INSTITUTE OF PHYSICS PUBLISHING INVERSE PROBLEMS

Inverse Problems **18** (2002) 1435–1460 PII: S0266-5611(02)32060-4

The inverse problem of emission tomography

Daniel Gourion[1] **and Dominikus Noll**[2]

[1] Université Paul Sabatier, Laboratoire Traceurs et Traitement de l'Image (EA 3033), 133 route de Narbonne, 31077 Toulouse, France

[2] Université Paul Sabatier, Mathématiques pour l'Industrie et la Physique (CNRS UMR 5640), 118 route de Narbonne, 31062 Toulouse, France

E-mail: gourion@cict.fr and noll@mip.ups-tlse.fr

Received 30 May 2002
Published 16 September 2002
Online at stacks.iop.org/IP/18/1435

Abstract

The image reconstruction process in emission computed tomography (ECT) is an inverse problem for the photon transport equation. For monochromatic emission sources it is closely related to the inversion of the attenuated Radon transform, a nonlinear ill posed inverse problem. Due to its practical importance for medical diagnostics, this problem has been addressed various times. Here we present the theoretical setting of ECT and discuss some new numerical strategies based on regularization techniques. We include experiments to compare some of the numerical approaches.

(Some figures in this article are in colour only in the electronic version)

1. Problem setting

Single-photon emission computed tomography (SPECT) is a non-invasive diagnostic technology in nuclear medicine. It is used to show blood flow in the heart muscle, the extent of damage in stroke patients, the presence and degree of malignancy of tumours and much else. SPECT images the function of the body through a *tracer*, a biochemical molecule labelled with radioactivity. The radioactive material is incorporated by the patient and either metabolized by the organ or simply trapped or bound in tissue. The emitted gamma radiation is recorded by a SPECT camera rotating around the patient, and a three-dimensional visualization is created from the two-dimensional projection data via an inverse mathematical method. The setup is schematically shown in figures 1 and 2. For additional introductory information see [77, ch 5, 14].

Positron emission tomography (PET) is a related technology with similar diagnostic goals. In PET, the radioactive agent is a positron emitter, attached to the tracer molecule. Positrons annihilate quasi-instantaneously with a nearby electron, and the energy equivalent to the two masses is shared by two photons, which radiate, each at 511 keV, in opposite directions. The events are recorded in coincidence mode on a cylindrical detector array surrounding the

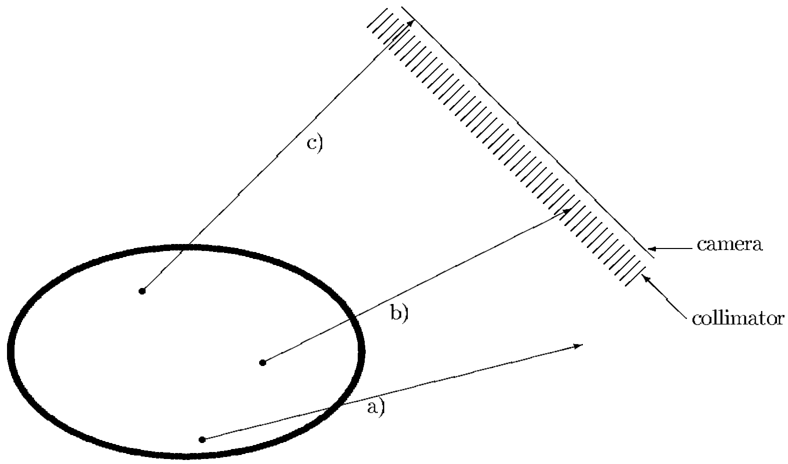

Figure 1. The principle of SPECT: photons emitted from the body are recorded by a rotating camera equipped with a parallel hole collimator. Photon (a) misses the camera, photon (b) is absorbed by the collimator, photon (c) travels perpendicular to the camera surface and therefore passes the collimator. Ideally, only photons which propagate perpendicular to the camera are recorded. Many systems use two or three camera heads.

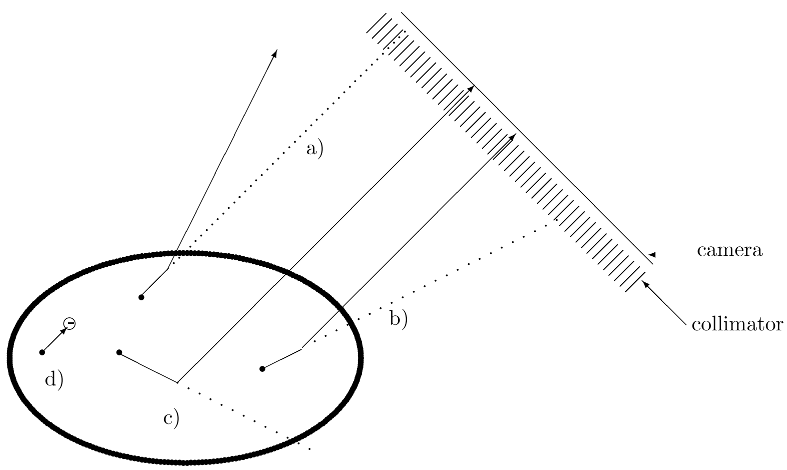

Figure 2. SPECT attenuation and scatter: photon (a) is deviated due to the scatter effect, and the count is 'missing'. The photon is attenuated. Photon (b) should be absorbed by the collimator, but due to scatter is 'erroneously' recorded. Photon (c) should miss the camera, but scatters, passes the collimator and hits the camera. Due to the large scattering angle, the energy loss (5) is recognized, and the event is not recorded. Photon (d) is attenuated due to the photoeffect and the count is 'missing'.

patient as schematically shown in figure 3. See [65], [77, ch 6] for the principles, and [13] for the history of PET. Emission computed tomography (ECT) is a generic term which refers to SPECT, PET and a more recent imaging modality, the Compton camera [69, 70], which is still in an experimental state. All techniques of ECT are governed by the photon transport equation and lead to typical inverse problems.

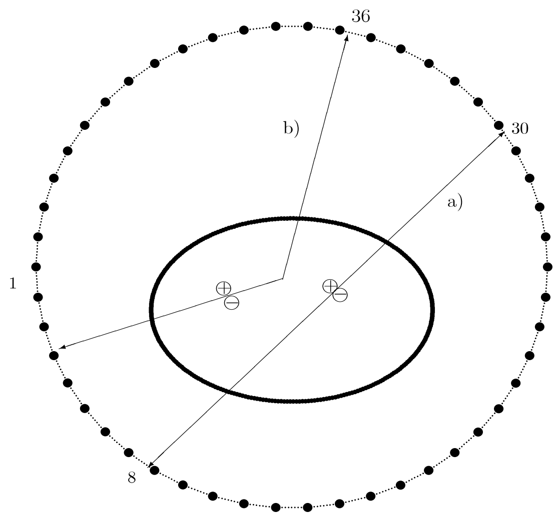

Figure 3. The principle of PET: the emitted positron annihilates with a nearby electron. The energy associated with the mass of both particles is shared by two photons, which part in opposite directions, each one at 511 keV. In case (a) the event is correctly identified as having originated on the line joining receptor tubes 8 and 30. In case (b), one of the photons Compton scatters and changes direction. If the loss of energy (5) is moderate, the event is 'erroneously' attributed to the line joining the tubes 3 and 36. For larger scattering angles, the Compton photon is outside the energy window about the nominal energy 511 keV, and the event is not recorded.

These inverse problems may be solved by two types of strategy. Either additional transmission tomography measurements are performed to alleviate the reconstruction, or the inversion uses the emission data only, in which case a difficult nonlinear inverse problem has to be solved, since two unknown parameters have to be reconstructed simultaneously. In this paper our emphasis is on the second strategy. We show that it may be competitive if diligently handled. This requires a full picture of the physical inverse problem, which we obtain in section 2, starting from the photon transport equation. We then derive a numerical strategy for simultaneous inversion, which we test numerically using a simulation and a physical phantom study.

2. Photon transport equation

2.1. General case

For stationary emission sources, gamma radiation and its interaction with tissue in the human body may be described by the linear stationary photon transport equation (cf [25, 63]):

$$\theta \cdot \nabla I(x, \theta, E) + \mu(x, E)\, I(x, \theta, E) = \int_{\mathbb{S}^2} \sigma(x, E' \to E, \theta' \to \theta)\, I(x, \theta', E')\, \mathrm{d}\theta' + f(x, E). \tag{1}$$

Here $I(x, \theta, E)$ is the specific intensity or energy transport at position x, in direction $\theta \in \mathbb{S}^2$ and at photon energy E, $f(x, E)$ is the unknown radioactive source at position x and energy E, $\mu(x, E)$ is the unknown linear attenuation coefficient and σ is the unknown scattering kernel.

We may assume that the attenuation coefficient decomposes into

$$\mu(x, E) = \rho(x)\,\mu_m(E), \tag{2}$$

with $\rho(x)$ the unknown electron density at position x, and $\mu_m(E)$ a known term representing all physical processes which end flights of photons at energy E. For the energies relevant to nuclear medicine, these include the photoeffect, Rayleigh and Compton scatter [63]. The unknown $\rho(x)$ describes the patient *anatomy*, while the unknown emission source f, the ultimate goal of the SPECT study, characterizes the *function* of the organ under investigation. The scattering kernel $\sigma(x, E' \to E, \theta' \to \theta)$ decomposes into

$$\sigma(x, E' \to E, \theta' \to \theta) = \rho(x)\,\sigma_m(E' \to E, \theta' \to \theta),$$

where the scattering cross-section σ_m is a known term, which accounts for all physical processes that cause photons with incident energy E' travelling in the incident direction θ' to interact with matter and scatter into the new direction of propagation θ with the energy loss $E' - E \geqslant 0$.

For incident energies E' between 80 and 1000 keV, and for the soft and bone tissue relevant to nuclear medicine, we may consider Compton scatter as the dominating effect and model the scattering kernel as

$$\sigma(x, E' \to E, \theta' \to \theta) = \rho(x)\,\sigma_m\big(E', E, \widehat{(\theta, \theta')}\big) \tag{3}$$

where $\widehat{(\theta, \theta')}$ is the angle between θ and θ', and $\sigma_m(E', E, \phi)$ is the Klein–Nishina scattering cross-section (cf [26]):

$$\sigma_m(E', E, \phi) = \frac{r_0^2}{2}\left(\frac{E}{E'}\right)^2\left(\frac{E'}{E} + \frac{E}{E'} - \sin^2\phi\right), \tag{4}$$

where r_0 is the classical electron radius (see [31]). Formula (4) describes the probability distribution of Compton scattering angles $\phi = \widehat{(\theta, \theta')}$ for the incident photon energy E' (see [31] for details).

Notice that due to the conservation of energy and momentum in the Compton scatter process, the energy E' of the incident photon scattered from direction θ' into the direction θ and the energy E of the scattered photon are linked through the formula

$$E' = \frac{E}{1 - (E/mc^2)(1 - \cos\phi)} \tag{5}$$

with $\phi = \widehat{(\theta', \theta)}$ and m the electron rest mass (cf [75]). In particular, $\sigma_m(E', E, \phi) = \sigma_m(E', \phi)$, obtained by substituting (5) into (4), is determined by the incident energy E' and the scattering angle ϕ alone. The reader is referred to [74] for a more complex model of σ_m, where in particular Rayleigh scatter is taken into account.

Remark 1. The decompositions (2) and (3) are approximations, which may be justified in nuclear medicine. On a closer look, the situation is much more complex, as the attenuation coefficient depends on the Z-number of the atom causing the interaction [31].

