies have shown that radiation exposure during Computed Tomography (CT) scanning is a significant factor in raising the total public risk of cancer deaths [3], [29], [34]. To balance image quality with these concerns, radiologists use the protocol As Low as Reasonably Achievable (ALARA). It meant to ensure that "... CT dose factors are kept to a point where risk is minimized for maximum diagnostic benefit..", where the dose can be determined by the product of the CT tube current and the time the patient has been exposed to the radiation (see [26] for an overview). Currently, there are several state-of-the-art technologies that attempt to achieve dose reduction. There are the Iterative Reconstruction (IR) methods which are successful in reducing artifacts, improving resolution and lowering the noise in the reconstructed images ([10], [35]). More recently, Model Based Iterative Reconstruction (MBIR) [2], [36] was introduced. It improves upon the IR methods by incorporating accurate system physics models coupled with statistical noise models and prior models.

However, dosage levels during CT exams are still at the focus of attention and any new method that can reduce them is considered highly valuable. This paper

M. Floater et al. (Eds.): MMCS 2012, LNCS 8177, pp. 13–33, 2014.

describes an adaptive acquisition model that theoretically is superior to existing non-adaptive acquisition methods and allows minimal and optimal dosage levels. The method can be considered a significant generalization of existing two step adaptive acquisition methods [20], [27] and can potentially use the same hardware configurations that are capable of changing their geometric configuration and acquisition protocols on-the-fly (see also [33]). For example, in [20], the authors describe an imaging C-arm system where a low-dose overview (OV) scan is used to dynamically identify an arbitrary Volume Of Interest (VOI). The OV and VOI scans are then registered and reconstructed together. In [27], the authors develop a flexible x-ray micro-CT system, named FaCT, capable of changing its geometric configuration and acquisition protocol in order to best suit an object being imaged for a particular diagnostic task. In their system, a fast, sparse-projection pre-scan is performed, the data are reconstructed, and the region of interest is identified. Next, a diagnostic-quality scan is performed where, given the region of interest, the control computer calculates an illumination window for on-line control of an x-ray source masking aperture to transmit radiation only through the region of interest throughout the scan trajectory. In going further, for our adaptive CT approach to work, it is required that the system will be able to configure on the fly an iterated projection scheme, beyond the existing predefined uniform parallel or fan beam acquisition patterns.

Observe that adaptive acquisition should not be confused with adaptive reconstruction. In the latter, the acquisition model is a non-adaptive uniform sampling scheme, where over a discrete set of pre-determined angles, projections lines are computed at equal intervals. In this setup, the adaptive elements, if exist, are part of the post-acquisition reconstruction step.

The outline of the algorithm is as follows: First, the system projects the object with an extreme low dose according to a uniform predetermined pattern and reconstructs an initial low quality image. Next, the system predicts from the reconstructed low quality image where the significant edges of the true objects are and projects along them. Then, the system iterates by incorporating the newly added line projections in order to obtain a refined approximation of the true image. The algorithm continues to iterate between estimation of locations of finer significant features, adaptive acquisition and reconstruction until a convergence criterion is met. The goal is to quickly converge until a high quality reconstruction is achieved with minimal dose. Moreover, by using the mathematical model of Ridgelets [4], the algorithm has a natural multiresolution capability, where the significance of edges is analyzed at different scales. We show, in the experimental results section that this approach yields significantly higher image reconstruction quality, when compared with known non-adaptive acquisition algorithms, for the same number of projection lines.

It is important to clarify the following fundamental assumption we make on the acquired images. To illustrate, let I be a bi-level image, i.e. with pixels that are either '0' or '1', where the '1' values are sparse. Even on this simple image, our approach would be rendered useless if the '1' values are scattered in random locations against the background of zeros. In such a case, as clearly explained in

[1], adaptive acquisition has absolutely no advantage over non-adaptive methods (e.g [7], [17]). However, if I is what is called a 'cartoon' image, where the '1' values are grouped into 'nicely' connected subdomains with piecewise smooth boundaries, then the situation changes dramatically. Our method relies on the mathematical theory of [4] which quantifies in the setup of Computed Tomography the geometric 'structure' of the image and how fast a Ridgelet approximation converges to the image. Our algorithm, whose goal is to acquire an unknown image, regards the adaptive Ridgelet approximation of this image as the 'optimal' benchmark and is designed to match its performance. This approach has strong ties with the waveform analysis presented in [31], that allowed the authors to classify singularities and quantify the 'stability' of limited angle tomography. Indeed, although in our work we limit the number of line projections, but do not limit the angles, the fundamental understanding of the relationship between a function's edge singularities and its Radon representation as explained in [31] is at the core of our algorithm (see Fig. 3 and the accompanying explanation).

The paper is organized as follows: Section 2 overviews necessary mathematical background. Section 3 describes in detail our adaptive acquisition algorithm. Experimental results and comparisons with non-adaptive methods are given in Section 4. In the last section we drew conclusions and discuss future work.

2 Preliminaries

2.1 Fast Algorithms for Total Variation Functionals with 'Sparse' Constraints

For a given image $I \in \mathbb{R}^{m\times m}$, with pixels values $\{I_{i,j}\}$, we define the gradient of I by $(\nabla I)_{i,j} = (I_{i,j} - I_{i-1,j}, I_{i,j} - I_{i,j-1})$. The Total Variation (TV) norm of the image is given by

$$|I|_{TV} := \sum_{i,j=2}^{m} (|I_{i,j} - I_{i-1,j}| + |I_{i,j} - I_{i,j-1}|).$$

Denote $N = m^2$, and let $x \in \mathbb{R}^N$, be a one-dimensional representation of I by concatenating the rows of I into a single column vector

$$x = (I_{1,1},\, I_{1,2}, \ldots, I_{1,m}, \ldots, I_{m,1}, \ldots I_{m,m})^T.$$

Given an $n \times N$, $n << N$, sampling matrix $A \in \mathbb{R}^{n\times N}$ and corresponding observations vector $y \in \mathbb{R}^n$, generated by $Ax = y$, the so-called TV-minimization is concerned with solving one of the following optimization problems

$$\min_U |U|_{TV} \quad s.t.\ Au = y, \quad \min_U |U|_{TV} + \mu \left\|Au - y\right\|_2^2, \tag{1}$$

where $u \in \mathbb{R}^N$ is the one-dimensional representation of $U \in \mathrm{R}^{m\times m}$ and μ is a given weight parameter. The right hand side minimization problem is applied in the presence of noise in the sampling process and the weight μ depends, in

part, on the expected noise level. This model is difficult to solve directly due to non-differentiability and non-linearity of the TV term. During the last few years there has been an explosion of new numeric iterative methods (see the papers in the "Compressive Sensing Recovery Algorithms" section of [8]).

