Research Article

Esther Klann*, Ronny Ramlau and Peng Sun

A Mumford–Shah-type approach to simultaneous reconstruction and segmentation for emission tomography problems with Poisson statistics

DOI: 10.1515/jiip-2016-0077
Received November 3, 2016; accepted December 17, 2016

Abstract: We propose a variational model to simultaneous reconstruction and segmentation in emission tomography. As in the original Mumford–Shah model [27] we use the contour length as penalty term to preserve edge information whereas a different data fidelity term is used to measure the information discrimination between the computed tomography data of the reconstructed object and the observed (or simulated) data. As data fidelity term we use the Kullback–Leibler divergence which originates from the Poisson distribution present in emission tomography. In this paper we focus on piecewise constant reconstructions which is a reasonable assumption in medical imaging. The segmenting contour as well as the corresponding reconstructions are found as minimizers of a Mumford–Shah-type functional over the space of piecewise constant functions. The numerical scheme is implemented by evolving the level-set surface according to the shape derivative of the functional. The method is validated for simulated data with different levels of noise.

Keywords: Ill-posed problem, regularization method, Mumford–Shah, segmentation, level-set method, tomography, Poisson statistics

MSC 2010: 44A12, 47A52, 65K10, 92C55, 94A08

1 Introduction

Tomography is a widely used technique in medical imaging. Emission tomography techniques such as SPECT (*Single Photon Emission Computerized Tomography*) or PET (*Positron Emission Tomography*) provide information about the *functional level* of a part of the body. A low-level radioactive chemical, called *radiotracer* or *radiopharmaceutical* travels in the bloodstream and accumulates, e.g., in the heart, or binds to tumor cells. The concentration of the radiopharmaceutical within the body is referred to as *activity distribution* and it will be modeled as a real-valued function f. In SPECT, the radioactive material emits photons whereas in PET the tracer emits positrons that annihilate with electrons, causing two gamma photons to be emitted in opposite directions. In SPECT, the interaction (attenuation of photons) depends on the mass density distribution of the surrounding tissue, again modeled as a real-valued function μ. This mass density is assumed to

***Corresponding author: Esther Klann:** Technical University Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany; and Industrial Mathematics Institute, Johannes Kepler University Linz, Altenbergerstraße 69, 4040, Linz, Austria, e-mail: klann@indmath.uni-linz.ac.at
Ronny Ramlau: Radon Institute for Computational and Applied Mathematics (RICAM) and Johannes Kepler University Linz, Industrial Mathematics Institute and Doctoral Program "Computational Mathematics", Altenbergerstraße 69, 4040 Linz, Austria, e-mail: ronny.ramlau@jku.at
Peng Sun: Johannes Kepler University Linz, Doctoral Program "Computational Mathematics", Altenbergerstraße 69, 4040 Linz, Austria, e-mail: nakatasai@gmail.com

be known, e.g., from a separate scan. Both in SPECT and PET, the photons are measured outside the body by a γ-camera, i.e., the measured data are actually photon counts. According to [14, 18] the observable data g, i.e., the detector counts, may be modelled as independent Poisson random variables, with the mean of each variable expressed by the attenuated Radon transform, see [28],

$$A_\mu f(s,\omega) = \int_{\boldsymbol{x}\cdot\omega=s} e^{-\boldsymbol{D}\mu(\boldsymbol{x},\omega^\perp)} f(\boldsymbol{x})\, d\boldsymbol{x} \tag{1.1}$$

with $\boldsymbol{D}\mu(\boldsymbol{x},\omega^\perp) = \int_0^\infty \mu(\boldsymbol{x}+t\omega^\perp)\,dt$. In a deterministic setting the inverse problem consists in reconstructing f from measured data $A_\mu f = g$. This is an ill-posed problem and regularization methods have to be applied [12]. We proceed with the statistical approach and wish to determine the activity distribution f that is most likely given the observed data g. Besides the reconstruction of the activity distribution itself other characteristics such as the number of objects and their respective boundaries might be of interest. Our main aim in this paper is to get a segmentation of the reconstructed activity distribution simultaneously with the reconstruction. The segmentation is connected to the singularity set of f which we denote by Γ. Using Bayes's Law and prior information to include spatial constraints and further properties of the function f, we derive in Section 2 the constrained minimization problem

$$\min_{f,\Gamma} J(f,\Gamma) = \min_{f,\Gamma}\{D_{\mathrm{KL}}(g\|A_\mu f) + \alpha\mathcal{P}(f) + \beta|\Gamma|\}, \quad f \geq 0 \tag{1.2}$$

where D_{KL} denotes the *Kullback–Leibler divergence*, see Definition 2.1, $\mathcal{P}(f)$ is a general penalty term on f, and $|\Gamma|$ denotes the Hausdorff measure of the singularity set Γ of f. We further assume that the activity distribution is piecewise constant, i.e., $f = \sum_{i=1}^m c_i\chi_{D_i}$, and we minimize the regularized functional with respect to the space of piecewise constant functions. The sets $\vec{D} := (D_1,\dots,D_m)$ together with the corresponding coefficients $\vec{c} := (c_1,\dots,c_m)$ are considered as argument of the functional in (1.2). In Section 2 we choose the ℓ_2-norm of the coefficients as regularization term and get the following functional:

$$J(\vec{c},\vec{D}) = \int_{\mathbb{R}\times S^1} \left[\sum_{i=1}^m c_i(A_\mu\chi_{D_i})(s,\omega) - g(s,\omega)\log\left(\sum_{i=1}^m c_i(A_\mu\chi_{D_i})(s,\omega)\right)\right] d(s,\omega) + \alpha\|\vec{c}\,\|_{\ell^2}^2 + \beta|\partial\vec{D}|, \quad \vec{c}\geq 0 \tag{1.3}$$

with $|\partial\vec{D}| = \sum_{i=1}^m |\partial D_i|$.

We compute a minimizing argument of this functional by an alternating algorithm similar to the one introduced in [33]. We use shape sensitivity analysis [8] to update the sets $\vec{D}$, and gradient projection to compute the coefficients $\vec{c}$. The numerical realization is done using level-sets [30] and this yields direct access to the segmenting contours. The functional (1.3) fits into the scheme of Mumford–Shah-type functionals, originally introduced in [27] for the problem of image denoising and segmentation (i.e., with the identity operator), and generalized to tomography problems in [20–22, 33, 34].

Bayesian image reconstruction with Kullback–Leibler data fidelity term is considered, for example, in [6, 14, 18, 23]. In [23], the authors especially consider a TV-prior. Using the total variation as penalty term is an image reconstruction approach that preserves singularities. It was introduced in [37] with the standard L_2-fidelity term. In [46] the discrete Kullback–Leibler divergence is used together with the TV penalty for deblurring. In [38] this technique is successfully used to evaluate cardiac ${H_2}^{15}O$ observation from PET (positron emission tomography) measurements, whereas in [4] the squared H^1 norm is chosen as penalty term.

However, none of these TV-based method gives direct access to the segmenting contours (as it is the case for Mumford–Shah-type methods) but provides reconstructions that are well fitted to serve as input for a step-by-step approach to gain a segmentation, namely first reconstructing and then segmenting this reconstruction, for references and a discussion of the drawbacks see [22, 33].

The paper is organized as follows. In Section 2 we derive the optimization problem from the MAP (*Maximum A Posteriori*) perspective and introduce the piecewise constant setting. In Section 3 we briefly introduce the level-set method. In Section 4 we describe the minimization algorithm together with the shape sensitivity analysis and coefficient update scheme respectively. In Section 5 we present numerical experiments with simulated data with different levels of noise.

2 Description of the variational model

In this section we present the variational model that we use to recover the activity distribution together with its segmentation. We start with a short description of the Bayesian setting for Poisson distributed data. We then define the Kullback–Leibler divergence and adapt Bayes' estimate to the SPECT setting, resulting in the constrained optimization problem (2.6). In order to get simultaneously a reconstruction and a segmentation, we model the unknown activity distribution as a piecewise constant function. We conclude this section with the constrained optimization problem (2.7) over space of piecewise constant functions.

2.1 Bayesian estimate for Poisson distributed data

In this subsection we use Bayes' law to motivate the objective functional (1.2) given in the introduction. We assume here that g is an observation caused by an unknown f. Later on, we will specify the connection of f and g, namely the attenuated Radon transform; for the moment we assume the identity, thus leading to f and g being defined on the same domain Ω. This part follows mainly [23]. Bayes' law states

$$P(f|g) = \frac{P(g|f)P(f)}{P(g)},$$

where f and g are events and

- $P(f|g)$ is the probability of f given that g is true,
- $P(g|f)$ is the probability of g given that f is true,
- $P(g)$ and $P(f)$ are the probabilities of g and f without regard to each other.

We wish to determine that f that is most likely given the observation g, thus we wish to maximize $P(f|g)$ (*MAP – Maximum A Posteriori probability*). Since g is the given observation, we can consider $P(g)$ as fixed and equivalently maximize $P(g|f)P(f)$. In the next step we incorporate that the observable g is Poisson distributed. Recall the Poisson distribution with mean and standard deviation μ:

$$P_\mu(n) = \frac{e^{-\mu}\mu^n}{n!}, \quad n \geq 0.$$

Assuming that g is Poisson distributed, for each $x \in \Omega$ we have

$$P(g(x)|f(x)) = P_{f(x)}(g(x)) = \frac{e^{-f(x)}f(x)^{g(x)}}{g(x)!}, \quad g(x) \geq 0.$$

We discretize the domain of definition and assume that the values of g at $x_i \in \Omega$ are independent. Then

$$P(g|f) = \prod_i \frac{e^{-f(x_i)}f(x_i)^{g(x_i)}}{g(x_i)!}.$$

Instead of maximizing $P(g, f)P(f)$, we minimize $-\log P(g, f)P(f)$, the negative log-likelihood. The result is that we seek a minimizer of

$$\sum_i (f(x_i) - g(x_i)\log f(x_i)) - \log P(f), \qquad f(x_i) \geq 0,$$

where $f(x_i) \geq 0$ assures the well-definedness of the expression. We regard the previous formula as a discrete approximation of the functional

$$E(f) = \int_\Omega (f - g\log f)\,dx - \log P(f), \quad f \geq 0 \tag{2.1}$$

where the condition $f \geq 0$ assures the well-definedness of the expression.

For the prior distribution several choices are possible and have been used in inverse problems. One classical choice is the TV-prior

$$P(f) = \exp\left(-\beta \int_\Omega |\nabla f|\right),$$

where β is a regularization parameter, see [23] for the reconstruction of images corrupted by Poisson noise, and [17] for PET. The motivation for using TV is that it has been shown to suppress noise effectively while capturing sharp edges [37]. Although using the TV-prior preserves edges or discontinuities it does not give an explicit access to these discontinuities as it is the case for our method. We follow the approach of [27] to use as prior a geometrical information of the sought after function f, namely the $(n-1)$-dimensional Hausdorff-measure of its discontinuity set $\Gamma = \Gamma(f)$,

$$P(f) = \exp(-\beta|\Gamma|), \tag{2.2}$$

where β is again a regularization parameter. This choice promotes solutions with a small total length of the edge or discontinuity set Γ and stabilizes the reconstruction by preventing oscillating boundaries, see [5] for a statistical interpretation of the piecewise smooth Mumford–Shah functional with this prior. Please note that we are abusing notation here since (2.2) is not to be understood as an actual prior probability distribution that assigns probabilities. Together, equations (2.2) and (2.1) yield

$$E(f) = \int_\Omega (f - g\log f)\,dx + \beta|\Gamma|, \quad f \geq 0. \tag{2.3}$$

We now define the *Kullback–Leibler divergence*, see, e.g., [39].

