# Attenuation Correction Using SPECT Emission Data Only

Daniel Gourion, Dominikus Noll, Pierre Gantet, Anna Celler, *Member, IEEE*, and Jean-Paul Esquerré, *Member, IEEE*

*Abstract*—A major step toward quantitative single photon emission computerized tomography (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 may 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 and test it in a simulated case study and a physical phantom experiment.

*Index Terms*—Attenuation correction, nonlinear optimization, radon transforms, regularization, single photon emission computerized tomography (SPECT).

## I. INTRODUCTION

I N ITS original understanding, the term scatter or attenuation correction referred to methods which tried to improve single photon emission computerized tomography (SPECT) or positron emission tomography (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 [1] 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 [2] for a discussion and references on at least five different source configurations.

Here we will be mainly concerned with a third class 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 an analytical attenuation-correction method using nonlinear optimization techniques. Two programs $(P)$ and $(G)$, based on a Poisson and a Gaussian likelihood function are compared and their viability is substantiated using a simulated case study and a phantom experiment.

## II. PHYSICAL ATTENUATION CORRECTION

Transmission-based 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 coexist 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 downscatter into the energy window of the transmission isotope, generating crosstalk between the two procedures (see [2]). This leads to artifacts in the reconstructed images. In [2], the authors suggest that if the transmission source used the higher energy isotope than the SPECT tracer, the impact of the crosstalk 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 system. However, this will complicate the protocol, and may lead to the nontrivial problem of coregistration of two images acquired with different geometries. In addition, if X-ray computer tomography (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.

The first class of methods, pioneered by Natterer [3], uses the Helgason consistency formula (see, e.g., [4, Theorem II.6.2]) to estimate the unknown attenuation map $\mu(x)$ prior to reconstructing the emission source $f(x)$. This idea has recently been revived in [5] and [6], see also [7], 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° camera rotation, since Helgason's formula has no substitute for different rotation angles. Notice that the artificial attenuation maps obtained by this type of methods will partly correct for some of the scatter effects. On the other hand, a rigorous scatter correction where the forward and backward projection operators include a Compton scatter model could hardly be based on the elliptical attenuating medium.

A second type of mathematical methods, also initiated by Natterer [8], tries to fit a template or reference attenuation map, along with a prespecified deformation procedure, to the individual case, using either the consistency formula, or by estimating $\mu$ 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

$$R[\mu]f(s, \theta) = \int_{-\infty}^{\infty} f(s\theta+t\theta^{\perp})\exp\left(-\int_{t}^{\infty} \mu(s\theta+\tau\theta^{\perp})d\tau\right)dt$$
$$= p(s, \theta) \qquad (1)$$

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 $(s, \theta)$. Equation (1) is used in [9], where the completely discretized version of the problem is studied and an iterative algorithm solving for $f$ and $\mu$ simultaneously is proposed. It is based on cyclic subgradient projections for convex–concave nonlinear systems of equations. A more recent approach is presented by Dicken [10], [11], who uses a Tychonov type regularization to invert

(1). This requires solving an optimization problem $(G)$ of the form

$$(G) \qquad \min_{f, \mu} \|R[\mu]f - p\|_2^2 + \alpha I[f, \mu] \qquad (2)$$

featuring an appropriate regularization term $\alpha I[f, \mu]$ which penalizes and thereby avoids highly irregular distributions $(f, \mu)$ that would match the data within the acceptable error tolerance (see Section IV-C). In $(G)$ we minimize the negative log likelihood of a Gaussian law, and possible choices of the norm $\|.\|$ will be discussed in Section IV-A. The regularizing term $I[f, \mu]$ will then play the role of a Bayesian prior, and possible choices of these regularizers are discussed in Section IV-D, while steering the penalty parameter $\alpha$ is discussed in Section IV-C.

An interesting way to solve $(G)$ was recently proposed by Bronnikov [12]–[14]. Exploiting the fact that $R[\mu]f$ is linear in $f$, the author first solves the inner linear least squares problem in $(G)$ with respect to the variable $f$, using the pseudoinverse $R[\mu]^+$. The remaining nonlinear least squares problem

$$\min_{\mu} \left\|R[\mu]R[\mu]^+p - p\right\|_2^2 + \alpha I[\mu] \qquad (3)$$

in the unknown $\mu$ 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 crosstalk between the reconstructions of $f$ and $\mu$ observed, e.g., by Dicken. Yet another approach, based on a direct inversion of (1), is that of Panin *et al.* [16], where the authors use a singular value decomposition to partially linearize the nonlinear dependence of $R[\mu]f$ on $\mu$.