Although conceptually our method may use such solvers as black boxes, its unique features allow us to apply critical modifications that not only accelerate the iterative methods, but also make them feasible in large datasets problems when N is large. In this work, we implemented a modified version of the TVAL3 solver [23], [37]. Our modified version utilizes the fact that in our special case the matrix A is highly sparse. This is in complete contrast to the usual setup in compressed sensing, where the theory typically promotes a dense matrix (usually of pseudo-random nature). As we shall see in Section 3, in our case the sparsity is due to the fact that each row of A is associated with an integration over a digital line in the image and therefore a vector of '0's and '1's. The values '1' are located in entries associated with the pixels of the digital line and thus each row in matrix A has $\leq \sqrt{2}m = \sqrt{2N}$ non-zero entries. We note that even if we use a more accurate model based interpolation, where the line is given some width and then the result is a weighted sum of pixels, the matrix A would remain sparse. This structure allows us to store, to adaptively update a sparse data structure for A and to implement fast linear algebra operations. This idea is not new to the CT community. Moreover, for practical clinical data sizes in 3D helical uniform acquisition, the matrix A can be too large to hold in memory and must be computed on the fly. Also, its form is carefully determined from the geometry of the focal spots and detectors [11]. In this work we focus on the 2D model and in future work we plan to investigate whether in the 3D case our smaller adaptive sampling set can be stored in memory or computed on the fly.

We now explain, for the sake of completeness, our modification of the TVAL3 algorithm. For the constrained optimization problem such as (1), there are a number of methods that approach the original constrained problem by a sequence of unconstrained subproblems. One of them is the Quadratic Penalty Method [9]. This method puts a quadratic penalty term instead of the constraint in the objective function where each penalty term is a square of the constraint violation with multiplier. However, this method requires to increase the multipliers to infinity so as to guarantee the convergence, which may cause the ill-conditioning problem, numerically. Another method concerning the constrained optimization problem is the Augmented Lagrangian method [15] (an augmented Lagrange method has been already used in CT reconstruction [32]). According to this method, the corresponding Augmented Lagrangian of the left-hand side minimization in (1), is given by

$$\begin{aligned} L_A(w,u,v,\lambda,\mu,\beta) := \sum_{s=1}^{N} \left(\|w_s\|_1 - \langle v_s, (Du)_s - w_s \rangle + \frac{\beta_s}{2} \|(Du)_s - w_s\|_2^2 \right) \\ - \langle \lambda^t, Au - y \rangle + \frac{\mu}{2} \|Au - y\|_2^2 , \end{aligned} \tag{2}$$

where $w_s, v_s \in \mathbb{R}^2$, $\|w_s\|_1 := |w_s(1)| + |w_s(2)|$, $(Du)_s := \nabla U_{i(s),j(s)}$, $1 \leq s \leq N$, and the two vectors λ,v, are the Lagrangian multipliers. To solve (2),

the following Alternating Direction scheme is used: Denote the approximate minimizers of (2) at the kth inner iteration by $w^{(k)}$ and $u^{(k)}$. Then $w^{(k+1)}$and $u^{(k+1)}$ can be attained by solving two separated subproblems. The first is the 'w-subproblem':

$$\begin{aligned} w^{(k+1)} &= \arg\min_w L_A(w, u^{(k)}) = \\ &\sum_{s=1}^{N} \left(\|w_s\|_1 - \left\langle v_s, \left(Du^{(k)}\right)_s - w_s \right\rangle + \tfrac{\beta_s}{2} \left\| \left(Du^{(k)}\right)_s - w_s \right\|_2^2 \right). \end{aligned} \tag{3}$$

Note that the 'w-subproblem' is separable with respect to each w_s, $1 \le s \le N$, and has a closed form solution [23]. The second subproblem, also known as the 'u-subproblem' is:

$$\begin{aligned} u^{(k+1)} &= \arg\min_u L_A(w^{(k+1)}, u) = \\ &\sum_{s=1}^{N} \left(\left\| w_s^{(k+1)} \right\|_1 - \left\langle v_s, (Du)_s - w_s^{(k+1)} \right\rangle + \tfrac{\beta_s}{2} \left\| (Du)_s - w_s^{(k+1)} \right\|_2^2 \right) \\ &- \langle \lambda^t, Au - y \rangle + \tfrac{\mu}{2} \|Au - y\|_2^2 . \end{aligned} \tag{4}$$

The 'u-subproblem' can be solved using a steepest decent method, but since this might be too costly for large scale problem, an aggressive 'one-step' of the steepest decent can be computed as an iteration (see the details in [23]). After attaining $w^{(k+1)}$ and $u^{(k+1)}$, the multiplier updating is performed based on the analysis of [18], [30]

$$\begin{aligned} v_s^{(k+1)} &= v_s^{(k)} - \beta_s \left(\left(Du^{(k+1)}\right)_s - w_s^{(k+1)} \right), \ 1 \le s \le N, \\ \lambda^{(k+1)} &= \lambda^{(k)} - \mu(Au^{(k+1)} - y). \end{aligned}$$

This second update step is exactly an example of where our modification accelerates significantly the TV minimization, by either storing and applying the matrix A in a sparse form or by computing and applying the sparse rows of A on the fly. Finally, choose new penalty parameters $\beta_s^{(k+1)} \ge \beta_s^{(k)}$ and $\mu^{(k+1)} \ge \mu^{(k)}$. The stopping criteria are one of the following:

(i) The quantities

$$\left| \nabla L_A(w^{(k)}, u^{(k)}, v^{(k)}, \lambda^{(k)}, \mu^{(k)}, \beta^{(k)}) \right|, \ \sum_{s=1}^{N} \left\| \left(Du^{(k)} \right)_s - w_s^{(k)} \right\|_2, \ \left\| Au^{(k)} - y \right\|_2 ,$$

are sufficiently small.

(ii) The relative change $\left\| u^{(k+1)} - u^{(k)} \right\|_2$, is sufficiently small.

Inside the main loop of the Alternating Direction scheme, the number of rows in A is increased by a predetermined fixed constant M at each iteration (see Section 3), where the rows are projection lines determined from Ridgelet analysis of the approximant of the image.

2.2 Fundamentals of Ridgelet Theory

Let $\psi \in L_2(\mathbb{R})$ be a wavelet [24]. For the purpose of this paper it is sufficient that the wavelet function has two properties: compact support and vanishing moments. The latter implies that for some $r \geq 1$,

$$\int_{\mathbb{R}} \psi(x)\, x^l dx = 0, \quad l = 0, \ldots, r-1.$$

The classical example for a wavelet function is the Haar wavelet with one vanishing moment

$$\psi(x) := \begin{cases} 1, & 0 \leq x \leq 1/2, \\ -1, & 1/2 < x \leq 1, \\ 0, & else. \end{cases} \tag{5}$$

A bivariate Ridgelet function [4], [14], is defined by

$$\psi_{a,b,\theta}(x_1, x_2) := a^{-1/2}\psi((x_1 \cos\theta + x_2 \sin\theta - b)/a),$$

where a, b and θ are the parameters determining the scale, transition and rotation of the Ridgelet function, respectively (see Fig. 1).