Definition 2.1 (Kullback–Leibler divergence). Let g_1 and g_2 be two probability densities on $\mathbb{R}^d$. The *Kullback–Leibler divergence* is defined as

$$D_{\mathrm{KL}}(g_1\|g_2) = \int_{\mathbb{R}^2} \left(g_2(y) - g_1(y) + g_1(y)\log\left\{\frac{g_1(y)}{g_2(y)}\right\}\right) dv(y) \tag{2.4}$$

with $v(y)$ denoting the Lebesgue measure. Here, $0\log(0/0)$ as well as $0\log(0/p)$ is defined as 0. The integrand is nonnegative, so that the integral is well-defined if the value $+\infty$ is admitted.

The finiteness of the Kullback–Leibler divergence can be guaranteed, e.g., by assuming $\operatorname{supp} g_1 \subseteq \operatorname{supp} g_2$, see Assumption 2.3. For the Kullback–Leibler divergence of g and f, where g is the given data, we get from (2.4)

$$D_{\mathrm{KL}}(g\|f) = \int (f - g + g\log g - g\log f)\,dx.$$

As the data g is considered as given and therefore fixed, the minimizers of the above formula are equivalent to the minimizers of the first term of (2.3). Expressing explicitly the dependence of the functional E on not only the functional variable f but also on its singularity set Γ as well as using the *Kullback–Leibler divergence*, we get the functional

$$J(f,\Gamma) = D_{\mathrm{KL}}(g\|f) + \beta|\Gamma|, \quad f \geq 0. \tag{2.5}$$

We would like to remark that (2.5) can be considered as Mumford–Shah-type functional for the simultaneous segmentation and denoising of images with Poisson statistics. We now adapt the functional to the SPECT setting and consider the space of piecewise constant functions as minimization space.

2.2 Adaptation to SPECT setting

In SPECT (Single Photon Emission Computed Tomography), one wants to recover the activity distribution $f : \Omega \to \mathbb{R}$ of a radiopharmaceutical in the body by measuring photon counts caused by said radiopharmaceutical. Radioactive decay is described as Poisson process [7] and this carries over to the photon counts measured outside the body. As the photons travel trough the surrounding tissue, modelled as real-valued function $\mu : \Omega \to \mathbb{R}$, they are attenuated. According to [14, 18] the observable data g, i.e., the detector counts, may be modelled as independent Poisson random variables, with the mean of each variable expressed by the attenuated Radon transform A_μ, see (1.1). Given an activity distribution f that generates photons, we do not distinguish between the random variable and the concentration. The Poisson probability function of the

measured data g is

$$P(g|f) = \prod_{(s_l,\omega_k)\in M} \frac{e^{-(A_\mu f)(s_l,\omega_k)}(A_\mu f)(s_l,\omega_k)^{g_{l,k}}}{g_{l,k}!},$$

where (s_l, ω_k) refers to the discretized detector area M and $g_{l,k}$ the number of photons captured by the corresponding area. Please note that this is an idealized view as for measured data also background and readout noise has to be taken into account [3].

Following the derivation of the MAP estimate in Section 2.1 and including the length prior we get

$$\min_f\{-\log(P(g|f)P(f))\} = \min_{f,\Gamma}\left\{\sum_{l,k}((A_\mu f)(s_l,\omega_k) - g_{l,k}\log(A_\mu f)(s_l,\omega_k)) + \beta|\Gamma|\right\},$$

where $|\Gamma|$ denotes the Hausdorff measure of the singularity set of f, and we assume $A_\mu f(s_l, \omega_k) \geq 0$. So far, this is still the MAP estimate for a semi-discrete setting since the measurement space is discretized via the angles and detector positions. Similar as in [26] we let the number of both the angles and the detector position go to infinity, i.e., $l, k \to \infty$ and get the continuous version

$$\min_{f,\Gamma}\left\{\int_{\mathbb{R}\times S^1} [A_\mu f - g\log(A_\mu f)](s,\omega)\, d\sigma(s,\omega) + \beta|\Gamma|\right\}, \quad A_\mu f \geq 0.$$

For the attenuated Radon transform the constraint $A_\mu f \geq 0$ is fulfilled if $f \geq 0$ which is a natural assumption in the context of the application. For a general operator K this might become an additional condition. In [35] boundedness conditions on an integral kernel are imposed to ensure the positivity. As explained in the previous part of this section we use the Kullback–Leibler divergence and get the equivalent minimization problem

$$\min_{f,\Gamma}\{D_{\mathrm{KL}}(g\|A_\mu f) + \beta|\Gamma|\}, \quad f \geq 0.$$

In order to get regularization results one would need an additional penalty on the function f. However, we do not specify this penalty for a general functional variable f but for the piecewise constant setting introduced in the next part. Thus, with a general penalty $\mathcal{P}$ we get the minimization problem as presented in the introduction, equation (1.2),

$$(f^{\mathrm{rec}}, \Gamma^{\mathrm{rec}}) = \arg\min_{f,\Gamma} J_{(\alpha,\beta)}(f,\Gamma) \quad \text{with} \quad \begin{cases} J_{(\alpha,\beta)}(f,\Gamma) = D_{\mathrm{KL}}(g\|A_\mu f) + \alpha\mathcal{P}(f) + \beta|\Gamma| \\ \text{subject to } f \geq 0. \end{cases} \tag{2.6}$$

The attenuated Radon transform with μ known is defined as operator $A_\mu : L^2(\mathbb{R}^2) \to L^2(\mathbb{R}^2 \times S^1)$. Analogously to [32] one can prove the following lemma which will be used later.

Lemma 2.2. *For $\mu \in L^\infty(\mathbb{R}^2)$ the attenuated Radon transform is a continuous map from $L^1(\mathbb{R}^2)$ to $L^1(\mathbb{R}^2 \times S^1)$.*

We make the following assumptions.

Assumption 2.3. Let $g \in L^\infty(\mathbb{R}^2 \times S^1)$ and $f \in L_1(\mathbb{R}^2)$ be compactly supported and
(a) $\mathrm{supp}(g) \subseteq \mathrm{supp}(A_\mu f)$,
(b) there exists $\tau > 0$ such that $A_\mu f \geq \tau$ holds on $\mathrm{supp}(A_\mu f)$.

Remark 2.4. The first part of Assumption 2.3 guarantees that the quotient $\frac{g}{A_\mu f}$ is well-defined. In addition, the second part guarantees $\frac{g}{A_\mu f} \in L^\infty(\mathbb{R}^2 \times S^1)$.

In Sections 4 and 5 we present an algorithm to compute a solution to (2.6) over piecewise constant functions and discuss a strategy to exclude the value ∞ introduced by discretization of boundaries and erroneous assignment of adjacent pixel values along the boundary in the numerical experiments.

2.3 Piecewise constant functions

We assume that f vanishes outside a bounded domain $\Omega \subset \mathbb{R}^2$ and can be represented as a piecewise constant function with respect to a partition of this domain. This is considered to be a reasonable assumption

in medical imaging [13, 15, 24] and is in itself a regularization [1] as it renders certain effects, e.g., streak artifacts, impossible in the achieved reconstructions.

Let $\Gamma \subset \mathbb{R}^2$ be any finite collection of pairwise disjoint, closed, bounded curves and let f be constant on each connected component of $\mathbb{R}^2 \setminus \Gamma$. As f is assumed to have compact support, it follows that f must vanish on the unbounded component of $\mathbb{R}^2 \setminus \Gamma$. The (finitely many) connected components of $\mathbb{R}^2 \setminus \Gamma$ are denoted by $\{D_i^\Gamma\}_{i=1}^{m(\Gamma)}$ and we say that $\{D_i^\Gamma\}_{i=1}^{m(\Gamma)}$ is a *partition* of $\mathbb{R}^2$. Here, $m(\Gamma) \in \mathbb{N}$ stands for the number of connected components, not counting the unbounded component. We define the space of piecewise constant functions with respect to the geometry Γ (or with respect to the partition $\{D_i^\Gamma\}_{i=1}^{m(\Gamma)}$) as

$$\mathrm{PC}(\mathbb{R}^2 \setminus \Gamma) = \left\{ \sum_{i=1}^{m(\Gamma)} c_i \chi_{D_i^\Gamma} : c_i \in \mathbb{R},\ \nu(D_i) > 0 \right\},$$

where χ_D denotes the characteristic function of the set D. The characteristic functions $\{\chi_{D_i^\Gamma}\}_{i=1}^{m(\Gamma)}$ form a basis in $\mathrm{PC}(\mathbb{R}^2 \setminus \Gamma)$. Note that the characteristic function of the unbounded component of $\mathbb{R}^2 \setminus \Gamma$ is not an element of the basis. To simplify notation we write $m = m(\Gamma)$ for the number of connected components, $\mathrm{PC}_m := \mathrm{PC}(\mathbb{R}^2 \setminus \Gamma)$, and $D_i = D_i^\Gamma$.

A function $f \in \mathrm{PC}_m$ is uniquely determined by the collection of sets $\vec{D} := (D_1, \dots, D_m)$ and the corresponding vector of coefficients $\vec{c} := (c_1, \dots, c_m)$ via the operator

$$f = \mathcal{F}(\vec{c}, \vec{D}) := \sum_{i=1}^{m} c_i \chi_{D_i}.$$

The singularity set of f is given by the boundaries of the sets D_i, i.e., $\Gamma = \bigcup_{i=1}^m \partial D_i$. As outer boundaries of one set in $\vec{D}$ can be inner boundaries of another set, we introduce as length penalty the *total perimeter of the partition*

$$|\partial \vec{D}| = \sum_{i=1}^{m} p_\Omega(\partial_* D_i) = |\Gamma|.$$

Here, $p_\Omega(A)$ denotes the perimeter of a measurable set $A \subset \Omega$ defined by

$$p_\Omega(A) = \sup\left\{ \int_A \operatorname{div}(\boldsymbol{v}(\boldsymbol{x}))\, d\boldsymbol{x} \;\middle|\; \boldsymbol{v} : \Omega \to \mathbb{R}^n,\ \boldsymbol{v} \text{ is smooth with } \operatorname{supp}(\boldsymbol{v}) \subset \Omega \text{ and } \|\boldsymbol{v}(\boldsymbol{x})\| \le 1 \text{ for all } \boldsymbol{x} \in \Omega \right\},$$

and $\partial_* A$ is defined as the set of all points in A that are neither Lebesgue points for A nor for $\Omega \setminus A$, see [2] for details.

We now rewrite (2.6) and get the piecewise constant constraint minimization problem,

$$(\vec{c}^{\,\mathrm{rec}}, \vec{D}^{\,\mathrm{rec}}) = \operatorname*{arg\,min}_{\vec{c},\vec{D}} J_{(\alpha,\beta)}(\vec{c}, \vec{D}) \quad \text{with} \quad \begin{cases} J_{(\alpha,\beta)}(\vec{c}, \vec{D}) = D_{\mathrm{KL}}(g \| A_\mu(\mathcal{F}(\vec{c}, \vec{D}))) + \alpha \|\vec{c}\,\|_{\ell^2}^2 + \beta |\partial \vec{D}| \\ \text{subject to } \vec{c} \ge 0. \end{cases} \tag{2.7}$$

3 Level sets for geometry evolution

We use the level-set methodology [31] to represent the segmenting contours. Although this can be found in previous works [22, 33], we give a short introduction for the sake of completeness and to explain some of the notation we use later on.

Introducing a Lipschitz continuous (level-set) function $\phi : \mathbb{R}^2 \to \mathbb{R}$, we set

$$\Gamma = \{\boldsymbol{x} \in \mathbb{R}^2 : \phi(\boldsymbol{x}) = 0\}.$$

The connected components of $\mathbb{R} \setminus \Gamma$, named $D_1, D_2, \dots, D_m$ with $m = m(\Gamma)$ then constitute a partition of $\mathbb{R}^2$. Please note that we only use one level-set function ϕ to define the partition. In the algorithm described in the next section the number of components is not fixed, and the level-set framework allows for topological changes such as splitting and merging of sets. We use the occurrence of a change of sign in the level-set function to distinguish between components that have a common boundary (adjacent components).

