Attenuation Correction Using SPECT Emission Data Only

Daniel Gourion[1], Dominikus Noll[2], Pierre Gantet[1], Anna Celler[3], member IEEE, Jean-Paul Esquerré[1,4], member IEEE

[1] Université Paul Sabatier, Laboratoire Traceurs et Traitement de l'Image, 133, route de Narbonne, 31077 Toulouse, France
[2] Université Paul Sabatier, Mathématiques pour l'Industrie et la Physique, 118 route de Narbonne, 31062 Toulouse, France
[3] Vancouver Hospital & Health Sciences Center, Vancouver, BC, Canada, V5Z 1M9
[4] Hôpital Toulouse Purpan, Service de Médecine Nucléaire, Chef de Service, Place du Docteur-Baylac, 31059 Toulouse, France

Abstract

A major step towards quantitative SPECT imaging may be achieved if attenuation, scatter and blurring effects are accounted for in the reconstruction process. Here we consider an approach which simultaneously estimates the unknown attenuation coefficient and the emission source using the emission data only. This leads to an inverse mathematical problem which could no longer be solved via iterative procedures like the well-known EM-algorithm. Instead, a regularization approach based on nonlinear optimization techniques is used. We present a successful strategy of the analytic type, and we test it in a simulated case study.

I. Introduction

In its original understanding, the term scatter or attenuation correction referred to methods which tried to improve SPECT or PET reconstructed images by correcting or modifying the emission data *prior to reconstruction*. This included uniform attenuation correction, or methods which estimate the Compton scattered photons from secondary energy peak information, and modify the emission data by subtracting the scattered contribution. The idea was to compensate for the fact that tissue attenuation and scatter were not included in the filtered backprojection algorithm, used until recently to reconstruct SPECT and PET images. While it is clear that this approach is not justified rigorously, some of these heuristics have been reported to work with considerable success. We refer to this family of methods as the *approach via modification of the data*. See [6] for an overview on such methods.

The meaning of the term attenuation and scatter correction has changed significantly over recent years, and is now generally used to delineate strategies, where the unknown tissue attenuation map $\mu(x)$ is estimated via transmission scans performed either simultaneously or in succession with the emission scanning. We refer to this as the transmission SPECT correction methods, or simply as the *physical approach* to attenuation and scatter correction. See [11] for a discussion and references on at least five different source configurations.

Here we will be mainly concerned with a third type of attenuation or scatter correction methods, which try to estimate the unknown attenuation coefficient *using the emission data only*. As compared to the physical methods, this approach has to get by with less information, and therefore leads to more complicated mathematical inversion procedures. We will refer to it as analytical or *mathematical attenuation and scatter correction*. The purpose of this work is to present and discuss several such analytical attenuation correction method, based on two nonlinear optimization programs, (P) and (G), to compare them, and to substantiate their viability using a simulated case study and a phantom study.

II. Physical Attenuation Correction

Transmission SPECT attenuation correction sets a benchmark for the analytical methods to be discussed here. However, transmission SPECT has its own limitations and drawbacks, and one may argue that in the future, physical and analytical attenuation and scatter correction will probably co-exist and complement one another.

Transmission based attenuation correction clearly increases the patient dose, and requires maintaining an additional radioactive source in the clinical environment. In the same vein, if the emission/transmission scanning are to be performed in parallel, the choice of the transmission isotope will restrict the choice of the compatible SPECT isotopes.

Even in successful approaches 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 (see [11]). This leads to artifacts in the reconstructed images. In [11], the authors suggest that if the transmission source used the higher energy isotope than the SPECT tracer, the impact of the cross-talk could be somewhat reduced. But even then, some of the indicated restrictions persist.

Spilling over of the higher energy isotope would not matter if the emission and transmission scans were performed in succession, using either the same or a different camera systems. However, this will complicate the protocol, and may lead to the nontrivial problem of co-registration of two images acquired with different geometries. In addition, if X-ray CT imaging is used for the transmission imaging, the attenuation map could not be entirely adapted to the SPECT tracer energy due to beam hardening.

These issues, which we have only touched upon here, make it seem interesting to have alternative procedures, which would allow to estimate the unknown tissue attenuation map using the emission data only. We will start investigating this possibility, by looking at some of the analytical methods proposed in the

past, and shall then present the optimization models (G) and (P) on which our present approach is based.

III. Analytical Attenuation Correction

Analytical attenuation correction has already a rich history, and the existing methods may roughly be grouped in three categories.

A first class of methods, pioneered by F. Natterer [26], uses the Helgason consistency formula (see e.g. [27, Theorem II.6.2]) to estimate the unknown attenuation map $\mu(x)$ prior to reconstructing the emission source. This idea has recently been revived in [35, 21], see also [16], leading to a method called ConTraSPECT, where the authors fit an elliptical dummy attenuation map featuring six parameters. This approach, which in many cases works surprisingly well, is only feasible for a 360^0 camera rotation, since Helgason's formula has no substitute for different cases. Notice further that the artificial attenuation maps obtained by this type of methods are unlikely to serve the purpose of scatter correction.

A second type of mathematical methods, also initiated by Natterer [28], tries to fit a template or reference attenuation map $\mu_0(x)$, along with a prespecified deformation procedure, to the individual case, using either the consistency formula, or by estimating μ and f simultaneously via the attenuated Radon transform (1) below. This approach could obviously be extended or refined by using a stack of model attenuation maps and applying automatic learning procedures when matching the reference object. The attenuation maps obtained by this class of methods are of better quality than in the first case, and may very well be used to include scatter correction.

