Efficient Implementation of 'Optimal' Algorithms in Computerized Tomography

F. Natterer

We describe three optimal algorithms for the reconstruction of a function in $\mathbb{R}^2$ from a finite number of line or strip integrals: Optimal recovery, Bayes estimate, Tikhonov-Phillips method. In the case of a rotationally invariant scanning geometry we show that the resulting linear system is a bloc-cyclic convolution. This observation leads to algorithms which are roughly as efficient as filtered backprojection which is one of the standard methods. The algorithms can be applied to the case of hollow and truncated projections.

Numerical examples are given.

1. Introduction

Let f_o be the density of a plane picture contained in a disk Ω of diameter 1. We want to recover f_o from a finite set of integrals

$$g_k = \int_{L_k} f_o(x)\,dx, \quad k = 1,\ldots,n .$$

For this system we write sometimes simply $g = Rf_o$. We may think of the L_k as strips, lines or narrow cones in Ω, depending on the ray geometry of the tomograph we want to model. See [1] for this picture reconstruction problem.

In general the system $g = Rf$ is underdetermined. In order to compute an approximation to f_o from g we thus need some kind of a priori-information on f_o.

Let $H^\alpha = H^\alpha(\mathbb{R}^2)$ be the Sobolev space of order $\alpha \geq 0$, i. e. the linear space of all tempered distributions f such that

$$\|f\|^2_{H^\alpha} = \int (1+|\xi|^2)^\alpha |\hat{f}(\xi)|^2 \, d\xi < \infty$$

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

$$H^\alpha_o(\Omega) = \{ f \in H^\alpha : \operatorname{supp}(f) \subseteq \bar{\Omega} \} .$$

Note that $R : H^\alpha_o(\Omega) \to \mathbb{R}^n$ is continuous for $\alpha \geq 0$ if the L_k are strips or cones of finite width and for $\alpha > 1/2$ if the L_k are lines. We always choose α as to make R continuous. The norm in $\mathbb{R}^n$ is the euclidean norm.

In [2] we have shown that $f \in H^\alpha_o(\Omega)$ with α close to $1/2$ is a reasonable assumption for picture densities. Thus we may use

$$f_o \in K = \{ f \in H^\alpha_o(\Omega) : \|f\|_{H^\alpha} \leq 1 \}$$

as a priori-information on f_o. This leads to the following optimal recovery problem:

Compute from

$$Rf_o = g, \qquad f_o \in K$$

an approximation f_M to f_o with minimal worst case error (in the sense of $H_o^\alpha(\Omega)$). Obviously, f_M is the minimal norm solution of $Rf = g$ in $H_o^\alpha(\Omega)$, see [3], i. e. f_M is the solution of $Rf = g$ for which $\|f\|_{H^\alpha}$ is smallest.

In the case $\alpha = 0$, f_M enjoys another optimality property: With the "backprojection" operator

$$R^* : \mathbb{R}^n \to L_2(\Omega),$$

$$(R^*g)(x) = \sum_{k=1}^{n} g_k \chi_k ,$$

$$\chi_k(x) = \begin{cases} 1, & x \in L_k \\ 0, & \text{otherwise} \end{cases}$$

we define a class of reconstruction methods, the backprojection class, by requiring that the final step in the reconstruction process be a backprojection, i. e. the reconstructed density is in range (R^*). The backprojections used in the picture reconstruction literature are essentially the operation defined above but may involve additional interpolations.

It is easy to see that for $\alpha = 0$, R^* is the adjoint of R. As $L_2(\Omega) = \ker(R) \oplus \text{range}(R^*)$, f_M can be characterized as the solution of $Rf = g$ which is in range (R^*). Consequently, if $f \in$ range (R^*) is any approximation to f_o, we have

$$\|f_M - f_o\|_{L_2(\Omega)} \leq \|f - f_o\|_{L_2(\Omega)} ,$$

i. e. f_M is the most accurate (in the L_2-sense) reconstruction of the backprojection class.

If data errors are involved, then the above procedure has to be modified. Instead of computing minimal norm solutions we minimize

$$\|Rf-g\|^2_{\mathbb{R}^n} + \omega^2 \|f\|^2_{H^\alpha}$$

in $H^\alpha_o(\Omega)$ where the euclidean norm in $\mathbb{R}^n$ is used. The minimizer f_{TP} (TP stands for Tikhonov-Phillips) depends on the regularization parameter ω. It is shown in [4] that f_{TP} has minimal worst case error if ω is chosen appropriately.

Instead of deterministic a priori-information of the form $f_o \in K$ with some set K we may also use stochastic assumptions on f_o. We think of f_o as a two-dimensional random process which is normally distributed with mean value $\bar{f}$ and covariance

$$E(f(x)-\bar{f}(x))\ (f(x')-\bar{f}(x')) = F(x,x').$$

The Bayes estimate of f having seen g is then defined to be

$$f_B = E(f \mid Rf = g) = \bar{f} + FR^*(RFR^*)^{-1}\ (g-R\bar{f}),$$

see e. g. [5], p.19 and 28 (see also [6], [7]), where F is the integral operator with kernel F(x,x'). The inverse is to be understood in the sense of Moore-Penrose.

In general, the direct computation of f_M, f_{TP} and f_B is far beyond the capacity of present day computers. However, if the L_k form a pattern which is in a certain sense rotationally invariant (and if F is of the displacement type in the case of f_B), then a very efficient computation is possible, as will be seen in the following section.

