DEPARTMENT OF STATISTICS

University of Wisconsin
Madison, WI 53706

TECHNICAL REPORT NO. 612

July 1980

A NEW APPROACH TO THE NUMERICAL EVALUATION OF THE INVERSE RADON TRANSFORM WITH DISCRETE, NOISY DATA

by

Grace Wahba

This work was supported in part by the Lady Davis Trust and in part by the Office of Naval Research under Grant No. N00014-77-C-0675.

A NEW APPROACH TO THE NUMERICAL EVALUATION OF THE INVERSE RADON TRANSFORM WITH DISCRETE, NOISY DATA

Grace Wahba
Department of Statistics
University of Wisconsin
Madison, WI 53706 USA

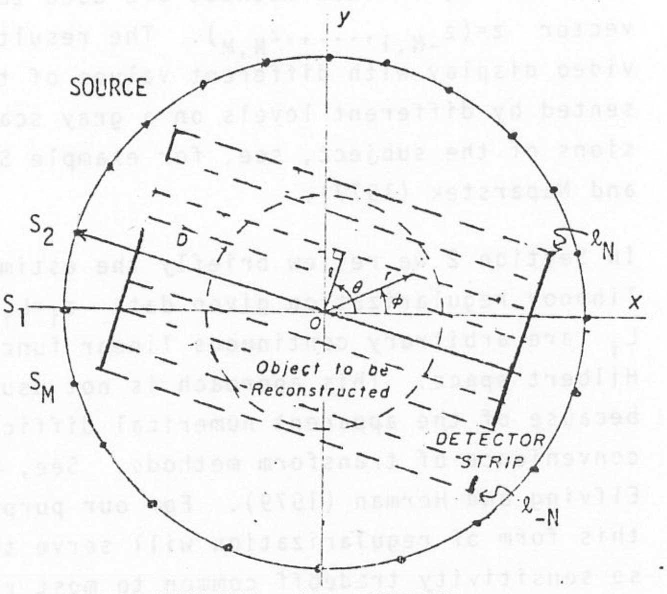
Abstract
The inner (singular) integral in the inverse Radon transform for parallel beam computerized tomography devices can be integrated analytically if the Radon transform considered as a function of the ray position along the detector, is a cubic polynomial spline. Furthermore by using some spline identities, large terms that cancel can be eliminated analytically and the calculation of the resulting expression for the inner integral done in a numerically stable fashion. We suggest using smoothing splines to smooth each set of projection data and by so doing obtain the Radon transform in the above spline form. The resulting analytic expression for the inner integral in the inverse transform is then readily evaluated, and the outer (periodic) integral is replaced by a sum. The work involved to obtain the inverse transform appears to be within the capability of existing computing equipment for typical large data sets. In this regularized transform method the regularization is controlled by the smoothing parameter in the splines. The regularization is directed against data errors and not to prevent unstable numerical operations. Strip integral as well as line integral data can be handled by this method. The method is shown to be closely related to the Tihonov form of regularization.

I. Introduction

Consider a thin "slice" of the human head. In modern computerized tomography (CT) with parallel beam geometry the equivalent of an array of $2N+1$ X-ray beams is directed through the slice and the amount of attenuation of each beam is measured. This procedure is repeated as the array is rotated through M positions, $s_1,\ldots,s_M$, about the head (see Fig.1) to give attenuation factors for a total of $n=(2N+1)M$ beams through the slice. The log of the attenuation

Figure 1

GEOMETRY OF DATA COLLECTION



factor for the ith beam when the array is in the jth position is given approximately by

$$L_{ij}f = \int w_{ij}(x,y)f(x,y)\,dxdy \qquad (1.1)$$

where $f(x,y)$ is the X-ray density of the head slice at the point (x,y) and w_{ij} is a non-negative weight function which is 0 outside a strip surrounding the ij th beam and represents the non-uniform effective distribution of the beam intensity across its narrow width. The formula makes the approximation that the X-ray attenuation coefficient is constant over the spread of energies present in the (nearly) monochromatic beam. In this report we model the data as

$$z_{ij} = L_{ij}f + \varepsilon_{ij}, \quad i=-N,\ldots,N, \quad j=1,2,\ldots,M,$$

where the ε_{ij} are independent zero mean random variables with approximately the same variance which model counting noise and any other (hopefully non-systematic) errors inherent in the measuring device and the approximations being made. The number $n(2N+1)M$ of data points may be of the order of magnitude of 10^5.

In practice, various methods are used to estimate f from the data vector $z=(z_{-N,1},\ldots,z_{N,M})$. The results are usually presented on a video display with different values of the estimate of $f(x,y)$ represented by different levels on a gray scale. For more detailed discussions of the subject, see, for example Shepp and Logan (1974), Herman and Naparstek (1977).

