LBL-37749
UC-405

Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

The Statistical Efficiency of Filtered Backprojection in Emission Tomography

A. Kuruc

September 1995

Donner Laboratory

Biology & Medicine Division

REFERENCE COPY
Does Not
Circulate
Bldg. 50 Library.
LBL-37749
Copy 1

Prepared for the U.S. Department of Energy under Contract Number DE-AC03-76SF00098

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

LBL–37749
UC–405

The Statistical Efficiency of Filtered Backprojection in Emission Tomography

Alvin Kuruc

Center for Functional Imaging
Life Sciences Division
Lawrence Berkeley National Laboratory
University of California
Berkeley, CA 94020

September 1995

This work was supported by the Director, Office of Energy Research, Office of Basic Energy Sciences, Medical Applications and Biophysical Research Division, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098, and in part by NIH training grant HL07367, and by National Heart, Lung and Blood Institute of the U.S. Department of Health and Human Services under grant HL25840.

The Statistical Efficiency of Filtered Backprojection in Emission Tomography

Alvin Kuruc

Center for Functional Imaging
Lawrence Berkeley National Laboratory
University of California, Berkeley, CA

September 25, 1995

Running head: Statistical Efficiency of Filtered Backprojection

Address for correspondence:

Alvin Kuruc
MS 55-121
Lawrence Berkeley National Laboratory
University of California
1 Cyclotron Road
Berkeley, CA 94720
A_Kuruc@lbl.gov
510-486-6868
510-486-4768 (FAX)

Abstract

While there has been much interest in developing tomographic reconstruction algorithms that are more statistically efficient than filtered backprojection (FB), the degree of improvement possible has not been well understood. We present an algorithm-independent theory of statistical accuracy attainable in emission tomography that provides a geometrical interpretation of the statistical efficiency of FB. Our analysis shows that, in general, one can build unbiased estimators with smaller variance than FB. The improvement in performance is obtained by exploiting the range properties of the Radon transform.

Keywords: Radon transform, nonparametric estimation, inverse problems, ill-posed problems, regions of interest

1 Introduction

Filtered backprojection (FB) is widely used as a reconstruction algorithm for computed tomography. It can be implemented in a computationally efficient manner and has been found to be reasonably robust in practical applications. While these properties of FB make it an attractive algorithm, its derivation does not explicitly take into account the statistical nature of the observations. This has lead to considerable interest in developing reconstruction algorithms based on explicit statistical models. While these algorithms can be expected to be more or less statistically optimal, they are computationally intensive. This leads to the question: Is the improvement in statistical efficiency worth the increased computational burden? To answer this question, it is necessary to understand the statistical efficiency of FB. In this paper, we propose a framework for understanding the statistical efficiency of FB in the context of emission tomography (ET).

Our analysis shows that FB is just one of many possible unbiased estimators for the image. This multiplicity of unbiased estimators is related to the fact that not all functions on the observation space are Radon transforms of functions on the image space. In general, one can use this fact to build unbiased estimators that are more statistically efficient than FB.

1.1 Outline of Paper

In section 2, we propose a simple statistical model of ET. In section 3, we review concepts related to FB. In section 4, we show that the FB estimator has the form of a linear estimator and derive its statistical properties. In section 5, we derive statistically efficient estimators for the ET problem. In section 6, we discuss concrete representations of the estimators constructed in section 5 with examples. Some concluding remarks are given in section 7.

Remark 1.1 Many of the results discussed in this paper are taken from a previous paper by the author [Kur95]. This paper is largely an attempt to explain these results and their significance in a more accessible way. We refer the reader to [Kur95] for mathematically precise versions of these results.

2 Statistical Model of ET

We start by proposing a simple statistical model of ET. The model is highly idealized in that it ignores numerous secondary physical effects that occur in practice. However, it abstracts the basic problem of ET.

In essence, ET is the problem of estimating the density, f, of a radioactive tracer in a subject as a function of position by external detection of emitted photons. We will consider the simple case where f describes the tracer density on the unit (radius) disk $D \subset \mathbb{R}^2$, where $\mathbb{R}^2$ denotes 2-dimensional Euclidean space. A radioactive disintegration occurring at $\mathbf{x} \in D$ results in the emission of one or two photons that travel along a random line through $\mathbf{x}$ with uniformly distributed orientation. (Positron emitters give off two photons that travel in antipodal directions, hence along the same line.) In most imaging systems, only the lines lying in the plane of D are detected. We will therefore consider the observations in ET to consist of these lines. In other words, we will ignore the 3-dimensional aspect of the problem and treat it as a problem in 2 dimensions.

We assume that f is normalized to integrate to 1. The locations of the radioactive disintegrations are modeled as independent, identically distributed (i.i.d.) random variables with probability density function (p.d.f.) f. The observations are modeled as random lines in $\mathbb{R}^2$ through the locations of the radioactive disintegrations with uniformly distributed orientation. It is then easy to show that the observations are i.i.d. random variables with p.d.f. at a given line proportional to the integral of f along that line.

Let $\mathbb{L}$ denote the set of lines in $\mathbb{R}^2$. If f is a real-valued function on $\mathbb{R}^2$ (we denote this by $\mathsf{f} : \mathbb{R}^2 \to \mathbb{R}$), we define the Radon transform of f to be the function $\mathsf{Rf}\colon \mathbb{L} \to \mathbb{R}$ whose value at $\mathbf{l} \in \mathbb{L}$ is the integral of f over $\mathbf{l}$. (We use boldface type for $\mathbf{l}$ since two coordinates are needed to describe the points in $\mathbb{L}$. It is therefore, in a sense, a vector-valued quantity.) We see that the observations in our model of ET are i.i.d. $\mathbb{L}$-valued random variables with p.d.f. proportional to Rf. In fact, the observations

are distributed according to $\pi^{-1}\mathsf{Rf}$ [JS90, sec. 2.1].

Remark 2.1 To avoid notation confusion, we emphasize that R is a linear function that takes functions on $\mathbb{R}^2$ to functions on $\mathbb{L}$. Such a function is sometimes called a linear operator. Using the standard notation for functions, one could write the result of applying the function R to f as $\mathsf{R}(\mathsf{f})$. However, we have chosen to write it as Rf to emphasize the analogy of the linear operator R acting on f to a matrix $\mathbf{A}$ acting on a vector $\mathbf{x}$. $\mathsf{Rf}(\mathbf{l})$ is therefore obtained by applying the linear operator R to f and evaluating the resulting function at $\mathbf{l} \in \mathbb{L}$.

Remark 2.2 Note that f is defined to be a p.d.f. on the locations of radioactive disintegrations; it contains no information about the absolute *rate* of disintegrations. (This is a result of our assumption that f is normalized to unit integral.) Similarly, the data are taken to be a sequence of elements of $\mathbb{L}$; there is no time information. Thus the way we have set up the problem defines away the problem of estimating the total count rate. This explains why the familiar Poisson distribution does not appear in our model. In practice, one would like to know the total count rate, but good estimates for this quantity are easy to construct. Curiously, this slight change in definition of the model seems to have a substantial effect on the form of the resulting analysis, cf. [VSK85].

3 Filtered Backprojection

In this section, we review the FB algorithm. We start by considering the filtering and backprojection operations for functions on $\mathbb{L}$.

3.1 Filtering on $\mathbb{L}$

The first step of the FB algorithm involves a filtering operation on the observations, which are considered as a function on $\mathbb{L}$. In preparation for this, we review the standard convention for performing filtering operations on functions on $\mathbb{L}$.

Definition 3.1 Define the unit vector $\boldsymbol{\theta} \equiv (\cos\theta, \sin\theta) \in \mathbb{R}^2$. Note the notational distinction between $\theta \in \mathbb{R}$ and the boldface $\boldsymbol{\theta} \in \mathbb{R}^2$. We put coordinates on $\mathbb{L}$ by assigning the coordinates (θ, s) to the line in $\mathbb{R}^2$ through $s\boldsymbol{\theta}$ that is perpendicular to the vector $\boldsymbol{\theta}$. We write the integral of a function g on $\mathbb{L}$ as

$$\int_{\mathbb{L}} \mathsf{g}(\mathbf{l})\, d\mathbf{l} \equiv \int_0^{\pi} \int_{-\infty}^{\infty} \mathsf{g}(\theta, s)\, ds\, d\theta.$$

The usual filtering operation for functions on $\mathbb{R}^2$ is described by the mathematical operation of convolution. For example, suppose f is a function on $\mathbb{R}^2$. A smoothed version of f can be obtained by taking a smooth function a on $\mathbb{R}^2$ and computing the convolution of f and a:

$$\mathsf{a} * \mathsf{f}(\mathbf{x}) \equiv \int_{\mathbb{R}^2} \mathsf{a}(\mathbf{y})\mathsf{f}(\mathbf{x} - \mathbf{y})\, d\mathbf{y}.$$

If a is concentrated around the origin, then $\mathsf{a}*\mathsf{f}(\mathbf{x})$ is a weighed average of the values of f around $\mathbf{x}$.