2. Efficient computation of f_M, f_{TP} and f_B

With η_k the generators of the functionals $f \to \int_{L_k} f\,dx$ in $H_o^\alpha(\Omega)$, i. e.

$$\forall f \in H_o^\alpha(\Omega) \qquad (\eta_k, f)_{H_o^\alpha(\Omega)} = \int_{L_k} f\,dx$$

it is immediately verified that

$$f_M = \sum_{k=1}^{n} r_k\,\eta_k = R^* r.$$

R^* is the $H_o^\alpha(\Omega)$ - adjoint of R, and the vector r solves $Sr = g$ with

$$S_{k,\ell} = (\eta_k, \eta_\ell)_{H_o^\alpha(\Omega)} \quad ,k,\ell = 1,\ldots n.$$

This system need not be uniquely solvable, but since $Sr = 0$ implies $R^*r = 0$ it doesn't matter which solution we take. Likewise, f_{TP} is given by

$$f_{TP} = \sum_{k=1}^{n} r_k\,\eta_k,$$

$$(S+\omega^2\,I)\,r = g.$$

For the Bayesian estimate we put

$$\eta_k(x) = \int_{L_k} F(x,x')\,dx' \ ,$$

$$S_{k,\ell} = \int_{L_k}\int_{L_\ell} F(x,x')\,dx\,dx'.$$

Then,

$$f_B = \bar{f} + \sum_{k=1}^{n} r_k\,\eta_k$$

$$Sr = g - R\bar{f}.$$

Here we have to take the minimum norm solution in the case of non-uniqueness.

In all three cases we end up with a huge linear system. Fortunately, the matrix of this system turns out to be a bloc-cyclic convolution if the L_k form a pattern which is invariant with respect to rotation. To fix ideas we restrict ourselves to the case of the parallel geometry, but it is obvious that the same reasoning applies to the fanbeam geometry and to other geometries as well. We remark that the basic idea is due to A. Lent [8].

Let $\omega_j = (\cos \varphi_j, \sin \varphi_j)$, $\varphi_j = j\pi/p$, $j = 0,\ldots, 2p-1$ be $2p$ directions and $s_i = ih$, $i = -q,\ldots, q, h = (2q)^{-1}$. Then we write (note the slight change in the notation) L_{ji} for the strip of width d with axis orthogonal to ω_j crossing ω_j at the point $s_i\omega_j$, i. e.

$$L_{ji} = \{x \in \Omega : s_i - d/2 < x\cdot\omega_j \leq s_i + d/2\} .$$

Thus the number of the strips is $n = 2p(2q+1)$, but since the closures of $L_{j+p,i}$ and $L_{j,-i}$ coincide, $j = 0,\ldots, p-1$, only $n/2$ data are needed.

Instead of considering all strips L_{ji}, $i = -q,\ldots,q$, we can remove strips with $|i| > q_o$ or $|i| < q_o$ where $q_o < q$; i. e. we can treat the cases of hollow and truncated projections as well.

Arranging the strip indices (j,i) in lexicographical order starting with (0,-q), we can decompose the vectors g, r into

$$g = \begin{pmatrix} g^0 \\ \vdots \\ g^{2p-1} \end{pmatrix} \quad , \quad r = \begin{pmatrix} r^0 \\ \vdots \\ r^{2p-1} \end{pmatrix} \quad , \quad g^j,\ r^j \in \mathbb{R}^{2q+1}$$

with

$$g_i^j = \int_{L_{ji}} f(x)\ dx$$

and the matrix S into

$$S = \begin{pmatrix} S_{0,0} & S_{0,1} \cdots & S_{0,2p-1} \\ S_{1,0} & S_{1,1} \cdots & S_{1,2p-1} \\ \vdots & & \vdots \\ S_{2p-1,0} & S_{2p-1,1} \cdots & S_{2p-1,\,2p-1} \end{pmatrix}$$

where in the case of f_M and f_{TP} the $(2q+1,2q+1)$-matrices $S_{j,j'}$ are defined by

$$(S_{j,j'})_{k,\ell} = (\eta_{jk}, \eta_{j'\ell})_{H_0^{\alpha}(\Omega)'}$$

$\eta_{j\ell}$ the generators of the functionals $f \to \int_{L_{j\ell}} f\,dx$. From the rotational symmetry of our strip pattern it follows that $S_{j,j'}$ depends only on $j-j'$ mod $2p$. Putting

$$S_j = S_{j,0}\ ,\ j = 0,\ldots,\ 2p-1$$

we thus obtain the decomposition

$$S = \begin{pmatrix} S_0 & S_1 & \cdots & S_{2p-1} \\ S_{2p-1} & S_0 & \cdots & S_{2p-2} \\ \vdots & & & \vdots \\ S_1 & S_2 & \cdots & S_0 \end{pmatrix},$$

i. e. S is a bloc cyclic convolution.

The same decomposition is possible in the Bayesian setting if F depends only on $|x-x'|$; we then have

$$(S_j)_{k\ell} = \int_{L_{ok}} \int_{L_{j\ell}} F(|x-x'|)\,dx\,dx'.$$

Systems of the form Sr = g with S, r, g decomposed as above can be solved by taking Fourier transforms: If $b^0, \ldots, b^{2p-1}$ is any set of vectors or matrices we have

$$b^j = \frac{1}{2p} \sum_{k=0}^{2p-1} e^{\pi ijk/p} \hat{b}^k ,$$

$$\hat{b}^k = \sum_{j=0}^{2p-1} e^{-\pi ijk/p} b^j .$$

The system at hand is equivalent to ($\omega \neq 0$ only for TP)

$$(\hat{S}_j + \omega^2 I)\hat{r}^j = \hat{g}^j, \quad j = 0, \ldots, 2p-1.$$

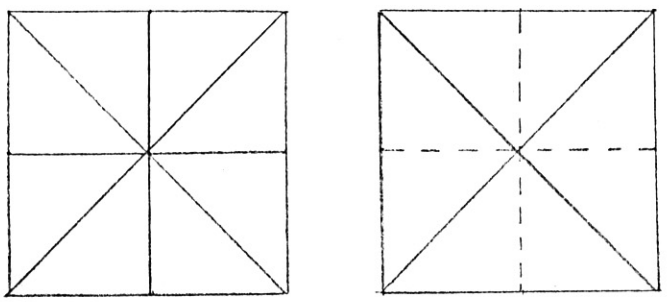
In order to find symmetries of $\hat{S}_j$ we write

$$\hat{S}_j = \sum_{k=0}^{p-1} e^{-\pi ijk/p} (S_k + (-1)^j S_{k+p})$$

S_{k+p} is obtained from S_k by reversing its rows, i. e.

$$(S_{k-p})_{i\ell} = (S_k)_{i,-\ell}.$$

It follows that $S_k + (-1)^j S_{k+p}$, hence $\hat{S}_j$, has the following symmetry axes:



j even j odd

A reflection of the elements of $\hat{S}_j$ is the identity if the axis is solid but changes the sign if it is dotted. As a matter of course, the inverses $\hat{S}_j^{-1}$ (and, if the inverses

fail to exist, the Moore-Penrose generalized inverses $\hat{S}_j^+$) enjoy the same symmetry properties.