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

$$(P) \quad \min_{f, \mu} \sum_{j=1}^{M}\sum_{k=1}^{S}\left\{\sum_{i=1}^{N} R_{ijk}[\mu]f_i - p_{jk} \log\left(\sum_{i=1}^{N} R_{ijk}[\mu]f_i\right)\right\}$$
$$+\alpha I[f, \mu] \quad (4)$$

which minimizes the negative Poisson log-likelihood function of the independent Poisson distributed random vector $p = (p_{jk})$ with

$$E(p_{jk}) = \sum_{i=1}^{N} R_{ijk}[\mu]f_i$$

augmented by a regularizing term $\alpha.I[f, \mu]$ as above.

Whenever required, we will 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 $i$th pixel, $\mu_i$ 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 $k$. Accordingly, $p = (p_{jk})$ represents the projection data, with $p = (p_{jk})$ 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., [17], and we may even choose different bases for $\mu$ and $f$ in order to account for their different spatial resolutions.

## IV. METHODS

In this paper, we solve the problems $(G)$ and $(P)$ directly using nonlinear optimization methods. This requires optimization software for nonlinear optimization with bound constraints. In our experiment, we have used the code l-bfgs-b described in [18]. In this section, we discuss some practical aspects of the different optimization strategies based on the schemes $(G)$ and $(P)$. This includes in particular the suitable choice of the regularizer $I[f, \mu]$, which is crucial for a good performance of the method.

### A. Nonlinear Least Squares

An important aspect of the nonlinear least squares approach $(G)$ is the correct choice of the norm $\|e\|$ of the forward error $e = R[\mu]f - p$ in data space. There is evidence that the Euclidean norm may *not* be an appropriate candidate. This point was already made in [4], [10], [11], while [13], [14] still uses the Euclidean norm. Here we will give some support for the choices suggested in [4], [10], [11] by comparing the approaches $(G)$ and $(P)$.

Notice that [4] shows that for fixed $\mu$, the linear operator $f \to R[\mu]f$ is continuous between the spaces $L^2(D)$ on the unit disk $D$ and $L^2(Z, w)$ on the cylinder $Z = [-1, 1] \, x \, D$ if the weighted measure $w(s)dsd\theta$, 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. The probabilistic model $(P)$ using the Poisson statistics supports this argument.

Let $l(w, p) = w - p \log w$, so that the negative Poisson log-likelihood function in $(P)$ may be written as

$$\sum_{j=1}^{M} \sum_{k=1}^{S} l(w_{jk}, p_{jk}) \quad \text{with} \quad w_{jk} = \sum_{i=1}^{N} R_{ijk}[\mu]f_i.$$

Taylor expansion gives the well-known estimate

$$l(w, p) - l(p, p) \approx \frac{1}{2} \frac{(w-p)^2}{p} \tag{5}$$

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 $\sum_{jk} l(p_{jk}, p_{jk})$, be close to the weighted norm expression

$$\frac{1}{2} \|w - p\|_p^2 = \frac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{S} \frac{(w_{jk} - p_{jk})^2}{p_{jk}} \tag{6}$$

which is then the norm we should use if we prefer a Gaussian model $(G)$. The norm (6) coincides with the norm on $L^2(Z, w)$ used above if the source $f$ is a constant function on the unit disk.

### B. Poisson Model

As in the case of $(G)$, we solve the Poisson model $(P)$ via nonlinear programming techniques. Due to the logarithmic term, this requires a hard positivity constraint $f_i > 0$. This excludes infeasible optimizers like some sequential quadratic programming codes, but does not pose any problem with most bound constrained solvers.

Notice that the standard expectation–maximization (EM)-algorithm [19], [20] and its modifications such as ordered subset expectation–maximization (OSEM) or rescaled block-iterative (RBI) [22] are based on the model $(P)$ with known attenuation map and therefore cannot be used to compute $f$ and $\mu$ simultaneously. An extension of the EM-algorithm to the case of unknown $\mu$ is given in [23]. A drawback of these EM-type methods is that regularizing terms $I[f, \mu]$ are difficult to include in the iterative procedure. This leads to the well-known observation that the iterates may deteriorate if the procedure is carried too far. Some regularization is obtained by stopping the EM-algorithm after a limited number of iterations. Another way in which regularization could still be included in an EM-algorithm, at the cost of a considerable slow down of the algorithm, is presented in [24]. In the present work, we shall concentrate on the optimization-based approach.

### C. Stopping and Scaling

An important practical aspect of both approaches, $(G)$ and $(P)$, concerns appropriate stopping rules, and suitable scaling of the variables $f$ and $\mu$.

