Fast Variance Prediction for Iteratively Reconstructed CT with Applications to Tube Current Modulation

by

Stephen Schmitt

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2015

Doctoral Committee:

Professor Jeffrey A. Fessler, Chair
Professor Anna C. Gilbert
Professor Mitchell M. Goodsitt
Associate Professor Clayton D. Scott

ACKNOWLEDGEMENTS

This work could not have been done without the help and support of many people throughout the last N years. First and foremost, I would like to thank my advisor, Jeff Fessler. His oversight on this research has been invaluable. On the many occasions when I became stuck, either needing a way to cut through a difficult problem or a new direction to take our research, Jeff had the answer. His constant feedback has also done wonderful things for my writing, which is immediately apparent looking at papers from years ago.

I would like to thank my committee, Anna Gilbert, Clay Scott, and Mitch Goodsitt, for their helpful comments and questions. I especially thank Mitch for his help collecting the data from a real CT scanner used in this work. Without this data, everything here would just be theoretical.

My friends here in Ann Arbor have all been wonderful and have made my time in Michigan a great one that I will always remember fondly. Ann Arbor has been the warmest place I've lived, both socially and meteorologically. Thanks, too, to friends in other cities who, while I couldn't spend a lot of time with them in person, have given me support from afar. Special thanks to Erin, my best friend and greatest supporter, for everything, but in particular the patience to read many parts of this work and listen to me talk about it, many times over.

Finally, I would like to thank my parents for getting me almost any book I could have wanted while I was growing up. The University of Michigan library system has given me a lot of reading material and taught me a lot (see Bibliography, page 113), but it is still a distant second to the people who gave me a lifetime appreciation for learning.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Fast Variance Prediction for Iteratively Reconstructed CT with Applications to Tube Current Modulation

by

Stephen Schmitt

Chair: Jeffrey Fessler

X-ray computed tomography (CT) is an important, widely-used medical imaging modality. A primary concern with the increasing use of CT is the ionizing radiation dose incurred by the patient. Statistical reconstruction methods are able to improve noise and resolution in CT images compared to traditional filter backprojection (FBP) based reconstruction methods, which allows for a reduced radiation dose. Compared to FBP-based methods, statistical reconstruction requires greater computational time and the statistical properties of resulting images are more difficult to analyze. Statistical reconstruction has parameters that must be correctly chosen to produce high-quality images. The variance of the reconstructed image has been used to choose these parameters, but this has previously been very time-consuming to compute. In this work, we describe a method for fast prediction of the variance that can be used to choose these parameters.

We derive an approximation to the local frequency response (LFR) of the operation combining CT projection, statistical weighting, and backprojection. This approximation is separable into one term that is not weighting dependent and a term that is weighting dependent but has one fewer dimension. Prior work has provided this LFR approximation for a limited set of specific geometries; here we extend this to arbitrary CT geometries.

We apply this LFR approximation to predict the variance of statistically reconstructed CT images. Compared to the empirical variance derived from multiple simulated reconstruction realizations, our method is as accurate as the currently available methods of variance prediction while being computable for thousands of voxels per

second, faster than these previous methods by a factor of over ten thousand. We also compare our method to empirical variance maps produced from an ensemble of reconstructions from real sinogram data. The LFR can also be used to predict the power spectrum of the noise and the local frequency response of the reconstruction.

Tube current modulation (TCM), the redistribution of X-ray dose in CT between different views of a patient, has been demonstrated to reduce dose when the modulation is well-designed. TCM methods currently in use were designed assuming FBP-based image reconstruction. We use our LFR approximation to derive fast methods for predicting the SNR of linear observers of a statistically reconstructed CT image. Using these fast observability and variance prediction methods, we derive TCM methods specific to statistical reconstruction that, in theory, potentially reduce radiation dose by 20% compared to FBP-specific TCM methods.

CHAPTER I

Introduction

1.1 Problem statement

The methods for X-ray computed tomography (CT) image reconstruction most often used in practice are filter back-projection (FBP) methods. These methods are able to reconstruct an image exactly, in theory, but in practice ignore multiple considerations that give rise to image artifacts.

By alternately modeling these non-ideal effects, such as discretization and observation noise and posing image reconstruction as a maximum likelihood problem, the resulting reconstructed images are more accurate, avoiding several classes of artifacts, such as:

- Cases where the imaged object contains X-ray opaque objects, causing metal artifacts [67],
- Limited view artifacts, caused by missing views due to e.g., a detector fault [1],
- Artifacts caused by beam hardening [10].

This framework, reconstructing a CT image by treating it as a maximum likelihood problem, is known as *statistical* or *iterative image reconstruction*. It has been used not just for CT, but also for MRI [58, 3] and terahertz [39, 40, 53] imaging. Statistical image reconstruction can also reduce noise and improve resolution compared to FBP-based methods [61, 78]. Using these statistical methods instead of FBP can thus reduce the X-ray dose required to obtain an equally accurate image. As the prevalence of CT imaging increases, concerns about radiation dose absorbed by patients are becoming more and more significant—it has been estimated that as many as 2% of new cancers in the United States are caused by CT radiation dose [4], and

several agencies, including governmental agencies, recommend following the "ALARA principle", standing for "as low as reasonably achievable", regarding X-ray radiation [37, 63, 64].

The downside of statistical image reconstruction is the much higher computational cost compared to the relatively simple FBP algorithm. Statistical reconstruction requires iterative methods to maximize the corresponding likelihood, which require computing a forward- and back-projection of the data for each iteration. Additionally, statistical reconstruction adds a regularization term that provides more control over the reconstructed image. This regularizer, though, gives us more parameters that we must design or select. Simple choices of regularizers can cause non-uniform or non-isotropic image resolution and fail to preserve edges in the reconstructed image, which are undesirable effects [55]. Therefore more complicated regularizers are often used, and having an estimate of the variance of the reconstruction of each voxel helps in formulating these regularizers in a way that can preserve edges while also eliminating noise [76, 77, 8].

Tube current modulation is a method for reducing X-ray dose in CT by adaptively changing the intensity of incident X-rays for different views. It has been verified to reduce dose both theoretically and empirically for images reconstructed using FBP-like algorithms [59, 29, 24] as well as for statistically reconstructed images [65, 38], and is used widely in practice. The tube current modulation strategies currently used are agnostic to the image reconstruction method used, and so modulate tube current in a way that is ideal for FBP-like reconstruction, but not *necessarily* ideal for statistical reconstruction. Since statistical reconstruction is able to produce higher-quality images than traditional methods with data that is more constrained in photon count or view angle, statistically reconstructed images allow a more aggressive dose modulation and reduction strategy than the tube current modulation strategies used for FBP images.

Figure 1.1 compares coronal slices of a clinical CT image using three different reconstruction methods. The leftmost image was produced using traditional FBP-based reconstruction methods. The center image was produced using ASiR, a proprietary GE Healthcare method that is statistical but not model-based [56]. The rightmost image was produced using the fully model-based statistical reconstruction methods that we consider in this work. The images are arranged in order of increasing visual quality, but also in order of increasing computation time.

Figure 1.1: Coronal slices of a CT image using FBP-based (left), ASiR (center), and statistical (right) reconstruction methods acquired using the same radiation dose. (Figure courtesy of GE)

1.2 Contributions

Predicting the variance map for iteratively reconstructed CT has been researched before, but previous methods are often computationally intractable for large images. Methods exist for more quickly generating a variance map for some CT geometries, including 2D fan-beam CT [76] and 3D step-and-shoot CT with very small steps [77].

In Chapter III, we derive a new approximation for the frequency response of the combination of projection, weighting, and backprojection, which is the Hessian of the data-fit term in the cost function for iterative CT reconstruction. This approximation is the basis for our novel method for fast prediction of the variance map (Chapter IV), which applies to arbitrary CT geometries, with further approximations for CT geometries with a small cone angle [50, 51, 52].

To our knowledge, the problem of adapting tube current modulation for statistically reconstructed images was previously an unexplored area. In Chapter V, we develop fast methods for planning tube current modulation strategies that are particular to CT scans that will be statistically reconstructed. As part of this, we also develop fast frequency domain approximations for the performance of several resolution and feature detectability metrics, including the contrast recovery coefficient (CRC) and SNR for linear image observers.

1.3 Organization

In Chapter II we present a simplified model of X-ray CT systems that we use in our methods, concluding with a linear system model and model for the noise statistics

of observed data. We also describe the existing closed-form methods for CT image reconstruction using this model. In Chapter III we begin our methods by deriving an approximation for the frequency response of the Hessian of the data-fit term. In Chapter IV we apply the results of Chapter III to produce an approximation of the variance of one reconstructed voxel and refine this approximation for some specific CT geometries in common use. We also compare the results of these methods to empirical variance maps produced by an ensemble of simulated reconstructions. Chapter V applies similar frequency domain methods, using the frequency response derived in Chapter III, to investigate automatic tube current modulation schemes for statistically reconstructed images. Finally, Chapter VI summarizes our results and describes potential ideas for future work.

CHAPTER II

Background

2.1 X-ray CT systems

A modern X-ray CT system consists of an X-ray source combined with an X-ray detector mounted on a gantry that rotates around a patient placed between the source and the detector. The X-ray source acts as a photon source, and emits photons of energies up to the voltage applied to the X-ray tube (for example, if 100kV is applied to the tube, photons with energy up to 100keV are emitted). The source is approximately a point source; in our work, we will model it as a point source, but the source not being a point is a source of blur that we do not account for. The photons are then probabalistically absorbed by the material in the scanner, between the source and detector, and the photons that pass through the scanned object are measured by the detector. By measuring these unabsorbed photons at many positions along the rotation of the source and detector, a 2D or 3D image of the X-ray attenuation of the object can be made.

Figure 2.1 shows a top view of a sample geometry considered in this paper with the notation used for angles and distances. We consider the origin to be the center point that the source and detector rotate around. The X-ray source is always the distance D_{so} from the origin, and for every ray, the distance from the source to the detector is D_{sd}. Note that this requires the detector to be an arc of the circle of radius D_{sd} centered around the source. This 2D projection shows the geometry for 2D fan-beam CT. In 3D multidetector row CT, the geometry is the same, except that the detector is extruded (but not curved) up and down from this plane in the z direction. With such 2D detectors, we can reconstruct a 3D image of the object. Here, σ is the angular position of the X-ray source, where the axes are chosen such that $\sigma = 0$ is at the negative x-axis and in increasing σ, the X-ray source travels

toward the negative y-axis.[1]

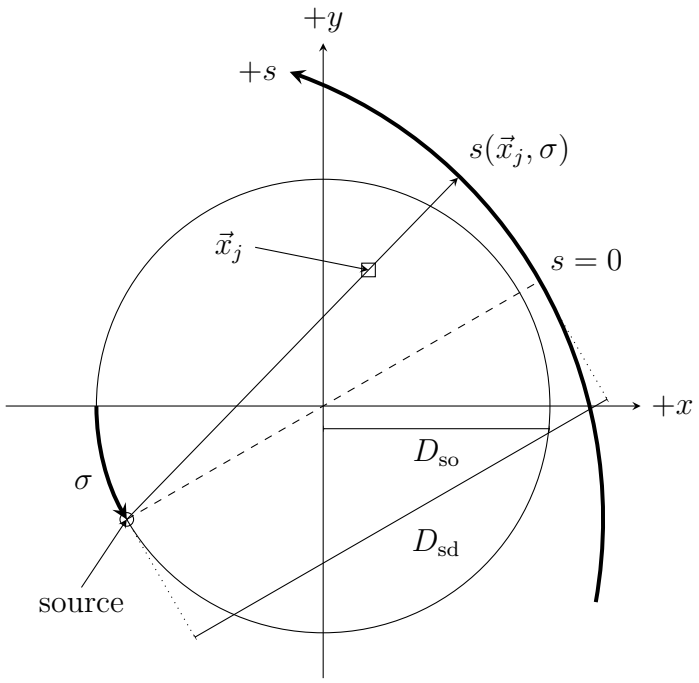


Figure 2.1: 2D fan-beam CT geometry.

Figure 2.2 (adapted from [18]) shows a view of 3DCT geometry where the detector is flat. In third-generation CT geometries, the detector is curved along the direction labeled s.

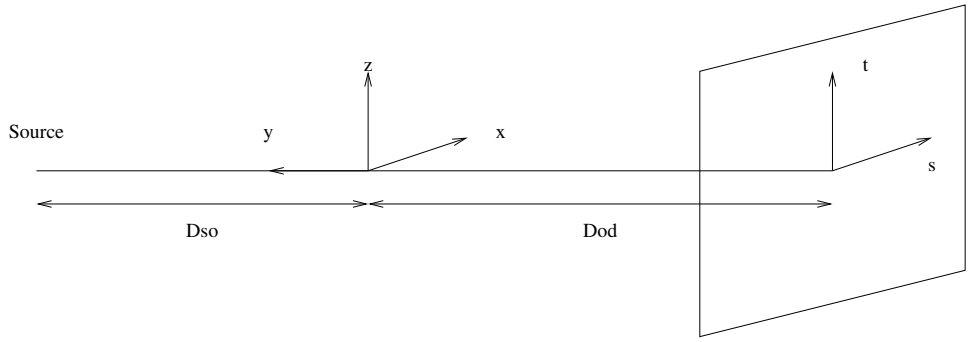


Figure 2.2: 3D cone-beam CT geometry with a flat detector.

Other 3D geometries should be described either because they have been the focus of prior work or are considered specifically in this work: In 3D step-and-shoot CT, the source and detector make a rotation around the object in the xy plane, move in

[1]This particular choice of location for $\sigma = 0$ is arbitrary, although conventionally the coordinate system is chosen to that the X-ray source begins at the $+y$ axis and moves toward the $-x$ axis.

the z direction, and repeat this process with multiple axial-CT-like shots translated in the z direction.[2] The resulting data can be used to reconstruct a 3D volume with a longer z-axis than in just axial CT. In 3D helical CT, the source and detector consistently rotate while also translating in the z-axis. Like step-and-shoot CT, this can reconstruct a larger volume. 3D axial CT is a special case of both step-and-shoot CT (with one step) and helical CT (with no z-axis source translation), in which the source rotates around the object with no translation.

2.2 Background on filter backprojection

Define the *ray transform* $\mathcal{R}$ of a function $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ as integrals over all lines in $\mathbb{R}^n$, which are represented as the pair $\vec{\theta} \in \mathbb{S}^{n-1}$, a direction on the unit sphere in $\mathbb{R}^n$, and $\vec{r} \in \vec{\theta}^{\perp}$, a point on the subspace perpendicular to $\vec{\theta}$:

$$(\mathcal{R}f)(\vec{r}, \vec{\theta}) \triangleq \int f(\vec{r} + t\vec{\theta})\, dt. \tag{2.1}$$

The domain of $\mathcal{R}f$ is denoted $\mathbb{T}^n$, where $\mathbb{T}^n = \{(\vec{r}, \vec{\theta}) : \vec{\theta} \in \mathbb{S}^{n-1}, \vec{r} \in \vec{\theta}^{\perp}\}$. We will focus primarily on the case of 3D CT, where $n = 3$. When $n = 2$, the ray transform is identical to the Radon transform, which is defined as the integral of f through all $(n-1)$-dimensional hyperplanes.

The Fourier transform of a ray transform g, $\mathcal{F}g$, is defined across $\vec{r}$:

$$(\mathcal{F}g)(\vec{u}, \vec{\theta}) \triangleq \int_{\vec{\theta}^{\perp}} \exp(-j2\pi \vec{r}^{\mathrm{T}} \vec{u}) g(\vec{r}, \vec{\theta})\, d\vec{r}, \quad \vec{u} \in \vec{\theta}^{\perp} \tag{2.2}$$

The adjoint $\mathcal{R}^*$ of the ray transform, or the *backprojection* operator is given by:

$$(\mathcal{R}^* g)(\vec{x}) = \int_{\mathbb{S}^{n-1}} g(\mathbf{P}_{\vec{\theta}} \vec{x}, \vec{\theta})\, d\vec{\theta}, \tag{2.3}$$

where $\mathbf{P}_{\vec{\theta}} = \mathbf{I}_n - \vec{\theta}\vec{\theta}^{\mathrm{T}}$ is an $n \times n$ matrix that projects $\vec{x}$ onto $\vec{\theta}^{\perp}$.

From [41] (pg. 18), we have the following theorem:

$$(\mathcal{F}\mathcal{R}f)(\vec{u}, \vec{\theta}) = (\mathcal{F}f)(\vec{u}), \tag{2.4}$$

[2] In the CT geometries described here that involve a z-direction translation, in practice often the patient is moved and not the source and detector gantry. We will consider a coordinate system, however, in which the origin is fixed relative to the patient. In these coordinates, regardless of whether the patient or the gantry is moving in physical space, the source and detector are translating.

that is, the Fourier transform of the ray transform of f is proportional to the Fourier transform of f. Note that the Fourier operators $\mathcal{F}$ are different on each side of (2.4); the former is in the space of $\vec{\theta}^{\perp}$, as in (2.2) and the latter in $\mathbb{R}^n$.

The Riesz potential $\mathcal{I}^{\alpha}$ is defined in $\mathbb{R}^n$ as a multiplication in frequency space:

$$(\mathcal{F}\mathcal{I}^{\alpha} f)(\vec{\nu}) \triangleq ||\vec{\nu}||^{-\alpha} (\mathcal{F} f)(\vec{\nu}), \tag{2.5}$$

and in projection space $\mathbb{T}^n$ as

$$(\mathcal{F}\mathcal{I}^{\alpha} g)(\vec{\nu}, \vec{\theta}) \triangleq ||\vec{\nu}||^{-\alpha} (\mathcal{F} g)(\vec{\nu}, \vec{\theta}). \tag{2.6}$$

With these definitions, we have the following theorem [41] (pg. 18): If $g = \mathcal{R} f$,

$$f = \frac{1}{|\mathbb{S}^{n-1}|} \mathcal{I}^{-\alpha} \mathcal{R}^{*} \mathcal{I}^{\alpha - 1} g. \tag{2.7}$$

Note that the first Riesz potential is in the domain of $\mathbb{R}^n$ and the second in $\mathbb{T}^n$.

In two dimensions, choosing $\alpha = 0$ or $\alpha = 1$ gives two different methods for exact reconstruction of an image f from its projections g:

- If $\alpha = 0$, (2.7) specifies the filtered backprojection (FBP) algorithm for 2DCT reconstruction, which first filters each projection $g(\cdot, \vec{\theta})$ with a *ramp filter* with frequency response $||\vec{\omega}||$, and then backprojects (multiplies with $\mathcal{R}^{*}$).

- If $\alpha = 1$, (2.7) describes the backproject-filter (BPF) method, which first backprojects and then filters in the spatial domain with a filter that has a frequency response of $||\omega||$, a *cone filter*.

In practice, implementation concerns with the latter algorithm (BPF) are larger than those for the former (FBP), and so FBP is used more often in practice. The cone filter used in BPF amplifies high-frequency noise; this can be fixed by multiplying the filter in the frequency domain with a window that rolls off higher frequencies. In the noise-free case, the resulting image would be the original image convolved with the impulse response of the windowing function. The backprojection $\mathcal{R}^{*} g$ has an infinite support, so for practical computation of BPF, $\mathcal{R}^{*} g$ must be stored for a very large area, increasing the computational cost.

The 1D ramp filter used in FBP also amplifies high-frequency noise, and so this is also apodized like the 2D cone filter. This filter also has a wide support in the spatial domain but because the signal being filtered has a finite support in the case of the

1D projection-domain filter, it is more computationally tractable than the 2D cone filtering of the infinitely supported backprojection.

The problem of 3D reconstruction is much more difficult than in 2D, first because not all of g is available, and secondly because the Riesz potentials are complicated. There exists a formula in [45] for exact recovery in 3D of f from a subset of $\mathcal{R}f$, but this subset itself as also not always available. The Feldkamp cone-beam algorithm (FDK) [16] is widely used for approximate 3DCT image reconstruction when the X-ray source has a circular path.

2.3 3D CT geometry

The $\mathbb{T}^n$ coordinates used in the previous section are useful for theoretical analysis of CT reconstruction, but for our purposes we prefer coordinates that are not specified in terms of a direction and offset, but rather of a source position σ and detector position $\vec{s}$ that index the subset of $\mathbb{T}^n$ that is observable by an actual CT system. It is assumed for these coordinates that a source trajectory is already defined as a function $\vec{p}_s : \Sigma \to \mathbb{R}^n$, that depends on a specific CT geometry, that maps a position σ along the trajectory $\Sigma = [\sigma_{\min}, \sigma_{\max}]$ to the spatial coordinates of the X-ray source at that position. We make an approximation that the source is a point; in practice this is not true, but it is close to true and greatly simplifies notation. With a source position σ, a detector position $\vec{s} \in \mathbb{R}^{n-1}$ defines a spatial detector coordinate in $\vec{p}_d(\vec{s}, \sigma) \in \mathbb{R}^n$ on which a particular X-ray lands. From $\vec{p}_s$ and $\vec{p}_d$, the coordinates $(\vec{r}, \vec{\theta}) \in \mathbb{T}^n$ corresponding to $(\vec{s}, \sigma)$ can be found. We denote the space of $(\vec{s}, \sigma) \in \mathbb{R}^{n-1} \times \Sigma$ as $\mathbb{D}$.

We define the *projection* of a function f to be:

$$(\mathcal{P}f)(\vec{s}, \sigma) \triangleq (\mathcal{R}f)(\vec{r}(\vec{s}, \sigma), \vec{\theta}(\vec{s}, \sigma)).$$

Since the functions that map $(\vec{s}, \sigma)$ to $\vec{r}$ and $\vec{\theta}$ are dependent on geometry, the projection operator $\mathcal{P}$ now depends on the geometry as well. Whereas a change to the source trajectory or detector shape does not change $\mathcal{R}f$, it will change $\mathcal{P}f$. The projection $\mathcal{P}f$ does not carry any additional information about f than the ray transform; it is simply a coordinate transformation of a subset of $\mathcal{R}f$. Under certain conditions it does contain enough information to analytically invert $\mathcal{P}$. However, within the statistical reconstruction framework we use, we are only concerned with using the forward model in reconstruction. We will not attempt to find or use $\mathcal{P}^*$ the way that (2.7) requires $\mathcal{R}^*$.

2.3.1 System discretization

We denote the X-ray attenuation as a function of spatial position $\vec{x}$ and photon energy $\mathcal{E}$ as $\mu(\vec{x}; \mathcal{E})$. Many system models, though, assume mono-energetic attenuation for the purposes of simplification, and so we will drop the $\mathcal{E}$ and represent the mono-energetic attenuation as simply $\mu(\vec{x})$. We approximate μ as a linear combination of N_{vox} spatial functions $R_i(\vec{x})$, $j \in \{1, 2, \ldots, N_{\text{vox}}\}$:

$$\mu(\vec{x}) = \sum_{j=1}^{N_{\text{vox}}} x_j R_j(\vec{x}). \tag{2.8}$$

The factors x_j are collected as a finite-dimensional vector $\mathbf{x} \in \mathbb{R}^{N_{\text{vox}}}$ that describes the continuous attenuation function $\mu(\vec{x})$. In this work, we will primarily consider regions R_j such that

$$R_j(\vec{x}) = R_{\text{basis}}(\mathbf{V}^{-1}(\vec{x} - \vec{x}_j)), \tag{2.9}$$

i.e., the regions are the same shape, given by a basis region R_{basis}, which is then scaled by the voxel-spacing[3] matrix $\mathbf{V}$ and positioned at $\vec{x}_j$. Furthermore, we assume that

$$\vec{x}_j = \mathbf{V}\vec{n}_j + \vec{x}_{\text{offset}}, \tag{2.10}$$

where $\vec{n}_j \in \mathbb{Z}^n$ is the position of region j specified on an integer grid, so that the positions of the regions are fixed to a lattice specified by the lattice spacing $\mathbf{V}$, and the common offset $\vec{x}_{\text{offset}}$. The most common choice for the voxel-spacing matrix in 2DCT is square pixels on a square lattice, given by $\mathbf{V} = \Delta\mathbf{I}_2$. In 3DCT geometries such as axial CT or helical CT where the source rotates around the z axis, the most common choice is a rectangular grid with voxels that are the same size in the x and y directions, but optionally differently sized in the z direction. In this case, $\mathbf{V} = \text{diag}(\Delta_x, \Delta_x, \Delta_z)$.

2.3.2 Physical X-ray model

For *monoenergetic* CT, we represent the incident X-ray photon intensity as a function $I(\vec{s}, \sigma)$, such that with no attenuating object present, the number of photons that land in a detector region $\mathcal{D}$ due to the source radiation is a Poisson random

[3]This name and notation suggests 3-dimensional CT, but this is in fact general to n-dimensional CT where $\mathbf{V} \in \mathbb{R}^{n \times n}$.

variable with mean:

$$\bar{I}(\mathcal{D}) = \int_{\mathcal{D}} I(\vec{s}, \sigma)\, d\vec{s}\, d\sigma.$$

According to the Beer-Lambert law[2], an object with attenuation $\mu(\vec{x})$ as a function of space $\vec{x}$ attenuates the beam in the direction represented by $(\vec{s}, \sigma)$ by a factor of $\exp(-(\mathcal{P}\mu)(\vec{s}, \sigma))$, such that with the object μ present, the number of photons landing in the detector region $\mathcal{D}$ is Poisson with a mean:

$$\bar{I}(\mathcal{D}; \mu) = \int_{\mathcal{D}} I(\vec{s}, \sigma) \exp(-(\mathcal{P}\mu)(\vec{s}, \sigma))\, d\vec{s}\, d\sigma. \tag{2.11}$$

An observation i is represented by a detector support function $D_i(\vec{s}, \sigma)$ such that we approximate the number of photons observed in observation i is Poisson with a mean:

$$\bar{I}_i = \int_{\mathbb{D}} I(\vec{s}, \sigma) D_i(\vec{s}, \sigma) \exp(-(\mathcal{P}\mu)(\vec{s}, \sigma))\, d\vec{s}\, d\sigma.$$

Note that if D_i is always equal to 0 or 1 for all $(\vec{s}, \sigma)$, then this is equivalent to (2.11), with $\mathcal{D} = \mathcal{D}_i \triangleq \{(\vec{s}, \sigma) : D_i(\vec{s}, \sigma) = 1\}$. When D_i is between 0 and 1, the representation of $\bar{I}_i$ as a Poisson random variable requires D_i representing the probability that a given photon is detected at that location. The mean ratio of detected photons at observation i to the expected number of photons with no object present is given by

$$\frac{\bar{I}_i}{\bar{I}(\mathcal{D}_i)} = \frac{\int_{\mathbb{D}} I(\vec{s}, \sigma) D_i(\vec{s}, \sigma) \exp(-(\mathcal{P}\mu)(\vec{s}, \sigma))\, d\vec{s}\, d\sigma}{\int_{\mathbb{D}} I(\vec{s}, \sigma) D_i(\vec{s}, \sigma)\, d\vec{s}\, d\sigma}. \tag{2.12}$$

It is commonly assumed that D_i has a small enough support around a point $(\vec{s}_i, \sigma_i)$ such that $(\mathcal{P}\mu)(\vec{s}, \sigma) D_i(\vec{s}, \sigma) \approx (\mathcal{P}\mu)(\vec{s}_i, \sigma_i) D_i(\vec{s}, \sigma)$. That is, that $(\mathcal{P}\mu)$ is slowly varying enough that it is approximately constant within the small support of D_i. This is a common approximation in CT, and the resulting "exponential edge-gradient effect" on reconstruction has been well-studied [28, 66]. With this approximation, (2.12) simply becomes

$$\frac{\bar{I}_i}{\bar{I}(\mathcal{D}_i)} = \exp(-(\mathcal{P}\mu)(\vec{s}_i, \sigma_i)), \tag{2.13}$$

and $(\mathcal{P}\mu)(\vec{s}_i, \sigma_i)$, the projection of μ evaluated at one point on the detector and one

detector position, becomes:

$$\begin{aligned}(\mathcal{P}\mu)(\vec{s}_i,\sigma_i) &= (\mathcal{R}\mu)(\underbrace{\vec{r}(\vec{s}_i,\sigma_i)}_{\triangleq\vec{r}_i},\underbrace{\vec{\theta}(\vec{s}_i,\sigma_i)}_{\triangleq\vec{\theta}_i}) \\ &= \int_{\mathbb{R}} \mu(\vec{r}_i+\tau\vec{\theta}_i)\,d\tau = \epsilon_j + \int_{\mathbb{R}} \sum_j x_j R_j(\vec{r}_i+\tau\vec{\theta}_i)\,d\tau \\ &= \sum_j x_j \int_{\mathbb{R}} R_j(\vec{r}_i+\tau\vec{\theta}_i)\,d\tau \\ &= [\mathbf{A}\mathbf{x}]_i. \end{aligned} \tag{2.14}$$

The matrix $\mathbf{A}$ is a *system matrix*, with elements defined as

$$[\mathbf{A}]_{ij} = \int_{\mathbb{R}} R_j(\vec{r}_i+\tau\vec{\theta}_i)\,d\tau = (\mathcal{P}R_j)(\vec{s}_i,\sigma_i). \tag{2.15}$$

Figure 2.3 shows an example of one of the elements of this system matrix. For the marked source and detector location pair σ_i, $\vec{s}_i$, and the region R_j (represented as a box), the element $[\mathbf{A}]_{ij}$ is the length of the line segment shown in red.

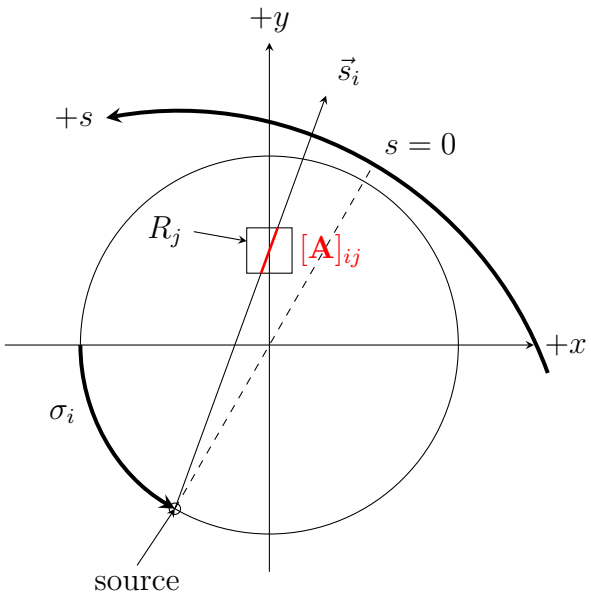


Figure 2.3: Schematic of a system matrix element for 2DCT.

In practice, the exponential edge-gradient effect is often ignored[4] for the sake of the convenience of arriving at the linear expression of (2.14). We define the vector

[4] e.g., in [55, 60] for CT, analogously [57] for PET

$\bar{\mathbf{y}} \triangleq \mathbf{A}\mathbf{x}$ to be the projections of μ, ignoring the exponential edge-gradient effect, evaluated at these observation points. This notation is chosen since $\mathbf{y}$ often represents a measurement of $\mathbf{A}\mathbf{x}$, and $\bar{\mathbf{y}}$ is then the noiseless, and ideally a mean, measurement.

2.4 Model-based image reconstruction

In general, model-based image reconstruction begins with a probabalistic model for observations[5] $\mathbf{Y}$ given data $\mathbf{x}$, such that the pdf of the distribution of $\mathbf{Y}$ is $f_{\mathbf{Y}|\mathbf{x}}(\mathbf{Y}|\mathbf{x})$. We could attempt to find the reconstruction $\hat{\mathbf{x}}$ from observations $\mathbf{Y}$ using a maximum likelihood estimator:

$$\hat{\mathbf{x}}(\mathbf{Y}) = \arg\max_{\mathbf{x}} f_{\mathbf{Y}|\mathbf{x}}(\mathbf{Y}|\mathbf{x}),$$

or equivalently, by minimizing the negative log-likelihood function (or data-fit term) $L(\mathbf{x};\mathbf{Y}) = -\ln f_{\mathbf{Y}|\mathbf{x}}(\mathbf{Y}|\mathbf{x})$:

$$\hat{\mathbf{x}}(\mathbf{Y}) = \arg\min_{\mathbf{x}} L(\mathbf{Y};\mathbf{x}). \tag{2.16}$$

The reconstruction of (2.16) is usually an underdetermined system, so a regularizer $R(\mathbf{x})$ is added to capture prior information about the data $\mathbf{x}$ and better condition the problem. Solving the regularized problem

$$\hat{\mathbf{x}}(\mathbf{Y}) = \arg\min_{\mathbf{x}} L(\mathbf{Y};\mathbf{x}) + \alpha R(\mathbf{x}) \tag{2.17}$$

corresponds to the maximum a posteriori estimate of $\mathbf{x}$ given a prior on $\mathbf{x}$ that is proportional to $f_{\mathbf{x}}(\mathbf{x}) \propto \exp(-\alpha R(\mathbf{x}))$.

2.4.1 Data negative log-likelihood

We will assume that the data-fit term can be represented generally as

$$L(\mathbf{Y};\mathbf{x}) = \sum_{i=1}^{N_d} L_i(Y_i;[\mathbf{A}\mathbf{x}]_i) = \sum_{i=1}^{N_d} L_i(Y_i;\bar{y}_i(\mathbf{x})). \tag{2.18}$$

This formulation implies that the observations $\mathbf{Y}$ are independent given an object $\mathbf{x}$, and that the likelihood is a function of the image only via the projection $\bar{y}_i$. This formulation captures both of the most common statistical models for the observation

[5] We use $\mathbf{Y}$ here to represent generalized observation of any form to distinguish them from observations meant to be a linear estimate of the image vector $\mathbf{x}$.

given the data: the model of observed photons as a Poisson random variable, and the model of the logarithm of observed photons as a Gaussian random variable.

It is important to note that any mismatch between L in (2.18) and a hypothetical "true" likelihood $\breve{L}$ that perfectly matches the true physics of a CT system does not affect our methods. Even if L is a poor approximation of $\breve{L}$, and the resulting reconstruction $\hat{\mathbf{x}}$ is a poor image, our methods can still predict the variance of $\hat{\mathbf{x}}$, so long as L is correctly characterized. In particular, our method accurately predicts the variance even in the usual case where there is model mismatch between the system matrix $\mathbf{A}$ and the actual scanner physics. An accurate characterization of the $\mathbf{A}$ used in a particular reconstruction algorithm is much more important to predicting the variance of that reconstruction than whether $\mathbf{A}$ accurately characterizes the physics. (Of course, the usefulness of the reconstructed image $\hat{\mathbf{x}}$ will depend on the accuracy of $\mathbf{A}$.)

Poisson Likelihood: A common model used follows from (2.13) if we treat the observed number of photons at observation i, I_i, to be a Poisson random variable with a mean of

$$\bar{I}_i = \bar{I}(\mathcal{D}_i) \exp(-(\mathcal{P}\mu)(\vec{s}_i, \sigma_i)) = \bar{I}(\mathcal{D}_i) e^{-\bar{y}_i}.$$

In this case, the negative log-likelihood term fits the template of (2.18) with

$$L_i(\bar{y}_i; I_i) = \bar{I}_i - I_i \ln \bar{I}_i; \tag{2.19}$$

note the implicit change of notation from Y_i, the generalized observation, to I_i, which is specifically the observed number of X-ray photons.

It is useful to note here for future derivation that

$$\left.\frac{\partial^2}{\partial y^2} L_i(I_i; y)\right|_{y=\bar{y}_i} = \operatorname{var}(I_i) \left.\frac{\partial^2}{\partial y \partial I_i} L_i(I_i; y)\right|_{y=\bar{y}_i} = \frac{I_i}{\bar{I}_i^2}. \tag{2.20}$$

Gaussian Likelihood: The other common model used considers, for an observed number of photons I_i at observation i,

$$y_i = -\ln(I_i / \bar{I}(\mathcal{D}_i)). \tag{2.21}$$

This y_i is meant to be a measurement of $\bar{y}_i$ (the mean of y_i, assuming the mean of I_i is $\bar{I}_i$, is very close to $\bar{y}_i$). If I_i is a Poisson random variable and $\bar{I}_i$ is large, then it can be approximated as a Gaussian random variable. For this large $\bar{I}_i$, the standard

deviation of I_i is small compared to its mean, so the transformation of (2.21) can be approximated well by a first-order Taylor expansion. Using this linear approximation, y_i can also be approximated as a Gaussian random variable. A detailed treatment of the statistics of y_i is given in Appendix A, but for the purposes of reconstruction it is often simply treated as Gaussian, and so for our methods we do the same.

Under this Gaussian approximation, the log-likelihood term L_i is given by:

$$L_i(\bar{y}_i; y_i) = \frac{1}{2}\frac{(\bar{y}_i - y_i)^2}{\text{var}(y_i)}. \tag{2.22}$$

Often, $\text{var}(y_i)$ is not exactly known, and so we choose a weighting w_i for each observation and use the data-fit term:

$$L_i(\bar{y}_i; y_i) = \frac{1}{2} w_i (\bar{y}_i - y_i)^2. \tag{2.23}$$

The sum of the data-fit terms for each view gives the data-fit term for all of the observations, given the data:

$$L(\mathbf{y}; \mathbf{x}) = \sum_i \frac{1}{2} w_i ([\mathbf{A}\mathbf{x}]_i - y_i)^2 = \frac{1}{2}||\mathbf{y} - \mathbf{A}\mathbf{x}||^2_{\mathbf{W}}, \tag{2.24}$$

where $\mathbf{W}$ is a diagonal *weighting matrix* where $[\mathbf{W}]_{ii} = w_i$. The weighting is often chosen such that $w_i \approx 1/\text{var}(y_i)$, since although $\text{var}(y_i)$ is not known, it can be estimated from the observation y_i. We leave this generalized, though, as w_i is also sometimes chosen deliberately to not approximate $1/\text{var}(y_i)$ [75]. As with the Poisson case, it is useful here to note for future derivation that:

$$\frac{\partial^2}{\partial y^2} L_i(y_i; y)\bigg|_{y=\bar{y}_i} = w_i, \tag{2.25}$$

and

$$\frac{\partial^2}{\partial y \partial y_i} L_i(y_i; y)\bigg|_{y=\bar{y}_i} = w_i^2. \tag{2.26}$$

2.4.2 Regularization

The majority of the regularizers used in practice take the form

$$R(\mathbf{x}) = \sum_{d=1}^{N_\mathbf{C}} r_d \sum_k \psi([\mathbf{C}_d \mathbf{x}]_k), \tag{2.27}$$

where ψ is some penalty function, $\mathbf{C}_d$ is a sequence of $N_{\mathbf{C}}$ matrices, and r_d is a weight for the penalty associated with the dth matrix. The penalty function ψ is usually chosen to be convex so that the term being minimized in (2.17) is convex. Often $\mathbf{C}_d$ is a matrix that takes the first difference between neighboring samples, pixels, or voxels along directions enumerated by d. For example, in 2D, it is common to penalize the difference between a pixel and its neighbors to the left, top-left, top, and top-right[6], which would require four differencing matrices $\mathbf{C}_1, \ldots, \mathbf{C}_4$ representing differencing along these four directions, respectively. However, we may want to penalize the difference between a pixel and its left neighbor more strongly than the difference between the pixel and its top-left neighbor, which we represent by making r_1 larger than r_2.

When $\mathbf{C}_d$ are first-differencing matrices, (2.27) can be equivalently represented as

$$R(\mathbf{x}) = \sum_{d=1}^{N_{\mathbf{C}}} \sum_{\vec{n} \in \mathbb{Z}^n} r_d \psi(x[\vec{n} + \vec{m}_d] - x[\vec{n}])\ \mathbf{1}_{\vec{n} \in \mathbb{I}, \vec{n} + \vec{m}_d \in \mathbb{I}}, \tag{2.28}$$

where $\vec{m}_d$ represents the lattice offset of the dth difference (e.g., in the 2D example above with 4 differences, $\vec{m}_{1,2,3,4} = \{(1,0),(1,1),(0,1),(-1,1)\}$). Solely in (2.28), we reference elements of the image vector by their location instead of their lexicographic position in $\mathbf{x}$, such that $x[\vec{n}_j] = x_j$. We define $\mathbb{I} \subset \mathbb{Z}^n$ for the purposes of (2.28) to be the set of locations $\vec{n}$ such that $\vec{n}$ is represented in the reconstructed image.