We assume that the unknown emission source and attenuation map are compactly supported with $\mathrm{supp}(f)$, $\mathrm{supp}(\mu) \subset Z$, where Z is a cylinder whose axis a_Z is the axis of rotation of the SPECT camera. Boundary information may then be added to (1), firstly by assuming that there is no incoming radiation:

$$I^- := I(x^-, \theta, E) = 0 \qquad \text{for all } \theta \in \mathbb{S}^2,\ E \geqslant 0 \tag{6}$$

at every entry boundary point $x^- \in \partial Z$; that is, $\theta \cdot \nu(x^-) < 0$ with $\nu(x^-)$ the exterior normal to Z at x^-. Secondly, the outgoing radiation is partially detected:

$$I^+ := I(x^+, \theta, E) = d(x^+, \theta, E) \qquad \text{for all } \theta \in \mathbb{S}^2 \text{ with } \theta \cdot a_Z = 0,\ E \geqslant 0 \tag{7}$$

at every exit boundary point $x^+ \in \partial Z$; that is, $\theta \cdot \nu(x^+) > 0$, but only along directions θ perpendicular to the cylinder axis a_Z. With these preparations, the inverse problem for SPECT may now be cast as follows.

Find the right-hand side f and the electron density map ρ provided that the energy transfer terms $I^\pm$ on the boundary of the cylindrical region are known.

The inverse problem, even when understood in a localized sense, is largely open. Some interesting partial results are presented in [39], and a uniqueness result which essentially covers the case of computed tomography (CT) is given in [5, 22]. To our knowledge, the earliest contributions are due to Anikonov (see e.g. [3]), who considers the even more general case, where the attenuation and scatter terms are assumed to be independent and not linked through the common unknown $\rho(x)$ as in our cast. In [4] it is shown under suitable assumptions that if the damping coefficient μ along with scattering kernels σ_1, σ_2 produce identical boundary transport $I^\pm$, then $\sigma_1 = \sigma_2$, $f_1 = f_2$, an interesting mathematical support for the physically motivated assumption (3). A more recent work on the inverse problem is [1], where a two-dimensional transport model is discussed.

2.2. *Monochromatic sources*

In many applications, the radioactive tracer is monochromatic, i.e. emits at a unique photon energy E_0, $f(x, E) = f(x)\,\delta(E - E_0)$. Under that hypothesis, and if we have an ideal SPECT camera with perfect spatial, angular and energy resolution, equipped with a perfect parallel hole collimator, which accepts only photons which travel perpendicular to the camera surface, the full three-dimensional transport model (1) simplifies. Indeed, as $f(x, E) = 0$ for $E > E_0$, and since no incoming radiation is allowed, there is no energy transport at energies $E > E_0$, i.e., $I(x, \theta, E) = 0$ for $E > E_0$. Therefore, at the nominal energy E_0, the scatter term vanishes, and the equation reduces to

$$\theta \cdot \nabla I(x, \theta, E_0) + \mu(x, E_0)\, I(x, \theta, E_0) = f(x), \tag{8}$$

with $\theta \in \mathbb{S}^2$, $x \in \mathbb{R}^3$. Clearly (8) decouples into a series of equations along lines perpendicular to the cylinder axis a_Z. Using the boundary information (6), we may integrate (8) on each line separately. The three-dimensional inverse problem then decomposes into a series of two-dimensional inversions in transaxial slices.

Let us restrict x to one of these two-dimensional slices, which we rename $\mathbb{R}^2$. Choose a standard parametrization (θ, s) for the lines in $\mathbb{R}^2$ by letting $\theta = (\cos\phi, \sin\phi)$, and $\theta^\perp = (-\sin\phi, \cos\phi) \in \mathbb{S}^1$, $s \in \mathbb{R}$: represent x on the line (s, θ) as $x = t\theta + s\theta^\perp$. On setting $\mu(x) := \mu(x, E_0)$, using the boundary information (7) and on letting $d(x^+, \theta) := I(x^+, \theta, E_0)$, the boundary value problem in that slice is equivalent to the nonlinear integral equation

$$\mathcal{R}[\mu]f(\theta, s) := \int_{-\infty}^{\infty} f(s\theta^\perp + t\theta)\, \exp\left(-\int_t^{\infty} \mu(s\theta^\perp + \tau\theta)\, \mathrm{d}\tau\right) \mathrm{d}t = \mathrm{d}\,(\theta, s), \tag{9}$$

to be solved for the unknown f and μ simultaneously. For fixed μ, the linear operator $f \to \mathcal{R}[\mu]f$ is known as the attenuated Radon transform (cf [56]).

Remark 2. Notice that (9) is usually introduced as a simplified version of (1), where the scatter terms have been omitted. Our argument shows that it is the *correct* model for monochromatic sources.

Following [58, 60], *inverting the attenuated Radon transform* refers to the inversion of $f \to \mathcal{R}[\mu]f$ for given μ. The problem of finding or estimating μ (and possibly f) from the emission data $d = \mathcal{R}[\mu]f$ prior to solving for the unknown f is sometimes referred to as the *identification problem*.

Even for monochromatic emission sources, very little is known about the inverse problem for (1). Some partial positive answers under strong additional assumptions are presented in [39]. Now, as we have just learned, if we neglect the scattering contribution and look at the nominal energy level E_0 only, it seems natural to ask whether the nonlinear operator $(f, \mu) \to \mathcal{R}[\mu]f$ is invertible, or whether at least appropriate classes of functions f and μ may be identified where such a statement is valid. The following argument, pointed out to the authors by an anonymous referee, shows that uniqueness in this sense fails without additional assumptions on f and μ.

Example. Suppose μ is rotationally invariant, $\mu(x) = \mu(|x|)$, and of class $\mathcal{C}^1$, and let f, μ be compactly supported on the unit disc. Following Quinto [66, theorem 3.1], we expand $f \in L^2$ into spherical harmonics,

$$f(r\theta) = \sum_{lm} f_{lm}(r)\, Y_{lm}(\theta), \qquad f_{lm}(r) = \int_{\mathbb{S}^1} f(r\theta)\, Y_{lm}(\theta)\, \mathrm{d}\theta.$$

Then the image also has a spherical harmonics expansion

$$\mathcal{R}[\mu]f(s,\theta) = \sum_{lm} (\mathcal{R}[\mu]f)_{lm}(s)\, Y_{lm}(\theta),$$

and the coefficients are obtained as

$$(\mathcal{R}[\mu]f)_{lm}(1/t) = 2\int_a^t T_l(s/t)\, f_{lm}(1/s)\, U((1-(s/t)^2)^{1/2}/s, 1/t)\, s^{-2}(1-(s/t)^2)^{-1/2}\, \mathrm{d}s \tag{10}$$

for $0 < a < 1$ fixed such that $f(x) = 0$ for $|x| > a$. Here $U(s,t) = \exp\{-\int_t^\infty \mu(s\theta + \tau\theta^\perp)\,\mathrm{d}\tau\}$ is the divergent beam transform, which due to our assumption on μ depends only on the parameters s, t. Moreover, T_l denotes the Tchebychev polynomial of degree l.

Now let $d = \mathcal{R}[\mu]f$, so $d = \sum_{lm} d_{lm}\, Y_{lm}$, then each f_{lm} is the unique solution $h(s)$ of a generalized Abel integral equation

$$g(t) = \int_a^t E(s,t)\,(t-s)^{-1/2}\, h(s)\, \mathrm{d}s \tag{11}$$

where the kernel E is of class $\mathcal{C}^1$ on $[a, 1]^2$, $E(s, s) \neq 0$, $g' \in L^2$, $g(a) = 0$. Indeed, reading (10) correctly gives the form (11) on setting $E_l(s,t) := 2T_l(s/t)U(s,t)t(t+s)^{-1/2}$ and $g_{lm}(t) = d_{lm}(1/t)$. The solution is then $h(s) = f_{lm}(1/s)$ on $[a, 1]$.

Now let $\tilde{\mu}$ be another rotationally invariant function of class $\mathcal{C}^1$, with the corresponding $\tilde{U}$ and $\tilde{E}_l$. Consider the Abel integral equations (10), (11), with the same data d_{lm}, but with kernels $\tilde{E}_l$. Let $\tilde{f}_{lm}$ be the corresponding unique solutions. Then, if convergent in L^2, the series $\tilde{f}(r\theta) = \sum_{lm} \tilde{f}_{lm}(r)\, Y_{lm}(\theta)$ gives rise to a second pair $(\tilde{f}, \tilde{\mu})$ such that $\mathcal{R}[\tilde{\mu}]\tilde{f} = \mathcal{R}[\mu]f = d$. Convergence issues may for instance be avoided by choosing f rotationally invariant as well.

This example shows that some assumptions of (f, μ) are needed to enable a uniqueness result for the operator $(f, \mu) \to \mathcal{R}[\mu]f$.

Recently, an important uniqueness result has been obtained by Novikov [58, 60], who presented an inversion formula for the attenuated Radon transform, which is structurally of the form $\mathcal{N}[\mu]d = f$ for $\mathcal{R}[\mu]f = d$. That is, it inverts $f \to \mathcal{R}[\mu]f$ for fixed μ. Invertibility of the attenuated Radon transform may also be derived from the result in [1], when the scattering kernel in the two-dimensional transport model is ignored. Local uniqueness results for generalized attenuated Radon transforms had been obtained before in [32, 53].

Remark 3. The situation in PET is similar to the monochromatic case in SPECT, as the pair of photons is always emitted at $E_0 = 511$ keV. With a perfect energy resolution, and with perfect electronic collimation, scattered photons are identified, and (1) decouples into equations along lines. In contrast with SPECT, the boundary information in PET is symmetric: $I(x^+, \theta, E_0) = I(x^{++}, -\theta, E_0)$, where x^{++} is the exit point opposite x^+ on the same line. This leads to the known nonlinear integral equation (cf [7])

$$\exp\{-\mathcal{R}\mu\}\mathcal{R}f = d \tag{12}$$

which is simpler than (9). The major advantage of (12) over its alter ego (9) is that as soon as $\mu(x)$ is known, the data may be corrected for attenuation, and the problem reduces to inverting the Radon transform.

3. Camera blurring

The transport equation (1) and its special case (9) assume an ideal camera with perfect spatial angular and energy resolution, a mathematical fiction. A real camera will of course have a finite spatial resolution and scan a finite number of directions only. In addition, a realistic model will have to account for the imperfections of the collimator, which will in practice have a nonzero opening angle. Geometric models assuming cylindrical holes are for instance given in [26, 74]. In reality, we have to add knowledge about the internal resolution of the camera photo-multiplier tubes and include effects like septal penetration in the lead collimator. Then we obtain the full camera response function, or *camera blurring*. It turns out that the response of a point source δ_p at a position p at distance d to the camera may very accurately be described as $\mathcal{R}_b[0]\delta_p = \eta\phi_{\sigma(d)}$, where $\phi_{\sigma(d)}$ is a two-dimensional Gaussian centred at the projection p' of p on the camera, with variance $\sigma(d) = \sigma_0 + md$ a known affine function of d, and where η is the efficiency of the collimator, which is of the order of 10^{-4} or even 10^{-5}; see [8, 48] for details. In other terms, a real camera blurs with a distance dependent Gaussian filter. Finally, a realistic camera will use an energy window $E_0 \pm \Delta E$ about the nominal photon energy E_0. This means that scatter effects re-enter the scene even for monochromatic emission sources, since the ideal model in section 3 was based on a fictional perfect energy resolution.

In order to account for these effects we will have to consider discretized versions of (1) or (9). In our numerical tests we use a standard discretization as for instance given in [6, 15, 71]. An interesting alternative discretization was proposed by Dicken [26, 27]. Another possible choice is a wavelet basis [76]. With the appropriate changes, our method may use any of these discretizations.

