A Moment-Based Variational Approach to Tomographic Reconstruction

Peyman Milanfar, *Member, IEEE*, William C. Karl, *Member, IEEE*, and Alan S. Willsky, *Fellow, IEEE*

***Abstract*— In this paper, we describe a variational framework for the tomographic reconstruction of an image from the maximum likelihood (ML) estimates of its orthogonal moments. We show how these estimated moments and their (correlated) error statistics can be computed directly, and in a linear fashion from given noisy and possibly sparse projection data. Moreover, thanks to the consistency properties of the Radon transform, this two-step approach (moment estimation followed by image reconstruction) can be viewed as a statistically optimal procedure.**

Furthermore, by focusing on the important role played by the moments of projection data, we immediately see the close connection between tomographic reconstruction of nonnegative-valued images and the problem of nonparametric estimation of probability densities given estimates of their moments. Taking advantage of this connection, our proposed variational algorithm is based on the minimization of a cost functional composed of a term measuring the *divergence* between a given prior estimate of the image and the current estimate of the image and a second quadratic term based on the error incurred in the estimation of the moments of the underlying image from the noisy projection data. We show that an iterative refinement of this algorithm leads to a practical algorithm for the solution of the highly complex *equality constrained* divergence minimization problem. We show that this iterative refinement results in superior reconstructions of images from very noisy data as compared with the classical filtered back-projection (FBP) algorithm.

I. Introduction

In this paper, we discuss the tomographic reconstruction of a function $f(x, y)$ from noisy measured values of its projections via the maximum likelihood (ML) estimation of the orthogonal moments of f. In particular, the fundamental result on which the algorithms in this paper rely is that the statistically optimal estimate of an image based on noisy samples of its Radon transform can be obtained in two distinct steps: the first step being the ML (or MAP) estimation of the moments of the underlying image from the noisy data and a second step focusing on the reconstruction of the image from its estimated moments. In this way, we demonstrate and take advantage of the natural utility of moments in solving tomographic reconstruction problems.

The first step in this two-tier algorithm is a simple *linear* estimation problem (allowing us also to determine error statistics with relative ease), whereas the second is a highly ill-posed inverse problem. In particular, by adapting this approach, we have transformed the problem of inverting the Radon transform into one of reconstructing a function from estimates of its moments. While the problem of reconstructing a function from a finite number of estimated moments is known to be highly ill-posed [40], by making contact with the field of statistics, and in particular the problem of nonparametric probability density estimation from estimated moments, we can take advantage of the many concepts that have been devised to deal with this ill-posedness in other contexts. Specifically, by using this connection, we adapt ideas from nonparametric probability density estimation resulting in efficient algorithms for reconstructing an image using a divergence-based variational criterion. This criterion allows us to use prior knowledge (obtained, for example using standard tomographic methods) to regularize the problem and defaults to a maximum entropy solution if no prior information is available.

We show that there are several advantages to our two-step approach. One is that the use of moments provides an explicit mechanism for controlling the degrees of freedom in the reconstructions, which is an issue of considerable importance in problems with very noisy or sparse projection data. Such situations arise, for instance, in nondestructive evaluation and ocean acoustic tomography, where due to various physical constraints, the gathered data can often be sparse and very noisy. In such circumstances, the reconstruction process must be regularized in order to yield an acceptable result. As we demonstrate in this paper, controlling the degrees of freedom of the reconstruction is an effective, robust, and efficient way to accomplish this.

Another advantage is the computational savings inherent to our approach, as compared to standard variational algorithms that involve the numerical solution of complex partial differential equations. A third is that by using these formulations, we can introduce prior information, in terms of prior estimates of reconstructions, or geometric information, in a very simple way with only minimal increase in computation. Finally, these features yield an overall efficient and versatile set of algorithms that yield reconstructions of excellent quality when compared to filtered back-projection (FPB) operating on data limited in quality and quantity.

Manuscript received December 19, 1993; revised March 28, 1995. This work was supported by the Advanced Research Projects Agency under Air Force Grant F49620-93-1-0604, the Office of Naval Research under Grant N00014-91-J-1004, the US Army Research Office under Contract DAAL03-92-G-0115, and by the Clement Vaturi Fellowship in Biomedical Imaging Sciences at MIT. The associate editor coordinating the review of this paper and approving it for publication was Prof. Ken D. Sauer.

P. Milanfar is with SRI International, Menlo Park, CA 94025 USA (e-mail: milanfar@unix.sri.com).

W. C. Karl is with the Department of Electrical, Computer, and Systems Engineering, Boston University, Boston, MA 02215 USA.

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 USA.

Publisher Item Identifier S 1057-7149(96)01799-X.

1057–7149/96$05.00 © 1996 IEEE

The reconstruction of images from their moments has not been a central topic in image processing theory since the use of moments in this setting has primarily focused on their *extraction* from images (for use as distinguishing features) rather than on their use in reconstruction [27]. Furthermore, there has been relatively little work in this area within the tomography community [34], [29], [19], [28], [10], [3]. On the other hand, the moment problem has been the subject of much work in the mathematics and statistics communities for many years [1], [35], [6]–[8]. However, while variants of variational/regularization methods developed here have been studied elsewhere in the literature [41], [40], [16], [26], [2], [8], [7], [14], [33], the precise combination of techniques we propose here has not been developed or investigated in the particular context of tomographic reconstruction from moments. We also propose novel and efficient numerical techniques for solving this variational problem and study some of their properties and extensions.

In Section II, we present an optimal (ML) algorithm for the estimation of the moments of an image from noisy measurements of its projections. In Section III, we describe how the underlying image may be reconstructed from these estimated moments via regularization. Section IV contains the explicit solution to this variational problem, and here, we also discuss those properties of this solution that make it attractive. In Section V, we discuss an iterative refinement of the divergence-based regularization approach and demonstrate how this refinement leads to efficient solution of a highly complex equality-constrained divergence minimization problem. Section VI contains our numerical simulation results including illustration of how prior information—in this case, that provided by the standard FBP solution—can be incorporated into our approach. Finally, in Section VII, we state our conclusions.

II. Estimating Moments from Projections

Let $f(x,y) \in L^2(D)$ denote a square-integrable function with support inside the unit disk D in the plane and further denote by $g(t,\theta) = \Re f$ the Radon transform of f defined as follows:

$$g(t,\theta) = \int\int_D f(x,y)\delta(t - \omega \cdot [x,y]^T)\, dx\, dy \tag{1}$$

where $\omega = [\cos(\theta),\ \sin(\theta)]$ and $\delta(\cdot)$ denotes the Dirac delta function; see Fig. 1.

The function $g(t,\theta) \in L^2([-1,1] \times [0,2\pi])$ [12] is defined for each pair (t,θ) as the integral of f over a line at angle $\theta + \frac{\pi}{2}$ with the x-axis and at radial distance t away from the origin. An elementary result [12], which follows from the definition of the Radon transform, states that if $F(t)$ is *any* square integrable function on $[-1,1]$, then the following relation holds true:

$$\int_{-1}^{1} g(t,\theta)F(t)dt = \int\int_D f(x,y)F(\omega \cdot [x,y]^T)dxdy. \tag{2}$$

By considering $F(t) = e^{-iwt}$, the celebrated *Projection Slice Theorem* [13] is obtained. What we wish to consider is the case where $F(t)$ is taken to range over a set of orthonormal basis functions over $[-1,1]$. In particular, we will consider

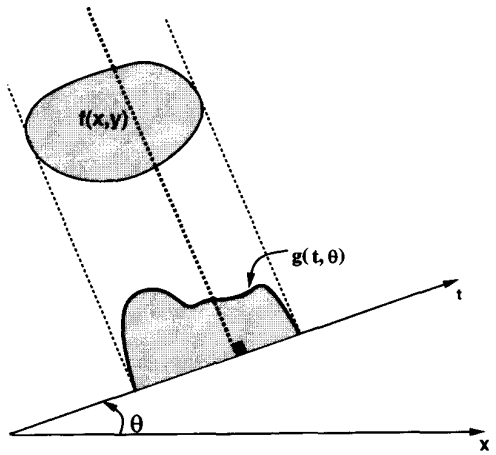

Fig. 1. Radon transform.

the case when $F(t) = P_k(t)$, where $P_k(t)$ is the kth-order normalized Legendre polynomial over $[-1,1]$ defined by

$$P_k(x) = \sum_{i=0}^{k} \beta_{ik}x^i = \sqrt{\frac{2k+1}{2}}\frac{1}{2^k k!}\frac{d^k}{dx^k}(x^2-1)^k. \tag{3}$$

In this basis, (2) relates the moments of the function f linearly to those of its Radon transform g, as we describe next.