We will exclusively consider regularizers of the form (2.27) with convex penalties ψ that are twice-differentiable around zero and scaled such that $\ddot{\psi}(0) = 1$. In this case, the Hessian of the regularization term is given by

$$\nabla^2 R(\mathbf{x}) = \sum_{d=1}^{N_{\mathbf{C}}} r_d \mathbf{C}_d^{\mathrm{T}} \ddot{\mathbf{\Psi}}_d(\mathbf{x}) \mathbf{C}_d, \tag{2.29}$$

where $\ddot{\mathbf{\Psi}}_d(\mathbf{x})$ is a diagonal matrix of second derivatives of the penalty function, evaluated for an image $\mathbf{x}$:

$$[\ddot{\mathbf{\Psi}}_d(\mathbf{x})]_{jk} \triangleq \begin{cases} \ddot{\psi}([\mathbf{C}_d \mathbf{x}]_k), & k = j \\ 0, & k \neq j \end{cases}. \tag{2.30}$$

[6]The other four neighbors are omitted, since the difference between the pixel of interest and it's neighbor to the right will be taken care of when penalizing the difference between the pixel's right neighbor and the pixel's right neighbor's left neigbor.

We define the matrix $\mathbf{P}$ as:

$$\mathbf{P} \triangleq \sum_{d=1}^{N_{\mathbf{C}}} r_d \mathbf{C}_d^{\mathrm{T}} \mathbf{C}_d. \tag{2.31}$$

This matrix has significance as the Hessian (2.29) evaluated where the penalty function is quadratic, i.e. $\psi(x) = x^2/2$, or where the differences are equal to zero, i.e. $\mathbf{C}_d\mathbf{x} = \mathbf{0}$.

2.4.3 Linear statistical reconstruction

In the case where we have a Gaussian penalty (2.24) and regularization of the form in (2.27) with a quadratic penalty ψ, the total cost function

$$L(\mathbf{y};\mathbf{x}) + \alpha R(\mathbf{x}) = \frac{1}{2}||\mathbf{y} - \mathbf{A}\mathbf{x}||_{\mathbf{W}}^2 + \frac{\alpha}{2}\sum_d r_d ||\mathbf{C}_d\mathbf{x}||_2^2$$

is quadratic in $\mathbf{x}$, and the resulting (unconstrained) reconstruction $\hat{\mathbf{x}}(\mathbf{y})$ is a linear function of $\mathbf{x}$ given simply by:

$$\hat{\mathbf{x}}(\mathbf{y}) = (\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} + \alpha\mathbf{P})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{y}. \tag{2.32}$$

This linear form, while not as common in practice as non-linear reconstruction, is useful for its simplicity.

2.5 Variance prediction methods

When we consider a statistical reconstruction method of the form given in (2.17), with a data-fit term of the form given in (2.18) and a regularization term of the form given in (2.27), the covariance matrix $\mathrm{cov}(\hat{\mathbf{x}})$ of the reconstruction $\hat{\mathbf{x}}$ can be found to a second-order approximation[17] given by:

$$\mathrm{cov}(\hat{\mathbf{x}}) \approx (\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} + \alpha\nabla^2 R(\breve{\mathbf{x}}))^{-1}\mathbf{A}^{\mathrm{T}}\hat{\mathbf{W}}\mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} + \alpha\nabla^2 R(\breve{\mathbf{x}}))^{-1}. \tag{2.33}$$

Here, the vector $\breve{\mathbf{x}} \triangleq \hat{\mathbf{x}}(\bar{\mathbf{Y}})$ represents the reconstruction via (2.17) of noiseless observations. The diagonal matrices $\mathbf{W}$ and $\hat{\mathbf{W}}$ are defined as:

$$[\mathbf{W}]_{ii} \triangleq \left.\frac{\partial^2}{\partial y^2} L_i(Y_i; y)\right|_{y=[\mathbf{A}\breve{\mathbf{x}}]_i} \tag{2.34}$$

$$[\hat{\mathbf{W}}]_{ii} \triangleq \mathrm{var}(Y_i) \cdot \left.\frac{\partial^2}{\partial y \partial Y_i} L_i(Y_i; y)\right|_{y=[\mathbf{A}\breve{\mathbf{x}}]_i} \tag{2.35}$$

The elements of $\mathbf{W}$ and $\hat{\mathbf{W}}$ are given above for the Poisson likelihood model in (2.20), and for the Gaussian model in (2.25) and (2.26). In the Gaussian case, $\mathbf{W}$ is equivalent to the weighting matrix $\mathbf{W}$ in (2.24).

In computing $\mathrm{cov}(\hat{\mathbf{x}})$ or expressions derived from it, $\mathbf{W}$ and $\hat{\mathbf{W}}$ as given in (2.34), (2.35) are not available exactly, since their evaluation relies on noiseless data. However, (2.33) is not very sensitive to the particular values of $\mathbf{W}$ or $\hat{\mathbf{W}}$ used [49], and so we will assume in our methods that at any point these matrices may be an estimate.

We can confirm that for the linear reconstruction of (2.32), (2.33) is actually an equality. Since (2.32) multiplies the observation $\mathbf{y}$ by the reconstructing matrix $(\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} + \alpha\mathbf{P})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}$, the covariance of $\hat{\mathbf{x}}$ is the covariance of $\mathbf{y}$ multiplied by this matrix on both sides:

$$\mathrm{cov}(\hat{\mathbf{x}}) = (\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} + \alpha\mathbf{P})^{-1}\mathbf{A}^{\mathrm{T}} \underbrace{\mathbf{W}\,\mathrm{cov}(\mathbf{y})\mathbf{W}}_{=\hat{\mathbf{W}}} \mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} + \alpha\mathbf{P})^{-1}. \tag{2.36}$$

2.5.1 Computation

The approximation of (2.33) is a useful starting point for producing a map of the image variance, but is not computationally tractable. The Hessian matrix $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} + \alpha\nabla^2 R(\breve{\mathbf{x}})$ is a square matrix with a dimension of the number of image voxels on each side, and so it is not reasonable to store for any realistic image size, and even less reasonable to invert. However, there are many approaches in previous work toward computing subsets of the image covariance in ways that are computationally possible.

2.5.1.1 Single-voxel inversion

We define $\mathbf{e}_j$ to be the unit vector consisting of all zeros except for a single one at element j. Define $\tilde{\mathbf{e}}_j$ to be:

$$\tilde{\mathbf{e}}_j \triangleq (\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} + \alpha\nabla^2 R(\breve{\mathbf{x}}))^{-1}\mathbf{e}_j,$$

possibly using some estimate for $\breve{\mathbf{x}}$ if it is unknown. The vector $\tilde{\mathbf{e}}_j$ is intractible to compute exactly, but we can approximate it by solving the system

$$(\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} + \alpha\nabla^2 R(\breve{\mathbf{x}}))\tilde{\mathbf{e}}_j = \mathbf{e}_j \tag{2.37}$$

using some method that does not require the inversion of $(\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} + \alpha\nabla^2 R(\breve{\mathbf{x}}))$. We can use this to compute one element of (2.36) as:

$$\begin{aligned}
[\mathrm{cov}(\hat{\mathbf{x}})]_{jk} &= \mathbf{e}_j^{\mathrm{T}} \mathrm{cov}(\hat{\mathbf{x}})\mathbf{e}_k \\
&= \mathbf{e}_j^{\mathrm{T}}(\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} + \alpha\nabla^2 R(\breve{\mathbf{x}}))^{-1}\mathbf{A}^{\mathrm{T}}\hat{\mathbf{W}}\mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} + \alpha\nabla^2 R(\breve{\mathbf{x}}))^{-1}\mathbf{e}_k \\
&= \tilde{\mathbf{e}}_j^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\hat{\mathbf{W}}\mathbf{A}\tilde{\mathbf{e}}_k.
\end{aligned} \tag{2.38}$$

We can therefore predict the variance of voxel j by solving (2.37) to find $\tilde{\mathbf{e}}_j$ and computing $\tilde{\mathbf{e}}_j^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\hat{\mathbf{W}}\mathbf{A}\tilde{\mathbf{e}}_j$. This is perhaps the most accurate of the methods for computing one element of $\mathrm{cov}(\hat{\mathbf{x}})$, but it requires solving (2.37) for each voxel of interest, which would take an unreasonable amount of computation to produce a variance map for a full image.

2.5.1.2 Discrete frequency-domain methods — prior work

Previous work has approximately evaluated (2.33) using frequency-domain approximations to $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$ and $\mathbf{P}$. The basis of these methods is replacing $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$ with a spatially shift-invariant approximation, which is then diagonalized by the DFT or DSFT (discrete space Fourier transform).

Suppose that the voxel indices of the image (in $\mathbb{Z}^n$) are contained within a finite volume of size $\vec{N}$, with volume $|\vec{N}| = N_1 N_2 \cdots N_n$, such that $[\vec{n}_j]_d \in \{0, 1, \ldots N_d - 1\}$ for each dimension $d = \{1, \ldots, n\}$. We define $\mathbf{E} \in \mathbb{R}^{|\vec{N}| \times N_{\mathrm{vox}}}$ as an operator that embeds the image vector $\mathbf{x}$ into this volume of size $\vec{N}$. This operator can be seen as a tall matrix defined such that

$$[\mathbf{E}\mathbf{x}][\vec{n}] = \begin{cases} x_j, & \text{if } \vec{n} = \vec{n}_j \\ 0, & \text{if } \vec{n} \neq \vec{n}_j \text{ for all } j \end{cases},$$

where the elements in $\mathbb{R}^{|\vec{N}|}$ are indexed not by orthographic position in a vector, but by their position $\vec{n} \in \mathbb{Z}_{N_1} \times \cdots \times \mathbb{Z}_{N_n}$. The transpose of this matrix, which selects specific voxel indices in this volume and puts them back in their vector order, can be defined by $[\mathbf{E}^{\mathrm{T}}\mathbf{z}]_j = z[\vec{n}_j]$.

We use this embedding matrix to factor the projection matrix $\mathbf{A}$ into $\mathbf{A} = \mathbf{A}_f\mathbf{E}$, where $\mathbf{A}_f \in \mathbb{R}^{N_d \times |\vec{N}|}$ is a matrix that projects the entire volume of size $\vec{N}$ into N_p observations. If $\mathbf{A}_f^{\mathrm{T}}\mathbf{W}\mathbf{A}_f$ were circularly shift-invariant in this volume, then we could represent it as

$$\mathbf{A}_f^{\mathrm{T}}\mathbf{W}\mathbf{A}_f = \mathbf{F}^{\mathrm{T}}\hat{\mathbf{H}}^W\mathbf{F}, \tag{2.39}$$

where $\mathbf{F}$ is an n-dimensional DFT (the operator $\oslash$ here represents element-wise division):

$$\mathbf{F}_{\vec{k},\vec{\ell}} = |\vec{N}|^{-1/2} \exp(-\jmath 2\pi \vec{k} \cdot \vec{\ell} \oslash \vec{N}),$$

and $\hat{\mathbf{H}}^W$ would be a diagonal matrix representing the spectrum of $\mathbf{A}_f^{\mathrm{T}}\mathbf{W}\mathbf{A}_f$. Here, we index the elements of the matrix by the corresponding positions $\vec{k}$ in discrete frequency space and $\vec{\ell}$ in image space. When we add the embedding operator on to each side of (2.39), it becomes

$$\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} = \mathbf{E}^{\mathrm{T}}\mathbf{A}_f^{\mathrm{T}}\mathbf{W}\mathbf{A}_f\mathbf{E} = \mathbf{E}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\hat{\mathbf{H}}^W\mathbf{F}\mathbf{E}. \tag{2.40}$$

We will use this frequency domain representation to approximate the inverse of $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$; since $(\mathbf{F}^{\mathrm{T}}\hat{\mathbf{H}}^W\mathbf{F})^{-1} = \mathbf{F}^{\mathrm{T}}(\hat{\mathbf{H}}^W)^{-1}\mathbf{F}$,

$$(\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A})^{-1} \approx \mathbf{E}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}(\hat{\mathbf{H}}^W)^{-1}\mathbf{F}\mathbf{E}. \tag{2.41}$$

Even if $\mathbf{A}_f^{\mathrm{T}}\mathbf{W}\mathbf{A}_f$ were circularly shift-invariant, (2.41) would be an approximation and not an exact inverse since it approximates $\mathbf{E}^{\mathrm{T}}\mathbf{E} = \mathbf{I}_{N_{\mathrm{vox}}}$ and $\mathbf{E}\mathbf{E}^{\mathrm{T}} = \mathbf{I}_{|\vec{N}|}$. This former claim is indeed true, but the effect of $\mathbf{E}\mathbf{E}^{\mathrm{T}}$ is to zero out any voxels in $\mathbf{R}^{|\vec{N}|}$ that are not represented by an element of the image vector. This effect is not the same as the identity matrix, but for the usefulness of (2.41), we approximate $\mathbf{E}\mathbf{E}^{\mathrm{T}} \approx \mathbf{I}_{|\vec{N}|}$.

The local impulse response (LIR) of $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$ for the voxel j is defined by

$$\mathbf{h}_j^W \triangleq \mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}\mathbf{e}_j, \tag{2.42}$$

where $\mathbf{e}_j$ is again defined as the unit vector with a single 1 at position j. This LIR can be written exactly as the impulse $\mathbf{e}_j$ operated on by a frequency-domain filter $\hat{\mathbf{H}}_j^W$:

$$\mathbf{h}_j^W = \mathbf{E}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\hat{\mathbf{H}}_j^W\mathbf{F}\mathbf{E}\mathbf{e}_j, \tag{2.43}$$

where $\hat{\mathbf{H}}_j^W$ is a filter local to voxel j. This filter is the DFT of the local impulse response centered around voxel j. The matrix $\mathbf{F}^{\mathrm{T}}\hat{\mathbf{H}}_j^W\mathbf{F}$ represents performing the same circular convolution that transforms the impulse $\mathbf{e}_j$ into the LIR $\mathbf{h}_j^W$ across the entire

image. Equations (2.42) and (2.43) together suggest the circulant approximation

$$\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} \approx \mathbf{E}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\hat{\mathbf{H}}_j^W\mathbf{F}\mathbf{E}, \tag{2.44}$$

which, in general, is not a reasonable approximation for the entire matrix $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$, but is a reasonable for a voxel k near voxel j:

$$[\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}]_{kj} \approx \mathbf{e}_k^{\mathrm{T}}\mathbf{E}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\hat{\mathbf{H}}_j^W\mathbf{F}\mathbf{E}\mathbf{e}_j.$$

Except near edges in the image, the same circulant approximation methods work for the regularizer Hessian:

$$\nabla^2 R(\mathbf{x}) \approx \mathbf{E}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\hat{\mathbf{R}}\mathbf{F}\mathbf{E}. \tag{2.45}$$

With $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$ and $\nabla^2 R(\mathbf{x})$ approximately diagonalized as in (2.44) and (2.45), the covariance expression in (2.33) becomes:

$$\mathrm{cov}(\hat{\mathbf{x}}) \approx \mathbf{E}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{S}_j\mathbf{F}\mathbf{E}, \tag{2.46}$$

where

$$\mathbf{S}_j \triangleq \hat{\mathbf{H}}_j^{\hat{W}}(\hat{\mathbf{H}}_j^W + \alpha\hat{\mathbf{R}})^{-2}. \tag{2.47}$$

The diagonal matrix $\mathbf{S}_j$ is analogous to a local noise power spectrum (LNPS). Predicting the variance of one voxel from (2.46) becomes:

$$\begin{aligned}\mathrm{var}(\hat{x}_j) = \mathbf{e}_j^{\mathrm{T}}\,\mathrm{cov}(\hat{\mathbf{x}})\mathbf{e}_j = \mathbf{e}_j^{\mathrm{T}}\mathbf{E}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{S}_j\mathbf{F}\mathbf{E}\mathbf{e}_j = \frac{1}{|\vec{N}|}\sum_{\vec{k}}\mathbf{S}_j[\vec{k}] \\ = \frac{1}{|\vec{N}|}\sum_{\vec{k}}\frac{\hat{\mathbf{H}}_j^{\hat{W}}[\vec{k}]}{(\hat{\mathbf{H}}_j^W[\vec{k}] + \alpha\hat{\mathbf{R}}[\vec{k}])^2}.\end{aligned} \tag{2.48}$$

In practice, the steps for using this method to approximate variance is to compute the local impulse response $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}\mathbf{e}_j$ using some implementation of the projector $\mathbf{A}$, find the n-dimensional DFT via the FFT of this impulse response, and sum (2.48). We will refer to these as "DFT-based" methods. These have been used in [47, 19, 57]. Their use is more computationally tractable than (2.38) for approximating the variance of one voxel of interest, but it requires the computation of $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}\mathbf{e}_j$ for each voxel of interest, which is still very time-consuming for variance approximation over many voxels.

2.5.1.3 Continuous frequency-domain methods — prior work

Further acceleration of variance approximation can be found in [76, 77], which obviate the need for computing $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}\mathbf{e}_j$ by using direct theory-based approximations for the local frequency response of $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$. These methods work in the continuous-frequency space of the DSFT instead of the discrete-frequency DFT. Analogous to (2.48), they integrate the continuous local noise power spectrum as opposed to summing the discrete local NPS, but by changing the coordinates of the integral, these methods are able to reduce the dimensionality of the NPS integral. This results in closed-form expressions for $\mathrm{var}(\hat{x}_j)$ that are much faster to compute than previous methods. These methods, though, are limited to specific CT geometries; 2D fan-beam CT in [76] and 3D step-and-shoot CT in [77]. We will take a similar continuous-frequency approach to [76, 77] to predict variance, but with approximations to the local frequency response of $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$ that apply in more geometries than that previous work.

We can analyze $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$ and $\mathbf{P}$ in using the DSFT in nearly the same way as in the previous section. We define an embedding operator $\mathcal{E}$ that embeds the image vector $\mathbf{x}$ into the infinite space $\mathbb{Z}^n$:

$$(\mathcal{E}\mathbf{x})[\vec{n}] = \begin{cases} x_j, & \text{if } \vec{n} = \vec{n}_j \\ 0, & \text{if } \vec{n} \neq \vec{n}_j \text{ for all } j \end{cases}; \tag{2.49}$$

the adjoint of $\mathcal{E}$, a selection operator, is defined as

$$(\mathcal{E}^* x)_j = x[\vec{n}_j]. \tag{2.50}$$

We use this embedding operator to factor the projection $\mathbf{A}$ into $\mathbf{A} = \mathcal{A}_f\mathcal{E}$, where $\mathcal{A}_f$ is an operator that projects not just the voxels in $\mathbf{x}$, but the entirety of $\mathbb{Z}^n$. Clearly, $\mathcal{A}_f$ is not implementable in any form, but it is a useful fiction. If $\mathcal{A}_f^*\mathbf{W}\mathcal{A}_f$ were shift-invariant, but unlike in (2.39), not circularly shift-invariant, it is diagonalized with an n-dimensional DSFT:

$$\mathcal{A}_f^*\mathbf{W}\mathcal{A}_f = \mathcal{F}_f^*\mathcal{D}\left\{H^W\right\}\mathcal{F}_f, \tag{2.51}$$

where $\mathcal{D}$ is a "diagonalization" operator:

$$\left(\mathcal{D}\left\{H\right\}X\right)(\vec{\nu}) = H(\vec{\nu})X(\vec{\nu}),$$

and the full DSFT of $\mathbb{Z}^n$ and its adjoint are defined as:

$$(\mathcal{F}_f x)(\vec{\nu}) = \sum_{\vec{n}\in\mathbb{Z}^n} x[\vec{n}] \exp\left(-\jmath 2\pi \vec{\nu}^{\mathrm{T}} \vec{n}\right)$$
$$(\mathcal{F}_f^* X)[\vec{n}] = \int\limits_{\left[-\frac{1}{2},\frac{1}{2}\right]^n} X(\vec{\nu}) \exp\left(\jmath 2\pi \vec{\nu}^{\mathrm{T}} \vec{n}\right)\, d\vec{\nu},$$

where $\vec{\nu}$ has units of cycles per sample. We mark these as the 'full' DSFT since we will also consider the DSFT with just the voxels in the image selected, which we represent as $\mathcal{F}$:

$$(\mathcal{F}\mathbf{x})(\vec{\nu}) = \sum_k x_k \exp\left(-\jmath 2\pi \vec{\nu}^{\mathrm{T}} \vec{n}_k\right)$$
$$[\mathcal{F}^* X]_k = \int\limits_{\left[-\frac{1}{2},\frac{1}{2}\right]^n} X(\vec{\nu}) \exp\left(\jmath 2\pi \vec{\nu}^{\mathrm{T}} \vec{n}_k\right)\, d\vec{\nu}. \tag{2.52}$$

For these definitions, the limited DSFT (2.52) is equivalent to an embedding followed by the full DSFT, that is, $\mathcal{F} = \mathcal{F}_f\mathcal{E}$. By adding the embedding and its adjoint to (2.51), we find that $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$ can be diagonalized by H^W:

$$\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} = \mathcal{F}^*\mathcal{D}\left\{H^W\right\}\mathcal{F}.$$

The local impulse response from (2.42) can be written exactly as the impulse $\mathbf{e}_j$ operated on by a frequency-domain filter $H_j^W(\vec{\nu})$:

$$\mathbf{h}_j^W = \mathcal{F}^*\mathcal{D}\left\{H_j^W\right\}\mathcal{F}\mathbf{e}_j, \tag{2.53}$$

where

$$H_j^W(\vec{\nu}) = \exp\left(\jmath 2\pi \vec{\nu}^{\mathrm{T}} \vec{n}_j\right)\ (\mathcal{F}\mathbf{h}_j^W)(\vec{\nu}). \tag{2.54}$$

The exponential term "centers" the transform at the jth voxel using the shift property of the DSFT.

We will refer to H_j^W as a *local* frequency response (LFR). Again, in the region near voxel j, $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$ is typically *approximately* spatially shift-invariant, in the sense that $\mathbf{A} = \mathcal{A}_f\mathcal{E}$ for an approximately shift-invariant $\mathcal{A}_f$. This leads us to the approximation

$$[\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}]_{kj} \approx \mathbf{e}_k^{\mathrm{T}}\mathcal{F}^*\mathcal{D}\left\{H_j^W\right\}\mathcal{F}\mathbf{e}_j, \tag{2.55}$$

for voxel k near voxel j, which is suggested by (2.42) and (2.53). Except at the edges of the reconstructed image, $\mathbf{P}$ can again also be represented by its frequency response $R(\vec{\nu})$:

$$\mathbf{P} = \mathcal{F}^* \mathcal{D} \left\{ R \right\} \mathcal{F}. \tag{2.56}$$

The regularizers that are representable in the form of (2.28) have a closed-form expression for $R(\vec{\nu})$, given in [77]:

$$R(\vec{\nu}) = \sum_{d=1}^{N_{\mathbf{C}}} 4 r_d \sin^2 \left(\pi \vec{m}_d \cdot \vec{\nu} \right). \tag{2.57}$$

We assume for variance prediction that $\ddot{\mathbf{\Psi}}_d(\breve{\mathbf{x}}) \approx \mathbf{I}$, such that $\nabla^2 R(\breve{\mathbf{x}}) \approx \mathbf{P}$. This assumption is based on the idea that the majority of neighboring-voxel differences in the reconstruction $\breve{\mathbf{x}}$ from noise-free projection data will be small, since the regularizer penalizes large neighboring-voxel differences. We hope that for these small differences, the second derivative of their penalties will be near 1. The utility of this assumption to our purposes of fast variance prediction is enormous. First, variance prediction using $\nabla^2 R(\breve{\mathbf{x}})$ requires foreknowledge of the noiseless reconstruction $\breve{\mathbf{x}}$. For non-phantom applications, $\breve{\mathbf{x}}$ is clearly unavailable. Second, even using a noisy reconstruction $\hat{\mathbf{x}}$ requires the time to compute $\hat{\mathbf{x}}$, which is much greater than the computation time of our fast methods. This would diminish the utility of our methods in application. Third, our methods require precomputation time to produce lookup tables for a particular shape of the regularization penalty. We could maintain the utility of our methods if we could, near an edge, approximate $\nabla^2 R(\breve{\mathbf{x}})$ with the original Hessian $\mathbf{P}$, but scaled by a factor determined by the proximity of the voxel to an edge. However, the presence of an edge not only scales the Hessian of the regularizer near the edge, but also changes its shape; if the Hessian were isotropic before, the edge will deform it. In our methods, the edge both breaks the assumption that $\nabla^2 R(\breve{\mathbf{x}})$ is shift-invariant, and our ability to use pre-computation to accelerate our algorithm. For our derivations, we will assume that $\nabla^2 R(\breve{\mathbf{x}}) = \mathbf{P}$.

With both of these matrices diagonalized, (2.33) simplifies, locally to a voxel j, to

$$\operatorname{cov}(\hat{\mathbf{x}}) \approx \mathcal{F}^* \mathcal{D} \left\{ S_j \right\} \mathcal{F}, \tag{2.58}$$

where

$$S_j(\vec{\nu}) \triangleq \frac{H_j^{\hat{W}}(\vec{\nu})}{(H_j^W(\vec{\nu}) + \alpha R(\vec{\nu}))^2} \tag{2.59}$$

is the local noise power spectrum (LNPS) of the reconstruction. Note here the difference between H_j^W and $H_j^{\hat{W}}$; $H_j^{\hat{W}}$ comes from the same derivation as (2.42)–(2.55), but with $\hat{\mathbf{W}}$ substituted for $\mathbf{W}$. As in the DFT case of the previous section, (2.58) requires approximating $\mathcal{E}^*\mathcal{E} = I_{N_{\text{vox}}}$ and $\mathcal{E}\mathcal{E}^* = \mathcal{I}$. The former is true, the latter has the effect of masking out any voxels in $\mathbb{Z}^n$ not represented in the image vector.

If $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$ were shift-invariant, so that (2.55) were exact and not a local approximation, (2.59) would be the global NPS. Extracting the variance of one voxel can be done by left- and right-multiplying the covariance matrix by unit vectors:

$$\operatorname{var}(\hat{x}_j) = \mathbf{e}_j^{\mathrm{T}} \operatorname{cov}(\hat{\mathbf{x}})\mathbf{e}_j;$$

plugging in the approximation (2.58) to this expression simplifies it to an integral of the LNPS:

$$\operatorname{var}(\hat{x}_j) \approx \int_{\left[-\frac{1}{2},\frac{1}{2}\right]^n} S_j(\vec{\nu})\, d\vec{\nu}, \tag{2.60}$$

analogous to (2.48).

Chapter III focuses on approximating the LFR H_j^W. Chapter IV then explores how these approximations are able to simplify (2.60) in a way that significantly reduces the computation needed.

CHAPTER III

Frequency Response Approximation

In this chapter, we develop an approximation to the local frequency response defined in (2.54) of $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$, the projection, weighting, and back-projection operator crucial to variance prediction and other frequency-domain methods. In (3.7), we find a Taylor expansion that approximates an arbitrary CT geometry locally as a parallel-beam geometry with a flat detector. This Taylor expansion allows us to approximate the correlation of the projections of two nearby voxels using the Fourier-slice theorem. In (3.6), we translate the effect of $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$ into this CT geometry, approximating its elements as the correlation of the projections of a pair of voxels, to which we can apply this Fourier-slice theorem based approximation to the correlation. These methods culminate in the expression (3.25), which is a separable approximation of the local frequency response of $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$, which has the property that the terms dependent on the scanned image are not dependent on the frequency magnitude, which will prove useful for accelerating variance prediction methods.

3.1 The ray and projection transforms

We begin by examining one element of $\mathbf{h}_j^W$, the LIR of $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$ applied to voxel j. Writing the matrix multiplication that defines this element as a sum over views i gives:

$$\left[\mathbf{h}_j^W\right]_k = \left[\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}\right]_{kj} = \sum_i \left[\mathbf{A}\right]_{ik} \left[\mathbf{W}\right]_{ii} \left[\mathbf{A}\right]_{ij}. \tag{3.1}$$

Recall the definition of the ray transform operator $\mathcal{R}$, that transforms a function f into line integrals through it:

$$(\mathcal{R}f)(\vec{x},\hat{\theta}) \triangleq \int_{\mathcal{R}} f(\vec{x}+\tau\hat{\theta})\, d\tau, \tag{3.2}$$

where $\vec{x}$ is a point on a ray and $\hat{\theta}$ is a unit vector representing a ray direction. Also recall that the projection operator $\mathcal{P}$ is simply a rebinning of $\mathcal{R}$:

$$(\mathcal{P}f)(\vec{s},\sigma) \triangleq (\mathcal{R}f)(\vec{x}(\vec{s},\sigma),\hat{\theta}(\vec{s},\sigma)); \tag{3.3}$$

this continuous-domain projection $\mathcal{P}f$ is defined over the $(n-1)$-dimensional detector position $\vec{s}$ and 1-dimensional source position σ. The function $\vec{x}(\vec{s},\sigma)$ maps a source position and detector position pair to a point on the ray that connects the pair. The function $\hat{\theta}(\vec{s},\sigma)$ maps these positions to a unit vector that lies along the ray direction. We will also use $\vec{s}(\vec{x},\sigma)$ to represent the detector coordinate corresponding to the ray that passes through the source at position σ and the spatial position $\vec{x}$. These functions are defined by the geometry of the CT system under consideration, and are assumed to be known.

With this definition for $\mathcal{P}$, the elements of $\mathbf{A}$ are approximated as discretized samples of the continuous projection:

$$[\mathbf{A}]_{ik} \approx (\mathcal{P}R_k)(\vec{s}_i,\sigma_i), \tag{3.4}$$

that is, the ith observation of a Kronecker impulse at voxel k is approximately the projection of the kth basis voxel, R_k, sampled at one detector position $\vec{s}_i$ with the source at one position σ_i. The approximation (3.4) for one element of the projection matrix is mismatched from the $\mathbf{A}$ that would be used in the implementation of the reconstruction (2.17) in two ways. First, the approximation assumes that the projection can be measured at exactly one point, which neglects detector blur. Second, (3.4) is defined to be an *exact* projection at this particular location, whereas an implemented system matrix $\mathbf{A}$ (e.g., [35]) will use approximations to the integral through a basis function.

In the same way that we replace $\mathbf{A}$ with a samples of a continuous function, we define $w(\vec{s},\sigma)$ to denote a continuous (i.e., interpolated) version of the elements of the weighting matrix $\mathbf{W}$. The actual method used for interpolation is relatively unimportant, but we assume that w is a function such that $w(\vec{s}_i,\sigma_i) = [\mathbf{W}]_{ii}$. With these continuous analogs for $\mathbf{A}$ and $\mathbf{W}$, we rewrite the sum (3.1) as:

$$\left[\mathbf{h}_j^W\right]_k = \sum_i (\mathcal{P}R_j)(\vec{s}_i,\sigma_i)\ (\mathcal{P}R_k)(\vec{s}_i,\sigma_i)\ w(\vec{s}_i,\sigma_i). \tag{3.5}$$

For typical clinical CT scans (but not for sparse view acquisitions), this sum is usually

sufficiently finely sampled that we can approximate it with an integral over $\vec{s}$ and σ:

$$\left[\mathbf{h}_j^W\right]_k \approx \int_{\Sigma} \int_{\mathbb{R}^{n-1}} (\mathcal{P}R_j)(\vec{s},\sigma)\ (\mathcal{P}R_k)(\vec{s},\sigma)\ u(\vec{s},\sigma)\, d\vec{s}\, d\sigma. \tag{3.6}$$

Here, Σ represents a continuous approximation to the domain of σ for which we collect observations; for all the system geometries we have examined this is a single interval $[\sigma_{\min}, \sigma_{\max}]$, but for applications such as gated CT this may be the union of multiple intervals. The term u in (3.6) is the product of w with a Jacobian term representing the "size" of a sample in the sum (3.5). In the usual case where the detector pixels have a constant area $\Delta_{\vec{s}}$ and the spacing (in radians) between source positions is a constant Δ_σ, then $u(\vec{s},\sigma) = w(\vec{s},\sigma)/\Delta_{\vec{s}}\Delta_\sigma$. However, for geometries where the spacing is nonuniform, u will not simply be a scaling of w. The approximation of (3.6) also includes expanding the domain of detector positions $\vec{s}$ to $\mathbb{R}^{n-1}$; to compensate for this we consider the weighting $w(\vec{s},\sigma)$, and hence $u(\vec{s},\sigma)$, to be zero in the regions where we have no observations.

3.2 Linearization of projection transform

If we fix a spatial position $\vec{x}_j$ and source position σ, this also fixes a detector location $\vec{s}_{j,\sigma} \triangleq \vec{s}(\vec{x}_j,\sigma)$ and ray direction $\hat{\theta}_{j,\sigma} \triangleq \hat{\theta}(\vec{s}_{j,\sigma},\sigma)$, such that $(\mathcal{R}f)(\vec{x}_j,\hat{\theta}_{j,\sigma}) = (\mathcal{P}f)(\vec{s}_{j,\sigma},\sigma)$. The coordinate mappings between $\mathcal{P}$ and $\mathcal{R}$ are usually regular enough that a first-order Taylor expansion is quite accurate for small perturbations $\vec{\Delta x}$:

$$(\mathcal{R}f)(\vec{x}_j + \vec{\Delta x}, \hat{\theta}_{j,\sigma}) \approx (\mathcal{P}f)(\vec{s}_{j,\sigma} + \mathbf{B}_{j,\sigma}\vec{\Delta x}, \sigma), \tag{3.7}$$

where

$$\mathbf{B}_{j,\sigma} \triangleq \nabla_{\vec{x}} \vec{s}(\vec{x},\sigma^*)|_{\vec{x}=\vec{x}_j,\sigma^*=\sigma}$$

is a $(n-1)\times n$ matrix that "linearizes" the system geometry, locally to $\vec{x}_j$ and σ, to a parallel-beam, flat-panel geometry. We can also reverse (3.7) to find a spatial shift that corresponds to a particular detector position shift:

$$\begin{aligned}(\mathcal{R}f)(\vec{x}_j + \mathbf{B}_{j,\sigma}^+\vec{\Delta s}, \hat{\theta}_{j,\sigma}) &\approx (\mathcal{P}f)(\vec{s}_{j,\sigma} + \mathbf{B}_{j,\sigma}\mathbf{B}_{j,\sigma}^+\vec{\Delta s}, \sigma) \\ &= (\mathcal{P}f)(\vec{s}_{j,\sigma} + \vec{\Delta s}, \sigma),\end{aligned} \tag{3.8}$$

where $\mathbf{B}_{j,\sigma}^+$ is the $n\times(n-1)$ pseudo-inverse of $\mathbf{B}_{j,\sigma}$.

We will also need in the QR factorization of $\mathbf{B}_{j,\sigma}$:

$$\mathbf{B}_{j,\sigma} = \mathbf{R}_{j,\sigma}\mathbf{Q}_{j,\sigma}, \tag{3.9}$$

where $\mathbf{R}_{j,\sigma}$ is a $(n-1) \times (n-1)$ matrix and $\mathbf{Q}_{j,\sigma}$ is a $(n-1) \times n$ matrix with orthonormal rows; each of these rows is also orthogonal to $\hat{\theta}_{j,\sigma}$.

3.3 Towards local shift invariance

We define

$$c_{kj,\sigma} \triangleq \int_{\mathbb{R}^{n-1}} (\mathcal{P}R_j)(\vec{s},\sigma)\ (\mathcal{P}R_k)(\vec{s},\sigma)\, d\vec{s} \tag{3.10}$$

to be the integral over the detector plane of the product of the continuous projections of the voxel basis functions R_j and R_k for a specific source position σ, and

$$\breve{u}_{kj,\sigma} \triangleq \frac{\int_{\mathbb{R}^{n-1}} (\mathcal{P}R_j)(\vec{s},\sigma)\ (\mathcal{P}R_k)(\vec{s},\sigma) u(\vec{s},\sigma)\, d\vec{s}}{\int_{\mathbb{R}^{n-1}} (\mathcal{P}R_j)(\vec{s},\sigma)\ (\mathcal{P}R_k)(\vec{s},\sigma)\, d\vec{s}}, \tag{3.11}$$

so that the approximation to the LIR in (3.6) becomes

$$\left[\mathbf{h}_j^W\right]_k \approx \int_{\Sigma} c_{kj,\sigma}\breve{u}_{kj,\sigma}\, d\sigma. \tag{3.12}$$

Next we simplify each term in the integrand using approximations that remove most of their dependence on k.

3.3.1 Simplifying $c_{jk,\sigma}$

We make the usual assumption that the basis functions for voxels k and j have the same shape, and differ only by translation, i.e.,

$$R_k(\vec{x}) = R_j(\vec{x} - (\vec{x}_k - \vec{x}_j)).$$

Given this spatial relationship, their ray transforms are related by translation:

$$\begin{aligned}(\mathcal{R}R_k)(\vec{x},\hat{\theta}) &= \int R_k(\vec{x}+\tau\hat{\theta})\,d\tau \\ &= \int R_j(\vec{x}+\tau\hat{\theta}-(\vec{x}_k-\vec{x}_j))\,d\tau \\ &= (\mathcal{R}R_j)(\vec{x}-(\vec{x}_k-\vec{x}_j),\hat{\theta}). \end{aligned} \tag{3.13}$$

Using the Taylor expansion in (3.7), the projection footprint of nearby voxels are approximately related by the following translation:

$$(\mathcal{P}R_k)(\vec{s},\sigma) \approx (\mathcal{P}R_j)(\vec{s}-\mathbf{B}_{j,\sigma}(\vec{x}_k-\vec{x}_j),\sigma); \tag{3.14}$$

that is, for a fixed σ, the projection of voxel k can be approximated as a shift of the projection of voxel j, so long as $\vec{x}_k$ is sufficiently close to $\vec{x}_j$. Using (3.14), we rewrite the correlation $c_{kj,\sigma}$ as the autocorrelation of just the projection of the jth voxel:

$$\begin{aligned} c_{kj,\sigma} &\triangleq \int_{\mathbb{R}^{n-1}} (\mathcal{P}R_j)(\vec{s},\sigma)(\mathcal{P}R_k)(\vec{s},\sigma)\,d\vec{s} \\ &\approx \int_{\mathbb{R}^{n-1}} (\mathcal{P}R_j)(\vec{s},\sigma)(\mathcal{P}R_j)(\vec{s}-\mathbf{B}_{j,\sigma}(\vec{x}_k-\vec{x}_j),\sigma)\,d\vec{s} \\ &= (\mathcal{A}\mathcal{P}R_j)(\mathbf{B}_{j,\sigma}(\vec{x}_k-\vec{x}_j),\sigma) \qquad (3.15) \\ &\triangleq \tilde{c}_{kj,\sigma}, \qquad (3.16) \end{aligned}$$

where $\mathcal{A}$ denotes an autocorrelation operator:

$$(\mathcal{A}f)(\vec{x}) \triangleq \int f(\vec{t})f(\vec{x}+\vec{t})\;d\vec{t}.$$

That is, this correlation between the projections of voxels j and k is approximately a function of the autocorrelation of the projection of just voxel j, and the dependence on k is only via its location relative to j.

3.3.2 Simplifying $\breve{u}$

For a given source position σ, the expression for $\breve{u}$ in (3.11) is the integral over the detector of the product of the continuous projections of R_j and R_k with u, normalized by the integral of just the projections of R_j and R_k. The projections of each of R_j and R_k have a small support, and so does their product; we can therefore approximate the

effect of (3.11) as a sifting that selects one value of u much like a Dirac impulse. We assume that u varies slowly with respect to $\vec{s}$, which is often true for typical choices of $\mathbf{W}$. The peak of $\mathcal{P}R_j$ is located near $\vec{s}(\vec{x}_j, \sigma)$, the detector location that corresponds to a ray originating at the detector at position σ and passing through the center of the jth voxel location, $\vec{x}_j$. For any other voxel k, if the integrand of (3.5) is non-zero, the peak of the projection of k must overlap the projection of j, meaning it is close enough to the peak of the projection of k that we can make the approximation:

$$\breve{u}_{kj,\sigma} \approx \tilde{u}_{j,\sigma} \triangleq u(\vec{s}(\vec{x}_j, \sigma), \sigma). \tag{3.17}$$

We compute $\tilde{u}$ by looking up the value of w for the detector pixel closest to $\vec{s}(\vec{x}_j, \sigma)$. With the approximations (3.15) and (3.17), the LIR in (3.12) becomes:

$$\left[\mathbf{h}_j^W\right]_k \approx \int \tilde{u}_{j,\sigma}(\mathcal{A}\mathcal{P}R_j)(\mathbf{B}_{j,\sigma}(\vec{x}_k - \vec{x}_j), \sigma)\, d\sigma. \tag{3.18}$$

This is our final "space domain" approximation to the original LIR expression in (3.1). Next we move to the frequency domain.

