# Hierarchical Reconstruction Using Geometry and Sinogram Restoration

Jerry L. Prince, *Member, IEEE,* and Alan S. Willsky, *Fellow, IEEE*

*Abstract*—**We describe and demonstrate a hierarchical reconstruction algorithm for use in noisy and limited-angle or sparse-angle tomography. The algorithm estimates an object's mass, center of mass, and convex hull from the available projections, and uses this information, along with fundamental mathematical constraints, to estimate a full set of smoothed projections. The mass and center of mass estimates are made using a least squares estimator derived from the principles of consistency of the Radon transform. The convex hull estimate is produced by first estimating the positions of support lines of the object from each available projection and then estimating the overall convex hull using prior shape information. Estimating the position of two support lines from a single projection is accomplished using a generalized likelihood ratio technique for estimating jumps in linear systems. We show results for simulated objects in a variety of measurement situations and discuss several possible extensions to the work.**

## I. Introduction

COMPUTED tomography (CT), the practice of reconstructing images of cross sections from measurements of their projections, has become an important tool in many areas of application, including nondestructive evaluation, sonar imaging, synthetic aperture imaging, and medical imaging. It is well known that in the full-data problem, where one is given enough high-quality projections over a 180° angular range, images of outstanding quality may be obtained. In many important practical cases, however, there is not enough data to obtain high-quality images using the usual techniques. In particular, the *limited-angle problem* occurs when projections are available over an angular range less than 180°, and the *sparse-angle problem* occurs when only a small number of angles evenly spaced over 180° are available.

The algorithm described in this paper uses a hierarchical approach to process the available measured projections in order to generate an estimate of the full set of projections, an image of which is called a *sinogram*. The estimated sinogram satisfies the two lowest-order constraints, which we call the mass and center of mass constraints, specified by the Ludwig–Helgason consistency conditions of the Radon transform [1], [2]. It is also consistent with an estimate of the convex support of the object derived directly from the measurements, in that its values are nearly zero for lines that miss the estimated convex support. Finally, it is smoothed by incorporating prior probabilistic knowledge using a Markov random field, optimally removing the contributions of noise. The object estimate is then produced using convolution back-projection applied to this estimated sinogram.

Limited-angle and sparse-angle problems occur often in practice. For example, a limited-angle problem occurs in cardiac CT imaging, where the carriage containing the X-ray emitters and detectors can only travel part of the way through the full angular range before significant heart motion occurs [3]. In combustion analysis, the time and energy constraints are often so severe that acquiring a sparse data set is unavoidable [4]. Noise is another problem that arises in different ways, depending on the application. For example, in X-ray applications a low dose or high sample attenuation will cause the measurements to be dominated by the photon statistics of the X-rays [5]. In synthetic aperture radar problems, receiver noise, clutter, and jamming all contribute to the overall noise that obscures the desired signal [6].

Many researchers have proposed solutions to the limited-angle and sparse-angle problems, although few have also dealt explicitly with noise. Solutions tend to fall into two categories: transform techniques that incorporate little *a priori* information and finite series expansion methods that may incorporate *a priori* information as constraints or probabilities. The transform techniques are usually single-pass direct reconstructions while the finite series expansion methods are usually iterative. Among the transform techniques are the pseudoinverse of the 2-D Radon transform [7], [8], angle-dependent rho-filters [9], analytic continuation in the Fourier plane [10], and the method of squashing [11]. The finite series expansion methods include linear minimum variance methods [12], [13], projection onto convex sets (POCS) [14], [15], the Gerchberg-Papoulis algorithm [16], [17], iterative pseudo-inverse methods [15], [18], and Bayesian methods [19], [20]. POCS in particular has shown great flexibility in dealing with known geometric constraints and with noise. The basic principle of POCS is that each piece of *a priori* knowledge must be represented by a convex set onto which the current image (or sinogram) estimate can be projected. It has been shown that if the intersection of these convex sets is nonempty then the sequence of cyclic projections will

Manuscript received February 20, 1992; revised January 6, 1993. This work was supported by the National Science Foundation Grant MIP-9015281, the Office of Naval Research Grant N00014-91-J-1004, and the U.S. Army Research Office Grant DAAL03-92-G-0115. In addition, the work of J. L. Prince was supported in part by a U.S. Army Research Office Fellowship. The associate editor coordinating the review of this paper and approving it for publication was Prof. Yoram Bresler.

J. L. Prince is with the Image Analysis and Communications Laboratory, Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD 21218.

A. S. Willsky is with the Laboratory for Information and Decision Systems, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139.

IEEE Log Number 9208876.

converge weakly to a point in this intersection [21]. In addition to these two general approaches there are other approaches that depend upon severely restricting the class of objects to be reconstructed. For example, Rossi and Willsky [22] use hierarchical maximum-likelihood methods to estimate the position, radius, and eccentricity of objects with a known unit profile such as the unit disk. Soumekh [23], Chang and Shelton [24], and Fishburn *et al.* [25] have investigated reconstruction of binary objects from a small number of projections.

The hierarchical algorithm described in this paper was designed to address several problems common to many of the existing algorithms mentioned above. For example, one of the major problems in methods that iterate between the object and projection spaces is reprojection error [26]. This problem is particularly bad when noise is present in the measurements [13]. One elegant way to avoid this problem proposed by Kim, Kwak, and Park [27], [28], is to iterate entirely in projection-space, incorporating certain object constraints mathematically, without reprojection. Our method is also a projection-space method, although it is not a POCS method as in [27] and [28].

Another problem with many, but not all, of the existing limited-angle algorithms is that they do not account for noise in an optimal sense. Although it is possible to modify POCS to account for noise processes [29], it does not make optimal (in a Bayesian sense) use of known noise statistics together with *a priori* knowledge. Minimum variance and Bayesian methods do take noise processes into account, but rarely have the capability to account for additional geometric knowledge as is done in POCS and our method. Parts of our method are Bayesian also, using the maximum *a posteriori* criterion, which specifies optimum solutions accounting for both the noise process and *a priori* knowledge.

Finally, a vital part of many POCS algorithms, the Gerchberg–Papoulis algorithm, and some iterative pseudoinverse approaches is the necessity of having known convex constraints, and in particular, having knowledge of the object support or convex support. Medoff [15] notes that one way to acquire this information is to have "a radiologist determine the outer boundary of the object." Aside from the fact that this is a potentially time-consuming process for a busy radiologist, the image that the radiologist uses to generate this boundary must be created before any correction has taken place, and therefore can be expected to be rife with artifacts from both noise and limited-angle measurement geometry. One has therefore introduced a potential source of error in this seemingly innocuous step. Our method instead estimates the convex support of the object directly from the projection measurements using prior probabilistic knowledge of the general shape to be expected. We then use this estimate to assist in a projection-space reconstruction algorithm.

## II. Overview of Hierarchical Approach

Referring to the geometry of Fig. 1, we define the 2-D Radon transform by [30]

$$g(t,\theta) = \int_{\boldsymbol{x}\in\mathbf{R}^2} f(\boldsymbol{x})\delta\left(t-\boldsymbol{\omega}^T\boldsymbol{x}\right)d\boldsymbol{x} \tag{1}$$

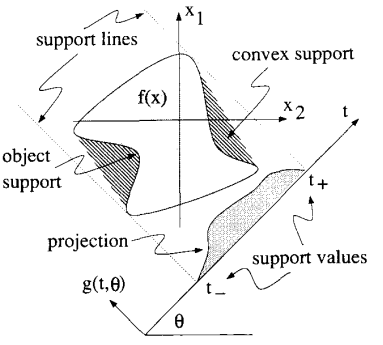



Fig. 1. The geometry of the 2-D Radon transform.

where $\boldsymbol{\omega} = [\cos\theta\ \sin\theta]^T$, $\theta$ is the angle measured counterclockwise from the horizontal-axis, $\delta(\cdot)$ is the Dirac delta function, and $f(\boldsymbol{x})$ is a function of the two-dimensional vector $\boldsymbol{x}$. For this paper, we assume $f(\boldsymbol{x})$ to be a real function defined on the disk of radius $T$ centered at the origin. For a particular $\theta$ the function $g(t,\theta)$ is a function of $t$ and is called a *parallel ray projection* or just a *projection*. A *sinogram* is an image of the 2-D Radon transform, where $t$ and $\theta$ form the vertical and horizontal axes, respectively, of a cartesian coordinate system. Because of the periodicity of the 2-D Radon transform and because of the assumed domain of $f(\boldsymbol{x})$, the sinogram is completely characterized by knowledge of $g(t,\theta)$ over the domain

$$\mathcal{Y}_T = \{(t,\theta)|t\in[-T,T],\ \theta\in[\pi/2],3\pi/2)\} \tag{2}$$

Two objects and their sinograms are displayed in Fig. 2. Note that the columns of the sinograms are projections, with the left-most projection arising from horizontal line integrals. In most real problems, we expect to have a discrete version of a sinogram, sampled for many values of $t$ and $\theta$. We define a *finest-grain sinogram* to be one that is known over the rectilinear lattice of $n_d$ (odd) uniformly spaced points in the $t$-direction and $n_v$ uniformly spaced points in the $\theta$-direction. Our observations, both limited-angle and sparse-angle, consist of measured (and therefore possibly noisy) sinogram values over a subset $\mathcal{Y}_O$ of this finest-grain lattice. More precisely, we assume that the observations are subject to noise and are given by

$$y(t_i,\theta_j) = g(t_i,\theta_j) + n(t_i,\theta_j) \tag{3}$$

where the indices $i\in\{1,\ldots,n_d\}$ and $j\in\mathcal{J}\subset\{1,\ldots,n_v\}$ index points on the regular rectangular lattice in the domain $\mathcal{Y}_T$. The set $\mathcal{J}$ contains $J$ indices of the angular positions of the observed projections. The noise samples $n(t_i,\theta_j)$ are zero-mean white Gaussian random variables, independent between lattice sites. Our goal is to reconstruct a good representation of $f(\boldsymbol{x})$ given these measured values.