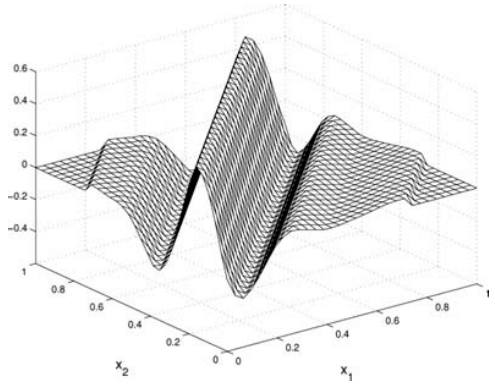


Fig. 1. A Ridgelet function $\psi_{a,b,\theta}(x_1, x_2)$

Given $f \in L_1(\mathbb{R}^2)$, its Continuous Ridgelet Transform (CRT) is defined by

$$CRT_f(a, b, \theta) := \int_{\mathbb{R}^2} \psi_{a,b,\theta}(x) f(x) dx. \tag{6}$$

The continuous Radon transform [19], [28] of a bivariate function f at direction θ is defined as

$$R_f(\theta, t) := \int_{\mathbb{R}^2} f(x_1, x_2)\delta(x_1 \cos\theta + x_2 \sin\theta - t)\, dx_1 dx_2, \tag{7}$$

where δ is the Dirac function. The Radon and the Ridgelet transforms are related by

$$CRT_f(a,b,\theta) = \int_{\mathbb{R}} \psi_{a,b}(t) R_f(\theta,t)\, dt, \tag{8}$$

where $\psi_{a,b}(x) := a^{-1/2}\psi((x-b)/a)$. In applications, this means that the Ridgelet transform can be computed by the application of the Radon transform at a given angle, followed by 1D fast wavelet transform.

It is interesting to point out that Ridgelets [4] did not previously find too many applications in image processing. Their 'descendants' Curvelets [5], [6] and Shearlets [21], [22], which capture directional information as well, were found to be more useful due to their better time-frequency localization. In the context of CT reconstruction, Curvelets have been used as a regularization tool [16]. However, we find that Ridgelets are the right mathematical tool in the setup of Computed Tomography, because the acquisition device is not able to capture through its sampling process, well localized functionals such as Curvelet coefficients.

From approximation theoretical perspective, the mathematical foundation of our adaptive algorithm follows the framework of characterizing the images by the appropriate function smoothness spaces and then providing an estimate for the order of convergence.

Definition 1. *[4] For $\alpha > 0$, and $p, q > 0$, we say that $f \in \dot{R}^{\alpha}_{p,q}\left(\mathbb{R}^2\right)$, if $f \in L_1\left(\mathbb{R}^2\right)$ and*

$$\|f\|_{\dot{R}^{\alpha}_{p,q}} := \left(\sum_{j=-\infty}^{\infty} 2^{j(\alpha+1/2)q} \left(\frac{1}{\pi} \int_0^{\pi} \left\| CRT_f\left(2^j, \cdot, \theta\right) \right\|_p^p d\theta \right)^{q/p} \right)^{1/q} < \infty.$$

We note that this definition requires certain conditions on the wavelet ψ. It is sufficient to assume ψ is compactly supported, is in C^r and has r vanishing moments, with $r > \max(2/p, \alpha + 5/2)$. These conditions ensure that membership in the smoothness space $\dot{R}^{\alpha}_{p,q}$ does not depend on the particular wavelet used in (6). A typical non trivial example for a function in $\dot{R}^{\alpha}_{p,q}$ is a function with a singularity along a line such as

$$f(x_1, x_2) = 1_{\{x_1 > 0\}}(x_1, x_2)\, (2\pi)^{-1/2}\, e^{-\left(x_1^2 + x_2^2\right)/2}.$$

This function is in the Besov class [12] $B^{\alpha}_{1,1}$ only for $\alpha < 1$, which means that it almost has a first derivative in the classical sense. In contrast, this function is contained in $\dot{R}^{\alpha}_{1,1}$, for any $\alpha < 3/2$ [4], which implies that it is smoother in the scale of Ridgelet spaces than in the scale of Classical Besov spaces. This is a direct consequence of the fact that its singularity has simple lower dimensional structure.

In this work we assume that the functions we analyze are compactly supported in a 'standard' compact domain such as $[-1,1]^2$ and attain the value zero on its boundary. Indeed, CT images satisfy this requirement (see the examples below). Therefore, by a simple zero extension argument, a function $f \in L_2\left([-1,1]^2\right)$

of this nature can also be regarded as a function in $L_1\left(\mathbb{R}^2\right) \cap L_2\left(\mathbb{R}^2\right)$. By sampling the CRT, one may obtain a discrete Ridgelet Frame system $\{\psi_\gamma\}$ with a dual system $\left\{\tilde{\psi}_\gamma\right\}$, for a countable index $\{\gamma = (a, b, \theta)\}$, such that for $f \in L_2\left([-1,1]^2\right)$,

$$f = \sum_\gamma \left\langle f, \tilde{\psi}_\gamma \right\rangle \psi_\gamma = \sum_\gamma \langle f, \psi_\gamma \rangle \, \tilde{\psi}_\gamma.$$

Recall, that the frame property guarantees 'stability' of the representation, in the sense that there exist constants $0 < A \leq B < \infty$, such that

$$A \left\| f \right\|_2^2 \leq \sum_\gamma \left| \langle f, \psi_\gamma \rangle \right|^2 \leq B \left\| f \right\|_2^2, \quad \forall f \in L_2\left([-1,1]^2\right).$$

Let us rearrange the Ridgelet coefficients based on the size of their absolute values

$$\left| \langle f, \psi_{\gamma_1} \rangle \right| \geq \left| \langle f, \psi_{\gamma_2} \rangle \right| \geq \cdots,$$

and denote the n-term adaptive approximation to f by

$$f_n := \sum_{i=1}^{n} \langle f, \psi_{\gamma_i} \rangle \, \tilde{\psi}_{\gamma_i}.$$

Then, we have the Jackson-type estimate [4] for $\alpha > 1/2$ and $1/\tau = \alpha - 1/2$,

$$\left\| f - f_n \right\|_{L_2\left([-1,1]^2\right)} \leq c n^{-\alpha/2} \left\| f \right\|_{\dot{R}^\alpha_{\tau,\tau}}.$$

Thus, under certain assumptions on the input function, not only is the convergence of the adaptive approximation is ensured, but its rate is also estimated. The outcome of the theory is that the approximation rate of an adaptive Ridgelet approximation depends on the smoothness of the function in a given Ridgelet smoothness space, much in the same manner that adaptive wavelet approximation is characterized by Besov space smoothness [12].

As we shall see in Section 3, our adaptive acquisition method relies on adaptive Ridgelet approximation to predict, at each iteration, the next significant acquisition set.

3 Adaptive Tomography Acquisition

Before presenting the details of the algorithm, we first provide an instructive and useful example: Assume we had access to an optimal 'oracle'. We then ask, how many line projections are needed as rows in the matrix A, such that the 'Square' image of Fig. 2 can be reconstructed with high precision, using the TV functional (1)?

In fact, equipped with an 'oracle', this image can be reconstructed with extremely high quality, where the matrix A in (1) contains only 8 rows associated with 8 line projections. Thus, the numbers of samples, satisfies $n = 0.000122N$,

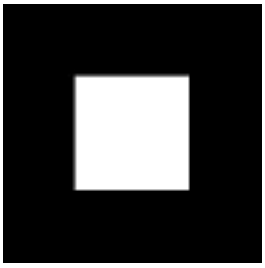


Fig. 2. : 'Square' image of size 256×256

which is a tiny fraction of the size of the image $N = 256 \times 256$. This is achieved by selecting the unique four pairs of line projections that are the immediate neighbors of each of the four lines associated with the edges of the white square. Fig. 3 shows the locations of the line projections and the reconstructed image.