An alternative description of this operation may be given using the Fourier transform. The Fourier transform of a function f on $\mathbb{R}^d$ is defined by

$$\widetilde{\mathsf{f}}(\boldsymbol{\xi}) \equiv \int_{\mathbb{R}^d} e^{-i2\pi\mathbf{x}\cdot\boldsymbol{\xi}} \mathsf{f}(\mathbf{x})\, d\mathbf{x},$$

where $\mathbf{x}\cdot\boldsymbol{\xi}$ denotes the inner product of $\mathbf{x}$ and $\boldsymbol{\xi}$. It is well-known that the Fourier transform of $\mathsf{a} * \mathsf{f}$ is equal to the product of the Fourier transforms of f and a. Thus

the operation of convolution with a is equivalent to the operation of multiplying the Fourier transform of f by the Fourier transform of a and taking the inverse Fourier transform. It is customary to refer to this operation as applying the frequency-domain filter $\tilde{\mathsf{a}}$, where $\tilde{\mathsf{a}}$ is the Fourier transform of a.

To define convolutions of functions on $\mathbb{L}$, we use the convention that the convolution is taken with respect to the second variable only, i.e.,

$$\mathsf{b} * \mathsf{g}(\theta, s) \equiv \int_{-\infty}^{\infty} \mathsf{b}(\theta, t)\mathsf{g}(\theta, s-t)\, dt.$$

Similarly, the Fourier transform of functions on $\mathbb{L}$ is taken with respect to the second variable only:

$$\tilde{\mathsf{g}}(\theta, \eta) \equiv \int_{-\infty}^{\infty} e^{-i2\pi s\eta}\mathsf{g}(\theta, s)\, ds.$$

With these conventions, the usual result that the Fourier transform of the convolution of two functions is equal to the product of their Fourier transforms holds for functions on $\mathbb{L}$.

Remark 3.2 For fixed θ, $\mathsf{Rf}(\theta, s)$ is a projection of f. Thus the convention of considering convolution and the Fourier transform with respect to the second variable only amounts to applying these operators to each projection.

3.2 Backprojection

The second step of the FB algorithm is backprojection.

Definition 3.3 The backprojection operator, which we denote by R^T, is a linear operator that takes functions on $\mathbb{L}$ to functions on $\mathbb{R}^2$, i.e., it goes in the opposite direction of R. It maps the function g on $\mathbb{L}$ to the function $\mathsf{R}^T\mathsf{g}$ on $\mathbb{R}^2$ defined by

$$\mathsf{R}^T\mathsf{g}(\mathbf{x}) \equiv \int_0^{\pi} \mathsf{g}(\theta, \mathbf{x}\cdot\boldsymbol{\theta})\, d\theta. \tag{3.1}$$

The notation R^T is used to indicate that the backprojection operator is the *adjoint* operator of R. This means that the equality

$$\int_{\mathbb{L}} \mathsf{Rf}(\mathbf{l})\, \mathsf{g}(\mathbf{l})\, d\mathbf{l} = \int_{\mathbb{R}^2} \mathsf{f}(\mathbf{x})\, \mathsf{R}^T\mathsf{g}(\mathbf{x})\, d\mathbf{x} \tag{3.2}$$

holds for all functions f on $\mathbb{R}^2$ and g on $\mathbb{L}$ (satisfying certain technical conditions that ensure the existence of the integrals) [Her83, p. 169].

Remark 3.4 The matrix analog to the adjoint of a linear operator is the transpose. The matrix analog of equation 3.2 is the matrix identity $\mathbf{y} \cdot \mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{y} \cdot \mathbf{x}$. In equation 3.2, the integral on the left side of the equation serves as the inner product of the functions Rf and g. Similarly, the integral on the right side of the equation serves as the inner product of the functions f and $\mathsf{R}^T\mathsf{g}$.

3.3 The FB algorithm

Definition 3.5 We define the ramp filter for functions on $\mathbb{L}$, which we denote by H, to be the frequency-domain filter $2\pi|\eta|$, i.e., $\widetilde{\mathsf{Hg}}(\theta, \eta) \equiv 2\pi|\eta|\tilde{\mathsf{g}}(\theta, \eta)$.

The FB algorithm is based on the following inversion formula for the Radon transform:

$$\mathsf{f} = (2\pi)^{-1}\mathsf{R}^T\mathsf{HRf}, \tag{3.3}$$

cf. [Nat86, thm II.2.1]. In words, f can be recovered from its Radon transform by ramp filtering each projection and then backprojecting.

In practice, Rf is not known exactly, but is measured with some statistical error. As a result, direct use of equation 3.3 is not feasible since the ramp filter H will result in unacceptable amplification of the high-frequency components of the statistical errors. To counteract this problem, the ramp filter H is usually combined in series with a low pass filter. We express this low pass filter as convolution by the function w on $\mathbb{L}$. The effect of adding the low-pass filter on the resulting reconstruction may be understood from the formula

$$\mathsf{a} * \mathsf{f} = (2\pi)^{-1}\mathsf{R}^T(\mathsf{Hw} * \mathsf{Rf}), \tag{3.4}$$

where

$$\mathsf{a} \equiv (2\pi)^{-1}\mathsf{R}^T\mathsf{Hw}, \tag{3.5}$$

cf. [Nat86, eq. V.1.2]. In other words, instead of estimating $\mathsf{f}(\mathbf{x})$, we estimate the weighted value $\mathsf{a} * \mathsf{f}(\mathbf{x})$ of f around $\mathbf{x}$. The weighting function a is referred to as the aperture function.

In what follows, we will assume that the filter function w is a smooth function on $\mathbb{L}$ that is symmetric in s and independent of θ. Then a is a radial function on $\mathbb{R}^2$, i.e., $\mathsf{a}(\mathbf{x})$ depends only on $|\mathbf{x}|$, where $|\mathbf{x}|$ denotes the Euclidean norm of $\mathbf{x}$. Under these conditions, we can invert the relationship between a and w. To do so, we first note that, under these conditions, w satisfies the well-known conditions for being in the range of the Radon transform [Nat86, thm. II.4.1]. We can therefore write $\mathsf{w} = \mathsf{R}\mathsf{a}'$ for some function a' on $\mathbb{R}^2$. Substituting this equation into equation 3.5 and applying equation 3.3 gives

$$\begin{aligned} \mathsf{a} &= (2\pi)^{-1}\mathsf{R}^T\mathsf{H}\mathsf{R}\mathsf{a}' \\ &= \mathsf{a}', \end{aligned}$$

and hence

$$\mathsf{w} = \mathsf{R}\mathsf{a}.$$

Example 3.6 Suppose a is a Gaussian density function on $\mathbb{R}^2$ centered at the origin with dispersion σ^2, i.e.,

$$\mathsf{a}(\mathbf{x}) = (2\pi\sigma^2)^{-1}e^{-|\mathbf{x}|^2/2\sigma^2}.$$

The corresponding $\mathsf{w} = \mathsf{R}\mathsf{a}$ is a Gaussian density function with respect to the second variable on $\mathbb{L}$, i.e.,

$$\mathsf{w}(\theta, s) = (2\pi\sigma^2)^{-1/2}e^{-s^2/2\sigma^2}$$

[Dea83, eq. 3.4.5]. This corresponds to the low-pass frequency-domain filter

$$\tilde{\mathsf{w}}(\theta, \eta) = e^{-\eta^2/2\tau^2},$$

where $\tau \equiv 1/2\pi\sigma$, which also has a Gaussian functional form.

Remark 3.7 In discrete versions of the FB algorithm, the low-pass filter often comes in implicitly by means of interpolation in the backprojection step. Detailed discussion of this point may be found in [Lew83, sec. III] and [Nat86, sec. V.1].

3.4 The FB algorithm in ET

We apply equation 3.4 to ET in the following way. We estimate Rf by $\pi n^{-1}\sum_{i=1}^{n}\delta(\mathbf{l}_i)$, where n is the number of observed lines, $\mathbf{l}_i = (\theta_i, s_i) \in \mathbb{L}$ is the ith observation, and $\delta(\mathbf{l}_i)$ denotes a unit point mass, i.e., a delta function, at $\mathbf{l}_i$. An estimate of $\mathsf{a} * \mathsf{f}$ is then obtained by substituting $\pi n^{-1}\sum_{i=1}^{n}\delta(\mathbf{l}_i)$ in for Rf in equation 3.4. Thus the FB estimate of $\mathsf{a} * \mathsf{f}(\mathbf{x})$ amounts to

$$\begin{aligned}(2\pi)^{-1}\mathsf{R}^T[\mathsf{Hw} * \pi n^{-1}\sum_{i=1}^{n}\delta(\mathbf{l}_i)](\mathbf{x}) &= 2^{-1}n^{-1}\sum_{i=1}^{n}\mathsf{R}^T[\mathsf{Hw} * \delta(\mathbf{l}_i)](\mathbf{x}) \\ &= 2^{-1}n^{-1}\sum_{i=1}^{n}\mathsf{HRa}(\theta_i, s_i - \mathbf{x}\cdot\boldsymbol{\theta}_i), \qquad (3.6)\end{aligned}$$

where in the last equality we use the assumed symmetry of $\mathsf{w} = \mathsf{Ra}$.

