

Tomography: mathematical aspects and applications

Paolo Facchi^{1,2}, Marilena Ligabò³ and Sergio Solimini³

¹Dipartimento di Fisica and MECENAS, Università di Bari, I-70126 Bari, Italy

²INFN, Sezione di Bari, I-70126 Bari, Italy

³Dipartimento di Meccanica, Matematica e Management, Politecnico di Bari, I-70125 Bari, Italy

E-mail: marilena.ligabo@poliba.it

Received 4 November 2014

Accepted for publication 16 December 2014

Published 11 June 2015



Abstract

In this article we present a review of the Radon transform and the instability of the tomographic reconstruction process. We show some new mathematical results in tomography obtained by a variational formulation of the reconstruction problem based on the minimization of a Mumford–Shah type functional. Finally, we exhibit a physical interpretation of this new technique and discuss some possible generalizations.

Keywords: Radon transform, image segmentation, calculus of variations, integral geometry

(Some figures may appear in colour only in the online journal)

1. Introduction

The primary goal of tomography is to determine the internal structure of an object without cutting it, namely using data obtained by methods that leave the object under investigation undamaged. These data can be obtained by exploiting the interaction between the object and various kinds of probes including x-rays, electrons, and many others. After its interaction with the object under investigation, the probe is detected to produce what we call a projected distribution or tomogram, see figure 1.

Tomography is a rapidly evolving field for its broad impact on issues of fundamental nature and for its important applications such as the development of diagnostic tools relevant to disparate fields, such as engineering, biomedical and archaeometry. Moreover, tomography can be a powerful tool for many reconstruction problems coming from many areas of research, such as imaging, quantum information and computation, cryptography, lithography, metrology and many others, see figure 2.

From the mathematical point of view the reconstruction problem can be formulated as follows: one wants to recover an unknown function through the knowledge of an appropriate family of integral transforms. It was proved by Radon [1] that a smooth function $f(x, y)$ on \mathbb{R}^2 can be determined explicitly by means of its integrals over the lines in \mathbb{R}^2 .

Let $\mathcal{R}f(X, \theta)$ denote the integral of f along the line $x \cos \theta + y \sin \theta = X$ (tomogram). Then

$$f(x, y) = (-\Delta)^{\frac{1}{2}} \int_0^{2\pi} \mathcal{R}f(x \cos \theta + y \sin \theta, \theta) \frac{d\theta}{4\pi}, \quad (1)$$

where $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the Laplacian on \mathbb{R}^2 , and its square root is defined by Fourier transform (see theorem 1). We now observe that the formula above has built in a remarkable duality: first one integrates over the set of points in a line, then one integrates over the set of lines passing through a given point. This formula can be extended to the N -dimensional case by computing the integrals of the function f on all possible hyperplanes. This suggests to consider the transform $f \mapsto \mathcal{R}f$ defined as follows. If f is a function on \mathbb{R}^N then $\mathcal{R}f$ is the function defined on the space of all possible $(N-1)$ -dimensional planes in \mathbb{R}^N such that, given a hyperplane λ , the value of $\mathcal{R}f(\lambda)$ is given by the integral of f along λ . The function $\mathcal{R}f$ is called *Radon transform* of f .

There exist several important generalizations of the Radon transform by John [2], Gel'fand [3], Helgason [4] and Strichartz [5]. More recent analysis has been boosted by Margarita and Volodya Man'ko and has focused on symplectic transforms [6], on the deep relationship with classical systems and classical dynamics [7–9], on the formalism of star product quantization [10–12], and on the study of marginals along curves that are not straight lines [13, 14].

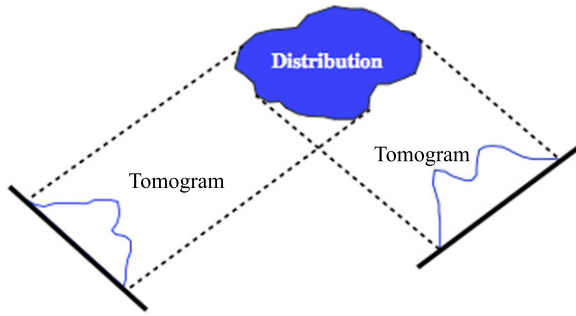


Figure 1. Distribution and two tomograms.

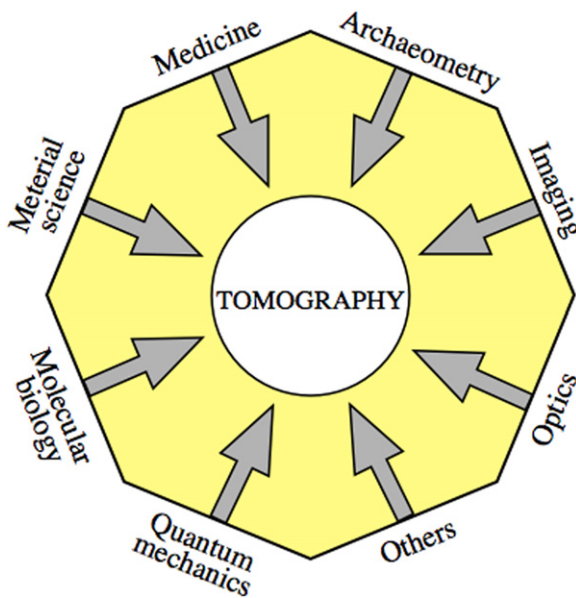


Figure 2. Reconstruction problems from diverse fields may be united within the framework of tomography.

In quantum mechanics the Radon transform of the Wigner function [15–17] was considered in the tomographic approach to the study of quantum states [18, 19] and experimentally realized with different particles and in diverse situations. For a review on the modern mathematical aspects of classical and quantum tomography see [20]. Good reviews on recent tomographic applications can be found in [21] and in [22], where particular emphasis is given on maximum likelihood methods, that enable one to extract the maximum reliable information from the available data.

As explained above, from the mathematical point of view, the internal structure of the object is described by an unknown function f (density), that is connected via an operator to some measured quantity g (tomograms). The tomographic reconstruction problem can be stated as follows: for given data g , the task is to find f from the operator equation $\mathcal{R}f = g$. There are many problems related to the implementation of effective tomographic techniques due to the instability of the reconstruction process. There are two principal reasons of this instability. The first one is the ill-posedness of the reconstruction problem: in order to obtain a

satisfactory estimate of the unknown function it is necessary an extremely precise knowledge of its tomograms, which is in general physically unattainable [23]. The second reason is the discrete and possibly imperfect nature of data that allows to obtain only an approximation of the unknown function.