For further use, let us introduce a mnemonic notation to discern among the different models. While $\mathcal{R}[\cdot]$ refers to the attenuated Radon transform (9), $\mathcal{R}_b[\cdot]$ is the modification of (9) which accounts for camera blurring and attenuation, but ignores scatter effects. Similarly, $\mathcal{R}_s[\cdot]$ stands for the operator form of (1), for instance [25, 74], so it includes attenuation and scatter effects, but ignores blurring. Finally, $\mathcal{R}_{sb}[\cdot]$ stands for the complete model [74], which accounts for both effects. We will use the operator symbols for both continuous and discrete models, as it will always be clear from the context which case we have in mind.

4. Physical attenuation and scatter correction

The current trend in SPECT image reconstruction is to alleviate the difficulties of the inversion of (1) or (9) by estimating the tissue attenuation μ prior to solving for f. This is done by performing a transmission scan either in succession or in parallel with the SPECT scanning, and the approach is referred to as transmission SPECT. Similar lines are taken in PET. Algorithmically, this may be cast as follows.

Method 1

1^0 Using the transmission data d_t, invert the Radon transform $\mathcal{R}_s\mu = d_t$ to obtain the attenuation map $\mu(x, E_1)$ at the energy level E_1 of the (usually monochromatic) transmission source.
2^0 Estimate $\mu(x, E_0)$ from $\mu(x, E_1)$.
3^0 With $\mu = \mu(\cdot, E_0)$ known, obtain f by inverting $\mathcal{R}_b[\mu]f = d_e$, where d_e are the emission data at the energy E_0 of the emission source.

The appealing aspect of method 1, which is used in practice, is that step 3^0 may be performed via iterative methods based on the EM algorithm and its variations, so that it may include attenuation, scatter *and* blurring effects: $\mathcal{R}_{sb}[\mu]f = d_e$. In the latter case, step 2^0 requires estimating $\mu(\cdot, E)$ from $\mu(\cdot, E_1)$ for every E in an appropriate energy window about E_0. Along with these positive points, this approach also encounters several problems, which we need to discuss in order to understand the complexity of the problem.

In configurations where the emission/transmission data are acquired simultaneously, it has been observed that the higher energy isotope, usually the SPECT tracer, will down scatter into the energy window of the transmission isotope, generating cross-talk between the two procedures. If not accounted for, this leads to artifacts in the reconstructed images. This has been studied in [18], and the authors suggest that if the transmission source used a higher energy isotope than the SPECT tracer the impact of the cross-talk could be somewhat reduced.

Formally, the cross-talk implies that the data acquired are $d_e(E) + d_t(E)$, and that we are not able to segregate them into $d_e(E_0)$ and $d_t(E_1)$. This could be addressed by the following.

Method 2

1^0 Solve $\mathcal{R}_s\mu + \mathcal{R}_{sb}[\mu]f = d_e + d_t$ simultaneously for the unknown f and μ (or ρ).
2^0 Keep the μ so obtained, and reconstruct f by solving $\mathcal{R}_{sb}[\mu]f = d_e$ for f.

While adapted to the physical nature of the problem, this approach encounters major numerical difficulties due to the nonlinearity in μ, and is hardly feasible at present. This means that method 1 has to be maintained. Reference [18] makes propositions how to deal with the cross-talk effect.

As a remedy to the spilling over effect of the higher energy isotope into the lower energy window, it has been proposed that the transmission scan be performed in succession with the SPECT scan. Unfortunately, this provokes a new problem: co-registration of two images acquired with different geometries, a nontrivial task. Moreover, if x-ray CT imaging or other sources with several energy peaks are used for the transmission part, step 2 in method 1 becomes a major problem. The phenomenon is sometimes referred to as the beam hardening effect (see e.g. [77]).

In cases where the mentioned drawbacks of transmission SPECT are tolerable, a new reconstruction method for 360° scans may be obtained via Novikov's inversion formula [58, 60].

Method 3

1^0 Use the transmission data d_t to invert the Radon transform $\mathcal{R}_s\mu = d_t$ and obtain $\mu(x, E_1)$ at the energy level E_1 of the transmission source.

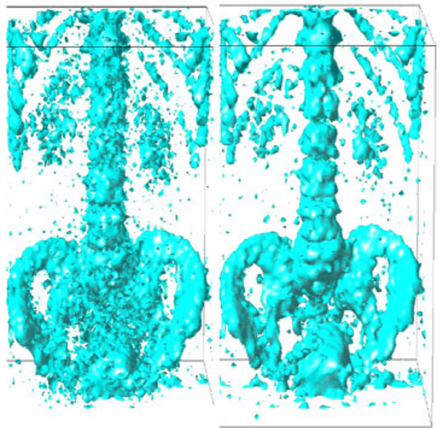

Figure 4. Effect of collimator blurring correction: the left-hand image uses parallel geometry; the right-hand image reconstructed with collimator response.

2^0 Estimate $\mu(\cdot, E_0)$ from $\mu(\cdot, E_1)$.

3^0 With $\mu = \mu(\cdot, E_0)$ known, divide the three-dimensional region of interest into transaxial slices $\nu = 1, \ldots, T$, and regroup μ and the emission data d_e accordingly. In each slice, invert the two-dimensional attenuated Radon transform through Novikov's formula $f^\nu = \mathcal{N}[\mu^\nu]d_e^\nu$.

4^0 Obtain f by stacking the f^ν, $\nu = 1, \ldots, T$ together.

Numerical approaches to $f = \mathcal{N}[\mu]d$ are proposed in [43, 58]. Unfortunately, despite its beauty, the inversion formula is unlikely to improve on existing approaches like method 1, as it is not clear in which way it could account for the collimator blurring effects. It is known that camera blurring *has to be accounted for* in order to improve the visual resolution of the reconstructed images, as demonstrated by figure 4. A competitive implementation of the inversion formula $f = \mathcal{N}[\mu]d$ would have to find ways to take this effect into account.

5. Mathematical attenuation and scatter correction

5.1. Methods

Transmission SPECT or PET clearly sets *the* benchmark for the other approaches to be discussed here, since it introduces additional information into the process. On the other hand, performing the additional transmission scan has several drawbacks. It clearly increases the radioactive impact of the procedure, and requires maintaining an additional source of radioactivity in the clinical environment. Moreover, it may produce artifacts in the reconstruction methods due to the cross-talk between the emission and transmission isotopes.

In consequence, it is attractive to estimate μ and f using the emission data only. We refer to these methods as analytical or *mathematical correction methods*, and the most general approach could be based on the following.

Method 4

1^0 Solve $\mathcal{R}_{sb}[\mu]f = d$ simultaneously for the unknown f and μ.

2^0 Keep the μ so obtained, and reconstruct f by solving $\mathcal{R}_{sb}[\mu]f = d$ for f using an iterative procedure.

The crucial part is step 1^0 here, which leads to a difficult large scale optimization problem. If we replace $\mathcal{R}_{sb}$ by the sparser $\mathcal{R}_b$, the problem is within the reach of large scale optimization procedures, even though matrix vector products require calculations on the fly. In order to reduce the numerical burden of method 4, we propose the following mixed strategy, which simplifies step 1^0 in method 4.

Method 5

1^0 Divide the three-dimensional region of interest into transaxial slices, $\nu = 1, \ldots, T$, and regroup the emission data accordingly. In each slice, invert the two-dimensional transform (9), $\mathcal{R}[\mu^\nu]f^\nu = d^\nu$, simultaneously with respect to f^ν and μ^ν.
2^0 Form a three-dimensional attenuation map μ by stacking the slices μ^ν.
3^0 Invert the three-dimensional transform $\mathcal{R}_b[\mu]f = d$ with respect to f.

This method has been used in our numerical experiments. The numerical details of step 1^0 will be discussed in sections 7.2 and 7.3.

5.2. The identification problem

Methods 4 and 5 suggest that we take a look at the various ways in which step 1^0 could be attacked. This is referred to as the identification problem, and the methods used in the past may roughly be divided into three groups. We recall the most important approaches and discuss whether they will be feasible in three dimensions (method 4), and whether they may serve to include scatter correction.

A first approach, initiated by Natterer [55], tries to estimate the unknown $\mu(x)$ using the Helgason consistency formula [56, theorem II. 6.2]. Structurally, it is of the form $\mathcal{C}[\mu]d = 0$, whenever $d = \mathcal{R}[\mu]f$. Unfortunately, it does not fully characterize the unknown μ, while its alter ego, the corresponding consistency formula for the PET projection data, does (see [56] for the details). A consistency formula for SPECT which fully characterizes the range of the attenuated Radon transform has been obtained by Novikov [61].

The Helgason consistency formula was used in [72] to obtain a rough estimate of the unknown attenuation map. The resulting method, called ConTraSPECT, was further tested in [35, 44]. The idea of ConTraSPECT is not to try to estimate μ exactly, but to use an approximation featuring a uniformly attenuating ellipse with five unknown parameters to compensate for the attenuation effects. The limitation of this approach is that the Helgason formula is only valid for a full 360° scan, while in clinical practice it is often preferable to acquire data on a 180° tour. Further limitations are that the attenuation map so obtained is not suitable to include scatter correction. Moreover, a three-dimensional analogue as for instance required in method 4 does not exist.

A second type of identification method, also pioneered by Natterer [57], tries to solve (9) by fitting a template or reference μ_0 to the unknown μ. Here a deformation procedure carrying μ_0 into μ has to be specified, and the parameters of that process are fitted using (9), the consistency formula or a combination of both. This idea has been revived various times, leading for instance to automatized learning procedures which, starting out with a list of known 'good' cases (f, μ), try to fit the template based on this prior knowledge. As opposed to the first type of method, the attenuation maps so obtained may very well be used to include scatter correction, and they are in principle open to three-dimensional optimization as method 4.

The third type of mathematical approach was initiated by Censor *et al* [19], who solved the discretized equation (9) directly for (f, μ) via an iterative procedure. A more recent approach of this type is Dicken [26, 27]. The author proposes a functional analytic setting for (9), and then derives regularization strategies in the sense of Tychonov motivated by the operator theoretic form of the problem. This leads to an optimization problem of the form

$$(G) \qquad \min_{f,\mu} \|\mathcal{R}[\mu]f - d\|^2 + \mathcal{S}_1[\mu] + \mathcal{S}_2[f]$$

where the smoothing terms $\mathcal{S}_1[\mu]$, $\mathcal{S}_2[f]$ are based on a Tychonov regularization in an appropriate Sobolev space, and where $\|\cdot\|$ stands for the corresponding Sobolev norm. Work in the same spirit is for instance Krol *et al* [41], where again an optimization problem of the form (G), but with the standard L^2-norm, is used to solve for f and μ. We notice that model (G) arises from a Gaussian law for the measurement noise.

An interesting way to solve (G) has recently been developed by Bronnikov [10–12]. Using the fact that the operator $\mathcal{R}[\mu]f$ is linear in the variable f, the approach of Golub and Pereira [34] allows the author to split the joint minimization (G) over (f, μ) into an outer minimization over μ and an inner minimization over f, solving the inner linear least squares problem using the pseudo-inverse $\mathcal{R}[\mu]^\dagger$. As a consequence, problem (G) may now be cast as an optimization over μ alone:

$$\min_{\mu} \|\mathcal{R}[\mu]\mathcal{R}[\mu]^\dagger d - d\|_2^2 + \mathcal{S}_1[\mu].$$

The inconvenience here is that derivatives will depend on the pseudo-inverse $\mathcal{R}[\mu]^\dagger$ of $\mathcal{R}[\mu]$, and that the objective may even be discontinuous at positions where $\mathcal{R}[\mu]^\dagger$ changes rank. Nonetheless, the method is reported to work well with synthetic data, and the reconstruction has the additional advantage that the cross-talk effect reported in [26, 27, 41] (see also section 7.2) is avoided. Notice in addition that the inner minimization may be stabilized by including a Tychonov regularizing term $\mathcal{S}_2[f] = \alpha\|f\|^2$, or more general quadratic expressions in f.