Let $\mathbf{x} \in \mathbb{R}^2$. We will now show that the FB estimate of $\mathsf{a} * \mathsf{f}(\mathbf{x})$ is linear in the sense that it can be written in the form $n^{-1}\sum_{i=1}^{n}\mathsf{b}_{\mathbf{x}}(\mathbf{l}_i)$ for a function $\mathsf{b}_{\mathbf{x}}$ on $\mathbb{L}$. Define the translation of the aperture function a by $\mathbf{x} \in \mathbb{R}^2$ by $\mathsf{a}_{\mathbf{x}}(\mathbf{x}') \equiv \mathsf{a}(\mathbf{x}' - \mathbf{x})$. Then, using the shifting property of the Radon transform, $\mathsf{Ra}_{\mathbf{x}}(\theta, s) = \mathsf{Ra}(\theta, s - \mathbf{x}\cdot\boldsymbol{\theta})$ [Dea83, eq. 3.5.1], equation 3.6 can be rewritten as

$$(2\pi)^{-1}\mathsf{R}^T[\mathsf{Hw} * \pi n^{-1}\sum_{i=1}^{n}\delta(\mathbf{l}_i)](\mathbf{x}) = 2^{-1}n^{-1}\sum_{i=1}^{n}\mathsf{HRa}_{\mathbf{x}}(\mathbf{l}_i).$$

Define the function $\mathsf{b}_{\mathbf{x}}\colon \mathbb{L} \to \mathbb{R}$ by

$$\mathsf{b}_{\mathbf{x}} \equiv 2^{-1}\mathsf{HRa}_{\mathbf{x}}. \qquad (3.7)$$

Then our estimate of $\mathsf{a} * \mathsf{f}(\mathbf{x})$ can be written as $n^{-1}\sum_{i=1}^{n}\mathsf{b}_{\mathbf{x}}(\mathbf{l}_i)$. In other words, our estimate of $\mathsf{a} * \mathsf{f}(\mathbf{x})$ is the average value of the function $\mathsf{b}_{\mathbf{x}}$ at the observation points. We shall call the function $\mathsf{b}_{\mathbf{x}}$ the observation-space representation of the FB estimator at $\mathbf{x}$.

Example 3.8 Continuing example 3.6, suppose the aperture function a is a Gaussian density function centered at the origin with dispersion σ^2. Then $\mathsf{a}_{\mathbf{x}}$ is a Gaussian density function centered at $\mathbf{x} \in \mathbb{R}^2$ with dispersion σ^2. A calculation, the details of which may be found in [Kur95, prop. 6.2], shows that

$$\mathsf{b}_{\mathbf{x}}(\theta, s) = (2\pi\sigma^2)^{-1} e^{-(s-\mathbf{x}\cdot\boldsymbol{\theta})^2/2\sigma^2} \boldsymbol{\Phi}(-1/2; 1/2; (s - \mathbf{x}\cdot\boldsymbol{\theta})^2/2\sigma^2), \tag{3.8}$$

where $\boldsymbol{\Phi}$ denotes the special function known as Kummer's confluent hypergeometric function [Sla72]. Defining the function $\chi_\sigma : \mathbb{R} \to \mathbb{R}$ by

$$\chi_\sigma(s) \equiv (2\pi\sigma^2)^{-1} e^{-s^2/2\sigma^2} \boldsymbol{\Phi}(-1/2; 1/2; s^2/2\sigma^2),$$

we can write

$$\mathsf{b}_{\mathbf{x}}(\theta, s) = \chi_\sigma(s - \mathbf{x}\cdot\boldsymbol{\theta}).$$

Thus for each fixed θ, $\mathsf{b}_{\mathbf{x}}$ as a function of s is a translate of the function χ_σ. In figure 1, we illustrate the function χ_σ for $\sigma = 0.1$. In figure 2, the graph on the upper left shows a Gaussian aperture function centered at the origin with $\sigma = 0.1$. The graph on the upper right shows the corresponding $\mathsf{b}_{\mathbf{x}}$. The lower half of figure 1 is similar to the upper half, except that the aperture function is now centered at $\mathbf{x} = (1, 0)$ instead of at the origin. Figure 3 is identical to figure 2 except that $\sigma = 0.5$ instead of $\sigma = 0.1$.

Remark 3.9 It is useful to compare the ET problem with the simpler problem where the observations are distributed according to f itself, i.e., the problem where we observe the locations of the radioactive disintegrations. We term this problem the planar imaging problem. For this problem, the natural estimator for the quantity $\mathsf{a} * \mathsf{f}(\mathbf{x})$ is the linear estimator $n^{-1}\sum_{i=1}^{n} \mathsf{a}_{\mathbf{x}}(\mathbf{x}_i)$, where the $\mathbf{x}_i$ are the observations.

4 Statistical Properties of FB

As discussed in section 3, it is impractical to estimate arbitrary features of f by filtered backprojection when Rf is measured with statistical errors. To obtain a statistically well-posed problem, we instead estimate $\mathsf{a} * \mathsf{f}$, where a is an aperture function. For $\mathbf{x} \in \mathbb{R}^2$, the value of the FB estimate at $\mathbf{x}$ may be thought of as an estimate of

$$\mathsf{a} * \mathsf{f}(\mathbf{x}) = \int_{\mathbb{R}^2} \mathsf{a}_{\mathbf{x}}(\mathbf{x}')\mathsf{f}(\mathbf{x}')\, d\mathbf{x}'. \tag{4.1}$$

For notational convenience, we define $\mathsf{a}_{\mathbf{x}} \cdot \mathsf{f} \equiv \int_{\mathbb{R}^2} \mathsf{a}_{\mathbf{x}}(\mathbf{x}')\mathsf{f}(\mathbf{x}')\, d\mathbf{x}'$, using the analogy of the integral $\int_{\mathbb{R}^2} \mathsf{a}_{\mathbf{x}}(\mathbf{x}')\mathsf{f}(\mathbf{x}')\, d\mathbf{x}'$ to an inner product.

To assess the statistical properties of FB, we consider the value of the FB estimate at $\mathbf{x} \in \mathbb{R}^2$ as an estimate of $\mathsf{a}_{\mathbf{x}} \cdot \mathsf{f}$. The representation of FB as a linear estimator given in section 3.4 makes it easy to compute the statistical properties of this estimate. We denote mathematical expectation when the true image density is f by E_{f}. Since the $\mathbf{l}_i$ are independent, the expected value of the estimator when the true image density is f is given by

$$\begin{aligned} E_{\mathsf{f}} n^{-1} \sum_{i=1}^{n} \mathsf{b}_{\mathbf{x}}(\mathbf{l}_i) &= (2\pi)^{-1} \int_{\mathbb{L}} \mathsf{HRa}_{\mathbf{x}}(\mathbf{l})\, \mathsf{Rf}(\mathbf{l})\, d\mathbf{l} \\ &= (2\pi)^{-1} \int_{\mathbb{R}^2} \mathsf{R}^T\mathsf{HRa}_{\mathbf{x}}(\mathbf{x}')\mathsf{f}(\mathbf{x}')\, d\mathbf{x}' \\ &= \int_{\mathbb{R}^2} \mathsf{a}_{\mathbf{x}}(\mathbf{x}')\, \mathsf{f}(\mathbf{x}')\, d\mathbf{x}' \\ &= \mathsf{a}_{\mathbf{x}} \cdot \mathsf{f}, \end{aligned}$$

where we used equations 3.2 and 3.3. Thus FB has the desirable property that it is an unbiased estimator of $\mathsf{a}_{\mathbf{x}} \cdot \mathsf{f}$.

