ON THE ASYMPTOTIC EQUIVALENCE BETWEEN THE RADON AND THE HOUGH TRANSFORMS OF DIGITAL IMAGES

RICCARDO ARAMINI, FABRICE DELBARY, MAURO C BELTRAMETTI AND CLAUDIO ESTATICO

Dipartimento di Matematica, Università di Genova, Genova, Italy

MICHELE PIANA AND ANNA MARIA MASSONE

Dipartimento di Matematica, Università di Genova, Genova, Italy
CNR - SPIN Genova, Genova, Italy

(Communicated by the associate editor name)

ABSTRACT. Although characterized by different mathematical definitions, both the Radon and the Hough transforms ultimately take an image as input and provide, as output, functions defined on a preassigned parameter space, i.e., the so-called either Radon or Hough *sinograms*. The parameters in these two spaces describe a family of curves, which represent either the integration domains considered in the Radon transform, or the kind of curves to be detected by the Hough transform.

It is heuristically known that the Hough sinogram converges to the corresponding Radon sinogram when the discretization step in the parameter space tends to zero. By considering generalized functions in multi-dimensional setting, in this paper we give an analytical proof of this heuristic rationale when the input grayscale digital image is described as a set of grayscale points, that is, as a sum of weighted Dirac delta functions. On these grounds, we also show that this asymptotic equivalence may have a valuable impact on the image reconstruction problem of inverting the Radon sinogram recorded by a medical imaging scanner.

1. Introduction. The Radon transform (RT) [1] is an important tool in harmonic analysis and inverse problems theory with significant impacts on both group theory and applied mathematics. The classical definition of this transform considers integrals over hyperplanes with specific orientation and distance from a reference hyperplane. In this case, many functional properties of the RT have been investigated, including the characterization of its kernel and range, the ill-posedness of the inverse problem, as well as several inversion formulas and algorithms. The RT has also been extended to integration on manifolds [2], although here important functional and computational problems are still open. In biomedical imaging, RT is the well-established mathematical model for data formation in X-ray Computerized Tomography (CT) and in Positron Emission Tomography (PET) [3, 4] and all software tools for image visualization implemented in current industrial CT and PET scanners must realize, at same stage, the numerical inversion of RT.

While RT plays a crucial role in image reconstruction, the Hough transform (HT) [5, 6, 7] provides an important computational technique in pattern recognition. Indeed, HT is widely used in image-processing to detect specific algebraic plane

2010 *Mathematics Subject Classification.* Primary: 44A12, 46F10; Secondary: 68U10, 92C55.

Key words and phrases. Radon Transform, Hough Transform, sinogram, image reconstruction.

curves, which are zero-loci of polynomials whose coefficients depend polynomially on a (small) set of parameters. The basic idea of this recognition procedure comes from the point-line duality in projective plane geometry. Indeed, any single point in the image space corresponds to a locus (which is its HT) in the parameter space, so that each parameter in the locus identifies a curve passing through the point. In turn, a curve in the image space corresponds to a single point in the parameter space, given by the intersection of all HTs of the points belonging to the curve. A histogram (the Hough counter) can therefore be constructed, representing an accumulator function defined on the discretized parameter space: for each cell in the discretized parameter space, the value of the accumulator corresponds to the number of HTs passing through that cell. This way, the position of the maximum in the Hough counter identifies the set of parameters characterizing the curve to be detected in the image space. Improved algorithms based on HT voting schemes and enabling very rapid and efficient recognition of specific geometric features in large images or data sets, have been introduced in [8, 9]. From a more theoretical point of view, a recent research [10] provided a rigorous mathematical foundation, based on algebraic-geometry arguments, for the case of algebraic plane curves of whatever degree, together with a key lemma stating equivalent conditions under which the existence and uniqueness of the intersection point of the HTs is guaranteed. This framework was then applied in [11, 12, 13, 14] to provide an atlas of algebraic curves used to recognize profiles in real astronomical and biomedical images.

This paper formally investigates a known relationship between RT and HT which has been so far discussed just from a heuristic perspective. Specifically, our aim is to prove analytically that, given a digital image represented as a sum of weighted Dirac delta functions, the corresponding Hough counter tends to become the RT of the image itself as the discretization of the parameter space becomes finer and finer. The basic idea dates back to 1981, when an IEEE letter [15] provided examples in which HT can be considered as a particular case of RT. Although influential and constructive, this letter is heuristic and does not consider any formal definition of HT. About ten years later, in [16], the limitations of [15] were noticed and the similarity between the two transforms was furtherly investigated by relying on such a formal definition. Later on, in 2004, an inspirational report [17] outlined a sort of research program to properly understand the link between the two computational tools. However, none of these studies provided a consistent and rigorous proof of the convergence of the Hough counter to the RT sinogram.

The present paper aims to fill this gap setting up a general formalism for a unified description of the two transforms based on well-established mathematical tools, providing a general theorem proving the asymptotic superposition of the Radon and Hough sinograms, and showing a first hint about how this result can be utilized in image reconstruction.

The structure of the paper is as follows. Section 2 provides a summary of the formalism necessary to prove the convergence theorem, which is illustrated in Section 3. Section 4 shows a first application of the theorem to image processing. Our conclusions are offered in Section 5.

2. The Radon Transform and the Hough Transform. This section provides an overview of basic definitions and results concerning RT and HT. The level of generality we adopted is coherent with the formal framework necessary to prove the convergence theorem discussed in Section 3, which is the core result of this paper.

2.1. The Radon Transform. The RT of a piecewise continuous function with compact support in $\mathbb{R}^n$ can be defined over a $(n-1)$ -dimension manifold in an open subset of $\mathbb{R}^n$, parameterized by a finite number t of parameters varying on a open subset of $\mathbb{R}^t$, according to the following.

Definition 1. For $n \in \mathbb{N} \setminus \{0, 1\}$ and $t \in \mathbb{N} \setminus \{0\}$, let W and E be non-empty open subsets of $\mathbb{R}^n$ and $\mathbb{R}^t$ respectively, and let $E' := \{\lambda' \in \mathbb{R}^{t-1} : \exists \lambda_t \in \mathbb{R} : \lambda = (\lambda', \lambda_t) \in E\}$ (with $E' = \emptyset$ if and only if $t = 1$). Moreover, let $f : W \times E \rightarrow \mathbb{R}$ be a function expressible in the λ_t -solvable form, i.e., as $f(x; \lambda) := \lambda_t - F(x; \lambda')$, being $F : W \times E' \rightarrow \mathbb{R}$ such that $F \in C^1(W \times E')$, and assume that

- (i) $\mathcal{S}(\lambda) := \{x \in W : f(x; \lambda) = 0\} \neq \emptyset \forall \lambda \in E$;
- (ii) $\text{grad}_x f(x; \lambda) \neq 0 \forall \lambda \in E, \forall x \in \mathcal{S}(\lambda)$.