As already observed in [19, 71], see also [15], there is strong reason to model radioactive emissions through a Poisson statistic. This means that we are led to replace the Gaussian optimization problem (G) by the Poisson likelihood function. After discretizing the operator (9), this leads to a minimization problem of the form

$$(P) \qquad \min_{\mu,f} \sum_{j=1}^{M}\sum_{k=1}^{S}\left(\sum_{i=1}^{N} R_{ijk}[\mu]f_i - d_{jk}\log\left(\sum_{i=1}^{N} R_{ijk}[\mu]f_i\right)\right) + \mathcal{S}[\mu, f]$$

where d_{jk} are the discrete data, and $R_{ijk}[\cdot]$ stands for the discretized operator (cf [6, 26, 71]). Notice that in this setting, the regularizing term $\mathcal{S}[\mu, f] = \mathcal{S}_1[\mu] + \mathcal{S}_2[f]$ may be interpreted as a Bayesian prior defined on the parameter spaces for μ and f. In our experiments we will compare model (P) to the Gaussian model (G), where both programs use suitable regularizers to address the ill-posedness of the problem. Except for the numerical difficulties, this could be extended to three dimensions (method 4), and the attenuation map μ so obtained may be used to include scatter correction.

The most prominent way to solve a maximum likelihood problem is the expectation maximization (EM) algorithm, and not surprisingly, there exists an abundance of literature on EM approaches to the inverse problem of ECT. The EM algorithm, [6, 15, 71], along with its modifications like OSEM [38] or RBI [14], works best on the linear inverse problem when μ is known. Moreover, due to the limited number of iterations, the EM algorithm is ideally suited for the full scatter and blurring model $\mathcal{R}_{sb}$ (see e.g. [54]). Its drawback is that regularizing terms reflecting a Bayesian approach are not easily included, or may only be added at the cost of a considerable slowdown. For a way how this could at least be partially overcome see [68].

In the case where both f and μ are unknown, extensions of the usual EM scheme are required. Three such approaches are [41, 62, 64], which use sophisticated definitions of the incomplete data space and/or suitable linearizations to address the nonlinear problem.

In our numerical experiments we have solved both the Poisson program (P) and the Gaussian program (G) via nonlinear optimization techniques. Our approach differs from previous work in two important points. We use physically motivated regularizers, and we optimize the variables (f, μ) jointly. Except [10–12], to our knowledge all previous approaches based on (P) or (G) alleviated the difficulties of a joint minimization by a succession of separate minimizations in μ and f. This coordinate descent approach has a bad reputation in optimization due to failure of convergence.

6. Regularizers—old and new

In this section we shall present possible choices of regularizers $\mathcal{S}_1[\mu]$ and $\mathcal{S}_2[f]$ for problems (G) and (P). Since there is no risk of ambiguity, we will continue to use the same operator symbols for both continuous and discrete models.

6.1. Traditional choices

Despite its merits as a classical and well analysed prototype, Tychonov regularization, appearing as $\mathcal{S}[f] = \alpha\|f\|_2^2$ or $\mathcal{S}[f] = \alpha\|\nabla f\|_2^2$, is known to have too strong a smoothing effect. As a remedy, the so-called flat zone regularizer $\mathcal{S}[f] = \alpha\|\nabla f\|_1$ has been proposed (cf [21]), and in numerical experiments, confirmed by several groups, turned out to be an interesting alternative, particularly so in cases where images with large regions of homogeneous grey level are reconstructed. Our own experience with the flat zone regularizer in dynamic tomography [9] confirms this behaviour.

Remark 4. We present a heuristic argument explaining the success of the flat zone regularizer. Consider for simplicity the one-dimensional penalty program

$$(P_\alpha) \qquad \min_f \|Rf - d\|_2^2 + \alpha\|f'\|_1$$

with a linear operator R. The related tolerance program is

$$(T_\epsilon) \qquad \begin{array}{ll} \text{minimize} & \|f'\|_1 \\ \text{subject to} & \|Rf - d\|_2 \leqslant \epsilon. \end{array}$$

Observe that a solution f_α to (P_α) is also a Kuhn–Tucker point f^ϵ for (T_ϵ) with $\epsilon = \epsilon(\alpha) := \|Rf_\alpha - d\|_2$. Conversely, if the inequality constraint is active at a solution f^ϵ of (T_ϵ), the latter turns out a solution f_α of (P_α) for an appropriate $\alpha = \alpha(\epsilon)$. So both approaches may be considered equivalent, since the inequality constraint will be active in all cases of practical interest. Now consider a finite difference discretization of (T_ϵ). Making the change of variables $g_i = f_i - f_{i+1}$, we recast (T_ϵ) as

$$(\tilde{T}_\epsilon) \qquad \begin{array}{ll} \text{minimize} & \|g\|_1 \\ \text{subject to} & \|RAg - d\|_2 \leqslant \epsilon \end{array}$$

where $f = Ag$ is that change of variables. This means that we minimize the one-norm of g over an elliptic cylinder, and the minimum is found by scaling the norm ball until it touches the cylinder from outside. Now recall that the one-norm ball has $2n$ extreme points, n the dimension of the discretized g, and it is highly likely that the contact is in one of the extreme points, or an extreme k-face. At any rate, the contact point is bound to have many differences g_i equal to zero, which produces the flat zones observed in numerical experiments with this regularizer. Needless to say, the same phenomenon could be obtained by many other choices of norms, which renders the use of $\|\nabla f\|_1$ somewhat accidental.

Remark 5. The equivalence of the tolerance model and the penalty model suggest a search strategy for the correct choice of the penalty parameter α. Suppose that the error terms $e_i = (Rf)_i - d_i$ are independent random square integrable. Then we have $\mathbb{E}\|Rf - d\|^2 = \sum_{i=1}^n \mathbb{E}(e_i^2) = \sum_{i=1}^n \mathbb{V}(d_i)$. Therefore, if for instance the d_i were Poisson distributed, $\mathbb{E}(d_i) = \mathbb{V}(d_i)$, a good estimate would be

$$\|Rf - d\|^2 \approx \sum_{i=1}^n d_i =: \epsilon^2. \tag{13}$$

This means that the correct choice of the penalty parameter is $\alpha = \alpha(\epsilon)$ for that ϵ. Since this formula is not explicit, a line search in α is required. The tolerance condition (13) may be used as a stopping criterion in the numerical procedures (cf [9]).

6.2. Problem oriented regularization

One may argue that the flat zone regularizer is appropriate if a bird's view of the situation is taken. On the other hand, regularizers specially adapted to the physical nature of the image generating process in ECT should be expected to give better results. In [37], several possibilities motivated by the nuclear medicine context have been discussed. Here we propose regularizers based on physically motivated penalty terms. A first example, proposed in [50, 51], uses high pass filtering to penalize high frequencies which may *a priori* be identified as noise contributions.

Observe that in the absence of tissue attenuation, the Fourier slice theorem $\widehat{\mathcal{R}f(\cdot,\theta)}(\sigma) = \hat{f}(\sigma\theta)$, cf [56], tells us that every detail present in the unknown image f should be visible in some of the projections $\mathcal{R}f(\cdot,\theta)$. In other terms, any detail finer than $2\pi/b$ in a candidate image f (b is the known bandwidth of the sinogram $\mathcal{R}f$) could *not* originate from the true source f, and should be attributed to a noise source. This statement remains at least approximately correct if the effect of tissue attenuation is taken into account without being predominant (see [47]). We therefore propose a regularizing high pass filter of the form

$$\mathcal{S}[f] = \alpha\|\mathcal{H}_b(f)\|_2^2 = \alpha\|\phi_b \cdot \hat{f}\|_2^2, \tag{14}$$

with cut-off bandwidth b fixed moderately above the known bandwidth of the sinogram b_0.

It was observed in [56, 67] that the two-dimensional spectrum $\widehat{Rf}(\sigma, k)$ of the Radon transform $\mathcal{R}f(s,\theta)$ is concentrated on a region of bowtie shape in the (σ, k) frequency plane (see figure 5). This suggests a cut-off function $\phi_{b,m}$ having $\phi_{b,m}(\sigma,k) \approx 1$ outside, and $\phi_{b,m}(\sigma,k) = 0$ inside a bowtie region of horizontal bandwidth $|\sigma| \leqslant b$, thickness $|k| \leqslant m$ at the axis $\sigma = 0$ and slope ≈ 1 for a source f supported on the unit disc. This leads to a structured high pass filter of the form

$$\mathcal{S}[f] = \alpha\|\mathcal{H}_{b,m}(\mathcal{R}f)\|_2^2 = \alpha\|\phi_{b,m} \cdot \widehat{\mathcal{R}f}\|_2^2. \tag{15}$$

Following the analysis in [47], a similar result is even correct for the attenuated Radon transform, suggesting the even better adapted joint filter

$$\mathcal{S}[\mu, f] = \alpha\|\mathcal{H}_{b,m}(\mathcal{R}[\mu]f)\|_2^2. \tag{16}$$

Due to the nonlinearity in μ, however, (16) is numerically demanding.

The spatial bandwidth b in (15) should be moderately above the known spatial bandwidth of the sinogram, while the thickness m of the bowtie region has to be adapted to the *dynamics* of the attenuation map μ. For details we refer to [47].

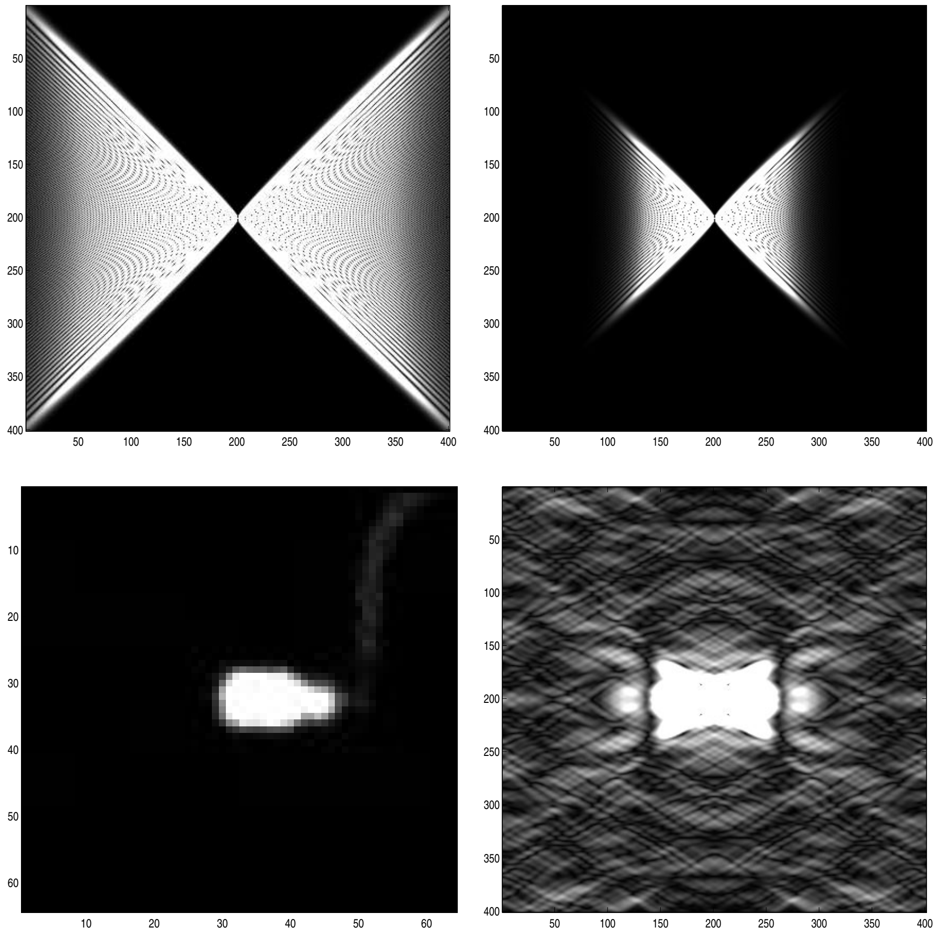

Figure 5. The two-dimensional spectrum of the Radon transform $\widehat{\mathcal{R}f}$ is concentrated on a region of bowtie shape in the frequency plane. This remains qualitatively correct for the attenuated Radon transform. The top left and top right images show the spectrum $\widehat{\mathcal{R}\phi_\sigma}$ of a Gaussian source with different resolution, corresponding to signals with different spatial bandwidth. The bottom right image shows the bowtie belonging to the spectrum of the Radon transform of a physical source defined by a small hot bottle shown bottom left.