Most optimization codes have built-in stopping tests, which dispense with the iteration as soon as either the necessary optimality conditions are satisfied, or no sizable progress is made. However, large-scale applications may greatly benefit from additional user-provided stopping tests, which will typically intervene before the internal tests. Here we propose to halt the optimization procedure as soon as the $l_2$-norm approaches the overall error in the data

$$\|R[\mu]f - p\|_2^2 \approx \sum_{jk} p_{jk}. \tag{7}$$

The rationale here is that with Poisson data $p$, iterates $(f, \mu)$ below the noise level (7) should not be allowed, as they may exhibit the same noise amplification phenomenon as observed, e.g., in the EM-algorithm. This test was already reported to work well in a different context [25]. Our present experiments confirm this guideline.

Notice that the choice of the penalty parameter $\alpha$ is closely related to this stopping test. If $\alpha$ is chosen too large, we may be unable to achieve the desired error margin (7). On the other hand, choosing $\alpha$ too small will give many candidates $(f, \mu)$ which match (7). In that case, the effect of our regularizer is too weak.

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

## D. Regularizers

The most important aspect of both models $(G)$ and $(P)$ is the choice of the regularizer $I[f, \mu]$, as it may greatly influence the performance of the method. This is where our approach differs most from previous work obtained with these models.

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 [26]. In the present section, we discuss possible choices of regularizers $I[f]$ and $I[\mu]$ adapted to our problem.

Using a high-pass filter

$$I[f] = \|H_b f\|_2^2 = \left\|\phi_b.\hat{f}\right\|_2^2 \tag{8}$$

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 in [27], [28], the Fourier slice theorem could give us a guideline on the choice of $b$. Observe that without tissue attenuation, $\widehat{Rf}(\sigma, \theta) = \hat{f}(\sigma\theta)$, 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 (8).

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, (8) works considerably well in practice (see also [25]).

An interesting variation of (8) uses the fact that the two–dimensional (2-D) spectrum of the attenuated Radon transform $p = R[\mu]f$ is concentrated on a bowtie-shaped region in the frequency plane [4], [29], [30]. This was first observed in the unattenuated case, but [29] shows that it remains qualitatively correct in the attenuated case. This suggests a regularizer of the form

$$I[f] = \|H_{m,b}(Rf)\|_2^2 = \left\|\phi_{m,b}.\widehat{Rf}\right\|_2^2 \tag{9}$$

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 $\theta$. For details see the above references. Notice that in both formulas (8) and (9) 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

$$I[f] = \|\nabla f\|_1 \tag{10}$$

which modifies the notorious Tychonov term, known to be too smoothing, replacing the Euclidean norm by the $l_1$-norm. This is reported to favor reconstructed images $f$ featuring flat zones with identical gray values. Our experiments confirm this phenomenon in the context of SPECT (see Section VII). The flat-zone regularizer may be justified by the following argument.

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

$$(\tilde{G}) \qquad \text{minimize } \|f'\|_1$$

$$\text{subject to } \|Rf - p\|_2 \leq \varepsilon$$

for a fixed $\varepsilon$. 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 $f^\varepsilon$ of $(\tilde{G})$ is also a local solution $f_\alpha$ of $(G)$ with a certain value $\alpha = \alpha(\varepsilon)$. Conversely, a local solution $f_\alpha$ of $(G)$ also locally solves $(\tilde{G})$ for the value $\varepsilon = \varepsilon(\alpha) = \|Rf - p\|_2$. For short, $f^\varepsilon = f_{\alpha(\varepsilon)}$ and $f_\alpha = f^{\varepsilon(\alpha)}$.

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

$$(\tilde{G}) \qquad \text{minimize } \|g\|_1$$

$$\text{subject to } \|RTg - p\|_2 \leq \varepsilon$$

where $f = Tg$ is that change of variables. This means that we minimize the $l_1$-norm of $f = Tg$ 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 $l_1$-norm ball has $2n$ extreme points, $n$ being 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 $l_1$-norm is somewhat accidental here, and that other norm balls could be used with equal rights.

How about regularizing $\mu$? In principle, we could use the same ideas as for $f$, even though the guideline for the cutoff frequency in (8) 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 $\mu(x)$ is practically wasted, as it barely improves 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 (9) for $\mu$, since the exponent in (1), known as the divergent beam transform $D\mu(x, \theta)$, see [4], is close to the Radon transform $R\mu$, exhibiting similar spectral properties. This means that regularizing $R\mu$ also helps to stabilize $D\mu(x, \theta)$. Looking at formula (1), it is clear that even though we try to estimate $\mu$ through our procedure, what is required to reconstruct $f$ is not $\mu$ but $\exp\{-D\mu(x, \theta)\}$. In particular, $\exp\{-D\mu(x, \theta)\}$ should have its spatial resolution (bandwidth) comparable to that of $f$, not $\mu$.