In [31] we presented an algorithm that restores a finest-grain sinogram given measurements on $\mathcal{Y}_O$. The method assumes knowledge of the convex support of the object to

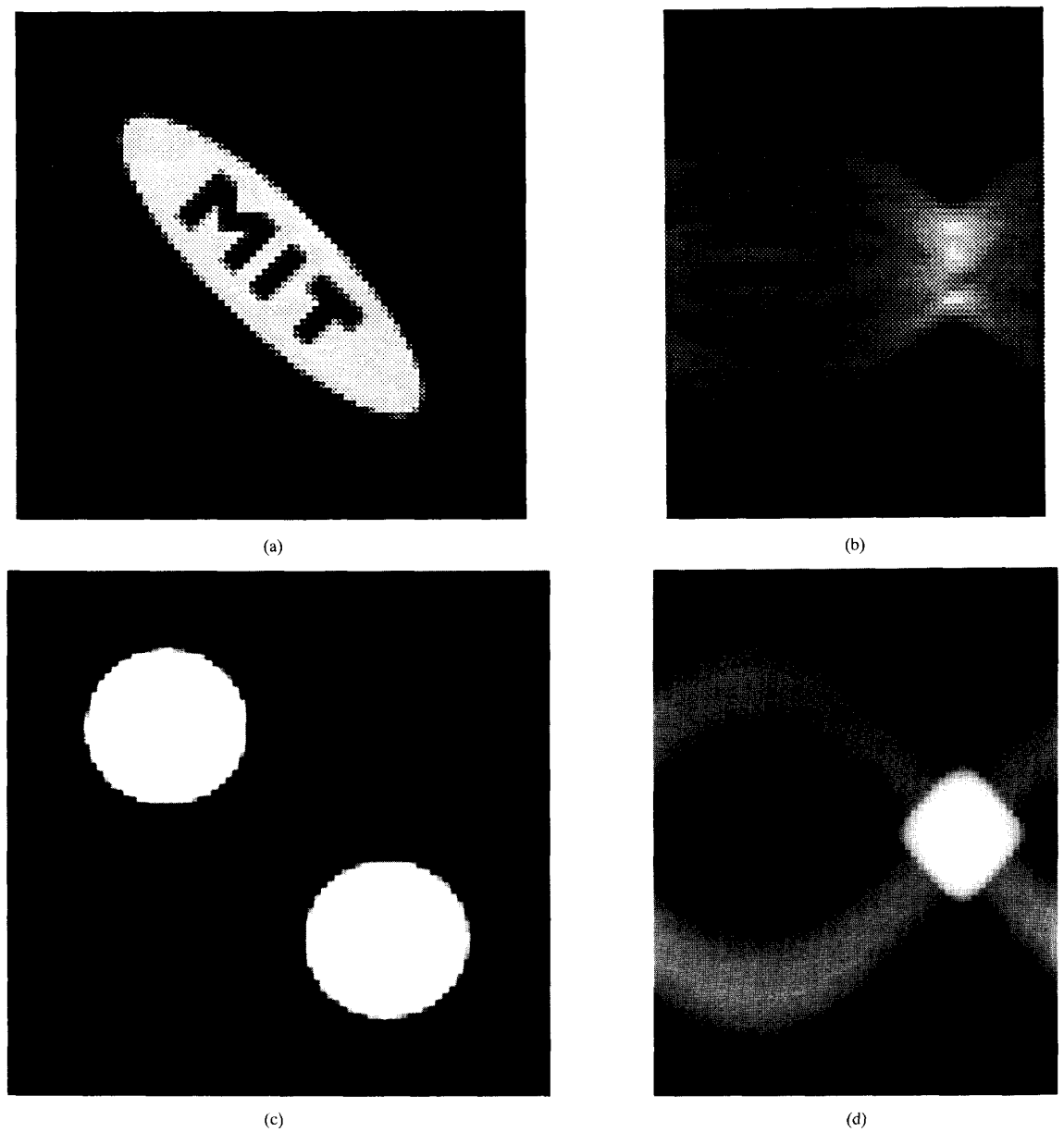


Fig. 2. (a) The MIT-ellipse object and (b) its sinogram. (c) The two-disk object and (d) its sinogram.

be reconstructed and assumes that the object is centered at the origin and has unit mass. It then specifies a Markov random field description of sinograms which incorporates this prior geometric knowledge and allows optimal recovery of the full sinogram using Bayesian techniques. The object itself is then reconstructed using convolution backprojection. The need for this detailed prior geometric information, however, precludes its use as a stand-alone procedure for processing raw projection data. This paper addresses this problem by presenting methods to estimate the mass, center of mass, and convex support directly from the available projection data, thus providing a complete hierarchical reconstruction algorithm.

An object generally is not centered at the origin and generally does not have unit mass. If the mass and center of mass are known, however, the raw projections can be easily modified to appear to have been acquired from such an object. In Section II we present methods to estimate the mass and center of mass from the available data and to scale and shift the projections to correspond to a unit mass object centered at the origin. The mass and center of mass estimation methods are based on the principle of least squares estimation.

In Section III we describe an approach that estimates those values of $t$ within each projection that prescribe the support lines of the object as viewed from the angle of the projection

(see Fig. 1). These values are called *support values,* and if measured perfectly for all angles $\theta$ they can be used to define the true convex support of the object. There are two problems, however: 1) because of noise, support values cannot be measured perfectly, and 2) projections are not available at all angles. The first problem can lead to inconsistent support value measurements for which there is no convex support that corresponds to the measured support values. The second problem implies that the convex support of the object (up to the resolution of the finest-grain sinogram) is not uniquely determined from the measured support values. These problems were addressed in [32] and [33], which presented methods to estimate the convex support from noisy and incomplete measurements of the support values of an object. In [32], the nature of support value inconsistency was demonstrated, and a maximum-likelihood method to estimate consistent support values given complete observations was presented. Extension was made to incomplete data in [33], where prior knowledge of the shape of the convex support was used in a maximum *a posteriori* formulation. The resultant collection of consistent support values is called a *support vector.* The largest convex set having support values determined by the support vector is then used as an estimate of the convex support of the object.

Given this background, it is easy to see how a hierarchical reconstruction algorithm can be put together. In the first phase, the mass and center of mass are calculated from the raw projection data. The center of mass is then used to adjust the sinogram data to correspond to an object centered at the origin; the measured projections are divided by the estimated mass to form a unit mass object. In the second phase, the support values are estimated from the available (adjusted) projections. These support values are used to produce a support vector which immediately gives an estimate of the object's support. The third phase restores the finest-grain sinogram using this derived geometric information, and the fourth phase reconstructs the object using convolution backprojection. The following three sections give a more detailed presentation of the ingredients comprising the hierarchical algorithm.

## III. Mass and Center of Mass Estimation

### A. Mass and Center of Mass

Conventional tomographic reconstruction algorithms, such as convolution backprojection (CBP) [5], which attempt to invert the Radon transform, require the availability of a complete set of projection data in order for the inversion to be possible. Consequently, their use in limited-angle or sparse-angle situations requires that some accommodation be made for the missing data. The simplest approach is in essence to set the missing measurement values to zero by applying the inversion operator only over the available measurement set. Such an approach is well known to produce a severely distorted reconstruction, and while other simple schemes typically lead to some improvement, serious degradations are still present. Similarly, the presence of significant measurement errors in the projection data can lead to pronounced distortions or artifacts in the resulting reconstructions [34]. One of the reasons for the presence and level of severity of degradations in each of these cases is that the inversion operation is being applied to a data set that could *not* be the Radon transform of *any* object. Specifically, a function $g(t,\theta)$ that is a valid 2-D Radon transform and satisfies certain mild regularity conditions must also satisfy the Ludwig–Helgason consistency conditions [1], [2], which specify an infinite set of constraints on certain moments of $g(t,\theta)$.

It is possible to use a large number of Ludwig–Helgason moment constraints in a reconstruction strategy (cf. [8], [35]), but in this paper we concentrate on just the two lowest order constraints. These constraints give rise to the *mass* and *center of mass* sinogram constraints, which are given by

$$m(\theta) = \int_{-\infty}^{\infty} g(t,\theta)\, dt = m \tag{4}$$

and

$$c(\theta) = \frac{1}{m}\int_{-\infty}^{\infty} g(t,\theta) t\, dt = \boldsymbol{c}^T\boldsymbol{\omega} = c_1\cos\theta + c_2\sin\theta \tag{5}$$

where

$$m = \int_{\boldsymbol{x}\in\mathbb{R}^2} f(\boldsymbol{x})\, d\boldsymbol{x} \tag{6}$$

and

$$\boldsymbol{c} = [c_1 \quad c_2]^T = \frac{1}{m}\int_{\boldsymbol{x}\in\mathbb{R}^2} \boldsymbol{x} f(\boldsymbol{x})\, d\boldsymbol{x}\,. \tag{7}$$

The sinogram restoration algorithm, which comprises a part of the overall hierarchical algorithm, uses these two constraints in order to guarantee a consistent sinogram estimate, which significantly reduces the degradations and artifacts caused by noisy and/or missing data. These constraints also have simple geometric interpretations, and $m$ and $\boldsymbol{c}$ can be easily and reliably estimated from the available projections, as described below.

### B. Mass and Center of Mass Estimation

The mass and center of mass of the object are estimated directly from the projections using least squares. We assume that the mass constraint given by (4) may be approximated by the summation

$$\frac{2T}{n_d}\sum_{i=1}^{n_d} g(t_i,\theta_j) = m \quad \forall\, j\,. \tag{8}$$

Substituting the observed projections $y(t_i,\theta_j)$ for the true projections $g(t_i,\theta_j)$ yields a system of equations that may be solved for $m$ using least squares. Accordingly, the mass estimate is

$$\hat{m} = \frac{2T}{n_d J}\sum_{j\in\mathcal{J}}\sum_{i=1}^{n_d} y(t_i,\theta_j) \tag{9}$$

which is proportional to the average of all the observed line integrals.

To estimate the center of mass, we first approximate the integral in (5) by the summation

$$c_j = \frac{1}{m}\frac{2T}{n_d}\sum_{i=1}^{n_d} t_i g(t_i,\theta_j)\,. \tag{10}$$

Substituting the mass estimate $\hat{m}$ for $m$, the observed projections $y(t_i, \theta_j)$ for $g(t_i, \theta_j)$, and $\boldsymbol{c} \cdot \boldsymbol{\omega}_j$ for $c_j$ yields a system of linear equations with unknown $\boldsymbol{c}$, the center of mass of the object. This system may also be solved using least squares, yielding the center of mass estimate

$$\hat{\boldsymbol{c}} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{b} \tag{11}$$

where

$$\boldsymbol{A} = \begin{bmatrix} \cos\theta_1 & \sin\theta_1 \\ \vdots & \vdots \\ \cos\theta_J & \sin\theta_J \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} \tilde{c}_1 \\ \vdots \\ \tilde{c}_J \end{bmatrix}$$

and

$$\tilde{c}_j = \frac{1}{\hat{m}} \frac{2T}{n_d} \sum_{i=1}^{n_d} t_i y(t_i, \theta_j) .$$

### C. Object Centering and Mass Normalization

Using the mass and center of mass estimates, the measured projections are shifted and scaled so that the center of mass of the object is at the origin and the mass is unity. The modified projections are given by

$$\tilde{y}(t, \theta) = \frac{1}{\hat{m}} y(t - \hat{\boldsymbol{c}} \cdot \boldsymbol{\omega}, \theta) . \tag{12}$$

This processing is done so that the convex support estimation stage and sinogram estimation stage may assume the object to have unit mass and to be centered at the origin. After the sinogram processing, the full set of estimated projections are shifted back using a similar equation.

## IV. Support Value Measurement

The convex support of an object is estimated in two stages. The first stage measures two support values from each available projection, and the second stage estimates a complete, consistent collection of support values. In this section we present an algorithm for support value measurement based on detection of jumps in linear systems. A method to adaptively estimate the error in these measurements is also presented, and simulation results are presented. The second stage, outlined in Section V, estimates a complete feasible support vector from these support value measurements using the methods of [32] or [33]. We begin with a more precise discussion of what is meant by convex support.