The first question is whether a partial information still determines the function uniquely. A negative answer is given by a theorem of Smith *et al* [24], that states: ‘a function f with compact support in the plane is uniquely determined by any infinite set, but by no finite set of its tomograms’. Therefore, it is clear that one has to abandon the request of uniqueness in the applications of tomography. Thus, due to the ill-posedness of reconstruction problem and to the loss of uniqueness in the inversion process, a regularization method has to be introduced to stabilize the inversion.

A powerful approach is the introduction of a Mumford–Shah (MS) functional, first introduced in a different context for image denoising and segmentation [25]. The main motivation is that, in many practical applications, one is not only interested in the reconstruction of the density distribution f , but also in the extraction of some specific features or patterns of the image. An example is the problem of the determination of the boundaries of inner organs. By minimizing the MS functional, one can find not only (an approximation of) the function but also its sharp contours. Very recently a MS functional for applications to tomography has been introduced in the literature [26–28]. Some preliminary results in this context are already available but there are also many interesting open problems and promising results in this direction, as we will try to explain in the second part of this article.

The article is organized as follows. Section 2 contains a short introduction to the Radon transform, its dual map and the inversion formula. Section 3 is devoted to a brief discussion on the ill-posedness of the tomographic reconstruction and to the introduction of regularization methods. In section 4 a MS functional is applied to tomography as a regularization method. In particular, in subsection 4.1 the piecewise constant model and known results are discussed together with a short list of some interesting open problems. Finally, in section 5 we present an electrostatic interpretation of the regularization method based on the MS functional, which motivates us to introduce an improved regularization method, based on the Blake–Zisserman functional [29], as a relaxed version of the previous one.

2. The Radon transform: definition and inversion formula

Consider a body in the plane \mathbb{R}^2 , and consider a beam of particles (neutrons, electrons, x-rays, etc) emitted by a source. Assume that the initial intensity of the beam is I_0 . When the particles pass through the body they are absorbed or scattered and the intensity of the beam traversing a length Δs decreases by an amount proportional to the density of the body μ ,

namely

$$\Delta I/I = -\Delta s \mu(s), \quad (2)$$

so that

$$I(s) = I_0 \exp\left(-\int_0^s \mu(r) dr\right). \quad (3)$$

A detector placed at the exit of the body measures the final intensity $I(s)$ and then from

$$-\ln \frac{I(s)}{I_0} = \int_0^s \mu(r) dr \quad (4)$$

one can record the value of the density μ integrated on a line. If another ray with a different direction is considered, with the same procedure one obtains the value of the integral of the density on that line.

The mathematical model of the above setup is the following: given a smooth function $f(x)$ on the plane, $x \in \mathbb{R}^2$, and a line λ , consider its tomogram, given by

$$\mathcal{R}f(\lambda) = \int_{\lambda} f(x) dm(x), \quad (5)$$

where dm is the Euclidean measure on the line λ . In this way, we have defined an operator \mathcal{R} that maps a smooth function f on the plane \mathbb{R}^2 into a function $\mathcal{R}f$ on \mathbb{P}^2 , the manifold of the lines in \mathbb{R}^2 .

We ask the following question: if we know the family of tomograms $(\mathcal{R}f(\lambda))_{\lambda \in \mathbb{P}^2}$, can we reconstruct the density function f ? The answer is affirmative and in the following we will see how to obtain this result.

Let us generalize the above definitions to the case of an N -dimensional space. Let f be a function defined on \mathbb{R}^N , integrable on each hyperplane in \mathbb{R}^N and let \mathbb{P}^N be the manifold of all hyperplanes in \mathbb{R}^N . The Radon transform of f is defined by equation (5), where dm is the Euclidean measure on the hyperplane λ . Thus we have an operator \mathcal{R} , the *Radon transform*, that maps a function f on \mathbb{R}^N into a function $\mathcal{R}f$ on \mathbb{P}^N , namely $f \mapsto \mathcal{R}f$. Its dual transform, also called *back projection operator*, $g \mapsto \mathcal{I}g$ associates to a function g on \mathbb{P}^N the function $\mathcal{I}g$ on \mathbb{R}^N given by

$$\mathcal{I}g(x) = \int_{x \in \lambda} g(\lambda) d\mu(\lambda), \quad (6)$$

where $d\mu$ is the unique probability measure on the compact set $\{\lambda \in \mathbb{P}^N | x \in \lambda\}$ which is invariant under the group of rotations around x .

Let us consider the following covering of \mathbb{P}^N

$$\mathbb{R} \times \mathbb{S}^{N-1} \rightarrow \mathbb{P}^N, \quad (X, \xi) \mapsto \lambda, \quad (7)$$

where \mathbb{S}^{N-1} is the unit sphere in \mathbb{R}^N . Thus, the equation of the hyperplane λ is

$$\lambda = \left\{ x \in \mathbb{R}^N \mid X - \xi \cdot x = 0 \right\}, \quad (8)$$

with $a \cdot b$ denoting the Euclidean inner product of $a, b \in \mathbb{R}^N$. See figure 3. Observe that the pairs (X, ξ) ,

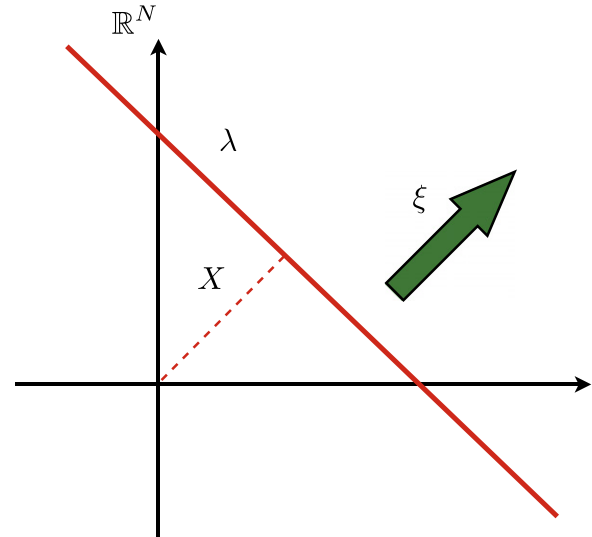


Figure 3. Parametrization of the hyperplane λ using its signed distance X from the origin and a unit vector ξ perpendicular to λ .

$(-X, -\xi) \in \mathbb{R} \times \mathbb{S}^{N-1}$ are mapped into the same hyperplane $\lambda \in \mathbb{P}^N$. Therefore (7) is a double covering of \mathbb{P}^N . Thus \mathbb{P}^N has a canonical manifold structure with respect to which this covering mapping is differentiable. We identify continuous (differentiable) functions g on \mathbb{P}^N with continuous (differentiable) functions g on $\mathbb{R} \times \mathbb{S}^{N-1}$ satisfying $g(X, \xi) = g(-X, -\xi)$.