Finally, let m be a piecewise continuous function with compact support defined on W , and denoted by $m \in PC_C^0(W)$. Then, the generalized Radon transform (GRT) of m is the function $(R_f m) : E \rightarrow \mathbb{C}$ defined by

$$(1) \quad (R_f m)(\lambda) := \int_W \delta(f(x; \lambda)) m(x) dx \quad \forall \lambda \in E.$$

Here $\delta(f(\cdot; \lambda))$ denotes the δ -Dirac distribution of the function $f(\cdot; \lambda) : W \rightarrow \mathbb{R}$, associated to the submanifold $\mathcal{S}(\lambda) \subset W$, $\forall \lambda \in E$ [18]. The λ_t -solvability condition for f in this definition is not necessary for the following results to hold, but we assume it since it will notably simplify the formalism.

It is also possible to define the GRT of a δ -Dirac distribution by first recalling that any locally integrable function $u \in L_{\text{loc}}^1(W)$ uniquely defines a corresponding distribution $\iota_k(u) \in (C_C^k)'(W)$, that is a continuous linear functional $\iota_k(u) : C_C^k(W) \rightarrow \mathbb{C}$, by setting

$$(2) \quad \langle \iota_k(u), \phi \rangle := \int_W u(x) \phi(x) dx \quad \forall \phi \in C_C^k(W).$$

Here, for any $k \in \mathbb{N}$ or $k = \infty$, $C_C^k(W)$ denotes the space of compactly supported k -times continuously differentiable functions or the space of compactly supported infinitely differentiable functions, respectively, defined on W and, accordingly, $(C_C^k)'(W)$ denotes the corresponding linear space of distributions. In this respect, $L_{\text{loc}}^1(W)$ can be identified as a subspace of $(C_C^k)'(W)$, and then consider ι_k as an inclusion map, i.e., $\iota_k : L_{\text{loc}}^1(W) \hookrightarrow (C_C^k)'(W)$. Moreover, we recall that any distribution $\ell \in (C_C^k)'(W)$ admits (distributional) partial derivatives $\partial^\alpha \ell \in (C_C^{k+|\alpha|})'(W)$ of any order $|\alpha| \geq 0$, according to the following

$$(3) \quad \langle \partial^\alpha \ell, \phi \rangle := (-1)^{|\alpha|} \langle \ell, \partial^\alpha \phi \rangle \quad \forall \alpha \in \mathbb{N}^n, \quad \forall \phi \in C_C^{k+|\alpha|}(W).$$

Then, the following theorem holds.

Theorem 1. (see [19] and lemma 2.1 in [20]) Let $W, \tilde{x} \in W$ and $K \subset W$ be a non-empty open subset of $\mathbb{R}^n$, a given point of W and a compact subset of W containing a neighborhood $U_{\tilde{x}}$ of $\tilde{x}$, respectively. Moreover, let $(m_i)_{i \in \mathbb{N}} \subset C_C^0(W)$ be sequences of functions such that $\text{supp } m_i \subset K \forall i \in \mathbb{N}$ and $\iota_0(m_i) \rightarrow \delta_{\tilde{x}}$ in $(C_C^0)'(W)$ as $i \rightarrow \infty$, i.e.,

$$\lim_{i \rightarrow \infty} \langle \iota_0(m_i), \phi \rangle = \langle \delta_{\tilde{x}}, \phi \rangle = \phi(\tilde{x}) \quad \forall \phi \in C_C^0(W).$$

Then, it holds that $\iota_1(R_f m_i) \rightarrow \delta(f(\tilde{x}; \cdot))$ in $(C_C^1)'(E)$ as $i \rightarrow \infty$, i.e.,

$$\lim_{i \rightarrow \infty} \langle \iota_1(R_f m_i), \psi \rangle = \langle \delta(f(\tilde{x}; \cdot)), \psi \rangle \quad \forall \psi \in C_C^1(E),$$

where $\delta(f(\tilde{x}; \cdot))$ is the Dirac delta of the function $f(\tilde{x}; \cdot) : E \rightarrow \mathbb{R}$.

From Theorem 1 it follows that the appropriate definition of the GRT of a single Dirac delta $\delta_{\tilde{x}} \in (C_C^0)'(W)$ is

$$(4) \quad (R_f \delta_{\tilde{x}})(\cdot) := \delta(f(\tilde{x}; \cdot)) \in (C_C^1)'(E).$$

In addition, the specific form of (4) for the (classical, i.e., not the generalized) Radon transform (RT) on hyperplanes $\mathcal{P}(\omega, \gamma) := \{x \in \mathbb{R}^n : \gamma - \omega \cdot x = 0\}$, $\forall (\omega, \gamma) \in (\mathbb{R}^n \setminus \{0\}) \times \mathbb{R}$, becomes

$$(5) \quad (R\delta_{\tilde{x}})(\omega, \gamma) := \delta(\gamma - \omega \cdot \tilde{x}),$$

where $\delta(\gamma - \omega \cdot \tilde{x})$ is the Dirac delta $\delta(f(\tilde{x}; \cdot)) \in (C_C^1)'(\mathbb{R}^{n+1})$ of the function mapping (ω, γ) into $f(\tilde{x}; \omega, \gamma) := \gamma - \omega \cdot \tilde{x}$.

In the following we will model a digital image m as the sum of a finite number ν of δ -Dirac functions, each one centered in the pixel center $x(P_j)$ and with amplitude $\mu_j \geq 0$ equal to the pixel content, $\forall j = 1, \dots, \nu$, i.e.

$$(6) \quad m(x) = \sum_{j=1}^{\nu} \mu_j \delta(x - x(P_j)).$$

Therefore, by linearity the classical RT of m is

$$(7) \quad (Rm)(\gamma, \omega) = \sum_{j=1}^{\nu} \mu_j \delta(\gamma - \omega \cdot x(P_j)),$$

and its GRT is

$$(8) \quad (R_f m)(\lambda) = \sum_{j=1}^{\nu} \mu_j \delta(f(x(P_j); \lambda)).$$

The visual representations of the intensity values of RT in (7) and of GRT in (8), with respect to the coordinates $\lambda \in E$, are called sinogram and generalized sinogram, respectively.