### A. Convex Support

One very useful way in which to interpret the Radon transform consistency constraints is that they provide prior information that in essence reduces the number of degrees of freedom that must be recovered from the measurement data. A second piece of prior information which can be of significant value for the same reason is knowledge about the support $\mathcal{F}$ of the function $f(\boldsymbol{x})$ to be reconstructed, i.e., the set of points where $f(\boldsymbol{x})$ may be nonzero. In particular, our algorithm makes use of information about the *convex support* of $f(\boldsymbol{x})$, i.e., the convex hull of $\mathcal{F}$, $\mathcal{F}_c = \mathrm{hul}(\mathcal{F})$. For any set $S$, the *support value* of $S$ at angle $\theta$ is the maximum projection of points in $S$ onto the $\boldsymbol{\omega}$-axis, i.e.,

$$h(\theta) = \sup_{\boldsymbol{x} \in S} \boldsymbol{x}^T\boldsymbol{\omega} . \tag{13}$$

Treated as a function of $\theta$, $h(\theta)$ is known as the *support function* of $S$.

As shown in Fig. 1 we see that for a fixed $\theta$, the projection $g(t, \theta)$ has support confined to the interval between the two points $t_+(\theta)$ and $t_-(\theta)$, which are related to the support function as follows:

$$h(\theta) = \begin{cases} t_+(\theta), & 0 \le \theta < \pi \\ -t_-(\theta - \pi) & \pi \le \theta < 2\pi . \end{cases} \tag{14}$$

This collection of support values, or equivalently the support function, segment a sinogram into a region of support $\mathcal{G}$ for $g(t, \theta)$ and its complement, $\overline{\mathcal{G}}$, where

$$\mathcal{G} = \{(t, \theta) \in \mathcal{Y}_T | t_-(\theta) \le t \le t_+(\theta)\} . \tag{15}$$

Therefore, for a given object support set $\mathcal{F}$, we think of $\mathcal{G}$ as the matching region of support in Radon space. This region is readily identified in Fig. 2(b) as the region that is not black—i.e., where the sinogram values are nonzero. If, however, the support of $f(\boldsymbol{x})$ is not a connected set, as in Fig. 2(c), it is possible for sinogram values within $\mathcal{G}$ to be zero, as in Fig. 2(d). Therefore, $\mathcal{G}$ is not necessarily the actual support of $g(t, \theta)$, but it does contain all the points $(t, \theta)$ for which $g(t, \theta)$ is nonzero. Although $\mathcal{F}$ uniquely determines $\mathcal{G}$, it is clear from the geometry (see Fig. 1) that $\mathcal{G}$ uniquely determines only $\mathrm{hul}(\mathcal{F})$, not $\mathcal{F}$ itself. This is why in this paper we are primarily concerned with the convex support of $f(\boldsymbol{x})$, since this is what may be determined directly from knowledge of $\mathcal{G}$, which may in turn be estimated directly from the projections.

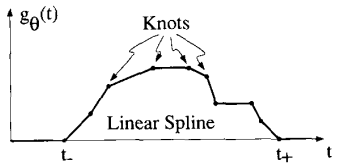



Fig. 3. A projection modeled as a linear spline with knots.

### B. Knot-Location Method

In this section we model a projection $g(t)$ as a continuous piecewise-linear waveform as shown in Fig. 3. Such a function is called a linear *spline*; it is composed of a set of linear functions that connect a series of points called *knots*, so that the resultant function is continuous but its slope has an abrupt change at each of the knots. Two support values for this projection are determined by estimating the positions, $t_-$ and $t_+$, of the two outer knots.

Our *knot-location method* is based on the generalized likelihood ratio (GLR) techniques developed by Willsky and Jones [36] for detecting abrupt changes in dynamic systems, and later applied to spline estimation by Mier-Muth and Willsky [37]. (It is worth noting that other computationally efficient methods

for maximum likelihood estimation of knots exist [38], [39].) Our approach is to run a Kalman filter, starting at $t = -T$, assuming an underlying signal model corresponding to a *linear ramp* waveform (initialized with slope = 0). The innovations of the Kalman filter, which should be a zero-mean white Gaussian sequence, provide the basis for estimating the point at which the true signal deviates from the assumed model—i.e., the point at which the slope changes suddenly. The first such point yields the point $t_-$. The point $t_+$ is found by running the filter backwards starting at $t = T$ and finding the first knot in the backwards direction.

As discussed in [37], a linear spline with a single knot can be described by the following two-dimensional discrete state equation

$$\boldsymbol{x}(i+1) = \boldsymbol{\Phi}\boldsymbol{x}(i) + \alpha\delta(i+1-k)\boldsymbol{f} \tag{16}$$

where $\delta(\cdot)$ is the discrete impulse function, $\alpha$ is the height of the discontinuity, $k$ is the discrete position of the knot $(1 \le k \le n_d)$, and

$$\boldsymbol{\Phi} = \begin{bmatrix} 1 & 2T/n_d \\ 0 & 1 \end{bmatrix}, \qquad \boldsymbol{f} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

The problem of fitting such a linear spline to an observed data sequence $y(i)$ corresponds to estimating the parameters of (16)—namely the initial conditions and slope, and the knot location $k$ and jump height $\alpha$—assuming that the data are noisy measurement of the spline, i.e.,

$$y(i) = \boldsymbol{h}^T\boldsymbol{x}(i) + v(i) \tag{17}$$

where $\boldsymbol{h}^T = [1\ 0]$, and $v(i)$ are zero-mean white jointly Gaussian random variables with variance $R = \sigma^2$. In our problem the $y(i)$ represent the projection measurements $y(t_i, \theta_j)$ as a function of $i$ for each fixed $\theta_j$.

The first step in the knot location algorithm is to run the following Kalman filter on the data:

$$\hat{\boldsymbol{x}}(i|i-1) = \boldsymbol{\Phi}\hat{\boldsymbol{x}}(i-1|i-1) \tag{18a}$$

$$\hat{\boldsymbol{x}}(i|i) = \hat{\boldsymbol{x}}(i|i-1) + \boldsymbol{K}(i)\gamma(i) \tag{18b}$$

$$\gamma(i) = y(i) - \boldsymbol{h}^T\hat{\boldsymbol{x}}(i|i-1) \tag{18c}$$

where $\hat{\boldsymbol{x}}(i|\,l)$ is the best estimate of $\boldsymbol{x}(i)$ given $y(1), \cdots, y(l)$, $\gamma(i)$ is the innovations sequence, and $\boldsymbol{K}(i)$ is the Kalman filter gain. Assuming that there is no jump (slope discontinuity), the innovations sequence is a zero-mean, white, jointly Gaussian random sequence whose variance $V(i)$ is given by

$$V(i) = \boldsymbol{h}^T\boldsymbol{P}(i|i-1)\boldsymbol{h} + R$$

where the error covariance $\boldsymbol{P}(i|i-1)$ may be computed together with the Kalman gain $\boldsymbol{K}(i)$ using the following recursive algorithm:

$$\boldsymbol{P}(i|i) = \left[\boldsymbol{I} - \boldsymbol{K}(i)\boldsymbol{h}^T\right]\boldsymbol{P}(i|i-1) \tag{19a}$$

$$\boldsymbol{K}(i) = \boldsymbol{P}(i|i-1)\boldsymbol{h}V^{-1}(i) \tag{19b}$$

$$\boldsymbol{P}(i+1|i) = \boldsymbol{\Phi}\boldsymbol{P}(i|i)\boldsymbol{\Phi}^T. \tag{19c}$$

Because a projection is zero outside the disk of radius $T$ and since we may take $T$ to be as large as necessary, we know with certainty the initial spline parameters and thus initialize the Kalman filter as follows:

$$\hat{\boldsymbol{x}}(0|0) = \hat{\boldsymbol{x}}(1|\,0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\boldsymbol{P}(0|0) = \boldsymbol{P}(1|0) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

In order to determine whether a jump has occurred we examine the innovations sequence which will deviate from the statistics given above if a jump takes place. In particular, given our single jump model of (16) it can be shown that the true innovations sequence takes the form [40]

$$\gamma(i) = \boldsymbol{G}(i,k)\boldsymbol{f}\alpha + \tilde{\gamma}(i), \tag{20}$$

where $\tilde{\gamma}(i)$ is the zero-mean, white, and Gaussian innovations if there is no jump and $\boldsymbol{G}(i,k)$ is the *jump signature matrix* given by

$$\boldsymbol{G}(i,k) = \boldsymbol{h}^T\left[\boldsymbol{\Phi}^{i-k} - \boldsymbol{\Phi}\boldsymbol{F}(i-1,k)\right] \tag{21a}$$

$$\boldsymbol{F}(i,k) = \boldsymbol{K}(i)\boldsymbol{G}(i,k) + \boldsymbol{\Phi}\boldsymbol{F}(i-1,k) \tag{21b}$$

where $\boldsymbol{G}(i,k)$ and $\boldsymbol{F}(i,k)$ are both 0 for $i < k$ and $\boldsymbol{F}(i,i) = \boldsymbol{K}(i)\boldsymbol{h}^T$.

Equation (20) is the key to the GLR knot-location method. Through this equation, we see how to form the ML estimate of $\alpha$, assuming a jump occurred at time $k$ in the filter's past for each current time index of $i$ of the Kalman filter. Actually, to reduce the required computation, at each point $i$ we look for possible jumps only over a trailing window $W(i) = \{i-1, i-2, \cdots, i-N\}$ of length $N$ in the filter's past. Then using the ML estimate of $\alpha$ for each $k \in W(i)$, we form the GLR for the hypothesis that a jump actually occurred at $k$. If the GLR exceeds a preset threshold then a jump is deemed to have occurred. The above calculations are given by the following equations each evaluated for all $k \in W(i)$ [37]

$$\boldsymbol{C}(i,k) = \sum_{j=1}^{i} \boldsymbol{G}^T(j,k)V^{-1}(j)\boldsymbol{G}(j,k) \tag{22a}$$

$$\boldsymbol{d}(i,k) = \sum_{j=k}^{i} \boldsymbol{G}^T(j,k)V^{-1}(j)\gamma(j) \tag{22b}$$

$$\hat{\alpha}(i,k) = \frac{\boldsymbol{f}^T\boldsymbol{d}(i,k)}{\boldsymbol{f}^T\boldsymbol{C}(i,k)\boldsymbol{f}} \tag{22c}$$

$$l(i,k) = \frac{\left(\boldsymbol{f}^T\boldsymbol{d}(i,k)\right)^2}{\boldsymbol{f}^T\boldsymbol{C}(i,k)\boldsymbol{f}} \tag{22d}$$

where $\hat{\alpha}(i,k)$ is the ML estimate of $\alpha$ assuming that a jump occurred at time $k$, and $l(i,k)$ is the logarithm of the generalized likelihood ratio for this event. The best estimate of the *location* of a jump is then given by

$$\hat{k}(i) = \underset{k \in W(i)}{\operatorname{argmax}}\, l(i,k). \tag{23}$$

Then, to decide whether a jump has actually taken place we use the following threshold rule

$$l(i,\hat{k}(i)) \begin{matrix} > \varepsilon & \text{Jump} \\ \le \varepsilon & \text{No Jump} \end{matrix} \tag{24}$$

### C. Jump Threshold

Specification of the GLR threshold $\varepsilon$ is an important consideration since if too low, noise will often produce an inaccurate knot estimate, and if too high, it is possible that no knot will be found in the entire projection. We now describe an adaptive method to choose this threshold that has worked well in practice.