We will momentarily work in the Schwartz space $\mathcal{S}(\mathbb{R}^N)$ of complex-valued rapidly decreasing functions on \mathbb{R}^N . In analogy with $\mathcal{S}(\mathbb{R}^N)$ we define $\mathcal{S}(\mathbb{R} \times \mathbb{S}^{N-1})$ as the space of C^∞ functions g on $\mathbb{R} \times \mathbb{S}^{N-1}$ which for any integers $m \geq 0$, any multiindex $\alpha \in \mathbb{N}^N$, and any differential operator D on \mathbb{S}^{N-1} satisfy

$$\sup_{X \in \mathbb{R}, \xi \in \mathbb{S}^{N-1}} \left| |X|^m \frac{\partial^\alpha (Dg)}{\partial X^\alpha}(X, \xi) \right| < +\infty. \quad (9)$$

The space $\mathcal{S}(\mathbb{P}^N)$ is then defined as the set of $g \in \mathcal{S}(\mathbb{R} \times \mathbb{S}^{N-1})$ satisfying $g(-X, -\xi) = g(X, \xi)$.

Now we want to obtain an inversion formula, namely we want to prove that one can recover a function f on \mathbb{R}^N from the knowledge of its Radon transform. In order to get this result we need a preliminary lemma, whose proof can be found in [20], which suggests an interesting physical interpretation.

Lemma 1. Let $f \in \mathcal{S}(\mathbb{R}^N)$ and $V(x) = 1/|x|$, $x \in \mathbb{R}^N$, $x \neq 0$. Then

$$\mathcal{I}(\mathcal{R}f)(x) = a_N f * V, \quad (10)$$

where a_N is a constant that depends only on the dimension N , and $*$ denotes the convolution product,

$$(f * g)(x) = \int_{\mathbb{R}^N} f(y) g(x - y) dy. \quad (11)$$

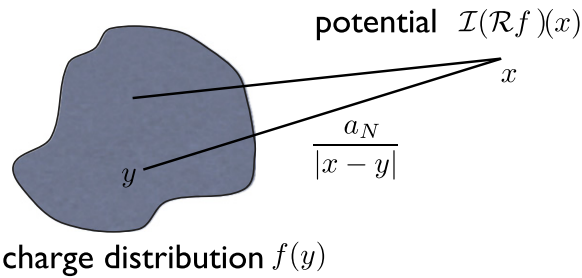


Figure 4. $\mathcal{I}(\mathcal{R}f)(x)$ is the potential at x generated by the charge distribution f .

A physical interpretation of lemma 1 is the following: if f is a charge distribution, then the potential at the point x generated by that charge is exactly $\mathcal{I}(\mathcal{R}f)(x)$, see figure 4. Notice, however, that the potential of a point charge scales always as the inverse distance *independently* of the dimension N , and thus it is Coulomb only for $N = 3$. The only dependence on N is in the strength of the elementary charge a_N . This fact is crucial: indeed, the associated Poisson equation involves an N -dependent (fractional) power of the Laplacian, which appears in the inversion formula for the Radon transform.

Theorem 1. Let $f \in \mathcal{S}(\mathbb{R}^N)$. Then

$$f(x) = \frac{1}{2(2\pi)^{N-1}} (-\Delta)^{\frac{N-1}{2}} \mathcal{I}(\mathcal{R}f)(x), \quad (12)$$

where $(-\Delta)^\alpha$, with $\alpha > 0$, is a pseudodifferential operator whose action is

$$((-\Delta)^\alpha f)(x) = \int_{\mathbb{R}^N} |k|^{2\alpha} \hat{f}(k) e^{ik \cdot x} \frac{dk}{(2\pi)^N}, \quad (13)$$

where \hat{f} is the Fourier transform of f ,

$$\hat{f}(k) = \int_{\mathbb{R}^N} f(x) e^{-ik \cdot x} dx. \quad (14)$$

The proof of theorem 1 can be found in [20]. Equation (12) says that, modulo the final action of $(-\Delta)^{\frac{N-1}{2}}$, the function f can be recovered from its Radon transform $\mathcal{R}f$ by the application of the dual mapping \mathcal{I} : first one integrates over the set of points in a hyperplane and then one integrates over the set of hyperplanes passing through a given point. Explicitly we get

$$f(x) = \frac{1}{2(2\pi)^{N-1}} (-\Delta)^{\frac{N-1}{2}} \int_{\mathbb{S}^{N-1}} \mathcal{R}f(\xi \cdot x, \xi) d\xi, \quad (15)$$

which has the following remarkable interpretation. Note that if one fixes a direction $\xi \in \mathbb{S}^{N-1}$, then the function $\mathcal{R}f(\xi \cdot x, \xi)$ is constant on each plane perpendicular to ξ , i.e. it is a (generalized) plane wave. Therefore, equation (15) gives a representation of f in terms of a continuous superposition of plane waves. A well-known analogous decomposition is given by Fourier transform. When $N = 2$,

one recovers the inversion formula (1) originally found by Radon [1].

3. Instability of the inversion formula with noisy data

We have defined the Radon transform of any function $f \in \mathcal{S}(\mathbb{R}^N)$ as $\mathcal{R}f$. The following theorem [4] contains the characterization of the range of the Radon linear operator \mathcal{R} and the extension of \mathcal{R} to the space of square integrable functions $L^2(\mathbb{R}^N)$.

Theorem 2. The Radon transform \mathcal{R} is a linear one-to-one mapping of $\mathcal{S}(\mathbb{R}^N)$ onto $\mathcal{S}_H(\mathbb{P}^N)$, where the space $\mathcal{S}_H(\mathbb{P}^N)$ is defined as follows: $g \in \mathcal{S}_H(\mathbb{P}^N)$ if and only if $g \in \mathcal{S}(\mathbb{P}^N)$ and for any integer $k \in \mathbb{N}$ the integral

$$\int_{\mathbb{R}} g(X, \xi) X^k dX \quad (16)$$

is a homogeneous polynomial of degree k in ξ_1, \dots, ξ_N . Moreover, the Radon operator \mathcal{R} can be extended to a continuous operator from $L^2(\mathbb{R}^N)$ to $L^2(\mathbb{P}^N)$.

In medical imaging, computerized tomography is a widely used technique for the determination of the density f of a sample from measurements of the attenuation of x-ray beams sent through the material along different angles and offsets. The measured data g are connected to the density f via the Radon transform \mathcal{R} . To compute the density distribution f the equation $g = \mathcal{R}f$ has to be solved. Unfortunately it is a well known fact that \mathcal{R} is not continuously invertible on $L^2(\mathbb{P}^N)$ [4], and this implies that the problem of inversion is ill-posed. For this reason, regularization methods have to be introduced to stabilize the inversion in the presence of data noise.