The variance at f is given by

$$\begin{aligned} &E_{\mathsf{f}} \left(n^{-1} \sum_{i=1}^{n} \mathsf{b}_{\mathbf{x}}(\mathbf{l}_i) - \mathsf{a}_{\mathbf{x}} \cdot \mathsf{f} \right)^2 \\ &= n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} E_{\mathsf{f}}\{[\mathsf{b}_{\mathbf{x}}(\mathbf{l}_i) - \mathsf{a}_{\mathbf{x}} \cdot \mathsf{f}][\mathsf{b}_{\mathbf{x}}(\mathbf{l}_j) - \mathsf{a}_{\mathbf{x}} \cdot \mathsf{f}]\} \end{aligned}$$

$$
\begin{aligned}
&= n^{-2} \sum_{i=1}^{n} E_{\mathsf{f}}\{[\mathsf{b}_{\mathbf{x}}(\mathbf{l}_i) - \mathsf{a}_{\mathbf{x}} \cdot \mathsf{f}]^2\} \\
&= n^{-2} \sum_{i=1}^{n} \{E_{\mathsf{f}} \mathsf{b}_{\mathbf{x}}^2(\mathbf{l}_i) - (\mathsf{a}_{\mathbf{x}} \cdot \mathsf{f})^2\} \\
&= n^{-1} \left(\pi^{-1} \int_{\mathbb{L}} \mathsf{b}_{\mathbf{x}}^2(\theta, s)\, \mathsf{Rf}(\theta, s)\, ds\, d\theta - (\mathsf{a}_{\mathbf{x}} \cdot \mathsf{f})^2 \right). \qquad (4.2)
\end{aligned}
$$

Example 4.1 Let f_u denote the uniform distribution on D, i.e., f_u is the constant function π^{-1}. Then $\mathsf{Rf}_u(\theta, s) = 2\pi^{-1}\sqrt{1-s^2}$ (cf. [Dea83, sec. 2.5, ex. 4]). We numerically evaluated equation 4.2 when $\mathsf{f}_0 = \mathsf{f}_u$, a is the Gaussian aperture function considered in example 3.8, $\mathbf{x}$ is the origin, and $n = 10^6$. The results are shown in the upper curve of figure 4 as function of the standard deviation of the aperture function, σ. For comparison, the variance of the estimator $n^{-1} \sum_{i=1}^{n} \mathsf{a}_{\mathbf{x}}(\mathbf{x}_i)$ for the planar-imaging problem is shown in the lower curve. The asymptotic behavior of these curves as $\sigma \to 0$ can be described very simply. For small σ, the variance of the FB estimate for $\mathsf{a}_{\mathbf{x}} \cdot \mathsf{f}$ is approximately $1/8\pi^{3/2}\sigma^3 n$. In comparison, the variance for the planar imaging problem is approximately $1/4\pi^2\sigma^2 n$. We refer the reader to [Kur95, prop. 6.10] for the details of the calculations.

Remark 4.2 The problem of x-ray computed tomography with a Gaussian aperture function was considered in [Tre78]. While the structure of the observations and the noise for this problem differ from that of the ET problem, it is interesting to note that the result in [Tre78] reduces to $1/8\pi^{3/2}\sigma^3 n$ by taking appropriate limits, where n now denotes the number of transmitted photons.

5 Efficient Estimators for ET

Now that we have characterized the statistical performance of FB, it is natural to ask how it compares to other estimators. To make this concrete, we will compare the performance of various estimators for the task of estimating $\mathsf{a}_{\mathbf{x}} \cdot \mathsf{f}$. We have just seen that FB is an unbiased estimator of $\mathsf{a}_{\mathbf{x}} \cdot \mathsf{f}$. A natural figure of merit for unbiased estimators is variance. Since the variance of estimators will vary as a function of f, it is necessary to fix some f_0. We then ask: among all unbiased estimators of $\mathsf{a}_{\mathbf{x}} \cdot \mathsf{f}$, which estimator has the smallest variance at f_0. This estimator is termed the efficient estimator at f_0. We can then define the statistical efficiency of FB at f_0 as the ratio of the variance of the efficient estimator to the variance of FB at f_0.

We saw in section 4 that the FB estimate for $\mathsf{a}_{\mathbf{x}} \cdot \mathsf{f}$ has the form of a linear estimator, i.e., the estimator is of the form $n^{-1} \sum_{i=1}^{n} \mathsf{b}(\mathbf{l}_i)$, where b is a function on $\mathbb{L}$. It turns out the efficient estimator is linear as well. The proof of this fact given in [Kur95, sec. 4] is rather long and technical. Instead of presenting this proof in its entirety, we shall restrict ourselves to proving that this estimator is an efficient linear estimator, i.e., an unbiased linear estimator whose variance at f_0 is minimal with respect to all unbiased linear estimators. Restricting the analysis to linear estimators minimizes mathematical technicalities and, in the author's opinion, gives the most insight into the problem.

We start our construction of an efficient estimator by characterizing the unbiased linear estimators. Consider the linear estimator generated by the function b on $\mathbb{L}$. Its expected value at f is given by

$$E_{\mathsf{f}} n^{-1} \sum_{i=1}^{n} \mathsf{b}(\mathbf{l}_i) = \pi^{-1} \int_{\mathbb{L}} \mathsf{b}(\mathbf{l}) \mathsf{R}\mathsf{f}(\mathbf{l})\, d\mathbf{l}.$$

Thus the condition that the estimator generated by b is unbiased amounts to the condition

$$\begin{aligned} \pi^{-1} \int_{\mathbb{L}} \mathsf{b}(\mathbf{l})\, \mathsf{R}\mathsf{f}(\mathbf{l})\, d\mathbf{l} &= \pi^{-1} \int_{\mathbb{R}^2} \mathsf{R}^T \mathsf{b}_{\mathbf{x}}(\mathbf{x}') \mathsf{f}(\mathbf{x}')\, d\mathbf{x}' \\ &= \int_{\mathbb{R}^2} \mathsf{a}_{\mathbf{x}}(\mathbf{x}') \mathsf{f}(\mathbf{x}')\, d\mathbf{x}' \end{aligned} \tag{5.1}$$

for all p.d.f.s f on D. A standard argument shows that the last equality holding for all p.d.f.s f on D is equivalent to the condition that

$$\pi^{-1}\mathsf{R}^T\mathsf{b} = \mathsf{a}_{\mathbf{x}} \tag{5.2}$$

on D, i.e., b backprojects to $\mathsf{a}_{\mathbf{x}}$ on D.

Since the function $\mathsf{b}_{\mathbf{x}}$ satisfies the equation $\pi^{-1}\mathsf{R}^T\mathsf{b} = \mathsf{a}_{\mathbf{x}}$ on $\mathbb{R}^2$, it *a fortiori* satisfies it on D and hence generates an unbiased estimator. However, it is not the only unbiased estimator. It is clear from equation 5.2 that we can add any function that backprojects to 0 to b and the resulting sum will generate an unbiased linear estimator. For example, we could add the function

$$b'(\theta, s) \equiv \begin{cases} 1 & \text{if } 0 \leq \theta < \pi/2 \\ -1 & \text{if } \pi/2 \leq \theta < \pi \end{cases} .$$

Thus there are many unbiased linear estimators for $\mathsf{a}_{\mathbf{x}} \cdot \mathsf{f}$ and we are thus lead to the problem of finding which one has the least variance .

Suppose b generates an unbiased linear estimator. Essentially the same calculation as that made in equation 4.2 shows that the variance of this estimator at f_0 is given by

$$n^{-1}\left(\pi^{-1}\int_{\mathbb{L}} \mathsf{b}^2(\mathbf{l})\mathsf{R}\mathsf{f}_0(\mathbf{l})\, d\mathbf{l} - (\mathsf{a}_{\mathbf{x}} \cdot \mathsf{f}_0)^2\right).$$

We see that finding the efficient linear estimator amounts to finding the b that minimizes $\pi^{-1}\int_{\mathbb{L}} \mathsf{b}^2(\mathbf{l})\mathsf{R}\mathsf{f}_0(\mathbf{l})\, d\mathbf{l}$ subject to the constraint $\pi^{-1}\mathsf{R}^T\mathsf{b} = \mathsf{a}_{\mathbf{x}}$ on D.

We now proceed to describe the solution to this constrained optimization problem. Let $\mathbb{L}(D)$ denote the subset of $\mathbb{L}$ consisting of lines that intersect D. For the remainder of the paper, we will view R as an operator that maps functions on D to functions on $\mathbb{L}(D)$. It is easy to show that the adjoint of R viewed as an operator taking functions on D to functions on $\mathbb{L}(D)$ is just R^T viewed as an operator taking functions on $\mathbb{L}(D)$ to functions on D, i.e., R^T satisfies

$$\int_{\mathbb{L}(D)} \mathsf{R}\mathsf{f}(\mathbf{l})\, \mathsf{g}(\mathbf{l})\, d\mathbf{l} = \int_D \mathsf{f}(\mathbf{x})\mathsf{R}^T\mathsf{g}(\mathbf{x})\, d\mathbf{x}$$

for all functions f on D and g on $\mathbb{L}(D)$. With these definitions, the unbiased condition reduces to $\pi^{-1}\mathsf{R}^T\mathsf{b} = \mathsf{a}_{\mathbf{x}}$, where we now view R^T as an operator taking functions on $\mathbb{L}(D)$ to functions on D. We denote the set of functions on $\mathbb{L}(D)$ whose elements satisfy the equation $\mathsf{R}^T\mathsf{b} = 0$ by $\mathcal{N}(\mathsf{R}^T)$, i.e., $\mathcal{N}(\mathsf{R}^T)$ is the nullspace of the operator R^T. (The reason we fuss about the domains of the functions is that the nullspace of R^T depends on the assumed domain of the functions. For example, there exist nonzero functions on $\mathbb{L}(D)$ that backproject to 0 on D, but not on $\mathbb{R}^2$.)