Let $G^{(k)}(\theta)$ denote the kth-order Legendre moment of $g(t,\theta)$ for each fixed θ. That is

$$G^{(k)}(\theta) = \int_{-1}^{1} g(t,\theta)P_k(t)dt. \tag{4}$$

In addition, denote by λ_{pq} the orthogonal moments of f defined as

$$\lambda_{pq} = \int\int_D f(x,y)P_p(x)P_q(y)dxdy. \tag{5}$$

By appealing to (2), it is easily shown that the kth orthogonal moment $G^{(k)}(\theta)$ of $g(t,\theta)$ is a linear combination of the orthogonal moments λ_{pq} of $f(x,y)$ of order[1] $p+q \le k$, which is a direct consequence of the consistency conditions for Radon transforms discussed in [12] and [29]. Defining $\mathcal{G}_N(\theta) = [G^{(0)}(\theta),\cdots,G^{(N)}(\theta)]^T$, $\lambda^{(k)} = [\lambda_{k,0},\lambda_{k-1,1},\cdots,\lambda_{0,k}]^T$ and $\mathcal{L}_N = [\lambda^{(0)^T},\cdots,\lambda^{(N)^T}]^T$, we can write

$$\mathcal{G}_N(\theta) = A_N(\theta)\mathcal{L}_N \tag{6}$$

where $A_N(\theta)$ is lower block triangular. When considering the complete (infinite) set of moments of f and g, we can write

$$\mathcal{G}(\theta) = A(\theta)\mathcal{L} \tag{7}$$

where $\mathcal{G}(\theta)$ and $\mathcal{L}$ contain *all* the moments of g and f, respectively, and $A(\theta)$ is a lower triangular linear operator. Note that since the infinite set of moments $\mathcal{L}$ and $\mathcal{G}(\theta)$ provide complete orthogonal decompositions of $f(x,y)$ and of $g(t,\theta)$, (7) provides us with a factorization of the Radon transform. Specifically, let A denote the operator taking $\mathcal{L}$ to the family

[1] In fact, for k even, $G^{(k)}(\theta)$ is a linear combination of λ_{pq} for $p+q = k,\ k-2,\ \cdots,\ 2,\ 0$, whereas for k odd, it is a linear combination of λ_{pq} for $p+q = k,\ k-2,\ \cdots,\ 3,\ 1$.

of functions $G^{(k)}(\theta)$ of θ according to (7), and define the moment operators $\Omega f = \mathcal{L}$ and $Mg = \mathcal{G}$ (where M maps the function $g(t,\theta)$ to the family of functions $G^{(k)}(\theta)$). Then, since $g = \Re f$ and since M and Ω are unitary, we see that

$$\Re = M^* A \Omega. \tag{8}$$

In [22], we have used this decomposition of the Radon transform to derive new interpretations of classical reconstruction algorithms such as FBP.

Suppose now that we are given noisy measurements of g at m distinct angles $\theta_1, \theta_2, \cdots, \theta_m$ in $[0,\pi)$ as

$$y(t,\theta_j) = g(t,\theta_j) + e(t,\theta_j) \tag{9}$$

where $e(t,\theta_j)$ are independent white noise processes in t with intensity σ^2, and where we assume that for each θ_j, $y(t,\theta_j)$ is available for all[2] $-1 \le t \le 1$. If for each θ_j we represent our data in terms of its orthogonal moments, we have

$$Y^{(k)}(\theta_j) = G^{(k)}(\theta_j) + e^{(k)}(\theta_j), \quad k = 0, 1, \cdots \tag{10}$$

where $Y^{(k)}(\theta_j)$ and $e^{(k)}(\theta_j)$ denote the L^2 inner products of $y(t,\theta_j)$ and $e(t,\theta_j)$ with the kth-order Legendre polynomial $P_k(t)$. Due to the orthonormality of the family $\{P_k(t),\ k \ge 0\}$ and the assumption of white noise, the error terms[3] $e^{(k)}(\theta_j) \sim \mathcal{N}(0,\sigma^2)$ are independent across both k and j. Thus, if we let $Y(\theta_j)$ denote the set of all $Y^{(k)}(\theta_j)$ for $k = 0,\ 1, \cdots$, and use analogous notation for $e(\theta_j)$, we see that thanks to (7)

$$Y(\theta_j) = A(\theta_j)\mathcal{L} + e(\theta_j), \quad j = 1, 2, \cdots, m. \tag{11}$$

Since the full set of moments $\mathcal{L}$ provides a complete characterization of $f(x,y)$, we can see that a sufficient statistic for the estimation of $f(x,y)$ is the ML estimate of $\mathcal{L}$, given the data in (11). However, given the fact that we only have a finite number of viewing angles, it is not surprising that (11) does *not* provide an invertible relation between the data $Y(\theta_j)$ and the full set of moments $\mathcal{L}$. In fact, we have the following.

Proposition I: Given line integral projections of $f(x,y)$ at m different angles θ_j in $[0,\pi)$, one can uniquely determine the first m moment vectors $\lambda^{(j)}$, $0 \le j < m$ of $f(x,y)$. This can be done using only the first m orthogonal moments $G^{(k)}(\theta_j)$, $0 \le k < m$ of the projections. Furthermore, moments of $f(x,y)$ of higher order *cannot* be uniquely determined from m projections.

What this result, which is proved in Appendix A and in [23], says is the following. Let $Y_N(\theta_j)$ denote the vector of the Legendre moments of $y(t,\theta_j)$ of order $k = 0, 1, \cdots, N$ so that $Y_N(\theta_j) = A_N(\theta_j)\mathcal{L}_N + e_N(\theta_j)$ (where $e_N(\theta_j)$ is defined analogously). Collecting all of the $Y_N(\theta_j)$ into a large column vector

$$\mathbf{Y}_N = [Y_N(\theta_1)^T,\ Y_N(\theta_2)^T,\ \cdots,\ Y_N(\theta_m)^T]^T \tag{12}$$

we have

$$\mathbf{Y}_N = \mathbf{A}_N \mathcal{L}_N + \mathbf{e}_N \tag{13}$$

[2]Clearly, in practice, as in our numerical experiments, $y(t,\theta)$ will be sampled in t as well as in θ.

[3]$\mathcal{N}(0,\sigma^2)$ denotes a zero-mean Gaussian random variable with variance σ^2.

where $\mathbf{A}_N$ and $\mathbf{e}_N \sim \mathcal{N}(0,\sigma^2 I)$ are defined in a corresponding fashion. Then, from Proposition 1, we have that $\mathbf{A}_N$ has full column rank so that a unique ML estimate of $\mathcal{L}_N$ exists, if and only if $N \le m-1$, and this estimate is given by

$$\widehat{\mathcal{L}}_N = \left(\mathbf{A}_N^T \mathbf{A}_N\right)^{-1} \mathbf{A}_N^T \mathbf{Y}_N \tag{14}$$

with the corresponding error covariance matrix given by $Q_N = \sigma^2\left(\mathbf{A}_N^T \mathbf{A}_N\right)^{-1}$. Moreover, thanks to the lower triangular relationship inherited from (6), we also have that the ML estimate of $\mathcal{L}_N$ in (14), based on the Legendre moments of the data of order $\le N$, is *identical* to the ML estimate of $\mathcal{L}_N$ based on the complete data, i.e., on all the Legendre moments as in (12).

Note further that for $N \ge m$, $\mathbf{A}_N$ will *not* have full column rank, implying that only some linear combination of the λ_{pq} for $p+q > m$ have well-defined ML estimates. In principle, optimal processing requires that *all* of these ML estimates be calculated. However, in practice, only a finite number of moments can be calculated. Furthermore, as one might expect, the estimates of the higher order moments are increasingly uncertain for a fixed amount of data. In fact, useful information is only provided for moments of order considerably less than m. As an example, Fig. 2 displays plots of the trace of the covariance matrices of the estimated orthogonal moment vectors[4] $\widehat{\lambda}^{(k)}$ up to order $k = 10$ versus k and for different SNR values. For the curves in this plot, $m = 60$ equally spaced projections in $[0,\pi)$ were considered. Consequently, for practical purposes, there is no significant information loss in using (14) for a value of $N < m$ as a sufficient statistic in place of the ML estimate of *all* moments. Thus, in the remainder of this paper, we consider the problem of reconstructing $f(x,y)$ given noisy measurements $\widehat{\mathcal{L}}_N$ of $\mathcal{L}_N$ with error covariance Q_N. Finally, note that because of the lower triangular structure of $A_N(\theta_j)$, Q_N is *not* block diagonal, i.e., the estimated moments of $f(x,y)$ of different order have correlated errors. The algorithm described in the sequel takes this into account in a statistically optimal fashion.

III. The Inverse Problem and Its Regularization

In this section, we propose a variational approach for the reconstruction of an image from noisy estimates of (a finite number of) its moments that regularizes the moment problem and at the same time takes into account the explicit structure of the corrupting noise. Our approach is founded on the principle of Minimum I-Divergence (MID) [6], [38], [39]. The principle states that of all the functions that satisfy a given set of moment constraints, one should pick the one $\widehat{f}$ with the least *I-Divergence* $D(f,f_0)$, relative to a given prior estimate f_0 of f, where this is defined as

$$D(f,f_0) = \int\int \Big(f(x,y)\log\Big(\frac{f(x,y)}{f_0(x,y)}\Big) + f_0(x,y) - f(x,y)\Big)\,dx\,dy. \tag{15}$$

[4]Note that for a given k, the covariance matrix of $\widehat{\lambda}^{(k)}$ is simply the $(k+1)$th, $(k+1)\times(k+1)$ diagonal block of the covariance matrix Q_N of $\widehat{\mathcal{L}}_N$

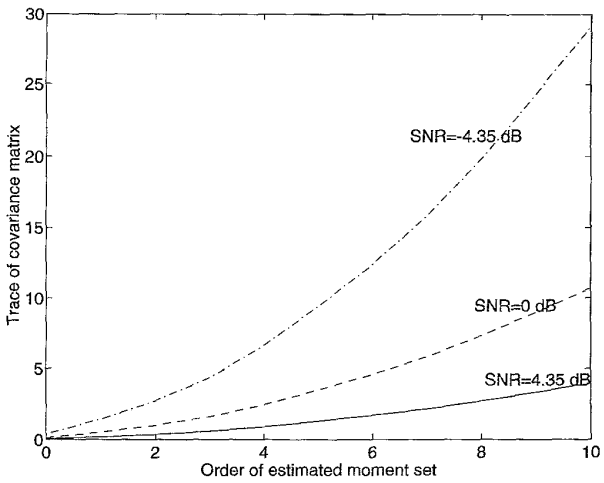

Fig. 2. Trace of covariance matrix versus moment order up to order 10.