## V. EXISTING METHODS

Our new optimization approach to inverting (1) has to be compared to some existing techniques. In particular, we implemented (cf. [7]) the ConTraSPECT method of [5], [6], which

corrects for attenuation using a dummy attenuation map $\mu_{ell}(x)$ of elliptical shape with constant attenuation. This leaves a total of six degrees of freedom, the constant attenuation coefficient, and five 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 [5], [6], the six variables are sometimes difficult to optimize simultaneously, and the best results are obtained by fixing the attenuation coefficient, and optimizing the five shape parameters subsequently. As mentioned in Section III, since Helgason's formula is only valid over $360°$, we can only compare our method to ConTraSPECT in this case. Notice that the often surprisingly good results of ConTraSPECT are understood from the observation that knowledge of $\mu$ is not required to 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, 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 recently been used to provide good starting points for the various optimizers (see Section VII). A detailed comparison of these two methods as attenuation-correction strategies in their own right is presented in [7].

We conclude this section by mentioning another inverse approach to (1), recently proposed by Novikov [31] (see also [32]). The author presents a mathematically appealing inversion formula for the attenuated Radon transform, (1), a curiosity, since this formula has been searched 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 $f = N[\mu]p$, if $p = R[\mu]f$, that is, it inverts $f \to R[\mu]f$, but not $(f, \mu) \to R[\mu]f$. It could therefore be employed in a physical attenuation-correction approach, but less straightforwardly in the analytical setting. However, if implemented as proposed in [32], [33], the formula is unlikely to improve on current SPECT reconstruction procedures, as it does not allow to take collimator blurring into account.

## VI. EXPERIMENTS

### A. Simulated Data

The simulated study uses a slice of the mathematical cardiac torso (MCAT) phantom at the level of the heart (Fig. 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 of Technetium (140 keV) were chosen as 0.210, 0.166, 0.0427, and 0.150 cm$^{-1}$, respectively. The emission source $f$ and attenuation map $\mu$ were both discretized into $64 \times 64$ pixels of size 6.25 mm $\times$ 6.25 mm. Parallel projections were simulated including the effects of photon attenuation. Scatter and collimator blurring were not simulated. The data were Poisson-noised in order to create a realistic signal-to-noise ratio. A total of 64 projections were
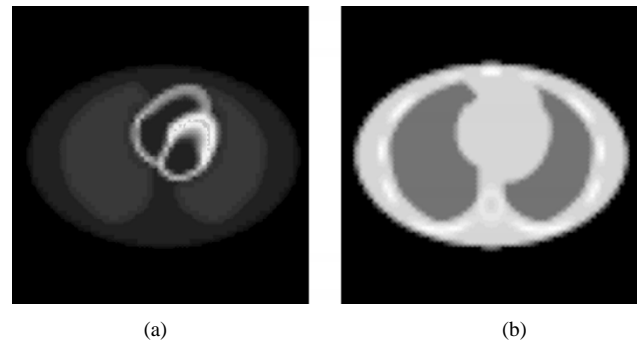


Fig. 1. A transaxial slice of the MCAT phantom. (a) Activity distribution. (b) Attenuation coefficient distribution.
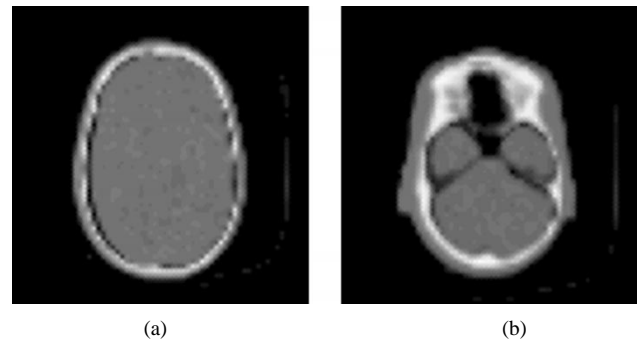


Fig. 2. X-ray CT images of the RSD phantom. (a) A cortical slice. (b) A cerebellar slice.

scanned over $180°$, and alternatively over $360°$. 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.

### B. Experimentally Acquired Data

The experimental study uses the physical Radiology Support Device (RSD) striatal phantom shown in Fig. 2. The phantom, an artificial 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 303 kBq/ml labeled with $^{99m}$Tc. The projection data were acquired using an Elscint dual head SPX gamma-camera, equipped with parallel low-energy high-resolution collimators. A total of 60 angular views, equally spaced over $360°$, were scanned over 15 s per view, and the projection data were sampled on a $128 \times 128$ grid with pixels of size 3.44 mm $\times$ 3.44 mm. The data were corrected for the decay of the tracer isotope. This resulted in approximately 400 000 counts per projection. A $\pm 10\%$ energy window about the primary photopeak at 140 keV was used. A second emission data set was acquired in a $\pm 3\%$ energy window about a secondary peak at 122 keV.