Using the level-set representation ϕ for the geometry Γ, we can describe the propagation of the geometry with the help of a propagating (time-dependent) level-set function,

$$\Gamma(t) = \{\boldsymbol{x} \in \mathbb{R}^2 : \phi(\boldsymbol{x}, t) = 0\}$$

where $\phi(\boldsymbol{x}, t)$ is the solution to

$$\phi_t + F|\nabla\phi| = 0 \quad \text{and} \quad \phi(\boldsymbol{x}, t) = \phi_0 \tag{3.1}$$

with $\Gamma_0 = \{\phi_0 = 0\}$. The scalar speed function F acts as the direction of perturbation for the update of the geometry. The time-evolution of the level-set equation (3.1) is closely related to a propagation of Γ along the flow generated by the vector field $\boldsymbol{v} = F\frac{\nabla\phi}{|\nabla\phi|}$, see [30] for more details. Here we denote the time-dependent level-set evolution procedure with respect to the scalar speed function F as $\Gamma(t) = \mathcal{D}_t(\Gamma_0; F)$.

In the algorithm described in the next section we use the shape derivative of the functional (2.7) to update the geometry. We denote by $\boldsymbol{n}_{D_i}$ the exterior unit normal vector field to the set D_i. It is known [8] that the derivative of a shape dependent functional depends on the direction of perturbation $\boldsymbol{v}$ only via its orthonormal projections $\langle \boldsymbol{v}, \boldsymbol{n}_{D_i}\rangle$. We set

$$s_i := -\operatorname{sign}(\phi(\boldsymbol{z})) \quad \text{with an arbitrary } \boldsymbol{z} \in D_i.$$

With this, we have

$$\langle \boldsymbol{v}, \boldsymbol{n}_{D_i}\rangle = s_i F \tag{3.2}$$

and we can express the directional derivative of a functional of the geometry Γ in terms of the level-set type speed function F.

In the following computation we use some well-known results from shape sensitivity analysis. The differentiation rules for domain and boundary functionals

$$J_d(D) = \int_D g\,dx \quad \text{and} \quad J_b(\Gamma) = \int_\Gamma h\,d\sigma$$

are given by

$$dJ_d(D; F) = \int_{\partial D} gF\,d\sigma \quad \text{and} \quad dJ_b(\Gamma; F) = \int_\Gamma (\langle \nabla h, \mathbf{n}\rangle + h\kappa)F\,d\sigma.$$

Here κ is the mean curvature of Γ and $\mathbf{n}$ is the unit exterior normal, see [8, 16, 41] for more details. The sign variables s_i defined in (3.2) become important when the functional depends on more than one set. In that case adjacent sets share a boundary and the *exterior* unit normal of one set is the *interior* unit normal of the other.

4 Minimization algorithm

In this section we describe the minimization algorithm for problem (2.7). An algorithm that updates both the geometry $\vec{D}$ and the coefficients $\vec{c}$ simultaneously is difficult to formulate as the geometry defines the underlying domains for the computation of the optimal coefficients. Hence, we use an *alternating procedure* described in Algorithm 1, see also [20, 22, 33]. For fixed sets $\vec{D}$ we solve the variational problem

$$\arg\min_{\vec{c}\,\geq 0} J(\vec{c}, \vec{D}).$$

This means, we compute the best possible coefficients for a given geometry. We denote the minimizer as $\vec{c}^{\,*}$ and the corresponding function as $f^* = \mathcal{F}(\vec{c}^{\,*}, \vec{D})$. With this minimizer we consider the *shape optimization problem*

$$\arg\min_{\vec{D}} \hat{J}(\vec{D}) \quad \text{with} \quad \hat{J}(\vec{D}) := J(\vec{c}^{\,*}, \vec{D}) \tag{4.1}$$

and update the geometry such that the value of the cost functional decreases.

Algorithm 1 (Minimization of the objective functional (2.7)). Proceed as follows.

- **Step 1:** Choose an initial estimate $\vec{D}^0$ for the geometry and $\vec{c}^{\,0}$ for the coefficients such that $f^0 = \mathcal{F}(\vec{c}^{\,0}, \vec{D}^0)$ satisfies Assumption 2.3.
- **Step 2:** For fixed geometry $\vec{D}$ minimize J with respect to $\vec{c} \geq 0$ by solving the corresponding minimization problem. Let the optimal solution of this problem be $\vec{c}^{\,*} = \vec{c}^{\,*}(\vec{D}) \geq 0$.
- **Step 3:** Consider the reduced functional

$$\hat{J}(\vec{D}) = J(\vec{c}^{\,*}, \vec{D}).$$

  Find a descent direction F for the functional $\hat{J}$ with respect to the sets $\vec{D}$.
- **Step 4:** Update $\vec{D}$ according to the chosen descent direction F using an appropriate line search rule. Use a level-set formulation for the update of the geometry as follows:
  - Set the initial step size as t_0, $0 < q < 1$ and $l = 0$. Given the old shape Γ_{old}, the new test shape $\Gamma_{\text{new}}^{t_l}$ is seen as a time-evolution process $\mathcal{D}_t(\Gamma_{\text{old}}; F)$ from the old shape Γ_{old} with time step t_l,

$$\Gamma_{\text{new}}^{t_l} = \mathcal{D}_{t_l}(\Gamma_{\text{old}}; F) \tag{4.2}$$

    which satisfies $\Gamma_{\text{old}} = \mathcal{D}_0(\Gamma_{\text{old}}; F)$.
  - Check the values of $\frac{\partial H}{\partial c_i}$ in (4.6):
    - If $\frac{\partial H}{\partial c_i} < \infty$, then accept the new shape $\Gamma_{\text{new}} = \Gamma_{\text{new}}^{t_l}$ and go to **Step 5**.
    - Else set $l = l + 1$, $t_l = q^l t_0$ and go to (4.2).
- **Step 5:** Check for optimality:
  - If the shape gradient is large, go to **Step 2**.
  - If the shape gradient is small, then determine the derivative of the cost functional with respect to the functional variable $f = \mathcal{F}(\vec{c}, \vec{D})$. If a significant maximum or minimum exists for the functional gradient, introduce a new set in the vicinity of the extremum. Go back to **Step 2**.
- **Step 6:** Repeat **Steps 2–5** until a pre-set number of iteration steps is reached.

We now describe the individual steps in details.

4.1 Step 1: Choosing an initial estimate

An initial estimate can be gained directly from the measured data by using filtered back-projection to find the convex hull of the support of the object [22]. In [20], lambda reconstruction is used to get a starting configuration. As bones divide the boundaries of inner organs one can also use structural information from an X-ray CT modality as hint to a proper initial estimate [10, 36]. In order to have well-definedness, Assumption 2.3 must hold for the initial estimate.

4.2 Step 2: Optimal solution $\vec{c}\,(\vec{D})$ for fixed geometry $\vec{D}$

For fixed geometry $\vec{D} = (D_1, \ldots, D_m)$, we want to determine the coefficient vector $\vec{c} = (c_1, \ldots, c_m)$ that minimizes

$$D_{\text{KL}}(g \| A_\mu \mathcal{F}(\vec{c}, \vec{D})) + \alpha \|\vec{c}\,\|_{\ell^2}^2, \quad \vec{c} \geq 0. \tag{4.3}$$

Setting $\alpha = 0$ and considering the functional variable $f = \mathcal{F}(\vec{c}, \vec{D})$ a standard method from statistics for finding the maximum-a-posteriori estimate is the iterative *expectation maximization* (EM) algorithm [9], see [40] for its application to PET (Positron Emission Tomography). More details about the EM algorithm with regularization term can be found in [46]. However, the EM iteration rule 'destroys' the structure of our solution space, i.e., the resulting minimizer is no longer piecewise constant.

In order to keep the structure of the solution we consider (4.3) as a function of the coefficient vector, i.e., we define

$$H : \mathbb{R}^m \to \mathbb{R}$$

by

$$\begin{aligned} H(\vec{c}\,) &= D_{\mathrm{KL}}(g\|A_\mu \mathcal{F}(\vec{c},\vec{D})) + \alpha\|\vec{c}\,\|_{\ell^2}^2 \\ &= \int_{\mathbb{R}\times S^1} \left[\sum_{i=1}^m c_i (A_\mu \chi_{D_i})(s,\omega) - g(s,\omega)\log\left(\sum_{i=1}^m c_i (A_\mu\chi_{D_i})(s,\omega)\right)\right] \\ &\qquad - [g(s,\omega) - g(s,\omega)\log g(s,\omega)]\, d(s,\omega) + \alpha\|\vec{c}\,\|_{\ell^2}^2 \end{aligned} \tag{4.4}$$

with $\vec{c} = (c_1, c_2, \ldots, c_m)$. We then solve the constrained optimization problem

$$\min_{\vec{c}\geq 0} H(\vec{c}\,). \tag{4.5}$$

Before proving the existence of a unique minimizer in the following lemma we recall that a real-valued functional H defined on $\Sigma \subset \mathbb{R}^m$ is called *coercive* if $H(\vec{c}\,) \to \infty$ whenever $\vec{c} \in \Sigma$ and $\|\vec{c}\,\|_{\ell^2} \to \infty$.

Lemma 4.1. *Let $\alpha > 0$ and H as in* (4.4). *The constraint minimization problem* (4.5) *has a unique minimizer $\vec{c}^{\,*}$.*

Proof. Assuming that H is coercive and strictly convex, existence and uniqueness of a minimizer of (4.4) follow along the same lines as in the proof of [11, Proposition 1.2], based on the fact that the set $\{\vec{c} \in \mathbb{R}^m \mid \vec{c} \geq 0\}$ is closed and convex.

Since $D_{\mathrm{KL}}(\cdot\|\cdot) \geq 0$ in [35], for $\alpha > 0$ the function $H(\vec{c}) \to \infty$ as $\|\vec{c}\,\|_{\ell^2}^2 \to \infty$ and hence the coercivity of (4.4) follows. It remains to show that H is strictly convex. According to [35, Lemma 3.4], $D_{\mathrm{KL}}(g\|A_\mu F(\vec{c},\vec{D}))$ is convex with respect to the vector variable $\vec{c}$ and therefore the combination of $D_{\mathrm{KL}}(g\|A_\mu F(\vec{c},\vec{D}))$ and the strictly convex term $\|\vec{c}\,\|_{\ell^2}^2$ is strictly convex. □

For the computation of the minimizer of functional (4.4) we use the *gradient projection method*, see Algorithm 2 and [19, Algorithm 5.4.1].

Algorithm 2 (Minimization of the functional (4.5) by gradient projection method). Proceed as follows.

- **Step 1:** Set initial guess $\vec{c}^{\,0} \geq 0$ and $k = 0$,
- **Step 2:** *(Gradient projection)*
  - Given $\vec{c}^{\,k}$, apply the gradient projection method [19, Algorithm 5.4.1], i.e.,

    $$\vec{c}^{\,k}(\lambda) = P_{\vec{c}\geq 0}(\vec{c}^{\,k} - \lambda \nabla H(\vec{c}^{\,k}))$$

    with $P_{\vec{c}\geq 0}(\cdot)$ projecting the iterate $\vec{c}^{\,k} - \lambda\nabla H(\vec{c}^{\,k})$ back to the constraints set $\{\vec{c} \in \mathbb{R}^m \mid \vec{c} \geq 0\}$. The step length $\lambda > 0$ is chosen such that it satisfies the Armijo condition

    $$H(\vec{c}^{\,k}(\lambda)) \leq H(\vec{c}^{\,k}) - m\lambda|\nabla H(\vec{c}^{\,k})|^2, \quad 0 < m < 1,$$

    as well as $|\nabla H(\vec{c}^{\,k}(\lambda))| < +\infty$; see Lemma 4.3.
  - Update $\vec{c}^{\,k} = \vec{c}^{\,k}(\lambda)$, set $k = k + 1$ and check the stopping criterion, e.g., [25]. If the criterion is satisfied, stop the iteration; otherwise, go to **Step 2**.