Define $L^2(\mathbb{L}(D), \mathsf{Rf}_0)$ to be the space of square integrable functions on $\mathbb{L}(D)$ with respect to the weighting function Rf_0, i.e., the set of functions b on $\mathbb{L}(D)$ such that

$$||\mathsf{b}||^2_{L^2(\mathbb{L}(D),\mathsf{Rf}_0)} \equiv \int_{\mathbb{L}} \mathsf{b}^2(\mathbf{l})\, \mathsf{Rf}_0(\mathbf{l})\, d\mathbf{l} < \infty.$$

Using this notation, the variance of the estimator generated by b at f_0 can be written as

$$n^{-1}\left(\pi^{-1}||\mathsf{b}||^2_{L^2(\mathbb{L}(D),\mathsf{Rf}_0)} - (\mathsf{a}_{\mathbf{x}} \cdot \mathsf{f}_0)^2\right).$$

Thus the efficient linear estimator is generated by the smallest $\mathsf{b} \in L^2(\mathbb{L}(D), \mathsf{Rf}_0)$ that satisfies the unbiased condition $\pi^{-1}\mathsf{R}^T\mathsf{b} = \mathsf{a}_{\mathbf{x}}$.

The key to the solution of this constrained optimization problem is the following orthogonal composition theorem for $L^2(\mathbb{L}(D), \mathsf{Rf}_0)$. If $\mathsf{b}, \mathsf{b}' \in L^2(\mathbb{L}(D), \mathsf{Rf}_0)$, then their inner product in $L^2(\mathbb{L}(D), \mathsf{Rf}_0)$ is defined by

$$\langle \mathsf{b}, \mathsf{b}' \rangle_{L^2(\mathbb{L}(D),\mathsf{Rf}_0)} \equiv \int_{\mathbb{L}(D)} \mathsf{b}(\mathbf{l})\mathsf{b}'(\mathbf{l})\mathsf{Rf}_0(\mathbf{l})\, d\mathbf{l}. \tag{5.3}$$

If $\langle \mathsf{b}, \mathsf{b}' \rangle_{L^2(\mathbb{L}(D),\mathsf{Rf}_0)} = 0$, then b and b' are said to be orthogonal in $L^2(\mathbb{L}(D), \mathsf{Rf}_0)$. We define the orthogonal complement of $\mathcal{N}(\mathsf{R}^T)$ in $L^2(\mathbb{L}(D), \mathsf{Rf}_0)$ to be the subspace of $L^2(\mathbb{L}(D), \mathsf{Rf}_0)$ whose elements are orthogonal to all $\mathsf{b}' \in \mathcal{N}(\mathsf{R}^T)$. We denote it by $\mathcal{N}(\mathsf{R}^T)^\perp$. We emphasize that, unlike $\mathcal{N}(\mathsf{R}^T)$, $\mathcal{N}(\mathsf{R}^T)^\perp$ depends on f_0 through the weighting function Rf_0. If $\mathsf{b} \in L^2(\mathbb{L}(D), \mathsf{Rf}_0)$, there is a unique decomposition

$$\mathsf{b} = p_{\mathcal{N}(\mathsf{R}^T)}\mathsf{b} + p_{\mathcal{N}(\mathsf{R}^T)^\perp}\mathsf{b}$$

of b such that $p_{\mathcal{N}(\mathsf{R}^T)}\mathsf{b} \in \mathcal{N}(\mathsf{R}^T)$ and $p_{\mathcal{N}(\mathsf{R}^T)^\perp}\mathsf{b} \in \mathcal{N}(\mathsf{R}^T)^\perp$. The functions $p_{\mathcal{N}(\mathsf{R}^T)}\mathsf{b}$ and $p_{\mathcal{N}(\mathsf{R}^T)^\perp}\mathsf{b}$ are termed the projections of b onto $\mathcal{N}(\mathsf{R}^T)$ and $\mathcal{N}(\mathsf{R}^T)^\perp$, respectively. The projection $p_{\mathcal{N}(\mathsf{R}^T)}\mathsf{b}$ (respectively $p_{\mathcal{N}(\mathsf{R}^T)^\perp}\mathsf{b}$) has the geometric interpretation of uniquely minimizing $||\mathsf{b}' - \mathsf{b}||^2_{L^2(\mathbb{L}(D),\mathsf{Rf}_0)}$ over $\mathsf{b}' \in \mathcal{N}(\mathsf{R}^T)$ (respectively $\mathsf{b}' \in \mathcal{N}(\mathsf{R}^T)^\perp$), i.e., it is the closest point in $\mathcal{N}(\mathsf{R}^T)$ (respectively $\mathcal{N}(\mathsf{R}^T)^\perp$) to b with respect to the metric $||\cdot||_{L^2(\mathbb{L}(D),\mathsf{Rf}_0)}$. An easy calculation gives a version of the Pythagorean theorem:

$$||\mathsf{b}||^2_{L^2(\mathbb{L}(D),\mathsf{Rf}_0)} = ||p_{\mathcal{N}(\mathsf{R}^T)}\mathsf{b}||^2_{L^2(\mathbb{L}(D),\mathsf{Rf}_0)} + ||p_{\mathcal{N}(\mathsf{R}^T)^\perp}\mathsf{b}||^2_{L^2(\mathbb{L}(D),\mathsf{Rf}_0)}.$$

We will now prove that the efficient linear estimator is generated by $p_{\mathcal{N}(\mathsf{R}^T)^\perp}\mathsf{b}_\mathbf{x}$. We first note that $p_{\mathcal{N}(\mathsf{R}^T)^\perp}\mathsf{b}_\mathbf{x}$ generates an unbiased estimator since it differs from $\mathsf{b}_\mathbf{x}$ by an element in $\mathcal{N}(\mathsf{R}^T)$ and hence satisfies the unbiasedness condition given by equation 5.2. If b is any other solution of equation 5.2, then $\mathsf{b} - p_{\mathcal{N}(\mathsf{R}^T)^\perp}\mathsf{b}_\mathbf{x} \in \mathcal{N}(\mathsf{R}^T)$. The uniqueness of the orthogonal decomposition and the Pythagorean theorem then gives

$$||\mathsf{b}||^2_{L^2(\mathbb{L}(D),\mathsf{Rf}_0)} = ||p_{\mathcal{N}(\mathsf{R}^T)^\perp}\mathsf{b}_\mathbf{x}||^2_{L^2(\mathbb{L}(D),\mathsf{Rf}_0)} + ||\mathsf{b} - p_{\mathcal{N}(\mathsf{R}^T)^\perp}\mathsf{b}_\mathbf{x}||^2_{L^2(\mathbb{L}(D),\mathsf{Rf}_0)}.$$

This is clearly minimized when $\mathsf{b} = p_{\mathcal{N}(\mathsf{R}^T)^\perp}\mathsf{b}_\mathbf{x}$.

We have just shown that the efficient estimator is generated by the projection of $\mathsf{b}_\mathbf{x}$ onto the subspace $\mathcal{N}(\mathsf{R}^T)^\perp$ of $L^2(\mathbb{L}, \mathsf{Rf}_0)$. There is an important alternative characterization of $\mathcal{N}(\mathsf{R}^T)^\perp$ as the range of the Radon transform. We define the normalized Radon transform $\mathsf{R}_{\mathsf{f}_0}$ as the operator that maps f to $\mathsf{Rf}/\mathsf{Rf}_0$. The statistical motivation for introducing this normalization is that $\mathsf{R}_{\mathsf{f}_0}\mathsf{f}(\mathbf{l})$ is the likelihood ratio of the observation $\mathbf{l} \in \mathbb{L}(D)$ under the statistical hypotheses f and f_0. Then $\mathcal{N}(\mathsf{R}^T)^\perp$ is equal to the range of the normalized transform $\mathsf{R}_{\mathsf{f}_0}$ in $L^2(\mathbb{L}, \mathsf{Rf}_0)$, $\mathcal{R}(\mathsf{R}_{\mathsf{f}_0})$. (Strictly speaking, $\mathcal{N}(\mathsf{R}^T)^\perp$ is actually the closure of $\mathcal{R}(\mathsf{R}_{\mathsf{f}_0})$, i.e., we need to include the limits of sequences of functions in $\mathcal{R}(\mathsf{R}_{\mathsf{f}_0})$ in addition to $\mathcal{R}(\mathsf{R}_{\mathsf{f}_0})$.) To prove this, we note that R^T is the adjoint of the operator $\mathsf{R}_{\mathsf{f}_0}$ when $\mathsf{R}_{\mathsf{f}_0}$ is viewed as an operator with range in the weighted space $L^2(\mathbb{L}, \mathsf{Rf}_0)$ since

$$\langle \mathsf{R}_{\mathsf{f}_0}\mathsf{f}, \mathsf{g}\rangle_{L^2(\mathbb{L}(D),\mathsf{Rf}_0)} = \int_{\mathbb{L}(D)} \frac{\mathsf{Rf}(\mathbf{l})}{\mathsf{Rf}_0(\mathbf{l})}\mathsf{g}(\mathbf{l})\mathsf{Rf}_0(\mathbf{l})\, d\mathbf{l}$$

$$= \int_{\mathbb{L}(D)} \mathsf{Rf}(\mathbf{l})\mathsf{g}(\mathbf{l})\, d\mathbf{l}$$
$$= \int_D \mathsf{f}(\mathbf{x})\mathsf{R}^T\mathsf{g}(\mathbf{x})\, d\mathbf{x}$$

(cf. 3.2). The result then follows from a general theorem that says the orthogonal complement of the adjoint of a linear operator is equal to the range of the linear operator [Con90, secs. I.2, II.2]. This theorem is the analogue of the familiar fact from linear algebra that the range space of a real matrix is orthogonal to the nullspace of its transpose.