## VII. RESULTS

### A. Simulated Data

In the simulated study, we have reconstructed the unknown $f$ and $\mu$ using the following 2-D algorithm:

**2-D Algorithm**

TABLE I
COMPARISON OF OPTIMIZATION METHODS

| | I1 | | I2 | | I3 | | I4 | |
|---|---|---|---|---|---|---|---|---|
| | 180° | 360° | 180° | 360° | 180° | 360° | 180° | 360° |
| -- | 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 | * |



(a)                                         (b)

Fig. 3. Reconstruction of the MCAT phantom with $(G_1)$, $I2$, 360°. (a) Activity distribution. (b) Attenuation coefficient distribution.

1) Generate an initial guess $(f_0, \mu_0)$, using one among four possible procedures $I1, \ldots, I4$.

2) Run the optimizer l-bfgs-b to solve $G_j$ or $P_j$ using one of the five possible regularizers $j = 0, 1, 2, h, 1+c$ until the stopping test (7) applies.

The results of our experiment are presented in Table I, which is to be read as follows. In each optimization scenario $(G)$ or $(P)$ we have started the reconstruction method with four different initial points $(f_0, \mu_0)$. $I1$ corresponds to choosing $f_0 = 0$, $\mu_0 = 0$. $I2$ corresponds to running the Gaussian optimization $(G)$ with a constant attenuation map on the contour, that is, $\mu(x) = \mu$ with $\mu$ a variable to be optimized. $I3$ chooses the ConTraSPECT reconstruction $(f_{ell}, \mu_{ell})$ as initial. Notice that since $\mu_{ell}$ has nonzero values outside the contour, the optimization procedure $(G)$ or $(P)$, too, will have to allow for nonzero values outside the contour. Finally, $I4$ chooses $f_0$ as the EM-reconstruction with $\mu_0 = \mu_{const}$ constant on the contour, and based on the best possible value of $\mu$.

Based on programs $(G)$ and $(P)$, we have used the optimization strategies $(G_j)$: $(G_0)$ uses $(G)$ without regularizer, $(G_1)$ uses the flat-zone regularizer (10) for $f$ and $\mu$, $(G_2)$ uses the Tychonov regularizer $I[f] = \|\nabla f\|_2^2$ for both $f$ and $\mu$, while $(G_h)$ uses a high-pass filter (8) for $f$ and $\mu$. Finally, $(G_{1+c})$ combines the flat-zone regularizer for $f$ with a specially adapted penalty term (12) to avoid the crosstalk phenomenon between $f$ and $\mu$. In the case of the Poisson program, the notation is analogous.

The numbers in Table I represent relative error terms

$$e_{rec} = \|f_{\text{true}} - f_{\text{rec}}\|_2 / \|f_{\text{true}}\|_2 \tag{11}$$

comparing the reconstructions $f_{\text{rec}}$ obtained via the various programs $(G_j)$, $(P_j)$ with the true source $f_{\text{true}}$. The first line shows the relative errors of the four possible initials $I\nu$.

Notice that if we reconstruct $f_{em, \text{true}}$ using the true attenuation map $\mu_{\text{true}}$ and an ML-EM algorithm, the relative errors are 18.1% for 180°, and 16.0% for a 360° tour. These errors are due to the random nature of the emission data. Since the signal-to-noise ratio in $f_{em, \text{true}}$ is only mildly inferior to the signal-to-noise ratio of the data, these errors may be considered close to the lowest possible noise level in *any* reconstruction.

Notice that $I3$ could only be tested on a 360° tour, which is indicated by the $-$ in the first subcolumn of $I3$. The symbol $*$ for regularizers $j = 1, 2, h$ indicates that the optimizer was not able to improve the error margin of the reconstruction without regularizer (i.e., $j = 0$). For regularizer $j = 1+c$, this symbol
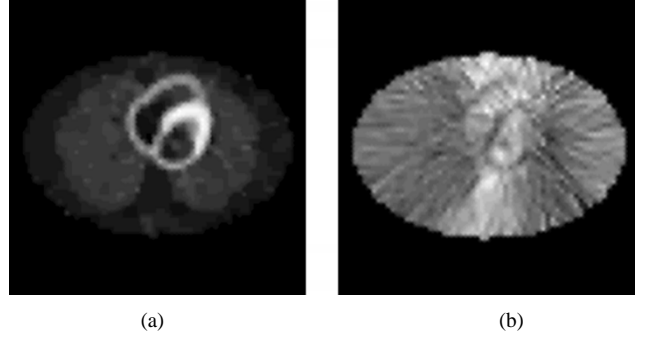
means that there is no improvement over the error margin of the reconstruction with $j = 1$. The displayed results correspond to the best choices of penalty constants $\alpha$ involved in the various regularizers. Those differ between the programs $(G)$ and $(P)$, and also between the regularizers $j = 1, 2, h, 1+c$.