Our present contribution belongs to a third form of mathematical attenuation correction methods, which uses the attenuated Radon transform (cf. [27]):

$$\begin{aligned}\mathcal{R}[\mu]f(s,\theta) &= \int_{-\infty}^{\infty} f(s\theta + t\theta^{\perp})\, e^{-\int_t^{\infty} \mu(s\theta+\tau\theta^{\perp})\,d\tau}\, dt \\ &= p(s,\theta) \qquad (1)\end{aligned}$$

to simultaneously estimate the unknown attenuation map $\mu(x)$ and emission source $f(x)$ from the emission data $p(s,\theta)$. Here $p(s,\theta)$ is the datum acquired on the line referenced by (s,θ), and $f(x)$ and $\mu(x)$ are the emission source and attenuation coefficient respectively. Equation (1) was used in [12], where the authors choose a Poisson model for the statistics of the emission data p. It has recently been revived by V. Dicken [13, 14], who uses a Tychonov type regularization to invert equation (1). This requires solving an optimization problem of the form

$$(G) \quad \min_{f,\mu} \quad \|\mathcal{R}[\mu]f - p\|^2 \,+\, \alpha\,\mathcal{I}[\mu,f]$$

featuring an appropriate regularization term $\alpha\mathcal{I}[\mu,f]$ which penalizes and thereby avoids highly irregular reconstructions μ, f that would match the data within the acceptable error tolerance (see Section V). An approach in the same spirit is [19]. In (G) we minimize the negative log-likelihood of a Gaussian law, and possible choices of the norm $\|\cdot\|$ will be discussed in the next section. The regularizing term $\mathcal{I}[\mu,f]$ will then play the role of a Bayesian prior, and possible choices of these regularizers are discussed in Section VII, while steering the penalty parameter α is discussed in Section V.

An interesting way to solve (G) was recently proposed by Bronnikov [3, 4, 5]. Exploiting the fact that $\mathcal{R}[\mu]f$ is linear in f, the author first solves the inner linear least squares problem in (G) with respect to the variable f, using the pseudo-inverse $\mathcal{R}[\mu]^{\dagger}$. The remaining nonlinear least squares problem

$$\min_{\mu} \quad \|\mathcal{R}[\mu]\mathcal{R}[\mu]^{\dagger}p - p\|^2 \,+\, \alpha\mathcal{I}[\mu]$$

in the unknown μ is then of smaller dimension. This is in fact a special case of an algorithm proposed by Golub and Pereira in [15]. The method is reported to work well on a simulated example. In particular, it is reported to avoid the undesired cross talk between the reconstructions of f and μ observed e.g. by Dicken. Yet another approach, based on a direct inversion of (1), is Zeng et al. [31], where the authors use a singular value decomposition to partially linearize the nonlinear dependence of $\mathcal{R}[\mu]f$ on μ.

It is tempting to try to solve problem (G) by an alternating procedure which optimizes with either μ or f fixed, iterating back and forth between these modes. Several approaches of this type have been presented recently. The reconstructions all seem to suffer from a strong cross-talk phenomenon between f and μ, not surprisingly so, as alternating procedures of this type are not convergent as a rule and are often reported to fail. Our present investigation shows that it is preferable to use a genuinely convergent optimization strategy, combined with a suitable choice of the regularizers as discussed in Section VII.

Another method to solve (1) uses a Poisson statistic for the emission data. This leads to an optimization problems of the form

$$\begin{aligned}(P) \qquad \min_{f,\mu} \quad & \sum_{j=1}^{M}\sum_{k=1}^{S}\Big\{\sum_{i=1}^{N} R_{ijk}[\mu]f_i \\ & -p_{jk}\log\Big(\sum_{i=1}^{N} R_{ijk}[\mu]f_i\Big)\Big\} \,+\, \alpha\,\mathcal{I}[\mu,f]\end{aligned}$$

which up to constant terms minimizes the negative Poisson log-likelihood function of the independent Poisson distributed random vector $p = (p_{jk})$ with expectation $\mathbb{E}(p_{jk}) = \sum_{i=1}^{N} R_{ijk}f_i$, augmented by a regularizing term $\alpha\mathcal{I}[\mu,f]$ as above. We solve (P) directly using nonlinear optimization methods. The nonlinearity of (1) respectively (P) in the variable μ foils using iterative procedures in the spirit of the notorious EM-algorithm. Nonetheless, there *is* an alternative way in which the EM-algorithm could still be used. We shall discuss this in Section VI.

We use the following standard notations: Let $i = 1,\ldots,N$ be the discretization of the emission image and attenuation map into pixels or voxels, f_i the activity of the ith element, μ_i its attenuation coefficient. Let $k = 1,\ldots,S$ be the angular positions or stops of the camera, and let $j = 1,\ldots,M$

enumerate camera bins. Then $R_{ijk}[\mu]$ may be understood as the conditional probability that a photon originating from voxel i is recorded in the camera bin j at the camera position k or θ_k. Accordingly, $p = (p_{jk})$ represents the projection data, with p_{jk} the number of counts detected in camera bin j during stop k. Clearly some modifications may be adopted. We may switch to different basis functions, see e.g. [38], and we may even choose different bases for μ and f in order to account for their different spatial resolutions.

IV. Nonlinear least squares

An important problem of the nonlinear least squares approach (G) is the correct choice of the norm $\|e\|^2$ of the forward error $e = \mathcal{R}[\mu]f - p$ in data space. There is evidence that the Euclidean norm may *not* be an appropriate candidate. This point was already made in [27, 13, 14], while [3, 4] still uses the Euclidean norm. Here we will give some support for the choices suggested by [27, 13, 14] by comparing the approaches (G) and (P).

Notice that [27] shows that for fixed μ, the linear operator $f \to \mathcal{R}[\mu]f$ is continuous between the spaces $L^2(D)$ on the unit disk D and $L^2(Z, w)$ on the cylinder $Z = [-1, 1] \times D$ if the weighted measure $w(s)\, ds\, d\theta$, with $w(s) = (1 - s^2)^{-1/2}$, is used on Z. This weighted Euclidean norm attributes high cost to mismatch in places with few counts, typically located at the ends of the camera cross section. The probabilistic model (P) using the Poisson statistics supports this argument.