It is near at hand to use the same type of regularizer for the unknown attenuation map μ, possibly with different cut-off frequencies b. In practice it is easy to obtain estimates for the bandwidth of the attenuation map μ by evaluating reconstructions obtained via transmission data processed in parallel with the SPECT scan (cf [16, 17]). As already observed by Natterer [57], and confirmed by our experiments here, if model (9) is used, it is not necessary to know μ to a high precision. What is required is knowledge of $\exp\{-\mathcal{D}\mu(x,\theta)\}$ at certain points x and along certain directions θ. In consequence, only mild regularization on μ is required, and the effect of different regularizers on the quality of the reconstructed f is marginal.

In our numerical tests we have compared the one-norm, the two-norm and the high-pass regularizer (14) using the Gaussian model (G) and the Poisson model (P). The bowtie regularizer was tested in [46] and in [9] for dynamic emission tomography (dSPECT). Notice that all the regularizers discussed either immediately generalize to the three-dimensional case, or have natural three-dimensional analogues.

6.3. Convergence theorem

In order to justify our choice of regularizer

$$\mathcal{S}[\mu, f] = \alpha(\|\mathcal{H}_{b_1}\mu\|_2^2 + \|\mathcal{H}_{b_1,0}(\mathcal{R}\mu)\|_2^2 + \|\mathcal{H}_{b_2}f\|_2^2 + \|\mathcal{H}_{b_2,0}(\mathcal{R}f)\|_2^2)$$

we proceed to prove a convergence theorem in the spirit of [29, 30] or [42]. In order to do this, we first have to give a precise functional analytic setting for problem (9).

We recall that $(\mu, f) \to \mathcal{R}[\mu]f$ is continuous as an operator defined on the positive cone $L^2(D)_+ \times L^2(D)_+$ of the Hilbert space $L^2(D) \times L^2(D)$, taking values in the Hilbert space $L^2(Z, w)$ with weight $w(s) = (1-s^2)^{-1/2}$. We shall use the notation $\|\cdot\|_w$ for the induced norm. Here we assume the unknown f and μ supported on the unit disc D in $\mathbb{R}^2$, which means that the attenuated Radon transform is defined on the cylinder $Z = [-1, 1] \times \mathbb{S}^1$, cf [26, 27, 56] for details.

Note that a singular value decomposition for the Radon transform $\mathcal{R}$ as an operator between $L^2(D)$ and $L^2(Z, w)$ was established by Louis [45], showing that $\mathcal{R}$ is ill posed on these spaces. However, $\mathcal{R}$ is also continuous as an operator mapping $L^2(D)$ in $L^2(Z)$ with the standard norm, and it seems convenient to measure data in this norm $\|\cdot\|_2$. In particular, $\mathcal{R}$ is still ill posed with respect to this norm, and the same observation is true for the attenuated Radon transform. Many numerical approaches are in fact based on the standard L^2-norm on the data space (see [10, 41]).

Let us now define the correct setting for the Fourier transform and for the corresponding high pass filters. Let Ω be a domain containing D, and let $\mathcal{F}$ be the two-dimensional Fourier transform restricted to $L^2(\Omega)$, as an operator with values in $L^2(\mathbb{R}^2)$. Then $\mathcal{F}^*$ is the inverse Fourier transform followed by the restriction operator. The high pass filter $\mathcal{H}_b$ may then be defined on $L^2(\Omega)$ as

$$\mathcal{H}_b u = \mathcal{F}^*\phi_b \cdot \mathcal{F}u, \tag{17}$$

where ϕ_b will for simplicity be the ideal cut-off function $\phi_b = 0$ for $|s| \leqslant b$, $\phi(s) = 1$ for $|s| > b$, defining a diagonal operator $\hat{u} \to \phi_b \cdot \hat{u}$ in the frequency plane $L^2(\mathbb{R}^2)$. Notice that if $\Omega = [-\omega, \omega]^2$, every $f \in L^2([-\omega, \omega]^2)$ may be written as a Fourier series

$$f = \sum_{k\in\mathbb{Z}^2} (f, u_k)\, u_k, \qquad u_k(x) = (2\omega)^{-1} \mathrm{e}^{\mathrm{i}\pi x\cdot k/\omega},$$

with $\hat{f}_k = (f, u_k)$ the Fourier coefficients of f on $[-\omega, \omega]^2$. In that case, the ideal high and low pass filters are defined as

$$\mathcal{H}_b f = \sum_{|k|>b} \hat{f}_k u_k, \qquad \mathcal{L}_b f = \sum_{|k|\leqslant b} \hat{f}_k u_k.$$

For fixed $b_1, b_2 > 0$, let $D \subset \Omega = [-\omega, \omega]^2$. Fix an upper bound $\bar{\mu} > 0$ for the unknown attenuation map, and a lower bound $-\underline{f}$ with $\underline{f} > 0$ for f. We consider the Tychonov or penalty type regularization problem

$$(P_{\alpha,d}) \qquad \begin{array}{ll} \text{minimize} & \|\mathcal{R}[\mu]f - d\|_w^2 + \alpha\big(\|\mathcal{H}_{b_1}f\|^2 + \|\mathcal{H}_{b_2}\mu\|^2\big) \\ \text{subject to} & f, \mu \in L^2(D),\ f \geqslant -\underline{f}, 0 \leqslant \mu \leqslant \bar{\mu}. \end{array}$$

We assume that there exist ideal data d_* which satisfy $d_* = \mathcal{R}[\mu_*]f_*$ for certain $0 \leqslant \mu_* \leqslant \bar{\mu}$ and $f_* \geqslant -\underline{f}$ in $L^2(D)$.

Theorem. *Let $\alpha_n \to 0^+$, $\delta_n \to 0^+$ and $\epsilon_n \to 0^+$, with $\delta_n^2/\alpha_n \to 0$ and $\epsilon_n/\alpha_n \to 0$. Let the observed data d_n satisfy $\|d_* - d_n\| = \delta_n$. Suppose (f_n, μ_n) is feasible for (P_{α_n,d_n}) and an ϵ_n-approximate solution for that problem. Suppose $\mathcal{H}_{b_1}\mu_* = 0$, $\mathcal{H}_{b_2} f_* = 0$. Then we have the following.*

(1) There exist subsequences $f_{n_k} \to \hat{f}$, $\mu_{n_k} \to \hat{\mu}$ convergent in $L^2(D)$, such that $(\hat{f}, \hat{\mu})$ is a solution to $\mathcal{R}[\mu]f = d_*$. Moreover, $\mathcal{H}_{b_1}\hat{f} = 0$ and $\mathcal{H}_{b_2}\hat{\mu} = 0$.

(2) If (f_*, μ_*) is the only low frequency solution of $\mathcal{R}[\mu]f = d_*$, the sequences $\mu_n \to \mu_*$ and $f_n \to f_*$ converge in $L^2(D)$. Moreover, if we let $\alpha_n \sim \delta_n$, $\epsilon_n \sim \delta_n^2$, then $f_n - f_* = \mathcal{O}(\delta_n^{1/2})$, $\mu_n - \mu_* = \mathcal{O}(\delta_n^{1/2})$.

Proof. Notice first that an ϵ_n-approximate solution to (P_{α_n,d_n}) means a pair (f_n, μ_n) satisfying

$$\|\mathcal{R}[\mu_n]f_n - d_n\|_w^2 + \alpha_n(\|H_{b_1}\mu_n\|^2 + \|H_{b_2} f_n\|^2) \\ \leqslant \|\mathcal{R}[\mu]f - d_n\|_w^2 + \alpha_n(\|H_{b_1}\mu\|^2 + \|H_{b_2} f\|^2) + \epsilon_n \tag{18}$$

for every feasible (f, μ). As a consequence, on substituting f_* and μ_* on the right-hand side of (18), we find $\mathcal{H}_{b_2} f_n \to 0$, $\mathcal{H}_{b_1}\mu_n \to 0$ for

$$\|\mathcal{H}_{b_1}\mu_n\|^2 + \|\mathcal{H}_{b_2} f_n\|^2 \leqslant \frac{\delta_n^2}{\alpha_n} + \frac{\epsilon_n}{\alpha_n} + \|\mathcal{H}_{b_1}\mu_*\|^2 + \|\mathcal{H}_{b_2} f_*\|^2 \to 0 \tag{19}$$

with the last two terms on the right-hand side vanishing by assumption.

Observe that μ_n is bounded in $L^2(D)$ due to the constraint $0 \leqslant \mu \leqslant \bar{\mu}$. We argue that f_n is bounded in $L^2(D)$, too. Clearly $\mathcal{R}[\mu_n]f_n$ is bounded in $L^2(Z, w)$, and since $f_n \geqslant -\underline{f}$ and $\mu_n \leqslant \bar{\mu}$, $\mathcal{R}f_n$ is bounded in $L^2(Z, w)$. Now since $\mathcal{H}_{b_2} f_n$ is bounded, it remains to show that $\mathcal{L}_{b_2} f_n$ is bounded. Observe that $\mathcal{R}(\mathcal{L}_{b_2} f_n)$ is bounded as well. Now by definition the low frequency space is finite dimensional, and $\mathcal{R}$, restricted to this space, has a bounded inverse, which means that $\mathcal{L}_{b_2} f_n$ is bounded as soon as its image sequence $\mathcal{R}(\mathcal{L}_{b_2} f_n)$ is.

Observe that $f_n = \mathcal{L}_{b_2} f_n + \mathcal{H}_{b_2} f_n$ and $\mathcal{H}_{b_2} f_n \to 0$ implies that f_n has a norm convergent subsequence $f_{n_k} \to \hat{f}$, the low frequency space being finite dimensional. Similarly $\mu_{n_k} \to \hat{\mu}$. Then $\mathcal{R}[\mu_{n_k}]f_{n_k} \to \mathcal{R}[\hat{\mu}]\hat{f}$ and $\mathcal{H}_{b_1}\hat{\mu} = 0$, $\mathcal{H}_{b_2}\hat{f} = 0$. But observe that (19) also implies $\mathcal{R}[\mu_n]f_n \to d_*$, and therefore $\mathcal{R}[\hat{\mu}]\hat{f} = d_*$, i.e., $(\hat{\mu}, \hat{f})$ is a solution. This proves (1).