The gradient projection method requires the derivative of the functional H with respect to the coefficients c_i. Straightforward computation yields

$$\frac{\partial H}{\partial c_i} = \int_{D_i} A_\mu^*\left(\mathbf{1} - \frac{g}{\sum_{j=1}^m c_j A_\mu \chi_{D_j}}\right)(\boldsymbol{x})\, d\boldsymbol{x} + 2\alpha c_i, \tag{4.6}$$

where $\mathbf{1}$ denotes the constant function which has value 1 everywhere. In the following we consider the existence and finiteness of (4.6) for given data g.

Lemma 4.2. *Let $g \in L^\infty(\mathbb{R}\times S^1)$ and let Assumption 2.3 hold for $f = \sum_{j=1}^m c_j\chi_{D_j}$. Then the derivative in* (4.6) *exists and is finite.*

Proof. As $g \in L^\infty(\mathbb{R}\times S^1)$, $\operatorname{supp} g \subseteq \operatorname{supp} A_\mu f$ and $A_\mu f$ is bounded away from zero, we have

$$(g/A_\mu f) \in L^\infty(\operatorname{supp}(A_\mu f)).$$

According to Lemma 2.2 we have

$$A_\mu : L^1(\mathbb{R}^2) \to L^1(\mathbb{R} \times S^1)$$

and thus, for the adjoint operator, $A_\mu^* : L^\infty(\mathbb{R} \times S^1) \to L^\infty(\mathbb{R}^2)$. □

Without causing ambiguities, we say that the pair $(\vec{c}, \vec{D})$ fulfills Assumption 2.3 if the corresponding function $f := \mathcal{F}(\vec{c}, \vec{D}) = \sum_{j=1}^m c_j \chi_{D_j}$ fulfills Assumption 2.3. We now show that there is a step-length λ that satisfies the conditions in Step 2 of Algorithm 2.

Lemma 4.3. *Let Assumption 2.3 hold for $(\vec{c}, \vec{D})$. Then there exists a step-length λ such that both the Armijo condition*

$$H(\vec{c}(\lambda)) \leq H(\vec{c}) - m\lambda |\nabla H(\vec{c})|^2, \quad 0 < m < 1,$$

and the condition $|\nabla H(\vec{c}(\lambda))| < +\infty$ are satisfied.

Proof. For the existence of the step-length that satisfies the Armijo condition see [29, Lemma 3.1]. Let $\mathcal{S}$ denote the set of step-lengths λ that satisfy the Armijo condition. Since Assumption 2.3 holds for $(\vec{c}, \vec{D})$, we have $H(\vec{c}) < \infty$ and $|\nabla H(\vec{c})| < \infty$. The projection $P_{\vec{c} \geq 0}$ guarantees that $A_\mu f$ is bounded away from zero on its support as long as there is at least one $c_j \neq 0$. The latter can be assumed as long as the data g does not vanish completely. We now assume that for all $\lambda \in \mathcal{S}$, $|\nabla H(\vec{c}^{\,k}(\lambda))| = \infty$. Rewriting the logarithmic terms in functional (4.4) as a quotient, we see that $|\nabla H(\vec{c}^{\,k}(\lambda))| = \infty$ then implies $H(\vec{c}^{\,k}(\lambda)) = \infty$ which is a contradiction to the inequality from the Armijo condition. □

The following lemma shows that Assumption 2.3 holds for all iterates of Algorithm 2.

Lemma 4.4. *Let Assumption 2.3 hold for $(\vec{c}^{\,0}, \vec{D}^{\,0})$. Then Assumption 2.3 holds for each iterate $(\vec{c}^{\,k}, \vec{D}^{\,0})$, $k \in \mathbb{N}$, in Algorithm 2.*

Proof. The projection operator $P_{\vec{c} \geq 0}$ guarantees that $A_\mu f$ is bounded away from zero on its support as long as there is at least one $c_j \neq 0$. The latter can be assumed as long as the data g does not vanish completely. It remains to show that part (a) of Assumption 2.3 holds for each iterate $(\vec{c}^{\,k}, \vec{D}^{\,0})$, $k \in \mathbb{N}$. As $(\vec{c}^{\,0}, \vec{D}^{\,0})$ fulfills Assumption 2.3, one has

$$H(\vec{c}^{\,0}) < +\infty \quad \text{and} \quad |\nabla H(\vec{c}^{\,0})| < +\infty.$$

According to Lemma 4.3, there exists $\lambda_1 > 0$ such that $\vec{c}^{\,1} = \vec{c}^{\,0}(\lambda_1)$ satisfies

$$H(\vec{c}^{\,1}) \leq H(\vec{c}^{\,0}) - m\lambda_1 |\nabla H(\vec{c}^{\,0})|^2, \quad 0 < m < 1,$$

and $|\nabla H(\vec{c}^{\,1})| < +\infty$. Assuming that Assumption 2.3 does not hold for $(\vec{c}^{\,1}, \vec{D}^{\,0})$, we get $H(\vec{c}^{\,1}) = \infty$. This contradicts the inequality above as both terms on the right-hand side have finite values, hence the assumption holds for $(\vec{c}^{\,1}, \vec{D}^{\,0})$. The same arguments are valid for general k. Hence, the lemma follows by induction. □

Remark 4.5. The input for Algorithm 2 is either the vector $\vec{c}^{\,0}$ from the initial guess or given by the output from Step 4 in Algorithm 1. Assumption 2.3 holds for both.

4.3 Step 3: Shape sensitivity analysis of the reduced functional and choice of descent direction

In the following we consider the shape optimization problem (4.1). For that we compute the shape derivative of the reduced functional $\hat{J}(\vec{D}) = J(\vec{c}^{\,*}(\vec{D}), \vec{D})$, where $\vec{c}^{\,*}(\vec{D})$ is the optimal solution for fixed geometry $\vec{D}$ that was obtained in the previous step. The derivative of the reduced functional $\hat{J}(\vec{D}) = J(\vec{c}^{\,*}(\vec{D}), \vec{D}))$ with respect to the geometric variable $\vec{D}$ formally reads as

$$d\hat{J}(\vec{D}; F) = \partial_{\vec{c}} J(\vec{c}^{\,*}, \vec{D})(\vec{c}^{\,*})'(\vec{D}; F) + d_{\vec{D}} J(\vec{c}^{\,*}, \vec{D}; F), \tag{4.7}$$

where $\partial_{\vec{c}}$ denotes the derivative with respect to $\vec{c}$ for fixed $\vec{D}$, and $(\vec{c}^{\,*})'(\vec{D}; F)$ is the shape derivative of $\vec{c}^{\,*}$ with respect to $\vec{D}$ in direction F and $d_{\vec{D}} J(\vec{c}^{\,*}, \vec{D}; F)$ denotes the Eulerian derivative of J in direction F for fixed $\vec{c}^{\,*}$.

Please note that in the case of an L_2-data fit term as considered in [22, 33] the first term of (4.7) vanishes due to the optimality conditions of the previous step of the alternating minimization algorithm. This is not the case for the Kullback–Leibler data fit term as it requires the non-negativity constraint on the coefficient vector $\vec{c}$. Note that whenever the positivity constraint is not active,i.e., if the unconstrained problem fulfills $\vec{c} > 0$, then the first part vanishes. As this yields anyway in most cases, we will neglect the first term in (4.7).

Proposition 4.6. *Let A_μ^* denote the adjoint of the attenuated Radon transform, and $\vec{c}^{\,*}$ the solution to* (4.5). *Let $\alpha, \beta > 0$ and $f^* = \mathcal{F}(\vec{c}^{\,*}, \vec{D})$ such that $g/(A_\mu f^*)$ is well-defined. The Eulerian shape derivative in direction F of the functional*

$$J(\vec{c}^{\,*}, \vec{D}) = D_{\mathrm{KL}}(g \| A_\mu \mathcal{F}(\vec{c}^{\,*}, \vec{D})) + \alpha \|\vec{c}^{\,*}\|_{\ell_2}^2 + \beta |\partial \vec{D}|$$

is given by

$$d_{\vec{D}} J(\vec{c}^{\,*}, \vec{D}; F) = \sum_{i=1}^{m} c_i^* s_i \int_{\boldsymbol{x} \in \partial D_i} \left(A_\mu^* \left(\mathbf{1} - \frac{g}{A_\mu f^*} \right) \right)(\boldsymbol{x}) \cdot F \, d\sigma + \beta \int_\Gamma \kappa F \, d\sigma,$$

with $f^* = \mathcal{F}(\vec{c}^{\,*}, \vec{D})$, $\Gamma = \bigcup_i \partial D_i$ and κ is the mean curvature of Γ. The s_i are the sign variables indicating the interior and exterior of the sets D_i, see equation (3.2).

Proof. To simplify notation we omit from now on the superscript * in $\vec{c}^{\,*}$ and f^*. We expand and split the functional J as follows:

$$\begin{aligned} J(\vec{c}, \vec{D}) &= \int_{\mathbb{R}\times S^1} \left\{ \left(A_\mu \left(\sum_{i=1}^m c_i \chi_{D_i} \right) \right)(s,\omega) - g(s,\omega) \log \left(\left(A_\mu \left(\sum_{i=1}^m c_i \chi_{D_i} \right) \right)(s,\omega) \right) \right\} d(s,\omega) + \alpha \sum_{i=1}^m c_i^2 + \beta \int_{\bigcup_i \partial D_i} \mathbf{1} \, d\sigma \\ &= \underbrace{\sum_{i=1}^m c_i \int_{\mathbb{R}\times S^1} A_\mu \chi_{D_i}(s,\omega) \, d(s,\omega)}_{=: I_1(\vec{c},\vec{D})} - \underbrace{\int_{\mathbb{R}\times S^1} g(s,\omega) \log \left(\sum_{i=1}^m c_i A_\mu \chi_{D_i}(s,\omega) \right) d(s,\omega)}_{=: I_2(\vec{c},\vec{D})} + \underbrace{\alpha \sum_{i=1}^m c_i^2}_{=: I_3(\vec{c})} + \underbrace{\beta \int_\Gamma \mathbf{1} \, d\sigma}_{=: I_4(\vec{c},\Gamma)} \end{aligned}$$

with $\Gamma = \bigcup_i \partial D_i$. The penalty on the coefficients does not depend on the geometric variable and hence does not contribute to the derivative, i.e., $d_{\vec{D}} I_3 = 0$. Straightforward application of the rule for a boundary functional yields

$$d_\Gamma I_4(\vec{c}, \Gamma; F) = \beta \int_\Gamma \kappa F \, d\sigma,$$

where κ is the mean curvature of Γ. We rewrite I_1 as

$$\begin{aligned} I_1(\vec{c}, \vec{D}) &= \sum_{i=1}^m c_i \int_{\mathbb{R}\times S^1} (A_\mu \chi_{D_i})(s,\omega) \, d(s,\omega) \\ &= \sum_{i=1}^m c_i \int_{\mathbb{R}\times S^1} ds \, d\omega \int_{\mathbb{R}} \chi_{D_i}(s\omega + t\omega^\perp) \exp\left(- \int_{\tau=t}^{\infty} \mu(s\omega + \tau\omega^\perp) \, d\tau \right) dt \\ &= \sum_{i=1}^{m)} c_i \int_{S^1} d\omega \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \chi_{D_i}(s\omega + t\omega^\perp) \exp\left(- \int_{\tau=t}^{\infty} \mu(s\omega + \tau\omega^\perp) \, d\tau \right) dt \right] ds \\ &= \sum_{i=1}^m c_i \int_{S^1} d\omega \int_{\mathbb{R}^2} \left[\chi_{D_i}(\boldsymbol{x}) \exp\left(- \int_{\tau=0}^{\infty} \mu(\boldsymbol{x} + \tau\omega^\perp) \, d\tau \right) \right] d\boldsymbol{x} \\ &= \sum_{i=1}^m c_i \int_{S^1} d\omega \int_{D_i} \left[\mathbf{1}(\boldsymbol{x}) \exp\left(- \int_{\tau=0}^{\infty} \mu(\boldsymbol{x} + \tau\omega^\perp) \, d\tau \right) \right] d\boldsymbol{x}, \end{aligned}$$