Since the mass of each projection is the same, we might expect that a projection with a small support width would rise rapidly at the support values in order to include the required mass, and this would correspond to a large value of $\alpha$ and a corresponding large value of $l(i, \hat{k}(i))$. In contrast, a projection with a larger support width might rise less rapidly, and would correspond to a smaller value of $l(i, \hat{k}(i))$. A simple estimate of the width of a projection is given by the approximate second moment of a normalized (to unit mass), shift-corrected projection as follows

$$m_2(\theta_j) = \sum_{i=1}^{n_d} t_i^2 \max\{0, \tilde{y}(t_i, \theta_j)\}. \tag{25}$$

The $\max\{\ \}$ function is included since it possible that elements of $\tilde{y}$ are negative, and therefore that $m_2(\theta)$ might otherwise be negative.

The second moment in (25) is roughly equivalent to a variance calculation, and the quantity

$$p(\theta) = \sqrt{m_2(\theta)} \tag{26}$$

is analogous to a standard deviation, which serves as an approximate measure of the width of the projection. If $p$ is large, then the projection is wide and the slope change at the support value is probably small. Therefore, we want to specify a GLR threshold $\varepsilon$ that is relatively small. Using similar reasoning we conclude that for small $p$, the threshold $\varepsilon$ should be large. After some experimentation we have chosen the function $\varepsilon(p)$ depicted in Fig. 4. Since $p$ is a measure of the width of the entire projection, $\varepsilon$ is used as the threshold value for both the forward and backward stages of the knot-location algorithm (i.e., for measuring both $t_-$ and $t_+$.)

### D. Performance

It is important in our hierarchical approach to be able to assess the performance of the support value measurement method so that this information may be used by the subsequent support vector estimation stage. The accuracy of the support value measurements clearly depends not only on the variance of the additive noise but also on the characteristics of the underlying projection, particularly at or around the true support value. Indeed if these characteristics were summarized, say in a template model, one could compute a Cramer–Rao bound specifying these dependencies in a quantitative manner. Unfortunately, these characteristics can vary widely from projection to projection even for the same object. Thus it is essential that we have a method for determining the quality of our support measurements directly from the projection data. In this section we present a method to obtain such

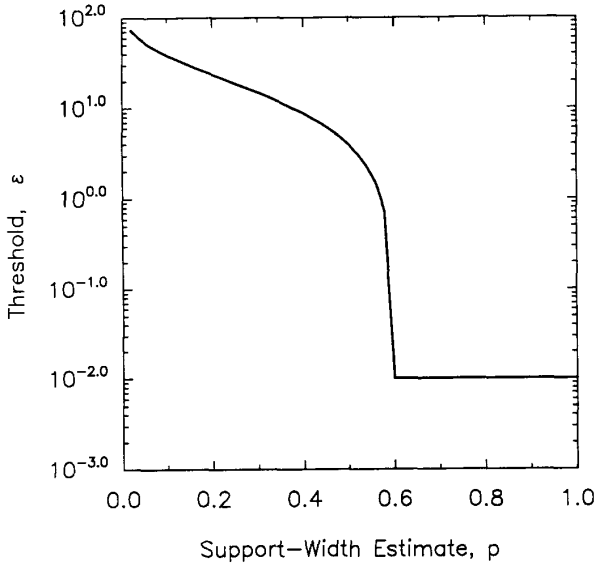



Fig. 4. Generalized likelihood ratio threshold selection curve.

an estimate of the error variance arising from the knot-location algorithm described above. In doing this we do not assume any prior shape information at this stage, so our estimate is made completely on the basis of statistics available during processing, and, in particular, on the shape of the log-likelihood function. If the log-likelihood function is sharply peaked at its maximum then we presume that it is a good estimate; if it has a shallow maximum then we presume that the estimate is not as good. Following this principle, we fit a downturned quadratic centered at $\hat{k}$ to the log-likelihood function that was evaluated over the window $W(i)$, and the coefficient of the quadratic term yields our error variance estimate. It is worth noting that this quadratic fitting method can be viewed as a signal-adaptive estimate of the Cramer–Rao bound.

Let $\hat{k} \in W(i)$ be our estimate of the knot location, made when the Kalman filter has progressed to the $i$th index. We wish to fit a downturned quadratic of the form

$$\hat{l}(k) = -a(k - \hat{k})^2 + c$$

to the data $l(i, k)$ so that, in particular, we may determine $a$. To make this fit, we minimize

$$\sum_{k=1}^{N} (\hat{l}(k) - l(i,k))^2 = \sum_{k=1}^{N} \left(-a(k - \hat{k})^2 + c - l(i,k)\right)^2$$

with respect to $a$ and $c$, yielding

$$a = \frac{\sum_{k=1}^{N} (c - l(i,k))(k - \hat{k})^2}{\sum_{k=1}^{N} (k - \hat{k})^4},$$

$$c = l(i, \hat{k}).$$

Our estimate of error variance $\sigma_e^2$ for the support estimate is

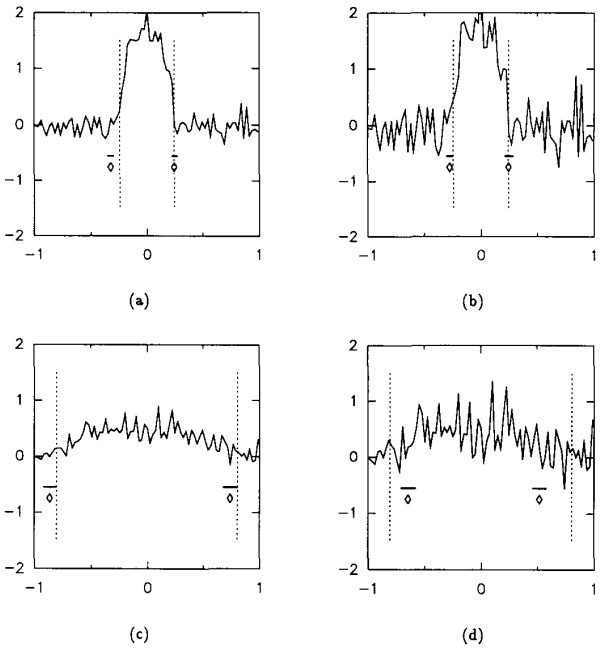



Fig. 5. Head-on projections of an ellipse with noise standard deviations (a) $\sigma = 0.1450$ and (b) $\sigma = 0.3245$. Broadside projections of an ellipse with noise standard deviations (c) $\sigma = 0.1450$ and (d) $\sigma = 0.3245$. Knot-location support value estimates are shown using diamond markers; the true support values are shown using vertical dotted lines.

then given by

$$\sigma_e^2 = \frac{1}{2|a|}. \tag{27}$$

### E. Simulations

Fig. 5 shows simulation results in which the knot-location algorithm is used to measure support values. All four panels show noisy projections of an ellipse centered at the origin, with major semiaxis radius of 0.806 and minor semiaxis radius of 0.242—these are the same dimensions as the MIT-ellipse shown in Fig. 2(a). Figs. 5(a) and 5(b) show the narrowest projection—the support values are $-0.242$ and $0.242$—with noise standard deviations 0.1450 and 0.3245, respectively. These noise levels represent the amount of noise required to achieve a signal-to-noise ratio (SNR) of 10 dB and 3 dB, respectively, with respect to the full sinogram, where SNR is defined in Section VI. Figs. 5(c) and 5(d) show the widest projection— the support values are $-0.806$ and $0.806$—also with noise standard deviations of 0.1450 and 0.3245. The positions of the true support values are indicated by the vertical dotted lines in each of the panels. The same underlying unit variance noise sequence was used for all four projections in each figure, which accounts for the similarity in the noise structure.

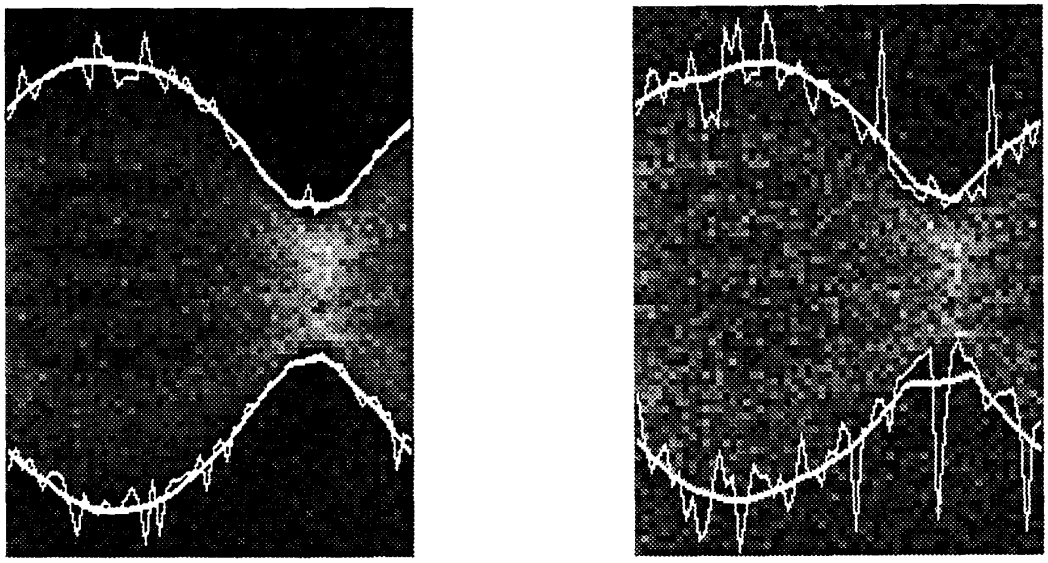
Support value measurements are indicated by diamond markers in each of the panels in Fig. 5. The error bar centered directly above each of these symbols has a length of two (error) standard deviations. In most cases, the estimated values are within three standard deviations of the true values; but, as one would expect, the size of the error increases as the noise variance grows. Also, the error bars are longer in Figs. 5(c) and 5(d) than those in 5(a) and 5(b). This agrees with our intuitive reasoning that it should be more difficult to detect the onset of a projection of the broadside of the ellipse versus the head-on projection, given the same noise variance (since the broadside projection has a more gradual rise than the head-on projection).

In Section II we noted that a collection of support value measurements may not be self-consistent. Fig. 6 shows two noisy MIT-ellipse sinograms together with the full set of knot-location support value measurements shown using thin white curves. The SNR of the underlying sinograms, using the definition of SNR given in Section VI, are given by (a) 10.0 dB and (b) 3.0 dB. The thick white curves represent the collection of support values prescribed by the support vector—i.e., the consistent set of support values—that is closest to the measured values. These curves also correspond to the constrained maximum likelihood (ML) support vector estimate [32]. Clearly, the thick curves are different than the underlying thin curves, which means that the measured support values are inconsistent in each panel of this figure. The set defined between the top and bottom thick white lines constitutes $\hat{\mathcal{G}}$, which should ideally contain all pixels with nonzero values. The performance of the knot-location algorithm is noticeably worse in Fig. 6(b), but the constrained ML support vector estimate does not show the same qualitative degradation. Since the constrained ML algorithm does not use any prior geometric information, this shows that utilization of support vector consistency alone adds considerable robustness to errors in the knot-location measurement process.

## V. Hierarchical Algorithm

A block diagram of our hierarchical algorithm is shown in Fig. 7. Each block represents a significant stage which either estimates a new parameter or set of parameters or transforms the data in some fashion; blocks that correspond to prior work are shaded. Where required, an estimate of the reliability of the information is also passed between the blocks. In this way, poor estimates are not viewed as perfect by subsequent processing stages, and extremely good estimates are given greater weight in subsequent blocks.