The moral of this example, which correlates well with the theory reviewed in Section 2.2, is that during the acquisition process, we should try to adaptively sample the line projections that are aligned and centered on the edges of the image. Obviously, the image to be acquired is unknown and we do not have access to an 'oracle'. As we shall see in the next subsection, this is exactly where the multiresolution nature of the Ridgelet model is useful.

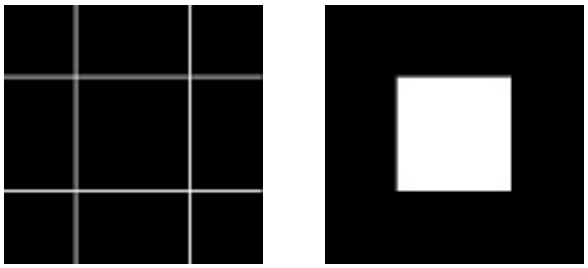


Fig. 3. 'Square' image: On the left, the acquired 8 line projections (using an 'oracle') and on the right, a reconstruction computed from the 8 projections using TV minimization, PSNR=61.85dB

We now present in detail the steps of the algorithm. After initialization, at the kth iteration, we have an adaptive sampling matrix $A^{(k)}$ whose last rows are the new samples obtained at the previous step. We use $A^{(k)}$ to solve a TV minimization problem and obtain the iterative approximation $U^{(k)}$.

3.1 Initialization

First, we create an initial sampling matrix $A^{(0)}$ by using a relatively very small uniform set of line projections. The number of line projections is relative to the image size. For example, in our experimental results, for images of size 256×256, we measured 8 uniformly spaced line integrals at eight uniformly spaced angles,

which gives a total of 64 initial measurements that are about 0.1% of the image size. In Fig. 5(a), we see an illustration of this non-adaptive sampling pattern for images of size 256×256. We also initialize the current approximation to the unknown image I with $U^{(-1)} = -\infty$ (e.g. each entry equal to -10^8).

3.2 TV Minimization Step

At the beginning of the kth iteration of our algorithm, we are equipped with the sampling matrix $A^{(k)}$ whose last rows contain a subset of newly acquired samples and the previous approximation $U^{(k-1)}$ as the initial guess. Therefore, we use this updated matrix and proceed with a TV minimization step (1) to compute $U^{(k)}$. Recall that in our setup, the sparse nature of $A^{(k)}$ allows us to achieve this computation on large images by using a sparse representation of the matrix $A^{(k)}$. Our modified TVAL algorithm (see Section 2.1) stores $A^{(k)}$ as a MATLAB sparse matrix of Boolean values, which reduces significantly the memory access overhead.

Here, we have an option to select a tradeoff between reconstruction quality and performance. We do not necessarily need to completely solve the TV minimization problem by iterating an algorithm such as in Section 2.1 until it converges. Instead, we may apply only a fixed and limited number of iterations of the TV solver, or terminate the iterations using a less demanding stop criterion and then proceed to the next step of the Ridgelet analysis. This will speed up the algorithm, but in some cases, its effect on the next analysis step will imply that we will need to acquire more line projections for the same reconstruction image quality. In any case, our adaptive acquisition process terminates at an iteration of this step if we obtain $\left\|U^{(k)} - U^{(k-1)}\right\|_2 \leq \varepsilon$, for some prescribed threshold ε.

3.3 Ridgelet Analysis Step

Now that we have, at the kth iteration, an improved approximation $U^{(k)}$ to I, we compute a discrete set of its Ridgelet coefficients. Recall, that a Ridgelet transform can be computed by the application of the Radon transform (7) followed by a wavelet transform, as shown in (8). Since in our application, we only require Ridgelets for analysis, we do not need to use an invertible transform as in [14], which simplifies the implementation. In practice, we found out that if we chose the number of angles to be a quarter of the image length, then our sampling scheme is sufficiently dense for the purpose of our algorithm, but not too dense so as to lead to subsequent unnecessary acquisition, as will become clear. Thus, for an images of size 256×256, the allowable set of projection lines corresponds to only 64 angles $\{0, \pi/64, ..., 63\pi/64\}$, with 256 line projections per direction. For our experimental results, we computed Ridgelet coefficients using the univariate discrete Haar wavelet (5). The discretization of angles is related to the scale parameter of the Ridgelets, so as to avoid subsequent unnecessary acquisition. Specifically, we calculate the Ridgelet coefficients $\alpha^{(k)}_{a,b,\theta} := \left\langle U^{(k)}, \psi_{a,b,\theta} \right\rangle$, using the Haar wavelet function $\psi_{a,b}(x)$, with $a = 2^j,\ j = 0, ..., J_\theta$, where J_θ depends

on the angle θ. The discrete sampling of the Ridgelet coefficients is controlled in the following way:

Angle, $0 \leq l < 8$	J_θ
$8l\pi/64$	3
$(8l+1)\,\pi/64$	0
$(8l+2)\,\pi/64$	1
$(8l+3)\,\pi/64$	0
$(8l+4)\,\pi/64$	2
$(8l+5)\,\pi/64$	0
$(8l+6)\,\pi/64$	1
$(8l+7)\,\pi/64$	0

3.4 Adaptive Sampling of New Line Projections

Based upon the analysis of the Ridgelet coefficients $\{\alpha^{(k)}_{a,b,\theta}\}$, computed at the previous step, we make our decision on which new line projections are added to $A^{(k)}$ as new rows to create the matrix $A^{(k+1)}$. Specifically, we chose these line projections to be associated with the M coefficients with largest absolute values that have not yet been marked as sampled by the algorithm. In our experiments, we select $M = 0.1n$, which is a tenth of the image row size.

The goal of the line projections is to roughly approximate (6) where ψ is the Haar wavelet. In Fig. 4 we see an illustration of a support of a Haar Ridgelet function (outer dotted lines) and the associated two line projections (inner lines) within its support that we compute on the unknown image .

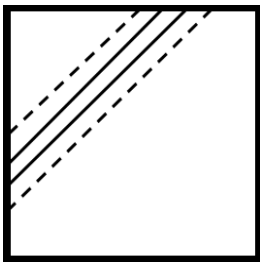


Fig. 4. Line integrals acquired per a significant Ridgelet coefficient: The support of the Ridgelet lies within the area bounded by the external dotted lines. The inner lines are the sampled line projections.