In view of these symmetries and since g^{j+p} is g^j reversed, we may assume that r^{j+p} is r^j reversed, i. e. the vectors $\hat{r}^j$ are even for j even and odd for j odd. Combining all these symmetries we find that the equations for $\hat{r}_j$ can be replaced by

$$2[\hat{S}_j+\omega^2 I][\hat{r}_j]^* = [\hat{g}_j],\ j = 0,..,\ 2p-1$$

where [b] denotes the part of the vector (or matrix) b with subscripts 0,... , q if j even and 1,...,q if j odd, and the asterik indicates that for j even the 0 component is multiplied by $\frac{1}{2}$.

For the work estimate it is important to note that the matrices $\hat{S}_j$ are real. This follows easily from $S_k = S_{2p-k}$, k = 0,..., p-1. Suppose the $[\hat{S}_j+\omega^2 I]^+$ are precomputed and stored. Then the computation of the $\hat{r}^j$, j = 0,.., 2p-1, requires essentially $4pq^2$ arithmetic operations (1 operation = 1 real addition + 1 real multiplication). The Fourier transformations can be done with $6pq\ell n_2(2p)$ operations if the Fast Fourier Transform is used; hence this part of the computation can be neglected. Thus we need about $4pq^2$ operations.

f is obtained from r by a generalized backprojection. For f_M and f_{TP}, this backprojection is the ordinary one if $\alpha = 0$. In that case, the evaluation of f on a m x m grid requires $O(m^2 \cdot p)$ operations. For $\alpha > 0$ sufficiently small the $\eta_{j\ell}$ do not have small support, but they decay away from $L_{j\ell}$ fast enough as not to impair essentially this operation count. For f_B we obtain the same work estimate provided that $F(x,x') = 0$ if $|x-x'| \geq C \cdot h$ with some constant C. As $4pq^2$ is usually of the same order as $m^2 \cdot p$ we see that our algorithm requires $O(m^2 p)$ operations. This coincides with the work estimate for the filtered backprojection or convolution method, which is considered to be the most economical

reconstruction technique from the computational point of view.

We conclude this section by determining the kernel of $[\hat{S}_j]$ in a practically important case: We assume that for each j the $L_{j\ell}, \ell = -q,\ldots,q$ cover Ω completely without overlapping, i. e. $d = h$. In this case, we have

$$\sum_{\ell=-q}^{q} (S_j)_{k\ell} = \sum_{\ell=-q}^{q} \int_{L_{j\ell}} \eta_{ok}(x)\, dx = \int_{\Omega} \eta_{ok}(x)\, dx,$$

hence $S_j e, e \in \mathbb{R}^{2q+1}$ the vector whose components are 1, is independent of j. It follows that $\hat{S}_j e = 0$ unless $j = 0$. Using the symmetry properties of $\hat{S}_j$ we find that

$$[\hat{S}_j][e]^* = 0, \quad j = 2,4,\ldots 2p-2.$$

In the Sobolev space setting we are able to show that there are no other elements in the null spaces of $[\hat{S}_j]$, i. e.

$[\hat{S}_j]$ invertible for $j = 0$ and j odd,

$\ker [\hat{S}_j] = \mathrm{sp}\,\{[e]^*\}$, j even.

It suffices to show that the deficiency of S in the space M of data vectors, i. e. $M = \{g : g^{j+p}$ is g^j reversed$\}$ is at most p-1 (hence precisely p-1). Let $Sr = 0$, $r \in M$. With R^* the $H_o^{\alpha}(\Omega)$-adjoint of R, $R^*g = \sum_{j\ell} g_\ell^j \eta_{j\ell}$, we have $S = R R^*$, hence $Sr = 0$ implies $R^*r = 0$. Define the operator $T : L_2(\Omega) \to H_o^{\alpha}(\Omega)$ by

$$\forall f \in H_o^{\alpha}(\Omega) \qquad (Tg,f)_{H_o^{\alpha}(\Omega)} = (g,f)_{L_2(\Omega)}$$

T is injective, and

$$\eta_{j\ell} = T\chi_{j\ell}$$

with the $\chi_{j\ell}$ the characteristic function of $L_{j\ell}$. Hence

$$R^*r = \sum_{j\ell} r_\ell^j \eta_{j\ell} = T \sum_{j\ell} r_\ell^j \chi_{j\ell}$$