2.2. The Hough Transform. The Hough transform (HT) is a pattern recognition technique for the automated detection of shapes (such as lines, curves or surfaces) in images [5, 7, 10, 16, 21, 22]. In our framework an appropriate definition of HT is as follows.

Definition 2. Let $f : W \times E \subset \mathbb{R}^n \times \mathbb{R}^t \rightarrow \mathbb{R}$ be a function such that, for each $\lambda \in E$, the map $f_{\lambda} := f(\cdot; \lambda) : W \rightarrow \mathbb{R}$ defined by $x \mapsto f(x; \lambda)$ satisfies the following three conditions

- (a) $\mathcal{S}(\lambda) := \{x \in W : f_{\lambda}(x) = 0\} \neq \emptyset$;
- (b) $f_{\lambda} \in C^1(W)$;
- (c) $(\text{grad}_x f_{\lambda})(x) \neq 0, \forall x \in \mathcal{S}(\lambda)$.

Let $P \in W$ be a point in the image space having coordinates $x = (x_1(P), \dots, x_n(P))$, and let $f_x : E \rightarrow \mathbb{R}$ be the map defined by $\lambda \mapsto f(x; \lambda)$. Then we say that the zero locus of f_x , defined as $\mathcal{H}(x) := \{\lambda \in E : f_x(\lambda) = 0\} \subset E$, is the Hough transform (HT) of the point P with respect to the function f . If no confusion will arise, we simply say that $\mathcal{H}(x)$ is the HT of x .

By considering $\mathcal{S}(\lambda) = \{x \in W : f_\lambda(x) = 0\} \subset W$ and $\mathcal{H}(x) = \{\lambda \in E : f_x(\lambda) = 0\} \subset E$, for each $(x, \lambda) \in W \times E$, the following *duality condition* [10] holds true:

$$(9) \quad x \in \mathcal{S}(\lambda) \Leftrightarrow 0 = f_\lambda(x) = f(x; \lambda) = f_x(\lambda) = 0 \Leftrightarrow \lambda \in \mathcal{H}(x).$$

This allows us to conclude that *the HT $\mathcal{H}(x)$ of a point $x \in W$ contains a point $\lambda \in E$ if and only if $\mathcal{S}(\lambda)$ passes through x .*

In the case of a digital image, a histogram, called weighted Hough counter (or weighted Hough accumulator), is constructed according to the following three steps.

I. *Discretization of the parameter space.* Identify a suitable (and bounded) investigation domain $\mathcal{T} \subset E$ in the parameter space $\mathbb{R}^t$. Next, choose an *initialization point* $\lambda^* = (\lambda_1^*, \dots, \lambda_t^*)$ in $\mathcal{T}$ and, for each $k = 1, \dots, t$, a *sampling distance* d_k with respect to the component λ_k . Then, set

$$(10) \quad \lambda_{k,n_k} := \lambda_k^* + n_k d_k, \quad k = 1, \dots, t, \quad n_k = 0, \dots, \mathcal{N}_k - 1,$$

where $\mathcal{N}_k$ is the number of considered samples for such component, and n_k the index labelling the sample. Moreover, denote by

$$(11) \quad C(\mathbf{n}) := \left\{ \lambda = (\lambda_1, \dots, \lambda_t) \in \mathcal{T} : \lambda_k \in \left[\lambda_{k,n_k} - \frac{d_k}{2}, \lambda_{k,n_k} + \frac{d_k}{2} \right) \forall k = 1, \dots, t \right\}$$

the rectangular cell with centre in the *sampling point* $\lambda_{\mathbf{n}} := (\lambda_{1,n_1}, \dots, \lambda_{t,n_t})$ of the discretized region $\mathcal{T}$, where $\mathbf{n} \in \mathbb{N}^t$ denotes the multi-index $(n_1, \dots, n_t)$ labelling the cells. Finally, denote by $C(\lambda)$ the cell containing the point $\lambda \in \mathcal{T}$. We point out that the discretization defined by (10) only depends on the choice of the initialization point $\lambda^* \in \mathcal{T}$ and the *discretization step* vector $d := (d_1, \dots, d_t) \in \mathbb{R}_+^t$. In the following, such a discretization will be denoted by $\{\lambda^*, d\}$.

II. *Definition of the HT kernel.* Similarly to Definition 1, we now focus again on a λ_t -solvable function $f : W \times E \rightarrow \mathbb{R}$, that is, a function f satisfying (a)-(b)-(c) of Definition 2 and the following two conditions:

- (d) f is λ_t -solvable, i.e., $f(x; \lambda)$ is of the form $f(x; \lambda_1, \dots, \lambda_t) = \lambda_t - F(x; \lambda_1, \dots, \lambda_{t-1})$;
- (e) for each $x \in W$, the function $f_x : E \rightarrow \mathbb{R}$ introduced in Definition 2 is continuously differentiable, i.e., in view of (d), $F_x \in C^1(E')$, where, for each $x \in W$, the map $F_x : E' \rightarrow \mathbb{R}$ is obviously defined as $\lambda' \mapsto F(x; \lambda')$.

To introduce the Hough kernel, we first consider the function c_k mapping the k -th parameter λ_k , for each $k = 1, \dots, t$, to its closest discretized value, which, according to (10) and (11), is defined as $c_k(\lambda_k) = \lambda_k^* + \left\lfloor \frac{1}{2} + \frac{\lambda_k - \lambda_k^*}{d_k} \right\rfloor d_k$, where $\lfloor \cdot \rfloor$ denotes the floor function. Moreover, by setting $\mathcal{T}' := \{\lambda' \in \mathbb{R}^{t-1} : \exists \lambda_t \in \mathbb{R} : (\lambda', \lambda_t) \in \mathcal{T}\}$, we also define the map $c' : \mathcal{T}' \rightarrow \mathbb{R}^{t-1}$ as $\lambda' = (\lambda_1, \dots, \lambda_{t-1}) \mapsto c'(\lambda') = (c_1(\lambda_1), \dots, c_{t-1}(\lambda_{t-1}))$. Then the Hough kernel is the map $p(\cdot, \cdot; \lambda^*, d) : W \times \mathcal{T} \rightarrow \{0, 1\}$, depending on both the function f and the discretization $\{\lambda^*, d\}$ of (10)-(11), defined as

$$(12) \quad p(x, \lambda; \lambda^*, d) := \begin{cases} 1 & (x, \lambda) \in C(x; \lambda) \\ 0 & \text{otherwise,} \end{cases}$$

where $C(x; \lambda) := \{(x, \lambda) \in W \times \mathcal{T} : -d_t/2 \leq \lambda_t - F(x; c'(\lambda')) < d_t/2\}$.