We discuss ill-posed problems only in the framework of linear problems in Hilbert spaces [30]. Let \mathcal{H}, \mathcal{K} be Hilbert spaces and let A be a linear bounded operator from \mathcal{H} into \mathcal{K} . The problem

$$\text{given } g \in \mathcal{K}, \text{ find } f \in \mathcal{H} \text{ such that } Af = g \quad (17)$$

is called well-posed if it is uniquely solvable for each $g \in \mathcal{K}$ and if the solution depends continuously on g . Otherwise (17), is called ill-posed. This means that for an ill-posed problem the operator A^{-1} either does not exist, or is not defined on all of \mathcal{K} , or is not continuous. The practical difficulty with an ill-posed problem is that even if it is solvable, the solution of $Af = g$ need not be close to the solution of $Af = g^\epsilon$ if g^ϵ is close to g .

In general A^{-1} is not a continuous operator. To restore continuity we introduce the notion of a regularization of A^{-1} . This is a family $(T_\gamma)_{\gamma>0}$ of linear continuous operators $T_\gamma: \mathcal{K} \rightarrow \mathcal{H}$ which are defined on all \mathcal{K} and for which

$$\lim_{\gamma \rightarrow 0} T_\gamma g = A^{-1}g \quad (18)$$

on the domain of A^{-1} . Obviously $\|T_\gamma\| \rightarrow +\infty$ as $\gamma \rightarrow 0$ if A^{-1}

is not bounded. With the help of a regularization we can solve (17) approximately in the following sense. Let $g^\epsilon \in \mathcal{H}$ be an approximation of g such that $\|g - g^\epsilon\| \leq \epsilon$. Let $\gamma(\epsilon)$ be such that, as $\epsilon \rightarrow 0$,

$$\gamma(\epsilon) \rightarrow 0, \quad \|T_{\gamma(\epsilon)}\| \epsilon \rightarrow 0. \quad (19)$$

Then, as $\epsilon \rightarrow 0$,

$$\begin{aligned} \|T_{\gamma(\epsilon)}g^\epsilon - A^{-1}g\| &\leq \|T_{\gamma(\epsilon)}(g^\epsilon - g)\| + \|(T_{\gamma(\epsilon)} - A^{-1})g\| \\ &\leq \|T_{\gamma(\epsilon)}\| \epsilon + \|(T_{\gamma(\epsilon)} - A^{-1})g\| \rightarrow 0. \end{aligned} \quad (20)$$

Hence, $T_{\gamma(\epsilon)}g^\epsilon$ is close to $A^{-1}g$ if g^ϵ is close to g . The number γ is called a *regularization parameter*. Determining a good regularization parameter is one of the crucial points in the application of regularization methods.

There are several methods for constructing a regularization as the truncated singular value decomposition, the method of Tikhonov–Phillips or some iterative methods [30]. In the following section we present a regularization method based on the minimization of a MS type functional.

4. MS functional for the simultaneous segmentation and reconstruction of a function

In many practical applications one is not only interested in the reconstruction of the density distribution f but also in the extraction of some specific features within the image which represents the density distribution of the sample. For example, a surgical planning might require the determination of the boundaries of inner organs like liver or lung or the separation of cancerous and healthy tissues. Segmenting a digital image means finding its *homogeneous regions* and its *edges*, or *boundaries*. Of course, the homogeneous regions are supposed to correspond to meaningful parts of objects in the real world, and the edges to their apparent contours. The MS variational model is one of the principal models of image segmentation. It defines the segmentation problem as a joint smoothing/edge detection problem: given an image $g(x)$, one seeks simultaneously a ‘piecewise smoothed image’ $u(x)$ with a set Γ of abrupt discontinuities, the ‘edges’ of g . The original MS functional [25], is the following:

$$\begin{aligned} MS(\Gamma, u) = &\|u - g\|_{L^2(D)}^2 + \alpha \int_{D \setminus \Gamma} |\nabla u(x)|^2 dx \\ &+ \beta \mathcal{H}^{N-1}(D \cap \Gamma), \end{aligned} \quad (21)$$

where

- $D \subset \mathbb{R}^N$ is an open set (*screen*);
- $\Gamma \subset \mathbb{R}^N$ is a closed set (*set of edges*);
- $u: D \rightarrow \mathbb{R}$ (*cartoon*);
- ∇u denotes the distributional gradient of u ;
- $g \in L^2(D)$ is the datum (*grey intensity levels of the image*);
- $\alpha, \beta > 0$ are parameters (*tuning parameters*);
- \mathcal{H}^{N-1} denotes the $(N - 1)$ -dimensional Hausdorff measure.



Figure 5. Left: image of an eye (g). Center: contours of the image in the Mumford–Shah model (edges Γ). Right: piecewise smooth function approximating the image (cartoon u) [32].

The squared L^2 -distance in (21) plays the role of a fidelity term: it imposes that the cartoon u approximate the image g . The second term in the functional imposes that the cartoon u be piecewise smooth outside the edge set Γ . In other words this term favors sharp contours rather than zones where a thin layer of gray is used to pass smoothly from white to black or vice versa. Finally the third term in the functional imposes that the contour Γ be ‘small’ and as smooth as possible. What is expected from the minimization of this functional is a sketchy, cartoon-like version of the given image together with its contours. See figure 5.

The minimization of the MS functional represents a compromise between accuracy and segmentation. The compromise depends on the tuning parameters α and β which have different roles. The parameter α determines how much the cartoon u can vary, if α is small some variations of u are allowed, while as α increases u tends to be a piecewise constant function. The parameter β represents a scale parameter of the functional and measures the amount of contours: if β is small, a lot of edges are allowed and we get a fine segmentation. As β increases, the segmentation gets coarser. For more details on the model see the original paper [25], and the book [31].

The minimization of the MS functional in (21) is performed among the admissible pairs (Γ, u) such that Γ is closed and $u \in C^1(D \setminus \Gamma)$. It is worth noticing that in this model there are two unknowns: a scalar function u and the set Γ of its discontinuities. For this reason this category of problems is often called ‘free-discontinuity problem’. Existence of minimizers of the MS functional in (21) was proven by De Giorgi et al in [33] in the framework of bounded variation functions without Cantor part (space SBV) introduced by Ambrosio and De Giorgi in [34]. Further regularity properties for optimal segmentation in the MS model were shown in [35–38].