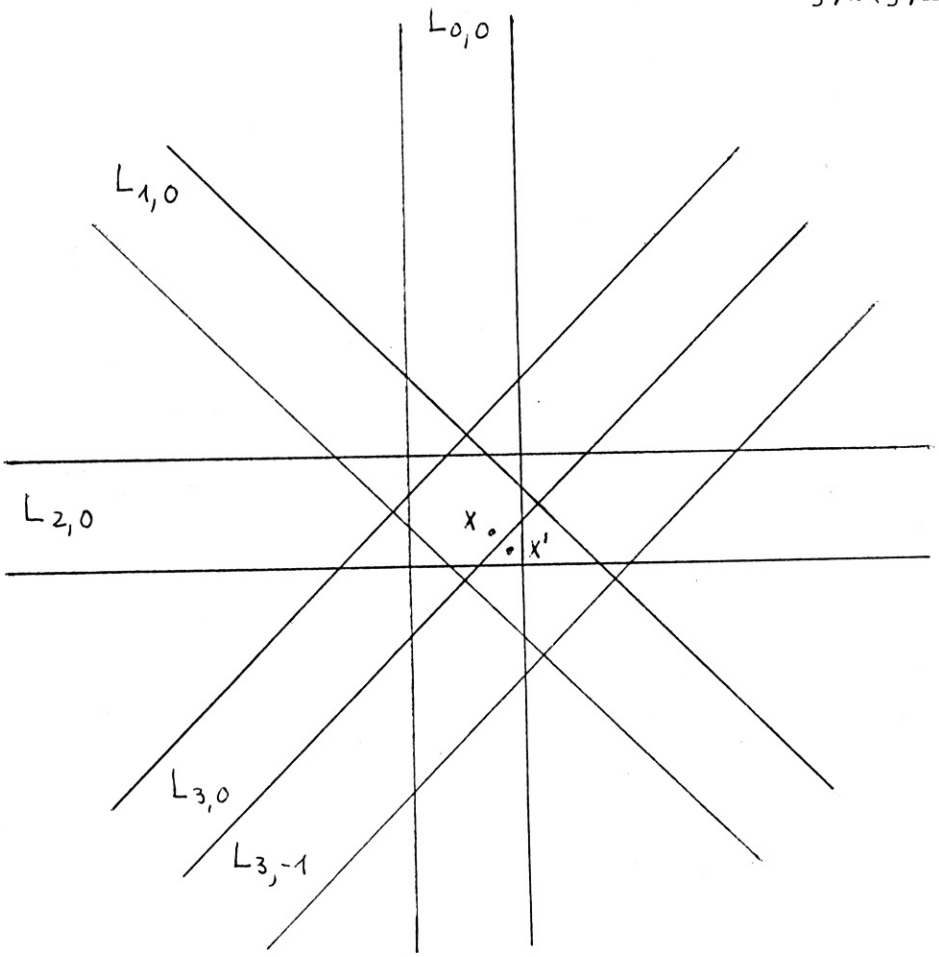
and $R^* r = 0$ implies

$$\sum_{j\ell} r_\ell^j \chi_{j\ell} = 0.$$

For x not on the boundary of any of the strips we obtain

$$\sum_{j=0}^{p-1} r^j_{\ell(j,x)} = 0$$

where $\ell(j,x)$ is uniquely determined by $x \in L_{j,\ell(j,x)}$ for $j < p$.



Choice of x,x'. Figure shows the case $p = 4$, $j_o=3, \ell(j_o,x)=0$

Now fix some $j = j_o$ and choose x close enough to one of the boundaries of $L_{j_o,\ell(j_o,x)}$ as to make sure that the reflection x' of x across this boundary is in $L_{j,\ell(j,x)}$ for $j \neq j_o$ and in $L_{j_o,\ell(j_o,x)+1}$ or $L_{j_o,\ell(j_o,x)-1}$ (see figure; the few cases in

which this can not be done are easily taken care of). Then

$$\sum_{\substack{j=0 \\ j\neq j_o}}^{p-1} r^j_{\ell(j,x)} + r^{j_o}_{\ell(j_o,x)\pm 1} = 0$$

hence

$$r^{j_o}_{\ell(j_o,x)\pm 1} = r^{j_o}_{\ell(j_o,x)}.$$

Proceeding in this fashion we see that each of the vectors r^j, $j = 0,\ldots, p-1$ must be constant, i. e. $r^j = \alpha_j e$, $j = 0,\ldots, p-1$. In addition,

$$\sum_{j=0}^{p-1} r^j_{\ell(j,x)} = \sum_{j=0}^{p-1} \alpha_j e = 0,$$

hence

$$\sum_{j=0}^{p-1} \alpha_j = 0.$$

This proves dim ker(S) $\leq$ p-1.

§ 3 Numerical experiments

For our numerical tests we created a phantom whose density f is given in fig. 1. For a 3D-display of the phantom see fig. 2.

For p = 50 directions we covered the phantom with 51 non-overlapping strips (i. e. q = 25, h = d) and computed the strip integrals analytically. We also computed the line integrals along the axes of the strips. The optimal reconstructions below have been computed from the strip integrals. We compared the optimal reconstructions with the filtered backprojection reconstruction f_{FB}. In order to make the comparison fair we used the line integrals as data for f_{FB}. Our implementation of the filtered backprojection method uses linear interpolation in the backprojection step and the ideal low pass filter whose cut-off-frequency has been chosen so as to minimize the L_2-error of the reconstruction, see [9] for implementations of this method. All reconstructions have been computed on a grid with width 1/64 inside the circle of radius 1/2. The L_2-errors are to be understood as root mean square errors on this grid.

With this phantom and these data we did four experiments.

1. We computed f_M for $\alpha = 0$ and compared it with f_{FB}. The L_2-errors of f_M, f_{FB} are 0.058 and 0.062, respectively. In the 3D-displays of fig. 3 and fig. 4, f_{FB} looks a little bit smoother than f_M. Obviously, there is not much difference between the reconstructions.

2. We computed f_M for $\alpha = 0$ from "hollow" projections. More precisely, we removed from our data the strips passing through the circle of diameter 0.375 around the midpoint of our phantom, i. e. we avoided to hit the three little objects in the interior of the annulus. The reconstructed phantom is displayed in fig. 5 where the density has been put equal to zero in the interior of the annulus. Comparing fig. 4 with fig. 3 we see that working with hollow projections instead of full projections only slightly deteriotes the reconstruction in the outer part of the phantom.

3. We computed f_M for $\alpha = 0$ from "truncated" projections. More precisely, we used only those strips which pass through the circle of diameter 0.375 around the midpoint of our object, i. e. we considered this circle as "region of interest" and avoided wasting radiation (and computer time) in that part of the phantom we are not interested in. A cross section of the reconstruction f_M along the line shown in fig. 1 is given in fig. 6, together with the cross section of the filtered backprojection reconstruction f_{FB} for the truncated data without any compensation. Both reconstructions are completely useless outside the region of interest. Inside the region of interest they approximate a lifted version of the phantom, but f_M is considerably better than f_{FB}.