The basic idea dates back to Kullback [17] and was later generalized by Csiszár [6] and includes the principle of *maximum entropy* [15] as a special case when f_0 is assumed to be a constant function. Entropy and, more recently, I-Divergence have a rich history of applications in pattern classification [37], spectral analysis [36], image processing [42], [11] and, recently, tomography [34], [31], [9], [25], [3], [4]. In most of these applications, the general problem has often been posed as the following type of equality constrained optimization problem:

$$\min_f D(f, f_0) \text{ subject to } \int\int f(x,y)\,\phi_{i,j}(x,y)\,dx\,dy = \widehat{s}_{i,j}. \tag{16}$$

In particular, in the context of tomography, the weight functions $\phi_{i,j}(x,y)$ have frequently been chosen as appropriate delta functions so that the constraints $\widehat{s}_{i,j}$ are the *noisy* measured values of the Radon transform $g(t_i, \theta_j)$ [34], [31], [9], [25]. That is to say, the constraints have the form

$$\int\int f(x,y)\,\delta(t_i - \omega_j \cdot [x,y]^T)\,dx\,dy = \widehat{s}_{i,j} \tag{17}$$

where ω_j is the unit direction vector making an angle θ_j with the x-axis. In fact, most of the tomography literature on the subject has been concerned with a very special case of *maximum entropy* reconstruction. Other variants of these algorithms allow for the equality constraints to be inequality constraints so that some notion of uncertainty in the measured values of $\widehat{s}_{i,j}$ can be taken into account [16].

Four important features distinguish our approach from other available algorithms mentioned above. The first concerns the incorporation of a prior estimate f_0. In particular, in most (but not all) other work using divergence-like criteria as in (16), the focus has been on maximum entropy methods corresponding to the trivial choice $f_0 = 1$. Not only do we allow for the possibility of an arbitrary (but positive) f_0, but we also demonstrate the use of particular methods for choosing f_0 that can enhance performance considerably by allowing for the incorporation of prior geometric and image information. The second is that we use the estimated Legendre moments instead of the actual measured values of the projections. This is to say that, in our case, the basis functions are $\phi_{i,j}(x,y) = P_i(x)P_j(y)$, where $P_i(\cdot)$ denotes the ith-order normalized Legendre polynomial over the interval $[-1, 1]$. Third, we do not use the estimated moments to form hard equality or inequality constraints but rather use these estimates, along with their computed covariance structure, to form a composite cost function that consists of the I-Divergence term plus a quadratic form in terms of the estimated moments. Finally, and perhaps most importantly, in addition to using the estimated moments, we also directly incorporate their estimated covariances, thus ensuring that these data are used in a statistically optimal way. That is, as we discussed in the preceding section, by using moments, we are able to *focus* the information in the raw projection data, via a simple linear processing step, identifying a much more compact set of statistically significant quantities capturing *most* information of use in reconstruction.[5]

Formally, we define the I-Divergence regularization (IDR) cost functional as

$$\begin{aligned} J_{IDR}(f, f_0) =& \gamma D(f, f_0) \\ &+ \frac{1}{2}(\mathcal{L}_N(f) - \widehat{\mathcal{L}}_N)^T \Sigma_N (\mathcal{L}_N(f) - \widehat{\mathcal{L}}_N) \end{aligned} \tag{18}$$

where $\gamma \in (0, \infty)$ is the regularization parameter, and $\Sigma_N = Q_N^{-1}$ is the inverse of the error covariance matrix for the estimate $\widehat{\mathcal{L}}_N$. To derive a probabilistic interpretation of the IDR cost functional, consider the MAP estimate of f based on noisy measurement of its moments up to order N. Assuming that $P(f)$ is some prior probability density function on the space of functions f, the MAP cost to be minimized is given by

$$\begin{aligned} J_{map}(f) =& -\log P(\widehat{\mathcal{L}}_N | f) - \log P(f) \\ =& \frac{1}{2}(\mathcal{L}_N(f) - \widehat{\mathcal{L}}_N)^T \Sigma_N (\mathcal{L}_N(f) - \widehat{\mathcal{L}}_N) - \log c P(f) \end{aligned} \tag{19}$$

where c is a normalizing constant depending only on N and Σ_N. Comparing (19) to $J_{IDR}(f)$, we conclude that if

$$P(f) = \frac{1}{c}\exp(-\gamma D(f, f_0)) \tag{20}$$

then $J_{IDR}(f, f_0) = J_{map}(f)$. For positive-valued functions f and f_0 (as in images), the functional $D(f, f_0)$ is, in fact, known as a *directed distance*[6] [16]. From this point of view, the probability density function given by (20) is quite analogous to the standard Gaussian density, the difference being that in the Gaussian case, the exponent is basically the L^2 norm of the difference $f - f_0$.

[5] As the error variances in the higher order moments become increasingly large, the information contained in moments beyond some order is quite small. Consequently, using a finite number of moments captures "most" of the information.

[6] Note that $D(\cdot, \cdot)$ is *not* a true metric since $D(f, f_0) \neq D(f_0, f)$.

IV. Solution of the Variational Problem and Its Properties

To make the presentation simpler, we define the vectors $\phi_k(x,y)$ for $k = 0, 1, \cdots, N$ as

$$\phi_k(x,y) = [P_k(x)P_0(y),\ P_{k-1}(x)P_1(y), \cdots, P_0(x)P_k(y)]^T \tag{21}$$

where $P_k(\cdot)$ is the kth-order normalized Legendre polynomial over the interval $[-1, 1]$. In addition, define

$$\Phi_N(x,y) = \left[\phi_0^T(x,y),\ \phi_1^T(x,y),\ \cdots,\ \phi_N^T(x,y)\right]^T. \tag{22}$$

With this notation, and absorbing the factor of $1/2$into γ, the cost functional $J_{IDR}(f)$ can be written as

$$\begin{aligned} J_{IDR}(f, f_0) =& \left(\int\int_{\mathcal{O}} f(x,y)\Phi_N(x,y)dxdy - \widehat{\mathcal{L}}_N\right)^T \\ & \cdot \Sigma_N\left(\int\int_{\mathcal{O}} f(x,y)\Phi_N(x,y)dxdy - \widehat{\mathcal{L}}_N\right) \\ & + \gamma D(f, f_0). \end{aligned} \tag{23}$$

The cost functional $J_{IDR}(f)$ has a unique minimum due to its convex nature [17], [16]. Furthermore, a straightforward variational calculation analogous to ones in other I-Divergence minimization problems [16], [38] (adapted here to deal with the explicit use of estimated moments and the uncertainties in them rather than hard equality or inequality constraints) yields the following implicit specification of f:

$$f(x,y) = f_0(x,y)\exp\left(\frac{-1}{\gamma}\Phi_N^T(x,y)\Sigma_N(\mathcal{L}_N(f) - \widehat{\mathcal{L}}_N)\right). \tag{24}$$

The above is now a nonlinear functional equation in f, which must be solved. (Note that f appears on the right-hand side in the form of the moment functional $\mathcal{L}_N(f)$.) The prior estimate f_0 enters the solution multiplicatively. We shall have more to say later about the choice of this prior.

Due to the form of the solution (24), we may convert (24) into a nonlinear *algebraic* equation in terms of the coefficient vector C_N defined as follows:

$$C_N = \frac{-1}{\gamma}\Sigma_N(\mathcal{L}_N(f) - \widehat{\mathcal{L}}_N). \tag{25}$$

Substituting the expression for $\mathcal{L}_N(f)$ using (24), we obtain an equation in terms of C_N as follows:

$$C_N = \frac{-1}{\gamma}\Sigma_N H(C_N) \tag{26}$$

$$\begin{aligned} & H(C_N) \\ & = \left(\int\int_{\mathcal{O}} f_0(x,y)\exp(\Phi_N^T(x,y)C_N)\Phi_N(x,y)dxdy - \widehat{\mathcal{L}}_N\right). \end{aligned} \tag{27}$$

What we now have is a set of nonlinear, algebraic equations which may be solved by any one of many techniques such as Newton's method or the conjugate-gradient method [5] to yield the unique solution

$$\widehat{f}_{IDR}(x,y) = f_0(x,y)\exp(\Phi_N^T(x,y)\widehat{C}_N). \tag{28}$$

In the experiments reported in this paper, we used a fixed point iteration described in [5] to arrive at the solution of (26).

Despite the seemingly complex nature of the cost functional J_{IDR}, the computation of the coefficient vector $\widehat{C}_N$ involves solving a set of nonlinear algebraic equations. When compared with most variational algorithms that involve the numerical solution of complex partial differential equations, iteratively solving a (relatively small) set of nonlinear algebraic equations makes the IDR approach a relatively computationally attractive one. In addition, note that if f_0 is a positive function of x and y, then the reconstruction $\widehat{f}_{IDR}$ is necessarily a positive function as well. This is clearly desirable since we are dealing with images.

V. Iterative Regularization (It-IDR)

In this section, we present an iterative refinement of the IDR algorithm that is based on redefining the prior. In this formulation, an initial prior is chosen, and using this prior, a solution to the IDR minimization problem is computed. This solution is then used as the prior for a new IDR cost functional and the minimization is carried out again. Therefore, the It-IDR algorithm involves two levels of iteration since at each iteration of this algorithm, a distinct IDR problem is solved iteratively as described in Section IV.

Formally, beginning with $\widehat{f}_0 = f_0$, we can iteratively define

$$\widehat{f}_{k+1} = \arg\min_{\mathrm{f}} \mathrm{J_k}(\mathrm{f}, \widehat{\mathrm{f}}_\mathrm{k}) \tag{29}$$

where the cost function J_k is as in (18) with f_0 replaced by $\widehat{f}_k$ and γ replaced by γ_k.