We observe that $I3$ and $I4$ 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 to reduce the error in $I1$ to a competitive value, so strategy $I1$ turns out insufficient.

Concerning regularization, the flat zone regularizer performed better than either the high-pass filtering or the Tychonov regularizer, probably due to the fact that the ideal source $f_{true}$ is piecewise constant, with edges sharper than in realistic situations. The results with this regularizer and initial points $I2$ or $I3$ came even close to the "minimum" error margin obtained with the true attenuation map.

In the heart study, it is possible to use a specially suited regularizer in order to avoid or at least reduce the crosstalk phenomenon between $f$ and $\mu$ reported in several approaches. The shadow of $f$ apparent in the reconstructed $\mu$ appears in a region where the correct value of $\mu$ is basically known. While correcting $\mu$ by hand is, of course, prohibitive, we recommend a regularizer of the form

$$I[f, \mu] = \sum_{i=1}^{N} f_i(\mu_{\max} - \mu_i) \tag{12}$$

which will penalize values $\mu_i$ too low at places $i$ with high activity $f_i$. Notice that this is a nonconvex function in $(f, \mu)$.

This approach works well, but barely improves the quality of the reconstructed $f_{\text{rec}}$ or the error $e_{\text{rec}}$. This seems to indicate that the damage of the shadow artifact is negligible anyway, so its only effect is that the reconstructed $\mu_{\text{rec}}$ is less fancy.

The reconstructed attenuation map in Fig. 3(b) should be compared to results obtained by Dicken [10], [11] and Bronnikov [12]–[14]. As the difference between the true [Fig. 1(b)] and reconstructed $\mu$ [Fig. 3(b)] seems significant, some explanation is in order. We argue that in the given case, the reconstruction in Fig. 3(b) is rather what we should expect. A reconstruction $\mu_{\text{rec}}$ resembling to the true $\mu_{\text{true}}$ could even be undesirable, since the optimization approach $(P)$, $(G)$ requires $\exp\{-D\mu\}$, and not $\mu$, to be accurately known. In particular, it tries to adjust $\exp\{-D\mu(x, \theta)\}$ at positions $x$ where the source $f(x)$ is sizable. Fig. 4 shows that $D\mu_{\text{rec}}(x, \theta)$
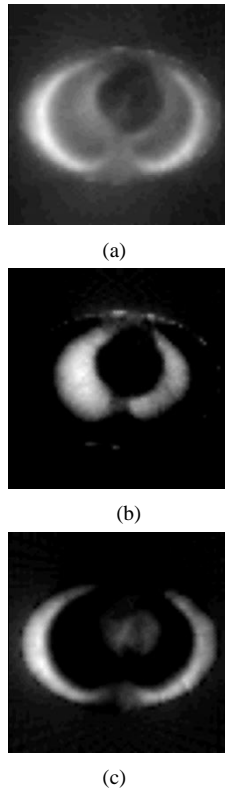
(a)

(b)

(c)

Fig. 4. (a) Map of the total error $\sum_{k=1,\ldots,S} |\exp\{-D[\mu_{\mathrm{rec}}](i,\,k)\} - \exp\{-D[\mu_{\mathrm{true}}](i,\,k)\}|$ with $(G_1)$, $I_2$, $360°$. (b) Negative part. (c) Positive part.

is very close to $D\mu_{\mathrm{true}}(x,\,\theta)$ at positions $x$ where the activity distribution is high, which means that the attenuation correction provided by our method is accurate in the heart.

Since, therefore, $D\mu(x,\,\theta)$ is only remotely reliable at positions $x$ where $f(x)$ is weak, we should not expect $\mu_{\mathrm{rec}}$ to be of good quality, as the passage from $D\mu$ to $\mu$ is by itself an ill-posed procedure. This argument is the 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 $\mu$ with a highly localized $f(x)$ is, of course, highlighted in the case of a point source $f = \delta_x$, where $\mu$ is not even theoretically uniquely determined [4]. This contradiction could be resolved if the operator $R[\mu]f$ included scatter effects and the optimization models $(G)$ and $(P)$ were solved in three dimensions.

### B. Experimentally Acquired Data

In the phantom study, the reconstructions were obtained via the following three–dimensional (3-D) algorithm:

### 3-D Algorithm

1) Divide the 3-D region of interest (ROI) into transaxial slices $\nu = 1,\,\ldots,\,T$. In each slice generate an initial guess using $I3$ or $I4$, and run the 2-D Algorithm $(G_0)$ or $(G_1)$ to obtain a reconstruction $(f_\nu,\,\mu_\nu)$.