Here we present a variation of the MS functional, adapted to the inversion problem of the Radon transform. More precisely, we consider a regularization method that quantifies the edge sets together with images, i.e. a procedure that gives simultaneously a reconstruction and a segmentation of f (assumed to be supported in $D \subset \mathbb{R}^N$) directly from the measured tomograms g , based on the minimization of the MS type functional

$$\begin{aligned} J_{MS}(\Gamma, f) = &\|\mathcal{R}f - g\|_{L^2(\mathbb{P}^N)}^2 \\ &+ \alpha \int_{D \setminus \Gamma} |\nabla f(x)|^2 dx + \beta \mathcal{H}^{N-1}(\Gamma). \end{aligned} \quad (22)$$

The only difference between the functionals MS and J_{MS} is the first term, i.e. the fidelity term, that ensures that the reconstruction for f is close enough to a solution of the equation $\mathcal{R}f = g$,

whereas the other terms play exactly the same role explained for the functional MS. As explained above, in addition to the reconstruction of the density f , we are interested in the reconstruction of its singularity set Γ , i.e. the set of points where the solution f is discontinuous. The main difference with respect to the standard MS functional (21) is that we have to translate the information about the set of sharp discontinuities of g (and hence on the space of the Radon transform) into information about the strong discontinuities of f .

4.1. The piecewise constant model

Here we will review the results obtained by Ramlau and Ring [26] concerning the minimization of (22) restricted to piecewise constant functions f , and then consider some interesting open problems. For medical applications, it is often a good approximation to restrict the reconstruction to densities f that are constant with respect to a partition of the body, as the tissues of inner organs, bones, or muscles have approximately constant density.

We introduce the space $PC_m(D)$ as the space of piecewise constant functions that attain at most m different values, where D is an open and bounded subset of \mathbb{R}^N . In other words, each $f \in PC_m(D)$ is a linear combination of m characteristic functions χ_{Ω_k} of sets $(\Omega_k)_{k=1, \dots, m}$ which satisfy

$$\sum_{k=1}^m \chi_{\Omega_k} = \chi_D \quad \text{a.e. .}$$

We assume that the Ω_k 's are relatively open in D and we set $\Gamma_k = \partial\Omega_k$ for the boundary of Ω_k with respect to the relative topology of the open domain D . In this situation the edge set will be given by the union of the boundaries of Ω_k 's. For technical reasons it is necessary to assume a *nondegeneracy* condition on the admissible partitions of D :

$$\Omega = (\Omega_k)_{k=1, \dots, m} \quad \text{is admissible if} \quad \mathcal{L}^N(\Omega_k) \geq \delta, \quad (23)$$

for some $\delta > 0$, for all $k = 1, \dots, m$, where \mathcal{L}^N denotes the Lebesgue measure on \mathbb{R}^N .

It turns out to be convenient to split the information encoded in a typical function,

$$f = \sum_{k=1}^m f_k \chi_{\Omega_k} \in PC_m(D), \quad (24)$$

into a 'geometrical' part described by the m -tuple of pairwise disjoint sets $\Omega = (\Omega_k)_{k=1, \dots, m}$ which cover D up to a set of measure zero and a 'functional' part given by the m -tuple of values $f = (f_k)_{k=1, \dots, m}$. We also use the notation $\Gamma = (\Gamma_k)_{k=1, \dots, m}$, for the boundaries $\Gamma_k = \partial\Omega_k$ of Ω_k .

As usual when dealing with inverse problems, we have to assume that the data g are not exactly known, but that we are only given noisy measured tomograms g^ϵ of a (hypothetical) exact data set g with $\|g^\epsilon - g\|_{L^2(\mathbb{P}^N)} \leq \epsilon$.

If we restrict the functional (22) to functions in $PC_m(D)$ we obtain that the second term (involving the derivatives of f)

disappears, therefore it remains to minimize the functional

$$J_\beta(\Omega, f) = \| \mathcal{R}f - g^\epsilon \|_{L^2(\mathbb{P}^N)}^2 + \beta \sum_{k=1}^m \mathcal{H}^{N-1}(\Gamma_k), \quad (25)$$

over $PC_m(D)$, with respect to the functional variable f (a vector of m components) and the geometric variable Ω (a partition of the domain D with at most m distinct regions satisfying the non degeneracy condition (23)). So the problem is to find $\tilde{f} \in PC_m(D)$ such that

$$\tilde{f} = \sum_{k=1}^m \tilde{f}_k \chi_{\tilde{\Omega}_k} \in PC_m(D), \quad (26)$$

where

$$(\tilde{\Omega}, \tilde{f}) = \arg \min_{(\Omega, f)} J_\beta(\Omega, f). \quad (27)$$

It is clear that \tilde{f} will depend on the regularization parameter β and on the error level ϵ .

Now we can state the results concerning the functional J_β in (25). There are several technical details necessary for the precise statement and proof of the theorems, for which we refer to the original paper [26]. Here we will give a simplified version of the theorems with the purpose of explaining the main goal, without too many technical details. The first result is about the existence of minimizers of the functional J_β in (25).

Theorem 3. For all $g^\epsilon \in L^2(\mathbb{P}^N)$ there exists a minimizer $(\tilde{\Omega}_\beta^\epsilon, \tilde{f}_\beta^\epsilon)$ of the functional J_β in (25), with $\beta > 0$.

The second result regards the stable dependence of the minimizers of the functional J_β in (25) on the error level ϵ .

Theorem 4. Let $(g^{\epsilon_n})_{n \in \mathbb{N}}$ be a sequence of functions in $L^2(\mathbb{P}^N)$ and let $g^\epsilon \in L^2(\mathbb{P}^N)$. For all $n \in \mathbb{N}$, let $(\tilde{\Omega}_\beta^{\epsilon_n}, \tilde{f}_\beta^{\epsilon_n})$ denote the minimizers of the functional J_β with initial data g^{ϵ_n} . If $g^{\epsilon_n} \rightarrow g^\epsilon$ in $L^2(\mathbb{P}^N)$, as $n \rightarrow +\infty$, then there exists a subsequence of $(\tilde{\Omega}_\beta^{\epsilon_n}, \tilde{f}_\beta^{\epsilon_n})$ such that

$$(\tilde{\Omega}_\beta^{\epsilon_{n_j}}, \tilde{f}_\beta^{\epsilon_{n_j}}) \rightarrow (\tilde{\Omega}_\beta^\epsilon, \tilde{f}_\beta^\epsilon)$$

as $j \rightarrow +\infty$, and $(\tilde{\Omega}_\beta^\epsilon, \tilde{f}_\beta^\epsilon)$ is a minimizer of J_β with initial data g^ϵ . Moreover, the limit of each convergent subsequence of $(\tilde{\Omega}_\beta^{\epsilon_n}, \tilde{f}_\beta^{\epsilon_n})$ is a minimizer of J_β with initial data g^ϵ .

Finally the last theorem is a regularization result.