3.3.3 Local frequency response of $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$

We use the form (3.18) for $\left[\mathbf{h}_j^W\right]_k$ to find the empirical LFR, in (3.20) through (3.24), using (2.54). In (3.20), we take (2.54) and insert the approximation (3.18) for $\mathbf{h}_j^W$. In (3.21), we interchange the sum over voxels k and the integral over source position σ, and move the $\tilde{u}_{j,\sigma}$ term out of the sum, as it does not depend on k. In (3.22), we approximate the sum over k with an integral. This replacement assumes that, first, the summand term is slowly varying enough that we can make an integral approximation, and second, that the summand decays rapidly enough beyond a region of interest that we can replace it with an integral not just over the image support, but over $\mathbb{R}^n$. This integral is the continuous Fourier transform of $(\mathcal{A}\mathcal{P}R_j)(\mathbf{B}_{j,\sigma}\mathbf{V}\vec{n})$ over $\vec{n}$ in $\mathbb{R}^n$, and so in (3.23), we replace it with its Fourier transform. We must be careful, though, since we are taking the n-dimensional transform of $\mathcal{A}\mathcal{P}R_j$, an $(n-1)$-dimensional function. The $\mathbf{R}_{j,\sigma}$ that appears here comes from the QR factorization (3.9). Finally, in (3.24), we use the Dirac impulse in (3.23) to sift out only a finite number of σ values in the integral such that $\hat{\theta}_{j,\sigma}^{\mathrm{T}}\mathbf{V}^{-\mathrm{T}}\vec{\nu} = 0$. We define this set to be $\mathcal{B}_j\left(\frac{\vec{\nu}}{||\vec{\nu}||}\right)$, as follows:

$$\mathcal{B}_j\left(\frac{\vec{\nu}}{||\vec{\nu}||}\right) \triangleq \left\{\sigma : \hat{\theta}_{j,\sigma}^{\mathrm{T}}\mathbf{V}^{-\mathrm{T}}\vec{\nu} = 0\right\}. \tag{3.19}$$

Note that $\mathcal{B}_j$ depends on spatial frequency variable $\vec{\nu}$ *only* via its direction, $\vec{\nu}/||\vec{\nu}||$, and not its magnitude.

$$
\begin{aligned}
H_j^W(\vec{\nu}) &= \mathcal{D}\left\{\exp\left(\jmath 2\pi \vec{\nu}^{\mathrm{T}} \vec{n}_j\right)\right\} \mathcal{F}\mathbf{h}_j^W \\
&= \exp\left(\jmath 2\pi \vec{\nu}^{\mathrm{T}} \vec{n}_j\right) \sum_k \left[\mathbf{h}_j^W\right]_k \exp\left(-\jmath 2\pi \vec{\nu}^{\mathrm{T}} \vec{n}_k\right) \\
&\approx \sum_k \left(\int \tilde{u}_{j,\sigma} (\mathcal{A}\mathcal{P}R_j)(\mathbf{B}_{j,\sigma}(\vec{x}_k - \vec{x}_j), \sigma)\, d\sigma \right) \cdot \exp\left(-\jmath 2\pi \vec{\nu}^{\mathrm{T}} (\vec{n}_k - \vec{n}_j)\right) \qquad (3.20) \\
&= \int \tilde{u}_{j,\sigma} \left(\sum_k (\mathcal{A}\mathcal{P}R_j)(\mathbf{B}_{j,\sigma}\mathbf{V}(\vec{n}_k - \vec{n}_j)) \exp\left(-\jmath 2\pi \vec{\nu}^{\mathrm{T}} (\vec{n}_k - \vec{n}_j)\right) \right) d\sigma \qquad (3.21) \\
&\approx \int \tilde{u}_{j,\sigma} \left(\int (\mathcal{A}\mathcal{P}R_j)(\mathbf{B}_{j,\sigma}\mathbf{V}(\vec{n} - \vec{n}_j)) \exp\left(-\jmath 2\pi \vec{\nu}^{\mathrm{T}} (\vec{n} - \vec{n}_j)\right)\, d\vec{n} \right) d\sigma \qquad (3.22) \\
&= \int \frac{\tilde{u}_{j,\sigma}}{\det(\mathbf{V}) \det(\mathbf{R}_{j,\sigma})} (\mathcal{F}^{n-1}\mathcal{A}\mathcal{P}R_j)(\mathbf{B}_{j,\sigma}^{+T}\mathbf{V}^{-\mathrm{T}}\vec{\nu})\ \delta(\hat{\theta}_{j,\sigma}^{\mathrm{T}}\mathbf{V}^{-\mathrm{T}}\vec{\nu})\, d\sigma \qquad (3.23) \\
&= \sum_{\sigma \in \mathcal{B}_j\left(\frac{\vec{\nu}}{||\vec{\nu}||}\right)} \frac{\tilde{u}_{j,\sigma}}{\det(\mathbf{V}) \det(\mathbf{R}_{j,\sigma})} (\mathcal{F}^{n-1}\mathcal{A}\mathcal{P}R_j)(\mathbf{B}_{j,\sigma}^{+T}\mathbf{V}^{-\mathrm{T}}\vec{\nu}) \left(\frac{\partial}{\partial \sigma} \hat{\theta}_{j,\sigma}^{\mathrm{T}} \mathbf{V}^{-\mathrm{T}} \vec{\nu} \right)^{-1} \qquad (3.24)
\end{aligned}
$$

3.4 Final LFR approximation

We further simplify (3.24) by approximating $\mathcal{F}^{n-1}\mathcal{A}\mathcal{P}R_j$. We use the Fourier identity for autocorrelation:

$$(\mathcal{F}^n \mathcal{A} x)(\vec{\nu}) = \left|(\mathcal{F}^n x)(\vec{\nu})\right|^2,$$

and a Taylor-expansion-based approximation for $\mathcal{F}^{n-1}\mathcal{P}\mu$ shown in (3.28) through (3.29). Using the approximation (3.29) in (3.24), we find that (3.24) simplifies to the

following LFR expression:

$$
\begin{aligned}
H_j^W(\vec{\nu}) &\approx \frac{\left|(\mathcal{F}^n R_j)(\mathbf{V}^{-\mathrm{T}}\vec{\nu})\right|^2}{\det(\mathbf{V})} \sum_{\sigma \in \mathcal{B}_j\left(\frac{\vec{\nu}}{||\vec{\nu}||}\right)} \frac{\tilde{u}_{j,\sigma}\det(\mathbf{R}_{j,\sigma})}{\left(\frac{\partial}{\partial\sigma}\hat{\theta}_{j,\sigma}^{\mathrm{T}}\mathbf{V}^{-\mathrm{T}}\vec{\nu}\right)} \\
&= \frac{\left|(\mathcal{F}^n R_j)(\mathbf{V}^{-\mathrm{T}}\vec{\nu})\right|^2}{\det(\mathbf{V})\,||\vec{\nu}||} \sum_{\sigma \in \mathcal{B}_j\left(\frac{\vec{\nu}}{||\vec{\nu}||}\right)} \frac{\tilde{u}_{j,\sigma}\det(\mathbf{R}_{j,\sigma})}{\left(\frac{\partial}{\partial\sigma}\hat{\theta}_{j,\sigma}^{\mathrm{T}}\mathbf{V}^{-\mathrm{T}}\frac{\vec{\nu}}{||\vec{\nu}||}\right)} \\
&= J(\vec{\nu}) E_j^W\left(\frac{\vec{\nu}}{||\vec{\nu}||}\right), \qquad (3.25)
\end{aligned}
$$

where we define the terms:

$$
J(\vec{\nu}) \triangleq \frac{\left|(\mathcal{F}^n R_j)(\mathbf{V}^{-\mathrm{T}}\vec{\nu})\right|^2}{\det(\mathbf{V})^2\,||\vec{\nu}||} \qquad (3.26)
$$

$$
E_j^W\left(\frac{\vec{\nu}}{||\vec{\nu}||}\right) \triangleq \sum_{\sigma \in \mathcal{B}_j\left(\frac{\vec{\nu}}{||\vec{\nu}||}\right)} \frac{\tilde{u}_{j,\sigma}\det(\mathbf{R}_{j,\sigma})\det(\mathbf{V})}{\left(\frac{\partial}{\partial\sigma}\hat{\theta}_{j,\sigma}^{\mathrm{T}}\mathbf{V}^{-\mathrm{T}}\frac{\vec{\nu}}{||\vec{\nu}||}\right)}. \qquad (3.27)
$$

$$
\begin{aligned}
\frac{(\mathcal{F}^{n-1}\mathcal{P}R_j)(\vec{u};\sigma)}{\exp(\jmath 2\pi\vec{u}^{\mathrm{T}}\vec{s}_{j,\sigma})} &= \int_{\mathbb{R}^{n-1}} (\mathcal{P}R_j)(\vec{s}_{j,\sigma}+\vec{\Delta s};\sigma)\exp(-\jmath 2\pi\vec{u}^{\mathrm{T}}\vec{\Delta s})\,d\vec{\Delta s} \qquad (3.28) \\
&\approx \int_{\mathbb{R}^{n-1}} (\mathcal{R}R_j)(\vec{x}_j+\mathbf{B}_{j,\sigma}^{+}\vec{\Delta s},\hat{\theta}_{j,\sigma})\exp\left(-\jmath 2\pi\vec{u}^{\mathrm{T}}\vec{\Delta s}\right)\,d\vec{\Delta s} \\
&= \int_{\mathbb{R}^{n-1}}\int_{\mathbb{R}} R_j(\vec{x}_j+\mathbf{B}_{j,\sigma}^{+}\vec{\Delta s}+\tau\hat{\theta}_{j,\sigma})\exp\left(-\jmath 2\pi\vec{u}^{\mathrm{T}}\vec{\Delta s}\right)\,d\tau d\vec{\Delta s} \\
&= \int_{\mathbb{R}^n} R_j(\vec{x}_j+\tilde{\mathbf{B}}_{j,\sigma}^{-1}\vec{w})\exp\left(-\jmath 2\pi\vec{u}^{\mathrm{T}}\mathbf{S}_{n-1}\vec{w}\right)\,d\vec{w} \\
&\qquad \left(\vec{w} \triangleq \begin{bmatrix} \vec{\Delta s} \\ \hline \tau \end{bmatrix}, \tilde{\mathbf{B}}_{j,\sigma} \triangleq \begin{bmatrix} \mathbf{B}_{j,\sigma} \\ \hline \hat{\theta}_{j,\sigma} \end{bmatrix}\right) \\
&= \det(\tilde{\mathbf{B}}_{j,\sigma})\exp\left(\jmath 2\pi\vec{u}^{\mathrm{T}}\mathbf{B}_{j,\sigma}\vec{x}_j\right)(\mathcal{F}^n R_j)(\mathbf{B}_{j,\sigma}^{\mathrm{T}}\vec{u}) \\
\left|(\mathcal{F}^{n-1}\mathcal{P}R_j)(\vec{u};\sigma)\right|^2 &\approx \det(\mathbf{R}_{j,\sigma})^2\left|(\mathcal{F}^n R_j)(\mathbf{B}_{j,\sigma}^{\mathrm{T}}\vec{u})\right|^2. \qquad (3.29) \\
&\qquad \left(\det(\tilde{\mathbf{B}}_{j,\sigma}) = \det(\mathbf{R}_{j,\sigma})\right)
\end{aligned}
$$

We make one last simplification of $J(\vec{\nu})$, by assuming that each R_j is equal to a basis unit voxel R_{basis} (for example, a unit cube) scaled by the voxel spacing $\mathbf{V}$ and

shifted by $\vec{x}_j$, that is, $R_j(\vec{x}) = R_{\text{basis}}(\mathbf{V}^{-1}(\vec{x} - \vec{x}_j))$. In this case, J becomes:

$$J(\vec{\nu}) = \frac{|(\mathcal{F}^n R_{\text{basis}})(\vec{\nu})|^2}{||\vec{\nu}||}.$$

It is important to note that in the factorization of (3.25), J is independent of voxel j, the weighting (via u), or even the specific CT geometry used. The E_j^W term depends on all of these factors, but depends only on the frequency via its direction, not its magnitude. This factorization is key to computational efficiency. To use (3.27), one must consider the CT geometry of interest, as shown in Section 4.3.

CHAPTER IV

Variance Prediction

In this chapter, we apply the frequency-domain approximation (3.25) to the fast evaluation of the variance prediction integral (2.60). This results in a variance prediction method that is much faster to evaluate than frequency-domain methods that rely on empirical estimates of the local frequency response of $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$. Because (3.25) is factored into (3.26) and (3.27), we accelerate the integral (2.60), which for n-dimensional CT would be an n-dimensional integral, into the $(n-1)$-dimensional form (4.3).

We explore applications of this general form to some specific CT geometries. For example, for 2DCT, this general form reduces to the form (4.7), which is equivalent to the previous work shown in [76]. We also show that for 3DCT geometries with a small cone-angle, we can make further approximations that reduce variance prediction to the one-dimensional integral (4.10).

4.1 Methods

Representing $\vec{\nu}$ in spherical coordinates, such that $\varrho \triangleq ||\vec{\nu}||$ is the frequency magnitude and $\vec{\Theta} \triangleq \vec{\nu}/||\vec{\nu}||$ is the frequency direction, we change the coordinates of (2.60) into n-dimensional spherical coordinates:

$$\mathrm{var}(\hat{x}_j) \approx \int_{\mathbb{S}^n} \int_0^{\varrho_{\max}} \frac{H_j^{\hat{W}}(\varrho, \vec{\Theta})}{(H_j^W(\varrho, \vec{\Theta}) + \alpha R(\varrho, \vec{\Theta}))^2} \, \varrho^{n-1} d\varrho d\vec{\Theta}, \tag{4.1}$$

where $\mathbb{S}^n$ represents the surface of a sphere in n-dimensions, or equivalently, the set of all $\vec{\Theta}$ that are unit vectors in $\mathbb{R}^n$. In 2D, we will parameterize $\vec{\Theta}$ with the mapping $\vec{\Theta}(\Phi) = (\cos\Phi, \sin\Phi)^{\mathrm{T}}$, which turns the integral over $\mathbb{S}^2$ into the 1D integral over Φ

from 0 to 2π. In 3D, we will parameterize $\vec{\Theta}$ with the spherical mapping

$$\vec{\Theta}(\Phi, \Theta) = \begin{pmatrix} \cos\Theta\cos\Phi \\ \cos\Theta\sin\Phi \\ \sin\Theta \end{pmatrix}; \tag{4.2}$$

the corresponding integral $\int_{\mathbb{S}^3} f(\vec{\Theta})\, d\vec{\Theta}$ becomes $\int_{-\pi/2}^{\pi/2}\int_0^{2\pi} f(\vec{\Theta}(\Phi,\Theta))\cos\Theta\ d\Phi\ d\Theta$.

We rearrange terms in (4.1) so that:

$$\mathrm{var}(\hat{x}_j) \approx \alpha^{-1} \int_{\mathbb{S}^n} \frac{E_j^{\hat{W}}(\vec{\Theta})}{E_j^{W}(\vec{\Theta})} G(\alpha^{-1} E_j^{W}(\vec{\Theta}), \vec{\Theta})\, d\vec{\Theta}, \tag{4.3}$$

where $G(\gamma, \vec{\Theta})$ is a function defined as

$$G(\gamma, \vec{\Theta}) \triangleq \int_0^{\varrho_{\max}(\vec{\Theta})} \frac{\gamma J(\varrho, \vec{\Theta})}{(\gamma J(\varrho, \vec{\Theta}) + R(\varrho, \vec{\Theta}))^2} \varrho^{n-1}\, d\varrho, \tag{4.4}$$

and where $\varrho_{\max}(\vec{\Theta})$ is the maximum extent of ϱ in $[-1/2, 1/2]^n$:

$$\varrho_{\max}(\vec{\Theta}) = \frac{1}{2||\vec{\Theta}||_\infty}.$$

In general, G cannot be computed in a closed form, but it is well-behaved and depends only on voxel shape (which determines $R_{\text{basis}}(\vec{x})$ and hence $J(\vec{\nu})$) and regularizer shape (which determines $R(\vec{\nu})$). We precompute a single table of values of G and use that table to predict variance maps via (4.3) for multiple voxels, any regularization parameter α, any weighting, any voxel spacing or scan geometry (so long as the dimensionality n does not suddenly change).

We will also consider the special case where we approximate the voxel shape to be a point impulse and the regularizer to be a purely quadratic penalty of $||\vec{\nu}||^2$. When the voxel shape is a point impulse, the data-independent frequency response term $J(\varrho, \vec{\Theta})$ loses its dependence on $\vec{\Theta}$, and is equal to $1/\varrho$. When the frequency response of the regularizer is also isotropic, such that $R(\varrho, \vec{\Theta}) = R_0\varrho^2$, G has a closed form for $n \geq 2$ (see Appendix B) of

$$G(\gamma, \vec{\Theta}) = \frac{\varrho_{\max}(\vec{\Theta})^{n+1}}{\gamma(n+1)} {}_2F_1\left(2, \frac{n+1}{3}; \frac{n+4}{3}; -\frac{R_0\varrho_{\max}(\vec{\Theta})^3}{\gamma}\right). \tag{4.5}$$

When $n = 2$, the arguments to the hypergeometric function ${}_2F_1$ simplify, and ${}_2F_1(2, 1; 2; z) = 1/(1 - z)$, leading to the simple form found for 2D fan-beam CT in [76]. For 3DCT, however, this no longer admits a "nice" expansion, and so the error inherent in the approximations of the voxel shape as a point impulse and the regularizer as isotropic are likely not worth the improvement in computational cost, which is smaller than the form for 2DCT. Surprisingly, (4.5) suggests that this integral would become easy to evaluate for 5DCT since ${}_2F_1(2, 2; 3; z)$ has a simple closed form, but this is beyond the scope of our research for now.

The integral of (4.3) has one fewer dimension than the integral of (2.60) or (4.1), and so it can be computed in significantly less time.

4.2 Application to 2DCT

In this section, we examine the results of simplifying (4.3) in the case of the common 2D fan-beam CT geometry where the X-ray source rotates circularly around the isocenter (which will be, in our coordinate system, at $\vec{x}_{\text{isocenter}} = (0, 0)^{\text{T}}$), and the X-ray detector forms an arc with its center located at the X-ray source.

In this geometry, the source position is given by

$$\vec{p}_s(\sigma) = -D_{\text{so}} \begin{pmatrix} \cos\sigma \\ \sin\sigma \end{pmatrix},$$

where D_{so} is the distance from the isocenter to the X-ray source, and σ is the angular representation of the position of the source. We will represent the position of a pixel j in question in rectangular and polar coordinates as

$$\vec{x}_j = \begin{pmatrix} x_{j,1} \\ x_{j,2} \end{pmatrix} = \begin{pmatrix} r_j \cos\phi_j \\ r_j \sin\phi_j \end{pmatrix}.$$

The ray from the source at σ to the position $\vec{x}_j$ of pixel j is then denoted

$$\vec{\ell}_{j,\sigma} = \vec{x}_j - \vec{p}_s(\sigma).$$

For a given σ the detector position $s(\vec{x}_j)$, in the geometry where the center of the X-ray detector arc is the X-ray source, is a function of voxel location $\vec{x}_j$:

$$s(\vec{x}_j; \sigma) = D_{\text{sd}} \operatorname{atan2}(\ell_{j,2,\sigma}, \ell_{j,1,\sigma}) - \sigma = D_{\text{sd}} \angle \begin{pmatrix} x_{j,1} + D_{\text{so}} \cos\sigma \\ x_{j,2} + D_{\text{so}} \sin\sigma \end{pmatrix} - \sigma.$$

The matrix $\mathbf{B}_{j,\sigma}$ is

$$\mathbf{B}_{j,\sigma} = \nabla_{\vec{x}} s = \frac{D_{\text{sd}}}{||\vec{\ell}_{j,\sigma}||^2} \begin{pmatrix} -\ell_{j,2,\sigma} \\ \ell_{j,1,\sigma} \end{pmatrix}^{\text{T}}.$$

In this 2D case, $\mathbf{B}_{j,\sigma}$ is simply a row vector that is perpendicular to the ray $\vec{\ell}_{j,\sigma}$ that points in the direction of increasing detector position, with a length representing a detector magnification factor for this source and pixel combination. This vector has the QR decomposition $\mathbf{B} = \mathbf{R}_{j,\sigma}\mathbf{Q}_{j,\sigma}$, where, in the 2D case, $\mathbf{R}_{j,\sigma}$ is the scalar length of $\mathbf{B}_{j,\sigma}$ and $\mathbf{Q}_{j,\sigma}$ is a unit row vector in the direction of $\mathbf{B}_{j,\sigma}$:

$$\begin{aligned} \mathbf{R}_{j,\sigma} &= \frac{D_{\text{sd}}}{||\vec{\ell}_{j,\sigma}||} \\ \mathbf{Q}_{j,\sigma} &= \frac{1}{||\vec{\ell}_{j,\sigma}||} \begin{pmatrix} -\ell_{j,2,\sigma} \\ \ell_{j,1,\sigma} \end{pmatrix}^{\text{T}}. \end{aligned}$$

Denote the $\vec{\theta}_{j,\sigma} = \vec{\ell}_{j,\sigma}/||\vec{\ell}_{\sigma}||$ to be a unit vector in the direction of the ray from the source at position σ to the pixel j. Let the voxel spacing be equal to $\mathbf{V} = \Delta\mathbf{I}_2$. Writing the frequency vector $\vec{\nu}$ in polar coordinates: $\vec{\nu} = (\rho\cos\Phi, \rho\sin\Phi)^{\text{T}}$, the argument to the Dirac impulse that appears in (3.23) is equal to

$$\hat{\theta}_{j,\sigma}^{\text{T}}\mathbf{V}^{-\text{T}}\frac{\vec{\nu}}{||\vec{\nu}||} = \frac{1}{\Delta||\vec{\ell}_{j,\sigma}||}\vec{\ell}_{j,\sigma}^{\text{T}}\frac{\vec{\nu}}{\rho} = \frac{1}{\Delta||\vec{\ell}_{j,\sigma}||}\left(r_j\cos(\Phi-\phi_j) + D_{\text{so}}\cos(\Phi-\sigma)\right). \quad (4.6)$$

The set of σ that set the argument in (4.6) equal to zero, as a function of frequency direction Φ and pixel j, is

$$\mathcal{B}_j(\Phi) = \Phi \pm \cos^{-1}\left(-\frac{r_j}{D_{\text{so}}}(\Phi-\phi_j)\right).$$

At $\sigma \in \mathcal{B}_j(\Phi)$,

$$\left|\frac{\partial}{\partial\sigma}\hat{\theta}_{j,\sigma}^{\text{T}}\mathbf{V}^{-\text{T}}\frac{\vec{\nu}}{||\vec{\nu}||}\right| = \frac{1}{\Delta||\vec{\ell}_{j,\sigma}||}D_{\text{so}}\sin(\Phi-\sigma) = \frac{D_{\text{so}}\sqrt{1-r_j\cos(\Phi-\phi_j)}}{\Delta||\vec{\ell}_{j,\sigma}||},$$

and so $E_j^W(\Phi)$ from the general form (3.27) simplifies to:

$$\begin{aligned} E_j^W(\Phi) &= \Delta^2 \sum_{\sigma \in \mathcal{B}_j(\Phi)} \tilde{u}_{j,\sigma} \frac{D_{\text{sd}}}{||\vec{\ell}_{j,\sigma}||} \left(\frac{D_{\text{so}} \sqrt{1 - r_j \cos(\Phi - \phi_j)}}{\Delta ||\vec{\ell}_{j,\sigma}||} \right)^{-1} \\ &= \Delta^3 \frac{D_{\text{sd}}}{D_{\text{so}}} \sum_{\sigma \in \mathcal{B}_j(\Phi)} \frac{\tilde{u}_{j,\sigma}}{\sqrt{1 - r_j \cos(\Phi - \phi_j)}}. \end{aligned}$$

If we make the approximation of $J(\vec{\nu}) = 1/\rho$ and $R(\vec{\nu}) = R_0 \rho^2$ that lead to (4.5), we can make the simplification that

$$G(\gamma, \Phi) = \frac{1}{3} \frac{\rho_{\max}(\Phi)^3}{\gamma + R_0 \rho_{\max}(\Phi)^3}.$$

If we use this form for G in the variance-prediction integral (4.3), in the case where $\hat{\mathbf{W}} = \mathbf{W}$,

$$\begin{aligned} \text{var}(\hat{x}_j) &\approx \frac{1}{3} \alpha^{-1} \int_0^{2\pi} \frac{\rho_{\max}(\Phi)^3}{\alpha^{-1} E_j^W(\Phi) + R_0 \rho_{\max}(\Phi)^3} \, d\Phi \\ &= \int_0^{2\pi} \frac{\rho_{\max}(\Phi)^3/3}{\Delta^3 \frac{D_{\text{sd}}}{D_{\text{so}}} \left(\sum_{\sigma \in \mathcal{B}_j(\Phi)} \frac{\tilde{u}_{j,\sigma}}{\sqrt{1 - r_j \cos(\Phi - \phi_j)}} \right) + \alpha R_0 \rho_{\max}(\Phi)^3} \, d\Phi. \end{aligned} \tag{4.7}$$

Except for choices of normalization and notation, (4.7) is equivalent to the single-integral variance prediction form in [76]. Additionally, we can take this to a parallel-beam, flat-detector case by taking $D_{\text{so}} \to \infty$ and $\frac{D_{\text{sd}}}{D_{\text{so}}} \to 1$. In this case, $\mathcal{B}_j(\Phi) = \Phi \pm \pi/2$, $\sqrt{1 - r_j \cos(\Phi - \phi_j)}$ converges to 1, and so

$$\text{var}(\hat{x}_j) \approx \frac{1}{3} \int_0^{2\pi} \frac{\rho_{\max}(\Phi)^3}{2\Delta^3 \tilde{u}_{j,\Phi} + \alpha R_0 \rho_{\max}(\Phi)^3} \, d\Phi.$$

The replacement of the sum over $\mathcal{B}_j(\Phi)$ relies on the fact that, for 2DCT, $\tilde{u}_{j,\sigma} = \tilde{u}_{j,\sigma+\pi}$, since these two quantities are both functions of the attenuation over the same line; one is just the attenuation on the line integrating the other direction. Since $\rho_{\max}(\Phi) = \rho_{\max}(\Phi + \pi/2)$, we can integrate over $\Phi' = \Phi + \pi/2$ to get the above form.

4.3 Application to 3D axial and helical cone-beam CT

In a third-generation GE helical CT system, the spatial source position $\vec{p}_s$ as a function of the source angle σ is given by

$$\vec{p}_s(\sigma) = \begin{pmatrix} -D_{\mathrm{so}} \cos\sigma \\ -D_{\mathrm{so}} \sin\sigma \\ p_3 \sigma \end{pmatrix},$$

where D_{so} is the source-isocenter distance, D_{sd} is the source-detector distance, and p_3 is the helical pitch, in units of length per radian; axial CT is a special case where $p_3 = 0$. We denote the position of voxel j in rectangular and cylindrical coordinates as

$$\vec{x}_j = \begin{pmatrix} x_{j,1} \\ x_{j,2} \\ x_{j,3} \end{pmatrix} = \begin{pmatrix} r_j \cos\phi_j \\ r_j \sin\phi_j \\ x_{j,3} \end{pmatrix}.$$

The ray from the source at σ to the position $\vec{x}_j$ is denoted $\vec{\ell}_{j,\sigma} = \vec{x}_j - \vec{p}_s(\sigma)$. The first coordinate of the detector position $s_1(\vec{x};\sigma)$ is given by $s_1(\vec{x}_j;\sigma) = D_{\mathrm{sd}} \operatorname{atan2}(\ell_{2,j,\sigma}, \ell_{1,j,\sigma}) - \sigma$. The second coordinate is given by

$$s_2(\vec{x}_j;\sigma) = \frac{D_{\mathrm{sd}}}{D_{2,j,\sigma}} \ell_{3,j,\sigma},$$

where $D_{2,j,\sigma} = ||\operatorname{diag}(1,1,0)\vec{\ell}_{j,\sigma}||$ is the length of the projection of $\vec{\ell}_{j,\sigma}$ into the xy-plane. From these coordinates, we find the matrix $\mathbf{B}_{j,\sigma}$ for this geometry:

$$\mathbf{B}_{j,\sigma} = D_{\mathrm{sd}} \begin{bmatrix} -\frac{\ell_{2,j,\sigma}}{D_{2,j,\sigma}^2} & \frac{\ell_{1,j,\sigma}}{D_{2,j,\sigma}^2} & 0 \\ -\frac{\ell_{1,j,\sigma}\ell_{3,j,\sigma}}{D_{2,j,\sigma}^3} & -\frac{\ell_{2,j,\sigma}\ell_{3,j,\sigma}}{D_{2,j,\sigma}^3} & \frac{1}{D_{2,j,\sigma}} \end{bmatrix},$$

which has a factorization

$$\begin{aligned} \mathbf{B}_{j,\sigma} &= \mathbf{R}_{j,\sigma}\mathbf{Q}_{j,\sigma} \\ \mathbf{R}_{j,\sigma} &= \begin{bmatrix} D_{\mathrm{sd}}/D_{2,j,\sigma} & 0 \\ 0 & D_{\mathrm{sd}}||\vec{\ell}_\sigma||/D_{2,j,\sigma}^2 \end{bmatrix} \\ \mathbf{Q}_{j,\sigma}^{\mathrm{T}} &= \begin{bmatrix} -\ell_{2,\sigma}/D_{2,j,\sigma} & -\ell_{1,\sigma}\ell_{3,\sigma}/||\vec{\ell}_\sigma||D_{2,j,\sigma} \\ \ell_{1,\sigma}/D_{2,j,\sigma} & -\ell_{2,\sigma}\ell_{3,\sigma}/||\vec{\ell}_\sigma||D_{2,j,\sigma} \\ 0 & D_{2,j,\sigma}/||\vec{\ell}_\sigma|| \end{bmatrix} \end{aligned}$$

For 3-dimensional CT, we use (Φ, Θ) for 3D spherical frequency coordinates, such that

$$\frac{\vec{\nu}}{||\vec{\nu}||} = \vec{\Theta} = \begin{pmatrix} \cos\Theta\cos\Phi \\ \cos\Theta\sin\Phi \\ \sin\Theta \end{pmatrix}.$$

If the voxel spacing is equal to the common choice of $\mathbf{V} = \mathrm{diag}(\Delta_x, \Delta_x, \Delta_z)$, the argument to the Dirac impulse in (3.23) is equal to

$$\begin{aligned}\hat{\theta}_{j,\sigma}^{\mathrm{T}}\mathbf{V}^{-\mathrm{T}}\vec{\Theta} =& \frac{1}{||\vec{\ell}_{j,\sigma}||}\left(\frac{\cos\Theta}{\Delta_x}(r\cos(\Phi-\phi)+\right. \\ & \left. D_{\mathrm{so}}\cos(\Phi-\sigma)) + \frac{\sin\Theta}{\Delta_z}(x_3 - p_3\sigma)\right). \end{aligned} \tag{4.8}$$

For typical clinical 3DCT system geometries, the majority of the noise power in LFR approximations to $S_j(\vec{\nu})$ is located in regions of spatial frequency with relatively small Θ. Since we are particularly interested in the region where Θ is small, and since $(x_3 - p_3\sigma)$ is often small, we make the following approximation to (4.8):

$$\hat{\theta}_{j,\sigma}^{\mathrm{T}}\mathbf{V}^{-\mathrm{T}}\vec{\Theta} \approx \frac{\cos\Theta}{\Delta_x||\vec{\ell}_{j,\sigma}||}(r_j\cos(\Phi-\phi_j) + D_{\mathrm{so}}\cos(\Phi-\sigma)),$$

so that $\mathcal{B}_j(\vec{\Theta}) \approx \Phi \pm \cos^{-1}(-r_j(\Phi-\phi_j)/D_{\mathrm{so}})$. This approximation is useful because (4.8) has no closed-form solution for $\mathcal{B}_j(\vec{\Theta})$. It also allows us to approximate $E_j^W(\Phi,\Theta)\cos\Theta \approx E_j^W(\Phi,0)$. This facilitates using a 1D integral instead of a 2D integral for variance prediction. Defining:

$$\begin{aligned} J_{\mathrm{cyl}}(\vec{\nu}) &\triangleq J(\vec{\nu})\sec\Theta \\ E_{j,\mathrm{cyl}}^W(\Phi) &\triangleq E^W(\Phi,0) \approx E_j^W(\Phi,\Theta)\cos\Theta, \end{aligned}$$

gives:

$$J(\vec{\nu})E_j^W(\Phi,\Theta) \approx J_{\mathrm{cyl}}(\vec{\nu})E_{j,\mathrm{cyl}}^W(\Phi). \tag{4.9}$$

With this geometry, the expression $E_{j,\mathrm{cyl}}^W(\Phi)$ simplifies to

$$E_{j,\mathrm{cyl}}^W(\Phi) \approx \frac{D_{\mathrm{sd}}^2\Delta_x^3\Delta_z}{D_{\mathrm{so}}}\sum_{\sigma\in\mathcal{B}_j(\Phi)}\tilde{u}_{j,\sigma}\frac{||\vec{\ell}_{j,\sigma}||^2}{D_2^3}|\csc(\Phi-\sigma)|.$$

This approximation is reasonably accurate until $|\Theta|$ approaches $\pi/2$ minus the largest cone angle used in the CT system. For the purposes of variance prediction, the

inaccuracy of the LFR in large-Θ regions has a negligible effect, since for small cone angle systems, this inaccurate region is a very small fraction of the entire frequency space that is integrated in (4.3).

Figure 4.1 shows estimates of local projection-weight-backprojection frequency responses H_j^W and the corresponding approximate noise power spectra S_j. Three cases are shown: the actual local frequency response $\mathcal{F}_j\left\{[\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}]_{\cdot j}\right\}$, the spherically separable approximation (3.25), and the cylindrically separable approximation (4.9). Comparing the first row of this figure to the second row, the approximate local frequency response in (3.25) closely matches the overall appearance of the DFT-based frequency response except for large-Θ regions. These are of less importance to approximating the noise power spectrum, as can be seen in the figures comparing them. Furthermore, comparing the third row, the result of (4.9) also matches the DFT LFR except for large Θ, but again this error has a diminished effect on the NPS.

Using approximation (4.9), we rearrange the variance prediction integral (2.60) by changing from Cartesian coordinates to cylindrical coordinates (ρ, Φ, ν_3), to be

$$\mathrm{var}(\hat{x}_j) \approx \alpha^{-1} \int_0^{2\pi} \frac{E_{j,\mathrm{cyl}}^{\hat{W}}(\Phi)}{E_{j,\mathrm{cyl}}^{W}(\Phi)} G_{\mathrm{cyl}}(\Phi, \alpha^{-1} E_{j,\mathrm{cyl}}^{W}(\Phi))\, d\Phi, \tag{4.10}$$

where we define another object-independent function G_{cyl}:

$$G_{\mathrm{cyl}}(\Phi, \gamma) \triangleq \int_0^{\rho_{\max}(\Phi)} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\gamma J_{\mathrm{cyl}}(\vec{\nu})}{(\gamma J_{\mathrm{cyl}}(\vec{\nu}) + R(\vec{\nu}))^2}\, \rho d\nu_3 d\rho. \tag{4.11}$$

In this case, $\rho_{\max} = 1/(2\max\{|\cos\Phi|, |\sin\Phi|\})$. Again, G_{cyl} has no closed form but is a well-behaved function of only two parameters that we precompute and tabulate. We compute this table only once for a given differencing matrix $\mathbf{C}$ and voxel shape; a given image, weighting, system geometry, or regularization parameter α does not change the table G_{cyl}. Using the table, variance prediction via (4.10) simply requires looking up values of G_{cyl} and numerically integrating them in 1D. This integration can be evaluated using a coarse discretization of Φ with reasonably accurate predicted variance, especially given that the integrand is periodic and integrated over one period, a case in which numerical integration converges quickly [68]. While the method of derivation is changed, this is the form for fast variance prediction given in [52, 51], which also reduces to the form given in [50] for quadratic regularization and an axial geometry.

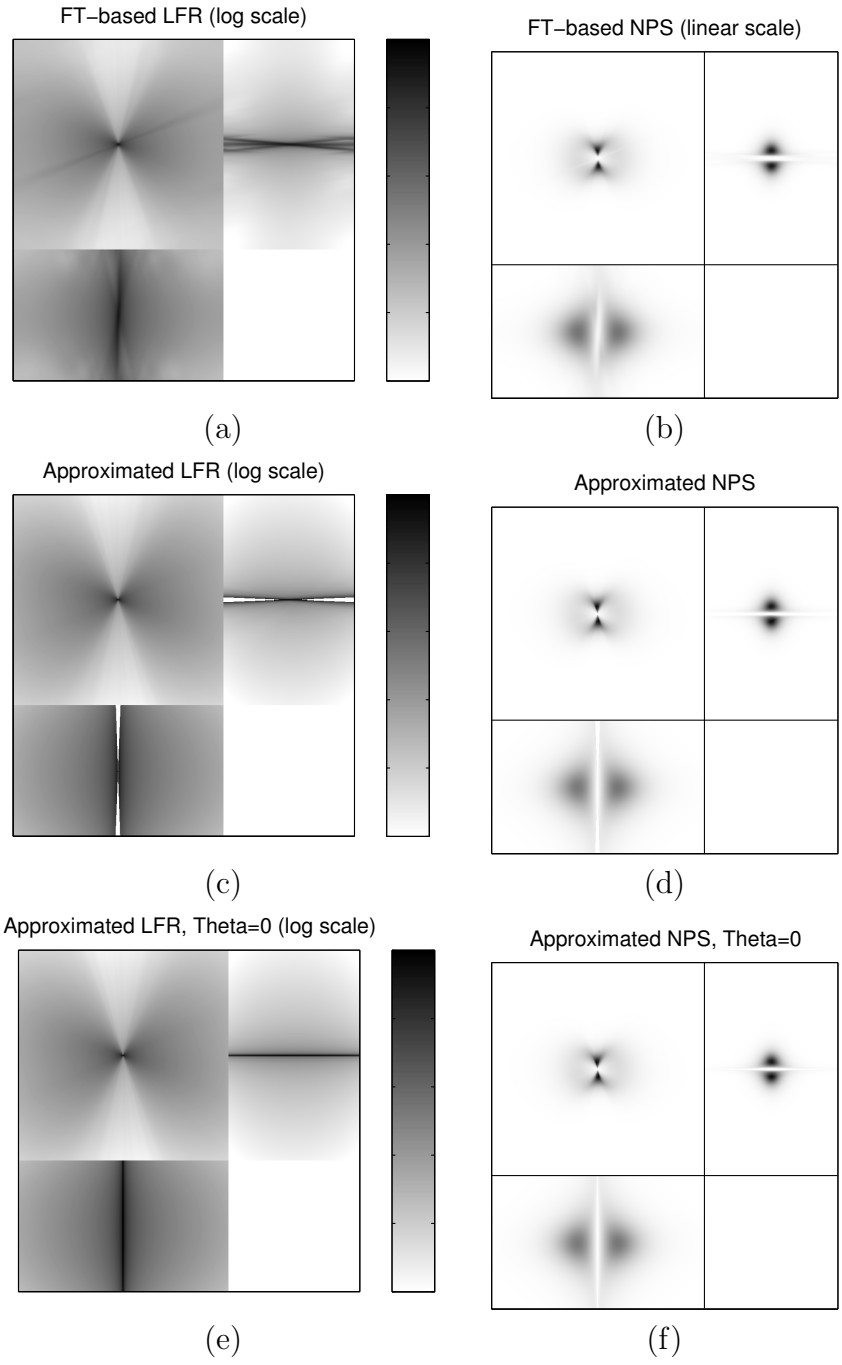


Figure 4.1: Planes cut through local frequency responses (left) and corresponding noise power spectra (right)

In each image, the upper-left portion is the (ν_1, ν_2) plane where $\nu_3 = 0$; top-right is $(\nu_3, \nu_2), \nu_1 = 0$, and bottom-left is $(\nu_1, \nu_3), \nu_2 = 0$.