By appealing to (28), the solution at each k may be written as

$$\widehat{f}_{k+1}(x,y) = \widehat{f}_k(x,y)\exp(\Phi_N^T(x,y)\widehat{C}_N^{(k+1)}) \tag{30}$$

where

$$\widehat{C}_N^{(k+1)} = \frac{-1}{\gamma_k}\Sigma_N(\mathcal{L}_N(\widehat{f}_{k+1}) - \widehat{\mathcal{L}}_N). \tag{31}$$

In terms of $\widehat{C}_N^{(k)}$, we may rewrite this as

$$\begin{aligned} \widehat{C}_N^{(k+1)} &= \frac{-1}{\gamma_k}\Sigma_N \\ & \left(\int\int_{\mathcal{O}} f_0(x,y)\Phi_N(x,y)\exp\left(\Phi_N^T(x,y)\sum_{j=1}^{k+1}\widehat{C}_N^{(j)}\right)dxdy - \widehat{\mathcal{L}}_N\right). \end{aligned} \tag{32}$$

Therefore, at each iteration, as before, an IDR solution is computed by solving an algebraic set of equations for $\widehat{C}_N^{(k+1)}$. There are several appealing features about this iterative approach. The first is that it allows us to control how strictly the estimated moment information is enforced in the final solution both through the sizes of the regularization parameter γ_k (which, as we discuss, may vary with iteration) and through the number of iterations performed. Second, as shown in Appendix B, if (29) is carried to convergence, $\widehat{f}_k$ converges to the solution of the following equality constrained problem

$$\min_f D(f, f_0), \quad \text{subject to} \quad \mathcal{L}_N(f) = \widehat{\mathcal{L}}_N^{(c)} \tag{33}$$

where $\widehat{\mathcal{L}}_N^{(c)}$ denotes the projection, defined with respect to the inner product $< l_1, l_2 >_{\Sigma_N} = l_1^T \Sigma_N l_2$, of $\widehat{\mathcal{L}}_N$ onto the range of the operator Ω_N. Here, Ω_N denotes the operator mapping a square-integrable function $f \in L^2(D)$ with support in the unit disk to its Legendre moments up to order N. Note that if $\widehat{\mathcal{L}}_N$ happens to be in the range $\mathcal{R}a(\Omega_N)$ of the operator Ω_N, the constraint simply becomes $\mathcal{L}_N(f) = \widehat{\mathcal{L}}_N$. If the estimated moments are not *consistent*, i.e., $\widehat{\mathcal{L}}_N \notin \mathcal{R}a(\Omega_N)$, the proposed iterative algorithm implicitly computes and enforces the projection of $\widehat{\mathcal{L}}_N$ onto the set of consistent moments as hard constraints. Hence, iterative regularization provides a method of converting the *soft-constrained* solutions $\widehat{f}_{IDR}$ to *hard-constrained* solutions. The fact that this is done automatically and implicitly is particularly appealing since no explicit description of the set $\mathcal{R}a(\Omega_N)$ is known to exist [35].

The idea of using iterative methods to solve divergence-based minimization problems has been considered in other contexts [6]–[8], [39], [3], [32], [14], [33]. Distinct features of our approach are the applications to tomography[7] using estimated moments and the explicit use of the error covariance matrix for these estimates in forming the penalty function to be minimized. Furthermore, to our knowledge, the specific nature of our iteration (using the finite-dimensional coefficients $\widehat{C}_N^{(k)}$) is also new. In addition, by explicitly taking into account noise, we have a rational mechanism for stopping the iteration based on the fidelity of the moment estimates.

Several results on convergence of iterative algorithms can be found in [6], [39], [3], and [8]. In Appendix C, we provide a convergence result for our specific context that, in particular, provides us with guidance on how the regularization parameter γ_k should be chosen at each iteration. In practice, however, finding the γ_k's according to this result is, in any particular case, a nontrivial task and, in fact, is computationally quite involved. Hence, in the experiments reported here, we have used heuristics based on the result of Appendix C to come up with the regularization parameters. A simple heuristic we found useful, and practical, was to start with a fairly large value for the regularization parameter (200 to 500) and to reduce this value to a relatively small value (5 to 30) after two or three iterations. After this, further reduction of the regularization parameter typically resulted in nonlinear instabilities. To avoid these, the rest of the iterations (to convergence) were carried out with a fixed regularization parameter in the typical range of 5 to 30.

Finally, note that, assuming that γ_k is chosen to ensure convergence, our result states that even if our estimated moments are inconsistent (i.e., they fall outside $\mathcal{R}a(\Omega_N)$), our iterative algorithm produces an estimate with consistent moments satisfying the equality constraints in (33).

VI. Numerical Examples

To demonstrate the potential of the algorithms presented in this paper for improving tomographic reconstruction, we provide simulated reconstructions of sample phantoms. In particular, in this section, we study the performance of the proposed IDR and It-IDR algorithms by applying these techniques in the tomographic reconstruction of two distinct phantoms. In the experiments to follow, we assume that samples of the projections $g(t, \theta)$ of these phantoms are given from m distinct directions in the interval $[0, \pi)$ and that in each direction θ_j, n samples of $g(t, \theta_j)$ are given and that these are corrupted by Gaussian white noise. We denote the data as follows:

$$y(t_i, \theta_j) = g(t_i, \theta_j) + e(t_i, \theta_j) \tag{34}$$

where $e(t_i, \theta_j)$ is a Gaussian white noise sequence with variance σ^2. To quantify the level of noise in relative terms, we define the following signal-to-noise ratio (SNR) *per sample*.

$$\text{SNR (dB)} = 10 \log_{10} \left(\frac{\sum_i \sum_j g^2(t_i, \theta_j)/(m \times n)}{\sigma^2} \right). \tag{35}$$

In addition, to quantify the quality of the reconstructions, we define the *percent* mean-squared-error (% MSE) as follows:[8]

$$\% \text{ MSE} = \frac{\int\int (f - \widehat{f})^2 \, dx \, dy}{\int\int f^2 \, dx \, dy} \times 100\% \tag{36}$$

In the experiments that follow, we define convergence as the point at which the MSE did not improve by more than 1% in three iterations.

Example I: The first phantom to be reconstructed is a 64 by 64 gray-scale image shown in the upper left corner of Fig. 3. Projections were generated from 64 equally spaced angles in $[0, \pi)$, and 64 equally spaced samples were collected in each projection. The projection data were then corrupted by

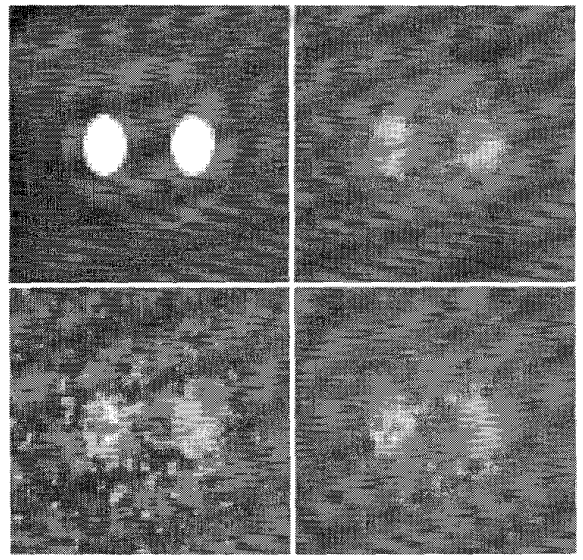

Fig. 3. Counterclockwise from upper left: Phantom, f_0 based on FBP (% MSE=69.1), It-IDR solution after 3 iter. (% MSE=38.1), It-IDR solution after 10 iter. (% MSE=11.1). Data: 64 proj. with 64 samples per proj. SNR = 4.35 dB; moments up to order 8 used.

[7] The problem of (emission) tomographic reconstruction is considered in [3], but with a different setup in which the effects of *measurement noise* are captured via a divergence term, in contrast to our use of it as a direct means of capturing prior information. In addition, no use is made of moment information in [3].

[8] Since there is no universally accepted measure of image quality, and since MSE is often used, we have used this performance measure here to make quantitative comparisons.

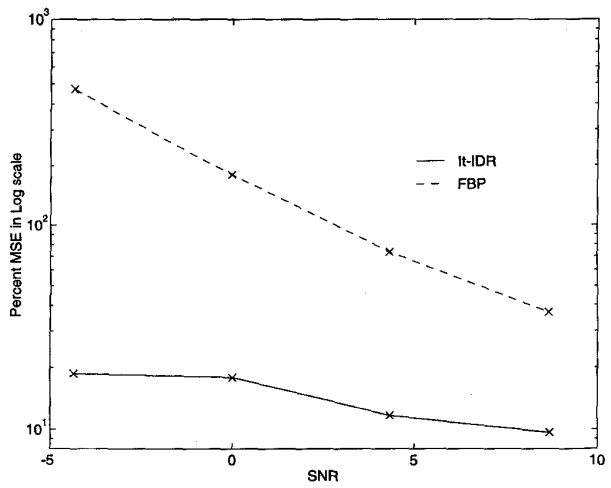

Fig. 4. MSE versus SNR (in decibels) in reconstructing the phantom of Fig. 3: Moments up to order 8 used.

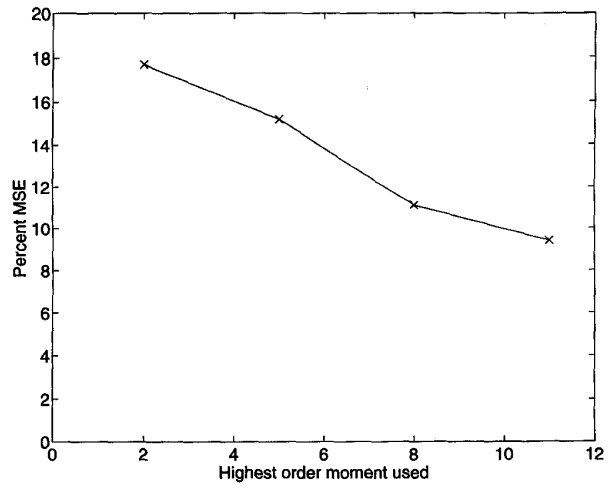

Fig. 5. MSE versus number of moments used in It-IDR reconstruction of the phantom of Fig. 3 with SNR=4.35 dB.