4. We multiplied the data by $1+c\cdot\ell$ where ℓ is equally distributed in $[0,1]$ and c was chosen in such a way that the relative L_2-error in the data was 10.4 %. Then we computed f_{TP} and f_{FB}. The regularization parameter ω in f_{TP} (as well as the cut-off-frequency in f_{FB}) have been chosen so as to minimize the L_2-error of the reconstruction. The L_2-errors for f_{TP} and f_{FB} are 0.075 and 0.094 respectively. From the 3D-display of f_{TP}, f_{FB} in fig. 7 and fig. 8 respectively we see that the small objects in the interior of the annulus can be made out in f_{TP} but are lumped together in f_{FB}. Thus f_{TP} not only is considerably better than f_{FB} in the L_2-sense but also contains more details. Even though f_{FB} looks nicer than f_{TP} we prefer f_{TP} over f_{FB}. Of course more details would appear in f_{FB} if we increased the cut-off-frequency, but then the L_2-error of f_{FB} would become even worse.

We did these experiments for quite a number of other phantoms, with results very similiar to the reported ones. So it seems we can draw the following conclusions:

For full projections without noise, the minimum norm reconstruction has no advantage over the filtered backprojection method. But note that the latter method requires the determination of an optimal cut-off-frequency. Note also that the former method does

not require any preprocessing of the data if the data are strip integrals rather than line integrals.

There is an obvious advantage of the minimum norm reconstruction over the filtered backprojection method in the case of hollow and truncated projections.

The Tikhonov-Phillips method is superior to the filtered backprojection method in the following sense: If we compare reconstructions with the same L_2-error, than the former reconstruction recovers details more reliably.

Acknowledgement: This work has been supported by the Deutsche Forschungsgemeinschaft. I would like to thank Dipl. Math. M. Engel for writing the programs.

References

[1] Budinger, I.F. - Gulberg, G. T.: Three-Dimensional Reconstruction in Nuclear Medicine Emission Imaging, IEEE Transactions on Nuclear Science 21, 2-20 (1974)

[2] Natterer, F.: A Sobolev Space Analysis of Picture Reconstruction, to appear in SIAM J. Appl. Math.

[3] Golomb, M. - Weinberger, H. F.: Optimal Approximation, in: Langer, R. E. (ed.): On Numerical Approximation, The University of Wisconsin Press, Madison (1959)

[4] Melkman, A. A. - Micchelli, C. A.: Optimal Estimation of Linear Operators in Hilbert Spaces From Inaccurate Data, to appear in SIAM J. Numer. Anal.

[5] Anderson, T. W.: An Introduction to Multivariate Statistical Analysis. Wiley 1957

[6] Hurwitz, H.: Entropy Reduction in Bayesian Analysis of Measurements, Phys. Rev., Sect. A 12, 698-706 (1975)

[7] Tasto, M.: Reconstruction of Objects from Noisy Projections, Computer Graphics and Image Processing 6, 103-122 (1977)

[8] Lent, A.: Seminar talk at the Biodynamic Research Unit, Mayo Clinic, 1975

[9] Rowland, S. W.: Computer Implementation of Image Reconstruction Formulas, in: G. T. Herman, (ed.): Image Reconstruction from Projections. Topics in Applied Physics, vol. 32, Springer 1979.

Fig. 1 Density distribution of phantom. Diameter = 1

Fig. 2 3D-display of phantom

Fig. 3 3D-display of f_M

Fig. 4 3D-display of f_{FB}

Fig. 5 3D-display of f_M for hollow projections

Fig. 6 Cross sections of f_M(Δ), f_{FB}(□) and original(+) along line shown in fig. 1