Theorem 5. Let $f^* \in PC_m(D)$ be given,

$$f^* = \sum_{k=1}^m f_k^* \chi_{\Omega_k^*},$$

and let $g^* = \mathcal{R}f^*$. Assume we have noisy data $g^\epsilon \in L^2(\mathbb{P}^N)$ with $\|g^\epsilon - g^*\|_{L^2(\mathbb{P}^N)} \leq \epsilon$. Let us choose the parameter $\beta = \beta(\epsilon)$ satisfying the conditions $\beta(\epsilon) \rightarrow 0$ and

$\epsilon^2/\beta(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$. For any sequence $\epsilon_n \rightarrow 0$, let $(\tilde{\Omega}^n, \tilde{f}^n)$ denote the minimizers of the functional $J_{\beta(\epsilon_n)}$ with initial data g^{ϵ_n} and regularization parameter $\beta = \beta(\epsilon_n)$. Then there exists a convergent subsequence of $(\tilde{\Omega}^n, \tilde{f}^n)$. Moreover, for every convergent subsequence with limit $(\tilde{\Omega}, \tilde{f})$ the function

$$\tilde{f} = \sum_{k=1}^m \tilde{f}_k \chi_{\tilde{\Omega}_k} \in PC_m(D)$$

is a solution of the equation $\mathcal{R}f = g^*$ with a minimal perimeter. Moreover if f^* is the unique solution of this equation then the whole sequence converges

$$\tilde{f}^n \rightarrow f^*, \quad \text{in } L^2(D),$$

when $n \rightarrow +\infty$.

Finally, let us list some open problems in this context:

- Is the nondegeneracy condition (23) necessary?
- Can one find an *a priori* optimal value for the number m of different values?
- Is it possible to give an *a priori* estimate on the L^∞ -norm of the solution (maximum principle)?
- And finally, it would be very important for applications to prove the existence of minimizers of the functional J_{MS} not restricted to piecewise constant functions f .

We observe that all these problems are quite natural, and have been completely solved in the case of the standard MS functional J_{MS} in (21), see e.g. [31, 39].

5. Electrostatic interpretation of J_{MS}

In this section we restrict our attention to the 3-dimensional case. We propose an electrostatic interpretation of the regularization method based on the functional J_{MS} discussed in the previous section. The intent is to give a physical interpretation of the fidelity term $\|\mathcal{R}f - g\|_{L^2(\mathbb{P}^3)}^2$ in the functional (22), that provide the intuition for an improved regularization method. For $N = 3$, the inversion formula (12) and the electrostatic identity (10) particularize, respectively, as follows: for all $f \in S(\mathbb{R}^3)$ one gets

$$f = -\frac{1}{2(2\pi)^2} \Delta I(\mathcal{R}f) \tag{28}$$

and

$$I(\mathcal{R}f) = a_3 f * V, \tag{29}$$

where $V(x) = 1/|x|$ and a_3 is a constant. We present two preliminary Lemmas.

Lemma 2. For all real valued $f \in S(\mathbb{R}^3)$ one has

$$\|\mathcal{R}f\|^2 = \frac{1}{2(2\pi)^2} \|\nabla I(\mathcal{R}f)\|^2. \tag{30}$$

Proof. We know that $f = -\frac{1}{2(2\pi)^2} \Delta I(\mathcal{R}f)$, therefore

$$\begin{aligned} & \int_{\mathbb{R} \times \mathbb{S}^2} |\mathcal{R}f(X, \xi)|^2 dX d\xi \\ &= \int_{\mathbb{R}^3} f(x) I(\mathcal{R}f)(x) dx \\ &= \int_{\mathbb{R}^3} \frac{1}{2(2\pi)^2} [-\Delta I(\mathcal{R}f)](x) I(\mathcal{R}f)(x) dx \\ &= \frac{1}{2(2\pi)^2} \int_{\mathbb{R}^3} |\nabla I(\mathcal{R}f)(x)|^2 dx. \end{aligned}$$

□

Lemma 3. For all real valued $f \in S(\mathbb{R}^3)$ define

$$E = -\frac{1}{2(2\pi)^2} \nabla I(\mathcal{R}f) \tag{31}$$

and

$$\varphi = \frac{1}{2(2\pi)^2} I(\mathcal{R}f). \tag{32}$$

Then

$$f = \nabla \cdot E = -\Delta \varphi. \tag{33}$$

Proof.

$$\begin{aligned} \nabla \cdot E &= -\frac{1}{2(2\pi)^2} \nabla \cdot \nabla I(\mathcal{R}f) = -\Delta \varphi \\ &= \frac{1}{2(2\pi)^2} (-\Delta) I(\mathcal{R}f) = f, \end{aligned}$$

where we used the inversion formula (28). □

Now we consider a measured tomogram $g : \mathbb{P}^3 \rightarrow \mathbb{R}$ and let us assume that $g = \mathcal{R}f_0$ for some $f_0 : \mathbb{R}^3 \rightarrow \mathbb{R}$. By lemmas 2–3 it follows immediately that the fidelity term $\|\mathcal{R}f - g\|_{L^2(\mathbb{P}^3)}^2$ can be rewritten as follows:

$$\begin{aligned} \|\mathcal{R}f - g\|_{L^2(\mathbb{P}^3)}^2 &= 2(2\pi)^2 \|E - E_g\|_{L^2(\mathbb{R}^3)}^2 \\ &= 2(2\pi)^2 \|\nabla \varphi - \nabla \varphi_g\|_{L^2(\mathbb{R}^3)}^2, \end{aligned} \tag{34}$$

where

$$E = -\frac{1}{2(2\pi)^2} \nabla I(\mathcal{R}f), \quad E_g = -\frac{1}{2(2\pi)^2} \nabla I(g) \tag{35}$$

are the corresponding electric fields, while

$$\varphi = \frac{1}{2(2\pi)^2} I(\mathcal{R}f), \quad \varphi_g = \frac{1}{2(2\pi)^2} I(g) \tag{36}$$

are the corresponding potentials.

With respect to the standard MS functional in (21), the new fidelity term in the functional J_{MS} in (22) controls the distance between the Radon transform of f and the tomographic data g . The relevant difference with respect to the original functional is that the function f and its Radon transform $\mathcal{R}f$ are defined in different spaces. Let us try to interpret the fidelity term $\|\mathcal{R}f - g\|_{L^2(\mathbb{P}^3)}^2$ from a physical point of view. A key ingredient for this goal is the electrostatics formulation of the Radon transform. This formulation can be summarized as follows: if we consider, in dimension 3, a function f , we can think at it as a charge distribution density; if we apply to f first the Radon operator \mathcal{R} and then its adjoint \mathcal{I} we obtain, up to a constant, the electrostatic potential generated by the charge distribution f . This formulation can be stated in any dimension N : the difference with general potential theory in dimension N is that, in tomography, the potential produced by a point charge always scales like $1/|x|$, which is the case of electrostatic potential only in dimension 3. From the electrostatic formulation of the Radon transform we can prove that the fidelity term in the functional J_{MS} actually imposes that the electric field produced by the charge distribution f must be close to the ‘measured electric field’. Therefore we conclude that the term $\|\mathcal{R}f - g\|_{L^2(\mathbb{P}^3)}^2$ is a fidelity term in this weaker sense.