Let $\ell(w, p) = w - p \log w$, so that the negative Poisson log-likelihood function in (P) may be written as $\sum_{j=1}^{M} \sum_{k=1}^{S} \ell(w_{jk}, p_{jk})$ with $w_{jk} = \sum_{i=1}^{N} R_{ijk}[\mu] f_i$. Taylor expansion gives the well-known estimate

$$\ell(w, p) - \ell(p, p) \leq \frac{1}{2} \frac{(w - p)^2}{p}$$

valid for $0 < w < 2p$. This means that as soon as the parametric forward estimate w_{jk} is close enough to the datum p_{jk}, in particular, $w_{jk} < 2p_{jk}$, the negative log-likelihood objective in (P) will, up to the constant term $\sum_{jk} \ell(p_{jk}, p_{jk})$, be close to the weighted norm expression

$$\frac{1}{2}\|w - p\|_p^2 := \frac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{S} \frac{(w_{jk} - p_{jk})^2}{p_{jk}}, \tag{2}$$

which is then the norm we should use if we prefer a Gaussian model (G). The norm (2) coincides with the norm on $L^2(Z, w)$ used above if the source f is a constant function on the unit disk.

V. Stopping and Scaling

While the choice of the regularizer $\mathcal{I}[f, \mu]$ will be discussed in Section VII, we presently dwell on practical aspects of (G) and (P), which concern appropriate stopping rules for the algorithms, as well as a suitable scaling of the variables f and μ.

We generally stop the optimization procedure as soon as the ℓ_2-norm $\|\mathcal{R}[\mu]f - p\|_2^2$ approaches the overall error in the data:

$$\|\mathcal{R}[\mu]f - p\|_2^2 \approx \sum_{jk} p_{jk} \tag{3}$$

This is clearly based on a Poisson statistic of the data p, and was already reported to work well in a different context [2]. Our present experiments confirm this guideline. If the weighted norm $\|\cdot\|_p$ is optimized, it may be preferable to replace (3) by the test

$$\|\mathcal{R}[\mu]f - p\|_p^2 \approx 1, \tag{4}$$

which works equally well in practice. In our experiments we use the rule (3).

Notice that the choice of the penalty parameter α is closely related to this stopping test. If α is chosen too large, we may be unable to achieve the desired error margin (3). On the other hand, choosing α too small will give many candidates (f, μ) which match (3). In that case, the effect of our regularizer is too weak.

Concerning scaling, observe that the highly nonlinear dependence of $\mathcal{R}[\mu]f$ on μ is in strong contrast with the linearity in f, and the gradients of the objective function in (G) or (P) in the variables μ and f may be orders of magnitude apart. This may cause serious problems in practice, and it is preferable to properly scale the nonlinear variable. In our experiments we found that absolute counts for f and the unit m^{-1} for μ worked best. The standard unit cm^{-1} on the contrary produced too strong gradients in μ.

VI. Poisson Model

In this section we discuss practical aspects of the Poisson model (P). In our experiments, we have solved (P) directly using nonlinear optimization methods, since the standard versions of the EM-algorithm, along with its modifications like OS-EM [18] or row action methods [7], are no longer applicable.

As we shall see, following the out-set of [34, 8], it is possible to obtain an extended version of the EM-algorithm, if we accept to perform an explicit optimization step at each iteration, at the cost of slowing down the procedure. This is already the case if Bayesian extensions of the EM-algorithm including regularizers are considered, as shown in [33].

Recall that the EM-algorithm for (P) generates a sequence (f^n, μ^n) of parameter estimates by alternatingly performing an E-step and an M-step [34, 8]. Given current parameters (f^n, μ^n), the E-step creates a new instance q^n of the so-called *complete data vector*, and the M-step then performs a maximum likelihood estimation in the complete data space, based on these hypothetical data q^n. This generates new parameter estimates f^{n+1}, μ^{n+1}.

In our case we assume that the complete data vector $q = (q_{ijk})$ is independent and Poison distributed with expectation $\mathbb{E}(q_{ijk}) = R_{ijk}[\mu] f_i$. This is a hypothetic quantity representing that part of the activity in pixel i which is emitted towards camera bin j during stop k. The incomplete data $p = (p_{jk})$, by definition always those who are actually

sampled, are linked to the complete data q through the forgetful connection $p_{jk} = \sum_{i=1}^{N} q_{ijk}$. With these preparations, the scheme becomes:

E-step. Given parameter estimates f^n and μ^n, form the complete data vector q^n by

$$q_{ijk}^n = p_{jk} \frac{R_{ijk}[\mu^n] f_i^n}{\sum_{i'=1}^{N} R_{i'jk}[\mu^n] f_{i'}^n} \tag{5}$$

M-step. Given the complete data vector q^n, obtain new parameter estimates f^{n+1}, μ^{n+1} by solving the maximum likelihood optimization problem

$$\min_{f,\mu} \quad \sum_{j=1}^{M} \sum_{k=1}^{S} \sum_{i=1}^{N} R_{ijk}[\mu] f_i - q_{ijk}^n \log \big(R_{ijk}[\mu] f_i \big) \tag{6}$$

Some remarks are in order here. Notice first that (5) in the E-step is the formula for the conditional expectation of the complete data vector q, given the data, p, and under the assumption that the current parameter estimates f^n and μ^n are the correct ones. This part is in fact exactly the same as in the well-known static case (see e.g. [34, 8]).

The difference is in the M-step. Since μ has become a variable, the is no hope to solving the minimization over (μ, f) explicitly. What seems worse, due to the variable μ, the M-step no longer splits into N problems of size $2S$, while it *did* split in a series of N problems of size S in the classical case (with μ known and held fixed).