As shown in Fig. 7, the overall processing is divided into four stages: (a) sinogram preconditioning, (b) sinogram restoration, (c) sinogram postconditioning, and (d) object reconstruction. In (a) the mass and center of mass are estimated directly from the available measured projections as described in Section III. These quantities are used to center the coordinate system by shifting each projection to correspond to an unit mass object centered at the origin, yielding $\tilde{y}$. Stage (b) forms the bulk of the processing, with support vector estimation and sinogram restoration—two methods reported in previous work—and two new steps required to support these procedures. Here, the block labeled *support*



(a) (b)

Fig. 6. (a) 10.0 dB and (b) 3.0 dB noisy MIT-ellipse sinograms overlayed with the knot-location support value measurement (thin white curves) and the constrained maximum likelihood support vector estimate (thick white curves).

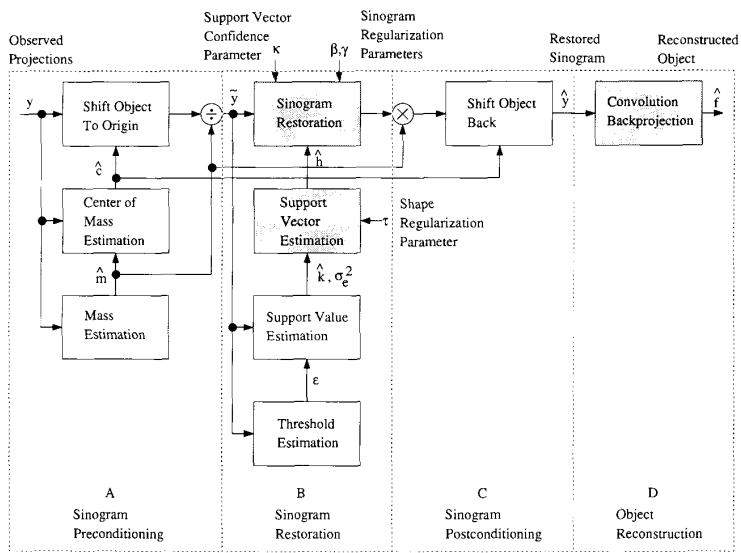



Fig. 7. Block diagram of the hierarchical reconstruction algorithm.

*vector estimation* estimates a support vector $\hat{h}$, which is used to define a corresponding segmentation of the sinogram. This segmentation is used by the block labeled *sinogram restoration* to estimate a full sinogram. Support value estimates, required for support vector estimation are provided by the new block labeled *support value estimation,* as described in Section IV. The knot-location threshold is determined adaptively in the block labeled *threshold estimation.* Stage (c) takes the restored sinogram as input and rescales its mass to the estimated mass and restores the coordinate system to the estimated center of mass. Finally, stage (d) performs convolution backprojection on the restored sinogram to reconstruct an object estimate $\hat{f}$.

The hierarchical algorithm requires several user inputs (see Fig. 7): $\gamma$, $\beta$, $\tau$, and $\kappa$. The parameters $\gamma$ and $\beta$ specify sinogram horizontal and vertical smoothness, respectively, as described below. Currently, these parameters must be fixed *a priori*, although in principle they could be estimated hierarchically using any one of several parameter estimation schemes [41], [42]. The parameter $\tau$ symbolizes prior shape information which is required for support vector estimation from incomplete data; it includes both the choice of a method and one or more parameters associated with the chosen method. The parameter $\kappa$ specifies a measure of confidence in the support vector estimate. A large $\kappa$ indicates great confidence

in $\hat{\mathcal{G}}$; a small $\kappa$ indicates little confidence in $\hat{\mathcal{G}}$. The overall performance of the hierarchical algorithm is affected by how well the parameters we use represent the truth; in our experiments they are empirically adjusted to match the class of objects and imaging geometry used in the experiments. (See [32] and [33] for detailed discussions of the effects of parameter selection on the MAP sinogram estimation algorithm and the support vector estimation algorithm, respectively.)

For completeness we now provide a brief description of two of the blocks that are derived from other works: sinogram restoration and support vector estimation.

### A. Sinogram Restoration [31]

We take the restored sinogram to be the function $g(t,\theta)$ that minimizes

$$I = \iint_{\mathcal{Y}_O} \frac{1}{2\sigma^2}(y-g)^2\,dt\,d\theta + \iint_{\overline{\mathcal{G}}} \kappa g^2\,dt\,d\theta + \iint_{\mathcal{Y}_T} \left[\beta\left(\frac{\partial g}{\partial t}\right)^2 + \gamma\left(\frac{\partial g}{\partial \theta}\right)^2\right] dt\,d\theta \quad (28)$$

subject to the condition that each projection has unit mass and center of mass at the origin, and subject to the boundary conditions $g(T,\theta) = g(-T,\theta) = 0$ and $g(t,0) = g(-t,\pi)$. Here, $\kappa$, $\beta$, and $\gamma$ are positive constants. The first term in $I$, which integrates over the set $\mathcal{Y}_O$, represents a penalty that seeks to keep the estimate close to the observations. The second term integrates over the complement of the estimated region of support $\overline{\mathcal{G}}$ to attempt to keep sinogram values outside the region of support small. The final integral contains two terms involving the square of the two partial derivatives of $g$, which provides a smoothing effect in both the $t$ and $\theta$ directions. The boundary conditions indicate that line integrals are expected to be zero outside a disk of radius $T$ centered at the origin, and that the sinogram is periodic as prescribed in the Ludwig–Helgason consistency conditions. The numerical solution to this variational problem was shown in [31] to be equivalent to the maximum *a posteriori* solution of a probabilistic formulation in which the sinogram is modeled as a certain Markov random field.

It was shown in [31] that the restored sinogram $g(t,\theta)$ must satisfy the original constraints and boundary conditions and the following second order partial differential equation (PDE)

$$\left(2\kappa\overline{\mathcal{X}}_G + \frac{1}{\sigma^2}\mathcal{X}_Y\right)g - 2\beta\frac{\partial^2 g}{\partial t^2} - 2\gamma\frac{\partial^2 g}{\partial \theta^2} = \frac{1}{\sigma^2}\mathcal{X}_Y y - \lambda_1(\theta) - \lambda_2(\theta)t \quad (29)$$

where $\overline{\mathcal{X}}_G$ and $\mathcal{X}_Y$ are indicator functions of the sets $\overline{\mathcal{G}}$ and $\mathcal{Y}_O$, respectively. The solution must also satisfy the additional boundary condition $\partial g(t,0)/\partial t = \partial g(-t,\pi)/\partial t$. Equation (29) contains three unknown functions: $g(t,\theta)$ and two Lagrange multiplier functions $\lambda_1(\theta)$ and $\lambda_2(\theta)$, one for each constraint. If the Lagrange multiplier functions were known then the restored sinogram would simply be the solution of the PDE (29), which may be solved by any of several well known numerical methods. Although these functions are not known in general, good initial estimates of these Lagrange multipliers often exist [43]. Primal-dual optimization methods can then be used to iteratively converge to the solution, as described in [31].

### B. Support Vector Estimation [32], [33]

Since there are $n_v$ projections in the finest-grain sinogram and two support values per projection, each full sinogram determines $M = 2n_v$ support values which represent samples of the support function of an object at angles $\theta_i = 2\pi(i-1)/M$, $i = 1,\cdots,M$. A *support vector* $\boldsymbol{h}$ is a vector made by organizing the values of a support function $h(\theta)$ sampled at these angles, yielding

$$\boldsymbol{h} = [h(\theta_1) \quad h(\theta_2) \quad \cdots \quad h(\theta_M)]^T. \quad (30)$$

In [32] it was shown that a support vector must belong to a particular convex cone $\mathcal{C} \subset \mathbb{R}^M$. Therefore, not all vectors constructed from the support values measured from an observed sinogram can be expected to be support vectors: they may be inconsistent. Furthermore, incomplete data problems yield only measurements of a subset of the elements of $\boldsymbol{h}$.

The formulations in [32] and [33] reconstruct an $M$-dimensional support vector $\boldsymbol{h}$ based on $K \leq M$ noisy measurements. The measurements are modeled as

$$\tilde{\boldsymbol{z}} = \boldsymbol{S}\boldsymbol{h} + \boldsymbol{\nu} \quad (31)$$

where $\boldsymbol{\nu}$ is a zero-mean jointly Gaussian vector with covariance $\mathrm{diag}[(\sigma_e^2)_j]$ and $\boldsymbol{S}$ is a $K \times M$ "selection" matrix specifying the elements of $\boldsymbol{h}$ that are measured. The log-likelihood $l(\boldsymbol{h})$ is readily formed from this observation model; however, it has a unique maximum only if $\boldsymbol{S} = \boldsymbol{I}$, i.e., complete observations. In this case, the maximum likelihood (ML) estimate is defined as the support vector in $\mathcal{C}$ that maximizes $l(\boldsymbol{h})$, which we call the *constrained* ML estimate. In the special case where the noise covariance is a scalar multiple of the identity matrix then the constrained ML estimate is exactly the vector in $\mathcal{C}$ that is closest to the measurement $\tilde{z}$. If $\boldsymbol{S} \neq \boldsymbol{I}$ then the observations are incomplete—e.g., the data are sparse-angle or limited-angle. In this case, a unique solution can be found using the maximum *a posteriori* (MAP) criterion, which requires the specification of a prior probability on support vectors. Several such choices are explored in [33].

In the examples of Section VI we use the *joint ellipse* (JE) algorithm to estimate support vectors from support value measurements [33]. This algorithm assumes that the true support vector is near that of an ellipse, but the parameters of the ellipse—i.e., position $\boldsymbol{v}$, orientation $\phi$, eccentricity $\varepsilon$, and size $t$—are unknown. The JE algorithm jointly estimates the ellipse parameters and the support vector by solving

$$\underset{\boldsymbol{h},\boldsymbol{v},t,\varepsilon,\phi}{\text{minimize}}\ \alpha\|\boldsymbol{S}\boldsymbol{h} - \tilde{\boldsymbol{z}}\|^2 + (1-\alpha)\|\boldsymbol{h} - \boldsymbol{e}(\boldsymbol{v},t,\varepsilon,\phi)\|^2$$

subject to $t \geq 0$, $0 \leq \varepsilon \leq 1$, and $\boldsymbol{h} \in \mathcal{C}$. Here $\boldsymbol{e}(\boldsymbol{v},t,\varepsilon,\phi)$ is the support vector of the ellipse with the given parameters and $\alpha$ is a constant in the interval $[0,1]$. Thus the optimal $\boldsymbol{h}$ represents a vector that is close to the available observations, but is also near the support vector of some ellipse. A tradeoff between how well each condition is matched is provided by $\alpha$, which we set to 0.5 in all of our experiments below.