Let us look closer at the implication of using only two line projections to approximate the value of a Haar Ridgelet. Assume that the Ridgelet coefficient $\alpha^{(k)}_{a,b,\theta} = \left\langle U^{(k)}, \psi_{a,b,\theta} \right\rangle$ has not been marked as sampled yet, but is significant enough to be sampled at the current iteration. Let $R_I(\theta, \cdot)$ be the Radon transform of the unknown image I at the fixed angle θ. In such a case, the two values

of the line projections that we acquire are $R_I(\theta, b + a/4)$ and $R_I(\theta, b + 3a/4)$. These values should be considered as the approximation

$$a^{-1/2}\left(R_I(\theta, b + a/4) - R_I(\theta, b + 3a/4)\right) \approx CRT_I(a, b, \theta).$$

We summarize the adaptive acquisition algorithm by 'pseudo-code':

1. *Initialize the approximated image $U^{(-1)}$ with entries corresponding to $-\infty$ (e.g. -10^8) . Obtain initial small number of line projection samples determined by a 'uniform' sampling matrix $A^{(0)}$.*
2. *For $k = 0, 1, \ldots$ Iterate:*
   (a) *Compute $U^{(k)}$ using $A^{(k)}$ by solving (1).*
   (b) *If $\left\|U^{(k)} - U^{(k-1)}\right\|_2 \leq \varepsilon$, go to step 3.*
   (c) *Apply Ridgelet analysis on $U^{(k)}$ to obtain next set of M new candidate projection lines.*
   (d) *Sample the (unknown) image I at the new M projection lines.*
   (e) *Add the new projection lines to the matrix $A^{(k)}$ to create an updated sampling matrix $A^{(k+1)}$.*
3. *Output the most updated approximated image $U^{(k)}$.*

3.5 Analysis and Examples

In Fig. 5, we see a few iterations of the adaptive acquisition algorithm on the Ellipse image. We see in (a) the small number of uniform, non-adaptive line projection measurements that are used for the initialization step. In (b) we see the reconstructed approximation $U^{(0)}$. In (c), we show the new set of line projections that were determined by the Ridgelet analysis on $U^{(0)}$, to be the most significant. The next subfigures show further iterations of newly acquired line projections associated with the next unsampled M largest Ridgelet coefficients and then the approximations $U^{(k)}$ produced by solving the TV functional after adding these new samples as last rows of $A^{(k)}$. Note that the algorithm quickly identifies the edges of the ellipse and only takes line measurements that are aligned with them, where more samples are taken along the longer axis first. Moreover, initially, when the approximation $U^{(k)}$ is still blurry, the algorithm finds through the Ridgelet analysis that it should first acquire line projections associated with low resolution Ridgelet coefficients. Only after the approximation contains sufficiently sharp edges, higher scale Ridgelet coefficients become significant and the line projections associated with them are acquired. In summary, the algorithm attempts to acquire only line projections around and aligned with edge singularities and ordered by scale.

Next, we demonstrate the effectiveness of the estimate for the significant Ridgelet coefficients of the unknown image. The test is conducted on the well-known 256×256 CT Zubal Head test image [38]. To this end, we use the standard Peak Signal to Noise Ratio (PSNR), measured in dB, to quantify an approximation $\tilde{I}$ to the image I where the images pixels take values in $[0, 1]$,

$$PSNR\left(I, \tilde{I}\right) := 10 \log_{10} \frac{1}{\frac{1}{N}\sum_{i,j}\left|I_{i,j} - \widetilde{I}_{i,j}\right|^2}. \tag{9}$$

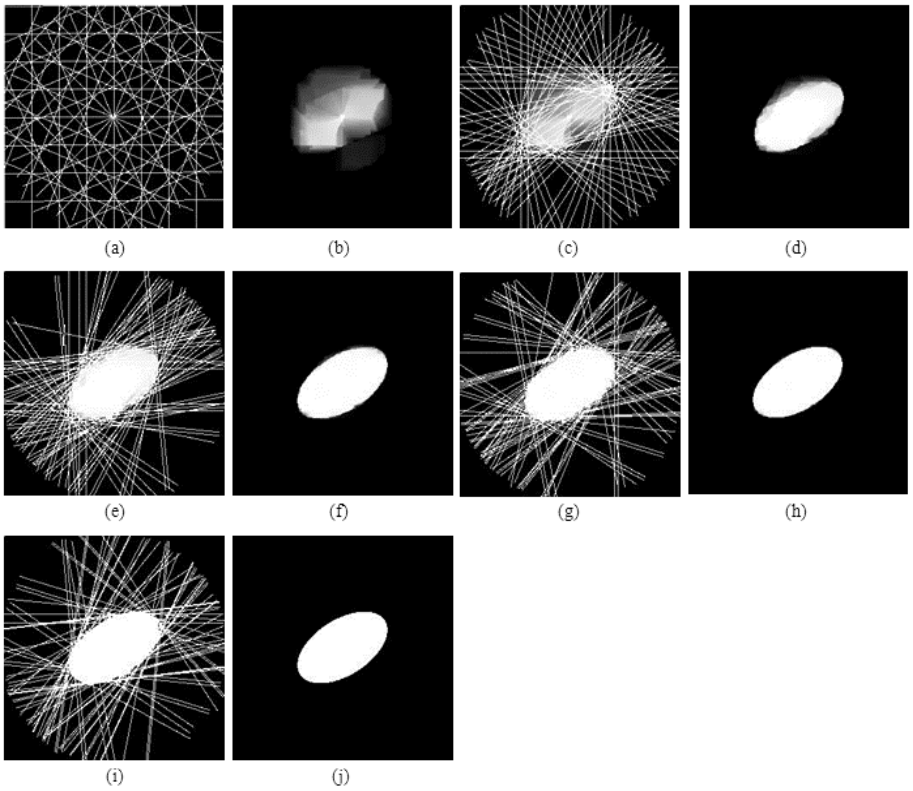


Fig. 5. Adaptive acquisition of the Ellipse image: Iterations of newly added projection lines and approximations $U^{(k)}$

In Fig. 6, we show a graph with number of line projections used by the adaptive algorithm and the PSNR obtained. We compare the performance of our adaptive algorithm with the performance of a benchmark method that uses Ridgelet analysis of the actual real image instead of using the Ridgelet analysis performed on the iterated image. We see that despite of not having the true image available at the time of acquisition, our algorithm manages to perform almost as well as an algorithm equipped with an 'oracle' that uses the Ridgelet analysis of the true image. In general, this property of the algorithm depends on the size M of the set newly sampled projection lines at each iteration. That is, the algorithm manages to trace and collect more accurately the significant line projections associated with the largest Ridgelet coefficients of the real image, if it runs in more iterations, adding each time a small set of new line projections.

We also see (Fig. 7) that the algorithm obtained perfect reconstruction using 3834 line projections. In standard CT acquisition models, 256 line projections are acquired at 256 orientations, a total of 65,536 line projections. In comparison, our algorithm achieves perfect reconstruction using about 6% of that total.

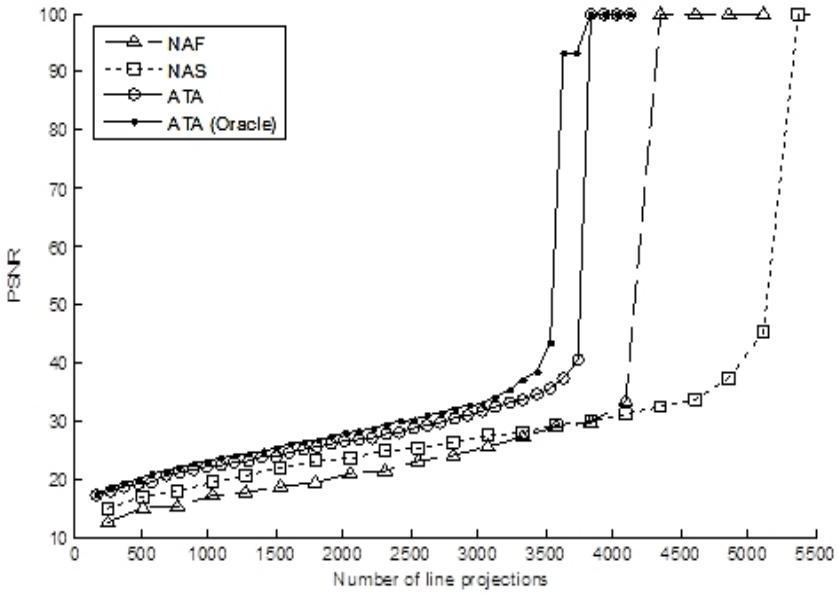


Fig. 6. Results of the adaptive algorithm on the Zubal Head. ATA - performance of the algorithm. ATA (oracle)- performance of the algorithm equipped with an 'oracle' (i.e. Ridgelet analysis of true image). PSNR=100 is actually a graphical cut-off line of perfect reconstruction (PSNR=∞). NAS - Non Adaptive Equally Spaced acquisition. NAF - Non Adaptive Uniform Fourier acqusition (See Section 4)