In order to establish the rate of convergence in the case where the low frequency solution (f_*, μ_*) is unique, observe that (19) implies $\mathcal{H}_{b_1}\mu_n = \mathcal{O}(\delta_n)$, $\mathcal{H}_{b_2} f_n = \mathcal{O}(\delta_n)$. Therefore, again from (19), $\mathcal{R}[\mu_n]f_n - d_* = \mathcal{O}(\delta_n)$. But now the argument proceeds along the classical lines (like [29, 30, 42]). □

Remark 6. Let us make a few points of importance for practice. Firstly, to make the theorem more realistic, we may include a discretization of the optimization problem, and by properly relating the discretization parameter h_n to the penalty parameter α_n, obtain a convergence theorem for the discretized program. This could be done along the lines of [59].

Secondly, we have to be aware that the observed data $d^{(n)}$ belonging to the discretized problem will never do us the favour of converging to d_*. What is needed to guarantee this is a regression procedure, the easiest example being a histogram estimator producing the sequence $d_n = \mathcal{S}_n d^{(n)} \to d_*$. Naturally, the filtering induced by $\mathcal{S}_n$, for instance a nearest neighbour filter, has to be adapted to the discretization and the constant α_n. For an example from a different setting of how this could be managed see [59].

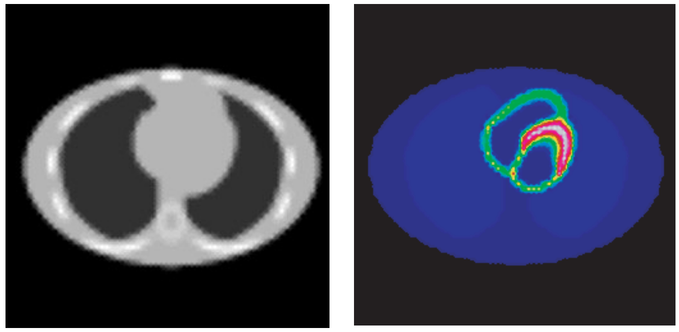

Figure 6. Transaxial slice through the MCAT phantom at the level of the heart. Right, ^{99m}Tc uptake distribution; left, attenuation map at 140 keV.

Remark 7. Note that our hypothesis of a low frequency signal $d_* = \mathcal{R}[\mu_*]f_*$ is somewhat opposite to the point of view in [26, 27], where the author presents a regularization approach valid even for discontinuous, or at least infinite bandwidth signals. Clearly in that case a high pass filter with fixed cut-off frequency could not be completely adapted to the situation, as we do in the theorem. As a consequence, a high pass filter would essentially act like the classical Tychonoff term $\mathcal{S}[f] = \alpha\|f\|_2^2$, and hypotheses as in [29, 30, 42] would be required to prove convergence.

7. Experiments

In order to validate method 5, we have performed two experiments, a simulated case study and a study using a physical phantom.

7.1. Data acquisition

The Mathematical Cardiac Torso (MCAT) phantom is a three-dimensional analytical phantom which models the attenuation coefficient and activity distribution in the upper human torso. For the simulated cardiac SPECT study we have selected a particular slice shown in figure 6. We simulated a ^{99m}Tc-based SPECT tracer and assumed relative concentrations of 75.0, 3.82 and 1.76 in heart, lungs and soft tissues. The attenuation coefficients in the cortical bones, trabecular bones, lungs and muscle at the nominal energy $E_0 = 140$ keV of technetium were chosen as 0.21, 0.17, 0.043 and 0.15 cm^{-1} respectively. The source f and attenuation map μ were both discretized in a 64×64 pixel map with pixels of 6.25×6.25 mm size. A perfect parallel hole collimator was assumed, and synthetic projection data were created via formula (9), which included tissue attenuation, but neither scatter nor blurring. The data were Poisson noised to create a realistic signal-to-noise ratio. A total of 64 projections were scanned over 180°, and alternatively over 360°. The size of the projection bins was 6.25 mm. The total number of detected counts in the slice was approximately 180 000.

The experimental study uses the radiology support device (RSD) striatal phantom, an artificial skull enclosed within material that mimics soft tissue, ears, nose and neck (see figure 10). It has one brain reservoir and four striatal containers. The phantom was filled with a uniform solution of 303 kBq ml^{-1} of ^{99m}Tc in water. The projection data were acquired with an Elscint dual-head spectral SPX camera, equipped with a parallel low energy high resolution

Table 1. Comparison of optimization methods.

	I_1		I_2		I_3		I_4	
—	100	100	28.1	26.6	—	21.0	23.4	22.3
G_0	61.8	53.4	27.4	24.4	—	20.0	22.7	20.5
G_1	*	*	21.2	18.6	—	18.8	21.0	19.3
G_2	*	*	24.4	23.5	—	*	22.5	*
G_h	61.2	*	24.4	22.6	—	*	22.6	20.4
G_{1+c}			20.5					
P_0	66.3	58.8	26.8	25.3	—	20.5	22.7	20.7
P_1	65.3	*	20.4	18.9	—	19.8	21.7	19.6
P_2	65.3	*	24.9	23.0	—	*	22.6	*
P_h	65.8	*	24.3	22.3	—	20.4	*	*
P_{1+c}							21.3	
	180°	360°	180°	360°	180°	360°	180°	360°

(LEHUR) collimator. A total of 60 angles, equally spaced over 360° were scanned, and the projection data were sampled on a 128×128 grid with pixels of size 3.44×3.44 mm^2. The data were sampled over 15 s per view, and corrected for the known decay of the radioisotope, resulting in approximately 400 000 counts per projection. A $\pm$ 10% energy window about the primary photo-peak at $E_0 = 140$ keV was used, and a second emission data set was acquired on a $\pm$ 3% energy window about a secondary photo-peak at 122 keV.

7.2. Results of the simulation

In the simulated study, the true model is described by (9), which is also used to reconstruct μ and f. We use the following.

Two-dimensional algorithm

1^0 Use an initial step I_k, with $k \in \{1, \dots, 4\}$, to calculate a starting point (f^0, μ^0).

2^0 Run the optimizer (G_j), respectively (P_j), with $j \in \{0, 1, 2, h, 1+c\}$ a suitable regularizer, until either the optimizer's internal stopping test or the statistical stopping test (13) applies.

3^0 If the optimizer stops for intrinsic reasons before the statistical stopping test is attained, reduce the penalty constant α and go back to step 2^0. If the statistical stopping test applies before the internal test stops the optimization, increase α and go back to step 2^0.

4^0 The optimization finished, keep the μ so obtained, and reconstruct f by inverting $R[\mu]f = d$ via the EM algorithm.

In a simulated study, the true emission source f_{true} is known, and we may check the quality of each reconstruction f_{rec}. This is done by calculating the relative error

$$e_{\text{rec}} = \frac{\|f_{\text{true}} - f_{\text{rec}}\|_2}{\|f_{\text{rec}}\|_2}. \tag{20}$$

Table 1 shows the results of this test for a considerable number of scenarios. The letters G and P indicate whether the Gauss or Poisson objectives have been used. The index $j \in \{0, 1, 2, h, 1+c\}$ indicates whether no regularizer, the one-norm regularizer, the two-norm regularizer, a high-pass filter or the one-norm regularizer combined with a special filter (21) avoiding cross-talk between the reconstructed f and μ has been used.

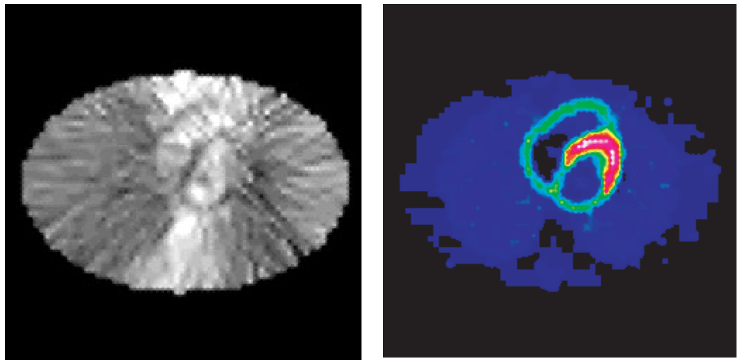

Figure 7. Same transaxial slice reconstructed.

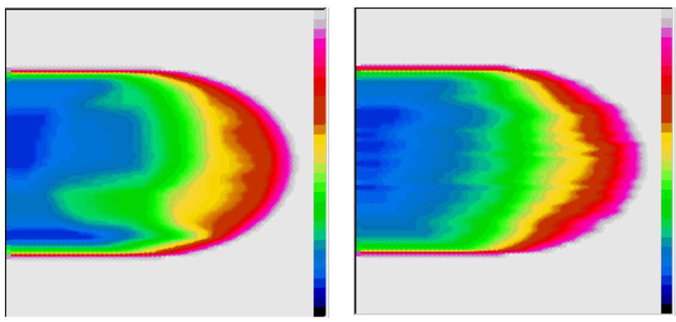

Figure 8. Comparison of the divergent beam transforms $\mathcal{D}[\mu](x, \theta_1)$ for μ_{rec} (left) and μ_{true} (right). As can be seen, despite the sizable difference between the maps, the integrated terms differ only slightly, and almost agree on the region around the heart.

In table 1, we have tested four different ways I_k, $k = 1, \ldots, 4$ to initialize the optimizer. I_1 corresponds to choosing the initial $f^0 = 0$, $\mu^0 = 0$. In I_2, μ^0 is chosen as constant on the contour. Then the linear inversion $\mathcal{R}[\mu^0]f = d$ is solved using program (G), and the result is called f^0. In I_3, the result (f, μ) of the ConTraSPECT method (based on the ideas of Natterer, see [44, 72]) is used as an initial guess. Finally, I_4 chooses μ^0 constant on the contour and obtains f^0 as the EM reconstruction of $\mathcal{R}[\mu^0]f = d$.

In the table, the symbol — means that the method did not apply to that case. The symbol $*$ indicates that the optimizer could not improve the figure of merit compared to the results G_0 or P_0 obtained without the regularizer. An empty field indicates that the method $1 + c$ did not improve over the method G_1 or P_1. In the simulated study, we observed that the flat zone regularizer worked slightly better than the Fourier regularizer, while models (G) versus (P) produced quite similar results.

Remark 8. Notice that if f is reconstructed with the *correct* attenuation map μ_{true} via the Poisson EM algorithm, the relative error amounts to $e_{\text{true}} = 18.1\%$ for a 180° scan, and $e_{\text{true}} = 16.0\%$ on 360°. These errors occur since the data are Poisson distributed random events. In the case of the simulated study, we can observe that these errors e_{true} are close to the possible minimum, by which we mean that the signal-to-noise ratio of f_{true} is only mildly worse than the signal-to-noise-ratio in the data. The latter in turn could be considered *the*

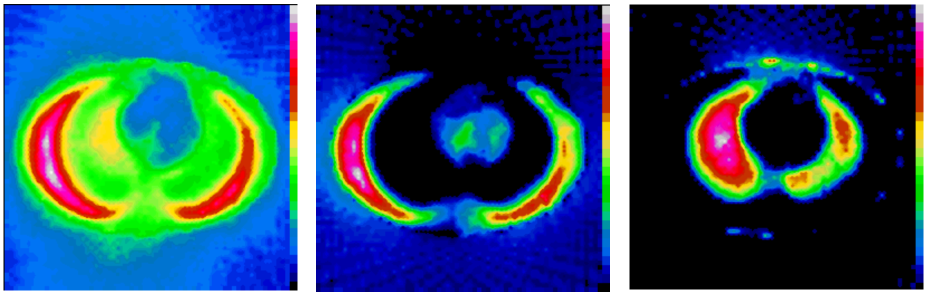

Figure 9. Map of the total error $\int_{\mathbb{S}^1} |\exp\{-\mathcal{D}[\mu_{\mathrm{rec}}](x,\theta)\} - \exp\{-\mathcal{D}[\mu_{\mathrm{true}}](x,\theta)\}|\,\mathrm{d}\theta$ (left); positive part, middle; negative part, right.

