Image Reconstruction from Truncated Data in Single-Photon Emission Computed Tomomgraphy with Uniform Attenuation

Frederic Noo, Michel Defrise, Jed Pack, Rolf Clackdoyle

IOP PUBLISHING INVERSE PROBLEMS

Inverse Problems **23** (2007) 645–667 doi:10.1088/0266-5611/23/2/011

Image reconstruction from truncated data in single-photon emission computed tomography with uniform attenuation

Frédéric Noo[1], Michel Defrise[2], Jed D Pack[1,3] and Rolf Clackdoyle[4]

[1] UCAIR, Department of Radiology, University of Utah, CAMT Building, 729 Arapeen Drive, Salt Lake City, Utah 84108, USA
[2] Department of Nuclear Medicine, Vrije Universiteit Brussels, Belgium
[3] CT Systems and Applications Lab, GE Global Research Center, New York, USA
[4] Laboratoire TSI, CNRS-UMR5516, Saint Etienne, France

E-mail: noo@ucair.med.utah.edu

Received 7 July 2006, in final form 29 December 2006
Published 8 March 2007
Online at stacks.iop.org/IP/23/645

Abstract

We present a mathematical analysis of the problem of image reconstruction from truncated data in two-dimensional (2D) single-photon emission computed tomography (SPECT). Recent results in classical tomography have shown that accurate reconstruction of some parts of the object is possible in the presence of truncation. We have investigated how these results extend to 2D parallel-beam SPECT, assuming that the attenuation map is known and constant in a convex region Ω that includes all activity sources. Our main result is a proof that, just like in classical tomography accurate SPECT reconstruction at a given location $\underline{x} \in \Omega$, does not require the data on all lines passing through Ω; some amount of truncation can be tolerated. Experimental reconstruction results based on computer-simulated data are given in support of the theory.

1. Introduction

Recently, several works have shown that classical two-dimensional (2D) tomography is not 'all or nothing' contrary to long-standing folklore [1–4]. That is, the measured data need not cover the whole object for accurate reconstruction to be possible in some regions-of-interest (ROIs). Or, in other words, each parallel-beam projection may be truncated without preventing accurate reconstruction of some ROIs. These statements hold for any finite-size object, and this paper investigates how they may be extended to single-photon emission computed tomography (SPECT).

In an idealized model, the 2D SPECT reconstruction problem amounts to finding a (tracer-activity) function $p(x, y)$ or $p(\underline{x})$ with $\underline{x} = (x, y)$ from its attenuated Radon transform

$$(A_a p)(\phi, s) = \int_{-\infty}^{\infty} p(s\underline{\theta} + t\underline{\theta}^{\perp})\, \mathrm{e}^{-\int_t^{\infty} a(s\underline{\theta}+t'\underline{\theta}^{\perp})\,\mathrm{d}t'}\,\mathrm{d}t, \tag{1}$$

where a is a known function, referred to as the attenuation map, $\underline{\theta} = (\cos\phi, \sin\phi)$ and $\underline{\theta}^{\perp} = (-\sin\phi, \cos\phi)$. For physical reasons, the support of p is bounded. Therefore, the integrand in (1) is non-zero over a bounded interval; this integral is written over $(-\infty, +\infty)$ for notational convenience.

In this paper, the attenuation map is assumed to be a constant μ_o over a convex region Ω that includes all the activity, that is a convex region outside which p is zero. Under this assumption, the SPECT reconstruction problem reduces to finding $p(\underline{x})$ from its exponential Radon transform

$$(E_{\mu_o} p)(\phi, s) = \int_{-\infty}^{\infty} p(s\underline{\theta} + t\underline{\theta}^{\perp})\, \mathrm{e}^{\mu_o t}\,\mathrm{d}t. \tag{2}$$

This reduction is achieved through a simple weighting of $(A_a p)(\phi, s)$. See [5–9] for details. Geometrically, $(E_{\mu_o} p)(\phi, s)$, just like $(A_a p)(\phi, s)$, represents a weighted integral of p along the line of direction $\underline{\theta}^{\perp}$ at signed distance s from the origin.

For a fixed ϕ, the values of $(E_{\mu_o} p)(\phi, s)$ obtained by varying s define an exponential parallel-beam projection of p. If these values are only known for a limited range of s, the projection is said to be truncated. For example, if Ω is the centred disc of radius R, the projection is truncated as long as $(E_{\mu_o} p)(\phi, s)$ is not known for all $s \in [-R, R]$. In practice, a projection may be considered known over a given interval whenever $(E_{\mu_o} p)(\phi, s)$ is measured over this interval with a sampling distance that is in agreement with the resolution targeted for p.

Until 2001, theoretically exact and stable reconstruction of p from its exponential Radon transform was investigated under the assumption that the projections of p are non-truncated and known over 360°. Various reconstruction methods were developed under this assumption, see for example [5–17]. A nice theory unifying most of these methods in a common framework was presented in [18, 19]. The 360° of data seemed to be necessary because the exponential Radon transform does not present the parity property of the Radon transform (in general, $(E_{\mu_o} p)(\phi + \pi, -s) \neq (E_{\mu_o} p)(\phi, s)$ for $\mu_o \neq 0$). However, in 2001, accurate reconstruction from non-truncated projections known only over 180° was shown to be possible for any value of μ_o [20]. The results in [20] were extended to various π-scheme data acquisitions in [21, 22], which allow the 180° of data to be distributed over a union of disjoint intervals. Furthermore, an explicit inversion formula was suggested in [23].

In this paper, we investigate image reconstruction from exponential parallel-beam projections that are only known over 180° and may each be truncated. We build this investigation on Rullgård's results in [23]. The angular range over which the projections are known is seen as a single interval, not a union of disjoint intervals, and, for convenience, this interval is chosen as $\phi \in [0, \pi]$. Given a truncation pattern, we examine where accurate reconstruction is possible inside Ω. We assume Ω is not only convex but also bounded.

The paper is organized in five sections. In section 2, we first reduce the inversion of the exponential Radon transform with truncated projections to the solution of a 1D integral equation. Then, we set up a condition under which this integral equation admits a unique and stable inverse and we present in section 3 a numerical procedure allowing verification that this condition is met for arbitrary values of μ_o. The theory is supported by simulation results in the SPECT context, which are given in section 4. Finally, section 5 summarizes our findings and discusses various related aspects.

2. Theory

To achieve SPECT reconstruction with truncation, we follow the steps of the theory developed in [2] for the non-attenuated case with truncated data. First, we introduce a differentiated backprojection (DBP) operation and show that this DBP yields a one-dimensional (1D) integral equation for the unknown p. Then, we investigate the inversion of this integral equation.

2.1. Differentiated backprojection

Usually, the backprojection operation in computed tomography represents the adjoint operation of the projection equation describing the data. For the exponential Radon transform, this (unweighted) adjoint is

$$\int_0^\pi \mathrm{e}^{\mu_o \underline{x}\cdot\underline{\theta}^\perp}(E_{\mu_o}p)(\phi, \underline{x}\cdot\underline{\theta})\,\mathrm{d}\theta. \tag{3}$$

However, Tretiak and Metz [8] and Gullberg and Budinger [5] noted that replacing μ_o by $-\mu_o$ in (3) is a key step in developing an exact reconstruction formula. Noting this point, we define the differentiated backprojection for the exponential Radon transform as

$$b(\underline{x}) = \int_0^\pi \mathrm{e}^{-\mu_o \underline{x}\cdot\underline{\theta}^\perp}(E'_{\mu_o}p)(\phi, \underline{x}\cdot\underline{\theta})\,\mathrm{d}\phi \tag{4}$$

with

$$(E'_{\mu_o}p)(\phi, s) = \frac{\partial}{\partial s}(E_{\mu_o}p)(\phi, s). \tag{5}$$

To guarantee the existence of this DBP, we assume that p is continuously differentiable. Note that our definition is consistent with the concept of DBP in the non-attenuated case: when μ_o is equal to zero, (4) reduces to the DBP expression in [2].

2.2. Link between the DBP and the image function p

We start from (5). Replacing $(E_{\mu_o}p)(\phi, s)$ by its integral expression (2), moving the partial derivative with respect to s within the integral and applying the chain rule, we obtain

$$(E'_{\mu_o}p)(\phi, s) = \int_{-\infty}^{\infty} \mathrm{e}^{\mu_o t'}\underline{\theta}\cdot(\vec{\nabla}p)(s\underline{\theta} + t'\underline{\theta}^\perp)\,\mathrm{d}t'. \tag{6}$$

Therefore, since $\underline{x} = (\underline{x}\cdot\underline{\theta})\underline{\theta} + (\underline{x}\cdot\underline{\theta}^\perp)\underline{\theta}^\perp$,

$$\begin{aligned}(E'_{\mu_o}p)(\phi, \underline{x}\cdot\underline{\theta}) &= \int_{-\infty}^{\infty} \mathrm{e}^{\mu_o t'}\underline{\theta}\cdot(\vec{\nabla}p)((\underline{x}\cdot\underline{\theta})\underline{\theta} + t'\underline{\theta}^\perp)\,\mathrm{d}t' \\ &= \int_{-\infty}^{\infty} \mathrm{e}^{\mu_o t'}\underline{\theta}\cdot(\vec{\nabla}p)(\underline{x} + (t' - \underline{x}\cdot\underline{\theta}^\perp)\underline{\theta}^\perp)\,\mathrm{d}t'.\end{aligned} \tag{7}$$

Or, equivalently, with $t = t' - \underline{x}\cdot\underline{\theta}^\perp$, and using a second time the chain rule,

$$\begin{aligned}(E'_{\mu_o}p)(\phi, \underline{x}\cdot\underline{\theta}) &= \mathrm{e}^{\mu_o \underline{x}\cdot\underline{\theta}^\perp}\int_{-\infty}^{\infty} \mathrm{e}^{\mu_o t}\underline{\theta}\cdot(\vec{\nabla}p)(\underline{x} + t\underline{\theta}^\perp)\,\mathrm{d}t \\ &= -\mathrm{e}^{\mu_o \underline{x}\cdot\underline{\theta}^\perp}\int_{-\infty}^{\infty} \frac{\mathrm{e}^{\mu_o t}}{t}\frac{\partial}{\partial\phi}\{p(\underline{x} + t\underline{\theta}^\perp)\}\,\mathrm{d}t.\end{aligned} \tag{8}$$

The insertion of this result in the DBP (4) with a change in the order of integration (ϕ first, t second) yields the sought link:

$$b(\underline{x}) = -\int_{-\infty}^{\infty} \frac{e^{\mu_o t}}{t} [p(\underline{x} + t\underline{\theta}^{\perp})]_0^{\pi} \, dt$$

$$= -\int_{-\infty}^{\infty} \frac{e^{\mu_o t}}{t} (p(x, y - t) - p(x, y + t)) \, dt \tag{9}$$

$$= -\fint_{-\infty}^{\infty} \frac{e^{\mu_o t}}{t} p(x, y - t) \, dt + \fint_{-\infty}^{\infty} \frac{e^{\mu_o t}}{t} p(x, y + t) \, dt \tag{10}$$

with $\underline{x} = (x, y)$.