III. *Construction of the weighted Hough accumulator.* For any given set of points of interest in the image space, say P_j , $j = 1, \dots, \nu$, denote by μ_j the grey level [23] associated with P_j . Let us introduce the *weighted Hough counter (with respect to the function f)*, also called *weighted Hough accumulator*, as the map $H(\cdot; \lambda^*, d) : \mathcal{T} \rightarrow \mathbb{N}$

defined by

$$(13) \quad H(\lambda; \lambda^*, d) := \sum_{j=1}^{\nu} \mu_j p(x(P_j), \lambda; \lambda^*, d),$$

where $x(P_j) \in \mathbb{R}^n$ denotes the n coordinates of the point P_j . Accordingly, we can now define the so-called rescaled Hough counter, which is the basis of the link between the HT and the GRT investigated in the following section.

Definition 3. Let $f : W \times E \rightarrow \mathbb{R}$ be a function satisfying conditions (a)–(e) of the step II above, and let $\mathcal{T} \subset E$ be the investigation domain introduced in item I. Then, the function mapping $\lambda \in \mathcal{T}$ to $H(\lambda; \lambda^*, d)/d_t$ (i.e., the ratio between the weighted Hough counter defined in (13) with HT kernel (12) and the sampling distance d_t with respect to the component λ_t) will be called rescaled (weighted) Hough counter. A visual representation of the intensity values of $H(\lambda; \lambda^*, d)/d_t$ in the coordinate system $(\lambda_1, \dots, \lambda_t)$ will be called rescaled Hough sinogram.

3. The convergence theorem. The main result of the present paper is the following

Theorem 2. Let $f : W \times E \rightarrow \mathbb{R}$ be a function satisfying the properties (a)–(e) already assumed in the previous two sections, i.e., (a) $\mathcal{S}(\lambda) := \{x \in W : f_\lambda(x) = 0\} \neq \emptyset \forall \lambda \in E$; (b) $f_\lambda \in C^1(W) \forall \lambda \in E$; (c) $(\text{grad}_x f_\lambda)(x) \neq 0 \forall x \in \mathcal{S}(\lambda), \forall \lambda \in E$; (d) $f(x; \lambda_1, \dots, \lambda_t) = \lambda_t - F(x; \lambda_1, \dots, \lambda_{t-1})$; (e) $F_x \in C^1(E') \forall x \in W$. Moreover, let $\{\lambda^*, d\} = \{\lambda^*, (d_1, \dots, d_t)\}$ be a discretization of the parameter space as explained in item I of the previous section, and define $D := \max\{d_1, \dots, d_t\}$. Finally, let m be a discrete image (6), $(R_f m)(\lambda)$ its generalized Radon transform GRT (8), and $H(\lambda; \lambda^*, d)/d_t$ the corresponding rescaled Hough counter of Definition 3 with HT kernel (12), defined on a bounded and open investigation domain $\mathcal{T} \subset E$.

Then the rescaled Hough counter tends to the GRT as the discretization step tends to zero, that is

$$(14) \quad \lim_{D \rightarrow 0^+} \iota_1 \left(\frac{H(\lambda; \lambda^*, d)}{d_t} \right) = (R_f m)(\lambda) \quad \text{in} \quad (C_C^1)'(\mathcal{T}),$$

where $\iota_1 : L_{\text{loc}}^1(\mathcal{T}) \rightarrow (C_C^1)'(\mathcal{T})$ is the inclusion map (2).

Proof. We first consider the simplest digital image m , i.e., a single Dirac delta function located at a fixed point $P \in W$. Then, from linearity properties, the generalization to a digital image m of kind (6) will straightforwardly conclude the proof.

In the case of a single delta function, the statement of the theorem becomes:

$$(15) \quad \lim_{D \rightarrow 0^+} \iota_1 \left(\frac{p(x(P), \lambda; \lambda^*, d)}{d_t} \right) = \delta(\lambda_t - F(x(P); \lambda')) \quad \text{in} \quad (C_C^1)'(E' \times \mathbb{R}),$$

where p is the HT kernel (12) used in the rescaled Hough counter of Definition 3. According to (12) and (2), the thesis (15) can be recast as

$$(16) \quad \begin{aligned} \lim_{D \rightarrow 0^+} \left\langle \iota_1 \left(\frac{p(x(P), \lambda; \lambda^*, d)}{d_t} \right), \psi \right\rangle = \\ = \left\langle \delta(\lambda_t - F(x(P); \lambda')), \psi \right\rangle \quad \forall \psi \in (C_C^1)'(E' \times \mathbb{R}). \end{aligned}$$

Note that $\delta(\lambda_t - F(x(P); \lambda')) \in (C_C^1)'(E' \times \mathbb{R})$ is well-defined because, for $\lambda = (\lambda', \lambda_t) \in E' \times \mathbb{R}$, we have [18]

$$(17) \quad \begin{aligned} \left\langle \delta(\lambda_t - F(x(P); \lambda')), \psi \right\rangle &= \int_{E' \times \mathbb{R}} \delta(\lambda_t - F(x(P); \lambda')) \psi(\lambda) d\lambda = \\ &= \int_{E'} \psi(\lambda'; F(x(P); \lambda')) \frac{\sqrt{1 + |\nabla_{\lambda'} F(x(P); \lambda')|^2}}{|\nabla_{\lambda} (\lambda_t - F(x(P); \lambda'))|} d\lambda' = \int_{E'} \psi(\lambda'; F(x(P); \lambda')) d\lambda'. \end{aligned}$$