We may nevertheless do one thing and perform the minimization over f first. This may be solved explicitly and gives a formula linking f^{n+1} and μ^{n+1}. To find it, notice that the objective (6) may be recast as

$$\sum_i \sigma_i[\mu] f_i - \tau_i^n \log f_i - \sum_{ijk} q_{ijk}^n \log \big(R_{ijk}[\mu] \big) \tag{7}$$

with the abbreviations $\sigma_i[\mu] = \sum_{jk} R_{ijk}[\mu]$ and $\tau_i^n = \sum_{jk} q_{ijk}^n$. The minimization over f only involves the first two terms, is separable in that variable, and has the explicit solution

$$f_i^{n+1} = \frac{\tau_i^n}{\sigma_i[\mu^{n+1}]} \tag{8}$$

which will of course come into action as soon as we will have obtained μ^{n+1} by solving the remaining optimization of (7) with respect to μ. After substituting (8) into (7), the remaining part of the M-step is, up to terms now constant in μ:

$$\min_{\mu} \quad \sum_i \tau_i^n \log \sigma_i[\mu] - \sum_{ijk} q_{ijk}^n \log \big(R_{ijk}[\mu] \big). \tag{9}$$

This optimization problem with solution μ^{n+1} is of size N, to be solved once per M-step. Altogether, we obtain an iterative procedure, which we may use to estimate μ, and which we cast as an iterative procedure in the parameters (μ, q):

Modified EM-Algorithm

1. Given the synthetic data $q = (q_{ijk})$, obtain a new parameter estimate μ^+ by solving (9).
2. Obtain f^+ via formula (8).
3. Obtain q^+ via formula (5).

Clearly, f may be eliminated from this scheme, which may in consequence be considered an estimation procedure for μ alone. Notice that this suggests ideas like using a coarse grid approximation to estimate μ, in order to accelerate this process.

VII. Regularizers

In a probabilistic setting, regularizing terms may be interpreted as Bayesian priors on the parameter spaces of the Gaussian or Poisson model under consideration, as shown in [17]. In the present section, we discuss possible choices of regularizers $\mathcal{I}[f]$ and $\mathcal{I}[\mu]$ adapted to our problem.

Using a high-pass filter

$$\mathcal{I}[f] = \|\mathcal{H}_b f\|_2^2 = \|\Phi_b \cdot \hat{f}\|_2^2 \tag{10}$$

seems natural, as we expect noise contributions to be of high frequency, which we should then penalize through the regularizing term. But how to choose the cutoff frequency b? As proposed by [24, 25], the Fourier slice theorem could give us a guideline on the choice of b. Observe that without tissue attenuation, $\widehat{\mathcal{R}f}(\sigma, \theta) = \hat{f}(\sigma\theta)$, which tells us that the spatial resolution of the unknown emission source is no better than the spatial resolution of the projections, or put differently, any detail present in the image f should be visible in some of the projections. Consequently, details finer than the known resolution b of the projection p should be attributed to noise sources and penalized through (10).

Clearly, in the presence of tissue attenuation, we have to be conservative about the proposed choice of b, as the Fourier slice theorem will only be approximately true. Nonetheless, (10) works considerably well in practice (see also [2]).

An interesting variation of (10) uses the fact that the 2D spectrum $\hat{p}$ of the attenuated Radon transform $p = \mathcal{R}[\mu] f$ is concentrated on a bowtie shaped region in the frequency plane (cf. [32, 27, 23]). This was first observed in the unattenuated case, but [23] shows that it remains qualitatively correct in the attenuated case. This suggests a regularizer of the form

$$\mathcal{I}[f] = \|\mathcal{H}_{m,b}(\mathcal{R}f)\|_2^2 = \|\Phi_{m,b} \cdot \widehat{\mathcal{R}f}\|_2^2 \tag{11}$$

where $\Phi_{m,b}$ is an appropriate cutoff operator adapted to a bowtie of width $2b$ in the direction of the frequency plane axis belonging to the variable s, and thickness $2m$ at the origin in direction of the frequency plane axis belonging to the variable θ (see Figure X). For details see the above references. Notice that in both formulas (10), (11) we exploit Parseval's identity, which allows us to implement the regularizer in the frequency domain.

A somewhat different regularizer with some popularity in the mathematical community is the so-called flat zone regularizer

$$\mathcal{I}[f] = \|\nabla f\|_1, \tag{12}$$

which modifies the notorious Tychonov term, known to be too smoothing, replacing the Euclidean norm by the 1-norm. This is reported to privilege reconstructed images f featuring flat zones with identical grey values. Some of our experiments confirm this phenomenon (see Section IX). We stop to give an optimizer's point of view explaining this behavior.

Consider for simplicity a 1D linear inverse problem for the abstract operator $\mathcal{R}$. Following (G), we solve $\min_f \|\mathcal{R}f - p\|_2^2 + \alpha\|f'\|_1$ for a fixed penalty constant α. Along with (G) consider the corresponding error tolerance optimization program

$$(\tilde{G}) \quad \begin{array}{ll} \text{minimize} & \|f'\|_1 \\ \text{subject to} & \|\mathcal{R}f - p\|_2 \leq \epsilon \end{array}$$

for a fixed ϵ. Observe that as long as the inequality constraint in $(\tilde{G})$ is active, (G) and $(\tilde{G})$ are equivalent in the following sense: every local solution f^ϵ of $(\tilde{G})$ is also a local solution f_α of (G) with a certain value $\alpha = \alpha(\epsilon)$. Conversely, a local solution f_α of (G) also locally solves $(\tilde{G})$ for the value $\epsilon = \epsilon(\alpha) = \|\mathcal{R}f_\alpha - p\|_2$. For short, $f^\epsilon = f_{\alpha(\epsilon)}$, and $f_\alpha = f^{\epsilon(\alpha)}$.

Now consider a discretized version of $(\tilde{G})$, where we replace the derivative f' by a finite difference approximation. Making a change of variables $g_i = f_i - f_{i+1}$, say, we recast the problem as

$$(\tilde{G}) \quad \begin{array}{ll} \text{minimize} & \|g\|_1 \\ \text{subject to} & \|\mathcal{R}Tg - p\|_2 \leq \epsilon \end{array}$$

where $f = Tg$ is that change of variables. This means that we minimize the 1-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 1-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 these extreme points, an extreme face, etc. As we can see, any one of these extreme elements has many differences g_i equal zero, which produces the mentioned flat zones. Notice, however, that this analysis shows that the choice of the 1-norm is somewhat accidental here, and that other norm balls could be used with equal rights.