where we have used the transformation

$$\boldsymbol{x} = s\omega + t\omega^\perp,$$

i.e., $s = \langle \boldsymbol{x}, \omega \rangle$ and $t = \langle \boldsymbol{x}, \omega^\perp \rangle$. We apply the rule for the shape derivative of a domain functional and get

$$\begin{aligned} d_{\vec{D}} I_1(\vec{c}, \vec{D}; F) &= \sum_{i=1}^{m} c_i \int_{S^1} d\omega \, d_\Gamma \Big[\int_{D_i} \mathbf{1}(\boldsymbol{x}) \exp\Big(- \int_{\tau=0}^{\infty} \mu(\boldsymbol{x} + \tau\omega^\perp) \, d\tau \Big) \, d\boldsymbol{x}; F \Big] \\ &= \sum_{i=1}^{m} c_i s_i \int_{S^1} d\omega \int_{\boldsymbol{x} \in \partial D_i} F \cdot \mathbf{1}(\boldsymbol{x}) \exp\Big(- \int_{\tau=0}^{\infty} \mu(\boldsymbol{x} + \tau\omega^\perp) \, d\tau \Big) \, d\sigma \\ &= \sum_{i=1}^{m} c_i s_i \int_{\boldsymbol{x} \in \partial D_i} F \int_{\omega \in S^1} \mathbf{1}(\boldsymbol{x}) \exp\Big(- \int_{\tau=0}^{\infty} \mu(\boldsymbol{x} + \tau\omega^\perp) \, d\tau \Big) \, d\sigma \\ &= \sum_{i=1}^{m} c_i s_i \int_{\boldsymbol{x} \in \partial D_i} (A_\mu^* \mathbf{1})(\boldsymbol{x}) \cdot F \, d\sigma. \end{aligned}$$

The derivative of I_2 is

$$\begin{aligned} d_{\vec{D}} I_2(\vec{c}, \vec{D}; F) &= d_{\vec{D}} \Big[\int_{\mathbb{R} \times S^1} g(s, \omega) \log((A_\mu f)(s, \omega)) \, d(s, \omega); F \Big] \\ &= \int_{\mathbb{R} \times S^1} g(s, \omega) \, d_\Gamma [\log((A_\mu f)(s, \omega)); F] \, d(s, \omega) \\ &= \int_{\mathbb{R} \times S^1} \frac{g(s, \omega)}{A_\mu f(s, \omega)} d_{\vec{D}} [(A_\mu f)(s, \omega); F] \, d(s, \omega) \\ &= \sum_{i=1}^{m} c_i \int_{\mathbb{R} \times S^1} \frac{g(s, \omega)}{A_\mu f(s, \omega)} \, d_{\vec{D}} [(A_\mu \chi_{D_i})(s, \omega); F] d(s, \omega). \end{aligned}$$

Similar as in the derivation of the shape derivative of I_2 we transform the expression into a domain functional. For that we expand the operator A_μ as well as interchange the order of differentiation (i.e., the shape derivative $d_{\vec{D}}$) and the integration for the variable s. In doing so we ignore the shape dependence of the term $g/A_\mu f$ in the differentiation (this terms actually stems from the differentiation that was already carried out) even though the term formally appears inside the range of the differential operator. Again using the transformations $\boldsymbol{x} = s\omega + t\omega^\perp$, i.e., $s = \langle \boldsymbol{x}, \omega \rangle$ and $t = \langle \boldsymbol{x}, \omega^\perp \rangle$, we obtain

$$\begin{aligned} d_{\vec{D}} I_2(\vec{c}, \vec{D}; F) &= \sum_{i=1}^{m} c_i \int_{\mathbb{R} \times S^1} \frac{g(s, \omega)}{A_\mu f(s, \omega)} \, d_{\vec{D}} \Big[\int_{t \in \mathbb{R}} \chi_{D_i}(s\omega + t\omega^\perp) \exp\Big(- \int_{\tau=t}^{\infty} \mu(s\omega + \tau\omega^\perp) \, d\tau \Big) \, dt; F \Big] \, d(s, \omega) \\ &= \sum_{i=1}^{m} c_i \int_{\omega \in S^1} d\omega \, d_{\vec{D}} \Big[\int_{s \in \mathbb{R}} \int_{t \in \mathbb{R}} \frac{g(s, \omega)}{A_\mu f(s, \omega)} \chi_{D_i}(s\omega + t\omega^\perp) \exp\Big(- \int_{\tau=t}^{\infty} \mu(s\omega + \tau\omega^\perp) \, d\tau \Big) \, dt \, ds; F \Big] \\ &= \sum_{i=1}^{m} c_i \int_{\omega \in S^1} d\omega \, d_{\vec{D}} \Big[\int_{\boldsymbol{x} \in \mathbb{R}^2} \chi_{D_i}(\boldsymbol{x}) \frac{g}{A_\mu f}(\langle \boldsymbol{x}, \omega \rangle, \omega) \exp\Big(- \int_{\tau=0}^{\infty} \mu(\boldsymbol{x} + \tau\omega^\perp) \, d\tau \Big) \, d\boldsymbol{x}; F \Big]. \end{aligned}$$

We now deal with the differentiation of a domain integral over D_i. Application of the corresponding rule, and using the adjoint operator, we get

$$\begin{aligned} d_{\vec{D}} I_2(\vec{c}, \vec{D}; F) &= \sum_{i=1}^{m} c_i s_i \int_{\omega \in S^1} d\omega \int_{\boldsymbol{x} \in \partial D_i} F \cdot \frac{g}{A_\mu f}(\langle \boldsymbol{x}, \omega \rangle, \omega) \exp\Big(- \int_{\tau=0}^{\infty} \mu(\boldsymbol{x} + \tau\omega^\perp) \, d\tau \Big) \, d\sigma \\ &= \sum_{i=1}^{m} c_i s_i \int_{\boldsymbol{x} \in \partial D_i} A_\mu^* \Big(\frac{g}{A_\mu f} \Big)(\boldsymbol{x}) \, d\sigma. \end{aligned}$$

The finiteness of the above expression is guaranteed by the same arguments as used in the proof of Lemma 4.2. Combining the results for I_1, I_2, I_3 and I_4 concludes the proof. □

Corollary 4.7 (Descent direction). *Let all conditions of Proposition 4.6 hold. One possible choice for a descent direction of the functional*

$$J(\vec{c}^{\,*}, \vec{D}) = D_{\mathrm{KL}}(g \| A_\mu \mathcal{F}(\vec{c}^{\,*}, \vec{D})) + \alpha \|\vec{c}^{\,*}\|_{\ell_2}^2 + \beta |\partial \vec{D}|$$

is

$$F(\boldsymbol{x}) = -\left(\sum_{i=1}^{m} c_i^* s_i \left(A_\mu^* \left(\mathbf{1} - \frac{g}{A_\mu f} \right) \right)(\boldsymbol{x}) + \beta \kappa(\boldsymbol{x}) \right), \quad \boldsymbol{x} \in \partial \vec{D},$$

if $\vec{c}^{\,*} > 0$ *is the solution to the unconstrained problem.*

4.4 Step 4: Update of the geometry

At this step we update the geometry by the level-set evolution $\mathcal{D}_t(\Gamma; F)$ introduced in Section 3, while guaranteeing that the support inclusion, see Assumption 2.3 stays valid. An initial time step-size t_0 is set in advance and the new test geometry is updated as $\Gamma_{\mathrm{new}}^{t_0} = \mathcal{D}_{t_0}(\Gamma_{\mathrm{old}}; F)$ in the chosen direction F from Corollary 4.7. For the new test geometry $\Gamma_{\mathrm{new}}^{t_0}$, we check the values of $\frac{\partial H}{\partial c_i}$. If the values are finite, then the new test geometry is adopted as the new geometry Γ_{new}, otherwise we decrease the time step-size by multiplying a factor $0 < q < 1$, i.e., $t_l = q^l t_0$ and update the new test geometry as $\Gamma_{\mathrm{new}}^{t_l} = \mathcal{D}_{t_l}(\Gamma_{\mathrm{old}}; F)$ for $l = 1, 2, \ldots$ until the derivative values are finite. The support-inclusion condition then holds as long as the condition is satisfied for the initial setup $(\vec{c}^{\,0}, \vec{D}^0)$ from Step 1.

4.5 Step 5: Check for optimality and insert strategy

The shape derivative computed so far describes the sensitivity of the cost functional with respect to (local) changes of the boundaries of the existing sets. The level-set method that is used for the numerical realization allows topology changes such as splitting or merging of sets. Introducing new components away from existing boundaries is not possible. Thus, the evolution of the geometry can get stuck in a local minimum. As an example we consider a large, homogeneous object with a small inclusion away from the outer object boundary. If this outer boundary is found, the shape gradient will be very small because a local change of the existing boundaries would not result in a decrease of the functional. In this situation we use the insert strategy from [33]: check if the functional gradient of J has a significant maximum or minimum and insert a new set Ω_{m+1} in the vicinity of this extremum.

Here, we use the derivative of the data fit term $D_{\mathrm{KL}}(g \| A_\mu f)$ with respect to $f = \mathcal{F}(\vec{c}, \vec{D})$ as indicator for a change in topology. We have

$$\partial_f D_{\mathrm{KL}}(g \| A_\mu f) = A_\mu^* \left(\mathbf{1} - \frac{g}{A_\mu f} \right). \tag{4.8}$$

For a numerical implementation we use a Gaussian filter to smooth (4.8) and compare the smoothed functional derivative with its mean. We say $\partial_f D_{\mathrm{KL}}(g \| A_\mu f)$ has a significant extremum if the maximum or minimum deviates from the mean by more than three times the standard deviation. When the functional derivative has a significant extremum, we insert a new component in its vicinity. A typical choice of the inserted component is a small circle centering at the extremum with a specific radius, which is small enough so that it would not intersect any existing boundaries. After that we restart the level-set evolution and proceed according to the alternating minimization steps.

4.6 Step 6: Stopping criteria and parameter choice rules

The iteration is stopped after a pre-defined number of iterations. Parameters are chosen empirically. Theoretically motivated stopping criteria and parameter choice rules are subject of further research.

5 Numerical results

In this section we apply the presented approach of simultaneous segmentation and reconstruction to simulated tomography data with Poisson statistics. In order to generate SPECT sinograms with Poisson noise of different levels we apply the MATLAB routine `imnoise`. For a double-precision image, the routine `imnoise` generates Poisson noise from the data as follows: The input pixel values are interpreted as means of Poisson distributions scaled up by 10^{12}. For example, if an input pixel has the value 5.5×10^{-12}, then the corresponding output pixel will be generated from a Poisson distribution with mean of 5.5 and then scaled back down by multiplying 10^{-12}. According to [6], the factor 10^{12} is fixed and denotes the maximal number of detectable photons. Thus in order to generate sinograms with different amounts of noise we rescale the initial sinogram with a proper factor before applying imnoise and then scale it back with the same factor. Using the rescale factor can be interpreted in two ways: keeping the concentration of the radiopharmaceutical constant and varying the exposure time, and keeping the exposure time constant and varying the concentration of the radiopharmaceutical. Both interpretations result in a different amount of captured/counted photons.