In order to prove (16), by considering (2) we have

$$(18) \quad \begin{aligned} \left\langle \iota_1 \left(\frac{p(x(P), \lambda; \lambda^*, d)}{d_t} \right), \psi \right\rangle &= \\ &= \int_{E' \times \mathbb{R}} \frac{p(x(P), \lambda; \lambda^*, d)}{d_t} \psi(\lambda) d\lambda = \int_{C_d} \frac{\psi(\lambda)}{d_t} d\lambda, \end{aligned}$$

where

$$C_d = \{(\lambda', \lambda_t) \in E' \times \mathbb{R} : -d_t/2 \leq \lambda_t - F(x(P); c'(\lambda')) < d_t/2\}$$

is the set of discretized cells uniquely determined by the submanifold $\lambda_t - F(x(P); \lambda') = 0$. The last integral of (18) has a key role in our proof. Indeed, its integration domain C_d is a multidimensional strip, in general non continuous, contained in $E \subset \mathbb{R}^t$, whose width with respect to the coordinate λ_t is d_t , which is also equal to the denominator of the integrand. This means that, as D goes to zero, hence d_t too, the value of the integral tends, locally, to the value of ψ on the submanifold of the parameter space E such that $\lambda_t - F(x(P); \lambda') = 0$, by virtue of Lagrange mean value theorem. The following part of the proof actually gives a mathematical evidence to this rationale. On these grounds, first we set

$$\begin{aligned} A_d &:= \{(\lambda', \lambda_t) \in E' \times \mathbb{R} : \lambda_t \geq -d_t/2 + F(x(P); c'(\lambda'))\}, \\ B_d &:= \{(\lambda', \lambda_t) \in E' \times \mathbb{R} : \lambda_t > d_t/2 + F(x(P); c'(\lambda'))\}, \end{aligned}$$

so that we can write

$$(19) \quad \int_{C_d} \frac{\psi(\lambda)}{d_t} d\lambda = \int_{A_d} \frac{\psi(\lambda)}{d_t} d\lambda - \int_{B_d} \frac{\psi(\lambda)}{d_t} d\lambda.$$

Both these two integration domains depend on the discretization step d . Since our aim is to relate these domains with the zero loci of f (which does not depend on the discretization step d), we formally consider the shift of the t -th coordinate as follows

$$(20) \quad \tilde{\lambda}_t = \lambda_t - F(x(P); c'(\lambda')) + F(x(P); \lambda')$$

so that we can rewrite the two domains with respect to this new coordinate as

$$\begin{aligned} \tilde{A}_d &:= \{(\lambda', \tilde{\lambda}_t) \in E' \times \mathbb{R} : \tilde{\lambda}_t \geq -d_t/2 + F(x(P); \lambda')\}, \\ \tilde{B}_d &:= \{(\lambda', \tilde{\lambda}_t) \in E' \times \mathbb{R} : \tilde{\lambda}_t > d_t/2 + F(x(P); \lambda')\}. \end{aligned}$$

Hence, since the determinant of the Jacobian of any pure shift coordinate transformation is 1, we can write

$$(21) \quad \begin{aligned} \int_{A_d} \frac{\psi(\lambda)}{d_t} d\lambda &= \int_{A_d} \frac{\psi(\lambda', \lambda_t)}{d_t} d\lambda' d\lambda_t = \\ &= \int_{\tilde{A}_d} \frac{\psi(\lambda', \tilde{\lambda}_t)}{d_t} d\lambda' d\tilde{\lambda}_t = \int_{\tilde{A}} \frac{\psi(\lambda', \tilde{\lambda}_t - d_t/2)}{d_t} d\lambda' d\tilde{\lambda}_t, \end{aligned}$$

$$(22) \quad \begin{aligned} \int_{B_d} \frac{\psi(\lambda)}{d_t} d\lambda &= \int_{B_d} \frac{\psi(\lambda', \lambda_t)}{d_t} d\lambda' d\lambda_t = \\ &= \int_{\tilde{B}_d} \frac{\psi(\lambda', \tilde{\lambda}_t)}{d_t} d\lambda' d\tilde{\lambda}_t = \int_{\tilde{B}} \frac{\psi(\lambda', \tilde{\lambda}_t + d_t/2)}{d_t} d\lambda' d\tilde{\lambda}_t, \end{aligned}$$

where the two sets

$$\begin{aligned} \tilde{A} &:= \left\{ (\lambda', \tilde{\lambda}_t) \in E' \times \mathbb{R} : \tilde{\lambda}_t - F(x(P); \lambda') \geq 0 \right\}, \\ \tilde{B} &:= \left\{ (\lambda', \tilde{\lambda}_t) \in E' \times \mathbb{R} : \tilde{\lambda}_t - F(x(P); \lambda') > 0 \right\}, \end{aligned}$$

coincide up to a zero-measure subset of $E' \times \mathbb{R}$, so that we could use as well the domain $\tilde{A}$ instead of $\tilde{B}$ in the last integral. Accordingly, from (18) and (19),

$$(23) \quad \begin{aligned} \left\langle \iota_1 \left(\frac{p(x(P), \lambda; \lambda^*, d)}{d_t} \right), \psi \right\rangle &= \\ &= \int_{C_d} \frac{\psi(\lambda)}{d_t} d\lambda = \int_{A_d} \frac{\psi(\lambda)}{d_t} d\lambda - \int_{B_d} \frac{\psi(\lambda)}{d_t} d\lambda = \\ &= \int_{\tilde{A}} \frac{\psi(\lambda', \tilde{\lambda}_t - d_t/2)}{d_t} d\lambda' d\tilde{\lambda}_t - \int_{\tilde{A}} \frac{\psi(\lambda', \tilde{\lambda}_t + d_t/2)}{d_t} d\lambda' d\tilde{\lambda}_t = \\ &= - \int_{\tilde{A}} \frac{\psi(\lambda', \tilde{\lambda}_t + d_t/2) - \psi(\lambda', \tilde{\lambda}_t - d_t/2)}{d_t} d\lambda' d\tilde{\lambda}_t. \end{aligned}$$