2) Form a 3-D attenuation map $\mu$ by stacking the $\mu_\nu$, $\nu = 1,\,\ldots,\,T$.

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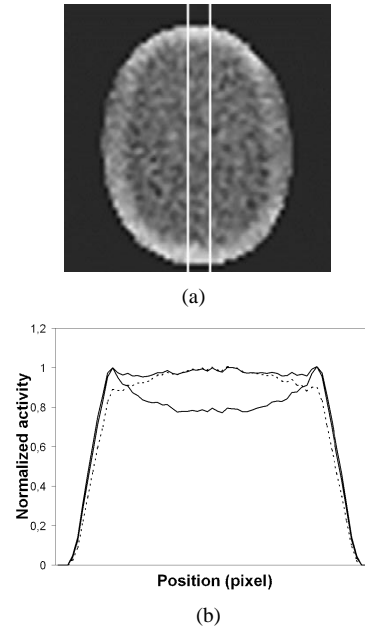


(a)

(b)

Fig. 5. Evaluation of phantom study in a cortical slice. (a) Attenuation map reconstructed by the optimizer and location of the 4-pixel-thick profile used in (b). (b) Reconstructed activity profiles; lower curve: without correction; dotted curve: initial $I4$; upper curve: result of optimizer $I4$, $(G_0)$.



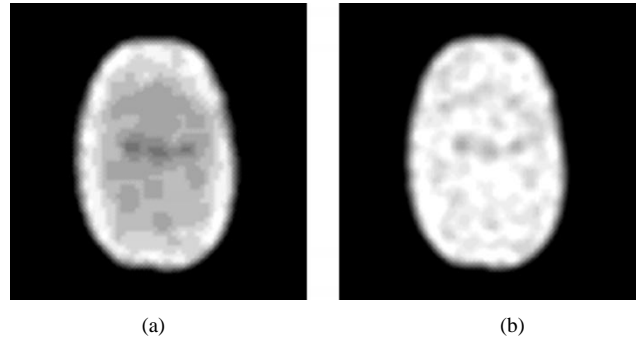(a)                                        (b)

Fig. 6. Reconstructions of the same cortical slice. (a) Without any correction. (b) With collimator and attenuation correction.

In order to estimate the head contour required in $I4$, we have reconstructed the data acquired about the secondary energy peak at 122 keV. This leads to a slightly enlarged contour (see Fig. 5).

Notice that the slice-by-slice estimation of $\mu$ in step 1) is necessary, since a 3-D inversion would lead to a difficult large-scale optimization problem with $2.64^3$ unknown variables.

Since the activities in the different containers of the phantom are the same, the true emission source is known up to a constant factor. However, the proportionality constant is difficult to estimate in practice and we have evaluated the quality of the reconstructions along 4-pixel-thick profiles in several slices. The correct profile is expected to be flat. Fig. 5 shows for a cortical slice that the initial $I4$ already improves over a reconstruction without attenuation correction, while the optimizer gave further improvement around the contour. Fig. 6 also shows that the reconstruction with our optimizer is almost uniform, while the reconstructed central activity without attenuation correction is seriously underestimated.

## VIII. CONCLUSION

We have presented an optimization-based attenuation-correction approach without transmission measurements using the emission data only. In the simulated study, several optimization strategies have been compared. We observed a strong dependence of the optimization on the starting point. The flat-zone regularizer performed slightly better than the Fourier-type regularizer. No significant difference between the Poisson and Gaussian objectives was observed. As was to be expected, the results on the $360°$ tour were slightly better than for the $180°$ acquisition, but the difference was not significant. Some of the reconstructions $f_{rec}$ came close to the error margin already present in the random data. In that situation, the result is close to optimal, as the preponderant fraction in the error comes from the reconstruction method itself. The phantom experiment shows improved relative quantitative attenuation correction over the ConTraSPECT and the contour method.

## REFERENCES

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

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

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

[4] F. Natterer, *The Mathematics of Computerized Tomography*. Stuttgart, Germany: Teubner Verlag, 1986.

[5] A. Welch, R. Clack, F. Natterer, and G. T. Gullberg, "Toward accurate attenuation correction in SPECT without transmission measurements," *IEEE Trans. Med. Imag.*, vol. 16, pp. 532–541, Oct. 1997.

[6] I. Laurette, R. Clackdoyle, A. Welch, and G. T. Gullberg, "Comparison of three applications of ConTraSPECT," in *Proc. IEEE NSS and MIC Conf. Rec.*, Toronto, ON, Canada, 1998.