How about regularizing μ? In principle we could use the same ideas as for f, even though the guideline for the cutoff frequency in (10) is no longer correct. What is observed in our experiments and confirmed in other approaches is that the resolution of the attenuation map $\mu(x)$ need not be as fine as that of the emission source $f(x)$. In particular, since the reconstruction method itself has some imperfections, the work required to improve say the ultimate 10% of resolution of $\mu(x)$ are practically wasted, as they barely improve the quality of the reconstructed image f. This suggests using a coarser bandwidth b for the signal $\mu(x)$.

Notice that we recommend using the filter (11) for μ, since the exponent in (1), known as the divergent beam transform $\mathcal{D}\mu(x,\theta)$, see [27], is close to the Radon transform $\mathcal{R}\mu$, exhibiting similar spectral properties. This means that regularizing $\mathcal{R}\mu$ also helps to stabilize $\mathcal{D}\mu(x,\theta)$. Looking at formula (1), it is clear that even though we try to estimate μ through our procedure, what is required to reconstruct f is not μ, but $\exp\{-\mathcal{D}\mu(x,\theta)\}$. In particular, $\exp\{-\mathcal{D}\mu(x,\theta)\}$ should have its spatial resolution (bandwidth) comparable to that of f, not μ.

VIII. Existing Methods

Our new optimization approach to inverting (1) has to be compared to some existing techniques. In particular, we implemented (cf. [16]) the ConTraSPECT method of [35, 21], which corrects for attenuation using a dummy attenuation map $\mu_{\text{ell}}(x)$ of elliptical shape with constant attenuation. This leaves a total of six degrees of freedom, the constant attenuation coefficient, and 5 geometric parameters fixing the shape and position of the ellipse. Using Helgason's consistency formula, the attenuation map is adjusted to the emission data using nonlinear least squares. As reported in [35, 21], the six variables are sometimes difficult to optimize simultaneously, and the best results are obtained by fixing the attenuation coefficient after some initial steps, and optimizing the 5 shape parameters subsequently. As mentioned in the introduction, since Helgason's formula is only valid over 360^0, we can only compare our method to ConTraSPECT in this case. Notice that the often surprisingly good results of ConTraSPECT are understood from our previous comments, making the point that μ is not required at a very high precision in order to improve the quality of the reconstructed image f.

A somewhat older approach, useful for instance in brain imaging, but known to fail in more complicated situations like a cardiac study, consists in automatically detecting the contour, and assuming a constant attenuation coefficient thereon. In our experimental brain study, we have estimated the head contour using emission data acquired at a secondary energy peak, representing scattered photons from the primary photo peak.

The ConTraSPECT and the contour method have presently been used to provide good starting points for the various optimizers (see the next section). A detailed comparison of these two methods as attenuation correction strategies of their own right is presented in [16].

We conclude this section by mentioning another inverse approach to (1), recently proposed by Novikov [30]; see also [29]. The author presents a mathematically appealing inversion formula for the attenuated Radon transform, (1), a curiosity, since this formula has been sought for like the holy grail since at least 1915, and many a valiant researcher ceased to believe in its existence. Structurally, it is of the form $f = \mathcal{N}[\mu]d$, if $d = \mathcal{R}[\mu]f$, that is, it inverts $f \to \mathcal{R}[\mu]f$, but not $(\mu, f) \to \mathcal{R}[\mu]f$. It could therefore be employed in a physical attenuation correction approach, but less straightforwardly in the analytical setting. However, the formula is unlikely to improve on current SPECT reconstruction procedures, as it does not allow to take collimator blurring into account. If implemented as proposed in [29, 20], Novikov's formula will play a role for SPECT similar to the role filtered back-projection plays for CT imaging.

IX. EXPERIMENTS

A first simulated study uses the MCAT phantom slice at the level of the heart shown in Figure 1. We assume a ^{99m}Tc-based tracer with relative concentrations of 75.0, 3.82 and 1.76 in heart, lungs and soft tissue. The attenuation coefficient in the cortical bones, trabecular bones, lungs and muscle at the nominal energy $E_0 = 140$keV of Technetium were chosen as 0.210, 0.166, 0.0427 and 0.150 cm^{-1} respectively. The emission source f and attenuation map μ were both discretized into 64×64 pixels of size 6.25×6.25mm. A SPECT camera with a perfect parallel hole collimator was assumed, and synthetic projection data using the attenuated Radon transform were calculated. The model includes tissue attenuation, but neither collimator blurring nor scatter. The data were Poisson noised in order to create a realistic signal-to-noise ratio. A total of 64 projections were scanned over 180^0, and alternatively over 360^0. The size of the camera projection bins was 6.25 mm. The total number of counts in the selected slice was of the order of 180,000.

A second experimental study uses the physical Radiology Support Device (RSD) neuro-receptor phantom shown in Figure 2. The phantom, an artifical skull enclosed within material that mimics soft tissue, ears, nose and neck, has one brain reservoir and four striatal containers. The chambers were filled with a homogeneous solution of 303kBq/ml labelled with ^{99m}Tc. The projection data were acquired with an Elscint dual head spectral SPX camera, equipped with a low energy high resolution (LEHUR) collimator. A total of 60 angular views, equally spaced over 180^0, and alternatively over 360^0, were scanned over 15s per view, and the projection data were sampled on a 128×128 grid with pixels of size 3.44×3.44mm. The data were scanned over 15s per view, and corrected for the known decay of the tracer isotope. This resulted in approximately 400,000 counts per projection. A $\pm 10\%$ energy window about the primary photo peak at 140keV was used. A second emission data set was acquired in a $\pm 3\%$ energy window about a secondary peak at 122keV.