Using this property based on the electrostatic interpretation of the tomographic reconstruction, we can try to minimize some appropriate functionals in the new variables E (electric field) or φ (electric potential) and then compute the corresponding f (charge density). We manipulate the functional J_{MS} as follows:

$$\begin{aligned} J_{MS}(\Gamma, f) &= \|\mathcal{R}f - g\|_{L^2(\mathbb{P}^3)}^2 + \alpha \int_{\mathbb{R}^3 \setminus \Gamma} |\nabla f(x)|^2 dx + \beta \mathcal{H}^2(\Gamma) \\ &= 2(2\pi)^2 \|E - E_g\|^2 + \alpha \int_{\mathbb{R}^3 \setminus \Gamma} |\nabla(\nabla \cdot E)(x)|^2 dx \\ &\quad + \beta \mathcal{H}^2(\Gamma) \\ &= 2(2\pi)^2 \|E - E_g\|^2 + \alpha \int_{\mathbb{R}^3 \setminus \Gamma} |\Delta E|^2 dx + \beta \mathcal{H}^2(\Gamma) \\ &= F(\Gamma, E), \end{aligned} \tag{37}$$

where F is a new functional depending on a vector function E and on a set Γ , and we used the fact that $\nabla \wedge E = 0$, since E is conservative.

We observe that the functional

$$\begin{aligned} F(\Gamma, E) &= 2(2\pi)^2 \|E - E_g\|^2 + \alpha \int_{\mathbb{R}^3 \setminus \Gamma} |\Delta E|^2 dx \\ &\quad + \beta \mathcal{H}^2(\Gamma) \end{aligned} \tag{38}$$

is a second order functional for a vector field E in which the set Γ is the set of discontinuities of f and thus is the set of discontinuities of $\nabla \cdot E$. In the functional F we recognize some similarities with a famous second-order free-discontinuity problem: the Blake–Zisserman model. This model is based on the minimization of the Blake–Zisserman

functional

$$\begin{aligned} BZ(\Gamma_0, \Gamma_1, v) &= \|v - v_0\|_{L^2(D)}^2 \\ &\quad + \alpha \int_{D \setminus (\Gamma_0 \cup \Gamma_1)} |\Delta v(x)|^2 dx \\ &\quad + \beta \mathcal{H}^{N-1}(\Gamma_0 \cap D) \\ &\quad + \gamma \mathcal{H}^{N-1}((\Gamma_1 \setminus \Gamma_0) \cap D), \end{aligned} \tag{39}$$

among admissible triples (Γ_0, Γ_1, v) , where

- $D \subset \mathbb{R}^N$ is an open set;
- $\Gamma_0, \Gamma_1 \subset \mathbb{R}^N$ are closed sets;
- Γ_0 is the set of discontinuities of v (jump set), and Γ_1 the set of discontinuities of ∇v (crease set);
- $v: D \rightarrow \mathbb{R}$, $v \in C^2(D \setminus (\Gamma_0 \cup \Gamma_1)) \cap C(D \setminus \Gamma_0)$ is a scalar function;
- Δv denotes the distributional Laplacian of v ;
- $v_0 \in L^2(D)$ is the datum (grey intensity levels of the given image);
- $\alpha, \beta, \gamma > 0$ are parameters;
- \mathcal{H}^{N-1} denotes the $(N - 1)$ -dimensional Hausdorff measure.

The Blake–Zisserman functional allows a more precise segmentation than the MS functional in the sense that also the curvature of the edges of the original picture is approximated. On the other hand, minimizers may not always exist, depending on the values of the parameters β, γ and on the summability assumption on v_0 . We refer to [29, 40–44] for motivation and analysis of variational approach to image segmentation and digital image processing. In particular see [40, 41, 45] for existence of minimizer results and [42] for a counterexample to existence and [46, 47] for results concerning the regularity of minimizers.

Equation (37) implies that the functional J_{MS} can be rewritten in terms of the vector field E and of the discontinuities set of $\nabla \cdot E$, i.e. the set of creases of E , using the terminology of the Blake–Zisserman model. The fact that in the functional F the discontinuities set of E is not present depends on the fact that we are assuming that the charge density f in the functional J_{MS} do not concentrate on surfaces or on lines. If we admit concentrated charge layers we can consider the Blake–Zisserman model for the vector function E as a relaxed version of the MS model for the charge f . In other words we propose to investigate the connections between minimizers of J_{MS} and minimizers of the higher order functional J_{BZ} :

$$\begin{aligned} J_{BZ}(\Gamma_0, \Gamma_1, E) &= \|E - E_g\|_{L^2(\mathbb{R}^3)}^2 \\ &\quad + \alpha \int_{\mathbb{R}^3 \setminus (\Gamma_0 \cup \Gamma_1)} |\Delta E(x)|^2 dx \\ &\quad + \beta \mathcal{H}^2(\Gamma_0) \\ &\quad + \gamma \mathcal{H}^2(\Gamma_1 \setminus \Gamma_0), \end{aligned} \tag{40}$$

with the additional constraint $\nabla \wedge E = 0$. The main advantage of this approach is that the functional J_{BZ} is a purely differential functional, while the functional J_{MS} is an integro-

differential one. We expect that some results about the Blake–Zisserman model that could be rephrased into tomographic terms would provide immediately new results in tomography. Conversely all the peculiar tomographic features, as the intrinsic vector nature of the variable E , the fact that its support cannot be bounded and the extra constraint $\nabla \wedge E = 0$, motivate new research directions in the study of free-discontinuity problem. For example, an interesting result in this context would be the determination of a good hypothesis on the datum E_g that ensure that the charge density f do not concentrate.

We conclude this section with some comments:

- we proved that the measured data g are actually the measured electric field produced by the unknown charge density, so the term $\|\mathcal{R}f - g\|_{L^2(\mathbb{P}^3)}^2$ in the functional is a fidelity term in a weak sense.
- The problem of the reconstruction of the charge can be rephrased into a reconstruction problem for the electric field. The electric field is an irrotational vector field, so the new minimization problem is actually a constrained minimization. In order to avoid this constraint one could reformulate the reconstruction problem in terms of the electric potential φ ($E = -\nabla\varphi$) obtaining a third-order functional in which the fidelity term is

$$\|\nabla\varphi - \nabla\varphi_g\|_{L^2(\mathbb{R}^3)}^2, \quad (41)$$

where the potentials are given by (36).