First row (a,b): from (2.54), the Fourier transform of the directly computed LIR

Middle row (c,d): from (3.25), a spherically separable approximation

Bottom row (e,f): from (4.9), a cylindrically separable approximation

Note: In figures (b,d,f), the bottom section ($\nu_2 = 0$) is scaled by a factor of 4 for visibility.

For 3DCT geometries where (4.9) is an inaccurate approximation, such as those where a voxel under consideration has rays passing through it in directions that cover much of $\mathbb{S}^3$, one must revert to (4.3) for fast variance prediction. Since the angles of the rays that pass through a voxel, though, have a 1-dimensional function (the source position) underlying them, we suspect that there is a strategy for generalizing the simplification of (4.3) to (4.10) into a 1-dimensional integral for any source trajectory. How this general strategy might work is an area of future work, discussed in section 6.2.6.

4.3.1 Spatially varying regularization

In this section we investigate the effect on reconstruction variance of using a spatially varying regularizer such as that defined in [21] designed to produce a reconstruction with uniform spatial resolution. For first-order differencing matrices $\mathbf{C}_d$, (2.27) can be equivalently written as

$$R(\mathbf{x}) = \sum_k \sum_{\ell \in \mathcal{N}_k} r_{k\ell} \psi(x_\ell - x_k), \tag{4.12}$$

where $\mathcal{N}_k$ is the set of voxels that neighbor k. The term $r_{k\ell}$ replaces r_d in (2.27), and $r_{k\ell} = r_d$ when the spatial relationship between voxels k and ℓ is the one represented by the differencing direction d.

In [21], each voxel j has an associated factor κ_j representing the 'certainty' of the voxel that multiplies the effect of α, giving a spatially varying regularizer of the form:

$$R_{\text{sv}}(\mathbf{x}) = \sum_k \sum_{\ell \in \mathcal{N}_k} r_{k\ell} \kappa_k \kappa_\ell \psi(x_\ell - x_k), \tag{4.13}$$

This modulates the blurring effect of the regularizer in otherwise less certain regions to ensure uniform resolution at the cost of increasing the voxel variance. To adapt our variance prediction method to this situation, we approximate $\kappa_k \approx \kappa_j$ and $\kappa_\ell \approx \kappa_j$ for voxels k and ℓ near j. Usually the κ factors are slowly varying as a function of position, so this approximation is acceptable. Therefore, near a voxel j, we approximate (4.13)

as

$$\begin{aligned} R_{\text{sv}}(\mathbf{x}) &= \sum_k \sum_{\ell \in \mathcal{N}_k} r_{k\ell} \kappa_j^2 \psi(x_\ell - x_k) \\ &\approx \sum_k \sum_{\ell \in \mathcal{N}_k} r_{k\ell} \kappa_j^2 \psi(x_\ell - x_k) \\ &= \kappa_j^2 R(\mathbf{x}). \end{aligned} \tag{4.14}$$

For the purposes of variance prediction near the voxel j, we define a per-voxel effective regularization parameter $\alpha_{\text{eff}} \triangleq \alpha \kappa_j^2$ and evaluate (4.10) with this α_{eff} and the spatially invariant regularizer $R(\mathbf{x})$ rather than R_{sv}.

For our results, we use κ factors given, for a statistical weighting matrix $\mathbf{W}$, by:

$$\kappa_j = \frac{\sum_i [\mathbf{A}]_{ij} [\mathbf{W}]_{ii}}{\sum_i [\mathbf{A}]_{ij}} \tag{4.15}$$

which is similar to the formula given in [21]; this form is commonly used since it is easier to compute.

The effect of using κ_j^2 in the regularizer is intuitive: assuming that the change in the argument to F is small when α_{eff} is varied compared to the change in the α^{-1} multiplying the integral, the approximate variance decreases inversely with increasing certainty κ_j^2.

4.3.2 Object support masking

Outside the support of the object there is significant approximation error because our method ignores the non-negativity constraint that is often used in solving the reconstruction problem (2.18). The empirical variance outside the object approaches zero, and so the relative error of our method explodes. We use a method similar to [25] to identify regions that are outside the support of the object and predict that the variance in these regions is zero.

4.4 Results

To evaluate our fast variance prediction approach (4.10), we compared it to an empirical variance map in two cases. In one case, the empirical variance was given by reconstructions from multiple realizations of simulations of noisy projection data of an XCAT phantom. In the other case, we repeatedly scanned a physical phantom

and produced the empirical variance from the reconstructions of these scans.

4.4.1 Simulation data

For the simulation study, we reconstructed a $512 \times 512 \times 320$ voxel section of the XCAT phantom [54] with voxel size $\Delta_x \times \Delta_z = 0.9764 \times 0.625$mm that covers an anatomical section between the neck and mid-lungs. We simulated a GE third-generation system geometry with a 888×64 quarter-offset detector with detector element size 1.0239×1.0964mm; the detector went through three turns with a pitch of 1, taking 2952 views. Each reconstruction used 80 iterations of an ordered-subset method [14] using 64 subsets. In the regularization used, $\mathbf{C}$ was a matrix that takes 3 first differences for each voxel, one each for the next voxel in each axis. These differences were penalized by a Huber potential function:

$$\psi(x) = \begin{cases} x^2/2, & |x| \leq \delta \\ \delta|x| - \delta^2/2, & |x| > \delta, \end{cases} \tag{4.16}$$

which satisfies our criteria for cost functions. The value of δ was 10 Hounsfield units. We looked at two separate cases for regularization, one with the spatially varying regularization described in section 4.3.1, and one without. The regularization parameter α was equal to 2^6 in the non-spatially-varying case, and 2^{14} in the spatially varying case. The weighting $\mathbf{W}$ was normalized so that unattenuated rays had a weight of 1. Noise was applied to each of the 93 simulated projections by realizing a Poisson random variable with a mean equal to the expected number of photons with an photon count incident on the phantom of 10^5 photons per view.

Figure 4.5 shows sample reconstructions of these simulated sinograms. Figure 4.5(a) is a reconstruction with spatially varying regularization and Figure 4.5(b) is a reconstruction with uniform, spatially invariant regularization.

Figures 4.2(a) (with spatially varying regularization) and 4.3(a) (uniform regularization) show axial, sagittal, and coronal slices of the 3D map of the empirical standard deviation from our simulated reconstructions. Since the empirical standard deviation maps were noisy and the ground truth standard deviation is spatially slowly varying, we blurred the empirical variance maps with a gaussian kernel with a FWHM of 3 voxels each in each direction. Figures 4.2(b) and 4.3(b) show the corresponding slices through the 3D predicted standard deviation map from (4.10). Since standard deviation varies slowly, we computed it once per $4 \times 4 \times 4$ block and used nearest-neighbor interpolation to fill in the rest. More sophisticated interpolation could be

used, but the interpolation error is minimal compared to the intrinsic approximation error of our method. Figures 4.2(c) and 4.3(c) show the absolute magnitude of the error of our approximated standard deviation compared to the empirical results. The gray scale in these figures is transformed to better show the dynamic range of the error. Figures 4.2(d) and 4.3(d) show the empirical and predicted standard deviation along a one-dimensional axial profile through the image, behind the center of the spine in the phantom, along with the standard deviation as computed directly from (2.60) using (2.54) as the LFR Sections near the end of the axial FOV have been omitted in all images; the empirical variance becomes extreme due to a suboptimal OS algorithm implementation that is somewhat unstable in regions where the helical sampling is poor. The OS algorithm in [33] would reduce this instability and reduce the empirical variance in the end slices.

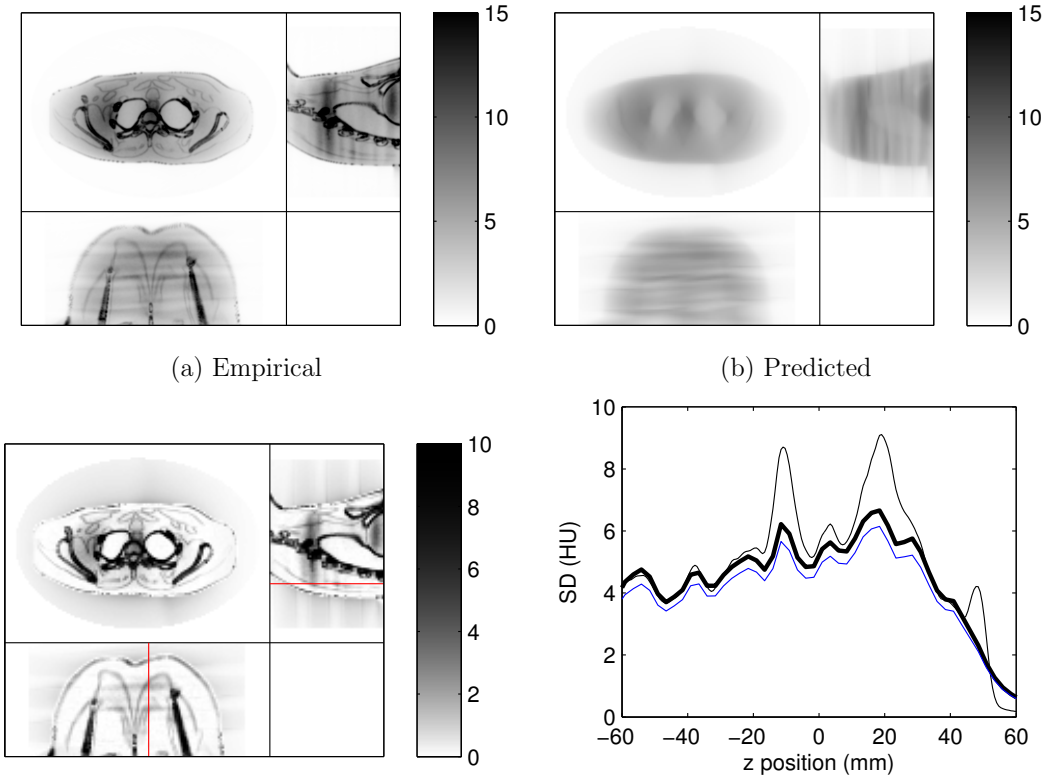
The computation time of our method for the entire volume using $4 \times 4 \times 4$ downsampling was 1207 CPU-seconds using one core of an Intel Core i7-860 with 16 GB of memory. The empirical reconstructions took an average of 1.71 days each using one core of an Intel X5650 processor also with 16GB of memory[1].

4.4.2 Real sinogram data

For our real-world dataset, we scanned a model lung phantom with spherical nodules 10 times with a GE Discovery CT750 HD scanner and reconstructed each of the 10 sinograms separately to produce an empirical variance map of the reconstruction. The geometry of the system is the same as the simulated geometry used in the previous section, with the exception of performing an axial scan using a 16-row detector and 984 views. Since we could not ensure that each scan began at the same starting angle, using multiple realizations of the same helical scan to produce an empirical variance was not possible with our physical CT scanner. With the axial scans, we used a projection matrix $\mathbf{A}$ that was correctly aligned to the starting angle of each scan so that each reconstruction was aligned to the same voxel grid. We used two different tube currents (40mA, 200mA) for a low-dose and a high-dose scan, and in all cases the tube voltage was 120 kVp and the scan time was 0.5 seconds.

We reconstructed each of the 10 sinograms using statistical reconstruction methods. The size of the reconstruction was $512 \times 512 \times 32$ voxels with voxel size $\Delta_x \times \Delta_z = 0.9764 \times 0.625$mm, as in the simulated phantom reconstructions. Each reconstruction used 100 iterations of an ordered-subset method [14] using 64 subsets.

[1] In a clinical setting, a reconstruction would use multiple processor cores, requiring much less than 1.71 days in real time.



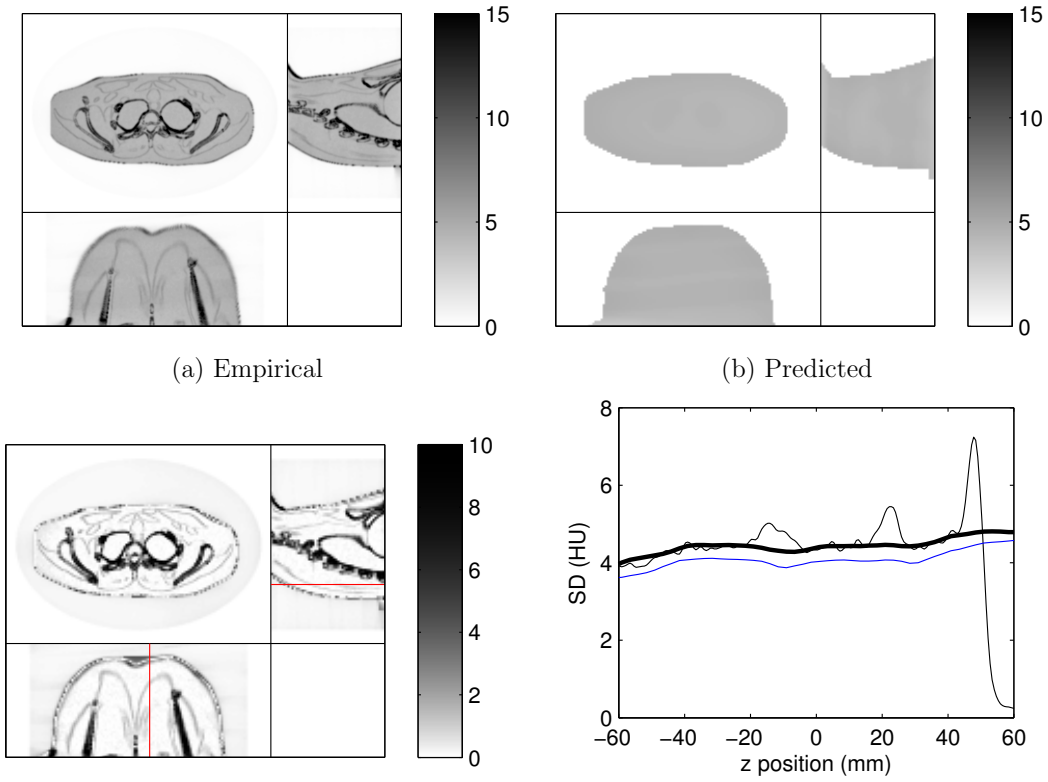
(a) Empirical

(b) Predicted

(c) Absolute Error

(d) z-profile

Figure 4.2: Three slices of standard deviation maps for simulated reconstructions using spatially varying regularization (Hounsfield units); the red line in (c) indicates the profile used in (d). In (a), (b), and (c), for representing a 3D volume, the top-left segment is an axial slice of the volume; the top-right segment is a sagittal slice; the bottom-left is a coronal slice. This convention will be used in all following images of a 3D volume. In (d), the thin black line is empirical standard deviation, the thick black line is predicted by our methods, and the blue line is the DFT-based prediction.



(a) Empirical

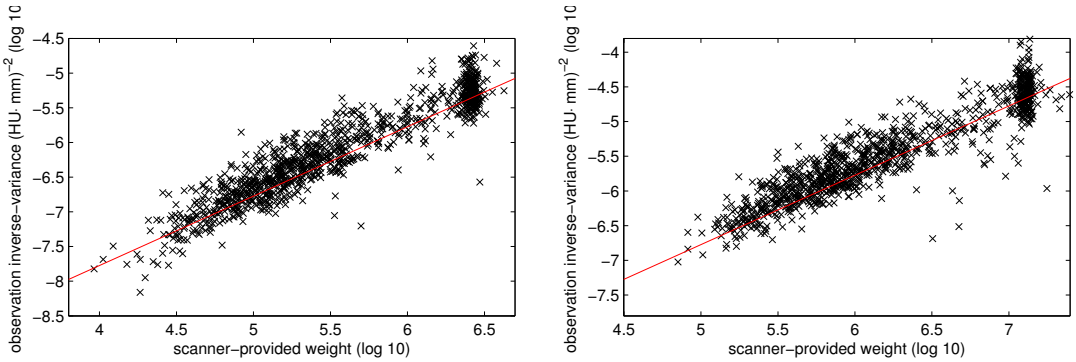
(b) Predicted

(c) Absolute Error

(d) z-profile

Figure 4.3: Three slices of standard deviation maps for simulated reconstructions using uniform regularization (Hounsfield units); the red line in (c) indicates the profile used in (d). In (d), the thin black line is empirical standard deviation, the thick black line is predicted by our methods, and the blue line is the DFT-based prediction.

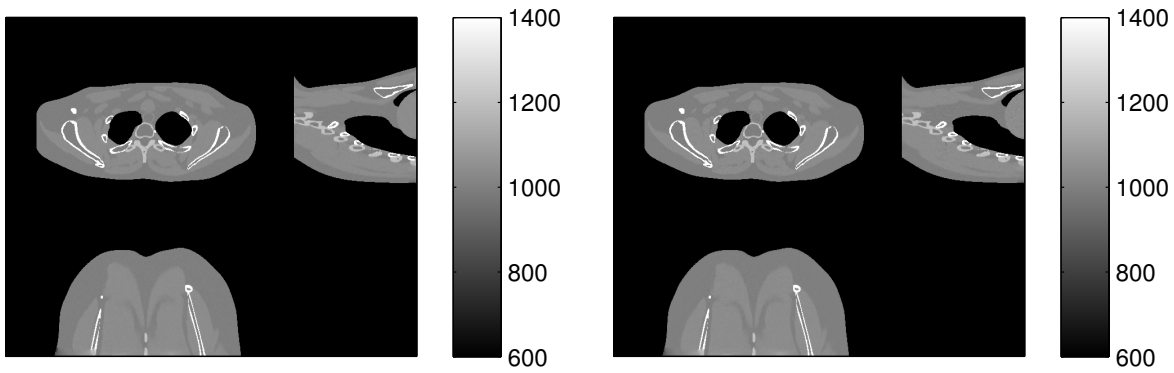
We performed the reconstructions using two different choices of regularization. In the first case, the regularization used a quadratic penalty and was spatially varying using the method of Section 4.3.1. In the second case, the penalty function used the Huber potential (4.16) with a threshold δ of 10 Hounsfield units, as used in [61], and was not spatially varying. In the quadratic-penalty case, the regularization parameter α was equal to 2^{-14}; in the Huber-penalty case, $\alpha = 2^{28}$. In both cases, the elements of the weighting matrix $\mathbf{W}$ were chosen to correspond to the CT scanner's estimate of the inverse of the variance of each ray given the scanner-specific corrections used [6]. Given that we have several repeated scans of the same object, we could find the empirical variance of the observations $\mathbf{y}$ from this data. Using this empirical observation variance for the purposes of evaluating variance prediction would be unrealistic, since in a clinical setting we do not have this data. When we require elements of the matrix $\hat{\mathbf{W}}$ for variance prediction, we estimate the observation variance from the inverse of the scanner-provided weight. Figure 4.4 shows, for a random subsample of observations, a plot of the scanner-provided weighting element versus the inverse of the empirical variance of that observation, along with the linear approximation we use to map scanner weights into predicted observation variance. Figure 4.6 shows sample reconstructions from one of these sinograms. Figure 4.6(a) is a reconstruction with quadratically penalized spatially varying regularization and Figure 4.6(b) is a reconstruction with Huber-penalized spatially invariant regularization.



(a) (b)

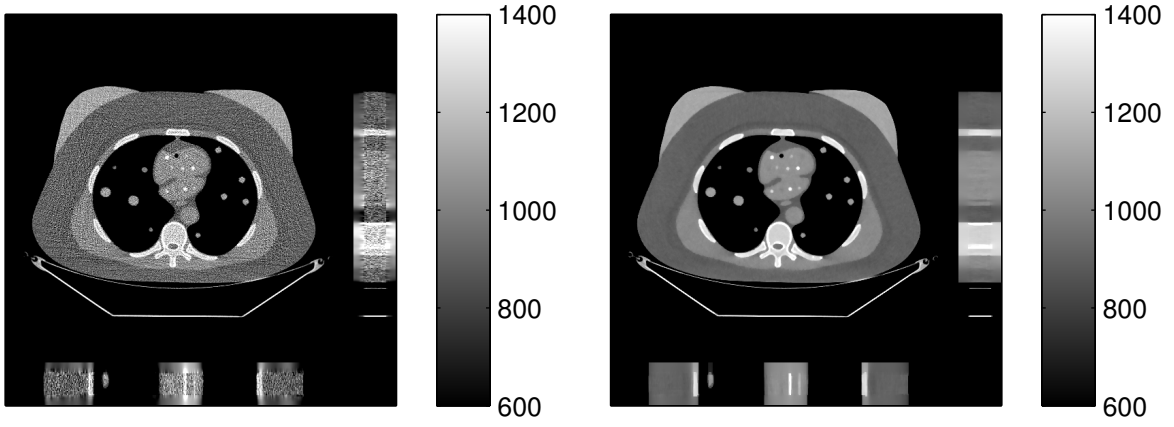
Figure 4.4: Scatterplot of scanner empirical observation inverse variance (black) with estimate for observation variance from scanner weight (red) used in variance prediction. A low dose scan is shown in (a); a high dose scan is shown in (b). Both variance estimations use the same scale factor.

Figures 4.7(a) (with spatially varying, quadratic regularization) and 4.8(a) (uniform, Huber-penalized regularization) show axial, sagittal, and coronal slices of the 3D map of the empirical standard deviation from our simulated reconstructions. As



(a) (b)

Figure 4.5: Three slices of reconstructions of simulated phantom data; with spatially varying regularization in (a), uniform regularization in (b). Scale in Hounsfield units, shifted so that 0HU represents no attenuation.



(a) (b)

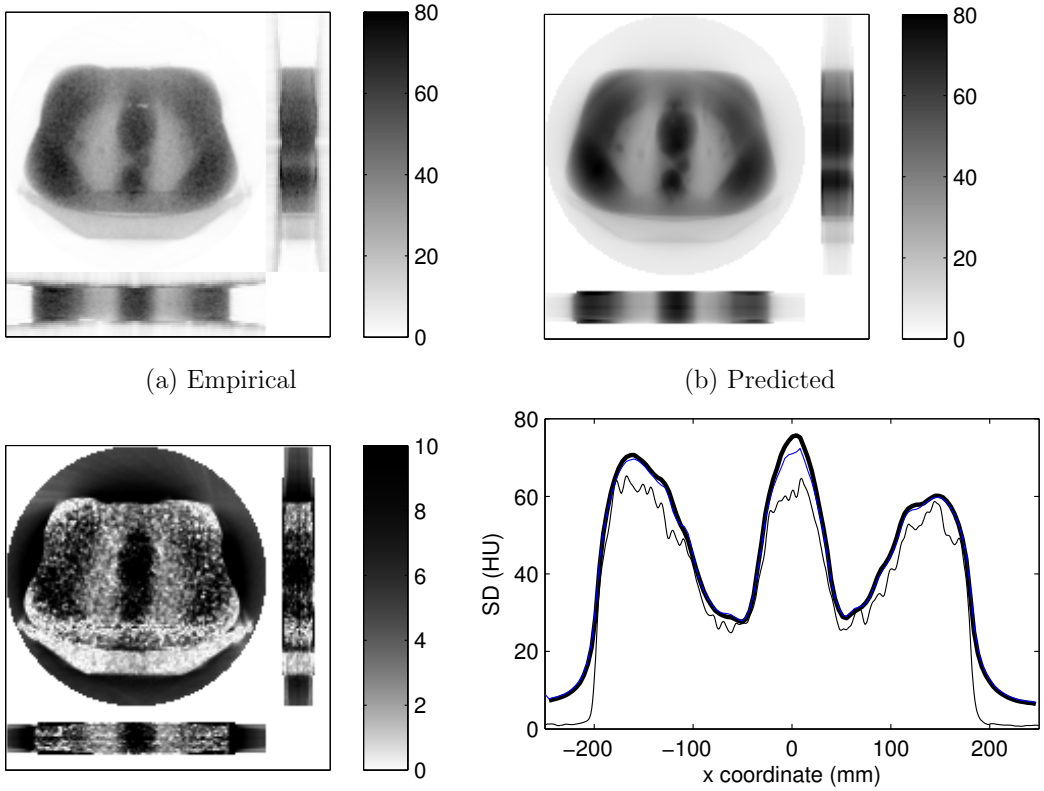
Figure 4.6: Three slices of reconstructions of real phantom data; with spatially varying, quadratic regularization in (a), uniform, Huber-penalized regularization in (b). Scale in Hounsfield units, shifted so that 0HU represents no attenuation. Coronal and sagittal slices were stretched in the trans-axial direction by a factor of two for visualization.

in the simulated empirical standard deviation maps, the empirical maps were noisy, so we blurred the empirical variance maps with a 2D gaussian kernel with a FWHM of 5 voxels each in each direction. Figures 4.7(b) and 4.8(b) show the corresponding slices through the 3D predicted standard deviation map from (4.10). We computed the standard deviation once per $4 \times 4 \times 1$ block and used nearest-neighbor interpolation to fill in the rest. Figures 4.7(c) and 4.8(c) show the absolute magnitude of the error of our approximated standard deviation compared to the empirical results. Figures 4.7(d) and 4.8(d) show the empirical and predicted standard deviation along a one-dimensional coronal profile through the center of the image, along with the standard deviation as computed directly from (2.60) using (2.54) as the LFR.

4.5 Discussion

The presented methods are able to predict the standard deviation of most voxels in the simulated reconstructed images within an error of one Hounsfield unit in both the spatially varying regularization case (about 85% within 1HU) and the uniform regularization case (about 95% within 1HU) in less time than empirical measurement by a factor of over 20000. Figure (4.9) shows, for the entire CT volume, the percentage of the image that had an error within a specified bound in both the spatially varying and uniform regularization cases. Whether the tradeoff for time at the expense of accuracy provided by our method is acceptable depends on the application. For the purposes of regularization design or tube current modulation design, which we will discuss in Chapter V, these methods will be useful. The methods would not necessarily be as useful for applying confidence intervals to a reconstruction.

Figure 4.7 shows very good qualitative agreement between empirical and predicted standard deviation, even away from the plane of X-ray source rotation, until the furthest end slices of the reconstruction, which have insufficient data coverage and would not be presented in a clinical reconstruction. Inside the phantom, the region with the highest error is the center-most section. Curiously, the DFT-based method has trouble with this region as well, although not as much as our prediction. The reason for this discrepancy is unknown. Figure 4.8 shows good qualitative agreement between the empirical and predicted standard deviation, except near edges. Except near the edges, the standard deviation is nearly constant in both the empirical and predicted reconstructions. The predicted standard deviation has a slight variation in the profile that we can see that seems like it follows a ground-truth slight variation that exists away from edges. The prediction slightly overestimates variance in 4.7



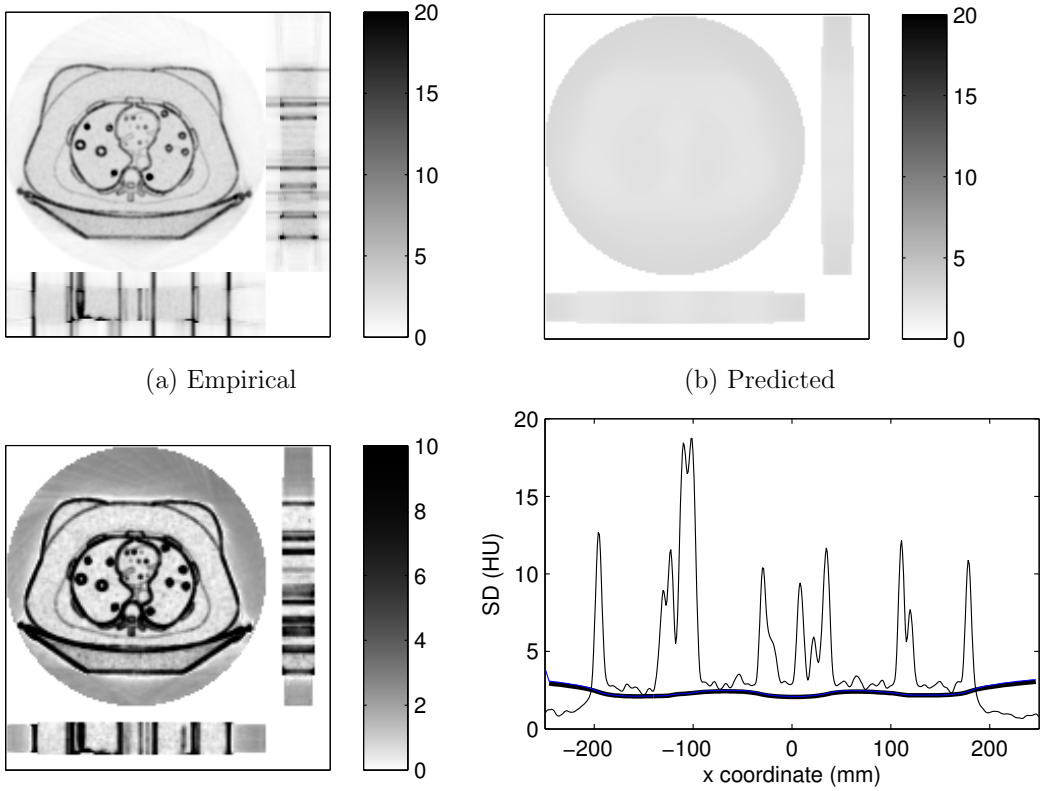
(a) Empirical

(b) Predicted

(c) Absolute Error

(d) Coronal Profile

Figure 4.7: Three slices of standard deviation maps for real reconstructions using spatially varying, quadratic regularization (Hounsfield units). Coronal and sagittal slices were stretched in the trans-axial direction by a factor of two for visualization. In (d), the thin black line is empirical standard deviation, the thick black line is predicted by our methods, and the blue line is the DFT-based prediction.



(a) Empirical

(b) Predicted

(c) Absolute Error

(d) Coronal Profile

Figure 4.8: Three slices of standard deviation maps for real reconstructions using spatially uniform, Huber-penalized regularization (Hounsfield units). Coronal and sagittal slices were stretched in the trans-axial direction by a factor of two for visualization. In (d), the thin black line is empirical standard deviation, the thick black line is predicted by our methods, and the blue line is the DFT-based prediction.

and underestimates 4.8; we are not sure what causes this discrepancy. It is possible that this is a function of the regularization parameter α (which is much higher in the Huber case, to compensate for the edge-preserving effect), and that for an α somewhere between these two cases the bias crosses zero.

The comparisons in Figures 4.2(d), 4.3(d), 4.7(d), and 4.8(d) demonstrate that the majority of the error incurred by our methods occurs in the assumptions of quadratic-like regularization and local shift-invariance that ultimately lead to (2.60), rather than our approximations that transform (2.60) into the more computationally tractable (4.10). There is almost no difference at all between the fast prediction and DFT-based prediction in the profile in Figure 4.8(d). This is reasonable considering that most of the error in the approximation (3.25) is in regions of high frequency magnitude $||\vec{\nu}||$, but the regularization frequency response $R(\vec{\nu})$ is still accurate. In the case displayed by Figure 4.8(d), since the regularization parameter is so large, the noise power spectrum $S_j(\vec{\nu})$ tends toward zero in large-$||\vec{\nu}||$ regions of the noise power spectrum.

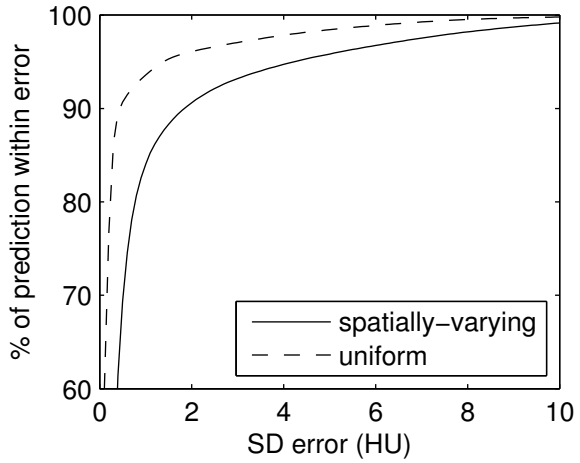


Figure 4.9: Cumulative distribution of error of predicted standard deviation vs. empirical standard deviation for simulated reconstructions.

Since our approximation of the regularization requires the penalty function to be in a quadratic region, it is not surprising that the main locations of error within the support of the object are near edges, where the quadratic-regularization assumption is invalid. This effect of edges on noise properties has also been seen for 2D fan-beam CT in [34], which postulates that the source of the high variance near edges is the uncertainty of edge position. It may be the case that accurate variance prediction near edges relies on *a priori* knowledge of edge location.

CHAPTER V

Tube Current Modulation for Statistical Reconstruction

5.1 Background

Reducing X-ray dose is an ongoing concern in CT. Iterative reconstruction techniques allow for reduced tube current, and hence X-ray dose, by producing higher-quality images from an equivalent dose to FBP-based reconstructions [61, 67, 10].

Other work has explored schemes for dose reduction via automatic tube current modulation (ATCM) [30]. Tube current modulation changes the current of an X-ray tube, and hence dose, in the middle of a scan to optimize noise properties. This technique is used in commercially-available CT scanners, such as GE's 'Smart mA'. This method uses the attenuation data from two initial radiographs to choose a sinusoid-like modulation of the tube current around a patient. The modulation is chosen such that it is constrained to a tube current range and produces a given reconstruction variance in a water phantom. ATCM methods have been shown to reduce dose when an image is reconstructed with FBP-like methods [59, 29, 24] as well as for statistically reconstructed images. In Section 5.1, we describe presently-used ATCM methods that are specific to FBP reconstructions.

Previous ATCM methods have focused on optimizing tube current under the assumption that the resulting sinogram will be reconstructed using FBP-like algorithms. In this section, we examine methods of improving ATCM that are specific to statistically reconstructed images. Tube current modulation for statistically reconstructed CT images is a more difficult problem than for FBP-reconstructed images—for FBP, modulation of the tube current affects the image quality through the noise variance, and optimization can be done by simply minimizing this noise. For statistical images, on the other hand, tube current modulation affects both the variance and resolution

of the resulting image, and the tradeoff must be considered when optimizing tube current. Because of the complexities of this tradeoff, we will consider multiple methods for "optimizing" TCM for different metrics. Each of these optimizations produces different results, and so it is important to consider what metric is appropriate for the task at hand.

In Section 5.1.2, we investigate tube current modulation methods for statistical reconstruction with the simple goal of minimizing image variance, as is done for TCM methods for FBP-reconstructed images. We demonstrate why, unlike for FBP images, this goal does not work for statistically reconstructed images. In Section 5.2, we derive fast prediction methods for the resolution of a reconstruction method and apply this to tube current modulation methods that jointly penalize variance and reward increased resolution. In Section 5.3, we examine several potential model observers and derive fast methods for predicting the performance of these observers. In Section 5.4, we take the predicted observer performance from Section 5.3 and apply this to tube current modulation schemes that optimize for predicted observer performance in statistically-reconstructed images. Tseng [62] has used model observer performance to study the dose reduction potential for statistically-reconstructed images. Finally, in Section 5.5, we find ATCM methods for the hypothetical case of statistical reconstruction where the local impulse response of the reconstruction is fixed by the regularizer.

5.1.1 Optimal tube current modulation for FBP

Gies [22] derives an analytic expression for the dose as a function of source angle that is optimized by minimizing the noise variance in the center voxel (the center voxel we will denote x_{center}) of the backprojection of a sinogram while keeping the total X-ray dose constant. They consider tube current modulation where the X-ray source makes one full rotation around the object, and the tube current on the two opposite sides of the object are equal. For our methods, we will assume an axial scan where the X-ray source makes one full rotation, and enumerate the views as $\{1, 2 \ldots, V\}$. We will denote the number of incident photons at view v as N_v and the attenuation of the ray in view v that goes through the center voxel as A_v, so that the expected number of photons on the center of the detector at view v is proportional to N_v/A_v. In [22], the variance of the noise of an FBP-based reconstruction at the center voxel is estimated to be

$$\sigma^2_{\text{center}} \propto \sum_{v=1}^{V} \frac{A_v}{N_v}. \tag{5.1}$$

The solution to the variance minimization problem

$$\mathbf{N}^* = \text{argmin}_{\mathbf{N}} \sum_{v=1}^{V} \frac{A_v}{N_v}, \quad \text{s.t.} \sum_{v=1}^{V} N_v = N_0, \tag{5.2}$$

is given by an incident photon count that is proportional to the square root of the attenuation:

$$N_v^* \propto A_v^{1/2},$$

assuming that there are no other restrictions on $\mathbf{N}$, that is, that tube current can be chosen freely. Since variance is inversely proportional to dose, the total dose can then be reduced while maintaining the same noise variance that would result from an unmodulated tube current.

To our knowledge, the problem of combining iterative reconstruction with ATCM is largely unexplored. We present an analysis of methods similar to [22] adapted to iterative reconstruction that uses frequency-domain techniques similar to those of Section IV to make the problem of ATCM tractable.

5.1.2 Variance-reduction methods

Optimizing tube current modulation to minimize the variance at the center voxel of an axial CT image, as in [22], does not work for iteratively-reconstructed images. For an axial 3DCT image with a full source rotation iteratively reconstructed via the minimization of (2.17), the variance at the center voxel of an image, as given by (4.10), simplifies to

$$\text{var}(\hat{x}_{\text{center}}) \approx \alpha^{-1} \int_0^{2\pi} G_{\text{cyl}} \left(\Phi, \frac{K w_\Phi}{\alpha} \right) \, d\Phi, \tag{5.3}$$

where w_Φ is the statistical weighting for the rays passing through the center at angles $\Phi \pm \pi/2$, and K is a constant equal to:

$$K = \frac{2|\mathbf{V}|\Delta_x}{\Delta_{\vec{s}}\Delta_\sigma} \frac{D_{\text{sd}}^2}{D_{\text{so}}^2}. \tag{5.4}$$

Denote the incident photon density in view v as N_v and the attenuation of the ray that goes through the center voxel of the object A_v, so that the expected number of photons on the center of the detector at view v is proportional to N_v/A_v. In the Gaussian statistical model, the variance of this observation is A_v/N_v, and so the ideal

weighting $w_v = N_v/A_v$. With this ideal weighting, we can estimate the variance of the center voxel as

$$\mathrm{var}(\hat{x}_{\mathrm{center}}) \approx \alpha^{-1}\frac{2\pi}{V}\sum_{v=1}^{V} G_{\mathrm{cyl}}\left(\phi_v, \frac{KN_v}{\alpha A_v}\right) \triangleq \mathrm{VP}(\mathbf{N}) \tag{5.5}$$

where ϕ_v is the source angle of view v. For a fixed object, and hence a fixed A_v, we denote this variance prediction as a function of incident photon intensities N_v as $\mathrm{VP}(\mathbf{N})$. The problem of minimizing dose while staying within a given variance bound σ^2 can then be expressed simply as

$$\mathbf{N}^* = \mathrm{argmin}_{\mathbf{N}} \sum_i N_i \quad \text{s.t.}\, \mathrm{VP}(\mathbf{N}) \leq \sigma^2. \tag{5.6}$$

However, there is a problem that exists with (5.6) for statistically-reconstructed images, but not for FBP-reconstructed images; the solution is $N_v^* = 0$! With no radiation dose, the regularizer is forced to take over and reconstruct the entire image. This image will be uninformative, but will indeed be noise-free. Therefore, in the case of statistically reconstructed images, we must consider methods more advanced than simply minimizing variance.