In Section 2 we review briefly the estimation of a function f by Tihonov regularization given data $z_i=L_i f+\varepsilon_i, i=1,2,\ldots,n$, where the L_i are arbitrary continuous linear functionals on some appropriate Hilbert space. This approach is not usual in human head and body CT because of the apparent numerical difficulty and the computational convenience of transform methods. See, however Natterer (1980), Artzy, Elfving and Herman (1979). For our purposes, a close examination of this form of regularization will serve to clarify the resolution-noise sensitivity tradeoff common to most regularization methods for dealing with discrete, noisy data. The method is highly appealing in many mildly ill posed problems (as is the CT problem) whenever it is feasible to implement it.

Most modern human CT devices use methods for estimating f based on an approximate numerical evaluation of a regularized inverse Radon transform. For a recent description of one such algorithm, see Herman Naparstek (1977), Chang and Herman (1978). In Section 3 we present a new approach for the approximate numerical integration of the inverse Radon transform from discrete, noisy data. The work was motivated by a study of, but is apparently quite different from the method described in the above two references. It is in fact quite close to the Tihonov form of regularization with moment discretization. The method entails using a cubic smoothing spline to obtain a smooth function representing each set of projection data, that is, each set $z_j=(z_{-N,j},z_{-(N-1),j},\ldots,z_{N,j})$ where j is fixed. Then the inner (singular) integral in the Radon transform can be evaluated analytically. After using certain relations between the coefficients in cubic splines, one obtains a computationally stable numerical inversion formula which appears feasible to implement with data sets with N and M of the order order of $10^{2.5}$.

The smoothing parameter in the cubic smoothing splines controls the resolution noise sensitivity tradeoff. The suggested approach bypasses most of the usual discretization, quadrature and aliasing errors common to other methods. Unlike smoothing approaches which are, at least in part, directed against numerical problems connected with evaluating a singular integrand, the present approach directs the smoothing against the noisy data in such a way that the singularity can be integrated out analytically.

In Section 4 we indicate the relationship between the transform method proposed, and Tihonov regularization.

2. Tihonov regularization

Let H be a Hilbert space of functions defined on some domain Ω, let $f\epsilon H$ and suppose one observes

$$z_i = L_i f + \varepsilon_i, \quad i=1,2,\ldots,n \qquad (2.1)$$

where the L_i are continuous linear functionals on H, and the ε_i are errors. It is supposed that the ε_i are uncorrelated zero mean random variables with common variance.

Having chosen H, the (Tihonov) regularized estimate $f_{n,\lambda}$ of f given the data $z=(z_1,\ldots,z_n)'$ is the solution to the problem: Find $f\epsilon H$ to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(L_i f - z_i)^2 + \lambda\|f\|^2 \quad . \qquad (2.2)$$

The first term represents the "infidelity" of the solution to the data and, assuming H is a space of "smooth" functions, $\|f_{n,\lambda}\|^2$ represents the "roughness" of the solution. The parameter λ controls this tradeoff. Equivalently λ controls the tradeoff between sensitivity to noise, and resolution. If λ is large $\|f_{n,\lambda}\|$ will be small, and the solution will have low resolution but the sensitivity to noise will also be less, since $\frac{1}{n}\sum_{i=1}^{n}(L_i f_{n,\lambda} - z_i)^2$ can be larger. A small λ will allow $\|f_{n,\lambda}\|$ to be large and corrrespondingly require $L_i f_{n,\lambda}$ to match the data better in mean square.

Since the L_i are continuous linear functionals on H, there exist representers $\eta_1,\ldots,\eta_n \in H$ such that $L_i f = \langle \eta_i, f \rangle$, where $\langle .\,,\,. \rangle$ is the inner product in H. Then the minimizer of (2.2) can be shown to be given by

$$f_{n,\lambda} = (\eta_1,\ldots,\eta_n)(Q+n\lambda I)^{-1} z \ , \qquad (2.3)$$

where Q is the $n \times n$ gram matrix of the representers, with ijth entry Q_{ij}

$$Q_{ij} = \langle \eta_i, \eta_j \rangle \ .$$

Equivalently,

$$f_{n,\lambda} = K_n^*(K_n K_n^* + n\lambda I)^{-1} z \ , \qquad (2.4)$$

where $K_n : H \to E_n$ is defined by $K_n f = (L_1 f,\ldots,L_n f)$, and K_n^* is the adjoint of K_n in the sense that $(z, K_n f) = \langle K_n^* z', f \rangle$, all $z \in E_n, f \in H$ where $(\ ,\ )$ is the Euclidean inner product. ($K_n K_n^*$ is the operator of matrix multiplication by Q) Results are available concerning the convergence of $f_{n,\lambda}$ when λ is chosen appropriately and are stated in a little more detail in Section 4.