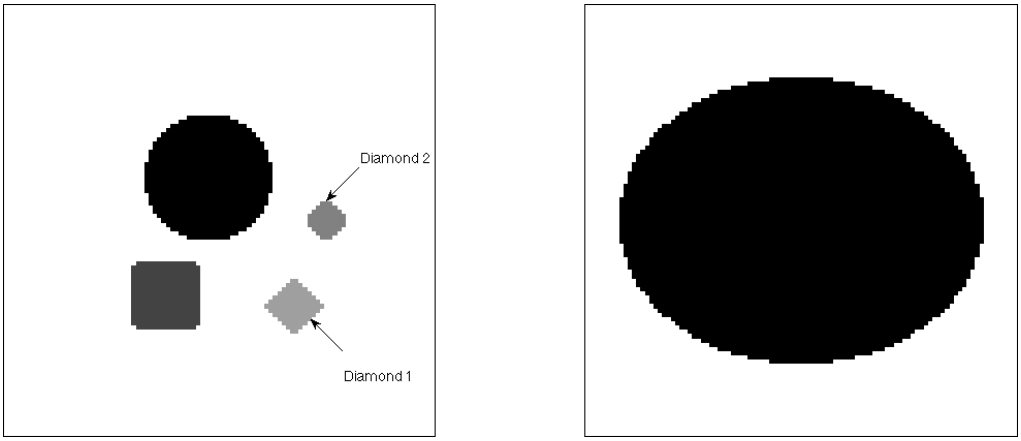

Figure 1. Test phantom: (Left) Four objects with different regular shapes and photon intensities, see Table 1; (Right) density μ.

As our first example we choose a phantom f that consists of four objects with different regular shapes (Figure 1, left). We assign to each of the four objects a value indicating the photon intensity, see Table 1. As density function μ we choose an ellipsoid with constant value 1 that includes all objects (Figure 1, right). SPECT data from this phantom are shown in Figure 2: the expectation $A_\mu f$ together with Poisson distributed data using the MATLAB routine `imnoise` and scale factors 10^9, 10^{10}, 10^{11}. Figure 2 indicates that the Poisson distributed data gets closer to its expectation $A_\mu f$ (and by that the deterministic model) the smaller the scale

Scale factor	Algorithm	Background	Square	Circle	Diamond 1	Diamond 2
	Ground truth	0	6.0000e-2	8.0000e-2	3.0000e-2	4.0000e-2
10^9	L^2 method	3.6000e-5	5.9992e-2	8.0194e-2	3.3287e-2	4.4749e-2
	KL-L^2 method	9.0422e-15	6.0698e-2	8.0090e-2	3.0495e-2	4.2159e-2
10^{10}	L^2 method	4.2000e-5	6.1282e-2	7.9824e-2	3.5050e-2	4.5392e-2
	KL-L^2 method	1.3773e-11	6.2597e-2	8.0264e-2	3.2378e-2	4.1810e-2
10^{11}	L^2 method	0	5.9101e-2	7.8531e-2	4.2231e-2	4.3856e-2
	KL-L^2 method	5.3112e-14	6.3027e-2	7.8189e-2	3.9621e-2	4.9993e-2

Table 1. Ground truth and reconstructed coefficient values from noisy data with scale factor = 10^9, 10^{10} and 10^{11}.

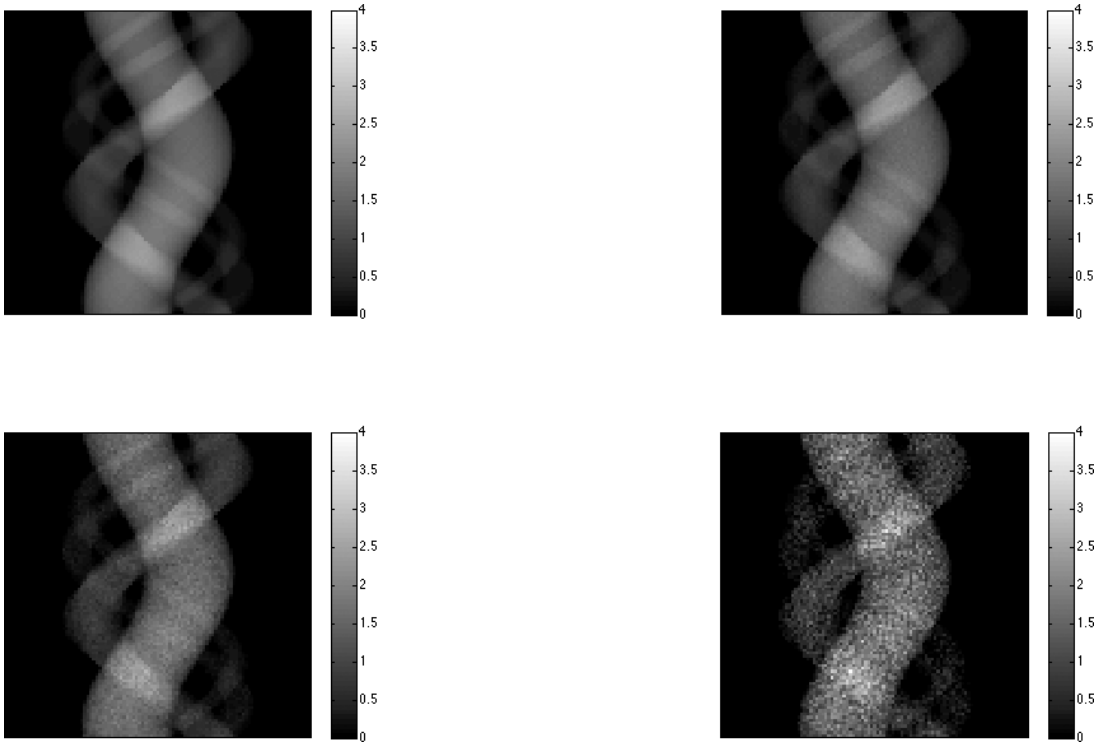

Figure 2. SPECT sinogram of artificial shapes with different level of noise: (Upper Left) expectation $A_\mu f$; (Upper Right) scale factor = 10^9, percentage of L^1-norm relative error = 1.98 %; (Lower Left) scale factor = 10^{10}, percentage of L^1-norm relative error = 6.15 %; (Lower Right) scale factor = 10^{11}, percentage of L^1-norm relative error = 19.96 %.

factor. As has been said, the scale factor influences the amount of captured photons. The larger the scale factor the less photons are captured. Whereas a scale factor of 10^9 results in a maximal amount of $10^{12}/10^9 = 10^3$ countable photons, a scale factor of 10^{10} only allows for 10^2 countable photons. Thus decreasing the scale factor results in a higher photon count which increases the signal quality and approaches the deterministic model. As a complement to the scale factor, we use the L^1-norm relative error $\|g^\delta - g\|_{L^1}/\|g\|_{L^1}$ to measure the noises in the data with Poisson statistics, since the expectation g and the measured data g^δ are in the Lebesgue space L^1.

As observed from the numerical experiments, the wrong assignment of a single pixel (where the boundary is passing through this pixel) can already result in severe outliers or undefined values (∞) for the shape derivative. Especially inclusions of zero intensity (as in the MCAT heart phantom example) can cause a breakdown of the whole process. Also, if one were to start with a very good approximation of the actual shape of the activity distribution that is simply located false (say, with an offset) the Kullback–Leibler direction would not be well-defined. In order to remedy this we suggest a **combined Kullback–Leibler-L^2 (KL-L^2) direction**.

Rewriting the descent direction (4.8) for the Kullback–Leibler data fit term and comparing it with the one for the standard L_2 data fit term with a general linear operator K (see [22, 33]) yields

$$F_{\mathrm{KL}} = -\sum_{i=1}^{m} c_i s_i \left(K^* \left(\frac{Kf - g}{Kf} \right) \right)(\boldsymbol{x}) + \cdots ,$$
$$F_{L_2} = -\sum_{i=1}^{m} c_i s_i (K^*(Kf - g))(\boldsymbol{x}) + \cdots .$$

Thus, the Kullback–Leibler descent direction differs from the L^2-descent direction in that, before backprojection, the data residual $Kf - g$ is divided by Kf. Hence, we propose to use the L^2 direction whenever the Kullback–Leibler direction is not well-defined.

In the following we present pictures of reconstructions, Figures 3–5, and segmentations, Figure 6, as well as a table with the reconstructed activity values for the segmented sets, Table 1. To access the quality of the reconstruction, we use two quality measures, namely the *PSNR* (Peak Signal-to-Noise Ratio) and the *MSSIM* (Mean Structural SIMilarity index) in Table 2.

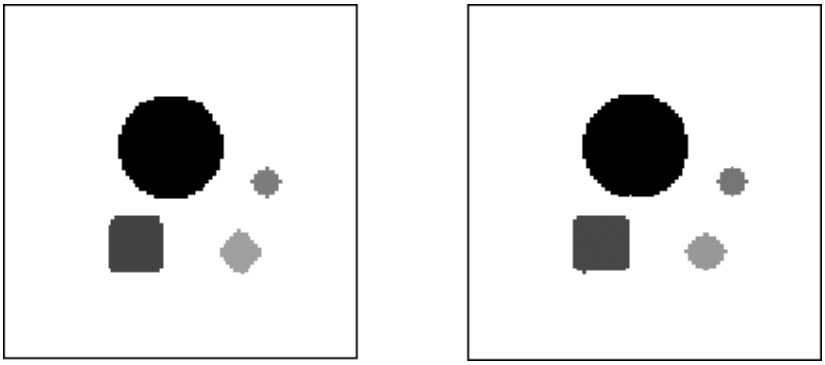

Figure 3. Reconstructed shapes by KL-L^2 method (Left) and L^2 method (Right) for scale factor 10^9, percentage of L^1-norm relative error = 1.98 %: (Left) KL-L^2 method with $\alpha = \beta = 10^{-14}$, MSSIM = 0.9884, PSNR = 32.4036; (Right) L^2 method with $\alpha = \beta = 10^{-14}$, MSSIM = 0.9847, PSNR = 32.4326;

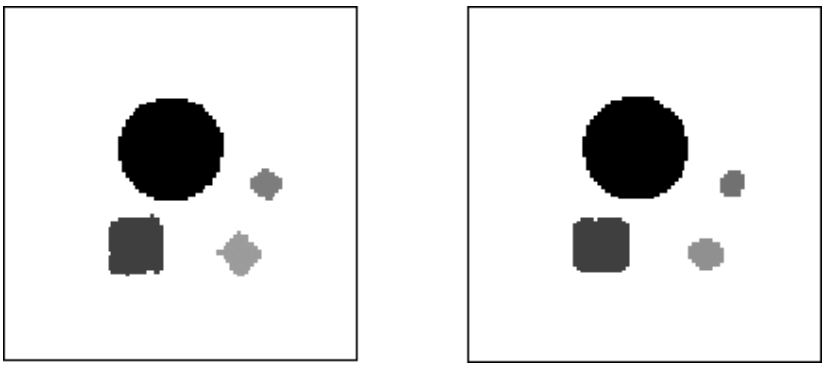

Figure 4. Reconstructed shapes by KL-L^2 method (Left) and L^2 method (Right) for scale factor 10^{10}, percentage of L^1-norm relative error = 6.15 %: (Left) KL-L^2 method with $\alpha = \beta = 10^{-13}$, MSSIM = 0.9846, PSNR = 29.4648; (Right) L^2 method with $\alpha = \beta = 10^{-13}$, MSSIM = 0.9661, PSNR = 27.9993;