5.2 Fast resolution prediction

To resolve this issue, one potential method is introducing a resolution penalty that should greatly penalize reconstructions with a very low resolution, such as the no-radiation, no-information image that results from (5.6). In this section, we will explore potential resolution penalties and their effect on designing a tube current modulation scheme. Similar to the computational issues with variance prediction, many resolution metrics of interest involve computation times that would be infeasible in a clinical setting. Consider a CT system that performs a scout scan to determine approximate attenuations through the patient, then computes a resulting tube current modulation profile from the scout scan, then performs the main scan; a long computation time in the middle step would be impractical. For this reason, we develop frequency-domain methods for the prediction of resolution metrics that we consider that resemble the methods we used previously for variance prediction.

5.2.1 Dispersion-based resolution measurement

5.2.1.1 Methods

The dispersion Δf of a function $f : \mathbb{R}^n \to \mathbb{R}$ is:

$$\Delta f \triangleq \frac{\inf_{\vec{x}_0} \int_{\mathbb{R}^n} ||\vec{x} - \vec{x}_0||^2 \, |f(\vec{x})|^2}{\int_{\mathbb{R}^n} |f(\vec{x})|^2}.$$

The Fourier uncertainty principle states that $(\Delta f) \cdot (\Delta \mathcal{F} f) \geq c_n$, for a constant c_n that depends on dimension n and the choice of Fourier transform normalization. For one potential resolution penalty, we would like to penalize the dispersion of the PSF of an iterative reconstruction. Finding this PSF is a computationally difficult task, but we can use frequency-domain information, via the Fourier uncertainty principle, to indirectly infer resolution information. If $\mathbf{e}_\mathrm{c}$ is the unit vector with a 1 representing the center voxel, and the center voxel is not near an edge (so that locally, its regularization is effectively quadratic), its reconstruction will be approximately

$$\hat{\mathbf{e}}_\mathrm{c} \approx (\mathbf{A}^\mathrm{T}\mathbf{W}\mathbf{A} + \alpha\mathbf{P})^{-1}\mathbf{A}^\mathrm{T}\mathbf{W}\mathbf{A}\mathbf{e}_\mathrm{c}. \tag{5.7}$$

Using the approximation of $\mathbf{A}^\mathrm{T}\mathbf{W}\mathbf{A}$ and $\mathbf{P}$ as local filters as in (2.55) and (2.56) allows us to represent (5.7) as a filtering of $\mathbf{e}_c$ by a filter with frequency response F_c^W:

$$\begin{aligned} \hat{\mathbf{e}}_\mathrm{c} &\approx \mathcal{F}^* \mathcal{D}\left\{F_\mathrm{c}^W\right\} \mathcal{F}\mathbf{e}_\mathrm{c} \\ F_\mathrm{c}^W(\vec{\nu}) &= \frac{H_\mathrm{c}^W(\vec{\nu})}{H_\mathrm{c}^W(\vec{\nu}) + \alpha R(\vec{\nu})}. \end{aligned} \tag{5.8}$$

By approximating the dispersion ΔF_c^W, we can find a lower bound on the dispersion of $\hat{\mathbf{e}}_\mathrm{c}$, the local PSF.

If we use the approximate factorization (4.9) of H_c^W in (5.8), as with variance prediction, we can rearrange terms to approximate the dispersion of F_c^W as the fraction of two sums over each view:

$$\Delta F_\mathrm{c}^W \approx \frac{\sum_{v=1}^V G_{\mathrm{cyl},2,2,2}(\phi_v, KN_v/\alpha A_v)}{\sum_{v=1}^V G_{\mathrm{cyl},0,2,2}(\phi_v, KN_v/\alpha A_v)}, \tag{5.9}$$

where $G_{\mathrm{cyl},m,n,d}$ is a generalization of G_cyl:

$$G_{\mathrm{cyl},m,n,d}(\Phi, \gamma) = \iint \frac{(\gamma J(\vec{\nu}))^n}{(\gamma J(\vec{\nu}) + R(\vec{\nu}))^d} (\rho^2 + \nu_3^2)^{m/2} \rho \, d\nu_3 \, d\rho. \tag{5.10}$$

Note that $G_{\text{cyl},0,1,2} = G_{\text{cyl}}$. The inverse of ΔF_{c}^{W} represents this lower bound on the dispersion of the PSF; we define it as $\text{RP}(\mathbf{N})$, a resolution penalty:

$$\text{RP}(\mathbf{N}) \triangleq \frac{\sum_{v=1}^{V} G_{\text{cyl},0,2,2}(\phi_v, KN_v/\alpha A_v)}{\sum_{v=1}^{V} G_{\text{cyl},2,2,2}(\phi_v, KN_v/\alpha A_v)} \tag{5.11}$$

Of course, as both an approximation and a lower bound, this provides no guarantees about the PSF resolution; however, we use it as a starting point to improve the conditioning of (5.6). The formulation that is the easiest to minimize is

$$\mathbf{N}^* = \text{argmin}_{\mathbf{N}}\, \beta\, \text{VP}(\mathbf{N}) + \lambda\, \text{RP}(\mathbf{N}) + \sum_{v} N_v. \tag{5.12}$$

That is, we consider optimizing the tube current for each view by minimizing a cost function consisting of terms corresponding to the predicted variance and inverse resolution at iso-center, and the total incident radiation dose.

For the purposes of evaluating our method, we also define the anisotropy prediction $\text{AP}(\mathbf{N})$, defined as:

$$\text{AP}(\mathbf{N}) \triangleq \text{var}_i \frac{G_{\text{cyl},0,2,2}(\phi_i, KN_i/\alpha A_i)}{\sum_{v=1}^{V} G_{\text{cyl},2,2,2}(\phi_v, KN_v/\alpha A_v)}. \tag{5.13}$$

For an isotropic PSF, $\text{AP} = 0$.

5.2.1.2 Results

To evaluate (5.12), we compared the estimated PSF that results from a variety of ATCM methods. The PSF was estimated using (5.8), with the local frequency response H_{c}^{W} estimated as 3D Fourier transform of $\mathbf{A}^{\text{T}}\mathbf{W}\mathbf{A}\mathbf{e}_{\text{c}}$, where the projection and back-projection are actually performed, not approximated via (4.9). These PSFs were simulated using an XCAT phantom and a third-generation GE system geometry. Figure 5.1 shows the estimated PSF of a reconstruction using a uniform tube current. Figure 5.2 shows the PSF of a reconstruction using a tube current matched to the attenuation ($N_v \propto A_v$), which [22] shows for FBP reconstruction should have an equal variance to the uniform case but an isotropic PSF.

Figure 5.3 shows, over a sampling of 41 views over a π-radian source arc (each of the remaining views uses the same tube current and has the same attenuation as the view opposite it), the attenuation of that view (in black), and the tube current that results from solving (5.12) (in red). On this scale, the uniform tube current is given by

100, so that the red line represents a percentage of the original tube current. Figure 5.4 shows the estimated PSF from the proposed ATCM scheme shown in Figure 5.3. Figure 5.5 shows profiles through the center of the PSF for the uniform (black) and proposed method (red). Table 5.1 shows the predicted variance penalty, resolution penalty, and anisotropy for each of these modulation methods.

5.2.1.3 Discussion

From Figures 5.1, 5.2, and Table 5.1, we note that the findings of [22] that the attenuation-matched modulation should produce an image with the same noise variance as a uniform modulation but with an isotropic PSF are largely confirmed for iterative reconstruction. The PSF is visibly isotropic, which is also reflected in its corresponding anisotropy prediction being smaller by orders of magnitude.

The tube current modulation scheme we have proposed seems to work, but only based on the metrics we have defined—the noise standard deviation is reduced by about 1/3, while reducing the resolution penalty and cutting the total tube current to about 1/4 that of the uniform dosage. However, the resulting PSF would be inappropriate in an image reconstruction, which calls the method and utility of the metrics into question. By going so far as to effectively turn the tube off for the 3/4 of the scan with the highest attenuation, our method's "proposed method" for dose reduction is a severely limited-angle CT that chooses the most efficient limited angles, which squeezes all of the PSF dispersion into one axis. This is reflected in the anisotropy prediction for this PSF being nearly an order of magnitude larger than the uniform PSF.

5.2.2 CRC prediction

5.2.2.1 Methods

Qi [47] measures resolution in statistical reconstructions of PET (although, as a measure of resolution, this is also applicable to CT images) using the *contrast recovery coefficient* (CRC), which is defined in terms of the local impulse response of the reconstruction. This reconstruction local impulse response (RLIR), not to be confused with the local impulse response of $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$, we define here as:

$$\hat{\mathbf{e}}^{j}(\mathbf{x}) \triangleq \lim_{\delta \to 0} \frac{\mathbb{E}\left[\hat{\mathbf{x}}(\mathbf{y}(\mathbf{x} + \delta \mathbf{e}_j))\right] - \mathbb{E}\left[\hat{\mathbf{x}}(\mathbf{y}(\mathbf{x}))\right]}{\delta}, \tag{5.14}$$

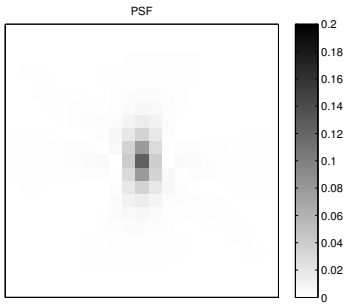


Figure 5.1: PSF of uniform tube current reconstruction.

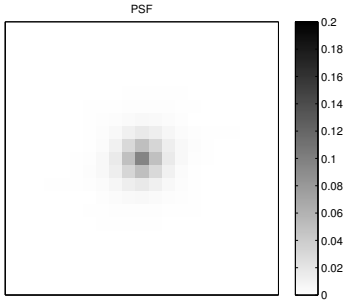


Figure 5.2: PSF of attenuation-matched tube current reconstruction.

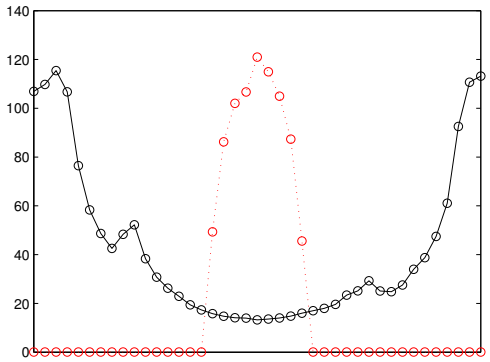


Figure 5.3: Attenuation (black) and proposed tube current modulation (red) for select views over a π-radian source arc.

Modulation	**VP** (HU)	**RP** (no units)	**AP** (no units)
Uniform	33.0	14.7	0.0206
Attenuation-Matched	31.8	19.1	$4.6 \cdot 10^{-6}$
Proposed	21.6	7.9	0.152

Table 5.1: PSF Measures for TCM methods; VP defined in (5.5), RP in (5.11), AP in (5.13).

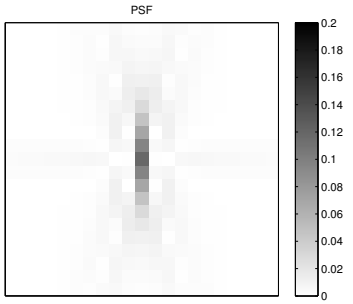


Figure 5.4: PSF of proposed TCM reconstruction.

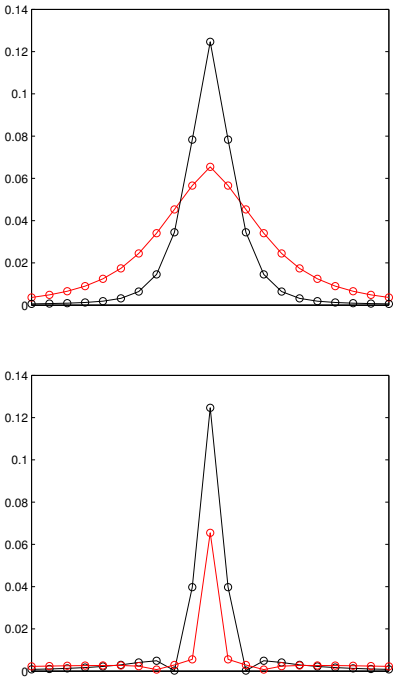


Figure 5.5: Profiles through the center of PSF profiles of uniform current (black) and designed TCM (red). Each marked point is one voxel. The top figure is a vertical profile through the PSF and the bottom figure is a horizontal profile.

where $\mathbf{x}$ is a true image, $\mathbf{y}$ is the projection data given an image $\mathbf{x}$, and $\hat{\mathbf{x}}$ represents the statistical reconstruction of $\mathbf{x}$ from the projection data $\mathbf{y}$. This RLIR we can think of as the derivative of the reconstruction $\hat{\mathbf{x}}$ with respect to a perturbation of one voxel of the original image. Under the assumptions on the form of the reconstruction we have already made for the purposes of variance prediction, the RLIR simplifies to

$$\hat{\mathbf{e}}^j(\mathbf{x}) \approx (\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} + \alpha\nabla^2 R(\hat{\mathbf{x}}))^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}\mathbf{e}_j. \tag{5.15}$$

We assume that the regularizer has the form of (2.27) with Hessian matrices given again by (2.29). If we follow this by the approximation that $\ddot{\boldsymbol{\Psi}} \approx \mathbf{I}$, as in the case where we simplified (2.29), the RLIR (5.15) becomes simply

$$\hat{\mathbf{e}}^j \approx (\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} + \alpha\mathbf{P})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}\mathbf{e}_j. \tag{5.16}$$

Under this approximation, the RLIR is no longer a function of the image being perturbed, since the RLIR is simply a linear function of the image $\mathbf{x}$. We interpret (5.16) as stating that the effect of a small one-voxel perturbation to the non-linear reconstruction $\hat{\mathbf{x}}$ of an image $\mathbf{x}$ can be approximated as the "reconstruction" of an impulse $\mathbf{e}_j$ under a quadratic approximation of the reconstruction algorithm. The CRC, then, is the peak of this RLIR, assumed to be at the same location of the impulse:

$$\mathrm{crc}_j \triangleq \hat{\mathbf{e}}^j_j \tag{5.17}$$

This has been used as a measure of resolution by [47], and [36] shows that the CRC has a significant rank correlation to the detectability of a lesion in an image.

If we apply the local shift-invariant approximation for $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$ and $\mathbf{P}$:

$$\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} \approx \mathcal{F}^*\mathcal{D}\left\{H_j^W(\vec{\nu})\right\}\mathcal{F}$$
$$\mathbf{P} \approx \mathcal{F}^*\mathcal{D}\left\{R(\vec{\nu})\right\}\mathcal{F},$$

then (5.16) becomes simply a filtering of the impulse, exactly as in (5.8):

$$\hat{\mathbf{e}}^j \approx \mathcal{F}^*\mathcal{D}\left\{F_j^W(\vec{\nu})\right\}\mathcal{F}\mathbf{e}_j \tag{5.18}$$

$$F_j^W(\vec{\nu}) \triangleq \frac{H_j^W(\vec{\nu})}{H_j^W(\vec{\nu}) + \alpha R(\vec{\nu})}, \tag{5.19}$$

and measurement of the CRC from (5.18) becomes an integral of the filter over n-

dimensional frequency space:

$$\mathrm{crc}_j \approx \int\limits_{\left[-\frac{1}{2},\frac{1}{2}\right]^n} F_j^W(\vec{\nu})\; d\vec{\nu}. \tag{5.20}$$

Since our approximations for H_j^W and R are strictly real and non-negative, it is clear from the definition of F_j^W that crc_j is between 0 and 1. In cases where $R(\vec{0}) = 0$, which includes the case where $\mathbf{C}$ is a first-differencing matrix, $F_j^W(\vec{0}) = 1$. Since this DC term of the RLIR is one, a CRC close to one suggests that most of the RLIR is concentrated at the center voxel, which is desirable for resolution. A lower CRC suggests that the RLIR must be spread out further from the center and that consequently the resolution is poorer.

Conveniently, the approximate factorizations of either (3.25) for generic CT geometries, or (4.9) for small-angle cone beam geometries, make the evaluation of (5.20) simple. In the general case, we define a generalization of G in (4.4):

$$G_{m,p,q,n}(\gamma, \vec{\Theta}) \triangleq \int\limits_0^{\varrho_{\max}(\vec{\Theta})} \frac{(\gamma J(\varrho, \vec{\Theta}))^p}{(\gamma J(\varrho, \vec{\Theta}) + R(\varrho, \vec{\Theta}))^q} \varrho^m \varrho^{n-1}\; d\varrho; \tag{5.21}$$

the G from (4.4) in this notation in n dimensions is the specific case $G_{0,1,2,n}$ of (5.21). In terms of this function, expanding F_j^W and rearranging terms, (5.20) becomes

$$\int\limits_{\mathbb{S}^n} G_{0,1,1,n}(E_j^W(\vec{\Theta})/\alpha, \vec{\Theta})\; d\vec{\Theta}. \tag{5.22}$$

In the 3D axial case, if we factor H_j^W using (4.9), (5.22) becomes

$$\mathrm{crc}_j \approx \int\limits_0^{2\pi} G_{\mathrm{cyl},0,1,1}(\Phi, E_j^W(\Phi)/\alpha)\, d\Phi. \tag{5.23}$$

For the center voxel, (5.23) simplifies to

$$\mathrm{crc}_\mathrm{c} \approx \frac{2\pi}{V} \sum_{v=1}^{V} G_{\mathrm{cyl},0,1,1}(\phi_v, KN_v/\alpha A_v). \tag{5.24}$$

5.2.2.2 Discussion

Preliminary results suggested that the tube current modulation scheme that results from the minimization of (5.12) using the CRC as the resolution penalty in place of the dispersion-based resolution penalty of (5.11) is overly anisotropic in a similar way to the results of the dispersion-based penalty. The CRC is somewhat similar to the dispersion-based penalty in that the PSF is 'allowed' to spread out a lot in one direction without penalizing the modulation scheme too greatly. For this reason we did not further consider the CRC-based penalty for the purposes of tube current modulation. However, the forms (5.22) and (5.23) may be useful for future work where fast prediction of the CRC is required.

5.3 Prediction of model observer performance

5.3.1 Background

The previous sections in this chapter have all focused on minimizing the variance of a image to improve its quality, although as we have shown in the no-radiation image reconstruction case, simply minimizing variance is not necessarily the best predictor of image quality for the purposes of some tasks, such as detection of a particular feature. As an alternative to measuring image variance and resolution, we consider the prediction of the performance of linear image observers for the purposes of analyzing tube current modulation methods, as wall as for other potential statistical analyses. We are interested in the fast computation of these performance indices using the local frequency approximations of Chapter III.

The main type of observer that we will evaluate are algorithms (referred to as *model observers* to differentiate from human observers) that distinguish which of two distributions a random vector $\mathbf{x}$ was drawn from, where the observer knows some information about the distributions. The correlation between the performance of model observers and human observers in detecting whether a known signal (e.g. a lesion) is or is not present on top of a known background image (e.g. a patient) has been well-studied. Model observers that show a high correspondence with the performance of human observers have been used to study many medical imaging methods without the need of live human observers, who are reportedly more difficult to use than purely algorithmic observers. For example, [71, 72] apply model observers to optimize tube current modulation for FBP-reconstructed images. In [62, 46], the non-prewhitened image observer (which we will use in Section 5.3.3.3) is used to

quantify the performance and dose reduction potential of statistical reconstruction. In [23], the performance of model observers is both correlated with that of human observers, and used to evaluate scatter-correction strategies for SPECT.

We will consider a few specific model observers. In Section 5.3.2, we describe a ratio test observer, which is the uniformly most powerful test for determining whether a feature is present, but is not a reasonable model of a human observer. In Section 5.3.3, we consider linear observers, which use a test statistic consisting of a linear combination of sinogram elements or reconstructed image voxels to determine the class of the sinogram or image. The primary observer we will apply to tube current modulation is the non-prewhitened image observer described in Section 5.3.3.3, which was shown in [23] to have the highest performance correlation with human observers in their radiological detection task.

5.3.2 Ratio-test observers

We first consider the simple case of a model observer performing a likelihood-ratio test to discriminate whether $\mathbf{z}$ was drawn from one distribution or another (call these Υ_0 and Υ_1). Suppose we compare the likelihood-ratio,

$$\Lambda(\mathbf{z}) \triangleq \frac{f_{\mathbf{z}}(\mathbf{z}|\Upsilon_1)}{f_{\mathbf{z}}(\mathbf{z}|\Upsilon_0)},$$

to a fixed value η and conclude that $\mathbf{z}$ was drawn from Υ_1 if $\Lambda(\mathbf{z}) > \eta$, and Υ_0 otherwise. Assuming this calculation is possible, it will be the uniformly most powerful test for distinguishing the two classes [42].

Consider the actual CT problem at hand, however. Consider two images, one a fixed background image $\mathbf{z}$, and the other the background image plus some feature $\mathbf{f}$ to detect. Once projected, these become probabalistic vectors, either a random vector of sinogram observations $\mathbf{y}(\mathbf{z})$ projected from a featureless image, or a vector $\mathbf{y}(\mathbf{z}+\mathbf{f})$ projected from the image containing $\mathbf{f}$. Then the problem of detecting the feature is distinguishing whether the reconstructed image, $\hat{\mathbf{z}}(\mathbf{y})$, which is now a random variable, was ultimately produced with or without the feature; call these distributions Υ_1 and Υ_0, respectively. Since $\hat{\mathbf{z}}$ cannot extract more information from $\mathbf{y}$ than was already present, the likelihood-ratio statistic on the reconstructed images, $\Lambda(\hat{\mathbf{z}}) = f_{\hat{\mathbf{z}}}(\hat{\mathbf{z}}|\Upsilon_1)/f_{\hat{\mathbf{z}}}(\hat{\mathbf{z}}|\Upsilon_0)$, cannot do better than the likelihood-ratio statistic produced directly from the sinogram, $\Lambda(\mathbf{y}) = f_{\mathbf{y}}(\mathbf{y}|\Upsilon_1)/f_{\mathbf{y}}(\mathbf{y}|\Upsilon_0)$.

This observer is not the most reasonable model of a human observer, since real-world human observers do not look directly at the projection data of a CT scan to

determine the presence of a relevant feature.

5.3.3 Hotelling and other linear observers

5.3.3.1 Background

In this section, we consider observers that produce scalar statistics that are linear functions of their input data.

Assume that a noisy vector $\mathbf{z}$ is drawn from one of two distributions, one of which has a mean $\bar{\mathbf{z}}_0$ and the other, $\bar{\mathbf{z}}_1$, and that the covariance matrices of the two distributions are given by $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$. We will denote the average of these covariance matrices $\boldsymbol{\Sigma}_{\mathbf{z}} = (\boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_1)/2$ and the difference in their means as $\bar{\mathbf{z}}_\Delta = \bar{\mathbf{z}}_1 - \bar{\mathbf{z}}_0$.

Given the vector

$$\mathbf{w}_{\text{Hot}} \triangleq \boldsymbol{\Sigma}_{\mathbf{z}}^{-1}\bar{\mathbf{z}}_\Delta,$$

the Hotelling observer statistic t for observed data $\mathbf{z}$ is given by:

$$t_{\text{Hot}} \triangleq \mathbf{w}_{\text{Hot}}^{\text{T}}\mathbf{z}. \tag{5.25}$$

If both of the distributions that $\mathbf{z}$ can be drawn from are multivariate Gaussian distributions with known and equal covariance matrices, then the log-likelihood-ratio[1] statistic is given by:

$$\begin{aligned}\ln\Lambda(\mathbf{z}) &= \frac{1}{2}(\mathbf{z}-\bar{\mathbf{z}}_0)^{\text{T}}\boldsymbol{\Sigma}_{\mathbf{z}}^{-1}(\mathbf{z}-\bar{\mathbf{z}}_0) - \frac{1}{2}(\mathbf{z}-\bar{\mathbf{z}}_1)^{\text{T}}\boldsymbol{\Sigma}_{\mathbf{z}}^{-1}(\mathbf{z}-\bar{\mathbf{z}}_1) \\ &= \bar{\mathbf{z}}_\Delta^{\text{T}}\boldsymbol{\Sigma}_{\mathbf{z}}^{-1}\mathbf{z} + C,\end{aligned}$$

where C is a constant that can be dropped. Consequently, in this case, the Hotelling observer is equivalent to a likelihood-ratio test, and so is the uniformly most powerful test for distinguishing which distribution $\mathbf{z}$ was drawn from [42]. With equal-covariance multivariate Gaussian distributions for both classes, the area under the ROC curve for the Hotelling observer test increases with the observer SNR, given by:

$$\text{SNR}_{\text{PWH}}^2 = \mathbf{w}_{\text{PWH}}^{\text{T}}\bar{\mathbf{z}}_\Delta = \bar{\mathbf{z}}_\Delta^{\text{T}}\boldsymbol{\Sigma}_{\mathbf{z}}^{-1}\bar{\mathbf{z}}_\Delta.$$

In the previous case we assume that the observer knows $\boldsymbol{\Sigma}_{\mathbf{z}}$; a linear non-prewhitening (NPW) observer may take the form of (5.25), where the linear combining vector $\mathbf{w}$

[1] The logarithm is a monotonic function, so any test that compares $\ln\Lambda \gtrless t$ is equivalent to the likelihood-ratio test $\Lambda \gtrless e^t$.

is simply

$$\mathbf{w}_{\text{NPW}} \triangleq \bar{\mathbf{z}}_{\Delta},$$

and the resulting observer statistic is $t_{\text{NPW}} = \mathbf{w}_{\text{NPW}}^{\text{T}}\mathbf{z}$. In this case,

$$\text{SNR}_{\text{NPW}}^2 = \frac{(\bar{\mathbf{z}}_{\Delta}^{\text{T}}\bar{\mathbf{z}}_{\Delta})^2}{\bar{\mathbf{z}}_{\Delta}^{\text{T}}\boldsymbol{\Sigma}_{\mathbf{z}}\bar{\mathbf{z}}_{\Delta}}$$

Suppose that we have a human observer with the task of determining whether a CT image $\mathbf{z}$ belongs to one of two classes; one class contains a notable feature and the other class does not. All of the assumptions in the above derivation of the Hotelling observer are unreasonable in the context of a human observer attempting to detect the presence or absence of feature in a CT image. The human observer does not know $\boldsymbol{\Sigma}_{\mathbf{z}}$, although the visual detection process may involve the unconscious creation of a noise model. The covariance matrices of the two classes may not be equal. Depending on the context of the detection process, the human observer may have limited or no knowledge of the class means.

Even discounting all of the previous objections, with a fully known noise model and means, a human observer would be unable to linearly synthesize all of an observed image into one statistic simply by visual observation. To aid in modeling the performance of a human observer, a common modification to the observer is the addition of *channels*, a set of linear combinations of the data that better represent the capabilities of the observer [73]. For example, for images, the channels may be banks of band-pass filters, as in [23], which are meant to simulate the frequency-selective nature of the human visual system[5]. Channelized observers that significantly reduce the dimensionality of the data also make it believable that an observer can estimate the covariance matrix of the channelized data.

These combinations can be represented simply by a wide matrix $\mathbf{U}$, such that for a vector $\mathbf{z}$, the observer makes their decision based on the channelized observation $\mathbf{U}\mathbf{z}$. Based only on these observations, the ideal prewhitened statistic is given by

$$t_{\text{CHO}} \triangleq \mathbf{w}_{\text{CHO}}^{\text{T}}\mathbf{U}\mathbf{z}, \qquad \mathbf{w}_{\text{CHO}} \triangleq (\mathbf{U}\boldsymbol{\Sigma}_{\mathbf{z}}\mathbf{U}^{\text{T}})^{-1}\mathbf{U}\bar{\mathbf{z}}_{\Delta},$$

where CHO is a common abbreviation for channelized Hotelling observer, and the SNR is given by:

$$\text{SNR}_{\text{CHO}}^2 = \bar{\mathbf{z}}_{\Delta}^{\text{T}}\mathbf{U}^{\text{T}}(\mathbf{U}\boldsymbol{\Sigma}_{\mathbf{z}}\mathbf{U}^{\text{T}})^{-1}\mathbf{U}\bar{\mathbf{z}}_{\Delta}.$$

A non-prewhitened analog to the channelized Hotelling observer may take the form

$$t_{\text{CNPW}} \triangleq \mathbf{w}_{\text{CNPW}}^{\text{T}}\mathbf{U}\mathbf{z}, \qquad \mathbf{w}_{\text{CNPW}} \triangleq \mathbf{U}\bar{\mathbf{z}}_{\Delta}, \qquad \text{SNR}_{\text{CNPW}}^2 = \frac{(\bar{\mathbf{z}}_{\Delta}^{\text{T}}\mathbf{U}^{\text{T}}\mathbf{U}\bar{\mathbf{z}}_{\Delta})^2}{\bar{\mathbf{z}}_{\Delta}^{\text{T}}\mathbf{U}^{\text{T}}\mathbf{U}\boldsymbol{\Sigma}_{\mathbf{z}}\mathbf{U}^{\text{T}}\mathbf{U}\bar{\mathbf{z}}_{\Delta}}.$$

5.3.3.2 Sinogram observers in CT

Consider the same detection problem in CT given in section 5.3.2, where we wish to distinguish a featureless background image $\mathbf{x}$ from a background image plus feature $(\mathbf{x}+\mathbf{f})$. Given our projection model, we will model the two noisy sinogram distributions as multivariate Gaussian with means $\mathbf{A}\mathbf{x}$ and $\mathbf{A}(\mathbf{x}+\mathbf{f})$, respectively. We will assume that they have the same known covariance matrix $\boldsymbol{\Sigma}_{\mathbf{y}}$ that is approximately the inverse of the weighting matrix $\mathbf{W}$ used in their reconstructions. Echoing notation used for variance approximation, we will denote $\hat{\mathbf{W}} \triangleq \mathbf{W}\boldsymbol{\Sigma}_{\mathbf{y}}\mathbf{W}$. Since the weighting matrix is usually chosen to be the inverse of some estimate for the sinogram covariance, $\hat{\mathbf{W}} \approx \mathbf{W}$. The assumption that the covariance matrices are equal is untrue, since adding the feature $\mathbf{f}$ changes the weighting matrix, but for simplicity we will assume that the effect on $\mathbf{f}$ on $\boldsymbol{\Sigma}_{\mathbf{y}}$ is small enough to be ignored.

We can examine the effect of an *approximately* pre-whitened observer, since we approximately know the noise covariance. Since the difference in sinogram means is $\mathbf{A}\mathbf{f}$, the projection of the added feature, an approximate pre-whitened Hotelling observer of a *sinogram* $\mathbf{y}$ is:

$$t_{\text{S-PW}} \triangleq \mathbf{w}_{\text{S-PW}}^{\text{T}}\mathbf{y}, \qquad \mathbf{w}_{\text{S-PW}} \triangleq \mathbf{W}\mathbf{A}\mathbf{f}.$$

We will abbreviate this as the S-PW observer, as a pre-whitened sinogram observer. The ideal Hotelling sinogram observer we will denote S-OPW, since it is a sinogram that is pre-whitened by an oracle that knows the covariance of the sinogram with and without the feature present. The S-PW observer has an SNR of

$$\text{SNR}_{\text{S-PW}}^2 = \frac{(\mathbf{f}^{\text{T}}\mathbf{A}^{\text{T}}\mathbf{W}\mathbf{A}\mathbf{f})^2}{\mathbf{f}^{\text{T}}\mathbf{A}^{\text{T}}\hat{\mathbf{W}}\mathbf{A}\mathbf{f}}.$$

If the covariance is known exactly, such that $\hat{\mathbf{W}} = \mathbf{W}$, then $\text{SNR}_{\text{S-PW}}^2 = \mathbf{f}^{\text{T}}\mathbf{A}^{\text{T}}\mathbf{W}\mathbf{A}\mathbf{f}$. Interestingly, like the likelihood-ratio test, this bypasses entirely the issue of image reconstruction. The only factors influencing the SNR of this observer is the sinogram covariance, which we could influence using tube current modulation, and the quality of the covariance approximation, which we assume is negligible.

This formulation is not extremely useful for predicting an optimal tube current

modulation scheme. Suppose we wish to identify a feature represented by an impulse at voxel j ($\mathbf{f} = \mathbf{e}_j$) and that the sinogram covariance is known. Then the SNR is given by:

$$\mathrm{SNR}^2_{\mathrm{S-PW}} = [\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}]_{jj} = \sum_k [\mathbf{W}]_{kk}[\mathbf{A}]^2_{kj}.$$

We can then approximate the ideal tube-current modulation by maximizing this SNR subject to a limit on total incident radiation. Let N_k represent the number of incident photons in observation k; then $[\mathbf{W}]_{kk}$ is proportional to $N_k e^{-[\mathbf{A}\mathbf{x}]_k}$. If we can control the incident radiation in each ray freely, then the solution to

$$\arg\max_{\mathbf{N}} \sum_k N_k e^{-[\mathbf{A}\mathbf{x}]_k}[\mathbf{A}]^2_{kj} \quad \text{s.t.} \quad N_k \geq 0, \quad \sum_k N_k = N_0$$

is to set $N_k = N_0$ if k maximizes $e^{-[\mathbf{A}\mathbf{x}]_k}[\mathbf{A}]^2_{kj}$ and $N_k = 0$ otherwise. How do we interpret this? The k that maximizes $e^{-[\mathbf{A}\mathbf{x}]_k}[\mathbf{A}]^2_{kj}$ is the ray that does the best job of both passing through voxel j and not getting attenuated. If we know exactly what the background image looks like and which voxel contains some feature, the best way to detect the feature's presence is to allocate all of our alloted photons into what we know is the clearest view of that one voxel. Even if we restrict the possible range of tube-current modulation to be one number of photons per angle, and not per ray, this formulation suggests putting all the radiation into one angle. Clearly, this is an unrealistic formulation for general CT tube-current modulation.

5.3.3.3 Image observers in CT

We will now look exclusively at linear observers of an image reconstruction $\hat{\mathbf{x}}$, since this is the actual vector undergoing observation in practice. We will consider the quadratically-penalized least squares reconstruction

$$\hat{\mathbf{x}}_{\mathrm{QPWLS}}(\mathbf{y}) = (\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} + \alpha\mathbf{P})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{y}.$$

We can think of this as having channels $\mathbf{U}_{\mathrm{QPWLS}} = (\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} + \alpha\mathbf{P})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}$. If we apply a Hotelling observer to $\hat{\mathbf{x}}$, the effect is essentially the same as using a pre-whitening observer on the sinogram data, as above, since the ideal linear observer will just 'undo' the multiplication by $\mathbf{U}_{\mathrm{QPWLS}}$. The only difference from the pre-whitened sinogram observer is that this pre-whitened image observer does not have access to data contained in the nullspace of $\mathbf{U}_{\mathrm{QPWLS}}$. In frequency-domain approximations of

the reconstruction, there are commonly no zeros of the LFR of the reconstruction filter, and so the prewhitened observers of the image and of the sinogram will be identical. We will abbreviate the prewhitened observer of the image as I-PW.

One non-prewhitening observer of $\hat{\mathbf{x}}_{\text{QPWLS}}$, which we abbreviate I-NPW, is given by the inner product of the reconstruction and $\hat{\mathbf{f}} = (\mathbf{A}^{\text{T}}\mathbf{W}\mathbf{A} + \alpha\mathbf{P})^{-1}\mathbf{A}^{\text{T}}\mathbf{W}\mathbf{A}\mathbf{f}$, the difference in the means of the QPWLS reconstructions of the image with and without the feature present. The observer statistic is given by

$$t_{\text{I-NPW}}(\hat{\mathbf{x}}) = \hat{\mathbf{f}}^{\text{T}}\hat{\mathbf{x}}.$$

The I-NPW observer has an SNR given by:

$$\begin{aligned}\text{SNR}^2_{\text{I-NPW}} &= \frac{(\mathbf{f}^{\text{T}}\mathbf{A}^{\text{T}}\mathbf{W}\mathbf{A}(\mathbf{A}^{\text{T}}\mathbf{W}\mathbf{A} + \alpha\mathbf{P})^{-2}\mathbf{A}^{\text{T}}\mathbf{W}\mathbf{A}\mathbf{f})^2}{\mathbf{f}^{\text{T}}\mathbf{A}^{\text{T}}\mathbf{W}\mathbf{A}(\mathbf{A}^{\text{T}}\mathbf{W}\mathbf{A} + \alpha\mathbf{P})^{-2}\mathbf{A}^{\text{T}}\hat{\mathbf{W}}\mathbf{A}(\mathbf{A}^{\text{T}}\mathbf{W}\mathbf{A} + \alpha\mathbf{P})^{-2}\mathbf{A}^{\text{T}}\mathbf{W}\mathbf{A}\mathbf{f}} \\ &= \frac{\hat{\mathbf{f}}^{\text{T}}\hat{\mathbf{f}}}{\hat{\mathbf{f}}^{\text{T}}\boldsymbol{\Sigma}_{\hat{\mathbf{x}}}\hat{\mathbf{f}}},\end{aligned}$$

where $\boldsymbol{\Sigma}_{\hat{\mathbf{x}}}$ is the reconstructed image covariance. Fast computation of this expression is possible if we apply frequency-domain approximations for $\mathbf{A}^{\text{T}}\mathbf{W}\mathbf{A}$ and $\mathbf{P}$. If the feature $\mathbf{f}$ has a DSFT of $F(\vec{\nu})$, then this observer's SNR is approximately:

$$\text{SNR}^2_{\text{I-NPW}} \approx \frac{\left(\int \frac{|F(\vec{\nu})|^2 H_j^W(\vec{\nu})^2}{(H_j^W(\vec{\nu}) + \alpha R(\vec{\nu}))^2}\, d\vec{\nu}\right)^2}{\int \frac{|F(\vec{\nu})|^2 H_j^W(\vec{\nu})^2 H_j^{\hat{W}}(\vec{\nu})}{(H_j^W(\vec{\nu}) + \alpha R(\vec{\nu}))^4}\, d\vec{\nu}}. \tag{5.26}$$

Fast computation of (5.26) is complicated by the presence of the arbitrary feature DSFT $F(\vec{\nu})$. However, we can simplify the evaluation of (5.26) for some classes of features. For example, in Appendix C we derive an approximation for integrals of the general form

$$\int\limits_{\left[-\frac{1}{2},\frac{1}{2}\right]^n} |F(\vec{\nu})|^2 \frac{H_j^W(\vec{\nu})^p H_j^{\hat{W}}(\vec{\nu})^{\hat{p}}}{(H_j^W(\vec{\nu}) + \alpha R(\vec{\nu}))^q}\, d\vec{\nu} \tag{5.27}$$

for Gaussian bump features, where $|F(\vec{\nu})| \propto \exp(-\tau||\vec{\nu}||_2^2)$. For these features, (5.26) can be computed by numerically integrating

$$\text{SNR}^2_{\text{I-NPW}} \approx \frac{\left(\int_{\mathbb{S}^n} G_{\text{gauss},2,2,n}(2\tau, R_0\alpha/E_j^W(\vec{\Theta}))\, d\vec{\Theta}\right)^2}{\int_{\mathbb{S}^n} \frac{E_j^{\hat{W}}(\vec{\Theta})}{E_j^W(\vec{\Theta})^2} G_{\text{gauss},3,4,n}(2\tau, R_0\alpha/E_j^W(\vec{\Theta}))\, d\vec{\Theta}}, \tag{5.28}$$

where $G_{\text{gauss},p,q,n}$ is a function that we can pre-compute tables of. When the bump is specified in terms of its FWHM, τ in (5.28) is equal to $(\pi^2/4\ln 2)\cdot \text{FWHM}^2$. Figure 5.6 shows values of the table of $G_{\text{gauss},p,q,2}(9\pi^2/2\ln 2, \gamma)$, corresponding to a Gaussian bump function with a FWHM of 3 pixels, for the six choices of p and q that are relevant to our methods. The most notable feature common to each of these tables is that each of them contains a region, for $\gamma < 1$, where the function is essentially a constant that does not depend on γ. This is also the case when $q = 0$. Both of these facts are understandable from the definition of the function in (C.3). Also notable is that in the cases where $p, q = 1, 1$ and $2, 2$, the function is nearly proportional to $\gamma^{-2/3}$ for $\gamma \geq 10^5$; in the cases of $p, q = 1, 2$ and $3, 4$, the function is nearly proportional to γ^{-1} for $\gamma \geq 10^5$.