[7] D. Gourion, X. Hatchondo, P. Gantet, D. Noll, and J.-P. Esquerré, "Comparison of two methods for SPECT attenuation correction without transmission measurements," in *Proc. IEEE NSS and MIC Conf. Rec.*, San Diego, CA, 2001.

[8] F. Natterer, "Determination of tissue attenuation in emission tomography of optically dense media," *Inverse Probl.*, vol. 9, pp. 731–736, 1993.

[9] Y. Censor, D. E. Gustafson, A. Lent, and H. Tuy, "A new approach on the emission computerized tomography problem: simultaneous calculation of attenuation and activity coefficients," *IEEE Trans. Nucl. Sci.*, vol. NS-26, pp. 2775–2779, Apr, 1979.

[10] V. Dicken, "Simultaneous activity and attenuation reconstruction in emission tomography," *Inverse Probl.*, vol. 15, pp. 931–960, 1999.

[11] ——, "Simultaneous activity and attenuation reconstruction in single photon emission computerized tomography, a nonlinear ill-posed problem," Ph.D. dissertation, Univ. Postdam, Potsdam, Germany, 1998.

[12] A. V. Bronnikov, "Numerical solution of the identification problem for the attenuated Radon transform," *Inverse Probl.*, vol. 15, pp. 1315–1324, 1999.

[13] ——, "Approximate reconstruction of attenuation map in SPECT imaging," *IEEE Trans. Nucl. Sci.*, vol. 42, pp. 1483–1488, Oct. 1995.

[14] ——, "Reconstruction of attenuation map using discrete consistency conditions," *IEEE Trans. Med. Imag.*, vol. 19, pp. 451–462, May 2000.

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

[16] V. Y. Panin, G. L. Zeng, and G. T. Gullberg, "A method of attenuation map and emission activity reconstruction from emission data," *IEEE Trans. Nucl. Sci.*, vol. 48, pp. 131–138, Feb. 2001.

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

[18] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "L-Bfgs-b:algorithm 778:l-bfgs-b, Fortran routines for large scale bound constrained optimization,," *ACM Trans. Math. Soft.*, vol. 23, no. 4, pp. 550–560, 1997.

[19] Y. Vardi, L. A. Shepp, and L. Kaufman, "A statistical model for Positron Emission Tomography," *J. Amer. Statist. Assoc.*, vol. 80, pp. 8–20, 1985.

[20] R. E. Carson and K. Lange, "The EM parametric image reconstruction algorithm," *J. Amer. Statist. Assoc.*, vol. 80, pp. 20–22, 1985.

[21] H. Hudson and R. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, vol. 13, pp. 601–609, Dec. 1994.

[22] C. L. Byrne, "Accelerating the EMML algorithm and related iterative algorithms by rescaled block-iterative methods," *IEEE Trans. Image Processing*, vol. 7, pp. 100–109, Jan. 1998.

[23] A. Krol, J. E. Bowsher, S. H. Manglos, D. H. Feiglin, M. P. Tornai, and F. D. Thomas, "An EM algorithm for estimating SPECT emission and transmission parameters from emission data only," *IEEE Trans. Med. Imag.*, vol. 20, pp. 218–232, Mar. 2001.

[24] B. Setzepfand, "ESNM: Ein rauschunterdrückendes EM-Verfahren für die Emissionstomographie," Ph.D. dissertation, Univ. Münster, Münster, Germany, 1992.

[25] C. Blondel, D. Noll, J. Maeght, A. Celler, and T. Farncombe, "Comparison of different figure of merit functions for dynamic single photon emission computed tomography (dSPECT)," in *Proc. Conf. Rec. IEEE TNS and MIC*, Lyon, France, 2000.

[26] D. M. Higdon, J. E. Bowsher, V. E. Johnson, T. G. Turkington, D. R. Gilland, and R. J. Jaszczak, "Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data," *IEEE Trans. Med. Imag.*, vol. 16, pp. 516–526, Oct. 1997.

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

[28] ——, "A new reconstruction methodology for computerized tomography: FRECT (Fourier Regularized Computed Tomography)," in *Proc. Conf. Rec. IEEE Medical Imaging Conf.*, Seattle, WA, 1999.

[29] J. Maeght and D. Noll, "Resolution in dynamic emission tomography," *SIAM J. Math. Anal.*, vol. 31, no. 5, pp. 1100–1120, 2000.

[30] P. A. Rattey and A. G. Lindgren, "Sampling the 2-D Radon transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, no. , pp. 994–1002, 1981.

[31] R. G. Novikov, "An inversion formula for the attenuated X-ray transform,," preprint.

[32] F. Natterer, "Inversion of the attenuated Radon transform," preprint.

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