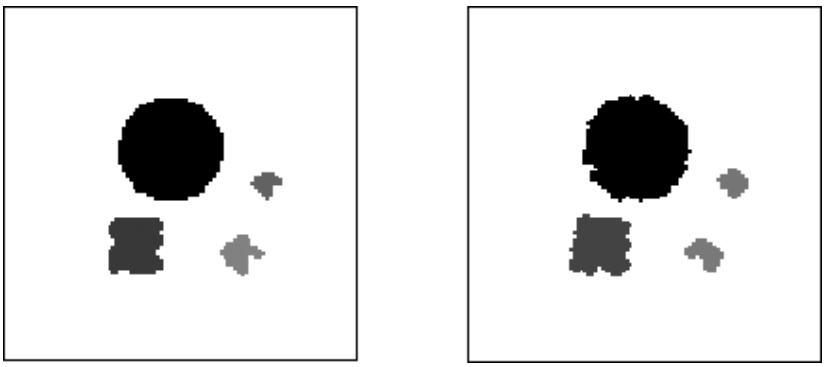

Figure 5. Reconstructed shapes by KL-L^2 method (Left) and L^2 method (Right) for scale factor 10^{11}, percentage of L^1-norm relative error = 19.96 %: (Left) KL-L^2 method with $\alpha = \beta = 10^{-11}$, MSSIM = 0.9642, PSNR = 25.7312; (Right) L^2 method with $\alpha = \beta = 10^{-11}$, MSSIM = 0.9280, PSNR = 21.9164;

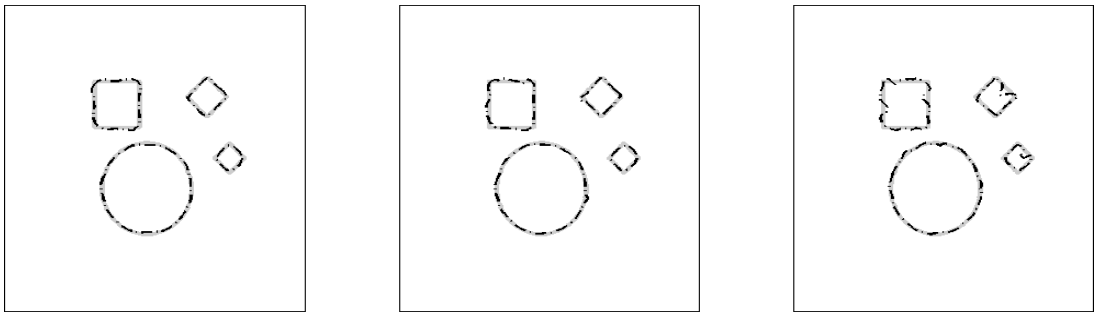

Figure 6. Reconstructed contours from data perturbed by different noise levels, using KL-L^2 method: (Left) for scale factor 10^9, percentage of L^1-norm relative error = 1.98 %; (Middle) for scale factor 10^{10}, percentage of L^1-norm relative error = 6.15 %; (Right) for scale factor 10^{11}, percentage of L^1-norm relative error = 19.96 %; original contour (grey, solid line); reconstructed contour (black, dash-dotted line).

Table 1 presents the true values for the activity coefficients (ground truth) as well as the reconstructed values for both the Kullback–Leibler and the L^2-data fit term. For both methods the coefficients are recovered sufficiently well. There is no clear indication that one method outperforms the other. The coefficients depend on the reconstructed sets. If a set is too small, the corresponding coefficient is too large and vice versa. This behavior can be observed clearly for the two smallest sets, diamond 1 and 2, when compared to their respective geometrical reconstructions given in Figures 3–5. The length penalty results in a regularized solution, prevents oscillating boundaries but also subdues smaller sets. One can see that his behavior acts out with increasing scale factor, corresponding to a lower quality of the noisy sinogram data.

To assess the quality of the overall reconstruction (depending on the geometry as well as on the coefficients), we use two measures, namely the *peak signal-to-noise ratio (PSNR)* and the *MSSIM* (Mean Structural SIMilarity index), a recently emerging index to perceive the image degradation as change in structural information, and assess the overall quality of simultaneous reconstruction and segmentation. The *peak signal-to-noise ratio (PSNR)* is widely used as a measure of subjective quality assessment and based on 'a byte-by-byte comparison of the data without considering what they actually represent' as the author argued in [45]. In order to overcome this disadvantage and get a quantitative comparison the *SSIM index* (Structural SIMilarity) [43] is an alternative.

Definition 5.1 (SSIM). Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two nonnegative image signals. The *SSIM index* of $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined as

$$\mathrm{SSIM}(\boldsymbol{x}, \boldsymbol{y}) := \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)},$$

where μ_* and σ_*^2 denote the mean and variance of $\boldsymbol{x}$ and $\boldsymbol{y}$, and σ_{xy} denotes the covariance of $\boldsymbol{x}$ and $\boldsymbol{y}$. The positive constants C_i, $i = 1, 2$, stabilize the denominator.

The SSIM index can be applied globally, resulting in one value for the image signals $\boldsymbol{x}$ and $\boldsymbol{y}$, or locally, taking into account that statistical features of images such as the mean and the standard deviation are usually not stationary in space. Thus, the local statistics μ_*, σ_* and σ_{xy} are computed within a window, e.g., an 8×8 area, which moves over the entire image. The SSIM index then becomes a map of indices. As we would like to have a single overall quality measure of the similarities of two images, we use the mean SSIM (MSSIM) index as introduced in [43],

$$\mathrm{MSSIM}(\boldsymbol{X}, \boldsymbol{Y}) := \frac{1}{M}\sum_{j=1}^{M} \mathrm{SSIM}(\boldsymbol{x}_j, \boldsymbol{y}_j),$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are the reference and the distorted images, respectively; $\boldsymbol{x}_j$ and $\boldsymbol{y}_j$ are the image contents at the jth local window; and M is the number of local windows. We use the MATLAB command `ssim` to estimate this index, which uses a default 11×11 window size.

Since we consider not only the values assigned to each domain but also the geometry of each domain, the MSSIM index seems to be a plausible measure of similarity. The MSSIM indices and the PSNR values are

Scale factor	10^9		10^{10}		10^{11}	
	L^2	KL-L^2	L^2	KL-L^2	L^2	KL-L^2
PSNR	32.4326	32.4036	27.9993	29.4648	21.9164	25.7312
MSSIM	0.9847	0.9984	0.9661	0.9846	0.9280	0.9642

Table 2. PSNR and MSSIM indices for different rescale factor = 10^9, 10^{10} and 10^{11}.

provided in Table 2 as well as in Figures 3–5 together with the reconstruction for both the L^2 and the KL method. Table 2 demonstrates that for both PSNR and MSSIM index the KL method obtains a larger quantity and accordingly a better reconstruction quality for each scale factor. Please note that the parameters here are chosen empirically for the KL-L^2 method to give a better performance. The experiment results also show that the L^2 method are extremely sensitive to the parameter choice for data with Poisson statistics, and the near-optimal reconstructions are obtained with parameters of the similar magnitude with those for the KL-L^2 method.

The MCAT-phantom

As second example we consider the standard *mathematical cardiac torso* (MCAT) phantom (see Figure 7 and [42, 44]). The piecewise constant activity function f represents a section through a simplified model of a human heart. The activity function depicts the blood supply of the myocardial muscle and two ventricles. In particular, it shows the clinical situation that the blood supply is interrupted, resulting in a gap in the outer left myocardial muscle. The left ventricle seems to be open at this point since the radiopharmaceuticum cannot reach this area. Hence, the background or outside area and the left ventricle are modeled as one connected domain. The value for the myocardial muscle is 10^{-6}. The density is assumed to be known and modelled as section through a simplified version of the human torso, with different values for the spine, spinal canal, left (or right) lung and body area. Figure 8 shows SPECT data of the MCAT phantom, perturbed by different levels of Poisson noise. From top to bottom and left to right, the noise level grows from a mild to a severely disturbed data set. For a moderately perturbed case (scale factor 10^4, Figure 8, left) the result of reconstruction and segmentation is given in Figure 9.

With the empirically chosen penalty parameters α and β our algorithm can generate a reasonable reconstruction and segmentation of the MCAT heart phantom. For a better visualization a zoom-in of the reconstructed contour is provided in Figure 9 (right). We would like to point out that especially the important feature, namely the gap in the myocardial muscle where the blood flow is interrupted is captured in the reconstruction. Table 3 provides the reconstructed values for the coefficients as well as the PSNR and the MSSIM index.

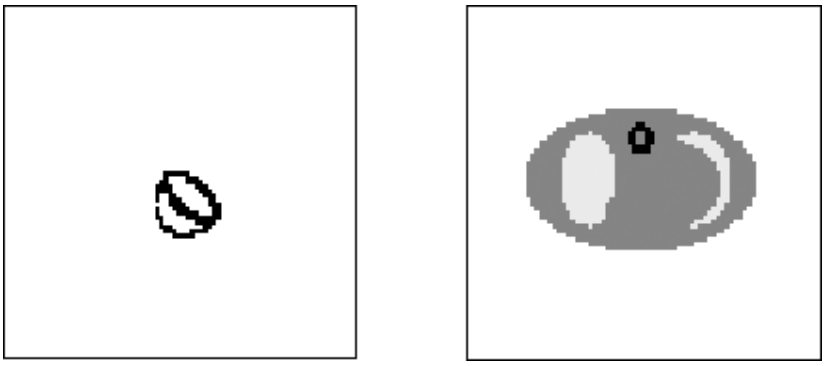

Figure 7. Standard MCAT phantom: (Left) activity f; (Right) density μ.

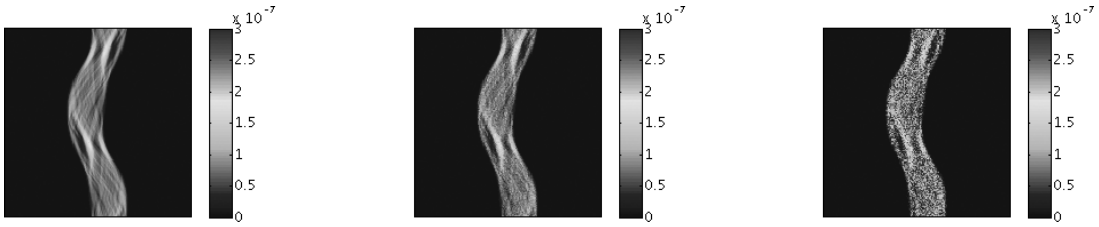

Figure 8. SPECT sinogram of MCAT with different level of noises: (Left) theoretic expectation numbers of captured photons; (Middle) scale factor = 10^3, percentage of L^1-norm relative error = 8.97 %; (Right) scale factor = 10^4, percentage of L^1-norm relative error = 28.31 %.

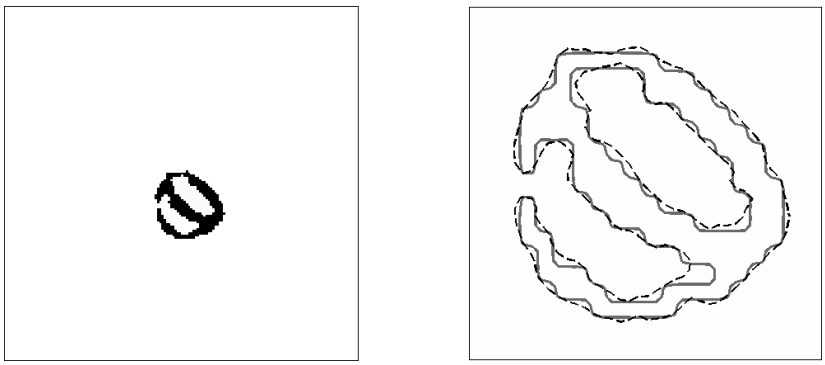

Figure 9. Reconstruction of MCAT phantom from SPECT data perturbed with moderate Poisson noise (scale factor = 10^4, percentage of L^1-norm relative error = 28.31 %), chosen parameter $\alpha = \beta = 10^{-12}$, MSSIM = 0.9890, PSNR = 25.9511: (Left) reconstructed heart phantom; (Right) zoom-in of the contour of reconstructed heart phantom: original contour (solid, grey), reconstructed contour
(dash line, black).

	left ventricle	right ventricle	myocardial	MSSIM	PSNR
ground truth	0	0	1e-6	1	∞
L^2-KL	2.2204e-16	5.8078e-8	9.1291e-7	0.9889	25.8012

Table 3. MCAT reconstruction: coefficient values, SSIM and PSNR index for noisy data with scale factor = 10^4.