Under the assumption that the algorithm manages to almost accurately identify the most significant Ridgelet coefficients using only the approximate iterated image, we may analyze the rate of convergence using the theory discussed in Section 2.2. Observe that a function belongs to the Radon 'smoothness' space of Definition 1, with a high value of 'smoothness' index α, if it has 'sparse' directional information, which decreases sufficiently fast as $a = 2^j$ is smaller. As an example, consider 'cartoon' functions which are piecewise constants over polygonal domains. At a fixed angle θ, the Radon transform $R_f(\theta, t)$ is a simple piecewise linear function of the variable t, with compact support and a bounded number of discontinuities, depending of the number of segments in the polygonal boundary. Assume the wavelet ψ has at least two vanishing moments, then for small values of a in (8), the Ridgelet transform $CRT_f(a, b, \theta)$ will be non-zero only in segments of total length $\leq ca$, where c is an absolute constant depending on the choice of ψ and the geometry of the polygonal boundary. Therefore, in the p-norm, for $a = 2^j$, $j < 0$, we get an estimate $\left\|CRT_f\left(2^j, \cdot, \theta\right)\right\|_p^p \leq c(f, \psi, p)\, 2^{j/2}$. This implies that $f \in \dot{R}_{p,q}^{\alpha}\left(\mathbb{R}^2\right)$, for any 'smoothness' $\alpha > 0$. Therefore, from the Jackson estimate we may conclude that the adaptive algorithm will converge for these simple prototype functions at the rate $n^{-\alpha/2}$ for any $\alpha > 0$, which matches the perfect reconstruction results we obtain for these functions

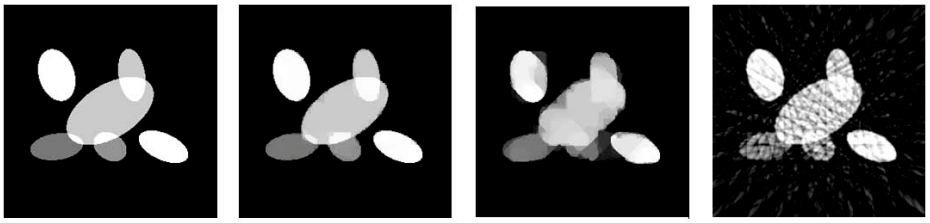
Fig. 7. Perfect reconstruction of the Zubal Head from 3834 adaptive line projections

in this work. For more complex images, the convergence analysis is similar, but more involved. One needs to estimate the Radon smoothness of the function in order to understand the rate of convergence. This type of analysis of adaptive methods has been carried out for wavelet image compression by characterizing images as functions in Besov spaces [13].

4 Experimental Results

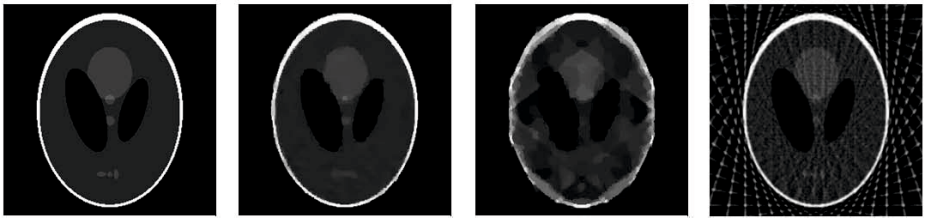
In this section we compare our adaptive approach with standard non-adaptive methods. We show that for a given number of projection lines measured on the unknown image I, our adaptive method provides a significantly better approximation to I. To this end, given an $m \times m$ image, we prescribe a target of n samples. Denote $d = n/m$ (assuming $n \bmod m = 0$). We compare four acquisition and reconstruction methods:

1. Filtered Back Projection (FBP): For the FBP method we sample $60 \times m$ line projections (regardless of the target limit), which are m equally spaced line integrals over the angles $0, \pi/60, ..., 59\pi/60$. We then used the MATLAB implementation ('iradon') to obtain an approximate image.
2. Non Adaptive Equally Spaced (NAS): We use equally spaced rotations and a fixed number of line integrals at each angle such that the total number of line integrals matched the prescribed budget. We then applied TV minimization to this sampled data. Specifically, $m/2$ (equally spaced) line projections are acquired over the angles $0, \pi/2d, 2\pi/2d, ..., (2d-1)\pi/2d$.
3. Non Adaptive Uniform Fourier (NAF): This method is used in [7]. It is mathematically equivalent to NAS, but produces slightly different results in digital implementation. In this mode, we uniformly select lines in the Fourier domain of the image and use Fourier coefficients on these lines as the entries of the sampling matrix A. Specifically, m Fourier coefficients were taken on the lines associated with the angles $0, \pi/d, 2\pi/d, ..., (d-1)\pi/d$.



(a) ATA (b) NAS (c) NAF (d) FBP

Fig. 8. Non-adaptive and adaptive acquisition on the 256 × 256 6-Ellipse image. (a) ATA. 971 line projections. Perfect reconstruction. (b) NAS. 1024 line projections. PSNR=29.73 dB. (c) NAF. 1024 line projections. PSNR=21.93 dB. (d) FBP. 5120 line projections. PSNR=19.64 dB.



(a) ATA (b) NAS (c) NAF (d) FBP

Fig. 9. Non-adaptive and adaptive acquisition on the 256 × 256 Shepp-Logan image. (a) ATA. 1630 line projections. Perfect reconstruction. (b) NAS. 1792 line projections. PSNR=26.44 dB. (c) NAF. 1792 line projections. PSNR=19.53 dB. (d) FBP. 5120 line projections. PSNR=18.04 dB.

4. Adaptive Tomography Acquisition (ATA): Our proposed adaptive method. Line projections were acquired adaptively as described in Section 3.