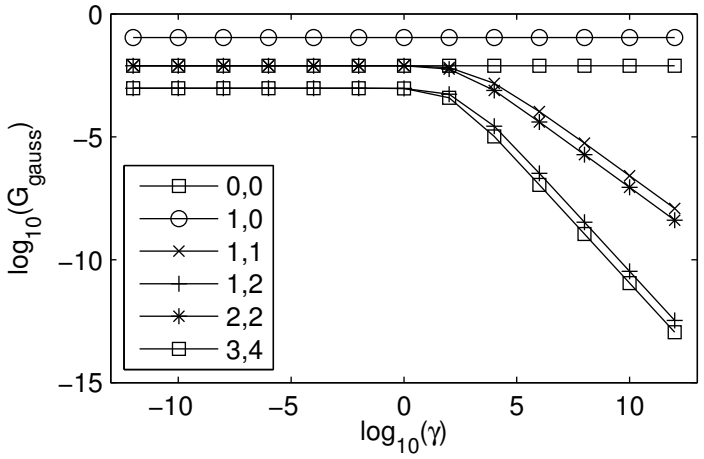


Figure 5.6: Values of $G_{\text{gauss},p,q,2}(\tau, \gamma)$ for a 3-pixel FWHM Gaussian bump.

We also analyze a non-prewhitened observer for which the observer statistic is

$$t_{\text{UR}-\text{NPW}}(\hat{\mathbf{x}}) = \mathbf{f}^{\text{T}}\hat{\mathbf{x}}.$$

We abbreviate this observer UR-NPW, for unknown-reconstruction, non-prewhitened. Some computational observers may know the original feature $\mathbf{f}$, but may be unable to know or compute the reconstruction $\hat{\mathbf{f}}$ of $\mathbf{f}$. In this case, they may use the UR-NPW observer. The SNR of this observer is:

$$\text{SNR}^2_{\text{UR}-\text{NPW}} = \frac{(\mathbf{f}^{\text{T}}\hat{\mathbf{f}})^2}{\mathbf{f}^{\text{T}}\boldsymbol{\Sigma}_{\hat{\mathbf{x}}}\mathbf{f}}.$$

The numerator of the UR-NPW SNR has the approximation

$$\mathbf{f}^{\mathrm{T}}\hat{\mathbf{f}} \approx \int\limits_{\left[-\frac{1}{2},\frac{1}{2}\right]^n} |F(\vec{\nu})|^2 \frac{H_j^W(\vec{\nu})}{H_j^W(\vec{\nu}) + \alpha R(\vec{\nu})}\, d\vec{\nu}. \tag{5.29}$$

The denominator relies on the unknown image covariance, but this has a frequency domain approximation

$$\mathbf{f}^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{\mathbf{x}}}\mathbf{f} \approx \int\limits_{\left[-\frac{1}{2},\frac{1}{2}\right]^n} |F(\vec{\nu})|^2 \frac{H_j^{\hat{W}}(\vec{\nu})}{(H_j^W(\vec{\nu}) + \alpha R(\vec{\nu}))^2}\, d\vec{\nu}. \tag{5.30}$$

For a Gaussian $F(\vec{\nu})$, we use the table-based approximation from Appendix C to approximate the denominator (5.31) as

$$\mathbf{f}^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{\mathbf{x}}}\mathbf{f} \approx \int\limits_{\mathbb{S}^n} \frac{E_j^{\hat{W}}(\vec{\Theta})}{E_j^W(\vec{\Theta})^2} G_{\mathrm{gauss},1,2,n}(2\tau, \alpha R_0 / E_j^W(\vec{\Theta}))\, d\vec{\Theta}. \tag{5.31}$$

5.3.3.4 Channelized image observers in CT

Many models of human observers use a channelized Hotelling observer with a very small number of channels N_c, potentially as few as five. In this case, we will enumerate the channels, represented as N_c column vectors $\mathbf{u}_1, \ldots, \mathbf{u}_{N_c}$, or in the frequency domain as filters $U_1(\vec{\nu}), \ldots, U_{N_c}(\vec{\nu})$. When represented in this way, the channel matrix $\mathbf{U}$ is given by:

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{u}_{N_c}^{\mathrm{T}} \end{bmatrix},$$

and the outputs of the channeled observations of $\mathbf{x}$, which has a frequency-domain representation $X(\vec{\nu})$, are equivalently given by either

$$\mathbf{c} \triangleq \mathbf{U}\mathbf{x} = \begin{bmatrix} \mathbf{u}_1^{\mathrm{T}}\mathbf{x} \\ \vdots \\ \mathbf{u}_{N_c}^{\mathrm{T}}\mathbf{x} \end{bmatrix} = \begin{bmatrix} \int U_1(\vec{\nu}) X(\vec{\nu})\, d\vec{\nu} \\ \vdots \\ \int U_{N_c}(\vec{\nu}) X(\vec{\nu})\, d\vec{\nu} \end{bmatrix},$$

where $\mathbf{c}$ denotes the channelized observations. The covariance matrix of these channelized observations, which we will denote $\boldsymbol{\Sigma}_{\mathbf{c}}$, is equal to $\boldsymbol{\Sigma}_{\mathbf{c}} = \mathbf{U}\boldsymbol{\Sigma}_{\mathbf{x}}\mathbf{U}^{\mathrm{T}}$. Given that this matrix is small, only $N_c \times N_c$, it is computationally feasible to examine all

of its elements separately using frequency-domain approximations. If the covariance $\mathbf{\Sigma_x}$ can be represented by a noise power spectrum $S(\vec{\nu})$, then one element of the covariance matrix is given by:

$$[\mathbf{\Sigma_c}]_{ij} = \mathbf{u}_i^{\mathrm{T}} \mathbf{\Sigma_x} \mathbf{u}_j = \int S(\vec{\nu})\, U_i(\vec{\nu})\, U_j(\vec{\nu})\; d\vec{\nu}. \tag{5.32}$$

5.3.3.5 Prewhitened channelized image observers

In this section, we describe fast frequency-domain approximations to the performance of channelized image observers. While we will not relate these to the problem of tube current modulation, we will present them for potential future applications of fast performance prediction.

The idea of a human observer intuitively prewhitening a signal when observing it becomes more reasonable when we model the observer only needing to estimate and invert the much smaller $N_c \times N_c$ matrix of channel covariances. In our case of quadratically-penalized CT reconstruction, we will denote the pre-whitened channelized Hotelling observer as C-PW. The C-PW SNR, again for a feature $\mathbf{f}$, is given by:

$$\mathrm{SNR}^2_{\mathrm{C-PW}} = \mathbf{f}^{\mathrm{T}} \mathbf{A}^{\mathrm{T}} \mathbf{W} \mathbf{A} (\mathbf{A}^{\mathrm{T}} \mathbf{W} \mathbf{A} + \alpha \mathbf{P})^{-1} \mathbf{U}^{\mathrm{T}} \mathbf{\Sigma}_{\hat{\mathbf{c}}}^{-1} \mathbf{U} (\mathbf{A}^{\mathrm{T}} \mathbf{W} \mathbf{A} + \alpha \mathbf{P})^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{W} \mathbf{A} \mathbf{f} \tag{5.33}$$

$$= \hat{\mathbf{f}}^{\mathrm{T}} \mathbf{U}^{\mathrm{T}} \mathbf{\Sigma}_{\hat{\mathbf{c}}}^{-1} \mathbf{U} \hat{\mathbf{f}} \tag{5.34}$$

$$\mathbf{\Sigma}_{\hat{\mathbf{c}}} = \mathbf{U} \mathbf{\Sigma}_{\hat{\mathbf{x}}} \mathbf{U}^{\mathrm{T}}, \tag{5.35}$$

where $\hat{\mathbf{c}}$ represents the channelized observations of the reconstruction; $\hat{\mathbf{c}} = \mathbf{U}\hat{\mathbf{x}}$. While this is unwieldy, we can simplify with frequency-domain approximations. One element of $\mathbf{\Sigma}_{\hat{\mathbf{c}}}$ can be found via (5.32), or equivalently from (5.35), with our usual approximation for the local noise power spectrum:

$$[\mathbf{\Sigma}_{\hat{\mathbf{c}}}]_{ij} \approx \int \frac{H^{\hat{W}}(\vec{\nu})}{(H^W(\vec{\nu}) + \alpha R(\vec{\nu}))^2} U_i(\vec{\nu}) U_j(\vec{\nu})\; d\vec{\nu}. \tag{5.36}$$

If we compute $\mathbf{\Sigma}_{\hat{\mathbf{c}}}$ and its inverse, the expression for SNR we can simplify somewhat

with the summation and frequency-domain approximation:

$$\mathrm{SNR}^2_{\mathrm{C-PW}} = \sum_{i,j} \left[\mathbf{\Sigma}_{\hat{\mathbf{c}}}^{-1}\right]_{ij} \left(\hat{\mathbf{f}}^{\mathrm{T}}\mathbf{u}_i\right)\left(\hat{\mathbf{f}}^{\mathrm{T}}\mathbf{u}_j\right) \tag{5.37}$$

$$\approx \sum_{i,j} \left[\mathbf{\Sigma}_{\hat{\mathbf{c}}}^{-1}\right]_{ij} \left(\int F(\vec{\nu}) \frac{H^W(\vec{\nu})}{H^W(\vec{\nu}) + \alpha R(\vec{\nu})} U_i(\vec{\nu})\, d\vec{\nu}\right) \cdot \left(\int F(\vec{\nu}) \frac{H^W(\vec{\nu})}{H^W(\vec{\nu}) + \alpha R(\vec{\nu})} U_j(\vec{\nu})\, d\vec{\nu}\right). \tag{5.38}$$

Some common choices of channels are non-overlapping in the frequency domain, such that $\int U_i(\vec{\nu}) U_j(\vec{\nu})\, d\vec{\nu} = 0$ if $i \neq j$. In this case, (5.38) simplifies greatly; since $\mathbf{\Sigma}_{\hat{\mathbf{c}}}$ becomes a diagonal matrix, the SNR can be approximated by:

$$\mathrm{SNR}^2_{\mathrm{C-PW}} \approx \sum_i \frac{\left(\int F(\vec{\nu}) \frac{H^W(\vec{\nu})}{H^W(\vec{\nu}) + \alpha R(\vec{\nu})} U_i(\vec{\nu})\, d\vec{\nu}\right)^2}{\int \frac{H^{\hat{W}}(\vec{\nu})}{(H^W(\vec{\nu}) + \alpha R(\vec{\nu}))^2} U_i(\vec{\nu})^2\, d\vec{\nu}}. \tag{5.39}$$

Whether or not the channels are overlapping, our separable approximation to the frequency response $H_j^W(\vec{\nu})$ enables rapid computation of either (5.38) or (5.39). First, we compute the elements of $\mathbf{\Sigma}_{\hat{\mathbf{c}}}$, using (5.36):

$$[\mathbf{\Sigma}_{\hat{\mathbf{c}}}]_{ij} \approx \frac{1}{\alpha} \int\limits_{\mathbb{S}^{n-1}} \frac{E^{\hat{W}}(\vec{\Theta})}{E^W(\vec{\Theta})} G_{\mathrm{chan},1,2,ij}(\vec{\Theta}, E^W(\vec{\Theta})/\alpha)\, d\vec{\Theta} \tag{5.40}$$

$$G_{\mathrm{chan},1,2,ij}(\vec{\Theta}, \gamma) \triangleq \int\limits_0^{\varrho_{\max}(\vec{\Theta})} \frac{\gamma J(\varrho, \vec{\Theta})}{(\gamma J(\varrho, \vec{\Theta}) + R(\varrho, \vec{\Theta}))^2} U_i(\varrho, \vec{\Theta}) U_j(\varrho, \vec{\Theta}) \varrho^{n-1}\, d\varrho. \tag{5.41}$$

Equation (5.40) can be computed quickly using tables for $G_{\mathrm{chan},1,2,ij}$, although this requires N_c^2 tables, which may become infeasible for a large number of channels. The matrix $\mathbf{\Sigma}_{\hat{\mathbf{c}}}$ is small enough that it is tractably invertible. We also make a frequency-domain approximation to the mean difference in the output of the channels with and without the feature present:

$$\hat{\mathbf{f}}^{\mathrm{T}}\mathbf{u}_i \approx \int\limits_{\mathbb{S}^{n-1}} G_{\mathrm{chan},1,1,i}(\vec{\Theta}, E^W(\vec{\Theta})/\alpha; F)\, d\vec{\Theta} \tag{5.42}$$

$$G_{\mathrm{chan},1,1,i}(\vec{\Theta}, \gamma; F) \triangleq \int\limits_0^{\varrho_{\max}(\vec{\Theta})} \frac{\gamma J(\varrho, \vec{\Theta})}{\gamma J(\varrho, \vec{\Theta}) + R(\varrho, \vec{\Theta})} F(\varrho, \vec{\Theta})\, U_i(\varrho, \vec{\Theta}) \varrho^{n-1}\, d\varrho. \tag{5.43}$$

This again allows for pre-computed tables of $G_{\text{chan},1,1,i}$, although only N_c of them. These tables, though, are only valid for one choice of feature spectrum $F(\vec{\nu})$, and must be recomputed for a different feature. We finally compute the SNR using the approximations in (5.40) and (5.42):

$$\text{SNR}^2_{\text{C}-\text{PW}} \approx (\mathbf{U}\hat{\mathbf{f}})^{\text{T}} \boldsymbol{\Sigma}_{\mathbf{c}}^{-1} (\mathbf{U}\hat{\mathbf{f}}).$$

Finally, we consider the non-prewhitened, channelized observer of $\hat{\mathbf{c}} = \mathbf{U}\hat{\mathbf{x}}$, which we will denote C-NPW. The C-NPW observer statistic is

$$t_{\text{C}-\text{NPW}}(\hat{\mathbf{x}}) = (\mathbf{U}\hat{\mathbf{f}})^{\text{T}} \hat{\mathbf{x}},$$

and the SNR of this observer is

$$\text{SNR}^2_{\text{C}-\text{PW}} = \frac{((\mathbf{U}\hat{\mathbf{f}})^{\text{T}} (\mathbf{U}\hat{\mathbf{f}}))^2}{(\mathbf{U}\hat{\mathbf{f}})^{\text{T}} \boldsymbol{\Sigma}_{\hat{\mathbf{c}}} (\mathbf{U}\hat{\mathbf{f}})}.$$

This SNR can be approximated in the frequency domain using the same methods as the C-PW observer. Table 5.2 is a summary of all of the linear observers we have considered, listing the observer statistic, the SNR, and the corresponding frequency domain approximation of the SNR.

5.3.4 Evaluation of observers

To evaluate our above frequency-domain methods for the fast computation of the SNR of the observers we consider, we compared the SNR of the observers computed by several methods for a scenario where the observer must distinguish whether or not a known lesion is present in the reconstruction of an image. The system considered for the test was a simulated 2D fan-beam geometry with one slice of an XCAT phantom used as the object present. In an edgeless region near the center of the phantom, we added a small feature represented by a Gaussian bump with a FWHM of 3 pixels and an amplitude chosen so that at the resulting sinogram noise level, the ideal S-OPW observer had an AUC of approximately 0.991. Figure 5.7 shows, for this scenario, the noiseless images with and without this feature present, and one sample reconstruction with and without the feature.

For each of the considered observers, we computed the SNR for each of these methods:

1. We computed the SNR given by the matrix-vector product form (the third

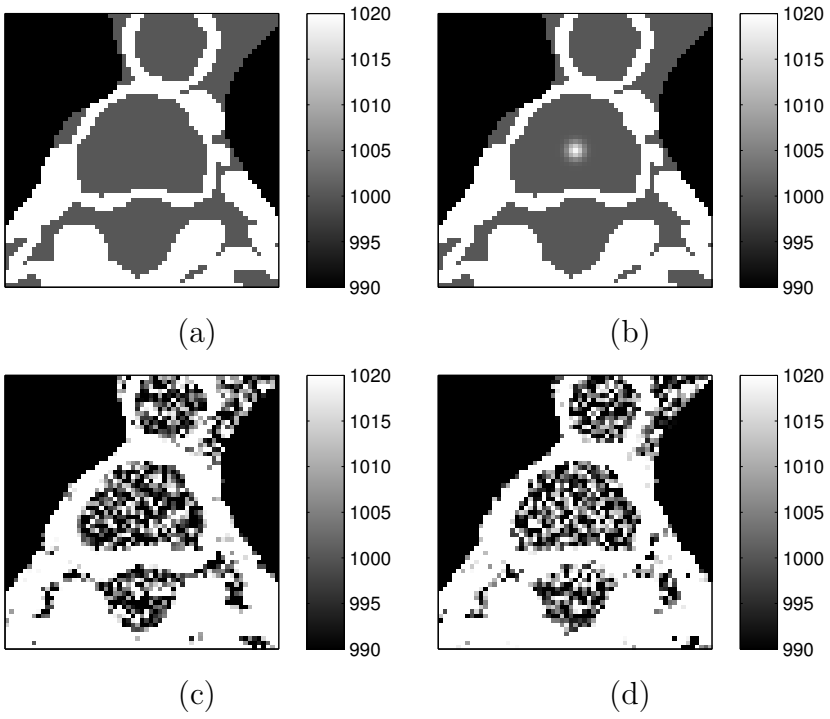


Figure 5.7: Segment of 2D XCAT phantom considered for observer evaluation. Scale in Hounsfield units. (a) Noiseless phantom without feature; (b) Noiseless phantom with feature; (c) Sample reconstruction without feature; (d) Sample reconstruction with feature.

column of Table 5.2). We reconstructed 100 images from simulated noisy scans of the phantom without a lesion present, and 100 images from scans of the phantom with the feature. For theoretical SNR computations that require an image covariance $\mathbf{\Sigma}_{\hat{\mathbf{x}}}$, we compute this empirically from these 100 reconstructed images.

2. We used the DFT-based local frequency response (2.54) to find H_j^W and numerically integrated our frequency-domain approximation a of the SNR.

3. We numerically integrated the same frequency-domain representations of the SNR using the approximation (3.25).

Table 5.3 summarizes the results of the above methods.

5.4 Tube current modulation using model observers

In this section, we apply the methods of the previous section for quickly predicting the performance of our model observers toward optimizing a tube current modulation

Observer	Statistic	$\mathbf{SNR}^2$	Freq. Domain Approximation
S-OPW	$t(\mathbf{y}) = \mathbf{f}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y}$	$\mathbf{f}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{A}\mathbf{f}$	$\int \lvert F(\vec{\nu})\rvert^2 H^{\Sigma_{\mathbf{y}}^{-1}}(\vec{\nu})\ d\vec{\nu}$
S-PW	$t(\mathbf{y}) = \mathbf{f}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{y}$	$\mathbf{f}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\hat{\mathbf{W}}\mathbf{A}\mathbf{f}$	$\int \lvert F(\vec{\nu})\rvert^2 \frac{H^W(\vec{\nu})^2}{H^{\hat{W}}(\vec{\nu})}\ d\vec{\nu}$
I-PW	$t(\hat{\mathbf{x}}) = \hat{\mathbf{f}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{\mathbf{x}}}^{-1}\hat{\mathbf{x}}$	$\hat{\mathbf{f}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{\mathbf{x}}}^{-1}\hat{\mathbf{f}}$	$\int \lvert F(\vec{\nu})\rvert^2 \frac{H^W(\vec{\nu})^2}{H^{\hat{W}}(\vec{\nu})}\ d\vec{\nu}$
I-NPW	$t(\hat{\mathbf{x}}) = \hat{\mathbf{f}}^{\mathrm{T}}\hat{\mathbf{x}}$	$\frac{(\hat{\mathbf{f}}^{\mathrm{T}}\hat{\mathbf{f}})^2}{\hat{\mathbf{f}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{\mathbf{x}}}\hat{\mathbf{f}}}$	$\frac{\left(\int \frac{\lvert F(\vec{\nu})\rvert^2 H_j^W(\vec{\nu})^2}{(H_j^W(\vec{\nu})+\alpha R(\vec{\nu}))^2}\ d\vec{\nu}\right)^2}{\int \frac{\lvert F(\vec{\nu})\rvert^2 H_j^W(\vec{\nu})^2 H_j^{\hat{W}}(\vec{\nu})}{(H_j^W(\vec{\nu})+\alpha R(\vec{\nu}))^4}\ d\vec{\nu}}$
UR-NPW	$t(\hat{\mathbf{x}}) = \mathbf{f}^{\mathrm{T}}\hat{\mathbf{x}}$	$\frac{(\mathbf{f}^{\mathrm{T}}\hat{\mathbf{f}})^2}{\mathbf{f}^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{\mathbf{x}}}\mathbf{f}}$	$\frac{\left(\int \lvert F(\vec{\nu})\rvert^2 \frac{H^W(\vec{\nu})}{H^W(\vec{\nu})+\alpha R(\vec{\nu})}\ d\vec{\nu}\right)^2}{\int \lvert F(\vec{\nu})\rvert^2 \frac{H^{\hat{W}}(\vec{\nu})}{(H^W(\vec{\nu})+\alpha R(\vec{\nu}))^2}\ d\vec{\nu}}$
C-PW	$t(\hat{\mathbf{x}}) = (\mathbf{U}\hat{\mathbf{f}})^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{\mathbf{c}}}^{-1}\mathbf{U}\hat{\mathbf{x}}$ $\boldsymbol{\Sigma}_{\hat{\mathbf{c}}} = \mathbf{U}\boldsymbol{\Sigma}_{\hat{\mathbf{x}}}\mathbf{U}^{\mathrm{T}}$	$(\mathbf{U}\hat{\mathbf{f}})^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{\mathbf{c}}}^{-1}(\mathbf{U}\hat{\mathbf{f}})$	$[\mathbf{U}\hat{\mathbf{f}}]_i \approx \int F(\vec{\nu})\frac{H^W(\vec{\nu})}{H^W(\vec{\nu})+\alpha R(\vec{\nu})}U_i(\vec{\nu})\ d\vec{\nu}$ $[\boldsymbol{\Sigma}_{\hat{\mathbf{c}}}]_{ij} \approx \int \frac{H^{\hat{W}}(\vec{\nu})}{(H^W(\vec{\nu})+\alpha R(\vec{\nu}))^2}U_i(\vec{\nu})U_j(\vec{\nu})\ d\vec{\nu}$
C-NPW	$t(\hat{\mathbf{x}}) = (\mathbf{U}\hat{\mathbf{f}})^{\mathrm{T}}\mathbf{U}\hat{\mathbf{x}}$	$\frac{((\mathbf{U}\hat{\mathbf{f}})^{\mathrm{T}}(\mathbf{U}\hat{\mathbf{f}}))^2}{(\mathbf{U}\hat{\mathbf{f}})^{\mathrm{T}}\boldsymbol{\Sigma}_{\hat{\mathbf{c}}}(\mathbf{U}\hat{\mathbf{f}})}$	(See above)

Table 5.2: Summary of linear observers.

Observer	Theoretical	LFR Appx.	Fast LFR Appx.
S-OPW	11.49	11.56	12.16
I-PW	(same as S-OPW)		
I-NPW	3.99	4.98	5.31
UR-NPW	4.32	5.09	4.75

Table 5.3: Squared SNR of linear observers compared to frequency domain approximations.

scheme. As we have previously noted, sinogram observers and the pre-whitened image observer are both impractical both in terms of their non-realistic modeling of a human observer and their simplistic suggestions for a tube current modulation scheme. Because of this, we will focus our attention on the non-prewhitened observers (I-NPW and UR-NPW). The I-NPW observer is the same as the "statistical decision theory model" from [36]. In their study, this model resulted in the highest rank correlation of detectability between human observers and model observers, given a task of detecting nylon beads in X-ray images.

5.4.1 Methods

For our two non-prewhitened image observers, I-NPW and UR-NPW, we use the approximations (5.28), (5.29), and (5.31) to suggest a tube current modulation scheme. We will analyze the case of optimizing observer performance at the center voxel of a 2D fan-beam reconstruction with the statistical weighting matrix chosen correctly so that $\mathbf{W} = \hat{\mathbf{W}}$.

With a regularization parameter α small or large enough such that the values of $G_{\text{gauss},p,q,n}$ fall within the ranges where they are well-approximated by a constant, $\gamma^{-2/3}$, or γ^{-1}, we can derive a tube current modulation scheme in closed form. First, we examine the case where α is small, such that we approximate $G_{\text{gauss},p,q,n}$ as a constant. For this case, the I-NPW performance integral (5.28) simplifies to

$$\text{SNR}^2_{\text{I-NPW}} \approx \frac{C}{\int_0^{2\pi} \frac{1}{E_j^W(\Phi)}\, d\Phi} \approx \frac{C}{\frac{2\pi}{V}\sum_{v=1}^{V} \frac{A_v}{N_v}}, \tag{5.44}$$

for some constant C. As previously, N_v and A_v represent the incident number of photons for the center pixel at view v, and A_v represents the attenuation of the ray passing through the center pixel from view v. Designing the tube current modulation is performed by the optimization

$$\mathbf{N}^* = \text{argmax}_{\mathbf{N}}\, \text{SNR}^2_{\text{I-NPW}}, \quad \text{s.t.} \sum_{v=1}^{V} N_v = N_0, \tag{5.45}$$

Maximizing the SNR (5.45) can be done by equivalently minimizing $\sum_v A_v/N_v$, which is exactly the same minimization in [22] for FBP-reconstructed images, and therefore the optimal incident radiation is given by $N_v \propto \sqrt{A_v}$. This is a useful check of the result if we think of the limit of statistical reconstruction as $\alpha \to 0$ to be similar to FBP reconstruction. The simplification (5.44) also applies in the same way for the

UR-NPW observer if we simplify (5.29) and (5.31) for small α, giving us the same tube current modulation result.

In the case where α is large, we will replace $G_{\text{gauss},1,1,2}$ and $G_{\text{gauss},2,2,2}$ with $\gamma^{-2/3}$, and $G_{\text{gauss},1,2,2}$ and $G_{\text{gauss},3,4,2}$ with γ^{-1}. The E_j^W terms in the denominator of (5.28) cancel out, turning the denominator into a constant, and the numerator becomes an integral of $E_j^W(\Phi)^{2/3}$. This simplifies the I-NPW performance integral into

$$\text{SNR}^2_{\text{I-NPW}} \approx C \left(\int_0^{2\pi} E_j^W(\Phi)^{2/3} \right)^2 \approx C \left(\frac{2\pi}{V} \sum_{v=1}^{V} \left(\frac{N_v}{A_v} \right)^{2/3} \right)^2, \tag{5.46}$$

where C is a constant (not the same as in (5.44)) that can be dropped. Maximizing (5.46) is then equivalent to maximizing $\sum_v (N_v/A_v)^{2/3}$. As a tube current modulation scheme, to solve the optimization

$$\mathbf{N}^* = \text{argmax}_{\mathbf{N}} \sum_v \left(\frac{N_v}{A_v} \right)^{2/3} \quad \text{s.t.} \quad \sum_v N_v = N_0, \tag{5.47}$$

we will use the Lagrangian

$$\Lambda(\mathbf{N}, \lambda) \triangleq \sum_v \left(\frac{N_v}{A_v} \right)^{2/3} + \lambda \left(N_0 - \sum_v N_v \right),$$

which has partial derivatives with respect to N_i

$$\frac{\partial \Lambda}{\partial N_i} = \frac{2}{3} A_i^{-2/3} N_i^{-1/3} - \lambda.$$

Setting this equal to zero and solving for N_i gives

$$N_i^* = c A_i^{-2}, \tag{5.48}$$

where $c = (3\lambda/2)^{-3}$. This result, that the optimal tube current should be proportional to the inverse of the square of the attenuation, is surprising but possibly understandable. It is very contrary to the small-α or FBP result in that it suggests more incident radiation through less-attenuating views instead of through more-attenuating views; the inverse being squared also causes even more extreme swings in tube current than in the square-root-attenuation tube current modulation scheme, which already suggested a modulation more extreme than X-ray tubes can currently carry out. With a regularization parameter large enough that we operate in this regime, it is possible

that we are considering reconstructed images where the regularization is essentially doing all of the work. Regardless, a choice of α so large that we are operating in this regime produces images that are both not clinically relevant and have a lower SNR than images with a correctly chosen α.

Most choices of regularization parameter α that produce reasonable images are not in either this pathologically small or pathologically large region. The combination of (5.44) and (5.48) seems to indicate that optimal tube current $\mathbf{N}^*$ for reconstructions in the middle fall into some interpolated middle ground (although not necessarily $N_v^* \propto A_v^p$ for some power p). In this case $G_{\text{gauss},p,q,2}$ is not easily replaced by a simple closed-form representation. We turn to numerical methods to design tube current modulation. We will restrict our attention to optimizing just three parameters jointly: the regularization parameter α and the amplitude and phase of a sinusoidal tube current modulation scheme. This both greatly reduces the dimensionality of the problem and forces us to only consider tube current modulation that would be more realizable in practice than the large current range of $N_v \propto A_v^p$. We represent possible continuous sinusoidal tube current modulation schemes as

$$N_v(A, \theta) = N_0(1 + A\cos(2\phi_v - \theta)), \tag{5.49}$$

where N_v is the number of photons in view v; ϕ_v is the angle of view v and N_0 is the number of photons per view in a baseline no-modulation comparison. The amplitude A is restricted to the range $[0, 1]$, since above 1 results in a negative N; the phase θ we will consider to be in the range $[0, 2\pi]$.

Under this restricted form, we jointly optimize the regularization strength α and the tube current modulation parameters by solving

$$\alpha^*, A^*, \theta^* = \text{argmax}_{\alpha,A,\theta}\, \text{SNR}^2_{\text{I-NPW}}(\alpha, A, \theta), \tag{5.50}$$

using the SNR approximation in (5.28) and tube current modulation of (5.49).

5.4.2 Results

With the fast computation method (5.28), it is reasonable, at least for our analysis, to find the solution of (5.50) by exhaustive search. Figure 5.8 shows the predicted squared SNR for the I-NPW given the simulated feature detection problem described in Section 5.3.4 using a wide range of regularization parameters α and all possible sinusoidal tube current modulation schemes. Each of the 32 panels represents one

value of α, in the first (top-left) panel $\alpha = 2^{-2}$ and the value of α increases by a factor of $2^{1/2}$ in each successive panel, until $\alpha = 2^{27/2}$ in the last (bottom-right) panel. In one panel, the amplitude A varies along the y-axis, from 0 (no modulation) at the bottom, to 1 (modulation to the point of N_v becoming 0 at 2 points). The phase θ varies along the x-axis from 0 to 2π.

Figure 5.9(a) shows the panel for the smallest α, and Figure 5.9(b) shows the tube current modulation scheme from (5.49) representing the amplitude and phase combination ($A = 0.52$, $\theta = \pi$) from this small-α plane corresponding to the highest predicted SNR ($\text{SNR}^2 \approx 4.56$). Figure 5.12(a) shows a small section of a simulated reconstruction using this tube current modulation scheme and no feature present, and Figure 5.12(b) shows another reconstruction realization with the feature present. Figure 5.12(c) shows the full simulated reconstruction of Figure 5.12(a). Figure 5.12(d) shows the PSF of the reconstruction.

Figure 5.11(a) shows the panel for the largest α, and Figure 5.11(b) shows the tube current modulation scheme from (5.49) representing the amplitude and phase combination ($A = 0.98$, $\theta = 0.04 \cdot \pi$) from this large-α plane corresponding to the highest predicted SNR ($\text{SNR}^2 \approx 4.49$). Figure 5.14(a) shows a full simulated reconstruction using this tube current modulation with no feature present. Figure 5.14(b) shows the PSF of the reconstruction.

Figure 5.10(a) shows the panel from Figure 5.8 for the value of α ($\alpha = 2^{15/2}$) that contained the largest predicted SNR ($\text{SNR}^2 \approx 7.85$) for the entire volume. In this plane, the amplitude and phase that corresponded to this largest SNR were $A = 0.44$, $\theta = 0$; the resulting proposed modulation scheme is shown in Figure 5.10(b). Figure 5.13(a) shows a small section of a simulated reconstruction using this tube current modulation scheme and no feature present, and Figure 5.13(b) shows another reconstruction realization with the feature present. Figure 5.13(c) shows the full simulated reconstruction of Figure 5.13(a). Figure 5.13(d) shows the PSF of the reconstruction.

We also consider solving (5.50) using a simple gradient ascent algorithm. The derivative of (5.28) with respect to regularization parameter α, modulation amplitude A, and modulation phase θ is easy to compute in a similar form using table lookups given the derivative of $G_{\text{gauss},n,p,q}$ shown in (C.5). For the feature detection problem of Section 5.3.4, this gradient ascent finds an optimal regularization and modulation of $\alpha \approx 209 \approx 2^{15.4/2}$, $A = 0.51$, $\theta = 0.015 \cdot \pi$, which is very close to the results shown in Figure 5.10.

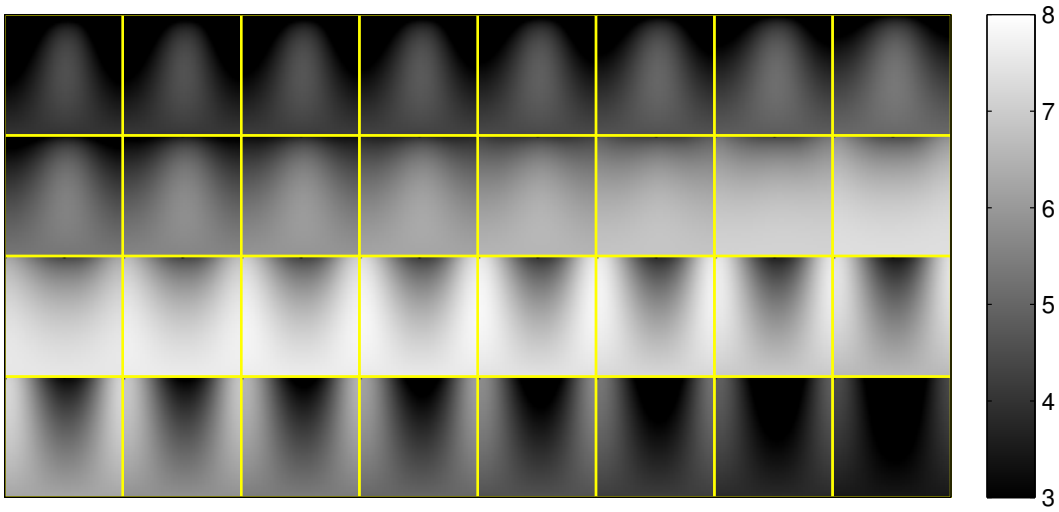
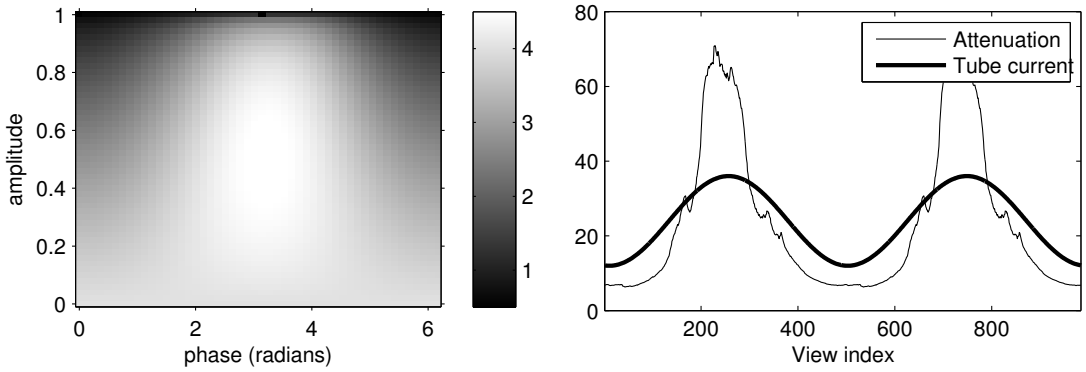


Figure 5.8: Predicted squared SNR of the I-NPW observer by α, tube current modulation amplitude and phase. Each of the 32 panels represents one value of α, in the first (top-left) panel $\alpha = 2^{-2}$ and the value of α increases by a factor of $2^{1/2}$ in each successive panel, until $\alpha = 2^{27/2}$ in the last (bottom-right) panel. In one panel, the amplitude A varies along the y-axis, from 0 (no modulation) at the bottom, to 1 (modulation to the point of $N(\phi)$ becoming 0 at 2 points). The phase θ varies along the x-axis from 0 to 2π.

5.4.3 Discussion

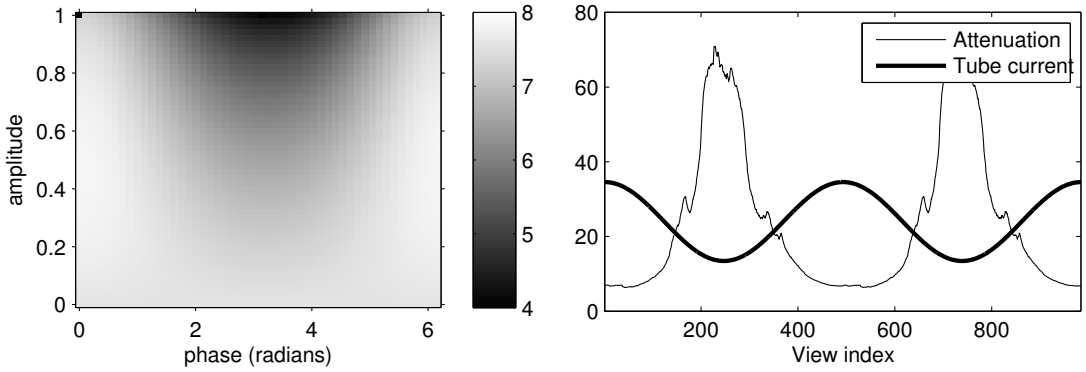
In the plane of Figure 5.9, the value of α is small enough that these reconstructions approach the FBP-like reconstruction regime, so it is not surprising that the resulting tube current modulation scheme is near the optimal square-root-attenuation scheme, as far as sinusoidal modulation allows. On the other end, for values of α high enough, as in Figure 5.11, such that we are within the regime where the SNR is approximately given by (5.46), it appears that the suggested modulation is indeed nearing the inverse-squared-attenuation modulation of (5.48). The highest point of the suggested tube current in 5.11(b) is about 100 times larger than the lowest point; for comparison, the highest squared attenuation is about 120 times larger than the smallest squared attenuation. Of course, the image produced by this TCM, shown in Figure 5.11, is not clinically useful.

It appears that the predicted optimal modulation scheme does transition smoothly between these two extremes; for example, in Figure 5.8, the rightmost panel of the second row represents $\alpha = 2^{11/2}$, for this regularization strength, no tube current modulation is perhaps best. For the highest predicted SNR in our entire search range,



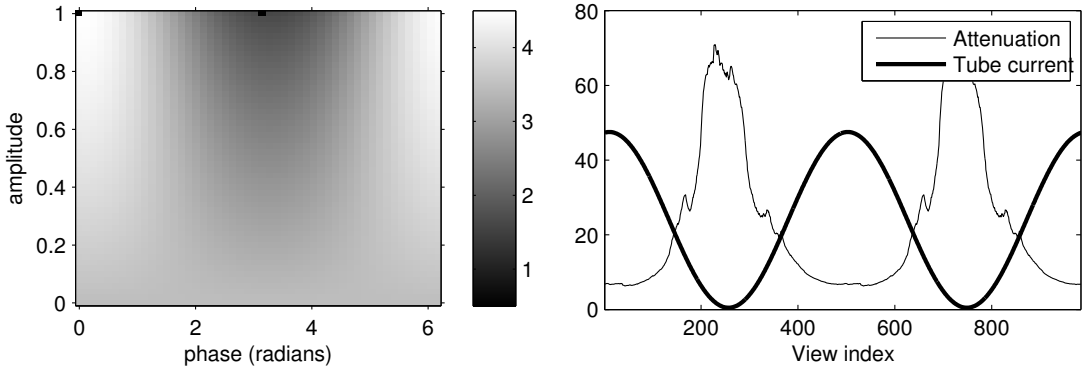
(a) (b)

Figure 5.9: (a): Plane of Figure 5.8 for small (FBP-like) α; (b): Tube current modulation corresponding to the largest value ($\mathrm{SNR}^2 = 4.56$) on this plane.