Thus, by Lagrange theorem, there exists $\eta_t \in (\tilde{\lambda}_t - d_t/2, \tilde{\lambda}_t + d_t/2)$ such that

$$\frac{\psi(\lambda', \tilde{\lambda}_t + d_t/2) - \psi(\lambda', \tilde{\lambda}_t - d_t/2)}{d_t} = \frac{\partial \psi}{\partial \lambda_t}(\lambda', \eta_t),$$

so that, by the continuity of $\partial \psi / \partial \lambda_t$ and of F and the fact that $\lim_{D \rightarrow 0^+} c'(\lambda') = \lambda'$, we find

$$(24) \quad \lim_{D \rightarrow 0^+} \frac{\psi(\lambda', \tilde{\lambda}_t + d_t/2) - \psi(\lambda', \tilde{\lambda}_t - d_t/2)}{d_t} = \frac{\partial \psi}{\partial \lambda_t}(\lambda', \lambda_t),$$

recalling that, since $\lim_{D \rightarrow 0^+} d_t = 0$, then $\lim_{D \rightarrow 0^+} \eta_t = \tilde{\lambda}_t$ and $\lim_{D \rightarrow 0^+} c'(\lambda') = \lambda'$, so that $\lim_{D \rightarrow 0^+} \tilde{\lambda}_t = \lambda_t$. This way, by considering an upper bound $\bar{D} > 0$ for the maximum sampling distance D , since $\psi \in C_C^1(\mathbb{R}^t)$, we have

$$(25) \quad \left| \frac{\psi(\lambda', \tilde{\lambda}_t + d_t/2) - \psi(\lambda', \tilde{\lambda}_t - d_t/2)}{d_t} \right| \leq \|\psi\|_{C^1} \mathbf{1}_{K_{\bar{D}}}(\lambda),$$

where $\mathbf{1}_{K_{\bar{D}}}(\cdot)$ denotes the characteristic function of the compact subset of $\mathbb{R}^t$ defined as $K_{\bar{D}} := \{(\lambda', \tilde{\lambda}_t + \bar{\lambda}_t) \in \mathbb{R}^t : (\lambda', \tilde{\lambda}_t) \in \text{supp } \psi \text{ and } |\bar{\lambda}_t| \leq \bar{D}\}$. Because $\mathbf{1}_{K_{\bar{D}}}(\cdot) \in L^1(\mathbb{R}^t)$, by (24)–(25) we can apply Lebesgue dominated convergence theorem in (23), thus obtaining

$$(26) \quad \lim_{D \rightarrow 0^+} \left\langle \iota_1 \left(\frac{p(x(P), \lambda; \lambda^*, d)}{d_t} \right), \psi \right\rangle = - \int_A \frac{\partial \psi}{\partial \lambda_t}(\lambda) d\lambda,$$

where

$$A := \{(\lambda', \lambda_t) \in E' \times \mathbb{R} : \lambda_t - F(x(P); \lambda') \geq 0\}.$$

The right hand side of (26) can be rewritten by means of the strict relationships between derivative of a Heaviside step function and the Dirac delta functional. Generally speaking, given a real function $g \in C^1(E)$, let $\Theta(g) : E \rightarrow \{0, 1\}$ be the characteristic function of the region $\{\lambda \in E : g(\lambda) \geq 0\}$, i.e.,

$$(27) \quad \Theta(g(\lambda)) := \begin{cases} 0 & \text{for } \lambda \in E : g(\lambda) < 0, \\ 1 & \text{for } \lambda \in E : g(\lambda) \geq 0. \end{cases}$$

Obviously, $\Theta(g) = \Theta \circ g$, where Θ is the Heaviside function, i.e., $\Theta(t) = 0$ for $t < 0$ and $\Theta(t) = 1$ for $t \geq 0$. Note that $\Theta(g) \in L^1_{\text{loc}}(E)$: then, recalling the inclusion map $\iota_k : L^1_{\text{loc}}(E) \hookrightarrow (C^k_C)'(E)$ defined by (2), $\Theta(f)$ can also be regarded as an element of $(C^k_C)'(E)$, i.e., as $\iota_k(\Theta(f))$, which admits partial derivatives, in a distributional sense, with respect to λ_i , for all $i = 1, \dots, t$ [18]. In particular, for any $k \in \mathbb{N}$ or $k = \infty$, by chain rule it can be proved that its partial derivatives are as follows

$$\frac{\partial \iota_k(\Theta(g))}{\partial \lambda_i} = \frac{\partial g}{\partial \lambda_i} \delta(g) \in (C^{k+1}_C)'(E) \quad \forall i \in \{1, \dots, t\}.$$

Hence, for $g(\lambda) = \lambda_t - F(x(P); \lambda')$, since the characteristic function of the domain A can be written as $\mathbb{R}^t \ni \lambda = (\lambda', \lambda_t) \mapsto \Theta(\lambda_t - F(x(P); \lambda'))$, by considering (3) with $\ell = \iota_0(\Theta(\lambda_t - F(x(P); \lambda')))$, $W = E$, $n = t$, $\phi = \psi$ and $\alpha = (0, \dots, 0, 1)$, we finally have

$$\begin{aligned} \int_A \frac{\partial \psi}{\partial \lambda_t}(\lambda) d\lambda &= \int_{\mathbb{R}^t} \Theta(\lambda_t - F(x(P); \lambda')) \frac{\partial \psi}{\partial \lambda_t}(\lambda) d\lambda = \\ &= \left\langle \iota_0(\Theta(\lambda_t - F(x(P); \lambda'))), \frac{\partial \psi}{\partial \lambda_t} \right\rangle = \\ &= - \left\langle \frac{\partial \iota_0(\Theta(\lambda_t - F(x(P); \lambda')))}{\partial \lambda_t}, \psi \right\rangle = \\ &= - \left\langle \frac{\partial}{\partial \lambda_t}(\lambda_t - F(x(P); \lambda')) \delta(\lambda_t - F(x(P); \lambda')), \psi \right\rangle = \\ &= - \langle \delta(\lambda_t - F(x(P); \lambda')), \psi \rangle. \end{aligned}$$

Now, a straightforward comparison between the two last equalities proves (16), equivalent to (15).