Note the vertically centred horizontal bar over the integral symbols in (10). Each integral in (10) is only meaningful in the sense of a Cauchy principal value, that is (10) should be understood as

$$b(\underline{x}) = -\lim_{\varepsilon \to 0, \varepsilon > 0} \int_{|t| > \varepsilon} \frac{e^{\mu_o t}}{t} p(x, y - t) \, dt + \lim_{\varepsilon \to 0, \varepsilon > 0} \int_{|t| > \varepsilon} \frac{e^{\mu_o t}}{t} p(x, y + t) \, dt. \tag{11}$$

The bar over each integral symbol in (10) emphasizes this understanding. On the other hand, no bar was needed over the integral symbol in (9) because the numerator in (9) behaves like a multiple of t near $t = 0$, so that the integrand in (9) is not singular at $t = 0$.

Equation (10) is easily rewritten as

$$b(x, y) = -2\pi \fint_{-\infty}^{\infty} \frac{\cosh \mu_o (y - y')}{\pi (y - y')} p(x, y') \, dy'. \tag{12}$$

We see thus that $b(x, y)$ is (up to a factor of -2π) a 1D convolution of $p(x, y)$ with $\cosh(\mu_o y)/(\pi y)$ at fixed x. For μ_o tending towards 0, this convolution kernel reduces to $1/(\pi y)$, and $b(x, y)$ becomes (up to the factor of -2π) the Hilbert transform of $p(x, y)$ at fixed x, in agreement with the results in [2].

In practice, (12) is only of interest for $(x, y) \in \Omega$, the bounded and convex region outside which p is known to be zero. Let (x, L_x) and (x, U_x) be the end points of the intersection of Ω with the line parallel to the y-axis through a point $(x, y) \in \Omega$. Using this notation, we have

$$b(x, y) = -2\pi \fint_{L_x}^{U_x} \frac{\cosh \mu_o (y - y')}{\pi (y - y')} p(x, y') \, dy' \qquad \text{for} \quad (x, y) \in \Omega. \tag{13}$$

This expression is preferred to (12) because it shows explicitly that the integral in y' has finite bounds. For example, if Ω is a centred disc of radius R, $U_x = -L_x = \sqrt{R^2 - x^2}$.

Equation (13) reduces the reconstruction problem to solving a 1D integral equation with a convolution kernel. However, note that p is not readily found from b using a simple 1D deconvolution, because the convolution kernel $\cosh(\mu_o y)/(\pi y)$ grows exponentially with increasing y.

At this stage, we would like to emphasize that (13) together with (4) is not an original contribution of this paper. Rullgård [23] recently developed this result, albeit using a somewhat different notation. Furthermore, Rullgård [23] showed that there exists a generalized function (distribution) ρ such that

$$p(x, y) = \fint_{y - D_x}^{y + D_x} \rho(y - y') b(x, y') \, dy' \qquad \text{for} \quad (x, y) \in \Omega, \tag{14}$$

where $D_x = U_x - L_x$. Note that the limits of integration in (14) far extend beyond the $[L_x, U_x]$ interval. A method to obtain a polynomial approximation of ρ was also described in [23].

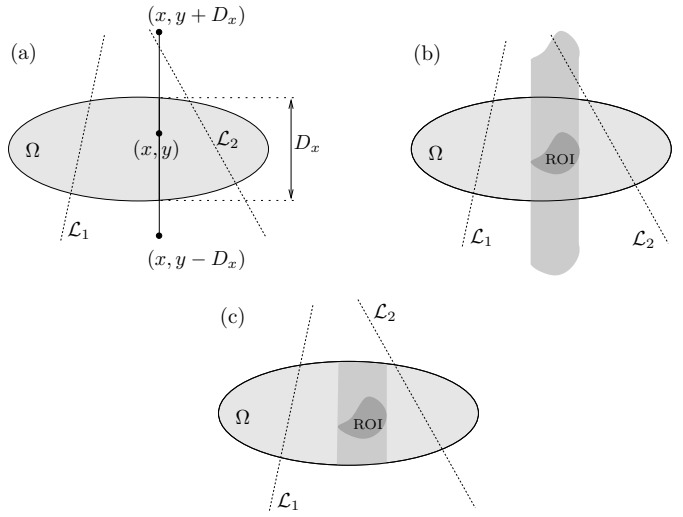

Figure 1. Illustration of allowed data truncation. The activity sources are within the ellipse Ω. (a) According to Rullgård's results [23], accurate reconstruction at (x, y) only requires the data on the lines that cross a neighbourhood of the vertical line segment through (x, y). Line $\mathcal{L}_2$ is thus needed but not line $\mathcal{L}_1$. (b) According to Rullgård's results [23], accurate reconstruction within the depicted ROI requires the data on all lines passing through the shaded vertical strip. (c) According to our results, accurate reconstruction within the depicted ROI only requires the data on all lines passing through the intersection of the shaded vertical strip with Ω. In particular, line $\mathcal{L}_2$ is not needed.

Although not pointed out in [23], (14) represents a first proof that reconstruction of an ROI in SPECT with uniform attenuation does not require knowledge of $E_{\mu_o} p$ on all lines passing through the region Ω where the activity lies. To obtain p at a given location $(x, y) \in \Omega$ according to (14), only the values of $E_{\mu_o} p$ required to compute $b(x, y')$ with $|y' - y| < D_x$ are needed. These values correspond to the lines that meet a neighbourhood of the line segment connecting $(x, y - D_x)$ to $(x, y + D_x)$, hence $E_{\mu_o} p$ need not be available for all lines passing through the activity region. See figure 1(a) for an illustration of this assertion and figure 1(b) for its extension to an ROI.

The methodology outlined in equations (4)–(14) for accurate reconstruction from truncated data is conceptually similar to the approach developed in [2] for classical tomography. However, we observe a fundamental difference in terms of allowed data truncation: following the theory in [2] when $\mu_o = 0$ only the values of $E_{\mu_o} p$ for the lines that intersect a neighbourhood of the line segment from (x, L_x) to (x, U_x) are needed to accurately reconstruct $p(x, y)$. We improve on Rullgård's result, equation (14), by discarding the convolution and relaxing the measurement requirement. Specifically, we show that

> A function p that is continuously differentiable on Ω can be accurately reconstructed at a given location $(x, y) \in \Omega$, if for each $\phi \in [0, \pi]$ the values of $E_{\mu_o} p$ are known (measured) for just those lines that intersect a neighbourhood of the line segment from (x, L_x) to (x, U_x).

This statement holds for any value of μ_o that satisfies a condition given in section 2.4 (equation (45)). Numerical evaluation of this condition reveals that it is satisfied for a fine sampling of a wide range of values of μ_o that effectively covers all medical applications of

SPECT. See figure 1(c) for a visual illustration of the statement in comparison with truncation allowed by Rullgård's formula.

To prove our statement, the inversion problem is recast in the following more compact form: reconstruct a function $f(\tau)$ with $|\tau| \leqslant 1$ assuming that

$$g(t) = \int\!\!\!\!\!\!-_{-1}^{1} \frac{\cosh \mu(t-\tau)}{\pi(t-\tau)} f(\tau)\, \mathrm{d}\tau \tag{15}$$

is known for $|t| \leqslant 1$ and that

$$m_\mu = \int_{-1}^{1} f(\tau) \cosh \mu\tau \, \mathrm{d}\tau \tag{16}$$

is also a known quantity, where μ is some constant. Solving this more compact problem implies proving our statement because, first, (13) is easily cast into (15) through the affine change of variables that maps the interval $[L_x, U_x]$ into $[-1, 1]$ and yields

$$\mu = \mu_o \frac{U_x - L_x}{2}, \tag{17}$$

$$f(\tau) = p\left(x, \frac{U_x - L_x}{2}\tau + \frac{U_x + L_x}{2}\right), \tag{18}$$

$$g(t) = \frac{1}{\pi(L_x - U_x)} b\left(x, \frac{U_x - L_x}{2}t + \frac{U_x + L_x}{2}\right). \tag{19}$$

Second, m_μ may be considered as a known quantity because $(E_{\mu_o} p)(0, x)$ and $(E_{\mu_o} p)(\pi, -x)$ are part of the data we require for reconstruction at $(x, y) \in \Omega$ and

$$m_\mu = \frac{2}{U_x - L_x}\left\{\mathrm{e}^{-\mu_o \frac{U_x + L_x}{2}} (E_{\mu_o} p(0, x)) + \mathrm{e}^{\mu_o \frac{U_x + L_x}{2}} (E_{\mu_o} p(\pi, -x))\right\}. \tag{20}$$

In medical applications of SPECT, the value of μ_o depends both on the radionuclide and on the type of tissue. For the most common nuclide used in SPECT, ^{99m}Tc, typical values of μ_o range between 0.15 cm^{-1} for soft tissues such as muscle and 0.21 cm^{-1} for bones. For a head scan the diameter of Ω would be about 20 cm, yielding $\mu = 1.5$ for $\mu_o = 0.15$ cm^{-1}. For a cardiac scan, the diameter of Ω might be as large as 80 cm, leading to $\mu = 6$ for $\mu_0 = 0.15$ cm^{-1}.