In summary, not all functions on $\mathbb{L}(D)$ are Radon transforms of functions on D. While the FB algorithm makes no use of this fact, the efficient estimator takes advantage of it by subtracting off components that are orthogonal to this subspace.

Remark 5.1 The efficient estimator at f_0 depends on f_0 since the projection operation depends on the weighting function Rf_0. Roughly speaking, this reflects the fact that Rf_0 is measured with a statistical uncertainty that varies over $\mathbb{L}$ with variance proportional to Rf_0. (Consider the usual discrete models for ET where the observations have a Poisson distribution.) The efficient estimator constructed above is not a practical estimator since the weighting function Rf_0 is not known *a priori.* One could construct a practical estimator by replacing Rf_0 with a suitable estimate. Of course, the estimator as a whole would then become nonlinear.

Remark 5.2 While the FB algorithm makes no use of the dependence of the statistical uncertainty of the observations on the underlying image, other algorithms, such as maximum likelihood and certain iterative approaches, do. The results here can be viewed as a way of quantifying how much improvement might be expected through use of this additional information.

Remark 5.3 We have just characterized how well the quantity $\mathsf{a}_{\mathbf{x}} \cdot \mathsf{f}$ may be estimated for a given aperture function a. In fact, the entire analysis remains valid if the function $\mathsf{a}_{\mathbf{x}}$ is replaced by any smooth function ϕ on $\mathbb{R}^2$. Defining $\psi \equiv 2^{-1}\mathsf{HR}\phi$, the linear

estimator generated by $p_{\mathcal{N}(\mathsf{R}^T)^\perp}\boldsymbol{\psi}$ is an efficient estimator of $\mathsf{a}_{\mathbf{x}} \cdot \boldsymbol{\phi}$ at f_0 with variance

$$n^{-1}\left(\pi^{-1}||p_{\mathcal{N}(\mathsf{R}^T)^\perp}\boldsymbol{\psi}||^2_{L^2(\mathbb{L}(D),\mathsf{Rf}_0)} - (\boldsymbol{\phi}\cdot\mathsf{f}_0)^2\right).$$

Thus our analysis quantifies how well a broad range of features in an image may be estimated.

6 Construction of Projection Operators

In section 5, we saw that the linear estimator generated by $p_{\mathcal{R}(\mathsf{R}_{\mathsf{f}_0})}\mathsf{b}_{\mathbf{x}} = p_{\mathcal{N}(\mathsf{R}^T)^\perp}\mathsf{b}_{\mathbf{x}}$ is an efficient estimator for $\mathsf{a}_{\mathbf{x}} \cdot \mathsf{f}$ at f_0 in the ET problem. In this section, we will express the projection operator $p_{\mathcal{R}(\mathsf{R}_{\mathsf{f}_0})}$ in a concrete way.

Recall from linear algebra that it is easy to compute the projection of a vector onto a subspace if one has an orthonormal basis for the subspace. The coefficients of the projection of a vector with respect to the orthonormal basis are just the inner products of the vector with the basis vectors. The same ideas apply to the construction of projection operators for $L^2(\mathbb{L}(D), \mathsf{Rf}_0)$ with the inner product defined in equation 5.3.

Definition 6.1 A subset of $L^2(\mathbb{L}(D), \mathsf{Rf}_0)$ is said to be orthonormal if each of its elements have unit norm and distinct elements are orthogonal to each other. An orthonormal subset $B \subset L^2(\mathbb{L}(D), \mathsf{Rf}_0)$ is said to be an orthonormal basis if any element in $L^2(\mathbb{L}(D), \mathsf{Rf}_0)$ can be expressed as a linear combination of elements in B. If $B = \{\mathsf{g}_i\}$ is an orthonormal basis for $L^2(\mathbb{L}(D), \mathsf{Rf}_0)$, then each $\mathsf{g} \in L^2(\mathbb{L}(D), \mathsf{Rf}_0)$ can be expressed as $\mathsf{g} = \sum_i \langle \mathsf{g}_i, \mathsf{g} \rangle_{L^2(\mathbb{L}(D),\mathsf{Rf}_0)} \mathsf{g}_i$ and $||\mathsf{g}||^2_{L^2(\mathbb{L}(D),\mathsf{Rf}_0)} = \sum_i \langle \mathsf{g}_i, \mathsf{g} \rangle^2_{L^2(\mathbb{L}(D),\mathsf{Rf}_0)}$.

We start in section 6.1 with the special case where f_0 is the uniform distribution f_u considered in example 4.1. It turns out that the analysis of this special case provides useful building blocks for the analysis of the general case, which is carried out in section 6.2.

6.1 The Uniform Distribution

We will now give orthonormal bases for $\mathcal{R}(\mathsf{R}_{\mathsf{f}_u})$ and $\mathcal{N}(\mathsf{R}^T)$ in $L^2(\mathbb{L}(D), \mathsf{Rf}_u)$. We begin by establishing some notation.

Definition 6.2 Let $\mathbb{N}$ and $\mathbb{N}^+$ denote the sets of nonnegative integers and positive integers, respectively. For $m \in \mathbb{N}$, define the functions $\mathsf{U}_m\colon [-1,1] \to \mathbb{R}$ by

$$\mathsf{U}_m(\cos\theta) \equiv \frac{\sin[(m+1)\theta]}{\sin\theta}.$$

The U_m are called the Chebyshev polynomials of the second kind [Dea83, sec. 7.6]. As the name implies, the U_m are indeed polynomials; U_m is a polynomial of order m. The first few are 1, $2s$, $4s^2-1$, and $8s^3-4s$. For $l \in \mathbb{N}^+$ and $m \in \mathbb{N}$ with $l+m$ even, define the functions $\mathsf{c}_{l,m} : \mathbb{L}(D) \to \mathbb{R}$ and $\mathsf{d}_{l,m} : \mathbb{L}(D) \to \mathbb{R}$ by

$$\mathsf{c}_{l,m}(\theta, s) \equiv \sqrt{2/\pi}\,\mathsf{U}_m(s)\cos(l\theta),$$

and

$$\mathsf{d}_{l,m}(\theta, s) \equiv \sqrt{2/\pi}\,\mathsf{U}_m(s)\sin(l\theta).$$

It turns out that

$$B_u \equiv \{\pi^{-1/2}\mathsf{U}_m : m \in 2\mathbb{N}\} \cup \{\mathsf{c}_{l,m}, \mathsf{d}_{l,m} : l \in \mathbb{N}^+, m \in l + 2\mathbb{N}\}$$

and

$$B'_u \equiv \{\mathsf{c}_{l,m}, \mathsf{d}_{l,m} : l \in \mathbb{N}^+,\ m \in \{l \bmod 2, l \bmod 2 + 2, \ldots, l-2\}\}.$$

are orthonormal bases for $\mathcal{R}(\mathsf{R}_{\mathsf{f}_u})$ and $\mathcal{N}(\mathsf{R}^T)$ in $L^2(\mathbb{L}(D), \mathsf{Rf}_u)$, respectively. A proof of this follows easily from some standard results on the Radon transform [Dea83, sec. 7.6]. The details may be found in [Kur95]. Thus $p_{\mathcal{R}(\mathsf{R}_{\mathsf{f}_0})}\mathsf{b}_{\mathbf{x}}$ has the expansion

$$p_{\mathcal{R}(\mathsf{R}_{\mathsf{f}_0})}\mathsf{b}_{\mathbf{x}} = \sum_{\mathsf{g}\in B} \langle \mathsf{g}, \mathsf{b}_{\mathbf{x}}\rangle_{L^2(\mathbb{L}(D),\mathsf{Rf}_u)} \tag{6.1}$$

and the variance of the linear estimator generated by $\mathsf{b}_{\mathbf{x}}$ is given by

$$n^{-1}\left(\sum_{\mathsf{g}\in B} \langle \mathsf{g}, \mathsf{b}_{\mathbf{x}}\rangle^2_{L^2(\mathbb{L}(D),\mathsf{Rf}_u)} - (\mathsf{b}_{\mathbf{x}} \cdot \mathsf{Rf}_u)^2\right), \tag{6.2}$$

where

$$\langle \mathsf{g}, \mathsf{b}_{\mathbf{x}}\rangle_{L^2(\mathbb{L}(D),\mathsf{Rf}_u)} = \frac{2}{\pi^2}\int_0^\pi \int_{-1}^1 \mathsf{g}(\theta, s)\mathsf{b}_{\mathbf{x}}(\theta, s)\sqrt{1-s^2}\,ds\,d\theta. \tag{6.3}$$