Fig. 7 3D-display of f_{TP} for noisy data

Fig. 8 3D-display of f_{FB} for noisy data

Fig. 1

$f = 0.30$

$f = 1.00$

$f = 0.66$

$f = 0.22$

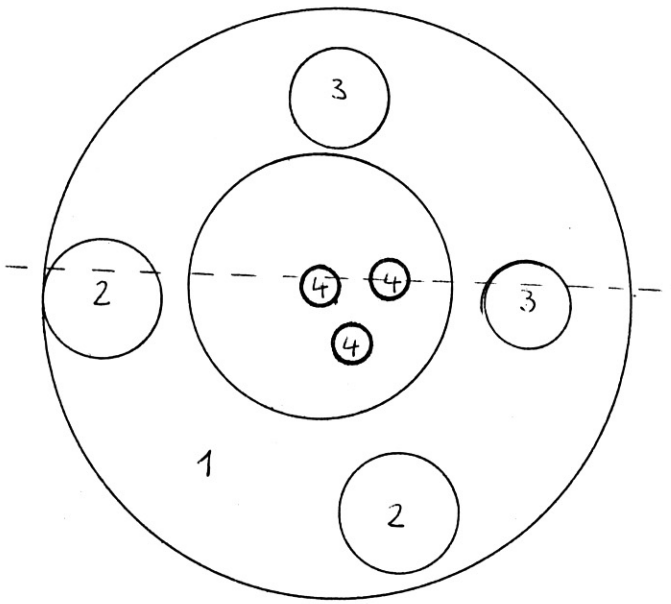