Gaussian white noise to produce an overall SNR of 4.35 dB per sample. In the lower left side of Fig. 3, the FBP reconstruction is shown where a Butterworth filter of order 3 with cut-off frequency of 0.25 (normalized) was used. This choice of filter and cut-off frequency was arrived at to produce the best FBP reconstruction possible, at least from a visual standpoint.

One of the significant features that we wish to demonstrate is that the algorithms we have developed here can significantly enhance noise rejection and feature delineation given an initial estimate f_0 of the underlying image. One obvious choice for that initial estimate is the FBP reconstruction, or rather a slight modification of the FBP solution. In particular, FBP is not guaranteed to produce a positive-valued reconstruction; hence, in order to use the FBP reconstruction as an initial estimate, we add a number to each pixel value in the FBP image in order to maintain positivity. Furthermore, to speed up the convergence of the It-IDR algorithm, we scaled the result to produce an initial estimate with integral equal to the estimated zeroth-order moment.

Using estimated moments up to order 8, the result of the It-IDR algorithm after only three iterations is shown in the lower right-hand side of Fig. 3, whereas the final It-IDR solution (reached after only 10 iterations) is shown in the upper right-hand side of the same figure. A drastic visual improvement in the reconstruction quality is seen both in terms of reduced noise and enhanced feature delineation. In fact, in terms of the % MSE, the improvement is equally striking. The %MSE for the (unnormalized) FBP is roughly 70%, whereas after only three iterations of the It-IDR, this number is reduced to 38.1%, and the final It-IDR reconstruction incurs only 11.1% error. Similar experiments were performed at various SNR's to demonstrate the robustness of and MSE reduction provided by the It-IDR solution to noise. A plot of % MSE versus SNR for the FBP and It-IDR solutions is shown in Fig. 4.

A second issue concerns the order of moments incorporated into the procedure, i.e., the value of N. As we have discussed, the quality of higher order moment estimates decreases rapidly, and thus, we would expect diminishing returns from the

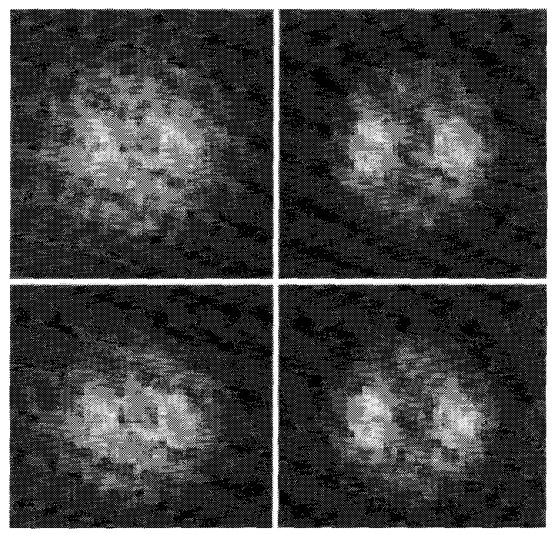

Fig. 6. Counterclockwise from upper left: Reconstructions using moments up to order 2, 5, 8, and 11. Data: 64 projections with 64 samples per projection at SNR=4.35 dB. Initial guess was based on FBP in every case.

inclusion of additional moments. This is illustrated in Fig. 5, which shows the MSE versus the order of the highest order moment used in the reconstructions. Fig. 6 shows the It-IDR reconstructions obtained using moments up to order 2, 5, 8, and 11, respectively, at SNR = 4.35 dB. Note that increasing the order of moments from 8 to 11 reduced the % MSE by only roughly 1%, and additional experiments showed even less improvement if even higher order moments are included. These small gains, however, are only achieved at a significant computational cost. Indeed, note that the number of moments of order k is $k + 1$, and thus, the dimension of $\mathcal{L}_N$ and thus C_N increases considerably as N increases (e.g., from dimension 45 for $N = 8$ to 78 for $N = 11$), increasing the complexity in solving the nonlinear equation (26). To choose the number of moments to be incorporated into the reconstruction process automatically, the minimum description

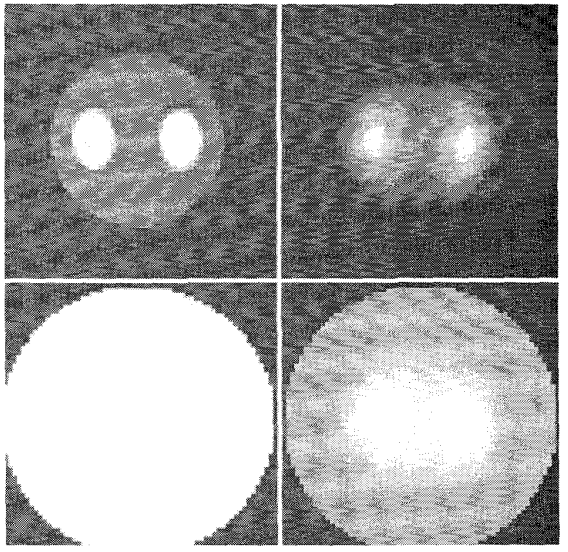

Fig. 7. Counterclockwise from upper left: Phantom, uniform initial estimate (% MSE=65.7), IDR solution (% MSE=55.9), It-IDR solution (% MSE=15.8). Data: 64 projections with 64 samples per projection at SNR=4.35 dB; moments up to order 8 were used.

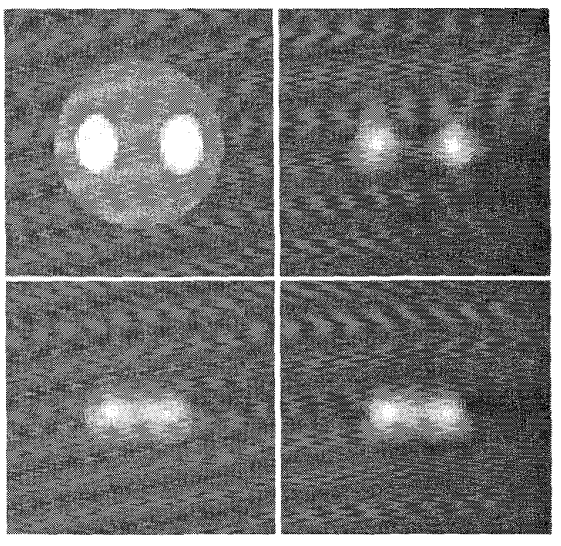

Fig. 8. Counterclockwise from upper left: Phantom, Initialization computed using Burg entropy (% MSE=39.8), IDR solution (% MSE=31), It-IDR solution (% MSE=10.3). Data: 64 projections, 64 samples per projection with SNR=4.35 dB; moments up to order 8 used.

length (MDL) criterion can be considered. The application of MDL would involve the inclusion of a term, involving the number of moments used, to the IDR or It-IDR cost functionals. The analysis of convergence of such an algorithm would, however, be significantly more complicated. We leave this for future research.

To show how the reconstructions change as a function of the choice of prior f_0, we next show the IDR and It-IDR reconstructions when two different priors are used. In Fig. 7, we show the reconstructions when a uniform prior is used. As we have pointed out previously, this corresponds to a

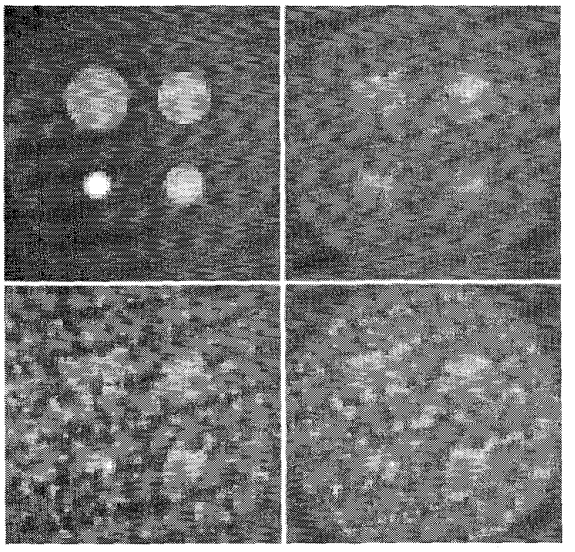

Fig. 9. Counterclockwise from top left: Phantom, Initial Estimate from FBP, IDR reconstruction, Final It-IDR reconstruction (64 views, 64 samples per view, SNR =4.35 dB, moments up to order 10 used).

maximum entropy-type criterion. In particular, in this case, the It-IDR solution to (33) is precisely the classical maximum entropy solution. Estimated moments up to order 8 were used in the reconstructions. As can be seen, the It-IDR reconstruction produces a rough estimate of the underlying image with smooth or "flattened" edge regions. This is essentially due to the fact that the maximum entropy prior seeks the "flattest" reconstruction that matches the data best. Fig. 8 shows the IDR and It-IDR reconstructions when the minimum *Burg entropy* solution is used as the prior and using estimated moments up to order 8. This prior is given by the solution of $f_0 = \arg\min_{\mathrm{f}} \gamma_0 \int\int_{\mathrm{D}} \mathrm{f} - \log(\mathrm{f})\mathrm{dxdy} + (\widehat{\mathcal{L}}_{\mathrm{N}} - \mathcal{L}_{\mathrm{N}}(\mathrm{f}))^{\mathrm{T}} \Sigma_{\mathrm{N}} (\widehat{\mathcal{L}}_{\mathrm{N}} - \mathcal{L}_{\mathrm{N}}(\mathrm{f}))$. As is apparent, in contrast to the maximum (Shannon) entropy solution in Fig. 7, the Burg entropy solution is known to give "peaked" or "spikey" results [16]. It is interesting to contrast the It-IDR solutions in the upper right-hand corners of Figs. 3, 7, and 8 corresponding to our three different choices of f_0. First of all, since the uniform and Burg entropy priors (in the lower left corners of Figs. 7 and 8) do *not* have high-frequency noise, the It-IDR reconstructions in these cases also do not exhibit such noise. This is in contrast to the FBP prior (in the lower left of Fig. 3). On the other hand, because it is far less constrained than the other two priors, the FBP not only exhibits noise but also far more accurate delineation of the features in the image. As a result, the It-IDR reconstruction using the FBP prior has far less distortion in the reconstruction of these figures. On a MSE basis for this example, the It-IDR solution for the Burg prior is slightly superior to that using FBP. However, which of these choices is preferable depends on the application.