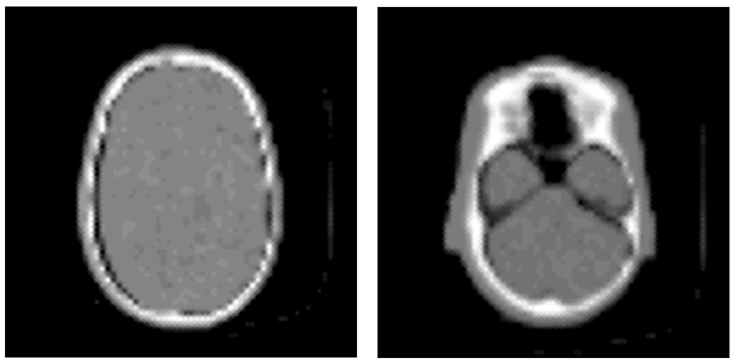

Figure 10. X-ray CT images of the RSD phantom. Left, cortical slice; right, cerebellar slice.

indicator of what could at best be achieved by *any* reconstruction method. The fact that some of the errors e_{rec} in table 1 come close to this margin indicates that our approach worked quite well, and that even a substantially more sophisticated regularization technique is unlikely to improve the quality of the results.

Remark 9. As first observed by Dicken [27] in cardiac studies, the reconstructed attenuation map may feature a shadow artifact of the heart region. We have observed a similar phenomenon here (figure 7, left), and we believe it is due to the fact that the optimization converges to a local minimum. Ironically, this shadow heart with incorrect attenuation coefficients occurred in a region where μ is basically known, and could even be corrected by hand. As this is clearly prohibitive, we have used the additional regularizing term

$$\mathcal{I}[\mu, f] = \alpha \sum_{i=1}^{N} f_i(\mu_{\max} - \mu_i) \tag{21}$$

which penalizes values of μ in places i where the activity f_i is sizeable. This approach works well and reduces the visibility of this shadow artifact quite significantly.

Remark 10. The reconstructed attenuation map in figure 7 (left) should be compared to results obtained by Dicken [26, 27] and Bronnikov [10–12]. As the difference between the true and reconstructed attenuation maps μ in figures 6 and 7 (left) seems significant, some explanation is indicated. We argue that in the given case the reconstruction in figure 7 (left)

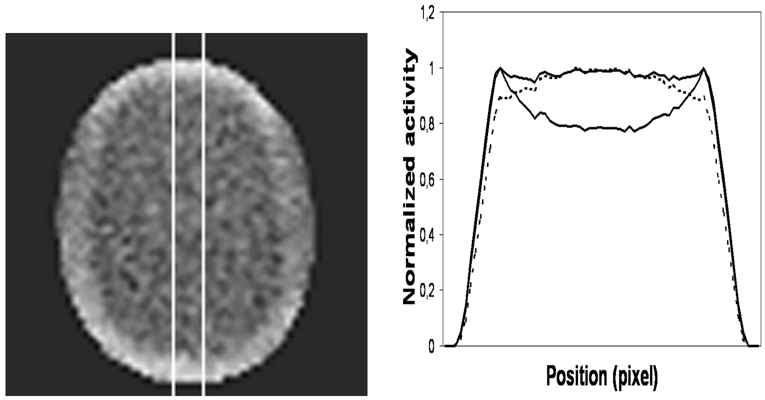

Figure 11. The left image shows the reconstructed attenuation map in the cortical slice. The contour is obtained from the reconstructed secondary photo-peak and mildly overestimated due to thresholding. The right-hand diagram shows reconstructed activity curves along the indicated cut profile: (a) without attenuation correction (lower curve), (b) curve obtained by initial procedure I_4 (dotted), (c) curve obtained via optimizer. The initial is already of good quality, but the optimizer still improves around the edges. The true activity curve is expected to be flat.

is rather 'correct', and that *we should not even expect a reconstruction* μ_{rec} *resembling the true* μ_{true}. Why is that? In fact, first observe that the optimization approach (P), (G) we have chosen requires $\exp\{-\mathcal{D}[\mu]\}$, and not μ, to be accurately known. Secondly, it tries to adjust $\exp\{-\mathcal{D}[\mu_{\text{rec}}](x,\theta)\}$ at positions x where the source $f(x)$ is sizable. Since therefore $\mathcal{D}[\mu_{\text{rec}}](x,\theta)$ is only remotely reliable at positions x where $f(x)$ is weak (see figures 8 and 9), we cannot expect μ_{rec} to be of good quality, as the passage from $\mathcal{D}[\mu]$ to μ is the inversion of an ill posed procedure. This argument is more applicable, the more concentrated the emission source, and in the present case we see that most of the activity is concentrated around the heart region. The impossibility of finding μ with a highly localized $f(x)$ is of course highlighted in the case of a point source $f = \delta_x$, where μ is not even theoretically uniquely determined (cf [36, 56], where this has been observed).

7.3. Results of the phantom experiment

The phantom data have been inverted by the following procedure based on method 5. After preprocessing the data, the ROI was divided into 30 transaxial slices, and the data have been regrouped accordingly.

Three-dimensional algorithm

1^0 In each slice, method I_4 was used to initialize the optimizer.
2^0 In each slice, the two-dimensional algorithm was used to calculate an optimal (f^ν, μ^ν).
3^0 The attenuation maps μ^ν were stacked to obtain a three-dimensional map μ.
4^0 The three-dimensional model $\mathcal{R}_b[\mu]f = d$ was inverted using the EM algorithm to estimate the emission source f.

Estimating the head contour in each slice is required in step 1^0. This is done by reconstructing the emission data acquired at the secondary energy peak, which consists mainly of scattered photons. This leads to a slightly enlarged contour, as seen in figures 11 (left) and 13 (left).

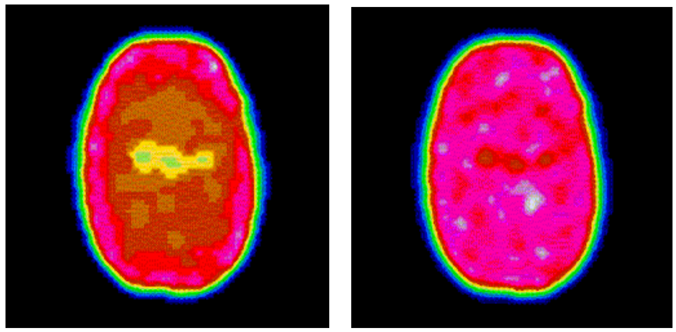

Figure 12. Activity in reconstructed cortical slice. Left-hand image without, right-hand image with attenuation correction according to the three-dimensional algorithm. Both images were obtained with collimator response taken into account. The true activity distribution is known to be homogeneous.

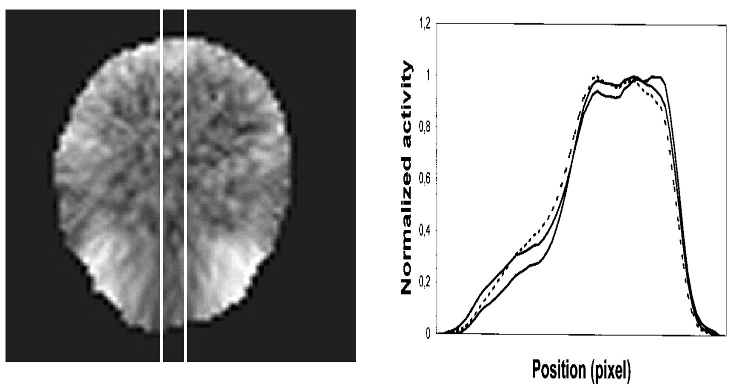

Figure 13. The left-hand image shows the reconstructed attenuation map in the cerebellar slice. The right-hand plot shows activity curves along the profile: (a) without attenuation correction, (b) curve obtained by initial I_4 (dotted), (c) curve obtained via optimizer. The difference is less marked than in the cortical slices.

Figure 12 shows a cortical slice reconstructed without (left) and with attenuation correction (right). Both reconstructions included the collimator response function. The visual inspection clearly shows the improvement obtained through attenuation correction, as the true activity distribution in the cortical slices is known to be homogeneous.

Figure 13 shows a reconstructed attenuation map in a cerebellar slice (left), and the curves along the profile as before show the case without attenuation correction, and the case obtained by the three-dimensional algorithm. The improvement obtained by attenuation correction is less marked in the cerebellar slices.

Finally, figure 14 shows a reconstructed cerebellar slice with (right) and without (left) attenuation correction. The activity in the brain container is expected to be homogeneous. Attenuation correction improves most on the upper part, where the true image is expected to show no activity.

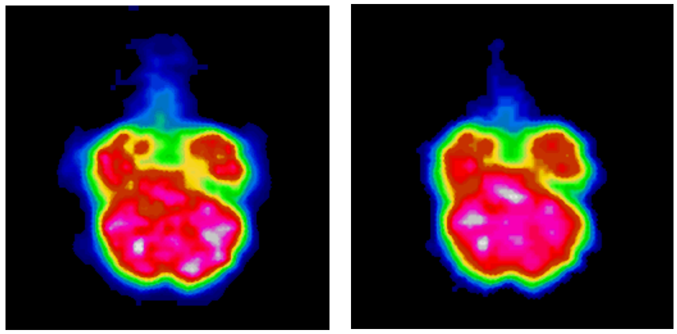

Figure 14. Activity in reconstructed cerebellar slice. Left-hand image without, right-hand image with attenuation correction according to the three-dimensional algorithm. Both images were obtained with collimator response taken into account.

8. Conclusion

We have derived the image reconstruction problem in ECT from the photon transport equation. Based on this model, two possible inversion strategies have been discussed. The first uses additional transmission measurements to alleviate the difficulty of the inverse problem. The second uses the emission data only and is mathematically demanding.

Our numerical experiments, which are based on the second strategy, use a variational approach in tandem with Tychonoff type regularization. They confirm that when diligently handled, mathematical attenuation and scatter correction remains a promising alternative or complement to the transmission based strategies currently privileged by the engineering community. The future will probably see a combination of both approaches.

Acknowledgments

We thank both referees for their valuable comments, which helped to improve the paper, and one of the referees for bringing [1] to our attention.