For the first set of noise-free phantom test images, we used in the iterations the left-hand side constraint in (1), $Au = y$, so that our solutions satisfy the sampling equations exactly. We see below results on well-known tests image. In Fig. 8 we see that for an equivalent number of line projections, our adaptive algorithm achieves prefect reconstruction while the uniform limited angle, non-adaptive acquisition algorithms, equipped with the same TV minimization solver achieve significantly lower image quality. Similar results are shown in Fig. 9 for the 'Shepp-Logan' phantom (see also the graphs in Fig. 10) and for the 'Zubal Head' in Fig. 11. We note that currently the running times of the adaptive acquisition Matlab simulations are about 7-10 times slower than the non-adaptive for the same number of line projections. This relates to the choice of M, the number of new line projections introduced at each iteration. So, for a given number of line projections n, the choice $M = 0.1n$, yields about 10 iterations, where the matrix $A^{(k)}$ contains about $0.1kn$, $k = 1, \ldots, 10$, rows. Solving these iterations

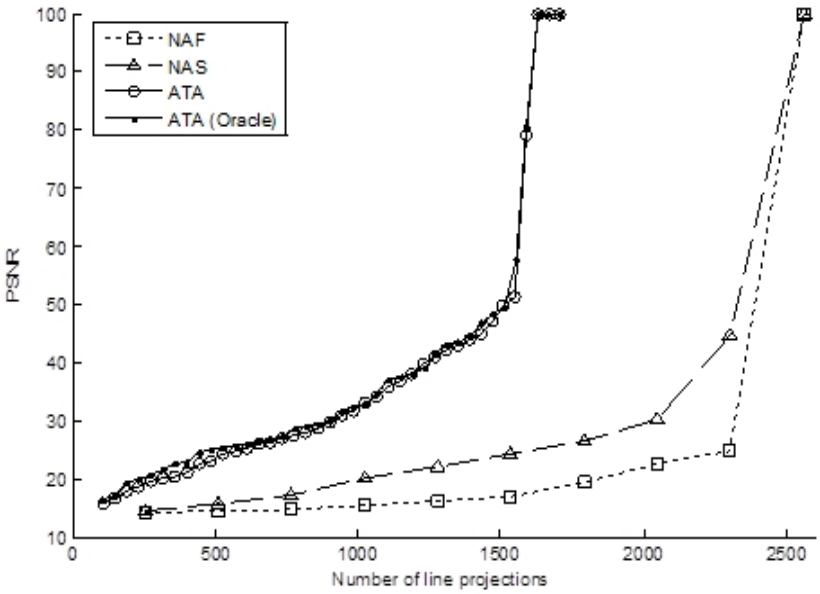
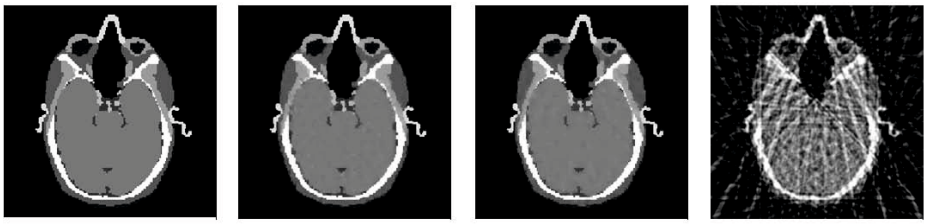


Fig. 10. Comparison of the adaptive acquisition algorithm (ATA) vs. the limited angle, non-adaptive methods of NAS and NAF on the Shepp-Logan image. PSNR=100 is actually a graphical cut-off line of perfect reconstruction (PSNR=∞).

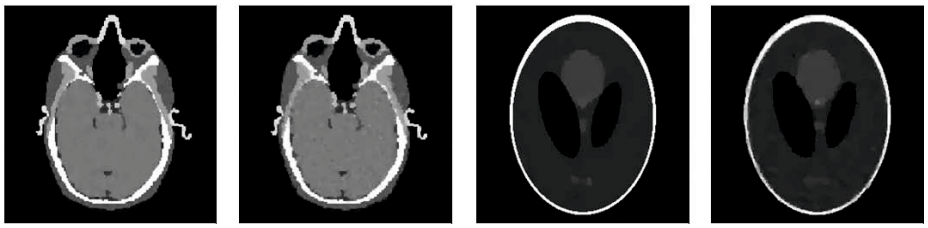


(a) ATA (b) NAS (c) NAF (d) FBP

Fig. 11. Non-adaptive and adaptive acquisition on the 256 × 256 Zubal-Head. (a) ATA. 3834 line projections. Perfect reconstruction. (b) NAS. 4096 line projections. PSNR=31.16 dB. (c) NAF. 4096 line projections. PSNR=33.31 dB. (d) FBP. 5120 line projections. PSNR=17.67 dB.

is about 5.5 slower than solving the TV-minimization of order n only once. The rest of the running time of the adaptive method is spent on the Ridgelet analysis computations that are performed at each iteration.

Next, we show results with simulated low dose as in [10]. For a selected parameter of incident photon count γ_I, the simulated detected photon counts $\tilde{\gamma}$, were chosen as Poisson distributed random variables with mean equal to $\gamma_I e^{-p}$, where p is a noiseless line projection. The simulated noisy projection, $\tilde{p}$, is then



(a) ATA (b) NAS (c) ATA (d) NAS

Fig. 12. Reconstruction from simulated incident photon counts (a) ATA. $\gamma_I = 1,000,000$. 3834 projection lines. PSNR=33.44 dB.(b) NAS. $\gamma_I = 1,000,000$. 4096 projection lines. PSNR=29.26 dB. (c) ATA. $\gamma_I = 250,000$. 1630 projection lines. PSNR=37.58 dB. (d) NAS. $\gamma_I = 250,000$. 1792 projection lines. PSNR=25.30 dB.

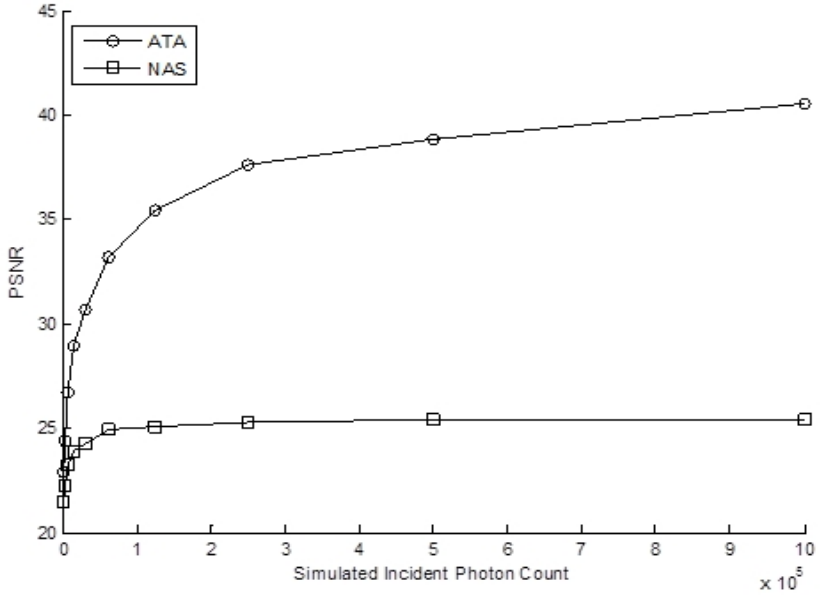


Fig. 13. Graph plot for Shepp-Logan with different simulated incident photon counts

determined by $\tilde{p} = -\log(\tilde{\gamma}/\gamma_I)$. This time, in our iterations, we used the right-hand side constraint in (1), which provides better regularity for noisy data. In Fig. 12 we see a comparison of adaptive and limited angle (non-adaptive) acquisitions using dose simulations. We see that the image quality of our adaptive is clearly higher for the same number of line integrals. In Fig. 13, we see a plot of the reconstructions at various levels of simulated dose levels.