X. RESULTS OF THE SIMULATION

In the simulated study, we have reconstructed the unknown μ and f using the following

2D Algorithm

1. Generate an initial guess (μ^0, f^0) using one among four possible procedures $I_1, \ldots, I_4$.
2. Run the optimizer (G_j) or (P_j) using one of the five possible regularizers $j \in \{0, 1, 2, h, 1+c\}$, until the stopping test (2) applies.
3. Keep the μ so obtained, and obtain $f_{\rm rec}$ by inverting $\mathcal{R}[\mu]f = d$ via the EM-algorithm.

In each optimization scenario (G) or (P) we have started the reconstruction method with four different initial points (f^0, μ^0). I_1 corresponds to choosing $f^0 = 0$, $\mu^0 = 0$. I_2 corresponds to running the Gaussian optimization (G) with a constant attenuation map on the contour, that is $\mu(x) = \mu$, with μ a variable to be optimized. I_3 chooses the ConTraSPECT reconstruction $(f_{\rm ell}, \mu_{\rm ell})$ as initial. Notice that since $\mu_{\rm ell}$ has nonzero values outside the contour, the optimization procedure (G) or (P), too, will have to allow for nonzero μ outside the contour. Finally, I_4 chooses f^0 as the EM-reconstruction with $\mu^0 = \mu_{\rm const}$ constant on the contour, and based on the best possible value of μ.

Based on programs (G) and (P), we have used the following optimization strategies (G_j): (G_0) uses (G) without regularizer, (G_1) uses (G) and the flat zone regularizer (12) for f and μ, (G_2) uses (G) with the Tychonov regularizer $\mathcal{I}[f] = \|\nabla f\|_2$ for both f and μ, while (G_h) uses a high pass filter (10) for f and μ. Finally, (G_{1+c}) we combines the flat zone regularizer with (14), a specially adapted penalty term to avoid the cross-talk phenomenon between μ and f. In the case of the Poisson program, the notation is analogous.

In Table 1 we have compared the results $f_{\rm rec}$ of various optimization based reconstructions. Using the true emission source we calculate the relative error terms:

$$e_{\rm rec} = \frac{\|f_{\rm true} - f_{\rm rec}\|_2}{\|f_{\rm true}\|_2} \tag{13}$$

The entries in Table 1 show relative error terms for the various regularizers. The first line shows the relative errors of the four possible initials.

Notice that if we reconstruct $f_{\rm em,true}$ using the true attenuation map $\mu_{\rm true}$, the relative errors are $e_{\rm em,true} = 18.1\%$ for 180^0, and $e_{\rm em,true} = 16.0\%$ for a 360^0 tour. These errors are due to the random nature of the emission data. Since the signal-to-noise ration in $f_{\rm em,true}$ is only mildly inferior to the signal-to-noise ration of the data, these errors may be considered a good indicator for the lowest possible noise level in *any* reconstruction.

	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	65.1	*	24.9	23.0	$-$	*	22.6	*
P_h	65.8	*	24.3	22.3	$-$	20.4	*	*
P_{1+c}							21.3	
	180^0	360^0	180^0	360^0	180^0	360^0	180^0	360^0

Table 1
Comparison of Optimization Methods

Notice that I_3 could only be tested on a 360^0 tour, which is indicated by the $-$ in the first sub-column of I_3. The symbol * indicates that the optimizer was not able to improve the error margin of the initial point (f^0, μ^0). The results displayed

correspond to the best choices of the penalty constants α involved in the various regularizers. Those differ between the programs (G) and (P), and also between the regularizers $j \in \{1, 2, h, 1+c\}$. As we observed, the correct choice of the penalties α may depend on the type of study (kidney, heart, brain, etc.), but once specified, is generally patient independent. This observation was already made in [22].

We observe that I_3 and I_4 provide already initial guesses with a good error margin, but generally optimizing improves over the initial values. Notice, however, a relatively strong dependence of the optimizers on the starting points. For instance, none of the (G)'s or (P)'s was able reduce the error in I_1 to a competitive value, so strategy I_1 turns out insufficient.

In the heart study it is possible to use a specially suited regularizer in order to avoid the cross-talk phenomenon between f and μ reported in several approaches. The shadow of f apparent in the reconstructed μ appears in a region where the correct value of μ is basically known. While correcting μ by hand is of course prohibitive, we recommend a regularizer of the form

$$\mathcal{I}[f, \mu] \;=\; \sum_{i=1}^{N} f_i \, (\mu_{\max} - \mu_i) \tag{14}$$

which will obviously penalize values μ_i too low at places i with high activity f_i. Notice that this is a non-convex function in (f, μ).

This approach works well, but the bad news is that it barely improves the quality of the reconstructed $f_{\rm rec}$ or the errors $e_{\rm rec}$. The good part of that bad news is that this seems to indicate that the damage of the shadow artifact is negligeable anyway, so its only effect is that the reconstructed $\mu_{\rm rec}$ is less fancy.

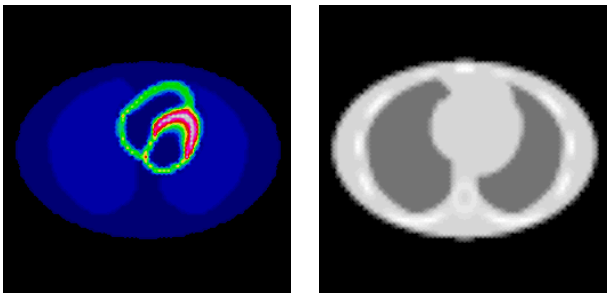

Fig. 1 Transaxial slice through the MCAT phantom at the level of the heart. Left ^{99m}Tc uptake distribution, right attenuation map at 140keV.

XI. Results of the Phantom Study

In the phantom study, the reconstructions were obtained via the following

3D Algorithm

1. Divide the 3D ROI into transaxial slices $\nu = 1, \ldots, T$. In each slice generate an initial guess using I_4, and run the 2D Algorithm to obtain a reconstruction (μ^ν, f^ν).