References

[1] Arbuzov E V, Bukhgeim A L and Kazantsev S G 1998 Two-dimensional tomography problems and the theory of A-analytic functions *Sib. Adv. Math.* **8** 1–20

[2] Anikonov Y E 1995 Multidimensional Inverse and Ill-Posed Problems for Differential Equations (*Inverse and Ill-Posed Problems Series*) (Utrecht)

[3] Anikonov D S 1974 Inverse problems for the transfer equation *Diff. Urav.* **10** 7–17 (Engl. transl. *Mathematics Institute, Siberian Division, Acad. Sci. USSR*)

[4] Anikonov D S 1984 Uniqueness of simultaneous determination of two coefficients of the transport equation *Sov. Math. Dokl.* **30** 149–51

[5] Anikonov D S 1985 Uniqueness of the determination of a coefficient of the transport equation for a special type of source *Sov. Math. Dokl.* **32** 511–5

[6] Bauschke H, Noll D, Celler A and Borwein J M 1999 An EM-algorithm for dynamic emission tomography *IEEE Trans. Med. Imaging* **18** 252–61

[7] Bilger K, Kupferschläger J, Müller-Schauenburg W, Nüsslin F and Bares R 2001 Threshold calculation for segmented attenuation correction in PET with histogram fitting *IEEE Trans. Nucl. Sci.* **48** 43–50

[8] Blinder S A, Celler A, Wells R G, Thomson D and Harrop R 2001 Experimental verification of 3D detector response compensation using the OSEM reconstruction method *IEEE Trans. Nucl. Sci. (TNS-MIC Conf. Record (San Diego, CA))*

[9] Blondel C, Noll D, Maeght J, Celler A and Farncombe T 2002 Comparison of different figure of merit functions for dynamic single photon emission computed tomography (dSPECT) *IEEE Trans. Nucl. Sci.* at press
[10] Bronnikov A V 1999 Numerical solution of the identification problem for the attenuated Radon transform *Inverse Problems* **15** 1315–24
[11] Bronnikov A V 1995 Approximate reconstruction of attenuation map in SPECT imaging *IEEE Trans. Nucl. Sci.* **42** 1483–8
[12] Bronnikov A V 2000 Reconstruction of the attenuation map using discrete consistency conditions *IEEE Trans. Med. Imaging* **19** 451–62
[13] Brownell G L *A History of Positron Imaging. Presentation in Celebration of the 50th Year of Services of the Author to the Massachusetts General Hospital (1999)*
[14] Byrne C L 1998 Accelerating the EMML algorithm and related iterative algorithms by rescaled block-iterative methods *IEEE Trans. Image Process.* **7** 100–9
[15] Carson R E and Lange K 1985 The EM parametric image reconstruction algorithm *J. Am. Stat. Assoc.* **80** 20–2
[16] Celler A, Sitek A and Harrop R 1996 Reconstruction of multiple line source attenuation maps *IEEE Nuclear Science Symp. Conf. Record (1996)* pp 1420–4
[17] Celler A, Sitek A, Stoub E, Lyster D, Dykstra C, Worsley D and Fung A 1997 Development of a multiple line source attenuation array for SPECT transmission scans *J. Nucl. Med.* **38** 215ff
[18] Celler A, Axen D, Togane D and El-Khatib J 2000 Investigation of scatter in SPECT transmission studies *IEEE Trans. Nucl. Sci.* **47** 1251–6
[19] Censor Y, Gustafson D E, Lent A and Tuy H 1979 A new approach on the emission computerized tomography problem: simultaneous calculation of attenuation and activity coefficients *IEEE Trans. Nucl. Sci.* **41** 1594–600
[20] Cercignani C 1988 *The Boltzmann Equation and its Applications (Applied Mathematical Sciences vol 67)* (Berlin: Springer)
[21] Chavent G and Kunisch K 1996 Regularization of linear least squares problems by total bounded variation *Technical Report* Karl-Franzens Univ., Graz
[22] Choulli M and Stefanov P 1999 An inverse boundary value for the stationary transport equation *Osaka J. Math.* **36** 87–104
[23] Csiszár I 1975 I-divergence geometry of probability distributions and minimization problems *Ann. Prob.* **3** 146–58
[24] Csiszár I and Tusnády G 1984 Information geometry and alternating minimization procedures *Statis. Decisions* 205–37 Suppl.
[25] Dautrey R and Lions J L 1984 *Analyse Mathématique et Calcul Numérique* vol 9 (Paris: Masson)
[26] Dicken V 1999 Simultaneous activity and attenuation reconstruction in emission tomography *Inverse Problems* **15** 931–60
[27] Dicken V 1998 Simultaneous activity and attenuation reconstruction in single photon emission computed tomography, a nonlinear ill-posed problem *PhD Thesis* Universität Potsdam
[28] DiPerna R J and Lions P L 1989 On the Cauchy problem for Boltzmann equations: global existence and weak stability *Ann. Math.* **130** 321–66
[29] Engl H W, Kunisch K and Neubauer A 1989 Convergence rates for Tikhonov regularisation of non-linear ill-posed problems *Inverse Problems* **5** 523–40
[30] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer)
[31] Evans R D 1955 *The Atomic Nucleus* (New York: McGraw-Hill)
[32] Finch D V 1986 Uniqueness for the attenuated x-ray transform in the physical range *Inverse Problems* **2** 197–203
[33] Frey E C and Tsui B M W 1993 A practical method for incorporating scatter in a projector–backprojector for accurate scatter compensation in SPECT *IEEE Trans. Nucl. Sci.* **40** 1107–16
[34] Golub G H and Pereira V 1973 The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate *SIAM J. Numer. Anal.* **10** 413–32
[35] Gourion D, Hatchondo X, Gantet P, Noll D and Esquerré J P 2001 Comparison of two methods for SPECT attenuation correction without transmission measurements *IEEE–MIC Conf. Record (San Diego, CA)*
[36] Hertle A 1988 The identification problem for the constantly attenuated Radon transform *Math. Z.* **197** 9–13
[37] Higdon D M, Bowsher J E, Johnson V E, Turkington T G, Gilland D R and Jaszczak R J 1997 Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data *IEEE Trans. Med. Imaging* **16** 516–26
[38] Hudson H M and Larkin R S 1994 Accelerated image reconstruction using ordered subsets of projection data *IEEE Trans. Med. Imaging* **13** 601–9
[39] Isakov V 1998 *Inverse Problems for Partial Differential Equations (Springer Series in Applied Math. Sci. vol 127)* (Berlin: Springer)

[40] Johnson C A and Sofer A 2000 A primal-dual method for large scale image reconstruction in emission tomography *SIAM J. Optim.* **11** 691–715

[41] Krol A, Manglos S H, Bowsher J E, Bassano D A and Thomas FD 2001 An EM algorithm for estimating SPECT emission and transmission parameters from emission data only *IEEE Trans. Med. Imaging* **20** 218–32

[42] Kunisch K and Ring W 1993 Regularization of nonlinear illposed problems with closed operators *Numer. Funct. Anal. Optim.* **14** 389–404

[43] Kunyansky L 2001 A new SPECT reconstruction algorithm based on the Novikov's explicit inversion formula *Inverse Problems* **17** 293–306

[44] Laurette I, Clackdoyle R, Welch A, Natterer F and Gullberg G T 1998 Comparison of three applications of ConTraSPECT *IEEE NSS and MIC Conf. Record (Toronto, 1998)*

[45] Louis A 1989 *Inverse und schlecht gestellte Probleme* (Stuttgart: Teubner)

[46] Maeght J 1999 Analyse et méthodes pour un problème inverse en tomographie dynamique *Thèse de Doctorat* l'Université Paul Sabatier, Toulouse, France

[47] Maeght J and Noll D 2000 Resolution in dynamic emission tomography *SIAM J. Math. Anal.* **31** 1100–20

[48] Majorel C, Gantet P, Gourion D, Payoux P and Esquerré J P 2001 Modélisation de la réponse du collimateur dans un algorithme de reconstruction OSEM en TEMP *Méd. Nucl.—Imagerie Fonctionnelle Métabol.* **25** 259–68

[49] Maréchal P and Lannes A 1997 Unification of some deterministic and probabilistic methods for the solution of linear inverse problems via the principle of maximum entropy on the mean *Inverse Problems* **13** 135–51

[50] Maréchal P, Togane D, Celler A and Borwein J M Numerical assessment of the stability of reconstruction processes for computed tomography

[51] Maréchal P, Togane D and Celler A 1999 A new reconstruction methodology for computerized tomography: FRECT (Fourier Regularized Computed Tomography) *Conf. Records IEEE Medical Imaging Conf. (Seattle, 1999)*

[52] Markoe A 1984 Fourier inversion of the attenuated radon transform *SIAM J. Math. Anal.* **15** 718–22

[53] Markoe A and Quinto E T 1985 An elementary proof of local invertibility for the generalized and attenuated radon transforms *SIAM J. Math. Anal.* **16** 1114–9

[54] Narayanan M V, King M A, Leppo J, Dahlberg S, Pretorius P H and Gifford H 2001 Optimization of regularization of attenuation and scatter-corrected ^{99m}Tc cardiac SPECT studies for defect detection using hybrid images *IEEE Trans. Nucl. Sci.* **48** 785–9

[55] Natterer F 1983 Computerized tomography with unknown sources *SIAM J. Appl. Math.* **43** 1201–12

[56] Natterer F 1986 *The Mathematics of Computerized Tomography* (Stuttgart: Teubner)

[57] Natterer F 1993 Determination of tissue attenuation in emission tomography of optically dense media *Inverse Problems* **9** 731–6

[58] Natterer F 2001 Inversion of the attenuated Radon transform *Inverse Problems* **17** 113–9

[59] Noll D 1997 Consistency of a nonlinear deconvolution method with applications to image restoration *Adv. Math. Sci. Appl.* **7** 789–808

[60] Novikov R G An inversion formula for the attenuated x-ray transform *Ark. Mat.*

[61] Novikov R G On the range characterization for the two-dimensional attenuated x-ray transformation *Preprint*

[62] Nuyts J, Dupont P, Stroobants S, Bennick R, Mortelmans L and Suetens P 1999 Simultaneous maximum *a posteriori* reconstruction of attenuation and activity distribution from emission sinograms *IEEE Trans. Med. Imaging* **18** 393–403

[63] Oxenius J 1986 *Kinetic Theory of Particles and Photons. Theoretical Foundations of Non-LTE Plasma Spectroscopy (Springer Series in Electrophysics vol 20)* (Berlin: Springer)

[64] Panin V Y, Zeng G L and Gullberg G T 2001 A method of attenuation and emission activity reconstruction from emission data *IEEE Trans. Nucl. Sci.* **48** 131–8

[65] Phelps M E, Mazziotta J C and Schelbert H (ed) 1986 *Positron Emission Tomography and Autoradiography: Principles and Applications in Brain and Heart* (New York: Raven)

[66] Quinto E T 1983 The invertibility of rotation invariant radon transforms *J. Math. Anal. Appl.* **91** 510–22

[67] Rattey P A and Lindgren A G 1981 Sampling the 2-D Radon transform *IEEE Trans. Acoust. Speech Signal (Proc. ASSP-29)* pp 994–1002

[68] Setzepfand B 1992 ESNM: ein rauschunterdrückendes EM-Verfahren für die emissionstomographie *Dissertation* Universität Münster

[69] Singh M 1983 An electronically collimated gamma camera for single photon emission computed tomography. I: theoretical considerations and design criteria *Med. Phys.* **10** 421–7

[70] Singh M and Doria D 1983 An electronically collimated gamma camera for single photon emission computed tomography. II: image reconstruction and preliminary experimental measurements *Med. Phys.* **10** 428–35

[71] Vardi Y, Shepp L A and Kaufman L 1985 *J. Am. Stat. Assoc.* **80** 8–20

[72] Welch A, Clark R, Natterer F and Gullberg GT 1997 Toward accurate attenuation correction in SPECT without transmission measurements *IEEE Trans. Med. Imaging* **16** 532–41

[73] Welch A, Gullberg G, Christian P and Datz F 1995 A transmission based scatter correction technique for SPECT in inhomogeneous media *Med. Phys.* **22** 1627–35

[74] Wells R G, Celler A and Harrop R 1998 Analytic calculation of photon distribution in SPECT projections *IEEE Trans. Nucl. Sci.* **45** 3202–14

[75] Young T 1995 Attenuation correction in SPECT by simultaneous reconstruction of emission and intrinsic attenuation information using an ART algorithm *PhD Thesis* Graduate School of Syracuse University

[76] Zhao S, Welland G and Wang G 1997 Wavelet sampling and localization schemes for the Radon transform in two dimensions *SIAM J. Appl. Math.* **57** 1749–62

[77] 1996 *Mathematics and Physics of Emerging Biomedical Imaging Committee on the Mathematics and Physics of Emerging Dynamic Biomedical Imaging* (Washington, DC: National Academy Press)