We remark that if $H = L_2$ then $K_n^*(K_n K_n^* + \lambda I)^{-1}$ is essentially a back projection operator, see Natterer (1980), however in this case λ should be thought of as controlling the scale or dynamic range of the solution rather than its smoothness, and it is then not very important parameter for tumor detection.

We make some remarks on choosing λ and the space H. Natterer (1980) has suggested that for computerized tomography, H should be chosen as the space $H^\alpha(\Omega)$,

$$H^\alpha(\Omega) = \{f : \iint (1+|\xi|^2)^\alpha \ |\hat{f}(\xi)|^2 d\xi < \infty, \ \text{supp } f \subset \bar{\Omega}\}$$

where $\hat{f}(\xi)$ is the Fourier transform of f and α is close to 1/2. Ideally, one should choose H so that it "just" contains the true solutions. If one looks at the problem in "frequency Space" (see Craven and Wahba (1979)) or "eigenfunction space" (see Wahba (1979a)), one can see that the regularized estimate $f_{n,\lambda}$ can be thought of as passing the data through a "low pass filter" where λ controls the half power point (or "bandwidth") of the filter and α controls the "shape", or steepness of the "roll off" of the filter. For H fixed the method of generalized cross validation (GCV) can be used to estimate a good value of λ, or in the case of computerized tomography, to obtain good starting values for human "fine tuning". See Wahba (1979b) and references cited there. In the typical tomography problem it will be necessary to utilize the special structure of the problem and possibly to do GCV on a subset of the data. See the appendix.

3. A novel regularized transform method using smoothing splines.

Herman and Naparstek (1977) and Chang and Herman (1978) have recently studied regularized transform methods for CT reconstruction for a fan beam device. In this section we suggest a new numerical approach to the regularized inversion of the Radon transform for a parallel beam device. A similar but more involved analysis can be carried out for the fan beam inverse transform discussed by Herman and Naparstek but we do not do it here. The method to be given appears to have the advantage of introducing discretization errors and quadrature approximations relatively late in the numerical procedure, and, intuitively, the regularization parameter of the method appears to affect the resolution - noise sensitivity tradeoff in an appropriate manner. The noise suppresion filtering acts directly on the raw data. The resulting smoothed data is in such a form that the singular integrand is evaluated analytically, and large terms which cancel are subtracted analytically. Unstable numerical calculations and further discretization do not appear and thus their effect does not have to be suppressed with further filtering.

The object to be reconstructed is assumed to be within a circle of radius D. The device can be considered to be the equivalent of a raster of parallel rays, which are rotated about the axis. Let Θ index the angular position of the raster, ℓ the distance from the axis to a parallel ray, and let the location of a point inside the circle be given in polar coordinates as (r,ϕ). (See Fig.1). Then $f(r,\phi)$ is the X-ray absorption coefficient at the point (r,ϕ). Let $p(\ell,\phi)$ be the line integral over f taken over the ray indexed by (ℓ,Θ).

We begin with the Radon inversion formula for parallel beams as quoted by Herman and Naparstek (1977), equ. (6).

$$f(r,\phi) = \lim_{\epsilon\to o} \frac{1}{4\pi^2} \int_o^{2\pi} \int_{-D}^{D} \frac{k_\varepsilon(\ell'-\ell)}{\ell'-\ell} \frac{d}{d\ell} p(\ell,\Theta)\, d\ell\, d\Theta \qquad (3.1)$$

where $k_\epsilon(u) =$ if $|u|\geq\epsilon$ and $k_\epsilon(u)=0$ otherwise, and

$$\ell' = r\cos(\Theta-\phi).$$

We first consider the idealized case where the beams are infinitely narrow. Then the data $\{z_{ij}\}$ consist of noise contaminated measurements of $p(\ell_i,\Theta_j)$, that is,

$$z_{ij} = p(\ell_i,\Theta_j) + \epsilon_{ij} \qquad \begin{array}{l} i=-N,\ldots,N \\ j=1,\ldots,M \end{array}$$

It is desired to estimate $f(r,\phi)$ on a grid of points $\{r_k,\phi_j\}$, from this data. It will simplify matters if we let $\phi_j=2\pi j/M$, $j=1,2,\ldots,M$. Without loss of generality, we may derive our formulas by setting $\phi=0$ since the formulas for $\phi=\Theta_k$ may be obtained by relabeling the data.