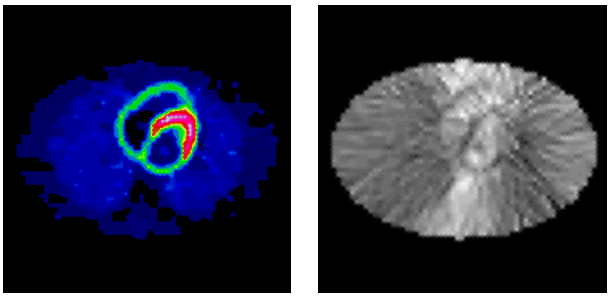

Fig. 2 Reconstruction of same transaxial MCAT slice. Activity distribution (left) is close to the truth. A shadow artifact of the heart is still visible in the reconstructed attenuation map (right).

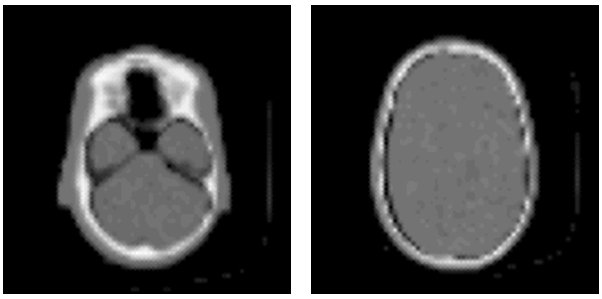

Fig. 3 X-ray CT images of the RSD phantom. Right cortical slice, left cerebellar slice.

2. Form a 3D attenuation map μ by stacking the μ^ν, $\nu = 1, \ldots, T$.

3. Obtain the emission source $f_{\rm rec}$ by a 3D inversion of $\mathcal{R}_b[\mu] f \;=\; d$ via the EM-algorithm, where the model includes attenuation and collimator blurring.

In order to estimate the head contour required in I_4, we have reconstructed the data acquired about the secondary energy peak at 122keV .

Notice that the slice-by-slice estimation of μ in step 1 is necessary, since a 3D inversion would lead to a difficult large scale optimization problem with $2 \cdot 64^3$ unknown variables.

Since the activities in the different containers of the phantom are known, the true emission source is known up to a constant factor. However, the proportionality constant is difficult to estimate in practice, and we have therefore decided to use a different strategy to evaluate the reconstructions $f_{\rm rec}$.

XII. Conclusion

Our experiments have shown that attenuation correction using SPECT emission data only is possible. In the simulated study, some of the reconstructions $f_{\rm rec}$ came close to the error margin already present in the random data. In that situation, the result is close to optimal, and it seems hard to improve e.g. by constructing more sophisticated regularizers. Put differently, the preponderant fraction in the error comes from the reconstruction method itself.

In the simulation we observed that the flat zone regularizer performed slightly better than the bandpass filtering, probably

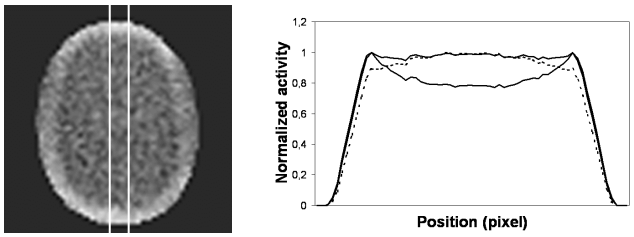

Fig. 4 Reconstructed attenuation map of same cortical slice with profile (left). Activity curves along profile (right) show no attenuation correction (lower curve), initial guess I_4 (broken line) result of optimizer (upper continuous line). The true activity curve is known to be flat.

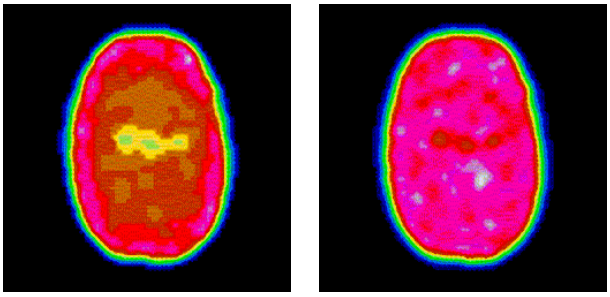

Fig. 5 Same reconstructed cortical slice shown without attenuation correction (left), and with attenuation correction (right). The true activity distribution is known to be homogeneous. Both reconstructions use the correct collimator response.

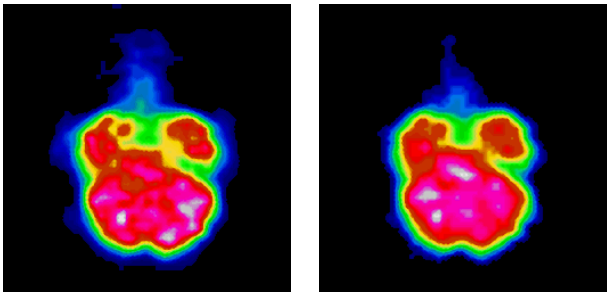

Fig. 6 Same reconstructed cerebellar slice shown without attenuation correction (left), and with attenuation correction (right). Both reconstructions use the correct collimator response.

due to the fact that the ideal source f_{true} is piecewise constant, with edges sharper than in realistic situations. An additional complication of the high pass filter is that two parameters b and α had to be adapted, but the guidelines we indicated worked well.

The experimental study shows that our method works well in a realistic situation. The proposed attenuation correction showed a significant improvement in the cortical slices, and a mild one in the cerebellar region.

XIII. References

[1] Bauschke H, Noll D, Celler A, Borwein JM, An EM-algorithm for dynamic emission tomography, *IEEE Trans. Med. Imaging* 18, 1999, no. 3, 252 – 261.

[2] Blondel C, Noll D, Maeght J, Celler A, Farncombe T, Comparison of different figure of merit functions for dynamic single photon emission computed tomography (dSPECT), *IEEE Trans. Nucl. Sci*, 2000 TNS-MIC Conference Record, Lyon.

[3] Bronnikov AV, Numerical solution of the identification problem for the attenuated Radon transform, *Inverse Problems* 15, 1999, 1315 – 1324.