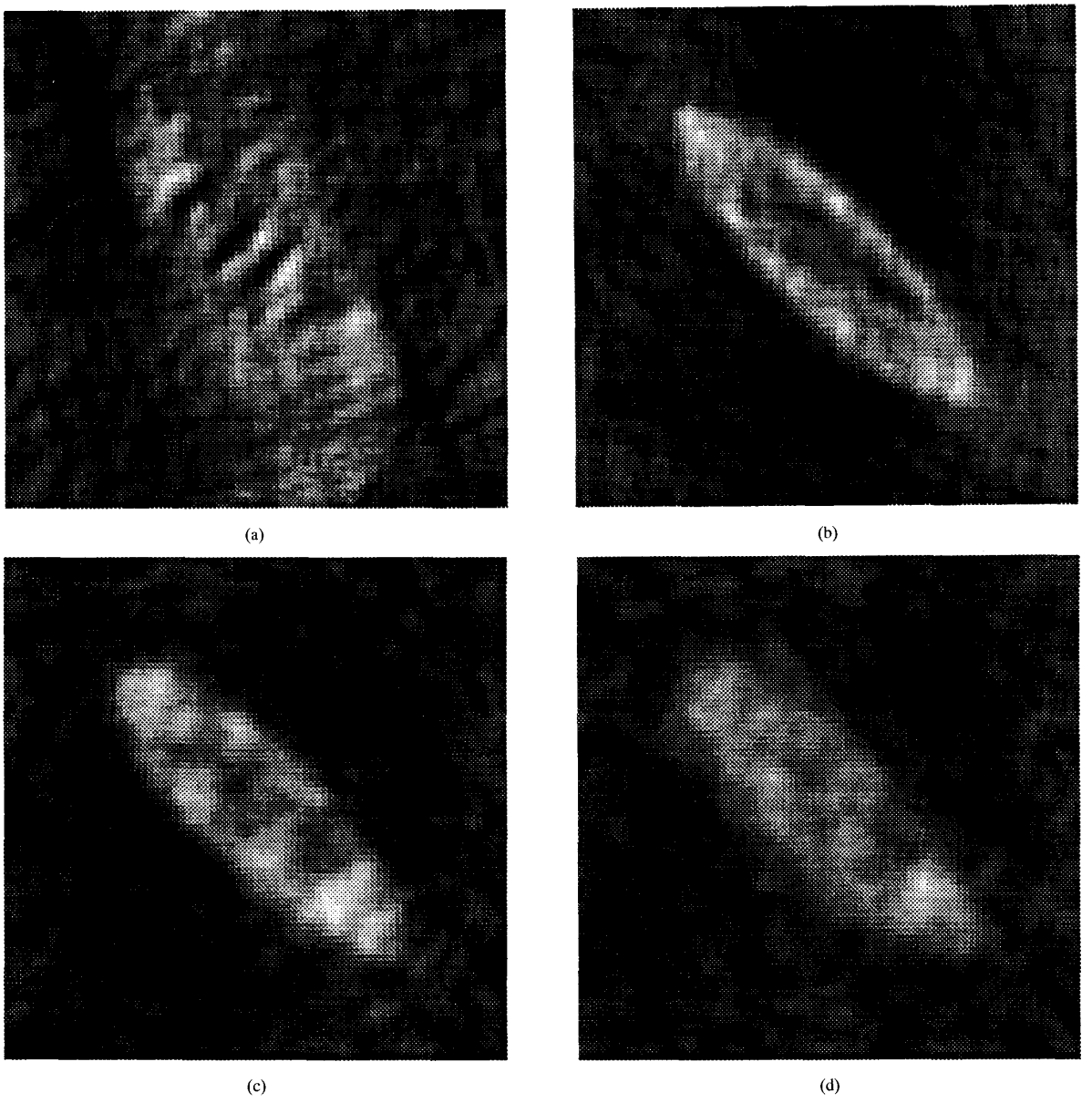



Fig. 8. Objects reconstructed from the 10.0 dB MIT-ellipse sinogram using ART followed by a 3 × 3 averaging filter using only (a) the left-most 40 projections, (b) the right-most 40 projections, (c) 15 sparse projections, and (d) 10 sparse projections.

## VI. Experimental Results

The MIT-ellipse shown in Fig. 2(a) was chosen for experimentation because the loss of data over different angular regions affects the reconstructions in different ways. For example, the absence of line integrals parallel to the long axis of the ellipse causes a lack of information related to the narrow dimension of the ellipse, but retains information about the letters inside the ellipse. In contrast, the absence of line integrals parallel to the short axis of the ellipse obscures the letters, but reveals the narrowness of the ellipse. To synthesize noisy observations we add independent samples of zero-mean Gaussian noise with variance $\sigma^2$ to each element of the true sinogram. The resulting sinogram has SNR defined as

$$\mathrm{SNR} = 10 \log \frac{\frac{\pi}{n_v} \frac{2T}{n_d} \sum_{j=1}^{n_v} \sum_{i=1}^{n_d} g^2(t_i, \theta_j)}{\sigma^2} \tag{32}$$

where $g(t_i, \theta_j)$ is the true sinogram. For example, 10 dB and 3 dB sinograms of the MIT-ellipse are shown in Figs. 6(a) and 6(b), respectively.

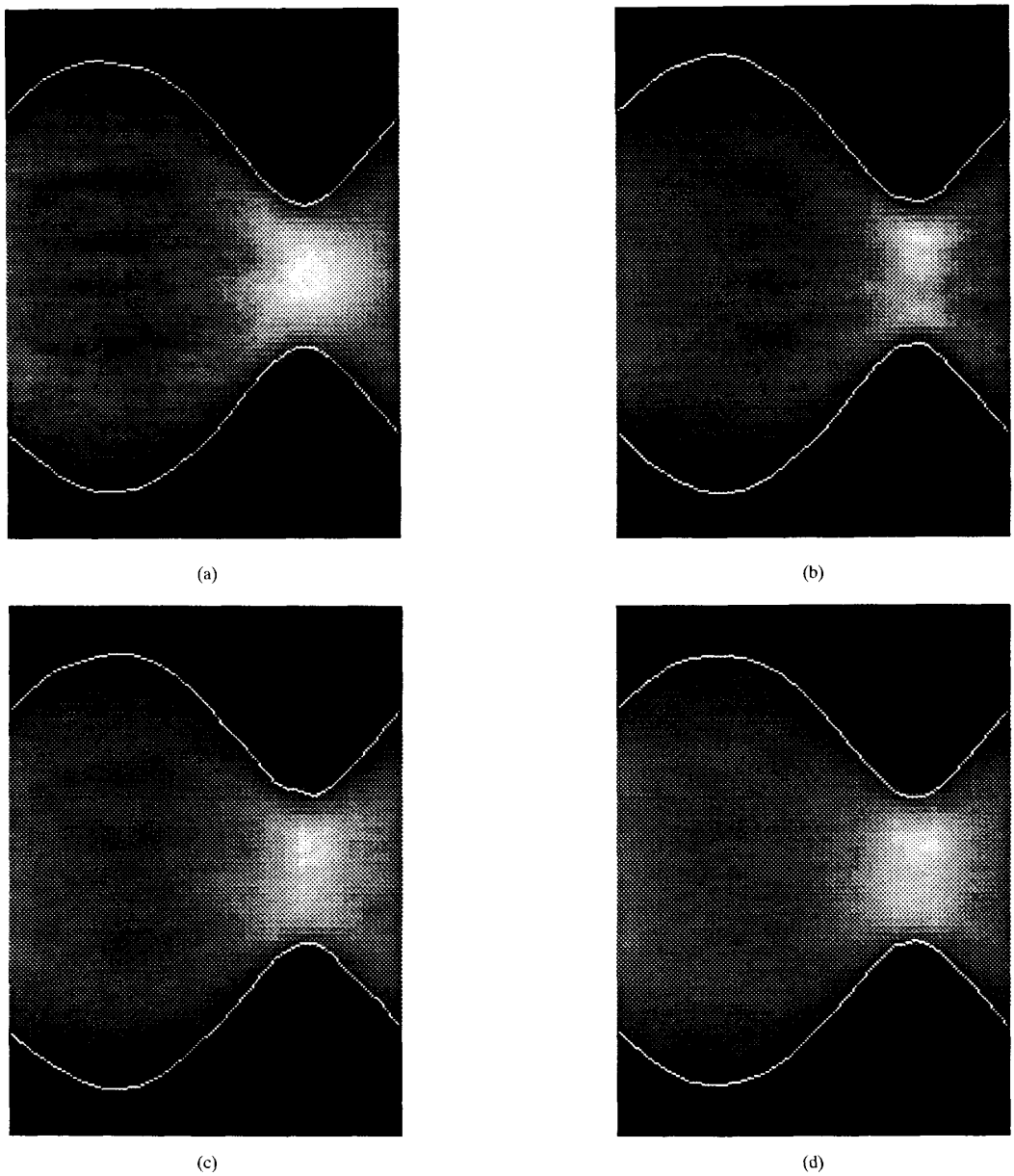
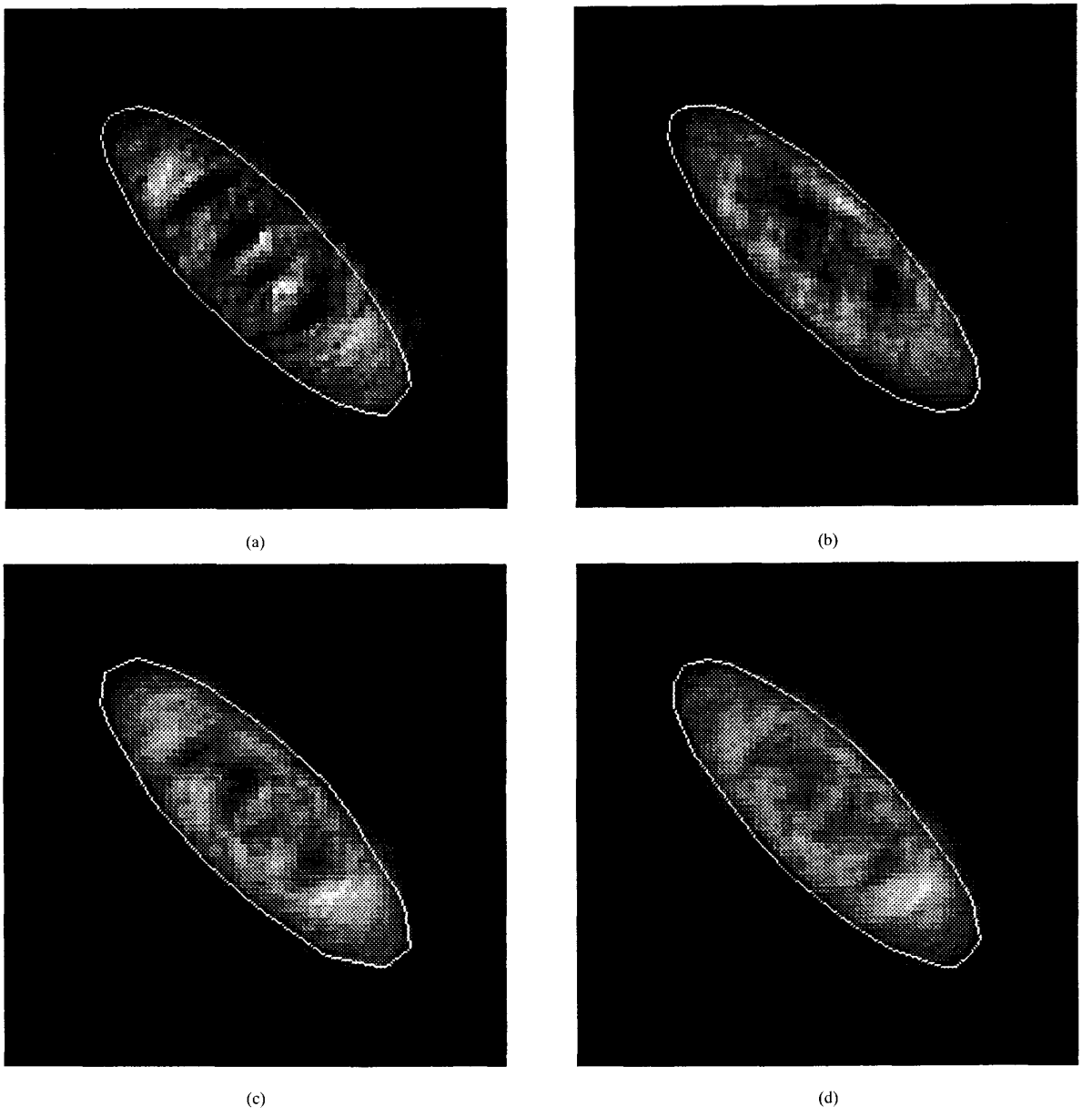



Fig. 9. Sinograms restored using the hierarchical algorithm applied to (a) the left-most 40 projections, (b) the right-most 40 projections, (c) 15 sparse projections, and (d) 10 sparse projections.