2.3. Conversion to a Fredholm equation of the second kind

We convert the problem defined by equations (15) and (16) to a Fredholm equation of the second kind, as explained below.

Let $g_0(t)$ be the expression of $g(t)$ in the particular case where $\mu = 0$. According to Söhngen [24] and Tricomi [25],

$$f(t) = \frac{1}{\sqrt{1-t^2}}\left\{-\int\!\!\!\!\!\!-_{-1}^{1} \frac{1}{\pi(t-\rho)} \sqrt{1-\rho^2} g_0(\rho)\, \mathrm{d}\rho + \frac{m_0}{\pi}\right\} \tag{21}$$

for $|t| < 1$, with

$$m_0 = \int_{-1}^{1} f(\tau)\, \mathrm{d}\tau. \tag{22}$$

Also, from (15),

$$g_0(\rho) = g(\rho) - \mu \int_{-1}^{1} c(\mu(\rho - \tau)) f(\tau)\, \mathrm{d}\tau \tag{23}$$

with

$$c(q) = \begin{cases} (\cosh q - 1)/(\pi q) & \text{if } q \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

Note that there is no Cauchy principal value symbol in the second term of the right-hand side of (23) because $c(q)$ behaves like $q/(2\pi)$ near $q = 0$, unlike $(\cosh \mu q)/(\pi q)$ in (15), which behaves like $1/(\pi q)$ near $q = 0$. Also note that $c(q)$ is a smooth function. We insert equation (23) into (21). This yields

$$f(t)\sqrt{1-t^2} = -\fint_{-1}^{1} \frac{1}{\pi(t-\rho)}\sqrt{1-\rho^2}g(\rho)\,\mathrm{d}\rho + \frac{m_0}{\pi} + \mu \int_{-1}^{1} \mathrm{d}\rho \frac{\sqrt{1-\rho^2}}{\pi(t-\rho)} \int_{-1}^{1} \mathrm{d}\tau f(\tau) c(\mu(\rho-\tau)). \tag{25}$$

Or, switching the order of integration in the last term (see appendix A) and introducing intermediate quantities,

$$h(t) = h_d(t) + \frac{m_0}{\pi} + \int_{-1}^{1} h(\tau) \frac{1}{\sqrt{1-\tau^2}} k_\mu(t,\tau)\,\mathrm{d}\tau, \tag{26}$$

with

$$h(t) = f(t)\sqrt{1-t^2}, \tag{27}$$

$$h_d(t) = -\fint_{-1}^{1} \frac{1}{\pi(t-\rho)}\sqrt{1-\rho^2}g(\rho)\,\mathrm{d}\rho, \tag{28}$$

$$k_\mu(t,\tau) = \mu \fint_{-1}^{1} \frac{\sqrt{1-\rho^2}}{\pi(t-\rho)} c(\mu(\rho-\tau))\,\mathrm{d}\rho. \tag{29}$$

Equation (26) is a Fredholm integral equation of the second kind for h, with h_d directly obtainable from the data. However, this equation is not completely satisfactory because m_0 is only known when $\mu = 0$. To circumvent the problem, we replace m_0 by the quantity m_μ of equation (16), incorporating the difference between m_0 and m_μ within the integral kernel. In operator notation, the integral equation becomes

$$h = h_d + \frac{1}{\pi} m_\mu + Kh, \tag{30}$$

with

$$(Kh)(t) = \int_{-1}^{1} h(\tau) \frac{1}{\sqrt{1-\tau^2}} \bar{k}_\mu(t,\tau)\,\mathrm{d}\tau, \tag{31}$$

where

$$\bar{k}_\mu(t,\tau) = k_\mu(t,\tau) + \frac{1}{\pi}(1 - \cosh \mu\tau). \tag{32}$$

In this equation, the two 'forcing terms', h_d and m_μ, are now both known quantities.

Recall from, e.g., [25] that

$$\fint_{-1}^{1} \frac{\sqrt{1-\rho^2}}{\pi(t-\rho)}\,\mathrm{d}\rho = t \qquad \text{if} \quad |t| \leqslant 1. \tag{33}$$

Using this result, $k_\mu(t,\tau)$ in (29) can be rewritten in the following form:

$$k_\mu(t,\tau) = \mu t c(\mu(t-\tau)) + \mu \int_{-1}^{1} \sqrt{1-\rho^2} \frac{c(\mu(\rho-\tau)) - c(\mu(t-\tau))}{\pi(t-\rho)}\,\mathrm{d}\rho \qquad \text{if} \quad |t| \leqslant 1. \tag{34}$$

This expression does not include a Cauchy principal value anymore and shows us since $c(r)$ is a smooth function that $k_\mu(t,\tau)$ and therefore $\bar{k}_\mu(t,\tau)$ in (32) are continuous over the region $(t,\tau)\in[-1,1]\times[-1,1]$.

2.4. Solution of the integral equation

Equation (30) is not easier to explicitly solve than (15) but the fact that the singular Cauchy kernel in (15) has been replaced by a continuous Fredholm kernel (equation (32)) greatly simplifies the analysis of existence, uniqueness and stability.

First, we observe that if p and therefore f are continuously differentiable, then h and h_d in equations (27) and (28) are in the Hilbert space $H=L^2_w(-1,1)$ with norm

$$\|h\|^2=\int_{-1}^{1}\frac{|h(t)|^2}{\sqrt{1-t^2}}\,\mathrm{d}t. \tag{35}$$

Furthermore, the operator K defined by (31) is compact in H, because the kernel (32) is continuous and thus square integrable, that is

$$\int_{-1}^{1}\mathrm{d}t\int_{-1}^{1}\mathrm{d}\tau\,\frac{1}{\sqrt{1-\tau^2}}\frac{1}{\sqrt{1-t^2}}|\bar{k}_\mu(t,\tau)|^2<\infty. \tag{36}$$

See for instance chapter 9 in [26] for a proof of compactness for operators with square integrable kernels.

Given the above observation, we analyse equation (30) in H, applying the Fredholm alternative theorem [27]. This theorem guarantees that (30) admits for any h_d+m_μ/π a unique solution that depends continuously on h_d+m_μ/π provided the parameter μ is such that the homogeneous equation $Kh=h$ has only the trivial solution $h=0$. In other words, the inverse operator $(I-K)^{-1}$ exists and is bounded when $Kh=h$ implies $h=0$. The bound $\|(I-K)^{-1}\|=C$ is such that

$$\|(I-K)^{-1}h\|\leqslant C\|h\| \qquad \text{for any} \quad h\in H \tag{37}$$

or, equivalently,

$$\|(I-K)h\|\geqslant C^{-1}\|h\| \qquad \text{for any} \quad h\in H. \tag{38}$$

Inequality (37) is important because it determines the stability property of the solution to (30). Specifically, suppose only some approximation $h_d^\varepsilon+m_\mu^\varepsilon/\pi$ of the forcing term h_d+m_μ/π is known, with

$$\left\|\left(h_d^\varepsilon+\frac{1}{\pi}m_\mu^\varepsilon\right)-\left(h_d+\frac{1}{\pi}m_\mu\right)\right\|\leqslant\varepsilon \tag{39}$$

for some $\varepsilon>0$. Then, the solution h^ε corresponding to $h_d^\varepsilon+m_\mu^\varepsilon/\pi$ approximates the unknown h as follows:

$$\|h^\varepsilon-h\|=\left\|(I-K)^{-1}\left(\left(h_d^\varepsilon+\frac{1}{\pi}m_\mu^\varepsilon\right)-\left(h_d+\frac{1}{\pi}m_\mu\right)\right)\right\|\leqslant C\varepsilon. \tag{40}$$

Assume a strictly positive quantity $\hat{C}$ is found such that

$$\|(I-K)h\|\geqslant\hat{C}^{-1}\|h\| \qquad \text{for any} \quad h\in H. \tag{41}$$

Then, the equation $Kh=h$ can only have $h=0$ for solution and, from the Fredholm alternative theorem, the inverse $(I-K)^{-1}$ exists and is bounded with $C\leqslant\hat{C}$. Hence, the existence and uniqueness of the solution to (30) and the stability of this solution in the presence of noise are all ensured by finding a strictly positive quantity $\hat{C}$ satisfying (41).

A quantity such as $\hat{C}$ may be obtained through a decomposition of K into the sum of two terms K_M and R_M, where K_M is a compact operator that approximates K and R_M is the difference between K and K_M. Suppose we can find positive numbers A_M and B_M such that for any $h \in H$

$$\|(I - K_M)h\| \geqslant A_M\|h\| \tag{42}$$

$$\|R_M h\| \leqslant B_M\|h\|. \tag{43}$$

The numbers A_M^{-1} and B_M represent, respectively, an upper bound on the norm of the operator $(I - K_M)^{-1}$ and on the norm of the approximation error R_M. Then, for any $h \in H$,

$$\|(I - K)h\| \geqslant \|(I - K_M)h\| - \|R_M h\| \geqslant (A_M - B_M)\|h\|, \tag{44}$$

which is similar to (41). Thus, provided $A_M > B_M$, equations (38) and (40) are satisfied with

$$C \leqslant \frac{1}{A_M - B_M}. \tag{45}$$

In section 3, we introduce a degenerate-kernel approximation K_M of K that comes with appropriate numbers A_M and B_M, allowing us to numerically verify existence, uniqueness and stability, according to the description above.