First fix $\Theta=\Theta_j$. The inner integral in (3.1) becomes

$$\lim_{\epsilon\to o} \int_{-D}^{\ell'-\epsilon} \frac{1}{\ell'-\ell}\frac{d}{d\ell}p(\ell,\Theta_j)\,d\ell + \int_{\ell'+\epsilon}^{D} \frac{1}{\ell'-\ell}\frac{d}{d\ell}p(\ell,\Theta_j)\,d\ell \qquad (3.2)$$

The idea is as follows. One first obtains a cubic smoothing spline approximation, call it $p_\lambda(\ell,\Theta_j)$ to $p(\ell,\Theta_j)$. $p_\lambda(\ell,\Theta_j)$ is the minimizer of

$$\frac{1}{N}\sum_{i=-N}^{N}(f(\ell_i)-z_{ij})^2 + \lambda\int_{-D}^{D}(f''(\ell))^2 d\ell \qquad (3.3)$$

in $W_2^2[-D,D]$. $p_\lambda(\ell,\Theta_j)$ has a representation

$$p_\lambda(\ell,\Theta_j) = a_k+b_k\ell+c_k\ell^2/2+d_k\ell^3/3\ , \quad \ell\epsilon[\ell_k,\ell_{k+1}]\ ,$$

where a_k,b_k,c_k and d_k are (for fixed λ), linear functions of the data z_{ij}, $i=-N,\ldots,N$, and p_λ has two continuous derivatives in ℓ. If λ is chosen well, then under circumstances that are likely to be satisfied here,1/ it is known that $\frac{d}{d\ell}p_\lambda(\ell,\Theta_j)$ is a good estimate of $\frac{d}{d\ell}p(\ell,\Theta_j)$. See Craven and Wahba (1979). We estimate $\frac{d}{d\ell}p(\ell,\Theta_j)$ by

$$\frac{d}{d\ell}p_\lambda(\ell,\Theta_j) = b_k+c_k\ell+d_k\ell^2\ , \qquad \ell\epsilon[\ell_k,\ell_{k+1}]\ . \qquad (3.4)$$

There exist coefficients w_{ki}^b, w_{ki}^c and w_{ki}^d independent of the data and depending on λ and $\{\ell_i\}$ such that

1/ but see last paragraph below.

$$b_k = \sum_{i=-N}^{N} w_{ki}^b \; z_{ij}$$

$$c_k = \sum_{i=-N}^{N} w_{ki}^c \; z_{ij} \qquad (3.5)$$

$$d_k = \sum_{i=-N}^{N} w_{ki}^d \; z_{ij} \quad .$$

These coefficients can be determined and stored once and for all (after λ is selected), requiring 3(2N)(2N+1) storage locations. The storage requirements can be reduced at the cost of time by exploiting recursion relations between the $\{w_{ki}\}$, we omit the details.

By substituting (3.4) into (3.2) the inner integral can be evaluated analytically and the limit as $\epsilon \to 0$ taken.

First, let $\ell' = r\cos\Theta_j$ be in the interior of $[\ell_m, \ell_{m+1}]$. Then (3.2) becomes

$$\sum_{\substack{i=-N \\ i \neq m}}^{N-1} \int_{\ell_i}^{\ell_{i+1}} \frac{b_i + c_i \ell + d_i \ell^2}{\ell' - \ell} \, d\ell$$

$$+ \lim_{\epsilon \to 0} \left[\int_{\ell_m}^{\ell' - \epsilon} \frac{b_m + c_m \ell + d_m \ell^2}{\ell' - \ell} \, d\ell + \int_{\ell' + \epsilon}^{\ell_{m+1}} \frac{b_m + c_m \ell + d_m \ell^2}{\ell' - \ell} \, d\ell \right] \qquad (3.6)$$

$$= J_r(\Theta_j) \text{ , say .}$$

Upon carrying out the indicated integrations, one obtains

$$J_r(\Theta_j) = \sum_{i=-N}^{N-1} (b_i + c_i \ell' + d_i \ell'^2) \ln \left(\frac{\ell' - \ell_i}{\ell' - \ell_{i+1}} \right) + (b_m + c_m \ell' + d_m \ell'^2) \ln \left(\frac{\ell' - \ell_m}{\ell_{m+1} - \ell'} \right)$$

$$+ \sum_{i=-N}^{N-1} (c_i + 2d_i \ell')(\ell_{i+1} - \ell_i) + \frac{1}{2} \sum_{i=-N}^{N-1} d_i ((\ell' - \ell_i)^2 - (\ell' - \ell_{i+1})^2) .$$

(3.7)

These calculations are all stable except possibly for the two terms involving $\ln(\frac{\ell' - \ell_m}{\ell_{m+1} - \ell'})$ and either $\ln(\frac{\ell' - \ell_{m-1}}{\ell' - \ell_m})$ (if ℓ' is near ℓ_m) or

$\ell n(\frac{\ell'-\ell_{m+1}}{\ell'-\ell_{m+2}})$ (if ℓ' is near ℓ_{m+1}). We give the details for ℓ' close to ℓ_m.