Two limited-angle and two sparse-angle cases are considered, both using selected projections from the 10 dB noisy MIT sinogram shown in Fig. 6(a). One limited-angle case observes the left-most 40 (out of 60) projections and the other observes the right-most 40 projections. These cases are generally considered to be severe limited-angle problems since even missing as little as 1% of the data can produce severe artifacts [15]. One sparse-angle case observes every 4th projection starting with the left column for a total of 15 projections. The second sparse-angle case observes every 6th projection for a total of 10 projections. All projections have 81 samples and all reconstructed images are $81 \times 81$ pixels.

Applying convolution backprojection (CBP) to the available noisy projections yields unacceptable reconstructions: they are very noisy and full of streak artifacts. ART, the algebraic reconstruction technique, yields only slightly better results, but can be improved by postfiltering each image with a $3 \times 3$ pixel averaging filter, as shown in Fig. 8. Similar results can be obtained using quadratic optimization and Bayesian methods [5]. The limited-angle reconstruction shown



(a) (b)

(c) (d)

Fig. 10. Objects reconstructed from restored sinograms in corresponding panels of Figure 9.

in Fig. 8(a) shows the most readable letters but loses the narrow dimension of the ellipse. In contrast, the limited-angle case of Fig. 8(b) captures the overall shape of the ellipse quite well, but the letters in the interior are illegible. The two sparse-angle cases show degradation of both the overall shape and legibility of the interior letters.

Figs. 9 and 10 show the results of applying the hierarchical algorithm to the same data. Fig. 9 shows restored sinograms with the estimated segmentation superposed. It is apparent from these images that the missing projection data are now filled in. Furthermore, all the sinogram data has been modified during restoration in order to remove the effects of noise, impose consistency, and reduce the amplitude outside the estimated region of support. Fig. 10 shows the objects reconstructed (using CBP) from the restored sinograms of Fig. 9 with outlines of the estimated convex support sets superposed. These images show dramatic improvement in contrast between the object and its background over those in Fig. 8. Also, the estimates of convex support are nearly perfect, even in the two severely limited-angle cases, and the letters in the interior are generally more legible. The one exception is Fig. 10(a), where the letters are less well defined

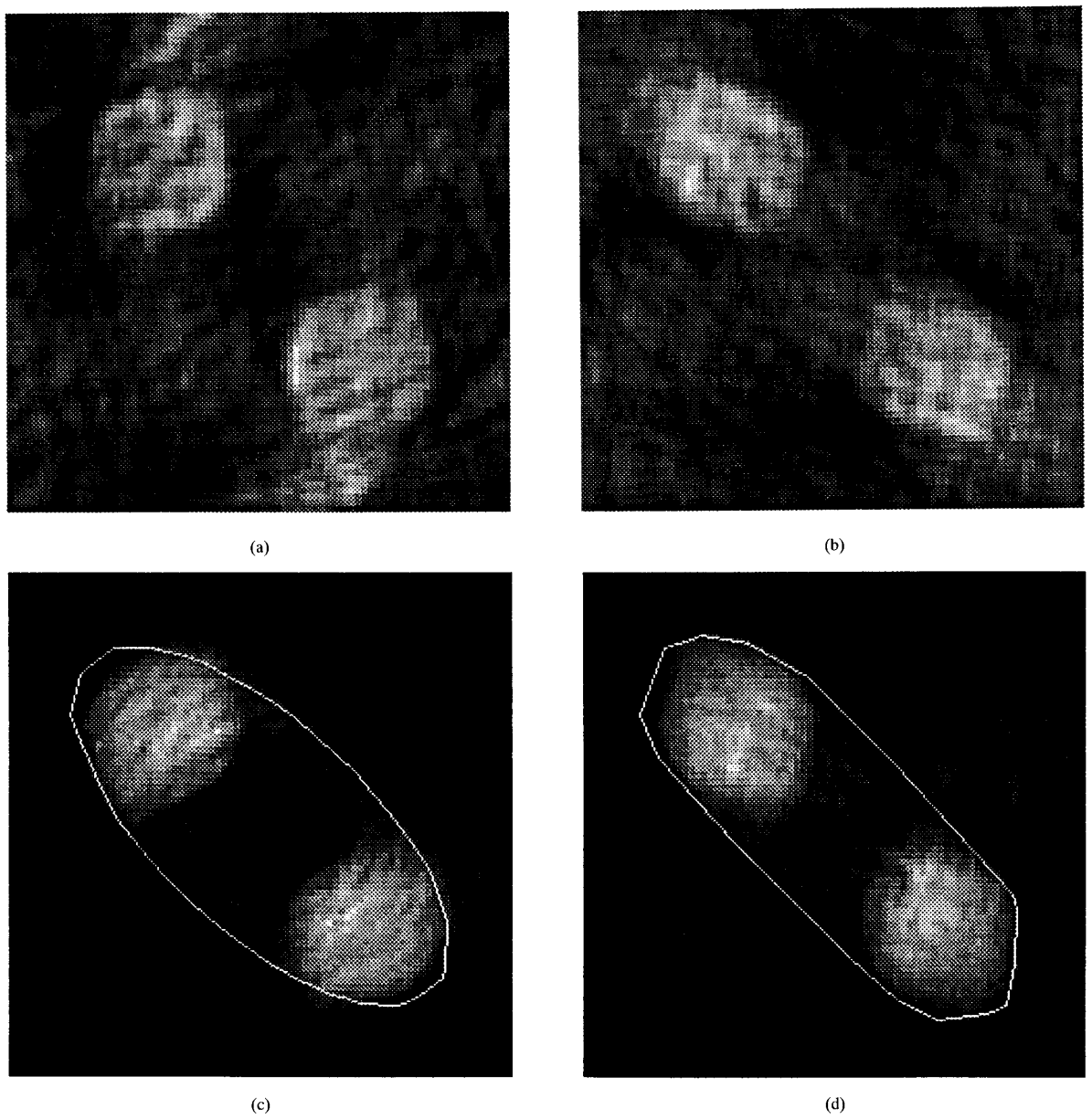


Fig. 11. Reconstructions from limited-angle observations of a 10.0 dB two-disk object: (a) ART from the left 40 projections; (b) ART from the right 40 projections; (c) hierarchical algorithm from the left 40 projections; and (d) hierarchical algorithm from the right 40 projections.

than those in Fig. 8(a), although they do have higher contrast. It should be noted that the application of a post-filter to the original ART reconstructions removes noise at the expense of greater blurring, which reduces legibility. Different filtering or alternate methods can improve legibility, but invariably allow greater noise and reduce the contrast between object and background.

Because the JE support vector estimation algorithm was used in the preceding experiments, all the reconstructed objects have nearly elliptical shape. It could therefore be argued that the success of the hierarchical algorithm is strongly dependent on the accuracy of this knowledge. To get a sense of the true degree of this effect we considered two limited-angle, noisy (SNR = 10.0 dB) observations of the two-disk sinogram shown in Fig. 2(d). Figs. 11(a) and 11(b) show ART reconstructions, again followed by a $3 \times 3$ pixel averaging filter, given observations of the left-most and right-most 40 noisy projections, respectively. Artifacts similar to the MIT-ellipse reconstructions of Figs. 8(a) and 8(b) are apparent, including streaks and low contrast.

The hierarchical algorithm produces the reconstructions and convex support estimates shown in Figs. 11(c) and 11(d). As in the previous cases, these results show dramatic improvement in contrast between the object and background, and the streak artifacts and background noise are largely removed. The two convex support estimates are quite different, however. In Fig. 11(c) the available projections do not provide views of the narrowest dimension; therefore, the JE algorithm makes the object shape more like that of an ellipse while approximately maintaining the support values measured from the observed projections. In Fig. 11(d), the observed projections contain information on both the narrow and wide dimensions, allowing a more accurate convex support reconstruction. Here, the convex support does not look elliptical because the available support value measurements do not agree with that of an ellipse. This is an important demonstration of the fact that prior shape knowledge is used primarily to remove nonuniqueness in the convex support estimation.

## VII. Discussion

We have demonstrated a method to estimate and hierarchically incorporate geometric information in a reconstruction algorithm designed for noisy and limited-angle or sparse-angle tomography. The method is based on estimation principles, incorporating prior probabilistic information and consistency conditions to overcome problems resulting from insufficient data.

Many variations of the basic algorithm may be considered, and several are discussed in [43]. One variation is to incorporate more specific information about the shape of the objects' convex hull. For example, with the additional knowledge that the object is an ellipse, but without knowing the size, orientation, or position or the ellipse, the algorithm produces nearly perfect support estimation in the experimental geometries and SNR used in Section VI. Another variation is to incorporate more than just the two consistency conditions given by the mass and center of mass constraints. One approach, which eliminates the requirement to specifically estimate the mass and center of mass and yet produce consistent sinograms, is presented in [35].

Another potential area of future research concerns the coefficient $\kappa$, which is used in the sinogram restoration algorithm to indicate the confidence in the given sinogram segmentation. A larger value indicates a higher degree of confidence, so a very large value of $\kappa$ could be used if the true segmentation were known and a smaller value for estimated segmentations. It may be reasonable to let $\kappa$ vary spatially to account for our varying degrees of confidence in the sinogram segmentation. In particular where there was significant interpolation, we would expect to make $\kappa$ smaller. Also, one might want $\kappa$ to be small near the estimated support value and increase with increasing $t$. The rate of increase might be related to the performance measure of the support value estimation algorithm. A heuristic approach has been reported in [44].

Finally, another area of future research would consider the possibility of spatially varying coefficients $\beta$ and $\gamma$. These coefficients specify the expected spatial smoothness of sinograms, which one might expect to be related to the spatial smoothness and shape of objects. Some results along these lines have been reported [45].

## References


[1] D. Ludwig, "The Radon transform on Euclidean space," *Comm. Pure Appl. Math.*, vol. 19, pp. 49–81, 1966.

[2] S. Helgason, *The Radon Transform.* Boston, MA: Birkhauser, 1980.