3. Numerical analysis

Successive approximations, degenerate-kernel approximation, quadrature methods and projection methods are standard textbook methods for the numerical solution of integral equations [27, 28]. The method using degenerate-kernel approximation is the simplest in terms of implementation and error analysis, but is often less efficient than either the quadrature method or the projection method. However, for kernels involving hyperbolic functions, Taylor expansions provide in some cases a way to achieve an efficient degenerate-kernel approximation. We report here on such an approximation and its use for the analysis of the inversion of (30). More specifically, we first explain below how Taylor expansions can be used to obtain a degenerate-kernel approximation for K. Next, we show that the combination of this approximation with the strategy outlined in section 2.4 yields a numerical method to check whether (30) has a unique and stable solution for a given μ. We have applied this method and proved thereby that (30) has a unique and stable solution for a fine sampling of μ that effectively covers all medical applications of SPECT.

3.1. Series expansion for K

Using Taylor series expansions as explained in appendix B, we found the following series expression for the kernel of K in equation (32):

$$\overline{k}_\mu(t, \tau) = \sum_{n=0}^{\infty} \frac{(-\mu)^n}{\pi n!} a_n(t) r_n(\mu\tau) \qquad \text{for} \quad (t, \tau) \in [-1, 1] \times [-1, 1] \tag{46}$$

where $a_n(t)$ and $r_n(\mu\tau)$ are both functions exhibiting attractive properties.

The function $a_n(t)$ is a polynomial of degree n that can be obtained by recurrence according to the relation

$$a_n(t) = t a_{n-1}(t) - \eta_{n-1} \qquad \text{for} \quad n \geqslant 1, \tag{47}$$

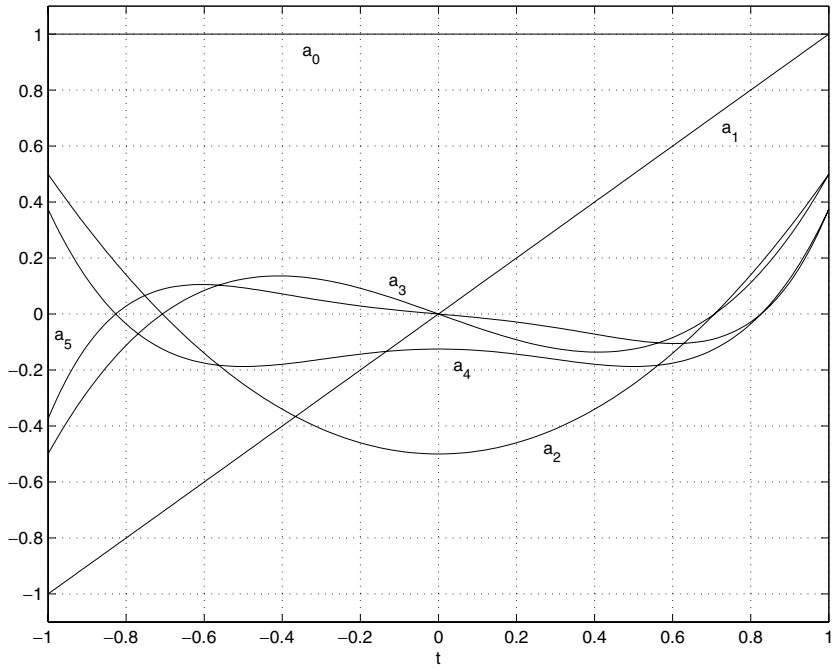

Figure 2. Plot of polynomials $a_n(t)$ for $n = 0, \ldots, 5$. The interval $|t| \leqslant 1$ is the region for which all $a_n(t)$ are required for reconstruction, independent of μ.

with $a_0(t) = 1$ and with $\eta_0 = 0$, $\eta_1 = 1/2$ and $\eta_{n+2} = n\eta_n/(n+3)$ for $n \geqslant 0$. The important feature of $a_n(t)$ is that

$$|a_n(t)| \leqslant 2 \qquad \text{for} \quad n \geqslant 0. \tag{48}$$

The function $r_n(q)$ is very similar to either $\cosh q$ or $\sinh q$ depending on the parity of n, with the $\sinh q$ behaviour exhibited in the case where n is odd. This function is best described by the following series expression:

$$r_n(q) = \begin{cases} 1 - \cosh q & \text{for} \quad n = 0 \\ \displaystyle\sum_{m=0}^{\infty} \frac{n}{2m+n} \frac{q^{2m}}{(2m)!} & \text{for} \quad n > 0 \quad \text{and even} \\ \displaystyle\sum_{m=0}^{\infty} \frac{n}{2m+n+1} \frac{q^{2m+1}}{(2m+1)!} & \text{for} \quad n > 0 \quad \text{and odd.} \end{cases} \tag{49}$$

The important feature of $r_n(q)$ is that

$$|r_n(q)| \leqslant \cosh q \qquad \text{for} \quad n \geqslant 0. \tag{50}$$

Recalling that $\sinh|q| < \cosh q$, this feature is easily derived from (49). Figures 2 and 3 give an illustration of functions $a_n(t)$ and $r_n(q)$, respectively.

The series expression (46) for the kernel of K is attractive because $a_n(t)$ and $r_n(q)$ are smooth functions that are easily computed and because, more importantly, $a_n(t)$ and $r_n(q)$ are uniformly bounded sequences of functions, so that (46) is a rapidly converging series. More

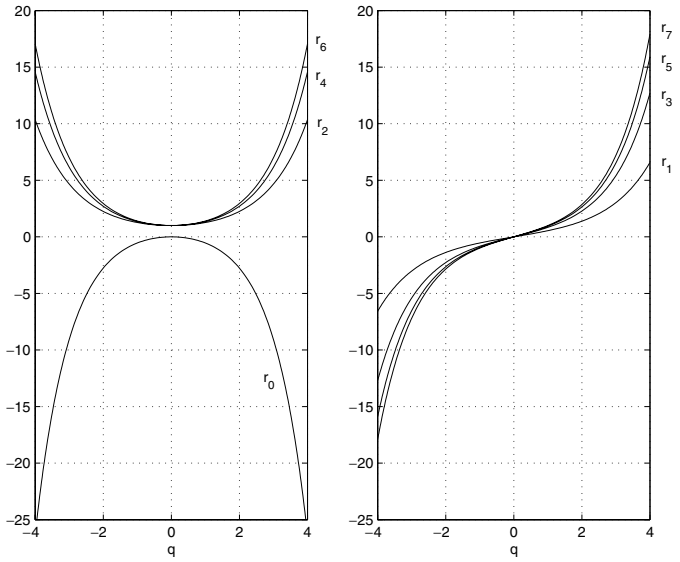

Figure 3. Plot of functions $r_n(q)$ for $n = 0, \ldots, 7$. Left: for n even. Right: for n odd. The interval $|z| \leqslant 4$ includes the domain over which $r_n(q)$ is required for reconstruction with any $\mu \leqslant 4$.

precisely, if $\overline{k}_{\mu,M}(t,\tau)$ is the result of truncating (46) to the first M terms, then (48) and (50) yield the following approximation error for $|t| \leqslant 1$ and $|\tau| \leqslant 1$:

$$|\overline{k}_{\mu}(t,\tau) - \overline{k}_{\mu,M}(t,\tau)| \leqslant \frac{2}{\pi}\cosh\mu \sum_{n=M}^{\infty} \frac{\mu^n}{n!}$$
$$\leqslant \frac{2}{\pi}\cosh\mu \left(\mathrm{e}^{\mu} - \sum_{n=0}^{M-1} \frac{\mu^n}{n!}\right). \tag{51}$$

This error is smaller than 10^{-8} for $M = 40$ and any $\mu \leqslant 8$.

We use $\overline{k}_{\mu,M}(t,\tau)$ to split K into the sum of two terms K_M and R_M and we assess the invertibility of (30) through this decomposition, following the theory in section 2.4. The specific definition for K_M is

$$(K_M h)(t) = \int_{-1}^{1} h(\tau) \frac{1}{\sqrt{1-\tau^2}} \overline{k}_{\mu,M}(t,\tau)\,\mathrm{d}\tau \tag{52}$$

for any $h \in H$. Thus, from (46),

$$(K_M h)(t) = \sum_{n=0}^{M-1} \frac{(-\mu)^n}{\pi n!} \beta_n a_n(t) \tag{53}$$

with

$$\beta_n = \int_{-1}^{1} \frac{r_n(\mu\tau)h(\tau)}{\sqrt{1-\tau^2}}\,\mathrm{d}\tau, \qquad n = 0, 1, \ldots, M-1. \tag{54}$$

By definition, K_M is a degenerate-kernel operator and as such is compact in H. Below, bounds A_M and B_M on $I - K_M$ and the residual R_M are given following equations (42) and (43), respectively.