Let $\ell_{m+1}-\ell_m=h$ and let

$$\ell' = \ell_m+\delta h \tag{3.8}$$

where $0\leq\delta<1/2$. The possibly offending terms from (3.7) are

$$(b_{m-1}+c_{m-1}\ell'+d_{m-1}\ell'^2)\ell n(\frac{\ell'-\ell_{m-1}}{\ell'-\ell_m})+(b_m+c_m\ell'+d_m\ell'^2)\ell n(\frac{\ell'-\ell_m}{\ell_{m+1}-\ell'}) \tag{3.9}$$

Since the cubic smoothing spline has continuous first and second derivatives at ℓ_m, we always have the relations

$$\begin{aligned}(b_m-b_{m-1})+(c_m-c_{m-1})\ell_m+(d_m-d_{m-1})\ell_m^2 &= 0\\ (c_m-c_{m-1}) \quad +2(d_m-d_{m-1})\ell_m &= 0\end{aligned} \tag{3.10}$$

Substituting (3.8) and (3.10) into (3.9) gives that (3.9) is equal to

$$2(d_m-d_{m-1})h^2\delta^2\ell n(\frac{\delta}{1+\delta}) + (b_m+c_m\ell'+d_m\ell'^2)\ell n(\frac{1+\delta}{1-\delta}) \ . \tag{3.11}$$

If $1/2\leq\delta<1$, a similar expression may be obtained by summing the mth and m+1 st terms.

Subtituting (3.11) into (3.7) one obtains, provided $0\leq\delta<1/2$, (and assuming the ℓ_i are equally spaced)

$$\begin{aligned}J_r(\Theta_j) = &\sum_{i=-N}^{N-1}(b_i+(i+\delta)hc_i+(i+\delta)^2h^2d_i)\ell n(\frac{\delta+(m-i)}{\delta+(m-i-1)})\\ &+h\sum_{i=-N}^{N-1}(c_i+2(i+\delta)d_i) + \frac{1}{2}\sum_{i=-N}^{N-1}d_i[2(m+\delta)h-(2i-1)h^2]\\ &+2(d_m-d_{m-1})h^2\delta^2\ell n(\frac{\delta}{1+\delta})\\ &+ b_m+c_m(m+\delta)h+d_m(m+\delta)^2\ell n(\frac{1+\delta}{1-\delta}) \ .\end{aligned} \tag{3.12}$$

Since $\delta^2\ell n(\frac{\delta}{1+\delta})$ and $\ell n(\frac{1+\delta}{1-\delta})\to 0$ as $\delta\to 0$, this expression is computed in a straightforward manner for $\epsilon_0<\delta\leq 1/2$, for some suitable

ϵ_0 , and set equal to 0 if $0 \leq \delta < \epsilon_0$. A similar expression is obtained for $1/2 \leq \delta < 1$.

Having evaluated $J_r(\Theta_j)$, the estimate of $f(r,0)$ is

$$f(r,0) \approx \frac{1}{M} \sum_{j=1}^{M} J_r(\Theta_j) . \quad (3.13)$$

Thus, one can process each set of projection data (i.e. the data for fixed Θ_j) in parallel. For each j one collects $z_j = (z_{-N,j}) \ldots z_{N,j})$, computes the $\{b_k\}$, $\{c_k\}$ and $\{d_k\}$ from (3.5) , $J_r(\Theta_j)$ from (3.12) or the corresponding expression for $1/2 \leq \delta < 1$, and $f(r,0)$ from (3.13). To obtain $f(r,\Theta_p)$, $\Theta_p \neq 0$, one repeats the calculations with each data set z_j relabeled as z_{j-p}. Note that the coeffients b_k, c_k and d_k depend only on z_j. They can be computed in parallel once for each set of projection data and then the projection data discarded.

The regularization parameter here is λ (the choice of ϵ_0, if reasonable, is secondary). If λ is fixed the w_{ki} of (3.5) can be stored.

The ultimate choice of λ (or several values of λ to provide alternative pictures), should, of course be chosen by examining pictures with competitive λ for their medical usefulness. Since λ controls the smoothness-fidelity tradeoff, varying λ is likely to have the visual effect of bringing the picture in and out of "focus". A too large λ should result in an oversmoothed, blurred picture while a too small λ should result in an overly grainy or "streaky" picture. A good set of candidate λ's should be obtainable at the design stage by using the method of generalized cross validation (GCV), on data from typical subjects with the parameters (e.g. number of photons, number (2N+1) of rays, etc.) that will be used in practice. Transportable code is available for doing this (Merz (1979), Utreras (1979), Fleisher ((1979)). Given λ, the coefficients w_{ki} may be obtained from standard spline theory (e.g. Reinsch (1967)). Numerical results on the estimation of the derivative from noisy data by this method may be found in Craven and Wahba (1979).