Fig. 2

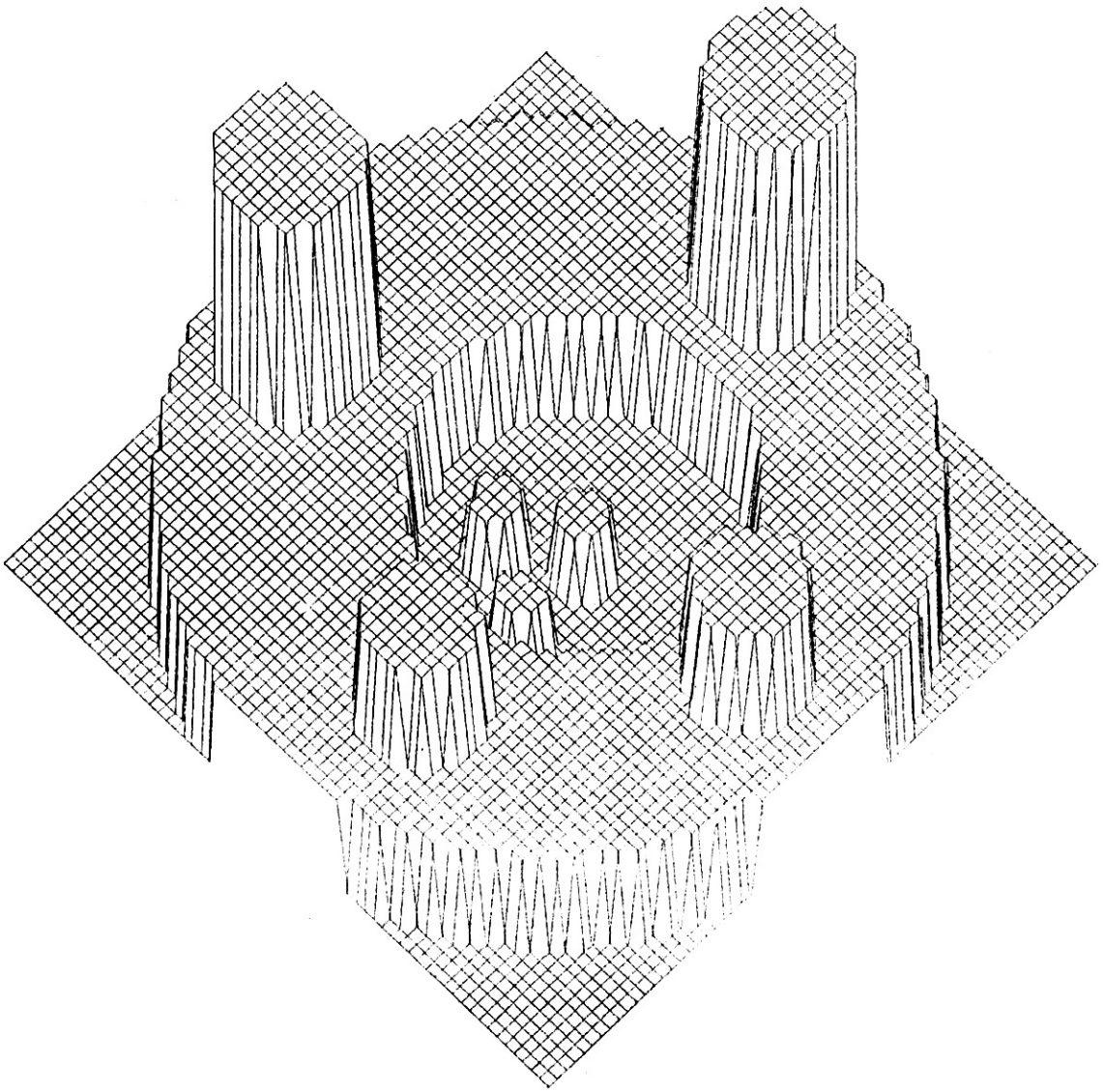


Fig. 3

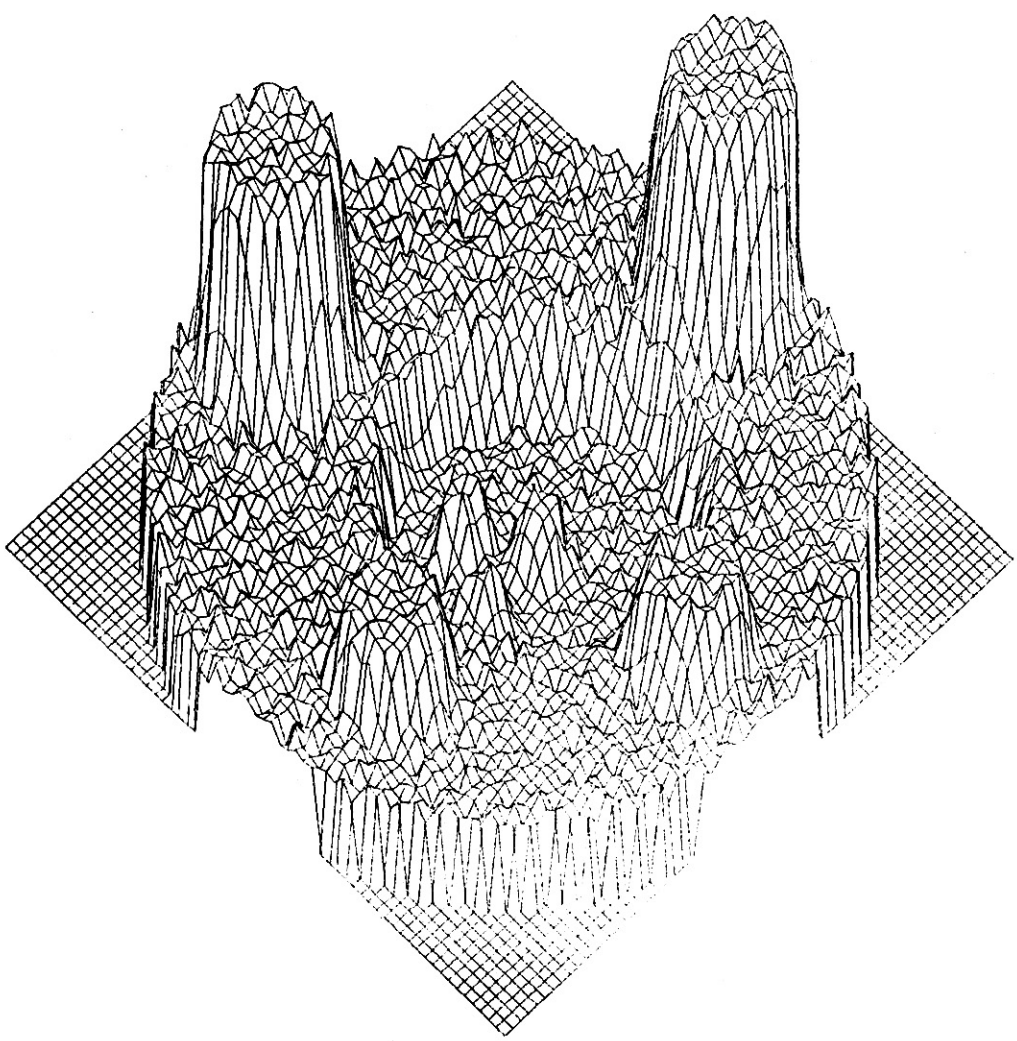


Fig. 4

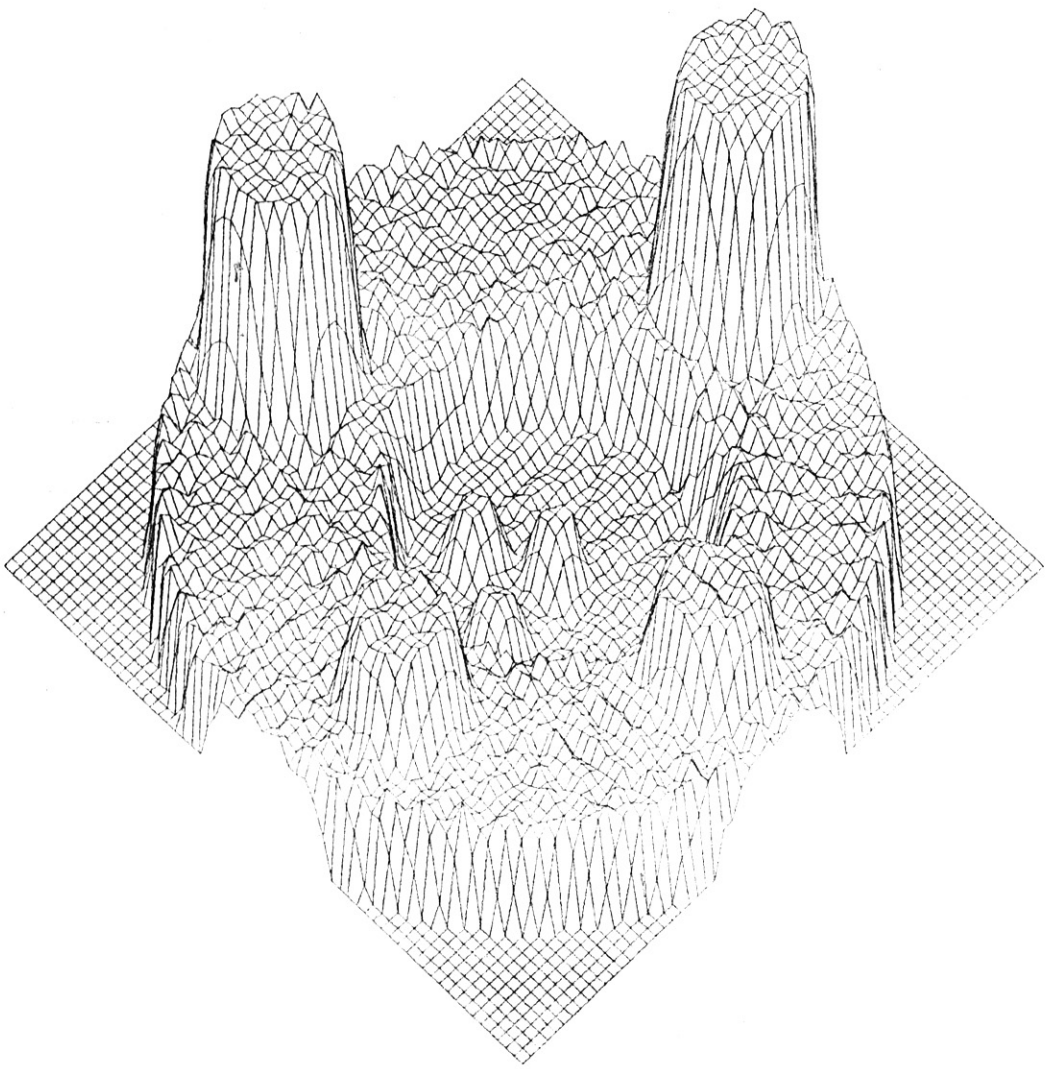


Fig. 5