3.2. Expression for A_M

According to (42), A_M exists only if the operator $I - K_M$ is invertible and A_M defines then a bound on the inverse of this operator. So, to find A_M, we need to investigate the problem of finding $h \in H$ such that

$$(I - K_M)h = h_g \tag{55}$$

for any given $h_g \in H$. Following standard theory for the solution of integral equations with degenerate kernels [27], we use (53) to replace this equation by

$$h(t) = h_g(t) + \sum_{n=0}^{M-1} \frac{(-\mu)^n}{\pi n!} \beta_n a_n(t) \tag{56}$$

with β_n given by (54). Then, we replace $h(\tau)$ on the right-hand side of equation (54) by its expression from (56), and obtain

$$\beta_m = \gamma_m + \sum_{n=0}^{M-1} B_{mn}\beta_n, \qquad \text{for} \quad m = 0, 1, \ldots, M-1 \tag{57}$$

with

$$B_{mn} = \frac{(-\mu)^n}{\pi n!} \int_{-1}^{1} \frac{r_m(\mu t)a_n(t)}{\sqrt{1-t^2}}\, \mathrm{d}t, \qquad m \geqslant 0, n \geqslant 0. \tag{58}$$

and

$$\gamma_m = \int_{-1}^{1} \frac{r_m(\mu t)h_g(t)}{\sqrt{1-t^2}}\, \mathrm{d}t. \tag{59}$$

Equation (57) defines a system of M linear equations for M unknowns, namely the coefficients β_n with $n = 0, \ldots, M-1$. Let D_M be the determinant of the matrix of this system of equations; when μ is zero, D_M equals 1 because then $B_{mn} = 0$ for $n > 0$ and $B_{00} = 0$ (recall from (49) that $r_0(0) = 0$). If $D_M \neq 0$, then $I - K_M$ has a bounded inverse. For now, we assume this condition holds and write the solution of (57) as

$$\beta_n = \sum_{m=0}^{M-1} C_{nm}\gamma_m. \tag{60}$$

Using this notation, we find an expression for A_M as follows. First, we note from (60) that

$$|\beta_n| \leqslant \sum_{m=0}^{M-1} |C_{nm}||\gamma_m|. \tag{61}$$

Next, we apply the Cauchy–Schwartz relation to (59), which gives

$$|\gamma_m| \leqslant \|h_g\| \cdot \|r_m(\mu t)\|, \tag{62}$$

so that from (61)

$$|\beta_n| \leqslant \|h_g\| \sum_{m=0}^{M-1} |C_{nm}|\|r_m(\mu t)\|. \tag{63}$$

Finally, we apply the triangle inequality to (56). This yields with (63)

$$\begin{aligned} \|h\| &\leqslant \|h_g\| + \sum_{n=0}^{M-1} \|a_n(t)\| \frac{\mu^n}{\pi n!} |\beta_n| \\ &\leqslant \|h_g\| \left(1 + \sum_{m,n=0}^{M-1} \|a_n(t)\| \frac{\mu^n}{\pi n!} |C_{nm}|\|r_m(\mu t)\|\right). \end{aligned} \tag{64}$$

From (55), this last result is equivalent to

$$\|(I - K_M)h\| \geqslant \left(1 + \sum_{m,n=0}^{M-1} \|a_n(t)\| \frac{\mu^n}{\pi n!} |C_{nm}| \|r_m(\mu t)\|\right)^{-1} \|h\|, \tag{65}$$

which yields, comparing with (42),

$$A_M = \left(1 + \sum_{m,n=0}^{M-1} \|a_n(t)\| \frac{\mu^n}{\pi n!} |C_{nm}| \|r_m(\mu t)\|\right)^{-1}. \tag{66}$$

Recall $D_m \neq 0$ was assumed earlier; expression (66) for A_M applies only when $D_m \neq 0$ so that the coefficients C_{mn}, which form the inverse of the matrix of the system of equations in (57), are well defined.

3.3. Expression for B_M

According to (43), B_M is any upper bound on the norm of the residual operator $R_M = K - K_M$. From (31) and (52), R_M is an operator in H with $\bar{k}_\mu(t, \tau) - \bar{k}_{\mu,M}(t, \tau)$ for kernel. Hence,

$$\|R_M\| \leqslant \left(\int_{-1}^{1} \mathrm{d}t \frac{1}{\sqrt{1-t^2}} \int_{-1}^{1} \mathrm{d}\tau \frac{1}{\sqrt{1-\tau^2}} |\bar{k}_\mu(t, \tau) - \bar{k}_{\mu,M}(t, \tau)|^2\right)^{1/2} \tag{67}$$

(see for example [26]). Applying (51), this relation yields

$$\|R_M\| \leqslant 2 \cosh \mu \left(\mathrm{e}^\mu - \sum_{n=0}^{M-1} \frac{\mu^n}{n!}\right), \tag{68}$$

so that the following expression may be taken for B_M:

$$B_M = 2 \cosh \mu \left(\mathrm{e}^\mu - \sum_{n=0}^{M-1} \frac{\mu^n}{n!}\right). \tag{69}$$

3.4. Existence, uniqueness and stability

Now that expressions have been found for A_M and B_M, the existence, uniqueness and stability of the solution to (30) can be proved for any value of μ through a numerical evaluation of these expressions. If a value of M can be found such that $D_M \neq 0$ and $A_M > B_M$, then (30) admits a unique and stable solution. Appendix C describes an efficient method to compute the matrix elements B_{mn} and thus to facilitate the numerical evaluation of A_M in (66).

We computed D_M, A_M and B_M for $M = 20$ and a fine uniform sampling of μ over the interval $[0, 8]$. The sampling distance was 0.0001 and the computation outcome is illustrated in figure 4. This figure shows that D_M was always larger than 1, thus (57) was always invertible. The figure also shows that A_M was larger than B_M for $\mu < 4.7$, thus proving (30) has a unique and stable solution for all sampled values of μ below 4.7.

Note that the result in figure 4 does not preclude (30) from having a unique and stable solution for μ larger than 4.7. To illustrate this point, we ran the experiment again with $M = 40$. The outcome of this experiment is illustrated in figure 5. This figure shows that the increase in M had little effect on the values of D_M and A_M but significantly changed those of B_M, so that B_M is now smaller than A_M up to at least $\mu = 8$.

The experimental results above prove that (30) admits a unique and stable solution for $\mu = k/10000$, $k = 0, \ldots, 80000$. The obtained plots also strongly suggest a stable solution

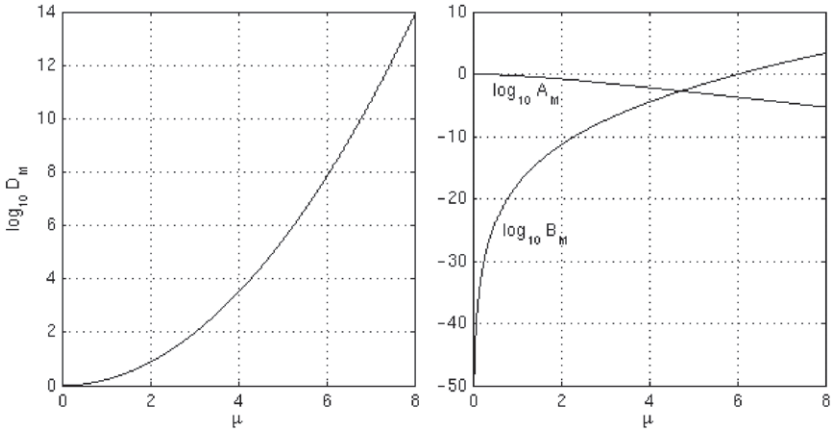

Figure 4. Left: plot of $\log_{10} D_M$ as a function of μ, where D_M is the determinant of the system of equations defined by (57); the plot shows that $D_M \geqslant 1$ for all sampled values of μ. Right: plot of $\log_{10} A_M$ and $\log_{10} B_M$ as functions of μ. When A_M and B_M are such that $A_M > B_M$, or equivalently $\log_{10} A_M > \log_{10} B_M$, existence, uniqueness and stability of the solution to (30) are guaranteed. Here, $M = 20$ and A_M is larger than B_M for all sampled values of μ below 4.7.

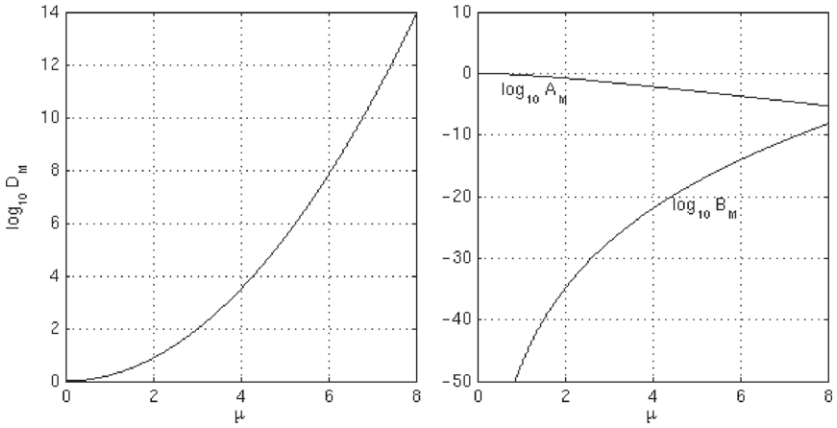

Figure 5. Same plots as in figure 4 but for $M = 40$. The graphs of D_M and A_M are similar to those in figure 4, but not the graph of B_M. For $M = 40$, we see that B_M remains smaller than A_M up to at least $\mu = 8$.

for all values of $\mu \in [0, 8]$. Note that A_M seems to decrease exponentially with increasing μ, which negatively affects the stability of the inversion, as outlined by equations (40) and (45) in section 2.4. The plot of $1/(A_M - B_M)$ is given in figure 6 for $M = 40$ and $\mu \in [0, 4]$, a range that covers most patients in medical applications of SPECT. In this figure, the stability appears clearly very robust for $\mu < 1$ but beyond that starts degrading exponentially. This degradation is not surprising given the exponential behaviour of the convolution kernel that relates the data to the sought function in the initial problem (equation (15)).