We now consider the case where a line integral approximation to the data is not adequate. Suppose it is more appropriate to assume

$$z_{ij} = \int_{\ell_i}^{\ell_{i+1}} w_i(\ell) p(\ell,\Theta_j) + \epsilon_{ij} ,$$

say. Then $p_\lambda(\ell,\Theta_j)$ is estimated by the minimizer of

$$\frac{1}{2N}\sum_{i=-N}^{N-1}\left(\int_{\ell_i}^{\ell_{i+1}} w_i(\ell)f(\ell)d\ell - z_{ij}\right)^2 + \lambda\int_{-D}^{D}(f''(\ell))^2 d\ell \ .$$

If $w_i(\ell)$ is taken as a constant, then $p_\lambda(\ell,\Theta_j)$ has a representation

$p_\lambda(\ell,\Theta_j) = \tilde{a}_k+\tilde{b}_k\ell+\tilde{c}_k\ell^2/2+\tilde{d}\ \ell^3/3+e_k\ell^4/4$, $\ell\in[\ell_k,\ell_{k+1}]$ where $\tilde{a}_k$, $\tilde{b}_k,\tilde{c}_k,\tilde{d}_k$ and $\tilde{e}_k$ are linear functions of the data and p_λ has 3 continuous derivatives. Expressions for the $\tilde{a}_k,\tilde{b}_k,\tilde{c}_k,\ \tilde{d}_k$ and $\tilde{e}_k$ can be obtained, for example by using the representation for splines given in Wahba and Wendelberger (1979). An expression for $J_r(\Theta_j)$ is obtained by adding terms corresponding to $\tilde{e}_k$, to (3.12). $w_i(\ell)$ can also be modelled as, e.g. a trapezoid, which will still result in a piecewise polynomial representation for p_λ, with a sufficient number of continuous derivatives to carry out a similar analysis. There will be more pieces to the piecewise polynomial, however.

4. On the relation between the spline transform method and Tihonov regularization

In section 3 we have discussed a new method for the numerical inversion of the Radon transform which essentially consists of smoothing the data in the range space and then inverting the transform analytically. Due to the circular symmetry in Θ, if one obtains $p_\lambda(\ell,\Theta)$ for $\Theta\neq\Theta_1,\ldots,\Theta_M$, by, e.g. any periodic spline interpolant in Θ through $p_\lambda(\ell,\Theta_k)$, $k=1,2,\ldots,M$, and then performs the integrations of (3.1) exactly, the result will be the same, namely (3.13).

We now discuss the relation of such methods to Tihonov regularization. Let H be any Hilbert space, let $L_t, t\varepsilon T$ be a family of linearly independent continuous linear functionals on H and define the operator K by

$$(Kf)(t) = g(t),\ g(t) = L_t f\ ,\quad t\varepsilon T\ .$$

Letting X be the range of K, we can make X a Hilbert space with the norm

$$||g||_X = \inf_{\substack{f\varepsilon H\\ Kf=g}} ||f||$$

Now consider the data smoothing problem: Find $g\varepsilon X$ to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(g(t_i)-z_i)^2+\lambda\|g\|_X^2 . \qquad (4.1)$$

Letting η_t be the representer of L_t, it can be shown that X is a reproducing kernel space with reproducing kernel $Q(s,t)=\langle\eta_s,\eta_t\rangle_H$. (See Nashed and Wahba (1974). It then follows that the minimizer $g_{n,\lambda}$ of (4.1) is given by

$$g_{n,\lambda}(t)=(Q(t_1,t),\ldots,Q(t_n,t))(Q+n\lambda I)^{-1}z , \qquad (4.2)$$

where Q is the $n\times n$ matrix with ijth entry $Q(t_i,t_j)=\langle\eta_{t_i},\eta_{t_j}\rangle_H$. Now $Q(t_i,t)=L_t\eta_{t_i}=Q_{t_i}(t)$, say , so that

$$Q_{t_i}=K\eta_{t_i} .$$

Letting $L_i=L_{t_i}$, we have by inspection of (2.3) that $f_{n,\lambda}$, the minimizer of

$$\frac{1}{n}\sum_{i=1}^{n}(L_if-z_i)^2+\lambda\|f\|_H^2$$

satisfies

$$Kf_{n,\lambda}=g_{n\lambda} .$$

Now η_t, $t\varepsilon T$ span the orthogonal compliment of the null space of K, since $\langle\eta_t,f\rangle=0, t\varepsilon T \Rightarrow Kf=0$. Thus $f_{n,\lambda}\varepsilon\mathcal{N}(K)^{\perp}$, so that $f_{n,\lambda}$ is the unique element of minimal norm satisfying $Kf=g$ and so (by definition of the generalized inverse K^+) $f_{n,\lambda}=K^+g_{n,\lambda}$. Thus minimal norm smoothing in the range space (endowed with the induced norm), with an exact inversion is equivalent to Tihonov regularization in the domain space. Furthermore $E\|g-g_{n,\lambda}\|_X^2=E\|K^+Kf-f_{n,\lambda}\|_H^2$, where the expectation is taken over the ε_i and convergence obtains as $n\to\infty$ under general conditions and $\lambda=\lambda(n)$ is chosen correctly. See Wahba (1977).