Example II: The second phantom to be reconstructed is a 64 by 64 gray-scale image[9] shown in the upper left corner of

[9] The use of only 64-by-64 images here is appropriate for proof of principle. More analysis and experimentation on larger images (512-by-512) will be required to ascertain the performance of the proposed algorithms in many practical situations.

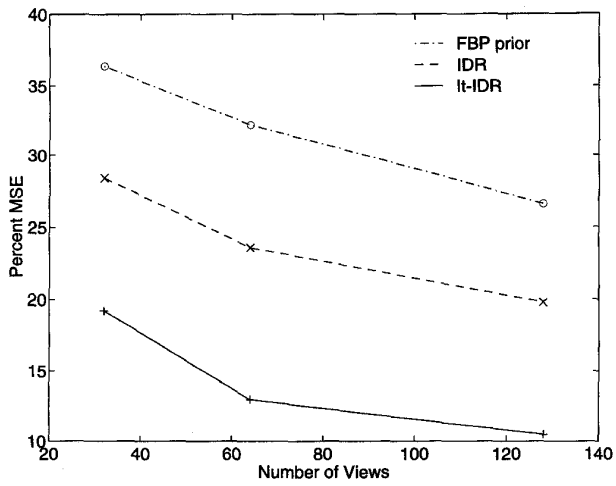

Fig. 10. MSE versus number of views for "adjusted" FBP prior, IDR solution and It-IDR solution, SNR=4.35 dB, moments up to order 10 used.

Fig. 9, which has been chosen to illustrate the capability of the FBP-initialized algorithm to delineate features of differing size and contrast. Projections were generated from 64 equally spaced angles in $[0, \pi)$, and 64 equally spaced samples were collected in each projection. The projection data were then corrupted by Gaussian white noise to produce an overall SNR of 4.35 dB per sample. In the lower left side, the FBP reconstruction is shown where a Butterworth filter of order 2 with cut-off frequency of 0.3 (normalized) was used. After proper normalization, the FBP reconstruction was then used as the initial prior f_0 in the It-IDR reconstruction algorithm. Using estimated moments up to order 10, the result of the It-IDR algorithm after only one iteration (i.e., the IDR solution) is shown in the lower right-hand side of Fig. 9, whereas the final It-IDR solution (reached after only 11 iterations) is shown in the upper right-hand side of the same figure. The drastic visual improvement in the reconstruction quality is again seen.[10]

The It-IDR algorithm performs well even when a much smaller number of projections is available. As shown in Fig. 10, the MSE in the reconstruction using 32 equally spaced views in $[0, \pi)$ at SNR = 4.35 dB is still significantly better than the corresponding MSE value for the normalized FBP reconstruction.

VII. Conclusions

In this paper, we have shown how the tomographic reconstruction problem can be naturally decomposed into a two-step process whereby we first compute ML estimates of the orthogonal moments of the underlying image directly from the projections and then use these estimated moments to obtain a reconstruction of the image. In particular, making a connection to the field of nonparametric probability density estimation, we took advantage of the I-Divergence criterion and its desirable properties to produce regularized reconstructions of images from noisy projection data that far exceed, in quality, those reconstructions produced by classical tomographic reconstruction techniques.

It should be pointed out that the domain of applicability of our proposed algorithms extends to emission tomography as well, where the assumption of Poisson noise statistics in the projections is more appropriate. In fact, the major variation needed here would be to develop statistically robust methods for estimating moments from projections, as well as estimating the uncertainty in those moments. In fact, we can, in general, use the least squares (LS) optimality criterion to estimate the moments whenever the Gaussian noise assumption is inappropriate. When the LS criterion is used, the error covariance for the estimated moments simply coincides with the expression given in the paper. Having these modifications, the rest of our algorithms would remain unchanged.

We demonstrated how our proposed algorithm provides an explicit mechanism for controlling the degrees of freedom in the reconstructions, hence resulting in better results. In addition, in contrast with other divergence- (or entropy-) based algorithms that use the directly measured projection data to form constraints, the use of moments results in a more efficient algorithm since typically, the number of moments needed (and used) is far less than the total number of projection measurements (in our examples, this resulted in a reduction in dimensionality by a factor of roughly 90). Moreover, in our approach, we calculate the error variances in estimating the moments and then make explicit use of this information in our reconstruction algorithm. Furthermore, and perhaps most importantly, we showed how our formulations allow for the explicit incorporation of prior information, in terms of prior estimates of reconstructions, in a very simple way and with minimal increase in computation. In particular, it is worth noting that other geometric information, beyond that used in our examples, can be directly incorporated. For instance, assume that after performing some geometric preprocessing on the data, such as extraction of support information [30], [18] or a preliminary parameterized reconstruction such as polygonal reconstructions [24], an estimate is obtained of the region of the plane where the object of interest may lie (i.e., the spatial support of the object). Then, according to this information, the prior f_0 can be chosen as essentially an indicator function over this estimated region. Due to the multiplicative nature of the solution $\widehat{f}_{IDR}$, the prior f_0 in effect nulls out the part of the reconstruction that the geometric preprocessor eliminated as not being part of the spatial support of the object. This feature of the IDR (and, hence, It-IDR) algorithm is uniquely well suited to situations where it is important to concentrate the reconstruction on a particular region of interest.

Since our proposed algorithms make explicit use of the covariance matrix of the estimated moments, higher order moments, the estimates of which are more inaccurate, are weighed less than lower order ones. Hence, our proposed algorithms essentially make use of a finite and modestly small number of moments to efficiently produce superior reconstructions.

[10] We note that the number of iterations to convergence depends on at least four (related) factors: 1) the noise level, 2) the number of moments used, 3) the choices of regularization parameters, and 4) the choice of prior. In particular, without significantly alterating the regularization regime, and assuming that convergence is not lost, we observed that increasing the noise level tends to increase the number of iterations to convergence. The same is observed when increasing the total number of moments used.

This feature, along with the the overall robustness of the It-IDR algorithm to noise and the number of available views, make it particularly useful for relatively efficient tomographic reconstruction for low signal-to-noise ratio scenarios and when the number of available projections may be small. More extensive testing would be needed in the future to ascertain the performance of the proposed algorithms in practical situations.

Appendix A
Proof of Proposition 1

This result is most easily proved using the nonorthogonal geometric moments

$$H^{(k)}(\theta) = \int_{-1}^{1} g(t,\theta)t^k dt \tag{37}$$

$$\mu_{p,q} = \int_D f(x,y)x^p y^q dxdy. \tag{38}$$

Define

$$\mathcal{H}_N(\theta) = [H^{(0)}(\theta),\ H^{(1)}(\theta),\ \cdots,\ H^{(N)}(\theta)]^T$$
$$\mu^{(k)} = [\mu_{k,0},\ \mu_{k-1,1},\ \cdots,\ \mu_{0,k}]^T$$

and

$$\mathcal{M}_N = [\mu^{(0)^T},\ \mu^{(1)^T},\ \cdots,\ \mu^{(N)^T}]^T$$

. Then, there is a (lower-triangular) invertible relationship between the geometric moments $\mathcal{H}_N(\theta)$ and the Legendre moments $\mathcal{G}_N(\theta)$ and an analogous one between $\mathcal{M}_N$ and $\mathcal{L}_N$. Thus, what we need to show is that given $\mathcal{H}_N(\theta_j)$ for $j = 1,\ 2, \cdots,\ m$, we can uniquely determine $\mathcal{M}_N$ if and only if $N \leq m-1$.

To begin, note that, thanks to (2), there is a *block-diagonal* relationship between the geometric moments of $g(t,\theta)$ and $f(x,y)$, namely

$$H^{(k)}(\theta) = D^{(k)}(\theta)\mu^{(k)} \tag{39}$$

$$D^{(k)}(\theta) = [\rho_{k,0}\cos^k(\theta),\ \rho_{k,1}\cos^{k-1}(\theta)\sin(\theta), \cdots \rho_{k,k-1}\cos(\theta)\sin^{k-1}(\theta),\ \rho_{k,k}\sin^k(\theta)] \tag{40}$$

where $\rho_{k,j} = \frac{k!}{j!(k-j)!}$ are the binomial coefficients. Because the kth-order geometric moment of $g(t,\theta)$ is only a function of the vector of kth-order geometric moments of $f(x,y)$, we need only show that $\mu^{(N)}$ is uniquely determined by $\mathcal{H}^{(N)} = [H^{(N)}(\theta_1),\ H^{(N)}(\theta_2),\ \cdots,\ H^{(N)}(\theta_m)]^T$ if and only if $N \leq m-1$.