- All these considerations hold true in dimension 3. In a generic dimension $n \geq 3$ the situation is quite different because the inversion formula for the Radon transform involves a (possibly fractional) power of the Laplacian. In this case the electrostatic description of tomography given in this section fails. In order to restore it, it is necessary to consider another Radon-type transform which involves integrals of f over linear manifolds with codimension d such that $(n - d)/2 = 1$, i.e. $d = n - 2$, see e.g. [4, 20].

Acknowledgments

We thank G Devillanova, G Florio and F Maddalena for helpful discussions. This work was supported by ‘Fondazione Cassa di Risparmio di Puglia’ and by the Italian National Group of Mathematical Physics (GNFM-INdAM).

References

- [1] Radon J 1917 *Bre. Sachsische Akad. Wiss.* **69** 262
- [2] John F 1955 *Plane Waves and Spherical Means: Applied to Partial Differential Equations* (New York: Wiley-Interscience)
- [3] Gel'fand I M and Shilov G E 1966 *Generalized Functions: Properties and Operations* vol 5 (New York: Academic)
- [4] Helgason S 1980 *The Radon Transform* vol 5 (Boston, MA: Birkhäuser)
- [5] Strichartz R S 1982 *Am. Math. Mon.* **89** 377
- [6] Mancini S, Man'ko V I and Tombesi P 1995 *Quantum Semiclass. Opt.* **7** 615
- [7] Man'ko O V and Man'ko V I 1997 *J. Russ. Laser Res.* **18** 407
- [8] Asorey M, Facchi P, Man'ko V I, Marmo G, Pascazio S and Sudarshan E C G 2007 *Phys. Rev. A* **76** 012117
- [9] Facchi P, Ligabò M and Pascazio S 2010 *J. Mod. Opt.* **57** 239
- [10] Man'ko O V, Man'ko V I and Marmo G 2000 *Phys. Scr.* **62** 446
- [11] Man'ko O V, Man'ko V I and Marmo G 2002 *J. Phys. A: Math. Gen.* **35** 699
- [12] Man'ko V I, Marmo G, Simoni A and Ventriglia F 2006 *Open Sys. Inf. Dyn.* **13** 239
- [13] Man'ko M A, Man'ko V I and Mendes R V 2001 *J. Phys. A: Math. Gen.* **34** 8321
- [14] Asorey M, Facchi P, Man'ko V I, Marmo G, Pascazio S and Sudarshan E C G 2008 *Phys. Rev. A* **77** 042115
- [15] Wigner E P 1932 *Phys. Rev.* **40** 749
- [16] Moyal J 1949 *Proc. Camb. Phil. Soc.* **45** 99
- [17] Hillary M, O'Connell R F, Scully M O and Wigner E 1984 *Phys. Rep.* **106** 121
- [18] Bertrand J and Bertrand P 1987 *Found. Phys.* **17** 397
- [19] Vogel K and Risken H 1989 *Phys. Rev. A* **40** 2847
- [20] Facchi P and Ligabò M 2010 *AIP Conf. Proc.* **1260** 3
- [21] Deans S R 2007 *The Radon Transform and Some of Its Applications* (New York: Dover)
- [22] Paris M G A and Řeháček J 2004 *Quantum state estimation (Lecture Notes in Physics* vol 649) (Berlin: Springer)
- [23] Asorey M, Facchi P, Florio G, Man'ko V I, Marmo G, Pascazio S and Sudarshan E C G 2011 *Phys. Lett. A* **375** 861
- [24] Smith K T, Solmon D C and Wagner S L 1977 *Bull. Am. Math. Soc.* **83** 1227
- [25] Mumford D and Shah J 1989 *Commun. Pure Appl. Math.* **42** 577
- [26] Ramlau R and Ring W 2010 *Inverse Problems* **26** 115001
- [27] Ramlau R and Ring W 2007 *J. Comput. Phys.* **221** 539
- [28] Rondi L and Santosa F 2001 *ESAIM Control Optim. Calc. Var.* **6** 517
- [29] Blake A and Zisserman A 1987 *Visual Reconstruction* (Cambridge, MA: MIT Press)
- [30] Natterer F 1986 *The Mathematics of Computerized Tomography* vol 32 (Philadelphia, PA: SIAM)
- [31] Morel J M and Solimini S 1995 *Variational Methods in Image Segmentation* vol 14 (Boston, MA: Birkhäuser)
- [32] Yuille A 2012 http://en.wikipedia.org/wiki/File:Mumford_shah-example.png (The original image has been partially modified here.)
- [33] De Giorgi E, Carriero M and Leaci A 1989 *Arch. Ration. Mech. Anal.* **108** 195
- [34] De Giorgi E and Ambrosio L 1992 *Atti Accad. Naz. Lincei* **8** 168 89
- [35] Dal Maso G, Morel J M and Solimini S 1988 *Acta Mat.* **82** 199
- [36] Ambrosio L, Fusco N and Pallara D 2000 *Functions of Bounded Variation and Free Discontinuity Problems* (Oxford: Oxford University Press)
- [37] Lops F A, Maddalena F and Solimini S 2001 *Annales de l'institut Henri Poincaré (C) Analyse non linéaire* **18** 639
- [38] Bonnet A and David G 2001 *Astérisque* **274** pp vi+ 259
- [39] Fusco N 2003 *Milan J. Math.* **71** 95
- [40] Carriero M, Leaci A and Tomarelli F 1996 *Variational methods for discontinuous structures (Como, 1994 Progress in Nonlinear Differential Equations and their Applications* vol 25) (Boston, MA: Birkhäuser) p 57
- [41] Carriero M, Leaci A and Tomarelli F 1997 *Ann. Sc. Norm. Super. Pisa Cl. Sci. (4)* **25** 257

- [42] Carriero M, Leaci A and Tomarelli F 2001 *Nonconvex Optimization and its Applications (From Convexity to Nonconvexity* vol 55) (Dordrecht: Kluwer Academic) p 381
- [43] Carriero M, Leaci A and Tomarelli F 2003 *J. Physiol. Paris* **97** 343
- [44] Carriero M, Leaci A and Tomarelli F 2008 *Calc. Var. Partial Differ. Equ.* **32** 81
- [45] Coscia A 1991 *Ann. Univ. Ferrara* **37** 185
- [46] Carriero M, Leaci A and Tomarelli F 2002 *C. R. Acad. Sci., Paris I* **334** 343
- [47] Carriero M, Leaci A and Tomarelli F 2002 *Progress in Nonlinear Differential equations and Their Applications* (Boston, MA: Birkhäuser) **51** 67