4. Experimental results

Numerical testing using a SPECT version of the Shepp–Logan phantom [29] was performed under various conditions. This phantom is displayed in figure 7, while table 1 gives a

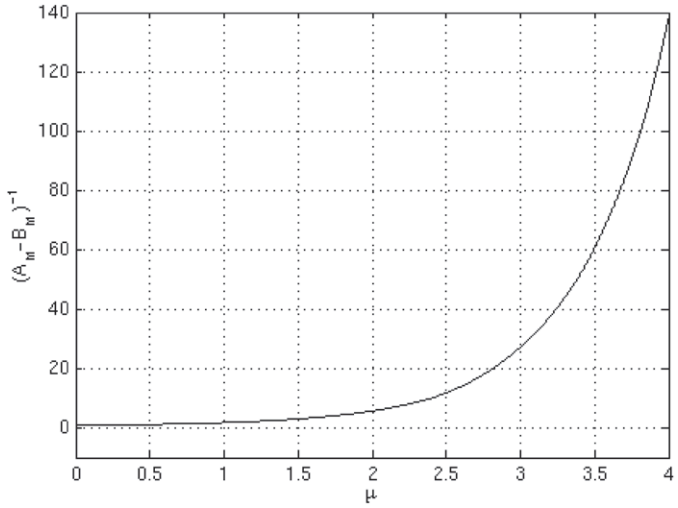

Figure 6. Plot of $1/(A_M - B_M)$ as a function of μ for $M = 40$. This quantity is an upper bound on how errors in the DBP result due to data noise are amplified through the inversion of (30), as explained by equations (40) and (45).

Figure 7. Display of the Shepp–Logan phantom used for testing. Grey scale: $[0.15, 0.45]$. The central rectangular box drawn over the phantom was used to truncate the data, as described in the text.

description of the ellipses that form this phantom by addition. Analytical formulae were used to sample the exponential Radon transform of the phantom for three values of μ_o, namely 0.0 cm^{-1}, 0.15 cm^{-1} and 0.3 cm^{-1}. A total of 1000 projections with 400 rays per projection was simulated for each value of μ_o. The samples in ϕ were at angles $\phi_k = k\pi/999$ with $k = 0, \ldots, 999$, and the samples in s were at positions $s = (-9.975 + 0.05l)$ cm with $l = 0, \ldots, 399$. None of the simulated projections was truncated.

Reconstructions from the simulated data were performed with and without added Poisson noise and also with and without introduced truncation. These reconstructions used a centred square of side 20 cm for Ω, the bounded and convex set that includes all activity sources. The first reconstruction step was the computation of the DBP of equation (4) onto a centred Cartesian grid of 400×400 square pixels of width 0.05 cm. This step reduced the reconstruction problem into 400 integral equations, one for each sampled value of x and given each by (13) with $U_x = -L_x = 10$ cm. Next, each integral equation was converted into

Table 1. Definition of the ellipses forming the SPECT version of the Shepp–Logan phantom by addition.

Centre (cm)	1st axis (cm)	2nd axis (cm)	Polar angle (°)	Intensity
(0, 0)	6.9	9.2	0	0.5
(0, −0.184)	6.624	8.74	0	−0.2
(2.2, 0)	3.1	1.1	72	−0.2
(−2.2, 0)	4.1	1.6	108	−0.2
(0, 3.5)	2.1	2.5	0	0.1
(0, 1)	0.46	0.46	0	0.1
(0, −1)	0.46	0.46	0	0.1
(−0.8, −6.05)	0.46	0.23	0	0.1
(0, −6.05)	0.23	0.23	0	0.1
(0.6, −6.05)	0.23	0.46	0	0.1

equation (30) using $\mu = 10\mu_o$, according to (17). The forcing term h_d in (30) was computed using a band-limited version of the Hilbert transform kernel, as explained in [2]. Finally, each instance of (30) was inverted using the degenerate-kernel approximation of equation (56) with $M = 20$ and $h_g = h_d + m_\mu/\pi$.

Truncation was introduced by setting to zero any sample that corresponded to a line not going through the rectangular box drawn over the phantom in figure 7. The width of this box was 4 cm. According to our theory, accurate reconstruction should be possible anywhere within this box.

Poisson noise was added following the standard procedure, assuming a total count of 10^9 photons and a uniform attenuation map of support identical to the support of the phantom. For completeness, we now summarize this procedure: (i) the samples of the exponential Radon transform were converted into samples of the attenuated Radon transform, (ii) each resulting sample was scaled so that the sum of the samples become 10^9, (iii) each scaled sample was replaced by one realization of a Poisson variable of mean equal to the value of the scaled sample, (iv) each noisy sample was scaled back down and then converted back to noisy samples of the exponential Radon transform. The conversion forth and back between the exponential and the attenuated Radon transform followed the equations in [5].

Figure 8 shows the reconstruction results from non-truncated data. As expected, the larger the value of μ_o the more noisy the reconstruction when noise is added to the data. Even in the absence of noise, we observe that a larger value of μ_o results in larger discretization errors.

Figure 9 shows the reconstruction results from truncated data. Outside the rectangular box, no accurate reconstruction was expected and this is clearly visible. However, inside the rectangular box, the reconstruction appears as accurate as in the non-truncated case, as predicted by the theory in this paper. Figure 10 compares the profiles along the line $x = 0$ and gives thus a quantitative picture of discretization errors.

In both figures 8 and 9, the noise in the reconstructions appears strongly spatially variant. We attribute this variance to the mean value of the exponential weight in the expression of the DBP. Consider for example the computation of the DBP at $y = 0$. In this case, the exponential weight is $\exp(-\mu x \sin\phi)$. Recalling that we use $\phi \in [0, \pi]$, $\sin\phi \geqslant 0$ and therefore this weight is larger than one for any positive x and smaller than one for any negative x. Hence, data noise gets more amplified in the reconstruction as x increases.

5. Conclusions and discussion

This paper presented an analysis of the problem of image reconstruction from truncated projections in SPECT. For this analysis, we assumed that the attenuation map is constant in

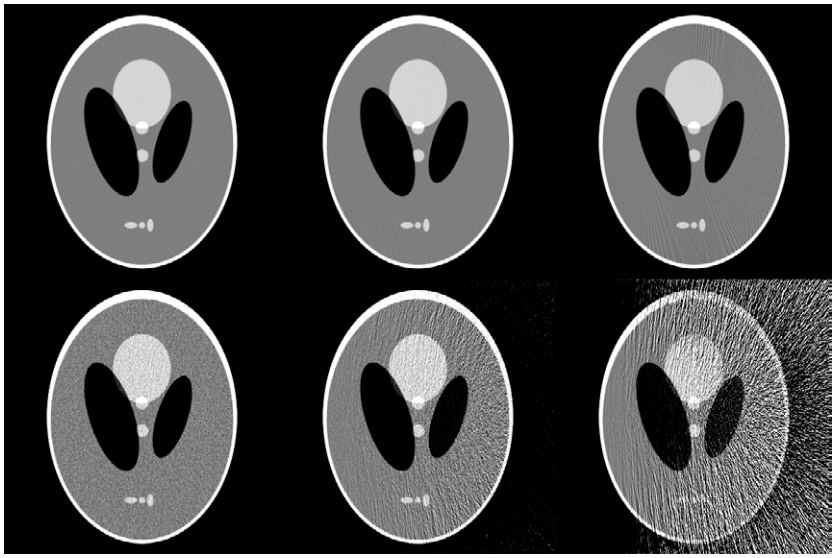

Figure 8. Reconstructions with no data truncation. Grey scale: [0.15, 0.45]. Top row: without noise. Bottom row: with added Poisson noise corresponding to a total count of 10^9 photons. From left to right, in each row, μ_o is successively 0 cm^{-1}, 0.15 cm^{-1} and 0.3 cm^{-1}.

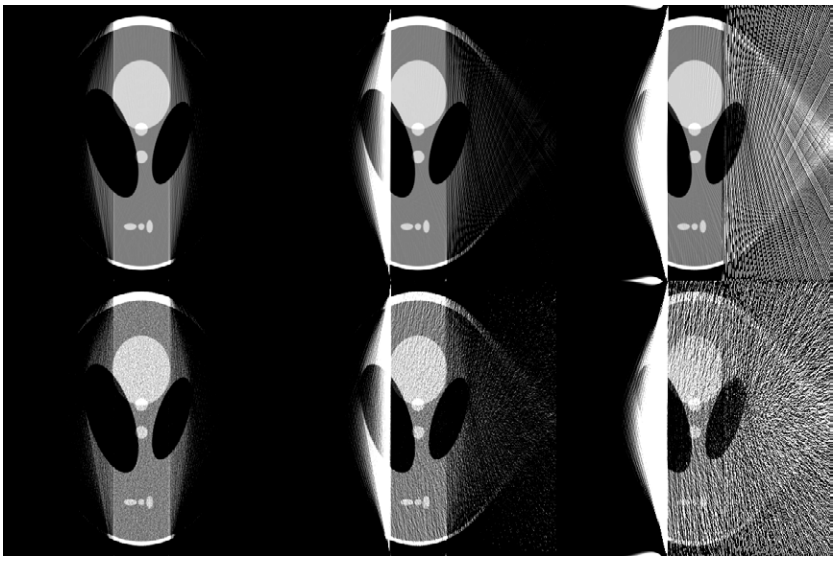

Figure 9. Reconstructions with data truncation such that any measurement corresponding to a line not passing through the rectangular box in figure 7 was discarded. Grey scale: [0.15, 0.45]. Top row: without noise. Bottom row: with added Poisson noise corresponding to a total count of 10^9 photons. From left to right, in each row, μ_o is successively 0 cm^{-1}, 0.15 cm^{-1} and 0.3 cm^{-1}. Accurate reconstruction is only expected within the rectangular box.

a convex region that includes all activity sources, so that the exponential Radon transform could be used to model the mathematical relation between the projections and the sought