Consider now a discrete image (6) as a weighted sum (or linear combination) of Dirac delta functions. From above we have that

$$\iota_1 \left(\frac{p(x(P_j), \lambda; \lambda^*, d)}{d_t} \right) \rightarrow \delta(\lambda_t - F(x(P_j); \lambda'))$$

in $(C^1_C)'(E' \times \mathbb{R})$ as $D \rightarrow 0^+$, for each $j = 1, \dots, \nu$. The limit

$$\lim_{D \rightarrow 0^+} \iota_1 \left(\frac{H(\lambda; \lambda^*, d)}{d_t} \right) = \lim_{D \rightarrow 0^+} \iota_1 \left(\sum_{j=1}^{\nu} \frac{p(x(P_j), \lambda; \lambda^*, d)}{d_t} \right) = \sum_{j=1}^{\nu} \beta_j \delta(\lambda_t - F(x(P_j); \lambda'))$$

immediately follows from the linearity of the inclusion map $\iota_1 : L^1_{\text{loc}}(E' \times \mathbb{R}) \hookrightarrow (C^1_C)'(E' \times \mathbb{R})$. Finally, it suffices to recall that the right-hand side of the last equalities is just the GRT of m as shown in (8). Accordingly, this is equivalent to relation (14), thus proving the theorem. $\square$

FIGURE 1. (a) The Shepp-Logan phantom. (b) Upper plot: the noise-free Radon sinogram of the Shepp-Logan phantom, computed for 629 values of $\vartheta \in [0, \pi)$ (horizontal axis) and 287 values of $\gamma \in [-\sqrt{2}, \sqrt{2}]$ (vertical axis). Lower plot: the noisy Radon sinogram of the Shepp-Logan phantom, obtained from the upper one by corrupting it with additive 100% level Gaussian noise.

4. Numerical applications. The theoretical result contained in Theorem 2 may have significant consequences on imaging applications. Just as an example, this section shows that dealing with a noisy Radon sinogram as a (rescaled) Hough sinogram allows a more accurate image reconstruction, particularly when the noise level affecting the data is high. We consider the well-known Shepp-Logan phantom [24], shown in panel (a) of Figure 1, as the piecewise constant image (contained in a square of sides ranging from -1 to 1 and formed by pixels with values ranging from 0 to 1) to be recovered from a very noisy Radon sinogram. To this end, we first compute the exact RT with respect to the family of straight lines of equation $\gamma - x_1 \cos \vartheta - x_2 \sin \vartheta = 0$, which is of the form $f(x; \lambda) = 0$, under the identifications $x = (x_1, x_2)$, $\lambda = (\lambda_1, \lambda_2) = (\vartheta, \gamma)$ and $f(x; \lambda) = \lambda_2 - x_1 \cos \lambda_1 - x_2 \sin \lambda_1$. Such computation can be easily performed as explained in [25], by considering I discretized values ϑ_i (with $i = 1, \dots, I$) of $\vartheta \in [0, \pi)$ and J discretized values γ_j (with $j = 1, \dots, J$) of $\gamma \in [-\sqrt{2}, \sqrt{2}]$. The corresponding noise-free sinogram, obtained for $I = 629$ and $J = 287$, is represented in the upper part of panel (b) in Figure 1.

The noise-free sinogram is then corrupted by additive Gaussian noise by using the formula $S_n(\vartheta_i, \gamma_j) = S_t(\vartheta_i, \gamma_j) + \ell \varepsilon S_t(\vartheta_i, \gamma_j)$, where $S_t(\vartheta_i, \gamma_j)$ is the true value of the sinogram at the point (ϑ_i, γ_j) , $S_n(\vartheta_i, \gamma_j)$ is the noisy value of the sinogram at the point (ϑ_i, γ_j) , ε is a realization of a normal Gaussian random variable, and ℓ is the relative noise level (here $\ell = 100\%$ is always assumed). The resulting noisy Radon sinogram is shown in the lower part of panel (b) in Figure 1.

Usually, the inversion of RT is numerically performed by using the filtered back-projection (FBP) algorithm, where the presence of a ramp filter (Ram-Lak filter) in the frequency domain attenuates the blurring effect of a naive unfiltered back-projection and where, at the same time, a second filtering function multiplying the ramp filter allows the attenuation of high frequency noise that can be present in

FIGURE 2. Reconstructions of the Shepp–Logan phantom, obtained from the noisy Radon sinogram (shown in the bottom part of Figure 1, panel (b)) by using the unfiltered back-projection, and the FBP algorithm with five different choices for the filtering function. Except for the case of the unfiltered back-projection, the internal structure of the phantom is almost completely lost.

the Radon sinogram. Common choices for this filtering function are [26, 4] the Shepp–Logan filter; the Cosine filter; the Hamming window; the Hann window. We have then applied both the unfiltered and the filtered back-projection to recover the Shepp–Logan image from its noisy Radon sinogram, using all cited filters in the case of the FBP algorithm. The corresponding results are shown in Figure 2. It is clear that, independently of the particular filter adopted, the FBP algorithm fails to recover the internal structure of the phantom, while the unfiltered back-projection can visualize its main features.

While dealing with the Radon sinogram as a (rescaled) Hough sinogram, then each pixel of the noisy Radon sinogram is a cell of centre (ϑ_i, γ_j) in the parameter space, and the value of the pixel, multiplied by the sampling distance d_2 with respect to the component $\lambda_2 = \gamma$ (cf. (10), for $t = 2$), is the number of straight lines characterized by parameters (ϑ_i, γ_j) in the image space. Note that this number need not be an integer. In fact, all pixel values in the image space $A^2_{(x_1, x_2)}(\mathbb{R})$ are initialized to zero and then, for any pixel centred at (ϑ_i, γ_j) in the parameter space $A^2_{(\vartheta, \gamma)}(\mathbb{R})$ and having value $S_n(\vartheta_i, \gamma_j)$, we trace back in $A^2_{(x_1, x_2)}(\mathbb{R})$ a straight line of equation $\gamma_j - x_1 \cos \vartheta_i - x_2 \sin \vartheta_i = 0$, and the value of each pixel crossed by this straight line is increased by $d_2 S_n(\vartheta_i, \gamma_j)$. The resulting visualization is shown in the bottom-right panel of Figure 3. It is also interesting to implement the above procedure by taking into account only the pixels whose values are larger than a certain threshold. Various thresholds are considered, as five different percentages of the maximum value of the pixels in the noisy Radon sinogram (multiplied by d_2). The corresponding visualizations are shown in the first five panels of Figure 3.

It is worth observing that, unlike Figure 2, the pixel values in the panels of Figure 3 are not related, in principle, to the true values of the Shepp–Logan phantom.