The variance of the FB algorithm at f_u is given by

$$n^{-1}\left(\sum_{\mathsf{g}\in B\cup B^{\perp}}\langle \mathsf{g},\mathsf{b}_{\mathbf{x}}\rangle^2_{L^2(\mathbb{L}(D),\mathsf{Rf}_u)}-(\mathsf{b}_{\mathbf{x}}\cdot\mathsf{Rf}_u)^2\right),$$

so the difference in variance at f_u between the FB and the efficient estimator is

$$n^{-1}\sum_{\mathsf{g}\in B^{\perp}}\langle \mathsf{g},\mathsf{b}_{\mathbf{x}}\rangle^2_{L^2(\mathbb{L}(D),\mathsf{Rf}_u)}. \tag{6.4}$$

Example 6.3 Continuing example 4.1, we consider the case of a Gaussian aperture function with $\sigma = 0.5$, $\mathsf{f}_0 = \mathsf{f}_u$, and $\mathbf{x} = (1,0) \in \mathbb{R}^2$. The FB estimate is generated by $\mathsf{b}_{\mathbf{x}}$, which is shown at the bottom right of figure 3. The efficient estimator at f_u is generated by $p_{\mathcal{R}(\mathsf{R}_{\mathsf{f}_u})}\mathsf{b}_{\mathbf{x}}$, which is illustrated at the left of figure 5. The difference $p_{\mathcal{N}(\mathsf{R}^T)}\mathsf{b}_{\mathbf{x}} = \mathsf{b}_{\mathbf{x}} - p_{\mathcal{R}(\mathsf{R}_{\mathsf{f}_u})}\mathsf{b}_{\mathbf{x}}$ is illustrated at the right of figure 5. In this particular case, the variance of the FB estimator is $0.087n^{-1}$ while the variance of the efficient estimator is $0.066n^{-1}$. Thus, in this case, the variance of the FB estimator is more than 30% higher than that of the efficient estimator.

We conclude this section by showing there is an important special case where the FB algorithm is efficient at f_u. This occurs when $\mathbf{x} = 0$, so that $\mathsf{a}_{\mathbf{x}} = \mathsf{a}$. To see this, recall from section 3 that we assume that the aperture function a is radial. It is then easy to verify that the observation space representation b_0 of a is independent of θ and can be written as an even function of s. It follows that, for $l + m \in 2\mathbb{N}^+$, the inner product $\langle \mathsf{c}_{l,m}, \mathsf{b}_0\rangle_{L^2(\mathbb{L}(D),\mathsf{Rf}_u)}$ reduces to

$$\frac{2}{\pi^2}\int_0^{\pi}\cos(l\theta)\,d\theta\int_{-1}^{1}\mathsf{U}_m(s)\mathsf{b}_0(s)\sqrt{1-s^2}\,ds = 0.$$

Similarly, $\langle \mathsf{d}_{l,m}, \mathsf{b}_0\rangle_{L^2(\mathbb{L}(D),\mathsf{Rf}_u)} = 0$. It follows that

$$\begin{aligned}\mathsf{b}_0 &= \sum_{m\in 2\mathbb{N}}\langle \mathsf{U}_m, \mathsf{b}_0\rangle_{L^2(\mathbb{L}(D),\mathsf{Rf}_u)}\mathsf{U}_m. \\ &= p_{\mathcal{R}(\mathsf{R}_{\mathsf{f}_u})}\mathsf{b}_0.\end{aligned} \tag{6.5}$$

6.2 The General Case

We now consider the general case where $\mathsf{f}_0 \neq \mathsf{f}_u$. The first thing to note is that the subspace $\mathcal{N}(\mathsf{R}^T)$ is defined independently of f_0. This means that B'_u will still be a basis for $\mathcal{N}(\mathsf{R}^T)$. However, it will not generally be an *orthonormal* basis in $L^2(\mathbb{L}(D), \mathsf{Rf}_0)$. To convert this basis to an orthonormal basis, it is necessary to apply the Gram-Schmidt procedure [Con90, I.4.6], which is analogous to the usual Gram-Schmidt orthonormalization procedure of linear algebra.

Similar considerations apply to finding an orthonormal basis for $\mathcal{R}(\mathsf{R}_{\mathsf{f}_0})$. The functions in $\mathcal{R}(\mathsf{R}_{\mathsf{f}_0})$ are of the form $\mathsf{Rf}/\mathsf{Rf}_0$. The functions in $\mathcal{R}(\mathsf{R}_{\mathsf{f}_u})$ are of the form $\mathsf{Rf}/\mathsf{Rf}_u$. Thus a basis for $\mathcal{R}(\mathsf{R}_{\mathsf{f}_0})$ is given by the functions of the form $\mathsf{g}\frac{\mathsf{Rf}_u}{\mathsf{Rf}_0}$ with $\mathsf{g} \in B_u$. Again, this basis will not generally be orthonormal in $L^2(\mathbb{L}(D), \mathsf{Rf}_0)$, but an orthonormal basis may be obtained by the Gram-Schmidt procedure.

Example 6.4 We consider the case of a Gaussian aperture function with $\sigma = 0.5$, $\mathsf{f}_0 = \mathsf{f}_u$, and . (This was illustrated at the top of figure 3.) In section 6.1, we saw that the FB estimator is efficient at $\mathbf{x} = 0$ for f_u. We now present an example that shows the FB estimator can be very suboptimal if $\mathsf{f}_0 \neq \mathsf{f}_u$. Suppose f_0 is highly concentrated about the point $(0, 0.63) \in D$. Then Rf_0 is highly concentrated about the curve $\theta \mapsto (\theta, 0.63 \sin\theta) \in \mathbb{L}$ and we can approximate the integral $\int_{\mathbb{L}(D)} \mathsf{g}(\theta, s)\mathsf{Rf}_0(\theta, s)\, ds\, d\theta$ by $\pi^{-1} \int_0^\pi \mathsf{g}(\theta, 0.63 \sin\theta)\, d\theta$ for any function g on $\mathbb{L}$. Using this approximation, the variance of the FB estimator at f_0 is $\approx 0.04683n^{-1}$. By the above results, the function $\mathsf{c}_{2,0}/||\mathsf{c}_{2,0}||_{L^2(\mathbb{L}(D),\mathsf{Rf}_0)}$ is a unit vector of $\mathcal{N}(\mathsf{R}^T)$ in $L^2(\mathbb{L}(D), \mathsf{Rf}_0)$. The squared inner product of this function with b_0 in $L^2(\mathbb{L}(D), \mathsf{Rf}_0)$ is ≈ 0.04594, which implies that the variance of the efficient estimator at f_0 is at most $0.00089n^{-1}$. Thus in this, admittedly extreme, example, the variance of the efficient estimator is less than the variance of the standard estimator by a factor of more than 50.

7 Discussion

To summarize, the fact that not all functions on the observation space are Radon transforms of functions on the image space means that ET is, in a sense, an overdetermined problem. To construct an efficient estimator, it is necessary to weight the information obtained from the observations according to its statistical uncertainty. The resulting estimator is analogous to a weighted least squares procedure. Moreover, since the statistical uncertainties depend on the unknown image, a practical estimator must estimate these uncertainties from the observations, making the overall estimation procedure nonlinear.

The results in this paper give best-possible lower bounds on the variance of unbiased estimators in ET. They can be used as a benchmark in assessing the performance of image reconstruction and quantification algorithms. Appropriately generalized, they can also be used as a design tool for assessing the performance that is achievable by new imaging devices.

The numerical results given in section 6 show that, at least in some cases, the efficient estimator has significantly less variance than the FB estimator. More extensive evaluation of the bound should help delineate the conditions under which significant improvement over FB is possible.

We want to say a few words about how one might construct a practical version of our efficient estimator. In practice, observations are usually collected as binned data, resulting in a discrete, finite-dimensional observation space. This finite-dimensional observation space decomposes into finite-dimensional analogs of $\mathcal{N}(\mathsf{T}^T)$ and $\mathcal{R}(\mathsf{T}_{\mathsf{f}_0})$. The analog of the efficient estimator would be obtained by projecting a finite-dimensional analog of $\mathsf{g}_{\mathsf{x}_0}$ onto the analog of $\mathcal{R}(\mathsf{T}_{\mathsf{f}_0})$. This projection operation amounts to a weighted least squares problem. Since the definition of $\mathsf{T}_{\mathsf{f}_0}$ and the projection operation depend on $\mathsf{T}\mathsf{f}_0$, we require an estimate of the analog of $\mathsf{T}\mathsf{f}_0$. The obvious estimate is the observation vector itself. We are currently investigating estimators based on this scheme.

Acknowledgements

The author thanks Ronald Huesman, Bryan Reutter, and Sundar Amartur for their helpful comments on preliminary drafts of this paper. This work was supported in part by the Director, Office of Energy Research, Office of Health and Environmental Research, Medical Applications and Biophysical Research Division of the U.S. Department of Energy under contract No. DE-AC03-76SF00098 and in part by NIH training grant HL07367 and National Heart, Lung, and Blood Institute grant P01-HL25840.

References

[Con90] J. B. Conway. *A Course in Functional Analysis.* Springer-Verlag, New York, 1990.

[Dea83] Stanley R. Deans. *The Radon Transform and Some of Its Applications.* John Wiley & Sons, New York, 1983.