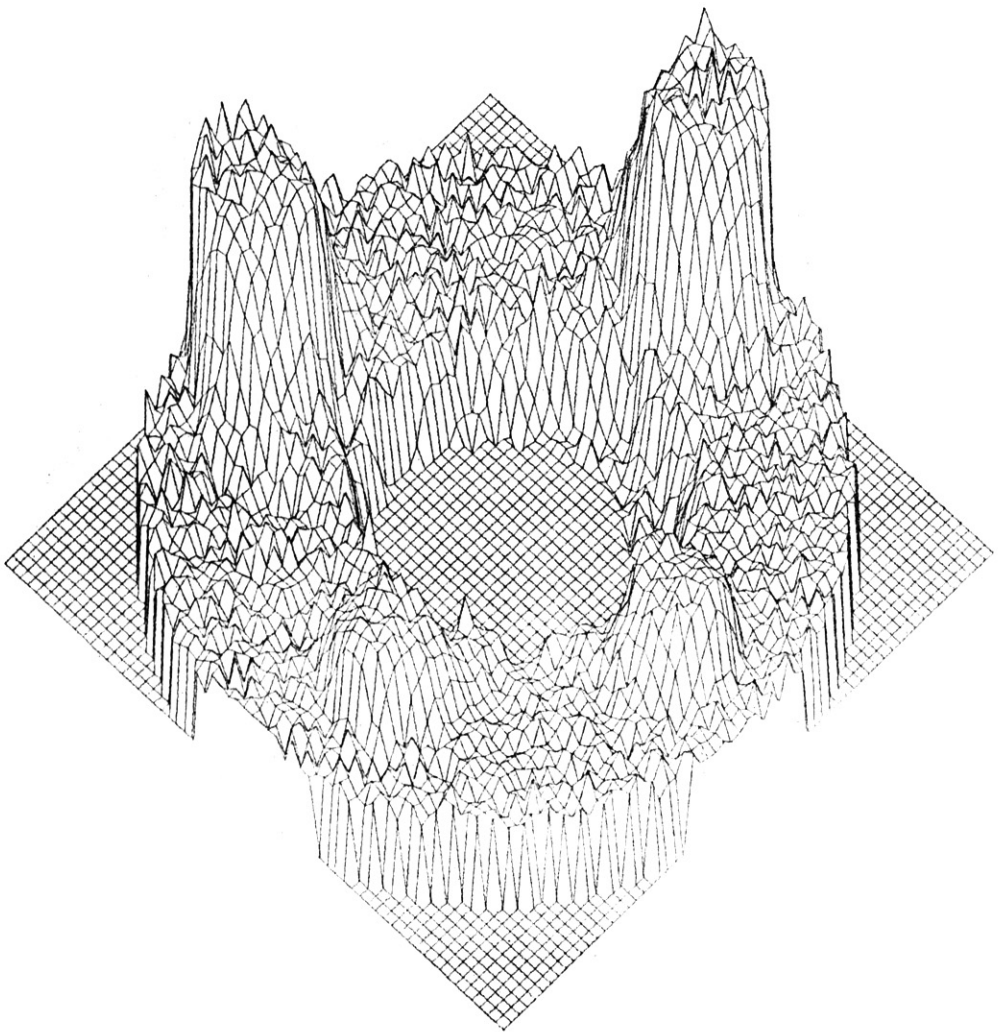
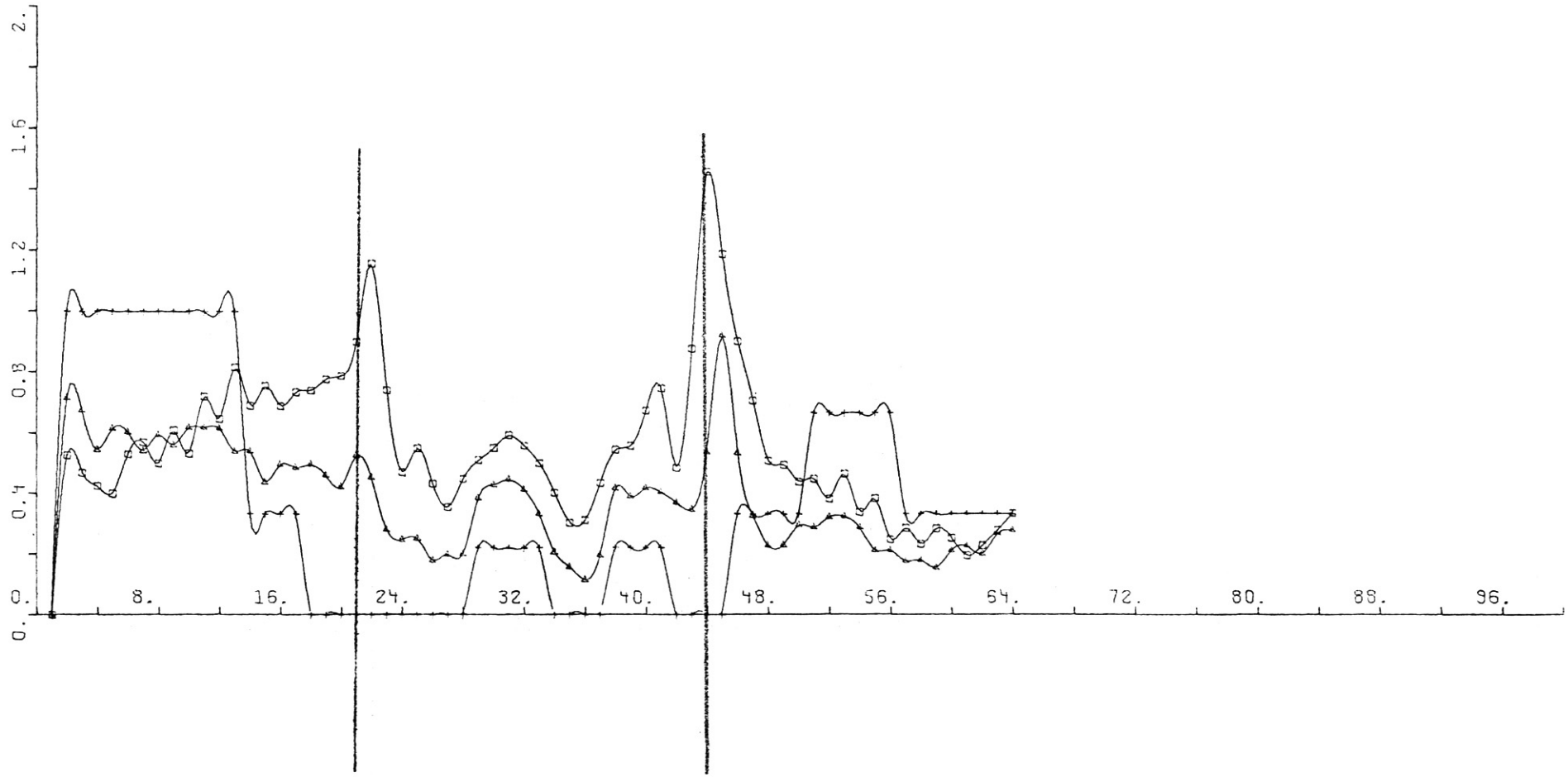


Fig. 6

SCHNITTBILD: ZEILE 30



○ RALA . R=0.1875
△ MNL . R=0.1875
\+ ORIGINAL, M=64

Fig. 7