Note that $\mathcal{H}^{(N)} = D_N\mu^{(N)}$, where the $m \times (N+1)$ matrix D_N has rows $D^{(N)}(\theta_1),\ D^{(N)}(\theta_2),\ \cdots,\ D^{(N)}(\theta_m)$. Note first that for D_N to have full column rank (equal to $N+1$), we must have $N \leq m-1$. Thus, we must only show that if $N \leq m-1$, then the columns of D_N are linearly independent. From (40), we find that this will be the case if and only if there is no set of α_i (not all zero) such that for $\theta = \theta_1,\ \cdots,\ \theta_m$:

$$p_N(\theta) = \alpha_0\cos^N(\theta) + \alpha_1\cos^{N-1}(\theta)\sin(\theta) + \cdots + \alpha_{N-1}\cos(\theta)\sin^{N-1}(\theta) + \alpha_N\sin^N(\theta) = 0. \tag{41}$$

To see that this cannot happen for any such $p_N(\theta)$ for any N and m satisfying $N \leq m-1$, we proceed by induction on N. Specifically, note first that for $N = 0$, $p_0(\theta) = \alpha_0$, which is nonzero for any nonzero choice of α_0. That is, the result is verified for $N = 0$. Thus, suppose by induction that the result is true for all $N \leq k-1$, where $k \leq m-1$. What we need to show is that it is also true for $N = k$. Therefore, take any nonzero $p_k(\theta)$. Note first that if $p_k(\pi/2) = 0$, then from (41) (with $N = k$), we see that

$$p_k(\pi/2) = \alpha_k\sin^k(\pi/2) = \alpha_k = 0. \tag{42}$$

Therefore, we have that

$$p_k(\theta) = \cos(\theta)(\alpha_0\cos^{k-1}(\theta) + \cdots + \alpha_{k-1}\sin^{k-1}(\theta)) = \cos(\theta)p_{k-1}(\theta). \tag{43}$$

If one of the θ_j, say $\theta_m = \pi/2$, then what we want is that $p_{k-1}(\theta_j)$ cannot vanish for all $j = 1, \cdots, m-1$, but this is exactly verified by part of the induction hypothesis. If none of the $\theta_j = \pi/2$, then we need to ensure that $p_{k-1}(\theta_j)$ cannot vanish for $j = 1, \cdots, m$, which is also part of the induction hypothesis.

Finally, if $p_k(\pi/2) \neq 0$, we can write $p_k(\theta)$ as

$$p_k(\theta) = \cos^k(\theta)q_k(\theta) \tag{44}$$

where

$$q_k(\theta) = \alpha_0 + \alpha_1\tan(\theta) + \cdots + \alpha_{k-1}\tan^{k-1}(\theta) + \alpha_k\tan^k(\theta). \tag{45}$$

Letting $u = \tan(\theta)$, we observe that the right-hand side of (45) is simply a polynomial of order k in u. By the Fundamental Theorem of Algebra [21], this polynomial has at most k real roots. Since $\tan(\theta)$ is one-to-one over the interval $[0, \pi)$, we have that $p_k(\theta_j)$ can vanish for at most k of the m values of θ_j, proving the result.

Appendix B
How It-IDR Solves (33)

In this Appendix, we show that if the It-IDR converges, it indeed solves (33). We first consider the case when $\widehat{\mathcal{L}}_N \in \mathcal{R}a(\Omega_N)$. It is well known [17], [38], [6] that the unique solution to (33) has the form[11]

$$f(x,y) = f_0(x,y)\exp(\Phi_N^T(x,y)K_N) \tag{46}$$

where the vector of constants K_N is chosen such that

$$\mathcal{L}_N(f) = \widehat{\mathcal{L}}_N. \tag{47}$$

In fact, if a function of the form (46) exists and satisfies the constraints given by (47), then it is necessarily the unique solution of (33). Hence, to show that $\widehat{f}_{It-IDR}$ solves (33), it suffices to show that it has the form given by (46) and moments given by (47). From (30), we see that

$$\widehat{f}_{k+1} = f_0\exp(\Phi_N^T(x,y)\sum_{i=1}^{k+1}\widehat{C}_N^{(i)}). \tag{48}$$

Recall that we have assumed that the It-IDR algorithm converges to a finite limit point. Now, through (32), this implies

[11] Note that the existence of the solution is guaranteed by the assumption that $\widehat{\mathcal{L}}_N \in \mathcal{R}a(\Omega_N)$

that the sum $\sum_{i=1}^{k+1} \widehat{C}_N^{(i)}$ also converges as $k \to \infty$. Hence, as $k \to \infty$, in the limit, the It-IDR solution has the same form as (46), with $K_N = \sum_{i=1}^{\infty} \widehat{C}_N^{(i)}$. At the fixed point of (30), the solution $\widehat{f}_{It-IDR}$ satisfies

$$\widehat{f}_{It-IDR}(x,y) = \widehat{f}_{It-IDR}(x,y)\exp(\Phi_N^T(x,y)\widehat{C}_N^{(\infty)}) \quad (49)$$

which, since the elements of the vector $\Phi_N(x,y)$ are linearly independent, implies that $\widehat{C}_N^{(\infty)} = 0$. This, in turn, through (31), implies that

$$\mathcal{L}(\widehat{f}_{It-IDR}) = \widehat{\mathcal{L}}_N. \quad (50)$$

Therefore, $\widehat{f}_{It-IDR}(x,y)$ is the unique solution of (33) in the case $\widehat{\mathcal{L}}_N \in \mathcal{R}a(\Omega_N)$.

If $\widehat{\mathcal{L}}_N$ is not in the range of Ω_N, we simply write $\widehat{\mathcal{L}}_N$ in terms of its orthogonal decomposition with respect to the inner product $< \cdot, \cdot >_{\Sigma_N}$

$$\widehat{\mathcal{L}}_N = \widehat{\mathcal{L}}_N^{(c)} + \widehat{\mathcal{L}}_N^{(i)} \quad (51)$$

where $\widehat{\mathcal{L}}_N^{(c)} \in \mathcal{R}a(\Omega_N)$, and $\widehat{\mathcal{L}}_N^{(i)}$ is orthogonal to $\mathcal{R}a(\Omega_N)$. Then, we may write

$$\widehat{f}_{k+1} = \arg\min_{\mathrm{f}} \gamma_{\mathrm{k}} \mathrm{D}(\mathrm{f}, \widehat{\mathrm{f}}_{\mathrm{k}}) + \|\mathcal{L}_{\mathrm{N}}(\mathrm{f}) - \widehat{\mathcal{L}}_{\mathrm{N}}^{(\mathrm{c})}\|_{\Sigma_{\mathrm{N}}}^2 + \|\widehat{\mathcal{L}}_{\mathrm{N}}^{(\mathrm{i})}\|_{\Sigma_{\mathrm{N}}}^2. \quad (52)$$

Now, clearly, $\mathcal{L}_N(\widehat{f}_k) \in \mathcal{R}a(\Omega_N)$ at every iteration k. Hence, the estimates $\widehat{f}_k$ do not depend on the inconsistent part of the estimated moments $\widehat{\mathcal{L}}_N^{(i)}$, and we may drop the last term on the right-hand side of (52) without changing the solution of the optimization problem (52). This implies that the It-IDR algorithm converges to the solution of (33).

Appendix C
A Convergence Result for It-IDR

We find a sufficient condition for the local asymptotic convergence of the It-IDR algorithm by first assuming that $\widehat{f}_k$ and γ_{k-1} are given for some $k \geq 1$. To solve for $\widehat{f}_{k+1}$, we compute $\widehat{C}_N^{(k+1)}$ by finding the solution of (26). Solving (26) iteratively, we have

$$C_N^{(k+1)}(j+1) = \frac{-1}{\gamma_k}\Sigma_N H(C_N^{(k+1)}(j)). \quad (53)$$

Linearizing $H(C_N^{(k+1)}(j))$ about $C_N = \underline{0}$, we have

$$\widehat{C}_N^{(k+1)}(j+1) \approx \frac{-1}{\gamma_k}\Sigma_N(\mathcal{L}_N(\widehat{f}_k) + D_k\widehat{C}_N^{(k+1)}(j) - \widehat{\mathcal{L}}_N) \quad (54)$$

where $D_k = \int\int_{\mathcal{O}} \widehat{f}_k(x,y)\Phi_N(x,y)\Phi_N^T(x,y)\,dx\,dy$. Hence, for the iteration (53) to be locally asymptotically stable about $C_N = \underline{0}$, it suffices that the eigenvalues of $\frac{1}{\gamma_k}\Sigma_N D_k$ have magnitude strictly less than one [20]. That is

$$\textbf{Condition 1:} \quad \left|\lambda\left(\frac{1}{\gamma_k}\Sigma_N D_k\right)\right| < 1. \quad (55)$$

If Condition 1 is satisfied, then from (54), we have a linear approximation to $C_N^{(k+1)}$ given by

$$\widehat{C}_N^{(k+1)} \approx (\gamma_k I + \Sigma_N D_k)^{-1}\Sigma_N(\widehat{\mathcal{L}}_N - \mathcal{L}_N(\widehat{f}_k)). \quad (56)$$

From the *definition* of $C_N^{(k)}$ from (31), we have that $\gamma_{k-1}\widehat{C}_N^{(k)} = \Sigma_N(\widehat{\mathcal{L}}_N - \mathcal{L}_N(\widehat{f}_k))$, which after substitution in (56) yields

$$\widehat{C}_N^{(k+1)} = \gamma_{k-1}(\gamma_k I + \Sigma_N D_k)^{-1}\widehat{C}_N^{(k)} = T_k\widehat{C}_N^{(k)}. \quad (57)$$

A sufficient condition [20] for the asymptotic convergence of (57) is that for all $k \geq 1$, some $c \geq 0$, and some $0 \leq \delta < 1$

$$\textbf{Condition 2:} \quad \|\prod_{j=1}^{k} T_j\|_2 \leq c\delta^k. \quad (58)$$

Therefore, by carefully choosing γ_k to satisfy conditions 1 and 2 simultaneously at each iteration, the overall It-IDR algorithm can be made locally asymptotically convergent.

References

[1] N. I. Akhiezer, *The Classical Moment Problem and Some Related Questions in Analysis*. New York: Hafner, 1965.