[Her83] Alexander Hertle. Continuity of the Radon transform and its inverse on Euclidean space. *Math. Z.*, 184:165–192, 1983.

[JS90] Iain M. Johnstone and Bernard W. Silverman. Speed of estimation in positron emission tomography and related inverse problems. *Ann. Statist.*, 18:251–280, 1990.

[Kur95] Alvin Kuruc. Estimation of smooth integral functionals in emission tomography. Technical Report LBL-37032, Lawrence Berkeley Laboratory, Berkeley, CA, 1995.

[Lew83] Robert M. Lewitt. Reconstruction algorithms: Transform methods. *Proc. IEEE*, 71:390–408, 1983.

[Nat86] F. Natterer. *The Mathematics of Computed Tomography.* John Wiley & Sons, New York, 1986.

[Sla72] Lucy Joan Slater. Confluent hypergeometric functions. In Milton Abramowitz and Irene A. Stegun, editors, *Handbook of Mathematical Functions*, pages 503–535. Dover Publications, New York, 1972.

[Tre78] O. J. Tretiak. Noise limitations in X-ray computed tomography. *J. Comp. Asst. Tomo.*, 2:477–480, 1978.

[VSK85] Y. Vardi, L. A. Shepp, and L. Kaufman. A statistical model for positron emission tomography. *J. Amer. Stat. Assn.*, 80:8–20, 1985.

Index

A

$*$, 7

a, 9

$\mathsf{a}_{\mathbf{x}}$, 11

adjoint, 8

aperture function, 10

B

$\mathsf{b}_{\mathbf{x}}$, 11

B_u, 22

B'_u, 22

C

χ_σ, 12

$\mathsf{c}_{l,m}$, 22

convolution, 7

D

D, 5

$\cdot$, 7

δ, 11

$\mathsf{d}_{l,m}$, 22

E

η, 8

E_{f}, 13

efficient estimator, 15

efficient linear estimator, 15

ET, 3

F

f, 5

f_u, 14

f_0, 15

FB, 3

Fourier transform, 7

G

g, 7

H

H, 9

I

$\int_{\mathbb{L}}$, 7

$\langle \cdot, \cdot \rangle$, 17

i.i.d., 5

L

$\mathbb{L}$, 5

$\mathbf{l}$, 5

$\mathbb{L}(D)$, 16

L^2, 17

N

$|\cdot|$, 10

n, 11

$\mathcal{N}$, 17

$||\cdot||$, 17

$\mathbb{N}$, 22

$\mathbb{N}^+$, 22

O

orthogonal, 17

orthogonal complement, 17

orthonormal, 21

orthonormal basis, 21

P

$\mathbf{\Phi}$, 12

$\perp$, 17

p, 18

p.d.f., 5

planar imaging, 12

R

$\mathbb{R}^2$, 5

R, 5

$\mathcal{R}$, 18

$\mathsf{R}_{\mathsf{f}_0}$, 18

Radon transform, 5

S

s, 7

σ, 10

T

$\boldsymbol{\theta}$, 7

θ, 7

$\tilde{\cdot}$, 7

T, 8

U

U_m, 22

W

w, 9

X

$\mathbf{x}$, 5

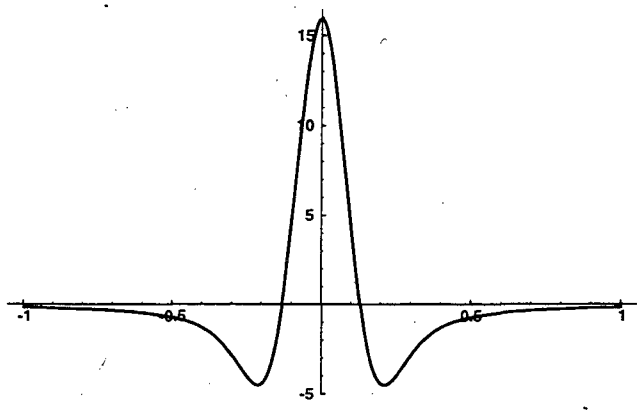

Figure 1: The function χ_σ with $\sigma = 0.1$

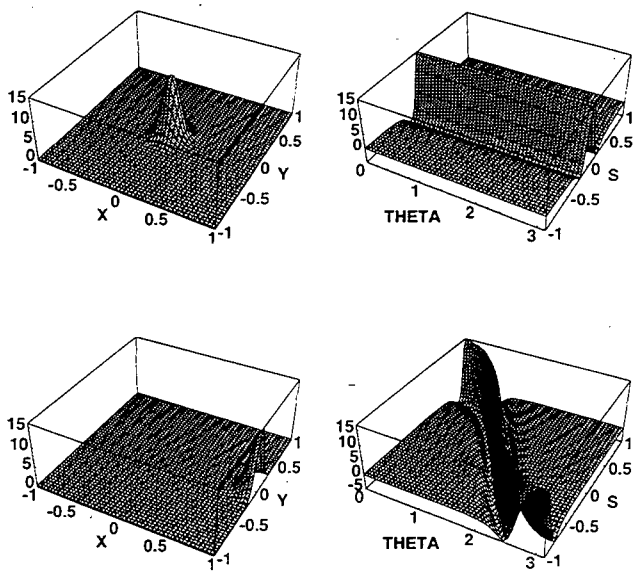

Figure 2: Observation-space representations of the FB estimator with a Gaussian aperture function. The aperture function translated by $\mathbf{x}_0$ is shown on the left and its observation-space representation $\mathbf{b}_{\mathbf{x}_0}$ is shown on the right. The upper pair is for $\mathbf{x}_0 = (0, 0)$ while the lower pair is for $\mathbf{x}_0 = (1, 0)$. For both pairs, $\sigma = 0.1$.

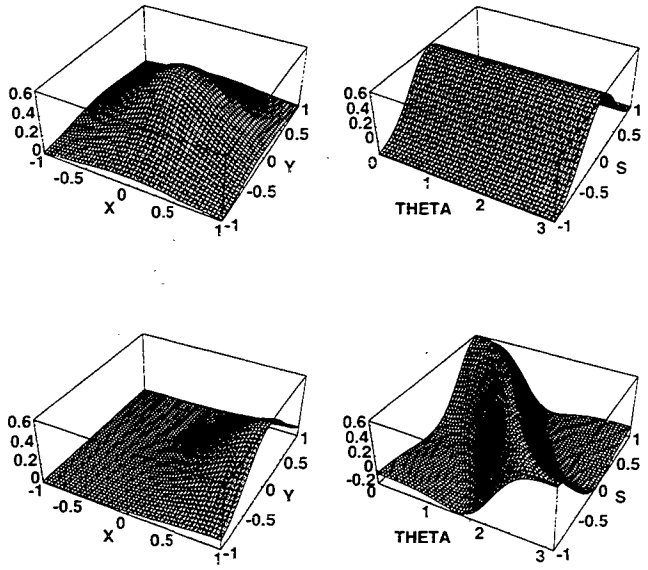

Figure 3: Observation-space representations of the FB estimator with a Gaussian aperture function. Everything is as in figure 2, except that $\sigma = 0.5$.

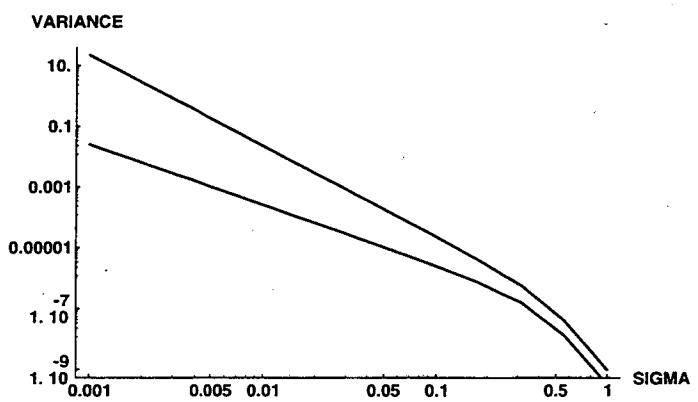

Figure 4: Variance of FB with Gaussian aperture function at the origin for uniform distribution with 10^6 photon pairs (upper curve). Lower curve is variance for planar imaging.

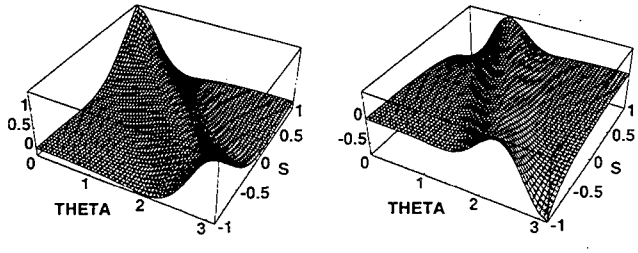

Figure 5: The function which generates the efficient estimator is shown on the left. The difference between the generators of the FB and the efficient estimators is shown on the right.

LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
TECHNICAL INFORMATION DEPARTMENT
BERKELEY, CALIFORNIA 94720