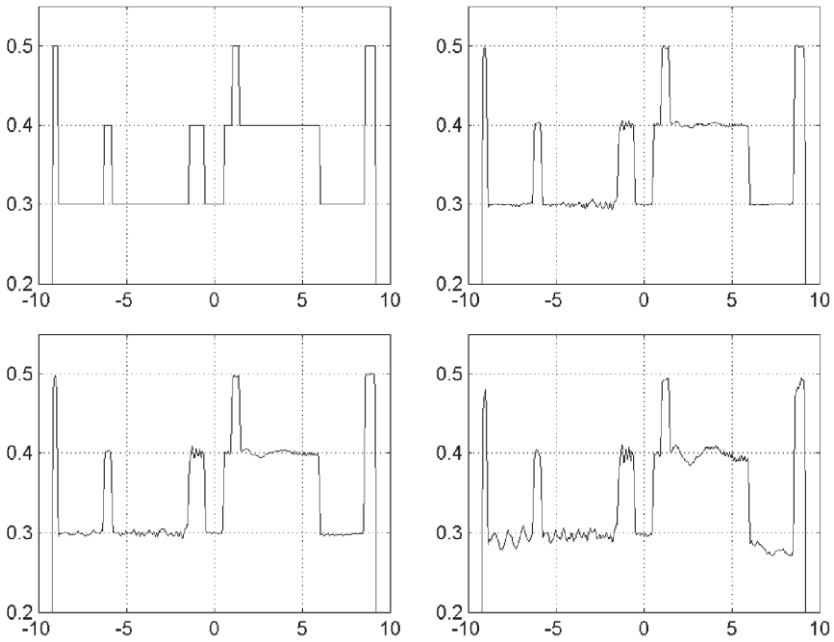

Figure 10. Plots along the line $x = 0$ in the Shepp–Logan phantom. Top left: ground truth. The other plots correspond to the reconstructions in the first row of figure 9, i.e. from non-noisy but truncated data. Top right: $\mu_o = 0$. Bottom left: $\mu_o = 0.15\ \text{cm}^{-1}$. Bottom right: $\mu_o = 0.3\ \text{cm}^{-1}$.

activity function. Also, the projections were assumed to be known only over 180°. The analysis was based on the concept of DBP, introduced earlier by Rullgård [23]. This concept reduces the reconstruction problem to a 1D integral equation for which Rullgård found a deconvolution-type inversion formula. We noticed first that Rullgård's result implies exact and stable reconstruction of a ROI is possible from truncated data. Next, we presented a shift-variant method as a substitute to Rullgård's deconvolution formula for inversion of the 1D integral equation resulting from the DBP. Using this shift-variant method, we proved that exact and stable reconstruction of a given ROI is possible with even less restriction on the data than required by Rullgård's formula. The proof was given for a broad range of values of the attenuation coefficient that covers all medical applications of SPECT. The theory was supported by simulated results in the SPECT context.

The improvement on the pattern of allowed truncation for reconstruction of a given ROI is the main result of this paper. Though any appropriate iterative reconstruction algorithm could be used to take advantage of this result, our shift-variant inversion method also provides an efficient reconstruction algorithm, the efficacy of which was demonstrated using simulated data. In this demonstration, the data were finely sampled and simulated with a high photon count since the primary goal was to support the theory with numerical examples. Evaluations with less ideal data are reported in [30].

The theory we presented defines an extension of results published in [2] for classical tomography. Compared to the non-attenuated case, the particular 180° angular range over which the projections are known seems to play a critical role in the pattern of truncation that is allowed for reconstruction of a given ROI. This difference arises from the fact that projections that are 180° apart differ from each other in the presence of attenuation, unlike in

the non-attenuated case where any 180° angular interval of measurements provides the same complete set of measurements.

As mentioned at the beginning of this section, our theory is not valid for arbitrary non-uniform attenuation maps. Remarkable findings have been published recently on a general inversion of the attenuated Radon transform [31–38], see [39] for an overview. These findings allow accurate reconstruction with attenuation maps that are non-uniform over the activity region. However, they assume that the projections are known over 360°, except in [36, 37] and [38] where early theories are laid down for reconstruction from data over 180°. And, more importantly, [31–39] assume non-truncated projections. The problem of inverting the attenuated Radon transform for regional reconstruction from truncated projections remains open.

The ability to handle truncation is significant for SPECT, particularly when using converging collimators to focus the data acquisition on a given ROI. Converging collimators offer an improved trade-off between detection efficiency and spatial resolution, but this improvement is usually offset by data truncation difficulties. The results in this paper offer a better understanding of conditions under which quantitative reconstruction with converging collimators is guaranteed in the presence of truncation, disregarding the analytical or iterative nature of the algorithm used for reconstruction.

Acknowledgments

This work was supported by the National Institutes of Health (NIH) under grant R21 EB002526. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIH.

Appendix A

We briefly justify here the switching in the order of integration that was used to step from (25) to (26). This switching is valid for any square integrable function f as can be seen by recalling from (24) that c is a smooth function and by rewriting (25) in the following form:

$$\begin{aligned}\fint_{-1}^{1} \mathrm{d}\rho \frac{\sqrt{1-\rho^2}}{\pi(t-\rho)} \int_{-1}^{1} \mathrm{d}\tau f(\tau) c(\mu(\rho-\tau)) \\ = \int_{-1}^{1} \mathrm{d}\rho \sqrt{1-\rho^2} \int_{-1}^{1} \mathrm{d}\tau f(\tau) \frac{c(\mu(\rho-\tau)) - c(\mu(t-\tau))}{\pi(t-\rho)} \\ + \left(\fint_{-1}^{1} \mathrm{d}\rho \frac{\sqrt{1-\rho^2}}{\pi(t-\rho)} \right) \left(\int_{-1}^{1} \mathrm{d}\tau f(\tau) c(\mu(t-\tau)) \right). \end{aligned} \tag{A.1}$$

In the first term, there is not singularity anymore at $t = \rho$ since c is a smooth function, so switching the order of integration is allowed for this term. In the second term, the two integrals are independent and their order can thus be switched. Recombining the two terms together after carrying order switching in each of them yields the same result as directly changing the order of integration in the initial double integral, even though the first integral is only defined as a Cauchy principal value.

Appendix B

This appendix explains how to derive the series expansion (46) of $\overline{k}_\mu(t, \tau)$, the kernel of K, that was given in section 3.1.

Consider the following Taylor series expansion of $c(\mu(\rho-\tau))$, which appears in the first term of $\bar{k}_\mu(t,\tau)$ according to equations (32) and (29):

$$c(\mu(\rho-\tau)) = \sum_{n=0}^{\infty} \frac{(\mu\rho)^n}{n!} c^{(n)}(-\mu\tau) \tag{B.1}$$

where $c^{(n)}$ is the nth derivative of c. From the definition (24) of c, $c^{(n)}(-q) = (-1)^{n+1}c^{(n)}(q)$ and defining

$$c^{(n)}(q) = \frac{r_{n+1}(q)}{\pi(n+1)}, \tag{B.2}$$

we obtain

$$c(\mu(\rho-\tau)) = -\sum_{n=0}^{\infty} \frac{(-\mu\rho)^n}{\pi(n+1)!} r_{n+1}(\mu\tau). \tag{B.3}$$

The insertion of this expansion into (29) gives

$$k_\mu(t,\tau) = \sum_{n=0}^{\infty} \frac{(-\mu)^{n+1}}{\pi(n+1)!} r_{n+1}(\mu\tau) a_{n+1}(t) \tag{B.4}$$

$$= \sum_{n=1}^{\infty} \frac{(-\mu)^n}{\pi n!} r_n(\mu\tau) a_n(t) \tag{B.5}$$

where

$$a_n(t) = \fint_{-1}^{1} \frac{\sqrt{1-\rho^2}}{\pi(t-\rho)} \rho^{n-1}\, \mathrm{d}\rho. \tag{B.6}$$

At this stage, we define $a_0(t) = 1$ and $r_0(q) = 1 - \cosh q$, so that from (32)

$$\bar{k}_\mu(t,\tau) = \sum_{n=0}^{\infty} \frac{(-\mu)^n}{\pi n!} r_n(\mu\tau) a_n(t). \tag{B.7}$$

Equation (B.7) has the same form as (46), the sought expansion. To complete the proof, we need to show that the introduced functions $r_n(q)$ and $a_n(t)$ are as defined in section 3.1. Using Taylor series expansion around $q = 0$, we have from (24)

$$c(q) = \frac{1}{\pi} \sum_{m=0}^{\infty} \frac{1}{2m+2} \frac{q^{2m+1}}{(2m+1)!} \tag{B.8}$$

and, consequently,

$$c^{(n)}(q) = \frac{1}{\pi} \sum_{m=(n-\bmod(n,2))/2}^{\infty} \frac{1}{2m+2} \frac{q^{2m+1-n}}{(2m+1-n)!} \tag{B.9}$$

$$= \frac{1}{\pi} \sum_{m=0}^{\infty} \frac{1}{2m+n+1+j} \frac{q^{2m+j}}{(2m+j)!}, \qquad j = \bmod(n+1,2). \tag{B.10}$$

The combination of this expression with (B.2) yields equation (49) for $r_n(q)$.