In the procedure we have discussed, smoothing in the Θ direction is not explicit and any periodic method will give the same result. Let $p_\lambda(\ell,\Theta)$ be obtained by, say, cubic spline interpolation given $p_\lambda(\ell,\Theta_k)$, $k=1,2,\ldots,M$. If $p_\lambda(\ell,\Theta)$ were the minimizer of, say

$$\frac{1}{N(2M+1)}\sum_{i,j}(g(\ell_i,\Theta_j)-z_{ij})^2+\int_0^{2\pi}\int_{-D}^{D}\left(\frac{\partial^4 q}{\partial\Theta^2\partial\ell^2}\right)^2 d\Theta d\ell \qquad (4.3)$$

in an appropriate space of functions periodic in Θ, then the method being proposed would be exactly equivalent to Tihonov regularization.

The minimizer of (4.3) and $p_\lambda(\ell,\Theta)$ do not appear to be exactly the same function, however, but the resulting inversion appears to be close.

As far as the choice of space is concerned, this method is appropriate for $p(\cdot,\Theta)\varepsilon W_2^2$,

$$W_2^2 = \{f:f,f' \text{ continuous}, \quad f''\varepsilon L_2[-N,N]\} .$$

Ho.. ver, it is more natural to assume $p(\cdot,\Theta)\varepsilon W_2^1=\{f:f \text{ continuous}, f'\varepsilon L_2[-N,N]\}$ as follows: Consider, for example, head sections $f(x,y)$ which are continuously differentiable functions of x and y plus a tumor which is the equivalent of adding a region of, say, constant higher density. If the boundary of this region is strictly convex and "smooth" then a little reflection will show that $p(\ell,\Theta)$ is a continuous function of ℓ and $\frac{\partial}{\partial\ell}p(\ell,\Theta)$ is piecewise continuous, so that $p(\cdot,\Theta)\varepsilon W_2^1$. The preceeding analysis with line integrals cannot be carried out to obtain a stable computational formula because the derivative of the linear spline is not continuous. However a similar analysis can be carried out with a double integral over a Θ-increment, or, alternatively, doing spline smoothing assuming $p(\cdot,\cdot)\varepsilon W_2^1(-N,N) \otimes W_2^1$ (periodic). This will appear se.. .ately. The ability of $\frac{\partial}{\partial\ell}p_\lambda(\ell,\Theta)$ to approximate $\frac{\partial}{\partial\ell}p(\ell,\Theta)$ in the W_2^1 . .. has yet to be established in a practical sense but may be quite satisfactory for the present purpose if there is L_2 convergence.

Appendix

The GCV estimate of λ is the minimizer of $V(\lambda)$ given by

$$V(\lambda) = \frac{1}{n}\sum_{k=1}^{n}(\hat{z}_k(\lambda)-z_k)^2/(1-\frac{1}{n}\sum_{i=1}^{n}a_k(\lambda))^2$$

where $\hat{z}_k(\lambda)=L_k f_{n,\lambda}$, and $a_k(\lambda)=\frac{\partial}{\partial z_k}L_k f_{n,\lambda}$. See Wahba (1979b) and references cited there. Letting $L_k=L_{ij}$ where i indexes the ray number, and j indexes the rotational position of the detector ($i=-N,\ldots,N$, $j=1,\ldots,M$ in the notation of Section 3), due to the rotational symmetry of the device one should have $a_{ij}(\lambda)=a_i(\lambda)$ independent of j, thus

$$V(\lambda) = \frac{1}{n}\sum_{i=-N}^{N}\sum_{j=1}^{M}(\hat{z}_{ij}(\lambda)-z_{ij})^2/(1-\frac{1}{2N+1}\sum_{i=-N}^{N}a_i(\lambda))^2$$

(I thank F. Natterer for this observation).

One way of obtaining the denominator is to compute $a_i(\lambda)$ as $L_i \delta^{i,l}_{n,\lambda}$, where $\delta^{il}_{n,\lambda}$ is the "picture" when the input is $z=(0,\ldots,0,1,0,\ldots,0)$ with 1 in the ilth position. As an approximation to $V(\lambda)$ one might consider

$$V(\lambda)=\frac{1}{(2N'+1)M'}\sum_{k=-N'}^{N'}\sum_{\ell=1}^{M'}(\hat{z}_{i_k j_\ell}(\lambda)-z_{i_k j_\ell})^2/(1-\frac{1}{2N'+1}\sum_{k=-N'}^{N'}a_{i_k}(\lambda))^2$$

where $\{i_k\}$ and $\{j_\ell\}$ are either representative or randomly selected subsets of the indices.