[3] G. S. Harell, D. F. Guthaner, R. S. Breiman, C. C. Morehouse, E. J. Seppi, W. H. Marshall, and L. Wexler, "Stop-action cardiac computed tomography," *Radiology*, vol. 123, pp. 515–517, 1977.

[4] C. K. Zoltani, K. J. White, and R. P. Kruger, "Results of feasibility study on computer assisted tomography for ballistic applications," U.S. Army Ballistic Research Laboratory, ARBRL-TR-02513, 1983.

[5] G. T. Herman, *Image Reconstruction from Projections.* New York: Academic, 1980.

[6] D. R. Wehner, *High Resolution Radar.* Norwood, MA: Artech House, Inc., 1987.

[7] M. Ein-Gal, *The Shadow Transformation: An Approach to Cross-Sectional Imaging,* Ph.D. dissertation, Stanford University, Dept. of Electr. Engr., 1974.

[8] A. K. Louis, "Picture reconstruction from projections in restricted range," *Math. Meth. Appl. Sci.*, vol. 2, pp. 209–220, 1980.

[9] M. E. Davison and F. A. Grunbaum, "Tomographic reconstructions with arbitrary directions," *Comm. Pure Appl. Math,* vol. 34, pp. 77–119, 1979.

[10] T. Inoye, "Image reconstruction with limited angle projection data," *IEEE Trans. Nucl. Sci.*, NS-26(2):2666–2669, 1979.

[11] J. A. Reeds and L. A. Shepp, "Limited angle reconstruction in tomography via squashing," *IEEE Trans. Med. Imaging,* vol. MI-6, pp. 89–97, June 1987.

[12] S. L. Wood, A. Macovski, and M. Morf, "Reconstruction with limited data using estimation theory," In *Computer Aided Tomography and Ultrasonics in Medicine,* Raviv, *et al.*, Ed. New York: North-Holland Publishing Co., pp. 219–233, 1979.

[13] M. H. Buonocore, *Fast Minimum Variance Estimators for Limited Angle Computed Tomography Image Reconstruction.* Ph.D. dissertation, Stanford University, 1981.

[14] M. I. Sezan and H. Stark, "Tomographic image reconstruction from incomplete view data by convex projections and direct Fourier inversion," *IEEE Trans. Med. Imag.*, vol. MI-3, pp. 91–98, 1984.

[15] B. P. Medoff, "Image reconstruction from limited data: theory and applications in computerized tomography," In *Image Recovery: Theory and Application,* H. Stark, Ed., chap. 9, Orlando: Academic, 1987, pp. 321–368.

[16] R. W. Gerchberg, "Super-resolution through error energy reduction," *Opt. Acta.*, vol. 21, no. 9, pp. 709–720, 1974.

[17] A. Papoulis, "A new algorithm in spectral analysis and band-limited extrapolation," *IEEE Trans. Circuits. Syst.,* CAS-22, pp. 735–742, 1975.

[18] J. L. C. Sanz and T. S. Huang, "Unified Hilbert space approach to iterative least squares linear signal restoration," *J. Opt. Soc. Am.*, vol. 73, no. 11, pp. 1455–1465, 1983.

[19] K. M. Hanson and G. W. Wecksung, "Bayesian approach to limited-angle reconstruction in computed tomography," *Appl. Optics,* vol. 24, pp. 4028–4039, Dec. 1980.

[20] S. Geman and D. E. McClure, "Bayesian image analysis: and application to single photon emission tomography," Brown Univ., 1985, Tech. Rep., Proc. Amer. Stat. Assoc. Statistical Computing.

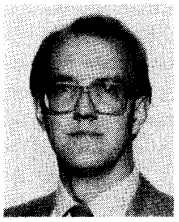
[21] D. C. Youla, "Mathematical theory of image restoration by the method of convex projections," In *Image Recovery: Theory and Application,* H. Stark, Ed., Orlando: Academic, 1987, chap. 2, pp. 29–78.

[22] D. J. Rossi and A. S. Willsky, "Reconstruction from projections based on detection and estimation of objects–parts I and II: Performance analysis and robustness analysis," *IEEE Trans. Acoust. Speech, Signal Processing,* ASSP-32, pp. 886–906, 1984.

[23] M. Soumekh, "Binary image reconstruction from four projections," In *Proc. 1988 Int. Conf. Acoust., Speech., Sig., Proc.,* Apr. 1988, vol. 2, pp. 1280–1283.

[24] S. K. Chang and G. L. Shelton, "Two algorithms for multiple-view binary pattern reconstruction," *IEEE Trans. Sys. Man. and Cyber.*, vol. SMC-1, pp. 90–94, Jan. 1971.

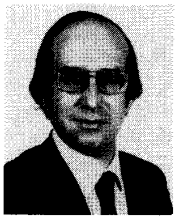
[25] P.C. Fishburn, J.C. Lagarias, J.A. Reeds, and L.A. Shepp, "Sets uniquely determined by projections on axes I. continuous case." *SIAM J. Applied Math.,* vol. 50, no. 1, pp. 288–306, Feb. 1990.
[26] H.J. Trussell, H. Orun–Ozturk, and M.R. Civanlar, "Errors in reprojection methods in computerized tomography," *IEEE Trans. Med. Imag.,* vol. MI-6, pp. 220–227, Sept. 1987.
[27] J.H. Park, K.Y. Kwak, and S.B. Park, "Iteractive reconstruction-reproduction in projection space," *Proc. IEEE,* vol. 73, pp. 1140–1141, June 1985.
[28] J.H. Kim, K.Y. Kwak, S.B. Park, and Z.H. Cho, "Projection space iteration reconstruction-reprojection," *IEEE Trans. Med. Imag.,* vol. MI-4, pp. 139–143, Sept. 1985.
[29] M.I. Sezan and H. Stark, "Image restoration by convex projections in the presence of noise," *App. Opt.,* vol. 22, no. 18, pp. 2781–2789, 1983.
[30] S.R. Deans, *The Radon Transform and Some of Its Applications.* New York: John Wiley and Sons, 1983.
[31] J.L. Prince and A.S. Willsky, "A geometric projection-space reconstruction algorithm," in *Linear Algebra and its Applications,* vol. 130, pp. 151–191, 1990; Special Issue on Linear Algebra in computed Tomography, G.T. Herman, Ed.
[32] J.L. Prince and A.S. Willsky, "Reconstructing convex sets from support line measurements," *IEEE Trans. Patt. Anal. Mach. Intell.,* vol. 12, pp. 377–389, Apr. 1990.
[33] J.L. Prince and A.S. Willsky, "Convex set reconstruction using prior geometric information," *CVGIP: Graphical Models and Image Processing,* vol. 53, no. 5, pp. 413–427, 1991.
[34] A.C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging,* New York: IEEE Press, 1988.
[35] J.L. Prince and A.S. Willsky, "Constrained sinogram restoration for limited-angle tomography," *Opt. Eng.,* pp. 535–544, May 1990.
[36] A.S. Willsky and H.L. Jones, "A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems," *IEEE Trans. Auto. Contr.,* pp. 108–112, Feb. 1976.
[37] A.M. Mier-Muth and A.S. Willsky, "A sequential method for spline approximation with variable knots," M.I.T. Electronic Systems Laboratory, Tech. Rep. ESL-P-759, 1977.
[38] A. Blake, "Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction," *IEEE Trans. Patt. and Anal. Mach. Intell.,* vol. 11, pp. 2–12, Jan. 1989.
[39] S.F. Yau and Y. Bresler, "Image restoration by complexity regularization via dynamic programming," In *IEEE Proc. ICASSP,* vol. 3, pp. 305–308, 1992.
[40] A.S. Willsky and H.L. Jones, "A generalized likelihood ratio approach to state estimation in linear systems subject to abrupt changes," M.I.T. Laboratory for Information and Decision Syst., Tech. Rep. LIDS-P-538, 1974.
[41] G. Wahba, *Spline Models for Observational Data* in SIAM, Philadelphia, PA, 1990.
[42] L. Younes, "Parametric inference for imperfectly observed Gibbsian fields," *Probability Theory and Related Fields,* vol. 82, pp. 625–645, 1989.
[43] J.L. Prince, *Geometric Model-Based Estimation From Projections.* PhD dissertation, M.I.T., Cambridge, MA, 1988.
[44] J.L. Prince, "Consistency and convexity in object reconstruction from projections," In *Proc. IEEE Int. Conf. Syst. Eng.,* Pittsburgh, PA, Aug. 1990, pp. 543–546. IEEE Catalog Number: 90CH2872-9
[45] J.L. Prince, "An iterative approach to sinogram restoration," In *Proc. 1990 IEEE Int. Conf. Eng. in Medicine and Biology,* 1990.



**Jerry L. Prince** (S'78–M'83) received the B.S. degree from the University of Connecticut in 1979 and the S.M., E.E., and Ph.D. degrees in 1982, 1986, and 1988, respectively, from M.I.T., all in electrical engineering.

From 1982 to 1983, he was employed at the Brigham and Women's Hospital in Boston, MA, where he developed instrumentation and reconstruction algorithms for ultrasonic imaging in medicine. From 1983 to 1988, he held both teaching and research assistantships at M.I.T., where he conducted research on geometric reconstruction methods in computed tomography. In 1988, he joined the technical staff at The Analytic Sciences Corp. (TASC) in Reading, MA, where he contributed to the design of an automated vision system for synthetic aperture radar imaging. He joined the faculty at the Johns Hopkins University in 1989, where he is currently an Assistant Professor in the Department of Electrical and Computer Engineering and holds a joint appointment in the Department of Radiology. His current research interests are in image processing and computer vision with primary application to medical imaging.

Dr. Prince is a member of IEEE and Sigma Xi professional societies and Tau Beta Pi, Eta Kappa Nu, and Phi Kappa Phi honor societies. He is currently an Associate Editor of *IEEE Transactions on Image Processing.*



**Alan S. Willsky** (S'70–M'73–SM'82–F'86) received both the S.B. degree and the Ph.D. degree from M.I.T. in 1969 and 1973, respectively.

He joined the M.I.T. faculty in 1973 and his present position is Professor of Electrical Engineering. From 1974 to 1981, Dr. Willsky served as Assistant Director of the M.I.T. Laboratory for Information and Decision Systems. He has been an associate editor of several journals, including the *IEEE Transactions on Automatic Control,* has served as a member of the Board of Governors and Vice President for Technical Affairs of the IEEE Control Systems Society, was program chairman for the 1981 Bilateral Seminar on Control Systems held in the People's Republic of China, and was special guest editor of the 1992 special issue of the *IEEE Transactions on Information Theory* on wavelet transforms and multiresolution signal analysis. Dr. Willsky is the author of the research monograph *Digital Signal Processing and Control Estimation Theory* and is co-author of the undergraduate text *Signals and Systems.* Dr. Willsky's present research interests are in problems involving multidimensional and multiresolution estimation and imaging, discrete-event systems, and the asymptotic analysis of control and estimation systems.

He is a founder and member of the board of directors of Alphatech, Inc. In 1975, he received the Donald P. Eckman Award from the American Automatic Control Council.

In 1988 he was made a Distinguished Member of the IEEE Control Systems Society. He was awarded the 1979 Alfred Noble Prize by the ASCE and the 1980 Browder J. Thompson Memorial Prize Award by the IEEE for a paper excerpted from his monograph.