Despite the very good match of the true contours and our segmentation we note that the coefficient value we have obtained for the right ventricle is too large. The zoom into the reconstructed heart phantom, Figure 9 (right), shows that the reconstructed myocardial muscle is a little bit too large as some details are smoothed out. The coefficient assigned to the myocardial muscle is a little too small (9.1291e-7 instead of 1e-6) which corresponds to the previously described compensation effect: too large sets go hand in hand with too small coefficients and vice versa. However, for the right ventricle, although the area of the reconstructed set might be a little too small the value of the reconstructed coefficient seems to be quite a bit too large. Hence, we considered the situation that the exact geometry is known and only computed the coefficients for data sets

	background & left ventricle	right ventricle	myocardial
ground truth	0	0	1e-6
scale factor 10^2	0	0	9.9973e-7
scale factor 10^3	0	2.4767e-9	9.9811e-7
scale factor 10^4	3.5527e-14	3.5143e-8	1.0816e-6

Table 4. MCAT reconstruction: coefficient values for fixed (true) geometry for noisy data with scale factor 10^2, 10^3 and 10^4.

with different noise levels, see Table 4. This experiment demonstrates the influence of the noise on the data. Whereas for a large scaling factor (10^2) the values of all sets are computed very accurately the result starts to get worse for increasing noise. Especially the value of the right ventricle seems to be affected by the increasingly noisy data.

Conclusion

In medical emission tomography, the collected data of photon counts is a realization of random variables, which in many cases are properly modeled by a Poisson distribution. In this paper, we have proposed a Mumford–Shah-type functional for the simultaneous reconstruction and segmentation of the activity distribution from tomographic data with Poisson statistics. We described and implemented an alternating minimization algorithm for the corresponding constrained minimization problem. The shape derivative and the gradient projection type method are used to update the geometry and functional variables respectively, while the level-set method is used to realize the evolution of the geometry. An intuitive insert scheme is provided to prevent that the algorithm stops prematurely in a local minimum. Numerical experiments with simulated data suggest that the proposed algorithm attains satisfactory reconstruction and segmentation of the activity distribution.

Funding: The work of E. Klann has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement no. 600209 (TU Berlin – IPODI) as well as from the Austrian Science Funds (FWF) under grant T 529-N18 and the MPNS COST Action MP1207. The work of R. Ramlau and P. Sun is supported by DK (Doktoratskolleg) program "Computational Mathematics" (W1214) granted by Austrian Science Funds (FWF). The work of P. Sun was also supported by CSC-FWF joint program.

References

[1] G. Alessandrini and S. Vessella, Lipschitz stability for the inverse conductivity problem, *Adv. in Appl. Math.* **35** (2005), no. 2, 207–241.

[2] L. Ambrosio, N. Fusco and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Clarendon Press, Oxford, 2000.

[3] J. M. Bardsley and C. R. Vogel, A nonnegatively constrained convex programming method for image reconstruction, *SIAM J. Sci. Comput.* **25** (2004), no. 4, 1326–1343.

[4] M. Benning, T. Kosters, F. Wubbeling, K. Schafers and M. Burger, A nonlinear variational method for improved quantification of myocardial blood flow using dynamic $H^2{}_{15}O$ PET, in: *Nuclear Science Symposium Conference Record* (NSS '08), IEEE Press, Piscataway (2008), 4472–4477.

[5] T. Brox and D. Cremers, On the statistical interpretation of the piecewise smooth Mumford–Shah functional, in: *Scale Space and Variational Methods in Computer Vision*, Lecture Notes in Comput. Sci. 4485, Springer, Berlin (2007), 203–213.

[6] M. Burger, J. Müller, E. Papoutsellis and C. B. Schönlieb, Total variation regularization in measurement and image space for PET reconstruction, *Inverse Problems* **30** (2014), no. 10, Article ID 105003.

[7] F. Cannizzaro, G. Greco, S. Rizzo and E. Sinagra, Results of the measurements carried out in order to verify the validity of the poisson-exponential distribution in radioactive decay events, *Int. J. Appl. Radiation Isotopes* **29** (1978), DOI 10.1016/0020-708X(78)90101-1.

[8] M. Delfour and J. Zolésio, *Shapes and Geometries*, 2nd ed., SIAM, Philadelphia, 2011.

[9] A. P. Dempster, N. M. Laird and D. B. Rubin, Maximum likelihood from incomplete data via the em algorithm, *J. R. Stat. Soc. Ser. B.* **39** (1977), no. 1, 1–38.

[10] M. Droske and W. Ring, A Mumford–Shah level-set approach for geometric image registration, *SIAM J. Appl. Math.* **66** (2006), no. 6, 2127–2148.

[11] I. Ekeland and R. Témam, *Convex Analysis and Variational Problems*, Class. Appl. Math. 28, SIAM, Philadelphia, 1999.

[12] H. W. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Math. Appl. (Dordrecht) 375, Springer, Dordrecht, 1996.
[13] S. Geman and D. McClure, Bayesian image analysis: An application to single photon emission tomography, *Proc. Statist. Comput. Sec. Amer. Statist. Assoc.* (1985), 12–18.
[14] S. A. Geman and D. E. McClure, *Statistical Methods for Tomographic Image Reconstruction*, MIT, Cambridge, 1987.
[15] G. Gerig, O. Kubler, R. Kikinis and F. A. Jolesz, Nonlinear anisotropic filtering of mri data, *IEEE Trans. Med. Imaging* **11** (1992), no. 2, 221–232.
[16] M. Hintermüller and W. Ring, An inexact newton-cg-type active contour approach for the minimization of the Mumford–Shah functional, *J. Math. Imaging Vision* **20** (2004), no. 1–2, 19–42.
[17] E. Jonsson, S.-C. Huang and T. Chan, Total-variation regularization in positron emission tomography, preprint (1998).
[18] J. Kay, Statistical models for PET and SPECT data, *Stat. Methods Med. Res.* **3** (1994), no. 1, 5–21.
[19] C. T. Kelley, *Iterative Methods for Optimization*, Front. Appl. Math., SIAM, Philadelphia, 1999.
[20] E. Klann, A Mumford–Shah-like method for limited data tomography with an application to electron tomography, *SIAM J. Imaging Sci.* **4** (2011), no. 4, 1029–1048.
[21] E. Klann and R. Ramlau, Regularization properties of Mumford–Shah-type functionals with perimeter and norm constraints for linear ill-posed problems, *SIAM J. Imaging Sci.* **6** (2013), no. 1, 413–436.
[22] E. Klann, R. Ramlau and W. Ring, A Mumford–Shah level-set approach for the inversion and segmentation of SPECT/CT data, *Inverse Probl. Imaging* **5** (2011), no. 1, 137–166.
[23] T. Le, R. Chartrand and T. J. Asaki, A variational approach to reconstructing images corrupted by Poisson noise, *J. Math. Imaging Vision* **27** (2007), no. 3, 257–263.
[24] S.-J. Lee, A. Rangarajan and G. Gindi, Bayesian image reconstruction in spect using higher order mechanical models as priors, *IEEE Trans. Med. Imaging* **14** (1995), no. 4, 669–680.
[25] J. J. Moré and G. Toraldo, On the solution of large quadratic programming problems with bound constraints, *SIAM J. Optim.* **1** (1991), no. 1, 93–113.
[26] H. N. Mülthei, B. Schorr and W. Törnig, On an iterative method for a class of integral equations of the first kind, *Math. Methods Appl. Sci.* **9** (1987), no. 1, 137–168.
[27] D. Mumford and J. Shah, Optimal approximations by piecewise smooth functions and associated variational problems, *Comm. Pure Appl. Math.* **42** (1989), no. 5, 577–685.
[28] F. Natterer, *The Mathematics of Computerized Tomography*, Class. Appl. Math., SIAM, Philadelphia, 2001.
[29] J. Nocedal and S. Wright, *Numerical Optimization*, Springer, New York, 2006.
[30] S. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*, Appl. Math. Sci. 153, Springer, New York, 2003.
[31] S. Osher and J. A. Sethian, Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton–Jacobi formulations, *J. Comput. Phys.* **79** (1988), no. 1, 12–49.
[32] E. T. Quinto, An introduction to X-ray tomography and Radon transforms, in: *The Radon Transform, Inverse Problems, and Tomography* (Atlanta 2005), Proc. Sympos. Appl. Math. 63, American Mathematical Society, Providence (2006), 1–23.
[33] R. Ramlau and W. Ring, A Mumford–Shah level-set approach for the inversion and segmentation of X-ray tomography data, *J. Comput. Phys.* **221** (2007), no. 2, 539–557.
[34] R. Ramlau and W. Ring, Regularization of ill-posed Mumford–Shah models with perimeter penalization, *Inverse Problems* **26** (2010), no. 11, Article ID 115001.
[35] E. Resmerita, H. W. Engl and A. N. Iusem, *The expectation-maximization algorithm for ill-posed integral equations: A convergence analysis*, *Inverse Problems* **23** (2007), no. 6, Article ID 2575.
[36] N. Roé-Vellvé, F. Pino, C. Falcon, A. Cot, J. D. Gispert, C. Marin, J. Pavía and D. Ros, Quantification of rat brain spect with 123i-ioflupane: Evaluation of different reconstruction methods and image degradation compensations using Monte Carlo simulation, *Phys. Med. Biol.* **59** (2014), no. 16, Article ID 4567.
[37] L. I. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Phys. D* **60** (1992), no. 1–4, 259–268.
[38] A. Sawatzky, C. Brune, F. Wubbeling, T. Kosters, K. Schafers and M. Burger, Accurate EM-TV algorithm in PET with low SNR, in: *Nuclear Science Symposium Conference Record* (NSS '08), IEEE Press, Piscataway (2008), 5133–5137.
[39] O. Scherzer, *Handbook of Mathematical Methods in Imaging*, Springer, Berlin, 2011.
[40] L. A. Shepp and Y. Vardi, Maximum likelihood reconstruction for emission tomography, *IEEE Trans. Med. Imaging* **1** (1982), no. 2, 113–122.
[41] J. Sokolowski and J. P. Zolesio, *Introduction to Shape Optimization: Shape Sensitivity Analysis*, Springer Ser. Comput. Math., Springer, Berlin, 2012.
[42] J. A. Terry, B. M. W. Tsui, J. R. Perry, J. L. Hendricks and G. T. Gullberg, The design of a mathematical phantom of the upper human torso for use in 3-d spect imaging research, in: *Biomedical Engineering: Opening New Doors*, New York University Press, New York (1990), 185–190.
[43] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, Image quality assessment: From error visibility to structural similarity, *IEEE Trans. Image Process.* **13** (2004), no. 4, 600–612.

[44] M. N. Wernick and J. N. Aarsvold, *Emission Tomography: The Fundamentals of PET and SPECT*, Elsevier Science, Amsterdam, 2004.

[45] S. Winkler and P. Mohandas, The evolution of video quality measurement: From psnr to hybrid metrics, *IEEE Trans. Broadcasting* **54** (2008), no. 3, 660–668.

[46] M. Yan, A. A. T. Bui, J. Cong and L. A. Vese, General convergent expectation maximization (EM)-type algorithms for image reconstruction, *Inverse Probl. Imaging* **7** (2013), no. 3, 1007–1029.