[2] U. Amato and W. Hughes, "Maximum entropy regularization of Fredholm integral equations of the first kind," *Inverse Problems*, vol. 7, pp. 793–808, 1991.

[3] C. L. Byrne, "Iterative image reconstruction algorithms based on cross-entropy minimization," *IEEE Trans. Image Processing*, vol. 2, no. 1, pp. 96–103, Jan. 1993.

[4] A. T. Chinwalla and J. A. O'Sullivan, "Image regularization using a divergence penalty method," in *Proc. Johns Hopkins Conf. Inform. Sci. Syst.*, Baltimore, MD, Mar. 1993, pp. 30-33.

[5] S. Conte, *Elementary Numerical Analysis*. New York: McGraw-Hill, 1965.

[6] I. Csiszár, "I-Divergence geometry of probability distributions and minimization problems," *Ann. Probab.*, vol. 3, pp. 146–158, 1975.

[7] ——, "A geometric interpretation of Darroch and Ratcliff's generalized iterative scaling," *Ann. Stat.*, vol. 17, no. 3, pp. 1409–1413, 1989.

[8] J. N. Darroch and D. Ratcliff, "Generalized iterative scaling for log-linear models," *Ann. Math. Stat.*, vol. 43, no. 5, pp. 1470–1480, 1972.

[9] N. J. Dusaussoy and I. E. Abdou, "The extended MENT algorithm: A maximum entropy type algorithm using prior knowledge for computerized tomography," *IEEE Trans. Signal Processing*, vol. 39, no. 5, pp. 1164–1180, May 1991.

[10] M. Ein-Gal, "The shadow transformation: an approach to cross-sectional imaging," Ph.D. dissertation, Dept. of Elect. Eng., Stanford Univ., Stanford, CA, 1974.

[11] S. F. Gull and G. J. Daniell, "Image reconstruction from incomplete and noisy data," *Nature*, vol. 272, pp. 686–690, Apr. 1978.

[12] S. Helgason, Radon Transform. Boston, MA: Birkhauser, 1980.

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

[14] S. Holte, P. Schmidlin, A Linden, G. Rosenqvist, and L. Eriksson, "Iterative image reconstruction for positron emission tomography: A study of convergence and quantitation problems," *IEEE Trans. Nucl. Sci.*, vol. 37, no. 2, pp. 629–635, Apr. 1990.

[15] E. T. Jaynes, "On the rationale of maximum entropy methods," *Proc. IEEE*, vol. 70, no. 9, pp. 939–952, Sept 1982.

[16] L. K. Jones and C. L. Byrne, "General entropy criteria for inverse problems, with applications to data compression, pattern classification, and cluster analysis," *IEEE Trans. Inform. Theory*, vol. 36, no. 1, pp. 23–30, 1990.

[17] S. Kullback, *Information Theory and Statistics*. New York: Wiley, 1959.

[18] A. Lele, "Convex set estimation from support line measurements," Master's thesis, Dep. Elect. Eng., Mass. Inst. of Technol., Cambridge, 1990.

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

[20] D. G. Luenberger, *Introduction to Dynamic Systems*. New York: Wiley, 1979.

[21] S. MacLane and G. Birkhoff, *Algebra*. New York: Chelsea, 1988.

[22] P. Milanfar, "Geometric estimation and reconstruction from tomographic data," Ph.D. dissertation, Dept. of Elect. Eng., Mass. Inst. Technol., Cambridge, June 1993.

[23] P. Milanfar, W. C. Karl, and A. S. Willsky, "Recovering the moments of a function from its Radon-transform projections: Necessary and sufficient conditions," LIDS Tech. Rep. LIDS-P-2113, Mass. Inst. Technol., Lab. for Inform. Decision Syst., June 1992.
[24] ——, "Reconstructing finitely parameterized objects from projections: A statistical view," *CVGIP: Graphical Models and Image Processing*, vol. 56, no. 5, pp. 371–391, Sept. 1994.
[25] G. Minerbo, "MENT: A maximum entropy algorithm for reconstructing a source from projection data," *Comput. Graphics Image Processing*, vol. 10, pp. 48–68, 1979.
[26] F. Natterer, "Regularization techniques in medical imaging," in *Mathematics and Computer Science in Medical Imaging*, vol. F39 of NATO ASI Series. New York: Springer-Verlag, 1988, pp. 127–141.
[27] M. Pawlak, "On the reconstruction aspects of moments descriptors," *IEEE Trans. Inform. Theory*, vol. 38, no. 6, pp. 1698–1708, Nov. 1992.
[28] A. Peres, "Tomographic reconstruction from limited angular data," *J. Comput. Assisted Tomography*, vol. 3, no. 6, pp. 800–803, 1979.
[29] J. L. Prince and A. S. Willsky, "Constrained sinogram restoration for limited-angle tomography," *Opt. Eng.*, vol. 29, no. 5, pp. 535–544, May 1990.
[30] ——, "Reconstructing convex sets from support line measurements," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 12, no. 4, pp. 377–389, 1990.
[31] M. L. Reis and N. C. Roberty, "Maximum entropy algorithms for image reconstruction from projections," *Inverse Problems*, vol. 8, pp. 623–644, 1992.
[32] W. H. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Amer.*, vol. 62, pp. 55–59, Jan. 1972.
[33] P. Schmidlin, "Iterative separation of sections in tomographic scintigrams," *Nuclear Med.*, vol. 15, no. 1, pp. 1–16, 1972.
[34] M. I. Sezan and H. Stark, "Incorporation of a priori moment information into signal recovery and synthesis problems," *J. Math. Anal. Appl.*, vol. 122, pp. 172–186, 1987.
[35] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*. New York: Amer. Math. Soc., 1943.
[36] J. E. Shore, "Minimum cross-entropy spectral analysis," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-29, no. 2, pp. 230–237, Apr. 1981.
[37] J. E. Shore and R. M. Gray, "Minimum cross-entropy pattern classification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-4, no. 1, pp. 11–17, Jan. 1982.
[38] J. E. Shore and R. W. Johnson, "Properties of cross-entropy minimization," *IEEE Trans. Inform. Theory*, vol. IT-27, no. 4, pp. 472–482, July 1981.
[39] D. L. Snyder, T. J. Shulz, and J. A. O'Sullivan, "Deblurring subject to nonnegativity constraints," *IEEE Trans. Signal Processing*, vol. 40, no. 5, pp. 1143–1150, May 1992.
[40] G. Talenti, "Recovering a function from a finite number of moments," *Inverse Problems*, vol. 3, pp. 501–517, 1987.
[41] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*. Washington, DC: Winston/Wiley, 1977.
[42] S. J. Wernecke and L. D'Addario, "Maximum entropy image reconstruction," *IEEE Trans. Comput.*, vol. C-26, no. 4, pp. 351–364, Apr. 1977.

Peyman Milanfar (S'90–M'93) received the B.S. degree in engineering mathematics from the University of California at Berkeley, in 1988, and the S.M., E.E., and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1990, 1992, and 1993, respectively.

In 1993, he joined Alphatech, Inc. as a member of the technical staff, where he conducted research in multiresolution image processing and compression and over-the-horizon radar signal processing. Since July 1994, he has been with SRI International's Applied Electromagnetics and Optics Laboratory, Menlo Park, CA, where his current research interests are in statistical signal and image processing and optimal estimation.

Dr. Milanfar is a member of Sigma Xi and the Mathematical Association of America.

William C. Karl (M'91) received the Ph.D. degree in electrical engineering and computer science in 1991 from the Massachusetts Institute of Technology, Cambridge, where he also received the S.M., E.E., and S.B. degrees.

He has held the position of Staff Research Scientist with the Brown–Harvard–M.I.T. Center for Intelligent Control Systems and the M.I.T. Laboratory for Information and Decision Systems from 1992 to 1994. He joined the faculty of Boston University where he is currently Assistant professor of Electrical, Computer, and Systems Engineering. Since January 1996 he also held a joint appointment in the Department of Biomedical Engineering. In 1993 he was the organizer and chair of the "Geometry and Estimation" session of the Conference on Information Sciences and Systems at Johns Hopkins University. In 1994 he was on the the technical committee for the Workshop on Wavelets in Medicine and Biology, part of Internation Conference of the IEEE Engineering in Medicine and Biology Society. He is special guest editor of the 1977 special issue of the *International J. Pattern Recognition, and Artificial Intelligence* on "Processing, Analysis, and Understanding of MR Images of the Human Brain." he is also associate editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING in the areas of tomography and MRI. His research interests are in the areas of multidimensional and multiscale signal and image processing and estimation, geometrical estimation, and medical signal and image processing.

Alan S. Willsky (S'70–M'73–SM'82–F'86) received both the S.B. and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, USA, in 1969 and 1973, respectively.

He joined the MIT faculty in 1973, and his present position is Professor of Electrical Engineering. From 1974 to 1981, he served as Assistant Director of the MIT Laboratory for Information and Decision Systems. He is also a founder and member of the Board of Directors of Alphatech, Inc. His 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.

In 1975, Dr. Willsky received the Donald P. Eckman Award from the American Automatic Control Council. He has held visiting positions at Imperial College, London, L'Université de Paris-Sud, and the Institut de Recherche en Informatique et Systèmes Aléatoires in Rennes, France. He was program chairman for the 17th IEEE Conference on Decision and Control, 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. In addition, in 1988, he was made a Distinguished Member of the IEEE Control Systems Society. He has also given several plenary lectures at major scientific meetings including the *20th IEEE Conference on Decision and Control, the 1991 IEEE International Conference on Systems Engineering, the SIAM Conference on Applied Linear Algebra* in 1991, and the *1992 Inaugural Workshop for the National Centre for Robust and Adaptive Systems*, Canberra, Australia. He is the author of the research monograph "Digital Signal Processing and Control and Estimation Theory" and is coauthor of the undergraduate text *Signals and Systems*. 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.