(a) (b)

Figure 5.10: (a): Plane of Figure 5.8 for α containing the largest SNR; (b): Tube current modulation corresponding to the largest value ($\mathrm{SNR}^2 = 7.85$) on this plane.



(a) (b)

Figure 5.11: (a): Plane of Figure 5.8 for large α; (b): Tube current modulation corresponding to the largest value ($\mathrm{SNR}^2 = 4.49$) on this plane.

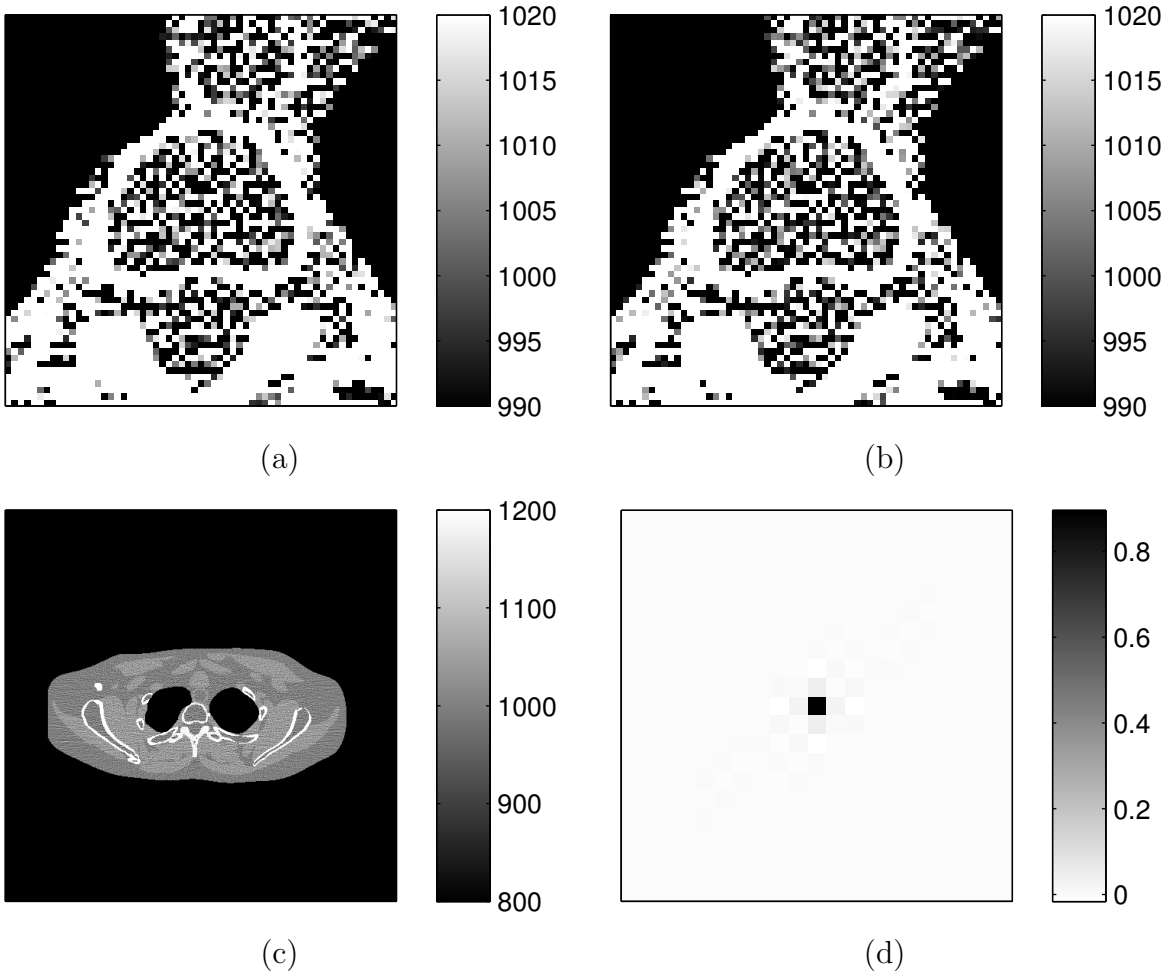


Figure 5.12: 2D phantom reconstructed using the tube current modulation of Figure 5.9(b); (a): Segment without feature, (b): Segment with feature, (c): Full image, (d): Reconstruction PSF.

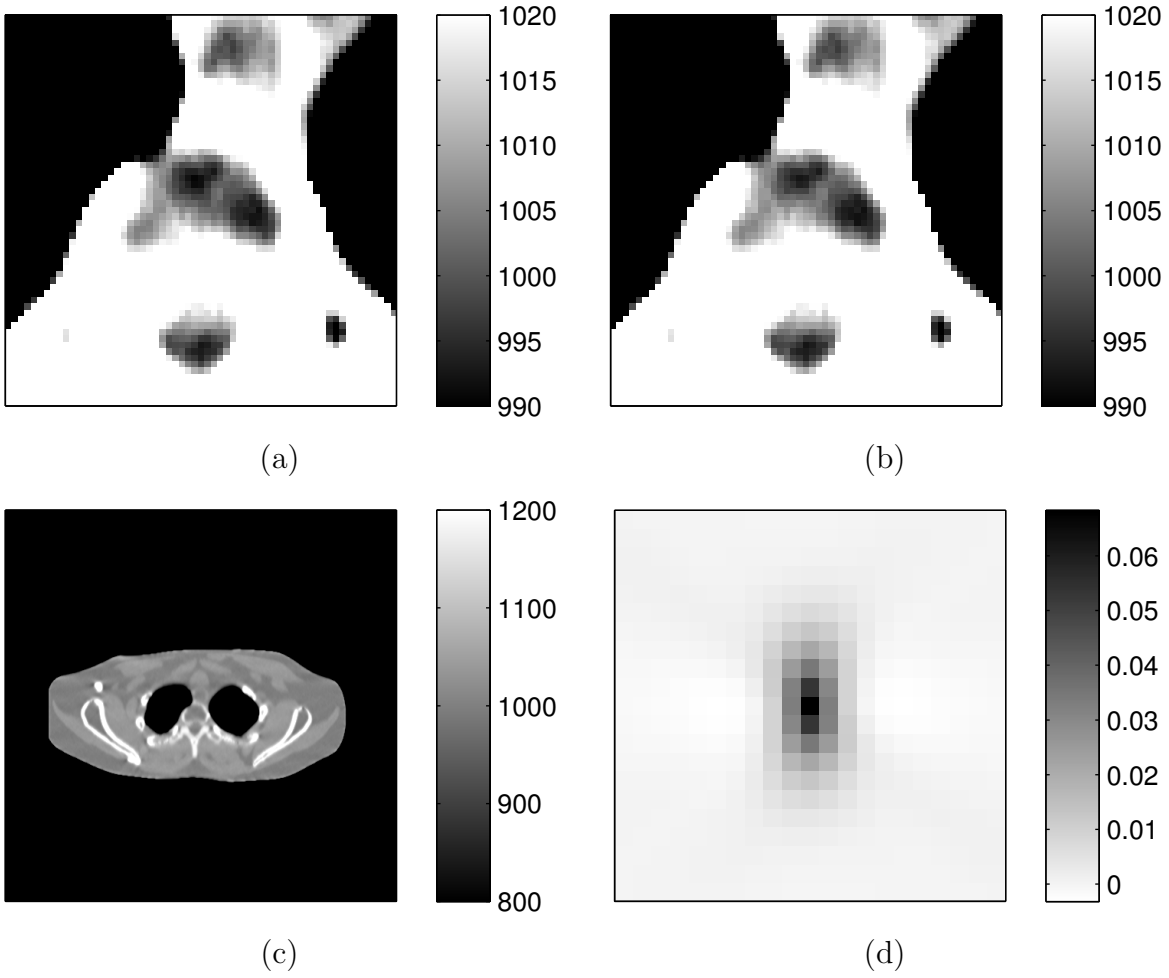
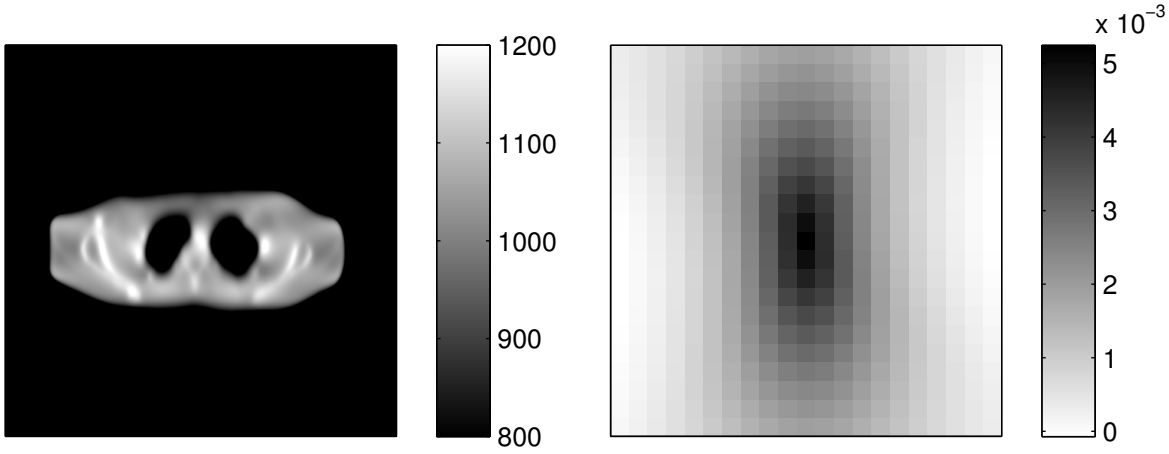


Figure 5.13: 2D phantom reconstructed using the tube current modulation of Figure 5.10(b); (a): Segment without feature, (b): Segment with feature, (c): Full image, (d): Reconstruction PSF.

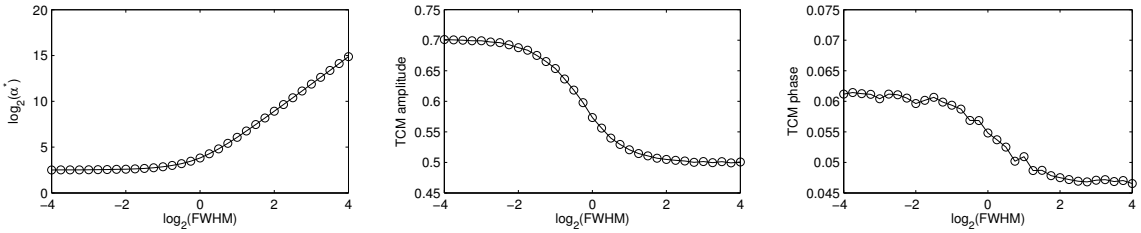
the suggested modulation is inversely related to attenuation, as opposed to ideal modulation for FBP-reconstructed images, but not inversely related to the extreme of Figure 5.11(b).

The resulting reconstruction shown in Figure 5.10, as we would predict via the SNR, does appear to have a visible increase in attenuation at the location of the feature in Figure 5.10(b) compared to Figure 5.10(a). However, this demonstrates some of the limitations of jointly designing regularization strength and tube current modulation via maximizing $\text{SNR}^2_{\text{I-NPW}}$. First, without prior knowledge of both the background image and the location and shape of the feature, the difference between the reconstructions with and without the feature would be indistinguishable. There is clearly a mismatch between the process of a human looking for a particular feature in an image with the location unknown, and the model of this process as a linear observer with a known background and known signal. In [36], the I-NPW SNR was found to correlate with human observers, but this was specific to the task of finding the presence and position of a known feature in one of 25 known locations. In the case where a feature is within some region, and the observer must find its location, we could potentially use the methods of [74]. Secondly, the resulting image is too blurry, due to over-regularization, compared to typical clinical images. This is particular to our choice of feature. If we knew beforehand in a signal detection task that we are to find a Gaussian bump with a 3 pixel FWHM, the image that is best for this task may be this blurry to match the resolution of our feature; the regularization and tube current modulation that we designed takes this into account. Figure 5.15 shows the regularization and tube current resulting from optimizing (5.50) for a range of feature full-width-half-maxima. Two features are evident from Figure 5.15(a). When the FWHM tends toward zero, converging toward a feature that is an impulse, the optimum regularization strength α^* tends toward a constant (in this case, $\alpha^* \approx 5.7$) that is particular to the base tube current N_0. When the FWHM tends toward infinity, the ideal α^* seems to be proportional to FWHM^3; the source of this cubic behavior is unknown. However, we are often interested in finding features for which the resolution is unknown. Figure 5.15(b) shows that the designed tube current modulation amplitude transitions smoothly from approximately 0.7 in the small-feature case to 0.5 in the large-feature case. Figure 5.15(c) shows that the phase does not change appreciably between the two feature sizes, for all feature sizes the phase is chosen so that the tube current is larger for views with lower attenuation and smaller for more attenuating views. In fact, the range of θ^* is small enough that the noise resulting from the table-based approximations to $\text{SNR}^2_{\text{I-NPW}}$ is visible.



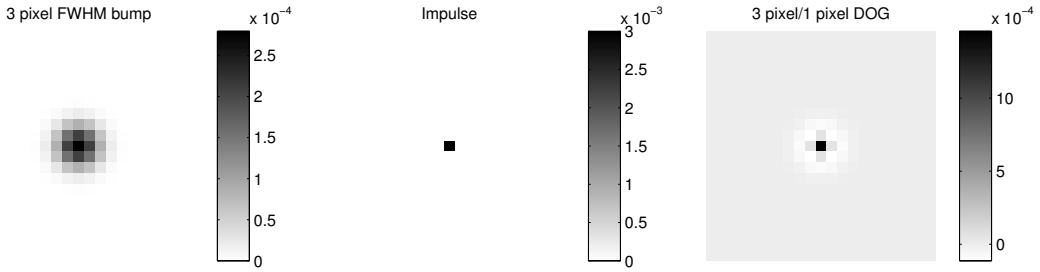
(c) (d)

Figure 5.14: 2D phantom reconstructed using the tube current modulation of Figure 5.10(b); (a): Full image, (b): Reconstruction PSF.



(a) (b) (c)

Figure 5.15: Optimized tube current modulation schemes as a function of feature FWHM. (a): regularization strength, (b): TCM amplitude, (c): TCM phase.



(a) (b)

Figure 5.16: Objects used for SNR tests. (a): 3 pixel FWHM Gaussian bump; (b): impulse; (c): difference of (a) and a 1 pixel FWHM bump.

		Traditional		Proposed, optimized for:		
		Uniform	**FBP Opt.**	**3 FWHM**	**Impulse**	**DOG**
	A	0	-0.52	0.51	0.70	0.99
	θ	—	0.02π	0.015π	0.02π	0.04π
Object	**3 FWHM**	7.63	6.78	**7.86**	7.82	7.55
	α	(a) 118	(d) 65.7	(g) 209	261	382
	Impulse	26.84	23.20	28.42	**28.57**	27.99
	α	(b) 3.25	(e) 2.17	4.87	(h) 5.73	7.67
	3,1 DOG	7.72	6.44	9.02	9.64	**10.85**
	α	(c) 2.85	(f) 1.46	5.25	6.24	(i) 7.20

Table 5.4: Predicted squared SNR of the I-NPW observer for no TCM, TCM optimized for FBP images, TCM optimized for a 3-pixel FWHM gaussian bump, and TCM optimized for an impulse. Features tested were a 3-pixel FWHM bump, an impulse, and the difference between a 3-pixel and a 1-pixel gaussian bump. Letters in parenthesis refer to corresponding PSFs in Figure 5.17. Largest SNR for each object is displayed in bold.

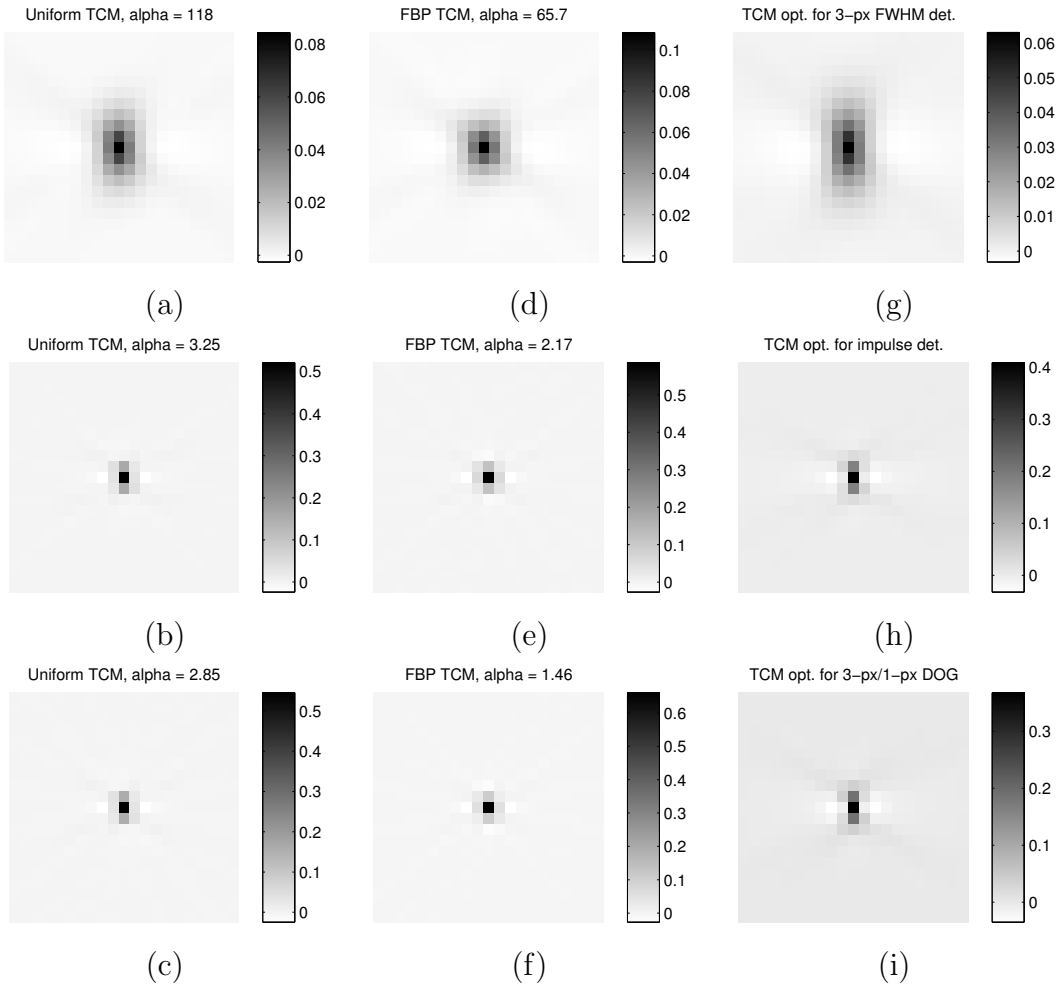


Figure 5.17: PSFs for the regularization and TCM parameters shown in Table 5.4.

Table 5.4 shows the squared SNR for the I-NPW observer for several choices of tube current modulation and regularization parameters and several test features: the Gaussian bump with a 3 pixel FWHM considered previously, a one-pixel impulse, and the difference between the 3 pixel FWHM Gaussian bump and a 1 pixel FWHM bump. These features are shown in Figure 5.16. In the first column, we measure the squared SNR for uniform (no) tube current modulation for each of these features. In the second column, we measure the squared SNR for the sinusoidal tube current modulation chosen to best match the square root of the attenuation. That is, A and θ are chosen to minimize

$$A^*, \theta^* = \text{argmin}_{A,\theta} \sum_v \frac{A_v}{1 + A\cos(2\phi_v - \theta)},$$

minimizing the variance at the center voxel of a hypothetical FBP-based reconstruction, as in (5.2), using tube current modulation restricted to be sinusoidal, as in (5.49). In the last three columns, the tube current modulation is chosen simultaneously with the regularization parameter α to optimize detection for a particular feature, the same three as those shown in Figure 5.16. In each case, the regularization parameter is optimized with the tube current modulation parameters fixed and displayed below the shown SNR. For some of the combinations of regularization and tube current modulation parameters, we show the corresponding PSF in Figure 5.17.

Table 5.4 emphasizes the importance of foreknowledge of the feature to optimizing tube current modulation for I-NPW observer detection. In each case, the TCM and regularization parameters jointly optimized for detection of a particular feature provided the highest SNR for detection of the same feature. Parameters optimized for a different feature give, in almost all cases, a higher SNR than uniform tube current modulation. Only in the case of detection of the 3-pixel FWHM bump where the TCM parameters are optimized for the difference of Gaussians object is the resulting SNR lower than no TCM. The TCM amplitude in the DOG-optimized case is very high, however, contributing to the poor performance with other features. In all cases, the traditional TCM optimized for a hypothetical FBP-based reconstruction gave a lower SNR than no tube current modulation.

The PSFs in Figure 5.17 demonstrate that, as is the case for FBP-reconstructed images, the FBP-optimized TCM (shown in Figure 5.17 (d)–(f)) results in a PSF that is more isotropic than the PSF given by uniform TCM, and our proposed TCM results in a PSF more anisotropic than uniform TCM. This is similar to the anisotropic PSF results of Section 5.2.

5.5 Variance minimization with fixed resolution

In this section, we consider optimization of TCM via minimizing the variance of a particular pixel or voxel of interest using a hypothetical regularizer designed to produce a particular local impulse response of the reconstruction. For example, for a statistical reconstruction we may choose the regularization parameter α to pick the resolution that we want in terms of FWHM and use a regularizer designed to make the LIR isotropic, as in [8, 7].

We again approximate the local impulse response of the effect of the statistical reconstruction by the filter $F_j^W(\vec{\nu})$ given in (5.19). Suppose that we *knew* the exact form of $H_j^W(\vec{\nu})$ and we *designed* a regularizer $R_{j,\text{design}}^W(\vec{\nu})$ such that the effect of reconstruction is filtering by a desired target filter $\hat{F}_{j,\text{target}}(\vec{\nu})$. In this case, we can solve (5.19) for the regularization filter:

$$\alpha R_{j,\text{design}}^W(\vec{\nu}) = H_j^W(\vec{\nu}) \frac{1 - F_{j,\text{target}}(\vec{\nu})}{F_{j,\text{target}}(\vec{\nu})}$$

If we plug this designed regularizer into the frequency-domain expression for the local noise NPS in (2.59), we find that

$$S_j(\vec{\nu}) = \frac{H_j^{\hat{W}}(\vec{\nu})}{(H_j^W(\vec{\nu}) + \alpha R_{j,\text{design}}(\vec{\nu}))^2} = F_{j,\text{design}}(\vec{\nu})^2 \frac{H_j^{\hat{W}}(\vec{\nu})}{H_j^W(\vec{\nu})^2}. \tag{5.51}$$

We then apply the approximate factorization (3.25) and integrate (5.51) as in (2.60) to find an approximate variance:

$$\begin{aligned}
\text{var}(\hat{x}_j) &\approx \int_{\left[-\frac{1}{2},\frac{1}{2}\right]^n} F_{j,\text{design}}^2(\vec{\nu}) \frac{H_j^{\hat{W}}(\vec{\nu})}{H_j^W(\vec{\nu})^2} \, d\vec{\nu} \approx \int_{\left[-\frac{1}{2},\frac{1}{2}\right]^n} \frac{F_{j,\text{design}}^2(\vec{\nu})}{J(\vec{\nu})} \frac{E_j^{\hat{W}}(\vec{\Theta})}{E_j^W(\vec{\Theta})^2} \, d\vec{\nu} \\
&= \int_{\mathbb{S}^{n-1}} \frac{E_j^{\hat{W}}(\vec{\Theta})}{E_j^W(\vec{\Theta})^2} \int_0^{\varrho_{\max}(\vec{\Theta})} \frac{F_{j,\text{design}}(\varrho, \vec{\Theta})^2}{J(\varrho, \vec{\Theta})} \varrho^{n-1} \, d\varrho \, d\vec{\Theta}
\end{aligned} \tag{5.52}$$

Alternatively, for 3DCT, using the cylindrical factorization (4.9) gives

$$\begin{aligned}
\mathrm{var}(\hat{x}_j) &\approx \int_{\left[-\frac{1}{2},\frac{1}{2}\right]^n} F_{j,\mathrm{design}}(\vec{\nu})^2 \frac{H_j^{\hat{W}}(\vec{\nu})}{H_j^W(\vec{\nu})^2}\, d\vec{\nu} \approx \int_{\left[-\frac{1}{2},\frac{1}{2}\right]^n} \frac{F_{j,\mathrm{design}}(\vec{\nu})^2}{J_{\mathrm{cyl}}(\vec{\nu})} \frac{E_{j,\mathrm{cyl}}^{\hat{W}}(\Phi)}{E_{j,\mathrm{cyl}}^W(\Phi)^2}\, d\vec{\nu} \\
&= \int_0^{2\pi} \frac{E_{j,\mathrm{cyl}}^{\hat{W}}(\Phi)}{E_{j,\mathrm{cyl}}^W(\Phi)^2} \left(\int_{-1/2}^{1/2} \int_0^{\rho_{\max}(\Phi)} \frac{F_{j,\mathrm{design}}(\vec{\nu})^2}{J_{\mathrm{cyl}}(\vec{\nu})} \rho\, d\rho\, d\nu_3 \right) d\Phi \\
&= \int_0^{2\pi} \frac{E_{j,\mathrm{cyl}}^{\hat{W}}(\Phi)}{E_{j,\mathrm{cyl}}^W(\Phi)^2} f_{j,\mathrm{cyl}}(\Phi)\, d\Phi \qquad (5.53) \\
&\stackrel{?}{=} \int_0^{2\pi} \frac{1}{E_{j,\mathrm{cyl}}^W(\Phi)} f_{j,\mathrm{cyl}}(\Phi) \qquad (5.54)
\end{aligned}$$

Where $f_{j,\mathrm{cyl}}(\Phi)$ is a function that depends on the designed local frequency response, but not the tube current:

$$f_{j,\mathrm{cyl}}(\Phi) \triangleq \int_{-1/2}^{1/2} \int_0^{\rho_{\max}(\Phi)} \frac{F_{j,\mathrm{design}}(\vec{\nu})^2}{J_{\mathrm{cyl}}(\vec{\nu})} \rho\, d\rho\, d\nu_3. \qquad (5.55)$$

The approximation (5.54) comes from (5.53) if we assume that the weighting matrix is chosen to be the ideal weighting, in which case $E_j^{\hat{W}} = E_j^W$.

As in [22], we will consider the center voxel specifically, for which $E_j^W(\Phi) = K w_\Phi$, where K is the constant defined in (5.4), and as in (5.3), w_Φ is the statistical weighting for the rays passing through the center at angles $\Phi \pm \pi/2$. In this case, (5.54) simplifies to

$$\mathrm{var}(\hat{x}_\mathrm{c}) \approx \frac{2\pi}{KV} \sum_{v=1}^{V} \frac{A_v}{N_v} f_{\mathrm{c,cyl}}(\Phi_v). \qquad (5.56)$$

The solution to the variance minimization problem,

$$\mathbf{N}^* = \mathrm{argmin}_{\mathbf{N}}\, \mathrm{var}(\hat{x}_c), \quad \text{s.t.} \sum_{v=1}^{V} N_v = N_0,$$

for the approximate $\mathrm{var}(\hat{x}_c)$ in (5.56), is:

$$N_v^* \propto \sqrt{A_v f_{\mathrm{c,cyl}}(\Phi_v)}. \qquad (5.57)$$

If the desired local impulse response $F_{j,\text{design}}(\vec{\nu})$ is cylindrically isotropic, i.e., independent of Φ, then (5.57) approximately becomes

$$N_v^* \propto \sqrt{A_v}, \tag{5.58}$$

since $J_{\text{cyl}}(\vec{\nu})$ is itself nearly isotropic, making $f_{\text{c,cyl}}(\Phi)$ no longer dependent on Φ. This is the same result for optimal tube current modulation presented for FBP-reconstructed images in [22], suggesting that that tube current modulation method, at least from the perspective of minimizing the variance with fixed resolution, is also suitable for statistically reconstructed images. However, (5.57) generalizes this to the option of non-isotropic resolution. We can interpret $f_{j,\text{cyl}}(\Phi)$ as roughly (ignoring J_{cyl}) the dispersion of the desired frequency response on the plane of a fixed Φ. If we want a local impulse response that is more concentrated along one axis, its corresponding local frequency response will have a higher dispersion, and (5.57) suggests increasing N_v^* in the views perpendicular to that axis. This makes some intuitive sense, and (5.57) provides a quantitative result for the tube current modulation as a function of the desired LFR dispersion.

5.6 Discussion

Tube current modulation for statistically reconstructed CT images is a more difficult problem than for FBP-reconstructed images. This is because for FBP, modulation of the tube current affects the image quality mostly through the noise variance, and optimization can be done by minimizing this noise. For statistical images, though, TCM affects both the variance and resolution of the resulting image, and the tradeoff must be considered when optimizing tube current. The results that we have found for tube current modulation can be divided into two broad classes, largely defined by whether or not a feature to be detected is known.

In Section 5.2, we investigated optimizing TCM by introducing resolution metrics that we can predict quickly and simultaneously optimizing for noise variance and these resolution metrics. In Section 5.4, we optimized TCM for the task of algorithmically detecting the presence of an *a priori* known feature in a known background, using the fast observer performance predictions developed in Section 5.3. In both of these formulations, optimizing for these metrics causes the tube current to be higher for less attenuating views and lower for more attenuating views, which is counter to the optimum TCM for FBP images. These two formulations are more connected than

they would seem at first. The CRC resolution metric in Section 5.2 is equivalent to maximizing the performance of the UR-NPW observer used in Section 5.3 if the known feature to be detected is an impulse. The UR-NPW observer, in turn, is very similar to the I-NPW observer.

The results of Section 5.3.3.2 suggest that, for the problem of detecting whether or not an impulse at a known location is present in a known background, the ideal photon allocation is all in one beam passing through the impulse's location in the least attenuating direction. In both Section 5.2 and Section 5.3, the resulting TCM is more like a one-view radiograph or tomosynthesis, a method for very limited-angle tomography [43], than traditional CT. Features in images reconstructed via tomosynthesis are more identifiable in the plane perpendicular to the rays corresponding to the views used, but the resolution along the rays is very limited. This is consistent with the PSF shown in Figure 5.4. In the horizontal direction, which is perpendicular to the direction of the incoming X-rays (the peak in Figure 5.3) the PSF is sharp; in the vertical direction, the PSF is not.

In practice, however, it is not a common scenario that we have the prior knowledge required to use either of these previous methods. If the feature is unknown, radiologists generally would be unlikely to prefer an anisotropic PSF that distorts shapes. In Section 5.5, we found ATCM methods for the hypothetical case of statistical reconstruction where the local impulse response of the reconstruction is fixed by the regularizer. In this case, we are not optimizing for detectability of a known feature, either explicitly (as in Section 5.4) or implicitly (as in Section 5.2). Because we no longer control the resolution, detectability of any feature is generally maximized using the tube current modulation that minimizes the reconstructed variance, and this is achieved, as in the FBP case of [22], with tube current proportional to the square root of the attenuation (when an isotropic PSF is desired).

CHAPTER VI

Conclusion

6.1 Summary

In this work, we have investigated a local frequency response approximation to $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$, the Hessian of the data-fit term in a statistical CT image reconstruction, that is separable into a data-independent term and a data-dependent term that has a lower dimensionality. This LFR approximation applies to the Hessian with arbitrary CT geometries. We have made further approximations to this separable form that are particular to 2D fan-beam, 3D axial, and 3D helical CT that reduce the data-dependent term to a function of only one dimension. We have shown that this LFR approximation is largely accurate for 3D helical CT, except in regions near the frequency-domain axis acorresponding to the axis around which the X-ray source rotates. We have shown that the LFR approximation, when applied to the expression for the local noise power spectrum, closely matches the local NPS derived from the DFT of the local impulse response of the Hessian $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$.

We have used the LFR approximation in the integral of the noise power spectrum to produce variance predictions for a statistically reconstructed CT image in a much faster time for 3D axial and helical geometries than any previous method. These variance predictions are very accurate when compared to the frequency-domain variance predictions produced by the DFT of the LIR of $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$, and mostly accurate compared to the empirical variance of an ensemble of reconstructions, depending on the particular reconstruction methods used. Our variance predictions are very accurate for quadratically regularized reconstructions, and accurate for non-quadratically regularized reconstructions in regions not near edges.

Finally, we used the LFR approximation, along with the variance predictions derived from it, to produce tube current modulation schemes specific to statistically reconstructed images. We have found results particular to this model-based framework

that suggest a tube current modulation design that may improve the detectability of certain features while simultaneously reducing dose. What form the results take depend on the choice of observer, but for many model observers often considered in practice to model human observers, we have derived a corresponding tube current modulation scheme.

6.2 Future Work

6.2.1 Verification of LFR approximation for broader geometries

In Chapter III, we derive an approximation to the local frequency response of $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$ for arbitrary CT geometries and compare the result of this approximation to the LFR of a simulated helical CT geometry. While the simulated helical CT geometry accounts for a wide range of CT geometries used clinically, there are other geometries for which we have not verified the applicability of the approximation. For example, CT using sparse views would disrupt the approximation used in (3.6) to replace a sum over views as an integral. We have also not verified the approximation for limited-angle methods such as tomosynthesis, although this is likely to be less harmful to the result as sparse views. For other geometries like those provided by a C-arm, in which the source trajectories can have a much wider range than the more fixed helical geometries, our methods may be more difficult to verify. Future work could verify the use of our LFR approximation for variance prediction by finding empirical variance maps using real sinograms from geometries beyond those we specifically considered.

6.2.2 Extension of methods to edge-preserving regularization

In Section 2.5.1.3, we use the assumption commonly held in prior work that the frequency-domain representation of the regularizer in (2.57) holds, even near edges in the image. As mentioned previously, this is not true both because the assumption of spatial shift-invariance and the approximation of the regularizer hessian as the closed form matrix $\mathbf{P}$ break down. As a result, our methods work only away from edges in the image. For some applications, this is undesirable. Potential methods for fixing this limitation in future work could be finding approximations for the exact reconstruction distribution in [20] or adapting the methods of [47], used for statistically reconstructed PET images.

6.2.3 Other applications of detectability indices

In Section 5.4, we demonstrate the use of fast approximations to model observer performance for joint optimization of regularization strength and sinusoidal tube current modulation. This is a fairly specific usage of the fast approximations in Section 5.3, but is also useful as an example of potentially broader usage of these detectability measures. A widening of the scope of Section 5.4 to more general tube current modulation schemes is a potential direction of future work. Alternatively, we could consider other possible domains of tube current modulation that more accurately capture the capabilities of specific scanners.

We could also consider the joint optimization of tube current modulation and more generalized regularization design. The tube current modulation scheme that results from (5.12), that has the PSF shown in Figure 5.4, is "optimized" using a particular regularizer $R(\vec{x})$ and regularization strength in mind. It is possible that by also considering the optimization over both the regularization parameter α and the relative per-direction penalties of the regularization (labeled r_d in (2.27)) we could better reduce the dose while making the resulting PSF less pathological. The tube current modulation found from maximizing (5.28) considers joint optimization over α, but again does not change the relative strength r_d of the regularization directions. Since tube current modulation changes the way in which the resulting PSF is anisotropic, regularization design is a strong candidate for a potential way to even out the effect of tube current. In the other extreme, however, where we have full control of the regularization for each voxel of interest, we find ourselves in the situation of Section 5.5.

One other potential use for our observability approximations is a direct statistical usage. For example, if a human observer is looking for a known feature in an image, we may consider providing an "AUC map" for detection of the feature similar to the variance maps we show in Chapter IV.

There are also likely other possible uses for fast prediction of observer performance measures that we have not considered.

6.2.4 Extension to unknown-background or unknown-location tasks

The derivations in Section 5.3 for observer performance are specific to an observer in the task of finding a known signal at a known location in known background. Although this is correlated to human observer performance for detecting a signal at an unknown location, this correlation is specific to a task where the potential

unknown locations are very spatially separated, so that their noise is nearly uncorrelated. Yendiki [74] provides potential methods for extending our results to fast approximations for performance in the task of localizing a feature where potential locations do not necessarily have uncorrelated noise.

6.2.5 Using SNR prediction for dose reduction

Optimizing tube current modulation by posing it as a problem of maximizing SNR, as we have done in (5.50), does not directly apply to reducing X-ray dose, as it does in [22]. In FBP-based reconstruction, a scheme to modulate tube current that decreases variance while maintaining the same average tube current can be easily transformed into a method for dose reduction while maintaining equal feature detectability. However, in statistical reconstruction the relationship is not as simple. A modification of (5.50) that would be more ideal for reducing dose would be:

$$\alpha^*, A^*, \theta^*, N_0^* = \operatorname{argmin} N_0 \quad \text{s.t.} \quad \mathrm{SNR}^2_{\mathrm{I-NPW}}(\alpha, A, \theta, N_0) \geq \mathrm{SNR}^2_{\mathrm{min}}. \tag{6.1}$$

That is, we wish to find the minimum average tube current N_0 such that we can find a potential tube current modulation that maintains an SNR above some minimum.

We have at least one simple method to transform an SNR increase into a dose reduction, even if it is suboptimal compared to the result of (6.1). If we maintain the same TCM method but scale N_0, the SNR scales as:

$$\mathrm{SNR}^2_{\mathrm{I-NPW}}(\alpha, A, \theta, c\, N_0) = c \cdot \mathrm{SNR}^2_{\mathrm{I-NPW}}(\alpha, A, \theta, N_0). \tag{6.2}$$

We can achieve a minimum SNR by reducing dose by choosing $c = \mathrm{SNR}^2_{\mathrm{min}}/\mathrm{SNR}^2_{\mathrm{I-NPW}}(\alpha, A, \theta, N_0)$ and then using an average current level of cN_0. The higher an $\mathrm{SNR}^2_{\mathrm{I-NPW}}$ we can find, the smaller c becomes and the lower the dose becomes. While this may not be the optimal solution to (6.1), it is *a* conversion from an SNR improvement into a dose reduction.

Another consideration here is that N_0, the average tube current, is not directly proportional to the X-ray dose absorbed by a patient. We could consider using more advanced methods for approximating the absorbed dose, such as [11]. Using these advanced methods, we could develop formulations more advanced than (6.1) for tube current modulation, possibly toward minimizing metrics such as potential patient harm (as some organs are more sensitive to radiation) rather than dose.