5 Conclusion and Future Work

In this paper we proposed a mathematical model for adaptive Computed Tomography acquisition whose theoretical goal is to radically reduce dosage levels. We presented numerical simulations that demonstrate the potential of the mathematical model of adaptive acquisition and compared our results to the state of the art non adaptive ones.

Our future research will focus on creating more realistic simulations to CT acquisition. We plan to enhance our algorithm to perform well on more realistic images and model more accurately adaptive low-dose radiation, beyond the simplistic model of the total number of line projections. The dose in a CT scan depends on the machine's flux intensity, with lower flux intensity implying lower dose, but higher Poisson-type noise in the detected measurements. We also plan to simulate true 3D scanning and add motion correction.

Lastly, in our work we used a form of the TVAL solver [37], adapted to our problem. It should be very interesting to test other TV solvers such as [25] and see if they (or a modified version of them) are better suited to the adaptive scheme.

References

1. Arias-Castro, E., Candès, E., Davenport, M.: On the fundamental limits of adaptive sensing. IEEE Trans Info. Theory 59, 472–481 (2013)
2. Barnes, E.: MBIR aims to outshine ASIR for sharpness, CT dose reduction. Article in AuntMinnie (May 18, 2010)
3. Brenner, D., Ellison, C.: Estimated risks of radiation-induced fatal cancer from Pediatric CT. American Journal of Roentgenology 176, 289–296 (2001)
4. Candès, E.: Ridgelets: Theory and applications. Ph.D. Thesis, Stanford University (1998)
5. Candès, E., Donoho, D.: New tight frames of Curvelets and optimal representations of objects with piecewise C^2 singularities. Comm. Pure App. Math. 57, 219–266 (2003)
6. Candès, E., Demanet, L., Donoho, D., Ying, L.: Fast Discrete Curvelet Transforms. Multiscale Modeling & Simulation 5, 861–899 (2006)
7. Candès, E., Romberg, J., Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inf. Theory 52, 489–509 (2006)
8. Compressive Sensing Resources home page, `http://dsp.rice.edu/cs`
9. Courant, R.: Variational methods for the solution of problems with equilibrium and vibration. Bull. Amer. Math. Soc. 49, 1–23 (1943)
10. Delaney, A., Bresler, Y.: Globally Convergent Edge-Preserving Regularized Reconstruction: An Application to Limited-Angle Tomography. IEEE Trans. Image Proc. 7, 204–221 (1998)
11. De Man, B., Basu, S.: Distance-driven projection and backprojection in three dimensions. Phys. Med. Biol. 49, 2463–2475 (2004)
12. DeVore, R.: Nonlinear approximation. Acta Numerica 7, 50–150 (1998)

13. DeVore, R., Jawerth, B., Lucier, B.: Image compression through wavelet transform coding. IEEE Trans. Inf. Theory 38, 719–746 (1992)
14. Do, M., Vetterli, M.: The Finite Ridgelet transform for image representation. IEEE Trans. Image Proc. 12, 16–28 (2003)
15. Fortin, M., Glowinski, R.: Augmented Lagrangian Methods: Application to the Numerical Solution of Boundary Value Problems. North-Holland, Amsterdam (1983)
16. Frikel, J.: A new framework for sparse regularization in limited angle x-ray tomography. In: Proc. IEEE Conf. Biomedical Imaging, pp. 824–827 (2010)
17. Frikel, J.: Sparse regularization in limited angle tomography. ACHA 34, 117–141 (2013)
18. Hestenes, M.: Multiplier and gradient methods. Journal of Optimization Theory and Applications 4, 303–320 (1969)
19. Hsieh, J.: Computed Tomography, principles, design, artifacts and recent advances, 2nd edn. SPIE and Wiley-Interscience (2009)
20. Kolditz, D., Kyriakou, Y., Kalender, W.: Volume-of-interest (VOI) imaging in C-arm flat-detector CT for high image quality at reduced dose. Medical Physics 37, 2719–2730 (2010)
21. Kutyniok, G., Labate, D.: Construction of Regular and Irregular Shearlet Frames. J. Wavelet Theory and Appl. 1, 1–10 (2007)
22. Labate, D., Lim, W.-Q., Kutyniok, G., Weiss, G.: Sparse multidimensional representation using shearlets. Wavelets XI, 254–262 (2005); SPIE Proc. 5914
23. Li, C.: An Efficient Algorithm For Total Variation Regularization with Applications to the Single Pixel Camera and Compressive Sensing. Master's thesis, Rice University (September 2009)
24. Mallat, S.: A Wavelet Tour of Signal Processing, 3rd edn. The Sparse Way. Academic Press (2009)
25. Mao, Y., Fahimian, B., Osher, S., Miao, J.: Development and optimization of regularized tomographic reconstruction algorithms. IEEE Trans. Image Proc. 19, 1259–1268 (2010)
26. McCollough, C., Bruesewitz, M., Kofler, J.: CT Dose Reduction and Dose Management Tools: Overview of Available Options. Radiographics 26, 503–512 (2006)
27. Moore, J., Barret, H., Furenlid, L.: Adaptive CT for high-resolution, controlled-dose, region-of-interest imaging. In: IEEE Nucl. Sci. Symp. Conf., pp. 4154–4157 (2009)
28. Natterer, F.: The mathematics of computerized tomography. Classics in applied mathematics, vol. 32. SIAM (1986)
29. Nelson, R.: Thousands of new cancers predicted due to increased use of CT. Medscape news (December 17, 2009)
30. Powell, M.: A method for nonlinear constraints in minimization problems. In: Fletcher, R. (ed.) Optimization, pp. 283–298. Academic Press, London (1969)
31. Quinto, T.: Singularities of the X-ray transform and limited data tomography in $\mathbb{R}^2$ and $\mathbb{R}^3$. SIAM J. Math. Anal. 24, 1215–1225 (1993)
32. Ramani, S., Fessler, J.: Parallel MR image reconstruction using augmented Lagrangian methods. IEEE Trans. Med. Imag. 30, 694–706 (2011)
33. Rutt, B.: Scanned projection radiography using high speed computed tomographic scanning system. U.S Patent 4,573,179
34. Shuryak, I., Sachs, R., Brenner, D.: Cancer Risks after Radiation Exposure in Middle Age. J. National Cancer Institute 102, 1606–1609 (2010)

35. Thibault, J., Sauer, K., Bouman, C., Hsieh, J.: A three-dimensional statistical approach to improved image quality for multislice helical CT. Med. Phys. 34, 4526–4544 (2007)
36. Thibault, J., Sauer, K., Bouman, C., Hsieh, J.: Fast Model-Based X-Ray CT Reconstruction Using Spatially Nonhomogeneous ICD Optimization. IEEE Trans. Image Proc. 20, 161–175 (2011)
37. Li, C., Yin, W., Zhang, Y.: TVAL3 homepage, `http://www.caam.rice.edu/~optimization/L1/TVAL3/`
38. Zubal, I., Harrell, C., Smith, E., Rattner, Z., Gindi, G., Hoffer, P.: Computerized three-dimensional segmented human anatomy. Med. Phys. 21, 299–302 (1994)