[4] Bronnikov AV, Approximate reconstruction of attenuation map in SPECT imaging, *IEEE Trans. Nucl. Sci.* 42, 1995, 1483 – 1488.

[5] Bronnikov AV, Reconstruction of attenuation map using discrete consistency conditions, *IEEE Trans. Med. Imag.*, vol. 19, no. 5, pp. 451- 462, 2000.

[6] Buvat I, Rodriguez-Villafuerte M, Todd-Pokropek A, Benali H, DiPaola R, Comparative assessment of nine scatter correction methods based on spectral analysis using Monte Carlo simulations, *J. Nucl. Med.* 36, 1995, 1476 – 1488.

[7] Byrne CL, Accelerating the EMML algorithm and related iterative algorithms by rescaled block-iterative methods, *IEEE Trans. Img. Process.* 7, no. !, 1998, 100 – 109.

[8] Carson RE, Lange K, The EM parametric image reconstruction algorithm, *J. Amer. Stat. Assoc.* 80, 1985, 20 – 22.

[9] Celler A, Sitek A, Harrop R, Reconstruction of multiple line source attenuation maps. *IEEE Nuclear Science Symposium Conference Record*, pp. 1420 – 1424, November 1996.

[10] Celler A, Sitek A, Stoub E, Lyster D, Dykstra C, Worsley D, Fung A, Development of a multiple line source attenuation array for SPECT transmission scans. *J. Nuclear Medicine* 38, 1997, 215ff.

[11] Celler A, Axen D, Togane D, El-Khatib J, Investigation of scatter in SPECT transmission studies, *IEEE Trans. Nucl. Sci.* 47, 2000, 1251 – 1256.

[12] Censor Y, Gustafson DE, Lent A, Tuy H, A new approach on the emission computerized tomography problem: simultaneous calculation of attenuation and activity coefficients. *IEEE Trans. Nucl. Sci.* NS26(2), 1979, 2775 – 2779.

[13] Dicken V, Simultaneous activity and attenuation reconstruction in emission tomography, *Inverse Problems*, 15, 1999, 931 – 960.

[14] Dicken V, Simultaneous activity and attenuation reconstruction in single photon emission computed tomography, a nonlinear ill-posed problem. *PhD Thesis, Universität Potsdam*, 1998.

[15] Golub GH, Pereira V, The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate, *SIAM J. Numer. Anal.*, 10, 1973, 413 – 432.

[16] Gourion D, Hatchondo X, Gantet P, Noll D, Esquerré JP, Comparison of two methods for SPECT attenuation correction without transmission measurements, *IEEE-MIC Conference Record* 2001, San Diego, CA, USA

[17] Higdon DM, Bowsher JE, Johnson VE, Turkington TG, Gilland DR, Jaszczak RJ: Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data, *IEEE TMI*, 16, 1997, 516 – 526.

[18] Hudson H, Larkin R, Accelerated image reconstruction using ordered subsets of projection data, *IEEE Trans. Med. Imaging* 13, 1994, 601 – 609.

[19] Krol A, Manglos SH, Bowsher JE, Bassano DA, Thomas FD, An EM algorithm for estimating SPECT emission and transmission parameters from emission data only. Technical Report, SUNY Health Science Center and Duke Medical Center, Dept. of Radiology, Syracuse, NY 13210, 1996.

[20] Kunyansky L, A new SPECT reconstruction algorithm based on the Novikov's explicit inversion formula. Preprint.

[21] Laurette I, Clackdoyle R, Welch A, Natterer F, Gullberg GT, Comparison of three applications of ConTraSPECT. *IEEE NSS and MIC Conference Record*, Toronto, 1998.

[22] Maeght, J, Analyse et méthodes pour un problème inverse en

tomographie dynamique. *Thèse de Doctorat de l'Université Paul Sabatier, Toulouse, France* 1999.

[23] Maeght J, Noll D, Resolution in dynamic emission tomography, *SIAM J. Math. Analysis 31, no. 5, 2000, pp. 1100 – 1120.*

[24] Maréchal P, Togane D, Celler A, Borwein JM, Numerical assessment of the stability of reconstruction processes for computed tomography.

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

[26] Natterer F, Computerized tomography with unknown sources, *SIAM J. Appl. Math.* 43, 1983, 1201 – 1212.

[27] Natterer F, The Mathematics of Computerized Tomography, Teubner Verlag, Stuttgart, 1986.

[28] Natterer F, Determination of tissue attenuation in emission tomography of optically dense media, *Inverse Problems* 9, 1993, 731 – 736.

[29] Natterer F, Inversion of the attenuated Radon transform. Preprint.

[30] Novikov RG, An inversion formula for the attenuated X-ray transform. Preprint.

[31] Panin VY, Zeng GL, Gullberg GT, A method of attenuation and emission activity reconstruction from emission data, *IEEE TNS*, vol. 48, no. 1, 2001, 131 – 138.

[32] Rattey PA, Lindgren AG, Sampling the 2-D Radon transform, *IEEE Trans. Acoustics, Speech, Signal Proc.* ASSP-29, 1981, 994 – 1002.

[33] Setzepfand B, ESNM: Ein rauschunterdrückendes EM-Verfahren für die Emissionstomographie, *Dissertation, Universität Münster*, 1992.

[34] Vardi Y, Shepp LA, Kaufman L, A statistical model for Positron Emission Tomography, *J. Amer. Stat. Assoc.* 80, 1985, 8 – 20.

[35] Welch A, Clark R, Natterer F, Gullberg GT, Toward accurate attenuation correction in SPECT without transmission measurements, *IEEE Trans. Med. Imag.*, 16(5), 1997, 532 – 541.

[36] Wells RG, Celler A, Harrop R, Analytic calculation of photon distribution in SPECT projections, *IEEE Trans. Nuclear Sci.* 45, no. 6, 1998, 3202 – 3214.

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

[38] Zhao S, Welland G, Wang G, Wavelet sampling and localization schemes for the Radon transform in two dimensions, *SIAM J. Appl. Math.* 57, no. 6, 1997, 1749 – 1762.