6.2.6 Generalized single-integral form

Recall that, prior to sifting an integral over Σ, our form for the local frequency response of $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$ is given by:

$$H_j(\vec{\nu}) \approx C|\mathbf{V}| \left|(\mathcal{F}^n \hat{R})(\vec{\nu})\right|^2 \int_{\Sigma} \tilde{w}_{j,\sigma} |\tilde{\mathbf{B}}_j| \delta(\vec{\theta}_j(\sigma)^{\mathrm{T}} \mathbf{V}^{-\mathrm{T}} \vec{\nu})\, d\sigma, \tag{6.3}$$

that, analogously to (3.25), we can factor into two terms, one of which remains with this integral un-sifted:

$$\begin{aligned} H_j(\vec{\nu}) &\approx J(\vec{\nu}) E_j^W(\vec{\Theta}) \\ J(\vec{\nu}) &\triangleq \frac{\left|(\mathcal{F}^n \hat{R})(\vec{\nu})\right|^2}{||\vec{\nu}||} \\ E_j^W(\vec{\Theta}) &\triangleq C|\mathbf{V}| \int_{\Sigma} \tilde{w}_{j,\sigma} |\tilde{\mathbf{B}}_j| \delta(\vec{\theta}_j(\sigma)^{\mathrm{T}} \mathbf{V}^{-\mathrm{T}} \vec{\Theta})\, d\sigma \end{aligned}$$

In section 4.3, we make approximations specific to certain 3DCT geometries that allow us to reduce the inherently $(n-1)$-dimensional integral of (4.3) into the $(n-2) = 1$-dimensional integral of (4.10). However, we speculate that, by delaying this sifting, there may be a general 1-dimensional integral form, where the integral is over Σ, that does not require the inelegant approximation (4.9). We begin by making some definitions to break $E_j^W(\vec{\Theta})$ apart further:

$$\begin{aligned} E_j^W(\vec{\Theta}) &= \int_{\Sigma} f_j^W(\sigma) \delta(\vec{v}_j(\sigma)^{\mathrm{T}} \vec{\Theta})\, d\sigma \\ f_j^W(\sigma) &\triangleq C|\mathbf{V}| \tilde{w}_{j,\sigma} |\tilde{\mathbf{B}}_j| \\ \vec{v}_j(\sigma) &\triangleq \mathbf{V}^{-1} \vec{\theta}_j(\sigma) \end{aligned} \tag{6.4}$$

In this case, we can use the first steps of (4.3), but then re-expand $E_j^{\hat{W}}$ as (6.4):

$$\begin{aligned}
\operatorname{var}(\hat{\mathbf{x}}_j) &\approx \int_{\mathbb{S}^n} \int_0^{\varrho_{\max}} \frac{J(\varrho,\vec{\Theta})E_j^{\hat{W}}(\vec{\Theta})}{(J(\varrho,\vec{\theta})E_j^{W}(\vec{\Theta}) + \alpha R(\varrho,\vec{\Theta}))^2}\, \varrho^{n-1} d\varrho d\vec{\Theta} \\
&= \alpha^{-1} \int_{\mathbb{S}^n} \frac{E_j^{\hat{W}}(\vec{\Theta})}{E_j^{W}(\vec{\Theta})} G(\alpha^{-1}E_j^{W}(\vec{\Theta}),\vec{\Theta})\, d\vec{\Theta}, \\
&= \alpha^{-1} \int_{\mathbb{S}^n} \left(\int_{\Sigma} f_j^{\hat{W}}(\sigma)\delta(\vec{v}_j(\sigma)^{\mathrm{T}}\vec{\Theta})\, d\sigma \right) \frac{G(\alpha^{-1}E_j^{W}(\vec{\Theta}),\vec{\Theta})}{E_j^{W}(\vec{\Theta})}\, d\vec{\Theta}. \qquad (6.5)
\end{aligned}$$

Now we wish to sift this impulse with respect to $\vec{\Theta}$ instead of σ:

$$\operatorname{var}(\hat{\mathbf{x}}_j) = \alpha^{-1} \int_{\Sigma} f_j^{\hat{W}}(\sigma) \int_{\mathbb{S}^n} \frac{G(\alpha^{-1}E_j^{W}(\vec{\Theta}),\vec{\Theta})}{E_j^{W}(\vec{\Theta})}\, \delta(\vec{v}_j(\sigma)^{\mathrm{T}}\vec{\Theta})\, d\vec{\Theta} d\sigma. \qquad (6.6)$$

If we define an orthonormal matrix

$$\mathbf{D} = \left[\begin{array}{c:c} \frac{\vec{v}_j(\sigma)}{||\vec{v}_j(\sigma)||} & \vec{v}_j(\sigma)^{\perp} \end{array}\right],$$

where $\vec{v}_j(\sigma)^{\perp} \in \mathbb{R}^{n\times(n-1)}$ is a basis for the subspace perpendicular to $\vec{v}_j(\sigma)$, then the inner integral of (6.6) can be written as an integral over $\mathbb{S}^{n-1}$, the sphere in one fewer dimension:

$$\begin{aligned}
\int_{\mathbb{S}^n} g(\vec{\Theta})\delta([\mathbf{D}^{\mathrm{T}}\vec{\Theta}]_1)\, d\vec{\Theta} &= \int_{\mathbb{S}^n} g(\mathbf{D}\vec{\Lambda})\delta([\vec{\Lambda}]_1)\, d\vec{\Lambda} && \left(\vec{\Lambda} = \mathbf{D}^{\mathrm{T}}\vec{\Theta}\right) \\
&= \int_{\mathbb{S}^{n-1}} g\left(\mathbf{D}\left[\begin{array}{c} 0 \\ \hdashline \vec{\Omega} \end{array}\right]\right) d\vec{\Omega} && \left(\vec{\Lambda} = \left[\begin{array}{c} \Lambda_1 \\ \hdashline \vec{\Omega} \end{array}\right]\right) \\
&= \int_{\mathbb{S}^{n-1}} g\left(\vec{v}_j(\sigma)^{\perp}\vec{\Omega}\right) d\vec{\Omega}.
\end{aligned}$$

Applying this to (6.6) gives:

$$\operatorname{var}(\hat{\mathbf{x}}_j) = \alpha^{-1} \int_{\Sigma} f_j^{\hat{W}}(\sigma) \int_{\mathbb{S}^{n-1}} \frac{G(\alpha^{-1}E_j^{W}(\vec{v}_j(\sigma)^{\perp}\vec{\Omega}),\vec{v}_j(\sigma)^{\perp}\vec{\Omega})}{E_j^{W}(\vec{v}_j(\sigma)^{\perp}\vec{\Omega})} d\vec{\Omega} d\sigma.$$

From here, we hope that this inner integral can be simplified further, but this is a topic for future work. In its current form without simplification, this is actually more

computationally intensive than (4.3), since it is also an $(n-1)$-dimensional integral, but Σ is "larger" than an additional dimension to the sphere, since in helical CT it can traverse a space larger than $[0, 2\pi]$.

6.2.7 Validation of tube current modulation with human observers

In Chapter V, we derive tube current modulation schemes that optimize feature detectability with respect to model observers. The performance of some of these model observers correlate well with the performance of human observers, but human observers are still what are used in practice when detecting a feature in medical imaging. The best standard, then, for a proposed tube current modulation combined with a regularizer, would be human performance in detecting a feature. Compared to our other potential directions for future work, validating proposed tube current modulation with human observers is more far-fetched, but it would provide a very strong justification for our methods.

APPENDICES

APPENDIX A

X-ray photon detector statistics

Compound Poisson Model While much previous work has been done in statistical reconstruction assuming that the number of measured photons has a Poisson distribution that can be measured directly (e.g. [31, 27]), X-ray CT detector elements are unable to directly count photons to measure I_j, and instead measure current created by lower-energy photons released by a scintillator. For a monoenergetic incident X-ray spectrum, [13, 69, 70] suggest a compound-Poisson distribution for the measurement of the detector caused by incoming x-ray photons. Specifically, if the number of X-ray photons incident on the detector is a Poisson random variable I, as above, and each incident photon i creates X_i lower energy photons, where

$$I \sim \text{Poisson}\left[\bar{I}\right] \tag{A.1}$$

$$X_i \sim \text{Poisson}\left[\bar{X}\right], \tag{A.2}$$

then the total number V of low-energy photons is

$$V = \sum_{i=1}^{I} X_i. \tag{A.3}$$

We can use moment-generating function for V given in [13] or the law of total expectation and the law of total variance to determine that

$$\mathbf{E}\left[V\right] = \bar{I}\bar{X} \tag{A.4}$$

$$\text{var}(V) = \bar{I}\bar{X}(1+\bar{X}). \tag{A.5}$$

If we instead hope to estimate the number of incident photons I as $V/\bar{X}$, this estimate is also a random variable Z, with

$$\begin{aligned} \mathbf{E}[Z] &= \bar{I} && \text{(A.6)} \\ \text{var}(Z) &= \bar{I}(1+\bar{X}^{-1}), && \text{(A.7)} \end{aligned}$$

which is the same mean as I, and the same variance plus a correction term of $\bar{I}/\bar{X}$. This correction term is related to the Fano factor [15], where it appears in several contexts involving charges produced secondary to high-energy photons, e.g. [48]. We hope to show that for a large enough $\bar{I}$ and $\bar{X}$, that V can be approximated by $\mathcal{N}\left(\bar{I}\bar{X}, \bar{I}\bar{X}(1+\bar{X})\right)$ and thus Z can be approximated by $\mathcal{N}\left(\bar{I}, \bar{I}(1+\bar{X}^{-1})\right)$, by investigating the similarity of the cumulants of V to the cumulants of a normal distribution.

Derivation of Compound-Poisson Cumulants The moment-generating function $M_V(t)$ of V can be found:

$$\begin{aligned} M_V(t) &= \mathbf{E}\left[e^{tV}\right] && \text{(A.8)} \\ &= \sum_i P(I=i)\mathbf{E}\left[e^{tV}|I=i\right]; && \text{(A.9)} \end{aligned}$$

noting here that by definition of V, and the known mgf of a Poisson distribution,

$$\mathbf{E}\left[e^{tV}|I=i\right] = \mathbf{E}\left[e^{tX}\right]^i = \exp\left(i\ln\mathbf{E}\left[e^{tX}\right]\right) = \exp\left(i\bar{X}(e^t-1)\right), \qquad \text{(A.10)}$$

and so

$$\begin{aligned} M_V(t) &= \sum_i P(I=i)\exp\left(i\bar{X}(e^t-1)\right) && \text{(A.11)} \\ &= \mathbf{E}_I\left[\exp\left(I\bar{X}(e^t-1)\right)\right] && \text{(A.12)} \\ &= \mathbf{E}_I\left[e^{uI}\right]\Big|_{u=\bar{X}(e^t-1)} && \text{(A.13)} \\ &= \exp\left(\bar{I}\left(\exp(\bar{X}(e^t-1)-1)\right)\right). && \text{(A.14)} \end{aligned}$$

The logarithm of the moment generating function for V,

$$h(t) \triangleq \bar{I}\left(\exp(\bar{X}(e^t-1))-1\right), \qquad \text{(A.15)}$$

can be used to find cumulants. Define $j(t) \triangleq e^t - 1$, $k(t) \triangleq \bar{X}(e^t - 1)$, so that $h(t) = \bar{I}(j \circ k)(t)$. We can then use Faà di Bruno's formula [26] to evaluate arbitrary derivatives of $(j \circ k)(t)$ at zero:

$$\begin{aligned}
(j \circ k)^{(n)}(0) &= \sum_{\pi \in \Pi(n)} j^{(|\pi|)}(k(0)) \prod_{B \in \pi} k^{(|B|)}(0) && \text{(A.16)} \\
&= \sum_{\pi \in \Pi(n)} \prod_{B \in \pi} \bar{X} && \text{(A.17)} \\
&= \sum_{\pi \in \Pi(n)} \bar{X}^{|\pi|} && \text{(A.18)} \\
&= \sum_{i=1}^{n} \left\{ {n \atop i} \right\} \bar{X}^i. && \text{(A.19)}
\end{aligned}$$

In (A.16), $\Pi(n)$ is the set of partitions of the labeled set $\{1, \ldots, n\}$, and π is one of these partitions. B are the parts in the partition π. For example, one element π of $\Pi(4)$ is the partition $\{\{1\}, \{2\}, \{3, 4\}\}$, and one element B of this π is $\{2\}$. We can then simplify this significantly by noting that $k(0) = 0$, and $|\pi| \geq 1$, and $j^{(1,2,3,\ldots)}(0) = 1$, and that $|B| \geq 1$, $k^{(1,2,3,\ldots)}(0) = \bar{X}$. Applying these values in (A.16) gives us (A.17). In (A.19) this is simplified to be in terms of Stirling numbers of the second kind.

Curiously, the n^{th} cumulant of V divided by $\bar{I}$,

$$\frac{\kappa_{V,n}}{\bar{I}} = \sum_{i=1}^{n} \left\{ {n \atop i} \right\} \bar{X}^i \tag{A.20}$$

is equal to the n^{th} moment of a Poisson distribution with mean $\bar{X}$ [32].

If we define $Z = (V - \bar{X}\bar{I})/\bar{I}^{1/2}\bar{X}$, so that Z has zero mean and near-unit variance, with the intention of approximating Z with a normal distribution, the cumulants of Z are

$$\kappa_{Z,n} = \frac{\bar{I}}{\bar{I}^{n/2}\bar{X}^n} \sum_{i=1}^{n} \left\{ {n \atop i} \right\} \bar{X}^i = \bar{I}^{1-n/2} \sum_{j=0}^{n-1} \left\{ {n \atop n-j} \right\} \bar{X}^{-j}, \quad n \geq 2. \tag{A.21}$$

We now want to see whether the series $\kappa_{Z,n}$ diverges. In the worst case, when $\bar{X} = 1$, then

$$\kappa_{Z,n} = \bar{I}^{1-n/2} B_n, \tag{A.22}$$

where B_n is the n^{th} Bell number. An asymptotic limit for the Bell numbers is given in [9] as

$$\frac{\ln B_n}{n} = \ln n - \ln\ln n - 1 + \frac{\ln\ln n}{\ln n} + \frac{1}{\ln n} + \frac{1}{2}\left(\frac{\ln\ln n}{\ln n}\right)^2 + O\left(\frac{\ln\ln n}{(\ln n)^2}\right), \quad \text{(A.23)}$$

which, as n increases, will exceed

$$-\frac{\ln \bar{I}^{-n/2}}{n} = \frac{1}{2}\ln \bar{I}, \quad \text{(A.24)}$$

which is fixed. Therefore, B_n will grow more rapidly than $\bar{I}^{1-n/2}$ can shrink, and the cumulants $\kappa_{Z,n}$ diverge. However, many of the low-order cumulants are vanishingly small, since initially $\bar{I}^{1-n/2}$ shrinks much faster than B_n grows. The effects of the high-order cumulants are yet to be investigated; for now, we assume for convenience that they don't significantly affect our gaussian approximation.

APPENDIX B

Closed-form integral of G and its properties

In this section we wish to derive a closed-form expression for the integral

$$\int \frac{x^n}{(1+x^3)^2}\,dx.$$

From the series $(1-z)^{-2} = 1 + 2z + 3z^2 + \ldots$, plugging in $z = -x^3$ and multiplying by x^n gives the formal power series

$$\frac{x^n}{(1+x^3)^2} = x^n \sum_{k=0}^{\infty} (-1)^k x^{3k}(k+1) = x^n \left(1 - 2x^3 + 3x^6 - 4x^9 + \ldots\right).$$

Integrating each term gives:

$$\begin{aligned}
\int \frac{x^n}{(1+x^3)^2}\,dx &= \sum_{k=0}^{\infty} \frac{k+1}{3k+n+1}(-1)^k x^{3k+n+1} \\
&= x^{n+1} \sum_{k=0}^{\infty} \frac{k+1}{3k+n+1}(-x^3)^k \\
&= \frac{x^{n+1}}{n+1} \sum_{k=0}^{\infty} a_k (-x^3)^k
\end{aligned}$$

where $a_k = \frac{(k+1)(n+1)}{3k+n+1}$. In general, when $a_0 = 1$ and

$$\frac{a_{k+1}}{a_k} = \frac{(k+A)(k+B)}{(k+1)(k+C)},$$

The sum of $a_k z^k$ is the hypergeometric function [44]:

$$\sum_{k=0}^{\infty} a_k z^k = {}_2F_1(A, B; C; z).$$

In our case

$$\frac{a_{k+1}}{a_k} = \frac{(k+2)(k+(n+1)/3)}{(k+1)(k+(n+4)/3)},$$

and so:

$$\int \frac{x^n}{(1+x^3)^2}\,dx = \frac{x^{n+1}}{n+1}\ {}_2F_1\left(2, \frac{n+1}{3}; \frac{n+4}{3}; -x^3\right).$$

From here it is trivial algebra to show that

$$\int_0^L \frac{\gamma x^n}{(\gamma + R_0 x^3)^2}\,dx = \frac{L^{n+1}}{\gamma(n+1)} {}_2F_1\left(2, \frac{n+1}{3}; \frac{n+4}{3}; -\frac{R_0}{\gamma}L^3\right).$$

We may wish to use the following series for ${}_2F_1$ around $z = 0$ and $z = \infty$:

$$\begin{aligned} {}_2F_1\left(2, \frac{n+1}{3}; \frac{n+4}{3}; -z^3\right) &= 1 - 2\frac{n+1}{n+4}z^3 + O(z^6) \\ &= \frac{1}{z^{n+1}}\Gamma\left(\frac{n+4}{3}\right)\Gamma\left(\frac{5-n}{3}\right) + O(z^{-6}), \qquad n \le 4. \end{aligned}$$

For example,

$$\lim_{R_0 L^3/\gamma \to \infty} \int_0^L \frac{\gamma x^n}{(\gamma + R_0 x^3)^2}\,dx = \frac{\gamma^{(n-2)/3}}{R_0^{(n+1)/3}} \frac{1}{n+1}\Gamma\left(\frac{n+4}{3}\right)\Gamma\left(\frac{5-n}{3}\right)$$

is useful when $\gamma \approx 0$ (when the statistical weighting is very small or regularization parameter very large), and

$$\lim_{R_0 L^3/\gamma \to 0} \int_0^L \frac{\gamma x^n}{(\gamma + R_0 x^3)^2}\,dx = \frac{L^{n+1}}{\gamma(n+1)}$$

is useful when γ is large, i.e., when the statistical weighting dominates the regularization.

APPENDIX C

Table-based evaluation for observer performance

In this section we consider evaluation of integrals of the form

$$\int\limits_{\left[-\frac{1}{2},\frac{1}{2}\right]^n} |F(\vec{\nu})|^2 \frac{H_j^W(\vec{\nu})^{p-\hat{p}} H_j^{\hat{W}}(\vec{\nu})^{\hat{p}}}{(H_j^W(\vec{\nu}) + \alpha R(\vec{\nu}))^q}\, d\vec{\nu},$$

as given in (5.27), for Gaussian bumps where $|F(\vec{\nu})| = \exp(-\tau||\vec{\nu}||_2^2)$ and rectangular voxels. We make the approximations that $R(\vec{\nu}) \approx R_0\varrho^2$ and $J(\vec{\nu}) \approx (1 - c\varrho^2)/\varrho$, where $c = \pi^2/3$. This contains the second-order Taylor expansion of $|(\mathcal{F}^n R_{\text{basis}})(\vec{\nu})|^2$ for rectangular voxels, which is

$$|(\mathcal{F}^n R_{\text{basis}})(\vec{\nu})|^2 = \prod_{i=1}^{n} \text{sinc}^2(\nu_i).$$

The second-order Taylor expansion of $\text{sinc}^2(\nu_i)$ is $1 - c\nu_i^2 + O(\nu_i^4)$, and collecting the second-order terms of

$$\prod_{i=1}^{n}(1 - c\nu_i^2 + O(\nu_i^4))$$

gives $1 - c\sum_{i=1}^{n} \nu_i^2 = 1 - c\varrho^2$. Since Gaussian bumps roll off very quickly, in many cases this second-order approximation for J and the second-order approximation for R are cut off by the decay of $|F|$ before their error grows too large.

With these approximations, evaluating (5.27) begins:

$$\int_{\left[-\frac{1}{2},\frac{1}{2}\right]^n} |F(\vec{\nu})|^2 \frac{H_j^W(\vec{\nu})^{p-\hat{p}} H_j^{\hat{W}}(\vec{\nu})^{\hat{p}}}{(H_j^W(\vec{\nu}) + \alpha R(\vec{\nu}))^q}\, d\vec{\nu} = \int_{\mathbb{S}^n} \int_0^{\varrho_{\max}(\vec{\Theta})} e^{-\tau\varrho^2} \frac{J(\varrho)^p E_j^W(\vec{\Theta})^{p-\hat{p}} E_j^{\hat{W}}(\vec{\Theta})^{\hat{p}}}{(J(\varrho)E_j^W(\vec{\Theta}) + \alpha R(\varrho))^q} \varrho^{n-1}\, d\varrho\, d\vec{\Theta}$$

$$= \int_{\mathbb{S}^n} \left(\frac{E_j^{\hat{W}}(\vec{\Theta})}{E_j^W(\vec{\Theta})}\right)^{\hat{p}} E_j^W(\vec{\Theta})^{p-q} \int_0^{\varrho_{\max}(\vec{\Theta})} e^{-\tau\varrho^2} \frac{J(\varrho)^p}{(J(\varrho) + (\alpha/E_j^W(\vec{\Theta}))R(\varrho))^q} \varrho^{n-1}\, d\varrho\, d\vec{\Theta}. \tag{C.1}$$

We then focus on the integral over ϱ above. We first replace its upper limit, $\varrho_{\max}(\vec{\Theta})$, with ∞, under the assumption that $|F|$ rolls off well before the upper limit of the integral. In doing this, the integrand is entirely independent of $\vec{\Theta}$:

$$\int_0^\infty e^{-\tau\varrho^2} \frac{J(\varrho)^p}{(J(\varrho) + (\alpha/E_j^W(\vec{\Theta}))R(\varrho))^q} \varrho^{n-1}\, d\varrho \approx \int_0^\infty e^{-\tau\varrho^2} \frac{(1-c\varrho^2)^p}{(1-c\varrho^2+\gamma\varrho^3)^q} \varrho^{n-1+q-p}\, d\varrho \Bigg|_{\gamma = R_0\alpha/E_j^W(\vec{\Theta})}$$

$$= G_{\text{gauss},p,q,n}(\tau, R_0\alpha/E_j^W(\vec{\Theta})), \tag{C.2}$$

where $G_{\text{gauss},p,q,n}$ is defined as

$$G_{\text{gauss},p,q,n}(\tau,\gamma) \triangleq \int_0^\infty e^{-\tau\varrho^2} \frac{(1-c\varrho^2)^p}{(1-c\varrho^2+\gamma\varrho^3)^q} \varrho^{n-1+q-p}\, d\varrho. \tag{C.3}$$

As we have done previously with functions for variance prediction, we precompute $G_{\text{gauss},p,q,n}$ for a range of potential τ and γ. The integral of (C.1) then becomes, in terms of this function,

$$\int_{\left[-\frac{1}{2},\frac{1}{2}\right]^n} |F(\vec{\nu})|^2 \frac{H_j^W(\vec{\nu})^{p-\hat{p}} H_j^{\hat{W}}(\vec{\nu})^{\hat{p}}}{(H_j^W(\vec{\nu}) + \alpha R(\vec{\nu}))^q}\, d\vec{\nu} \approx \int_{\mathbb{S}^n} \left(\frac{E_j^{\hat{W}}(\vec{\Theta})}{E_j^W(\vec{\Theta})}\right)^{\hat{p}} E_j^W(\vec{\Theta})^{p-q} G_{\text{gauss},p,q,n}(\tau, R_0\alpha/E_j^W(\vec{\Theta}))\, d\vec{\Theta}. \tag{C.4}$$

Computing the integral of (5.27) then becomes this lookup-based numerical integral (C.4).

We can get away with only needing a one-dimensional table if we use an impulse

basis for the voxels instead of a rectangular basis, in which case $J(\vec{\nu}) = 1/\varrho$, and

$$\int_0^{\varrho_{\max}(\vec{\Theta})} e^{-\tau\varrho^2} \frac{J(\varrho)^p}{(J(\varrho) + (\alpha/E_j^W)R(\varrho))^q} \varrho^{n-1}\, d\varrho \approx \int_0^{\infty} e^{-\tau\varrho^2} \varrho^{n-1+q-p}(1+\gamma\varrho^3)^{-q}\, d\varrho \Big|_{\gamma=R_0\alpha/E_j^W(\vec{\Theta})}$$

$$= \tau^{-(n+q-p)/2} \int_0^{\infty} e^{-z^2} z^{n-1+q-p}(1+\gamma z^3/\tau^{3/2})^{-q}\, dz$$

$$= \tau^{-(n+q-p)/2} G^{(1\mathrm{D})}_{\mathrm{gauss},p,q,n}(\gamma/\tau^{3/2}),$$

where

$$G^{(1\mathrm{D})}_{\mathrm{gauss},p,q,n}(a) \triangleq \int_0^{\infty} e^{-z^2} z^{n-1+q-p}(1+az^3)^{-q}\, dz.$$

For the purposes of tube current optimization, tables of the value of the derivative of $G_{\mathrm{gauss},p,q,n}$ with respect to its second argument γ are useful. Conveniently, we can use existing tables:

$$G^{(\gamma)}_{\mathrm{gauss},n,p,q}(\tau,\gamma) \triangleq \frac{\partial}{\partial\gamma} G_{\mathrm{gauss},p,q,n}(\tau,\gamma) = \int_0^{\infty} e^{-\tau\varrho^2} \frac{(1-c\varrho^2)^p}{(1-c\varrho^2+\gamma\varrho^3)^{q+1}}(-q)\varrho^3\varrho^{n-1+q-p}\, d\varrho \tag{C.5}$$

$$= -q\; G_{\mathrm{gauss},p,q+1,n+2}(\tau,\gamma). \tag{C.6}$$

We represent the derivative of $G_{\mathrm{gauss},p,q,n}$ as $G^{(\gamma)}_{\mathrm{gauss},p,q,n}$ when we wish to emphasize that the derivative is being used, but will evaluate the derivative using (C.6).

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] A. H. Andersen. Algebraic reconstruction in CT from limited views. *IEEE Trans. Med. Imag.*, 8(1):50–5, March 1989.

[2] August Beer. Bestimmung der absorption des rothen lichts in farbigen flüssigkeiten. *Ann. Phys. Chem*, 86(2):78–90, 1852.

[3] Kai Tobias Block, Martin Uecker, and Jens Frahm. Undersampled radial MRI with multiple coils. iterative image reconstruction using a total variation constraint. *Magnetic resonance in medicine*, 57(6):1086–1098, 2007.

[4] D. J. Brenner and E. J. Hall. Computed tomography — an increasing source of radiation exposure. *New England Journal of Medicine*, 357(22):2277–2284, 2007.

[5] Arthur E Burgess. From light to optic nerve: optimization of the front end of visual systems. In *Medical Imaging'98*, pages 2–13. International Society for Optics and Photonics, 1998.

[6] Z. Chang, J-B. Thibault, K. Sauer, and C. Bouman. Statistical X-ray computed tomography from photon-starved measurements. In *Proc. SPIE 9020 Computational Imaging XII*, page 90200G, 2014.

[7] J. Cho and J. Fessler. Regularization designs for uniform spatial resolution and noise properties in statistical image reconstruction for 3D X-ray CT. *Medical Imaging, IEEE Transactions on*, PP(99):1–1, 2014.

[8] J. H. Cho and J. A. Fessler. Quadratic regularization design for 3D axial CT. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 78–81, 2013.

[9] N.G. de Bruijn. *Asymptotic methods in analysis*. Dover, 1981.

[10] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens. An iterative maximum-likelihood polychromatic algorithm for CT. *IEEE Trans. Med. Imag.*, 20(10):999–1008, October 2001.

[11] B. De Man, M. Wu, P. FitzGerald, Y. Jin, Z. Yin, P. M. Edic, Y. Yao, X. Tian, and E. Samei. Dose reconstruction for real-time patient-specific dose estimation in CT. In *Proc. 3rd Intl. Mtg. on image formation in X-ray CT*, pages 17–20, 2014.

[12] NIST Digital Library of Mathematical Functions. http://dlmf.nist.gov/, Release 1.0.6 of 2013-05-06. Online companion to [44].

[13] I. A. Elbakri and J. A. Fessler. Efficient and accurate likelihood for iterative image reconstruction in X-ray computed tomography. In *Proc. SPIE 5032, Medical Imaging 2003: Image Proc.*, pages 1839–50, 2003.

[14] H. Erdogan and J. A. Fessler. Ordered subsets algorithms for transmission tomography. *Phys. Med. Biol.*, 44(11):2835–51, November 1999.

[15] U. Fano. Ionization yield of radiations. II. The fluctuations of the number of ions. *Phys. Rev.*, 72(1):26–29, Jul 1947.

[16] L. A. Feldkamp, L. C. Davis, and J. W. Kress. Practical cone-beam algorithm. *J. Opt. Soc. Am. A*, 1(6):612–619, Jun 1984.

[17] J. A. Fessler. Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography. *IEEE Trans. Im. Proc.*, 5(3):493–506, March 1996.

[18] J. A. Fessler. *Image reconstruction: Algorithms and analysis*. ?, 2006. Book in preparation.

[19] J. A. Fessler and S. D. Booth. Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction. *IEEE Trans. Im. Proc.*, 8(5):688–99, May 1999.

[20] J. A. Fessler, H. Erdogan, and W. B. Wu. Exact distribution of edge-preserving MAP estimators for linear signal models with Gaussian measurement noise. *IEEE Trans. Im. Proc.*, 9(6):1049–56, June 2000.

[21] J. A. Fessler and W. L. Rogers. Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs. *IEEE Trans. Im. Proc.*, 5(9):1346–58, September 1996.

[22] M. Gies, W. A. Kalender, H. Wolf, and C. Suess. Dose reduction in CT by anatomically adapted tube current modulation. I. Simulation studies. *Med. Phys.*, 26(11), 1999.

[23] Howard C Gifford, Michael A King, Daniel J de Vries, and Edward J Soares. Channelized Hotelling and human observer correlation for lesion detection in hepatic SPECT imaging. *Journal of Nuclear Medicine*, 41(3):514–521, 2000.

[24] Holger Greess, Anton Nmayr, Heiko Wolf, Ulrich Baum, Michael Lell, Bernhard Bwing, Willi Kalender, and WernerA. Bautz. Dose reduction in CT examination of children by an attenuation-based on-line modulation of tube current (CARE dose). *European Radiology*, 12(6):1571–1576, 2002.

[25] J. Gregor. Data-driven problem reduction for image reconstruction from projections using gift wrapping. *Nuclear Science, IEEE Transactions on*, 58(3):724–729, 2011.

[26] Michael Hardy. Combinatorics of partial derivatives. *The Electronic Journal of Combinatorics*, 13(1), 2006.

[27] H. M. Hudson and R. S. Larkin. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans. Med. Imag.*, 13(4):601–9, December 1994.

[28] P. M. Joseph and R. D. Spital. The exponential edge-gradient effect in X-ray computed tomography. *Phys. Med. Biol.*, 26(3):473–87, May 1981.

[29] W. A. Kalender, H. Wolf, and C. Suess. Dose reduction in CT by anatomically adapted tube current modulation. II. Phantom measurements. *Med. Phys.*, 26(11):2248–53, November 1999.

[30] Willi A. Kalender, Stefanie Buchenau, Paul Deak, Markus Kellermeier, Oliver Langner, Marcel van Straten, Sabrina Vollmar, and Sylvia Wilharm. Technical approaches to the optimisation of CT. *Physica Medica*, 24(2):71 – 79, 2008. Highlights of the X EFOMP Congress - First European Conference on Medical Physics" Il Ciocco, Italy, 20-22 September.

[31] C. Kamphuis and F. J. Beekman. Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm. *IEEE Trans. Med. Imag.*, 17(6):1001–5, December 1998.

[32] C. Stuart Kelley. Moments of the Poisson distribution. *Phys. Rev. B*, 20(8):3221–3222, Oct 1979.

[33] D. Kim, D. Pal, J-B. Thibault, and J. A. Fessler. Improved ordered subsets algorithm for 3D X-ray CT image reconstruction. In *Proc. 2nd Intl. Mtg. on image formation in X-ray CT*, pages 378–81, 2012.

[34] T. Koehler and R. Proksa. Noise properties of maximum likelihood reconstruction with edge-preserving regularization in transmission tomography. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 263–6, 2009.

[35] Yong Long, J.A. Fessler, and J.M. Balter. 3D forward and back-projection for X-Ray CT using separable footprints. *Medical Imaging, IEEE Transactions on*, 29(11):1839–1850, 2010.

[36] L. N. D. Loo, K. Doi, and C. E. Metz. A comparison of physical image comparison indices and observer performance in the radiographic detection of nylon beads. *Phys. Med. Biol.*, 29(7):837–57, July 1984.

[37] J R Mayo, K P Whittall, A N Leung, T E Hartman, C S Park, S L Primack, G K Chambers, M K Limkeman, T L Toth, and S H Fox. Simulated dose reduction in conventional chest ct: validation study. *Radiology*, 202(2):453–457, 1997.

[38] Kevin P Murphy, L Crush, PD McLaughlin, Hilary S OSullivan, Maria Twomey, Sylvia Lynch, J Bye, Sean E McSweeney, Owen J OConnor, F Shanahan, et al. The role of pure iterative reconstruction in conventional dose CT enterography. *Abdominal imaging*, pages 1–7, 2014.

[39] Malakeh A Musheinesh, Charles J Divin, Jeffrey A Fessler, and Theodore B Norris. Time-reversal and model-based imaging in a THz waveguide. In *Conference on Lasers and Electro-Optics*. Optical Society of America, 2009.

[40] Malakeh A Musheinesh, Charles J Divin, Jeffrey A Fessler, and Theodore B Norris. Model-based THz imaging for 2D reflection-mode geometry. In *2010 Conference on Lasers and Electro-Optics (CLEO) and Quantum Electronics and Laser Science Conference (QELS)*, pages 1–2. IEEE, 2010.

[41] F. Natterer and F. Wübbeling. *Mathematical methods in image reconstruction*. Soc. Indust. Appl. Math., Philadelphia, 2001.

[42] J Neyman and ES Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society*, 231:289–337, 1933.

[43] Loren T Niklason, Bradley T Christian, Laura E Niklason, Daniel B Kopans, Donald E Castleberry, BH Opsahl-Ong, Cynthia E Landberg, Priscilla J Slanetz, Angela A Giardino, Richard Moore, et al. Digital tomosynthesis in breast imaging. *Radiology*, 205(2):399–406, 1997.

[44] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, editors. *NIST Handbook of Mathematical Functions*. Cambridge University Press, New York, NY, 2010. Print companion to [12].

[45] S. S. Orlov. Theory of three-dimensional reconstruction II. the recovery operator. *Sov. Phys. Crystallogr.*, 20:429, 1976.

[46] Julien G Ott, Fabio Becce, Pascal Monnin, Sabine Schmidt, François O Bochud, and Francis R Verdun. Update on the non-prewhitening model observer in computed tomography for the assessment of the adaptive statistical and model-based iterative reconstruction algorithms. *Physics in medicine and biology*, 59(15):4047, 2014.

[47] J. Qi and R. M. Leahy. A theoretical study of the contrast recovery and variance of MAP reconstructions from PET data. *IEEE Trans. Med. Imag.*, 18(4):293–305, April 1999.

[48] Victor V. Samedov. Theoretical basis for the experimental determination of the intrinsic resolution of a strip detector. *X-Ray Spectrometry*, 40(1):7–11, 2011.

[49] K. Sauer and C. Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Trans. Sig. Proc.*, 41(2):534–48, February 1993.

[50] S. Schmitt and J. A. Fessler. Fast variance computation for quadratically penalized iterative reconstruction of 3D axial CT images. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, pages 3287–92, 2012.

[51] S. M. Schmitt and J. A. Fessler. Fast variance computation for iterative reconstruction of 3D helical CT images. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 162–5, 2013.

[52] S. M. Schmitt and J. A. Fessler. Fast variance prediction for iteratively reconstructed CT images. *IEEE Trans. Med. Imag.*, 2015. In revsion.

[53] S. M. Schmitt, J. A. Fessler, G. D. Fichter, and D. A. Zimdars. Model-based, one-sided, time-of-flight terahertz image reconstruction. In *Proc. SPIE 9020 Computational Imaging XII*, page 90200N, 2014.

[54] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui. Realistic CT simulation using the 4D XCAT phantom. *Med. Phys.*, 35(8):3800–8, August 2008.

[55] H.R. Shi and J.A. Fessler. Quadratic regularization design for 2-D CT. *Medical Imaging, IEEE Transactions on*, 28(5):645 –656, may 2009.

[56] Alvin C Silva, Holly J Lawder, Amy Hara, Jennifer Kujak, and William Pavlicek. Innovations in CT dose reduction strategy: application of the adaptive statistical iterative reconstruction algorithm. *American Journal of Roentgenology*, 194(1):191–199, 2010.

[57] J. W. Stayman and J. A. Fessler. Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction. *IEEE Trans. Med. Imag.*, 19(6):601–15, June 2000.

[58] Bradley P Sutton, Douglas C Noll, and Jeffrey A Fessler. Fast, iterative image reconstruction for MRI in the presence of field inhomogeneities. *Medical Imaging, IEEE Transactions on*, 22(2):178–188, 2003.

[59] D. Tack, V. De Maertelaer, and P. A. Gevenois. Dose reduction in multidetector CT using attenuation-based online tube current modulation. *American Journal of Roentgenology*, 181:331–4, 2003.

[60] J-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh. A recursive filter for noise reduction in statistical iterative tomographic imaging. In *Proc. SPIE 6065, Computational Imaging IV*, page 60650X, 2006.

[61] J-B. Thibault, K. Sauer, C. Bouman, and J. Hsieh. A three-dimensional statistical approach to improved image quality for multi-slice helical CT. *Med. Phys.*, 34(11):4526–44, November 2007.

[62] Hsin-Wu Tseng, Jiahua Fan, Matthew A Kupinski, Paavana Sainath, and Jiang Hsieh. Assessing image quality and dose reduction of a new X-ray computed

tomography iterative reconstruction algorithm using model observers. *Medical physics*, 41(7):071910, 2014.

[63] United Kingdom Health and Safety Executive. ALARP "at a glance". `http://www.hse.gov.uk/risk/theory/alarpglance.htm`.

[64] United States Nuclear Regulatory Commission. Code of Federal Regulations, Title 10, Section 20.1003. Last amended December 2009.

[65] Varut Vardhanabhuti, Robert Loader, and Carl A Roobottom. Assessment of image quality on effects of varying tube voltage and automatic tube current modulation with hybrid and pure iterative reconstruction techniques in abdominal/pelvic ct: a phantom study. *Investigative radiology*, 48(3):167–174, 2013.

[66] J. G. Verly and R. N. Bracewell. Blurring in tomograms made with X-Ray beams of finite width. *Journal of Computer Assisted Tomography*, 3(5):662–78, october 1979.

[67] G. Wang, M. W. Vannier, and P. C. Cheng. Iterative X-ray cone-beam tomography for metal artifact reduction and local region reconstruction. *Microscopy and Microanalysis*, 5(1):58–65, January 1999.

[68] J. A. C. Weideman. Numerical integration of periodic functions: A few examples. *The American Mathematical Monthly*, 109(1):pp. 21–36, 2002.

[69] B. R. Whiting. Signal statistics in X-ray computed tomography. In *Proc. SPIE 4682, Medical Imaging 2002: Med. Phys.*, pages 53–60, 2002.

[70] B. R. Whiting, P. Massoumzadeh, O. A. Earl, J. A. O'Sullivan, D. L. Snyder, and J. F. Williamson. Properties of preprocessed sinogram data in X-ray computed tomography. *Med. Phys.*, 33(9):3290–303, September 2006.

[71] A. Wunderlich and Frederic Noo. Evaluation of the impact of tube current modulation on lesion detectability using model observers. In *Proc. Int'l. Conf. IEEE Engr. in Med. and Biol. Soc.*, pages 2705–8, 2008.

[72] Adam Wunderlich and Frédéric Noo. Image covariance and lesion detectability in direct fan-beam X-ray computed tomography. *Physics in medicine and biology*, 53(10):2471, 2008.

[73] Jie Yao and Harrison H Barrett. Predicting human performance by a channelized Hotelling observer model. In *San Diego'92*, pages 161–168. International Society for Optics and Photonics, 1992.

[74] A. Yendiki. *Analysis of signal detectability in statistically reconstructed tomographic images*. PhD thesis, Univ. of Michigan, Ann Arbor, MI, 48109-2122, Ann Arbor, MI, 2005.

[75] K. Zeng, B. De Man, and J-B. Thibault. Correction of iterative reconstruction artifacts in helical cone-beam CT. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 242–5, 2009.

[76] Y. Zhang-O'Connor and J. A. Fessler. Fast predictions of variance images for fan-beam transmission tomography with quadratic regularization. *IEEE Trans. Med. Imag.*, 26(3):335–46, March 2007.

[77] Y. Zhang-O'Connor and J. A. Fessler. Fast variance predictions for 3D cone-beam CT with quadratic regularization. In *Proc. SPIE 6510, Medical Imaging: Phys. Med. Im.*, pages 65105W:1–10, 2007.

[78] A. Ziegler, Th. Köhler, and R. Proksa. Noise and resolution in images reconstructed with FBP and OSC algorithms for CT. *Medical Physics*, 34(2):585–598, 2007.