Now, the case of $a_n(t)$ is discussed. The recurrence relation of equation (47) is first proved, then the bound (48). First, we know from Tricomi [25] that $a_1(t) = t$ for $|t| < 1$,

which validates (47) for $n = 1$. For $n > 1$, the integral in (B.6) is split into two integrals using the relation $\rho^{n-1} = (\rho - t + t)\rho^{n-2}$:

$$a_n(t) = -\frac{1}{\pi}\int_{-1}^{1}\sqrt{1-\rho^2}\rho^{n-2}\,\mathrm{d}\rho + t\fint_{-1}^{1}\frac{\sqrt{1-\rho^2}}{\pi(t-\rho)}\rho^{n-2}\,\mathrm{d}\rho$$
$$= -\eta_{n-1} + t a_{n-1}(t) \tag{B.11}$$

with

$$\eta_n = \frac{1}{\pi}\int_{-1}^{1}\sqrt{1-\rho^2}\rho^{n-1}\,\mathrm{d}\rho, \qquad n \geqslant 1. \tag{B.12}$$

From (B.12), one finds $\eta_1 = 1/2$, $\eta_2 = 0$ and $\eta_{n+2} = n\eta_n/(n+3)$ for $n > 0$, as expected. Finally, to get the bound (48), note that (B.11) implies

$$|a_n(t)| \leqslant |a_{n-1}(t)| + \eta_{n-1} \qquad \text{for} \quad |t| \leqslant 1 \quad \text{and} \quad n \geqslant 1. \tag{B.13}$$

Given that $a_0(t) = 1$ and $\eta_0 = 0$ by definition, successive applications of this relation yield

$$|a_n(t)| \leqslant 1 + \sum_{q=0}^{n-1}\eta_q \leqslant 1 + \sum_{q=1}^{\infty}\eta_q, \tag{B.14}$$

i.e., from (B.12)

$$|a_n(t)| \leqslant 1 + \int_{-1}^{1}\frac{\sqrt{1-\rho^2}}{\pi}\sum_{q=1}^{\infty}\rho^{q-1}\,\mathrm{d}\rho \leqslant 1 + \int_{-1}^{1}\sqrt{\frac{1+\rho}{1-\rho}}\,\mathrm{d}\rho. \tag{B.15}$$

Hence, $|a_n(t)| \leqslant 2$ for $|t| \leqslant 1$ and $n \geqslant 0$.

Appendix C

This appendix briefly describes an efficient method to compute the integral in the definition (58) of the matrix element B_{mn}. First, let

$$A_{pn} = \int_{-1}^{1}\frac{t^p a_n(t)}{\sqrt{1-t^2}}\,\mathrm{d}t. \tag{C.1}$$

From the recurrence relation (47) for $a_n(t)$,

$$A_{pn} = A_{p+1,n-1} - \pi\rho_p\eta_{n-1} \tag{C.2}$$

with

$$\rho_p = \frac{1}{\pi}\int_{-1}^{1}\frac{t^p}{\sqrt{1-t^2}}\,\mathrm{d}t. \tag{C.3}$$

From (C.3), $\rho_0 = 1$, $\rho_1 = 0$ and $\rho_{p+2} = \rho_p(p+1)/(p+2)$. Therefore, noting that $A_{p,0} = \pi\rho_p$ since $a_0(t) = 1$, (C.2) may be used to quickly compute a range of values of A_{pn}.

Now, recall from (49) that a rapidly converging power series expansion is available for $r_m(\mu t)$. Using the following general notation for this series,

$$r_m(\mu t) = \sum_{p=0}^{\infty} b_p^{(m)} t^p, \tag{C.4}$$

(58) is rewritten as

$$B_{mn}(\mu\tau) = \frac{(-\mu)^n}{\pi n!}\sum_{p=0}^{\infty} b_p^{(m)} A_{pn}, \tag{C.5}$$

which may be computed accurately with few terms, using pre-computed values of A_{pn}.

References

[1] Clackdoyle R, Noo F, Guo J and Roberts J A 2004 Quantitative reconstruction from truncated projections in classical tomography *IEEE Trans. Nucl. Sci.* **51** 2570–78

[2] Noo F, Clackdoyle R and Pack J D 2004 A two-step Hilbert transform method for 2D image reconstruction *Phys. Med. Biol.* **49** 3903–23

[3] Zou Y, Pan X and Sidky E Y 2005 Image reconstruction in regions-of-interest from truncated projections in a reduced fan-beam scan *Phys. Med. Biol.* **50** 13–28

[4] Gelfand I M and Graev M I 1991 Crofton's function and inversion formulas in real integral geometry *Funct. Anal. Appl.* 1–5

[5] Gullberg G T and Budinger T F 1981 The use of filtering methods to compensate for constant attenuation in single-photon emission computed tomography *IEEE Trans. Biomed. Eng.* **28** 142–57

[6] Bellini S, Piacenti M, Caffario C and Rocca F 1979 Compensation of tissue absorption in emission tomography *IEEE Trans. Acoust. Speech Signal Process.* **27** 213–8

[7] Clough A and Barrett H H 1980 Attenuated Radon and Abel transforms *J. Opt. Soc. Am.* **39** 341–54

[8] Tretiak O and Metz C 1980 The exponential Radon transform *SIAM J. Appl. Math.* **39** 341–54

[9] Markoe A 1984 Fourier inversion of the attenuated x-ray transform *SIAM J. Math. Anal.* **15** 718–22

[10] Finch D V and Hertle A 1987 The exponential radon transform *Integral Geometry (Contemporary Mathematics* vol 63 ed Bryant, Guillemin, Helgason and Wells*)* (Providence, RI: American Mathematical Society) pp 67–73

[11] Hawkins W G, Leichner P K and Yang N C 1988 The circular harmonic transform for SPECT reconstruction and boundary conditions on the Fourier transform of the sinogram *IEEE Trans. Med. Imag.* **7** 135–48

[12] Inouye T, Kose K and Hasegawa A 1989 Image reconstruction algorithm for single-photon-emission computed tomography with uniform attenuation *Phys. Med. Biol.* **34** 299–304

[13] Hazou I A and Solmon D C 1988 Inversion of the exponential x-ray transform: I: analysis *Math. Methods Appl. Sci.* **10** 561–74

[14] Hazou I A and Solmon D C 1989 Filtered-backprojection and the exponential Radon transform *J. Math. Anal. Appl.* **141** 109–19

[15] Liang Z, Ye J and Harrington D P 1994 An analytical approach to quantitative reconstruction of non-uniform attenuated brain SPECT *Phys. Med. Biol.* **39** 2023–41

[16] Liang Z, Ye J, Cheng J and Harrington D P 1996 The inversion of the exponential Radon transform for quantitative brain SPECT *Phys. Med. Biol.* **41** 1227–32

[17] Clarkson E 1999 Projections onto the range of the exponential Radon transform and reconstruction algorithms *Inverse Problems* **15** 563–71

[18] Metz C E and Pan X 1995 A unified analysis of exact methods of inverting the 2D exponential Radon transform, with implications for noise control in SPECT *IEEE Trans. Med. Imag.* **14** 643–58

[19] Pan X and Metz C E 1995 Analysis of noise properties of a class of exact methods of inverting the 2D exponential Radon transform *IEEE Trans. Med. Imag.* **14** 659–68

[20] Noo F and Wagner J-M 2001 Image reconstruction in 2D SPECT with 180° acquisition *Inverse Problems* **17** 1357–71

[21] Pan X, Kao C-M and Metz C E 2002 A family of π-scheme exponential Radon transforms and the uniqueness of their inverses *Inverse Problems* **18** 825–36

[22] Wagner J-M 2002 Méthodes analytiques pour la correction d'atténuation en tomographie tridimensionelle par émission monophotonique *PhD Thesis* Collection des Publications de la Faculté des Sciences Appliquées, University of Liège, Belgium p 214

[23] Rullgård H 2004 An explicit inversion formula for the exponential Radon transform using data from 180° *Ark. Mat.* **42** 353–62

[24] Söhngen H 1937 Die Lösungen der Integralgleichung $g(x) = \frac{1}{2\pi}\int_{-a}^{a}\frac{f(\xi)}{x-\xi}\mathrm{d}\xi$ and deren Anwendung in the Tragflügeltheorie *Math. Z.* **45** 245–64

[25] Tricomi F G 1957 *Integral Equations* (New York: Dover)

[26] Hochstadt H 1973 *Integral Equations* (New York: Wiley)

[27] Kress R 1999 *Linear Integral Equations (Applied Mathematical Sciences* vol 82*)* (New York: Springer)

[28] Atkinson K, Ablowitz M J, Davis S H and Hinch E J 1997 *The Numerical Solution of Integral Equations of the Second Kind* (Cambridge: Cambridge University Press)

[29] Li T, You J, Wen J and Liang Z 2005 An efficient reconstruction method for nonuniform attenuation compensation in nonparallel beam geometries based on Novikov's explicit inversion formula *IEEE Trans. Med. Imag.* **24** 1357–68

[30] Noo F, Defrise M, Pack J D and Clackdoyle R 2006 Image reconstruction from truncated projections in SPECT *Conference Record of the 2006 IEEE Medical Imaging Conference*

[31] Arbuzov E V, Bukhgeim A L and Kazantsev S G 1998 Two-dimensional tomography problems and the theory of A-analytic functions *Siberian Adv. Math.* **8** 1–20
[32] Novikov R G 2002 An inversion formula for the attenuated x-ray transformation *Ark. Mat.* **40** 145–67
[33] Natterer F 2001 Inversion of the attenuated Radon transform *Inverse Problems* **17** 113–19
[34] Kunyansky L A 2001 A new SPECT reconstruction algorithm based on the Novikov explicit inversion formula *Inverse Problems* **17** 293–306
[35] Boman J and Strömberg J-O 2004 Novikov's inversion formula for the attenuated Radon transform—a new approach *J. Geom. Anal.* **14** 185–98
[36] Bal G 2004 On the attenuated Radon transform with full and partial measurements *Inverse Problems* **20** 399–418
[37] Sidky E Y and Pan X 2002 Variable sinograms and redundant information in single-photon emission computed tomography with non-uniform attenuation *Inverse Problems* **18** 1483–97
[38] Rullgård H 2004 Stability of the inverse problem for the attenuated Radon transform with 180° data *Inverse Problem* **20** 781–97
[39] Finch D V 2003 The attenuated x-ray transform: recent developments *Inside Out: Inverse Problems* vol 47 (MSRI Publications) pp 47–66