FIGURE 3. Visualizations of the Shepp-Logan phantom, obtained by drawing straight lines identified by pairs of parameters corresponding to cells in the Hough counter (i.e., the noisy Radon sinogram shown in the bottom part of Figure 1, panel (b)) whose values are higher than a fixed percentage of the maximum value. Five different thresholds are chosen, while “no threshold” means that all the pairs of parameters related to non-empty cells are used to identify straight lines in the image space.

However, a visual comparison between Figure 2 and 3 suggests that, for appropriate values of the threshold, the Hough-based approach can provide visualizations that are more informative and accurate than those provided by the (Radon-based) FBP algorithm. This is confirmed by a more quantitative analysis, where we have 1) rescaled the grey levels of all the visualizations in the range $[0, 1]$, as in the original Shepp-Logan phantom; 2) masked the pixels of the background in order to compare just the values of the pixels inside the phantom; 3) computed the visualization error as the Frobenius norm of the matrices given by the differences between each visualization and the original Shepp-Logan phantom. The results of this analysis are in Figure 4, which shows that the Hough-based reconstruction presents a semi-convergent behavior and that it is more accurate than the Radon-based one.

5. Conclusions and future perspectives. This paper provides a rigorous description of the formal relationship between RT and HT. Specifically, the main theoretical result of the paper is concerned with the forward problem associated to RT, i.e., the proof that the rescaled Hough counter of a digital image tends to the Radon sinogram as the discretization step in the parameter space vanishes. Here a digital image is modeled by a finite linear combination of Dirac delta functions, each centered in the center of its corresponding pixel. Moreover, we briefly discussed how the relationship between RT and HT may have impacts on the inverse problem associated to image reconstruction in the case of modalities in which the data formation process is modeled by the RT. Indeed, application perspectives of

FIGURE 4. Visualization errors committed in the inversion of a noisy Radon sinogram by using unfiltered/filtered back projection with different filtering functions (grey dashed plot and grey axes) and by using the Hough-based procedure (black plot and axes).

this paper are concerned with the possibility to invert a Radon sinogram by regarding it as a rescaled Hough sinogram. The simple synthetic example considered in the previous section suggests that the Hough-based procedure with an optimal threshold on the Hough counter may be effectively exploited to invert noisy Radon sinograms. Further studies should be aimed at testing this computational method in the case of experimental tomographic data.

REFERENCES

- [1] S. Helgason, "Integral geometry and Radon transforms", Springer, New York, 2011.
- [2] P. Kuchment, Generalized transforms of Radon type and their applications in "Proceedings of Symposia in Applied Mathematics, vol. 63" (eds. G. Olafsson and E T. Quinto), AMS, 2006.
- [3] F. Natterer, "The Mathematics of Computerized Tomography", Wiley, New York, 1986.
- [4] F. Natterer and F. Wübbeling, "Mathematical Methods in Image Reconstruction", SIAM, Philadelphia, 2001.
- [5] P. V. C. Hough, US Patent 3069654, 1962.
- [6] V. F. Leavers, "Shape Detection in Computer Vision Using the Hough Transform", Springer, London, 1992.
- [7] P. Mukhopadhyay and B. B. Chaudhuri, A survey of Hough Transform, Pattern Recogn., 48 (2015), 993-1010.
- [8] L. A. F. Fernandes and M. M. Oliveira, Real-time line detection through an improved Hough transform voting scheme, Pattern Recogn., 41 (2008), 299-314.
- [9] F. A. Limberger and M. M. Oliveira, Real-time detection of planar regions in unorganized point clouds, Pattern Recogn., 48 (2015), 2043-2053.
- [10] M. C. Beltrametti, A. M. Massone and M. Piana, Hough transform of special classes of curves, SIAM J. Imaging Science, 6 (2013), 391-412.
- [11] A. M. Massone, C. Campi C, A. Perasso A and M. C. Beltrametti, Profile detection in medical and astronomical images by means of the Hough transform of special classes of curves, J. Math. Imaging Vis., 51 (2015), 296-310.
- [12] A. Perasso, C. Campi, A. M. Massone and M. C. Beltrametti, Spinal canal and spinal marrow segmentation by means of the Hough transform of special classes of curves, in "Image Analysis and Processing ICIAP 2015" (eds. V. Murino and E. Puppo), Springer, 2015.

- [13] G. Ricca, M. C. Beltrametti and A. M. Massone, Detecting Curves of Symmetry in Images Via Hough Transform, *Math. Comput. Sci.*, 10 (2016), 179-205.
- [14] C. Marini, S. Morbelli et al., Interplay between spinal cord and cerebral cortex metabolism in amyotrophic lateral sclerosis, *Brain*, 141 (2018), 2272-2279.
- [15] S. R. Deans, Hough transform from the Radon transform, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 3 (1981), 185-188.
- [16] J. Princen, J. Illingworth and J. Kittler, A formal definition of the Hough transform: Properties and relationships, *J. Math. Imaging Vis.*, 1 (1992), 153-168.
- [17] M. van Ginkel, C. L. L. Hendriks and L. J. van Vliet, "The Quantitative Image Group Technical Report Series", N. QI-2004-01, 2004.
- [18] I. M. Gel'fand, M. I. Graev and N. Y. Vilenkin, "Generalized Functions. Volume 5, Integral Geometry and Representation Theory", Academic Press, New York, 1966.
- [19] A. H. Zemanian, "Distribution Theory and Transform Analysis", Dover, New York, 1987.
- [20] M. A. S. Pereira Vasconcelos, "On the Radon Transform of Sampled Functions", PhD Thesis, Harvard University, Cambridge (MA).
- [21] D. H. Ballard, Generalizing the Hough transform to detect arbitrary shapes, *Pattern Recogn.*, 13 (1981), 111-122.
- [22] R. O. Duda and P. E. Hart, Use of the Hough transformation to detect lines and curves in pictures, *Comm. ACM*, 15 (1972), 11-15.
- [23] R. C. Gonzalez and R. E. Woods, "Digital Image Processing", Prentice Hall, Upper Saddle River, 2002.
- [24] L. Shepp and B. F. Logan, The Fourier reconstruction of a head section, *IEEE Trans. Nucl. Sci.*, 21 (1974), 21-43.
- [25] E. W. Weisstein, "Radon Transform Square" From MathWorld - A Wolfram Web Resource. <http://mathworld.wolfram.com/RadonTransformSquare.html> accessed: 2016-07-28
- [26] M. Bertero and P. Boccacci, "Introduction to Inverse Problems in Imaging", IOP, Bristol, 1998.