We note that the use of GCV is appropriate to estimate the smoothing parameter for low pass filtering methods other than Tihonov regularization, see Craven and Wahba (1979), Golub, Heath and Wahba (1979).

References

1. E. Artzy, T. Elfving, G.T. Herman (1979). Quadratic Optimization for image reconstruction, II. Computer Graphics and Image Processing 11, 242-261.
2. T. Chang and G.T. Herman (1978). Filter selection for the fan beam convolution reconstruction algorithm. TR No. MIPG 6, Dept. of Computer Science, SUNY Buffalo.
3. P. Craven and G. Wahba (1979). Smoothing noisy data with spline functions. Numer. Math. 31, 377-403.
4. J. Fleisher (1979) Spline smoothing routines reference manual for the 1110. Academic Computing Center, The University of Wisconsin-Madison.
5. G. Golub, M. Heath and G. Wahba (1979). Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics 21,2, 215-223.
6. G.T. Herman and A. Naparstek (1977). Fast image reconstruction based on a Radon inversion formula appropriate for rapidly collected data. SIAM J. Applied Mathematics 33, 511-533.
7. P.H. Merz (1979). Spline smoothing by generalized cross-validation, a technique for data smoothing. Chevron Research Corp., Richmond, CA.
8. M.Z. Nashed and G. Wahba (1974). Generalized inverses in reproducing kernel spaces. SIAM J. Math. Anal. 5, 974-987.
9. F. Natterer (1979). A Sobolev space analysis of picture reconstruction, to appear, SIAM J. Applied Math.
10. F. Natterer (1980). Efficient implementation of 'optimal' algorithms in computerized tomography, manuscript.
11. C.M.Reinsch (1967). Smoothing by spline functions. Num. Math. 10, 177-183.

12. L.A. Shepp and B.F. Logan (1974). The Fourier reconstruction of a head section, IEEE Transactions on Nuclear Science NS-21, 21-43.

13. F. Utreras (1979). Cross validation techniques for smoothing spline functions in one or two dimensions. In "Smoothing Techniques for Curve Estimation", T. Gasser and M. Rosenblatt, eds. Lecture notes in Mathematics, No.757, Springer-Verlag, Berlin.

14. G. Wahba (1977). Practical approximate solutions to linear operator equations when the data are noisy. SIAM J. Numer. Anal. 14,A, 651-667.

15. G. Wahba (1979a). Spline interpolation and smoothing on the sphere. University of Wisconsin-Madison Statistics Dept. TR No.584.

16. G. Wahba (1979b). Ill posed problems: Numerical and statistical methods for mildly, moderately,and severely ill posed problems with noisy data. University of Wisconsin-Madison Statistics Dept. TR No.595.

17. G. Wahba and J. Wendelberger (1979). Some new mathematical methods for variational objective analysis using splines and cross-validation. University of Wisconsin-Madison Statistics Dept. TR No.578.

Acknowledgments

This work was supported in part by the Lady Davis Trust and in part by the Office of Naval Research (USA) under Grant No.N00014-77-C-0675. The author wishes to acknowledge the hospitality of the Mathematics Department of the Technion, Haifa, where the work was completed.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 612	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A NEW APPROACH TO THE NUMERICAL EVALUATION OF THE INVERSE RADON TRANSFORM WITH DISCRETE, NOISY DATA		5. TYPE OF REPORT & PERIOD COVERED Scientific Interim
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Grace Wahba		8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0675
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of Wisconsin Madison, WI 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research		12. REPORT DATE July 1980
		13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) computerized tomography, inverse Radon transform, smoothing splines		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) (see reverse side)		

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

The inner (singular) integral in the inverse Radon transform for parallel beam computerized tomography devices can be integrated analytically if the Radon transform considered as a function of the ray position along the detector, is a cubic polynomial spline. Furthermore by using some spline identities, large terms that cancel can be eliminated analytically and the calculation of the resulting expression for the inner integral done in a numerically stable fashion. We suggest using smoothing splines to smooth each set of projection data and by so doing obtain the Radon transform in the above spline form. The resulting analytic expression for the inner integral in the inverse transform is then readily evaluated, and the outer (periodic) integral is replaced by a sum. The work involved to obtain the inverse transform appears to be within the capability of existing computing equipment for typical large data sets. In this regularized transform method the regularization is controlled by the smoothing parameter in the splines. The regularization is directed against data errors and not to prevent unstable numerical operations. Strip integral as well as line integral data can be handled by this method. The method is shown to be closely related to the Tihonov form of regularization.